UNIVERSITÀ DEGLI STUDI DI MILANO STATALE

PhD School in Molecular and Cellular Biology – XXXVI cycle

Dipartimento di Bioscienze

Investigating Genome Protection at G-Quadruplexes: Structural-Functional Analysis of *S. cerevisiae* Domesticated Transposase Vid22

BIO/11 Molecular Biology

Author: Giulia Maria BERNINI (R12993)

Supervisor: Prof. Federico Lazzaro, Prof. Marco Muzi Falconi

PhD Coordinator: Prof. Roberto Mantovani

Academic year 2022-2023

Index

1 Sommario

L'instabilità del genoma è una caratteristica distintiva dell'invecchiamento e di varie patologie umane, tra cui il cancro. L'instabilità genomica è provocata da agenti esogeni che danneggiano il DNA e da molteplici fattori endogeni, tra cui le G-tetradi, strutture secondarie non canoniche del DNA. Queste strutture eterogenee sono state implicate nella modulazione di numerosi processi biologici, inclusi la trascrizione e la replicazione. Di conseguenza, la perturbazione dell'omeostasi delle G-tetradi è associata ad alterazioni nell'espressione genica, a stress replicativo, a rotture del DNA e, in ultima analisi, a un aumento dell'instabilità del genoma. Abbiamo recentemente scoperto che la proteina Vid22 di *Saccharomyces cerevisiae*, un presunto membro della famiglia hAT di DNA trasposasi in base alla predizione di un motivo a dito di zinco di tipo BED e un ripiegamento simile alla ribonucleasi H, lega alcuni siti formanti G-tetradi e contrasta i loro effetti negativi sulla stabilità del genoma. In questo lavoro, ho mutato vari residui della proteina Vid22 o la sequenza di DNA di uno dei suoi siti di legame nel genoma per investigare i requisiti intrinseci ed estrinseci per il suo reclutamento alla cromatina. L'arricchimento di Vid22 in certi loci dipende dalla presenza di una G-tetrade e contribuisce al legame al DNA di Tbf1, un fattore regolativo generale multifunzionale. Inoltre, il ruolo di Vid22 nella protezione del genoma è correlato alla sua interazione fisica con Tbf1. La distorsione del motivo a dito di zinco di tipo BED, putativo dominio di legame al DNA, affligge solo parzialmente l'interazione di Vid22 con il DNA e Tbf1, mentre la mutazione di residui C-terminali abolisce entrambi i processi e correla con ipersensibilità cellulare a stress genotossici. Inoltre, Vid22 forma omodimeri *in vivo*. Il modello computazionale e la struttura a bassa risoluzione del dimero di Vid22 mostrano somiglianze sorprendenti con quello formato da Hermes, trasposasi attiva della famiglia hAT, nonostante la loro distanza evolutiva e la divergenza di sequenza. Analogamente a Hermes, il C-terminale di Vid22 sembra essere cruciale nel mantenimento della corretta struttura terziaria della proteina. La mutazione del C-terminale abolisce anche l'interazione Vid22-Vid22 a causa della sua prossimità spaziale a una seconda regione putativa di legame al DNA direttamente coinvolta nella dimerizzazione (residui 170-210), inferita per omologia con Hermes, la cui

esistenza è corroborata dalla compromissione dell'auto-interazione nel mutante vid22 F195A. Pertanto, Vid22 origina evolutivamente dalla domesticazione di una DNA trasposasi della famiglia hAT a supporto della stabilità del genoma e della regolazione delle G-tetradi nel lievito gemmante. Poiché anche il genoma umano codifica proteine poco caratterizzate che sono predette possedere un'organizzazione di domini e un'origine evolutiva simili, la ricerca su Vid22 potrebbe gettare le basi per il loro studio, oltre a fornire nuove informazioni sul riconoscimento e sulla biologia delle G-tetradi.

2 Abstract

Genome instability is a characteristic feature of ageing and several human pathologies, including cancer. Genome instability is elicited by exogenous DNA damaging agents and by multiple endogenous factors, among which G-quadruplex non canonical DNA secondary structures. These structurally heterogeneous nucleic acid folds have been implicated in the modulation of multiple biological processes, including transcription and replication. Consequently, perturbation of G-quadruplex homeostasis is linked to alterations in gene expression, replication stress, DNA breakage, and, ultimately, enhanced genome instability. We recently found that *Saccharomyces cerevisiae* Vid22, putative member of the hAT family of DNA transposases by prediction of a BED-type zinc finger domain and a RNase H fold, binds at G-quadruplex forming sites and counteracts their detrimental effects on genome stability. In this work, I mutated various residues of Vid22 protein or the DNA sequence of one of its binding loci in the genome to investigate the intrinsic and extrinsic requirements for its recruitment to chromatin. Vid22 enrichment at certain loci depends on the presence of a G-quadruplex and contributes to the binding of Tbf1, a multifunctional general regulatory factor, to DNA. Moreover, Vid22 role in genome protection correlates with its physical interaction with Tbf1. Disruption of the putative DNA-binding BED domain only partially affects Vid22 interaction with DNA and Tbf1, whereas mutation of C-terminal residues abolishes both processes and correlates with cell hypersensitivity to genotoxic stress. Moreover, Vid22 forms homodimers *in vivo*. The computational model and low-resolution structure of Vid22 dimer display striking similarities to that formed by the active hAT Hermes transposase dimer despite their evolutionary distance and sequence divergence. Similarly to Hermes, Vid22 C-terminus seems to be crucial in maintaining the proper protein tertiary structure. Mutation of the C-terminus also abolishes Vid22 self-interaction due to its spatial proximity to a second putative DNA-binding region directly involved in dimerisation (residues 170-210) inferred by homology to Hermes, which existence is corroborated by the impairment of self-interaction in the vid22 F195A mutant. Thus, Vid22 has evolved from domestication of a hAT DNA transposase to support genome stability and G-quadruplex regulation

in budding yeast. Because the human genome also encodes uncharacterised proteins predicted to possess similar domain organisation and evolutionary origin, research on Vid22 may pave the way for their study, besides providing novel insights into G-quadruplex recognition and biology.

3 Aim of the Project

Genomic instability, denoting the occurrence of a wide spectrum of mutations in the genetic material of a cell, entails physiological alterations driving ageing and multiple diseases in humans. An endogenous threat to genome integrity is represented by G-quadruplexes, secondary structures formed at numerous naturally occurring G-rich nucleic acid sequences. G-quadruplexes have multiple biological roles, so they cause detrimental effects linked to pathogenesis when their homeostasis is perturbed. Specific spatio-temporal regulation of these non-canonical architectures is accomplished by G-quadruplex interacting proteins, which can act as structural stabilisers/destabilisers or effectors of G-quadruplex functions. The characterisation of such proteins is, therefore, of fundamental importance to clarify the physiological and pathological mechanisms associated with these structures and develop effective therapeutic strategies. The study of G-quadruplex regulation and the mechanisms preserving genome integrity, as well as their relevance to human health, is greatly facilitated by the use of *Saccharomyces cerevisiae* as a model organism. The attractiveness of employing budding yeast for these purposes is due to its genetic tractability, rapid growth, and the conservation of key factors involved in fundamental cellular processes in higher eukaryotes, among other advantages. Multiple genetic screenings performed in budding yeast identified *VID22* as a factor involved in genome stability maintenance, as cells experience endogenous DNA damage during unperturbed growth and accumulate genetic aberrations in its absence. Vid22 is reported to contribute to the regulation of chromatin organization both at promoters, in the presence of its two validated physical interactors Tbf1 and Env11, and at double-strand DNA break sites to facilitate repair, exclusively with Tbf1. Moreover, we recently showed that Vid22 binds DNA G-quadruplexes and counteracts their potential to induce genome instability. However, this factor has not yet been fully characterised at the biochemical and functional levels. Therefore, the main objective of this work was to validate Vid22 recruitment at G-quadruplex DNA *in vivo*, exploring the relevance of its computationally predicted DNA binding domain and its potential dependence on other players in this context, with a particular focus on Tbf1. Our research also aims

to clarify the molecular mechanisms underlying Vid22 role in genome protection at G-quadruplexes and its broader cellular effects. To do so, I adopted computational, genetic, and molecular approaches involving both random mutagenesis screenings and targeted mutation of Vid22 residues conserved among its homologues, chromatin immunoprecipitation experiments coupled with quantitative PCR analysis to address protein recruitment to particular DNA sequences, as well as co-immunoprecipitation assays to analyse protein-protein interactions. These phenotypes were correlated to genome stability through cellular sensitivity assays to genotoxic agents and interpreted through *in silico* structural analyses.

4 Introduction

4.1 Overview of genome instability

Genome instability reflects the incidence of mutations in any given genome over time. Mutations are an important aspect in evolution, as they represent a source of variability and are subjected to natural selection[1,2]. Such genetic alterations can manifest themselves in different forms and magnitudes[3] (**Figure I-1 D**). For instance, point mutations are the result of the substitution (*i.e.* transitions or transversions), addition, or deletion of a single nucleotide[3]. Small insertions/deletions, collectively referred to as indels, are represented by the expansion or absence of short sequences[3]. Gross chromosomal rearrangements, instead, comprise duplication or loss of extensive portions of chromosomes, as well as translocations and inversions[3]. Finally, aneuploidies define changes in the number of whole chromosomes[3] (**Figure I-1 D**). At the cellular level, the mutational load can lead to unfaithful transmission of the genetic information, altered metabolism, abnormal growth, and, ultimately, impaired viability (**Figure I-1 E**). Genome instability characterises ageing[1,4–6] and numerous human pathological disorders, including most cancers[2,7–11] and rare genetic diseases, among which Seckel syndrome, Ataxia telangiectasia, Nijmegen breakage syndrome, Fanconi anemia, Xeroderma pigmentosum, Bloom and Werner syndromes, to name a few[1,12–14] (**Figure I-1 E**). In fact, genetic instability is not only a feature of tumorigenesis, but also a causative event[8,11,15–18]. Consequently, the research of both the factors that increase genome instability and the cellular mechanisms aimed at its prevention or suppression is of paramount importance for the rational design of effective therapeutic strategies[19,20].

The integrity of the genome is constantly challenged by a multitude of environmental factors and endogenous sources, which collectively induce different DNA lesions (**Figure I-1 A-B**). Such exogenous and intrinsic agents can induce chemical modifications to nucleotides (*e.g.* isomerisation, depurination, depyrimidination, deamination, oxidation, alkylation, and formation of covalent bonds with chemically active chemicals, **Figure I-1 B**).

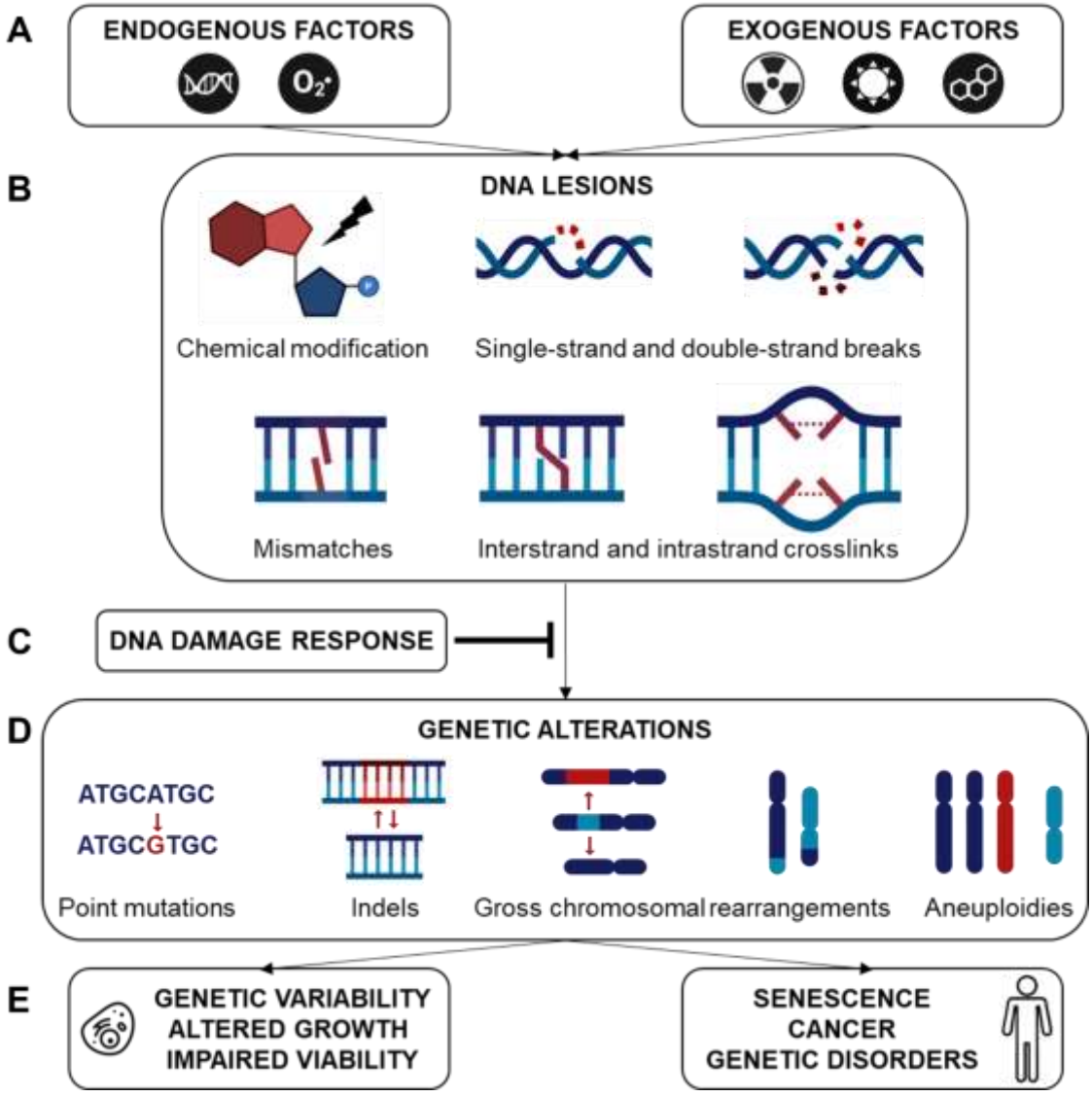


Figure I-1 – Genome instability is triggered by DNA lesions and counteracted by the DNA damage response

Graphic summary of the sources and types of DNA lesions, of the genetic alterations characterising genome instability and their consequences. **(A)** Organisms are continuously exposed to a variety of endogenous (*e.g.* replication problems, reactive oxygen species) and exogenous (*e.g.* ionising radiation, ultraviolet light, several chemicals) factors that can threaten the integrity of the genome. **(B)** Endogenous and exogenous factors can collectively induce different DNA lesions (*e.g*. chemical modification to nucleotides, single-strand and double-strand breaks, intrastrand and interstrand crosslinks). **(C)** The DNA damage response, comprising multiple molecular pathways (*e.g.* DNA damage checkpoint, tolerance, and repair mechanisms), allows cells to perceive, bypass, or resolve DNA lesions, ultimately trying to prevent damaged DNA from being converted into fixed mutations. **(D)**

If the DNA damage response systems fail to properly counteract DNA lesions, then these might be converted into a wide array of genetic alterations (*e.g.* point mutations, indels, gross chromosomal rearrangements, aneuploidies). Genome instability refers to the accumulation of such genetic alterations. **(E)** At the cellular level, the mutational load can lead to genetic variability, altered growth and impaired viability. In humans, genome instability is associated with senescence, cancer, and several genetic disorders.

In addition, exposure to these external and internal sources of DNA damage can lead to establishment of covalent bonds between bases within the same DNA filament or between opposite strands (*i.e.* intrastrand or interstrand crosslinking), and to ruptures of the phosphate backbone involving a single or both DNA chains (*i.e.* single-strand and double-strand breaks, **Figure I-1 B**)[13,21,22]. Extrinsic genotoxins include both physical sources, such as ionising radiation and ultraviolet light, and a plethora of chemical compounds[13,22] (**Figure I-1 A**). DNA can be damaged also by physiological products, like reactive oxygen species, and during the normal DNA dynamics, among which replication[13,22] (**Figure I-1 A**). Indeed, replication stress is among the major endogenous sources of genomic instability[18,23–26].

Replication stress defines the condition in which the progression of the replication forks is decelerated or stalled. This situation arises from a variety of conditions and barriers. One of the most common causes of replication stress is the restriction in the supply of deoxyribonucleoside triphosphate (dNTPs) substrates for DNA polymerization during replication or repair processes[1,23,27–29]. The availability of dNTPs can be affected, for instance, by treatment with a widely used drug, hydroxyurea (HU)[30]. This small molecule primarily targets the ribonucleotide reductase enzyme of various organisms, reversibly inhibiting the transfer of electrons during the reduction of ribonucleoside diphosphates to their respective deoxyribonucleotides[31,32]. Besides dNTPs shortage, also the occurrence of hotspots on the template acting in *cis* as obstacles during DNA synthesis, and dysfunctions in *trans*-acting elements mediating genome dynamics are linked to the onset of replication stress[1,23,27–29]. *Cis*-acting elements that can hinder the movement of the replication forks are heterogeneous, comprising not only the abovementioned DNA lesions, but also ribonucleotides embedded in the DNA filament[33–36], as well as non-

nucleosomal proteins tightly bound to DNA[37–39]. Moreover, replication stress is observed at highly transcribed loci due to both clashes of the replicative and transcriptional molecular machineries and the formation of R-loops, DNA:RNA hybrids resulting from the annealing of the transcript to its DNA template[40–47]. Additional hindering hotspots are repetitive sequences, which are prone to instability despite their length[48,49], and non-canonical DNA structures, including Z-DNA, triplex or H-DNA, hairpins, cruciforms, i-motifs, and, most relevant for this work, G-quadruplexes[50–55], which will be described in greater detail in the following section. Finally, several proteins participating in multiple biological processes act in *trans* to suppress replication stress and genomic instability. In fact, the stability of a genome depends on the proper functioning of: (i) proteins involved in replication (*e.g.* DNA polymerases, helicases, nucleases, topoisomerases, and ligases), (ii) chromatin factors (*e.g.* histone chaperones and modifiers, nucleosome remodelers), (iii) cohesins and players involved in the structural maintenance of chromosomes, (iv) proteins mediating mRNA biogenesis, processing, and export, (v) checkpoint kinases and mediators, as well as (vi) factors involved in DNA damage tolerance and repair[1,23,28,56].

Checkpoint, tolerance and repair factors are particularly beneficial for upholding genome integrity as they can prevent replication stress or address the harms it may entail. Indeed, these factors are considered to be the main players of the DNA damage response because their combined activity allows cells to specifically perceive, bypass and resolve the different kinds of lesions that are continuously formed in DNA[13,22,57,58] (**Figure I-1 C**). Among the mechanisms involved in DNA damage repair, which rely on the concerted activity of various proteins, several cell cycle phase-specific pathways exist for repairing DNA double-strand breaks, with error-free homologous recombination and error-prone non-homologous end joining being the primary ones, whereas single-strand breaks are mended through a dedicated pathway[13,22,57]. Nucleotide excision repair addresses bulky DNA adducts capable of distorting the double helix, such as UV photoproducts and intrastrand crosslinks, for the latter of which alternative repair mechanisms also exist[13,22,57]. Mismatch repair proteins correct mispaired nucleotides, and base excision repair factors are dedicated to resolve small chemical alterations in DNA bases[13,22,57].

Besides the above mentioned repair pathways, cells also possess additional systems to temporarily bypass unrepaired DNA lesions during replication, including translesion DNA synthesis performed by peculiar polymerases or template-switching[59–61].

Replication is a fundamental process for life propagation, yet it's inherently vulnerable to complications and aberrations. Cells have thus evolved a network of specific molecular pathways, including checkpoint factors and specialized enzymatic activities, to perceive the insurgence of replication problems, stabilise perturbed replication forks, and assist in their restart (*e.g.* by fork repriming, reversal, degradation and backtracking)[27,62–67]. Undoubtedly, the main pathways that protect the genome from insults resulting from DNA lesions, as well as by challenges posed by DNA replication, are represented by DNA damage checkpoint signal transduction cascades. DNA damage checkpoint proteins monitor the status of the genome, and, in response to replication stress or DNA lesions, they coordinate various metabolic processes—like (i) local changes in the chromatin structure, (ii) expression and activity of DNA repair factors, (iii) replisome stabilisation and restart, and (iv) regulation of dNTP levels—with cell cycle progression and origin firing[13,68–70]. Sensor proteins recognise the DNA structures that form as byproducts of DNA damage or replication stress, including single-strand DNA and DNA breaks. Then, evolutionarily conserved apical phosphatidylinositol 3-kinase-related kinases (PIKKs, like Mec1 and Tel1 in budding yeast, and ATM, ATR, and DNA-PKcs in mammals) are activated and phosphorylate a number of protein substrates, like adaptors (yeast Mrc1 and Rad9 and human Claspin, MDC1 and 53BP1) and downstream effector kinases (Rad53, Dun1 and Chk1 in *S. cerevisiae*, CHK1 and CHK2 in *H. sapiens*), effectively amplifying the signal[13,68–74].

Nonetheless, when such mechanisms fail to properly rescue replication defects, DNA polymerases may stall persistently, leading to desynchronization of the DNA unwinding and synthesis processes and, thus, to the hazardous exposure of long stretches of single-stranded DNA[23,28,67,75]. Blocked forks may also regress, causing the generation of a four-way DNA junction or cruciform, or, in the worst case, collapse, leading to the formation of a DNA double-strand break[23,28,67,75]. Single-stranded DNA, regressed forks, and DNA double-strand breaks represent the

prevalent intermediates formed during replication stress that have the potential to compromise genome integrity, for instance in cases when their sensing and proper resolution are defective or when they are aberrantly processed by molecular pathways not dedicated to their management[1,23]. The identity of the intermediate generated after the primary perturbation, the nature of the mechanism processing such intermediate, and the phase of the cell cycle during which these events take place, are among the factors that influence the specific types of genetic mutations resulting from the failure to revert replication stress[1,23].

In conclusion, it is evident that the mechanisms influencing genome stability are complex and heterogeneous, warranting continuous investigation given their significant implications for human health.

4.2 The yin and yang of G-quadruplex structures

As anticipated in the previous section, among the endogenous factors that can potentially elicit DNA lesions and genomic instability, G-quadruplexes, also known as G-tetraplexes (G4s), have aroused increasing interest from the scientific community in the last decades[50,53,76–78]. G-quadruplexes are a peculiar class of structurally heterogeneous secondary structures that can be formed by diverse G-rich DNA or RNA motives. G-quadruplex-forming motives are usually characterised by four stretches of at least two guanines, separated by loops of variable length, generalisable in the $G_{\geq 2}N_XG_{\geq 2}N_XG_{\geq 2}N_XG_{\geq 2}$ consensus[78–80] (**Figure I-2 A**). The original concept that only canonical tetraplex architectures exist, formed at sequences following the $G_3N_{1\text{-}7}G_3N_{1\text{-}7}G_3N_{1\text{-}7}G_3$ rule (*i.e.* with three guanines per each G-tract and with a maximum loop length of seven nucleotides, **Figure I-2 A**), has been surpassed thanks to accumulating evidence of the presence *in vivo* of a variety of non-canonical structures. Such non-canonical G-quadruplexes (*e.g.* structures with long loops, bulges in the G-tracts, or vacant guanine spots filled thanks to snapback loops)[81,82] escape this conservative consensus motif which has been used by most algorithm for the genome-wide prediction of the G-quadruplex forming potential[83], thus likely leading to an underestimation of the actual count of these formations in the genome. The G-quadruplex assembly results from the piling up, via π-π interactions, of two or more planar tetradic arrangements, named G-quartets or G-tetrads, defined by the interaction of four guanine bases, each from a distinct G-tract, through Hoogsteen-type hydrogen bonds, forming a four-stranded right-handed helix[78–80] (**Figure I-2 B**). The oxygen atoms of the guanine residues typically coordinate a monovalent cation, which is placed in between the planar tetrads and, by neutralisation of their electrostatic repulsion, results in different degrees of structural stabilisation depending on the ion identity[84] (**Figure I-2 B**). The intrinsic stability, specific structure, and other properties of these tetrahelical architectures are also influenced by the length of the contiguous guanine repeats forming the G-tracts, as well as by the length and sequence composition of the intervening loops[85–90].

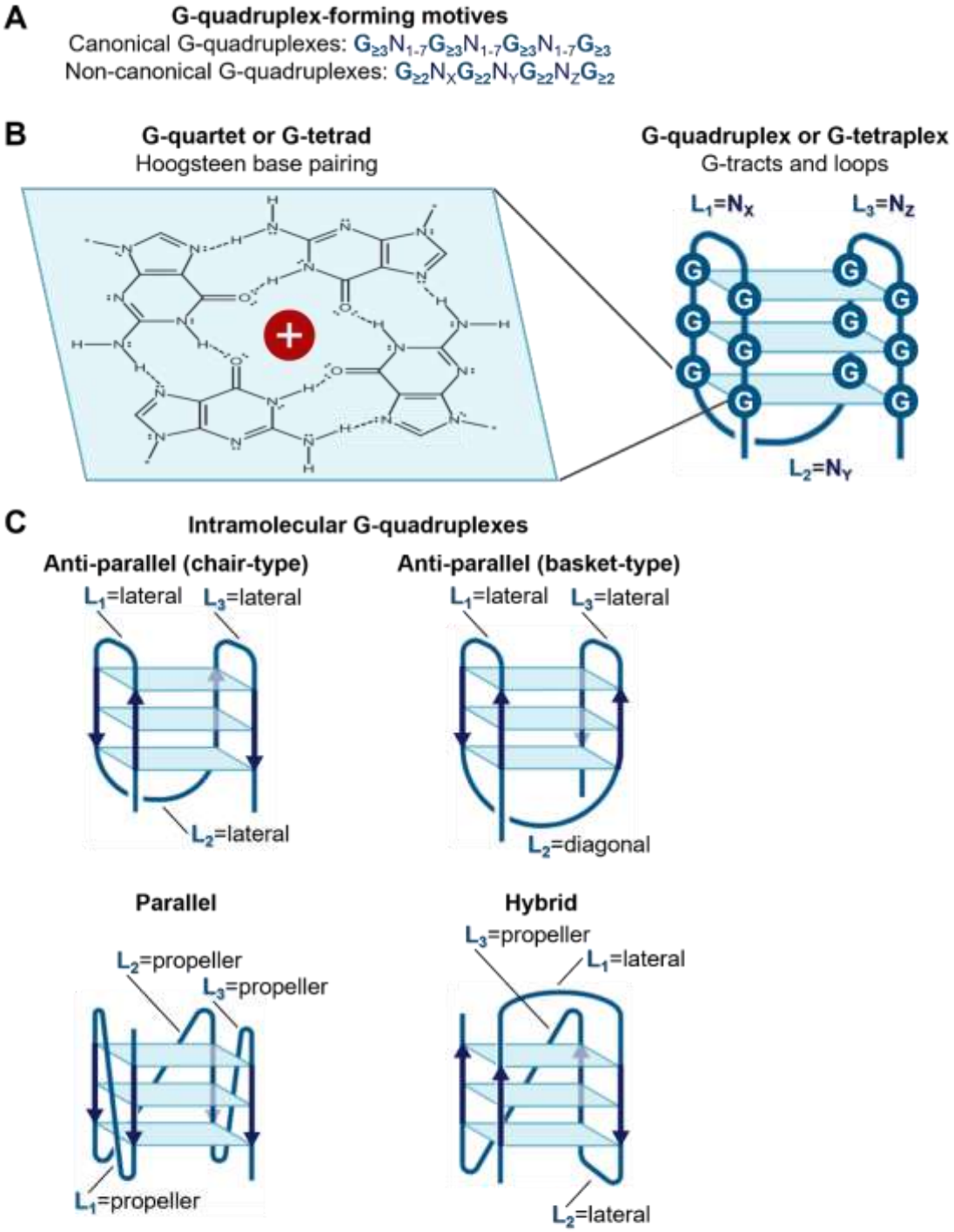

A
G-quadruplex-forming motives
Canonical G-quadruplexes: G≥3N1-7G≥3N1-7G≥3N1-7G≥3
Non-canonical G-quadruplexes: G≥2NXG≥2NYG≥2NZG≥2
B
G-quartet or G-tetrad
Hoogsteen base pairing
G-quadruplex or G-tetraplex
G-tracts and loops
L1=NX
L3=NZ
L2=NY
C
Intramolecular G-quadruplexes
Anti-parallel (chair-type)
L1=lateral
L3=lateral
L2=lateral
Anti-parallel (basket-type)
L1=lateral
L3=lateral
L2=diagonal
Parallel
L2=propeller
L3=propeller
L1=propeller
Hybrid
L3=propeller
L1=lateral
L2=lateral
D
Intermolecular G-quadruplexes
bimolecular
tetramolecular

Figure I-2 – G-quadruplex folds are structurally heterogeneous

(A) Canonical and non-canonical DNA or RNA motives that have the potential to fold into G-quadruplexes. G-tracts or G-runs are constituted by at least 2, commonly 3, guanines (**$G_{\geq 2}$**) and intervening loops longer than 7 nucleotides have been observed (N_X). **(B)** A planar G-quartet or G-tetrad is formed when four guanines, each from a distinct G-tract, interact through Hoogsteen-type hydrogen bonds. The oxygen atoms of the guanines coordinate a monovalent cation. The G-quadruplex or G-tetraplex is formed by the stacking of 2 or more G-quartets. Loops **L_1**, **L_2**, and **L_3** connect the G-tracts in the core structure. **(C)** Examples of intramolecular G-quadruplex topologies. G-quadruplexes structures vary according to the polarity of the phosphodiester backbone of the G-tracts (*e.g.* anti-parallel, parallel, hybrid) and the directionality of the loops connecting the G-tracts (*e.g.* lateral, diagonal, propeller). (**D**) Examples of intermolecular G-quadruplexes formed when the G-tracts are provided by distinct nucleic acid strands (*e.g.* bimolecular, tetramolecular).

The extensive structural heterogeneity of these folds arises from the fact that G-quadruplexes can display different topologies considering the polarity of the phosphodiester backbone linking the guanines in the core structure (**Figure I-2 C**). The parallel conformation is achieved when all the G-tracts share the same directionality, whereas, in the anti-parallel configuration, two strands have opposite orientations, leading either to the basket-type or to the chair-type tetraplexes[79–81]. In the hybrid architecture, instead, a single strand has opposite orientation with respect to the other three[79–81]. The organizational variability of these tetrahelical assemblies is further expanded by the directionality through which the loops connect the four G-runs, which can be classified in diagonal, lateral or propeller[80,81] (**Figure I-2 C**). Moreover, G-quadruplexes can be polymorphic, such that the same sequence can adopt more than one configuration in solution[91–95]. Finally, these folds can be assembled from runs of guanines present on a single polynucleotide chain, resulting in intramolecular structures, also known as uni- or mono-molecular G-quadruplexes (**Figure I-2 C**), or provided by two, three or four different strands, forming intermolecular multimeric structures, such as bi-, tri- and tetra-molecular architectures, respectively[79,96,97] (**Figure I-2 D**). These latter have been comparatively less studied with respect to the former, due to difficulties in their computational prediction and experimental detection, but which existence and

relevance are increasingly acknowledged[97–102]. Adding to the complexity of G-quadruplex structures, also intermolecular DNA:RNA hybrid tetraplex assemblies can be formed both co- and post-transcriptionally[103,104].

The establishment of these non-conventional tetrahelical secondary structures in a genomic context is in competition with the right-handed double helical structure of B-form DNA[78,105]. As a result, G-quadruplex folding is favoured during processes that lead to transient exposure of single-strand DNA, for instance during transcription, replication, and repair[106–108]. Actually, the dynamic equilibrium between the double-stranded DNA form and the G-quadruplex fold is influenced by many intrinsic and extrinsic factors[81,108–113]. Sequence-specific features and the context of the G-rich motif—like (i) the type of nucleic acid, (ii) the number of G-tetrads, (iii) loop length and composition, and (iv) the nature of the flanking regions—impact G-quadruplex stability. Tetraplex formation is also affected by environmental parameters such as (i) the species and concentrations of cations, (ii) temperature, (iii) pH, and (iv) molecular crowding. Moreover, G-quadruplex folding is driven by the ease of DNA denaturation, which depends on, for example, (i) DNA topology and supercoiling state, and (ii) by nucleosome organization and chromatin accessibility. G-quadruplex-interacting proteins play a particularly relevant role in the regulation of the G-quadruplex landscape *in vivo*. To maintain the homeostasis of these heterogeneous tetrahelical architectures, G-quadruplex-interacting proteins collectively display a wide array of activities. The action of such proteins ranges from promoting tetraplex folding and stabilisation (*e.g.* budding yeast Rap1, Hop1, Est1; mammalian nucleolin, thrombin, topoisomerases, TLS/FUS, LARK, APE1), or recruitment of additional proteins to the structure (*e.g.* Rif1 in yeast and BRCA1, p53 and TRF2 in humans), to their destabilisation (*e.g.* yeast Cdc13, human POT1, ATRX, MAZ), unwinding (*e.g.* fungal Pif1, Dna2, Sgs1, mammalian RHAU/DHX36, BLM, FANCJ, WRN, FEN1, XPD/XPB, RPA) and nucleolytic processing (*e.g.* Mre11, Kem1/Xrn1 in yeast or human GQN1)[108,114–127]. Proteins able to interact with G-quadruplexes bind these secondary structures with different specificities, for instance displaying preferences towards certain topologies or non-canonical architectures, and positions, such as by sitting on top of the tetrad stack, or recognising the loops or groove of the structure[125]. The number of validated G-

quadruplex-interacting proteins is in constant expansion, providing specific spatio-temporal regulation of these structures *in vivo* in the context of multiple molecular pathways[127].

Through a combination of computational and experimental studies, G-quadruplexes have been ubiquitously found in viruses[128,129], prokaryotes[130,131] and eukaryotes[132–136], and their distribution was correlated with functional regions. Indeed, in eukaryotic genomes, tetraplexes are enriched at relevant sites and nucleosome-depleted positions including (i) promoters, (ii) enhancers, (iii) origins of replication, (iv) ribosomal repeats, (v) telomeric sequences, (vi) immunoglobulin switch regions, (vii) positions corresponding to 5' untranslated regions in messenger RNAs, and (viii) splicing sites, among other loci[77,132,135–142]. Similarly to the nuclear DNA, also the mitochondrial genome harbours numerous G-quadruplex-forming sequences and provide favourable conditions for their folding[143–146]. Consequently, these tetrahelical arrangements have the potential to regulate fundamental biological processes[78,106,110,147,148]. DNA G-tetraplexes can exert both stimulating and inhibitory effects on transcription in a context-dependent manner, depending on their position with respect to the transcription start site, their strand orientation, and the identity of the recruited transcriptional regulatory proteins[149–153]. Related concepts in gene expression are nucleosome positioning, chromatin remodeling and the establishment of epigenetic marks by histone post-translational modifications, which have also been shown to be influenced by such secondary structures, as they can mediate recruitment of key players or exclude nucleosome deposition[139,152,154–157]. In addition, G-quadruplexes can act as signals for initiation of DNA replication, by specification of replication origins and regulation of their binding by relevant factors[158–161]. The abundant tetraplexes found in ribosomal DNA were proposed to regulate processes related to cell growth and proliferation[162,163]. Telomeric tetrahelical architectures, easily formed in the single-stranded overhang, but also present in the double-stranded upstream regions, can regulate elongation by telomerase, possess protective capping functions against unwanted activity of nucleases and mediate inter-telomeric chromosomal contacts for their structural organization and tethering to the nuclear envelope[164–166]. Moreover, G-quadruplexes were correlated with genome organization within the nuclear volume,

possibly by influencing higher-order chromatin architecture, mediating long-range chromosomal interactions and chromatin loops, and participating in the definition of topologically associating domains[167–171]. Lastly, DNA tetraplex assemblies were shown to affect the function and efficiency of DNA repair pathways, modulating recruitment of repair factors, for example positively affecting nucleotide excision repair[77,172]. To be comprehensive, it must be mentioned that the involvement of G-quadruplexes in the regulation of biological processes extends to those formed in RNA. Briefly, they have been found in different transcripts, both coding and noncoding (*e.g.* messenger, telomere-associated, ribosomal, tRNA-derived stress-induced, micro, PIWI-interacting, and long-noncoding RNAs), modulating RNA biogenesis, processing, and localisation, as well as influencing translation, alternative splicing, and polyadenylation, among other processes[110,161,173–176].

Hence, unsurprisingly, genetic or chemical perturbation of tetraplex homeostasis, due to mutations of regulatory proteins or treatment with stabilising ligands, is linked to enhanced genome instability[76–78,148,177,178]. For instance, as mentioned earlier, persistent G-quadruplexes sterically hinder the progression of the replisome causing replication stress, which, if not properly resolved, can also lead to formation of DNA double strand breaks[160,179–182] (**Figure I-3 A**). Analogously, the movement of the RNA polymerase may be stalled by the presence of a stabilised tetraplex structure in its template strand, causing alterations in gene expression alongside other repercussions[147,150,153,183–185] (**Figure I-3 B**). In addition, an enduring G-quadruplex in the coding strand can promote co-transcriptional R-loop formation, resulting in the establishment of the G-loop structure and increased gene transcription[186–188] (**Figure I-3 C**). High steady-state levels of G-loop structures are associated with DNA damage and enhanced genome instability[187,188]. Importantly, in humans, G-tetraplexes have been correlated to cancer, ageing, and a multitude of genetic and neurodegenerative diseases, like, for example, amyotrophic lateral sclerosis and frontal-temporal dementia, as well as others pathologies previously mentioned among those characterised by marked genetic instability[126,162,173,189–202]. In particular, G-rich motifs able to fold into G-tetraplexes are significantly enriched in oncogene promoters[137,203] and tumour cells bear a higher content of folded tetrahelical assemblies compared to normal cells[204].

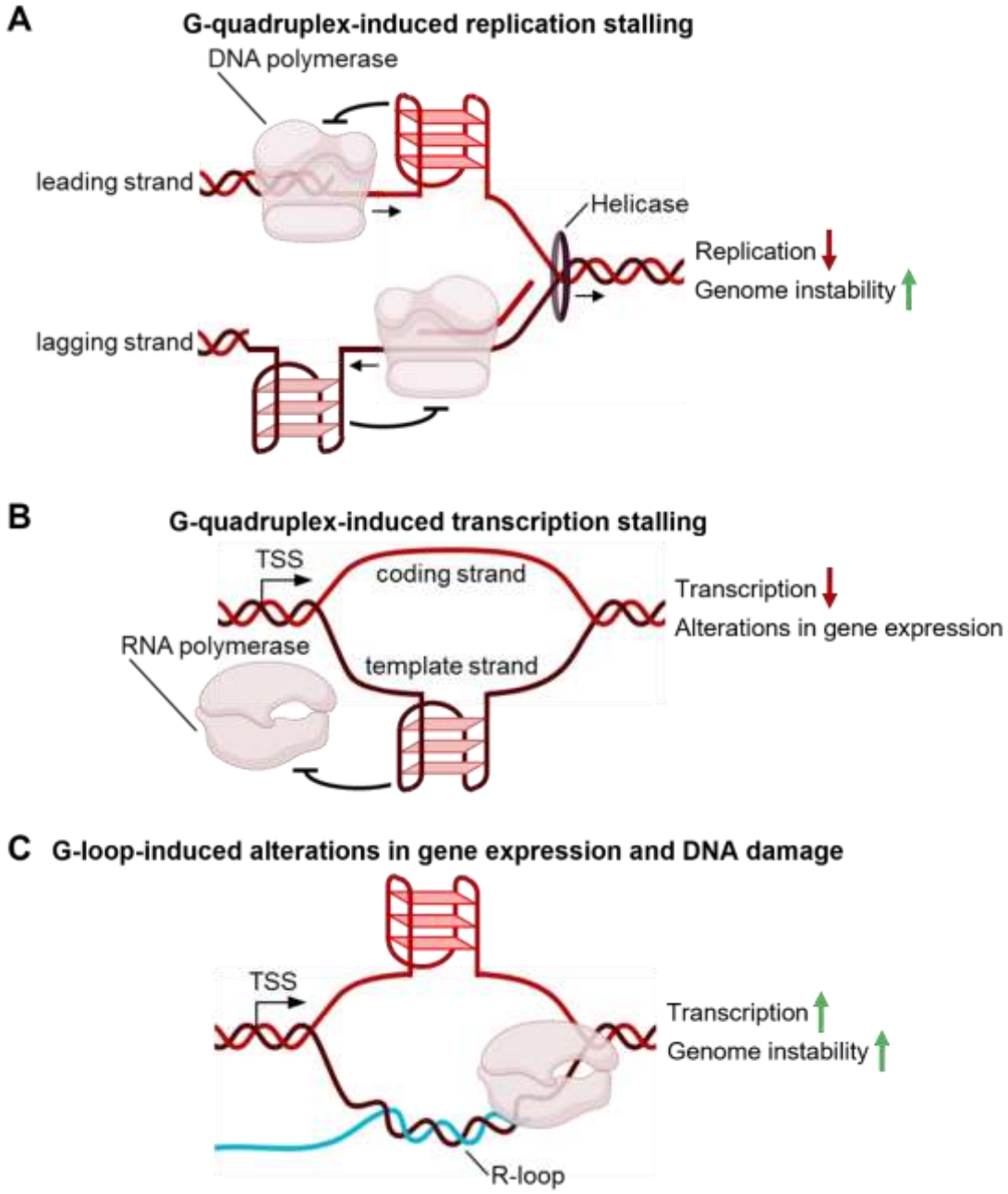


Figure I-3 – Chemical or genetic perturbation of G-quadruplex homeostasis enhances genome instability

Persistent G-quadruplexes, due to chemical stabilisation or mutations in G-quadruplex-interacting proteins, can cause replication stress, alterations in gene expression and DNA damage. **(A)** G-quadruplexes on both the leading and the lagging strands have the potential to sterically hinder the progression of DNA polymerases, inducing replication stress, and enhance genome instability. **(B)** G-quadruplexes on the template strand can block the movement of RNA polymerases, hampering gene transcription, and alter gene expression.

(C) G-quadruplexes on the coding strand can promote co-transcriptional R-loop formation, resulting in the G-loop structure, and increase gene transcription. High steady-state levels of G-loop structures are associated with DNA damage and enhanced genome instability.

Moreover, stabilization of certain topologies of telomeric G-quadruplexes was shown to inhibit telomerase activity[165,166,205–208]. Such and additional evidence have revealed the potential of these non-canonical secondary nucleic-acid structures as molecular targets both for cancer diagnosis and therapeutics. Consequently, a great effort is devoted to design effective and structure-specific G-quadruplex binders, as well as to discover novel factors regulating their metabolism and gain mechanistic insights into tetraplex biology. Indeed, the catalogue of G-quadruplex ligands is consistently growing[209,210], in the search for those potentially useful in cancer therapy[211,212]. One of such compounds is the aromatic 5,10,15,20-tetra(N-methyl-4-pyridyl)porphin, abbreviated as TMPyP4, which has been shown to selectively bind and stabilize a variety of tetraplex structures, including those found at human telomeres and in the c-Myc promoter, with antitumor activity[213–227], and has been widely used despite some controversies regarding its mechanisms of action, affinities for different DNA G-quadruplex topologies and effect on RNA tetraplexes[98,226,228–233]. In contrast to the plethora of studies dedicated to the development of drugs interacting with G-quadruplexes, relatively few of their regulatory proteins have been mechanistically or biochemically characterised[120,122]. The study of G-quadruplex-interacting proteins is not only biologically significant but also holds therapeutic relevance, providing an alternative to the direct targeting of these miscellaneous DNA structures, allowing for the exploitation of their evolutionarily selected specificity[122,124].

Overall, the multitude of structures and functions associated so far with G-tetraplexes adds an additional layer of biological complexity, presenting both an opportunity to modulate cellular processes and a potential source of complications that may lead to mutations and pathologies, thus substantiating the necessity for their research, especially, but not only, to elucidate the mechanisms behind the prevention, formation, and resolution of the DNA damages induced by G-quadruplexes, and dissect their contribution in physiology and disease.

4.3 Identification of *S. cerevisiae VID22* as a genome instability suppressing gene

To date, considerable progress has been made in the identification of human genes related to the maintenance of the integrity of the genome[1,23], and their inventory is continuously being updated. Nonetheless, our knowledge about the mechanisms and players related to genome instability suppression is still incomplete[234]. The study of processes preserving genome stability, and of the aetiology of human diseases that arise upon failure of such mechanisms, has been delayed due to the difficulties of working with mammalian systems, for example because of the limitations in performing genetic screenings[234]. Despite these limitations being partially surmounted by the recent advent of the CRISPR/Cas9 technology[235], the use over the last decades of the ascomycete *Saccharomyces cerevisiae* as a model organism has allowed to overcome some of the constraints associated with animal models. The effectiveness of using budding yeast in research relates to (i) the evolutionary conservation of a fair number of its proteins and molecular mechanisms in higher eukaryotes, (ii) the efficiency of its transformation and genetic manipulation, and (iii) the development of a vast repertoire of accessible experimental techniques and assays to probe fundamental biological aspects such as genome stability maintenance[234,236–246]. This microscopic single-celled eukaryote has a compact 16-chromosome genome which encodes ~6000 genes and is roughly 1/500 of the size of the human genome. Budding yeast has simple and affordable growth requirements. In addition, the investigation of the phenotypic effect of mutations and the performance of genetic analyses is facilitated by this yeast's ability to live in both haploid and diploid states, its propensity to integrate exogenous sequences in the genome via homologous recombination, and its capacity to carry and propagate plasmids[247]. Budding yeast can also be instrumental in the investigation *in vivo* of the biological roles of G-quadruplex structures[248].

Owing to the heterogeneity of the processes leading to genome instability and involved in its suppression, the experiments devoted to the discovery of related factors, especially in yeast, are diverse, and provide complementary and partially overlapping results depending on the specific design of the assay[243,249]. The onset and extent of genetic instability can be assessed by monitoring either (i) the loss of

genetic markers due to gross chromosomal rearrangements, (ii) the rate of accumulation of particular mutations or (iii) the incidence of variations in the length of the telomeric sequence, in a collection of knockout mutant strains compared to the wild-type yeast[19,234,250–258]. Direct evaluation of the mutational landscape on a genome-wide scale in cells in which certain gene functions have been removed can be achieved by whole-genome sequencing[259]. The occurrence of perturbations potentially challenging the integrity of the genome can also be inferred by detection of the biomarkers indicative of the activation of checkpoint pathways in response to DNA damage[249] or by following the formation of microscopically-visible subnuclear foci due to accumulation of specific proteins or post-translational modifications in consequence of the induction of DNA double-strand breaks, which are instrumental for their repair[260–262]. Lastly, the existence of differences in the fitness of strains in which a single gene has been disrupted with respect to double-mutant yeasts with the combined absence of another gene, revealed by synthetic genetic arrays, can be informative of functional interactions among the factors involved in genome stability maintenance[240]. While the identification of nearly all non-essential *S. cerevisiae* genes associated with suppressing genomic instability has been essentially accomplished, a thorough functional and biochemical characterisation is still lacking for a number of them[19,234].

VID22 (*Vacuolar Import and Degradation 22, YLR373C*), found on chromosome XII and encoding a 901-residues-long protein, is one of such cases. Indeed, the involvement of this gene in safeguarding genome integrity was evidenced by multiple independent genetic screenings. Our research group performed a synthetic genetic array screen based on the galactose-inducible overexpression of *DDC2* in haploid yeast knockout strains carrying start-to-stop deletions of all nonessential open reading frames. We found that DDC2 overexpression combined to VID22 gene deletion produced clones exhibiting a reduced colony size, indicative of synthetic dosage fitness defects[263]. Ddc2, also known as Lcd1 or Pie1, the yeast putative homolog of human ATRIP, is an essential protein physically interacting with the DNA damage checkpoint apical kinase Mec1 and contributing to its activation upon DNA lesions or replication stress[264–266]. The overexpression of *DDC2* does not affect the vitality of unperturbed cells, unless in the presence of DNA lesions, when it impairs

yeast survival by leading to prolonged cell cycle arrest due to hyperactivation of the DNA damage checkpoint response and persistent Rad53 phosphorylation[267]. Therefore, only the mutant clones experiencing spontaneous endogenous DNA damage as a consequence of the deletion of a genome maintenance gene, like *vid22*Δ cells, are sensitized to *DDC2* overexpression and exhibit defective fitness[263]. A distinct genome-wide synthetic genetic interaction screen aimed at the discovery of players involved in the DNA damage checkpoint and, specifically, participating in the rescue of stalled replication forks, revealed that the combination of the deletion of *MUS81,* or *MMS4,* with the absence of *VID22* causes cell death[268]. The two former gene products, also known as Slx3 and Slx2, homologs of human MUS81 and EME1, respectively, constitute the subunits of a heterodimeric structure-selective endonuclease reported to function during DNA repair, able to ensure proper metabolism of recombination intermediates and involved in the cleavage of substrates formed by arrested replication forks[269–272]. Moreover, in an additional work cataloguing genes required to maintain genome structure, two complementary assays, based on marker loss due to spontaneous mutagenesis, detected chromosome instability in cells depleted of *VID22*[19]. *VID22* knockout mutants were also shown to be prone to gross chromosomal rearrangements in three analogous assays based on marker loss, indicative of its role in suppressing genome instability[234]. The final publication I wish to bring to attention here employed a synthetic genetic array-high content screening strategy, based on the automated imaging of DNA damage foci marked by the repair factor Rad52 fused to GFP, for the discovery of non-essential genes participating in the DNA damage response[262]. Cells lacking *VID22* manifested elevated levels of Rad52-GFP foci, denoting the formation of DNA double-strand and single-strand breaks, exposure of stretches of single-strand DNA, and activation of the homologous recombination pathway[260,273,274], in combination with *SGS1* deletion[262]. Sgs1—an ATP-dependent 3'-5' DNA helicase belonging to the RecQ-like family and homologous to human BLM, ABCG4 and WRN—has multiple roles in genome integrity maintenance[275], being required during replication stress to prevent collapse of stalled replication forks[1,243,276–279], participating in the processing of homologous recombination intermediates[243,272,280,281], and having the ability to unfurl G-quadruplex folds[282,283].

The existence of a negative genetic interaction between *VID22* and *SGS1*, *MUS81-MMS4* or any other gene, in the context of the worsening of genome instability phenotypes, may be indicative of their functional interconnection, as it can be explained by at least three possible scenarios[243]. The gene products operate in distinct parallel redundant pathways to mend DNA in response to the same kind of lesion (*i.e.* the biochemical model)[243]. Alternatively, one protein works to repair the DNA damage, whereas the other either inhibits the onset of the injury itself (*i.e.* the suppression/repair model) or prevents its aberrant downstream processing (*i.e.* the damage/response model)[243]. Indeed, the BioGRID biomedical interaction repository, constantly compiling data on physical and genetic interactions, reports a collection of over 400 gene products potentially related to Vid22 function[284]. These comprise factors annotated to be involved in DNA damage checkpoint, repair, transcription, and chromatin remodeling processes, as well as those participating in mitochondrial, ribosomal, and telomeric functions.

Collectively, these data establish the importance of Vid22 in maintenance of genomic integrity, reporting it as a factor engaged in preventing impairments during replication or homologous recombination, and which activity is necessary to avoid the spontaneous onset of DNA damage and accumulation of severe genetic alterations at the intersection of different molecular pathways, but do not provide insights on the mechanistic details about its function.

4.4 Vid22 and its physical interactor Tbf1 are implicated in multiple processes compatible with the preservation of genomic integrity

The physical interactions between proteins are intricately linked to their cellular function and provide insights into the organization of biochemical pathways. To date, the BioGRID database reports 17 proteins involved in diverse functions that may directly associate with Vid22[284]. However, the only two validated physical interactions are with Vid22 uncharacterised paralog Env11, arising from a whole-genome duplication event that occurred around 10 million years ago during the evolutionary history of the Saccharomycotina subphylum[285], and the general regulatory factor Tbf1[286–291]. Env11, encoded by *YGR071C*, was named *late ENdosome and Vacuole interface function* due to its potential involvement in vacuolar functions[292,293]. Tbf1 (expressed from *YPL128C*), which stands for TTAGGG repeat-Binding Factor 1[294], recognizes this motif, thanks to a C-terminal DNA-binding Myb domain (related to human TRF1 and TRF2, subunits of the shelterin complex involved in telomere protection)[294–296], within subtelomeric anti-silencing regions and telomeric repeats[294,297,298]. At these loci, Tbf1 displays different modulatory roles involved in maintenance of telomere length, including control of chromatin structure and insulation, telomeric ends capping, and regulation of telomerase recruitment[294,298–303]. This helix-turn-helix transcription factor, able to homodimerise[303], also binds its consensus sequence present in the majority of the promoters of small nucleolar RNA genes and other genes involved in ribosome biogenesis, where it allows to boost their expression[298,304–306]. These RNAs participate in the chemical processing of both ribosomal and small nuclear RNAs to regulate ribosomal and spliceosomal biogenesis and functions[307,308]. Tbf1 is also found at over 130 additional promoter sequences[304], where it colocalizes specifically with both Vid22 and Env11, and affects histone occupancy promoting formation of a nucleosome free region[304]. Indeed, Tbf1 was recognised to act as a weak nucleosome displacing factor, able to effectively prevent nucleosome deposition when multiple Tbf1 motifs are located closely, but not over a single binding site[309]. Vid22 and Env11, instead, were implicated in stabilising Tbf1 binding to chromatin both at promoters and at subtelomeres[304,310]. Although Tbf1,

Vid22, and Env11 binding at several promoters is validated, precise quantification of their impact on gene expression, for instance by RNA-sequencing, is still lacking. Whereas *TBF1* is fundamental to support cell viability[311], possibly because of its essential transcriptional regulatory role rather than its functions at telomeres[298,304,312], both *VID22* and *ENV11* are dispensable for yeast survival. Nonetheless, the absence of *VID22* negatively impacts cell growth efficiency[312,313]. Additionally, deletion of *VID22*, but not of *ENV11*, causes hypersensitivity to a variety of genotoxic compounds[263,312,314,315]—among which (i) G-quadruplex stabilizing ligands (*e.g.* TMPyP4), (ii) drugs inducing DNA double-strand breaks (*e.g.* bleomycin and phleomycin), (iii) molecules causing replication stress (*e.g.* hydroxyurea), (iv) DNA-alkylating agents (*e.g.* methyl methanesulfonate, MMS), (v) crosslinkers (*e.g.* formaldehyde), and (vi) reactive oxygen species (*e.g.* H_2O_2)—possibly indicative of a Vid22-specific and Env11-independent role in the maintenance of genome stability. Furthermore, it was reported the existence of a temperature-sensitive mutant *tbf1-2* allele, insensitive to DNA alkylation or breakage upon exposure to MMS or phleomycin and, *vice versa*, a different *tbf1-3* mutant allele able to grow at restrictive temperatures but not in presence of the same genotoxic agents[312]. On top of that, the artificially diminished expression of a hypomorphic *TBF1* allele, containing a deletion in the 3' untranslated region of its messenger RNA, in a *vid22Δ* background worsens the growth impairments observed in the individual mutants[286]. These data are compatible with Tbf1 having a Vid22-associated non-essential function in preserving genome integrity, genetically separable from its essential role[312].

A link with between Vid22, Tbf1, epigenetics and genome stability may be evinced by the presence of Hst3 histone deacetylase at the same genomic locations[316]. The three proteins form an independent meta-assemblage and seem to bind the same TAGGG(C/T) consensus usually found at promoter midpoints, about 150-200 bp upstream of the transcription start site[316]. Hst3, and its redundant paralog Hst4, are involved in the deacetylation of H3-K56 throughout most of the genome[317,318], a modification established on newly synthesized histone H3 molecules during S phase by the acetyltransferase Rtt109, along with the histone chaperones Asf1 and Vps75, before their incorporation into chromatin[319–325]. The regulation of this histone

modification influences chromatin compaction[326–328], as well as checkpoint signalling in response to DNA damage and replication stress[329], as denoted by the phenotypes of *hst3*Δ *hst4*Δ double mutants having a striking resemblance to those arising in absence of *VID22* in terms of high incidence of spontaneous DNA damage, elevated genomic instability, hypersensitivity to genotoxic stress and synthetic lethality in combination with *SGS1* deletion[317,318,321,330,331]. Acetylated H3-K56 accumulates at the sites of DNA damage thanks to Mec1-mediated Hst3 targeting for proteolytic degradation, allowing proper checkpoint signal transduction, stabilisation of stalled replication forks, and recruitment of DNA repair and cohesion factors[332]. Therefore, when either acetylation or deacetylation are defective, this modification no longer identifies locations of DNA lesions, thereby impairing damage sensing and repair[332]. Curiously, a positive genetic interaction was detected between *VID22* and *VPS75*, as their combined absence alleviates the defects of the single mutants[333], suggestive of either suppression or masking mechanisms, usually arising when the gene products act in the same or in an alternative, yet related, pathway[333,334].

Moreover, Vid22 and Tbf1 were localised at sites of induced DNA double-strand breaks and directly associated with chromatin remodeling during processing of the broken ends by both the homologous recombination and the non-homologous end joining pathways, ultimately increasing repair efficiency[312]. Briefly, *vid22Δ* and *tbf1-3* mutants showed strong negative interactions upon phleomycin-induced DNA double strand break generation with key repair factors (including homologous recombination specific *RAD51* and *RAD52* and non-homologous end joining related *DNL4*), as well as with players involved in nucleosome mobilisation (like *ARP8*, *RSC2* and *SWR1*)[312]. They were also characterised by reduced 5'–3' end resection and impaired recruitment of Mre11 during homologous recombination, as well as defective Dnl4 recruitment and relegation of the broken ends during non-homologous end joining[312]. Curiously, the combination of the temperature and DNA damage sensitive *tbf1-1* allele with deletion of *RPD3*, encoding a histone H3 and H4 deacetylase[335], rescues cell sensitivity to genotoxic agents but not the temperature sensitivity[312], reinforcing the link between Tbf1 non-essential role in genome stability maintenance and regulation of histone acetylation[336]. Vid22 recruitment to

the break sites, but not to promoters, necessitates the presence of the Sgs1 helicase, which is particularly relevant for the maintenance of the integrity of the ribosomal DNA array[262]. Intriguingly, the genomic loci bound by Vid22 show significant overlap with sequences predicted to form G-quadruplexes[262,263] and, in fact, we showed that Vid22 functions to prevent G-quadruplex-dependent genome instability[263], which will be discussed in greater detail in the following paragraph.

4.5 Vid22 suppresses G-quadruplex-induced genome instability

Following the identification of *VID22* from the genome-wide screen for synthetic dosage fitness defects with *DDC2* overexpression, our research group collaborated to characterise *vid22*Δ phenotypes[263]. Firstly, we observed constitutive phosphorylation of both Ddc2, activator of DNA damage checkpoint apical kinase Mec1, and of the checkpoint effector kinase Rad53, denoting the spontaneous onset of endogenous DNA lesions in the absence of *VID22*. Moreover, *vid22*Δ cells exhibit an altered cell cycle distribution, with a significant fraction of clones blocked in the G1 phase, which is not only due to checkpoint activation. Visual analysis of chromosome profiles by pulsed field gel electrophoresis in distinct independently isolated populations of yeast cells lacking *VID22* revealed frequent and variable genomic abnormalities compatible with duplication of whole chromosomes and gross chromosomal rearrangements. Sequencing of genomic DNA from aberrant clones detected smaller duplication/deletion events occurring at over 200 loci, significantly overlapping or close to both computationally predicted and experimentally verified intramolecular G-quadruplexes. Accordingly, the mitochondrial genome, telomeres, and the ribosomal DNA array, which are enriched in G-tetraplex structures, exhibit high genetic instability in *vid22*Δ cells, which tend to have lower copies of mitochondrial DNA (in agreement with a previous publication[259]), longer telomeric ends[263], and altered ribosomal DNA repeats[262,337]. Consistently with our data, a parallel work monitoring genomic alterations caused by the deletion of non-essential yeast genes reported the insurgence of aneuploidies, instability of the *CUP1* repetitive locus, and several secondary mutations in cells lacking *VID22*[259]. Through an assay allowing to estimate gross chromosomal rearrangement rates based on marker loss associated to the presence of specific sequences, we observed a ~15-fold increase in the frequency of marker loss in *vid22*Δ cells only when combined with the insertion of G-quadruplex forming motives, whereas in *env11Δ* mutants, the rate of rearrangements was indistinguishable from that of wild-type cells independently of the potential of the sequence to fold into tetraplexes[263]. Therefore, it's apparent that cells devoid of *VID22* experience problematic complications during unperturbed growth,

especially linked to G-quadruplex structures, causing endogenous DNA damage, and ultimately leading to genomic alterations of variable magnitude.

We also mapped Vid22 binding sites genome-wide, which mostly fall in promoters and other intergenic regulatory regions, and confirmed their significant overlap or proximity to predicted G-quadruplexes[263], as was previously reported[262]. To validate these results, additional assays were performed. Focusing on one mutation hotspot in the promoter sequence of *SKN7*, we demonstrated the folding *in vitro* of the non-canonical potential G-tetraplex forming sequence by the profile of the thermal difference spectrum. The thermal difference spectrum of the G-quadruplex forming sequence in *SKN7* promoter, obtained by recording the ultraviolet absorbance spectra of the unfolded and folded states at temperatures above and below its melting temperature, displays the characteristic negative peak at 295 nm and positive values at about 240 nm and 270 nm[338,339], similarly to the parallel G-quadruplex formed by the yeast telomeric sequence in presence of potassium ions[263]. The specific binding of the anti-DNA G-quadruplex structures antibody BG4 to the wild-type *SKN7* promoter and yeast telomeric sequences provides additional evidence of their folding into tetrahelical arrangements[263]. In accordance, purified Vid22 specifically interacts with the G-quadruplexes formed by the *SKN7* promoter and yeast telomeric sequences, showing no binding towards their mutated forms designed to abolish G-tetraplex formation. Similarly, Vid22 is enriched at the G-quadruplex forming locus in *SKN7* promoter *in vivo,* as revealed by chromatin immunoprecipitation coupled to quantitative PCR analysis, yet it's virtually undetectable at the G-quadruplex mutated locus.

In summary, these findings suggest that Vid22 interacts with naturally occurring loci predisposed to forming DNA G-quadruplex structures, thereby mitigating their potential to induce genetic instability[263]. Nonetheless, further research is needed to elucidate the mechanistic details of Vid22 role in safeguarding genome integrity at sequences folding into G-tetraplex architectures. Moreover, a possible Tbf1 involvement in G-quadruplex metabolism has not been addressed yet.

4.6 Vid22 protein sequence is predicted to be related to hAT transposable elements, like human ZBEDs

Vid22 evolutionary history, its molecular mechanisms, and the identity of its candidate human homologs, may be hinted from the *in silico* prediction of Vid22 domains. The InterPro integrated database, collecting annotations about protein families and functional domains, and allowing knowledge transfer to uncharacterised sequences thanks to predictive diagnostic models[340], reports that Vid22 contains a single N-terminal BED-type zinc finger (IPR003656, residues 66-122) and a central RNase H-like superfamily domain (IPR012337, residues 276-682, **Figure I-4**). The BED domain is present, as a singular instance or in multiple copies, in over 24 thousand entries encoded in the genomes of about 2500 eukaryotic species, including fungi, plants and metazoans[340]. Its appellative derives from the name of two proteins that were found to contain such domain, that are *D. melanogaster*'s BEAF (boundary element- associated factor) and DREF (DNA replication-regulated element binding factor), insulators interacting with chromatin boundary elements and involved in the regulation of gene expression[341]. The domain is described to fold in a β-β-α secondary structure from sequences approximately 50 to 60 amino acid long, characterised by a specific pattern of Cys and His residues. The latter are organized in a $CX_2CX_nHX_{3-5}$[H/C] signature and predicted to be involved in chelation of a Zn^{2+} ion responsible for stabilizing the structural arrangement, with aromatic residues, like Trp, Phe, and Tyr, upstream of the motif[341]. This metal-coordinating fold, evolutionarily derived from the larger class of Cys_2-His_2 Zn finger domains[342–347], is proposed to mediate DNA interactions because it is present in characterised DNA-binding proteins[341], and indeed it was demonstrated for some cases[348–350]. The BED domain might also participate in protein-protein interactions, akin to certain Cys_2-His_2 Zn fingers[342,345,351–358]. In addition to being present in transcriptional regulators, for instance in those involved in the regulation of chromatin structure, the domain is found in all transposases of the hAT family[359].

The RNase H-like superfamily domain constitutes the catalytic fold present in all nucleic acid-processing enzymes of this clade sharing a common ancestry[360,361].

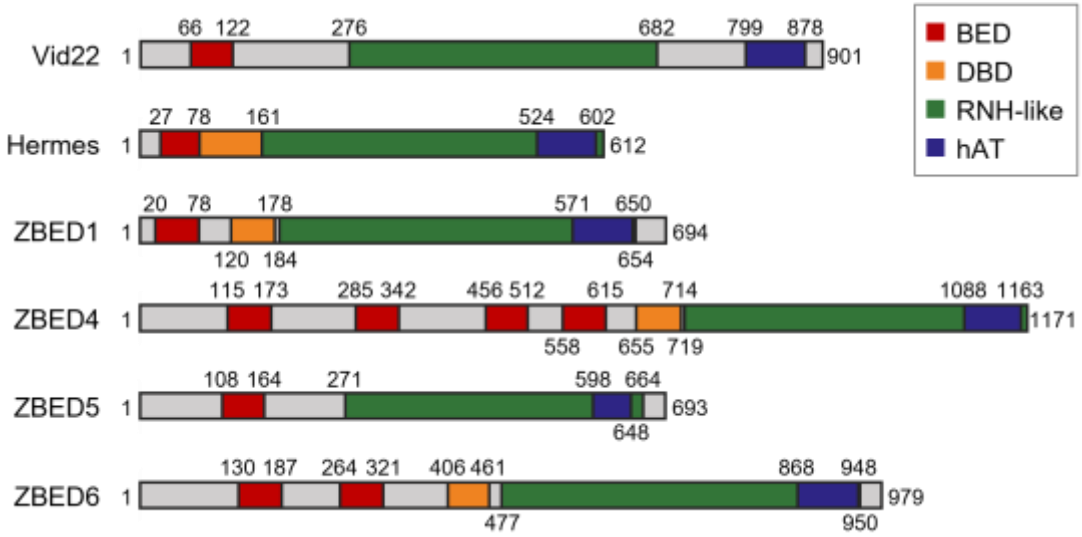


Figure I-4 – Vid22 and ZBEDs share a similar predicted domain organisation with respect to Hermes, an active hAT DNA transposase

Schematic representation of *S. cerevisiae* Vid22, *M. domestica* Hermes DNA transposase, and *H. sapiens* ZBED1, ZBED4, ZBED5, and ZBED6 protein sequences, annotated of the BED-type Zn finger (BED, IPR003656), Hermes transposase DNA-binding domain (DBD, IPR018473 or SSF140996), RNase H-like superfamily (RNH-like, IPR012337), and C-terminal hAT dimerisation domain (hAT, IPR008906 or PF05699).

The RNase H-like superfamily, owing its name to the first protein from *E. coli* in which this polypeptide chain architecture was observed[362,363], is also known as the retroviral integrase superfamily and comprises approximately 1 million of evolutionarily related gene products, belonging to over 45 thousand viral, bacterial, and eukaryotic genomes[340]. These proteins retain barely detectable, if any, sequence similarity, displaying disparate functions and having different substrate specificities, but still possessing striking structural and mechanistic resemblances, owing to analogies of the global three-dimensional architecture and the arrangement of the catalytic core[360,361]. The RNase H-like fold is characterised by a mixed five-stranded β-sheet, flanked and interrupted by α-helices with varying positions and arrangements[360,361]. The initial three strands exhibit an anti-parallel orientation, typically proceed without any insertions, and are usually followed by a conserved α-helix, while the shorter fourth and fifth strands run parallel to the first strand[360,361]. Typically, this fold is connected to other functional domains that play

roles in nucleic-acid binding, provide other enzymatic activities, and/or mediate protein-protein interactions, including self-interactions, as certain retroviral integrase proteins are reported to function as homodimers[360,361]. Despite the high level of sequence divergence between member of this superfamily, the identity and position of the catalytic amino acids is fairly conserved, especially for the first ones. These catalytic residues, which are involved in coordinating divalent metal cations essential for catalysis, typically are an Asp placed in strand β1, Asp/Glu/His in β4 or in the α-helix preceding the last β strand, and Asp/Glu/His placed after the fifth strand, normally in the C-terminal α-helix[360,361]. Proteins with an active RNase H-like fold carry out two broad categories of reactions, such as nicking of a nucleic acid filament by hydrolytic cleavage of a phosphodiester bond and strand-transfer involving rejoining of the broken filament at a different locus[360]. Retroviral integrase superfamily members include resolvases, DNA transposases, DNA polymerases, reverse transcriptases, helicases, and deadenylases, besides exonucleases, endonucleases, and integrases, classified in over 150 families and implicated in multiple cellular processes, such as DNA replication and repair, homologous recombination, transposition, splicing and RNA interference[360,361].

Curiously, the SUPERFAMILY sequence search algorithm assigns Vid22 RNase H-like domain to the Hermes transposase-like family (E-value 0.044)[364,365], after a highly characterised *Musca domestica* transposase (Figure I-4) for which the crystal structure has been determined[350,366–369]. This family of cut-and-paste DNA transposable elements and their derived proteins is also called DD[E/D], from the identity of the residues constituting the catalytic triad of the RNase H-like domain, or hAT, after three famous active members, namely the *hobo* element from *Drosophila melanogaster*, the Activator element from *Zea mays*, and the Tam3 element from *Antirrhinum majus*[359,370–372]. hAT elements form an ancient, large, diverse, and wide-spread eukaryotic protein family, which includes both active autonomous class II transposases, catalysing excision and mobilisation of their encoding sequence, and related inactive sequences from fungi, plants, and animals[359,373–377]. Notably, within this family, there is a number of examples of proteins that, over the course of evolution, were co-opted to acquire significant functions for the host genome through a process referred to as domestication or

exaptation[349,374,378–392]. Active and domesticated hAT transposases usually form dimers and oligomers[349,367,368,393], and are characterised by (i) an N-terminal BED domain, followed by (ii) a specific DNA binding domain, also mediating homodimerisation (DBD), and (iii) the catalytic RNase H-like domain with a β1-β2-β3-α1-β4-α2/3-β5-α4-α5 secondary structure[359,367,368,371] (**Figure I-4**). The latter is occasionally interrupted by an additional non-conserved domain after the fifth β-strand, commonly larger than 250 residues and participating in self-interaction[359,367,368,371]. Although all hAT members analysed so far possess a BED domain conforming to the signature presented earlier and involved in DNA interaction, its general sequence varies according to the small subterminal repeats flanking the ends of the transposon sequence, which are specific to each element[359,373]. The same variability is observed in the second DNA binding domain, which crucially participates in dimerisation of Hermes through intertwining α-helices and presents few conserved residues including a Pro-Phe couple preceded by Leu/Ile/Val residues and positively charged Arg amino acids compatible with DNA contacts[359,373]. By contrast, three regions of the RNase H-like fold display relatively high sequence conservation because of their fundamental structural and functional importance: one spanning the initial β-strand and including the first catalytic carboxylate, the second encompassing the α-helices separating strands β4 and β5 with a distinguishable CxxH and RW motives, and finally the three C-terminal α-helices with a Trp couple and the third catalytic residue in an ERF motif[359,373]. These latter represent the most conserved region among hAT elements[359,373,394,395] and were identified as being required for the self-association of certain proteins[393,396,397], leading to the annotation of the hAT family C-terminal dimerisation domain (*i.e.* IPR008906, **Figure I-4**)[340]. However, in the Hermes structure, these C-terminal helices are, in fact, crucial for maintaining the overall architecture of the transposase[367–369]. Nonetheless, this feature is recognized in Vid22 (PF05699, residues 799-878, **Figure I-4**)[340], reinforcing its relationship with hAT members, as well as in other ~54 thousand entries of approximately 2000 species, mainly eukaryotic[340].

Standard homology searches based on sequence similarity do not detect any Vid22 homolog outside the Saccharomycotina subphylum, unsurprisingly, given the

divergence of RNase H-like and hAT members despite their evolutionarily relationship[359–361,373]. Indeed, the rationale for identifying domesticated hAT proteins is the preservation of one or more domains characteristic of such transposon family[359]. Therefore, candidate human homologs may be identified by structural similarity as those proteins sharing the same BED, RNase H-like and hAT annotations. Interrogation of the SUPERFAMILY database for entries from the human proteome with predicted BED-type zinc finger domains, Hermes-like RNase H folds, or hAT dimerisation features finds 6, 17, and 5 hits, respectively[364,365]. Among these, the best candidates are ZBED proteins 1, 4, 5 and 6, displaying all of the three annotations[384] (**Figure I-4**). ZBEDs are hAT domesticated sequences[384,398], which roles in human cells are still poorly characterised, as denoted by the relatively few related articles published so far. ZBED1, also known as hDREF, is the most studied. It was reported to self-associate through its C-terminal hAT region in order to be imported in the nucleus and bind the promoters of genes encoding ribosomal proteins and factors involved in cell proliferation with a positive transcriptional regulatory role mediated by its SUMO ligase activity influencing nucleosome remodeling[393,399–401]. ZBED6 was published to work as a nucleolar transcription factor in mammals, negatively modulating the expression of many genes and, thus, globally influencing cell proliferation, organism development and growth[380]. ZBED4 function is less clear. It's a retinal protein found in human cone photoreceptors and glial Müller cells distributed between the nuclear and cytoplasmic compartments[402,403]. As ZBED1, it also dimerises through its C-terminus to reach the nucleus[393]. Most interestingly, *in vitro* experiments probing its nucleic acid binding properties revealed that it preferentially recognises G-rich sequences[404], some of which having the potential to fold into G-quadruplex structures. To date, ZBED5 is still uncharacterised. Other possibly related elements are ZNF618, which only lacks the prediction of the BED domain. On the contrary, the BED domain is the only common feature detected in ZBED2 and ZBED3, whereas the remaining ZBEDs, namely number 7, 8, and 9, group with other proteins predicted to have only the Hermes-like RNase H fold, such as FAM200A, FAM200B, GTF2IRD2A, GTF2IRD2B, KIAA1568, MYM, PRDM11, THAP12, and ZNF862.

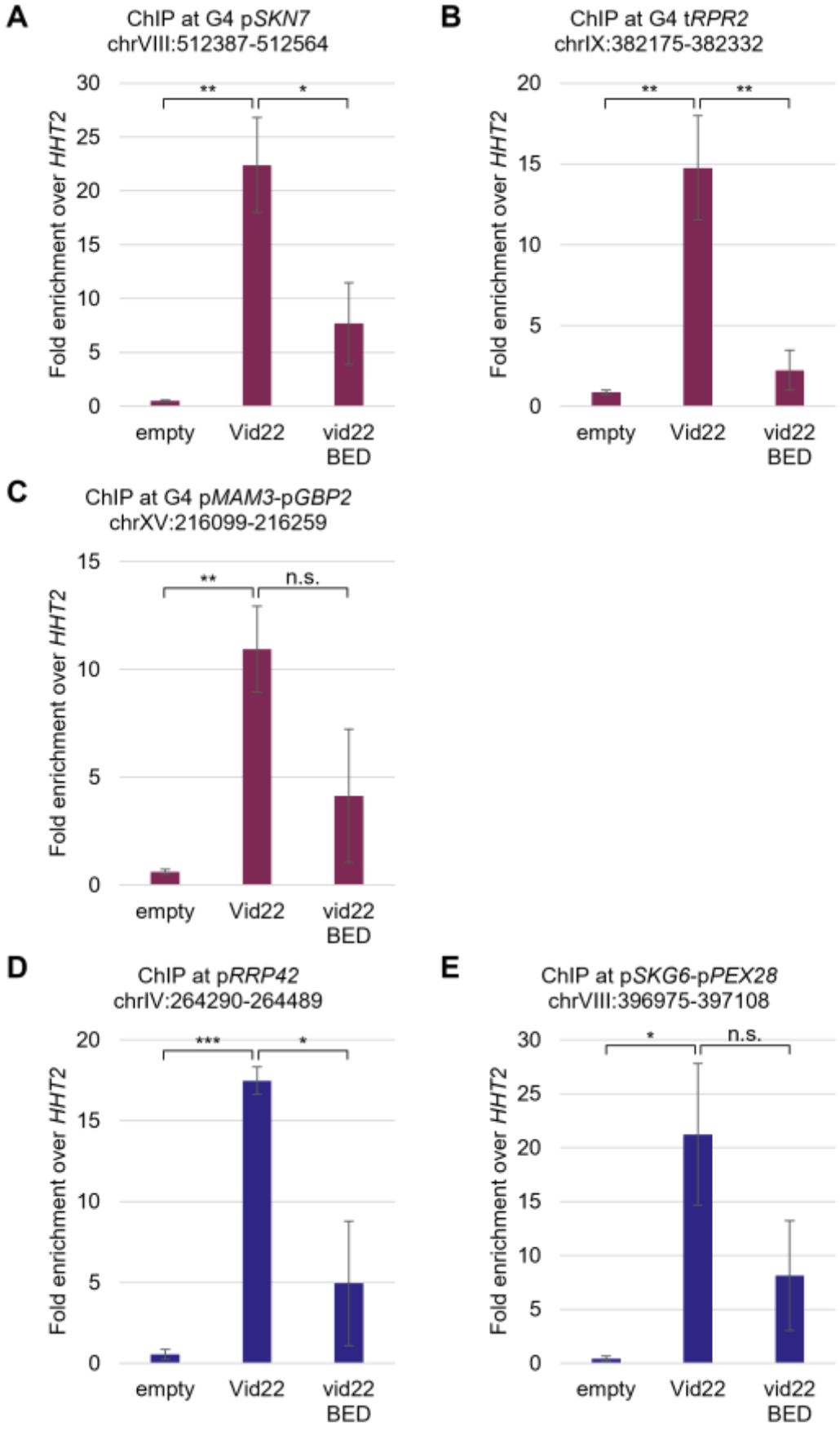
Therefore, Vid22 is computationally predicted to be evolutionarily derived from a hAT transposase, analogously to mammalian ZBEDs, among which potential functional homologs might be present. However, neither the extent of biochemical similarities with hAT transposases nor the functionality of Vid22 predicted domains has been explored yet.

5 Results and Discussion

5.1 Vid22 DNA binding is partially mediated by its BED-type zinc finger

Vid22 is a DNA-binding protein acting to prevent the spontaneous rising of unrestrained genomic instability in the budding yeast model system[19,234,262,263,268], and, in particular, able to interact with G4 structures[263]. Yet, the molecular details of its role remain unresolved. Understanding the nature of a protein's interaction with DNA provides useful functional insights and an important contribution to our knowledge of cellular and biochemical processes. It may also hold practical implications for biotechnology and therapeutic interventions[122,124]. Therefore, to unravel the complexity of Vid22-DNA interactions and better characterise its nuclear function in genome protection, we aimed to identify the protein domain responsible for the observed interaction with G-quadruplex structures. Bioinformatics resources, dedicated to the functional analysis of protein sequences, predict that a portion of Vid22 N-terminus (residues 66-122) folds into a BED-type zinc finger domain (IPR003656)[340], identified *in silico* in DNA transposases and chromatin regulatory proteins of multiple species of fungi, plants, and animals[341]. The BED finger, which has a putative DNA binding function, is characterised by a signature motif formed by a pair of Cys and a pair of His, closely spaced and preceded by aromatic residues[341]. These conserved Cys and His dyads, corresponding to Vid22 positions C87, C90, H110, and H115, are expected to be involved in coordination of a Zn^{2+} ion[341], which is essential for stabilizing the β-β-α secondary structure typical of the large class of Cys_2-His_2 Zn finger domains[342–346], from which the BED finger is evolutionarily derived[347]. Consequently, to elucidate the functional significance of this predicted domain, we distorted Vid22 BED finger by targeted mutagenesis of the metal-chelating residues. Cys to Ala or to Arg substitutions have been shown to hamper the function of the BED and other Cys_2-His_2 type zinc finger domains[349,405,406], possibly by disrupting the fold. Furthermore, His to Tyr mutation in the Cys_2-His_2 zinc finger of the yeast Adr1 transcription factor prevents DNA binding[407]. In addition, the same kind of substitution in the Cys_2-His_2 zinc finger domains of several human transcription factors occurs significantly more frequently than expected in skin cutaneous melanoma[346].

Accordingly, I examined the *in vivo* effect of C87A C90R H110Y H115Y mutations affecting Vid22 BED finger. I thus generated the desired mutations *in vitro* on a *pVID22:VID22-Myc* construct, cloned in a centromeric plasmid and subsequently transformed in *vid22*Δ haploid cells, to isolate vid22 BED mutant (vid22 BED). Firstly, I assessed by ChIP-qPCR the recruitment of the Myc-tagged vid22 C87A C90R H110Y H115Y mutant protein at a subset of target loci on different chromosomes, previously identified by Vid22 ChIP-seq[263] (**Figure 1 A-E**). The selected Vid22 binding sites include both loci containing computationally predicted and/or experimentally verified potential G quadruplex-forming motives[132,135] (G4, **Figure 1 A-C**) and devoid of such sequences (**Figure 1 D-E**). These loci fall in the promoter region of various genes (p*SKN7*, p*MAM3*-p*GBP2*, p*RRP42,* p*SKG6*-p*PEX28*), except for the G-tetraplex in chromosome IX, which is found in the terminal region of the *RPR2* gene (t*RPR2*). As a reference, samples obtained from *vid22*Δ complemented with Myc-tagged wild-type Vid22 (Vid22) and the negative control carrying the empty vector (empty) were also analysed. As a result, whereas Vid22-Myc is enriched at the selected sequences approximately 10 to 20-fold above background signal, the ability of the BED mutant to bind DNA is reduced on average of ~70% with respect to the wild-type at every locus (**Figure 1 A-E**). Importantly, the reduced enrichment of vid22 C87A C90R H110Y H115Y (*vid22*Δ + vid22 BED^{Myc}) was not due to a diminished accumulation of the protein, which levels are comparable to the wild-type Myc-tagged form encoded in the same plasmid (*vid22*Δ + Vid22^{Myc}) as shown by Western blot analysis (**Figure 1 F**). These data imply that, unexpectedly, mutation of the BED-type zinc finger considerably affects, but does not abolish, Vid22 binding to all tested sequences, irrespective of the presence of a potential G-quadruplex, suggesting that this domain is not specific for interaction with these tetrahelical architectures. To further explore the relevance of the BED domain for Vid22 function at G-tetraplexes, I verified whether the partial loss of the DNA binding function displayed by vid22 BED mutant would be sufficient to heighten cell sensitivity to the G-quadruplex stabilising ligand TMPyP4, towards which cells deleted of *VID22* are sensitive[263].


A
ChIP at G4 pSKN7
chrVIII:512387-512564
Fold enrichment over HHT2
30
25
20
15
10
5
0
**
*
empty
Vid22
vid22 BED
B
ChIP at G4 tRPR2
chrIX:382175-382332
Fold enrichment over HHT2
20
15
10
5
0
**
**
empty
Vid22
vid22 BED
C
ChIP at G4 pMAM3-pGBP2
chrXV:216099-216259
Fold enrichment over HHT2
15
10
5
0
**
n.s.
empty
Vid22
vid22 BED
D
ChIP at pRRP42
chrIV:264290-264489
Fold enrichment over HHT2
20
15
10
5
0
***
*
empty
Vid22
vid22 BED
E
ChIP at pSKG6-pPEX28
chrVIII:396975-397108
Fold enrichment over HHT2
30
25
20
15
10
5
0
*
n.s.
empty
Vid22
vid22 BED

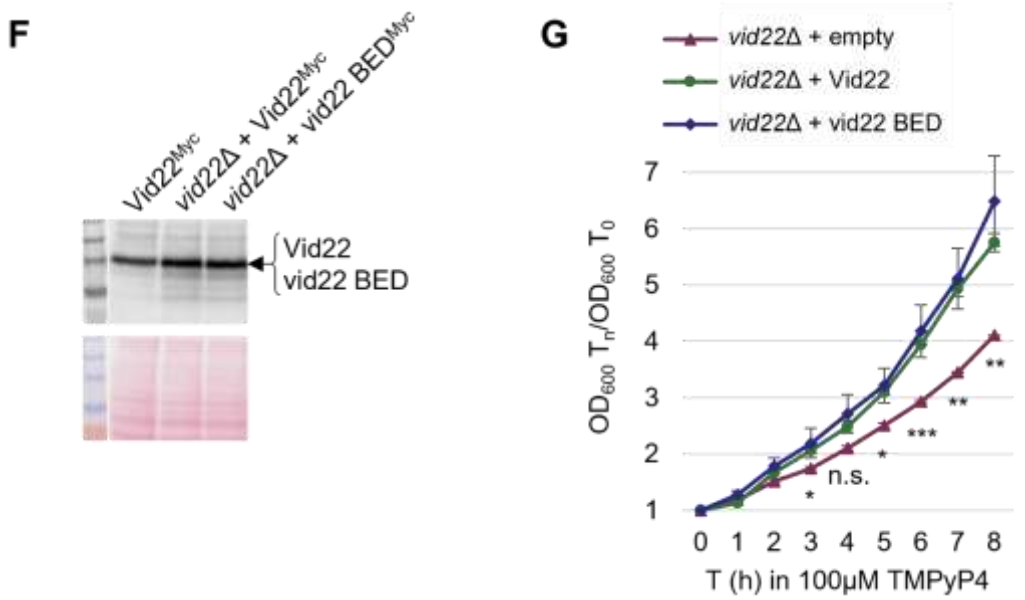


Figure 2 – Partial DNA binding of vid22 BED mutant is sufficient to retain tolerance to G-quadruplex stabilisation

(A-E) ChIP-qPCR of wild-type Vid22 or mutated in BED domain at three G4-forming[132,135,263] (G4, A-C) and two non-G4 (D-E) loci. ChIP was performed in wild-type cells transformed with the empty vector (empty, strain YGE877), *vid22*Δ complemented with Vid22-Myc (Vid22, strain YFL3902), and *vid22*Δ complemented with vid22 C87A C90R H110Y H115Y-Myc (vid22 BED, strain YFL3904) with 9E10 Myc antibodies. Fold enrichment at the different loci was calculated relative to *HHT2* internal standard. Data are represented as mean ± standard error of N=3 independent experiments and the p-value of pairwise t-tests is indicated (p>0,05 n.s.; p<0,05 *; p<0,01 **; p<0,001 ***). **(F)** Western blot analysis of the levels of wild-type Vid22 or mutated in BED domain. Total proteins were extracted from normalised cultures of Vid22-Myc (Vid22^{Myc}, strain YFL3870), *vid22*Δ complemented with Vid22-Myc (*vid22*Δ + Vid22^{Myc}, strain YFL3902), and *vid22*Δ complemented with vid22 C87A C90R H110Y H115Y-Myc (*vid22*Δ + vid22 BED^{Myc}, strain YFL3904) grown in plasmid-selective media. The blot was incubated with 9E10 Myc antibodies. The arrow points to the expected band. Ponceau staining is shown as loading control. Representative blot of N=3 independent experiments. **(G)** Sensitivity of *vid22*Δ complemented with wild-type Vid22 or mutated in the BED domain, compared to *vid22*Δ. Curve of the normalized optical density from *vid22*Δ transformed with the empty vector (*vid22*Δ + empty, strain YGE880), *vid22*Δ complemented with Vid22-Myc (*vid22*Δ + Vid22, strain YFL3902), and *vid22*Δ complemented with vid22 C87A C90R H110Y H115Y-Myc (*vid22*Δ + vid22 BED, strain YFL3904) grown in plasmid-selective media supplemented with 100µM TMPyP4. Data are

represented as mean ± standard error of N=2 independent experiments and the p-value of pairwise t-tests is indicated (p>0,05 n.s.; p<0,05 *; p<0,01 **; p<0,001 ***).

Cell sensitivity to G-quadruplex stabilisation was estimated by monitoring the optical density of a culture of *vid22*Δ cells complemented either with Myc-tagged wild-type Vid22 (*vid22*Δ + Vid22) or vid22 BED mutant (*vid22*Δ + vid22 BED), or the empty vector (*vid22*Δ + empty), growing in plasmid-selective media supplemented with 100µM TMPyP4 (**Figure 1 G**). The curve of the normalized optical density from cells expressing vid22 BED mutant resembles that from cells complemented with wild-type Vid22, reflecting that the growth of both strains is comparably affected by TMPyP4 treatment (**Figure 1 G**). The growth of *vid22*Δ cells carrying the empty plasmid, instead, is significantly affected by G-tetraplex stabilisation, as expected[263] (**Figure 1 G**). Thus, despite the recruitment to DNA of vid22 BED mutant being diminished, cells can still endure G-quadruplex stabilisation.

These findings provide compelling evidence supporting a general DNA binding function of this annotated domain, and only a marginal role in mediating the protein's interaction with G-tetraplexes *in vivo*. This possibility is supported by the observation that purified vid22 C87A C90R H110Y H115Y is still able to bind the tetrahelical secondary structures formed by the G-rich motives in the *SKN7* promoter and of the yeast telomeric sequences *in vitro*[263]. Potentially, other portions of the protein might engage in interaction with the B-form of DNA, which would support the residual enrichment observed at the various loci, and a distinct region may be more specific for the binding to G-quadruplex structures. In fact, members of the hAT family, to which Vid22 is computationally assigned to[364,365], generally possess a second DNA binding domain downstream of the BED[359,367,368,371]. Alternatively, the BED domain could be involved in protein-protein contacts, as was shown for some Cys_2-His_2 Zn finger proteins interacting with other transcription factors[345,355–358]. If Vid22 recruitment to DNA relies on BED-mediated physical association with other proteins, and this interaction is impaired by mutation of the Cys and His dyads, it would explain the diminished enrichment of the mutant. Indeed, Cys_2-His_2 zinc finger proteins are documented to operate both autonomously and to engage in cooperation with other DNA-binding proteins[345].

Therefore, to discriminate between these two possibilities, we questioned whether other proteins could be involved in mediating Vid22 recruitment to DNA.

5.2 Vid22 recruitment to G-quadruplexes is required for Tbf1 binding and is Sgs1-independent

During the characterisation of Vid22 role in genome protection, in parallel to the attempt to identify Vid22 domain responsible for G-quadruplex interaction, we sought to discern whether Vid22 operates independently in binding DNA or if it relies on the involvement of other physical interactors to mediate its recruitment, or, alternatively, if it acts as part of a larger protein complex.

In a previous work, it was evidenced by ChIP-qPCR that Vid22 recruitment to both sides of an induced targeted DNA double-stranded break was completely abolished in the absence of the ATP-dependent RecQ-like 3'-5' DNA helicase Sgs1[262]. Conversely, Vid22 binding at selected promoters, both lacking and containing G-tetraplex forming sequences, was shown to be Sgs1-independent[262]. Moreover, several high and low throughput works detected a negative genetic interaction between *SGS1* and *VID22*, as summarized in the BioGRID database[408]. For instance, it was found that their combined gene deletion exacerbates the formation of DNA damage foci, impairs stability of the ribosomal DNA array, and severely aggravates cell sensitivity to genotoxic stress induced by hydroxyurea and MMS compared to the single mutants[262], suggesting a possible interconnected role in genome stability maintenance. Vid22 is not known to physically interact with Sgs1, but because the latter is involved, both directly and by recruiting a variety of factors, in a multitude of nuclear processes, acting on a wide range of DNA substrates, and, in particular, it is able to unwind G-quadruplex folds[275], we sought to explore a possible interconnection between Vid22 and this helicase in prevention of G-tetraplex associated genome instability. Therefore, I tested if Sgs1 is required for Vid22 recruitment at the representative G-quadruplex forming locus in the promoter region of the *SKN7* gene on chromosome VIII (G4 p*SKN7*). This locus was chosen because we confirmed that the non-canonical G-rich motif embedded in the promoter (chrVIII:512269-512329)[132] folds into a G-quadruplex *in vitro*, and showed that purified Vid22 specifically binds this sequence but not its mutated form in which few crucial guanines are substituted to cytosines abolishing the formation of the secondary structure (G4 mutated)[263]. Similarly, Vid22 recruitment at this locus *in vivo* depends on the ability of the sequence to form a G-tetraplex[263].

Curiously, proximal regions, downstream to the site of the G-quadruplex, are frequently altered in *vid22*Δ clones[263]. In fact, the G-quadruplex forming sequence, but not its G4 mutated form, causes a significant increase in gross chromosomal rearrangement frequency in *vid22*Δ cells, indicating that the genetic instability is dependent of the establishment of the tetrahelical fold[263]. Curiously, despite its recruitment at the *SKN7* promoter, Vid22 does not seem to be involved in *SKN7* transcriptional control. In fact, I also verified that, during unperturbed exponential growth, there is no significant difference in the quantity of the *SKN7* target in the total RNAs extracted and retrotranscribed from the p*SKN7* G4 mutated or *vid22Δ* strains, compared to that in the wild type (**Figure 2 A-B**), despite Vid22 not being enriched at the mutated sequence[263].

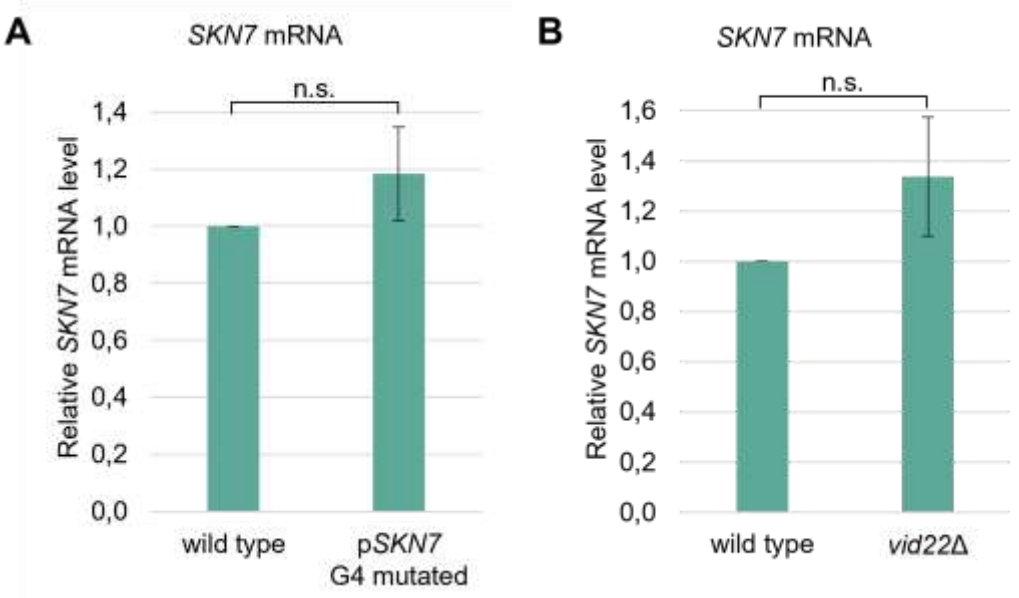


Figure 2 – Vid22 does not affect SKN7 expression

(A) *SKN7* expression in the strain mutated in the G4-forming G triplets in *SKN7* promoter as indicated in Figure 4 B (p*SKN7* G4 mutated, strain YGE1014), relative to its expression in the wild type (wild type, strain BY4741). *SKN7* expression was normalised over *HHT2* internal standard. Data are represented as mean ± standard error of N=3 independent experiments and the p-value of pairwise t-tests is indicated (p>0,05 n.s.; p<0,05 *; p<0,01 **; p<0,001 ***). **(B)** *SKN7* expression in *vid22*Δ Tbf1-Myc (*vid22*Δ, strain YGE671) relative to its expression in the wild-type (wild type, strain BY4741). *SKN7* expression was normalised over *HHT2* internal standard. Data are represented as mean ± standard error of

N=3 independent experiments and the p-value of pairwise t-tests is indicated (p>0,05 n.s.; p<0,05 *; p<0,01 **; p<0,001 ***).

Hence, for all these reasons, the G-quadruplex locus in the *SKN7* promoter could serve as a representative model through which we could gain comprehensive insights into Vid22 role in genome protection at these secondary structures. I thus performed ChIP-qPCR experiments to assay the binding at this locus of a Myc-tagged version of Vid22 when *SGS1* is deleted (Vid22 *sgs1*Δ) with respect to wild-type conditions (Vid22; **Figure 3 A**). Vid22-Myc is enriched in a comparable manner in the two samples, ~25-fold over background (**Figure 3 A**), indicating that recruitment of Vid22 at this locus is unaffected by the absence of *SGS1*.

