PhD degree in Molecular Medicine (curriculum in Computational Biology)

European School of Molecular Medicine (SEMM),

University of Milan and University of Naples "Federico II"

Settore disciplinare: BIO/11

Changes of Replication Timing induced by PML-RARα

Francesco Santaniello

IEO,Milan

Matricola: R10341

Supervisor: Pier Giuseppe Pelicci
IEO,Milan

Added supervisor: Lucilla Luzi
IEO,Milan

Gaetano Ivan Dellino
IEO,Milan

Academic Year 2015/2016

INDEX

Tables

FIGURES

Table of abbreviations

Abbreviation	Meaning	Abbreviation	Meaning
DNA	Deoxyribonucleic acid	SNS	Short nascent strands
ORIs	replication origins	TSS	Transcription Start Site
ORC	Origin Recognition Complex	ncRNA	non-coding RNAs
pre-RC	pre-Replication Complex	BrdU	5-bromo-2'-deoxyuridine
pre-IC	pre-Initiation Complex	APL	Acute promyelocytic leukemia
RNA	Ribonucleic acid	FAB	French-American-British
ODP	Origin Decision Point	ATRA	All-trans retinoic acid
MCM	Minichromosome Maintenance	RA	Retinoic acid
PCNA	(Proliferating Cell Nuclear Antigen	DAPI	4',6-diamidino-2-phenylindole
DHFR	dihydrofolate reductase	BWA	Barrows-Wheeler Aligner
LCR	Locus Control Region	FACS	Fluorescence-activated cell sorting
UTR	Untranslated Region	TMM	Trimmed Mean of M-values
HDAC	histone Deacetylase	OCRs	Origin Containing Regions
HBO1	Human Acetylase Binding to ORC1	ROC	Receiver Operating Characteristic
dUMP	Deoxyuracilmonophosphates	AUC	Area Under the Curve
dTMP	Deoxythyminemonophosphates	DRT	Differential Replication Threshold
mESCs	murine Embryonic Stem Cells	SRT	Stable Replication Threshold
hESCs	human Embryonic Stem Cells	TPM	Transcripts per Million
piPSCs	partially Induced Pluripotent Stem Cells	DRVs	Differentially Replicating Valleys
mEpiSCs	mouse Epiblast-derived Stem Cells	UGs	Up-regulated genes
CDKs	Cyclin-dependent kinases	DGs	Down-regulated genes
HP1	Heterochromatin Protein 1		

ABSTRACT

DNA replication is the cellular process that ensures the loyal and faithful inheritance of the genetic instructions contained in the double-strand DNA molecule from one parental cell to each daughter cell. It starts from precise genomic loci known as DNA Replication Origins (ORIs). Several tens of thousands ORIs are “licenced” during the M/G1 phase of cell cycle to become effective sites of replication start, but only a portion of them are really “activated” in S phase. Due to the complexity and the crucial importance of the DNA replication, these two regulatory processes must be tightly regulated in both space and time.

Different works have recently shed some light onto the positional identification of ORIs and their activation, also thanks to the big contribution given recently by the application of the Repli-seq technology, a novel method by which it is possible to follow the progression of DNA replication both in space and time.

Up to now, the time-related features of DNA replication, together with the factors that might impact the temporal dimension of this system, are yet poorly studied and described.

To also try to fill the gap for the lack of standard methods able to recognize differences in Replication Timing, we developed a bioinformatic method (DART; Differential Analysis of Replication Timing) to accomplish this task.

Indeed the application of this procedure to our Repli-seq data was instrumental to answer our main question:

Might an oncogene (e.g. the initiating leukemogenic factor PML-RARα) fulfil, at least in part, its tumorigenic potential eliciting an alteration of the normal replication timing pace in cells?

To identify genomic regions carrying a potential difference in Replication Timing caused by the oncogene endeavour and characterize them, we performed Repli-seq, RNA-seq and ChIP-seq experiments using an inducible cellular system, the promonocytic U937PR9 cells which, if treated with Zinc, ectopically express the PML-RARα fusion protein and, consequently, we integrated these data.

PML-RARα is the initiator oncogene of the Acute Promyelocytic Leukaemia and it works as an aberrant transcriptional factor also recruiting chromatin modifier enzymes to promoters of its target genes.

As a result, we found that, after its expression, PML-RARα indeed exerts a deregulative effect on Replication Timing, inducing some regions to replicate earlier, and some other later, than in control cells. Characterizing them, we observed a close and specific association between these differentially replicated regions and both pre-existing, and PML-RARα-related, transcriptional status and chromatin structure.

We learnt that regions presenting a EtoL-shifted replication coincide with 'active' chromatin foci enriched for direct Down-regulated targets of PML-RARα; at the opposite, after expression of the oncogene, regions with a LtoE-shifted Replication Timing show moderate epigenetic 'active' features and are enriched for indirect Up-regulated transcriptional targets of PML-RARα.

It is difficult, at this stage, to envisage more clearly the mechanistic route that might explain priorities or temporal precedencies of the processes (transcription regulation, chromatin status and Replication Timing) that PML-RARα seems to harm.

Surely it will be interesting and probably easier, now that a new dowel has been added, to better understand the entire picture represented by the complex interconnected actions implemented by the oncogene over its malicious assault plan and then figure out how to combat them.

INTRODUCTION

Background

The DNA Replication process

DNA replication is a biological process that allows the faithful and exact inheritance of the complete genome from a mother cell to two daughter cells.

The starting point of such process is the opening of the double helix at precise genomic regions, called replication origins (ORIs) which, in turn, are subsequently recognized by ORC (Origin Recognition Complex), a protein complex composed by six subunits, ORC1~ORC6. At the beginning of the G1 phase, ORC recognizes ORIs and bind to DNA via its ORC1 subunit.

In middle G1 phase, another protein complex, the pre-Replication Complex (pre-RC) is loaded the selected ORIs, which are therefore "licensed" (that is, are made "eligible" for further activation). In S phase, other factors are bound to the pre-RC, forming the pre-Initiation Complex (pre-IC).

Formation of the pre-IC will then lead to the activation (known as "origin firing") of the licensed origins, allowing DNA synthesis machinery to copy each DNA strand, with the DNA synthesis progressing from ORIs in a bidirectional manner (Bell and Dutta, 2002; Donley and Thayer, 2013) (Figure 1).

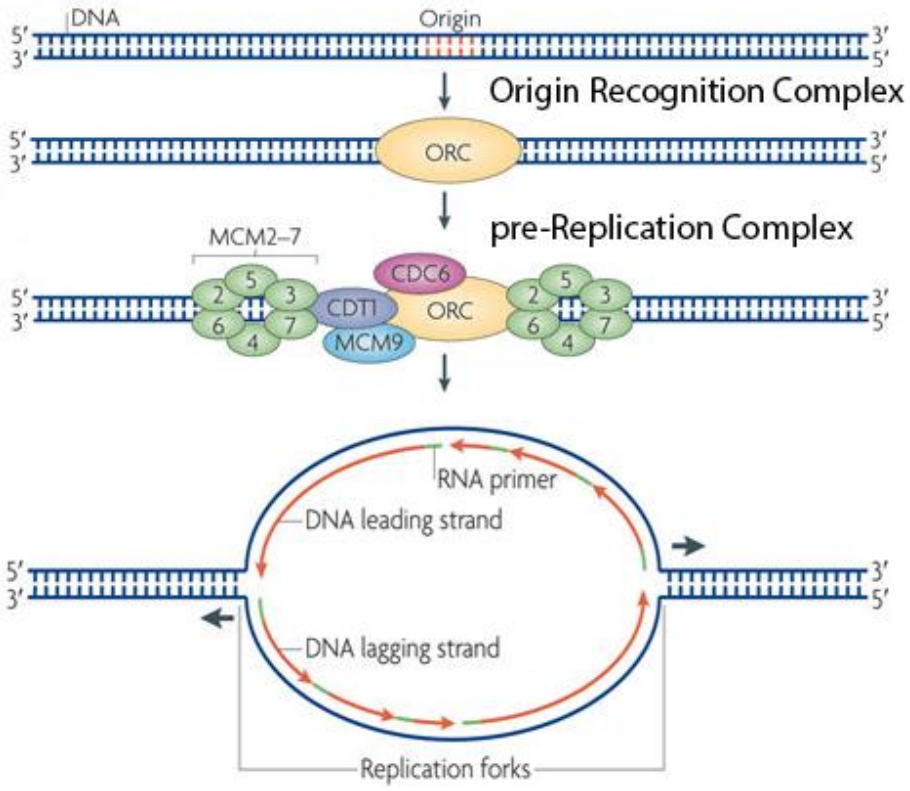

Figure 1. Main steps of Eukaryotic DNA Replication - edited from (Méchali, 2010)
DNA Replication takes place in well-defined genomic regions, called DNA Replication Origins (ORIs). As a first step, ORIs are recognized and bound by a protein complex named Origin Recognition Complex (ORC). Subsequently, another group of proteins, the pre-Replication Complex (pre-RC) bound to ORC. Only when the pre-RC and ORC are released, the DNA replication begins and replication forks initiate to move in a bi-directional manner replicating leading and lagging DNA on opposite strands.

Since the process of DNA replication involves the separation and duplication of two DNA strands, it constitutionally puts the genomic DNA into a precarious situation. In addition, whereas in bacteria it is merely coupled with cell growth and proliferation, in metazoan DNA replication is linked also to cell differentiation. Thus, to make things more complex, along with an exact copy of the genome, there is also the necessity to faithfully reproduce all of the chromatin features that will be the principal determinants of the cell identity (Fragkos et al., 2015).

Indeed, the progression of the replicative machinery might be hindered by both internal and external factors (Figure2).

As an example, the advancing replicative forks might encounter unusual secondary structures, such as dinucleotide or trinucleotide repeats, hairpins, quadruplexes or specific

G-rich nucleotide repeats or DNA-RNA hybrids (R-loops).

On its path, toward the replicating DNA strand, the replicative machinery might also clash with other DNA-binding proteins or chromatin proteins, or its advance being hampered by a reduction of the nucleotide pool or a depletion in other factors needed for the progression of the fork (Hyrien, 2000) .

Among the endogenous factors that can impair DNA replication, a major threat is posed by the collision between the replication and transcription machineries.

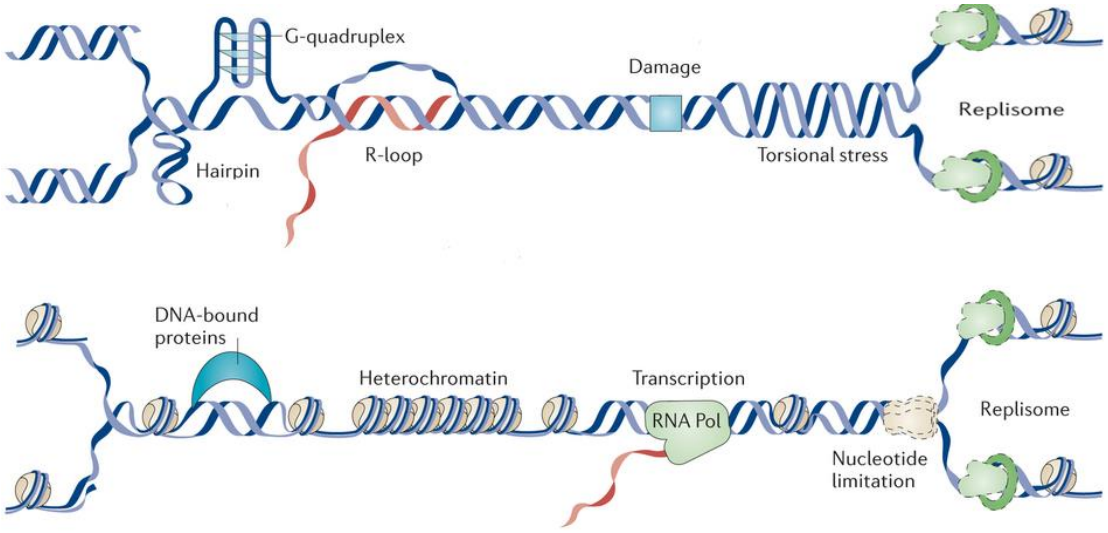

Figure 2. Factors impairing progression of Replication Fork – edited from (Gaillard et al., 2015b).
DNA replication can be hampered by several endogenous factors, such as DNA secondary structures like G-quadruplex and Hairpins, torsional stress induced by nearby transcription or replicative events, or DNA-RNA hybrids. Moreover, progression of the replication forks throughout the genome might be slowed or stalled by other circumstances, such as DNA-bound proteins, transcriptional events (which may also lead to collision between DNA polymerase and RNA polymerase), or heterochromatin.

It is known that ORIs are commonly found within genomic territories usually enriched for transcriptional activity, such as promoters, DNase-hypersensitivity sites and open chromatin regions and that replication timing and transcriptional activity are influenced by the same alteration in chromatin status, such as histones acetylation or methylation.

The event of a collision between DNA polymerase and RNA polymerase might have particularly dire consequences on the fate of a cell. Such accidents, indeed, might give rise to cell death or, worse, to aberrant chromosomal events (*e.g.* recombinations), DNA

damage, and in general DNA instability, rising the probability of malignancies. (Helmrich et al., 2013).

Taking into consideration the external insults that might interfere with a correct replication of DNA, an important contribution is given by oncogenes. As an example, it is known that oncogenes might activate a cascade of signals resulting in the deregulation of replication origin activation, leading to cellular hyper-proliferation (Hills and Diffley, 2014).

A hyper-activation of ORIs might induce a stalling of the replicative fork or to a fork-to-fork collision leading, in turn, to DNA damage, senescence and cellular death. Such events, in general, are however successfully managed by the DNA damage repair system. When these systems fail to prevent the damage, though, cells might undergo DNA instability and oncogenic transformations (Di Micco et al., 2006; Jones et al., 2013) (Figure 3).

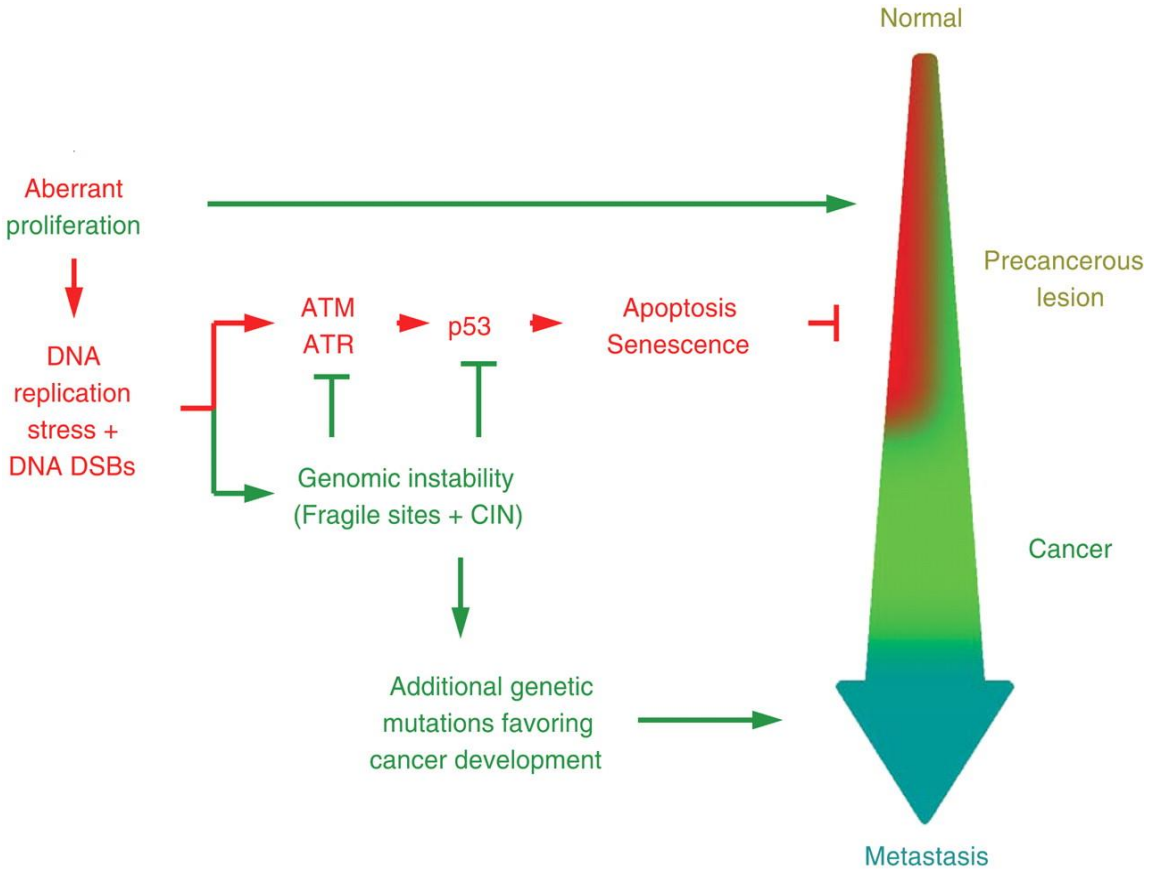

Figure 3. From aberrant DNA replication to cancer – edited from (Halazonetis et al., 2008) DNA replication is a fine-tuned process, which starts at the beginning of the cell cycle and terminates at the beginning of the G2/M phase, when cell is ready to divide into two daughter cells. During this process, DNA double helix is open, stretched, twisted and scanned by different protein complexes. In the meanwhile, other processes, such as transcription, take place in regions nearby the one being replicated. For this reason, if a tight regulation fails to be achieved, a series of events (*e.g.* DNA damage, nucleotide insertions, duplications)

might lead to an aberrant replicative process which, in turn, might lead to genomic instability and, possibly, an oncogenic transformation of the cell.

Therefore, cells have implemented a series of strategies to regulate the DNA replicative process, in order to dramatically reduce the possibility of undesired accidents.
The main approach adopted by the cells to overcome such events and ensure transmission of an exact copy of the genome from one cell to another mainly resides in a tight regulation of DNA replication.

These control mechanisms translate mostly in two different interdependent aspects of the DNA replication that are the "space" and the "time".

In particular,

the Spatial regulation refers to the:

- Recognition of replication origins by the Origin Recognition Complex (ORC) during the transition from M to G1 phase.
- Assembly of a pre-Replication Complex onto the selected origins in mid-G1 phase.

the Temporal Regulation applies to the:

- sequential and temporally ordered activation of licensed origins throughout the S-phase. By this regulative program, every genomic region will be replicated according to the temporary pattern attributed to the licensed ORIs contained in it and, therefore, by the specific *replication timing* of the region itself.

Control mechanisms of DNA Replication

Spatial Regulation

Recognition of Origins

While in bacteria, the initiation of DNA replication is achieved by the binding of a single initiation protein, DnaA, to the *oriC,* that is the only origin of replication present in these organisms, studies in *Saccharomyces cerevisiae* led to the discovery that the DnaA equivalent in eukaryotes is a protein complex: the Origin Recognition Complex (ORC), a protein complex consisting of 6 different subunits: ORC1~ORC6.

Differently from bacteria, where binding of DnaA to *oriC* is sufficient to start replication, the recognition of replication origins in yeast depends by both the binding of ORC to a 11 bp AT-rich consensus sequence (ACS, ARS consensus sequence), and the presence of a precise chromatin context (Bell and Stillman, 1992; Eaton et al., 2010; Leonard and Grimwade, 2010).

On the other hand, in metazoan the number of ORIs is much larger than in lower eukaryotes and ORC does not exhibit any sequence specificity (Schaarschmidt et al., 2004; Vashee et al., 2003).

ORC binding to DNA, however, is not random, and studies on elements required for ectopic replication from mammalian replicators suggested that recognition of replication origins might be strongly context-dependent, relying on regions enriched for dinucleotide repeats, matrix attachment regions (MARs), and AT-rich sequences (Altman and Fanning, 2004; Debatisse et al., 2004; Paixão et al., 2004; Wang et al., 2004).

Assembly of the pre-Replication Complex and its regulation

Soon after ORC binding to DNA during the transition from M to G1 phase, other proteins are sequentially assembled on the selected ORIs to form a protein complex known as pre-Replication Complex (pre-RC).

The completion of the pre-RC on an ORI is known as "origin licensing" and may be followed, in S phase, by the "activation" of that origin and its subsequent replication.

Furthermore, as explained further on, origins can only be licensed and activated once per each cell cycle. Licensing is achieved during the middle part of the G1 phase, a period also referred as "Origin Decision Point" (ODP) (Wu and Gilbert, 1997).

As a first step, Cdc6, complexed with ATP, is bound to ORC. Successively, Cdt1, followed by MCM2-7, join the Cdc6-ORC-DNA complex (Figure 4).

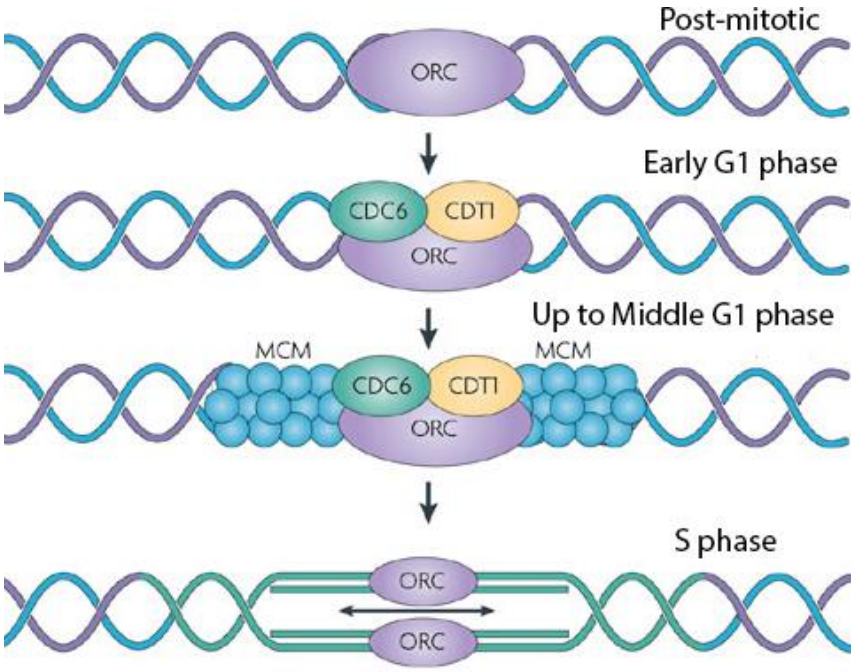

Figure 4. Formation of a pre-Replication Complex onto a DNA Replication Origin – edited from (Aladjem, 2007)
After ORI recognition by the ORC, ATP-bound Cdc6 binds to ORC, followed in rapid succession by Cdt1 and MCM2-7. Binding of MCM2-7 to the DNA-ORC-Cdc6-Cdt1 complex is known as "Origin Licensing" and it is a 'conditio sine qua non' for a genomic region to serve as a replication origin.

The ATP bound to Cdc6 is hydrolysed, and this reaction stimulates the dissociation of Cdc6 from the complex. In turn, this catalyses the disengagement of Cdt1, which consequently prompt the stable binding of the MCM (Minichromosome Maintenance) complex to DNA.

Successively, the ORC-bound ATP is hydrolysed, allowing the dissociation of ORC and the reiterative loading of the MCM complex. Loading of the MCM complex onto the pre-RC brings to origin licensing, and it is the last step of the pre-RC formation, subsequently followed by the activation of the origin of replication (Bowers et al., 2004).

Upon entering the S-phase, part of Cdt1 is bound by Geminin, preventing the loading of the MCM complex on the DNA. The remaining Cdt1 is degraded by either a CDK-mediated ubiquitin-dependent proteolysis or by interaction with PCNA (Proliferating Cell Nuclear Antigen). In this way, Cdt1 will not be present during the S phase, in order not to cause re-licensing of replication origins, which will in turn lead to re-replication of the same genomic region which would, in turn, result into genomic instability (Lee et al., 2004; Wohlschlegel et al., 2000).

Activation of the pre-Replication Complex

Activation of the pre-RC is achieved by the assembly of additional replication factors onto the licensed ORIs, forming the pre-Initiation Complex (pre-IC), necessary for the activation of the MCM2-7 helicase, unwinding of ORI, formation of the Replisome Progression Complexes and the bi-directional progression of the replication fork.

Such process is under the control of two kinases, CDK (Cyclin Dependend Kinase) and DDK (Dpf4-Dependent Kinase) which, in turn, trigger a cascade of signal resulting in the recruitment of two additional factors, Cdc45 and GINS (a heterotetrametric complex formed by the association of Psf1,2,3 and Sld5) .

Upon their recruitment, Cdc45 and GINS are tightly bound to MCM2-7, forming the

Cdc45-MCM-GINS (CMG) Complex.(Symeonidou et al., 2012; Tanaka and Araki, 2013; Walter and Araki, 2006)

The loading of Cdc45 on the pre-IC is critical for at least two reasons:

- Upon formation of the Cdc45-MCM complex, DNA unwinding is initiated
- Loading of Cdc45 on pre-IC complex allows the recruitment onto the replisome of the three DNA-polymerases: DNA-polδ, DNA-polε, and DNA-polα

It is plausible that the formation of the pre-IC itself might dictate the replication timing of licensed origins. Indeed, whereas all the potential (licensed) ORIs are pre-RC-bound, the formation of the pre-IC is limited by the availability of the factors being assembled. Thus, ORIs with a higher affinity for those factors might have an earlier Replication Timing than the ones with a lower affinity. (Tanaka and Araki, 2013)

Activation of the pre-RC is a finely tuned process, with the pre-IC being target of several kinases involved in the regulation of the cell-cycle checkpoints, as seen for Rad53 (Lopez-Mosqueda et al., 2010; Zegerman and Diffley, 2010) and Chk1 (Boos et al., 2011) in studies involving buddying yeast.

Other factors contributing to selection of Replication Initiation Sites

Among the other factors that might strongly contribute to the process of origin selection, the most mentionable are *distal elements, chromatin structures and transcription*.

Distal elements

Two particular cases exemplify the role that genomic sequences distant from the ORIs play for the genomic replication of those loci: Chinese hamster DHFR locus (dihydrofolate

reductase)(Kalejta et al., 1998; Mesner and Hamlin, 2005) and the Human β-globin locus (Aladjem, 2007; Aladjem et al., 1995; Cimbora et al., 2000)

Human β-globin locus

Replication of the β-globin locus depends on an origin replication region placed in the intergenic region between the β-like and β-globin genes.

A natural form of β-thalassemia (known as Hispanic Thalassemia) deletes a region located ~55kb upstream of the β-globin locus. Deletion of the entire region, or part of it, whereas leaving the β-globin locus intact, caused a repression of transcription of β-globin genes in erythroid cells (Forrester et al., 1990). For such reason, that genomic region is known as Locus Control Region.

Further on, it was showed that, in transgenic mice bringing the same LCR deletion caused by Hispanic Thalassemia (Aladjem et al., 1995; Cimbora et al., 2000), also Replication Timing was altered, with a shift from early to late replication, most probably because replication of the β-globin locus was caused by an initiation region located many kilobases downstream (Aladjem, 2007).

DHFR in Chinese hamster ovary cells

In 1998, Kalejta and colleagues (Kalejta et al., 1998), and as successively reported by Mesner and Hamlin in 2005 (Mesner and Hamlin, 2005), found a 55kb intergenic region enriched with replication origins between DHFR and *2BE2121* genes in Chinese Hamster Ovary cells (CHOC).

In wild-type CHOCs, DHFR 3' UTR region successfully stops transcription of the gene at the 5' boundary of the initiation site, and a deletion of 45kb portion of the intergenic region (>90%) showed no effects on initiation of replication (Kalejta et al., 1998).

On the other hand, after the deletion of the 3' UTR of the DHFR gene, it can be appreciated an elongation of transcription of the locus up to 35kb within the intergenic

region, with dramatic reduction of the efficiency of the initiation of replication, and a replication timing undertaking the whole S-phase in opposition to the original 3-4 hours.

By restoring the deleted DHFR gene 3' UTR segment, the original efficiency of initiation of replication is fully restored, thus showing that the wild-type 3' end of the DHFR gene constitute a regulatory element not only for the gene, but also for the local cluster of replication origins (Mesner and Hamlin, 2005).

Chromatin structure

There are evidences showing that modification of chromatin architecture can affect the aggregation of the pre-RC which, in turn, might result in an impairment of the replication initiation.

Indeed, recent studies addressed the importance of chromatin structure not only for the transcriptional program, but also to regulate the events leading to the selection and activation of replication origins.

As showed in studies on yeast, initiation or replication is facilitated by mutation in histone deacetylases (HDACs), as in the case of Sir2 (Pappas et al., 2004).

In another case, in human cells, mutations on HBO1, an acetyl-transferase, negatively regulated the assembly of the pre-RC. Indeed, after the inhibition of HBO1 expression, the MCM complex failed to load onto chromatin, thus affecting the correct assembly of the pre-RC on replication origin, suggesting an active role of Acetylation and Deacetylation in the correct formation of the pre-RC (Burke et al., 2001; Iizuka et al., 2006).

Transcription and topology

Transcription can be regarded as both as a positive or negative regulator of initiation of replication of certain loci.

Indeed, actively transcribed gene usually coincide with a negative regulation of replication, enacted by either silencing of replication origins inside the gene (not to have RNA and DNA polymerases to occupy the same genomic region and thus avoiding collisions) or by reducing the size of the initiation of replication region (Mesner and Hamlin, 2005; Sasaki et al., 2006).

Conversely, transcription might be indirectly linked to activation of origins by the topological changes induced by the transcriptional machinery.

As reported in different works (Ghosh et al., 2004; Haase et al., 1994; Kohzaki and Murakami, 2005; Nieduszynski et al., 2005), transcription factors might aid selection of Initiation Sites by altering local chromatin structure. The chromatin structure generally inhibits DNA replication as well as transcription since it reduces accessibility of proteins to DNA. In transcription, transcription factors counteract this inhibitory effect of chromatin by changing the local chromatin structure surrounding their binding sites. The case of the β-globin LCR discussed above can be cited as an example. In fact, this region contains several binding sites for different transcription factors, whose binding might change the chromatin status in a way which is also favourable for replication.

As an example negative supercoiling is induced as a consequence of the progresses of the transcriptional bubble on the DNA and, on the other hand, negative supercoiling favours the unwinding of the DNA double helix.

In this way, the negative supercoiling caused by two transcriptional machineries progressing in diverting ways, might increase accessibility of chromatin and favour the binding of the pre-RC. This hypothesis seems to be supported by a recent finding

illustrating that divergent transcription takes place in most of transcriptional active genes. In such way, then, the genomic regions between the divergently moving RNA polymerases might be kept nucleosome free, allowing a more efficient origin activation by assembly of pre-RC (Seila et al., 2009).

TEMPORAL REGULATION OF DNA REPLICATION

Two are the main properties, which describe the temporal control of the replication route: the plasticity of the replication origins selection and their sequential activation.

Plastic Activation of Selected DNA replication origins

Involving only a portion of the licenced ORIs, at a first glance activation of replication origins might look like a very inefficient process.

Indeed, in yeast, only 50% of origins seems to be activated and, in metazoan, this efficiency sensibly drops to a mere 5-20% (Hamlin et al., 2008; Heichinger et al., 2006).

Though, what might be regarded as an ineffective process, actually, is one of the most powerful mechanisms to regulate the replicative process: the plasticity of the origin selection.

