S-phase induced RNAs initiate formation of DNA replication origins

Simone Ummarino[1ǂ], Alex K. Ebralidze[1ǂ], Emanuele Monteleone[2], Yhanzou Zhang[1], Bon Q. Trinh[1], Annalisa Di Ruscio*[1, 3§], and Daniel G. Tenen*[1,4,5 §].

*,ǂ These two authors equally contributed to the work.

[1]Harvard Medical School Initiative for RNA Medicine, Harvard Medical School, Boston, MA, 02115, USA

[2]Department of Molecular Biotechnology and Health Sciences, and GenoBiToUS, Genomics and Bioinformatics Service, University of Turin, Turin, 10126, Italy

[3]University of Eastern Piedmont, Department of Translational Medicine, Novara, 28100, Italy

[4]Cancer Science Institute, National University of Singapore, 117456, Singapore

[5]Harvard Stem Cell Institute, Harvard Medical School, Boston, MA, 02115, USA

[§]Contact:

Daniel G. Tenen
Center for Life Sciences
3 Blackfan Circle
Room 437
Boston, MA 02115
Tel: 617-735-2235
FAX: 617-735-2222
daniel.tenen@nus.edu.sg

Annalisa Di Ruscio
Center for Life Sciences
3 Blackfan Circle
Room 407
Boston, MA 02115
Tel: 617-735-2217
FAX: 617-735-2222
adirusci@bidmc.harvard.ed

Abstract

The mechanistic events leading to alterations of DNA replication in normal and diseased cells remain unknown. The present study uncovers the role of a new class of non-coding RNAs induced in early S-phase, able to activate and repress replication origins formation. The discovery of a RNA-mediated mechanism governing the replication of c-MYC locus offers novel strategies to control DNA replication by modulating transcriptional activity, and pave the way for future clinical applications to treat cancer and genetic disorders.

Introduction

In human cells, more than 6 billion base pairs of DNA need to be replicated and packaged into chromatin in every cell division cycle. Failures lead to mutations, epigenetic changes and chromosomal aberrations that ultimately cause genetic diseases and cancer (Gaillard, Garcia-Muse et al. 2015).

Therefore, modern genomic approaches are trying to identify the "DNA replication timing (RT)", which is a cell-type specific feature, and refers to the temporal order whereby defined units of chromosomes replicate during the course of the S-phase in the cell cycle.

Interestingly, normal cell differentiation involves reproducible changes in RT, that is disrupted in malignant transformation. It is also known, that a single gene deletion can affects global replication timing and chromatin architecture in mammals *(Marchal, Sima et al.* ***2019****)*. Comparison RT profiles between normal and diseased tissues enable the identification of replication defects at specific gene loci: i.e. *ROR1, CCNE1* and *c-MYC* that would not be identified by standard transcriptome analyses *(Macheret and Halazonetis* ***2018****, Rivera-Mulia, Sasaki et al.* ***2019****)*.

With the introduction of "*click chemistry",* numerous replication origins have been identified in the mammalian genome (*Macheret and Halazonetis **2019**)*. However, the lack of methods to re-direct the replication of targeted genes, hinder the development of protocols for clinical correction of replication-related diseases.

In mammals, the recognition of the origin site is mediated by a hexameric protein subunits, termed the Origin Recognition Complex (ORCs) that plays a pivotal role in the initiation of DNA replication.

In this study we have investigated interaction between long noncoding RNAs (l*ncRNAs*) and ORCs. We have shown that these interactions are essential in the recruitment of the complexes at the origin sites, and define a regulatory mechanism to drive the replication timing of cells.

Herein, we provide for the first time a molecular mechanism to describe how "*transcriptional activity regulates DNA Replication"* and we demonstrate how modulating *lncRNAs*-ORCs association on chromatin opens the possibility of induction or halting the formation of the replicating origins along mammalian chromosomes.

Translating this knowledge to a disease model of cancer progression driven by dysregulated DNA replication such as Chronic myelogenous leukemia (CML), that is a clonal myeloproliferative disease characterized by excessive proliferation of progenitor cells with massive accumulation of mature myeloid lineage cells (*Sawyers **1999***), we developed a novel therapeutic strategy to re-direct gene-specific replication, in order to correct deleterious replication timing patterns.

Identification and characterization of replication origins in leukemia models.

While in prokaryotes, the DNA replicate through the well-defined single origin of replication, the DNA in eukaryotes forms multiple origins of replication across the genome, and majority of them are not conserved among different cell-types or tissues (*Akerman, Kasaai et al.* ***2020***).

Although, the replication origin, of few genes including: *TOP1*, *LMN B2* and *c-MYC* has been clearly defined within the 300 to 500 nucleotides region within the gene locus *(Giacca, Zentilin et al.* ***1994****, Keller, Ladenburger et al.* ***2002****, Leman, Noguchi et al.* ***2010****)*, the molecular mechanism for the activation of those replication origins remained unknown.

By applying a combination of ORCs-Chromatin immunoprecipitation analyses ChIP-qPCR and/or ChIP-seq), we identified precise position of ORCs in K562, HL60, and HEK293 cell lines. The experiment was performed using ORC1, ORC2 and replacement histone H2A.Z (including its acetylated form - acH2A.Z) antibodies. H2A.Z and acH2A.Z were included in this study as recently identified markers of replication origin: deposition of H2A.Z in precise location of the genome facilitates licensing and activation of early replication origin (*Long, Zhang et al.* ***2020****)*. Figure 1 demonstrates the overlap of ORC1 and ORC2 binding to chromatin with the deposition of the both H2A.Z and acH2A.Z at the replication origin of the *c-MYC* locus.