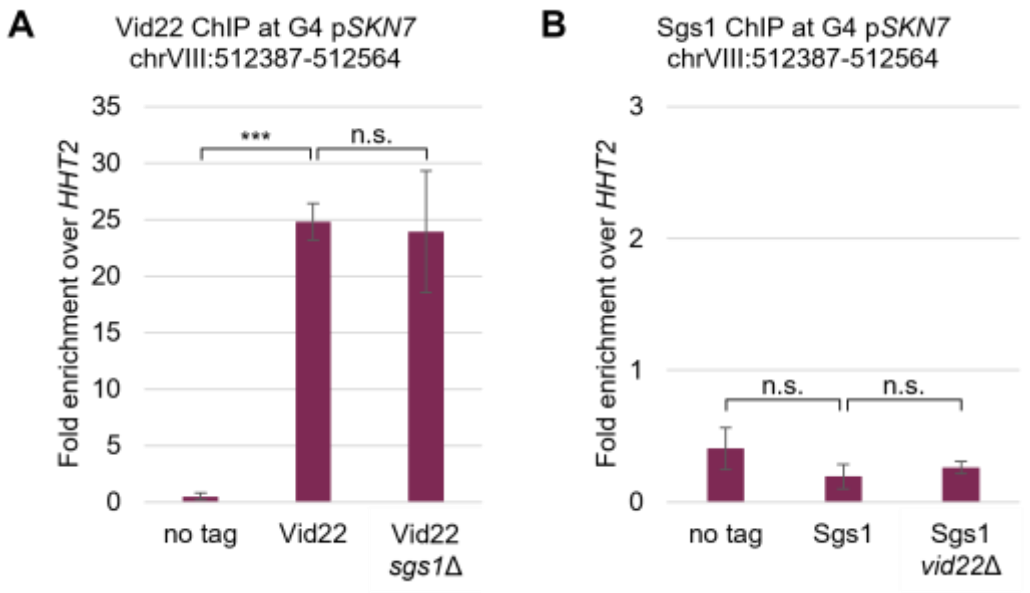


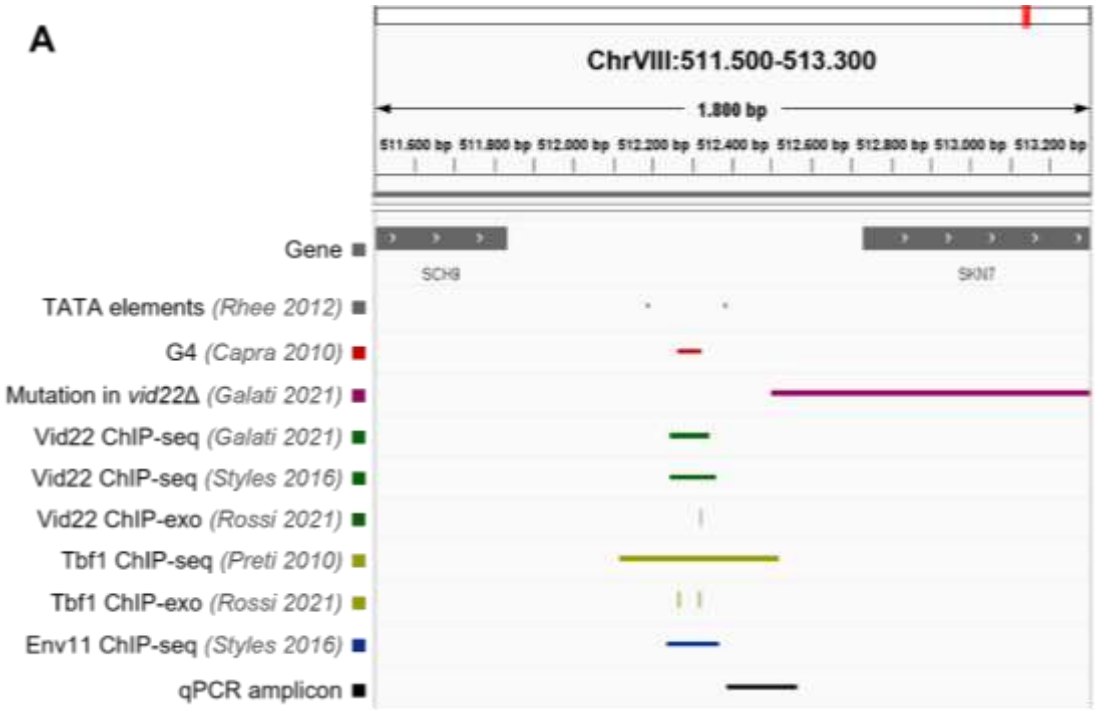
Figure 3 – Vid22 recruitment at the G4 locus in SKN7 promoter is Sgs1-independent
(A) ChIP-qPCR of Vid22 at the G4-forming[132,263] locus in p*SKN7* (G4 p*SKN7*) in presence or in absence of *SGS1*. ChIP was performed in wild-type cells (no tag, strain BY4741), Vid22-Myc (Vid22, strain YGE651.2) and Vid22-Myc *sgs1*Δ (Vid22 *sgs1*Δ, strain YFL2987) with 9E10 Myc antibodies. Fold enrichment was calculated relative to *HHT2* internal standard. Data are represented as mean ± standard error of N=3 independent experiments and the p-value of pairwise t-tests is indicated (p>0,05 n.s.; p<0,05 *; p<0,01 **; p<0,001 ***). (**B**) ChIP-qPCR of Sgs1 at the G4-forming[132,263] locus in p*SKN7* (G4 p*SKN7*) in presence or in absence of *VID22*. ChIP was performed in wild-type cells (no tag, strain BY4741), Sgs1-Myc (Sgs1,

strain YFL2955) and Sgs1-Myc *vid22*Δ (Sgs1 *vid22*Δ, strain YFL2994) with 9E10 Myc antibodies. Fold enrichment was calculated relative to *HHT2* internal standard. Data are represented as mean ± standard error of N=3 independent experiments and the p-value of pairwise t-tests is indicated (p>0,05 n.s.; p<0,05 *; p<0,01 **; p<0,001 ***).

We questioned whether Sgs1 may bind the G-quadruplex in *SKN7* promoter in the absence of *VID22*. Consequently, I also investigated by additional ChIP-qPCR assays the recruitment at this locus of Myc-tagged Sgs1 in a wild-type background (Sgs1) or combined with *VID22* deletion (Sgs1 *vid22*Δ), compared to the negative control (no tag; **Figure 3 B**). There is no statistically significant difference in the signals from the tested strains, denoting that Sgs1 is not enriched at the G4 p*SKN7* locus in any of the tested conditions (**Figure 3 B**). Consequently, despite the existence of a strong genetic interaction between *VID22* and *SGS1* in the context of the maintenance of genome stability[262,312], it is noteworthy that this helicase is not recruited at the G-tetraplex fold in *SKN7* promoter during unperturbed exponential growth, not even in the absence of *VID22*, when the secondary structure acquires mutagenic potential[263]. Additionally, Vid22 recruitment at the same locus appears to be Sgs1-independent, consistent with the observations made for other promoters[262]. Based on these findings, we postulate that Sgs1 may play a parallel non-redundant role in maintaining genome integrity at a distinct subset of genomic G-quadruplexes or be implicated in other conditions.

We next considered other potential factors that might be relevant in Vid22 recruitment to DNA, especially for its binding at the representative G-quadruplex forming locus in *SKN7* promoter. Publicly available and our unpublished genome-wide data reveal that Vid22 colocalizes with its nonessential paralog Env11 and the essential telobox-containing general regulatory factor Tbf1 at over a hundred promoters, both associated with and deprived of potential G-quadruplex forming sequences, including the *SKN7* promoter, displaying overlapping annotations[262,263,304,316] (**Figure 4 A**). Furthermore, the ability of the three proteins to physically interact forming a stable complex is well established[286–291]. Because Tbf1 is known to bind at TAGGG core consensus sequences[294,297,298,304,316,409,410] through a C-terminal Myb-type helix-turn-helix DNA-binding domain[295]

(IPR017930)[340], and *pSKN7* contains two such Tbf1 consensus binding sites embedded in the G-quadruplex motif, Vid22 and Env11 binding to this locus might be primarily dependent upon Tbf1 interaction and/or the TAGGG tracts. In fact, it was shown that the related Myb-domain in Rap1 can recognize TGGG repeats both in the canonical B-DNA and when folded into G-tetraplexes[411]. We chose to centre our investigation on the analysis of Vid22 and Tbf1 in this context, temporarily excluding Env11. Despite Env11 being able to complex with Vid22 and Tbf1, *env11Δ* cells show no increase in the sensitivity to genotoxic stress[263,312], nor they exhibit gross chromosomal rearrangements dependent on G-quadruplex forming sequences[263], suggesting that this protein has a minor or no impact on preserving genome stability. Therefore, as an approach to explore the extent of Vid22 reliance on Tbf1 for its recruitment, we opted for mutagenesis of the TAGGG motifs, in the attempt to abolish direct Tbf1 recruitment and genetically dissect the requirements for their binding. By means of the Delitto Perfetto strategy[412,413], I disrupted both Tbf1 TAGGG binding sites in *SKN7* promoter by mutagenesis of the first bases in TCGGG and CCGGG, respectively, thus maintaining unaltered the ability of the sequence to form the G-tetraplex (TBS mutated; **Figure 4 B**).



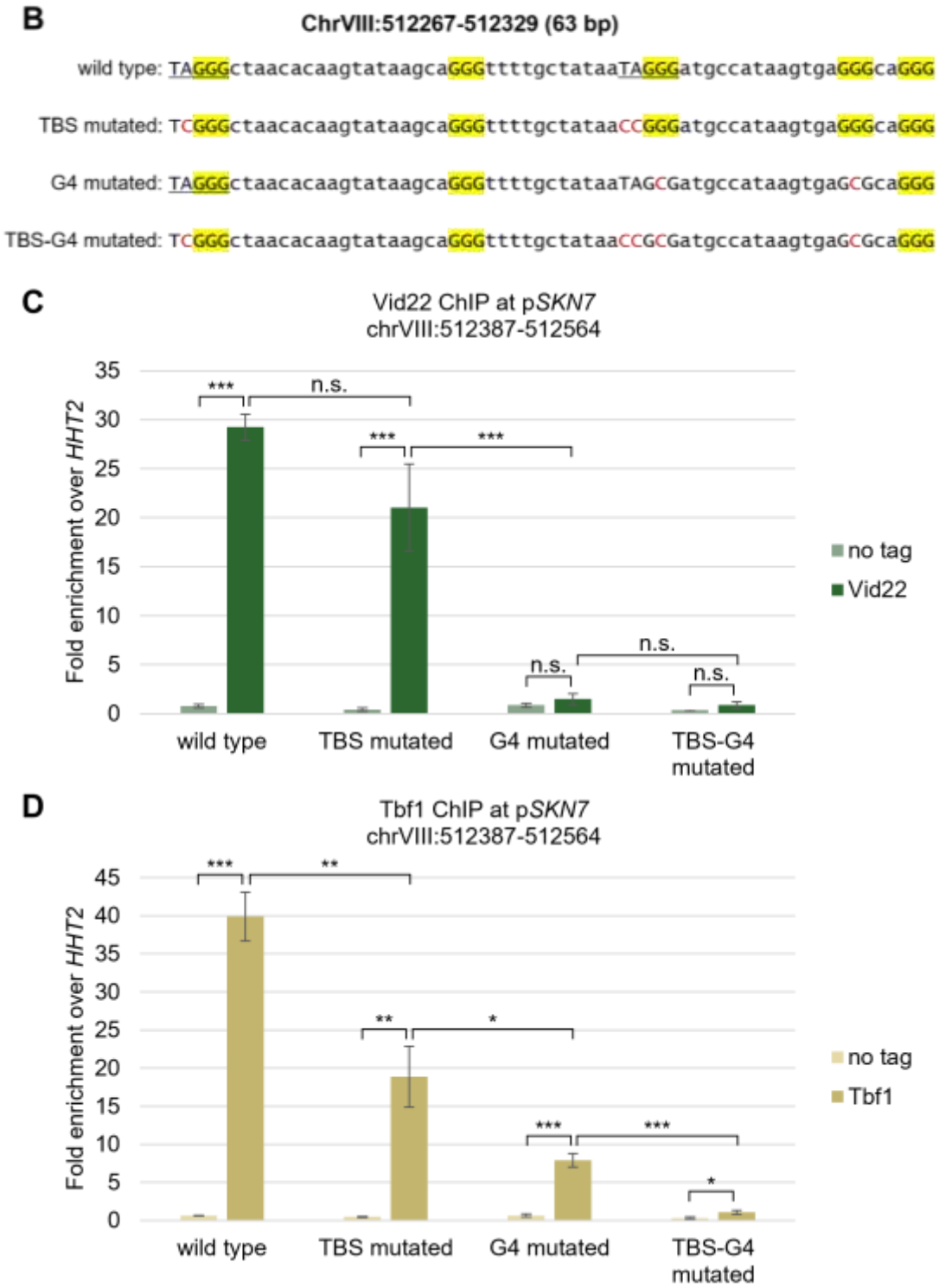


Figure 4 – Vid22 recruitment at SKN7 promoter is G4-dependent and supports Tbf1-DNA binding

(A) *S. cerevisiae* ChrVIII:511.500-513.300 genomic region with gene annotations centred over Vid22[262,263,316], Tbf1[304,316] and Env11[262] ChIP peaks. The position of TATA elements[414], of the G4[132], of the chromosomal alterations in *vid22*Δ strains[263] and of the qPCR amplicon

analysed in this work is also shown. **(B)** Sequence of the G4-forming motif in p*SKN7*. The G triplets available for G4 folding are highlighted in yellow. The Tbf1 consensus binding sites are underlined. The mutations designed to target Tbf1 consensus binding sites (TBS mutated) or to abolish G4 formation (G4 mutated) are red coloured. **(C)** ChIP-qPCR of Vid22 at the wild-type G4-forming^{132,263} locus in p*SKN7* (wild type) or mutated in the Tbf1 consensus binding sites (TBS mutated), in the G4-forming G triplets (G4 mutated), or with the combined mutations (TBS-G4 mutated). ChIP was performed in wild-type cells (wild type no tag, strain BY4741), Vid22-Myc (wild type Vid22, strain YGE651.2), cells mutated in Tbf1 consensus binding sites in pSKN7 (TBS mutated no tag, strain YFL3770), Vid22-Myc in the same background (TBS mutated Vid22, strain YFL3793), cells mutated in the G4-forming G triplets (G4 mutated no tag, strain YGE1014.2), Vid22-Myc in the same background (G4 mutated Vid22, strain YFL2922), cells mutated both in Tbf1 consensus binding sites and the G4-forming G triplets (TBS-G4 mutated no tag, strain YFL3797), and Vid22-Myc in the same background (TBS-G4 mutated Vid22, strain YFL3800) with 9E10 Myc antibodies. Fold enrichment at the different loci was calculated relative to *HHT2* internal standard. Data are represented as mean ± standard error of N=3 independent experiments and the p-value of pairwise t-tests is indicated (p>0,05 n.s.; p<0,05 *; p<0,01 **; p<0,001 ***). **(D)** ChIP-qPCR of Tbf1 at the wild-type G4-forming^{132,263} locus in p*SKN7* (wild type) or mutated in the Tbf1 consensus binding sites (TBS mutated), in the G4-forming G triplets (G4 mutated), or with the combined mutations (TBS-G4 mutated). ChIP was performed in wild-type cells (wild type no tag, strain BY4741), Tbf1-Myc (wild type Tbf1, strain YGE654.1), cells mutated in Tbf1 consensus binding sites in pSKN7 (TBS mutated no tag, strain YFL3770), Tbf1-Myc in the same background (TBS mutated Tbf1, strain YFL3794), cells mutated in the G4-forming G triplets (G4 mutated no tag, strain YGE1014.2), Tbf1-Myc in the same background (G4 mutated Tbf1, strain YGE1033), cells mutated both in Tbf1 consensus binding sites and the G4-forming G triplets (TBS-G4 mutated no tag, strain YFL3797) and in Tbf1-Myc in the same background (TBS-G4 mutated Tbf1, strain YFL3801) with 9E10 Myc antibodies. Fold enrichment at the different loci was calculated relative to *HHT2* internal standard. Data are represented as mean ± standard error of N=3 independent experiments and the p-value of pairwise t-tests is indicated (p>0,05 n.s.; p<0,05 *; p<0,01 **; p<0,001 ***).

I also combined these mutations with the base substitutions designed to abolish G-quadruplex folding (TBS-G4 mutated; **Figure 4 B**) and independently tagged both

Vid22 and Tbf1 with the Myc epitope in both backgrounds to perform ChIP-qPCR experiments. To complete the analysis and explore what effect would the mutation of the G-quadruplex alone (G4 mutated; **Figure 4 B**) have on the binding of Tbf1, since it abolishes Vid22 enrichment[263], the above-mentioned strains were tested by ChIP-qPCR in parallel to those expressing either Vid22-Myc or Tbf1-Myc both in the G-tetraplex mutated and in the wild-type backgrounds, along with their no tag controls **Figure 4 C-D**). The Vid22-Myc samples showed approximately 30-fold and 20-fold enrichment at the wild-type and TBS mutated *SKN7* promoter sequences over their respective background signals (**Figure 4 C**), indicating that Vid22-Myc can bind both loci *in vivo*, despite being partially but not significantly affected by the mutation of the TAGGG sites maintaining the G4 folding potential unaltered. Tbf1-Myc is enriched ~40-fold at wild-type *SKN7* promoter and ~20-fold at TBS mutated, implying that the mutation of the binding consensus significantly hampers, but does not abolish, Tbf1 recruitment (**Figure 4 D**). Because neither Vid22 nor Tbf1 recruitment is abolished in the TBS mutated conditions, the observed enrichment likely results from the sole contribution of the G4 assembly. Instead, mutagenesis of the guanines involved in G-quadruplex formation (G4 mutated), leaving one Tbf1 binding site unaltered (**Figure 4 B**), virtually abolishes Vid22-Myc recruitment[263] (**Figure 4 C**) and severely impairs, yet does not prevent, Tbf1-Myc binding, which displays a decreased enrichment to about 10-fold over background (**Figure 4 D**). Finally, when the mutations disrupting Tbf1 binding consensus sequence are combined with those abolishing G-tetraplex folding (TBS-G4 mutated), neither Vid22-Myc nor Tbf1-Myc are detectable at the locus by ChIP-qPCR (**Figure 4 C-D**), indicating that their binding *in vivo* is completely abolished. These data would exclude the possibility of Tbf1 being responsible for the recruitment of Vid22, because if this were the case, Vid22 would still be enriched at the G4 mutated locus, where Tbf1 is present. It might be, however, that the ChIP-qPCR assay is not sensitive enough to detect the residual binding of Vid22. Alternatively, these results may indicate that Vid22 is important for stabilising the binding of Tbf1 to chromatin, as previously suggested[304]. Therefore, taken together, these data support the interpretation that G-quadruplex folding, and not the exact nucleotide sequence, is the major determinant for Vid22 and Tbf1 recruitment at the *SKN7* promoter and

that, probably, Vid22 maintains Tbf1 bound to DNA. Similarly to what observed in the G4 mutated condition, additional ChIP-qPCR experiments revealed that, in a *vid22Δ* background, Tbf1-Myc enrichment at wild-type *SKN7* promoter is severely impaired compared to wild-type conditions (**Figure 5 A**), despite its increased transcript and protein levels, evaluated by qPCR and Western blot, respectively (**Figure 5 B-C**), reinforcing the model in which Tbf1 is stabilised at the locus by G-quadruplex-dependent recruitment of Vid22.

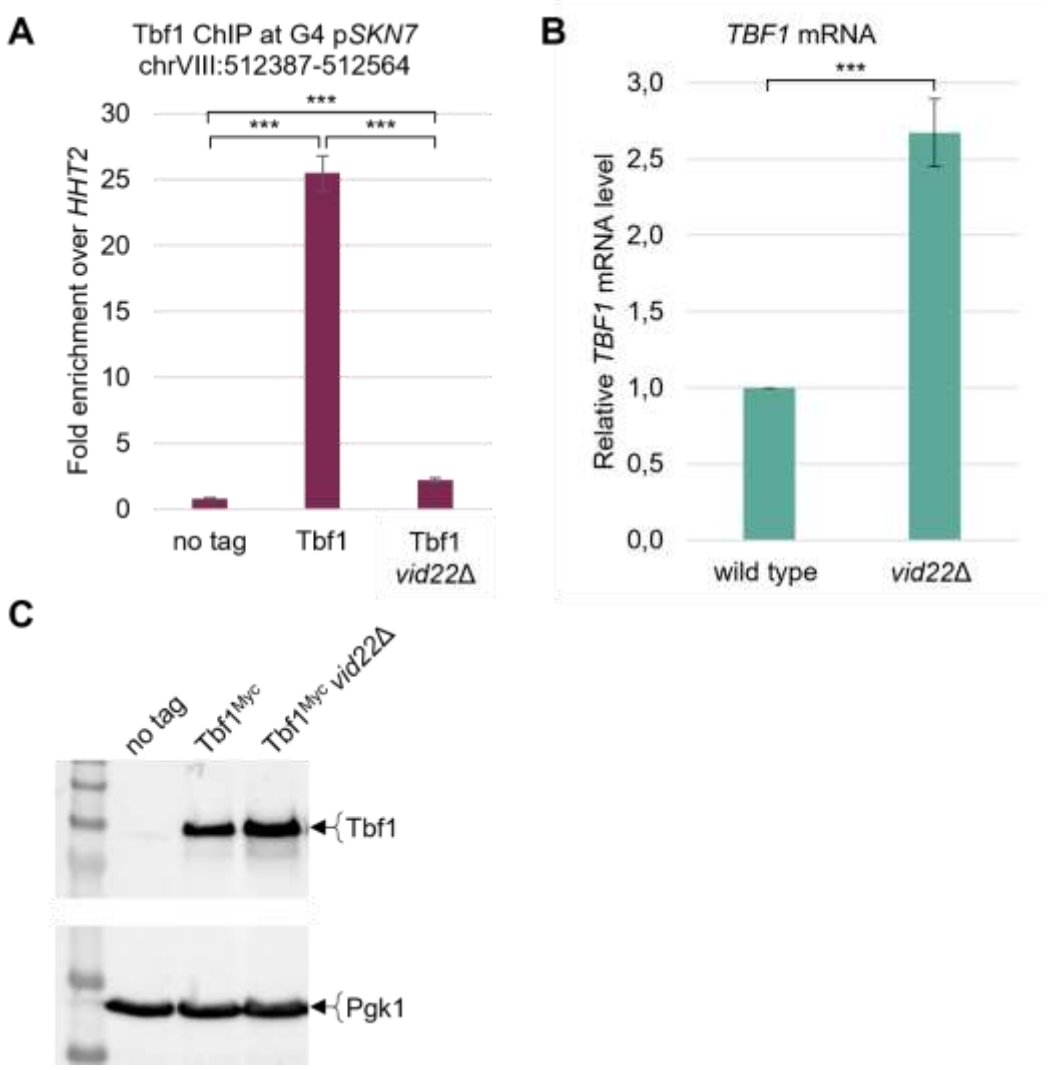


Figure 5 – Tbf1 recruitment at the G4 locus in SKN7 promoter is severely affected in absence of VID22

(A) ChIP-qPCR of Tbf1 at the G4-forming[132,263] locus in p*SKN7* (G4 p*SKN7*) in presence or in absence of *VID22*. ChIP was performed in wild-type cells (no tag, strain BY4741), Tbf1-Myc

(Tbf1, strain YGE654) and Tbf1-Myc *vid22*Δ (Tbf1 *vid22*Δ, strain YGE671) with 9E10 Myc antibodies. Fold enrichment was calculated relative to *HHT2* internal standard. Data are represented as mean ± standard error of N=3 independent experiments and the p-value of pairwise t-tests is indicated (p>0,05 n.s.; p<0,05 *; p<0,01 **; p<0,001 ***). **(B)** *TBF1* expression in *vid22*Δ Tbf1-Myc (*vid22*Δ, strain YGE671) relative to its expression in the wild-type (wild type, strain BY4741). *TBF1* expression was normalised over *HHT2* internal standard. Data are represented as mean ± standard error of N=3 independent experiments and the p-value of pairwise t-tests is indicated (p>0,05 n.s.; p<0,05 *; p<0,01 **; p<0,001 ***). **(C)** Western blot analysis of the levels of Tbf1 in presence or in absence of *VID22*. Total proteins were extracted from wild-type cells (no tag, strain BY4741), Tbf1-Myc ($Tbf1^{Myc}$, strain YGE654), Tbf1-Myc *vid22*Δ ($Tbf1^{Myc}$ *vid22*Δ, strain YGE671). The top blot was incubated with 9E10 Myc antibodies, the bottom blot with Pgk1 antibodies. Pgk1 was used as loading control. Blot of N=3 independent experiments.

A compatible explanation could be that Vid22 only exists bound to Tbf1 and the Vid22-Tbf1 complex is recruited to G-quadruplexes with a genome-protective function, giving the signal observed in the TBS mutated background, whereas the fraction of Tbf1 not bound to Vid22 can recognise TAGGG sites in B-DNA throughout the genome and carry out its essential Vid22-independent role, unrelated to genome stability maintenance[312], accounting for the residual Tbf1 binding observed in the G4 mutated or *vid22Δ* scenarios. Yet, since Vid22 and Tbf1 can physically interact, but mutation of Tbf1 TAGGG binding consensus is not enough to abolish its recruitment, the phenotypic analysis of strains expressing Vid22 mutant proteins unable to interact with Tbf1 may possibly help to dissect their binding requirements and clarify their individual contributions to the maintenance of genome stability at G-quadruplexes.

5.3 Vid22-Tbf1 interaction is partially compromised by BED domain mutations and completely abolished by C-terminal deletions

Due to the ambiguity regarding the relationship between Vid22 and Tbf1 recruitment, complex formation, and the specificity towards either B-DNA or G4 interaction, we strived to obtain Vid22 mutants that disrupt the physical interaction between these two proteins, in the effort to facilitate the dissection of their individual contributions to genome protection, in particular at the level of G-tetraplex structures. At the time, no bioinformatic, structural, or experimental data providing details of Vid22-Tbf1 interaction were available. Nonetheless, Vid22 BED domain might mediate the interaction with Tbf1, as some Cys_2-His_2 fingers are reported to be involved in protein-protein contacts[342,345,351–358]. Therefore, to further explore the relevance of this domain and possibly discriminate its role towards protein or DNA binding, I assayed the interaction of Myc-tagged Vid22 BED mutant with Flag-tagged Tbf1 through co-IP experiments (**Figure 6**).

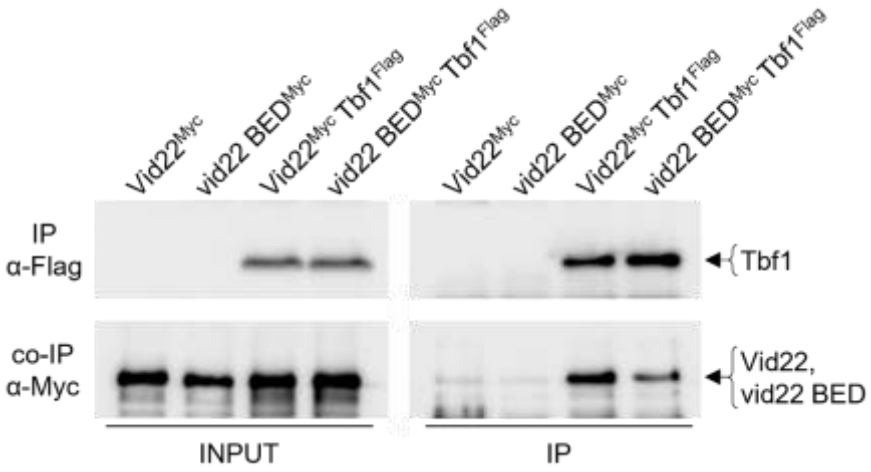


Figure 6 – Mutation of Vid22 BED domain partially impairs its interaction with Tbf1

Western blot analysis of co-IP experiments of wild-type Vid22 or mutated in the BED domain with Tbf1. Normalised protein extracts of Vid22-Myc ($Vid22^{Myc}$, strain YFL3902), vid22 BED-Myc ($vid22\ BED^{Myc}$, strain YFL3904), Vid22-Myc Tbf1-Flag ($Vid22^{Myc}\ Tbf1^{Flag}$, strain YGE1071.1), and vid22 BED-Myc Tbf1-Flag ($vid22\ BED^{Myc}\ Tbf1^{Flag}$, strain YGE1073.1) were incubated with M2 Flag antibodies. Blots were incubated individually with M2 Flag

(top panels) or 9E10 Myc (bottom panels) antibodies. The arrows point to the expected bands. Representative immunoblots of N=2 independent experiments.

Protein extracts derived from *vid22*Δ Tbf1-Flag cells complemented with a centromeric plasmid expressing either vid22 C87A C90R H110Y H115Y-Myc (Tbf1 vid22 BED) or Vid22-Myc (Tbf1 Vid22) were incubated with anti-Flag-conjugated beads. Samples from Tbf1-untagged negative control strains (vid22 BED and Vid22) were analysed in parallel to confirm specificity of the interaction. Subsequently, the co-immunoprecipitated fraction was subjected to Western blot analysis using anti-Myc antibodies (**Figure 6**). Notably, Myc-tagged vid22 BED mutant consistently co-precipitates with Tbf1-Flag to a lesser extent compared to wild-type Vid22-Myc, despite the levels of both proteins in the input fractions being comparable (**Figure 6**). These results imply that the interaction between vid22 BED mutant and Tbf1 is partially compromised *in vivo*. If Vid22 depends on Tbf1 for its recruitment to TAGGG DNA sequences and mutation of the BED domain impairs their physical interaction (**Figure 6**), it could explain the overall reduction in DNA enrichment observed in the previous ChIP experiments (**Figure 1 A-E**). However, this interpretation contrasts with the evidence from the ChIP experiments at the G4 mutated *SKN7* promoter, at which Vid22 is not present despite Tbf1 recruitment (**Figure 4 C-D**), unless Vid22 can only be found in complex with Tbf1 and the Vid22-Tbf1 complex is specifically recruited by G-quadruplex structures. Therefore, Vid22-Tbf1 complex formation and their recruitment to DNA seem to be intimately linked, encouraging the search for Vid22 mutants unable to interact with Tbf1 to clarify the interplay between these two proteins and B-/G4-DNA interactions.

Therefore, to identify Vid22 mutants that would abolish interaction with Tbf1, we considered adopting a random mutagenesis strategy based on the following rationale. On one hand, Tbf1 was proposed to possess both an essential function, potentially unrelated to Vid22, in transcriptional regulation[304,312], as well as a non-essential role associated with Vid22 and linked to genome stability maintenance[286,312]. On the other hand, it had been previously observed in our laboratory that overexpression of *VID22* in a haploid background has a detrimental effect on cellular viability. While the precise cause of the observed phenotype

remains uncertain, we formulated the following hypothesis. Given the documented physical interaction between Vid22 and Tbf1[286–291] and the requirement of a balanced protein complex stoichiometry for the execution of proper biological function, we reasoned that the cytotoxicity resulting from *VID22* overexpression might be attributed to the sequestration of free Tbf1 by the excess of Vid22 molecules. As a result, this interference could potentially hinder the execution of Tbf1 essential function. Consequently, we envisaged that Vid22 mutants which do not induce toxicity upon overexpression may have lost their ability to interact with Tbf1. To isolate *VID22* alleles non-lethal upon overexpression, we thus conducted a genetic screening in which its coding sequence was first subjected to random mutagenesis through PCR, employing unbalanced dNTP concentrations. Taking advantage of the efficiency of the yeast homologous recombination system, the mutated sequences were subsequently cloned *in vivo* in a centromeric plasmid under the control of a strong Gal-inducible promoter, and N-terminally tagged with the HA epitope. Cells which reconstituted these vectors were then plated on a galactose-containing medium to stimulate overexpression. Approximately 150 strains viable under overexpression conditions were further examined by Western-blot analysis to evaluate the dimension and abundance of mutant proteins. The six strains producing mutant proteins of a size and to a level comparable to that of the wild-type protein were chosen for vector extraction and sequencing (**Figure 7 A**), whereas the other strains produced mutant proteins with an exceedingly lower molecular weight and/or at a reduced level compared to that of the wild-type. The only two full-length mutants recovered in the screening, vid22 #29 and #44, displayed 8 and 10 missense mutations, respectively, because of single nucleotide changes resulting in the coding of a different amino acid during translation, which may have completely altered Vid22 structure and/or function and were thus ineligible for the dissection of the contribution of each mutation to the phenotype (**Figure 7 A**). A frame-shift mutation, due to a single-nucleotide deletion in *VID22* gene, characterises mutant #76, having an altered sequence from residue 834 and truncated of the last 61 amino acids (**Figure 7 A**).

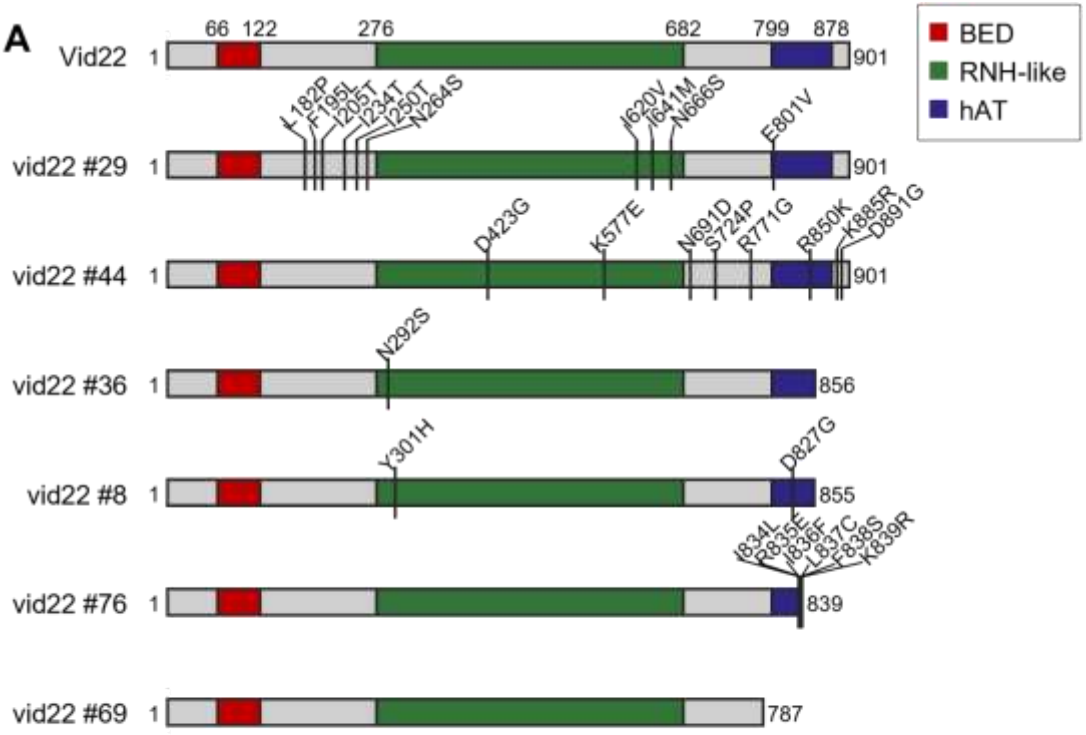

A
Vid22
vid22 #29
vid22 #44
vid22 #36
vid22 #8
vid22 #76
vid22 #69
66 122 276 682 799 878
1 901 856 855 839 787
L182P F195L I205T I234T I250T N264S
I620V I641M N666S E801V
D423G K577E N691D S724P R771G R850K K885R D891G
N292S
Y301H D827G
I834L R835E I836F L837C F838S K839R
BED
RNH-like
hAT
B
control
induced
empty
Vid22
vid22 #8
vid22 #29
vid22 #36
vid22 #44
vid22 #69
vid22 #76
C
empty
Vid22
vid22 #8
vid22 #29
vid22 #36
vid22 #44
vid22 #69
vid22 #76
Vid22, vid22 #29, #44
vid22 #36, #8, #76, #69

Figure 7 – Vid22 alleles non-toxic upon overexpression produce full-length proteins with several substitutions or truncated proteins with few substitutions

(A) Schematic representation of wild-type Vid22 protein or of the randomly mutated variants obtained from the screening for alleles non-toxic upon overexpression, annotated of the predicted BED-type Zn finger (BED, IPR003656), RNase H-like superfamily (RNH-like, IPR012337) and C-terminal hAT dimerisation domain (hAT, PF05699). Nonsynonymous substitutions generated in vid22 mutants are reported. **(B)** Yeast sensitivity to overexpression of *VID22* or the randomly mutated alleles evaluated by spot assay. Log-phase wild-type cells transformed with the empty vector (empty, strain YGE722.1), transformed with the overexpression vector of wild-type Vid22 (Vid22, strain YGE723.1) or of the randomly mutated variants vid22 #8 (strain YFL3376), vid22 #29 (strain YFL3377), vid22 #36 (strain YFL3378), vid22 #44 (strain YFL3379), vid22 #69 (strain YFL3380), or vid22 #76 (strain YFL3381) were diluted to an optical density at 600 nm (OD_{600}) of 0,06 in H_2O and spotted, along with three 10-fold serial dilutions, on solid plasmid-selective media with 2% glucose as a control (control) or 2% raffinose, 2% galactose to induce allele overexpression (induced). The plates were photographed after 3 days of incubation at 28°C. Representative outcome of N=3 independent experiments. **(C)** Western blot analysis of the levels of wild-type Vid22 or of the randomly mutated variants obtained from the screening for alleles non-toxic upon overexpression. Total proteins were extracted from normalised cultures of wild-type cells transformed with the empty vector (empty, strain YGE722.1), transformed with the overexpression vector of wild-type N-terminally HA-tagged Vid22 (Vid22, strain YGE723.1) or of the randomly mutated variants vid22 #8 (strain YFL3376), vid22 #29 (strain YFL3377), vid22 #36 (strain YFL3378), vid22 #44 (strain YFL3379), vid22 #69 (strain YFL3380), or vid22 #76 (strain YFL3381) grown in plasmid-selective media with 2% raffinose, 2% galactose to induce allele overexpression. The blot was incubated with 12CA5 HA antibodies. Arrows point to the expected bands. Ponceau staining is shown as loading control. Representative blot of N=3 independent experiments.

The three remaining mutants—vid22 #36, #8, and #69—exhibited up to two aminoacidic substitutions, and C-terminal truncations of 45 to 114 residues in length, due to the introduction of a premature stop codon (**Figure 7 A**). I validated these results by examining the rescue of the cytotoxicity upon overexpression of these mutant alleles, following their reintroduction in wild-type haploid yeasts,

through spot assays (**Figure 7 B**). The growth on non-inductive and inductive media of serial dilutions of these six strains was qualitatively compared to the positive control conditionally overexpressing wild-type *VID22* (Vid22), and the negative control transformed with the empty plasmid (empty). Whereas the galactose-induced overexpression of wild-type *VID22* is lethal, all strains overexpressing the mutated alleles are viable on galactose, comparably to the negative control (**Figure 7 B**), despite the high and comparable accumulation of the mutant and wild-type proteins (**Figure 7 C**). Therefore, we deduced that the C-terminus may be relevant for Vid22 interaction with Tbf1.

To test this possibility, I thus mutated *VID22* encoded at its endogenous locus to generate a mutant expressing a truncated protein, vid22 856-901Δ (vid22 CtΔ45), with the minimal deletion that emerged in the screening and C-terminally tagged with the HA epitope in the Tbf1-Myc background. I next investigated by co-IP the physical interaction between vid22 856-901Δ and Tbf1 (**Figure 8 A**). Tbf1 was immunoprecipitated via Myc-specific antibodies and the co-immunoprecipitated fraction probed for full-length Vid22 (Vid22^{HA} Tbf1^{Myc}) or its C-terminally truncated variant (vid22 CtΔ45^{HA} Tbf1^{Myc}). The truncated vid22 856-901Δ protein is undetectable in Tbf1 co-immunoprecipitated fraction, whereas wild-type Vid22 co-precipitates with Tbf1 as expected (**Figure 8 A**), implying that the established experimental conditions allow to detect the specific association between these two proteins and that the integrity of Vid22 C-terminus may be required for such interaction. However, it must be noted that, despite normalization to the same total protein concentration, the level of vid22 856-901Δ in the input samples is lower with respect to that of full-length Vid22 (**Figure 8 A**). Assuming that the experimental conditions allow to avert spontaneous protein degradation, this implies that the truncated protein may be intrinsically unstable *in vivo*. Given this observation, it cannot be excluded that the absence of vid22 856-901Δ signal from Tbf1 co-immunoprecipitated fraction is attributable to its levels being below detection sensitivity, rather than to an actual loss of interaction.

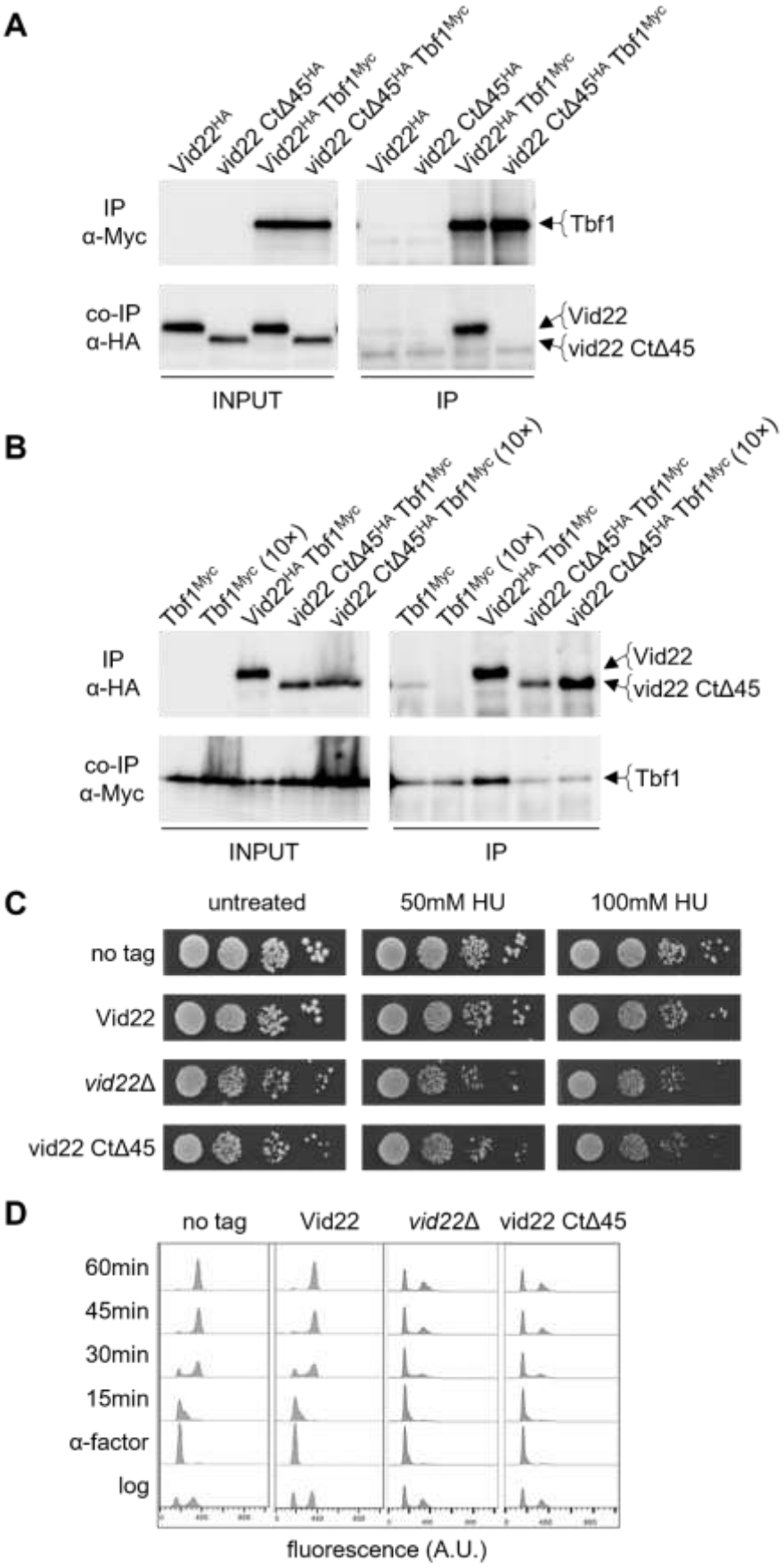

A
Vid22HA
vid22 CtΔ45HA
Vid22HA Tbf1Myc
vid22 CtΔ45HA Tbf1Myc
IP
α-Myc
Tbf1
co-IP
α-HA
Vid22
vid22 CtΔ45
INPUT
IP
B
Tbf1Myc
Tbf1Myc (10×)
Vid22HA Tbf1Myc
vid22 CtΔ45HA Tbf1Myc
vid22 CtΔ45HA Tbf1Myc (10×)
IP
α-HA
Vid22
vid22 CtΔ45
co-IP
α-Myc
Tbf1
INPUT
IP
C
untreated
50mM HU
100mM HU
no tag
Vid22
vid22Δ
vid22 CtΔ45
D
no tag
Vid22
vid22Δ
vid22 CtΔ45
60min
45min
30min
15min
α-factor
log
fluorescence (A.U.)

Figure 8 – Truncation of Vid22 C-terminus abolishes interaction with Tbf1, renders cells HU-sensitive, hampers cell cycle progression but impairs protein stability

(A) Western blot analysis of co-IP experiments of wild-type Vid22 or the C-terminally truncated mutant with Tbf1. Normalised protein extracts of Vid22-HA (Vid22^{HA}, strain YFL3161), vid22 856-901Δ-HA (vid22 CtΔ45^{HA}, strain YFL3170), Vid22-HA Tbf1-Myc (Vid22^{HA} Tbf1^{Myc}, strain YFL3164), and vid22 856-901Δ-HA Tbf1-Myc (vid22 CtΔ45^{HA} Tbf1^{Myc}, strain YFL3111) were incubated with 9E10 Myc antibodies. Blots were incubated individually with 9E10 Myc (top panels) or 12CA5 HA (bottom panels) antibodies. The arrows point to the expected bands. Representative immunoblots of N=2 independent experiments. **(B)** Western blot analysis of co-IP experiments of Tbf1 with wild-type Vid22 or the C-terminally truncated mutant. Normalised protein extracts of Tbf1-Myc (Tbf1^{Myc}, strain YGE654.1), Vid22-HA Tbf1-Myc (Vid22^{HA} Tbf1^{Myc}, strain YFL3164), and vid22 856-901Δ-HA Tbf1-Myc (vid22 CtΔ45^{HA} Tbf1^{Myc}, strain YFL3111) were incubated with 12CA5 HA antibodies. Protein concentration was normalised to 2mg/mL or 20mg/mL (10×) prior to immunoprecipitation. Blots were incubated individually with 12CA5 HA (top panels) or 9E10 Myc (bottom panels) antibodies. The arrows point to the expected bands. Representative immunoblots of N=2 independent experiments. **(C)** Sensitivity to hydroxyurea (HU) of wild-type Vid22 or the C-terminally truncated mutant, compared to *vid22*Δ, evaluated by spot assay. Log-phase wild-type cells (no tag, strain BY4741), Vid22-HA (Vid22, strain YFL3161), *vid22*Δ (*vid22*Δ, strain YNOV124), and vid22 856-901Δ-HA (vid22 CtΔ45, strain YFL3170) were diluted to an optical density at 600 nm (OD_{600}) of 0,06 in H_2O and spotted, along with three 10-fold serial dilutions, on solid complete media devoid of HU (untreated) or supplemented with 50mM or 100mM HU. The plates were photographed after 3 days of incubation at 28°C. Representative outcome of N=3 independent experiments. **(D)** Cell cycle analysis in wild-type Vid22 or the C-terminally truncated mutant, compared to *vid22*Δ. Cell cycle distribution of log-phase, α-factor synchronised and released cultures at different time-points of wild-type cells (no tag, strain BY4741), Vid22-HA (Vid22, strain YFL3161), *vid22*Δ (*vid22*Δ, strain YNOV124), and vid22 856-901Δ-HA (vid22 CtΔ45, strain YFL3170) was determined by flow cytometry quantitation of DNA stained with Sytox Green[415]. Lower fluorescence peaks correspond to G1-phase cells with a haploid DNA content, and higher fluorescence peaks to cells with a doubled DNA content as they have undergone S-phase and replicated their genome. Representative outcome of N=2 independent experiments.

Therefore, to discriminate between these two possibilities, I also assayed co-immunoprecipitation of Tbf1-Myc with vid22 856-901Δ-HA and Vid22-HA (**Figure 8 B**). To overcome the issue of vid22 CtΔ45 instability and precipitate a comparable amount of the full-length and truncated proteins, the experiment was also performed with extracts from Tbf1-Myc vid22 856-901Δ-HA and Tbf1-Myc 10 times more concentrated (10×, 20mg/mL). Tbf1 is virtually absent in both the co-immunoprecipitated fractions of vid22 CtΔ45 (**Figure 8 B**), consistently with the outcome of the previous experiment. Altogether, co-IP analysis of the physical interaction between vid22 CtΔ45 and Tbf1 revealed that the deletion of Vid22 C-terminus does indeed abrogate complex formation, but also impairs the protein's stability. Nonetheless, to investigate the functional relevance of the loss of Vid22-Tbf1 interaction in the context of genome stability maintenance, I assayed the sensitivity of the truncated mutant to replication stress induced by hydroxyurea (HU, **Figure 8 C**). Serial dilutions of mutant yeasts expressing vid22 CtΔ45 were spotted on solid media in absence or in presence of different concentrations of hydroxyurea (50mM and 100mM HU), and their growth compared to that of cells devoid of *VID22* (*vid22*Δ) or with wild-type Vid22 either untagged (no tag) or HA-tagged (Vid22). Qualitative examination of the growth patterns under these conditions reveals noticeable differences in colony size and density between vid22 CtΔ45 and wild-type cells in the hydroxyurea-containing plates, with the former displaying a sensitivity to replication stress comparable to that of *vid22*Δ cells. Finally, to better characterise vid22 856-901Δ phenotypes, I also followed cell cycle progression by flow cytometry, monitoring the DNA content in cells at different time-points after yeast culture synchronisation with α-factor in G1-phase and release from cell cycle arrest (**Figure 8 D**). Cell cycle analysis revealed that vid22 CtΔ45 mutants are impaired in the G1-S transition, evidenced by the high number of cells still in G1 phase after 45min from the release, as *vid22*Δ cells[263] (**Figure 8 D**). Overall, these results suggest that the integrity of Vid22 C-terminus is required for the interaction with Tbf1, which correlates with cell tolerance to genotoxic stress, as well as for normal cell cycle progression. However, deletion of the C-terminal region impairs protein homeostasis, and because of the experimental challenges of dealing with an inherently unstable protein, the data are not fully conclusive. Yet,

because the C-terminus is highly conserved among hAT elements and has been implicated, despite some controversies, in protein-DNA and protein-protein contacts[359,373,393,395,396,416], we deemed worth the pursuit of stable Vid22 C-terminal mutants disrupting its physical association with Tbf1 to better address the relevance of Vid22-Tbf1 complex in genome stability maintenance.

5.4 Targeted mutation of Vid22 conserved C-terminal residues impairs Tbf1 interaction, tolerance to genotoxins and DNA binding

Despite successful in preventing Vid22-Tbf1 interaction, the truncation of Vid22 C-terminus rendered the mutant protein markedly unstable *in vivo*, thereby limiting its utility for subsequent experiments. We envisaged that missense mutations altering the protein primary sequence may have a milder impact on protein homeostasis. Consequently, we redirected our efforts towards the generation of Vid22 point mutants with targeted aminoacidic substitutions that could selectively abolish the formation of the Vid22-Tbf1 complex without affecting protein stability, also opening new avenues for the investigation of the nuanced roles of specific residues in the observed interactions. Given the previous indications of the relevance of Vid22 C-terminus, to identify candidate residues for mutagenesis, we decided to investigate the conservation of this region among Vid22 homologs. Multiple sequence alignments (MSA) can provide important guidance in the identification of the conserved sites for performing site-directed mutagenesis, as they allow to estimate the functional or structural relevance of each residue in a protein sequence[417]. Indeed, amino acids in a protein are subjected to different selection pressures imposed by the requirements of three-dimensional structure and of biological function. Consequently, residues of functional or structural importance tend to be more conserved among homologous sequences, displaying a high frequency in the alignment. A reliable MSA, nonetheless, requires an appropriate input set of evolutionarily related sequences. By the time this work was performed, 220 new yeast genomes belonging to the subphylum Saccharomycotina were sequenced[285], adding up to the 112 genomes already publicly available. To include such recently published genomes in the homology search, I compiled a BLAST database of the protein sequences annotated in these 332 budding yeast genomes and used the *S. cerevisiae* Vid22 protein sequence as the query to run a BLASTp against the local database. As a result, 60 non redundant subject sequences that align locally to the query at or above the standard threshold score were identified as candidate Vid22 homologues (among which its paralogue Env11), deriving from 34 unique yeast species with a paraphyletic distribution according to the published phylogenetic tree[285]. Nine species exhibited a singular homologue,

whereas the remaining 24 species displayed at least two paralogous sequences, stemming from the whole-genome duplication event that occurred ~10 Mya during yeast evolution[285]. I next applied the MUSCLE algorithm[418] to generate the multiple sequence alignment of Vid22 and its yeast homologues (**Figure 9**). Given the study's focus on Vid22 and because of space considerations, a concise version of the multiple sequence alignment is presented, illustrating Vid22 sequence without gaps. In the alignment, conserved amino acids are highlighted, indicating positions with high identity (*i.e.* the fraction of sequences in a column with the same residue) and occupancy (*i.e.* the number of aligned sequences in a column) scores. The reported consensus sequence logo represents the relative frequency of residues per column, which can be estimated by their size in the logo, aiding in the identification of the most common amino acid at each position. The alignment quality annotation is an automatically calculated measure of the likelihood of observing the mutations, if any, in a particular column of the alignment. The alignment quality score is inversely proportional to the average cost of all pairs of mutations observed in each column, so a high quality score implies that there are no or few mutations, or that most mutations observed are favourable (**Figure 9**). Scanning the alignment from the N-terminus, the first appreciable conserved feature in most sequences is the signature Cys and His dyads of the BED domain predicted to form the zinc finger[341], with a strong $CKHCX_nHLX_3H$ consensus in these yeasts, anticipated by conserved aromatic positions (Trp, Phe, Tyr; **Figure 9** block 2 of 18, residues 50-99; and block 3 of 18, residues 100-149). Vid22 C87, C90, and H115 display 89% identity in the MSA, H110 90% identity (**Figure 10 A**). The consistency between the residues in Vid22 and Env11 forming the BED signature identified here and previously published[341] provides a good indication of the reliability of the alignment.

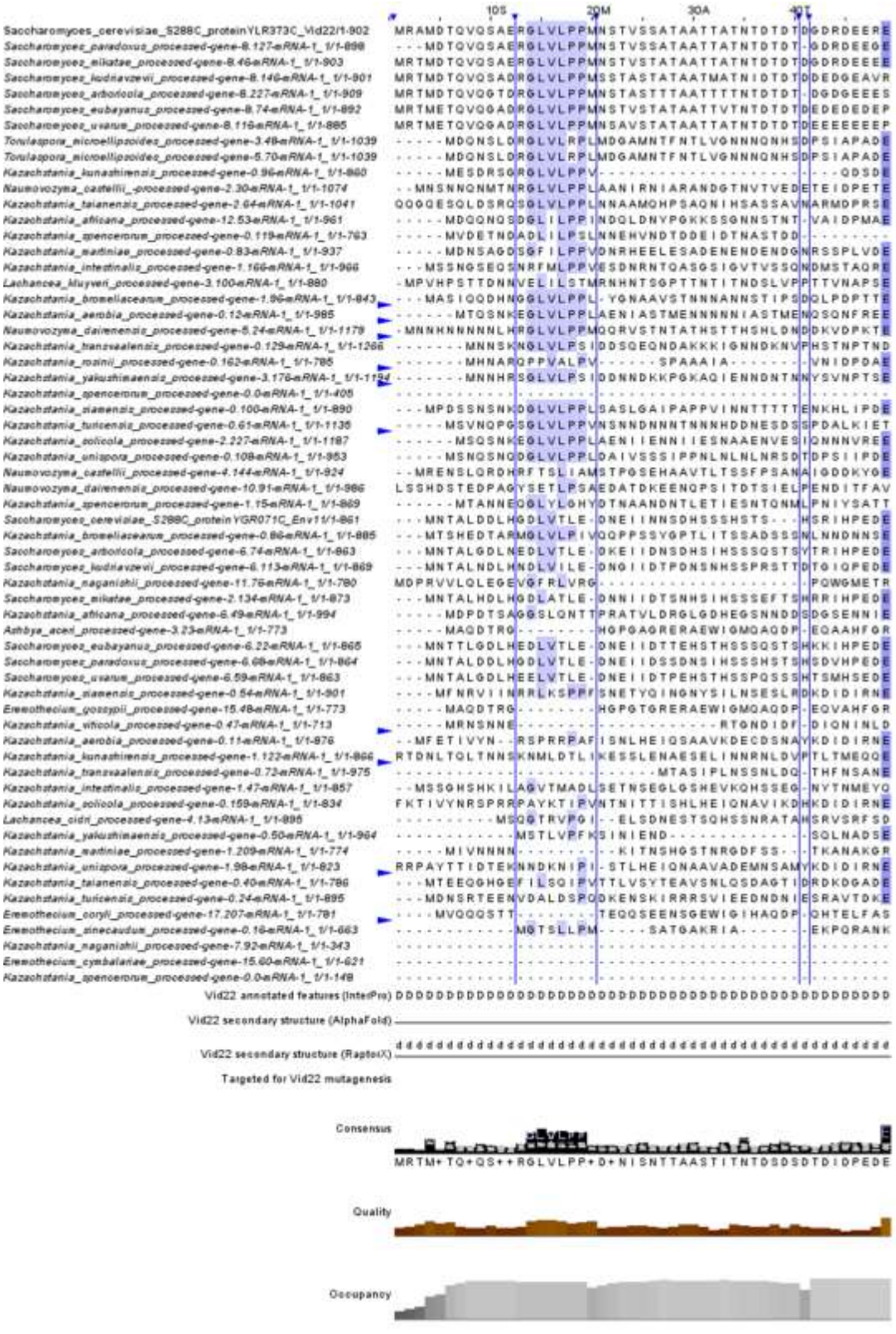


Figure 9 – Multiple sequence alignment of Vid22 and its homologues in Saccharomycotina species (block 1 of 18, residues 1-49)

(Figure legend at page 80)

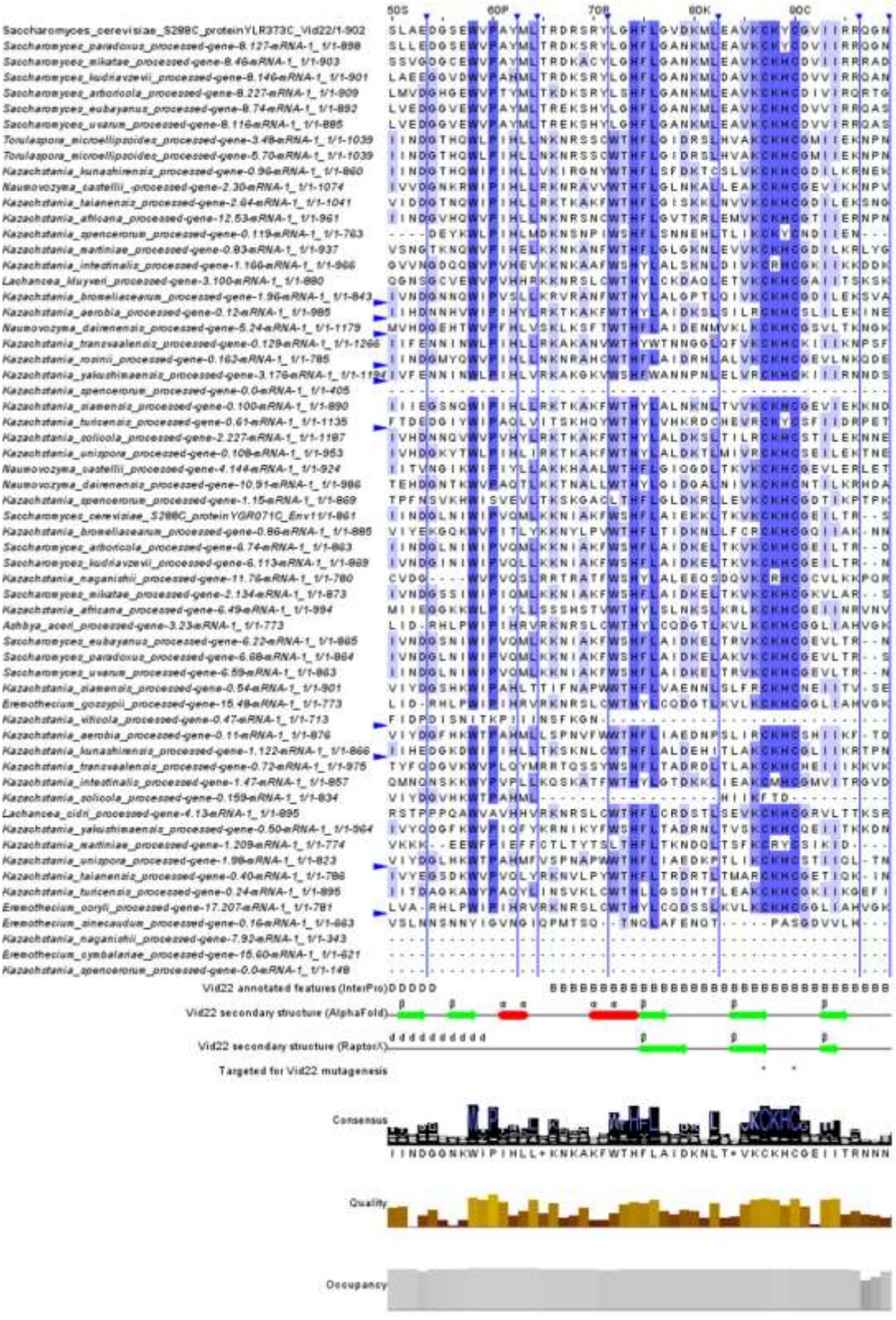


Figure 9 – Multiple sequence alignment of Vid22 and its homologues in Saccharomycotina species (block 2 of 18, residues 50-99)

(Figure legend at page 80)

Figure 9 – Multiple sequence alignment of Vid22 and its homologues in Saccharomycotina species (block 3 of 18, residues 100-149)

(Figure legend at page 80)

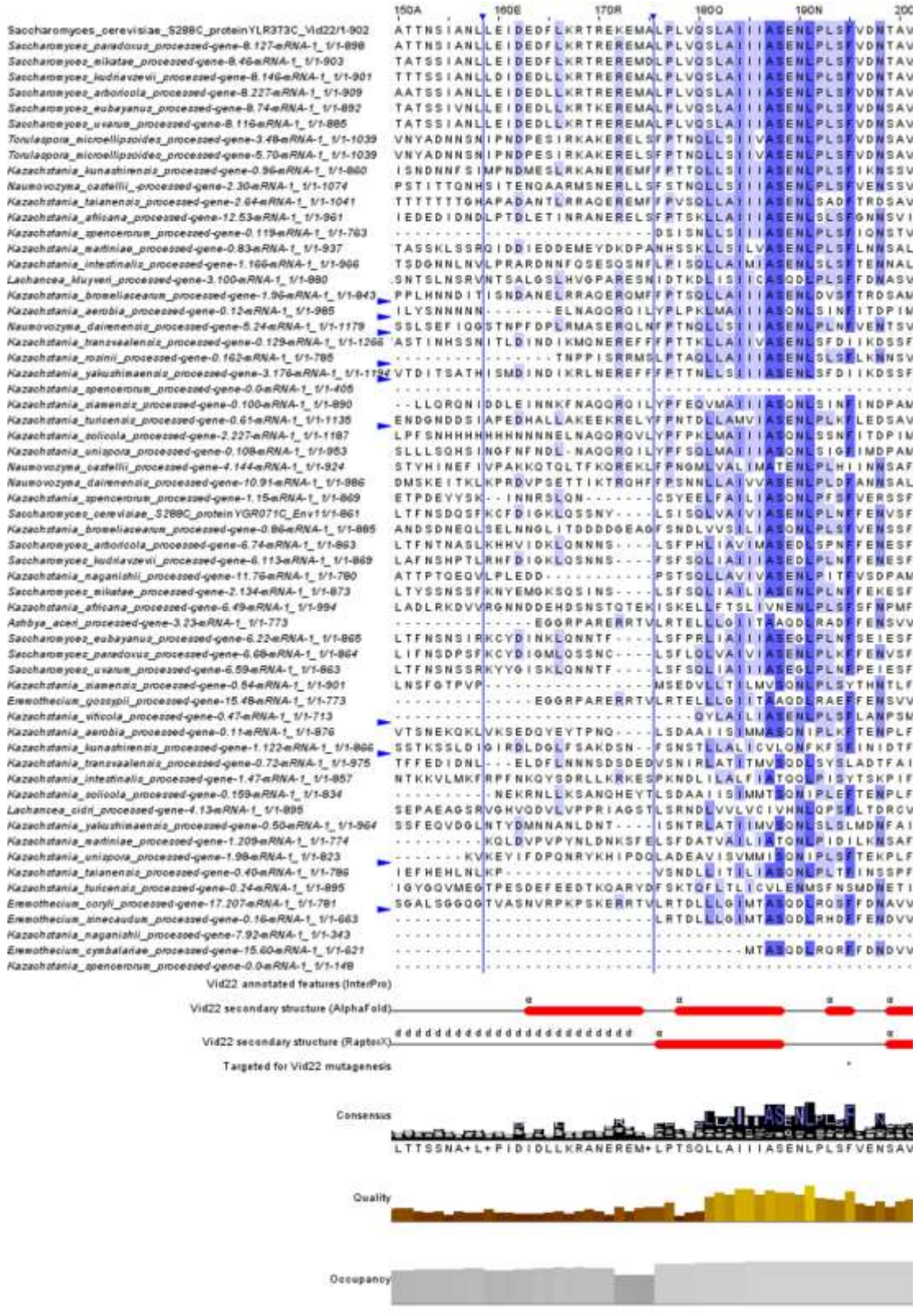


Figure 9 – Multiple sequence alignment of Vid22 and its homologues in Saccharomycotina species (block 4 of 18, residues 150-199)

(Figure legend at page 80)

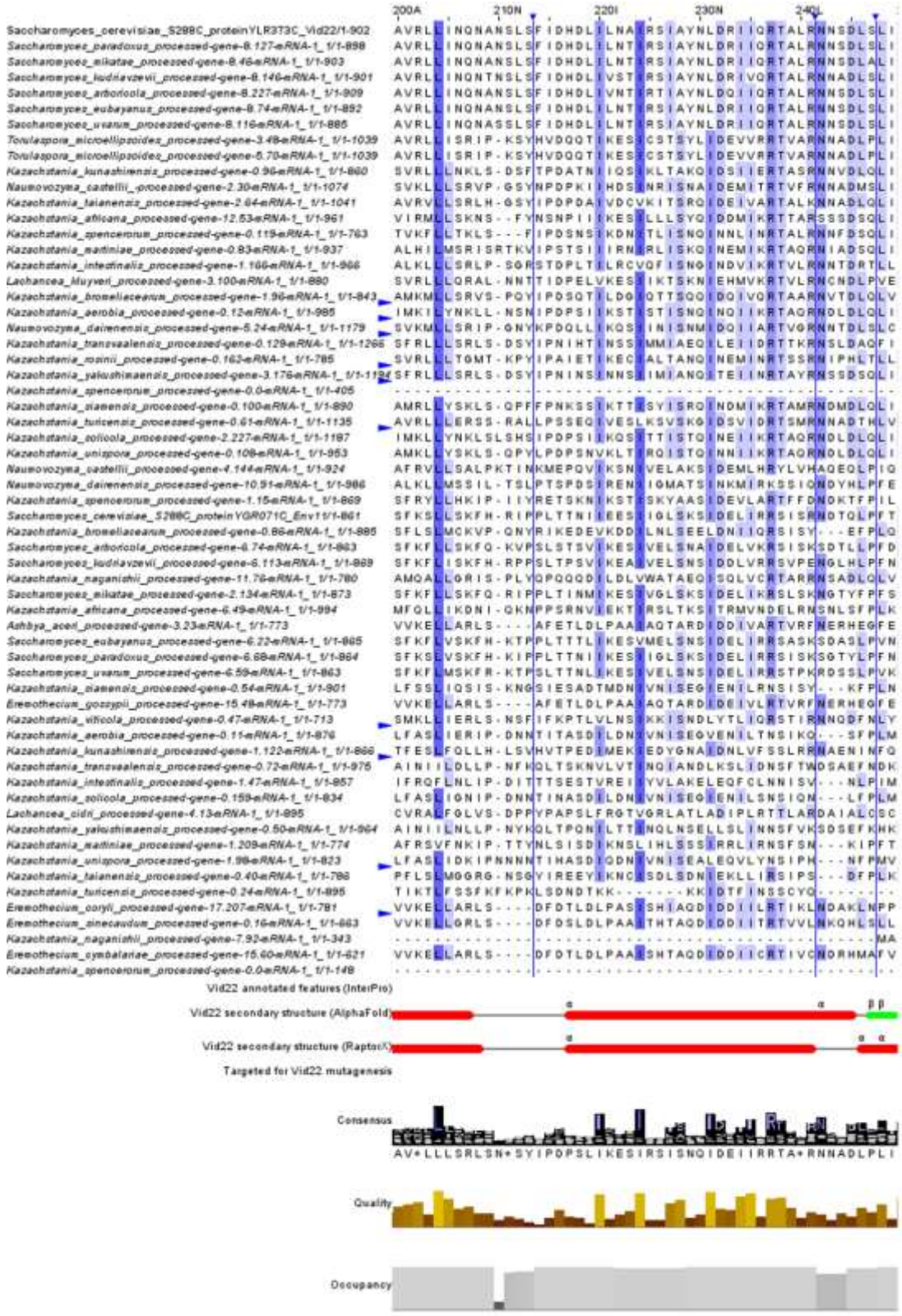


Figure 9 – Multiple sequence alignment of Vid22 and its homologues in Saccharomycotina species (block 5 of 18, residues 200-249)

(Figure legend at page 80)

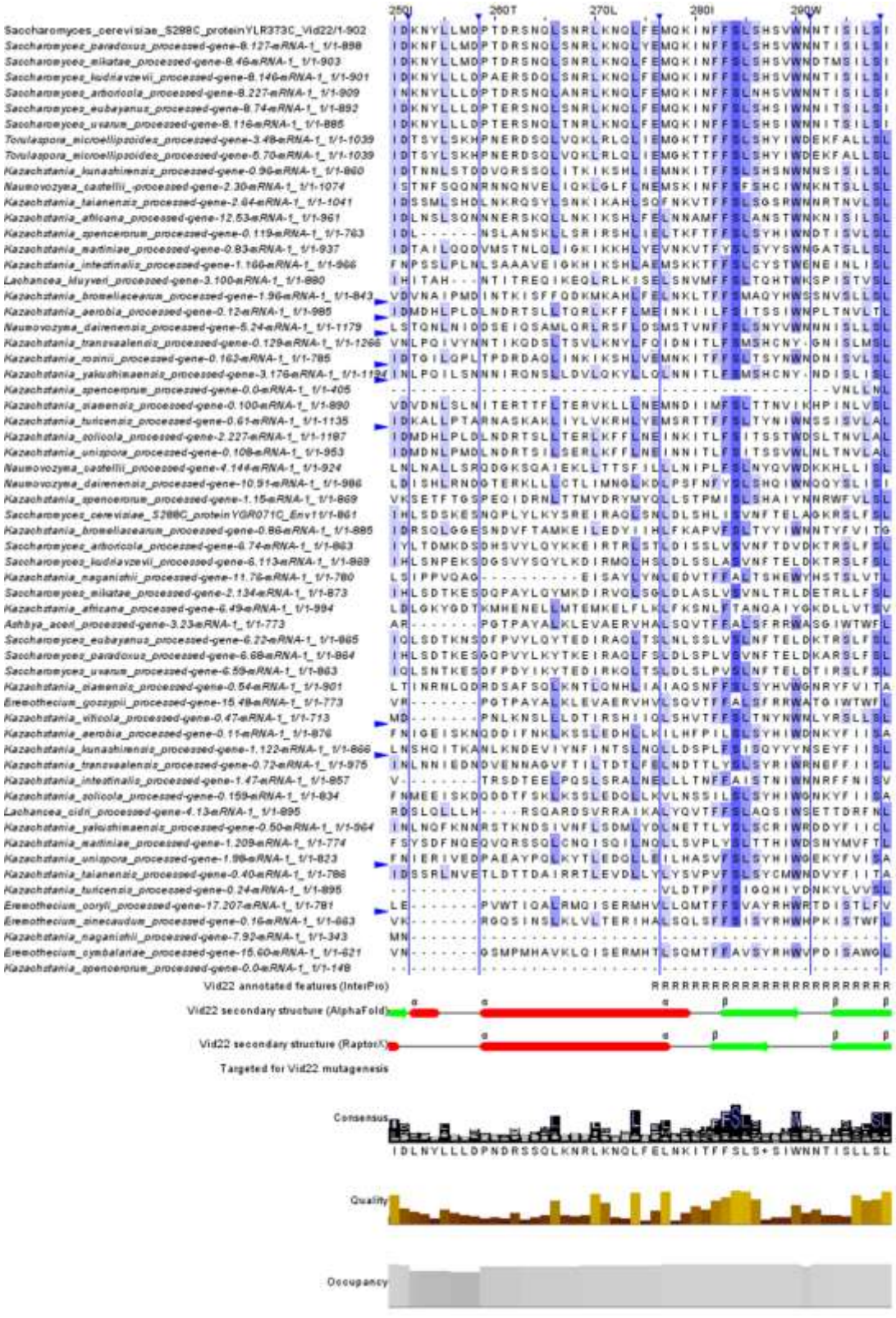


Figure 9 – Multiple sequence alignment of Vid22 and its homologues in Saccharomycotina species (block 6 of 18, residues 250-299)

(Figure legend at page 80)

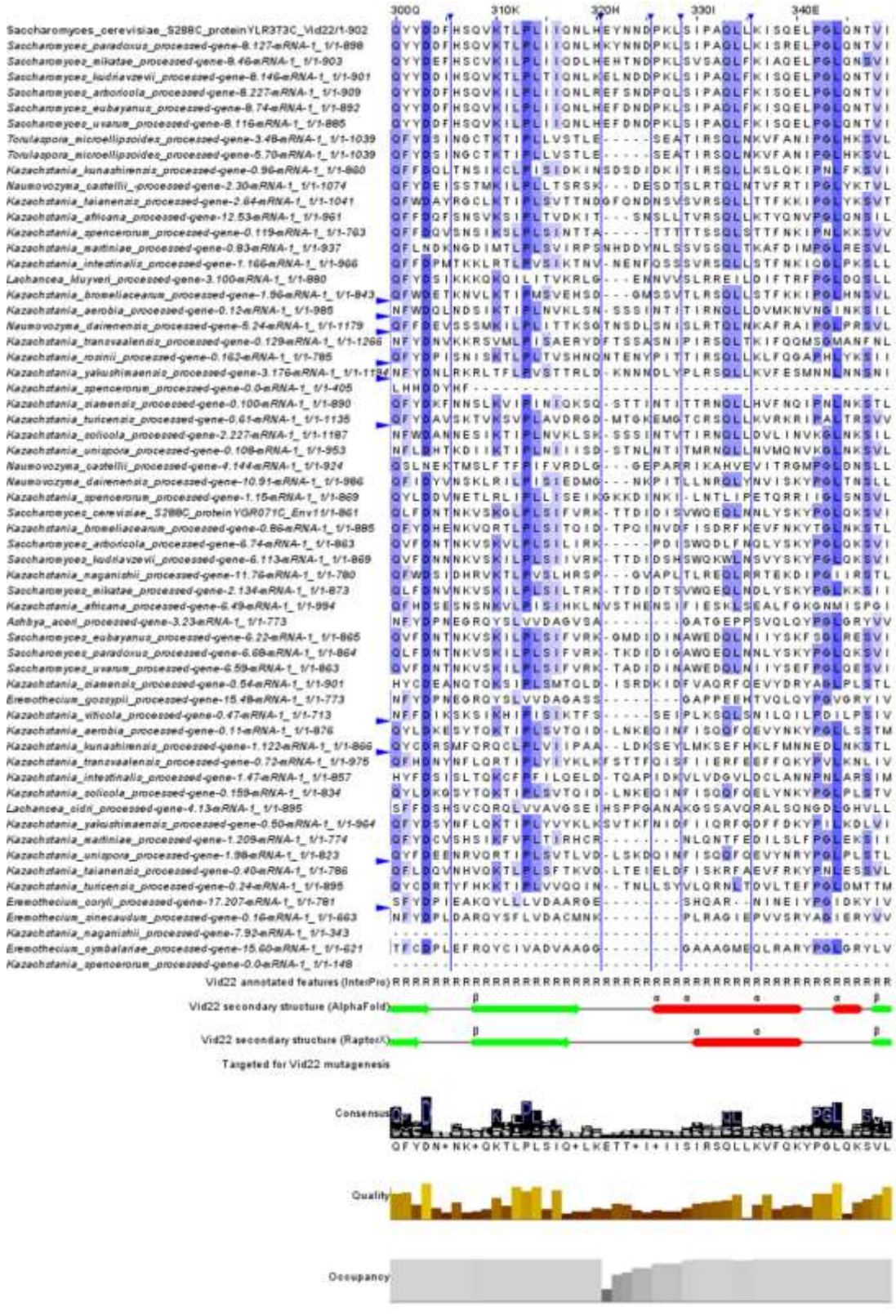


Figure 9 – Multiple sequence alignment of Vid22 and its homologues in Saccharomycotina species (block 7 of 18, residues 300-349)

(Figure legend at page 80)

Figure 9 – Multiple sequence alignment of Vid22 and its homologues in Saccharomycotina species (block 8 of 18, residues 350-399)

(Figure legend at page 80)

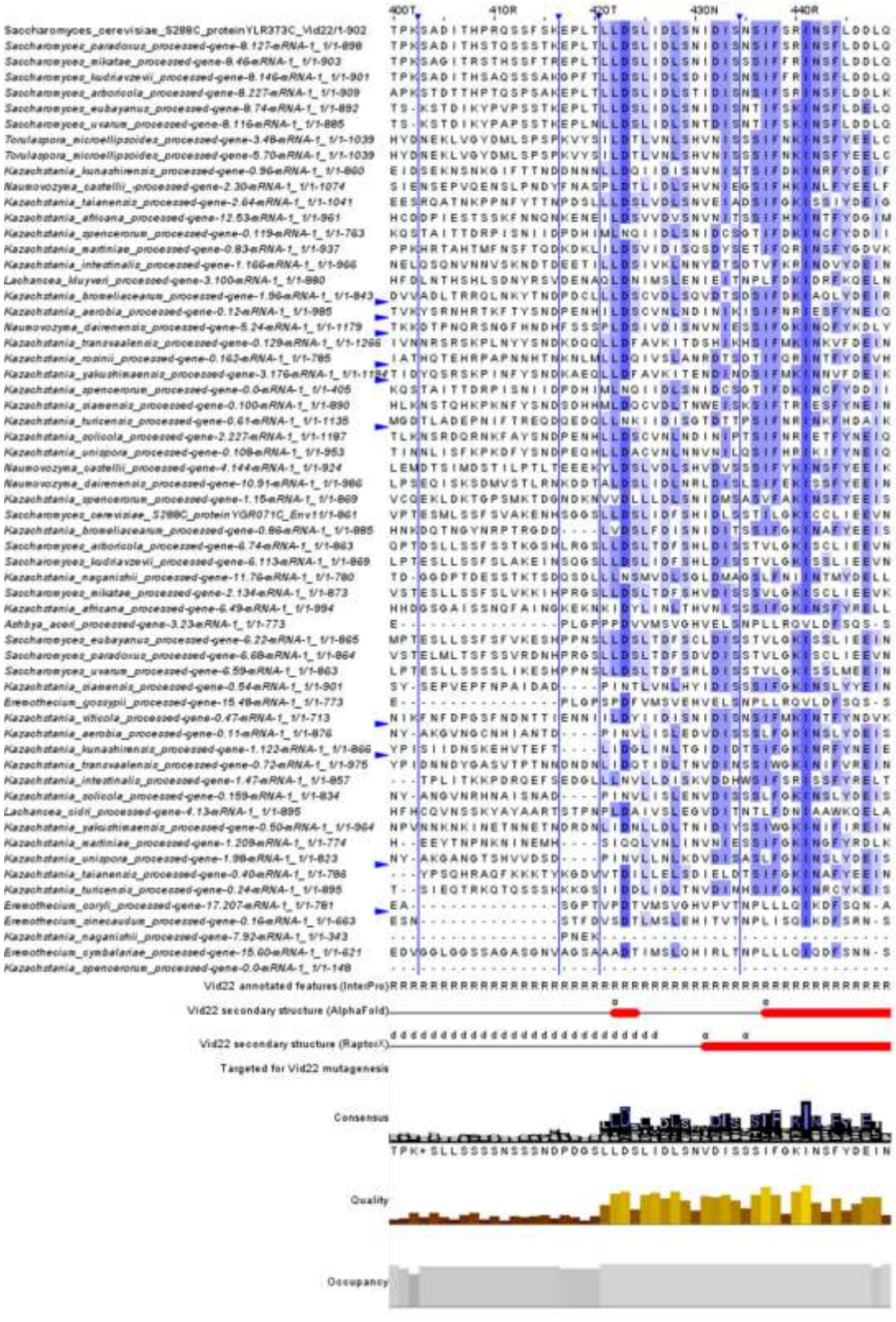


Figure 9 – Multiple sequence alignment of Vid22 and its homologues in Saccharomycotina species (block 9 of 18, residues 400-449)

(Figure legend at page 80)

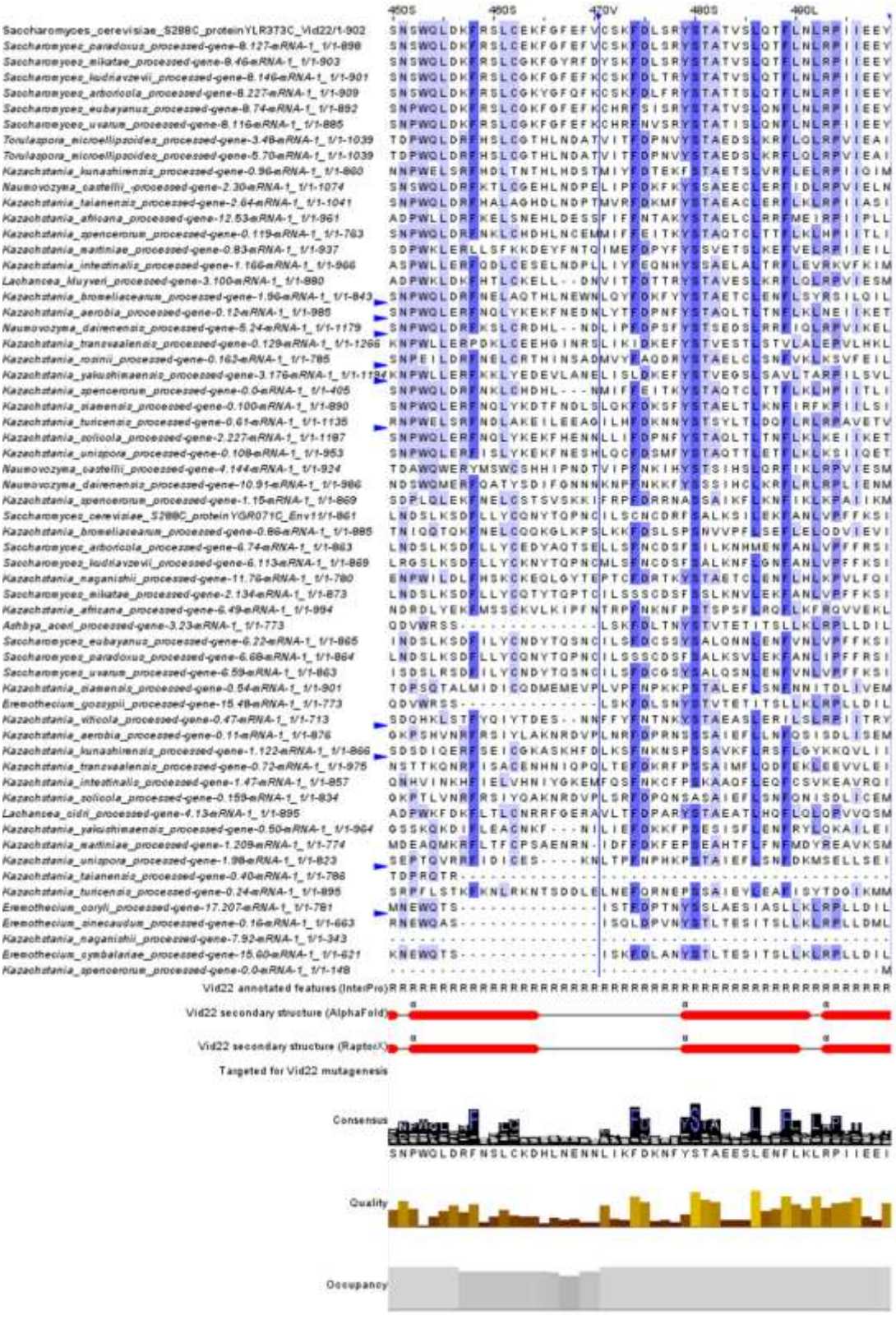


Figure 9 – Multiple sequence alignment of Vid22 and its homologues in Saccharomycotina species (block 10 of 18, residues 450-499)

(Figure legend at page 80)

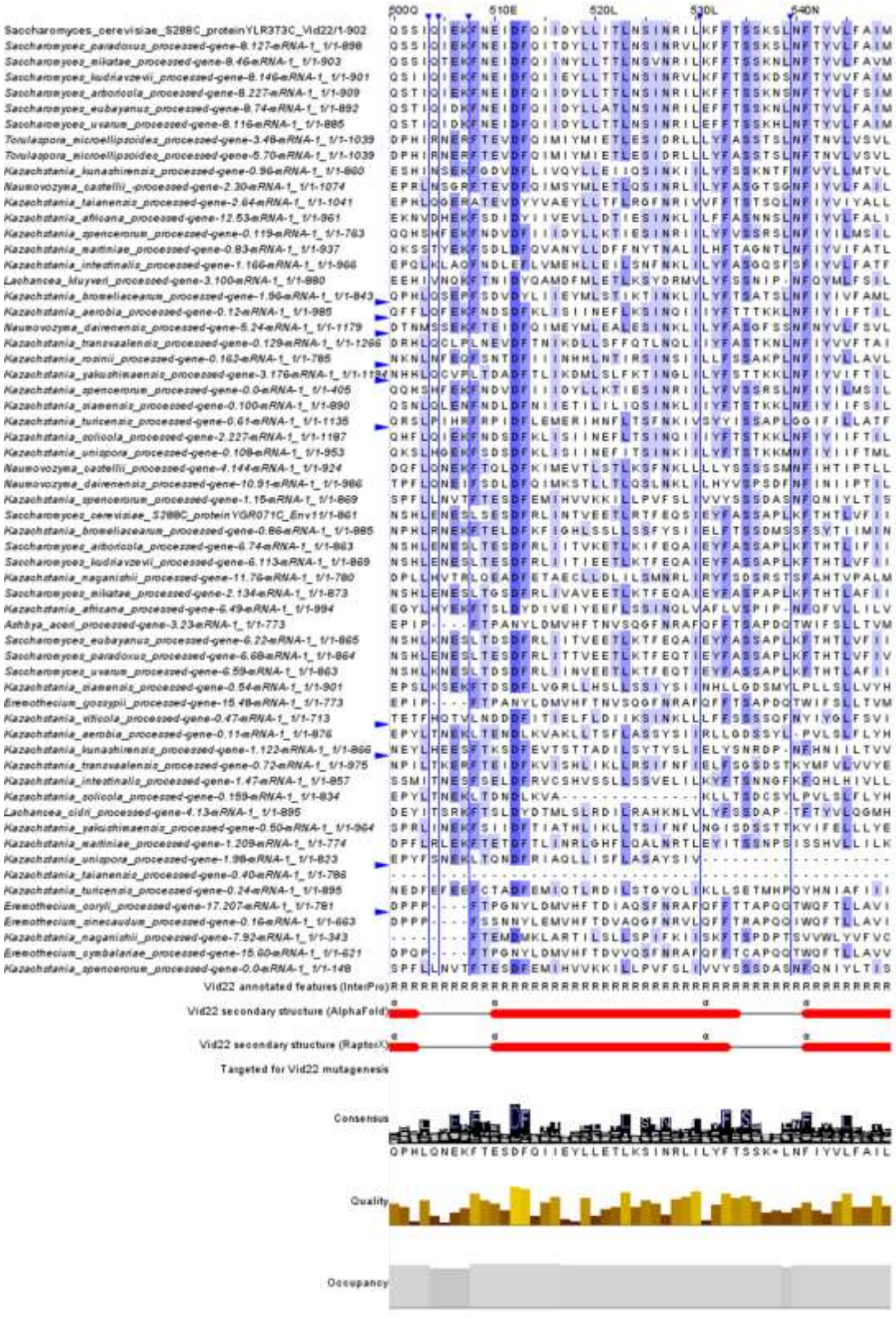


Figure 9 – Multiple sequence alignment of Vid22 and its homologues in Saccharomycotina species (block 11 of 18, residues 500-549)

(Figure legend at page 80)

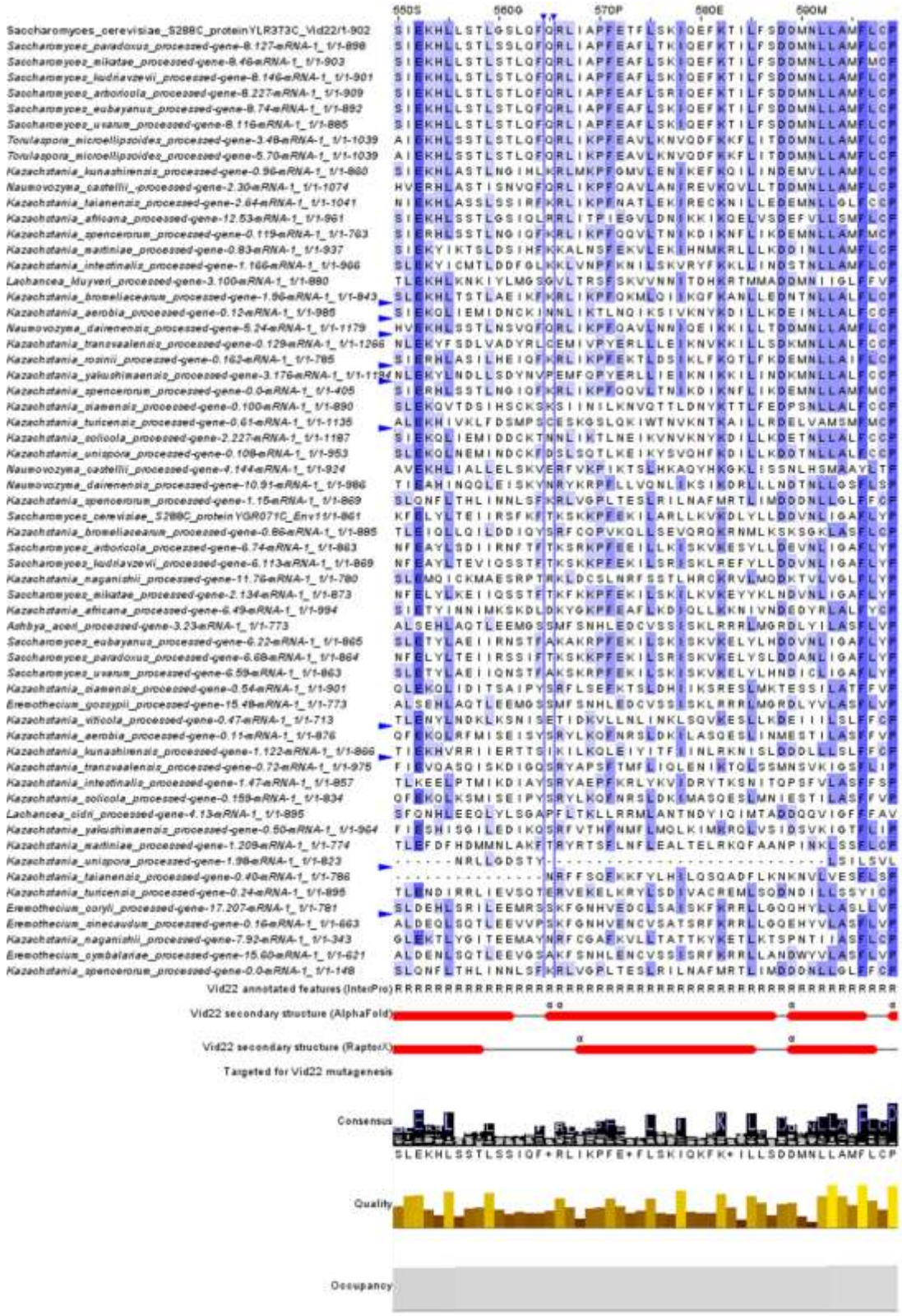


Figure 9 – Multiple sequence alignment of Vid22 and its homologues in Saccharomycotina species (block 12 of 18, residues 550-599)

(Figure legend at page 80)

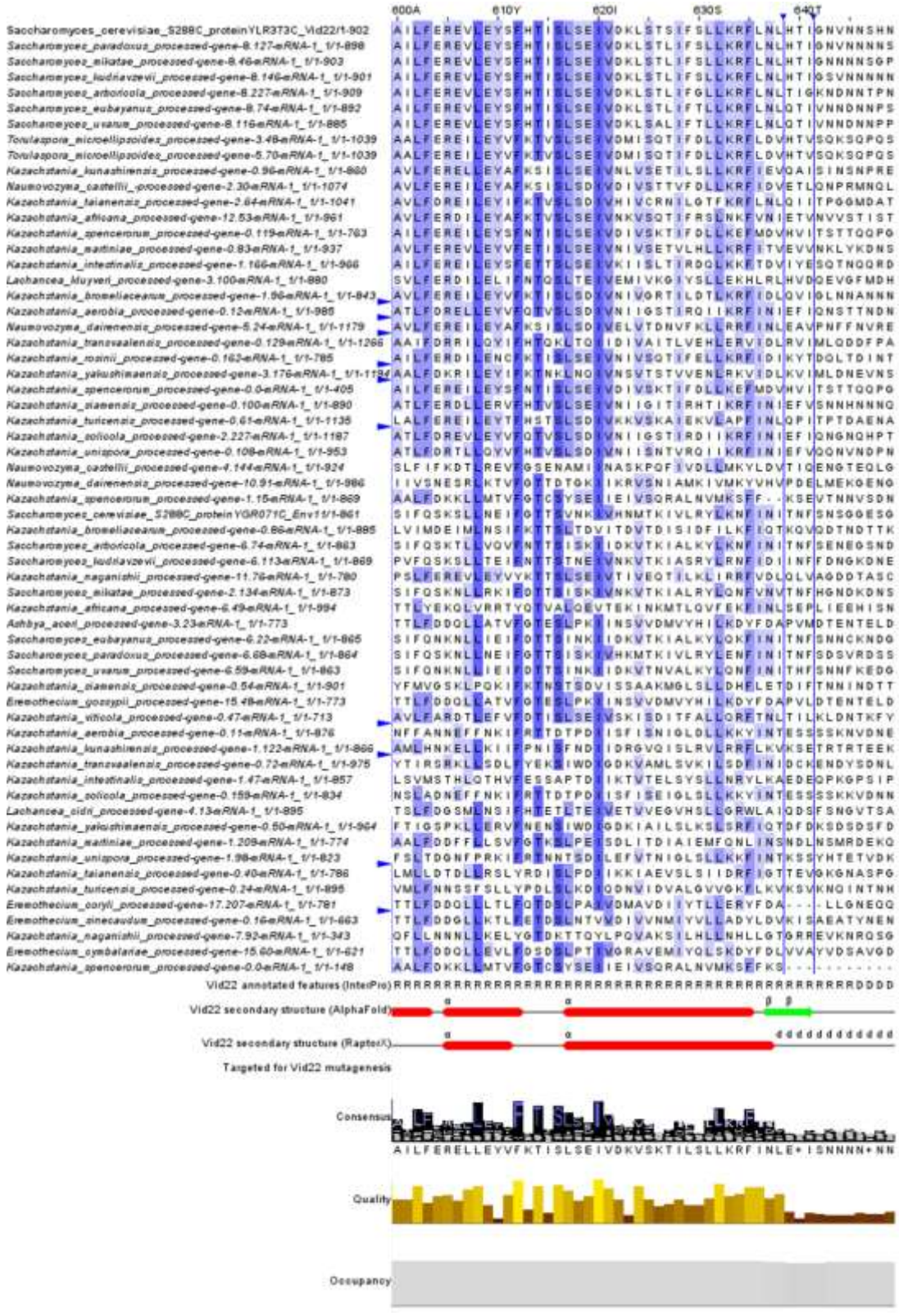


Figure 9 – Multiple sequence alignment of Vid22 and its homologues in Saccharomycotina species (block 13 of 18, residues 600-649)

(Figure legend at page 80)

Figure 9 – Multiple sequence alignment of Vid22 and its homologues in Saccharomycotina species (block 14 of 18, residues 650-699)

(Figure legend at page 80)

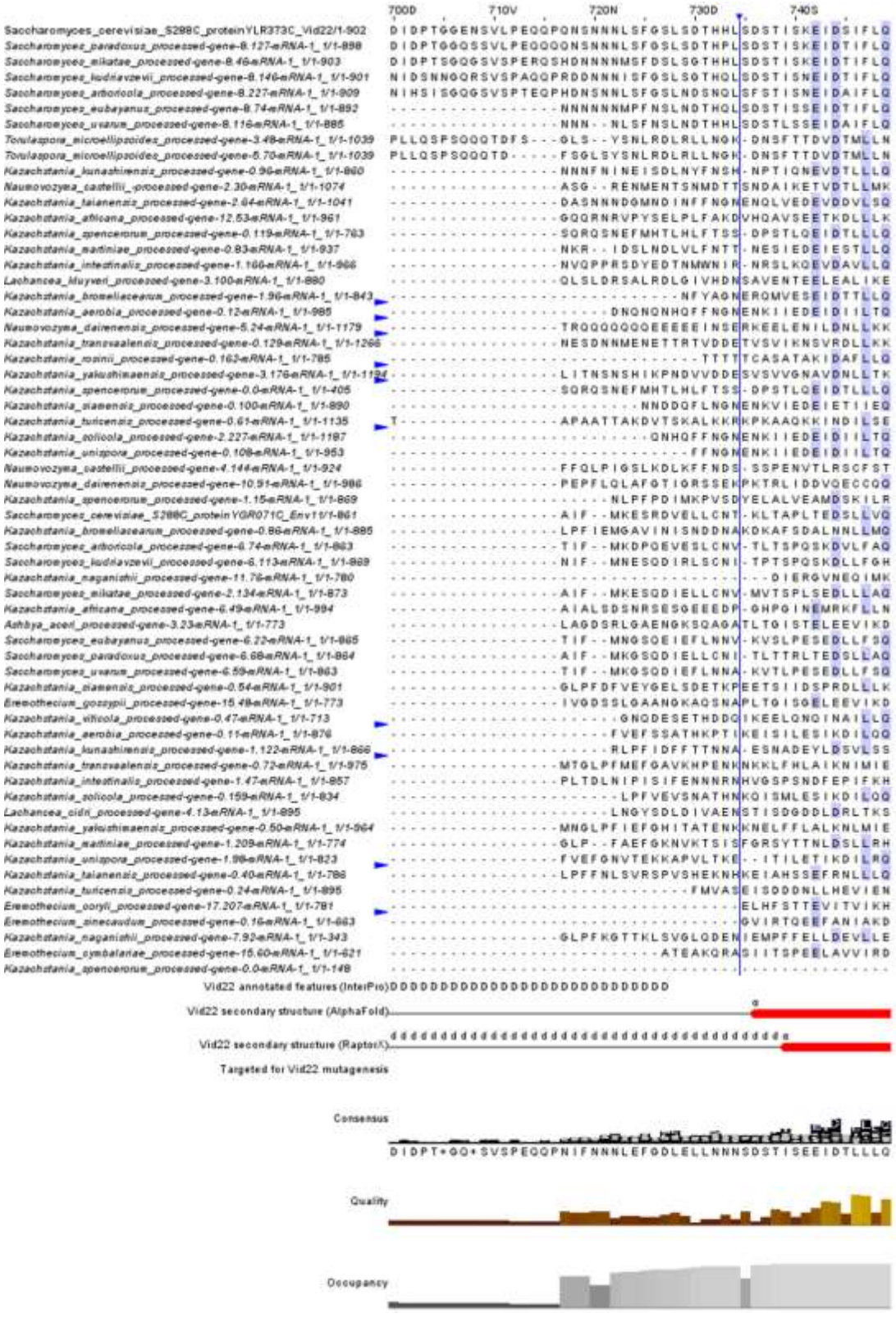


Figure 9 – Multiple sequence alignment of Vid22 and its homologues in Saccharomycotina species (block 15 of 18, residues 700-749)

(Figure legend at page 80)

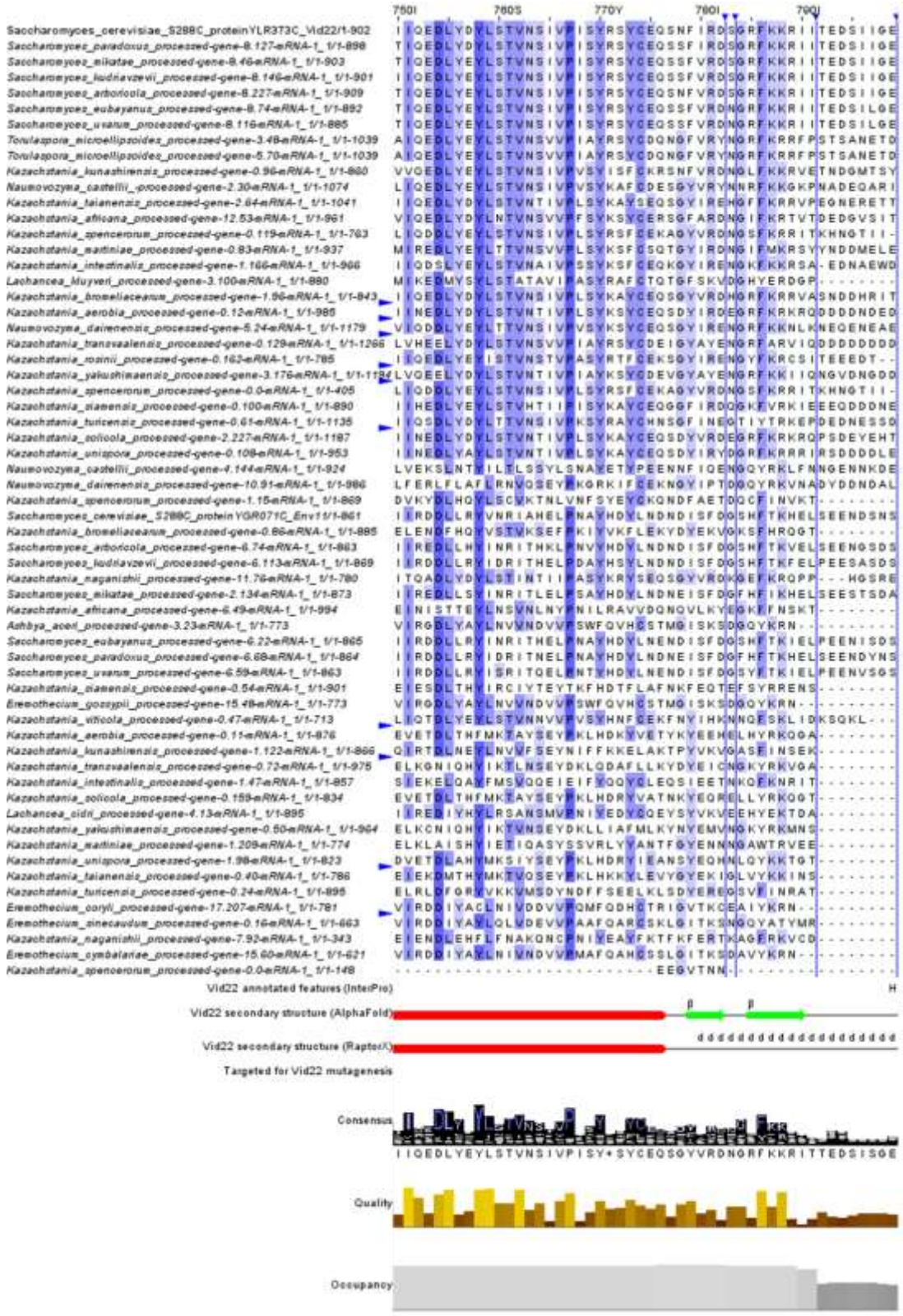


Figure 9 – Multiple sequence alignment of Vid22 and its homologues in Saccharomycotina species (block 16 of 18, residues 750-799)

(Figure legend at page 80)

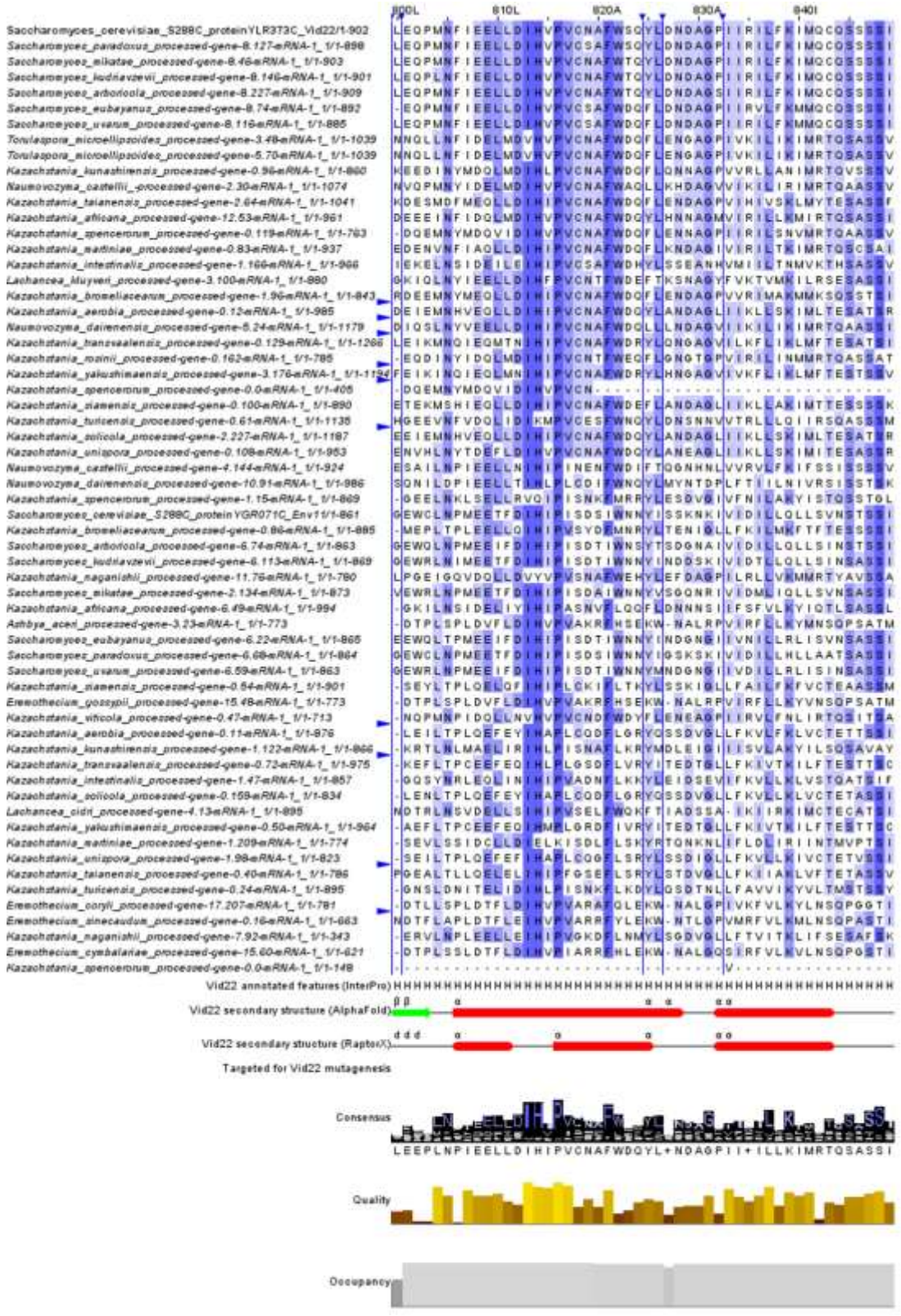


Figure 9 – Multiple sequence alignment of Vid22 and its homologues in Saccharomycotina species (block 17 of 18, residues 800-849)

(Figure legend at page 80)

Figure 9 – Multiple sequence alignment of Vid22 and its homologues in Saccharomycotina species (block 18 of 18, residues 850-901)

Multiple sequence alignment (MSA) of Vid22 protein sequence and its 60 homologous se-

quences in Saccharomycotina species. Residues are automatically blue coloured proportionally to percentage identity in each alignment column. Columns containing gaps in Vid22 and rows with redundant sequences were hidden. Vid22 annotated features (disordered=D; BED=B; RNH-like=R; hAT=H), secondary structure predictions (alpha helix=α; beta sheet=β), residues targeted for Vid22 mutagenesis (point mutation=*; deletion=-), the alignment consensus logo (automatic calculation of the relative amount of residues per column, proportional to its size in the logo), the alignment quality (automatic measure of the likelihood of observing the mutations in a particular column of the alignment calculated from BLOSUM62 scores), and occupancy (number of residues aligned at each column) are reported below.

