A functional map of lncRNAs in adaptive phenotypes of breast cancer cells

Bianca Giuliani
Italian Institute of technology
University of Milan
Matricola n. R12115

Tutor*: Dr Francesco Nicassio*
Italian Institute of technology

Added supervisor: *Dr Matteo Marzi*
Italian Institute of technology

Table of Contents

Table of figures

List of abbreviations

AI, Aromatase Inhibitors

ALDH, aldehyde-dehydrogenase

AS, anti-sense

CD, Cluster of Differentiation

ceRNA, competing endogenous RNA

CRISPRa, CRISPR*activation*

CRISPRi, CRISPR*interference*

CSC, Cancer Stem Cells

dCas9, "dead"-Cas9

DTP, Drug-Tolerant Persister

FDR, False Discovery Rate

EMT, epithelial to mesenchymal transition

ER, Estrogen Receptor

ES, Enrichment Score

FBS, Fetal Bovine Serum

HER2, Human Epidermal Growth Factor Receptor 2

IFN-γ, interferon-γ

KD, Knock-Down

KRAB, Kruppel-associated box

lncRNAs, long non-coding RNAs

MaSC, Mammary Stem Cell

mESC, mouse Embryonic Stem Cell

miRNA, micro-RNA

MLE, Maximum-Likelihood estimation

MRD, Minimal Residual Disease

NACT, Neo-adjuvant chemotherapy

NES, Negative Enrichment Score

NET-cage, Native elongating transcript–cap analysis of gene expression

NTC, Non-Targeting control

PAM50, Prediction analysis of Microarray 50

PAM, Protospacer Adjacent Motif

PBS, Phosphate Buffer Saline

PCG, Protein Coding Gene

pEMT, partial EMT

PR, Progesterone Receptor

RBP, RNA binding protein

RNApol2, RNA polymerase II

scRNA-seq, single-cell RNA-seq

SFE, Sphere-Forming Efficiency

sgRNA, single guide RNA

shRNA, short hairpin RNA

snoRNA, Small Nucleolar RNA

snRNA, Small Nuclear RNAs

TAD, Topologically Associating Domain

TF, Transcription Factor

TGF-β, Transforming-Growth Factor-β

TLDU, terminal lobular-ductal unit

TNBC, Triple Negative Breast Cancer

TSS, Transcription Start Site

Abstract

In this project, we investigate the contribution of long non-coding RNAs (lncRNAs) in the acquisition of adaptive phenotypes in breast cancer. LncRNAs have multiple regulatory functions that define these transcripts as fine regulators of gene expression, engaging in different oncogenic networks. However, the functional classification of these RNA molecules is still limited. We aim at building a functional map of lncRNAs in the context non-genetic adaptive mechanisms occurring in breast cancer by means of pooled CRISPR*interference* (CRISPRi) screenings. We selected 620 lncRNAs for their positive association with models of adaptive responses in breast cancer. Specifically, we evaluated the expression of adaptive response signatures in primary tumors, modeled the chemotolerance to the neo-adjuvant drug Paclitaxel and evaluated the inherent plasticity of the MaSC-like population of the HMLE cell line.

We designed a CRISPRi library for the simultaneous perturbation of 620 lncRNAs. To precisely map the TSS of candidates lncRNAs, the design of the lentiviral library was aided by NET-Cage data.

We tested how the perturbation of lncRNAs modulates the adaptive properties of the TNBC cell line SUM159PT by growing cells in increasingly stringent conditions (2D, 3D, chemoresponse, in vivo). In this setting cells must address different responses to survive.

Overall, we could identify 18.8% of genes DROP-OUTs, and 19% of genes DROP-INs in one or multiple screenings. We will broaden the characterization of lncRNAs hits by evaluating the transcriptomic response upon their perturbation.

With this work, we outline a sound and reliable approach for identifying lncRNAs that are relevant in cancer and provide new insights into the regulatory roles lncRNAs play in cancer cell plasticity.

1. Introduction

1.1 Breast Cancer and adaptation mechanisms during cancer evolution

1.1.1 The mammary gland: an intrinsically plastic organ

The mammary gland is composed of a glandular epithelium responsible for lactation. This organ peculiarly completes its development post-natal and remains prone to large remodeling through several decades. The mammary gland is organized in a ductal-alveolar tree that digs its branches into the stroma of the organ (the fat pad) and opens in the nipple. The functional unit of the gland is the terminal lobular-ductal unit (TLDUs), generally composed of two types of cells: secretory inner cuboidal luminal cells and basal myoepithelial cells in direct contact with the basal membrane (**Figure 1**). During embryogenesis the mammary gland consists in a rudimental ductal structure, a "primordium". After birth, the development of the gland arrests until puberty when the ovarian hormones re-launch the development of the gland. The rudimental ducts differentiate in the multilayered structure that elongates and produce side branches forming the entire ductal tree. During pregnancy, the inner epithelial layer differentiates into milk-secretory alveoli. The gland is then subject to involution that partially repristinate the status of the gland before pregnancy. The gland remains susceptible to multiple remodeling steps for each pregnancy until menopause (**Figure 2**) (Gjorevski and Nelson, 2011). The dynamic re-arrangement in the architecture of the gland implies a stem compartment able to maintain the organ plasticity and its cellular complexity. It was indeed demonstrated the existence of mammary stem cells (MaSCs) that own the basic features of adult stem cells: they can regenerate the mammary gland when transplanted *in vivo* into cleared mammary fat pad of mice (Shackleton et al., 2006; Stingl et al., 2006), they divide by asymmetric divisions and are normally placed in a quiescent state. The MaSCs are intrinsically resistant to anoikis and can grow anchorage-independently (Lloyd-Lewis et al., 2017). The precise hierarchical structure of progenitors in the mammary gland remains to be clarified (Koren and Bentires-Alj, 2015) (**Figure 3**). Nonetheless, these characteristics define the mammary gland as an heterogenous group of epithelial cells that are prone to plastic changes.
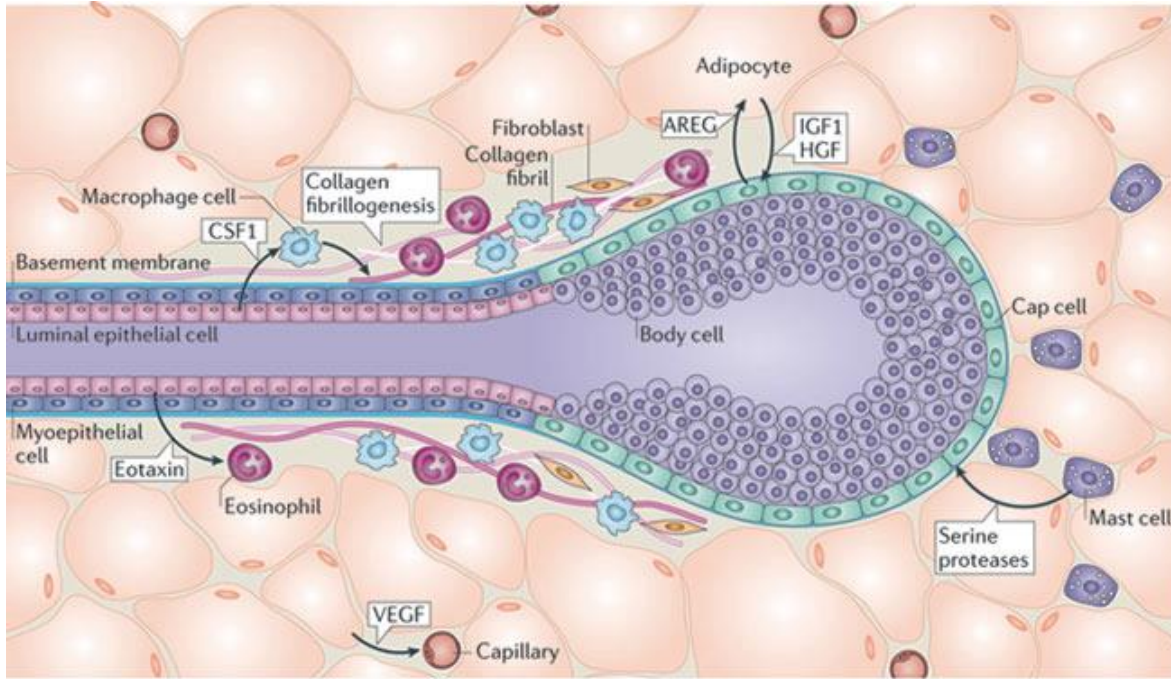
Figure 1: Representation of the functional unit of the mammary gland – The picture represents the components of mammary stroma and a terminal lobular-ductal unit of luminal epithelial cells (in pink) and basal myoepithelial cells (in violet), the duct terminates into a terminal end bud. Figure from (Gjorevski and Nelson, 2011).
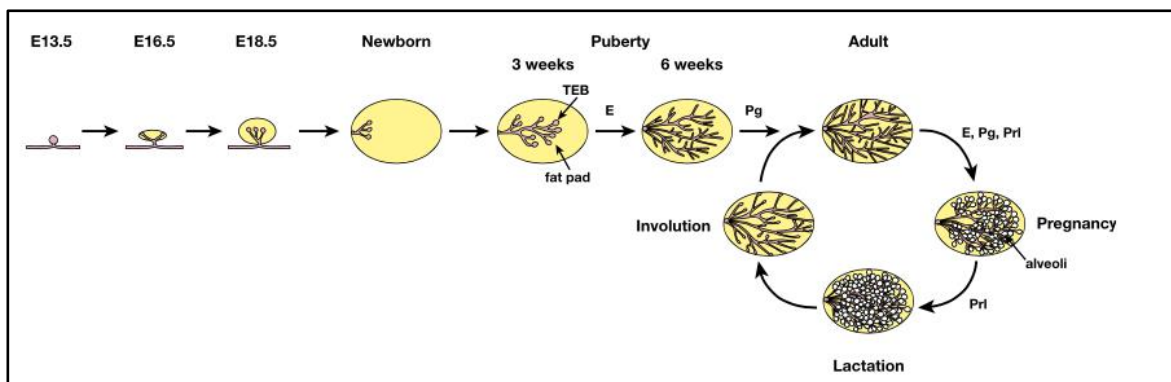
Figure 2: Development of the mammary gland – Different stages of mammary development from embryonic to adult life in mice. TEB= Terminal End Buds, formation at the extremities of the TLDU from which the elongation and branching of the ducts starts. E=Estrogen, Pg= Progesterone, Prl= Prolactin. Figure from (Visvader and Stingl, 2014).
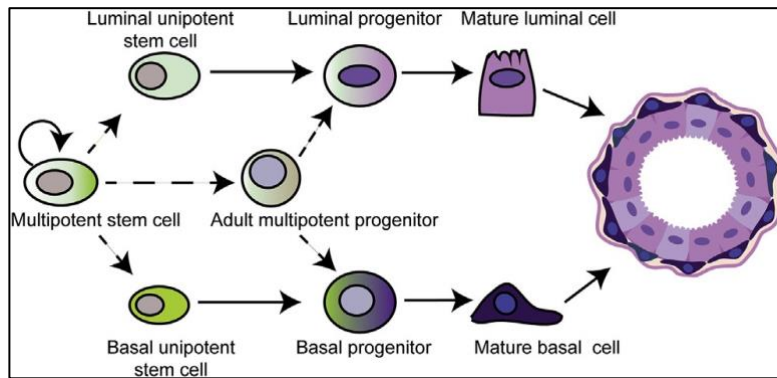
Figure 3: Model of mammary cells hierarchy – It is still unclear whether the homeostasis of the mammary gland is maintained by an adult multipotent stem cell giving rise to both luminal and basal progenitors or the two lineages are sustained in the adult life by basal and luminal unipotent stem cells. Figure from (Koren and Bentires-Alj, 2015).

1.1.2 The heterogeneity of breast cancer

The heterogeneity and plasticity of the mammary gland is reflected by its main disease: the breast cancer. Breast cancer is the leading cause of death among adult women, it is a complex and multi-faceted disease that was historically classified according to the immunohistochemical status of the hormonal receptors. This classification distinguishes ER+ tumors (tumors immunoreactive for progesterone (PR) and estrogen receptors (ER)), HER2+ tumors (tumors positive for the expression of the human epidermal growth factor receptor 2 (HER2)) and triple-negative breast cancer (TNBC) that lacks the expression of the three receptors (ER-, PR-, HER2-). This classification expanded with the molecular characterization provided by the introduction of the PAM50 (Prediction analysis of Microarray 50) (Hu et al., 2006), which introduced a molecular signature of 50 expressed genes that further distinguishes the tumors in 5 different categories: Luminal A, luminal B, HER2 amplified, Basal-like and Claudin-low (Parker et al., 2009; Prat et al., 2012; Prat and Perou, 2011) . This characterization shows that different tumor subtypes resemble different gene expression profiles of the mammary gland. Luminal A and Luminal B tumors approximate more closely the transcriptome of mature luminal cells, while Basal-like and claudin-low tumors are more akin to the transcriptome of basal and less-differentiated MaSCs (**Figure 4**). This nomenclature does not refer to the cell-of-orgin of the tumor. It was demonstrated indeed that luminal progenitors are equally able to give rise to basal-like tumors (Molyneux et al., 2010). The molecular subtypes classification helps describing the phenotype of the different breast tumors, suggests prognosis and operatively defines the eligibility to different therapeutic options.

Therapies for breast cancer involve either local treatment (surgery and radiotherapy) as well as systemic treatments that include both targeted therapies as well as cytotoxic chemotherapy

(Cardoso et al., 2019). Upon detection, specialists evaluate the size of the primary mass and involvement of proximal lymph nodes. Smaller and less invasive masses are surgically resected, and patients are followed-up with different pharmacological options according to the histological and molecular diagnosis with the goal of treating distant micro metastasis and prolong survival. Larger and higher-grade lesions are treated first with Neo-Adjuvant Chemotherapy (NACT) with the aim of reducing the size and downgrading tumors to increase the success of surgery (Loibl et al., 2021).

Hormone therapy is the standard pharmacological option for ER+/PR+ breast tumors. Tamoxifen is the leading drug used for the treatment of such malignancy. Tamoxifen interferes with the pro-survival activity of estrogen hormone. Conspicuous data support that this drug reduces the risk of recurrence and metastasis regardless of the age of patients (Regan et al., 2016). Tamoxifen is administered for 10 years after primary treatment and recent studies suggest the beneficial effect of the combination with other hormone agonist (Luteinizing hormone agonist and releasing hormone agonist) and aromatase inhibitors (AI) (Regan et al., 2016).

The treatment of HER2+ breast cancer has also benefit of the introduction of targeted agents such as Trastuzumab. This drug is a monoclonal antibody that specifically targets the HER2 receptor inhibiting the downstream signaling. Recent evidence supports the efficacy of the combinatorial treatment with trastuzumab and cytotoxic agents leading to greatly improved survival for patients suffering of this form of breast cancer (von Minckwitz et al., 2017).

Patients displaying TN tumors are not suitable for targeted regimens and treated with cytotoxic chemotherapy and display the worst prognosis (Eliyatkın et al., 2015). The first-line chemotherapeutic option consists in combinations of taxanes and anthracycline in

sequence (Del Mastro et al., 2015) or alternatively cyclophosphamide, methotrexate, 5-fluorouracil (CMF) can be proposed (Cardoso et al., 2019).
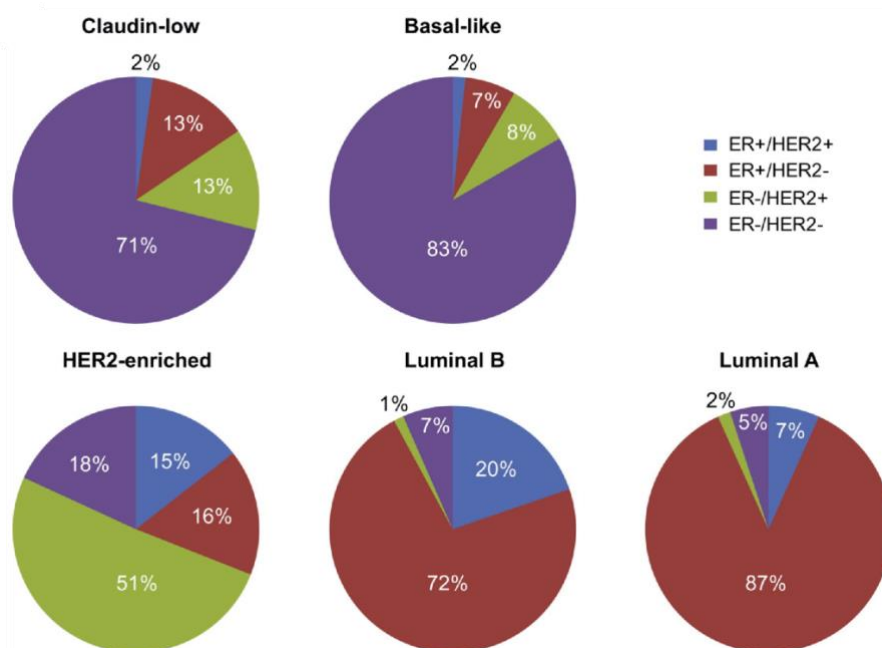
Figure 4: Distribution of immunological clinical groups per molecular subtype – The pie chart shows that claudin-low and basal-like tumors collect the majority of TN (ER-/HER2-). Figure adapted form (Prat and Perou, 2011).

However, despite similar diagnosis, the response to therapies for breast cancer shows great variability (Ramos and Bentires-Alj, 2019). This variability can be partially explained by the intrinsic phenotypic and functional heterogeneity existing in breast cancer (De Palma and Hanahan, 2012; Meacham and Morrison, 2013). The heterogeneity comes from different sources (**Figure 5**). Inter-tumor heterogeneity among breast cancers derives from the unique combination of driver mutation and the status of the cell-of-origin where it occurs (Melchor et al., 2014). The intra-tumor heterogeneity can be genetic, deriving from the accumulation of diverse mutations in relevant genes in sub-clones of cells that result genetically heterogenous (Gerlinger et al., 2012). Alternatively, cancer cells display phenotypic heterogeneity regardless of their genetic background but due to cell plasticity that place cancer cells in transcriptionally heterogenous states. Indeed, in breast tumors not all cells have the same tumor-initiating capacity, but this potential is restricted to a population of Cancer Stem Cells (CSCs) that, just like normal mammary stem cells have self-renewal capacity and the ability to regrow the whole tumor. CSCs are in phenotypic equilibrium within the tumor bulk and preserve the heterogeneity of the tumor (D'Amato et al., 2012; Gupta et al., 2011; Van Keymeulen et al., 2015). Lastly, the microenvironment plays a role

in the development of the tumor and so, cancer cells display heterogenic phenotype depending on microenvironmental cues (Ghajar et al., 2013). All sources of heterogeneity coexist in breast tumors and provide cancer cells with multiple escapes from therapy and other challenging environments.
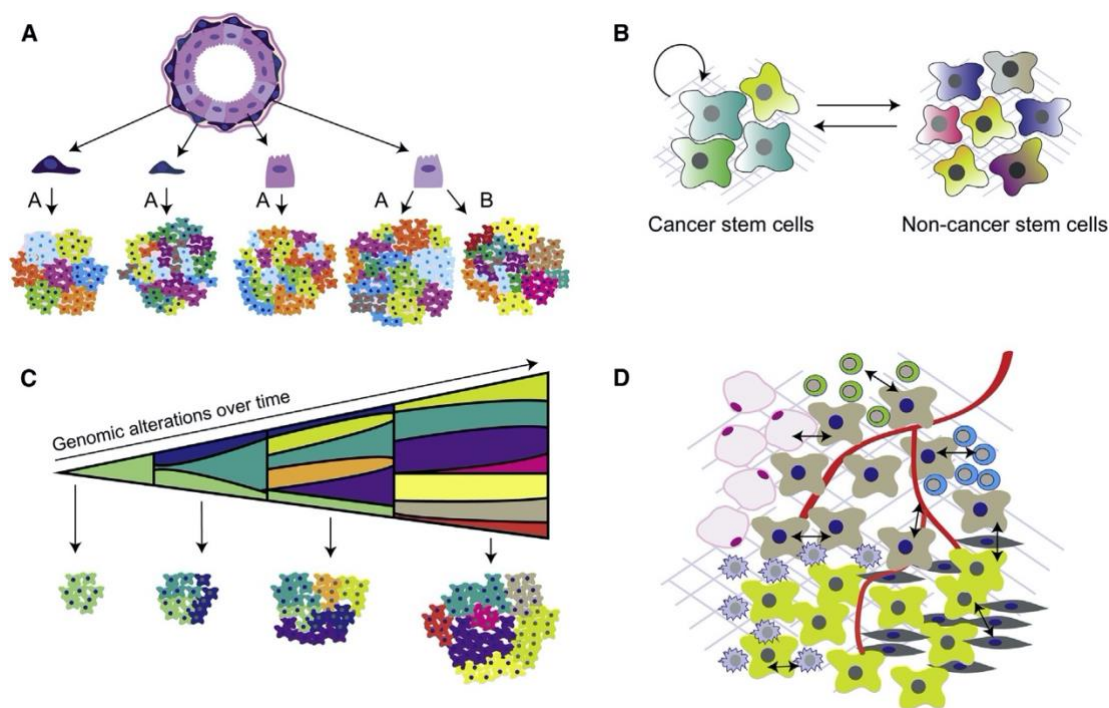
Figure 5: Sources of breast cancer heterogeneity – A) The combination of the tumor initiating genetic alteration and the cell of origin produce phenotypically distinct tumors. **B)** Cell plasticity place cells in transcriptionally heterogeneous states. **C)** Tumoral multi-step genomic evolution produces genetically distinct subclones. **D)** The interaction with the microenvironment impact on tumor diversity. Figure adapted from (Koren and Bentires-Alj, 2015).

1.1.3 Plasticity drives adaptation mechanisms in cancer

Despite treatment, some tumors progress by acquisition of therapy resistance, invasiveness, and metastatic capacity. Tumor evolution is a dynamic process in which cancer cells are exposed to many challenges, such as nutritional limitations, hypoxia, cytotoxic stress. Plasticity is the ability of cells to transiently alter their phenotypes and adapt to changing circumstances (Bakir et al., 2020).

Beside the genetic diversity, tumors are structured in phenotypically heterogeneous components (Brock et al., 2009; Pogrebniak and Curtis, 2018) (See **1.1.2**). The introduction of lineage-tracing tools paired with single-cell RNA-seq (scRNA-seq) technologies allowed unprecedent knowledge of complex phenotypes such as resistance to therapies (Echeverria et al., 2019; Oren et al., 2021) and metastasis formation (Echeverria et al., 2018; Simeonov

et al., 2021). Both phenotypes are ascribable to the selection of rare pre-existing clones genetically similar to the primary tumor but transcriptionally distinct. These observations make space for the non-genetic interpretation of disease progression.

All therapeutic options for breast cancer leave minimal residual disease (MRD), from which tumor unavoidably re-grows causing relapse (Jansen et al., 2013; Magnani et al., 2013; Thomas et al., 2011). This is particularly true for TNBC patients which are usually treated, before primary tumor resection, with NACT (Asselain et al., 2017).

Lacking molecular surface targets, NACT hits the tumor with a combination of taxanes (mitotic inhibitors) and anthracyclines (DNA intercalators). Nonetheless 30%-50% of patients show resistance and development of distant metastasis which remain largely incurable (Foulkes et al., 2010; Liedtke et al., 2008).

The mutational profile and transcriptomic response to NACT has been investigated pre- and post-treatment, in a cohort of TNBC patients (Kim et al., 2018a). In some cases, mutations were pre-existing to the treatment and the surviving clones actively adapted to the treatment through chemo-therapy induced transcriptional reprogramming, converging the transcriptome of post-treatment cells towards common pathways such as the acquisition of the epithelial to mesenchymal (EMT) transcription program, AKT1 survival pathway, hypoxia, angiogenesis and extracellular matrix (ECM) degradation that were previously associated with chemo-resistance (Kim et al., 2007; Lu et al., 2012; Oskarsson, 2013; Petit et al., 2016). With similar importance, the transcriptional adaptation to NACT was reported to be reversible in terms of transcriptomic response and drug-sensitivity (Echeverria et al., 2019). This is consistent with the clinical evidence that some tumors respond to the same therapy following a "drug holiday" (Nardi et al., 2004).

The non-genetic mechanisms of drug-response have been reported in multiple cancers (Bell et al., 2019; Shaffer et al., 2018; Sharma et al., 2010). This process has been accounted to the survival of drug-tolerant persisters (DTPs). In reaction to the therapeutic pressure, some cells survive the treatment by entering in a less proliferative and drug-resilient state, not driven by mutations (Shen et al., 2020). Having remained undetected, DTPs re-establish proliferation, being ultimately responsible for relapse. The acquisition of a less proliferative state in DTPs establishes a difference with genetic resistance in which, due to the acquired mutations, cells have unaltered proliferation in presence of the drug. Therefore, the process of altered proliferation in response to the drug is more precisely referred to as chemo-tolerance (Marine et al., 2020).

The origin of DTPs can be accounted to the presence, in the heterogenous primary tumors, of cells in a primed transcriptional program of drug-tolerance (Emert et al., 2021) or,

alternatively, by an active drug-induced reprogramming occurring in rare clones (Echeverria et al., 2019; Rambow et al., 2018) (**Figure 6**).
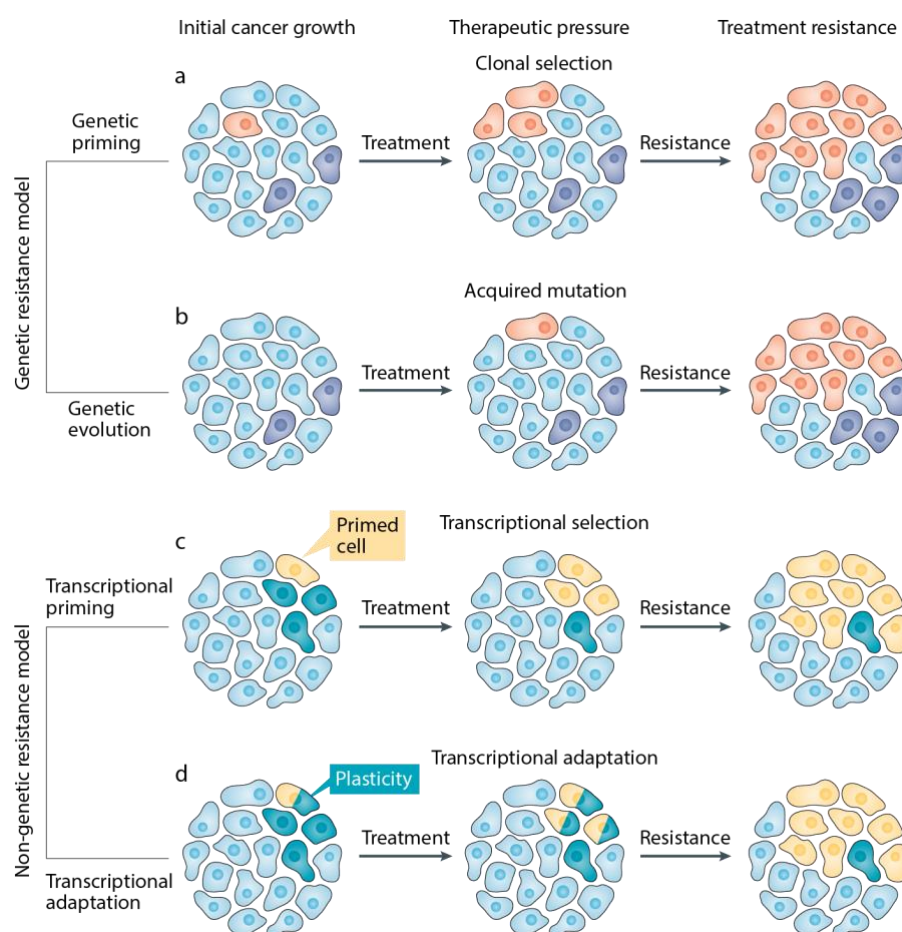
Figure 6: Origin of treatment resistance – A) The treatment induce the clonal selection of rare pre-existing clones (pre-existing therapy resistance or genetic priming). **B)** The drug treatment allows the accumulation of novel mutation that can be causative of treatment resistance (acquired resistance). **C)** Clones are selected by the treatment because of their pre-existing transcriptional drug-tolerant state (transcriptional priming). **D)** Cancer cells adapt their phenotype in response to the therapeutic pressure. Figure from (Marine et al., 2020).

In vitro DTPs can be generated by exposing cells to largely cytotoxic concentration of therapeutic drugs (Sharma et al., 2010). Drug-tolerance is realized by different means that are just being uncovered. DTP cells are placed in a dormant G0 state of reversible cell cycle arrest by activating quiescence programs and thus, eluding the anti-proliferative activity of many chemotherapeutic agents (Recasens and Munoz, 2019). Importantly, disseminated cancer cells undergo dormancy when they leave the primary tumor site and seed in a new niche, where they can remain quiescent for a long time (Sosa et al., 2014). This shared phenotype creates a conceptual link between the two events.

The acquisition of a less proliferative state can be acquired thanks to the coordinate work of different stimuli and cell responses. In breast cancer, the acquisition of a quiescent state has been linked to the activation of Notch signalling (Abravanel et al., 2015), hypoxia (Samanta

et al., 2014), Unfolded Protein Response (Ranganathan et al., 2006) and activation of the Nuclear Factor-κB (NFκB) inflammatory response (El-Shennawy et al., 2018).

Microenvironmental stimuli, such as the production of the Transforming-Growth Factor-β (TGF-β) (Prunier et al., 2018) and the inflammatory cytokine interferon-γ (IFN-γ) secreted by senescent cells, can activate dormancy (Müller-Hermelink et al., 2008).

DTPs undergo metabolic changes to sustain the quiescent state. This is achieved by reducing the glucose consumption and shifting through a mitochondrial oxidative respiration and peroxisomal fatty acid β-oxydation (Viale et al., 2014). Moreover, the activation of autophagy, by recycling substrates fuels these metabolisms (Li et al., 2019).

The dormant state of DTPs resembles the inner quiescent and xenobiotic-resistant state of CSCs (Raha et al., 2014; Sharma et al., 2010). Indeed, dormant cells express common stemness markers: the CD44 surface marker (Hangauer et al., 2017) and increased activity of the de-toxifying enzyme aldehyde-dehydrogenase (ALDH) (Raha et al., 2014; Sharma et al., 2010). It has been widely demonstrated that the exploitation of the epithelial to mesenchymal transcription program (EMT) promotes the acquisition of a de-differentiated state endowing cells with CSCs features and increased drug-resistance properties such as the upregulation of antioxidant responses (Del Vecchio et al., 2014; Feng et al., 2014) and activation of pro-survival pathways (Wu et al., 2015).

EMT is a trans-differentiation program physiologically implied during embryonic development (Hay, 2005). EMT allows the reversible trans-differentiation from an epithelial to a mesenchymal state, in a complex transcriptional mechanism that realize profound phenotypic changes. EMT implies the repression of epithelial genes (e.g. E-cadherin) and coordinate expression of genes associated to a mesenchymal identity (e.g. vimentin). Importantly, cells undergoing EMT are placed in a spectrum of intermediate phenotypes characterized by the simultaneous expression of epithelial and mesenchymal markers (partial EMT, pEMT) as opposed to discrete uniquely "epithelial" or "mesenchymal" states. The hybrid pEMT status is highly tumorigenic and dynamic (Kröger et al., 2019; Simeonov et al., 2021). In pEMT the aggressiveness of both phenotypic states is maximized (**Figure 7**). DTPs exploits the mesenchymal drug-resistant and less-proliferating traits to survive the cytotoxic insult and restore more epithelial characteristics to re-instate tumor growth (Shen et al., 2020).
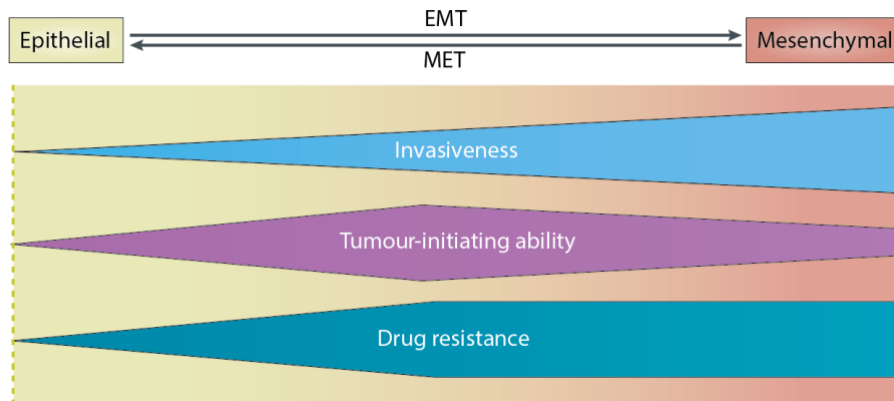
Figure 7: Properties of the EMT phenotypic transition – EMT is responsible for enabling invasiveness, metastatic seeding and tumor resistance of cancer cells. The tumor-initiating ability of EMT-undergoing cells peaks when cells acquire an intermediate phenotype. Figure adapted from (Shibue and Weinberg, 2017).

A recent study showed that DTPs exploit the conserved developmental program of the "embryonic diapause", defined as the paused state of embryonic development triggered by unfavorable environmental conditions including nutrient deprivation. The implication of the embryonic diapause has been demonstrated in in breast cancer (Dhimolea et al., 2021) and colorectal cancer (CRC) (Rehman et al., 2021). In response to cytotoxic drugs, cells undergoing diapause suppress MYC and mTOR pathways, slow the biosynthetic and metabolic activities, limit oxidative stress, and reduce the pro-apoptotic signalling, favoring the survival of these cells.

Besides the relevance in the acquisition of chemo-tolerant phenotypes the mechanisms here presented show how cancer cells can adapt and survive to changing circumstances by orchestrating complex changes in their metabolism and phenotypes. These changes are dynamic and transient and often occur through the out-of-context expression of pre-existing pathways. Such events can be explained by the acquisition of epigenetic and transcriptional plasticity.

In normal cells, gene expression is tightly regulated by the coordinate activity of gene-regulatory elements, such as promoters and enhancers, that are spatially and functionally engaged in topologically associating domains (TAD) (Dixon et al., 2016). TADs are independently regulated regions restricted by the binding of CTCF and cohesin at insulator regions, that avoid promiscuous co-regulation of nearby genes (**Figure 8**).

The activity of a locus depends on its accessibility to transcription factors (TF) and the transcriptional machinery. Therefore, active genes are placed in regions of "open" chromatin whereas genes not expressed are retained in inaccessible regions of highly compact chromatin. The chromatin status is defined by the deposition of histone marks: active regions are enriched in H3K27ac and H3K4me3 while, compact repressed regions are characterized

by H3K27me3, H3K9me3 and DNA methylation (Margueron and Reinberg, 2010). During development, the deposition of hereditable repressive chromatin marks by the Polycomb repressive complex 2 (PRC2) allows the establishment of cellular lineages (Comet et al., 2016). The commitment through specific lineages restricts the gene expression to lineage-specific genes, despite preserving a degree of plasticity responsible for homeostasis and physiological processes such as tissue regeneration (Rajagopal and Stanger, 2016) (**Figure 8**).

In cancer, the chromatin regulation can be compromised at every level. Mutations that impair the activity of epigenetic remodeling complexes (e.g. loss-of-function mutation in PRC2 components and SWI/SNF nucleosome remodeling complexes) (Hodges et al., 2016; Kim and Roberts, 2016), or mutations occurring in regulatory regions (e.g. impaired binding of CTCF at insulators due to mutations in its binding motif) (Hnisz et al., 2016) are among the most frequently occurring in cancer. The loosening of regulatory boundaries collectively contributes to the establishment of a more "permissive" chromatin status that allows cell-to-cell variabilities in gene expression and the interaction among genes normally not expressed from a specific cell type (Flavahan et al., 2017). The aberrant expression of genes and pathways in epigenetically and transcriptionally plastic cells might confer an adaptive advantage in specific contexts.
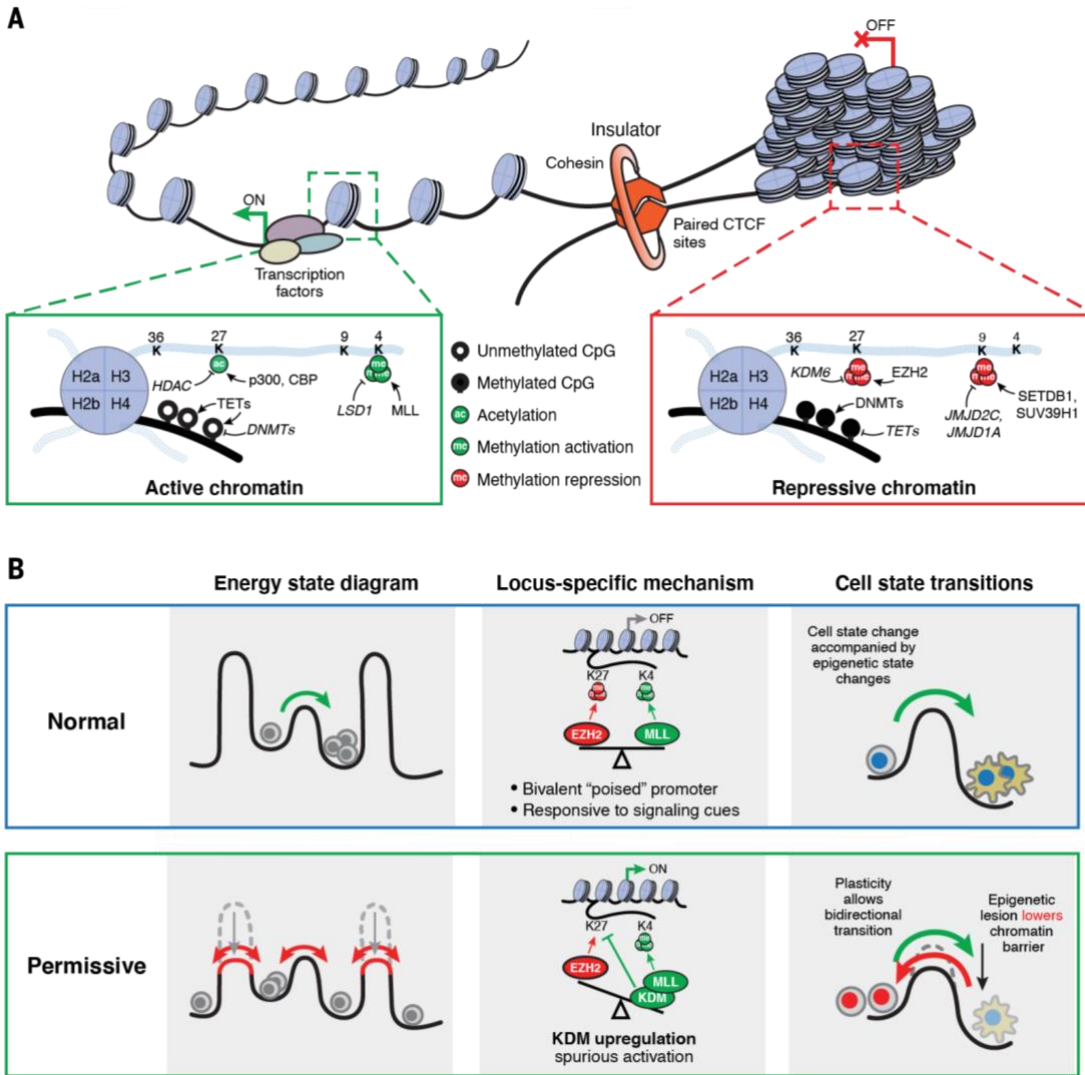
Figure 8: Epigenetic alterations favor the acquisition of epigenetic and transcriptional plasticity – A) Schematic representation of regions of open and closed chromatin showing main epigenetic regulators of chromatin. **B)** In normal cells, chromatin networks establish specific cell states. Plastic cells transition through different cellular states thanks to mutations impairing chromatin regulation. Figure adapted from (Flavahan et al., 2017).

Plasticity is not a uniquely cell-autonomous condition but different stimuli collaborate in establishing this status. The metabolic condition of cells influences the availability of intermediates necessary for the epigenetic regulation (Shyh-Chang et al., 2013). Microenvironmental clues collaborate in corrupting the epigenetic regulation. For instance, hypoxia suppresses DNA de-methylases thus promoting hypermethylation (and therefore, repression) of tumor suppressors in breast cancer (Thienpont et al., 2016).

Despite providing potential harm for patients, increasing evidence suggests that plasticity can be favorably exploited in therapies. For instance, in the context of breast cancer, it has been shown that inhibition of the pro-mitogenic signal through MEK inhibitor trametinib, favors the acquisition of a pEMT state in DTPs through the upregulation of different epigenetic modifying enzymes including BET proteins. The contextual targeting of BET proteins and MEK-pathway promotes cell-death (Risom et al., 2018).

1.2 Regulatory functions of lncRNAs

1.2.1 A non-coding perspective

The past decades have been years of huge changes in our understanding of RNA-based mechanisms (**Figure 9**). RNA has been detached from the "protein-centric" view of being solely responsible for shuttling the information from genes to proteins, with the discovery of catalytic functions of RNA in splicing, of regulatory RNAs such as Xist (Brockdorff et al., 1992; Brown et al., 1992) and of the RNA interference pathway (Fire et al., 1998).

The introduction of high-throughput technologies and the development of bioinformatic strategies allowed the comprehension that eucaryotic genomes are largely transcribed (Carninci et al., 2005; Guttman et al., 2009; Mercer et al., 2011) and in this multitude of transcripts, just a small proportion encodes for protein coding genes (PCGs) (1.5%) (Michelini et al., 2018). The transcriptome includes several non-coding RNAs with recognized features and a plethora of different functions; just to mention some: small nuclear RNAs (snRNAs) take part in the formation of the Spliceosome; small nucleolar RNAs (snoRNAs) guide the maturation of ribosomal RNAs (Peculis, 2000); microRNAs (miRNA) represent cytoplasmic regulators of mRNAs expression (Ghildiyal and Zamore, 2009). Discovering new functions of RNAs unveils intricate regulatory schemes in gene regulation in which non-coding RNAs are undiscussed players.
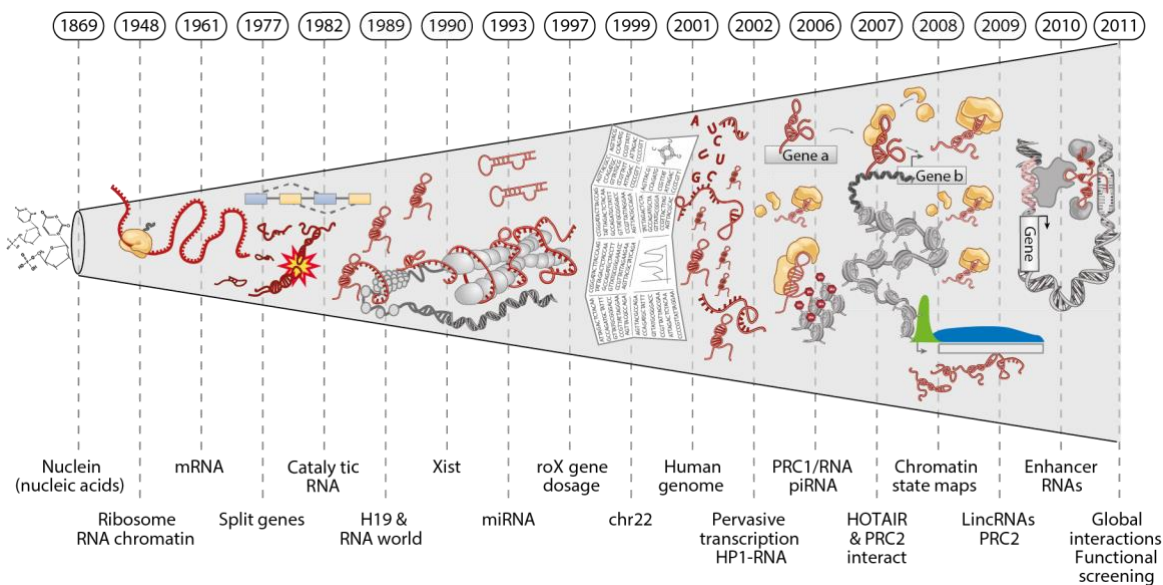
Figure 9: Breakthroughs in the discovery of RNA functions - Figure adapted from (Rinn and Chang, 2012).

1.2.2 Features of lncRNAs

Long non-coding RNAs (lncRNAs) are broadly defined as RNA molecules longer than 200 bp with no or limited coding potential (Kopp and Mendell, 2018); lncRNAs are ubiquitously transcribed from more than 50000 loci in the human genome (Hon et al., 2017; Iyer et al., 2015).

LncRNAs share some features with PCGs. At the epigenetic level, lncRNAs loci respond to the same histone code of PCGs, with peaks of H3K4me3 in promoter regions and activation by H3K27Ac (Guttman et al., 2011). Being transcribed by RNA polymerase II, lncRNAs most often are capped, spliced (despite less efficiently) and polyadenylated (Kopp and Mendell, 2018). LncRNAs are generally shorter than PCGs and exons in lncRNAs are usually less but longer compared to PCGs (Cabili et al., 2011; Derrien et al., 2012). Usually, lncRNAs are expressed at lower levels compared to PCGs (Kopp and Mendell, 2018), frequently in a cell-type specific fashion (Cabili et al., 2011; Hon et al., 2017). LncRNAs represent a class of newly evolved transcripts as they are poorly conserved outside mammals (Derrien et al., 2012; Djebali et al., 2012). Instead, lncRNAs promoters show the same conservation compared to PCGs and are potentially similarly regulated in the response of transcriptional programs of the cell (Carninci et al., 2005; Derrien et al., 2012).The majority of lncRNAs are retained in the nucleus (Werner and Ruthenburg, 2015), or otherwise exported in the cytoplasm or in specific subcellular compartments (Derrien et al., 2012; Gudenas and Wang, 2018), and exosomes (Li et al., 2018); their localization is closely linked to their function (see next paragraph).

Some lncRNAs fold into secondary structures important for their function, however they are not predicted to be more structured transcripts compared to PCGs and they can indeed play many structure-independent functions (Managadze et al., 2011).

The most common way to refer to lncRNAs is their location relative to PCGs (**Figure 10**). Thus, lncRNAs can be classified as "antisense" when they are transcribed from the opposite strand of a PCG, "intronic" when the sequence of the lncRNA is entirely embedded in an intron of a PCG, "divergent" when the transcription of the lncRNA is initiated by a shared promoter with a PCG, from the opposite strand. Lastly, "intergenic" or "intervening" lncRNAs (sometimes referred to as "lincRNAs") have independent promoters and do not overlap the sequence of a PCG (Rinn and Chang, 2012). However, this classification is obsolete and misleading as the RNApol2 transcription generates different classes of short-lived bidirectional RNAs, as well as lncRNAs can be transcribed from enhancers (Wu et al., 2017). Perhaps the most faithful definition of lncRNAs is the one based on their function.
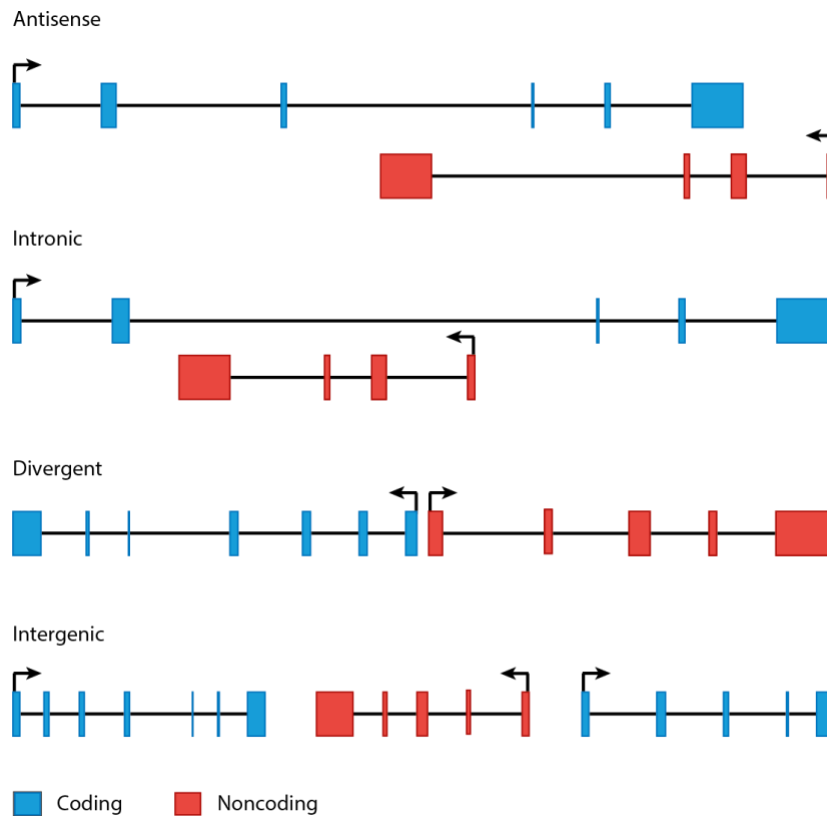
Figure 10: Organization of lncRNAs loci – LncRNAs can be classified taking into account their relative position to neighbor PCGs. Figure adapted from (Rinn and Chang, 2012).

1.2.3   Functions of lncRNAs

The information about the functional role of the vast majority of lncRNAs is still missing. Nevertheless, the knowledge about few hundreds of well-characterized lncRNAs pictures a heterogeneous group of transcripts with a plethora of functions and involves them in fine regulatory mechanisms of gene expression.

One of the challenges in understanding lncRNAs mechanisms is that multiple features can drive the lncRNA function. Indeed, the act of the transcription or RNA processing itself might influence the nearby genes; a lncRNA can interact with other transcripts or DNA through base-paring and cooperate with proteins through secondary folding. Alternatively, a DNA regulatory element might confer the functional properties to a specific locus and the transcription of the lncRNA can be irrelevant (Paralkar et al., 2016).

The activity of lncRNAs usually occurs in cis or in trans. The activity of a cis-acting lncRNA depends on the locus of their transcription and influences other elements in proximity, while trans-acting lncRNAs, resembling more closely mRNAs, regulate genes in loci distant from their site of transcription (Gil and Ulitsky, 2019). Executing their function in the vicinity of the locus of their transcription, cis-acting lncRNAs are usually expressed at very low levels, even few copies per cell (Derrien et al., 2012).

One of the most extensively studied lncRNAs is *Xist*. *Xist* is a 17 kb transcript, responsible for dosage compensation through X chromosome inactivation. Upon its allele-specific expression, *Xist* coats the X chromosome to be inactivated, and recruits the transcriptional repressors SPEN and PRC2 complex, leading to the deposition of H3K27Me3 and driving heterochromatin spreading and repression of the whole chromosome. In addition, *Xist* tethers the inactive X chromosome to the nuclear periphery via direct interaction with the nuclear membrane protein laminin B (Chen et al., 2016). In this scenario, *Xist* acts as "guide" for the repressive complexes through specific repeats in its sequence (Rinn and Chang, 2020). *Xist* can be considered a cis-acting lncRNAs (as it is usually defined), as its activity influences the expression of genes nearby its locus but also influences the activity of genes at Mb distance and importantly, the repressive activity of *Xist* is maintained when the lncRNAs is ectopically expressed from another locus (Engreitz et al., 2013), failing one of the definitions of cis-acting lncRNAs. This is just an example of how the functions of lncRNAs elude strict definitions and one transcript may have multiple properties difficult to discern.

LncRNAs are involved in different mechanisms depending also on their localization. In the nucleus, the activity of lncRNAs is mainly devoted to transcription regulation and organization of nuclear functionally distinct domains, while, in the cytosol, lncRNAs are engaged in various post-transcriptional regulatory mechanisms. Here are presented some of these mechanisms.

Transcription regulation

The transcriptional-regulatory activity of lncRNAs interest the activation as well as the repression of other units occurring in physiological conditions, development, and disease (Perry and Ulitsky, 2016; Schmitt and Chang, 2016).

One example is the involvement of lncRNA *Airn* in the genomic imprinting of the *Igf2r* locus in mouse Embryonic Stem Cells (mESC). *Airn* is transcribed antisense of *Igf2r* in the paternal allele. Its transcription produces transcriptional interference by steric hindrance that impairs the assembly of the RNApol2 complex at the *Igf2r* locus that results repressed. Moreover, *Airn* transcript recruits PRC2 complex and guides the repression of two distally imprinted loci (*Scl22a2* and *Slc22a3*) (Latos et al., 2012; Schertzer et al., 2019).

LncRNAs can coordinate multiple chromatin regulatory activities in one locus. One leading example is the activity of *HOTAIR* in the repression of HOX genes, crucial regulators of development. *HOTAIR* is transcribed from the HOXC locus, and it is responsible of the repression of the distal *HOXD* genes in trans. The silencing of *HOTAIR* by siRNAs is associated to decreased repressive chromatin mark H3K27me3 in the HOXD locus

corroborating its RNA-dependent function (Rinn et al., 2007). *HOTAIR* coordinates two repressive activities: the demethylation of the activating H3K4me3 histone mark and deposition of H3K27me3 by contextual interaction with PRC2 and KDM1A (histone demethylase), realizing the repression of the HOXD locus (**Figure 11**). Importantly, the exact mechanisms of PRC2 recruitment remains to be fully elucidated (Almeida et al., 2020). LncRNAs can act also as positive regulators of expression: for example, genes in the HOXA cluster are regulated by the lncRNA *HOTTIP*. HOTTIP locus is located 40 kb away from the HOXA cluster. *HOTTIP* interacts with MLL1 complex through an adapter protein and drives the specific deposition of H3K4me3 and *HOXA* transcription (Wang et al., 2011). Other lncRNAs influence gene expression by recruiting inhibitory proteins in favor of the expression of PCGs. This is the case of the lncRNA *PACER* that recruits the transcriptional repressor p50 avoiding the inhibition of COX2.
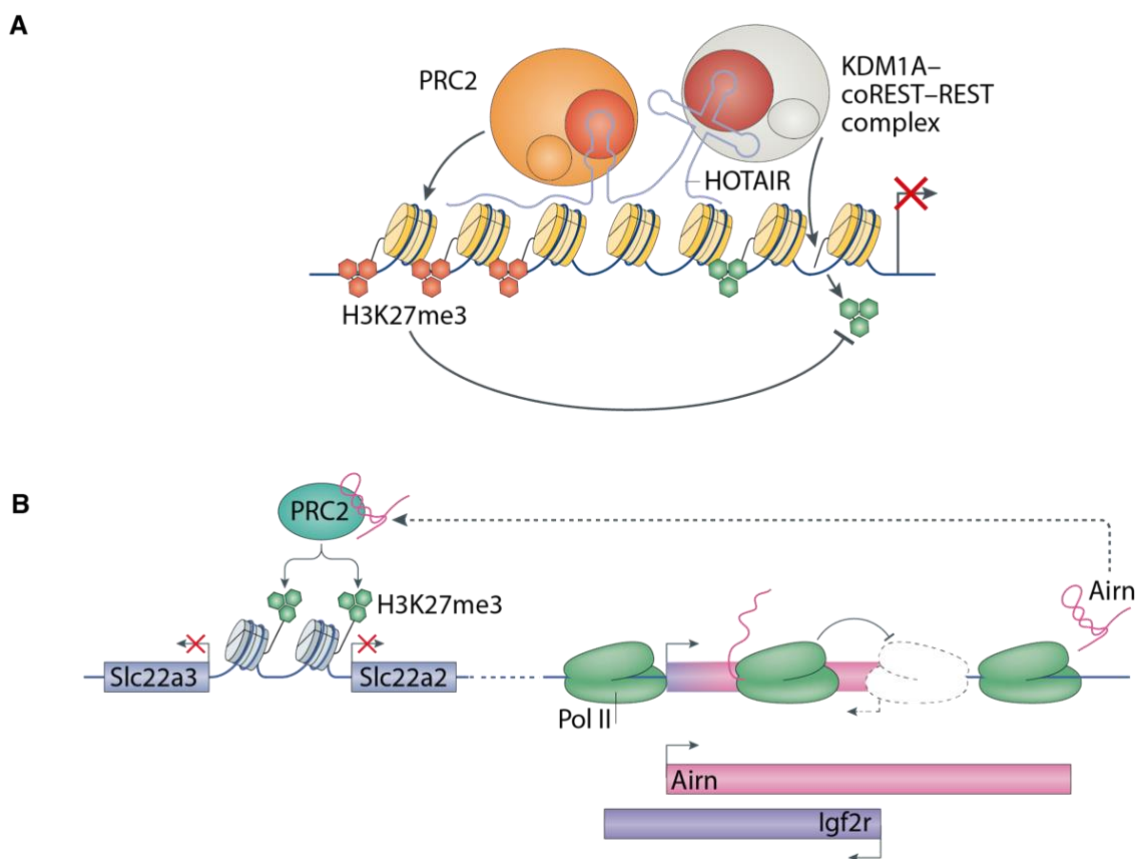
Figure 11: Transcriptional repression by lncRNAs – A) Coordinate activity of PRC2 (deposition of H3K27me3) and KDM1A-coREST-REST complex (removal of H3Kme3, activator histone mark) realizing the repression of the HOXD locus. Figure adapted from (Ransohoff et al., 2017). **B)** Contextual cis and trans repressive activity of *Airn*. *Airn* is responsible of transcriptional interference in cis towards the imprinted Igf2r locus and epigenetic repression in trans of Slc22a3 and Slc22a2 loci. Figure adapted from (Statello et al., 2021).

As mentioned before, the activity of a lncRNAs is led by multiple features. The presence of the transcriptional machinery and spliceosome influences nearby genes (Ali and Grote, 2020; Gil and Ulitsky, 2019). This is the case of the transcription of the lncRNA gene *Blustr*. The premature abrogation of *Blustr* transcription by insertion of a PolyA signal impairs the expression of the nearby gene *Sfmbt2,* associated with an increased deposition of the repressive chromatin mark H3K27me3 and a reduced level of the activating mark H3K4me3. The effect is independent of *Blustr* sequence (Engreitz et al., 2016) (**Figure 12**). This effect can be explained by lncRNAs participating in enhancer activities. Active enhancers are transcribed in short, bidirectional enhancerRNAs (eRNAs), as well as longer, unidirectional spliced transcripts, sometimes referred to elncRNAs, acting as potential mediators of enhancers long-range activity and interaction with chromatin (Kim et al., 2018b; Statello et al., 2021).
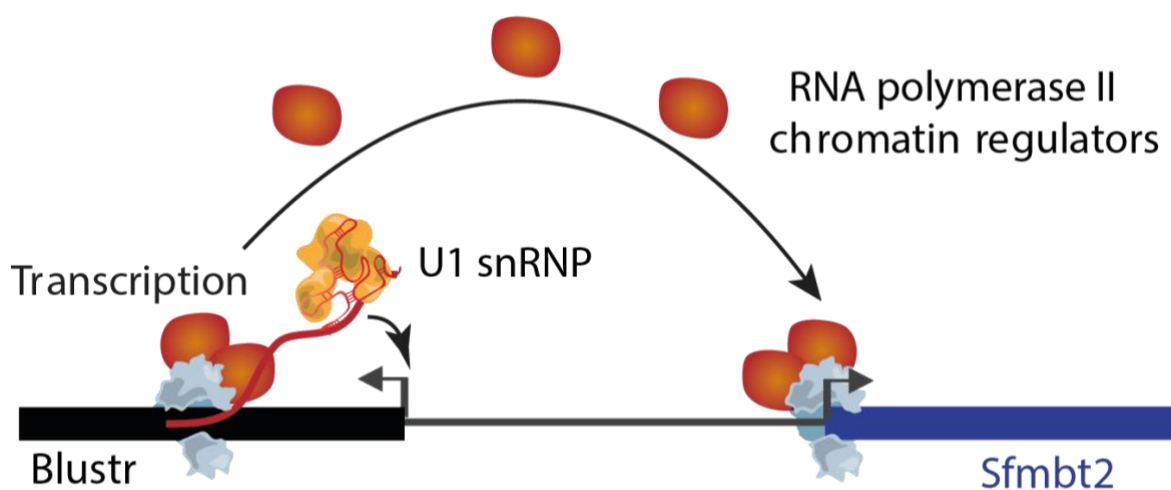
Figure 12: Cis-activating effect of *Blustr* transcription – Transcription and/or splicing of a lncRNA locus influence the activity of neighboring genes. In this example, the expression and processing of the lncRNA *Blustr* is associated with the epigenetic regulation and expression of the PCG *Sfmbt2*. Figure adapted from (Engreitz et al., 2016).

One example of such association of enhancer and lncRNAs is the regulation of the locus of *Hand2*, a TF important for the heart development, whose expression needs to be finely dosed to avoid heart malformations. Adjacent to *Hand2* locus there are two lncRNAs: *Upperhand (Uph),* transcribed bidirectionally from the same promoter of *Hand2* and *Handsdown (Hdn),* located 7.2 kb downstream. *Uph* sustains the expression of *Hand2* thanks to the presence of an enhancer element in its sequence. Alteration of *Uph* expression by insertion of premature polyadenylation signal impairs the epigenetic activation of the locus and *Hand2* expression, supporting the *Uph* transcriptional-dependent regulation of the gene. Nonetheless, the genetic ablation of *Uph* promoter abrogates *Uph* transcription but just mildly affects *Hand2*.

The downstream *Hdn* instead, competes with *Hand2* for the interaction with upstream regulatory elements. The activation of *Hdn* promoter leads to chromatin looping that place *Hdn* close to these elements, thus, excluding *Hand2* from their activity and reducing its expression (**Figure** *13*) (Anderson et al., 2016; Ritter et al., 2019). The effect of the removal of *Uph* promoter might argue against the involvement of the transcript in the regulatory functions mentioned. Other studies though, highlights the importance of the transcript itself in mediating enhancer-activities. This is the case of the lncRNA *A-ROD* and its down-stream spatially associated gene *DKK*. It has been demonstrated that impairing the dissociation of *A-ROD* from chromatin negatively impacts on *DKK* expression. The dissociation of lncRNAs from chromatin depends on the complete processing of lncRNAs through transcription and splicing (Ntini et al., 2018). Indeed, the activity of enhancers appears to be correlated to these two processes (Tan et al., 2020).

The precise involvement of transcription in enhancer activity is still a matter of discovery. However, the fact that lncRNAs loci are frequently overlapping enhancer regions, their cell-type specific expression and the examples just shown suggest lncRNAs to be close participant in enhancer regulatory functions.
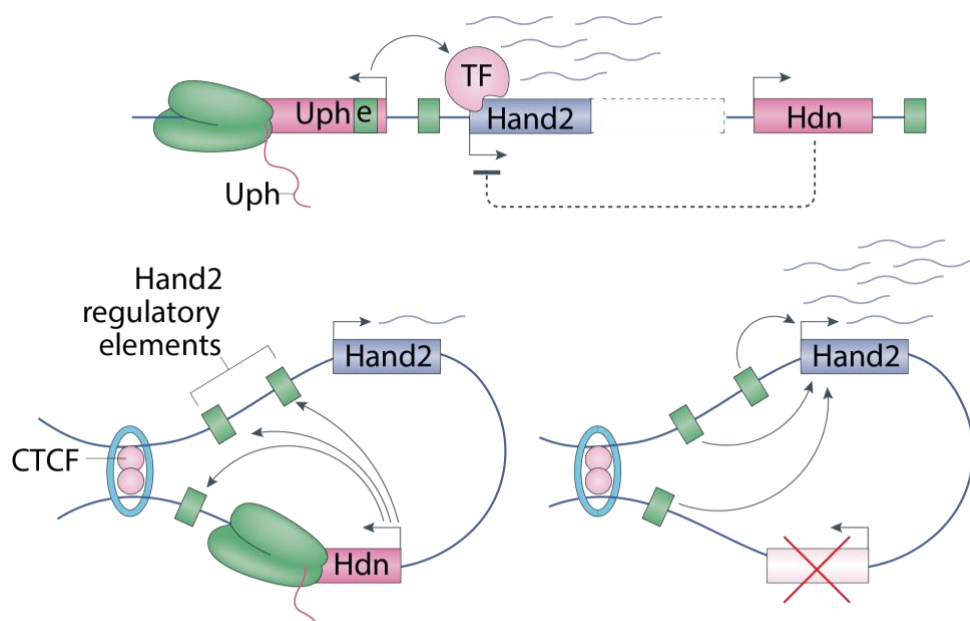
Figure 13: Multiple lncRNAs regulating the same locus – Different lncRNAs contribute to the expression of the *Hand2* transcription factor during mouse cardiac development. *Uph* transcription, from a bidirectional promoter with *Hand2* triggers the expression of the TF thanks to a DNA regulatory element in its sequence (e). Hdn instead, restricts the expression of Hand2 through promoter competition of Hand2 regulatory elements. Figure adapted from (Statello et al., 2021).

Organization of nuclear architecture

Some lncRNAs are abundantly expressed in the nucleus and coordinate the activity of functionally distinct nuclear compartments. This is the case of *NEAT1* and *MALAT1*. These two lncRNAs are transcribed from proximal loci but then localized in separate compartments. *MALAT1* localizes in speckles, where components of the spliceosome concentrate (Spector and Lamond, 2011). The suggested function for *MALAT1* in this compartment is allowing the positioning of the paraspeckles towards actively transcribed genes (West et al., 2014). *NEAT1* instead, is localized in paraspeckle, a dynamic compartment responsible of transcription and RNA processing. Like *MALAT1*, *NEAT1* associates with actively transcribed genes (West et al., 2014). While *MALAT1* is dispensable for speckle organization, *NEAT1* is essential for the assembly of paraspeckles. Thanks to different protein interacting domains, NEAT1 realizes the precise localization of proteins in the compartment (West et al., 2016).

Furthermore, in response to heat and chemical stress, the highly repetitive lncRNA *HSATII*, is transcribed and together with heat-shock transcription factor 1, drives the formation of nuclear stress bodies. These condensates retain transcription factor and splicing factors and promote stress-induced intron retention (Ninomiya et al., 2020).

Alternatively, the lncRNA *Firre* is expressed from the X chromosome and interacts with the nuclear membrane protein hnRNPU at the level of repeat regions in its sequence, establishing trans-chromosomal interactions that are abrogated upon *Firre* knock-down (KD) (Hacisuleyman et al., 2014).
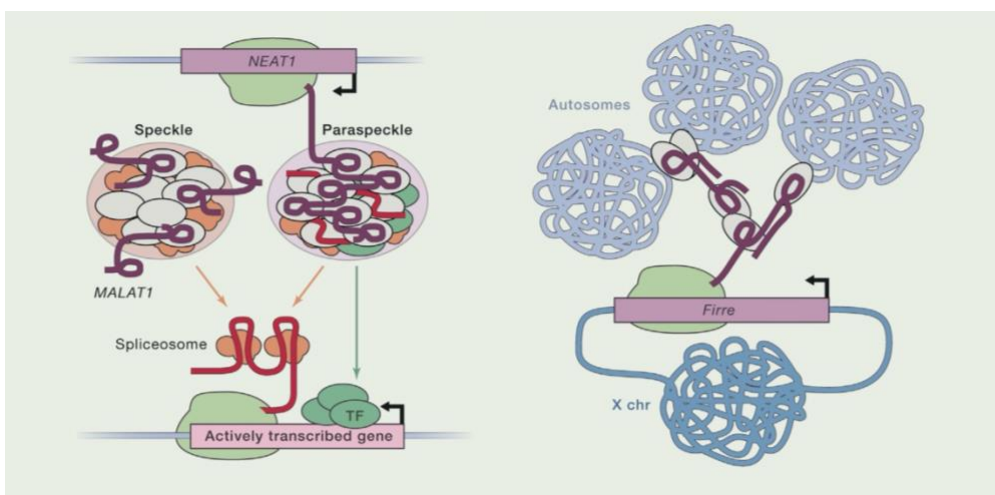
Figure 14: lncRNAs in sub-nuclear domains – LncRNAs *MALAT1* and *NEAT1* (left) contribute to the function and organization of speckles and paraspeckles, respectively. Nuclear domains responsible for RNA splicing and processing; Firre lncRNA establishes trans-chromosome interaction and the nuclear membrane. Figure adapted from (Kopp and Mendell, 2018).

Post-transcriptional regulation

On the other hand, the features of lncRNAs allow multiple post-transcriptional regulatory activities in the cytoplasm, that exploit the ability of lncRNAs of interacting both with other nucleic acids and proteins.

One well-characterized cytoplasmic lncRNA is *NORAD*. The sequence of *NORAD* contains multiple PUMILIO-responsive elements (PREs). PUMILIO proteins (PUM1 and PUM2) are RNA binding proteins (RBP) that triggers the degradation and translational inhibition of their targets. PUMILIO targets are involved in mitosis and genomic stability. In this context, *NORAD* functions as a molecular decoy for PUMILIO proteins, limiting their availability in the cytoplasm. Repression of *NORAD* releases PUMILIO proteins and triggers the degradation of their targets, leading to chromosomal instability and aneuploidy. For this reason, PUMILIO levels needs to be strictly regulated. NORAD is the primary target of PUMILIO and it is regulated by different stress stimuli such as DNA damage and hypoxia (Lee et al., 2016; Tichon et al., 2016) (**Figure 15A**).

Besides regulating PCGs, lncRNAs interact with miRNAs as competing endogenous RNAs (ceRNA). miRNAs bind precise miRNA Responsive Elements (MRE) in the 3' end of their targets, leading to mRNA degradation. Often, miRNA targets are functionally linked, so their activity orchestrates networks of post-transcriptional regulation. ceRNA regulate miRNA availability through multiple MRE embedded in their sequence. This is the case of linc-ROR. MiR-145 targets key factors involved in the maintenance of pluripotency such as SOX2, OCT4 and Nanog. Linc-ROR exerts its ceRNA activity by reducing the availability of miR-145, allowing complete expression miR-145 targets. Linc-ROR expression decreases during differentiation, allowing activation of miR-145, degradation of its targets and a shift towards more differentiated states (Wang et al., 2011)(**Figure 15B**).

Moreover, lncRNAs controls stability of mRNAs. Mammalian cells regulate the expression of mRNAs containing repeats through Stau-mediated mRNA decay. STAU1 protein recognizes the presence of double-strands RNA in the 3' end of transcripts and triggers mRNA degradation. During epidermal differentiation, the lncRNA TINCR binds specific regions in epidermal differentiation genes avoiding STAU1-mediated degradation and stabilizing these genes (Kretz et al., 2013)(**Figure 15C**).

