UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Università degli Studi di Padova

Dipartimento di Biologia

CORSO DI DOTTORATO DI RICERCA IN BIOSCIENZE E BIOTECNOLOGIE

CURRICOLO GENETICA E BIOLOGIA MOLECOLARE DELLO SVILUPPO

CICLO XXIX

THE FINE MODULATION OF MAMMALIAN DNA REPLICATION IN RESPONSE TO ENDOGENOUS AND EXOGENOUS STRESS CONDITIONS

Coordinatore: Ch.mo Prof. Paolo Bernardi

Supervisore: Ch.mo Prof. Antonella Russo

Dottoranda: Martina Stevanoni

TABLE OF CONTENTS

RIASSUNTO

La replicazione del DNA è essenziale per consentire l'accurata trasmissione del materiale genetico durante le successive divisioni cellulari. Nelle cellule di mammifero i siti di inizio sono raggruppati in regioni genomiche di 200-400 kb, definite *cluster* replicativi, che sono a loro volta racchiusi in più ampi domini di replicazione (Méchali 2010; Cayrou et al. 2011). Questa organizzazione gerarchica consente di controllare il processo replicativo nello spazio e nel tempo tramite la regolazione dell'attivazione delle origini sia a livello locale all'interno dei *cluster* che a livello globale nei domini di replicazione (Yekezare et al. 2013).

Durante la fase G_1 del ciclo cellulare le origini vengono "licenziate", ma solo una parte di esse viene poi attivata nella successiva fase S (Ge et al. 2007; Blow & Ge 2009; Méchali 2010). Inoltre, le origini si attivano in modo stocastico all'interno dei *cluster* e sono diversamente distribuite nel genoma tra le diverse cellule (Hyrien et al. 2003; Gilbert 2007; Méchali 2010). Pertanto, le origini di replicazione nei mammiferi sono ridondanti e flessibili (Hyrien et al. 2003; Méchali 2010), caratteristiche indispensabili per fronteggiare non solo condizioni di stress replicativo, ma anche cambiamenti dell'organizzazione cromatinica quali quelli che avvengono durante lo sviluppo e il differenziamento cellulare (Cortez 2015; Alver et al. 2014; Palumbo et al. 2013; Zeman & Cimprich 2014). La regolazione temporale dell'attivazione delle origini è stabilita a livello dei domini replicativi, conservati in successivi cicli cellulari e classificati come regioni a replicazione precoce, intermedia e tardiva in base al loro *timing* di attivazione durante la fase S (Hiratani et al. 2008; Pope & Gilbert 2013; Rivera-Mulia & Gilbert 2016b).

I profili replicativi determinano il programma di replicazione specifico di ogni tipo cellulare e sono regolati durante le diverse fasi dello sviluppo e del differenziamento, così come in risposta a stress (Palumbo et al. 2013; Courbet et al. 2008; Anglana et al. 2003). I profili di replicazione sono definiti da vari parametri: oltre alla posizione e alla scelta delle origini, anche la velocità di progressione delle forche e i diversi pattern replicativi alternativi, quali forche unidirezionali, asincrone e in pausa, sono finemente regolati. Per la corretta interpretazione dei complessi fenomeni associati alla replicazione del DNA tutti questi pattern devono essere attentamente considerati (Prioleau & MacAlpine 2016; Hyrien 2015; Palumbo et al. 2013). Pertanto, per riuscire ad apprezzare la plasticità e la variabilità intrinseche del processo replicativo, le tecniche di analisi su singola molecola risultano essere particolarmente appropriate (Tuduri et al. 2010; Técher et al. 2016; Prioleau & MacAlpine 2016).

In questo lavoro ho valutato come vengono modulati i profili di replicazione in presenza di stress replicativi endogeni ed esogeni. In particolare, utilizzando diversi modelli cellulari ho analizzato come venga regolata la replicazione del DNA

in regioni genomiche instabili, in particolari condizioni di crescita e durante il differenziamento.
In primo luogo, ho determinato gli effetti associati alla presenza di sequenze trinucleotidiche ripetute. Le triplette ripetute sono tra le regione del genoma più instabili e variazioni della loro lunghezza sono coinvolte in diverse malattie neurodegenerative (McMurray 2010; Lee & McMurray 2014). La loro caratteristica principale, strettamente collegata alla loro instabilità, è la capacità di formare strutture secondarie insolite (Mirkin & Mirkin 2007; Krasilnikova & Mirkin 2004), che inducono stress replicativo impedendo la normale progressione delle forche (Zeman & Cimprich 2014; Magdalou et al. 2014; León-Ortiz et al. 2014). Si pensa che tra i vari processi metabolici associati al DNA la replicazione abbia un ruolo fondamentale nel promuovere l'espansione e l'instabilità delle triplette ripetute, tra cui anche la ripetizione GAA associata all'atassia di Friedreich (Pearson et al. 2005; Cleary & Pearson 2005). Per capire meglio il meccanismo che sta alla base della loro instabilità sono stati utilizzati diversi modelli sperimentali, dal lievito a cellule umane ingegnerizzate (Follonier et al. 2013; Chandok et al. 2012; Kim et al. 2008). Tuttavia, rispetto ai lunghi tratti ripetuti osservati in cellule di pazienti, il limitato numero di triplette presente in questi modelli non consente uno studio approfondito degli effetti sulla sintesi del DNA. Per questo motivo, ho analizzato i profili replicativi in cellule linfoblastoidi umane derivate da pazienti affetti da atassia di Friedreich e che presentano in omozigosi un'espansione della tripletta GAA nel primo introne del gene *Fratassina*. In presenza dell'espansione ho osservato alterazioni del *timing* di replicazione del gene tramite la tecnica della FISH su nuclei in interfase, identificando vari cambiamenti nei profili replicativi della regione analizzata. Nello specifico, in base alla proporzione dei segnali di ibridazione duplicati ho potuto definire un ritardo nella replicazione del gene mutato durante la prima metà della fase S. Questo ritardo viene però recuperato nella parte finale della fase S, in quanto la replicazione del gene viene completata sia negli alleli normali che in quelli con tripletta espansa. Grazie alla tecnica del *molecular combing* ho poi monitorato le dinamiche di replicazione di una regione genomica di 850 kb contenente il gene *Fratassina*. Dalle mie analisi è emerso che l'effetto principale associato alla presenza della mutazione è l'attivazione di origini dormienti aggiuntive, situate a valle della tripletta ripetuta e all'interno del gene. Esse possono essere considerate come un meccanismo di recupero per assicurare il completamento della replicazione negli alleli mutati. In conseguenza dell'attivazione di queste origini dormienti, la direzione con cui la tripletta espansa viene replicata cambia, dimostrando che il modello di *origin-switch*, proposto per descrivere i meccanismi di instabilità delle triplette ripetute, è conforme al caso dell'espansione GAA nel gene *Fratassina*, come precedentemente osservato anche nel locus *FMR1* (Gerhardt et al. 2014). In maniera indipendente un altro gruppo di ricerca è giunto alle mie stesse conclusioni (Gerhardt et al. 2016). Nel corso di queste analisi è emersa una significativa diminuzione della lunghezza delle forche unidirezionali negli alleli espansi, che può essere considerata come un secondo

effetto associato alla mutazione nel gene *Fratassina*. Similmente a quanto osservato nel locus *FMR1* (Gerhardt et al. 2014), in questo lavoro sono stati identificati eventi di arresto/pausa delle forche in corrispondenza della corta ripetizione GAA in una delle due linee di controllo analizzate, osservazione che suggerisce un possibile ruolo del tratto ripetuto non patologico sulla progressione delle forche. Recentemente è stato dimostrato che la sequenza GAA espansa è un sito di blocco delle forche replicative, ed è stato ipotizzato che gli eventi di arresto siano dovuti alla collisione tra i complessi proteici coinvolti nella replicazione e nella trascrizione del DNA (Gerhardt et al. 2016). Indagini future devono mirare a capire se il meccanismo di *origin-switch* osservato in presenza dell'espansione GAA sia la causa dell'espansione stessa o ne sia una conseguenza.

Un secondo aspetto che ho considerato riguarda la modulazione del programma replicativo in seguito a innalzamento e sbilanciamento dei pool dei nucleotidi. Sintesi, degradazione e consumo dei precursori del DNA devono essere accuratamente controllati per assicurare la replicazione completa del genoma e per evitare instabilità (Rampazzo et al. 2010; Chabes & Stillman 2007; Chabosseau et al. 2011; Bester et al. 2011). In presenza di quantità limitate di nucleotidi le forche replicative rallentano drasticamente e inducono l'attivazione della risposta al danno del DNA, sottoponendo le cellule a stress replicativo (Anglana et al. 2003; Courbet et al. 2008). Al contrario, le conseguenze di un elevato supporto di nucleotidi sulle dinamiche di replicazione sono state osservate solo in lievito, dove in seguito a sovraespressione dell'enzima ribonucleotide reduttasi è stato dimostrato un aumento della velocità di progressione delle forche (Poli et al. 2012). Gli effetti in cellule di mammifero sono invece ancora sconosciuti. Per questo motivo ho utilizzato come modelli sperimentali fibroblasti primari umani e monociti della linea THP1 con pool nucleotidici alti e sbilanciati a causa della mancanza della proteina SAMHD1 (Franzolin et al. 2013; Miazzi et al. 2014). Con la tecnica del *molecular combing* mi è stato possibile studiare i profili di replicazione a livello di intero genoma. Ho dimostrato che lo sbilanciamento dei pool dei nucleotidi non influisce sulle dinamiche di replicazione dei fibroblasti primari, né in condizioni fisiologiche né sotto stress replicativo. Al contrario, i monociti THP1 privi di SAMHD1 mostrano un inatteso rallentamento delle forche replicative con conseguente aumento delle origini attivate in ogni *cluster*. La diversa risposta cellulare osservata in seguito allo sbilanciamento dei pool può essere spiegata da un ruolo specifico di SAMHD1 in funzione del tipo cellulare. In alternativa, è possibile che i fibroblasti abbiano sviluppato un fenotipo adattativo per compensare la mancanza della proteina. Analisi future avranno come scopo quello di verificare la validità di queste due ipotesi e di valutare possibili effetti associati a blocco e ripartenza delle forche replicative.

Infine, ho valutato come cambiano i profili di replicazione durante il differenziamento. È noto che *timing* e dinamiche di replicazione sono specifici del tipo cellulare e sono regolati in base a cambiamenti dell'organizzazione cromatinica che avvengono durante lo sviluppo e il differenziamento (Palumbo et al. 2013;

Hiratani et al. 2010; Hiratani et al. 2008). Quindi, ho valutato se la riattivazione forzata del ciclo cellulare in cellule terminalmente differenziate influenzi la replicazione del DNA. Come modello sperimentale ho usato miotubi di topo forzati a rientrare nel ciclo cellulare e li ho analizzati con *molecular combing* rispetto a mioblasti proliferanti. In seguito a riattivazione forzata, ho osservato una significativa riduzione della velocità delle forche replicative, similmente a quanto visto in condizioni di stress replicativo (Anglana et al. 2003; Palumbo et al. 2010). Questo risultato è in accordo con il fallimento del processo replicativo osservato nei miotubi riattivati e si pensa sia dovuto a un'incapacità di queste cellule di espandere in modo corretto i pool dei nucleotidi (Pajalunga et al. 2010; Pajalunga et al. 2017). In effetti, dopo l'aggiunta dei precursori del DNA nei miotubi riattivati si osserva un parziale miglioramento della velocità delle forche. Nonostante ciò però, il numero di origini attivate in ogni *cluster* è simile tra miotubi e mioblasti, perciò è ragionevole pensare che regioni del genoma restino non replicate. In conclusione, l'incapacità di completare la replicazione del DNA osservata nei miotubi dopo riattivazione forzata del ciclo cellulare non solo è dovuta alla limitata disponibilità di nucleotidi, ma anche a un difetto nel reclutare origini di replicazione aggiuntive per compensare la riduzione delle velocità. Un secondo aspetto interessante è l'aumentata proporzione di forche unidirezionali nei miotubi riattivati rispetto ai mioblasti, simile a quanto precedentemente osservato in fibroblasti primari umani (Palumbo et al. 2013). In questo caso, le forche unidirezionali visualizzate possono essere considerate come un residuo della modalità con cui alcuni domini replicano quando le cellule vanno verso il differenziamento terminale. Resta interessante capire se anche il *timing* della replicazione sia alterato dopo riattivazione forzata del ciclo cellulare.

La replicazione del DNA è un processo molto complesso caratterizzato da una grande plasticità, pertanto non è consigliato trarre conclusioni generali da studi basati su singoli tipi cellulari. I risultati ottenuti in questo lavoro contribuiscono a interpretare la grande complessità del processo replicativo, offrendo anche nuove prospettive per approfondimenti futuri.

SUMMARY

DNA replication is essential to allow faithful inheritance of the genome. In mammalian cells, many replication origins are grouped within 200-400 kb regions called replication clusters, which are in turn enclosed in large replication domains (Méchali 2010; Cayrou et al. 2011). This hierarchical organisation is required for the temporal and spatial control of DNA replication and it allows modulating origin activation locally within clusters and globally at the level of replication domains (Yekezare et al. 2013).

During G_1 several initiation sites are licensed, but only a subset is activated in the following S-phase (Ge et al. 2007; Blow & Ge 2009; Méchali 2010). Furthermore, origins fire stochastically within clusters and in different positions among cells in a population (Hyrien et al. 2003; Gilbert 2007; Méchali 2010). Remarkably, origin redundancy and plasticity are intrinsic features of the mammalian replication process necessary to face changes in chromatin organisation occurring during development and cell differentiation and to overcome replication stress (Cortez 2015; Alver et al. 2014; Palumbo et al. 2013; Zeman & Cimprich 2014). The temporal regulation of origin firing is established at the level of replication domains, which are conserved among subsequent cell cycles and classified as early, mid or late replicating according to their timing of activation throughout the S-phase (Hiratani et al. 2008; Pope & Gilbert 2013; Rivera-Mulia & Gilbert 2016b).

Replication profiles define the replication program in each cell type and are modulated according to the diverse developmental and differentiation cellular stages, as well as in response to replication stress (Palumbo et al. 2013; Courbet et al. 2008; Anglana et al. 2003). Replication profiles are determined by several parameters, including fork rates, origin choice and alternative replication patterns (i.e. unidirectional and asynchronous forks and pause/arrest events), and their regulation is implied by the flexible nature of the mammalian replication process. Hence, not only origin position but the whole replication dynamics must be taken into consideration to better elucidate the complex phenomena associated with mammalian DNA replication (Prioleau & MacAlpine 2016; Hyrien 2015; Palumbo et al. 2013). In this frame, single-molecule techniques represent the most appropriate tool to detect the intrinsic plasticity and cell-to-cell variability of mammalian genomes (Tuduri et al. 2010; Técher et al. 2016; Prioleau & MacAlpine 2016).

Here, I evaluated how replication profiles are regulated in response to endogenous and exogenous replication stress conditions. In particular, I considered the modulation of the mammalian replication program in relation to sequence organisation, cell growth and differentiation. For this purpose, different cellular models were used.

First, I focused on the effects of a trinucleotide repeat (TNR) expansion on DNA replication. TNR are among the most unstable genomic regions and variations of

their length are implicated in many human neurodegenerative disorders (McMurray 2010; Lee & McMurray 2014). Expanded repeats are prone to form stable unusual DNA secondary structures (Mirkin & Mirkin 2007; Krasilnikova & Mirkin 2004), which are a well-known source of replication stress (Zeman & Cimprich 2014; Magdalou et al. 2014; León-Ortiz et al. 2014). A replication-based mechanism is largely accepted to be at the origin of expansion and instability of several TNR, including also the GAA-repeat expansion responsible for Friedreich's ataxia (FRDA). To better understand this mechanism, different experimental systems from yeast to transfected or engineered human cells were employed (Follonier et al. 2013; Chandok et al. 2012; Kim et al. 2008), but none of them displayed the huge amount of repeats observed in patients' cells. Thus, to understand how replication profiles are modulated in the presence of long repetitive sequences in the endogenous context, I used human lymphoblastoid cell lines derived from FRDA patients carrying an homozygous GAA-repeat expansion within intron 1 of the *Frataxin* (*FXN*) gene. In the presence of the GAA-repeat expansion an alteration of the replication timing was observed by interphase FISH and wide changes in the replication profiles were demonstrated. Indeed, in mutant cells, according to the frequency of duplicated FISH spots, the replication of the *FXN* gene was slowed or delayed during the first half of the S-phase. This effect was normalised in the latter part when both normal and expanded alleles complete their replication. Further, by molecular combing replication dynamics was monitored in a large genomic region harbouring *FXN*. By this approach I verified that the most relevant effect associated with the presence of the GAA-repeat expansion was the recruitment of additional dormant origins firing downstream the repeat, which can be considered as a rescue mechanism to assure replication of mutant alleles. As a consequence of dormant origin activation, a switch of the prevalent direction by which the expanded GAA-repeat is replicated was observed, indicating that the origin-switch model for TNR instability conform to the case of the GAA-repeat expansion, similarly to what observed at the *FMR1* locus (Gerhardt et al. 2014). Same conclusions were reached in parallel in another study (Gerhardt et al. 2016). Remarkably, a strong reduction of unidirectional fork length was observed in mutant alleles with respect to the normal sequence and this may be considered as a second effect of the GAA-repeat expansion. In line with results obtained at the *FMR1* locus (Gerhardt et al. 2014), in the present study recurrent paused/arrested forks were recorded in proximity of the short GAA-repeat in one of the normal cell lines analysed, suggesting a possible impact of the non-pathological GAA-repeat on fork progression. Interestingly, it was recently demonstrated an high occurrence of fork pausing at the expanded GAA-repeat in both undifferentiated and differentiated FRDA cells and these events were hypothesised to be caused by a collision between the replication and transcription machineries (Gerhardt et al. 2016). Understanding whether the origin-switch is the cause or instead a consequence of the GAA-repeat expansion will be necessary and future investigations will primarily contribute to clarify this aspect.

Secondly, I evaluated the effects of increased and unbalanced dNTP pools on the replication program. To ensure completion of DNA replication and to avoid genome instability, the balance among dNTP supply, degradation and consumption must be tightly controlled according to the rate of DNA synthesis (Rampazzo et al. 2010; Chabes & Stillman 2007; Chabosseau et al. 2011; Bester et al. 2011). Indeed, the limiting availability of nucleotide precursors has been widely demonstrated to induce replication stress by slowing fork progression and leading to activation of the DNA damage response (Anglana et al. 2003; Courbet et al. 2008). Conversely, the consequences of an increased supply of dNTPs were only partially depicted in yeast mutants, where replication forks move faster upon ribonucleotide reductase overexpression (Poli et al. 2012). Instead, the effects in mammalian cells are still unknown. Thus, here I used human primary fibroblasts and THP1 monocytes with increased and unbalanced dNTP pools due to depletion of SAMHD1, a protein involved in nucleotide metabolism (Franzolin et al. 2013; Miazzi et al. 2014). By this experimental models the replication profiles were analysed genome-wide by molecular combing in comparison to normal cells. My results indicated that independently of the dNTP pool imbalances, DNA replication proceeds mostly undisturbed in mutated and control fibroblasts, both in physiological and perturbed growth conditions. In contrast, an unexpected slow down of replication forks and a consequent increase in origin firing were detected in SAMHD1-depleted THP1 cells with respect to the wildtype cell line. This differential response to the high availability of nucleotide precursors may be ascribed either to a supposed cell type-specific activity of SAMHD1 or to the development of an adaptive phenotype in mutant fibroblasts compensating for the SAMHD1 depletion. Future goals will be to confirm these hypotheses and to evaluate the effects associated with fork stalling and restart.

Finally, I considered how replication dynamics are modulated during cell differentiation. It is well known that replication timing and replication profiles are cell type-specific and regulated according to changes in chromatin organisation occurring during development and cell differentiation (Palumbo et al. 2013; Hiratani et al. 2010; Hiratani et al. 2008). Hence, I assessed whether DNA replication is affected upon forced cell cycle reactivation. In particular, terminally differentiated mouse myotubes forced to re-enter the cell cycle were analysed by molecular combing in comparison to proliferating myoblasts. A significant reduction of fork rates was detected in reactivated myotubes with respect to proliferating muscle cells, resembling the effects seen under replication stress. This result is in line with previous evidence indicating that myotubes fail to complete their replication, as they are not able to properly expand their dNTP pools (Pajalunga et al. 2010; Pajalunga et al. 2017). Accordingly, after addition of nucleotide precursors the reduction of fork rate was partially ameliorated. However, the number of activated origins in each replication cluster was comparable between myotubes dosed with deoxynucleosides and myoblasts, suggesting that some regions along the genome are left under-replicated. Thus, the

replication failure detected upon forced cell cycle re-entry may be ascribed not only to the depletion of nucleotide precursors, but also to the inability of these cells to recruit additional origins compensating for the reduction of fork speed. Interestingly, I found an increased proportion of unidirectional forks in reactivated myotubes when compared to myoblasts, in line with previous data obtained in human primary fibroblasts (Palumbo et al. 2013). Thus, unidirectional forks could be viewed as a remnant of the modality by which some replication domains are replicated when cells move toward terminal differentiation. In this perspective, understanding whether replication timing may be implicated in the replication impairment observed upon forced cell cycle reactivation remains an intriguing issue to be further unravelled.

The complexity of mammalian DNA replication may be ascribed to the intrinsic plasticity of the process, and drawing general conclusions from studies based on individual cellular models may be incautious. The results obtained in this study add new evidence for interpret this complexity, and offer insights for future investigations.

INTRODUCTION

DNA replication is among the main cellular processes. Organisms have evolved different mechanisms to ensure the completion of DNA replication according to the length of their genome and of the cell cycle. Preliminary studies in *E. coli* indicate that DNA replication starts at a specific initiation site called replication origin, whose sequence is recognised by a replicator element responsible for the recruitment of all components of the replication machinery. Then, replication forks move bidirectionally along the entire bacterial circular chromosome till they meet at the termination site (O'Donnell 2006; Leonard & Méchali 2013; Hyrien et al. 2013; Evertts & Coller 2012). When dealing with the larger genomes of Eukaryotes, a much complex scenario emerges, as cells have to complete replication in a narrow temporal window of their cell cycle, the S-phase. For this reason, eukaryotic cells duplicate their genome from many replication origins differently distributed and activated in a time-specific manner (Evertts & Coller 2012; Méchali 2010; Fragkos et al. 2015; Rivera-Mulia & Gilbert 2016a).

The tight temporal and spatial control of the replication process is necessary to avoid under- and over-replication (McIntosh & Blow 2012; Zeman & Cimprich 2014; Alver et al. 2014; Magdalou et al. 2014; Berti & Vindigni 2016). Origins are activated by a two-step mechanism (Remus & Diffley 2009; Yekezare et al. 2013): first they are "licensed" during late mitosis and G_1 (Remus & Diffley 2009; Méchali 2010) and then they are activated throughout the S phase. Origin licensing involves the binding of the Origin Recognition Complex (ORC) at initiation sites and the consequent recruitment and assembly of the replicative helicase MCM2-7 mediated by various licensing factors (Remus & Diffley 2009; DePamphilis et al. 2006; Lutzmann et al. 2008; Ilves et al. 2010). Once this pre-replicative complex (pre-RC) is stably bound to DNA, two S-phase kinases (CDK and DDK) are recruited by different firing factors for its activation (origin firing). Many studies demonstrate that there are much more MCM2-7 bound to DNA than the amount of ORC in eukaryotic cells, suggesting that one single ORC is able to assemble more pre-RCs (Hyrien et al. 2003; Ticau et al. 2015). Although recent single-cell and single-molecule studies identify the minimum set of proteins required for origin recognition, selection and activation (Ticau et al. 2015; Duzdevich et al. 2015; Yeeles et al. 2015; Deegan & Diffley 2016), many aspects concerning the dynamics by which helicase is loaded onto DNA and then activated remain still elusive. Differently from yeast origins that are characterised by a specific *consensus* sequence, called Autonomous Replicating Sequences (ARS), mammalian ORC have no sequence-specificity and origins seem to be defined also by epigenetic components (Dhar et al. 2012; Méchali 2010; Hyrien 2015; Prioleau & MacAlpine 2016). CpG islands, G-quadruplexes, transcriptional promoters and markers of the open chromatin state are all common elements found to be associated with

mammalian origins, but none of them is sufficient for origin specification (Cadoret et al. 2008; Cayrou et al. 2011; Cayrou et al. 2015; Sequeira-Mendes et al. 2009; Martin et al. 2011; Besnard et al. 2012). The lack of a specific marker and the heterogeneity of elements related to replication origins might be in part explained by the high structural complexity of mammalian genomes.

Temporal and spatial regulation of DNA replication is required to overcome all potential impediments to fork progression, as well as to face changes in chromatin organisation or growth conditions occurring during development and cell differentiation (Blow et al. 2011; Yekezare et al. 2013; Zeman & Cimprich 2014; Rivera-Mulia & Gilbert 2016b; Berti & Vindigni 2016). In mammalian cells, replication origins are grouped in 200-400 kb regions called replication clusters and many of these clusters are in turn enclosed in larger genomic domains (i.e. replication domains) subjected to a strict temporal regulation (Courbet et al. 2008; Cayrou et al. 2011; Guilbaud et al. 2011; Mesner et al. 2011; Ostankovitch & Debatisse 2013; Hyrien et al. 2013). This hierarchical organisation allows cells to control origin activation locally within clusters and globally at the level of replication domains (Yekezare et al. 2013).

It is well documented that the amount of licensed origins is higher than the number of those activated (Woodward et al. 2006; Ge et al. 2007; Blow & Ge 2009; Plosky 2015). This means that within each replication cluster there are many licensed origins that can potentially be activated, but among them only a restricted group is chosen (Fig. 1). Furthermore, they fire stochastically in each cell cycle (Fig. 1). Hence, the number of potential origins is redundant and origin choice varies among cells in a population (Hyrien et al. 2003; Gilbert 2007; Méchali 2010). Origin flexibility and redundancy are essential in mammalian cells to successfully complete DNA replication in the presence of replication stress (Anglana et al. 2003; Alver et al. 2014; Aguilera & García-Muse 2013; Cortez 2015). Indeed, in unperturbed conditions the supply of licensed but inactive dormant origins is never used and therefore forks from nearby activated initiation sites passively replicate dormant origins (Fig. 1). In contrast, under replication stress conditions replication forks stall, leading to activation of the DNA damage checkpoint. As a consequence, within the active replication cluster dormant origins are recruited and fire to avoid formation of gaps; in parallel, origins in still inactive replication domains are inhibited (Fig. 1). Thus, the DNA damage checkpoint acts locally on origin choice and globally preventing firing in inactive replication domains (Pope & Gilbert 2013; Yekezare et al. 2013; Berti & Vindigni 2016; Alver et al. 2014).

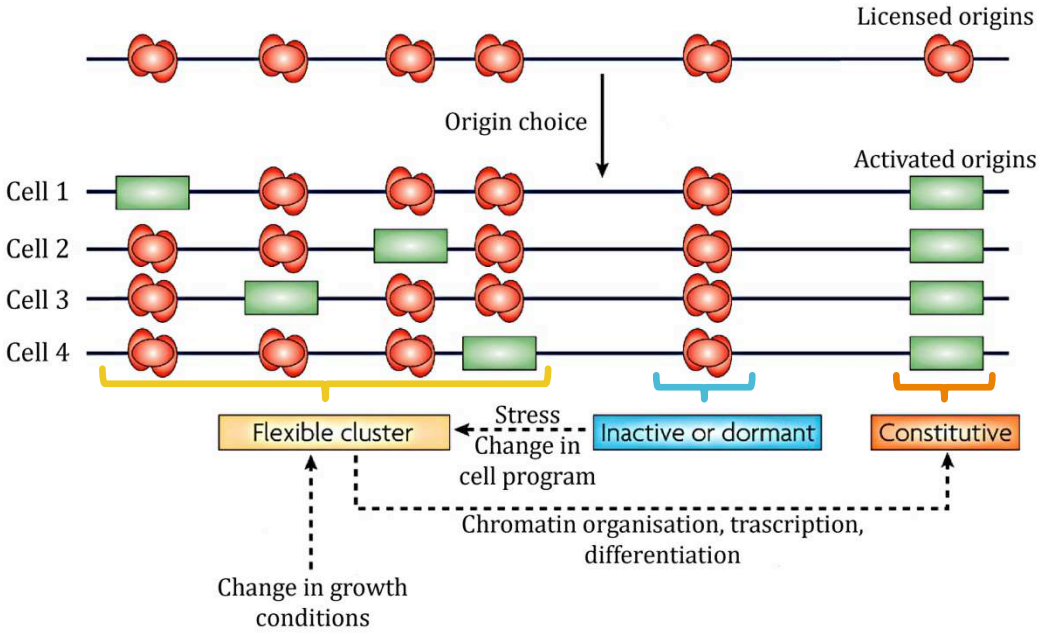


Fig. 1. Origin choice is regulated according to changes in chromatin organisation occurring during development and cell differentiation, or in response to replication stress. The majority of initiation sites fires stochastically at different positions within replication clusters and their use may increase or decrease according to physiological or perturbed growth conditions (flexible origins). A small fraction of origins are constitutively activated at fixed positions within clusters because of chromatin or transcriptional constraint (constitutive origins). Instead, inactive or dormant origins are never used in unperturbed conditions, but they are recruited under replication stress via activation of the DNA damage response. Adapted from (Méchali 2010).

Differently from the flexible nature of origin selection, replication clusters and domains are conserved among subsequent cell cycles and they are subjected to temporal regulation (Hiratani et al. 2008; Ryba et al. 2011; Pope & Gilbert 2013; Pope et al. 2014; Rivera-Mulia & Gilbert 2016b). Indeed, replication domains are classified as early, mid or late replicating according to their temporal order of activation during the S-phase (Fig. 2). Long origin-poor transition regions, replicated by unidirectional running forks, separate early and late domains (Fig. 2). Transition regions have been demonstrated to be more prone to DNA damages (Hiratani et al. 2008), as exemplified by the human common fragile site *FRA6E* (Palumbo et al. 2010). The importance of temporally regulating origin activation within clusters is associated with the need to ration the limited amount of resources, such as nucleotides and firing factors, to assure completion of genome duplication (Douglas & Diffley 2012; Rivera-Mulia & Gilbert 2016a). For this reason, the temporal program is set at the timing decision point in early G_1 before origin specification (Fig. 2; (Clifford 2014; Yekezare et al. 2013; Pope & Gilbert 2013; Rivera-Mulia & Gilbert 2016a)).

Replication timing is a peculiar feature of eukaryotic cells and it is strictly correlated with chromatin organisation, gene expression and nuclear position (Hiratani et al. 2008; Foti et al. 2016; Pope & Gilbert 2013; Helmrich et al. 2013;

Hoffman et al. 2015; Rivera-Mulia & Gilbert 2016b; Göndör & Ohlsson 2009). In addition, it is cell type-specific and modulated during development and cell differentiation (Hiratani et al. 2008; Ryba et al. 2011; Pope et al. 2014). For instance, during the neuronal differentiation of embryonic stem cells large replication domains are assembled by merging smaller units according to chromatin modifications (Hiratani et al. 2008). This process is called domain consolidation and it has been demonstrated to be lineage-specific (Ryba et al. 2011). Furthermore, to avoid collisions between replication and transcription, eukaryotic cells maintain these two processes temporally separated, so that genes that are actively transcribed are early replicated, as in the case of histone genes (Helmrich et al. 2013; Hoffman et al. 2015). Interestingly, a tight correlation between replication domains and their nuclear position has been proven (Fig. 2; (Pope et al. 2014; Rivera-Mulia & Gilbert 2016b; Foti et al. 2016)). Indeed, late replicating regions enriched in heterochromatic markers are mainly associated with the nuclear lamina as demonstrated by the co-localisation of Rif1 with Lamin B1 in mouse embryonic stem cells (Foti et al. 2016); in contrast, early domains are prevalently euchromatic and localised in the nuclear interior (Pope & Gilbert 2013; Pope et al. 2014; Göndör & Ohlsson 2009). However, understanding whether genomic functions, such as replication timing and gene transcription, are the causes or instead the consequences of changes in chromatin structure is a critical issue to be further resolved (Rivera-Mulia & Gilbert 2016b).

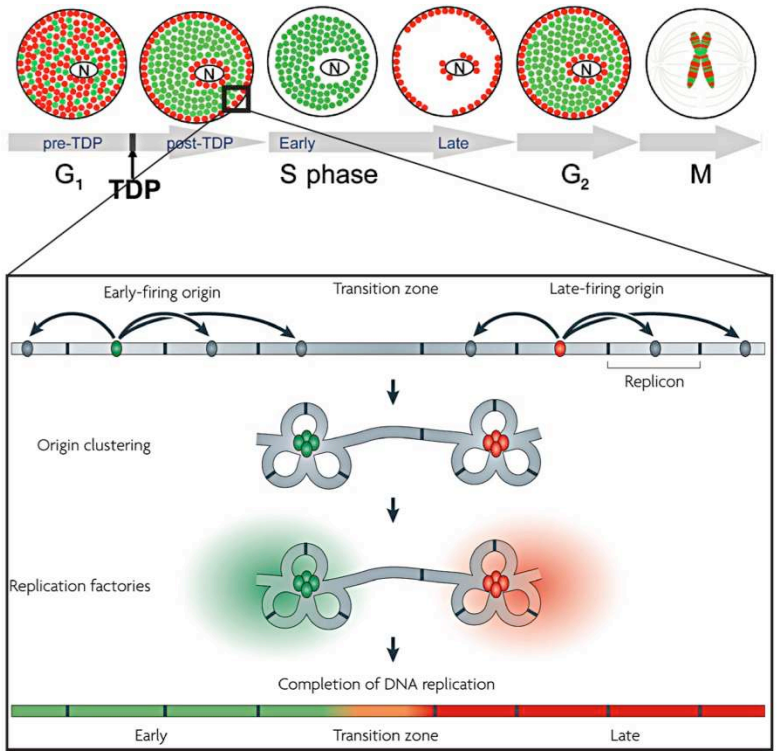


Fig. 2. The Timing Decision Point (TDP) is set in early G_1 and it represents the time when chromatin and replication timing are reorganised. Replication domains replicating early (green spots) in S-phase are localised in the nuclear interior, whereas late (red spots) domains are anchored to the nuclear lamina and to nucleolus. During mitosis early and late replicating regions are visualised as R and G bands on chromosomes, respectively. In the following G_1 phase, the spatial organisation of replication domains is restored at the TDP before origin specification. In the enlargement, large origin-poor transition regions separating early and late replication domains can be appreciated. Origin clusters are grouped within domains and their activation occurs in a temporally regulated manner via recruitment of firing factors forming replication factories. In this frame, the limited amount of resources is rationed among clusters with different timing, ensuring the completion of DNA replication. Adapted from (Göndör & Ohlsson 2009; Gilbert et al. 2010).

The flexibility of mammalian replication origins and the lack of *consensus* sequences or specific epigenetic markers are limiting aspects for origin identification (Dellino & Pelicci 2014; Gilbert 2012; Hyrien 2015; Prioleau & MacAlpine 2016). Indeed, all approaches used to map yeast origins have revealed challenging in mammals. Thus, in recent years exploiting the emergence of high-throughput sequencing technologies, several methods were developed to adapt to the mammalian system (Hyrien 2015; Prioleau & MacAlpine 2016). However, to date an exhaustive map of mammalian replication origins is still missing. Hence, data obtained by different approaches must be integrated to thoroughly describe the mammalian replication landscape (Dellino & Pelicci 2014; Hyrien 2015; Tuduri et al. 2010; Gilbert 2012; Prioleau & MacAlpine 2016).

To achieve precise origin mapping several techniques rely on the detection of characteristic replication intermediates, such as small nascent strand DNA (SNS), replication bubbles and Okazaki fragments. Among them, SNS sequencing (SNS-seq) allows to map origins at nucleotide-precision by the determination of sites of SNS enrichment. By SNS-seq a strong correlation between G-quadruplexes and replication origins has been observed (Besnard et al. 2012), however this result is restricted to early replicating regions and it is partially biased by the inability of λ-exonuclease to digest G-quadruplexes (Foulk et al. 2015; Dellino & Pelicci 2014; Prioleau & MacAlpine 2016). In comparison to SNS-seq, sequencing of replication bubbles is more sensitive but less precise (Dellino et al. 2013; Prioleau & MacAlpine 2016) so that only few broad initiation zones have been identified; interestingly, they are prevalently in intergenic or non transcribed regions (Mesner et al. 2013; Petryk et al. 2016). Recently, a novel technique (Ini-seq) based on a cell-free system has been developed (Langley et al. 2016). It allows not only to detect initiation sites with a precision similar to that of SNS-seq, but also to test the role of several co-factors on origin specification. However, this *in vitro* technique is really far away from the endogenous cellular context and the required synchronisation of cell cultures can give biased results. Finally, by sequencing Okazaki fragments both the activation of replication origins and the direction of replication forks can be detected (Petryk et al. 2016). Indeed, 6-150 kb initiation zones bordering active genes have been identified and the presence of large genomic regions replicated almost exclusively by unidirectional moving forks has also been determined (Petryk et al. 2016). The weakness of all these genome-wide approaches is the failure in identifying rare and inefficient replication origins, which are crucial events responsible for the cell-to-cell variability peculiar of mammalian genomes. This effect may be ascribed to the fact that genome-wide techniques rely on averaged data collected from whole cell populations (Tuduri et al. 2010; Dellino & Pelicci 2014; Hyrien 2015; Gilbert 2012; Prioleau & MacAlpine 2016).

An alternative way to identify replication origins in mammalian cells is based on the recognition of components of the pre-replicative complex. In particular, ORC1 and ORC2 subunits of the pre-RC are immunoprecipitated after their binding to DNA and the resulting genomic fragments are sequenced (ChIP-seq; (Dellino et al. 2013; Miotto et al. 2016)). Results poorly overlap with those obtained by SNS- or bubble-seq, but this is not surprising because ChIP-seq is not able to discriminate between inactive and active initiation sites and thus it detects all licensed origins (MacAlpine 2016; Dellino et al. 2013; Prioleau & MacAlpine 2016). Hence, this approach may be useful to evaluate epigenetic markers associated to replication origins, such as CpG islands or histone modifications, although it has been remarked that results must be carefully analysed to avoid biased interpretation (Teytelman et al. 2013).

For the full comprehension of the phenomena associated to mammalian DNA duplication, not only origin position but also replication dynamics must be taken into consideration, as the flexible nature of the process implies the modulation of

both fork rates and origin choice (Tuduri et al. 2010; Técher et al. 2013; Kaykov et al. 2016; Palumbo et al. 2013; Hyrien 2015; Norio & Schildkraut 2001; Prioleau & MacAlpine 2016). In this frame, single-molecule techniques represent the most appropriate tool to appreciate the cell-to-cell variability present in cellular populations of different types. Among them, molecular combing is the most accurate with a 1-5 kb resolution thanks to its constant rate of DNA elongation at difference with other fiber assays (Tuduri et al. 2010; Kaykov et al. 2016; Bensimon et al. 1994; Michalet et al. 1997). Molecular combing can be applied both for genome-wide and single-locus analyses. It allows not only the detection of fork rates, distances between activated origins and cluster size, but also the identification of alternative replication patterns, such as forks with unidirectional progression or with arms running at different speed (asynchronous forks) and pause/arrest sites (Tuduri et al. 2010; Técher et al. 2013; Herrick & Bensimon 2008; Prioleau & MacAlpine 2016). When performed at single loci, FISH probes are hybridised to depict the region of interest and bidirectional origins are also mapped. For single-locus analysis the Single Molecule Analysis of Replicated DNA (SMARD) assay is also used because it is less laborious than DNA combing. However, its main constraint is that it focuses on quite limited genomic regions obtained by endonuclease digestion, which can in turn influence the endogenous epigenetic background (Norio & Schildkraut 2001; Norio & Schildkraut 2004; Norio et al. 2005; Gerhardt et al. 2014; Gerhardt et al. 2016). Although single-molecule approaches do not provide an accurate genome-wide map of all replication origins because of the limited number of analysed molecules in comparison to high-throughput techniques, their major advantage is the ability to determine rare events variable among cells and prevalently associated to genome instability, such as activation of dormant origins (Tuduri et al. 2010; Técher et al. 2013; Hyrien 2015; Prioleau & MacAlpine 2016). Indeed, molecular combing has been crucial for understanding the replication-based mechanism underlying the fragility at human common fragile sites *FRA3B*, *FRA6E* and *FRA16C* (Letessier et al. 2011; Palumbo et al. 2010; Ozeri-Galai et al. 2011; Palumbo et al. 2013). Their intrinsic instability is caused by different mechanisms: a paucity of efficient initiation sites at the core of *FRA3B* (Letessier et al. 2011; Palumbo et al. 2013), a failure in activating additional origins in response to fork stalling at *FRA16C* (Ozeri-Galai et al. 2011) and the presence of a transition zone mostly replicated by unidirectional forks at *FRA6E* (Palumbo et al. 2010). Interestingly, while replication forks seem to move slower at *FRA16C* than in the whole genome, in both *FRA3B* and *FRA6E* loci fork rate remains constant (Letessier et al. 2011; Palumbo et al. 2010; Ozeri-Galai et al. 2011). Therefore, to better elucidate the events related to the mammalian replication process, an integration of data obtained by genome-wide high-throughput and single-molecule techniques is required (Hyrien 2015; Dellino & Pelicci 2014; Gilbert 2012; Tuduri et al. 2010; Prioleau & MacAlpine 2016).

In the light of all above considerations and being aware that replication profiles are cell type-specific and vary during development and cell differentiation (Palumbo et al. 2013; Hiratani et al. 2008; Ryba et al. 2011), my PhD project aims at understanding how replication profiles are regulated in response to endogenous and exogenous stress conditions. In particular, I considered whether and how an unstable DNA sequence may influence the replication program. In this frame, the effects of unusual DNA secondary structures caused by the expansion of a trinucleotide repeat on replication timing and replication dynamics were evaluated and human lymphoblastoid cell lines carrying a GAA-repeat expansion were used as experimental model. Analyses were carried out by applying different single-molecule approaches and conclusions were deduced by integrating all data. Furthermore, I focused on two other general aspects concerning modulation of mammalian DNA replication. In collaboration with Prof. Bianchi (Dept. of Biology, University of Padova), the effects of an increased and unbalanced supply of nucleotide precursors on replication dynamics were evaluated. Human primary fibroblasts and human acute monocytic leukemia THP1 cells lacking SAMHD1, a protein involved in nucleotide metabolism, were used as models for genome-wide molecular combing analyses. Finally, I considered how DNA replication responds to the forced reactivation of the cell cycle in terminally differentiated cells. In the frame of a collaboration with Dr. Crescenzi (Istituto Superiore di Sanità, Rome), mouse myotubes were used as experimental model for terminal differentiation; after they were forced to re-enter the cell cycle, replication dynamics are analysed by molecular combing.

1. Trinucleotide repeats and DNA replication: a single-locus perspective from the *Frataxin* gene

Variation in the number of tandem repeats between individuals at different loci usually occurs without any negative consequences. An exception is represented by a class of microsatellites, and among them trinucleotide repeats (TNR) are the most unstable (Mirkin 2007; Dion & Wilson 2009; Nelson et al. 2013). Indeed, variations of TNR length are implicated in human neurodegenerative and neuromuscular disorders, and strongly affect age-of-onset and severity of these diseases (McMurray 2010; Lee & McMurray 2014). The expansion of TNR is a dynamic process occurring during intergenerational transmission, but in most disorders it has also a somatic counterpart (Pearson et al. 2005). In fact, evidences of somatic TNR instability have been reported in several tissues of patients affected by some of these expansion disorders, such as Huntington's disease and Friedreich's ataxia (Gomes-Pereira et al. 2014; Bidichandani et al. 1999; Hellenbroich et al. 2001).
TNR instability is tightly correlated with the propensity of expanded repeats to form unusual DNA secondary structures. DNA replication, repair, recombination and gene transcription are all implicated in affecting TNR stability (Cleary &

Pearson 2005; Gomes-Pereira et al. 2014; Kim & Mirkin 2013), although the interplay among these pathways is still unclear. DNA secondary structures are well-known sources of replication stress as they represent an impediment to replication fork progression (Alver et al. 2014; Zeman & Cimprich 2014; Magdalou et al. 2014; León-Ortiz et al. 2014). Different models have been proposed to describe mechanisms underlying their formation during DNA synthesis (Pearson et al. 2005; Cleary & Pearson 2005; Kim & Mirkin 2013). In several experimental systems including bacteria, yeast, transfected or engineered human cells, the stability of DNA secondary structures has been demonstrated to strongly depend on the orientation by which TNR are replicated (Mirkin & Mirkin 2007; Follonier et al. 2013; Gerhardt et al. 2014; Kim et al. 2008). For instance, GAA- and CGG-repeats preferentially form stable secondary structures when synthesised from the leading strand (Gerhardt et al. 2014; Follonier et al. 2013; Chandok et al. 2012; Kim et al. 2008). Furthermore, to explain TNR instability and its variability among different TNR repeats, the balancing act hypothesis has been proposed on the basis of analyses carried out on yeast mutants (Kim & Mirkin 2013). This hypothesis states that depending on the stability of DNA secondary structures TNR can undergo small- or large-scale expansions via two different pathways relying either on a strand-specific slippage of DNA polymerase or on a fork deviation mechanism such as the template switch during DNA synthesis (Kim & Mirkin 2013). According to this hypothesis, strand slippage may describe the variability of repeat threshold lengths of different human expansion disorders, while the somatic instability may be explained by large-scale expansions caused by template switching (Kim & Mirkin 2013; Shishkin et al. 2009). However, further work is required to confirm this hypothesis in human cell lines.

In order to describe the replication-based mechanisms underlying TNR expansion, three models were proposed on the basis of data derived from yeast (Cleary & Pearson 2005; Pearson et al. 2005): the origin-shift and the origin-switch mechanisms are dependent on replication origin position relative to TNR repeat, whereas the fork-shift model links the epigenetic background with a replication fork slippage at TNR repeat. In recent years, the fine-tuning of protocols to obtain induced pluripotent stem cells (iPSC), the availability of mouse models and the improvement of single-molecule and single-locus techniques for origin detection allow to evaluate whether these proposed models can also conform to mammalian and human cells. For instance, analysis of mouse embryonic stem cells carrying a CGG-repeat expansion in the *FMR1* locus reveals a change in the position of the activated replication origins with respect to the repeats, indicating the origin-switch as the model at the basis of the CGG-expansion (Gerhardt et al. 2014). During my study, I obtained similar results on lymphoblastoid cell lines carrying a GAA-repeat expansion (Stevanoni et al. 2016) and in parallel the same conclusions were reached in an independent study carried out on iPSC (Gerhardt et al. 2016). A switch in the progression of forks through the repeat was demonstrated in both

cellular systems, however if this mechanism is actually the cause of the repeat expansion remain not cleared.

There is a great deal of evidence supporting a replication-based mechanism at the origin of expansion and instability of several TNR, but most of the data were obtained by using model systems in which TNR length never reached the huge amount of repeats observed in patients (Follonier et al. 2013; Chandok et al. 2012; Mirkin & Mirkin 2007; Kim et al. 2008; Krasilnikova & Mirkin 2004). Thus, it is unknown whether the presence and the size of a TNR expansion may influence DNA replication profile and dynamics in the endogenous context. To investigate this aspect during my PhD activity, human Friedreich's ataxia (FRDA) cell lines were used as experimental model. FRDA patients are homozygous for a GAA-repeat expansion within intron 1 of the *Frataxin* (*FXN*) gene (Campuzano et al. 1996; Pandolfo 2008). The mutation is causing the silencing of the gene by inhibition of either transcription initiation or elongation and as a consequence of chromatin remodelling sequence heterochromatinisation was also observed in the proximity of the expanded GAA-repeat (Bidichandani et al. 1999; Silva et al. 2015; Evans-Galea et al. 2013; Chan et al. 2013; Li et al. 2015; Yandim et al. 2013). Interestingly, recent works demonstrate that in the presence of the GAA-repeat expansion stable RNA/DNA hybrids (R-loops) are formed in model systems as well as *in vivo* at the endogenous *FXN* gene (Groh et al. 2014; Grabczyk et al. 2007; Reddy et al. 2014). R-loops correlate with repressive chromatin marks and induce *FXN* silencing by promoting sequence heterochromatinisation (Groh et al. 2014). For this reason, preventing R-loop formation represents an important goal for clinical therapies (Corey et al. 2016). However, it is not clear whether R-loops are formed before chromatin mark establishment or instead nearby histone modifications promote their formation. Furthermore, a very recent work by Gerhardt and colleagues (Gerhardt et al. 2016) speculates on a possible role of the collision between replication and transcription in promoting GAA-repeat expansions. In mammalian cells conflicts between these two processes are considered uncommon, because gene expression and replication timing are temporally separated with the only proven exception represented by long genes at CFS (Helmrich et al. 2013; Hoffman et al. 2015; Rivera-Mulia & Gilbert 2016b). Thus, to demonstrate the occurrence of replication-transcription collisions at the expanded GAA-repeat, it is crucial to prove the coexistence of these two processes at the repeat. However, information concerning the transcriptional profile of the *FXN* gene are lacking, hence any conclusions are yet incautious.