Such level of regulation was revealed in studies by Taylor in 1977 (Taylor, 1977), in which Chinese hamster ovary cells were subjected to a treatment by fluorodeoxyuridine, a compound that inhibits the conversion from dUMP (deoxyuracilmonophosphates) to dTMP (deoxythyminemonophosphates) thus causing a reduction in the available thymidine pool.

After treatment, DNA synthesis was analysed by labelling newly synthesized DNA with tritiated thymidine and visualized by DNA fibre autoradiography.

The results showed that, consequently to starvation for thymidine, distances between the labelled segments on DNA fibres were reduced, as were also the rates of fork movements (Gilbert, 2007).

Thus, when fork progression is impaired by external factors (like, in this case, a depletion of the nucleotide pool), cells try to compensate by increasing the number of active origins, in order to achieve the completion of the S phase within the normal time.

Conversely, by adding adenine and uridine, replication fork speed was accelerated, with a consequent reduction of active origins (Conti et al., 2007).

Accordingly, DNA replication origins can be divided into three classes: *constitutive*, *flexible* and *dormant* (DePamphilis, 1993) (Figure 5).

'*Constitutive'* are origins that are always used in any cell type and represent the minority of all of the DNA origins in a cell.

'*Flexible'*, instead, are origins which are stochastically used in different cells and/or in peculiar conditions (as the change of chromatin context during cell development and differentiation), and that might account for the low efficiency of origin activation in eukaryotes described above in this paragraph.

'*Dormant'*, on the other hand, are inactive origins which are set as a "back-up plan", in case the normal replicative process is hindered by either endogenous or exogenous events. What might seem a "surplus" of origins might instead play a role in the maintenance of genomic integrity. In fact, given the impossibility of the selection of new DNA replication origins in S phase, the activation of such origins might rescue situations in which forks are stalled or slowed down (Blow et al., 2011; Woodward et al., 2006).

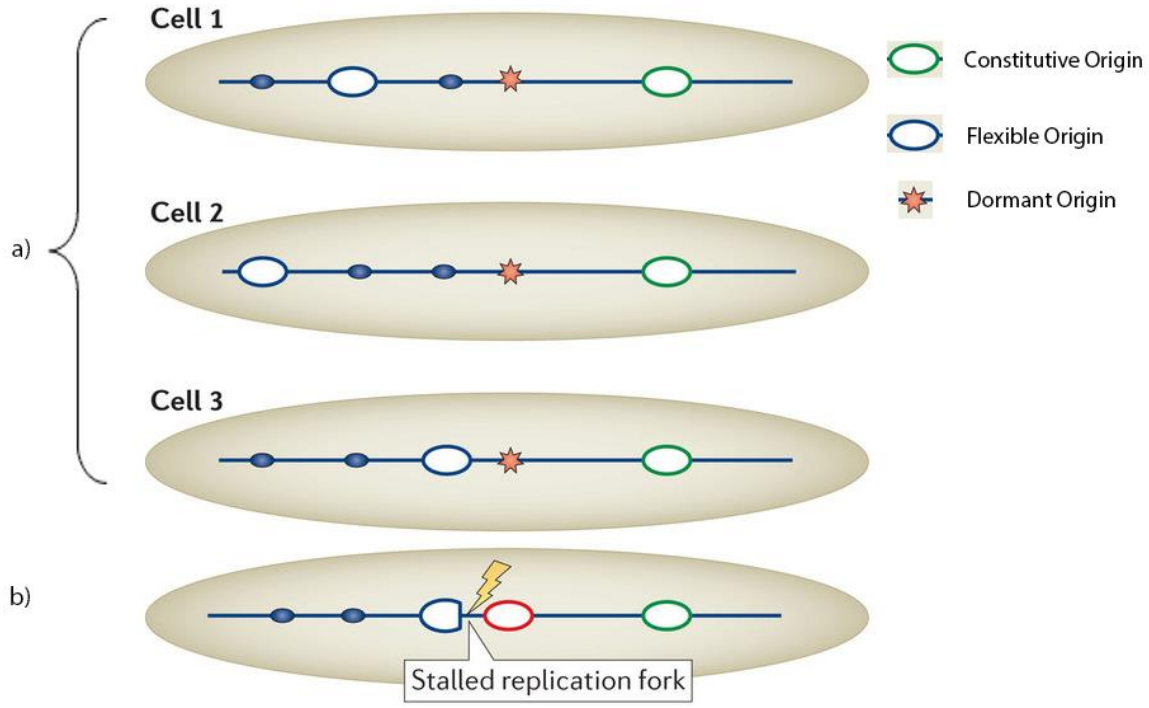

Figure 5. Different types of DNA Replication Origins – edited from (Fragkos et al., 2015)
a) DNA Replication Origins can be divided into three classes:
Constitutive ORIs: DNA Replication Origins which are always active without differences between cells in the same pool or even between cell lines
Flexible ORIs: These ORIs are activated in a stochastic way, so that some cells will "fire" them, whereas in same other cells they will be kept inactive
Dormant ORIS: This class of DNA Replication Origins are always keep inactive, and activated only in case nearby ORIs are unable to be activated.
b) Activation of a Dormant ORI: Dormant ORIs are intended as a "back-up plan" for cells in case something interferes with correct DNA replication. As an example, in case of a replication fork being stalled, nearby Dormant ORIs can be activated in order to resume the replication for that particular genomic region.

Sequential activation of Replication Origins – Replication Timing

The second layer of temporal regulation of DNA replication is the activation of replication origins at different times of the S-phase. Depending on which temporal point of the S phase they are activated, origins are classified as early, middle and late replicated ones (Figure 6).

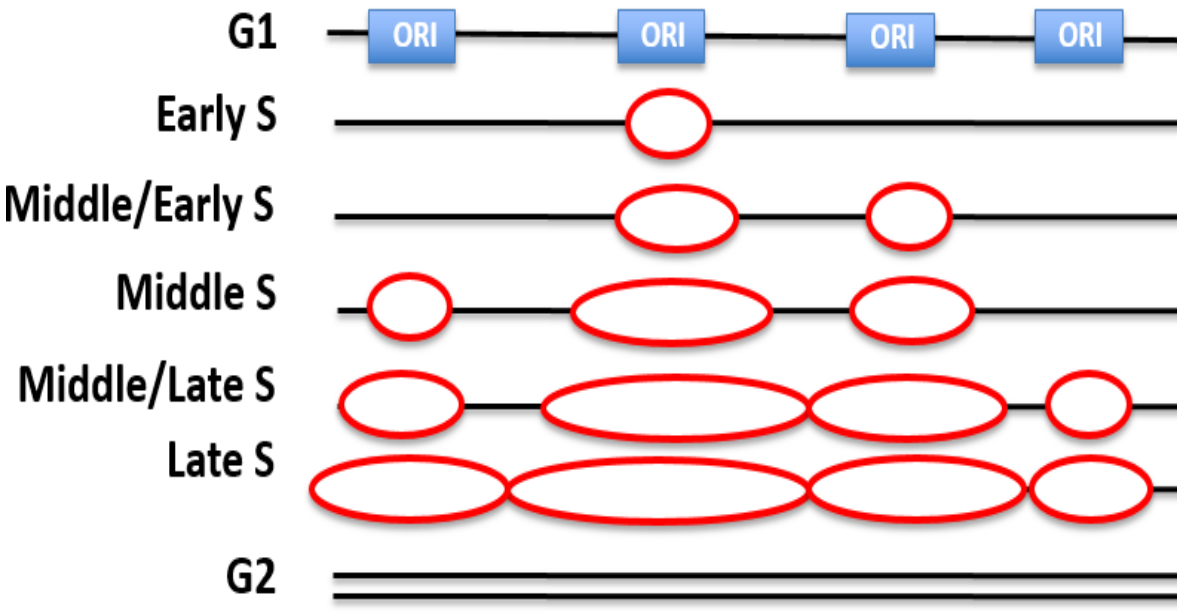

Figure 6. Spatial and Temporal regulation of DNA replication.
DNA replication must be tightly regulated in order to allow a faithful inheritance of the genome from a mother cell to two daughter cells. Such regulation exerts in two, interconnected, ways. First, a spatial regulation is carried out, with the selection of DNA Replication Origins. As a next step, following a temporally regulated pattern, selected ORIs will be activated in a sequential way, each with its own timing (Replication timing), throughout the S phase.

Biological importance of Replication Timing

As reported in a review by Pope, Hiratani and Gilbert (Pope et al., 2010a), the existence of a Replication Timing might be explained by the interconnection between replication and changes in the chromatin structure of the locus being replicated.

Indeed, if replication might change the chromatin status of the genomic region interested being actively replicated, then different chromatin conformations might be favoured during different parts of the S phase. To cite an example already reported, β-globin genes are replicated, and transcribed, in early S phase in erythroid cells, whereas in non-erythroid cells the same genes are transcribed and replicated in late S phase.

Experimental evidences seem to support this model. Two studies (Lande-Diner et al., 2009; Zhang et al., 2002) are reported as an example.

In one study (Zhang et al., 2002) the investigators injected a reporter plasmid in asynchronously rat cells, and consequently measured the expression of the plasmid, together with the nucleosome composition, in early and late replicating cells. What they found was that efficiency of transcription was much higher in early replicating cells, than in those in late replication, and that early S phases correlated with a higher histone acetylation.

In another study, (Lande-Diner et al., 2009), DNA plasmids were microinjected in replicating cells and their acetylation levels studied. As a consequence, plasmids injected in early replicating cells, that therefore did not undergo any further replication, remained acetylated, whereas those microinjected in late S phase were found packed with histones and, therefore, presenting a deacetylated state.

Another study conducted on Drosophila (Bell et al., 2010), showed a positive association between active and repressive histone modifications with early and late replication timing.

Indeed, Bell and colleagues found that, whereas early replicating regions showed an enrichment for acetylation (H4K16Ac), a notorious marker for “open chromatin”, histones

in late replicating regions were “decorated” by methylated residues (H3K27me3), showing reduced accessibility and chromatin compaction.

Replication Timing has also a fundamental role in cell development and differentiation.
As reported in different studies in literature (Hiratani et al., 2008, 2010; Ryba et al., 2010) conducted on both murine and human systems, chromatin reorganization during cellular development and differentiation is associated with a change of replication timing of large (200kb-400kb) replication domain.
During differentiation of both mESCs (murine Embryonic Stem Cells) and hESCs (human Embryonic Stem Cells), early to late shifts of replication timing can be observed in regions associated with down-regulation mESC-specific genes and chromatin inaccessibility, whereas late to early shifts were seen in regions associated with up-regulated genes associated with a particular cellular “differentiated state”.
An “extreme” example to an early-to-late shift in replication timing associated with stable lineage-dependent repression of chromatin status can be found during the inactivation of the X chromosome (Xi, “X inactivation”).

Interestingly, stable repression of chromatin status occurred during Xi is maintained in partially induced Pluripotent Stem Cells (piPSCs) which reprogramming was obtained from partially differentiated mESCs which already underwent early-to-late shifts in replication timing (which, in females, includes Xi). Conversely, except for those replication domains interested in Xi, all the others early-to-late lineage-dependent shifts in replication timing could be reverted (Hiratani and Gilbert, 2010).

Moreover, Replication Timing is an evolutionary conserved feature among closely related cell lines, as observed in comparative studies in mESC, hESC and differentiated human cell lines (Ryba et al., 2010).

In their work, the authors underlined how close similarities in Replication Timing, for the same replication domains, existed between in different species for similar cell lineages as for hESC, mESC and mEpiSCs (mouse Epiblast-derived Stem Cells), and between mouse and human lymphoblasts.

On the other hand, Ryba and colleagues observed significant differences in Replication Timing could be observed between cell lines in different developmental stages, such as between hESC/mESC and lymphoblastoid cells, with changes involving more than one third of the genome.

In addition, Replication Timing correlates with spatial compartmentalization of chromatin in nuclear territories, with a radial symmetry between late replicating loci, localized on the nucleus periphery, and early replicating domains, placed more toward its centre.

Coherently, shifts in replication timing during development correlate with chromatin spatial reorganization, with late-to-early and early-to-late shifts in Replication Timing causing a "swapping" of positions in replication domains for nuclear territories (Hiratani et al., 2008, 2010).

Another plausible option for the existence the replication timing associates the regulation of replication timing with preservation of important genomic regions.

For example, it is known that early replicating regions have a lower mutational rate when compared to late-replicating regions (Chen et al., 2010; Lang and Murray, 2011; Stamatoyannopoulos et al., 2009).

Therefore, in a scenario where late replication timing is directly associated with mutation rates, the replication of gene-rich genomic regions in the early moments of the S phase might reflect an evolutionary pressure to concentrate the nucleotide pools and DNA-repair capacity in early S phase, to lessen the probability of mutational events in those regions.

As an example, early replicated regions might be the first to have the possibility to undergo DNA repair by homologous recombination, being their sister-chromatid template immediately available as long as the replication process proceeds.

Causes and consequences of changes in Replication Timing

Given the vast association that Replication Timing has with transcription, and chromatin structures, and given the pivotal regulative role Replication Timing has on cellular differentiation and development, it comes natural that a deregulation of such process would have dramatic consequences on the cell fate is largely expected.

Therefore, defects in replication timing might be caused, or mirrored by, genetic changes, changes in gene expression and epigenetic changes.

Genetic changes

A first association between genetic changes and deregulation of replication timing can be appreciated by exposing cells to different DNA damaging agents. Indeed, cells treated with hydrogen peroxide or ionizing radiation, show an aberrant replication timing, up to a whole-chromosome delay of replication timing (Brás et al., 2008; Breger et al., 2004; Korenstein-Ilan et al., 2008).

A chromosome-wide delay might, in turn, become a cataclysmic genome-wide event. Indeed, a delay in replication timing interesting a whole chromosome might result in chromosome structure instability, with a possible higher rate of chromosomal rearrangement.

Even though such structural rearrangements are usually seen only on the concerned chromosome, other chromosomes might engage in inter-chromosomal translocation with the delayed chromosome. Such event closely resembles *chromothripsis*, a phenomenon

present in some cancer malignancies, in which one or multiple chromosomes, or chromosome arms, are fragmented and then reassembled at random (Breger et al., 2005; Stephens et al., 2011).

Changes in gene expression

As already extensively addressed before, there is a strong mutual interconnection between replication and transcription. In addition, different studies have shown that the deregulation of the expression of key-genes involved in replication (such as proteins of the ORC complex and CDKs) is followed by abnormalities in replication timing.

Indeed, in Drosophila, it was observed that a mutation in ORC2, one of the six subunits of the ORC complex, is responsible for an aberrant loading of the pre-RC onto DNA replication origins, causing an elongation of the S phase with a consequent early-to-late shift of Replication Timing, with some regions of euchromatin replicating later than heterochromatin (which usually is replicated in the very last moments of the S phase) (Loupart et al., 2000).

In another study conducted on Xenopus Laevis, it was seen that mutations targeting Cdc7 and Cdk2, two S phase–promoting kinases (Shechter et al., 2004),resulted in an anomalous origin licensing and activation, with consequent deregulation of Replication Timing.

Moreover, replication timing is impacted by deregulation of expression of genes involved in chromatin modification. In fact, the deregulation of expression of chromatin-modifying genes can cause long-range repercussions, with a change of replication timing of distant regions, even on chromosome different from the one being directly interested. This situation has proved to be true by a study conducted on Drosophila, where the deregulation of HP1 (Heterochromatin Protein 1) caused the change of replication timing in 5-10% of total genomic loci (Schwaiger et al., 2010).

Epigenetic changes

Different studies have proven that chromatin acetylation or methylation profoundly impacts on replication timing.

As a proof of concept, a study can be cited (Goren et al., 2008), where the investigators measured acetylation levels in the β-globin domain in erythroid and non-erythroid cells, retrieving and hyper- and hypo-acetylated chromatin status in the former and in the latter, respectively.

In addition, to test if acetylation occurring at replication origins could be representing an epigenetic mark for early replication, they forcefully acetylated the β-globin domain in non-erythroid cells, and caused deacetylation in the domain present in erythroid cells, and found that, consequent to acetylation and deacetylation, late-to-early and early-to-late shifts could be observed, respectively, in non-erythroid and erythroid cells. (Goren et al., 2008).

In the same fashion, other different studies have also proved that changes of the histone marks signatures inevitably follow a change in replication timing (Aparicio et al., 2004; Vogelauer et al., 2002a; Zappulla et al., 2002).

Aberrations in Replication Timing and cancer

Defects or abnormalities in replication timing seem to be linked with various types of cancer, such as breast (Fritz et al., 2013; Grinberg-Rashi et al., 2010; Kazanov et al., 2015), prostate (Dotan et al., 2004) and lung cancer (Kazanov et al., 2015), and different scenarios can be hypothesized to explain this phenomenon. As an example, as previously addressed, DNA damage might bring deregulation of replication timing which, in turn, might cause genomic instability. Genomic region particularly susceptible to DNA damages related to shifts in replication timing are the so-called transition regions, genomic loci

placed between early replicating and late replication domains. In particular, whereas in replication domains forks move bidirectionally, transition regions are replicated only by one replicative fork. Thus, a stall or an aberrant progression of the fork throughout the genome might be reflected into an aberrantly replicated transition regions, with either short or long-distance repercussions on nearby genomic loci (Watanabe and Maekawa, 2010).
Or, alternatively, an early-to-late shift in a particular genomic region might result in the silencing of a tumour-suppressor gene (De and Michor, 2011).
Indeed, as briefly described before, early replicating regions tend to have a lesser amount of mutations than late-replicating ones, with late replicating regions presenting an higher degree of deletions than their early-replicating counterpart.

Another way to promote genomic instability and cancer transformation is achieved by disrupting the temporal program of replication origins activation, such as increasing the number of active origins per genomic region. Such is the case, as an example, of Myc, Cyclin-E and H-Ras which, if overexpressed, induce DNA damage by a hyper-activation of DNA replication origins (Di Micco et al., 2006; Dominguez-Sola et al., 2007; Jones et al., 2013).
Indeed, an hyperactivation of replication origins might create conflicts between replication and transcription machineries, leading to DNA damage and fork stalling. On its turn, fork stalling might induce activation of dormant origins, thus increasing the chances of fork-to-fork collisions and further accumulation of DNA damage.

Moreover, it has been showed that changes in replication timing play a critical role in leukemic cells, with 10-20% of genome being replicated with different replication timing when compared with normal controls (Ryba et al., 2012).
Change of replication timing was not restricted to one particular genomic region, but

interested all chromosomes, and was evenly spread throughout the genome. Furthermore, even with slight differences depending on the type of leukaemia, changes in replication timing were common to all collected samples.

Many of the shifts in replication timing occurred near regions undertaking genomic rearrangements, with one change in replication timing occurring in a common site of translocation in leukemic cells.

Even if all of the samples shared the same change of replication timing associated with translocations, not all of them actually showed the translocational event, thus suggesting that changes in replication timing might possibly predispose chromosomes to translocation.

Finally, all of the replication changes were generally late-to-early shifts, with only fewer early-to-late shifts in replication timing.

Genome-wide studies on DNA Replication – Identification of ORIs

Some noteworthy efforts have been made to identify ORIs and characterize their features. Unfortunately, as reported in a review by Gilbert (Gilbert, 2010), all of the studies conducted so far show a lack of reproducibility. Up to now, the only two methodologies use to gain information about ORIs are based on the isolation and successive hybridization of short nascent strands (SNS) or onto the isolation of the replication bubble.

Isolation and hybridization of short nascent strands

This technique is based on the knowledge that origins move bi-directionally and therefore, for each origin, there are two protruding newly-replicating single-stranded DNA strands. In theory, if applied on asynchronously growing cells, the abundance of short nascent strands (SNSs) should be proportioned to the number of ORIs being activated throughout the

whole duration of the S phase. In practice, this technique has different practical drawbacks. The major is that, since SNSs just represents origins being fired, it might not be a complete map of all (or most of) the licensed origins in the cell population.

Two studies using this technique in Hela Cells (Cadoret et al., 2008; Karnani et al., 2010) identified nearly one thousand ORIs, covering approximatively 1% of the human genome as mapped by the ENCODE (https://www.genome.gov/10005107/) database. Though, even if using the same methodology, they showed an overlap of just 14%, even if the identified ORIs were all enriched for TSSs. (Transcription Start Sites).

More recently, another study used the SNSs technique coupled with next-generation sequencing (Martin et al., 2011), similarly describing an enrichment of ORIs near TSSs and hypothesizing a possible interconnection between DNA transcription and replication.

Bubble trap

This technique relies on the hypothesis that actively replicating ORIs, given the bi-directional progress of the replication fork, are virtually circular DNA molecules. Such molecules could be then "trapped" in an agarose gel, the DNA fragments purified, cloned into a vector and successively analysed (Mesner and Hamlin, 2009). Differently from the SNSs technique, this method can potentially allow not only the identification of actively replicating origins, but of entire clusters (Mesner et al., 2011).

However, even though also the bubble-trap technique shows an enrichment for ORIs near TSSs and genebodies, this methodology displayed little accordance with the studies conducted by the SNSs technique, with an overlap of ~10-33% (Gilbert, 2010) (Figure 7).

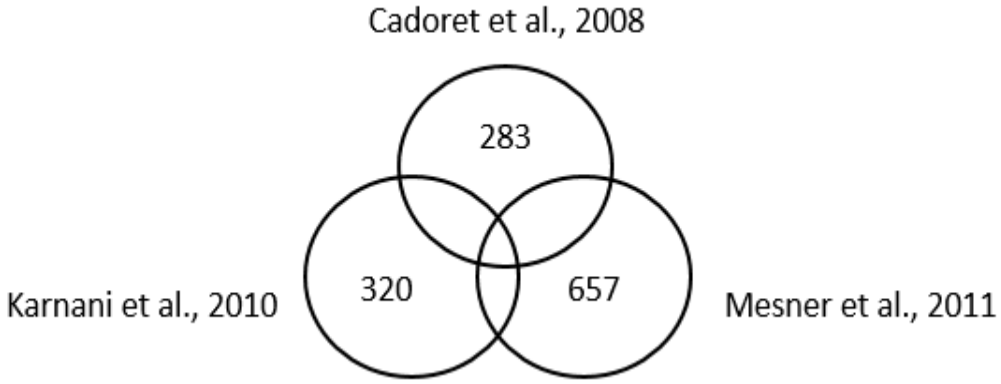

Figure 7 – Overlap between different studies aimed at the identification of ORIs.

Mainly, two methodologies have been used attempting to the identification of ORIs: SNS and Bubble-trap. Unfortunately, experiments run with both techniques not only have uncovered only ~1% of the genome but also showed a poor overlap among them.

Orc1 as a proxy for genome-wide of ORIs

A novel approach, different by the aforementioned two, is the one performed in HeLa cells by Dellino and colleagues (Dellino et al., 2013a).

By putting together an ultracentrifugation step of cross-linked chromatin, anti-ORC1 ChIP-sequencing and Repli-seq (a methodology which will be described shortly) into a novel approach, it was possible not only to exactly pinpoint the position of ORIs all over the genome, but also to know in which fraction of the S phase such origins were activated.

In this study, the investigators found ~13,000 ORIs, more than ten times the number of DNA replication origins found by the two aforementioned techniques. Furthermore, they were able to link both the number and the efficiency of ORIs with replication timing, showing that as the number of ORIs progressively decreases during the S phase, with ~4,000 ORIs in S1 and ~700 in S6, so was the height of ORC1 peaks.

In addition, in concordance with literature, they found a strict association with ORIs and expressed genes, with ~17% and ~28% of ORC1 sites co-localizing near highly and

moderately expressed genes (high efficiency ORIs) and with the remaining ~55% being localized near TSSs of ncRNAs (non-coding RNAs) or within the genebody (low efficiency ORIs).

Genome-wide studies on DNA Replication - Repli-seq

Characterization of Replication Timing as a proxy to study of the DNA replicative process and its regulation, and effects on it deregulation, offers several advantages than the identification of Replication Origins.

As an example, characterization of Replication Timing offers a better reproducibility (Yue et al., 2014). In addition, contrarily to studies conducted on ORIS, it allows a genome-wide characterization of DNA Replication and its different features (Chen et al., 2010; Farkash-Amar and Simon, 2010; Pope et al., 2011; Gilbert, 2010)

A huge contribution to methodologies concerning the study of Replication Timing has been given by Hansen and colleagues, in 2010 (Hansen et al., 2010a), when they coupled part of an already existing methodology used for the study of Replication Timing (Hansen et al., 1993) with next-generation technology.

By this new technique, named Repli-seq, they were able to study changes in Replication Timing across different cell types, for the same genomic region.

By this technique, as in the STS method from which it derives, asynchronously growing cells are subjected to Bromodeoxyuridine (5-bromo-2'-deoxyuridine, BrdU), a synthetic analogue of thymidine. Such pulse allows labelling of newly-replicated DNA, which incorporates this compound in place of thymidine. Successively, labelled cells are sorted into different sub-fractions, spanning all of the S phase (the cell cycle DNA synthesis phase). Sorted cells are therefore lysed, and BrdU-hybridized DNA is fragmented and immunoprecipitated by the mean of an anti-BrdU antibody. Finally, BrdU-labelled DNA

fragments are sequenced.

Once plotted on Genome Browser, the ordered profiles of BrdU enriched S-phase fractions visually shape in space and time the progression of the DNA replication throughout the genome. At low magnification (MegaBases scale) the whole Repli-seq profile is characterized by a series of early-to-late transitions (also called “Inverted-Vs”, because of their shape), being originated from presumably ORIs enriched regions (Figure 8).

In his work, Hansen and colleagues compared Repli-seq profiles among different cell lines, studying how the Replication Timing for the same regions might change depending on the cell type or the cell differentiated status (Figure 9). Moreover, they took advantage of the Repli-seq technique to study, among the other things, an association between Replication Timing, transcriptional activity and chromatin status.

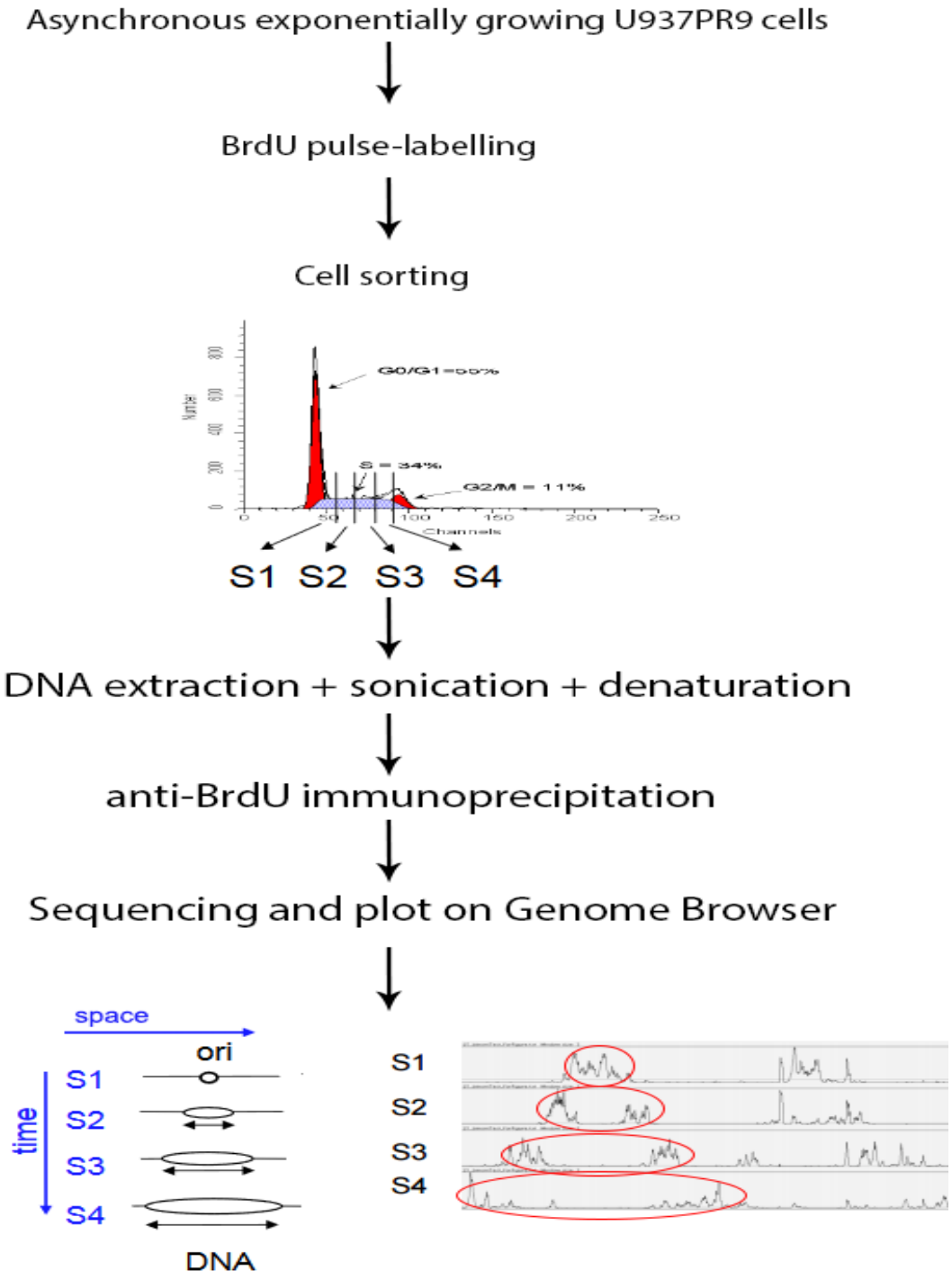

Figure 8. Repli-seq workflow– edited from

(https://www.normalesup.org/~vorgogoz/BioInfoCourses/37-recombination-Cheng.pdf; Dellino et al., 2013)
In Repli-seq technique, asynchronously growing cells are subjected to a BrdU pulse, then sorted based on their DNA content**.** Subsequently, BrdU-labelled DNA is fragmented and sequenced. By aligning sequenced reads to the genome, for each S phase fraction in which cells have been sorted, is possible to appreciate the progression of the replication forks throughout the genome in both space and time.

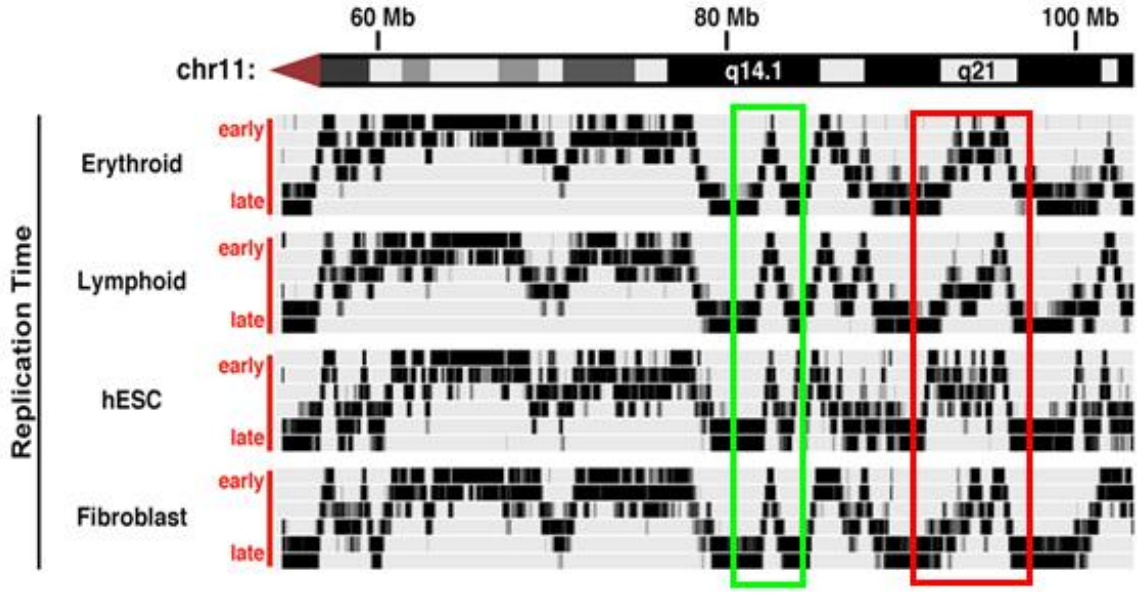

Figure 9. Visual comparison of Repli-seq profiles among different cell lines – edited from (Hansen et al., 2010a)
In their work, Hansen and colleagues showed that Replication Timing is a feature conserved across several cell lines (green square). Differences in Replication Timing among cell lines, are usually characteristic of a particular cell lines, thus making Replication Timing a pivotal feature for cell development and differentiation.