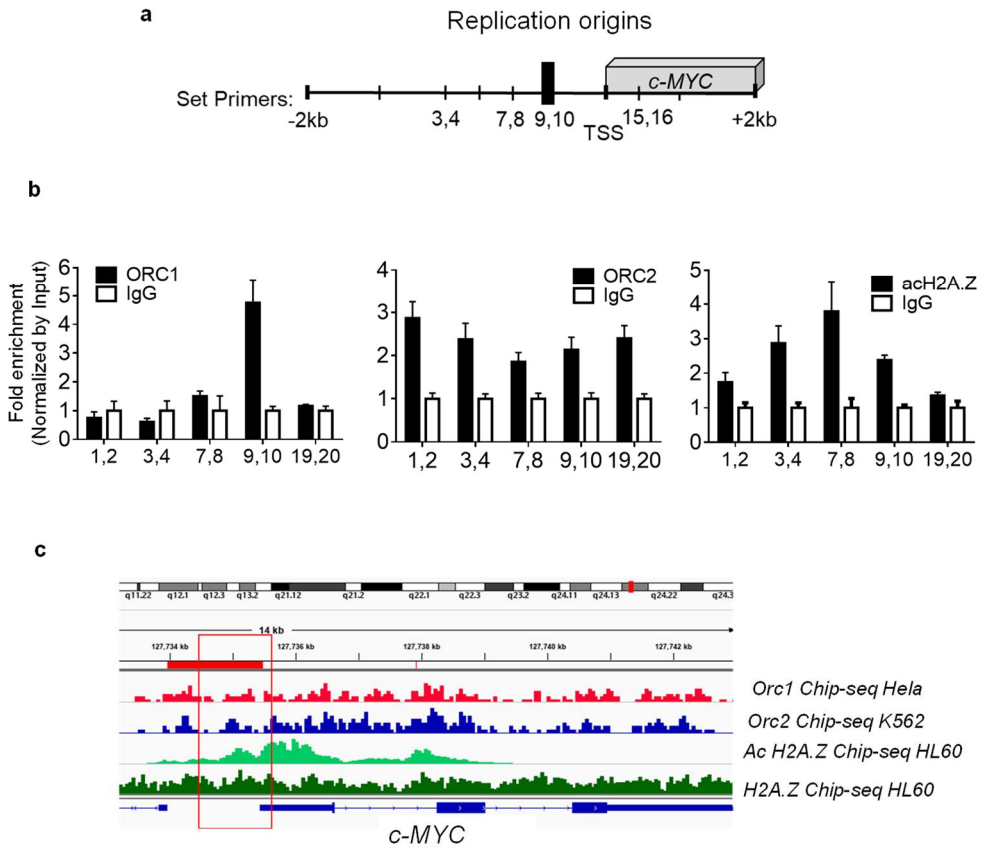


Figure 1. **Enrichment of ORC1, ORC2 and H2A.Z in the origin of *c-MYC*. a**, schematic of *c-MYC* locus. Black bars indicate the screened regions upstream of *c-MYC* TSS. The replication origin localizes in region 9,10; **b**, enrichment of ORC1, ORC2 and acH2A.Z **across the 2 kb region upstream** *c-MYC* TSS; **c**, ChIP-seq peak tracks comparison of ORC1, ORC2, H2A.Z and acH2A.Z. The red square indicates the origin of replication.

We found a diverse behavior of the three proteins acH2A.Z displayed the strongest enrichment in the origin (please refer to region 7,8 and 9,10 in **Fig. 1 a,b**) of almost 3-fold in comparison to the other analyzed sites. ORC1 which is the pioneer protein, shows almost 4.5-fold enrichment in the origin, while ORC2 pattern seems to be more spread across the locus (*Jaremko, On et al.* **2020**). Using already published database (*Dellino, Cittaro et al.* **2013**, *Miotto, Ji et al.* **2016**), we further compared the ChIP sequencing data from three different cell lines previously published (*Dellino,*

Cittaro et al. **2013**, *Miotto, Ji et al.* **2016**) and detected a consistent overlap for ORC1, ORC2 and H2A.Z peaks within the *c-MYC* origin.

Perturbation of transcriptional activity correlates with anomalous replication patterns.

Previously we identified and characterized a specific class of lncRNAs induced in the early S-Phase, which drive the modification and deposition of the acetylated form of histone H2A.Z by bringing together the replacement histone and the histone acetyl transferase TIP60. These lncRNAs were defined by performing nasRNA-seq (nascent RNA sequencing) of nuclear RNAs isolated from the synchronized HL60 cells (*Ebralidze*, *Ummarino; et al.* **2021**)

In this study a RIP (RNA immunoprecipitation) followed by qRT-PCR analyses demonstrates the interaction between "ORC proteins and *c-MYC*-l*ncRNAs"* (**Fig. 2a**) occurs in the proximity of *c-MYC* origin of replication, and is favored in S phase.