As a consequence of the stable unusual DNA secondary structure formed by the expanded repeats in mammalian cells, a replication impairment is expected, but how this would affect the replication program of the genomic segment cannot be easily predictable. Here, replication timing and dynamics of the *FXN* gene in the presence/absence of the GAA-repeat expansion were considered to elucidate this issue.

2. Altered dNTP pools and modulation of DNA replication

A balanced supply of deoxynucleotide triphosphates (dNTPs) is required to assure completion of DNA replication and to avoid genome instability (Magdalou et al. 2014; Zeman & Cimprich 2014; Evertts & Coller 2012). dNTP pools are differently regulated according to the proliferative status of cells. Indeed, quiescent cells require dNTPs only for mitochondrial DNA synthesis and for DNA repair so that pools are maintained at their basal level (Rampazzo et al. 2010; Pontarin et al. 2012; Nordlund & Reichard 2006). Instead, in proliferating cells the sufficient amount of nucleotides must be provided for DNA synthesis, thus dNTP pools greatly increase upon entry into S-phase compared to the G_1 levels (Rampazzo et al. 2010; Franzolin et al. 2013; Chabes et al. 2003; Kumar et al. 2010). This process is mediated by the overexpression of the ribonucleotide reductase (RNR) enzyme, which is responsible for the synthesis of all four deoxynucleoside triphosphates by the catalysis of ribonucleoside di- or triphosphates (Nordlund & Reichard 2006). RNR activity is maintained throughout S-phase and it is counteracted by different nucleotidases and phosphohydrolases, so that in the G_1 phase dNTP pool sizes are restored. Therefore, the concentration of each dNTP fluctuates in a strictly regulated manner during the cell cycle by the balanced action of synthetic and catabolic enzymes (Rampazzo et al. 2010).

It is essential to tightly control the balance among dNTP supply, degradation and consumption according to replication fork rates: indeed, dNTP pool imbalances have been demonstrated to interfere with DNA replication leading to mutagenesis in yeast (Chabes & Stillman 2007; Buckland et al. 2014; Watt et al. 2015; Kumar et al. 2010) and promoting tumorigenesis in human cells (Bester et al. 2011; Chabosseau et al. 2011; Gemble et al. 2015). Whereas the effects of decreased dNTP pools on the replication process are known (Anglana et al. 2003; Bester et al. 2011; Courbet et al. 2008), the consequences of increased and unbalanced pools are only partially depicted in yeast (Chabes & Stillman 2007; Buckland et al. 2014; Poli et al. 2012; Watt et al. 2015; Kumar et al. 2010). In particular, when the availability of nucleotide precursors is limiting, replication forks are markedly slowed down and in turn the DNA damage checkpoint is activated leading to recruitment of dormant origins (Anglana et al. 2003; Courbet et al. 2008; Kumar et al. 2010; Bester et al. 2011). After addition of exogenous nucleotide precursors, replication fork rates are fully restored, indicating that dNTPs availability is crucial for modulation of fork speed and origin choice (Anglana et al. 2003; Courbet et al. 2008). Furthermore, dNTP depletion has also been demonstrated to promote tumorigenesis in human cells via oncogenic deregulation of the retinoblastoma pathway (Bester et al. 2011). In contrast, the effects of a high supply of dNTPs have been mainly described in yeast. Increased and unbalanced pools affect replication fidelity of both leading and lagging strand particularly in coding or late replicating genomic regions (Buckland et al. 2014; Watt et al. 2015). In addition, constitutively high concentrations of dNTPs in yeast have been demonstrated to block cell cycle progression in late G_1 as

a consequence of a reduced recruitment of firing factors at replication origins (Chabes & Stillman 2007). Finally, a single study in yeast considers the effects of an high dNTP availability on replication dynamics and suggests that fork rates increase also under replication stress, but without activating the DNA damage checkpoint (Poli et al. 2012).

Whether and how altered dNTP pools influence replication dynamics in mammalian cells is unknown, and my work aims at elucidating these aspects. For this purpose, human cell lines with increased and unbalanced dNTP pools due to the lack of SAMHD1, a protein involved in nucleotide precursor metabolism (Franzolin et al. 2013; Miazzi et al. 2014), were used as experimental model. SAMHD1 is a triphosphohydrolase converting all four dNTPs in deoxynucleosides both in proliferating and quiescent cells (Franzolin et al. 2013; Miazzi et al. 2014). SAMHD1 by the hydrolysis of DNA precursors represent the major counterpart for RNR activity and it is responsible for the maintenance of the correct level of dNTP pools in G_1 (Franzolin et al. 2013; Miazzi et al. 2014). Indeed, its silencing in normal human lung fibroblasts leads to a great increase in dNTP pools and to a subsequent accumulation of cells in G_1 (Franzolin et al. 2013), similarly to the yeast model (Poli et al. 2012). Mutations of the *SAMHD1* gene are involved in the Aicardi-Goutièrres syndrome (AGS), an inflammatory disorder (Rice et al. 2009), as well as in several types of cancer (Clifford et al. 2014; Kohnken et al. 2015).

Here, replication dynamics was assessed genome-wide in human SAMHD1-mutated skin fibroblasts derived from AGS patients and in human acute monocytic leukemia THP1 cells depleted of SAMHD1 (KO THP1). THP1 cells have been widely used as cellular model to study modulation of monocyte and macrophage functions (Chanput et al. 2014) and differently from other monocytic cell lines they have a normal endogenous expression of SAMHD1 (Laguette et al. 2011). The depletion of SAMHD1 has been employed to explore the effects of SAMHD1 silencing on cell proliferation, cell cycle progression and apoptosis (Bonifati et al. 2016). KO THP1 cells have increased dNTP levels, accumulate in G_1 and show more robust cell viability associated with less activation of apoptotic pathways (Bonifati et al. 2016). Thus, both AGS fibroblasts and THP1 cells may be good cellular models to evaluate the possible effects of increased and unbalanced dNTP pools on DNA replication. Beyond the expected consequences on replication fork rates, understanding whether high dNTP pools may influence the cellular response to replication stress and all the events associated with fork stalling and restart is an intriguing aspect to be considered.

3. Forced cell cycle reactivation and DNA replication

Replication profiles of mammalian cells are not only cell type-specific but also different along the genome (Palumbo et al. 2013; Pope & Gilbert 2013; Pope et al. 2014; Ryba et al. 2011; Letessier et al. 2011; Rivera-Mulia et al. 2015). This large heterogeneity may be ascribed to the random use of replication origins and the variability of fork rates, but it is also defined by the diverse proportions of alternative replication patterns, such as unidirectional forks, pause/arrest events and forks with arms running at different rates (Palumbo et al. 2013; Bianco et al. 2012; Courbet et al. 2008; Kaykov et al. 2016). For instance, analysis of the replication profile of human primary fibroblasts revealed a marked contribution of unidirectional forks during DNA replication in comparison with lymphocytes or normal lymphoblastoid cell lines (Palumbo et al. 2013). Furthermore, in the same cell lines it has also been demonstrated that proceeding along population divisions toward senescence the proportion of unidirectional forks increases (Palumbo et al. 2013). This suggests that unidirectional forks are important features of mammalian replication profiles. Unidirectional patterns are seen also in large transition zones dividing early and late replication domains (Hiratani et al. 2008) and at the *FRA6E* locus in lymphocytes (Palumbo et al. 2010), however their biological meaning is still elusive. Beyond replication dynamics variability, changes in the replication timing are also cell type-specific and differ during cell differentiation and development (Hiratani et al. 2008; Hiratani & Gilbert 2009; Ryba et al. 2011; Rivera-Mulia & Gilbert 2016b; Rivera-Mulia et al. 2015). Thus, the identification of specific features defining the replication programs in different cell lines and at different developmental and differentiation stages is necessary to better characterise the mammalian process.

In this frame, taking advantage of a collaboration with Dr. Crescenzi, my goal is to understand how replication is modulated in response to forced cell cycle reactivation of terminally differentiated mammalian cells.

The so-called terminal differentiation state is defined by the persistent exit of cells from the cell cycle occurring during acquisition of their functional specification. Terminally differentiated (TD) cells are characterized by their inability to proliferate due to different (and still elusive) mechanisms that permanently arrest cell cycle in G_0. However, terminal differentiation has been demonstrated to be a reversible state (Tiainen et al. 1996). Indeed, TD cells can be forced to renter the cell cycle by different mechanisms, including the use of oncogenic viruses (Latella et al. 2000) or the knock down of cyclin-dependent kinase inhibitors (CKIs; (Pajalunga et al. 2007)). After cell cycle reactivation by p21 and p27 CKI depletion, quiescent mouse fibroblasts can extensively proliferate (Pajalunga et al. 2007). Instead, skeletal muscle myotubes, which are a good model for terminal differentiation, fail to complete DNA replication and accumulate DNA damages upon forced cell cycle re-entry (Pajalunga et al. 2010). To explain these discrepancies, the proposed working hypothesis ascribes the DNA replication

failure to the inability of reactivated myotubes to properly expand their dNTP pools (Pajalunga et al. 2017), with respect to the basal level typical of quiescence (see p. 19). Therefore, mouse myotubes forced to renter cell cycle by depletion of p21 and p27 CKIs and proliferating myoblasts were used as experimental models. Moreover, the effects of the exogenous nucleotide precursor addition were tested in TD muscle cells after their forced cell cycle reactivation (Pajalunga et al. 2017). In the light of above considerations and previous studies (Palumbo et al. 2013; Courbet et al. 2008), fork rate, origin distance and replication cluster size as well as proportions of unidirectional, paused/arrested and asynchronous forks were considered. Understanding whether changes in the replication timing occur in reactivated myotubes will contribute to further clarify mechanisms underlying the DNA replication impairment observed.

MATERIALS AND METHODS

1. Cell cultures

Epstein Barr virus-transformed lymphoblastoid cell lines from two unrelated FRDA patients GM15850 and GM16227 and from the healthy brother GM15851 of patient GM15850 were obtained by the Human Genetic Cell Repository of the Coriell Institute (Camden, New Jersey, USA). The H691 cell line is an Epstein Barr virus-transformed lymphoblastoid cell line established from a young healthy male adult. All lymphoblastoid cell lines used in this study were grown in RPMI 1640 medium supplemented with 15% foetal bovine serum (FBS; EuroClone, Italy) and 100 U/ml penicillin/streptomycin antibiotics (Gibco, Life Technologies). Cells were sub-cultured three times a week at the density of 0.5 x 10^6 cell/ml and the estimated duplication time ranges between 29–32 h.
The THP1 wildtype cell line is a human acute leukemia monocytic cell line and the knockout of the SAMHD1 protein was obtained by CRISPR/Cas9 technology. THP1 cells were grown in RPMI 1640 medium (Gibco, Life Technologies) supplemented with 10% FBS (EuroClone, Italy) and 100 U/ml penicillin/streptomycin antibiotics (Gibco, Life Technology). Cells were sub-cultured three times a week at the density of 0.2 x 10^6 cell/ml. Human skin fibroblasts derived from Aicardi-Goutièrres syndrome patients GR295 and MG282 and from healthy subjects C63 and Goteborg were cultured in DMEM medium (Gibco, Life Technologies) supplemented with 10% FBS (EuroClone, Italy), nonessential amino acids (Gibco, Life Technology) and 100 U/ml penicillin/streptomycin (Gibco, Life Technology). Fibroblasts were split (1:2) when at confluence to avoid cell growth arrest. Both THP1 cells and fibroblast cell lines were supplied by the laboratory of Prof. Bianchi (Dept. Biology, University of Padova).
All cells were grown at 37°C and 5% CO_2.

2. Genotype and transcriptional analysis in lymphoblastoid cell lines

The number of GAA-repeats within the intron 1 of the *FXN* gene is variable among FRDA alleles. In the cell line GM15850 alleles carry 650 and 1030 GAA-repeats, respectively; in GM16227 cells alleles carry 630 and 830 GAA-repeats, respectively. To confirm genotypes reported by the Coriell cell repository, the length of the repetitive tracts was assessed in the four cell lines by long-range PCR (QIAGEN Long range PCR kit, Qiagen) using a primer set specific for the amplification of the GAA-repeat (forward: 5'-GGAGGGATCCGTCTGGGCAAAGG-3', reverse: 5'-CAATCCAGGACAGTCAGGGCTT-3'; normal amplicon length: 1,5 kb) (Campuzano et

al. 1996). 50 ng of genomic DNA purified from each cell line and 10 pg of BAC clone RP11-265B8 harbouring the wildtype *FXN* gene were used as template.
For the expression analysis, total RNA was isolated from the four cell lines using the RNeasy mini kit (Qiagen). 1 µg of RNA was retro-transcribed (EuroScript M-MLW Reverse Transcriptase (RNase H-), EuroClone) and according to manufacturer's instructions, 2 µl of the product of the reaction (cDNA) were used as template for the semiquantitative RT-PCR reactions. For *FXN* transcript primer pairs were forward: 5'-CCTTGCAGACAAGCCATACA-3', reverse: 5'-GGTCCACTGGATGGAGAAGA-3' (amplicon length: 153 bp). *GAPDH* was chosen as normalising gene and the transcript was amplified by forward: 5'-CCTCAACGACCACTTTGTCA-3', reverse: 5'-TTCCTCTTGTGCTCTTGCTG-3' (amplicon length: 143 bp).

3. Proliferation assays and FACS sorting

Cell cycle distributions of the different cell lines were determined by flow cytometry after propidium iodide staining with a BD FACSCanto II flow cytometer (Becton Dickinson) according to standard protocols. The Cell Quest software (Becton Dickinson) was used for the analyses.
Cell sorting of GM15851, GM15850 and GM16227 cell lines was carried out on using a BD FACSAria (BD Biosciences), according to the procedure displayed in Fig. 3 and in agreement with (Hansen et al. 2010). Per each experiment at least 150 x 10^6 cells were labelled with a 30 min pulse of 100 µM 5-Chloro-2'-deoxyuridine (CldU; Sigma-Aldrich) (Fig. 3). After harvesting, cells were prepared for FACS analysis according to standard procedure. On the basis of the observed cell cycle distribution, S-phase intervals were set in order to collect four fractions of identical size spanning the entire S-phase (Fig. 3). The purity of each S-phase fraction was assessed at the end of the experiment (Fig. 3).

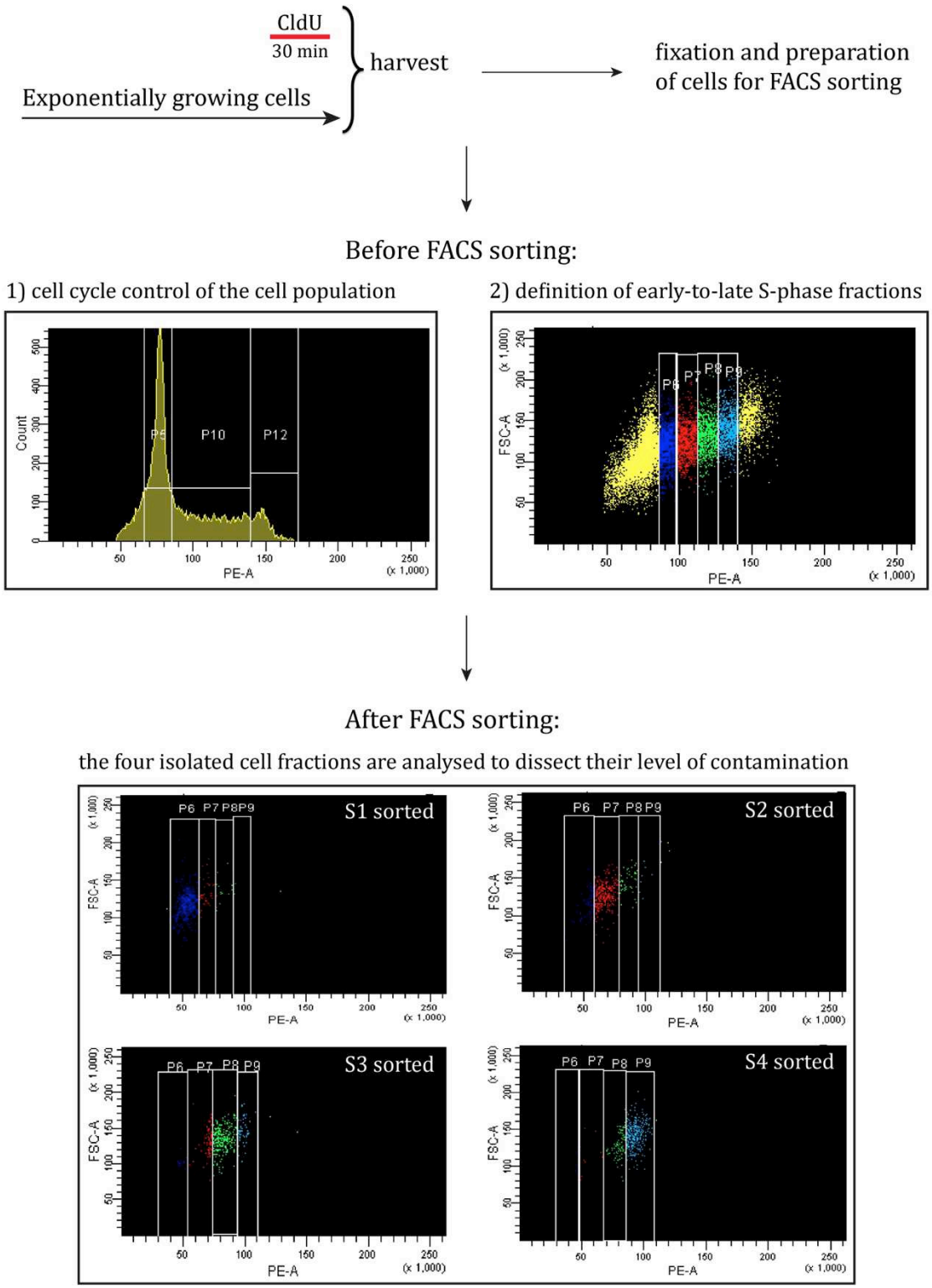


Fig. 3. Workflow of the FACS sorting procedure. Before harvesting, exponentially growing cells are labelled with CldU for 30 min. This allows confirming the accuracy of FACS sorting by CldU-immunodetection. Harvested cells are fixed and nuclei are stained with propidium iodide to allow separation according to DNA content. Before fractionation, the cell cycle distribution of the cell population is controlled (G_1-, S- and G_2/M-phases are displayed as P5, P10 and P12, respectively) and the four consecutive temporal windows of S-phase are defined (P6-9 in the graphs). After FACS sorting, early-to-late S-phase (S1-S4) isolated fractions are analysed to verify their level of contamination.

4. Interphase FISH

The interphase FISH approach was used to determine the replication timing of normal and expanded *FXN* alleles in asynchronous and sorted GM15851, GM15850 and GM16227 cell line populations.

4.1 BAC clone preparation and labelling

The BAC clones used in this study to identify the *FXN* locus were obtained by the Children's Hospital Oakland Institute (CHORI, USA). BAC DNA was extracted and purified according to a standard procedure. Briefly, bacterial cells were resuspended in Solution I (50 mM glucose, 10 mM EDTA and 25 mM Tris-HCl) supplemented with 4 mg/ml lysozime (Sigma-Aldrich, Italy) and then lysed through incubation in Solution II (0.2 N NaOH, 1%SDS) for 4 min at room temperature. The reaction was stopped by addition of cold Solution III (3M CH_3COOK, 11.5% CH_3COOH). After 20 min at 14000 rpm, 4°C, DNA was extracted by the standard phenol:chlorophorm:isoamyl alcohol protocol, resuspended in 1X TE buffer and treated with 40 µg/ml RNase A (Sigma-Aldrich, Italy) for 30 min at 37°C.
Purified BAC DNA was labelled with biotin-16-dUTP by a nick translation kit (Roche Biochemicals). In particular, 1 µg of template DNA was mixed with the Nick Translation mix and with a solution containing a mixture of dNTPs (dATP, dCTP and dGTP 0.25 mM each, 0.17 mM dTTP, 0.08 mM Biotin-16-dUTP). The reaction was carried out for 90 min at 16°C and stopped with 1 µl of 0.5 M EDTA pH 8. The success of the labelling procedure (fragments in the range 200-500 bp and absence of smear) was monitored by gel electrophoresis.

4.2 Interphase FISH protocol

Slides were prepared by using Cytospin 3 (Shandon Scientific Limited, UK), following a standard procedure. 30000-50000 cells were laid down on each slide.
Slides were post-fixed in 3:1 ethanol/acetic acid solution on ice, digested with 5 µg/ml Pepsin in 0.01M HCl, pH 3.0 at 37°C for 10 min and dehydrated in 70%, 90% and 100% ethanol. After denaturation at 72°C for 4 min in 70% formamide, 2X SSC pH 7.0, slides were dehydrated in ethanol solutions (70%, 90%, 100%).
The probe mix (100 ng of BAC DNA, 50X human Cot1 DNA (Invitrogen, Life Technologies), 5X Salmon Sperm DNA (Invitrogen, Life Technologies)) was denatured 10 min at 70°C and pre-annealed at 37°C for 90 min.
Per slide, 10 µl of probe mix were applied under a 22x22 mm coverslip and hybridisation was carried out overnight in a humidified chamber.
Post-hybridisation washes were 50% formamide, 2X SSC pH 7.0 at 42°C (3 times), 2X SSC pH 7.0 (3 times). Slides were incubated 30 min at 37°C in 1X blocking

solution (Roche Biochemicals) before immunodetection. CldU was detected by a rat monoclonal anti-BrdU antibody specifically cross-reacting with CldU (1:40, Abcam) and a 594 Alexa Fluor-conjugated anti-rat IgG (1:100; Molecular Probes, Life Technologies). At the same time, the biotin-labelled probe was detected by 488 Alexa Fluor-conjugated streptavidin (1:300, Molecular Probes, Life Technologies), followed by polyclonal biotin-conjugated anti-streptavidin (1:300; Rockland) and 488 Alexa Fluor-conjugated streptavidin (1:300; Molecular Probes, Invitrogen). Detection steps are reported in Table 1. Slides were counterstained with DAPI (2 µg/ml) in Vectashield Mounting medium (Vector, USA).

Table 1. Affinity molecules and antibodies employed for FISH analysis on interphase nuclei. The relative dilutions are indicated. In brackets host species are displayed. SAV=streptavidin; biot=biotin.

	I layer	**II layer**	**III layer**
Biotin	SAV- 488 1:300	Anti-SAV biot 1:300	SAV- 488 1:300
CldU	Anti-BrdU (rat) 1:40	Anti-rat IgG 594 1:100	

5. Molecular combing

Molecular combing allows performing replication analysis of both the whole eukaryotic genome (Fig. 5) and single genomic loci (Fig. 6).

5.1 Cell labelling and preparation of agarose plugs

Exponentially growing cells were labelled with two sequential 30 min pulses of 50 µM 5-Iodo-2'-deoxyuridine (IdU; Sigma-Aldrich) and 100 µM 5-Chloro-2'-deoxyuridine (CldU; Sigma-Aldrich) (Figs. 5A and 6A). An exception to this labelling scheme is represented by mouse myoblasts and reactivated myotubes (rMt), where both IdU and CldU were supplied at a final concentration of 100 µM and the length of pulses was: 30 min each for myoblasts and deoxynucleoside-dosed rMt; 60 min each for rMt. After harvesting, 1-2 x 10^5 cells were immobilised in LMP agarose plugs (1.5% w/v in 1X PBS, pH 7.4; Sigma-Aldrich, Italy) and incubated overnight at 50°C in 2 mg/ml Proteinase K solution (1% N-laurosylsarcosine, 0.1M EDTA pH 8.0, 0.01M Tris-HCl pH 8.0, 0.02M NaCl). Labelling and plug preparations of myoblasts and reactivated myotubes were done in Dr. Crescenzi's lab and the material sent in Padova for molecular combing. After digestion with 3 U of β-agarase I (New England BioLabs), high molecular weight DNA was released from 1–2 plugs in 0.1M MES, pH 6.1. According to a validated procedure (Palumbo et al. 2013; Palumbo et al. 2010), DNA combing was performed on silanised surfaces (homemade and from Genomic Vision); the main steps are depicted in Fig. 4.

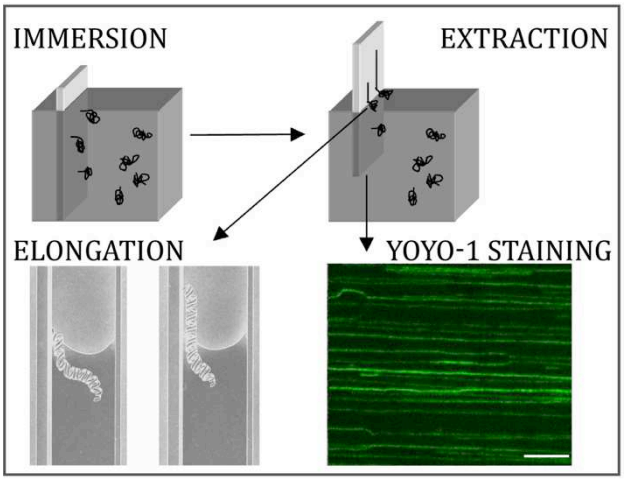


Fig. 4. Molecular combing procedure. Silanised coverslips are immersed into a solution containing random-coil DNA molecules (IMMERSION), whose extremities bound to surfaces. Coverslips are then pulled out at a constant speed of 300 µm/s (EXTRACTION) and the receding air-water meniscus allows the uniform stretching and alignment of DNA molecules on the surface (ELONGATION). The quality of combed molecules is monitored by DNA counterstaining with the fluorescent intercalating dye YOYO-1 (YOYO-1 STAINING). Calibration bar = 50 kb. Adapted from (Lebofsky & Bensimon 2003).

5.2 Genome-wide replication analysis

To study genome-wide replication profiles, combed DNA was denatured in 50 mM NaOH, 1 M NaCl for 15 min before being dehydrated in 70%, 90% and 100% ethanol. Detection steps are displayed in Table 2. Replication tracks were detected by immunofluorescence after two consecutive 30 min incubation steps: anti-BrdU antibodies cross-reacting with IdU (Becton Dickinson, developed in mouse, 2:7 dilution) and CldU (Abcam, developed in rat, 1:40 dilution) were recognised by a mixture of Alexa Fluor 488-conjugated anti-mouse IgG and Alexa Fluor 594-conjugated anti-rat IgG (all from Life Technologies, 1:50 dilution), respectively. To monitor DNA molecule integrity, slides were incubated with three additional layers of antibodies: primary mouse IgG2A anti-ssDNA (Millipore, Clone 3034, 1:25 dilution, 90 min), Alexa Fluor 350-conjugated anti-mouse IgG2A developed in goat, and Alexa Fluor 350-conjugated anti goat IgG developed in donkey (both secondary antibodies from Life Technologies, 1:50 dilution, 30 min). Therefore, replication forks are visualised in green and red fluorescence, while DNA was counterstained with blue fluorescence (Fig. 5).

Table 2. Antibodies employed for genome-wide replication analysis on combed DNA. In brackets host species are displayed. The relative dilutions are indicated.

	I layer	II layer	III layer	IV layer	V layer
IdU	Anti-BrdU (mouse) 2:7	Anti-mouse IgG 488 1:50			
CldU	Anti-BrdU (rat) 1:40	Anti-rat IgG 594 (donkey) 1:50			
DNA			Anti-ssDNA (mouse) 1:25	Anti-mouse IgG2A 350 (goat) 1:50	Anti-goat IgG 350 (donkey) 1:50

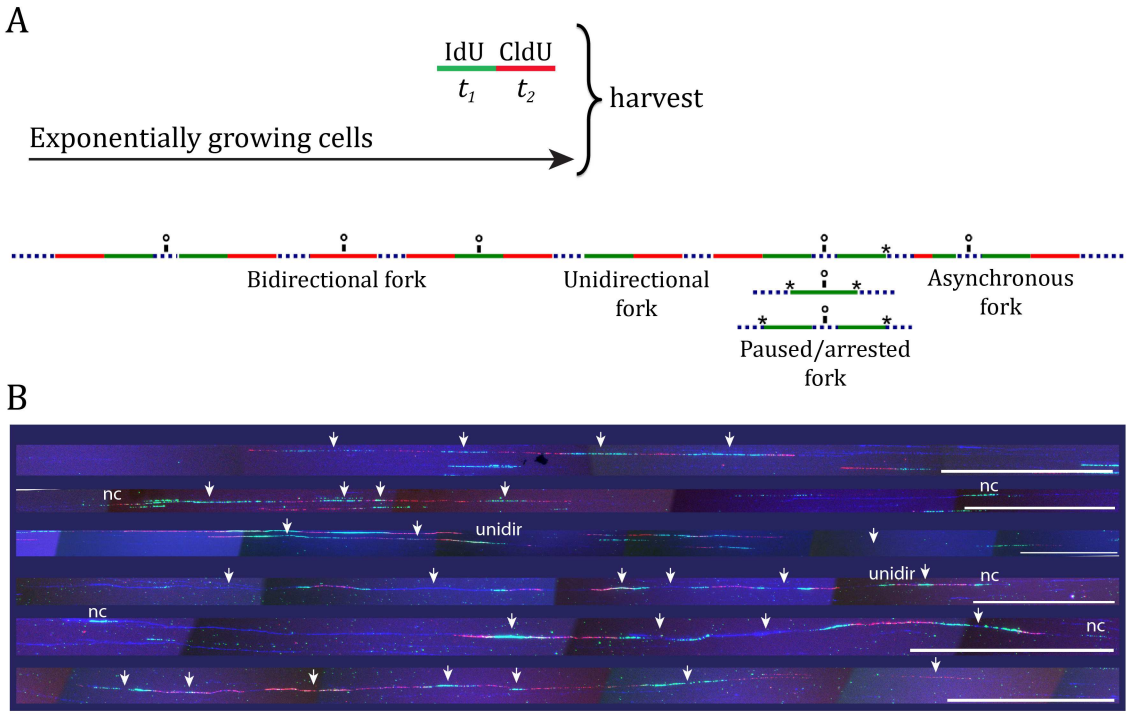


Fig. 5. (A) Experimental design and criteria for genome-wide molecular combing analyses. A two-pulse labelling scheme is applied for detection of replication forks. During the first pulse (t_1), IdU is incorporated in newly synthesized DNA and detected by green fluorescence; in the second pulse (t_2), CldU is supplied for the synthesis of the nascent strands and labelled DNA is detected by red fluorescence. To monitor the integrity of the molecules, DNA is counterstained and immunodetected in blue fluorescence (dotted lines). The expected replication patterns in a genome-wide molecular combing analysis include bidirectional origins (o), which may be mapped either in the middle of the distance between two opposite arms of a replication fork or in the middle of a green- or red-only track (which are observed when forks fire during the first or second pulse, respectively). Alternative replication patterns can be observed: forks with a single arm classified as unidirectional forks; paused/arrested forks (*) identified by bidirectional origins lacking the red fluorescence in one or both arms; forks firing from the origin with different arm rate classified as asynchronous forks. Prerequisite to include these replication patterns in the analysis is the presence of upstream/downstream fluorescent signals from adjacent forks or DNA counterstaining. (B) Examples of replicating molecules: green and red fluorescence correspond to replication tracks (IdU and CldU, respectively), while DNA is counterstained and visualised by immunodetection in blue. Arrows represent bidirectional replication origins. Unidirectional forks are also indicated (unidir). nc= tracts not considered in the course of the analysis. Calibration bar = 200 kb.

5.3 Single-locus replication analysis

For single-locus replication analysis, differentially spaced probes covering about 1 Mb are usually employed to identify the genomic region of interest (Palumbo et al. 2013; Palumbo et al. 2010). In this study, replication profiles of normal and expanded *FXN* alleles were assessed in a 850 kb-region identified by RP11-203L2, RP11-265B8 and RP11-548B3 BAC genomic clones (Children's Hospital Oakland Institute; CHORI, USA) and centred on the *FXN* gene (S4 Fig. in Appendix A).

5.3.1 Random priming

250 ng of BAC DNA were biotin-labelled by random priming (BioPrime DNA labelling System, Invitrogen, Life Technologies), according to manufacturer's instructions. The central BAC RP11-265B8 harbouring *FXN* was also labelled with a custom-made nucleotide mix containing Cy5-AP3-dUTP (GE Healthcare) to allow its identification and orientation in molecules showing a probe pair only instead than the whole set of three probes. The success of the labelling procedure was monitored by gel electrophoresis. Amplified fragments were in a range of 100-1000 bp.

5.3.2 FISH on combed DNA

Per slide, 250 ng of each probe were mixed in 20 µl of hybridization solution (50% formamide, 1% N-Laurosylsarcosine, 10mM NaCl, 2X SSC in BlockAid (Invitrogen, Life Technologies) containing an excess (13X) of human Cot-1 DNA (Invitrogen, Life Technologies) and 10 µg Salmon Sperm DNA (Invitrogen, Life Technologies)). Denaturation was carried out at 80°C for 10 min. Slides were denatured for 15 min in 1M NaCl, 0.05M NaOH, immediately dehydrated in ethanol solutions (70%, 90%, 100%) on ice, hybridized with the probe mix (20 µl under 22 x 22 mm coverslips) for 19 hours at 37°C in a humidified chamber. Stringency washes were: 3 x 5 min in 50% formamide, 2X SSC pH 7.0 followed by 3 x 5 min in 2X SSC pH 7.0. To localize hybridisation signals along with replication tracks, a three-colour scheme of immunodetection was used: biotinylated probes were detected in green, whereas IdU and CldU in blue and red, respectively (Fig. 6). Three layers of antibodies were applied (30 min each) and reported in Table 3: in the first one, 488 Alexa Fluor-conjugated streptavidin (1:50; Molecular Probes, Invitrogen) allowed probe detection and two primary anti-BrdU antibodies were used cross-reacting respectively with IdU (2:7; Becton-Dickinson, developed in mouse) and CldU (1:40; Abcam, developed in rat). In the second layer, a polyclonal biotin-conjugated anti-streptavidin antibody (1:50; Rockland, USA), a 350 Alexa Fluor-conjugated anti-mouse IgG (1:50; Molecular Probes, produced in goat) and a 594 Alexa Fluor-conjugated anti-rat IgG (1:50; Molecular Probes, produced in donkey) were mixed. In the third layer, 488 Alexa Fluor-conjugated streptavidin mixed with 350 Alexa Fluor-conjugated anti-goat IgG (1:50; Molecular Probes, made in donkey) were used to complete the amplification steps. Cy5-labelled probes do not require amplification of the hybridisation signal.

Table 3. Antibodies employed for the single-locus replication analysis on combed DNA. In brackets host species are displayed. The relative dilutions are indicated. SAV=streptavidin; biot=biotin.

	I layer	**II layer**	**III layer**
IdU	Anti-BrdU (mouse) 2:7	Anti-mouse IgG 350 (goat) 1:50	Anti-goat IgG 350 (donkey) 1:50
CldU	Anti-BrdU (rat) 1:40	Anti-rat IgG 594 (donkey) 1:50	
Biotin	SAV-488 1:50	Anti-SAV biot 1:50	SAV-488 1:50

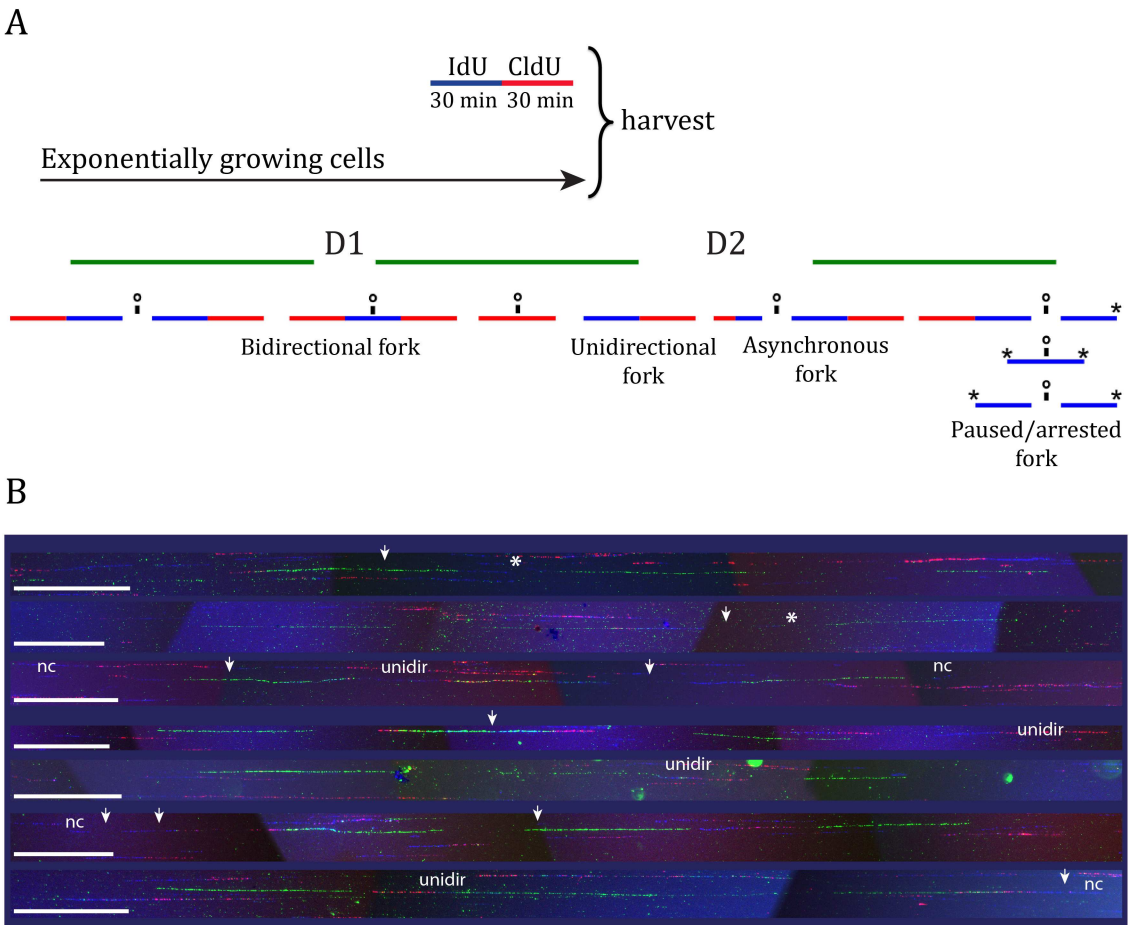


Fig. 6. (A) Experimental design and criteria for molecular combing analyses performed at single loci. A two-pulse labelling scheme is applied for the detection of replication forks. In the first pulse, IdU is incorporated in the nascent strands to be visualised by blue fluorescence; during the second pulse, CldU is supplied for DNA synthesis and it will be detected by red fluorescence. Three probes differentially spaced (D1 and D2) are detected by FISH with green fluorescence. The FISH pattern allows defining the centromere-telomere orientation and the integrity of the molecule. When only two hybridization signals are visible, the centromere-telomere orientation can be inferred because the central probe is co- labelled with Cy5 and biotin. Bidirectional origins (o) may be mapped in the middle of the two opposite arms or in the middle of a blue- or red-only track if a replication fork fires during the first or the second pulse, respectively. Forks with a single arm moving with the same orientation than upstream or downstream replication tracks are classified as unidirectional forks. Paused/arrested forks (*) are identified by bidirectional origins lacking the red fluorescence in one or both arms. Forks firing from the origin with different arm rate are classified as asynchronous forks. (B) Representative images of the *FXN* locus as detected by FISH and molecular combing. Three probes (green fluorescence) differentially spaced are used to identify the genomic region harbouring *FXN*. Replication tracks are visualized in blue (IdU) and red (CldU) fluorescence. Bidirectional origins are indicated by arrows, asterisks correspond to paused/arrested forks. Unidirectional forks are also displayed (unidir). nc = tracks not considered in the course of the analysis. Calibration bar = 100 kb.

6. Image analysis

A motorized fluorescence microscope (Zeiss Axio Imager.M1) equipped with a CCD camera (Photometrix, Coolsnap HQ2) was used for microscope analyses.

6.1 Interphase FISH analysis

Interphase FISH analysis was carried out under a 100X oil immersion objective (N.A. = 1.30) and more than 250 nuclei were scored per each experiment. Replication timing was determined on the basis of the observed hybridisation signals: non-replicated alleles are represented by single spots (S), whereas duplicated signals (D) referred to replicated alleles (Fig. 7). Thus, nuclei were classified as SS when both alleles were not replicated, SD when only one allele completed replication and DD if both alleles were already replicated (Fig. 7). In parallel and by blind analysis, CldU-positive nuclei were scored and classified as early, mid and late S-phase according to their fluorescent patterns (O'Keefe et al. 1992). The hybridisation efficiency was calculated by the formula: [SS+SD+DD+1/2 (S+D)/total number of scored nuclei] x 100 (Karnani et al. 2010).

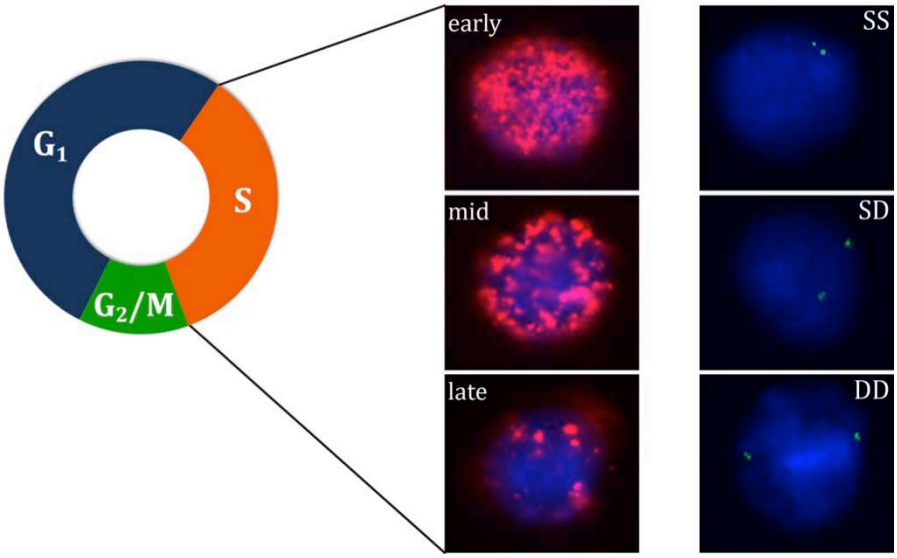


Fig. 7. The replication timing of single genomic loci is assessed by the interphase FISH approach. Cells are labelled with CldU for 30 min before harvesting. Replicating interphase nuclei are identified by CldU-immunodetection and classified in early, mid and late S-phase according to the observed fluorescent patterns. In the examples, CldU is displayed by red fluorescence, whereas DNA is counterstained with DAPI (blue fluorescence). In parallel, the replication timing of individual genomic regions is determined by hybridisation of locus-specific probes (green fluorescence), under the assumption that single FISH spots (S) define non-replicated alleles, while sequences that have completed their replication are represented by duplicated FISH signals (D). In the examples: SS = nuclei with two single FISH spots (non-replicated alleles); SD = nuclei with one single and one duplicated FISH signal (only one allele has been replicated); DD= nuclei with two duplicated FISH signals (both alleles have been replicated).

6.2 Molecular combing analysis

Molecular combing analyses were performed using a 40X oil immersion objective (N.A. = 1.30). Adjacent fields were acquired under adequate filter sets, then merged and aligned using the Adobe Photoshop CS2 software, allowing the reconstruction of whole DNA molecules spanning several Mb. Fluorescent patterns were measured by the Metavue Research Imaging System (Molecular Devices), according to the molecular combing calibration factor (1 μm = 2 kb) and the magnification provided by the objective and the CCD camera (1 pixel = 0.16125 μm).

6.2.1 Genome-wide molecular combing analysis

The lengths of replication tracks and their spatial organisation define the replication profile (fork rate, inter-origin distances, initiation sites, size of the replication clusters). Expected fluorescent patterns are depicted in Fig. 5 and stringent criteria were applied for their correct interpretation. In brief, only complete bidirectional forks were used to calculate fork speed, while alternative replication patterns including asynchronous and paused/arrested forks were discarded from this calculation. Unidirectional forks were recorded only if they moved with the same orientation of upstream/downstream replication tracks. Upstream/downstream replication forks are also requested features to classify unidirectional forks. Bidirectional forks displaying no red fluorescence in one or both arms may be regarded as either normal pausing events or abnormal arrest of fork progression, but unfortunately they cannot be discriminated. More details can be found in (Palumbo et al. 2013).

6.2.1 Single-locus molecular combing analysis

To orientate molecules, probe length and probe-to-probe distances were determined. Only molecules showing the hybridisation of at least two probes were considered informative for the detection of the replication activity within *FXN* and the surrounding 850 kb genomic region. The fraction of replicating molecules was calculated as the ratio between the number of molecules displaying replication signals and the total number of observed molecules. Fork rates, inter-origin distances and replication origin positioning define replication profiles. To correctly interpret fluorescent signals, stringent criteria were applied according to those described in details elsewhere (Palumbo et al. 2013; Palumbo et al. 2010). Expected replication patterns are shown in Fig. 6. Briefly, genomic DNA was not counterstained, hence only replication signals in a linear array and framed by fluorescent probe were considered, as they belong with maximum confidence to the same single molecule. Similar to genome-wide replication analyses, fork rates were calculated only with complete bidirectional forks and all other patterns,

including unidirectional, asynchronous and paused/arrested forks, were considered only when upstream or downstream replication tracks supported the presence of uninterrupted filaments. Blue-only tracks were interpreted, according to the whole replication pattern along the molecule, as termination events occurring during the first pulse OR paused/arrested forks. More information in (Palumbo et al. 2013).

7. Small nascent strand abundance assay and quantitative real-time PCR

Small Nascent Strand (SNS) abundance assay was employed with some modifications to (Gray et al. 2007), as shown in Fig. 8. In particular, 70 x 10^6 cells derived from the GM15850, GM15851, GM16227 and H691 lymphoblastoid cell lines were washed with 1X PBS and collected in 240 µl of 10% glycerol/PBS. 60 µl of each cell sample were lysed directly into a well of a 1.25% alkaline agarose gel (50mM NaOH, 1mM EDTA) for 15 min at 4°C (Fig. 8B). The electrophoresis was carried out under the same denaturing conditions for 5–6 h at 30 V, neutralised in 1X TAE and stained with ethidium bromide (Fig. 8B). DNA fragments ranging 0.5–1.5 kb were purified from the gel using a QIAEX II Gel extraction kit (Qiagen) (Fig. 8B). Genomic DNA from the same cell populations used for the SNS isolation was purified by applying the standard phenol-chloroform-isoamyl alcohol method, digested with 0.4 mg/ml Proteinase K and treated with 20 µg/ml RNase A. SNS and genomic DNA were quantified by NanoDrop 1000 (ThermoScientific). Quantitative real-time PCR (qPCR) was performed with 0.2 µM of each primer and the Power SYBR Green PCR Master Mix (Life Technologies) in an Applied BioSystems 7500 Real-Time PCR System (Life Technologies). Primer binding sites span the whole *FXN* gene and two upstream and downstream regions chosen as controls on chromosome 9. In addition, origin/non origin sites previously characterised around the *LAMIN B2* gene (Abdurashidova et al. 2000; Giacca et al. 1997) were used as further controls to test the reliability of the quantitative PCR assay. Primer pairs are listed in (Stevanoni et al. 2016) and their positions relative to the *FXN* gene are shown in Figs. 3 and 7A in Appendix A. For each primer set, a 6 point standard curve derived from 1:3 serial dilutions of genomic DNA was amplified in each plate and for each cell line, starting from 4.5 x10^4 copies of genomic equivalents. Standard curves were run in triplicate, while SNS samples were run in five-eight replicas. The mean quantity of genomic equivalents in the short nascent DNA samples was determined in comparison to the correspondent standard curve. Amount of SNS at each primer set was normalized both to the average quantity of SNS estimated for each cell line in the control regions on chromosome 9 (primer sets C1-C3) and to the non-origin site *LB2C1*, and further corrected according to the enrichment estimated at the *LAMIN B2* origin. This latter normalisation procedure was used to set a threshold establishing the enrichment fold for each primer set in *FXN* and control regions (primer sets F1-F4, C1-C3).