In the same year, Chen and colleagues refined the Repli-seq method, by estimating a Replication Timing quantitative measure computed from the signal of the S-phase fractions (Chen et al., 2010), which alone was capable to summarize in single value the timing of the Replication process of each portion of the genome (Figure 10).

Such profile derives from a series of values, named "Replication Time estimator S50" (from now on referred as "S50" or "S50 Value"), ranging from 0 to 1. An S50 value is defined as "the fraction of the S phase ($0 < S50 < 1$) at which 50% of DNA is replicated (50% of the cumulative enrichment) in a defined genome region ".

S50 values were computed by linearly interpolating the signals coming from all of the S phase fractions, as described in a previous work by Jeon et colleagues, in 2006 (Jeon et al., 2005).

Thus, early replication corresponds to low S50 values, whereas high values correspond to a later Replication Timing.

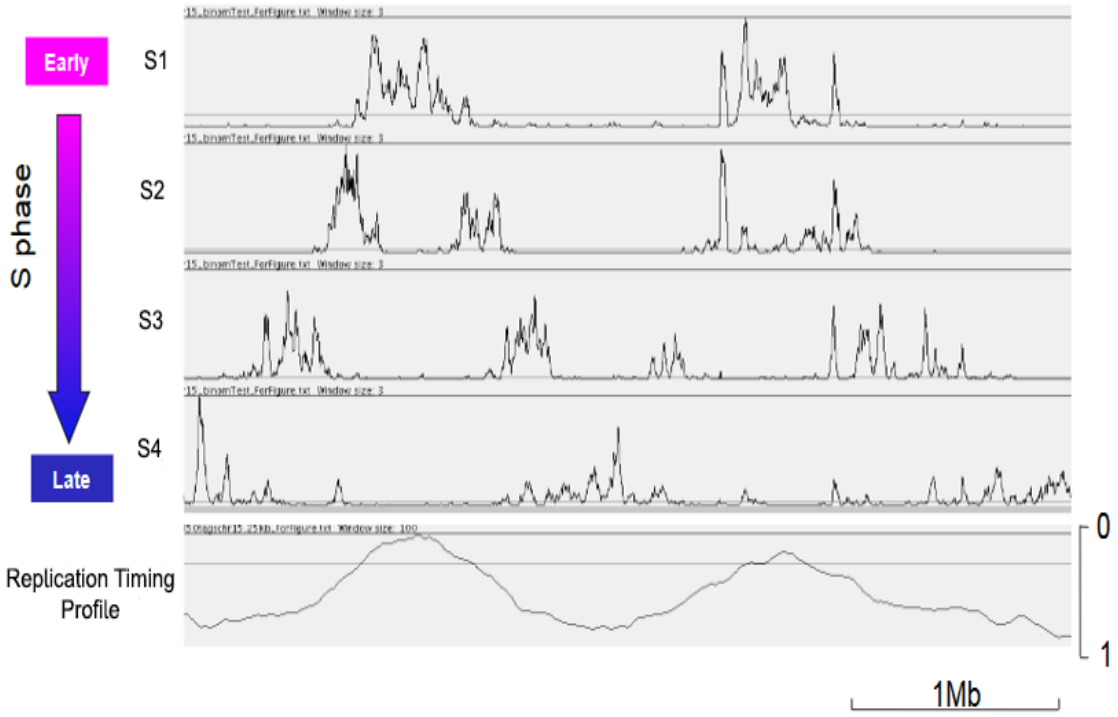

Figure 10. Replication Timing profile – edited from https://www.normalesup.org/~vorgogoz/BioInfoCourses/37-recombination-Cheng.pdf
From different S phase fraction signals, it is possible to calculate a unique Replication Timing Profile. Such profile derives from consecutive values (S50 Values). Each value, consisting in the 50% of the amount of the DNA replicated in a genomic interval, is an indirect measurement of the Replication Timing for each interval. Furthermore, being S50 values associating amount of DNA being replicated with Replication Timing, small S50 values will be addressing early replicating regions, whereas to high S50 values will be corresponding late replicating regions.

Up to now, Repli-seq has been used in different studies (Barlow et al., 2013; Debatisse et al., 2012; Liu et al., 2015; Polak et al., 2015). In particular, one study (Pope et al., 2014) implied massively use of the Repli-seq technique, taking advantage of fifteen different Repli-seq experiments performed on just as many different cell lines.

Unfortunately, though, up to now Repli-seq technique has not been used as a main tool of analysis, but only as a way to characterize and compare results with other techniques (see below).

Leading Hypothesis

The hypothesis underlying this study has been driven by different evidences linking both deregulation of transcription and disruption of the normal chromatin status with deregulation of DNA replication.

Indeed, earlier in this chapter, it has been shown that deregulation of genes involved in the first steps of the replicative process is usually followed by deregulation of the replication timing (Loupart et al., 2000; Shechter et al., 2004).

Replication Timing might also be altered by a deregulation of normal chromatin status, with hyperacetylation bringing a late-to-early shift and deacetylation causing instead a late replication in regions normally being replicated earlier in the S phase (Göndör and Ohlsson, 2009; Goren et al., 2008)

Furthermore, it has been shown that 10-20% of the genome in leukemic cells is replicated with a Replication Timing different from control cells. Importantly, deregulation of Replication Timing appears to be diffused along the genome, and not restricted to particular genomic areas (Ryba et al., 2012).

Given this, we hypothesized an active role for PML-RARα, the initiator oncogene in APL (Acute Promyelocytic Leukaemia), in the deregulation of normal Replication Timing.

Indeed, given the close relationships among DNA Replication, Chromatin status and transcription, we hypothesize that PML-RARα might exert part of its oncogenic activity by deregulating the normal Replication Timing of a cell, as a consequence of the deregulation of transcription and change of normal chromatin status.

To test this hypothesis, we took advantage of U937PR9 cells (Casini and Pelicci, 1999) as our model system. U937PR9 cell line is a stable clone obtained from monoblastic U937 by transfection of PML-RARα cDNA under the control of a Zinc-inducible metallothionein promoter. In normal conditions,U937PR9 cells do not express PML-RARα, whereas the oncogene is expressed after induction by Zinc, in concentrations comparable to the ones

found in APL patients.

To study changes in Replication Timing after expression of PML-RARα, we will take full advantage of the Repli-seq technique, measuring replication timing before and after induction by Zinc in cells asynchronously replicating for 24 hours.

Moreover, to study whether the deregulation of the normal transcriptional programme or the normal chromatin status by PML-RARα is able to modify Replication Timing, we will take advantage of RNA-sequencing (from now on referred as RNA-seq) and anti-H3K4me3, anti-H3K4me1 and H3K27Ac ChIP-sequencing (from now on referred as ChIP-seq) performed on both normal samples and cells retrieved after 8 hours of induction by Zinc.

Acute Promyelocytic Leukaemia

Acute promyelocytic leukemia (APL), a sub-type of Acute Myeloid Leukemia, is one of the most studied and understood at molecular level blood malignancies. It is characterized by a block of differentiation of myeloid precursors, which show a promyelocytic cytomorphology, and by a reciprocal translocation between chromosome 15q24.1 and chromosome 17q21.2 (Figure11).

The reciprocal t(15;17) (q24.1;q21.2) translocation found in APL gives rise to the expression of a fusion protein, resulting from the fusion of the PML (ProMyelocitic Leukemia) locus with the RARα (Retinoic Acid Receptor alpha) one.

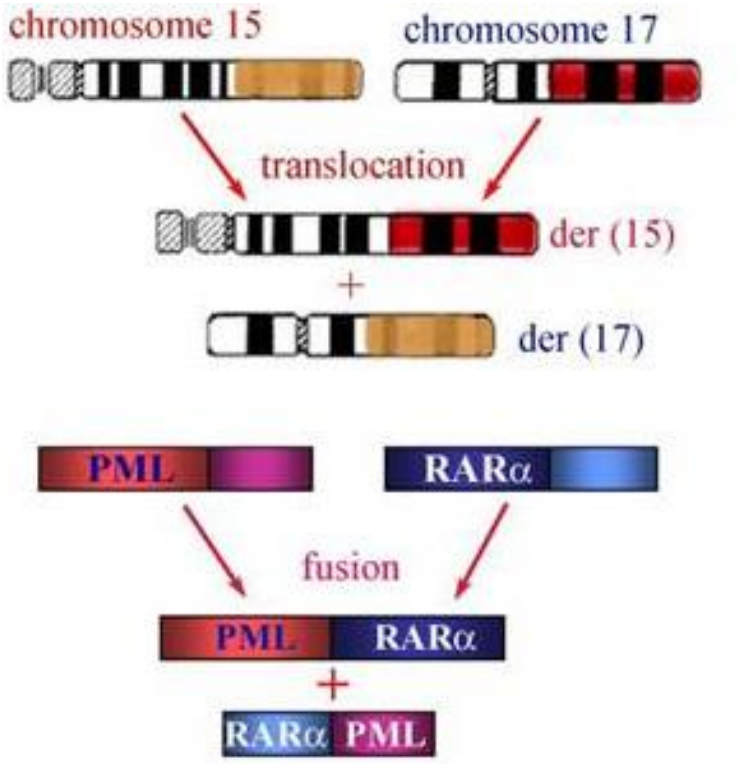

Figure 11. Reciprocal translocation between chromosome 15 and chromosome 17 results in PML-RARα fusion protein – edited from (Wang and Chen, 2008)
Reciprocal translocation between chr15q24.1 and chr17q21.2. The main result of such translocation is a functioning fusion protein, PML-RARα, together with RARα-PML and an aberrant PML isoform.

The deriving fusion protein, PML-RARα, is believed to suppress PML activity (which normally has a pro-apoptotic and growth suppressor activity) and is known to block differentiation of myeloid progenitors by repressing genes normally targeted by retinoic acid (Kakizuka et al., 1991; de Thé et al., 1991).

APL is morphologically categorized by the French-American-British (FAB) classification as AML-M3. Up to now, two morphologic variants of APL are known: a "typical" (hypergranular) and a "microgranular" (hypogranular) APL.

Hypergranular APL is the most common form, with ~70% of cases of APL showing its morphological features, such as cells presenting bilobed nuclei, and numerous large, red to purple cytoplasmic granules (Sainty et al., 2000).

Typical APL is traditionally treated with ATRA (all-trans retinoic acid) (Grimwade et al., 2000), which is capable to revert of the block of differentiation induced by PML- RARα fusion protein (Figure 12).

Untreated APL has a median survival of less than one month. On the other hand, prognosis for APL patients is the better among all the other AML subtypes, with higher chances to achieve complete remission after treatment.

Moreover, clinical trials have proved that 90% of patients affected by APL with classic t(15;17) achieve full remission, with low risks of relapse , after treatment consisting of ATRA, ATO (arsenic trioxide) and chemotherapy (Wang and Chen, 2008)

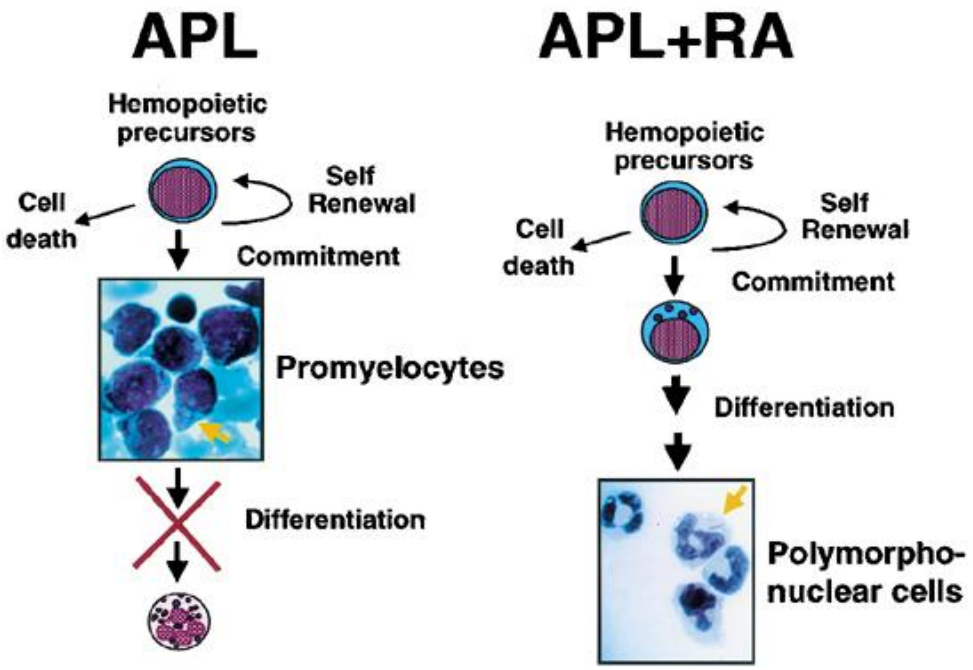

Figure 12 – Block of differentiation in APL and reinstate of myeloid differentiation after treatment by Retinoic acid – edited from (He et al., 1999)

Expression of PML- RARα causes block of differentiation and abnormal proliferation in Myeloid Progenitors (left). After treating leukemic cells with Retinoic Acid (RA), cells revert block of differentiation and resume cellular differentiation (right).

PML-RARα

In 1993, Grignani and colleagues (Grignani et al., 1993a) identified PML- RARα as the initiator oncogene in APL. The identification was brought by several findings. As an example, the (15q24.1;17q21.2.) translocation is found in all of the APL cases, and in 90% of times it is the only anomaly discernible in APL patients (Biondi et al., 1991; Pandolfi et al., 1992; Tallman and Kwaan, 1992) and that, despite the fact that the translocation might imply longer or shorter chromosomal fragments, there will be always formed a functional PML- RARα fusion, with a PML and a RARα domains retaining their wild-type functions (*i.e.* DNA-binding domain for PML retinol and DNA-binding domain for RARα) gene (Pandolfi et al., 1992). In addition to PML- RARα, other two proteins, aberrant PML and RARα-PML, are generated by the translocation, but won't be discussed here.

Usually, oncogenic transformation is acquired after a series of consecutive events.

In APL, instead, it looks like a single translocation is able to drive an oncogenic transformation, alone. This property might be given by the arising, from this translocation, of three functioning aberrant proteins (Grignani et al., 1993a), which might greatly reduce the number of additional events in order to bring oncogenic transformation.

In addition, since to PML- RARα retains both PML and RARα functional domains, the fusion protein alone is able to tamper with the endogenous PML and RARα pathways.

Indeed, evidences show that PML- RARα, in vitro, has an affinity and binding specificity to retinoids equal to the wild-type RARα (Nervi et al., 1992) and that it is able to regulate RARα target genes (Nervi et al., 1992; de Thé et al., 1991, Casini and Pelicci, 1999).

As been showed by Grignani et al, treating APL patients with Retinoic Acid induces differentiation of RA PML/RARα-expressing cell, with a depletion of PML/RARα expression. Such phenomenon, might be indeed due to the retain of the wild-type RARα properties by PML/RARα induces differentiation of PML/RARα-expressing cells, both in

the patients (Casini and Pelicci, 1999; Grignani et al., 1994)

Furthermore, PML- RARα is able to dimerize with wild-type PML (Kastner et al., 1992). PML is the main component of the so-called “PML Nuclear Bodies”, nuclear structures which are supposed to have a role into transcription, DNA replication or post-translational modifications (Lallemand-Breitenbach and de Thé, 2010). Therefore, dimerization of PML- RARα with endogenous PML might contribute to the deregulation of the cellular processes regulated by the PML Nuclear Bodies.

Mechanistically, PML- RARα exerts its block of differentiation and suppression of transcription of retinoic acid target genes by recruiting the histone deacetylase (HDAC) complex (Grignani et al., 1998) and, furthermore, is has been shown that PML- RARα is also able to cause hypermethylation of target promoters by actively recruiting methyltransferases (Croce et al., 2002), such as Dnmt1 or Dnmt3a.

In addition to its role as a repressor, Alcalay and colleagues (Alcalay et al., 2003) found that PML- RARα might also play as a positive regulator of transcription. Indeed, PML-RARα seems to induce transcription in genes already targeted by other fusion proteins commonly expressed in Acute Myeloid Leukaemia, such as AML1-ETO and PLZF-RARα.

MATERIALS AND METHODS

MODEL SYSTEM – U937PR9 CELLS

U937 is a hematopoietic cell line derived from a lymphoid malignancy of a 37-years old patient suffering from histiocytic lymphoma. However, these cells retain myeloid features, rather than lymphoid.

As an example, these cells have the ability to secrete lysozymes, usually produced by monocytes and macrophages. Indeed, high levels of lysozyme are usually found in monocytic and myeloid malignancies and not in lymphoid leukaemias (Sundström and Nilsson, 1976).

In addition, further phenotypical, cytochemical, and functional analyses have proved that U937 cells closely resemble myeloid precursor with a blocked differentiation at the promonocytic stage. (Nilsson et al., 1981).

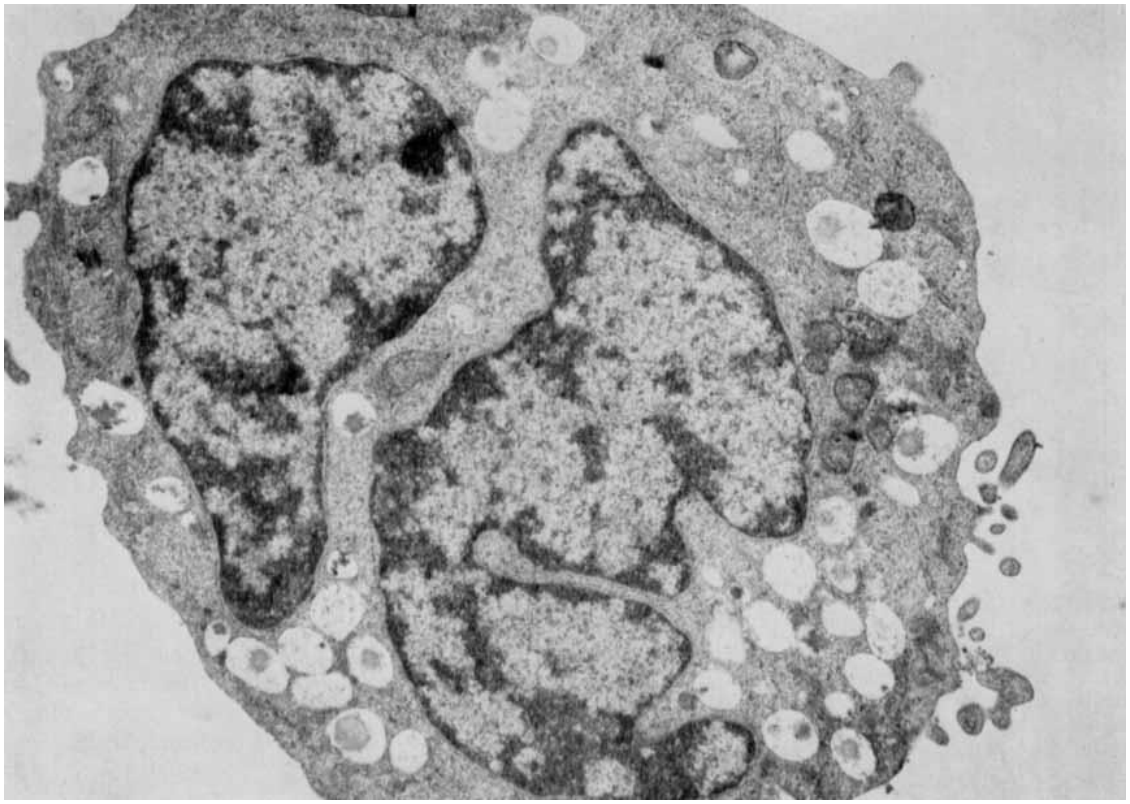

Figure13. A U937 cell (Sundström and Nilsson, 1976)
In this picture, taken by transmission electron microscope (x12,000 enhancement), it is possible to appreciate a lobated nucleus, together with numerous cytoplasmatic vesicles.

Furthermore, when stimulated by dihydroxyvitamin D3 and transforming growth factor β1, U937 cells are able to revert the differentiation block in myeloblastic stage and proceed toward monocytic terminal differentiation (Testa et al., 1993).

Thus, U937 are per se a good model to test the effects of expression of PML-RARα on cellular differentiation.

In 1993, Grignani and colleagues (Grignani et al., 1993) successfully engineered U937 cells by transducing them with a vector carrying PML-RARα cDNA under control of the inducible mouse metallothionein 1 (MT-1) promoter (Figure 14).

Figure 14 – Original Schema of pMTPR plasmid used to transduce U937 cells - from (Grignani et al., 1993b).
Mo-MT-1: mouse MT-1 promoter; PML-RARα: cDNA of fusion protein PML-RARα; SVpA: SV40 (Simian Virus 40) polyadenylation (termination) region; SV40neo: Neomycin (Neo) gene under control of SV40 promoter-enhancer.

Upon infection and selection, the PR9 clone was chosen, based on the high levels of PML-RARα expressed upon induction by ZnSO4 (Zinc) already after 24 hours.

More importantly, PML-RARα levels were comparable with the ones expressed in fresh APL blasts collected from patients.

Therefore, for the aim of this project, in order to study possible changes in Replication Timing after expression of PML-RARα, we took advantage of a pool of exponentially growing U937PR9 cells inducted by Zinc for 24 hours, and a control pool of U937PR9 cells without any induction.

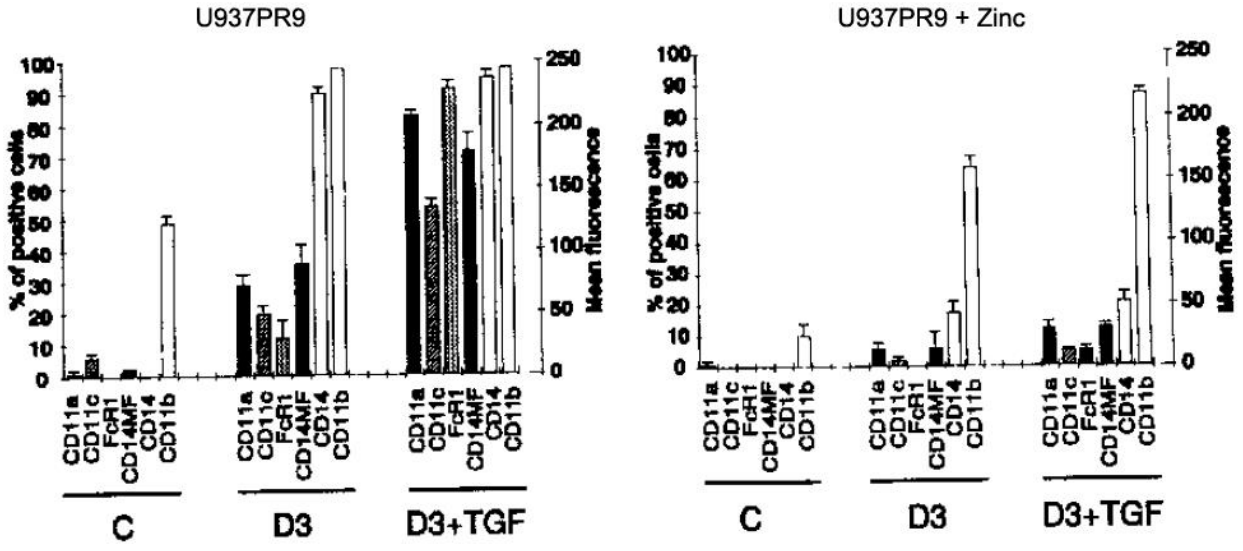

Figure 15. Expression of differentiation antigen before (left) and after expression of PML-RARα induced by Zinc (right) – edited from (Grignani et al., 1993b)**.**

The expression of several differentiation factors (CD11a, CD11b, CD11c, CDC14 and FcRI) was analysed in U937PR9 cells with and without induction of PML-RARα expression by 100 μM ZnSO4, in control cells (C), cells treated with dihydroxy vitamin D3 (D3) and cells treated with both dihydroxy vitamin D3 and Transforming Growth Factor β1 (D3+TGF).

Results are expressed as the percentage of cells positive for the antigens for CD11a, CD11b, CD11c, CDC14 and FcRI (left) or the mean fluorescence for the CD14 marker (CD14 Mean Fluorescence, CD14MF, right).

In the left panel, U937PR9 resume differentiation after treatment with both D3 or D3+TGF, whereas in U937PR9 expressing PML-RARα after induction by Zinc differentiation is blocked even after treatment with either D3 or D3+TGF.

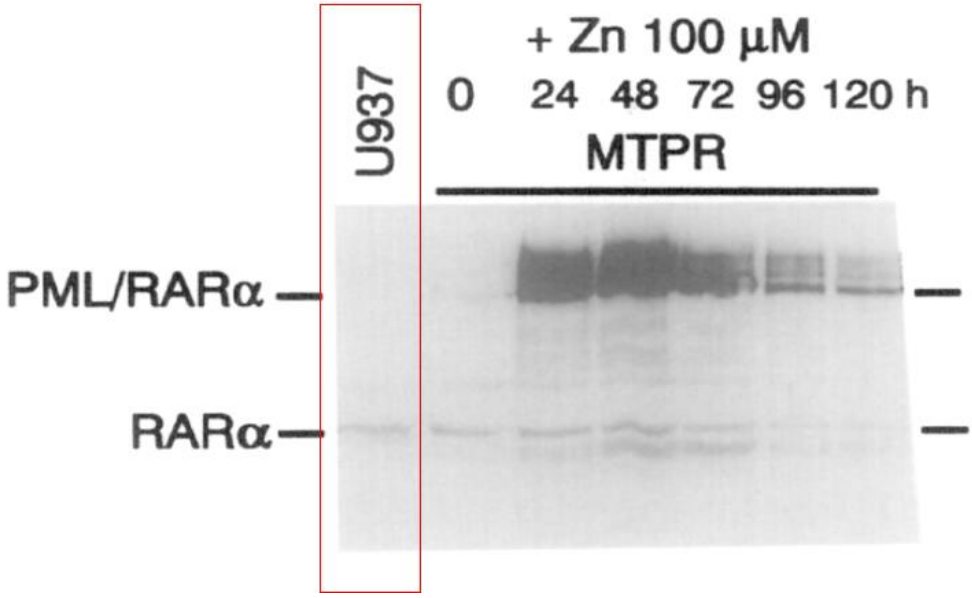

Figure 16 – Expression of PML-RARα in U937PR9 after induction by Zinc.

Expression of PML-RARα is promoted already after 24 hours of induction by Zinc. Expression of PML-RARα remains constant for almost 3 days, and begins degrading after 120 hours since adding Zinc to the cell colture.

Repli-seq

Asynchronously exponentially growing U937PR9 cells were collected and incubated for 45 minutes with 50 µM 5-bromo-2-deoxyuridine (BrdU) (i.e. BrdU pulse).

For the "Zinc-treated" condition, expression of PML-RARα was induced upon addition of 100 µM ZnSO4 to cell culture and cells were left growing for 24 hours, prior to the BrdU pulse.

Subsequently, and for each condition separately, cells were DAPI-labelled and sorted by FACS into six fractions spanning throughout the S phase (S1, S2, S3, S4, S5, S6. (Table 1)

	FACS Sorting Gates		Manually reassigned*	
S phase fractions	Start	End	Start	End
S1	2.09	2.29		
S2	2.3	2.6		2.61
S3	2.62	2.94		
S4	2.95	3.26		3.275
S5	3.29	3.59		3.68
S6	3.77	4.08	3.68	4

Table 1. FACS Sorting Gates according to cell DNA content quantified by DAPI
* gaps among gates were manually reassigned by equally splitting gap distance among adjacent gates

Following sorting, cells were lysed, and DNA extracted and fragmented. BrdU-labelled DNA fragments were immunoprecipitated using an anti-BrdU monoclonal antibody (Becton Dickinson Cat.no. 347580), assembled into separate Illumina libraries and sequenced.

BrdU-labelled DNA fragments were sequenced by Illumina Hiseq 2000 with 36bp single-end short reads. Successively, reads were aligned to the reference genome (Hg18) using the Barrows-Wheeler Aligner (BWA) (Li and Durbin, 2009), obtaining an average of ~20 million and ~14 million uniquely mapped reads in Replicate1 for the Zinc-treated and Untreated conditions, and ~16 million for both Zinc-treated and Untreated conditions in Replicate 2 (Table 2).

	Unique reads mapped to genome (Hg18)			
	Replicate1		Validation Replicate	
S phase fractions	**Zinc-treated**	**Untreated**	**Zinc-treated**	**Untreated**
S1	16,452,864	8,684,378	18,026,962	9,447,101
S2	24,519,006	18,116,761	13,227,563	15,895,253
S3	18,525,159	12,779,129	28,767,155	19,594,275
S4	19,603,429	19,607,396	12,081,052	17,884,233
S5	24,215,342	13,751,038	11,517,206	15,776,376
S6	18,666,357	12,034,952	13,284,630	19,060,496

Table 2. Uniquely mapped reads for Zinc-treated and Untreated conditions in Replicate1 and Validation Replicate.

RNA-sequencing and Differential Expression Analysis

RNA-sequencing

RNA-sequencing libraries were prepared by collecting U937PR9 cells (Untreated and Zinc-treated, separately). For Zinc-treated condition, cells were incubated with 100 µM ZnSO4 for 8 hours.

RNA was extracted using Qiagen RNeasy Mini Kit according to manufacturer's practice and libraries were prepared following protocols described in Illumina TruSeq® RNA Version 2 and RNA fragments were sequenced with Illumina Hiseq 2000 by 50bp paired-end short reads.

RNA-Seq data analysis

The sequenced samples were aligned on the reference genome Hg18 (*Homo sapiens* NCBI Assembly 36.1 (GCF_000001405.12)) using TopHat software (Trapnell et al., 2009).

Mapped sequences were processed with HTSeq software(Anders et al., 2015) with parameters: (-m) intersection strict, (-a) skip quality reads less than 1.

Differential gene expression analyses including size-factor normalization, shrinkage estimation for the distribution's variance and negative binomial distribution were performed using the R package edgeR (Robinson et al., 2010).