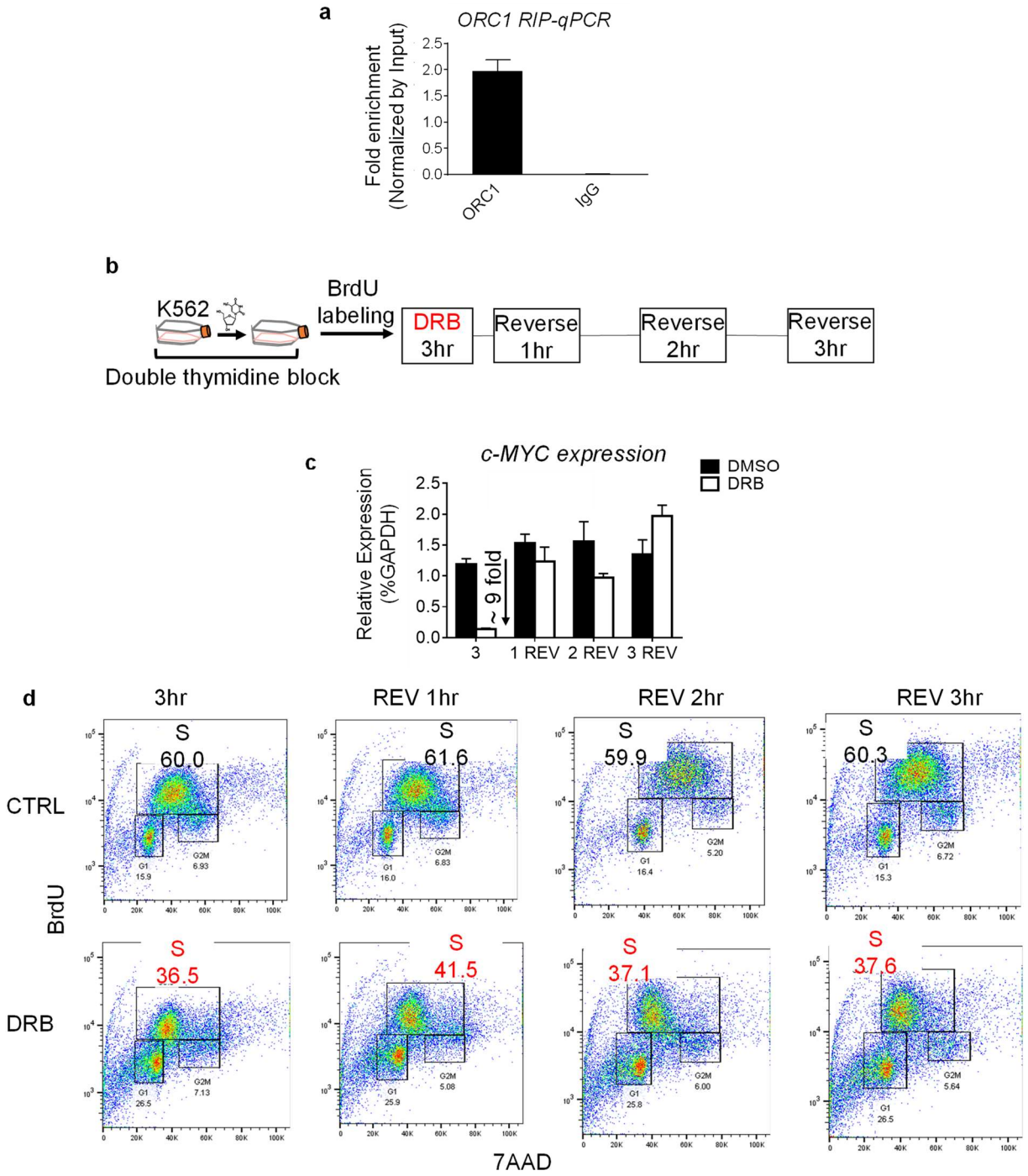


Figure 2. **Functional role of S-phase ncRNAs. a**, RIP-qPCR analysis of *c-MYC-ncRNAs* performed in K562 (respectively, black bar ORC1 antibody and white bar IgG antibody); **b,** A schematic of the experiment is shown in the upper part, release into the S-phase of synchronized K562, followed by BrdU labelling of nasDNA. **c**,**d** K562 were then treated with DRB for 3hr and washed with fresh media to reverse the inhibition for additional 3hr. Samples were collected every

hour to perform, RTPCR **c**, and FACS analysis **d**. Impaired transcription is shown only during DRB treatment, and impaired DNA synthesis was observed in all the time points.

DNA replication must be precisely regulated during cell cycle in order to ensure genome stability, and an extensive body of work has clearly demonstrated that induced activation of oncogenes provokes replication stress at susceptible genomic sites through different molecular mechanisms *(Macheret and Halazonetis **2018**)*. We used a similar approach to perturb globally ORC-*lncRNA* interactions, an outline of the experiment is presented in **Fig. 2a**.

K562 were firstly synchronized by double thymidine block *(Tonini, Radzioch et al. **1987**)* and during their release into the S phase, the reversible inhibitor of Polymerases (Pol II and Pol III), 5,6-Dichloro-1-β- d-ribofuranosylbenzimidazole (DRB) was added to the cell cultures.

Ultimately, we measured the abundance of nasDNA by BrdU staining and FACS analysis. The technique was used to describe the dynamic formation and/or depletion of replication origins together with replication fork movement **(Fig 2b-d).**

Interestingly, we discovered that transient inhibition of Pol II/Pol III provokes a reversible impairment of global transcription (**Fig. 2c**) but irreversible impairment of DNA replication (**Fig. 2d**) demonstrating ORCs-*lncRNA*s association strictly regulate the replication origins formation and disruption of several gene loci.

***c-MYC lncRNAs* modulate c-MYC locus replication.**

Using two targeted approaches, we demonstrated modulation of the ORCs-l*ncRNAs* interaction, in order to interfere with the ORCs recruitment at single gene locus.

Loss-of-Function Approach. Perturbation of the replication for a single gene was performed by repressing or up-regulating selected *ncRNAs* with CRISPR interference (CRISPRi, CRISPR-

dCAS9 system) *(Larson, Gilbert et al.* ***2013***), and CRISPR activator (CRISPRa, CRISPR-VP64) *(Perez-Pinera, Kocak et al.* ***2013)*** respectively.

After a preliminary screening of the gRNAs, the CRISPRi system was optimized in K562, to downregulate c-MYC *lncRNA*, and a Tet-inducible promoter was introduced in the guide-expressing vector (a schematic of the system is reported in **Fig 3a**).

At fixed concentration of 5µM of doxycycline, we precisily defined specific intervals of time by which gRNA2 expression reach is maximum value, and accordingly, c-MYC *ncRNAs* transcription drop at minimum value (**Fig 3a,b**).

Exploiting "*click chemistry*", nascent DNA (nasDNA) was then used to quantify replication origin activation along *c-MYC* promoter and interestingly, the abundance of nasDNA strands was found strongly decreased in proximity of the origin of replication (~2.4 and 4.5 folds), but not to the other regions of the promoter (**Fig 3c**).