Focusing on the C-terminus, the segment of the multiple sequence alignment covering Vid22 residues 856-901 (**Figure 9** block 18 of 18, residues 850-901) is well conserved, evident from the overall high sequence similarity and occupancy scores, with a recognizable consensus sequence, suggesting its potential functional and/or structural importance. L856, at position -45 from the end of Vid22 sequence, exhibits 90% identity, and, following a short stretch of low sequence similarity, other conserved amino acids with high identity include the negatively charged K873 (85% identity, with an additional 2% of the sequence conserving the overall physicochemical properties with R substitutions) and K875 (84% identity, with 11% of the sequences displaying R), interspersed with non-polar amino acids, like Ile and Leu (**Figure 9** block 18 of 18, residues 850-901). Further on, other highlighted positions with identity higher than 75% consist of polar residues, such as N878 (79% identity) and Q880 (97% identity), and the aromatic F881 (77% identity, and an additional 11% of sequences with Y), followed at short distance by charged residues, for instance D887 (77% identity, and 2% with E). L893 marks the final conserved position in this region, possessing hydrophobic properties (84% identity, plus 10% with I and 2% with V; **Figure 9** block 18 of 18, residues 850-901). Therefore, we selected for mutagenesis most residues from this pool of highly conserved positions in the last 45 amino acids of Vid22 C-terminus, targeting (i) the two close K873 and K875 simultaneously, (ii) N878, (iii) Q880, (iv) D887 along with the proximal D889 (43% identity, and 23% of sequences with E) and D891 (64% identity), and finally (v) L893 (**Figure 10 B**). All targeted amino acids were

substituted to A, except L893 (**Figure 10 B**). The latter was mutated to P instead, because it was reported that mutation of Leu residues in Hermes C-terminus to Ala, had no phenotypic effect, possibly because of their similar biochemical properties, contrarily to what observed for Pro substitutions[397]. While we decided to initially focus on this 45-amino acid window to select residues for site-specific mutagenesis based on the results of the previous genetic screening, it's interesting to note that other blocks of the alignment are characterized by high sequence similarity, providing the opportunity to explore the functionality of other regions of the protein in a targeted manner **(Figure 9**).

Following the choice of the aminoacidic substitutions for targeted mutagenesis, I first assessed whether the overexpression of such mutant alleles would rescue lethality (**Figure 10 C**). To this end, *VID22* coding sequence, cloned in a centromeric plasmid in frame with an N-terminal HA epitope sequence and under the control of a galactose-inducible promoter, was specifically mutagenized by PCR[419] to introduce the desired aminoacidic substitutions by the means of custom oligos. Following sequencing, the overexpression vectors were transformed in wild-type haploid yeasts. I then evaluated by spot assay the survival of cells conditionally producing N-terminally HA-tagged vid22 K873A K875A (vid22 KK), vid22 N878A (vid22 N), vid22 Q880A (vid22 Q), vid22 D887A D889A D891A (vid22 DDD), and vid22 L893P (vid22 L/P), along with the control strains expressing the wild-type (Vid22) or with the empty vector (empty), plated on a galactose-containing medium to induce overexpression and on a raffinose-containing medium for the repressed condition (**Figure 10 C**). Whereas none of the mutant alleles appreciably affects cell growth on raffinose-containing non-inducing medium, yeasts overproducing vid22 KK or vid22 DDD mutant proteins are viable on galactose-containing inducing medium, comparably to yeasts transformed with the empty vector, whereas cells accumulating vid22 L/P display an intermediate phenotype (**Figure 10 C**).

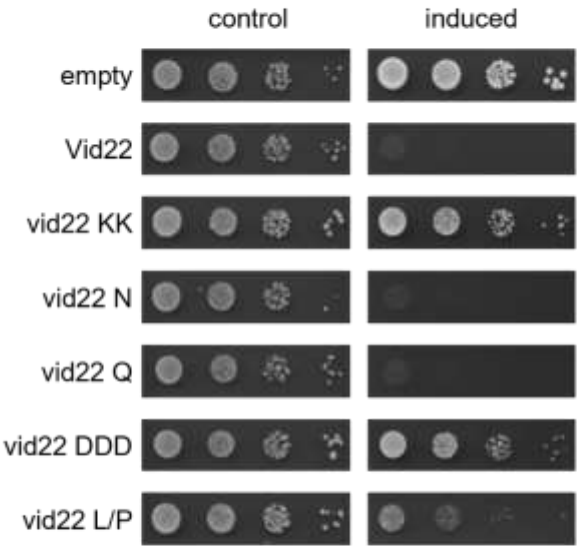
A

Mutant	*Residue*	*Identity*	*Similarity*	*Substitution*
vid22 BED	C87	89%	-	A
	C90	89%	-	R
	H110	90%	-	Y
	H115	89%	-	Y

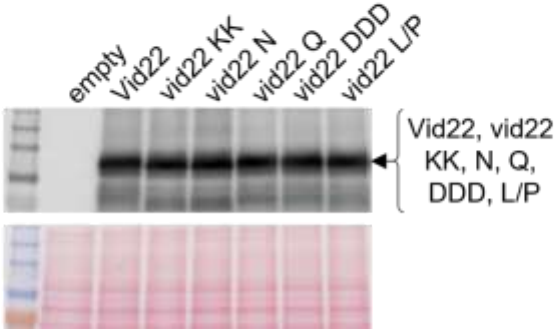
B

Mutant	*Residue*	*Identity*	*Similarity*	*Substitution*
vid22 KK	K873	85%	87% (R)	A
	K875	84%	95% (R)	A
vid22 N	N878	79%	82% (S)	A
vid22 Q	Q880	97%	-	A
vid22 DDD	D887	77%	79% (E)	A
	D889	43%	65% (E)	A
	D891	64%	-	A
vid22 L/P	L893	84%	96% (I, V)	P

C



D



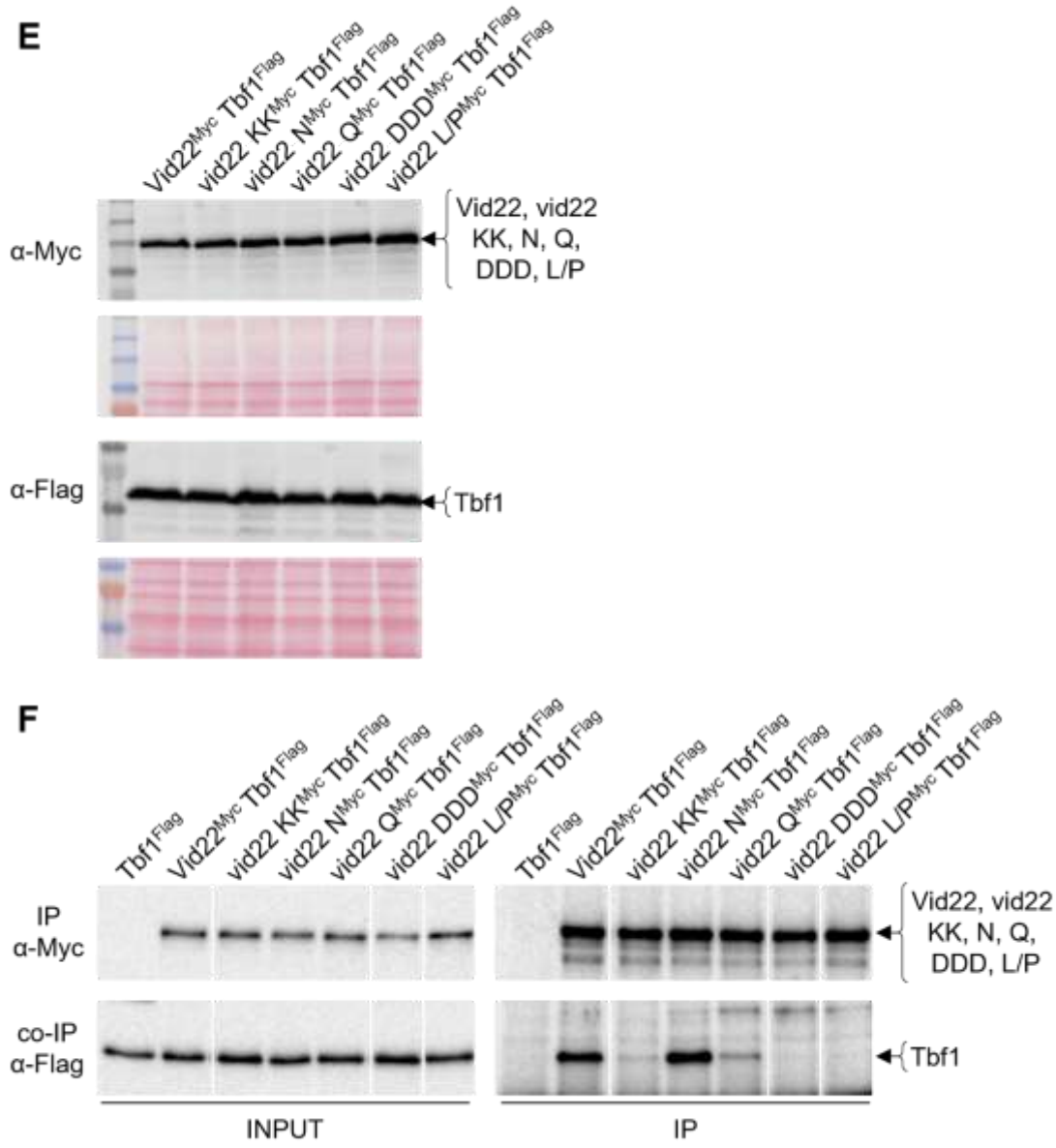


Figure 10 – Vid22 C-terminal mutations targeting conserved residues that suppress overexpression toxicity also compromise interaction with Tbf1

(A) Conservation of the residues mutated in Vid22 BED domain among Vid22 homologues in Saccharomycotina species. **(B)** Conservation of the residues mutated in Vid22 C-terminal region among Vid22 homologues in Saccharomycotina species. **(C)** Yeast sensitivity to overexpression of *VID22* or mutant alleles targeting conserved residues in the C-terminal region evaluated by spot assay. Log-phase wild-type cells transformed with the empty vector (empty, strain YGE722.1), transformed with the overexpression vector of wild-type Vid22 (Vid22, strain YGE723.1) or of vid22 K873A K875A (vid22 KK, strain YFL3683.1), vid22 N878A (vid22 N, strain YFL3684.1), vid22 Q880A (vid22 Q, strain YFL3685.1), vid22 D887A D889A D891A (vid22 DDD, strain YFL3687.1), vid22 L893P (vid22 L/P, strain

YFL3689.1) were diluted to an optical density at 600 nm (OD_{600}) of 0,06 in H_2O and spotted, along with three 10-fold serial dilutions, on solid plasmid-selective media with 2% glucose as a control (control) or 2% raffinose, 2% galactose to induce allele overexpression (induced). The plates were photographed after 3 days of incubation at 28°C. Representative outcome of N=3 independent experiments. **(D)** Western blot analysis of the levels of wild-type Vid22 or mutated in conserved C-terminal residues. Total proteins were extracted from normalised cultures of wild-type cells transformed with the empty vector (empty, strain YGE722.1), transformed with the overexpression vector of wild-type HA-tagged Vid22 (Vid22, strain YGE723.1) or of vid22 K873A K875A (vid22 KK, strain YFL3683.1), vid22 N878A (vid22 N, strain YFL3684.1), vid22 Q880A (vid22 Q, strain YFL3685.1), vid22 D887A D889A D891A (vid22 DDD, strain YFL3687.1), vid22 L893P (vid22 L/P, strain YFL3689.1) grown in plasmid-selective media with 2% raffinose, 2% galactose to induce allele overexpression. The blot was incubated with 12CA5 HA antibodies. Arrows point to the expected bands. Ponceau staining is shown as loading control. Representative blot of N=3 independent experiments. **(E)** Western blot analysis of the levels of wild-type Vid22 or mutated in conserved C-terminal residues and Tbf1. Total proteins were extracted from normalised cultures of Vid22-Myc Tbf1-Flag ($Vid22^{Myc}$ $Tbf1^{Flag}$, strain YFL3993), vid22 K873A K875A-Myc Tbf1-Flag ($vid22\ KK^{Myc}$ $Tbf1^{Flag}$, strain YFL3995), vid22 N878A-Myc Tbf1-Flag ($vid22\ N^{Myc}$ $Tbf1^{Flag}$, strain YFL3996), vid22 Q880A-Myc Tbf1-Flag ($vid22\ Q^{Myc}$ $Tbf1^{Flag}$, strain YFL3997), vid22 D887A D889A D891A-Myc Tbf1-Flag ($vid22\ DDD^{Myc}$ $Tbf1^{Flag}$, strain YFL3999), and vid22 L893P-Myc Tbf1-Flag ($vid22\ L/P^{Myc}$ $Tbf1^{Flag}$, strain YFL4001) grown in complete media. The top blot was incubated with 9E10 Myc antibodies, the bottom blot with M2 Flag antibodies. Arrows point to the expected bands. Ponceau staining is shown as loading control. Representative blot of N=3 independent experiments. **(F)** Western blot analysis of co-IP experiments of Tbf1 with wild-type Vid22 or mutated in conserved C-terminal residues. Normalised protein extracts of Tbf1-Flag ($Tbf1^{Flag}$, strain YFL3128), Vid22-Myc Tbf1-Flag ($Vid22^{Myc}$ $Tbf1^{Flag}$, strain YFL3993), vid22 K873A K875A-Myc Tbf1-Flag ($vid22\ KK^{Myc}$ $Tbf1^{Flag}$, strain YFL3995), vid22 N878A-Myc Tbf1-Flag ($vid22\ N^{Myc}$ $Tbf1^{Flag}$, strain YFL3996), vid22 Q880A-Myc Tbf1-Flag ($vid22\ Q^{Myc}$ $Tbf1^{Flag}$, strain YFL3997), vid22 D887A D889A D891A-Myc Tbf1-Flag ($vid22\ DDD^{Myc}$ $Tbf1^{Flag}$, strain YFL3999), and vid22 L893P-Myc Tbf1-Flag ($vid22\ L/P^{Myc}$ $Tbf1^{Flag}$, strain YFL4001) were incubated with 9E10 Myc antibodies. Blots were incubated individually with 9E10 Myc (top panels) or M2 Flag (bottom panels) antibodies. The arrows point to the expected bands. Representative immunoblots of N=3 independent experiments.

Accumulation of vid22 N or vid22 Q mutants, by contrast, causes lethality as for wild-type Vid22. Semi-quantitative Western blot analysis of protein accumulation in cells grown in galactose confirmed that all mutant proteins displayed a signal comparable to that of Vid22 (**Figure 10 D**).

To understand if suppression of overexpression toxicity by the mutant alleles correlates with the resulting protein having impaired interaction with Tbf1, I next introduced the point mutations targeting the chosen Vid22 C-terminal residues, along with the Myc tag, in *VID22* encoded at its endogenous locus under the control of its native promoter by gene replacement through homologous recombination, providing cells with PCR products amplified from plasmid templates subjected to *in vitro* mutagenesis. Sequencing of the genomic locus allowed to validate the integration of the desired mutations both in wild-type and in Tbf1-Flag backgrounds. Moreover, steady-state levels of all mutant proteins were comparable to that of Vid22, as evaluated by Western blot (**Figure 10 E**), indicating that, unlike the deletion of the C-terminal portion, the selected mutations do not affect protein stability. In addition, Tbf1 levels are unaltered by Vid22 mutation (**Figure 10 E**). I thus proceeded to analyse co-immunoprecipitation of Tbf1-Flag with Myc-tagged wild-type Vid22 or with the C-terminally mutated variants in whole cell extracts from the obtained strains (**Figure 10 F**). Despite immunoprecipitation levels of wild-type or C-terminally mutated Vid22 being similar, Tbf1 is virtually undetectable in the coprecipitated fractions of vid22 K873A K875A, vid22 D887A D889A D891A, and vid22 L893P mutants, implying that these substitutions abrogate the proteins' physical interaction with Tbf1. Reduced Tbf1 levels in the co-immunoprecipitated fraction of vid22 Q880A is indicative of an impaired interaction. On the contrary, mutation of N878 does not affect protein-protein interactions, as Tbf1 co-precipitates with vid22 N878A to a comparable level to that co-precipitated with wild-type Vid22 (**Figure 10 F**). The comparison of these data with the results of the previous spot assays (**Figure 10 C**) suggests the existence of a correlation between overexpression toxicity and Tbf1 physical interaction phenotypes, as the mutant alleles non-lethal when overexpressed produce proteins completely incompetent for Tbf1 interaction, whereas the lethal ones give rise to proteins that retain either full or partial interaction ability.

I continued with the phenotypic characterization of Vid22 C-terminal mutants by assessing cellular sensitivity to genotoxic stress by spot assay (**Figure 11**). We decided to compare the effect on viability of two drugs to which *vid22*Δ cells are sensitive: hydroxyurea (HU), which, by causing replicative stress, also indirectly promotes the folding of G-quadruplexes in the single-strand DNA stretches generated by the deceleration in the DNA synthesis process[31,420], and bleomycin (BM), which causes DNA breaks in addition to damaging other cellular components[421]. Therefore, serial dilution of cell cultures producing Myc-tagged vid22 K873A K875A (vid22 KK), vid22 N878A (vid22 N), vid22 Q880A (vid22 Q), vid22 D887A D889A D891A (vid22 DDD), or vid22 L893P (vid22 L/P) were plated on solid complete media in absence and in presence of increasing concentrations of hydroxyurea (50mM, 100mM, and 200mM HU) or bleomycin (125ng/µL, 250ng/µL, and 500ng/µL BM), along with an untagged control (no tag), cells expressing Vid22-Myc (Vid22) and a strain in which *VID22* is deleted (*vid22*Δ). Whereas all strains grow comparably in untreated conditions, the growth of vid22 KK, vid22 DDD, and vid22 L/P is impaired by both drugs in a dosage-dependent manner, indicative of heightened cell sensitivity to genotoxic stress, similarly to what is observed for *vid22*Δ. Instead, the sensitivity to hydroxyurea and bleomycin of vid22 N and vid22 Q mutants is unaffected, as they grow like Vid22 or the no tag strains (**Figure 11**). Therefore, sensitivity to genotoxic stress also correlates with overexpression toxicity and interaction with Tbf1: the same mutations that completely abolish complex formation—such as K873A K875A, D887A D889A D891A, and L893P—render cells more susceptible to genotoxic stress, and overexpression of their alleles is not lethal. Instead, total or partial retention of Tbf1 interaction in vid22 N878A and vid22 Q880A appears to be sufficient to ensure cellular tolerance to genotoxic stress and correlates with cytotoxicity upon allele overexpression. Collectively, these data support the relevance of Vid22-Tbf1 complex formation in genome stability maintenance.

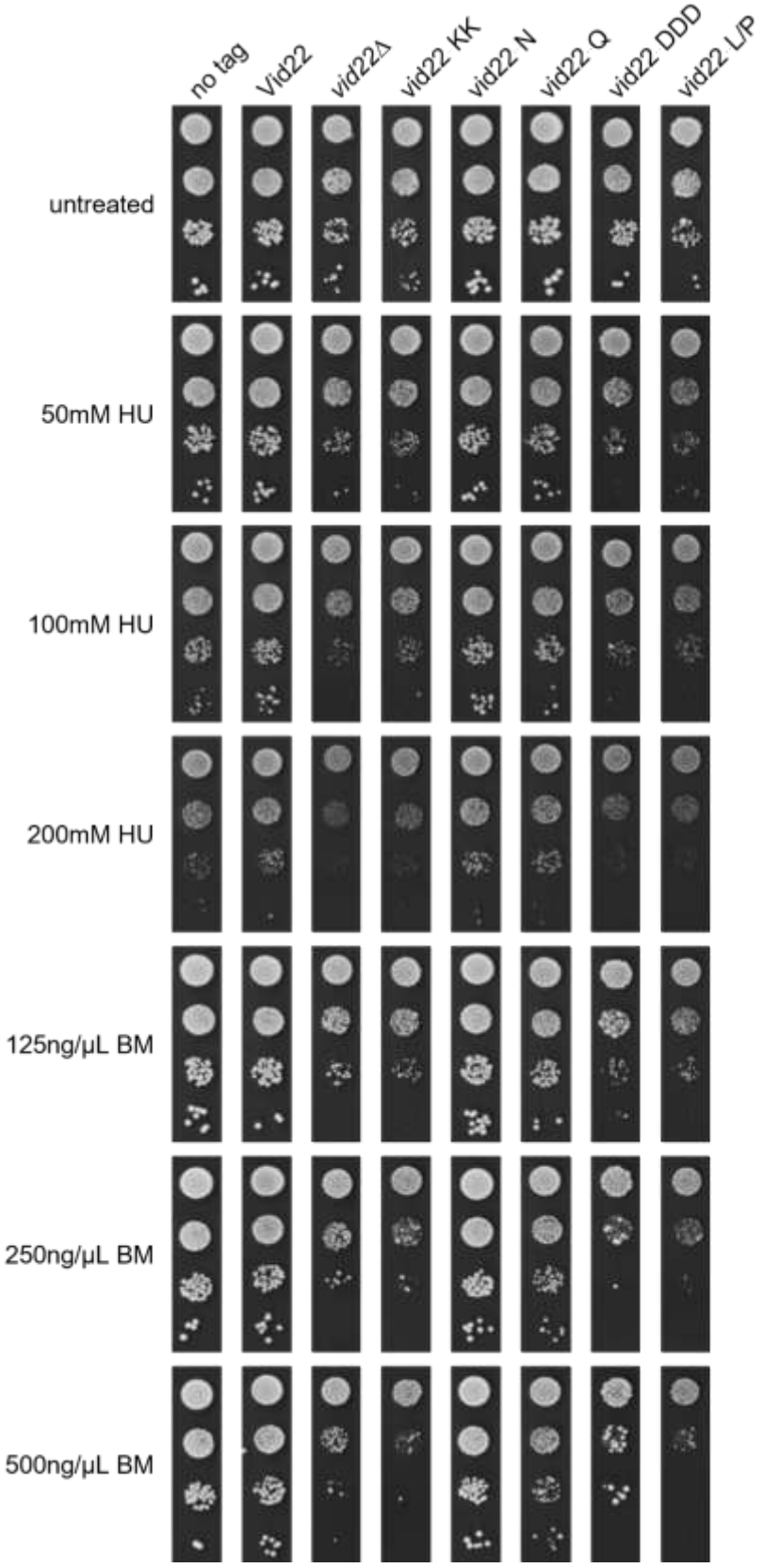

no tag
Vid22
vid22Δ
vid22 KK
vid22 N
vid22 Q
vid22 DDD
vid22 L/P
untreated
50mM HU
100mM HU
200mM HU
125ng/µL BM
250ng/µL BM
500ng/µL BM

Figure 11 – Vid22 C-terminal mutations that compromise interaction with Tbf1 also heighten cell sensitivity to genotoxic stress

Sensitivity to hydroxyurea (HU) and bleomycin (BM) of wild-type Vid22 or mutated in conserved C-terminal residues, compared to *vid22*Δ, evaluated by spot assay. Log-phase wild-type cells (no tag, strain BY4741), Vid22-Myc (Vid22, strain YFL3870), *vid22*Δ (*vid22*Δ, strain YNOV124), vid22 K873A K875A-Myc (vid22 KK, strain YFL3895), vid22 N878A-Myc (vid22 N, strain YFL3897), vid22 Q880A-Myc (vid22 Q, strain YFL3871), vid22 D887A D889A D891A-Myc (vid22 DDD, strain YFL3900), and vid22 L893P-Myc (vid22 L/P, strain YFL3876) were diluted to an optical density at 600 nm (OD_{600}) of 0,06 in H_2O and spotted, along with three 10-fold serial dilutions, on solid complete media devoid of drugs (untreated) or supplemented with 50mM HU, 100mM HU, 200mM HU, 125ng/µL BM, 250ng/µL BM, or 500ng/µL BM. The plates were photographed after 3 days of incubation at 28°C. Representative outcome of N=3 independent experiments.

I finally analysed the effect of K873A K875A and L893P mutations, chosen as representatives among those associated with the most severe phenotypes, in both Vid22 and Tbf1 binding at the G-quadruplex forming locus in the promoter of *SKN7* by ChIP-qPCR (**Figure 12 A-B**). To evaluate the enrichment of both proteins in the same context, extracts from cells producing Myc-tagged Vid22, or its C-terminal mutant variants, in combination with Flag-tagged Tbf1 were separately incubated with Myc- or Flag-specific antibodies to perform immunoprecipitation, along with extracts from the untagged negative control. qPCR analysis of the co-precipitated chromatin fractions revealed that vid22 KK and vid22 L/P are fundamentally unable to bind at the G4 p*SKN7* locus, displaying an enrichment comparable to the negative control (**Figure 12 A**), and Tbf1 enrichment is reduced of ~90% in the same backgrounds with respect to the Vid22-Myc Tbf1-Flag strain (**Figure 12 B**). From this analysis, it can be inferred that Vid22 is necessary to stabilize or enhance Tbf1 binding to DNA, consistent with ChIP-qPCR data at the p*SKN7* mutated loci (**Figure 4 C-D**). These results, combined with Tbf1 interaction data and the outcome of cell sensitivity assay, can also be interpreted in support of the model in which the Vid22-Tbf1 complex has a relevant function in genomic protection, because of the correlation between mutation of Vid22 C-terminus, loss of Tbf1 interaction,

increased sensitivity to genotoxic stress, and impairment of the recruitment of both proteins to a G-quadruplex forming locus.

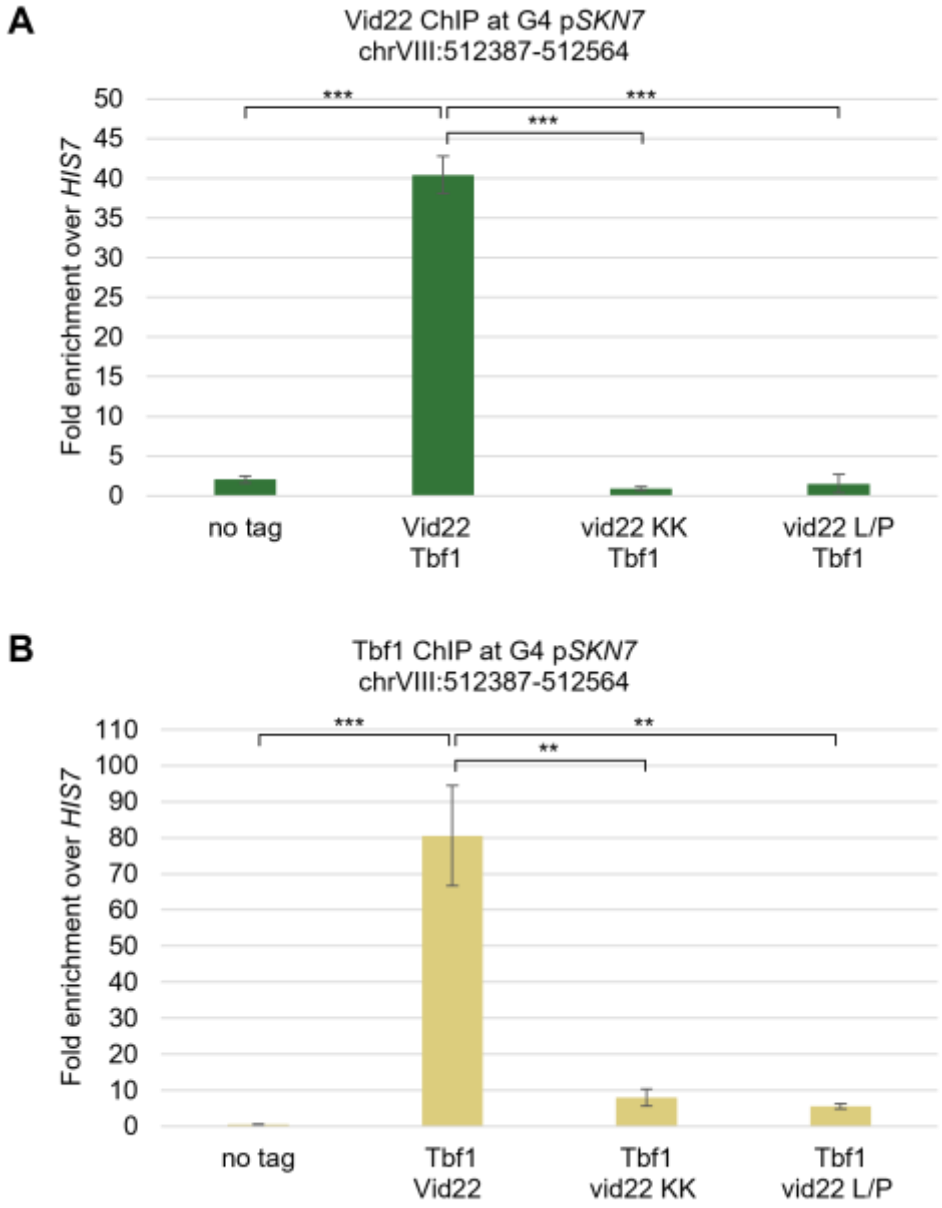


Figure 12 – Vid22 and Tbf1 recruitment at the G4 locus in SKN7 promoter is highly impaired by mutation of Vid22 conserved C-terminal residues

(A) ChIP-qPCR of wild-type Vid22 or mutated in conserved C-terminal residues at the G4-forming[132,263] locus in p*SKN7* (G4 p*SKN7*). ChIP was performed in wild-type cells (no tag, strain BY4741), Vid22-Myc Tbf1-Flag (Vid22 Tbf1, strain YFL3993), vid22 K873A K875A-Myc Tbf1-Flag (vid22 KK Tbf1, strain YFL3995), and vid22 L893P-Myc Tbf1-Flag (vid22 L/P

Tbf1, strain YFL4001) with 9E10 Myc antibodies. Fold enrichment was calculated relative to *HIS7* internal standard. Data are represented as mean ± standard error of N=3 independent experiments and the p-value of pairwise t-tests is indicated (p>0,05 n.s.; p<0,05 *; p<0,01 **; p<0,001 ***). **(B)** ChIP-qPCR of Tbf1 in a wild-type Vid22 or mutated in conserved C-terminal residues background at the G4-forming[132,263] locus in p*SKN7* (G4 p*SKN7*). ChIP was performed in wild-type cells (no tag, strain BY4741), Vid22-Myc Tbf1-Flag (Tbf1 Vid22, strain YFL3993), vid22 K873A K875A-Myc Tbf1-Flag (Tbf1 vid22 KK, strain YFL3995), and vid22 L893P-Myc Tbf1-Flag (Tbf1 vid22 L/P, strain YFL4001) with M2 Flag antibodies. Fold enrichment was calculated relative to *HIS7* internal standard. Data are represented as mean ± standard error of N=3 independent experiments and the p-value of pairwise t-tests is indicated (p>0,05 n.s.; p<0,05 *; p<0,01 **; p<0,001 ***).

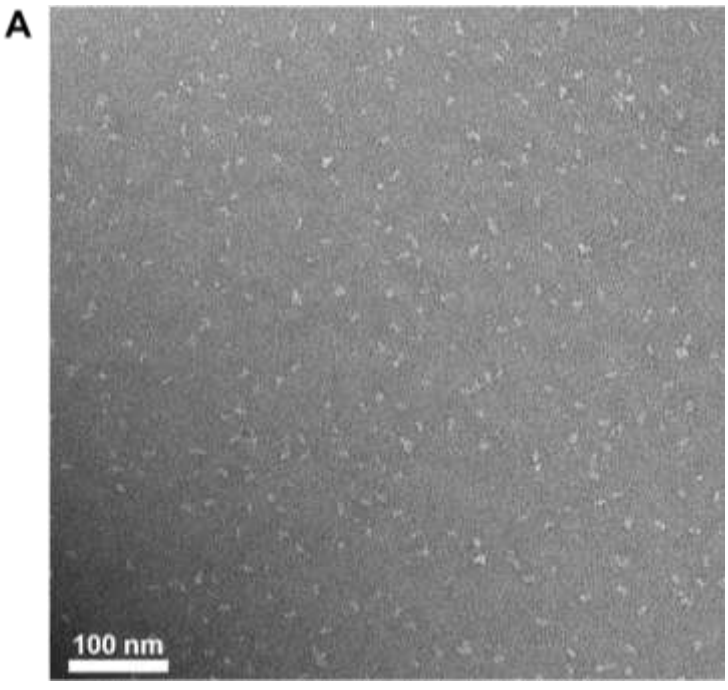
Preliminary investigation of the effect of the Q880A and D887A D889A D891A mutations on Vid22 and Tbf1 recruitment at G4 p*SKN7* by ChIP-qPCR indicated that whereas vid22 DDD is fundamentally unable to interact with DNA and correlates with a similar decrease in Tbf1 enrichment as vid22 KK and vid22 L/P, vid22 Q displays an intermediate phenotype with a decreased enrichment of ∼70% compared to that of Vid22, mirrored by a parallel decrease of Tbf1 enrichment of ∼75% (data not shown, single replicate). Overall, K873A K875A, D887A D889A D891A, and L893P mutations essentially abolish Vid22 interaction with Tbf1 and its DNA binding at the p*SKN7* locus, which also correlates with impaired Tbf1-DNA recruitment and heightened cell sensitivity to hydroxyurea and bleomycin. The Q880A mutation, instead, impairs without abolishing protein-protein and protein-DNA interactions, which is probably sufficient to retain cell tolerance to genotoxic stress.

The analysis of Vid22 mutated in the C-terminus or in the BED domain does not resolve whether Vid22 recruitment to DNA depends on Tbf1, because, in fact, the phenotypes of loss of interaction with DNA and Tbf1 are correlated (**Figures 1 A-E, 6, 10 F, 12 A-B**). Tbf1 appears to be able to bind the G4-forming locus in p*SKN7* independently of Vid22, but, in the absence of the latter, its observed enrichment is severely decreased. The most convincing explanation is that the Vid22-Tbf1 complex is recruited at the locus and performs a different function compared to free Tbf1, likely in genome protection, as abolition of Vid22-Tbf1 interaction correlates

with heightened cell sensitivity to replicative stress, G4 stabilisation and DNA breakage. However, even a partial interaction of Vid22-Tbf1 with DNA is functional for maintaining cellular tolerance to genotoxic stress.

5.5 Vid22 is structurally homologous to the Hermes transposase

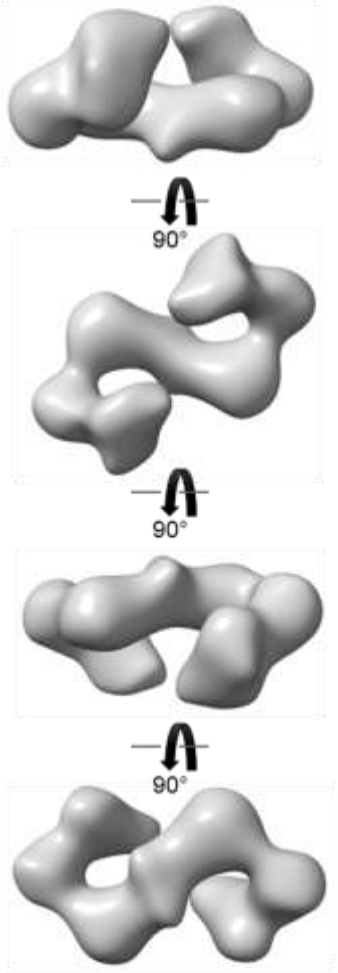
For the characterisation of Vid22 biochemical and structural features via *in vitro* experiments, *VID22* was His-tagged, heterologously expressed in a bacterial system, purified by affinity chromatography[263] and subjected to a subsequent purification step via size exclusion chromatography with multi-angle static light scattering to precisely estimate the molecular weight of the eluted proteins and detect the potential formation of aggregates, in collaboration with professor S. Ricagno. We noted that the main peak in the elution profile recoded at 280 nm was at 248 kDa (data not shown), compatible with Vid22 forming homodimers, as the monomer has a molecular weight of ~102,5 kDa, and there was no evidence of high molecular weight aggregates. Indeed, Vid22 is computationally predicted to be related to proteins belonging to the RNase H-like superfamily[340], which are known to self-interact to perform their functions[360,361], especially those of the hAT family[349,350,359,367,368,373,393]. Consistently, preliminary 16Å low-resolution structural reconstructions from negative staining electron microscopy analysis of the purified protein, performed in collaboration with A. Chaves Sanjuan, revealed that Vid22 forms homogeneous particles with a volume sufficient to accommodate two monomers (**Figure 13 A-D**).



B


10 nm

C


90°
90°
90°

D

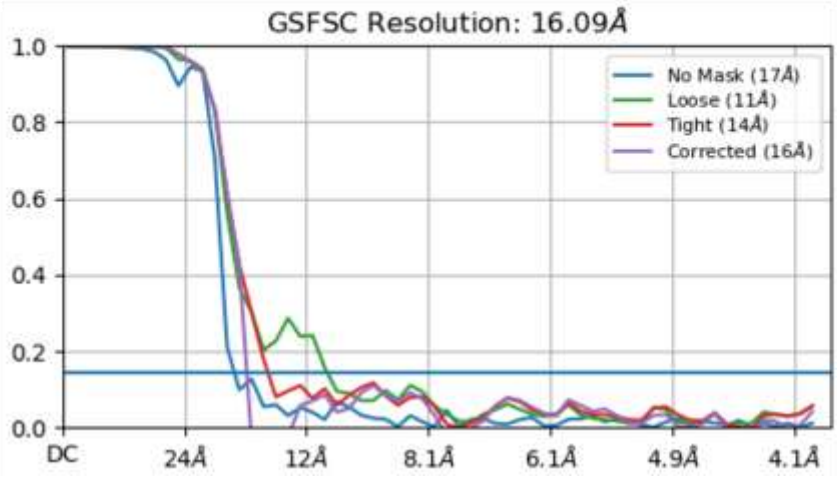

GSFSC Resolution: 16.09Å
1.0
0.8
0.6
0.4
0.2
0.0
DC
24Å
12Å
8.1Å
6.1Å
4.9Å
4.1Å
No Mask (17Å)
Loose (11Å)
Tight (14Å)
Corrected (16Å)

Figure 13 – Purified Vid22 forms homogeneous particles compatible with dimer formation

(A) Negative staining electron microscopy image of purified Vid22. Observations were carried out on a Talos L120C microscope (Thermo Scientific) with an accelerating voltage of 120 kV, and the image acquired with a Ceta 4k×4k camera at a nominal magnification of 73.000 corresponding to a pixel size of 1.9Å/pix. **(B)** Representative particles out of 13.128 particles selected for 2D classification in CRYOSPARC 4.2.1[422] to obtain the final 3D reconstruction model of Vid22 structure. **(C)** Final 3D reconstruction model of Vid22 structure with a 16Å resolution. **(D)** Resolution of the final 3D reconstruction model of Vid22 structure according to the gold standard FSC of 0.143.

To investigate the possibility of Vid22 dimerisation *in vivo*, I initially conducted yeast two-hybrid assays (**Figure 14 A**). *VID22* was cloned in both the bait and prey vectors, fused to the LexA DNA-binding domain (B-Vid22) or to the B42 transcriptional activation domain (P-Vid22), respectively, and transformed simultaneously or in combination with the empty vectors in the yeast strain carrying the *lacZ* reporter system[423]. Tbf1 bait (B-Tbf1) and prey (P-Tbf1) constructs were also generated and combined with Vid22 prey and bait, respectively, to be used as positive controls, as well as with the empty vectors. Proper cloning and accumulation of the constitutively expressed bait fusions and of the galactose-inducible prey chimaeras was confirmed by sequencing and Western blot, respectively (data not shown). Interaction between the fusion proteins was detected by *lacZ* expression in X-gal plate assays (**Figure 14 A**). Vid22 bait (B-Vid22) successfully interacts with Vid22 prey (P-Vid22), as reported by the blue pigmentation of the cell patch, which is indicative of β-galactosidase catalysed hydrolysis of the lactose analogue X-gal. The same coloration is displayed by B-Vid22 P-Tbf1 and B-Tbf1 P-Vid22 patches, as expected. The reporter gene is not expressed when Vid22 or Tbf1 bait constructs are co-transformed with the empty prey vector (P-empty), nor when the prey protein fusions are combined with the bait empty vector (B-empty; **Figure 14 A**).

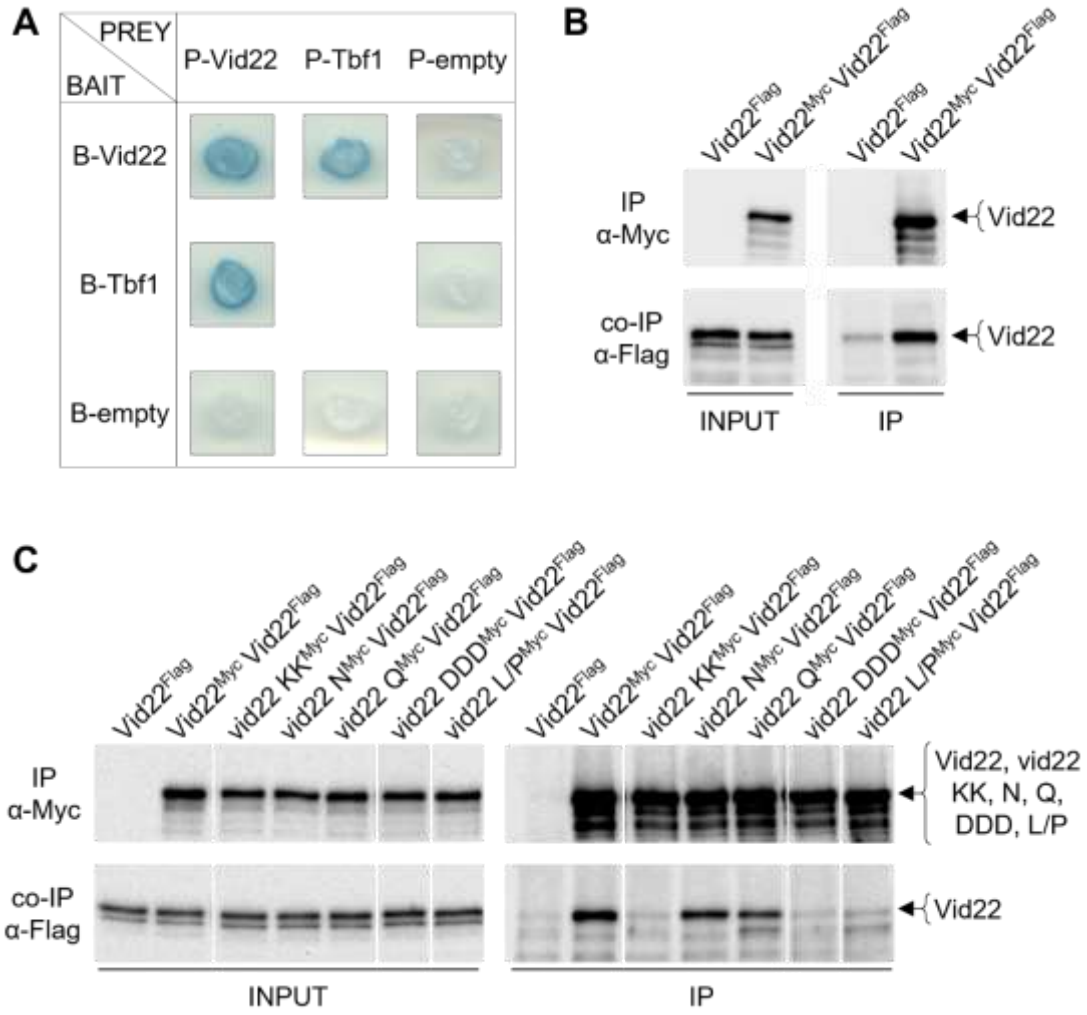


Figure 14 – In vivo Vid22 dimerisation is impaired by mutation of conserved C-terminal residues

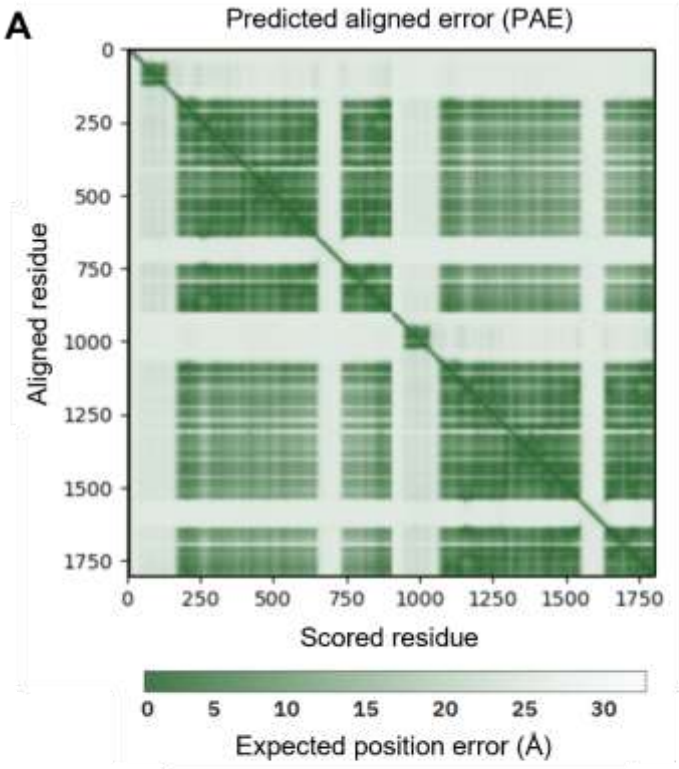
(A) Yeast two hybrid X-gal plate assay probing Vid22 self-interaction. Bait constructs were constitutively expressed as fusions to LexA DNA-binding domain. Gal-inducible prey constructs were expressed as fusions to B42 transcriptional activation domain-HA epitope tag. Cells expressing Vid22 bait and prey constructs (B-Vid22 P-Vid22, strain YFL3271), Vid22 bait and Tbf1 prey (B-Vid22 P-Tbf1, strain YFL3280), Tbf1 bait and Vid22 prey (B-Tbf1 P-Vid22, strain YFL3283), their combinations with the empty vectors (B-Vid22 P-empty, strain YFL3255; B-empty P-Vid22, strain YFL3259; B-Tbf1 P-empty, strain YFL3263; B-empty P-Tbf1, strain YFL3267), and the negative control (B-empty P-empty, strain YGE253) were plated on solid synthetic medium supplemented with X-gal to evidence lacZ reporter gene expression by formation of blue pigmented colonies, and 2% raffinose, 2% galactose to induce prey overexpression. The plate was photographed after 1 days of incubation at 28°C. Representative outcome of N=3 independent experiments. **(B)** Western

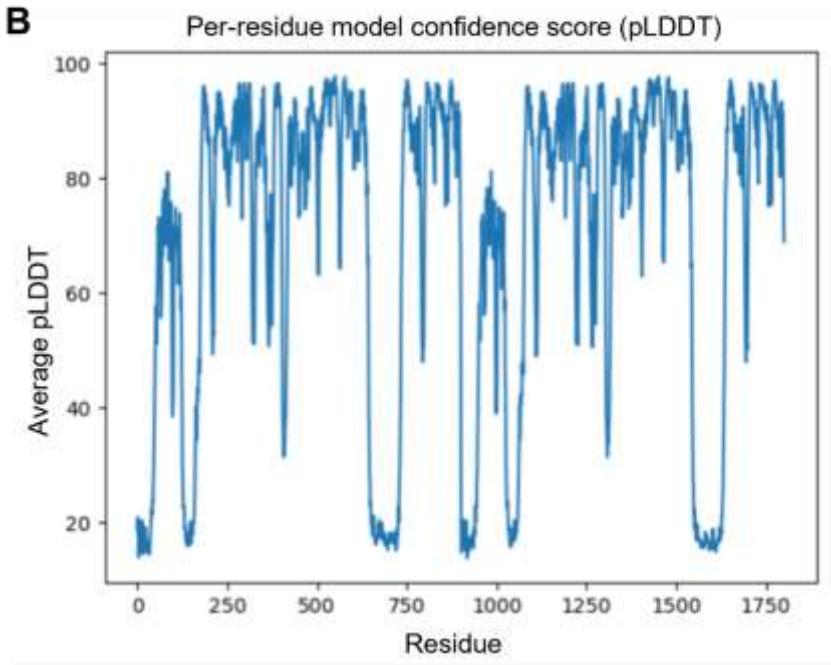
blot analysis of co-IP experiments of wild-type Vid22 with itself. Normalised protein extracts of Vid22-Flag cells transformed with the empty vector (Vid22^{Flag}, strain YFL3989) or transformed with Vid22-Myc (Vid22^{Myc} Vid22^{Flag}, strain YFL3987) were incubated with 9E10 Myc antibodies. Blots were incubated individually with 9E10 Myc (top panels) or M2 Flag (bottom panels) antibodies. The arrows point to the expected bands. Representative immunoblots of N=3 independent experiments. **(C)** Western blot analysis of co-IP experiments of wild-type Vid22 with wild-type Vid22 or mutated in conserved C-terminal residues. Normalised protein extracts of Vid22-Flag cells transformed with the empty vector (Vid22^{Flag}, strain YFL3989) or transformed with Vid22-Myc (Vid22^{Myc} Vid22^{Flag}, strain YFL3987), vid22 K873A K875A (vid22 KK^{Myc} Vid22^{Flag}, strain YFL4014), vid22 N878A (vid22 N^{Myc} Vid22^{Flag}, strain YFL4015), vid22 Q880A (vid22 Q^{Myc} Vid22^{Flag}, strain YFL4016), vid22 D887A D889A D891A (vid22 DDD^{Myc} Vid22^{Flag}, strain YFL4018), or vid22 L893P (vid22 L/P^{Myc} Vid22^{Flag}, strain YFL4020) were incubated with 9E10 Myc antibodies. Blots were incubated individually with 9E10 Myc (top panels) or M2 Flag (bottom panels) antibodies. The arrows point to the expected bands. Representative immunoblots of N=3 independent experiments.

These data indicate that the assay allows to specifically detect the interaction between the proteins of interest and, thus, that Vid22 can homodimerise. To validate the results of the yeast two-hybrid assay, I tested Vid22 self-interaction by co-IP (**Figure 14 B**). To this end, a centromeric plasmid carrying the p*VID22*:*VID22-13MYC* insert was provided to haploid yeasts producing Flag-tagged Vid22 encoded at its endogenous locus. Vid22-Flag transformed with the empty vector serves as the negative control. Whole protein extracts from these strains were incubated with Myc-specific antibodies and the co-immunoprecipitated fraction probed with Flag-specific antibodies. Vid22^{Flag} is co-precipitated with Vid22^{Myc} confirming self-interaction (**Figure 14 B**). We next chose to test if Vid22 C-terminal mutations characterised so far would affect dimerisation because the C-terminal region was reported to be directly involved in self-association of the hAT derived human ZBED1 protein[393]. Moreover, despite Hermes transposase C-terminus not being directly engaged in multimerisation, as revealed by its crystal structures[350,367,368,424], mutation of conserved C-terminal residues prevents its self-interaction[397]. Therefore, Flag-tagged Vid22 cells were transformed with centromeric vectors in

which the p*VID22*:*VID22-13MYC* insert was mutated *in vitro* to introduce the desired nucleotide substitutions. Subsequently, Myc-tagged vid22 KK, vid22 N, vid22 Q, vid22 DDD, and vid22 L/P were immunoprecipitated and the co-precipitated fraction probed for Vid22-Flag (**Figure 14 C**). Despite no observable differences in the immunoprecipitation efficiencies, Vid22-Flag is only detectable in the co-precipitated fraction of vid22 N878A and vid22 Q880A, albeit at a marginally reduced level in the latter (**Figure 14 C**). This outcome indicates that the K873A K875A, D887A D889A D891A and L893P substitutions abrogate not only Vid22-Tbf1 interaction, but also Vid22 self-association. Taken these and the previous results together, the C-terminus might not be directly responsible for Tbf1 interaction. The loss of Vid22-Tbf1 complex formation caused by mutation of Vid22 C-terminal residues (**Figure 10 F**) might result from the impairment in Vid22 homodimerisation (**Figure 14 C**). If Tbf1 can only interact with Vid22 dimers, any condition abolishing Vid22 self-interaction would prevent their physical association.

Thanks to recent advancements in computational methods for predicting the structure of proteins based on their amino acid sequence and, particularly, to the development of AlphaFold artificial intelligence[425,426] and its implementations[427], we have now access to a relatively accurate model of Vid22 three-dimensional fold (C7GRW6) and extension programs like AlphaFold2 Multimer specifically designed to predict protein-protein complexes. Therefore, from Vid22 full-length sequence, we built a model of Vid22 homodimer using ColabFold[427] implementation of AlphaFold2-multimer, in collaboration with professor C. Camilloni (**Figure 15 A-D**). The analysis of metrics indicating the reliability of Vid22 monomer (which is superimposable to the prediction deposited in the AlphaFold Protein Structure Database C7GRW6) and dimer structure predictions, like the predicted aligned error (PAE) and the per-residue model confidence score (pLDDT), reveal that, except for the N-terminal region (residues 1-161) and the disordered loop (residues 648-723) in each monomer, the relative orientation of the protein secondary structure elements and domains is predicted with high confidence (**Figure 15 A-B**).


A
Predicted aligned error (PAE)
Aligned residue
Scored residue
0 250 500 750 1000 1250 1500 1750
0 5 10 15 20 25 30
Expected position error (Å)


B
Per-residue model confidence score (pLDDT)
Average pLDDT
Residue
100 80 60 40 20
0 250 500 750 1000 1250 1500 1750

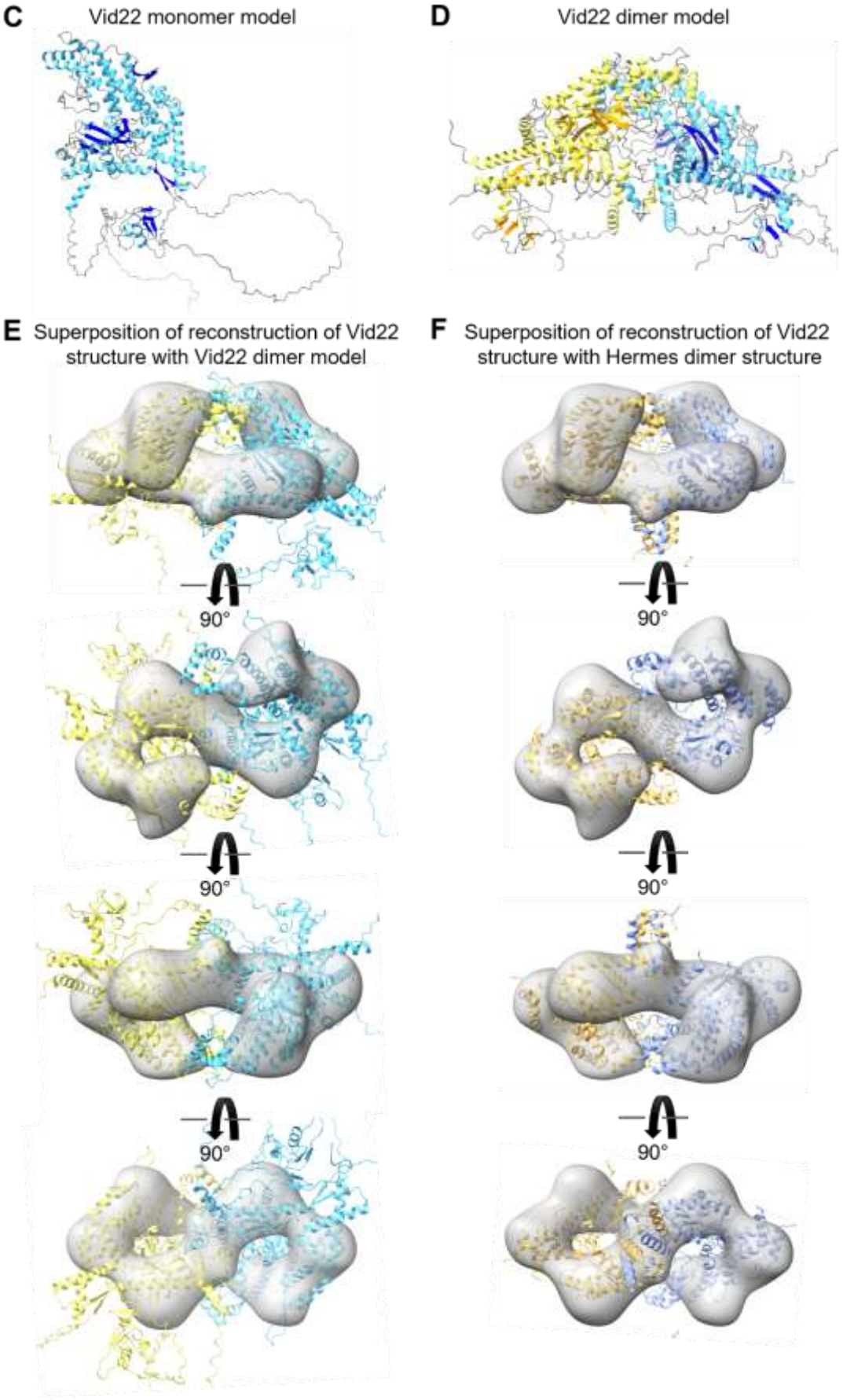
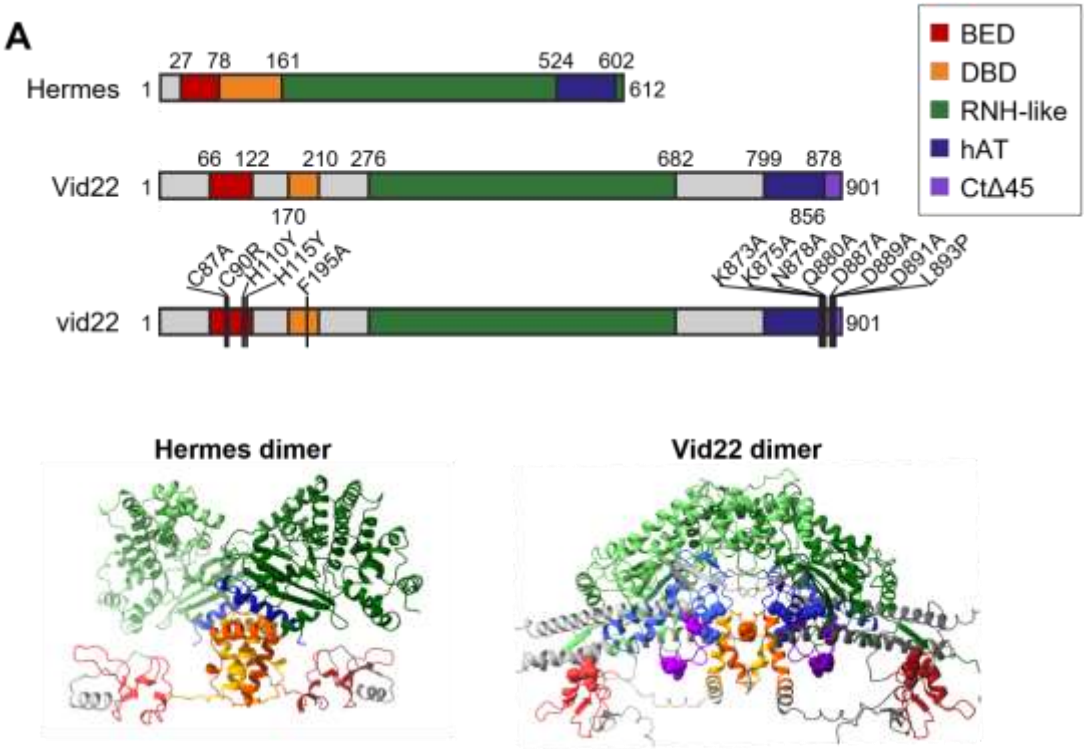

C
Vid22 monomer model
D
Vid22 dimer model
E
Superposition of reconstruction of Vid22 structure with Vid22 dimer model
F
Superposition of reconstruction of Vid22 structure with Hermes dimer structure
90°
90°
90°
90°
90°
90°

Figure 15 – Vid22 dimer model and Hermes dimer structure overlap to the 3D reconstruction model of Vid22 structure

(A) Predicted aligned error (PAE) plot of Vid22 dimer model. Residues 1-901 correspond to one Vid22 monomer, residues 902-1802 correspond to the second monomer. PAE estimates the expected positional error for each residue in the predicted protein dimer structure if it were aligned to a corresponding residue in the true protein dimer structure and provides an indication of the level of confidence in the relative positions and orientations of different parts of the predicted model. Lower PAE values for residue pairs from different domains suggest well-defined relative positions and orientations in the prediction model, while higher PAE values indicate that the relative positions and/or orientations of these domains in the 3D structure are uncertain and should not be interpreted. **(B)** Per-residue model confidence score (pLDDT) plot of Vid22 dimer model. Residues 1-901 correspond to one Vid22 monomer, residues 902-1802 correspond to the second monomer. pLDDT corresponds to the model's predicted score on the LDDT-Cα metric[425,428] and estimates per-residue confidence on 0-100 scale. Regions with pLDDT>90 are expected to be modelled to high accuracy, regions with 70<pLDDT<90 are expected to be modelled with a generally good backbone prediction, regions with 50<pLDDT<70 are modelled with low confidence and should be treated with caution, regions with pLDDT<50 are usually disordered (either unstructured in physiological conditions or only structured as part of a complex), often have a ribbon-like appearance and should not be interpreted. **(C)** Model of Vid22 monomer. Alpha helices are coloured in light blue (helix), beta strands are coloured in dark blue (strand), coils are grey coloured (coil). **(D)** Model of Vid22 dimer. Vid22 monomers are differently coloured in yellow and blue respectively. Alpha helices are coloured in yellow in one monomer (helix) and light blue in the other (helix), beta strands are coloured in orange (strand) and dark blue (strand), coils are grey coloured in both monomers (coil). **(E)** Superposition of Vid22 dimer model to the 3D reconstruction model of Vid22 structure obtained by negative staining electron microscopy analysis with a correlation coefficient of 0,65. In Vid22 dimer model, one monomer is coloured in yellow, the other in light blue. **(F)** Superposition of Hermes dimer structure[350], trimmed of the BED domains, to the 3D reconstruction model of Vid22 structure obtained by negative staining electron microscopy analysis with a correlation coefficient of 0,94. In Hermes dimer structure, one monomer (residues 81-609) is coloured in light orange, the other in light purple.

Two domains are clearly recognisable in Vid22 monomer: the smaller BED-type zinc finger and a transposase-derived domain formed by the bulk of the protein (**Figure 15 C**). The top three predictions are consistent in the relative positioning of the monomers and the best model is reported (**Figure 15 D**). Superposition of the best model of Vid22 homodimer generated *in silico* to the low-resolution three-dimensional reconstruction obtained by negative staining electron microscopy shows a correlation coefficient of 0,65 (**Figure 15 E**). The overlap could be improved by removing the disordered regions, as the PAE for these two regions is poor relative to the rest of the protein, so they may be misplaced and hinder complex prediction. Strikingly, the correlation coefficient reaches 0,94 when Vid22 volume reconstruction is superposed to the experimentally determined structure of the Hermes dimer[350] trimmed of the BED domains (**Figure 15 F**), despite its inability to fill the entire reconstruction because of a smaller size. This indicates that the model of Vid22 homodimer can be improved and probably its actual structure is more similar to the Hermes transposase than expected from sequence comparison.

Given the structural homology between Vid22 and the extensively characterized Hermes[350,367,368], their comparison may allow to interpret and predict Vid22 structural-functional properties (**Figure 16 A-B**).



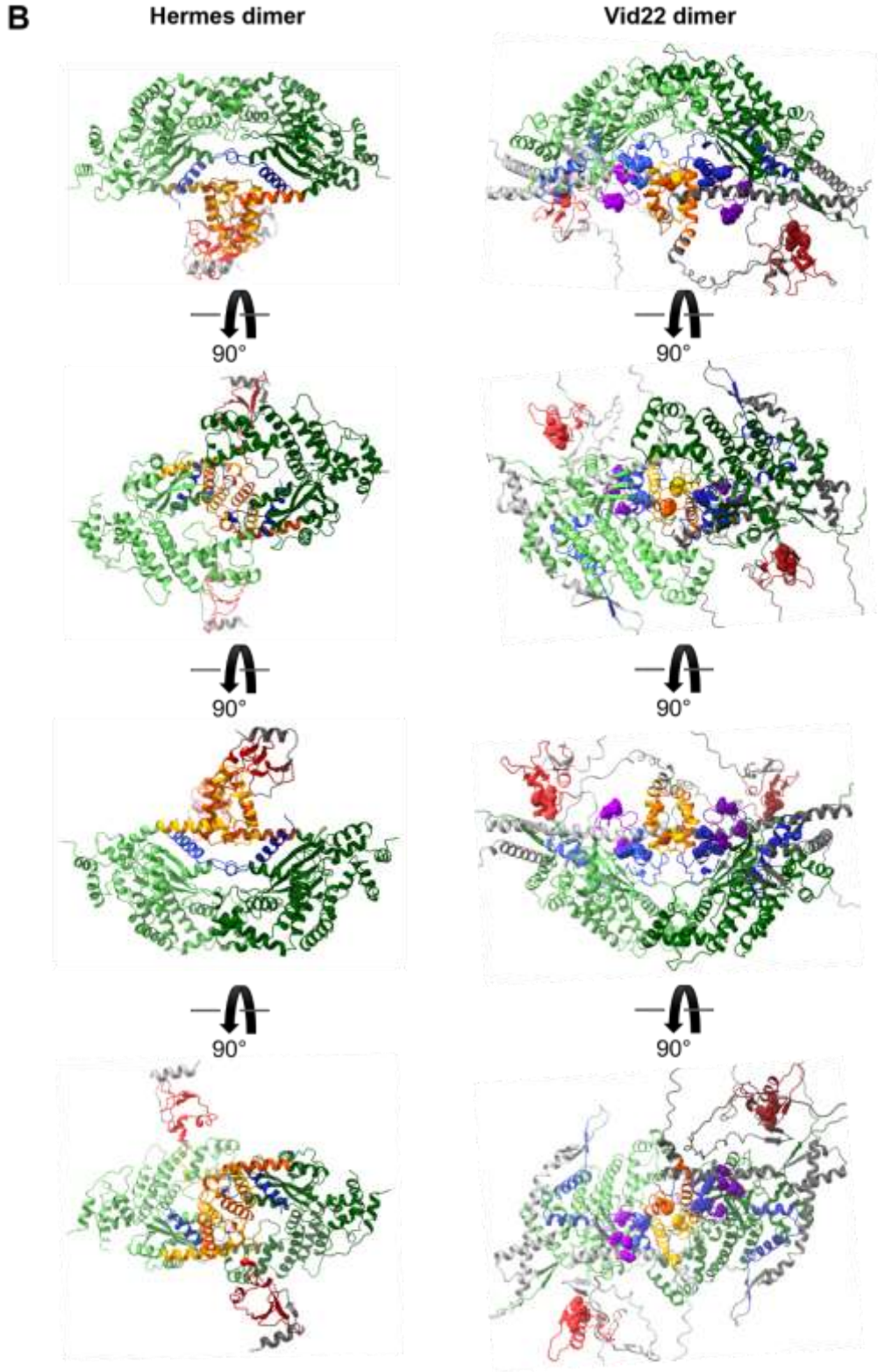


Figure 16 – Comparison between the domain annotation in Vid22 dimer model and Hermes dimer structure

(A) Schematic representation of wild-type Vid22 protein sequence or mutated in conserved

residues and the Hermes transposase sequence from *M. domestica*, annotated of the BED-type Zn finger (BED, IPR003656), Hermes transposase DNA-binding domain (DBD, IPR018473) and the putative corresponding region in Vid22 based on sequence conservation analysis (Figure 9) and Vid22 dimer model, RNase H-like superfamily (RNH-like, IPR012337) and C-terminal hAT dimerisation domain (hAT, PF05699). The C-terminal portion truncated in vid22 856-901Δ mutant (CtΔ45) is also annotated. Substitutions targeting conserved residues in vid22 mutants are reported. Hermes dimer structure[350] and Vid22 dimer model are shown below maintaining the annotations colours. The side chains of the residues mutated in Vid22 are displayed in sphere style. **(B)** Comparison of Hermes dimer structure[350] and Vid22 dimer model, maintaining the annotations colours shown in panel A and displayed in the same orientations of Figure 15 E-F.

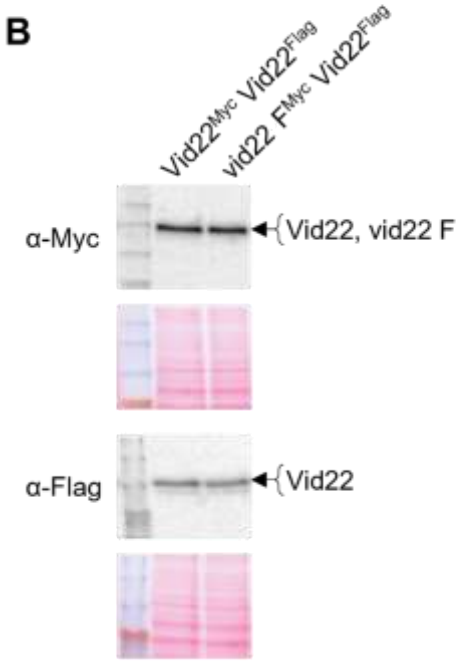
Vid22 BED domains extend towards the edges of the quaternary structure and, similarly to Hermes[350,367,368], they contribute to DNA interaction, although they are not essential, as mutation of the zinc coordinating residues decreases without abolishing DNA binding *in vivo* (**Figure 1 A-E**). We next examined the location and potential interactions of the conserved C-terminal residues that were mutated in Vid22 (**Figure 10 B**) with respect to the global quaternary structure of Vid22 homodimer (**Figure 16 A-B**). Whereas none of the mutated amino acids is directly involved in Vid22 self-association in the obtained Vid22 dimer model, they are close in space to residues 170-210 encompassing the putative dimerisation interface and to an additional potential DNA binding site inferred by homology to Hermes DBD[350,367,368] (**Figure 16 A-B**). Therefore, Vid22 C-terminus is key to maintain the overall tertiary and, thus, quaternary structures, providing another piece of evidence in support of the resemblance between these two proteins despite their evolutionary distance and sequence divergence. The structural relevance of Vid22 C-terminus is evidenced by the high conservation among homologous sequences (**Figure 9** block 18 of 18, residues 850-901). Indeed, this could explain the phenotypes of loss-of-function Vid22 C-terminal mutants, stressing the importance of the functionality of this factor in maintenance of genome stability. Based on the position of the C-terminus in Vid22 structural model, it is also unlikely that this region constitutes the interface for Tbf1 interaction. The abolition of Vid22-Tbf1 physical association by mutation of Vid22 C-terminal residues (**Figure 10 F**) might

either be explained by the possibility that Tbf1 only interacts with Vid22 dimers, since these mutations also prevent Vid22 self-interaction (**Figure 14 C**), or by the spatial proximity of the C-terminus to the actual site of Vid22-Tbf1 contact, which might have been deformed by the C-terminal mutations.

To explore the possibility of the existence of a second, not yet annotated, DNA binding region encompassing residues 170-210, which may be directly involved both in Vid22-DNA/G-quadruplex interaction and in its self-association, we selected the candidate residue F195 for targeted mutagenesis from the structural model of Vid22 dimer (**Figure 16 A-B**) and the evolutionary conservation data of Vid22 sequence (**Figure 9** block 4 of 18, residues 150-199). In addition, this residue was mutated in vid22 #29 (**Figure 7 A**). I thus substituted Vid22 F195, with 80% identity in the MSA, to A (**Figure 17 A**) by in vitro mutagenesis of p*VID22*:*VID22*-Myc cloned in a centromeric vector, which was subsequently transformed into Vid22-Flag cells.

A

Mutant	*Residue*	*Identity*	*Similarity*	*Substitution*
vid22 F	F195	80%	83% (Y)	A



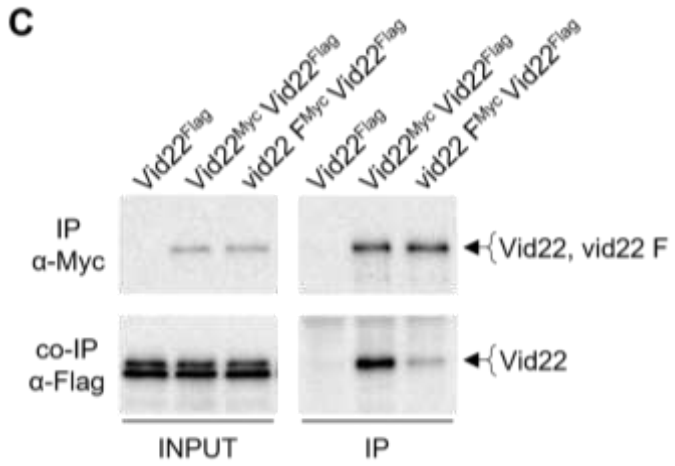


Figure 17 – Mutation of Vid22 F195, expected to be directly involved in self-interaction in Vid22 dimer model, impairs dimerisation in vivo

(A) Conservation of F195 found in Vid22 putative dimerisation interface among Vid22 homologues in Saccharomycotina species. **(B)** Western blot analysis of the levels of wild-type Vid22 or mutated in F195. Total proteins were extracted from normalised cultures of Vid22-Flag cells transformed with Vid22-Myc ($Vid22^{Myc}$ $Vid22^{Flag}$, strain YFL3987), or with vid22 F195A-Myc (vid22 F^{Myc} $Vid22^{Flag}$, strain YFL4158) grown in plasmid-selective media. The top blot was incubated with 9E10 Myc antibodies, the bottom blot with M2 Flag antibodies. Arrows point to the expected bands. Ponceau staining is shown as loading control. Representative blot of N=3 independent experiments. **(C)** Western blot analysis of co-IP experiments of wild-type Vid22 with wild-type Vid22 or mutated in F195. Normalised protein extracts of Vid22-Flag cells transformed with the empty vector ($Vid22^{Flag}$, strain YFL3989) or transformed with Vid22-Myc ($Vid22^{Myc}$ $Vid22^{Flag}$, strain YFL3987), or vid22 F195A-Myc (vid22 F^{Myc} $Vid22^{Flag}$, strain YFL4158) were incubated with 9E10 Myc antibodies. Blots were incubated individually with 9E10 Myc (top panels) or M2 Flag (bottom panels) antibodies. The arrows point to the expected bands. Representative immunoblots of N=3 independent experiments.

The stability of the mutant protein was verified by Western blot analysis (**Figure 17 B**). Therefore, I tested the ability of vid22 F195A-Myc (vid22 F) to interact with Vid22-Flag by co-IP experiments (**Figure 17 C**). Despite comparable immunoprecipitation levels of the Myc-tagged mutant and wild-type proteins, the signal of Vid22-Flag is greatly reduced in the co-immunoprecipitated fraction of vid22 F compared to that of Vid22 (**Figure 17 C**). This result evidences that the

F195A mutation severely affects the protein's physical interaction with its wild-type form, and supports the validity of Vid22 dimer model.

6 Conclusions and Future Perspectives

Vid22 has evolved from domestication of a DNA transposase of the hAT family, as was suggested by computational prediction of its domain composition and experimentally evidenced by its quaternary structure (**Figures 13 C, 14 A-B**) and structural homology with the Hermes homodimer (**Figures 15 E-F, 16 A-B**), to support genome stability and G4 regulation in budding yeast[263]. Similarly to what observed in this active transposase[350,367,368], Vid22 C-terminus seems to have a structural role in maintaining a functional protein fold, as evidenced by the high sequence conservation of this region (**Figure 9** block 18 of 18, residues 850-901) and by the severity of the phenotypes of the C-terminal mutants (**Figures 8 A-D, 10 F, 11, 12 A, 14 C**), despite it likely not being directly involved in homodimerisation (**Figure 16 A-B**) or interaction with Tbf1 and DNA. Compatibly with the obtained ChIP-qPCR data (**Figure 1 A-E**), Vid22 BED domain may not be the only region of the protein mediating DNA interaction. Hermes possesses a second DNA binding domain (residues 78-161) which is directly involved in dimerisation through intertwining α-helices[350,359,367,368,373], possibly corresponding to Vid22 residues 170-210. Indeed, this region has compatible secondary structure, localisation in Vid22 homodimer model (**Figure 16 A-B**), and sequence features, like a closely spaced Pro-Phe pair (P191 and F195 in Vid22) preceded by hydrophobic (*e.g.* I, L, A, V) and positively charged (*e.g.* K, R) amino acids which may participate in Vid22-DNA contacts (**Figure 9** blocks 4 and 5 of 18, residues 150-249). The observation that dimerisation is highly impaired in vid22 F195A (**Figure 17 C**) supports this possibility, warranting further characterisation of the phenotypes of this and related mutants to investigate the involvement of this protein region in B-DNA and G-quadruplex interactions. Aside from the generation and analysis of a model of Vid22 dimer in presence of a DNA/G4 sequence driven by homology to Hermes to aid in mutant design, we are also attempting to obtain a reliable model of the Vid22-Tbf1 complex, which would allow to obtain a preliminary indication of where their interaction interface could be located. Moreover, Vid22, which recruitment to certain DNA loci likely depends on G4 structures (**Figure 4 C**), significantly supports Tbf1-DNA interaction (**Figures 4 C-D, 5 A, 12 A-B**). The formation of the Vid22-Tbf1

complex, which is partially affected by mutation of Vid22 BED domain (**Figure 6**), is likely functional in genome stability maintenance, since analysis of Vid22 C-terminal loss-of-function mutants revealed that Vid22 dimerisation, interaction with Tbf1, their concerted recruitment to a G4-forming DNA locus, and cell tolerance to genotoxic stress are intercorrelated *in vivo* (**Figures 10 F, 11, 12 A-B, 14 C**). Further comparisons between Vid22 structural models and sequence conservation data will enable to rationalise the design of additional candidate hypomorphic mutations, for instance selectively abrogating interaction with Tbf1, potentially allowing to dissect the contribution of these factors in genome protection with unprecedented precision.

The biochemical characterisation of Vid22-G-quadruplex binding specificity and its effects on these secondary structures is ongoing at present. To refine the models generated *in silico* and clarify the stoichiometry of the Vid22-Tbf1-DNA/G4 complex, we are currently undertaking additional negative staining and cryo-electron microscopy analysis of purified Vid22 in presence of a G-quadruplex forming sequence and/or Tbf1. Structural data would also provide useful insights into Vid22-Tbf1 interaction mode and enable to verify whether Vid22 dimerisation is necessary for their physical association. Additionally, Tbf1 co-immunoprecipitation with vid22 F195A will be assayed. Moreover, *in vitro* circular dichroism experiments of G-quadruplex forming sequences combined with purified Vid22 and/or Tbf1 protein titration would allow to determine if Vid22 and its related protein complex exert a positive, negative or neutral effect on the G-tetrahelical assembly by detection of potential changes of the characteristic spectrum[429], as well as explore whether Vid22 RNase H-like domain conserves any catalytic activity, and determine the contribution of Tbf1 in Vid22 G-quadruplex-associated function. If Vid22 displays any catalytic activity, targeted mutagenesis of the residues forming the putative DDE/D triad, combined with an *in vitro* endonuclease assay, would allow to validate and characterize Vid22 mechanism of action. Electrophoretic mobility shift assays, or reflective phantom interface analysis, could also enable the identification of the amino acids specifically mediating Vid22 interactions with B-form DNA and/or G-quadruplex folds and investigate Vid22 affinity towards different tetrahelical topologies. A better understanding of the molecular features that

mediate G-tetraplex binding by native proteins may present the opportunity to design molecules able to selectively modulate these interactions as an alternative therapeutic strategy to overcome the specificity issue of current G-quadruplex ligands[114,120,430,431].

We are also attempting to clarify the pathway in which Vid22 functions to prevent G-quadruplex-associated genome instability. Since Vid22 preferentially bind at promoters site together with Tbf1[262,263,304,316], to quantify its impact on gene expression we performed RNA-Seq experiments in unperturbed wild-type and *vid22*Δ cells, which are presently being analysed. In addition, given Vid22 and Tbf1 implication in nucleosome positioning[309,310,312,432–437], and their link to histone de-acetylation[312,316,333], ATAC-Seq analysis[438] has also been planned. In addition to their involvement in transcription and nucleosome organization, Vid22 and Tbf1 have also been shown to impact DNA double-strand break repair, telomere and ribosomal DNA stability, and replication[262,263,439–442,268,301,303,304,309,312,337,409]. The function of all these processes, which can be potentially affected by G-quadruplex structures (*cfr.* Paragraph 4.2), has been linked to the three-dimensional organization of genomic DNA in the nuclear space in both budding yeast and mammalian cells[443–445]. Within the non-random chromosome territories, chromatin is organized in eu- and hetero-chromatic compartments involved in abundant intrachromosomal contacts, especially at the level of self-associating chromatin domains, as well as interchromosomal contacts[446]. Proper nuclear positioning and contact of certain DNA sequences (*e.g.* ribosomal DNA array, telomeres) is physiologically relevant, promoting their stability and epigenetic signature[447]. Transcription or replication factories, as well as DNA repair foci, formed by co-localisation and interaction of specific DNA elements and proteins in discrete nuclear sites, have been described[448,449]. Because both transcription factors[450–452] and G-quadruplexes[167–171] have the potential to bridge chromosome interactions and participate in the spatial organization of the genome, we are currently investigating the tempting possibility of Vid22 being involved in these aspects by *in silico* approaches and targeted 3C-qPCR experiments (*e.g.* exploring the contact between two regions of chromosome VIII detected in a previous work[444], one containing p*SKN7* and the other p*MDM31*, both bound by Vid22 and Tbf1) and HiC genome-wide analysis.

Vid22 functional characterisation exploiting mutants targeted in conserved residues, rationally designed from structural models, is a useful approach to clarify the details of its role in suppression of G-quadruplex induced genomic instability and aid in the identification of candidate functional human homologues. *In vivo* and *in vitro* structural-functional correlation will also provide new insights into G-quadruplex-protein recognition and into how these tetrahelical architectures influence biological processes.