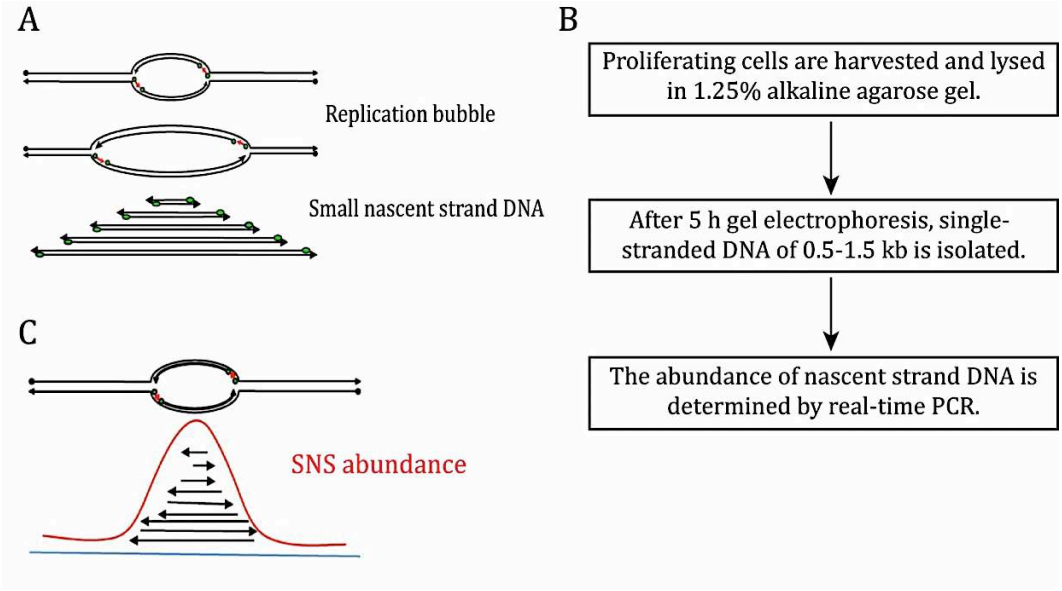


Fig. 8. (A) When a replication origin is activated, the MCM complex unwinds the double helix and DNA polymerases start to synthesise new DNA from the two single strand templates, forming the so called replication bubbles. Replication intermediates formed at replication origins include small nascent strand DNA, still bound to the RNA primer (green dot). (B) Brief flowchart of the SNS isolation procedure. Exponentially growing cells are harvested and directly lysed into a well of an alkaline agarose gel. Denatured fragmented DNA ranging 0.5-1.5 kb is isolated and the abundance of nascent strand is defined by quantitative real-time PCR. (C) The position of replication origins may be mapped by determining the enrichment in small nascent strand DNA. A peak in SNS abundance may be found at active replication origins. Adapted from (Hyrien 2015).

RESULTS

1. Single-locus analysis of the *Frataxin* gene in the presence of a GAA-repeat expansion

It is largely accepted that trinucleotide repeat expansions lead to formation of stable unusual DNA secondary structures, which in turn represent a well-known source of replication stress (Mirkin & Mirkin 2007; Kim & Mirkin 2013; Cleary & Pearson 2005; Alver et al. 2014; Magdalou et al. 2014; León-Ortiz et al. 2014; Zeman & Cimprich 2014). The effects of an expanded trinucleotide repeat on DNA replication in the endogenous context are still unknown. Here, I evaluate how replication dynamics is modulated in the presence of a GAA-repeat expansion in intron 1 of the *Frataxin* (*FXN*) gene causing the Friedreich's ataxia disease.
A manuscript was submitted to Plos Genetics on June 2015 and additional experiments were carried out during revision. The revised paper was accepted on June 2016 and it is included in the Appendix A (Stevanoni et al. 2016).

1.1 Characterisation of the cellular model

Two human cell lines derived from unrelated Friedreich's ataxia (FRDA) patients carrying a homozygous GAA-repeat expansion (GM15850 and GM16227) were used as experimental model and analysed in comparison to GM15851 control cells derived from the healthy brother of GM15850. All cell lines were obtained from the Coriell cell line repository (Camden, New Jersey, USA). Genotype, transcriptional, cell cycle and replication analyses were carried out for the comprehensive characterisation of the three cell samples (S1 Fig. in Appendix A). Somatic instability of the GAA-repeat expansion has been reported in several tissues of FRDA patients (Sharma et al. 2002; Hellenbroich et al. 2001; De Biase et al. 2007) and it has also been observed in mutated lymphoblastoid cell lines (Bidichandani et al. 1999; Ditch et al. 2009). Hence, I determined the length of the GAA-repeat expansion in GM15850 and GM16227 patients' cells by long-range PCR both at the beginning and the conclusion of my work. Results were in line with the information provided by the Coriell cell repository (S1A and S1B Fig. in Appendix A) and the lack of multiple bands relative to the amplification of the expanded GAA-repeats allowed to exclude the presence of somatic instability in FRDA cells used in this study (S1B Fig. in Appendix A). Furthermore, the expected transcriptional silencing of the expanded *FXN* gene was observed in patients' cell lines (S1C Fig. in Appendix A). No detectable differences of flow cytometry-based cell cycle distributions were found between FRDA and control cells (S1D Fig. in Appendix A). In addition, evaluation of replication profiles by molecular combing (S1E Fig. in Appendix A)

displayed fork rates and inter-origin distances (IOD) falling into the ranges known for lymphoblastoid cell lines (Palumbo et al. 2013; Cadoret et al. 2008).
According to the requests advanced during the revision of the manuscript, genotype and transcriptional analyses were carried out on a further control cell line (S1B and S1C Fig. in Appendix A), the EBV-immortalized B lymphoblastoid H691 cells, derived from a male subject and previously employed in our laboratory for replication studies (Palumbo et al. 2013).

1.2 Evaluation of the replication timing of the *FXN* gene

Replication timing along S-phase of mammalian cells has a crucial role in the modulation of the process. According to data provided by the ENCODE project, *FXN* is harboured in a mid-late replicating domain (Birney et al. 2007). However, the effects of long GAA-repeat expansion (ranging 630–1030 repeats in my samples; S1A Fig. in Appendix A) on replication timing were never considered. In this frame, interphase FISH experiments were carried out to assess the replication timing of the *FXN* gene in both mutant cells (GM15850 and GM16227 cell lines) and in the control cell line GM15851, under the assumption that nuclei showing two single FISH spots (SS) carried non-replicated alleles, while cells showing a pair of duplicated FISH signals (DD) had already completed the replication of both alleles (Fig. 7). Asynchronous patterns with a single and a duplicated FISH signal (SD) are also expected (Fig. 7; (Karnani et al. 2007)). In FRDA and control cells, these patterns were similarly represented independent of the presence of the GAA-repeat expansion, indicating that no differences exist between mutant and normal *FXN* alleles (Appendix A, Fig. 1 and S1 Table). In parallel, the late replicating sequence of the common fragile site *FRA3B* was evaluated as a positive control and less than 25% nuclei displayed a DD pattern at this locus (Appendix A, Fig. 1 and S1 Table). Taken together these results suggested that *FRA3B* replication is slightly postponed with respect to *FXN*, which showed indeed a mid-late replication timing, in agreement with previous data (Birney et al. 2007).
However, the sensitivity of the interphase FISH approach could not be sufficient to detect mild temporal shifts, in particular if mid-late or late replicating regions are investigated. Indeed, mid and late replication domains are characterised by the activation of less efficient and more stochastic origins with respect to those in early genomic regions, leading to an increased cell-to-cell variability (Dellino et al. 2013; Hansen et al. 2010). Therefore, I reasoned that coupling interphase FISH with FACS sorting experiments could allow examining in depth the *FXN* replication timing (Fig. 3). Four cell fractions of identical size and corresponding to early-to-late S-phase (S1-S4) were isolated with a very low level of contamination recorded by post-sorting FACS analysis (S2 Fig. in Appendix A). In addition, sorted cells were CldU-labelled immediately before harvesting, so that in the course of the following FISH analyses nuclei could be classified as early, mid or late S-phase by CldU-immunodetection, giving a second control of the accuracy of the cell separation (S3

Fig., S2 Table in Appendix A). Consistent with the post-sorting FACS analyses (S2 Fig. in Appendix A), in each fraction the large majority of the cells were positive to CldU labelling; moreover, early-mid nuclei were found in S1-S2 fractions, whereas mid-late CldU-labelling patterns were associated with S3-S4 S-phase substages.

Interphase FISH data concerning the normal cell line GM15851 were obtained in two independent sorting experiments and were remarkably reproducible (S2 Table in Appendix A for raw data). In Table 1 in Appendix A the summary of these analyses was reported. According to the percentage of DD nuclei observed in S2, only 25% of GM15851 cells have completed the replication of the wildtype *FXN* locus within the first half of the S-phase (Table 1 in Appendix A), further supporting the indication of a mid-late replication timing of this sequence. In S1 and S2 fractions the replication patterns of normal and mutant cells were significantly different when compared by chi-square analysis ($P<0.05$ for early S-phase cells; $P<0.001$ for mid S-phase cells). In particular, the excess of SS nuclei persisting in mutant S2 cells demonstrated a faster replication progression of normal than expanded *FXN* alleles (Appendix A, Table 1 and S2 Table). When the S3 fraction was considered, a statistical difference still existed between replication patterns of late S-phase wildtype and mutant cells ($P<0.05$). In the S4 fraction replication patterns of mid S-phase mutant and normal cells were significantly different ($P<0.001$). These latter observations can be interpreted as a downstream effect of the earlier shift of the replication timing, as the proportions of SS nuclei were not involved in these variations (Table 1 in Appendix A). In fact, in the second half of the S-phase both normal and expanded *FXN* alleles are undergoing and completing their replication.

As a positive control, the late replicating sequence at *FRA3B* was analysed also in GM15851 control and GM15850 mutant sorted cell populations (S3 Table in Appendix A). *FRA3B* was confirmed to replicate later than the wildtype *FXN* locus, as in S1 and S2 fractions less than 15% nuclei displayed a pair of replicated *FRA3B* alleles (DD pattern). Furthermore, no differences were detected between control and mutant cells in all sorted S-phase fractions, thus indicating that replication patterns of expanded *FXN* alleles were strictly comparable to those of the late replicating *FRA3B* (Appendix A, Table 1 and S3 Table).

To evaluate whether the effects of the GAA-repeat expansion are restricted to the *FXN* locus, I analysed the replication patterns of a region located about 170 kb downstream the *FXN* locus and identified by BAC RP11-548B3 (S4C Fig. in Appendix A). S2 and S3 nuclei displayed the same replication patterns in normal GM15851 and mutant GM15850 cells (S4 Table in Appendix A). This led me to conclude that the shift of replication timing occurring in the presence of an expanded repeat did not involve a wide genomic region.

1.3 Single-locus analysis of the replication profile of *Frataxin*

To evaluate whether and how replication profiles are affected by the presence of the GAA-repeat expansion in the endogenous genomic context, molecular combing was performed in normal (GM15851 and H691 cell lines) and mutant (GM15850 and GM16227 cell lines) cells. Origin firing and replication dynamics were monitored within a 850 kb region centred on the *FXN* gene (S4 Fig. in Appendix A). The fraction of replicating molecules was higher than 50% in all cell samples and it was comparable between normal and mutant cell lines (Table 2 in Appendix A). At least 100 replication forks were scored and classified per each cell sample (Table 2 in Appendix A). Fork rates and Inter-Origin Distances (IOD) were similar in all cell lines, as confirmed by the Kruskal-Wallis non-parametric test (Appendix A, Table 2 and Fig. 2).

An intrinsic feature of mammalian DNA replication is the high flexibility of origin choice. Indeed, the position of activated replication origins vary among cells in a population during subsequent cell cycles (Fig. 1; (Méchali 2010; Hyrien et al. 2003)). In agreement, within the investigated genomic region origin firing occurred with wide molecule-to-molecule variability, as can be appreciated per each cell line in S5-S8 Figs. in Appendix A. A synoptical view showing the position of all the bidirectional origins mapped for the four cell lines under study is shown in Fig. 3 in Appendix A.

Looking at the whole 850 kb genomic region delimited by the three BAC probes, changes in origin choice and a differential distribution of activated bidirectional origins were detected in mutated versus normal cells (Fig. 3 in Appendix A). Dealing with each probe separately, origin selection changed both upstream and downstream *FXN* in the expanded alleles (Fig. 3 in Appendix A). However, the most intriguing differences among samples emerged when focusing on the region I am more interested in, the central BAC RP11-265B8 harbouring the *FXN* gene. Indeed, in normal GM15851 cells 18 replicating molecules had at least one bidirectional origin firing within that region, 20 origins were mapped in total, but none of them fired inside the *FXN* gene (Appendix A, Fig. 3 and S5 Fig.). The same pattern was found also in the second control, the H691 cells, where a total of 20 bidirectional origins in 16 replicating molecules were detected inside the region identified by BAC RP11-265B8, but no origin firing was found within *FXN* (Appendix A, Fig. 3 and S6 Fig.). By considering the orientation of the replication forks running through the short GAA-repeat it appeared that this sequence was prevalently, but not exclusively, the template for the lagging strand. In contrast, in both mutant cell lines several molecules showed one origin firing within the *FXN* allele carrying the expanded GAA-repeat (Appendix A, Figs. 3-4 and S7–S9 Figs.). In particular, in GM15850 cells I found 14 molecules showing at least one bidirectional origin in the region identified by BAC RP11- 265B8, for a total of 19 origins mapped within this genomic sequence (Appendix A, Fig. 3 and S7 Fig.). Seven of these origins, each of them firing in an independent molecule, were located within the *FXN* gene in

different positions (Appendix A, Fig. 3 and S7 Fig.). A similar profile was seen in GM16227 cells, where 17 bidirectional origins (from 11 molecules) were mapped within the central BAC RP11- 265B8, and 9 of them (from 9 diverse molecules) were located in different positions of the *FXN* gene (Appendix A, Fig. 3 and S8 Fig.). In consequence of dormant origin activation within *FXN*, the proportion of forks replicating the GAA-repeat from a downstream origin, i.e. from the leading strand, is higher than in the wildtype alleles (S5–S8 Figs. in Appendix A). Furthermore, in the region identified by BAC RP11-265B8 the number of bidirectional origins per replicating molecule was respectively 1.11 in GM15851, 1.25 in H691, 1.36 in FRDA GM15850 and 1.55 in FRDA GM16227. These data suggest a mild increase of origin firing in the presence of the GAA-repeat expansion, indicating that dormant origins do not substitute initiations occurring in the normal alleles, while they fire as additional events.

Beyond the detection of bidirectional origins, also unidirectional forks and pause/arrest sites were evaluated by molecular combing in the 850 kb genomic region under study. Replication forks with unidirectional progression were observed in proportions ranging from 19 to 32.5% in the different cell lines (Table 2 in Appendix A) and they appeared evenly distributed along the genomic region investigated (Fig. 5A in Appendix A). However, a prevalence of short unidirectional forks was noted in FRDA cells compared to the average length detected in normal ones and this difference was particularly evident in the central segment harbouring the *FXN* gene (Fig. 5A in Appendix A). Since the actual origin position cannot be defined when forks run unidirectionally, it is not correct to calculate their speed. Hence, the length of unidirectional tracks entirely running in the region including the central BAC and the flanking probe-to-probe distances were measured (Fig. 5A in Appendix A). A marked length reduction was detected in FRDA cells when compared to the normal GM15851 and H691 cell lines and this difference is statistically significant according to the Kruskal-Wallis non-parametric test ($P<0.01$; Fig. 5B in Appendix A). In particular, average lengths with standard errors were respectively: 113.7 ± 12.41 kb in normal GM15851 cells, 102.8 ± 8.56 kb in normal H691 cells, 67.5 ± 9.81 kb in FRDA GM15850 cells, 70.5 ± 11.83 kb in FRDA GM16227 cells (Fig. 5B in Appendix A). In addition, according to the coefficient of variation (CV), unidirectional fork length measures are less dispersed in control cells (CV about 35%) than in FRDA ones (CV about 60%). The statistical difference in the length distribution of unidirectional forks persisted between FRDA and wildtype cells also when the whole 850 kb genomic region was considered ($P<0.005$ Kruskal-Wallis non-parametric test). In this case the average lengths with standard errors were respectively: 115.9 ± 9.02 kb in normal GM15851 cells, 103.6 ± 5.67 kb in normal H691 cells, 76.9 ± 5.43 kb in FRDA GM15850 cells, 84.3 ± 7.30 kb in FRDA GM16227 cells. The magnitude of the CVs associated with the distribution of unidirectional forks in the whole 850 kb region remains higher in FRDA cells (although CVs decrease to values of about 45%) than in the controls (about 35%). Finally, several events of pause/arrest of the fork were observed

within the *FXN* locus and in the adjacent sequences. In GM15851 control cells paused/arrested forks were frequently recorded in the proximity of the short GAA-repeat, while less sites of pause/arrest were found in H691 cells as well as in both mutant cell lines at the long GAA-repeat (Fig. 6 in Appendix A).
On the whole, these results demonstrate that the normal *FXN* sequence is passively replicated by incoming replication forks and the short GAA-repeat is the prevailing template for the lagging strand synthesis. Instead, in the presence of the expanded GAA-repeat, several changes of the replication profile, including recruitment of additional origins within the gene, widespread changes in origin choice, a differential distribution of unidirectional forks, provide the basis for assuring the completion of DNA replication. As a consequence of dormant origin activation in the expanded alleles, a switch of the direction by which replication forks proceed through the GAA-repeat is frequently observed with respect to the normal sequence.

1.4 Alternative strategies to assess the replication activity of the *Frataxin* locus

As already described in the Introduction, to map mammalian replication origins several techniques based on the detection of replication intermediates have been developed (Hyrien 2015; Prioleau & MacAlpine 2016). During the revision of the *FXN* manuscript (Appendix A) the referees requested to confirm the replication profiles observed at *FXN* in normal and mutant cells, and in particular the activation of dormant origins within the expanded *FXN* alleles, by using techniques different from molecular combing. For this reason, I performed small nascent strand (SNS) abundance assay coupled with quantitative real-time PCR both in the normal and the expanded *FXN* locus (Fig. 8 in Material and Methods).
To carry out the SNS abundance assay under optimal conditions an origin-free region should be used to normalise the SNS amounts obtained per each primer set (Abdurashidova et al. 2000; Giacca et al. 1997; Gray et al. 2007). According to molecular combing data (Fig. 3 in Appendix A), initiation events are widespread within the 850 kb sequence harbouring the *FXN* gene and an origin-free region shared by the four cell lines cannot be firmly identified. Hence, in these experiments qRT-PCR quantities were normalised versus an origin-positive sequence, a validated alternative approach to analyse SNS abundance experiments (Gray et al. 2007; Cadoret et al. 2008). Recurrent pattern of origin activation among all cell lines may be inferred by the molecular combing analysis in two positions located upstream and downstream the *FXN* gene respectively, and were employed to design primer sets C1-C3 (S5 Table and Figs. 3 and 7A in Appendix A). Concerning the *FXN* gene, four primer sets (F1-F4; S5 Table and Figs. 3 and 7A in Appendix A) were designed to analyse two sites downstream the GAA repeat. The enrichment fold at the *LAMIN B2* origin (Abdurashidova et al. 2000; Giacca et al. 1997), calculated as a quality control of each SNS isolation experiment, ranged 5–

119 (two independent experiments for each cell line) and these values were used to set the threshold to estimate the SNS enrichment in FRDA and control cells as described in Material and Methods (p. 36). When the average SNS quantities in the control regions C1-C3 of chromosome 9 were used as the normalising factor for the values estimated within the *FXN* gene (F1-F4), no differential patterns were detectable between FRDA and control cells (Fig. 7B in Appendix A). However, when data are normalised against a sequence with origins, a flattening effect can be produced resulting in the prevalence of the background noise over real differences in origin activation, particularly if low-efficient events are considered as in this case (Dellino & Pelicci 2014). To overcome this limitation, which could be responsible for the lack of differentiation visible in Fig. 7B in Appendix A, SNS quantities were normalised to the non-origin site *LB2C1*. This value was further corrected by the mean enrichment of *LAMIN B2* origin, estimated for each cell line, as described in Materials and Methods (p. 36). Despite the lack of any specific trends when considering each cell line separately (Fig. 7C in Appendix A), by pooling data of FRDA or normal cell lines it appeared a weak differential response coherent with a lack of initiation events within the normal *FXN* alleles: indeed, SNS quantities detected in normal cells within the *FXN* gene (primers F1-F4) remained under the threshold, while the threshold was reached at the control sites (C1-C3) in all cell lines and at *FXN* in FRDA cells (Fig. 7D in Appendix A). On the whole, data shown in Fig. 7 in Appendix A indicate that the SNS abundance assay is not sensitive enough to confirm the activation of dormant origins during the replication of the expanded *FXN* alleles.

During the revision of the *FXN* manuscript (Appendix A), data obtained by the novel OK-seq approach, based on Okazaki fragment sequencing, were published providing a global description of the replication landscape in a normal lymphoblastoid cell line (GM06990) and in HeLa cells (Petryk et al. 2016). Thus, to further confirm my results concerning the activation of replication origins, I analysed OK-seq profiles in the genomic region harbouring *FXN* as represented in S10 Fig., Appendix A. By sequencing Okazaki fragments, the proportions of rightward- and leftward-moving forks are determined and displayed in OK-seq profiles as red and blue spots, respectively (S10 Fig. in Appendix A). According to replication fork directionality, zones of replication initiation and termination can then be defined as upward and downward slopes, respectively (Petryk et al. 2016). I found that in both the lymphoblastoid cell line and the HeLa cells, *FXN* is associated with a termination region (downward slope) bordered by two initiation zones (upward slope) upstream and downstream the gene, displayed as b and c in S10 Fig. in Appendix A. This OK-Seq profile strictly agreed with my molecular combing analysis of GM15851 and H691 cell lines and according to the model described in (Petryk et al. 2016) it correlated with the presence of forks emanating from the surrounding initiation zones and converging at different positions within the *FXN* gene body. Furthermore, according to the published criteria (Petryk et al. 2016), the large termination zone delimited by the initiation zones a and b

corresponded to a cascade of termination events associated with the firing of background origins (S10 Fig. in Appendix A). This is again coherent with my molecular combing results (S5–S8 Figs. in Appendix A). Thus, the analysis of OK-Seq profiles verified that in the absence of the GAA-repeat expansion *FXN* is passively replicated by nearby incoming replication forks. Unfortunately and similarly to all genome-wide high-throughput techniques, OK-seq is weak efficient in identifying rare, isolated and widespread initiation events, such as activation of dormant origins, because their firing is not strong enough to produce a detectable initiation zone (Hyrien, personal communication). Hence, OK-seq cannot be successfully performed in FRDA cells to have a direct validation of my results.

Therefore, I can conclude that single-molecule approaches and in particular molecular combing are the most appropriate tools to identify rare and stochastic events characterising locus-specific changes of the mammalian replication program.

2. Modulation of DNA replication in response to alterations in dNTP pools

To evaluate the effects of increased and unbalanced deoxynucleotide triphosphates (dNTPs) pools on DNA replication in mammalian cells, human cell lines lacking SAMHD1, a protein involved in nucleotide precursor metabolism (Franzolin et al. 2013; Miazzi et al. 2014), were used as experimental model. In particular, two human skin fibroblast cell lines derived from Aicardi-Goutièrres syndrome (AGS) patients carrying mutations affecting the *SAMHD1* gene were analysed in comparison to C63 and Goteborg control cells (Table 4). In collaboration with the laboratory of Prof. Bianchi, the dNTP pool size was estimated in AGS and control cells: in AGS cell lines an higher dNTP supply was detected (both in quiescent and proliferating status) when compared to control C63 cells (Table 4). In parallel, replication dynamics and dNTP pools were also evaluated in human acute monocytic leukemia THP1 cells, by comparing the effects of SAMHD1 depletion (THP1 KO) with respect to wildtype cells (THP1 wt) (Table 4).

Table 4. Cell types and mutations affecting the SAMHD1 gene in the cell lines under study.

Cell line	**Cell type**	**Phenotype**	**Mutation in the *SAMHD1* gene**	**dNTP pool fold increase[#]**			
				dGTP	**dATP**	**dCTP**	**dTTP**
C63	Human skin fibroblast	wildtype	-	-	-	-	-
Goteborg	Human skin fibroblast	wildtype	-	-	-	-	-
GR295	Human skin fibroblast	Aicardi-Goutièrres syndrome	c. 1273_*2del	5.10	2.82	1.24	2.02
MG282	Human skin fibroblast	Aicardi-Goutièrres syndrome	c. 433C>T + 490C>T	3.44	1.88	0.83	1.09
THP1 wt	Human monocyte	Acute leukemia	-	-	-	-	-
THP1 KO	Human monocyte	Acute leukemia	CRISPR/Cas9 depletion of *SAMHD1*	4.06	3.98	2.39	1.27

[#] Fold increase is calculated as the ratio of dNTP pool sizes in AGS or THP1 KO cells with respect to the correspondent control.

2.1 Whole genome molecular combing analysis of AGS fibroblasts

To assess whether replication dynamics is affected by increased and unbalanced dNTP pools, two AGS cell lines were analysed genome-wide by molecular combing. Data are summarised in Table 5.

No differences were detected between AGS and control cells in terms of fork rates, inter-origin distances and cluster lengths (Table 5) and also their distributions were comparable, as indicated by the Kruskal-Wallis non-parametric test that returned not significant results (Fig. 9A-C). Furthermore, proportions of unidirectional forks and pause/arrest events were similar among all cell samples according to the chi-square analysis (Table 5) and they fell into the ranges observed in other mammalian cell lines (Palumbo et al. 2013). It is important to note that in genome-wide experiments the quality of DNA fibers was monitored by DNA counterstaining. In these analyses, mean and median lengths of analysed molecules were above the critical threshold for the correct estimation of replication parameters (Técher et al. 2013; Bianco et al. 2012) and they fell into the range known for mammalian cells (Table 5; (Kaykov et al. 2016)). Molecule length distributions were similar in normal and mutant cell lines (Fig. 9D), further indicating that the estimation of replication parameters was accurate. In conclusion, this set of data suggested that an increased availability of nucleotide precursors might not directly influence genome duplication.

On the basis of the yeast model, it seems that the expansion of dNTP pools promotes replication fork progression and improves cell survival under replication stress conditions (Poli et al. 2012). However, it is unknown in mammalian cells whether a high supply of nucleotide precursors may improve their tolerance to replication stress. Hence, I evaluated the replication profile of the GR295 AGS cell line in comparison to C63 control cells after treatment with aphidicolin (APH), an inhibitor of the replicative DNA polymerases. In agreement with previous studies (Palumbo et al. 2013; Courbet et al. 2008), after treatment with aphidicolin a significant reduction in fork rates, IODs and cluster size was observed both in GR295 mutant and C63 control cells with respect to their corresponding values under unperturbed growth conditions ($P<0.001$ for all parameters by Kruskal-Wallis non-parametric test; Table 5 and Fig. 11A-C). As expected (Courbet et al. 2008; Palumbo et al. 2013), in response to replication stress an higher number of origins was activated within each replication cluster in both cell lines when compared with their unperturbed counterparts ($P<0.001$ linear regression analysis; Table 5 and Fig. 11E-F). Interestingly, also the proportions of unidirectional forks increased under replication stress in both cell lines ($P<0.05$ by chi-square analysis; Table 5), as previously demonstrated (Palumbo et al. 2013). However, any differential response to aphidicolin could be appreciated when comparing AGS and control cells (Table 5 and Fig. 10A-C). When molecule lengths were monitored for the accuracy of the combing analyses (Table 5) a slight statistical difference was observed between their distributions in AGS and control

cells (P<0.05 Kruskal-Wallis non-parametric test; Fig. 10D), but this can be ascribed to an intrinsic variability among different cell populations. Importantly, in the APH-treated AGS and in control cell lines molecule lengths were 8-fold larger than their corresponding average IODs (Table 5). In agreement with published studies (Bianco et al. 2012; Kaykov et al. 2016) and based on the laboratory expertise in molecular combing analyses, to avoid biases, mean molecule length must be at least 3-4 fold larger than the estimated average IOD, which is together with cluster size the most susceptible parameter to fiber length variations. Thus, in spite of the slightly different molecule lengths found between cell samples, robust conclusions can be drawn by this set of data. Furthermore, fiber lengths were strongly comparable to that observed in the two cell lines analysed under physiological conditions (Table 5), as confirmed by the Kruskal-Wallis non-parametric test on their distributions (Fig. 11D). On the whole, results indicated that independently of the high availability of nucleotide precursors DNA replication proceeds mostly undisturbed both in AGS and control cells.

2.2 Whole genome molecular combing analysis in SAMHD1-depleted cells

In parallel to the molecular combing analysis of AGS fibroblasts, replication dynamics were also evaluated in THP1 wildtype and knockout cells. Raw data are reported in Table 5. Similarly to fibroblasts, the accuracy of replication parameters was verified by monitoring molecule lengths according to the critical threshold reported (Técher et al. 2013; Bianco et al. 2012) and on the basis of the laboratory expertise in molecular combing analyses. Molecule length distributions in THP1 wt and KO cells were comparable, as demonstrated by the Kruskal-Wallis non-parametric test (Fig. 12D). In contrast, when dealing with fork rates, IODs and cluster size, a pronounced reduction was detected in THP1 KO cells with respect to wildtype (Table 5). These differences were highly significant for all replication parameters, as determined by the statistical analysis of their distributions (P<0.001 Kruskal-Wallis non-parametric test; Fig. 12A-C). In addition, consistent with the cellular response to replication stress in mammals (Anglana et al. 2003; Courbet et al. 2008; Blow et al. 2011), the number of activated origins per cluster significantly increased in THP1 KO than wt cells (P<0.01 by linear regression analysis; Fig. 12E). Notably, the proportion of paused/arrested forks in THP1 KO cells was twice than that observed in the wildtype cell line, despite the lack of the statistical significance (chi-square analysis; Table 5).
Taken together, these results indicated that SAMHD1 depletion and the associated imbalance of the dNTP pools strongly affect replication dynamics of THP1 cells, differently from AGS fibroblasts.

Table 5. Molecular combing analyses in AGS fibroblasts and THP1 cells.

		Control			Aicardi-Goutièrres syndrome			THP1 cells	
		C63		Goteborg	GR295		MG282	wildtype	Knock-out
		APH			APH				
		-	+		-	+			
Replication forks	Total	166	208	221	171	207	150	228	229
	Unidirectional forks (N)	42	78	52	41	69	35	58	54
	Unidirectional forks (%)	25.3	37.5	23.5	24	33.3	23.3	25.4	23.6
	Paused/arrested bidirectional forks (N)[ç]	15	19	18	15	22	18	8	16
	Paused/arrested bidirectional forks (%)	9	9.1	8.1	8.8	10.6	12	3.5	7
	Asynchronous bidirectional forks (N)	6	1	4	5	2	3	1	2
	Asynchronous bidirectional forks (%)	3.6	0.5	1.8	2.9	1	2	0.4	0.9
Fork rate (kb/min)[§]	Mean ± SE	2.07 ± 0.073	0.59 ± 0.017	2.03 ± 0.053	2.09 ± 0.069	0.62 ± 0.018	2.11 ± 0.086	1.37 ± 0.033	0.68 ± 0.019
	Median	2.10	0.56	2.03	2.09	0.64	2.10	1.34	0.66
	N	108	110	147	110	114	94	161	157
	Min	0.25	0.16	0.55	0.70	0.21	0.36	0.43	0.23
	Max	3.88	1.18	3.96	4.01	1.11	4.18	2.80	1.55
Inter-origin distance (kb)	Mean ± SE	228.1 ± 11.57	144.6 ± 15.48	236.5±12.30	259.2 ± 14.88	157.3 ± 16.21	207.6±11.96	186.8 ± 8.97	127.1 ± 9.70
	Median	204.1	102.4	202.1	229.7	108.1	178.7	168.1	89.5
	N	130	66	147	101	90	123	159	132
	Min	25.1	22.3	29.4	47.8	10.8	28.4	21.3	14.9
	Max	668.6	645.2	884.4	814.7	809.2	707.5	592.8	542.0

		Control			Aicardi-Goutièrres syndrome			THP1 cells	
		C63		Goteborg	GR295		MG282	wildtype	Knock-out
		APH			APH				
		-	+		-	+			
Cluster size (kb)	Mean ± SE	818.0±53.03	489.2±51.62	987.7±61.49	957.7±61.49	571.9±47.95	968.8±70.76	895.1±61.95	652.7±60.23
	Median	754.6	411.1	969.8	864.4	570.7	801.5	822.9	538.7
	N	71	56	67	62	54	46	59	52
	Min	223.4	86.7	228.6	248.8	50.1	392.1	183.8	106.5
	Max	2310.9	2016.6	2689.4	2432.0	1349.0	2098.6	2114.7	1774.9
Molecule length (kb)	Mean ± SE	1149.1±52.81	1172.4±60.77	1534.3±63.44	1319.4±76.97	1379.6±67.31	1360.0±81.61	1147.6±55.98	1082.3±62.83
	Median	1072.6	1104.7	1505.8	1183.2	1406.9	1318.6	1026.6	994.9
	N	71	56	67	62	54	46	59	52
	Min	349.8	309.5	502.8	305.0	421.3	538.9	533.2	319.7
	Max	2329.9	2287.8	3029.0	2953.4	2536.6	2643.5	2164.1	2250.1
Number of activated origins per cluster	Mean	3	3	4	3	4	5	5	5
	N	241	191	274	215	217	212	296	268
	Min	2	2	2	2	2	2	2	2
	Max	8	10	9	9	14	10	13	17

§ Only bidirectional forks, excluding asynchronous and paused/arrested forks.

ç Replication forks with bilateral pause/arrest are considered as a single event because associated with the firing of one origin.

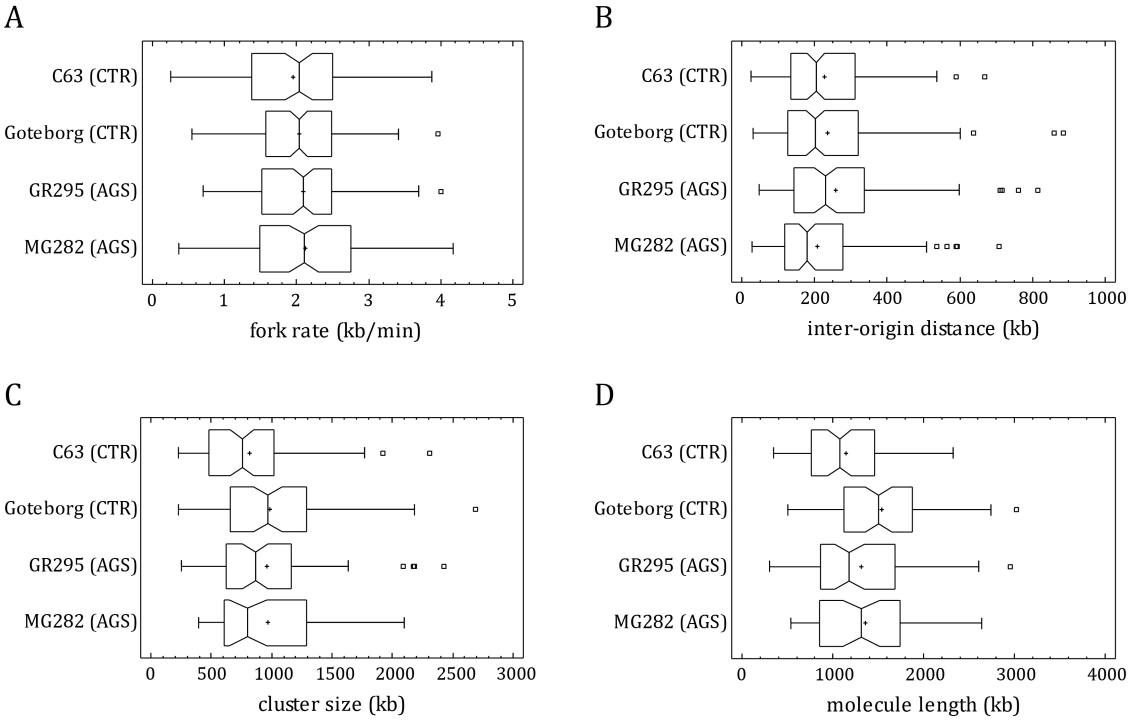


Fig. 9. Distributions of fork rates (A), inter-origin distances (B), cluster size (C) and molecule lengths (D) in AGS and control cells in unperturbed physiological conditions. No significant differences were observed for all parameters (Kruskal-Wallis non-parametric test).

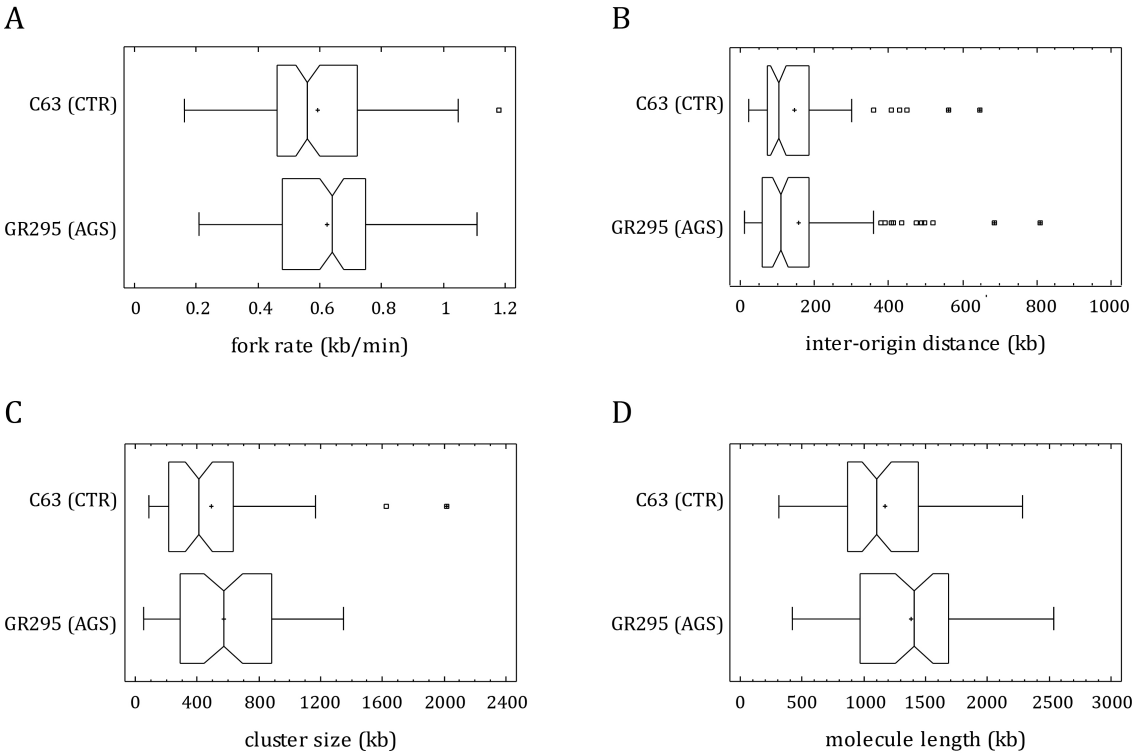


Fig. 10. Distribution of fork rates (A), inter-origin distances (B), cluster size (C) and molecule lengths (D) in GR295 AGS cells and in C63 control cell line exposed to aphidicolin. A statistical difference was detected between molecule lengths of AGS and control cells ($P<0.05$ Kruskal-Wallis non-parametric test), whereas no differences were found for all other variables.

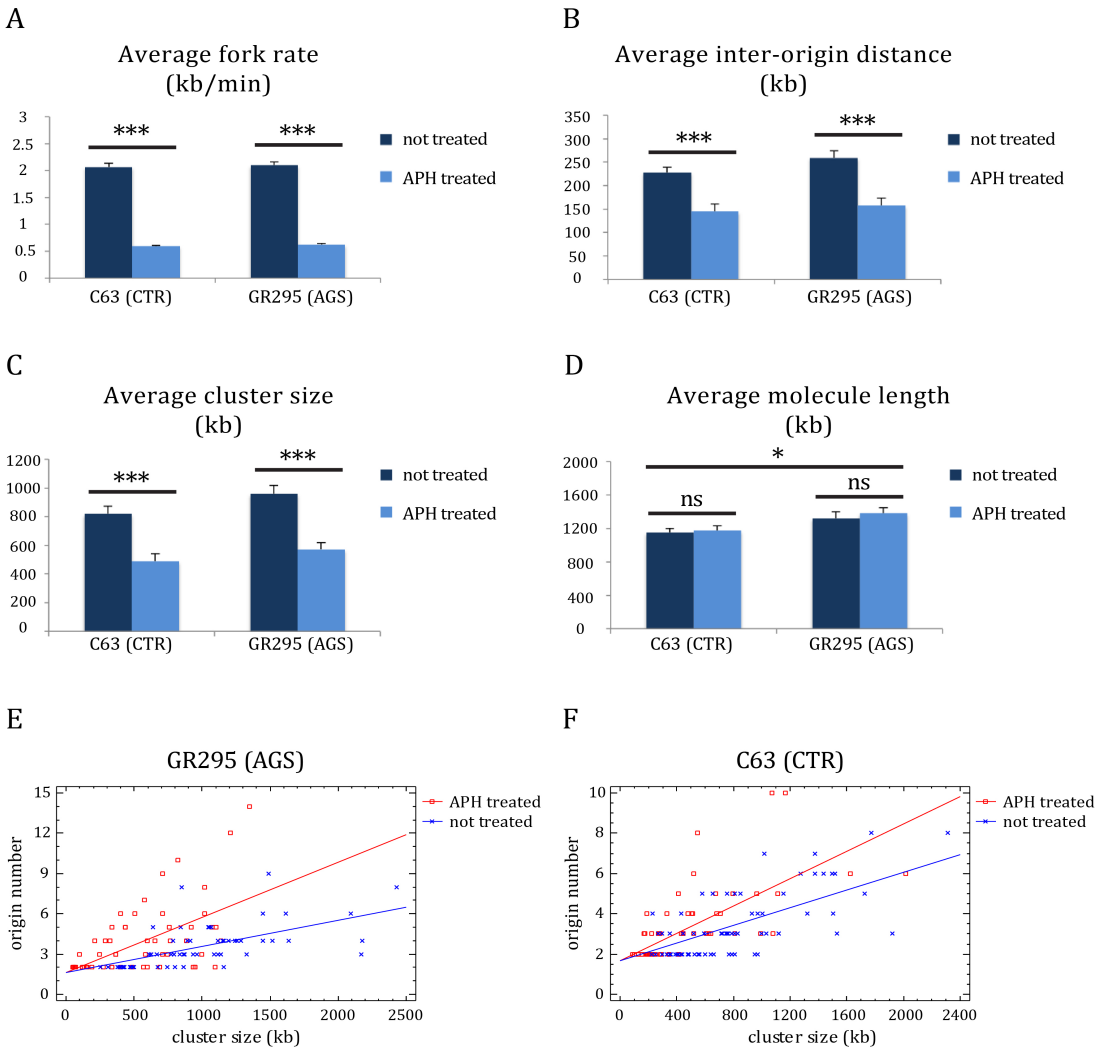


Fig. 11. Average values of fork rates (A), inter-origin distances (B), cluster size (C) and molecule lengths (D) in untreated and aphidicolin (APH)-treated GR295 and C63 cell lines. Error bars indicate standard error of the mean. In each histogram, the statistical significance is also displayed. * P<0.05; ** P<0.01; *** P<0.001; ns=not significant. The number of activated replication origins per cluster was determined by linear regression analysis under both normal and perturbed conditions in GR295 AGS cells (E) and in C63 control cells (F). By comparison of regression lines a significant increase in origin firing was detected under replication stress in both cell lines (P<0.001 for the difference between regression slopes).

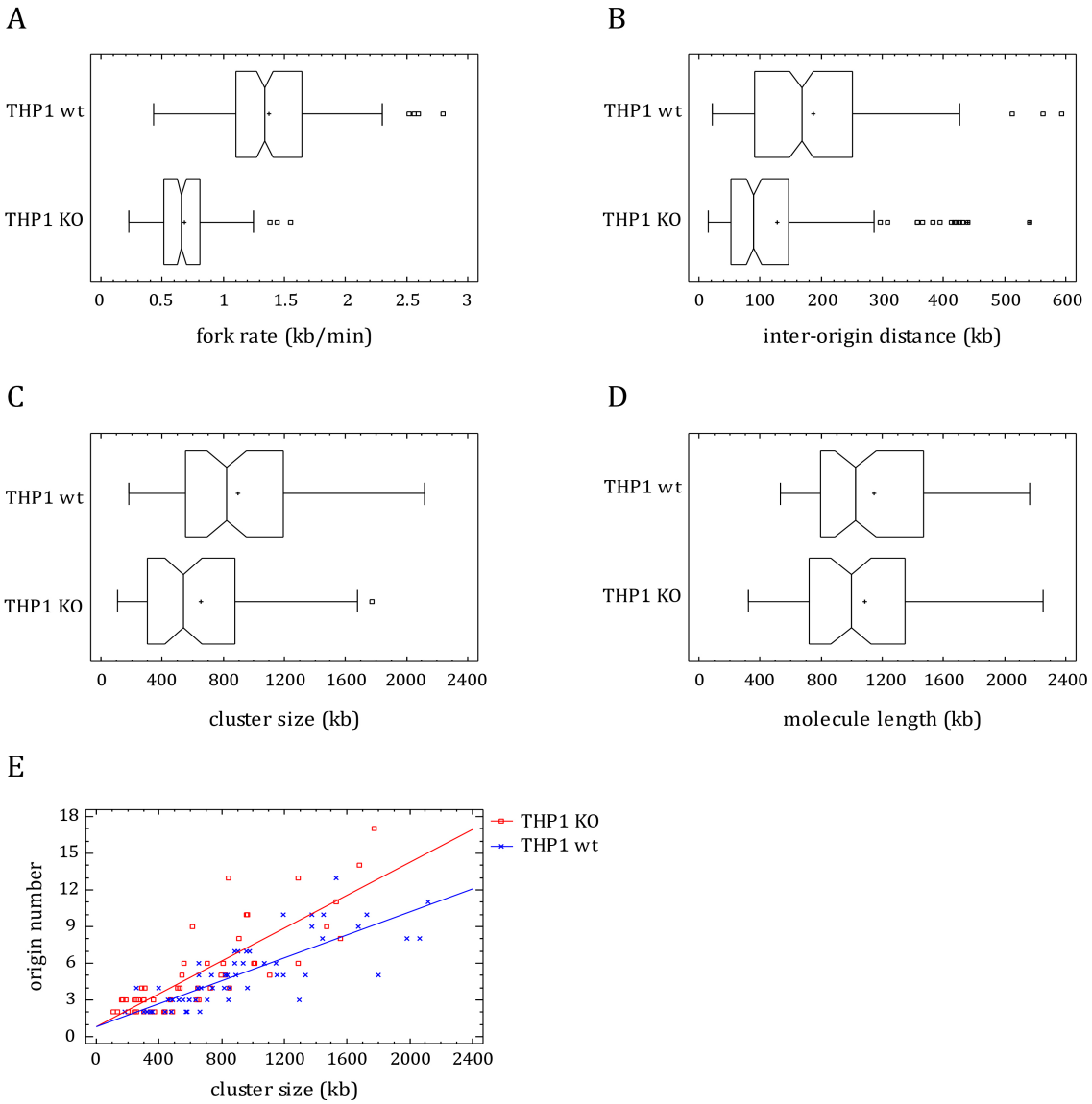


Fig. 12. Distributions of fork rates (A), inter-origin distances (B), cluster size (C) and molecule lengths (D) in THP1 KO and wildtype cells in physiological conditions. Molecule lengths were highly comparable among cell samples, while a significant reduction of all other replication parameters was observed in THP1 KO cells with respect to wildtype ($P<0.001$ Kruskal-Wallis non-parametric test). The number of activated origins in each replication cluster was determined by linear regression analysis (E) and by comparing regression lines a significant increase in origin firing was observed in THP1 KO than wildtype cells ($P<0.01$ for the difference between regression slopes).