ChIP-sequencing

ChIP-sequencing libraries were prepared by collecting 1-5 ng of ChIP DNA from U937PR9 cells (Untreated and Zinc-treated, separately) and following instructions as per an *in-house* protocol. For Zinc-treated condition, cells were incubated with 100 µM ZnSO4 for 8 hours.

ChIP was performed for H3K4me3, H3K4me1 and H3K27Ac taking advantage of the following antibodies:

H3K4me3: anti-H3K4me3 Active Motif 39159 (Rabbit, polyclonal)

- H3K4me1: anti-H3K4me1 Abcam ab8895 (Rabbit, polyclonal)
- H3K27Ac: anti-H4K27Ac Abcam ab4729 (Rabbit, polyclonal)

Immunoprecipitated DNA fragments were sequenced with Illumina Hiseq 2000 by 51bp single-end short reads.

Sequenced reads were aligned on the reference genome hg18 with BWA software (version 0.6.2-r126) (Li and Durbin, 2009) using presets end-to-end parameters, and discarding reads with more than one mapping.

Duplicate reads were removed with Samtool rmdup version 0.1.19 (Li et al., 2009).

ChIP-Seq signals were identified using MACS v.1.4.10 (Zhang et al., 2008) with shifting

model and dynamic lambda disabled, using a tag size of 51 and the to-large option to scale the IP samples to the input sample.

Uniquely mapped reads for H3K4me3, H3K4me1 and H3K27Ac samples for both Zinc-treated and Untreated conditions and for input are reported in Table 3.

	Zinc-treated	**Untreated**
H3K4me3	14,410,659	16,615,404
H3K4me1	22,195,611	26,620,576
H3K27Ac	42,066,749	25,671,866
Input	22,168,088	

Table 3 – Uniquely mapped reads in ChIP-seq Histone marks data

Alignment of Repli-seq and ChIP-seq data, and alignment of RNA-seq data and DE analyses were performed by taking advantage of an in-house pipeline developed by Luciano Giacò.

DART Pipeline

DART was computed by taking advantage of three different programming languages: shell scripting, R (version 3.2.2) and Python (version 2.7.3).

Binning of reference genome

Hg18 was binned in 100bp intervals by command “makewindows” from the BEDTools suite (Quinlan and Hall, 2010), with parameters -g <reference genome> -w 100

Calculation of coverage

This multi-step process was performed by taking advantage of different tools.

At first, each S phase fraction signal was converted from bigwig to wig format by bigWigToWig (Kent et al., 2010) and, successively, signal underlying each 100bp genomic interval was retrieved by “intersectBed” and “groupby”, other two tools of the BEDTools suite.

Internal normalization

TMM normalization was performed by command “calcNormFactors” from edgeR and normalizing factors were applied to each S phase fraction, for each condition separately.

Inter-condition normalization

Quantile normalization was performed by “normalizeQuantiles” in Limma (Ritchie et al., 2015) R library.

Calculation of S50 values

Computation of S50 values was performed by a custom script kindly provided by Davide Cittaro, and described in (Dellino et al., 2013a).

RESULTS

Different studies have showed a strict association of Replication Timing with transcription (Huvet et al., 2007; Rivera-Mulia et al., 2015), chromatin structure (Donaldson, 2005; Lubelsky et al., 2014; Vogelauer et al., 2002), cell development and differentiation (Pope et al., 2010).

In 2010, Hansen and colleagues coupled pre-existing Replication Timing approaches with next-generation sequencing technology, devising a novel method known as Repli-seq (Hansen et al., 2010). Thanks to Repli-seq, it is possible to study the progression of the replication forks throughout the genome and to identify regions containing, or enriched for, ORIs.

In addition, Hansen and colleagues made a first attempt to compare DNA replication timing in different cell lines.

However, in the past years Repli-seq technology has been mainly used to characterize the Replication Timing *per se*, with no studies taking advantage of such method to measure and to assign statistically significant differences, at genome scale, in Replication Timing between different cell lines or experimental conditions.

Thus, this work aimed to: i) generate a bioinformatic and statistical procedure to identify regions showing differential replication timing between two conditions, and ii) apply this method to study possible changes in Replication Timing of promonocytic U937-PR9 cells before and after the expression of the PML-RARα oncogene.

DART (Differential Analysis of Replication Timing) pipeline

In order to evaluate the presence of differences in Replication Timing before and after PML-RARα expression, we developed a bioinformatic method to compare, pairwise, Repli-seq experiments.

Such approach, after normalization between BrdU signals of each S sub-phases, to account for technical and biological intra- and inter-condition variability, allows genome-wide calculation of Replication Timing for each condition and the identification of genomic regions showing a statistically significant differences in Replication Timing. The workflow of the pipeline is schematized in Figure 17 and will be briefly discussed below. It consists of the following main steps:

- Alignment and coverage calculation
- Internal Normalization: TMM
- Signal Transformation
- Removal of nonspecific signal
- Normalization between conditions: Quantile Normalization
- Calculation of Replication Timing
- Identification of Regions enriched for ORIs (Valleys)
- Identification of Valleys showing Different Replication Timing

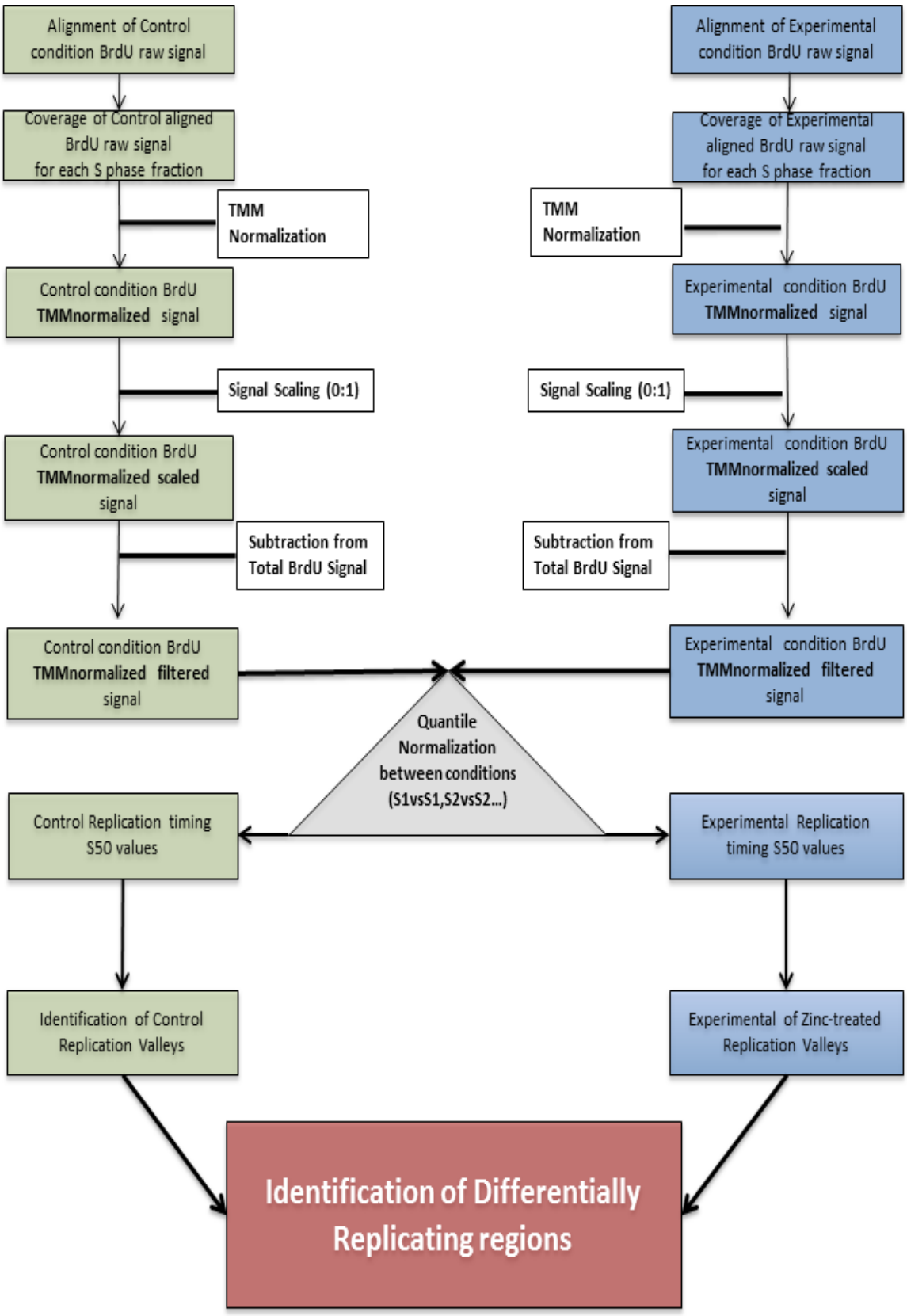

Figure 17. DART pipeline workflow

The diagram shows, step-wise, the operations' flow that lead to the identification of regions enriched for ORIs with a different Replication Timing between two experimental conditions.

The Repli-seq at glance

During Repli-seq procedure, asynchronously growing cells are pulse-labelled by Bromodeoxyuridine (BrdU), a thymidine analogue that is actively incorporated in nascent DNA strands, thus labelling the newly-replicated DNA.

Cells are then stained with Propidium Iodide (PI), a fluorescent molecule that binds to DNA by intercalating among the bases, and analysed by Fluorescence-activated cell sorting (FACS).

FACS measures the DNA content of each cell by means of their PI fluorescence signal and fractionates the analysed cells in "n" subpopulations with different DNA content (where "n" is the number of the sorting gates which have been chosen). Since total DNA content of each cell increases during the progression of the S-phase, each sorted subpopulation of cells corresponds to a defined time interval of the S-phase; in our experiments n=6 (S1 to S6).

Sorted cells are then lysed, the DNA is fragmented and immunoprecipitated with an anti-BrdU antibody and, finally, the immunoprecipitated DNA fragments undergo high-throughput sequencing.

In our cell model, a normal cell cycle lasts ~24 hours and one third of them (~8 hours) are used for DNA synthesis (S-phase). Since the S-phase was divided into six fractions of equal length, each S-phase fraction corresponds to slightly less than 1.5 hours.

Alignment:

For each experimental condition and for each S-phase fraction, the first step of the bioinformatic approach is the alignment, on the reference human genome (in this study we used the human Hg18 release), of BrdU-positive DNA sequences.

Visualization of the S1-to-S6 ordered raw alignment on the Genome Browser immediately

allows the identification of the expected progression of DNA replication, in space and time, as revealed by the Repli-seq assay (Figure 18a).

Progression of the replication forks is represented by a series of early-to-late transitions, called “Inverted-Vs”, which originate from the summit of each “Inverted-V”, known as “Inverted-V apex”, regions which are, probably, enriched ORIs.

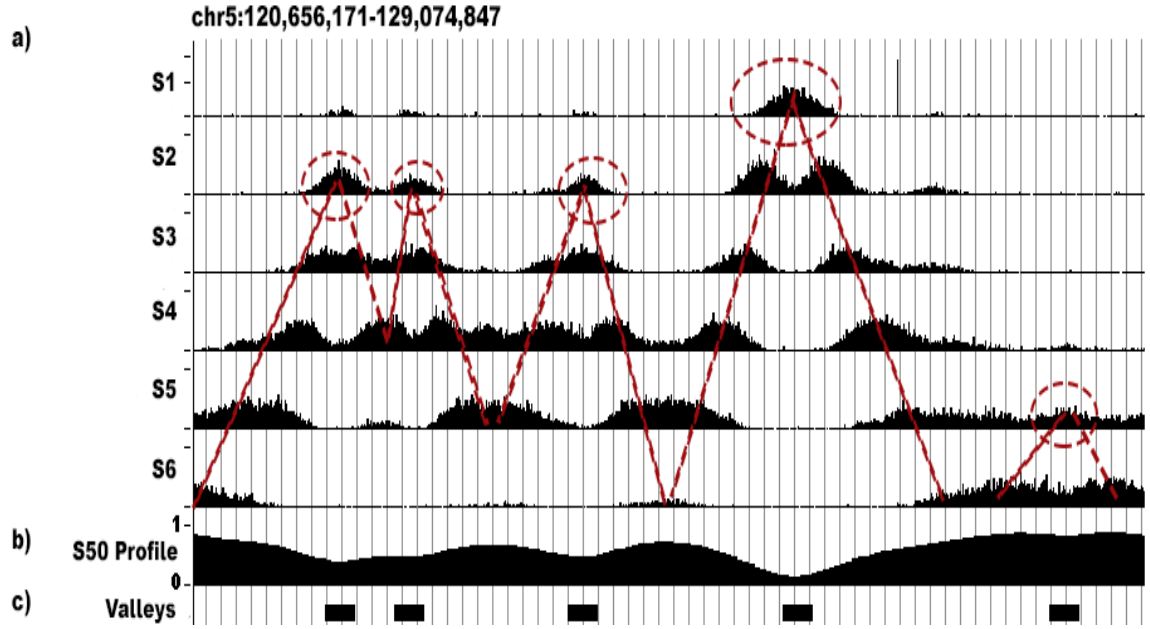

Figure 18 – From alignment of Repli-seq to Replication Valleys
a) Aligned Repli-seq S-phase fractions (S1-S6) (raw data), with the characteristic Inverted-Vs (red dotted line) and Inverted-V apexes (red dotted circles)
b) S50 profile of the region, with values ranging from 0 to 1
c) Replication Valleys.

Calculation of Coverage:

After sequencing and alignment, the reference genome was divided in 100 bp-long consecutive and non-overlapping bins; the underlying BrdU signal of each S-phase fraction was calculated for each bin and for each experimental condition.

Internal Normalization:

Given the nature of the Repli-seq assay, each S-phase fraction (S1-S6) derives from an independent sequencing run, but is strictly dependent upon the others since was obtained

from the whole S-phase of the same cell population.

Therefore, an intra-condition normalization step among the S-phase fractions allowed the elimination of possible biases introduced by different sequencing depth across samples. For this purpose, we decided to apply a TMM (Trimmed Mean of M-values) normalization to each experimental condition, separately.

The TMM is a normalization method firstly described by Robinson and colleagues (Robinson and Oshlack, 2010) and largely used for the normalization of RNA-seq libraries (Li et al., 2015). By this normalization, it is possible to apply a scaling factor to the library size of the samples being studied, thus allowing their direct comparison.

Signal transformation:

Since each region of the genome is replicated only once and only in a specific moment of the S-phase, the total signal of the six fractions, in principle, should be the same for each bin. As a further normalization step, we therefore transformed the normalized counts of each genomic bin in a value expressing the relative quantity of signal found in each specific S fraction, with respect to the total signal.

To this end, for each 100 bp interval, we divided the normalized counts of each Sphase fraction by the total Sphase signal (S1 to S6), thus converting the signal from each 100 bp interval in six values ranging from 0 to 1.

This allowed a precise measure of the contribution of each Sphase fraction to the completion of the DNA replication for each genomic interval of 100 bp.

Subtraction of Total BrdU signal:

To reduce the background signal in each fraction of the S-phase, we first created an *in-silico* pseudo-Total BrdU signal: for each 100 bp interval, we summed up the values from each S-phase fraction, and then calculated the mean value. This pseudo-Total BrdU signal was then subtracted from each S-phase fraction.

Normalization between conditions by Quantile Normalization:

TMM normalization might be sufficient to study the Replication Timing of a single experiment *per se*. However, in order to compare two Repli-seq experiments performed in control and perturbed experimental conditions, another run of normalization was performed, aimed to remove variability between the two conditions due to technicalities and not deriving from true biological differences.

A Quantile Normalization was applied in pairwise combinations between the same S-phase fraction deriving from the two experimental conditions (*e.g.*, $S1_{Exp}$ vs $S1_{Ctrl}$, $S2_{Exp}$ vs $S2_{Ctrl}$, etc.), before the identification of true biological variations between control cells (Control) and cells expressing the oncogene (Exp).

Calculation of Replication Timing (S50 values):

In order to transform the complex replication pattern deriving from BrdU signals from six independent S-phase fractions, in a single quantitative value which measures the Replication Timing, we used the S50 value, defined as the fraction of S-phase at which half of the total DNA from a genomic region is replicated**.**

As previously reported (Dellino et al., 2013), the reference genome was divided in 50 kb consecutive and non-overlapping intervals, obtaining ~65,000 genomic windows.

For each condition, and for each 50 kb window, an intermediate cumulative sum (S1,

S1+S2, S1+S2+S3, etc.) and a final sum M (M: S1+S2+S3+S4+S5+S6) were computed and a S50 value was calculated by a linear interpolation of the two cumulative sums closest to M/2.

Consequently, S50 values close to zero correspond to early-replicating regions, while S50 values close to "1" correspond to late-replicating regions (Figure 18b).

Identification of Regions enriched for ORIs:

We primarily focused on Differential Replication Timing at Inverted-V apexes, that show smaller S50 values than their flanking regions.

For Exp and Control cells separately, at each Inverted-V apex we methodologically defined an origin-containing region (OCR) as the local minimum value in the S50 profile (corresponding to a 50 kb region) extended by two windows (50 kb each) both upstream and downstream, for a total 250 kb-long genomic region (Figure 18c).

Since OCRs appear as depressions in the S50 Profile, as shown in Figure 19, where the S50 profile of the entire chromosome 10 is reported, they are also called "Replication Valleys" or, simply, "Valleys".

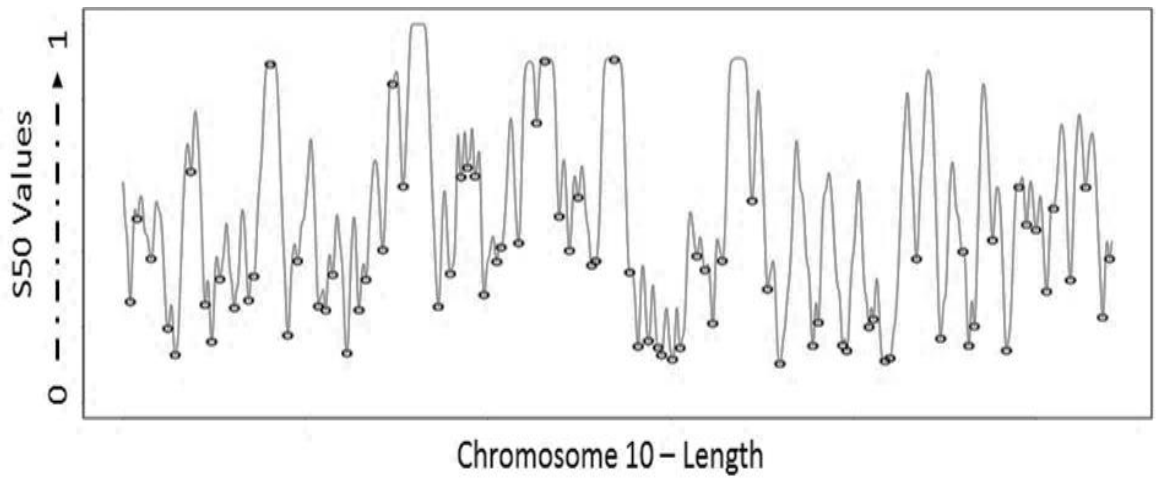

Figure 19 - Examples of S50 Profile and Valleys.
S50 profile (solid line) calculated for the whole chromosome 10 and Valleys (circles)

Identification of Valleys sharing the same genomic location in the two Samples:

In order to measure differences in replication timing between Exp and Control samples, we first focused on Valleys common to the two experimental conditions. At this purpose, we calculated the Jaccard index, according to the following formula:

$$JI = (A \cap B / A \cup B)$$

thus corresponding to the intersection between A and B, divided by the union of A and B, and selected Valleys showing Jaccard Index equal to or higher than 0.6 in a range from 0 to 1 (*i.e.*, overlapping for 4, out of 5, 50 kb-long intervals).

Calculation of difference in Replication Timing - the Diff_S50 values:

We then calculated, for each Valley, the difference between the S50 values of Exp and Control.

Such values, termed Diff_S50, were obtained by the formula:

$$Diff\_S50 = min(S50_{Exp}) - min(S50_{Ctrl})$$

A negative Diff_S50 value indicates a late-to-early shift (LtoE) of Replication Timing ($S50_{Exp}$ values smaller than $S50_{Ctrl}$), whereas a positive Diff_S50 value reveals an early-to-late shift (EtoL) of the Replication Timing ($S50_{Exp}$ values bigger than $S50_{Ctrl}$).

Identification of Valleys with a significant different Replication Timing:

EtoL- and LtoE-shifted Valleys were tested for the statistical significance of the observed change of Replication Timing (|Diff_S50 | > 0). At this purpose, we took advantage a paired t-test and, based on the results, Valleys were subsequently divided in three classes:

LtoE-shifted Valleys, showing in the Exp sample S50 values significantly smaller than in the Control sample (adjusted p-value ≤0.05); these Valleys showed a late-to-early shift in Replication Timing following PML-RAR expression;

EtoL-shifted Valleys, showing in the Exp sample S50 values significantly higher than the Control sample (adjusted p-value <=0.05); these Valleys showed an early-to-late shift in Replication Timing following PML-RAR expression.

Stable Valleys – Valleys without any significant difference in Replication Timing following PML-RAR expression.

DART Pipeline applied to Repli-seq experiments in U937-PR9 cells upon PML-RARα expression

We run the DART pipeline on our PML-RARα-expressing (Zinc-treated) and Control (Untreated) samples and found 1,717 and 1,707 Valleys for the Zinc-treated and Untreated conditions, respectively, with ~91% (n=1,560) overlapping Valleys (Figure 20a).

Overlapping Valleys with a Jaccard Index ≥ 0.6 were 1,379/1,560 (~88%), which represented about the 80% of both total Zinc-treated and Untreated Valleys (Figure 20b).

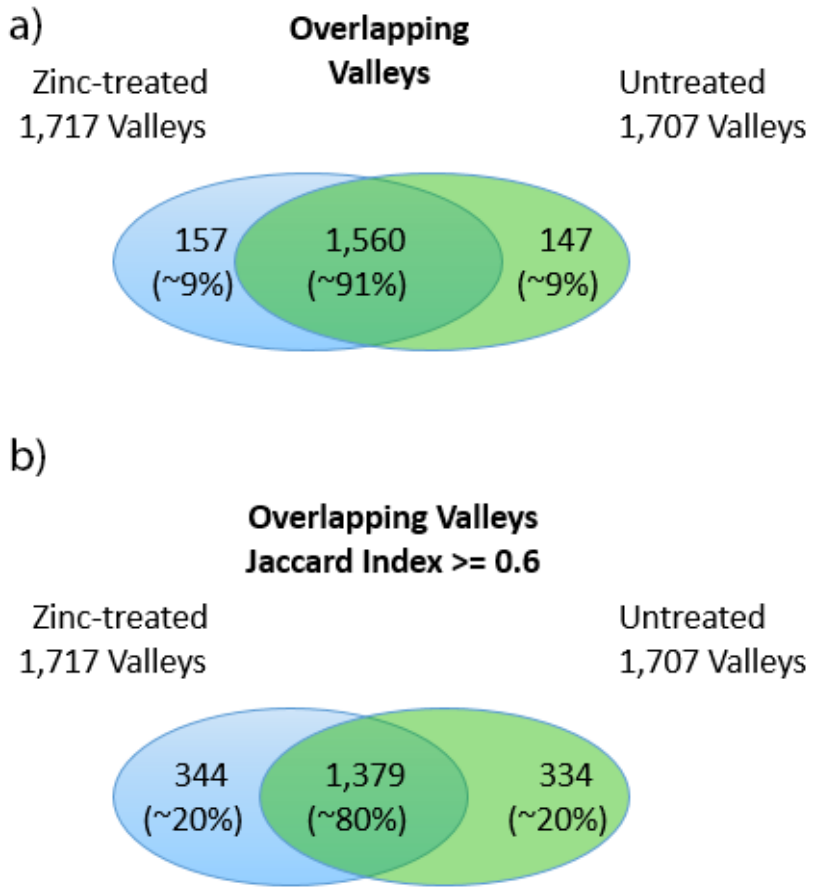

Figure 20. Overlapping Valleys in Zinc-treated and Untreated conditions
a) Venn Diagram showing the high degree of overlap (~91%) between Valleys found in Zinc-treated and Untreated conditions; **b)** Even if Venn Diagram showing the high degree of overlap (~91%) between Valleys found in Zinc-treated and Untreated conditions;

Students' t-test, performed to measure statistical significance, revealed the presence of:

- **536** LtoE-shifted Valleys
- **414** EtoL-shifted Valleys
- **429** Stable Valleys

Notably, while we observed a comparable number of Stable, LtoE- or EtoL-shifted Valleys, the distribution of their Diff_S50 values was clearly different (Figure 21), with LtoE-shifted Valleys showing a broader dispersion than the EtoL-shifted Valleys which, instead, showed Diff_S50 values generally closer to zero.

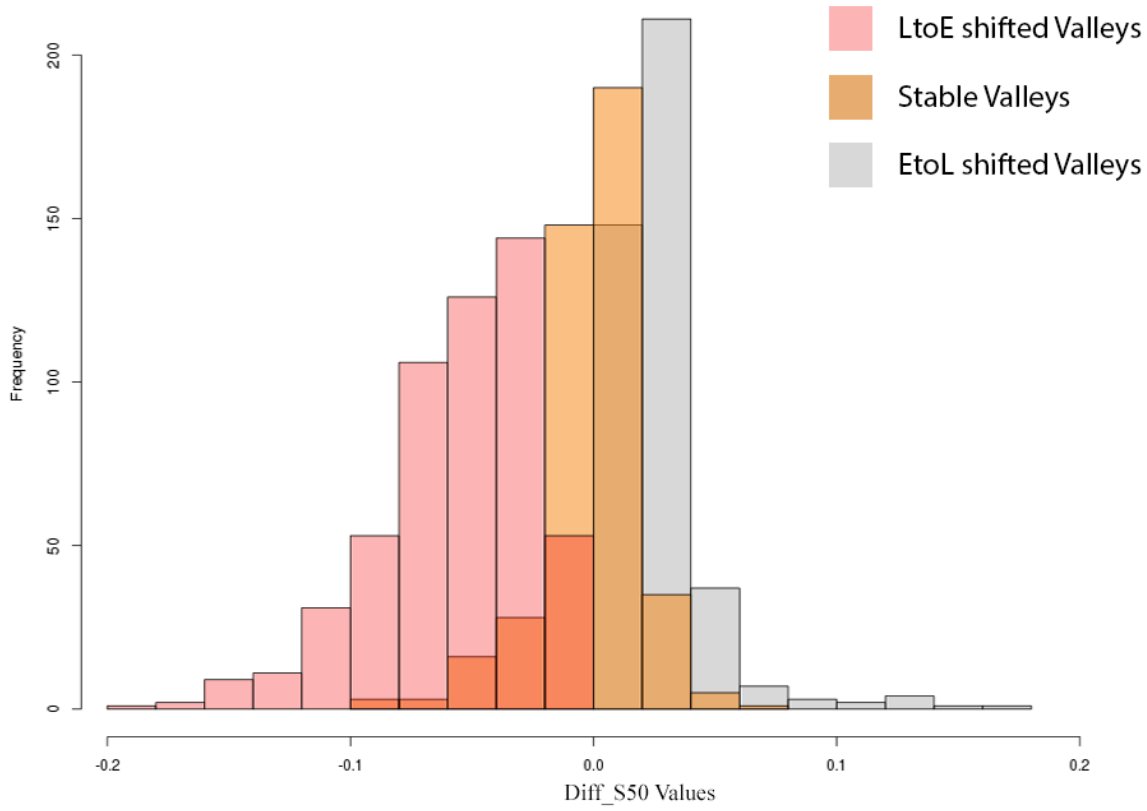

Figure 21 - Distribution of Diff_S50 values in LtoE shifted, Stable and EtoL shifted Valleys.
These overlaid histograms show the distribution of Diff_S50 values in LtoE shifted (red histogram), Stable (orange histogram) and EtoL shiftedValleys (grey histogram).

ROC Curve and assessment of performance

To assess the performance goodness of our DART pipeline, we performed a ROC curve analysis.

A ROC (Receiver Operating Characteristic) curve is usually employed to predict the robustness of diagnostic tests; on the typical ROC graph, each point represents a ratio between sensitivity and specificity (Zweig and Campbell, 1993) of the predictions.
The robustness of a ROC curve is measured by the Area Under the Curve (AUC), with values ranging between 0 and 1, representing the accuracy of the method being tested. Intervals of AUC values correspond to specific test evaluation, as follows:

- 0.0 - 0.6 = Failed
- 0.6 - 0.7 = Poor
- 0.7 - 0.8 = Fair
- 0.8 - 0.9 = Good
- 0.9 - 1.0 = Excellent

To perform the ROC analysis using our pipeline, we randomly took a set of replicating regions by visually inspecting and comparing raw profile signals of Zinc-treated and Untreated Repli-Seq BrdU samples in the Genome Browser. We identified 115 regions to use as control dataset, 57 of which showing clear differences in Replication Timing (positive controls – all LtoE-shifted) and 58 showing a stable Replication Timing upon treatment with Zinc (Figure 22).
Notably, while we easily recognized Valleys with a clear LtoE shift of the Replication Timing, none of the EtoL shifted valleys were obviously identifiable, either because they do not exist or, more likely, because the difference in Replication Timing was too small.

Indeed, as shown in Figure 21, EtoL shifted Valleys had Diff_S50s generally closer to zero, making therefore the identification by eye extremely difficult.

However, the analysis returned an AUC of ~0.95, thus indicating an excellent predictive power of the DART pipeline. The ROC curve is reported in Figure 23.