7 Materials and Methods

7.1 Tables

7.1.1 Yeast strains

Name	*Background*	*Genotype*	*Plasmid*	*Source*
BY4741	S288C	Mat a *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*	-	Euro-scarf
EGY48	EGY48	Mat α *ura3Δ his3Δ trp1Δ* *leu2::6LexAop-LEU2*	-	R. Brent
YFL2922	YGE1014	*VID22-13Myc:HIS3MX6*	-	our lab
YFL2955	BY4741	*SGS1-13Myc:HIS3MX6*	-	our lab
YFL2987	BY4741	*sgs1Δ::HPHMX6* *VID22-13Myc:HIS3MX6*	-	our lab
YFL2994	BY4741	*vid22Δ::KANMX6* *SGS1-13Myc:HIS3MX6*	-	our lab
YFL3111	BY4741	*vid22(856-901Δ)-3HA:KANMX6* *TBF1-13Myc:HIS3MX6*	-	this work
YFL3128	BY4741	*TBF1-3Flag:KANMX6*	-	our lab
YFL3161	BY4741	*VID22-3HA:KANMX6*	-	our lab
YFL3164	BY4741	*VID22-3HA:KANMX6* *TBF1-13Myc:HIS3MX6*	-	our lab
YFL3170	BY4741	*vid22(856-901Δ)-3HA:KANMX6*	-	this work
YFL3255	YHN93	-	pGELE59 pJG4-5	this work
YFL3259	YHN93	-	pEG202 pGELE60	this work
YFL3263	YHN93	-	pGELE61 pJG4-5	this work
YFL3267	YHN93	-	pEG202 pGELE62	this work
YFL3271	YHN93	-	pGELE59 pGELE60	this work
YFL3280	YHN93	-	pGELE59 pGELE62	this work
YFL3283	YHN93	-	pGELE61 pGELE60	this work
YFL3376	BY4741	-	pGELE53	this work

Name	*Background*	*Genotype*	*Plasmid*	*Source*
YFL3377	BY4741	-	pGELE54	this work
YFL3378	BY4741	-	pGELE55	this work
YFL3379	BY4741	-	pGELE56	this work
YFL3380	BY4741	-	pGELE57	this work
YFL3381	BY4741	-	pGELE58	this work
YFL3683	BY4741	-	pFL230	this work
YFL3684	BY4741	-	pFL228	this work
YFL3685	BY4741	-	pFL231	this work
YFL3687	BY4741	-	pFL232	this work
YFL3689	BY4741	-	pFL234	this work
YFL3770	BY4741	*chrVIII 512268A:C 512304T:C 512305A:C*	-	this work
YFL3793	YFL3770	*VID22-13Myc:HIS3MX6*	-	this work
YFL3794	YFL3770	*TBF1-13Myc:HIS3MX6*	-	this work
YFL3797	BY4741	*chrVIII 512268A:C 512304T:C 512305A:C 512307G:C 512323G:C*	-	this work
YFL3800	YFL3797	*VID22-13Myc:HIS3MX6*	-	this work
YFL3801	YFL3797	*TBF1-13Myc:HIS3MX6*	-	this work
YFL3870	BY4741	*VID22-13Myc:LEU2MX6*	-	this work
YFL3871	BY4741	*vid22(Q880A)-13Myc:LEU2MX6*	-	this work
YFL3876	BY4741	*vid22(L893P)-13Myc:LEU2MX6*	-	this work
YFL3895	BY4741	*vid22(K873A-K875A)-13Myc:LEU2MX6*	-	this work
YFL3897	BY4741	*vid22(N878A)-13Myc:LEU2MX6*	-	this work

Name	*Background*	*Genotype*	*Plasmid*	*Source*
YFL3900	BY4741	*vid22(D887A-D889A-D891A)-13Myc:LEU2MX6*	-	this work
YFL3902	BY4741	*vid22Δ::KANMX6*	pGELE36	this work
YFL3904	BY4741	*vid22Δ::KANMX6*	pFL276	this work
YFL3987	BY4741	*VID22-3Flag:KANMX6*	pFL241	this work
YFL3991	BY4741	-	pFL241	this work
YFL3993	BY4741	*VID22-13Myc:LEU2MX6 TBF1-3Flag:KANMX6*	-	this work
YFL3995	BY4741	*vid22(N878A)-13Myc:LEU2MX6 TBF1-3Flag:KANMX6*	-	this work
YFL3996	BY4741	*vid22(Q880A)-13Myc:LEU2MX6 TBF1-3Flag:KANMX6*	-	this work
YFL3997	BY4741	*vid22(D887A-D889A-D891A)-13Myc:LEU2MX6 TBF1-3Flag:KANMX6*	-	this work
YFL3999	BY4741	*vid22(L893P)-13Myc:LEU2MX6 TBF1-3Flag:KANMX6*	-	this work
YFL4001	BY4741	*vid22(L893P)-13Myc:LEU2MX6 TBF1-3Flag:KANMX6*	-	this work
YFL4014	BY4741	*VID22-3Flag:KANMX6*	pFL247	this work
YFL4015	BY4741	*VID22-3Flag:KANMX6*	pFL243	this work
YFL4016	BY4741	*VID22-3Flag:KANMX6*	pFL248	this work
YFL4018	BY4741	*VID22-3Flag:KANMX6*	pFL245	this work
YFL4020	BY4741	*VID22-3Flag:KANMX6*	pFL246	this work
YFL4158	BY4741	*VID22-3Flag:KANMX6*	pFL330	this work
YGE1014	BY4741	*chrVIII 512307G:C 512323G:C*	-	our lab
YGE1033	YGE1014	*TBF1-13Myc:HIS3MX6*	-	our lab
YGE1071	BY4741	*vid22Δ::HPHMX6 TBF1-3Flag:KANMX6*	pGELE36	our lab
YGE1073	BY4741	*vid22Δ::HPHMX6 TBF1-3Flag:KANMX6*	pFL276	our lab

Name	*Background*	*Genotype*	*Plasmid*	*Source*
YGE253	YHN93	-	pEG202 pJG4-5	our lab
YGE651	BY4741	*VID22-13Myc:HIS3MX6*	-	our lab
YGE654	BY4741	*TBF1-13Myc:HIS3MX6*	-	our lab
YGE671	BY4741	*vid22Δ::KANMX6 TBF1-13Myc:HIS3MX6*	-	our lab
YGE722	BY4741	-	pGREG 535	this work
YGE723	BY4741	-	pGELE23	this work
YGE877	BY4741	-	pRS415	our lab
YGE880	BY4741	*vid22Δ::KANMX6*	pRS415	our lab
YHN93	EGY48	-	pSH18-34	R. Brent
YNOV124	BY4741	*vid22Δ::KANMX6*	-	our lab

7.1.2 Plasmids

Name	*Insert*	*Selection in E. coli, yeast*	*Additional features*	*Size (bp)*	*Source*
pEG202	*pADH1:lexA*	*bla, HIS3*	*E. coli* ori, *S. cerevisiae* 2µ replication origin	10166	R. Brent
pFL228	*pGAL1:7HA-vid22(N878A)*	*bla, LEU2 HIS3 KANMX6*	*E. coli* ori, *S. cerevisiae CEN6 ARS*	11227	this work
pFL230	*pGAL1:7HA-vid22(K873A-K875A)*	*bla, LEU2 HIS3 KANMX6*	*E. coli* ori, *S. cerevisiae CEN6 ARS*	11227	this work
pFL231	*pGAL1:7HA-vid22(Q880A)*	*bla, LEU2 HIS3 KANMX6*	*E. coli* ori, *S. cerevisiae CEN6 ARS*	11227	this work
pFL232	*pGAL1:7HA-vid22(D887A-D889A-D891A)*	*bla, LEU2 HIS3 KANMX6*	*E. coli* ori, *S. cerevisiae CEN6 ARS*	11227	this work
pFL234	*pGAL1:7HA-vid22(L893P)*	*bla, LEU2 HIS3 KANMX6*	*E. coli* ori, *S. cerevisiae CEN6 ARS*	11227	this work
pFL241	*pVID22:VID22-13Myc*	*bla, LEU2*	*E. coli* ori, *S. cerevisiae CEN6 ARS*	9987	this work
pFL243	*pVID22:vid22 (N878A)-13Myc*	*bla, LEU2*	*E. coli* ori, *S. cerevisiae CEN6 ARS*	9987	this work
pFL245	*pVID22:vid22 (D887A-D889A-D891A)-13Myc*	*bla, LEU2*	*E. coli* ori, *S. cerevisiae CEN6 ARS*	9987	this work
pFL246	*pVID22:vid22 (L893P)-13Myc*	*bla, LEU2*	*E. coli* ori, *S. cerevisiae CEN6 ARS*	9987	this work
pFL247	*pVID22:vid22 (K873A-K875A)-13Myc*	*bla, LEU2*	*E. coli* ori, *S. cerevisiae CEN6 ARS*	9987	this work
pFL248	*pVID22:vid22 (Q880A)-13Myc*	*bla, LEU2*	*E. coli* ori, *S. cerevisiae CEN6 ARS*	9987	this work
pFL276	*pVID22:vid22 (C87A-C90R-H110Y-H115Y)-13Myc*	*bla, LEU2*	*E. coli* ori, *S. cerevisiae CEN6 ARS*	9987	this work
pFL330	*pVID22:vid22 (F194A)-13Myc*	*bla, LEU2*	*E. coli* ori, *S. cerevisiae CEN6 ARS*	9987	this work
pGELE23	*pGAL1:7HA-VID22*	*bla, LEU2 HIS3 KANMX6*	*E. coli* ori, *S. cerevisiae CEN6 ARS*	11227	our lab
pGELE36	*pVID22:VID22-13Myc*	*bla, LEU2*	*E. coli* ori, *S. cerevisiae CEN6 ARS*	9987	our lab
pGELE53	*pGAL1:7HA-vid22#8*	*bla, LEU2 HIS3 KANMX6*	*E. coli* ori, *S. cerevisiae CEN6 ARS*	11089	this work
pGELE54	*pGAL1:7HA-vid22#29*	*bla, LEU2 HIS3 KANMX6*	*E. coli* ori, *S. cerevisiae CEN6 ARS*	11227	this work
pGELE55	*pGAL1:7HA-vid22#36*	*bla, LEU2 HIS3 KANMX6*	*E. coli* ori, *S. cerevisiae CEN6 ARS*	11092	this work

Name	*Insert*	*Selection in E. coli, yeast*	*Additional features*	*Size (bp)*	*Source*
pGELE56	*pGAL1:7HA-vid22#44*	*bla, LEU2 HIS3 KANMX6*	*E. coli* ori, *S. cerevisiae CEN6 ARS*	11227	this work
pGELE57	*pGAL1:7HA-vid22#69*	*bla, LEU2 HIS3 KANMX6*	*E. coli* ori, *S. cerevisiae CEN6 ARS*	10885	this work
pGELE58	*pGAL1:7HA-vid22#76*	*bla, LEU2 HIS3 KANMX6*	*E. coli* ori, *S. cerevisiae CEN6 ARS*	11044	this work
pGELE59	*pADH1:lexA-VID22*	*bla, HIS3*	*E. coli* ori, *S. cerevisiae* 2µ replication origin	12807	this work
pGELE60	*pGAL1:NLS-B42-HA-VID22*	*bla, TRP1*	*E. coli* ori, *S. cerevisiae* 2µ replication origin	8994	this work
pGELE61	*pADH1:lexA-TBF1*	*bla, HIS3*	*E. coli* ori, *S. cerevisiae* 2µ replication origin	11790	this work
pGELE62	*pGAL1:NLS-B42-HA-TBF1*	*bla, TRP1*	*E. coli* ori, *S. cerevisiae* 2µ replication origin	7977	this work
pGREG 535	*pGAL1:7HA*	*bla, LEU2 HIS3 KANMX6*	*E. coli* ori, *S. cerevisiae CEN6 ARS*	9689	Euroscarf
pJG4-5	*pGAL1:NLS-B42-HA*	*bla, TRP1*	*E. coli* ori, *S. cerevisiae* 2µ replication origin	6449	R. Brent
pRS415	-	*bla, LEU2*	*E. coli* ori, *S. cerevisiae CEN6 ARS*	6021	R. S. Sikorski
pSH18-34	*8LexAop:lacZ*	*bla, URA3*	*E. coli* ori, *S. cerevisiae* 2µ replication origin	10484	R. Brent

7.1.3 Oligos

Name	*5'-3' sequence*	*Direction*	*Length (bases)*	*T_m* (°C)*	*Purpose*
chrVIII:512387_F	ATGAGCAAAATGTGGTCAGC	Forward	20	56	qPCR analysis
chrVIII:512564_R	ACCCAAACAAAAGCAGCAAG	Reverse	20	57	qPCR analysis
chrIX:382175_F	AACCAAGACATCACTGCTTCC	Forward	21	57	qPCR analysis
chrIX:382332_R	GATCGGCGCCAATATTAAAACG	Reverse	22	57	qPCR analysis
chrXV:216099_F	GCCTGATCGTGACCTAGACC	Forward	20	58	qPCR analysis
chrXV:216259_R	CACATAGGCTTTCCCCATTCG	Reverse	21	58	qPCR analysis
chrIV:264290_F	GGAAGAAAAGCACACCCTAGG	Forward	21	57	qPCR analysis
chrIV:264489_R	CGCATCATTTTCACGTCATACTC	Reverse	23	56	qPCR analysis
chrVIII:396975_F	TTCAAAACGTGTCAGGGCTAC	Forward	21	58	qPCR analysis
chrVIII:397108_R	CGCGGCGGAAATAATAAAAGC	Reverse	21	59	qPCR analysis
SKN7_CDS_F	GGCGAACAAAGTTGGGAGTT	Forward	20	59	qPCR analysis
SKN7_CDS_R	TGTTGGGGTTAGTGCCTTCA	Reverse	20	57	qPCR analysis
TBF1_CDS_F	GTCGGTAACAACTCCCTGGA	Forward	20	58	qPCR analysis
TBF1_CDS_R	CGTTCGTCTCGTCCTGAAAC	Reverse	20	59	qPCR analysis
HHT2_F	GCGTGAATAGCAGCCAGATTAGT	Forward	23	59	qPCR analysis
HHT2_R	TCAATCTTCTGCTATCGGTGCTT	Reverse	23	59	qPCR analysis
HIS7_F	AGTTGTCCAGGTTCGATGATTC	Forward	22	58	qPCR analysis
HIS7_R	TGCAATTTTCCAACCGTCATTTTC	Reverse	24	57	qPCR analysis
VID22_C87A-C90R_F	ATGCTGGAGGCAGTAAAAGCCAAGTACCGCGGTGTGATAATAAGACGG	Forward	48	62	QuikChange site-directed mutagenesis
VID22_C87A-C90R_R	CCGTCTTATTATCACACCGCGGTACTTGGCTTTTACTGCCTCCAGCAT	Reverse	48	64	QuikChange site-directed mutagenesis
VID22_H110Y-H115Y_F	GAAGCCTCGCAAACTTATTTGTGGAGCACGTATAAGATAGACCCGA	Forward	46	66	QuikChange site-directed mutagenesis
VID22_H110Y-H115Y_R	TCGGGTCTATCTTATACGTGCTCCACAAATAAGTTTGCGAGGCTTC	Reverse	46	67	QuikChange site-directed mutagenesis
VID22_F195A_F	CAGAAAACTTGCCTCTGTCGGCCGTAGACAACACTGCCGTACG	Forward	43	66	QuikChange site-directed mutagenesis
VID22_F195A_R	CGTACGGCAGTGTTGTCTACGGCCGACAGAGGCAAGTTTTCTG	Reverse	43	66	QuikChange site-directed mutagenesis

Name	*5'-3' sequence*	*Direction*	*Length (bases)*	*T_m* (°C)*	*Purpose*
VID22_K873A-K875A_F	GATTTAACTCAAGAGATTATTGCAATTGCATTATTCAATGATCAATTCG	Forward	49	50	QuikChange site-directed mutagenesis
VID22_K873A-K875A_R	CGAATTGATCATTGAATAATGCAATTGCAATAATCTCTTGAGTTAAATC	Reverse	49	50	QuikChange site-directed mutagenesis
VID22_N878A_F	GATTATTAAAATTAAATTATTCGCAGATCAATTCGTGGCAAGTAAAG	Forward	47	54	QuikChange site-directed mutagenesis
VID22_N878A_R	CTTTACTTGCCACGAATTGATCTGCGAATAATTTAATTTTAATAATC	Reverse	47	53	QuikChange site-directed mutagenesis
VID22_Q880A_F	GATTATTAAAATTAAATTATTCAATGATGCATTCGTGGCAAGTAAAGTGG	Forward	50	56	QuikChange site-directed mutagenesis
VID22_Q880A_R	CCACTTTACTTGCCACGAATGCATCATTGAATAATTTAATTTTAATAATC	Reverse	50	57	QuikChange site-directed mutagenesis
VID22_D887A-D889A-D891A_F	CGTGGCAAGTAAAGTGGCATACGCATTGGCAACTTTACAAACTGCAAG	Forward	48	55	QuikChange site-directed mutagenesis
VID22_D887A-D889A-D891A_R	CTTGCAGTTTGTAAAGTTGCCAATGCGTATGCCACTTTACTTGCCACG	Reverse	48	54	QuikChange site-directed mutagenesis
VID22_L893P_F	GTGGATTACGACTTGGACACTCCACAAACTGCAAGTCAGTATCTTC	Forward	46	62	QuikChange site-directed mutagenesis
VID22_L893P_R	GAAGATACTGACTTGCAGTTTGTGGAGTGTCCAAGTCGTAATCCAC	Reverse	46	63	QuikChange site-directed mutagenesis
DP_VIII_NEW_F	AAGTCGATTAAAAGTAGGGCTAACACAAGTATAAGCAGGGGAGCTCGTTTTCGACACTGG	Forward	60	58	Delitto Perfetto (CORE cassette integration)
DP_VIII_NEW_R	TGAGCCCTGCCCTCACTTATGGCATCCCTATTATAGCAAATCCTTACCATTAAGTTGATC	Reverse	60	49	Delitto Perfetto (CORE cassette integration)
TBS_mut_B_F	AAAGTCGATTAAAAGCCGGGCTAACACAAGTATAAGCAGGGTTTTGCTATAACCGGGATG	Forward	60	-	Delitto Perfetto (replacement cassette)
TBS_mut_C_R	TGAGCCCTGCCCTCACTTATGGCATCCCGGTTATAGCAAAACCCTGCTTATACTTGTGTT	Reverse	60	-	Delitto Perfetto (replacement cassette)

Name	*5′-3′ sequence*	*Direction*	*Length (bases)*	T_m* (°C)	*Purpose*
TBS_mut_A_F	AAAAAAAAAAAAAAAAAGTTTTTCTTAAATGCATAGGTTTAAAGTCGATTAAAAGTCGGG	Forward	60	56	Delitto Perfetto (replacement cassette)
TBS_mut_D_R	AAAAATTCAGAATTTCGGAAAATCCATGTACGCGCATCGATGAGCCCTGCCCTCACTTAT	Reverse	60	60	Delitto Perfetto (replacement cassette)
TBS-G4_mut_B_F	AAAGTCGATTAAAAGTCGGGCTACACAAGTATAAGCAGGGTTTTGCTATAACCGCGATG	Forward	60	-	Delitto Perfetto (replacement cassette)
TBS-G4_mut_C_R	TGAGCCCTGCGCTCACTTATGGCATCGCGGTTATAGCAAAACCCTGCTTATACTTGTGTT	Reverse	60	-	Delitto Perfetto (replacement cassette)
TBS-G4_mut_A_F	AAAAAAAAAAAAAAAAAGTTTTTCTTAAATGCATAGGTTTAAAGTCGATTAAAAGTCGGG	Forward	60	56	Delitto Perfetto (replacement cassette)
TBS-G4_mut_D_R	AAAAATTCAGAATTTCGGAAAATCCATGTACGCGCATCGATGAGCCCTGCGCTCACTTAT	Reverse	60	60	Delitto Perfetto (replacement cassette)
VID22_F1-Δ	GCTGGCAGCATCCACAGGATAAACAAGCACGAGAGCACATCGGATCCCCGGGTTAATTAA	Forward	60	55	Deletion
VID22_F2-HA/Myc	CGACTTGGACACTTTACAAACTGCAAGTCAGTATCTTCCACGGATCCCCGGGTTAATTAA	Forward	60	55	C-terminal epitope tagging
VID22_F2-Flag	CGACTTGGACACTTTACAAACTGCAAGTCAGTATCTTCCATCTCATCATCATCATCATGG	Forward	60	50	C-terminal epitope tagging
VID22_856-901Δ_F	ATGCAATGTCAGTCATCTTCCTCCATTAGGGGAGAGTACTCATTTCGGATCCCCGGGTTAATTAA	Forward	65	54	Mutagenesis and C-terminal epitope tagging
VID22_R1	CTAAAATTTAAAACAAACAGACTTTGTTCCTTTGTATGTCGAATTCGAGCTCGTTTAAAC	Reverse	60	53	Deletion, epitope tagging
VID22_R1-pFL241	CTAAAATTTAAAACAAACAGACTTTGTTCCTTTGTATGTCTTTCCTTATCACGTTGAGCC	Reverse	60	54	Mutagenesis and C-terminal epitope tagging
TBF1_F2-Myc	ACATCGGACAATACAGGTTTTGATCCTCATTTAGAAGATGGGATGCGGATCCCCGGGTTAATTAA	Forward	65	55	C-terminal epitope tagging
TBF1_F2-Flag	ACATCGGACAATACAGGTTTTGATCCTCATTTAGAAGATGGGATGTCTCATCATCATCATCATGG	Forward	65	50	C-terminal epitope tagging

Name	*5'-3' sequence*	*Direction*	*Length (bases)*	*T_m* (°C)*	*Purpose*
TBF1_R1	TGGTTCAATAGAAAATTAGAAGTCTGTTCATTTAAGCGCGGAATTCGAGCTCGTTTAAAC	Reverse	60	53	C-terminal epitope tagging
Vid22_F_pEG202	TGGCGGTTGGGGTTATTCGCAACGGCGACTGGCTGATGAGAGCGATGGACACACAGGTAC	Forward	60	62	Yeast two hybrid plasmid construction
Vid22_F_pJG4-5	GCCTCCTACCCTTATGATGTGCCAGATTATGCCATGAGAGCGATGGACACACAGGTACAG	Forward	63	63	Yeast two hybrid plasmid construction
Vid22_TH_R	CTTATTTAATAATAAAAATCATAAATCATAAGAAATTCGCCTATGGAAGATACTGACTTGCAG	Reverse	63	56	Yeast two hybrid plasmid construction
Tbf1_F_pEG202	GCGGTTGGGGTTATTCGCAACGGCGACTGGCTGATGGATTCGCAAGTGCCCAATAATAAC	Forward	60	60	Yeast two hybrid plasmid construction
Tbf1_F_pJG4-5	GCCTCCTACCCTTATGATGTGCCAGATTATGCCATGGATTCGCAAGTGCCCAATAATAAC	Forward	60	60	Yeast two hybrid plasmid construction
Tbf1_TH_R	CTTATTTAATAATAAAAATCATAAATCATAAGAAATTCGCCTACATCCCATCTTCTAAATGA	Reverse	62	50	Yeast two hybrid plasmid construction

Underlined portions in oligos sequence do not anneal to template DNA. *T_m of the annealed portion of the oligo was retrieved from SnapGene which assumes a 50mM Na^+ concentration and a 0,25µM oligo concentration.

Oligos were purchased from Eurofins Genomics.

7.1.4 Antibodies

Antibody	*Specificity*	*Clonality*	*Host*	*Dilution*	*Source*
9E10 Myc	Myc tag	monoclonal	mouse	1:30	supernatant from hybridoma cultures
12CA5 HA	HA tag	monoclonal	mouse	1:30	supernatant from hybridoma cultures
ANTI-FLAG® M2 (F1804)	Flag tag	monoclonal	mouse	1:2000	Sigma-Aldrich
LexA Antibody 2-12	LexA	monoclonal	mouse	1:300	Santa Cruz
Pgk1 (CSB-PA00035A0Rb)	Pgk1	polyclonal	rabbit	1:2000	Cusabio
Goat anti-mouse HRP	mouse IgG	polyclonal	goat	1:30000	Thermo Fisher
Goat anti-rabbit HRP	rabbit IgG	polyclonal	goat	1:30000	Thermo Fisher

7.2 Yeast strain development, growth and treatment conditions

Yeast strains were obtained by transformation with the LiAc/ssDNA/PEG method[453–455]. The one-step PCR-mediated technique[456–458] was used to replace by homologous recombination target genomic sequences for mutagenesis and epitope tagging purposes. All plasmids and yeast strains were sequenced (Eurofins Genomics) at targeted sites to confirm proper cloning, tagging and/or mutagenesis. *VID22* cloned in different plasmids (pGELE36, pGELE23, pFL241) was subjected to *in vitro* site-directed mutagenesis with the QuikChange approach[419]. The obtained plasmids were either directly transformed in cells or used as PCR templates for the generation of recombination cassettes. Plasmid pFL241 was obtained from pGELE36 by PvuII digestion and *in vitro* cloning to invert p*VID22*:*VID22-13Myc* direction such that *LEU2* selection marker is placed downstream. Mutagenesis of the *SKN7* promoter was accomplished through the 'Delitto Perfetto' strategy[412,459] with oligos listed in Table 7.1.3. *VID22* random mutagenesis was performed by 5 independent PCR amplifications unbalancing one or two dNTPs at a time (0,05mM-1mM dCTP, 0,05mM-1mM dTTP, 2mM dGTP, 2mM dATP) and with tailed primers

carrying homology towards pGREG535. Mutant alleles were cloned *in vivo* by gap repair[460] in the recipient vector linearised with SalI. Cells were plated on solid 2% raffinose, 2% galactose plasmid-selective media and ~150 surviving clones were selected for protein size and accumulation analysis. The obtained plasmids were purified from the selected yeasts, amplified in *E. coli* and sequenced. The strains used in the related spot assay were obtained by transformation of wild-type cells with purified vectors. Yeast two hybrid vectors were obtained by PCR amplification of wild-type *VID22* and *TBF1* with tailed primers carrying vector-homologous sequence and cloned *in vivo* by gap repair[460] in both bait pEG202 and prey pJG4-5 plasmids linearised with BamHI + EcoRI and EcoRI + XhoI, respectively.

Yeast cells were grown on liquid or solid (2% agar) complete (1% yeast extract; 2% peptone; 2% glucose/raffinose/galactose as specified) or synthetic-defined (0,17% Yeast Nitrogen Base Difco™; 0,5% ammonium sulfate; 0,05mg/mL L-Thr; 0,025mg/mL L-Phe; 0,025mg/mL L-Tyr; 0,025mg/mL L-Lys; 0,025mg/mL L-Ile; 0,025mg/mL L-Arg; 0,025mg/mL L-Met; 0,025 mg/ml adenine; 2% glucose/raffinose/galactose as specified) media. To maintain selection, proper combination of additional amino acids and bases were supplemented to synthetic-defined media (0,05mg/mL L-Trp; 0,05mg/mL L-Leu; 0,05mg/mL L-His; 0,05mg/mL uracil). When antibiotics were added to the synthetic-defined medium (50µg/mL geneticin, G418), ammonium sulfate was replaced by L-Glu. For spot assays testing sensitivity to genotoxic agents, 50mM, 100mM or 200mM hydroxyurea (HU, USBiological) or 125ng/µL, 250ng/µL or 500ng/µL bleomycin sulfate (BM, Sigma) were supplemented to the solid medium. Strains for the yeast two hybrid X-Gal plate assay were plated on 2% agar; 7mg/mL sodium hydrogen phosphate hydrate; 3mg/mL sodium dihydrogen phosphate; 0,17% Yeast Nitrogen Base Difco™; 0,5% ammonium sulfate; 0,05mg/mL L-Thr; 0,025mg/mL L-Phe; 0,025mg/mL L-Tyr; 0,025mg/mL L-Lys; 0,025mg/mL L-Ile; 0,025mg/mL L-Arg; 0,025mg/mL L-Met; 0,05mg/mL L-Leu; 0,025mg/mL adenine; 2,5% galactose; 2,5% raffinose; 0,08mg/mL 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside. For TMPyP4 (Sigma) treatment, overnight cultures grown in liquid glucose-containing synthetic-defined media were diluted in fresh media supplemented with 100µM TMPyP4 and grown at 28°C in the dark. The OD_{600} of the treated cultures was

measured every hour for 8 hours. Cell cycle progression was monitored by flow cytometry[415] and analysed with FlowJo v10 software[33].

7.3 Spot assay

Qualitative growth tests were performed starting from dilution of log-phase yeast cultures to 2×10^6 cells/mL (OD_{600} 0,06). Three 10-fold serial dilutions were prepared for each strain and 8µL of the four dilutions plated on solid complete or synthetic-defined media containing the specified carbon source. Growth medium was supplemented with 50mM, 100mM or 200mM hydroxyurea (HU, USBiological) or 125ng/µL, 250ng/µL or 500ng/µL bleomycin sulfate (BM, Sigma) where indicated. Plates were incubated at 28°C for 3 days and scanned for documentation.

7.4 RNA extraction and RT-qPCR analysis

Total RNAs were isolated from 2×10^8 log-phase yeast cells with High Pure RNA Isolation Kit (Roche) following the manufacturer protocol. Following sample quantification with NanoDrop™ UV-Vis Spectrophotometer (Thermo Scientific), 500ng RNA were retrotranscribed with iScript™ cDNA Synthesis Kit (Bio-Rad) in a reaction mix final volume of 20µL/sample by 5min incubation at 25°C, 30min at 42°C, and 5min at 85°C. cDNA was stored at -20°C prior qPCR analysis. cDNAs were analysed using primer pairs listed in Table 7.1.3 producing amplicons within 120-200bp size range. qPCR was performed in 20µL reactions containing ready-to-use Quantitative Master Mix with SYBR® Green GeneSpin (with Xtra Taq Pol, dNTPs, $MgCl_2$ and stabilizers optimized for use in real time PCR amplification), primers (130-150nM) and template (2µL 1:5 cDNA dilution). All components were mixed in 96-well hard-shell PCR plates with clear wells (Bio-Rad), which were sealed with optically clear adhesive Microseal® 'B' seals (Bio-Rad) and centrifuged at 1000×g for 10s. The thermal conditions during reaction were 3min at 95°C followed by 39 thermal cycles at 62°C for 30s and 95°C for 10s. qPCR was performed on a CFX Connect Real Time PCR System (Bio-Rad). PCR efficiencies were 90–105% calculated from standard curves of serial dilutions of input samples (1:5, 1:25, 1:125, 1:625 and 1:3125). All reactions were performed in technical triplicates, along with no template controls. Relative gene expression was calculated for the target gene

relative to the *HHT2* internal standard by subtracting their respective cycle thresholds (ΔCt). This difference was then subtracted to that of the reference sample (ΔΔCt) and the obtained value utilized as the negative exponent of 2 ($2^{-\Delta\Delta Ct}$). Means of relative gene expression and standard errors were calculated for biological replicates and unpaired Student's t-test was used to determine statistical significance between samples.

7.5 Chromatin immunoprecipitation

ChIP was performed as described[263]. Yeast strains used in ChIP experiments were obtained by C-terminally tagging the proteins of interest with either 13Myc or 3Flag epitopes. Overnight cell cultures, pre-grown in complete or synthetic-defined media, were diluted into 50mL fresh medium to a cell density of 4×10^6 cell/mL and grown at 28°C until they reached a density of $1.5–2\times10^7$ cell/mL before being cross-linked with 1% formaldehyde (Sigma) for 10 min at 25°C. Cross-linking was quenched by addition of 125 mM glycine (Sigma) prior to cell harvesting, washing and freezing at -80°C until further analysis. Approximately 5×10^8 log-phase cell pellets were resuspended in 600µL in ChIP lysis buffer (1% NP-40 for ChIP anti-Myc, or 1% Triton X-100 for ChIP anti-Flag; 140mM NaCl; 50mM HEPES pH 7,5; 1mM EDTA pH 8,0; 0,1% sodium deoxycholate; 1mM phenylmethylsulfonyl fluoride Sigma; cOmplete Protease Inhibitor Tablets Roche of appropriate size and quantity) and 600µL acid-treated glass beads (425-600µm, Sigma) and mechanically lysed by bead beating using the high-speed benchtop homogenizer FastPrep-24™ Classic Instrument (MP Biomedicals) for six 15s cycles of vibrations at a speed of 5.5m/s, each followed by a 1min sample incubation on ice. Following centrifugation at 16000×g for 30min at 4°C, pellets were resuspended in 300µL ChIP lysis buffer and sonicated for 15min (10 cycles, 30s on, 60s off) at high power at 4°C with a Bioruptor Plus sonicator (Diagenode). Following addition of 300µL ChIP lysis buffer and two refrigerated centrifugation steps at 9000×g for 5min and 15min, respectively, input samples were collected (1/20 of total volume). Supernatants were then incubated with 30µL/sample Dynabeads™ Protein G (Invitrogen) functionalised with 12,5µg appropriate primary antibody (Table 7.1.4) for 2h at 4°C on rotation. Washing steps and reverse crosslinking were performed as previously described[461]. Input and

immunoprecipitated DNA was purified using Wizard® SV Gel and PCR Clean-Up System (Promega) and elution performed in 60–80µL nuclease-free water. DNA was quantified with Qubit 4 Fluorometer using Qubit dsDNA HS Assay kit (Invitrogen) to approximately 0.5–1 ng/µL in the input and 0.1–0.5ng/µL in the immunoprecipitated samples. Samples were analysed by qPCR and stored at −20°C.

7.6 qPCR analysis

qPCR analysis of ChIP samples was performed as described[263]. Input and immunoprecipitated DNA were analysed using primer pairs listed in Table 7.1.3 producing amplicons within 80-200bp size range. qPCR was performed in 25µL reactions containing ready-to-use Quantitative Master Mix with SYBR® Green 2× GeneSpin (with Xtra Taq Pol, dNTPs, $MgCl_2$ and stabilizers optimized for use in real time PCR amplification), primers (130-200nM) and template (2µL 1:100 input or undiluted immunoprecipitated DNA) or in 20µL reactions containing ready-to-use Luna® Universal qPCR Master Mix 2× NEB (allowing quantitation of target DNA sequences using the SYBR®/FAM channel and containing Hot Start *Taq* DNA Polymerase), primers (350-390nM) and template (5µL 1:100 input or undiluted immunoprecipitated DNA). All components were mixed in 96-well hard-shell PCR plates with clear wells (Bio-Rad), which were sealed with optically clear adhesive Microseal® 'B' seals (Bio-Rad) and centrifuged at 1000×g for 10s. The thermal conditions during reaction were 3min at 95°C followed by 39 thermal cycles at 62°C for 30s and 95°C for 10s with GeneSpin Master Mix or 1min 15s at 95°C followed by 39 thermal cycles at 58°C for 45s and 95°C for 15s with LUNA Master Mix. qPCR was performed on a CFX Connect Real Time PCR System (Bio-Rad). PCR efficiencies were 90–105% calculated from standard curves of serial dilutions of input samples (1:5, 1:25, 1:125, 1:625 and 1:3125). All reactions were performed in technical triplicates, along with no template controls. Fold enrichment was calculated using the PfaffI method[462] from the difference between immunoprecipitated and input cycle threshold values (after correcting for input sample dilution) of the target gene and the *HHT2* or *HIS7* internal standards. Means of fold enrichment and standard errors were calculated for biological replicates and unpaired Student's t-test was used to determine statistical significance between samples.

7.7 Protein extraction, SDS-PAGE, and Western blot analysis

Protein extraction, SDS-PAGE and detection by Western blot analysis were performed for all tagged proteins. Denatured total protein extracts were obtained by trichloroacetic acid extraction[463]. Protein extracts were resolved by SDS-PAGE (10% polyacrylamide resolving gel). Western blots were performed with the indicated tag-specific primary antibodies listed in Table 7.1.4., and goat anti-mouse HRP as secondary antibody, by standard procedures. Chemiluminescence signal was acquired using a ChemiDoc™ Touch Imaging System (Bio-Rad).

7.8 Co-immunoprecipitation

Yeast strains used in co-IP experiments were obtained by C-terminally tagging the proteins of interest with either 13Myc, 3HA, or 3Flag epitopes. Overnight cell cultures, pre-grown in complete or synthetic-defined media, were diluted into 50mL fresh medium to a cell density of 4×10^6 cell/mL and grown at 28°C until they reached a density of $1.5–2\times10^7$ cell/mL before being harvested, washed and frozen at -80°C until further analysis. Approximately 5×10^8 log-phase cell pellets were resuspended in 500µL Inhibitor Buffer (50mM Tris-HCl pH 7,5; 150mM NaCl; 5mM EDTA pH 8,0; 0,1% Triton X-100; 50mM NaF; 10mM β-glycerophosphate; 1mM phenylmethylsulfonyl fluoride Sigma; 1mM Na_3VO_4; cOmplete Protease Inhibitor Tablets Roche of appropriate size and quantity) and 600µL acid-treated glass beads (425-600µm, Sigma) and mechanically lysed by bead beating using the high-speed benchtop homogenizer FastPrep-24™ Classic Instrument (MP Biomedicals) for six 15s cycles of vibrations at a speed of 5.5m/s, each followed by a 1min sample incubation on ice. Following lysate clarification by centrifugation at 16000×g for 30 min at 4°C, total protein concentration of the supernatant was estimated through a Bradford protein assay by measurement of the absorbance at 595nm of sample aliquots stained with Coomassie Brilliant Blue (Bio-Rad) and normalised by means of a standard BSA calibration curve to 2mg/mL unless stated otherwise in 1mL final volume. Following collection of input sample (1/20 of total volume), ~2mg of normalised proteins were incubated with 40µg free tag-specific primary antibody (Table 7.1.4) for 2h at 4°C on rotation, and, subsequently, with 40µL/sample Agarose Protein G (GeneSpin) bead suspension in 1×PBS, pre-saturated with

0,1mg/mL BSA. Following collection of 50µL immunodepleted sample controls and washing of the unbound material with inhibitor buffer, the immunoprecipitated fractions were recovered in Laemmli buffer (60mM Tris-HCl pH 6,8; 2% sodium dodecyl sulfate; 10% glycerol; 100mM dithiothreitol; 0,025% bromophenol blue) at 95°C for 5min. Input, immunodepleted and immunoprecipitated samples were analysed by Western blot.

7.9 Yeast two hybrid assay

The bait proteins were fused to the *E. coli* sequence-specific DNA-binding LexA repressor, which recognizes a 16 bp palindromic operator sequence (5'-CTGTATATATATACAG-3'), while the preys were joined to the *E. coli* B42 acidic sequence acting as a transcriptional activation domain. Yeast two hybrid constructs and control vectors were transformed in the appropriate combinations in the yeast strain YHN93 carrying the reporter plasmid pSH18-34 with four LexA operators upstream of *lacZ*. The obtained strains were grown in both inductive and non-inductive selective media (-His, -Trp, -Ura, 2% raffinose, 2% galactose as inductor) for 4h. Proper accumulation of the protein chimaeras was confirmed by Western blot. The obtained strains and appropriate controls were patched on agar plates containing the substrate X-gal for assaying β-galactosidase activity, grown at 28°C over-night in the dark and scanned for documentation.

7.10 Purification and negative staining electron microscopy

Vid22 was purified from BL21(DE3)pLysS *E. coli* cells. Cells were grown in LD with 50g/mL ampicillin at 30°C until OD_{600nm} 0,5 prior to induction of protein expression by addition of 0,2mM IPTG at 30°C for 3-4h. For protein purification, cells were disrupted in Lysis buffer (300mM NaCl, 50mM sodium phosphate pH 7) supplemented with 30mM imidazole by sonication for 15min (10 cycles, 30s on, 60s off) at high power at 4°C with a Bioruptor Plus sonicator (Diagenode). Protein extracts were clarified by centrifugation at 27200×g for 1h at 4°C. Clarified extracts were then loaded into a Ni-NTA Agarose matrix (Qiagen) previously washed with Lysis buffer supplemented with 50mM imidazole. Proteins were eluted in Lysis buffer with 250mM imidazole. Proteins were dialysed in 200mM NaCl, 50mM Tris-

HCl pH 7,5 using Thermo Scientific™ SlideA-Lyzer™ Dialysis Cassettes 3500 MWCO. For biochemical assays, proteins were subjected to a second purification step through size exclusion chromatography with multi-angle static light scattering. 500µL of the 1mg/mL sample were loaded on a Superdex® 200 Increase 10/300 GL column (10-600 kDa, Sigma) with a 0,4mL/min flow. Main peak corresponding to Vid22 dimer was eluted at ~27min. For negative staining electron microscopy analysis, 4 µL of Vid22 at 0,02mg/mL in 200mM NaCl, 50mM Tris-HCl pH 7,5 buffer were transferred to glow-discharged 400 mesh carbon coated Cu grids (Agar Scientific) and grids were blotted after 1min with Whatman filter paper. After washing, grids were contrasted with 2% uranyl acetate for 1min and blotted. Observations were carried out on a Talos L120C microscope (Thermo Scientific) with an accelerating voltage of 120 kV, and images acquired with a Ceta 4k×4k camera at a nominal magnification of 73.000 corresponding to a pixel size of 1.9 Å/pix. 169 micrographs were recorded within a defocus range of -1,0µm to -3,5µm, and then imported in RELION 3.1[464] for data preprocessing. CTF estimation was carried out with CTFFIND4[465]. 70 micrographs were selected with a defocus range within -1,3µm to -2,5µm and 23.664 particles were picked with a Laplacian-of-Gaussian protocol with a radius range within 100Å-200Å. Particles were extracted with a box size of 128 pix and imported in CRYOSPARC 4.2.1[422]. After two rounds of 2D classification, 13.128 particles were selected and used for *ab initio* reconstruction with one single class and C1 symmetry. The initial model was used for a homogeneous refinement with C2 symmetry and a final resolution of 16Å, according to the gold standard FSC of 0.143.

7.11 Evolutionary conservation analysis

Protein sequences of 332 budding yeasts species were retrieved from published data[285]. A local BLAST database of 1.689.424 proteins from 332 budding yeasts genomes was compiled from the FASTA file of the aminoacidic sequences using the NCBI *makeblastdb* application via Microsoft Windows Command Prompt. Using BLAST2.7.1+ suite of command-line tools, homologous sequences to Vid22 query were searched in the local database with a protein-protein BLAST. The multiple sequence alignment of Vid22 homologs was performed with MUSCLE progressive

alignment algorithm[418], using SeaView user interface[466]. No manual corrections were carried out. Jalview[467] was used to visualise and export the multiple sequence alignment after masking redundant sequences. Only the columns of the multiple sequence alignments corresponding to Vid22 sequence devoid of gaps are shown. Domain predictions and secondary structure annotations are reported. Consensus logo displayed below the alignment indicates the relative number of residues per column, which can be estimated by its size in the logo, and gaps in columns were included by default in this calculation. Alignment Quality quantitative annotation was automatically calculated and represents is an ad-hoc measure of the likelihood of observing the mutations (if any) in a particular column of the alignment. The quality score is inversely proportional to the average cost of all pairs of mutations observed in a particular column of the alignment. A high alignment quality score for a column suggests that there are no mutations, or most mutations observed are favourable. The quality score is calculated for each column in an alignment by summing, for all mutations, the ratio of the two BLOSUM62 scores for a mutation pair and each residue's conserved BLOSUM62 score (which is higher). This value is normalised for each column, and then plotted on a scale from 0 to 1. Alignment occupancy reflects the number of residues aligned at each column.

7.12 Structural models

Vid22 structure predictions were performed using the ColabFold (version 1.3.0) implementation of AlphaFold2-multimer (version 2) on a local server equipped with an NVIDIA A100 GPU with 80GB RAM. Vid22 homodimer was predicted using its full-length sequence. Five models were generated using 12 recycles. The top three predictions are within 2Å of root mean square deviation (excluding the disordered regions) from each other, with the two remainder predictions showing the same monomeric conformation but a different quaternary structure. The conformation of the monomer can be superimposed on the prediction deposited in the AlphaFold Protein Structure Database (C7GRW6). The Predicted Aligned Error (PAE) and the per-residue model confidence score (pLDDT) are reported.

8 List of Abbreviations

BM	Bleomycin sulfate
dNTP	deoxyribonucleotide
G4	G-quadruplex, G-tetraplex
HU	hydroxyurea
MSA	Multiple sequence alignment
TMPyP4	*meso*-5,10,15,20-Tetrakis-(N-methyl-4-pyridyl)porphine

9 Bibliography

1. Aguilera, A. & Gómez-González, B. Genome instability: a mechanistic view of its causes and consequences. *Nat. Rev. Genet.* **9**, 204–217 (2008).
2. Tubbs, A. & Nussenzweig, A. Endogenous DNA Damage as a Source of Genomic Instability in Cancer. *Cell* **168**, 644–656 (2017).
3. Johnston, M. O. Mutations and New Variation: Overview. in *Encyclopedia of Life Sciences* (Wiley, 2003). doi:10.1038/npg.els.0001723.
4. Gravel, S. & Jackson, S. P. Increased Genome Instability in Aging Yeast. *Cell* **115**, 1–2 (2003).
5. McMurray, M. A. & Gottschling, D. E. Aging and genetic instability in yeast. *Curr. Opin. Microbiol.* **7**, 673–679 (2004).
6. Lemoine, M. The Evolution of the Hallmarks of Aging. *Front. Genet.* **12**, (2021).
7. Lengauer, C., Kinzler, K. W. & Vogelstein, B. Genetic instabilities in human cancers. *Nature* **396**, 643–649 (1998).
8. Greenman, C. *et al.* Patterns of somatic mutation in human cancer genomes. *Nature* **446**, 153–158 (2007).
9. Hanahan, D. & Weinberg, R. A. A. Hallmarks of Cancer: The Next Generation. *Cell* **144**, 646–674 (2011).
10. Vogelstein, B. *et al.* Cancer genome landscapes. *Science (80-.).* **340**, 1546–1558 (2013).
11. Pikor, L., Thu, K., Vucic, E. & Lam, W. The detection and implication of genome instability in cancer. *Cancer Metastasis Rev.* **32**, 341–352 (2013).
12. Vessey, C. J., Norbury, C. J. & Hickson, I. D. Genetic disorders associated with cancer predisposition and genomic instability. *Progress in nucleic acid research and molecular biology* vol. 63 189–221 (1999).
13. Ciccia, A. & Elledge, S. J. The DNA Damage Response: Making It Safe to Play with Knives. *Mol. Cell* **40**, 179–204 (2010).
14. Terabayashi, T. & Hanada, K. Genome instability syndromes caused by impaired DNA repair and aberrant DNA damage responses. *Cell Biol. Toxicol.* **34**, 337–350 (2018).
15. Sieber, O. M., Heinimann, K. & Tomlinson, I. P. M. Genomic instability - The engine of tumorigenesis? *Nat. Rev. Cancer* **3**, 701–708 (2003).
16. Negrini, S., Gorgoulis, V. G. & Halazonetis, T. D. Genomic instability — an evolving hallmark of cancer. *Nat. Rev. Mol. Cell Biol.* **11**, 220–228 (2010).
17. Schmitt, M. W., Prindle, M. J. & Loeb, L. A. Implications of genetic heterogeneity in

cancer. *Ann. N. Y. Acad. Sci.* **1267**, 110–116 (2012).

18. Gaillard, H., García-Muse, T. & Aguilera, A. Replication stress and cancer. *Nat. Rev. Cancer* **15**, 276–289 (2015).
19. Yuen, K. W. Y. *et al.* Systematic genome instability screens in yeast and their potential relevance to cancer. *Proc. Natl. Acad. Sci.* **104**, 3925–3930 (2007).
20. Pilié, P. G., Tang, C., Mills, G. B. & Yap, T. A. State-of-the-art strategies for targeting the DNA damage response in cancer. *Nat. Rev. Clin. Oncol.* 1 (2018) doi:10.1038/s41571-018-0114-z.
21. Ames, B. N. Mutagenesis and carcinogenesis: Endogenous and exogenous factors. *Environ. Mol. Mutagen.* **14**, 66–77 (1989).
22. Chatterjee, N. & Walker, G. C. Mechanisms of DNA damage, repair, and mutagenesis. *Environ. Mol. Mutagen.* **58**, 235–263 (2017).
23. Aguilera, A. & García-Muse, T. Causes of genome instability. *Annu. Rev. Genet.* **47**, 1–32 (2013).
24. Lambert, S. & Carr, A. M. Replication stress and genome rearrangements: lessons from yeast models. *Curr. Opin. Genet. Dev.* **23**, 132–139 (2013).
25. Zheng, D.-Q., Zhang, K., Wu, X.-C., Mieczkowski, P. A. & Petes, T. D. Global analysis of genomic instability caused by DNA replication stress in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci.* **113**, E8114–E8121 (2016).
26. Maffia, A., Ranise, C. & Sabbioneda, S. From R-Loops to G-Quadruplexes: Emerging New Threats for the Replication Fork. *Int. J. Mol. Sci.* **21**, 1506 (2020).
27. Branzei, D. & Foiani, M. Maintaining genome stability at the replication fork. *Nat. Rev. Mol. Cell Biol.* **11**, 208–219 (2010).
28. Zeman, M. K. & Cimprich, K. A. Causes and consequences of replication stress. *Nat. Cell Biol.* **16**, 2–9 (2014).
29. Gadaleta, M. C. & Noguchi, E. Regulation of DNA Replication through Natural Impediments in the Eukaryotic Genome. *Genes (Basel).* **8**, 98 (2017).
30. Slater, M. L. Effect of reversible inhibition of deoxyribonucleic acid synthesis on the yeast cell cycle. *J. Bacteriol.* **113**, 263–270 (1973).
31. Singh, A. & Xu, Y.-J. J. The Cell Killing Mechanisms of Hydroxyurea. *Genes (Basel).* **7**, 99 (2016).
32. Musiałek, M. W., Rybaczek, D. & Davidson, T. M. Hydroxyurea—The Good, the Bad and the Ugly. *Arch. Otolaryngol. Neck Surg.* **123**, 18 (2021).
33. Lazzaro, F. *et al.* RNase H and postreplication repair protect cells from ribonucleotides incorporated in DNA. *Mol. Cell* **45**, 99–110 (2012).

34. Reijns, M. A. M. *et al.* Enzymatic removal of ribonucleotides from DNA is essential for mammalian genome integrity and development. *Cell* **149**, 1008–1022 (2012).
35. Pizzi, S. *et al.* Reduction of hRNase H2 activity in Aicardi–Goutières syndrome cells leads to replication stress and genome instability. *Hum. Mol. Genet.* **24**, 649–658 (2015).
36. Meroni, A. *et al.* The Incorporation of Ribonucleotides Induces Structural and Conformational Changes in DNA. *Biophys. J.* **113**, 1373–1382 (2017).
37. Mirkin, E. V. & Mirkin, S. M. Replication Fork Stalling at Natural Impediments. *Microbiol. Mol. Biol. Rev.* **71**, 13–35 (2007).
38. Hizume, K. & Araki, H. Replication fork pausing at protein barriers on chromosomes. *FEBS Lett.* **593**, 1449–1458 (2019).
39. Ruggiano, A. & Ramadan, K. DNA–protein crosslink proteases in genome stability. *Commun. Biol.* **4**, 11 (2021).
40. Brambati, A., Colosio, A., Zardoni, L., Galanti, L. & Liberi, G. Replication and transcription on a collision course: eukaryotic regulation mechanisms and implications for DNA stability. *Front. Genet.* **6**, 1–8 (2015).
41. Santos-Pereira, J. M. & Aguilera, A. R loops: New modulators of genome dynamics and function. *Nat. Rev. Genet.* **16**, 583–597 (2015).
42. Hamperl, S. & Cimprich, K. A. Conflict Resolution in the Genome: How Transcription and Replication Make It Work. *Cell* **167**, 1455–1467 (2016).
43. Hamperl, S., Bocek, M. J., Saldivar, J. C., Swigut, T. & Cimprich, K. A. Transcription-Replication Conflict Orientation Modulates R-Loop Levels and Activates Distinct DNA Damage Responses. *Cell* **170**, 774-786.e19 (2017).
44. Lin, Y.-L. L. & Pasero, P. Transcription-Replication Conflicts: Orientation Matters. *Cell* **170**, 603–604 (2017).
45. Allison, D. F. & Wang, G. G. R-loops: formation, function, and relevance to cell stress. *Cell Stress* **3**, 38–46 (2019).
46. Kemiha, S., Poli, J., Lin, Y.-L., Lengronne, A. & Pasero, P. Toxic R-loops: Cause or consequence of replication stress? *DNA Repair (Amst).* **107**, 103199 (2021).
47. St Germain, C., Zhao, H. & Barlow, J. H. Transcription-Replication Collisions—A Series of Unfortunate Events. *Biomolecules* **11**, 1249 (2021).
48. Branzei, D. & Foiani, M. Leaping forks at inverted repeats. *Genes Dev.* **24**, 5–9 (2010).
49. Khristich, A. N. & Mirkin, S. M. On the wrong DNA track: Molecular mechanisms of repeat-mediated genome instability. *J. Biol. Chem.* **295**, 4134–4170 (2020).
50. Zhao, J., Bacolla, A., Wang, G. & Vasquez, K. M. Non-B DNA structure-induced genetic

instability and evolution. *Cell. Mol. Life Sci.* **67**, 43–62 (2010).

51. Saini, N., Zhang, Y., Usdin, K. & Lobachev, K. S. When secondary comes first – The importance of non-canonical DNA structures. *Biochimie* **95**, 117–123 (2013).
52. Zell, J., Sperti, F. R., Britton, S. & Monchaud, D. DNA folds threaten genetic stability and can be leveraged for chemotherapy. *RSC Chem. Biol.* **2**, 47–76 (2021).
53. Duardo, R. C., Guerra, F., Pepe, S. & Capranico, G. Non-B DNA structures as a booster of genome instability. *Biochimie* **214**, 176–192 (2023).
54. Makova, K. D. & Weissensteiner, M. H. Noncanonical DNA structures are drivers of genome evolution. *Trends Genet.* **39**, 109–124 (2023).
55. Wang, G. & Vasquez, K. M. Dynamic alternative DNA structures in biology and disease. *Nat. Rev. Genet.* **24**, 211–234 (2023).
56. Mikolaskova, B. *et al.* Maintenance of genome stability: the unifying role of interconnections between the DNA damage response and RNA-processing pathways. *Curr. Genet.* **64**, 971–983 (2018).
57. Friedberg, E. C., McDaniel, L. D. & Schultz, R. A. The role of endogenous and exogenous DNA damage and mutagenesis. *Curr. Opin. Genet. Dev.* **14**, 5–10 (2004).
58. Giglia-Mari, G., Zotter, A. & Vermeulen, W. DNA Damage Response. *Cold Spring Harb. Perspect. Biol.* **3**, a000745–a000745 (2011).
59. Bi, X. Mechanism of DNA damage tolerance. *World J. Biol. Chem.* **6**, 48 (2015).
60. Ashour, M. E. & Mosammaparast, N. Mechanisms of damage tolerance and repair during DNA replication. *Nucleic Acids Res.* **49**, 3033–3047 (2021).
61. Ler, A. A. L. & Carty, M. P. DNA Damage Tolerance Pathways in Human Cells: A Potential Therapeutic Target. *Front. Oncol.* **11**, 822500 (2022).
62. Lopes, M. *et al.* The DNA replication checkpoint response stabilizes stalled replication forks. *Nature* **412**, 557–561 (2001).
63. Tercero, J. A. & Diffley, J. F. X. Regulation of DNA replication fork progression through damaged DNA by the Mec1/Rad53 checkpoint. *Nature* **412**, 553–557 (2001).
64. Cha, R. S. & Kleckner, N. ATR Homolog Mec1 Promotes Fork Progression, Thus Averting Breaks in Replication Slow Zones. *Science (80-.).* **297**, 602–606 (2002).
65. Puddu, F., Piergiovanni, G., Plevani, P. & Muzi-Falconi, M. Sensing of Replication Stress and Mec1 Activation Act through Two Independent Pathways Involving the 9-1-1 Complex and DNA Polymerase ε. *PLoS Genet.* **7**, e1002022 (2011).
66. Cortez, D. Preventing replication fork collapse to maintain genome integrity. *DNA Repair (Amst).* **32**, 149–157 (2015).
67. Berti, M. & Vindigni, A. Replication stress: Getting back on track. *Nat. Struct. Mol. Biol.*

23, 103–109 (2016).

68. Branzei, D. & Foiani, M. The Rad53 signal transduction pathway: Replication fork stabilization, DNA repair, and adaptation. *Exp. Cell Res.* **312**, 2654–2659 (2006).
69. Lazzaro, F. *et al.* Checkpoint mechanisms at the intersection between DNA damage and repair. *DNA Repair (Amst).* **8**, 1055–1067 (2009).
70. Lanz, M. C., Dibitetto, D. & Smolka, M. B. DNA damage kinase signaling: checkpoint and repair at 30 years. *EMBO J.* **38**, (2019).
71. Myung, K. & Kolodner, R. D. Suppression of genome instability by redundant S-phase checkpoint pathways in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci.* **99**, 4500–4507 (2002).
72. Kolodner, R. D., Putnam, C. D. & Myung, K. Maintenance of Genome Stability in *Saccharomyces cerevisiae*. *Science (80-.).* **297**, 552–557 (2002).
73. Harrison, J. C. & Haber, J. E. Surviving the Breakup: The DNA Damage Checkpoint. *Annu. Rev. Genet.* **40**, 209–235 (2006).
74. Putnam, C. D., Jaehnig, E. J. & Kolodner, R. D. Perspectives on the DNA damage and replication checkpoint responses in Saccharomyces cerevisiae. *DNA Repair (Amst).* **8**, 974–982 (2009).
75. Sogo, J. M., Lopes, M. & Foiani, M. Fork Reversal and ssDNA Accumulation at Stalled Replication Forks Owing to Checkpoint Defects. *Science (80-.).* **297**, 599–602 (2002).
76. van Kregten, M. & Tijsterman, M. The repair of G-quadruplex-induced DNA damage. *Exp. Cell Res.* **329**, 178–183 (2014).
77. Pavlova, A. V., Kubareva, E. A., Monakhova, M. V., Zvereva, M. I. & Dolinnaya, N. G. Impact of G-Quadruplexes on the Regulation of Genome Integrity, DNA Damage and Repair. *Biomolecules* **11**, 1284 (2021).
78. Dell'Oca, M. C. *et al.* Spotlight on G-Quadruplexes : From Structure and Modulation to Physiological and Pathological Roles. *Int. J. Mol. Sci.* **25**, 3162 (2024).
79. Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K. & Neidle, S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* **34**, 5402–5415 (2006).
80. Ma, Y., Iida, K. & Nagasawa, K. Topologies of G-quadruplex: Biological functions and regulation by ligands. *Biochem. Biophys. Res. Commun.* **531**, 3–17 (2020).
81. Lightfoot, H. L., Hagen, T., Tatum, N. J. & Hall, J. The diverse structural landscape of quadruplexes. *FEBS Lett.* **593**, 2083–2102 (2019).
82. Jana, J., Mohr, S., Vianney, Y. M. & Weisz, K. Structural motifs and intramolecular interactions in non-canonical G-quadruplexes. *RSC Chem. Biol.* **2**, 338–353 (2021).
83. Puig Lombardi, E. & Londoño-Vallejo, A. A guide to computational methods for G-

quadruplex prediction. *Nucleic Acids Res.* **48**, 1–15 (2020).

84. Bhattacharyya, D., Mirihana Arachchilage, G. & Basu, S. Metal Cations in G-Quadruplex Folding and Stability. *Front. Chem.* **4**, (2016).
85. Hazel, P., Huppert, J., Balasubramanian, S. & Neidle, S. Loop-Length-Dependent Folding of G-Quadruplexes. *J. Am. Chem. Soc.* **126**, 16405–16415 (2004).
86. Guédin, A., Gros, J., Alberti, P. & Mergny, J.-L. How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res.* **38**, 7858–7868 (2010).
87. Sattin, G. *et al.* Conformation and stability of intramolecular telomeric G-quadruplexes: Sequence effects in the loops. *PLoS One* **8**, e84113 (2013).
88. Tippana, R., Xiao, W. & Myong, S. G-quadruplex conformation and dynamics are determined by loop length and sequence. *Nucleic Acids Res.* **42**, 8106–8114 (2014).
89. Cheng, M. *et al.* Loop permutation affects the topology and stability of G-quadruplexes. *Nucleic Acids Res.* **46**, 9264–9275 (2018).
90. Dvorkin, S. A., Karsisiotis, A. I. & Webba Da Silva, M. Encoding canonical DNA quadruplex structure. *Sci. Adv.* **4**, eaat3007 (2018).
91. Phan, A. T., Kuryavyi, V., Burge, S., Neidle, S. & Patel, D. J. Structure of an Unprecedented G-Quadruplex Scaffold in the Human c-kit Promoter. *J. Am. Chem. Soc.* **129**, 4386–4392 (2007).
92. Dai, J., Carver, M., Punchihewa, C., Jones, R. A. & Yang, D. Structure of the Hybrid-2 type intramolecular human telomeric G-quadruplex in K+ solution: insights into structure polymorphism of the human telomeric sequence. *Nucleic Acids Res.* **35**, 4927–4940 (2007).
93. Greco, M. L. *et al.* Coexistence of two main folded G-quadruplexes within a single G-rich domain in the EGFRpromoter. *Nucleic Acids Res.* **45**, 10132–10142 (2017).
94. Marquevielle, J. *et al.* Structure of two G-quadruplexes in equilibrium in the KRAS promoter. *Nucleic Acids Res.* **48**, 9336–9345 (2020).
95. Grün, J. T. & Schwalbe, H. Folding dynamics of polymorphic G-quadruplex structures. *Biopolymers* **113**, 1–15 (2022).
96. Kolesnikova, S., Hubálek, M., Bednárová, L., Cvačka, J. & Curtis, E. A. Multimerization rules for G-quadruplexes. *Nucleic Acids Res.* **45**, 8684–8696 (2017).
97. Kolesnikova, S. & Curtis, E. A. Structure and Function of Multimeric G-Quadruplexes. *Molecules* **24**, 3074 (2019).
98. Kim, M.-Y., Gleason-Guzman, M., Izbicka, E., Nishioka, D. & Hurley, L. H. The different biological effects of telomestatin and TMPyP4 can be attributed to their selectivity for interaction with intramolecular or intermolecular G-quadruplex structures.

Cancer Res. **63**, 3247–3256 (2003).

99. Xu, Y., Kimura, T. & Komiyama, M. Human telomere RNA and DNA form an intermolecular G-quadruplex. *Nucleic Acids Symp. Ser.* **52**, 169–170 (2008).
100. Víglaský, V., Bauer, L. & Tlučková, K. Structural Features of Intra- and Intermolecular G-Quadruplexes Derived from Telomeric Repeats. *Biochemistry* **49**, 2110–2120 (2010).
101. Liano, D., Monti, L., Chowdhury, S., Raguseo, F. & Di Antonio, M. Long-range DNA interactions: inter-molecular G-quadruplexes and their potential biological relevance. *Chem. Commun.* **58**, 12753–12762 (2022).
102. Frasson, I., Pirota, V., Richter, S. N. & Doria, F. Multimeric G-quadruplexes: A review on their biological roles and targeting. *Int. J. Biol. Macromol.* **204**, 89–102 (2022).
103. Zheng, K. W. *et al.* Co-transcriptional formation of DNA:RNA hybrid G-quadruplex and potential function as constitutional cis element for transcription control. *Nucleic Acids Res.* **41**, 5533–5541 (2013).
104. Berlyoung, A. S. & Armitage, B. A. Assembly and Characterization of RNA/DNA Hetero-G-Quadruplexes. *Biochemistry* **59**, 4072–4080 (2020).
105. Chalikian, T. V., Liu, L., Macgregor, R. B. & Macgregor Robert B., J. Duplex-tetraplex equilibria in guanine- and cytosine-rich DNA. *Biophys. Chem.* **267**, 106473 (2020).
106. Maizels, N. Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nat. Struct. Mol. Biol.* **13**, 1055–1059 (2006).
107. Rhodes, D. & Lipps, H. J. G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.* **43**, 8627–8637 (2015).
108. Zhang, Z.-H. H., Qian, S. H., Wei, D. & Chen, Z.-X. X. In vivo dynamics and regulation of DNA G-quadruplex structures in mammals. *Cell Biosci.* **13**, 117 (2023).
109. Hardin, C. C., Perry, A. G. & White, K. Thermodynamic and kinetic characterization of the dissociation and assembly of quadruplex nucleic acids. *Biopolymers* **56**, 147–194 (2000).
110. Varshney, D., Spiegel, J., Zyner, K., Tannahill, D. & Balasubramanian, S. The regulation and functions of DNA and RNA G-quadruplexes. *Nat. Rev. Mol. Cell Biol.* **21**, 459–474 (2020).
111. Yuan, W. F. *et al.* The influencing factors and functions of DNA G-quadruplexes. *Cell Biochem. Funct.* **38**, 524–532 (2020).
112. Nishio, M., Tsukakoshi, K. & Ikebukuro, K. G-quadruplex: Flexible conformational changes by cations, pH, crowding and its applications to biosensing. *Biosens. Bioelectron.* **178**, 113030 (2021).

113. Jana, J. & Weisz, K. Thermodynamic Stability of G-Quadruplexes: Impact of Sequence and Environment. *ChemBioChem* **22**, 2848–2856 (2021).
114. Fry, M. Tetraplex DNA and its interacting proteins. *Front. Biosci.* **12**, 4336 (2007).
115. Kaushik, M., Kaushik, S., Bansal, A., Saxena, S. & Kukreti, S. Structural Diversity and Specific Recognition of Four Stranded G-Quadruplex DNA. *Curr. Mol. Med.* **11**, 744–769 (2011).
116. Brázda, V., Hároníková, L., Liao, J. C. C. & Fojta, M. DNA and RNA Quadruplex-Binding Proteins. *Int. J. Mol. Sci.* **15**, 17493–17517 (2014).
117. Moye, A. L. & Bryan, T. M. Proteins that interact with DNA and RNA G-quadruplexes. in *Biological Relevance & Therapeutic Applications of DNA- & RNA-Quadruplexes* 100–117 (Future Science Ltd, 2015). doi:10.4155/fseb12013.14.227.
118. Mishra, S. K., Tawani, A., Mishra, A. & Kumar, A. G4IPDB: A database for G-quadruplex structure forming nucleic acid interacting proteins. *Sci. Rep.* **6**, 38144 (2016).
119. Mendoza, O., Bourdoncle, A., Boulé, J.-B. B., Brosh, R. M. & Mergny, J.-L. L. G-quadruplexes and helicases. *Nucleic Acids Res.* **44**, 1989–2006 (2016).
120. McRae, E. K. S. S., Booy, E. P., Padilla-Meier, G. P. & McKenna, S. A. On characterizing the interactions between proteins and guanine quadruplex structures of nucleic acids. *J. Nucleic Acids* **2017**, 1–11 (2017).
121. Brázda, V. *et al.* The amino acid composition of quadruplex binding proteins reveals a shared motif and predicts new potential quadruplex interactors. *Molecules* **23**, 2341 (2018).
122. Sun, Z.-Y., Wang, X.-N., Cheng, S.-Q., Su, X.-X. & Ou, T.-M. Developing Novel G-Quadruplex Ligands: from Interaction with Nucleic Acids to Interfering with Nucleic Acid–Protein Interaction. *Molecules* **24**, 396 (2019).
123. Sengupta, A., Roy, S. S. & Chowdhury, S. Non-duplex G-Quadruplex DNA Structure: A Developing Story from Predicted Sequences to DNA Structure-Dependent Epigenetics and beyond. *Acc. Chem. Res.* **54**, 46–56 (2021).
124. Shu, H., Zhang, R., Xiao, K., Yang, J. & Sun, X. G-Quadruplex-Binding Proteins: Promising Targets for Drug Design. *Biomolecules* **12**, 648 (2022).
125. Meier-Stephenson, V. G4-quadruplex-binding proteins: review and insights into selectivity. *Biophys. Rev.* **14**, 635–654 (2022).
126. Vijay Kumar, M. J., Morales, R. & Tsvetkov, A. S. G-quadruplexes and associated proteins in aging and Alzheimer's disease. *Front. Aging* **4**, 1164057 (2023).
127. Dai, Y., Teng, X., Zhang, Q., Hou, H. & Li, J. Advances and challenges in identifying and characterizing G-quadruplex–protein interactions. *Trends Biochem. Sci.* **48**, 894–909

(2023).

128. Lavezzo, E. *et al.* G-quadruplex forming sequences in the genome of all known human viruses: A comprehensive guide. *PLOS Comput. Biol.* **14**, e1006675 (2018).
129. Li, Z. *et al.* G-quadruplexes in genomes of viruses infecting eukaryotes or prokaryotes are under different selection pressures from hosts. *J. Genet. Genomics* **49**, 20–29 (2022).
130. Bartas, M. *et al.* The Presence and Localization of G-Quadruplex Forming Sequences in the Domain of Bacteria. *Molecules* **24**, 1711 (2019).
131. Yadav, P. *et al.* G-quadruplex structures in bacteria: Biological relevance and potential as an antimicrobial target. *J. Bacteriol.* **203**, (2021).
132. Capra, J. A., Paeschke, K., Singh, M. & Zakian, V. A. G-quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in Saccharomyces cerevisiae. *PLoS Comput. Biol.* **6**, 9 (2010).
133. Yadav, V., Hemansi, Kim, N., Tuteja, N. & Yadav, P. G Quadruplex in Plants: A Ubiquitous Regulatory Element and Its Biological Relevance. *Front. Plant Sci.* **8**, (2017).
134. Griffin, B. D. & Bass, H. W. Review: Plant G-quadruplex (G4) motifs in DNA and RNA; abundant, intriguing sequences of unknown function. *Plant Sci.* **269**, 143–147 (2018).
135. Marsico, G. *et al.* Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Res.* **47**, 3862–3874 (2019).
136. Wu, F. *et al.* Genome-wide analysis of DNA G-quadruplex motifs across 37 species provides insights into G4 evolution. *Commun. Biol.* **4**, 98 (2021).
137. Eddy, J. & Maizels, N. Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.* **34**, 3887–3896 (2006).
138. Hershman, S. G. *et al.* Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in Saccharomyces cerevisiae. *Nucleic Acids Res.* **36**, 144–156 (2008).
139. Wong, H. M. & Huppert, J. L. Stable G-quadruplexes are found outside nucleosome-bound regions. *Mol. Biosyst.* **5**, 1713 (2009).
140. Chambers, V. S. *et al.* High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat. Biotechnol.* **33**, 877–881 (2015).
141. Hänsel-Hertsch, R. *et al.* G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.* **48**, 1267–1272 (2016).
142. Vannutelli, A., Perreault, J.-P. & Ouangraoua, A. G-quadruplex occurrence and

conservation: more than just a question of guanine–cytosine content. *NAR Genomics Bioinforma.* **4**, lqac010 (2022).

143. Falabella, M., Fernandez, R. J., Johnson, F. B. & Kaufman, B. A. Potential Roles for G-Quadruplexes in Mitochondria. *Curr. Med. Chem.* **26**, 2918–2932 (2018).
144. Falabella, M. *et al.* G-quadruplex dynamics contribute to regulation of mitochondrial gene expression. *Sci. Rep.* **9**, 5605 (2019).
145. Sarkar, S. & Nambiar, M. G-quadruplexes in the mitochondrial genome – a cause for instability. *FEBS J.* **289**, 117–120 (2022).
146. Sahayasheela, V. J., Yu, Z., Hidaka, T., Pandian, G. N. & Sugiyama, H. Mitochondria and G-quadruplex evolution: an intertwined relationship. *Trends Genet.* **39**, 15–30 (2023).
147. Bochman, M. L., Paeschke, K. & Zakian, V. A. DNA secondary structures: Stability and function of G-quadruplex structures. *Nat. Rev. Genet.* **13**, 770–780 (2012).
148. Teng, F.-Y. Y. *et al.* G-quadruplex DNA: a novel target for drug design. *Cell. Mol. Life Sci.* **78**, 6557–6583 (2021).
149. Rigo, R., Palumbo, M. & Sissi, C. G-quadruplexes in human promoters: A challenge for therapeutic applications. *Biochim. Biophys. Acta - Gen. Subj.* **1861**, 1399–1413 (2017).
150. Kim, N. The Interplay between G-quadruplex and Transcription. *Curr. Med. Chem.* **26**, 2898–2917 (2017).
151. Spiegel, J. *et al.* G-quadruplexes are transcription factor binding hubs in human chromatin. *Genome Biol.* **22**, 117 (2021).
152. Fang, S. *et al.* Decoding regulatory associations of G-quadruplex with epigenetic and transcriptomic functional components. *Front. Genet.* **13**, (2022).
153. Armas, P., David, A. & Calcaterra, N. B. Transcriptional control by G-quadruplexes: In vivo roles and perspectives for specific intervention. *Transcription* **8**, 21–25 (2017).
154. Guilbaud, G. *et al.* Local epigenetic reprogramming induced by G-quadruplex ligands. *Nat. Chem.* **9**, 1110–1117 (2017).
155. Varizhuk, A., Isaakova, E. & Pozmogova, G. DNA G-Quadruplexes (G4s) Modulate Epigenetic (Re)Programming and Chromatin Remodeling: Transient Genomic G4s Assist in the Establishment and Maintenance of Epigenetic Marks, While Persistent G4s May Erase Epigenetic Marks. *BioEssays* **41**, 1900091 (2019).
156. Oyoshi, T. & Masuzawa, T. Modulation of histone modifications and G-quadruplex structures by G-quadruplex-binding proteins. *Biochem. Biophys. Res. Commun.* **531**, 39–44 (2020).

157. Reina, C. & Cavalieri, V. Epigenetic Modulation of Chromatin States and Gene Expression by G-Quadruplex Structures. *Int. J. Mol. Sci.* **21**, 4172 (2020).
158. Besnard, E. *et al.* Unraveling cell type–specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat. Struct. Mol. Biol.* **19**, 837–844 (2012).
159. Hoshina, S. *et al.* Human Origin Recognition Complex Binds Preferentially to G-quadruplex-preferable RNA and Single-stranded DNA. *J. Biol. Chem.* **288**, 30161–30171 (2013).
160. Valton, A.-L. & Prioleau, M.-N. G-Quadruplexes in DNA Replication: A Problem or a Necessity? *Trends Genet.* **32**, 697–706 (2016).
161. Masai, H. & Tanaka, T. G-quadruplex DNA and RNA: Their roles in regulation of DNA replication and other biological functions. *Biochem. Biophys. Res. Commun.* **531**, 25–38 (2020).
162. Datta, A. *et al.* G-Quadruplex Assembly by Ribosomal DNA: Emerging Roles in Disease Pathogenesis and Cancer Biology. *Cytogenet. Genome Res.* **161**, 285–296 (2021).
163. Green, A. T. *et al.* Computational and Experimental Characterization of rDNA and rRNA G-Quadruplexes. *J. Phys. Chem. B* **126**, 609–619 (2022).
164. Smith, J. S. *et al.* Rudimentary G-quadruplex–based telomere capping in Saccharomyces cerevisiae. *Nat. Struct. Mol. Biol.* **18**, 478–485 (2011).
165. Bryan, T. M. G-Quadruplexes at Telomeres: Friend or Foe? *Molecules* **25**, 3686 (2020).
166. Oganesian, L. & Bryan, T. M. Physiological relevance of telomeric G-quadruplex formation: a potential drug target. *BioEssays* **29**, 155–165 (2007).
167. Hou, Y. *et al.* Integrative characterization of G-Quadruplexes in the three-dimensional chromatin structure. *Epigenetics* **14**, 894–911 (2019).
168. Li, L. *et al.* YY1 interacts with guanine quadruplexes to regulate DNA looping and gene expression. *Nat. Chem. Biol.* **17**, 161–168 (2021).
169. Robinson, J. *et al.* DNA G-quadruplex structures: More than simple roadblocks to transcription? *Nucleic Acids Res.* **49**, 8419–8431 (2021).
170. Tikhonova, P. *et al.* DNA G-Quadruplexes Contribute to CTCF Recruitment. *Int. J. Mol. Sci.* **22**, 7090 (2021).
171. Pavlova, I., Iudin, M., Surdina, A., Severov, V. & Varizhuk, A. G-Quadruplexes in Nuclear Biomolecular Condensates. *Genes (Basel).* **14**, 1076 (2023).
172. Linke, R., Limmer, M., Juranek, S., Heine, A. & Paeschke, K. The Relevance of G-Quadruplexes for DNA Repair. *Int. J. Mol. Sci.* **22**, 12599 (2021).
173. Kharel, P., Balaratnam, S., Beals, N. & Basu, S. The role of RNA G-quadruplexes in

human diseases and therapeutic strategies. *WIREs RNA* **11**, e1568 (2020).

174. Tassinari, M., Richter, S. N. & Gandellini, P. Biological relevance and therapeutic potential of G-quadruplex structures in the human noncoding transcriptome. *Nucleic Acids Res.* **49**, 3617–3633 (2021).
175. Wang, S.-K., Wu, Y. & Ou, T.-M. RNA G-Quadruplex: The New Potential Targets for Therapy. *Curr. Top. Med. Chem.* **15**, 1947–1956 (2015).
176. Sahayasheela, V. J. & Sugiyama, H. RNA G-quadruplex in functional regulation of noncoding RNA: Challenges and emerging opportunities. *Cell Chem. Biol.* **31**, 53–70 (2024).
177. Tarsounas, M. & Tijsterman, M. Genomes and G-Quadruplexes: For Better or for Worse. *J. Mol. Biol.* **425**, 4782–4789 (2013).
178. Sato, K. & Knipscheer, P. G-quadruplex resolution: From molecular mechanisms to physiological relevance. *DNA Repair (Amst).* **130**, 103552 (2023).
179. Kruisselbrink, E. *et al.* Mutagenic Capacity of Endogenous G4 DNA Underlies Genome Instability in FANCJ-Defective C. elegans. *Curr. Biol.* **18**, 900–905 (2008).
180. Paeschke, K., Capra, J. A. & Zakian, V. A. DNA Replication through G-Quadruplex Motifs Is Promoted by the Saccharomyces cerevisiae Pif1 DNA Helicase. *Cell* **145**, 678–691 (2011).
181. Lemmens, B., Van Schendel, R. & Tijsterman, M. Mutagenic consequences of a single G-quadruplex demonstrate mitotic inheritance of DNA replication fork barriers. *Nat. Commun.* **6**, 8909 (2015).
182. Bryan, T. M. Mechanisms of DNA Replication and Repair: Insights from the Study of G-Quadruplexes. *Molecules* **24**, 3439 (2019).
183. Broxson, C., Beckett, J. & Tornaletti, S. Transcription Arrest by a G Quadruplex Forming-Trinucleotide Repeat Sequence from the Human c-myb Gene. *Biochemistry* **50**, 4162–4172 (2011).
184. Smestad, J. A. & Maher, L. J. Relationships between putative G-quadruplex-forming sequences, RecQ helicases, and transcription. *BMC Med. Genet.* **16**, 91 (2015).
185. Tang, W. *et al.* The Werner syndrome RECQ helicase targets G4 DNA in human cells to modulate transcription. *Hum. Mol. Genet.* **25**, 2060–2069 (2016).
186. Duquette, M. L., Handa, P., Vincent, J. A., Taylor, A. F. & Maizels, N. Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev.* **18**, 1618–1629 (2004).
187. De Magis, A. *et al.* DNA damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 816–825

(2019).

188. Miglietta, G., Russo, M. & Capranico, G. G-quadruplex-R-loop interactions and the mechanism of anticancer G-quadruplex binders. *Nucleic Acids Res.* **48**, 11942–11957 (2020).
189. Maizels, N. G4-associated human diseases. *EMBO Rep.* **16**, 910–922 (2015).
190. Wu, Y. & Brosh, R. M. G-quadruplex nucleic acids and human disease: G-quadruplex nucleic acids and human disease. *FEBS J.* **277**, 3470–3488 (2010).
191. Düchler, M. G-quadruplexes: targets and tools in anticancer drug design. *J. Drug Target.* **20**, 389–400 (2012).
192. Bidzinska, J., Cimino-Reale, G., Zaffaroni, N. & Folini, M. G-Quadruplex Structures in the Human Genome as Novel Therapeutic Targets. *Molecules* **18**, 12368–12395 (2013).
193. Simone, R., Fratta, P., Neidle, S., Parkinson, G. N. & Isaacs, A. M. G-quadruplexes: Emerging roles in neurodegenerative diseases and the non-coding transcriptome. *FEBS Lett.* **589**, 1653–1668 (2015).
194. Asamitsu, S., Obata, S., Yu, Z., Bando, T. & Sugiyama, H. Recent progress of targeted G-quadruplex-preferred ligands toward cancer therapy. *Molecules* **24**, 1–29 (2019).
195. Carvalho, J., Mergny, J.-L. L., Salgado, G. F., Queiroz, J. A. & Cruz, C. G-quadruplex, Friend or Foe: The Role of the G-quartet in Anticancer Strategies. *Trends Mol. Med.* **26**, 848–861 (2020).
196. Nakanishi, C. & Seimiya, H. G-quadruplex in cancer biology and drug discovery. *Biochem. Biophys. Res. Commun.* **531**, 45–50 (2020).
197. Sanchez-Martin, V., Lopez-Pujante, C., Soriano-Rodriguez, M. & Garcia-Salcedo, J. A. An updated focus on quadruplex structures as potential therapeutic targets in cancer. *Int. J. Mol. Sci.* **21**, 1–24 (2020).
198. Manas, K. & Hoeve, C. Guanine quadruplexes and disease: Insight into mechanisms and therapeutic potential. 11 (2020).
199. Antcliff, A., McCullough, L. D. & Tsvetkov, A. S. G-Quadruplexes and the DNA/RNA helicase DHX36 in health, disease, and aging. *Aging (Albany. NY).* **13**, 25578–25587 (2021).
200. Kosiol, N., Juranek, S., Brossart, P., Heine, A. & Paeschke, K. G-quadruplexes: a promising target for cancer therapy. *Mol. Cancer* **20**, 40 (2021).
201. Wang, E., Thombre, R., Shah, Y., Latanich, R. & Wang, J. G-Quadruplexes as pathogenic drivers in neurodegenerative disorders. *Nucleic Acids Res.* **49**, 4816–4830 (2021).
202. Xu, J., Huang, H. & Zhou, X. G-Quadruplexes in Neurobiology and Virology: Functional

Roles and Potential Therapeutic Approaches. *JACS Au* **1**, 2146–2161 (2021).

203. Balasubramanian, S., Hurley, L. H. & Neidle, S. Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat. Rev. Drug Discov.* **10**, 261–275 (2011).
204. Biffi, G., Tannahill, D., Miller, J., Howat, W. J. & Balasubramanian, S. Elevated Levels of G-Quadruplex Formation in Human Stomach and Liver Cancer Tissues. *PLoS One* **9**, e102711 (2014).
205. Zahler, A. M., Williamson, J. R., Cech, T. R. & Prescott, D. M. Inhibition of telomerase by G-quartet DMA structures. *Nature* **350**, 718–720 (1991).
206. Sun, D. *et al.* Inhibition of Human Telomerase by a G-Quadruplex-Interactive Compound. *J. Med. Chem.* **40**, 2113–2116 (1997).
207. Oganesian, L., Graham, M. E., Robinson, P. J. & Bryan, T. M. Telomerase Recognizes G-Quadruplex and Linear DNA as Distinct Substrates. *Biochemistry* **46**, 11279–11290 (2007).
208. Asamitsu, S., Shioda, N. & Sugiyama, H. Telomeric quadruplexes as therapeutic targets. in *Annual Reports in Medicinal Chemistry* vol. 54 77–99 (Elsevier, 2020).
209. Wang, Y.-H. *et al.* G4LDB 2.2: a database for discovering and studying G-quadruplex and i-Motif ligands. *Nucleic Acids Res.* **50**, D150–D160 (2022).
210. Li, Q. *et al.* G4LDB: a database for discovering and studying G-quadruplex ligands. *Nucleic Acids Res.* **41**, D1115–D1123 (2013).
211. Sanchez-Martin, V., Soriano, M. & Garcia-Salcedo, J. A. Quadruplex Ligands in Cancer Therapy. *Cancers (Basel).* **13**, 3156 (2021).
212. Awadasseid, A., Ma, X., Wu, Y. & Zhang, W. G-quadruplex stabilization via small-molecules as a potential anti-cancer strategy. *Biomed. Pharmacother.* **139**, 111550 (2021).
213. Izbicka, E. *et al.* Effects of cationic porphyrins as G-quadruplex interactive agents in human tumor cells. *Cancer Res.* **59**, 639–644 (1999).
214. Han, F. X., Wheelhouse, R. T. & Hurley, L. H. Interactions of TMPyP4 and TMPyP2 with Quadruplex DNA. Structural Basis for the Differential Effects on Telomerase Inhibition. *J. Am. Chem. Soc.* **121**, 3561–3570 (1999).
215. Grand, C. L. *et al.* The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth in vivo. *Mol. Cancer Ther.* **1**, 565–573 (2002).
216. Siddiqui-Jain, A., Grand, C. L., Bearss, D. J. & Hurley, L. H. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c- MYC transcription. *Proc. Natl. Acad. Sci.* **99**, 11593–11598 (2002).

217. Dexheimer, T. S., Sun, D. & Hurley, L. H. Deconvoluting the Structural and Drug-Recognition Complexity of the G-Quadruplex-Forming Region Upstream of the bcl-2 P1 Promoter. *J. Am. Chem. Soc.* **128**, 5404–5415 (2006).
218. Yamashita, T., Uno, T. & Ishikawa, Y. Stabilization of guanine quadruplex DNA by the binding of porphyrins with cationic side arms. *Bioorg. Med. Chem.* **13**, 2423–2430 (2005).
219. Wei, C., Jia, G., Yuan, J., Feng, Z. & Li, C. A Spectroscopic Study on the Interactions of Porphyrin with G-Quadruplex DNAs. *Biochemistry* **45**, 6681–6691 (2006).
220. Mikami-Terao, Y. *et al.* Antitumor activity of G-quadruplex–interactive agent TMPyP4 in K562 leukemic cells. *Cancer Lett.* **261**, 226–234 (2008).
221. Mikami-Terao, Y. *et al.* Antitumor activity of TMPyP4 interacting G-quadruplex in retinoblastoma cell lines. *Exp. Eye Res.* **89**, 200–208 (2009).
222. Martino, L., Pagano, B., Fotticchia, I., Neidle, S. & Giancola, C. Shedding Light on the Interaction between TMPyP4 and Human Telomeric Quadruplexes. *J. Phys. Chem. B* **113**, 14779–14786 (2009).
223. Wei, C., Jia, G., Zhou, J., Han, G. & Li, C. Evidence for the binding mode of porphyrins to G-quadruplex DNA. *Phys. Chem. Chem. Phys.* **11**, 4025 (2009).
224. Wei, C. *et al.* The binding mode of porphyrins with cation side arms to (TG4T)4 G-quadruplex: Spectroscopic evidence. *Biophys. Chem.* **143**, 79–84 (2009).
225. Prakash, A., Kieken, F., Marky, L. A. & Borgstahl, G. E. O. O. Stabilization of a G-Quadruplex from Unfolding by Replication Protein A Using Potassium and the Porphyrin TMPyP4. *J. Nucleic Acids* **2011**, 1–13 (2011).
226. Andrew, E. J. *et al.* Pentose Phosphate Pathway Function Affects Tolerance to the G-Quadruplex Binder TMPyP4. *PLoS One* **8**, e66242 (2013).
227. Alniss, H., Zamiri, B., Khalaj, M., Pearson, C. E. & Macgregor, R. B. Thermodynamic and spectroscopic investigations of TMPyP4 association with guanine- and cytosine-rich DNA and RNA repeats of C9orf72. *Biochem. Biophys. Res. Commun.* **495**, 2410–2417 (2018).
228. Ren, J. & Chaires, J. B. Sequence and Structural Selectivity of Nucleic Acid Binding Ligands. *Biochemistry* **38**, 16067–16075 (1999).
229. Weisman-Shomer, P. The cationic porphyrin TMPyP4 destabilizes the tetraplex form of the fragile X syndrome expanded sequence d(CGG)n. *Nucleic Acids Res.* **31**, 3963–3970 (2003).
230. Monchaud, D., Allain, C. & Teulade-Fichou, M.-P. Development of a fluorescent intercalator displacement assay (G4-FID) for establishing quadruplex-DNA affinity

and selectivity of putative ligands. *Bioorg. Med. Chem. Lett.* **16**, 4842–4845 (2006).

231. Rog, O. & Cooper, J. P. Telomeres in Drag: dressing as DNA damage to engage telomerase. *Curr. Opin. Genet. Dev.* **18**, 212–220 (2008).
232. Morris, M. J., Wingate, K. L., Silwal, J., Leeper, T. C. & Basu, S. The porphyrin TmPyP4 unfolds the extremely stable G-quadruplex in MT3-MMP mRNA and alleviates its repressive effect to enhance translation in eukaryotic cells. *Nucleic Acids Res.* **40**, 4137–4145 (2012).
233. Zamiri, B., Reddy, K., Macgregor, R. B. & Pearson, C. E. TMPyP4 porphyrin distorts RNA G-quadruplex structures of the disease-associated r(GGGGCC)n repeat of the C9orf72 gene and blocks interaction of RNAbinding proteins. *J. Biol. Chem.* **289**, 4653–4659 (2014).
234. Putnam, C. D. *et al.* A genetic network that suppresses genome rearrangements in Saccharomyces cerevisiae and contains defects in cancers. *Nat. Commun.* **7**, 11256 (2016).
235. Bock, C. *et al.* High-content CRISPR screening. *Nat. Rev. Methods Prim.* **2**, 8 (2022).
236. Botstein, D., Chervitz, S. A. & Cherry, M. Yeast as a Model Organism. *Science (80-.).* **277**, 1259–1260 (1997).
237. Kolodner, R. D., Putnam, C. D. & Myung, K. Maintenance of genome stability in Saccharomyces cerevisiae. *Science (80-.).* **297**, 552–557 (2002).
238. Putnam, C. D., Pennaneach, V. & Kolodner, R. D. Saccharomyces cerevisiae as a Model System To Define the Chromosomal Instability Phenotype. *Mol. Cell. Biol.* **25**, 7226–7238 (2005).
239. Smith, M. G. & Snyder, M. Yeast as a Model for Human Disease. *Curr. Protoc. Hum. Genet.* **48**, 1–8 (2006).
240. Boone, C., Bussey, H. & Andrews, B. J. Exploring genetic interactions and networks with yeast. *Nat. Rev. Genet.* **8**, 437–449 (2007).
241. Petranovic, D. & Nielsen, J. Can yeast systems biology contribute to the understanding of human disease? *Trends Biotechnol.* **26**, 584–590 (2008).
242. Pereira, C. *et al.* New insights into cancer-related proteins provided by the yeast model. *FEBS J.* **279**, 697–712 (2012).
243. Putnam, C. D. & Kolodner, R. D. Pathways and Mechanisms that Prevent Genome Instability in *Saccharomyces cerevisiae*. *Genetics* **206**, 1187–1225 (2017).
244. Skrzypek, M. S. *et al.* Saccharomyces genome database informs human biology. *Nucleic Acids Res.* **46**, D736–D742 (2018).
245. Srivatsan, A., Putnam, C. D. & Kolodner, R. D. Analyzing Genome Rearrangements in

Saccharomyces cerevisiae. in *Genome Instability* (eds. Muzi-Falconi, M. & Brown, G. W.) vol. 1672 43–61 (Springer New York, 2018).

246. Kachroo, A. H., Vandeloo, M., Greco, B. M. & Abdullah, M. Humanized yeast to model human biology, disease and evolution. *Dis. Model. Mech.* **15**, dmm049309 (2022).
247. Duina, A. A., Miller, M. E. & Keeney, J. B. Budding Yeast for Budding Geneticists: A Primer on the Saccharomyces cerevisiae Model System. *Genetics* **197**, 33–48 (2014).
248. Johnson, J. E., Smith, J. S., Kozak, M. L. & Johnson, F. B. In vivo veritas: Using yeast to probe the biological functions of G-quadruplexes. *Biochimie* **90**, 1250–1263 (2008).
249. Hendry, J. A., Tan, G., Ou, J., Boone, C. & Brown, G. W. Leveraging DNA Damage Response Signaling to Identify Yeast Genes Controlling Genome Stability. *G3 Genes|Genomes|Genetics* **5**, 997–1006 (2015).
250. Ouspenski, I. I., Elledge, S. J. & Brinkley, B. R. New yeast genes important for chromosome integrity and segregation identified by dosage effects on genome stability. *Nucleic Acids Res.* **27**, 3001–3008 (1999).
251. Myung, K., Chen, C. & Kolodner, R. D. Multiple pathways cooperate in the suppression of genome instability in Saccharomyces cerevisiae. *Nature* **411**, 1073–1076 (2001).
252. Huang, M.-E., Rio, A.-G., Nicolas, A. & Kolodner, R. D. A genomewide screen in Saccharomyces cerevisiae for genes that suppress the accumulation of mutations. *Proc. Natl. Acad. Sci.* **100**, 11529–11534 (2003).
253. Huang, D. & Koshland, D. Chromosome integrity in Saccharomyces cerevisiae: The interplay of DNA replication initiation factors, elongation factors, and origins. *Genes Dev.* **17**, 1741–1754 (2003).
254. Askree, S. H. *et al.* A genome-wide screen for Saccharomyces cerevisiae deletion mutants that affect telomere length. *Proc. Natl. Acad. Sci.* **101**, 8658–8663 (2004).
255. Smith, S. *et al.* Mutator genes for suppression of gross chromosomal rearrangements identified by a genome wide screening in Saccharomyces cerevisiae. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 9039–9044 (2004).
256. Andersen, M. P., Nelson, Z. W., Hetrick, E. D. & Gottschling, D. E. A Genetic Screen for Increased Loss of Heterozygosity in *Saccharomyces cerevisiae*. *Genetics* **179**, 1179–1195 (2008).
257. Stirling, P. C. *et al.* The Complete Spectrum of Yeast Chromosome Instability Genes Identifies Candidate CIN Cancer Genes and Functional Roles for ASTRA Complex Components. *PLoS Genet.* **7**, e1002057 (2011).
258. Alver, B. *et al.* A Whole Genome Screen for Minisatellite Stability Genes in Stationary-Phase Yeast Cells. *G3 Genes|Genomes|Genetics* **3**, 741–756 (2013).

259. Puddu, F. *et al.* Genome architecture and stability in the Saccharomyces cerevisiae knockout collection. *Nature* **573**, 416–420 (2019).
260. Alvaro, D., Lisby, M. & Rothstein, R. Genome-wide analysis of Rad52 foci reveals diverse mechanisms impacting recombination. *PLoS Genet.* **3**, 2439–2449 (2007).
261. Cheng, E. *et al.* Genome Rearrangements Caused by Depletion of Essential DNA Replication Proteins in *Saccharomyces cerevisiae*. *Genetics* **192**, 147–160 (2012).
262. Styles, E. B. B. *et al.* Exploring Quantitative Yeast Phenomics with Single-Cell Analysis of DNA Damage Foci. *Cell Syst.* **3**, 264-277.e10 (2016).
263. Galati, E. *et al.* VID22 counteracts G-quadruplex-induced genome instability. *Nucleic Acids Res.* **49**, 12785–12804 (2021).
264. Rouse, J. & Jackson, S. P. LCD1: An essential gene involved in checkpoint control and regulation of the MEC1 signalling pathway in Saccharomyces cerevisiae. *EMBO J.* **19**, 5801–5812 (2000).
265. Paciotti, V., Clerici, M., Lucchini, G. & Longhese, M. P. The checkpoint protein Ddc2, functionally related to S. pombe Rad26, interacts with Mec1 and is regulated by Mec1-dependent phosphorylation in budding yeast. *Genes Dev.* **14**, 2046–2059 (2000).
266. Wakayama, T., Kondo, T., Ando, S., Matsumoto, K. & Sugimoto, K. Pie1, a Protein Interacting with Mec1, Controls Cell Growth and Checkpoint Responses in Saccharomyces cerevisiae. *Mol. Cell. Biol.* **21**, 755–764 (2001).
267. Clerici, M. *et al.* Hyperactivation of the yeast DNA damage checkpoint by TEL1 and DDC2 overexpression. *EMBO J.* **20**, 6485–6498 (2001).
268. Bellaoui, M. *et al.* Elg1 forms an alternative RFC complex important for DNA replication and genome integrity. *EMBO J.* **22**, 4304–4313 (2003).
269. Haber, J. E. & Heyer, W. D. The fuss about Mus81. *Cell* vol. 107 551–554 (2001).
270. Kaliraman, V., Mullen, J. R., Fricke, W. M., Bastin-Shanower, S. A. & Brill, S. J. Functional overlap between Sgs1-Top3 and the Mms4-Mus81 endonuclease. *Genes Dev.* **15**, 2730–2740 (2001).
271. Mullen, J. R., Kaliraman, V., Ibrahim, S. S. & Brill, S. J. Requirement for three novel protein complexes in the absence of the Sgs1 DNA helicase in Saccharomyces cerevisiae. *Genetics* **157**, 103–118 (2001).
272. Fabre, F., Chan, A., Heyer, W. D. & Gangloff, S. Alternate pathways involving Sgs1/Top3, Mus81/ Mms4, and Srs2 prevent formation of toxic recombination intermediates from single-stranded gaps created by DNA replication. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 16887–16892 (2002).

273. Raderschall, E., Golub, E. I. & Haaf, T. Nuclear foci of mammalian recombination proteins are located at single-stranded DNA regions formed after DNA damage. *Proc. Natl. Acad. Sci. U. S. A.* **96**, 1921–1926 (1999).
274. Lisby, M., Mortensen, U. H. & Rothstein, R. Colocalization of multiple DNA double-strand breaks at a single Rad52 repair center. *Nat. Cell Biol.* **5**, 572–577 (2003).
275. Gupta, S. V. & Schmidt, K. H. Maintenance of yeast genome integrity by recq family dna helicases. *Genes (Basel).* **11**, 205 (2020).
276. Cobb, J. A., Bjergbaek, L., Shimada, K., Frei, C. & Gasser, S. M. DNA polymerase stabilization at stalled replication forks requires Mec1 and the RecQ helicase Sgs1. *EMBO J.* **22**, 4325–4336 (2003).
277. Cobb, J. A. *et al.* Replisome instability, fork collapse, and gross chromosomal rearrangements arise synergistically from Mec1 kinase and RecQ helicase mutations. *Genes Dev.* **19**, 3055–3069 (2005).
278. Calzada, A., Hodgson, B., Kanemaki, M., Bueno, A. & Labib, K. Molecular anatomy and regulation of a stable replisome at a paused eukaryotic DNA replication fork. *Genes Dev.* **19**, 1905–1919 (2005).
279. Liberi, G. *et al.* Rad51-dependent DNA structures accumulate at damaged replication forks in sgs1 mutants defective in the yeast ortholog of BLM RecQ helicase. *Genes Dev.* **19**, 339–350 (2005).
280. Heyer, W. D., Ehmsen, K. T. & Liu, J. Regulation of homologous recombination in eukaryotes. *Annu. Rev. Genet.* **44**, 113–139 (2010).
281. Heyer, W.-D. Regulation of Recombination and Genomic Maintenance. *Cold Spring Harb. Perspect. Biol.* **7**, a016501 (2015).
282. Sun, H., Bennett, R. J. & Maizels, N. The Saccharomyces cerevisiae Sgs1 helicase efficiently unwinds G-G paired DNAs. *Nucleic Acids Res.* **27**, 1978–1984 (1999).
283. Huber, M. D., Lee, D. C. & Maizels, N. G4 DNA unwinding by BLM and Sgs1p: Substrate specificity and substrate-specific inhibition. *Nucleic Acids Res.* **30**, 3954–3961 (2002).
284. Oughtred, R. *et al.* The BIOGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci.* **30**, 187–200 (2021).
285. Shen, X.-X. X. *et al.* Tempo and Mode of Genome Evolution in the Budding Yeast Subphylum. *Cell* **175**, 1533-1545.e20 (2018).
286. Krogan, N. J. *et al.* Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. *Nature* **440**, 637–643 (2006).
287. Gavin, A.-C. C. *et al.* Functional organization of the yeast proteome by systematic

analysis of protein complexes. *Nature* **415**, 141–147 (2002).