LncRNAs activity may also impact on translation. The *AS-Uchl1* is a SINEB2 containing transcript, produced antisense to *Uchl2* and has sequence-complementarity with the 5' end of *Uchl2* mRNA. The pairing of the lncRNA and mRNA enhances the association with polysomes and *Uchl2* translation. Importantly, *AS-Uchl1* shuttles from the nucleus to cytoplasm in response to different stress stimuli such as the inhibition of mTORC nutrient

pathway (Carrieri et al., 2012). Moreover, the presence of functional translation-enhancing domains is a shared feature with other lncRNAs highlighting a common mechanism of regulation (Zucchelli et al., 2015) (**Figure 15D**).
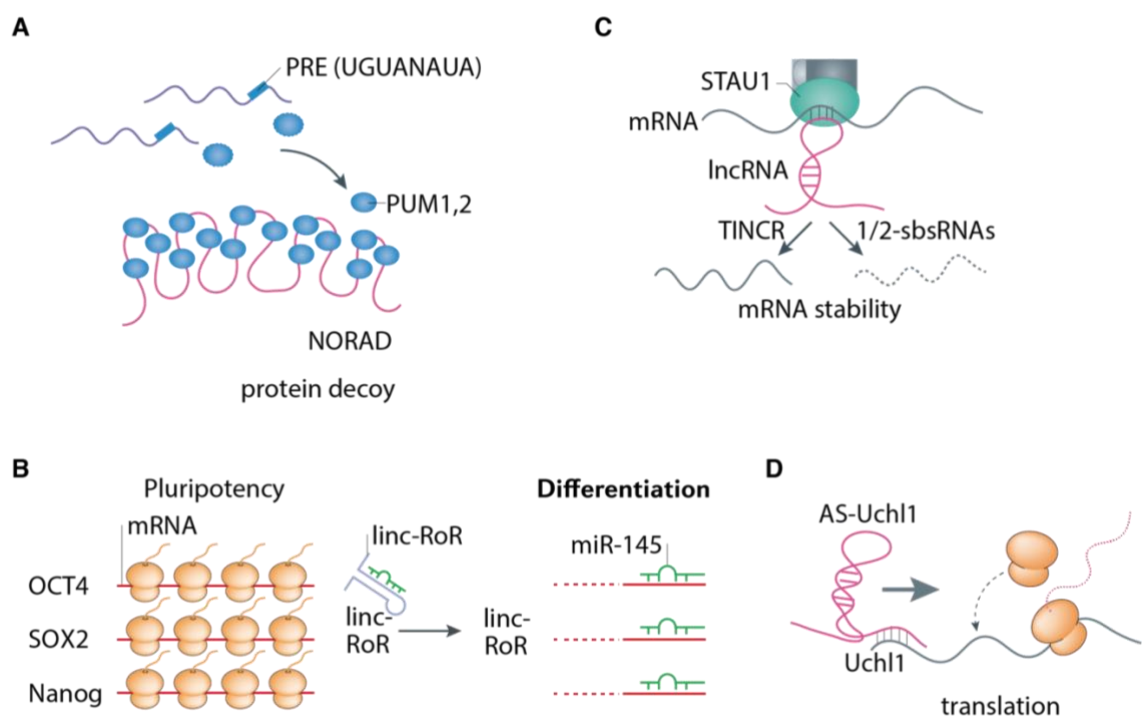
Figure 15: Post-transcriptional regulatory functions of lncRNAs – A) Model of NORAD decoy of PUMILIO proteins. NORAD has multiple PRE binding sites and regulates levels of available PUM1 and PUM2 proteins. Figure adapted from (Ransohoff et al., 2017). B) ceRNA activity on miR-145 of linc-ROR maintains a pluripotent state during differentiation. Figure adapted from (Ransohoff et al., 2017). **C)** TINCR lncRNAs stabilizes the mRNAthrough paring at the TINCR-box of epidermal-differentiation genes. Figure adapted from (Statello et al., 2021). **D)** Pro-translation activity of the lncRNA AS-Uchl1 through 5' paring with Uchl1. Figure adapted from (Statello et al., 2021).

1.3 LncRNAs in cancer

LncRNAs exert both oncogenic and tumor-suppressor activities. It is not surprising that cancer cells exploit the functions of lncRNAs to sustain tumor growth. One example is represented by the regulation of MYC locus. MYC is a master regulator of cell proliferation and one of the main oncogenes. MYC expression is regulated by the lncRNA *CCAT1-L* which is expressed from an enhancer region 500kb distant from MYC locus. CCAT1-L comes in contact with MYC promoter through chromatin loops between the two regions that enhance MYC expression. This looping also involves the promoter of another lncRNA called PVT1. PVT1 is frequently co-amplified with MYC in breast cancer. Indeed, PVT1 has oncogenic functions by stabilizing MYC protein in trans (Tseng et al., 2014). However, this is contrasting with the fact the PVT1 locus harbors structural rearrangements that abrogate

the expression of the lncRNA. Similarly to the regulation of the Hands2 locus (see **Transcription regulation**), PVT1 promoter competes with MYC for the binding of enhancer elements able to regulate MYC as well. Therefore, the loss of PVT1 has a pro-proliferative effect due to enhancer rewiring towards MYC (Cho et al., 2018). This example shows the contextual presence of enhancer-like functions of lncRNAs, the trans-activity of PVT1 transcript as well as the contribution of a DNA-regulatory element.

As MYC, also many other oncogenes and cancer pathways are regulated by intricated networks of coding and non-coding elements.
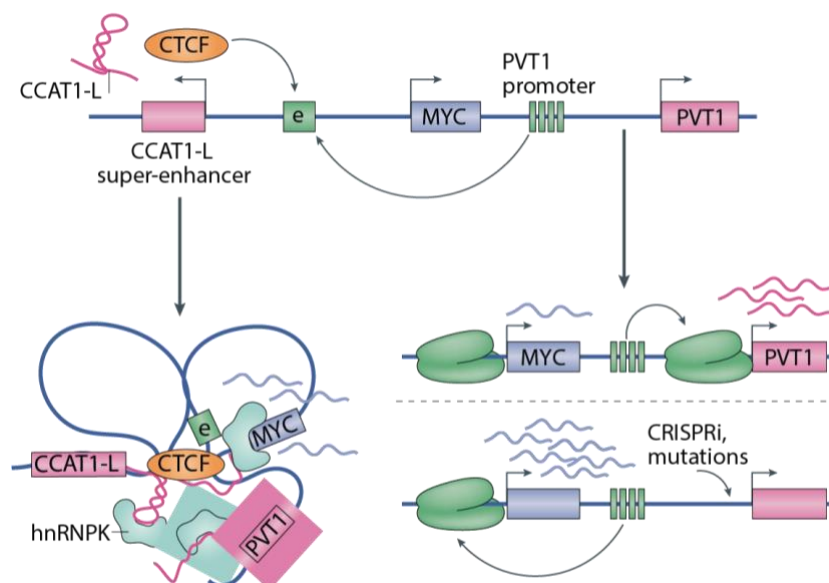
Figure 16: lncRNAs regulators of MYC expression – Regulation of the MYC locus is dependent on different non-coding activities. *CCAT1-L* promotes chromatin looping through the interaction with CTCF and hnRNPK. PVT1 competes with MYC for the activity of enhancers. Cancer cells suppress PVT1 promoter to favor the rewiring of these enhancers towards MYC. Figure adapted from (Statello et al., 2021).

Another example of a pro-tumorigenic lncRNA activity is the expression of the lncRNA *PNUTS*. The sequence of this transcript contains multiple MRE for miR-205, responsible for the silencing of ZEB1 and ZEB2, key regulators of the EMT. Upregulation of *PNUTS* sequesters miR-205 from the cytoplasm resulting in upregulation of these genes and their pro-invasive, pro-metastatic activity (Grelet et al., 2017).

LncRNAs expression can be triggered by specific environmental stimuli. For instance, *GAS5* is a lncRNA induced in cell-cycle arrested cells. *GAS5* interferes with glucocorticoid response, forcing the expression of pro-apoptotic signals. Breast cancer cell, undergoing nutrient deprivation, suppress the expression of *GAS5* to elude its anti-proliferative activity (Hudson et al., 2014).

Being ubiquitous regulators of gene expression, lncRNAs have been involved in the acquisition of all the hallmarks of cancer (Hanahan and Weinberg, 2011; Schmitt and Chang, 2016).

Besides being adjuvant in the establishment of cancer phenotypes, lncRNAs may represent biomarkers of disease states (Bhan et al., 2017) and a therapeutic opportunity. RNA is, by nature, easier to target and degrade compared to protein and the high tissue-specificity of their expression, makes lncRNAs suitable for targeted approaches (Statello et al., 2021).
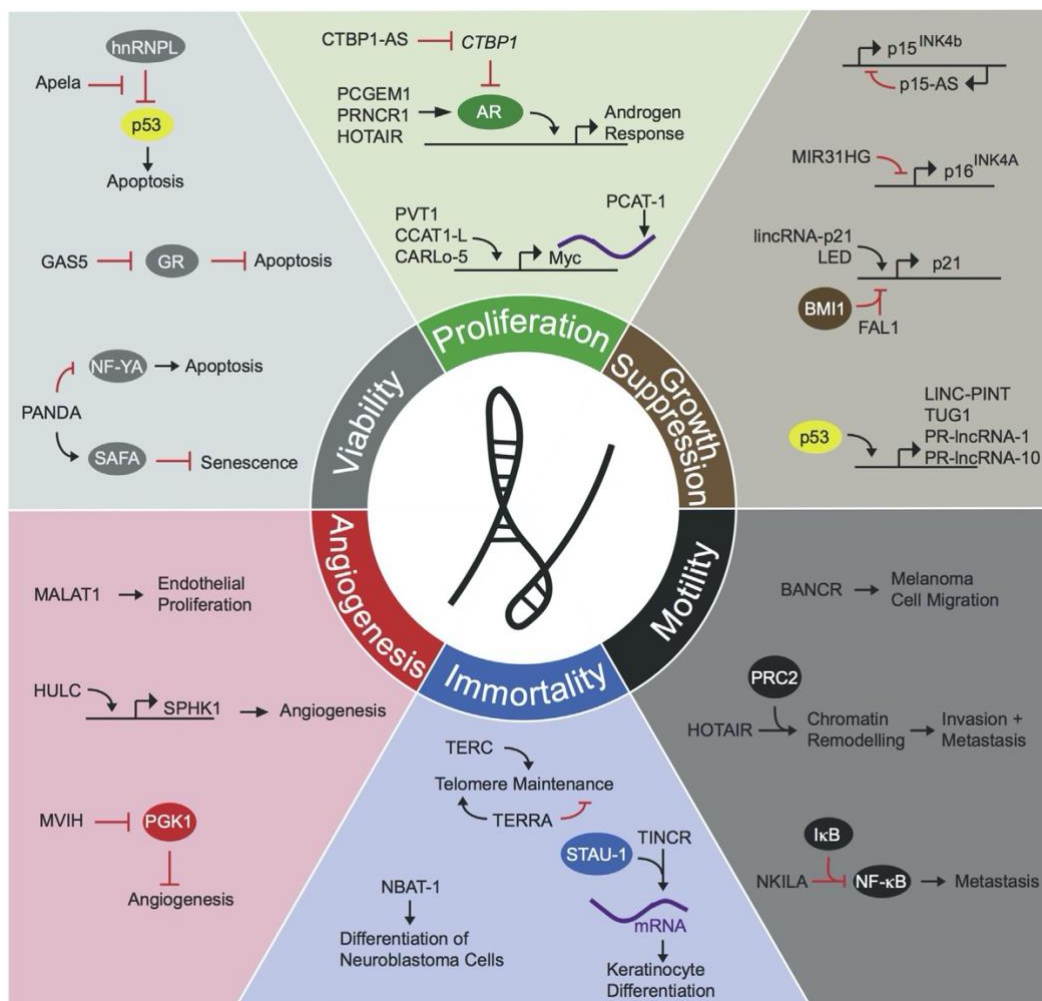
Figure 17: lncRNAs and the hallmarks of cancer – LncRNAs, participating in several regulatory pathways have been involved in cancer phenotypes through the acquisition of all the hallmarks of cancer. Figure adapted from (Schmitt and Chang, 2016).

1.4 Understanding lncRNAs function

LncRNAs have heterogenous functions, that depends on different features and localization of the lncRNA. Therefore, techniques to investigate lncRNAs functions must take into consideration this diversity. One first assessment of lncRNAs functions depends on evaluating their cis- and trans- activity. This can be done by perturbing the lncRNA and evaluating the effect on genes nearby. The different techniques to perturb lncRNAs expression are distinct depending on the feature targeted (Figure 18). In this context, CRISPR/Cas9 techniques find large use. The CRISPR/Cas9 is composed of an endonuclease (Cas9) and a single-guide RNA (sgRNA) composed of a constant region and a variable region. The RNA-protein complex can be driven to virtually any genomic locus thanks to the complementarity between the variable region in the sgRNA and a 20 bp region adjacent to a short 3-nucleotide motif called protospacer-adjacent motif (PAM). When correctly positioned, the Cas9 produces a double strand cut of the DNA, in proximity of the PAM (Dominguez et al., 2015; Sander and Joung, 2014).

In the lncRNAs investigation, the native CRISPR/Cas9 system can be used to achieve the ablation of the locus, producing a dual cut around the lncRNA sequence (usually across the first exon). The opposite ends of the cut are re-joined through the DNA repair systems. In this way, potentially all functions of the locus are abrogated. The CRISPR/Cas9 system has been also advanced by producing inactive versions of the protein Cas9 thanks to inactivating mutations in the catalytic site. This "dead-Cas9" (dCas9) retains the same binding specificities of the native Cas9 but binds the DNA without cutting. The protein has been functionalized through fusions with effector domains, producing the CRISPR*activation* (CRISPRa) and CRISPR*interference* (CRISPR*i*) systems. In CRISPR*activation,* the dCas9 carries to its target sites activator domains. Different versions have been proposed of the CRISPRa system, one of these is the VPR. In this system, the dCas9 is fused with a tripartite domain composed of VP64 (a synthetic tetramer of the herpes simplex VP16 transcriptional activator domain), p65 (subunit of NF-kB) and Rta (Epstein Barr virus transactivator protein) (Chavez et al., 2016). In the CRISPRi system instead, the dCas9 protein is fused with a Kruppel-associated box (KRAB) repressive domain. The CRISPRi system repress transcription by provoking steric hindrance to RNApol2 elongation and by the deposition of H3K9me3 chromatin repressive marks through the recruitment of KRAB-Associated Protein 1 (KAP1), Heterochromatin Protein 1-Alpha (HP1α) (Gilbert et al., 2013; Larson et al., 2013).

The sgRNAs target a narrow window close to the transcription start site (TSS) and realize site-specific control of gene expression. Both the activity of CRISPRi and CRISPRa interest

the activation/repression of the entire locus and thus, simultaneous modulation of DNA and RNA dependent functions.

The direct targeting of the RNA transcript can be done at different levels. For instance, with minimal alteration of the lncRNA, it is possible to induce premature transcription termination by inserting a transcriptional-terminating sequence, such as a polyA sequence or the insertion of a self-cleaving ribozyme, the latter having the advantage of being reversible. This approach preserves putative DNA regulatory elements included in the lncRNA sequence, while abrogating lncRNA transcription and processing.

Other approaches involve direct RNA targeting, such as the use of Antisense Oligonucleotides (ASOs), RNAinterference and the application of CRISPR-Cas13. Antisense Oligonucleotides (ASOs) are DNA oligos that pair with RNA through formation of an heteroduplex DNA:RNA which triggers the activity of RNase H. Conversely to RNAinterference-based approaches, ASOs realize an equivalently efficient knock-down of the RNA pool both in the nucleus and cytosol (Goff and Rinn, 2015; Liang et al., 2017). Recently, the tools for direct targeting of RNA expanded with a new tool in the CRISPR box. The Cas13 indeed targets directly the RNA, thanks to a complementary sequence of 60 bp to it guide RNA. Cas13-mediated repression is highly specific and targets the mature RNA. However, a complete set of design tools and reagents is still missing compared to other techniques (Gil and Ulitsky, 2019).
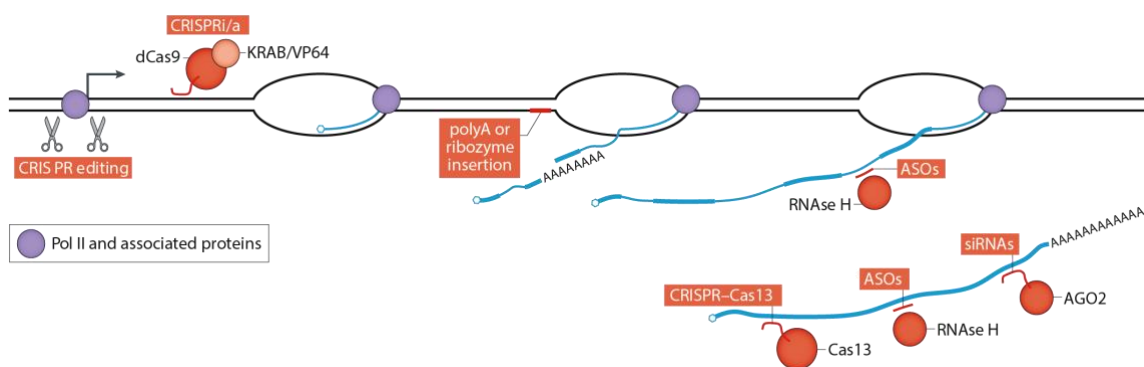
Figure 18: Tools for lncRNAs investigation – Different features are responsible for lncRNAs activities and different techniques are needed to define their specific function. CRISPR/Cas9 genome editing, by producing genetic ablation of a lncRNA locus impairs functions depending on DNA elements, transcription initiation, elongation, and the mature transcript. CRISPRi/a are CRISPR/based genomic tools for sequence-specific modulation of gene expression. The activity interest the locus, not only the transcript. Alternatively, the lncRNA locus can be modified through the insertion of a short polyA sequence or ribozyme. Both tools allow transcriptional initiation, namely the assembly of the transcriptional machinery but avoid transcriptional elongation. ASOs, siRNAs and Cas13 target the nascent RNA (mostly ASOs) and mature transcript. Figure adapted from (Gil and Ulitsky, 2019).
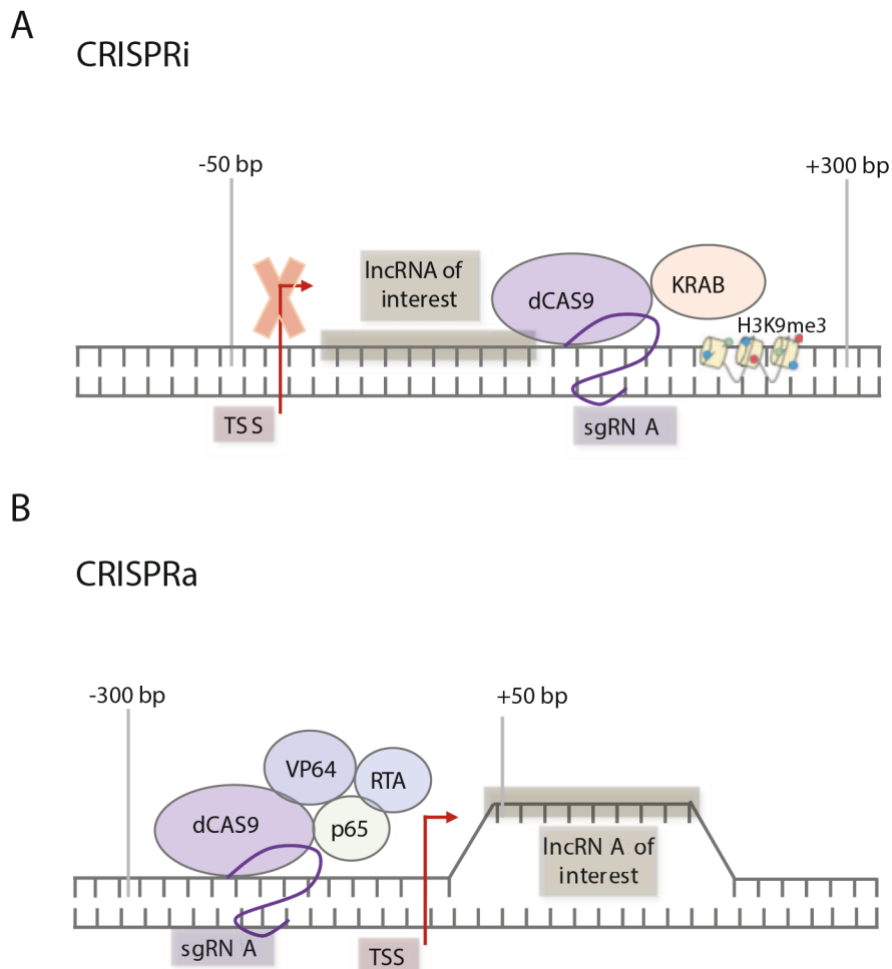
Figure 19: CRISPRi/a tools for modulating lncRNAs expression – The fusion of dCas9 with effector proteins allows the sequence-specific modulation of gene expression. These systems have the same requirements of native CRISPR/Cas9 system. The specificity depends on the presence of a PAM site and sequence complementarity of the sgRNA with the target region within a narrow window from the TSS. For CRISPRi this genomic window is located from -50 to +300 bp from the TSS; while, for CRISPRa is from -300 to +50 bp from the TSS. Figure adapted from (Morelli et al., 2021).

1.5 Screening-based approaches for the functional investigation of lncRNAs

Genetic screenings represent a forward-genetic unbiased approach to understand which genes contribute to a specific phenotype. Screenings allow to multiplex perturbation of multiple genes at the same time, up to genome-wide approaches. Usually, the perturbation is delivered to the cells in forms of lentiviral libraries containing a pool of constructs, such as sgRNAs in the case of CRISPR/Cas9 screenings, thus generating a population with multiple different genetic backgrounds. Lentiviral libraries have the advantage that can be calibrated to introduce just one perturbation per cell. Once the phenotype of interest has been observed, the sequence of the perturbation is traced back to identify genes influencing that

phenotype. The representation of sgRNAs is calculated by sequencing the integrated cassette before and after the screening. The statistical significance of the observations in the screening, depends on reaching the correct *library coverage*. Each construct must be carried by an adequate number of cells in the starting population (hundreds to thousands) so that it is possible to identify significant fluctuation in their representation and the same effect must be scored by multiple independent constructs targeting the same gene.

In a classic setting, cellular subpopulations carrying different constructs, are in competition with each other and at the end of the screening, some subpopulations will be over-represented (DROP-IN), meaning that the perturbation delivered to these cells has positive effects towards the selected phenotype. In the case of CRISPR/Cas9 screenings, the identification of DROP-INs points out putative negative regulators of the selected phenotype. Conversely, under-represented subpopulations (DROP-OUT) carry a perturbation that is detrimental towards the selected phenotype and the corresponding DROP-OUT genes should be considered as positive regulators of the phenotype. For instance, in the context of cell proliferation, cells carrying perturbations in tumor-suppressor genes are expected to proliferate more and these genes will be identified as DROP-INs. Instead, upon the perturbation of oncogenes, cells will proliferate less and the genes will behave as DROP-OUTs.

The perturbation of genes of interest can be delivered by RNAi, thanks to the delivery of lentiviral libraries of short hairpin RNAs (shRNAs) or by delivering the CRISPR/Cas9 system and pools of different sgRNAs. The Cas9 delivers the perturbation by cutting the target sequence that gets repaired introducing mutations that affect the function of PCGs or allow the degradation of the mRNA through non-sense mediated decay (NMD).

These approaches are functional in characterizing PCGs but quite limited in functional studies for lncRNAs. While the shRNAs offer the advantage of being selective for RNA-dependent phenotypes they are not effective in targeting nuclear-retained lncRNAs (a possible solution is to use ASOs, but in this case the approach is only suitable for arrayed screens (Ramilowski et al., 2020). Classic CRISPR/Cas9 systems are difficult to apply to lncRNAs because of the lack of functional targetable domains. One possibility to overcome this issue is by targeting the locus with two sgRNAs and delete larger fractions of the gene. This approach has proven useful in functionalizing lncRNAs that impar cell growth (Zhu et al., 2016). However, the effect of the deletion might preserve some transcription, responsible for the function of the locus. One alternative possibility is the targeting of lncRNAs promoters. However, this approach is dangerous and often unfeasible due to the targeting of embedded elements as well as affecting other genes sharing the same promoter.

One possible alternative is to apply the targeting of splice-sites in lncRNAs sequences. In this screening setting, the targeting of >10000 lncRNAs identified 2% of targets that impact cell growth (Liu et al., 2018b). This approach despite being quite specific, relies on the presence of PAMs around splicing sites and furthermore, the activity of the transcript might be not affected. Finally, the CRISPRi/a system offers a complementary approach to target lncRNAs, as they rely on the same specificity of CRISPR/Cas9 system and allows repression/activation of lncRNAs from their endogenous loci. These techniques are not devoid of problems since they can affect overlapping genes and highlight functions which are not dependent on the lncRNA transcription, but rather due to the activation/repression of other features (including DNA elements) embedded in their sequence. However, the targeting of lncRNAs by CRISPRi successfully identified genes involved in proliferation and allowed an unprecedent understanding of lncRNAs essentiality and cell-type specificity for a wide set of targets (>16000) in a large panel of different cells lines (Liu et al., 2017a). The possible phenotypic outputs of a screening are multiple. Beside the effect on proliferation, which is easy to carry out by passing cells in culture, other screenings have been effective in functionalizing lncRNAs in drug resistance (Bester et al., 2018; Esposito et al., 2019; Szlachta et al., 2018) and for their effect in vivo (Bossi et al., 2016; Han et al., 2020; Michels et al., 2020). Competitive screenings are more and more challenging at increasing level of selective pressure imposed by the phenotype. This is particularly true for *in vivo* settings. *In vivo* screenings have the advantage of recapitulating more faithfully the interactions within a specific organ, highlighting specific biological processes impossible to be properly captured *in vitro,* and allow the identification of therapeutic relevant targets. Usually, *in vivo* screening assessing the tumor initiating process are performed by transplanting a pool of human cells in immunocompromised animals. This process is usually highly selective and dramatically challenges the *library coverage* requirements.

Recently, screenings have found a new application by coupling specific perturbation with scRNA-seq technologies, allowing the dissection of the molecular mechanisms driving a specific phenotype. This technological advancement has been possible thanks to the introduction of techniques that allow the capture of sgRNAs in single-cell frameworks, such as CROP-seq (Datlinger et al., 2017). This approach make possible the definition of regulatory networks for protein coding genes as well as other regulatory elements, such as enhancers (Gasperini et al., 2019; Lopes et al., 2021).

Aim of the project

Non-genetic mechanisms of cancer adaptation allow cells to acquire the biological properties needed to survive in hostile conditions. The phenotypic adaptation often occurs in rare population of cells and depends on the acquisition of a plastic state in which cancer cells are able to exploit pre-existing transcriptional programs, pathways and developmental strategies to cope with the challenging environment. For instance, cancer cells exploit the epithelial to mesenchymal transition to shift between these two states and acquire CSCs properties and drug-tolerance. This dynamic shift requires a tight but flexible control of gene expression.

LncRNAs represent a large fraction of the mammalian transcriptome still not entirely characterized. Nonetheless, lncRNAs have been functionally implicated in mechanisms of chromatin, transcriptional and post-transcriptional regulation and they are generally considered as fine tuners of gene expression. Their role as oncogenes and tumor suppressors has been validated in different types of cancer.

In this project we aim at identifying lncRNAs that have a role in the acquisition of adaptive phenotypes in breast cancer; a malignancy characterized by high heterogeneity and inherent plasticity. We propose multiple screenings to test how the perturbation of lncRNAs impact on adaptive properties in the TNBC cell line SUM159PT.

Give the large number of lncRNAs, the application of genome-wide libraries is unfeasible when investigating phenotypes occurring in rare populations. For this purpose, we selected a list of lncRNAs according to their positive association in models displaying features of adaptation. Specifically, we evaluated the expression of adaptive response signatures in primary tumors, modeled the chemo-tolerance to the NACT drug Paclitaxel and evaluated the plasticity of the MaSC-like population of the HMLE cell line. Overall, from this analysis we selected 620 lncRNAs.

We used NET-Cage data produced in SUM159PT to refine the annotation of the TSS of candidates and designed a compact library of sgRNAs to test the function of these lncRNAs by means of multiple pooled CRISPRi screenings. The library also included 23 protein coding genes, mostly essential genes, and reporters of dCas9-KRAB activity as well as non-targeting controls for a library complexity of 3451 sgRNAs.

We tested how the perturbation of these genes modulate the adaptive properties of the cell line by testing the ability of the parental population to adapt its proliferation in 2D, 3D, in response to Paclitaxel and when injected in the mammary fat pad of NSG mice. In these conditions cells must adapt different strategies to survive.

We aim at identifying lncRNAs common and specific modulators of these phenotypes and ultimately provide the molecular characterization of their perturbation to find clues about their function and the mechanisms in which they are involved in.

2. Results

2.1 Transcriptional characterization of lncRNAs associated with breast cancer adaptative mechanisms

We defined a set of lncRNAs based on their transcriptional association with models of adaptive phenotypes in breast cancer that have been developed in the lab. Namely, we included as potentially relevant targets the lncRNAs that were highly expressed in: *i)* primary tumors characterized by high activity of "adaptive pathways" (**2.1.1**), *ii)* the CD44^{High}/CD24^{Low} "stem-like" population of the normal-mammary epithelial cell line HMLE (**2.1.2**) and *iii)* chemo-tolerant colonies of SUM159PT (**2.1.3**).

2.1.1   Adaptive responses in primary tumors

Cancer datasets such as The Cancer Genome Atlas (TCGA) are usually obtained using tumour biopsies that often containing a relevant proportion of immune and stromal cells, which particularly impact the analysis of low-expressed transcripts, such as lncRNAs, and on the analysis of pathways related to EMT and inflammation.

We evaluated the expression of lncRNAs in a panel of 26 breast primary cultures obtained in the lab of prof. Pier Paolo Di Fiore at IEO. The culture were established from 26 primary tumors from different histological subtypes and grading (**Figure 20**), which were was digested, dis-aggregated and cultured in adhesion for short-term (48h) (**Figure 2**). Non-epithelial contaminants such as fibroblast, endothelial and inflammatory components, were also removed to obtain a population of pure-epithelial breast cancer cells that we characterized by strand-specific RNA-sequencing. This dataset is extremely relevant for lncRNA expression analysis in breast cancer as it is kept pure in the epithelial composition and, at the same time, close to original tumour context (short term culture).

Histological subtype	Grading			Total
	G1	G2	G3	
Luminal A	2	2	/	4
Luminal B	/	5	4	9
HER2/Basal	/	1	4	5
HER2/Luminal	/	1	4	5
TNBC	/	1	3	4

Figure 20: Subtypes and Grading of the 27 breast primary tumors employed in this study

Tumor samples were stratified according to the expression of signatures of "adaptive response pathways" that have been associated with transcriptional plasticity, namely the acquisition of an EMT/partial EMT signature (Kim et al., 2018a; Kröger et al., 2019), hypoxic response (Rohwer and Cramer, 2011), stemness(Shibue and Weinberg, 2017), inflammation (Bhola et al., 2013) and Unfolded Protein Response (UPR) (Wang and Kaufman, 2016). For each signature, we calculated a scoring system (based on the sum of the Z-scores of the genes in the signature, see Methods for details), that was used to stratify tumor samples into three groups with "high", "mid", "low" signature score, respectively. Two sets of tumors coordinately displayed a consistent "high" or "low" activation score for most of the adaptive signatures (**Figure 21**). Of note, even if we noticed an association of "high" tumors with G3 grade and TNBC/basal classification and "low" with G1/G2 grade and Luminal, the separation was not perfect (**Figure 21**), reflecting either heterogeneity of patients or the effect of culturing in 2D or both.

We found a total of 1348 coding genes and 285 lncRNAs (padj <0.05 and log2FC >0.5). Using Gene Set Enrichment Analysis (GSEA), we confirmed a strong enrichment for all the hallmark pathways of adaptation, as expected (EMT, Hypoxia, Inflammation, Stemness; data not shown). In addition, we found strong enrichment for genes involved in invasive breast cancer, multi-cancer invasion and extra-cellular matrix remodeling, but not for genes related to tumor cell growth (**Figure 22**). This finding is consistent with our initial hypothesis to enrich for features of high tumor aggressiveness, which are mostly not related to proliferation capacity in basal-like breast cancer. Following a "guilty-by-association" approach, we imputed that the set of 285 co-associated lncRNAs might be involved in the same pathways.
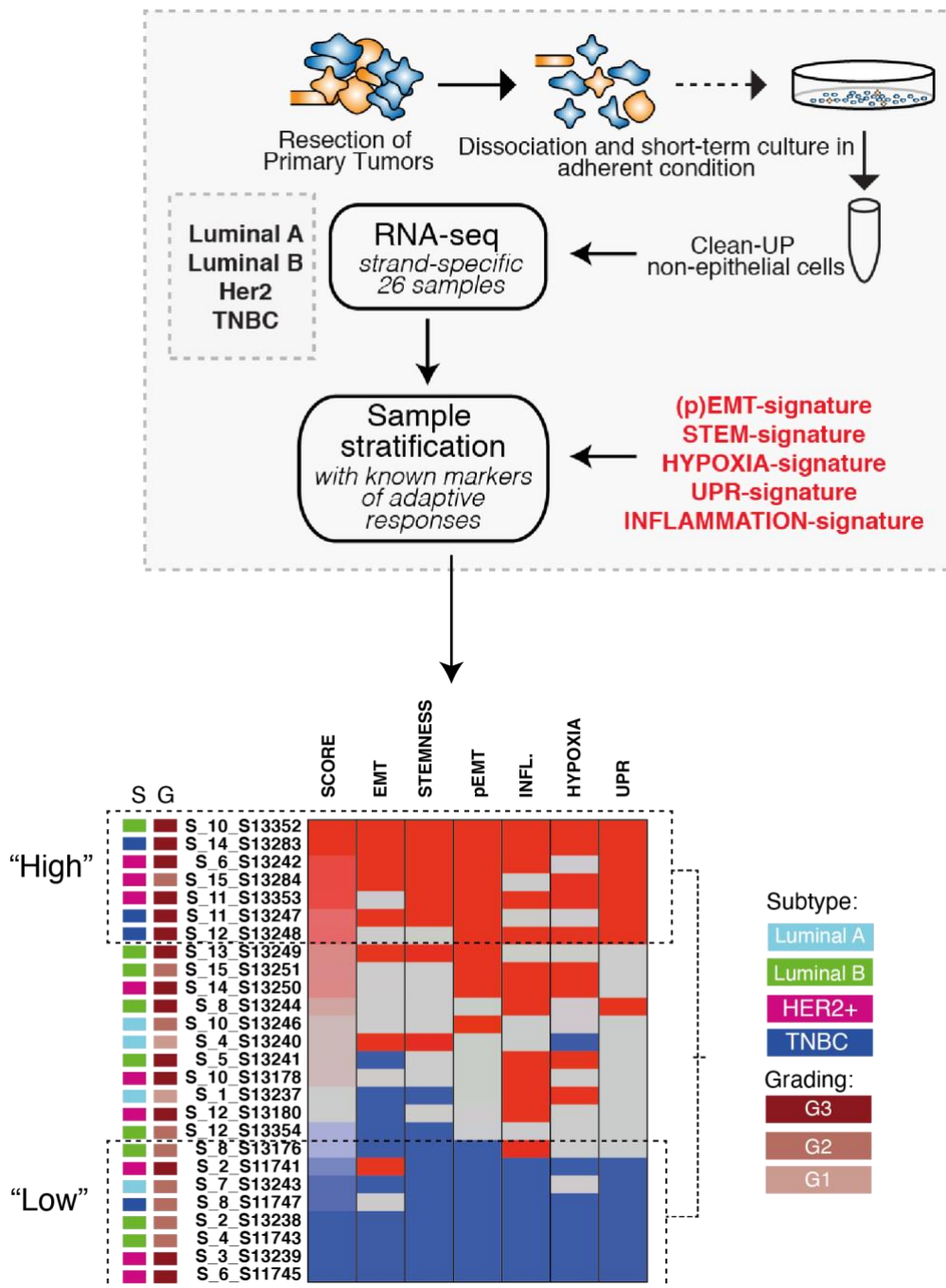
Figure 21: Classification of primary tumors according to Adaptive Response signatures – Outline of the experimental procedure for tumors short-term culture and cancer cells purification. Below representation of activation scores for all the adaptive signatures considered. Tumor subtype and grading are indicated. Data produced by Chiara Tordonato (Di Fiore Lab, Istituto Europeo di Oncologia, IEO) and Matteo Marzi (Nicassio Lab, Istituto Italiano di Tecnologia, IIT).
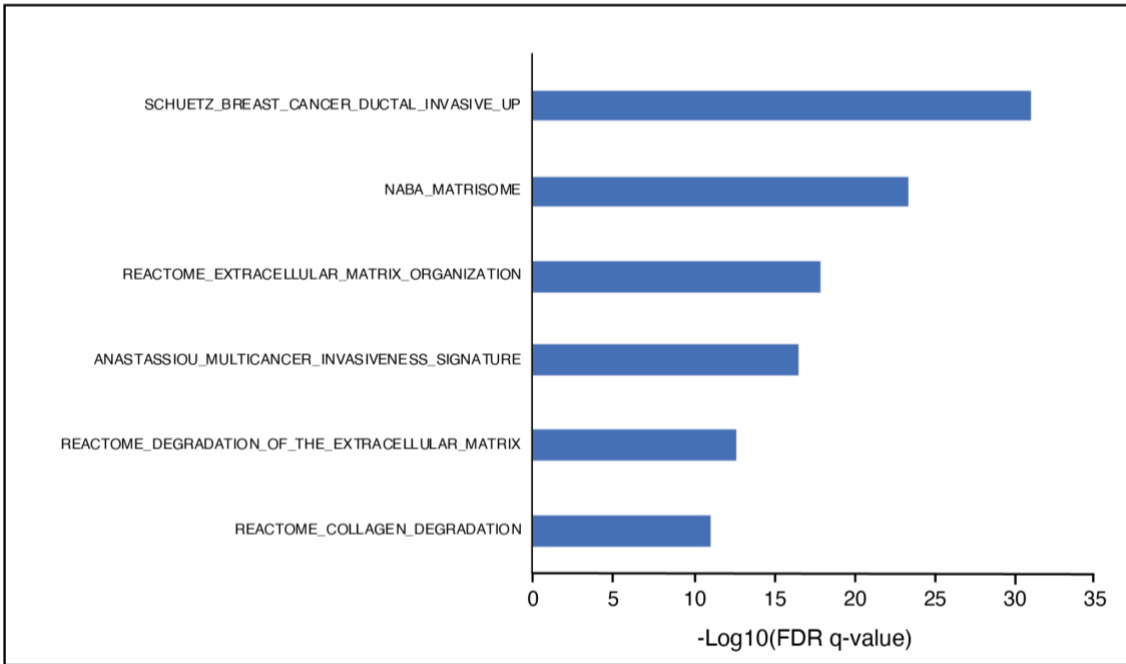
Figure 22: Enriched signatures in "High" vs "Low" tumors – Selected enriched sets by GSEA analysis of top 500 PCGs DEGs in "High" vs "Low" tumors.

2.1.2 Mammary Stem Cell (MaSC) -like population in HMLE

The HMLE is a breast normal epithelial cell line characterized by features that recall the heterogeneity of the mammary gland. The vast majority of HMLE cells (>90%) is immunophenotypically characterized by CD24^{high}/CD44^{low} profile and resembles luminal progenitors and differentiated cells, as they express epithelial markers (such as cytokeratins, E-cadherin and EpCAM,) and are unable to grow in non-adherent conditions. A small fraction of the bulk population (0,4 – 2%) is CD24^{low}/CD44^{high} and resembles mammary stem cells (MaSC-like cells), as they express high levels of mesenchymal markers (N-cadherin, Vimentin) and can survive in non-adherent conditions forming spheroids called as "mammospheres" (Al-Hajj et al., 2003) (**Figure *23***). The molecular and phenotypic properties of this sub-population have been widely investigated in the past and often correlated to aggressive properties of tumours, as the proportion of cells within this sub-population increases during oncogenic transformation and correlates with in vivo tumorigenesis, highlighting a parallelism between the normal and the cancer counterpart of stem cells (CSCs) (Chaffer et al., 2011).

With the aim of characterizing transcripts (coding and lncRNAs) associated to MaSC-like features, we isolated by cell-sorting the MaSC-like CD24^{low}/CD44^{high} population as opposed to the more differentiated luminal-like CD24^{high}/CD44^{low} population and performed strand-specific RNA-sequencing from three independent biological experiments. Using DESeq2, we found a total of 341 coding genes and 82 lncRNAs significantly enriched in the

"CD24^{low}/CD44^{high}" group (padj <0.05). We observed a good overlap of the CD44^{high} signature with the Adaptive Response signature described above (overlap of 114/341, 33.4%, expected overlap 11.8%, pval<0.05 by Fisher's test). Coherently, we also observed enrichment for signatures related to invasive breast cancer and extra-cellular matrix remodeling (**Figure *24***). Among the hallmark pathways, EMT was the most prominently enriched (p=6.8e-40), as expected by previous literature (Mani et al., 2008). In conclusion, the "CD44^{high}" signature captures similar transcriptional features of the "Adaptive" one, but also provides a different perspective and several additional lncRNA candidates.
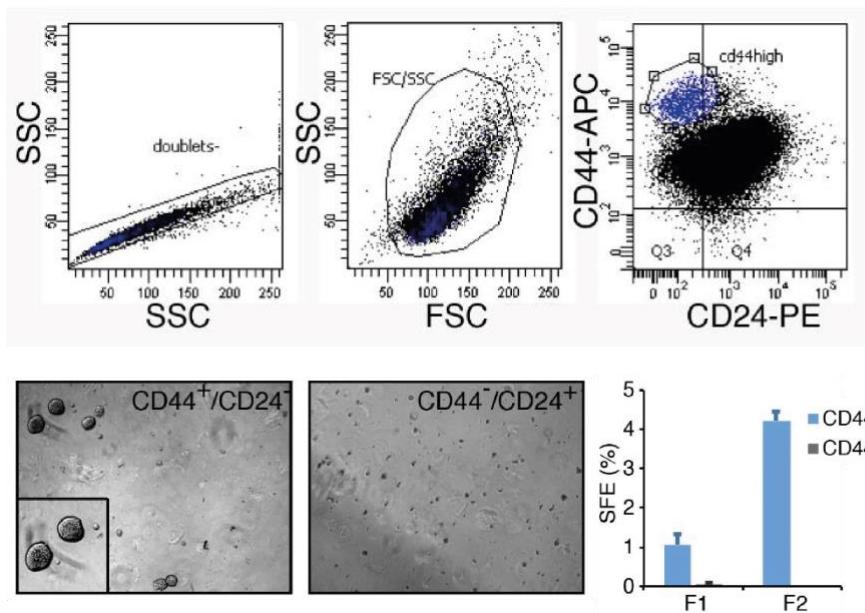
Figure 23: Differential properties of CD24^{low}/CD44^{high} vs CD44^{low}/CD24^{high} subpopulations of HMLE cell line – Dot plots of flow-cytometric analysis of the CD24 vs CD44 profile of the HMLE cell line. CD24^{low}/CD44^{high} represents a smaller fraction of the bulk population. When sorted, only this fraction of cells forms countable mammospheres 7 days after seeding in a 3D matrix. These cells maintain sphere-forming capability after the first generation of mammosphere. Data by Chiara Tordonato (Di Fiore Lab, Istituto Europeo di Oncologia, IEO)
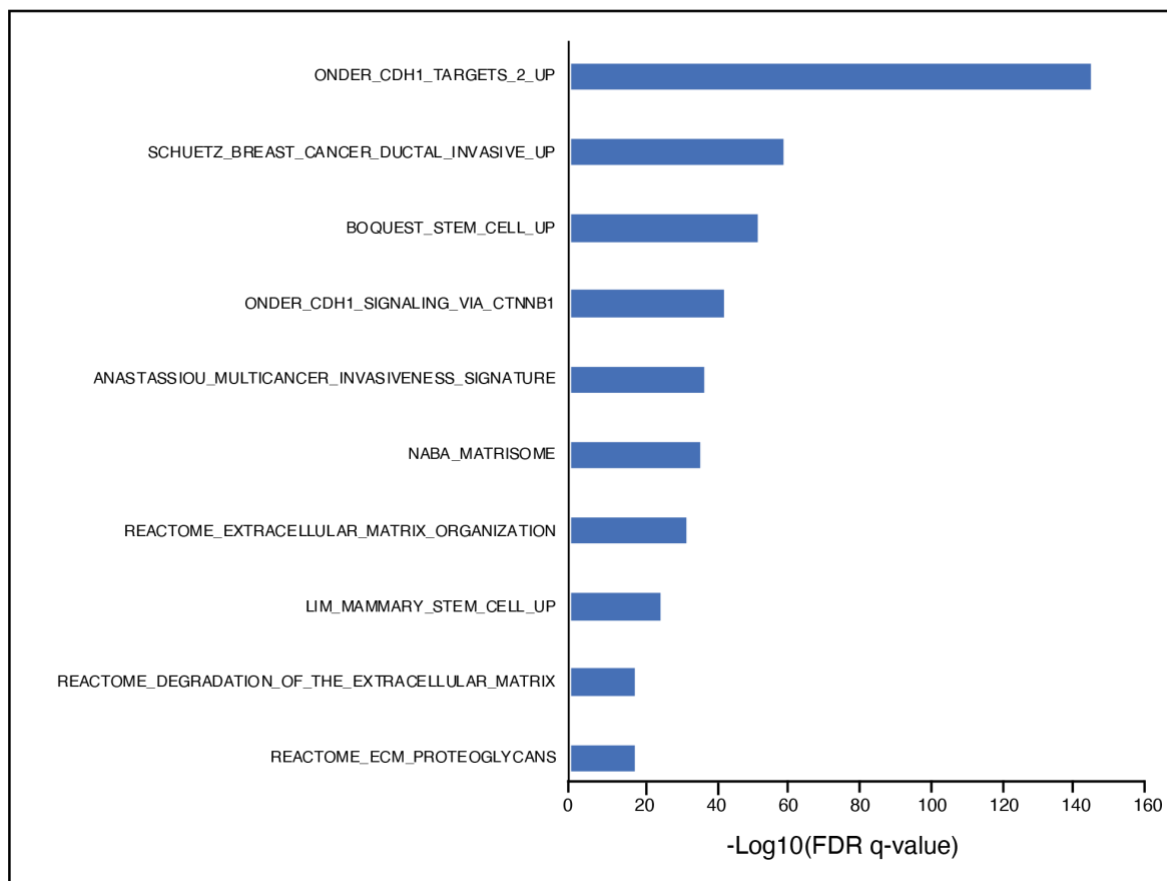
Figure 24: Enriched signatures of DEGS in CD24^{low}/CD44^{high} vs CD44^{low}/CD24^{high} - Selected enriched sets by GSEA analysis of 341 PCGs in The CD24^{low}/CD44^{hi} (stem-like) vs CD24^{high}/CD44^{low} (differentiated epithelial cells).

2.1.3 Model of chemo-adaptation

The persistence of cancer cells after treatments represents a major challenge in the therapy of breast cancer, especially for women suffering of TNBC. To study the transcriptional and epigenetic mechanisms associated to drug tolerance, we generated a model of chemo-adaptation by exposing the cell line SUM159PT, a highly aggressive claudin-low TNBC cell line, to Paclitaxel, a G2M-blocker commonly used in Neoadjuvant Chemotherapy regimens (NACT). Specifically, we treated cells with 50nM Paclitaxel for 72h, a dose that kill most cells but allows the emergence of rare drug-tolerant colonies. Upon treatment, cells stop proliferating and die due to mitotic catastrophe in a try to escape from the G2-M block induced by the drug. Only a small fraction of the cells (3%) survives and after few days reinstate proliferation growing into clonal colonies (**Figure 25A**). The chemo-adapted population emerges typically at 10-14 days from the treatment, and it shows biological properties typical of aggressive cancer cells, with an increased expression of markers and features of stemness (increased ALDH+ cells and increased frequency of cells capable of mammosphere formation, **Figure 25C**). Of note, the drug-tolerant state of SUM159PT is

44

transient and totally reversible. After 21-25 days from the treatment, cells spontaneously revert to their original state, losing all the properties (molecular and phenotypic) which were acquired along the adaptation. Such behavior is highly reproducible in SUM159PT and suggests that in this model system the acquisition of drug tolerance is mostly dependent on transcriptional/epigenetic reprogramming, a contention which is in line with recent findings obtained by others while investigating drug-adaptation mechanisms in TNBC (Echeverria et al., 2019; Kim et al., 2018a).

With the aim of characterizing transcripts (coding and lncRNAs) associated to the transcriptional/epigenetic reprogramming of drug tolerant cells, we compared the expression profiles of SUM159PT drug-tolerant colonies (at D12) with their parental counterpart (D0). To overcome the intrinsic experimental variability, we analysed 8 replicates per sample type (16 samples in total) and, then, identified 459 protein coding genes (>1 FPKM) and 152 lncRNAs (>0.1 FPKM) in the "Pacli-resistance" signature (padj < 0.05 and log2FC >0.5). Also in this case, the most enriched hallmark pathways were linked to adaptive pathways such as inflammation, EMT, Hypoxia, plus KRAS signalling and apoptosis. Furthermore, we found an overlap with basal breast tumours, invasive ductal carcinoma, cancer stem cells and invasiveness among cancer signatures (**Figure 26**). We also found a significant overlap with the other two sets (overlap with "Adaptive"=104/459, 22.7%; expected 1348/11357=11.8%) (observed overlap with "CD44"=50/459, 10.9%, expected 341/11357=3.0%).
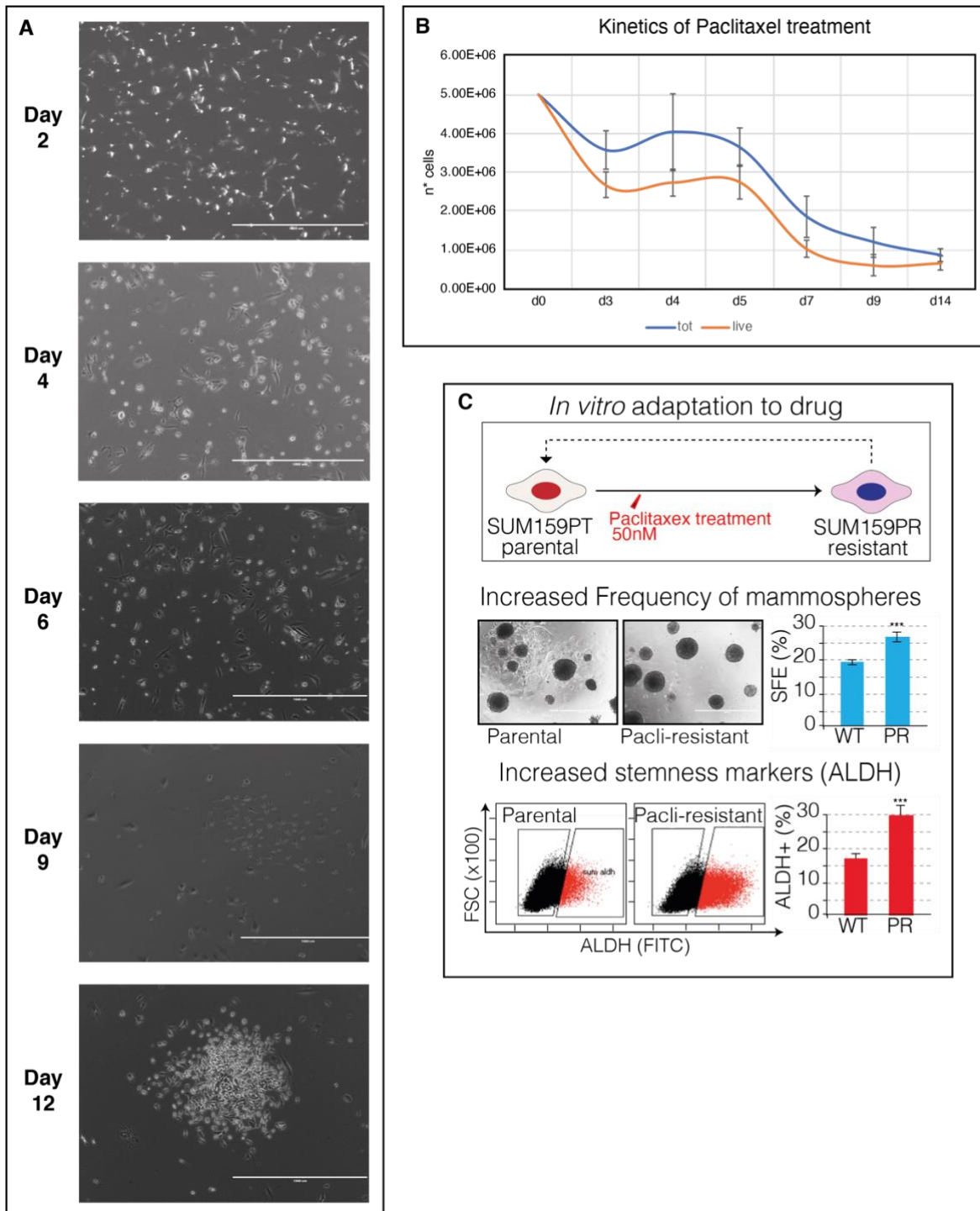
Figure 25: Emergence of Paclitaxel-tolerant SUM159PT – A) Time-course showing the extensive cell-death after the treatment with 50 nM of Paclitaxel and growth of colonies. Pictures acquired by EVOS, scale bar= 1000 μm. **B)** Plot showing the numbers of total and live cells during the paclitaxel response. **C)** Features of Paclitaxel-tolerant SUM159PT. Tolerant cells show boosted CSCs properties as measured by increased numbers of cells positive for the de-toxifying protein ALDH and increased SFE. Error bars represents SEM (n=3), statistical significance by Student's t-test. Data by Paola Bonetti and Maria Pirra Piscazzi (Nicassio Lab, IIT).
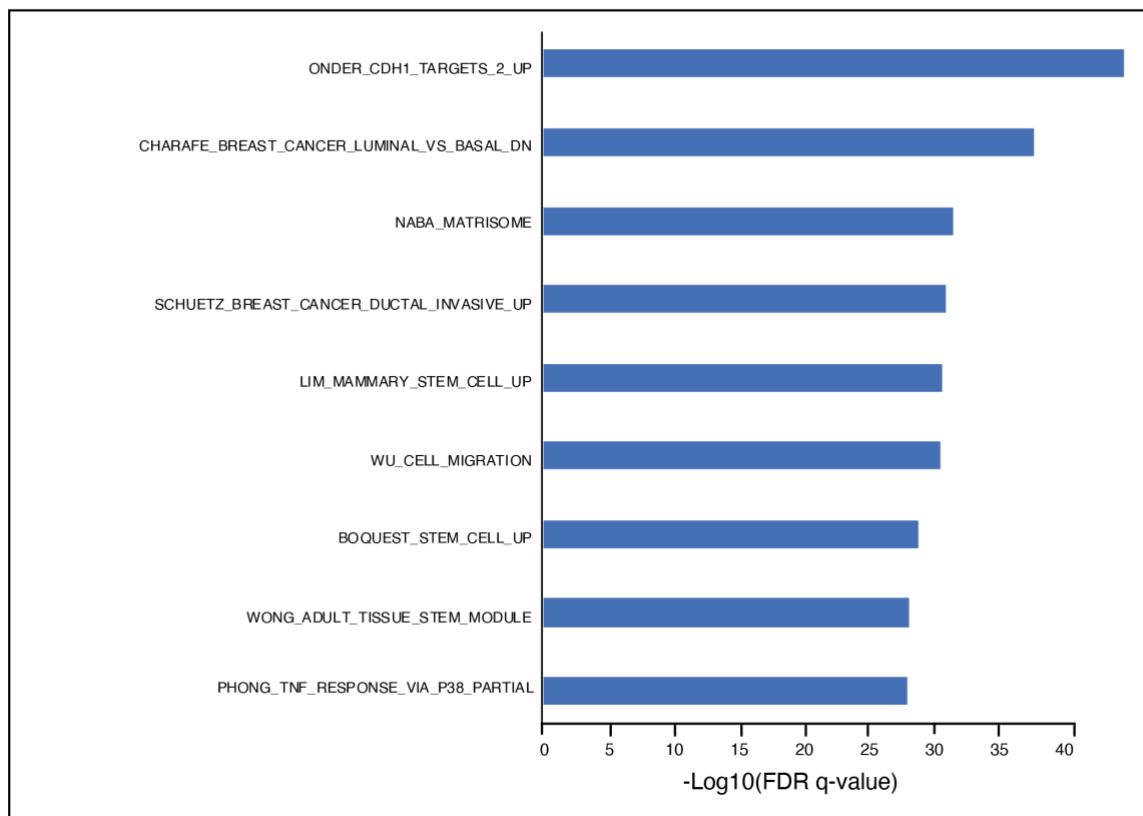
Figure 26: Enriched signatures of DEGS in Paclitaxel-tolerant vs Parental SUM159PT - Selected enriched sets by GSEA analysis of 459 PCGs upregulated in Paclitaxel-tolerant colonies of SUM159PT.

2.1.4 Selection of lncRNAs candidates

All the different models we used are capable of capturing transcriptional features of breast cancer adaptation starting from different perspectives: the intrinsic plasticity of MaSC-like cells (HMLE model), the transcriptional adaptation occurring upon acute drug treatment and the emergence of chemo-tolerant cells (SUM159PT Paclitaxel model) and the activity of adaptive-response signatures in primary tumors. We therefore combined the lists of candidate lncRNAs to obtain a final set to be subjected to functional analysis. Furthermore, we exploited scRNA-seq profiles generated in the lab from both i) CD24^{low}/CD44^{high} HMLE cells and ii) Paclitaxel tolerant SUM159PT (data not shown); to provide additional lncRNAs candidate associated to specific subpopulations. Finally, we complemented our dataset by selecting lncRNAs with literature-supported role in drug-resistance, EMT, survival and metastasis in breast cancer or multi-tumor as annotated in Lnc2Cancer (Ning et al., 2016) or by the analytical framework "Longhorn" (Chiu et al., 2018).

In all the above-mentioned analyses, we focused only on those candidates positively associated to phenotypes (DEG-UPs), whose ablation by dCAS9-KRAB was expected to negatively impact on mechanisms of cancer adaptation. Additionally, to maximize feasibility

in the screenings, we filtered out those candidates that could not be tested in the cellular model of the screenings (i.e. lncRNAs which are not detected in SUM159PT). Below is reported the final composition of the lncRNA candidates, while the complete list of lncRNAs candidates is provided in **Addendum**.
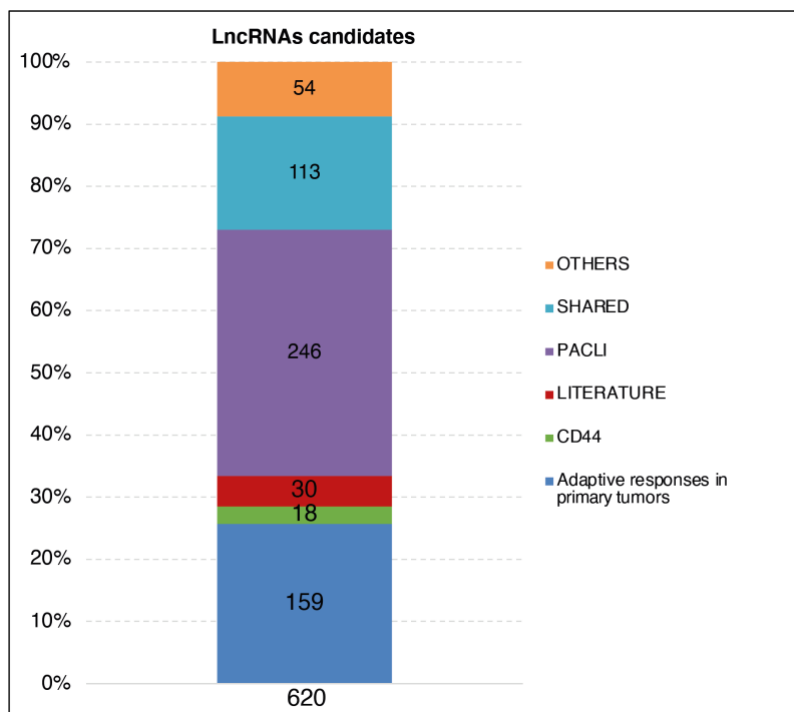
Figure 27: LncRNAs candidates – Bar chart showing the 620 candidates lncRNAs with indication of the models of their selection. The larger fractions of lncRNAs candidates were chosen from the analysis of adaptive response signatures in primary tumors (25%) and adaptation to Paclitaxel (39%). Only 3% of candidates lncRNAs was specific of CD24[low]/CD44[high], while 8.7% of the selected genes was common to more than one model ("shared"); the group "others" collects lncRNAs selected for their single-cell expression; 4.8% lncRNAs have a literature-supported role in cancer.

2.2 A CRISPRinterference platform

2.2.1   SUM159PT: a versatile TNBC model system

The first step to perform a screening is the selection of an appropriate model system, recapitulating the biological properties under investigation. Breast cancer is a widely studied malignancy and multiple cell lines have been stabilized during the years, representative of the many sub-types and heterogeneity of this tumor (Dai et al., 2017; Neve et al., 2006). Our lab has acquired long-standing experience in breast cancer cellular models and in particular of TNBC cell lines, among which SUM159PT represents a versatile and useful model for cancer studies. This cell line was initially isolated from a primary anaplastic breast tumor and belongs to the claudin-low subtype, a particularly aggressive form of Triple Negative Breast Cancer (TNBC) (Ethier et al., 1993; Flanagan et al., 1999). These cells indeed, lack the expression of Estrogen Receptor (ER), Progesterone Receptor (PR) and HER2 receptor (Bianchini et al., 2016). Genetically, the main mutations borne by SUM159PT interest HRAS, PIK3CA, TP53 and MYC amplification, mutations common in breast cancer, which sustain both the indefinite growth and the aggressive phenotype of these cells (Saunus et al., 2018). SUM159PT cells, as evidence of their epithelial origin, express low levels of cytokeratin 18 but they retain no expression of the epithelial cell-junction protein E-cadherin. In fact, they show high expression of the intermediate filament protein vimentin, displaying a mesenchymal-like phenotype (**Figure 28A**). Moreover, SUM159PT cells are highly tumorigenic *in vivo*. When injected in mammary fat pads of immunodeficient mice they form primary tumours with high efficiency and also produce distant metastasis (Flanagan et al., 1999). SUM159PT cells display also an intrinsic heterogeneity, being composed of different subpopulations described by their surface markers: a $CD44^{low}/CD24^{high}$ population, expressing genes of luminal differentiation and a large fraction (~90%) of $CD44^{high}/CD24^{low}$ population, expressing markers of basal lineage (Fillmore and Kuperwasser, 2008). The latter population appears as enriched in Cancer Stem Cells (CSCs), as it is able to form tumours *in vivo*, re-create all the different subpopulations when isolated by FACS-sorting and it shows a higher degree of resistance to chemotherapeutics (Fillmore and Kuperwasser, 2008; Gupta et al., 2011).

A frequently used *in vitro* assay to monitor the number of CSCs in cancer cellular population (i.e. a cancer cell line) is the mammosphere assay. (Dontu et al., 2003; Tordonato et al., 2021). In such assay, cells are plated non-adherently in methylcellulose matrices in absence of serum. In these conditions, only cells with stem-like features exploit their self-renewal ability and grow in clonal spheroids (**Figure 28B**). Therefore, the Sphere-Forming Efficiency (SFE), defined as the ratio between the number of spheroids obtained and the

number of seeded cells, is considered a proxy of the number of CSCs in the original population. When calculating the SFE of various breast cancer cell lines, we observed that SUM159PT cells showed the highest SFE (about 20%), maintained in multiple generations of mammosphere growth (Bonetti et al., 2019; Tordonato et al., 2021). These and other features, such as easy growing conditions and fast doubling time (21h, see **Figure 28**C), together with the intrinsic heterogeneity and plasticity, made this cell line a largely studied model in our lab and a suitable candidate for our CRISPR*interference* screening to probe lncRNA function in breast cancer adaptation phenotypes. Furthermore, our lab performed several molecular characterizations of SUM159PT exploiting multiple high resolution-omic approaches, including strand specific total RNA sequencing, small RNA sequencing, single-cell RNA sequencing, single-cell Assay for Transposase-Accessible Chromatin (ATAC) sequencing, native elongating transcript–cap analysis of gene expression (NET-CAGE) (Hirabayashi et al., 2019) and Nanopore direct RNA sequencing. The sum of these molecular characterizations provided an extremely helpful resource to map the transcriptional and epigenetic blueprint of these cells and to precisely characterize the type of lncRNAs and their expression regulation in this system.
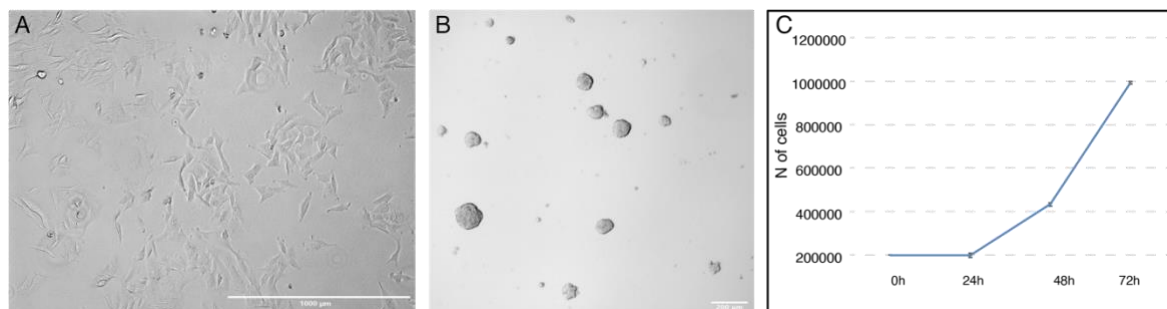
Figure 28: Properties of SUM159PT- A) Mesenchymal morphology of SUM159PT in 2D adherent culture. Pictures acquired by EVOS. Scale bar 1000 µm. **B)** First generation of SUM159PT spheroids in 3D non-adherent methylcellulose culture. Pictures acquired by Nikon SMZ25 stereomicroscope. Scalebar 200 µm. **C)** Growth curve of SUM159PT. Line shows average N of cells in 3 technical replicates. Error bar=SD. (Growth curve by Carmela Rubolino, Nicassio Lab, IIT).

2.2.2   Delivery of dCas9-KRAB

The completion of the adaptation screenings requires the silencing of the expression of candidates lncRNAs robustly and continuously throughout the assay using a reliable CRISPRi system. We sought an inducible system in which the expression of the dCas9-KRAB can be turned ON and OFF, as it offers the advantages of *i*) obtaining an acute repression of the targeted genes limited to the experimental setting; *ii*) avoiding unwanted or uncontrolled phenotypic effects when handling the cellular population before the

screening, and *iii*) decreasing chances of silencing and counterselection of the transgene during culture. To this end, we tested both a lentiviral and a transposable system. In the lentiviral system (pHage TRE dCas9-KRAB, Addgene #50917, (Kearns et al., 2014)), the expression of the dCas9-KRAB fusion protein occurs upon the Doxycycline-dependent binding of the transactivator protein (rtTA) to the upstream TRE promoter. In the transposable system, a PiggyBac transposon contains the dCas9-KRAB fusion protein, similarly regulated by a TRE promoter, and Hygromycin resistance gene, surrounded by Inverted Terminal Repeats (ITRs). (**Figure 30**, see **Methods**).
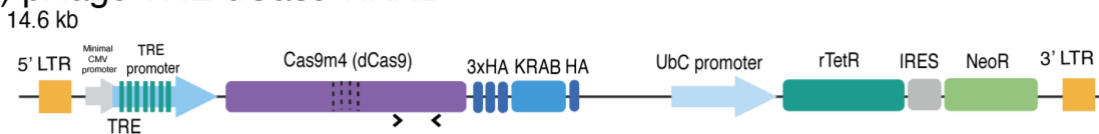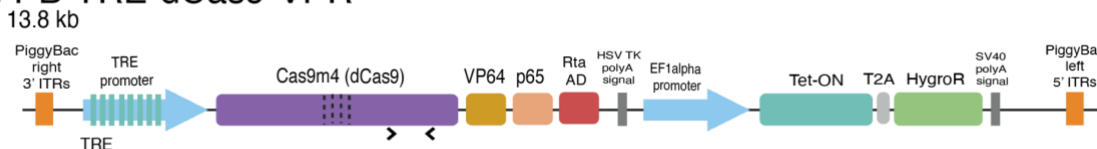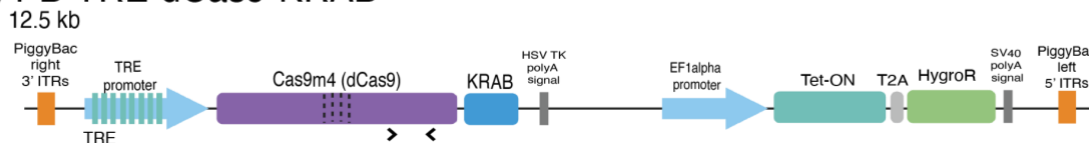
Figure 29: Schematic of constructs for dCas9 delivery - Dotted lines represents mutations inactivating the endonucleolytic sites of the Cas9 protein. Black arrows represent the location of qPCR primers for transgene detection.

The CRISPRi/a systems were tested by measuring the level of expression and inducibility of the transgenes (dCas9-VPR or dCas9-KRAB) and the effects on representative target genes. In case of the CRISPRa system, dCas9-VPR was highly induced upon Doxycycline treatment, producing a robust upregulation of CXCR4, targeted with a combination of validated sgRNAs (Chavez et al., 2016) (**Figure 30**)**.** In case of the CRISPRi, the two systems were compared by measuring the transgene expression and the effects on the expression of the lncRNA PVT1, using previously validated sgRNAs (Cho et al., 2018). As expected from literature (Chavez et al., 2016; Chen et al., 2015), the dCas9-KRAB was expressed by both systems, resulting in the repression of the target PVT1. However, the transposable system reached both higher levels of expression of the transgene and stronger repression of the target as compared to the lentiviral system, where we measured a limited repression of the target gene (**Figure 30**). Hence, the PB-TRE-dCas9-KRAB CRISPRi system was selected and several stocks of SUM159PT were generated to perform the

lncRNA screenings (hereafter named SUM-dCas9-KRAB). It is worth mentioning that a degree of leakiness was observed, as dCas9-KRAB was expressed (at low level) in absence of doxycycline treatment resulting in a partial repression of the target in absence of induction (**Figure 30**).
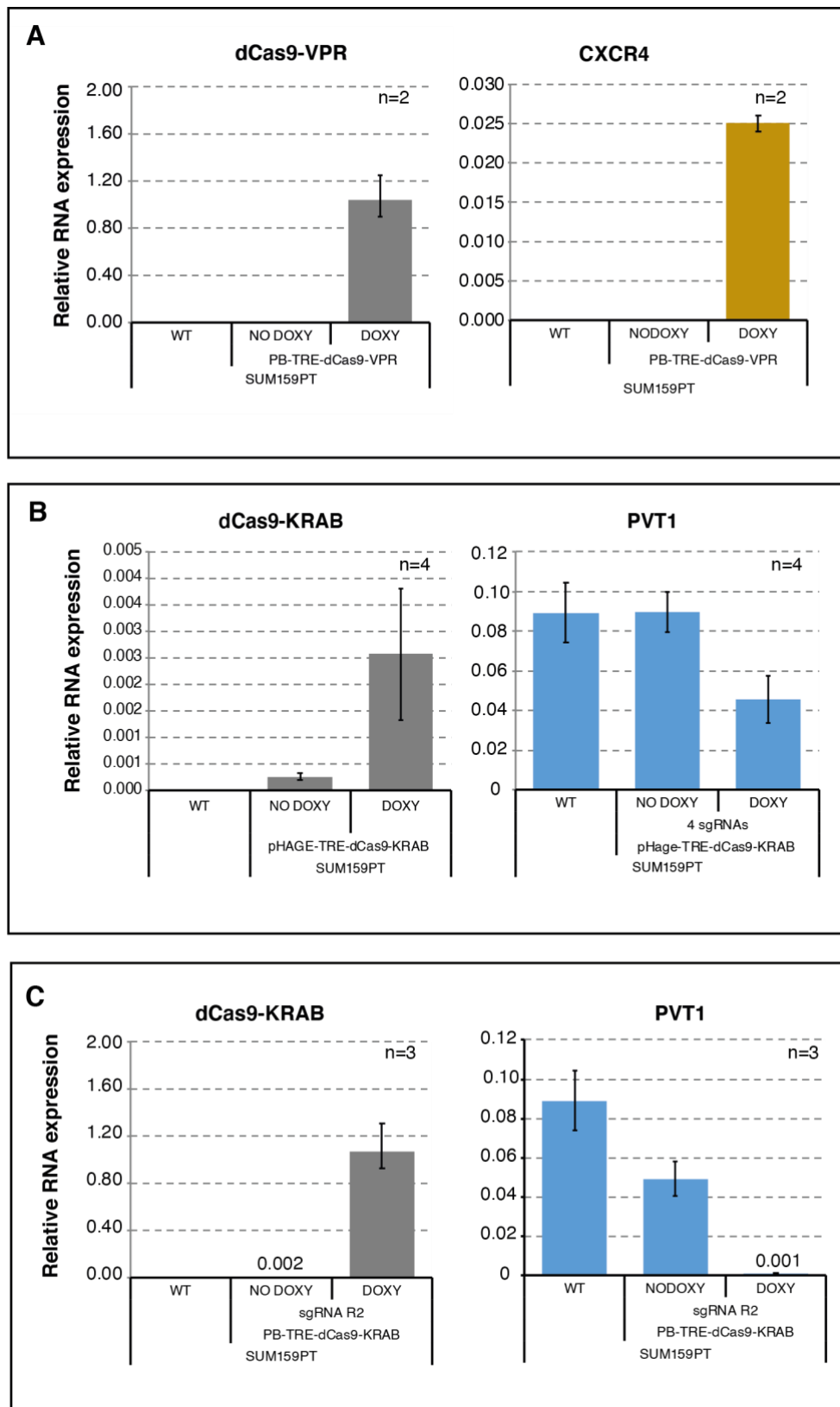
Figure 30: Performance of CRISPRi/a constructs – qRT-PCR expression data showing: **A)** Expression of dCas9-VPR from PiggyBac transposable construct and induced expression of target gene *CXCR4* (25X expression). **B)** Expression of dCas9-KRAB from lentiviral pHage contruct and expression of target gene *PVT1* (50% residual RNA expression compared to average WT). **C)** Expression of dCas9-KRAB from PiggyBac construct and expression of target gene PVT1 (1% residual RNA expression). Bars show average; error bars=SD (n in figure). Data normalized vs *RPLP0*. Expression measured after 48h Doxycyline treatment. (sgRNAs were cloned in LentiGuide-Puro sgRNA backbone, see Methods).