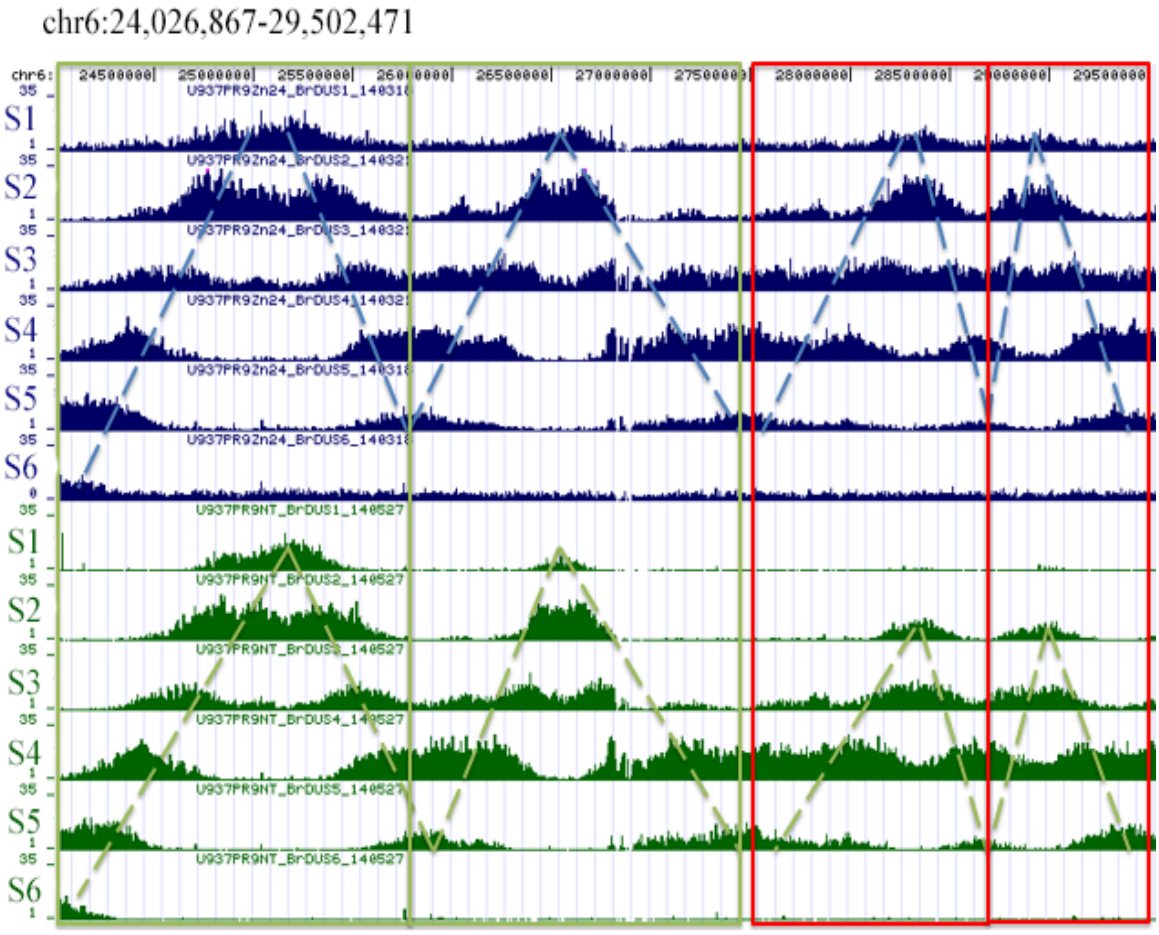

Figure 22 - Example of Control Regions used for the computation of the ROC Curve. In this picture it is shown an example of regions used as controls for the computation of the ROC Curve. Blue tracks: Zinc-treated Repli-seq S-phase fractions; Emerald tracks: Untreated Repli-seq S-phase fractions; Cyan dotted lines: Zinc-treated Repli-seq Inverted-Vs; Green dotted lines: Untreated Repli-seq Inverted-Vs; Green squares: Negative control regions retrieved as Inverted V-s starting replication at the same S-phase fraction between Zinc-treated and Untreated conditions; Red squares: Positive control regions retrieved as Inverted V-s starting replication at different S-phase fraction between Zinc-treated and Untreated conditions (LtoE shif of Replication Timing)

Diff_S50 values as a proxy for deregulation of Replication Timing

We took advantage of the ROC curve to estimate the Diff_S50 threshold that guaranteed the best trade-off between specificity and sensitivity for the call of stably- or differentially-replicated regions following PML-RARα expression.

The Diff_S50 cut-off value for the LtoE shifted regions was equal to -0.025, corresponding to specificity and sensitivity scores of 0.88 and 0.95, respectively.

Due to the lack of recognisable positive controls showing a EtoL of Replication Timing, we could not compute a Diff_S50 cut-off value for the EtoL shifted regions; therefore, we applied the same value (suitably changed in sign) as threshold for both LtoE- and EtoL-shifted Valleys, although we knew it would have been very stringent for the latter. We called this value Differential Replication Threshold, DRT.

Furthermore, due to its relevance for our following analyses, we also calculated the threshold for Valleys that did not show any significant change in their Replication Timing (Stable Replication Threshold, SRT).

The SRT, corresponding to a specificity value close to the sensitivity of the DRT (0.95), and a sensitivity value close to the specificity of the DRT (0.825), was equal to 0.0096 (Table 4, Figure 23).

	Threshold Diff_S50 (absolute value)	Specificity	Sensitivity
Differential Replication Threshold	0.025	0.88	0.95
Stable Replication Threshold	0.0096	0.95	0.825

Table4 – DRT and SRT with related Diff_S50 values, sensitivity and specificity of DRT and SRT

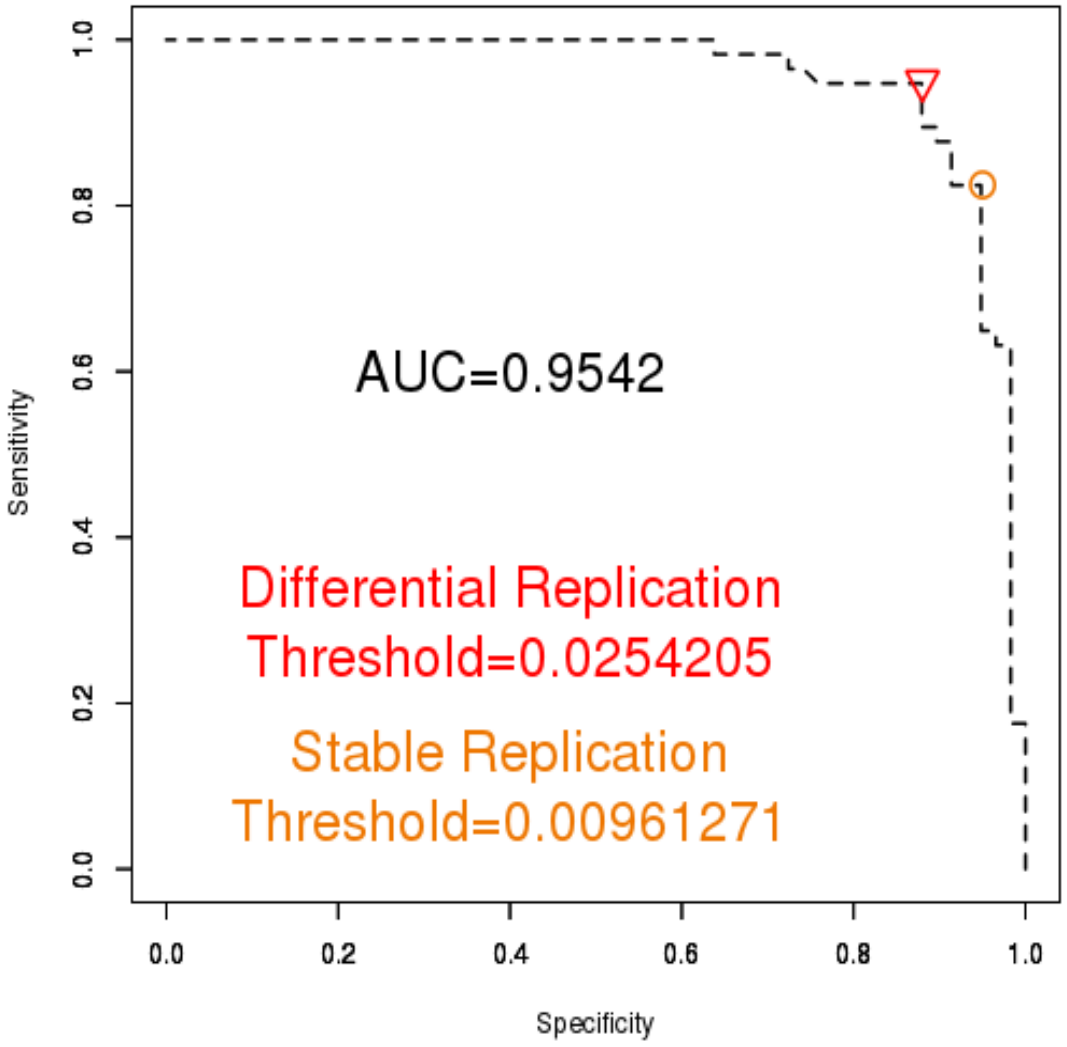

Figure 23. ROC Curve, DRT and SRT

ROC Curve assessing the robustness of the DART pipeline. Each point in the curve represent a Sensitivity/Specificity ratio.

Red triangle: Sensitivity/Specificity ratio chosen for DRT, equal to a Diff_S50 Value of 0.0252405

Orange Circle: Sensitivity/Specificity ratio chosen for SRT, equal to a Diff_S50 Value of 0.00961271

Differentially Replicating Valleys

Using the DRT and SRT thresholds, we refined the subsets of relevant Valleys, obtaining 1,074 Valleys classified as follows:

440 LtoE-shifted Valleys with: 1) a significant difference in S50 values between Zinc-treated and Untreated Samples and 2) a Diff_S50 value smaller or equal to the negative value of the DRT identified by the ROC curve;

183 EtoL-shifted Valleys with: 1) a significant difference in S50 values between Zinc-treated and Untreated Samples, and 2) a Diff_S50 value higher or equal to the positive value of the DRT identified by the ROC curve;

451 Stable Valleys without significant difference in S50 values between Zinc-treated and Untreated Samples, or with a Diff_S50 value higher than the positive, and smaller than the negative, absolute value of the SRT identified by the ROC curve.

In summary, our genome-wide analysis identified 1,717 and 1,707 Valleys for Zinc-treated and Control samples, respectively, covering about 14% of the human genome.
After the application of the DRT and SRT threshold, we obtained 1,074 Valleys, covering approximately 9% of the genome, of which: 3.8% (n=451 Valleys) showed a stable Replication Timing, 3.7% (n=440 Valleys) and 1.5% (n=183 Valleys) showed, respectively, a late-to-early and an early-to-late shift of Replication Timing following PML-RARα expression.

Validation of DART pipeline

To test the reproducibility of the results obtained in our experimental data, we applied the pipeline to a biological replicate of the Repli-seq experiment (from now on referred as "Replicate #2").

1,730 and 1,748 Valleys were identified in the Zinc-treated and Untreated sample, respectively (Table 5), with a high degree of overlap with the Valleys identified in the first experiment (Figure 24): ~84% (n=1,243/1,482) of the Valleys identified in the two Untreated replicates and ~83% (n=1,266/1,520) of those identified in the two Zinc-treated replicates overlapped for at least 80% of their length (i.e., four intervals of 50 kb out of five). We interpreted this result as the consequence of the reproducibility of the experimental procedure and of the robustness of the bioinformatic pipeline.

	Zinc-treated Valleys	Untreated Valleys
Replicate 1	1,717	1,707
Replicate 2	1,730	1,748

Table 5. Valleys identified in the Experiment1 and in the Replicate #2

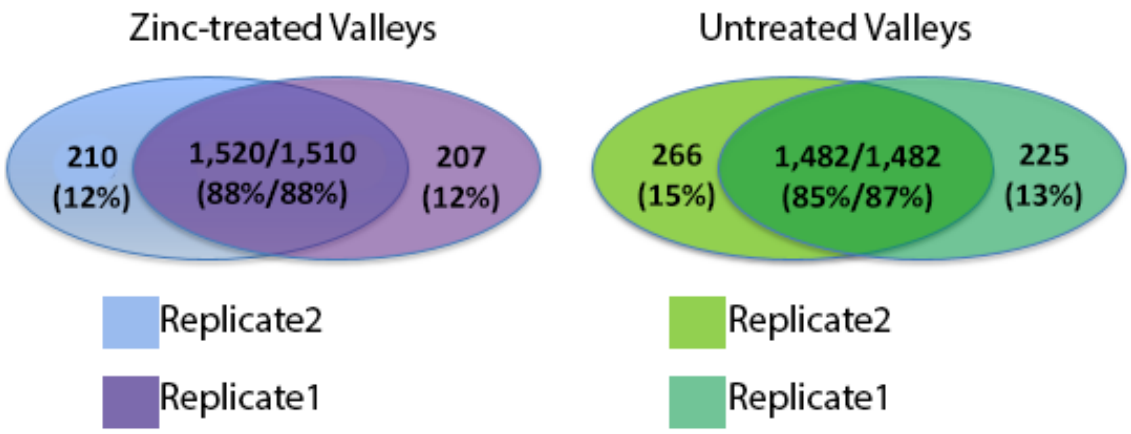

Figure 24 - Overlap between Valleys in Zinc-treated and Untreated Samples in Replicate1 and Replicate#2.
Both replicates show a reciprocal overlap of more than 80%, for both Zinc-treated and Untreated condition, as shown by the Venn Diagrams above.

Next, we computed a ROC curve using the same visually identified control regions used for the Replicate #1; 50/57 positive controls and 52/58 negative controls were found to overlap with the Valleys of Replicate #2.

The AUC of Replicate #2 was equal to ~0.8, lower than the original one, but still scoring a good predictive power.

DRT and SRT were obtained with a sensitivity of 0.77 and 0.73, respectively, and a specificity of 0.70 and 0.76, as summarized in Table 6.

	Replicate2			**Replicate1**		
	Threshold Diff_S50	**Specificity**	**Sensitivity**	**Threshold Diff_S50**	**Specificity**	**Sensitivity**
DRT	0.0205851	0.7	0.77	0.0254205	0.88	0.95
SRT	0.00855118	0.76	0.73	0.00961271	0.95	0.825

Table 6. Comparison of DRTs and SRTs and corresponding sensitivity and specificity values, in Replicate1 and Replicate2.

Thus, applying the DRT and SRT for **Replicate #2** we found:

- **417 LtoE-shifted Valleys**
- **153 EtoL-shifted Valleys**
- **673 Stable Valleys**

Comparison of Valleys identified in the two replicates, we observed different levels of overlap in the three different classes, ranging from 47-44% in the **LtoE-** to 20-17% in **EtoL-shifted** Valleys (Figure 25).

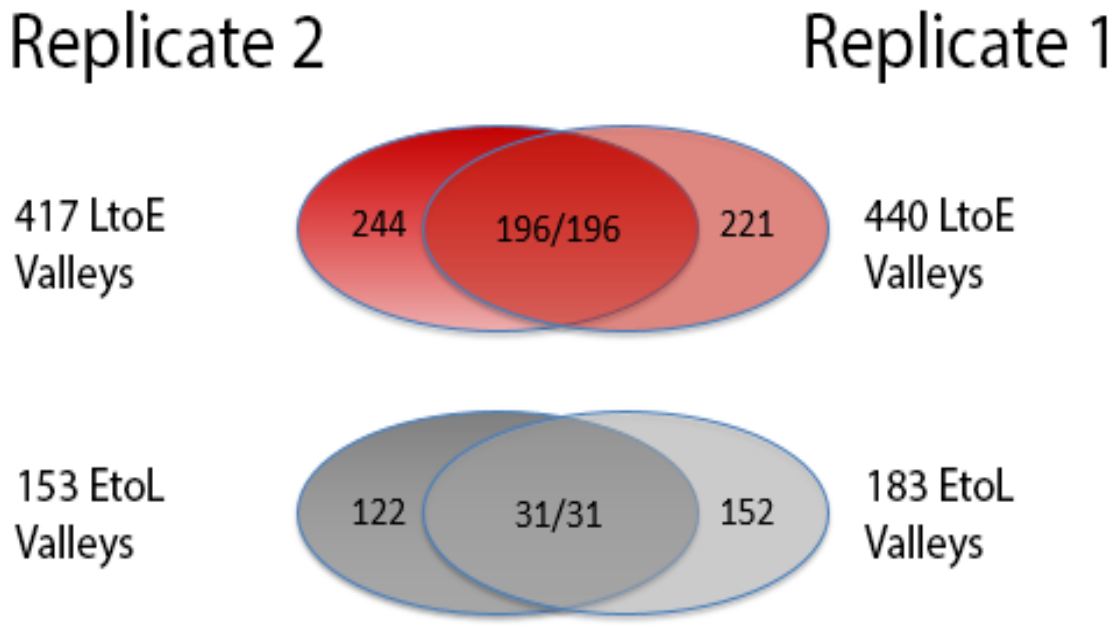

Figure 25. Overlap of DRVs and Stable Valleys between Replicate#1 and Replicate#2.
The Venn Diagrams show the mutual overlap among class of Valley in Replicate#1 and Replicate#2.
Red: LtoE-shifted Valleys; Grey: EtoL-shiftedValleys.

Given i) the complexity of the experimental procedure being studied and ii) the high stringency we applied to the analyses, we considered this overlap a good sign of

consistency and hence we decided to proceeded with the characterization of the differentially-replicated regions identified in Replicate #1.

Characterization of Differentially Replicating Valleys

Differences in Replication timing after PML-RARα expression are not evenly distributed throughout the genome but occur preferentially in regions replicated in specific timing of the S-phase.

As a first approach to the characterization of the Differentially Replicating Valleys, we studied whether differences in Replication Timing were uniformly distributed throughout the S-phase or were concentrated in some particular S-phase fractions.

At this purpose, we first transformed each of the six consecutive gates used during the cell sorting (and based on DNA content) into a specific range of S50 values (S1-S6; Table 7). The lowest S50 value among the five S50 values of the 50 kb windows that constitute each single Valley was then used to call each Valley as "Sn Valley" (where Sn is "S1" to "S6"), if its minimum S50 value fell within the range corresponding to that particular Sn. The distribution of replication timing (S50) of the three classes of Valleys (i.e., EtoL, LtoE or Stable) was finally studied in Zinc-treated and Untreated cells (Figure 26).

S50 interval	S sub-phase
0<= S50 <0.145	S1
0.145<= S50 <0.3050	S2
0.3050<= S50 <0.47	S3
0.47<= S50 <0.6375	S4
0.6375<= S50 <0.84	S5
0.84<= S50 <=1	S6

Table 7. Re-association of S50 interval to S sub-phases

As clearly shown in Figure 26, LtoE shifts of Replication Timing, globally, took part in Middle-S phase. Indeed, whereas in Untreated cells the higher number of Valleys was found in Middle-S phase (S3-S4), after PML-RARα induction we observed a general shift of Replication Timing toward earlier S-phase fractions, with a peak of replicating Valleys in Early/Middle-S phase (S2-S3).

Conversely, EtoL shifts of Replication Timing mostly involved regions replicated very early in the S-phase, with Valleys shifting from S1 in Untreated cells to S2 following PML-RARα expression. In addition, while LtoE shifts arose in Middle S-phase and persisted until late-S phase, EtoL occurrences were mostly confined in S1 and S2 sub-phases.

We then concentrated on LtoE- and EtoL-shifted Valleys, and observed that changes in replication timing did not necessarily resulted in a net shift from one S-phase fraction to another. To appreciate to what extent LtoE- or EtoL shifts (both within the same, or between different S-phase fractions) were distributed throughout the S-phase, we

constructed a square matrix in which it is shown the S-phase fraction in which each Valley showing a significant change of replication timing, was replicated before and after PML-RARα expression. A separate matrix was built for LtoE and EtoL shifts.

As shown in Figure 27, LtoE and EtoL shifts took place in Early/Middle-S phase and Early-S phase, respectively; however, most of changes in Replication Timing occurred within the same S-phase fraction (~62% and ~49% for LtoE- and EtoL-shifted Valleys, respectively).

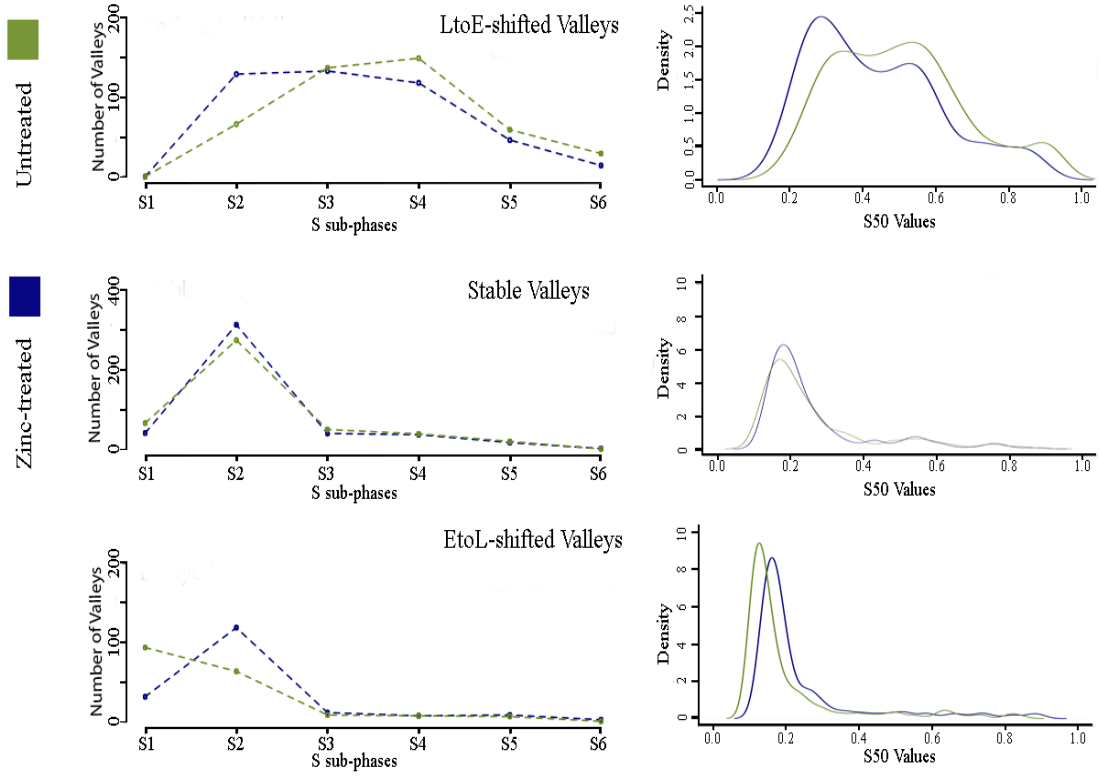

Figure 26 – Shift of Replication Timing in LtoE shifted and EtoL shiftedValleys.

DRVs replicating in different S-phase fractions after induction by Zinc; Right panel: S50 values in DRVs before and after expression of PML-RARα

Figure 27 - Differences in Replication Timing in LtoE shifted (green) and EtoL shifted(red) Valleys

Two heatmaps showing the percentage of those LtoE-shifted Valleys (left) and EtoL-shiftedValleys (right) replicating in different S-phase fractions and those replicating in the same S-phase fraction.
Heatmap was obtained by calculating, for each Valley, the S-phase in which they replicated before and after expression of PML-RARα and successively plotting the percentage of Valleys being replicated in each S sub-phase before and after induction by Zinc.

Overall, these data suggest that the Valleys showing changes in Replication Timing following PML-RARα expression, are not uniformly replicated during the S-phase: LtoE-shifted Valleys are normally replicated mainly during Early/Middle or Middle-S phase, while the EtoL-shifted ones are normally replicated almost exclusively during very Early-S phase.

EtoL shifted Valleys are enriched for PML-RARα binding sites

We then asked whether PML-RARα binds to the DRVs, using ChIP-seq experiments performed in the same U937PR9 cellular system.

Anti-PML ChIP-seq experiments in Zinc-treated (t=8h) and Untreated U937PR9 cells were analysed as explained in "Materials and Methods". Anti-PML antibodies did not recognize any enriched site or "peak" in Control cells, arguing against a direct DNA-binding activity of the endogenous PML protein. In the Zinc-treated samples, instead, we found ~23,600 peaks corresponding to *bona fide* PML-RARα binding sites (Martens et al., 2010; Wang et al., 2010).

Analysis of the distribution of PML-RARα peaks within the DRVs showed that EtoL-shifted Valleys were significantly enriched for PML-RARα binding sites when compared to Stable Valleys (about 9.0 and 5.6 binding sites on average per Valley, respectively); conversely, LtoE-shifted Valleys were significantly depleted (about 2.5 binding sites per Valley) (Figure 28, Table 8). Statistical significance was tested by a Mann-Whitney test.

N. PML-RARα peaks in Valleys	
1,091	**LtoE shifted Valleys (440)**
2,548	**Stable Valleys (451)**
1,654	**EtoL shifted Valleys (183)**

Table 8. Number of PML-RARα peaks in the three categories of Valleys

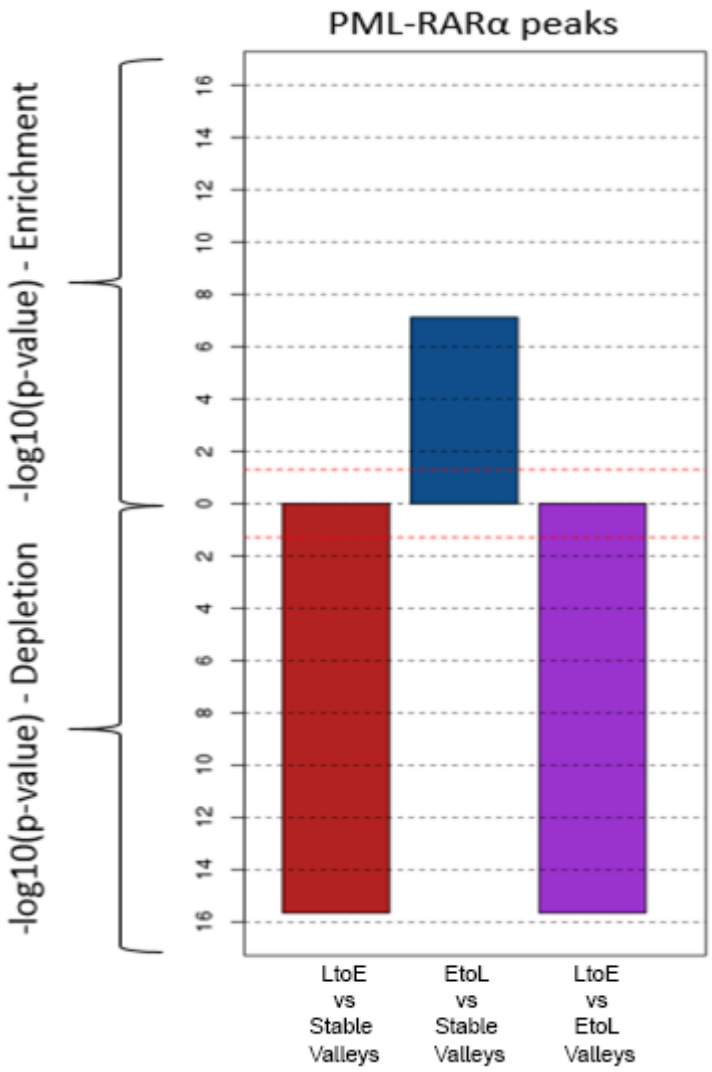

Figure 28 - Statistical significance for enrichment of PML-RARα binding sites in Valleys.

Results of Mann-Whitney statistical significance for depletion or enrichment of PML-RARα binding sites are represented by barplots.
Significance is expressed as -log10(p-value). From left to right: comparison of PML-RARα peaks in Valleys between LtoE shifted and EtoL shiftedValleys, between EtoL shifted and Stable Valleys and, finally, between LtoE shifted and EtoL shifted Valleys.

The distribution of gene expression level in regions with Different Replication Timing mirrors the distribution of Replication Timing and their changes through the S-phase

Several studies (Dellino et al., 2013; Fraser, 2013; Huvet et al., 2007; Méchali, 2010; Miotto et al., 2016; Rivera-Mulia et al., 2015) revealed a strict correlation between Replication Timing and transcription, showing that intense transcriptional activity positively correlates with early replication timing.

We thus investigated the relationship between basal transcription levels, or transcription regulation imposed by PML-RARα few hours after its expression (at t=8 hours), and the observed changes of replication timing observed 16 hours later (at t=24 hours).
For this purpose, we took advantage of RNA-seq experiments carried out in control U937PR9 cells and 8 hours after PML-RARα induction, and performed Differential Expression (DE) analyses according to standard procedures and computed by a custom pipeline (see Materials and Methods for details).

Henceforth, we mapped expressed genes within the DRVs and calculated their average expression (measured as Transcripts per Million, TPM) in each Valley; finally, for each S-phase fraction in the Control condition, we studied, for each class of Valleys, the distribution of the expression levels before and after the expression of PML-RARα (Figure 29).

Thus, we found that the distributions of gene expression of the three classes of Valleys was significantly different, with Stable Valleys containing highly expressed genes in early-S, EtoL-shifted Valleys in very early-S and LtoE-shifted Valleys in middle-S phase.

Furthermore, in LtoE-shifted Valleys we found a significant increase of TPM distribution in S3 (Figure 29, boxplots), consistent with the shift in Replication Timing in

Differentially Replicating Valleys, (which, in Control condition took place in Valleys replicating in S3-S4 S sub-phases), as previously depicted in Figure 26.

On the other hand, in EtoL shifted Valleys we observed a significant decrease in TPM distributions in Early-S fractions (S1-S2), consistent with the EtoL shift of Replication Timing found in the same S-phase fractions (for comparison, Figure 26).

At the same time, no significant variations in TPM distributions were found in Stable Valleys, thus confirming that a different level of expression was coherent with a different Replication Timing in LtoE- and EtoL-shifted Valleys.

Together, these results suggest that transcription levels measured at early time points (t=8 hours after PML-RARα induction) might contribute to the changes of replication timing observed 16 hours later (at t=24h).

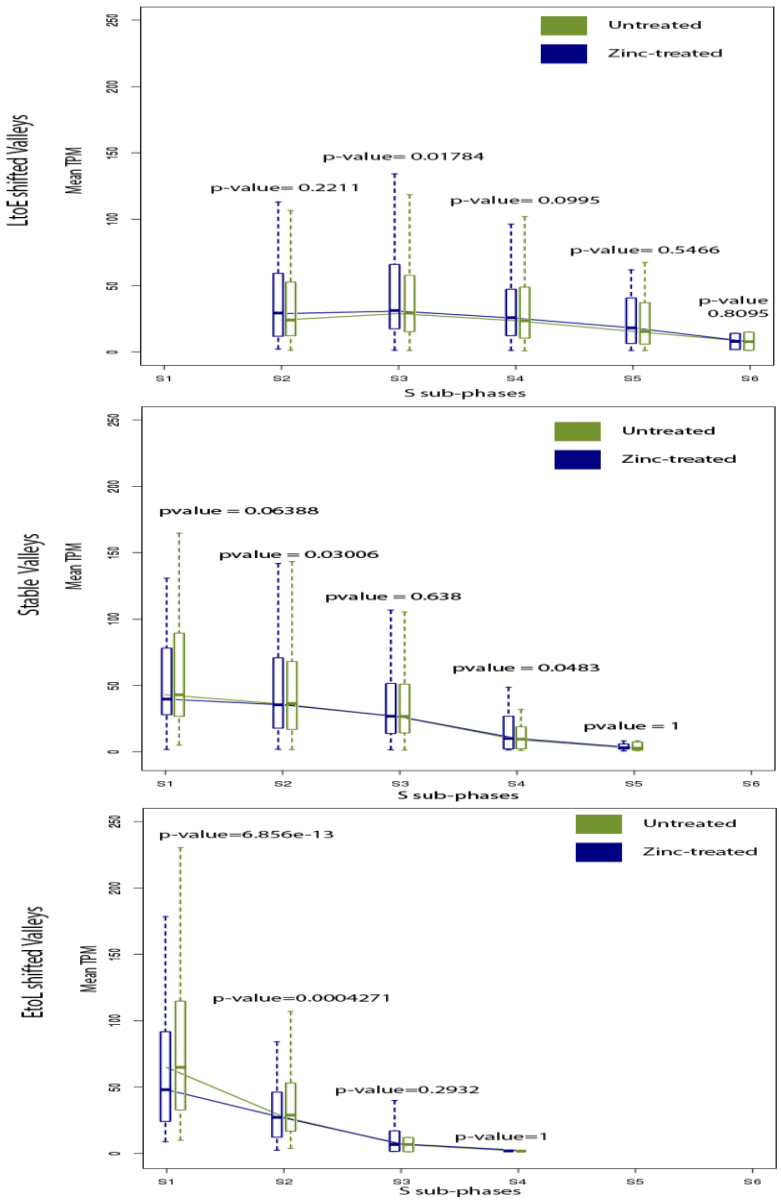

Figure 29 - DRVs are characterized by different levels of expression throughout the S-phase and differences in expression levels after induction of PML-RARα mirror changes in Replication timing.
EtoL-, LtoE-shifted and Stable Valleys differ for expression levels in the various S-phase fractions. Transcriptional levels in Untreated Cells (green lines) change after induction of PML-RARα (blue lines). Differences in expression (boxplots) were statistically tested by the mean of a Mann-Whitney test.