288. Bandyopadhyay, S., Kelley, R., Krogan, N. J. & Ideker, T. Functional Maps of Protein Complexes from Quantitative Genetic Interaction Data. *PLoS Comput. Biol.* **4**, e1000065 (2008).
289. Benschop, J. J. *et al.* A Consensus of Core Protein Complex Compositions for Saccharomyces cerevisiae. *Mol. Cell* **38**, 916–928 (2010).
290. Babu, M. *et al.* Interaction landscape of membrane-protein complexes in Saccharomyces cerevisiae. *Nature* **489**, 585–589 (2012).
291. Michaelis, A. C. *et al.* The social and structural architecture of the yeast protein interactome. *Nature* **624**, 192–200 (2023).
292. Wilson, W. A., Wang, Z. & Roach, P. J. Systematic Identification of the Genes Affecting Glycogen Storage in the Yeast Saccharomyces cerevisiae. *Mol. Cell. Proteomics* **1**, 232–242 (2002).
293. Ricarte, F. *et al.* A Genome-Wide Immunodetection Screen in S. cerevisiae Uncovers Novel Genes Involved in Lysosomal Vacuole Function and Morphology. *PLoS One* **6**, e23696 (2011).
294. Brigati, C. *et al.* An Essential Yeast Gene Encoding a TTAGGG Repeat-Binding Protein. *Mol. Cell. Biol.* **13**, 1306–1314 (1993).
295. Bilaud, T. *et al.* The Telobox, a Myb-Related Telomeric DNA Binding Motif Found in Proteins from Yeast, Plants and Human. *Nucleic Acids Res.* **24**, 1294–1303 (1996).
296. Pitt, C. W., Valente, L. P., Rhodes, D. & Simonsson, T. Identification and Characterization of an Essential Telomeric Repeat Binding Factor in Fission Yeast. *J. Biol. Chem.* **283**, 2693–2701 (2008).
297. Liu, Z. P. & Tye, B. K. A yeast protein that binds to vertebrate telomeres and conserved yeast telomeric junctions. *Genes Dev.* **5**, 49–59 (1991).
298. Koering, C. E. *et al.* Identification of high affinity Tbf1p-binding sites within the budding yeast genome. *Nucleic Acids Res.* **28**, 2519–2526 (2000).
299. Fourel, G., Revardel, E., Koering, C. E. & Gilson, É. Cohabitation of insulators and silencing elements in yeast subtelomeric regions. *EMBO J.* **18**, 2522–2537 (1999).
300. Brevet, V. *et al.* The number of vertebrate repeats can be regulated at yeast telomeres by Rap1-independent mechanisms. *EMBO J.* **22**, 1697–1706 (2003).
301. Hediger, F., Berthiau, A.-S. S., Van Houwe, G., Gilson, E. & Gasser, S. M. Subtelomeric factors antagonize telomere anchoring and Tel1-independent telomere length regulation. *EMBO J.* **25**, 857–867 (2006).
302. Arnerić, M. & Lingner, J. Tel1 kinase and subtelomere-bound Tbf1 mediate

preferential elongation of short telomeres by telomerase in yeast. *EMBO Rep.* **8**, 1080–1085 (2007).

303. Fukunaga, K., Hirano, Y. & Sugimoto, K. Subtelomere-binding protein Tbf1 and telomere-binding protein Rap1 collaborate to inhibit localization of the Mre11 complex to DNA ends in budding yeast. *Mol. Biol. Cell* **23**, 347–359 (2012).
304. Preti, M. *et al.* The Telomere-Binding Protein Tbf1 Demarcates snoRNA Gene Promoters in Saccharomyces cerevisiae. *Mol. Cell* **38**, 614–620 (2010).
305. Bosio, M. C., Fermi, B. & Dieci, G. Transcriptional control of yeast ribosome biogenesis: A multifaceted role for general regulatory factors. *Transcription* **8**, 254–260 (2017).
306. Bosio, M. C. *et al.* Abf1 and other general regulatory factors control ribosome biogenesis gene expression in budding yeast. *Nucleic Acids Res.* **45**, 4493–4506 (2017).
307. Bratkovič, T., Božič, J. & Rogelj, B. Functional diversity of small nucleolar RNAs. *Nucleic Acids Res.* **48**, 1627–1651 (2020).
308. Huang, Z., Du, Y., Wen, J., Lu, B. & Zhao, Y. snoRNAs: functions and mechanisms in biological processes, and roles in tumor pathophysiology. *Cell Death Discov.* **8**, 259 (2022).
309. Yan, C., Chen, H. & Bai, L. Systematic Study of Nucleosome-Displacing Factors in Budding Yeast. *Mol. Cell* **71**, 294-305.e4 (2018).
310. Ribaud, V., Ribeyre, C., Damay, P. & Shore, D. DNA-end capping by the budding yeast transcription factor and subtelomeric binding protein Tbf1. *EMBO J.* **31**, 138–149 (2012).
311. Liu, G. *et al.* Gene Essentiality Is a Quantitative Property Linked to Cellular Evolvability. *Cell* **163**, 1388–1399 (2015).
312. Bonetti, D., Anbalagan, S., Lucchini, G., Clerici, M. & Longhese, M. P. Tbf1 and Vid22 promote resection and non-homologous end joining of DNA double-strand break ends. *EMBO J.* **32**, 275–289 (2013).
313. Sendra, K., Jakubowska, A. & Korona, R. Comparing the rate of growth and metabolic efficiency of yeast experiencing environmental stress or genetic damage. *Acta Biochim. Pol.* **60**, 665–669 (2013).
314. de Graaf, B., Clore, A. & McCullough, A. K. Cellular pathways for DNA repair and damage tolerance of formaldehyde-induced DNA-protein crosslinks. *DNA Repair (Amst).* **8**, 1207–1214 (2009).
315. Novarina, D. Mechanisms preserving genome integrity in Saccharomyces cerevisiae.

228 (2012) doi:air.unimi.it › bitstream › phd_unimi_R08175.

316. Rossi, M. J. *et al.* A high-resolution protein architecture of the budding yeast genome. *Nature* **592**, 309–314 (2021).
317. Celic, I. *et al.* The Sirtuins Hst3 and Hst4p Preserve Genome Integrity by Controlling Histone H3 Lysine 56 Deacetylation. *Curr. Biol.* **16**, 1280–1289 (2006).
318. Maas, N. L., Miller, K. M., DeFazio, L. G. & Toczyski, D. P. Cell Cycle and Checkpoint Regulation of Histone H3 K56 Acetylation by Hst3 and Hst4. *Mol. Cell* **23**, 109–119 (2006).
319. Schneider, J., Bajwa, P., Johnson, F. C., Bhaumik, S. R. & Shilatifard, A. Rtt109 Is Required for Proper H3K56 Acetylation. *J. Biol. Chem.* **281**, 37270–37274 (2006).
320. Driscoll, R., Hudson, A. & Jackson, S. P. Yeast Rtt109 Promotes Genome Stability by Acetylating Histone H3 on Lysine 56. *Science (80-.).* **315**, 649–652 (2007).
321. Han, J. *et al.* Rtt109 Acetylates Histone H3 Lysine 56 and Functions in DNA Replication. *Science (80-.).* **315**, 653–655 (2007).
322. Han, J., Zhou, H., Li, Z., Xu, R.-M. & Zhang, Z. Acetylation of Lysine 56 of Histone H3 Catalyzed by RTT109 and Regulated by ASF1 Is Required for Replisome Integrity. *J. Biol. Chem.* **282**, 28587–28596 (2007).
323. Selth, L. & Svejstrup, J. Q. Vps75, A New Yeast Member of the NAP Histone Chaperone Family. *J. Biol. Chem.* **282**, 12358–12362 (2007).
324. Tsubota, T. *et al.* Histone H3-K56 Acetylation Is Catalyzed by Histone Chaperone-Dependent Complexes. *Mol. Cell* **25**, 703–712 (2007).
325. Fillingham, J. *et al.* Chaperone Control of the Activity and Specificity of the Histone H3 Acetyltransferase Rtt109. *Mol. Cell. Biol.* **28**, 4342–4353 (2008).
326. Braunstein, M., Rose, A. B., Holmes, S. G., Allis, C. D. & Broach, J. R. Transcriptional silencing in yeast is associated with reduced nucleosome acetylation. *Genes Dev.* **7**, 592–604 (1993).
327. Thompson, J. S., Ling, X. & Grunstein, M. Histone H3 amino terminus is required for telomeric and silent mating locus repression in yeast. *Nature* **369**, 245–247 (1994).
328. Brachmann, C. B. *et al.* The SIR2 gene family, conserved from bacteria to humans, functions in silencing, cell cycle progression, and chromosome stability. *Genes Dev.* **9**, 2888–2902 (1995).
329. Gershon, L. & Kupiec, M. The Amazing Acrobat: Yeast's Histone H3K56 Juggles Several Important Roles While Maintaining Perfect Balance. *Genes (Basel).* **12**, 342 (2021).
330. Celic, I., Verreault, A. & Boeke, J. D. Histone H3 K56 Hyperacetylation Perturbs

Replisomes and Causes DNA Damage. *Genetics* **179**, 1769–1784 (2008).

331. Hachinohe, M., Hanaoka, F. & Masumoto, H. Hst3 and Hst4 histone deacetylases regulate replicative lifespan by preventing genome instability in Saccharomyces cerevisiae: Hst3 and Hst4 histone deacetylases regulate replicative lifespan. *Genes to Cells* **16**, 467–477 (2011).
332. Thaminy, S. *et al.* Hst3 Is Regulated by Mec1-dependent Proteolysis and Controls the S Phase Checkpoint and Sister Chromatid Cohesion by Deacetylating Histone H3 at Lysine 56. *J. Biol. Chem.* **282**, 37805–37814 (2007).
333. Collins, S. R. *et al.* Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* **446**, 806–810 (2007).
334. van Leeuwen, J., Pons, C., Boone, C. & Andrews, B. J. Mechanisms of suppression: The wiring of genetic resilience. *BioEssays* **39**, (2017).
335. Rundlett, S. E. *et al.* HDA1 and RPD3 are members of distinct yeast histone deacetylase complexes that regulate silencing and transcription. *Proc. Natl. Acad. Sci.* **93**, 14503–14508 (1996).
336. Aricthota, S., Rana, P. P. & Haldar, D. Histone acetylation dynamics in repair of DNA double-strand breaks. *Front. Genet.* **13**, 926577 (2022).
337. Kobayashi, T. & Sasaki, M. Ribosomal DNA stability is supported by many 'buffer genes'-introduction to the Yeast rDNA Stability Database. *FEMS Yeast Res.* **17**, (2017).
338. Mergny, J.-L. Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Res.* **33**, e138–e138 (2005).
339. Małgowska, M., Gudanis, D., Teubert, A., Dominiak, G. & Gdaniec, Z. How to study G-quadruplex structures. *BioTechnologia* **4**, 381–390 (2012).
340. Paysan-Lafosse, T. *et al.* InterPro in 2022. *Nucleic Acids Res.* **51**, D418–D427 (2023).
341. Aravind, L. The BED finger, a novel DNA-binding domain in chromatin-boundary-element-binding proteins and transposases. *Trends Biochem. Sci.* **25**, 421–423 (2000).
342. Klug, A. & Schwabe, J. W. Protein motifs 5. Zinc fingers. *FASEB J.* **9**, 597–604 (1995).
343. Iuchi, S. & Kuldell, N. *Zinc finger proteins: from atomic contact to cellular function*. *Molecular biology intelligence unit* vol. 28 (Landes Bioscience ; Kluwer Academic/Plenum Publishers, 2005).
344. Iuchi, S. & Kuldell, N. *Zinc finger proteins. European Journal of Cancer* vol. 28 (1992).
345. Wolfe, S. A., Nekludova, L. & Pabo, C. O. DNA recognition by Cys2His2 zinc finger proteins. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 183–212 (2000).
346. Munro, D., Ghersi, D. & Singh, M. Two critical positions in zinc finger domains are

heavily mutated in three human cancer types. *PLoS Comput. Biol.* **14**, e1006290 (2018).

347. Babu, M. M., Iyer, L. M., Balaji, S. & Aravind, L. The natural history of the WRKY-GCM1 zinc fingers and the relationship between transcription factors and transposons. *Nucleic Acids Res.* **34**, 6505–6520 (2006).
348. Mack, A. M. & Crawford, N. M. The arabidopsis TAG1 transposase has an N-terminal zinc finger DNA binding domain that recognizes distinct subterminal motifs. *Plant Cell* **13**, 2319–2331 (2001).
349. Chiruvella, K. K. *et al.* Biochemical Characterization of Kat1: A Domesticated hAT-Transposase that Induces DNA Hairpin Formation and MAT-Switching. *Sci. Rep.* **6**, 21671 (2016).
350. Lannes, L., Furman, C. M., Hickman, A. B. & Dyda, F. Zinc-finger BED domains drive the formation of the active Hermes transpososome by asymmetric DNA binding. *Nat. Commun.* **14**, 4470 (2023).
351. Tapia-Ramírez, J., Eggen, B. J. L., Peral-Rubio, M. J., Toledo-Aral, J. J. & Mandel, G. A single zinc finger motif in the silencing factor REST represses the neural-specific type II sodium channel promoter. *Proc. Natl. Acad. Sci. U. S. A.* **94**, 1177–1182 (1997).
352. Ha, I. *et al.* Multiple functional domains of human transcription factor IIB: distinct interactions with two general transcription factors and RNA polymerase II. *Genes Dev.* **7**, 1021–1032 (1993).
353. Geisberg, J. V, Lee, W. S., Berk, A. J. & Ricciardi, R. P. The zinc finger region of the adenovirus E1A transactivating domain complexes with the TATA box binding protein. *Proc. Natl. Acad. Sci.* **91**, 2488–2492 (1994).
354. Zhou, Q., Gedrich, R. W. & Engel, D. A. Transcriptional repression of the c-fos gene by YY1 is mediated by a direct interaction with ATF/CREB. *J. Virol.* **69**, 4323–4330 (1995).
355. Lee, J. S., Galvin, K. M. & Shi, Y. Evidence for physical interaction between the zinc-finger transcription factors YY1 and Sp1. *Proc. Natl. Acad. Sci.* **90**, 6145–6149 (1993).
356. Perkins, N. D., Agranoff, A. B., Pascal, E. & Nabel, G. J. An Interaction between the DNA-Binding Domains of RelA(p65) and Sp1 Mediates Human Immunodeficiency Virus Gene Activation. *Mol. Cell. Biol.* **14**, 6570–6583 (1994).
357. Merika, M. & Orkin, S. H. Functional Synergy and Physical Interactions of the Erythroid Transcription Factor GATA-1 with the Krüppel Family Proteins Sp1 and EKLF. *Mol. Cell. Biol.* **15**, 2437–2447 (1995).
358. Gregory, R. C. *et al.* Functional interaction of GATA1 with erythroid Krüppel-like

factor and Sp1 at defined erythroid promoters. *Blood* **87**, 1793–1801 (1996).

359. Atkinson, P. W. *hAT* Transposable Elements. *Microbiol. Spectr.* **3**, 3.4.05 (2015).
360. Nowotny, M. Retroviral integrase superfamily: the structural perspective. *EMBO Rep.* **10**, 144–151 (2009).
361. Majorek, K. A. *et al.* The RNase H-like superfamily: New members, comparative structural analysis and evolutionary classification. *Nucleic Acids Res.* **42**, 4160–4179 (2014).
362. Katayanagi, K. *et al.* Three-dimensional structure of ribonuclease H from E. coli. *Nature* **347**, 306–309 (1990).
363. Yang, W., Hendrickson, W. A., Crouch, R. J. & Satow, Y. Structure of Ribonuclease H Phased at 2 Å Resolution by MAD Analysis of the Selenomethionyl Protein. *Science (80-.).* **249**, 1398–1405 (1990).
364. Gough, J., Karplus, K., Hughey, R. & Chothia, C. Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J. Mol. Biol.* **313**, 903–919 (2001).
365. Pandurangan, A. P., Stahlhacke, J., Oates, M. E., Smithers, B. & Gough, J. The SUPERFAMILY 2.0 database: a significant proteome update and a new webserver. *Nucleic Acids Res.* **47**, D490–D494 (2019).
366. Perez, Z. N., Musingarimi, P., Craig, N. L., Dyda, F. & Hickman, A. B. Purification, crystallization and preliminary crystallographic analysis of the Hermes transposase. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **61**, 587–590 (2005).
367. Hickman, A. B. *et al.* Molecular architecture of a eukaryotic DNA transposase. *Nat. Struct. Mol. Biol.* **12**, 715–721 (2005).
368. Hickman, A. B. B. *et al.* Structural Basis of hAT Transposon End Recognition by Hermes, an Octameric DNA Transposase from Musca domestica. *Cell* **158**, 353–367 (2014).
369. Hickman, A. B. *et al.* Structural insights into the mechanism of double strand break formation by Hermes, a hAT family eukaryotic DNA transposase. *Nucleic Acids Res.* **46**, 10286–10301 (2018).
370. Yuan, Y. W. & Wessler, S. R. The catalytic domain of all eukaryotic cut-and-paste transposase superfamilies. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 7884–7889 (2011).
371. Hickman, A. B., Chandler, M. & Dyda, F. Integrating prokaryotes and eukaryotes: DNA transposases in light of structure. *Crit. Rev. Biochem. Mol. Biol.* **45**, 50–69 (2010).
372. Kojima, K. K. Structural and sequence diversity of eukaryotic transposable elements. *Genes Genet. Syst.* **94**, 233–252 (2019).

373. Rubin, E., Lithwick, G. & Levy, A. A. Structure and Evolution of the hAT Transposon Superfamily. *Genetics* **158**, 949–957 (2001).
374. Zhou, L. *et al.* Transposition of hAT elements links transposable elements and V(D)J recombination. *Nature* **432**, 995–1001 (2004).
375. Nesmelova, I. V. & Hackett, P. B. DDE transposases: Structural similarity and diversity. *Adv. Drug Deliv. Rev.* **62**, 1187–1195 (2010).
376. Arensburger, P. *et al.* Phylogenetic and functional characterization of the hAT transposon superfamily. *Genetics* **188**, 45–57 (2011).
377. Sarilar, V., Bleykasten-Grosshans, C. & Neuveglise, C. Evolutionary dynamics of hAT DNA transposon families in saccharomycetaceae. *Genome Biol. Evol.* **7**, 172–190 (2014).
378. Volff, J.-N. Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. *BioEssays* **28**, 913–922 (2006).
379. Feschotte, C. & Pritham, E. J. DNA Transposons and the Evolution of Eukaryotic Genomes. *Annu. Rev. Genet.* **41**, 331–368 (2007).
380. Markljung, E. *et al.* ZBED6, a Novel Transcription Factor Derived from a Domesticated DNA Transposon Regulates IGF2 Expression and Muscle Growth. *PLoS Biol.* **7**, e1000256 (2009).
381. Sinzelle, L., Izsvák, Z. & Ivics, Z. Molecular domestication of transposable elements: From detrimental parasites to useful host genes. *Cell. Mol. Life Sci.* **66**, 1073–1093 (2009).
382. Volff, J. Tame affairs: domesticated transposase and domestic pigs. *EMBO Rep.* **11**, 241–242 (2010).
383. Alzohairy, A. M., Gyulai, G., Jansen, R. K. & Bahieldin, A. Transposable elements domesticated and neofunctionalized by eukaryotic genomes. *Plasmid* **69**, 1–15 (2013).
384. Hayward, A., Ghazal, A., Andersson, G., Andersson, L. & Jern, P. ZBED Evolution: Repeated Utilization of DNA Transposons as Regulators of Diverse Host Functions. *PLoS One* **8**, e59940 (2013).
385. Riordan, J. D. & Dupuy, A. J. Domesticated transposable element gene products in human cancer. *Mob. Genet. Elements* **3**, e26693 (2013).
386. Rajaei, N., Chiruvella, K. K., Lin, F. & Åström, S. U. Domesticated transposase Kat1 and its fossil imprints induce sexual differentiation in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 15491–15496 (2014).
387. Makałowski, W., Kischka, T. & Makałowska, I. Contribution of Transposable Elements

to Human Proteins. in *Encyclopedia of Life Sciences* 1–10 (Wiley, 2017). doi:10.1002/9780470015902.a0020793.pub2.

388. Bischerour, J. *et al.* Six domesticated piggybac transposases together carry out programmed DNA elimination in paramecium. *Elife* **7**, e37927 (2018).
389. Somerville, T. D. D. D. *et al.* ZBED2 is an antagonist of interferon regulatory factor 1 and modifies cell identity in pancreatic cancer. *Proc. Natl. Acad. Sci.* **117**, 11471–11482 (2020).
390. Capy, P. Taming, domestication and exaptation: Trajectories of transposable elements in genomes. *Cells* **10**, 3590 (2021).
391. Yellan, I., Yang, A. W. H. H. & Hughes, T. R. Diverse Eukaryotic CGG-Binding Proteins Produced by Independent Domestications of hAT Transposons. *Mol. Biol. Evol.* **38**, 2070–2075 (2021).
392. Almeida, M. V., Vernaz, G., Putman, A. L. K. K. & Miska, E. A. Taming transposable elements in vertebrates: from epigenetic silencing to domestication. *Trends Genet.* **38**, 529–553 (2022).
393. Yamashita, D. *et al.* Human DNA Replication-related Element Binding Factor (hDREF) Self-association via hATC Domain Is Necessary for Its Nuclear Accumulation and DNA Binding. *J. Biol. Chem.* **282**, 7563–7575 (2007).
394. Calvi, B. R., Hong, T. J., Findley, S. D. & Gelbart, W. M. Evidence for a common evolutionary origin of inverted repeat transposons in Drosophila and plants: hobo, activator, and Tam3. *Cell* **66**, 465–471 (1991).
395. Feldmar, S. & Kunze, R. The ORFa protein, the putative transposase of maize transposable element Ac, has a basic DNA binding domain. *EMBO J.* **10**, 4003–4010 (1991).
396. Essers, L., Adolphs, R. H. & Kunze, R. A Highly Conserved Domain of the Maize Activator Transposase Is Involved in Dimerization. *Plant Cell* **12**, 13 (2000).
397. Michel, K., O'Brochta, D. A., Atkinson, P. W., O'Brochta, D. A. & Atkinson, P. W. The C-terminus of the Hermes transposase contains a protein multimerization domain. *Insect Biochem. Mol. Biol.* **33**, 959–970 (2003).
398. Mukherjee, K. & Moroz, L. L. Transposon-derived transcription factors across metazoans. *Front. Cell Dev. Biol.* **11**, (2023).
399. Ohshima, N., Takahashi, M. & Hirose, F. Identification of a Human Homologue of the DREF Transcription Factor with a Potential Role in Regulation of the Histone H1 Gene. *J. Biol. Chem.* **278**, 22928–22938 (2003).
400. Yamashita, D. *et al.* hDREF Regulates Cell Proliferation and Expression of Ribosomal

Protein Genes. *Mol. Cell. Biol.* **27**, 2003–2013 (2007).

401. Jin, Y. *et al.* ZBED1/DREF: A transcription factor that regulates cell proliferation (Review). *Oncol. Lett.* **20**, 1 (2020).
402. Saghizadeh, M. *et al.* ZBED4, a BED-Type Zinc-Finger Protein in the Cones of the Human Retina. *Investig. Opthalmology Vis. Sci.* **50**, 3580 (2009).
403. Farber, D. B., Theendakara, V. P., Akhmedov, N. B. & Saghizadeh, M. ZBED4, A Novel Retinal Protein Expressed in Cones and Müller Cells. in *Retinal Degenerative Diseases* (eds. Anderson, R. E., Hollyfield, J. G. & LaVail, M. M.) vol. 664 79–87 (Springer New York, 2010).
404. Mokhonov, V. V, Theendakara, V. P., Gribanova, Y. E., Ahmedli, N. B. & Farber, D. B. Sequence-Specific Binding of Recombinant Zbed4 to DNA: Insights into Zbed4 Participation in Gene Transcription and Its Association with Other Proteins. *PLoS One* **7**, e35317 (2012).
405. Thukral, S. K., Morrison, M. L. & Young, E. T. Alanine scanning site-directed mutagenesis of the zinc fingers of transcription factor ADR1: Residues that contact DNA and that transactivate. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 9188–9192 (1991).
406. Tapia-Ramirez, J., Eggen, B. J. L., Peral-Rubio, M. J., Toledo-Aral, J. J. & Mandel, G. A single zinc finger motif in the silencing factor REST represses the neural-specific type II sodium channel promoter. *Proc. Natl. Acad. Sci.* **94**, 1177–1182 (1997).
407. Cook, W. J. *et al.* Mutations in the zinc-finger region of the yeast regulatory protein ADR1 affect both DNA binding and transcriptional activation. *J. Biol. Chem.* **269**, 9374–9379 (1994).
408. Oughtred, R. *et al.* The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci.* **30**, 187–200 (2021).
409. Badis, G. *et al.* A Library of Yeast Transcription Factor Motifs Reveals a Widespread Function for Rsc3 in Targeting Nucleosome Exclusion at Promoters. *Mol. Cell* **32**, 878–887 (2008).
410. Geertz, M., Shore, D. & Maerkl, S. J. Massively parallel measurements of molecular interaction kinetics on a microfluidic platform. *Proc. Natl. Acad. Sci.* **109**, 16540–16545 (2012).
411. Traczyk, A., Liew, C. W., Gill, D. J. & Rhodes, D. Structural basis of G-quadruplex DNA recognition by the yeast telomeric protein Rap1. *Nucleic Acids Res.* **48**, 4562–4571 (2020).
412. Storici, F., Lewis, L. K. & Resnick, M. A. In vivo site-directed mutagenesis using oligonucleotides. *Nat. Biotechnol.* **19**, 773–776 (2001).

413. Stuckey, S. & Storici, F. Gene Knockouts, in vivo Site-Directed Mutagenesis and Other Modifications Using the Delitto Perfetto System in Saccharomyces cerevisiae. in *Methods in Enzymology* vol. 533 103–131 (Elsevier, 2013).
414. Rhee, H. S. & Pugh, B. F. Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* **483**, 295–301 (2012).
415. Haase, S. B. Cell Cycle Analysis of Budding Yeast Using SYTOX Green. *Curr. Protoc. Cytom.* **26**, (2003).
416. Calvi, B. R., Hong, T. J., Findley, S. D. & Gelbart, W. M. Evidence for a common evolutionary origin of inverted repeat transposons in Drosophila and plants: hobo, activator, and Tam3. *Cell* **66**, 465–471 (1991).
417. Hoberman, R., Klein-Seetharaman, J. & Rosenfeld, R. Inferring property selection pressure from positional residue conservation. *Appl. Bioinformatics* **3**, 167–179 (2004).
418. Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
419. Agilent Technologies. QuikChange II Site-Directed Mutagenesis Kit. *Europe* **200523**, 19 (2011).
420. Papadopoulou, C., Guilbaud, G., Schiavone, D. & Sale, J. E. E. Nucleotide Pool Depletion Induces G-Quadruplex-Dependent Perturbation of Gene Expression. *Cell Rep.* **13**, 2491–2503 (2015).
421. Chen, J. & Stubbe, J. Bleomycins: towards better therapeutics. *Nat. Rev. Cancer* **5**, 102–112 (2005).
422. Punjani, A., Rubinstein, J. L., Fleet, D. J. & Brubaker, M. A. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods* **14**, 290–296 (2017).
423. Van Criekinge, W. & Beyaert, R. Yeast two-hybrid: State of the art. *Biol. Proced. Online* **2**, 1–38 (1999).
424. Mceachern, M. J., Krauskopf, A. & Blackburn, E. H. Elomeres and. *Mol. Microbiol.* (2000).
425. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
426. Skolnick, J., Gao, M., Zhou, H. & Singh, S. AlphaFold 2: Why It Works and Its Implications for Understanding the Relationships of Protein Sequence, Structure, and Function. *J. Chem. Inf. Model.* **61**, 4827–4831 (2021).
427. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods* **19**,

679–682 (2022).

428. Mariani, V., Biasini, M., Barbato, A. & Schwede, T. lDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics* **29**, 2722–2728 (2013).
429. Vorlíčková, M. *et al.* Circular dichroism and guanine quadruplexes. *Methods* **57**, 64–75 (2012).
430. Sissi, C., Gatto, B. & Palumbo, M. The evolving world of protein-G-quadruplex recognition: A medicinal chemist's perspective. *Biochimie* **93**, 1219–1230 (2011).
431. Sun, Z. Y., Wang, X. N., Cheng, S. Q., Su, X. X. & Ou, T. M. Developing novel G-quadruplex ligands: From interaction with nucleic acids to interfering with nucleic acid–protein interaction. *Molecules* **24**, (2019).
432. Fourel, G., Revardel, E., Koering, C. E. & Gilson, É. Cohabitation of insulators and silencing elements in yeast subtelomeric regions. *EMBO J.* **18**, 2522–2537 (1999).
433. Charoensawan, V., Janga, S. C., Bulyk, M. L., Babu, M. M. & Teichmann, S. A. DNA Sequence Preferences of Transcriptional Activators Correlate More Strongly than Repressors with Nucleosomes. *Mol. Cell* **47**, 183–192 (2012).
434. Machné, R. & Murray, D. B. The yin and yang of yeast transcription: Elements of a global feedback system between metabolism and chromatin. *PLoS One* **7**, (2012).
435. Tsui, K. Defining nucleosome occupancy and positioning: Evolution and the role of trans-acting factors. 176 (2013).
436. van Bakel, H. *et al.* A Compendium of Nucleosome and Transcript Profiles Reveals Determinants of Chromatin Architecture and Transcription. *PLoS Genet.* **9**, e1003479 (2013).
437. Yan, C., Zhang, D., Raygoza Garay, J. A., Mwangi, M. M. & Bai, L. Decoupling of divergent gene regulation by sequence-specific DNA binding factors. *Nucleic Acids Res.* **43**, 7292–7305 (2015).
438. Hendrickson, D. G., Soifer, I., Wranik, B. J., Botstein, D. & Scott McIsaac, R. Simultaneous Profiling of DNA Accessibility and Gene Expression Dynamics with ATAC-Seq and RNA-Seq. in 317–333 (2018). doi:10.1007/978-1-4939-8618-7_15.
439. Radchenko, E. A. *et al.* Partners in crime: Tbf1 and Vid22 promote expansions of long human telomeric repeats at an interstitial chromosome position in yeast. *PNAS Nexus* **1**, pgac080 (2022).
440. Aksenova, A. Y. & Mirkin, S. M. At the Beginning of the End and in the Middle of the Beginning: Structure and Maintenance of Telomeric DNA Repeats and Interstitial Telomeric Sequences. *Genes (Basel).* **10**, 118 (2019).

441. Fourel, G. *et al.* An activation-independent role of transcription factors in insulator function. *EMBO Rep.* **2**, 124–132 (2001).
442. Berthiau, A. S. *et al.* Subtelomeric proteins negatively regulate telomere elongation in budding yeast. *EMBO J.* **25**, 846–856 (2006).
443. Taddei, A. & Gasser, S. M. Structure and Function in the Budding Yeast Nucleus. *Genetics* **192**, 107–129 (2012).
444. Duan, Z. *et al.* A three-dimensional model of the yeast genome. *Nature* **465**, 363–367 (2010).
445. Misteli, T. The Self-Organizing Genome: Principles of Genome Architecture and Function. *Cell* **183**, 28–45 (2020).
446. Szabo, Q., Bantignies, F. & Cavalli, G. Principles of genome folding into topologically associating domains. *Sci. Adv.* **5**, eaaw1668 (2019).
447. Mekhail, K. & Moazed, D. The nuclear envelope in genome organization, expression and stability. *Nat. Rev. Mol. Cell Biol.* **11**, 317–328 (2010).
448. Misteli, T. Beyond the Sequence: Cellular Organization of Genome Function. *Cell* **128**, 787–800 (2007).
449. Stanic, M. & Mekhail, K. Integration of DNA damage responses with dynamic spatial genome organization. *Trends Genet.* **38**, S0168952521002614 (2021).
450. Kim, S. & Shendure, J. Mechanisms of Interplay between Transcription Factors and the 3D Genome. *Mol. Cell* **76**, 306–319 (2019).
451. Brickner, D. G. *et al.* The Role of Transcription Factors and Nuclear Pore Proteins in Controlling the Spatial Organization of the Yeast Genome. *Dev. Cell* **49**, 936-947.e4 (2019).
452. Randise-Hinchliff, C. & Brickner, J. H. Transcription factors dynamically control the spatial organization of the yeast genome. *Nucleus* **7**, 369–374 (2016).
453. Gietz, R. D. & Woods, R. A. 4 Transformation of Yeast by the Lithium Acetate/Single-Stranded Carrier DNA/PEG Method. *Methods Microbiol.* **26**, 53–66 (1998).
454. Daniel Gietz, R. & Woods, R. A. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. in *Methods in Enzymology* vol. 350 87–96 (Elsevier, 2002).
455. Gietz, R. D. Yeast transformation by the LiAc/SS carrier DNA/PEG method. *Methods Mol. Biol.* **1163**, 33–44 (2014).
456. Bähler, J. *et al.* Heterologous modules for efficient and versatile PCR-based gene targeting in Schizosaccharomyces pombe. *Yeast* **14**, 943–951 (1998).
457. Longtine, M. S. *et al.* Additional modules for versatile and economical PCR-based gene

deletion and modification in Saccharomyces cerevisiae. *Yeast* **14**, 953–961 (1998).

458. Petracek, M. E. & Longtine, M. S. PCR-based engineering of yeast genome. in *Methods in Enzymology* vol. 350 445–469 (Elsevier, 2002).
459. Storici, F. & Resnick, M. A. The Delitto Perfetto Approach to In Vivo Site-Directed Mutagenesis and Chromosome Rearrangements with Synthetic Oligonucleotides in Yeast. *Methods Enzymol.* **409**, 329–345 (2006).
460. Ma, H., Kunes, S., Schatz, P. J. & Botstein, D. Plasmid construction by homologous recombination in yeast. *Gene* **58**, 201–216 (1987).
461. Ferrari, M. *et al.* Functional Interplay between the 53BP1-Ortholog Rad9 and the Mre11 Complex Regulates Resection, End-Tethering and Repair of a Double-Strand Break. *PLoS Genet.* **11**, (2015).
462. Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, 45e – 45 (2001).
463. Giannattasio, M. *et al.* DNA decay and limited Rad53 activation after liquid holding of UV-treated nucleotide excision repair deficient S. cerevisiae cells. *DNA Repair (Amst).* **3**, 1591–1599 (2004).
464. Zivanov, J. *et al.* New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, (2018).
465. Rohou, A. & Grigorieff, N. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* **192**, 216–221 (2015).
466. Gouy, M., Guindon, S. & Gascuel, O. Sea view version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221–224 (2010).
467. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).

10 Appendix

10.1 *VID22* counteracts G-quadruplex-induced genome instability

I contributed to this publication with main figures 1 F-G, supplementary figures S1 A-B, S4 B, S6 A-C, and to additional data which were only presented to the referees during the revision of the manuscript and are shown in this thesis in figures 1 A-F, 3 B, 4 C-D, 5 A and 6. I also participated in text revision and editing during the revision of the manuscript.

Published online 6 December 2021 *Nucleic Acids Research, 2021, Vol. 49, No. 22* **12785–12804**
https://doi.org/10.1093/nar/gkab1156

VID22 counteracts G-quadruplex-induced genome instability

Elena Galati[1,†], Maria C. Bosio[1,†], Daniele Novarina[1,†], Matteo Chiara[1,2], Giulia M. Bernini[1], Alessandro M. Mozzarelli[1], Maria L. García-Rubio[3], Belén Gómez-González[3], Andrés Aguilera[3], Thomas Carzaniga[4], Marco Todisco[4], Tommaso Bellini[4], Giulia M. Nava[1], Gianmaria Frigè[5], Sarah Sertic[1], David S. Horner[1,2], Anastasia Baryshnikova[6], Caterina Manzari[2], Anna M. D'Erchia[2,7], Graziano Pesole[2,7], Grant W. Brown[8], Marco Muzi-Falconi[1,*] and Federico Lazzaro[1,*]

[1]Department of Biosciences, Università degli Studi di Milano, Via Celoria 26, 20133 Milan, Italy, [2]Istituto di Biomembrane, Bioenergetica e Biotecnologie Molecolari, Consiglio Nazionale delle Ricerche, Bari, Italy, [3]Centro Andaluz de Biología Molecular y Medicina Regenerativa-CABIMER, Universidad de Sevilla, Seville, Spain, [4]Dipartimento di Biotecnologie Mediche e Medicina Traslazionale, Università degli Studi di Milano, via Vanvitelli 32, 20129 Milan, Italy, [5]Department of Experimental Oncology, IEO, European Institute of Oncology IRCCS, Via Adamello 16, 20139 Milan, Italy, [6]Department of Molecular Genetics and Donnelly Centre, University of Toronto, Toronto, Canada, [7]Dipartimento di Bioscienze, Biotecnologie e Biofarmaceutica, Università di Bari 'A. Moro', Bari, Italy and [8]Department of Biochemistry and Donnelly Centre, University of Toronto, Ontario M5S 3E1, Toronto, Canada

Received June 03, 2021; Revised October 19, 2021; Editorial Decision November 05, 2021; Accepted November 08, 2021

ABSTRACT

Genome instability is a condition characterized by the accumulation of genetic alterations and is a hallmark of cancer cells. To uncover new genes and cellular pathways affecting endogenous DNA damage and genome integrity, we exploited a Synthetic Genetic Array (SGA)-based screen in yeast. Among the positive genes, we identified *VID22*, reported to be involved in DNA double-strand break repair. *vid22Δ* cells exhibit increased levels of endogenous DNA damage, chronic DNA damage response activation and accumulate DNA aberrations in sequences displaying high probabilities of forming G-quadruplexes (G4-DNA). If not resolved, these DNA secondary structures can block the progression of both DNA and RNA polymerases and correlate with chromosome fragile sites. Vid22 binds to and protects DNA at G4-containing regions both *in vitro* and *in vivo*. Loss of *VID22* causes an increase in gross chromosomal rearrangement (GCR) events dependent on G-quadruplex forming sequences. Moreover, the absence of Vid22 causes defects in the correct maintenance of G4-DNA rich elements, such as telomeres and mtDNA, and hypersensitivity to the G4-stabilizing ligand TMPyP4. We thus propose that Vid22 is directly involved in genome integrity maintenance as a novel regulator of G4 metabolism.

INTRODUCTION

DNA molecules are intrinsically unstable (1) and are often damaged by exposure to a variety of endogenous and exogenous genotoxic agents (2). When not correctly recognized and repaired, lesions can impede DNA duplication and endanger faithful transmission of the genetic material to the progeny. To preserve genome integrity, eukaryotic cells possess evolutionarily conserved mechanisms that act by handling problems or errors arising during DNA replication, repairing DNA lesions, monitoring chromosome segregation and ensuring proper coordination of all these processes with cell cycle progression (3–7). Failure of any of these DNA integrity pathways leads to a condition known

*To whom correspondence should be addressed. Tel: +39 0250314827; Fax: +39 025031544; Email: federico.lazzaro@unimi.it
Correspondence may also be addressed to Marco Muzi-Falconi. Email: marco.muzifalconi@unimi.it
†The authors wish it to be known that, in their opinion, the first three authors should be regarded as Joint First Authors.
Present addresses:
Daniele Novarina, European Research Institute for the Biology of Ageing, University of Groningen, University Medical Center Groningen, Groningen, The Netherlands.
Alessandro M. Mozzarelli, Humanitas University, Department of Biomedical Sciences, Pieve Emanuele, MI, Italy.
Anastasia Baryshnikova, Calico Life Sciences LLC, South San Francisco, CA, USA.

as 'genomic instability', characterized by the accumulation of unrepaired DNA damage and genetic alterations ranging from point mutations, insertions/deletions of few nucleotides, or expansion/contraction of repeated sequences, to gross chromosomal rearrangements and aneuploidy (8). Genomic instability impairs cell viability and compelling experimental evidence obtained in recent years in human cells shows that it acts as a driving force during tumorigenesis (9–12).

In addition to genotoxic agents, genome integrity is also threatened by non-canonical nucleic acid secondary structures. Indeed, certain genomic sequences have the potential to form DNA secondary structures such as triplexes, cruciforms, hairpins, DNA:RNA hybrids and G-quadruplexes (G4-DNA) that may interfere with physiological DNA metabolism (13). G4-DNA is one of the most characterized alternatives to the classical double helical DNA conformation and can occur when a stretch of two or more guanines is repeated four times at short distance. Hoogsteen hydrogen bonding between four guanines and monovalent cations induces an extremely stable planar structure called a G-quartet (14,15). Two or more parallel G-quartets form the so called G-quadruplex conformation that can be either intra or intermolecular (involving two or more strands) and each type can display different configurations (16,17).

Single-stranded DNA (ssDNA) formation can spontaneously induce the generation of G-quadruplexes. For this reason, DNA transactions that expose ssDNA, such as replication, transcription and telomere metabolism, promote the folding of G4-DNA (18). If not promptly resolved, these structures can jeopardize DNA replication and can compromise telomere homeostasis. In these contexts, cells have evolved specialized helicases that recognize and unwind G4-DNA to preserve genetic stability. Several G4-helicases have been identified *in vitro* and *in vivo*, including the yeast Sgs1 and Pif1 proteins, and mammalian BLM, WRN, PIF1, and FANCJ (19–23). In addition to potentially detrimental effects, G4-DNA structures may play important regulatory roles. While G4-DNA may be an obstacle for the progression of DNA polymerases, they also appear to be active components of metazoan replication origins (reviewed in (24)). Moreover, about half of all human genes contain sequences prone to G4-DNA formation in their promoter regions (25,26), suggesting a role for these structures in regulating gene expression. In accordance, the promoter regions of KRAS and c-MYC genes can form G4-DNA and, if stabilized, these structures repress transcription (27,28).

The identification and analysis of the factors that contribute to preserving genome integrity is a major source of information for understanding tumorigenesis and defining new potential therapeutic targets. While many molecular mechanisms underlying DNA metabolism are now well understood, several regulatory aspects are yet unclear, hinting at the involvement of still unidentified players. Genetic screens in budding yeast represent a powerful strategy for the identification of such factors. From a Synthetic Genetic Array (SGA) screen to identify new genes required to prevent spontaneous DNA damage, we identified *VID22*, which encodes a non-essential protein that contributes to a heterotrimeric complex with its paralog Env11 and the essential Myb domain telomere-binding protein Tbf1 (29,30). Vid22 is a nuclear protein (31) containing two domains whose functions are not well described, a BED-type zinc finger domain and a RNaseH-like domain (32,33). Previous reports indicate its involvement in double strand break (DSB) repair, and a role with Tbf1 in modulating histone occupancy in proximity of DSB has been suggested (34,35). Moreover, Vid22 chromatin immunoprecipitates are significantly enriched in genes associated with predicted G4 regions (35), suggesting that Vid22 could interact with G4-DNA *in vivo*.

Here, we delineate a role for Vid22 in promoting the stability of G4-DNA regions *in vivo*. Whole genome sequencing of *vid22Δ* mutant strains allowed the characterization of the genomic loci and the types of mutations that accumulate spontaneously in the absence of Vid22. We observed frequent gross chromosomal rearrangements (GCRs) in the proximity of G4-DNA structures as well as alterations in telomeric regions and loss of mitochondrial DNA, all of which are genomic regions that present high densities of G4 motifs. Accordingly, loss of Vid22 increases cell sensitivity to the TMPyP4 (5,10,15,20-tetratkis-(*N*-methyl-4-pyridyl)-21,23-*H*-porphyrin) ligand that binds and stabilizes G4 structures (36). We find that purified Vid22 is able to bind directly DNA G4-forming sequences *in vitro*, and that Vid22 is enriched on chromatin at G4-DNA and suppresses chromosomal aberrations at G4-DNA structures. Together, our data indicate that Vid22 controls G4-DNA metabolism to maintain genome integrity.

MATERIALS AND METHODS

Yeast strains, growth conditions and plasmids

All the strains and plasmids used in this work are listed in Supplementary Table S1. All the yeast strains were derived from BY4741, BY4742 and BY4743 (37), except strains used for GCR assays that derive from the YPH500 background (38). Yeast strains were obtained by standard procedures of transformation and tetrad dissection. A one-step PCR approach was used to delete genes and to tag proteins at the C-terminus, as described in (39). The eleven unrelated *vid22Δ* strains used in the PFGE analysis were obtained by independent one-step deletions. For the analysis shown in Supplementary Figure S2B, wild-type and *vid22Δ* mutants were obtained by tetrad dissection of diploid YNOV344 (*VID22*/*vid22Δ*) and grown in YEPD medium in the same conditions so that cells would go through the same number of cell divisions. Plateau-phase cultures were diluted with fresh medium to allow re-growth into plateau phase three times. Finally, cells were plated on YEPD to collect isolated clones for the PFGE analysis.

In the yeast strain YFL2922, the G4 in position Chr VIII 512264–512325 (17) was mutated using the 'Delitto Perfetto' strategy (40). Briefly, the pCORE cassette was amplified by PCR using G4_VIII_DPfor (5′-ATA TAA TCA GGG CTT AAG TAA ACG CTT CGC TGT GAT TTC CGA GCT CGT TTT CGA CAC TGG-3′) and G4_VIII_DPrev (5′-ATT TAA GAA AAA CTT TTT TTT TTT TTT TTT CGG ATG ATA TTC CTT ACC ATT AAG TTG ATC-3′) oligos and transformed in YGE651.2. The 'Delitto Perfetto' cassette was then replaced by ds-DNA containing the desired mutations obtained as described below. First, we amplified genomic DNA frag-

ment 1 by PCR using oligos G4_VIII_Afor (5′-CCT CAT CTA TAT ATA ATC AG-3′) and G4_VIII_Brev (5′-CGC TCA CTT ATG GCA TCG CTA TTA TAG CAA AAC CCT GCT T-3′), and fragment 2 using oligo G4_VIII_Cfor (5′-AAC ACA AGT ATA AGC AGG GTT TTG CTA TAA TAG CGA TGC CAT AAG TGA GCG CAG GGC TCA-3′) and G4_VIII_Drev (5′-TTG ACA AAT GTT TCA GAT CC-3′). Then, the PCR fragments 1 and 2 were annealed and amplified using oligos G4_VIII_Afor and G4_VIII_Drev. The oligos Brev and Cfor contain the desired mutations. A similar strategy was used to mutagenized Tbf1-BS, using the following oligos: DP_VIII_NEW_for (5′-AAG TCG ATT AAA AGT AGG GCT AAC ACA AGT ATA AGC AGG GGA GCT CGT TTT CGA CAC TGG-3′), DP_VIII_NEW_rev (5′-TGA GCC CTG CCC TCA CTT ATG GCA TCC CTA TTA TAG CAA ATC CTT ACC ATT AAG TTG ATC-3′), TBF1_BS_mut_Bfor (5′-AAA GTC GAT TAA AAG CCG GGC TAA CAC AAG TAT AAG CAG GGT TTT GCT ATA ACC GGG ATG-3′), TBF1_BS_mut_Crev (5′-TGA GCC CTG CCC TCA CTT ATG GCA TCC CGG TTA TAG CAA AAC CCT GCT TAT ACT TGT GTT-3), Tbf1_BS_mut_Afor (5′-AAA AAA AAA AAA AAA AAG TTT TTC TTA AAT GCA TAG GTT TAA AGT CGA TTA AAA GTC GGG-3′), TBF1_BS_mut_Drev (5′-AAA AAT TCA GAA TTT CGG AAA ATC CAT GTA CGC GCA TCG ATG AGC CCT GCC CTC ACT TAT-3′), Tbf1_BS_G4_mut_Bfor (5′-AAA GTC GAT TAA AAG TCG GGC TAA CAC AAG TAT AAG CAG GGT TTT GCT ATA ACC GCG ATG-3′), Tbf1_BS_G4_mut_Crev (5′-TGA GCC CTG CGC TCA CTT ATG GCA TCG CGG TTA TAG CAA AAC CCT GCT TAT ACT TGT GTT-3′), Tbf1_BS_mut_Afor (5′-AAA AAA AAA AAA AAA AAG TTT TTC TTA AAT GCA TAG GTT TAA AGT CGA TTA AAA GTC GGG-3′), Tbf1_BS_G4_mut_Drev (5′-AAA AAT TCA GAA TTT CGG AAA ATC CAT GTA CGC GCA TCG ATG AGC CCT GCG CTC ACT TAT-3′).

For all the experiments, cells were grown at 28°C in YEP medium (1% yeast extract, 2% peptone) with 2% glucose (YEPD), with 2% raffinose (YEPR), or with 2% raffinose and 2% galactose (YEPRG). For strains carrying plasmids, cells were grown in Synthetic-Complete (SC) medium (6.7 g/l yeast nitrogen base) containing the appropriate sugar(s) at 2% concentration and nutrients to maintain the selection. For TMPyP4 treatment, overnight cultures grown in SC medium supplemented with 2% glucose were treated with 0–20–30–40 μM of TMPyP4 for 2 h and then plated on YPED. The cell cycle phase was evaluated using flow cytometric analysis using FlowJo v10 software (41). Plasmids pNOV1.4 and pFL82.4 were obtained by inserting the *NAT*^{R} marker gene amplified from pAG25 plasmid using NAT1_F (5′-GTC GAC ACA TGG AGG CCC AGA ATA CCC-3′) and NAT1_R (5′-GTC GAC CAG TAT AGC GAC CAG CAT TCA C-3′) oligos, into the SalI digested pRS426 and pFL60 vectors respectively. pFL60 was obtained by cloning *GAL1pr-DDC2* SacI/XhoI fragment obtained by digestion of pML105 (42) into pRS426. Plasmid pFL183.1 was obtained by performing a standard PCR protocol for site-directed mutagenesis using plasmid KP118 as a template (oligoF 5′-CCA AGC GGT AAA ACT TAC ATG CGA TGG TGG CGT CAC ATG GGT GGT CCA AAG-3′; oligoR 5′-CTT TGG ACC ACC CAT GTG ACG CCA CCA TCG CAT GTA AGT TTT ACC GCT TGG-3′). Plasmid pFL178.1 was obtained by cloning Chr VIII from coordinates 512240 to 512755 amplified with ChrVIII_for (5′-CTA GTC TAG AAA TGC ATA GGT TTA AAG TCG-3′) and ChrVIII_rev (5′-CTA GTC TAG ACT ATT TAT GGT GGA AAA GCT C-3′) oligos into XbaI digested pRS415, while pGELE50.1 was obtained by amplifying the G4 mutated DNA version from yeast strain YFL2922 as the template using the same strategy described for the pFL178.1.

pGELE39.1 plasmid was obtained by cloning *VID22* sequence amplified from yeast genomic DNA using Vid22BamHI_F (5′-CGC GGA TCC GAT GAG AGC GAT GGA CAC ACA G-3′) and Vid22XhoI_R (5′-AAC CGC TGC AGC TAT GGA AGA TAC TGA CTT GC-3′) oligos in BamHI/XhoI digested pRSETb vector.

Plasmid pGELE48 carrying mutation in *VID22 BED* domain C87A-C90R-H110Y-H115Y (32) was obtained by performing a site directed mutagenesis (Quik Change Agilent) in two steps on pGELE39.1 using for step 1 oligos VBED_C87A_C90R_FW (5′-ATG CTG GAG GCA GTA AAA GCC AAG TAC CGC GGT GTG ATA ATA AGA CGG-3′) and VBED_C87A_C90R_REV (5′-CCG TCT TAT TAT CAC ACC GCG GTA CTT GGC TTT TAC TGC CTC CAG CAT-3′), and for step 2 oligos VBED_H110Y_H115Y_Fw (5′-GAA GCC TCG CAA ACT TAT TTG TGG AGC ACG TAT AAG ATA GAC CCG A-3′) and VBED_H110Y_H115Y_Rev (5′-TCG GGT CTA TCT TAT ACG TGC TCC ACA AAT AAG TTT GCG AGG CTT C-3′).

DDC2 synthetic dosage lethality screen

The genetic screen was performed using SGA technology (43). Briefly, the query strain carrying the inducible *DDC2*-overexpressing plasmid (pFL82.4) was crossed to an ordered array of all the viable yeast deletion strains. Diploid cells were transferred to a sporulation-inducing medium, after which the germinated spores were selected for the simultaneous presence of the gene deletion and the plasmid. Cells were then transferred to a galactose-containing medium to induce *DDC2* overexpression, and colony size was analyzed as a measure of fitness, as described in (43). The whole procedure was performed in parallel with a control query containing the same vector devoid of the *GAL1pr-DDC2* gene (pNOV1.4). The screening was performed in triplicate and SGA scores and p-values were calculated as indicated in (43). For each replicate, an intermediate cut-off ($\varepsilon < -0.08$, $P < 0.05$ (43)) was applied and a total of 52 mutant strains identified in at least two replicates were selected for further analysis. The colony size and SGA score data are presented in Supplementary Table S2.

Spatial analysis of functional enrichment

Network annotations were made with the Python implementation of spatial analysis of functional enrichment (SAFE) ((44); https://github.com/baryshnikova-lab/safepy). The yeast genetic interaction similarity network

and its functional domain annotations were obtained from (45).

Serial dilution growth tests

Log-phase yeast cultures were diluted to 2 × 10^6 cells/ml. A series of 10-fold dilutions were prepared and spotted on YEP plates or selective plates containing the appropriate sugar. Plates were incubated at 28°C for 2–3 days. Where indicated, the plates were irradiated with 50 J/m^2 using a UV Stratalinker 2400 (Stratagene) or 100mM HU (USBiological, cat. H9120) was added.

Yeast genomic sequencing

Yeast DNA was prepared using MasterPure™ Yeast DNA Purification Kit (Epicentre, MPY80200) and quantified by fluorimetric assay using the Quant-iTTM PicoGreen® dsDNA Kit (Invitrogen, Carlsbad, CA, USA). Libraries were prepared using the Nextera XT library preparation workflow (Illumina, San Diego, CA, USA) to obtain DNA fragments ranging in size from 600 to 1400 bp approximately. Sequencing was carried out on the Illumina MiSeq platform generating about 8 million 250 × 2 paired-end reads for each sample. Raw sequencing data for the 11 *vid22* mutants and one control *S. cerevisiae* BY4741 strain have been deposited in the Short Read Archive (SRA) under Bioproject PRJNA646604. Data can be accessed through the following link: https://www.ncbi.nlm.nih.gov/sra/PRJNA646604.

SDS-PAGE and western blotting

Protein extracts were prepared in trichloroacetic acid (TCA) as described in (46) and separated by SDS-PAGE. Western blotting was performed with anti-Rad53 (gift from C. Santocanale) or anti-HA (12CA5) as primary antibody, and Goat anti-Mouse HRP (ThermoFisher-Scientific, cat.31430) or Goat anti-Rabbit HRP (ThermoFisher-Scientific, cat.31460) as secondary antibody using standard techniques. Anti-Rad53 signal was detected using film (Amersham Hyperfilm ECL), while anti-HA signal was acquired using a ChemiDoc™ Touch Imaging System (Bio-Rad).

Chromatin immunoprecipitation (ChIP) assays

Yeast strains used for direct ChIP measurements were obtained by 13Myc tagging of *VID22*. Overnight cell cultures pre-grown in YEP medium containing the appropriate sugar were diluted into fresh medium to a cell density of 4 × 10^6 cells/ml in 50 ml and grown at 28°C until they reached a density of 1.5–2 × 10^7 cells/ml before being collected for further analysis. Cells were cross-linked with 1% formaldehyde (Sigma) for 10 min at 25°C. Cross-linking was quenched by addition of 125 mM glycine. After harvesting cells, the pellet was resuspended in NP-40 ChIP lysis buffer (1% NP-40, 140 mM NaCl, 50 mM HEPES (pH 7.5), 1 mM EDTA (pH 8.0), 0,1% sodium deoxycholate, 1 mM phenylmethylsulfonyl fluoride, cocktail proteases inhibitors (Roche)). Yeast whole cell extracts were prepared by FastPrep® FP120 Cell Disrupter (Thermo) in NP-40 ChIP lysis buffer. Following lysate clarification by centrifugation at 16 000 × g for 30 min at 4°C, pellets were resuspended in NP-40 ChIP lysis buffer and samples were sonicated for 15 min, 30 s on, 60 s off, at high power at 4°C with a Bioruptor® Plus sonicator (Diagenode). Immunoprecipitation was carried out with anti-MYC (9E10) mouse monoclonal antibody following incubation with Dynabeads M-280 Sheep Anti-Mouse IgG (Life Technologies). Washing steps and reverse crosslinks were performed as reported in (47). Inputs and IPs were purified using Wizard® SV Gel and PCR Clean-Up System (Promega, A9281); DNA was eluted in 80–100 µl of nuclease-free water. Nucleic acids were quantified with Qubit 4 Fluorometer using Qubit dsDNA HS Assay kit (Invitrogen Q32854), and the quantity obtained was 0.5–1 ng/µl for Inputs and 0.1–0.3 ng/µl for IPs. Samples were directly analyzed by qPCR, then stored at −20°C.

ChIP-seq analysis

Strains used for ChIP-seq protocol were wild type (no tag) as negative control and Vid22-13Myc tag. ChIP was performed as described above, starting from 150 ml of cultures with a cell density of 1.5–2 × 10^7 cells/ml. DNA obtained from two independent pooled ChIP experiment was quantified using the Qubit dsDNA High Sensitivity Assay Kit (Invitrogen, Q32851). ChIP-seq libraries were generated starting from 5 ng DNA using an in house protocol (48), quality checked on a Bioanalyzer High-Sensitivity DNA Chip (Agilent) and sequenced on an Illumina HiSeq 2000 at 50 bp read length. Reads were aligned to the reference assembly of the *Saccharomyces cerevisiae* genome using the bowtie (49) program with the following parameters '-m 1 –best –strata' to allow up to 1 mismatch and report only the best scoring alignment. PCR duplicates were removed by means of the rmdup utility from the samtools package (50). Peaks were called by means of the MACS2 program with default parameters. A FDR threshold of 10E-10 was applied (51).

qPCR analysis

For qPCR analysis of ChIP samples, input and immunoprecipitated (IP) DNA were analyzed using primer pairs producing a 89 bp amplicon for the *HHT2* locus (coordinates Chr XIV 575707–575796) (52) and a 177 bps amplicon for the Chr VIII-G4 at *SKN7* locus (coordinates Chr VIII 512387–512564): Chr VIII_F (5'-ATG AGC AAA ATG TGG TCA GC-3') Chr VIII_R (5'-ACC CAA ACA AAA GCA GCA AG-3') and HHT2_CDS_A (5'-TCA ATC TTC TGC TAT CGG TGC TT-3') HHT2_CDS_B (5'-GCG TGA ATA GCA GCC AGA TTA GT-3'). Oligos for qPCR were purchased from Eurofins Genomics.

PCR reactions were performed in 25 µl total volumes containing 12,5 µl 2× Quantitative Master Mix with SYBR® Green (GeneSpin proprietary formulation ready-to-use containing Xtra Taq Pol, dNTPs, $MgCl_2$ and stabilizers optimized for use in real time PCR amplification), primers (200 nM Chr VIII; 130 nM *HHT2*) and template (2 µl 1:100 INPUT or undiluted IP). All components were mixed in 96-well hard-shell PCR plates with clear wells (Bio-Rad), which were sealed with optically clear adhesive Microseal® 'B' seals (Bio-Rad) and centrifuged at 1000×*g*

for 10 s. The thermal conditions during reaction were 3 min at 95°C followed by 39 thermal cycles at 62°C for 30 s and 95°C for 10 s. All PCR reactions were assembled manually and qPCR experiments performed on a CFX Connect Real-Time PCR System (Bio-Rad). We accepted PCR efficiencies between 95–105%. All PCR reactions were performed in technical triplicates and all biological experiments were performed at least in triplicates. ΔC_T (Cycle Threshold) was first calculated, as the difference between IP and input C_T values (after correcting for input sample dilution), for both the target gene and the *HHT2* gene (chosen as an internal standard). The target gene enrichment in IP DNA was then calculated using Pfaffl method (53). Means of fold enrichment and standard errors of the mean were calculated for biological replicates and unpaired Student's *t*-test was used to determine statistical significance between samples.

Pulsed field gel electrophoresis

PFGE was performed using the Pulsaphor system with a hexagonal electrode array (Amersham Pharmacia Biotech). Agarose plugs with yeast chromosomal DNAs were prepared as previously described (54). For standard chromosome separation, plugs were loaded in a 1% agarose gel in 0.5× TBE and sealed in the gel using LMP agarose 0.5% in 0.5× TBE. The running conditions were 165 V with 60 s pulses for the first 12 h and 90 s pulses for the last 12 h at 8°C. To visualize DNA, the gel was stained in a solution with 2 μg/ml Ethidium Bromide in 0.5× TBE for 30 min; VersaDoc or Chemidoc (Bio-Rad) were used to acquire images of the stained gels.

Measurement of telomere length using Southern blotting

Yeast genomic DNAs were extracted using standard methods starting from plateau-phase cultures. DNA samples were digested with XhoI or SalI restriction enzymes; the obtained fragments were separated by electrophoresis in 0.8% agarose 1× TBE. Radiolabeled (DECAprime™ II Kit Invitrogen) specific probes were used to visualize telomeric poly(GT) tails (55) and subtelomeric regions (56). A standard protocol was used for hybridization and detection (57).

Identification of structural variants and copy number alterations, and GC normalization

Raw reads were subjected to quality trimming using the sliding window operation from the Trimmomatic program (58) (average quality 20, window length 8). Assembly was performed with the SPAdes program using the following set of kmers: 33, 55, 77 and 99. Detection of large indels and structural variants was performed by aligning the final assemblies to the reference sacCer3 genome, (downloaded from http://hgdownload.soe.ucsc.edu/goldenPath/sacCer3/bigZips/) by means of the Mummer4 program (59). Large scale structural rearrangements were identified directly from the alignment by using a custom Perl script. Raw reads were aligned to the reference sacCer3 assembly of the yeast genome using the bowtie2 program. Coverage profiles were computed on sliding genomic windows of 200 bp, overlapped by 100 bp, by using bedtools coverage tool. Variant calling was performed by the means of the varscan2 software (60), with default parameters. The rlm function from the R MASS package (61) was applied to perform GC composition normalization. Scaling factors for GC normalization and estimates of the sensitivity of our CNV detection assay before and after the application of GC content normalization are reported in Supplementary Tables S3 and S4 respectively. Copy Number Variants were identified by pairwise comparisons of coverage profiles of matched genomic windows between the wild-type strain and each *vid22Δ* strain by means of the chi-squared test. The Benjamini-Hochberg correction was applied to control the false discovery rate. A cut-off FDR of 0.05 was used for the identification of genomic windows showing significantly altered coverage. Finally, overlapping windows were merged using bedtools merge, to derive larger genomic intervals. Intervals of more than 500 bp in size, formed by at least three or more windows showing significantly altered coverage were considered as bona fide CNVs. Mitochondrial genome (mtDNA) copy numbers were estimated by comparing the proportion of NGS reads mapped to the nuclear and mitochondrial genome. To avoid confounding factors for the nuclear genome, only single copy *S. cerevisiae* as defined in https://www.nature.com/articles/s41586-018-0030-5 was considered. Coverage levels were normalized by applying the RPKM normalization. Mitochondrial copy number was inferred as the ratio between the average RPKM of single copy nuclear genes and with the mtDNA RPKM.

Intersection of coordinates of genomic features were performed using the bedtools intersect utility (49).

Gross chromosomal rearrangements

Gross chromosomal rearrangement GCR assays were performed as previously described (62). Different cassettes were integrated into the *PRB1* locus near the *CAN1* gene. GCR rates were measured using the webtool http://flucalc.ase.tufts.edu/ (63) and the MMS maximum likelihood statistical method (64). Briefly, 5 colonies for each strain were grown at 28°C for 3 days to obtain saturated cultures. Cells were plated on SC medium with 5-FOA and canavanine to select GCR events and SC medium to measure the total cell number. The G4 forming sequences tested in the GCR assay are: 5′-**GGG** TCC TCC AAG CGG TAA AAC TTA CAT **GGG** ATG GT**G GGG** TCA CAT **GGG**-3′ (Figure 5B and (21)) and 5′-**GGG** TTT TGC TAT AAT A**GG** GAT GCC ATA AGT GA**G GGC AGG G**-3′ (ChrVIII-G4 Figure 6F)

Reflective phantom interface (RPI) sensor preparation and measurement

Reflective phantom interface (RPI), is an optical label-free biosensor enabling the studying a variety of interactions, including antigen-antibody (65), protein-glycan (66), DNA–DNA (67), RNA–DNA (68) and protein–DNA (69). RPI analysis was employed for the characterization of the interaction between BG4 (70) or Vid22 and different DNA strands immobilized on the RPI sensing surface. The BG4 antibodies were produced with the support of the Protein Purification Facility of the Biosciences Department at University of Milan starting from pSANG10-3F-BG4 plasmid (70). Wild type Vid22 and a mutant in the BED domain

were purified from BL21(DE3)pLysS *Escherichia coli* cells. Cells were grown in LD with 50 μg/ml Ampicillin at 30°C until $OD_{600nm} = 0.5$, then proteins expression was induced by adding 0.2 mM IPTG at 30°C for 3–4 h. For protein purification, cells were disrupted in Lysis buffer (300 mM NaCl, 50 mM sodium phosphate pH 7) with 30 mM imidazole by sonication (10 cycles 15 s ON/30 s ice). Protein extracts were clarified by centrifugation at 15 000 rpm for 1 h at 4°C. Clarified extracts were then loaded into a Ni-NTA Agarose (QIAGEN 1018244), matrix was washed with lysis buffer supplemented with 50 mM imidazole, then proteins were eluted in lysis buffer with 250 mM imidazole. Proteins were dialysed in 300 mM NaCl, 50 mM Tris–HCl pH 7.5, 10% Glycerol using Thermo Scientific™ Slide-A-Lyzer™ Dialysis Cassettes 3500 MWCO. The DNA sequences used in this study are reported in Supplementary Table S5. All oligonucleotides were purchased from Integrated DNA Technologies (IDT) with Ultramer synthesis. All buffers and reagents were purchased from Sigma-Aldrich and prepared according to common protocols using Milli-Q pure water. DNA probe strands were covalently immobilized on the surface of RPI sensing chips in spots having 150–200 μm diameter following the procedure described in (71). To allow the formation of G4 structures and coupling with the complementary sequence, the oligonucleotides were previously denatured by heating for 5 min at 95°C and renatured by slow cooling at room temperature in 10 mM Tris–HCl (pH 8.0), 100 mM KCl, 1 mM EDTA (pH 8.0). Before use, the BG4 antibody was suspended in measuring buffer A (20 mM Tris-HCl (pH 7.5), 1 mM DTT, 0.1 mg/ml BSA, 30 mM KCl) while Vid22 was suspended in measuring buffer B (20 mM Tris–HCl (pH 7.5), 1 mM DTT, 0.1 mg/ml BSA, 150 mM KCl). The RPI measurements were performed by using the apparatus and the analysis algorithm described in (71). The sensor cartridges were filled with 1.3 ml of measuring buffer A or B. The cartridges were kept at 25°C during the measurement by a thermalized holder and rapid mixing of the solution was provided by a magnetic stirring bar. Sample spikes of target BG4 or Vid22 were performed by adding 50 μl measuring buffer containing different amounts of target molecules to a final concentration in the cartridge of 0.08, 0.4, 2, 10 and 50 nM (only for Vid22). Time sequences of RPI images of the spotted surface were analyzed by a custom Matlab program in order to obtain the surface density of the target molecules $\sigma(t)$ binding to the surface immobilized probes (71). The binding curves $\sigma_i(t)$ obtained by the RPI measurements were analyzed assuming a simple Langmuir model (72). The binding curves of at least six spots with identical composition were averaged and then fitted with an exponential growth function. Supplementary Table S6 reports the values for K_d, k_{on} and k_{off} obtained from the fits of the curves in Figure 6B, E and Supplementary Figure S6D.

UV spectrophotometric analysis

To collect information on the structure adopted by a DNA sequence, we measured and analyzed a series of thermal difference spectra of the oligos studied in this work. A thermal difference spectrum (TDS, (73)) can be computed subtracting the absorbance spectrum of a DNA oligonucleotide measured at low temperature (below melting) from the spectrum of the same oligonucleotide measured at high temperature (above melting). TDS recapitulates how temperature affects the spectral properties of the studied species and shows features that can be used to determine whether an oligonucleotide has transitioned from a G-quadruplex structure to a single stranded state upon thermal melting.

TDS is based on the widely studied hyperchromic effect of nucleic acids: the unstacking of nitrogenous bases results in an increase of their UV absorbance and this variation in the extinction coefficient is usually monitored to determine the melting temperature of oligonucleotides. This phenomenon effectively links the structure of the oligonucleotides to their absorbance and aside from monitoring the variation of amplitude at 260 nm, where the maximum typically lies, other wavelengths can be monitored to gather valuable information on the state in which the nitrogenous bases lie. While DNA duplexes show an increase of extinction coefficient at every typically measured wavelength when a melting experiment is performed, this is not true for G-quadruplex forming sequences: for the latter, as temperature increases above the G-quadruplex melting, the absorbance for wavelengths $\lambda < 290$ nm increases, while it decreases for $\lambda > 290$ nm (74). An isosbestic point in the TDS with value equal to zero at $\lambda = 290$ nm (black arrow in Figure 6C) corresponds to an inversion of temperature dependence of the absorbance and is a signature of G-quadruplex producing sequences.

All oligonucleotides used to measure TDS were purchased from Eurofins Genomics with no terminal modification using EXTREmers synthesis. The lyophilized powder was resuspended in Milli-Q pure water and for each DNA sequence a sample at ≈1 $OD_{260\ nm}$ was prepared with 100 mM KCl and degassed under vacuum. A series of complete spectra was acquired using a Thermo Scientific Evolution 300 UV-Vis Spectrophotometer in the 220–320 nm range using a 1 cm path quartz cuvette. The melting experiments were performed from 20°C to 90°C controlling the temperature with the instrument peltier module to obtain a rate of 0.4°C/min.

The signals of the thermal difference spectra (ΔA) were computed subtracting the spectrum measured at 20°C from the spectrum measured at 90°C and normalized for the maximum absorbance at low temperature (A_{max}) to be directly compared and to highlight how a point mutation can heavily affect the behavior of Chr VIII-G4 sequence.

DNA:RNA immunoprecipitation (DRIP) assays

DNA:RNA hybrids were immunoprecipitated using the S9.6 antibody from nucleic acid gently extracted and overnight digested with 50U of HindIII, EcoRI, BsrGI, XbaI and SspI, 2 mM spermidine and 2.5 ml BSA 10mg/ml genomic DNA; samples were then treated or not with RNase H, as described (75). Quantitative PCR was performed at the indicated loci. *SKN7* oligos (5′-CCG TTA ATT TCG CGA GCT TAT ACC TCA CCA TTC CAT TG-3′) and (5′-CCG TTA ATT AGC GCT GAT GTT GGA AGA TAG TAA GGT GA-3′). *GCN4* oligos (5′-TTG TGC CCG AAT CCA GTG A-3′) and (5′-TGG CGG CTT CAG TGT TTC TA-3′).

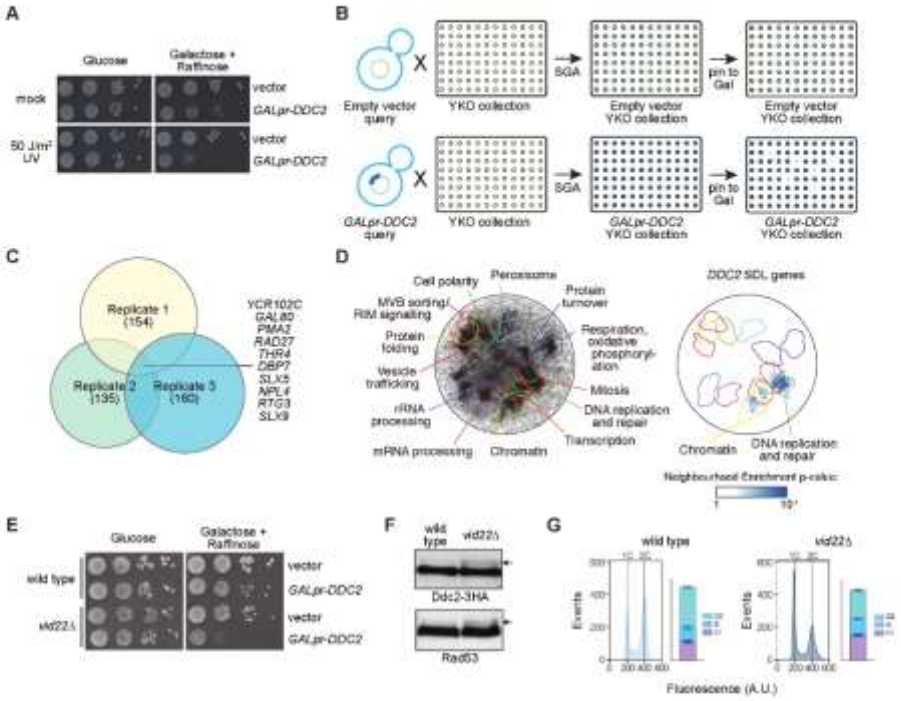


Figure 1. Genome-scale screen for synthetic dosage fitness defects with *DDC2* overexpression identifies *VID22*. Effect of *DDC2* overexpression on cell fitness. (**A**) Tenfold serial dilutions of exponentially growing wild-type cells carrying empty vector or *DDC2* under the control of the *GAL1* promoter were spotted on rich medium plates containing either glucose or galactose and raffinose as the carbon source. Cells were exposed to UV (50 J/m^2), or mock-treated. Images were taken after three days of incubation at 28°C. (**B**) Schematic representation of the *DDC2* synthetic dosage lethality screen. The galactose-inducible *DDC2* gene (*GAL1pr-DDC2*), or the empty vector, was introduced into the yeast deletion collection (YKO) by crossing the collection with query strains containing the plasmids. Haploid strains containing each gene deletion and the plasmid were isolated using SGA methodology. Each strain was pinned to media containing galactose to induce *DDC2* expression. Synthetic dosage fitness defects were evident when colonies were smaller in *GAL1pr-DDC2* than in empty vector. (**C**) The overlap of the *DDC2* SDL genes for the three replicate screens is plotted as a Venn diagram. The number of positives in each replicate is indicated, as are the 10 genes identified in all three screens. (**D**) Spatial analysis of functional enrichment. On the left, the yeast genetic interaction similarity network is annotated with GO biological process terms to identify major functional domains (45). Thirteen of the 17 domains are labeled and delineated by colored outlines. On the right, the network is annotated with the 52 *DDC2* SDL genes. The overlay indicates the functional domains annotated on the left. Only nodes with statistically supported enrichments (SAFE neighborhood enrichment *P*-value < 0.05) are colored. (**E**) The *DDC2*–*VID22* SDL interaction is validated by growth analysis. Wild-type and *vid22*Δ cells were transformed with the empty vector or with the *GAL1pr-DDC2* plasmid. Tenfold serial dilutions of exponentially growing cultures were spotted on glucose and galactose plus raffinose plates to repress or induce Ddc2 overexpression, respectively. Images were taken after three days of incubation at 28°C. (**F**) Spontaneous DNA damage checkpoint activation was evaluated by monitoring phosphorylation of the checkpoint proteins Ddc2 (Ddc2-3HA) and Rad53. Protein extracts from wild-type or *vid22*Δ cells expressing Ddc2-3HA were analyzed by western blotting with anti-HA and with anti-Rad53 antibodies. Arrows indicate the phosphorylated forms of Ddc2 and Rad53. (**G**) The cell cycle distribution of exponentially growing cultures was determined by flow cytometry of logarithmic phase wild-type (G1-25.9%; S-19%; G2-54.2%) and *vid22*Δ (G1-34.3%; S-22.4%; G2-38.3%) cultures from three independent replicates. The positions of cells with 1C and 2C DNA contents are indicated.

RESULTS

A genome-wide screen to identify new genes affecting genome stability

S. cerevisiae Ddc2 is a binding partner and activator of the DNA damage checkpoint (DDC) apical kinase Mec1 (human ATR) (76). *DDC2* overexpression has only a minor effect in unperturbed cells, while it causes hyperactivation of the checkpoint response after DNA damage, leading to prolonged cell cycle arrest and increased cell lethality ((42) and Figure 1A). We hypothesized that yeast mutants spontaneously accumulating DNA damage might be sensitized to *DDC2* overexpression (77), and therefore would show synthetic dosage fitness defects (78) with *DDC2*. We

developed a screen to identify genes and pathways involved in genome integrity maintenance based on their sensitivity to *DDC2* overexpression. Using Synthetic Genetic Array (SGA) technology (43,79), we introduced a multicopy plasmid carrying *DDC2* under the control of a galactose-inducible promoter, and a control empty vector, into the yeast knockout (YKO) collection (Figure 1B). We identified mutants that, on galactose-containing medium, exhibited reduced fitness, often referred to as synthetic dosage lethality (SDL), when the *DDC2*-containing plasmid was present compared to controls with the empty vector. From three independent replicate screens, we identified 52 genes that were positive in at least two replicates (Figure 1C and Table 1). Consistent with our initial hypothesis, several of these 52 genes were previously identified in at least one systematic screen for increased spontaneous DNA damage (35,80–82).

Twelve genes were previously reported to exhibit synthetic dosage growth defects when combined with *DDC2* overexpression (*Saccharomyces* Genome Database, https://www.yeastgenome.org, accessed 12 July 2020). These include *HOG1*, *RPD3* and *SLT2* (83,84) and *DUN1* (85), which were identified in screens where the given deletion mutant was the query. An SDL screen in a pooled format with *GAL1pr-DDC2* as the query identified *ASF1*, *CSM3*, *CTF4*, *MMS1*, *RAD27*, *RTT101*, *RTT109* and *TOF1* (86). Of these 12 genes, *RAD27* and *CTF4* were identified in our screens. Thus, we identified 50 putative synthetic dosage interactions that had not been reported previously.

To assess the functional properties of the 52 gene *DDC2* SDL set, we applied spatial analysis of functional enrichment (SAFE) (44) to determine if any regions of the functional genetic interaction similarity yeast cell map (45) are over-represented for the SDL gene set (Figure 1D). We found a statistically supported over-representation of the *DDC2* SDL genes in the DNA replication and repair neighborhood of the genetic interaction cell map, indicating that defects in DNA replication and repair sensitize cells to *DDC2* overexpression. Over-representation of the *DDC2* SDL genes in the chromatin and transcription neighborhoods of the genetic interaction cell map was also evident, consistent with a role for *DDC2* in sensing chromatin dysfunction.

VID22 suppresses spontaneous DNA damage

Of the 52 genes identified in our screens, we focused our attention on *VID22*. Although this gene shows conflicting functional annotations, including vacuole import and degradation (87), it was previously identified also in other genetic screens searching for players involved in genome integrity maintenance (87–89). *VID22* encodes a nuclear protein (31,90) computationally predicted to contain a BED-type zinc finger domain and a RNaseH-like domain (32,33) and it was suggested to play a role in DNA DSB repair (34,35).

Previous observations in strains experiencing spontaneous DNA damage (29,91,92)—as well as the rationale behind our genetic screen—led us to suspect that, in a *vid22* mutant, the DDC might be chronically activated even in the absence of external genotoxic insults albeit to a low level. After confirming the synthetic dosage lethality due to *DDC2* overexpression in a *vid22Δ* strain (Figure 1E), we verified our hypothesis that both Ddc2 and Rad53, markers of DDC activation (93,94), are partially phosphorylated in unperturbed *vid22Δ* cells (Figure 1F), similarly to other mutants known to spontaneously accumulate genomic instability such as *slx5Δ* and *slx8Δ* that also emerged in our SDL screen (Supplementary Figure S1A, (77) and Table 1). Thus, loss of *VID22* causes spontaneous activation of the DNA damage checkpoint response, which is compatible with an increased formation of spontaneous DNA lesions. Consistently with a previous screen (95), analysis of the cell cycle distribution of *vid22Δ* mutants shows a significant accumulation in the G1 subpopulation (Figure 1G). This observation excludes the possibility that Ddc2 and Rad53 phosphorylation is due to a G2/M accumulation (76), supporting instead the notion that the checkpoint response might be chronically alerted as a consequence of the elevated spontaneous formation of endogenous DNA damage. We further investigated whether the accumulation of cells in G1 observed in Figure 1G may be due to this chronic activation of the DDC, delaying entry into S-phase. However, ablation of the checkpoint kinase Mec1 does not rescue this defect, suggesting that the cell cycle delay observed in *vid22Δ* mutants may be, at least partially, of different origin (Supplementary Figure S1B).

Loss of *VID22* leads to chromosomal alterations

Chronic exposure to endogenous DNA damage frequently leads to genomic instability. To assess whether loss of *VID22* causes the accumulation of chromosomal aberrations, we analyzed the chromosomes of 11 independent unperturbed *vid22Δ* clones by PFGE and compared them to wild-type strains (Figure 2 and Supplementary Figure S2A). Eight of the eleven *vid22Δ* clones exhibited at least one detectable chromosomal aberration (Figure 2), which did not occur in the 10 wild-type control clones (Supplementary Figure S2A). In *vid22Δ* mutants we identified either entire chromosomal duplications (e.g. Chr XI and I in strain 1) or increased chromosome length (e.g. Chr XI in strain 2 or Chr V or VIII in strain 4) (Figure 2). To exclude that the observed abnormalities might be due to different numbers of cell generations, we repeated the analysis starting from the sporulation of a heterozygous diploid *VID22/vid22Δ*. As shown in Supplementary Figure S2B, after the same number of cell divisions, 4 of 7 *vid22Δ* mutants show at least 1 major chromosomal abnormality, not observed in any of the 13 wild-type clones tested.

To investigate whether particular chromosomal or sequence contexts are associated with the observed genome instability, we sequenced the genomes of the 11 *vid22Δ* strains shown in Figure 2. While *de novo* assemblies of the mutant strains' genomes did not indicate the presence of large structural genomic alterations (insertions, inversion, or translocations), prediction of Copy Number Variants (CNVs) based on the depth of coverage analyses were consistent with complete duplications for chromosomes I (strain 1), III (9 and 11), XI (1 and 7) and XIII (8 and 9). In addition, we observed 468 smaller CNVs not detected by PFGE, associated with 226 distinct genomic loci (Supplementary Table S7). Importantly, patterns and the total num-

Table 1. *DDC2* synthetic dosage lethality genes

Genes with epsilon < −0.06 in at least 2 replicates[a]	Systematic name	Replicate 1 epsilon[b]	Replicate 2 epsilon	Replicate 3 epsilon	Mean epsilon	Std Dev	Brief description[c]
YCR102C	*YCR102C*	−0.34595	−0.20615	−0.22037	−0.2775	0.0524	Putative quinone oxidoreductase
GAL80	*YML051W*	−0.36714	−0.16483	−0.25393	−0.2426	0.0578	Transcriptional regulator involved in the repression of GAL genes
PYC2	*YBR218C*	0	−0.38435	−0.25274	−0.2058	0.1582	Pyruvate carboxylase isoform
PMA2	*YPL036W*	−0.23367	−0.11122	−0.1458	−0.1686	0.0609	Plasma membrane H+-ATPase
RAD27	*YKL113C*	−0.16946	−0.08376	−0.16573	−0.2403	0.0588	5′ to 3′ exonuclease; 5′ flap endonuclease
VID22	*YLR373C*	−0.07129	−0.10449	−0.21956	−0.1318	0.0835	Glycosylated integral membrane protein localized to plasma membrane
MMS22	*YLR320W*	−0.13772	−0.08625	−0.23148	−0.1252	0.0929	Subunit of E3 ubiquitin ligase complex involved in replication repair
VPS68	*YOL129W*	−0.02521	−0.23802	−0.10864	−0.1240	0.0876	Vacuolar membrane protein of unknown function
THR4	*YCR053W*	−0.12709	−0.11031	−0.12381	−0.1204	0.0073	Threonine synthase
DBP7	*YER026C*	−0.08875	−0.16435	−0.10039	−0.1186	0.0543	Putative ATP-dependent RNA helicase of the DEAD-box family
SLX5	*YDL013W*	−0.132	−0.1192	−0.12084	−0.1173	0.0038	Subunit of the Slx5-Slx8 SUMO-targeted Ub ligase (STUbL) complex
LRO1	*YNR008W*	0.08822	−0.15822	−0.15615	−0.1127	0.0875	Acyltransferase that converts diacylglycerol to triacylglycerol (TG/A)
NKP2	*YLR311W*	−0.23546	−0.02581	−0.09278	−0.1314	0.0785	Central kinetochore protein and subunit of the CBF3 complex
NPL4	*YBR170C*	−0.16522	−0.12311	−0.10379	−0.1107	0.0088	Substrate-recruiting cofactor of the Cdc48p-Npl4p-Ufd1p segregase
SAT2	*YLR412C*	−0.15523	−0.1565	−0.01084	−0.1108	0.0841	GTPase-activating protein for Gpa1p
IRC4	*YDR540C*	−0.1842	−0.02255	−0.106	−0.1043	0.0868	Protein of unknown function
RPL22A	*YLR061W*	−0.12807	0.00318	−0.19649	−0.1038	0.0831	Ribosomal 60S subunit protein L22A
NCS2	*YBL093C-A*	−0.09108	−0.07698	−0.13178	−0.0999	0.0232	Protein involved in regulation of phospholipid metabolism
EKI1	*YDR147W*	−0.12455	−0.07955	−0.09191	−0.0987	0.0190	Ethanolamine kinase
SLX8	*YER116C*	−0.15266	0.01681	−0.15558	−0.0972	0.0886	Subunit of Slx5-Slx8 SUMO-targeted ubiquitin ligase (STUbL) complex
MNN10	*YDR245W*	−0.09849	−0.11437	−0.07617	−0.0963	0.0157	Subunit of a Golgi mannosyltransferase complex
PTC4	*YBR125C*	−0.23321	−0.11238	0.05822	−0.0958	0.1196	Cytoplasmic type 2C protein phosphatase (PP2C)
BYE6	*YBL103C*	−0.09982	−0.08377	−0.09895	−0.0942	0.0074	bHLH/Zip transcription factor for retrograde (RTG) and TOR pathways
PAC10	*YGR078C*	0.0254	−0.13436	−0.16326	−0.0921	0.0839	Part of the heterotetrameric co-chaperone GimC/prefoldin complex
SLX9	*YGR031C*	−0.0811	−0.08407	−0.10033	−0.0915	0.0122	Protein required for pre-rRNA processing
GUP1	*YGL084C*	−0.02287	−0.13467	−0.09719	−0.0912	0.0536	Plasma membrane protein involved in remodeling GPI anchors
KCC4	*YCL024W*	−0.13904	−0.08448	−0.04869	−0.0909	0.0374	Protein kinase of the bud neck involved in the septin checkpoint
MRS2	*YOR334W*	−0.1389	−0.15801	0.02559	−0.0898	0.0820	Mitochondrial inner membrane Mg^{2+} channel
RTS1	*YOR042W*	−0.06642	−0.09715	−0.09979	−0.0895	0.0142	B-type regulatory subunit of protein phosphatase 2A (PP2A)
TIM18	*YDR507C*	0.09103	−0.14955	−0.20481	−0.0879	0.1283	Component of the mitochondrial TIM22 complex
SAC7	*YDR389W*	−0.09922	−0.07025	−0.08967	−0.0864	0.0121	GTPase activating protein (GAP) for Rho1p
HTZ1	*YOL012C*	−0.12845	−0.09077	−0.07869	−0.0859	0.0571	Histone variant H2A.Z
CTF4	*YPR135W*	−0.08526	−0.10866	−0.14951	−0.0853	0.0599	Chromatin-associated protein
LYS2	*YBR115C*	0	−0.13134	−0.1298	−0.0840	0.0596	Alpha aminoadipate reductase
YDL102C	*YDL102C*	−0.01608	−0.08916	−0.1529	−0.0832	0.0557	Overlaps with CDC9 promoter
MRC1	*YCL061C*	−0.10752	−0.05381	−0.09651	−0.0823	0.0224	S-phase checkpoint protein required for DNA replication
PHO5	*YBR092C*	−0.14925	−0.12745	0.03574	−0.0803	0.0823	Repressible acid phosphatase
MRX10	*YDR282C*	−0.08352	−0.0926	−0.06064	−0.0756	0.0247	Mitochondrial inner membrane protein of unknown function
RPL40A	*YIL148W*	0.10598	−0.16715	−0.16521	−0.0749	0.1227	Ubiquitin-ribosomal 60S subunit protein L40A fusion protein
YDJ1	*YLR090W*	0.0167	−0.08585	−0.15132	−0.0722	0.0689	Chaperone with a role in facilitating mitochondrial protein import
YNL140C	*YNL140C*	−0.02494	−0.08911	−0.10192	−0.0720	0.0337	Protein of unknown function
PEX9	*YMR033W*	0.08997	0.09198	0.02484	0.0705	0.0521	Peroxisomal membrane signal receptor for peroxisomal matrix proteins
UBP6	*YFR010W*	−0.09206	−0.06376	−0.11433	−0.0701	0.0478	Ubiquitin-specific protease
DST1	*YGL043W*	−0.09769	0.00767	−0.11859	−0.0689	0.0547	General transcription elongation factor TFIIS
SPO73	*YER046W*	−0.13322	−0.09846	0.02037	−0.0671	0.0927	Meiosis-specific protein required for prospore membrane morphogenesis
BRE1	*YNR051C*	−0.13275	−0.1059	0.0386	−0.0667	0.0752	Ubiquitin protein ligase
MCA1	*YOR197W*	−0.0821	−0.10252	−0.06329	−0.0626	0.0428	Ca^{2+}-dependent cysteine protease
SIF2	*YBR283W*	0.01601	−0.09958	−0.09643	−0.0600	0.0538	WD40 repeat-containing subunit of Set3C histone deacetylase complex
YOS9	*YLR161C*	0.05837	−0.12347	−0.10298	−0.0593	0.0888	Putative transcription factor, contains Forkhead Associated domain
YAL030W	*YAL030W*	0.04237	−0.09285	−0.11481	−0.0580	0.0894	Dubious open reading frame
SCS7	*YMR272C*	0.20929	−0.13347	−0.0807	−0.0083	0.1767	Sphingolipid alpha-hydroxylase
AVA3	*YGR097W*	0.58527	−0.11335	−0.10438	0.1202	0.3727	Regulator of the Yvc1p glycerol channel

[a] Genes identified in all three replicates are indicated in **bold**.
[b] Epsilon values with $P > 0.05$ are entered as '0'.
[c] As annotated in YeastMine (https://yeastmine.yeastgenome.org/).

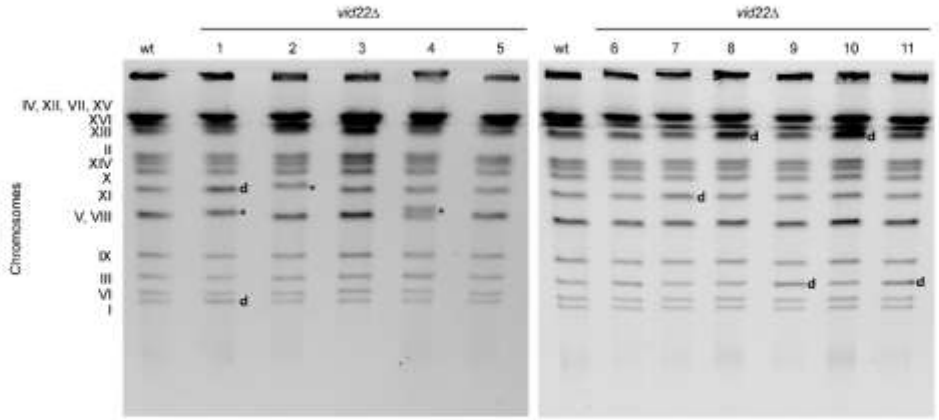


Figure 2. *vid22Δ* cells accumulate gross chromosomal rearrangements and chromosome duplications. DNA was prepared from wild-type and eleven independent *vid22Δ* mutants and chromosomes were fractionated by pulsed-field gel electrophoresis. The positions of the 16 chromosomes are indicated on the left. * indicates chromosome size changes; **d** indicates chromosome duplications.

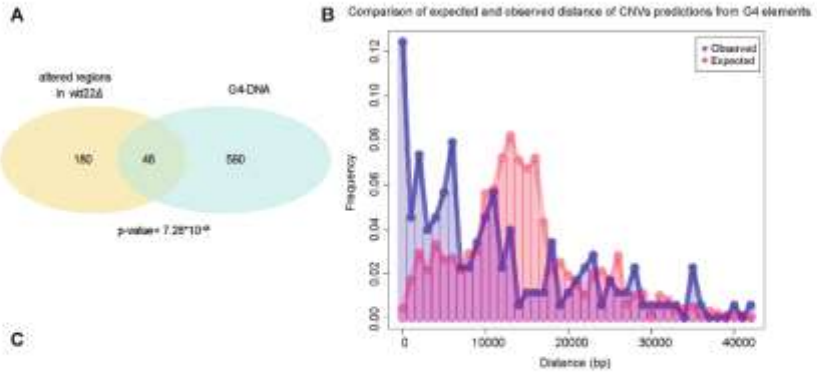
ber of observed single nucleotide substitutions and small indels (less than 5 base pair in size) to the reference assembly of the *S. cerevisiae* genome, were highly consistent between wild type BY4741 and *vid22* mutant strains (Supplementary Table S8), with only a marginal increase observed in strain 9. Overall this observation might suggest that *vid22* mutants are not systematically associated with an increased mutational burden and/or systematic defects in DNA repair mechanisms. Intriguingly, the CNVs detected from genome sequencing of *vid22Δ* strains show both statistically supported overlap and an association with G4 elements predicted by Capra and colleagues (17) (Figure 3A and B, hypergeometric $P < 7.28E{-}48$), suggesting that disruption of *VID22* is linked to increased genomic instability at G4-DNA rich genomic regions. In addition, overlap between the observed CNVs and G4s is still significant when interrogating different databases of predicted G4 regions (Supplementary Tables S9, S10 and Figure S3 (17,96,97)). Consistent with this model, we observed that *VID22* deletion induces great instability of the mitochondrial genome, which is particularly rich in potential G-quadruplex forming regions (17) (Figure 3C).

VID22 maintains telomere length

Telomeres are among the regions that show the highest density of sequences potentially forming G4-DNA structures (17,98). Budding yeast telomeres terminate with repetitive TG_{1-3} ends and contain an adjacent X core conserved region; an additional Y' subtelomeric region is also present at several telomeres, as schematized in Figure 4 (99,100). Telomeric sequences are highly repetitive, hampering their analysis by short sequencing reads. We directly monitored the telomeres of the 11 *vid22Δ* mutants using a Southern blot approach. Digestion of genomic DNA with XhoI generates a population of TG_{1-3} repeats of XY' type telomeres with a size of 1–1.5 kb (Figure 4A). In the X-only telomeres, XhoI cuts in variable positions, generating a wide spectrum of digestion fragments with sizes between 3 and 12 kb ((56) and Figure 4A). Using a TG_{1-3} specific probe, we found that the deletion of *VID22* generally causes an increase in the average length of telomeric ends (Figure 4A). This phenomenon is clearly detected in the XY'-type telomeres (bottom of panel 4A), and is also evident in the X-only fragments (top of the panel 4A). To detect X-only telomeres specifically, we used the strategy described in (56) and schematized in Figure 4B, which allowed us to visualize I_L, III_L, XI_R and XV_L telomeres. As shown in Figure 4B, in the absence of *VID22*, we were able to distinguish telomere alterations that are particularly evident at telomeres III_L and XI_R.

VID22 suppresses G4-dependent genome instability

To explore the hypothesis that *vid22Δ* mutants are susceptible to genome instability in the presence of G4-DNA structures, we exploited an assay for gross chromosomal rearrangements (GCRs) (21,101). Briefly, we used a modified chromosome V that harbours the *CAN1* and *URA3* markers in the left arm (Figure 5A). This configuration allows simultaneous selection on canavanine and 5-FOA to identify chromosome arm loss and large interstitial deletions. To analyse G4-induced GCRs, we inserted a cassette predicted to form G4-DNA (21) at the *PRB1* locus on the modified chromosome and monitored the GCR frequency. When a G4-forming motif is employed, deletion of *VID22* leads to a 15-fold increase in GCRs relative to the corresponding wild-type strain (Figure 5B). By contrast, deletion of *VID22* has little effect when the *PRB1* locus is unchanged (Figure 5B). To test whether this effect



Strain	Genotype	Strain number	mtDNA copy number relative to WT
BY4741	*WT*		1.00
DMA3420	*vid22Δ*	1	0.39
YNOV106	*vid22Δ*	2	0.42
YNO107	*vid22Δ*	3	0.02
YNOV125	*vid22Δ*	4	0.29
YGE377	*vid22Δ*	5	0.27
YGE384	*vid22Δ*	6	0.47
YGE850	*vid22Δ*	7	0.57
YGE851	*vid22Δ*	8	0.62
YGE852	*vid22Δ*	9	0.44
YGE853	*vid22Δ*	10	0.89
YGE854	*vid22Δ*	11	0.33

Figure 3. Genome sequence analysis of *vid22Δ* mutants reveals preferential genome instability at G4-rich loci. (**A**) The intersection of predicted G4 elements (17) with small (less than 2 Kb) copy number alterations detected by genome sequencing of eleven *vid22Δ* mutants. The overlap of chromosome coordinates of small CNVs detected in genome shotgun sequences of *vid22Δ* mutants and G4 elements in the yeast genome predicted by (17) are plotted as a Venn diagram. The indicated *P*-value of the intersection is from the Fisher's exact test. (**B**) Comparison of the distribution of expected an observed distance of CNV predictions from G4 elements. Frequency distribution of observed distance of predicted CNV from G4 elements is represented in blue. The expected distance distribution, estimated by 1000 independent random resampling of a matched number of genomic windows of identical size, is represented in red. Distances in base pairs (bp) are represented on the X axis, frequencies on the Y axis. (**C**) The inferred mtDNA copy number of eleven *vid22Δ* mutant strains. The copy number of mtDNA relative to the wild-type strain was inferred by comparing ratios of mitochondrial read counts, using quantile normalization and GC content normalization.

was indeed caused by the G4 structure, we mutagenized the G4 motif by substituting two GGGs with GCGs (Figure 5C), maintaining the same GC content but reducing the sequence propensity to form G4 structures. When the G4 mutated motif was introduced at the *PRB1* locus, we did not observe a statistically supported difference in GCRs between the *vid22Δ* and wild-type strain (Figure 5B), indicating that the chromosomal instability observed in the absence of *VID22* strain is strongly related to the presence of G4 structures.

