UNIVERSITA' DEGLI STUDI DI PADOVA

Dipartimento di Biologia

SCUOLA DI DOTTORATO DI RICERCA IN BIOSCIENZE

INDIRIZZO GENETICA E BIOLOGIA MOLECOLARE DELLO SVILUPPO

XXI CICLO

REPLICATION DYNAMICS AT COMMON FRAGILE SITE *FRA6E* (6q26)

Direttore della Scuola : Ch.mo Prof. Tullio Pozzan

Supervisore : Ch.mo Prof. Antonella Russo

Dottorando : Laura Matricardi

Table of contents

Riassunto

I siti fragili sono regioni instabili del genoma umano che mostrano una tendenza preferenziale a formare rotture cromosomiche o *gap* acromatici, visibili in cromosomi di cellule esposte, in coltura, a parziale inibizione della replicazione del DNA. Data la modalità di induzione, si ritiene che queste regioni siano costituite da DNA non replicato, formatosi in seguito allo stallo delle forche replicative. La presenza di DNA a singolo filamento può portare, quindi, alla formazione di rotture cromosomiche. L'instabilità a livello di questi siti è stata correlata alla formazione e allo sviluppo di diverse forme tumorali (Arlt *et al.*, 2003). In base alla frequenza di espressione all'interno della popolazione umana, i siti fragili vengono suddivisi in due categorie: i siti fragili rari e i siti fragili comuni.

I siti fragili rari, espressi in circa il 5% della popolazione umana, sono caratterizzati dalla presenza di ripetizioni di- o tri- nucleotidiche ($CGG_{(n)}$ e $AT_{(n)}$). L'espressione dei siti fragili rari, ereditati con modalità mendeliana, è direttamente associata all'espansione degli elementi ripetuti che li caratterizzano. Il meccanismo è alla base di diverse patologie associate alle mutazioni dinamiche, di cui la sindrome dell'X fragile è un esempio. Invece i siti fragili comuni (CFS) sono sequenze costitutive dei cromosomi che presentano un'elevata instabilità strutturale espressa *in vitro* in condizioni di stress replicativo. Essendo una componente normale della struttura cromosomica, i CFS sono virtualmente espressi in tutti gli individui, anche se non con la stessa frequenza. Sono regioni ricche in AT e siti preferenziali di riarrangiamento, amplificazione genica, rotture cromosomiche. Essi occupano regioni molto ampie nel genoma, dell'ordine delle Mb, e presentano spesso al loro interno grandi geni con sequenze intronic molto lunghe (Glover *et al.*, 2005), come ad esempio *FHIT*, *WWOX* e *PARK2* presenti rispettivamente in *FRA3B*, *FRA16D* e *FRA6E*. Per alcuni di questi geni, inoltre, è stato dimostrato il ruolo di oncosoppressore (Schwartz *et al.*, 2005).

Dai dati presenti in letteratura, è noto che la replicazione dei siti fragili comuni avviene tardivamente all'interno della fase S (Le Beau *et al.*, 1998), e si ritiene,

quindi, che la formazione di rotture a doppio filamento in questi siti sia la conseguenza dello stallo delle forche replicative, riconducibile ad una perturbazione del processo replicativo.

In questo studio, l'attenzione è stata focalizzata sul sito fragile comune *FRA6E*, il terzo CFS maggiormente espresso nella popolazione umana. *FRA6E* mappa sul cromosoma 6, in posizione q26, in una regione che presenta un'alta frequenza di riarrangiamenti in diversi tipi di tumori. All'interno di questo sito fragile comune mappano diversi geni ed alcuni di essi presentano lunghe sequenze introniche, come ad esempio *ARID1B* e *PARK2*; per quest'ultimo si suppone un ruolo di oncosoppressore. Nel laboratorio in cui ho lavorato, è stato dimostrato in precedenza che *FRA6E* si estende per circa 9 Mb (Russo *et al.*, 2006) e può essere suddiviso in tre sottoregioni, in base al suo differente *pattern* di fragilità: la sottoregione centrale è meno fragile rispetto a quella centromerica, in cui mappa il gene *ARID1B* e a quella telomerica, in cui mappa *PARK2*.