Differentially Replicated Valleys are enriched for direct PML-RARα target genes

Combining the identified PML-RAR binding sites and gene expression at Stable, EtoL- or LtoE-shifted Valleys, we then investigated the relationship between changes of replication timing and transcriptional regulation following PML-RARα induction.

As previously observed (Alcalay et al., 2003; Gaillard et al., 2015a), our RNA-seq experiments showed that PML-RARα induces both gene down-regulation and up-regulation (n=2,503 and 1,962 genes, respectively). A concise summary of RNA-seq experiment results is shown in Table 9.

Expressed and regulated genes before and after expression of PML-RARα				
Expressed Untreated	**Expressed after Zinc treatment (8h)**	**Total deregulated**	**Up-regulated**	**Down-regulated**
10,671	**10,805**	**4,465**	**1,962**	**2,503**

Table 9 - Expressed and regulated genes before and after induction by Zinc
Genes were considered expressed if TPM>=0.8 and regulated if adjusted p-value was equal to or higher than 0.2 and either the Untreated or the Zinc-treated sample had an expression equal to or higher to 0.8 TPM

In order to investigate whether, following PML-RARα induction, DRVs were enriched for regulated genes compared to Stable Valleys, we counted the fraction of up- and down-regulated genes and compared them to the total number of expressed genes falling into Differentially Replicating or Stable Valleys.

We found that the LtoE-shifted Valleys showed a two-fold enrichment of up-regulated genes compared to Stable Valleys: ~26% (n=169/644 genes) in LtoE-shifted versus ~14% (n=150/1,031 genes) in Stable Valleys, and a two-fold depletion of down-regulated genes: ~8% (n=50/644) in LtoE-shifted versus ~14% (n=146/1,031) in Stable Valleys (Table 10).

The opposite behaviour was observed for the EtoL-shifted Valleys, which were significantly enriched for down-regulated genes: ~40% (n=237/600) in EtoL-shifted *vs.* ~14% (n=146/1,031) in Stable Valleys, and depleted for up-regulated genes: ~4% (n=23/600) in EtoL-shifted *vs.* ~14%, (n=150/1,031) in Stable Valleys. (Table 10 and Figure 30).

Interestingly, also the proportion of up-regulated genes *vs.* down-regulated genes in the different classes of Valleys showing altered replication timing, follows the same behaviour: up-regulated genes in LtoE-shifted Valleys were 77%, and down-regulated genes in EtoL shifted Valleys were 91% of the total deregulated genes (Table 10).

Valleys	Deregulated genes	Number	% of DeReg genes on Total DeReg genes	Number of Total expressed genes	% of DeReg genes on Total expressed genes
LtoE_shifted	Up-regulated	**169**	77.2	**644**	26.2
	Down-regulated	**50**	22.8		7.8
	Total	**219**			34.0
Stable	Up-regulated	**150**	50.7	**1,031**	14.5
	Down-regulated	**146**	49.3		14.2
	Total	**296**			28.7
EtoL_shifted	Up-regulated	**23**	8.8	**600**	3.8
	Down-regulated	**237**	91.2		39.5
	Total	**260**			43.3

Table 10 - Number of Up- and Down- regulated genes on the total number of expressed and total deregulated genes in DRVs

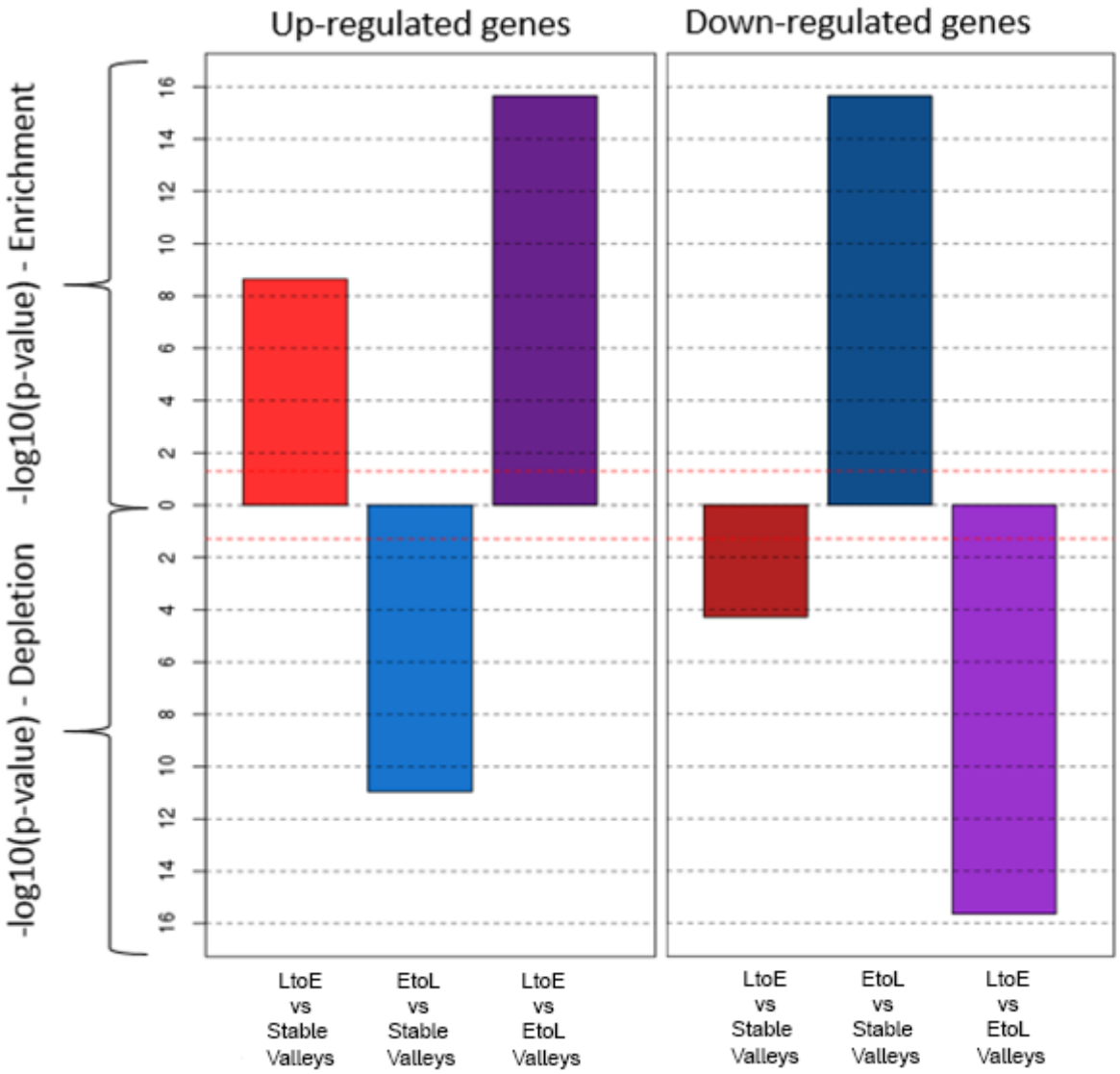

Figure 30 - Enrichment and depletion study for Up- and Down-regulated genes in DRVs
Results of Chi-square statistical significance for depletion or enrichment of deregulated genes in Valleys are represented as barplots. Significance is expressed as -log10(p-value). From left to right: comparison of up- (left panel) and down-regulated genes (right panel) in Valleys between "LtoE shifted versus Stable Valleys", "EtoL shifted versus Stable Valleys" and, finally, between "LtoE shifted versus EtoL shifted Valleys".

Next, we asked: 1) how many PML-RARα peaks were found on deregulated genes within Valleys, 2) how many of the deregulated genes were direct targets of PML-RARα (defined as genes showing PML-RARα bound either to their promoter or to gene body), and 3) how many of the direct target genes showed PML-RARα peaks at their promoter.

In total, we found 1,654, 2,548, and 1,091 PML-RARα peaks in EtoL-, LtoE-shifted and Stable Valleys, respectively. Interestingly, the vast majority of the 302 PML-RARα peaks

mapping at deregulated genes in LtoE-shifted Valleys bound to up-regulated genes: ~82% (n=247/302) *vs.* ~18% (n=55/302) at up- and down-regulated genes, respectively.

On the other hand, EtoL-shifted Valleys were enriched for PML-RARα peaks at down-regulated genes: ~95% (n=452/478) of the 478 peaks which were annotated to deregulated genes in EtoL-shifted Valleys, were bound to down-regulated genes and only ~5% (n=26/478) to up-regulated ones (Table 11, Figure 31).

In Stable Valleys, instead, PML-RARα showed no preferential binding, with ~46% (n=260/569) and ~54% (n=309/569) peaks annotated on, respectively, up- and down-regulated genes.

Such differences were evaluated by a Chi-square test, and were statistically significant when compared to Stable Valleys, or by comparing LtoE- versus EtoL-shifted Valleys (Figure 32).

Valleys	PML-RAR peaks	number	% of PML-RARa peaks on DeReg genes	Number of Total PML-RAR peaks	% of PML-RARa peaks on Total
LtoE_shifted	on Up-regulated genes	**247**	81.8	**1,091**	22.6
	on Down-regulated genes	**55**	18.2		5.0
	Total	**302**			27.7
Stable	on Up-regulated genes	**260**	45.7	**2,458**	10.6
	on Down-regulated genes	**309**	54.3		12.6
	Total	**569**			23.1
EtoL_shifted	on Up-regulated genes	**26**	5.4	**1,654**	1.6
	on Down-regulated genes	**452**	94.6		27.3
	Total	**478**			28.9

Table 11 - The table summarized the proportions of PML-RARα peaks annotated on Up- and Down-regulated genes on the total number of PML-RARα peaks observed in the three classes of Replication Valleys.

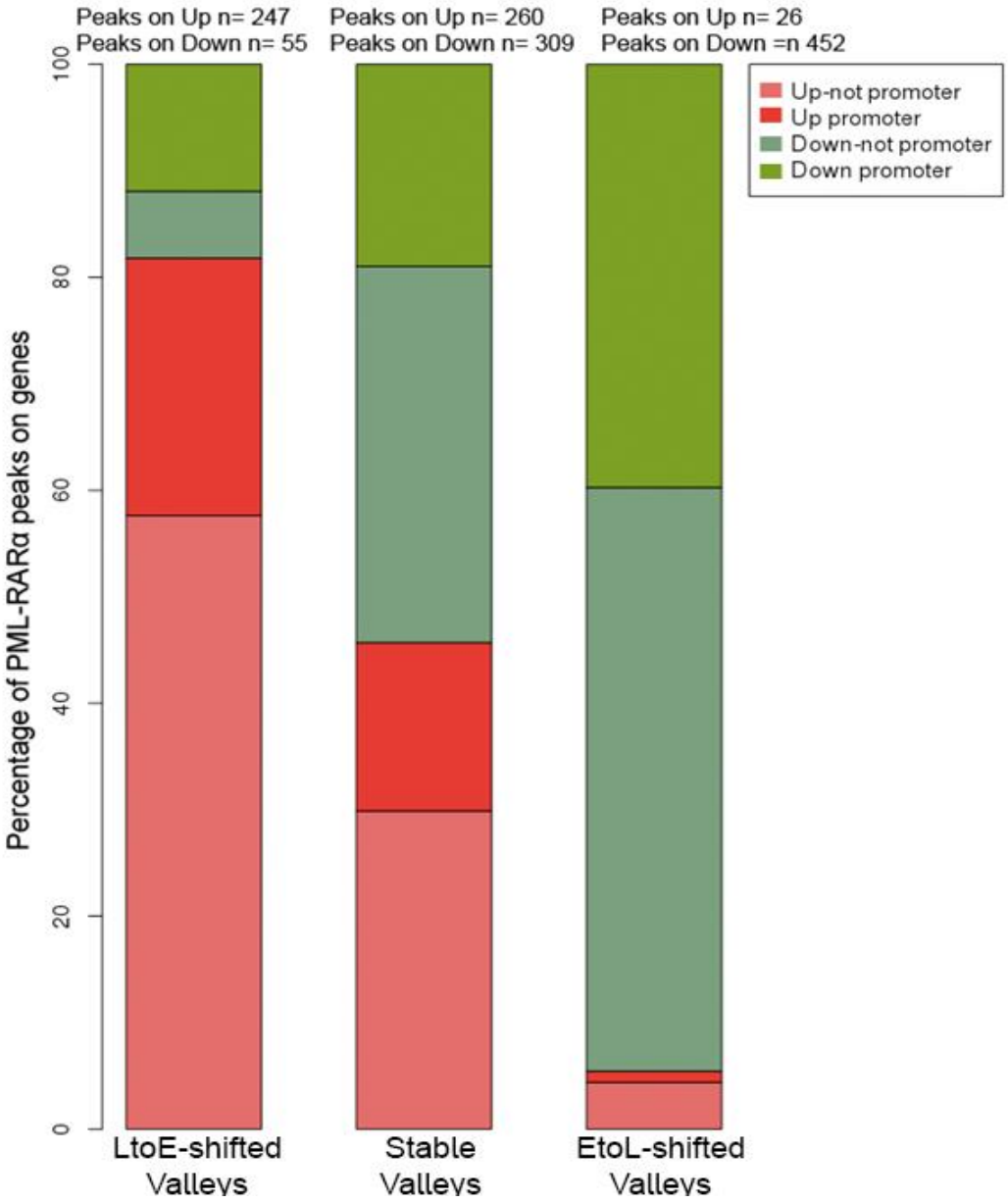

Figure 31 - Locations of PML-RARα peaks on deregulated genes in Stable and DRVs. The barplot illustrates the percentages of PML-RARα peaks found in UP- and DOWN-regulated genes by class of Stable and Differently Replicated Valleys; sub-groups of peaks that map or not to gene promoters are also shown.

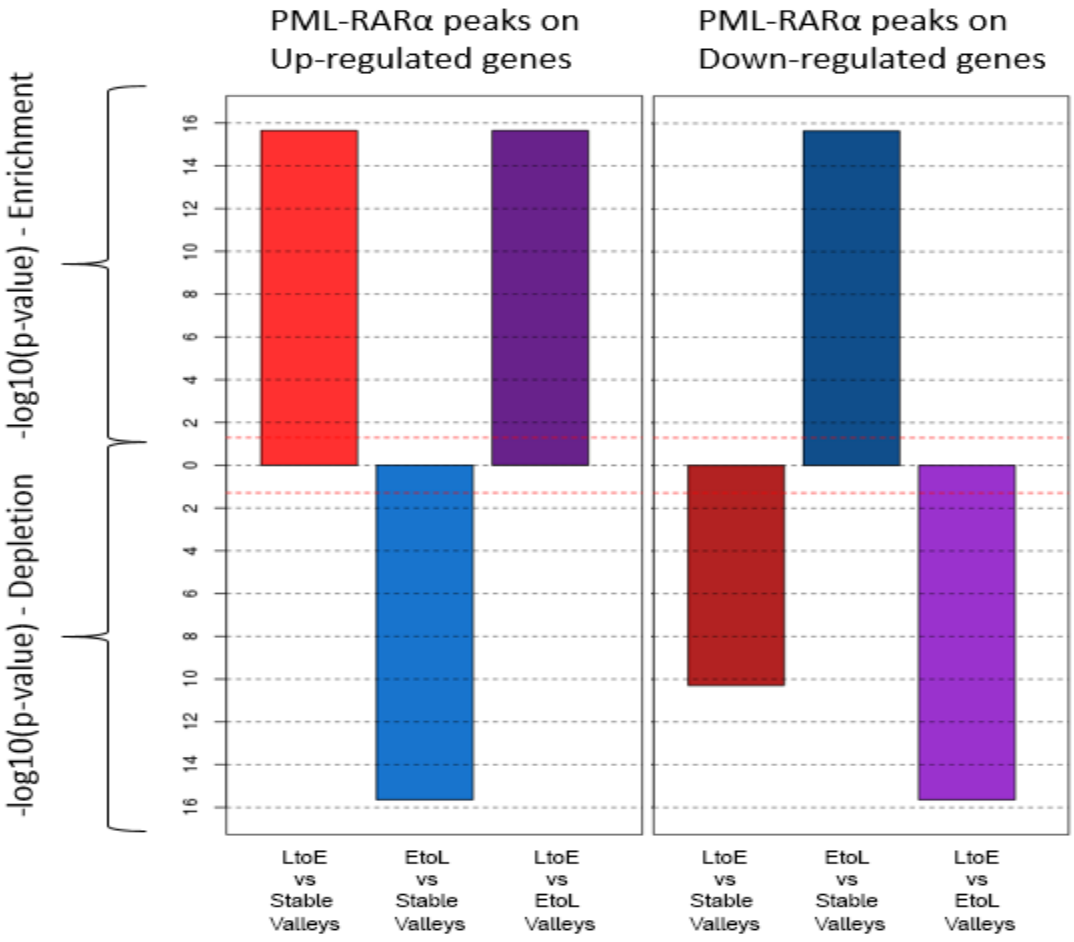

Figure 32 – Enrichment and depletion for PML-RARα peaks mapped on deregulated genes in DRVs
Results of chi-square statistical significance for depletion or enrichment of peaks on deregulated genes in Valleys are represented as barplots. Significance is expressed as -log10(p-value). From left to right: comparison of PML-RARα binding sites on up- (left panel) and down-regulated genes (right panel).
Comparisons were performed as "LtoE shifted versus Stable Valleys", "EtoL shifted versus Stable Valleys" and, finally, "LtoE shifted versus EtoL shifted Valleys".

Moreover, we found that, both in LtoE- and EtoL-shifted Valleys, PML-RARα preferentially bound to gene promoters (± 1 kb from TSS) of down-regulated genes compared to up-regulated genes: ~65% vs 30% and ~42% vs ~19% in LtoE- and EtoL-shifted Valleys, respectively, as shown in Table 12 and Figure 31.

Valleys	PML-RAR peaks	number	% of PML-RARa peaks on DeReg genes	Number of Total PML-RAR peaks	% of PML-RARa peaks on Total
LtoE_shifted	on Up-regulated genes	**247**	81.8	**1,091**	22.6
	on Down-regulated genes	**55**	18.2		5.0
	Total	**302**			27.7
Stable	on Up-regulated genes	**260**	45.7	**2,458**	10.6
	on Down-regulated genes	**309**	54.3		12.6
	Total	**569**			23.1
EtoL_shifted	on Up-regulated genes	**26**	5.4	**1,654**	1.6
	on Down-regulated genes	**452**	94.6		27.3
	Total	**478**			28.9

Table 12. Proportion of PML-RARα peaks on promoter of deregulated genes in DRVs.

Considering the deregulated genes falling in EtoL-shifted Valleys, we found that the most of the down-regulated genes (n=211/237, ~89%) were bound by PML-RARα; in an opposite fashion, the proportion of the up-regulated that were also bound by PML-RARα was smaller (n=11/23, ~48%).

On the other hand, in LtoE-shifted Valleys, up-regulated (n=102/169, ~60%) and down-regulated genes (n=36/50, ~72%) were directly bound by PML-RARα in comparable proportions (Table 13).

Though, when studying the location of PML-RARα binding sites within the genes, we found that, considering the proportion of deregulated genes with at least one peak of PML-

RARα on promoter compared the total deregulated genes, we observed that down-regulated genes were mostly bound on promoters of down-regulated genes (LtoE: n=33/50, ~66%; EtoL: n=176/237, ~74%) than the up-regulated ones (LtoE: n=63/169, ~37%; EtoL: n=5/23, ~22%), in both DRVs;

On the other hand, if we evaluated the number of genes with a PML-RARα binding site on promoter out of the number of genes with at least one peak of PML-RARα, wherever located in the locus, we noticed that nearly all of the down-regulated genes are bound on promoters, in both LtoE- shifted- (n=33/36, ~92%) and EtoL-shifted Valleys (n=176/211, ~83%). In addition, a lower fraction of up-regulated genes, whereas definitively not negligible, presented a PML-RARα binding site on their promoters (n=63/102, ~62% and n=5/11, ~46%) in both LtoE- and EtoL-shifted Valleys, respectively (Table 13).

Valleys	Deregulated genes	Tot Number	Number of DeReg genes bound by PML-RARa	% DeReg genes bound by PML-RARa / Total DeReg genes	Number of DeReg genes bound by PML-RARa on promoter	% DeReg genes bound on promoter by PML-RARa / bound by PML-RARa	% DeReg genes bound on promoter by PML-RARa / Total DeReg genes
LtoE shifted	Up-regulated	**169**	**102**	60.4	**63**	61.8	37.3
	Down-regulated	**50**	**36**	72.0	**33**	91.7	66.0
EtoL shifted	Up-regulated	**23**	**11**	47.8	**5**	45.5	21.7
	Down-regulated	**237**	**211**	89.0	**176**	83.4	74.3

Table 13. Up-Regulated and Down-regulated PML-RARα direct targets.

Overall these results suggest that the two opposite phenotypes of Replication Timing observed after the ectopic expression of PML-RARα (i.e. EtoL and LtoE shifts of the Replication Timing) might have different mechanistic links.

In EtoL-shifted regions PML-RARα exerts its function by mainly down-regulating its target genes, binding nearly always at their promoters. In addition, in these Valleys it is also possible to appreciate some promoter-bound up-regulated genes that, though, being in

much lower number than the down-regulated ones, are probably not able to give a contribution to the deregulation of such genomic regions.

Differently, in regions showing a LtoE shift of Replication Timing, PML-RARα seems to act in a dual manner: i) eliciting the down-regulation of few targets, again directly binding onto their promoters (such as for the down-regulated genes in EtoL-shifted regions) and ii) increasing the expression of several specific genes, either through binding to their promoters or to regulatory sites located in the gene bodies (e.g enhancer or both mapped and unmapped alternative TSSs).

PML-RARα plays a direct role in the deregulation of processes fundamental for oncogenic transformation in both LtoE- and EtoL-shifted Valleys

To better portrait the deregulative role of the oncogene on normal Replication Timing, we performed two Gene Ontology (GO) analyses by taking advantage of the gene list enrichment analysis tool EnrichR (GO database: "GO_Biological_Process_2015") (Chen et al., 2013). One analysis was performed by taking into consideration all PML-RARα target genes in Differentially Replicating Valleys whereas in the other one we focused only on direct targets. Both analyses were performed separately for deregulated LtoE- and EtoL-shifted Valleys

Interestingly for us, we noticed that both in LtoE- and EtoL-shifted Valleys (Tables 14 and 15) the majority of the top ten biological processes (adjusted p-value <=0.005) involving deregulated genes were enriched also in the analysis performed only on direct PML-RARα targets, thus suggesting that PML-RARα might impact these processes through both direct and indirect mechanisms. Moreover, depending if in LtoE- or EtoL-shifted Valleys, PML-RARα seems to enhance or suppress processes being already known as often deregulated

in cancer, and particularly in leukemia, such as autophagy, glucose metabolism, cell ageing in LtoE-shifted Valleys and myeloid cell differentiation and regulation of telomere maintenance in EtoL-Shifted Valleys.

GO_term	Pvalue	Adj_pval	UpReg	DownReg
regulation of autophagy (GO:0010506)	1.38E-06	0.000188	**ABL2**,**ZKSCAN3**,WDFY3,WAC,**LARP1**,**KIF25**,**SIRT1**	**PLEKHF1**,**MTDH**
positive regulation of autophagy (GO:0010508)	4.09E-06	0.00028	WDFY3,WAC,**LARP1**,**SIRT1**	**PLEKHF1**,**MTDH**
positive regulation of macroautophagy (GO:0016239)	3.10E-05	0.001416	WDFY3,WAC,**LARP1**,**SIRT1**	none
positive regulation of glucose metabolic process (GO:0010907)	5.16E-05	0.001768	PMAIP1,KAT2B,**IRS1**,**IRS2**,**TCF7L2**	none
positive regulation of response to extracellular stimulus (GO:0032106)	9.13E-05	0.002084	WDFY3,WAC,**LARP1**,**SIRT1**	none
positive regulation of response to nutrient levels (GO:0032109)	9.13E-05	0.002084	WDFY3,WAC,**LARP1**,**SIRT1**	none
regulation of mesenchymal cell proliferation (GO:0010464)	0.000184	0.003597	SHOX2,**TGFBR2**,**IRS1**,**IRS2**	**PHF14**
regulation of cellular response to stress (GO:0080135)	0.000233	0.003994	**PMAIP1**,**GAB1**,**ZKSCAN3**,WDFY3,WAC,**LARP1**,**SIRT1**,**DUSP10**,**CDK6**,**PLCB1**,IGF1R	RIPK2,**SH3RF3**,TLR4
regulation of macroautophagy (GO:0016241)	0.000522	0.004553	WDFY3,WAC,**LARP1**,**SIRT1**	none
negative regulation of cell aging (GO:0090344)	0.000446	0.004553	**ZKSCAN3**,**SIRT1**,**CDK6**	none

Table 14 – GO terms enriched for deregulated genes in LtoE-shifted Valleys.
Deregulated genes in LtoE-shifted Valleys are enriched in GO terms relative to cell survival and adaptation. In red, GO terms being enriched also in the direct PML-RARα target genes GO analysis (Adjusted pvalue <=0.05). In bold: direct PML-RARα target genes

GO_term	Pvalue	Adj_pval	UpReg	DownReg
cAMP metabolic process (GO:0046058)	0.0002337	0.0072	CACNB4,PDE4D	ADCY4,ADCY7,PDE4B
DNA catabolic process, endonucleolytic (GO:0000737)	0.0004881	0.0072	NEIL3	ERCC1,ARRB1,DNASE2,AIFM1,NTHL1
myeloid cell differentiation (GO:0030099)	0.0005089	0.0072	TET2	KLF1,CEBPA,PRDX3,VEGFA,DNASE2,PPARG,IRF8,PIK3CD
NIK/NF-kappaB signaling (GO:0038061)	0.0005533	0.0072	none	RIPK3,IKBKE,MAP3K14
negative regulation of cysteine-type endopeptidase activity involved in apoptotic process (GO:0043154)	0.000655	0.0072	HGF	ARRB1,PRDX3,VEGFA,DNAJA3,BIRC5
negative regulation of telomere maintenance (GO:0032205)	0.0007118	0.0072	none	ERCC1,ACD,TINF2
negative regulation of cysteine-type endopeptidase activity (GO:2000117)	0.0007535	0.0072	HGF	ARRB1,PRDX3,VEGFA,DNAJA3,BIRC5
negative regulation of protein modification process (GO:0031400)	0.0009641	0.0072	PDE4D	DDIT4,ANAPC1,EPHA1,BCOR,PGS1,UBC,ARRB1,ENG,MAS1,DNAJA3,TINF2,PRKCD,ATG5,SEMA4D
mitochondrion (GO:0005739)	0.001061	0.0072	ZDHHC8	OPA3,DDIT4,RIPK3,CTSD,PDP2,SDS,MRPL18,CERK,PTRH1,PGS1,APOO,ALAS1,AK4,MFN2,SDHB,SDSL,H5D17B10,PRDX3,DNAJA3,DRG2,GCDH,COX4I1,FAM32A,FPGS,AIFM1,PRKCD,ECI1,CD3EAP,H5H2D,NTHL1,GRAMD4
regulation of chromosome organization (GO:0033044)	0.001076	0.0072	none	ANAPC1,BCOR,ERCC1,ACD,ARRB1,PADI2,VEGFA,TINF2,CDK9

Table 15 – GO terms enriched for deregulated genes in EtoL-shifted Valleys.
Deregulated genes in EtoL-shifted Valleys are enriched in GO terms relative to cell survival and adaptation. In red, GO terms being enriched also in the direct PML-RARα target genes GO analysis (Adjusted pvalue <=0.05). In bold: direct PML-RARα target genes

A different pre-existing chromatin asset characterizes the three classes of Replication Valleys

PML-RARα induces transcription repression of its target genes by inducing chromatin remodelling through recruitment of chromatin modifiers, such as DNA methylases and histone deacetylases directly to PML-RARα - bound sites (Croce et al., 2002; Grignani et al., 1998; Hoemme et al., 2008; Saeed et al., 2011; Villa et al., 2007).

To study how pre-existing chromatin structure is involved in the observed changes of replication timing describing above, anti-H3K4me3, anti-H3K4me1 and anti-H3K27Ac

ChIP-seq experiments were performed in Zinc-treated (t=8 hours) and Untreated U937PR9 cells (t=0 hour).

For each ChIP-seq experiment, we first calculated the normalized coverage (measured as the number of reads in each Valley divided by the total number of reads of each experiment) of the three histone marks within each Differentially Replicating Valley, and then compared the signal intensities.

As showed in Figure 33, we found that before induction of the PML-RARα expression, there were significant differences in the distribution of the three histone marks compared to Stable Valleys, with LtoE-shifted ones showing lower levels for all tested histone modifications and EtoL-shifted Valleys showing significantly higher signal intensities.

Altogether, these observations suggest that the pre-existing chromatin status within DRVs plays a role in the observed behaviour following PML-RARα induction.

Expression of PML-RARα significantly changes chromatin status in Differentially Replicating Valleys

We finally compared the signal intensity of the three histone-modifications (H3K4me3, H3K4me1 and H3K27Ac), before and after PML-RARα induction, at Stable and Differentially Replicating Valleys, by analysing ChIP-seq experiments performed in Zinc-treated (t=8 hours) and Untreated U937PR9 cells (t=0 hour). Although negligible compared to the differences among the three classes of Valleys in the signal distributions prior to PML-RAR induction (as discussed above), we observed for all the histone marks significant differences in distribution of coverage within both Differentially Replicating and Stable Valleys after PML-RARα induction (Figure 33).