We then tested the sensitivity of the *vid22Δ* mutant to TMPyP4, a widely used G4-binding molecule that causes stabilization of these nucleic acid structures (36,102). As shown in Figure 5D, deletion of *VID22* significantly enhances cell sensitivity to the G4-ligand, strongly corroborating the hypothesized role of Vid22 in controlling G4 stability.

In yeast cells, a Vid22 paralog, Env11, is present (29). Since frequently paralogs exhibit partially overlapping functions, we tested whether Env11, which also interacts with Vid22, is similarly implicated in preserving genome stability at G4-prone regions. In the *env11Δ* mutant, the G4 dependent GCR rate is indistinguishable from that of a wild type (Supplementary Figure S4A). Moreover, *ENV11* deletion does not affect the sensitivity of the *vid22Δ* mutant to replication stress agents as Hydroxyurea or MMS (Supplementary Figure S4B and (34)), suggesting that Vid22 plays a specific role in maintaining genome stability, independently of *ENV11*.

Vid22 binds to and controls the stability of G4 regions

To determine the *in vivo* sites that were bound by Vid22, we performed a Vid22 ChIP-seq analysis. A total of 413 ChIP-

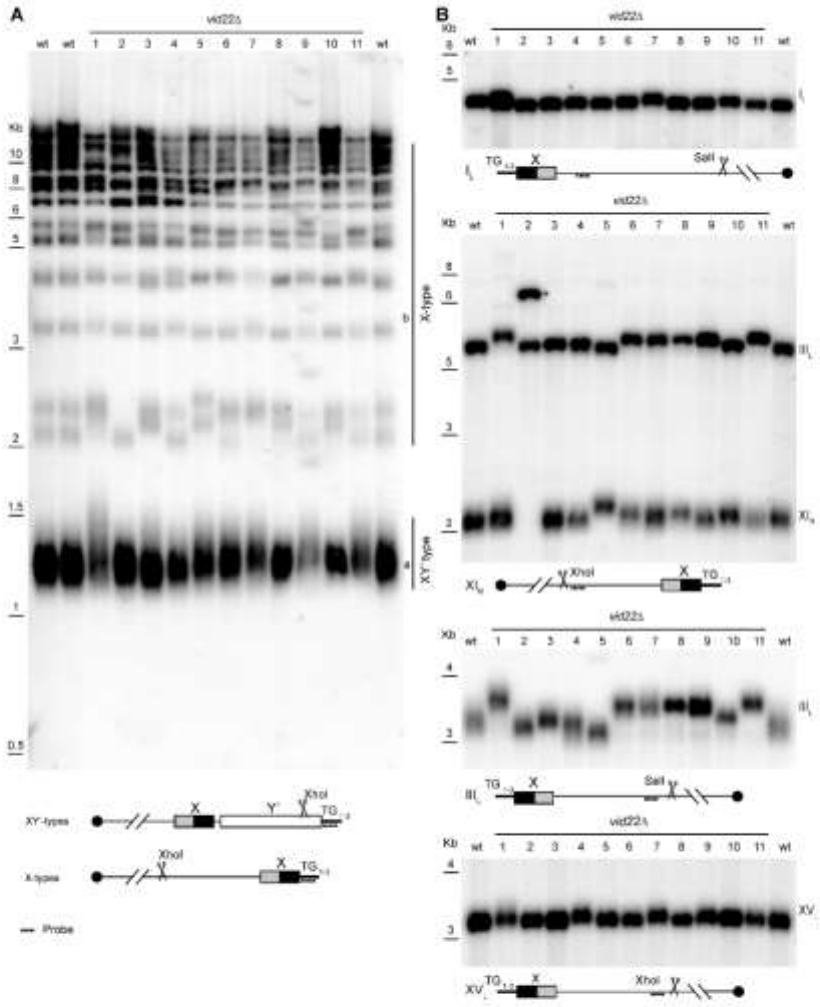


Figure 4. Vid22 is important for the maintenance of telomere homeostasis. Genomic DNA was prepared from wild-type and the eleven independent *vid22*Δ strains. DNA was digested with XhoI or SalI, fractionated on agarose gel, and hybridized with the indicated probes. (**A**) Southern blot and schematic for XY'-type telomeres. (**B**) Southern blot and schematic for 4 different X-type telomeres: I_L, XI_R, III_L, XV_L. The probe for the XI_R telomere cross-reacts with telomere III_L (as described in (56)). The band corresponding to XI_R of clone number 2 is indicated with *.

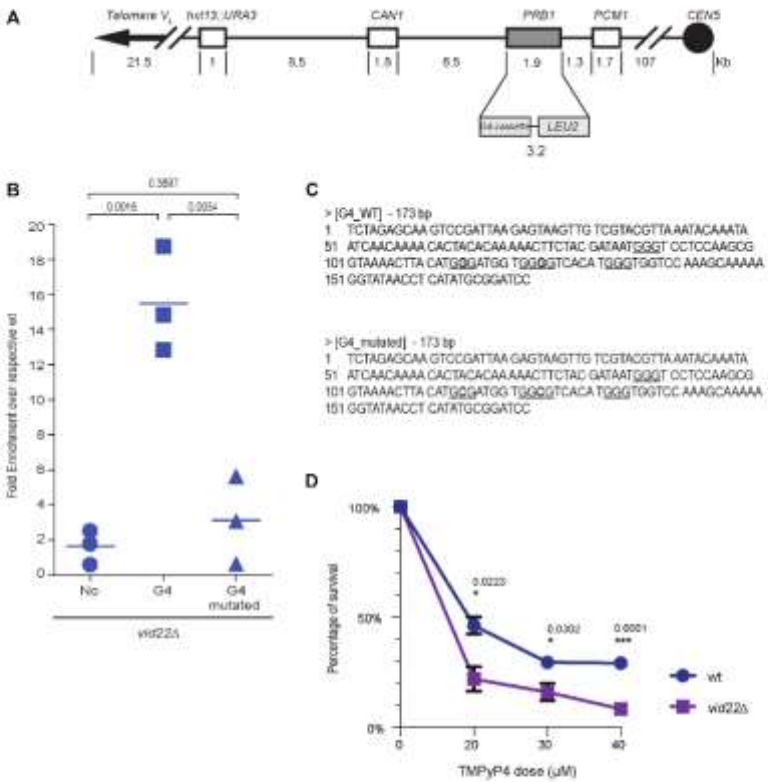


Figure 5. Gross chromosomal rearrangements and cell lethality induced by G4 DNA in *vid22Δ*. (**A**) Schematic representation of the left arm of Chr V in strains used for GCR assay. The G4 or the G4-mutated cassettes were inserted at the *PRB1* locus. *PCM1* is the essential gene nearest to the left telomere; *URA3* and *CAN1* are the two genetic markers used to select for chromosome arm loss or interstitial deletions. (**B**) The plot represents the fold enrichment obtained with GCR experiments. The fold enrichment is the ratio of the GCR rate between *vid22Δ* and wild type containing the same cassette. Each data point is from an independent fluctuation test, with $n \geq 3$ for each strain. The horizontal bars indicate the mean CGR rate for each strain ($N = 3$ independent experiments). An unpaired Student's *t*-test was used to compare the means of measurements and the p-value is indicated. (**C**) Partial sequences of the two cassettes inserted at *PRB1* locus. The underlined Gs are essential for G4-DNA formation; bold Gs in G4-mutated sequence were substituted with C to abolish G4 formation. (**D**) Survival curve after TMPyP4 treatment. Wild-type and *vid22Δ* strains were treated for 2 hours with TMPyP4 at the indicated concentration and plated on YEPD. The percentage of survival was reported ($N = 3$ independent experiments). An unpaired Student's *t*-test was used to compare the means of measurements and the p-value is indicated.

seq peaks were recovered, and in agreement with data published in Styles *et al.* (35) they are preferentially associated with gene promoters (Fisher Exact test *P*-value 1.49E–07) and significantly enriched (~ 2 fold) in the proximity (0 and 200 bp) of predicted G4 elements (Fisher exact test *P*-value 1.03E–06) (Supplementary Table S11 and Supplementary Figure S5). Importantly, the levels of enrichment in G4 elements were consistent across different types of annotations (TSS, promoters, exons) (Supplementary Figure S5A), suggesting that the preferential co-localization of Vid22 with G4 is not associated with specific genomic features.

Analyzing genomic sequencing data described above, we could identify the rearrangement hotspots in *vid22Δ* cells. We found that the *SKN7* locus, located on Chr VIII, is altered in 4 of 11 strains sequenced (Supplementary Table S7 and scheme Figure 6A). Consistently with our data, the region located upstream of *SKN7* is predicted to form G4 structures ((17) and Supplementary Figure S6A) and was enriched in our ChIP-seq analysis (Supplementary Table S11) and in Styles *et al.* (35).

To confirm experimentally the ability of the *SKN7* upstream sequence to form G4 structures *in vitro*, we

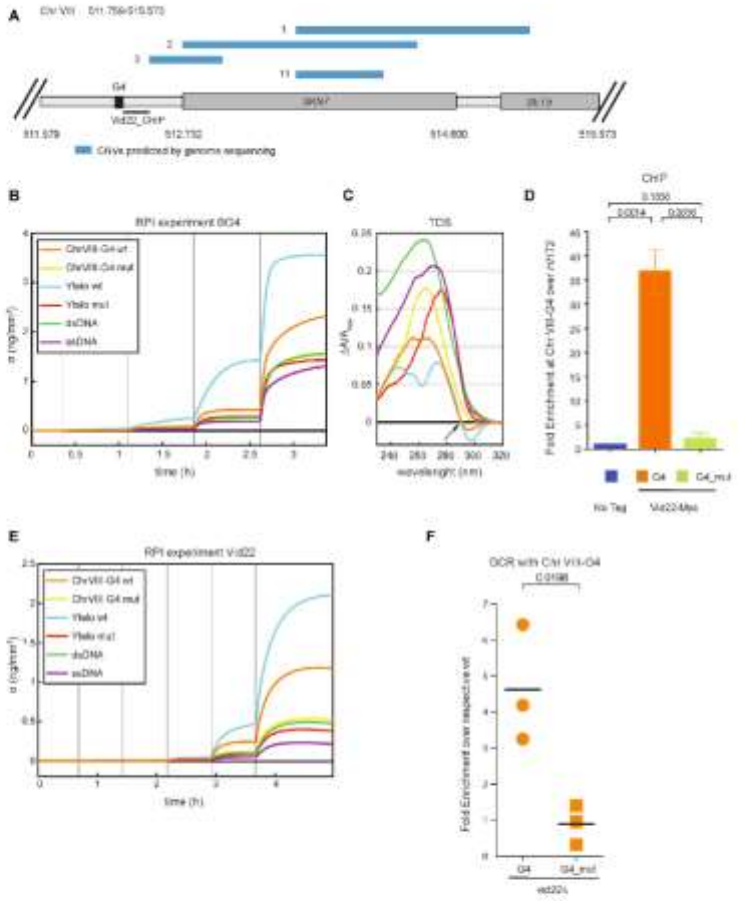


Figure 6. Vid22 binds to and controls the stability of Chr VIII-G4 region. (**A**) Graphical representation of genomic features associated with the Chr VIII *511.759–515.573* genomic locus. A G4 element predicted by (17) is indicated. Copy Number Variations (CNVs) detected at this region in *vid22Δ* strains in this study are indicated with blue bars. The region analyzed by qPCR for ChIP is indicated with a solid black line. (**B**) *In vitro* ability of the predicted Chr VIII-G4 sequence to form G4 structure. Analysis of the interaction of the BG4 antibody with different DNA sequences using the Reflective Phantom Interface (RPI) technique: control G4-forming sequence (Ytelo) and the mutated form Ytelo mut (123), unrelated ssDNA or dsDNA sequences. ChrVIII-G4 and ChrVIII-G4 mut. Increase of the surface density (σ) was measured over time upon antibody binding for the rising concentration of BG4 in solution ($T = 25°C$, and $c = 10$ μM). The vertical lines mark the time where BG4 concentration was increased step-wise from 0 to 10 nM. (**C**) Thermal difference spectra (TDS) of sequences used in the RPI experiment. Each spectrum is computed subtracting the absorbance spectrum at 20°C from the absorbance spectrum at 90°C and then normalized for the maximum amplitude at low temperature to allow a direct comparison among the different sequences. A neat isosbestic point associated with G-quadruplex forming species at 290nm is present (black arrow) both in Ytelo and ChrVIII-G4, while being absent in every other species studied. Y axis is stretched to fit the high TDS amplitude of ChrVIII-G4 mut, ssDNA and dsDNA. (**D**) ChIP-qPCR of Vid22 at the G4 predicted region (17) at the Chr VIII *SKN7* locus. ChIP was performed in wild type (No Tag), Vid22-13Myc (G4) and Vid22-13Myc harbouring two point mutations in the G4 predicted region (G4-mutated). Fold enrichment of Vid22 at Chr VIII was calculated relative to the internal standard *HHT2*. Data are represented as mean ± SEM of $N = 3$ independent experiments. (**E**) Analysis of interaction of Vid22 with DNA forming G4 structures using the Reflective Phantom Interface (RPI) as reported in panel B. Vid22 concentration was increased step-wise from 0 to 50 nM. (**F**) The plot represents the fold enrichment obtained with GCR assays in which *PRB1* locus was substituted with Chr VIII region (Figure 5A). The fold enrichment is the ratio of the GCR rate between *vid22Δ* and wild-type strain containing the same cassette. Each data point is from an independent fluctuation test, with $n \geq 3$ for each strain. The horizontal bars indicate the mean GCR rate for each strain ($N = 3$ independent experiments). (D–F) An unpaired Student's *t*-test was used to compare the means of measurements and the p-value is indicated.

analyzed the conformation of synthetic DNA sequences by UV spectrophotometric analysis and monitored binding of the G4-specific BG4 antibody by Reflective Phantom Interface (RPI). We compared a wild-type *SKN7* sequence (ChrVIII-G4 wt) to a mutant *SKN7* (ChrVIII-G4 mut) that was altered in its predicted capacity to form a G4 structure. As controls, we included a yeast telomeric sequence (Ytelo wt), known to form G4s, a mutated Ytelo, a ssDNA sequence and a dsDNA sequence. The results shown in Figure 6B-C and Supplementary Table S5, demonstrate that Ytelo wt and ChrVIII-G4 wt do indeed form G4 structures, as revealed by the absorbance spectrum and by the binding of the BG4 antibody, as detected by RPI analysis. Both features were lost in the mutated sequences.

We next directly tested *in vivo* binding of Vid22 at the *SKN7* genomic site by ChIP analysis and found that Chr VIII-G4 nt 512387–512565 is enriched in Vid22 chromatin immunoprecipitates (Figure 6D). Vid22 binding to Chr VIII-G4 *in vivo* is dependent on the ability of the target sequence to form G4 structures. Indeed, when we converted the chromosomal G4-forming sequence to a mutated version that loses the ability to form G4s (Supplementary Figure S6A and Figure 6B-C), binding of Vid22, as measured by ChIP, was also lost (Figure 6D).

It is important to note that two consensus sites for Tbf1 binding (TAGGG) (30) are present in proximity of the *SKN7* G4-forming sequence (Supplementary Figure S6A). Thus, we cannot exclude that the observed Vid22 chromatin enrichment might depend on the binding of Tbf1 to these sites. However, mutation of both Tbf1 consensus sites in TA nucleotides ((103) and Supplementary Figure S6A) does not significantly alter the ability of Vid22 to bind this region, which is still totally dependent on the G4 structure (Supplementary Figure S6B).

By ChIP, we excluded that the observed Vid22 binding is mediated by the Sgs1 protein (Supplementary Figure S6C), as was previously observed for other loci (35).

To verify that Vid22 can directly bind G4 structures *in vitro*, we analyzed by RPI the binding of recombinant Vid22 purified from *E. coli* to the synthetic sequences described above. As shown in Figure 6E, Vid22 shows a strong affinity to Ytelo and *SKN7* G4-forming sequences; the binding depends on the sequences' ability to fold into G4, as it is lost when Ytelo or SKN7 are mutated. Surprisingly, Vid22 direct binding to G4s is not mediated by the protein's BED domain as the interaction is not altered by mutation of the conserved CCHH signature (32) (Supplementary Figure S6D).

To verify that Vid22 binding to this G4-forming sequence is important to protect Chr VIII from GCRs, we introduced the wild-type or mutated *SKN7* G4 region into the *PRB1* locus in the GCR reporter strain and measured GCR rates in wild-type and *vid22Δ* strains. Deletion of *VID22* induces nearly five times more rearrangements in proximity to the *SKN7* G4 region compared to a *VID22* wild-type strain and this instability is dependent upon G4 formation, similarly to what was observed for Vid22 binding (Figure 6F). Overall, these data indicate that Vid22 binds regions that are likely to form physiological G4-DNA structures and counteracts G4-induced genome instability.

DISCUSSION

Genomic instability is a hallmark of cancer (9,11,104) and it has been proposed to act as a key driving force in tumorigenesis (12,104,105). Given the evolutionary conservation of the pathways that preserve DNA integrity, yeast research has provided valuable insights into the regulation of genome stability in humans. Multiple screens in *S. cerevisiae* have identified crucial players in genome stability maintenance (80,82,106–114). Nonetheless, the results of these screens are only partially overlapping, likely due to the different approaches used, and each study provides a distinct perspective of the genome integrity network. To identify new genes implicated in protecting cells from endogenous DNA damage, we used a modified version of the SGA technology (79), and screened the *S. cerevisiae* deletion collection for mutants that exhibited chronic accumulation of spontaneous DNA damage. The strategy was based on the observation that overexpression of the cell cycle protein Ddc2 in cells experiencing DNA damage causes hyperactivation of the DNA damage checkpoint and impaired growth ((42) and Figure 1A). We exploited the sensitivity to Ddc2 overexpression as a readout for the spontaneous accumulation of DNA lesions that activate the DNA damage checkpoint, thus broadening the range of genomic instability marks detected in a single screen.

In our screens, we identified two genes with reported synthetic dosage growth defects when combined with *DDC2* overexpression and 50 additional genes not previously known to have SDL interactions with *DDC2*. The 52 genes identified in our screens included 18 genes known to associate with genome instability phenotypes, and were enriched for genetic interactions in the DNA replication and DNA repair neighbourhood of the genetic interaction cell map (45), indicating that genes displaying an SDL interaction with *DDC2* are likely to be involved in genome maintenance.

Among the identified genes identified, we characterized *VID22* (*YLR373C*). Originally reported to have a vesicle trafficking function (87,115), it was later found in two independent screens for genome instability causing DNA mutations (35,89) and was implicated in DNA repair (34,35). Direct observation of DNA damage response markers in a *vid22Δ* strain indicates that *VID22* plays a relevant role in protecting genomic DNA from spontaneous damage: in particular, constitutive Ddc2 and Rad53 phosphorylation (Figure 1F) in the absence of Vid22 are indicative of chronic accumulation of DNA lesions (77,116). The involvement of *VID22* in preserving genome stability is also supported by a recent genomic study (117). The continuous exposure to spontaneous DNA damage probably underlies the remarkably high frequency of gross chromosomal rearrangements observed in *vid22Δ* strains (Figure 2 and S2B). Moreover, analysis of genomic sequencing data revealed that these chromosomal aberrations are enriched in regions displaying a high propensity to form G4 structures, such as telomeres and mitochondrial DNA (Figures 3 and 4 and Supplementary Table S7). This evidence is in agreement with data indicating that Vid22 binds in the proximity of predicted G4 structures (Figure 3B, Supplementary Figure S4, Table

S11 and (35)). Overall, these observations suggest a new role for Vid22 in preventing G4s from generating chromosomal aberrations.

This hypothesis is corroborated by the fact that chromosomal rearrangements at G4-prone sequences occur 15 times more frequently in the absence of *VID22* than in a wild-type background. Conversely, the GCR rate in *vid22Δ* is comparable to the wild type when the G4 motifs are absent or mutated (Figures 5B and 6F). Furthermore, loss of Vid22 sensitizes cells to the G4-stabilizing ligand TMPyP4 (Figure 5D). From the sequencing of *vid22Δ* strains genomes, we identified the *SKN7* locus on chromosome VIII as a mutational hotspots. Indeed the *SKN7* sequence is altered in 4 of 11 independent *vid22Δ* clones. The *SKN7* locus contains a G4-forming sequence that is responsible for the instability in *vid22Δ* cells. Indeed, removal of the G4 structure by point mutations suppresses the instability of the locus. Indeed, when the G4-forming sequence from *SKN7* is inserted in the *PRB1* site of the GCR reporter strain, GCRs increase nearly five-fold if *VID22* is lost, while if the G4 sequence is mutated, the GCR rate remains similar to the wild type (Figure 6F). This G4-forming sequence at *SKN7* promoter is also enriched by Vid22 ChIP and this binding is dependent on its ability to form G4, while it is independent on the presence of Tbf1-binding sites.

Interestingly, our Vid22 ChIP-seq analysis, revealed the enrichment of several additional regions that have no apparent relationship to G4s (Supplementary Figure S5), suggesting that Vid22 could bind at different genomic regions using different strategies and possibly different partners.

In the absence of *VID22*, unresolved G4-DNA structures could induce recombination intermediates, which are generally processed by specific enzymes such as Mus81 or Sgs1/Top3. Consistently, *vid22Δ* cells exhibit an increased recombination rate and are synthetic lethal/sick with *mus81Δ*, *mms4Δ* and *sgs1Δ* mutants (35,45,118). Partial or defective resolution of recombination intermediates could explain the frequent chromosomal aberrations observed in *vid22Δ* strains. Moreover, if these structures involve large chromosomal regions, their persistence could result in incorrect chromosome segregation leading to aneuploidy, which was observed in 6 of 11 *vid22Δ* strains tested (Figure 2). Curiously, we observed seven cases of whole chromosome duplication, mostly of chromosomes III, XI, and XIII. One possible explanation is that the stabilization of intermolecular-G4s between sister chromatids in the absence of *VID22* could alter chromosome segregation allowing aneuploidy. These chromosomes could have more of these structures, resulting in more susceptible alterations.

Of note, it has been shown that stabilization of G4 structures can increase the levels of DNA:RNA hybrids proximal to the G4 region, elevating genome instability (119). It is possible that, in *vid22Δ* cells, increased levels of R-loops could stimulate genomic rearrangements. However, the deletion of *VID22* does not alter the level of DNA:RNA hybrids either at the G4-prone region at the *SKN7* locus nor at the *GCN4* locus, which was reported to be enriched in DNA:RNA hybrids in several strains that accumulate stable R-loops (120) (Supplementary Figure S7). We infer from these data that the mechanism through which Vid22 protects G4-containing genomic regions from instability is not by restricting DNA:RNA hybrid accumulation.

Here, we suggest that Vid22 is able to bind G4 forming sequences both *in vitro* and *in vivo* (Figure 6D-E). Nevertheless, additional investigations are required to understand whether it contains an enzymatic activity to resolve G4 DNAs, or if it acts as a platform for specialized helicases. The existence of a genetic interactions with *SGS1* (35), which encodes a helicase capable of unwinding G4 structures, suggests that Vid22 might also contribute to G4 resolution in a distinct pathway.

The role we uncovered for *VID22* in preventing G4-dependent genome instability could have a great impact on our understanding of cancer biology. Indeed, there is a close correlation between G4 and cancer cells. Unresolved G4 structures can induce more DNA breaks and alter DNA replication, transcription and telomeric structures (121,122) severely increasing genome instability. Moreover, it is known that cancer cells present more stable G4-structures than normal cells (26). While a mammalian *VID22* orthologue has not been identified, at least eight human proteins contain a similar domain organization, that could be evolutionarily related and possess a homologous role.

DATA AVAILABILITY

Raw sequencing data for the 11 *vid22* mutants and one control *S. cerevisiae* BY4741 strains has been deposited in the Short Read Archive (SRA) under Bioproject PRJNA646604. Data can be accessed through the following link: https://www.ncbi.nlm.nih.gov/sra/PRJNA646604.

The Vid22 ChIP-seq dataset analyzed during the current study has been deposited at the Gene Expression Omnibus (GEO) under the accession number GSE174688. Data can be accessed through the following link https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE174688.

All the data, including different annotations of G4 elements, Vid22 and Tbf1 peaks, and predictions of genomic rearrangements can be visualized and queried along with the complete annotation of the sacCer3 assembly of the *S. cerevisiae* genome at https://genome.ucsc.edu/s/pantaleoM/sacCer3_MC_NAR.

To facilitate the navigation through this large amount of data, we also report a series of genomic coordinates/loci which provide sensible examples of the main findings of this work:

chrVIII:511 823–515 376
chrII:4877–11 025
chrII:756 777–762 925
chrIV:1 523 402–1 525 800
chrVIII:561 295–562 643
chrXII:4927–8975
chrXII:450 697–468 827
chrXIV:6486–6948
chrXIV:782 969–784 208

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Charlie Boone and members of the Boone lab for assistance with the Ddc2 SDL screens. We thank all members of the genome instability and human pathologies lab for critical discussions; C. Santocanale for the Rad53 antibodies; V.A. Zakian for providing us strains and plasmids. pSANG10-3F-BG4 was a gift from Shankar Balasubramanian (Addgene plasmid # 55756; http://n2t.net/addgene:55756; RRID:Addgene_55756).

FUNDING

Associazione Italiana per la Ricerca sul Cancro (AIRC) [15631, 21806 to M.M.F.]; MIUR [PRIN 2015-2015SJLMB9; PRIN 2017-2017KSZZJW to M.M.F.]; Telethon [GGP15227 to M.M.F.]; F.L. was supported by the University of Milano: "Piano di Sviluppo dell'Ateneo per la Ricerca. Linea B: Supporto per i Giovani Ricercatori"; M.C.B. was supported by Fondazione Veronesi; Research at the laboratory of A.A. was funded by the Spanish Ministry of Economy and Competitiveness [BFU2016-75058-P]; B.G.G. was funded by the Spanish Association Against Cancer; MIUR [PRIN2017-2017Z55KC to T.B.]; M.C., D.S.H. are supported by MIUR [PRIN 2017] and CNRbiomics [PIR01_00017]; H2020 Projects ELIXIR-EXCELERATE, EOSC-Life, EOSC-Pillar and Elixir-IIB; G.W.B. was supported by the Canadian Institutes of Health Research [FDN-159913]. Funding for open access charge: Associazione Italiana per la Ricerca sul Cancro (AIRC) [21806].

Conflict of interest statement. None declared.

REFERENCES

1. Lindahl,T. (1993) Instability and decay of the primary structure of DNA. *Nature*, **362**, 709–715.
2. Hoeijmakers,J.H.J. (2009) DNA damage, aging, and cancer. *N. Engl. J. Med.*, **361**, 1475–1485.
3. Bartek,J. and Lukas,J. (2007) DNA damage checkpoints: from initiation to recovery or adaptation. *Curr. Opin. Cell Biol.*, **19**, 238–245.
4. Suijkerbuijk,S.J.E. and Kops,G.J.P.L. (2008) Preventing aneuploidy: the contribution of mitotic checkpoint proteins. *Biochim. Biophys. Acta - Rev. Cancer*, **1786**, 24–31.
5. Lazzaro,F., Giannattasio,M., Puddu,F., Granata,M., Pellicioli,A., Plevani,P. and Muzi-Falconi,M. (2009) Checkpoint mechanisms at the intersection between DNA damage and repair. *DNA Repair (Amst.)*, **8**, 1055–1067.
6. Ciccia,A. and Elledge,S.J. (2010) The DNA damage response: making it safe to play with knives. *Mol. Cell*, **40**, 179–204.
7. Lanz,M.C., Dibitetto,D. and Smolka,M.B. (2019) DNA damage kinase signaling: checkpoint and repair at 30 years. *EMBO J.*, **38**, e101801.
8. Aguilera,A. and Gómez-González,B. (2008) Genome instability: a mechanistic view of its causes and consequences. *Nat. Rev. Genet.*, **9**, 204–217.
9. Lengauer,C., Kinzler,K.W. and Vogelstein,B. (1998) Genetic instabilities in human cancers. *Nature*, **396**, 643–649.
10. Halazonetis,T.D., Gorgoulis,V.G. and Bartek,J. (2008) An oncogene-induced DNA damage model for cancer development. *Science*, **319**, 1352–1355.
11. Hanahan,D. and Weinberg,R.A. (2011) Hallmarks of cancer: the next generation. *Cell*, **144**, 646–674.
12. Tubbs,A. and Nussenzweig,A. (2017) Endogenous DNA damage as a source of genomic instability in cancer. *Cell*, **168**, 644–656.
13. Saini,N., Zhang,Y., Usdin,K. and Lobachev,K.S. (2013) When secondary comes first – the importance of non-canonical DNA structures. *Biochimie*, **95**, 117–123.
14. Gellert,M., Lipsett,M.N. and Davies,D.R. (1962) Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. U.S.A.*, **48**, 2013–2018.
15. Sen,D. and Gilbert,W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, **334**, 364–366.
16. Burge,S., Parkinson,G.N., Hazel,P., Todd,A.K. and Neidle,S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.*, **34**, 5402–5415.
17. Capra,J.A., Paeschke,K., Singh,M. and Zakian,V.A. (2010) G-Quadruplex DNA sequences are evolutionarily conserved and associated with distinct genomic features in *Saccharomyces cerevisiae*. *PLoS Comput. Biol.*, **6**, e1000861.
18. Maizels,N. (2006) Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nat. Struct. Mol. Biol.*, **13**, 1055–1059.
19. Sun,H., Karow,J.K., Hickson,I.D. and Maizels,N. (1998) The Bloom's syndrome helicase unwinds G4 DNA. *J. Biol. Chem.*, **273**, 27587–27592.
20. Fry,M. and Loeb,L.A. (1999) Human Werner syndrome DNA helicase unwinds tetrahelical structures of the fragile X syndrome repeat sequence d(CGG)n. *J. Biol. Chem.*, **274**, 12797–12802.
21. Paeschke,K., Bochman,M.L., Garcia,P.D., Cejka,P., Friedman,K.L., Kowalczykowski,S.C. and Zakian,V.A. (2013) Pif1 family helicases suppress genome instability at G-quadruplex motifs. *Nature*, **497**, 458–462.
22. Mendoza,O., Bourdoncle,A., Boulé,J.-B., Brosh,R.M. and Mergny,J.-L. (2016) G-quadruplexes and helicases. *Nucleic Acids Res.*, **44**, 1989–2006.
23. Sun,H., Bennett,R.J. and Maizels,N. (1999) The *Saccharomyces cerevisiae* Sgs1 helicase efficiently unwinds G-G paired DNAs. *Nucleic Acids Res.*, **27**, 1978–1984.
24. Rhodes,D. and Lipps,H.J. (2015) G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res.*, **43**, 8627–8637.
25. Huppert,J.L. and Balasubramanian,S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.*, **35**, 2105–2105.
26. Hänsel-Hertsch,R., Beraldi,D., Lensing,S. V., Marsico,G., Zyner,K., Parry,A., Di Antonio,M., Pike,J., Kimura,H., Narita,M. *et al.* (2016) G-quadruplex structures mark human regulatory chromatin. *Nat. Genet.*, **48**, 1267–1272.
27. Siddiqui-Jain,A., Grand,C.L., Bearss,D.J. and Hurley,L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. U.S.A.*, **99**, 11593–11598.
28. Cogoi,S. and Xodo,L.E. (2006) G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.*, **34**, 2536–2549.
29. Krogan,N.J., Cagney,G., Yu,H., Zhong,G., Guo,X., Ignatchenko,A., Li,J., Pu,S., Datta,N., Tikuisis,A.P. *et al.* (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature*, **440**, 637–643.
30. Preti,M., Ribeyre,C., Pascali,C., Bosio,M.C., Cortelazzi,B., Rougemont,J., Guarnera,E., Naef,F., Shore,D. and Dieci,G. (2010) The telomere-binding protein Tbf1 demarcates snoRNA gene promoters in *Saccharomyces cerevisiae*. *Mol. Cell*, **38**, 614–620.
31. Huh,W.-K., Falvo,J. V, Gerke,L.C., Carroll,A.S., Howson,R.W., Weissman,J.S. and O'Shea,E.K. (2003) Global analysis of protein localization in budding yeast. *Nature*, **425**, 686–691.
32. Aravind,L. (2000) The BED finger, a novel DNA-binding domain in chromatin-boundary-element-binding proteins and transposases. *Trends Biochem. Sci.*, **25**, 421–423.
33. Nowotny,M. (2009) Retroviral integrase superfamily: the structural perspective. *EMBO Rep.*, **10**, 144–151.
34. Bonetti,D., Anbalagan,S., Lucchini,G., Clerici,M. and Longhese,M.P. (2013) Tbf1 and Vid22 promote resection and non-homologous end joining of DNA double-strand break ends. *EMBO J.*, **32**, 275–289.
35. Styles,E.B., Founk,K.J., Zamparo,L.A., Sing,T.L., Altintas,D., Ribeyre,C., Ribaud,V., Rougemont,J., Maybew,D., Costanzo,M. *et al.* (2016) Exploring quantitative yeast phenomics with single-cell analysis of DNA damage foci. *Cell Syst.*, **3**, 264–277.

36. Yamashita,T., Uno,T. and Ishikawa,Y. (2005) Stabilization of guanine quadruplex DNA by the binding of porphyrins with cationic side arms. *Bioorg. Med. Chem.*, **13**, 2423–2430.
37. Baker Brachmann,C., Davies,A., Cost,G.J., Caputo,E., Li,J., Hieter,P. and Boeke,J.D. (1998) Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: a useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast*, **14**, 115–132.
38. Sikorski,R.S. and Hieter,P. (1989) A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. *Genetics*, **122**, 19–27.
39. Longtine,M.S., Mckenzie,A. III, Demarini,D.J., Shah,N.G., Wach,A., Brachat,A., Philippsen,P. and Pringle,J.R. (1998) Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast*, **14**, 953–961.
40. Storici,F., Lewis,L.K. and Resnick,M.A. (2001) *In vivo* site-directed mutagenesis using oligonucleotides. *Nat. Biotechnol.*, **19**, 773–776.
41. Lazzaro,F., Novarina,D., Amara,F., Watt,D.L., Stone,J.E., Costanzo,V., Burgers,P.M., Kunkel,T.A., Plevani,P. and Muzi-Falconi,M. (2012) RNase H and postreplication repair protect cells from ribonucleotides incorporated in DNA. *Mol. Cell*, **45**, 99–110.
42. Clerici,M. (2001) Hyperactivation of the yeast DNA damage checkpoint by *TEL1* and *DDC2* overexpression. *EMBO J.*, **20**, 6485–6498.
43. Baryshnikova,A., Costanzo,M., Kim,Y., Ding,H., Koh,J., Toufighi,K., Youn,J.-Y., Ou,J., San Luis,B.-J., Bandyopadhyay,S. *et al.* (2010) Quantitative analysis of fitness and genetic interactions in yeast on a genome scale. *Nat. Methods*, **7**, 1017–1024.
44. Baryshnikova,A. (2016) Systematic functional annotation and visualization of biological networks. *Cell Syst.*, **2**, 412–421.
45. Costanzo,M., VanderSluis,B., Koch,E.N., Baryshnikova,A., Pons,C., Tan,G., Wang,W., Usaj,M., Hanchard,J., Lee,S.D. *et al.* (2016) A global genetic interaction network maps a wiring diagram of cellular function. *Science*, **353**, aaf1420.
46. Giannattasio,M., Lazzaro,F., Plevani,P. and Muzi-Falconi,M. (2004) DNA decay and limited Rad53 activation after liquid holding of UV-treated nucleotide excision repair deficient *S. cerevisiae* cells. *DNA Repair (Amst.)*, **3**, 1591–1599.
47. Ferrari,M., Dibitetto,D., De Gregorio,G., Eapen,V. V., Rawal,C.C., Lazzaro,F., Tsabar,M., Marini,F., Haber,J.E. and Pellicioli,A. (2015) Functional interplay between the 53BP1-ortholog Rad9 and the Mre11 complex regulates resection, end-tethering and repair of a double-strand break. *PLoS Genet.*, **11**, e1004928.
48. Blecher-Gonen,R., Barnett-Itzhaki,Z., Jaitin,D., Amann-Zalcenstein,D., Lara-Astiaso,D. and Amit,I. (2013) High-throughput chromatin immunoprecipitation for genome-wide mapping of in vivo protein-DNA interactions and epigenomic states. *Nat. Protoc.*, **8**, 539–554.
49. Langmead,B., Trapnell,C., Pop,M. and Salzberg,S.L. (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.*, **10**, R25.
50. Li,H., Handsaker,B., Wysoker,A., Fennell,T., Ruan,J., Homer,N., Marth,G., Abecasis,G., Durbin,R., Project,G. *et al.* (2009) The Sequence Alignment/Map format and SAMtools. *Bioinforma. Appl. NOTE*, **25**, 2078–2079.
51. Zhang,Y., Liu,T., Meyer,C.A., Eeckhoute,J., Johnson,D.S., Bernstein,B.E., Nussbaum,C., Myers,R.M., Brown,M., Li,W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
52. Fermi,B., Bosio,M.C. and Dieci,G. (2016) Promoter architecture and transcriptional regulation of Abf1-dependent ribosomal protein genes in *Saccharomyces cerevisiae*. *Nucleic Acids Res.*, **44**, 6113–6126.
53. Pfaffl,M.W. (2001) A new mathematical model for relative quantification in real-time RT–PCR. *Nucleic Acids Res.*, **29**, e45.
54. Giannattasio,M., Follonier,C., Tourrière,H., Puddu,F., Lazzaro,F., Pasero,P., Lopes,M., Plevani,P. and Muzi-Falconi,M. (2010) Exo1 competes with repair synthesis, converts NER intermediates to long ssDNA gaps, and promotes checkpoint activation. *Mol. Cell*, **40**, 50–62.
55. Viscardi,V., Baroni,E., Romano,M., Lucchini,G. and Longhese,M.P. (2003) Sudden telomere lengthening triggers a Rad53-dependent checkpoint in *Saccharomyces cerevisiae*. *Mol. Biol. Cell*, **14**, 3126–3143.
56. Craven,R.J. and Petes,T.D. (1999) Dependence of the regulation of telomere length on the type of subtelomeric repeat in the yeast *Saccharomyces cerevisiae*. *Genetics*, **152**, 1531–1541.
57. Church,G.M. and Gilbert,W. (1984) Genomic sequencing. *Proc. Natl. Acad. Sci. U.S.A.*, **81**, 1991–1995.
58. Bolger,A.M., Lohse,M. and Usadel,B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, **30**, 2114–2120.
59. Marçais,G., Delcher,A.L., Phillippy,A.M., Coston,R., Salzberg,S.L. and Zimin,A. (2018) MUMmer4: a fast and versatile genome alignment system. *PLoS Comput. Biol.*, **14**, e1005944.
60. Koboldt,D.C., Zhang,Q., Larson,D.E., Shen,D., McLellan,M.D., Lin,L., Miller,C.A., Mardis,E.R., Ding,L. and Wilson,R.K. (2012) VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res.*, **22**, 568–576.
61. Venables,W.N. and Ripley,B.D. (2002) In: *Modern Applied Statistics with S*. Springer Science & Business Media.
62. Putnam,C.D., Hayes,T.K. and Kolodner,R.D. (2010) Post-replication repair suppresses duplication-mediated genome instability. *PLoS Genet.*, **6**, e1000933.
63. Radchenko,E.A., McGinty,R.J., Aksenova,A.Y., Neil,A.J. and Mirkin,S.M. (2018) Quantitative analysis of the rates for repeat-mediated genome instability in a yeast experimental system. In: *Methods in Molecular Biology*. Humana Press Inc., Vol. **1672**, pp.421–438.
64. Hall,B.M., Ma,C.-X., Liang,P. and Singh,K.K. (2009) Fluctuation AnaLysis CalculatOR: a web tool for the determination of mutation rate using Luria-Delbruck fluctuation analysis. *Bioinformatics*, **25**, 1564–1565.
65. Tagliabue,G., Faoro,V., Rizzo,S., Shlattero,D., Saccani,A., Riccio,G., Bellini,T., Salina,M., Buscaglia,M. and Marcello,A. (2017) A label-free immunoassay for Flavivirus detection by the reflective phantom interface technology. *Biochem. Biophys. Res. Commun.*, **492**, 558–564.
66. Zilio,C., Bernardi,A., Palmioli,A., Salina,M., Tagliabue,G., Buscaglia,M., Consonni,R. and Chiari,M. (2015) New 'clickable' polymeric coating for glycan microarrays. *Sens. Actuators, B Chem.*, **215**, 412–420.
67. Vanjur,L., Carzaniga,T., Casiraghi,L., Chiari,M., Zanchetta,G. and Buscaglia,M. (2020) Non-Langmuir kinetics of DNA surface hybridization. *Biophys. J.*, **119**, 989–1001.
68. Zanchetta,G., Carzaniga,T., Vanjur,L., Casiraghi,L., Tagliabue,G., Morasso,C., Bellini,T. and Buscaglia,M. (2021) Design of a rapid, multiplex, one-pot miRNA assay optimized by label-free analysis. *Biosens. Bioelectron.*, **172**, 112751.
69. Carzaniga,T., Zanchetta,G., Frezza,E., Casiraghi,L., Vanjur,L., Nava,G., Tagliabue,G., Dieci,G., Buscaglia,M. and Bellini,T. (2021) A bit stickier, a bit slower, a lot stiffer: Specific vs. nonspecific binding of gal4 to dna. *Int. J. Mol. Sci.*, **22**, 3813.
70. Biffi,G., Tannahill,D., McCafferty,J. and Balasubramanian,S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat. Chem.*, **5**, 182–186.
71. Salina,M., Giavazzi,F., Lanfranco,R., Ceccarello,E., Sola,L., Chiari,M., Chini,B., Cerbino,R., Bellini,T. and Buscaglia,M. (2015) Multi-spot, label-free immunoassay on reflectionless glass. *Biosens. Bioelectron.*, **74**, 539–545.
72. Nava,G., Ceccarello,E., Giavazzi,F., Salina,M., Dumin,F., Chiari,M., Buscaglia,M., Bellini,T. and Zanchetta,G. (2016) Label-free detection of DNA single-base mismatches using a simple reflectance-based optical technique. *Phys. Chem. Chem. Phys.*, **18**, 13395–13402.
73. Mergny,J.-L. (2005) Thermal difference spectra: a specific signature for nucleic acid structures. *Nucleic Acids Res.*, **33**, e138.
74. Mergny,J.-L., Phan,A.-T. and Lacroix,L. (1998) Following G-quartet formation by UV-spectroscopy. *FEBS Lett.*, **435**, 74–78.
75. Garcia-Rubio,M., Barroso,S.I. and Aguilera,A. (2018) Detection of DNA-RNA hybrids in vivo. In: *Methods in Molecular Biology*. Humana Press Inc., Vol. **1672**, pp. 347–361.
76. Paciotti,V., Clerici,M., Lucchini,G. and Longhese,M.P. (2000) The checkpoint protein Ddc2, functionally related to *S. pombe* Rad26, interacts with Mec1 and is regulated by Mec1-dependent phosphorylation in budding yeast. *Genes Dev.*, **14**, 2046–2059.

77. Zhang,C., Roberts,T.M., Yang,J., Desai,R. and Brown,G.W. (2006) Suppression of genomic instability by *SLX5* and *SLX8* in *Saccharomyces cerevisiae*. *DNA Repair (Amst.)*., **5**, 336–346.
78. Kroll,E.S., Hyland,K.M., Hieter,P. and Li,J.J. (1996) Establishing genetic interactions by a synthetic dosage lethality phenotype. *Genetics*, **143**, 95–102.
79. Measday,V., Baetz,K., Guzzo,J., Yuen,K., Kwok,T., Sheikh,B., Ding,H., Ueta,R., Hoac,T., Cheng,B. *et al.* (2005) Systematic yeast synthetic lethal and synthetic dosage lethal screens identify genes required for chromosome segregation. *Proc. Natl. Acad. Sci. U.S.A.*, **102**, 13956–13961.
80. Hendry,J.A., Tan,G., Ou,J., Boone,C. and Brown,G.W. (2015) Leveraging DNA damage response signaling to identify yeast genes controlling genome stability. *G3 (Bethesda)*., **5**, 997–1006.
81. Alvaro,D., Lisby,M. and Rothstein,R. (2007) Genome-wide analysis of Rad52 foci reveals diverse mechanisms impacting recombination. *PLoS Genet.*, **3**, 2439–2449.
82. Stirling,P.C., Bloom,M.S., Solanki-Patil,T., Smith,S., Sipahimalani,P., Li,Z., Kofoed,M., Ben-Aroya,S., Myung,K. and Hieter,P. (2011) The complete spectrum of yeast chromosome instability genes identifies candidate CIN cancer genes and functional roles for ASTRA complex components. *PLoS Genet.*, **7**, e1002057.
83. Sharifpoor,S., Van Dyk,D., Costanzo,M., Baryshnikova,A., Friesen,H., Douglas,A.C., Youn,J.Y., VanderSluis,B., Myers,C.L., Papp,B. *et al.* (2012) Functional wiring of the yeast kinome revealed by global analysis of genetic network motifs. *Genome Res.*, **22**, 791–801.
84. Kaluarachchi Duffy,S., Friesen,H., Baryshnikova,A., Lambert,J.P., Chong,Y.T., Figeys,D. and Andrews,B. (2012) Exploring the yeast acetylome using functional genomics. *Cell*, **149**, 936–948.
85. Douglas,A.C., Smith,A.M., Sharifpoor,S., Yan,Z., Durbic,T., Heisler,L.E., Lee,A.Y., Ryan,O., Göttert,H., Surendra,A. *et al.* (2012) Functional analysis with a barcoder yeast gene overexpression system. *G3 Genes Genomes Genet.*, **2**, 1279–1289.
86. Pan,X., Yuan,D.S., Xiang,D., Wang,X., Sookhai-Mahadeo,S., Bader,J.S., Hieter,P., Spencer,F. and Boeke,J.D. (2004) A robust toolkit for functional profiling of the yeast genome. *Mol. Cell*, **16**, 487–496.
87. Brown,C.R., McCann,J.A., Hung,G.G., Elco,C.P. and Chiang,H.-L. (2002) Vid22p, a novel plasma membrane protein, is required for the fructose-1,6-bisphosphatase degradation pathway. *J. Cell Sci.*, **115**, 655–666.
88. Yuen,K.W.Y., Warren,C.D., Chen,O., Kwok,T., Hieter,P. and Spencer,F.A. (2007) Systematic genome instability screens in yeast and their potential relevance to cancer. *Proc. Natl. Acad. Sci. U.S.A.*, **104**, 3925–3930.
89. Putnam,C.D., Srivatsan,A., Nene,R. V., Martinez,S.L., Clotfelter,S.P., Bell,S.N., Somach,S.B., E.S. de Souza,J., Fonseca,A.F., de Souza,S.J. *et al.* (2016) A genetic network that suppresses genome rearrangements in *Saccharomyces cerevisiae* and contains defects in cancers. *Nat. Commun.*, **7**, 11256.
90. Miyata,N., Miyoshi,T., Yamaguchi,T., Nakazono,T., Tani,M. and Kugo,O. (2015) *VID22* is required for transcriptional activation of the *PSD2* gene in the yeast *Saccharomyces cerevisiae*. *Biochem. J.*, **472**, 319–328.
91. Duro,E., Vaisica,J.A., Brown,G.W. and Rouse,J. (2008) Budding yeast Mms22 and Mms1 regulate homologous recombination induced by replisome blockage. *DNA Repair (Amst.)*., **7**, 811–818.
92. Driscoll,R., Hudson,A. and Jackson,S.P. (2007) Yeast Rtt109 promotes genome stability by acetylating histone H3 on Lysine 56. *Science*, **315**, 649–652.
93. Pellicioli,A., Lucca,C., Liberi,G., Marini,F., Lopes,M., Plevani,P., Romano,A., Di Fiore,P.P. and Foiani,M. (1999) Activation of Rad53 kinase in response to DNA damage and its effect in modulating phosphorylation of the lagging strand DNA polymerase. *EMBO J.*, **18**, 6561–6572.
94. Lazzaro,F., Sapountzi,V., Granata,M., Pellicioli,A., Vaze,M., Haber,J.E., Plevani,P., Lydall,D. and Muzi-Falconi,M. (2008) Histone methyltransferase Dot1 and Rad9 inhibit single-stranded DNA accumulation at DSBs and uncapped telomeres. *EMBO J.*, **27**, 1502–1512.
95. Hoose,S.A., Rawlings,J.A., Kelly,M.M., Leitch,M.C., Ababneh,Q.O., Robles,J.P., Taylor,D., Hoover,E.M., Hailu,B., McEnery,K.A. *et al.* (2012) A systematic analysis of cell cycle regulators in yeast reveals that most factors act independently of cell size to control initiation of division. *PLoS Genet.*, **8**, e1002590.
96. Marsico,G., Chambers,V.S., Sahakyan,A.B., Mccauley,P., Boutell,J.M., Antonio,M. Di and Balasubramanian,S. (2019) Whole genome experimental maps of DNA G-quadruplexes in multiple species. *Nucleic Acids Res.*, **47**, 3862–3874.
97. Kudlicki,A.S. (2016) G-Quadruplexes involving both strands of genomic DNA are highly abundant and colocalize with functional sites in the human genome. *PLoS One*, **11**, e0146174.
98. Čutová,M., Manta,J., Porubiaková,O., Kaura,P., Šlaství,J., Jagelská,E.B., Goswami,P., Bartas,M. and Brázda,V. (2019) Divergent distributions of inverted repeats and G-quadruplex forming sequences in *Saccharomyces cerevisiae*. *Genomics*, **112**, 1897–1901.
99. Louis,E.J. (1995) The chromosome ends of *Saccharomyces cerevisiae*. *Yeast*, **11**, 1553–1573.
100. Wellinger,R.J. and Zakian,V.A. (2012) Everything you ever wanted to know about *Saccharomyces cerevisiae* telomeres: beginning to end. *Genetics*, **191**, 1073–1105.
101. Schmidt,K.H., Pennaneach,V., Putnam,C.D. and Kolodner,R.D. (2006) Analysis of gross-chromosomal rearrangements in *Saccharomyces cerevisiae*. In *Methods Enzymol.* **409**, 462–476.
102. Andrew,E.J., Merchan,S., Lawless,C., Banks,A.P., Wilkinson,D.J. and Lydall,D. (2013) Pentose phosphate pathway function affects tolerance to the G-quadruplex binder TMPyP4. *PLoS One*, **8**, e66242.
103. Koering,C.E., Fourel,G., Binet-Brasselet,E., Laroche,T., Klein,F. and Gilson,E. (2000) Identification of high affinity Tbf1p-binding sites within the budding yeast genome. *Nucleic Acids Res.*, **28**, 2519–2526.
104. Loeb,L.A. (2001) A mutator phenotype in cancer. *Cancer Res.*, **61**, 3230–3239.
105. Negrini,S., Gorgoulis,V.G. and Halazonetis,T.D. (2010) Genomic instability — an evolving hallmark of cancer. *Nat. Rev. Mol. Cell Biol.*, **11**, 220–228.
106. Smith,S., Hwang,J.-Y., Banerjee,S., Majeed,A., Gupta,A. and Myung,K. (2004) Mutator genes for suppression of gross chromosomal rearrangements identified by a genome-wide screening in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 9039–9044.
107. Ouspenski,I.I., Elledge,S.J. and Brinkley,B.R. (1999) New yeast genes important for chromosome integrity and segregation identified by dosage effects on genome stability. *Nucleic Acids Res.*, **27**, 3001–3008.
108. Myung,K., Chen,C. and Kolodner,R.D. (2001) Multiple pathways cooperate in the suppression of genome instability in *Saccharomyces cerevisiae*. *Nature*, **411**, 1073–1076.
109. Huang,M.-E., Rio,A.-G., Nicolas,A. and Kolodner,R.D. (2003) A genomewide screen in *Saccharomyces cerevisiae* for genes that suppress the accumulation of mutations. *Proc. Natl. Acad. Sci. U.S.A.*, **100**, 11529–11534.
110. Andersen,M.P., Nelson,Z.W., Hetrick,E.D. and Gottschling,D.E. (2008) A genetic screen for increased loss of heterozygosity in *Saccharomyces cerevisiae*. *Genetics*, **179**, 1179–1195.
111. Duffy,S., Fam,H.K., Wang,Y.K., Styles,E.B., Kim,J.-H., Ang,J.S., Singh,T., Larionov,V., Shah,S.P., Andrews,B. *et al.* (2016) Overexpression screens identify conserved dosage chromosome instability genes in yeast and human cancer. *Proc. Natl. Acad. Sci. U.S.A.*, **113**, 9967–9976.
112. Alvaro,D., Lisby,M. and Rothstein,R. (2007) Genome-wide analysis of Rad52 foci reveals diverse mechanisms impacting recombination. *PLoS Genet.*, **3**, e228.
113. Askree,S.H., Yehuda,T., Smolikov,S., Gurevich,R., Hawk,J., Coker,C., Krauskopf,A., Kupiec,M. and McEachern,M.J. (2004) A genome-wide screen for *Saccharomyces cerevisiae* deletion mutants that affect telomere length. *Proc. Natl. Acad. Sci. U.S.A.*, **101**, 8658–8663.
114. Cheng,E., Vaisica,J.A., Ou,J., Baryshnikova,A., Lu,Y., Roth,F.P. and Brown,G.W. (2012) Genome rearrangements caused by depletion of essential DNA replication proteins in *Saccharomyces cerevisiae*. *Genetics*, **192**, 147–160.
115. Brown,C.R., Cui,D.-Y., Hung,G.G.-C. and Chiang,H.-L. (2001) Cyclophilin A mediates Vid22p function in the import of

fructose-1,6-bisphosphatase into Vid vesicles. *J. Biol. Chem.*, **276**, 48017–48026.
116. Schmidt,K.H. and Kolodner,R.D. (2006) Suppression of spontaneous genome rearrangements in yeast DNA helicase mutants. *Proc. Natl. Acad. Sci. U.S.A.*, **103**, 18196–18201.
117. Piddle,F., Herzog,M., Selivanova,A., Wang,S., Zhu,J., Klein-Lavi,S., Gordon,M., Meirman,R., Millan-Zambrano,G., Ayestaran,I. *et al.* (2019) Genome architecture and stability in the *Saccharomyces cerevisiae* knockout collection. *Nature*, **573**, 416–420.
118. Bellaoui,M., Chang,M., Ou,J., Xu,H., Boone,C. and Grant W.Brown (2003) Elg1 forms an alternative RFC complex important for DNA replication and genome integrity. *EMBO J.*, **22**, 4304–4313.
119. De Magis,A., Manzo,S.G., Russo,M., Marinello,J., Morigi,R., Sordet,O. and Capranico,G. (2019) DNA damage and genome instability by G-quadruplex ligands are mediated by R loops in human cancer cells. *Proc. Natl. Acad. Sci. U.S.A.*, **116**, 816–825.
120. Herrera-Moyano,E., Mergui,X., Garcia-Rubio,M.L., Barroso,S. and Aguilera,A. (2014) The yeast and human FACT chromatin-reorganizing complexes solve R-loop-mediated transcription-replication conflicts. *Genes Dev.*, **28**, 735–748.
121. Piazza,A., Adrian,M., Samazan,F., Heddi,B., Hamon,F., Serero,A., Lopes,J., Teulade-Fichou,M., Phan,A.T. and Nicolas,A. (2015) Short loop length and high thermal stability determine genomic instability induced by G-quadruplex-forming minisatellites. *EMBO J.*, **34**, 1718–1734.
122. Hänsel-Hertsch,R., Di Antonio,M. and Balasubramanian,S. (2017) DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nat. Rev. Mol. Cell Biol.*, **18**, 279–284.
123. Giraldo,R., Suzuki,M., Chapman,L. and Rhodes,D. (1994) Promotion of parallel DNA quadruplexes by a yeast telomere binding protein: a circular dichroism study. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 7658–7662.

SUPPLEMENTARY DATA

Supplementary Table S1. List of the yeast strains and plasmids used in the present work.

Supplementary Table S2. Colony size and SGA scores from the DDC2 overexpression screens.

Supplementary Table S3. GC composition normalization. This table reports the scaling factors applied for the normalization of depth of coverage values based on GC composition. Bins of %GC composition are indicated in the first column. Corresponding normalization factors are reported in a separate column for every strain

Supplementary Table S4. Estimates of levels of sensitivity of CNV detection. For strains showing a complete chromosomal duplication, levels of sensitivity of our in silico CNV detection assay were estimated as the ratio between the total number of overlapped genomic windows associated to duplicated chromosomes and the total number of windows showing a significantly altered coverage according to our statistical test. Strains are indicated in the first column. Estimates of sensitivity before and after the application of the GC normalization correction are reported on the second column and third column respectively.

Supplementary Table S5. DNA sequences used in Figure 6B-C-E and in Figure S6D.

Supplementary Table S6. Kinetic and equilibrium coefficients for BG4 and Vid22 binding to various probe strands. kon, koff, and Kd: values extracted from data in Figure 6B-E and S6D

Supplementary Table S7. Genomic coordinates of copy number variants (CNVs) detected in the 11 *vid22*Δ yeast strains. The table contains the complete list of predicted copy number alterations in the 11 *vid22*Δ strains analyzed in the present study. Genomic coordinates in the form of chromosome name, start coordinate and end coordinate, are reported in columns 1 to 3. Column 4 indicates the number of predicted CNVs associated with each genomic interval. Column 5 reports a comma separated list of predicted CNVs. Each CNV is indicated according to the following format <Chr Name>_<start coordinate>_<End_coordinate>_<CNV_type>_ <strain>. For CNV types "d" indicates decrease (depletion) and "D" increase (duplication).

Supplementary Table S8. Single nucleotide variants and small indels in BY and *vid22* mutants. Total number of single nucleotide substitutions and small indels (less than 5 bp in size) are reported for the wild type BY4741 yeast strain, and for each of the 11 *vid22*Δ mutants used in this study (columns). For each strain counts of indels and single nucleotide substitutions are reported on the rows. Cumulative numbers are reported in the penultimate row (Tot). The last row (Wilcoxon P) indicates the p-value, according to a Wilcoxon sum and rank test, for the over-representation of small indels and SNVs in *vid22* mutants compared to the wild type strain.

Supplementary Table S9. Enrichment of G4 elements at predicted structural genomics rearrangements. p-values according to the hyper-geometric distribution. Fold changes of

enrichment are computed by comparison with a matched number of simulated random genomic intervals of the same size. The following datasets were used for the prediction/annotation of G4 elements in the yeast genome: Capra et al. predictions according to Capra et al. (17). PDS_Marsico_et_al, and noPDS_Marsico_et_al. G4 as established by G4-seq in Marsico et al 2019 (96) with and without pyridostatin (PDS) respectively. G4_AllQuads_Kudlicki et al. G4 as predicted by the AllQuads method, using "G3 N1-7 G3 N1-7 G3 N1-7 G3 " as definition of the intra-G4 motif (97). Capra et al. was considered since it provides a highly used resource which is based on non-stringent criteria for the prediction of G4. The predictions by ALLQuad were based on more canonical and very strict criteria, while in Marsico et al they provide experimental validation of G4 forming structures by a large-scale NGS assay and it is not purely based on in-silico predictions.

Supplementary Table S10. Complete list of the genomic regions altered in *vid*22 mutants and their intersections with G4 elements. The table contains the list of predicted copy number alterations in the 11 *vid22*Δ strains as described in Table_S7 intersected with predicted G4 elements. The G4 prediction dataset were described in Table_S9

Supplementary Table S11. Detailed annotation of ChIP-seq Peaks. Chr: chromosome. Start: start coordinate. End: end coordinate. LOG10-P: -log10 peak call p-value; Annotation: peak annotation. Distance to TSS: distance to nearest TSS in base pairs. Nearest Refseq: nearest refseq transcript id; Gene name: gene name.

Figure S1. Endogenous phosphorylation of Rad53 checkpoint kinase in mutants that show genomic instability. (**A**) Protein extracts from wild type, *vid22*Δ, *slx5*Δ *or slx8*Δ cells were analysed by western blotting with anti-Rad53 antibodies. The arrow indicates the phosphorylated forms of Rad53. (**B**) The cell cycle distribution of exponentially growing cultures was determined by flow cytometry of logarithmic phase *sml1*Δ, *sml1*Δ *vid22*Δ, *sml1*Δ *mec1*Δ and *sml1*Δ *mec1*Δ *vid22*Δ cultures. *SML1* deletion is necessary to suppress *mec1*Δ cell lethality. The positions of cells with 1C and 2C DNA contents are indicated.

Figure S2. Chromosomal pattern of wild type and *vid*22Δ strains. (**A**) Chromosomes were prepared from N=10 independent wild type strains and separated by PFGE. (**B**) DNA was prepared from 13 wild type and 7 independent *vid22Δ* mutants derived from the sporulation of a heterozygous diploid *VID22/vid22*Δ and chromosomes were fractionated by pulsed-field gel electrophoresis. The positions of the 16 chromosomes are indicated. * indicates chromosome size changes; **d** indicates chromosome duplications.

Figure S3. Comparison of expected and observed distance of CNVs prediction from G4 elements predicted from different. dataset. Frequency distribution of observed distance of predicted CNV from G4 elements is represented in blue. The expected distance distribution, estimated by 1000 independent random resampling of a matched number of genomic windows of identical size, is represented in red. Distances in base pairs (bp) are represented on the X axis, frequencies on the Y axis. The three different dataset of G4 prediction were described in Table S9.

Figure S4. Loss of Env11 does not lead to increased GCRs at G4 loci and Vid22 HU sensitivity. (**A**) The GCR rates for wt and *env11*Δ are plotted. Each data point is from an independent fluctuation test, with n ≥ 3 for each strain. The horizontal bars indicate the mean CGR rate for each strain (N=3 independent experiments). The GCR rate is calculated with the

experiments described and represented in Fig. 5. The *PRB1* locus was replaced with the G4 cassette as described in the text. An unpaired Student's *t*-test was used to compare the means of measurements and the p-value is indicated. (**B**) Tenfold serial dilutions of exponentially growing cultures of the indicated strains were plated on YPED, YPED + 100 mM HU. Images were taken after three days incubation at 28°C.

Figure S5. Summary of ChIP-seq peaks annotation and distance from G4 predicted elements. (A) The table reports a summary of the annotation of Vid22 ChIP-seq peaks. Annotation: type of genomic element. Total size: cumulative size in the genome. Vid22 Peaks: total number of Vid22 peaks associated with each element. Simulated Peaks: equivalent information for a matched number of simulated ChIP-seq peaks (see materials and methods). Log2 Enrichment: log2 fold enrichment (real vs simulated). p-value (Fisher Test): p-value for the enrichment according to a 1 tailed Fisher exact test. G4 real: number of Vid22 peaks overlapping predicted G4 elements. G4 simulated: equivalent for simulated peaks. Log2 G4 enrichment: log2 fold enrichment for proximity to G4 (real vs simulated), p-value G4 (Fisher Test): 1 tailed Fisher exact test for enrichment. (**B**) For every peak the distance to the closest element is reported. Dark purple: observed distribution. Light purple: expected distribution as established by analysis of a matched number of random genomic intervals of equivalent size. Distances, summarized in bins of 500 bps are reported on the X axis. Frequencies on the Y axis. (**C**) Zoom at a closer resolution. Only distances within 1000 bp from predicted G4 are displayed. Bins of 100 bps are used.

Figure S6. Vid22 binding to G4-DNA is independent of Tbf1 binding site, Sgs1 and its BED-domain. (**A**) List of DNA sequences of wild type and mutated Chr VIII *SKN7* locus; modified nucleotides are indicated in red; TAGGG represents the predicted Tbf1 binding site (TBS). (**B**) ChIP-qPCR of Vid22 at Chr VIII-G4 mutated in Tbf1 binding site. ChIP was performed in wild type (No Tag), Vid22-13Myc carrying G4 wild type sequence and Vid22-13Myc harbouring G4-mutated sequence; all strains used have both Tbf1-binding sites mutated as indicated in panel A. Fold enrichment of Vid22 at Chr VIII-G4 was calculated relative to the internal standard *HHT2*. Data are represented as mean ± SEM of N=3 independent experiments and the p-value is indicated. (**C**) ChIP-qPCR of Vid22 at Chr VIII-G4in presence or absence of Sgs1. Fold enrichment of Vid22 at Chr VIII-G4 was calculated as described in panel B. (**D**) Analysis of interaction of Vid22 BED-domain mutated protein with DNA forming G4 structures using the Reflective Phantom Interface (RPI) as reported in Figure 6B.

Figure S7. Vid22 loss does not lead to altered levels of DNA:RNA hybrids at the *GCN4* or *SKN7* loci. DNA:RNA hybrid immunoprecipitation (DRIP) with the S9.6 antibody in asynchronous culture of the wild-type, *vid22*Δ and *pob3-7* strains. Samples were treated (+) or not (−) in vitro with RNase HI (RNH) prior to the immunoprecipitation. Relative units are the ratio between *vid22Δ or pob3-7* and wild type percentage of input; N=3. The means and SEM of N=3 are plotted in both panels. An unpaired Student's *t*-test was used to compare the means of measurements and the p-value is indicated.

FIGURE S1

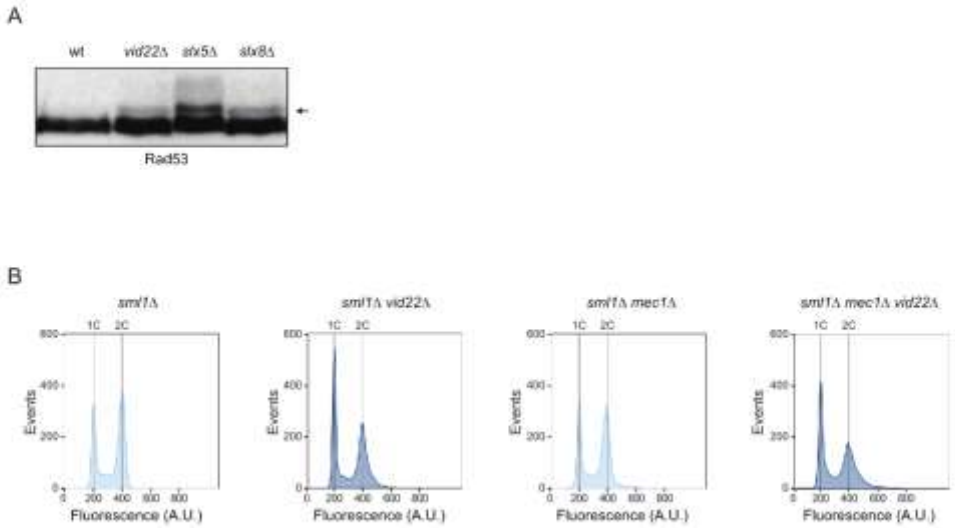


Figure S1. Endogenous phosphorylation of Rad53 checkpoint kinase in mutants that show genomic instability.
(A) Protein extracts from wild type, *vid22Δ*, *slx5Δ or slx8Δ* cells were analysed by western blotting with anti-Rad53 antibodies. The arrow indicates the phosphorylated forms of Rad53. (B) The cell cycle distribution of exponentially growing cultures was determined by flow cytometry of logarithmic phase *sml1Δ*, *sml1Δ vid22Δ*, *sml1Δ mec1Δ* and *sml1Δ mec1Δ vid22Δ* cultures. *SML1* deletion is necessary to suppress *mec1Δ* cell lethality.
The positions of cells with 1C and 2C DNA contents are indicated.

FIGURE S2

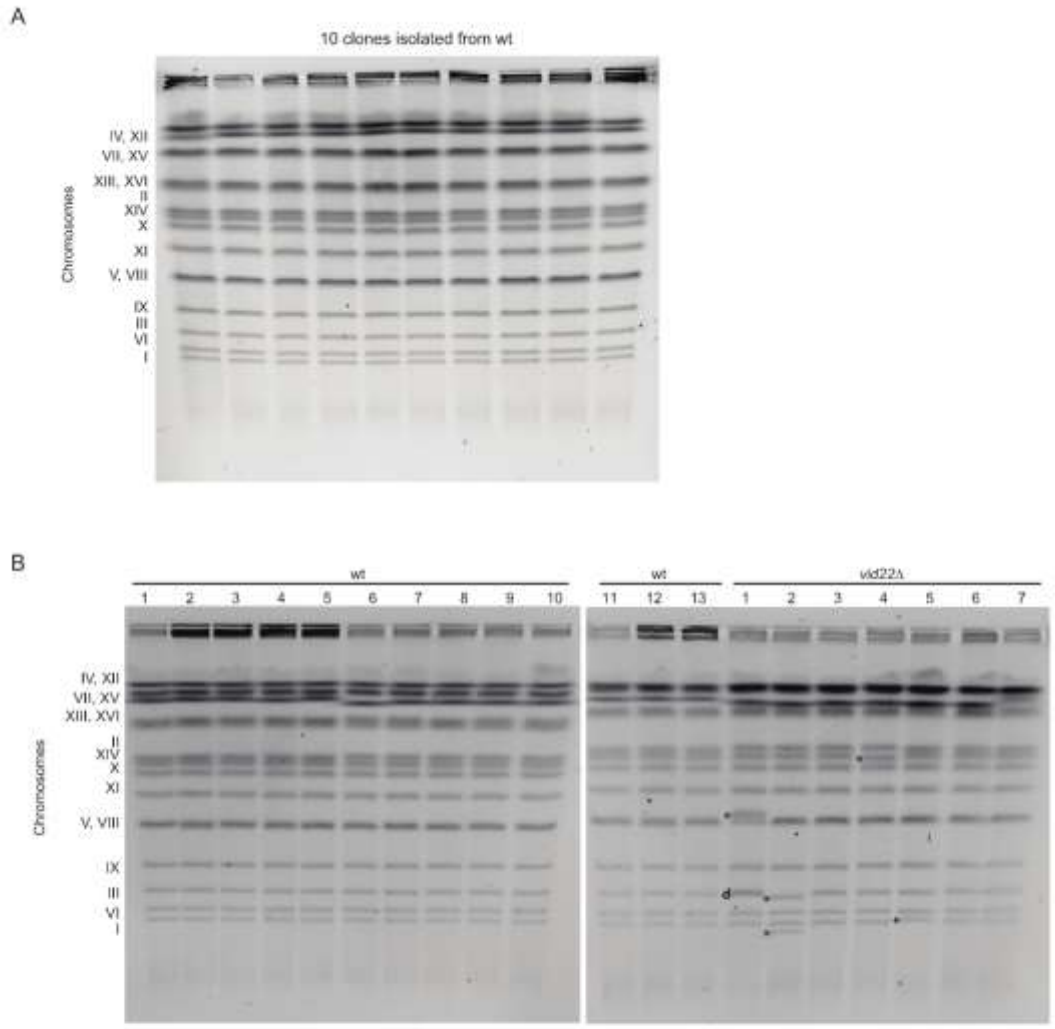


Figure S2. Chromosomal pattern of wild type and *vid22Δ* strains. (**A**) Chromosomes were prepared from N=10 independent wild-type strains and separated by PFGE. (**B**) DNA was prepared from 13 wild-type and 7 independent *vid22Δ* mutants derived from the sporulation of a heterozygous diploid *VID22/vid22Δ* and chromosomes were fractionated by pulsed-field gel electrophoresis. The positions of the 16 chromosomes are indicated. * indicates chromosome size changes; **d** indicates chromosome duplications.

FIGURE S3

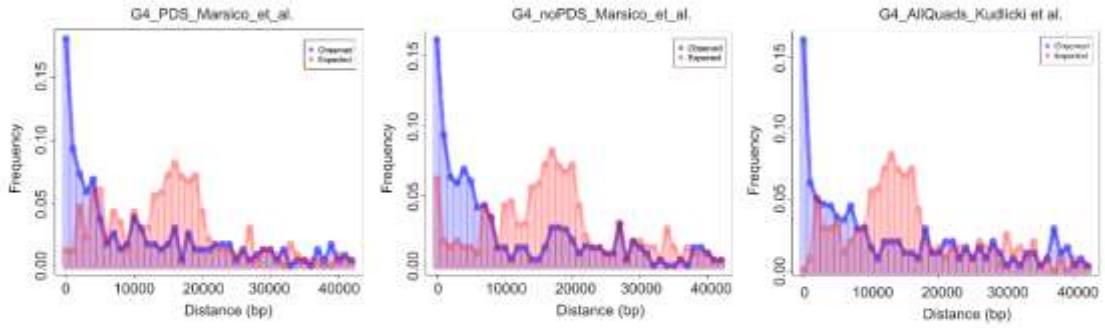


Figure S3. Comparison of expected and observed distance of CNVs prediction from G4 elements predicted from different dataset. Frequency distribution of observed distance of predicted CNV from G4 elements is represented in blue. The expected distance distribution, estimated by 1000 independent random resampling of a matched number of genomic windows of identical size, is represented in red. Distances in base pairs (bp) are represented on the X axis, frequencies on the Y axis. The three different dataset of G4 prediction were described in Table S9.

FIGURE S4

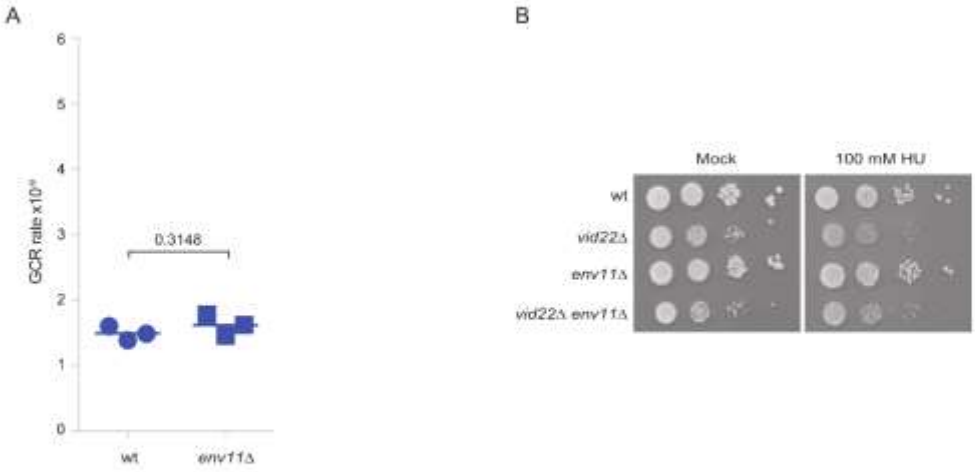


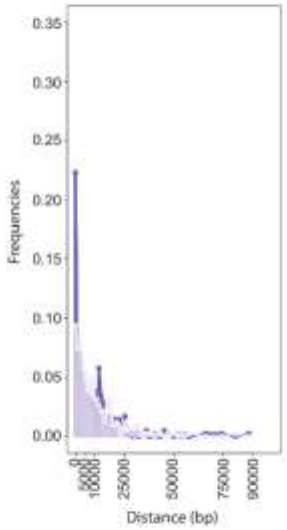
Figure S4. Loss of Env11 does not lead to increased GCRs at G4 loci and Vid22 HU sensitivity.
(**A**)The GCR rates for wt and *env11Δ* are plotted. Each data point is from an independent fluctuation test, with n ≥ 3 for each strain. The horizontal bars indicate the mean CGR rate for each strain (N=3 independent experiments). The GCR rate is calculated with the experiments described and represented in Fig. 5. The *PRB1* locus was replaced with the G4 cassette as described in the text. An unpaired Student's t-test was used to compare the means of measurements and the p-value is indicated.(**B**) Tenfold serial dilutions of exponentially growing cultures of the indicated strains were plated on YPED, YPED + 100 mM HU. Images were taken after three days incubation at 28°C.

FIGURE S5

A

Annotation	Total size (bp)	Vid22 Peaks	Simulated Peaks	Log2 Enrichment	Pvalue (Fisher Test)	G4(real)	G4 (simulated)	Log2 G4 enrichment	Pvalue (Fisher Test)
TTS	2928468	62	101	-0.7040152	9.98E-01	3	2	2.44	3.68E-01
Exon	3365496	11	125	-3.5363527	1.00E+00	1	3	5.68	1.80E-01
Intron	11804	1	0	NA	NA	0	0	NA	NA
Intergenic	2725287	108	70	0.6296045	6.03E-03	11	1	7.13	2.60E-02
Promoter	3033316	231	117	0.9613843	1.49E-07	25	2	6.33	2.27E-03
					Total	40	7	5.71	1.03E-08

B



C

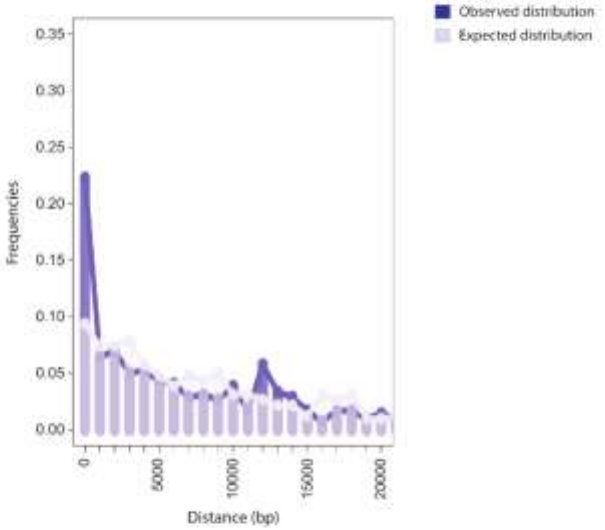


Figure S5. Summary of ChIP-seq peaks annotation and distance from G4 predicted elements.
(**A**) The table reports a summary of the annotation of Vid22 ChIP-seq peaks. Annotation: type of genomic element. Total size: cumulative size in the genome. Vid22 Peaks: total number of Vid22 peaks associated with each element. Simulated Peaks: equivalent information for a matched number of simulated ChIP-seq peaks (see materials and methods). Log2 Enrichment: log2 fold enrichment (real vs simulated). Pvalue (Fisher Test): p-value for the enrichment according to a 1 tailed Fisher exact test. G4 real: number of Vid22 peaks overlapping predicted G4 elements. G4 simulated: equivalent for simulated peaks. Log2 G4 enrichment: log2 fold enrichment for proximity to G4 (real vs simulated), Pvalue G4 (Fisher Test): 1 tailed Fisher exact test for enrichment. (**B**) For every peak the distance to the closest element is reported. Dark purple: observed distribution. Light purple: expected distribution as established by analysis of a matched number of random genomic intervals of equivalent size. Distances, summarized in bins of 500 bps are reported on the X axis. Frequencies on the Y axis. (**C**) Zoom at a closer resolution. Only distances within 20000 bp from predicted G4 are displayed.

FIGURE S6

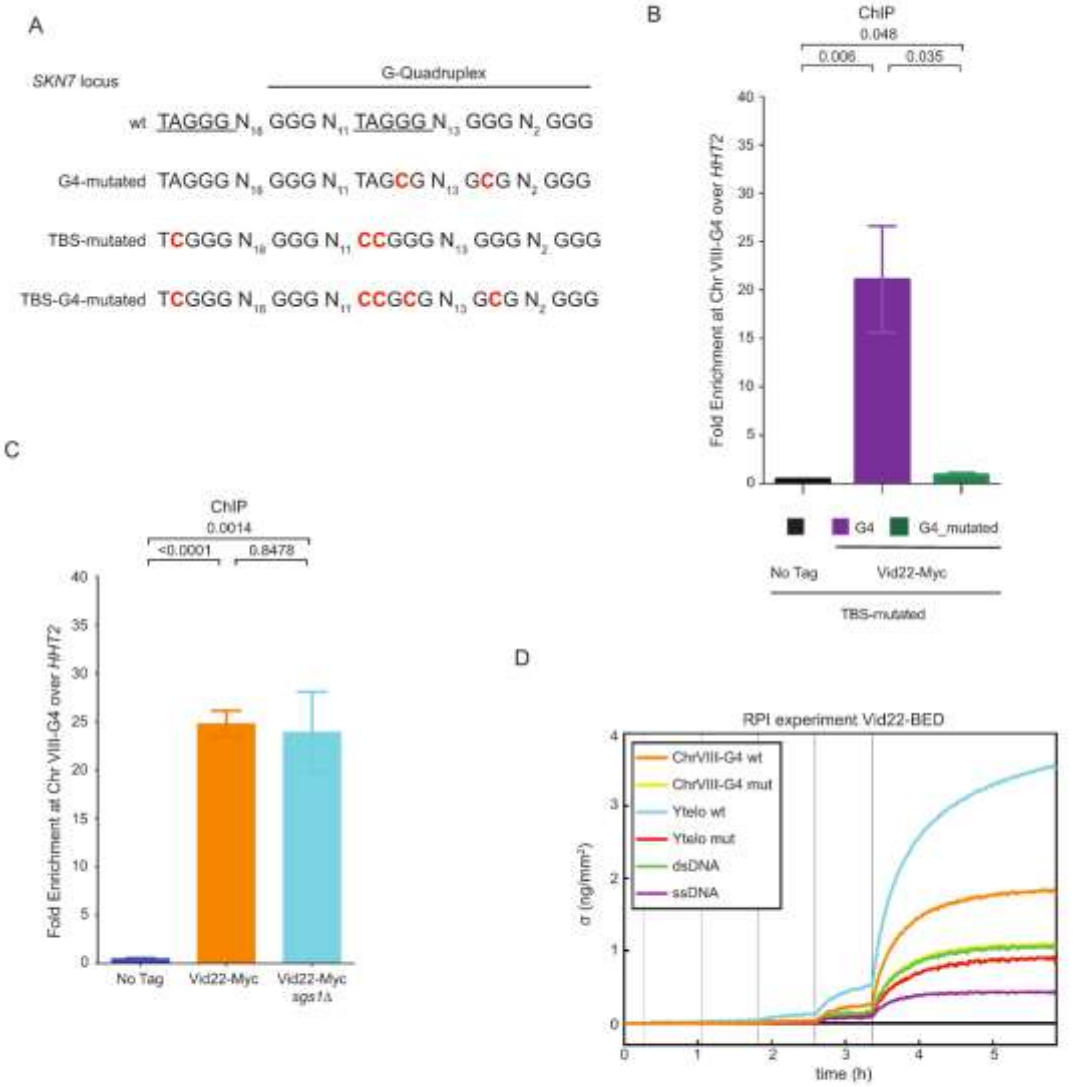


Figure S6. Vid22 binding to G4-DNA is independent of Tbf1 binding site, Sgs1 and its BED-domain. (**A**) List of DNA sequences of wild type and mutated Chr VIII *SKN7* locus; modified nucleotides are indicated in red; TAGGG represents the predicted Tbf1 binding site (TBS). (**B**) ChIP-qPCR of Vid22 at Chr VIII-G4 mutated in Tbf1 binding site. ChIP was performed in wild type (No Tag), Vid22-13Myc carrying G4 wild-type sequence and Vid22-13Myc harbouring G4-mutated sequence; all strains used have both Tbf1-binding sites mutated as indicated in panel A. Fold enrichment of Vid22 at Chr VIII-G4 was calculated relative to the internal standard *HHT2*. Data are represented as mean ± SEM of N=3 independent experiments and the p-value is indicated. (**C**) ChIP-qPCR of Vid22 at Chr VIII-G4 in presence or absence of Sgs1. Fold enrichment of Vid22 at Chr VIII-G4 was calculated as described in panel B. (**D**) Analysis of interaction of Vid22 BED-domain mutated protein with DNA forming G4 structures using the Reflective Phantom Interface (RPI) as reported in Figure 6B.

FIGURE S7

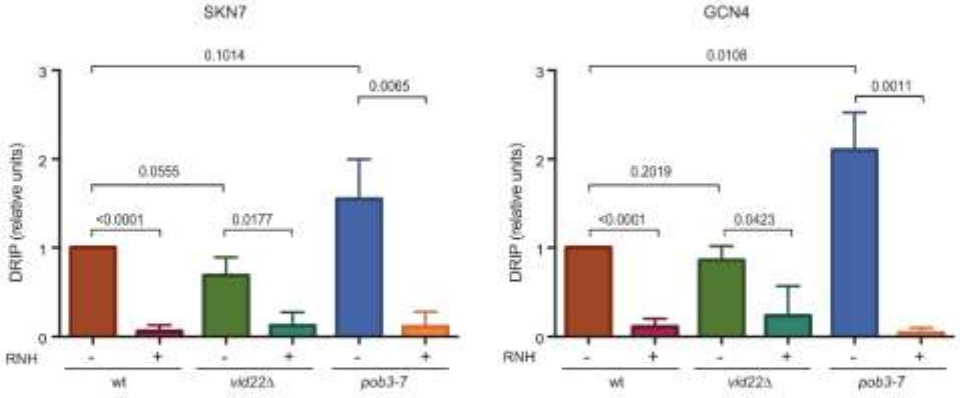


Figure S7. Vid22 loss does not lead to altered levels of DNA:RNA hybrids at the *GCN4* or *SKN7* loci. DNA:RNA hybrid immunoprecipitation (DRIP) with the S9.6 antibody in asynchronous culture of the wild type, *vid22*Δ and *pob3-7* strains. Samples were treated (+) or not (−) in vitro with RNase HI (RNH) prior to the immunoprecipitation. Relative units is the ratio between *vid22Δ* or *pob3-7* and wild type percentage of input; N=3. The means and SEM of N=3 are plotted in both panels. An unpaired Student's t-test was used to compare the means of measurements and the p-value is indicated.

10.2 Spotlight on G-quadruplexes: from structure and modulation to physiological and pathological roles

I contributed to this work by text revision and editing throughout, especially in paragraphs 2 and 3. I also performed an accurate revision of the literature cited in paragraph 2.

International Journal of
Molecular Sciences

MDPI

Review

Spotlight on G-Quadruplexes: From Structure and Modulation to Physiological and Pathological Roles

Maria Chiara Dell'Oca, Roberto Quadri, Giulia Maria Bernini, Luca Menin, Lavinia Grasso, Diego Rondelli, Ozge Yazici, Sarah Sertic, Federica Marini, Achille Pellicioli, Marco Muzi-Falconi and Federico Lazzaro *

Department of Biosciences, University of Milan, 20133 Milan, Italy; mariachiara.delloca@studenti.unimi.it (M.C.D.); giulia.bernini@unimi.it (G.M.B.); luca.menin@unimi.it (L.M.); lavinia.grasso@unimi.it (L.G.); diego.rondelli@unimi.it (D.R.); ozge.yazici@unimi.it (O.Y.); sarah.sertic@unimi.it (S.S.); federica.marini@unimi.it (F.M.); achille.pellicioli@unimi.it (A.P.); marco.muzifalconi@unimi.it (M.M.-F.)
* Correspondence: federico.lazzaro@unimi.it

Abstract: G-quadruplexes or G4s are non-canonical secondary structures of nucleic acids characterized by guanines arranged in stacked tetraplex arrays. Decades of research into these peculiar assemblies of DNA and RNA, fueled by the development and optimization of a vast array of techniques and assays, has resulted in a large amount of information regarding their structure, stability, localization, and biological significance in native systems. A plethora of articles have reported the roles of G-quadruplexes in multiple pathways across several species, ranging from gene expression regulation to RNA biogenesis and trafficking, DNA replication, and genome maintenance. Crucially, a large amount of experimental evidence has highlighted the roles of G-quadruplexes in cancer biology and other pathologies, pointing at these structurally unique guanine assemblies as amenable drug targets. Given the rapid expansion of this field of research, this review aims at summarizing all the relevant aspects of G-quadruplex biology by combining and discussing results from seminal works as well as more recent and cutting-edge experimental evidence. Additionally, the most common methodologies used to study G4s are presented to aid the reader in critically interpreting and integrating experimental data.

Keywords: G4; nucleic acids structures; DNA; RNA

Citation: Dell'Oca, M.C.; Quadri, R.; Bernini, G.M.; Menin, L.; Grasso, L.; Rondelli, D.; Yazici, O.; Sertic, S.; Marini, F.; Pellicioli, A.; et al. Spotlight on G-Quadruplexes: From Structure and Modulation to Physiological and Pathological Roles. *Int. J. Mol. Sci.* **2024**, *25*, 3162. https://doi.org/10.3390/ijms25063162

Academic Editor: Daniela Montesarchio

Received: 8 February 2024
Revised: 27 February 2024
Accepted: 5 March 2024
Published: 9 March 2024

1. Introduction

Nucleic acids are the biopolymers responsible for storage of genetic information, its heritability, and decoding to effectively build the components of an organism. While most of their functions have been linked to their canonical secondary structures, mounting experimental evidence has revealed the existence and functional relevance of a wide variety of alternative conformations of DNA and RNA, among which the G-quadruplex has been the most extensively studied. The following paragraphs expand upon seminal reviews in the field [1–3] by providing a summary of relevant and more recent literature pertaining to the structure, stability, dynamics, and physiological and pathological relevance of this non-canonical secondary structure.

2. Overview of the G-Quadruplex

Discovery and Structure of the G-Quadruplex

Research on G-quadruplexes and their structural unit, the G-tetrad or G-quartet, is a half-a-century old endeavor (Figure 1). The first experimental evidence of the self-assembly capabilities of guanines dates back to the 1910s, when it was reported that guanylic acid solutions at high concentrations formed a gel [4–6].

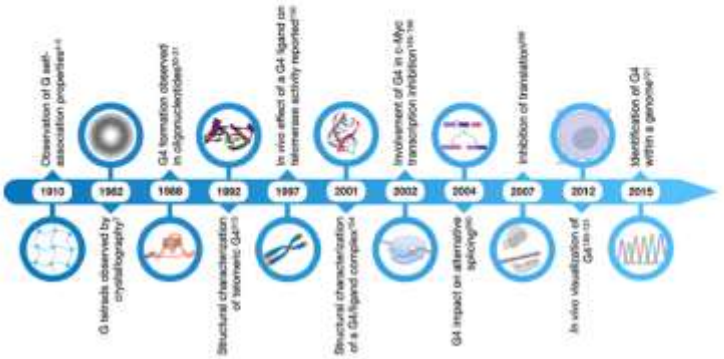


Figure 1. Selected milestones in G-quadruplex research.

Five decades later, the mechanism behind this curious phenomenon was revealed in an X-ray crystallography experiment, where the guanines of 5′-GMP were found to associate into a helical higher-order structure, presumed to be formed by the stacking of planar structures of guanine quartets held in place by hydrogen bonds [7]. This structural hypothesis of the G-tetrad was later built upon using X-ray diffraction data of polyinosinic acid [8,9], which is structurally similar to poly(G). In the late 1980s, new studies using oligonucleotides corresponding to the guanine-rich immunoglobulin switch regions and telomeric sequences postulated that these formed four-stranded assemblies compatible with stacked G-tetrads [10,11]. In addition to using sequences from real genomic regions notorious for their tandem GC repeats, the authors of these studies provided new compelling hypotheses on the involvement of this poorly characterized secondary structure in key biological processes, such as meiotic chromosome pairing before crossing over [10] and telomere regulation [11]. Since then, new insights into the structure of G-quadruplexes have been made thanks to a wide variety of experimental techniques (reviewed in ref. [12]), while the function of this assembly has been probed in biological systems. Indeed, shortly after in a seminal work it was found that oligonucleotides, bearing ciliate *Oxytricha nova* telomeric repeats folded into G-quadruplexes, inhibited telomerase activity in vitro [13], thus linking this structure to cancer biology and anticancer therapy. Nowadays, a consensus for the intramolecular G-quadruplex (also known as G4 or tetraplex) has been reached (Figure 2a), and its structure is understood to amount to at least two stacked planar G-tetrads, each consisting of four guanines interacting with one another forming non-canonical Hoogsteen bonds. More specifically, within the plane of a G-tetrad, four guanines are placed according to a four-fold rotation axis, exposing to one another compatible pairs of chemical groups that function as electron donor and acceptor. In this configuration, the electron-poor hydrogens of the -NH groups of the guanine base establish hydrogen bonds with the electron rich oxygens and nitrogens of the nearby guanine. In this planar array, stacked G-tetrads are further linked by van der Waals forces thanks to the π electrons of the aromatic rings from each base (reviewed in ref. [14,15]). The whole structure is additionally stabilized by metal ions, particularly monovalent alkali cations such as Na^+ and K^+, which are positioned at the central axis and coordinate the electron-rich oxygen atoms of the nearby guanilyl bases [16] (Figure 2b,d). Several in vitro studies have reported the ability of both monovalent and divalent cations to aid G4 folding and stability, albeit to different extents (reviewed in ref. [17]). The presence of four bases per G-tetrad implies that the G-quadruplex is a four-stranded structure, with each strand either being part of the same nucleic acid molecule (intramolecular G4s) or belonging to different molecules (intermolecular G4s). Additionally, the phosphodiester backbone, linking guanilyl bases of consecutive G-tetrads, has intrinsic directionality, meaning that a strand can be denoted as having a parallel or antiparallel

orientation to each of the neighboring ones (Figure 2c). Thus, G-quadruplexes can be referred to as parallel, antiparallel, or hybrid, the latter corresponding to only one G-strand running in the opposite direction with respect to the other three.

A notorious example of a parallel G-quadruplex structure comes from the determination of the human telomeric repeat assembly in the presence of potassium [18]. Noteworthily, in the case of intramolecular G4s, two strands participating in a G-quadruplex must be linked by a loop. Numerous loop connections have been described, with the simplest ones being the propeller, lateral, and diagonal loops (reviewed in ref. [14,19]; Figure 2c). Interestingly, it has been demonstrated that loop length and sequence affect the stability and folding of the tetraplex structure [20–25]. An additional element to be considered when describing the structure of a G-quadruplex is the type of grooves of the assembly, which are defined by the *anti* or *syn* conformations of the ribonucleotides or deoxyribonucleotides linked to the guanines participating in the G-tetrad. These conformations refer to the torsion angles formed by the glycosidic bond between the sugar and the nucleobase moieties. Depending on the configurations of the nucleotides of adjacent guanines in a tetrad, grooves are defined as wide, medium, or narrow, with the first and last of these corresponding to arrangements where aligned nucleotides have opposite glycosidic bond angles [15,26–28]. Although all the topologically possible combinations of N-glycosidic bond conformations and canonical loops have been determined [26], only a subset of these have been experimentally verified [29–32]. To complicate things even further, a number of non-canonical G-quadruplex assemblies have been observed over the years (reviewed in ref. [33]). While most G-quadruplexes have been reported to have a right-handed helical twist, left-handed G4s [34] and even hybrid G4s [35] have also been observed. Left-handed G-quadruplexes can be formed by at least two distinct minimal sequence motives of 12 nt [36,37]. Interestingly, the GTGGTGGTGGTG motif, which is highly abundant in the human genome, was not only shown to independently form left-handed G4 structures, but also to drive the formation of left-handed conformations from several other sequences, when attached to them [36]. Machine learning methods to classify right- and left-handed G4s based on torsional angles have been recently explored [38]. Moreover, bulges or protrusions of nucleotides from the G-tract [39,40], vacant guanine spots inside G-tetrads, snapback loops that fill in the empty spot [41–46], and D-shaped loops connecting guanines that are not contiguous in sequence but that are part of the same G-tract [47,48], are only some examples of our current understanding of unusual G-quadruplex topologies. All in all, depending on the combination of the number of nucleic acid molecules that participate in the formation of a G-quadruplex, the relative orientation of its backbone strands, and the presence and topology of connecting loops and bulges, it is evident that a wide variety of G-quadruplex conformations is possible. Crucially, it has been determined that the same G-quadruplex-forming sequence can give rise to two or more coexisting G4 structures in solution [30,49–51]. Interestingly, the type of cation in a buffer was shown to influence quadruplex conformation in vitro (reviewed in ref. [17]). In addition, different structures show variable affinity towards G-quadruplex-binding proteins [52,53] and small molecules [54–56], and this can be the basis for the selection of molecules that are specific for different structures. Given the abundance of polymorphic G-quadruplex structures, it is thus tempting to assert that cells can discriminate between different G4 topologies, which would then be linked to distinct biological roles in live cells [19], although the extent to which this is true is yet to be elucidated.