2.2.3 Biological properties of the SUM-dCas9-KRAB CRISPRi system

Before using the SUM-dCas9-KRAB cell line in the screenings, a series of experiments were performed to test robustness of dCas9-KRAB expression in different conditions and verify that the biological properties of the parental SUM159PT were preserved in this subline. A possible issue with Doxycycline-based inducible systems is represented by the alteration of the cellular metabolism and the slower proliferation of mammalian cells induced by Doxycycline *per se* (Ahler et al., 2013). For this reason, it is advisable to use Doxycycline at low concentrations to limit side-effects, while maintaining the effects on target genes. We calibrated the dose-response of dCas9-KRAB upon decreasing Doxycycline concentrations and measured the effect on the targeted gene PVT1. We observed that all tested doses were adequate to induce the expression of dCas9-KRAB at a sufficient level to repress PVT1 (**Figure 31**).
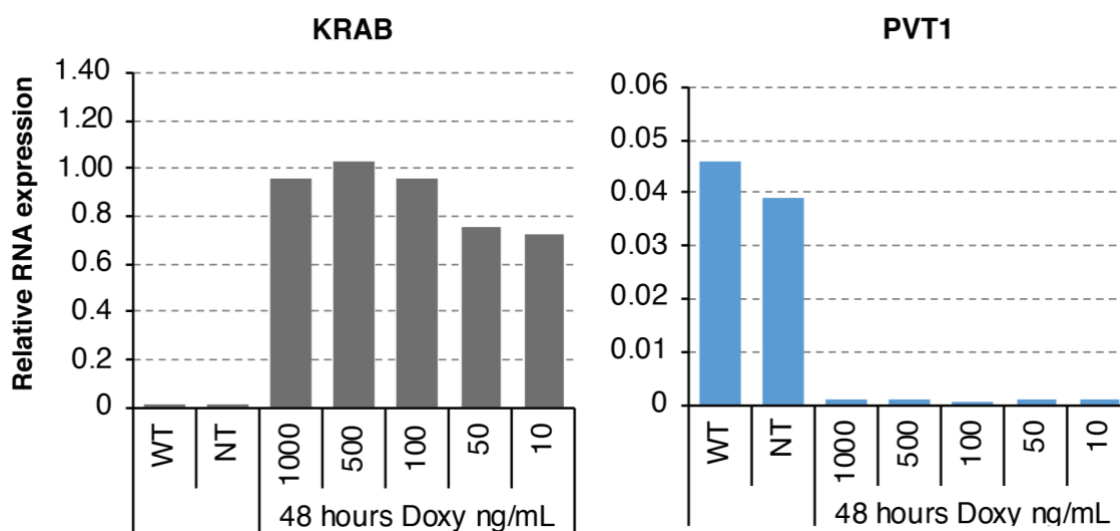
Figure 31: Dose response of dCas9-KRAB – qRT-PCR measuring the expression of dCas9-KRAB upon 48h induction with different concentration of Doxycyine and effect on target gene PVT1. Data normalized vs *RPLP0*.

Next, we checked that some peculiar biological properties of the parental SUM159PT were preserved in the SUM-dCas9-KRAB CRISPRi system. Specifically, we measured *i*) the content and the properties of CSCs by performing a mammosphere assay and measuring the Sphere Forming Efficiency (SFE %) and *ii*) the drug-sensitivity in response to paclitaxel measuring the IC50 (**Figure 32**). In both cases, we observed similar SFE% and IC50 when comparing SUM-dCas9-KRAB with parental cells. In summary, we are confident that SUM-dCas9-KRAB are truly mirroring the properties of their parental counterpart.
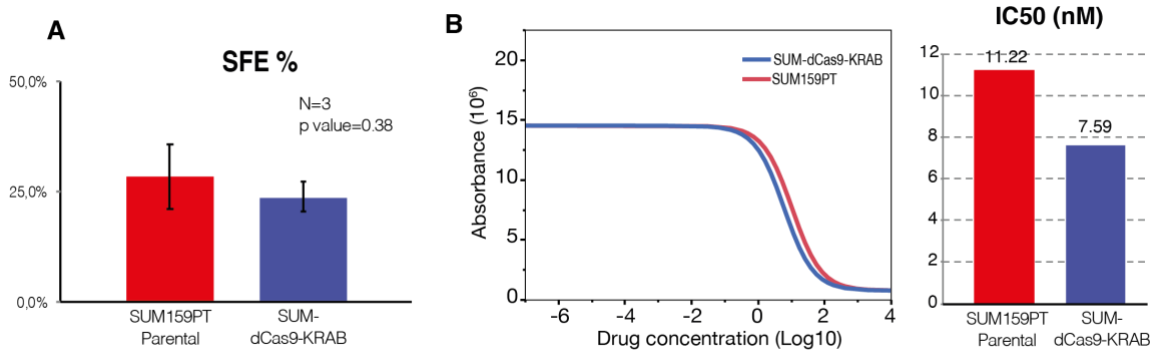
Figure 32: **Properties of SUM-dCas9-KRAB - A**) SFE % measured at the first generation (F1) of parental SUM159PT and SUM-dCas9-KRAB (Bars=average, error bars= s.e.m. from three replicates), p-values by Student's t-test; **B**) Cell viability was measured by ADP-Glo Max Assay in SUM159PT/SUM159-dCas9-KRAB cells treated for 3 days with Paclitaxel at different concentrations (1000-0.1 nM). Shown is a single representative experiment.

2.2.4   Modulation of lncRNAs expression by CRISPRi

Our next goal in setting up the system for the adaptation screenings was the design and production of sgRNAs able to interfere with the expression of the target.

We selected 7 lncRNAs from the list of candidates (See **2.1**) and STAT3, a pro survival transcription factor, involved in cell transformation and a key player in the maintenance of a stem identity. We exploited the interference of this PCG for proof-of-principle experiments that will be discussed later on. SUM-dCas9-KRAB cells express these genes at different basal levels (**Figure 33**). We selected as guide design tool the CRISPick from the Broad Institute. This web tool is optimized for the different CRISPR/Cas9 applications including CRISPRi. The tool searches for PAM sites in the genomic region that goes from -50 to + 300 bp from the annotated TSS or the corresponding genomic coordinates of the selected gene. The picking algorithm (Doench et al., 2016), searches for the sgRNAs that have the best on-target score and the lowest off-target effect, returning a ranked list of the best 20-bp protospacers. Using this tool, we designed 6-8 sgRNAs per gene, for a total of 55 guides that were individually cloned in the PerturbSeq sgRNA backbone (pBA439, Addgene #85967, (Adamson et al., 2016), (Dixit et al., 2016)). We measured the expression of the target genes after 72h of 100 ng/mL Doxycycline treatment for dCas9-KRAB induction. Of the tested sgRNAs, 37 guides (67%) produced a KD of the target genes to 20% or less of their expression in non-targeting controls cells; for 14 guides (25%) we measured a residual RNA expression between 20 and 80%, 4 guides (7%) failed to yield a KD of the target gene. Overall, we measured a mean residual RNA expression of 25% and for each tested gene but

55

one (RP-1154H7.4), we were able to identify at least 2 sgRNAs able to KD the expression of the target gene to a residual 20% (i.e. a repression of 80%). We concluded that the selected design tool and CRISPRi system are reliable tools to induce strong KD of targeted lncRNA genes.
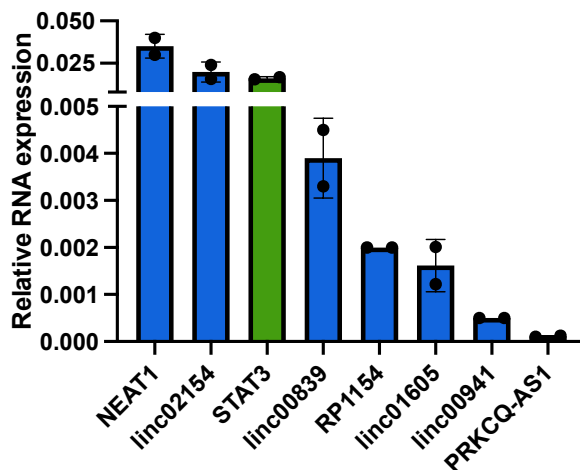
Figure 33: Baseline expression of genes tested in CRISPRi – qRT-PCR for relative RNA levels of genes KD by CRISPRi (below) measured in non-targeting control cells. Bars show mean (n=2), error bar= ± SD. Normalization vs *RPLP0*.

.
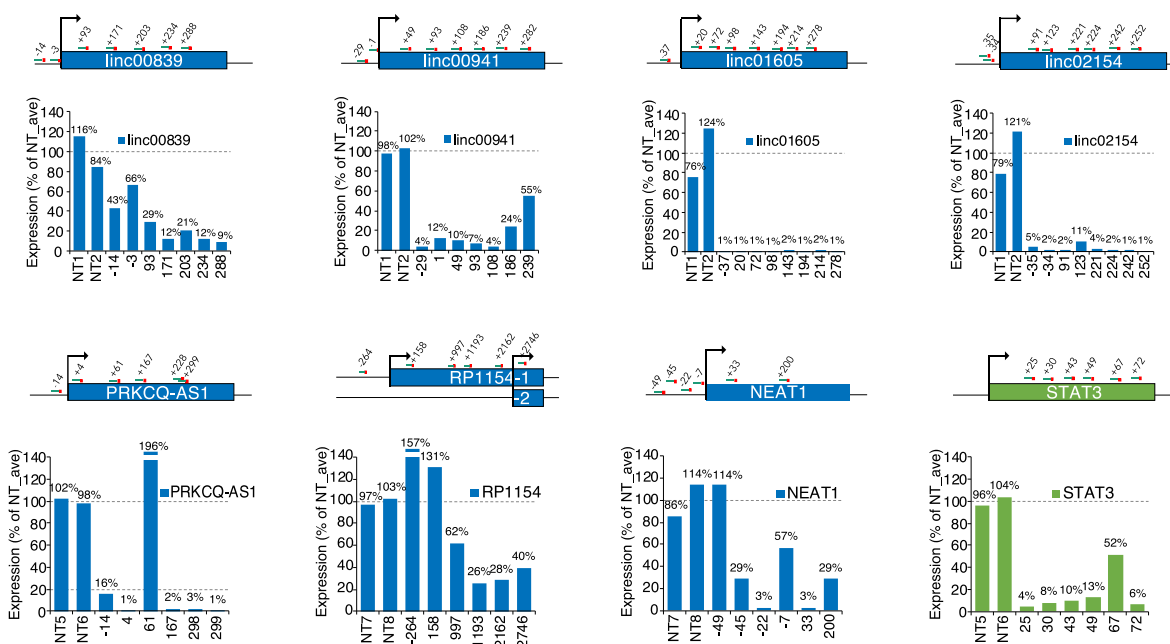
Figure 34: KD by CRISPRi of lncRNAs and the transcription factor STAT3 – Bars represents residual RNA expression in each sample as % of the average of the non-targeting controls. The cartoons show the location from the TSS of sgRNAs employed for this test. sgRNAs for this experiment were cloned in Perturb-Seq backbone (see **Methods**).
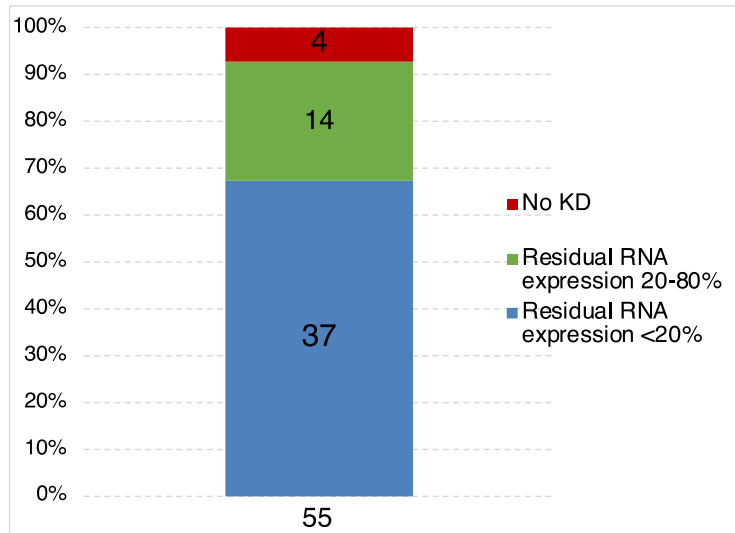
Figure 35: Summary of CRISPRi KD efficiency for lncRNAs targeting – Bar chart summarizing the effect of 55 sgRNAs tested.

2.3 LNC-Library design and production of reference P0s

2.3.1 Definition of a Transcription Start Site

The CRISPRi system relies on the targeting of gene promoters by sgRNAs. The most active sgRNAs are located just after the TSS, where the sgRNAs-dCas9-KRAB complex most efficiently interferes with transcription initiation and elongation (Gilbert et al., 2013; Horlbeck et al., 2016; Qi et al., 2013).
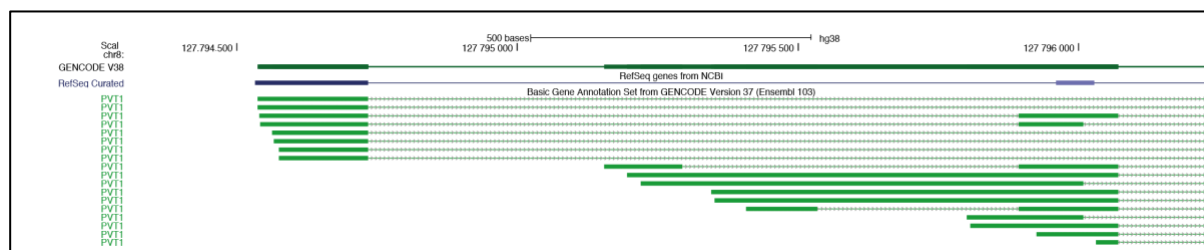
Figure 36: Multiple TSS for one lncRNA – Screenshot of the UCSC genome browser showing the locus of the lncRNA PVT1 and the different transcripts produced. The expression of this gene is known to be initiated from distant alternative promoters (Cho et al., 2018).

LncRNAs are often transcribed in multiple transcripts, which might originate from different promoters that act in a tissue- or cell line-specific fashion (Liu et al., 2017b; Mattioli et al., 2019). Distinguishing which of these transcripts are expressed in our model system is mandatory for an efficient perturbation of the target gene. For this purpose, we exploited in-house data of native elongating transcript–cap analysis of gene expression (NET-CAGE), obtained in SUM159PT to precisely define the TSS of our candidate lncRNAs. This dataset was produced in collaboration with the lab of Yasuhiro Murakawa at the RIKEN institute. NET-Cage is a strand-specific 5' mapping technique that combines the isolation of chromatin-retained nascent transcripts with cap-trapping allowing precise detection of 5' transcription start site of PCGs and lncRNAs (Hirabayashi et al., 2019). With the help of these data, we refined the TSS of the candidate genes in our library based on two approaches that are summarized below (see also **Figure 37** and Methods):

1) Approach 1: we used the annotation of the TSS defined by the RIKEN NET-Cage pipeline (Hirabayashi et al., 2019), where each peak is connected to an ENSG_ID or ENST_ID. Unfortunately, this approach provided an association only for a limited fraction (10%) of the lncRNAs candidates.

2) Approach 2: we defined a genomic window of 3kb around the TSS (-2kb; +1kb) of the candidates, based on their annotation (Gencode Release 26 was used for

consistency with the analysis performed for selecting the candidates (See **2.1**). We intersected these genomic windows with the NET-cage peaks and selected as TSS the one that was closest to the Gencode Basic annotation (GC26). This approach allows the selection of peaks quite distant from the annotated locus but could lead to the incorrect annotation of a distinct gene if it is overlapping in the same genomic region.
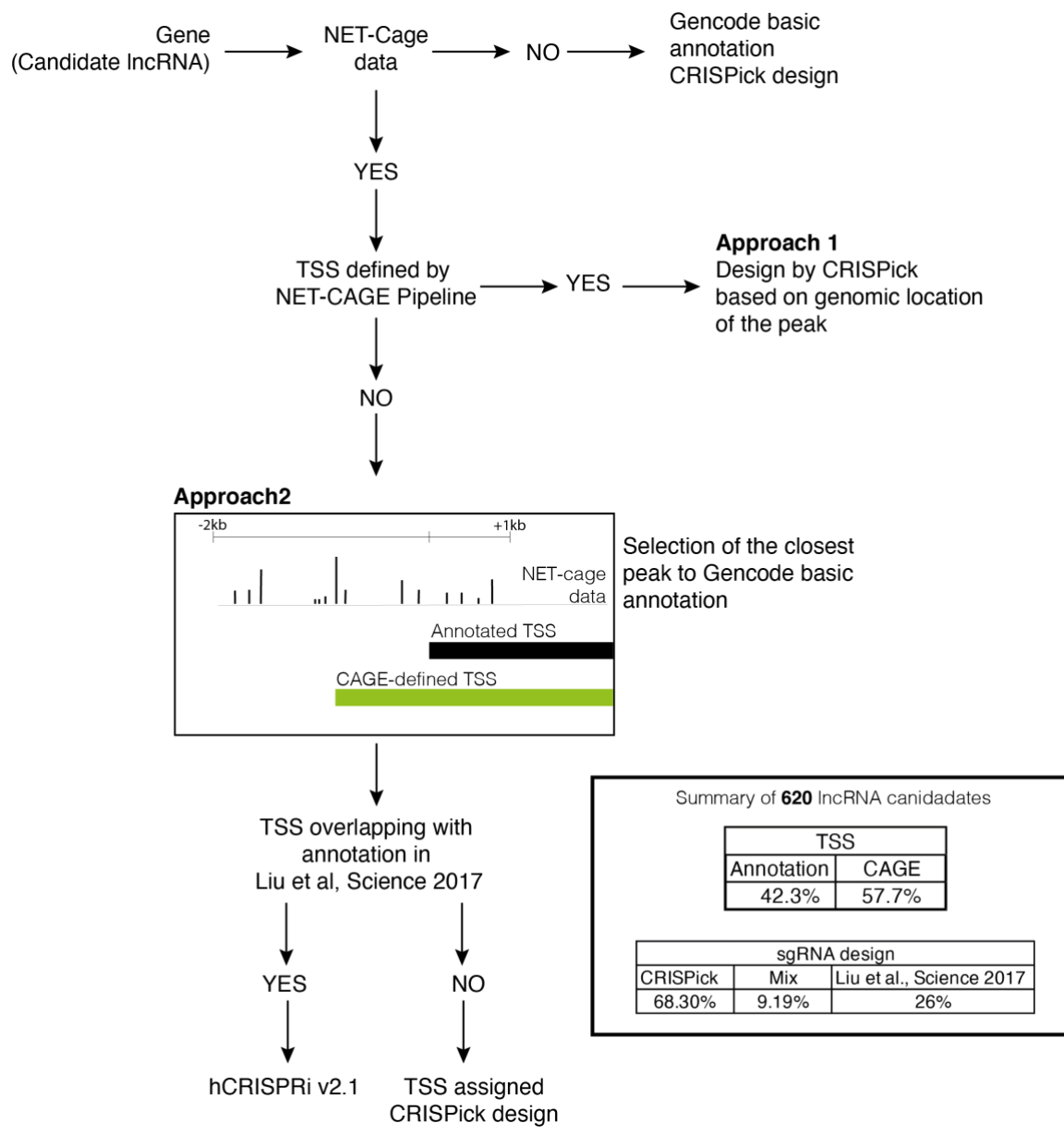
Figure 37: Schematic of TSS definition and library design – Workflow of TSS definition for lncRNAs candidates. Box summarizes the number of lncRNAs candidate with TSS defined by CAGE or their annotation and the tool used for guide design.

For 57.7% of lncRNA candidates we could identify a NET-Cage supported TSS and, for most of them, the NET-cage confirmed the TSS in a position very close to the original Gencode annotation. On the contrary, for 20% of them the NET-Cage pointed out a TSS more than 350 bp distant. In those cases, guide design would have been significantly affected, leading to inefficient targeting (**Figure 38**). For the remaining 41% of the lncRNAs candidates we could not associate any NET-Cage peak and thus, we designed guides for these genes keeping the original annotation.

In order to further support the procedure, we compared the NET-CAGE refined TSS list with an external reference, i.e. the one used for the design of the CRISPRi Non-Coding Library (CRiNCL), a large genome-wide library targeting 16401 TSS of lncRNAs genes (Liu et al., 2017b). In that paper, the authors exploited CAGE to define TSSs: we found that in 57% of the cases the two CAGE peaks (in house and external) overlapped (195/342). Besides validating these 195 TSS, this analysis also allowed us to pick sgRNA from the CriNCL library. The CRiNCL library was designed using the hCRISPRi-v2.1 algorithm, an optimized tool for the design of CRISPRi-specific guides that integrates different type of data, including the FANTOM5 TSS annotation (Andersson et al., 2014), chromatin accessibility, sequence content and sgRNAs position to design guides highly active in repressing the target gene and that were extensively validated (Horlbeck et al., 2016). In conclusion, when the TSS were overlapping, we picked sgRNAs targeting candidates from this library. Conversely, if the gene was not present in the CRiNCL library or the TSS were differently annotated, we designed the sgRNAs using the CRISPick tool (**Figure 37**).
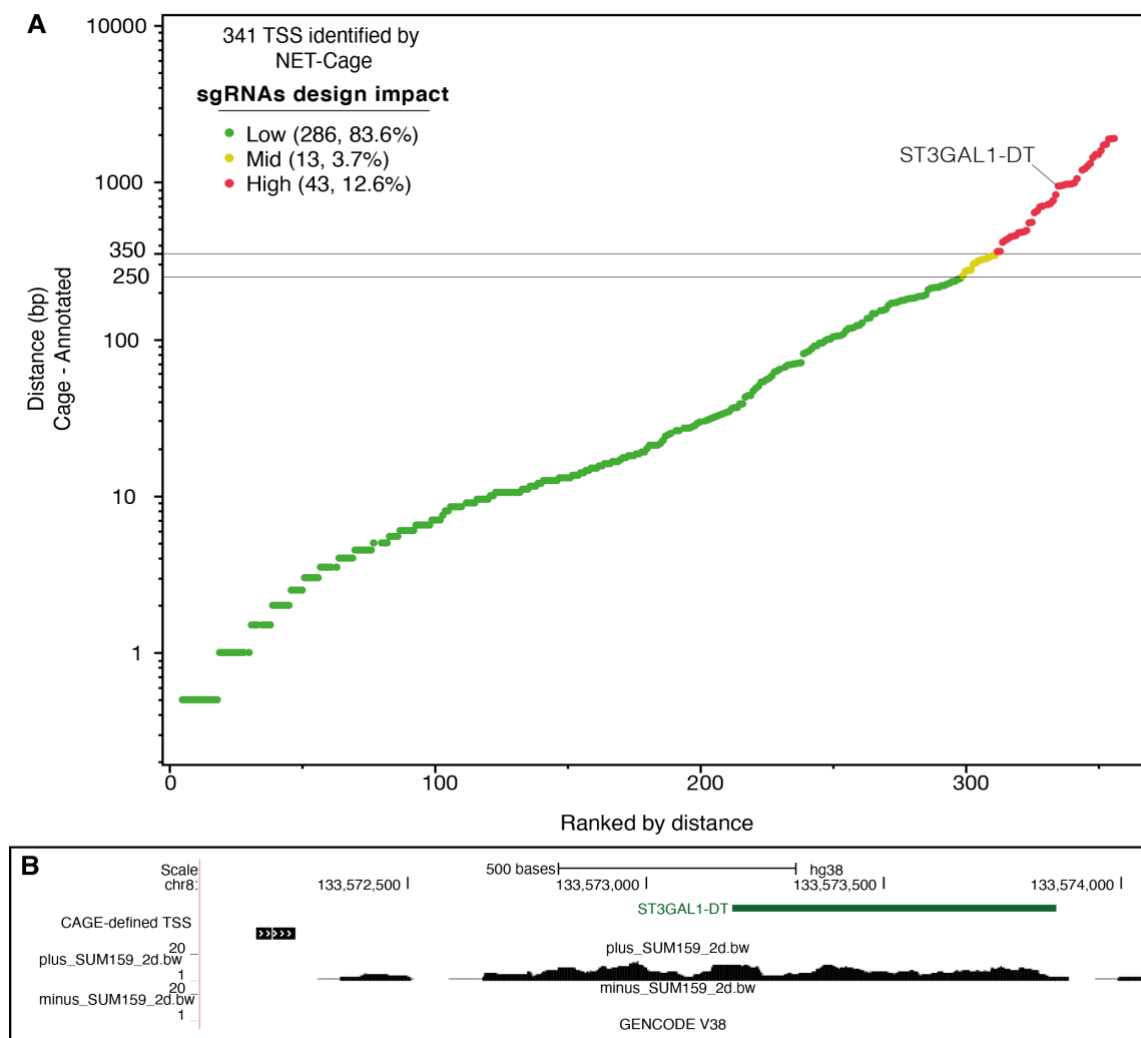
Figure 38: TSS definition by CAGE assisted guide design – **A)** Ranked representation of all 341 sgRNAs defined by CAGE data (in house NET-cage or available). **B)** Representative image of the locus of the candidate lncRNA S3GAL4-DT and the location of CAGE-defined TSS. Below, the RNA-seq track in SUM159PT confirms transcription starting several hundred base pairs upstream of the GENCODE annotated TSS (in green).

2.3.2 Modules of the Library

We aimed at designing a compact library, a necessary requirement when screening for phenotypes with strong selection, such as drug resistance. It was demonstrated that compact libraries with 5 highly active sgRNAs per target gene are equally performing to larger libraries targeting each gene with 10 different sgRNAs (Horlbeck et al., 2020). The 37% of the sgRNAs in our screening were previously validated in (Liu et al., 2017b) and we measured an efficient KD for most of the guides that we designed with the CRISPick tool (See **2.2.4**). Overall, the library (hereafter referred to as "LNC-library") has a designed library size of 3451 different sgRNAs that represents the *complexity* of the system (**Figure 39**).

The design of the library included the following modules:

Candidates lncRNAs

The sgRNAs targeting the TSS of the 620 lncRNA candidates, selected in the models of adaptive responses in breast cancer, represent the majority of the guides in the library. (583 DEGs and 37 lncRNAs selected from literature).

Essential genes

We selected 23 genes reported to be essential in a genome-wide CRISPRi screening (Horlbeck et al., 2020). Some of the genes in this module have a direct role in basal properties of the cell (e.g. cell cycle, DNA replication) while others impact cell proliferation in specific contexts (e.g. STAT3, specific inhibitor of 3D growth). The targeting of these genes acts as proxy of the function of the dCas9-KRAB. During the experimental framework, the activity of dCas9-KRAB towards the promoter of these genes should lead to a decreased proliferation rate and consequently to a progressive reduction of the representation of these guides.

Non-Targeting controls

The outcome of the screenings can only be evaluated with the help of non-targeting controls. These guides were selected from (Horlbeck et al., 2020; Liu et al., 2017b) and have no targets in the human genome or target desert regions. Their representation should remain unchanged throughout the screening. The non-targeting guides (NTCs) represent a considerable fraction of the library (236 guides, 7% of the total) and their distribution helps calling hits in the screening.
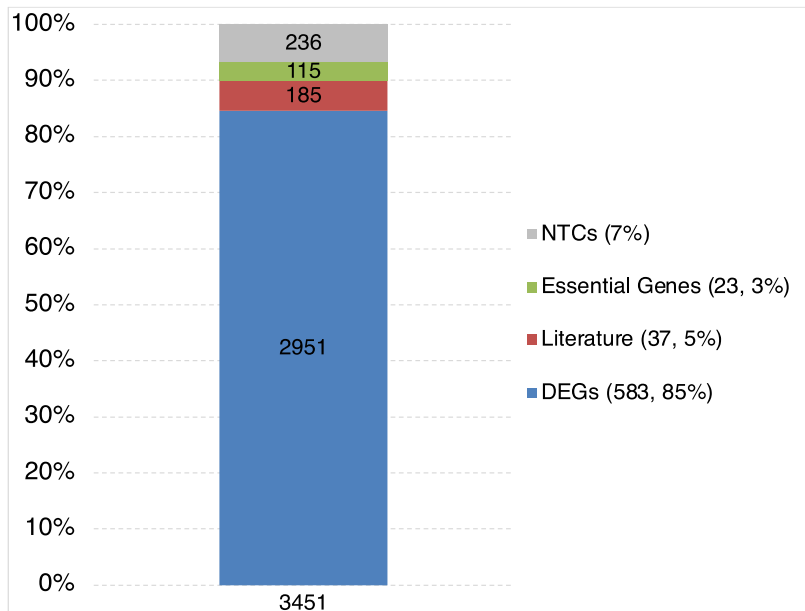
Figure 39: Composition of the LNC-library

2.3.3 Construct for sgRNA library delivery

We externalized the synthesis and cloning of the LNC-Library to Cellecta. We selected as sgRNA delivery system the pRGScribe1, a 3rd generation lentiviral plasmid, developed by the company for CROP-Seq approaches (Datlinger et al., 2017) (**Figure 40**). Briefly, this type of plasmid allows two levels of analysis. The expression of the sgRNA cassette is orchestrated by the U6 promoter. On the other side, the transcript encoding the RFP fluorescent reporter and Puromycin resistance genes embeds the sgRNAs cassette before its polyA tail. In this way, the sgRNAs cassette operates also as cellular barcode, allowing the recovery of sgRNAs identity and transcriptional perturbation of each cell in sc-RNAseq settings.

Details about amplification of the sgRNA cassette, sequencing and sgRNAs deconvolution are provided in the **Methods** section.
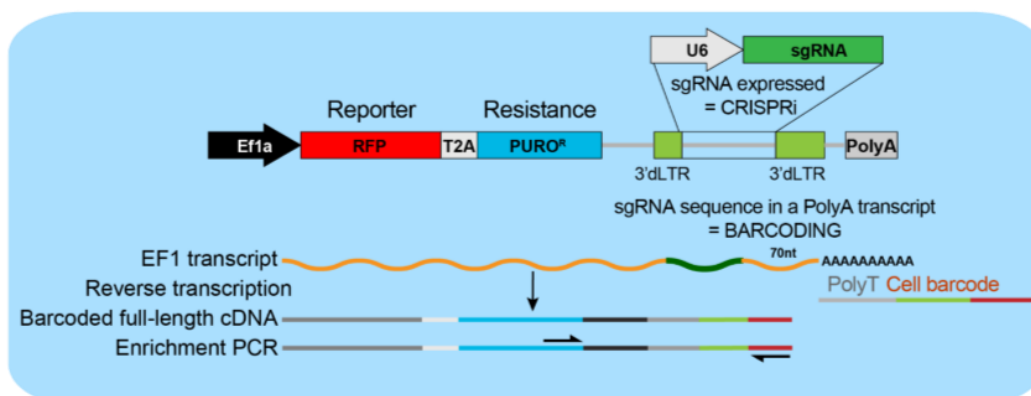
Figure 40: Schematic of plasmid for LNC-Library delivery

2.3.4 Library quality controls

Cellecta synthetized the pool of oligos and cloned them in the pRGScribe1 plasmid producing a pooled lentiviral DNA library. The performance of competition screenings depends in the first place, on the equal representation of all sgRNAs in the parental population, which is dictated by a similar representation of sgRNAs in the pooled DNA library. Cellecta sequenced the DNA pool by NGS and provided quality controls indicating that *i*) 3450 sgRNAs could be detected (3450/3451 designed sgRNAs); *ii*) the 90% of the sequences were equally represented (**Figure 41**). They also picked random clones that were individually sequenced by Sanger Sequencing and estimated that >95% of them have no mutation in the sgRNA insert (calculated mutation rate <0.25%).
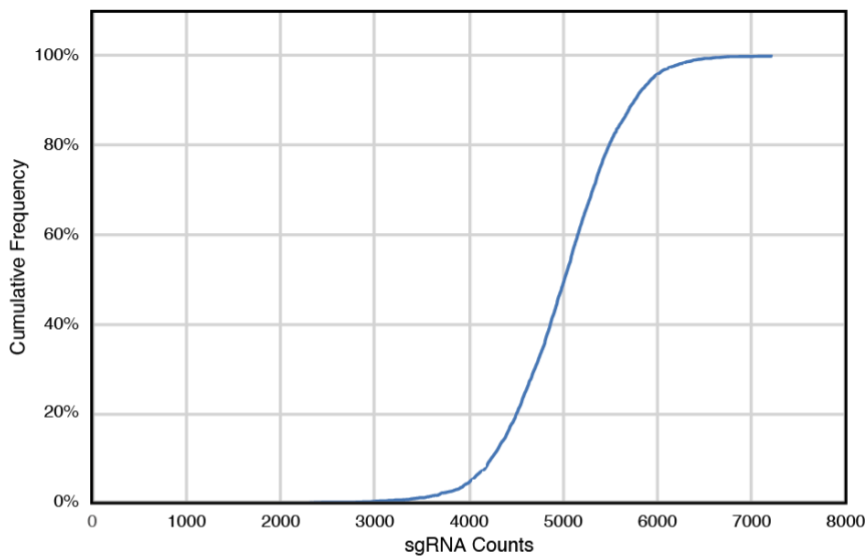
NGS distribution percentile DNA library (Counts)	
5th%	4008
10th%	4228
90th%	5744
95th%	5932
90th%/10th% (80%)	1.36
95th%/5th% (90%)	1.48

Figure 41: Cumulative distribution of sgRNAs in the LNC-library DNA – The plasmid LNC-library DNA was sequenced by Cellecta (Illumina Next-Seq 500) obtaining a total of 17196386 reads. Plot shows cumulative frequency and summary of sgRNAs distribution.

2.3.5 Production of P0s

Given the correct representation of sgRNAs in the pooled DNA library, we had to secure a similar equal representation of the sgRNAs in the parental population (P0). To do so, there are two interconnected considerations that need to be done. First, each cell must carry a single sgRNA: multiple integrations increase chances of unwanted insertions in relevant

genomic loci (e.g., insertions disrupting the sequence of oncogenes or tumor-suppressors); even more importantly, multiple integrations might lead to ambiguous interpretation of screening results. Secondly, each sgRNA needs to be represented multiple times in the P0 so that the interpretation of screening data is statistically supported. For these reasons, when producing the P0 the appropriate *Multiplicity of Infection* (*MOI*) must be set, guaranteeing single lentiviral integration and an appropriate number of cells must be infected, reaching the planned *library coverage* (i.e., the number of cells carrying a specific sgRNA in the P0). *MOI* is an empirical parameter that illustrates the ratio between the viral particle and the target cells. To obtain cells carrying only one construct we transduced cells at low *MOI*. *MOI* can be empirically monitored by transducing cells at different dilutions of the viral stock and analyzing by cytofluorimetry the percentage of cells expressing the fluorescent reporter. Thus, to produce the P0 we infected cells at serial dilutions of the lentiviral stock and measured the percentage of RFP+ cells at 36h post infection. We selected a dilution that gave us ~10% of RFP+ cells (1:100). To gain the planned *library coverage* (>1000 cells for each single sgRNA), we transduced 60 million cells. So, with 10% infection rate, around 6M cells were transduced and each construct was represented in the P0 >1500 times. Finally, we produced two independent P0s (these parental populations will be referred to as LNC1-P0 and LNC2-P0). We sequenced the two P0 parental population by NGS and observed the distribution of sgRNAs: similarly to the pooled library DNA, most of the constructs were similarly represented (**Figure 44**).
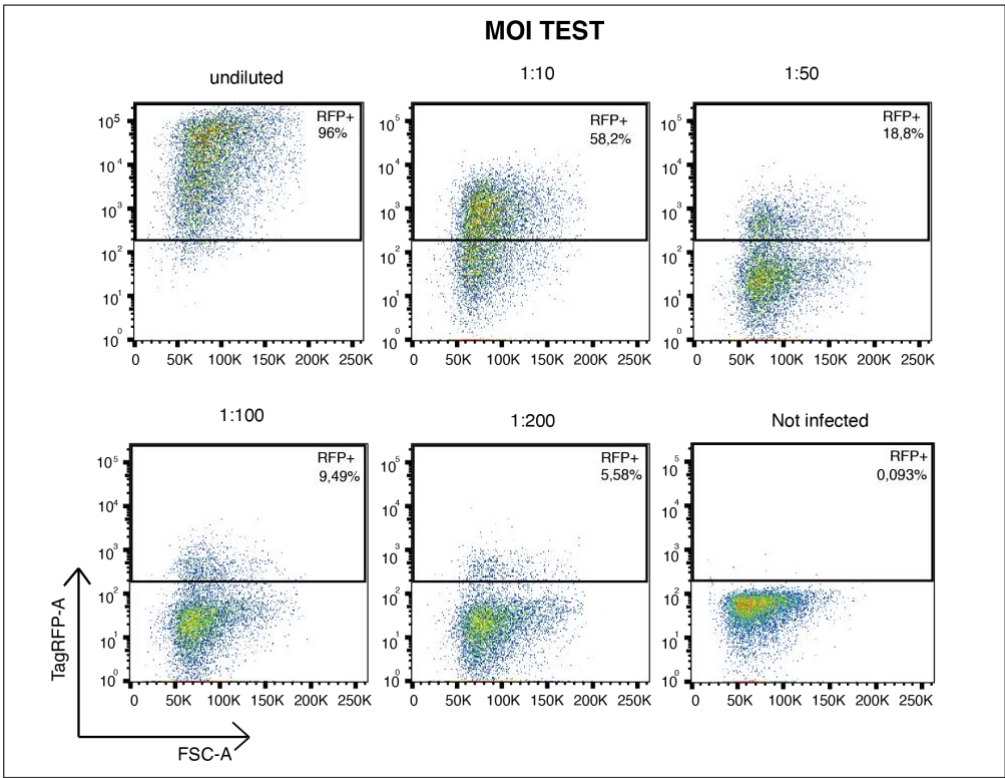
Figure 42: MOI test – The lentiviral particles encapsulating the LNC library were used to transduce SUM-dCas9-KRAB at different dilutions. Percentage of RFP+ cells were monitored 36h post infection by flow-cytometry. Flow cytometry by BD Facs Celesta.
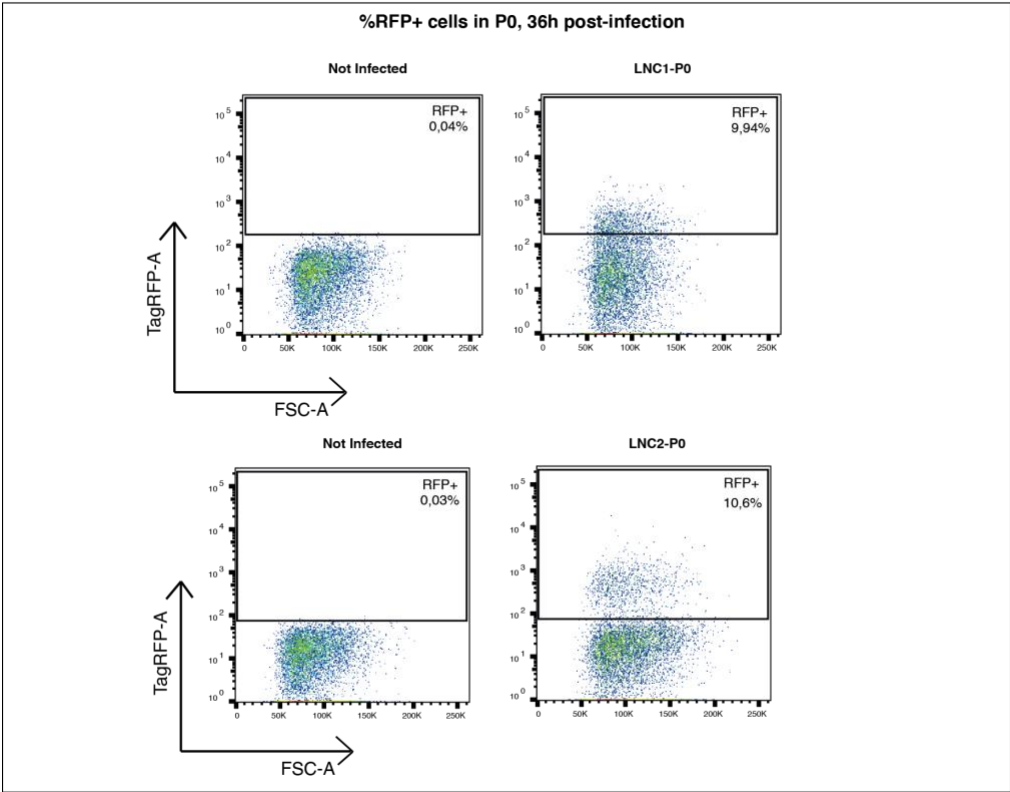
Figure 43: %RFP+ cells in LNC1-P0 and LNC2-P0 – Dot plots showing % of RFP+ cells 36h post infection in P0s before selection. Flow cytometry by BD FACS Celesta.

	LNC1-P0	LNC2-P0
5th%	1095	441
10th%	1172	476
90th%	1765	751
95th%	1854	799
90th%/10th% (80%)	1.51	1.58
95th%/5th% (90%)	1.69	1.81
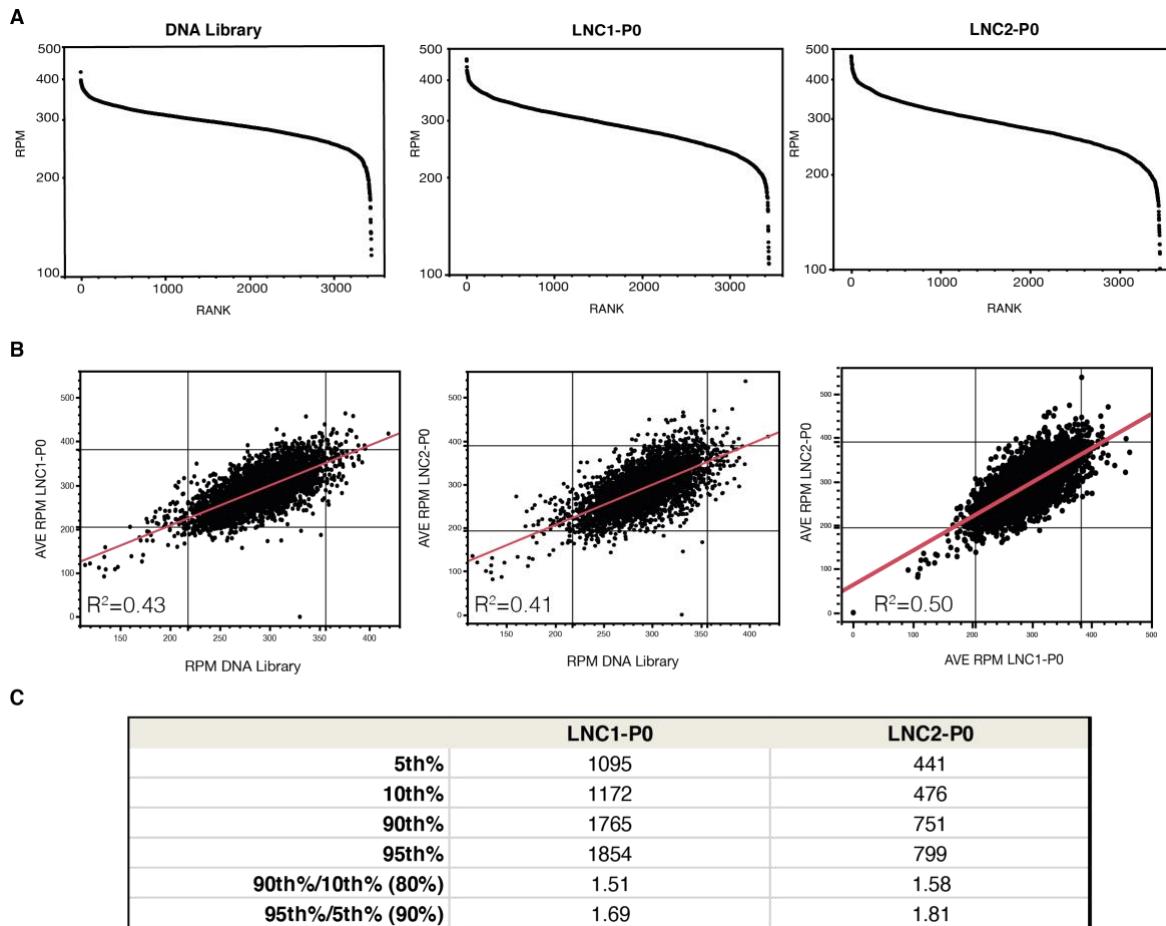
Figure 44: Distribution of sgRNAs in LNC1-P0 and LNC2-P0 – A) Ranked representation of sgRNAs in DNA library and LNC1-P0 and LNC2-P0. **B)** Scatter plot showing correlation of distribution of sgRNAs in the DNA library and parental reference population. **C)** Summary of the distribution of sgRNAs in LNC1-P0 and LNC2-P0.

2.4 2D Proliferation Screening

2.4.1 2D proliferation screening set-up
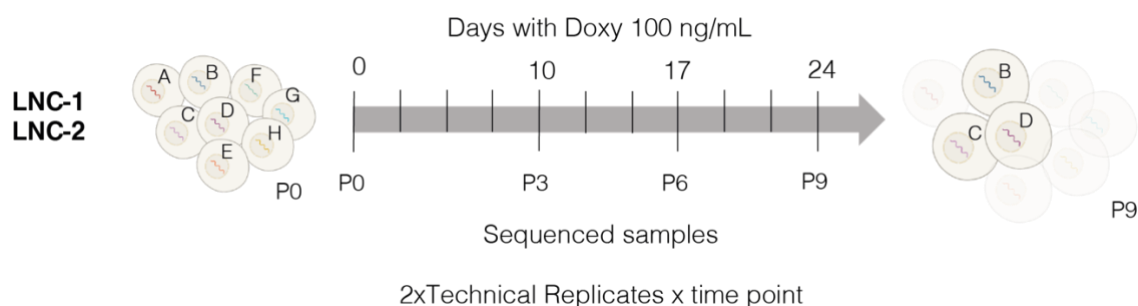
Figure 45: Outline of the proliferation screening – Cells were kept in culture for 24 days with regular split 3 times per week. We collected cell pellets from the reference (P0) and after 10 days Doxycycline treatment (P3), 17 days (P6) and 24 days (P9). In the cartoon representing population at P9, the enriched sgRNAs are highlighted while the depleted guides are shaded.

We analysed the growth modifying properties of the perturbation of the genes in the LNC-library in cells plated in a 2D monolayer. We performed the 2D screening for the following reasons: *i*) it was less difficult to set-up and helped us gather information about the correct performance of the CRISPRi model; *ii*) it was used as a reference to interpret the results of subsequent screenings, as alterations of the basal growth are a proxy of altered key cellular functions (e.g. cell cycle regulation, basal metabolism, etc…; (Liu et al., 2017b)).

We propagated the LNC1-P0 and LNC2-P0 cells (i.e. two biological replicates) for 24 days under continuous supplement of Doxycycline to induce the expression of dCas9-KRAB (**Figure 45**). We analysed the baseline (P0) and three different time points (P3, P6, P9; P = "passages"). We extracted gDNA from 3.5M cells (*library coverage*=1000X) and amplified the sgRNAs cassette to analyse enriched and depleted guides compared to the P0. First, as a quality control step, we analysed the distribution of non-targeting guides and of those targeting essential genes (See **2.4.2** and **2.4.3**). Then, we searched for hits by modelling the collective behaviour of sgRNAs at gene-level with the widely used tool MAGeCK (Li et al., 2014) (see **2.4.5**).

2.4.2 Analysis of non-targeting controls

The module of non-targeting guides (236 sgRNAs) should be consistently represented across the screening time-points. Large variation of their representation is indicative of a reduced *library coverage* due to several reasons, such as cells undergoing strong selective pressure or incorrect amplification of the sgRNAs cassette. We measured the Log2FC of the sgRNAs distribution compared to the respective P0 (LNC1-P0 and LNC2-P0). We observed that the

distribution of non-targeting guides remains unaltered across time points and biological replicates, with a median value proximal to zero. The 75% of non-targeting guides was consistently distributed with Log2FC values ranging from 0.2 to -0.1 Log2FC with only a few outliers guides more strongly and conservatively depleted (**Figure 46**). Overall, given the small variation of non-targeting guides across all timepoints, we concluded that the screening was performed with an adequate coverage and that the library was correctly amplified and faithfully sequenced.
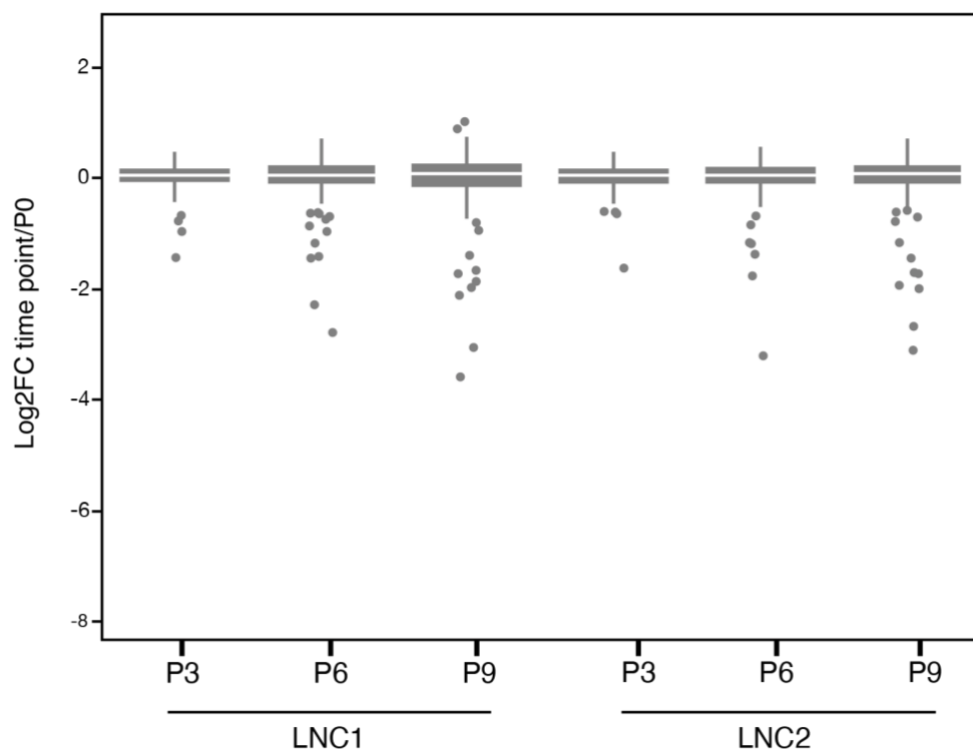
Figure 46: Distribution of 236 non-targeting guides in throughout the proliferation screening – Box plot showing the distribution of non-targeting guides in each time point (P3, P6, P9) and biological replicate of the proliferation screening (LNC1 and LNC2).

2.4.3 Analysis of depletion of sgRNAs targeting essential genes

The analysis of the behavior of non-targeting guides gives no actual clue about the correct interference on target genes. For this purpose, we analyzed the behavior of the "Essential genes" module (115 sgRNAs). The abrogation of the expression of these genes should strongly impair the fitness of the cells and hence, the representation of these guides should be progressively depleted. The boxplot in **Figure 47** describes the distribution of Log2FC of the sgRNAs targeting essential genes. When compared with non-targeting controls, it is immediate that, in both replicates, the median value for this group of guides decreases across the analyzed time points (e.g., at P9 Log2FC -1.8 for Exp1 and -1.76 for Exp2). Overall, the

"Essential genes" distribution at P9 in LNC1 and LNC2 samples was significantly different to the distribution of non-targeting guides (p-value<0.0001*, Student's paired t-test).
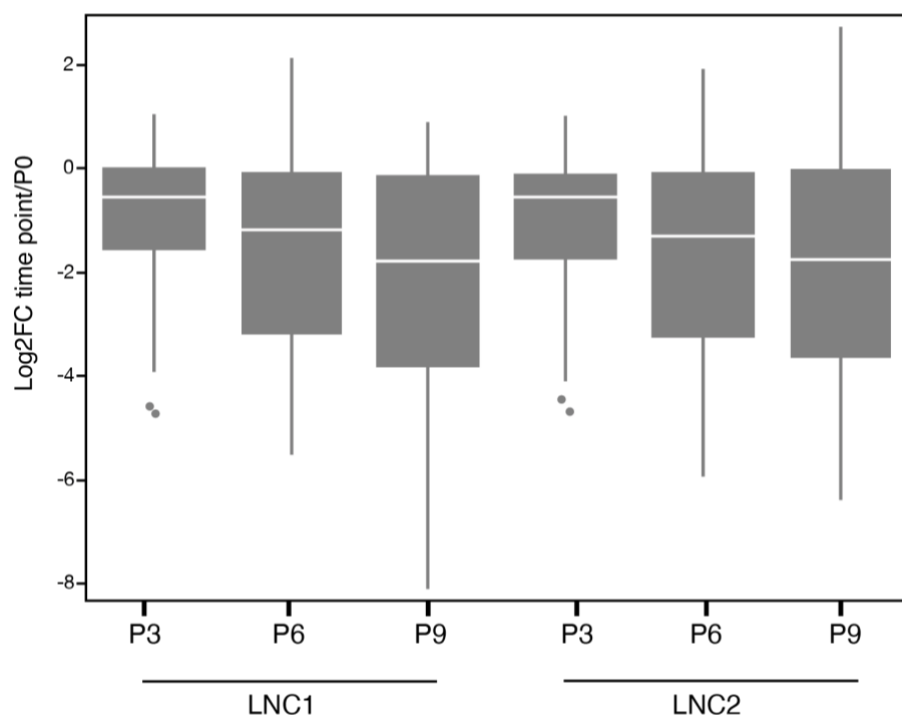
Figure 47: Distribution of guides targeting essential genes throughout the proliferation screening - Box plot showing distribution of guides targeting essential genes in each time point (P3, P6, P9) and biological replicate of the proliferation screening (LNC1 and LNC2).

We also noticed that not all the sgRNAs targeting "Essential genes" were depleted. This might be the consequence of two different scenarios: *i)* the gene annotated as "essential" was not behaving as such in SUM159PT; *ii)* the sgRNAs were not functioning in repressing the target gene. Therefore, we referenced the activity of each sgRNA to their target gene and we exploited the distribution of non-targeting controls to identify a confidence interval to properly call the guides as DROP-OUTs or DROP-INs. We defined as confidence interval at each time point, a range that covered the 5th – 95th percentile of the distribution of non-targeting guides and called hits in the "Essential gene" set with an empirical False Discovery Rate (FDR) of 10%. Those guides showing a Log2FC value that fell below or above this interval were respectively called DROP-OUTs or DROP-INs. We noticed a highly reproducible behavior of the sgRNAs in the two independent experiments.
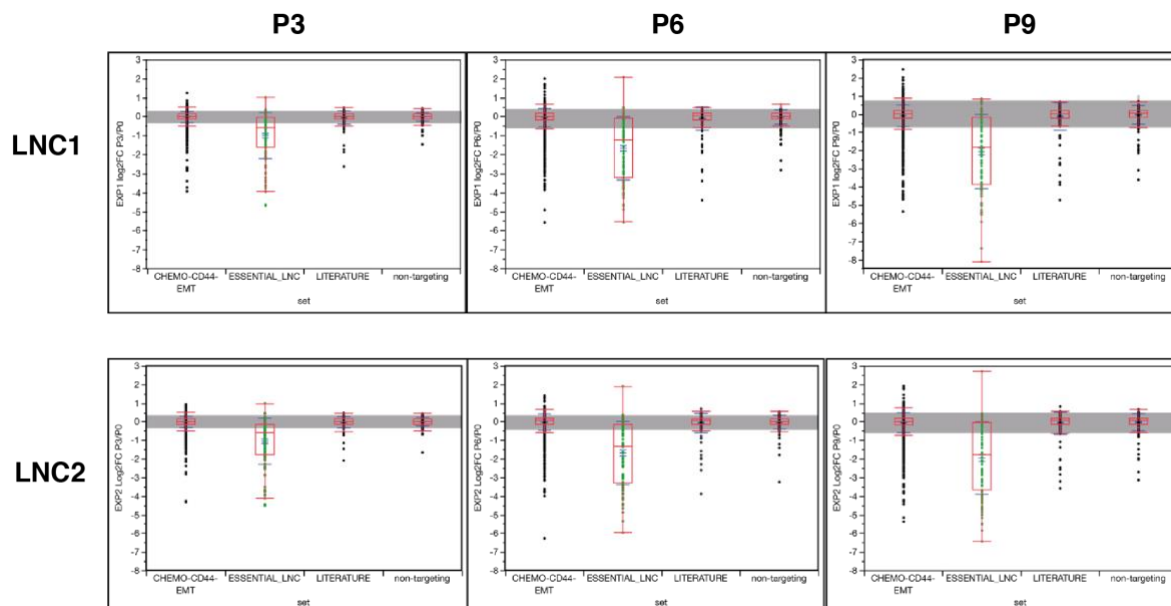
Figure 48: Confidence Intervals – Box plot showing the distribution of modules of the LNC-library in all samples of 2D proliferation screening. Grey area represents the calculated confidence interval for each sample (5th – 95th percentile of non-targeting distribution).
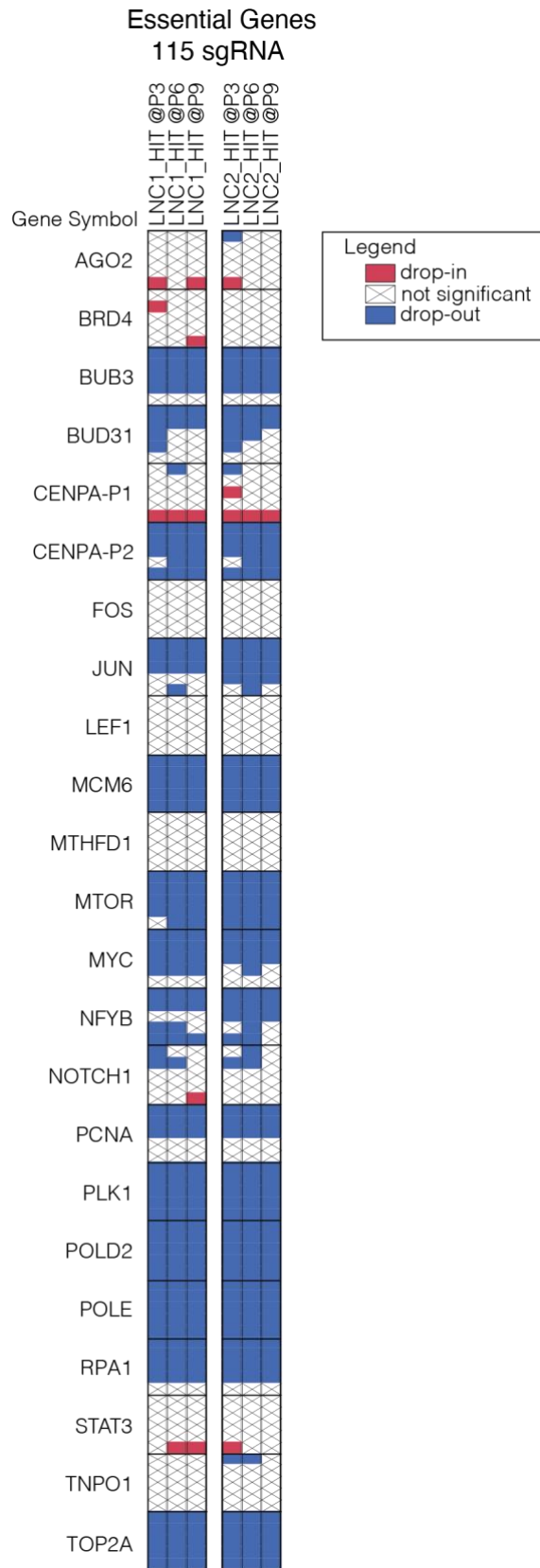
Figure 49 – Cell plot of Essential Genes sgRNAs – Digital map of sgRNAs targeting essential genes. Each cell summarizes the behavior of sgRNAs at every time point (P3, P6, P9) in the two biological replicates (LNC1 and LNC2).

We identified 14 genes that had multiple guides DROP-OUT at all time points. We compared their annotation in the DepMap portal, a repository of genome-wide genetic screening data in multiple cell lines (Meyers et al., 2017). Some of them are annotated as common essential genes (BUB3, BUD31, MCM6, MTOR, MYC, NFYB, PCNA, PLK1, POLD2, POLE, RPA1, TOP2A, CENPA-P2). We also defined as DROP-OUT hit the kinase JUN, reported in DepMap as selectively essential. Importantly, the genes that were not identified as DROP-OUT hits in this screening (AGO2, FOS, NOTCH1, LEF1, TNPO1) are reported to have a cell-line dependent selectively essential phenotype and were not reported to be essential in SUM159PT. Among the not-significantly depleted hits we also found STAT3 and MTHFD1, genes regulating other features of cancer cells (Tordonato et al., 2021; Yu et al., 2014) and dispensable during proliferation in 2D. In conclusion, these data soundly suggest that the CRISPRi system is correctly perturbing the expression of target genes.

2.4.4 Reproducibility of biological replicates

We evaluated the reproducibility of the two biological replicates, performed with LNC1-P0 and LNC2-P0 (**Figure 50**). We calculated the Log2FC per each time point by normalizing to the representation of each guide in their respective P0. We correlated the Log2FC at each time point in the two replicates. Non-targeting guides reproducibly clustered towards Log2FC values close to zero, while essential genes were reproducibly and progressively lost during passages. Overall, the samples displayed a good correlation in Log2FC values, with an R^2 of 0.63 at P3, 0.74 at P6 and 0.72 at P9.
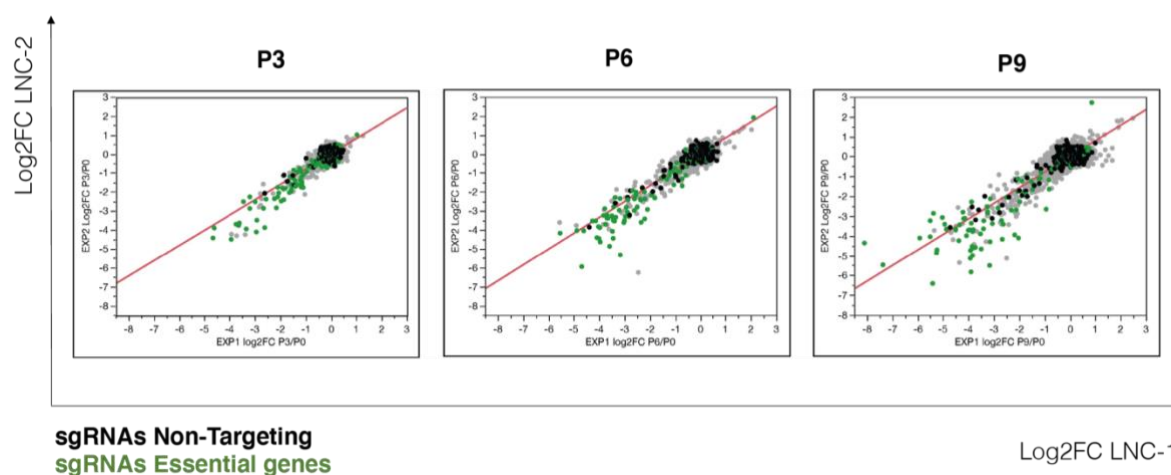
Figure 50: Correlation of biological replicates of 2D proliferation screening – Scatter plot showing the Log2FC of the 3450 sgRNA of the LNC-library in the two biological replicates of the screening. Black dots=NTCs, green dots=sgRNAs targeting PCGs of the essential genes set.

2.4.5 Analysis of the screening: MAGeCK

In **2.4.3** we showed that it is possible to take advantage of the non-targeting distribution to define confidence intervals to call hit genes. In **Figure 51**, we see highlighted the distribution of Log2FC at P9 of the sgRNAs targeting the lncRNAs genes PVT1 (literature), CKMT2-AS1 (LNC candidates) and the essential gene MYC. All the five sgRNAs targeting PVT1 are clustered together in the upper right corner of the plot and would be consistently called as DROP-INs, when considering the non-targeting distribution. For the other two genes, only two sgRNAs out of five would be called as DROP-OUTs. To bypass the arbitrary selection of screening hits based on a certain number of significant sgRNAs, we decided to model the collective behavior of all sgRNAs for a specific gene as summarized by their β-score calculated by MAGeCK. This tool applies a Maximum-Likelihood-Estimation (MLE) algorithm that, by taking into consideration the distribution (read counts) of non-targeting guides, sgRNAs targeting a specific gene in the P0 and each condition of the experiment, computes a unique β-score (and paired statistical analysis) that expresses the degree of selection of each gene. The β-score is corrected by calculating the KD efficiency of each sgRNA and thus, minimizing the effect of inefficient sgRNAs. This algorithm is also suitable for complex experimental designs that include multiple biological replicates and multiple time points, such as the 2D proliferation screening. (**Figure 52**).
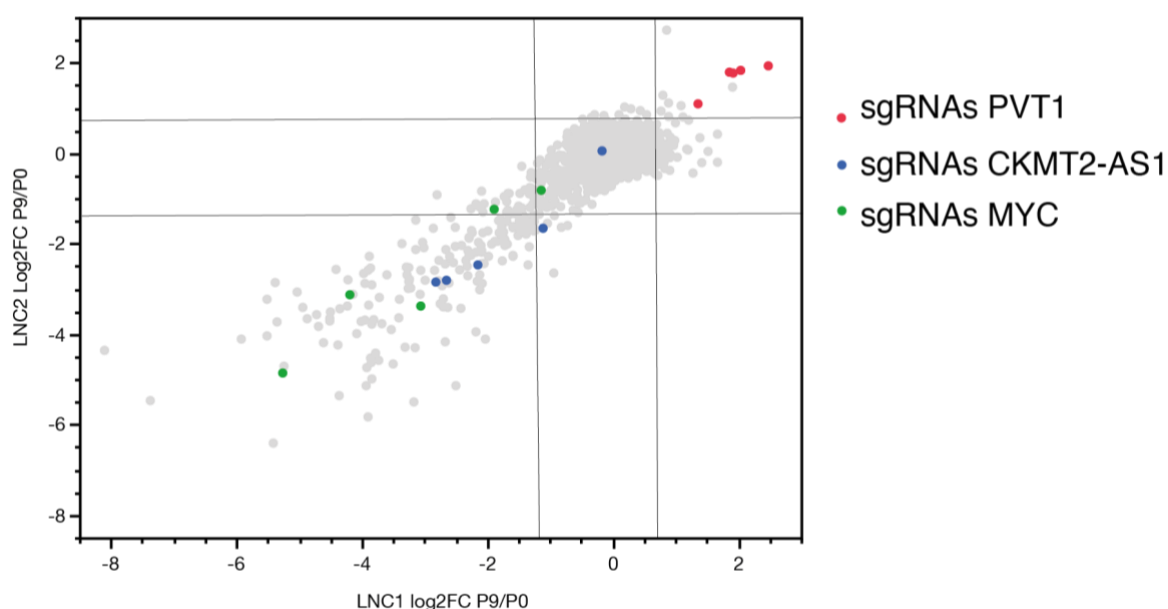
Figure 51 : Examples of depleted genes at LNC1 and LNC2 – Scatter plot of LNC1 and LNC2 Log2FC at P9. Highlighted the distribution of guides targeting PVT1, CKMT2-AS1 and MYC. Black bars represent the 5th – 95th percentile of non-targeting distribution (confidence interval) in each sample.
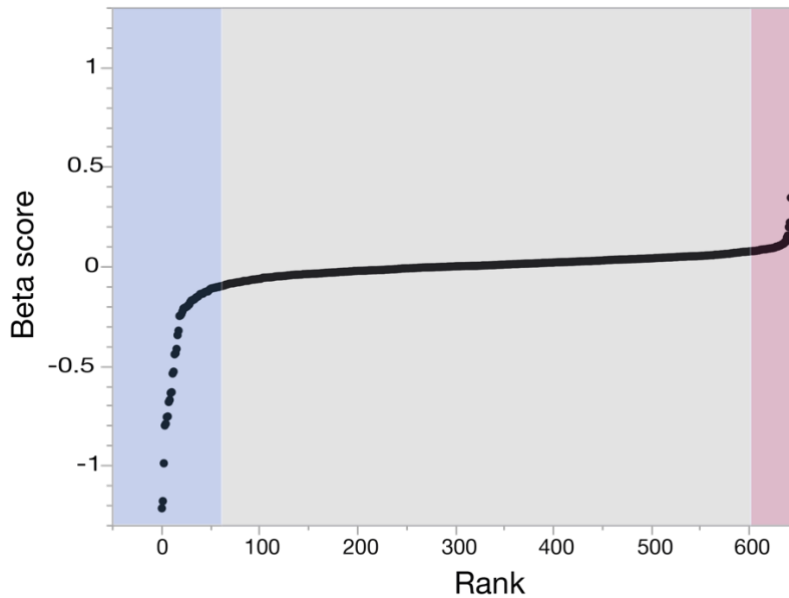
Figure 52 – Raked β-score for 2D proliferation screening. Ranked representation of β-scores for the 643 genes in the LNC-library. The blue area represents genes with negative β-score (negative selection), in red genes under positive selection. Grey area indicates genes under "neutral" selection.