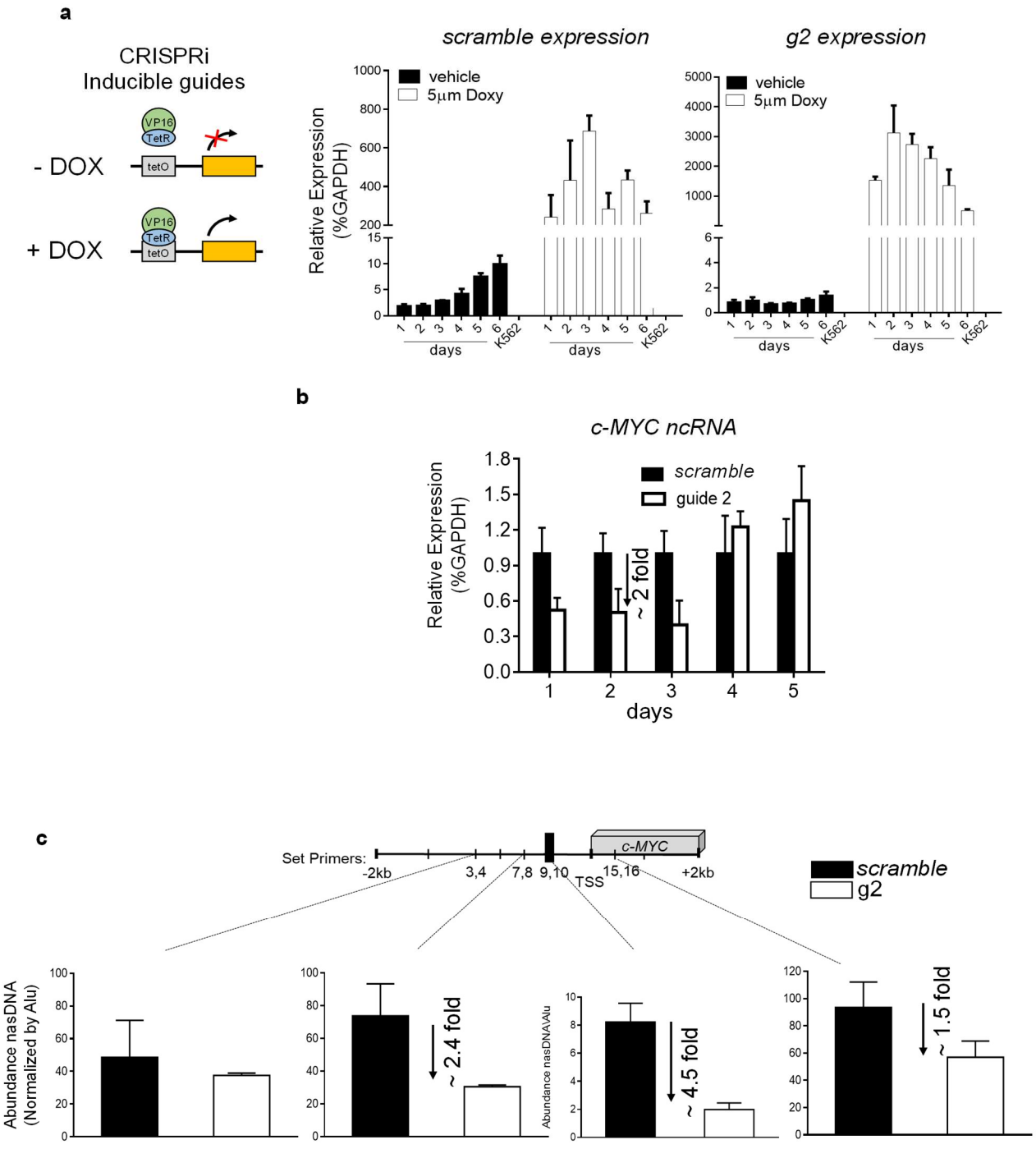


Figure 3. Local inhibition of c-MYC ncRNAs abrogates replication origin. a, Left panel: schematic representation of the inducible TetON promoter. **a, Right panel:** qPCR analysis of the guide expression (scramble and guide RNA 2) in different time points (1-6 days). The color of the bars indicate respectively vehicle (black bar) and 5μM doxycycline (white bar) added to the cell cultures **b,** c-MYC ncRNA expression of scramble (black bar), and guide RNA 2 (white bar) **c,** nasDNA quantification along c-MYC promoter for scramble (black bar), and guide RNA 2 (white bar)

Gain-of-Function Approach. In cells stably expressing dCas9-VP64 **(Fig. 4a-b)** we achieved a consistent up-regulation of *c-MYC* lncRNA (analyzed at different time points: from 1 to 5 weeks) **(Fig. 4a).**

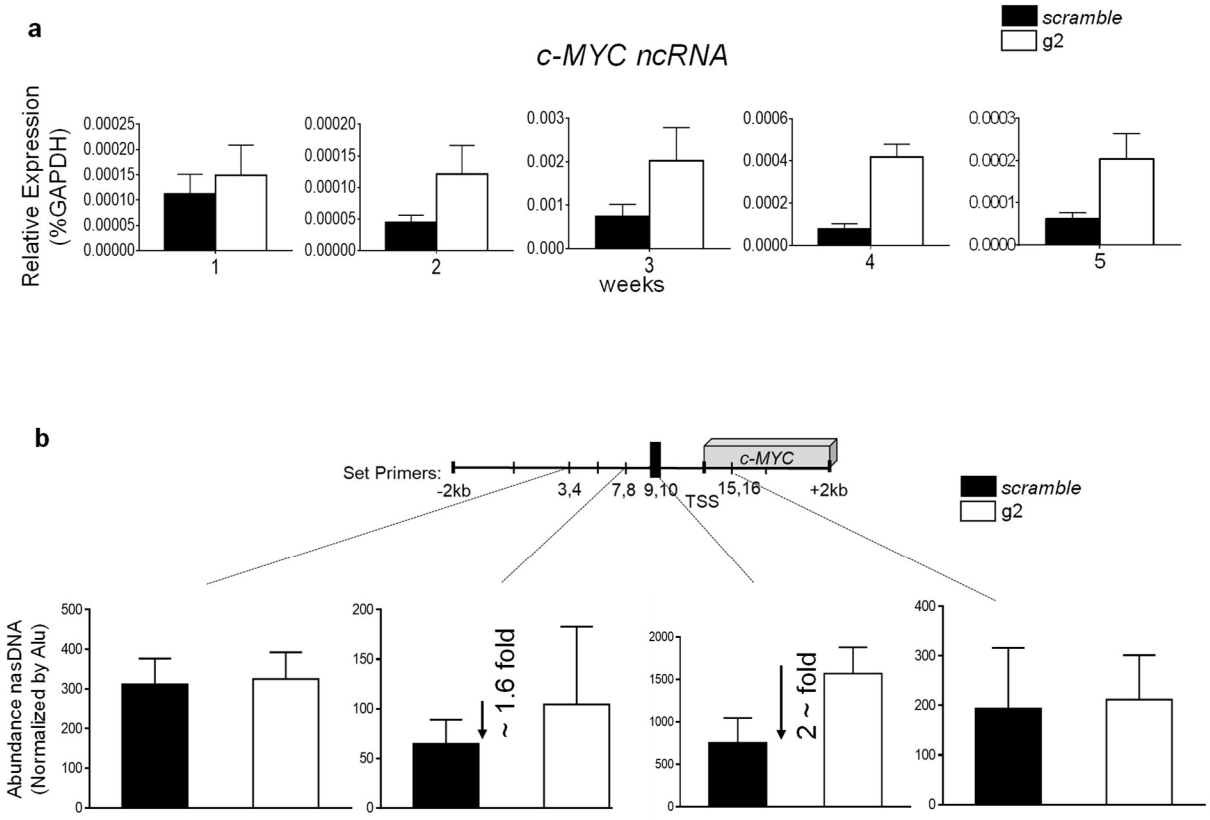


Figure 4. Induced c-MYC SPEARs effect replication origin. a, right qPCR analysis of c-MYC ncRNA expression scramble (black bar) and guide RNA 2 (white bar) in different time points (1-6 days). nasDNA quantification along c-MYC promoter for scramble (black bar), and guide RNA 2 (white bar)

The same time point was used to collect and isolate nasDNA. Analyses of nasDNA **Fig. 4b** confirmed the enrichment of newly synthesized DNA strands at the origin of replication, in comparison to the other regions of the promoter.