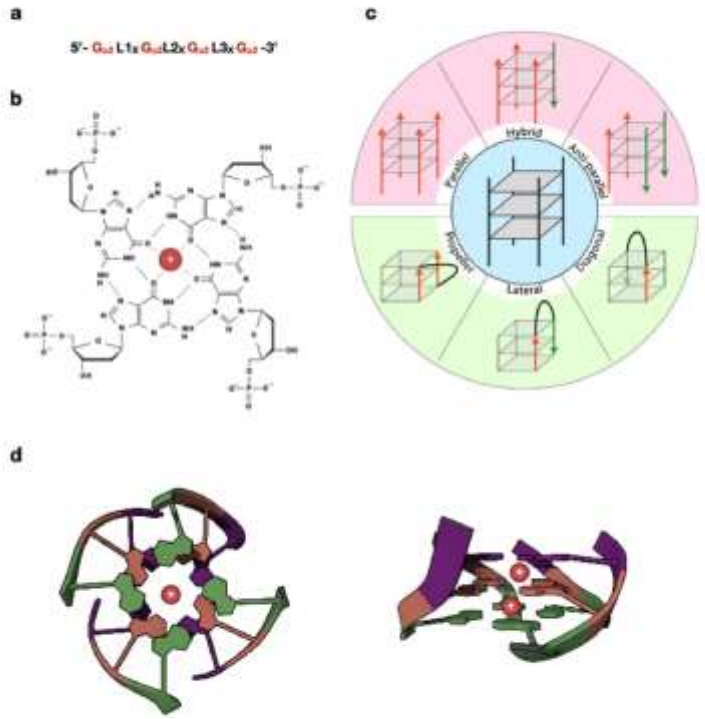


Figure 2. Structural organization of G-quadruplexes. (**a**) Basic sequence of an intramolecular G-quadruplex. (**b**) Structure of a DNA G-quartet, the basic structural element of a canonical G-quadruplex. Four coplanar guanines establish Hoogsteen bonds between their complementary surfaces, while the electron-rich O atoms are stabilized through coordination of the central monovalent cation. Multiple G-quartets are stacked upon one another via π-π orbital interaction of their aromatic systems. (**c**) Schematic representation of possible strand and loop orientations within a G-quadruplex. G4s are categorized as parallel when all four strands of their G-quartets have the same orientations, antiparallel when two pairs of strands run in opposite directions, or hybrid when three out of four strands are codirectional. G-quadruplex strands are connected by loop structures, the most common being the propeller (**bottom left**), the lateral (**center bottom**), and the diagonal loops (**bottom right**). (**d**) Top (**left panel**) and lateral (**right panel**) views of human telomere structure (PDB ID: 2HY9 [57]) modified with Mol* (https://doi.org/10.1093/nar/gkab314). Different colors represent different G-quartets.

Over the years, significant strides have been made in providing structural information of sequences harboring G-quadruplexes. As was expected from their G-rich repeats, structural studies have revealed that human telomeric repeats fold into G-quadruplexes in vitro [18]. Besides telomeres, sequences shown to fold into G-quadruplexes in vitro were found in promoters of key genes, particularly in the nuclease-hypersensitive element III of human *c-myc* [58–60] and a regulatory element 87 nucleotides upstream of the transcription start site of human *c-kit* [46,61]. In addition, in human minisatellites CEB1 and CEB25, G4 structures were solved by NMR [62,63]. Despite mounting evidence of G4 structures from G-rich sequences, it is not guaranteed that tetraplexes can fold at corresponding genomic sites inside cells. A key aspect of most in vitro studies is that

G-quadruplex structures are determined using single-stranded oligonucleotides. It is important to consider that in chromosomal DNA, the tetraplex structure is in competition with the standard double-helix conformation, with the displaced C-rich strand available to re-establish Watson–Crick complementarity. Indeed, it was found that, in the presence of the complementary strand, some oligonucleotides prone to G-quadruplex formation tended to form duplexes instead [21,64–72]. This suggests that G-quadruplexes do not fold automatically in chromosomal DNA unless the double helix is unwound or denatured to expose the guanines for G-tetrad formation. One possibility of that occurring is when DNA is subjected to negative supercoiling; however, it has been demonstrated that this alone cannot induce G-quadruplex folding in plasmids [73]. Contrastingly, single-molecule imaging experiments in cells revealed that G-quadruplexes are indeed formed upon dsDNA unwinding during DNA replication [74]. Furthermore, when the complementary C-rich strand is independently stabilized, G-quadruplex formation is favored. This was demonstrated in vivo with the so-called G-loop, a structure where R-loop formation on the transcribed strand is coupled to G-quadruplex folding on the non-template strand [75]. However, the opposite is true when considering the case of i-motifs. Similarly to guanine stretches, a series of cytidine residues in DNA or RNA can also associate in four-stranded structures called i-motifs. These are secondary nucleic acid structures consisting of four strands stabilized by hemi-protonated and intercalated cytosine base pairs (C:C+) [76,77]. It has recently been widely demonstrated that the complementary strand of any G-quadruplex-forming sequence is prone to forming i-motifs (reviewed in ref. [76,78]). Importantly, the stabilization of G-quadruplexes using small molecules was shown to destabilize the i-motifs, and vice versa. This suggests that these structures are interdependent [79]. Owing to their distinctive physicochemical properties, these i-motif structures have garnered considerable attention as novel targets for drug development (reviewed in ref. [80]); however, a thorough discussion of these secondary structures is beyond the scope of this review. Another factor that ties G4 folding to polymerase activity is the increased stability of the tetraplex in nanocages mimicking the confined space in the exit channel of polymerases [81]. G-quadruplex folding was also found to be favored in molecular crowding conditions [82–84], likely more representative of the in vivo environment of chromosomal DNA. Another aspect to consider is the local chromatin accessibility, a feature tightly regulated by chromatin factors and nucleosome positioning. Indeed, it was found that G-quadruplexes in native chromatin of human cells often colocalized with markers of euchromatin (H3K4me3 and RNApol II), which are mostly nucleosome-depleted [85], and that G4 formation likely displaces nucleosomes in vitro [86]. While the concept of non-B DNA structures being excluded by nucleosomes and accumulating in open chromatin is compelling, it needs further investigation. Overall, mounting evidence indicates that G4 folding in vivo is possible and facilitated by exposure of ssDNA, stabilization of the complementary C-rich strand, the crowded conditions genomic DNA can find in cells, confined spaces resembling the exit channel of polymerases, and possibly nucleosome-depleted accessible chromatin. As for RNA G-quadruplexes, in vitro studies determined that ribonucleotides could also fold in tetraplex assemblies. Intriguingly, given the single-stranded nature of most RNAs, these structures were predicted to fold more easily than in DNA, as well as being exceptionally stable [87–89]. In addition to proving that G-quadruplex folding is feasible in cells, and thus allowing further investigations into the biological function of these structures, most of the abovementioned conditions are compatible with transcription, telomere elongation, and DNA replication. The biological relevance of G-quadruplexes will be discussed in a later section.

3. Tools to Study G-Quadruplexes

Depending on the biological question, a plethora of protocols, instruments, and probes can be used to determine structural features or biological functions of G-quadruplexes. Understanding the type of data each one produces, as well as advantages and disadvantages, is fundamental when trying to piece evidence together into a comprehensive view

of G-quadruplex biology. Knowledge of the most common types of assays can also help in validating results or when developing or improving an experimental protocol. The following sections will provide a brief description of the techniques that have provided hallmark results in the past, as well as those that are widespread in today's scientific reports pertaining G4 biology (Table 1). Newly developed probes and protocols will also be briefly discussed to provide insight into how the field is being improved.

Table 1. List of in vitro, in silico, and in vivo techniques commonly used to study G-quadruplexes.

Technique	Data Obtained	Type	Advantages	Disadvantages
X-ray crystallography	Structure Ligands	in vitro	Angstrom level resolution	Requires suitable G-quadruplex crystals Impossible to study dynamics
NMR spectroscopy	Structure Stability (time) Ligands	in vitro in vivo (adapted)	Physiological-closed conditions Dynamic studies Can detect multiple G4 at the same time	Limited sensitivity
CD spectroscopy	Structure Stability (temperature)	in vitro	Can discriminate between parallel/antiparallel/hybrids G4s Cost-effective study of ligand stabilization/destabilization	Less informative than NMR or X-ray Susceptible to non-canonical conformations Susceptible to the presence of A-form duplexes
UV melting	Stability (temperature) Ligands	in vitro	Dynamic studies	Low resolution
FRET	Stability Distance between 3′ and 5′ ends of ssDNA	in vitro	Absolute distance measurement Single molecule resolution	Fluorophores might affect G4 folding
Bioinformatic prediction	Prediction of G4 within the genome/transcriptome	in silico	Cost-effective Genome-wide analyses	Can only predict G4s matching the model used to generate the predictor Requires validation
G4-seq and rG4-seq	Distribution in genome and transcriptome	in vivo	Can identify non-canonical G4s	Require validation Susceptible on the type of the used molecule
Antibody-based methods	G4 spatial distribution ChIP	in vivo	Direct visualization of G4s in cells	Susceptible on the type of the used molecule

3.1. In Vitro Structural Studies

In vitro studies of G-quadruplex structure and stability remain a fundamental step in probing the morphology and stability of a tetraplex. As mentioned, biochemical and structural investigations were first used to discover the structure of telomeric G-quadruplexes. These types of studies were fundamental to assess the wide variety of canonical and non-canonical G-quadruplex structures, paving the way for the concept of G-quadruplex polymorphism. Furthermore, seminal investigations into the dynamics of G-quadruplex forming oligonucleotides were the basis for understanding the structural elements and environmental conditions (cations, pH, and molecular crowding) that impact on G-quadruplex dynamics. These insights are still valuable in current research, as structural information can be used for rational ligand design and molecular simulations. Among all in vitro assays that have been used to probe G-quadruplex morphology, NMR spectroscopy (reviewed in ref. [90]) and X-ray crystallography (reviewed in ref. [91]) are the most common. The latter allows structural determination of DNA or RNA G-quadruplexes, alone or in complex with proteins or ligands, at Angstrom resolution. Despite the potential of this technique, its major bottleneck is the formation of G-quadruplex-containing crystals suitable for analysis. X-ray crystallography requires that the target molecule is found in a highly ordered

crystalline array, which is challenging to obtain. Several conditions need to be probed in G-quadruplex crystallization protocols to obtain a viable crystal [91,92]. On the other hand, the need for crystallization is completely bypassed in NMR spectroscopy, which offers the unique advantages of assaying G-quadruplex structure and dynamics in solution at close to physiological conditions [90,93]. In the NMR experiments, the appearance of an imino peak in the 10.0–12.5 ppm range is indicative of the G-tetrad formation [93] and can be used to detect more than one G-quadruplex conformation at the same time [94], although this may result in extensive spectral overlap. When the objective is to study a single structure from a sequence capable of folding into multiple ones, such a sequence is modified accordingly, for example, by changing the loop length or the sequences flanking the G-quadruplex [90]. Besides the oligo sequence, a desired G-quadruplex morphology can be selected by adjusting buffer and experimental conditions [90] or by removing all undesired structures from the sample using size exclusion chromatography [95]. NMR is also suited to monitoring G-quadruplex stability over time or structural changes upon addition of a ligand, allowing a deeper insight into the kinetics and dynamics of a particular conformation, as well as aiding ligand design [90]. Interestingly, some NMR protocols can also be performed in living cells, as is the case for a ^{19}F NMR spectroscopy study on human telomeric RNA G-quadruplexes injected into *X. laevis* oocytes [96]. Crucially, the study where this technique was described also unequivocally reported the existence of folded RNA G-quadruplexes in vivo [96], which had been previously debated [97]. Despite the widespread use of crystallography and NMR structural studies, their applicability is currently beyond the reach of investigating higher-order G-quadruplex structures due to their high topological flexibility, a feat that would provide unvaluable data on G4 positioning into a close-to-native DNA duplex. A recent study [98] provided a possible solution to this problem. By combining small angle X-ray scattering (SAXS), molecular dynamics simulations, previously solved G-quadruplex structures, and a 7.4 Å resolution cryo-EM model, the integrative approach yielded a plausible structural model of a parallel G4 embedded into duplex DNA with a polyd(T) ssDNA stretch. This tertiary DNA structure closely mimics that of promoter G-quadruplexes and suggests that guanine tetraplexes are found stacked coaxially to the dsDNA duplex surface, rather than protruding outside of the double helix as previously thought [98].

An alternative to X-ray crystallography and NMR spectroscopy that does not require expensive equipment is circular dichroism (CD) spectroscopy, which was a pioneering method used to study DNA secondary structures. Although providing much less informative than the two techniques described above, CD spectroscopy can be used to distinguish between parallel and antiparallel tetraplex conformations based on characteristic CD spectra of the two conformations; parallel G-quadruplexes exhibit a sharp ellipticity maximum at 260 nm and a minimum at 240 nm, while antiparallel structures show a maximum and minimum at about 290 nm and 265 nm, respectively. These stark differences in CD spectra are the result of the different N-glycosidic torsion angles of the strand participating in G-quadruplex formation (all *anti* in parallel G-quadruplexes, and alternate *anti-syn* in antiparallel ones), leading to differently stacked G-tetrads [14,99]. CD spectra are easily interpreted using these criteria, especially when studying telomeric G-quadruplexes; however, non-canonical conformations and structures that do not conform to telomeric G-quadruplexes may yield different CD spectra and prevent scientists from assigning strand orientation to the structure. Nevertheless, outside of these cases, CD spectroscopy offers the possibility to calculate the melting temperature, thus providing additional information on the stability of the assembly [14]. On the basis of the characteristic transition of UV-visible absorbance for G-quadruplexes at around 295 nm, three biophysical methods for the characterization of G4 structures in vitro are commonly used: isothermal differential spectrum (IDS), thermal differential spectrum (TDS), and UV melting (reviewed in ref. [100]). In particular, UV melting experiments are performed to assay G-quadruplex thermal melting by measuring UV absorbance of the sample at 295 nm at increasing temperatures [101]. In combination with these techniques, studies using ultraviolet resonance Raman spectroscopy [102] and

differential scanning calorimetry [103] can provide important information on the structure and dynamics of G-quadruplexes.

Alternatively, G-quadruplex unfolding can be monitored with fluorescent probes linked to the extremities of oligonucleotides. For example, Fluorescence Resonance Energy Transfer (FRET) has been adapted for G-quadruplex stability experiments; FRET efficiency (acceptor fluorescence intensity over donor fluorescence emission) can be used to measure the distance of the 3′ and 5′ ends of a ssDNA, which is minimal when a G-quadruplex is present [65,104,105]. Although particular care should be taken in the choice of fluorophores and in minimizing their impact on G-quadruplex foldability, FRET provides the unique advantage of absolute distance measurement as direct evidence of tetraplex folding and unfolding [104] and has evolved to reach single molecule resolution [66,106]. Recently, evolutions of this technique, such as FRET-MC (melting competition) assay and its isothermal version, Iso-FRET, have been proposed to analyze quadruplex formation in vitro [107,108]. Moreover single-molecule force spectroscopy techniques, include optical tweezers (OT), magnetic tweezers (MT), and atomic force microscopy (AFM) allow real-time detection of folding/unfolding dynamics and discovery of different G4 topologies (reviewed in ref. [109]).

3.2. Bioinformatic Prediction of G-Quadruplexes and Polymerase Stop Assays

In vitro biophysical studies have provided instrumental information on G-quadruplex structure variability, stability of different tetraplex forms, and how it is impacted by sequence elements. This allowed the definition of rules to predict whether a given sequence can fold into a G-quadruplex, laying the bases of bioinformatic predictions. The earliest works in this direction used a simple sequence model built on the assumption that intramolecular G-quadruplexes require four short stretches of at least three consecutive guanines separated by short spacers that would become the loops of the assembly [110,111]. In other words, the basic sequence defined as having the potential to fold into a G-quadruplex had the following structure: $G_{3-5}L1_{1-7}G_{3-5}L2_{1-7}G_{3-5}L3_{1-7}G_{3-5}$, where L1, L2, and L3 are the three loops connecting the guanines that are assembled into G-tetrads. The maximum length of these connecting sequences was capped at seven nucleotides to reduce complexity [111]. Even with this low complexity model, the Quadparser algorithm indicated that as many as about 375,000 non-overlapping sequences in the human genome could potentially form a G-quadruplex [110,111]. In the following years, this number has risen thanks to improvements in algorithms and putative quadruplex sequences (PQS) models (reviewed in ref. [112]). Despite the widespread adoption of the original strategy based on the abovementioned consensus motif [113–115], new methods have since been developed to allow the detection of non-canonical G-quadruplex structures. For example, the presence of a limited number of bulges, mismatches, and long loops in a sequence does not automatically exclude it from G-quadruplex propensity evaluation in pqsfinder, although a score is assigned to each identified sequence to penalize such imperfections [116]. Other detection tools, such as G4Hunter, assign the score using G richness and G skewness [117]. More recent algorithms have incorporated machine learning models [118,119] and/or are also trained on a dataset of sequences confirmed to form G-quadruplexes in vitro [116,117]. Overall, as new tools are developed to more accurately predict putative canonical and non-canonical G4 sequences, it is evident that the abovementioned sequence model used by the Quadparser algorithm has become obsolete.

Additionally, other bioinformatic tools have been used to predict the three-dimensional conformation of G-quadruplexes from a given sequence. Indeed, inter- and intramolecular G-quadruplex structures can be generated in silico using web tools such as 3D-NuS [120]. This modeling could be used to explore the dynamics of different G-quadruplexes and the docking of proteins and ligands.

Furthermore, the efficacy of some PQS prediction algorithms was tested using the output of the high-throughput G4-seq method [121], which consists of a modified Illumina sequencing protocol where polymerase arrest upon encountering a folded G-quadruplex

in the template is identified by a drop in sequencing quality scores. Only the dip in Q-scores specific for G-quadruplexes was selected thanks to the comparison between a sequencing run in standard conditions and one where G-quadruplex stabilizing factors (either K^+ or the small-molecule ligands such as pyridostatin or PhenDC3 were added to the sequencing buffer). G4-seq of the human genome identified more than 700,000 G-quadruplex forming sites, of which about 70% were not predicted by the standard G4 motif-based Quadparser algorithm [110,121], possibly comprising non-canonical G4 structures escaping the algorithm folding rule ($G_3+N_{1-7}G3+N_{1-7}G3+N_{1-7}G3+$), like those with loops longer than seven bases, bulges in the G-tracts, or only two G-tetrads. On the contrary, the original G4-seq method failed to detect 27% in PDS and 40% in K+ of Quadparser-predicted canonical G4-forming motifs [121]. Even when considering predicted potential G4-forming motifs with loop lengths up to 12 nucleotides, 37% of them still evaded G4-seq in PDS outputs [122]. This inconsistency may be mainly explained by inadequate sequencing coverage in certain GC-rich genomic regions (due to inefficient amplification at stable G4s in Na+ and PCR biases during library preparation), PDS stabilization performed in Na+, and low resolution of the observed G4 motifs and consequent merging of proximal G4 motifs of the original experiment. Additionally, limited G4 stability, binding specificities of the employed G4 ligands, and the in vitro experimental conditions may account for a fraction of false negatives. To overcome these limitations, improvements have been introduced in the second-generation method that have successfully increased the specificity of the assay in K^+ PDS. In fact, the refined protocol detected ~95% of human canonical Quadparser G4s and 84% of the ~706 k potential G4-forming sequences with loops as long as 12 nt [122].

The assay was also adapted for transcriptomic analysis; in rG4-seq reverse transcriptase stalling is induced when G-quadruplexes are folded in vitro on template mRNA upon addition of K^+ or K^+ and pyridostatin (PDS) [123]. After retrotranscription and Illumina sequencing, the sequences that form G-quadruplexes are detected as a drop in coverage when compared to the same assay performed with Li^+, a known G-quadruplex destabilizing agent [123]. Analysis of the HeLa transcriptome revealed that as much as 88% of detected reverse transcriptase arrests correspond to non-canonical G-quadruplexes [123]. It is evident that both G4-seq and rG4-seq can potentially detect tetraplex structures without prior knowledge of their sequence motif or overall fold, while traditional bioinformatics tools are based on a model that cannot be applied reliably to all possible G-quadruplexes. Despite this, both polymerase-stop assays and bioinformatic predictions of G-quadruplexes offer a limited view of the role of these structures in genomes and transcriptomes, because the presence of these folded structures must be validated in vivo to provide any meaningful insight. In addition, both approaches offer potentially biased results, since most algorithms are built on a sequence model that cannot take into account G-quadruplex structures that are yet to be studied, while G4-seq and rG4-seq offer only indirect proof of G-quadruplex formation and use small-molecule ligands that may have binding preference for certain folds over others. These issues, however, were partially addressed in the original studies; rG4-seq results showed more pronounced reverse transcriptase stalling in the K^+-PDS condition with respect to the K^+ only experiment [123], while G4-seq data with PDS was largely overlapping with that of the alternative G4 stabilizer molecule PhenDC3 [121,124].

A further improvement may come in the next years from innovative sequencing technologies such as nanopore sequencing. Indeed, a recent work proved that non-B DNA structures, including G-quadruplexes, can be predicted from whole genome nanopore sequencing data, based on the timing of DNA translocation of non-B compared with B DNA [125]. These findings represent a crucial step towards direct G4 detection in cellulo and in vivo, which could add an additional step in the study of these structures.

3.3. *Antibody-Based Methods for G4 Detection*

While in vitro and in silico studies of G-quadruplexes are fundamental to understand the structure, stability, folding dynamics, and prevalence of G4-forming sequences in

nucleic acids, the need for a dissection of their functional relevance has fueled the development of tools to detect these structures in vivo. Within cells, the complex interplay between guanine tetraplexes, native chromatin context, G-quadruplex-binding proteins, and those involved in genome maintenance and RNA biogenesis results in G-quadruplexes modulating a molecular process or pathway to ultimately produce a biologically relevant effect on the system. To link such effects to G-quadruplexes, the most common strategies rely on designing a probe, antibody, or small molecule that specifically binds the tetraplex assembly. Over the years, a small number of antibodies and single-chain variable fragments (scFv) have been used and validated to detect G-quadruplexes directly in cells. The latter correspond to structures where the heavy-chain and light-chain variable portions of a traditional antibody, which, when combined, form the antigen recognition surface, have been synthesized into a single polypeptide chain. Among these, the probes Sty3 and Sty49 were originally selected by ribosome display from the Human Combinatorial Antibody Library for high affinity binding of the ciliate *Stylonychia lemnae* telomeric repeat $d(G_4(T_4G_4)_4)$ and provided the first evidence of G-quadruplexes in isolated nuclei [126]. Sty3 was found to specifically recognize parallel G-quadruplex structures at picomolar affinity [126], while Sty49 showed comparable nanomolar affinity for both parallel and antiparallel conformations [126]. Probe binding was validated in vitro by in situ immunofluorescence on isolated *S. lemnae* macronuclei, where Sty3 produced no signal, while Sty49 binding was shown to be specific for folded G-quadruplexes [126,127]. Importantly, Sty49 immunostaining provided the first direct evidence that G-quadruplexes are absent from replicating telomeres, which was expected due to the tetraplex assembly potentially inducing polymerase arrest during replication [126]. It is important to note that these studies were performed on isolated macronuclei from hypotrichous ciliates, where telomeric DNA concentration is far higher compared to other species [128,129]. Two other single-chain variable fragments were validated by immunofluorescence experiments on human cells, proving to be appropriate solutions to explore G-quadruplex biology in metazoans. BG4 was isolated by phage display and confirmed to bind various G-quadruplex conformations in vitro [130]. BG4 staining in multiple human cancer and non-cancer cell lines shows mostly nuclear foci, with fluorescent signals increasing upon G-quadruplex stabilization induced by PDS treatment [130] and being abrogated in DNase I-treated or G4-folded oligos-transfected cells [130]. Remarkably, it was revealed that BG4 foci mostly do not colocalize with telomeres and that their number peaks in S phase, in accordance with the hypothesis that ssDNA exposure during replication favors G-quadruplex formation [130]. In addition, BG4 immunostaining was proven to detect cytoplasmic RNA G-quadruplexes [131]. Although the original study showed that BG4 could bind parallel, antiparallel, and mixed propeller G-quadruplexes in vitro [130], an independent group demonstrated that the single-chain antibody preferentially binds parallel conformations by EMSA experiments [132]. A less commonly employed ScFv designed to bind to G-quadruplexes is D1. Like BG4, D1 was identified by phage display and it was specifically chosen for its marked binding preference to parallel G-quadruplex structures in ELISA assays, as well as for its clear BG4-like foci signals in immunofluorescence [133]. Co-staining with the telomere binding protein TRF2 in human SiHa cells showed that telomeric repeats were mostly folded in parallel G-quadruplex assemblies in vivo [133]. In addition to BG4 and D1, the traditional monoclonal antibody 1H6 was generated using vertebrate and ciliate telomeric G-quadruplexes as immunogens [134]. 1H6 was shown to recognize most G-quadruplex conformations in vitro and to specifically bind folded tetraplexes in immunofluorescence and immunohistochemistry experiments, with an increase in fluorescent signals upon treatment with the G4 stabilizer TMPyP4 or knockout of G-quadruplex helicase FANCJ [134]. Despite the possibility of performing isotype controls in experiments, 1H6 was later found to exhibit cross-reactivity with adjacent thymines in both folded G-quadruplexes and denatured DNA while not recognizing G-quadruplex structures bearing less than three adjacent thymines [135], thus complicating the interpretation of 1H6 experiments and casting significant doubt on the results obtained with this tool [136]. Overall, owing to 1H6 cross-reactivity and D1 specificity

for parallel G-quadruplexes, BG4 has been the most frequently used G-quadruplex probe used to assay multiple tetraplex conformations at the same time. This does not exclude the possibility that BG4 may exhibit a bias towards some G-quadruplex folds over others; thus, particular care must be taken in not generalizing results. Despite this, BG4 has been used to adapt several other protocols for G-quadruplex studies in cells. Folded G-quadruplexes in cells were previously assayed indirectly by ChIP-seq of proteins known to recognize such structures [114,137–139]. Direct evidence of genomic loci bearing folded tetraplexes in native chromatin was obtained by BG4 ChIP-seq, which showed that only a small fraction of the quadruplexes observed in vitro are present in vivo in human cell lines and that the pattern of folded G-quadruplexes is cell-line- and cell-state-specific [85,140,141]. The original protocol was based on BG4 incubation after chromatin extraction [140] but it was subsequently combined with the CUT&TAG technology to allow in situ BG4 binding and to increase the signal-to-noise ratio [142,143]. Additionally, while both assay methods folded G-quadruplexes across a population of cells, BG4 CUT&TAG was adapted to capture genomic location of guanine tetraplex at the single cell level [144], thus allowing unprecedented insight into G-quadruplex heterogeneity in a cell population. The advantage of the adaptability and good performance of BG4 in these assays is counterbalanced by the fact that only this specific single-chain variable fragment has been widely used to study G-quadruplex prevalence in chromatin. In addition to concerns over BG4 preferential binding to certain G-quadruplex conformations over others, it is important to note that BG4 specificity is yet to be thoroughly tested. Usually, the specificity of an antibody for a ChIP experiment is evaluated in cells depleted of the protein of interest by knockout or knockdown approaches [145], which cannot really be realized with DNA secondary structures such as G4s [1]. Moreover, G-quadruplexes recognized by BG4 in crosslinked chromatin might not correspond to those actually found in live cells, skewing BG4 ChIP results even further from in vivo condition and favoring CUT&TAG-based approaches. To bypass this problem, expression of the G-quadruplex specific probe directly in cells before formaldehyde treatment should be performed. Unfortunately, this could be difficult with BG4, given that the reducing conditions in cytosol and the nucleus prevent the formation of key disulfide bonds necessary for proper scFv functionality [146–148]. On the other hand, the D1 probe was successfully expressed by transfected SiHa cells to perform ChIP-seq [133]. While D1 principally binds parallel G-quadruplexes, a recent study reported the production of a G-quadruplex-specific nanobody that was expressed in live cells to assay several G-quadruplex conformers in native chromatin by CUT&TAG [149]. Even though they are still far from perfect, the use of these probes, together with the more established BG4, should yield a less biased picture of G-quadruplex prevalence in cellulo.

3.4. *Small-Molecule G4 Stabilizers and Destabilizers*

Another fundamental tool to study G-quadruplexes in their native environment is based on small-molecule ligands. Historically, the search for compounds capable of binding to guanine tetraplexes had been linked to their possible use in anticancer therapy, given that telomerase activity was found to be inhibited by G-quadruplexes [13]. Indeed, the first G4 ligand to be reported in the literature, 2,6-diamidoanthraquinone, was shown to inhibit telomerase through stabilization of G-quadruplexes [150]. Since that discovery, the number of compounds either specifically designed to bind G-quadruplexes or repurposed from other fields has grown substantially (for some noteworthy examples, see Figure 3). Currently, about 3700 molecules are listed in the G4LDB 2.2 database as G-quadruplex binders [151,152], forming a structurally heterogeneous group of compounds that mostly stabilize these tetraplex structures. According to a recent review [153], organic G4 ligands exhibit multiple planar aromatic rings arranged into three main types of architectures: (a) fused aromatic polycyclic systems, (b) macrocycles, and (c) non-fused aromatic systems, mostly populated by modular G-quadruplex ligands. The rationale behind most ligand designs is the generation of a large planar surface capable of establishing π-π stacking with the external G-tetrads of the target (as exemplified by the NMR-derived structure of

the drug RHPS4 in complex with a parallel G-quadruplex [154]), together with cationic groups that form hydrogen bonds directly with the grooves and loops of the tetraplex assembly [155]. Different rules have been established for metallo-organic complexes, a rapidly expanding group of G-quadruplex-binding compounds reviewed in ref. [156]. Early chemistry studies had the explicit objective of developing G4 ligands that could function as antitumor drugs rather than simply studying these structures. As an example, in the 2000s, the trisubstituted acridine derivative BRACO-19 was shown to selectively recognize G-quadruplexes and inhibit telomerase in in vitro assays, achieving limited cytotoxicity in cancer cell lines, partial tumor xenograft regression in nude mice, and inhibition of the HIV-1 reverse transcription process [157–160]. Similarly, the cationic porphyrin TMPyP4 exhibited low cytotoxicity while inhibiting telomerase in several human cancer cell lines [161]. TMPyP4-mediated stabilization of G-quadruplexes in vitro occurred at slightly different affinities depending on tetraplex topology [162], although its selectivity towards quadruplex DNA was challenged in some studies [163,164]. Nevertheless, cell lines treated with the drug showed transcriptional downregulation of *c-myc* and other downstream genes in a manner dependent on G-quadruplex stabilization [165,166]. Another family of G-quadruplex-binding compounds is that of bisquinolinum derivatives, with its widely used member Phen-DC3. Its selective binding to G-quadruplexes in vitro was apparently similar to that of previously mentioned compounds and no changes in affinity to different G-quadruplex conformations were reported [124,167,168]. Interestingly, Phen-DC3 treatment of HeLa cells resulted in transcriptional changes in several genes bearing at least one sequence predicted to fold into a G-quadruplex [169], suggesting that the ligand could stabilize these secondary structures in vivo. An additional example of G4 ligand widely used in the literature is pyridostatin (PDS). After confirming its selectivity for human telomeric G-quadruplexes in vitro, PDS was shown to cause dissociation of POT1, a component of the human shelterin complex, from duplex telomeric DNA [170]. Moreover, the compound was demonstrated to induce DNA damage in cells at genomic sites enriched for putative G-quadruplex-forming sequences, to reduce the expression of genes close to DNA damage sites, and to inhibit the growth of several cancer cell lines [171,172].

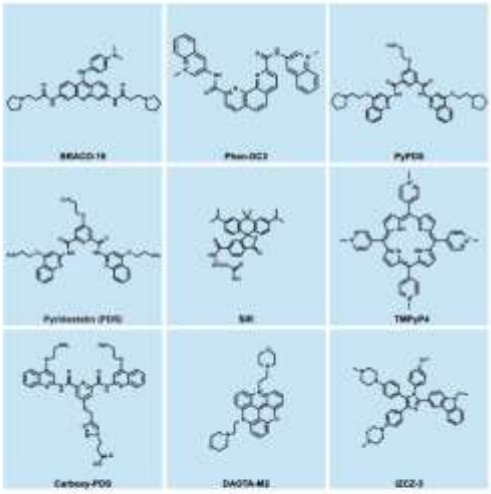


Figure 3. Chemical structures of G-quadruplex ligands mentioned in the text, as retrieved from PubChem [173], such as Braco-19 (CID: 9808666; https://pubchem.ncbi.nlm.nih.gov/compound/9808666), Phen-DC3 (CID: 44449504; https://pubchem.ncbi.nlm.nih.gov/compound/44449504), PDS

(CID: 25227847; https://pubchem.ncbi.nlm.nih.gov/compound/25227847), TMPyP4 (CID: 135442972; https://pubchem.ncbi.nlm.nih.gov/compound/135442972) and IZCZ-3 (CID: 137628645; https://pubchem.ncbi.nlm.nih.gov/compound/137628645); PyPDS and SiR have been successfully fused into SiR-PyPDS [174]. Carboxy-PDS, while structures of Carboxy-PDS and DAOTA-M2 were taken from their respective papers [131,175].

A key advantage of small-molecule G-quadruplex ligands is the possibility of using them on live cells to explore the effects of tetraplex stabilization or to just exploit their ability to recognize folded G4s in native chromatin. Several compounds have been modified or purposefully designed to function as fluorescent probes in cells, establishing a new way to image G-quadruplexes without fixing cells and reducing the likelihood of preventing tetraplex-interacting proteins from binding the structure. For example, DAOTA-M2 has been recently described as a fluorophore capable of binding G-quadruplex in live and fixed cells [175]. G-quadruplex dynamics with this probe was assessed by Fluorescence Lifetime Imaging Microscopy (FLIM) with limited cytotoxic effects [175]. Similarly, the pyridostatin derivative SiR-PyPDS has been synthesized by tethering the fluorophore silicon-rhodamine (SiR) to the PyPDS scaffold [174]. This probe showed specific binding to several G-quadruplex oligonucleotides without significantly inducing G-quadruplex folding [174], a condition that would have altered tetraplexes in cells and complicate interpretation of results. Low nanomolar concentrations of SiR-PyPDS were used to perform single-molecule imaging in live cells, allowing scientists to study folding and unfolding events of single G-quadruplexes in situ [174]. Alternatively, click chemistry on G4 ligands already inside cells can be used to generate a fluorescent probe. A derivative of Phen-DC3 was labeled with the fluorophore Cy5 in fixed cells to show drug accumulation in the nucleus [176], while a similar feat was accomplished by linking a pyridostatin scaffold to an Alexafluor azide moiety [171]. A similar strategy was exploited to capture G-quadruplex-interacting proteins in cellulo. Two studies reported a screening approach where a small tetraplex ligand is photo-crosslinked to nearby proteins which are later identified by mass spectrometry analysis [177,178]. These techniques have proven successful in identifying new candidate proteins that modulate G-quadruplexes in native chromatin, expanding the current view of how cells regulate these structures [177,178]. Moreover, the same approach can be used without much modification by changing the G-quadruplex ligand, thus allowing researchers to reduce bias and to target specific nucleic acid structures or a subset of their conformations. Indeed, thanks to the availability of numerous G-quadruplex structures showing different conformations and topologies, small molecules can be designed and optimized for selective recognition of one or few of the known structures, enabling modulation and investigation of only tetraplexes with a specific conformation to provide insight into the biological relevance of G-quadruplex polymorphism. This strategy was successfully applied for various compounds, although target binding for most of them was evaluated only in vitro [54,55,179–181]. A successful example is IZCZ-3, a carbazole/imidazole derivative that specifically binds a parallel *c-myc* G-quadruplex over other tetraplex conformations [182]. The compound was shown to downregulate *c-myc* transcription through stabilization of its G-quadruplex in cells and to inhibit cancer cell growth by inducing cell cycle arrest [182]. On the other hand, optimization of the pyridostatin scaffold led to the synthesis of carboxyPDS, which was shown to selectively bind RNA G-quadruplexes over DNA tetraplexes in vitro and in cellulo [131,183], allowing precise investigation of the biological role of RNA tetraplexes. Overall, chemical biology methods to study G-quadruplexes have flourished in the past two decades thanks to an increase in the available G-quadruplex structures. Time and time again, small-molecule ligands have proven fundamental in investigating G-quadruplex prevalence and dynamics in live cells, which would have been impossible to do with antibody-based methods. Additionally, the modularity of certain scaffolds, the possibility of performing click chemistry in situ, and the rational design and optimization of these probes allowed the synthesis of thousands of molecules, some specifically targeting certain G-quadruplex conformations over others. This vast collection of compounds was shown to stabilize guanine tetraplexes in vitro and

in cells, providing new insight into the effects of altered G-quadruplex dynamics. Considerable effort has gone into the analysis and optimization of G-quadruplex stabilizing probes, thus leaving the interesting avenue of G4-destabilizing compounds mostly unexplored. To date, only a few drugs have been shown to alter and unfold G-quadruplexes; however, the conclusions drawn from these experiments are still debated (reviewed in ref. [184,185]). Interestingly, one such compound was shown to upregulate *c-kit* transcription [186], which was previously shown to be reduced upon G-quadruplex stabilization [187]. In light of the possibility of investigating the effects of reduced G-quadruplex levels, more effort should be devoted to the discovery and synthesis of these compounds.

4. Prevalence of Guanine Tetraplexes in Genomes

After having discovered G-quadruplexes and probed their structure in numerous studies, the next logical step consisted of understanding where they are found in genomes and if their position has with a possible function. As explored previously, the most straightforward method used to identify G-quadruplexes in genomes is based on bioinformatic prediction. By virtue of knowing the structural requirements of G-quadruplex formation, it is possible to identify which sequences have the propensity to fold into G-quadruplexes and identify them in a genome of interest. Even the most conservative algorithms for putative G-quadruplex-forming sequence (PQS) yielded more than 375,000 hits in the human genome [110,111], with more recent strategies taking into account non-canonical G-quadruplex structures and machine learning to improve sensitivity [112], thus increasing the number of genomic sites that could potentially fold into canonical and non-canonical G-quadruplexes. Furthermore, the combination of bioinformatic prediction and high-throughput polymerase stop assays coupled to next generation sequencing [121,123] greatly expanded the view of G-quadruplex-forming sites in the human genome, revealing that transcription start sites, 5′UTR, and splicing regions have the highest enrichment of these structures, while comparatively few of these structures are found in coding sequences [121,188,189]. The rG4-seq dataset on the human transcriptome revealed that G-quadruplexes in mRNA have the highest density in 5′ and 3′ UTRs compared to CDS (in accordance with G4-seq data) [123].

Importantly, the same strategy was applied to various genomes across diverse species to provide insight into taxa-specific characteristics and the evolution of G4 motifs. Putative G-quadruplexes were found to be enriched in gene regulatory regions of several prokaryotic genomes, indicating a role in transcriptional regulation of specific gene classes [190–197]. For example, this link was confirmed for the specific case of radioresistance genes in *Deinococcus radiodurans*, whereby upon treatment with the small-molecule G4 ligand N-methyl mesoporphyrin (NMM), a drop in target gene transcription and in resistance to radiation was observed [198]. Bioinformatic analysis of human viral genomes revealed that putative G-quadruplex sequences were found in most virus categories, particularly in single-stranded genome viruses, with the dsDNA *Herpesviridae* family being a notable exception [199]. An example of the functional relevance of G-quadruplexes in viral genomes was provided in a reporter assay of PQS-containing human cytomegalovirus (HCMV) regulatory sequences, where chemically induced G-quadruplex stabilization resulted in transcriptional suppression of some of the tested viral promoters [200], suggesting that these secondary structures can be used as targets for specific antiviral drugs (reviewed in ref. [201,202]). Interestingly, a fraction of the predicted G4-forming sequences from the unicellular eukaryote *Saccharomyces cerevisiae* was found to be conserved in other fungal species [203], with these sequences being enriched in promoters, telomeres, rDNA, mtDNA, and sites of frequent double-strand breaks [203,204]. In contrast, among several plant species, the prevalence of potential G-quadruplex-forming motifs shows great variability, with the highest density observed in monocots and lycophytes compared to dicots and bryophytes (although this could be due to the difference in GC content between plant genomes) [205]. Moreover, G4 motifs with stretches of two guanines are strikingly more abundant than the standard G_3 sequences, and putative G-quadruplex location dramati-

cally favors intergenic regions over coding regions [205,206]. Despite this, about a fifth of annotated *Arabidopsis* genes harbor at least a putative G-quadruplex in their promoter, as well as 68% of gene models (i.e., all possible ORFs, including splice variants) [206], while in *Zea mays*, enrichment of G4 motifs was found to be particularly high in the proximity of the transcription start site [207]. Thus, the possibility of G-quadruplexes regulating gene expression at the transcriptional or post-transcriptional levels in plants cannot be excluded. Similarly, several metazoan species reportedly show a significant proportion of annotated genes with putative G-quadruplex-forming motifs found in the 2 kb upstream region [208], although the gene fraction in representative genomes from warm-blooded animals was much higher than that from cold-blooded animals [209]. A more dramatic difference between these animal groups was also reported when considering the frequency of predicted G-quadruplexes in a 1000 bp window around the transcription start site [209]. These findings were later confirmed and expanded upon using an optimized G4-seq protocol, which showed that sequences that can fold into G-quadruplexes in near physiological conditions are more common in promoters and 5′UTRs in human, mice, and *Trypanosoma* genomes, while these observed G-quadruplexes are not significantly enriched in *C. elegans*, *D. rerio*, or *D. melanogaster* [122]. Overall, pattern-matching algorithms for putative G-quadruplex searches revealed that these structures are widespread across evolutionarily distant genomes, in all domains of life and viral species alike. These studies show that the overall G-quadruplex density between genic and intergenic regions changes significantly across taxa [122], as does the G4 motif, particularly in loop length and variability of the core sequence [122,205,208]. Despite this, multiple reports underline that putative G-quadruplexes, as well as assemblies validated by G4-seq, are present at a significant level in gene regulatory regions, although the extent of this enrichment is extremely variable across taxa [122,190,199,204,205,208,209]. Importantly, one study [208] found that the location of putative G4 sequences shifted from evenly distributed to clustered at different chromosomal loci from basal eukaryotic species to higher organisms such as humans and *Gallus gallus*. Moreover, G4 motifs within the 2 kb regulatory region upstream of genes tended to accumulate across evolution in loci coding for transcription factors. This suggests that, during the evolution of eukaryotic species, G-quadruplexes gradually acquired a relevant function in gene expression regulation, particularly at the transcriptional level, and, therefore, G-quadruplex-forming sequences were slowly concentrated at strategic regulatory regions close to genes. The fact that genes coding for transcription factors show an abundance of putative G-quadruplex sequences [208] would then explain why G-quadruplexes have been linked to the wide variety of biological functions [208], which will be discussed in a later section. In the literature, various cases of specific G4 loci that are phylogenetically conserved have been reported. To report some examples, a validated G4 locus in the *RPB1* gene coding for the large subunit of RNA polymerase II has been found to be conserved in the Archaeplastida plant supergroup [210]. Additionally, a G-quadruplex-forming sequence was found in the 5′UTR of the *NRAS* proto-oncogene's mRNA in at least six different mammalian species, where it may have a role in translational regulation [211]. To sum up, G-quadruplexes are a feature common to all genomes analyzed so far, irrespective of species complexity. Moreover, these structures do not appear to have been selected against during evolution, but rather were progressively co-opted, likely to serve as a new layer of gene expression regulation.

5. Recognition and Modulation of G-Quadruplexes in Cells

As mentioned previously, only a small fraction of all sequences capable of folding into G-quadruplexes actually forms tetraplexes in vivo [85]. Indeed, the cell type and differentiation state are correlated to a different G4 repertoire [140,141]. In addition, since G-quadruplexes are not under negative evolutionary pressure to be eliminated, they have likely been domesticated to perform some kind of biological function [208]. This implies that the cell must be able to recognize these structures and induce their folding or unfolding when and where needed. G-quadruplex polymorphism may thus be biologically relevant,

meaning that cells would be capable of discriminating between different tetraplex topologies and conformations. It is therefore not surprising that a large number of proteins have been discovered that are capable of G-quadruplex binding.

5.1. Approaches to Identify G-Quadruplex-Binding Proteins

Among the several strategies that have been used for the identification of tetraplex-binding proteins, the most common relies on affinity pull-down of cell extracts using DNA or RNA G-quadruplex oligonucleotides [212–214]. This approach heavily depends on the type of bait being used and cannot capture protein-binding events that are modulated by flanking sequences or the chromatin context. These issues have been partially addressed with the development of co-binding mediated protein profiling assays, where proteins interacting with G-quadruplexes in native chromatin are identified through photo-crosslinking to a G4 specific probe [177,178]. On the other hand, an alternative indirect method consists of identifying proteins whose binding sites, assessed by ChIP, contain a putative G-quadruplex forming sequence [137,215]. However, additional assays are required to validate the actual binding to folded G-quadruplexes rather than the corresponding dsDNA sequence. A third type of screening is based on bioinformatic analysis of protein sequences from a wide collection of previously identified tetraplex interacting proteins to reveal common motifs. A recent study highlighted the enrichment of RGG motifs in G4-binding proteins [216], which was shown to mediate interaction with the nucleic acid structure through hydrogen bonding and π-stacking provided by key arginine and phenylalanine residues [217]. The authors propose that putative G-quadruplex-binding proteins could be more rapidly identified through the presence of the RGG domain in their sequence. Although compelling, this approach heavily relies on a specific protein domain to predict G-quadruplex binding of uncharacterized proteins. Considering the current state of G-quadruplex–protein interaction studies, it is safe to assume that new domains and motifs will be identified as mediators of tetraplex binding in the future, leading to less biased predictions. Nevertheless, candidate protein–G4 binding must be confirmed in in vitro assays, while the biological relevance of the interaction should be explored via protein depletion or overexpression in cell lines or other biological systems.

5.2. Established G-Quadruplex-Binding Proteins

As of 26 February 2024, the G4IPD database for G-quadruplex interacting proteins [218] (available at https://iiti.ac.in/people/~amitk/bsbe/ipdb/index.php, accessed on 26 February 2024) lists about 60 and 40 entries that recognize DNA and RNA G-quadruplexes, respectively. This collection of factors was demonstrated to either stabilize or destabilize the tetraplex structure (reviewed in ref. [219] and visually grouped in Figure 4). The latter is mainly performed by G-quadruplex-unwinding helicases (reviewed in ref. [220]). Among these, the mammalian RecQ helicases WRN and BLM have been studied for more than two decades. Both proteins were shown to unwind both dsDNA substrates, as well as G-quadruplex forming oligos in vitro in 3′ to 5′ directionality and in the presence of ATP, provided they have a centrally located ssDNA region and a short 3′ ssDNA overhang [221–224]. In a single molecule FRET assay, the mechanism of BLM binding to the substrate was found to involve both the RecQ C-terminal (RQC) and Helicase-RNase D C-terminal (HRDC) domains, with the latter recognizing the ssDNA region and aiding RQC binding and unfolding of the G-quadruplex in an ATP-independent manner [224]. In another study using smFRET, it was demonstrated that BLM could recognize and unwind substrates regardless of tetraplex conformation, while WRN was shown to selectively destabilize only telomeric G-quadruplex substrates [53]. Helicases are not only involved in modulation of genomic G-quadruplexes, but also process RNA tetraplexes (reviewed in ref. [225]), as in the case of DEAH box protein 36 (DHX36). Also known as RNA helicase associated with AU-rich element (RHAU) and G4 Resolvase 1 (G4R1), the helicase unwinds both RNA and DNA G4s in vitro [226]. It was found to selectively recognize parallel tetraplex structures through its short N-terminal domain interacting with the

exposed hydrophobic surface of a terminal G-tetrad and electrostatic interactions with backbone phosphate groups [227,228]. Mechanistic analysis of DHX36 activity revealed that it requires loading on a 3′ ssDNA overhang and then progressively unwinds substrate G4s in a 3′ to 5′ directionality through ATP hydrolysis, similar to other DEAH-box helicases [229]. Overall, BLM, WRN, DHX36, and all the other G-quadruplex helicases have been found to destabilize guanine tetraplexes in vitro and in vivo, where the unwinding activity has been linked to DNA replication and gene expression regulation. Indeed, it has long been hypothesized that G-quadruplex formation in the template strand would induce DNA or RNA polymerase stalling, leading to downregulated gene expression [230–232] and genome instability [233–235]. These consequences will be discussed extensively in following sections.

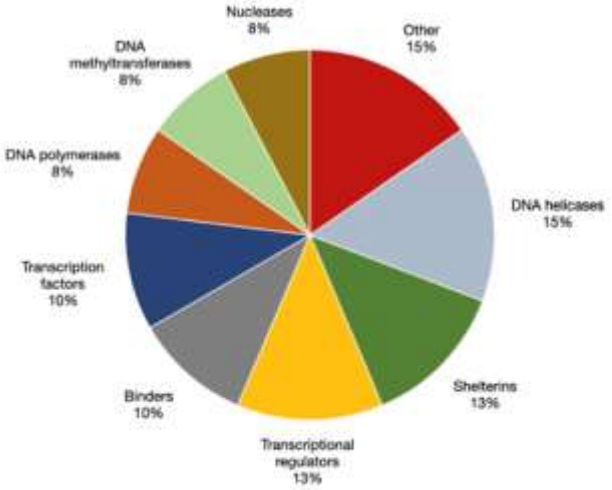


Figure 4. Classes of DNA G-quadruplex binding proteins listed in G4IPD.

Besides helicases, other proteins participate in G-quadruplex binding and regulation. A curious example is nucleolin, a multifunctional phosphoprotein commonly found in nucleoli that was shown to bind and stabilize G-quadruplexes in the *c-myc* upstream regulatory element $NHEIII_1$ [236]. Interestingly, nucleolin binding in vitro was shown to be dependent on the presence of at least one long loop (more than 3 nt) in the structure, independently of the underlying sequence or tetraplex conformation [52]. This suggests that the chaperone activity of nucleolin towards G-quadruplex-forming sequences is targeted mainly or exclusively at those loci where these long loops are formed [236]. Moreover, nucleolin-mediated stabilization of G-quadruplexes was reported to stimulate or inhibit transcription of *c-myc* in vitro [236,237], and VEGF in vitro and in vivo [238], two genes containing guanine tetraplexes in their promoters. Nucleolin is not alone in stabilizing G-quadruplexes; LARK and apurinic/apyrimidinic endonuclease 1 (APE1) were found to participate in G-quadruplex folding and stability, a function that was closely tied to transcriptional regulation [239,240]. Besides transcriptional modulation, G-quadruplex stabilization is involved in telomere DNA maintenance.

Not all G-quadruplex-binding proteins mainly function to regulate the stability of these secondary structures. An increasing proportion of these are recruited at G-quadruplex-forming sites through direct binding to the tetraplex to perform their molecular functions on other proteins or on nearby nucleic acids. For example, the telomeric repeat binding protein TRF2 was shown to bind simultaneously telomeric dsDNA and tetraplex telomeric

RNA TERRA in vitro, opening the possibility of TRF2 functioning in vivo as glue between the two and mediating TERRA localization in telomeric DNA [241]. Moreover, in vitro binding assays showed that three human DNA methyltransferases (DNMT1, DNMT3A, and DNMT3B) could selectively bind to G-quadruplexes [242], while DNMT1 occupancy at chromatin sites in K562 cells was revealed to be enriched at hypomethylated G-quadruplex forming sites, as evaluated by BG4 ChIP-seq [243]. Based on these results, the authors proposed that folded G-quadruplexes in chromosomal DNA sequester DNMT1, inhibiting its activity and protecting nearby sequences from methylation [243]. While this inhibitory mechanism is yet to be demonstrated in vivo, this compelling evidence links guanine tetraplexes to epigenetic regulation through protein recruitment at folded G-quadruplexes in the genome, a topic that will be discussed further in a later paragraph.

To summarize, a wide variety of proteins has been shown to physically interact with G-quadruplexes, both in vitro and in vivo. A large fraction of G-quadruplex-interacting proteins identified so far regulate the stability of the structure by actively unwinding it or promoting its folding, showing that the cell is capable of regulating G-quadruplex dynamics according to its needs. Interestingly, recent studies have revealed the existence of proteins that function as G-quadruplex readers, in that they bind selectively to these folded structures to concentrate their activity at specific genomic regions. The cell's ability to modulate and recognize its guanine tetraplexes in a conformation- and topology-specific manner confirms that these are bona fide structures in nucleic acids. It is important to note that a deeper understanding of G-quadruplex biology relies on the identification and characterization of new G-quadruplex-interacting proteins. Several proteome screening methods have been developed to expand the list of these proteins, each with its own advantages and drawbacks. As new G4-interacting proteins are identified and studied, our understanding of G-quadruplex biology will inevitably change. As of now, hypotheses and models on the functions of tetraplex assemblies inevitably paint an incomplete picture, which is, however, constantly and rapidly updated as more research is performed in this area.

6. Physiological and Pathological Roles of G-Quadruplexes

6.1. G-Quadruplexes in Transcriptional, Post-Transcriptional, and Epigenetic Regulation

Having established that G-quadruplexes are biologically relevant, the main functions they are involved in will be presented and are recapitulated in Figure 5. One of the most extensively studied areas in this context is transcription. As early as the start of the millennium, G-quadruplex involvement in transcriptional regulation of oncogenes was identified [166]. Additionally, as presented previously, the first computational predictions of guanine tetraplexes from newly sequenced genomes revealed that these putative G4-forming sequences were abundant in gene promoters [110,111]. This feature was confirmed in BG4 ChIP-seq experiments [85,141,244], further strengthening the link between guanine tetraplexes and transcription. The original molecular model considered folded G-quadruplexes simply as obstacles to RNA polymerases, thus causing a reduction in transcript levels. This idea was built on data from in vitro transcription assays showing RNA polymerase arrest when elongating G-rich or G-quadruplex-forming templates [245,246]. Although this model was reinforced by the demonstration that genes bearing G-quadruplexes near their transcription start sites were downregulated upon G4 stabilization by small-molecule ligands in cells [165,166,187,247], this could be the result of indirect consequences of global guanine tetraplex stabilization [248,249]. Over the years, the role of G-quadruplexes in transcription has been expanded greatly, painting a more nuanced picture of where G-quadruplexes exert transcriptionally stimulating or silencing effects in a context-dependent manner. A key finding was the characterization of transcription factors that bind folded G-quadruplexes and thus could be recruited at certain genomic sites [215,250–252]. Moreover, folded G4s as detected by BG4 ChIP-seq in HaCaT cells were almost exclusively found in nucleosome-depleted regions of genomic DNA [85], a chromatin feature that is common at eukaryotic transcription start sites and that more broadly

modulates DNA accessibility to transcription factors (reviewed in ref. [253]). Although a clear causal link has yet to be established, it is possible that G-quadruplex formation at selected genomic sites prevents nucleosome assembly, hinting at another mechanism of gene expression regulation by these secondary structures [254]. Overall, while G-quadruplex involvement in all stages of transcription has been extensively reported and several potentially compatible models of the underlying mechanisms have been proposed, solid evidence of direct modulation of chromatin states by guanine tetraplexes is lacking. Nevertheless, exploration of the transcriptional role of these structures is currently underway, especially considering the clinical importance of this mechanism.

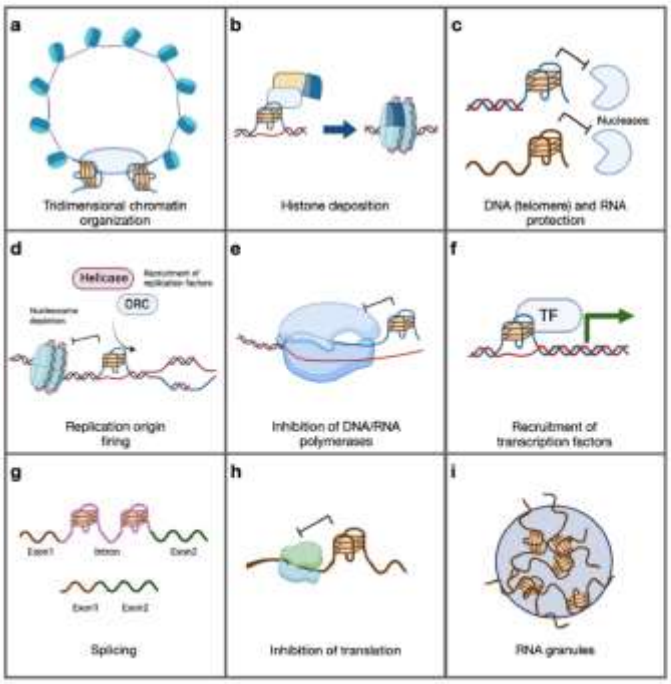


Figure 5. Cellular roles of G-quadruplexes. Thanks to their capability to distort the nucleic acid structure and to interact with a plethora of proteins, G-quadruplexes have been ascribed a role in several key cellular processes, comprising, but not limited to organization of chromatin in topologically associating domains (TADs) (**a**), deposition of non-replicative histone variants (**b**), inhibition of DNA and RNA nucleases (**c**), firing of replication origin (**d**), inhibition of replication and transcription (**e**), regulation of transcription (**f**), RNA maturation (**g**), regulation of translation (**h**), and formation of RNA stress granules (**i**).

The role of G-quadruplexes in gene expression regulation is not limited to transcription, as evidence of their involvement in RNA biology and translation has been reported [2,255]. Although mainly studied in the context of mRNA, G-quadruplexes have also been detected in diverse classes of non-coding RNAs, such as lncRNAs [256], pri-miRNAs [257], and piRNAs [258]. Guanine tetraplex assemblies have been reported to play an important role in pre-mRNA maturation, particularly in the context of alternative splicing, as shown for FM1, p53, and hTERT primary transcripts [259–262]. It has been demonstrated that G-quadruplexes recruit splicing factors, such as hnRNPH [263] and hnRNPF [264], and that

global destabilization of RNA guanine tetraplexes results in widespread splicing deregulation [265]. Besides mRNA maturation, its positioning within the cell is also regulated by G4s, for example, functioning as a signal for neurite targeting of selected transcripts [266]. Another intriguing example is their involvement in phase separation, as structured G-quadruplexes in C9ORF72 mRNA were found to induce stress granule assembly [267]. Similarly, a recent study by Wang and Xu demonstrated that short RNA molecules bearing CGG or telomeric repeats aggregate into solid state foci, both in vitro and in vivo, through the formation of intermolecular higher-order G-quadruplex structures [268]. Furthermore, RNA G-quadruplexes have been shown to have a direct effect on translation itself, for example, by preventing recruitment of the 43S ribosome and reducing translation efficiency unless unwound by the translation initiation factor eIF4A [269]. Another mechanism that has been reported in the literature is translational repression of mRNAs harboring rG4s that induce ribosome positioning at uORFs [270]. In one of the first instances where this phenomenon was observed, the RNA G-quadruplex in the 5′UTR of *NRAS* was shown to inhibit translation of a reporter gene in vitro [211]. To generalize, tetraplex assemblies in 5′UTRs of mature transcripts have been shown to modulate translation rates, albeit in a context-dependent manner rather than based solely on sequence or 3D structure [271].

Another layer to gene expression regulation where G-quadruplexes have been implicated is in chromatin organization and epigenetics (reviewed in ref. [6]). As discussed previously, G4s are enriched in nucleosome-depleted regions of transcriptionally active genes, suggesting a role in nucleosome positioning. This concept has been explored in the context of epigenetic heritability during DNA replication, where chemical G-quadruplex stabilization was tied to loss of H3K4me3 and subsequently addition of methylated cytosines. This epigenetic reprogramming was maintained throughout cell generations and was shown to directly repress transcription of nearby genes [272]. Non-replicative histone deposition is also influenced by G-quadruplexes through recruitment of histone chaperones, as has been demonstrated for ATRX-mediated deposition of H3.3 at telomeres [114,273] and the removal of the H2A-H2B dimer performed by G-quadruplex-binding nucleolin [274]. Moreover, histone post-translational modifications can be modulated in a G-quadruplex-dependent manner through the recruitment of histone modifying enzymes at sites of folded G-quadruplexes. Besides the cases of DNMT1, 3A, and 3B [242,243], which were explored in a previous section, this has additionally been shown by G4-mediated recruitment of the LSD1–CoREST complex at the hTERT gene promoter, where it catalyzes the demethylation of H3K4/9 [275]. On the whole, multiple studies have directly or indirectly linked folded G-quadruplexes to chromatin remodeling and epigenetic factor recruitment to affect local chromatin structure. In recent years, it has been suggested that the concept of guanine tetraplexes as protein recruitment hubs could also be expanded to include regulation of long-distance chromatin contacts, particularly pertaining to distal promoter–enhancer interactions [249]. The validity of this hypothesis was strengthened when the combination of BG4 ChIP-seq and chromatin conformation capture data revealed an enrichment of folded G-quadruplexes at topologically associated domain (TAD) boundaries [276], which correspond to regions of gDNA that are distant in linear sequence but are found to be closely associated in the three-dimensional organization of chromatin thanks to the DNA looping factors cohesin and CTCF [277,278]. Furthermore, binding sites of several architectural proteins were enriched in G-quadruplex-containing TAD boundaries, pointing to G-quadruplex-dependent loop formation via their recruitment [276]. These correlations were confirmed by Li and colleagues in a 2021 paper, where the transcription factor and DNA looping protein YY1 was shown to bind G-quadruplexes in cells and dimerize to generate long-distance chromatin contacts partly due to guanine tetraplexes to then mediate gene expression regulation [279].

Overall, G-quadruplexes have been implicated in both transcriptional and post-transcriptional regulation by a variety of proposed or demonstrated mechanisms, making the original claim that these structures work simply as physical blocks to RNA polymerase obsolete. While more data on the native context of these processes are needed to faithfully

explore G4 involvement in gene expression regulation, it can be nonetheless stated that a critical though not exclusive component of their role relies on recruiting proteins (transcription factors, splicing factors, translation factors, etc.) in selected genomic or RNA regions. This highlights further that understanding G-quadruplex biology at a deep level necessarily requires the identification and characterization of G-quadruplex-binding proteins.

6.2. Impact of G-Quadruplexes on DNA Replication

As for other non-canonical secondary structures, the presence and dynamics of G4s in nucleic acids have a significant impact in the context of DNA integrity and replication (reviewed in ref. [280]). As mentioned earlier, dsDNA unwinding during transcription or replication exposes single-strand DNA, favoring G-quadruplex folding, which creates obstacles to DNA polymerase progression, as shown in in vitro polymerase stop assays [281]. The concept of guanine tetraplexes causing DNA polymerase stalling in vivo has been strongly hinted at in numerous studies, with the underlying molecular mechanisms and the strategies adopted by biological systems to prevent and resolve these blocks being hypothesized or experimentally proven. Evidence that folded guanine tetraplexes induce replication fork arrest in vivo has been described in various biological systems, ranging from *X. laevis* egg extracts [282] to *S. cerevisiae* yeast cells [139], a patient-derived cell line [234], and whole *C. elegans* animals [283]. The common observation of all these studies is that DNA replication is negatively affected by the presence of these secondary DNA structures. When cells are unable to unwind G-quadruplexes or resume replication fork progression, DNA deletions were observed, with sizes ranging from 60–300 nt (similar to Okazaki fragments) [283] to up to 80 kb [234]. In mechanistic explorations of this phenomenon, G-quadruplex-unwinding helicases are often implicated as critical factors in resolving the tetraplex and avoiding replication fork stalling [139,282]. Additionally, dedicated DNA polymerases are specifically recruited at the stalled fork to resume elongation downstream of the folded structure. This has been shown for REV1 [284] and PrimPol, the latter having G-quadruplex-binding ability in vitro and being able to reprime the template downstream of the guanine assembly [285]. However, when these factors are depleted or inactive, the G-quadruplex becomes an obstacle to replicative fork progression, leaving an exposed ssDNA stretch that eventually generates a double-strand break [286,287]. Similarly, chemical stabilization of G-quadruplexes using the ligand pyridostatin was correlated to induction of double-strand breaks as evaluated by neutral comet assay [171]. Moreover, it was reported that a single unresolved G4 remains stable throughout mitosis and is inherited by daughter cells, which, when undergoing S phase, use the same G4-containing strand as a template and thus are exposed to the same type of DNA lesion [286].

Having just explored the negative influence of folded G-quadruplexes on replication fork progression, it may be surprising to learn that tetraplex assemblies were shown to participate in replication origin firing in eukaryotes. Indeed, deep sequencing of short nascent strands (SNSs) in human cell lines revealed that active replication origins are enriched at transcription start sites of CpG-rich promoters. Importantly, most of the identified origins map to sites of putative G-quadruplex-forming sequences (PQS), with PQS density and origin efficiency showing a positive correlation [288]. Confirmation of a causal relationship was obtained by studying origin firing in a normally late-replicated locus where the PQS-containing chicken β^A origin was inserted. In this genetic background, the transplanted origin induced an enrichment in SNS signal while replication timing remained unaffected. Moreover, point mutations that destabilized the predicted G-quadruplex caused a drop in origin firing efficiency, directly tying folded tetraplexes to replicon activity. Importantly, G4 orientation in this system determined the direction of the replication initiation site [289], consistently with previous observations of SNS peak enrichment within 200 bp 3′ of a PQS [290]. Although these results established that G-quadruplexes regulate origin activity in metazoans, the molecular mechanism behind this phenomenon remains obscure. Currently, three main hypotheses have been proposed, using models of replication start sites derived from yeast studies as a foundation. Firstly, it has been observed that

origins are enriched in nucleosome-depleted regions [291,292], a chromatin feature that has also been associated with G-quadruplexes [85,293] (although the causal relationship is yet to be defined). Another possibility lies in the fact that folded G-quadruplexes may function as recruitment hubs for factors involved in the initiation of replication. Indeed, the human Origin Recognition Complex (ORC) was found to specifically recognize tetraplex assemblies in RNA and ssDNA [294]. Furthermore, it is possible that G-quadruplex-unwinding helicases are recruited in folded tetraplexes near eukaryotic origins in order to remove the guanine assemblies and later unwind dsDNA to allow DNA replication to begin. This model would explain why folded G-quadruplexes determine the position of origin firing, since helicases have a defined unwinding directionality [280]. Overall, the idea that guanine tetraplexes play a role in replication origin activation and in determining the directionality of the initial dsDNA denaturation is intriguing. This phenomenon is in contrast with the simplistic notion that these secondary DNA structures are simply obstacles to DNA replication; instead, it highlights a more nuanced function in the wider context of genome maintenance. Further research is needed to elucidate how G-quadruplexes contribute to replication origin firing at the molecular level, particularly with regards to protein factors that are selectively recruited at those sequences.

6.3. G-Quadruplexes and the DNA Damage Response

Understanding how cells repair lesions caused by G-quadruplexes is important to comprehend and predict the consequences of impaired G4 removal. In general, the threat DNA lesions pose to genome integrity is avoided thanks to a complex system where the wide variety of repair pathways are specifically activated at damaged genomic sites through the DNA damage response (DDR) (reviewed in ref. [295–298]). In most models of DDR processes, the DNA lesion is represented as a signal that is detected by sensor proteins (upstream phosphorylation events), which then directly or indirectly transduce and amplify the signal to activate a wide variety of effectors to efficiently repair the damage or activate cell cycle arrest and apoptosis when that is not possible. The DDR signaling cascade is extremely complex in human cells; however, it is worth mentioning a few of its key players. Three phosphoinositide 3-kinase (PI3K)-related kinases (PIKKs) are principally involved in initiating the signaling cascade in eukaryotes: Ataxia-Telangiectasia Mutated protein (ATM), ATM- and Rad3-related protein (ATR), and DNA-dependent protein kinase (DNA-PK). Although acting in slightly different scenarios, all three apical kinases are specifically recruited at sites of DNA damage by the co-factors NBS1 (a MRN complex component), ATRIP, and Ku80, respectively. After autophosphorylation and other activating modifications, PIKKs trigger the signaling cascade by phosphorylating a wide range of targets, thus transmitting the signal to eventually activate terminal effectors. Among the earliest temporally, H2A.X phosphorylation at Ser139, principally mediated by ATM at DSB sites and referred to as γH2A.X, is a robust marker for DNA damage in cells [299,300]. H2A.X is a variant of the canonical core histone H2A that is central to most DDR pathways in that it sustains DDR signaling and spreads to nearby chromatin to organize an efficient response. In the literature, γH2A.X has often been described as a recruitment hub for several DNA damage-response proteins, as has been shown for BRCA1, P53BP1 and MDC1 [301–303]. This serves as both a mechanism to locally increase the concentration of repair factors at sites of DNA lesions and tether broken DNA ends together, as well as for generating a positive feedback loop to further enhance DDR signaling. Moreover, the response is articulated so as to activate cell cycle checkpoints and stop its progression. ATR and ATM phosphorylate the checkpoint kinases Chk1 and Chk2, respectively [304,305]. Both effector kinases target the protein phosphatase Cdc25A and induce its degradation, thus preventing removal of inhibitory phosphates from CDK2 [306,307]. Crucially, both ATM and ATR were shown to directly or indirectly stabilize the oncosuppressor p53 [308–311], which then up-regulates genes involved in cell cycle checkpoint activation and apoptosis. Overall, through activation of apical kinases, DDR signaling triggers a multitude of parallel and partially redundant pathways to efficiently activate the appropriate DNA-repair pathway and coor-

dinate a cell cycle arrest and apoptotic response to preserve genome integrity. The system is particularly robust since inactivation of a single pathway does not generally prevent the activation of many others with similar outcomes. However, an increased DNA lesion rate caused by the inability to remove obstacles to polymerases, such as stable G-quadruplexes, would increase the burden on this system, ultimately resulting in chromosome fragility.

In a 2014 paper, a dedicated DNA-repair pathway was implicated in the processing of damaged DNA resulting from failed G-quadruplex removal in *C. elegans*. According to the authors, the resulting double-strand break is repaired through single-nucleotide micro-homology and subsequent extension performed specifically by DNA polymerase θ, resulting in small deletions [287]. In addition, the presence of G4 structures can influence DNA-repair processes. For example, in the proximity of DSB breaks, these structures can inhibit ends processing, altering the balance between the Homologous Recombination (HR) and Non-Homologous End Joining (NHEJ) repair pathways [312,313]. Furthermore, G-quadruplexes promote genomic instability when combined with R-loops on the opposite DNA strand, forming a G-loop [75]. It has been shown that global G4 stabilization in a cancer cell line using two chemically distinct compounds leads to activation of the DNA damage response, as evaluated by H2A.X and ATM phosphorylation, as well as RAD51 and p53BP1 nuclear foci formation [314]. Interestingly, the same study reports a concomitant increase in R-loop levels 2 to 10 min after treatment with G4 stabilizers, showing that overexpression of RNase H1 in this system abrogates DNA damage induction as indicated by stable levels of γH2A.X [314]. Although the molecular mechanism behind G-loop-dependent DNA lesions has yet to be elucidated, the authors propose that the stabilizing effect G-quadruplexes have on R-loops could increase the risk of transcription–replication conflicts or of activating a recombination pathway that produces double-strand breaks [314,315]. Overall, guanine tetraplexes have been shown to block DNA (and RNA) polymerases in DNA replication and transcription, thus increasing the risk of damaging DNA [233]. Several factors have been discovered that prevent G4-mediated instability through tetraplex unwinding, repriming, or translesion synthesis. However, when one of these structures cannot be removed, DNA lesions are formed through a yet undefined molecular mechanism, therefore activating DNA damage-repair and recombination pathways. Research in this area is focused on precisely dissecting how G4-dependent DNA lesions are formed, as well as on analyzing the pathways that counteract genomic instability induced by guanine tetraplexes. Aberrant responses to tetraplex-induced DNA damage are particularly relevant in pathological conditions, a concept that will be explored later in this review.

6.4. G-Quadruplexes in Telomeric Regions

It is noteworthy to consider that the earliest studies on G-quadruplexes were focused on telomeric sequences. Telomeric DNA is located at the ends of eukaryotic chromosomes and one of its defining features is the presence of short species-specific G-rich tandem repeats. Importantly, in order to prevent improper recombination events through chromosome termini, telomere ends are protected from DNA damage-repair proteins. In vertebrates, this is achieved through the formation of T-loop structures, consisting of the ssDNA 3′ overhang invading an upstream telomeric repeat dsDNA thanks to the shelterin protein complex [316]. Owing to the presence of repeats rich in guanines, several three-dimensional structures of telomeric G-quadruplexes have been reported in the literature. These studies originally used ciliate telomeric repeats, but have since been expanded to include human telomeric sequences [18,317,318]. Eventually, telomeric DNA G-quadruplexes were observed in cells thanks to BG4 and D1 immunostaining [130,133], prompting scientists to study the influence of these secondary structures on telomere biology. Indeed, it was found that a collection of telomeric proteins can recognize folded G-quadruplexes, namely, Rif1 [138] and the shelterin component Telomeric Repeat-binding Factor 2 (TRF2) [241]. When G-quadruplex impact on telomere end protection was explored, it was discovered that folded tetraplex assemblies provide weak telomere capping through recruitment of

G-quadruplex-binding telomeric proteins [319]. In addition, DNA and RNA G4s were shown to participate in telomeric chromatin maintenance through recruitment of the human TLS/FUS protein, which then regulates local levels of H4K20me3 [320]. Besides their impact on telomeric chromatin structure, G-quadruplex levels modulate telomere lengthening processes. Part of the driving force behind research into G-quadruplex biology originated from early evidence that folded G-quadruplexes inhibit telomerase activity [13,150]. Importantly, the effect on telomerase is topology-specific; intramolecular antiparallel G-quadruplexes effectively block enzyme activity [13], while intermolecular parallel tetraplexes can be resolved and extended by telomerase [321]. Furthermore, as observed for non-telomeric DNA, unusually high levels of folded G-quadruplexes hamper DNA replication through telomeres. In fact, telomere shortening and fragility was observed in cells depleted from the G4 unwinding helicases BLM, WRN, and RTEL1 [322–324]. To further support the idea of G-quadruplexes as genome instability factors at telomeres, treatment with a collection of small-molecule G4 stabilizers results in similar phenotypes [325,326], which are further exacerbated when performed on cells with reduced G-quadruplex-unwinding capabilities [327–329]. Overall, folded G-quadruplexes have been shown to work as important structural elements in telomeres, possibly providing rudimentary telomere end capping and functioning as protein recruitment hubs to shape telomeric chromatin. Crucially, the three-dimensional structure of guanine tetraplexes determines their effect on telomerase, showing that structural polymorphism has biological significance. Furthermore, telomere replication and lengthening are negatively affected upon loss of regulation of G-quadruplex levels, inducing genome instability similarly to the roadblock model that was explored for DNA replication and transcription.

6.5. G-Quadruplexes in Human Disease and Therapy

As just explored, G-quadruplex formation and dynamics have been connected to a number of cellular functions. It is therefore possible that abnormally high or low levels of guanine tetraplexes in a cell lead to deregulation of key cellular processes, a scenario that can be pathological if the effects exceed the tolerance of an organism. Indeed, a collection of disease-causing genetic mutations have been shown to increase the number and/or stability of G-quadruplexes at specific loci. A notable example is the GGGGCC tandem repeat expansion in the first intron of C9ORF72, which has been characterized as the most common mutation in the neurodegenerative disorders Amyotrophic Lateral Sclerosis (ALS) and Frontotemporal Dementia (FTD). While healthy individuals possess no more than 23 repeats, patient DNA samples were found to harbor hundreds of them [330,331]. Importantly, the G_4C_2 motif was shown to fold into a G-quadruplex structure from RNA molecules in vitro [332,333], suggesting a causal link between higher levels of this secondary structure in the transcript and pathology. Indeed, it was demonstrated that ALS-associated G4 levels in C9ORF72 mRNA drive the assembly of cytosolic membrane-less organelles, both in vitro and in cells. The over-representation of RNA granules was proposed to sequester key proteins and alter RNA metabolism, ultimately leading to neurodegeneration [267]. This hypothesis was demonstrated for hnRNP H, which was found to colocalize with G-quadruplex structures and accumulate in insoluble aggregates in C9-ALS patient cells and brain samples [263]. Importantly, this recruitment was correlated to an abnormal level of incorrectly spliced transcripts in ALS patient brain samples, supporting the idea that abnormally high levels of RNA granules associated with critical protein factors lead to aberrant splicing events in transcripts of genes previously implicated in ALS pathophysiology [263]. An additional hypothesis indicates that enzymatic degradation of long transcripts bearing expanded CGG repeats may not be sufficient to repress cytotoxicity. Indeed, it was recently demonstrated that even short G-rich RNAs drive liquid-to-solid phase transition through the self-assembly capabilities of G-quadruplex structures. The resulting RNA foci induced cell dysfunction and death, suggesting an alternative molecular pathway behind the pathophysiology of repeat expansion disorders [268]. Overall, expansion of G-quadruplexes was clearly designated as the causal factor in ALS and FTD using cellular models and

patient-derived tissue samples. G4 motif repeat expansion in pathogenic alleles was shown to induce widespread deregulation of RNA metabolism, particularly in the context of splicing, through excessive RNA aggregate formation and sequestration of RNA-binding proteins. While alternative mechanisms linked to disease onset should not be overlooked, the current model reveals the central role of G-quadruplex structures in neurodegeneration, also pointing at G4s as amenable targets in ALS/FTD therapy.

The involvement of G-quadruplexes in human disease is especially relevant when considering these guanine assemblies as threats to genome integrity. In fact, as analyzed previously, defects in key components of the cellular response to G-quadruplex-induced polymerase stalling promote the formation of DNA lesions and increase chromosome fragility at pathological levels. It is no wonder that several conditions related to increased genome instability have been linked to impaired regulation of G-quadruplexes, as is the case of Bloom syndrome. Its symptoms include primordial dwarfism, sunlight sensitivity, impaired fertility, immunodeficiency, and a high incidence of cancers [221,334]. This rare genetic condition is caused by mutations in the *BLM* gene, which encodes a RecQ family helicase that is involved in multiple steps of DNA metabolism including G-quadruplex unwinding [221,223,224]. Importantly, a transcriptomics study of BLM-deficient fibroblasts revealed a positive correlation between upregulated genes and transcripts bearing putative G4 motifs (PQS), suggesting that BLM depletion results in a G-quadruplex-dependent global deregulation of transcription [230]. An alternative model of Bloom syndrome pathology proposes that the unusually high rates of recombination events and sister chromatid exchanges observed in BLM-deficient patients are caused by G4-mediated genomic instability, as has been shown for transcribed regions [235]. It is important to note that Bloom syndrome defects have not been exclusively linked to deregulation of G-quadruplexes. In fact, the BLM helicase was shown to contribute to several other G4-independent DNA-repair and recombination processes [335–337]. Despite this, it is undoubtedly possible that genomic loci prone to form G-quadruplexes are much harder to replicate in the absence of this critical helicase, thus ascribing part of the genome instability observed in patient cells to deregulated guanine tetraplexes.

Most of the research into G-quadruplexes as drug targets, however, has been performed in the context of cancer. It has long been demonstrated that genomic instability is a hallmark of cancer, a desirable trait that dramatically increases genetic variability among the tumor population. This critical element enables the insurgence of mutations that are favorable to cancer cells and confer selective advantage, particularly in the six hallmarks of cancer postulated by Hanahan and Weinberg [338]: replicative immortality, resistance to cell death, self-sustaining proliferative signaling, evasion of growth suppressors, angiogenesis induction, active tissue invasion, and metastasis. At the heart of all this, poorly functional DNA damage response, cell cycle checkpoints, and/or apoptosis mechanisms are often associated with tumorigenesis. Remarkably, global G-quadruplex levels are generally higher in tumor or immortalized cells than in normal ones [85,339]. Although it is still unclear if guanine tetraplexes are themselves causal agents of cancer or simply a byproduct of neoplastic transformation, several lines of evidence point at these structures being amenable therapeutic targets. Firstly, as discussed above, specific conformations of folded G4s have been shown to inhibit telomerase [13], which is activated in cancer cells to maintain telomeres. Secondly, as explored previously, G-quadruplexes are endogenous factors that promote polymerase stalling and genomic instability, particularly in genetic backgrounds where G4 unwinding or bypassing has been compromised. Moreover, several promoters of oncogenes bear G4 motifs and folded G-quadruplex structures that have been linked to transcriptional regulation of these genes, including *c-myc*, *c-kit*, SRC, and KRAS [61,166,171,240]. It is no surprise that significant effort has been devoted to the discovery and design of small molecules capable of recognizing G-quadruplexes, as already discussed in a previous section. Indeed, a subset of these compounds successfully impaired telomerase activity, reduced oncogene expression, induced lethal DNA lesions and/or cell death, and counteracted cancer-specific phenotypes in various cell lines and a

xenograft model [150,159,171,182] (reviewed in ref. [340]). Despite the growth in characterized G-quadruplex binders and the increased cytotoxicity of newer compounds, few have entered phase I clinical trials owing to their limited potency [341]. The candidate drug CX-5461 was recently shown to elicit a response in 14% of study participants carrying germline mutations of homologous recombination (HR) factors (BRCA1, BRCA2, or PALB2). Although limited, the data suggest that chemical G-quadruplex stabilization may have higher efficiency in patients with impaired HR [341,342]. On the whole, three main caveats might explain the poor performance of G-quadruplex binders in drug development. Firstly, from a pharmacological point of view, the decade-old drug design strategy that favors π-π stacking of ligands onto G-quartets fundamentally hinders the ability of the drug to be soluble in aqueous solutions, thus compromising the pharmacokinetics of most G-quadruplex ligands. Moreover, until recently, the limited availability of structural data and of specific assays made the design of molecules targeting a specific G-quadruplex within a genomic locus of interest (i.e., oncogene promoters) extremely difficult. In fact, it is not easy to demonstrate that transcriptional downregulation of a target gene is caused by stabilization of a precise G-quadruplex in its regulatory sequence rather than an indirect effect of global G4 persistence [248]. Luckily, the structural gap has partially been filled to allow the drug optimization process to be tailored to a specific G-quadruplex, provided it possesses unique topological characteristics [182]. On the other hand, non-selective G4 stabilization might still prove useful to globally deregulate transcription and induce DNA damage at such levels that induce cell death. In this case, however, it must be remembered that our understanding of G-quadruplex biology is still limited; therefore, predicting the outcome of guanine tetraplex stabilization is difficult and the process may produce unwanted consequences. Surely, expanding our knowledge on the topic will also help in defining new therapeutic strategies that target G-quadruplexes. A noteworthy example is the recently proposed link between chemical G-quadruplex stabilization and activation of the cGAS/STING pathway as a possible exploit to trigger innate immunity against tumors [341,343]. Overall, the progression of the G-quadruplex biology field is the foundation upon which new DNA- and RNA-targeting therapeutic strategies against cancer and genetic disorders may be discovered and optimized. Hence, once again the need to identify and characterize the factors and pathways that co-operate to modulate G-quadruplex levels in healthy cells is crucial in advancing the whole field.

7. Conclusions and Future Perspectives

Throughout seven decades of research, the relevance of the G-quadruplex in nucleic acid metabolism has been greatly expanded upon. Technological advancements, particularly next generation sequencing (G4-seq [121]), coupled to in-depth structural and dynamics investigations of these structures, both in vitro and in cells, have revealed crucial aspects of these polymorphic non-canonical secondary structures. G4s have been observed or predicted to exist in all currently sequenced genomes—from viruses to humans [110,111,122,190,199,203,205,208,209]. When combining all the experimental evidence, G-quadruplex biology has been implicated in all major processes of nucleic acid metabolism and to several layers of gene expression regulation, acting both as a positive and negative influences in any of those. Broadly speaking, G-quadruplexes influence cell biology in two major ways, one of which entails protein recruitment at specific genomic loci. The specificity of G-quadruplexes as scaffolding hubs for protein localization could, in theory, be guaranteed by the fact that G-quadruplexes are folded and unfolded in a controlled manner in cells. Additionally, multiple G4 binding proteins were shown to have higher binding affinity for a specific conformation of guanine tetraplexes, pointing at the relevance of the extensive structural polymorphism of these assemblies. However, it is currently unclear whether cells can dictate the topology of folding G-quadruplexes, for example, by locally regulating the pH or availability of stabilizing cations. Nevertheless, the spectrum of physiological functions regulated by G-quadruplexes is destined to expand as the G-quadruplex interactome does, again pointing at the central importance of discov-

ering and characterizing new G4 binding factors in this research field. On the other hand, G-quadruplexes were shown to work independently of protein recruitment, in particular as obstacles to polymerase activity. In this sense, G-quadruplexes become endogenous determinants of DNA lesions and genomic instability. In this case, the molecular mechanisms behind the prevention, formation, and resolution of G4-dependent DNA damage are yet to be elucidated completely. In the future, the discovery of factors that recognize G-quadruplexes threatening genome integrity and/or guide DNA-damage-processing factors into a specific repair pathway will be paramount in dissecting the role of guanine tetraplexes in physiology and disease.

While the current comprehensive view of G-quadruplex biology requires key information on molecular mechanisms, therapeutic strategies either targeting or employing guanine tetraplexes have been explored since the 1990s. Their inhibitory effect on either oncogene expression or directly on telomerase activity has propelled anticancer research into designing increasingly more specific and cytotoxic G4 ligands, with limited clinical applicability so far. With time, however, alternative strategies have been proposed, from targeting G4 binding proteins as a drug delivery system [344] to stabilizing guanine tetraplexes for the stimulation of innate immunity mechanisms [341,343]. An intriguing avenue yet to be fully explored is G4 unwinding by way of destabilizing ligands [185], whose applicability in cases of aberrantly higher tetraplex levels—especially in genetic conditions such as ALS, FTD, or Bloom syndrome—is predicted to be beneficial. On the whole, the potential translational applications of G-quadruplex studies are intriguing. The limited success of recent clinical trials of G4 ligands underlines the crucial need to dissect G4 biology at the molecular level in order to identify clinically relevant cellular pathways to be more efficiently targeted in affected patients. In this context, expanding the list of factors that specifically interact with guanine tetraplexes will provide new insight into how these peculiar secondary structures can influence the wide variety of biological pathways that have been presented in this review. An additional avenue for future research involves the precise deciphering of the relevance of G-quadruplex polymorphism in cells, which could be theoretically achieved thanks to the expanding list of G4 ligands capable of recognizing a specific tetraplex topology or of inducing a switch between these conformations [184].