With the analysis by MAGeCK we selected DROP-OUT and DROP-INs gene hits (p<0.05). We found, as the strongest depleted genes (with the highest negative β-score), the essential genes that we already characterized as DROP-OUTs in **2.4.3**. Conversely, the gene with the highest β-score was PVT1. The CRISPRi towards the promoter of PVT1 has been reported to increase the expression of MYC and therefore boosts its pro-proliferative activity (Cho et al., 2018) (**Figure 53**). Therefore, this observation and the depletion of essential genes made us confident the screening is measuring biological properties of the perturbed genes. In conclusion, we identified 32 DROP-OUT and 11 DROP-IN lncRNA hits that are potential modifiers of SUM159PT growth in 2D.
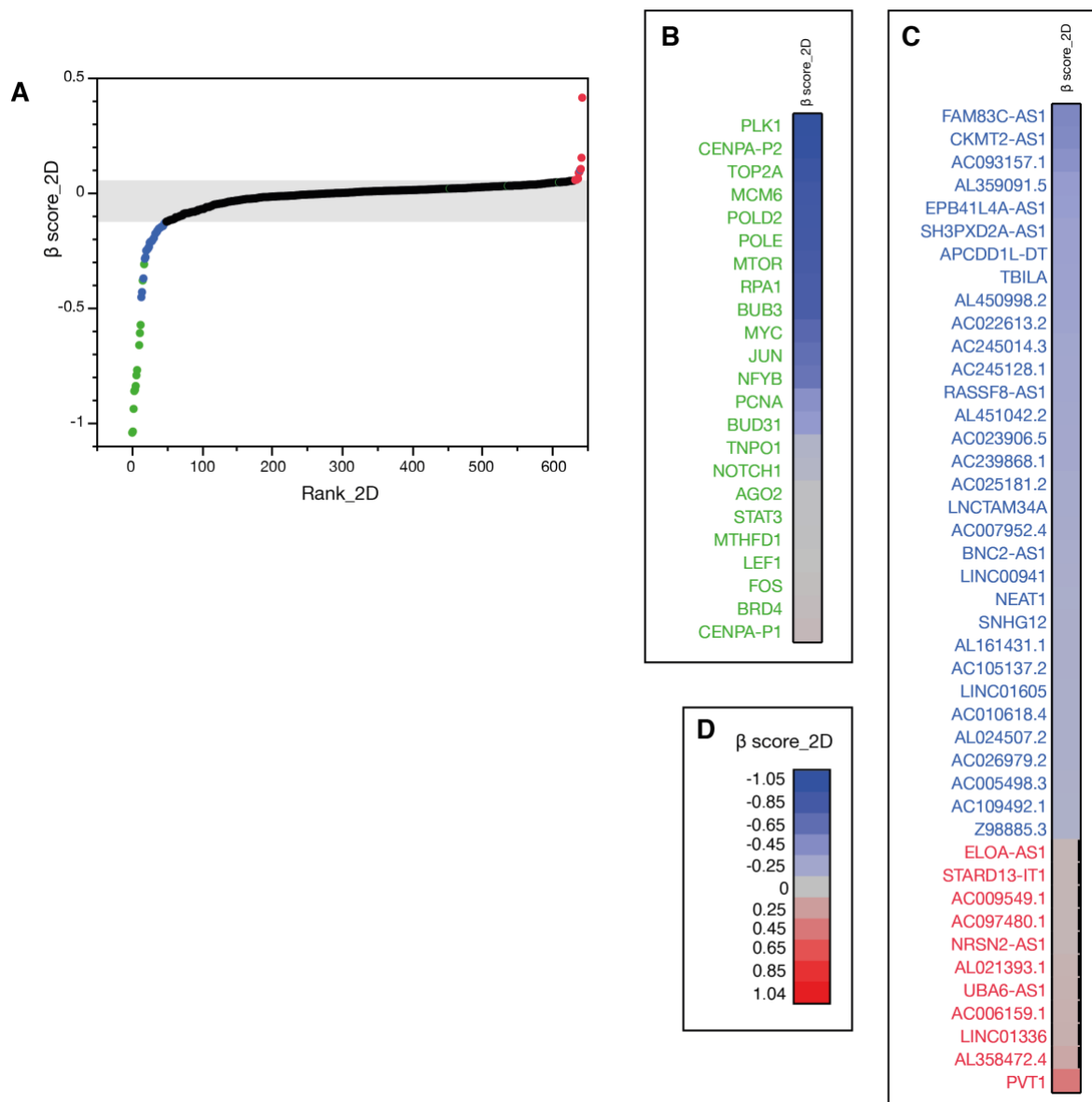
Figure 53: Hits of 2D proliferation screening by MAGeCK-MLE – **A)** Ranked list of β-score for all genes, highlighted hit genes, in grey area of non-significant β-scores. **B)** Heatmap of β-scores of all essential genes. **C)** Heatmap showing β-scores of all lncRNAs hits. **D)** Heatmap legend. Green: DROP-OUTs hits in "Essential Genes" set; Blue: DROP-OUTs lncRNAs Red: DROP-INs LNCRNAs.

2.5 3D growth screening

2.5.1 3D Screening set-up
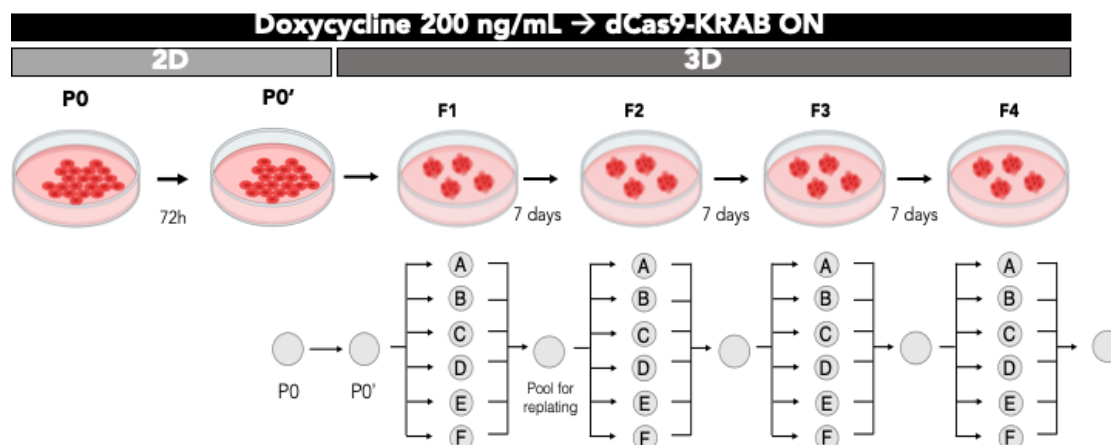
Figure 54: Outline of 3D growth screening – Before starting the 3D culture, we plated SUM159PT in presence of Doxycycline in 2D for 72h to obtain a pre-repressed state of the target lncRNAs once the non-adherent growth started. In each generation we collected data from separate pools of samples with the aim of evaluating potential issues with library coverage.

We assayed the effect of the perturbation of the genes in the LNC-library in modifying the ability of SUM159PT to grow in 3D. Non-adherent 3D growth, as monoclonal spheroids, more closely resembles primary tumors and enriches for cells with TIC properties (Dontu et al., 2003). In this screening, we activated the CRISPRi system in 2D for 72h (P0') and then we seeded cells in single cell suspension in a matrix of 1% methylcellulose in absence of serum. In these conditions, cells exploit their self-renewal ability and grow in spheroids. After one week, the spheroids were disaggregated and replated, for a total of 4 generations (F1, F2, F3, F4). In each generation, we produced 6 experimental groups ("sub-pools") to gain information about the correct *library coverage*. At each generation, after collection of the cells required for gDNA extraction, sub-pools were pooled in equal ratios to produce a representative pool for re-plating. This screening was performed with the LNC1-P0.

2.5.2 3D Screening set-up

Expression of dCas9-KRAB in 3D

During the 2D proliferation screening we plated cells in fresh Doxycycline at every passage (every 2 – 3 days). In 3D culture, after seeding cells needs to stay still for 7 days to avoid aggregation, thus media replenishments are unadvisable. We wondered if the progressive decreased availability of Doxycycline in the culture media would lead to a decreased

expression of dCas9-KRAB. We examined this issue by measuring the expression of the transgene in four different conditions in 3D culture (**Figure 55A**). We plated cells at two different dilutions (10000 and 20000 cells/mL) and with two different Doxycycline concentrations (200 ng/mL and 500 ng/mL). We observed that, indeed, in all tested conditions, there was a strong decrease in dCas9-KRAB expression moving from 2D to 3D (**Figure 55B**).
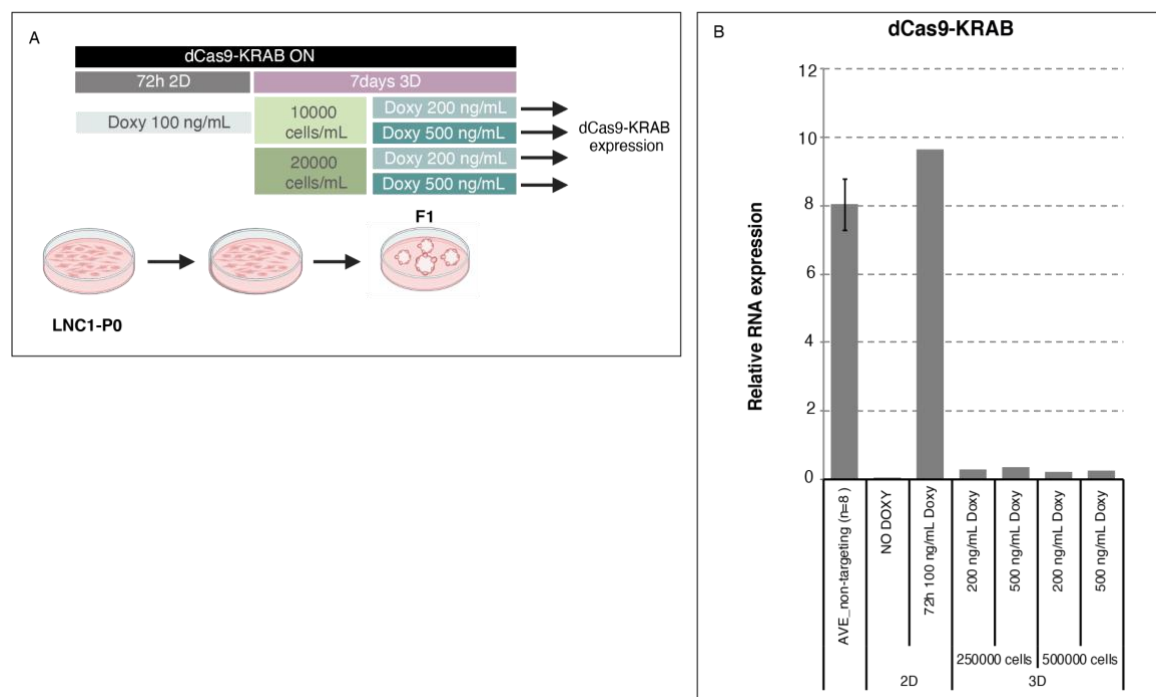
Figure 55: Expression of dCas9-KRAB in 3D culture – A) schematic representation of the evaluation of dCas9-KRAB expression in different 3D culture setting. B) Quantification of dCas9-KRAB expression by RT-qPCR. Error bar=standard error (n=8). Data normalized vs *RPLP0*.

Taking this into consideration, we wished to validate if the CRISPRi mediated knock-down of genes was preserved. We compared the expression of the LINC01605 (a 2D hit) in non-targeting cells (NT8) and in cells carrying the sgRNA LINC01605 d214 that we previously validated for producing a strong KD of the target gene (See **2.2.4**). We first induced the expression of dCas9-KRAB in 2D for 72h and propagated spheroids for two generations in media containing 200 ng/mL Doxycyline. After collecting the 2nd generation, we re-plated cells in 2D to evaluate if the expression of dCas9-KRAB was re-established (**Figure 56A**), proving that the 3D culture was not selecting sub-populations of cells expressing lower basal level of dCas9-KRAB. As observed before, moving from 2D to 3D culture the expression of dCas9-KRAB decreased (**Figure 56B**), but LINC01605 was still repressed at all

considered time-points (**Figure 56C**) and the expression of dCas9-KRAB was re-boosted (even if, not at the same level as the first time point) once re-plated in 2D (**Figure 56B**).
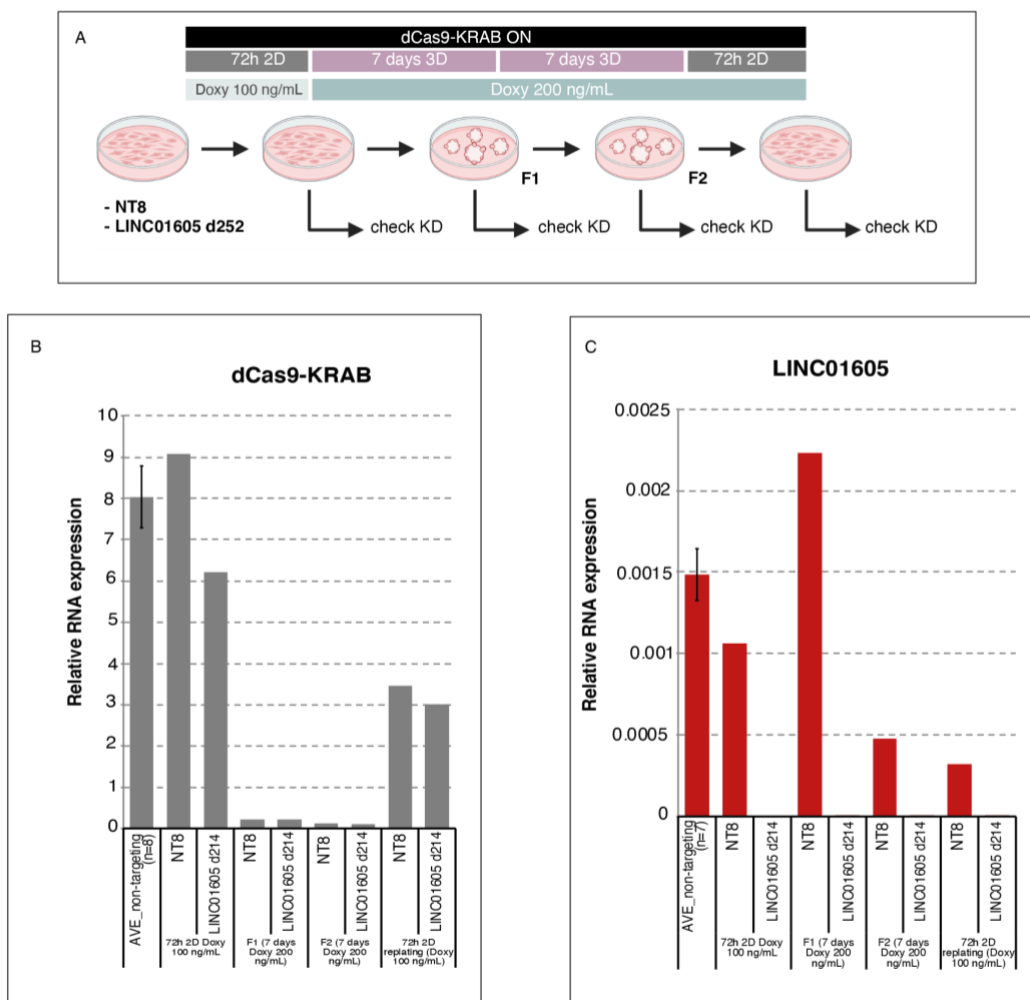
Figure 56: KD of target gene in 3D culture – A) Scheme representing validation of KD of LINC01605 in 3D culture. Expression of **B)** dCas9-KRAB and **C)** LINC01605 in NT8 and LINC01605 d214 cells in 2D before 3D culture, for two generations in 3D and upon re-plating in 2D.

Library coverage

The 3D growth introduces a new parameter to take into consideration when setting-up the screening: the selective pressure of this specific kind of growth. As previously described (See **2.2.1**), SUM159PT is a heterogeneous cell line and not all cells have the same capability of clonal growth. Normally, the SFE of SUM159PT ranges from 10 to 20% depending on the conditions of the assay. In other words, only 1:10 – 1:5 cells are able to grow. This factor affects the complexity of the system and therefore the number of cells screened must be increased from 5 to 10 times. We previously established that in a 2D setting, 3.5M cells, represent our preferential *library coverage* of 1000X (*Library size=3450*). In spheroids with an SFE of 20%, we expect that 18M cells (3.5M X 5) would allow to maintain the same

representation of sgRNAs. Furthermore, to assure clonal growth it is essential to limit cell aggregation. For this reason, cells are usually seeded in sphere-forming assays at a very low concentration (e.g. 200 cells/mL). Of course, these conditions are experimentally unfeasible for our *coverage* needs. Thus, we tested the cell aggregation at higher cell concentrations, by plating equal ratios of RFP and GFP expressing cells at a concentration of 10000 cells/mL and 20000 cells/mL in 15cm dishes (for a total of 250K and 500K seeded cells) (See **Methods**). We counted the number of RFP+, GFP+ and mixed spheroids. We observed that the spheroids in both conditions were mostly single-colored, with a percentage of mixed spheroids of 6% at 10000 cells/mL (**Figure 57B**) and 21% at 20000 cells/mL (**Figure 57C**). This approach does not take into account the aggregation of cells of the same colour. We decided to perform the screening at 20000 cells/mL for feasibility reasons, even if we observed some aggregation at this concentration. We had the additional accuracy to divide the cells in six sub-pools for parallel analysis (See **Figure 54** and **Figure 62**), allowing a better resolution of the actual coverage needed for future experiments and the dissection of the possible noise due to cell-aggregation.
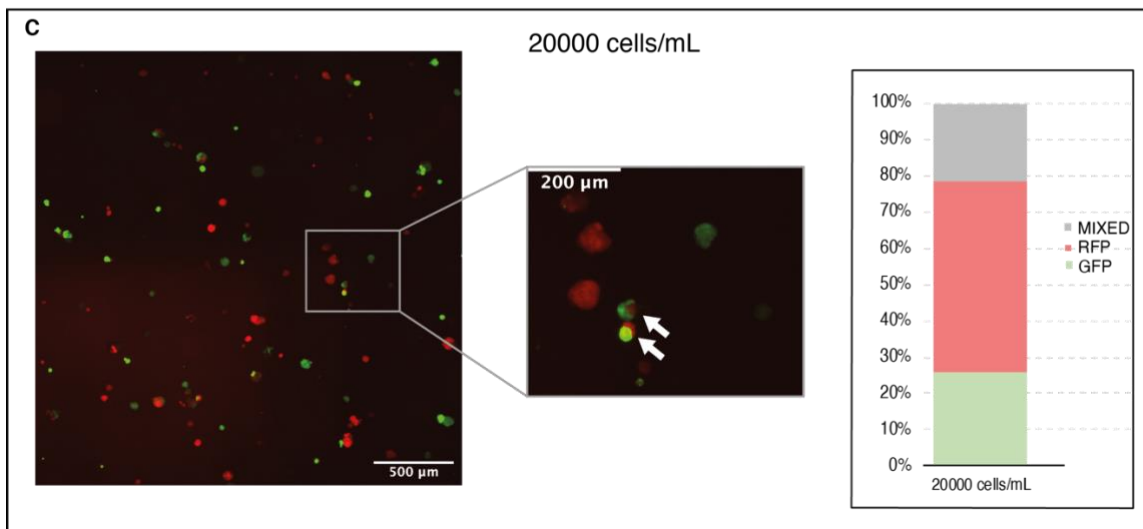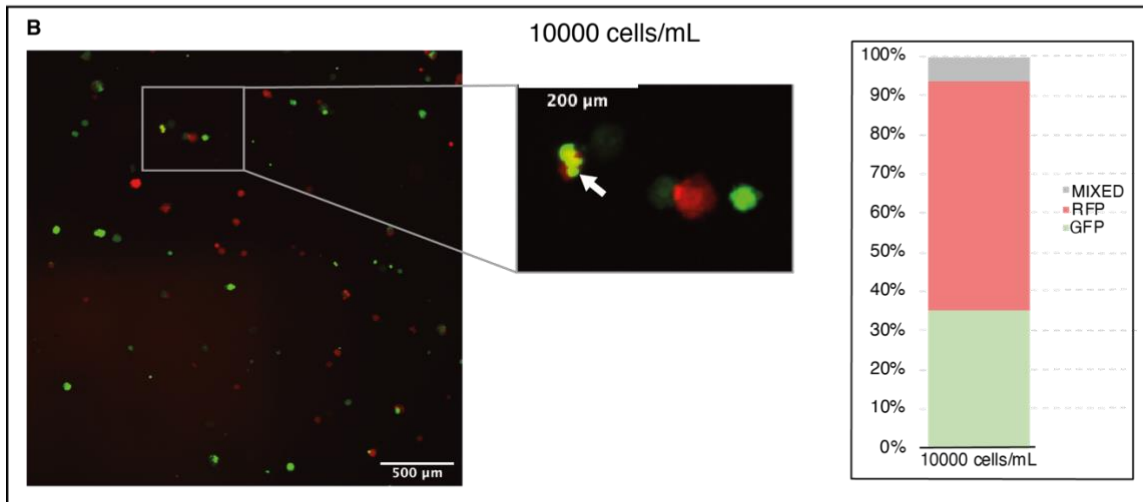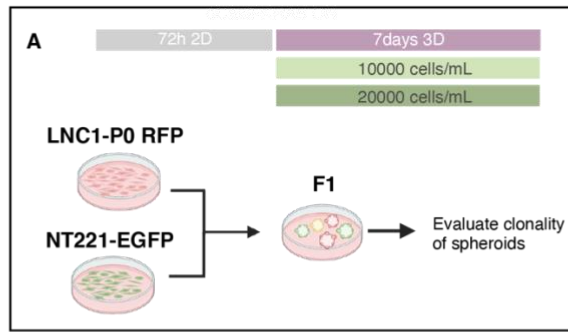
Figure 57: Evaluation of clonality of spheroids – A) Outline of the experiment; **B)** Left: Spheroids seeded at the density of 10000 cells/mL. The magnification is showing monoclonal or mixed spheroids (arrowhead). Right: Quantification of RFP+, GFP+ and mixed spheroids. **C)** Left: Spheroids seeded at the density of 20000 cells/mL, as in B. Pictures acquired by Nikon SMZ25 stereomicroscope.

2.5.3 Quality controls of 3D growth screening

Distribution of non-targeting guides

Similarly to the 2D screening, the distribution of non-targeting controls provides an indication that the *library coverage* has been preserved in this screening. We calculated the Log2FC of non-targeting guides in pooled samples compared to the LNC1-P0 and we observed that NTC guides are distributed with median values proximal to zero (P0'=0.01; F1=0.01, F2=0.025; F3=-0.02; F4=-0.035). When compared to the 2D proliferation screening, we observed limited but increased variability across generations. This is indicative of the stronger selective pressure.
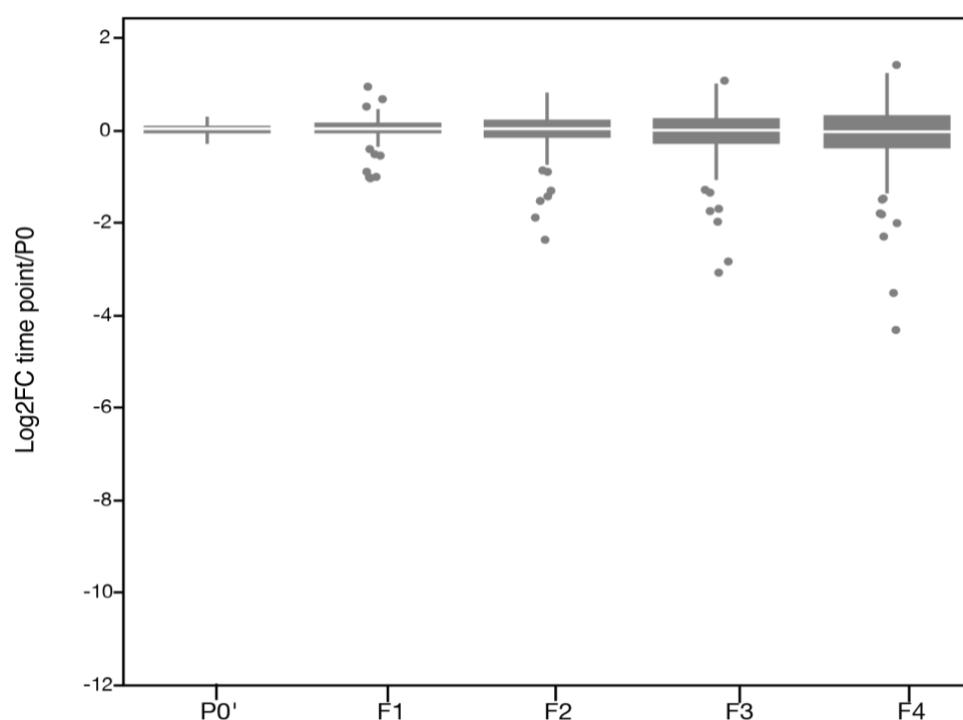
Figure 58: NTC distribution in pools of 3D screening – Box plot showing the distribution of 236 non-targeting guides at P0' (72h Doxycycline treatment in 2D) and for subsequent generation of 3D spheroid (F1, F2, F3, F4).

Distribution of guides targeting essential genes

As previously described, the 115 guides targeting essential genes represent an internal control for the activity of the dCas9-KRAB. As for the non-targeting controls, we calculated the log2FC for this sgRNA module at all time points of the screening. We saw that, as expected, the median log2FC decreases through generations (P0'=0; F1= -0.87, F2= -1.92; F3= -2.09; F4= -2.16).
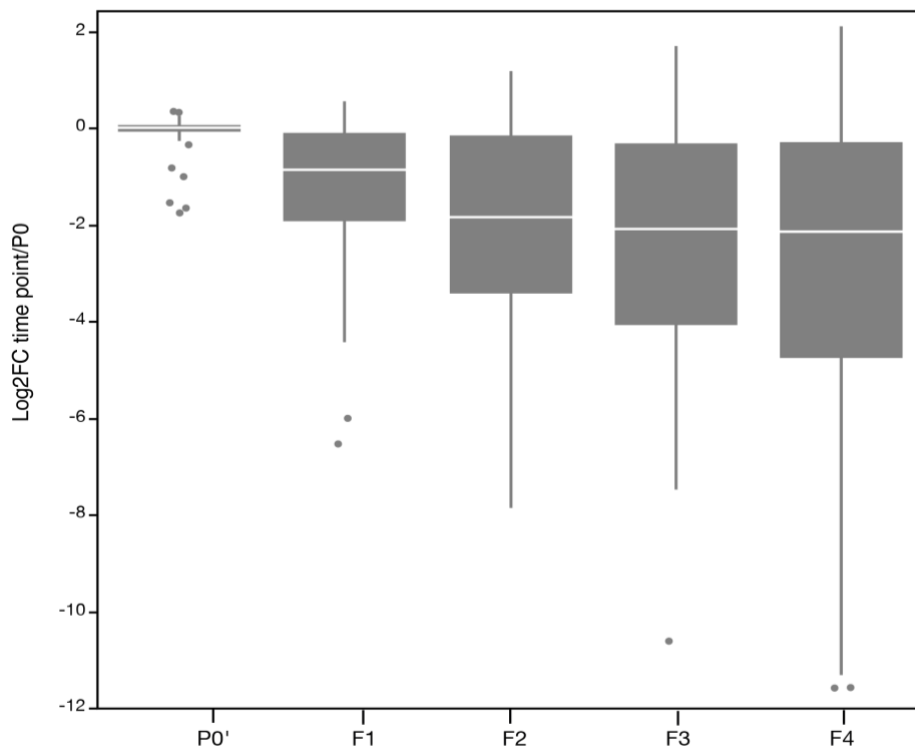
Figure 59: Distribution of guides targeting Essential Genes in 3D screening – Box plot showing progressive depletion of 115 guides targeting essential genes at P0', F1, F2, F3, F4.

2.5.4 Analysis of the 3D screening by MAGeCK

We applied MAGeCK to compute the β-score for each gene through the screening in pooled samples. We selected genes with significant β-score (p-value <0.05) as DROP-OUT or DROP-IN hits (**Figure 60**). We identified as DROP-OUT 17 of the 23 "Essential genes". While 14 of those genes were common hits also in the 2D screening, 3 were specific of the 3D setting (**Figure 61A**). Importantly, these 3D specific hits included STAT3 and MTHFD1, which were previously reported to have a role in stem-cell biology (Tordonato et al., 2021; Yu et al., 2014). These two genes show a sensible shift in rank from 2D to 3D (**Figure 60C e D**). Furthermore, we demonstrated that the KD by CRISPRi of STAT3 in SUM159PT caused a strong significative reduction of the SFE (**Figure 60E**). Overall, these data are indicative that the 3D screening can identify relevant genes that selectively control such kind of growth. Among the lncRNAs genes we identified 71 lncRNAs hits (25% of the tested lncRNAs): 48 lncRNAs were DROP-OUTs and 23 DROP-INs. 26 DROP-OUTs and 3 DROP-INs were in common with the 2D screening (**Figure 61B**).
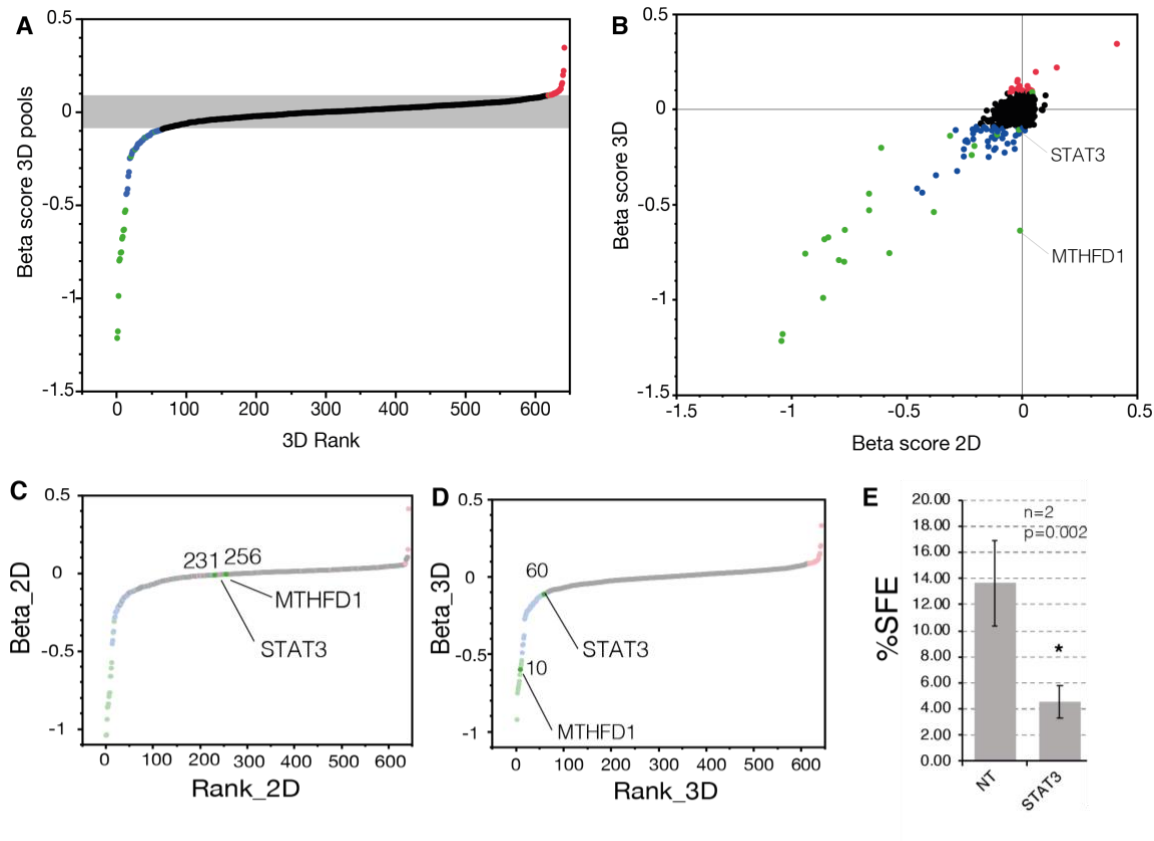
Figure 60: MAGeCK analysis of pools of the 3D growth screening – A) Ranked representation of β-scores. Grey range identifies non-significant β-scores. **B)** Scatter plot of 2D screening β-scores and 3D screening β-scores. **C)** Ranked representation 2D β-scores. **D)** Ranked representation of 3D β-scores. The position of MTHFD1 and STAT3 is highlighted. **E)** Reduced SFE in cells KD for STAT3. (Error Bars = standard error, N=2, p-value by Student's unpaired t-test). Figure legend: in blue significant lncRNAs DROP-OUTs; in red significant lncRNAs DROP-INs. Essential genes DROP-OUTs in green.
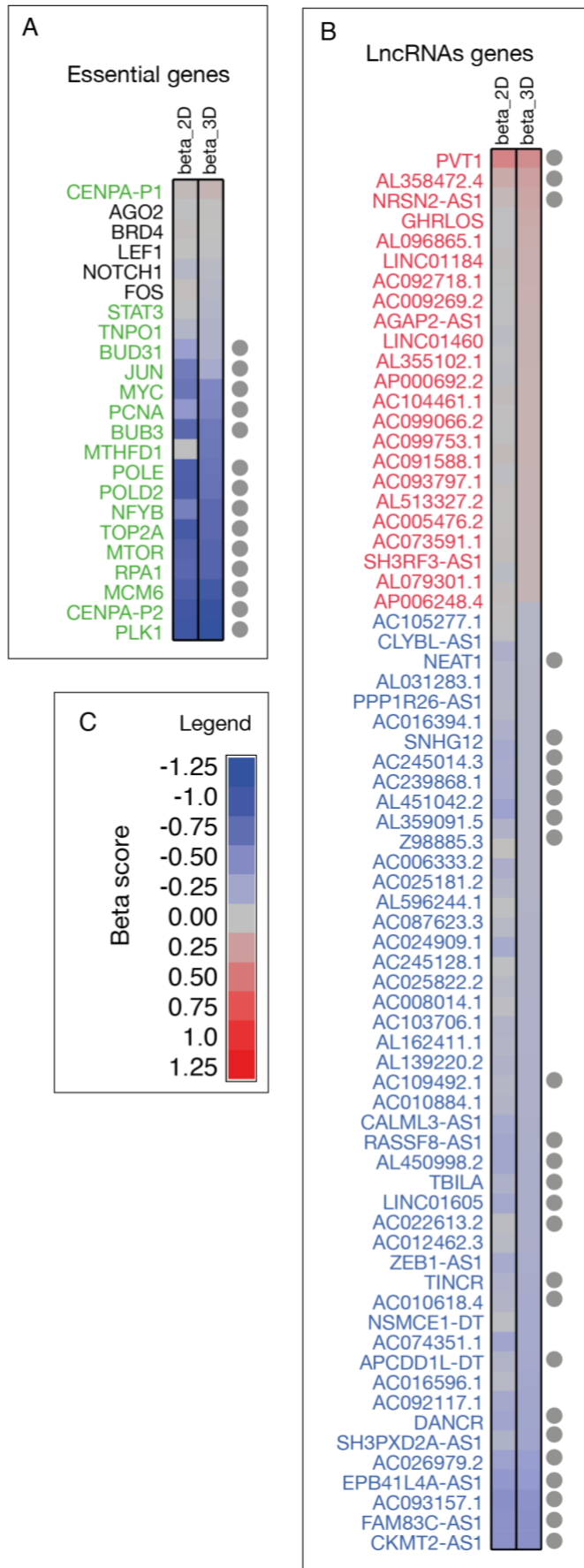
Figure 61: Cell plot of β-scores in 2D and 3D of 3D screening hits – A) β-scores for all essential genes. **B)** β-scores for lncRNAs (only hits). **C)** Figure legend. Genes are ordered from highest to lowest 3D β-score. Grey circles indicate common hits in 2D and 3D screening.

2.5.5 Sub-pool analysis of the 3D screening

The 3D screening was performed with a total of 18M cells, plated in 36 different dishes. This setting is laborious and demanding in terms of time and cost. To gain additional indications about the minimum library coverage that is required in order to perform a 3D screening, we isolated and sequenced the gDNA also from sub-pools of 3M cells (1/6 of the original coverage; sequencing strategy summarized in **Figure 62**, see **Methods**). An insufficient coverage would result in stochastic drop in complexity in the sub-pools, with poor reproducibility of the results obtained from sub-pools. When we analyzed the correlation of pools and sub-pools, as log2FC Vs P0 cells (**Figure 63**), we observed a strong correlation (average R^2 value = 0.77). Furthermore, the NTC guides were similarly distributed around zero either in pools or sub-pools. A multivariate analysis of the correlations between all the 3D screening samples clearly shows a high degree of similarity between samples of the same time point, regardless to the different coverage (**Figure 64**). Moreover, this plot gives some information about the biology of the system. Indeed, we can appreciate that the greatest differences are established from P0 to F1 and from F1 to F2. Samples of F3 and F4 resemble F2. In conclusion, for future 3D screenings, we can conclude that *i*) *library coverage* would be guaranteed also by processing samples derived from the plating of 3M cells and *ii*) most of the biological conclusions would be acquired by propagating spheroids for only two generations (F1, F2).
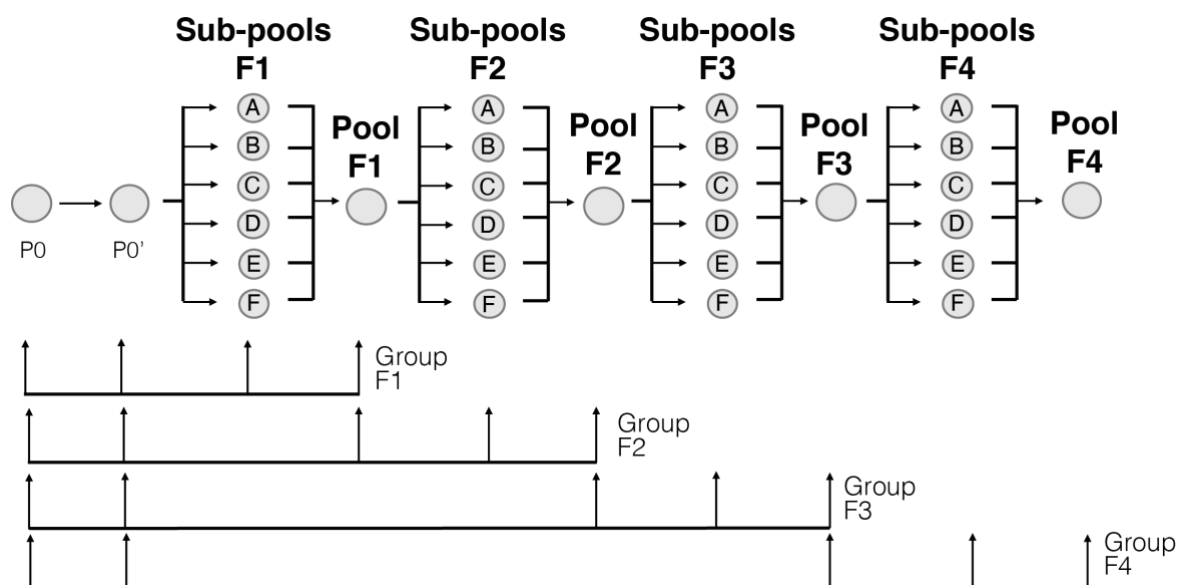
Figure 62: Sequencing strategy of pools and sub-pools of 3D screening – The samples of the 3D growth screening were sequenced in four pools. In each pool we sequenced P0 and pooled samples from previous generation to allow comparisons.
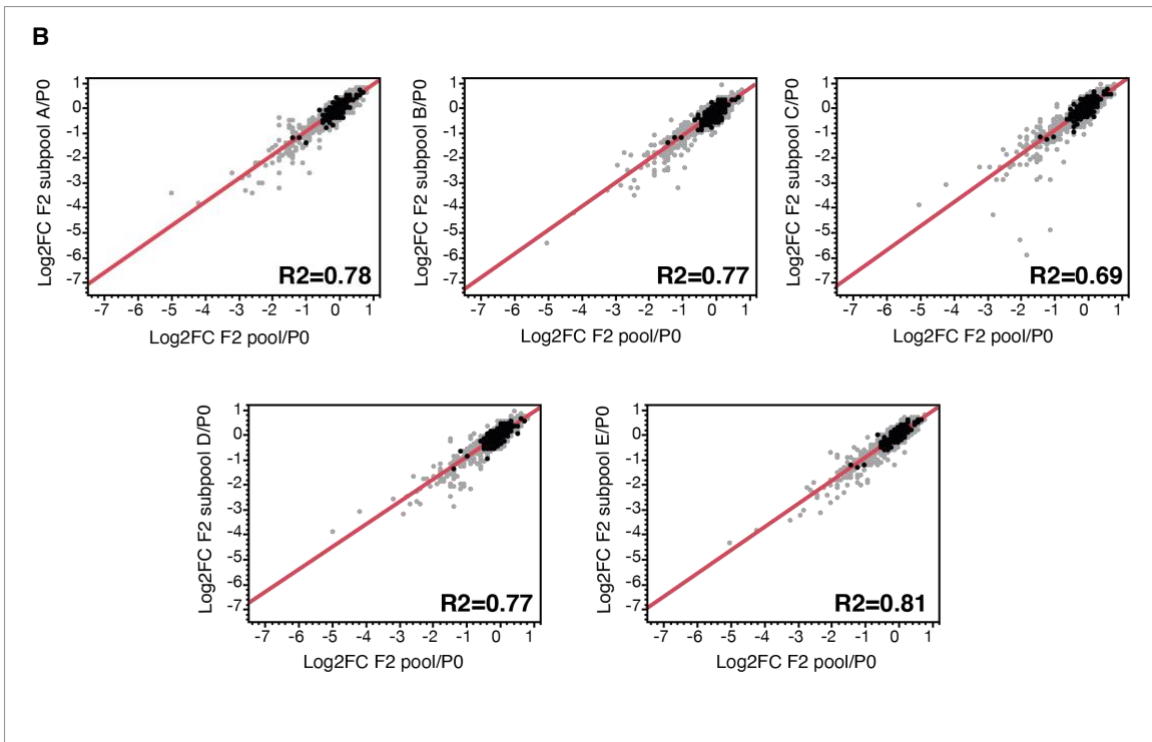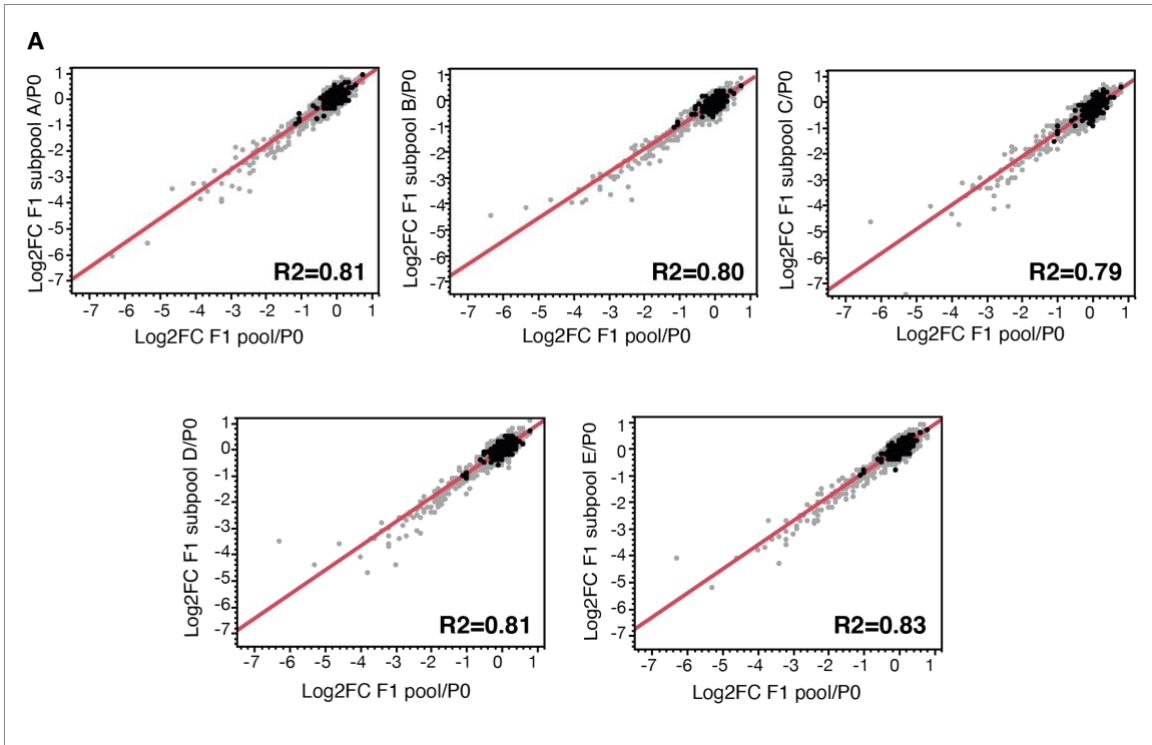
Figure 63: Distribution of NTCs in subpools vs pooled samples in F1 and F2 – Scatterplot showing that NTCs (black dots) are similarly distributed in pools and subpools with Log2FC values around zero in F1 and F2 samples.
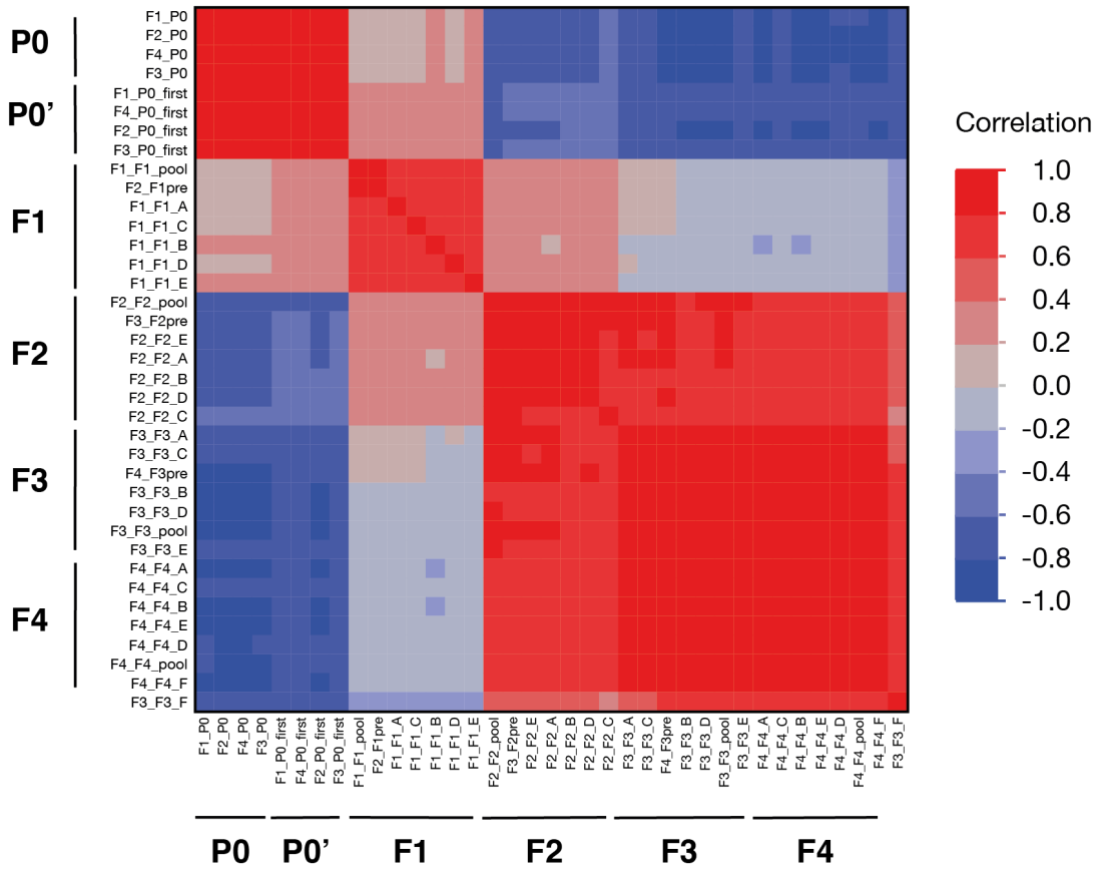
Figure 64: Clustered correlations on 3D screening samples – The heatmap displays the clustered correlations between all 3D samples, as estimated by the Row-wise method (JMP).

2.6 Paclitaxel screening

2.6.1 Outline of the Paclitaxel screening
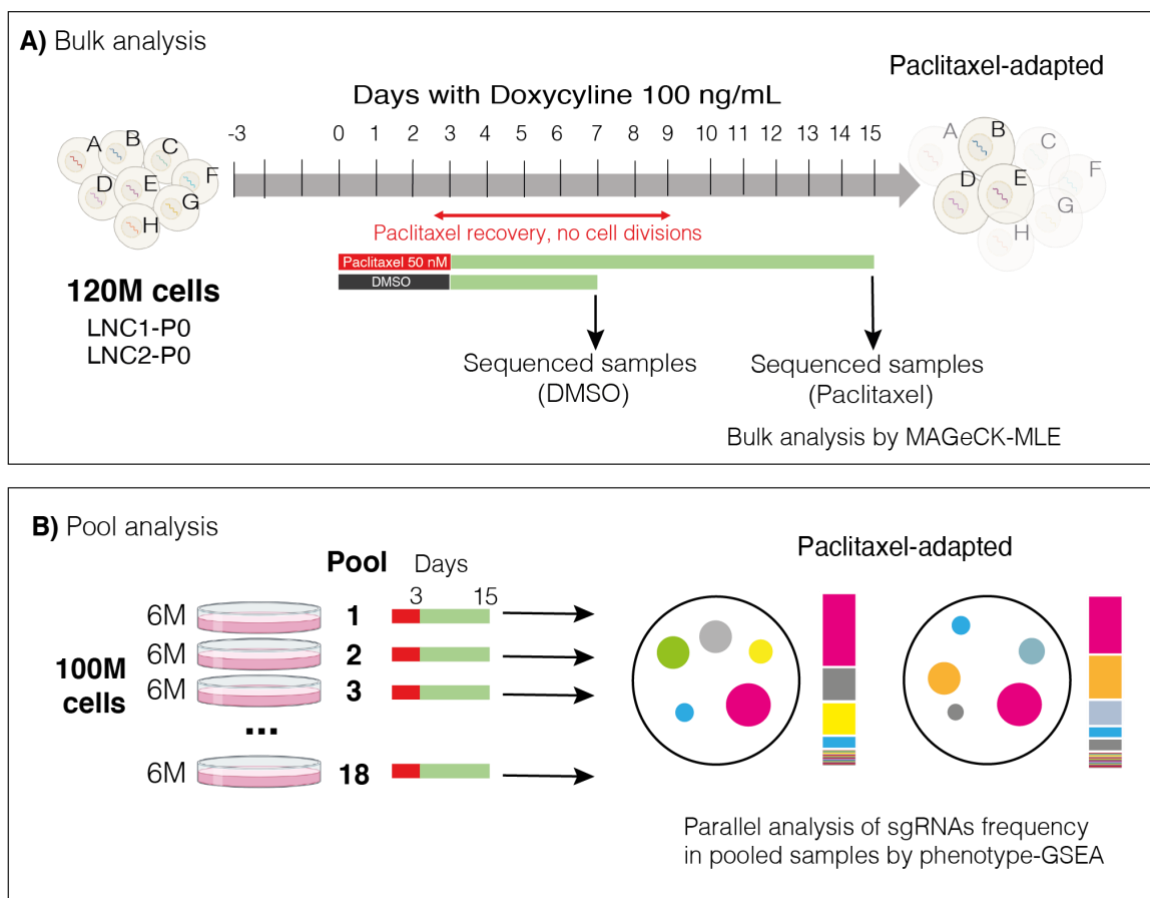
Figure 65: Outline of the Paclitaxel screening – A) Bulk analysis of the Paclitaxel screening by MAGeCK-MLE. **B)** Parallel analysis of clonal composition from 18 individual samples of 6×10^6 cells treated with the drug.

Paclitaxel is a drug used in neoadjuvant chemotherapy that impairs mitotic spindle assembly and blocks cells in G2-M. SUM159PT cells are sensitive to this drug, with an IC50 value of 8nM (see **2.2.3**). We performed a drug screening with both P0s of the LNC-library in order to isolate modifiers of the response after treatment. The standard analysis of replicate screenings highlighted an intrinsic noise in the results, due to the strong biological selection applied which led to a stochastic loss of library complexity (See **2.6.3**). For this reason, we also developed a new analytical strategy based on parallelized observations of sgRNAs occurrence in multiple independent samples at low coverage, coupled with a novel analytical framework inspired by Gene-Set Enrichment Analysis (GSEA), in which the ranked observations of sgRNAs frequencies in each sample are used to determine enriched and depleted guides (See **2.6.5**).

2.6.2 Set-up of the screening

As described in (See **2.1.3**), we previously modeled the response to Paclitaxel of SUM159PT. In this screening, we pre-activated the CRISPRi system for 72h and then, we treated cells acutely with Paclitaxel (IC97=50nM) or DMSO (See **Methods**) for 72h and waited for the emergence of drug-tolerant colonies.

In the set-up of the Paclitaxel screening, we need to take into consideration the strong drop of complexity induced by the drug (if 3% of cells survive, then the expected drop is ~33X). Therefore, in order to preserve the *library coverage,* we increased the number of cells treated to 120M cells, factoring the drop due to cell death (3.5M X 33 = 115.5M) and in theory preserving a 1000X *library coverage.* We checked the expression of the dCas9-KRAB at the end of the experiment and observed that, despite the strong treatment, LNC1 and LNC2 cells were still expressing the transgene even if at a slightly lower levels compared to the respective P0 after 72h of Doxycyline treatment.
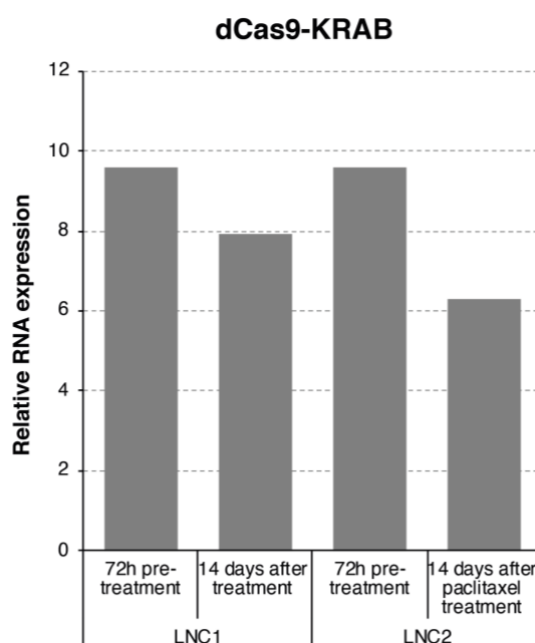
Figure 66: Expression of dCas9-KRAB in Paclitaxel screening – Expression of dCas9-KRAB in the LNC1 and LNC2 cells after 72 hours treatment with 100 ng/mL doxycycline before the drug-treatment and at the end of the screening.

2.6.3 Reproducibility of the Paclitaxel screening
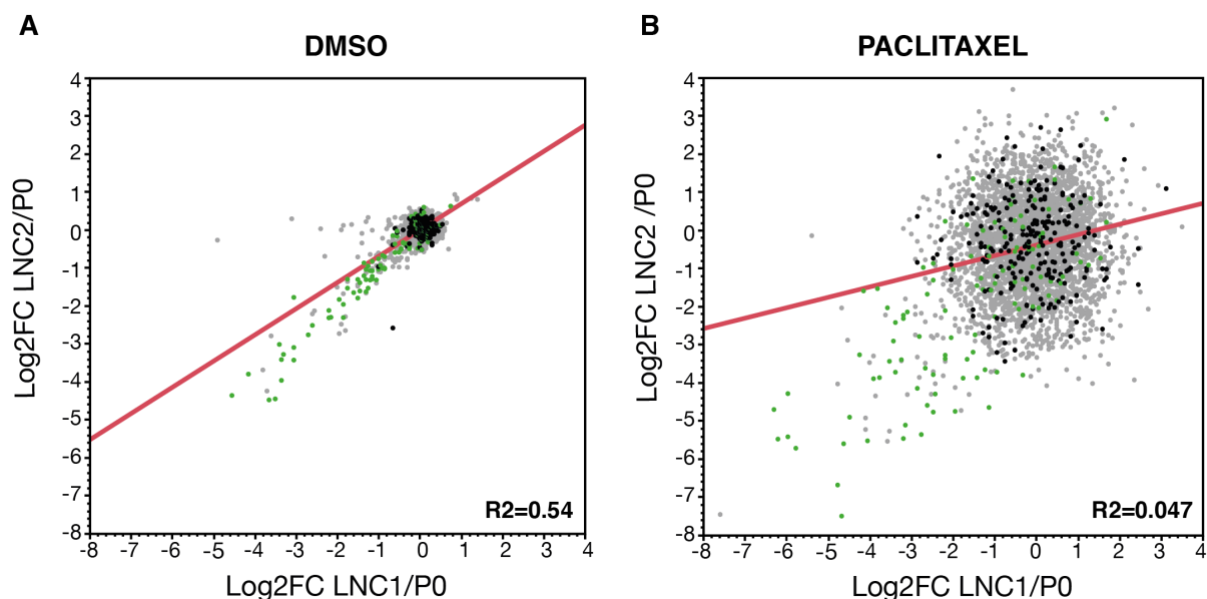
Figure 67: Reproducibility of the Paclitaxel Screening – Scatter plot of the Log2FC of **A)** DMSO and **B)** Paclitaxel samples for both LNC1 and LNC2 biological replicates. Figure legend: Black dots=NTC guides; green dots=sgRNAs targeting essential genes; grey dots=sgRNAs targeting lncRNA genes.

We compared the log2FC for both biological replicates (with LNC1 and LNC2 cells) of the two branches of the experiment (DMSO and Paclitaxel). We observed that in DMSO-treated cells, the sgRNAs behaved as expected from previous screenings, with NTC guides tightly distributed with median log2FC values close to zero (LNC1=0.05; LNC2=0.07). Conversely, in Paclitaxel-treated cells, we observed that the NTC were more dispersed and less consistently distributed in the two biological replicates, with a negative median Log2FC values (LNC1= -0.25; LNC2= -0.47) (**Figure 68**). This is symptomatic of a noisier system due to the strong selection applied and occurred regardless of our attempt to increase the coverage. We observed that, despite the noise, several guides were coherently depleted in the two replicates. Most of these guides targeted essential genes that indeed resulted depleted as a category (Median Log2FC LNC1= -1,5; LNC2= -1.75). This observation is indicative of an overall integrity of the CRISPRi system (**Figure 69**).
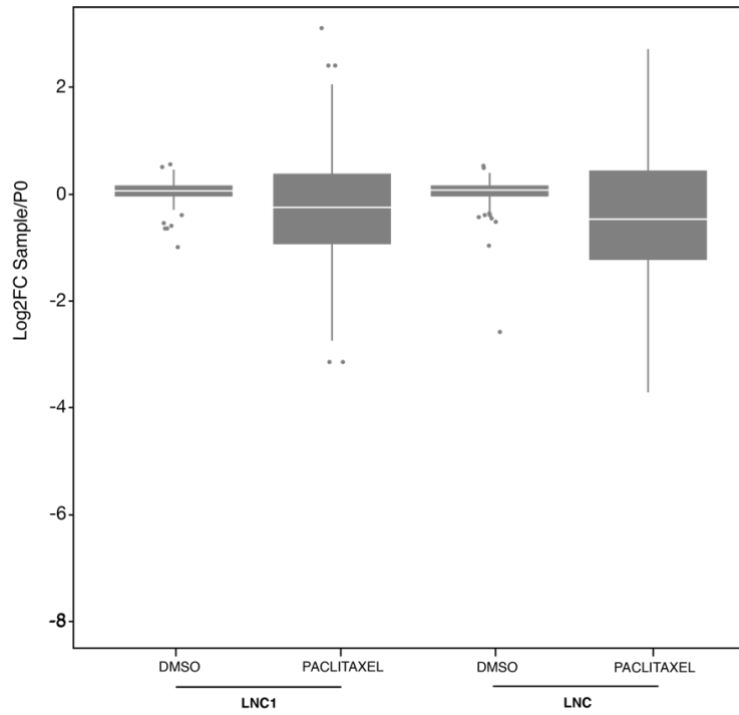
Figure 68: Distribution of NTC in Paclitaxel screening – Boxplot showing distribution of 236 NTC guides in both biological replicates. The paclitaxel-treated cells show larger dispersion compared to DMSO-treated cells.
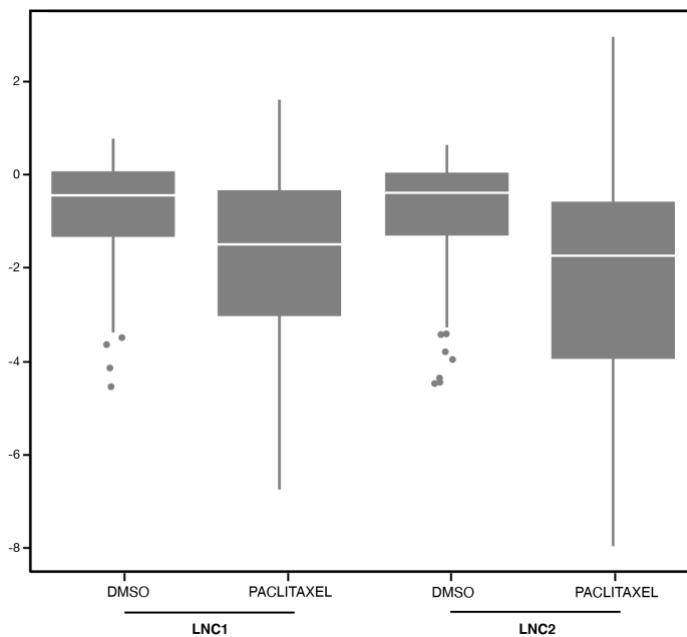
Figure 69: Distribution of sgRNAs targeting Essential Genes in Paclitaxel Screening – Distribution of 115 sgRNAs targeting the "essential genes" module in both biological replicates of the paclitaxel screening.

2.6.4 MAGeCK analysis of the Paclitaxel screening

We applied MAGeCK to model the gene-level effect of the perturbation of the genes in the LNC-library upon Paclitaxel treatment and called as hits those genes with significant p-value ($p<0.05$). The tool identified as depleted 13 of the 23 Essential genes called as hits in the 2D screening. Moreover, we identified 8 lncRNAs genes as DROP-INs and 26 lncRNAs genes as DROP-OUTs. Among the DROP-INs, the gene with the highest β-score was PVT1, further validating the pro-proliferative effect of the perturbation of this gene. Among the DROP-OUTs 10 were previously called as HITs in the 2D screening (**Figure 71**).
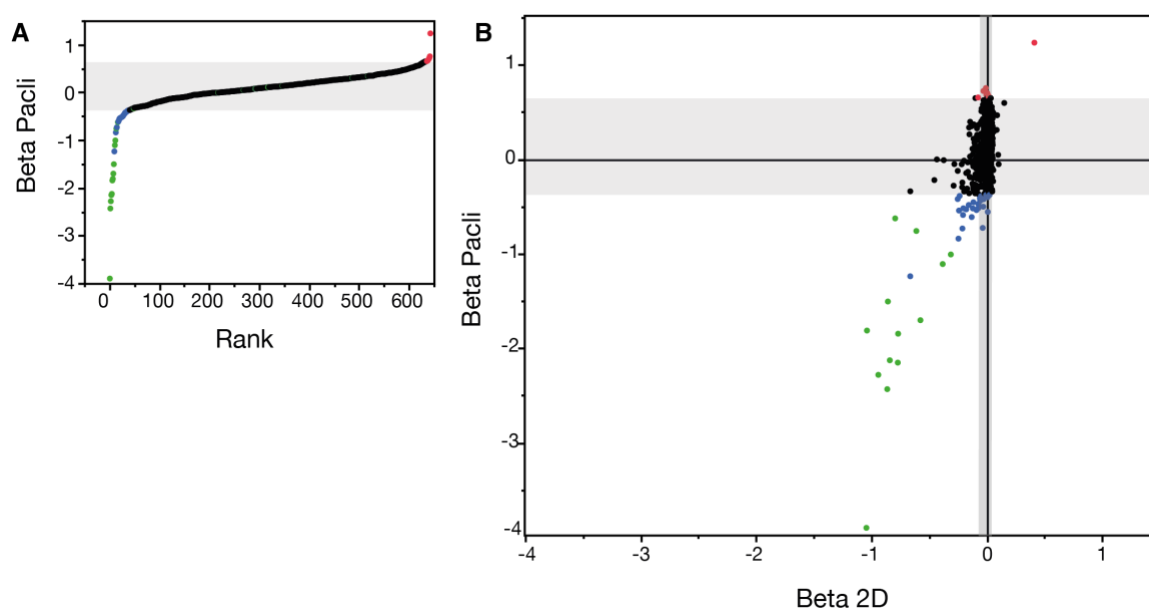
Figure 70: β-scores of Paclitaxel screening by MAGeCK – A) Ranked representation of β-scores for all genes in the LNC-library. **B)** Scatter plot of β-scores for 2D and paclitaxel screening for all genes in the LNC-library. Figure legend: green dots= all essential genes; Red dots= significant lncRNAs DROP-INs. Blue dots: Significant lncRNAs DROP-OUTs. Gray area identifies ranges of non-significant β-scores (p-value>0.05)
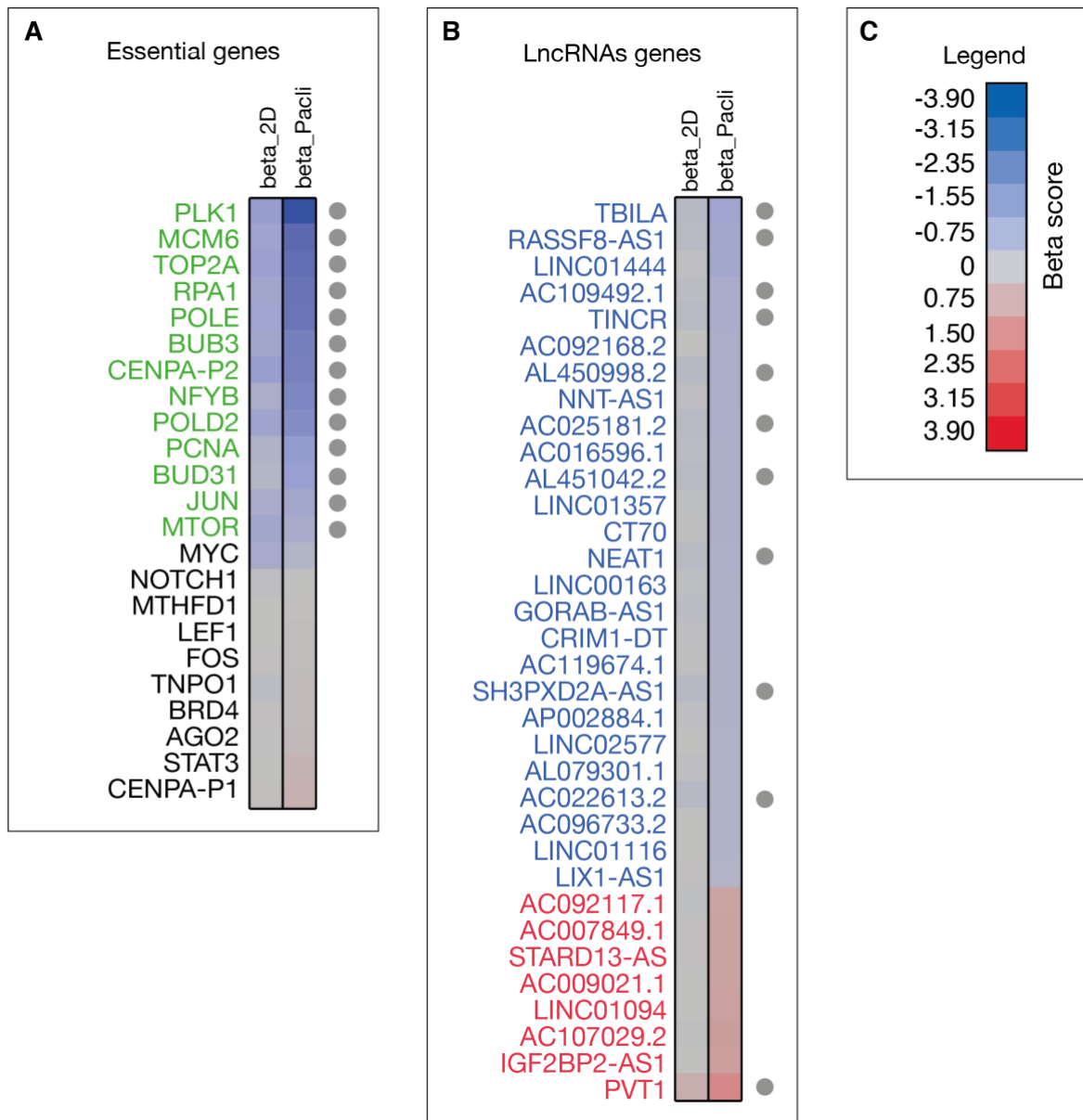
Figure 71: Cell plot of β-score values of Paclitaxel screening hits in 2D and Paclitaxel screening - A) β-scores for all essential genes (in green essential genes with significant β-score). **B)** β-scores for lncRNAs (hits only) (in blue DROP-OUT lncRNAs, in red DROP-INs). **C)** Figure legend. Genes are ordered from lowest to highest β-score in Paclitaxel screening. Grey circles indicate common hits in 2D and 3D screening.

2.6.5 Increasing the resolution of the screening

Altough MAGeCK identified genes with significant β-scores, we wondered if we approached this screening with the right perspective. Indeed, the trend of the NTCs highlighted that we have stochastically lost some *library coverage*. The computation of β-scores by MAGeCK starts from the raw counts of sgRNAs: the growth of colonies is a clonal phenotype (see **2.8.6**) and therefore the final frequency of sgRNAs is strongly influenced by their association with large monoclonal colonies. Thus, even if 3% of cells survives, only a fraction of these cells enters a highly proliferative state while other cells survive, but, at least initially, they display a low proliferation rate. In conclusion, the results obtained from pooling cells from many dishes are biased by the differential proliferative capacity of the colonies (**Figure 72**). We reasoned that, to properly dissect the biological outcome of the perturbations, we could *i*) increase the number of statistical observations and *ii*) apply an analytical framework that is not based on sgRNAs frequency but on sgRNAs ranking. To this aim, we repeated this screening with the LNC1-P0 but this time, we divided the samples in 18 pools of 6M cells that were independently processed during treatment and production of gDNA libraries. We observed that the frequency of sgRNAs in the pools closely mirrored the number of large clonal colonies, with an average of 227 highly frequent (dominant) sgRNAs (SD=±13.5) (>1000CPM) (**Figure 73A**), that accounted for 40% of the total normalized reads. The remaining reads were assigned to >1000 guides that were associated with small colonies (**Figure 73B**). To limit this issue, we ranked the sgRNAs by their frequency in each of the 18 separate pools and then we merged these ranked lists in a unified overall list, which contained each sgRNA a number of times equal to the number of experimental observations (in this case 18). Then we performed the analysis with a GSEA approach (Subramanian et al., 2005) ("Phenotype-GSEA", see **Methods**) in which "gene sets", instead of being composed of different genes, are composed of sgRNAs targeting the same gene. In summary, per each gene, we created a "gene-set" composed of 90 observations (5 sgRNAs per each gene*18 samples= 90 Gene Set length). We called as DROP-INs and DROP-OUTs the genes showing a Normalized Enrichment Score (NES) of respectively >2 or < -2 and we further filtered the hits based on a 5% FDR, as calculated by the GSEA algorithm. We also calculated an empirical FDR based on 46 randomly created gene-sets of 5 NTCs of which only 2 fulfilled the criteria for DROP-OUTs and 0 those for DROP-INs (see **Methods**).
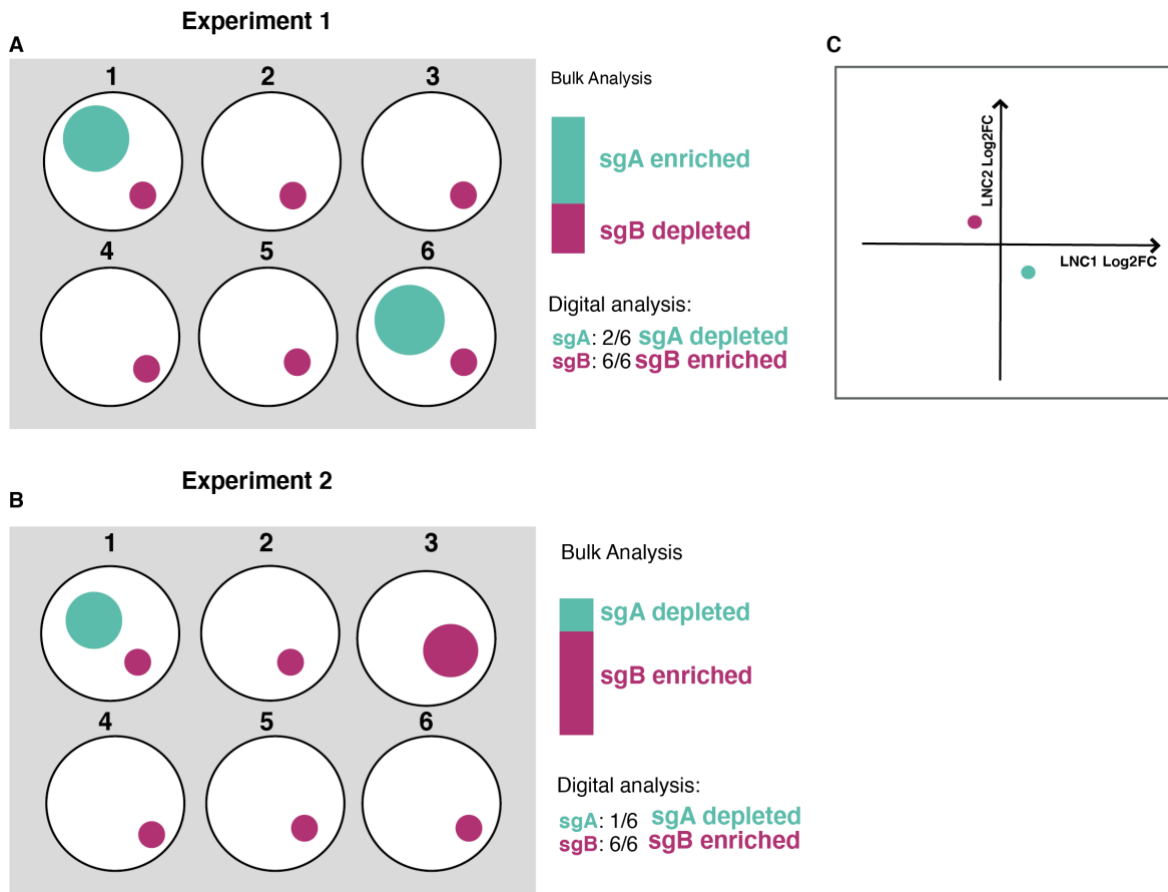
Figure 72: Influence of large colonies on the outcome of the screening – Scheme explaining a possible problem when performing the Paclitaxel Screening in bulk. **A)** and **B)** are representative cartoons of clonal growth for cells carrying 2 sgRNAs. In **A)** sgA grows in two large colonies in plates 1 and 6. SgB grows in small colonies in all plates. In **B)** sgA is present in a large colony only in plate 1, while sgB proliferates in small and medium size colonies in all plates. **C)** Scatter plot representing the log2FC of the two sgRNAs, compared to the P0, in the bulk analysis of the experiments pictured in **A)** and **B)**. The bulk analysis of the two experiments would lead to different and discordant results among the two experiments, due to the leading effect of the larger sgA colonies in the experiment 1. If the results of the experiment are expressed digitally, as presence/absence of colonies per each sgRNA, the phenotype of the guides is clearer and more reproducible among the different experiments with sgA depleted and sgB enriched in all samples.