3. Replication dynamics in terminally differentiated cells forced to re-enter the cell cycle

As already demonstrated (Palumbo et al. 2013; Ryba et al. 2011), replication profiles are cell type-specific and modulated during development and cell differentiation. To assess how replication dynamics is regulated in response to forced cell cycle re-entry, terminally differentiated mouse myotubes reactivated by depletion of p21 and p27 CKIs (rMt) were used as experimental model in comparison to proliferating myoblasts. As already described by the laboratory of Dr. Crescenzi (Pajalunga et al. 2010), rMt fail to complete DNA replication and accumulate DNA damages. Further evidences supported the hypothesis that the observed replication impairment may be ascribed to the inability of rMt to properly expand their dNTP pools. Indeed, after addition of nucleotide precursors, replication seemed to be slightly enhanced by DNA content detection and damages were partially rescued as observed by immunofluorescence experiments (Pajalunga et al. 2017). Thus, in the frame of a collaboration with Dr. Crescenzi, rMt with or without deoxynucleosides were analysed by molecular combing in comparison to proliferating myoblasts. A manuscript was submitted to Cell Death and Differentiation on July 2016 and the revised paper was accepted on January 2017 (Pajalunga et al. 2017).

3.1 Replication dynamics in terminally differentiated mouse myotubes upon forced cell cycle re-entry

Molecular combing experiments were carried out in rMt in the presence/absence of nucleotide precursors with respect to proliferating muscle cells. Data are summarised in Table 6. By DNA counterstaining, fiber integrity was monitored in order to verify the accuracy of measurements of examined replication parameters in each cell sample. Molecule lengths were comparable between rMt supplemented with deoxynucleosides and proliferating myoblasts (Table 6), as indicated by the Kruskal-Wallis non-parametric test and their mean and median values were above the critical threshold for the correct estimation of fork rates, IODs and cluster size, in agreement with published studies (Técher et al. 2013; Bianco et al. 2012). Instead, when unsupplemented rMt were considered, a significant reduction of fiber lengths was detected ($P<0.01$ Kruskal-Wallis non-parametric test; Table 6 and Fig. 13B), suggesting an intrinsic DNA fragility coherent with the high prevalence of DNA breaks observed in these cells (Pajalunga et al. 2010). Furthermore, the short average value of rMt molecules prevented the measurements of IODs and cluster size.

A remarkable reduction of fork rates was detected in rMt when compared to proliferating myoblasts ($P<0.001$ Kruskal-Wallis non-parametric test; Fig. 13A and Table 6), resembling the effects observed under replication stress (Palumbo et al.

2013; Anglana et al. 2003; Courbet et al. 2008). The addition of deoxynucleosides only partially rescued fork speed, as it remained 2-fold lower than in proliferating cells ($P<0.001$ Kruskal-Wallis non-parametric test; Fig. 13A and Table 6). These results were consistent with the hypothesis that dNTP depletion represents the primary cause of the replication failure observed in rMt. Similarly, both inter-origin distances and cluster size were markedly reduced in deoxynucleoside-treated rMt with respect to proliferating myoblasts ($P<0.001$ Kruskal-Wallis non-parametric test; Fig. 13C-D and Table 6). However, the number of activated origins per replication cluster was similar between the two cell samples, as indicated by the linear regression analysis, which was not significant (Fig. 13E). These data not only confirmed the notion that origins fire more closely under replication stress (Courbet et al. 2008; Anglana et al. 2003), but also suggested that some genomic regions in myotubes are left under-replicated.

A slight increase in the proportion of unidirectional forks was observed in rMt dosed with deoxynucleosides with respect to myoblasts ($P<0.05$ chi-square analysis; Fig. 13F and Table 6), although their values fell within ranges already determined in other mammalian cell lines (Palumbo et al. 2013). Conversely, no differences were detected when fork pause/arrest and asynchronous patterns were compared (Table 6). On the whole, results suggested that upon forced cell cycle reactivation fork rates are mainly affected as a consequence of dNTP depletion.

Table 6. Whole genome replication profiles of proliferating and TD mouse muscle cells by molecular combing upon forced cell cycle reactivation.

		Proliferating myoblasts	Reactivated myotubes	Reactivated myotubes supplemented with deoxynucleosides
Replication forks	Total	174	-	270
	Unidirectional forks (N)	47	-	98
	Unidirectional forks (%)	27.0	-	36.3
	Paused/arrested bidirectional forks (N)[ç]	22	-	28
	Paused/arrested bidirectional forks (%)	12.6	-	10.4
	Asynchronous bidirectional forks (N)	8	-	18
	Asynchronous bidirectional forks (%)	4.6	-	6.7
Fork rate (kb/min)[§]	Mean ± SE	2.35 ± 0.087	0.38 ± 0.020	1.06 ± 0.057
	Median	2.33	0.35	0.91
	N	96	66	125
	Min	0.58	0.09	0.19
	Max	5.72	0.77	3.55
Inter-origin distance (kb)	Mean ± SE	186.7 ± 10.41	-	139.1 ± 9.08
	Median	154.8	-	113.4
	N	154	-	115
	Min	24.1	-	13.5
	Max	809.2	-	505.6
Cluster size (kb)	Mean ± SE	681.2 ± 34.59	-	481.4 ± 33.89
	Median	631.7	-	414.5
	N	93	-	89
	Min	93.8	-	66.5
	Max	1850.6	-	1610.3
Molecule length (kb)	Mean ± SE	1165.1 ± 51.85	526.1 ± 20.09	1069.1 ± 60.28
	Median	1052.0	488.1	996.5
	N	93	65	89
	Min	323.4	249.4	166.2
	Max	2567.8	967.5	2869.1

		Proliferating myoblasts	Reactivated myotubes	Reactivated myotubes supplemented with deoxynucleosides
Number of activated origins per cluster	Mean	3	-	3
	N	304	-	281
	Min	2	-	2
	Max	9	-	10

§ Only bidirectional forks, excluding asynchronous and paused/arrested forks.

ç Replication forks with bilateral pause/arrest are considered as a single event because associated with the firing of one origin.

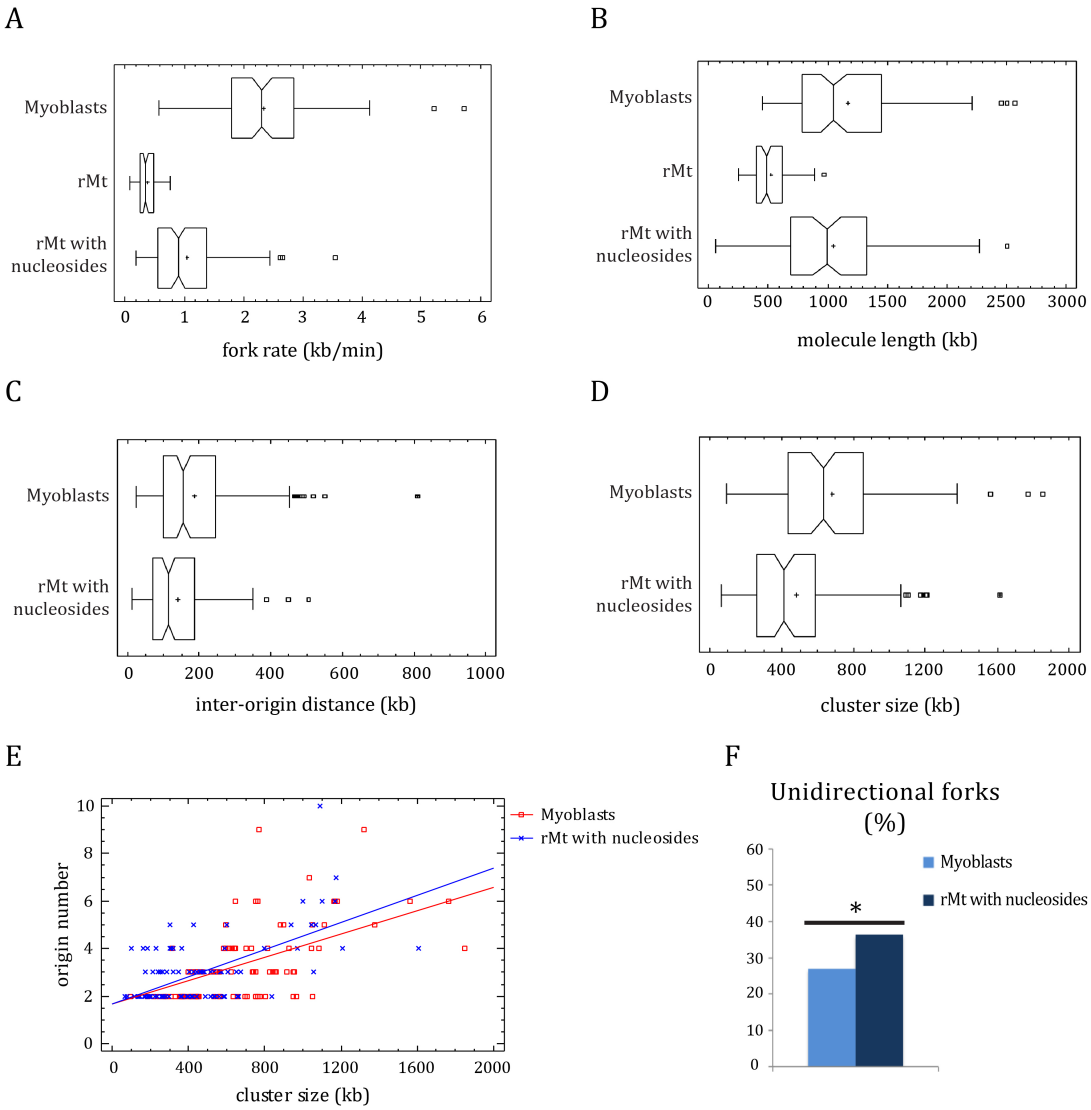


Fig. 13. Distributions of fork rates, inter-origin distances (IOD), cluster size and molecule lengths in reactivated myotubes and proliferating myoblasts. (A) Fork rate distributions were significantly different between rMt in the presence/absence of deoxynucleotide treatment and proliferating myoblasts ($P<0.001$ Kruskal-Wallis non-parametric test). (B) Distributions of molecule lengths were highly comparable between rMt dosed with deoxynucleosides and proliferating cells, whereas a significant reduction was detected in the fiber length of rMt ($P<0.001$ Kruskal-Wallis non-parametric test). (C) and (D) Distributions of IOD and cluster size in rMt dosed with nucleosides and proliferating myoblasts were significantly different ($P<0.001$ Kruskal-Wallis non-parametric test). (E) By the comparison of regression slopes the same number of activated origins per replication cluster was detected in both cell samples (not significant regression slope comparison). (F) A significant increase in the proportion of unidirectional forks was detected by molecular combing in rMt dosed with deoxynucleosides in comparison to proliferating myoblasts ($P<0.05$ chi-square analysis). The statistical significance is indicated. More details in (Pajalunga et al. 2017).

DISCUSSION

Replication timing and replication dynamics are cell type-specific. They are regulated according to changes in chromatin organisation and gene expression occurring during development and cell differentiation (Hiratani et al. 2010; Ryba et al. 2011; Pope & Gilbert 2013; Palumbo et al. 2013). Replication profiles are defined by different parameters, including fork rates, inter-origin distances and cluster size as well as alternative replication patterns, such as unidirectional and paused/arrested forks (Tuduri et al. 2010; Palumbo et al. 2013; Técher et al. 2013; Prioleau & MacAlpine 2016).

In this study, genome-wide molecular combing analysis revealed that replication fork rates were largely comparable among all the mammalian cell lines analysed (S1 Table in Appendix A; Tables 5 and 6), independently of species and in line with published data (Palumbo et al. 2013; Técher et al. 2013). The estimated mean fork speed ranged from about 1 to 2.3 kb/min in unperturbed conditions (S1 Table in Appendix A; Tables 5 and 6) and these values are consistent with measurements reported in other different mammalian cells (Herrick & Bensimon 2008; Palumbo et al. 2013). Despite the high level of similarity among observed fork rates, slight differences can be appreciated according to cell types (S1 Table in Appendix A; Tables 5 and 6). In particular, fork speeds in the order of 2-2.3 kb/min were detected in human primary skin fibroblasts and in mouse myoblasts (Tables 5 and 6), in agreement with average values previously estimated in IMR90 (Palumbo et al. 2013) and in transformed MRC5 human cells (Lemée et al. 2010; Guilbaud et al. 2011). In lymphoblastoid cell lines, replication forks run slower than in fibroblasts with mean fork speed ranging 1.7-1.9 kb/min (S1 Table in Appendix A; Tables 5 and 6), coherently with data obtained from other mammalian lymphocytes, lymphoblastoid cell lines (Palumbo et al. 2013) and JEFF cells (Letessier et al. 2011; Naim et al. 2013). Instead, the average fork rate of wildtype THP1 monocytes was 1.37 kb/min (Table 5), quite far from the mean fork speed estimated in the other cell types analysed, although it remained within the range previously determined in different mammalian fibroblasts and epithelial cells (Herrick & Bensimon 2008; Técher et al. 2013; Kaykov et al. 2016). The THP1 cell line derives from patients affected by acute leukemia, and noteworthy slow fork rates were detected also in other transformed human cell lines, such as in the chronic myelogenous leukemia K562 lymphoblast (Rimmelé et al. 2010) and in the adenocarcinoma epithelial HeLa cells (Malinsky et al. 2001; Guilbaud et al. 2011). Unfortunately, the molecular mechanism underlying the slow down of replication forks upon cell transformation is unclear as a direct comparison between primary cells and their transformed counterpart has never been provided. Remarkably, the comparisons between replication parameters evaluated in the whole genome and those detected at single loci by molecular combing are discouraged as it was demonstrated that slow

replication forks and particular replication patterns discriminate the locus-specific program from genome-wide replication profiles (Palumbo et al. 2013). Accordingly, fork rates estimated in the 850 kb region harbouring *FXN* were lower than the corresponding genome-wide values in the lymphoblastoid cell lines analysed (Appendix A, Table 2 and S1 Table).

The regulation of the replication process is reached by the coordinated modulation of fork rates and origin choice both in unperturbed and stress conditions (Méchali 2010; Yekezare et al. 2013; Zeman & Cimprich 2014). In particular, it is stated that a reduction of inter-origin distances corresponding to an enhanced origin firing is usually associated with a global slow-down of replication forks (Anglana et al. 2003; Courbet et al. 2008). This positive correlation was confirmed in all the cell lines under study both in physiological conditions and under replication stress (Tables 5 and 6). Furthermore, fork rates and IODs were positively modulated not only genome-wide, but also when dealing with individual genomic regions as demonstrated by the analysis at the *FXN* locus and in other four loci (Palumbo et al. 2013). Similarly to fork rates, average inter-origin distances were comparable among all the cell lines considered (S1 Table in Appendix A; Tables 5 and 6) and their values fell into the range already observed in other mammalian cells (Herrick & Bensimon 2008; Palumbo et al. 2013).

Replication profiles are also defined by variations in alternative replication patterns, including forks with unidirectional progression or with arms moving with different rates and paused/arrested sites. Indeed, the increased proportions of both unidirectional and stalled forks were demonstrated to be constitutive events characterising the replication program of human IMR90 fibroblasts (Palumbo et al. 2013). The proportions of alternative replication patterns determined in this study (S1 Table in Appendix A; Tables 5 and 6) fell within ranges previously identified in other lymphoblastoid cell lines and in fibroblasts (Palumbo et al. 2013). Interestingly, the proportion of unidirectional forks increased in response to replication stress in different cell lines evaluated. First at all, the percentages of forks with unidirectional progression were higher in aphidicolin-treated than in untreated C63 and GR295 fibroblasts (Table 5), consistent with published results in primary human lymphocytes (Palumbo et al. 2010). Furthermore, I detected an increased proportion of unidirectional forks in terminally differentiated mouse myotubes upon their forced cell cycle reactivation in comparison to proliferating myoblasts (Table 6), suggesting that these events are also implicated in the modulation of the replication program during cell differentiation. Similarly, it was previously demonstrated that in human primary fibroblasts the proportion of unidirectional forks progressively increased moving toward senescence (Palumbo et al. 2013). Moreover, changes in their length were also identified in the *FXN* locus in the presence of a replication fork impediment (in the Results section p. 43 and in Appendix A).

Therefore, one general conclusion of my study is that replication profiles are cell type-specific and this specificity may be ascribed not only to changes in fork rates

and origin choice, but also to variations in alternative replication patterns, in agreement with previous data (Palumbo et al. 2013). Concerning how mammalian replication profiles are regulated in response to endogenous and exogenous stress conditions, the following sections discuss in depth my results.

1. DNA replication in the presence of an expanded trinucleotide repeat: a single-molecule view of the *Frataxin* gene

1.1 The origin-switch model and the GAA-repeat expansion at the *FXN* locus

To determine the effects of trinucleotide repeat expansions on DNA replication in the endogenous context, human cell lines carrying an homozygous GAA-repeat expansion within the *FXN* gene were used as experimental model, and the replication program of the locus was examined. First at all, I evaluated the replication timing of the normal *FXN* sequence confirming that it is enclosed in a mid-late replication domain (Fig. 1 in Appendix A). Conversely, in the presence of the GAA-repeat expansion the timing of the *FXN* locus was altered. Indeed, the replication of the *FXN* gene was slowed or delayed during the first half of the S phase in mutant alleles when compared to the wildtype sequence, even though this effect was further normalised in the latter part of the S-phase, when both expanded and normal alleles are undergoing and completing their replication. Subsequently, by molecular combing the replication profile of a large genomic region harbouring *FXN* was evaluated. I found that in normal cells, origins never fire within the gene, which is thus passively replicated by incoming replication forks. These results are in full agreement with an exploratory analysis on human primary lymphocytes (Palumbo et al. 2013). In contrast, origin choice widely changed both upstream and downstream the expanded GAA-repeat and the predominant effect was the activation of additional dormant origins downstream the GAA-repeat expansion, which may be considered as a rescue mechanism to assure replication of mutant alleles (discussed more in depth in the following paragraph p. 64). Interestingly, firing of additional origins within the expanded *FXN* gene was confirmed in induced pluripotent stem cells (iPSC) derived from FRDA patients (Gerhardt et al. 2016). The main implication of dormant origin activation is an increase in the number of forks firing downstream the GAA-repeat in FRDA cells (S7 and S8 Figs. in Appendix A) with respect to normal ones (S5 and S6 Figs. in Appendix A). In turn, while in normal cells the GAA-repeat is prevalently in the lagging strand template, in mutant cells the expansion is frequently located in the leading strand template. Again, this observation is consistent with results obtained in FRDA iPSC by using a different single molecule technique (Gerhardt et al. 2016).
Hence, the origin-switch model for trinucleotide repeat instability (Pearson et al. 2005; Cleary & Pearson 2005; Mirkin & Smirnova 2002) conforms to the case of the

GAA-repeat expansion in the *FXN* locus, similarly to what found for the CGG-expansion at *FMR1* (Gerhardt et al. 2014).

1.2 Alternative replication patterns play a role in the replication of expanded *FXN* alleles

Cell type-specific replication profiles are determined not only by the variability of fork rates and origin choice but also by variations in the proportions of alternative replication patterns, including unidirectional forks and pause/arrest sites (Palumbo et al. 2013). For instance, it was demonstrated that in human primary fibroblasts wide regions of the genome are replicated by forks with unidirectional progression and this seems to be a peculiar feature of the replication program of this cell type (Palumbo et al. 2013). In FRDA cells, a marked reduction of the average length of unidirectional forks and a broader distribution of their values were observed by molecular combing with respect to the control cell lines (Fig. 5 in Appendix A). Furthermore, their frequencies in all cell samples fell into the range known for mammalian cell lines (Palumbo et al. 2013). Therefore, although the biological meaning of unidirectional forks is still not completely understood, their length reduction represents a major effect associated with the GAA-repeat at the expanded *FXN* gene.

Different model systems from yeast to transfected or engineered human cells were used to monitor the progression of DNA replication in the presence of a GAA-repeat expansion. However, their main constraint is the limited number of repetitive tracts that can be examined with respect to those detected in cells derived from FRDA patients. For this purpose, I analysed the replication profiles of human cell lines carrying from 630 to 1030 GAA-repeats. In FRDA cells, the functional loss of Frataxin does not affect cell proliferation and is not associated with global changes in the replication program (S1 Fig. in Appendix A). Furthermore, long stretches of GAA-repeats in these cell lines do not represent a strong impediment for the replication process, in agreement with published works (Krasilnikova & Mirkin 2004; Kim et al. 2008; Chandok et al. 2012; Follonier et al. 2013). Indeed, in the expanded *FXN* alleles the shift of the replication timing is rapidly solved in the second half of the S-phase (Table 1 in Appendix A) and the replication of the mutated *FXN* gene is completed via activation of dormant origins. This can be viewed as a rescue mechanism as it is well known that in mammalian cells dormant origins are usually recruited to solve fork stalling associated with replication impediments (Méchali 2010; Yekezare et al. 2013; Zeman & Cimprich 2014). Previous studies applied different strategies to investigate the replication of GAA-repeat *in vivo*. By cloning stretches of different length in yeast plasmid (Krasilnikova & Mirkin 2004), fork stalling and the consequent inhibition of fork progression in the order of 1.5 times were detected in long repetitive tracts ranging the size of premutated or mutated alleles, but not in the presence of short (<40) repeats. In a SV40-based plasmid transfected in human cells, several abnormal

replication intermediates were observed when expansions in the range of 33-90 GAA-repeats are replicated (Follonier et al. 2013). In particular, replication forks transiently pause at GAA-repeat longer than 66 units, while in longest repetitive tracts (>90) fork reversal was found to be associated with fork pausing (Follonier et al. 2013). Moreover, both in yeast and transfected human cells the most significant increase of fork pausing has been demonstrated to preferentially occur when the GAA-repeat is located in the lagging strand template (Shishkin et al. 2009; Follonier et al. 2013). Recently, in human induced pluripotent stem cells derived from FRDA patients an high occurrence of fork stalling was reported at the expanded GAA-repeat in the leading strand template and the impairment of fork progression was suggested to be caused by a collision between transcription and replication at the expanded GAA-repeat (Gerhardt et al. 2016). In addition, Gerhard and colleagues speculated that this conflict may have a role in promoting further expansions (Gerhardt et al. 2016).

By molecular combing, I found several dispersed and widespread pause/arrest events along the 850 kb region harbouring *FXN* (Fig. 6 in Appendix A). Remarkably, recurrent events of pause/arrest of the fork were recorded in the proximity of the short GAA-repeat in the GM15851 control cell line. This result is in line with that obtained at the *FMR1* locus, where replication forks stalled also in normal cells (Gerhardt et al. 2014). Instead, in the second control cell line and in FRDA cells a less intense occurrence of pause/arrest events was observed. To explain these differences the global response to the replication impairment associated with the presence of the GAA-repeat expansion must be considered. In particular, activation of dormant origins and changes in unidirectional fork progression play the major role and both events are correlated with long-lasting replication patterns. In contrast, in most of the cases, including the GAA-repeat (Follonier et al. 2013), fork pausing are transient events. Thus, while firing of dormant origins and reduction of unidirectional fork length are the most evident and easily detectable effects associated with the expanded GAA-repeat, the possibility to determine paused replication forks is affected by their transient nature and it is defined as a function of the labelling scheme adopted by different techniques. In particular, the chance to identify these events is enhanced by applying SMARD because it relies on a longer labelling period covering the entire S-phase (Norio & Schildkraut 2004; Gerhardt et al. 2016) with respect to that used in molecular combing experiments (Stevanoni et al. 2016). Hence, it is not surprising that differently from my results, using SMARD the expanded GAA-repeat was indicated as a pausing site both in undifferentiated and differentiated FRDA cell lines (Gerhardt et al. 2016).

1.3 Alternative experimental strategies fail to detect the occurrence of rare initiation events

It is largely accepted that trinucleotide repeat expansion are prone to form DNA secondary structures, which are among the most common sources of replication

stress (Krasilnikova & Mirkin 2004; Mirkin & Mirkin 2007). To overcome replication impediments causing fork slowing and stalling, mammalian cells recruit and activate adjacent dormant origins, which are rare and stochastic events (Blow et al. 2011; Alver et al. 2014; Ge et al. 2015). Differently from the induced massive response observed when cells are treated with DNA replication inhibitors (Anglana et al. 2003; Woodward et al. 2006; Courbet et al. 2008), my results suggested the occurrence of a physiological event restricted to a narrow genomic region at *FXN*. The frequency of dormant origin firing within the gene, estimated on the total number of replicating molecules, ranged 14-21% in the FRDA cell lines (S7 and S8 Figs. in Appendix A). Moreover, the activation of a dormant origin within *FXN* was never associated with the firing of additional dormant origins and its location was not restricted to a steady position within the gene. Thus, this event is a peculiar feature of the expanded *FXN* allele, but it occurs stochastically within the gene and among cells. To identify such rare, stochastic and widespread initiation events occurring as a change of the locus-specific replication program, single-molecule techniques appear the most appropriate tool because of their intrinsic ability to describe the high cell-to-cell variability peculiar of mammalian genomes. Indeed, activation of dormant origins at *FXN* in FRDA cells was demonstrated independently and in parallel by applying molecular combing or SMARD (Stevanoni et al. 2016; Gerhardt et al. 2016). Conversely, it is reasonable that genome-wide techniques are weakly effective in identifying rare and inefficient initiation events. Notwithstanding, following a referee's request, to validate the replication profiles of FRDA and normal cell lines determined by molecular combing I considered alternative approaches. The SNS abundance assay provided a weak trend supporting the activation of dormant origins within the expanded *FXN* gene (Fig. 7D in Appendix A). Most importantly, the availability of OK-seq data from a normal lymphoblastoid cell line and from HeLa cells (Petryk et al. 2016) gave me the opportunity to confirm the passive modality of replication of the normal *FXN* gene in agreement with my molecular combing results. *FXN* is associated with a termination region and replicated by forks emanating from the two adjacent initiation zones (S10 Fig. in Appendix A). Thus, OK-seq profiles not only strengthened results obtained by molecular combing but also demonstrated that my observations can be extended to other cell lines and to diverse cell types. Unfortunately, the identification of dormant origins in FRDA cells would be really challenging by OK-seq as their firing is not strong enough to generate a detectable initiation zone.

1.4 Future perspectives

By molecular combing, the replication of the *FXN* gene was monitored in its endogenous genomic context and a fine picture of the events associated with the trinucleotide repeat has been provided.

Understanding if the origin-switch effect represents the cause or instead is a consequence of the GAA-repeat expansion is an important issue to be further investigated. In this frame, tracking all the alterations of the replication program occurring during development will be necessary.
Furthermore, despite the weak evidence available from my molecular combing results, the detection of recurrent fork arrest events in the proximity of the short GAA-repeat in one of the control cell lines must be taken into further consideration as a possible impact of the short non-pathological GAA-repeat on fork progression, similarly to the case of the CGG-repeat expansion at *FMR1* (Gerhardt et al. 2014).
Moreover, information concerning the transcriptional profile of the *FXN* gene is scanty, preventing to prove the hypothesis of collision between the replication and transcription machineries at the expanded GAA-repeat (Gerhardt et al. 2016). In this perspective, further work is necessary to determine when *FXN* is transcribed during the cell cycle and in turn whether it is likely that the supposed replication-transcription conflict can occur in the presence of the GAA-repeat expansion.
Recently, several evidence both in prokaryotic and eukaryotic systems proved that replication-transcription collisions are a natural source of genomic instability in many neurodegenerative disorders and in different types of cancer (Helmrich et al. 2013; Hamperl & Cimprich 2016; García-Muse & Aguilera 2016). Understanding whether FRDA patients display chromosomal rearrangements and if these events are correlated to a replication-transcription conflict are intriguing issues to be further elucidated.

2. Imbalance of dNTP pools affects the replication program and may cause replication stress

2.1 Effects of dNTP pool imbalances on DNA replication in different model systems

Imbalances of dNTP pools have several implications in the progression of DNA replication leading to mutagenesis in yeast (Chabes & Stillman 2007; Kumar et al. 2010; Buckland et al. 2014; Watt et al. 2015; Williams et al. 2015) and tumorigenesis in human cells (Bester et al. 2011; Chabosseau et al. 2011; Gemble et al. 2015). The effects of the limiting availability of dNTPs on replication dynamics are known: they imply a global slow-down of replication fork speed and the consequent recruitment of dormant origins by activation of the DNA damage response (Anglana et al. 2003; Courbet et al. 2008; Kumar et al. 2010; Bester et al. 2011). In contrast, the consequences of increased and unbalanced dNTP pools are only partially depicted in yeast and mainly affect replication fidelity (Buckland et al. 2014; Watt et al. 2015). Only a single work in yeast reported that the high supply of dNTPs leads to an increase of fork rates and promotes cell survival under

replication stress conditions (Poli et al. 2012). Furthermore, the above study proposed that cells improve their tolerance to replication stress by expanding their dNTP pools (Poli et al. 2012). In parallel, this hypothesis was confirmed in another published work carried out on different yeast mutants (Davidson et al. 2012). The evaluation of the replication profiles in different human cell lines depleted of SAMHD1 and thus showing increased and unbalanced dNTP pools (Table 4) returned unexpected results in comparison to the yeast model. Indeed, no differences emerged between AGS and control fibroblasts both in physiological conditions (Fig. 9) and under replication stress (Fig. 10), suggesting that independent of the high availability of nucleotide precursors DNA replication proceeds mostly undisturbed. In contrast, in SAMHD1-depleted THP1 cells an unexpected reduction of fork rates and an increased activation of replication origins were observed when compared to the wildtype THP1 cell line (Fig. 12). The contradictory cellular responses to the increased availability of dNTPs in yeast and in mammalian cells may be explained not only by intrinsic differences among the two model systems, but also by the diverse mechanisms underlying the dNTP pool expansion. Indeed, in yeast mutants dNTP pool increase was due to overexpression/upregulation of the ribonucleotide reductase (Poli et al. 2012; Davidson et al. 2012), whereas different mutations affecting the *SAMHD1* gene and its complete depletion are responsible of the dNTP pool imbalances in AGS and THP1 KO cells, respectively (Table 4).

2.2 Hypotheses to explain the differential response of AGS and THP1 cell lines

Beyond the differences observed when yeast mutants were compared to mammalian cells, THP1 and AGS cell lines displayed diverse replication profiles in response to increased and unbalanced dNTP pools. Indeed, replication dynamics in AGS fibroblasts were highly comparable to control cells (Fig. 9), whereas a pronounced effect resembling the typical response to replication stress conditions (Anglana et al. 2003; Courbet et al. 2008; Yekezare et al. 2013; Zeman & Cimprich 2014) was detected in THP1 KO versus wildtype cells (Fig. 12). To explain this discrepancy, I propose the following hypotheses.

First, it is known that SAMHD1 is ubiquitously expressed in different cell types and tissues (Li et al. 2000; Descours et al. 2012; Clifford et al. 2014). Its role as triphosphohydrolase is well established in quiescent cells, where the supply of deoxynucleotides is maintained at the basal level (Rampazzo et al. 2010; Franzolin et al. 2013), but this activity was largely demonstrated to be essential also in proliferating cells, in particular to counteract viral infection by limiting dNTPs (Laguette et al. 2011; Franzolin et al. 2013). However, according to its ability to bind single stranded nucleic acids (Tüngler et al. 2013) a secondary role of SAMHD1 as nuclease has been supposed in proliferating cells (Beloglazova et al. 2013; Ryoo et al. 2014), although this aspect is still widely debated. In this frame,

SAMHD1 may be implicated in the digestion of replication intermediates formed at stalling sites, where the accumulation of single stranded fragments leads to the activation of the DNA damage response (Sogo et al. 2002; Edenberg et al. 2014). This hypothesis may be consistent with the double amount of paused/arrested forks detected in THP1 KO cells and associated with the activation of the cellular response to a replication stress, in comparison to THP1 wildtype cells (Table 5). Furthermore, in a recent work a ribonuclease role of SAMHD1 was demonstrated in THP1 cells, but also in other monocyte-derived cell lines (Ryoo et al. 2014). On the contrary, this nuclease activity was never reported in fibroblasts, where the only function determined for SAMHD1 was dNTPase (Rampazzo et al. 2010; Franzolin et al. 2013). The proportions of stalled forks were comparable in AGS and control fibroblast cell lines (Table 5). Therefore, the differential responses observed in AGS and THP1 cells may support the hypothesis of a cell type-specific activity of the SAMHD1 protein.
Alternatively, the similar replication behaviour of AGS and control fibroblasts may be associated with the development of an adaptive phenotype. Indeed, the endogenous higher supply of dNTPs caused by SAMHD1 mutations is a constitutive condition for AGS fibroblasts. Thus, it is possible that AGS cells compensate for the lack of the SAMHD1 protein and the consequent increase and imbalance of their dNTP pools. On the other hand, the effects on replication dynamics observed in THP1 KO cells may also be associated with cell transformation and therefore they may be independent of SAMHD1 depletion. To confirm these hypotheses, further work is required.

2.3 Future perspectives

The cellular responses to the increased and unbalanced dNTP pools were different between AGS and THP1 KO cells when compared to their corresponding control cell lines. Understanding whether the effects seen in THP1 KO cells are caused by SAMHD1 depletion or are instead associated with cell transformation is crucial. In this frame, possible experimental strategies could consider the depletion of SAMHD1 in normal fibroblasts, for instance by using the CRISPR/Cas9 technology (Sander & Joung 2014) or by infection with the viral protein Vpx, which leads to the endogenous ubiquitin-dependent degradation of SAMHD1 (Hrecka et al. 2011; Schaller et al. 2012). This latter procedure could be also performed on THP1 wildtype cells, so that it would be possible to clarify whether there is a cell type-specific response.
Molecular combing analyses in GR295 AGS and C63 control cells indicated that the treatment with aphidicolin has same consequences on replication dynamics (Fig. 10 and Table 5). However, it will be intriguing to evaluate how these cells will respond to different sources of replication stress. In particular, the treatment with hydroxyurea (HU), a ribonucleotide reductase inhibitor (Skoog & Nordenskj 1971), would contribute to elucidate whether AGS fibroblasts have adapted to the

increased availability of nucleotide precursors, as they may be differently sensitive to the depletion of dNTP pools with respect to normal cells. Furthermore, consistent with the yeast model (Poli et al. 2012; Davidson et al. 2012), the dNTP pool expansion in AGS fibroblasts might confer resistance to replication fork slowing/stalling induced by HU.
Moreover, understanding whether an increased supply of dNTPs may have an effect on the efficiency of replication fork restart will be an interesting issue to further consider, also in the light of the potential nuclease activity supposed for SAMHD1 (Beloglazova et al. 2013; Ryoo et al. 2014). In this perspective, the most reasonable expectation is that AGS cells will have higher rate of fork restart than normal fibroblasts as a consequence of their high availability of nucleotide precursors. Even so, the fidelity of DNA synthesis may also be taken into further considerations as it was largely demonstrated in yeast that increased dNTP pools promote genomic instability and mutagenesis by affecting replication fidelity (Buckland et al. 2014; Watt et al. 2015; Williams et al. 2015).

3. Modulation of replication profiles in terminally differentiated cells after forced cell cycle re-entry

3.1 Replication profiles upon cell cycle reactivation

By molecular combing I monitored replication dynamics after forced cell cycle re-entry in mouse myotubes. Fork rates were significantly reduced in reactivated myotubes (rMt) with respect to proliferating myoblasts (Fig. 13A and Table 6), resembling the effects seen under replication stress, such as after treatment with aphidicolin (Palumbo et al. 2013) or by nucleotide depletion mediated by hydroxyurea (Anglana et al. 2003; Courbet et al. 2008). This result is in agreement with previous data demonstrating that reactivated myotubes fail to complete replication of their genome (Pajalunga et al. 2010) and attributing this failure to the inability of these terminally differentiated cells to properly expand their dNTP pools (Pajalunga et al. 2017). Indeed, fork rates were partially rescued after addition of nucleotide precursors into rMt, but average fork speed remained 2-fold lower than that of proliferating myoblasts (Fig. 13A and Table 6). Although the depletion of dNTPs is strongly implicated in the replication impairment observed in rMt, it may be not the unique aspect involved, as fork speed was not completely restored after deoxynucleosides addition.
The effects on fork rates seen in rMt supplemented with deoxynucleosides highly correlated with the reduction of inter-origin distances and cluster size when compared to those detected in myoblasts (Fig. 13C-D and Table 6), as already described in the first section of the Discussion (p. 62). However, the number of activated origins in each cluster did not change significantly between the two cell

samples (Fig. 13E and Table 6). This result indicated that some genomic regions remained under-replicated in reactivated myotubes suggesting that these cells are not able to properly activate the DNA damage response. Therefore, the replication impairment observed in myotubes may be ascribed not only to the dNTP pool depletion affecting fork rates, but also to the inability of these terminally differentiated cells to recruit additional replication origins compensating for the reduction of fork speed. A possible explanation of this effect may concern variations in the chromatin context associated with the cellular differentiation stage. Indeed, during development and cell differentiation replication domains change in relation to nuclear positioning, chromatin condensation and gene expression (Hiratani et al. 2008; Hiratani et al. 2010; Therizols et al. 2014). For example, during development replication timing undergoes either early-to-late or late-to-early switches according to changes in the transcriptional activity of genes enclosed within the same replication domain (Hiratani et al. 2008; Hiratani et al. 2010). In addition, mouse partially reprogrammed induced pluripotent stem cells (piPSC), which fail to express many pluripotency genes, display some early-to-late replication domains that are difficult to be reprogrammed back to the embryonic stem cell state because of their highly condensed chromatin environment (Hiratani et al. 2010). Hence, in rMt the unsuccessful recruitment and activation of replication origins may be associated with chromatin modifications occurring during the differentiation process and representing a potential impediment to the correct loading of pre-replicative complexes and/or firing factors.
Interestingly, the proportion of unidirectional forks in rMt dosed with deoxynucleosides was slightly increased with respect to proliferating myoblasts (Table 6). Although their biological meaning is still unknown, unidirectional forks are peculiar of large origin-poor transition zones dividing early and late replication domains (Hiratani et al. 2008) and it was also demonstrated that the proportion of unidirectional forks in asynchronous cell populations increased along with subsequent cell divisions (as already discussed at p. 62; (Palumbo et al. 2013)). Therefore, it may be possible that moving toward terminal differentiation cells change the replication dynamics within replication domains, which are progressively replicated by unidirectional forks emanating from the surrounding genomic regions.

3.2 Future perspectives

In the light of above considerations, evaluating whether replication timing changes have a role in affecting DNA replication in terminally differentiated cells is an interesting issue to be further considered. In particular, to understand if the effects of the forced cell cycle reactivation on genome duplication are preferentially associated with certain genomic regions, it will be intriguing to determine the replication profiles at different loci both in early and late replication domains as well as in some difficult-to-replicate sequences. In parallel, to identify potential

correlations with replication timing also chromatin environment and gene expression profiles should be analysed.

CONCLUSIONS

In Eukaryotes the plasticity of DNA replication is required for a tight control of the process during development and cell differentiation and to overcome replication stress. However, the lack of a *consensus* sequence and of specific chromatin markers of mammalian replication origins makes challenging their identification with respect to the sequence-specific yeast model. In addition, it was demonstrated that replication profiles are modulated according to cell types and during cell differentiation (Hiratani et al. 2008; Palumbo et al. 2013). Despite precision and coverage of high-throughput genome-wide techniques in defining origin positions, cell-to-cell variability cannot be determined and this important issue would be better appreciated using single-molecule approaches (Tuduri et al. 2010; Técher et al. 2013; Hyrien 2015; Prioleau & MacAlpine 2016). In this study, I monitored whether and how mammalian cells modulate their replication program in response to endogenous and exogenous replication stress conditions.

By molecular combing, I was able to prove that the expansion of a trinucleotide repeat leading to the formation of unusual DNA secondary structures, which are well-known impediments to replication fork progression, induces wide changes in the single-locus replication program of human lymphoblastoid cells. In particular, dormant origin recruitment was required to assure replication of the *FXN* gene in the presence of a GAA-repeat expansion. In addition, a significant reduction of unidirectional fork length was observed in relation to the mutation. Moreover, the possibility to monitor fork progression in both normal and expanded *FXN* alleles allowed the identification of an origin-switch mechanism associated with the GAA-repeat expansion. The same conclusions were reached in a parallel study applying another single-molecule approach, namely SMARD (Gerhardt et al. 2016). Furthermore, I also demonstrated that high-throughput techniques were not sensitive enough to detect physiological, dynamic and non massive events associated with the presence of the trinucleotide repeat expansion (Stevanoni et al. 2016).

Similarly, replication dynamics was evaluated genome-wide by molecular combing in the presence of increased and unbalanced dNTP pools and in diverse stages of cell differentiation. In agreement with the notion that replication stress and altered cell growth conditions lead to modulation of the mammalian replication program (Palumbo et al. 2010; Méchali 2010; Yekezare et al. 2013; Zeman & Cimprich 2014; Alver et al. 2014), wide changes in the replication profiles of human primary fibroblasts were detected after treatment with aphidicolin, an inhibitor of DNA polymerases. In particular, the increased number of activated replication origins correlating with fork rate reduction suggested that fibroblasts activate the DNA damage response to counteract replication stress. However, in normal growth conditions, DNA replication proceeds mostly undisturbed independently of the

dNTP pool imbalances. Alterations of the replication program in human SAMHD1-depleted THP1 monocytes resemble the cellular response to replication stress conditions and interestingly an increased proportion of paused/arrested forks was observed in this cell type. Finally, I assessed genome-wide replication profiles in mouse muscle cells at diverse differentiation stages and under forced cell-cycle reactivation: as in previous models, changes in fork rates and inter-origin distances were involved, but unidirectional forks were also implicated. Indeed, in agreement with previous observations in human primary fibroblasts (Palumbo et al. 2013), an excess of unidirectional forks was found in terminally differentiated mouse myotubes upon forced cell cycle reactivation with respect to proliferating myoblasts. This may be viewed as a potential remnant of the modality by which some genomic regions are replicated when cells move toward terminal differentiation. On the whole, a relevant role of unidirectional forks emerges suggesting that not only bidirectional but also unidirectional initiation events contribute to the determination of mammalian replication profiles. Detection and accurate evaluation of unidirectional forks are crucial because besides a possible role during cell differentiation and development (p. 71) unidirectional forks may be associated with genomic instability, as in the case of CFS or in response to replication stress such as the treatment with aphidicolin (Palumbo et al. 2010).
The complexity of mammalian DNA replication may be ascribed to the intrinsic plasticity of the process, and drawing general conclusions from studies based on individual cellular models may be incautious. The results obtained in this study add new evidence for interpret this complexity, and offer insights for future investigations.

REFERENCES

Abdurashidova, G. et al., 2000. Start sites of bidirectional DNA synthesis at the human lamin B2 origin. *Science (New York, N.Y.)*, 287(5460), pp.2023–2026.

Aguilera, A. & García-Muse, T., 2013. Causes of genome instability. *Annual review of genetics*, 47, pp.1–32.

Alver, R.C., Chadha, G.S. & Blow, J.J., 2014. The contribution of dormant origins to genome stability: From cell biology to human genetics. *DNA Repair*, 19, pp.182–189.

Anglana, M. et al., 2003. Dynamics of DNA replication in mammalian somatic cells: Nucleotide pool modulates origin choice and interorigin spacing. *Cell*, 114(3), pp.385–394.

Beloglazova, N. et al., 2013. Nuclease activity of the human SAMHD1 protein implicated in the Aicardi-Goutièrres syndrome and HIV-1 restriction. *Journal of Biological Chemistry*, 288(12), pp.8101–8110.

Bensimon, A. et al., 1994. Alignment and sensitive detection of DNA by a moving interface. *Science*, 265(5181).

Berti, M. & Vindigni, A., 2016. Replication stress: getting back on track. *Nature Structural & Molecular Biology*, 23(2), pp.103–109.

Besnard, E. et al., 2012. Unraveling cell type–specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nature Structural & Molecular Biology*, 19(8), pp.837–844.

Bester, A.C. et al., 2011. Nucleotide Deficiency Promotes Genomic Instability in Early Stages of Cancer Development. *Cell*, 145(3), pp.435–446.

Bianco, J.N. et al., 2012. Analysis of DNA replication profiles in budding yeast and mammalian cells using DNA combing. *Methods*, 57(2), pp.149–157.

De Biase, I. et al., 2007. Somatic instability of the expanded GAA triplet-repeat sequence in Friedreich ataxia progresses throughout life. *Genomics*, 90(1), pp.1–5.

Bidichandani, S.I. et al., 1999. Somatic sequence variation at the Friedreich ataxia locus includes complete contraction of the expanded GAA triplet repeat, significant length variation in serially passaged lymphoblasts and enhanced mutagenesis in the flanking sequence. *Human molecular genetics*, 8(13), pp.2425–2436.

Birney, E. et al., 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 447(7146), pp.799–816.

Blow, J.J. & Ge, X.Q., 2009. A model for DNA replication showing how dormant origins safeguard against replication fork failure. *EMBO reports*, 10(4), pp.406–412.

Blow, J.J., Ge, X.Q. & Jackson, D.A., 2011. How dormant origins promote complete genome replication. *Trends in Biochemical Sciences*, 36(8), pp.405–414.

Bonifati, S. et al., 2016. SAMHD1 controls cell cycle status, apoptosis and HIV-1 infection in monocytic THP-1 cells. *Virology*, 495, pp.92–100.

Buckland, R.J. et al., 2014. Increased and Imbalanced dNTP Pools Symmetrically Promote Both Leading and Lagging Strand Replication Infidelity. *PLoS genetics*, 10(12), p.e1004846.