Lo scopo principale di questo progetto è stato quello di indagare se l'instabilità del sito fragile comune *FRA6E* può essere correlata a difetti nella sua replicazione. Lo studio della replicazione è stato effettuato grazie all'utilizzo di una tecnica innovativa, il *molecular combing*, che permette un'analisi molecolare ad alta risoluzione, in quanto le singole molecole di DNA vengono elongate in maniera uniforme su di un vetrino funzionalizzato. Per le molecole distese si ottiene la relazione 1 μm = 2 Kb, e questo permette di effettuare delle misurazioni molto accurate. Combinando l'ibridazione di sonde specifiche, che permettono di visualizzare le regioni di interesse, è possibile valutarne il *pattern* replicativo, rilevando mediante immunofluorescenza il DNA in replicazione: infatti le cellule sono state marcate con due *pulse* consecutivi di analoghi di nucleotidi.

La prima fase del progetto ha previsto la messa a punto del protocollo del *molecular combing* e delle varie procedure inerenti, in modo tale da poterla applicare efficacemente al nostro studio.

La seconda fase del progetto si è focalizzata sull'analisi del *pattern* replicativo del sito fragile comune *FRA6E*, in relazione ad altre due regioni: il ben caratterizzato sito fragile comune *FRA3B* e il *locus HPRT*, utilizzato come regione di controllo. Per quanto riguarda *FRA6E* mi sono concentrata sullo studio della replicazione

nelle due regioni di maggiore fragilità, dove mappano i geni *ARID1B* e *PARK2*, mentre per quanto riguarda *FRA3B* ho analizzato il *core* di fragilità, dove mappa il gene *FHIT*.

Per condurre le analisi ad alta risoluzione del *pattern* replicativo delle diverse regioni di interesse sono stati utilizzati linfociti primari, estratti da sangue di due diversi donatori.

Per ogni regione sono stati inoltre selezionati, attraverso analisi bioinformatiche, *set* di cloni genomici, da poter impiegare come sonde negli esperimenti di FISH su DNA elongato. Per descrivere le dinamiche di replicazione di ogni *locus* sono state determinati la velocità media delle forche replicative e gli eventuali eventi deregolativi (forche unidirezionali, eventi di arresto e asincronia nella velocità di progressione); è stata inoltre effettuata la mappatura delle origini di replicazione.

I dati ottenuti suggeriscono che le velocità medie delle forche replicative sono simili in tutte le regioni analizzate. Tuttavia in *FRA6E*, la velocità media è più alta nella regione dove mappa il gene *ARID1B* rispetto a quella dove mappa *PARK2*, pur essendo entrambe in accordo con i dati pubblicati a livello dell'intero genoma (Conti *et al.*, 2007). Inoltre, analizzando la distribuzione dei valori relativi alle velocità, rispetto a *PARK2* la regione di *ARID1B* mostra una minore variabilità: infatti, la maggior parte delle osservazioni si collocano in una sottoregione di 200 Kb. Questa regione ha tutte le caratteristiche tipiche di una sequenza a replicazione tardiva.

Nelle regioni di *FRA6E-PARK2* e di *FRA3B-FHIT* la frequenza delle forche unidirezionali è risultata più elevata rispetto a quella osservata nelle regioni di *FRA6E-ARID1B* e *HPRT*. Inoltre, sono state mappate le posizioni delle origini attive di replicazione nelle varie regioni e sono state stimate le relative distanze, le quali risultano essere in accordo con i dati già pubblicati (Conti *et al.*, 2007).

Per determinare come l'induzione dello stress replicativo influisca sulla replicazione a livello delle regioni fragili, sono state stimate la velocità delle forche replicative e la dimensione dei repliconi, in popolazioni cellulari di controllo ed in seguito al trattamento con afidicolina (APH), un inibitore della DNA polimerasi, valutando diverse dosi e diverse tempistiche. Anzitutto l'analisi è stata affrontata a livello dell'intero genoma, e quindi dei *loci* di mio interesse.