Collectively, these results will delineate the role of ncRNA in the formation and/or depletion of replication origins at selected genes.

Discussion

There are no available agents approved for the therapeutic correction of abnormal RT patterns; with particular relevance for diseases in which malfunctional of DNA replications happens as results by excessive proliferation likewise in myeloproliferative diseases. Indeed, the lack of knowledge in how the replication of specific genes is regulated and progressively advances, severely limit their clinical intervention. This study envisages the possibility to use RNA as gene-specific tool to tune timing of replication in leukemia. The ability to control in loco gene replication has crucial potential in clinical applications. In the short term this research can lead to the identification of novel key regulators of leukemogenesis and new targets for therapeutic treatments; in the long term pave the way for the development of RNA-based gene-specific replicating agents for cancer treatment. Additionally, identifying a newly *ncRNAs* "signature" that initiate DNA replication will enable: 1) further stratification of leukemias on the basis of the diverse RT observed; 2) refinement of classifications, particularly when integrated with other "signatures"; e.g. miRNA, histone modifications; 3) identification of new molecular and prognostic factors involved in leukemia development and progression; and 4) opens the possibility of correction replication-related diseases.

Author Contributions: DGT, ADR supervised the project. DGT, ADR, AKE and SU conceived and designed the study and wrote the manuscript; SU, AKE, YZ, BQT performed experiments; EM provided bioinformatic and statistical analysis of the data. BQT, EM and AKE provided

valuable suggestions about the project highlighting the weakness and providing possible solutions, and DGT, AKE critically reviewed the manuscript.

Funding: This work was supported by the award of NCI R35 CA197697, R01DK103858, W81XWH-15-1-0161, and P01HL131477-01A1 to DGT; and R50 CA211304 to AKE; the NCI R00 CA188595, the DOD award number W81XWH-20-1-0518, the Italian Association for Cancer Research (AIRC) Start-up Grant N.15347, the Giovanni Armenise-Harvard Foundation Career Development Award to ADR; the NCI K01CA222707 was awarded to BQT.

Conflicts of Interest:

The authors declare no conflict of interests.

References:

Akerman, I., B. Kasaai, A. Bazarova, P. B. Sang, I. Peiffer, M. Artufel, R. Derelle, G. Smith, M. Rodriguez-Martinez, M. Romano, S. Kinet, P. Tino, C. Theillet, N. Taylor, B. Ballester and M. Mechali (2020). "A predictable conserved DNA base composition signature defines human core DNA replication origins." Nat Commun **11**(1): 4826.

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

Ebralidze, A. K., *S. Ummarino;*, *M. A. Bassal;*, *H. Zhang;*, *B. Budnik;*, *E. Monteleone;*, *D. Kappei;*, *Y. V. Liu;*, *D. E.Tenen;*, *R. Coffey;*, *M. R. Sheen;*, *Y. Zhang;*, *A. Wanet;*, *B. Q. Trinh;*, *V. Poli;*, *V. E. Angarica;*, *R. T. Magallanes;*, *T. Benoukraf;*, *C. CraneRobinson;*, *A. D. Ruscio;* and *D. G. Tenen;* (**2021**). "Formation and recycling of an active epigenetic mark mediated by cell cycle-specific RNAs." Research square.

Gaillard, H., T. Garcia-Muse and A. Aguilera (2015). "Replication stress and cancer." Nat Rev Cancer **15**(5): 276-289.

Giacca, M., L. Zentilin, P. Norio, S. Diviacco, D. Dimitrova, G. Contreas, G. Biamonti, G. Perini, F. Weighardt, S. Riva and et al. (1994). "Fine mapping of a replication origin of human DNA." Proc Natl Acad Sci U S A **91**(15): 7119-7123.

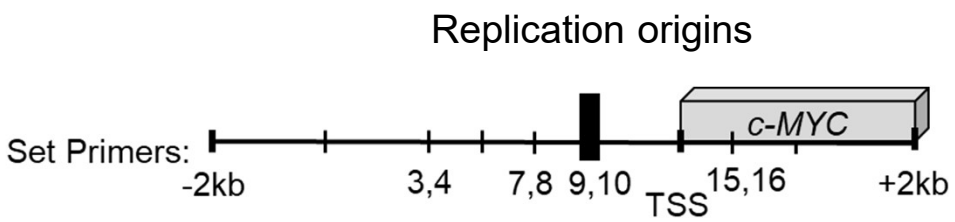
Jaremko, M. J., K. F. On, D. R. Thomas, B. Stillman and L. Joshua-Tor (2020). "The dynamic nature of the human origin recognition complex revealed through five cryoEM structures." Elife **9**.
Keller, C., E. M. Ladenburger, M. Kremer and R. Knippers (2002). "The origin recognition complex marks a replication origin in the human TOP1 gene promoter." J Biol Chem **277**(35): 31430-31440.
Larson, M. H., L. A. Gilbert, X. Wang, W. A. Lim, J. S. Weissman and L. S. Qi (2013). "CRISPR interference (CRISPRi) for sequence-specific control of gene expression." Nat Protoc **8**(11): 2180-2196.
Leman, A. R., C. Noguchi, C. Y. Lee and E. Noguchi (2010). "Human Timeless and Tipin stabilize replication forks and facilitate sister-chromatid cohesion." J Cell Sci **123**(Pt 5): 660-670.
Long, H., L. Zhang, M. Lv, Z. Wen, W. Zhang, X. Chen, P. Zhang, T. Li, L. Chang, C. Jin, G. Wu, X. Wang, F. Yang, J. Pei, P. Chen, R. Margueron, H. Deng, M. Zhu and G. Li (2020). "Author Correction: H2A.Z facilitates licensing and activation of early replication origins." Nature **578**(7793): E8.
Macheret, M. and T. D. Halazonetis (2018). "Intragenic origins due to short G1 phases underlie oncogene-induced DNA replication stress." Nature **555**(7694): 112-116.
Macheret, M. and T. D. Halazonetis (2019). "Monitoring early S-phase origin firing and replication fork movement by sequencing nascent DNA from synchronized cells." Nat Protoc **14**(1): 51-67.
Marchal, C., J. Sima and D. M. Gilbert (2019). "Control of DNA replication timing in the 3D genome." Nat Rev Mol Cell Biol **20**(12): 721-737.
Miotto, B., Z. Ji and K. Struhl (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**(33): E4810-4819.
Perez-Pinera, P., D. D. Kocak, C. M. Vockley, A. F. Adler, A. M. Kabadi, L. R. Polstein, P. I. Thakore, K. A. Glass, D. G. Ousterout, K. W. Leong, F. Guilak, G. E. Crawford, T. E. Reddy and C. A. Gersbach (2013). "RNA-guided gene activation by CRISPR-Cas9-based transcription factors." Nat Methods **10**(10): 973-976.
Rivera-Mulia, J. C., T. Sasaki, C. Trevilla-Garcia, N. Nakamichi, D. Knapp, C. A. Hammond, B. H. Chang, J. W. Tyner, M. Devidas, J. Zimmerman, K. N. Klein, V. Somasundaram, B. J. Druker, T. A. Gruber, A. Koren, C. J. Eaves and D. M. Gilbert (2019). "Replication timing alterations in leukemia affect clinically relevant chromosome domains." Blood Adv **3**(21): 3201-3213.
Sawyers, C. L. (1999). "Chronic myeloid leukemia." N Engl J Med **340**(17): 1330-1340.
Tonini, G. P., D. Radzioch, A. Gronberg, M. Clayton, E. Blasi, G. Benetton and L. Varesio (1987). "Erythroid differentiation and modulation of c-myc expression induced by antineoplastic drugs in the human leukemic cell line K562." Cancer Res **47**(17): 4544-4547.