Cadoret, J.-C. et al., 2008. Genome-wide studies highlight indirect links between human replication origins and gene regulation. *Proceedings of the National Academy of Sciences*, 105(41), pp.15837–15842.
Campuzano, V. et al., 1996. Friedreich's Ataxia: Autosomal Recessive Disease Caused by an Intronic GAA Triplet Repeat Expansion. *Science*, 271(5254).
Cayrou, C. et al., 2011. Genome-scale analysis of metazoan replication origins reveals their organization in specific but flexible sites defined by conserved features. *Genome Research*, 21(9), pp.1438–1449.
Cayrou, C. et al., 2015. The chromatin environment shapes DNA replication origin organization and defines origin classes. *Genome research*, 25(12), pp.1873–85.
Chabes, A. & Stillman, B., 2007. Constitutively high dNTP concentration inhibits cell cycle progression and the DNA damage checkpoint in yeast Saccharomyces cerevisiae. *Proceedings of the National Academy of Sciences*, 104(4), pp.1183–1188.
Chabes, A.L. et al., 2003. Mouse ribonucleotide reductase R2 protein: a new target for anaphase-promoting complex-Cdh1-mediated proteolysis. *Proceedings of the National Academy of Sciences*, 100(7), pp.3925–9.
Chabosseau, P. et al., 2011. Pyrimidine pool imbalance induced by BLM helicase deficiency contributes to genetic instability in Bloom syndrome. *Nature communications*, 2(May), p.368.
Chan, P.K. et al., 2013. Heterochromatinization induced by GAA-repeat hyperexpansion in Friedreich's ataxia can be reduced upon HDAC inhibition by vitamin B3. *Human Molecular Genetics*, 22(13), pp.2662–2675.
Chandok, G.S. et al., 2012. Effects of Friedreich's ataxia GAA repeats on DNA replication in mammalian cells. *Nucleic Acids Research*, 40(9), pp.3964–3974.
Chanput, W., Mes, J.J. & Wichers, H.J., 2014. THP-1 cell line: An in vitro cell model for immune modulation approach. *International Immunopharmacology*, 23(1), pp.37–45.
Cleary, J.D. & Pearson, C.E., 2005. Replication fork dynamics and dynamic mutations: the fork-shift model of repeat instability. *Trends in Genetics*, 21(5), pp.272–280.
Clifford, R. et al., 2014. SAMHD1 is mutated recurrently in chronic lymphocytic leukemia and is involved in response to DNA damage. *Blood*, 123(7), pp.1021–1031.
Corey, D.R. et al., 2016. Synthetic Nucleic Acids and Treatment of Neurological Diseases. *JAMA Neurology*, 21(9), pp.1115–1142.
Cortez, D., 2015. Preventing replication fork collapse to maintain genome integrity. *DNA Repair*, 32, pp.149–157.
Courbet, S. et al., 2008. Replication fork movement sets chromatin loop size and origin choice in mammalian cells. *Nature*, 455(7212), pp.557–560.
Davidson, M.B. et al., 2012. Endogenous DNA replication stress results in expansion of dNTP pools and a mutator phenotype. *The EMBO journal*, 31(4), pp.895–907.
Deegan, T.D. & Diffley, J.F., 2016. MCM: one ring to rule them all. *Current Opinion in Structural Biology*, 37, pp.145–151.
Dellino, G.I. et al., 2013. Genome-wide mapping of human DNA-replication origins: levels of transcription at ORC1 sites regulate origin selection and replication timing. *Genome research*, 23(1), pp.1–11.
Dellino, G.I. & Pelicci, P.G., 2014. Next-generation sequencing and DNA replication

in human cells: the future has arrived. *Future oncology*, 10(4), pp.683–693.
DePamphilis, M.L. et al., 2006. Regulating the licensing of DNA replication origins in metazoa. *Current Opinion in Cell Biology*, 18(3), pp.231–239.
Descours, B. et al., 2012. SAMHD1 restricts HIV-1 reverse transcription in quiescent CD4+ T-cells. *Retrovirology*, 9(1), p.87.
Dhar, M.K., Sehgal, S. & Kaul, S., 2012. Structure, replication efficiency and fragility of yeast ARS elements. *Research in Microbiology*, 163(4), pp.243–253.
Dion, V. & Wilson, J.H., 2009. Instability and chromatin structure of expanded trinucleotide repeats. *Trends in Genetics*, 25(7), pp.288–297.
Ditch, S. et al., 2009. Progressive GAA·TTC Repeat Expansion in Human Cell Lines H. Orr, ed. *PLoS Genetics*, 5(10), p.e1000704.
Douglas, M.E. & Diffley, J.F.X., 2012. Replication timing: The early bird catches the worm. *Current Biology*, 22(3), pp.R81–R82.
Duzdevich, D. et al., 2015. The Dynamics of Eukaryotic Replication Initiation: Origin Specificity, Licensing, and Firing at the Single-Molecule Level. *Molecular Cell*, 58(3), pp.483–494.
Edenberg, E.R., Downey, M. & Toczyski, D., 2014. Polymerase stalling during replication, transcription and translation. *Current Biology*, 24(10), pp.R445–R452.
Evans-Galea, M. V. et al., 2013. Epigenetic modifications in trinucleotide repeat diseases. *Trends in Molecular Medicine*, 19(11), pp.655–663.
Evertts, A.G. & Coller, H.A., 2012. Back to the Origin: Reconsidering Replication, Transcription, Epigenetics, and Cell Cycle Control. *Genes & Cancer*, 3(11–12), pp.678–696.
Follonier, C. et al., 2013. Friedreich's ataxia-associated GAA repeats induce replication-fork reversal and unusual molecular junctions. *Nature structural & molecular biology*, 20(4), pp.486–94.
Foti, R. et al., 2016. Nuclear Architecture Organized by Rif1 Underpins the Replication-Timing Program. *Molecular Cell*, 61(2), pp.260–273.
Foulk, M.S. et al., 2015. Characterizing and controlling intrinsic biases of lambda exonuclease in nascent strand sequencing reveals phasing between nucleosomes and G-quadruplex motifs around a subset of human replication origins. *Genome Research*, 25(5), pp.725–735.
Fragkos, M. et al., 2015. DNA replication origin activation in space and time. *Nature reviews. Molecular cell biology*, 16(6), pp.360–74.
Franzolin, E. et al., 2013. The deoxynucleotide triphosphohydrolase SAMHD1 is a major regulator of DNA precursor pools in mammalian cells. *Proceedings of the National Academy of Sciences*, 110(35), pp.14272–14277.
García-Muse, T. & Aguilera, A., 2016. Transcription–replication conflicts: how they occur and how they are resolved. *Nature Reviews Molecular Cell Biology*, 17(9), pp.553–563.
Ge, X.Q. et al., 2015. Genome against Replication Stress. *Stem Cell Reports*, 5, pp.1–10.
Ge, X.Q., Jackson, D. a & Blow, J.J., 2007. Dormant origins licensed by excess Mcm2-7 are required for human cells to survive replicative stress. *Genes & development*, 21(24), pp.3331–41.
Gemble, S. et al., 2015. Pyrimidine Pool Disequilibrium Induced by a Cytidine Deaminase Deficiency Inhibits PARP-1 Activity, Leading to the Under Replication of DNA. *PLoS genetics*, 11(7), p.e1005384.

Gerhardt, J. et al., 2016. Stalled DNA Replication Forks at the Endogenous GAA Repeats Drive Repeat Expansion in Friedreich's Ataxia Cells. *Cell Reports*, 16(5), pp.1218–1227.

Gerhardt, J. et al., 2014. The DNA Replication Program Is Altered at the FMR1 Locus in Fragile X Embryonic Stem Cells. *Molecular Cell*, 53(1), pp.19–31.

Giacca, M., Pelizon, C. & Falaschi, a, 1997. Mapping replication origins by quantifying relative abundance of nascent DNA strands using competitive polymerase chain reaction. *Methods (San Diego, Calif.)*, 13(3), pp.301–312.

Gilbert, D.M., 2007. Replication origin plasticity, Taylor-made: inhibition vs recruitment of origins under conditions of replication stress. *Chromosoma*, 116(4), pp.341–347.

Gilbert, D.M., 2012. Replication origins run (ultra) deep. *Nature Structural & Molecular Biology*, 19(8), pp.740–742.

Gilbert, D.M., Takebayashi, S. & Ryba, T., 2010. Space and Time in the Nucleus : Developmental Control of Replication Timing and Chromosome Architecture. , LXXV, pp.143–153.

Gomes-Pereira, M. et al., 2014. Disease-associated CAG·CTG triplet repeats expand rapidly in non-dividing mouse cells, but cell cycle arrest is insufficient to drive expansion. *Nucleic Acids Research*, 42(11), pp.7047–7056.

Göndör, A. & Ohlsson, R., 2009. Replication timing and epigenetic reprogramming of gene expression: a two-way relationship? *Nature Reviews Genetics*, 10(4), pp.269–276.

Grabczyk, E., Mancuso, M. & Sammarco, M.C., 2007. A persistent RNA.DNA hybrid formed by transcription of the Friedreich ataxia triplet repeat in live bacteria, and by T7 RNAP in vitro. *Nucleic acids research*, 35(16), pp.5351–9.

Gray, S.J. et al., 2007. An origin of DNA replication in the promoter region of the human fragile X mental retardation (FMR1) gene. *Molecular and cellular biology*, 27(2), pp.426–437.

Groh, M. et al., 2014. R-loops Associated with Triplet Repeat Expansions Promote Gene Silencing in Friedreich Ataxia and Fragile X Syndrome. *PLoS Genetics*, 10(5), p.e1004318.

Guilbaud, G. et al., 2011. Evidence for sequential and increasing activation of replication origins along replication timing gradients in the human genome. *PLoS computational biology*, 7(12), p.e1002322.

Hamperl, S. & Cimprich, K.A., 2016. Conflict Resolution in the Genome: How Transcription and Replication Make It Work. *Cell*, 167(6), pp.1455–1467.

Hansen, R.S. et al., 2010. Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proceedings of the National Academy of Sciences*, 107(1), pp.139–144.

Hellenbroich, Y., Schwinger, E. & Zühlke, C., 2001. Limited somatic mosaicism for Friedreich's ataxia GAA triplet repeat expansions identified by small pool PCR in blood leukocytes. *Acta neurologica Scandinavica*, 103(3), pp.188–192.

Helmrich, A. et al., 2013. Transcription-replication encounters, consequences and genomic instability. *Nature structural & molecular biology*, 20(4), pp.412–8.

Herrick, J. & Bensimon, A., 2008. Global regulation of genome duplication in eukaryotes: an overview from the epifluorescence microscope. *Chromosoma*, 117(3), pp.243–260.

Hiratani, I. et al., 2010. Genome-wide dynamics of replication timing revealed by in vitro models of mouse embryogenesis. *Genome research*, 20(2), pp.155–69.

Hiratani, I. et al., 2008. Global Reorganization of Replication Domains During Embryonic Stem Cell Differentiation S. M. Gasser, ed. *PLoS Biology*, 6(10), p.e245.

Hiratani, I. & Gilbert, D.M., 2009. Replication timing as an epigenetic mark. *Epigenetics*, 4(2), pp.93–97.

Hoffman, E.A. et al., 2015. Break-seq reveals hydroxyurea-induced chromosome fragility as a result of unscheduled conflict between DNA replication and transcription. *Genome Research*, 25(3), pp.402–412.

Hrecka, K. et al., 2011. Vpx relieves inhibition of HIV-1 infection of macrophages mediated by the SAMHD1 protein. *Nature*, 474(7353), pp.658–661.

Hyrien, O. et al., 2013. From simple bacterial and archaeal replicons to replication N/U-domains. *Journal of Molecular Biology*, 425(23), pp.4673–4689.

Hyrien, O., 2015. Peaks cloaked in the mist: The landscape of mammalian replication origins. *The Journal of Cell Biology*, 208(2), pp.147–160.

Hyrien, O., Marheineke, K. & Goldar, A., 2003. Paradoxes of eukaryotic DNA replication: MCM proteins and the random completion problem. *BioEssays*, 25(2), pp.116–125.

Ilves, I. et al., 2010. Activation of the MCM2-7 helicase by association with Cdc45 and GINS proteins. *Molecular cell*, 37(2), pp.247–58.

Karnani, N. et al., 2010. Genomic study of replication initiation in human chromosomes reveals the influence of transcription regulation and chromatin structure on origin selection. *Molecular biology of the cell*, 21(3), pp.393–404.

Karnani, N. et al., 2007. Pan-S replication patterns and chromosomal domains defined by genome-tiling arrays of ENCODE genomic areas. *Genome research*, 17(6), pp.865–76.

Kaykov, A. et al., 2016. Molecular Combing of Single DNA Molecules on the 10 Megabase Scale. *Scientific Reports*, 6, p.19636.

Kim, H.H.-M. et al., 2008. Chromosome fragility at GAA tracts in yeast depends on repeat orientation and requires mismatch repair. *The EMBO journal*, 27(21), pp.2896–906.

Kim, J.C. & Mirkin, S.M., 2013. The balancing act of DNA repeat expansions. *Current Opinion in Genetics and Development*, 23(3), pp.280–288.

Kohnken, R., Kodigepalli, K.M. & Wu, L., 2015. Regulation of deoxynucleotide metabolism in cancer: novel mechanisms and therapeutic implications. *Molecular cancer*, 14, p.176.

Krasilnikova, M.M. & Mirkin, S.M., 2004. Replication stalling at Friedreich's ataxia (GAA)n repeats in vivo. *Molecular and cellular biology*, 24(6), pp.2286–95.

Kumar, D. et al., 2010. Highly mutagenic and severely imbalanced dNTP pools can escape detection by the S-phase checkpoint. *Nucleic acids research*, 38(12), pp.3975–83.

Laguette, N. et al., 2011. SAMHD1 is the dendritic- and myeloid-cell-specific HIV-1 restriction factor counteracted by Vpx. *Nature*, 474(7353), pp.654–7.

Langley, A.R. et al., 2016. Genome-wide identification and characterisation of human DNA replication origins by initiation site sequencing (ini-seq). *Nucleic acids research*, p.gkw760.

Latella, L., Sacchi, a & Crescenzi, M., 2000. Long-term fate of terminally differentiated skeletal muscle cells following E1A-initiated cell cycle reactivation. *Cell death and differentiation*, 7(2), pp.145–54.

Lebofsky, R. & Bensimon, A., 2003. Single DNA molecule analysis: applications of

molecular combing. *Briefings in Functional Genomics & Proteomics*, 1(4), pp.385–396.
Lee, D.-Y. & McMurray, C.T., 2014. Trinucleotide expansion in disease: why is there a length threshold? *Current Opinion in Genetics & Development*, 26, pp.131–140.
Lemée, F. et al., 2010. DNA polymerase theta up-regulation is associated with poor survival in breast cancer, perturbs DNA replication, and promotes genetic instability. *Proceedings of the National Academy of Sciences*, 107(30), pp.13390–5.
León-Ortiz, A.M., Svendsen, J. & Boulton, S.J., 2014. Metabolism of DNA secondary structures at the eukaryotic replication fork. *DNA Repair*, 19, pp.152–162.
Leonard, A.C. & Méchali, M., 2013. DNA replication origins. *Cold Spring Harbor perspectives in biology*, 5(10), p.a010116.
Letessier, A. et al., 2011. Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. *Nature*, 470(7332), pp.120–123.
Li, N., Zhang, W. & Cao, X., 2000. Identification of human homologue of mouse IFN-γ induced protein from human dendritic cells. *Immunology Letters*, 74(3), pp.221–224.
Li, Y. et al., 2015. Expanded GAA repeats impede transcription elongation through the *FXN* gene and induce transcriptional silencing that is restricted to the *FXN* locus. *Human Molecular Genetics*, 24(24), p.ddv397.
Lutzmann, M. et al., 2008. MCM9 binds Cdt1 and is required for the assembly of prereplication complexes. *Molecular cell*, 31(2), pp.190–200.
MacAlpine, D.M., 2016. ORChestrating the human DNA replication program. *Proceedings of the National Academy of Sciences*, 113(33), pp.9136–9138.
Magdalou, I. et al., 2014. The causes of replication stress and their consequences on genome stability and cell fate. *Seminars in Cell and Developmental Biology*, 30, pp.154–164.
Malinsky, J. et al., 2001. The supply of exogenous deoxyribonucleotides accelerates the speed of the replication fork in early S-phase. *Journal of Cell Science*, 114(4).
Martin, M.M. et al., 2011. Genome-wide depletion of replication initiation events in highly transcribed regions. *Genome Research*, 21(11), pp.1822–1832.
McIntosh, D. & Blow, J.J., 2012. Dormant origins, the licensing checkpoint, and the response to replicative stresses. *Cold Spring Harbor perspectives in biology*, 4(10), p.a012955-.
McMurray, C.T., 2010. Mechanisms of trinucleotide repeat instability during human development. *Nature reviews. Genetics*, 11(11), pp.786–99.
Méchali, M., 2010. Eukaryotic DNA replication origins: many choices for appropriate answers. *Nature Reviews Molecular Cell Biology*, 11(10), pp.728–738.
Mesner, L.D. et al., 2011. Bubble-chip analysis of human origin distributions demonstrates on a genomic scale significant clustering into zones and significant association with transcription. *Genome Research*, 21(2), pp.377–389.
Mesner, L.D. et al., 2013. Bubble-seq analysis of the human genome reveals distinct chromatin-mediated mechanisms for regulating early- and late-firing origins. *Genome Research*, 23(11), pp.1774–1788.
Miazzi, C. et al., 2014. Allosteric regulation of the human and mouse

deoxyribonucleotide triphosphohydrolase sterile α-motif/histidine-aspartate domain-containing protein 1 (SAMHD1). *Journal of Biological Chemistry*, 289(26), pp.18339–18346.
Michalet, X. et al., 1997. Dynamic Molecular Combing: Stretching the Whole Human Genome for High-Resolution Studies. *Science*, 277(5331).
Miotto, B., Ji, Z. & Struhl, K., 2016. Selectivity of ORC binding sites and the relation to replication timing, fragile sites, and deletions in cancers. *Proceedings of the National Academy of Sciences*, 113(33), pp.E4810–E4819.
Mirkin, E. V & Mirkin, S.M., 2007. Replication fork stalling at natural impediments. *Microbiology and Molecular Biology Reviews : MMBR*, 71(1), pp.13–35.
Mirkin, S.M., 2007. Expandable DNA repeats and human disease. *Nature*, 447(7147), pp.932–940.
Mirkin, S.M. & Smirnova, E. V, 2002. Positioned to Expand. *Nature Genetics*, 31(1), pp.5–6.
Naim, V. et al., 2013. ERCC1 and MUS81–EME1 promote sister chromatid separation by processing late replication intermediates at common fragile sites during mitosis. *Nature Cell Biology*, 15(8), pp.1008–1015.
Nelson, D.L., Orr, H.T. & Warren, S.T., 2013. The Unstable Repeats—Three Evolving Faces of Neurological Disease. *Neuron*, 77(5), pp.825–843.
Nordlund, P. & Reichard, P., 2006. Ribonucleotide Reductases. *Annual Review of Biochemistry*, 75(1), pp.681–706.
Norio, P. et al., 2005. Progressive activation of DNA replication initiation in large domains of the immunoglobulin heavy chain locus during B cell development. *Molecular Cell*, 20, pp.575–587.
Norio, P. & Schildkraut, C.L., 2004. Plasticity of DNA replication initiation in Epstein-Barr virus episomes. *PLoS biology*, 2(6), p.e152.
Norio, P. & Schildkraut, C.L., 2001. Visualization of DNA replication on individual Epstein-Barr virus episomes. *Science (New York, N.Y.)*, 294(5550), pp.2361–2364.
O'Donnell, M., 2006. Replisome architecture and dynamics in Escherichia coli. *The Journal of biological chemistry*, 281(16), pp.10653–6.
O'Keefe, R.T., Henderson, S.C. & Spector, D.L., 1992. Dynamic organization of DNA replication in mammalian cell nuclei: spatially and temporally defined replication of chromosome-specific alpha-satellite DNA sequences. *The Journal of Cell Biology*, 116(5), pp.1095–110.
Ostankovitch, M. & Debatisse, M., 2013. From the Replicon to Replication Programs in Space and Time: Regulation of DNA Replication and Implications for Genomic Instability. *Journal of Molecular Biology*, 425(23), pp.4659–4662.
Ozeri-Galai, E. et al., 2011. Failure of Origin Activation in Response to Fork Stalling Leads to Chromosomal Instability at Fragile Sites. *Molecular Cell*, 43(1), pp.122–131.
Pajalunga, D. et al., 2017. A defective dNTP pool hinders DNA replication in cell cycle-reactivated terminally differentiated cells. *Cell Death & Differentiation*, Accepted on January 2017. doi: 10.1038/cdd.2017.4
Pajalunga, D. et al., 2010. DNA replication is intrinsically hindered in terminally differentiated myotubes. *PLoS ONE*, 5(7), p.e11559.
Pajalunga, D. et al., 2007. Non-Proliferation as an Active State: Conceptual and Practical Implications. *Cell Cycle*, 6(12), pp.1414–1417.
Palumbo, E. et al., 2010. Replication dynamics at common fragile site FRA6E.

Chromosoma, 119(6), pp.575–587.
Palumbo, E., Tosoni, E. & Russo, A., 2013. General and specific replication profiles are detected in normal human cells by genome-wide and single-locus molecular combing. *Experimental Cell Research*, 319(20), pp.3081–3093.
Pandolfo, M., 2008. Friedreich Ataxia. *Archives of Neurology*, 65(10), pp.589–620.
Pearson, C.E., Edamura, K.N. & Cleary, J.D., 2005. Repeat instability: mechanisms of dynamic mutations. *Nature Reviews Genetics*, 6(10), pp.729–742.
Petryk, N. et al., 2016. Replication landscape of the human genome. *Nature Communications*, 7, p.10208.
Plosky, B.S., 2015. Replication Origin Specification Gets a Push. *Molecular Cell*, 60(5), pp.711–712.
Poli, J. et al., 2012. dNTP pools determine fork progression and origin usage under replication stress. *The EMBO Journal*, 31(4), pp.883–894.
Pontarin, G. et al., 2012. Mammalian ribonucleotide reductase subunit p53R2 is required for mitochondrial DNA replication and DNA repair in quiescent cells. *Proceedings of the National Academy of Sciences*, 109(33), pp.13302–7.
Pope, B.D. et al., 2014. Topologically associating domains are stable units of replication-timing regulation. *Nature*, 515(7527), pp.402–405.
Pope, B.D. & Gilbert, D.M., 2013. The replication domain model: Regulating replicon firing in the context of large-scale chromosome architecture. *Journal of Molecular Biology*, 425(23), pp.4690–4695.
Prioleau, M.-N. & MacAlpine, D.M., 2016. DNA replication origins-where do we begin? *Genes & development*, 30(15), pp.1683–97.
Rampazzo, C. et al., 2010. Regulation by degradation, a cellular defense against deoxyribonucleotide pool imbalances. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, 703(1), pp.2–10.
Reddy, K. et al., 2014. Processing of double-R-loops in (CAG)·(CTG) and C9orf72 (GGGGCC)·(GGCCCC) repeats causes instability. *Nucleic acids research*, 42(16), pp.10473–87.
Remus, D. & Diffley, J.F., 2009. Eukaryotic DNA replication control: Lock and load, then fire. *Current Opinion in Cell Biology*, 21(6), pp.771–777.
Rice, G.I. et al., 2009. Mutations involved in Aicardi-Goutières syndrome implicate SAMHD1 as regulator of the innate immune response. *Nature genetics*, 41(7), pp.829–32.
Rimmelé, P. et al., 2010. Spi-1/PU.1 Oncogene Accelerates DNA Replication Fork Elongation and Promotes Genetic Instability in the Absence of DNA Breakage. *Cancer Research*, 70(17).
Rivera-Mulia, J.C. et al., 2015. Dynamic changes in replication timing and gene expression during lineage specification of human pluripotent stem cells. *Genome Research*, 25(8), pp.1091–1103.
Rivera-Mulia, J.C. & Gilbert, D.M., 2016a. Replicating Large Genomes: Divide and Conquer. *Molecular Cell*, 62(5), pp.756–765.
Rivera-Mulia, J.C. & Gilbert, D.M., 2016b. Replication timing and transcriptional control: Beyond cause and effect - part III. *Current Opinion in Cell Biology*, 40, pp.168–178.
Ryba, T. et al., 2011. Replication timing: a fingerprint for cell identity and pluripotency. *PLoS computational biology*, 7(10), p.e1002225.
Ryoo, J. et al., 2014. The ribonuclease activity of SAMHD1 is required for HIV-1 restriction. *Nature Medicine*, 20(8), pp.936–941.

Sander, J.D. & Joung, J.K., 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature biotechnology*, 32(4), pp.347–55.

Schaller, T., Goujon, C. & Malim, M.H., 2012. HIV Interplay with SAMHD1. *Science*, 335(6074).

Sequeira-Mendes, J. et al., 2009. Transcription Initiation Activity Sets Replication Origin Efficiency in Mammalian Cells W. A. Bickmore, ed. *PLoS Genetics*, 5(4), p.e1000446.

Sharma, R. et al., 2002. The GAA triplet-repeat sequence in Friedreich ataxia shows a high level of somatic instability in vivo, with a significant predilection for large contractions. *Human Molecular Genetics*, 11(18), pp.2175–87.

Shishkin, A.A. et al., 2009. Large-Scale Expansions of Friedreich's Ataxia GAA Repeats in Yeast. *Molecular Cell*, 35(1), pp.82–92.

Silva, A.M. et al., 2015. Expanded GAA repeats impair FXN gene expression and reposition the FXN locus to the nuclear lamina in single cells. *Human Molecular Genetics*, 24(12), pp.3457–3471.

Skoog, L. & Nordenskj, B., 1971. Effects of Hydroxyurea and 1-beta-D-Arabinofuranosyl-cytosine on Deoxyribonucleotide Pools in Mouse Embryo Cells. *European Journal of Biochemistry*, 19(1), pp.81–89.

Sogo, J.M., Lopes, M. & Foiani, M., 2002. Fork Reversal and ssDNA Accumulation at Stalled Replication Forks Owing to Checkpoint Defects. *Science*, 297(5581).

Stevanoni, M., Palumbo, E. & Russo, A., 2016. The Replication of Frataxin Gene Is Assured by Activation of Dormant Origins in the Presence of a GAA-Repeat Expansion. *PLOS Genetics*, 12(7), p.e1006201.

Técher, H. et al., 2013. Replication dynamics: Biases and robustness of DNA fiber analysis. *Journal of Molecular Biology*, 425(23), pp.4845–4855.

Técher, H. et al., 2016. Signaling from Mus81-Eme2-Dependent DNA Damage Elicited by Chk1 Deficiency Modulates Replication Fork Speed and Origin Usage. *Cell Reports*, 14(5), pp.1114–1127.

Teytelman, L. et al., 2013. Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. *Proceedings of the National Academy of Sciences*, 110(46), pp.18602–18607.

Therizols, P. et al., 2014. Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells. *Science*, 346(6214), pp.1238–1242.

Tiainen, M et al., 1996. Terminally differentiated skeletal myotubes are not confined to G0 but can enter G1 upon growth factor stimulation. *Cell growth & differentiation : the molecular biology journal of the American Association for Cancer Research*, 7(8), pp.1039–50.

Ticau, S. et al., 2015. Single-Molecule Studies of Origin Licensing Reveal Mechanisms Ensuring Bidirectional Helicase Loading. *Cell*, 161(3), pp.513–525.

Tuduri, S., Tourrière, H. & Pasero, P., 2010. Defining replication origin efficiency using DNA fiber assays. *Chromosome Research*, 18(1), pp.91–102.

Tüngler, V. et al., 2013. Single-stranded nucleic acids promote SAMHD1 complex formation. *Journal of Molecular Medicine*, 91(6), pp.759–770.

Watt, D.L. et al., 2015. Genome-wide analysis of the specificity and mechanisms of replication infidelity driven by imbalanced dNTP pools. *Nucleic acids research*, 44(4).

Williams, L.N. et al., 2015. dNTP pool levels modulate mutator phenotypes of error-prone DNA polymerase ε variants. *Proceedings of the National Academy of*

Sciences, 112(19), pp.E2457-66.
Woodward, A.M. et al., 2006. Excess Mcm2-7 license dormant origins of replication that can be used under conditions of replicative stress. *The Journal of cell biology*, 173(5), pp.673–83.
Yandim, C., Natisvili, T. & Festenstein, R., 2013. Gene regulation and epigenetics in Friedreich's ataxia. *Journal of Neurochemistry*, 126(SUPPL.1), pp.21–42.
Yeeles, J.T.P. et al., 2015. Regulated eukaryotic DNA replication origin firing with purified proteins. *Nature*, 519(7544), pp.431–435.
Yekezare, M., Gómez-González, B. & Diffley, J.F.X., 2013. Controlling DNA replication origins in response to DNA damage - inhibit globally, activate locally. *Journal of Cell Science*, 126(Pt 6), pp.1297–1306.
Zeman, M.K. & Cimprich, K. a, 2014. Causes and consequences of replication stress. *Nature cell biology*, 16(1), pp.2–9.

ABBREVIATIONS

AGS = Aicardi-Goutièrres syndrome
APH = aphidicolin
ARS = Autonomous Recognition Sequence
CDK = Cyclin-Dependent Kinase
CFS = common fragile site
ChIP-seq = Chromatin ImmunoPrecipitation sequencing
CKI = Cyclin-dependent Kinase Inhibitor
CldU = 5-Chloro-2'-deoxyuridine
DDK = Dbf4-dependent kinase
dNTP = deoxynucleotide triphosphate
ESC = Embryonic Stem Cell
FRDA = Friedreich's ataxia
FMR1 = *Fragile X mental retardation 1* gene
FXN = *Frataxin* gene
HU = hydroxyurea
IdU = 5-Iodo-2'-deoxyuridine
Ini-seq = initiation site sequencing
iPSC = induced Pluripotent Stem Cell
MCM2-7 = Mini-Chromosome Maintenance complex
OK-seq = Okazaki fragment sequencing
ORC = Origin Recognition Complex
piPSC = partially reprogrammed induced Pluripotent Stem Cell
pre-RC = pre-replicative complex
qRT-PCR = quantitative retro-transcribed polymerase chain reaction
rMt = reactivated myotubes
RNR = ribonucleotide reductase
SAMHD1 = Sterile alpha motif and HD-domain containing protein 1
SMARD = Single-Molecule Analysis of Replicated DNA
SNS = small nascent strand
SNS-seq = small nascent strand sequencing
ssDNA = single-stranded DNA
TD = terminal differentiated
TDP = Timing Decision Point
TNR = trinucleotide repeat
THP1 KO = Knock-out THP1 cells

ACKNOWLEDGEMENTS

Special thanks to Elisa Palumbo (Dept. of Biology, University of Padova) for her precious and useful advices during my whole PhD activity, for supporting and sustaining me during the job and also for the amazing free-time in the lab.

Thanks to the Prof. Bianchi's lab (Dept. of Biology, University of Padova) and in particular to Chiara Rampazzo, Elisa Franzolin, Giovanna Pontarin and Paola Ferraro for the great collaboration and for their suggestions.

Thanks also to Deborah Pajalunga from the Dr. Crescenzi's lab for the fruitful collaboration and for her willingness.

I acknowledge the skilled assistance of Anna Cabrelle (Venetian Institute of Molecular Medicine (VIMM), Padova) in the FACS sorting procedure.
Thanks to Chiara Romualdi (Dept. of Biology, University of Padova) for her advice in the statistical analysis of qRT-PCR results.
I am grateful to Olivier Hyrien (Institut de Biologie de l'Ecole Normale Supérieure, Paris, France) and Chunlong Chen (Institute Curie, Paris, France) for the helpful discussion on OK-seq data.

APPENDIX A. Reprint and supplementary material of the article "The Replication of Frataxin Gene Is Assured by Activation of Dormant Origins in the Presence of a GAA-Repeat Expansion" by Stevanoni *et al*.

The complete reference is:
Stevanoni M, Palumbo E, Russo A (2016) The Replication of Frataxin Gene Is Assured by Activation of Dormant Origins in the Presence of a GAA-Repeat Expansion. PLoS Genet 12(7): e1006201. doi:10.1371/journal.pgen.1006201

The resolution of the supplementary material is reduced because of space limitation for thesis upload. The full article and high resolution images are deposited at:
http://dx.doi.org/10.1371/journal.pgen.1006201

PLOS | GENETICS

RESEARCH ARTICLE

The Replication of *Frataxin* Gene Is Assured by Activation of Dormant Origins in the Presence of a GAA-Repeat Expansion

Martina Stevanoni[☯], **Elisa Palumbo**[☯], **Antonella Russo***

Department of Biology, University of Padova, Padova, Italy

☯ These authors contributed equally to this work.
* antonella.russo@unipd.it

OPEN ACCESS

Citation: Stevanoni M, Palumbo E, Russo A (2016) The Replication of *Frataxin* Gene Is Assured by Activation of Dormant Origins in the Presence of a GAA-Repeat Expansion. PLoS Genet 12(7): e1006201. doi:10.1371/journal.pgen.1006201

Editor: Christopher E. Pearson, The Hospital for Sick Children and University of Toronto, CANADA

Received: June 26, 2015

Accepted: June 27, 2016

Published: July 22, 2016

Data Availability Statement: All relevant data are within the paper and its Supporting Information files. Microscope images reconstructed for molecular combing analyses are available in Dryad (http://dx.doi.org/10.5061/dryad.f12cg).

Funding: This work was supported by the Italian Ministry of University and Research (PRIN2009, 2009MS78BE) and by the Department of Biology, University of Padova. EP was the recipient of a post-doc position from the University of Padova (CPDR095917). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

It is well known that DNA replication affects the stability of several trinucleotide repeats, but whether replication profiles of human loci carrying an expanded repeat differ from those of normal alleles is poorly understood in the endogenous context. We investigated this issue using cell lines from Friedreich's ataxia patients, homozygous for a GAA-repeat expansion in intron 1 of the *Frataxin* gene. By interphase, FISH we found that in comparison to the normal *Frataxin* sequence the replication of expanded alleles is slowed or delayed. According to molecular combing, origins never fired within the normal *Frataxin* allele. In contrast, in mutant alleles dormant origins are recruited within the gene, causing a switch of the prevalent fork direction through the expanded repeat. Furthermore, a global modification of the replication profile, involving origin choice and a differential distribution of unidirectional forks, was observed in the surrounding 850 kb region. These data provide a wide-view of the interplay of events occurring during replication of genes carrying an expanded repeat.

Author Summary

The expansion of trinucleotide repeats (TNR) is associated with a large number of human neurodegenerative and neuromuscular diseases, among which the most known are Friedreich's ataxia (GAA/TTC), Huntington's disease (CAG/CTG), and myotonic dystrophy (CTG/CAG). TNR are among the most unstable DNA regions in the genome, and an important step in their expansion is the attainment of a threshold-length. This process occurs during paternal or maternal gametogenesis, leading to an earlier onset of the disease in the next generation. The severity of the disease is strictly related to the TNR length. The repeat instability results from non-B DNA secondary structures formed during DNA replication, repair, recombination and gene transcription. However, the pathways leading to expansion remain poorly understood. Here, we describe the effects of the GAA-expansion on the DNA replication of *Frataxin* gene. By analyzing the replication profile of mutated and normal genotypes, we have found that the replication of the expanded gene is slowed or delayed in comparison with the non-expanded condition. Interestingly, this

Competing Interests: The authors have declared that no competing interests exist.

observation accords to a global modification of the replication profile affecting the usage of both replication forks and origins, which can be referred as the functional units of any replication program.

Introduction

During DNA replication the cell must be ready to face diverse potential obstacles to fork progression, including changes in chromatin organization, variations in cellular environment, formation of secondary structures [1–3]. To deal with these adverse conditions and ensure accurate genome duplication, mammalian cells rely on the plasticity of the replication process, which can be appreciated both at the global and local level [4–6].

It is well-known that DNA replication may affect the stability of several trinucleotide repeats [7–9]. Evidence was accumulated by a wide range of experimental systems, including bacteria, yeast, transfected or engineered human cells [8,10–13]. However, whether replication profiles of human loci carrying an expanded repeat differ from those of normal alleles is poorly understood in the endogenous context. A fine characterization of the replication profiles of loci involved in trinucleotide-expansion human diseases could be of general interest, because knowledge concerning the replication dynamics at unstable genomic regions is still limited [4,14]. In addition, this information could help to define the replication-based mechanisms causing instability of trinucleotide repeats [7,15].

In relation to the orientation of the repeat and the distance from a replication origin, secondary structures may have a diverse potential to be formed, to be stable, and eventually to cause replication impediments and trinucleotide length variations [16]. One model called origin-switch predicts that a change in the position of a replication origin across the repeat may lead to opposite orientations of normal and expanded alleles in the two template strands [17–19]. A recent study describing the replication profile of the *FMR1* locus, which is involved in fragile X syndrome when a CGG-repeat is expanded, strongly support an origin-switch mechanism at the basis of the CGG-repeat expansion in early developmental stages [20].

Human subjects affected by Friedreich's ataxia (FRDA) are homozygous for a GAA-repeat expansion in intron 1 of *Frataxin* (*FXN*) gene [21], a mutation causing the transcriptional inhibition of the gene [22–24]. In proximity of the expansion the chromatin is remodeled leading to sequence heterochromatinization [25–28]. Somatic instability of the expansion has been reported in several tissues of FRDA patients and the same effect may be observed in mutated lymphoblastoid cell lines [29–31]. There is large agreement concerning the involvement of a replication-based mechanism at the origin of GAA-repeat instability in FRDA patients. However, evidence was mainly derived from model systems, and by tracking the events occurring at the repeated sequence only [10–12,32].

To verify if mammalian cells modulate origin usage and fork rates in the presence of long stretches of GAA repeats, we used cell lines derived from FRDA patients. A mild shift of *FXN* replication timing was detected in patients' cells. By monitoring fork progression in a wide genomic segment surrounding *FXN*, we found evidence for recruitment of dormant origins, which may be consistent with an origin-switch effect at the GAA-expanded repeat.

Results

Cell line characterization

GM15850 and GM16227 cell lines (from two unrelated FRDA patients), and GM15851 cells derived from the healthy brother of patient GM15850, were obtained from the Coriell cell line

repository. In the lack of available cell samples from healthy relatives of the second patient, the EBV-immortalized B lymphoblastoid H691 cell line, derived from a male subject and previously used in our laboratory for replication studies [14], was used as a further control.

The three Coriell cell lines were thoroughly characterized by genotype, transcriptional, cell cycle and replication analysis (S1 Fig). *FXN* genotype and transcriptional activity were assessed also in the H691 cell line (S1B and S1C Fig).

The size of the GAA-repeat expansion in the two patients' cell lines was evaluated by long-range PCR at the beginning and the conclusion of the study. The results were in agreement with the information provided by the Coriell cell repository; furthermore, somatic instability of the expansion was excluded on the basis of the lack of multiple bands relative to the amplification of the expanded GAA-repeats (S1A and S1B Fig).

As expected, the transcriptional inhibition of the mutated alleles was observed in the patients' cell lines (S1C Fig).

According to flow cytometry-based cell cycle distributions (S1D Fig), no detectable differences were found between mutated and normal cells. In addition, when DNA replication profiles were assayed by molecular combing (S1E Fig), both replication fork rates and Inter-Origin Distances (IOD) fell into the ranges known for lymphoblastoid cells [14,33].

Replication timing of the *FXN* gene

According to data provided by the Encode project *FXN* is harbored in a mid-late replicating domain [34]. We wondered if the long GAA-repeat expansion found in mutant alleles (ranging 630–1030 repeats in our samples) could affect the replication timing of the gene. To answer this question, interphase FISH experiments were performed in FRDA cells (GM15850 and GM16227 cell lines) and in control cells GM15851, under the assumption that nuclei showing two single FISH spots (SS) carried non-replicated alleles, while cells showing a pair of duplicated FISH signals (DD) had already completed the replication of both alleles. Asynchronous patterns with a single and a duplicated FISH signal (SD) can also be observed [35]. In the case of *FXN* gene, these patterns were similarly represented in the three cell populations, independent of the presence of a mutated or a normal pair of alleles (Fig 1; S1 Table). This result might indicate that no differences exist between mutant and normal *FXN* alleles. In parallel, the late replicating sequence of the common fragile site *FRA3B* was evaluated as a positive control. This locus displayed a DD pattern in less than 25% nuclei (Fig 1; S1 Table), suggesting that *FRA3B* replication is slightly postponed with respect to the *FXN* timing. From this data we could confirm a mid-late replication timing for *FXN* locus.

However, we reasoned that the sensitivity of the interphase approach could not be sufficient to detect mild temporal replication shifts, in particular if mid-late or late replicating regions are investigated. Indeed, in mid-late replicating domains the activation of origins is less efficient and more stochastic than in early domains, leading to an increase of cell-to-cell variability [36,37]. Thus, to examine in depth the replication timing of *FXN* we performed FACS sorting experiments coupled with interphase FISH. Four cell fractions of identical size and corresponding to early-to-late S-phase (S1-S4) were isolated, and the level of contamination recorded by post-sorting FACS analysis appeared very small (S2 Fig).

As sorted cells have been CldU-labeled immediately before harvesting, in the course of the following microscope analyses nuclei could be classified as early, mid or late S-phase by CldU-immunodetection, giving a second control of the accuracy of cell separation (S3 Fig, S2 Table). Coherently with the post-sorting FACS analyses (S2 Fig), in the 4 fractions the large majority of the cells were positive to CldU labeling and belonged to the expected substage of the S-phase (early-mid in S1-S2 fractions, mid-late in S3-S4 fractions).

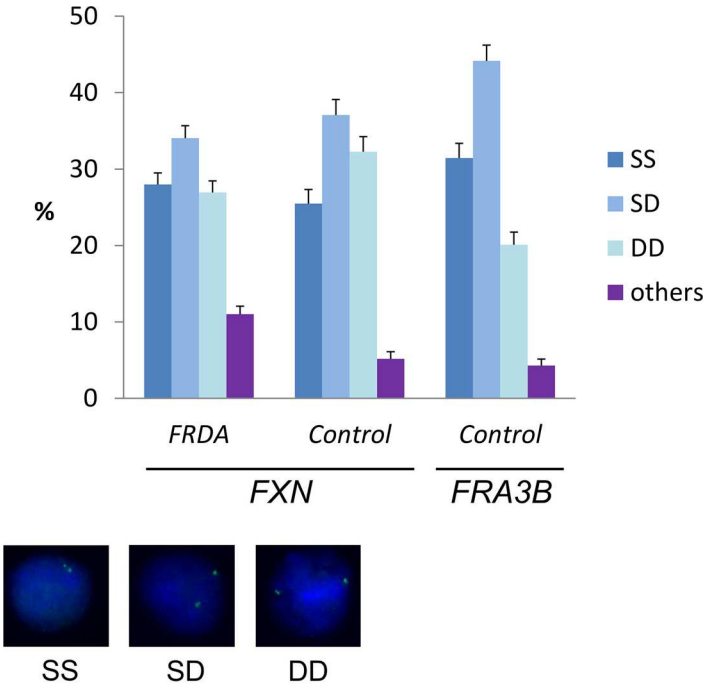


Fig 1. Replication timing of mutant and normal *FXN* alleles. Percentages are calculated from pooled data obtained with *FRDA* (GM15850 and GM16227) and control GM15851 cells. Per each group, at least 550 nuclei were analyzed from at least two independent replicated experiments (Raw data in S1 Table). SS = nuclei with two single FISH spots (non-replicated alleles); SD = nuclei with one single and one duplicated FISH signal (one allele has been replicated); DD = nuclei with two duplicated FISH signals (both alleles have been replicated); others = nuclei with one or none FISH signals. Error bars indicate standard errors of proportions. The probe used in these experiments is BAC RP11-265B8. For comparison, the replication timing of a late replication sequence (*FRA3B*, probe RP11-468L11) in normal GM15851 cells is shown. Examples of FISH replication patterns are shown in the bottom of the Figure.

doi:10.1371/journal.pgen.1006201.g001

Interphase FISH data concerning the normal cell line GM15851 were obtained in two independent sorting experiments and were remarkably reproducible (see S2 Table for raw data). Table 1 gives the summary of these analyses, further supporting the indication of a mid-late replication timing for the wildtype *FXN* locus: indeed, according to the percentage of DD nuclei observed in S2, within the first half of the S-phase only 25% of GM15851 cells have completed the replication of this sequence (Table 1). The replication patterns observed in the S1 and S2 fractions isolated from normal and mutant cells were significantly different when compared by chi-square analysis ($P < 0.05$ for early S-phase cells; $P < 0.001$ for mid S-phase cells). The observed trends indicate a faster replication progression of the wildtype than the mutant *FXN* allele, demonstrated in particular by the excess of SS nuclei persisting in mutant S2 cells (Table 1, S2 Table). In the S3 fraction a significant variation was recorded between replication patterns of late S-phase mutant and normal cells ($P < 0.05$); in the S4 fraction a statistical difference was detected when comparing the replication patterns of mid S-phase mutant with that of the normal cells ($P < 0.001$). These observations can be interpreted as a downstream effect of the earlier shift of the replication timing, as the proportions of SS nuclei were not involved in these variations (Table 1). In fact, in the second half of the S-phase both normal and expanded *FXN* alleles are undergoing and completing their replication.

When the late replicating sequence at *FRA3B* was considered (S3 Table), less than 15% of the S1-S2 cells carried a pair of replicated *FRA3B* alleles (DD patterns), confirming that this locus is replicating later than a wildtype *FXN* allele. Moreover, no differences were detected

between control (GM15851) and mutant (GM15850) cells in all sorted S-phase fractions. Therefore, replication patterns of expanded *FXN* alleles were strictly comparable to that of the late replicating *FRA3B* (Table 1 and S3 Table).

Table 1. Replication timing of normal and mutant *FXN* alleles according to interphase FISH after FACS cell sorting.

Cell sample	Cell fraction	Replication patterns#	Total cells	S-phase cells# N	early S-phase# N (% ± SE)§	mid S-phase# N (% ± SE)§	late S-phase# N (% ± SE)§
Control (GM15851 cells)	S1	SS	244	201	193 (61.3 ± 2.75)	8 (4.3 ± 1.50)	0
		DD	44	32	0	24 (13.0 ± 2.48)	8
		SD	278	249	111 (35.2 ± 2.69)	138 (75.0 ± 3.19)	0
		Others	31	25	11 (3.5 ± 1.03)	14 (7.6 ± 1.96)	0
		Total	597	507	315	184	8
	S2	SS	141	115	105 (47.3 ± 3.35)	10 (3.9 ± 1.20)	0
		DD	108	90	2 (0.9 ± 0.63)	67 (25.9 ± 2.72)	21 (80.8 ± 7.73)
		SD	300	279	108 (48.6 ± 3.36)	166 (64.1 ± 2.98)	5 (19.2 ± 7.73)
		Others	26	23	7 (3.2 ± 1.17)	16 (6.2 ± 1.50)	0
		Total	575	507	222	259	26
	S3	SS	48	30	20 (24.7 ± 4.79)	10 (2.7 ± 0.84)	0
		DD	209	179	1 (1.2 ± 1.23)	142 (38.3 ± 2.52)	36 (90.0 ± 4.74)
		SD	281	264	55 (67.9 ± 5.19)	205 (55.3 ± 2.58)	4 (10.0 ± 4.74)
		Others	24	19	5 (6.2 ± 2.67)	14 (3.8 ± 0.99)	0
		Total	562	492	81	371	40
	S4	SS	22	11	4 (23.5 ± 10.29)	7 (2.1 ± 0.78)	0
		DD	346	263	1 (5.9 ± 5.71)	182 (54.7 ± 2.73)	80 (93.0 ± 2.75)
		SD	183	149	11 (64.7 ± 11.59)	133 (39.9 ± 2.68)	5 (5.8 ± 2.52)
		Others	18	13	1 (5.9 ± 5.71)	11 (3.3 ± 0.98)	1 (1.2 ± 1.16)
		Total	569	436	17	333	86
FRDA (GM15850, GM16227 cells)	S1	SS	311	270	251 (62.1 ± 2.41)	19 (13.9 ± 2.95)	0
		DD	21	13	0	10 (7.3 ± 2.22)	3
		SD	212	201	119 (29.5 ± 2.27)	80 (58.4 ± 4.21)	2
		Others	77	62	34 (8.4 ± 1.38)	28 (20.4 ± 3.45)	0
		Total	621	546	404	137	5
	S2	SS	214	193	178 (55.5 ± 2.77)	15 (8.7 ± 2.14)	0
		DD	49	34	2 (0.6 ± 0.44)	26 (15.0 ± 2.72)	6
		SD	238	221	102 (31.8 ± 2.60)	116 (67.1 ± 3.57)	3
		Others	67	56	39 (12.1 ± 1.82)	16 (9.2 ± 2.20)	1
		Total	568	504	321	173	10
	S3	SS	33	28	19 (67.9 ± 8.83)	9 (2.8 ± 0.92)	0
		DD	239	204	1 (3.6 ± 3.51)	100 (31.1 ± 2.58)	103 (69.1 ± 3.79)
		SD	257	231	5 (17.9 ± 7.24)	192 (59.6 ± 2.73)	34 (22.8 ± 3.44)
		Others	49	36	3 (10.7 ± 5.85)	21 (6.5 ± 1.38)	12 (8.1 ± 2.23)
		Total	578	499	28	322	149
	S4	SS	9	4	3	1 (0.7 ± 0.68)	0
		DD	436	309	0	65 (44.2 ± 4.10)	244 (83.3 ± 2.18)
		SD	144	102	1	70 (47.6 ± 4.12)	31 (10.6 ± 1.80)
		Others	42	29	0	11 (7.5 ± 2.17)	18 (6.1 ± 1.40)
		Total	631	444	4	147	293

Replication patterns are based on features of the FISH signals of BAC RP11-265B8; S-phase cells are classified according to CldU-labeling. All details in Materials and Methods.

§ Percentages and SE of percentages were calculated only if > 10 total cells were observed per each S-phase substage.

doi:10.1371/journal.pgen.1006201.t001

To confirm the biological significance of our observations, we evaluated the replication pattern of a genomic region located about 170 kb downstream the *FXN* locus, and identified by BAC RP11-548B3 (S4C Fig). By analyzing S2 and S3 fractions of the cell samples used before, the same replication pattern was found for this region in normal and mutant cells (S4 Table). This led us to conclude that the shift of replication timing occurring in the presence of an expanded repeat did not involve a wide genomic region.

A single-molecule view of the replication profile of Frataxin

The replication profiles of normal and mutated *Frataxin* alleles were evaluated in the endogenous genomic context, by monitoring origin firing and replication fork dynamics within a 850 kb region centred on the *FXN* gene (S4 Fig). According to the estimated fractions of replicating molecules, having values higher than 50%, mutant and normal cell lines displayed comparable replication activity within the region (Table 2). At least 100 replication forks were scored and classified per each cell sample (Table 2). Fork rates and Inter-Origin Distances were comparable in normal and mutant cell line, as confirmed by the Kruskal-Wallis non-parametric test, which returned not significant results (Table 2; Fig 2).