Dopo aver verificato che il trattamento con APH non arresta la progressione del ciclo cellulare, se non dopo 24 ore di trattamento con la dose più elevata, le analisi ad alta risoluzione hanno mostrato che, dopo due ore di trattamento, vi è un forte decremento della velocità di progressione delle forche replicative ed una conseguente diminuzione delle dimensioni dei repliconi..

Alla luce di questi dati ho quindi valutato la dinamica di replicazione all'interno delle diverse regioni di interesse, e in parallelo al *molecular combing* ho utilizzato anche tecniche di citogenetica molecolare classiche, quali la FISH su nuclei in interfase. Quest'ultima è stata importante per comprendere meglio quali siano le regioni a replicazione tardiva, e come atteso è emerso che i due siti fragili replicano più tardivamente delle regioni di controllo. Dall'analisi dei nuclei in interfase è emerso anche che il trattamento con afidicolina provoca un generale rallentamento della progressione della replicazione in tutti i *loci* considerati, tenuto conto delle differenti modalità di ciascuna regione.

Per quanto riguarda lo studio condotto tramite l'utilizzo del *molecular combing*, queste analisi si sono concentrate sulla regione di *FRA6E-PARK2* e del *locus* di controllo *HPRT*. I dati ottenuti mostrano una forte riduzione della velocità delle forche replicative, in accordo con i dati ottenuti a livello dell'intero genoma, e un aumento della frequenza delle forche unidirezionali in entrambe le regioni. Tuttavia, la risposta del sito fragile allo stress replicativo si manifesta soprattutto come un blocco molto evidente delle origini replicative, mentre nella regione di *HPRT* si assiste ad una aumentata attivazione di origini, come dimostrato dal fatto che la distanza stimata tra origini è significativamente inferiore a quella osservata nei linfociti di controllo.

Nel corso del progetto ho valutato la possibilità di poter studiare la replicazione delle regioni di mio interesse in popolazioni cellulari arricchite per la fase S del ciclo cellulare: a questo scopo ho effettuato degli esperimenti di separazione delle cellule per centrifugazione elutriale. In questo modo si evita l'utilizzo di sostanze chimiche, che possono interferire con l'espressione delle regioni fragili. Questi esperimenti sono stati condotti utilizzando linee cellulari linfoblastoidi, perché era già stato verificato che il metodo non è utilizzabile con linfociti primari. L'isolamento di sottopopolazioni arricchite in fase S è stato possibile e consentirà

di sviluppare ulteriormente le analisi sulle fasi tardive della replicazione del DNA, allo scopo di meglio comprendere la modalità con cui i siti fragili si replicano e rispondono allo stress replicativo.

Summary

Fragile sites are non-random chromosomal regions prone to breakage, which are expressed as decondensations, gaps or breaks under conditions in which DNA synthesis is partially inhibited. It has been hypothesised that they could represent uncondensed regions of the chromosome, for example due to unreplicated or single-stranded DNA. Moreover, there is some evidences that these *loci* are involved in chromosomal rearrangements related to tumours (Arlt *et al.*, 2003).

Depending on the frequency of expression in the population and on their inheritance pattern, these regions can be classified in two main categories: rare and common fragile sites.