CPM	P0	pool1	pool2	pool3	pool4	pool5	pool6	pool7	pool8	pool9	pool10	pool11	pool12	pool13	pool14	pool15	pool16	pool17	pool18
>1000	1	203	218	233	223	241	221	241	252	226	245	219	243	213	223	206	231	227	225
>100	3432	1325	1315	1298	1246	1333	1195	1329	1475	1248	1360	1361	1334	1258	1285	1271	1378	1418	1292
>10	3443	3077	3031	3015	3007	3029	2901	3095	3292	3249	3155	3158	3090	3014	3077	3116	3084	3119	3114

Figure 73: Clonal composition of pools upon Paclitaxel treatment – A) Bar chart showing the number of detected sgRNAs at >10 CPM; >100 CPM; >1000 CPM (CPM, copies per million reads). **B)** Percentage contribution of each sgRNA class to the total read count. **C)** Cumulative distribution of sgRNAs frequency in P0 and pools. All these plots show that there is a selection of clones that, from being equally distributed in the P0, become strongly skewed after treatment.

2.6.6 Analysis of screening hits by phenotype-GSEA

For each gene, the GSEA software produces enrichment plots that illustrate the profile of the "genes" (in our case, "sgRNAs") in the set (**Figure 74 B, C, D**).

For instance, considering the enrichment plot of the essential gene PLK1 (**Figure 74B**), a negative ES was observed, with the leading edge sgRNAs skewed towards the bottom of the ranked list. This means that the sgRNAs for PLK1 were positioned low in the ranked lists of each experimental pool. Similar conclusions can be driven for the lncRNA DROP-OUT TBILA (**Figure 74C**). Conversely, the lncRNA gene LINC01194 was the gene with the highest positive NES and displayed a leading-edge distribution of the sgRNAs shifted towards the top of the ranked list.

Overall, the phenotype-GSEA analysis led to the identification of 52 HIT genes, 18 DROP-INs and 34 DROP-OUTs (**Figure 74A**). Although very intuitive and powerful, a limitation of this this analysis is that it does not take into consideration the distribution of the sgRNAs in the LNC1-P0. Even if we clearly showed that P0s are very well balanced (See **2.3.5**), yet there are a few sgRNAs that are under-represented from the beginning. To monitor the possible confounding action played by the starting representation, we measured the average initial representation of the sgRNAs for each gene and compared this value with the NES (**Figure 75**). We observed that for a few hits (either DROP-INs or DROP-OUTs) the behavior might be ascribable to their representation in the P0 (respectively higher and lower than the average). It is worth mentioning that the top-represented gene is only 2-fold more represented than the lowest-represented gene. In any case, we suggest that the interpretation of these hit genes should be carefully evaluated.

Overall, we identified 22 DROP-OUT hits and 16 DROP-INs among the candidates set. Among the DROP-OUT genes, 7 were similarly identified by the initial MAGeCK analysis (See **2.6.4**). Of the essential gene set, 11 genes were called as DROP-OUTs. Within the essential genes set, the behaviour of MYC dragged our attention. MYC has been called as DROP-OUT in the 2D and 3D screening while, despite a negative β-score, it failed to reach a significative level in the MAGeCK analysis. On the same line the lncRNA PVT1, negative regulator of MYC expression, has been called as DROP-IN in 2D, 3D and Pacli-MAGeCK analysis. With the phenotype-GSEA analysis, MYC and PVT1 displayed an opposite behavior, with MYC behaving as a DROP-IN and PVT1 as a DROP-OUT. This result might be suggestive of the capacity of the Phenotype-GSEA approach to capture different biological phenomena. In fact, MYC suppression has been recently observed in chemo-persister cells, which survive by reducing redox stress and apoptotic priming (Dhimolea et al., 2021).
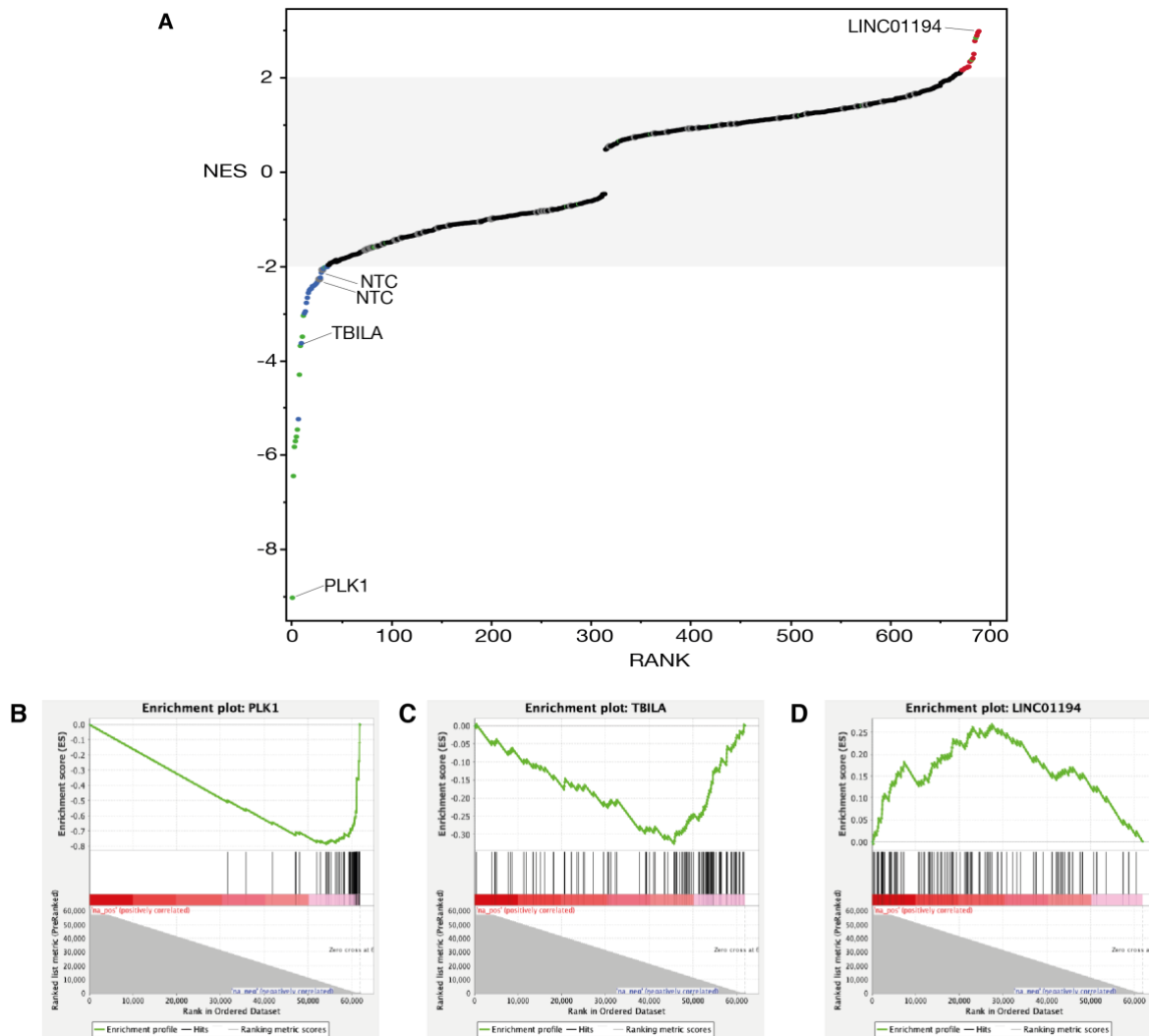
Figure 74: Normalized Enrichment Scores (NES) of Paclitaxel screening –A) Ranked representation of NES for all genes in the LNC-library (643) and 46 sets of NTC. In green highlighted essential genes. In blue DROP-OUT lncRNAs, in red DROP-IN lncRNAs. NTC sets in grey. Shaded area represents the area of non-significant NES. Two NTC sets are below the 5% FDR threshold. Enrichment plots for **B)** PLK1**; C)** TBILA**; D)** LINC01194. The upper part of the plot shows the running sum of the ranked list in which a positive value is given every time an sgRNA is over-represented and a negative value when the sgRNA is under-represented. The Enrichment Score (ES) is defined as the value most distant from zero. The bar below pictures the leading-edge sgRNAs namely the position in rank of the sgRNAs that contribute the most to the ES for that set. The NES is calculated from the actual ES normalized by mean ES calculated from 1000 permutations of the ranked list. The NES is the appropriate statistical measurement to compare analysis for different gene-sets (i.e. lncRNAs).
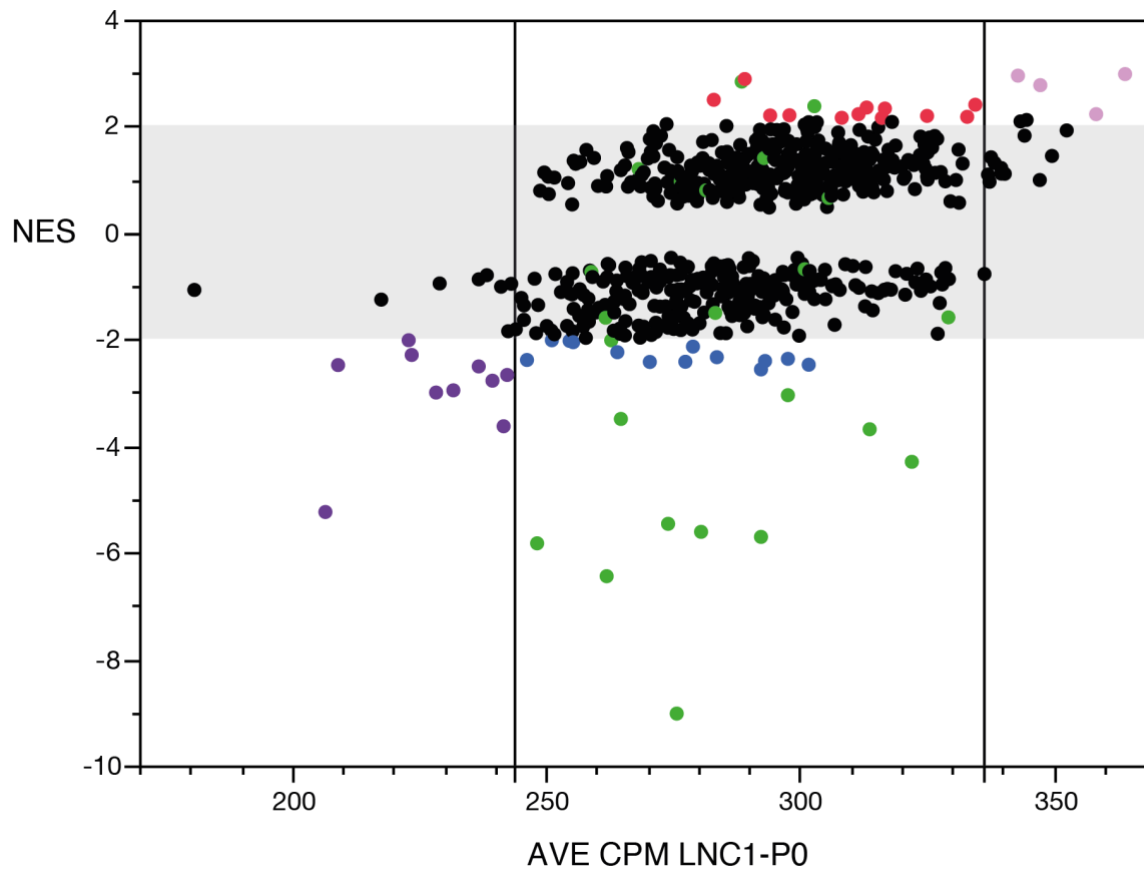
Figure 75: Impact of sgRNAs representation in LNC1-P0. Scatter plot showing the average gene-level representation of genes in the LNC-Library and relative NES value. Bars show 2.5-97.5 percentile of the distribution. Shaded area non-significant NES. Figure legend: Green= Essential Genes; Blue= lncRNAs DROP-OUT; Red= lncRNAs DROP-IN; Purple= lncRNAs DROP-OUT with low average CPM in the LNC1-P0. Pink= lncRNAs DROP-IN with high average CPM in the LNC1-P0.
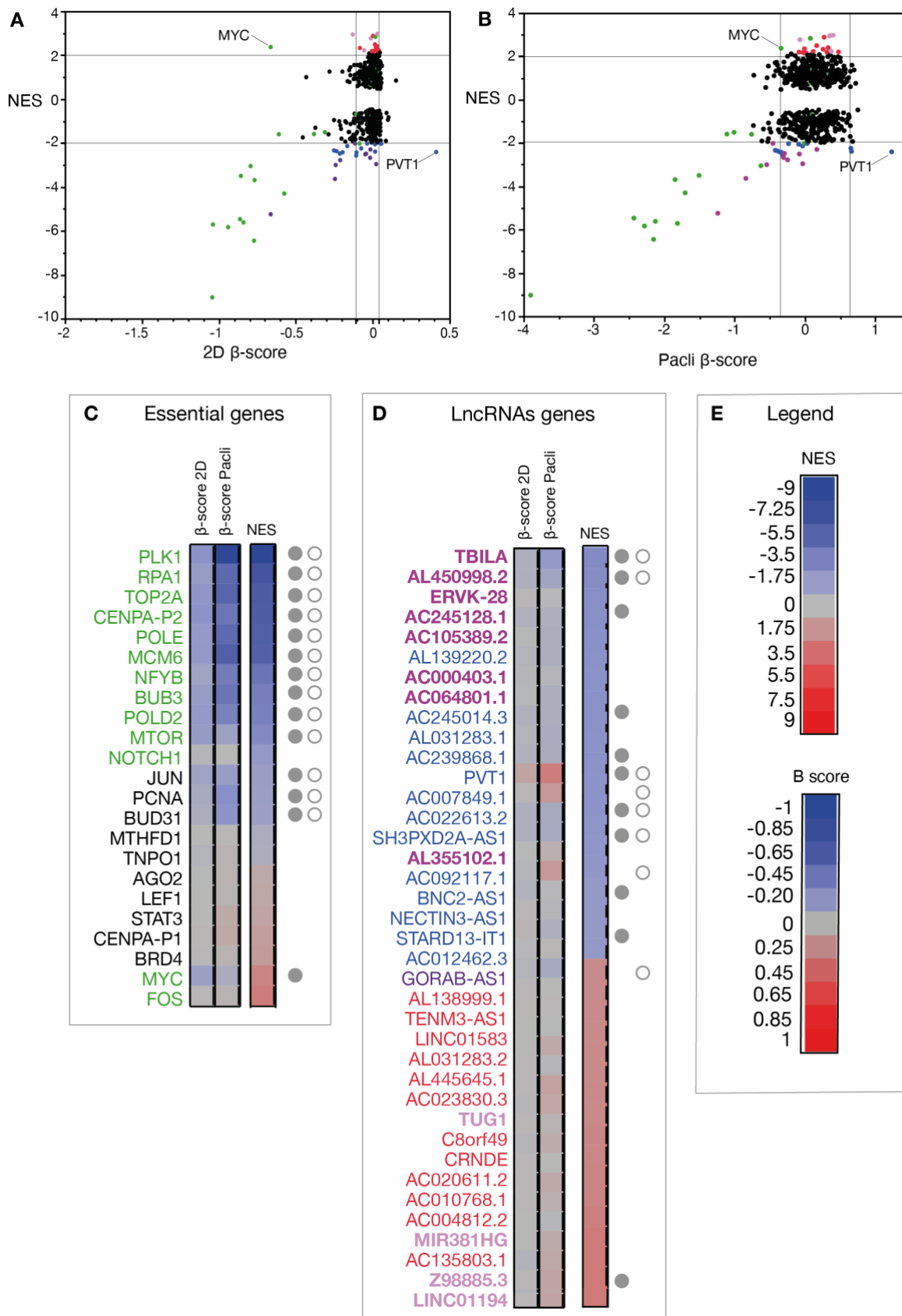
Figure 76: Hits by GSEA analysis – Representation of NES scores vs **A)** 2D β-scores or **B)** Paclitaxel β-scores. **C)** Comparison of β-scores and NES for **C)** all "Essential genes" set and **D)** only hits defined by phenotype-GSEA. Figure legend: green dots= essential genes; blue dots= lncRNAs DROP-OUT, in red= lncRNAs drop-in by phenotype-GSEA. Purple dots= DROP-OUT hits with low average CPM in the LNC1-P0. Pink= lncRNAs DROP-IN with high average CPM in the LNC1-P0.

2.7 *In vivo* screening

2.7.1 Outline of the *in vivo* screening
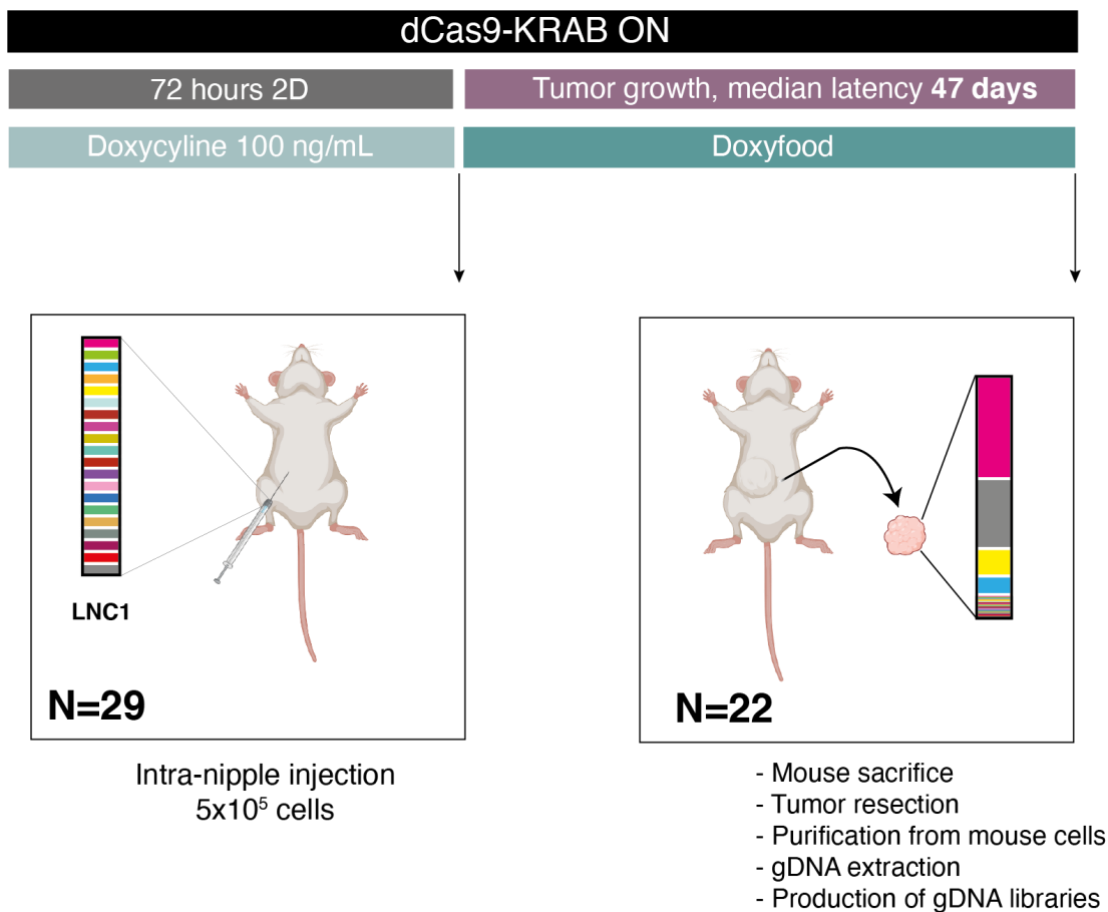
Figure 77: Outline of the *in vivo* screening – We injected NSG mice with 500000 cells of the LNC1-P0 after 72h doxycycline induction in 2D. We injected orthotopically 29 mice and we produced 22 libraries from the purified xenograft, representative of sgRNAs distribution in individual tumor.

In the last screening presented, we aimed at identifying lncRNAs whose perturbation impact the TIC properties of the cell line. To this purpose, we injected LNC1 cells intra-nipple in the mammary fat pad of NSG mice. This kind of xenograft more precisely recapitulates the location of the disease, and more accurately describes the challenges of growing in the mammary gland. We injected a total of 29 mice, and we manage to produce 22 representative gDNA libraries. We evaluated the drop of complexity of the system, and we computed DROP-OUTs and DROP-INs genes by evaluating their ranked representation in each tumor with the same phenotypic-GSEA approach presented for the Paclitaxel screening.

2.7.2 Analysis of clonal composition of tumors

SUM159PT is a tumorigenic cell-line, with the potential to generate solid tumors *in vivo*. This feature has been established in literature (Fillmore and Kuperwasser, 2008) and in the lab, the Tumor-Initiating Frequency of the cell line has been defined by limiting-dilution experiments (**Figure 78**).

	Dose	Response
CTRL	100000	3 /3
	10000	3 /3
	1000	9/12
	TIC-frequency (upper/lower)	1:721 (1: 356- 1:1461)

Figure 78: TIC Frequency of SUM159PT – TIC frequency of SUM159PT by limiting dilution experiments. Data from (Tordonato et al., 2021).

The limiting-dilution experiments address the capacity of a cellular population to form a tumor but gives no indication about the clonal composition of the mass. In the lab, we also previously established tracking approaches based on passive barcodes to define the clonality of solid tumors. We labelled SUM159PT with a barcode library of ~5000 barcodes (BCs) and transplanted intra-nipple 100.000 cells of the barcoded population in 6 mice. We waited for the development of palpable tumors of the maximum diameter of 1-1.2 cm, that grew with a median latency of 40 days. Mice were sacrificed and tumors purified from mice cells. We amplified the barcode cassette by PCR and sequenced the libraries. We normalized the total reads in CPM to describe the tumor as a "virtual mass" of the size of a million counts. In each tumor, we detected an average of 760 BCs (15.2% of the library). A subset of these BCs (average 100 BCs, ~3%) were classified as dominant barcodes (>1000CPM), accounting for most of the virtual tumor mass (**Figure 79B**). In particular, the TOP3 barcodes often occupied a sensible fraction of the virtual tumor mass (up to 60%).

Taken as a whole, these data show that tumor initiation process by SUM159PT cells is characterized by a strong selective pressure, with only few cells that can seed, proliferate, and contribute to the tumor mass. Stochastic events lead to the outgrowth of very large clones, in a process which is, in some way, reminiscent of that described for highly proliferative clones after paclitaxel (See **2.6.5**).
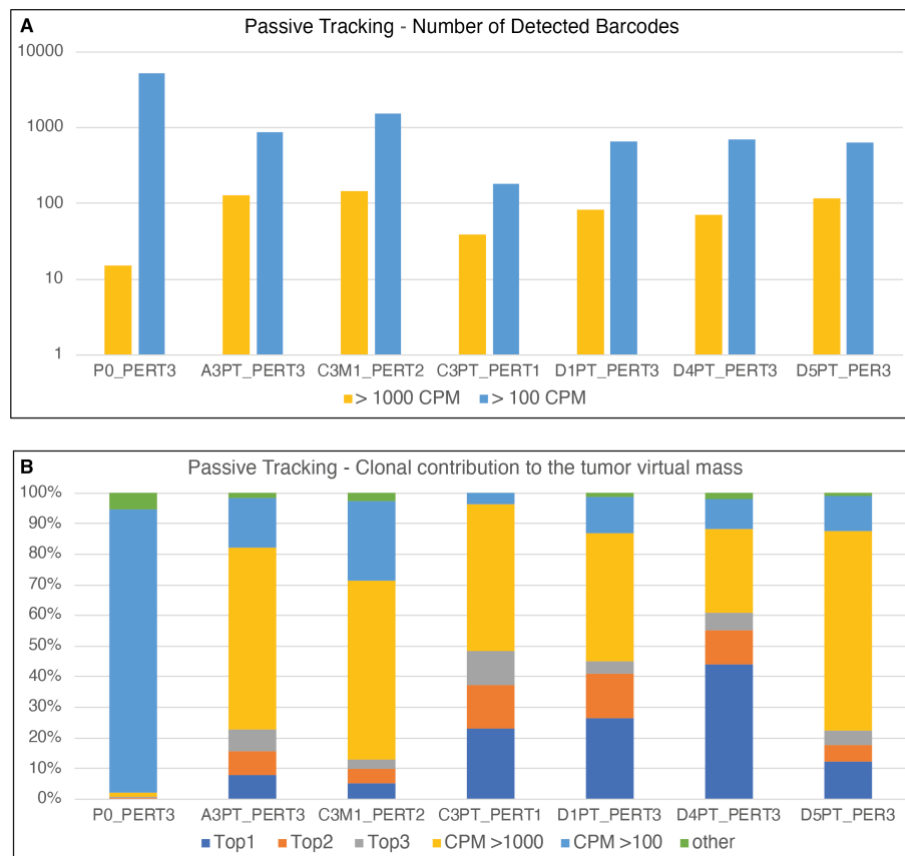
Figure 79: Clonal composition of SUM159PT xenograft - A) Number of barcodes detected at >1000 CPM or more than >100 CPM. B) Clonal composition expressed in percentage. The plot shows that a sensible portion of tumors is occupied by the top 3 most frequent sgRNAs. Data by Patricio Fuentes and Matteo Marzi, Nicassio Lab, IIT.

2.7.3 *In vivo* pilot experiment

We performed a pilot experiment in which we injected 100K, 500K and 1M cells of the LNC1-P0 in the mammary fat pad of 12 NSG mice. Prior to the injection, cells were treated with Doxycycline for 72h in 2D in order to induce the expression of dCas9-KRAB and start the repression of the target genes. We evaluated the latency of the tumor development, and we sampled the expression of the transgene from the collected tumors (mice were fed with Doxy-food). We observed an average latency of 47 days for the tumors (SD ±4.96 days), slightly longer than the latency compared to the passive barcodes experiment. The injection failed to form a tumor in only one mouse. Tumors were restricting the expression of dCas9-KRAB to lower levels compared to the expression in 2D (**Figure 80B**). We had previously measured that the KD is preserved at a similar expression levels in 3D (See **2.5.2**). No difference in latency was observed with different number of cells, therefore we performed the rest of injections with 500K cells. Based on the clonal diversity assayed by passive-

barcodes, we injected a total of 29 mice, with an expectation of observing a total of 3450 (sgRNA number) * 0.15 (engraftment frequency) * 29 (mice) = 15000 clones.
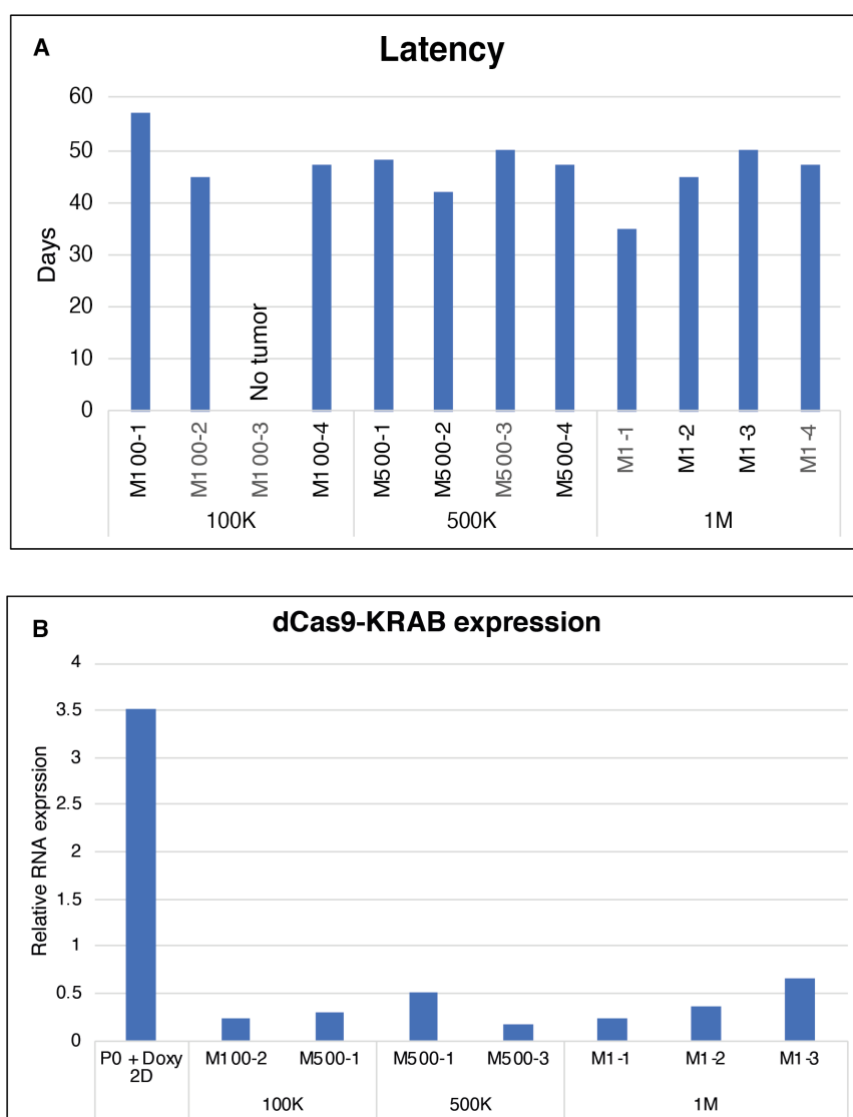
Figure 80: Pilot in vivo experiment – A) Tumors latency for pilot experiment. **B)** dCas9-KRAB expression from SUM-dCas9-KRAB- NC1 xenografts.

2.7.4 Clonal composition of SUM-dCas9-KRAB-LNC1 tumors

Of the 29 injected mice, 24 developed a tumor with similar latency to the pilot experiment. The LNC1 cells were purified from mouse tissues, and we experienced different recoveries, with an average 2.66 x10^6 cells (SD±2.47x10^6) (**Figure 81**) (tumor latency and N of human cells recovered per each tumor are reported in the **Methods** section). However, for two tumors, we failed to produce a PCR product probably due to residual mouse gDNA and overestimation of the actual SUM-dCas9-KRAB gDNA content. When possible, we produced technical replicates, observing very good reproducibility and limited noise (**Figure 82**). We analysed the clonal composition of tumors, and we monitored a similar trend compared to the passive barcode experiment with a clonal drift that led to the accumulation

of dominant sgRNAs that mainly contribute to the virtual tumor mass. Specifically, we detected an average of 1784 sgRNAs in every tumor (>10 CPM), representing only the 51.7% of the library. This means that, within each tumor, half of the library sgRNAs could not be detected due to the intrinsic selection of the process. We found an average of 122 dominant barcodes (>1000 CPM) in each tumor. For some samples, only few sgRNAs contributed to the "virtual tumor mass". Given the very low N of clones detected in mouse M3H3 and the different cumulative distribution of sgRNAs for mouse M3C we excluded these samples from further analysis (**Figure 83**).
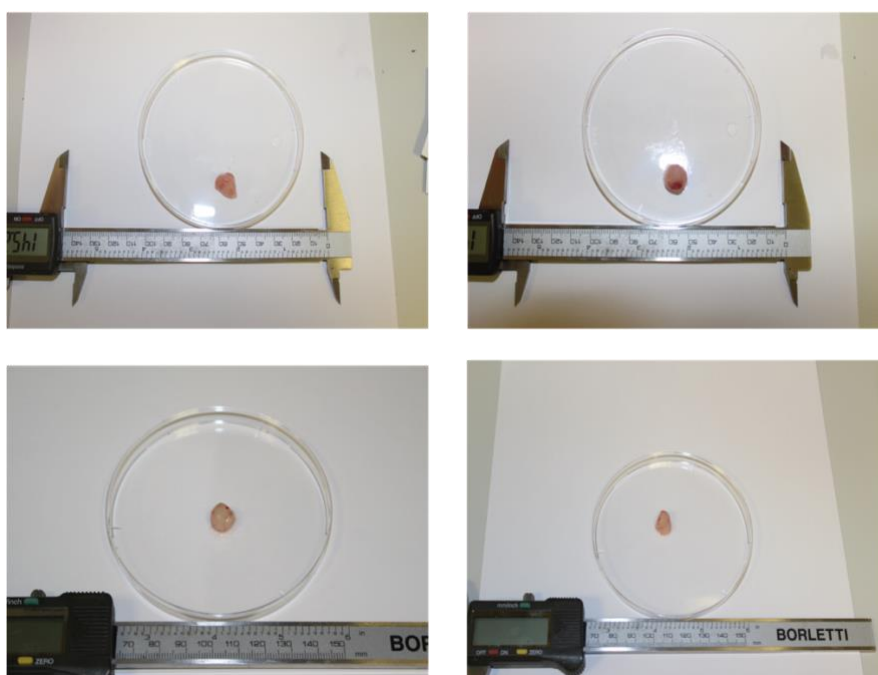
Figure 81: SUM-dCas9-KRAB tumors – Representative pictures of resected tumors.
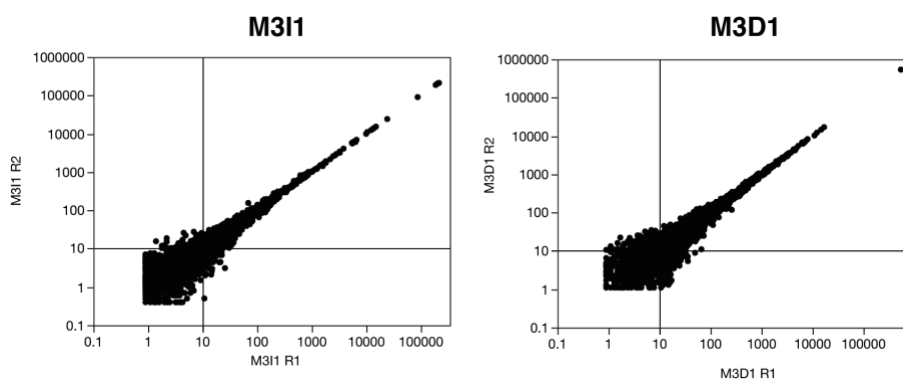
Figure 82: Reproducibility of technical replicates – Data shows the correlation of the distribution of raw sequencing reads in two technical replicates for two tumors.
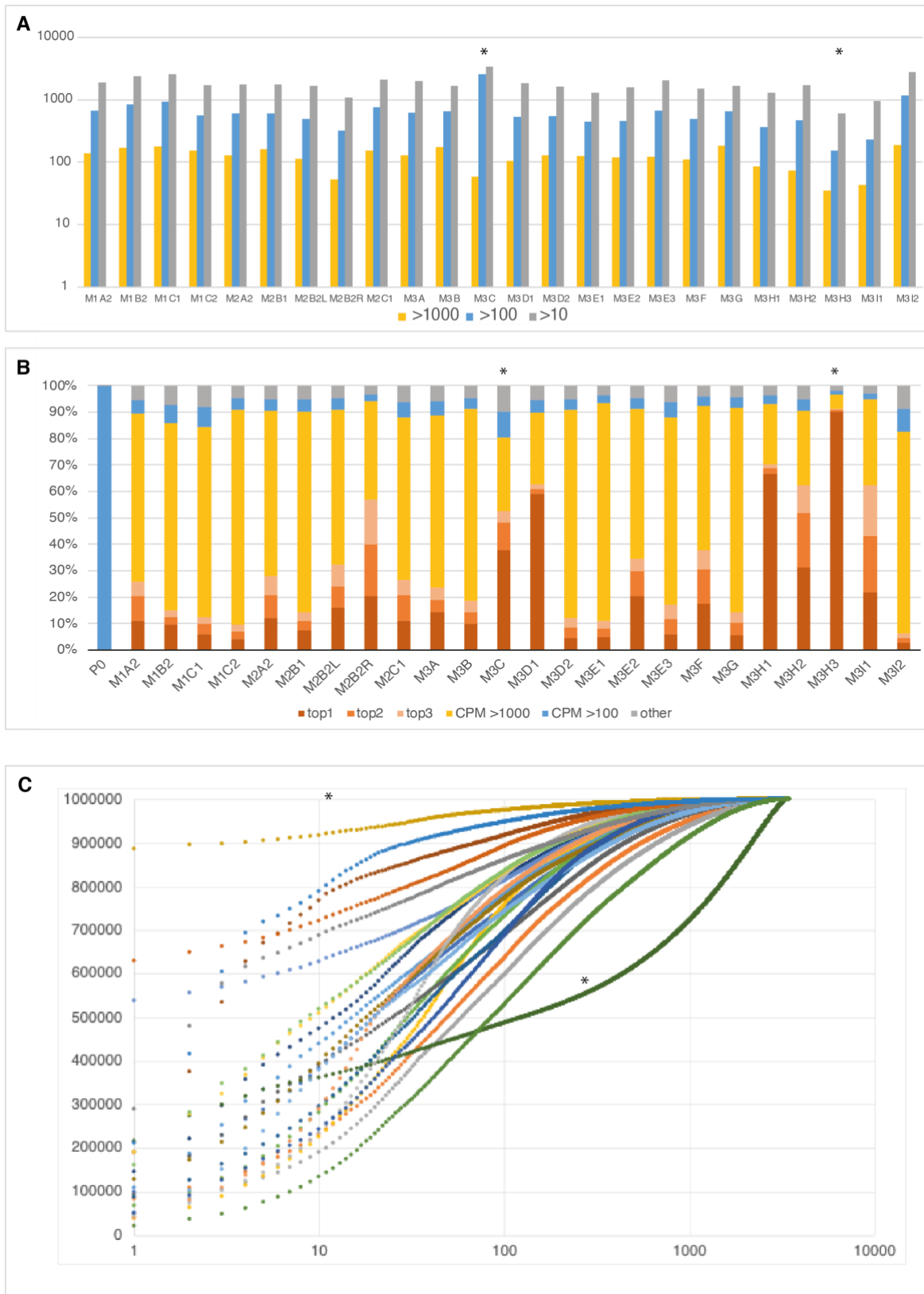
Figure 83: Distribution of sgRNAs in SUM-dCas9-KRAB-LNC1 tumors - A) N of sgRNAs detected >10 CPM, >100 CPM and >1000 CPM. **B)** Contribution of sgRNAs to the tumor mass. **C)** Cumulative distribution of sgRNAs frequencies in each tumor. Stars indicates excluded samples.

2.7.5 In vivo hits by phenotype-GSEA

With this level of selective pressure, it is not possible to maintain sufficient *library coverage* unless pooling samples from many different mice, as it is customary in *in vivo* screenings (Han et al., 2020). We previously showed that the selective pressure induced by Paclitaxel increases the noise of the system. To overcome this issue, we applied a phenotype-GSEA approach to model the drop-of-complexity observed (See **2.6.6**). In the *in vivo* screening we experience a similar need, and we applied the same analytical framework. We parallelized the analysis of the *in vivo* screening by calling DROP-IN and DROP-OUT genes based on their NES, calculated from the ranked frequencies of sgRNAs in every tumor (**Figure 84**). The behavior of NTCs was defined by creating 46 random sets of NTCs each composed by 5 non-targeting sgRNAs. We called DROP-IN and DROP-OUT hits based on their NES >2 or < -2 and a 5% FDR as computed by the GSEA tool. With these thresholds, we called 2 NTC sets as DROP-OUT and 4 NTC sets as DROP-INs, leading to an empirical FDR of 4% for DROP-OUTs and 8.7% for DROP-INs. We called as DROP-OUTs 13 of the Essential Genes and 55 lncRNAs genes and as DROP-INs 78 genes. Compared to the other screenings, we identified many hit genes, in particular in the DROP-IN class (see **Discussion**). We evaluated the representation of lncRNAs DROP-INs and observed that, similarly to the Paclitaxel screening, there is a limited effect of hist called as DROP-OUTs or DROP-INs that were under- or over- represented in the P0 (**Figure 84B**).
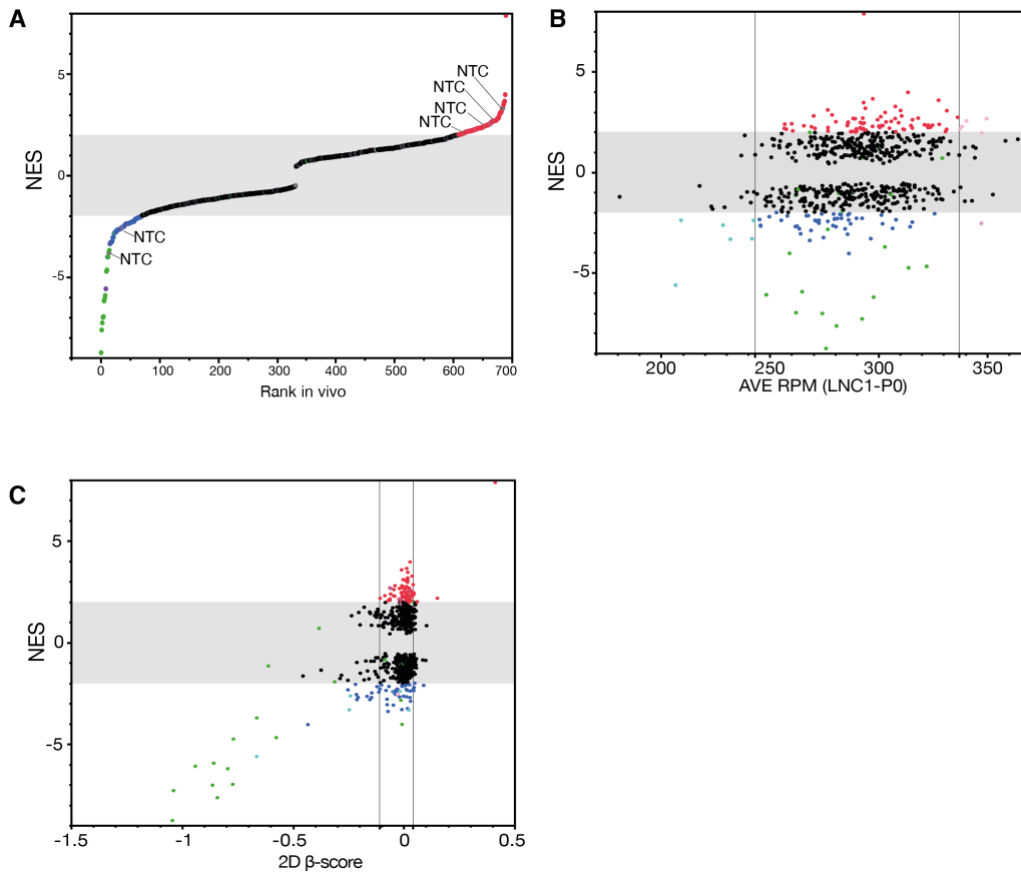
Figure 84: NES of the *in vivo* screening – A) Ranked representation of NES score. **B)** Scatter plot showing the average gene-level representation of genes in the LNC-Library and relative NES value *in vivo*. **C)** Scatter plot showing the average gene-level representation of genes in the LNC-Library and relative NES value. Figure legend: Bars show 2.5-97.5 percentile of the distribution. Shaded area non-significant NES. Figure legend: Green=Essential Genes; Blue=lncRNAs DROP-OUT; Red= lncRNAs DROP-IN; Light blue=lncRNAs DROP-OUT with low average CPM in the LNC1-P0. Pink=lncRNAs DROP-IN with high average CPM in the LNC1-P0.
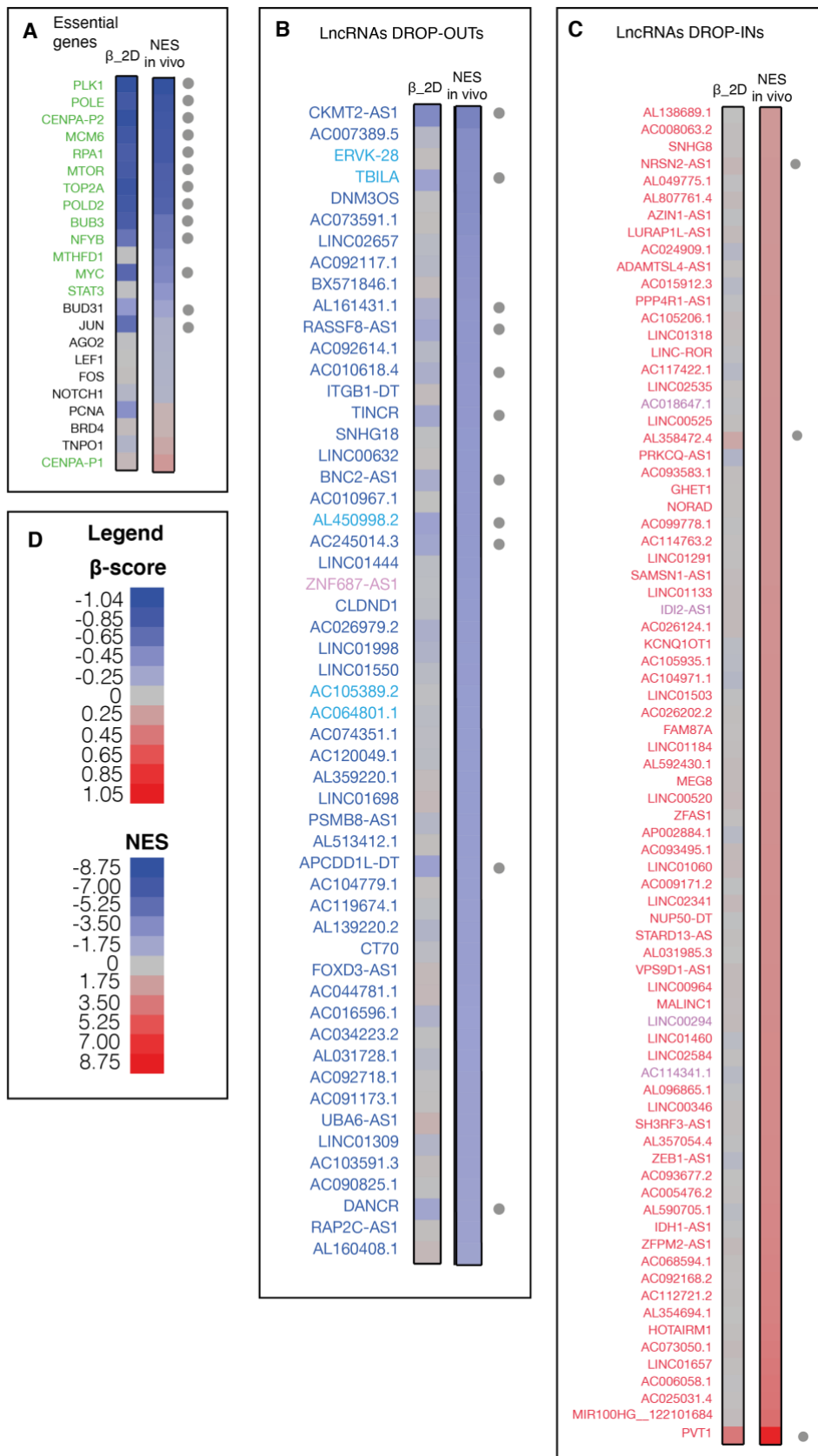
Figure 85: In vivo hits by GSEA analysis – Heatmaps showing comparison of β-scores and NES for **A)** all "Essential genes" set (green= genes of the set DROP-OUT in this screening) and **B)** only DROP-OUT hits defined by phenotype-GSEA; **C)** DROP-IN hits by phenotype GSEA. **D)** Cell legend. In blue= lncRNAs DROP-OUTs; light blue= DROP-OUT with low average CPM in the LNC1-P0. Red= lncRNAs DROP-IN; Pink= lncRNAs DROP-IN with high average CPM in the LNC1-P0.

2.8   Hits characterization & Validation

2.8.1   Summary of screenings hits

In this work we presented an approach aimed at identifying lncRNAs acting as regulators of adaptive phenotypes through the exploitation of multiple CRISPRi-based functional screenings. In particular, we tested 620 lncRNAs for their ability to modulate context-specific growth and/or survival under increasingly stringent growth conditions (2D, 3D, chemo-response, in vivo).

The LNC-library included also 23 coding genes picked as controls for the phenotypes under investigation (essential genes and cancer genes). As expected, most of the control coding genes (82%) were also found as hits, with 11 out of 23 of the "essential" genes found as common DROP-OUTs in all screenings. **Figure 86** summarizes the hits obtained through the different screenings. Overall, 253 genes were found as hits in at least one phenotypic screen (39%). 9.5% genes were found as hist in multiple screenings (>1), of which only a small part (10 genes, 1.6%) were incoherent. 14 hits were common to all screenings (2.1%). Usually, hits were shared between 3D and 2D. For DROP OUTs most of the hits were called in 3D and in vivo screens, while for the DROP INs the majority were called in the in vivo screening. Several hits were found in just one screening (29.5%). In next the paragraphs we present a preliminary characterization of hits and the first step towards the validation of the results obtained in the screenings.
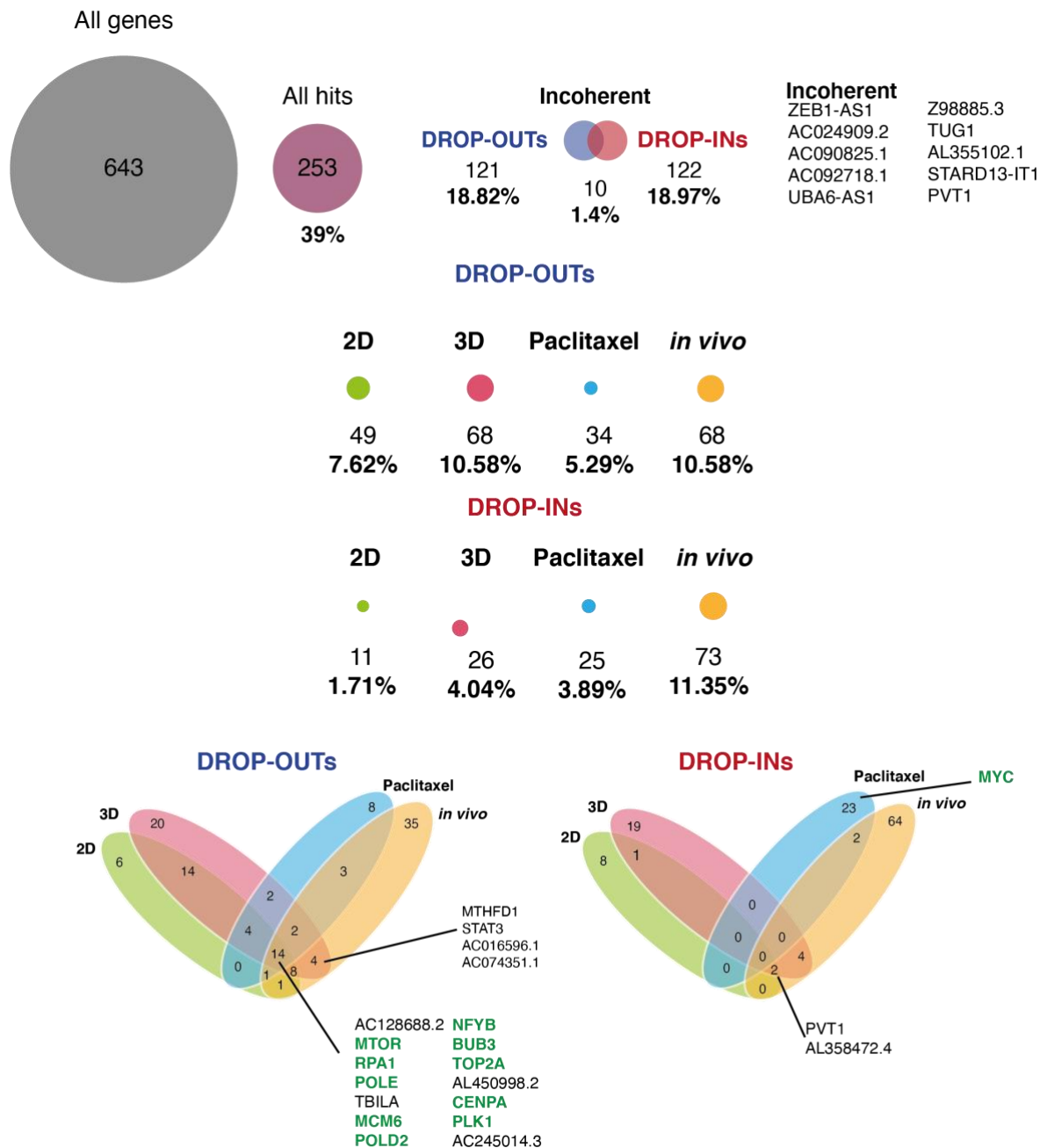
Figure 86: Summary of screening hits – Scheme summarizing all hit genes (PCGs and lncRNAs). Circles express the percentage of overall DROP-OUTs and DROP-INs with indication of hit genes behaving incoherently in different screenings. Below the representation of % of DROP-OUTs and DROP-INs in each screening. Reported hits by phenotype-GSEA in Paclitaxel and *in vivo* screening. Venn diagram showing overlap of hits among screenings.

2.8.2 Features of hit genes

TSS definition and sgRNAs guide design

Having defined the hits, we checked for possible correlation that may support the approach we used for the TSS definition and for the sgRNA design. Indeed, efficacy of CRISPRi-based screenings may depend on the design of the guides and on the precise identification of TSS of target genes, which is not trivial for lncRNAs. As said in previous section (see **2.3.1**), we could use CAGE data (in house generated and publicly available) to define the TSS. Interestingly, we observed an enrichment in DROP-OUTs for lncRNAs that were defined in such way (69% of Hits as compared to 55% of the starting library) (**Figure 87A**). However, we did not observe a similar behavior for DROP INs, which were in similar proportion with the library (58%) (**Figure 87A**).
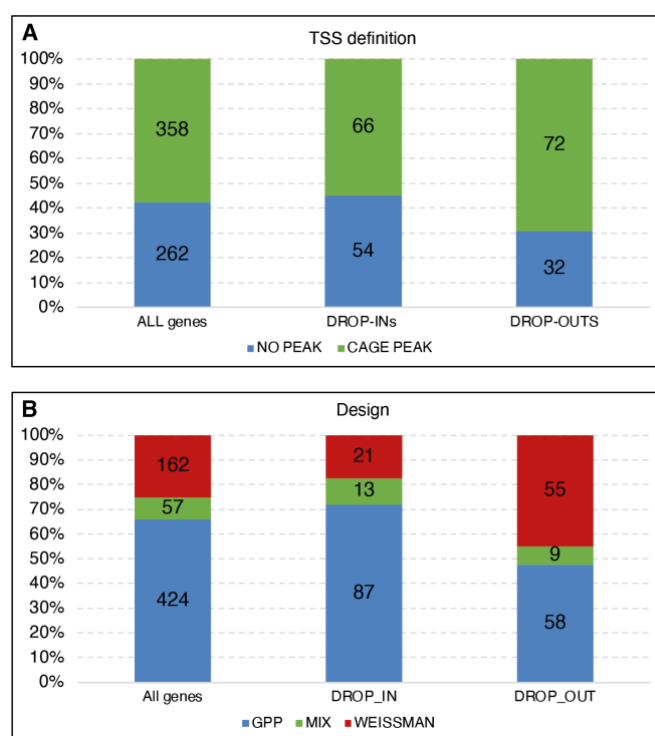
Figure 87: Hits by TSS definition and sgRNA design: A) Proportion of TSSs supported by CAGE in LNC-library and in DROP-IN and DROP-OUT hits. **B)** Hits by design strategy.

Structure and genomic features

Our lncRNAs candidates are also characterized by different genomic features. According to their GENCODE definition, we broadly defined lncRNAs into categories, as antisense lncRNAs (AS, 30% of the library) or intergenic lncRNAs (50%), with remaining transcripts (Others, 11%) having a different genomic organization (e.g. intronic or overlapping-transcript lncRNAs). We observed that these proportions were generally maintained in DROP-IN and DROP-OUT hits (**Figure 88A**), suggesting that such genomic features are

not segregating with the functions investigated through the screening. Next, we stratified lncRNA hits with their exon composition, as it might hint to the selection of specific functions (mono-exonic lncRNAs are often considered as a separate class). We observed no major changes in the proportion of mono-exonic and multi-exon lncRNAs in hit genes (**Figure 88B**).
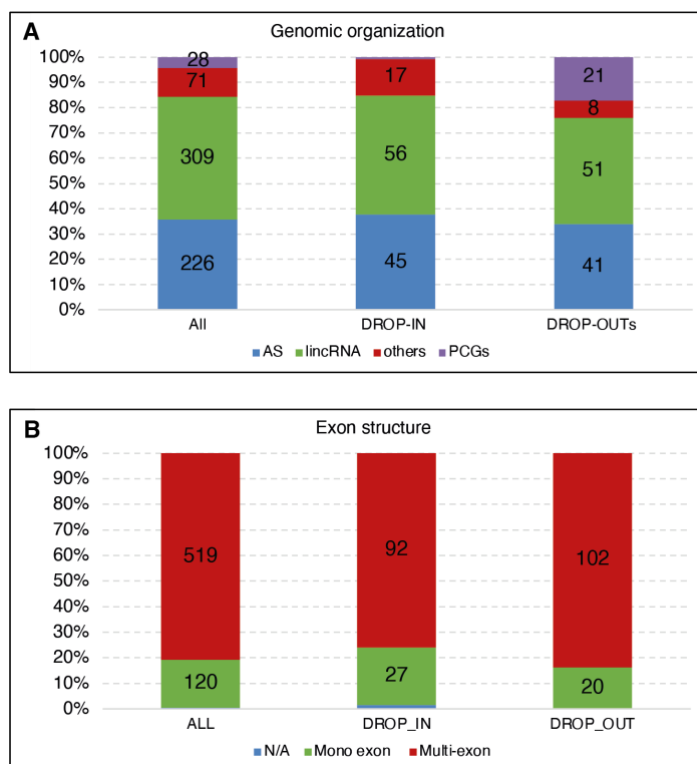
Figure 88: Genomic features and structure of hits – A) Genomic organization of genes in the LNC-library as AS, intergenic or "other", PCGs included are also displayed and represent, as expected, a fraction of the DROP-OUTs. **B)** Structure of genes in the LNC-library as "Monoexonic" or "multi-exon".

Distance from PCGs
The CRISPRi effect is unavoidably bi-directional and involves genes near the target. There is therefore a possibility that some of the lncRNA hits have been selected due to local repression of essential genes in the proximity. Therefore, we examined more into details the existing spatial relationships between lncRNAs and PCGs. We retrieved from ENSEMBL the list of all PCGs located within 1Mb from the TSS of lncRNAs (-500kb; +500kb) and annotated essential genes in this set according to their definition in DepMap. We split the guides according to the hit categories (DROP IN, DROP OUT, Incoherent and No effects) and measured their distance to PCGs and essential genes. Results are summarized in **Figure 89**. Overall, we did not find a clear skewing of Essential Genes for the DROP OUT category as compared to "No effects", discouraging the possibility of a major bias in the screening results. However, we could notice specific cases where essential genes are proximal to

114

guides. We found nine essential genes overlapping the window of sgRNAs design (~500 bp). These genes are likely to be directly affected by the activity of the dCas9-KRAB. In two cases (ZCCHC9 and EIF6), we also could confirm that guides directed towards the TSS of lncRNAs CKMT2-AS1 and FAM83C-AS1 exert a co-repressive effect on these genes (See **2.8.3**). We haven't checked for the other 7 genes, but we can assume that even in these cases a co-repression is occurring. Furthermore, we found four essential genes located between 500 and 2000 bp from the TSS of DROP-OUT lncRNA hits, a range where the activity of dCas9-KRAB cannot be excluded (Gilbert et al., 2014; Nuñez et al., 2021) and deserve further investigation. Of note, we also found eight essential genes that mapped in close proximity of lncRNAs whose perturbation showed no effect (4 within 500 bp and 4 within 2000 bp, respectively; **Figure 89)**. In conclusion, the analysis of distance from essential genes revealed that only a minority of DROP OUTs are in proximity of essential genes (their identity is reported in the table below). These hits deserve more attention as it is reasonable to assume that phenotypes scored for guide RNAs targeting their genomic loci are likely due to the combined targeting of both the coding and the non-coding gene.

Figure 89: Distance of lncRNAs in the LNC-Library from PCGs – A) Plot showing distance of all lncRNA genes of the library from protein coding genes. Red crosses= essential genes, black dots = not essential PCGs according to their definition in DepMap. In bold essential genes whose KD has been validated by qRT-PCR. **B)** Summary of essential protein coding genes located within 500 bp from the TSS of lncRNAs hits.

Cellular Localization

Often the function of lncRNAs is linked to their cellular localization. Information on localization helps also in the selection of other approaches rather than CRISPRi (such as ASOs or shRNAs), necessary to address functional analysis of specific lncRNAs. We distinguished lncRNAs into broad cellular categories (nuclear vs cytoplasmic) by exploiting in house data produced by the sequencing of RNA isolated from cellular fractionation in SUM159PT cells, in a chromatin and cytoplasmic fraction (See **Methods**). Fractionation was performed in three experimental replicates. As quality control, we evaluated the relative proportion of markers of fractions by qRT-PCR (**Figure 90**), including i) well known nuclear transcripts, such as primary miRNA transcripts (which are short-lived and chromatin retained) and NEAT1 lncRNA; or ii) frank cytosolic RNAs, such as mRNAs of PCGs

(*RPLP0* and *GAPDH*) and *H19* lncRNAs. Using RNA-seq data, we calculated the total FPKMs per each detected transcript in every replicate and defined the average percentages per each fraction. Overall, lncRNAs showed an intermediate distribution between nucleus and cytoplasm and were slightly more nuclear than PCGs (**Figure 91A**). Of the 620 lncRNAs in the LNC-Library we could detect 410 lncRNAs in all the replicates (66%) (**Figure 91B**). To classify their localization, we deemed as nuclear those species with an average nuclear distribution >60% and cytoplasmic those <40%. In total we classified 80 lncRNAs as nuclear (13%), and 260 (42%) as cytoplasmic while the remaining 233 (37%) showed no clear preference but were similarly distributed between nucleus and cytoplasm. When looking at the lncRNA hits in the screening, we noted that most hits were classified as cytoplasmic lncRNAs, in particular those from the 2D and 3D screenings (**Figure 91C, D**).
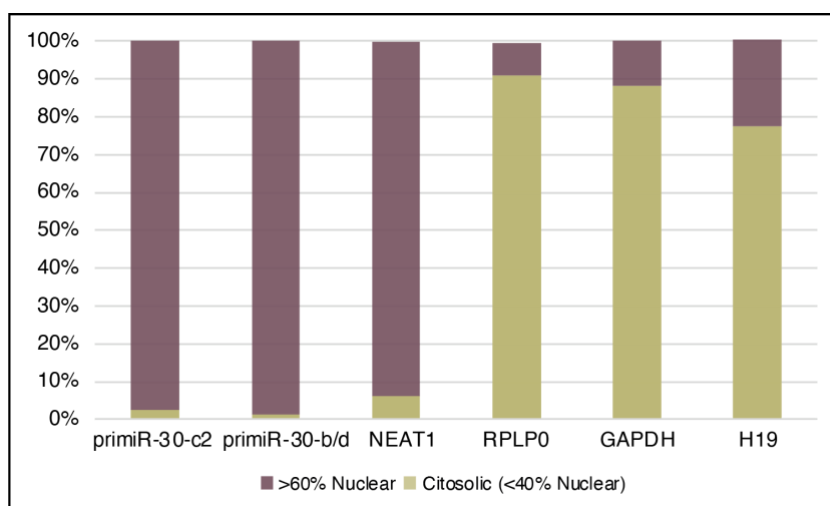
Figure 90: Quality control of cellular fractionation – Representative qRT-PCR showing the distribution of markers of nuclear and cytoplasmic fraction. Pri-miRNAs and the structural lncRNA *NEAT1* are enriched in the nuclear fraction while the ribosomal RNA *RPLP0*, the metabolic gene *GAPDH* and the lncRNA *H19* are enriched in the cytoplasmic fraction.
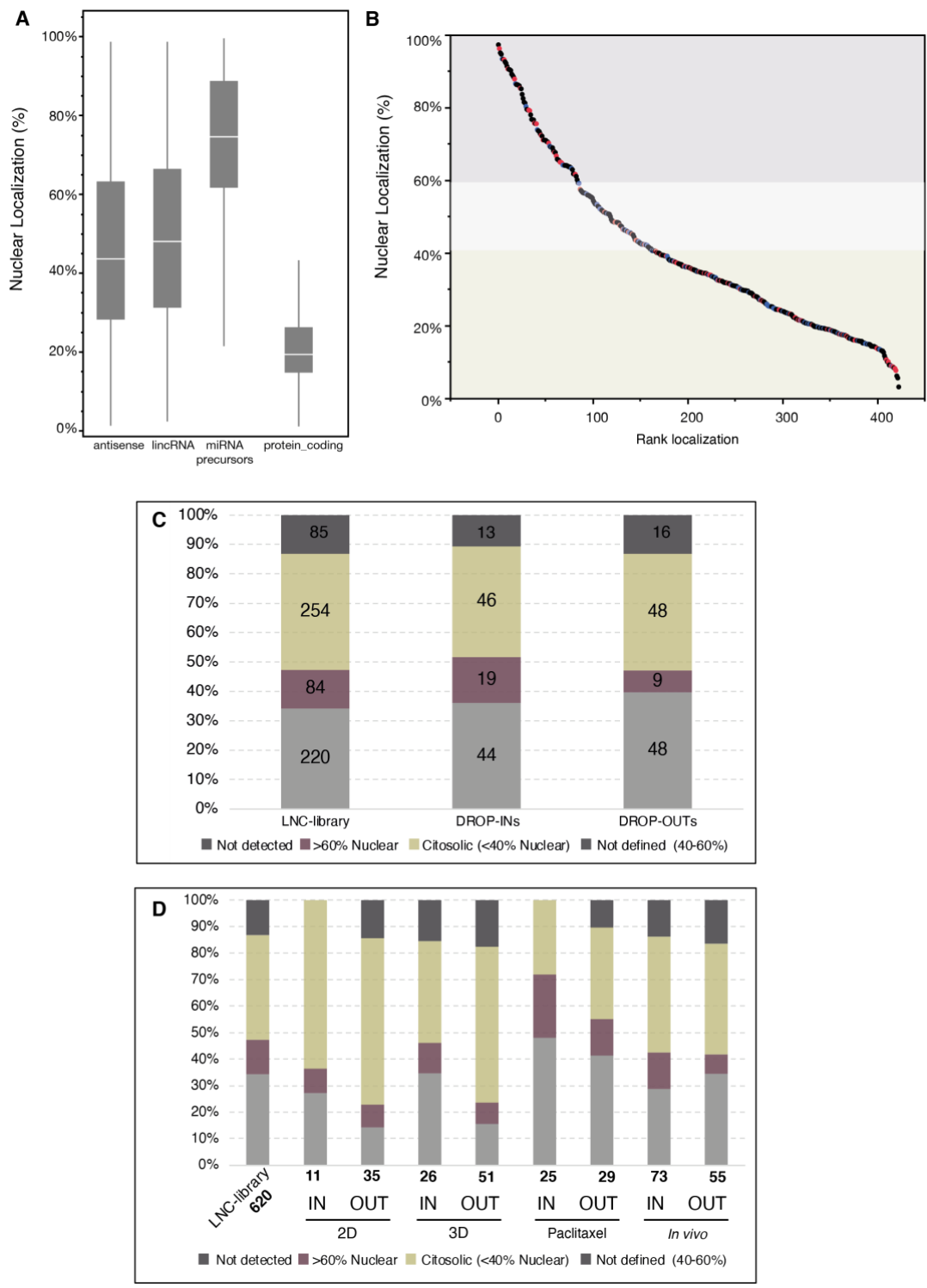
Figure 91: Cellular localization of library lncRNAs – A) Box plots showing the percentage distribution in the nuclear fraction of AS or intergenic lncRNAs, miRNA precursors and PCGs. The distribution is separately represented for genes expressed <1 or >1 FPKM. **B)** Ranked representation of percentage distribution in the nuclear fraction of all lncRNAs detected by RNA-seq of fractions. Blue dots and red dots indicate respectively DROP-OUTs and DROP-INs lncRNAs in any screen. In purple area of "Nuclear" genes and in yellow "cytosolic" genes. **C)** Stacked bar chart showing the distribution of all lncRNAs in the LNC-library, only lncRNAs DROP-OUTs or lncRNAs DROP-INs in any screen. **D)** Detailed distribution of lncRNAs hits in every screen. Data by Chiara Tordonato (Di Fiore Lab, IEO and Matteo Marzi, Nicassio lab, IIT).

Hits by model of selection

We also checked for any potential preference for lncRNA hits with respect to the breast cancer model that we used for the selection of lncRNA candidates (See **2.1**). Results are summarized in **Figure 92**, with hits stratified according to the model. Overall, the average rate of hit calling of DROP-OUT and DROP-IN hits were quite similar for all sets and not far from the original representation in the LNC-Library with just few exception that will be directed in discussion.
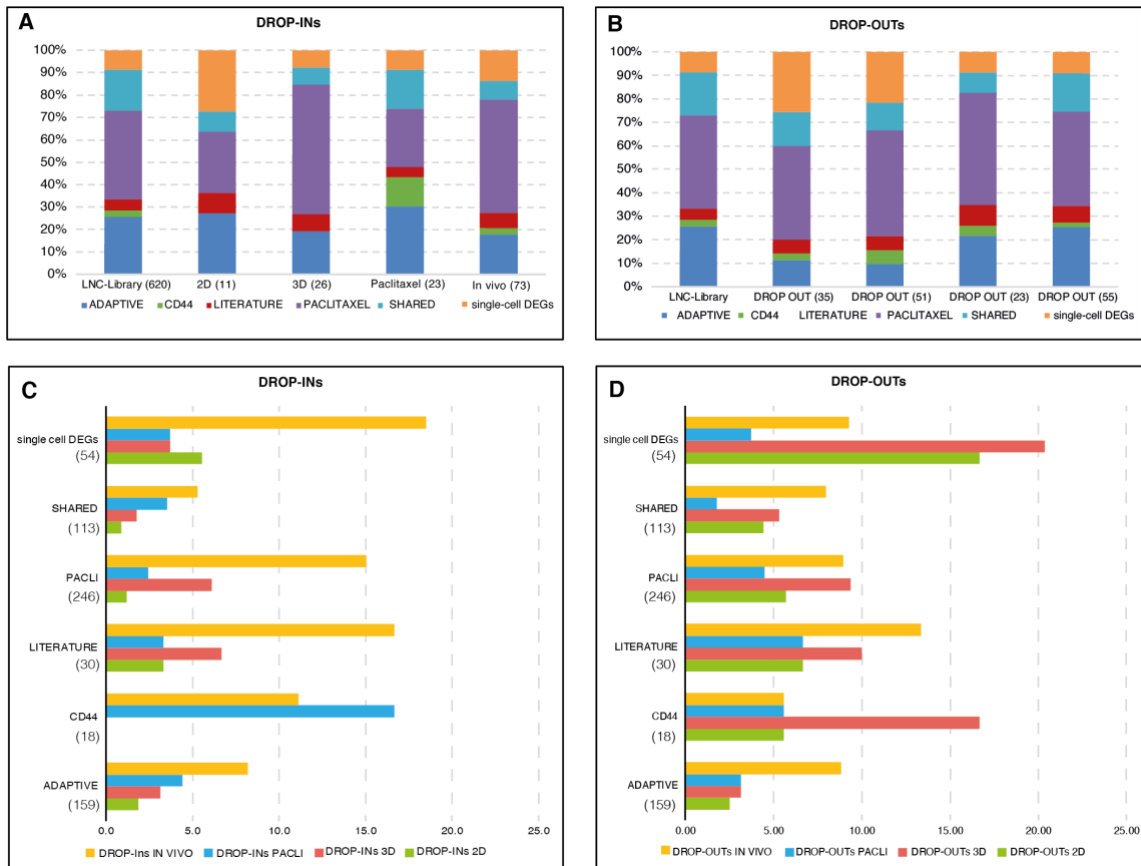
Figure 92: Hits by selection model – A) Percentage composition of DROP-IN hits lncRNAs genes in every screening according to the selection model. **B)** Percentage composition of DROP-OUT hit lncRNAs genes in every screening according to the selection model. **C)** Percentage of DROP-IN hits lncRNAs per selection model. **D)** Percentage of DROP-OUT hits lncRNAs per selection model in every screening.