It is accepted that activation of mammalian replication origins does not occur at steady genomic positions [5,6]. In agreement, within the investigated region origin firing occurred with wide molecule-to-molecule variability. The results obtained per single molecule and per each cell line are shown in S5–S8 Figs. A synoptical view showing the position of all the bidirectional origins mapped for the four cell lines investigated is shown in Fig 3. Looking at the whole genomic region delimited by the three BAC probes, changes in origin choice and a

Table 2. Replication profile of the 850 kb genomic region harboring *Frataxin*.

		GM15851	H691	GM15850	GM16227
		Control	Control	FRDA	FRDA
Replicating molecules	Total molecules	61	124	124	63
	Molecules with replication tracks	39	80	70	42
	Fraction of replicating molecules (%)	63.9	64.5	56.5	66.7
Replication forks	Total	122	203	151	101
	Unidirectional forks (N)	23	53	49	26
	Unidirectional forks (%)	18.9	26.1	32.5	25.7
	Paused/arrested bidirectional forks (N)[ç]	26	29	21	18
	Paused/arrested bidirectional forks (%)	21.3	14.3	13.9	18.0
	Asynchronous bidirectional forks (N)	8	6	11	1
	Asynchronous bidirectional forks (%)	6.6	3.0	7.3	1.0
Fork rate (kb/min)[§]	Mean ± SE	1.30 ± 0.166	1.63± 0.090	1.40 ± 0.123	1.36 ± 0.114
	Median	1.23	1.60	1.30	1.29
	N	24	56	29	18
	Min	0.18	0.28	0.35	0.42
	Max	3.56	3.25	2.96	2.60
Inter-Origin Distance (kb)	Mean ± SE	156.9 ± 25.63	181.7 ± 18.92	176.9 ± 31.38	126.4 ± 13.22
	Median	90.8	137.6	113.2	115.0
	N	54	78	45	40
	Min	16.6	11.1	19.4	19.0
	Max	1005.7	904. 5	1010.3	403.9

[ç] Replication forks with bilateral pause/arrest are calculated as a single event, because correlated to the firing of one origin.

[§] Only bidirectional forks (asynchronous and paused/arrested forks not included).

doi:10.1371/journal.pgen.1006201.t002

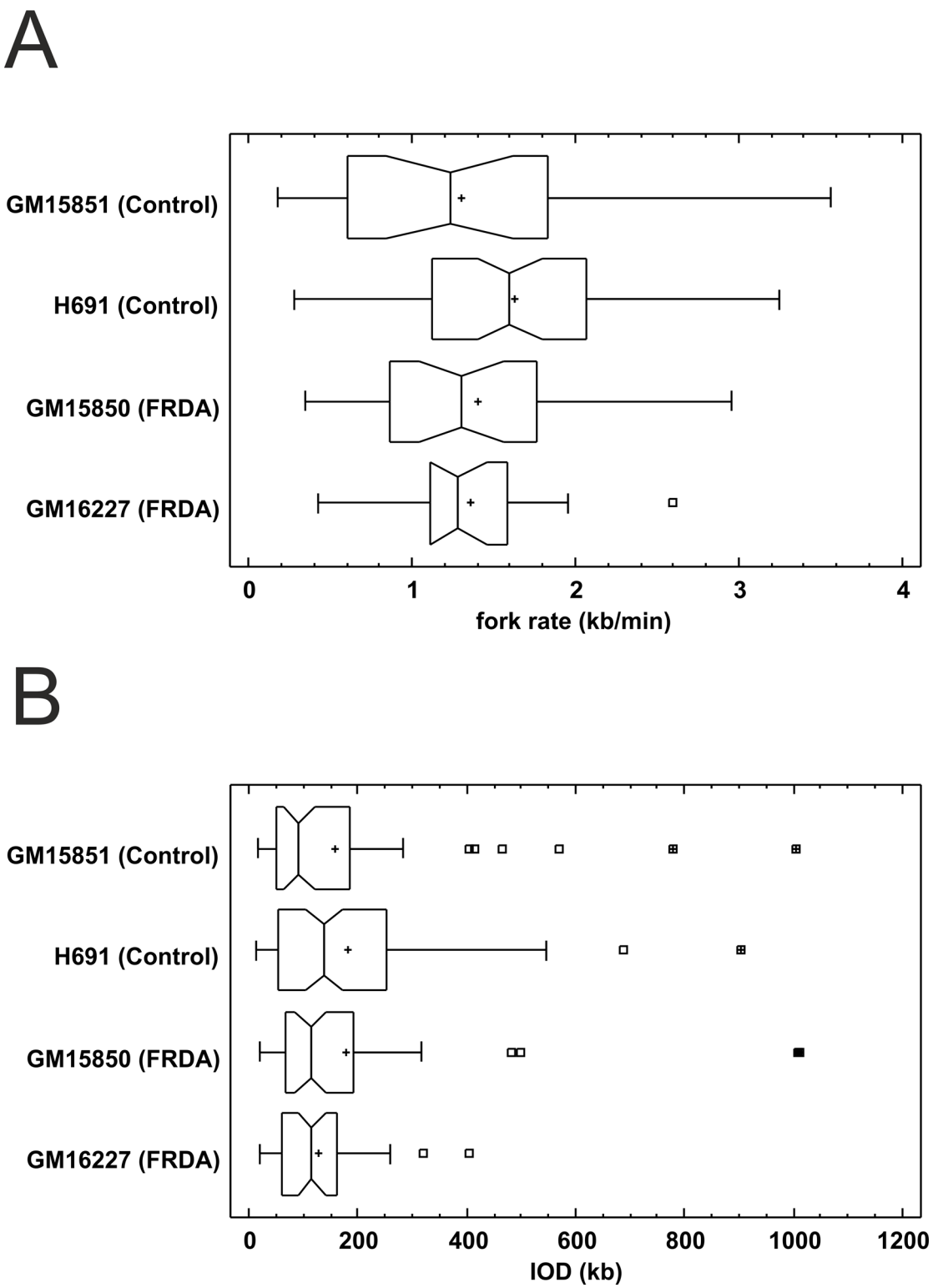


Fig 2. Distributions of fork rates and Inter-Origin Distances (IOD) in a 850 kb region centred around *FXN*. (A) Fork rate distributions observed in two normal (GM15851 and H691) and two FRDA (GM15850 and GM16227) cell lines. (B) IOD distributions observed in the same cell lines as above. No significant differences were detected for both variables described (Kruskal-Wallis non-parametric test).

doi:10.1371/journal.pgen.1006201.g002

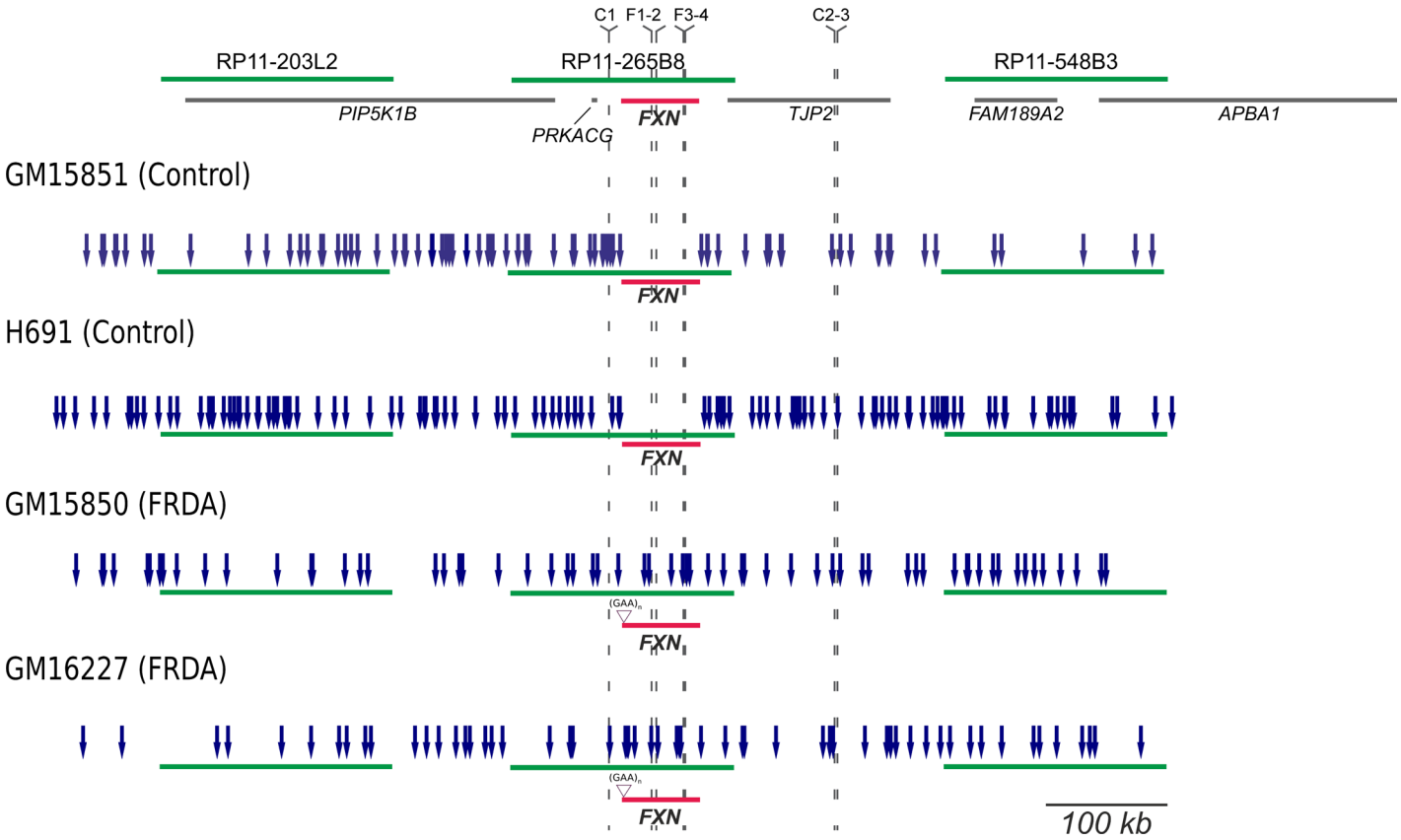


Fig 3. Synoptical view of the initiation events observed in a 850 kb region surrounding *FXN*. Top: the region investigated in the study. Green lines represent the three BAC clones used as probes, gray lines display genes, *FXN* is highlighted in red. The positions of primer sets used in the Short Nascent Strand (SNS) abundance assay are indicated in the map. In the four following schemes, blue arrows indicate the position of each bidirectional origin identified within normal (GM15851, H691) and mutant (GM15850, GM16227) alleles, relatively to the three probes (green) and the *FXN* gene (red). The position of the GAA-repeat expansion in the mutated alleles is also displayed.

doi:10.1371/journal.pgen.1006201.g003

differential distribution of activated bidirectional origins were detected in mutated versus normal cells (Fig 3). Dealing with each probe separately, we could appreciate that in the mutated alleles origin choices changed both upstream and downstream *FXN* (Fig 3). However, the most intriguing differences among samples emerged when focusing on the region we are more interested in, the central BAC RP11-265B8 harboring the *FXN* gene. In the normal cell line GM15851, 18 replicating molecules had at least one bidirectional origin firing within that region, 20 origins were mapped in total, but none of them fired inside the *FXN* gene (Fig 3, S5 Fig). The same pattern was found in our second control, the H691 cells: inside the region identified by BAC RP11-265B8 we detected 27 molecules with replication tracks, 16 of them carrying at least one bidirectional origin. No origin fired within the *FXN* gene, although 20 bidirectional origins were detected outside its sequence (Fig 3, S6 Fig). By considering the orientation of the replication forks running through the short GAA-repeat it appeared that this sequence was prevalently, but not exclusively, the template for the lagging strand. Remarkably, in both mutant cell lines several molecules showed one origin firing within the *FXN* allele with the expanded GAA-repeat (Figs 3 and 4; S7–S9 Figs). In particular, in GM15850 cells we found 14 molecules showing at least one bidirectional origin in the region identified by BAC RP11-265B8, for a total of 19 origins mapped within this genomic sequence (Fig 3, S7 Fig). Seven of these origins, each of them firing in an independent molecule, were located within the *FXN*

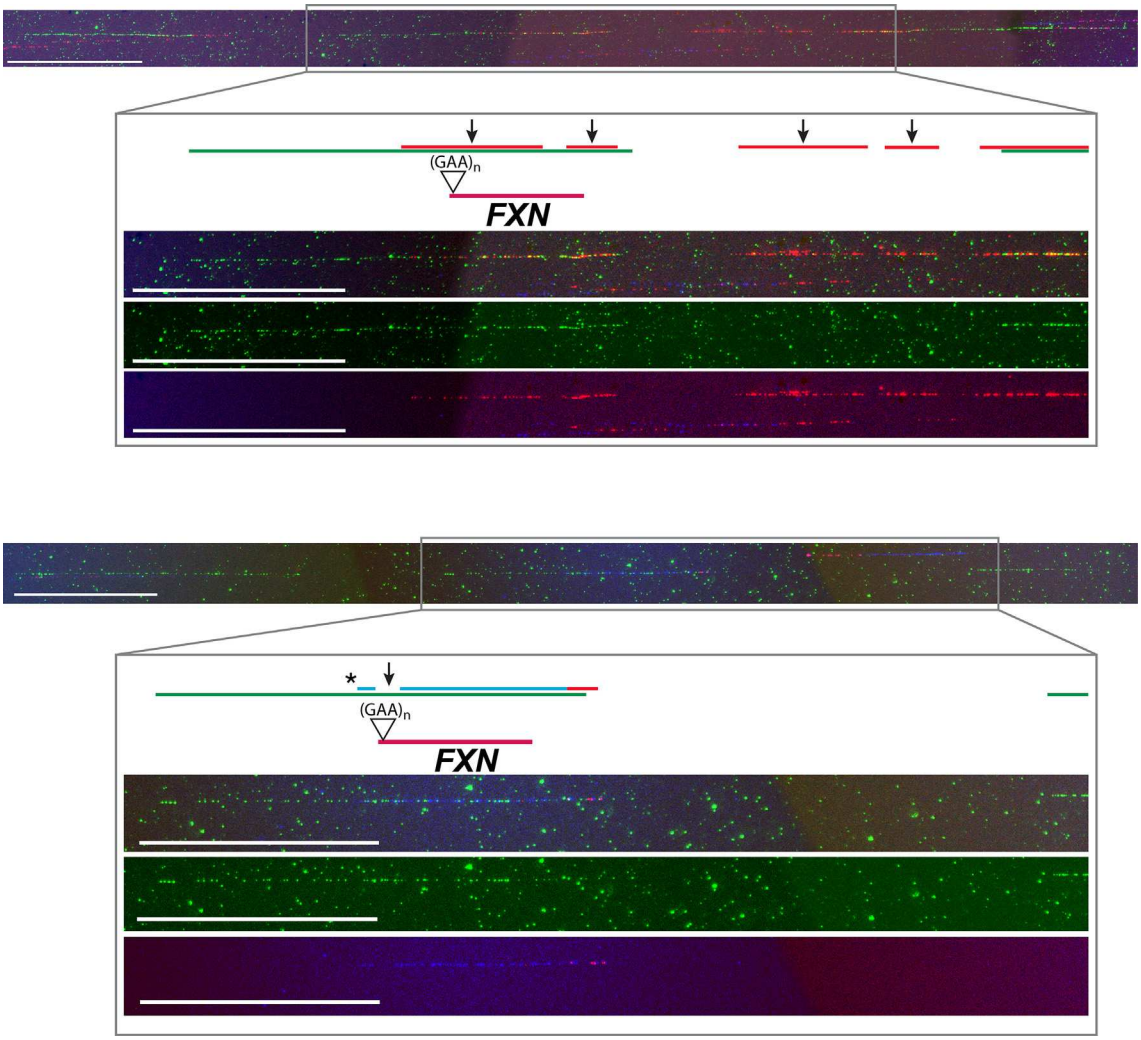


Fig 4. Two representative images of the *FXN* locus as detected by FISH and molecular combing. Three probes (green) are used to detect the 850 kb region harboring *FXN* (red). The position of the GAA-repeat expansion in the mutated alleles is also displayed. Replication tracks are visualized in blue (IdU) and red (CldU), arrows indicate origin positions, the asterisk corresponds to a paused/arrested fork. The two images refer to S8 Fig, molecules 41 and 5 respectively. The region identified by BAC RP11-265B8, where *FXN* maps, is enlarged to allow a better visualization of origin firing and a sharp interpretation of the replication signals. For each enlargement, the first frame corresponds to the merged image, the second frame shows the probes in green fluorescence, the third one displays the blue and red tracks coinciding with the replicative patterns. Calibration bar = 100 kb.

doi:10.1371/journal.pgen.1006201.g004

gene in different positions (Fig 3, S7 Fig). A similar profile was seen in GM16227 cells, where 17 bidirectional origins (from 11 molecules) were mapped within the central BAC RP11-265B8, and 9 of them (from 9 diverse molecules) were located in different positions of the *FXN* gene (Fig 3, S8 Fig). In consequence of dormant origin activation within the *FXN* gene sequence, the proportion of forks replicating the GAA-repeat from a downstream origin, and therefore from the leading strand, becomes higher than in the wildtype allele (S5–S8 Figs).

Replication forks with unidirectional progression were observed in proportions ranging from 19 to 32.5% in the different cell lines (Table 2). In all cell lines, they appeared evenly distributed along the genomic region investigated but a prevalence of short unidirectional forks was noted in FRDA cells compared to the average length detected in normal ones; this

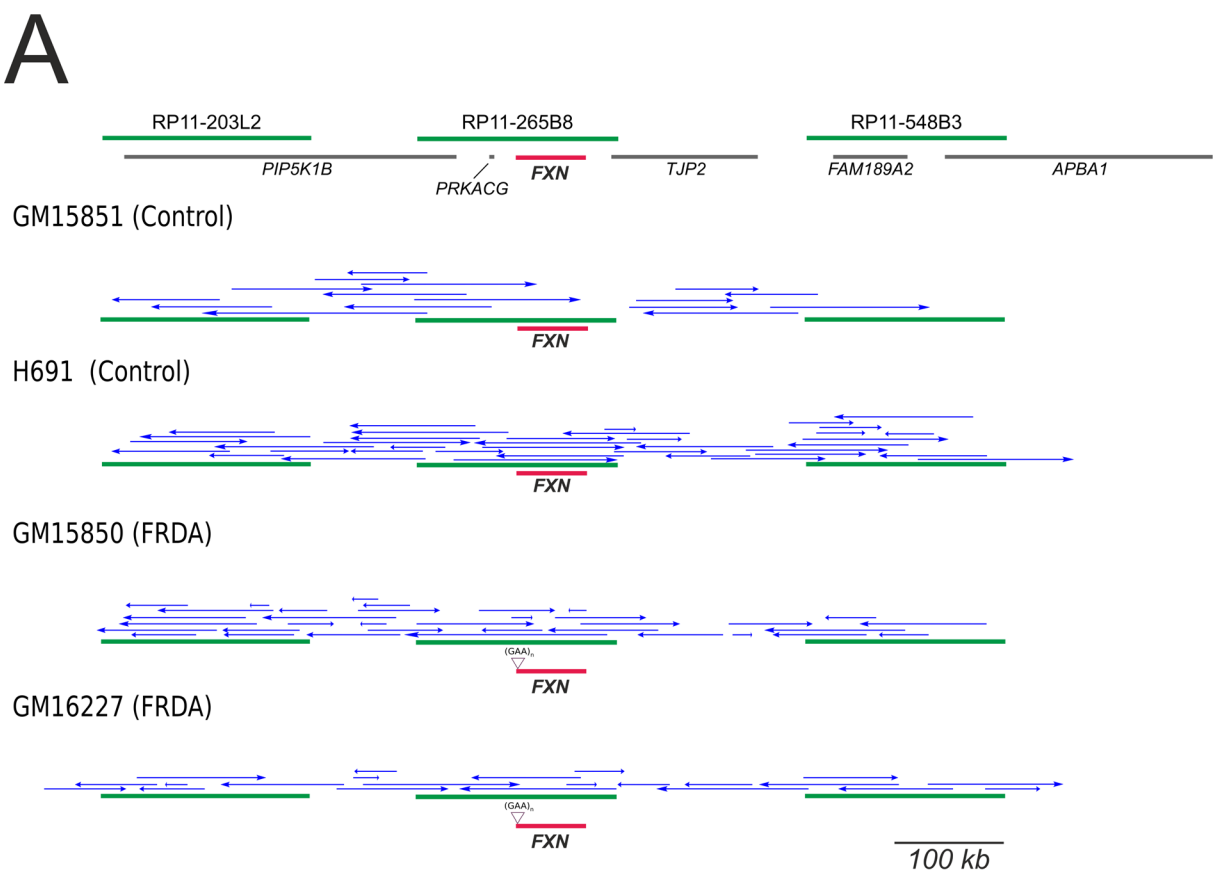
difference is particularly evident in the central segment that is harboring the *FXN* gene. Since the actual origin position cannot be defined when forks run unidirectionally, it is not correct to calculate their speed. Hence, we measured the length of unidirectional tracks entirely running in the region including the central BAC and the flanking probe-to-probe distances (Fig 5A). Statistically significant difference was detected among the four distributions, indicating the presence of a marked length reduction in FRDA cells when compared to the normal cell lines GM15851 and H691 (Fig 5B, $P < 0.01$, Kruskal-Wallis non-parametric test). Average lengths with standard errors were respectively: 113.7 ± 12.41 kb in normal GM15851 cells, 102.8 ± 8.56 kb in normal H691 cells, 67.5 ± 9.81 kb in FRDA GM15850 cells, 70.5 ± 11.83 kb in FRDA GM16227 cells (Fig 5B). In addition, according to the coefficient of variation (CV), unidirectional fork length measures are less dispersed in control cells (CV about 35%) than in FRDA ones (CV about 60%). Length distributions of unidirectional forks were significantly different in FRDA and control cells also when the whole panel of unidirectional forks was evaluated by Kruskal-Wallis test ($P < 0.005$). In this case the average lengths with standard errors were respectively: 115.9 ± 9.02 kb in normal GM15851 cells, 103.6 ± 5.67 kb in normal H691 cells, 76.9 ± 5.43 kb in FRDA GM15850 cells, 84.3 ± 7.30 kb in FRDA GM16227 cells. The magnitude of the CVs associated with the distribution of unidirectional forks in the whole 850 kb region remain higher in FRDA cells (although CVs decrease to values of about 45%) than in the controls (about 35%).

Finally, several events of pause/arrest of the fork were observed within the *FXN* locus and in the adjacent sequences. Frequent events of pause/arrest of the fork were detected in proximity of the short repeat in the GM15851 cells, while less intense occurrence of pause/arrest of the fork was found in H691 cells as well as in both mutant cell lines at the position of the long GAA-repeat (Fig 6).

Together, these data indicate that a passive modality of replication is favored within the normal *FXN* sequence, in which the short GAA repeat is the prevailing template for the lagging strand synthesis. In the presence of an expanded repeat, several changes of the replication profile, including recruitment of additional origins within the gene, widespread changes in origin choice, a differential distribution of unidirectional forks, provide the basis for assuring the completeness of DNA replication. In consequence of the activation of dormant origins in the expanded alleles, a switch of the direction by which the replication forks proceed through the GAA-repeat is frequently observed with respect to the normal sequence.

Short Nascent Strand (SNS) abundance assay

The pattern of origin activation in FRDA versus control cells was investigated also by the Short Nascent Strand (SNS) abundance assay and quantitative real-time PCR. To carry out the assay under optimal conditions an origin-free region should be used to normalize the SNS amounts obtained per each primer set [38–40]. Based on our molecular combing data (Fig 3), initiation events are widespread within the 850 kb sequence harboring the *FXN* gene and an origin-free region shared by the four cell lines cannot be firmly identified. Hence, qRT-PCR quantities were normalized versus an origin-positive sequence, a validated alternative approach to analyze SNS abundance experiments [40,41]. Two positions with recurrent pattern of origin activation among the four cell lines, located upstream and downstream the *FXN* gene respectively, may be inferred by the molecular combing analysis and were chosen to design primer sets C1-C3 (S5 Table, Figs 3 and 7A). Concerning the *FXN* gene, by using four primer sets (F1-F4, S5 Table, Figs 3 and 7A) we analyzed two sites downstream the GAA repeat. The enrichment fold at the *LAMIN B2* origin [38,39], calculated as a quality control of each SNS isolation experiment, ranged 5–119 (two independent experiments for each cell line). These values were used



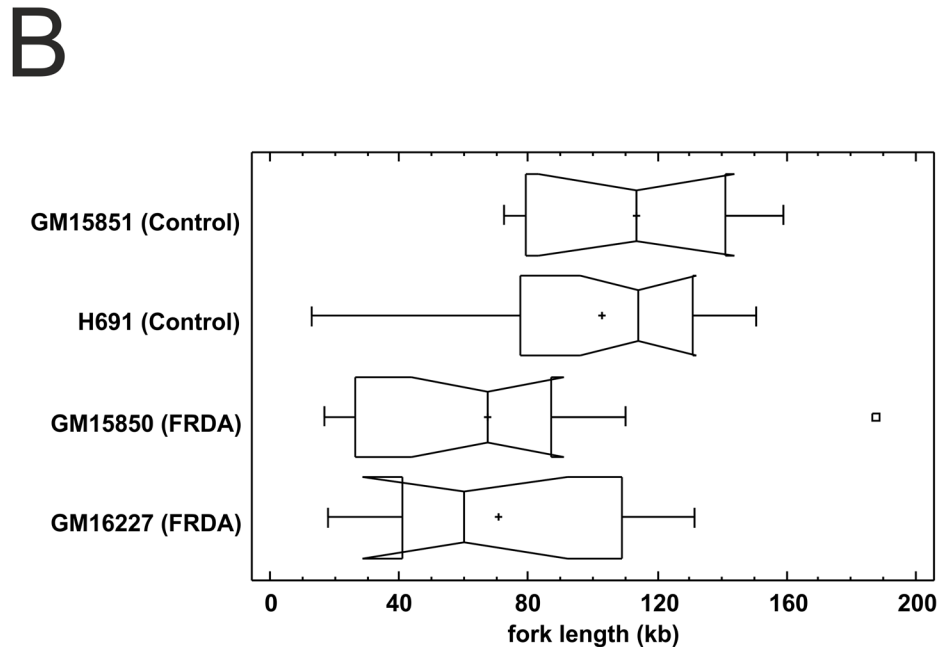


Fig 5. Position, direction and extension of the replication forks with unidirectional pattern observed in the genomic region under study. (A) The region investigated with green lines representing the three BAC clones used as probes, gray lines displaying genes, *FXN* (highlighted in red). In the following schemes, unidirectional fork progression within normal (GM15851, H691) and mutated (GM15850, GM16227) alleles is represented by blue arrows relatively to the three probes (green) and the *FXN* gene (red).

The position of the GAA-repeat expansion in the mutated alleles is also displayed. (B) Length distributions of the unidirectional running forks observed in normal and mutant cell lines, in the region including the central BAC and the flanking probe-to-probe distances. There is a significant length reduction in FRDA cells with respect to normal ones ($P < 0.01$, Kruskal-Wallis non-parametric test).

doi:10.1371/journal.pgen.1006201.g005

to set the threshold to estimate the SNS enrichment in FRDA and control cells as described in Material and Methods.

When the average quantity of SNS in the control regions C1-C3 on chromosome 9 (representing initiation zones) were used as the normalizing factor for the values estimated within the *FXN* gene (F1-F4), no differential patterns were detectable between FRDA and control cells (Fig 7B). Normalization of data versus a sequence with origins can produce a flattening effect and the background noise could be prevalent over true differences in origin activation, especially when dealing with low-efficient events as in this case [42]. To overcome this limitation, which could be responsible for the lack of differentiation visible in Fig 7B, we normalized SNS quantities to the non-origin site *LB2C1*. This value was further corrected by the mean enrichment of *LAMIN B2* origin, estimated for each cell line, as described in Materials and Methods. Although no specific trends emerged when considering each cell line separately (Fig 7C), by pooling data of FRDA or normal cell lines it appeared that SNS quantities detected in normal cells within the *FXN* gene (primers F1-F4) remained under the threshold, while the threshold

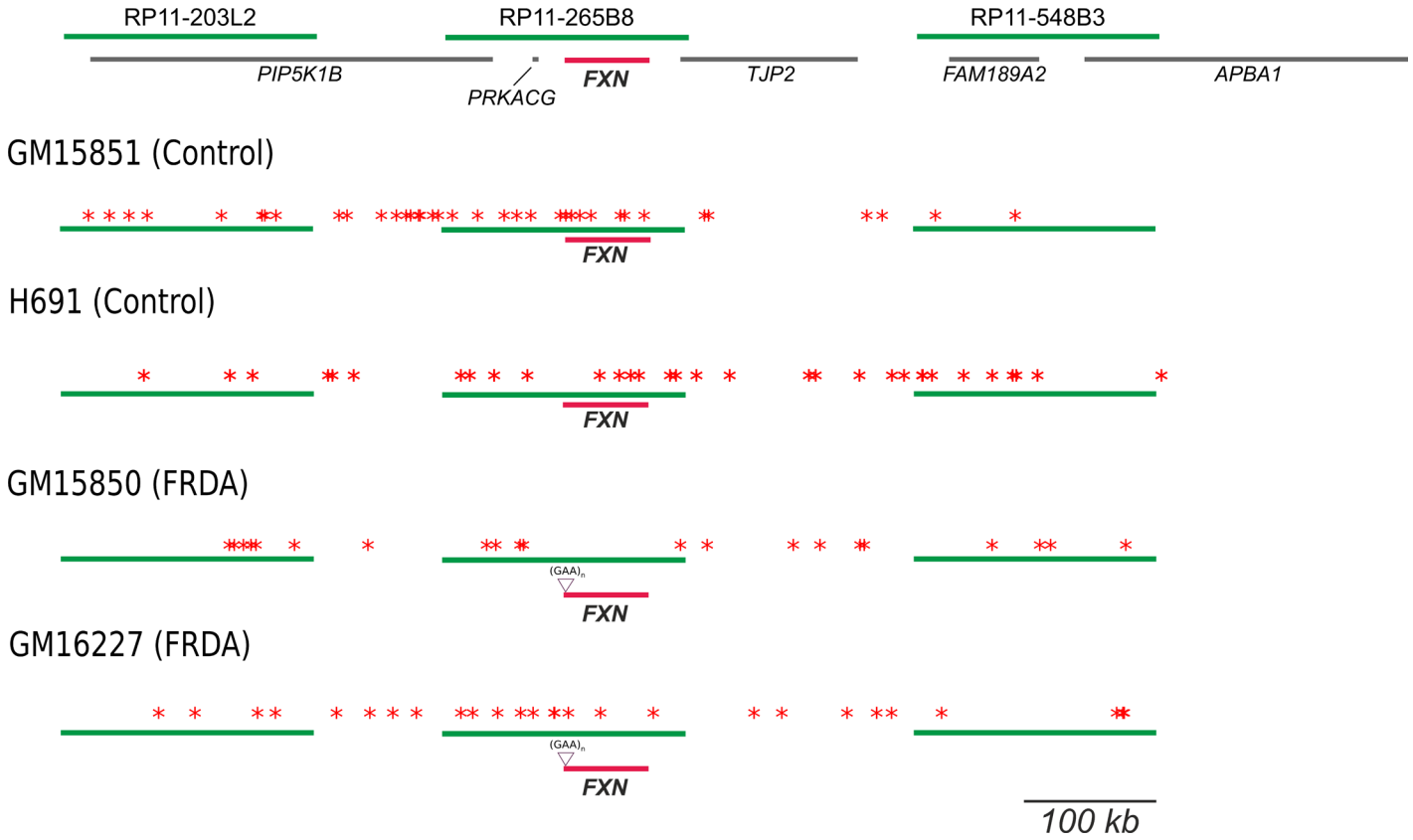


Fig 6. Synoptical view of the events of pause/arrest of the replication fork recorded in the genomic region under study. The region investigated is shown on the top: green lines represent the three BAC clones used as probes, gray lines display genes, *FXN* is highlighted in red. In the four following schemes, each red asterisk indicates a block in fork progression within normal (GM15851, H691) and mutated (GM15850, GM16227) alleles, relatively to the three probes (green) and the *FXN* gene (red). The position of the GAA-repeat expansion in the mutated alleles is also displayed.

doi:10.1371/journal.pgen.1006201.g006

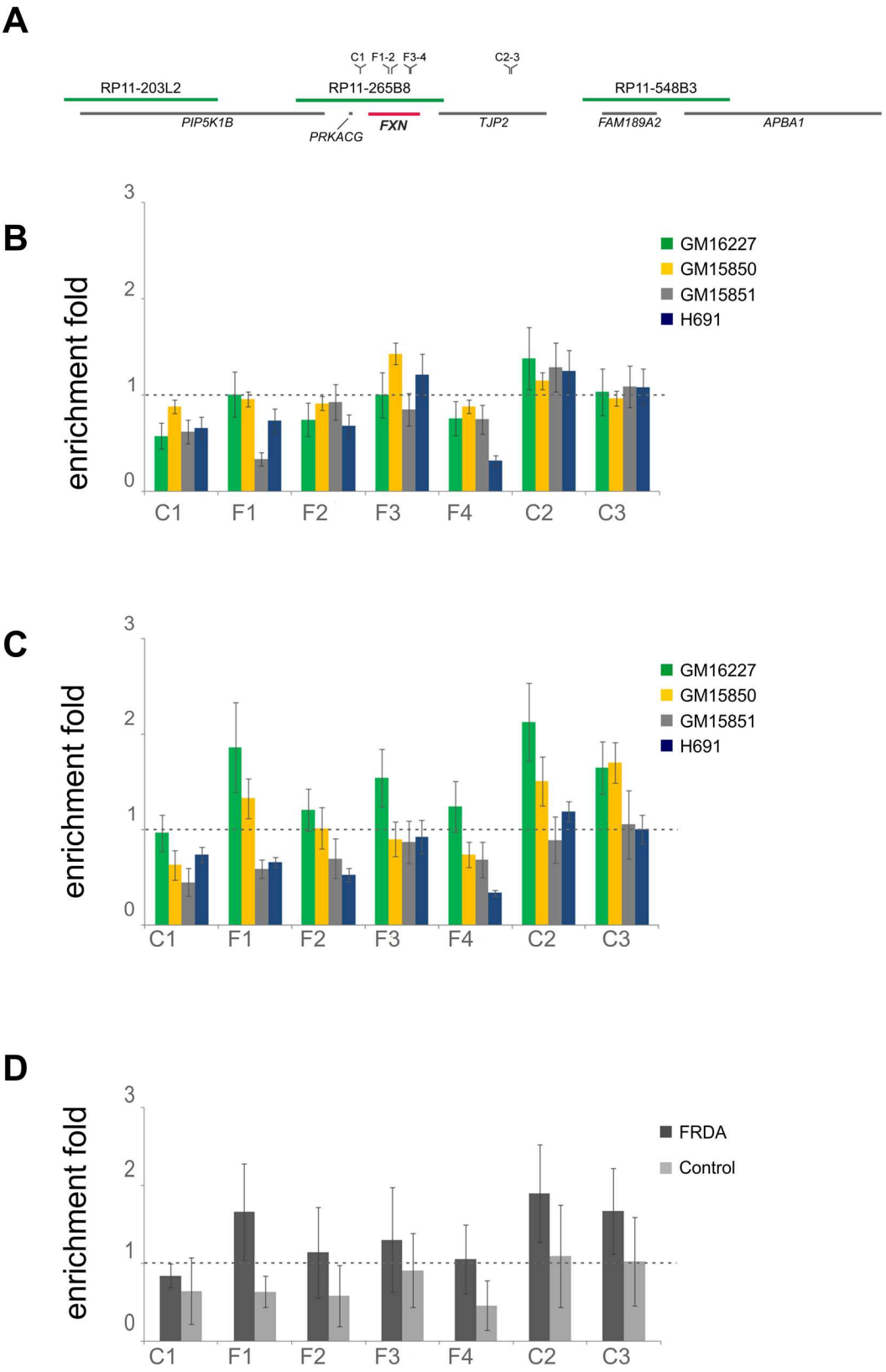


Fig 7. SNS abundance assay results. (A) The map position of primer sets used in the Short Nascent Strand (SNS) abundance assay. Green lines represent the three BAC clones used as probes in molecular combing

analyses; gray lines display genes, *FXN* is highlighted in red. (B) Abundance of *FXN* sequences in nascent DNA from two independent isolation experiments for each cell line. The mean quantities of SNS at *FXN* region (primers F1-F4) were normalized to the average quantities determined in the control sites on chr. 9 (primers C1-C3). (C) Abundance of SNS at each primer set on chr. 9 was normalized to *LB2C1*, and further corrected according to the enrichment at the *LAMIN B2* origin, which was set as threshold. (D) Abundance of SNS at each primer set on chr. 9 is shown as pooled data for FRDA (GM16227, GM15850) and control cells (GM15851 and H691); the analysis was performed as in C. Error bars are standard errors of the mean.

doi:10.1371/journal.pgen.1006201.g007

was reached at the control sites (C1-C3) in all cell lines and at *FXN* in FRDA cells (Fig 7D). This differential response is coherent with a lack of initiation events within the normal *FXN* alleles, but taking together the data shown in Fig 7 it must be concluded that the SNS abundance assay is not sensitive enough to confirm the activation of dormant origins during the replication of *FXN* expanded alleles.

Discussion

In this study we defined the replication program of the *FXN* gene in human cells, providing for the first time a wide view of origin firing and fork progression within an endogenous genomic context harboring an expanded GAA-repeat. In comparison to the normal *FXN* sequence, we found an altered replication timing of the mutated alleles. According to our results, the replication of expanded *FXN* alleles is slowed or delayed during the first half of the S-phase as compared with the wildtype sequence, while a normalization of this effect can be inferred in the second part of the S-phase.

By evaluating the replication profile of normal cells by molecular combing we found that *FXN* is passively replicated from incoming replication forks. Indeed, origins were never observed within the normal *FXN* sequence, both in this study and in our exploratory analysis of primary human lymphocytes from a healthy subject [14]. Changes in origin choice occurred even several kb upstream and downstream the expanded GAA-repeat, and the most relevant effect was the activation of origins downstream the GAA-expanded repeat, which can be considered dormant origins recruited to assure the replication of the mutated allele. By looking at the number of bidirectional origins per replicating molecule in the region identified by BAC RP11-265B8 (1.11 for GM15851, 1.25 for H691, 1.36 for FRDA GM15850, 1.55 for FRDA GM16227), the trend suggests an enhanced firing associated with the presence of the GAA-repeat expansion, indicating that the activation of the dormant origins does not substitute the initiations occurring in the normal alleles, while they are additional events. The activation of dormant origins at *FXN* has important implications to achieve replication of this gene, because the number of forks firing downstream the GAA-repeat is increased in mutant cells (S7 and S8 Figs) than in normal ones (S5 and S6 Figs). This implies that while in normal cells the GAA-repeat is prevalently located in the lagging strand template, in mutant cells the expansion is often in the leading strand template. This would be in agreement with the predictions of the origin-switch model for trinucleotide repeat instability [17–19], which was recently demonstrated to conform to the case of the CGG-expansion at the *FMR1* locus [20].

The activation of dormant origins is a rare and stochastic event occurring when cells are exposed to replication stress conditions, in order to respond to fork slowing and stalling [43–45]. Differently from the majority of published works in this field [1,46,47], in this study we observed the occurrence of a physiological event restricted to a narrow genomic region. Therefore, the activation of dormant origins at *FXN* is not comparable to an induced massive response as observed when cells are treated with DNA replication inhibitors. Indeed, looking at S7 and S8 Figs. the frequency of dormant origin firing, estimated on the total number of replicating molecules, ranged 14–21%. Moreover, when a dormant origin was found to be activated

within *FXN*, it was never associated with additional dormant origins and its location was not restricted to a steady position within the gene. Thus, firing of dormant origins is a peculiar feature of the *FXN* expanded allele, but the event occurs stochastically within the gene and among cells. Detecting such events as enrichment in origin activity through application of the SNS abundance assay is challenging. Moreover, it has been demonstrated that this approach is weakly effective when dealing with mid-late replicating loci [48–51] as in the case of the *FXN* locus, although it is considered accurate/stringent in the characterization of efficient origins, which are activated in early replicating regions. In view of the above considerations it is not surprising that only a weak trend was observed by applying the SNS abundance assay to the *FXN* locus (Fig 7D). However, these results are consistent with the molecular combing data supporting the activation of dormant origins within *FXN* in FRDA cells.

While this manuscript was under revision, data were published by applying a novel approach (OK-Seq), based on Okazaki fragment sequencing, which provides a global description of the replication landscape in a normal lymphoblastoid cell line (GM06990) and in HeLa cells [6]. OK-Seq data clearly indicate that *FXN* is associated with a termination region bordered by two initiation zones located upstream and downstream the gene (displayed as b and c in S10 Fig) in strict agreement with the replication profile obtained by molecular combing for the normal cell lines GM15851 and H691. More precisely, according to the model described in [6] the OK-Seq replication profile of the short termination zone associated with the *FXN* gene fits the scenario of forks emanating from the surrounding initiation zones and converging at different positions within the gene body. Additionally, Petryk et al. [6] identified a large termination region delimited by the initiation zones a and b (S10 Fig) which, according to their criteria, corresponds to a cascade of terminations associated with the firing of background origins. Again, this is coherent with our combing data (S5–S8 Figs). Thus, OK-Seq analysis strengthens the evidence of passive replication of the *FXN* gene in the absence of the expanded repeat, demonstrating that our observation can be extended to other cell lines and to diverse cell types. Despite the high resolution and precision of OK-Seq in identifying also broad and disperse initiation/termination zones [6], in the case of the expanded *FXN* allele it would be very hard to demonstrate the isolated, rare and widespread activation of dormant origins, because their firing may be not strong enough to generate a detectable initiation zone (Hyrien O. personal communication). On the whole, we can conclude that single molecule approaches, such as molecular combing, are the most appropriate tool to identify rare and stochastic firing events occurring as a change of the locus-specific replication program. In agreement with this opinion, for testing the reliability of genome wide approaches (e.g. ChIP-Seq) Dellino and Pellicci [42] recommend the application of single molecule techniques, because they provide relatively high-resolution origin maps in a large number of DNA molecules within the chromosomal region of interest [52–54].

In this study, a marked reduction of the average length of unidirectional running forks was observed in FRDA cells by molecular combing (Fig 5). Length values were also more broadly distributed in FRDA cells than in controls, in particular in the region delimited by the central BAC. We previously demonstrated that unidirectional forks are frequently detected in human cells by molecular combing and their frequencies ranged in the same interval observed here [14]. The biological meaning of the unidirectional forks is still unraveled, however in this case the length reduction observed in FRDA cells could be regarded as an additional effect of the replication impairment associated with the expanded GAA-repeat.

Using human cell lines from FRDA patients we were able to follow up the replication behavior of sequences carrying 630 to 1030 GAA-repeats, which are rather long arrays with respect to those evaluated in transfected/engineered cells. In these cell lines, we found that the functional loss of Frataxin does not have major effects on cell proliferation activity, and it is not

associated with global changes during the replication process (S1 Fig). In agreement with the published results [10–12,32], here we found that long stretches of GAA-repeats do not represent a strong impediment for the replication process. Indeed, although a shift of the replication timing was detected in mutated *FXN* alleles with respect to the normal ones, a rapid normalization of this effect occurs in the second half of the S-phase (Table 1). Furthermore, our molecular combing results suggest that in patients' cells the replication of the *FXN* gene is completed through the activation of origins that would not fire under normal conditions. This can be viewed as a rescue mechanism, as it is well known that in mammalian cells fork stalling related to a replication impediment may be solved by activating adjacent dormant origins [4,5].

Previously, different strategies were applied to evaluate the replication of GAA-repeats in vivo. By cloning stretches of different length in yeast plasmids [10], replication stalling was detected in long tracts ranging the size of premutated or mutated human alleles, while it was not observed in the presence of short (< 40) repeats. In consequence of fork stalling, inhibition of the fork progression, in the order of about 1.5 times, was reported [10]. More recently, a model was developed with a SV40-based plasmid transfected in human cells [11]. DNA replication could progress through expansions in the range of 33–90 GAA repeats, although several abnormal intermediates were found [11]. Transient pausing of the replication fork was detected at GAA-repeats longer than 66 trinucleotides, and in the case of longest tracts (> 90) fork reversal was found to be associated with fork pausing [11]. Moreover, both in yeast and transfected human cells it has been demonstrated that the most significant increase of fork pausing occurs when the GAA-repeat is located in the lagging strand template [11,55]. In the present study, we recorded several and widespread pause/arrest events in the 850 kb region harboring *FXN*. According to our previous results [14], in human cells these patterns often represent normal events of the replication program and DNA combing does not allow to distinguish a physiologically pausing fork from an event caused by a replication impairment. On the other hand, because their position can be mapped precisely and trinucleotide repeats could represent an obstacle for the progression of the replication forks, the preferential localization of pause/arrest events at the GAA-repeat can be checked. Recurrent events of pause/arrest of the fork were recorded in proximity of the short repeat in the GM15851 cells but not in the second control line and FRDA cells. To explain these differences the global response to the replication impairment associated with the presence of the GAA-repeat must be considered in its complexity. Here we demonstrated that in FRDA cells a major role is played by activation of dormant origins, and changes in unidirectional fork progression may be also involved. It is noteworthy that both activated dormant origins and unidirectional running forks are long-lasting replication patterns, while in most of the cases fork pausing are transient events; this feature has been reported also for the GAA-repeat by Follonier et al. [11]. In this frame, the activation of the dormant origins could be the most evident and easily detectable effect that can be ascribed to the presence of the expanded trinucleotide repeat. The reduction of the length of the unidirectionally running forks in FRDA cells is a second evident effect associated with the expanded GAA-repeat. In contrast, the chance to detect paused replication forks is affected by their transient nature. Moreover in FRDA cells the activation of dormant origins acts a safeguarding mechanism assuring the replication of the *FXN* gene. The consequence is that the number of forks running in the opposite direction through the GAA-repeat is increased (S7 and S8 Figs) and this sequence is replicated preferentially from the leading strand template, an orientation not frequently involved in fork pausing according to model systems [11,55]. Thus, in spite of the weak evidence available from this study, the detection of fork arrest events in proximity of the short GAA-repeat in GM15851 control cell line must be taken into further consideration as a possible impact of the short non-pathological GAA-repeat on fork

progression. Interestingly this result is in line with that obtained by a single molecule replication assay monitoring about 350 kb at the *FMR1* locus, where fork stalling at the CGG-repeat was found also in normal cells [20].

By molecular combing, we had the opportunity to monitor the *FXN* gene together with the surrounding genomic segment and a fine picture of the events associated with the replication of DNA tracts carrying trinucleotide repeats has been provided. In the present study, *FXN* replication profiles were evaluated on differentiated cells; understanding if the occurrence of origin-switch near the GAA-repeat may cause trinucleotide expansion in FRDA families, or it is instead a consequence of the expansion and/or the associated epigenetic phenomena, remain to be unraveled by further investigations.

Materials and Methods

Cell cultures and growth curves

Epstein Barr virus-transformed lymphoblastoid cell lines from two unrelated FRDA patients GM15850, (carrier of alleles with 650 and 1030 GAA-repeats respectively) and GM16227 (carrier of alleles with 630 and 830 GAA-repeats respectively), and from the healthy brother GM15851 of patient GM15850, were obtained by the Human Genetic Cell Repository of the Coriell Institute (USA). The H691 cell line is an Epstein Barr virus-transformed lymphoblastoid cell line established from a young healthy male adult. Cells were grown in RPMI 1640 medium supplemented with 15% foetal bovine serum (EuroClone, Italy) and penicillin/streptomycin antibiotics (Gibco, Life Technologies). The estimated duplication times range between 29–32 h for all cell lines.