Rare fragile sites occur in less than 5% of the human population and they are characterized by the presence of di- o trinucleotide repeats ($CGG_{(n)}$ e $AT_{(n)}$). The rare fragile site expression, segregating as mendelian traits, can be associated with pathological phenotypes. In most of the cases breakages at these sites are due to nucleotide repeat expansions, as occur in the fragile X syndrome. The major group of rare fragile sites is folate-sensitive *in vitro*, a vitamin involved in the nucleotides biosynthesis, and it can be hypothesized that alterations on the DNA replication can cause the fragility expression within these regions

Common fragile sites (CFS) are thought to be a normal feature of chromosomes and they show an high structural instability as a consequence of a replication stress. These sequences are seen in all individuals, although their expression vary and it is visible only in a subset of cells. CFS are characterised by the presence of AT-rich sequences and show several features of unstable DNA: high frequency of deletions, translocations, hotspots for sister chromatid exchanges (SCE). Moreover, these sites usually span along very large genomic regions and they contain large genes, presenting long intronic sequences (Glover *et al.*, 2005), as demonstrated for *FHIT*, *WWOX* and *PARK2* located in *FRA3B*. *FRA16D* and *FRA6E*, respectively. Some of them were found to function as tumour suppressor genes (Schwartz *et al.*, 2005) and this evidence supports the hypothesis that the genomic instability can play a role in cancer development.

It has been demonstrated that common fragile sites are late replicating regions, as reported for *FRA3B* (Le Beau *et al.*, 1998), and the presence of double strand breaks at these sites can appear as a consequence of stalled or collapsed replication forks.

In this study, analysis were focused on common fragile site *FRA6E*, one of the most frequently CFS expressed in humans. It is located on chromosome 6 (q26) in a susceptibility region frequently found rearranged in cancer. Many genes are located within *FRA6E* and some of them presented long intronic sequences, as *ARID1B* and *PARK2*. Concerning the last one, a putative role of oncosuppressor gene has been suggested. Previous data obtained in our laboratory indicated that *FRA6E* is a large genomic region of instability, spanning approximately 9 Mb (Russo *et al.*, 2006) and it can be sub-divided in three sub-regions, depending on the flexibility features: the central sub-region is less fragile than the centromeric one, where *ARID1B* gene is mapping, and than the telomeric one, where *PARK2* gene is located.

The aim of this study was to investigate if *FRA6E* instability could be related to replication defects at this site. Analysis were performed by using an innovative technique, the molecular combing, which allows the uniform elongation of single DNA molecules on silanised glass surfaces, leading to the defined equivalence of 1 μm = 2 kb. Combining FISH, with specific probes, and immunodetection of replicating DNA on single molecules, it is possible evaluate the replication pattern at specific regions. The first phase of the project consisted in setting up the molecular combing and other related experimental procedures.

The second part of the project was focused on the study of the replication pattern of *FRA6E*, compared to other two regions: the common fragile site *FRA3B* and the *HPRT locus*, as a control region. Concerning *FRA6E*, the two more fragile regions were considered, where *ARID1B* and *PARK2* genes are located, whereas in *FRA3B*, the most expressed common fragile site in humans, the core of fragility was analysed, where *FHIT* gene is mapping.

In order to evaluate the replication pattern at high resolution we used human primary lymphocytes, isolated from two different donors.

For each region, genomic clones were selected by bioinformatics analysis, to be employed as probes in FISH experiments on combed DNA. We focused the analyses in order to map replication origins, to determine fork density, replication rates and deregulative events (unidirectional forks, fork arrest events and asynchronous forks) .

Data collected showed that the mean fork rates seemed to be similar among the regions. In *FRA6E* the mean fork rates were higher in *ARID1B* than in *PARK2* region, even if they can not be considered significantly different when compared to published data relative to the whole genome (Conti *et al.*, 2007). Looking at the distribution of fork rate values, in *ARID1B* we noticed a lower variability than in *PARK2*, and notably forks have been detected on a 200 kb sub-region.

The frequency of unidirectional events was found to be more homogeneous and higher among *FRA6E-PARK2* and *FRA3B-FHIT* than *FRA6E-ARID1B* and *HPRT*. Also active origins were mapped and the inter-origin distances were evaluated, obtaining mean values in agreement with already published data (Conti *et al.*, 2007).

Moreover, in order to better understand in which way stress conditions can influence the replication process into the fragile regions, cells were treated with aphidicolin (APH), an inhibitor of the DNA polymerase, at different doses and times. Initially, we verified that the high concentration APH resulted in accumulation of cells in S phase after 24 h of treatment, whereas the cell cycle progression seemed not significantly modified at the other tested conditions.