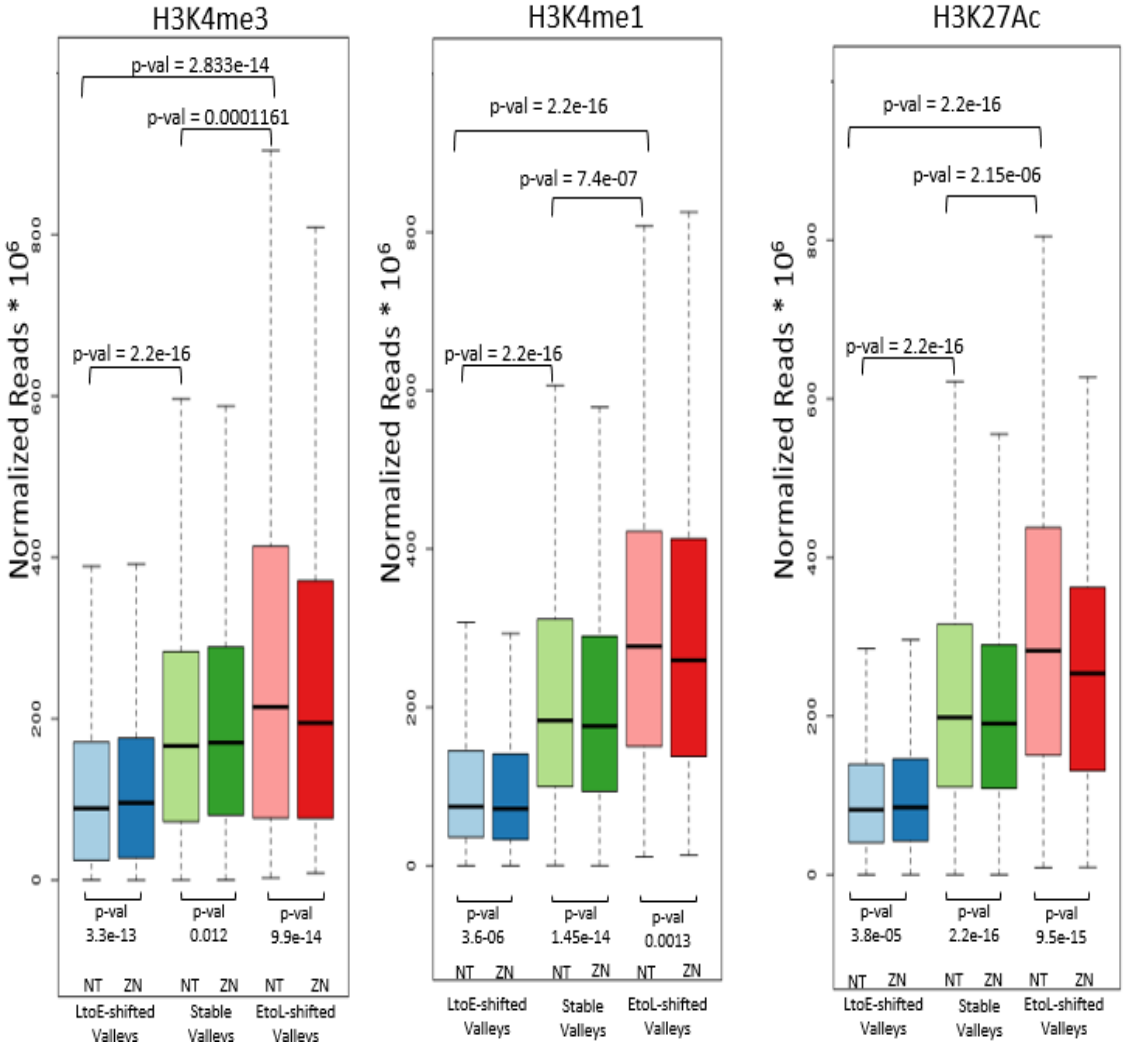

Figure 33 - Distribution of H3K4me3 (left), H3K4me1 (centre) and H3K27Ac (right) normalized coverage before ("NT") and after ("ZN") expression PML-RARα in DRVs. Differences between distributions were tested by the mean of a Mann-Whitney test. P-values are reported for each comparison.

We also compared the extent of these differences in all the three classes of Valleys, by measuring changes in signal intensity following PML-RARα inductions as log2ratio between the normalized coverages measured in the Zinc-treated cells *vs.* the Untreated ones, and statistical significance was assessed by the mean of a Mann-Whitney test (Figure 34).

While a slight decreased of H3K4me1 signal was observed for all the three classes of DRVs, with no significant differences among them, changes in H3K4me3 and H3K27Ac levels revealed strong differences. In particular, LtoE-shifted Valleys showed a dramatic increase of both histone modifications when compared to Stable and EtoL-shifted Valleys, while the latter displayed significant decrease in H3K4me3 and H3K27Ac signals when compared to LtoE and Stable Valleys (Figure 34).

In summary, the last data show that, among the three classes of Valleys, the EtoL-shifted ones are chacterized, at basal level, by the highest levels of mono- and tri-methylated H3K4, and acetylated H3K27, all generally associated with open, transcriptionally active chromatin; LtoE-shifted Valleys, instead, are characterized by the lowest levels of the three histone modifications.

Following PML-RARα induction, to some, but significant, extent, EtoL-shifted Valleys show decreased H3K4me3 and H3K27ac signals, while LtoE-shifted Valleys show increased signals of both histone marks.

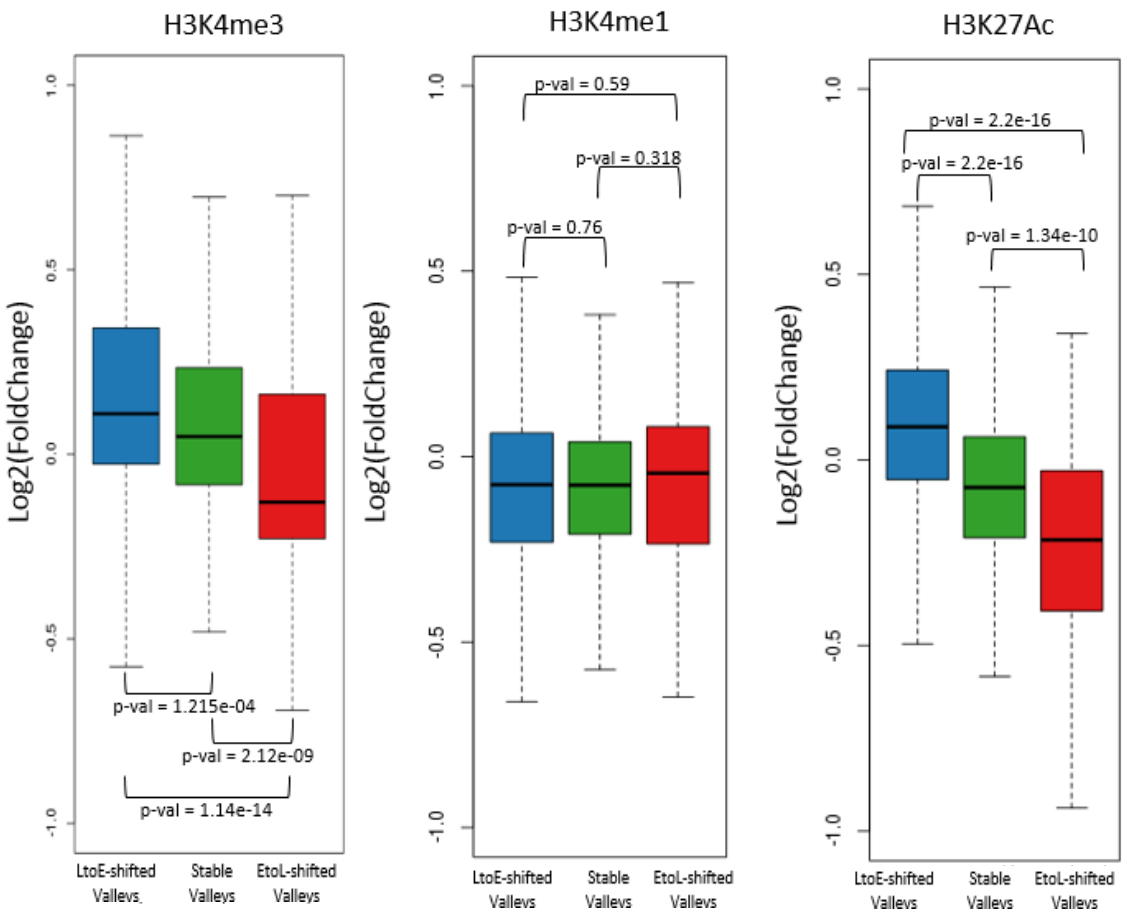

Figure 34 - Distributions of differences in coverage for H3K4me3 (left), H3K4me1 (centre) and H3K27Ac (right).

Differences for normalized coverage were calculated as log2($Coverage_{Zinc}$/$Coverage_{Untreated}$) in each DRV and differences for log2ratio were tested for significance among Valleys by the mean of a Mann-Whitney test. P-values are reported for each test.

DISCUSSION

DART is an excellent predictive method, comparing two Repli-seq experiments, to identify genomic regions that have undergone a change in the Replication Timing

Repli-seq (Hansen et al., 2010a) is one of the main techniques employed to study the dynamics of of DNA Replication. It allows to study the progression of DNA replication both in space and time (*i.e.* when, during the S phase (time), each specific region of the genome (space) is replicated).

However, up to now, all current studies that used Repli-seq, have analysed Replication Timing in one specific experimental condition, or have compared Repli-seq data with other methodologies (Barlow et al., 2013; Chen et al., 2010; Debatisse et al., 2012; Dellino et al., 2013a; Hansen et al., 2010a; Liu et al., 2015; Polak et al., 2015; Pope et al., 2014).

There are no studies, to our knowledge, that use Repli-seq experiments to measure genome-wide differences in Replication Timing between different cell lines, or experimental conditions in the same cell context, and test their statistically significance.

In this work, we developed a bioinformatic method that, comparing two different experimental conditions, allows the identification of genomic regions showing statistically significant differences in their Replication Timing.

The method, named DART (Differential Analysis of Replication Timing), is a bioinformatic pipeline which, starting from the Repli-seq experiments from two experimental conditions (or cell types), applies a series of manipulations to the data (normalization, scaling, transformation, etc) to measure the Replication Timing for each experimental condition or cell types (i.e., the S50 value) that can be visualized as a S50

profile in a Genome Browser and, more importantly, provides a list of genomic regions, enriched in DNA replication origins, showing statistically significant differences in Replication Timing.

Examining the predictive power of our method, we performed a ROC Curve analysis (Zweig and Campbell, 1993) using a control dataset of regions showing different Replication Timing, obtained by visual inspection of Repli-seq experiments on the UCSC Genome Browser.

We obtained an AUC equal to 0.95, which demonstrates an excellent predictive robustness of our method in identifying genomic regions that have undergone a change in the Replication Timing between two experimental conditions.

The identification of this bioinformatic procedure was used to ask whether and where an oncogene (we used PML-RARα, as prototype) regulates the timing of the DNA replication. We then characterized these differently replicated regions, by integrating Repli-seq with RNA-seq and ChIP-sequencing experiments.

For this study, we used the U937 cells as model system, a promonocytic cell line transduced with a vector carrying the PML-RARα cDNA under control of the inducible mouse metallothionein 1 (MT-1) promoter. The PR9 clone, used here and in several other publications to characterize the tumour initiation properties of PML-RARα, expresses the oncogene at levels comparable to the ones of APL fresh blasts, when induced with100 mM of Zinc.

Therefore, we applied DART to study changes in Replication Timing in U937PR9 cells before and after the expression of PML-RARα.

Expression of PML- RARα induces specific changes in normal Replication Timing

Different studies have shown a positive correlation between transcription and regions enriched for ORIs (Dellino et al., 2013; Martin et al., 2011; Mesner et al., 2011; Miotto et al., 2016), and a close association between Replication Timing and transcription, with regions with high transcriptional activity usually being replicated early in S phase, and silenced genes or gene-poor regions in late-S phase (Fraser, 2013; Huvet et al., 2007; Rivera-Mulia et al., 2015).

One of the best studied mechanisms through which PML-RARα exerts its oncogenic function is its repressive activity on transcription that is achieved by its direct binding to DNA and by the recruitment of chromatin modifiers, such histone deacetylases (HDACs) and DNA methyltransferases (DNMTs) to the promoter of its target genes (Grignani et al., 1998; Croce et al., 2002).

Since Repli-seq profiles did not show macroscopic differences between the two experimental condition, in an attempt to easily highlight reliable differences, we decided to focus our attention on changes of Replication Timing occurring within ORI-containing Regions ("Valleys") common to the two experimental conditions.

By applying the DART pipeline to our Repli-seq datasets, we retrieved 1,717 and 1,707 Valleys for, respectively, the Zinc-treated and Untreated cells.

We concentrated our further analyses on 1,379 common Valleys showing an overlap higher than 60% of their contigs (Jaccard index ≥0.66), in order to test them for differences in distributions of S50 values (*i.e.* our quantitative measure of the Replication Timing).

Among the common Valleys, we identified 536 which showed a late-to-early (LtoE-shifted Valleys) shift of Replication Timing, 414 showing an early-to-late shift of Replication Timing (EtoL-shifted Valleys) and 429 without differences in Replication timing when

comparing cells expressing PML-RARα and Untreated cells (Stable Valleys).

After a further step of filtering, based on thresholds defined with a ROC curve analysis, we finally came out with our final set of differently replicated regions: 440 LtoE-, 183 EtoL-shifted (that altogether covered about 5.5% of the genome) and 451 Stable Valleys, for a total of 1,074 regions.

Following PML-RARα expression, we therefore observed measurable changes of Replication Timing giving rise to both late-to-early and early-to-late shifts.

These results are in line with the ones obtained by an earlier study conducted by Ryba and colleagues on several leukemic paediatric patients (Ryba et al., 2012), where the authors showed that almost all patients showed deregulation of Replication Timing, with ~8% and ~5% of the genome showing late-to-early and early-to-late shifts in Replication Timing.

The novelty of our results - besides enhancing the resolution of the Replication Timing of several layers (from two to six different S-sub-phases) - is the possibility to investigate at the molecular level one of the mechanisms underlying the observed changes following oncogene expression.

Interestingly, characterization of Differently Replicated regions revealed that LtoE- and EtoL-shifted Valleys showed peculiar traits.

EtoL shift amplitudes in Replication Timing are smaller than LtoE ones

The extent of the timing variation (Diff_S50) was wider in the LtoE-shifted Valleys than in the EtoL-shifted ones. Probably for that reason, the last stringent filtering step we applied drops the number of originally identified EtoL-shifted Valleys of more than 50%.
Additional replicates may improve sensitivity for the identification of this specific class of regions, telling us if they exist or were just a peculiar feature of the samples taken in study.

Changes in Replication Timing induced by the PML-RARα expression involve regions normally replicated in different moments of the S phase

We also observed that LtoE and EtoL shifts of the Replication Timing took place in different time windows of the S-phase and, notably, both are not evenly spread. Indeed, LtoE shifts mainly involved Valleys replicating in Middle-S phase (S3-S4 S-phase fractions). On the other hand, early-to-late shifts were mainly observed in regions replicating during Early-S phase. Due to fractionation of S-phase in 6 temporal windows (based on the ranges of the sorting gates), we observed a comparable number of shifts in Replication Timing within the same S-phase fraction, or from one S-phase fraction to closest (towards either earlier- or later-S).

EtoL-shifted Valleys are more enriched for PML-RARα binding sites than LtoE-shifted ones.

A very interesting feature of LtoE- and EtoL-shifted Valleys was their different content in PML-RARα peaks.

We found that, when compared to Stable and LtoE-shifted Valleys, EtoL-shifted ones were significantly enriched for PML-RARα binding sites, with 1,654 peaks in 183 Valleys, with an average of 9 peaks per Valleys.

LtoE-shifted Valleys, on the other hand, were significantly depleted for PML-RARα peak, showing 1,091 binding sites in 440 Valleys, with an average of 2 peaks per Valley.

Transcription of genes lying in Differently Replicating regions reflects their Replication timing and changes

This observation allowed the study of a possible direct involvement of PML-RARα in the deregulation of Replication Timing, by integration of Repli-seq and ChIP-seq data. Consistent with previous observations, and with several studies that already demonstrated a strict mechanistically connection between replication and transcription ((Dellino et al., 2013b, 2013b; Fraser, 2013; Huvet et al., 2007; Miotto et al., 2016; Rivera-Mulia et al., 2015), we found that mean expression levels of genes located in the three classes of Valleys reflected their distribution along the S phase, with early-replicating regions containing high level of transcription.

More interestingly, variations in Replication Timing in Valleys replicating during a certain time window of the S phase, associated to specific changes in expression levels of the genes mapping within the same Valleys following PML-RARα induction.

Indeed, genes within LtoE-shifted-Valleys, mainly in S3 were up-regulated, whereas genes

within EtoL-shifted Valleys were Down-regulated at very early S sub-phases.

Overall, these data show that EtoL-shifted Valleys (at t=24 hours) are associated with PML-RARα - mediated transcription repression which takes place 16 hours earlier (at t=8 hours) in normally early-replicating regions, while in LtoE-shifted Valleys (at t=24 hours) PML-RARα causes transcription induction (at t=8 hours) in regions normally replicating later in Middle-S phase.

EtoL-Shifted Valleys are enriched for direct, promoter-bound, Down-regulated PML-RARα targets; LtoE-Shifted Valleys contain both up- and down-regulated direct PML-RARα targets

When we determined if the transcriptionally deregulated genes present in Differently Replicated Valleys were also directly bound by PML-RARα, we observed a striking segregation of states.

Indeed, in regions showing a EtoL-shifted Replication timing about, ~74% (176/237) of Down-regulated genes (which account for 91% of total deregulated genes in these regions – 237/260) were also bound by PML-RARα on their promoter, against only a ~ 22% (n=5/23) of up-regulated genes.

Conversely, in LtoE-shifted Valleys, where up-regulated genes account for ~77% (169/219) of total deregulated genes, there was not a clear difference in PML-RARα binding for up-regulated (~60% - 102/169) or down-regulated (~72% - 36/50) genes.

Interestingly though, also for this class of Valleys, down-regulated genes seemed to have

an higher affinity for PML-RARα binding onto their promoters (~66% - 33/50) than up-regulated genes (~37%, 63/169).

Thus, simplifying, we could state that relationship between Replication Timing changes and transcriptional deregulation fostered by the oncogene could be differently enacted for EtoL- and LtoE-shifted Valleys.

In EtoL-shifted Valleys, PML-RARα might elicit its oncogenic role by directly binding onto the promoters its target genes, inducing a transcriptionally repressed state, rather the inducing an increased expression. Indeed, since up-regulated genes in EtoL-shifted Valleys are present in such a small number, their contribution to deregulation of Replication Timing might be negligible when compared to the one given by down-regulated genes.

In LtoE-shifted Valleys, instead, oncogenic activity of PML-RARα exerted in a dual manner, with i) PML-RARα binding directly to the promoters of its down-regulated targets and ii) with a direct binding of the oncogene upon the regions located within gene-bodies of up-regulated genes, thus possibly involving binding to enhancers and/or both mapped and unmapped alternative TSSs.

LtoE and EtoL shifts in Replication timing associate with PML-RARα-mediated enhancement and suppression of biological processes known to be associated with tumorigenic transformation.

To gain the status of cancer cell, normal cells need to undergo a series of aberrant events which will confer them “hallmark” tumorigenic traits, such as immortalization, increased proliferation and invasiveness. (Hanahan and Weinberg, 2011).

Such events can be interpreted as modifications in the balance between gene networks or

alteration of whole biological pathways.

In accordance with this, when performing the GO Analysis on our samples, we found that several key biological processes were altered after the expression of PML-RARα.

In addition, we found that deregulated biological processes involved direct PML-RARα targets, thus suggesting both a direct and indirect role of the oncogene in the deregulation of such processes.

Among the altered biological processes, we could observe an induction of those conferring survival advantages in LtoE-shifted Valleys, such as autophagy, glucose metabolism and negative regulation of cell ageing.

Indeed, in literature, there are several evidences showing how aberrancy in such processes might confer an advantage in cancer cells.

Due to its cytoprotective role, abnormal autophagy has been shown to play a consistent role in cancer and oncogenic transformation by protecting the cell from apoptotic signals and preventing cell death due to low nutrients, metabolic stress and deprivation of growth factors (Boya et al., 2005; Jin and White, 2007; Lum et al., 2005). Particularly, Huang and colleagues (Huang et al., 2011), showed that PML-RARα expression results in a constitutive activation of autophagy, given as an increased formation of autophagosomes.

On the other hand, different studies have shown that cancer cells have an enhanced glucose metabolism (Kroemer and Pouyssegur, 2008), and in a recent investigation (Chen et al., 2014) it was shown that patients affected by Acute Myeloid Leukemia presented an altered glucose metabolism signature.

Finally, regarding the negative regulation of cell ageing, it is known that a necessary condition toward oncogenic transformation is the acquisition by the cell of the ability to escape senescence (Hanahan and Weinberg, 2011).

When performing the GO Analysis in EtoL-shifted Valleys, we found a negative regulation of leukaemia key processes, such as myeloid differentiation process and regulation of chromosome organization (being Acute Promyelocytic Leukemia characterized by a block of myeloid cells in the promyelocytic state and a reciprocal translocation between chromosomes 15 and 17).

Other important processes being inhibited were the negative regulation of telomere maintenance and NIK/NF-kappaB signalling, both being pervasively described in literature as key processes for oncogenic transformation (Cong et al., 2002; Hoesel and Schmid, 2013; Lopez-Guerra and Colomer, 2010; Luo et al., 2005; Reddel, 2014; Samassekou et al., 2009).

Taken together, these results possibly confirm the trend observed for deregulated genes in LtoE- and EtoL-shifted Valleys, resulting in PML-RARα acting by both direct and indirect mechanisms, with the abnormal activation in LtoE-shifted Valleys, or inhibition in EtoL-shifted Valleys, of biological processes which have been described as fundamental to achieve an oncogenic transformation.

LtoE-, EtoL-shifted and Stable Valleys are characterized by pre-existing different assets of epigenetic modifications, which coherently change after the expression of the PML-RARα oncogene

Several authors have shown that PML-RARα exerts part of its oncogenic function by modifying epigenetic markers (Croce et al., 2002; Grignani et al., 1998; Hoemme et al., 2008; Saeed et al., 2011; Villa et al., 2007), while others have shown an association between chromatin structure and Replication Timing (Aparicio et al., 2004; Bell et al., 2010; Gilbert et al., 2004; Goren et al., 2008; Hiratani and Gilbert, 2010; Ryba et al., 2010; Vogelauer et al., 2002a; Zappulla et al., 2002). Therefore, it is not surprising that within LtoE- and EtoL-shifted Valleys, ChIP-sequencing signals of three histone marks classically linked to an "active state" of the chromatin (H3K4me3, H3K4me1 and H3K27Ac), were truly different: for all the three markers, compared to the Stable Valleys, the LtoE- and EtoL-shifted Valleys showed significantly lower and significantly higher levels, respectively.

Moreover, consistent with the observed changes in transcription (at t=8 hours) and replication timing (at t=24 hours), following PML-RARα induction, we observed a significant increase, and a parallel significant decrease, for both H3K4me3 and H3K27Ac markers in LtoE- and EtoL-shifted Valleys.

Our proposed model

Altogether, the evidences collected in this study, although preliminary, prompt us to propose a possible model by which PML-RARα might impact on DNA Replication, eliciting both late-to-early, in greater extent, and early-to-late shifts of the Replication Timing.

Our hypothesis is that PML-RARα expression, by changing gene transcription and chromatin organization, alters normal DNA Replication Timing mainly in a direct manner, through its binding to open chromatin regions marked by histone modifications such as H3K4me3, H3K4me1 and H3K27ac.

It is difficult, at this stage, to conclude that a clear epistatic relationship exists between changes of transcription/chromatin organization induced by the oncogene and shifts of replication timing measured in our experimental model; however, alterations of transcription and chromatin structure following PML-RAR expression at early time points, observed at the same Valleys (i.e., DNA replication origin-containing regions) that, later on, show significant changes of Replication Timing, deserve further investigations.

Furthermore, identification of genomic regions showing changes of replication timing, now possible thanks to DART analysis, is instrumental for the identification, in space and time, at the genome scale, of potential sites of genomic instability deriving from unscheduled collisions between transcription and replication machineries.

Bibliography

Aladjem, M.I. (2007). Replication in context: dynamic regulation of DNA replication patterns in metazoans. Nat. Rev. Genet. *8*, 588–600.

Aladjem, M.I., Groudine, M., Brody, L.L., Dieken, E.S., Fournier, R.E., Wahl, G.M., and Epner, E.M. (1995). Participation of the human beta-globin locus control region in initiation of DNA replication. Science *270*, 815–819.

Alcalay, M., Meani, N., Gelmetti, V., Fantozzi, A., Fagioli, M., Orleth, A., Riganelli, D., Sebastiani, C., Cappelli, E., Casciari, C., et al. (2003). Acute myeloid leukemia fusion proteins deregulate genes involved in stem cell maintenance and DNA repair. J. Clin. Invest. *112*, 1751–1761.

Altman, A.L., and Fanning, E. (2004). Defined Sequence Modules and an Architectural Element Cooperate To Promote Initiation at an Ectopic Mammalian Chromosomal Replication Origin. Mol. Cell. Biol. *24*, 4138–4150.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinforma. Oxf. Engl. *31*, 166–169.

Aparicio, J.G., Viggiani, C.J., Gibson, D.G., and Aparicio, O.M. (2004). The Rpd3-Sin3 histone deacetylase regulates replication timing and enables intra-S origin control in Saccharomyces cerevisiae. Mol. Cell. Biol. *24*, 4769–4780.

Barlow, J.H., Faryabi, R.B., Callén, E., Wong, N., Malhowski, A., Chen, H.T., Gutierrez-Cruz, G., Sun, H.-W., McKinnon, P., Wright, G., et al. (2013). Identification of Early Replicating Fragile Sites that Contribute to Genome Instability. Cell *152*, 620–632.

Bell, S.P., and Dutta, A. (2002). DNA replication in eukaryotic cells. Annu. Rev. Biochem. *71*, 333–374.

Bell, S.P., and Stillman, B. (1992). ATP-dependent recognition of eukaryotic origins of DNA replication by a multiprotein complex. Nature *357*, 128–134.

Bell, O., Schwaiger, M., Oakeley, E.J., Lienert, F., Beisel, C., Stadler, M.B., and Schübeler, D. (2010). Accessibility of the Drosophila genome discriminates PcG repression, H4K16 acetylation and replication timing. Nat. Struct. Mol. Biol. *17*, 894–900.

Biondi, A., Rambaldi, A., Alcalay, M., Pandolfi, P.P., Lo Coco, F., Diverio, D., Rossi, V., Mencarelli, A., Longo, L., and Zangrilli, D. (1991). RAR-alpha gene rearrangements as a genetic marker for diagnosis and monitoring in acute promyelocytic leukemia. Blood *77*, 1418–1422.

Blow, J.J., Ge, X.Q., and Jackson, D.A. (2011). How dormant origins promote complete genome replication. Trends Biochem. Sci. *36*, 405–414.

Boos, D., Sanchez-Pulido, L., Rappas, M., Pearl, L.H., Oliver, A.W., Ponting, C.P., and Diffley, J.F.X. (2011). Regulation of DNA replication through Sld3-Dpb11 interaction is conserved from yeast to humans. Curr. Biol. CB *21*, 1152–1157.

Bowers, J.L., Randell, J.C.W., Chen, S., and Bell, S.P. (2004). ATP hydrolysis by ORC catalyzes reiterative Mcm2-7 assembly at a defined origin of replication. Mol. Cell *16*, 967–978.

Boya, P., González-Polo, R.-A., Casares, N., Perfettini, J.-L., Dessen, P., Larochette, N., Métivier, D., Meley, D., Souquere, S., Yoshimori, T., et al. (2005). Inhibition of Macroautophagy Triggers Apoptosis. Mol. Cell. Biol. *25*, 1025–1040.

Brás, A., Cotrim, C.Z., Vasconcelos, I., Mexia, J., Léonard, A., Sanzhar, I., Akhmatullina, N., and Rueff, J. (2008). Asynchronous DNA replication detected by fluorescence in situ hybridisation as a possible indicator of genetic damage in human lymphocytes. Oncol. Rep. *19*, 369–375.

Breger, K.S., Smith, L., Turker, M.S., and Thayer, M.J. (2004). Ionizing radiation induces frequent translocations with delayed replication and condensation. Cancer Res. *64*, 8231–8238.

Breger, K.S., Smith, L., and Thayer, M.J. (2005). Engineering translocations with delayed replication: evidence for cis control of chromosome replication timing. Hum. Mol. Genet. *14*, 2813–2827.

Burke, T.W., Cook, J.G., Asano, M., and Nevins, J.R. (2001). Replication factors MCM2 and ORC1 interact with the histone acetyltransferase HBO1. J. Biol. Chem. *276*, 15397–15408.

Cadoret, J.-C., Meisch, F., Hassan-Zadeh, V., Luyten, I., Guillet, C., Duret, L., Quesneville, H., and Prioleau, M.-N. (2008). Genome-wide studies highlight indirect links between human replication origins and gene regulation. Proc. Natl. Acad. Sci. U. S. A. *105*, 15837–15842.

Casini, T., and Pelicci, P.G. (1999). A function of p21 during promyelocytic leukemia cell differentiation independent of CDK inhibition and cell cycle arrest. Oncogene *18*, 3235–3243.

Chen, C.-L., Rappailles, A., Duquenne, L., Huvet, M., Guilbaud, G., Farinelli, L., Audit, B., d'Aubenton-Carafa, Y., Arneodo, A., Hyrien, O., et al. (2010). Impact of replication timing on non-CpG and CpG substitution rates in mammalian genomes. Genome Res. *20*, 447–457.

Chen, E.Y., Tan, C.M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, N.R., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics *14*, 128.

Chen, W.-L., Wang, J.-H., Zhao, A.-H., Xu, X., Wang, Y.-H., Chen, T.-L., Li, J.-M., Mi, J.-Q., Zhu, Y.-M., Liu, Y.-F., et al. (2014). A distinct glucose metabolism signature of acute myeloid leukemia with prognostic value. Blood *124*, 1645–1654.

Chen CL C-L CHEN - spatio temporal replication.pdf.

Cimbora, D.M., Schübeler, D., Reik, A., Hamilton, J., Francastel, C., Epner, E.M., and Groudine, M. (2000). Long-distance control of origin choice and replication timing in the human beta-globin locus are independent of the locus control region. Mol. Cell. Biol. *20*, 5581–5591.

Cong, Y.-S., Wright, W.E., and Shay, J.W. (2002). Human Telomerase and Its Regulation. Microbiol. Mol. Biol. Rev. *66*, 407–425.

Conti, C., Saccà, B., Herrick, J., Lalou, C., Pommier, Y., and Bensimon, A. (2007). Replication Fork Velocities at Adjacent Replication Origins Are Coordinately Modified during DNA Replication in Human Cells. Mol. Biol. Cell *18*, 3059–3067.

Croce, L.D., Raker, V.A., Corsaro, M., Fazi, F., Fanelli, M., Faretta, M., Fuks, F., Coco, F.L., Kouzarides, T., Nervi, C., et al. (2002). Methyltransferase Recruitment and DNA Hypermethylation of Target Promoters by an Oncogenic Transcription Factor. Science *295*, 1079–1082.

De, S., and Michor, F. (2011). DNA replication timing and long-range DNA interactions predict mutational landscapes of cancer genomes. Nat. Biotechnol. *29*, 1103–1108.

Debatisse, M., Toledo, F., and Anglana, M. (2004). Replication initiation in mammalian cells: changing preferences. Cell Cycle Georget. Tex *3*, 19–21.

Debatisse, M., Le Tallec, B., Letessier, A., Dutrillaux, B., and Brison, O. (2012). Common fragile sites: mechanisms of instability revisited. Trends Genet. *28*, 22–32.

Dellino, G.I., Cittaro, D., Piccioni, R., Luzi, L., Banfi, S., Segalla, S., Cesaroni, M., Mendoza-Maldonado, R., Giacca, M., and Pelicci, P.G. (2013a). Genome-wide mapping of human DNA-replication origins: levels of transcription at ORC1 sites regulate origin selection and replication timing. Genome Res. *23*, 1–11.

Dellino, G.I., Cittaro, D., Piccioni, R., Luzi, L., Banfi, S., Segalla, S., Cesaroni, M., Mendoza-Maldonado, R., Giacca, M., and Pelicci, P.G. (2013b). Genome-wide mapping of human DNA-replication origins: levels of transcription at ORC1 sites regulate origin selection and replication timing. Genome Res. *23*, 1–11.