S-phase induced RNAs initiate formation of DNA replication origins

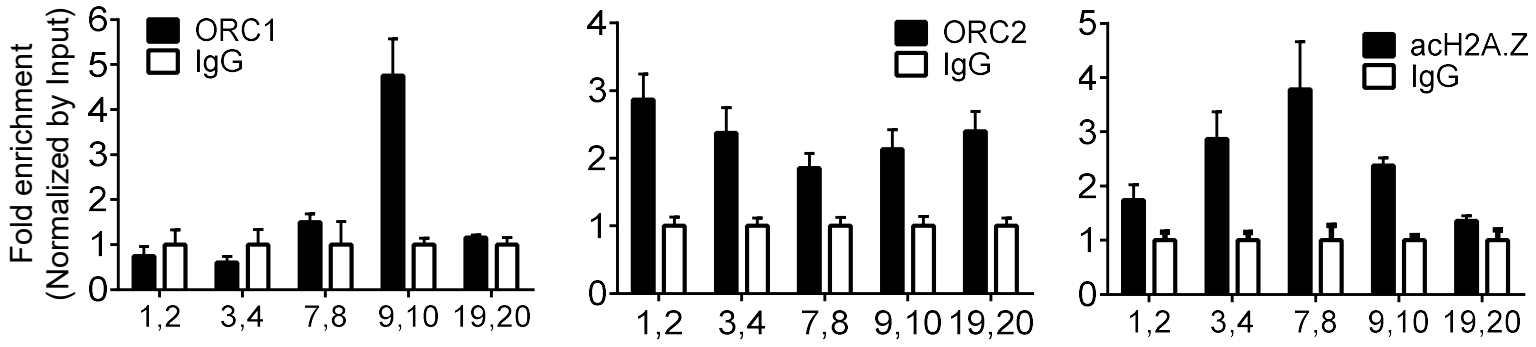
Figures

Identification and characterization of DNA replication origins

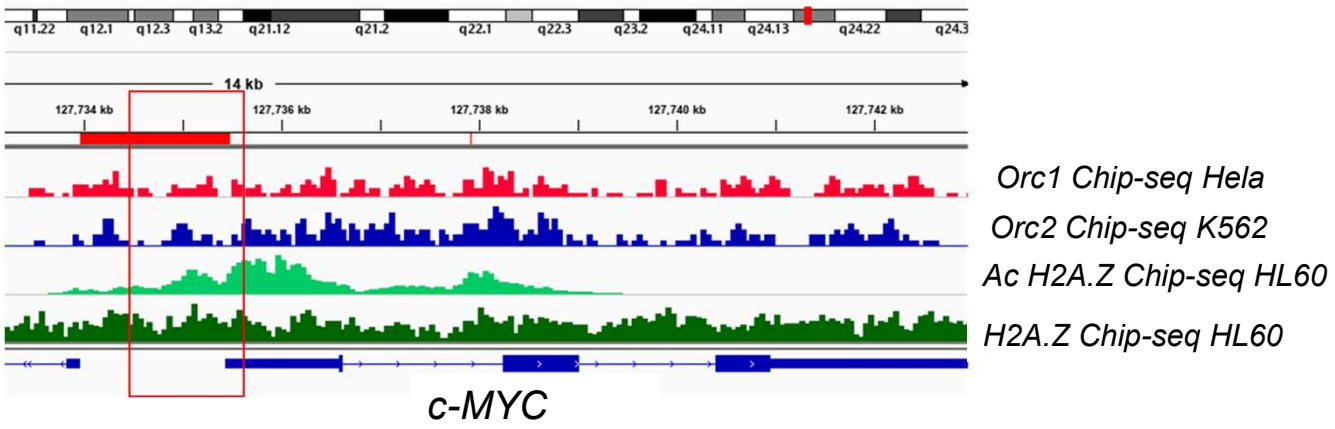
a



b

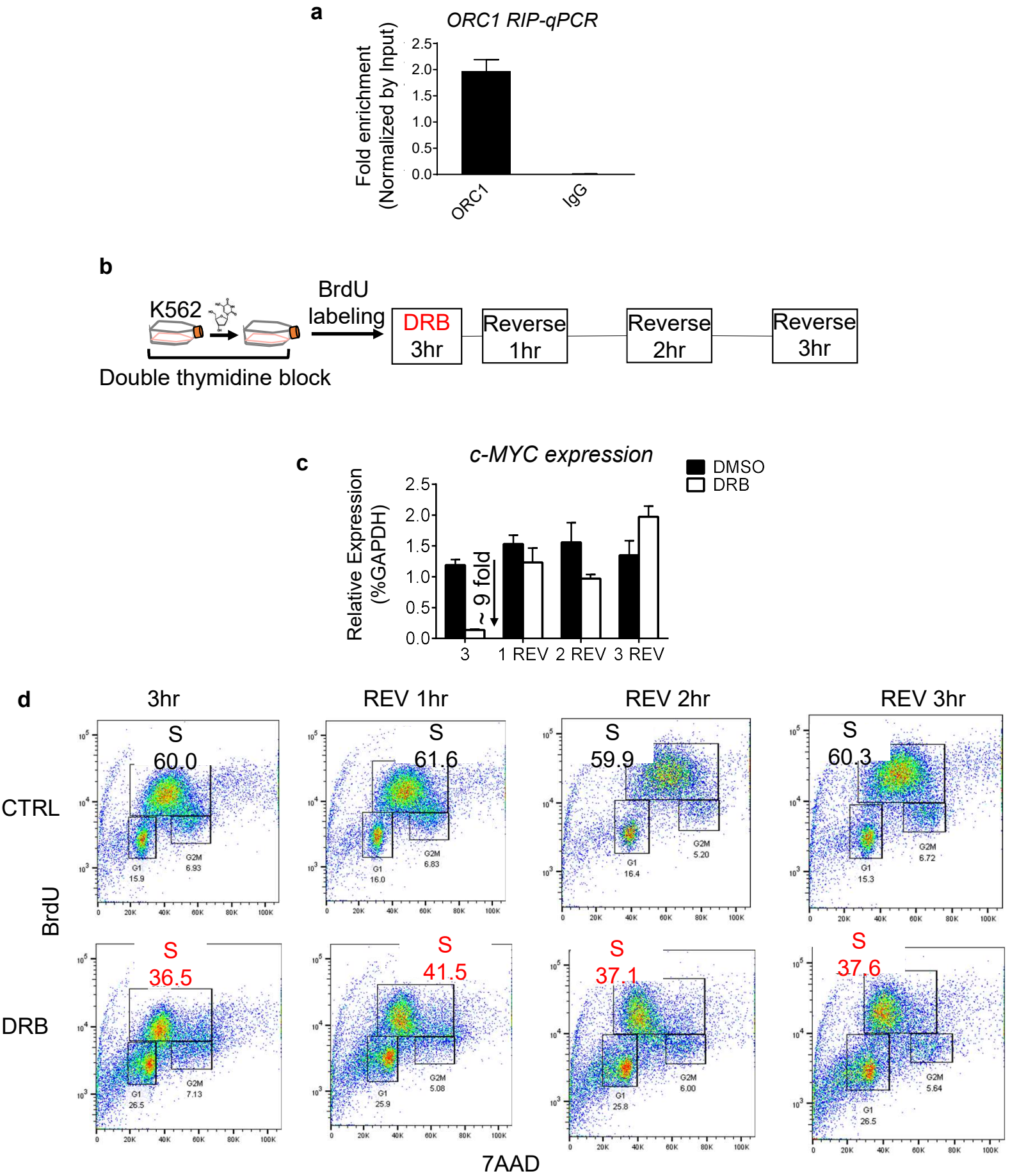