Analyses at whole genome level performed by molecular combing indicated a strong delaying effect on the replication process after 2 h of treatment, with a strong decrease of the mean fork speed. We also analysed the replication pattern at *FRA3B-FHIT*, *FRA6E-PARK2* and *HPRT loci* by FISH on interphase nuclei of APH-treated and control cells. APH treatment was found to be effective in slowing replication process, with different extent with respect to the *locus* considered.

Single *locus* analysis on APH treated combed DNA, carried out to evaluate the replication pattern of *FRA6E-PARK2* and *HPRT locus*, showed that the mean fork speed in both regions strongly decreased, if compared to the controls, in

agreement with the data obtained at whole genome level. At both *loci* unidirectional forks were found to increase with respect to the controls. Also in this case the inter-origin distances ware evaluated and they were found to be differently affected from APH treatment in *HPRT* region than in *PARK2*. Infact, in *HPRT* the replication stress induced the activation of replication origins that seemed to be inactive in control condition, correlated to a reduced interorigin distance, whereas in *PARK2* region the effect of the APH treatment was a strong inactivation of the replication origins.

In order to study the replication of these regions specifically during the S phase, we performed the Centrifugal Elutriation, to obtain synchronous cell sub-populations without any chemical or drug treatments, which may interfere with the fragile site expression. These experiments were performed by using lymphoblastoid cell lines, because the method cannot be used with primary lymphocytes. The achievement of enriched S-phase cell sub-populations will be useful to extend the analyses to the replication process at common fragile sites during the late S-phase, in order to better understand the response to stress conditions.

1. Introduction

1.1. The fragile sites

The first evidence of a "chromosome weakness" was reported on 1965 by Dekaban, who found an high frequency of chromosomal abnormality in cells of a woman previously irradiated (Dekaban, 1965).

However, the original definition of "fragile site" was coined in 1970 by Magenis and colleagues to describe recurrent chromosome breaks on the long arm of chromosome 16, which presented a Mendelian segregation in a large family and showed linkage to the haptoglobin *locus* (Magenis *et al.*, 1970).

At the present, over 120 fragile sites have been identified all over human chromosomes and they are named according to the chromosome band they are observed in (Debacker and Kooy, 2007; Lukusa and Fryns, 2008).

Chromosomal fragile sites are heritable specific *loci* that preferentially show instability, visible as non-random gaps and breaks on metaphase chromosomes. These regions are normally stable in cultured cells, but they can be expressed under certain culture conditions or by treatment with specific chemical agents (Durkin and Glover, 2007). Importantly, the cytogenetically visible fragile sites may appear or may be not appear broken, therefore it has been proposed that they could represent uncondensed regions of the chromosome, for example due to unreplicated or single-stranded DNA (Freudenreich, 2007). Fragile sites are evolutionary conserved in primates and other mammalian species (Arlt *et al.*, 2003), even in the yeast *S. cerevisiae* (Lemoine *et al.*, 2005), and for some of them it has been demonstrated that orthologs are present in primates and mice, suggesting a functional role in genome reorganization (Debacker and Kooy, 2007). In fact, there is evidence that certain chromosomal regions in the human genome have been repeatedly used in the evolutionary process. As a consequence, the genome seems to be composite of fragile regions prone to reorganization, that have been conserved in different lineages, and genomic sequences that do not show the same evolutionary plasticity (Ruiz-Herrera *et al.*, 2006).

Depending on their population frequency and on their inheritance pattern, these regions can be classified in two main categories: rare and common fragile sites. Each class has been also further divided according to their specific mode of induction *in vitro*.

1.2. The rare fragile sites

The rare fragile sites are expressed in less than 5% of human population, they segregate in a Mendelian manner and in most of the cases breakages at these sites are due to nucleotide repeat expansions (Durkin and Glover, 2007).

They can be further subdivided, depending on the conditions in which they are induced in cells cultures.