Expression of hit lncRNAs in cell lines and primary tumors

To provide an expression map of the lncRNAs hits, we exploited in house RNA-sequencing data produced from different breast cancer cells and conditions. In details, we characterized the expression transcripts in SUM159PT during unchallenged proliferation, 3D mammosphere growth, upon Paclitaxel treatment in acute conditions (50 nM for 72h) or after long-term adaptation (~100 days with cells treated with increasing concentration up to 100 nM of Paclitaxel). Moreover, we examined the expression of lncRNAs hits in the HMLE cell line (a stable human mammary epithelial cell line) during unchallenged proliferation, 3D mammosphere growth; and in other different breast cell lines, comprising all the molecular subtypes of mammary tumors, including: i) quasi-normal mammary cell lines (MCF10A) and their HRAS (Gly12Val) derivative (named MCF10A-DCIS), ii) luminal-type breast cancer cell lines (MCF7, MDA-MB-175-VII, TD47D, BT483); and basal-like breast cancer cell lines (BT549, MDA-MB-231, MDA-MB-436). Furthermore, we included also RNAseq data from primary cultures from 26 breast tumors, displayed as average expression (ALL) or 90th percentile value of the expression in the dataset (90%). **Figure 93** summarizes these results of lncRNA expression as an heatmap, highlighting for each hit the specific patterns of expression in the models. Overall, most lncRNA display a low but consistent expression in most cellular models, with some expression peaks in some conditions or cell lines. Some lncRNAs show higher expression in basal-like as compared to luminal models and other lncRNAs peaked in conditions associated to drug treatment. Of note, most lncRNAs showed higher expression in primary samples, suggesting that their expression is not limited to cell culture models, but their expression could be a common feature of breast cancer.
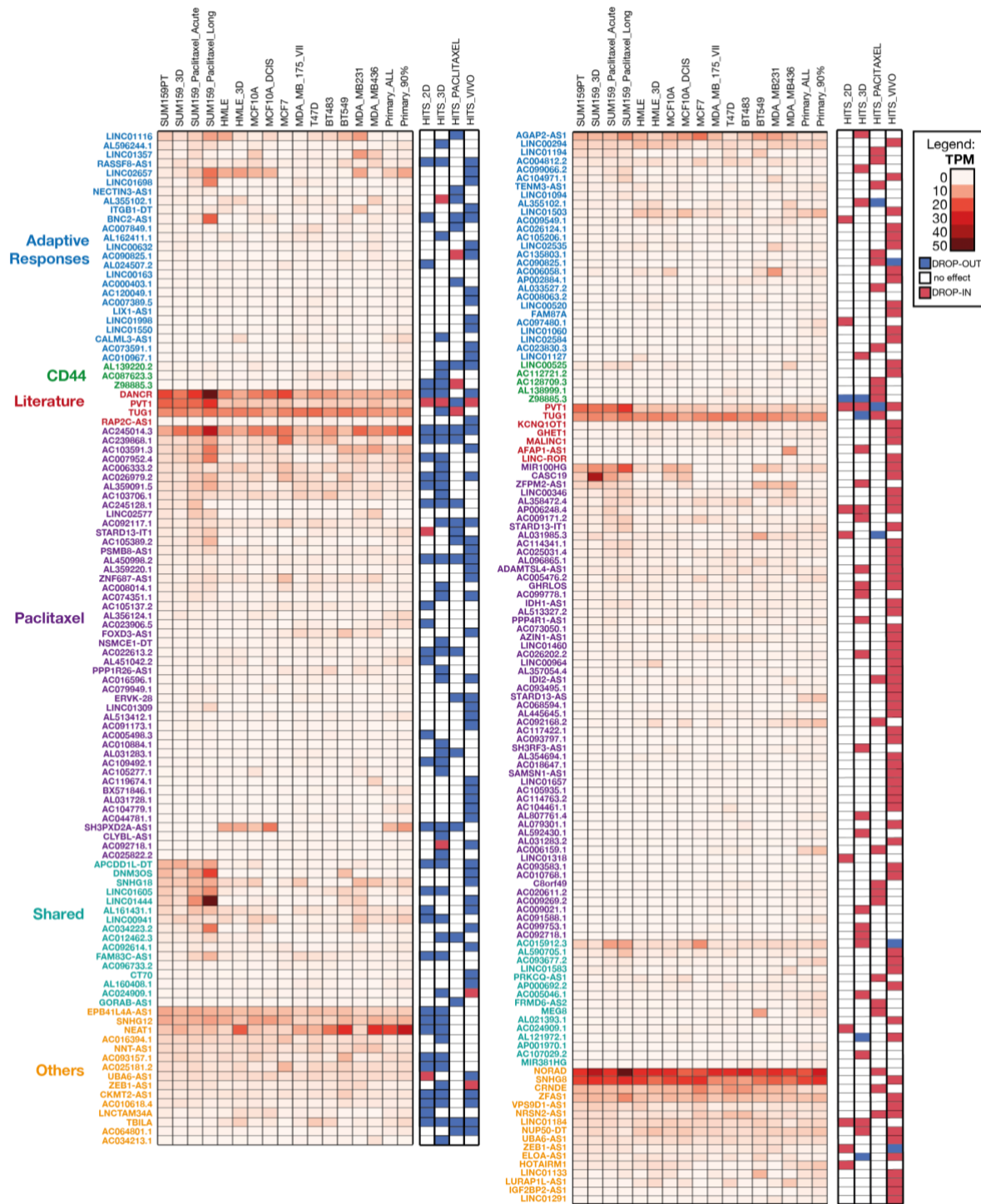
Figure 93 : Expression of hit lncRNAs in breast cell lines and primary tumors – Heatmaps showing the expression of DROP-OUT (left) and DROP-IN hits (right). The digital maps show the results of the screening per each gene displayed. Maps are divided by model of selection (others = single-cell DEGs). Primary tumor data are displayed as average expression in 26 breast tumors or 90th percentile value of the expression in the dataset. TPM values are ranked according to the expression levels in native SUM159PT.

2.8.3 KD validation of screening hits

So far, we have discussed the overall features of the lncRNA hits that we have identified through the screenings. In this section, instead, we want to provide details on the validity of the results obtained. Hence, we selected guides targeting several hits of the screenings, cloned them individually and tested experimentally both to confirm the specificity of the knock-down and to subsequently validate the phenotype (See **2.8.4**).

Initially, we selected 8 lncRNA gene hits identified in multiple screenings (summarized in **Figure 94**) and checked the on-target knock-down. We also included the PCG MTHFD1. For each target, we selected two independent sgRNAs from those used in the screening. We evaluated knock-down measuring the expression of the target gene/lncRNA. Furthermore, for antisense lncRNAs, we also measured the expression of the matched sense coding genes (data are summarized in **Figure 95** and extensively shown in **Figure 96**).

GENE NAME	2D	3D	PACLI-MAGeCK	PACLI GSEA	IN VIVO GSEA	KD test	competition
AC016596.1	no effect	DROP-OUT	DROP-OUT	DROP-OUT	DROP-OUT	YES	NO
AC093157.1	DROP-OUT	DROP-OUT	no-effect	no-effect	no-effect	YES	YES
AC245014.3	DROP-OUT	DROP-OUT	no-effect	DROP-OUT	DROP-OUT	YES	YES
CKMT2-AS1	DROP-OUT	DROP-OUT	no-effect	no-effect	DROP-OUT	NO	NO
EPB41L4A-AS1	DROP-OUT	DROP-OUT	no-effect	no-effect	no-effect	YES	YES
FAM83C-AS1	DROP-OUT	DROP-OUT	no-effect	no-effect	no-effect	YES	YES
LINC01444	no-effect	no-effect	DROP-OUT	no-effect	DROP-OUT	YES	YES
MTHFD1	no-effect	DROP-OUT	no-effect	no-effect	DROP-OUT	YES	YES
RASSF8-AS1	DROP-OUT	DROP-OUT	DROP-OUT	no-effect	DROP-OUT	NO	YES
SH3PXD2A-AS1	DROP-OUT	DROP-OUT	DROP-OUT	DROP-OUT	no-effect	NO	YES
TBILA	DROP-OUT	DROP-OUT	DROP-OUT	DROP-OUT	DROP-OUT	YES	YES

Figure 94: Summary table of lncRNAs candidates selected for validation.
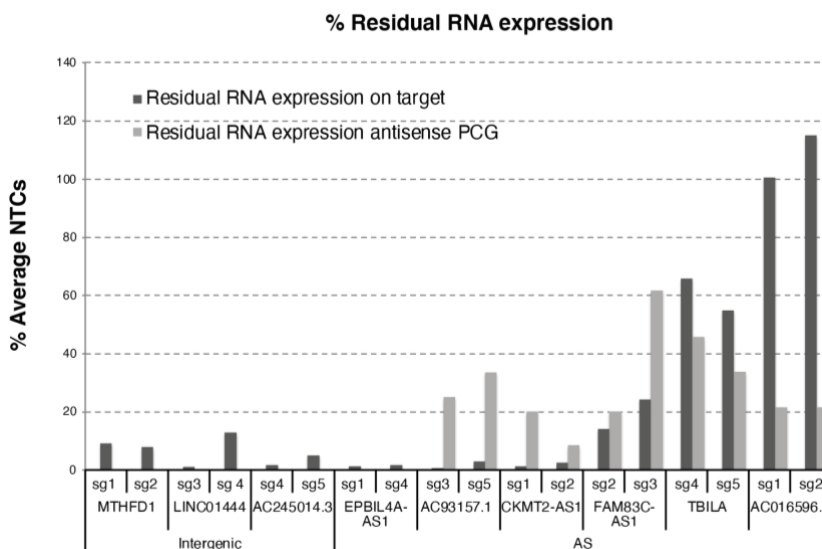
Figure 95: KD of screening hits – Residual RNA expression as % of the average expression assayed in two NTC samples. KD measured 72 upon induction of dCas9-KRAB by Doxycycline treatment.
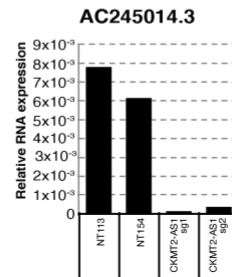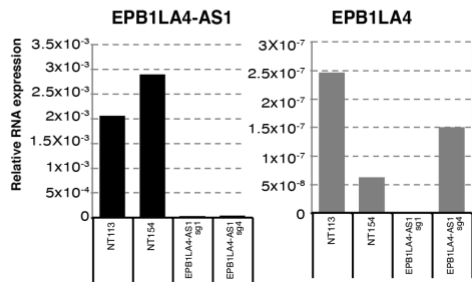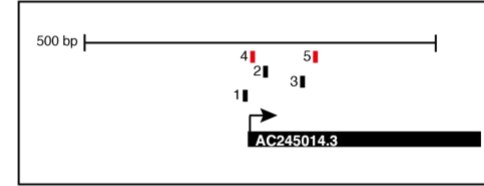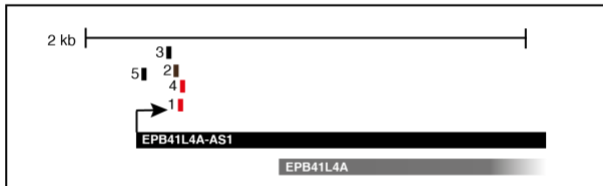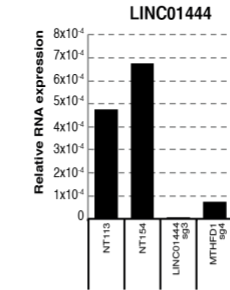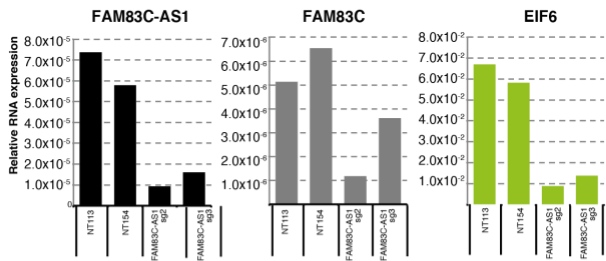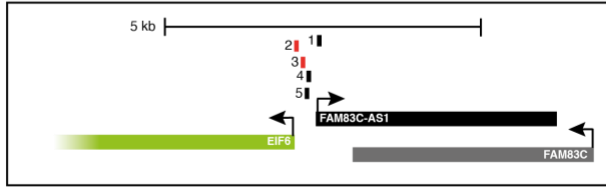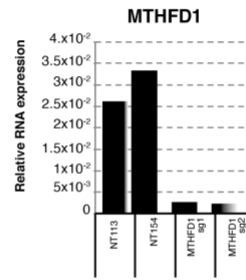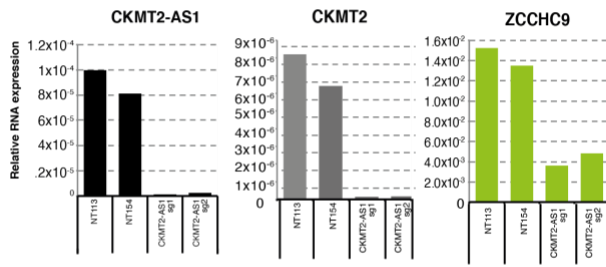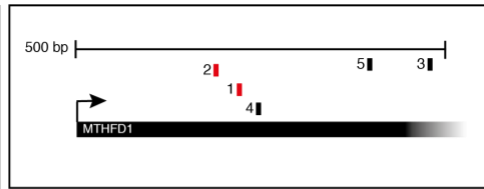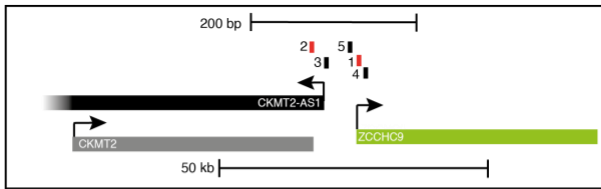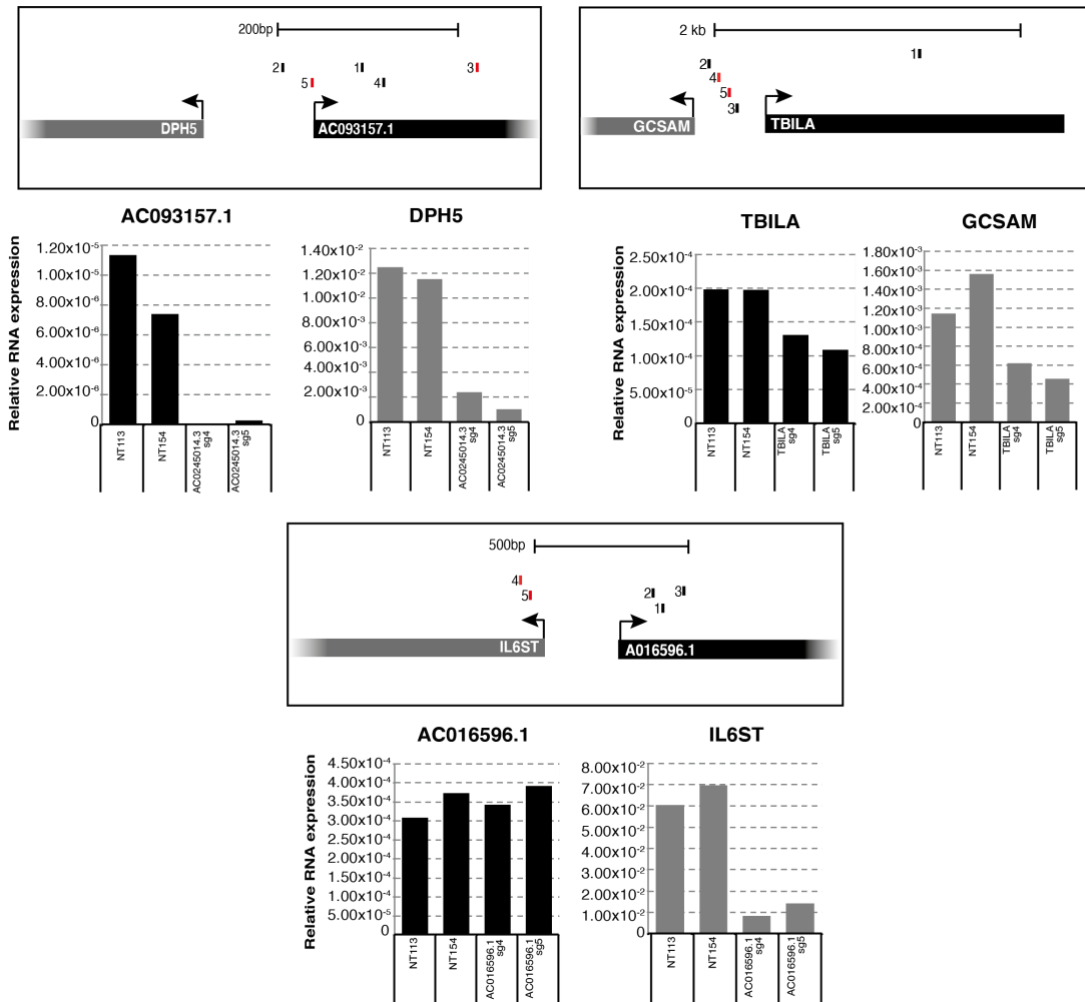
Figure 96: Relative RNA expression of target genes and closest antisense PCGs - Analysis of gene expression by qRT-PCR of target lncRNAs (in black) and genes in proximity (in grey, green bars highlights essential genes). The upper box shows genome browser screenshots of the genomic organization of genes with location of sgRNAs (in red sgRNAs tested for KD). Data normalized vs *RPLP0*.

We observed that in most cases (7 out of 8 of the tested target genes) we could effectively reduce the RNA expression with both guides. On average, we achieved a knock down efficiency of >80%. Only in one case, the mono-exonic lncRNA TBILA, we obtained a knock down of 50%. Only one lncRNA (AC016596.1) was unaffected, likely due to a miss-annotation of the transcription start site. Indeed, the guides were still able to repress the PCG IL6ST, which is close to the A016596.1 locus. As expected, we noticed that the repression effect induced by KRAB also involved to the nearby genes, when present. It is worth mentioning that for the CKMT2-AS1 locus, we observed the repression of the target lncRNA (CKMT2-AS1) together with repression of the sense transcript (CKMT2) and the divergent gene ZCCHC9, an essential gene according to the reference (DepMap). Similarly, guides directed to the lncRNA FAM83C-AS1 induced the repression of FAM83C and of the essential gene EIF6. Therefore, as anticipated in the previous section (Distance from PCGs),

the interpretation of the phenotypes deserves particular attention as it might be due to a combination of effects of lncRNAs and their neighboring essential genes.

2.8.4 Validation of effects in 2D proliferation

We set up competition assays comparing over time two cell populations grown in the very same experimental setting, one population (marked in red) bearing the targeting sgRNA and one population (marked in green) bearing a non-targeting control. In particular, we aimed at confirming the hits and, at the same time, at understanding how much the screening results are reliable in terms of hits definition and of size effects. For each target gene, we measured the effect of individual sgRNAs in terms of log2FC and, then, selected for validation 2 NTCs controls and 2 sgRNAs per target gene (**Figure 97**). We cloned guides in the same library plasmid (Cellecta pRGSCribe1) expressing fluorescent markers (RFP for targets and GFP for controls) and transduced SUM-dCas9-KRAB cells. To perform the competition, we mixed equal amounts of RFP+ and GFP+ cells at the starting time and then monitored the proportion of cells at 10 days of culture (**Figure 98** and **Figure 99**). When using only non-targeting controls, we observed that the two populations did not change their representation over time (average delta start-end at 10 days NT113= -1.30%, NT154= -1.50%). Conversely, when using sgRNAs targeting DROP-OUT hits, we observed that the GFP population was underrepresented after 10 days of culture. The sgRNAs behavior confirmed the one measured at P3 (10 days) in proliferation screenings. For instance, in case of AC245014.3, we observed, in the competition, the very same behavior in terms of size effect for both the two sgRNAs (**Figure 97)**, with sgRNA4 strongly depleted at P3 (10 days) while sgRNA5 depleted with a modest effect. In the competition experiment, sg4 RFP+ cells decreased down to 1.44% (strong effect), while sg5 RFP+ cells decreased to 38.4% (modest effect). In case of lncRNA EPB41L4A-AS1 and AC093157.1, we could also reproduce, in the validation experiment, this slower depletion kinetics observed in the screening (**Figure 99**). Overall, these data strongly support the results from the screening in 2D conditions, as we could validate not only the phenotype of target genes, but also the magnitude and the difference in the kinetics of individual sgRNAs.
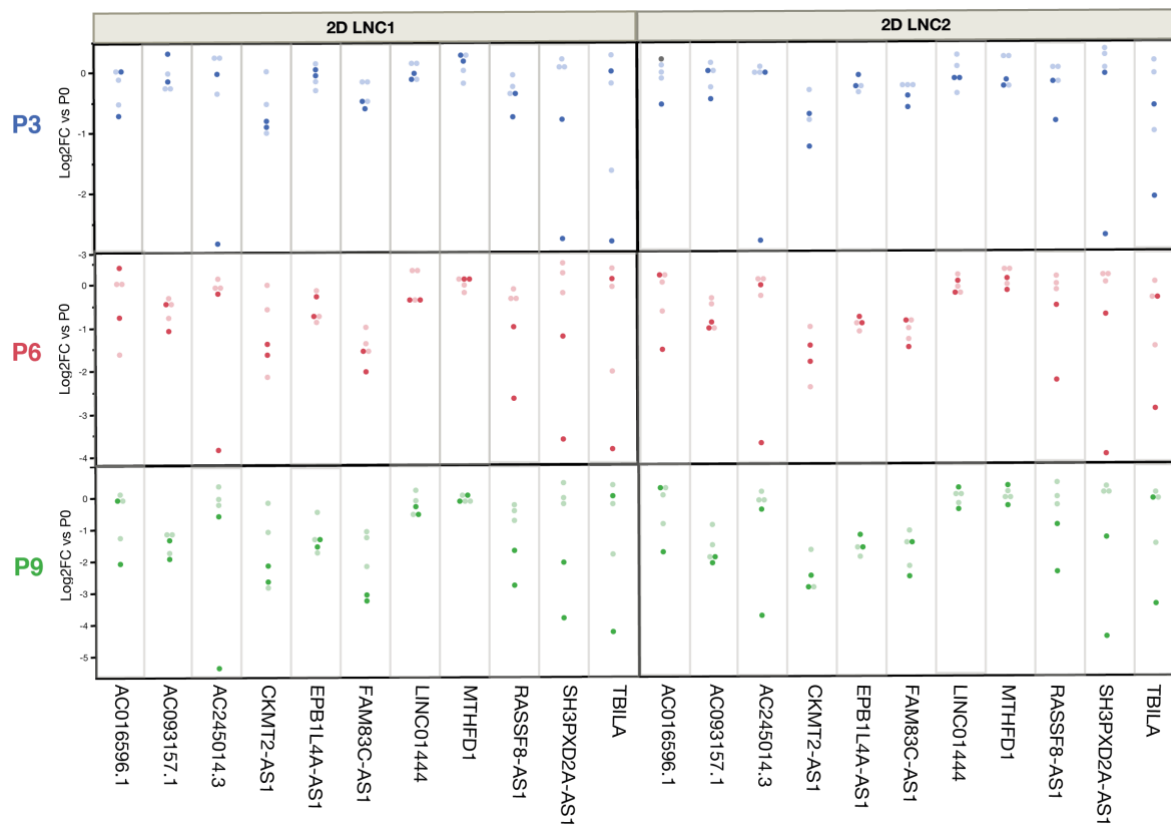
Figure 97: Log2FC of genes selected for validation - Figure expressing the log2FC of sgRNAs targeting genes for single competition experiments at every time point and biological replicate of the 2D proliferation screening. Highlighted sgRNAs selected. CKMT2-AS1 and AC016596.1 were not yet tested in this framework.
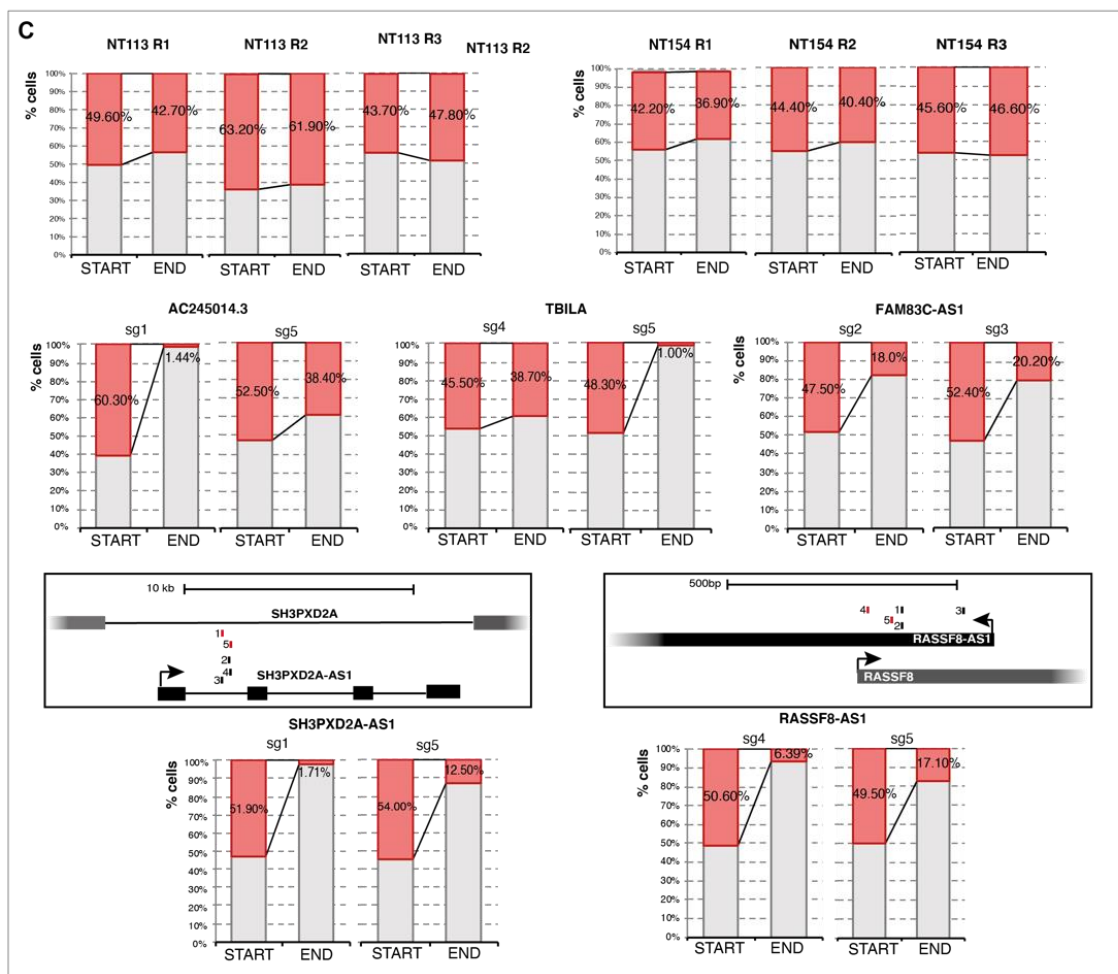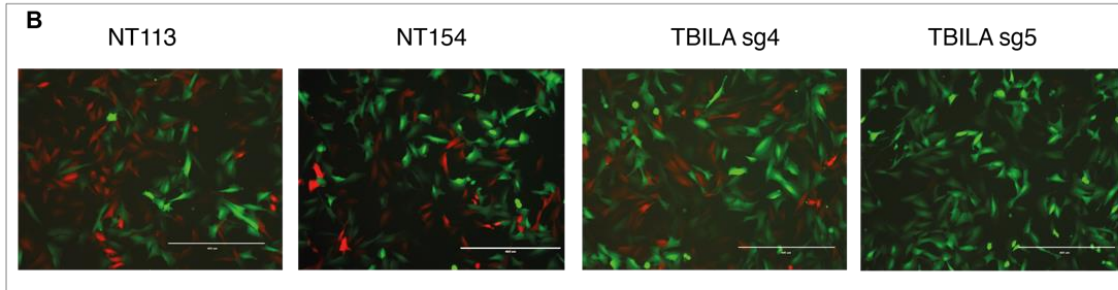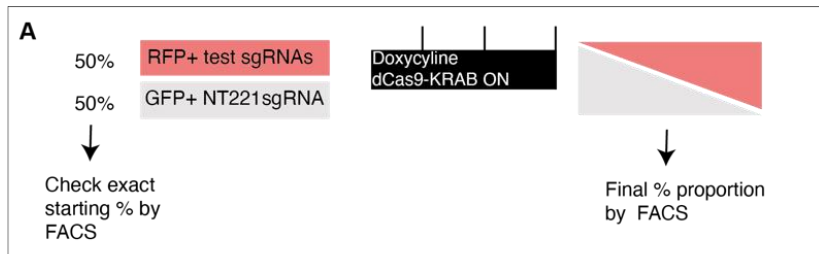
Figure 98: Single Competition experiment (P3 - 10 days) – A) Outline of the competition experiments. **B)** Representative pictures at day 10 of the competition experiment for samples NT113, NT154, TBILA sg4, TBILA sg5. **C)** Percentage quantification of GFP+ and RFP+ positive cells. All samples were tested also for the KD of the target gene (see above). For genes SH3PXD2A-AS1 and RASSF8-AS1 we are currently developing qRT-PCR assays (Box shows location of sgRNAs). Figure legend: grey bar indicates % of GFP+ cells, red bar indicates % of RFP+ cells with % indicated inside the bar.
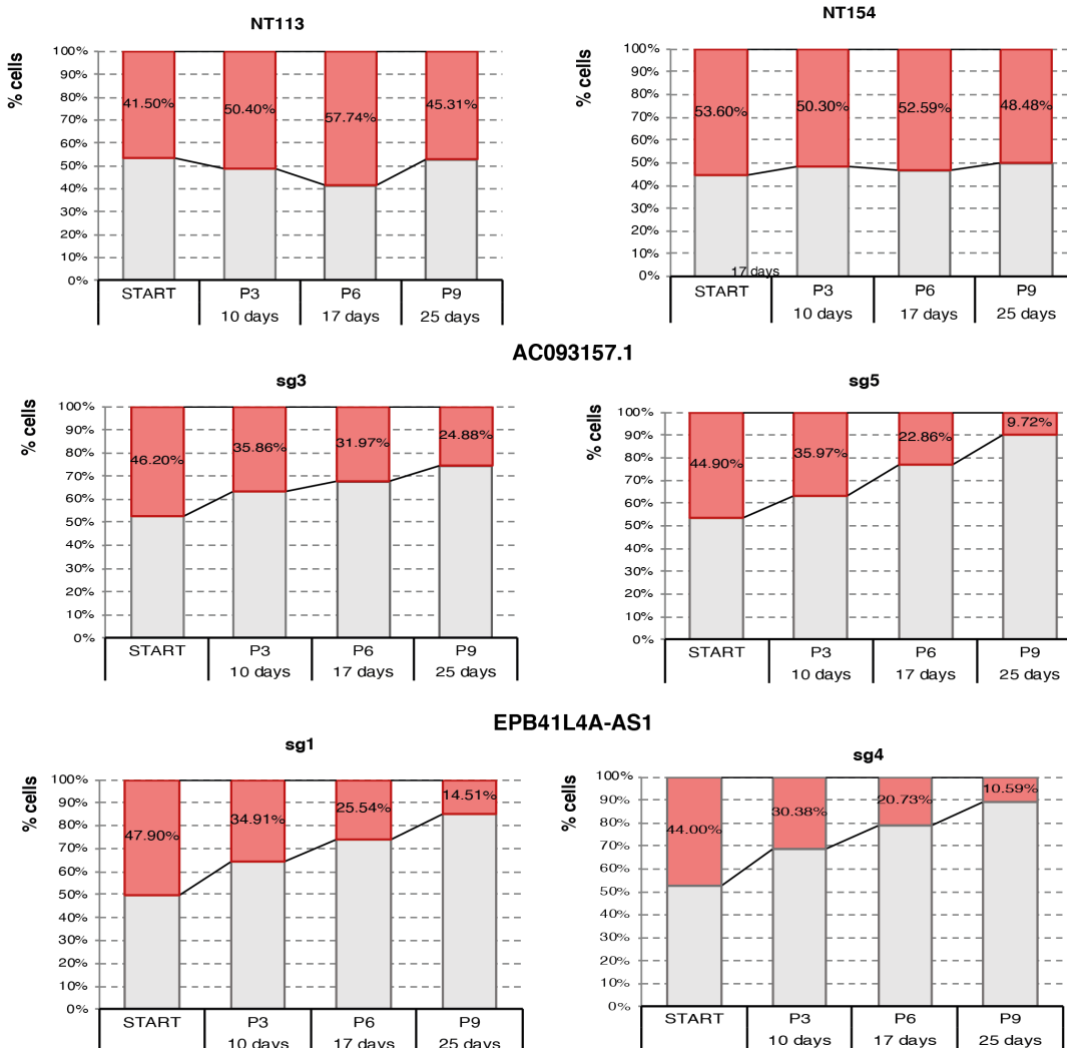
Figure 99: Single Competition experiment (P9 - 25 days) - Percentage quantification of GFP+ and RFP+ positive cells at 10 days (P3), 17 days (P6) and 24 days (P9). Samples were tested for the KD of the target gene (see above).

2.8.5 Validation of hits in 3D growth

Next, we moved to the 3D screening, using a similar approach aimed at confirming the hits and the screening results. We selected the PCG MTHFD1, which resulted DROP OUT in the 3D screening (and in vivo) but without any effect in 2D, a suitable case to evaluate if the differential effects in 3D vs 2D are confirmed using the experimental setting of single competition assays (see **2.8.4**). The sgRNAs used are highlighted in **Figure 101A**. We used two approaches to monitor the 3D growth: i) we measured the proportion of RFP+ and GFP+ cells in disaggregated spheroids by flow-cytometry (i.e., Total Cell Number); ii) we counted the number of spheroids by microscopy (i.e., Total Sphere Number and Size). We observed that the guides targeting MTHFD1 were not depleted during 2D growth, while both sgRNAs showed a decrease in their representation in the 3D growth (delta F2 – START: sg1 RFP+ cells -16.06% and sg2 RFP+ -17.08%). The non-targeting control RFP+ cells instead showed an increased percentage representation (delta F2 – START: sg1= +9.01%; sg2= +13.34),

which suggest that the effect size of MTHFD1 KD could be even greater. In conclusion, this experiment demonstrates that the perturbation of MTHFD1 selectively impairs 3D growth.
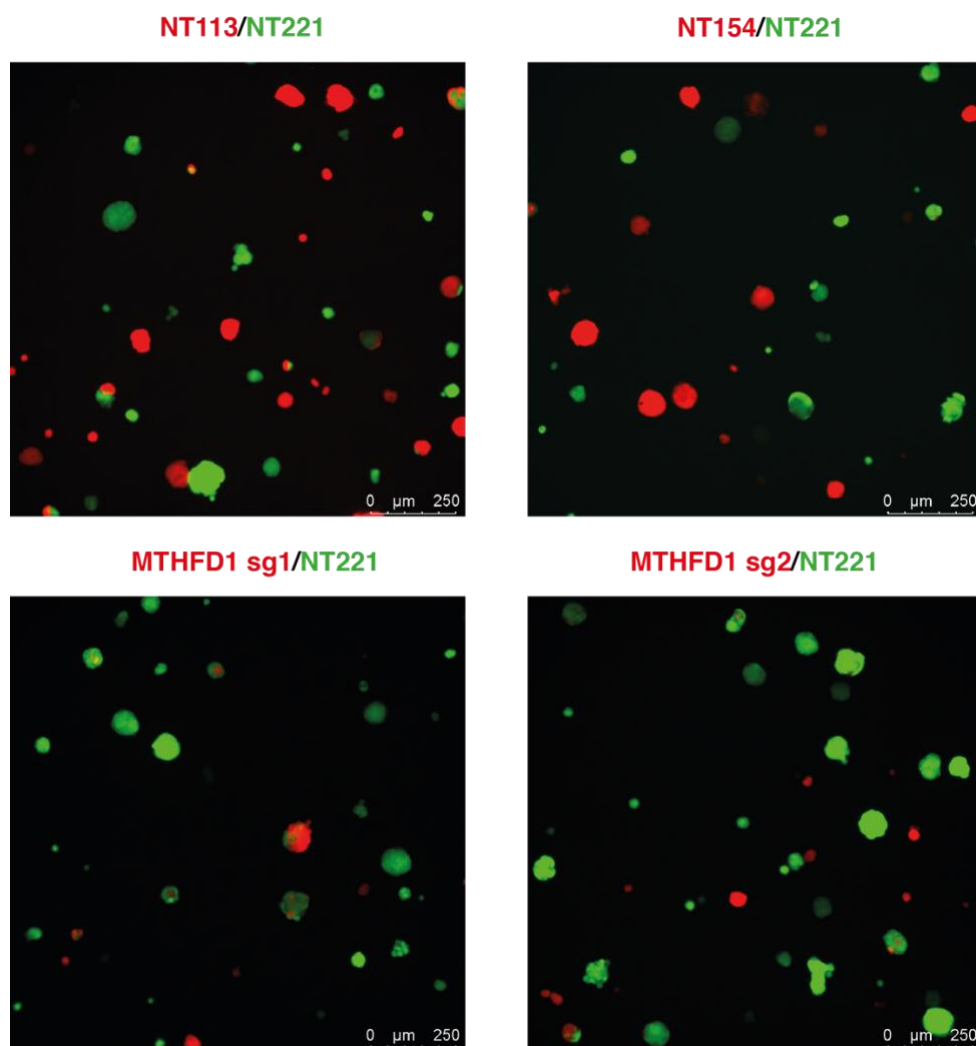
Figure 100: Competition assay of 3D growth – Representative pictures of the 2nd generation of spheroids (F2) produced in competition assays. Red spheroids express NTCs (NT113 or NT154) or sgRNAs targeting MTHFD1 (sg1 and sg2). Images acquired by Leica Thunder fluorescence widefield microscope, 10X magnification. Scale bar 250 μm. (Data in collaboration with Virigina Brancato, Nicassio Lab, IIT). Automatic counting of spheroids with ImageJ is ongoing.
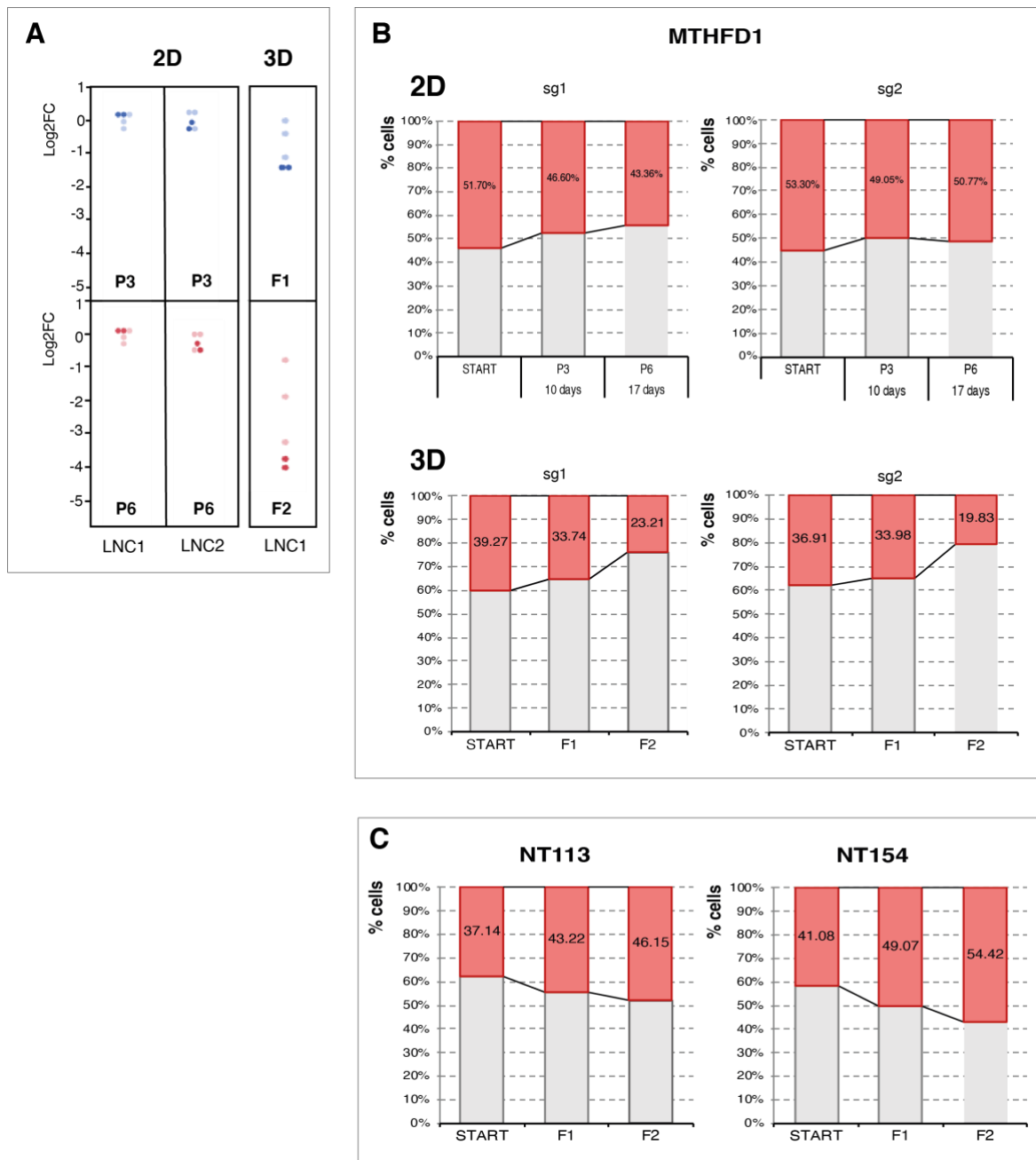
Figure 101: Specific effect of MTHFD1 pertubation in 3D – A) Log2FC values for sgRNAs targeting MTHFD1 in 2D and 3D growth screening. Highlighted sgRNAs employed in single competition experiments. **B)** Percentage quantification of GFP+ and RFP+ cells in 2D and 3D for MTHFD1 samples. **C)** Percentage quantification of GFP+ and RFP+ cells in 3D for NTC controls NT113 and NT154. Samples were tested for the KD of the target gene MTHFD1 (See above).

2.8.6   Set-up of competition experiments for drug-tolerance phenotypes

We wish to validate the impact of lncRNAs perturbation on drug-tolerant surviving cells upon Paclitaxel treatment by competition assays. As the adaptation to Paclitaxel is a process with intrinsic noise, a classic "single-colored" colony competition might result inconclusive for validation. Therefore, we focused on developing a competition assay similar for the 2D and 3D phenotypes. We performed a set-up experiment, mixing equal ratios of NTCs RFP+ and GFP+ cells (the same used before) and treating with Paclitaxel 50 nM for 72h. Then, we measured the proportion of GFP+ and RFP+ cells after replating, by flow cytometry. We observed a slight variation in the replicates for the negative controls NT113 and NT154 (**Figure 102A**). In addition, we monitored surviving colonies by imaging. We designed a macro (using ImageJ) with the help of the Imaging Unit at the European Institute of Oncology to obtain the automatic count of RFP+ and GFP+ colonies. This analysis measured no significant differences between RFP+ and GFP+ cells in terms of number of colonies, mean area and total area in 6 technical replicates (**Figure 102B, C**).

This competition assay provides multiple readouts and offers the advantage of an internal control represented by the rate of survival of GFP+ cells and an external control represented by the RFP+ NTCs, therefore we are confident that we can monitor true differences in drug response by lncRNA perturbation.
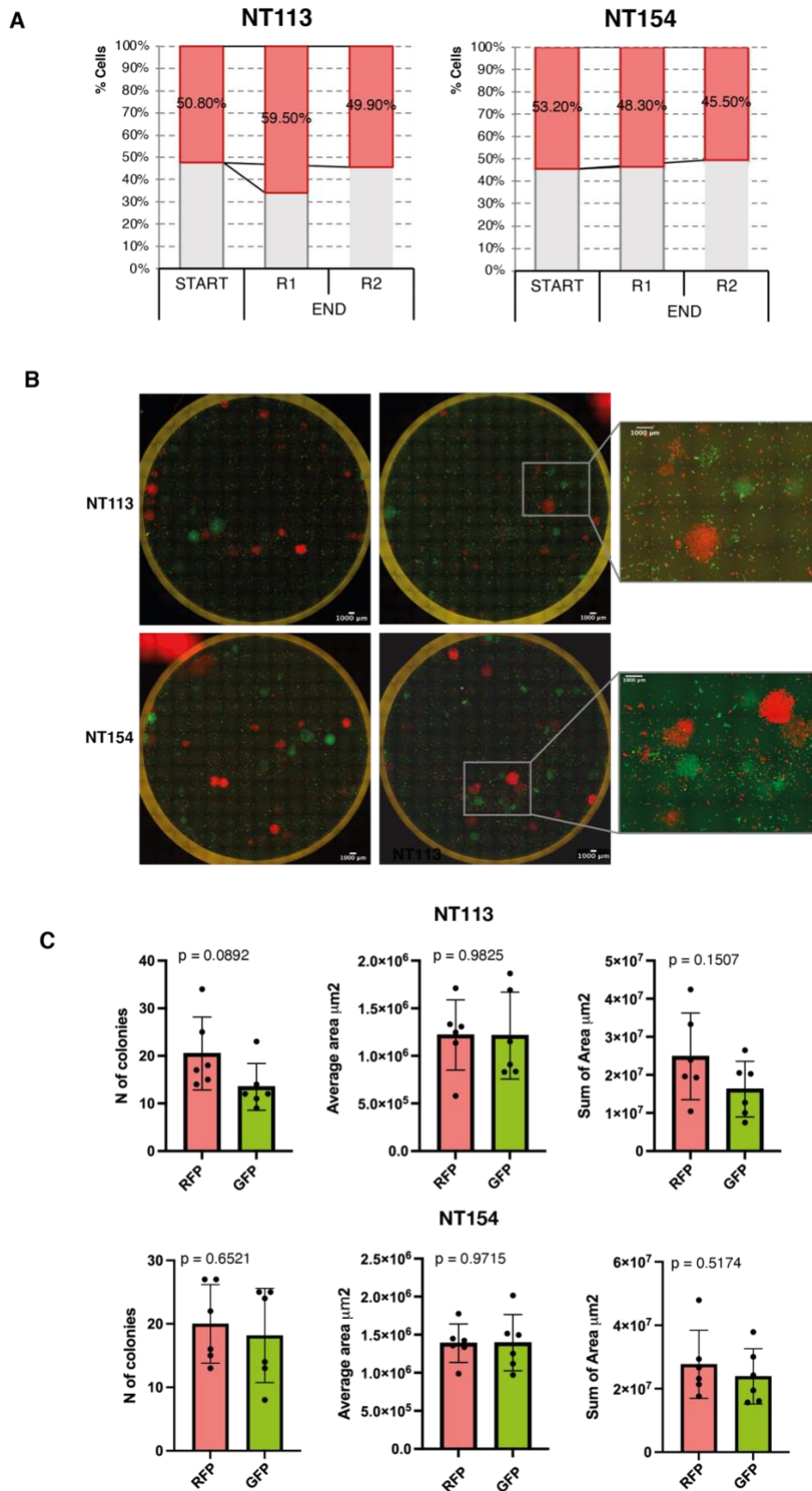
Figure 102: Competition assays to monitor the survival upon Paclitaxel treatment -

A) Percentage quantification of GFP+ and RFP+ cells upon Paclitaxel treatment by flow-cytometry (see **Methods**). **B)** Representative pictures of Paclitaxel-tolerant colonies. Picture shows that large colonies are either GFP+ or RFP+. Images acquired by Leica Thunder fluorescence widefield microscope, 4X magnification. Scale bar 1000 μm. **C)** Quantification of total number of colonies, average area, and sum of total area of GFP+ or RFP+ colonies. Error bars represent mean ±SD (n=6, technical replicates). P value by unpaired Student's t-test.

2.8.7 Set-up of lncRNA perturbation in other cell lines

In the next future, we wish to further investigate lncRNAs hits also in other cell lines, in order to understand if the phenotypes observed as result of their perturbation is specific for SUM159PT or could be reproduced in other TNBC cell lines (and more generally any other breast cancer cell line). Hence, we generated a series of 7 breast cancer cell lines, from different sub-types of breast cancer, bearing the PB-TRE-dCas9-KRAB system. In the figure below, the expression of dCas9-KRAB was evaluated in all cell lines, and it was found in the range of dCas9-KRAB activity (See **2.5.2**).
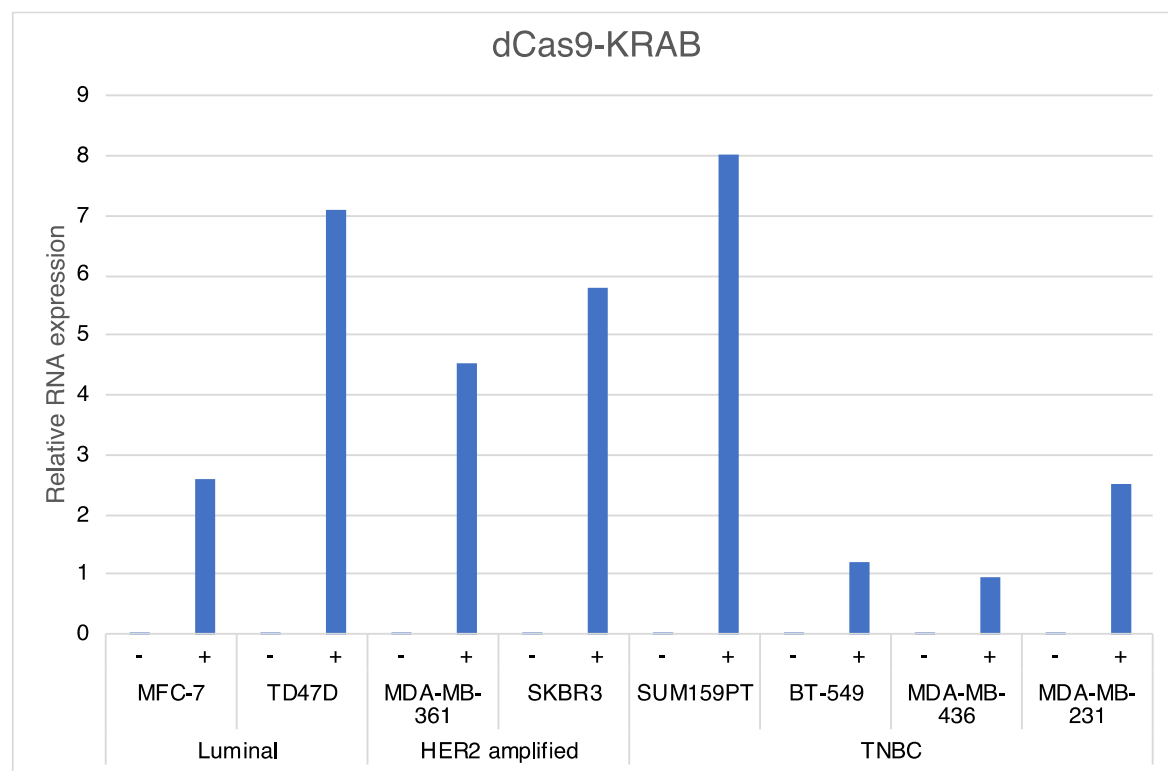
Figure 103: Expression of dCas9-KRAB in a panel of breast cancer cell lines – Expression of dCas9-KRAB measured by qRT-PCR. Data show expression of the transgene in untreated (-) and in cells treated with 1 µg/mL doxycycline for 72h. Data normalized vs *RPLP0*.

2.8.8 Transcriptomic effects of lncRNAs perturbation

To provide hints into their mechanism of action, we plan to perform a characterization of the transcriptional effects resulting from the perturbation of each lncRNA hit. The transcriptional fingerprint associated to each lncRNA hit will help i) to investigate transcriptional mechanisms related to lncRNA expression (i.e. the definition of cis- or trans-gene-regulatory functions) or more generally related to the repression of the lncRNA locus; and, at the same time, ii) to group into functional groups lncRNAs that share similar transcriptional endpoints.

Of note, we cloned the lncRNA library in a CROP-seq compatible backbone that allows to capture at the same time both sgRNA sequence and the transcriptome of individual cells by single cell RNA-sequencing (Datlinger et al., 2017). We can, therefore, map the transcriptional effects following the perturbation of lncRNA also at the single cell level, which may be critical to distinguish their effects on different or specific cellular subpopulations, such as cancer stem cells.

As a preliminary test, we analyzed by bulk and single cell RNAseq the transcriptional response following the perturbation by CRISPRi system of two DROP-OUT hits: STAT3 (a PCG and a transcription factor) and LINC01605 (**Figure 104A**). We used multiple independent sgRNAs targeting STAT3 or LINC01605 and different non-targeting sgRNAs. For bulk sequencing, we prepared a different RNA-seq library from each sample and defined the differentially expressed genes (DEGs) as usual, comparing expression data of targets vs controls (non-targeting sgRNAs) (**Figure 104A**). For sc-RNAseq, we mixed in the same single cell library equal numbers of cells carrying non-targeting sgRNAs, sgRNAs targeting STAT3 and the lncRNA LINC01605 and retrieved the type of perturbation of each individual transcriptome during data analysis (**Figure 104B**). We grouped the transcriptome according to the type of perturbation to measure transcriptional effects. As expected, we could detect a strong and specific silencing of STAT3 in those cells carrying the corresponding sgRNAs (**Figure 104C, D**), validating the procedure. Analysis on lncRNA LINC01605 effects at single cell level are still ongoing.
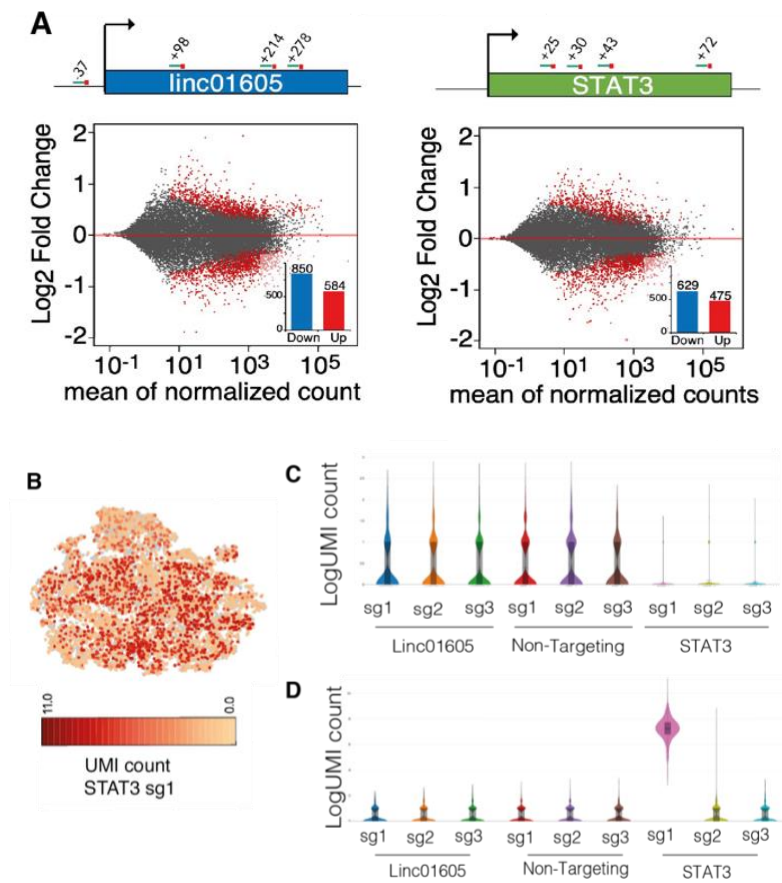
Figure 104: **Transcriptomic effects of lncRNAs perturbation** - **A)** Transcriptomic effect of the knock-down by CRISPRi of LINC01605 and STAT3 measured by bulk RNAseq. Cartoons above show the location of sgRNAs used. **B)** t-SNE showing the UMI counts for cells expressing STAT3 sg1. **C)** Expression of STAT3 in 9 groups of cells used for this experiment. Knock-down of STAT3 is specific in STAT3 sgRNAs cell populations. **D)** Expression of STAT3 sg1 is limited to one cell population.

3. Discussion

3.1 Challenges in lncRNAs characterization

In the last few years, new technological tools allowed an unprecedent understanding of complex biological phenomena and provided evidence about the pivotal regulatory role of non-coding RNA and DNA elements. Even if much effort has been put by the FANTOM and ENCODE projects to define the functional units in the genome (Abugessaisa et al., 2017; Carninci et al., 2005; Harrow et al., 2012), the interpretation of the massive transcriptional diversity is still a complicated issue, as pervasive transcription is also the source of non-functional transcripts (Guttman et al., 2009). LncRNAs are involved in a great spectrum of functions but regardless of the vast number of transcripts included in this class, the definition of their functional implication *a priori* is limited to few hints, such as the presence of conserved sequences at promoters or splice sites (Carninci et al., 2005; Derrien et al., 2012; Schlackow et al., 2017) or their cellular localization (Djebali et al., 2012). In absence of clear information about the relationship between structure and function, the identification of candidate lncRNAs involved in a selected phenotype can be extremely difficult and it is usually driven by selecting molecules according to their expression level. In general, this is a reasonable approach, since lncRNAs are expressed in a highly specific cell-type fashion. Furthermore, the high level of intra-sample heterogeneity that is typically observed in tumors or during the activation of adaptive processes (e.g., chemotolerance), strongly suggests the use of experimental approaches able to dissect their intrinsic heterogeneity, such as the characterization of transcriptional and epigenetic blueprints that are activated even in rare cellular sub-populations. Thus, even if large multiomics datasets are already available from international consortia (like TCGA, FANTOM or ENCODE), they show some limitations: the cell purity of bulk samples is variable, RNA-seq data is usually unstranded (e.g. those of TCGA) and cell sub-populations are not considered. For example, in several cancer types, samples display a variable level of infiltrate (stromal and immune cells), which can act as a confounding factor in the analysis of low expressed molecules, such as lncRNAs.

Taking all these points together, in this work we have generated in-house a series of transcriptomic profiles either characterizing at high-resolution specific cell populations (e.g. cancer stem cells; chemo-adapted clones, tumors showing high-activity of adaptive pathways), or directly resolving sample heterogeneity by single-cell analysis. Consistent literature supports the notion that cancer evolution occurs through genetic and non-genetic driving events (Marine et al., 2020). The non-genetic component of tumor progression is multilayered (epigenetic, transcriptional, metabolic components) and collectively define cell plasticity (da Silva-Diz et al., 2018; Gupta et al., 2019). Plasticity allows the adaptive

features of cancer cells. In this context, we selected lncRNAs positively associated with transcriptional determinants of adapted states in breast cancer.

We did so, by integrating the transcriptional output in models that approach the adaptation of breast cancer cells from different perspectives: the heterogeneous adaptive strategies of primary tumors, the innate adaptive properties of MaSC and the *in vitro* adaptation to the NACT drug Paclitaxel, deriving a list of candidates lncRNAs. The models selected are not comprehensive of the possible adaptive response that breast tumors accomplish. This analysis does not take into consideration the function of negative regulators of these phenotypes or incoherent activities among different models and most importantly, the expression level in a specific context is not necessarily predictive of functional engagement. However, regardless of their diversity, the models show a functional convergence towards pathways of stemness, EMT, hypoxia and inflammation and might hide common regulatory mechanisms in which lncRNAs take part. Furthermore, the selection of lncRNAs expressed across different models has practical implication as one possible long-term goal of this project is identifying markers and therapeutically exploitable targets.

3.2 Functional screening to assay adaptive properties

Genome-wide screenings can be seen as favorable approaches for the functional study of lncRNAs, applying an unbiased and high-throughput strategy for dissecting their biological relevance. Such screenings have been already successfully used to study lncRNAs involved in growth phenotypes (Liu et al., 2017b) and drug resistance (Bester et al., 2018). LncRNAs functions are critical in establishing complex phenotypes that require fine-tuning of gene expression as other works highlight (Guttman et al., 2011; Kretz et al., 2013; Loewer et al., 2010). These phenotypes are often occurring in rare population of cells and therefore, the application of large genome-wide libraries results unfeasible due to the large number of cells required to maintain the correct library representation. In this project we tried to conciliate the investigation of a conspicuous number of lncRNAs (620) in complex settings related to the acquisition of adaptive phenotypes in breast cancer cells. To this end, we exploited a pooled CRISPRi screening. CRISPRi simultaneously abrogates transcription and interferes with the function of DNA regulatory elements (Gilbert et al., 2013).

We applied NET-Cage data produced in our model system to precisely refine the TSS of candidates lncRNAs. This allowed us to design a compact library in which each candidate was targeted with 5 different sgRNAs, for an overall library size of 3451 sgRNAs. The design of the library was supported by evaluating the KD on target genes and exploiting sgRNAs previously validated in a different study (Liu et al.*,* 2017b). We retrospectively

checked the efficiency of KD of 9 target genes in the LNC-library that were called as hits in one or multiple screenings. We observed that all sgRNAs elicited the repression of the expected target, that, for 12/18 sgRNAs tested was above 80% (See **2.8.3**). Therefore, on average, we expect that 3 out of 5 sgRNAs that we designed should be able to produce strong on target effects.

The model system selected for the screenings is the TNBC cell line SUM159PT, known for being composed of heterogeneous populations of cells and displaying phenotypic plasticity (Gupta et al., 2011). We tested how the perturbation of the 620 lncRNAs affected growth and/or survival in context-specific conditions, including permissive growth on plastic dishes and with complete medium (2D) and more challenging conditions related to cancer adaptation (3D growth, chemo-response, *in vivo* tumor growth). For the sake of clarity, many issues of the set-up and analysis of the screenings have been addressed in each chapter. To sum up, in each system we characterized the drop-of-complexity introduced by the selective growth and suggested tailored approaches to interpret it. The screening pointed out common and specific hits for each screening. 11 of the PCGs included were depleted in every screening and appeared as "common essential genes" in our system. This is true also for some lncRNAs (AC128688.2, AL4050998.2, AC245014.3, TBILA). Across screenings, the observed rate of DROP-INs and DROP-OUTs was similar, and this was quite unexpected given the criteria of inclusion for lncRNAs (see **2.1**). Nonetheless, most of the DROP-INs were called only by the phenotype-GSEA analysis as *in vivo* hits (64/114) (See below).

2D and 3D screenings

The lncRNAs called as DROP-OUTs in 2D and 3D were largely shared (40/49 of the 2D hits were also 3D hits), underlying a possible common effect on proliferation for these transcripts. Nonetheless, the guides targeting the two protein coding genes MTHFD1 and STAT3 were selectively depleted in 3D and *in vivo*. We showed that the CRISPRi towards STAT3 severely impairs the sphere forming efficiency of SUM159 in the mammosphere assay (See **2.5.3**); furthermore STAT3 and MTHFD1 have a literature-supported role in stem-cell biology (Tordonato et al., 2021; Yu et al., 2014). In conclusion, the specific capture of STAT3 and MTHFD1 as hits in 3D and *in vivo* suggests that our screenings can also identify context-specific phenotypes. Intriguingly, two lncRNAs show the same specific behavior as these two PCGs. It will be interesting to address if they play a a similar function in maintenance of stem-identity.

From a technical point of view, we demonstrated how, for the 3D experiment, we largely covered the complexity of the library. Given that a certain level of cell aggregation was

observed at the used cell density (**Figure 57**) we speculate that replicating the 3D experiment with a lower cellular density might intensify the observed phenotypical effects.

Notably, we validated the growth modifying properties of a set of individual sgRNAs identified in 2D and 3D screening. In each and every case, we could precisely recapitulate the effects measured in the screening. This data, even if not formally providing an independent validation of the effects measured, as we used the same reagents of the screening, recapitulates in single the effects observed in the multiplexed setting of the screenings, supporting the soundness of MAGeCK β-scores and allowing to establish a parallelism between β-scores and biologically measurable effect sizes.

Chemo-tolerance screening

In the Paclitaxel screening, we observed that that the selective pressure introduced by the drug caused a strong drop of complexity and a less reproducible behavior of sgRNAs (See 2.6). The drug-tolerant phenotype is led by the proliferation of clonal colonies (Oren et al., 2021) that can be stochastically selected regardless of the effect of the carried sgRNA. To overcome this issue, we modified the screening to increase the number of observations collecting multiple individual samples (N=18). At the level of individual plates, we observed that a few sgRNAs became prevalent (from $6x10^6$ cells plated we retrieved 227 dominant sgRNAs on average). We built ranked sets for each gene, considering the contribution of each sgRNA in every sample in a GSEA-based analytical framework (phenotype-GSEA) (Subramanian et al., 2005). The "phenotype-GSEA" preserves the information of the contribution of each sgRNA in the clonal composition of parallel samples although bypassing differences in proliferation rates, dimming the effect of random large colonies that increase the noise of the system but rather defining hits on how often sgRNAs have "high" rank (large and small colonies) Vs "low" rank (individual cells or depletion).

It was very interesting to notice that PVT1 and MYC showed an opposite behavior in the Paclitaxel screening compared to the other screenings, with PVT1 acting as DROP-OUT and MYC as DROP-IN. This apparently incoherent behavior can be explained by the detrimental effect of MYC activity during chemo-adaptation. It has been described that cells undergoing a "diapause-like" state shut down MYC expression (Dhimolea et al., 2021). Therefore, CRISPRi-mediated inhibition of MYC might act favourably in the acquisition of the quiescence state to overcome the G2M block induced by Paclitaxel. In our CRISPRi system, the inhibition of the PVT1 promoter elicited a pro-proliferative effect in 2D, 3D and *in vivo*, due to an enhancer rewiring occurring to sustain the activity of MYC promoter (Cho et al., 2018). The behaviour of PVT1 as a DROP-OUT in the Paclitaxel screening further sustains the noxious effect of high MYC activity in the response to the drug.

Despite the need of a comprehensive validation, this observation encourages the idea that the phenotype-GSEA analytical framework provides an accurate description of the Paclitaxel screening data.

In vivo screening

The *in vivo* screening measures the TIC properties and more faithfully recapitulates he different adaptation mechanisms occurring during tumorigenesis (e.g., activation of the hypoxic response due to decreased oxygen availability compared to *in vitro* culture). The *in vivo* screening challenges the experimental reproducibility due to the strong selective pressure occurring, of which we were educated from previous clonal tracking data produced in the lab. Therefore, we decided to keep separated the gDNA of individual tumors (instead of pooling, as it is often done for such screenings), using the same phenotype-GSEA as in the Paclitaxel screening.

Overall, in this setting we called 13 genes of the "essential genes" set as DROP-OUTs. Of those, 11 were common with the 2D screening. We identified 55 DROP-OUT genes and 78 DROP-INs. The *in vivo* screening scored the highest number of DROP-INs compared to the other screenings. The empirical FDR, calculated on NTCs random sets, suggest that we experienced the stochastic DROP-IN of guides that were not-influencing TIC properties of the cell line and whose accumulation probably only reflects the strong pressure to lose guides that are detrimental for tumor growth.

Literature supported lncRNA hits

Overall, the soundness of our screening results is sustained by the identification of lncRNAs with a literature supported role. For instance, DANCR is a lncRNA DROP-OUT in 2D, 3D and *in vivo* screening. Studies involve this transcript in proliferation and stem-like properties in cell lines and xenograft models (Hu et al., 2020; Ramilowski et al., 2020; Zhen et al., 2018). TBILA (TGFb-induced lncRNA) was a common DROP-OUT in every screening and was reported as an oncogenic lncRNA in lung cancer through the activation of pro-survival pathways of the inflammatory response (Lu et al., 2018). The lncRNA EPB41L4A-AS1 is a DROP-OUT in 2D and 3D screening and the KD of this gene has been reported to inhibit proliferation, migration and EMT in colorectal cancer. LINC01605 is a 2D and 3D common DROP-OUT and literature shows that this lncRNA is involved in DNA-damage response and its KD sensitizes cells to apoptotic signals in colorectal cancer (Forrest et al., 2018).

Features of lncRNA hits

We looked at the expression of hit genes in a panel of breast cell lines and primary tumors to highlight possible specificities. Overall, most lncRNA hits display a low but consistent expression in most cellular models, Some lncRNAs show higher expression in basal-like lines as compared to luminal models, while other lncRNAs peaked in conditions associated to drug treatment.

The localization is a useful indication for lncRNAs function. We provide a global characterization of the nuclear or cytoplasmic retention of lncRNAs, as quantified from cellular fractionation. We could find lncRNAs hits that were localized either in the cytoplasmic or in the nucleus, with a slight enrichment in the case of 2D and 3D hits for the cytoplasmic compartment, which might relate to specific functions (e.g., ceRNA activity). CRISPRi has bidirectional effects (Gilbert et al., 2013; Qi et al., 2013). The observed lncRNAs hits were both intergenic and antisense. A selection of mostly anti-sense lncRNAs would have been indicative of phenotypes potentially depending on the perturbation of cognate PCGs to target lncRNAs. As mentioned before, we checked the KD of some lncRNAs hits. For those lncRNA that were in proximity with PCGs, we checked the effect on these genes as well, and observed co-repression in most cases (See **2.8.3**). Two of these genes (ZCCHC9 and EIF6) are defined as essential in DepMap (DepMap, 2019). In this regard, we checked the distance of lncRNAs hits from any PCGs. We observed that 10 lncRNAs DROP-OUTs are in the proximity of essential genes and we speculate that their phenotype might be ascribable to the targeting of these genes. However, we observed that not all PCGs that were supposed to be essential acted as such in all or some screenings, suggesting that essentiality is context-dependent and that lncRNAs proximal to putative essential PCGs does not have the be excluded *a priori* from further characterization.

3.3 Limitations of the study and future perspectives

Conversely to other interference techniques, such as antisense oligonucleotides (ASOs) that target the RNA and abrogates RNA-dependent functions of lncRNAs (Ramilowski et al., 2020; Roux et al., 2017), the simultaneous action of CRISPRi towards RNA and DNA-dependent functions, can be considered as a drawback and a limitation in the interpretation of the screening results. However, the functions of lncRNAs are often intimately connected to the activity of the locus from which they are transcribed (Gil and Ulitsky, 2019). Thus, picturing the effect of their perturbation might highlight novel functions of that locus that involve either the lncRNA or nearby genes. Therefore, at this stage we better refer to the

phenotypic characterization produced as generally dependent on the activity of the locus.