c



Ummarino *et al*_ Fig. 1

Global inhibition of SPEARs perturbs the replication firing

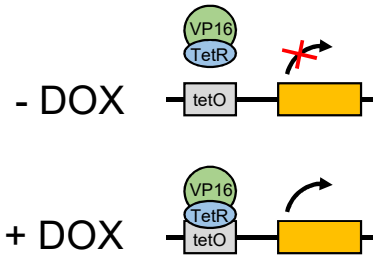


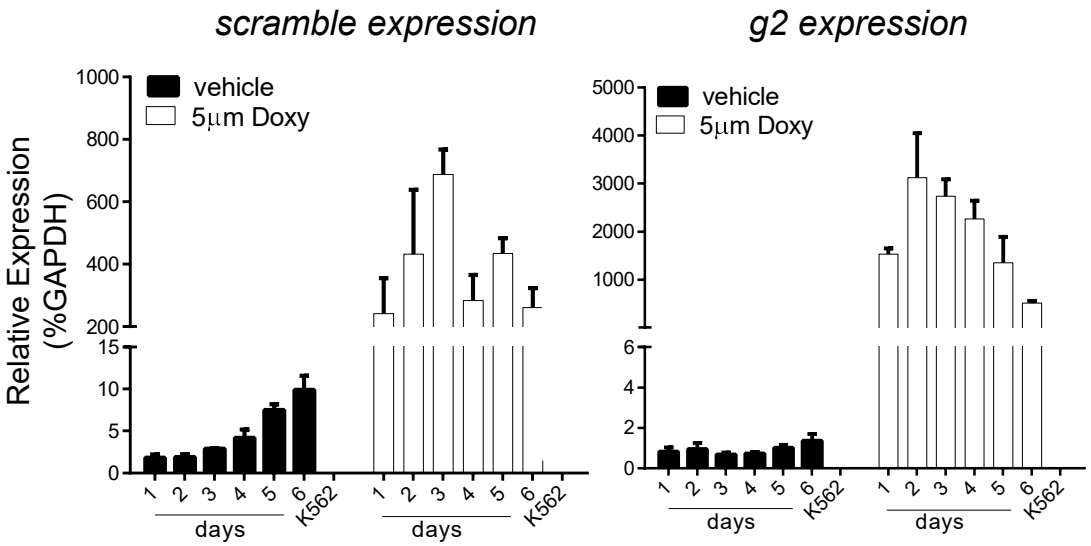
Ummarino *et al*_ Fig. 2

Local inhibition of c-MYC SPEARs abrogates replication origin