Genotyping and RT-PCR

50 ng of genomic DNA and 10 pg of BAC clone RP11-265B8 were amplified in long-range PCR reactions (QIAGEN Long range PCR kit, Qiagen) using a primer set specific for the amplification of the GAA-repeat (forward: 5'-GGAGGGATCCGTCTGGGCAAAGG-3', reverse: 5'-CAATCCAGGACAGTCAGGGCTT-3'; normal amplicon length: 1,5 kb) [21]. For the expression analysis total RNA was isolated using the RNeasy mini kit (Qiagen). 1 μg of RNA was retrotranscribed (EuroScript M-MLW Reverse Transcriptase (RNase H-), Euroclone) and 2 μl of cDNA were used as template for PCR. The primer pairs used for *FXN* transcript were forward: 5'-CCTTGCAGACAAGCCATACA-3', reverse: 5'-GGTCCACTGGATG GAGAAGA-3' (amplicon length: 153 bp). *GAPDH* was chosen as a normalizing gene and its transcript amplified by forward: 5'-CCTCAACGACCACTTTGTCA-3', reverse: 5'-TTCCTC TTGTGCTCTTGCTG-3' (amplicon length: 143 bp).

Proliferation assays and FACS sorting

Cell cycle distribution of GM15851, GM15850 and GM16227 cell lines was monitored by flow cytometry (FACS Canto II; Becton Dickinson) after propidium iodide staining according to standard protocols, and analyzed with the Cell Quest software (Becton Dickinson).

Cell sorting was carried out in agreement with [37] using a BD FACSAria (BD Biosciences). Briefly, per each experiment 150 x 10^6 cells, pre-labeled with a 30 min pulse of 100 μM 5-Chloro-2'-deoxyuridine (CldU; Sigma-Aldrich), were harvested and prepared for FACS analysis. On the basis of the observed cell cycle distribution, intervals were set in order to collect four fractions of identical size spanning the entire S-phase. The purity of each S-phase fraction was assessed at the end of the experiment.

Interphase FISH analysis

Replication timing of normal and expanded *FXN* alleles was determined by interphase FISH in asynchronous and sorted cell line populations. Slides were prepared by using Cytospin 3 (Shandon Scientific Limited, UK) following a standard procedure. BAC clones were obtained by Children's Hospital Oakland Institute (CHORI, USA); BAC DNA was labeled with biotin-16-dUTP by nick translation kit (Roche Biochemicals).

Slides were post-fixed in 3:1 ethanol/acetic acid solution on ice, digested with 5 μg/ml Pepsin in 0.01 M HCl, pH 3.0 at 37°C for 10 min, and dehydrated in 70%, 90% and 100% ethanol. After denaturation at 72°C for 4 min in 70% formamide, 2X SSC pH 7.0 slides were dehydrated in ethanol solutions. The probe mix (100 ng of BAC DNA, 50X human Cot1 DNA (Invitrogen, LifeTechnologies) 5X Salmon Sperm DNA (Invitrogen, LifeTechnologies)) was denatured 10 min at 70°C and pre-annealed at 37°C for 90 min. Per slide, 10 μl were applied under a 22x22 mm coverslip; hybridisation was carried out overnight in a humidified chamber. Post-hybridisation washes were 50% formamide, 2X SSC pH 7.0 at 42°C (3 times), 2X SSC pH 7.0 (3 times). Slides were incubated 30 min at 37°C in 1X blocking solution (Roche Biochemicals) before immunodetection. CldU was detected by a rat monoclonal anti-BrdU antibody specifically cross-reacting with CldU (1:40, Abcam) and a 594 Alexa Fluor-conjugated anti-rat IgG (1:100; Molecular Probes, Life Technologies). At the same time the biotin-labeled probe was detected by 488 Alexa Fluor-conjugated streptavidin (1:300, Molecular Probes, Life Technologies), followed by polyclonal biotin-conjugated anti-streptavidin (1:300; Rockland) and 488 Alexa Fluor-conjugated streptavidin (1:300; Molecular Probes, Invitrogen). Slides were counterstained with DAPI solution (2 μg/ml) in Vectashield Mounting medium (Vector, USA).

Molecular combing

Exponentially growing lymphoblastoid cells were labeled with two sequential 30 min pulses of 5-Iodo-2'-deoxyuridine (IdU; Sigma-Aldrich) and 5-Chloro-2'-deoxyuridine (CldU; Sigma-Aldrich) (S4 Fig). 1-2x10^5 cells were immobilized in agarose plugs and incubated overnight at 50°C in 2 mg/ml Proteinase K solution (1% N-laurosylsarcosine, 0.1 M EDTA pH 8.0, 0.01 M Tris-HCl pH 8.0, 0.02 M NaCl). After digestion with β-agarase I (BioLabs), high molecular weight DNA from 1–2 plugs was delivered in 0.1 M MES pH 6.1. According to a standard procedure, DNA combing was performed on silanised surfaces. For further details see [56].

Single-locus replication analysis

A region spanning 850 kb, identified by the three differentially spaced BAC genomic clones RP11-203L2, RP11-265B8 and RP11-548B3 (Children's Hospital Oakland Institute; CHORI, USA), was used for the single-locus replication analysis (S4 Fig). Probes were biotin-labeled by random priming (BioPrime DNA labeling System, Invitrogen, Life Technologies); the central BAC RP11-265B8, which harbour *FXN*, was also labeled with a custom-made nucleotide mix containing Cy5-AP3-dUTP (GE Healthcare) to allow its identification and orientation also in molecules showing a probe pair instead than the whole probe set. Per slide, 250 ng of each probe were mixed in 20 μl of hybridization solution (50% formamide, 1% N-Laurosylsarcosine, 10 mM NaCl, 2X SSC in BlockAid (Invitrogen, Life Technologies) containing 13X human Cot-1 DNA (Invitrogen) and 10 μg Salmon Sperm DNA (Invitrogen). Denaturation was carried out at 80°C for 10 min. Slides were denatured for 15 min in 1 M NaCl, 0.05 M NaOH, immediately dehydrated in ethanol solutions (70%, 90%, 100%) on ice, hybridized with the probe mix (20 μl under 22x22 mm coverslips) for 19 hours at 37°C in a humidified chamber. Stringency washes were: 3x5 min in 50% formamide, 2X SSC pH 7.0 followed by 3x5 min in 2X SSC pH 7.0. A three-colour scheme of immunodetection was used to localize hybridisation signals together

with replication tracks: biotinylated probes were detected in green, whereas IdU and CldU in blue and red, respectively (S4 Fig). Three layers of antibodies were applied (30 min each): in the first one, 488 Alexa Fluor-conjugated streptavidin (1:50; Molecular Probes, Invitrogen) allowed probe detection and two primary anti-BrdU antibodies were used cross-reacting respectively with IdU (2:7; Becton-Dickinson, developed in mouse) and CldU (1:40; Abcam, developed in rat). In the second layer a polyclonal biotin-conjugated anti-streptavidin antibody (1:50; Rockland, USA), a 350 Alexa Fluor-conjugated anti-mouse IgG (1:50; Molecular Probes, produced in goat) and a 594 Alexa Fluor-conjugated anti-rat IgG (1:50; Molecular Probes, produced in donkey) were mixed. In the third layer, 488 Alexa Fluor-conjugated streptavidin mixed with 350 Alexa Fluor-conjugated anti-goat IgG (1:50; Molecular Probes, made in donkey) were used to complete the amplification steps. Cy5-labeled probes do not require amplification of the hybridization signal.

Genome-wide replication of lymphoblastoid cell lines was assessed by IdU and CldU immunodetection, according to the protocol described in [14].

Image analysis

A motorized fluorescence microscope (Zeiss Axio Imager.M1) equipped with a CCD camera (Photometrix, Coolsnap HQ2) was used for microscope analyses.

Interphase FISH analysis was carried out under a 100X oil immersion objective (N.A. = 1.30) and more than 250 nuclei were scored per each experiment. Replication timing was evaluated according to the observed hybridization signals. Single spots (S) can be referred to unreplicated alleles, whereas duplicated signals (D) to replicated alleles. Thus, nuclei were classified as SS when both alleles were not replicated, SD when only one allele completed replication and DD if both alleles were already replicated. In parallel and by blind analysis, CldU-positive nuclei were recorded and classified according to their fluorescent patterns in early, mid and late S-phase [57]. The hybridization efficiency was calculated by the formula: [SS+SD+DD+1/2 (S+D)/total number of scored nuclei] x100.

Molecular combing analyses were performed using a 40X oil immersion objective (N.A. = 1.30). DNA molecules may span several kilobases, therefore adjacent fields were acquired under adequate filter sets, then merged and aligned using Adobe Photoshop CS2 software. Fluorescent signals corresponding to replication tracks and hybridized probes were measured by the Metavue Research Imaging System (Molecular Devices), according to the molecular combing calibration factor (1 μm = 2 kb) and to the magnification features of objectives and CCD camera (1 pixel = 0.16125 μm = 0.3225 kb).

Probe length and probe-to-probe distances were determined in order to orientate molecules and to detect the replication activity within *FXN* genomic region. Moreover, only molecules showing the hybridization of at least two probes were considered informative and were used to calculate the fraction of replicating molecules, as the ratio between the number of molecules displaying replication signals and the total number of observed molecules.

Fork rates, inter-origin distances and replication origin positioning define replication profiles. In order to correctly interpret fluorescent signals, stringent criteria were applied according to those described in details elsewhere [14,58]. Briefly, as genomic DNA was not counterstained, only fluorescent replication signals in a linear array and framed by probe signals were considered, as they belong with maximum confidence to the same single molecule. Replication rates were calculated with complete bidirectional forks only and all other patterns, including forks with unidirectional progression and possible deregulation events, such as asynchronous and paused/arrested forks, were recorded only when upstream or downstream replication tracks supported the presence of uninterrupted filaments. Blue-only tracks emanating from

origins fired during the first pulse were interpreted according to the whole replication pattern along the molecule as termination events occurring during the first pulse OR paused/arrested forks. Isolated tracks, not allowing a non-ambiguous interpretation, were excluded from the analysis. More information in [14].

Short Nascent Strand (SNS) abundance assay and real-time PCR

Short Nascent Strand (SNS) abundance assay [40] was employed with some modifications. 70×10^6 cells derived from the four cell lines (GM15850, GM15851, GM16227, H691) were washed with 1X PBS and collected in 240 μl of 10% glycerol/PBS. 60 μl of each cell sample were lysed for 15 min a denaturating 1.25% agarose gel (50 mM NaOH, 1 mM EDTA), 4°C. The electrophoresis was carried out under the same denaturing conditions for 5–6 h at 30 V, neutralized in 1X TAE and stained with ethidium bromide. DNA fragments 0.5–1.5 kb in length were purified from the gel using a QIAEX II Gel extraction kit (Qiagen). Genomic DNA was purified by phenol-chloroform-isoamylalchol method from cells derived by the same cultures used for the isolation of SNS samples and digested with 0.4 mg/ml Proteinase K and 20 μg/ml RNase A. SNS and genomic DNA were quantified by NanoDrop 1000 (ThermoScientific).

Quantitative real-time PCR was carried out with 0.2 μM of each primer and the Power SYBR Green PCR Master Mix (Life Technologies) in an Applied Biosystems 7500 Real-Time PCR System (Life Technologies). The primer binding sites span the whole *FXN* gene and two regions located upstream and downstream the gene, chosen as controls on the chromosome 9. In addition, origin/non origin sites previously characterized around the *LAMIN B2* gene [38,39] were used as a further control site to test the reliability of the qPCR assay. All primers pairs used are listed in S5 Table and their positions on the *FXN* gene are shown in Figs 3 and 7A. For each primer set, a 6 point standard curve derived from 1:3 serial dilutions of the total DNA of each cell line was amplified in each plate, starting from 4.5 $x10^4$ copies of genomic equivalents. Standard curves were run in triplicate, SNS samples were run in five-eight replicas and the mean quantity of genomic equivalents in the short nascent DNA samples was determined in comparison to the correspondent standard curve. Amount of SNS at each primer set was normalized:

1. to the average quantity of SNS estimated in the control regions on the chromosome 9 (primer sets C1-C3) for each cell line.
2. To the non-origin site *LB2C1*, and further corrected according to the enrichment estimated at the *LAMIN B2* origin, used as a threshold to establish the enrichment fold for each primer set in the *FXN* and control regions (primer sets F1-F4, C1-C3).

Supporting Information

S1 Fig. Genotype, transcriptional, and proliferation activity of the cell lines used in this study. (A) The cell lines used in the study; information is provided by the Coriell repository apart from H691 cells; (B) the length of *FXN* alleles in the four cell lines, as evaluated by long-range PCR analysis; (C) transcriptional activity of *FXN* in the four cell lines, as evaluated by semiquantitative RT-PCR; (D) flow cytometry-based cell cycle profiles of the cell lines obtained from the Coriell repository; (E) genome-wide molecular combing analysis in the three cell lines obtained from the Coriell repository.
(JPG)

S2 Fig. Pre- and post-FACS sorting cell cycle distribution. FACS sorting experiments aimed to separate cells in consecutive temporal windows of the S-phase (S1-S4). Experiments 3 and 4

represent two independent biological replicates in which GM15851 cells were separated by FACS sorting.
(JPG)

S3 Fig. Quality control and reproducibility of the FACS sorting experiments. Before fractionation in four consecutive temporal windows of the S-phase (S5–S8, S2 Figs), cells were exposed to CldU for 30 min immediately before harvesting to confirm the accuracy of cell sorting procedure by CldU-immunodetection. Cells were classified in early, mid, late S-phase according to the observed fluorescent pattern (see F). (A-D) Quality control: data in the graphs were collected during the FISH analyses of *FXN* replication timing (respectively: experiments 1–4 summarized in S2 Fig; raw data in Supplementary Table 2). In all of the experiments each cell fraction is enriched for the expected S-phase stage (error bars = errors of percentages). (E) Reproducibility: average proportions of cells belonging to different stages of the S-phase, as observed in the course of three FISH experiments carried out independently on the same S2 and S3 cell samples (respectively: exp. 2 for GM15850 cells, exp. 3 for GM15851). (F) Examples (left to right) of early, mid, late S-phase nuclei according to the observed fluorescent CldU pattern.
(JPG)

S4 Fig. Experimental design and criteria of analysis in molecular combing experiments. (A) The two-pulse labeling scheme for detection of replication forks. In the first pulse IdU is incorporated in the nascent strands and labeled DNA is detected by blue fluorescence; during the second pulse CldU is available for the synthesis of DNA, and labeling is detected by red fluorescence. (B) Examples of normal and altered replication patterns expected in single-locus replication analyses; replication tracks and probes are represented slightly displaced for simplicity. Three probes differentially spaced (D1 and D2) are detected by green fluorescence: the FISH pattern allows us to define the centromere-telomere orientation and the integrity of the molecule (a Cy5-labeled central probe is cohybridized with the biotin-labeled probes, to allow the centromere-telomere orientation when only two hybridization signals can be visualized). Bidirectional origins (o) may be mapped in the middle of the two arms or in the middle of a blue track of a replication fork. Paused/arrested forks (*) may be unilateral or bilateral events, as illustrated. Asynchronous forks (as.) fire from the origin with different rates. Unidirectional forks (unidir.) are identified when a single arm with blue/red pattern is progressing with same orientation than the upstream or downstream track (which is the case represented in this example). (C) The genomic region investigated in this study. Green lines represent three probes covering about 850 kb in the bp interval 68643,187–69477,097 at 9q21. *Frataxin* (*FXN)* is shown in red and the other genes mapping in the flanking regions are shown in grey.
(JPG)

S5 Fig. Fork progression and origin distribution in single molecules spanning the region investigated in GM15851 (control) cells. Dotted lines indicate the boundaries of the *FXN* sequence. A scheme of the genomic region is shown in S4C Fig Original reconstructed images are deposited in the Dryad Digital Repository at http://dx.doi.org/10.5061/dryad.f12cg.
(JPG)

S6 Fig. Fork progression and origin distribution in single molecules spanning the region investigated in H691 (control) cells. Dotted lines indicate the boundaries of the *FXN* sequence. A scheme of the genomic region is shown in S4C Fig Original reconstructed images are deposited in the Dryad Digital Repository at http://dx.doi.org/10.5061/dryad.f12cg.
(JPG)

S7 Fig. Fork progression and origin distribution in single molecules spanning the region investigated in GM15850 (FRDA) cells. Dotted lines indicate the boundaries of the *FXN* sequence. A scheme of the genomic region is shown in S4C Fig Original reconstructed images are deposited in the Dryad Digital Repository at http://dx.doi.org/10.5061/dryad.f12cg.
(JPG)

S8 Fig. Fork progression and origin distribution in single molecules spanning the region investigated in GM16227 (FRDA) cells. Dotted lines indicate the boundaries of the *FXN* sequence. A scheme of the genomic region is shown in S4C Fig Original reconstructed images are deposited in the Dryad Digital Repository at http://dx.doi.org/10.5061/dryad.f12cg.
(JPG)

S9 Fig. Representative images of replication at *FXN* (RP11-265B8 probe) as detected by FISH and molecular combing in FRDA and control cells. Images are selected enlargements from complete reconstructed molecules (from top to bottom: molecules 33, 42, 15 in S5 Fig, molecule 13 in S6 Fig, molecules 36 and 7 in S8 Fig, molecule 42 in S7 Fig) The probe (green), containing the *FXN* gene (red), and the flanking regions (probe-to-probe distances, S4C Fig) are shown. Replication tracks are visualized in blue (IdU) and red (CldU), arrows indicate origin positions, the asterisk corresponds to a paused/arrested fork, unidirectional forks are indicated. GAA-repeat expansion is also displayed. Calibration bar = 100kb.
(JPG)

S10 Fig. Replication fork directionality profiles (RFD) of GM06990 and HeLa cells in the region harboring the *FXN* gene (GRCh37, hg19). (A) RFD profile of the GM06990 lymphoblastoid cell line. Green lines represent the three BAC clones used as probes in molecular combing experiments, gray lines display genes, *FXN* is highlighted in red. Pink bars (a, b, c) represent initiation zones derived by OK-Seq analysis [6]. Two replicas are shown. (B) RFD profile of HeLa cell line. Green lines represent the three BAC clones used as probes in molecular combing experiments, gray lines display genes, *FXN* is highlighted in red. Pink bars (a, b, c, d) represent initiation zones derived by OK-Seq analysis [6]. Two replicas are shown. RFD were downloaded from http://157.136.54.88/cgi-bin/gbrowse/gbrowse/okazaki_ref/.
(JPG)

S1 Table. Raw data of the replication timing analysis of *FXN* carried out by interphase FISH.
(DOCX)

S2 Table. Raw data of the replication timing analysis of *FXN* carried out by interphase FISH after FACS cell sorting.
(DOCX)

S3 Table. Replication timing of the late replicating sequence *FRA3B* according to interphase FISH after FACS cell sorting.
(DOCX)

S4 Table. Replication timing of a genomic domain located about 170 kb downstream the *FXN* allele, according to interphase FISH after FACS cell sorting.
(DOCX)

S5 Table. Oligonucleotide primer sequences.
(DOCX)

Acknowledgments

We acknowledge the skilled assistance of Anna Cabrelle (Venetian Institute of Molecular Medicine (VIMM), Padova,) during FACS sorting procedures. Chiara Romualdi (Dept. of Biology, University of Padova, Italy) gave us important advices in the analysis of qRT-PCR results. We are grateful to Olivier Hyrien (Institut de Biologie de l'Ecole Normale Supérieure, Paris, France) and Chunlong Chen (Institute Curie, Paris, France) for helpful discussion concerning OK-Seq data.

Author Contributions

Conceived and designed the experiments: EP AR. Performed the experiments: MS EP. Analyzed the data: MS EP AR. Wrote the paper: MS EP AR.

References

1. Anglana M, Apiou F, Bensimon A, Debatisse M. Dynamics of DNA replication in mammalian somatic cells: nucleotide pool modulates origin choice and interorigin spacing. Cell 2003 Aug 8; 114(3):385–394. PMID: 12914702
2. Poli J, Tsaponina O, Crabbe L, Keszthelyi A, Pantesco V, Chabes A, et al. dNTP pools determine fork progression and origin usage under replication stress. EMBO J 2012 Feb 15; 31(4):883–894. doi: 10.1038/emboj.2011.470 PMID: 22234185
3. Zeman MK, Cimprich KA. Causes and consequences of replication stress. Nat Cell Biol 2014 Jan; 16 (1):2–9. doi: 10.1038/ncb2897 PMID: 24366029
4. Yekezare M, Gomez-Gonzalez B, Diffley JF. Controlling DNA replication origins in response to DNA damage—inhibit globally, activate locally. J Cell Sci 2013 Mar 15; 126(Pt 6):1297–1306. doi: 10.1242/jcs.096701 PMID: 23645160
5. Mechali M. Eukaryotic DNA replication origins: many choices for appropriate answers. Nat Rev Mol Cell Biol 2010 Oct; 11(10):728–738. doi: 10.1038/nrm2976 PMID: 20861881
6. Petryk N, Kahli M, d'Aubenton-Carafa Y, Jaszczyszyn Y, Shen Y, Silvain M, et al. Replication landscape of the human genome. Nat Commun 2016 Jan 11; 7:10208. doi: 10.1038/ncomms10208 PMID: 26751768
7. McMurray CT. Mechanisms of trinucleotide repeat instability during human development. Nat Rev Genet 2010 Nov; 11(11):786–799. doi: 10.1038/nrg2828 PMID: 20953213
8. Mirkin EV, Mirkin SM. Replication fork stalling at natural impediments. Microbiol Mol Biol Rev 2007 Mar; 71(1):13–35. PMID: 17347517
9. Mirkin EV, Mirkin SM. To switch or not to switch: at the origin of repeat expansion disease. Mol Cell 2014 Jan 9; 53(1):1–3. doi: 10.1016/j.molcel.2013.12.021 PMID: 24411078
10. Krasilnikova MM, Mirkin SM. Replication stalling at Friedreich's ataxia (GAA)n repeats in vivo. Mol Cell Biol 2004 Mar; 24(6):2286–2295. PMID: 14993268
11. Follonier C, Oehler J, Herrador R, Lopes M. Friedreich's ataxia-associated GAA repeats induce replication-fork reversal and unusual molecular junctions. Nat Struct Mol Biol 2013 Apr; 20(4):486–494. doi: 10.1038/nsmb.2520 PMID: 23454978
12. Chandok GS, Patel MP, Mirkin SM, Krasilnikova MM. Effects of Friedreich's ataxia GAA repeats on DNA replication in mammalian cells. Nucleic Acids Res 2012 May; 40(9):3964–3974. doi: 10.1093/nar/gks021 PMID: 22262734
13. Liu G, Chen X, Gao Y, Lewis T, Barthelemy J, Leffak M. Altered replication in human cells promotes DMPK (CTG)(n). (CAG)(n) repeat instability. Mol Cell Biol 2012 May; 32(9):1618–1632. doi: 10.1128/MCB.06727-11 PMID: 22354993
14. Palumbo E, Tosoni E, Russo A. General and specific replication profiles are detected in normal human cells by genome-wide and single-locus molecular combing. Exp Cell Res 2013 Dec 10; 319(20):3081–3093. doi: 10.1016/j.yexcr.2013.10.001 PMID: 24126019
15. Lopez Castel A, Cleary JD, Pearson CE. Repeat instability as the basis for human diseases and as a potential target for therapy. Nat Rev Mol Cell Biol 2010 Mar; 11(3):165–170. doi: 10.1038/nrm2854 PMID: 20177394
16. Kim JC, Mirkin SM. The balancing act of DNA repeat expansions. Curr Opin Genet Dev 2013 Jun; 23 (3):280–288. doi: 10.1016/j.gde.2013.04.009 PMID: 23725800

17. Cleary JD, Nichol K, Wang YH, Pearson CE. Evidence of cis-acting factors in replication-mediated trinucleotide repeat instability in primate cells. Nat Genet 2002 May; 31(1):37–46. PMID: 11967533
18. Mirkin SM, Smirnova EV. Positioned to expand. Nat Genet 2002 May; 31(1):5–6. PMID: 11984556
19. Pearson CE, Nichol Edamura K, Cleary JD. Repeat instability: mechanisms of dynamic mutations. Nat Rev Genet 2005 Oct; 6(10):729–742. PMID: 16205713
20. Gerhardt J, Tomishima MJ, Zaninovic N, Colak D, Yan Z, Zhan Q, et al. The DNA replication program is altered at the FMR1 locus in fragile X embryonic stem cells. Mol Cell 2014 Jan 9; 53(1):19–31. doi: 10.1016/j.molcel.2013.10.029 PMID: 24289922
21. Campuzano V, Montermini L, Molto MD, Pianese L, Cossee M, Cavalcanti F, et al. Friedreich's ataxia: autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science 1996 Mar 8; 271(5254):1423–1427. PMID: 8596916
22. Bidichandani SI, Ashizawa T, Patel PI. The GAA triplet-repeat expansion in Friedreich ataxia interferes with transcription and may be associated with an unusual DNA structure. Am J Hum Genet 1998 Jan; 62(1):111–121. PMID: 9443873
23. Ohshima K, Montermini L, Wells RD, Pandolfo M. Inhibitory effects of expanded GAA.TTC triplet repeats from intron I of the Friedreich ataxia gene on transcription and replication in vivo. J Biol Chem 1998 Jun 5; 273(23):14588–14595. PMID: 9603975
24. Silva AM, Brown JM, Buckle VJ, Wade-Martins R, Lufino MM. Expanded GAA repeats impair FXN gene expression and reposition the FXN locus to the nuclear lamina in single cells. Hum Mol Genet 2015 Jun 15; 24(12):3457–3471. doi: 10.1093/hmg/ddv096 PMID: 25814655
25. Evans-Galea MV, Hannan AJ, Carrodus N, Delatycki MB, Saffery R. Epigenetic modifications in trinucleotide repeat diseases. Trends Mol Med 2013 Nov; 19(11):655–663. doi: 10.1016/j.molmed.2013.07.007 PMID: 23953480
26. Chan PK, Torres R, Yandim C, Law PP, Khadayate S, Mauri M, et al. Heterochromatinization induced by GAA-repeat hyperexpansion in Friedreich's ataxia can be reduced upon HDAC inhibition by vitamin B3. Hum Mol Genet 2013 Jul 1; 22(13):2662–2675. doi: 10.1093/hmg/ddt115 PMID: 23474817
27. Yandim C, Natisvili T, Festenstein R. Gene regulation and epigenetics in Friedreich's ataxia. J Neurochem 2013 Aug; 126 Suppl 1:21–42. doi: 10.1111/jnc.12254 PMID: 23859339
28. Li Y, Lu Y, Polak U, Lin K, Shen J, Farmer J, et al. Expanded GAA repeats impede transcription elongation through the FXN gene and induce transcriptional silencing that is restricted to the FXN locus. Hum Mol Genet 2015 Dec 15; 24(24):6932–6943. doi: 10.1093/hmg/ddv397 PMID: 26401053
29. Ditch S, Sammarco MC, Banerjee A, Grabczyk E. Progressive GAA.TTC repeat expansion in human cell lines. PLoS Genet 2009 Oct; 5(10):e1000704. doi: 10.1371/journal.pgen.1000704 PMID: 19876374
30. De Biase I, Rasmussen A, Monticelli A, Al-Mahdawi S, Pook M, Cocozza S, et al. Somatic instability of the expanded GAA triplet-repeat sequence in Friedreich ataxia progresses throughout life. Genomics 2007 Jul; 90(1):1–5. PMID: 17498922
31. Sharma R, Bhatti S, Gomez M, Clark RM, Murray C, Ashizawa T, et al. The GAA triplet-repeat sequence in Friedreich ataxia shows a high level of somatic instability in vivo, with a significant predilection for large contractions. Hum Mol Genet 2002 Sep 1; 11(18):2175–2187. PMID: 12189170
32. Kim HM, Narayanan V, Mieczkowski PA, Petes TD, Krasilnikova MM, Mirkin SM, et al. Chromosome fragility at GAA tracts in yeast depends on repeat orientation and requires mismatch repair. EMBO J 2008 Nov 5; 27(21):2896–2906. doi: 10.1038/emboj.2008.205 PMID: 18833189
33. Techer H, Koundrioukoff S, Azar D, Wilhelm T, Carignon S, Brison O, et al. Replication dynamics: biases and robustness of DNA fiber analysis. J Mol Biol 2013 Nov 29; 425(23):4845–4855. doi: 10.1016/j.jmb.2013.03.040 PMID: 23557832
34. ENCODE Project Consortium, Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 2007 Jun 14; 447(7146):799–816. PMID: 17571346
35. Karnani N, Taylor C, Malhotra A, Dutta A. Pan-S replication patterns and chromosomal domains defined by genome-tiling arrays of ENCODE genomic areas. Genome Res 2007 Jun; 17(6):865–876. PMID: 17568004
36. Dellino GI, Cittaro D, Piccioni R, Luzi L, Banfi S, Segalla S, et al. Genome-wide mapping of human DNA-replication origins: levels of transcription at ORC1 sites regulate origin selection and replication timing. Genome Res 2013 Jan; 23(1):1–11. doi: 10.1101/gr.142331.112 PMID: 23187890
37. Hansen RS, Thomas S, Sandstrom R, Canfield TK, Thurman RE, Weaver M, et al. Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. Proc Natl Acad Sci U S A 2010 Jan 5; 107(1):139–144. doi: 10.1073/pnas.0912402107 PMID: 19966280

38. Giacca M, Zentilin L, Norio P, Diviacco S, Dimitrova D, Contreas G, et al. Fine mapping of a replication origin of human DNA. Proc Natl Acad Sci U S A 1994 Jul 19; 91(15):7119–7123. PMID: 8041756

39. Abdurashidova G, Deganuto M, Klima R, Riva S, Biamonti G, Giacca M, et al. Start sites of bidirectional DNA synthesis at the human lamin B2 origin. Science 2000 Mar 17; 287(5460):2023–2026. PMID: 10720330

40. Gray SJ, Gerhardt J, Doerfler W, Small LE, Fanning E. An origin of DNA replication in the promoter region of the human fragile X mental retardation (FMR1) gene. Mol Cell Biol 2007 Jan; 27(2):426–437. PMID: 17101793

41. Cadoret JC, Meisch F, Hassan-Zadeh V, Luyten I, Guillet C, Duret L, et al. Genome-wide studies highlight indirect links between human replication origins and gene regulation. Proc Natl Acad Sci U S A 2008 Oct 14; 105(41):15837–15842. doi: 10.1073/pnas.0805208105 PMID: 18838675

42. Dellino GI, Pelicci PG. Next-generation sequencing and DNA replication in human cells: the future has arrived. Future Oncol 2014 Mar; 10(4):683–693. doi: 10.2217/fon.13.182 PMID: 24754597

43. Blow JJ, Ge XQ, Jackson DA. How dormant origins promote complete genome replication. Trends Biochem Sci 2011 Aug; 36(8):405–414. doi: 10.1016/j.tibs.2011.05.002 PMID: 21641805

44. Alver RC, Chadha GS, Blow JJ. The contribution of dormant origins to genome stability: from cell biology to human genetics. DNA Repair (Amst) 2014 Jul; 19:182–189.

45. Ge XQ, Han J, Cheng EC, Yamaguchi S, Shima N, Thomas JL, et al. Embryonic Stem Cells License a High Level of Dormant Origins to Protect the Genome against Replication Stress. Stem Cell Reports 2015 Aug 11; 5(2):185–194. doi: 10.1016/j.stemcr.2015.06.002 PMID: 26190528

46. Woodward AM, Gohler T, Luciani MG, Oehlmann M, Ge X, Gartner A, et al. Excess Mcm2-7 license dormant origins of replication that can be used under conditions of replicative stress. J Cell Biol 2006 Jun 5; 173(5):673–683. PMID: 16754955

47. Conti C, Leo E, Eichler GS, Sordet O, Martin MM, Fan A, et al. Inhibition of histone deacetylase in cancer cells slows down replication forks, activates dormant origins, and induces DNA damage. Cancer Res 2010 Jun 1; 70(11):4470–4480. doi: 10.1158/0008-5472.CAN-09-3028 PMID: 20460513

48. Guilbaud G, Rappailles A, Baker A, Chen CL, Arneodo A, Goldar A, et al. Evidence for sequential and increasing activation of replication origins along replication timing gradients in the human genome. PLoS Comput Biol 2011 Dec; 7(12):e1002322. doi: 10.1371/journal.pcbi.1002322 PMID: 22219720

49. Hyrien O. Peaks cloaked in the mist: the landscape of mammalian replication origins. J Cell Biol 2015 Jan 19; 208(2):147–160. doi: 10.1083/jcb.201407004 PMID: 25601401

50. Mesner LD, Valsakumar V, Cieslik M, Pickin R, Hamlin JL, Bekiranov S. Bubble-seq analysis of the human genome reveals distinct chromatin-mediated mechanisms for regulating early- and late-firing origins. Genome Res 2013 Nov; 23(11):1774–1788. doi: 10.1101/gr.155218.113 PMID: 23861383

51. Besnard E, Babled A, Lapasset L, Milhavet O, Parrinello H, Dantec C, et al. Unraveling cell type-specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. Nat Struct Mol Biol 2012 Aug; 19(8):837–844. doi: 10.1038/nsmb.2339 PMID: 22751019

52. Guan Z, Hughes CM, Kosiyatrakul S, Norio P, Sen R, Fiering S, et al. Decreased replication origin activity in temporal transition regions. J Cell Biol 2009 Nov 30; 187(5):623–635. doi: 10.1083/jcb.200905144 PMID: 19951913

53. Norio P, Kosiyatrakul S, Yang Q, Guan Z, Brown NM, Thomas S, et al. Progressive activation of DNA replication initiation in large domains of the immunoglobulin heavy chain locus during B cell development. Mol Cell 2005 Nov 23; 20(4):575–587. PMID: 16307921

54. Schultz SS, Desbordes SC, Du Z, Kosiyatrakul S, Lipchina I, Studer L, et al. Single-molecule analysis reveals changes in the DNA replication program for the POU5F1 locus upon human embryonic stem cell differentiation. Mol Cell Biol 2010 Sep; 30(18):4521–4534. doi: 10.1128/MCB.00380-10 PMID: 20647538

55. Shishkin AA, Voineagu I, Matera R, Cherng N, Chernet BT, Krasilnikova MM, et al. Large-scale expansions of Friedreich's ataxia GAA repeats in yeast. Mol Cell 2009 Jul 10; 35(1):82–92. doi: 10.1016/j.molcel.2009.06.017 PMID: 19595718

56. Palumbo E, Matricardi L, Tosoni E, Bensimon A, Russo A. Replication dynamics at common fragile site FRA6E. Chromosoma 2010 Dec; 119(6):575–587.

57. O'Keefe RT, Henderson SC, Spector DL. Dynamic organization of DNA replication in mammalian cell nuclei: spatially and temporally defined replication of chromosome-specific alpha-satellite DNA sequences. J Cell Biol 1992 Mar; 116(5):1095–1110. PMID: 1740468

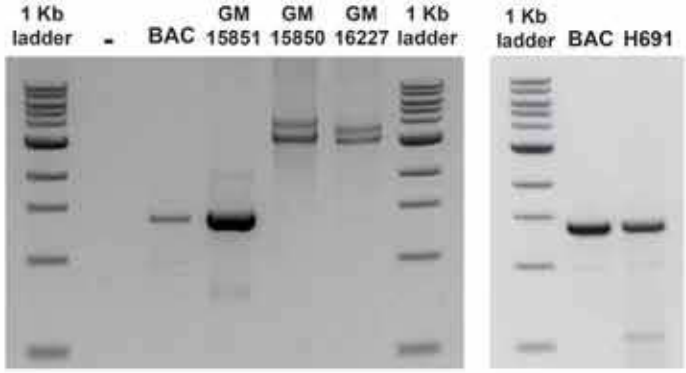
58. Conti C, Sacca B, Herrick J, Lalou C, Pommier Y, Bensimon A. Replication fork velocities at adjacent replication origins are coordinately modified during DNA replication in human cells. Mol Biol Cell 2007 Aug; 18(8):3059–3067. PMID: 17522385

S1 Fig.

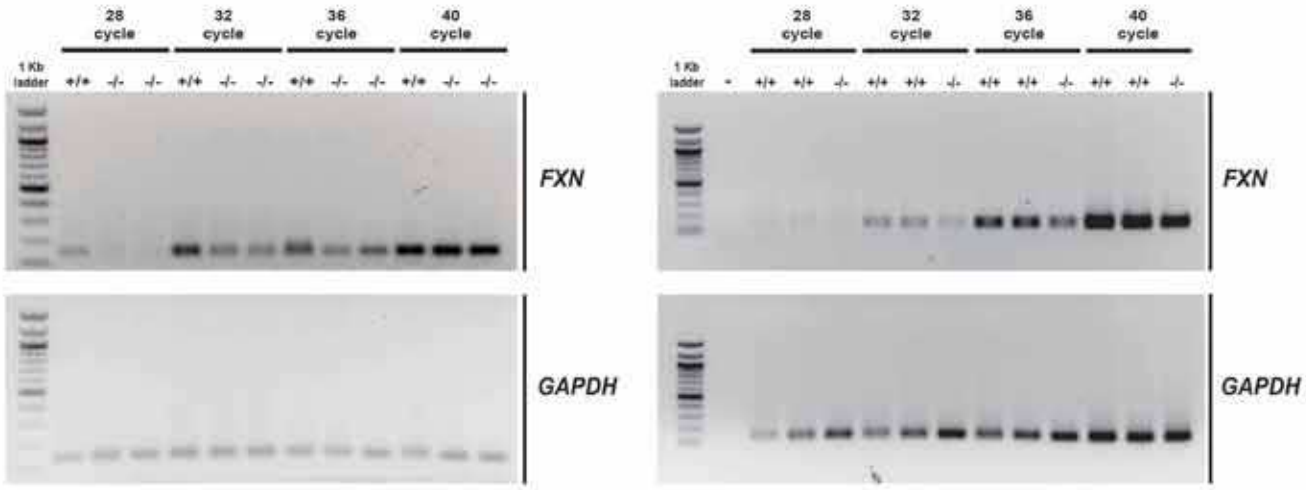
A

Cell lines	Genotype	GAA/TTC repeats
GM15851	+/+	wt/wt
GM15850	-/-	650/1030
GM16227	-/-	630/830
H691	+/+	wt/wt

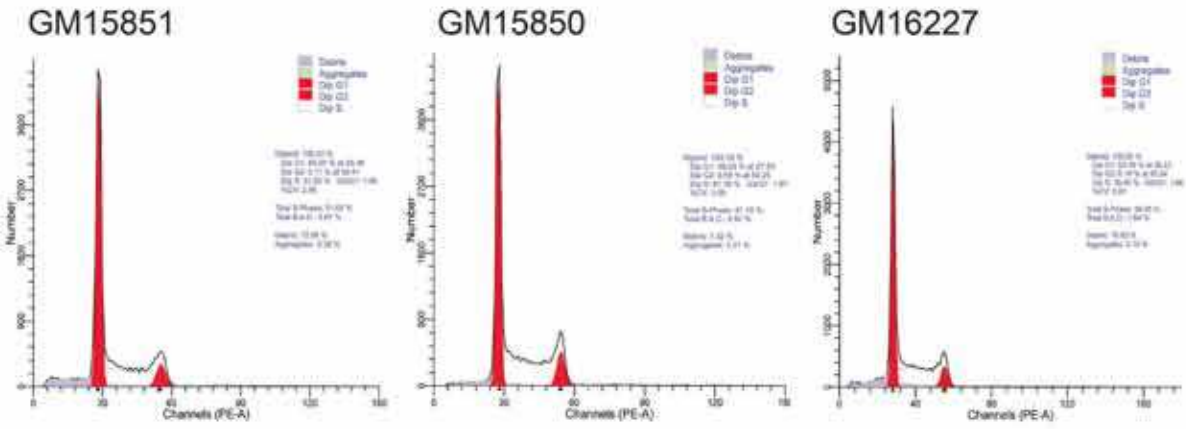
B



C



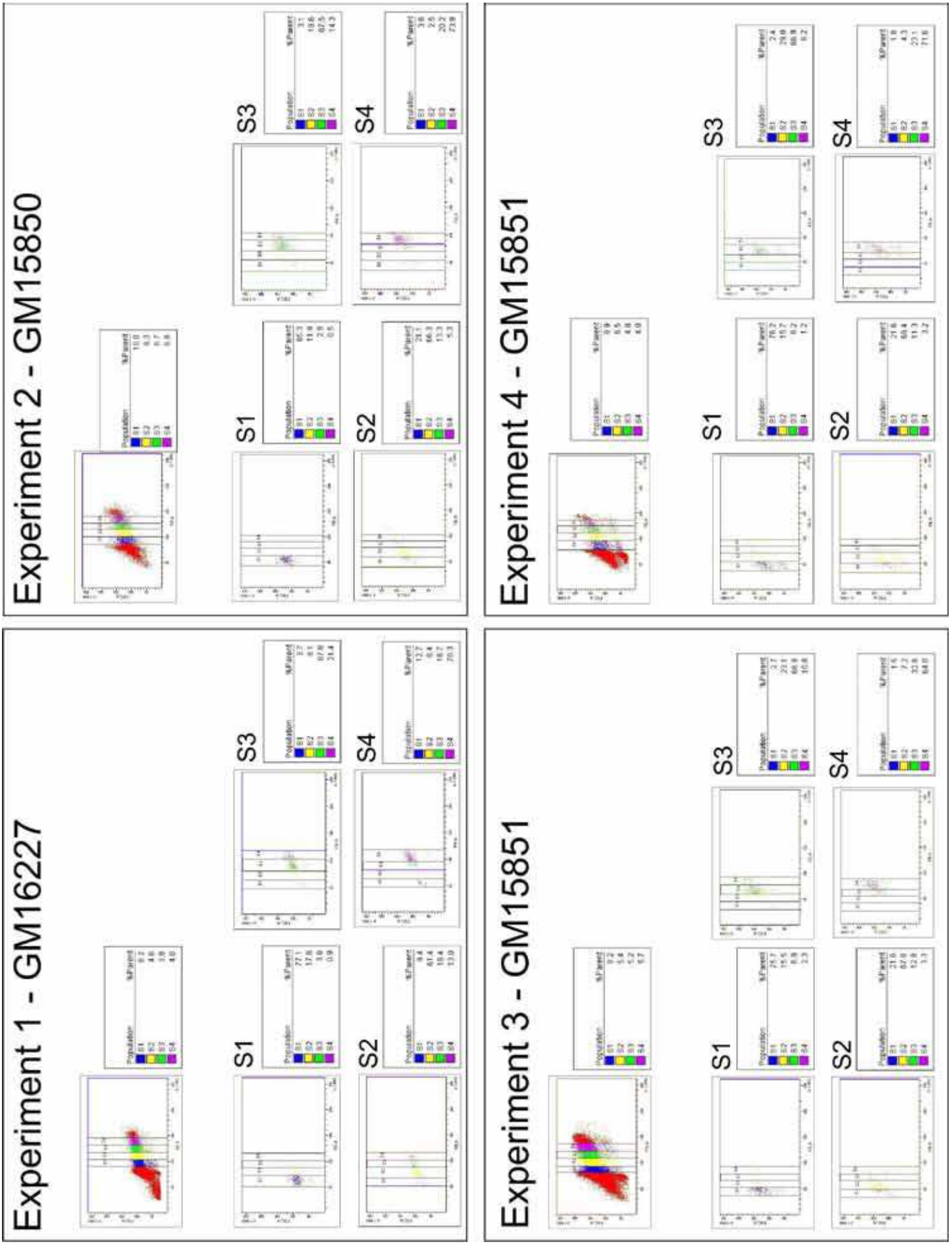
D



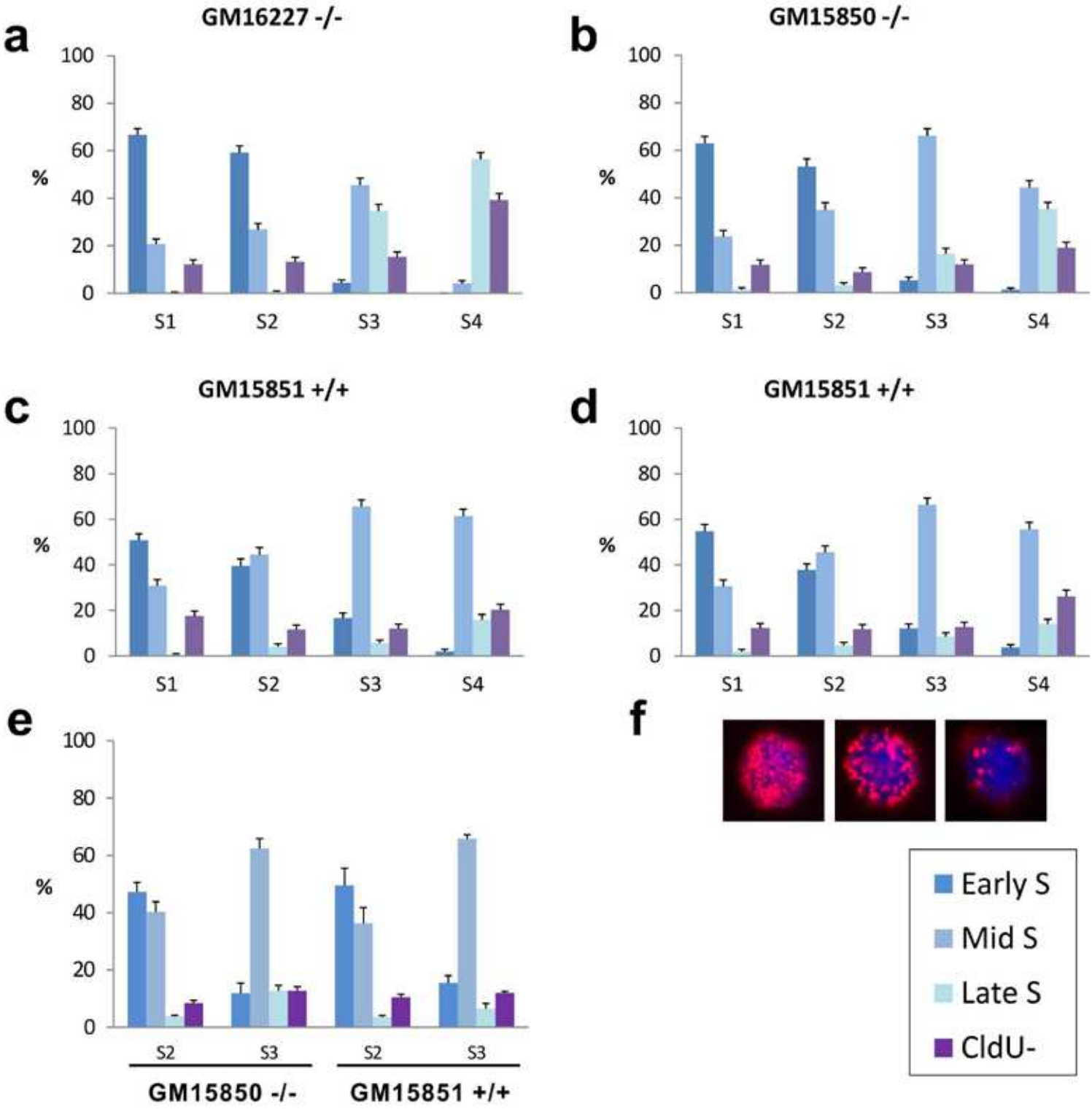
E

Cell line		GM15851 +/+	GM15850 -/-	GM16227 -/-
Replication forks	Total	133	145	67
	Unidirectional forks (%)	33.8	31.0	28.4
	Paused/arrested forks (%)	12.8	8.3	3.0
	Asynchronous forks (%)	7.5	7.6	6.0
Fork rate (kb/min)	Mean± SE	1.89 ±0.90	1.78 ± 0.062	1.70± 0.09
	N	88	100	47
IOD (kb)	Mean ± SE	220.8 ± 15.86	222.9 ± 22.84	171.4 ± 22.08
	N	54	13	23

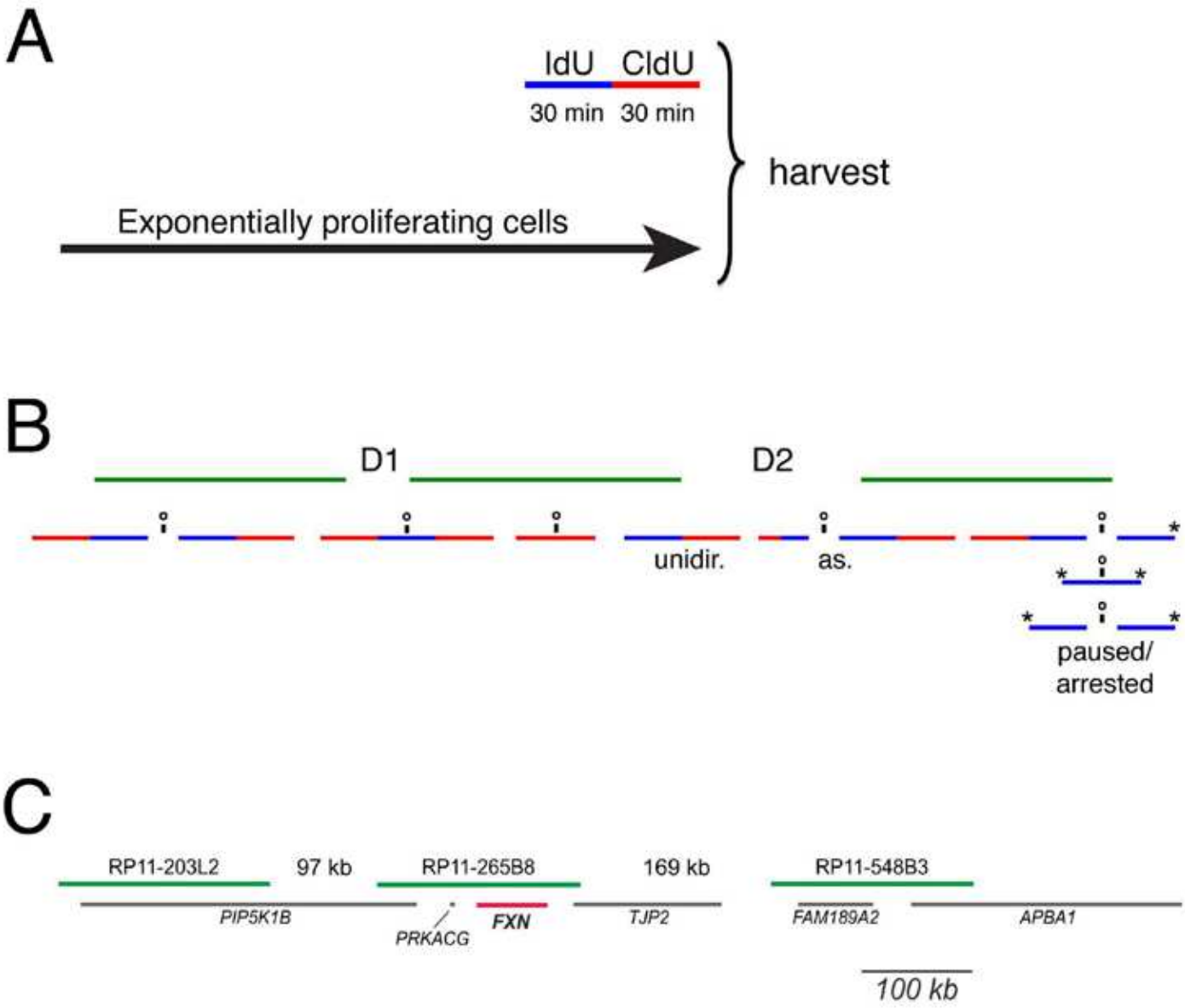
S2 Fig.