We provided an extensive characterization of lncRNAs in models of adaptive phenotypes in breast cancer. Currently our observations are limited to phenotypes related to the ability of cells of selectively proliferate in specific contexts. In the future it will be intriguing to challenge less proliferation-dependent phenotypes such as the selection of cells expressing markers of MaSCs (e.g. CD44) in basal and in transforming conditions, for instance by TGFb exposure or by testing their invasive capacity. Approaches based on cell sorting have been used in other contexts (Liu et al., 2017a; 2018a; Torre et al., 2021) but are equally challenging in preserving library complexity. A possible practical solution in order to further investigate lncRNA hits in more sophisticated biological assays would be to create a smaller sub-library containing only the relevant sgRNAs. Such sub-library might be used also to extend the results beyond the SUM159PT cell line. To this respect, we have already set-up a panel of breast cancer cell lines modified for CRISPRi where we can recapitulate the screenings or define the cell-type specificity of our screening hits.

We set up competition assays to measure in single sgRNA assays the effect that the lncRNAs perturbation have on 2D proliferation, 3D growth and Paclitaxel response. The advantage of these type of approach is that it is easily scalable thanks to imaging, and it can be used as well with orthogonal techniques of lncRNAs interference (e.g. ASOs). The panel of CRISPRi breast cancer cell lines can be used also to validate some of the results obtained in SUM159PT by these single-candidate assays.

Finally, the data here reported show that some lncRNAs are involved in the processes screened but gives no indication about their function. Our next goal is providing transcriptomic profiles upon the perturbation of selected lncRNAs, exploiting either bulk or scRNA-seq, according to the specific features of each hit. We will also characterize the corresponding genomic loci by in house available ATAC-seq/scATAC-seq data produced in SUM159PT as well as publicly available multicancer ATAC-seq data (Corces et al., 2018). This kind of molecular characterization will provide indications about the effect of the lncRNA perturbation at the level of local (within the area of activity of the dCas9-KRAB), regional or distal genomic level, allowing a better understanding of the lncRNA function within its locus as well as a deeper picture showing how the phenotypes observed are established, underling novel networks in control of breast cancer cell adaptation.

The scientific community has done great efforts in mapping genes causally implicated in cancer (e.g., DepMap). These genes are potential cancer vulnerabilities and promising therapeutic targets, but often they are not directly 'druggable'. Therefore, it is important to map their regulatory elements, which are currently largely unknown. In this scenario,

lncRNAs would be not only promising markers but also potential therapeutic molecules, whose modulation should increase the potency and specificity of other more generic treatments (like chemotherapy or epigenetic drugs).

4. Materials and Methods

4.1 Cell culture

4.1.2 SUM159pt

SUM159PT (triple negative human primary breast cancer cell lines) cells were cultured in Ham's F12 medium, supplemented with 5% Fetal Bovine Serum (FBS, South American origin), 2mM Glutamine, 5 µg/mL insulin, 1 µg/mL hydrocortisone, 10mM HEPES, 100 µg/mL of Penicillin/Streptomycin and and were cultured under an atmosphere of 10% CO_2 at 37°C.

4.1.3 SUM-KRAB

SUM159pt carrying the PB-TRE-dCas9-KRAB construct, referred in the text as SUM-KRAB (see PiggyBac transposition) were cultured in the same media of parental SUM159pt with 5% TETFREE FBS and 100 µg/mL Hygromycyn B.

4.1.4 Other lines

HMLE cells were cultured at 37°C with 5% CO2 in MEGM diluted 1:2 with Ham's F12 medium added L-Glutamine, 0.5 µg/mL Hydrocortisone, 10 µg/mL Insulin, Penicillin/Streptomycin, and 10 ng/mL of fresh EGF.

Cell lines PB-TRE-dCas9-KRAB transposed:

CELL LINE	ER	HER2	SUBTYPE	MEDIUM	Hygromycin SELECTION
SUM159PT	-	-	TNBC	Ham's F12, 5% Fetal Bovine Serum (FBS), 2mM Glutamine, 5 ug/mL insulin, 1 ug/mL hygrocortisone, 10mM HEPES, 100ug/ml of P/S	200 µg/mL
MDA-MB-231	-	-	TNBC	DMEM+10%FBS+2mM L-Glutamine+1%P/S	200 µg/mL
BT-549	-	-	TNBC	DMEM+10%FBS+2mM L-Glutamine+1%P/S	100 µg/mL
MDA-MB-436	-	-	TNBC	DMEM:Ham's F12 medium (1:1 mixture) supplemented with 2 mM L-glutamine and 10%FBS+1%P/S	100 µg/mL
T47D	+	-	LUMINAL A	DMEM+10%FBS+2mM L-Glutamine+1%P/S	100 µg/mL
MCF-7	+	-	LUMINAL A	DMEM+10%FBS+2mM L-Glutamine+1%P/S	100 µg/mL
MDA-MB-361	+	+	HER2 amplified	DMEM+10%FBS+2mM L-Glutamine+1%P/S	100 µg/mL
SK-BR-3	-	+	HER2 amplified	DMEM+10%FBS+2mM L-Glutamine+1%P/S	100 µg/mL

FBS was substituted with TETFREE FBS after transposition.

4.2 Production of stable dCas9-KRAB cell line

pHAGE-KRAB

The lentiviral delivery of dCas9-KRAB system was performed by infections with the pHage TRE-dCas9-KRAB (Addgene #50917), according to the lentiviral productions and transduction protocols provided below. Upon lentiviral infection, the construct is stably integrated. Transduced cells are selected by Neomycin resistance.

Cloning of PB-TRE-dCas9-KRAB

For the generation of the PB-TRE-dCas9-KRAB plasmid, the DNA sequence of KRAB repressor domain was amplified by PCR from the pHAGE TRE dCas9-KRAB (Addgene plasmid #50917) and cloned in frame into the PB-TRE-dCas9-VPR backbone (Addgene plasmid #63800) within the AscI/AgeI sites. The cloning was sequence-verified by Sanger Sequencing.

PiggyBac Transposition

Cells were seeded at 60-70% confluency the day before transposition in 6well plates. The following day, cells were transfected according to Lipo3000 transfection protocol (ThermoFisher Scientific) with 500 ng of transposon DNA (PB-TRE-dCas9-KRAB or PB-TRE-dCas9-vpr) and 200 ng of SuperPiggyBac transposase helper plasmid (Systems Bioscience). Specifically, the DNA was diluted in 125μL of Opti-MEM (ThermoFisher Scientific) and 5μL of P3000 transfection reagent (DNA mix), 7.5 μL of Lipofectamine3000 were diluted in 125 μL of Opti-MEM and vortex thoroughly (Lipo Mix). The DNA mix and Lipo mix were mixed 1:1 and incubated for 20 min at room temperature, then added dropwise to cells with fresh media. After at least 72h from transfection, cells were selected with 200 μg/mL Hygromycin B.

4.3 Single sgRNAs delivery

Lentiguide cloning

LentiGuide Puro (Addgene#52963) cloning occurs within the BsmBI (Esp3I) sites. Restriction digestion was performed as follows:

Starting conc.	Component	Quantity
	Plasmid	5 µg
10X	Tango Buffer (Thermo)	5 µL
10U/µL	Esp3I (Thermo)	3 µL
	DTT	20 mM
	H2O	up to 50 µL

The reaction was incubated for 2h at 37°c. Digested plasmid was purified from a 1.5% agarose gel and dephosphorylated as follows:

Starting conc.	Component	Quantity
	Plasmid gel purified	up to 1 µg
10X	Phosphatase buffer (Roche)	2 µL
1 U/µL	Alkaline Phosphatase (Roche)	1 U
	H2O	up to 20 µL

35°c 10 min, 75°C 2 min, 4°c

sgRNAs oligos were purchased as oligos with compatible ends for BsmBI cut sites by Sigma-Aldrich. Forward and Reverse oligos were annealed:

Starting conc.	Component	Quantity
10X	T4 ligation buffer (NEB)	1 µL
10U/µL	T4 Polynucleotide Kinase (PNK) (NEB)	0.5 µL
100 µM	Oligo FWD	1 µL
100 µM	Oligo RV	1 µL
	H2O	6.5 µL

Step	Temperature (°c)	Time
Phosphorylation	37	30 min
Denaturation	95	5 min
Oligo annealing	95 to 25	0.1°c/s
	4°c	Hold

Annealed Oligos were diluted in H2O 1:200 for subsequent ligation.

The ligation reaction was set-up as:

Starting conc.	Component	Quantity
	Plasmid dephos.	up to 1 µg
	Oligo annealed diluted (1:200)	1 µL
5X	DNA dilution buffer (Roche)	2 µL
	H2O	Up to 10 µL
	Mix well	
2X	T4 ligation buffer (Roche)	10 µL
5U/µL	T4 ligase (Roche)	5U

RT for 5 min.

2 µL of ligation mix were then used for subsequent transformation in Stbl3 cells. Cloning was sequence-verified by Sanger Sequencing.

PerturbSeq Cloning

The PerturbSeq GBC library couples a tracking system of passive barcodes with a sgRNAs scaffold (Addgene #85968, (Adamson et al., 2016; Dixit et al., 2016)). SgRNA cloning occurs within BlpI – BstxI sites. Restriction digestion was performed as follows:

Starting conc.	Component	Quantity
	Plasmid	5 µg
10X	Anza Buffer (Thermo)	2 µL
10U/µL	BlpI (Thermo)	5 µL
10U/µL	BstXI (Thermo)	3 µL
	H2O	up to 20 µL

Oligo annealing, plasmid dephosphorylation and ligation followed the same protocol as the LentiGuide Cloning.

pRGScribe1 Gibson cloning

We modified the library plasmid provided by Cellecta for single-sgRNAs delivery (see below). We re-transformed the library and isolated a single colony. Restriction digestion was set as follows:

Starting conc.	Component	Quantity
	Plasmid	5 µg
10X	Buffer O (Thermo)	2 µL
10U/µL	AgeI (Thermo)	5 µL
10U/µL	NdeI (Thermo)	3 µL
	H2O	up to 20 µL

2h at 37°c

The ligation occurs by Gibson cloning. We purchased by IDT eBlocks of dsDNA of 300 bp that contain the sequence of the sgRNA and 50 bp at both ends of overlapping region with the linearized plasmid. The Gibson reaction was:

Starting conc.	Component	Quantity
	Plasmid	200 ng (0.05 pmol)
10 ng/µL	eBlock	20 ng (0.1 pmol)
10X	NEBuilder HiFi Mix	10 µL
	H20	up to 20 µL

50°c for 15 min.

1/10 of the reaction is then used for subsequent transformation in Stbl3 cells. Each plasmid was sequence-verified by Sanger Sequencing.

Lentiviral production

HEK293T cells were seeded roughly 24 hours before transfection at 70-80% confluency in 6well plates in DMEM media with Glutamax, supplemented with 10% TET FREE FBS and 100 μg/mL of Penicillin/Streptomycin. The sgRNAs were packaged with psPAX2 (gag&pol) and VSV-G (envelope) plasmid transfected according to Lipofectamine3000 protocol. Below are indicated the quantities for the transfection of a well of a MW6 plates:

Lipo Mix

Component	Quantity
Opti-MEM	125 ul
Lipofectamine 3000	4 ul

DNA Mix

Component	Quantity
Optimem	125 μL
Plasmid sgRNA	1.8 μg
psPAX2	1.35 μg
vsv-g	0.45 μg
p3000	5 μL

Lipo and DNA mixes were united 1:1 and incubated at room temperature (RT) for 20 min. Mix was then added dropwise to cells with 1.5 mL of fresh media that was replaced after 12-16 hours. Supernatant was collected 24 hours after media replacement and centrifuged to remove cell debris, filtered through 0.22 μm syringe filters and used for subsequent infection.

Lentiviral transduction

150000 SUM-KRAB cells were seeded per well of 6W plates 12-16 hours prior to transduction. Polybrene was added to the undiluted lentiviral supernatant to a final concentration of 1 μg/mL. Transduction was usually carried out over-day and then media was changed. Selection was started 24 hours after the end of infection.

4.4 Evaluation of gene expression levels

Total RNA extraction

Total RNA extraction was performed with miRNeasy Micro Kit (Qiagen). Samples were prepared according to recommendations in the miRNeasy Micro Handbook. Cells were spun down and resuspended in 700 μL of QIAzol Lysis Reagent and incubated at RT for 5 min;

then 140 μL of Chloroform were added. The solution was then vortexed for 15 s and incubated at room temperature for 2-3 min. Each sample was centrifuged for 15 min at 12,000 x g at 4°C and the upper aqueous phase was transferred into a new collection tube. Then 1.5 volumes of 100% ethanol were added. Samples were transferred to RNeasy MinElute spin column and centrifuged at ≥8000xg for 15 s. The flow-through was discarded. Columns were washed with 350 μL of RWT buffer. DNase treatment was performed on columns. DNase I was prepared diluting 10 μL of DNase I (Qiagen) in 70 μL of RDD buffer and incubated at RT for 15 min. 350 μL of RWT were used for washing. Columns were further washed with 500 μL of RPE buffer in two centrifugation steps at ≥8000xg of 15 s and 2 min. Empty columns were further centrifuged to remove residual ethanol. RNA was extracted in RNase free H_2O and quantified by NanoDrop UV-Vis Spectrophotometer (Thermo Scientific). RNA was stored at -80°c.

mRNA retrotranscription

Total RNA was retro-transcribed in cDNA according to SuperScript VILO cDNA synthesis kit. The reaction was set-up as follows:

Component	Quantity
RNA	100 – 1 μg
5X Vilo Reaction Mix	4 μL
10X SuperScript Enzyme Mix	2 μg
H2O	To 20 μL

Retro-transcription reaction was carried-out in a thermal cycler with the following conditions:

Step	Temperature (°c)	Time (min)
Annealing	25	10
Extension	42	60
Enzyme inactivation	85	5
Incubation	4	hold

cDNA was stored at -20°c for subsequent amplification.

Real time quantitative PCR

Expression levels of target lncRNAs or protein coding genes were detected by Real-Time quantitative PCR (RT-qPCR) with Fast SYBR Green reagents (Life Technologies). For each cDNA sample, a reaction was set-up as follows:

Component	Quantity
2X Fast SYBR Master Mix	10 µL
Forward and Reverse primers 3.3µM	2 µL
Template cDNA	1 - 5 ng/µL
Nuclease-free H2O	up to 20 µL

Reactions were carried out in BIORAD CFX Real-Time PCR detection system under the following cycling conditions:

Step	Temperature (°c)	Time	Cycles
Initial Denaturation	95	15 min	
Denaturation	94	15 s	39
Annealing	55	30 s	39
Extension	70	30 s	39

Reactions were performed in triplicates. Rough data in terms of Ct were analyzed with Microsoft Excel (Microsoft). Data were normalized on Ct values of the housekeeping gene *Rplp0* by calculating their ΔCt. Data in the text are expressed in Relative RNA expression calculated according to the formula $2^{-\Delta Ct}$. Primer pairs were designed through computer assisted primer design software (Primer3). For lncRNAs, primer pairs were tested for the linear amplification of their target with serial dilutions of cDNA.

4.5 sgRNAs Library Design

The sgRNAs in the library come from 3 different sources:
1) sgRNAs targeting the "essential genes set" were picked from published sgRNAs libraries (Horlbeck et al., 2016);
2) lncRNAs with CAGE peaks overlapping the annotation provided in (Liu et al., 2017b) were selected from their sgRNAs set;
3) Otherwise, we designed sgRNAs with the CRISPick tool by providing genomic coordinates of the predicted TSS using standard parameters and selecting sgRNAs among the first 10 defined by the tool.

4.6 LNC-library common protocols

Production of lentiviral particles of the LNC-library

HEK293T cells were seeded in DMEM media with Glutamax, supplemented with 10% TET FREE FBS and 100 µg/mL of Penicillin/Streptomycin roughly 24 hours before transfection in 4x10cm plates. The sgRNAs library was packaged with psPAX2 (gag&pol) and VSV-G

(envelope) plasmids and transfected according to Lipofectamine3000 protocol, as suggested by the manufacturer, Cellecta. Two mixes were prepared:

Lipo Mix

Component	Quantity	
	1x10 cm	4x10 cm
Opti-MEM	737.5 µL	2940 µL
Lipofectamine 3000	23.6 µL	94.4 µL

DNA Mix

Component	Quantity	
	1x10 cm	4x10 cm
Opti-MEM	737.5 µL	2950 µL
DNA library	10.62 µg	42.48 µg
psPAX2	7.965 µg	31.86 µg
VSV-G	2.655 µg	10.62 µg
P3000	29.5 µL	118 µL

Lipo Mix and DNA Mix were mixed 1:1 and incubated at RT for 20 min. The mix was then added to HEK293T cells with fresh media dropwise (1547 µL per 10cm dish). The following day, 6 mL of media were replaced per dish. Media was supplemented with DNase I 1U/mL. 24 hours after media replacement, supernatant was collected and filtered through 0.22 µm syringe filters and ultracentrifuged at 50000xg for 2h and 10 minutes at 4°c. The final viral stock was resuspended in 240 µL of Phosphate Buffer Saline (PBS) (100X), aliquoted in 20 µL and stored at -80°c.

MOI test of the lentiviral stock

150000 SUM-KRAB cells were seeded in 6 well plates 16 hours before infection and transduced with different dilutions of the viral stock starting from the undiluted lentiviral particles (20 µL of the stock diluted in 2 mL of media to restore the original concentration), which was further diluted 1:10, 1:50, 1:100 and 1:200. The expression of tagRFP was checked 36 hours later by cytofluorimetry with BD FACSCelesta Cell analyzer.

Production of LNC1-P0 and LNC2-P0 cells

60 million SUM-KRAB cells were seeded 16 hours before infection in 15 cm dishes at a density of 3M cells per dish. Library infection was carried out in two independent events (LNC1-P0 and LNC2-P0). Cells were transduced for 7 hours in 15 mL of media per dish, with a 1:100 dilution of the stock with 1 µg/mL of polybrene to enhance infection. After 36 hours, one of the dishes was detached to check the expression of the reporter gene by

cytofluorimetry. Cells were then selected with 2 μg/mL of Puromycin, amplified and stocked in identical vials of 5M cells. P0s were always handled and stocked to maintain an average representation that guarantees a *library coverage* of 1000X. Practically, this means that no less than 3.5M cells were ever frozen or split (1000 cells x sgRNA, *library size*=3450 sgRNAs).

gDNA extraction of screening samples

For all screenings (when not otherwise indicated) each sequencing sample gDNA was collected from a frozen cell pellet of 5M cells (which was stored at -80°c until processing). gDNA extraction was performed with NucleoSpin Tissue gDNA extraction kit (Macherey-Nagel). Specifically, pellets were equilibrated at RT and resuspended in 180 μL of buffer T1 and 25 μL of Proteinase K. Samples were vigorously vortexed before overnight incubation at 56°c in thermomixer with mild shaking. 200 μL of B3 buffer was then added to samples and vortexed before incubation for 10 minutes at 70°c. Samples were then centrifuged (11000xg) to remove insoluble particles. Supernatant was collected and 210 μL of 100% ethanol were added. The sample was then placed in an extraction column and gDNA bound by centrifugation at 1100g. The membrane was washed in two subsequent centrifugation steps (11000g) with 500 μL of Buffer BW and 600 μL of buffer B5. Membrane was cleared from residual buffers in an additional centrifugation step at 11000xg. DNA was eluted in 100 μL of pre-warmed (70°c) buffer BE that was placed on column for 3 min before centrifugation. Elution was carried out in two steps to increase yield.

Amplification of the sgRNA cassette
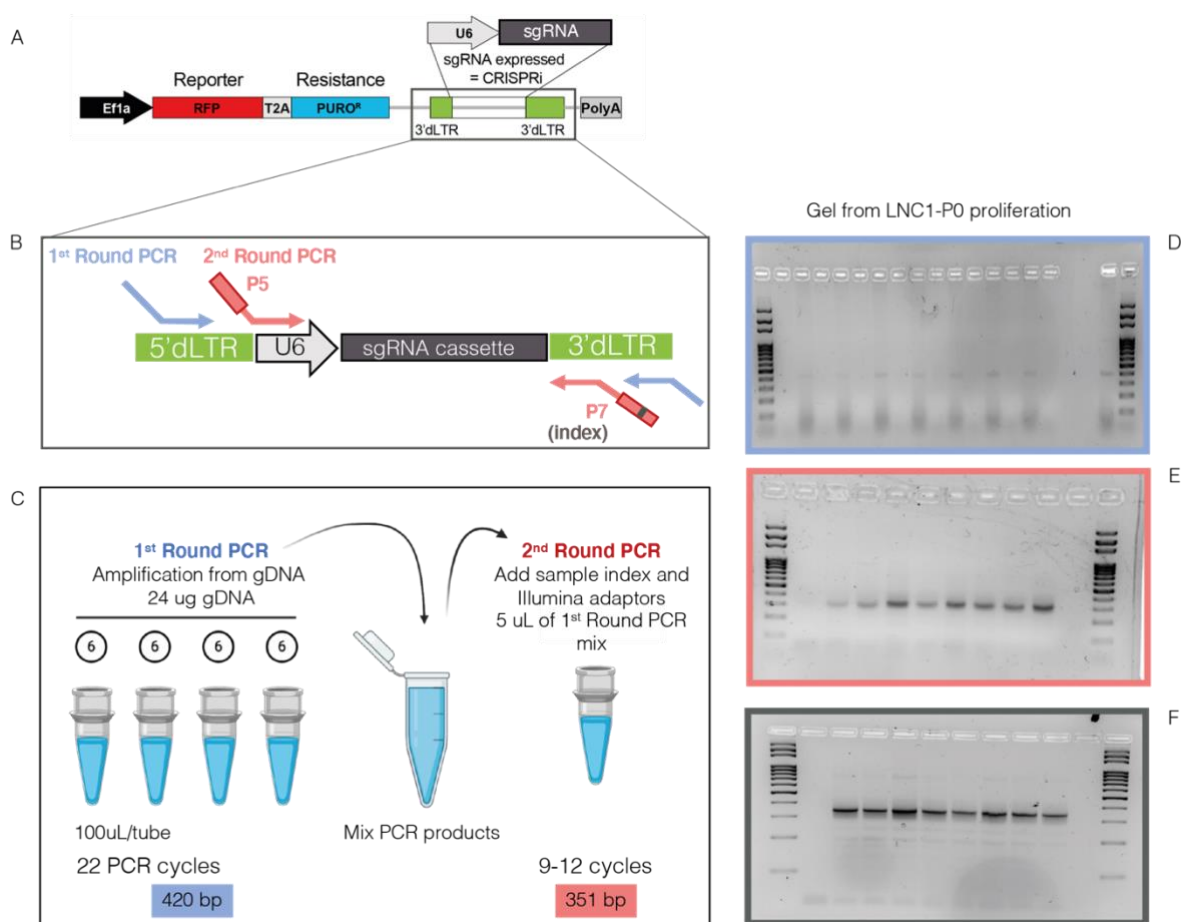
Figure 105: Schematic of sgRNAs cassette amplification from gDNA. A) and **B)** show details of the pRGScribe1 construct in which the library is cloned. **C)** The first PCR occurs in 4 parallel reactions, products are then mixed and an aliquote of the mixed products is used as the substrate for the 2nd round PCR. **D – F)** Representative PCR from one of the replicates of the proliferation screening: **D)** 1st Round PCR; **E)** 2nd Round PCR; F) 2nd round PCR after cycles optimization.

The readout of the screening(s) is the evaluation of differences in sgRNAs frequency compared to the respective P0. This is done by amplifying the cassette from the gDNA of the screening samples. As in the handling of the library, it is important to maintain the correct *library coverage* when amplifying the sgRNAs cassette to faithfully represent the occurrence of sgRNAs from the culture dishes to the sequencing data. Each cell carries on average 6.6 pg of gDNA, thus we calculated that 3.5M cells (*library coverage* =1000X) carry around 24 µg of gDNA. The amplification of the sgRNAs cassette occurs through 2 PCR steps (**Figure 105**). For the first PCR (1st Round PCR), the 24 µg of gDNA are divided in 4 identical reactions with a fixed N of cycles. For the second PCR (2nd Round PCR), the products of the 1st Round PCRs per each sample, are joined and serve as template. The 2nd Round PCR introduces the Illumina adapters for NGS sequencing. After few cycles (8) the intensity of each band is evaluated by gel-electrophoresis to optimize the number of cycles needed to

reach similar intensities and avoid over-cycling. After cycles optimization, the PCR products are gel-purified, and samples are pooled together equimolarly, thanks to a sample barcode present in the P7 adapter.

1ST Round PCR

For the 1st amplification step, for each sample, 4 equal PCR reactions were set as follows:

Starting Concentration	Component	Quantity	Final concentration
	gDNA	6 µg	60 ng/µL
10 µM	Oligo FWD	3 µL	0.625 µM
10 µM	Oligo RV	3 µL	0.625 µM
10 mM	dNTPs	2 µL	0.2 mM
10X	TitaniumTaq Buffer (Takara)	10 µL	1X
50X	TitaniumTaq polymerase (Takara)	2 µL	1X
	H2O	up to 100 µL	

Reaction was performed in a thermal cycler with the following cycling conditions:

Step	Temperature (°c)	Time	Cycles
Initial Denaturation	95	2 min	
Denaturation	95	30 s	22
Annealing	65	30 s	22
Extension	68	2 min	22
Final Extension	68	2 min	
Hold	4		

At the end of the reaction, the 4 PCRs of each sample were pooled and 10 µL (1/10 of the single reaction) were run on a 2% agarose gel to check the presence of a band at 420 bp which is usually faint (see **Figure 105**). After this quality control, 5 µL of the 1st Round PCRs served as templated for the 2nd Round PCR.

2nd Round PCR

This PCR step introduces the Illumina Adapters and ensures equal representation of each sample. After 8 cycles, the thermocycler is paused, and samples are placed on ice. An aliquot (5 µL) of each reaction is run on a 2% agarose gel. Additional PCR cycles are defined evaluating the intensity of each band. The reaction is set as indicated below:

Starting Concentration	Component	Quantity	Final concentration
	1st Round PCR product	5 µL	
10 µM	NFWD	5 µL	0.5 µM
10 µM	NRev Index	5 µL	0.5 µM
10 mM	dNTPs	2 µL	0.2 mM
10X	TitaniumTaq Buffer (Takara)	10 µL	1X
50X	TitaniumTaq polymerase (Takara)	2 µL	1X
	H2O	71 µL	

Thermal profile:

Step	Temperature (°c)	Time	Cycles
Initial Denaturation	95	2 min	
Denaturation	95	30 s	8 +
Annealing	65	30 s	8 +
Extension	68	2 min	8 +
Final Extension	68	2 min	
Hold	4		

The oligo used for these PCRs are reported below, red highlights sample indexes:

Oligo Name	Sequence 5' - 3'	STEP	Purification
Forward-XPG	CAAGGCTGTTAGAGAGATAATTGGAA	1st Round PCR	Desalt
Reverse-XPG	tagttagccagagagctcccag	1st Round PCR	Desalt
NFwd-XPG_P5	AGATACGGCGACCACCGAGATCTACACGAGATGGACTATCATATGCTTACCGTAACTTGAA	2nd Round PCR	PAGE
Nrev-XPG_P7_A	CAAGCAGAAGACGGCATACGAGATGCACGACGAGACGCAGACGAATACGACAGAGTGGTCTAAC CAGAGAGACCCAGTA	2nd Round PCR	PAGE
Nrev-XPG_P7_B	CAAGCAGAAGACGGCATACGAGATGCACGACGAGACGCAGACGAACTGATGAGAGTGGTCTAAC CAGAGAGACCCAGTA	2nd Round PCR	PAGE
Nrev-XPG_P7_C	CAAGCAGAAGACGGCATACGAGATGCACGACGAGACGCAGACGAAGCATCAAGAGTGGTCTAAC CAGAGAGACCCAGTA	2nd Round PCR	PAGE
Nrev-XPG_P7_D	CAAGCAGAAGACGGCATACGAGATGCACGACGAGACGCAGACGAAAGTCGTAGAGTGGTCTAAC CAGAGAGACCCAGTA	2nd Round PCR	PAGE
Nrev-XPG_P7_E	CAAGCAGAAGACGGCATACGAGATGCACGACGAGACGCAGACGAATCGCATAGAGTGGTCTAAC CAGAGAGACCCAGTA	2nd Round PCR	PAGE
Nrev-XPG_P7_F	CAAGCAGAAGACGGCATACGAGATGCACGACGAGACGCAGACGAACATAGCAGAGTGGTCTAAC CAGAGAGACCCAGTA	2nd Round PCR	PAGE
Nrev-XPG_P7_G	CAAGCAGAAGACGGCATACGAGATGCACGACGAGACGCAGACGAAAGCGTAAGAGTGGTCTAAC CAGAGAGACCCAGTA	2nd Round PCR	PAGE
Nrev-XPG_P7_H	CAAGCAGAAGACGGCATACGAGATGCACGACGAGACGCAGACGAAGTAGGCAGAGTGGTCTAAC CAGAGAGACCCAGTA	2nd Round PCR	PAGE
Nrev-XPG_P7_I	CAAGCAGAAGACGGCATACGAGATGCACGACGAGACGCAGACGAATTCAAGAGAGTGGTCTAAC CAGAGAGACCCAGTA	2nd Round PCR	PAGE
Nrev-XPG_P7_J	CAAGCAGAAGACGGCATACGAGATGCACGACGAGACGCAGACGAAGGATTCAGAGTGGTCTAAC CAGAGAGACCCAGTA	2nd Round PCR	PAGE
Nrev-XPG_P7_K	CAAGCAGAAGACGGCATACGAGATGCACGACGAGACGCAGACGAACCTGGAAGAGTGGTCTAAC CAGAGAGACCCAGTA	2nd Round PCR	PAGE
Nrev-XPG_P7_L	CAAGCAGAAGACGGCATACGAGATGCACGACGAGACGCAGACGAAAAGCCTAGAGTGGTCTAAC CAGAGAGACCCAGTA	2nd Round PCR	PAGE
Seq-XPG	CTTGGCTTTATATATCTTGTGGAAAGGACGAAACACCG	Sequencing	HPLC
Index-XPG	AGATGCACGACGAGACGCAGACGAA	Sequencing	HPLC

Gel extraction & pooling of screening samples

Purification of libraries occurs by gel extraction. Half of the volume of each 2nd round PCR

(50 µL) is run on a 2% agarose gel. The band at 350 bp is excised and gel extracted with

Qiagen Gel Extraction Kit. Specifically, the band was weighted and 6X volumes of the band weight of QG buffer were added (e.g. 0.6 µg → 360 µL of buffer QG). Samples were incubated on a thermomixer at 50°c for 10 minutes with mild shaking. Then, one gel volume of isopropanol was added, the mix was loaded on a column and centrifuged at 17900xg for 1 min and flow-through discarded. 500 µL of QG buffer were placed on column and centrifuged (17900xg, 1 min). 750 µL of PE buffer were then loaded on column and incubated at RT for 5 min and then centrifuged and discarded. Samples were then centrifuged to remove residual buffer. For the elution step, 30 µL of samples were then placed on column and incubated for 4 min. Samples were then centrifuged (17900xg, 1 min) and the eluted volume was used for a new round of elution with the same incubation and centrifugation. Eluted samples were then quantified by Qubit in duplicate and samples were pooled equimolarly to 10 nM (corresponding to a concentration of 1.42 ng/µL).

NGS & sgRNAs deconvolution

sgRNAs libraries were sequenced Paired-End by Novaseq 6000 with a sequencing depth of 10M reads/sample. Raw reads were processed with a custom pipeline in bash that is illustrated here step by step.

1) Given that the amplified sgRNA cassette does not contain canonical Illumina indexes, the demultiplexing step was performed manually. In order to do so, "undetermined" R1 and R2 reads were aligned side by side with the following command:

```
paste <(cat Undetermined_R1_.fastq|paste - - - -) <(cat
Undetermined_R2_001.fastq|paste - - - -) > all_samples.txt
```

2) The reads of each individual sample were retrieved by identifying its specific barcode (GCATCA in the shown example) within the R2 reads, and then saving the corresponding R1 reads, which contains the sgRNA sequence:

```
awk '{if ($8 ==
"GCATCAAGAGTGGTCTAACCAGAGAGACCCAGTACAAGCAAAAAGCAGACC") {print
$3}}' all.txt > P0_sgRNA_reads.txt
```

3) The identical unique reads were then counted, and the counted_list of all samples were merged. Finally, low counts sgRNAs were filtered out (in the proposed example those with less than 5 raw reads):

```
sort P0_sgRNA_reads.txt| uniq -c > counted_P0_sgRNA.txt
grep "" *.txt > all_count.txt
```

```
awk '$2>=5' all_count.txt > filtered.txt
```

4) The filtered list at this point contained all the detected sequences, including those originating from PCR or sequencing errors. Therefore, as a final step, we retained only the reads corresponding with sgRNA annotated in the lncRNA library. Typically, more than 80% of the raw reads could be correctly assigned to expected sgRNAs.

4.7 Screenings

Proliferation Screening

LNC1-P0 and LNC2-P0 cells were thawed and seeded in SUM-KRAB media with 100 ng/mL of Doxycycline to induce the expression of dCas9-KRAB. The screening started with $3X10^7$ cells that were split regularly 3 times per week. Multiple cell pellets of $5x10^6$ cells were collected after 10 days (P3), 17 days (P6), 24 days (P9). Pellet were processed at the end of the screening as reported above (**Amplification of the sgRNA cassette**). Here is a list of all samples sequenced for either biological (EXP1 and EXP2) and technical (e.g. P0_1 and P0_2) replicates of the screening, with the number of mapped reads obtained:
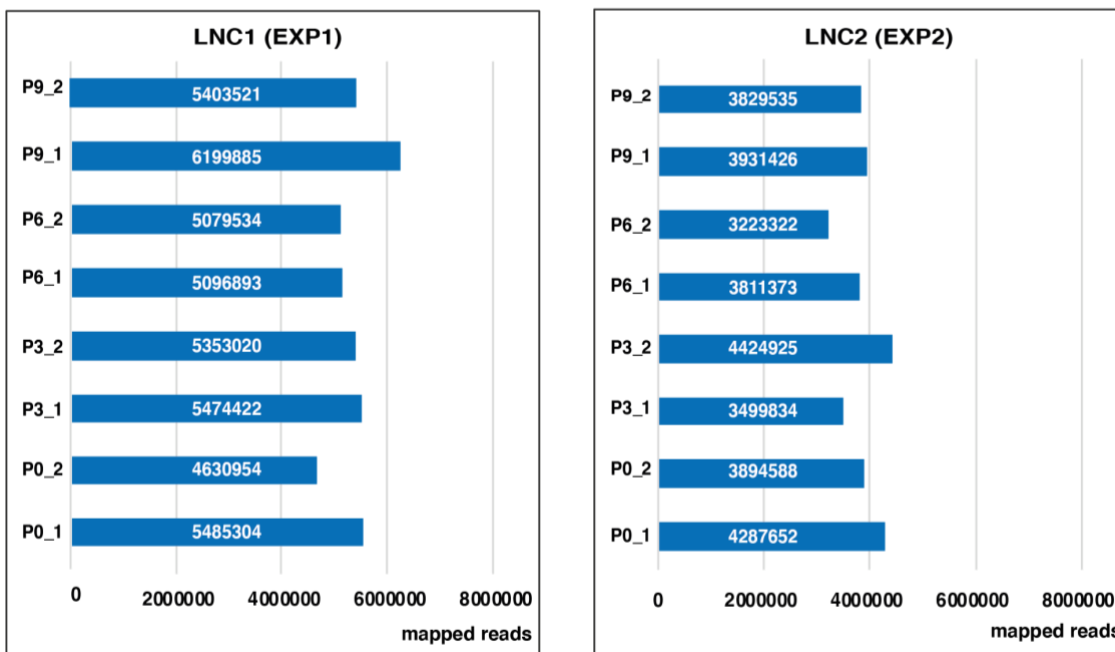
After verifying that technical replicates were nearly identical, they were merged and the resulting merged reads were used for MAGeCK analysis. β-scores for each gene were modeled by MAGeCK according to the following design matrix in a paired statistical analysis:

Samples	Baseline	time	EXP1_vs_EXP2
P0_exp1	1	0	0
P3_exp1	1	1	0
P6_exp1	1	2	0
P9_exp1	1	3	0
P0_exp2	1	0	1
P3_exp2	1	1	1
P6_exp2	1	2	1
P9_exp2	1	3	1

3D growth screening

Preparation of methylcellulose 2%

10g of Methylcellulose powder (Sigma) were autoclaved and dissolved in MEBM (Lonza) base medium for 2 hours on a warm magnetic stirrer (37°c). Solution was then polymerized on ice for 30 min. Methylcellulose was then transferred in 50 mL tubes and centrifuged at 2700xg to remove fibers and then divided in 25 mL aliquots and stored at -20°c.

MEBM Base	
Component	Final
P/S	0.01
L-Glutammine	2mM
Insuline	5 ug/mL
Hydro	0.5 µg/mL
Heparin	4 µg/mL

Seeding of F1 cells

LNC1-P0 cells were thawed and amplified. Cells were treated in 2D with 100 ng/mL of Doxycycline for 72h before 3D plating to allow an initial repression of library target genes. Cells were then detached, centrifuged, and washed with PBS before resuspension in STEM medium 2X. Cells were then filtered through a 40 µm cell strainer. Cells were mixed with the methylcellulose aliquots 1:1 to a final concentration at plating of 20000 cells/mL and 1% Methylcellulose. 5×10^5 cells were seeded in 36 polyhema dishes (15cm) for a total of 1.8×10^7 cells.

Stock	Component	Final	STEM 1X
	MEBM Base	500 mL	
50X	B27	4%	2%
100 ug/mL	EGF	40 ng/mL	20 ng/mL
100 ug/mL	FGF	40 ng/mL	20 ng/mL
1mg/mL	Doxycycline	400 ng/mL	200 ng/mL

Collection of spheroids and propagation

In each generation a total of 36 dishes were plated. Groups of spheroids from 6 dishes (i.e. "sub-pools" corresponding to $3x10^6$ seeded cells, 1000x coverage) were joined before dissociation and processed separately until plating. Briefly, each plate was diluted with 25 mL of PBS and spheroids were collected in falcon tubes. Plates were washed with additional 50 mL of PBS (methylcellulose remains attached to the dishes). After centrifugation spheroids of the same sub-pool were resuspended in 1 mL of PBS, joined in the same falcon tube and re-centrifuged. Pellet was resuspended in 1mL of Trypsin diluted 1:10 in PBS and incubated for 5 minutes at 37°c. After this time, spheroids were mechanically dissociated by pipetting gently up and down with a P200 micropipette for 3 min and placed back at 37°c. Mechanical dissociation was repeated for 1 minute and incubated for 2 more minutes at 37°c. At this point, spheroid dissociation was checked under the microscope and trypsin activity was inhibited with Trypsin inhibitor (1:1). Cells in each sub-pool were precisely counted and 2 cell pellets ($5x10^6$ cells) were collected per each sub-pool. (E.g. F1A, F1B, F1C, F1D, F1E, F1F). Subpools were finally joined in order to get a single "pool", by mixing $5x10^6$ cells from each subpool. Cell pellets were collected also from the joined pool. Cells were then filtered through 40 μm cell strainers and plated for the subsequent sphere generation, as before. The same procedure was repeated at every passage, until F4.

Sequencing and analysis of 3D samples

Samples were processed in four rounds according to the sequencing strategy shown in (See **2.5**). In each sequencing round, we re-sequenced P0, P0' (i.e. cells after 72h of doxy) and the pooled sample of the previous generation (marked by "Pre" below), in order to monitor the reproducibility of the library generation and sequencing steps. Here is a list of all samples sequenced and reads obtained:

	Sample	Reads
F1	F1_P0	8547337
	F1_P0_first	9535280
	F1_F1_pool	9721568
	F1_F1_A	8453565
	F1_F1_B	10246838
	F1_F1_C	9784766
	F1_F1_D	8993925
	F1_F1_E	10900385
F2	F2_P0	9573949
	F2_P0_first	9830708
	F2_F1pre	9167544
	F2_F2_pool	10845846
	F2_F2_A	9735526
	F2_F2_B	9462841
	F2_F2_C	10024152
	F2_F2_D	10238277
	F2_F2_E	10757890
F3	F3_F3_A	6927631
	F3_F3_B	7453961
	F3_F3_C	6370199
	F3_F3_D	7090160
	F3_F3_E	6977661
	F3_F3_F	6601852
	F3_F2pre	6490031
	F3_P0_first	6525009
	F3_P0	6324737
	F3_F3_pool	6319470
F4	F4_F4_A	6912692
	F4_F4_B	7468371
	F4_F4_C	7809800
	F4_F4_D	6778575
	F4_F4_E	7016193
	F4_F4_F	7155364
	F4_F3pre	6797192
	F4_P0_first	7209535
	F4_P0	7403609
	F4_F4_pool	7769834

β-scores for each gene were modeled by MAGeCK according to the following design matrixes (here is shown the analysis for pool samples):

Samples	Baseline	time
F1_P0	1	0
F1_P0_first	1	0
F1_F1_pool	1	1
F2_P0	1	0
F2_P0_first	1	0
F2_F1pre	1	1
F2_F2_pool	1	2
F3_P0	1	0
F3_P0_first	1	0
F4_P0	1	0
F4_P0_first	1	0
F3_F3_pool	1	3
F4_F4_pool	1	4
F3_F2pre	1	2
F4_F3pre	1	3

Paclitaxel screening

Paclitaxel screening in bulk

LNC1-P0 and LNC2-P0 cells were thawed and seeded in SUM-KRAB media with 100 ng/mL of Doxycycline to induce the expression of dCas9-KRAB for 72h. After that, cells were detached and plated in 15 cm culture dishes at a density of $3x10^6$ cells per dish (for a total of 120×10^6). Cells were treated with 50 nM Paclitaxel or DMSO 24 hours after seeding. After 72h media was changed and replenished regularly with fresh doxycycline. Paclitaxel-tolerant colonies were detached at day 12 and replated. This step was performed in order to remove the remaining dying cells. After 48 h cells were collected and stored for subsequent sgRNAs cassette amplification. DMSO treated cells were collected after 7 days, a period of time corresponding to a proliferative-window equivalent to that of Paclitaxel treated cells.

Sample	Reads	Sample	Reads
LNC1-P0	4561489	LNC2-P0	5131942
LNC1-P0	4009648	LNC2-P0	5590078
LNC1-DMSO	4117411	LNC2-DMSO	5469253
LNC1-DMSO	5359041	LNC2-DMSO	4649248
LNC1-PACLI	5356278	LNC2-PACLI	5162084
LNC1-PACLI	5406294	LNC2-PACLI	5931075

β-scores for each gene were modeled by MAGeCK according to the following design matrixes:

Samples	Baseline	Time	Pacli	EXP2_vs_EXP1
P0_EXP1	1	0	0	0
PACLI_EXP1	1	1	1	0
DMSO_EXP1	1	1	0	0
P0_EXP2	1	0	0	1
PACLI_EXP2	1	1	1	1
DMSO_EXP2	1	1	0	1

Paclitaxel screening in pools

The seeding of cells and treatment was performed just like the previous paragraph. At day 12 when colonies were detached, cells from two dishes were joined, for a total of 18 independent pools. Cells were grown for 48h and representative pellets were recovered from each pool for parallel processing. PCR amplification and preparation of samples for sequencing were performed as before. Samples were sequenced in two batches, including the same reference P0 to allow comparison. Here is a list of all samples sequenced for this experiment:

Sample	Reads	Sample	Reads
P0_b	2701597	P0f_b	3776321
P0	5991602	P0f	5651683
pool1	6163730	pool10	4487521
pool2	5623534	pool11	4713712
pool3	5508853	pool12	4996913
pool4	5399909	pool13	5192659
pool5	5612198	pool14	5330721
pool6	6447930	pool15	4466893
pool7	5261556	pool17	5476901
pool8	6921095	pool18	5654763
pool9	6376569	pool16	4239457

Data analysis

In order to apply the GSEA framework to this context, we first produced ranked lists of sgRNAs in each sample, ordering sgRNA from low to high reads count (rank1 was given to all the sgRNA without any read; a rank3450 was given the sgRNA with the highest read count). This choice was done in order to obtain in the GSEA analysis a positive NES for sgRNAs behaving as Drop-Ins and a negative NES for sgRNAs behaving as Drop-Outs. After, the 18 individual ranked lists (one for each sample, pool1-pool18) were merged into a single dataset. A unique name was attributed to each sgRNA in a given sample (e.g. sgRNA1_pool1; sgRNA1_pool2, etc…). This unique name was used in order to create the "gene-set", which included all the sgRNA targeting a given gene in all the samples. For example, in this particular experiment, having 5 sgRNA per each gene and 18 different samples, each gene-set had a size of 5*18=90. The phenotype-GSEA was finally performed by exploiting the GSEA v4.1.0 java application, using the GSEApreranked tool and using as "gene-set database" a file containing all the sgRNA defined gene-sets and as "ranked list" the merged ranked list of sgRNAs in all samples. As for non-targeting sgRNAs, given that their number was vastly exceeding that of targeting sgRNAs (i.e. 236 NTC vs 5 TC), we created 47 random subsets of NTCs of size 5. Having multiple gene-sets corresponding to negative controls allowed us to empirically measure the FDR, by counting the number of NTC sets that were considered at varying thresholds of the NES and of the FDR calculated within the GSEA framework.

In vivo screening

Intra-nipple injection

For the *in vivo* screening we injected NSG mice (NOD-*scid* IL2Rgamma[null]). The final set of 22 tumors is the result of three independent injections performed with LNC1-P0 cells which were thawed from identical cryovials. Cells were cultured for one passage after thawing and then dCas9-KRAB was induced with 100 ng/mL doxycycline in 2D for 72h. The day of the injection cells were harvested and counted. 500000 cells per mouse were then resuspended in 14 μL of PBS and mixed with 7 μL of Matrigel. Mice were anesthetized with Avertin and prepared for injection that occurred in the 4th nipple with a Hamilton syringe. The tumor development was monitored by frequent caliper measurements.

Tissue harvest and dissociation

Mice were euthanized with CO_2 when the tumor reached the approximate diameter of 1-1.2 cm. The solid tissue was rinsed with PBS, minced with a scalpel, and dissociated in digestion mix (DMEM+F12 1:1, supplemented with collagenase, hyaluronidase, 5 μg ml-1 insulin, 10 mM Hepes, 1 μg/mL hydrocortisone) for 2h on a rotating wheel at 37°c. Following a wash with base medium (DMEM:F12, 5 μg ml-1 insulin, 10 mM Hepes, 1 μg/mL hydrocortisone), cell suspension was consecutively passed through 100, 70 and 40 μm filters to remove any undigested tissue. Primary tumor cells were treated with ACK lysis buffer (Lonza) followed by resuspension in 1% BSA/PBS and processed using the Mouse Cell depletion kit (Miltenyi).

Processing of tumor samples

We obtained different numbers of cells and gDNA from each tumor (see table below). PCR were performed taking into consideration the quantity of gDNA (i.e., when we eluted >5 μg of gDNA we produced technical replicates, when we eluted less than 5 μg of gDNA we optimized the PCR cycles to obtain similarly intense PCR products (e.g., for mouse M3B, we extracted 5.97 μg of gDNA from 438000 cells. In the first round PCR we increased cycles from 8 to 12 and proceeded with 2nd round PCR as previously described).

Data Analysis

We obtained successful gDNA libraries from 24 out of 29 injected animals. We assessed the complexity of each tumor by normalizing sgRNA reads as copies-per-million and then counting the number of sgRNA detected at different CPM thresholds (>10, >100, >1000).

Set	Name	Latency	Nr Cells	Reads	Nr sgRNA >100CPM	Nr sgRNA > 10CPM
EXP1	M1C1	35	300000	6398989	935	2596
	M1B2	42	2060000	6977361	836	2405
	M1A2	45	4640000	5740471	672	1884
	M1C2	45	2560000	5181158	558	1691
EXP2	M2A2	47	2100000	6383333	603	1742
	M2B2R	47	1840000	4613101	322	1074
	M2B2L	47	430000	6258444	490	1654
	M2B1	50	3450000	7243047	605	1736
	M2C1	50	2740000	5837118	756	2096
EXP3	M3D1	42	5590000	4631384	526	1837
	M3I2	42	11220000	4705134	1158	2773
	M3E1	50	3450000	5207815	451	1299
	M3E3	50	3300000	5125183	668	2064
	M3A	50	3580000	5919034	624	1970
	M3I1	50	5800000	12142596	232	960
	M3D2	52	900000	5027948	541	1643
	M3H1	52	2800000	4603378	362	1288
	M3H2	55	3480000	4919829	471	1710
	M3B	55	438000	3756312	648	1664
	M3F	55	972000	3586072	498	1493
	M3C	57	444000	3589817	2583	3372
	M3H3	57	1220000	4126205	152	601
	M3E2	57	240000	4415905	454	1605
	M3G	57	312000	4093251	658	1673

Mice M3C and M3H3 were excluded from the analysis since they behaved as outliers in the cumulative distribution of the sgRNAs.

As for the phenotype-GSEA analysis, it was performed from raw counts, similarly as described for the Paclitaxel screening (see above).

4.8 Competition Assays

2D competition

SUM-dCas9-KRAB cells were infected with a modified version of the Cellecta pRGScribe1, in which the tagRFP cassette was substituted with a EGFP cassette by Gibson cloning.

We separately thawed and amplified RFP+ cells (sgTEST) and GFP+ cells (NT221 control). Cells were counted and mixed in equal ratios. Relative proportion of RFP and GFP+ cells were measured by FACS to define starting proportions. Cells were seeded in 6W plates, 100000 cells/well and split regularly. Proportions of RFP+ and GFP+ cells were monitored by FACS and EVOS at 10 days (P3), 17 days (P6), 24 days (P9).

3D competition

RFP+ (sgTEST) and GFP+ (NT221 control) were separately treated with 100 ng/mL doxycycline for 72h. Then cells were mixed in equal ratios and plated in 10 cm polyhema dishes for FACS analysis and replating at the density of 10000 cells/mL. Cells were also seeded in 24W plates for imaging (6000 cells/mL). For each condition and timepoint, 12 well were seeded. Initial proportion of RFP/GFP cells was acquired by FACS. Mammosphere were allowed to grow for 7 days. Spheroids from the 10 cm dishes were recovered and dissociated for replating (F2), 24W dishes were used for imaging. 3D culturing conditions and F2 propagation performed according to the protocol above.

24 well dishes were acquired by fluorescence widefield microscope (Thunder, Leica) and representative images of each condition were acquired at 10x magnification.

Paclitaxel competition

RFP+ (sgTEST) and GFP+ (NT221 control) were separately treated with 100 ng/mL doxycycline for 72h, cells were then mixed and seeded in 10 cm dishes ($2x10^6$ cells) and 6W (250000 cells/well). For each condition 2x10 cm and 6 wells were seeded. Proportion of RFP and GFP+ cells were acquired by FACS. The day after seeding, cells were treated with 50 nM Paclitaxel for 72h. At day 12, colonies in MW6 plates were acquired by fluorescence widefield microscope (Thunder, Leica) while colonies in 10 cm dishes were detached and replated in 6 wells separately from the two dishes to allow technical replicate for subsequent FACS analysis (48h later).

4.9. Cell viability analysis (IC50 estimation)

Cells were plated in 96-well plate, treated with paclitaxel (1000 nM, 100 nM, 10 nM, 1 nM, 0.1 nM) and analysed 72 hours later by measuring drugs sensitivity with ADP-GloTM Max Assay (promega, V7001) following the manufacturers' instructions. Luminescence was read using a luminometer and correlated to ADP concentrations by using an ATP- to-ADP conversion curve. Measurement of absorbance was plotted against drug concentration in a logistic regression curve. The inflection point of the curve was used to calculate the IC50 value.

4.10 Cell fractionation

This protocol was adapted from (Bhatt et al., 2012). Cells were initially harvested by removing the cell culture medium, washed once with PBS 1X and dissociated from the plates with Trypsin-EDTA solution. After detachment, cells were resuspended in culture medium and centrifuged at 1400 rpm for 5 min at 4°C. Cell pellets were then washed once with PBS 1X and resuspended with 1 volume of cytoplasmic lysis buffer. Samples were incubated on ice for 10 minutes and then laid on top of a chilled sucrose cushion (24% w/v sucrose dissolved in cytoplasmic lysis buffer without NP-40). Afterwards, samples were centrifuged at 13.000 rpm for 10 minutes at 4°C and the supernatants containing the cytoplasmic fractions were collected on fresh Eppendorf tubes. A 10% aliquot of the cytoplasmic fractions was taken for subsequent protein analysis. The nuclei pellets were then resuspended in cold glycerol buffer followed by the addition of cold nuclei lysis buffer. The samples were vortexed, incubated on ice for 1 minute and centrifuged at 13.000 rpm for 2 minutes at 4°C. Afterwards, the supernatants containing the nuclear fractions were collected in fresh Eppendorf tubes and 10% of nuclear extracts was taken for subsequent protein analysis. Finally, the resulting pellets were resuspended in 1 volume of cold PBS pipetting up and down. In this case, 10% of the solutions was taken for protein analysis. The other 90% of the resuspended chromatin pellets were prepared for RNA analysis and eventually RNA sequencing. At least one volume of TRIzolTM LS reagent (ThermoFisher Scientific) was added to the samples and the RNAs extracted according to the RNA extraction protocol above.

Buffers:

Cytoplasmic Lysis Buffer	
Reagent	**Final Concentration**
Tris-HCl pH 7.5 1M	10 mM
NaCl 5M	150 mM
NP-40 10%	0.0015
H_2O	
Before usage	
RNase Inhibitor 200 U/ml	20 U/µl
Protease Inhibitor 50X	1X

Glycerol Buffer	
Reagent	**Final Concentration**
Tris-HCl pH 7.9 1M	20 mM
NaCl 5M	75 mM
EDTA 0.5 M	0.5 mM
Glycerol 50%	96.4 mL
Before usage	
RNase Inhibitor 200 U/mL	20 U/µL
Protease Inhibitor 50X	1X
DTT 200 mM	0.85 mM

Nuclei Lysis Buffer	
Reagent	**Final Concentration**
HEPES pH 7.6 1M	20 mM
MgCl$_2$ 5M	7.5 mM
EDTA 0.5 M	0.2 mM
NaCl 5M	300 mM
Urea 7M	1M
NP-40 10%	0.01
H$_2$O	
Before usage	
RNase Inhibitor 200 U/ml	20 U/μl
Protease Inhibitor 50X	1X
DTT 200 mM	0.1 mM

4.11 Sequencing

NET-Cage

Library preparation and sequencing were performed in the lab of Yasuhiro Murakawa and peaks were called according to the published pipeline (Hirabayashi et al., 2019).

Bulk RNA-seq – Illumina (Cell fractionation and Models sequencing)

Libraries were generated according to TruSeq® Stranded Total RNA Sample Preparation Guide (Illumina), quantified by Qubit™ 2.0 Fluorometer (Invitrogen) and validated by High Sensitivity DNA Assay (2100 Bioanalyzer Instrument, Agilent), verifying samples goodness in terms of concentration and length of the fragments. RNA-sequencing (RNAseq) was performed on an Illumina HiSeq 2000 at 50 bp single-read mode with a depth of 30 million reads. Reads were then aligned to the gh38 human reference genome and differentially expressed genes identified using the Bioconductor package DESeq2 (Love et al., 2014). In order to stratify primary tumors according to the activity of "*adaptive response pathways*" we calculated a score for each signature that we considered. First, we downloaded the list of genes from Hallmark GSEA gene-sets. The "partial-EMT" set was derived from Puram et al. (Cell 2017), while the "stem" set was obtained from Lawson DA et al. (Nature 2015). Gene expression was transformed in a Z-score matrix and a score was obtained based on the sum of the Z-scores of the genes in each signature. Of note, the biological role of individual genes was considered whenever possible: the Z-score of positive regulators (e.g. EMT: CDH2, ZEB1, SLUG, etc…) was kept unchanged, while the sign of Z-score of negative regulators was changed (e.g. EMT: CDH1, EPCAM, etc…).

Bulk RNA-seq Quant-Seq Lexogen (CRISPRi KD)

RNA-seq libraries were prepared according to manufacturer's protocol. Endpoint PCR was exploited to precisely determine cycles for library amplification. Samples were sequenced on Novaseq 6000 at 50 bp pair-end. Differentially expressed genes were called according to Bluebee pipeline.

scRNA-seq

2000 cells for each condition (3 separate sgRNAs for CTRL, STAT3 KD and Linc-01605 KD), for a total of nine samples and 18000 cells, were subjected to the 10X Single Cell Protocol for transcriptome determination through droplet-based single-cell RNA-seq methodology (10X Genomics Chromium). Cells were separated into droplet emulsion using the Chromium Single Cell 3' Solution (V3.1) and Single-cell RNA-seq libraries were prepared according to the Single Cell 3' Reagent Kits User Guide (V3.1). Libraries were sequenced on a Novaseq 6000 flowcell (Illumina), with a minimum depth of 50K reads/cell. FASTQ reads were aligned, filtered and counted through the Cell Ranger pipeline (v4.0). using standard parameters. In order to specifically enrich for the sgRNA expressed by each cell population we amplified by PCR the sgRNA cassette from the library cDNA, following the protocol provided by Cellecta as follows:

1. Split in half the adaptor-ligated cDNA products generated after the adaptor ligation step.
2. Use the one half of the adaptor-ligated reaction to amplify cDNA with the P5 and P7 indexed primers following the standard scRNA protocol.
3. Use the other half of the adaptor-ligated cDNA to amplify only the CloneTracker barcoded cDNA product. Use the CloneTracker **P7-Index-Adapter-FBP1 primer sequence** in combination with the P5-Read1 primer. Run the same number of cycles as for the standard cDNA reaction, plus an additional 4 cycles. The 4 additional PCR cycles will increase the yield of the CloneTracker barcoded cDNA sequences to approximately 16-fold higher levels than other transcripts in the scRNA reaction.

   P7 i7 index Adapter 5'-
   CAAGCAGAAGACGGCATACGAGATNNNNNNNNGTGACTGGAGTTCAGACGTGTGCTC

Addendum

lncRNAs target (LNC-library)

GENE NAME	ENSG stable ID	Model of selection
ABCA9-AS1	ENSG00000231749	SHARED
AC000403.1	ENSG00000278727	ADAPTIVE

AC002398.1	ENSG00000267049	PACLI
AC002401.4	ENSG00000276851	SHARED
AC003092.1	ENSG00000236453	SHARED
AC003092.2	ENSG00000236938	PACLI
AC004585.1	ENSG00000266088	SHARED
AC004812.2	ENSG00000277283	ADAPTIVE
AC004817.3	ENSG00000269927	SHARED
AC004825.2	ENSG00000274818	ADAPTIVE
AC004982.1	ENSG00000272732	PACLI
AC005046.1	ENSG00000273055	SHARED
AC005224.3	ENSG00000266709	SHARED
AC005280.1	ENSG00000251393	PACLI
AC005332.1	ENSG00000265100	PACLI
AC005381.1	ENSG00000267243	ADAPTIVE
AC005387.2	ENSG00000269191	PACLI
AC005476.2	ENSG00000259146	PACLI
AC005498.3	ENSG00000269696	PACLI
AC005534.1	ENSG00000224903	PACLI
AC005840.3	ENSG00000256433	PACLI
AC005899.8	ENSG00000279762	single cell DEGs
AC006017.1	ENSG00000229591	PACLI
AC006058.1	ENSG00000261786	ADAPTIVE
AC006058.3	ENSG00000272121	ADAPTIVE
AC006159.1	ENSG00000235427	PACLI
AC006333.2	ENSG00000272686	PACLI
AC006538.2	ENSG00000261342	ADAPTIVE
AC007064.2	ENSG00000237689	PACLI
AC007278.2	ENSG00000236525	PACLI
AC007389.5	ENSG00000281920	ADAPTIVE
AC007405.3	ENSG00000239467	ADAPTIVE
AC007773.1	ENSG00000267213	SHARED
AC007849.1	ENSG00000242795	ADAPTIVE
AC007952.4	ENSG00000262202	PACLI
AC008014.1	ENSG00000257261	PACLI
AC008063.2	ENSG00000233397	ADAPTIVE
AC008115.3	ENSG00000275560	PACLI
AC008147.2	ENSG00000257298	ADAPTIVE
AC008267.5	ENSG00000237310	single cell DEGs
AC008632.1	ENSG00000253141	PACLI
AC008708.2	ENSG00000254187	SHARED
AC008937.1	ENSG00000225230	PACLI
AC009021.1	ENSG00000260905	PACLI
AC009053.3	ENSG00000261170	PACLI
AC009093.6	ENSG00000277999	SHARED
AC009152.1	ENSG00000275927	ADAPTIVE
AC009171.2	ENSG00000263105	PACLI

AC009269.2	ENSG00000259532	PACLI
AC009509.4	ENSG00000276261	ADAPTIVE
AC009549.1	ENSG00000270607	ADAPTIVE
AC010173.1	ENSG00000258101	PACLI
AC010327.4	ENSG00000267577	ADAPTIVE
AC010420.1	ENSG00000251206	SHARED
AC010618.4	ENSG00000269439	single cell DEGs
AC010768.1	ENSG00000254693	PACLI
AC010884.1	ENSG00000224509	PACLI
AC010967.1	ENSG00000228033	ADAPTIVE
AC011468.1	ENSG00000260160	PACLI
AC012213.4	ENSG00000271830	ADAPTIVE
AC012462.3	ENSG00000230695	SHARED
AC012640.2	ENSG00000259802	PACLI
AC013400.1	ENSG00000271991	SHARED
AC015912.3	ENSG00000274213	SHARED
AC015967.1	ENSG00000274400	CD44
AC015987.1	ENSG00000224746	CD44
AC016394.1	ENSG00000227540	single cell DEGs
AC016596.1	ENSG00000227908	PACLI
AC016924.1	ENSG00000250934	ADAPTIVE
AC016999.1	ENSG00000229915	PACLI
AC018647.1	ENSG00000227544	PACLI
AC018845.3	ENSG00000261173	PACLI
AC019117.2	ENSG00000236318	PACLI
AC019186.1	ENSG00000234584	PACLI
AC019205.1	ENSG00000229852	ADAPTIVE
AC019209.3	ENSG00000256686	PACLI
AC020611.2	ENSG00000255886	PACLI
AC020910.4	ENSG00000274104	PACLI
AC020916.1	ENSG00000267519	PACLI
AC020978.4	ENSG00000261469	PACLI
AC022537.1	ENSG00000231132	PACLI
AC022613.2	ENSG00000259523	PACLI
AC023421.2	ENSG00000267193	PACLI
AC023509.3	ENSG00000270175	PACLI
AC023830.3	ENSG00000278434	ADAPTIVE
AC023906.5	ENSG00000259712	PACLI
AC024230.1	ENSG00000248515	PACLI
AC024243.1	ENSG00000272969	ADAPTIVE
AC024909.1	ENSG00000274021	SHARED
AC025031.1	ENSG00000257496	PACLI
AC025031.4	ENSG00000275481	PACLI
AC025048.2	ENSG00000267248	ADAPTIVE
AC025181.2	ENSG00000272086	single cell DEGs
AC025822.2	ENSG00000236990	PACLI

AC026124.1	ENSG00000250280	ADAPTIVE
AC026202.2	ENSG00000233912	PACLI
AC026401.3	ENSG00000280206	PACLI
AC026979.2	ENSG00000271869	PACLI
AC027097.2	ENSG00000267787	single cell DEGs
AC027335.1	ENSG00000250001	ADAPTIVE
AC034213.1	ENSG00000250509	single cell DEGs
AC034223.2	ENSG00000251281	SHARED
AC037198.1	ENSG00000276107	PACLI
AC044781.1	ENSG00000229751	PACLI
AC055736.1	ENSG00000257500	SHARED
AC064801.1	ENSG00000278017	single cell DEGs
AC064875.1	ENSG00000225649	ADAPTIVE
AC068057.1	ENSG00000228528	ADAPTIVE
AC068594.1	ENSG00000263718	PACLI
AC068672.2	ENSG00000253377	PACLI
AC068672.3	ENSG00000254095	PACLI
AC068831.6	ENSG00000278514	ADAPTIVE
AC069544.1	ENSG00000272853	PACLI
AC073050.1	ENSG00000228222	PACLI
AC073072.1	ENSG00000179428	SHARED
AC073263.2	ENSG00000279070	SHARED
AC073365.1	ENSG00000223812	single cell DEGs
AC073591.1	ENSG00000257835	ADAPTIVE
AC074351.1	ENSG00000234707	PACLI
AC079298.1	ENSG00000278981	ADAPTIVE
AC079949.1	ENSG00000256001	PACLI
AC079949.2	ENSG00000278266	PACLI
AC087257.1	ENSG00000248100	PACLI
AC087623.3	ENSG00000272159	CD44
AC087721.1	ENSG00000259396	PACLI
AC089983.1	ENSG00000257732	single cell DEGs
AC090023.2	ENSG00000256915	ADAPTIVE
AC090152.1	ENSG00000167912	ADAPTIVE
AC090200.1	ENSG00000254357	PACLI
AC090220.1	ENSG00000266965	PACLI
AC090409.1	ENSG00000267279	SHARED
AC090559.1	ENSG00000255197	ADAPTIVE
AC090673.1	ENSG00000256083	SHARED
AC090825.1	ENSG00000259363	ADAPTIVE
AC091173.1	ENSG00000251127	PACLI
AC091182.2	ENSG00000253746	SHARED
AC091564.6	ENSG00000255680	ADAPTIVE
AC091588.1	ENSG00000264012	PACLI
AC091729.3	ENSG00000229043	single cell DEGs
AC091946.1	ENSG00000253766	PACLI

AC092117.1	ENSG00000276791	PACLI
AC092168.2	ENSG00000232034	PACLI
AC092614.1	ENSG00000227542	SHARED
AC092718.1	ENSG00000245059	PACLI
AC092807.3	ENSG00000282057	ADAPTIVE
AC092910.3	ENSG00000242622	PACLI
AC093157.1	ENSG00000233184	single cell DEGs
AC093388.1	ENSG00000272979	PACLI
AC093495.1	ENSG00000228242	PACLI
AC093583.1	ENSG00000242048	PACLI
AC093677.2	ENSG00000269559	SHARED
AC093730.1	ENSG00000250775	PACLI
AC093791.1	ENSG00000250038	ADAPTIVE
AC093797.1	ENSG00000233110	PACLI
AC096564.1	ENSG00000245293	PACLI
AC096708.3	ENSG00000273584	CD44
AC096733.2	ENSG00000273472	SHARED
AC096921.2	ENSG00000261468	PACLI
AC097451.1	ENSG00000250657	ADAPTIVE
AC097480.1	ENSG00000250064	ADAPTIVE
AC097634.1	ENSG00000270562	PACLI
AC098818.2	ENSG00000260278	SHARED
AC099066.2	ENSG00000227496	ADAPTIVE
AC099506.1	ENSG00000260862	SHARED
AC099518.6	ENSG00000278389	PACLI
AC099684.2	ENSG00000262445	SHARED
AC099753.1	ENSG00000282987	PACLI
AC099778.1	ENSG00000260236	PACLI
AC100786.1	ENSG00000246731	SHARED
AC100810.3	ENSG00000282021	SHARED
AC100858.3	ENSG00000255491	SHARED
AC103591.3	ENSG00000273338	PACLI
AC103706.1	ENSG00000261220	PACLI
AC104461.1	ENSG00000230623	PACLI
AC104779.1	ENSG00000249304	PACLI
AC104827.1	ENSG00000250062	SHARED
AC104971.1	ENSG00000267226	ADAPTIVE
AC105137.2	ENSG00000276744	PACLI
AC105206.1	ENSG00000177725	ADAPTIVE
AC105277.1	ENSG00000232453	PACLI
AC105389.2	ENSG00000249216	PACLI
AC105411.1	ENSG00000259867	ADAPTIVE
AC105935.1	ENSG00000228008	PACLI
AC106772.1	ENSG00000249650	SHARED
AC107029.2	ENSG00000244464	SHARED
AC108136.1	ENSG00000255325	ADAPTIVE

AC108142.1	ENSG00000248266	SHARED
AC109492.1	ENSG00000249061	PACLI
AC109587.1	ENSG00000244513	PACLI
AC110995.1	ENSG00000236120	ADAPTIVE
AC112206.2	ENSG00000249364	PACLI
AC112721.1	ENSG00000222022	SHARED
AC112721.2	ENSG00000222032	CD44
AC113192.2	ENSG00000255133	ADAPTIVE
AC113346.1	ENSG00000251144	SHARED
AC113383.1	ENSG00000250320	ADAPTIVE
AC114284.1	ENSG00000248927	SHARED
AC114341.1	ENSG00000278075	PACLI
AC114760.2	ENSG00000272211	PACLI
AC114763.2	ENSG00000230569	PACLI
AC116366.2	ENSG00000238160	PACLI
AC117422.1	ENSG00000248607	PACLI
AC119674.1	ENSG00000260971	PACLI
AC120024.1	ENSG00000260369	PACLI
AC120049.1	ENSG00000267414	ADAPTIVE
AC124276.2	ENSG00000255400	ADAPTIVE
AC124798.1	ENSG00000260196	ADAPTIVE
AC128709.3	ENSG00000237167	CD44
AC129507.4	ENSG00000262920	PACLI
AC131254.1	ENSG00000248964	SHARED
AC132192.2	ENSG00000268403	PACLI
AC135803.1	ENSG00000272864	ADAPTIVE
AC138331.1	ENSG00000258066	ADAPTIVE
AC144652.1	ENSG00000273117	PACLI
AC234772.2	ENSG00000269902	PACLI
AC239868.1	ENSG00000264207	PACLI
AC242426.2	ENSG00000237188	ADAPTIVE
AC244153.1	ENSG00000276170	SHARED
AC245014.3	ENSG00000276216	PACLI
AC245128.1	ENSG00000215765	PACLI
AC245128.3	ENSG00000268734	PACLI
AC245595.1	ENSG00000232527	single cell DEGs
ADAMTSL4-AS1	ENSG00000203804	PACLI
AF001548.2	ENSG00000263335	single cell DEGs
AF117829.1	ENSG00000251136	single cell DEGs
AF165147.1	ENSG00000232855	ADAPTIVE
AF250324.1	ENSG00000272566	CD44
AFAP1-AS1	ENSG00000272620	LITERATURE
AGAP2-AS1	ENSG00000255737	ADAPTIVE
AGBL1-AS1	ENSG00000260125	ADAPTIVE
AL008726.1	ENSG00000271984	ADAPTIVE
AL021393.1	ENSG00000226772	SHARED

AL021578.1	ENSG00000275894	PACLI	
AL022068.1	ENSG00000228412	PACLI	
AL023806.1	ENSG00000270638	SHARED	
AL024507.2	ENSG00000272476	ADAPTIVE	
AL031283.1	ENSG00000228140	PACLI	
AL031283.2	ENSG00000233485	PACLI	
AL031666.2	ENSG00000267882	PACLI	
AL031728.1	ENSG00000231105	PACLI	
AL031985.3	ENSG00000260920	PACLI	
AL033504.1	ENSG00000227681	ADAPTIVE	
AL033527.2	ENSG00000236546	ADAPTIVE	
AL035665.1	ENSG00000229771	PACLI	
AL049775.1	ENSG00000205562	ADAPTIVE	
AL079301.1	ENSG00000230922	PACLI	
AL096865.1	ENSG00000271857	PACLI	
AL109763.1	ENSG00000228708	PACLI	
AL117190.1	ENSG00000258399	SHARED	
AL118558.3	ENSG00000271780	PACLI	
AL121603.2	ENSG00000258738	single cell DEGs	
AL121672.3	ENSG00000273289	PACLI	
AL121772.1	ENSG00000274414	PACLI	
AL121821.1	ENSG00000258636	ADAPTIVE	
AL121972.1	ENSG00000229862	SHARED	
AL133415.1	ENSG00000234961	SHARED	
AL136146.2	ENSG00000255256	PACLI	
AL138689.1	ENSG00000274204	PACLI	
AL138789.1	ENSG00000233589	single cell DEGs	
AL138900.2	ENSG00000271736	PACLI	
AL138999.1	ENSG00000277386	CD44	
AL139220.2	ENSG00000230615	CD44	
AL157392.1	ENSG00000225112	SHARED	
AL157394.1	ENSG00000261438	SHARED	
AL157834.2	ENSG00000234026	PACLI	
AL158206.1	ENSG00000260912	ADAPTIVE	
AL160408.1	ENSG00000228044	SHARED	
AL160408.4	ENSG00000241475	SHARED	
AL161421.1	ENSG00000275202	PACLI	
AL161431.1	ENSG00000275216	SHARED	
AL162411.1	ENSG00000236924	ADAPTIVE	
AL162727.1	ENSG00000233817	SHARED	
AL353708.1	ENSG00000260360	PACLI	
AL353708.3	ENSG00000272906	PACLI	
AL353719.1	ENSG00000260475	SHARED	
AL354694.1	ENSG00000273056	PACLI	
AL355102.1	ENSG00000258412	ADAPTIVE	
AL355102.4	ENSG00000258793	ADAPTIVE	