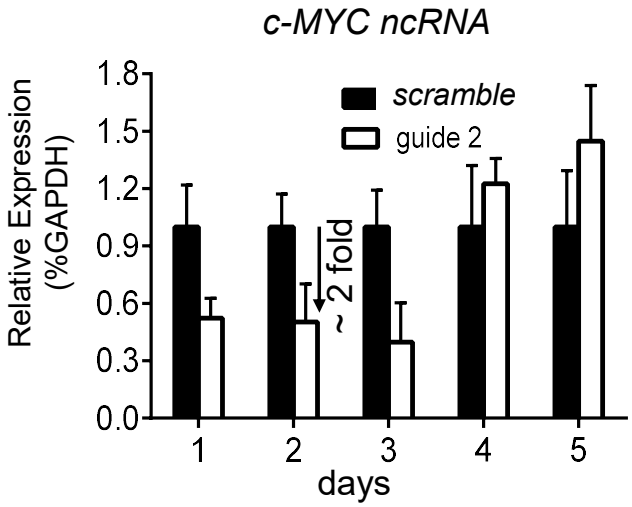
a

CRISPRi
Inducible guides

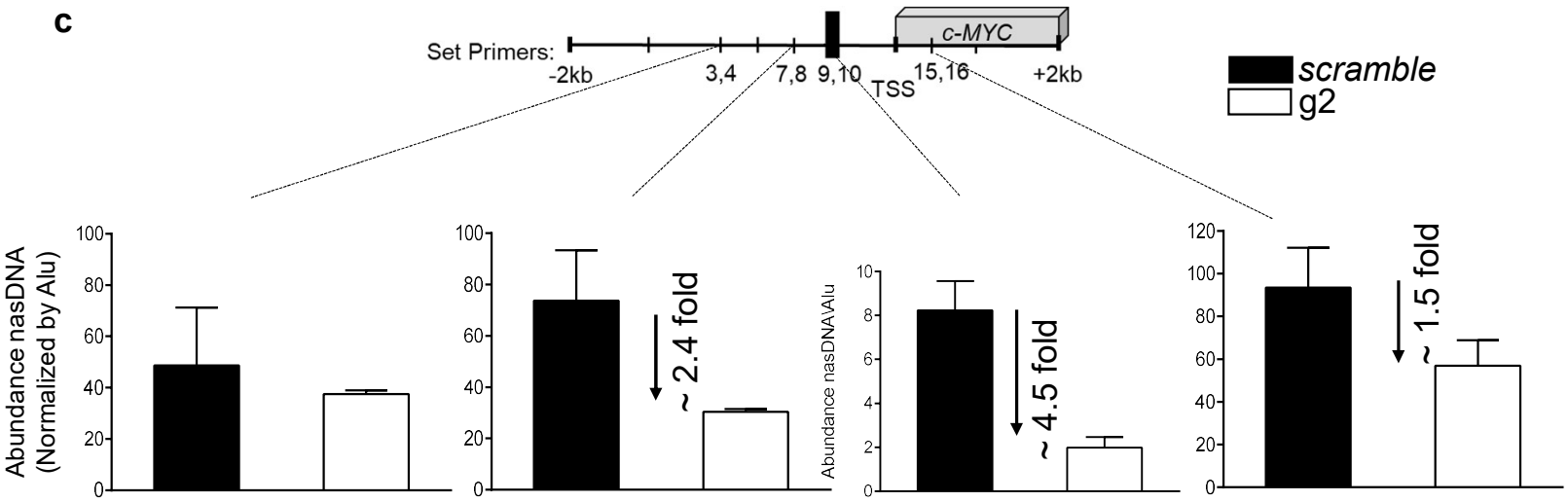




b

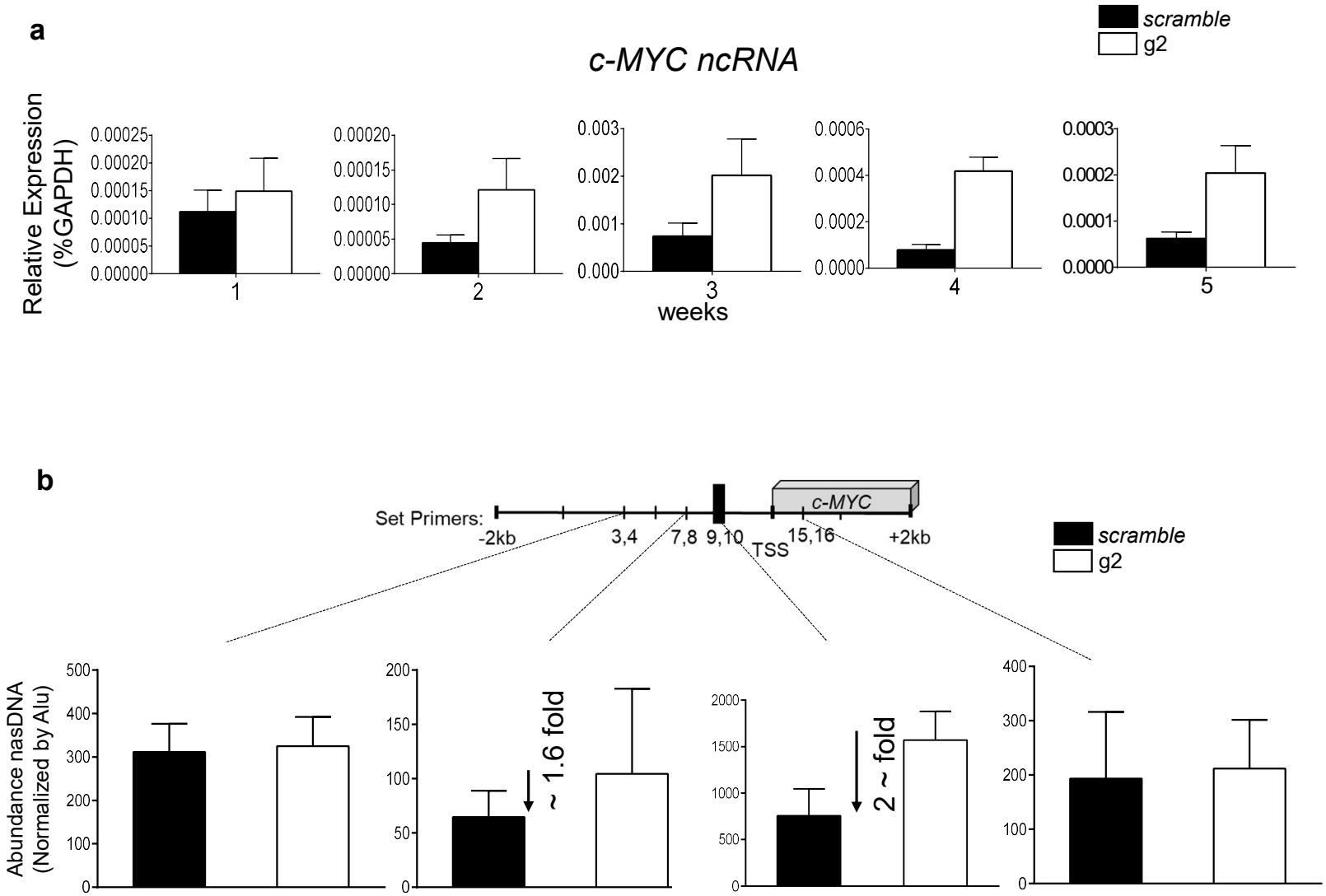


c



Ummarino *et al*_ Fig. 3

Induced of c-MYC SPEARs affects replication origin



Ummarino *et al*_ Fig. 4