Experiment 1 - GM16227
Experiment 2 - GM15850
Experiment 3 - GM15851
Experiment 4 - GM15851
S1
S2
S3
S4

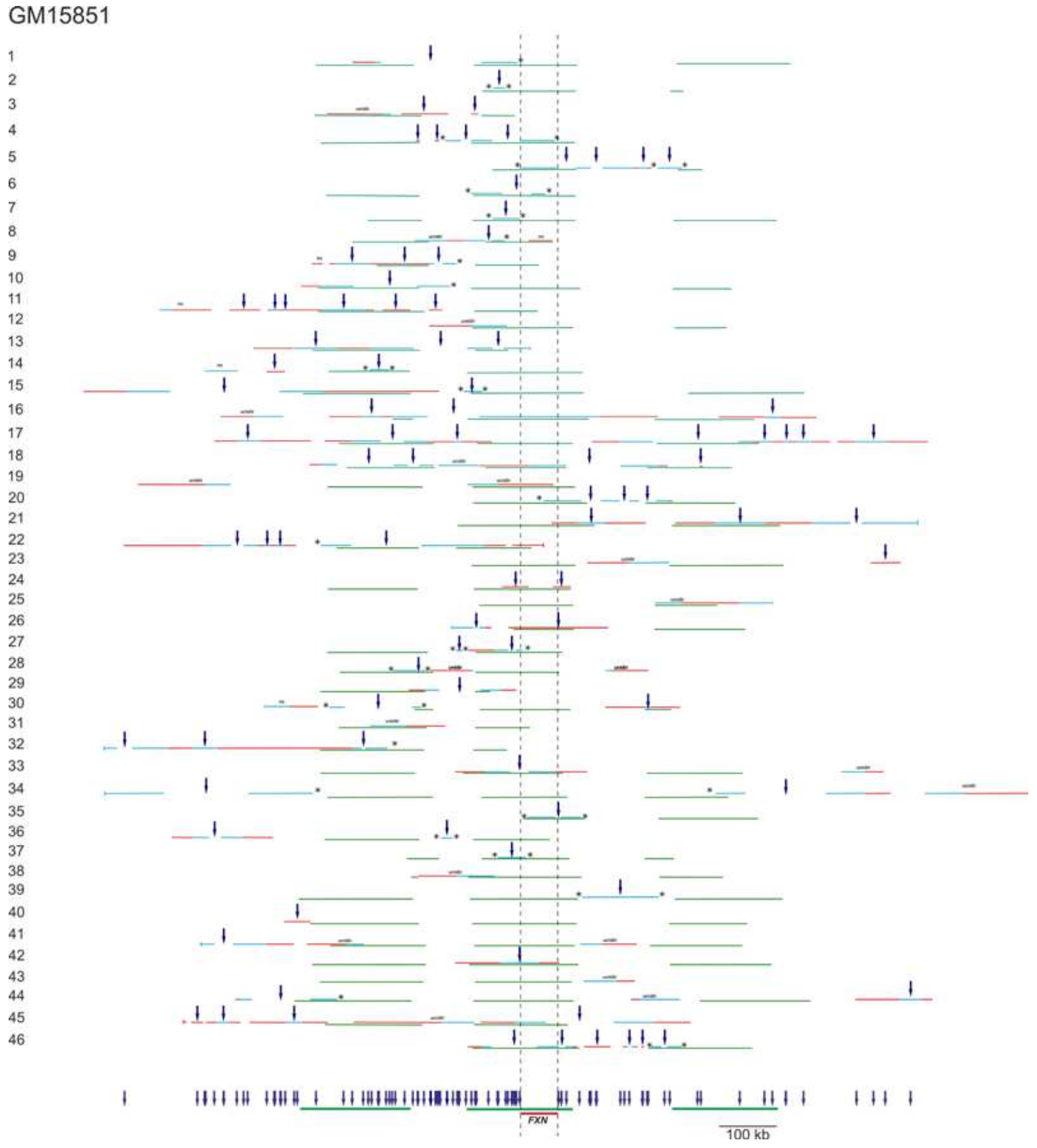
S3 Fig.


a
GM16227 -/-
%
100
80
60
40
20
0
S1
S2
S3
S4
b
GM15850 -/-
c
GM15851 +/+
d
GM15851 +/+
e
S2
S3
S2
S3
GM15850 -/-
GM15851 +/+
f
Early S
Mid S
Late S
CldU-

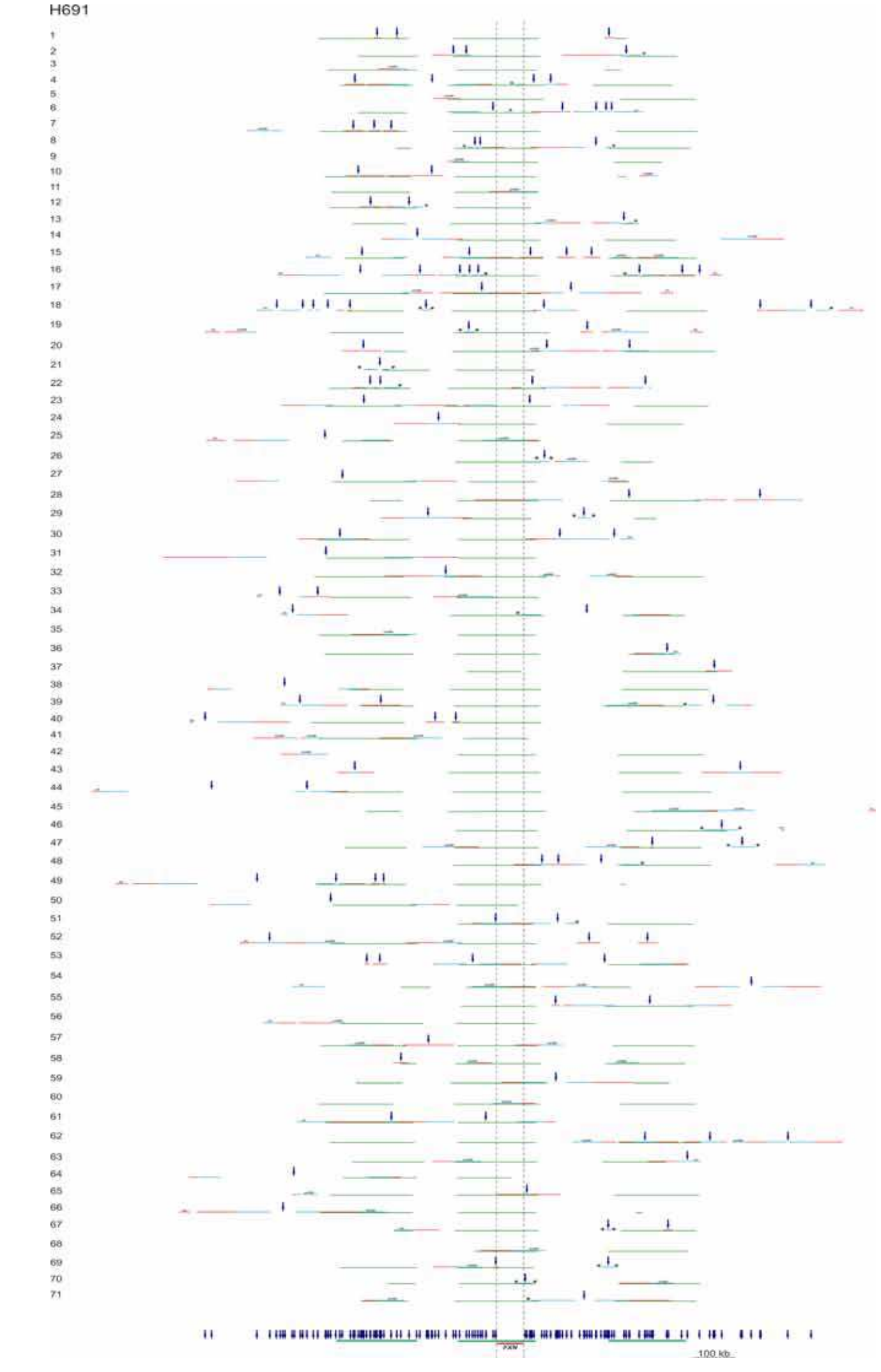
S4 Fig.


A
IdU
CldU
30 min
30 min
harvest
Exponentially proliferating cells
B
D1
D2
unidir.
as.
paused/
arrested
C
RP11-203L2
97 kb
RP11-265B8
169 kb
RP11-548B3
PIP5K1B
PRKACG
FXN
TJP2
FAM189A2
APBA1
100 kb

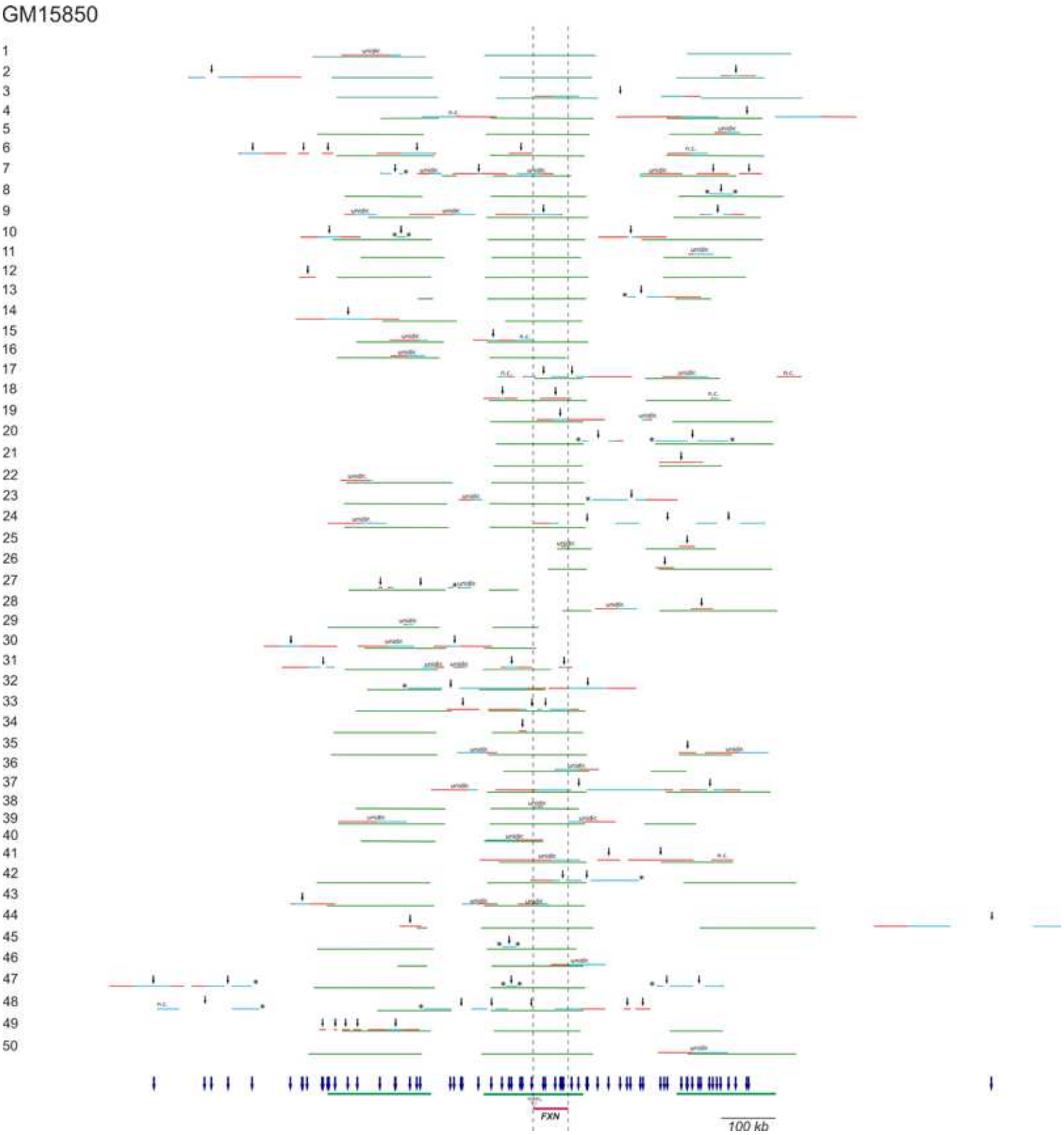
S5 Fig.


GM15851
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
FXN
100 kb

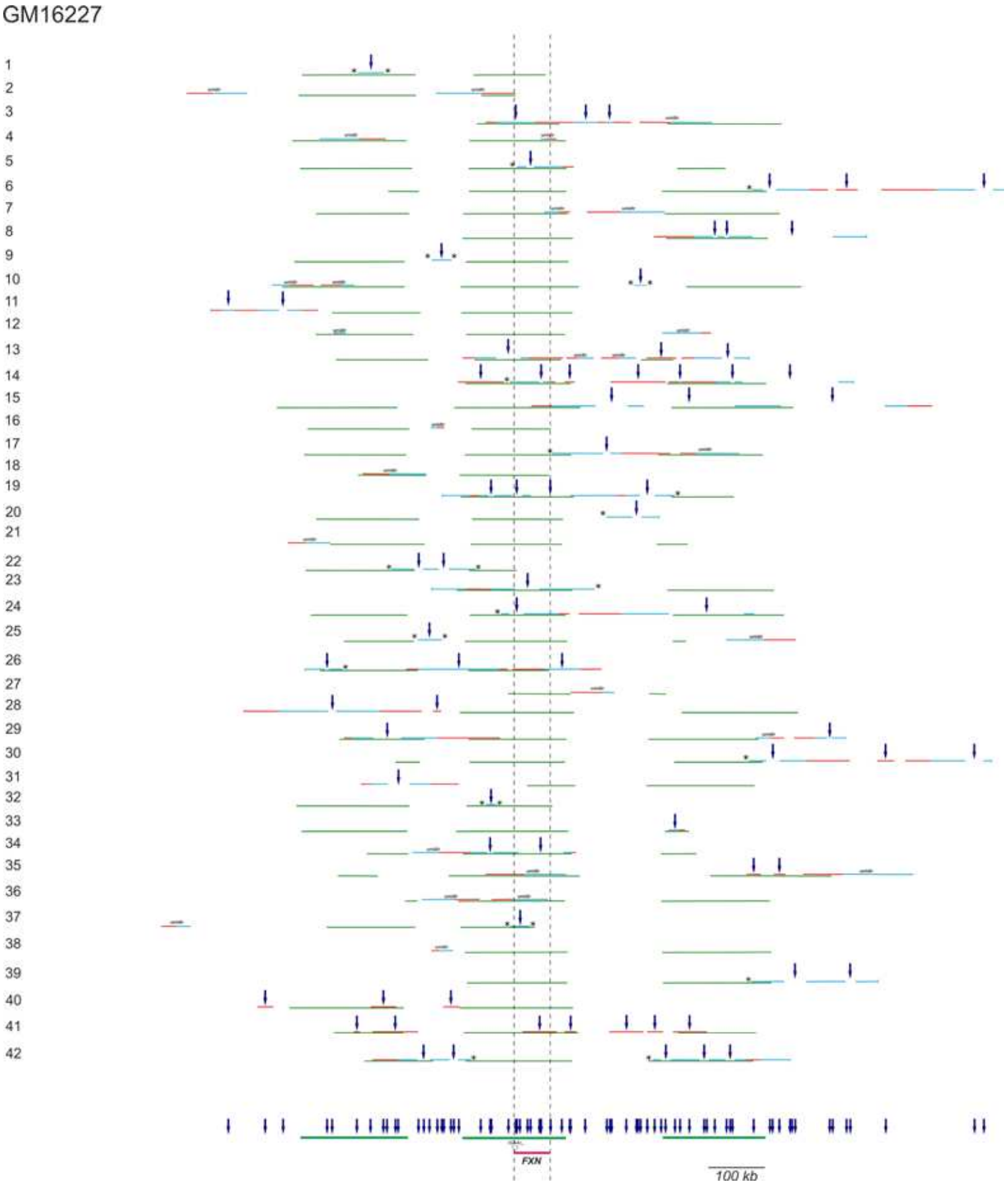
S6 Fig.


H691
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
PAX7
100 kb

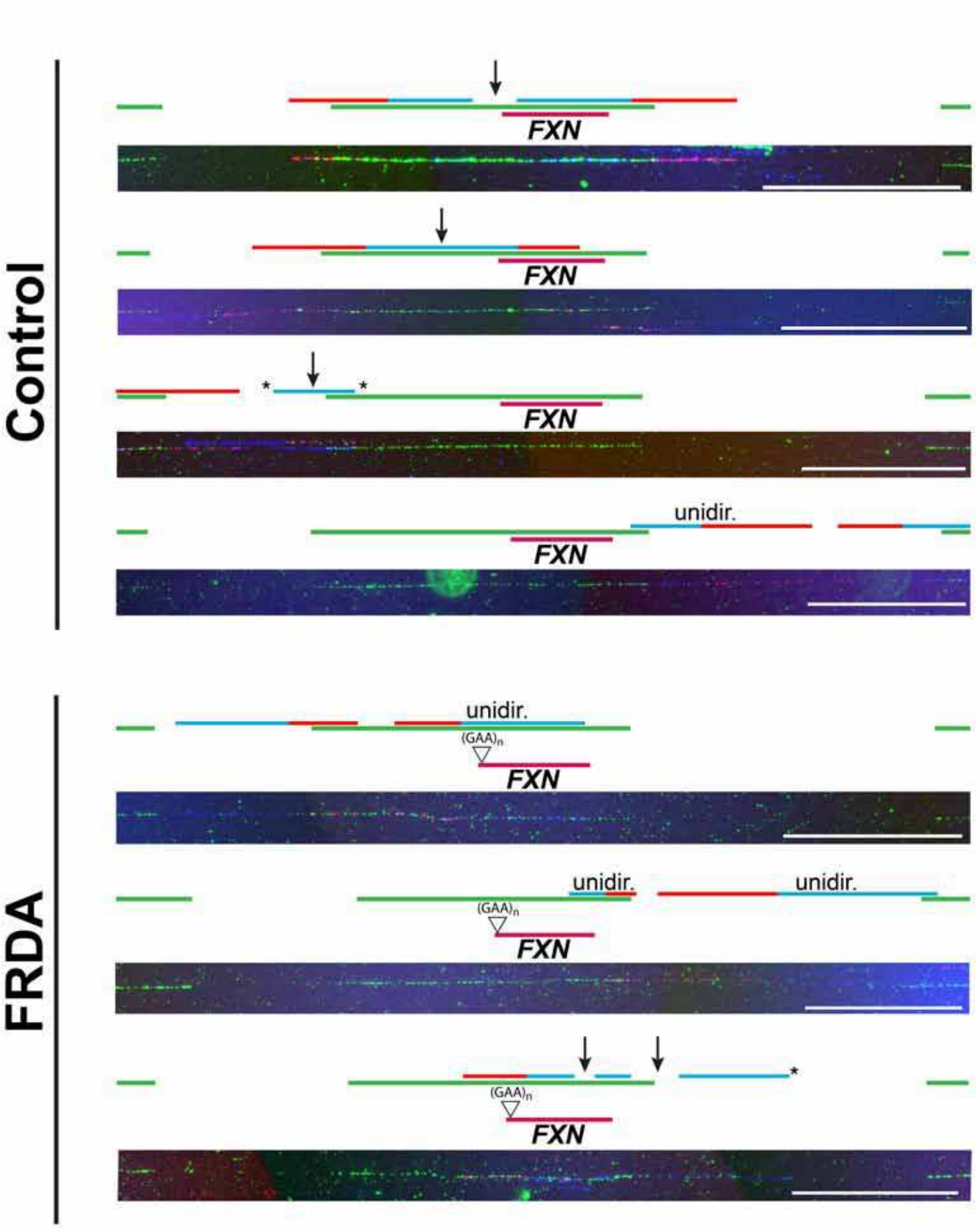
S7 Fig.


GM15850
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
FXN
100 kb

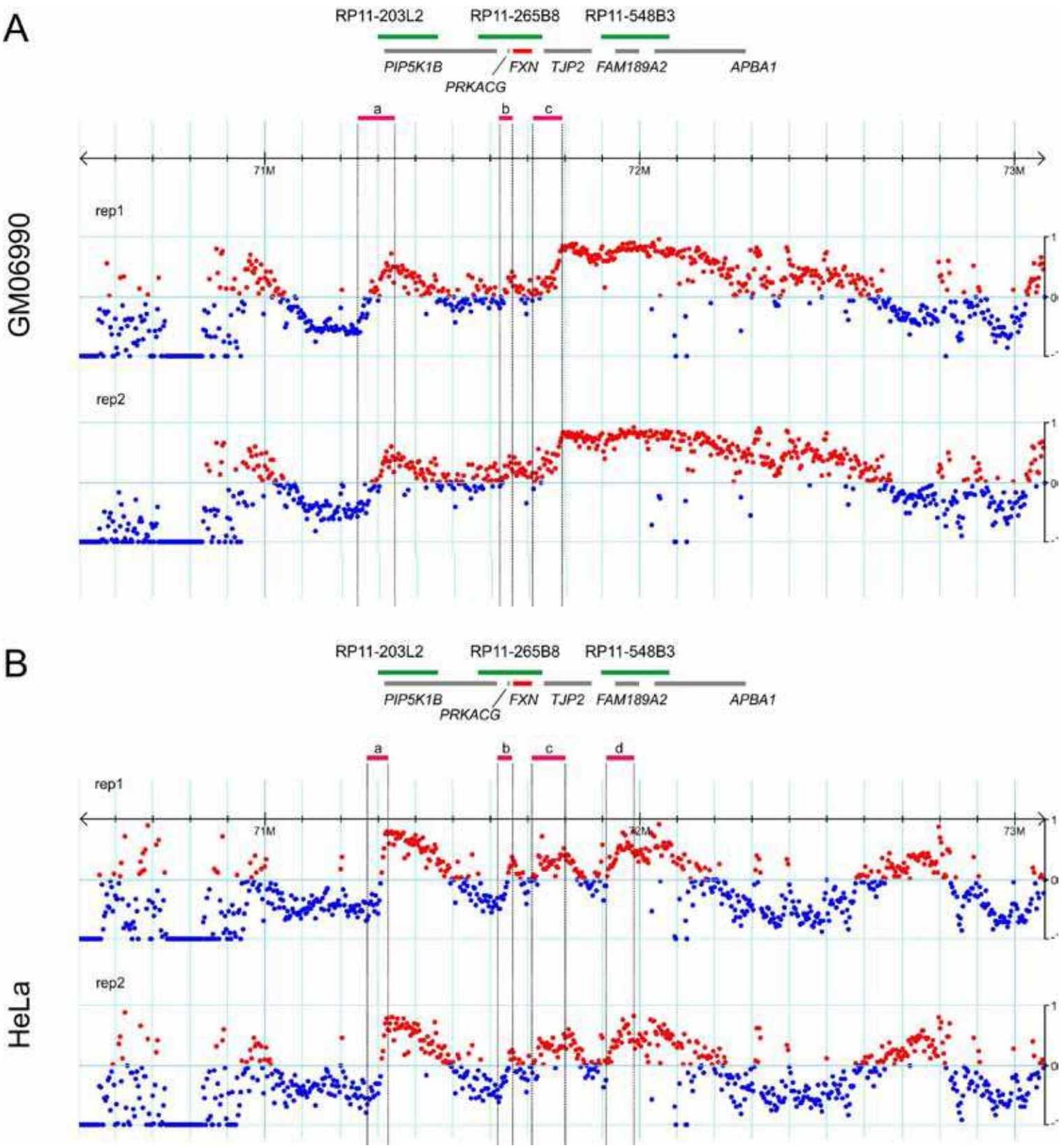
S8 Fig.


GM16227
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
FXN
100 kb

S9 Fig.


Control
FXN
FXN
FXN
unidir.
FXN
FRDA
unidir.
(GAA)n
FXN
unidir.
unidir.
(GAA)n
FXN
(GAA)n
FXN

S10 Fig.


A
RP11-203L2
RP11-265B8
RP11-548B3
PIP5K1B
PRKACG
FXN
TJP2
FAM189A2
APBA1
a
b
c
71M
72M
73M
rep1
rep2
GM06990
B
RP11-203L2
RP11-265B8
RP11-548B3
PIP5K1B
PRKACG
FXN
TJP2
FAM189A2
APBA1
a
b
c
d
71M
73M
rep1
rep2
HeLa

S1 Table. Raw data of the replication timing analysis of *FXN* carried out by interphase FISH.

Locus	Cell line, *FXN* genotype	Total cells	Cells with SS pattern N (% ± SE)	Cells with SD pattern N (% ± SE)	Cells with DD pattern N (% ± SE)	Cells with other patterns N (% ± SE)
FXN	GM15851, control[§]	288	65 (22.6 ± 2.46)	110 (38.2 ± 2.86)	97 (33.7 ± 2.78)	16 5.6 ± 1.35
	GM15851, control[ç]	273	78 (28.6 ± 2.73)	98 (35.9 ± 2.90)	84 (30.8 ± 2.79)	13 (4.8 ± 1.29)
	GM16227, FRDA	252	76 (30.2 ± 2.89)	81 (32.1 ± 2.94)	36 (14.3 ± 2.20)	59 (23.4 ± 2.67)
	GM15850, FRDA[°]	247	74 (30.0 ± 2.91)	80 (32.4 ± 2.98)	82 (33.2 ± 3.00)	11 (4.5 ±1.31)
	GM15850, FRDA	373	94 (25.2 ± 2.25)	136 (36.5 ± 2.49)	117 (31.4 ± 2.40)	26 (7.0 ± 1.32)
FRA3B[#]	GM15851, control[§]	294	97 (33.0 ± 2.74)	137 (46.6 ± 2.90)	46 (15.6 ± 2.12)	14 (4.8 ± 1.24)
	GM15851, control[ç]	288	86 (29.9 ± 2.70)	120 (41.7± 2.91)	71 (24.7 ± 2.54)	11 (3.8 ± 1.13)

[§] Pre-sorting control population of Experiment 3 (see Supplementary Figure 2)

[ç] Pre-sorting control population of Experiment 4 (see Supplementary Figure 2)

[°] Pre-sorting control population of Experiment 1 (see Supplementary Figure 2)

[#] The late replicating common fragile site *FRA3B* has been evaluated as positive control

S2 Table. Raw data of the replication timing analysis of *FXN* carried out by interphase FISH after FACS cell sorting.

Sorting experiment[ç]	Cell fraction	Replication patterns[#]	Total cells	S-phase cells[#] N	early S-phase[#] N (% ± SE)[§]	mid S-phase[#] N (% ± SE)[§]	late S-phase[#] N (% ± SE)[§]	Hybridization efficiency
Experiment 1 (GM16227, FRDA cells)	S1	SS	175	154	146 (63.8 ± 3.18)	8 (11.3 ± 3.75)	0	91.1
		DD	10	6	0	6 (8.5 ± 3.30)	0	
		SD	108	100	58 (25.3 ± 2.87)	41 (57.7 ± 5.86)	1	
		Others	50	41	25 (10.9 ± 2.06)	16 (22.5 ± 4.96)	0	
		Total	343	301	229	71	1	
	S2	SS	109	89	82 (43.9 ± 3.63)	7 (8.2 ± 2.98)	0	89.9
		DD	21	15	2 (1.1 ± 0.75)	12 (14.1 ± 3.78)	1	
		SD	137	130	76 (40.6 ± 3.59)	53 (62.4 ± 5.26)	1	
		Others	49	40	27 (14.4 ± 2.57)	13 (15.3 ± 3.90)	0	
		Total	316	274	187	85	2	
	S3	SS	21	16	9 (69.2 ± 12.80)	7 (5 ± 1.92)	0	93.0
		DD	115	95	0	32 (23.9 ± 3.68)	63 (61.8 ± 4.81)	
		SD	120	108	1 (7.7 ± 7.39)	79 (59.0 ± 4.25)	28 (27.5 ± 4.42)	
		Others	38	30	3 (23.1 ±11.68)	16 (11.9 ± 2.80)	11 (10.8 ± 3.07)	
		Total	294	249	13	134	102	

Sorting experiment[ç]	Cell fraction	Replication patterns[#]	Total cells	S-phase cells[#] N	early S-phase[#] N (% ± SE)[§]	mid S-phase[#] N (% ± SE)[§]	late S-phase[#] N (% ± SE)[§]	Hybridization efficiency
	S4	SS	1	0	0	0	0	94.6
		DD	247	155	0	1 (7.1 ± 6.88)	154 (82.4 ± 2.79)	
		SD	55	29	0	8 (57.1 ± 13.23)	21 (11.2 ± 2.31)	
		Others	28	17	0	5 (35.7 ± 12.81)	12 (6.4 ± 1.79)	
		Total	331	201	0	14	187	
Experiment 2 (GM15850, FRDA cells)	S1	SS	136	116	105 (60.0 ± 3.70)	11 (17 ± 4.59)	0	94.8
		DD	11	7	0	4 (6 ± 2.94)	3	
		SD	104	101	61 (34.9 ± 3.60)	39 (59.1 ± 6.05)	1	
		Others	27	21	9 (5.1 ± 1.67)	12 (18.2 ± 4.75)	0	
		Total	278	245	175	66	4	
	S2	SS	105	104	96 (71.6 ± 3.89)	8 (9.1 ± 3.06)	0	96.0
		DD	28	19	0	14 (15.9 ± 3.90)	5	
		SD	101	91	26 (19.4 ± 3.42)	63 (71.6 ± 4.81)	2	
		Others	18	16	12 (9.0 ± 2.47)	3 (3.4 ± 1.93)	1	
		Total	252	230	134	88	8	
	S3	SS	12	12	10 (67.0 ± 12.17)	2 (1.1 ± 0.75)	0	97.9

Sorting experiment[ç]	Cell fraction	Replication patterns[#]	Total cells	S-phase cells[#] N	early S-phase[#] N (% ± SE)[§]	mid S-phase[#] N (% ± SE)[§]	late S-phase[#] N (% ± SE)[§]	Hybridization efficiency
		DD	124	109	1 (7.0 ± 6.44)	68 (36.2 ± 3.50)	40 (85 ± 5.19)	
		SD	137	123	4 (26.7 ± 11.42)	113 (60.1 ± 3.57)	6 (12.8 ± 4.87)	
		Others	11	6	0	5 (2.7 ± 1.17)	1 (2.1 ±2.10)	
		Total	284	250	15	188	47	
	S4	SS	8	4	3	1 (0.8 ± 0.75)	0	97.7
		DD	189	154	0	64 (48.1 ± 4.33)	90 (85.0 ± 3.48)	
		SD	89	73	1	62 (46.6 ± 4.33)	10 (9.4 ± 2.84)	
		Others	14	12	0	6 (4.5 ± 1.80)	6 (5.7 ± 2.24)	
		Total	300	243	4	133	106	
Experiment 3 (GM15851, controls)	S1	SS	121	96	90 (57.7 ± 3.96)	6 (6.3 ± 2.50)	0	96.1
		DD	21	17	0	15 (15.8 ± 3.74)	2	
		SD	142	121	56 (35.9 ± 3.84)	65 (68.4 ± 4.77)	0	
		Others	23	19	10 (6.4 ± 1.96)	9 (9.5 ± 3.00)	0	
		Total	307	253	156	95	2	
	S2	SS	67	57	50 (47.6 ± 4.87)	7 (5.9 ± 2.17)	0	97.5
		DD	51	39	1 (1.0± 0.95)	29 (24.6 ± 3.96)	9 (81.8 ± 11.63)	

Sorting experiment[ç]	Cell fraction	Replication patterns[#]	Total cells	S-phase cells[#] N	early S-phase[#] N (% ± SE)[§]	mid S-phase[#] N (% ± SE)[§]	late S-phase[#] N (% ± SE)[§]	Hybridization efficiency
		SD	136	128	51 (48.6 ± 4.88)	75 (63.6 ± 4.43)	2 (18.2 ± 11.63)	
		Others	11	10	3 (2.9 ± 1.63)	7 (5.9 ± 2.17)	0	
		Total	265	234	105	118	11	
	S3	SS	24	15	9 (19.1 ± 5.74)	6 (3.2 ± 1.30)	0	97.0
		DD	95	83	1 (2.1 ± 2.10)	70 (37.8 ± 3.57)	12 (75.0 ± 10.83)	
		SD	146	136	33 (70.2± 6.67)	99± (53.5 ± 3.67)	4 (25.0 ± 10.83)	
		Others	17	14	4 (8.5 ± 4.07)	10 (5.4 ± 1.66)	0	
		Total	282	248	47	185	16	
	S4	SS	9	8	2	6 (3.5 ± 1.40)	0	97.5
		DD	176	139	0	97 (56.4 ± 3.78)	42 (93.3 ± 3.72)	
		SD	84	68	4	62 (36.0 ± 3.66)	2 (4.4 ± 3.07)	
		Others	11	8	0	7 (4.1 ± 1.51)	1 (2.2 ± 2.20)	
		Total	280	223	6	172	45	
Experiment 4 (GM15851, controls)	S1	SS	123	105	103 (64.8 ± 3.78)	2 (2.2 ± 1.57)	0	98.4
		DD	23	15	0	9 (10.1 ± 3.20)	6	
		SD	136	128	55 (34.6 ± 3.77)	73 (82.0± 4.07)	0	

Sorting experiment[ç]	Cell fraction	Replication patterns[#]	Total cells	S-phase cells[#] N	early S-phase[#] N (% ± SE)[§]	mid S-phase[#] N (% ± SE)[§]	late S-phase[#] N (% ± SE)[§]	Hybridization efficiency
		Others	8	6	1 (0.6 ± 0.63)	5 (5.6 ± 2.44)	0	
		Total	290	254	159	89	6	
	S2	SS	74	58	55 (47.0 ± 4.61)	3 (2.1 ± 1.22)	0	97.3
		DD	57	51	1 (0.9 ± 0.85)	38 (27.0 ± 3.74)	12 (80.0 ± 10.33)	
		SD	164	151	57 (48.7 ± 4.62)	91 (64.5 ± 4.03)	3 (20.0 ± 10.33)	
		Others	15	13	4 (3.4 ± 1.68)	9 (6.4 ± 2.06)	0	
		Total	310	273	117	141	15	
	S3	SS	24	15	11 (32.4 ± 8.02)	4 (2.2 ± 1.06)	0	98.8
		DD	114	96	0	72 (38.7 ± 3.57)	24	
		SD	135	128	22 (64.7 ± 8.20)	106 (57.0 ± 3.63)	0	
		Others	7	5	1 (2.9 ± 2.90)	4 (2.2 ± 1.06)	0	
		Total	280	244	34	186	24	
	S4	SS	13	3	2 (18.2 ± 11.63)	1 (0.6 ± 0.62)	0	98.8
		DD	170	124	1 (9.1 ± 8.67)	85 (52.8 ± 3.93)	38 (92.7 ± 4.07)	
		SD	99	81	7 (63.6 ± 14.50)	71 (44.1 ± 3.91)	3 (7.3 ± 4.07)	
		Others	7	5	1 (9.1 ±8.67)	4 (2.5 ± 1.23)	0	

Sorting experiment[ç]	Cell fraction	Replication patterns[#]	Total cells	S-phase cells[#] N	early S-phase[#] N (% ± SE)[§]	mid S-phase[#] N (% ± SE)[§]	late S-phase[#] N (% ± SE)[§]	Hybridization efficiency
		Total	289	213	11	161	41	

[ç] As described in Supplementary Figure 2

Replication patterns are based on features of the FISH signals; S-phase cells are classified according to CldU-labelling. All details in Materials and Methods

[§] Percentages and SE of percentages were calculated only if > 10 total cells were observed per each S-phase substage

S3 Table. Replication timing of the late replicating sequence *FRA3B* according to interphase FISH after FACS cell sorting.

Cell sample	Cell fraction	Replication patterns[#]	Total cells	S-phase cells[#] N	early S-phase[#] N (% ± SE)[§]	mid S-phase[#] N (% ± SE)[§]	late S-phase[#] N (% ± SE)[§]
Control (GM15851 cells)	S1	SS	131	97	90 (61.2 ± 4.02)	7 (9.9 ±3.54)	0
		DD	20	16	1 (0.7 ± 0.68)	10 (14.1 ±4.1)	5
		SD	125	102	49 (33.3 ± 3.89)	53 (74.6 ±5.16)	0
		Others	10	8	7 (4.8 ± 1.76)	1 (1.4 ±1.40)	0
		Total	286	223	147	71	5
	S2	SS	101	94	86 (61.9 ± 4.12)	8 (7.3 ± 2.48)	0
		DD	27	24	1 (0.7 ± 0.72)	15 (13.6 ± 3.27)	8 (66.7 ± 13.61)
		SD	145	132	49 (35.3 ± 4.05)	80 (72.7 ± 4.25)	3 (25.0 ± 12.50)
		Others	12	11	3 (2.2 ± 1.23)	7 (6.4 ± 2.33)	1 (8.3 ± 7.98)
		Total	285	261	139	110	12
	S3	SS	39	27	21 (37.5 ± 6.47)	6 (3.2 ± 1.30)	0
		DD	85	68	1 (1.8 ± 1.78)	57 (30.6 ± 3.38)	10 (83.3 ± 10.76)
		SD	156	147	31 (55.4 ± 6.64)	115 (61.8 ± 3.56)	1 (8.3 ± 7.98)
		Others	12	12	3 (5.4 ± 3.01)	8 (4.3 ± 1.49)	1 (8.3 ± 7.98)
		Total	292	254	56	186	12

Cell sample	Cell fraction	Replication patterns[#]	Total cells	S-phase cells[#] N	early S-phase[#] N (% ± SE)[§]	mid S-phase[#] N (% ± SE)[§]	late S-phase[#] N (% ± SE)[§]
FRDA (15850 cells)	S4	SS	11	3	1 (8.3 ± 7.98)	2 (1.1 ± 0.76)	0
		DD	164	127	0	85 (46.2 ± 3.68)	42 (84.0 ± 5.19)
		SD	130	110	11 (91.7 ± 7.98)	92 (50.0 ± 3.69)	7 (14.0 ± 4.91)
		Others	10	6	0	5 (2.7 ± 1.20)	1 (2.0 ± 1.98)
		Total	315	246	12	184	50
	S1	SS	144	116	111 (77.1 ± 3.50)	5 (6.8 ± 2.96)	0
		DD	7	2	0	0	2
		SD	94	88	28 (19.4 ± 3.30)	60 (82.2 ± 4.48)	0
		Others	16	13	5 (3.5 ± 1.53)	8 (11.0 ± 3.66)	0
		Total	261	219	144	73	2
	S2	SS	84	74	66 (64.7 ± 4.73)	8 (7.0 ± 2.37)	0
		DD	18	17	0	11 (9.6 ± 2.74)	6 (54.5 ± 15.01)
		SD	125	121	28 (27.5 ± 4.42)	89 (77.4 ± 3.90)	4 (36.4 ± 14.50)
		Others	18	16	8 (7.8 ± 2.66)	7 (6.1 ± 2.23)	1 (9.1 ± 8.67)
		Total	245	228	102	115	11

Cell sample	Cell fraction	Replication patterns[#]	Total cells	S-phase cells[#] N	early S-phase[#] N (% ± SE)[§]	mid S-phase[#] N (% ± SE)[§]	late S-phase[#] N (% ± SE)[§]
	S3	SS	26	18	16 (48.5 ± 8.70)	2 (1.2 ± 0.86)	0
		DD	73	60	0	38 (23.3 ± 3.31)	22 (84.6 ± 7.08)
		SD	134	130	16 (48.5 ± 8.70)	111 (68.1 ± 3.65)	3 (11.5 ± 6.27)
		Others	16	14	1 (3.0 ± 2.98)	12 (7.4 ± 2.05)	1 (3.8 ± 3.77)
		Total	249	222	33	163	26
	S4	SS	3	2	1	1 (0.7 ± 0.70)	0
		DD	138	108	0	47 (33.1 ± 3.95)	61 (92.4 ± 3.26)
		SD	112	92	0	89 (62.7 ± 4.06)	3 (4.5 ± 2.56)
		Others	11	8	1	5 (3.5 ± 1.55)	2 (3.0 ± 2.11)
		Total	264	210	2	142	66

Replication patterns are based on features of the FISH signals of BAC RP11-468L11; S-phase cells are classified according to CldU-labelling. All details in Materials and Methods

[§] Percentages and SE of percentages were calculated only if > 10 total cells were observed per each S-phase substage

S4 Table. Replication timing of a genomic domain located about 170 kb downstream the *FXN* allele, according to interphase FISH after FACS cell sorting.

Cell sample	Cell fraction	Replication patterns[#]	Total cells	S-phase cells[#] N	early S-phase[#] N (% ± SE)[§]	mid S-phase[#] N (% ± SE)[§]	late S-phase[#] N (% ± SE)[§]
Control (GM15851 cells)	S2	SS	114	95	95 (50.0 ± 3.67)	0	0
		DD	40	29	1 (0.5 ± 0.53)	20 (24.7 ± 4.79)	8
		SD	152	146	90 (47.4 ± 3.62)	56 (69.1 ± 5.13)	0
		Others	9	9	4 (2.1 ± 1.04)	5 (6.2 ± 2.67)	0
		Total	315	279	190	81	8
	S3	SS	16	11	10 (29.4 ± 7.81)	1 (0.5 ± 0.47)	0
		DD	134	111	0	82 (38.3 ± 3.32)	29 (93.5 ± 4.41)
		SD	159	153	22 (64.7 ± 8.20)	129 (60.3 ± 3.35)	2 (6.5 ± 4.41)
		Others	4	4	2 (5.9 ± 4.04)	2 (0.9 ± 0.66)	0
		Total	313	279	34	214	31
FRDA (15850 cells)	S2	SS	89	77	76 (53.5 ± 4.19)	1 (0.8 ± 0.84)	0
		DD	48	39	2 (1.4 ± 0.99)	26 (22.0 ± 3.82)	11 (91.7 ± 7.98)
		SD	151	143	59 (41.5 ± 4.14)	83 (70.3 ± 4.21)	1 (8.3 ± 7.98)
		Others	14	13	5 (3.5 ± 1.55)	8 (6.8 ± 2.31)	0
		Total	302	272	142	118	12

Cell sample	Cell fraction	Replication patterns[#]	Total cells	S-phase cells[#] N	early S-phase[#] N (% ± SE)[§]	mid S-phase[#] N (% ± SE)[§]	late S-phase[#] N (% ± SE)[§]
	S3	SS	27	14	13 (26.0 ± 6.20)	1 (0.6 ± 0.62)	0
		DD	115	95	0	65 (40.1 ± 3.85)	30 (90.9 ± 5.00)
		SD	135	127	35 (70.0 ± 6.48)	90 (55.6 ± 3.90)	2 (6.1 ± 4.15)
		Others	13	9	2 (4.0 ± 2.77)	6 (3.7 ± 1.48)	1 (3.0 ± 2.98)
		Total	290	245	50	162	33

Replication patterns are based on features of the FISH signals of BAC RP11-548B3; S-phase cells are classified according to CldU-labelling. All details in Materials and Methods

[§] Percentages and SE of percentages were calculated only if > 10 total cells were observed per each S-phase substage

S5 Table. Oligonucleotide primer sequences.

		Primer	Sequence	Position (GRCh38.p2)	Product (bp)
Chr.9 **CTR**	upstream *FXN*	C1_for	5'-AGGCAAGAGACAAGGCAAGACG-3'	69027030 - 69027051	114
		C1_rev	5'-TCTTCTCTGCCTCTGACCTGATG-3'	69027121 - 69027143	
	downstream *FXN*	C2_for	5'-AGGTTGTAGTCACTGGTGGGTT-3'	69205381 - 69205402	104
		C2_rev	5'-ACAGCAGCCTTTGATCCACAGA-3'	69205463 - 69205484	
		C3_for	5'-TGCAGACCATCACCAAGACAAA-3'	69206988 - 69207010	65
		C3_rev	5'-AGGAGGCAGAGTATGGTGGGAG-3'	69207031 - 69207052	
Chr.9 ***FXN***		F1_for	5'-TCCTGTCACCACTTTCCTTCCA-3'	69060619 - 69060640	91
		F1_rev	5'-GCTCAACTTCCTCCCACCAGT-3'	69060689 - 69060709	
		F2_for	5'-AACAAACCAGCAGTCCCAGATG-3'	69064602 - 69064623	112
		F2_rev	5'-GAGCTTTCCCTGTTCCCTGTGAG-3'	69064691 - 69064713	
		F3_for	5'-AACCTGGGATTCTAGCAGCCTG-3'	69086527 - 69086548	83
		F3_rev	5'-AGGGATGAGGGGAACAGGGATC-3'	69086588 - 69086609	
		F4_for	5'-ACTCTCCTGGCCTACTAGCTC-3'	69087712 - 69087732	100
		F4_rev	5'-GCTAGGTGATGTTATGAGGGGTCC-3'	69087789 - 69087812	
Chr.19 ***LAMIN B2***		LMNB2_for	5'-GGCTGGCATGGACTTTCATTTCA-3'	2428139 - 2428162	280
		LMNB2_rev	5'-CTCAGGAATAAACTCAGAGGCAGA-3'	2428288 - 2428310	
		LB2C1_for §	5'-GTTAACAGTCAGGCGCATGGG-3'	2431926 - 2431947	240
		LB2C1_rev §	5'-CCATCAGGGTCACCTCTGGTTC-3'	2432142 - 2432166	

§ according to: Giacca et al. Proc Natl Acad Sci U S A. 1994 91:7119-7123.