AL355304.1	ENSG00000232618	ADAPTIVE
AL355512.1	ENSG00000273143	ADAPTIVE
AL355601.1	ENSG00000234692	ADAPTIVE
AL355607.2	ENSG00000260454	SHARED
AL355916.2	ENSG00000258926	PACLI
AL356124.1	ENSG00000226149	PACLI
AL356489.2	ENSG00000260947	ADAPTIVE
AL357054.4	ENSG00000272463	PACLI
AL357060.1	ENSG00000237499	SHARED
AL357146.1	ENSG00000236013	SHARED
AL358472.4	ENSG00000282386	PACLI
AL359091.5	ENSG00000273186	PACLI
AL359182.1	ENSG00000226334	ADAPTIVE
AL359220.1	ENSG00000247287	PACLI
AL359504.2	ENSG00000271576	ADAPTIVE
AL390755.1	ENSG00000275830	ADAPTIVE
AL391097.1	ENSG00000223492	PACLI
AL445250.1	ENSG00000225096	SHARED
AL445645.1	ENSG00000224992	PACLI
AL450306.1	ENSG00000228748	PACLI
AL450998.2	ENSG00000179743	PACLI
AL451042.2	ENSG00000227959	PACLI
AL513323.1	ENSG00000228560	SHARED
AL513327.2	ENSG00000233246	PACLI
AL513412.1	ENSG00000224972	PACLI
AL589986.2	ENSG00000236427	ADAPTIVE
AL590004.3	ENSG00000260604	CD44
AL590428.1	ENSG00000231652	SHARED
AL590617.2	ENSG00000225177	PACLI
AL590705.1	ENSG00000203279	SHARED
AL592430.1	ENSG00000232682	PACLI
AL596244.1	ENSG00000261534	ADAPTIVE
AL606469.1	ENSG00000224215	ADAPTIVE
AL606760.1	ENSG00000226754	PACLI
AL645608.7	ENSG00000272512	PACLI
AL645608.8	ENSG00000273443	PACLI
AL663074.1	ENSG00000233069	SHARED
AL691403.2	ENSG00000273565	PACLI
AL807761.4	ENSG00000230782	PACLI
AP000442.1	ENSG00000255008	PACLI
AP000692.2	ENSG00000273199	SHARED
AP000695.1	ENSG00000230479	ADAPTIVE
AP000755.1	ENSG00000255015	ADAPTIVE
AP001042.1	ENSG00000205622	ADAPTIVE
AP001542.3	ENSG00000267480	PACLI
AP001970.1	ENSG00000254710	SHARED

AP002026.1	ENSG00000246090	ADAPTIVE	
AP002075.1	ENSG00000251309	PACLI	
AP002478.1	ENSG00000266401	PACLI	
AP002748.3	ENSG00000255517	single cell DEGs	
AP002884.1	ENSG00000250303	ADAPTIVE	
AP005230.1	ENSG00000263745	PACLI	
AP005436.1	ENSG00000255102	ADAPTIVE	
AP006248.4	ENSG00000279932	PACLI	
APCDD1L-DT	ENSG00000231290	SHARED	
ARLNC1	ENSG00000260896	PACLI	
ASAP1-IT2	ENSG00000280543	ADAPTIVE	
ATP2B1-AS1	ENSG00000271614	PACLI	
ATXN2-AS	ENSG00000258099	PACLI	
AZIN1-AS1	ENSG00000253320	PACLI	
BACH1-IT3	ENSG00000234293	ADAPTIVE	
BCYRN1	ENSG00000236824	LITERATURE	
BNC2-AS1	ENSG00000234779	ADAPTIVE	
BX571846.1	ENSG00000225393	PACLI	
C1QTNF1-AS1	ENSG00000265096	SHARED	
C8orf49	ENSG00000255394	PACLI	
C20orf197	ENSG00000176659	SHARED	
CADM3-AS1	ENSG00000225670	PACLI	
CAHM	ENSG00000270419	PACLI	
CALML3-AS1	ENSG00000205488	ADAPTIVE	
CASC9	ENSG00000249395	LITERATURE	
CASC19	ENSG00000254166	PACLI	
CD44-AS1	ENSG00000255443	PACLI	
CDKN2B-AS1	ENSG00000240498	LITERATURE	
CFLAR-AS1	ENSG00000226312	PACLI	
CKMT2-AS1	ENSG00000247572	single cell DEGs	
CLDND1	ENSG00000080822	LITERATURE	
CLYBL-AS1	ENSG00000234303	PACLI	
CPNE8-AS1	ENSG00000257718	ADAPTIVE	
CPSF6	ENSG00000111605	LITERATURE	
CRIM1-DT	ENSG00000279215	ADAPTIVE	
CRNDE	ENSG00000245694	single cell DEGs	
CT70	ENSG00000230013	SHARED	
CYTOR	ENSG00000222041	LITERATURE	
DANCR	ENSG00000226950	LITERATURE	
DEPDC1-AS1	ENSG00000234264	ADAPTIVE	
DLG5-AS1	ENSG00000282770	ADAPTIVE	
DNM3OS	ENSG00000230630	SHARED	
DPYD-AS2	ENSG00000235777	PACLI	
DSCAM-AS1	ENSG00000235123	ADAPTIVE	
DSCAM-IT1	ENSG00000233756	ADAPTIVE	
EFNA3	ENSG00000143590	PACLI	

EGOT	ENSG00000235947	CD44
ELOA-AS1	ENSG00000236810	single cell DEGs
EPB41L4A-AS1	ENSG00000224032	single cell DEGs
ERVK-28	ENSG00000267696	PACLI
FALEC	ENSG00000228126	LITERATURE
FAM83C-AS1	ENSG00000235214	SHARED
FAM87A	ENSG00000182366	ADAPTIVE
FGD5-AS1	ENSG00000225733	LITERATURE
FLG-AS1	ENSG00000237975	ADAPTIVE
FOXC2-AS1	ENSG00000260944	SHARED
FOXD3-AS1	ENSG00000230798	PACLI
FOXG1-AS1	ENSG00000257126	SHARED
FOXP1-IT1	ENSG00000242094	PACLI
FRMD6-AS2	ENSG00000258537	SHARED
FTX	ENSG00000230590	LITERATURE
GABPB1-AS1	ENSG00000244879	single cell DEGs
GASAL1	ENSG00000253669	PACLI
GHET1	ENSG00000281189	LITERATURE
GHRLOS	ENSG00000240288	PACLI
GIHCG	ENSG00000262349	ADAPTIVE
GORAB-AS1	ENSG00000231407	SHARED
H19	ENSG00000130600	SHARED
HAS2-AS1	ENSG00000248690	SHARED
HCG18	ENSG00000228894	LITERATURE
HHIP-AS1	ENSG00000248890	SHARED
HMGA2-AS1	ENSG00000197301	SHARED
HOMER3-AS1	ENSG00000269019	ADAPTIVE
HOTAIRM1	ENSG00000233429	single cell DEGs
HOXC-AS3	ENSG00000251151	PACLI
HSD11B1-AS1	ENSG00000227591	SHARED
IDH1-AS1	ENSG00000231908	PACLI
IDI2-AS1	ENSG00000232656	PACLI
IER3-AS1	ENSG00000272273	SHARED
IGF2BP2-AS1	ENSG00000163915	single cell DEGs
IGFL2-AS1	ENSG00000268621	PACLI
IQCJ-SCHIP1-AS1	ENSG00000241211	PACLI
ITGB1-DT	ENSG00000229656	ADAPTIVE
ITPRIP-AS1	ENSG00000228261	PACLI
KAZN-AS1	ENSG00000234593	SHARED
KCNQ1OT1	ENSG00000269821	LITERATURE
KDM4A-AS1	ENSG00000236200	PACLI
KMT2E-AS1	ENSG00000239569	PACLI
LINC00163	ENSG00000234880	ADAPTIVE
LINC00167	ENSG00000233220	PACLI
LINC00242	ENSG00000280868	ADAPTIVE
LINC00294	ENSG00000280798	ADAPTIVE

LINC00339	ENSG00000218510	single cell DEGs	
LINC00346	ENSG00000255874	PACLI	
LINC00442	ENSG00000232685	PACLI	
LINC00460	ENSG00000233532	PACLI	
LINC00476	ENSG00000175611	SHARED	
LINC00518	ENSG00000183674	LITERATURE	
LINC00520	ENSG00000258791	ADAPTIVE	
LINC00525	ENSG00000146666	CD44	
LINC00574	ENSG00000281305	LITERATURE	
LINC00589	ENSG00000251191	CD44	
LINC00632	ENSG00000203930	ADAPTIVE	
LINC00649	ENSG00000237945	SHARED	
LINC00662	ENSG00000261824	single cell DEGs	
LINC00667	ENSG00000263753	single cell DEGs	
LINC00698	ENSG00000244342	SHARED	
LINC00702	ENSG00000233117	ADAPTIVE	
LINC00703	ENSG00000224382	CD44	
LINC00707	ENSG00000238266	PACLI	
LINC00839	ENSG00000185904	ADAPTIVE	
LINC00877	ENSG00000241163	PACLI	
LINC00882	ENSG00000242759	PACLI	
LINC00911	ENSG00000259107	ADAPTIVE	
LINC00941	ENSG00000235884	SHARED	
LINC00960	ENSG00000242516	PACLI	
LINC00964	ENSG00000249816	PACLI	
LINC01003	ENSG00000261455	PACLI	
LINC01060	ENSG00000249378	ADAPTIVE	
LINC01094	ENSG00000251442	ADAPTIVE	
LINC01111	ENSG00000254300	PACLI	
LINC01116	ENSG00000163364	ADAPTIVE	
LINC01126	ENSG00000279873	PACLI	
LINC01127	ENSG00000281162	ADAPTIVE	
LINC01133	ENSG00000224259	single cell DEGs	
LINC01151	ENSG00000253819	ADAPTIVE	
LINC01179	ENSG00000249500	SHARED	
LINC01184	ENSG00000245937	single cell DEGs	
LINC01191	ENSG00000234199	CD44	
LINC01194	ENSG00000248131	ADAPTIVE	
LINC01202	ENSG00000280776	SHARED	
LINC01291	ENSG00000204792	single cell DEGs	
LINC01293	ENSG00000230836	ADAPTIVE	
LINC01303	ENSG00000250548	PACLI	
LINC01309	ENSG00000234551	PACLI	
LINC01318	ENSG00000237790	PACLI	
LINC01322	ENSG00000244128	ADAPTIVE	
LINC01336	N/A	ADAPTIVE	

LINC01340	ENSG00000250331	ADAPTIVE
LINC01357	ENSG00000224167	ADAPTIVE
LINC01444	ENSG00000264301	SHARED
LINC01460	ENSG00000205334	PACLI
LINC01465	ENSG00000221949	PACLI
LINC01480	ENSG00000280540	ADAPTIVE
LINC01503	ENSG00000233901	ADAPTIVE
LINC01546	ENSG00000228459	ADAPTIVE
LINC01550	ENSG00000246223	ADAPTIVE
LINC01583	ENSG00000259518	SHARED
LINC01588	ENSG00000214900	PACLI
LINC01605	ENSG00000253161	SHARED
LINC01615	ENSG00000223485	ADAPTIVE
LINC01630	ENSG00000227115	ADAPTIVE
LINC01657	ENSG00000224718	PACLI
LINC01679	ENSG00000237989	ADAPTIVE
LINC01698	ENSG00000231648	ADAPTIVE
LINC01704	ENSG00000231666	ADAPTIVE
LINC01705	ENSG00000232679	ADAPTIVE
LINC01748	ENSG00000226476	PACLI
LINC01806	ENSG00000227403	PACLI
LINC01852	ENSG00000236914	PACLI
LINC01912	ENSG00000264699	ADAPTIVE
LINC01940	ENSG00000227744	PACLI
LINC01998	ENSG00000243321	ADAPTIVE
LINC02029	ENSG00000241544	PACLI
LINC02056	ENSG00000248371	SHARED
LINC02057	ENSG00000249279	SHARED
LINC02084	ENSG00000272282	ADAPTIVE
LINC02104	ENSG00000271334	SHARED
LINC02120	ENSG00000248279	PACLI
LINC02154	ENSG00000235385	ADAPTIVE
LINC02158	ENSG00000225611	LITERATURE
LINC02241	ENSG00000251629	PACLI
LINC02265	ENSG00000249241	SHARED
LINC02328	ENSG00000258733	ADAPTIVE
LINC02341	ENSG00000283554	ADAPTIVE
LINC02376	ENSG00000256292	PACLI
LINC02434	ENSG00000248370	ADAPTIVE
LINC02438	ENSG00000248238	SHARED
LINC02454	ENSG00000256268	SHARED
LINC02475	ENSG00000251350	PACLI
LINC02535	ENSG00000234155	ADAPTIVE
LINC02541	ENSG00000230943	ADAPTIVE
LINC02551	ENSG00000254842	ADAPTIVE
LINC02577	ENSG00000228742	PACLI

LINC02580	ENSG00000230587	ADAPTIVE
LINC02584	ENSG00000254417	ADAPTIVE
LINC02595	ENSG00000231566	PACLI
LINC02657	ENSG00000242147	ADAPTIVE
LINC02660	ENSG00000226005	ADAPTIVE
LINC-ROR	ENSG00000258609	LITERATURE
LINP1	ENSG00000223784	SHARED
LIVAR	ENSG00000266304	PACLI
LIX1-AS1	ENSG00000251513	ADAPTIVE
LNCOG	ENSG00000257219	ADAPTIVE
LNCTAM34A	ENSG00000234546	single cell DEGs
LPP-AS2	ENSG00000270959	PACLI
LRRC8C-DT	ENSG00000231999	single cell DEGs
LUCAT1	ENSG00000248323	LITERATURE
LURAP1L-AS1	ENSG00000235448	single cell DEGs
MAGI2-AS3	ENSG00000234456	ADAPTIVE
MALAT1	ENSG00000251562	single cell DEGs
MALINC1	ENSG00000245146	LITERATURE
MAN1B1-DT	ENSG00000268996	single cell DEGs
MANCR	ENSG00000231298	ADAPTIVE
MAST4-AS1	ENSG00000229666	SHARED
MEG3	ENSG00000214548	SHARED
MEG8	ENSG00000225746	SHARED
MEG9	ENSG00000223403	SHARED
MIAT	ENSG00000225783	CD44
MIR100HG__122101684	ENSG00000255248	PACLI
MIR100HG__122422858	ENSG00000255248	PACLI
MIR137HG	ENSG00000225206	ADAPTIVE
MIR155HG	ENSG00000234883	ADAPTIVE
MIR222HG	ENSG00000270069	PACLI
MIR381HG	ENSG00000258861	SHARED
MIR503HG	ENSG00000223749	ADAPTIVE
MIR4435-2HG	ENSG00000172965	LITERATURE
MSC-AS1	ENSG00000235531	SHARED
NEAT1	ENSG00000245532	single cell DEGs
NECTIN3-AS1	ENSG00000242242	ADAPTIVE
NKILA	ENSG00000278709	ADAPTIVE
NNT-AS1	ENSG00000248092	single cell DEGs
NORAD	ENSG00000260032	single cell DEGs
NRSN2-AS1	ENSG00000225377	single cell DEGs
NSMCE1-DT	ENSG00000245888	PACLI
NUP50-DT	ENSG00000226328	single cell DEGs
ODC1-DT	ENSG00000257135	PACLI
PANTR1	ENSG00000233639	PACLI
PPP1R26-AS1	ENSG00000225361	PACLI
PPP4R1-AS1	ENSG00000263627	PACLI

PRKCQ-AS1	ENSG00000237943	SHARED	
PSMB8-AS1	ENSG00000204261	PACLI	
PVT1	ENSG00000249859	LITERATURE	
RAP2C-AS1	ENSG00000232160	LITERATURE	
RASSF8-AS1	ENSG00000246695	ADAPTIVE	
RBMS3-AS2	ENSG00000203506	SHARED	
RGMB-AS1	ENSG00000246763	SHARED	
RNF139-AS1	ENSG00000245149	PACLI	
RP11-488P3.1	N/A	ADAPTIVE	
SAMSN1-AS1	ENSG00000223662	PACLI	
SBF2-AS1	ENSG00000246273	single cell DEGs	
SCAT1	ENSG00000267123	PACLI	
SCAT8	ENSG00000236345	SHARED	
SENCR	ENSG00000254703	ADAPTIVE	
SFTA1P	ENSG00000225383	SHARED	
SH3PXD2A-AS1	ENSG00000280693	PACLI	
SH3RF3-AS1	ENSG00000259863	PACLI	
SIX3-AS1	ENSG00000236502	CD44	
SLC25A21-AS1	ENSG00000258708	single cell DEGs	
SLIT2-IT1	ENSG00000248228	SHARED	
SMILR	ENSG00000255364	ADAPTIVE	
SNHG7	ENSG00000233016	single cell DEGs	
SNHG8	ENSG00000269893	single cell DEGs	
SNHG10	ENSG00000247092	LITERATURE	
SNHG12	ENSG00000197989	single cell DEGs	
SNHG14	ENSG00000224078	LITERATURE	
SNHG15	ENSG00000232956	single cell DEGs	
SNHG18	ENSG00000250786	SHARED	
SNHG19	ENSG00000260260	LITERATURE	
SNHG26	ENSG00000228649	ADAPTIVE	
SOX2	ENSG00000181449	LITERATURE	
STARD13-AS	ENSG00000236581	PACLI	
STARD13-IT1	ENSG00000230300	PACLI	
STXBP5-AS1	ENSG00000233452	SHARED	
SUCLA2-AS1	ENSG00000227848	ADAPTIVE	
TARID	ENSG00000227954	SHARED	
TBILA	ENSG00000261488	single cell DEGs	
TENM3-AS1	ENSG00000177822	ADAPTIVE	
TFAP2A-AS1	ENSG00000229950	PACLI	
THAP9-AS1	ENSG00000251022	single cell DEGs	
TINCR	ENSG00000223573	ADAPTIVE	
TMEM147-AS1	ENSG00000236144	single cell DEGs	
TMEM202-AS1	ENSG00000261423	PACLI	
TMPO-AS1	ENSG00000257167	ADAPTIVE	
TP73-AS1	ENSG00000227372	LITERATURE	
TRHDE-AS1	ENSG00000236333	SHARED	

TTLL11-IT1	ENSG00000237548	ADAPTIVE
TUG1	ENSG00000253352	LITERATURE
UBA6-AS1	ENSG00000248049	single cell DEGs
UBAC2-AS1	ENSG00000228889	PACLI
UCA1	ENSG00000214049	LITERATURE
USP2-AS1	ENSG00000245248	SHARED
VIM-AS1	ENSG00000229124	PACLI
VPS9D1-AS1	ENSG00000261373	single cell DEGs
Z69733.1	ENSG00000234405	PACLI
Z98885.3	ENSG00000279345	CD44
Z99572.1	ENSG00000213062	PACLI
ZEB1-AS1	ENSG00000237036	single cell DEGs
ZEB2-AS1	ENSG00000238057	PACLI
ZFAS1	ENSG00000177410	single cell DEGs
ZFPM2-AS1	ENSG00000251003	PACLI
ZNF32-AS2	ENSG00000230565	PACLI
ZNF32-AS3	ENSG00000223910	PACLI
ZNF295-AS1	ENSG00000237232	PACLI
ZNF385D-AS1	ENSG00000225542	SHARED
ZNF687-AS1	ENSG00000232671	PACLI

Bibliography

Abravanel, D.L., Belka, G.K., Pan, T.-C., Pant, D.K., Collins, M.A., Sterner, C.J., and Chodosh, L.A. (2015). Notch promotes recurrence of dormant tumor cells following HER2/neu-targeted therapy. J Clin Invest *125*, 2484–2496.

Abugessaisa, I., Noguchi, S., Hasegawa, A., Harshbarger, J., Kondo, A., Lizio, M., Severin, J., Carninci, P., Kawaji, H., and Kasukawa, T. (2017). FANTOM5 CAGE profiles of human and mouse reprocessed for GRCh38 and GRCm38 genome assemblies. Sci Data *4*, 170107.

Adamson, B., Norman, T.M., Jost, M., Cho, M.Y., Nuñez, J.K., Chen, Y., Villalta, J.E., Gilbert, L.A., Horlbeck, M.A., Hein, M.Y., et al. (2016). A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. Cell *167*, 1867–1873.e21.

Ahler, E., Sullivan, W.J., Cass, A., Braas, D., York, A.G., Bensinger, S.J., Graeber, T.G., and Christofk, H.R. (2013). Doxycycline alters metabolism and proliferation of human cell lines. PLoS ONE *8*, e64561–e64561.

Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J., and Clarke, M.F. (2003). Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci USA *100*, 3983–3988.

Ali, T., and Grote, P. (2020). Beyond the RNA-dependent function of LncRNA genes. Elife *9*, 1–14.

Almeida, M., Bowness, J.S., and Brockdorff, N. (2020). ScienceDirect The many faces of Polycomb regulation by RNA. Current Opinion in Genetics & Development *61*, 53–61.

Anderson, K.M., Anderson, D.M., McAnally, J.R., Shelton, J.M., Bassel-Duby, R., and Olson, E.N. (2016). Transcription of the non-coding RNA upperhand controls Hand2 expression and heart development. Nature *539*, 433–436.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al. (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455–461.

Asselain, B., Barlow, W., Bartlett, J., Bergh, J., m, E.B.-N., Bliss, J., Boccardo, F., Boddington, C., Bogaerts, J., Bonadonna, G., et al. (2017). Articles Long-term outcomes for neoadjuvant versus adjuvant chemotherapy in early breast cancer: meta-analysis of individual patient data from ten randomised trials. Lancet Oncology *19*, 27–39.

Bakir, B., Chiarella, A.M., Pitarresi, J.R., and Rustgi, A.K. (2020). EMT, MET, Plasticity, and Tumor Metastasis. Trends Cell Biol *30*, 764–776.

Bell, C.C., Fennell, K.A., Chan, Y.-C., Rambow, F., Yeung, M.M., Vassiliadis, D., Lara, L., Yeh, P., Martelotto, L.G., Rogiers, A., et al. (2019). Targeting enhancer switching overcomes non- genetic drug resistance in acute myeloid leukaemia. Nature Communications 1–15.

Bester, A.C., Lee, J.D., Chavez, A., Lee, Y.-R., Nachmani, D., Vora, S., Victor, J., Sauvageau, M., Monteleone, E., Rinn, J.L., et al. (2018). An Integrated Genome-wide CRISPRa Approach to Functionalize lncRNAs in Drug Resistance. Cell *173*, 649–652.e20.

Bhan, A., Soleimani, M., and Mandal, S.S. (2017). Long Noncoding RNA and Cancer: A New Paradigm. Cancer Res *77*, 3965–3981.

Bhatt, D.M., Pandya-Jones, A., Tong, A.-J., Barozzi, I., Lissner, M.M., Natoli, G., Black, D.L., and Smale, S.T. (2012). Transcript Dynamics of Proinflammatory Genes Revealed by Sequence Analysis of Subcellular RNA Fractions. Cell *150*, 279–290.

Bhola, N.E., Balko, J.M., Dugger, T.C., Kuba, M.G., Sánchez, V., Sanders, M., Stanford, J., Cook, R.S., and Arteaga, C.L. (2013). TGF-β inhibition enhances chemotherapy action against triple-negative breast cancer. J Clin Invest *123*, 1348–1358.

Bianchini, G., Balko, J.M., Mayer, I.A., Sanders, M.E., and Gianni, L. (2016). Triple-negative breast cancer: challenges and opportunities of a heterogeneous disease. Nat Rev Clin Oncol 1–17.

Bonetti, P., Climent, M., Panebianco, F., Tordonato, C., Santoro, A., Marzi, M.J., Pelicci, P.G., Ventura, A., and Nicassio, F. (2019). Dual role for miR-34a in the control of early progenitor proliferation and commitment in the mammary gland and in breast cancer. Oncogene 1–15.

Bossi, D., Cicalese, A., Dellino, G.I., Luzi, L., Riva, L., D'Alesio, C., Diaferia, G.R., Carugo, A., Cavallaro, E., Piccioni, R., et al. (2016). In Vivo Genetic Screens of Patient-Derived Tumors Revealed Unexpected Frailty of the Transformed Phenotype. Cancer Discovery 6, 650–663.

Brock, A., Chang, H., and Huang, S. (2009). Non-genetic heterogeneity- a mutation-independent driving force for the somatic evolution of tumours. Nat Rev Genet 10, 336–342.

Brockdorff, N., Ashworth, A., Kay, G.F., McCabe, V.M., Norris, D.P., Cooper, P.J., Swift, S., and Rastan, S. (1992). The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. Cell 71, 515–526.

Brown, C.J., Hendrich, B.D., Rupert, J.L., Lafrenière, R.G., Xing, Y., Lawrence, J., and Willard, H.F. (1992). The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell 71, 527–542.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev 25, 1915–1927.

Cardoso, F., Kyriakides, S., Ohno, S., Penault-Llorca, F., Poortmans, P., Rubio, I.T., Zackrisson, S., Senkus, E., and Committee, O.B.O.T.E.G. (2019). Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Triple-Negative Breast Cancer - Clinical Results and Biomarker Analysis of GeparNuevo Study 30, 1194–1220.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., et al. (2005). The transcriptional landscape of the mammalian genome. Science 309, 1559–1563.

Carrieri, C., Cimatti, L., Biagioli, M., Beugnet, A., Zucchelli, S., Fedele, S., Pesce, E., Ferrer, I., Collavin, L., Santoro, C., et al. (2012). Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. Nature 491, 454–457.

Chaffer, C.L., Brueckmann, I., Scheel, C., Kaestli, A.J., Wiggins, P.A., Rodrigues, L.O., Brooks, M., Reinhardt, F., Su, Y., Polyak, K., et al. (2011). Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state. Proc Natl Acad Sci U S A 108, 7950–7955.

Chavez, A., Tuttle, M., Pruitt, B.W., Ewen-Campen, B., Chari, R., Ter-Ovanesyan, D., Haque, S.J., Cecchi, R.J., Kowal, E.J.K., Buchthal, J., et al. (2016). Comparison of Cas9 activators in multiple species. Nat Methods 13, 563–567.

Chen, C.-K., Blanco, M., Jackson, C., Aznauryan, E., Ollikainen, N., Surka, C., Chow, A., Cerase, A., McDonel, P., and Guttman, M. (2016). Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. Science *354*, 468–472.

Chen, X., Cui, J., Yan, Z., Zhang, H., Chen, X., Wang, N., Shah, P., Deng, F., Zhao, C., Geng, N., et al. (2015). Sustained high level transgene expression in mammalian cells mediated by the optimized piggyBac transposon system. Genes Dis *2*, 96–105.

Chiu, H.-S., Somvanshi, S., Patel, E., Chen, T.-W., Singh, V.P., Zorman, B., Patil, S.L., Pan, Y., Chatterjee, S.S., Network, T.C.G.A.R., et al. (2018). Pan-Cancer Analysis of lncRNA Regulation Supports Their Targeting of Cancer Genes in Each Tumor Context. CellReports *23*, 297–312.e12.

Cho, S.W., Xu, J., Sun, R., Mumbach, M.R., Carter, A.C., Chen, Y.G., Yost, K.E., Kim, J., He, J., Nevins, S.A., et al. (2018). Promoter of lncRNA Gene PVT1 Is a Tumor-Suppressor DNA Boundary Element. Cell 1–38.

Comet, I., Riising, E.M., Leblanc, B., and Helin, K. (2016). Maintaining cell identity: PRC2-mediated regulation of transcription and cancer. Nature Reviews Cancer *16*, 803–810.

Corces, M.R., Granja, J.M., Shams, S., Louie, B.H., Seoane, J.A., Zhou, W., Silva, T.C., Groeneveld, C., Wong, C.K., Cho, S.W., et al. (2018). The chromatin accessibility landscape of primary human cancers. Science *362*, eaav1898–15.

D'Amato, N.C., Ostrander, J.H., Bowie, M.L., Sistrunk, C., Borowsky, A., Cardiff, R.D., Bell, K., Young, L.J.T., Simin, K., Bachelder, R.E., et al. (2012). Evidence for phenotypic plasticity in aggressive triple-negative breast cancer: human biology is recapitulated by a novel model system. PLoS ONE *7*, e45684.

da Silva-Diz, V., Lorenzo-Sanz, L., Bernat-Peguera, A., Lopez-Cerda, M., and Muñoz, P. (2018). Cancer cell plasticity: Impact on tumor progression and therapy response. Seminars in Cancer Biology 1–55.

Dai, X., Cheng, H., Bai, Z., and Li, J. (2017). Breast Cancer Cell Line Classification and Its Relevance with Breast Tumor Subtyping. J Cancer *8*, 3131–3141.

Datlinger, P., Rendeiro, A.F., Schmidl, C., Krausgruber, T., Traxler, P., Klughammer, J., Schuster, L.C., Kuchler, A., Alpar, D., and Bock, C. (2017). Pooled CRISPR screening with single-cell transcriptome readout. 1–10.

De Palma, M., and Hanahan, D. (2012). The biology of personalized cancer medicine: facing individual complexities underlying hallmark capabilities. Mol Oncol *6*, 111–127.

Del Mastro, L., De Placido, S., Bruzzi, P., De Laurentiis, M., Boni, C., Cavazzini, G., Durando, A., Turletti, A., Nisticò, C., Valle, E., et al. (2015). Fluorouracil and dose-dense chemotherapy in adjuvant treatment of patients with early-stage breast cancer: an open-label, 2 × 2 factorial, randomised phase 3 trial. The Lancet *385*, 1863–1872.

Del Vecchio, C.A., Feng, Y., Sokol, E.S., Tillman, E.J., Sanduja, S., Reinhardt, F., and Gupta, P.B. (2014). De-differentiation confers multidrug resistance via noncanonical PERK-Nrf2 signaling. PLoS Biol *12*, e1001945.

DepMap, B. (2019). DepMap 19Q3 Public.

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. *22*, 1775–1789.

Dhimolea, E., de Matos Simoes, R., Kansara, D., Al'Khafaji, A., Bouyssou, J., Weng, X., Sharma, S., Raja, J., Awate, P., Shirasaki, R., et al. (2021). An Embryonic Diapause-like Adaptation with Suppressed Myc Activity Enables Tumor Treatment Persistence. Cancer Cell *39*, 240–256.e11.

Dixit, A., Parnas, O., Li, B., Chen, J., Fulco, C.P., Jerby-Arnon, L., Marjanovic, N.D., Dionne, D., Burks, T., Raychowdhury, R., et al. (2016). Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell *167*, 1853–1857.e17.

Dixon, J.R., Gorkin, D.U., and Ren, B. (2016). Chromatin Domains: The Unit of Chromosome Organization. Molecular Cell *62*, 668–680.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. Nature *489*, 101–108.

Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nature Biotechnology *34*, 184–191.

Dominguez, A.A., Lim, W.A., and Qi, L.S. (2015). Beyond editing: repurposing CRISPR–Cas9 for precision genome regulation and interrogation. Nature Publishing Group 1–11.

Dontu, G., Abdallah, W.M., Foley, J.M., Jackson, K.W., Clarke, M.F., Kawamura, M.J., and Wicha, M.S. (2003). In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. Genes Dev *17*, 1253–1270.

Echeverria, G.V., Ge, Z., Seth, S., Zhang, X., Jeter-Jones, S., Zhou, X., Cai, S., Tu, Y., McCoy, A., Peoples, M., et al. (2019). Resistance to neoadjuvant chemotherapy in triple-negative breast cancer mediated by a reversible drug-tolerant state. Science Translational Medicine *11*, eaav0936.

Echeverria, G.V., Powell, E., Seth, S., Ge, Z., Carugo, A., Bristow, C., Peoples, M., Robinson, F., Qiu, H., Shao, J., et al. (2018). High-resolution clonal mapping of multi-organ metastasis in triple negative breast cancer. Nature Communications 1–17.

El-Shennawy, L., Dubrovskyi, O., Kastrati, I., Danes, J.M., Zhang, Y., Whiteley, H.E., Creighton, C.J., and Frasor, J. (2018). Coactivation of Estrogen Receptor and IKKβ Induces a Dormant Metastatic Phenotype in ER-Positive Breast Cancer. Cancer Res *78*, 974–984.

Eliyatkın, N., Yalçın, E., Zengel, B., Aktaş, S., and Vardar, E. (2015). Molecular Classification of Breast Carcinoma: From Traditional, Old-Fashioned Way to A New Age, and A New Way. J Breast Health *11*, 59–66.

Emert, B.L., Cote, C.J., Torre, E.A., Dardani, I.P., Jiang, C.L., Jain, N., Shaffer, S.M., and Raj, A. (2021). Variability within rare cell states enables multiple paths toward drug resistance. 1–21.

Engreitz, J.M., Haines, J.E., Perez, E.M., Munson, G., Chen, J., Kane, M., McDonel, P.E., Guttman, M., and Lander, E.S. (2016). Local regulation of gene expression by lncRNA promoters, transcription and splicing. Nature Publishing Group *539*, 452–455.

Engreitz, J.M., Pandya-Jones, A., McDonel, P., Shishkin, A., Sirokman, K., Surka, C., Kadri, S., Xing, J., Goren, A., Lander, E.S., et al. (2013). The Xist lncRNA exploits three-dimensional genome architecture to spread across the X chromosome. Science *341*, 1237973.

Esposito, R., Bosch, N., Lanzós, A., Polidori, T., Pulido-Quetglas, C., and Johnson, R. (2019). Hacking the Cancer Genome: Profiling Therapeutically Actionable Long Non-coding RNAs Using CRISPR-Cas9 Screening. Cancer Cell *35*, 545–557.

Ethier, S.P., Mahacek, M.L., Gullick, W.J., Frank, T.S., and Weber, B.L. (1993). Differential Isolation of Normal Luminal Mammary Epithelial Cells and Breast Cancer Cells from Primary and Metastatic Sites Using Selective Media. Cancer Res *53*, 627.

Feng, Y.-X., Sokol, E.S., Del Vecchio, C.A., Sanduja, S., Claessen, J.H.L., Proia, T.A., Jin, D.X., Reinhardt, F., Ploegh, H.L., Wang, Q., et al. (2014). Epithelial-to-mesenchymal transition activates PERK-eIF2α and sensitizes cells to endoplasmic reticulum stress. Cancer Discovery *4*, 702–715.

Fillmore, C.M., and Kuperwasser, C. (2008). Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy. Breast Cancer Res *10*, R25–13.

Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806–811.

Flanagan, L., Van Weelden, K., Ammerman, C., Ethier, S.P., and Welsh, J. (1999). SUM-159PT cells: a novel estrogen independent human breast cancer model system. Breast Cancer Res Treat *58*, 193–204.

Flavahan, W.A., Gaskell, E., and Bernstein, B.E. (2017). Epigenetic plasticity and the hallmarks of cancer. Science *357*, 1–10.

Forrest, M.E., Saiakhova, A., Beard, L., Buchner, D.A., Scacheri, P.C., LaFramboise, T., Markowitz, S., and Khalil, A.M. (2018). Colon Cancer-Upregulated Long Non-Coding RNA lincDUSP Regulates Cell Cycle Genes and Potentiates Resistance to Apoptosis. Scientific Reports 1–12.

Foulkes, W.D., Smith, I.E., and Reis-Filho, J.S. (2010). Triple-negative breast cancer. N Engl J Med *363*, 1938–1948.

Gasperini, M., Hill, A.J., McFaline-Figueroa, J.L., Martin, B., Kim, S., Zhang, M.D., Jackson, D., Leith, A., Schreiber, J., Noble, W.S., et al. (2019). A Genome-wide Framework for Mapping Gene Regulation via Cellular Genetic Screens. Cell *176*, 377–390.e19.

Gerlinger, M., Rowan, A.J., Horswell, S., Math, M., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., et al. (2012). Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. N Engl J Med *366*, 883–892.

Ghajar, C.M., Peinado, H., Mori, H., Matei, I.R., Evason, K.J., Brazier, H., Almeida, D., Koller, A., Hajjar, K.A., Stainier, D.Y.R., et al. (2013). The perivascular niche regulates breast tumour dormancy. Nature Cell Biology *15*, 807–817.

Ghildiyal, M., and Zamore, P.D. (2009). Small silencing RNAs: an expanding universe. Nat Rev Genet *10*, 94–108.

Gil, N., and Ulitsky, I. (2019). Regulation of gene expression by cis-acting long non-coding RNAs. Nature Publishing Group 1–16.

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647–661.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-Mediated Modular RNA-Guided Regulation of Transcription in Eukaryotes. Cell *154*, 442–451.

Gjorevski, N., and Nelson, C.M. (2011). Integrated morphodynamic signalling of the mammary gland. 1–13.

Goff, L.A., and Rinn, J.L. (2015). Linking RNA biology to lncRNAs. Genome Res. *25*, 1456–1465.

Grelet, S., Link, L.A., Howley, B., Obellianne, C., Palanisamy, V., Gangaraju, V.K., Diehl, J.A., and Howe, P.H. (2017). A regulated PNUTS mRNA to lncRNA splice switch mediates EMT and tumour progression. Nature Cell Biology *19*, 1105–1115.

Gudenas, B.L., and Wang, L. (2018). Prediction of LncRNA Subcellular Localization with Deep Learning from Sequence Features. Scientific Reports *8*, 16385.

Gupta, P.B., Fillmore, C.M., Jiang, G., Shapira, S.D., Tao, K., Kuperwasser, C., and Lander, E.S. (2011). Stochastic State Transitions Give Rise to Phenotypic Equilibrium in Populations of Cancer Cells. Cell *146*, 633–644.

Gupta, P.B., Pastushenko, I., Skibinski, A., Blanpain, C., and Kuperwasser, C. (2019). Phenotypic Plasticity: Driver of Cancer Initiation, Progression, and Therapy Resistance. Stem Cell *24*, 65–78.

Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature *458*, 223–227.

Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B., Ach, R., Bruhn, L., et al. (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. Nature Publishing Group 1–11.

Hacisuleyman, E., Goff, L.A., Trapnell, C., Williams, A., Henao-Mejia, J., Sun, L., McClanahan, P., Hendrickson, D.G., Sauvageau, M., Kelley, D.R., et al. (2014). Topological organization of multichromosomal regions by the long intergenic noncoding RNA Firre. Nat Struct Mol Biol *21*, 198–206.

Han, K., Pierce, S.E., Li, A., Spees, K., Anderson, G.R., Seoane, J.A., Lo, Y.-H., Dubreuil, M., Olivas, M., Kamber, R.A., et al. (2020). CRISPR screens in cancer spheroids identify 3D growth-specific vulnerabilities. Nature 1–30.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646–674.

Hangauer, M.J., Viswanathan, V.S., Ryan, M.J., Bole, D., Eaton, J.K., Matov, A., Galeas, J., Dhruv, H.D., Berens, M.E., Schreiber, S.L., et al. (2017). Drug-tolerant persister cancer cells are vulnerable to GPX4 inhibition. Nature *551*, 247–250.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. *22*, 1760–1774.

Hay, E.D. (2005). The mesenchymal cell, its role in the embryo, and the remarkable signaling mechanisms that create it. Dev Dyn *233*, 706–720.

Hirabayashi, S., Bhagat, S., Matsuki, Y., Takegami, Y., Uehata, T., Kanemaru, A., Itoh, M., Shirakawa, K., Takaori-Kondo, A., Takeuchi, O., et al. (2019). NET-CAGE characterizes the dynamics and topology of human transcribed cis-regulatory elements. Nature Genetics 1–17.

Hnisz, D., Weintraub, A.S., Day, D.S., Valton, A.-L., Bak, R.O., Li, C.H., Goldmann, J., Lajoie, B.R., Fan, Z.P., Sigova, A.A., et al. (2016). Activation of proto-oncogenes by disruption of chromosome neighborhoods. Science *351*, 1454–1458.

Hodges, C., Kirkland, J.G., and Crabtree, G.R. (2016). The Many Roles of BAF (mSWI/SNF) and PBAF Complexes in Cancer. Cold Spring Harb Perspect Med *6*, a026930.

Hon, C.-C., Ramilowski, J.A., Harshbarger, J., Bertin, N., Rackham, O.J.L., Gough, J., Denisenko, E., Schmeier, S., Poulsen, T.M., Severin, J., et al. (2017). An atlas of human long non-coding RNAs with accurate 5' ends. Nature *543*, 199–204.

Horlbeck, M.A., Gilbert, L.A., Villalta, J.E., Adamson, B., Pak, R.A., Chen, Y., Fields, A.P., Park, C.Y., Corn, J.E., Kampmann, M., et al. (2016). Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. Elife *5*, 1–20.

Horlbeck, M.A., Liu, S.J., Chang, H.Y., Lim, D.A., and Weissman, J.S. (2020). Fitness effects of CRISPR/Cas9-targeting of long noncoding RNA genes. Nature Biotechnology 1–7.

Hu, X., Peng, W.-X., Zhou, H., Jiang, J., Zhou, X., Huang, D., Mo, Y.-Y., and Yang, L. (2020). IGF2BP2 regulates DANCR by serving as an N6-methyladenosine reader. Cell Death Differ *27*, 1782–1794.

Hu, Z., Fan, C., Oh, D.S., Marron, J.S., He, X., Qaqish, B.F., Livasy, C., Carey, L.A., Reynolds, E., Dressler, L., et al. (2006). The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics *7*, 96.

Hudson, W.H., Pickard, M.R., de Vera, I.M.S., Kuiper, E.G., Mourtada-Maarabouni, M., Conn, G.L., Kojetin, D.J., Williams, G.T., and Ortlund, E.A. (2014). Conserved sequence-specific lincRNA-steroid receptor interactions drive transcriptional repression and direct cell fate. Nature Communications *5*, 5395.

Iyer, M.K., Niknafs, Y.S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., Barrette, T.R., Prensner, J.R., Evans, J.R., Zhao, S., et al. (2015). The landscape of long noncoding RNAs in the human transcriptome. Nature Genetics *47*, 199–208.

Jansen, M.P.H.M., Knijnenburg, T., Reijm, E.A., Simon, I., Kerkhoven, R., Droog, M., Velds, A., van Laere, S., Dirix, L., Alexi, X., et al. (2013). Hallmarks of aromatase inhibitor drug resistance revealed by epigenetic profiling in breast cancer. Cancer Res *73*, 6632–6641.

Kearns, N.A., Genga, R.M.J., Enuameh, M.S., Garber, M., Wolfe, S.A., and Maehr, R. (2014). Cas9 effector-mediated regulation of transcription and differentiation in human pluripotent stem cells. Development *141*, 219–223.

Kim, C., Gao, R., Sei, E., Brandt, R., Hartman, J., Hatschek, T., Crosetto, N., Foukakis, T., and Navin, N.E. (2018a). Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. Cell *173*, 879–887.e13.

Kim, K.H., and Roberts, C.W.M. (2016). Targeting EZH2 in cancer. Nat Med *22*, 128–134.

Kim, S.-H., Juhnn, Y.-S., and Song, Y.-S. (2007). Akt involvement in paclitaxel chemoresistance of human ovarian cancer cells. Ann N Y Acad Sci *1095*, 82–89.

Kim, Y.J., Xie, P., Cao, L., Zhang, M.Q., and Kim, T.H. (2018b). Global transcriptional activity dynamics reveal functional enhancer RNAs. Genome Res. *28*, 1799–1811.

Kopp, F., and Mendell, J.T. (2018). Functional Classification and Experimental Dissection of Long Noncoding RNAs. Cell *172*, 393–407.

Koren, S., and Bentires-Alj, M. (2015). Breast Tumor Heterogeneity: Source of Fitness, Hurdle for Therapy. Molecular Cell *60*, 537–546.

Kretz, M., Siprashvili, Z., Chu, C., Webster, D.E., Zehnder, A., Qu, K., Lee, C.S., Flockhart, R.J., Groff, A.F., Chow, J., et al. (2013). Control of somatic tissue differentiation by the long non-coding RNA TINCR. Nature *493*, 231–235.

Kröger, C., Afeyan, A., Mraz, J., Eaton, E.N., Reinhardt, F., Khodor, Y.L., Thiru, P., Bierie, B., Ye, X., Burge, C.B., et al. (2019). Acquisition of a hybrid E/M state is essential for tumorigenicity of basal breast cancer cells. Proc Natl Acad Sci USA *116*, 7353.

Larson, M.H., Gilbert, L.A., Wang, X., Lim, W.A., Weissman, J.S., and Qi, L.S. (2013). CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nature Protocols *8*, 2180–2196.

Latos, P.A., Pauler, F.M., Koerner, M.V., Şenergin, H.B., Hudson, Q.J., Stocsits, R.R., Allhoff, W., Stricker, S.H., Klement, R.M., Warczok, K.E., et al. (2012). Airn transcriptional overlap, but not its lncRNA products, induces imprinted Igf2r silencing. Science *338*, 1469–1472.

Lee, S., Kopp, F., Chang, T.-C., Sataluri, A., Chen, B., Sivakumar, S., Yu, H., Xie, Y., and Mendell, J.T. (2016). Noncoding RNA NORAD Regulates Genomic Stability by Sequestering PUMILIO Proteins. Cell *164*, 69–80.

Li, S., Li, Y., Chen, B., Zhao, J., Yu, S., Tang, Y., Zheng, Q., Li, Y., Wang, P., He, X., et al. (2018). exoRBase: a database of circRNA, lncRNA and mRNA in human blood exosomes. Nucleic Acids Research *46*, D106–D112.

Li, S., Song, Y., Quach, C., Guo, H., Jang, G.-B., Maazi, H., Zhao, S., Sands, N.A., Liu, Q., In, G.K., et al. (2019). Transcriptional regulation of autophagy-lysosomal function in BRAF-driven melanoma progression and chemoresistance. Nature Communications *10*, 1693.

Li, W., Xu, H., Xiao, T., Cong, L., Love, M.I., Zhang, F., Irizarry, R.A., Liu, J.S., Brown, M., and Liu, X.S. (2014). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. Genome Biol. *15*, 554.

Liang, X.-H., Sun, H., Nichols, J.G., and Crooke, S.T. (2017). RNase H1-Dependent Antisense Oligonucleotides Are Robustly Active in Directing RNA Cleavage in Both the Cytoplasm and the Nucleus. Molecular Therapy *25*, 2075–2092.

Liedtke, C., Mazouni, C., Hess, K.R., André, F., Tordai, A., Mejia, J.A., Symmans, W.F., Gonzalez-Angulo, A.M., Hennessy, B., Green, M., et al. (2008). Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. Jco *26*, 1275–1281.

Liu, S.J., Horlbeck, M.A., Cho, S.W., Birk, H.S., Malatesta, M., He, D., Attenello, F.J., Villalta, J.E., Cho, M.Y., Chen, Y., et al. (2017a). CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. Science *355*, 1–16.

Liu, S.J., Horlbeck, M.A., Cho, S.W., Birk, H.S., Malatesta, M., He, D., Attenello, F.J., Villalta, J.E., Cho, M.Y., Chen, Y., et al. (2017b). CRISPRi-based genome-scale identification of functional long noncoding RNA loci in human cells. Science *355*, eaah7111–eaah7116.

Liu, Y., Yu, C., Daley, T.P., Wang, F., Cao, W.S., Bhate, S., Lin, X., Still, C.2., Liu, H., Zhao, D., et al. (2018a). CRISPR Activation Screens Systematically Identify Factors that Drive Neuronal Fate and Reprogramming. Cell Stem Cell *23*, 758–771.e758.

Liu, Y., Cao, Z., Wang, Y., Guo, Y., Xu, P., Yuan, P., Liu, Z., He, Y., and Wei, W. (2018b). Genome-wide screening for functional long noncoding RNAs in human cells by Cas9 targeting of splice sites. Nature Biotechnology.

Lloyd-Lewis, B., Harris, O.B., Watson, C.J., and Davis, F.M. (2017). Mammary Stem Cells: Premise, Properties, and Perspectives. Trends Cell Biol *27*, 556–567.

Loewer, S., Cabili, M.N., Guttman, M., Loh, Y.-H., Thomas, K., Park, I.H., Garber, M., Curran, M., Onder, T., Agarwal, S., et al. (2010). Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. Nature Genetics *42*, 1113–1117.

Loibl P.S.L., MD, P.P.P., MD, P.M.M., MD, P.C.D., and MD, P.G.C. (2021). Breast cancer. The Lancet *397*, 1750–1769.

Lopes, R., Sprouffske, K., Sheng, C., Uijttewaal, E.C.H., Wesdorp, A.E., Dahinden, J., Wengert, S., Diaz-Miyar, J., Yildiz, U., Bleu, M., et al. (2021). Systematic dissection of transcriptional regulatory networks by genome-scale and single-cell CRISPR screens. Sci Adv *7*, 1–16.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Lu, P., Weaver, V.M., and Werb, Z. (2012). The extracellular matrix: a dynamic niche in cancer progression. J Cell Biol *196*, 395–406.

Lu, Z., Li, Y., Che, Y., Huang, J., Sun, S., Mao, S., Lei, Y., Li, N., Sun, N., and He, J. (2018). The TGFβ-induced lncRNA TBILA promotes non-small cell lung cancer progression in vitro and in vivo via cis-regulating HGAL and activating S100A7/JAB1 signaling. Cancer Letters *432*, 156–168.

Magnani, L., Stoeck, A., Zhang, X., Lánczky, A., Mirabella, A.C., Wang, T.-L., Gyorffy, B., and Lupien, M. (2013). Genome-wide reprogramming of the chromatin landscape underlies endocrine therapy resistance in breast cancer. Proc Natl Acad Sci USA *110*, E1490–E1499.

Managadze, D., Rogozin, I.B., Chernikova, D., Shabalina, S.A., and Koonin, E.V. (2011). Negative correlation between expression level and evolutionary rate of long intergenic noncoding RNAs. Genome Biol Evol *3*, 1390–1404.

Mani, S.A., Guo, W., Liao, M.-J., Eaton, E.N., Ayyanan, A., Zhou, A.Y., Brooks, M., Reinhard, F., Zhang, C.C., Shipitsin, M., et al. (2008). The Epithelial-Mesenchymal Transition Generates Cells with Properties of Stem Cells. Cell *133*, 704–715.

Margueron, R., and Reinberg, D. (2010). Chromatin structure and the inheritance of epigenetic information. Nat Rev Genet *11*, 285–296.

Marine, J.-C., Dawson, S.-J., and Dawson, M.A. (2020). Non-genetic mechanisms of therapeutic resistance in cancer. Nature Reviews Cancer 1–14.

Mattioli, K., Volders, P.-J., Gerhardinger, C., Lee, J.C., Maass, P.G., Melé, M., and Rinn, J.L. (2019). High-throughput functional analysis of lncRNA core promoters elucidates rules governing tissue specificity. Genome Res. *29*, 344–355.

Meacham, C.E., and Morrison, S.J. (2013). Tumour heterogeneity and cancer cell plasticity. Nature *501*, 328–337.

Melchor, L., Molyneux, G., Mackay, A., Magnay, F.-A., Atienza, M., Kendrick, H., Nava-Rodrigues, D., López-García, M.Á., Milanezi, F., Greenow, K., et al. (2014). Identification of cellular and genetic drivers of breast cancer heterogeneity in genetically engineered mouse tumour models. J Pathol *233*, 124–137.

Mercer, T.R., Gerhardt, D.J., Dinger, M.E., Crawford, J., Trapnell, C., Jeddeloh, J.A., Mattick, J.S., and Rinn, J.L. (2011). Targeted RNA sequencing reveals the deep complexity of the human transcriptome. Nature Biotechnology *30*, 99–104.

Meyers, R.M., Bryan, J.G., McFarland, J.M., Weir, B.A., Sizemore, A.E., Xu, H., Dharia, N.V., Montgomery, P.G., Cowley, G.S., Pantel, S., et al. (2017). Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. Nature Genetics *49*, 1779–1784.

Michelini, F., Jalihal, A.P., Francia, S., Meers, C., Neeb, Z.T., Rossiello, F., Gioia, U., Aguado, J., Jones-Weinert, C., Luke, B., et al. (2018). From "Cellular" RNA to "Smart" RNA: Multiple Roles of RNA in Genome Stability and Beyond. Chem Rev *118*, 4365–4403.

Michels, B.E., Mosa, M.H., Streibl, B.I., Zhan, T., Menche, C., Abou-El-Ardat, K., Darvishi, T., Członka, E., Wagner, S., Winter, J., et al. (2020). Pooled In Vitro and In Vivo CRISPR-Cas9 Screening Identifies Tumor Suppressors in Human Colon Organoids. Cell Stem Cell *26*, 782–792.e787.

Minckwitz, von, G., Procter, M., de Azambuja, E., Zardavas, D., Benyunes, M., Viale, G., Suter, T., Arahmani, A., Rouchet, N., Clark, E., et al. (2017). Adjuvant Pertuzumab and Trastuzumab in Early HER2-Positive Breast Cancer. N Engl J Med 377, 122–131.

Molyneux, G., Geyer, F.C., Magnay, F.-A., McCarthy, A., Kendrick, H., Natrajan, R., Mackay, A., Grigoriadis, A., Tutt, A., Ashworth, A., et al. (2010). BRCA1 basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. Cell Stem Cell *7*, 403–417.

Morelli, E., Gulla, A., Amodio, N., Taiana, E., Neri, A., Fulciniti, M., and Munshi, N.C. (2021). CRISPR Interference (CRISPRi) and CRISPR Activation (CRISPRa) to Explore the Oncogenic lncRNA Network. In Long Non-Coding RNAs in Cancer, A. Navarro, ed. (New York, NY: Springer US), pp. 189–204.

Müller-Hermelink, N., Braumüller, H., Pichler, B., Wieder, T., Mailhammer, R., Schaak, K., Ghoreschi, K., Yazdi, A., Haubner, R., Sander, C.A., et al. (2008). TNFR1 signaling and IFN-gamma signaling determine whether T cells induce tumor dormancy or promote multistage carcinogenesis. Cancer Cell *13*, 507–518.

Nardi, V., Azam, M., and Daley, G.Q. (2004). Mechanisms and implications of imatinib resistance mutations in BCR-ABL. Curr Opin Hematol *11*, 35–43.

Neve, R.M., Chin, K., Fridlyand, J., Yeh, J., Baehner, F.L., Fevr, T., Clark, L., Bayani, N., Coppe, J.-P., Tong, F., et al. (2006). A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. Cancer Cell *10*, 515–527.

Ning, S., Zhang, J., Wang, P., Zhi, H., Wang, J., Liu, Y., Gao, Y., Guo, M., Yue, M., Wang, L., et al. (2016). Lnc2Cancer: a manually curated database of experimentally supported lncRNAs associated with various human cancers. Nucleic Acids Research *44*, D980–D985.

Ninomiya, K., Adachi, S., Natsume, T., Iwakiri, J., Terai, G., Asai, K., and Hirose, T. (2020). LncRNA-dependent nuclear stress bodies promote intron retention through SR protein phosphorylation. Embo J *39*, e102729.

Ntini, E., Louloupi, A., Liz, J., Muino, J.M., Marsico, A., and Ørom, U.A.V. (2018). Long ncRNA A-ROD activates its target gene DKK1 at its release from chromatin. Nature Communications *9*, 1636.

Nuñez, J.K., Chen, J., Pommier, G.C., Cogan, J.Z., Replogle, J.M., Adriaens, C., Ramadoss, G.N., Shi, Q., Hung, K.L., Samelson, A.J., et al. (2021). Genome-wide programmable transcriptional memory by CRISPR-based epigenome editing. Cell 1–35.

Oeffinger, K.C., Fontham, E.T.H., Etzioni, R., Herzig, A., Michaelson, J.S., Shih, Y.-C.T., Walter, L.C., Church, T.R., Flowers, C.R., LaMonte, S.J., et al. (2015). Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update From the American Cancer Society. Jama *314*, 1599–1614.

Oren, Y., Tsabar, M., Cuoco, M.S., Amir-Zilberstein, L., Cabanos, H.F., Hütter, J.-C., Hu, B., Thakore, P.I., Tabaka, M., Fulco, C.P., et al. (2021). Cycling cancer persister cells arise from lineages with distinct programs. Nature 1–23.

Oskarsson, T. (2013). Extracellular matrix components in breast cancer progression and metastasis. Breast *22 Suppl 2*, S66–S72.

Paralkar, V.R., Taborda, C.C., Huang, P., Yao, Y., Kossenkov, A.V., Prasad, R., Luan, J., Davies, J.O.J., Hughes, J.R., Hardison, R.C., et al. (2016). Unlinking an lncRNA from Its Associated cis Element. Molecular Cell *62*, 104–110.

Parker, J.S., Mullins, M., Cheang, M.C.U., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., et al. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. J Clin Oncol *27*, 1160–1167.

Peculis, B.A. (2000). RNA-binding proteins: if it looks like a sn(o)RNA.. Curr Biol *10*, R916–R918.

Perry, R.B.-T., and Ulitsky, I. (2016). The functions of long noncoding RNAs in development and stem cells. Development *143*, 3882–3894.

Petit, C., Gouel, F., Dubus, I., Heuclin, C., Roget, K., and Vannier, J.P. (2016). Hypoxia promotes chemoresistance in acute lymphoblastic leukemia cell lines by modulating death signaling pathways. BMC Cancer *16*, 746.

Pogrebniak, K.L., and Curtis, C. (2018). Harnessing Tumor Evolution to Circumvent Resistance. Trends in Genetics *34*, 639–651.

Prat, A., Parker, J.S., Fan, C., and Perou, C.M. (2012). PAM50 assay and the three-gene model for identifying the major and clinically relevant molecular subtypes of breast cancer. Breast Cancer Res Treat *135*, 301–306.

Prat, A., and Perou, C.M. (2011). Deconstructing the molecular portraits of breast cancer. Mol Oncol *5*, 5–23.

Prunier, C., Baker, D., Dijke, ten, P., and Ritsma, L. (2018). TGF-β Family Signaling Pathways in Cellular Dormancy. TRENDS in CANCER 1–13.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-Guided Platform for Sequence-Specific Control of Gene Expression. Cell *152*, 1173–1183.

Raha, D., Wilson, T.R., Peng, J., Peterson, D., Yue, P., Evangelista, M., Wilson, C., Merchant, M., and Settleman, J. (2014). The cancer stem cell marker aldehyde dehydrogenase is required to maintain a drug-tolerant tumor cell subpopulation. Cancer Res *74*, 3579–3590.

Rajagopal, J., and Stanger, B.Z. (2016). Plasticity in the Adult: How Should the Waddington Diagram Be Applied to Regenerating Tissues? Developmental Cell *36*, 133–137.

Rambow, F., Rogiers, A., Marin-Bejar, O., Aibar, S., Femel, J., Dewaele, M., Karras, P., Brown, D., Chang, Y.H., Debiec-Rychter, M., et al. (2018). Toward Minimal Residual Disease-Directed Therapy in Melanoma. Cell *174*, 843–855.e19.

Ramilowski, J.A., Yip, C.W., Agrawal, S., Chang, J.-C., Ciani, Y., Kulakovskiy, I.V., Mendez, M., Ooi, J.L.C., Ouyang, J.F., Parkinson, N., et al. (2020). Functional annotation of human long noncoding RNAs via molecular phenotyping. Genome Res. *30*, 1060–1072.

Ramos, P., and Bentires-Alj, M. (2019). Mechanism-based cancer therapy: resistance to therapy, therapy for resistance. 1–10.

Ranganathan, A.C., Zhang, L., Adam, A.P., and Aguirre-Ghiso, J.A. (2006). Functional coupling of p38-induced up-regulation of BiP and activation of RNA-dependent protein kinase-like endoplasmic reticulum kinase to drug resistance of dormant carcinoma cells. Cancer Res *66*, 1702–1711.

Ransohoff, J.D., Wei, Y., and Khavari, P.A. (2017). The functions and unique features of long intergenic non-coding RNA. Nature Publishing Group *19*, 143–157.

Recasens, A., and Munoz, L. (2019). Targeting Cancer Cell Dormancy. Trends in Pharmacological Sciences *40*, 128–141.

Regan, M.M., Francis, P.A., Pagani, O., Fleming, G.F., Walley, B.A., Viale, G., Colleoni, M., Láng, I., Gomez, H.L., Tondini, C., et al. (2016). Absolute Benefit of Adjuvant Endocrine Therapies for Premenopausal Women With Hormone Receptor-Positive, Human Epidermal Growth Factor Receptor 2-Negative Early Breast Cancer: TEXT and SOFT Trials. Jco *34*, 2221–2231.

Rehman, S.K., Haynes, J., Collignon, E., Brown, K.R., Wang, Y., Nixon, A.M.L., Bruce, J.P., Wintersinger, J.A., Mer, A.S., Lo, E.B.L., et al. (2021). Colorectal Cancer Cells Enter a Diapause-like DTP State to Survive Chemotherapy. Cell *184*, 226–242.e21.

Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. Annu Rev Biochem *81*, 145–166.

Rinn, J.L., and Chang, H.Y. (2020). Long Noncoding RNAs: Molecular Modalities to Organismal Functions. Annu Rev Biochem *89*, 283–308.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., et al. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell *129*, 1311–1323.

Risom, T., Langer, E.M., Chapman, M.P., Rantala, J., Fields, A.J., Boniface, C., Alvarez, M.J., Kendsersky, N.D., Pelz, C.R., Johnson-Camacho, K., et al. (2018). Differentiation-state plasticity is a targetable resistance mechanism in basal-like breast cancer. Nature Communications 1–17.

Ritter, N., Ali, T., Kopitchinski, N., Schuster, P., Beisaw, A., Hendrix, D.A., Schulz, M.H., Müller-McNicoll, M., Dimmeler, S., and Grote, P. (2019). The lncRNA Locus Handsdown Regulates Cardiac Gene Programs and Is Essential for Early Mouse Development. Developmental Cell *50*, 644–657.e648.

Rohwer, N., and Cramer, T. (2011). Hypoxia-mediated drug resistance: novel insights on the functional interaction of HIFs and cell death pathways. Drug Resist Updat *14*, 191–201.

Roux, B.T., Lindsay, M.A., and Heward, J.A. (2017). Knockdown of Nuclear-Located Enhancer RNAs and Long ncRNAs Using Locked Nucleic Acid GapmeRs. Methods Mol Biol *1468*, 11–18.

Samanta, D., Gilkes, D.M., Chaturvedi, P., Xiang, L., and Semenza, G.L. (2014). Hypoxia-inducible factors are required for chemotherapy resistance of breast cancer stem cells. Proc Natl Acad Sci USA *111*, E5429–E5438.

Sander, J.D., and Joung, J.K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. Nature Biotechnology *32*, 347–355.

Saunus, J.M., Smart, C.E., Kutasovic, J.R., Johnston, R.L., Kalita-de Croft, P., Miranda, M., Rozali, E.N., Vargas, A.C., Reid, L.E., Lorsy, E., et al. (2018). Multidimensional phenotyping of breast cancer cell lines to guide preclinical research. Breast Cancer Res Treat *167*, 289–301.

Schertzer, M.D., Braceros, K.C.A., Starmer, J., Cherney, R.E., Lee, D.M., Salazar, G., Justice, M., Bischoff, S.R., Cowley, D.O., Ariel, P., et al. (2019). lncRNA-Induced Spread of Polycomb Controlled by Genome Architecture, RNA Abundance, and CpG Island DNA. Molecular Cell *75*, 523–537.e10.

Schlackow, M., Nojima, T., Gomes, T., Dhir, A., Carmo-Fonseca, M., and Proudfoot, N.J. (2017). Distinctive Patterns of Transcription and RNA Processing for Human lincRNAs. Molecular Cell *65*, 25–38.

Schmitt, A.M., and Chang, H.Y. (2016). Long Noncoding RNAs in Cancer Pathways. Cancer Cell *29*, 452–463.

Shackleton, M., Vaillant, F., Simpson, K.J., Stingl, J., Smyth, G.K., Asselin-Labat, M.-L., Wu, L., Lindeman, G.J., and Visvader, J.E. (2006). Generation of a functional mammary gland from a single stem cell. Nature *439*, 84–88.

Shaffer, S.M., Dunagin, M.C., Torborg, S.R., Torre, E.A., Emert, B., Krepler, C., Beqiri, M., Sproesser, K., Brafford, P.A., Xiao, M., et al. (2018). Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance. Nature Publishing Group 1–22.

Sharma, S.V., Lee, D.Y., Li, B., Quinlan, M.P., Takahashi, F., Maheswaran, S., McDermott, U., Azizian, N., Zou, L., Fischbach, M.A., et al. (2010). A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. Cell *141*, 69–80.

Shen, S., Vagner, S., and Robert, C. (2020). Persistent Cancer Cells: The Deadly Survivors. Cell *183*, 860–874.

Shibue, T., and Weinberg, R.A. (2017). EMT, CSCs, and drug resistance: the mechanistic link and clinical implications. Nature Publishing Group *14*, 611–629.

Shyh-Chang, N., Locasale, J.W., Lyssiotis, C.A., Zheng, Y., Teo, R.Y., Ratanasirintrawoot, S., Zhang, J., Onder, T., Unternaehrer, J.J., Zhu, H., et al. (2013). Influence of threonine metabolism on S-adenosylmethionine and histone methylation. Science *339*, 222–226.

Simeonov, K.P., Byrns, C.N., Clark, M.L., Norgard, R.J., Martin, B., Ben Z Stanger, Shendure, J., McKenna, A., and Lengner, C.J. (2021). Single-cell lineage tracing of metastatic cancer reveals selection of hybrid EMT states. Cancer Cell 1–32.

Sosa, M.S., Bragado, P., and Aguirre-Ghiso, J.A. (2014). Mechanisms of disseminated cancer cell dormancy: an awakening field. Nature Reviews Cancer *14*, 611–622.

Spector, D.L., and Lamond, A.I. (2011). Nuclear speckles. Cold Spring Harb Perspect Biol *3*.

Statello, L., Guo, C.-J., Chen, L.L., and Huarte, M. (2021). Gene regulation by long noncoding RNAs and its biological functions. Nature Publishing Group 1–23.

Stingl, J., Eirew, P., Ricketson, I., Shackleton, M., Vaillant, F., Choi, D., Li, H.I., and Eaves, C.J. (2006). Purification and unique properties of mammary epithelial stem cells. Nature *439*, 993–997.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA *102*, 15545–15550.

Szlachta, K., Kuscu, C., Tufan, T., Adair, S.J., Shang, S., Michaels, A.D., Mullen, M.G., Fischer, N.L., Yang, J., Liu, L., et al. (2018). CRISPR knockout screening identifies combinatorial drug targets in pancreatic cancer and models cellular drug response. Nature Communications *9*, 7–13.

Tan, J.Y., Biasini, A., Young, R.S., and Marques, A.C. (2020). Splicing of enhancer-associated lincRNAs contributes to enhancer activity. Life Sci Alliance *3*.

Thienpont, B., Steinbacher, J., Zhao, H., D'Anna, F., Kuchnio, A., Ploumakis, A., Ghesquière, B., Van Dyck, L., Boeckx, B., Schoonjans, L., et al. (2016). Tumour hypoxia causes DNA hypermethylation by reducing TET activity. Nature *537*, 63–68.

Thomas, S., Thurn, K.T., Biçaku, E., Marchion, D.C., and Münster, P.N. (2011). Addition of a histone deacetylase inhibitor redirects tamoxifen-treated breast cancer cells into apoptosis, which is opposed by the induction of autophagy. Breast Cancer Res Treat *130*, 437–447.

Tichon, A., Gil, N., Lubelsky, Y., Havkin Solomon, T., Lemze, D., Itzkovitz, S., Stern-Ginossar, N., and Ulitsky, I. (2016). A conserved abundant cytoplasmic long noncoding RNA modulates repression by Pumilio proteins in human cells. Nature Communications *7*, 12209.

Tordonato, C., Marzi, M.J., Giangreco, G., Freddi, S., Bonetti, P., Tosoni, D., Di Fiore, P.P., and Nicassio, F. (2021). miR-146 connects stem cell identity with metabolism and pharmacological resistance in breast cancer. J Cell Biol *220*, 1–21.

Torre, E.A., Arai, E., Bayatpour, S., Jiang, C.L., Beck, L.E., Emert, B.L., Shaffer, S.M., Mellis, I.A., Fane, M.E., Alicea, G.M., et al. (2021). Genetic screening for single-cell variability modulators driving therapy resistance. Nature Genetics 1–30.

Tseng, Y.-Y., Moriarity, B.S., Gong, W., Akiyama, R., Tiwari, A., Kawakami, H., Ronning, P., Reuland, B., Guenther, K., Beadnell, T.C., et al. (2014). PVT1 dependence in cancer with MYC copy-number increase. Nature *512*, 82–86.

Van Keymeulen, A., Lee, M.Y., Ousset, M., Brohée, S., Rorive, S., Giraddi, R.R., Wuidart, A., Bouvencourt, G., Dubois, C., Salmon, I., et al. (2015). Reactivation of multipotency by oncogenic PIK3CA induces breast tumour heterogeneity. Nature *525*, 119–123.

Viale, A., Pettazzoni, P., Lyssiotis, C.A., Ying, H., Sánchez, N., Marchesini, M., Carugo, A., Green, T., Seth, S., Giuliani, V., et al. (2014). Oncogene ablation-resistant pancreatic cancer cells depend on mitochondrial function. Nature *514*, 628–632.

Visvader, J.E., and Stingl, J. (2014). Mammary stem cells and the differentiation hierarchy: current status and perspectives. Genes Dev *28*, 1143–1158.

Wang, K.C., Yang, Y.W., Liu, B., Sanyal, A., Corces-Zimmerman, R., Chen, Y., Lajoie, B.R., Protacio, A., Flynn, R.A., Gupta, R.A., et al. (2011). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature *472*, 120–124.

Wang, M., and Kaufman, R.J. (2016). Protein misfolding in the endoplasmic reticulum as a conduit to human disease. Nature *529*, 326–335.

Werner, M.S., and Ruthenburg, A.J. (2015). Nuclear Fractionation Reveals Thousands of Chromatin-Tethered Noncoding RNAs Adjacent to Active Genes. CellReports *12*, 1089–1098.

West, J.A., Davis, C.P., Sunwoo, H., Simon, M.D., Sadreyev, R.I., Wang, P.I., Tolstorukov, M.Y., and Kingston, R.E. (2014). The long noncoding RNAs NEAT1 and MALAT1 bind active chromatin sites. Molecular Cell *55*, 791–802.

West, J.A., Mito, M., Kurosaka, S., Takumi, T., Tanegashima, C., Chujo, T., Yanaka, K., Kingston, R.E., Hirose, T., Bond, C., et al. (2016). Structural, super-resolution microscopy analysis of paraspeckle nuclear body organization. J Cell Biol *214*, 817–830.

Wu, H., Yang, L., and Chen, L.L. (2017). The Diversity of Long Noncoding RNAs and Their Generation. Trends in Genetics *33*, 540–552.

Wu, S., Liu, S., Liu, Z., Huang, J., Pu, X., Li, J., Yang, D., Deng, H., Yang, N., and Xu, J. (2015). Classification of circulating tumor cells by epithelial-mesenchymal transition markers. PLoS ONE *10*, e0123976.

Yu, H., Lee, H., Herrmann, A., Buettner, R., and Jove, R. (2014). Revisiting STAT3 signalling in cancer: new and unexpected biological functions. Nature Publishing Group *14*, 736–746.

Zhen, Q., Gao, L.-N., Wang, R.-F., Chu, W.-W., Zhang, Y.-X., Zhao, X.-J., Lv, B.-L., and Liu, J.-B. (2018). LncRNA DANCR Promotes Lung Cancer by Sequestering miR-216a. Cancer Control *25*, 1073274818769849.

Zhu, S., Li, W., Liu, J., Chen, C.-H., Liao, Q., Xu, P., Xu, H., Xiao, T., Cao, Z., Peng, J., et al. (2016). Genome-scale deletion screening of human long non-coding RNAs using a paired-guide RNA CRISPR-Cas9 library. Nature Biotechnology *34*, 1279–1286.

Zucchelli, S., Cotella, D., Takahashi, H., Carrieri, C., Cimatti, L., Fasolo, F., Jones, M.H., Sblattero, D., Sanges, R., Santoro, C., et al. (2015). SINEUPs: A new class of natural and synthetic antisense long non-coding RNAs that activate translation. RNA Biol *12*, 771–779.