Author Contributions: Conceptualization, M.C.D., R.Q. and F.L.; writing—original draft preparation, M.C.D.; resources, S.S., A.P., M.M.-F. and F.L.; writing—review and editing, M.C.D., L.M., L.G., G.M.B., O.Y., S.S., F.M. and F.L.; visualization, R.Q. and D.R.; supervision, M.M.-F., A.P. and F.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by MUR and EU, grant number PRIN_2022JA8JY5 to F.L., PRIN_2022H3MJXX to A.P., PRIN_2022KJHC7S to M.M.F., PRIN_20228S45E to S.S.; Telethon, grant number GGP20001 to S.S.; Associazione Italiana Ricerca sul Cancro, grant number AIRC_IG-19917 to A.P., IG-21806 to M.M.F.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: The figures were created with https://www.biorender.com/.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Johnson, F.B. Fundamentals of G-Quadruplex Biology. *Annu. Rep. Med. Chem.* **2020**, *54*, 3–44. [CrossRef] [PubMed]
2. Varshney, D.; Spiegel, J.; Zyner, K.; Tannahill, D.; Balasubramanian, S. The Regulation and Functions of DNA and RNA G-Quadruplexes. *Nat. Rev. Mol. Cell Biol.* **2020**, *21*, 459–474. [CrossRef] [PubMed]
3. Spiegel, J.; Adhikari, S.; Balasubramanian, S. The Structure and Function of DNA G-Quadruplexes. *Trends Chem.* **2020**, *2*, 123–136. [CrossRef]
4. Levene, P.A.; Jacobs, W.A. Über Guanylsäure. *Berichte Dtsch. Chem. Ges.* **1909**, *42*, 2469–2473. [CrossRef]

5. Bang, I. Untersuchungen Über Die Guanylsäure. *Biochem. Z.* **1910**, *26*, 293–311.
6. Reina, C.; Cavalieri, V. Epigenetic Modulation of Chromatin States and Gene Expression by G-Quadruplex Structures. *Int. J. Mol. Sci.* **2020**, *21*, 4172. [CrossRef] [PubMed]
7. Gellert, M.; Lipsett, M.N.; Davies, D.R. Helix Formation by Guanylic Acid. *Proc. Natl. Acad. Sci. USA* **1962**, *48*, 2013–2018. [CrossRef] [PubMed]
8. Arnott, S.; Chandrasekaran, R.; Marttila, C.M. Structures for Polyinosinic Acid and Polyguanylic Acid. *Biochem. J.* **1974**, *141*, 537–543. [CrossRef]
9. Zimmerman, S.B.; Cohen, G.H.; Davies, D.R. X-Ray Fiber Diffraction and Model-Building Study of Polyguanylic Acid and Polyinosinic Acid. *J. Mol. Biol.* **1975**, *92*, 181–192. [CrossRef] [PubMed]
10. Sen, D.; Gilbert, W. Formation of Parallel Four-Stranded Complexes by Guanine-Rich Motifs in DNA and Its Implications for Meiosis. *Nature* **1988**, *334*, 364–366. [CrossRef]
11. Sundquist, W.I.; Klug, A. Telomeric DNA Dimerizes by Formation of Guanine Tetrads between Hairpin Loops. *Nature* **1989**, *342*, 825–829. [CrossRef]
12. Kwok, C.K.; Merrick, C.J. G-Quadruplexes: Prediction, Characterization, and Biological Application. *Trends Biotechnol.* **2017**, *35*, 997–1013. [CrossRef] [PubMed]
13. Zahler, A.M.; Williamson, J.R.; Cech, T.R.; Prescott, D.M. Inhibition of Telomerase by G-Quartet DMA Structures. *Nature* **1991**, *350*, 718–720. [CrossRef] [PubMed]
14. Burge, S.; Parkinson, G.N.; Hazel, P.; Todd, A.K.; Neidle, S. Quadruplex DNA: Sequence, Topology and Structure. *Nucleic Acids Res.* **2006**, *34*, 5402–5415. [CrossRef] [PubMed]
15. Webba da Silva, M. Geometric Formalism for DNA Quadruplex Folding. *Chem. Eur. J.* **2007**, *13*, 9738–9745. [CrossRef] [PubMed]
16. Pinnavaia, T.J.; Marshall, C.L.; Mettler, C.M.; Fisk, C.L.; Miles, H.T.; Becker, E.D. Alkali Metal Ion Specificity in the Solution Ordering of a Nucleotide, 5′-Guanosine Monophosphate. *J. Am. Chem. Soc.* **1978**, *100*, 3625–3627. [CrossRef]
17. Bhattacharyya, D.; Mirihana Arachchilage, G.; Basu, S. Metal Cations in G-Quadruplex Folding and Stability. *Front. Chem.* **2016**, *4*, 38. [CrossRef] [PubMed]
18. Parkinson, G.N.; Lee, M.P.H.; Neidle, S. Crystal Structure of Parallel Quadruplexes from Human Telomeric DNA. *Nature* **2002**, *417*, 876–880. [CrossRef] [PubMed]
19. Ma, Y.; Iida, K.; Nagasawa, K. Topologies of G-Quadruplex: Biological Functions and Regulation by Ligands. *Biochem. Biophys. Res. Commun.* **2020**, *531*, 3–17. [CrossRef]
20. Risitano, A.; Fox, K.R. Influence of Loop Size on the Stability of Intramolecular DNA Quadruplexes. *Nucleic Acids Res.* **2004**, *32*, 2598–2606. [CrossRef]
21. Risitano, A.; Fox, K.R. Stability of Intramolecular DNA Quadruplexes: Comparison with DNA Duplexes. *Biochemistry* **2003**, *42*, 6507–6513. [CrossRef]
22. Hazel, P.; Huppert, J.; Balasubramanian, S.; Neidle, S. Loop-Length-Dependent Folding of G-Quadruplexes. *J. Am. Chem. Soc.* **2004**, *126*, 16405–16415. [CrossRef]
23. Cheng, M.; Cheng, Y.; Hao, J.; Jia, G.; Zhou, J.; Mergny, J.-L.; Li, C. Loop Permutation Affects the Topology and Stability of G-Quadruplexes. *Nucleic Acids Res.* **2018**, *46*, 9264–9275. [CrossRef]
24. Guédin, A.; Gros, J.; Alberti, P.; Mergny, J.-L. How Long Is Too Long? Effects of Loop Size on G-Quadruplex Stability. *Nucleic Acids Res.* **2010**, *38*, 7858–7868. [CrossRef]
25. Tippana, R.; Xiao, W.; Myong, S. G-Quadruplex Conformation and Dynamics Are Determined by Loop Length and Sequence. *Nucleic Acids Res.* **2014**, *42*, 8106–8114. [CrossRef]
26. Dvorkin, S.A.; Karsisiotis, A.I.; Webba da Silva, M. Encoding Canonical DNA Quadruplex Structure. *Sci. Adv.* **2018**, *4*, eaat3007. [CrossRef]
27. Webba da Silva, M.; Trajkovski, M.; Sannohe, Y.; Ma'ani Hessari, N.; Sugiyama, H.; Plavec, J. Design of a G-Quadruplex Topology through Glycosidic Bond Angles. *Angew. Chem. Int. Ed.* **2009**, *48*, 9167–9170. [CrossRef]
28. Karsisiotis, A.I.; Hessari, N.M.; Novellino, E.; Spada, G.P.; Randazzo, A.; Webba da Silva, M. Topological Characterization of Nucleic Acid G-Quadruplexes by UV Absorption and Circular Dichroism. *Angew. Chem. Int. Ed.* **2011**, *50*, 10645–10648. [CrossRef] [PubMed]
29. Luu, K.N.; Phan, A.T.; Kuryavyi, V.; Lacroix, L.; Patel, D.J. Structure of the Human Telomere in K+ Solution: An Intramolecular (3 + 1) G-Quadruplex Scaffold. *J. Am. Chem. Soc.* **2006**, *128*, 9963–9970. [CrossRef] [PubMed]
30. Phan, A.T.; Kuryavyi, V.; Luu, K.N.; Patel, D.J. Structure of Two Intramolecular G-Quadruplexes Formed by Natural Human Telomere Sequences in K + Solution †. *Nucleic Acids Res.* **2007**, *35*, 6517–6525. [CrossRef] [PubMed]
31. Marušič, M.; Šket, P.; Bauer, L.; Viglasky, V.; Plavec, J. Solution-State Structure of an Intramolecular G-Quadruplex with Propeller, Diagonal and Edgewise Loops. *Nucleic Acids Res.* **2012**, *40*, 6946–6956. [CrossRef] [PubMed]
32. Lim, K.W.; Ng, V.C.M.; Martín-Pintado, N.; Heddi, B.; Phan, A.T. Structure of the Human Telomere in Na+ Solution: An Antiparallel (2+2) G-Quadruplex Scaffold Reveals Additional Diversity. *Nucleic Acids Res.* **2013**, *41*, 10556–10562. [CrossRef] [PubMed]
33. Jana, J.; Mohr, S.; Vianney, Y.M.; Weisz, K. Structural Motifs and Intramolecular Interactions in Non-Canonical G-Quadruplexes. *RSC Chem. Biol.* **2021**, *2*, 338–353. [CrossRef] [PubMed]

34. Chung, W.J.; Heddi, B.; Schmitt, E.; Lim, K.W.; Mechulam, Y.; Phan, A.T. Structure of a Left-Handed DNA G-Quadruplex. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 2729–2733. [CrossRef] [PubMed]
35. Winnerdy, F.R.; Bakalar, B.; Maity, A.; Vandana, J.J.; Mechulam, Y.; Schmitt, E.; Phan, A.T. NMR Solution and X-Ray Crystal Structures of a DNA Molecule Containing Both Right- and Left-Handed Parallel-Stranded G-Quadruplexes. *Nucleic Acids Res.* **2019**, *47*, 8272–8281. [CrossRef] [PubMed]
36. Bakalar, B.; Heddi, B.; Schmitt, E.; Mechulam, Y.; Phan, A.T. A Minimal Sequence for Left-Handed G-Quadruplex Formation. *Angew. Chem. Int. Ed.* **2019**, *58*, 2331–2335. [CrossRef] [PubMed]
37. Das, P.; Winnerdy, F.R.; Maity, A.; Mechulam, Y.; Phan, A.T. A Novel Minimal Motif for Left-Handed G-Quadruplex Formation. *Chem. Commun.* **2021**, *57*, 2527–2530. [CrossRef]
38. Li, K.; Yatsunyk, L.A.; Neidle, S. Machine Learning Shows Torsion Angle Preferences in Left-Handed and Right-Handed Quadruplex DNAs. *Biophys. J.* **2022**, *121*, 4874–4881. [CrossRef]
39. Mukundan, V.T.; Phan, A.T. Bulges in G-Quadruplexes: Broadening the Definition of G-Quadruplex-Forming Sequences. *J. Am. Chem. Soc.* **2013**, *135*, 5017–5028. [CrossRef]
40. Das, P.; Ngo, K.H.; Winnerdy, F.R.; Maity, A.; Bakalar, B.; Mechulam, Y.; Schmitt, E.; Phan, A.T. Bulges in Left-Handed G-Quadruplexes. *Nucleic Acids Res.* **2021**, *49*, 1724–1736. [CrossRef]
41. Heddi, B.; Martin-Pintado, N.; Serimbetov, Z.; Kari, T.M.A.; Phan, A.T. G-Quadruplexes with (4n − 1) Guanines in the G-Tetrad Core: Formation of a G-Triad·water Complex and Implication for Small-Molecule Binding. *Nucleic Acids Res.* **2016**, *44*, 910–916. [CrossRef]
42. Phan, A.T.; Kuryavyi, V.; Gaw, H.Y.; Patel, D.J. Small-Molecule Interaction with a Five-Guanine-Tract G-Quadruplex Structure from the Human MYC Promoter. *Nat. Chem. Biol.* **2005**, *1*, 167–173. [CrossRef] [PubMed]
43. Li, X.; Zheng, K.; Zhang, J.; Liu, H.; He, Y.; Yuan, B.; Hao, Y.; Tan, Z. Guanine-Vacancy–Bearing G-Quadruplexes Responsive to Guanine Derivatives. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 14581–14586. [CrossRef] [PubMed]
44. Wang, K.-B.; Dickerhoff, J.; Wu, G.; Yang, D. PDGFR-β Promoter Forms a Vacancy G-Quadruplex That Can Be Filled in by dGMP: Solution Structure and Molecular Recognition of Guanine Metabolites and Drugs. *J. Am. Chem. Soc.* **2020**, *142*, 5204–5211. [CrossRef] [PubMed]
45. Chen, Y.; Agrawal, P.; Brown, R.V.; Hatzakis, E.; Hurley, L.; Yang, D. The Major G-Quadruplex Formed in the Human Platelet-Derived Growth Factor Receptor β Promoter Adopts a Novel Broken-Strand Structure in K^+ Solution. *J. Am. Chem. Soc.* **2012**, *134*, 13220–13223. [CrossRef] [PubMed]
46. Phan, A.T.; Kuryavyi, V.; Burge, S.; Neidle, S.; Patel, D.J. Structure of an Unprecedented G-Quadruplex Scaffold in the Human c-Kit Promoter. *J. Am. Chem. Soc.* **2007**, *129*, 4386–4392. [CrossRef] [PubMed]
47. Marušič, M.; Veedu, R.N.; Wengel, J.; Plavec, J. G-Rich VEGF Aptamer with Locked and Unlocked Nucleic Acid Modifications Exhibits a Unique G-Quadruplex Fold. *Nucleic Acids Res.* **2013**, *41*, 9524–9536. [CrossRef] [PubMed]
48. Phan, A.T.; Kuryavyi, V.; Darnell, J.C.; Serganov, A.; Majumdar, A.; Ilin, S.; Raslin, T.; Polonskaia, A.; Chen, C.; Clain, D.; et al. Structure-Function Studies of FMRP RGG Peptide Recognition of an RNA Duplex-Quadruplex Junction. *Nat. Struct. Mol. Biol.* **2011**, *18*, 796–804. [CrossRef]
49. Greco, M.L.; Kotar, A.; Rigo, R.; Cristofari, C.; Plavec, J.; Sissi, C. Coexistence of Two Main Folded G-Quadruplexes within a Single G-Rich Domain in the EGFR Promoter. *Nucleic Acids Res.* **2017**, *45*, 10132–10142. [CrossRef]
50. Marquevielle, J.; Robert, C.; Lagrabette, O.; Wahid, M.; Bourdoncle, A.; Xodo, L.E.; Mergny, J.-L.; Salgado, G.F. Structure of Two G-Quadruplexes in Equilibrium in the KRAS Promoter. *Nucleic Acids Res.* **2020**, *48*, 9336–9345. [CrossRef]
51. Dai, J.; Carver, M.; Punchihewa, C.; Jones, R.A.; Yang, D. Structure of the Hybrid-2 Type Intramolecular Human Telomeric G-Quadruplex in K+ Solution: Insights into Structure Polymorphism of the Human Telomeric Sequence. *Nucleic Acids Res.* **2007**, *35*, 4927–4940. [CrossRef]
52. Lago, S.; Tosoni, E.; Nadai, M.; Palumbo, M.; Richter, S.N. The Cellular Protein Nucleolin Preferentially Binds Long-Looped G-Quadruplex Nucleic Acids. *Biochim. Biophys. Acta BBA—Gen. Subj.* **2017**, *1861*, 1371–1381. [CrossRef]
53. Tippana, R.; Hwang, H.; Opresko, P.L.; Bohr, V.A.; Myong, S. Single-Molecule Imaging Reveals a Common Mechanism Shared by G-Quadruplex–Resolving Helicases. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 8448–8453. [CrossRef]
54. Zhang, L.; Er, J.C.; Ghosh, K.K.; Chung, W.J.; Yoo, J.; Xu, W.; Zhao, W.; Phan, A.T.; Chang, Y.-T. Discovery of a Structural-Element Specific G-Quadruplex "Light-Up" Probe. *Sci. Rep.* **2014**, *4*, 3776. [CrossRef]
55. Asamitsu, S.; Bando, T.; Sugiyama, H. Ligand Design to Acquire Specificity to Intended G-Quadruplex Structures. *Chem.—Eur. J.* **2019**, *25*, 417–430. [CrossRef]
56. Biver, T. Discriminating between Parallel, Anti-Parallel and Hybrid G-Quadruplexes: Mechanistic Details on Their Binding to Small Molecules. *Molecules* **2022**, *27*, 4165. [CrossRef] [PubMed]
57. Dai, J.; Punchihewa, C.; Ambrus, A.; Chen, D.; Jones, R.A.; Yang, D. Structure of the Intramolecular Human Telomeric G-Quadruplex in Potassium Solution: A Novel Adenine Triple Formation. *Nucleic Acids Res.* **2007**, *35*, 2440–2450. [CrossRef] [PubMed]
58. Simonsson, T.; Pecinka, P.; Kubista, M. DNA Tetraplex Formation in the Control Region of C-Myc. *Nucleic Acids Res.* **1998**, *26*, 1167–1172. [CrossRef] [PubMed]
59. Ambrus, A.; Chen, D.; Dai, J.; Jones, R.A.; Yang, D. Solution Structure of the Biologically Relevant G-Quadruplex Element in the Human c-MYC Promoter. Implications for G-Quadruplex Stabilization. *Biochemistry* **2005**, *44*, 2048–2058. [CrossRef] [PubMed]

60. Yang, D.; Hurley, L.H. Structure of the Biologically Relevant G-Quadruplex in The c-MYC Promoter. *Nucleosides Nucleotides Nucleic Acids* **2006**, *25*, 951–968. [CrossRef] [PubMed]
61. Rankin, S.; Reszka, A.P.; Huppert, J.; Zloh, M.; Parkinson, G.N.; Todd, A.K.; Ladame, S.; Balasubramanian, S.; Neidle, S. Putative DNA Quadruplex Formation within the Human C-Kit Oncogene. *J. Am. Chem. Soc.* **2005**, *127*, 10584–10589. [CrossRef]
62. Adrian, M.; Ang, D.J.; Lech, C.J.; Heddi, B.; Nicolas, A.; Phan, A.T. Structure and Conformational Dynamics of a Stacked Dimeric G-Quadruplex Formed by the Human CEB1 Minisatellite. *J. Am. Chem. Soc.* **2014**, *136*, 6297–6305. [CrossRef]
63. Amrane, S.; Adrian, M.; Heddi, B.; Serero, A.; Nicolas, A.; Mergny, J.-L.; Phan, A.T. Formation of Pearl-Necklace Monomorphic G-Quadruplexes in the Human CEB25 Minisatellite. *J. Am. Chem. Soc.* **2012**, *134*, 5807–5816. [CrossRef] [PubMed]
64. Shirude, P.S.; Okumus, B.; Ying, L.; Ha, T.; Balasubramanian, S. Single-Molecule Conformational Analysis of G-Quadruplex Formation in the Promoter DNA Duplex of the Proto-Oncogene C-Kit. *J. Am. Chem. Soc.* **2007**, *129*, 7484–7485. [CrossRef]
65. Kumar, N.; Sahoo, B.; Varun, K.A.S.; Maiti, S.; Maiti, S. Effect of Loop Length Variation on Quadruplex-Watson Crick Duplex Competition. *Nucleic Acids Res.* **2008**, *36*, 4433–4442. [CrossRef]
66. Ying, L.; Green, J.J.; Li, H.; Klenerman, D.; Balasubramanian, S. Studies on the Structure and Dynamics of the Human Telomeric G Quadruplex by Single-Molecule Fluorescence Resonance Energy Transfer. *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 14629–14634. [CrossRef] [PubMed]
67. Deng, H.; Braunlin, W.H. Duplex to Quadruplex Equilibrium of the Self-Complementary Oligonucleotide d(GGGGCCCC). *Biopolymers* **1995**, *35*, 677–681. [CrossRef] [PubMed]
68. Kumar, N.; Maiti, S. Quadruplex to Watson–Crick Duplex Transition of the Thrombin Binding Aptamer: A Fluorescence Resonance Energy Transfer Study. *Biochem. Biophys. Res. Commun.* **2004**, *319*, 759–767. [CrossRef] [PubMed]
69. Phan, A.T.; Mergny, J. Human Telomeric DNA: G-quadruplex, i-motif and Watson–Crick Double Helix. *Nucleic Acids Res.* **2002**, *30*, 4618–4625. [CrossRef] [PubMed]
70. Li, W.; Wu, P.; Ohmichi, T.; Sugimoto, N. Characterization and Thermodynamic Properties of Quadruplex/Duplex Competition. *FEBS Lett.* **2002**, *526*, 77–81. [CrossRef]
71. Li, W.; Miyoshi, D.; Nakano, S.; Sugimoto, N. Structural Competition Involving G-Quadruplex DNA and Its Complement. *Biochemistry* **2003**, *42*, 11736–11744. [CrossRef] [PubMed]
72. Kumar, N.; Maiti, S. The Effect of Osmolytes and Small Molecule on Quadruplex–WC Duplex Equilibrium: A Fluorescence Resonance Energy Transfer Study. *Nucleic Acids Res.* **2005**, *33*, 6723–6732. [CrossRef]
73. Sekibo, D.A.T.; Fox, K.R. The Effects of DNA Supercoiling on G-Quadruplex Formation. *Nucleic Acids Res.* **2017**, *45*, 12069–12079. [CrossRef]
74. Lee, W.T.C.; Yin, Y.; Morten, M.J.; Tonzi, P.; Gwo, P.P.; Odermatt, D.C.; Modesti, M.; Cantor, S.B.; Gari, K.; Huang, T.T.; et al. Single-Molecule Imaging Reveals Replication Fork Coupled Formation of G-Quadruplex Structures Hinders Local Replication Stress Signaling. *Nat. Commun.* **2021**, *12*, 2525. [CrossRef]
75. Duquette, M.L.; Handa, P.; Vincent, J.A.; Taylor, A.F.; Maizels, N. Intracellular Transcription of G-Rich DNAs Induces Formation of G-Loops, Novel Structures Containing G4 DNA. *Genes Dev.* **2004**, *18*, 1618–1629. [CrossRef] [PubMed]
76. Abou Assi, H.; Garavís, M.; González, C.; Damha, M.J. I-Motif DNA: Structural Features and Significance to Cell Biology. *Nucleic Acids Res.* **2018**, *46*, 8038–8056. [CrossRef]
77. Snoussi, K.; Nonin-Lecomte, S.; Leroy, J.-L. The RNA I-Motif1. *J. Mol. Biol.* **2001**, *309*, 139–153. [CrossRef] [PubMed]
78. Brown, S.L.; Kendrick, S. The I-Motif as a Molecular Target: More Than a Complementary DNA Secondary Structure. *Pharmaceuticals* **2021**, *14*, 96. [CrossRef]
79. King, J.J.; Irving, K.L.; Evans, C.W.; Chikhale, R.V.; Becker, R.; Morris, C.J.; Peña Martinez, C.D.; Schofield, P.; Christ, D.; Hurley, L.H.; et al. DNA G-Quadruplex and i-Motif Structure Formation Is Interdependent in Human Cells. *J. Am. Chem. Soc.* **2020**, *142*, 20600–20604. [CrossRef]
80. Luo, X.; Zhang, J.; Gao, Y.; Pan, W.; Yang, Y.; Li, X.; Chen, L.; Wang, C.; Wang, Y. Emerging Roles of I-Motif in Gene Expression and Disease Treatment. *Front. Pharmacol.* **2023**, *14*, 1136251. [CrossRef]
81. Shrestha, P.; Jonchhe, S.; Emura, T.; Hidaka, K.; Endo, M.; Sugiyama, H.; Mao, H. Confined Space Facilitates G-Quadruplex Formation. *Nat. Nanotechnol.* **2017**, *12*, 582–588. [CrossRef] [PubMed]
82. Miyoshi, D.; Karimata, H.; Sugimoto, N. Hydration Regulates Thermodynamics of G-Quadruplex Formation under Molecular Crowding Conditions. *J. Am. Chem. Soc.* **2006**, *128*, 7957–7963. [CrossRef] [PubMed]
83. Dhakal, S.; Cui, Y.; Koirala, D.; Ghimire, C.; Kushwaha, S.; Yu, Z.; Yangyuoru, P.M.; Mao, H. Structural and Mechanical Properties of Individual Human Telomeric G-Quadruplexes in Molecularly Crowded Solutions. *Nucleic Acids Res.* **2013**, *41*, 3915–3923. [CrossRef]
84. Aznauryan, M.; Birkedal, V. Dynamics of G-Quadruplex Formation under Molecular Crowding. *J. Phys. Chem. Lett.* **2023**, *14*, 10354–10360. [CrossRef]
85. Hänsel-Hertsch, R.; Beraldi, D.; Lensing, S.V.; Marsico, G.; Zyner, K.; Parry, A.; Di Antonio, M.; Pike, J.; Kimura, H.; Narita, M.; et al. G-Quadruplex Structures Mark Human Regulatory Chromatin. *Nat. Genet.* **2016**, *48*, 1267–1272. [CrossRef]
86. Pavlova, I.; Barinov, N.; Novikov, R.; Severov, V.; Iudin, M.; Vedekhina, T.; Larin, A.; Babenko, V.; Aralov, A.; Gnuchikh, E.; et al. Modeling G4s in Chromatin Context Confirms Partial Nucleosome Exclusion and Reveals Nucleosome-Disrupting Effects of the Least Selective G4 Ligands. *Biochimie* **2023**, *204*, 8–21. [CrossRef] [PubMed]

87. Saccà, B.; Lacroix, L.; Mergny, J.-L. The Effect of Chemical Modifications on the Thermal Stability of Different G-Quadruplex-Forming Oligonucleotides. *Nucleic Acids Res.* **2005**, *33*, 1182–1192. [CrossRef]
88. Cheong, C.; Moore, P.B. Solution Structure of an Unusually Stable RNA Tetraplex Containing G- and U-Quartet Structures. Available online: https://pubs.acs.org/doi/pdf/10.1021/bi00151a003 (accessed on 4 November 2023).
89. Joachimi, A.; Benz, A.; Hartig, J.S. A Comparison of DNA and RNA Quadruplex Structures and Stabilities. *Bioorg. Med. Chem.* **2009**, *17*, 6811–6815. [CrossRef]
90. Adrian, M.; Heddi, B.; Phan, A.T. NMR Spectroscopy of G-Quadruplexes. *Methods* **2012**, *57*, 11–24. [CrossRef]
91. Campbell, N.H.; Parkinson, G.N. Crystallographic Studies of Quadruplex Nucleic Acids. *Methods* **2007**, *43*, 252–263. [CrossRef]
92. Parkinson, G.N.; Collie, G.W. X-Ray Crystallographic Studies of G-Quadruplex Structures. In *G-Quadruplex Nucleic Acids: Methods and Protocols*; Yang, D., Lin, C., Eds.; Methods in Molecular Biology; Springer: New York, NY, USA, 2019; pp. 131–155, ISBN 978-1-4939-9666-7. [CrossRef]
93. Lin, C.; Dickerhoff, J.; Yang, D. NMR Studies of G-Quadruplex Structures and G-Quadruplex-Interactive Compounds. In *G-Quadruplex Nucleic Acids: Methods and Protocols*; Yang, D., Lin, C., Eds.; Methods in Molecular Biology; Springer: New York, NY, USA, 2019; pp. 157–176, ISBN 978-1-4939-9666-7. [CrossRef]
94. Phan, A.T.; Modi, Y.S.; Patel, D.J. Two-Repeat Tetrahymena Telomeric d(TGGGGTTGGGGT) Sequence Interconverts Between Asymmetric Dimeric G-Quadruplexes in Solution. *J. Mol. Biol.* **2004**, *338*, 93–102. [CrossRef]
95. Dailey, M.M.; Miller, M.C.; Bates, P.J.; Lane, A.N.; Trent, J.O. Resolution and Characterization of the Structural Polymorphism of a Single Quadruplex-Forming Sequence. *Nucleic Acids Res.* **2010**, *38*, 4877–4888. [CrossRef]
96. Bao, H.-L.; Ishizuka, T.; Sakamoto, T.; Fujimoto, K.; Uechi, T.; Kenmochi, N.; Xu, Y. Characterization of Human Telomere RNA G-Quadruplex Structures in Vitro and in Living Cells Using 19F NMR Spectroscopy. *Nucleic Acids Res.* **2017**, *45*, 5501–5511. [CrossRef]
97. Guo, J.U.; Bartel, D.P. RNA G-Quadruplexes Are Globally Unfolded in Eukaryotic Cells and Depleted in Bacteria. *Science* **2016**, *353*, aaf5371. [CrossRef]
98. Monsen, R.C.; Chua, E.Y.D.; Hopkins, J.B.; Chaires, J.B.; Trent, J.O. Structure of a 28.5 kDa Duplex-Embedded G-Quadruplex System Resolved to 7.4 Å Resolution with Cryo-EM. *Nucleic Acids Res.* **2023**, *51*, 1943–1959. [CrossRef]
99. Vorlíčková, M.; Kejnovská, I.; Sagi, J.; Renčiuk, D.; Bednářová, K.; Motlová, J.; Kypr, J. Circular Dichroism and Guanine Quadruplexes. *Methods* **2012**, *57*, 64–75. [CrossRef] [PubMed]
100. Luo, Y.; Granzhan, A.; Marquevielle, J.; Cucchiarini, A.; Lacroix, L.; Amrane, S.; Verga, D.; Mergny, J.-L. Guidelines for G-Quadruplexes: I. In Vitro Characterization. *Biochimie* **2023**, *214*, 5–23. [CrossRef] [PubMed]
101. Mergny, J.-L.; Lacroix, L. UV Melting of G-Quadruplexes. *Curr. Protoc. Nucleic Acid Chem.* **2009**, *37*, 17.1.1–17.1.15. [CrossRef] [PubMed]
102. Fonzo, S.D.; Amato, J.; D'Aria, F.; Caterino, M.; D'Amico, F.; Gessini, A.; Brady, J.W.; Cesàro, A.; Pagano, B.; Giancola, C. Ligand Binding to G-Quadruplex DNA: New Insights from Ultraviolet Resonance Raman Spectroscopy. *Phys. Chem. Chem. Phys.* **2020**, *22*, 8128–8140. [CrossRef] [PubMed]
103. Pagano, B.; Randazzo, A.; Fotticchia, I.; Novellino, E.; Petraccone, L.; Giancola, C. Differential Scanning Calorimetry to Investigate G-Quadruplexes Structural Stability. *Methods* **2013**, *64*, 43–51. [CrossRef] [PubMed]
104. Juskowiak, B.; Takenaka, S. Fluorescence Resonance Energy Transfer in the Studies of Guanine Quadruplexes. In *Fluorescent Energy Transfer Nucleic Acid Probes: Designs and Protocols*; Didenko, V.V., Ed.; Methods in Molecular Biology™; Humana Press: Totowa, NJ, USA, 2006; pp. 311–341, ISBN 978-1-59745-069-0. [CrossRef]
105. Simonsson, T.; Sjöback, R. DNA Tetraplex Formation Studied with Fluorescence Resonance Energy Transfer. *J. Biol. Chem.* **1999**, *274*, 17379–17383. [CrossRef] [PubMed]
106. Deniz, A.A.; Dahan, M.; Grunwell, J.R.; Ha, T.; Faulhaber, A.E.; Chemla, D.S.; Weiss, S.; Schultz, P.G. Single-Pair Fluorescence Resonance Energy Transfer on Freely Diffusing Molecules: Observation of Förster Distance Dependence and Subpopulations. *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 3670–3675. [CrossRef]
107. Luo, Y.; Granzhan, A.; Verga, D.; Mergny, J.-L. FRET-MC: A Fluorescence Melting Competition Assay for Studying G4 Structures in Vitro. *Biopolymers* **2021**, *112*, e23415. [CrossRef] [PubMed]
108. Luo, Y.; Verga, D.; Mergny, J.-L. Iso-FRET: An Isothermal Competition Assay to Analyze Quadruplex Formation In Vitro. *Nucleic Acids Res.* **2022**, *50*, e93. [CrossRef]
109. Cheng, Y.; Zhang, Y.; You, H. Characterization of G-Quadruplexes Folding/Unfolding Dynamics and Interactions with Proteins from Single-Molecule Force Spectroscopy. *Biomolecules* **2021**, *11*, 1579. [CrossRef]
110. Huppert, J.L.; Balasubramanian, S. Prevalence of Quadruplexes in the Human Genome. *Nucleic Acids Res.* **2005**, *33*, 2908–2916. [CrossRef] [PubMed]
111. Todd, A.K.; Johnston, M.; Neidle, S. Highly Prevalent Putative Quadruplex Sequence Motifs in Human DNA. *Nucleic Acids Res.* **2005**, *33*, 2901–2907. [CrossRef]
112. Puig Lombardi, E.; Londoño-Vallejo, A. A Guide to Computational Methods for G-Quadruplex Prediction. *Nucleic Acids Res.* **2020**, *48*, 1–15. [CrossRef]
113. Piazza, A.; Adrian, M.; Samazan, F.; Heddi, B.; Hamon, F.; Serero, A.; Lopes, J.; Teulade-Fichou, M.-P.; Phan, A.T.; Nicolas, A. Short Loop Length and High Thermal Stability Determine Genomic Instability Induced by G-Quadruplex-Forming Minisatellites. *EMBO J.* **2015**, *34*, 1718–1734. [CrossRef]

114. Law, M.J.; Lower, K.M.; Voon, H.P.J.; Hughes, J.R.; Garrick, D.; Viprakasit, V.; Mitson, M.; De Gobbi, M.; Marra, M.; Morris, A.; et al. ATR-X Syndrome Protein Targets Tandem Repeats and Influences Allele-Specific Expression in a Size-Dependent Manner. *Cell* **2010**, *143*, 367–378. [CrossRef]
115. Kudlicki, A.S. G-Quadruplexes Involving Both Strands of Genomic DNA Are Highly Abundant and Colocalize with Functional Sites in the Human Genome. *PLoS ONE* **2016**, *11*, e0146174. [CrossRef]
116. Hon, J.; Martínek, T.; Zendulka, J.; Lexa, M. Pqsfinder: An Exhaustive and Imperfection-Tolerant Search Tool for Potential Quadruplex-Forming Sequences in R. *Bioinform. Oxf. Engl.* **2017**, *33*, 3373–3379. [CrossRef]
117. Bedrat, A.; Lacroix, L.; Mergny, J.-L. Re-Evaluation of G-Quadruplex Propensity with G4Hunter. *Nucleic Acids Res.* **2016**, *44*, 1746–1759. [CrossRef]
118. Sahakyan, A.B.; Chambers, V.S.; Marsico, G.; Santner, T.; Di Antonio, M.; Balasubramanian, S. Machine Learning Model for Sequence-Driven DNA G-Quadruplex Formation. *Sci. Rep.* **2017**, *7*, 14535. [CrossRef]
119. Garant, J.-M.; Perreault, J.-P.; Scott, M.S. Motif Independent Identification of Potential RNA G-Quadruplexes by G4RNA Screener. *Bioinform. Oxf. Engl.* **2017**, *33*, 3532–3537. [CrossRef]
120. Patro, L.P.P.; Kumar, A.; Kolimi, N.; Rathinavelan, T. 3D-NuS: A Web Server for Automated Modeling and Visualization of Non-Canonical 3-Dimensional Nucleic Acid Structures. *J. Mol. Biol.* **2017**, *429*, 2438–2448. [CrossRef] [PubMed]
121. Chambers, V.S.; Marsico, G.; Boutell, J.M.; Di Antonio, M.; Smith, G.P.; Balasubramanian, S. High-Throughput Sequencing of DNA G-Quadruplex Structures in the Human Genome. *Nat. Biotechnol.* **2015**, *33*, 877–881. [CrossRef]
122. Marsico, G.; Chambers, V.S.; Sahakyan, A.B.; McCauley, P.; Boutell, J.M.; Antonio, M.D.; Balasubramanian, S. Whole Genome Experimental Maps of DNA G-Quadruplexes in Multiple Species. *Nucleic Acids Res.* **2019**, *47*, 3862–3874. [CrossRef]
123. Kwok, C.K.; Marsico, G.; Sahakyan, A.B.; Chambers, V.S.; Balasubramanian, S. rG4-Seq Reveals Widespread Formation of G-Quadruplex Structures in the Human Transcriptome. *Nat. Methods* **2016**, *13*, 841–844. [CrossRef] [PubMed]
124. De Cian, A.; DeLemos, E.; Mergny, J.-L.; Teulade-Fichou, M.-P.; Monchaud, D. Highly Efficient G-Quadruplex Recognition by Bisquinolinium Compounds. *J. Am. Chem. Soc.* **2007**, *129*, 1856–1857. [CrossRef] [PubMed]
125. Hosseini, M.; Palmer, A.; Manka, W.; Grady, P.G.S.; Patchigolla, V.; Bi, J.; O'Neill, R.J.; Chi, Z.; Aguiar, D. Deep Statistical Modelling of Nanopore Sequencing Translocation Times Reveals Latent Non-B DNA Structures. *Bioinformatics* **2023**, *39*, i242–i251. [CrossRef]
126. Schaffitzel, C.; Berger, I.; Postberg, J.; Hanes, J.; Lipps, H.J.; Plückthun, A. In Vitro Generated Antibodies Specific for Telomeric Guanine-Quadruplex DNA React with *Stylonychia lemnae* Macronuclei. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 8572–8577. [CrossRef]
127. Paeschke, K.; Simonsson, T.; Postberg, J.; Rhodes, D.; Lipps, H.J. Telomere End-Binding Proteins Control the Formation of G-Quadruplex DNA Structures In Vivo. *Nat. Struct. Mol. Biol.* **2005**, *12*, 847–854. [CrossRef]
128. Lauth, M.R.; Spear, B.B.; Heumann, J.; Prescott, D.M. DNA of Ciliated Protozoa: DNA Sequence Diminution during Macronuclear Development of Oxytricha. *Cell* **1976**, *7*, 67–74. [CrossRef]
129. Swanton, M.T.; Heumann, J.M.; Prescott, D.M. Gene-Sized DNA Molecules of the Macronuclei in Three Species of Hypotrichs: Size Distributions and Absence of Nicks. *Chromosoma* **1980**, *77*, 217–227. [CrossRef]
130. Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative Visualization of DNA G-Quadruplex Structures in Human Cells. *Nat. Chem.* **2013**, *5*, 182–186. [CrossRef] [PubMed]
131. Biffi, G.; Di Antonio, M.; Tannahill, D.; Balasubramanian, S. Visualization and Selective Chemical Targeting of RNA G-Quadruplex Structures in the Cytoplasm of Human Cells. *Nat. Chem.* **2014**, *6*, 75–80. [CrossRef] [PubMed]
132. Javadekar, S.M.; Nilavar, N.M.; Paranjape, A.; Das, K.; Raghavan, S.C. Characterization of G-Quadruplex Antibody Reveals Differential Specificity for G4 DNA Forms. *DNA Res.* **2020**, *27*, dsaa024. [CrossRef] [PubMed]
133. Liu, H.-Y.; Zhao, Q.; Zhang, T.-P.; Wu, Y.; Xiong, Y.-X.; Wang, S.-K.; Ge, Y.-L.; He, J.-H.; Lv, P.; Ou, T.-M.; et al. Conformation Selective Antibody Enables Genome Profiling and Leads to Discovery of Parallel G-Quadruplex in Human Telomeres. *Cell Chem. Biol.* **2016**, *23*, 1261–1270. [CrossRef] [PubMed]
134. Henderson, A.; Wu, Y.; Huang, Y.C.; Chavez, E.A.; Platt, J.; Johnson, F.B.; Brosh, R.M.; Sen, D.; Lansdorp, P.M. Detection of G-Quadruplex DNA in Mammalian Cells. *Nucleic Acids Res.* **2014**, *42*, 860–869. [CrossRef] [PubMed]
135. Kazemier, H.G.; Paeschke, K.; Lansdorp, P.M. Guanine Quadruplex Monoclonal Antibody 1H6 Cross-Reacts with Restrained Thymidine-Rich Single Stranded DNA. *Nucleic Acids Res.* **2017**, *45*, 5913–5919. [CrossRef] [PubMed]
136. Hoffmann, R.F.; Moshkin, Y.M.; Mouton, S.; Grzeschik, N.A.; Kalicharan, R.D.; Kuipers, J.; Wolters, A.H.G.; Nishida, K.; Romashchenko, A.V.; Postberg, J.; et al. Guanine Quadruplex Structures Localize to Heterochromatin. *Nucleic Acids Res.* **2016**, *44*, 152–163. [CrossRef] [PubMed]
137. Gray, L.T.; Vallur, A.C.; Eddy, J.; Maizels, N. G Quadruplexes Are Genomewide Targets of Transcriptional Helicases XPB and XPD. *Nat. Chem. Biol.* **2014**, *10*, 313–318. [CrossRef] [PubMed]
138. Kanoh, Y.; Matsumoto, S.; Fukatsu, R.; Kakusho, N.; Kono, N.; Renard-Guillet, C.; Masuda, K.; Iida, K.; Nagasawa, K.; Shirahige, K.; et al. Rif1 Binds to G Quadruplexes and Suppresses Replication over Long Distances. *Nat. Struct. Mol. Biol.* **2015**, *22*, 889–897. [CrossRef] [PubMed]
139. Paeschke, K.; Capra, J.A.; Zakian, V.A. DNA Replication through G-Quadruplex Motifs Is Promoted by the Saccharomyces Cerevisiae Pif1 DNA Helicase. *Cell* **2011**, *145*, 678–691. [CrossRef] [PubMed]

140. Hänsel-Hertsch, R.; Spiegel, J.; Marsico, G.; Tannahill, D.; Balasubramanian, S. Genome-Wide Mapping of Endogenous G-Quadruplex DNA Structures by Chromatin Immunoprecipitation and High-Throughput Sequencing. *Nat. Protoc.* **2018**, *13*, 551–564. [CrossRef] [PubMed]
141. Zyner, K.G.; Simeone, A.; Flynn, S.M.; Doyle, C.; Marsico, G.; Adhikari, S.; Portella, G.; Tannahill, D.; Balasubramanian, S. G-Quadruplex DNA Structures in Human Stem Cells and Differentiation. *Nat. Commun.* **2022**, *13*, 142. [CrossRef] [PubMed]
142. Li, C.; Wang, H.; Yin, Z.; Fang, P.; Xiao, R.; Xiang, Y.; Wang, W.; Li, Q.; Huang, B.; Huang, J.; et al. Ligand-Induced Native G-Quadruplex Stabilization Impairs Transcription Initiation. *Genome Res.* **2021**, *31*, 1546–1560. [CrossRef]
143. Lyu, J.; Shao, R.; Kwong Yung, P.Y.; Elsässer, S.J. Genome-Wide Mapping of G-Quadruplex Structures with CUT&Tag. *Nucleic Acids Res.* **2022**, *50*, e13. [CrossRef]
144. Hui, W.W.I.; Simeone, A.; Zyner, K.G.; Tannahill, D.; Balasubramanian, S. Single-Cell Mapping of DNA G-Quadruplex Structures in Human Cancer Cells. *Sci. Rep.* **2021**, *11*, 23641. [CrossRef]
145. Struhl, K. Interpreting Chromatin Immunoprecipitation Experiments. In *Evaluating Techniques in Biochemical Research*; Cell Press: Cambridge, MA, USA, 2007; pp. 29–33.
146. Biocca, S.; Ruberti, F.; Tafani, M.; Pierandrei-Amaldi, P.; Cattaneo, A. Redox State of Single Chain Fv Fragments Targeted to the Endoplasmic Reticulum, Cytosol and Mitochondria. *Nat. Biotechnol.* **1995**, *13*, 1110–1115. [CrossRef]
147. Wörn, A.; Plückthun, A. Stability Engineering of Antibody Single-Chain Fv Fragments. *J. Mol. Biol.* **2001**, *305*, 989–1010. [CrossRef] [PubMed]
148. Böldicke, T. Single Domain Antibodies for the Knockdown of Cytosolic and Nuclear Proteins. *Protein Sci.* **2017**, *26*, 925–945. [CrossRef] [PubMed]
149. Galli, S.; Melidis, L.; Flynn, S.M.; Varshney, D.; Simeone, A.; Spiegel, J.; Madden, S.K.; Tannahill, D.; Balasubramanian, S. DNA G-Quadruplex Recognition In Vitro and in Live Cells by a Structure-Specific Nanobody. *J. Am. Chem. Soc.* **2022**, *144*, 23096–23103. [CrossRef] [PubMed]
150. Sun, D.; Thompson, B.; Cathers, B.E.; Salazar, M.; Kerwin, S.M.; Trent, J.O.; Jenkins, T.C.; Neidle, S.; Hurley, L.H. Inhibition of Human Telomerase by a G-Quadruplex-Interactive Compound. *J. Med. Chem.* **1997**, *40*, 2113–2116. [CrossRef] [PubMed]
151. Li, Q.; Xiang, J.-F.; Yang, Q.-F.; Sun, H.-X.; Guan, A.-J.; Tang, Y.-L. G4LDB: A Database for Discovering and Studying G-Quadruplex Ligands. *Nucleic Acids Res.* **2013**, *41*, D1115–D1123. [CrossRef] [PubMed]
152. Wang, Y.-H.; Yang, Q.-F.; Lin, X.; Chen, D.; Wang, Z.-Y.; Chen, B.; Han, H.-Y.; Chen, H.-D.; Cai, K.-C.; Li, Q.; et al. G4LDB 2.2: A Database for Discovering and Studying G-Quadruplex and i-Motif Ligands. *Nucleic Acids Res.* **2022**, *50*, D150–D160. [CrossRef] [PubMed]
153. Duarte, A.R.; Cadoni, E.; Ressurreição, A.S.; Moreira, R.; Paulo, A. Design of Modular G-Quadruplex Ligands. *ChemMedChem* **2018**, *13*, 869–893. [CrossRef] [PubMed]
154. Gavathiotis, E.; Heald, R.A.; Stevens, M.F.G.; Searle, M.S. Recognition and Stabilization of Quadruplex DNA by a Potent New Telomerase Inhibitor: NMR Studies of the 2:1 Complex of a Pentacyclic Methylacridinium Cation with d(TTAGGGT)4. *Angew. Chem. Int. Ed.* **2001**, *40*, 4749–4751. [CrossRef]
155. Paulo, A.; Castillo, C.C.; Neidle, S. 5.09—Targeting Promoter Quadruplex Nucleic Acids for Cancer Therapy. In *Comprehensive Medicinal Chemistry III*; Chackalamannil, S., Rotella, D., Ward, S.E., Eds.; Elsevier: Oxford, UK, 2017; pp. 308–340, ISBN 978-0-12-803201-5.
156. Cao, Q.; Li, Y.; Freisinger, E.; Qin, P.Z.; Sigel, R.K.O.; Mao, Z.-W. G-Quadruplex DNA Targeted Metal Complexes Acting as Potential Anticancer Drugs. *Inorg. Chem. Front.* **2017**, *4*, 10–32. [CrossRef]
157. Read, M.; Harrison, R.J.; Romagnoli, B.; Tanious, F.A.; Gowan, S.H.; Reszka, A.P.; Wilson, W.D.; Kelland, L.R.; Neidle, S. Structure-Based Design of Selective and Potent G Quadruplex-Mediated Telomerase Inhibitors. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 4844–4849. [CrossRef]
158. Moore, M.J.B.; Schultes, C.M.; Cuesta, J.; Cuenca, F.; Gunaratnam, M.; Tanious, F.A.; Wilson, W.D.; Neidle, S. Trisubstituted Acridines as G-Quadruplex Telomere Targeting Agents. Effects of Extensions of the 3,6- and 9-Side Chains on Quadruplex Binding, Telomerase Activity, and Cell Proliferation. *J. Med. Chem.* **2006**, *49*, 582–599. [CrossRef]
159. Burger, A.M.; Dai, F.; Schultes, C.M.; Reszka, A.P.; Moore, M.J.; Double, J.A.; Neidle, S. The G-Quadruplex-Interactive Molecule BRACO-19 Inhibits Tumor Growth, Consistent with Telomere Targeting and Interference with Telomerase Function. *Cancer Res.* **2005**, *65*, 1489–1496. [CrossRef] [PubMed]
160. Perrone, R.; Butovskaya, E.; Daelemans, D.; Palù, G.; Pannecouque, C.; Richter, S.N. Anti-HIV-1 Activity of the G-Quadruplex Ligand BRACO-19. *J. Antimicrob. Chemother.* **2014**, *69*, 3248–3258. [CrossRef] [PubMed]
161. Izbicka, E.; Wheelhouse, R.T.; Raymond, E.; Davidson, K.K.; Lawrence, R.A.; Sun, D.; Windle, B.E.; Hurley, L.H.; Von Hoff, D.D. Effects of Cationic Porphyrins as G-Quadruplex Interactive Agents in Human Tumor Cells1. *Cancer Res.* **1999**, *59*, 639–644. [PubMed]
162. Dexheimer, T.S.; Sun, D.; Hurley, L.H. Deconvoluting the Structural and Drug-Recognition Complexity of the G-Quadruplex-Forming Region Upstream of the Bcl-2 P1 Promoter. *J. Am. Chem. Soc.* **2006**, *128*, 5404–5415. [CrossRef] [PubMed]
163. Ren, J.; Chaires, J.B. Sequence and Structural Selectivity of Nucleic Acid Binding Ligands. *Biochemistry* **1999**, *38*, 16067–16075. [CrossRef] [PubMed]

164. Monchaud, D.; Allain, C.; Teulade-Fichou, M.-P. Development of a Fluorescent Intercalator Displacement Assay (G4-FID) for Establishing Quadruplex-DNA Affinity and Selectivity of Putative Ligands. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 4842–4845. [CrossRef] [PubMed]
165. Grand, C.L.; Han, H.; Muñoz, R.M.; Weitman, S.; Von Hoff, D.D.; Hurley, L.H.; Bearss, D.J. The Cationic Porphyrin TMPyP4 Down-Regulates c-MYC and Human Telomerase Reverse Transcriptase Expression and Inhibits Tumor Growth In Vivo. *Mol. Cancer Ther.* **2002**, *1*, 565–573.
166. Siddiqui-Jain, A.; Grand, C.L.; Bearss, D.J.; Hurley, L.H. Direct Evidence for a G-Quadruplex in a Promoter Region and Its Targeting with a Small Molecule to Repress c-MYC Transcription. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11593–11598. [CrossRef]
167. Tran, P.L.T.; Largy, E.; Hamon, F.; Teulade-Fichou, M.-P.; Mergny, J.-L. Fluorescence Intercalator Displacement Assay for Screening G4 Ligands towards a Variety of G-Quadruplex Structures. *Biochimie* **2011**, *93*, 1288–1296. [CrossRef] [PubMed]
168. De Rache, A.; Mergny, J.-L. Assessment of Selectivity of G-Quadruplex Ligands via an Optimised FRET Melting Assay. *Biochimie* **2015**, *115*, 194–202. [CrossRef] [PubMed]
169. Halder, R.; Riou, J.-F.; Teulade-Fichou, M.-P.; Frickey, T.; Hartig, J.S. Bisquinolinium Compounds Induce Quadruplex-Specific Transcriptome Changes in HeLa S3 Cell Lines. *BMC Res. Notes* **2012**, *5*, 138. [CrossRef] [PubMed]
170. Rodriguez, R.; Müller, S.; Yeoman, J.A.; Trentesaux, C.; Riou, J.-F.; Balasubramanian, S. A Novel Small Molecule That Alters Shelterin Integrity and Triggers a DNA-Damage Response at Telomeres. *J. Am. Chem. Soc.* **2008**, *130*, 15758–15759. [CrossRef] [PubMed]
171. Rodriguez, R.; Miller, K.M.; Forment, J.V.; Bradshaw, C.R.; Nikan, M.; Britton, S.; Oelschlaegel, T.; Xhemalce, B.; Balasubramanian, S.; Jackson, S.P. Small-Molecule–Induced DNA Damage Identifies Alternative DNA Structures in Human Genes. *Nat. Chem. Biol.* **2012**, *8*, 301–310. [CrossRef] [PubMed]
172. Müller, S.; Sanders, D.A.; Di Antonio, M.; Matsis, S.; Riou, J.-F.; Rodriguez, R.; Balasubramanian, S. Pyridostatin Analogues Promote Telomere Dysfunction and Long-Term Growth Inhibition in Human Cancer Cells. *Org. Biomol. Chem.* **2012**, *10*, 6537–6546. [CrossRef]
173. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B.A.; Thiessen, P.A.; Yu, B.; et al. PubChem 2023 Update. *Nucleic Acids Res.* **2023**, *51*, D1373–D1380. [CrossRef] [PubMed]
174. Di Antonio, M.; Ponjavic, A.; Radzevičius, A.; Ranasinghe, R.T.; Catalano, M.; Zhang, X.; Shen, J.; Needham, L.-M.; Lee, S.F.; Klenerman, D.; et al. Single-Molecule Visualization of DNA G-Quadruplex Formation in Live Cells. *Nat. Chem.* **2020**, *12*, 832–837. [CrossRef]
175. Summers, P.A.; Lewis, B.W.; Gonzalez-Garcia, J.; Porreca, R.M.; Lim, A.H.M.; Cadinu, P.; Martin-Pintado, N.; Mann, D.J.; Edel, J.B.; Vannier, J.B.; et al. Visualising G-Quadruplex DNA Dynamics in Live Cells by Fluorescence Lifetime Imaging Microscopy. *Nat. Commun.* **2021**, *12*, 162. [CrossRef]
176. Lefebvre, J.; Guetta, C.; Poyer, F.; Mahuteau-Betzer, F.; Teulade-Fichou, M.-P. Copper–Alkyne Complexation Responsible for the Nucleolar Localization of Quadruplex Nucleic Acid Drugs Labeled by Click Reactions. *Angew. Chem. Int. Ed.* **2017**, *56*, 11365–11369. [CrossRef]
177. Zhang, X.; Spiegel, J.; Martínez Cuesta, S.; Adhikari, S.; Balasubramanian, S. Chemical Profiling of DNA G-Quadruplex-Interacting Proteins in Live Cells. *Nat. Chem.* **2021**, *13*, 626–633. [CrossRef] [PubMed]
178. Su, H.; Xu, J.; Chen, Y.; Wang, Q.; Lu, Z.; Chen, Y.; Chen, K.; Han, S.; Fang, Z.; Wang, P.; et al. Photoactive G-Quadruplex Ligand Identifies Multiple G-Quadruplex-Related Proteins with Extensive Sequence Tolerance in the Cellular Environment. *J. Am. Chem. Soc.* **2021**, *143*, 1917–1923. [CrossRef] [PubMed]
179. Lai, H.; Xiao, Y.; Yan, S.; Tian, F.; Zhong, C.; Liu, Y.; Weng, X.; Zhou, X. Symmetric Cyanovinyl-Pyridinium Triphenylamine: A Novel Fluorescent Switch-on Probe for an Antiparallel G-Quadruplex. *Analyst* **2014**, *139*, 1834–1838. [CrossRef] [PubMed]
180. Kataoka, Y.; Fujita, H.; Kasahara, Y.; Yoshihara, T.; Tobita, S.; Kuwahara, M. Minimal Thioflavin T Modifications Improve Visual Discrimination of Guanine-Quadruplex Topologies and Alter Compound-Induced Topological Structures. *Anal. Chem.* **2014**, *86*, 12078–12084. [CrossRef] [PubMed]
181. Răsădean, D.M.; Sheng, B.; Dash, J.; Pantoș, G.D. Amino-Acid-Derived Naphthalenediimides as Versatile G-Quadruplex Binders. *Chem.—Eur. J.* **2017**, *23*, 8491–8499. [CrossRef] [PubMed]
182. Hu, M.-H.; Wang, Y.-Q.; Yu, Z.-Y.; Hu, L.-N.; Ou, T.-M.; Chen, S.-B.; Huang, Z.-S.; Tan, J.-H. Discovery of a New Four-Leaf Clover-Like Ligand as a Potent c-MYC Transcription Inhibitor Specifically Targeting the Promoter G-Quadruplex. *J. Med. Chem.* **2018**, *61*, 2447–2459. [CrossRef] [PubMed]
183. Di Antonio, M.; Biffi, G.; Mariani, A.; Raiber, E.-A.; Rodriguez, R.; Balasubramanian, S. Selective RNA Versus DNA G-Quadruplex Targeting by In Situ Click Chemistry. *Angew. Chem. Int. Ed.* **2012**, *51*, 11073–11078. [CrossRef]
184. O'Hagan, M.P.; Morales, J.C.; Galan, M.C. Binding and Beyond: What Else Can G-Quadruplex Ligands Do? *Eur. J. Org. Chem.* **2019**, *2019*, 4995–5017. [CrossRef]
185. Lejault, P.; Mitteaux, J.; Sperti, F.R.; Monchaud, D. How to Untie G-Quadruplex Knots and Why? *Cell Chem. Biol.* **2021**, *28*, 436–455. [CrossRef]
186. Waller, Z.A.E.; Sewitz, S.A.; Hsu, S.-T.D.; Balasubramanian, S. A Small Molecule That Disrupts G-Quadruplex DNA Structure and Enhances Gene Expression. *J. Am. Chem. Soc.* **2009**, *131*, 12628–12633. [CrossRef]

187. Bejugam, M.; Sewitz, S.; Shirude, P.S.; Rodriguez, R.; Shahid, R.; Balasubramanian, S. Trisubstituted Isoalloxazines as a New Class of G-Quadruplex Binding Ligands: Small Molecule Regulation of c-Kit Oncogene Expression. *J. Am. Chem. Soc.* **2007**, *129*, 12926–12927. [CrossRef]
188. Maizels, N.; Gray, L.T. The G4 Genome. *PLoS Genet.* **2013**, *9*, e1003468. [CrossRef]
189. Eddy, J.; Maizels, N. Gene Function Correlates with Potential for G4 DNA Formation in the Human Genome. *Nucleic Acids Res.* **2006**, *34*, 3887–3896. [CrossRef]
190. Rawal, P.; Kummarasetti, V.B.R.; Ravindran, J.; Kumar, N.; Halder, K.; Sharma, R.; Mukerji, M.; Das, S.K.; Chowdhury, S. Genome-Wide Prediction of G4 DNA as Regulatory Motifs: Role in *Escherichia Coli* Global Regulation. *Genome Res.* **2006**, *16*, 644–655. [CrossRef] [PubMed]
191. Bartas, M.; Čutová, M.; Brázda, V.; Kaura, P.; Šťastný, J.; Kolomazník, J.; Coufal, J.; Goswami, P.; Červeň, J.; Pečinka, P. The Presence and Localization of G-Quadruplex Forming Sequences in the Domain of Bacteria. *Molecules* **2019**, *24*, 1711. [CrossRef]
192. Yadav, P.; Kim, N.; Kumari, M.; Verma, S.; Sharma, T.K.; Yadav, V.; Kumar, A. G-Quadruplex Structures in Bacteria: Biological Relevance and Potential as an Antimicrobial Target. *J. Bacteriol.* **2021**, *203*, e0057720. [CrossRef]
193. Holder, I.T.; Hartig, J.S. A Matter of Location: Influence of G-Quadruplexes on Escherichia Coli Gene Expression. *Chem. Biol.* **2014**, *21*, 1511–1521. [CrossRef] [PubMed]
194. Yadav, V.K.; Abraham, J.K.; Mani, P.; Kulshrestha, R.; Chowdhury, S. QuadBase: Genome-Wide Database of G4 DNA—Occurrence and Conservation in Human, Chimpanzee, Mouse and Rat Promoters and 146 Microbes. *Nucleic Acids Res.* **2008**, *36*, D381–D385. [CrossRef] [PubMed]
195. Perrone, R.; Lavezzo, E.; Riello, E.; Manganelli, R.; Palù, G.; Toppo, S.; Provvedi, R.; Richter, S.N. Mapping and Characterization of G-Quadruplexes in Mycobacterium Tuberculosis Gene Promoter Regions. *Sci. Rep.* **2017**, *7*, 5743. [CrossRef]
196. Mishra, S.K.; Shankar, U.; Jain, N.; Sikri, K.; Tyagi, J.S.; Sharma, T.K.; Mergny, J.-L.; Kumar, A. Characterization of G-Quadruplex Motifs in espB, espK, and Cyp51 Genes of Mycobacterium Tuberculosis as Potential Drug Targets. *Mol. Ther.—Nucleic Acids* **2019**, *16*, 698–706. [CrossRef]
197. Kota, S.; Dhamodharan, V.; Pradeepkumar, P.I.; Misra, H.S. G-Quadruplex Forming Structural Motifs in the Genome of Deinococcus Radiodurans and Their Regulatory Roles in Promoter Functions. *Appl. Microbiol. Biotechnol.* **2015**, *99*, 9761–9769. [CrossRef]
198. Beaume, N.; Pathak, R.; Yadav, V.K.; Kota, S.; Misra, H.S.; Gautam, H.K.; Chowdhury, S. Genome-Wide Study Predicts Promoter-G4 DNA Motifs Regulate Selective Functions in Bacteria: Radioresistance of *D. Radiodurans* Involves G4 DNA-Mediated Regulation. *Nucleic Acids Res.* **2013**, *41*, 76–89. [CrossRef]
199. Lavezzo, E.; Berselli, M.; Frasson, I.; Perrone, R.; Palù, G.; Brazzale, A.R.; Richter, S.N.; Toppo, S. G-Quadruplex Forming Sequences in the Genome of All Known Human Viruses: A Comprehensive Guide. *PLOS Comput. Biol.* **2018**, *14*, e1006675. [CrossRef]
200. Ravichandran, S.; Kim, Y.-E.; Bansal, V.; Ghosh, A.; Hur, J.; Subramani, V.K.; Pradhan, S.; Lee, M.K.; Kim, K.K.; Ahn, J.-H. Genome-Wide Analysis of Regulatory G-Quadruplexes Affecting Gene Expression in Human Cytomegalovirus. *PLOS Pathog.* **2018**, *14*, e1007334. [CrossRef] [PubMed]
201. Ruggiero, E.; Richter, S.N. Targeting G-Quadruplexes to Achieve Antiviral Activity. *Bioorg. Med. Chem. Lett.* **2023**, *79*, 129085. [CrossRef] [PubMed]
202. Pathak, R. G-Quadruplexes in the Viral Genome: Unlocking Targets for Therapeutic Interventions and Antiviral Strategies. *Viruses* **2023**, *15*, 2216. [CrossRef] [PubMed]
203. Capra, J.A.; Paeschke, K.; Singh, M.; Zakian, V.A. G-Quadruplex DNA Sequences Are Evolutionarily Conserved and Associated with Distinct Genomic Features in Saccharomyces Cerevisiae. *PLoS Comput. Biol.* **2010**, *6*, e1000861. [CrossRef] [PubMed]
204. Hershman, S.G.; Chen, Q.; Lee, J.Y.; Kozak, M.L.; Yue, P.; Wang, L.-S.; Johnson, F.B. Genomic Distribution and Functional Analyses of Potential G-Quadruplex-Forming Sequences in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **2008**, *36*, 144. [CrossRef]
205. Garg, R.; Aggarwal, J.; Thakkar, B. Genome-Wide Discovery of G-Quadruplex Forming Sequences and Their Functional Relevance in Plants. *Sci. Rep.* **2016**, *6*, 28211. [CrossRef] [PubMed]
206. Mullen, M.A.; Olson, K.J.; Dallaire, P.; Major, F.; Assmann, S.M.; Bevilacqua, P.C. RNA G-Quadruplexes in the Model Plant Species Arabidopsis Thaliana: Prevalence and Possible Functional Roles. *Nucleic Acids Res.* **2010**, *38*, 8149–8163. [CrossRef]
207. Andorf, C.M.; Kopylov, M.; Dobbs, D.; Koch, K.E.; Stroupe, M.E.; Lawrence, C.J.; Bass, H.W. G-Quadruplex (G4) Motifs in the Maize (*Zea Mays* L.) Genome Are Enriched at Specific Locations in Thousands of Genes Coupled to Energy Status, Hypoxia, Low Sugar, and Nutrient Deprivation. *J. Genet. Genomics* **2014**, *41*, 627–647. [CrossRef] [PubMed]
208. Wu, F.; Niu, K.; Cui, Y.; Li, C.; Lyu, M.; Ren, Y.; Chen, Y.; Deng, H.; Huang, L.; Zheng, S.; et al. Genome-Wide Analysis of DNA G-Quadruplex Motifs across 37 Species Provides Insights into G4 Evolution. *Commun. Biol.* **2021**, *4*, 98. [CrossRef] [PubMed]
209. Zhao, Y.; Du, Z.; Li, N. Extensive Selection for the Enrichment of G4 DNA Motifs in Transcriptional Regulatory Regions of Warm Blooded Animals. *FEBS Lett.* **2007**, *581*, 1951–1956. [CrossRef] [PubMed]
210. Volná, A.; Bartas, M.; Karlický, V.; Nezval, J.; Kundrátová, K.; Pečinka, P.; Špunda, V.; Červeň, J. G-Quadruplex in Gene Encoding Large Subunit of Plant RNA Polymerase II: A Billion-Year-Old Story. *Int. J. Mol. Sci.* **2021**, *22*, 7381. [CrossRef] [PubMed]
211. Kumari, S.; Bugaut, A.; Huppert, J.L.; Balasubramanian, S. An RNA G-Quadruplex in the 5′ UTR of the NRAS Proto-Oncogene Modulates Translation. *Nat. Chem. Biol.* **2007**, *3*, 218–221. [CrossRef]
212. Williams, P.; Li, L.; Dong, X.; Wang, Y. Identification of SLIRP as a G Quadruplex-Binding Protein. *J. Am. Chem. Soc.* **2017**, *139*, 12426–12429. [CrossRef]

213. Makowski, M.M.; Gräwe, C.; Foster, B.M.; Nguyen, N.V.; Bartke, T.; Vermeulen, M. Global Profiling of Protein–DNA and Protein–Nucleosome Binding Affinities Using Quantitative Mass Spectrometry. *Nat. Commun.* **2018**, *9*, 1653. [CrossRef]
214. von Hacht, A.; Seifert, O.; Menger, M.; Schütze, T.; Arora, A.; Konthur, Z.; Neubauer, P.; Wagner, A.; Weise, C.; Kurreck, J. Identification and Characterization of RNA Guanine-Quadruplex Binding Proteins. *Nucleic Acids Res.* **2014**, *42*, 6630–6644. [CrossRef]
215. Raiber, E.-A.; Kranaster, R.; Lam, E.; Nikan, M.; Balasubramanian, S. A Non-Canonical DNA Structure Is a Binding Motif for the Transcription Factor SP1 In Vitro. *Nucleic Acids Res.* **2012**, *40*, 1499–1508. [CrossRef]
216. Brázda, V.; Červeň, J.; Bartas, M.; Mikysková, N.; Coufal, J.; Pečinka, P. The Amino Acid Composition of Quadruplex Binding Proteins Reveals a Shared Motif and Predicts New Potential Quadruplex Interactors. *Mol. J. Synth. Chem. Nat. Prod. Chem.* **2018**, *23*, 2341. [CrossRef]
217. Huang, Z.-L.; Dai, J.; Luo, W.-H.; Wang, X.-G.; Tan, J.-H.; Chen, S.-B.; Huang, Z.-S. Identification of G-Quadruplex-Binding Protein from the Exploration of RGG Motif/G-Quadruplex Interactions. *J. Am. Chem. Soc.* **2018**, *140*, 17945–17955. [CrossRef] [PubMed]
218. Mishra, S.K.; Tawani, A.; Mishra, A.; Kumar, A. G4IPDB: A Database for G-Quadruplex Structure Forming Nucleic Acid Interacting Proteins. *Sci. Rep.* **2016**, *6*, 38144. [CrossRef] [PubMed]
219. Zhang, Z.-H.; Qian, S.H.; Wei, D.; Chen, Z.-X. In Vivo Dynamics and Regulation of DNA G-Quadruplex Structures in Mammals. *Cell Biosci.* **2023**, *13*, 117. [CrossRef] [PubMed]
220. Mendoza, O.; Bourdoncle, A.; Boulé, J.-B.; Brosh, R.M., Jr.; Mergny, J.-L. G-Quadruplexes and Helicases. *Nucleic Acids Res.* **2016**, *44*, 1989–2006. [CrossRef] [PubMed]
221. Sun, H.; Karow, J.K.; Hickson, I.D.; Maizels, N. The Bloom's Syndrome Helicase Unwinds G4 DNA. *J. Biol. Chem.* **1998**, *273*, 27587–27592. [CrossRef] [PubMed]
222. Gray, M.D.; Shen, J.-C.; Kamath-Loeb, A.S.; Blank, A.; Sopher, B.L.; Martin, G.M.; Oshima, J.; Loeb, L.A. The Werner Syndrome Protein Is a DNA Helicase. *Nat. Genet.* **1997**, *17*, 100–103. [CrossRef]
223. Mohaghegh, P.; Karow, J.K.; Brosh, R.M., Jr.; Bohr, V.A.; Hickson, I.D. The Bloom's and Werner's Syndrome Proteins Are DNA Structure-Specific Helicases. *Nucleic Acids Res.* **2001**, *29*, 2843–2849. [CrossRef]
224. Chatterjee, S.; Zagelbaum, J.; Savitsky, P.; Sturzenegger, A.; Huttner, D.; Janscak, P.; Hickson, I.D.; Gileadi, O.; Rothenberg, E. Mechanistic Insight into the Interaction of BLM Helicase with Intra-Strand G-Quadruplex Structures. *Nat. Commun.* **2014**, *5*, 5556. [CrossRef]
225. Caterino, M.; Paeschke, K. Action and Function of Helicases on RNA G-Quadruplexes. *Methods* **2022**, *204*, 110–125. [CrossRef]
226. Creacy, S.D.; Routh, E.D.; Iwamoto, F.; Nagamine, Y.; Akman, S.A.; Vaughn, J.P. G4 Resolvase 1 Binds Both DNA and RNA Tetramolecular Quadruplex with High Affinity and Is the Major Source of Tetramolecular Quadruplex G4-DNA and G4-RNA Resolving Activity in HeLa Cell Lysates. *J. Biol. Chem.* **2008**, *283*, 34626–34634. [CrossRef]
227. Heddi, B.; Cheong, V.V.; Martadinata, H.; Phan, A.T. Insights into G-Quadruplex Specific Recognition by the DEAH-Box Helicase RHAU: Solution Structure of a Peptide–Quadruplex Complex. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 9608–9613. [CrossRef]
228. Chen, M.C.; Tippana, R.; Demeshkina, N.A.; Murat, P.; Balasubramanian, S.; Myong, S.; Ferré-D'Amaré, A.R. Structural Basis of G-Quadruplex Unfolding by the DEAH/RHA Helicase DHX36. *Nature* **2018**, *558*, 465–469. [CrossRef]
229. Yangyuoru, P.M.; Bradburn, D.A.; Liu, Z.; Xiao, T.S.; Russell, R. The G-Quadruplex (G4) Resolvase DHX36 Efficiently and Specifically Disrupts DNA G4s via a Translocation-Based Helicase Mechanism. *J. Biol. Chem.* **2018**, *293*, 1924–1932. [CrossRef]
230. Johnson, J.E.; Cao, K.; Ryvkin, P.; Wang, L.-S.; Johnson, F.B. Altered Gene Expression in the Werner and Bloom Syndromes Is Associated with Sequences Having G-Quadruplex Forming Potential. *Nucleic Acids Res.* **2010**, *38*, 1114–1122. [CrossRef] [PubMed]
231. Smestad, J.A.; Maher, L.J. Relationships between Putative G-Quadruplex-Forming Sequences, RecQ Helicases, and Transcription. *BMC Med. Genet.* **2015**, *16*, 91. [CrossRef] [PubMed]
232. Tang, W.; Robles, A.I.; Beyer, R.P.; Gray, L.T.; Nguyen, G.H.; Oshima, J.; Maizels, N.; Harris, C.C.; Monnat, R.J. The Werner Syndrome RECQ Helicase Targets G4 DNA in Human Cells to Modulate Transcription. *Hum. Mol. Genet.* **2016**, *25*, 2060–2069. [CrossRef] [PubMed]
233. Lopes, J.; Piazza, A.; Bermejo, R.; Kriegsman, B.; Colosio, A.; Teulade-Fichou, M.-P.; Foiani, M.; Nicolas, A. G-Quadruplex-Induced Instability during Leading-Strand Replication. *EMBO J.* **2011**, *30*, 4033–4046. [CrossRef] [PubMed]
234. London, T.B.C.; Barber, L.J.; Mosedale, G.; Kelly, G.P.; Balasubramanian, S.; Hickson, I.D.; Boulton, S.J.; Hiom, K. FANCJ Is a Structure-Specific DNA Helicase Associated with the Maintenance of Genomic G/C Tracts. *J. Biol. Chem.* **2008**, *283*, 36132–36139. [CrossRef] [PubMed]
235. van Wietmarschen, N.; Merzouk, S.; Halsema, N.; Spierings, D.C.J.; Guryev, V.; Lansdorp, P.M. BLM Helicase Suppresses Recombination at G-Quadruplex Motifs in Transcribed Genes. *Nat. Commun.* **2018**, *9*, 271. [CrossRef] [PubMed]
236. González, V.; Guo, K.; Hurley, L.; Sun, D. Identification and Characterization of Nucleolin as a C-Myc G-Quadruplex-Binding Protein. *J. Biol. Chem.* **2009**, *284*, 23622–23635. [CrossRef]
237. González, V.; Hurley, L.H. The C-Terminus of Nucleolin Promotes the Formation of the c-MYC G-Quadruplex and Inhibits c-MYC Promoter Activity. *Biochemistry* **2010**, *49*, 9706–9714. [CrossRef]
238. Uribe, D.J.; Guo, K.; Shin, Y.-J.; Sun, D. Heterogeneous Nuclear Ribonucleoprotein K and Nucleolin as Transcriptional Activators of the Vascular Endothelial Growth Factor Promoter through Interaction with Secondary DNA Structures. *Biochemistry* **2011**, *50*, 3796–3806. [CrossRef] [PubMed]

239. Niu, K.; Xiang, L.; Jin, Y.; Peng, Y.; Wu, F.; Tang, W.; Zhang, X.; Deng, H.; Xiang, H.; Li, S.; et al. Identification of LARK as a Novel and Conserved G-Quadruplex Binding Protein in Invertebrates and Vertebrates. *Nucleic Acids Res.* **2019**, *47*, 7306–7320. [CrossRef] [PubMed]
240. Pramanik, S.; Chen, Y.; Song, H.; Khutsishvili, I.; Marky, L.A.; Ray, S.; Natarajan, A.; Singh, P.K.; Bhakat, K.K. The Human AP-Endonuclease 1 (APE1) Is a DNA G-Quadruplex Structure Binding Protein and Regulates KRAS Expression in Pancreatic Ductal Adenocarcinoma Cells. *Nucleic Acids Res.* **2022**, *50*, 3394–3412. [CrossRef] [PubMed]
241. Biffi, G.; Tannahill, D.; Balasubramanian, S. An Intramolecular G-Quadruplex Structure Is Required for Binding of Telomeric Repeat-Containing RNA to the Telomeric Protein TRF2. *J. Am. Chem. Soc.* **2012**, *134*, 11974. [CrossRef]
242. Cree, S.L.; Fredericks, R.; Miller, A.; Pearce, F.G.; Filichev, V.; Fee, C.; Kennedy, M.A. DNA G-Quadruplexes Show Strong Interaction with DNA Methyltransferases In Vitro. *FEBS Lett.* **2016**, *590*, 2870–2883. [CrossRef]
243. Mao, S.-Q.; Ghanbarian, A.T.; Spiegel, J.; Cuesta, S.M.; Beraldi, D.; Di Antonio, M.; Marsico, G.; Hänsel-Hertsch, R.; Tannahill, D.; Balasubramanian, S. DNA G-Quadruplex Structures Mould the DNA Methylome. *Nat. Struct. Mol. Biol.* **2018**, *25*, 951–957. [CrossRef]
244. Li, G.; Su, G.; Wang, Y.; Wang, W.; Shi, J.; Li, D.; Sui, G. Integrative Genomic Analyses of Promoter G-Quadruplexes Reveal Their Selective Constraint and Association with Gene Activation. *Commun. Biol.* **2023**, *6*, 625. [CrossRef]
245. Broxson, C.; Beckett, J.; Tornaletti, S. Transcription Arrest by a G Quadruplex Forming-Trinucleotide Repeat Sequence from the Human c-Myb Gene. *Biochemistry* **2011**, *50*, 4162–4172. [CrossRef]
246. Zheng, K.; Xiao, S.; Liu, J.; Zhang, J.; Hao, Y.; Tan, Z. Co-Transcriptional Formation of DNA:RNA Hybrid G-Quadruplex and Potential Function as Constitutional Cis Element for Transcription Control. *Nucleic Acids Res.* **2013**, *41*, 5533–5541. [CrossRef]
247. Marchetti, C.; Zyner, K.G.; Ohnmacht, S.A.; Robson, M.; Haider, S.M.; Morton, J.P.; Marsico, G.; Vo, T.; Laughlin-Toth, S.; Ahmed, A.A.; et al. Targeting Multiple Effector Pathways in Pancreatic Ductal Adenocarcinoma with a G-Quadruplex-Binding Small Molecule. *J. Med. Chem.* **2018**, *61*, 2500–2517. [CrossRef]
248. Boddupally, P.V.L.; Hahn, S.; Beman, C.; De, B.; Brooks, T.A.; Gokhale, V.; Hurley, L.H. Anticancer Activity and Cellular Repression of C-MYC by the G-Quadruplex-Stabilizing 11-Piperazinylquindoline Is Not Dependent on Direct Targeting of the G-Quadruplex in the c-MYC Promoter. *J. Med. Chem.* **2012**, *55*, 6076–6086. [CrossRef]
249. Robinson, J.; Raguseo, F.; Nuccio, S.P.; Liano, D.; Di Antonio, M. DNA G-Quadruplex Structures: More than Simple Roadblocks to Transcription? *Nucleic Acids Res.* **2021**, *49*, 8419–8431. [CrossRef]
250. Cogoi, S.; Paramasivam, M.; Membrino, A.; Yokoyama, K.K.; Xodo, L.E. The KRAS Promoter Responds to Myc-Associated Zinc Finger and Poly(ADP-Ribose) Polymerase 1 Proteins, Which Recognize a Critical Quadruplex-Forming GA-Element. *J. Biol. Chem.* **2010**, *285*, 22003–22016. [CrossRef] [PubMed]
251. Gao, J.; Zybailov, B.L.; Byrd, A.K.; Griffin, W.C.; Chib, S.; Mackintosh, S.G.; Tackett, A.J.; Raney, K.D. Yeast Transcription Co-Activator Sub1 and Its Human Homolog PC4 Preferentially Bind to G-Quadruplex DNA. *Chem. Commun. Camb. Engl.* **2015**, *51*, 7242–7244. [CrossRef] [PubMed]
252. Spiegel, J.; Cuesta, S.M.; Adhikari, S.; Hänsel-Hertsch, R.; Tannahill, D.; Balasubramanian, S. G-Quadruplexes Are Transcription Factor Binding Hubs in Human Chromatin. *Genome Biol.* **2021**, *22*, 117. [CrossRef]
253. Bai, L.; Morozov, A.V. Gene Regulation by Nucleosome Positioning. *Trends Genet.* **2010**, *26*, 476–483. [CrossRef]
254. Kim, N. The Interplay between G-Quadruplex and Transcription. *Curr. Med. Chem.* **2019**, *26*, 2898–2917. [CrossRef]
255. Kharel, P.; Balaratnam, S.; Beals, N.; Basu, S. The Role of RNA G-Quadruplexes in Human Diseases and Therapeutic Strategies. *WIREs RNA* **2020**, *11*, e1568. [CrossRef] [PubMed]
256. Jayaraj, G.G.; Pandey, S.; Scaria, V.; Maiti, S. Potential G-Quadruplexes in the Human Long Non-Coding Transcriptome. *RNA Biol.* **2012**, *9*, 81–89. [CrossRef]
257. Rouleau, S.G.; Garant, J.-M.; Bolduc, F.; Bisaillon, M.; Perreault, J.-P. G-Quadruplexes Influence Pri-microRNA Processing. *RNA Biol.* **2018**, *15*, 198–206. [CrossRef] [PubMed]
258. Balaratnam, S.; Hettiarachchilage, M.; West, N.; Piontkivska, H.; Basu, S. A Secondary Structure within a Human piRNA Modulates Its Functionality. *Biochimie* **2019**, *157*, 72–80. [CrossRef] [PubMed]
259. Schaeffer, C.; Bardoni, B.; Mandel, J.-L.; Ehresmann, B.; Ehresmann, C.; Moine, H. The Fragile X Mental Retardation Protein Binds Specifically to Its mRNA via a Purine Quartet Motif. *EMBO J.* **2001**, *20*, 4803–4813. [CrossRef] [PubMed]
260. Didiot, M.-C.; Tian, Z.; Schaeffer, C.; Subramanian, M.; Mandel, J.-L.; Moine, H. The G-Quartet Containing FMRP Binding Site in FMR1 mRNA Is a Potent Exonic Splicing Enhancer. *Nucleic Acids Res.* **2008**, *36*, 4902–4912. [CrossRef]
261. Marcel, V.; Tran, P.L.T.; Sagne, C.; Martel-Planche, G.; Vaslin, L.; Teulade-Fichou, M.-P.; Hall, J.; Mergny, J.-L.; Hainaut, P.; Van Dyck, E. G-Quadruplex Structures in TP53 Intron 3: Role in Alternative Splicing and in Production of P53 mRNA Isoforms. *Carcinogenesis* **2011**, *32*, 271–278. [CrossRef]
262. Gomez, D.; Lemarteleur, T.; Lacroix, L.; Mailliet, P.; Mergny, J.; Riou, J. Telomerase Downregulation Induced by the G-quadruplex Ligand 12459 in A549 Cells Is Mediated by hTERT RNA Alternative Splicing. *Nucleic Acids Res.* **2004**, *32*, 371–379. [CrossRef]
263. Conlon, E.G.; Lu, L.; Sharma, A.; Yamazaki, T.; Tang, T.; Shneider, N.A.; Manley, J.L. The C9ORF72 GGGGCC Expansion Forms RNA G-Quadruplex Inclusions and Sequesters hnRNP H to Disrupt Splicing in ALS Brains. *eLife* **2016**, *5*, e17820. [CrossRef]
264. Huang, H.; Zhang, J.; Harvey, S.E.; Hu, X.; Cheng, C. RNA G-Quadruplex Secondary Structure Promotes Alternative Splicing via the RNA-Binding Protein hnRNPF. *Genes Dev.* **2017**, *31*, 2296–2309. [CrossRef]

265. Zhang, J.; Harvey, S.E.; Cheng, C. A High-Throughput Screen Identifies Small Molecule Modulators of Alternative Splicing by Targeting RNA G-Quadruplexes. *Nucleic Acids Res.* **2019**, *47*, 3667–3679. [CrossRef]
266. Subramanian, M.; Rage, F.; Tabet, R.; Flatter, E.; Mandel, J.-L.; Moine, H. G-Quadruplex RNA Structure as a Signal for Neurite mRNA Targeting. *EMBO Rep.* **2011**, *12*, 697–704. [CrossRef]
267. Fay, M.M.; Anderson, P.J.; Ivanov, P. ALS/FTD-Associated C9ORF72 Repeat RNA Promotes Phase Transitions in Vitro and in Cells. *Cell Rep.* **2017**, *21*, 3573–3584. [CrossRef] [PubMed]
268. Wang, S.; Xu, Y. RNA Structure Promotes Liquid-to-Solid Phase Transition of Short RNAs in Neuronal Dysfunction. *Commun. Biol.* **2024**, *7*, 137. [CrossRef] [PubMed]
269. Wolfe, A.L.; Singh, K.; Zhong, Y.; Drewe, P.; Rajasekhar, V.K.; Sanghvi, V.R.; Mavrakis, K.J.; Jiang, M.; Roderick, J.E.; Van der Meulen, J.; et al. RNA G-Quadruplexes Cause eIF4A-Dependent Oncogene Translation in Cancer. *Nature* **2014**, *513*, 65–70. [CrossRef]
270. Murat, P.; Marsico, G.; Herdy, B.; Ghanbarian, A.; Portella, G.; Balasubramanian, S. RNA G-Quadruplexes at Upstream Open Reading Frames Cause DHX36- and DHX9-Dependent Translation of Human mRNAs. *Genome Biol.* **2018**, *19*, 229. [CrossRef]
271. Bhattacharyya, D.; Morris, M.J.; Kharel, P.; Mirihana Arachchilage, G.; Fedeli, K.M.; Basu, S. Engineered Domain Swapping Indicates Context Dependent Functional Role of RNA G-Quadruplexes. *Biochimie* **2017**, *137*, 147–150. [CrossRef]
272. Guilbaud, G.; Murat, P.; Recolin, B.; Campbell, B.C.; Maiter, A.; Sale, J.E.; Balasubramanian, S. Local Epigenetic Reprograming Induced by G-Quadruplex Ligands. *Nat. Chem.* **2017**, *9*, 1110–1117. [CrossRef]
273. Udugama, M.; Chang, F.T.M.; Chan, F.L.; Tang, M.C.; Pickett, H.A.; McGhie, J.D.R.; Mayne, L.; Collas, P.; Mann, J.R.; Wong, L.H. Histone Variant H3.3 Provides the Heterochromatic H3 Lysine 9 Tri-Methylation Mark at Telomeres. *Nucleic Acids Res.* **2015**, *43*, 10227–10237. [CrossRef] [PubMed]
274. Angelov, D.; Bondarenko, V.A.; Almagro, S.; Menoni, H.; Mongélard, F.; Hans, F.; Mietton, F.; Studitsky, V.M.; Hamiche, A.; Dimitrov, S.; et al. Nucleolin Is a Histone Chaperone with FACT-like Activity and Assists Remodeling of Nucleosomes. *EMBO J.* **2006**, *25*, 1669–1679. [CrossRef]
275. Saha, D.; Singh, A.; Hussain, T.; Srivastava, V.; Sengupta, S.; Kar, A.; Dhapola, P.; Dhople, V.; Ummanni, R.; Chowdhury, S. Epigenetic Suppression of Human Telomerase (hTERT) Is Mediated by the Metastasis Suppressor NME2 in a G-Quadruplex–Dependent Fashion. *J. Biol. Chem.* **2017**, *292*, 15205–15215. [CrossRef]
276. Hou, Y.; Li, F.; Zhang, R.; Li, S.; Liu, H.; Qin, Z.S.; Sun, X. Integrative Characterization of G-Quadruplexes in the Three-Dimensional Chromatin Structure. *Epigenetics* **2019**, *14*, 894–911. [CrossRef]
277. Dixon, J.R.; Selvaraj, S.; Yue, F.; Kim, A.; Li, Y.; Shen, Y.; Hu, M.; Liu, J.S.; Ren, B. Topological Domains in Mammalian Genomes Identified by Analysis of Chromatin Interactions. *Nature* **2012**, *485*, 376–380. [CrossRef] [PubMed]
278. Li, Y.; Haarhuis, J.H.I.; Sedeño Cacciatore, Á.; Oldenkamp, R.; van Ruiten, M.S.; Willems, L.; Teunissen, H.; Muir, K.W.; de Wit, E.; Rowland, B.D.; et al. The Structural Basis for Cohesin–CTCF-Anchored Loops. *Nature* **2020**, *578*, 472–476. [CrossRef] [PubMed]
279. Li, L.; Williams, P.; Ren, W.; Wang, M.Y.; Gao, Z.; Miao, W.; Huang, M.; Song, J.; Wang, Y. YY1 Interacts with Guanine Quadruplexes to Regulate DNA Looping and Gene Expression. *Nat. Chem. Biol.* **2021**, *17*, 161–168. [CrossRef] [PubMed]
280. Valton, A.-L.; Prioleau, M.-N. G-Quadruplexes in DNA Replication: A Problem or a Necessity? *Trends Genet.* **2016**, *32*, 697–706. [CrossRef] [PubMed]
281. Weitzmann, M.N.; Woodford, K.J.; Usdin, K. The Development and Use of a DNA Polymerase Arrest Assay for the Evaluation of Parameters Affecting Intrastrand Tetraplex Formation. *J. Biol. Chem.* **1996**, *271*, 20958–20964. [CrossRef] [PubMed]
282. Castillo Bosch, P.; Segura-Bayona, S.; Koole, W.; van Heteren, J.T.; Dewar, J.M.; Tijsterman, M.; Knipscheer, P. FANCJ Promotes DNA Synthesis through G-Quadruplex Structures. *EMBO J.* **2014**, *33*, 2521–2533. [CrossRef] [PubMed]
283. Kruisselbrink, E.; Guryev, V.; Brouwer, K.; Pontier, D.B.; Cuppen, E.; Tijsterman, M. Mutagenic Capacity of Endogenous G4 DNA Underlies Genome Instability in FANCJ-Defective C. Elegans. *Curr. Biol.* **2008**, *18*, 900–905. [CrossRef] [PubMed]
284. Sarkies, P.; Reams, C.; Simpson, L.J.; Sale, J.E. Epigenetic Instability Due to Defective Replication of Structured DNA. *Mol. Cell* **2010**, *40*, 703–713. [CrossRef]
285. Schiavone, D.; Jozwiakowski, S.K.; Romanello, M.; Guilbaud, G.; Guilliam, T.A.; Bailey, L.J.; Sale, J.E.; Doherty, A.J. PrimPol Is Required for Replicative Tolerance of G Quadruplexes in Vertebrate Cells. *Mol. Cell* **2016**, *61*, 161–169. [CrossRef]
286. Lemmens, B.; van Schendel, R.; Tijsterman, M. Mutagenic Consequences of a Single G-Quadruplex Demonstrate Mitotic Inheritance of DNA Replication Fork Barriers. *Nat. Commun.* **2015**, *6*, 8909. [CrossRef]
287. Koole, W.; van Schendel, R.; Karambelas, A.E.; van Heteren, J.T.; Okihara, K.L.; Tijsterman, M. A Polymerase Theta-Dependent Repair Pathway Suppresses Extensive Genomic Instability at Endogenous G4 DNA Sites. *Nat. Commun.* **2014**, *5*, 3216. [CrossRef]
288. Besnard, E.; Babled, A.; Lapasset, L.; Milhavet, O.; Parrinello, H.; Dantec, C.; Marin, J.-M.; Lemaitre, J.-M. Unraveling Cell Type–Specific and Reprogrammable Human Replication Origin Signatures Associated with G-Quadruplex Consensus Motifs. *Nat. Struct. Mol. Biol.* **2012**, *19*, 837–844. [CrossRef]
289. Valton, A.-L.; Hassan-Zadeh, V.; Lema, I.; Boggetto, N.; Alberti, P.; Saintomé, C.; Riou, J.-F.; Prioleau, M.-N. G4 Motifs Affect Origin Positioning and Efficiency in Two Vertebrate Replicators. *EMBO J.* **2014**, *33*, 732–746. [CrossRef] [PubMed]
290. Cayrou, C.; Coulombe, P.; Puy, A.; Rialle, S.; Kaplan, N.; Segal, E.; Méchali, M. New Insights into Replication Origin Characteristics in Metazoans. *Cell Cycle* **2012**, *11*, 658–667. [CrossRef] [PubMed]

291. Eaton, M.L.; Galani, K.; Kang, S.; Bell, S.P.; MacAlpine, D.M. Conserved Nucleosome Positioning Defines Replication Origins. *Genes Dev.* **2010**, *24*, 748–753. [CrossRef] [PubMed]
292. Berbenetz, N.M.; Nislow, C.; Brown, G.W. Diversity of Eukaryotic DNA Replication Origins Revealed by Genome-Wide Analysis of Chromatin Structure. *PLOS Genet.* **2010**, *6*, e1001092. [CrossRef]
293. Kouzine, F.; Wojtowicz, D.; Baranello, L.; Yamane, A.; Nelson, S.; Resch, W.; Kieffer-Kwon, K.-R.; Benham, C.; Casellas, R.; Przytycka, T.M.; et al. Permanganate/S1 Nuclease Footprinting Reveals Non-B DNA Structures with Regulatory Potential across a Mammalian Genome. *Cell Syst.* **2017**, *4*, 344–356.e7. [CrossRef]
294. Hoshina, S.; Yura, K.; Teranishi, H.; Kiyasu, N.; Tominaga, A.; Kadoma, H.; Nakatsuka, A.; Kunichika, T.; Obuse, C.; Waga, S. Human Origin Recognition Complex Binds Preferentially to G-Quadruplex-Preferable RNA and Single-Stranded DNA. *J. Biol. Chem.* **2013**, *288*, 30161–30171. [CrossRef]
295. Podhorecka, M.; Skladanowski, A.; Bozko, P. H2AX Phosphorylation: Its Role in DNA Damage Response and Cancer Therapy. *J. Nucleic Acids* **2010**, *2010*, e920161. [CrossRef]
296. Huang, R.-X.; Zhou, P.-K. DNA Damage Response Signaling Pathways and Targets for Radiotherapy Sensitization in Cancer. *Signal Transduct. Target. Ther.* **2020**, *5*, 60. [CrossRef]
297. Groelly, F.J.; Fawkes, M.; Dagg, R.A.; Blackford, A.N.; Tarsounas, M. Targeting DNA Damage Response Pathways in Cancer. *Nat. Rev. Cancer* **2023**, *23*, 78–94. [CrossRef] [PubMed]
298. Blackford, A.N.; Jackson, S.P. ATM, ATR, and DNA-PK: The Trinity at the Heart of the DNA Damage Response. *Mol. Cell* **2017**, *66*, 801–817. [CrossRef] [PubMed]
299. Siddiqui, M.S.; François, M.; Fenech, M.F.; Leifert, W.R. Persistent γH2AX: A Promising Molecular Marker of DNA Damage and Aging. *Mutat. Res. Mutat. Res.* **2015**, *766*, 1–19. [CrossRef] [PubMed]
300. Rogakou, E.P.; Pilch, D.R.; Orr, A.H.; Ivanova, V.S.; Bonner, W.M. DNA Double-Stranded Breaks Induce Histone H2AX Phosphorylation on Serine 139 *. *J. Biol. Chem.* **1998**, *273*, 5858–5868. [CrossRef] [PubMed]
301. Paull, T.T.; Rogakou, E.P.; Yamazaki, V.; Kirchgessner, C.U.; Gellert, M.; Bonner, W.M. A Critical Role for Histone H2AX in Recruitment of Repair Factors to Nuclear Foci after DNA Damage. *Curr. Biol.* **2000**, *10*, 886–895. [CrossRef]
302. Celeste, A.; Petersen, S.; Romanienko, P.J.; Fernandez-Capetillo, O.; Chen, H.T.; Sedelnikova, O.A.; Reina-San-Martin, B.; Coppola, V.; Meffre, E.; Difilippantonio, M.J.; et al. Genomic Instability in Mice Lacking Histone H2AX. *Science* **2002**, *296*, 922–927. [CrossRef]
303. Stucki, M.; Clapperton, J.A.; Mohammad, D.; Yaffe, M.B.; Smerdon, S.J.; Jackson, S.P. MDC1 Directly Binds Phosphorylated Histone H2AX to Regulate Cellular Responses to DNA Double-Strand Breaks. *Cell* **2005**, *123*, 1213–1226. [CrossRef]
304. Guo, Z.; Kumagai, A.; Wang, S.X.; Dunphy, W.G. Requirement for Atr in Phosphorylation of Chk1 and Cell Cycle Regulation in Response to DNA Replication Blocks and UV-Damaged DNA in Xenopus Egg Extracts. *Genes Dev.* **2000**, *14*, 2745–2756. [CrossRef]
305. Ahn, J.-Y.; Schwarz, J.K.; Piwnica-Worms, H.; Canman, C.E. Threonine 68 Phosphorylation by Ataxia Telangiectasia Mutated Is Required for Efficient Activation of Chk2 in Response to Ionizing Radiation1. *Cancer Res.* **2000**, *60*, 5934–5936.
306. Falck, J.; Mailand, N.; Syljuåsen, R.G.; Bartek, J.; Lukas, J. The ATM–Chk2–Cdc25A Checkpoint Pathway Guards against Radioresistant DNA Synthesis. *Nature* **2001**, *410*, 842–847. [CrossRef] [PubMed]
307. Sørensen, C.S.; Syljuåsen, R.G.; Falck, J.; Schroeder, T.; Rönnstrand, L.; Khanna, K.K.; Zhou, B.-B.; Bartek, J.; Lukas, J. Chk1 Regulates the S Phase Checkpoint by Coupling the Physiological Turnover and Ionizing Radiation-Induced Accelerated Proteolysis of Cdc25A. *Cancer Cell* **2003**, *3*, 247–258. [CrossRef]
308. Tibbetts, R.S.; Brumbaugh, K.M.; Williams, J.M.; Sarkaria, J.N.; Cliby, W.A.; Shieh, S.-Y.; Taya, Y.; Prives, C.; Abraham, R.T. A Role for ATR in the DNA Damage-Induced Phosphorylation of P53. *Genes Dev.* **1999**, *13*, 152–157. [CrossRef] [PubMed]
309. Shieh, S.-Y.; Ahn, J.; Tamai, K.; Taya, Y.; Prives, C. The Human Homologs of Checkpoint Kinases Chk1 and Cds1 (Chk2) Phosphorylate P53 at Multiple DNA Damage-Inducible Sites. *Genes Dev.* **2000**, *14*, 289–300. [CrossRef] [PubMed]
310. Maya, R.; Balass, M.; Kim, S.-T.; Shkedy, D.; Leal, J.-F.M.; Shifman, O.; Moas, M.; Buschmann, T.; Ronai, Z.; Shiloh, Y.; et al. ATM-Dependent Phosphorylation of Mdm2 on Serine 395: Role in P53 Activation by DNA Damage. *Genes Dev.* **2001**, *15*, 1067–1077. [CrossRef] [PubMed]
311. Craig, A.; Scott, M.; Burch, L.; Smith, G.; Ball, K.; Hupp, T. Allosteric Effects Mediate CHK2 Phosphorylation of the P53 Transactivation Domain. *EMBO Rep.* **2003**, *4*, 787–792. [CrossRef] [PubMed]
312. Jimeno, S.; Camarillo, R.; Mejias-Navarro, F.; Fernández-Ávila, M.J.; Soria-Bretones, I.; Prados-Carvajal, R.; Huertas, P. The Helicase PIF1 Facilitates Resection over Sequences Prone to Forming G4 Structures. *Cell Rep.* **2018**, *24*, 3262–3273.e4. [CrossRef]
313. Linke, R.; Limmer, M.; Juranek, S.A.; Heine, A.; Paeschke, K. The Relevance of G-Quadruplexes for DNA Repair. *Int. J. Mol. Sci.* **2021**, *22*, 12599. [CrossRef] [PubMed]
314. De Magis, A.; Manzo, S.G.; Russo, M.; Marinello, J.; Morigi, R.; Sordet, O.; Capranico, G. DNA Damage and Genome Instability by G-Quadruplex Ligands Are Mediated by R Loops in Human Cancer Cells. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 816–825. [CrossRef] [PubMed]
315. Miglietta, G.; Russo, M.; Capranico, G. G-Quadruplex–R-Loop Interactions and the Mechanism of Anticancer G-Quadruplex Binders. *Nucleic Acids Res.* **2020**, *48*, 11942–11957. [CrossRef]
316. Lange, T. de Shelterin: The Protein Complex That Shapes and Safeguards Human Telomeres. *Genes Dev.* **2005**, *19*, 2100–2110. [CrossRef]

317. Kang, C.; Zhang, X.; Ratliff, R.; Moyzis, R.; Rich, A. Crystal Structure of Four-Stranded Oxytricha Telomeric DNA. *Nature* **1992**, *356*, 126–131. [CrossRef] [PubMed]
318. Gill, M.L.; Strobel, S.A.; Loria, J.P. 205Tl NMR Methods for the Characterization of Monovalent Cation Binding to Nucleic Acids. *J. Am. Chem. Soc.* **2005**, *127*, 16723–16732. [CrossRef] [PubMed]
319. Smith, J.S.; Chen, Q.; Yatsunyk, L.A.; Nicoludis, J.M.; Garcia, M.S.; Kranaster, R.; Balasubramanian, S.; Monchaud, D.; Teulade-Fichou, M.-P.; Abramowitz, L.; et al. Rudimentary G-Quadruplex–Based Telomere Capping in Saccharomyces Cerevisiae. *Nat. Struct. Mol. Biol.* **2011**, *18*, 478–485. [CrossRef] [PubMed]
320. Takahama, K.; Takada, A.; Tada, S.; Shimizu, M.; Sayama, K.; Kurokawa, R.; Oyoshi, T. Regulation of Telomere Length by G-Quadruplex Telomere DNA- and TERRA-Binding Protein TLS/FUS. *Chem. Biol.* **2013**, *20*, 341–350. [CrossRef] [PubMed]
321. Moye, A.L.; Porter, K.C.; Cohen, S.B.; Phan, T.; Zyner, K.G.; Sasaki, N.; Lovrecz, G.O.; Beck, J.L.; Bryan, T.M. Telomeric G-Quadruplexes Are a Substrate and Site of Localization for Human Telomerase. *Nat. Commun.* **2015**, *6*, 7643. [CrossRef] [PubMed]
322. Drosopoulos, W.C.; Kosiyatrakul, S.T.; Schildkraut, C.L. BLM Helicase Facilitates Telomere Replication during Leading Strand Synthesis of Telomeres. *J. Cell Biol.* **2015**, *210*, 191–208. [CrossRef]
323. Crabbe, L.; Verdun, R.E.; Haggblom, C.I.; Karlseder, J. Defective Telomere Lagging Strand Synthesis in Cells Lacking WRN Helicase Activity. *Science* **2004**, *306*, 1951–1953. [CrossRef]
324. Vannier, J.-B.; Pavicic-Kaltenbrunner, V.; Petalcorin, M.I.R.; Ding, H.; Boulton, S.J. RTEL1 Dismantles T Loops and Counteracts Telomeric G4-DNA to Maintain Telomere Integrity. *Cell* **2012**, *149*, 795–806. [CrossRef]
325. Rizzo, A.; Iachettini, S.; Zizza, P.; Cingolani, C.; Porru, M.; Artuso, S.; Stevens, M.; Hummersone, M.; Biroccio, A.; Salvati, E.; et al. Identification of Novel RHPS4-Derivative Ligands with Improved Toxicological Profiles and Telomere-Targeting Activities. *J. Exp. Clin. Cancer Res.* **2014**, *33*, 81. [CrossRef]
326. Gauthier, L.R.; Granotier, C.; Hoffschir, F.; Etienne, O.; Ayouaz, A.; Desmaze, C.; Mailliet, P.; Biard, D.S.; Boussin, F.D. Rad51 and DNA-PKcs Are Involved in the Generation of Specific Telomere Aberrations Induced by the Quadruplex Ligand 360A That Impair Mitotic Cell Progression and Lead to Cell Death. *Cell. Mol. Life Sci.* **2012**, *69*, 629–640. [CrossRef]
327. Uringa, E.-J.; Lisaingo, K.; Pickett, H.A.; Brind'Amour, J.; Rohde, J.-H.; Zelensky, A.; Essers, J.; Lansdorp, P.M. RTEL1 Contributes to DNA Replication and Repair and Telomere Maintenance. *Mol. Biol. Cell* **2012**, *23*, 2782. [CrossRef]
328. Lin, W.; Sampathi, S.; Dai, H.; Liu, C.; Zhou, M.; Hu, J.; Huang, Q.; Campbell, J.; Shin-Ya, K.; Zheng, L.; et al. Mammalian DNA2 Helicase/Nuclease Cleaves G-Quadruplex DNA and Is Required for Telomere Integrity. *EMBO J.* **2013**, *32*, 1425–1439. [CrossRef]
329. Zhang, M.; Wang, B.; Li, T.; Liu, R.; Xiao, Y.; Geng, X.; Li, G.; Liu, Q.; Price, C.M.; Liu, Y.; et al. Mammalian CST Averts Replication Failure by Preventing G-Quadruplex Accumulation. *Nucleic Acids Res.* **2019**, *47*, 5243–5259. [CrossRef]
330. DeJesus-Hernandez, M.; Mackenzie, I.R.; Boeve, B.F.; Boxer, A.L.; Baker, M.; Rutherford, N.J.; Nicholson, A.M.; Finch, N.A.; Gilmer, H.F.; Adamson, J.; et al. Expanded GGGGCC Hexanucleotide Repeat in Non-Coding Region of C9ORF72 Causes Chromosome 9p-Linked Frontotemporal Dementia and Amyotrophic Lateral Sclerosis. *Neuron* **2011**, *72*, 245–256. [CrossRef]
331. Renton, A.E.; Majounie, E.; Waite, A.; Simón-Sánchez, J.; Rollinson, S.; Gibbs, J.R.; Schymick, J.C.; Laaksovirta, H.; van Swieten, J.C.; Myllykangas, L.; et al. A Hexanucleotide Repeat Expansion in C9ORF72 Is the Cause of Chromosome 9p21-Linked ALS-FTD. *Neuron* **2011**, *72*, 257–268. [CrossRef]
332. Reddy, K.; Zamiri, B.; Stanley, S.Y.R.; Macgregor, R.B.; Pearson, C.E. The Disease-Associated r(GGGGCC)n Repeat from the C9orf72 Gene Forms Tract Length-Dependent Uni- and Multimolecular RNA G-Quadruplex Structures. *J. Biol. Chem.* **2013**, *288*, 9860–9866. [CrossRef]
333. Fratta, P.; Mizielinska, S.; Nicoll, A.J.; Zloh, M.; Fisher, E.M.C.; Parkinson, G.; Isaacs, A.M. C9orf72 Hexanucleotide Repeat Associated with Amyotrophic Lateral Sclerosis and Frontotemporal Dementia Forms RNA G-Quadruplexes. *Sci. Rep.* **2012**, *2*, 1016. [CrossRef] [PubMed]
334. Ellis, N.A.; German, J. Molecular Genetics of Bloom's Syndrome. *Hum. Mol. Genet.* **1996**, *5*, 1457–1463. [CrossRef] [PubMed]
335. Wu, L.; Hickson, I.D. The Bloom's Syndrome Helicase Suppresses Crossing over during Homologous Recombination. *Nature* **2003**, *426*, 870–874. [CrossRef] [PubMed]
336. Davies, S.L.; North, P.S.; Hickson, I.D. Role for BLM in Replication-Fork Restart and Suppression of Origin Firing after Replicative Stress. *Nat. Struct. Mol. Biol.* **2007**, *14*, 677–679. [CrossRef] [PubMed]
337. Karow, J.K.; Constantinou, A.; Li, J.-L.; West, S.C.; Hickson, I.D. The Bloom's Syndrome Gene Product Promotes Branch Migration of Holliday Junctions. *Proc. Natl. Acad. Sci. USA* **2000**, *97*, 6504–6508. [CrossRef] [PubMed]
338. Hanahan, D.; Weinberg, R.A. Hallmarks of Cancer: The Next Generation. *Cell* **2011**, *144*, 646–674. [CrossRef]
339. Biffi, G.; Tannahill, D.; Miller, J.; Howat, W.J.; Balasubramanian, S. Elevated Levels of G-Quadruplex Formation in Human Stomach and Liver Cancer Tissues. *PLoS ONE* **2014**, *9*, e102711. [CrossRef] [PubMed]
340. Savva, L.; Georgiades, S.N. Recent Developments in Small-Molecule Ligands of Medicinal Relevance for Harnessing the Anticancer Potential of G-Quadruplexes. *Molecules* **2021**, *26*, 841. [CrossRef] [PubMed]
341. Miglietta, G.; Marinello, J.; Russo, M.; Capranico, G. Ligands Stimulating Antitumour Immunity as the next G-Quadruplex Challenge. *Mol. Cancer* **2022**, *21*, 180. [CrossRef]

342. Hilton, J.; Gelmon, K.; Bedard, P.L.; Tu, D.; Xu, H.; Tinker, A.V.; Goodwin, R.; Laurie, S.A.; Jonker, D.; Hansen, A.R.; et al. Results of the Phase I CCTG IND.231 Trial of CX-5461 in Patients with Advanced Solid Tumors Enriched for DNA-Repair Deficiencies. *Nat. Commun.* **2022**, *13*, 3607. [CrossRef]
343. Miglietta, G.; Russo, M.; Duardo, R.C.; Capranico, G. G-Quadruplex Binders as Cytostatic Modulators of Innate Immune Genes in Cancer Cells. *Nucleic Acids Res.* **2021**, *49*, 6673–6686. [CrossRef]
344. Lopes-Nunes, J.; Oliveira, P.A.; Cruz, C. G-Quadruplex-Based Drug Delivery Systems for Cancer Therapy. *Pharmaceuticals* **2021**, *14*, 671. [CrossRef]