DePamphilis, M.L. (1993). Eukaryotic DNA replication: anatomy of an origin. Annu. Rev. Biochem. *62*, 29–63.

Di Micco, R., Fumagalli, M., Cicalese, A., Piccinin, S., Gasparini, P., Luise, C., Schurra, C., Garre', M., Nuciforo, P.G., Bensimon, A., et al. (2006). Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. Nature *444*, 638–642.

Dominguez-Sola, D., Ying, C.Y., Grandori, C., Ruggiero, L., Chen, B., Li, M., Galloway, D.A., Gu, W., Gautier, J., and Dalla-Favera, R. (2007). Non-transcriptional control of DNA replication by c-Myc. Nature *448*, 445–451.

Donaldson, A.D. (2005). Shaping time: chromatin structure and the DNA replication programme. Trends Genet. TIG *21*, 444–449.

Donley, N., and Thayer, M.J. (2013). DNA Replication Timing, Genome Stability and Cancer. Semin. Cancer Biol. *23*, 80–89.

Dotan, Z.A., Dotan, A., Ramon, J., and Avivi, L. (2004). Altered mode of allelic replication accompanied by aneuploidy in peripheral blood lymphocytes of prostate cancer patients. Int. J. Cancer *111*, 60–66.

Eaton, M.L., Galani, K., Kang, S., Bell, S.P., and MacAlpine, D.M. (2010). Conserved nucleosome positioning defines replication origins. Genes Dev. *24*, 748–753.

Farkash-Amar, S., and Simon, I. (2010). Genome-wide analysis of the replication program in mammals. Chromosome Res. Int. J. Mol. Supramol. Evol. Asp. Chromosome Biol. *18*, 115–125.

Forrester, W.C., Epner, E., Driscoll, M.C., Enver, T., Brice, M., Papayannopoulou, T., and Groudine, M. (1990). A deletion of the human beta-globin locus activation region causes a major alteration in chromatin structure and replication across the entire beta-globin locus. Genes Dev. *4*, 1637–1649.

Fragkos, M., Ganier, O., Coulombe, P., and Méchali, M. (2015). DNA replication origin activation in space and time. Nat. Rev. Mol. Cell Biol. *16*, 360–374.

Fraser, H.B. (2013). Cell-cycle regulated transcription associates with DNA replication timing in yeast and human. Genome Biol. *14*, R111.

Fritz, A., Sinha, S., Marella, N., and Berezney, R. (2013). Alterations in replication timing of cancer-related genes in malignant human breast cancer cells. J. Cell. Biochem. *114*, 1074–1083.

Gaillard, C., Tokuyasu, T.A., Rosen, G., Sotzen, J., Vitaliano-Prunier, A., Roy, R., Passegué, E., de Thé, H., Figueroa, M.E., and Kogan, S.C. (2015a). Transcription and methylation analyses of preleukemic promyelocytes indicate a dual role for PML/RARA in leukemia initiation. Haematologica *100*, 1064–1075.

Gaillard, H., García-Muse, T., and Aguilera, A. (2015b). Replication stress and cancer. Nat. Rev. Cancer *15*, 276–289.

Ghosh, M., Liu, G., Randall, G., Bevington, J., and Leffak, M. (2004). Transcription factor binding and induced transcription alter chromosomal c-myc replicator activity. Mol. Cell. Biol. *24*, 10193–10207.

Gilbert, D.M. (2007). Replication origin plasticity, Taylor-made: inhibition vs recruitment of origins under conditions of replication stress. Chromosoma *116*, 341–347.

Gilbert, D.M. (2010). Evaluating genome-scale approaches to eukaryotic DNA replication. Nat. Rev. Genet. *11*, 673–684.

Gilbert, N., Boyle, S., Fiegler, H., Woodfine, K., Carter, N.P., and Bickmore, W.A. (2004). Chromatin architecture of the human genome: gene-rich domains are enriched in open chromatin fibers. Cell *118*, 555–566.

Göndör, A., and Ohlsson, R. (2009). Replication timing and epigenetic reprogramming of gene expression: a two-way relationship? Nat. Rev. Genet. *10*, 269–276.

Goren, A., Tabib, A., Hecht, M., and Cedar, H. (2008). DNA replication timing of the human beta-globin domain is controlled by histone modification at the origin. Genes Dev. *22*, 1319–1324.

Grignani, F., Fagioli, M., Ferrucci, P.F., Alcalay, M., and Pelicci, P.G. (1993a). The molecular genetics of acute promyelocytic leukemia. Blood Rev. *7*, 87–93.

Grignani, F., Ferrucci, P.F., Testa, U., Talamo, G., Fagioli, M., Alcalay, M., Mencarelli, A., Grignani, F., Peschle, C., Nicoletti, I., et al. (1993b). The acute promyelocytic

leukemia-specific PML-RARα fusion protein inhibits differentiation and promotes survival of myeloid precursor cells. Cell *74*, 423–431.

Grignani, F., Fagioli, M., Alcalay, M., Longo, L., Pandolfi, P.P., Donti, E., Biondi, A., Lo Coco, F., Grignani, F., and Pelicci, P.G. (1994). Acute promyelocytic leukemia: from genetics to treatment. Blood *83*, 10–25.

Grignani, F., De Matteis, S., Nervi, C., Tomassoni, L., Gelmetti, V., Cioce, M., Fanelli, M., Ruthardt, M., Ferrara, F.F., Zamir, I., et al. (1998). Fusion proteins of the retinoic acid receptor-α recruit histone deacetylase in promyelocytic leukaemia. Nature *391*, 815–818.

Grimwade, D., Biondi, A., Mozziconacci, M.J., Hagemeijer, A., Berger, R., Neat, M., Howe, K., Dastugue, N., Jansen, J., Radford-Weiss, I., et al. (2000). Characterization of acute promyelocytic leukemia cases lacking the classic t(15;17): results of the European Working Party. Groupe Français de Cytogénétique Hématologique, Groupe de Français d'Hematologie Cellulaire, UK Cancer Cytogenetics Group and BIOMED 1 European Community-Concerted Action "Molecular Cytogenetic Diagnosis in Haematological Malignancies." Blood *96*, 1297–1308.

Grinberg-Rashi, H., Cytron, S., Gelman-Kohan, Z., Litmanovitch, T., and Avivi, L. (2010). Replication timing aberrations and aneuploidy in peripheral blood lymphocytes of breast cancer patients. Neoplasia N. Y. N *12*, 668–674.

Haase, S.B., Heinzel, S.S., and Calos, M.P. (1994). Transcription inhibits the replication of autonomously replicating plasmids in human cells. Mol. Cell. Biol. *14*, 2516–2524.

Halazonetis, T.D., Gorgoulis, V.G., and Bartek, J. (2008). An oncogene-induced DNA damage model for cancer development. Science *319*, 1352–1355.

Hamlin, J.L., Mesner, L.D., Lar, O., Torres, R., Chodaparambil, S.V., and Wang, L. (2008). A revisionist replicon model for higher eukaryotic genomes. J. Cell. Biochem. *105*, 321–329.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of Cancer: The Next Generation. Cell *144*, 646–674.

Hansen, R.S., Canfield, T.K., Lamb, M.M., Gartler, S.M., and Laird, C.D. (1993). Association of fragile X syndrome with delayed replication of the FMR1 gene. Cell *73*, 1403–1409.

Hansen, R.S., Thomas, S., Sandstrom, R., Canfield, T.K., Thurman, R.E., Weaver, M., Dorschner, M.O., Gartler, S.M., and Stamatoyannopoulos, J.A. (2010a). Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. Proc. Natl. Acad. Sci. U. S. A. *107*, 139–144.

Hansen, R.S., Thomas, S., Sandstrom, R., Canfield, T.K., Thurman, R.E., Weaver, M., Dorschner, M.O., Gartler, S.M., and Stamatoyannopoulos, J.A. (2010b). Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. Proc. Natl. Acad. Sci. U. S. A. *107*, 139–144.

He, L.Z., Merghoub, T., and Pandolfi, P.P. (1999). In vivo analysis of the molecular pathogenesis of acute promyelocytic leukemia in the mouse and its therapeutic implications. Oncogene *18*, 5278–5292.

Heichinger, C., Penkett, C.J., Bähler, J., and Nurse, P. (2006). Genome-wide characterization of fission yeast DNA replication origins. EMBO J. *25*, 5171–5179.

Helmrich, A., Ballarino, M., Nudler, E., and Tora, L. (2013). Transcription-replication encounters, consequences and genomic instability. Nat. Struct. Mol. Biol. *20*, 412–418.

Hills, S.A., and Diffley, J.F.X. (2014). DNA Replication and Oncogene-Induced Replicative Stress. Curr. Biol. *24*, R435–R444.

Hiratani, I., and Gilbert, D.M. (2010). Autosomal lyonization of replication domains during early Mammalian development. Adv. Exp. Med. Biol. *695*, 41–58.

Hiratani, I., Ryba, T., Itoh, M., Yokochi, T., Schwaiger, M., Chang, C.-W., Lyou, Y., Townes, T.M., Schübeler, D., and Gilbert, D.M. (2008). Global Reorganization of Replication Domains During Embryonic Stem Cell Differentiation. PLOS Biol *6*, e245.

Hiratani, I., Ryba, T., Itoh, M., Rathjen, J., Kulik, M., Papp, B., Fussner, E., Bazett-Jones, D.P., Plath, K., Dalton, S., et al. (2010). Genome-wide dynamics of replication timing revealed by in vitro models of mouse embryogenesis. Genome Res. *20*, 155–169.

Hoemme, C., Peerzada, A., Behre, G., Wang, Y., McClelland, M., Nieselt, K., Zschunke, M., Disselhoff, C., Agrawal, S., Isken, F., et al. (2008). Chromatin modifications induced by PML-RARα repress critical targets in leukemogenesis as analyzed by ChIP-Chip. Blood *111*, 2887–2895.

Hoesel, B., and Schmid, J.A. (2013). The complexity of NF-κB signaling in inflammation and cancer. Mol. Cancer *12*, 86.

Huang, Y., Hou, J.-K., Chen, T.-T., Zhao, X.-Y., Yan, Z.-W., Zhang, J., Yang, J., Kogan, S.C., and Chen, G.-Q. (2011). PML-RARα enhances constitutive autophagic activity through inhibiting the Akt/mTOR pathway. Autophagy *7*, 1132–1144.

Huvet, M., Nicolay, S., Touchon, M., Audit, B., d'Aubenton-Carafa, Y., Arneodo, A., and Thermes, C. (2007). Human gene organization driven by the coordination of replication and transcription. Genome Res. *17*, 1278–1285.

Hyrien, O. (2000). Mechanisms and consequences of replication fork arrest. Biochimie *82*, 5–17.

Iizuka, M., Matsui, T., Takisawa, H., and Smith, M.M. (2006). Regulation of replication licensing by acetyltransferase Hbo1. Mol. Cell. Biol. *26*, 1098–1108.

Jeon, Y., Bekiranov, S., Karnani, N., Kapranov, P., Ghosh, S., MacAlpine, D., Lee, C., Hwang, D.S., Gingeras, T.R., and Dutta, A. (2005). Temporal profile of replication of human chromosomes. Proc. Natl. Acad. Sci. U. S. A. *102*, 6419–6424.

Jin, S., and White, E. (2007). Role of autophagy in cancer: management of metabolic stress. Autophagy *3*, 28–31.

Jones, R.M., Mortusewicz, O., Afzal, I., Lorvellec, M., García, P., Helleday, T., and Petermann, E. (2013). Increased replication initiation and conflicts with transcription underlie Cyclin E-induced replication stress. Oncogene *32*, 3744–3753.

Kakizuka, A., Miller, W.H., Umesono, K., Warrell, R.P., Frankel, S.R., Murty, V.V., Dmitrovsky, E., and Evans, R.M. (1991). Chromosomal translocation t(15;17) in human acute promyelocytic leukemia fuses RAR alpha with a novel putative transcription factor, PML. Cell *66*, 663–674.

Kalejta, R.F., Li, X., Mesner, L.D., Dijkwel, P.A., Lin, H.B., and Hamlin, J.L. (1998). Distal sequences, but not ori-beta/OBR-1, are essential for initiation of DNA replication in the Chinese hamster DHFR origin. Mol. Cell *2*, 797–806.

Karnani, N., Taylor, C.M., Malhotra, A., and Dutta, A. (2010). Genomic study of replication initiation in human chromosomes reveals the influence of transcription regulation and chromatin structure on origin selection. Mol. Biol. Cell *21*, 393–404.

Kastner, P., Perez, A., Lutz, Y., Rochette-Egly, C., Gaub, M.P., Durand, B., Lanotte, M., Berger, R., and Chambon, P. (1992). Structure, localization and transcriptional properties of two classes of retinoic acid receptor alpha fusion proteins in acute promyelocytic leukemia (APL): structural similarities with a new family of oncoproteins. EMBO J. *11*, 629–642.

Kazanov, M.D., Roberts, S.A., Polak, P., Stamatoyannopoulos, J., Klimczak, L.J., Gordenin, D.A., and Sunyaev, S.R. (2015). APOBEC-Induced Cancer Mutations Are Uniquely Enriched in Early-Replicating, Gene-Dense, and Active Chromatin Regions. Cell Rep. *13*, 1103–1109.

Kent, W.J., Zweig, A.S., Barber, G., Hinrichs, A.S., and Karolchik, D. (2010). BigWig and BigBed: enabling browsing of large distributed datasets. Bioinforma. Oxf. Engl. *26*, 2204–2207.

Kohzaki, H., and Murakami, Y. (2005). Transcription factors and DNA replication origin selection. BioEssays News Rev. Mol. Cell. Dev. Biol. *27*, 1107–1116.

Korenstein-Ilan, A., Barbul, A., Hasin, P., Eliran, A., Gover, A., and Korenstein, R. (2008). Terahertz radiation increases genomic instability in human lymphocytes. Radiat. Res. *170*, 224–234.

Kroemer, G., and Pouyssegur, J. (2008). Tumor cell metabolism: cancer's Achilles' heel. Cancer Cell *13*, 472–482.

Lallemand-Breitenbach, V., and de Thé, H. (2010). PML Nuclear Bodies. Cold Spring Harb. Perspect. Biol. *2*.

Lande-Diner, L., Zhang, J., and Cedar, H. (2009). Shifts in replication timing actively affect histone acetylation during nucleosome reassembly. Mol. Cell *34*, 767–774.

Lang, G.I., and Murray, A.W. (2011). Mutation rates across budding yeast chromosome VI are correlated with replication timing. Genome Biol. Evol. *3*, 799–811.

Lee, C., Hong, B., Choi, J.M., Kim, Y., Watanabe, S., Ishimi, Y., Enomoto, T., Tada, S., Kim, Y., and Cho, Y. (2004). Structural basis for inhibition of the replication licensing factor Cdt1 by geminin. Nature *430*, 913–917.

Leonard, A.C., and Grimwade, J.E. (2010). Regulating DnaA complex assembly: it is time to fill the gaps. Curr. Opin. Microbiol. *13*, 766–772.

Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics *25*, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. Bioinforma. Oxf. Engl. *25*, 2078–2079.

Li, P., Piao, Y., Shon, H.S., and Ryu, K.H. (2015). Comparing the normalization methods for the differential analysis of Illumina high-throughput RNA-Seq data. BMC Bioinformatics *16*, 347.

Liu, F., Ren, C., Li, H., Zhou, P., Bo, X., and Shu, W. (2015). De novo identification of replication-timing domains in the human genome by deep learning. Bioinformatics btv643.

Lopez-Guerra, M., and Colomer, D. (2010). NF-kappaB as a therapeutic target in chronic lymphocytic leukemia. Expert Opin. Ther. Targets *14*, 275–288.

Lopez-Mosqueda, J., Maas, N.L., Jonsson, Z.O., Defazio-Eli, L.G., Wohlschlegel, J., and Toczyski, D.P. (2010). Damage-induced phosphorylation of Sld3 is important to block late origin firing. Nature *467*, 479–483.

Loupart, M.L., Krause, S.A., and Heck, M.S. (2000). Aberrant replication timing induces defective chromosome condensation in Drosophila ORC2 mutants. Curr. Biol. CB *10*, 1547–1556.

Lubelsky, Y., Prinz, J.A., DeNapoli, L., Li, Y., Belsky, J.A., and MacAlpine, D.M. (2014). DNA replication and transcription programs respond to the same chromatin cues. Genome Res. *24*, 1102–1114.

Lum, J.J., Bauer, D.E., Kong, M., Harris, M.H., Li, C., Lindsten, T., and Thompson, C.B. (2005). Growth factor regulation of autophagy and cell survival in the absence of apoptosis. Cell *120*, 237–248.

Luo, J.-L., Kamata, H., and Karin, M. (2005). IKK/NF-κB signaling: balancing life and death – a new approach to cancer therapy. J. Clin. Invest. *115*, 2625–2632.

Martens, J.H.A., Brinkman, A.B., Simmer, F., Francoijs, K.-J., Nebbioso, A., Ferrara, F., Altucci, L., and Stunnenberg, H.G. (2010). PML-RARα/RXR Alters the Epigenetic Landscape in Acute Promyelocytic Leukemia. Cancer Cell *17*, 173–185.

Martin, M.M., Ryan, M., Kim, R., Zakas, A.L., Fu, H., Lin, C.M., Reinhold, W.C., Davis, S.R., Bilke, S., Liu, H., et al. (2011). Genome-wide depletion of replication initiation events in highly transcribed regions. Genome Res. *21*, 1822–1832.

Méchali, M. (2010). Eukaryotic DNA replication origins: many choices for appropriate answers. Nat. Rev. Mol. Cell Biol. *11*, 728–738.

Mesner, L.D., and Hamlin, J.L. (2005). Specific signals at the 3′ end of the DHFR gene define one boundary of the downstream origin of replication. Genes Dev. *19*, 1053–1066.

Mesner, L.D., and Hamlin, J.L. (2009). Isolation of restriction fragments containing origins of replication from complex genomes. Methods Mol. Biol. Clifton NJ *521*, 315–328.

Mesner, L.D., Valsakumar, V., Karnani, N., Dutta, A., Hamlin, J.L., and Bekiranov, S. (2011). Bubble-chip analysis of human origin distributions demonstrates on a genomic scale significant clustering into zones and significant association with transcription. Genome Res. *21*, 377–389.

Miotto, B., Ji, Z., and Struhl, K. (2016). Selectivity of ORC binding sites and the relation to replication timing, fragile sites, and deletions in cancers. Proc. Natl. Acad. Sci. U. S. A. *113*, E4810-4819.

Nervi, C., Poindexter, E.C., Grignani, F., Pandolfi, P.P., Lo Coco, F., Avvisati, G., Pelicci, P.G., and Jetten, A.M. (1992). Characterization of the PML-RAR alpha chimeric product of the acute promyelocytic leukemia-specific t(15;17) translocation. Cancer Res. *52*, 3687–3692.

Nieduszynski, C.A., Blow, J.J., and Donaldson, A.D. (2005). The requirement of yeast replication origins for pre-replication complex proteins is modulated by transcription. Nucleic Acids Res. *33*, 2410–2420.

Nilsson, K., Forsbeck, K., Gidlund, M., Sundström, C., Tötterman, T., Sällström, J., and Venge, P. (1981). Surface characteristics of the U-937 human histiocytic lymphoma cell line: specific changes during inducible morphologic and functional differentiation in vitro. Haematol. Blood Transfus. *26*, 215–221.

Paixão, S., Colaluca, I.N., Cubells, M., Peverali, F.A., Destro, A., Giadrossi, S., Giacca, M., Falaschi, A., Riva, S., and Biamonti, G. (2004). Modular structure of the human lamin B2 replicator. Mol. Cell. Biol. *24*, 2958–2967.

Pandolfi, P.P., Alcalay, M., Fagioli, M., Zangrilli, D., Mencarelli, A., Diverio, D., Biondi, A., Lo Coco, F., Rambaldi, A., and Grignani, F. (1992). Genomic variability and alternative splicing generate multiple PML/RAR alpha transcripts that encode aberrant PML proteins and PML/RAR alpha isoforms in acute promyelocytic leukaemia. EMBO J. *11*, 1397–1407.

Pappas, D.L., Frisch, R., and Weinreich, M. (2004). The NAD+-dependent Sir2p histone deacetylase is a negative regulator of chromosomal DNA replication. Genes Dev. *18*, 769–781.

Polak, P., Karlić, R., Koren, A., Thurman, R., Sandstrom, R., Lawrence, M.S., Reynolds, A., Rynes, E., Vlahoviček, K., Stamatoyannopoulos, J.A., et al. (2015). Cell-of-origin chromatin organization shapes the mutational landscape of cancer. Nature *518*, 360–364.

Pope, B.D., Hiratani, I., and Gilbert, D.M. (2010a). Domain-wide regulation of DNA replication timing during mammalian development. Chromosome Res. Int. J. Mol. Supramol. Evol. Asp. Chromosome Biol. *18*, 127–136.

Pope, B.D., Hiratani, I., and Gilbert, D.M. (2010b). Domain-wide regulation of DNA replication timing during mammalian development. Chromosome Res. Int. J. Mol. Supramol. Evol. Asp. Chromosome Biol. *18*, 127–136.

Pope, B.D., Tsumagari, K., Battaglia, D., Ryba, T., Hiratani, I., Ehrlich, M., and Gilbert, D.M. (2011). DNA Replication Timing Is Maintained Genome-Wide in Primary Human Myoblasts Independent of D4Z4 Contraction in FSH Muscular Dystrophy. PLoS ONE *6*.

Pope, B.D., Ryba, T., Dileep, V., Yue, F., Wu, W., Denas, O., Vera, D.L., Wang, Y., Hansen, R.S., Canfield, T.K., et al. (2014). Topologically associating domains are stable units of replication-timing regulation. Nature *515*, 402–405.

Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

Reddel, R.R. (2014). Telomere Maintenance Mechanisms in Cancer: Clinical Implications. Curr. Pharm. Des. *20*, 6361–6374.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47.

Rivera-Mulia, J.C., Buckley, Q., Sasaki, T., Zimmerman, J., Didier, R.A., Nazor, K., Loring, J.F., Lian, Z., Weissman, S., Robins, A.J., et al. (2015). Dynamic changes in replication timing and gene expression during lineage specification of human pluripotent stem cells. Genome Res. *25*, 1091–1103.

Robinson, M.D., and Oshlack, A. (2010). A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. *11*, R25.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinforma. Oxf. Engl. *26*, 139–140.

Ryba, T., Hiratani, I., Lu, J., Itoh, M., Kulik, M., Zhang, J., Schulz, T.C., Robins, A.J., Dalton, S., and Gilbert, D.M. (2010). Evolutionarily conserved replication timing profiles predict long-range chromatin interactions and distinguish closely related cell types. Genome Res. *20*, 761–770.

Ryba, T., Battaglia, D., Chang, B.H., Shirley, J.W., Buckley, Q., Pope, B.D., Devidas, M., Druker, B.J., and Gilbert, D.M. (2012). Abnormal developmental control of replication-timing domains in pediatric acute lymphoblastic leukemia. Genome Res. *22*, 1833–1844.

Saeed, S., Logie, C., Stunnenberg, H.G., and Martens, J.H.A. (2011). Genome-wide functions of PML–RARα in acute promyelocytic leukaemia. Br. J. Cancer *104*, 554–558.

Sainty, D., Liso, V., Cantù-Rajnoldi, A., Head, D., Mozziconacci, M.J., Arnoulet, C., Benattar, L., Fenu, S., Mancini, M., Duchayne, E., et al. (2000). A new morphologic classification system for acute promyelocytic leukemia distinguishes cases with underlying PLZF/RARA gene rearrangements. Blood *96*, 1287–1296.

Samassekou, O., Ntwari, A., Hébert, J., and Yan, J. (2009). Individual Telomere Lengths in Chronic Myeloid Leukemia. Neoplasia N. Y. N *11*, 1146–1154.

Sasaki, T., Ramanathan, S., Okuno, Y., Kumagai, C., Shaikh, S.S., and Gilbert, D.M. (2006). The Chinese Hamster Dihydrofolate Reductase Replication Origin Decision Point Follows Activation of Transcription and Suppresses Initiation of Replication within Transcription Units. Mol. Cell. Biol. *26*, 1051–1062.

Schaarschmidt, D., Baltin, J., Stehle, I.M., Lipps, H.J., and Knippers, R. (2004). An episomal mammalian replicon: sequence-independent binding of the origin recognition complex. EMBO J. *23*, 191–201.

Schwaiger, M., Kohler, H., Oakeley, E.J., Stadler, M.B., and Schübeler, D. (2010). Heterochromatin protein 1 (HP1) modulates replication timing of the Drosophila genome. Genome Res. *20*, 771–780.

Seila, A.C., Core, L.J., Lis, J.T., and Sharp, P.A. (2009). Divergent transcription: a new feature of active promoters. Cell Cycle Georget. Tex *8*, 2557–2564.

Shechter, D., Costanzo, V., and Gautier, J. (2004). ATR and ATM regulate the timing of DNA replication origin firing. Nat. Cell Biol. *6*, 648–655.

Shen, Z.-X., Shi, Z.-Z., Fang, J., Gu, B.-W., Li, J.-M., Zhu, Y.-M., Shi, J.-Y., Zheng, P.-Z., Yan, H., Liu, Y.-F., et al. (2004). All-trans retinoic acid/As2O3 combination yields a high quality remission and survival in newly diagnosed acute promyelocytic leukemia. Proc. Natl. Acad. Sci. U. S. A. *101*, 5328–5335.

Stamatoyannopoulos, J.A., Adzhubei, I., Thurman, R.E., Kryukov, G.V., Mirkin, S.M., and Sunyaev, S.R. (2009). Human mutation rate associated with DNA replication timing. Nat. Genet. *41*, 393–395.

Stephens, P.J., Greenman, C.D., Fu, B., Yang, F., Bignell, G.R., Mudie, L.J., Pleasance, E.D., Lau, K.W., Beare, D., Stebbings, L.A., et al. (2011). Massive genomic rearrangement acquired in a single catastrophic event during cancer development. Cell *144*, 27–40.

Sundström, C., and Nilsson, K. (1976). Establishment and characterization of a human histiocytic lymphoma cell line (U-937). Int. J. Cancer *17*, 565–577.

Symeonidou, I.-E., Taraviras, S., and Lygerou, Z. (2012). Control over DNA replication in time and space. FEBS Lett. *586*, 2803–2812.

Tallman, M.S., and Kwaan, H.C. (1992). Reassessing the hemostatic disorder associated with acute promyelocytic leukemia. Blood *79*, 543–553.

Tanaka, S., and Araki, H. (2013). Helicase activation and establishment of replication forks at chromosomal origins of replication. Cold Spring Harb. Perspect. Biol. *5*, a010371.

Taylor, J.H. (1977). Increase in DNA replication sites in cells held at the beginning of S phase. Chromosoma *62*, 291–300.

Testa, U., Masciulli, R., Tritarelli, E., Pustorino, R., Mariani, G., Martucci, R., Barberi, T., Camagna, A., Valtieri, M., and Peschle, C. (1993). Transforming growth factor-beta potentiates vitamin D3-induced terminal monocytic differentiation of human leukemic cell lines. J. Immunol. Baltim. Md 1950 *150*, 2418–2430.

de Thé, H., Lavau, C., Marchio, A., Chomienne, C., Degos, L., and Dejean, A. (1991). The PML-RARα fusion mRNA generated by the t(15;17) translocation in acute promyelocytic leukemia encodes a functionally altered RAR. Cell *66*, 675–684.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinforma. Oxf. Engl. *25*, 1105–1111.

Vashee, S., Cvetic, C., Lu, W., Simancek, P., Kelly, T.J., and Walter, J.C. (2003). Sequence-independent DNA binding and replication initiation by the human origin recognition complex. Genes Dev. *17*, 1894–1908.

Villa, R., Pasini, D., Gutierrez, A., Morey, L., Occhionorelli, M., Viré, E., Nomdedeu, J.F., Jenuwein, T., Pelicci, P.G., Minucci, S., et al. (2007). Role of the polycomb repressive complex 2 in acute promyelocytic leukemia. Cancer Cell *11*, 513–525.

Vogelauer, M., Rubbi, L., Lucas, I., Brewer, B.J., and Grunstein, M. (2002a). Histone acetylation regulates the time of replication origin firing. Mol. Cell *10*, 1223–1233.

Vogelauer, M., Rubbi, L., Lucas, I., Brewer, B.J., and Grunstein, M. (2002b). Histone Acetylation Regulates the Time of Replication Origin Firing. Mol. Cell *10*, 1223–1233.

Walter, J.C., and Araki, H. (2006). 4 Activation of Pre-replication Complexes. Cold Spring Harb. Monogr. Arch. *47*, 89–104.

Wang, Z.-Y., and Chen, Z. (2008). Acute promyelocytic leukemia: from highly fatal to highly curable. Blood *111*, 2505–2515.

Wang, K., Wang, P., Shi, J., Zhu, X., He, M., Jia, X., Yang, X., Qiu, F., Jin, W., Qian, M., et al. (2010). PML/RARalpha targets promoter regions containing PU.1 consensus and RARE half sites in acute promyelocytic leukemia. Cancer Cell *17*, 186–197.

Wang, L., Lin, C.-M., Brooks, S., Cimbora, D., Groudine, M., and Aladjem, M.I. (2004). The Human β-Globin Replication Initiation Region Consists of Two Modular Independent Replicators. Mol. Cell. Biol. *24*, 3373–3386.

Watanabe, Y., and Maekawa, M. (2010). Spatiotemporal regulation of DNA replication in the human genome and its association with genomic instability and disease. Curr. Med. Chem. *17*, 222–233.

Wohlschlegel, J.A., Dwyer, B.T., Dhar, S.K., Cvetic, C., Walter, J.C., and Dutta, A. (2000). Inhibition of eukaryotic DNA replication by geminin binding to Cdt1. Science *290*, 2309–2312.

Woodward, A.M., Göhler, T., Luciani, M.G., Oehlmann, M., Ge, X., Gartner, A., Jackson, D.A., and Blow, J.J. (2006). Excess Mcm2-7 license dormant origins of replication that can be used under conditions of replicative stress. J. Cell Biol. *173*, 673–683.

Wu, J.R., and Gilbert, D.M. (1997). The replication origin decision point is a mitogen-independent, 2-aminopurine-sensitive, G1-phase event that precedes restriction point control. Mol. Cell. Biol. *17*, 4312–4321.

Yue, F., Cheng, Y., Breschi, A., Vierstra, J., Wu, W., Ryba, T., Sandstrom, R., Ma, Z., Davis, C., Pope, B.D., et al. (2014). A comparative encyclopedia of DNA elements in the mouse genome. Nature *515*, 355–364.

Zappulla, D.C., Sternglanz, R., and Leatherwood, J. (2002). Control of replication timing by a transcriptional silencer. Curr. Biol. CB *12*, 869–875.

Zegerman, P., and Diffley, J.F.X. (2010). Checkpoint-dependent inhibition of DNA replication initiation by Sld3 and Dbf4 phosphorylation. Nature *467*, 474–478.

Zhang, J., Xu, F., Hashimshony, T., Keshet, I., and Cedar, H. (2002). Establishment of transcriptional competence in early and late S phase. Nature *420*, 198–202.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zweig, M.H., and Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin. Chem. *39*, 561–577.