The major group of rare fragile sites is folate-sensitive and their expression is induced when cells are cultured in folate-deficient media or in the presence of inhibitors of the folate metabolism. Many folate-sensitive fragile sites have been cloned: *FRAXA* (Xq27.3), located in the *FMR1*gene and associated with the fragile X syndrome (Verkerk *et al.*, 1991), *FRAXE* (Xq28), in the *FMR2*gene, related to non-specific mental retardation (Knight *et al.*, 1993), *FRAXF* (Xq28) (Parrish *et al.*, 1994), *FRA16A* (16p13.11) (Nancarrow *et al.*, 1994), *FRA11B* (11q23.3), associated to the Jacobsen syndrome (Jones *et al.*, 1995), *FRA10A* (10q23.3) (Sarafidou *et al.*, 2004), *FRA12A* (12q13.1), in the *DIP2B* gene, recently associated with mental retardation (Winnepenninckx *et al.*, 2007) and *FRA11A* (11q13.3) (Debacker *et al.*, 2007). All of them are caused by large expansions of CGG-repeat (more than 200 repeats in *FRAXA*), which give rise to hypermethylation of the sequence and of the surrounding CpG island, followed by the transcriptional silencing of the gene located at that site. Intermediate CGG-repeat expansions (from 50 to 200 in *FRAXA*) are called premutations.

The non-folate-sensitive fragile sites are distamycin A-sensitive or bromodeoxyuridine (BrdU) inducible and this group, at the moment, includes *FRA10B* (10q25.2) and *FRA16B* (16q22.1), which have been cloned. Both of them are caused by expansion of a AT-rich minisatellite repeats.

Rare fragile sites are either inherited from one of the parents, predominantly in a maternal way, or appear *de novo* and in the fragile X syndrome it has been first observed that younger generations have an higher risk to be affected, a phenomenon called anticipation. At molecular level, this occurrence can be explained by the unstable inheritance of repeat sequences, that in some cases exceed a threshold size, causing a remarkable instability across generations (Debacker and Kooy, 2007).

Not only CGG-repeat, but also AT-rich minisatellite expansions are able to form unusual secondary structures, such as hairpins, triplex and tetraplex structures, depending on the length of the repeat tract, which can stall the replication fork progression and perturb the whole replication process (Pearson *et al.*, 2005). Moreover, it has been also demonstrated that the elongation of CGG-repeat tract strongly decreases the efficiency of nucleosome assembly *in vitro* (Wang *et al.*, 1996) and this can result in decondensation defects at these sites, which cytogenetically manifest as gaps or breaks (Lukusa and Fryns, 2008).

1.3. The common fragile sites

Common fragile sites (CFSs) is the most represented class of fragile sites and up today about 90 CFSs have been recorded on human chromosomes (Smith *et al.*, 2007), with variable expression level. They can be considered a component of the normal chromosome structure, without any nucleotide repeat expansion. These sequences are seen in all individuals, but the proportion of cells with a cytogenetically expression varies from individual to individual, reaching in some cases levels of expression up to 30% (Sutherland and Richards, 1995).

Depending on their specific mode of induction, CFSs can be divided in three main classes:

1. Aphidicolin (APH) inducible fragile sites: APH is a specific inhibitor of the DNA polymerase and is the major inducing drug of common fragile sites. Their induction is achieved by low concentrations of APH, which

partially inhibit the replication process without arresting the cell cycle progression. The majority of common fragile sites belongs to this group, but the most APH-sensitive sites are represented by *FRA3B* (3p14.2), *FRAXB* (Xp21.1), *FRA16D* (16q23.2) and *FRA6E* (6q26);

2. Common BrdU-inducible fragile sites: at least seven CFSs are specifically induced by BrdU, after 6-12 h of exposure. They map at different genomic region than the BrdU-inducible rare fragile sites;
3. Common 5-azacytidin inducible fragile sites: the four CFSs *FRA1H* (1q42), *FRA1J* (1q12), *FRA9F* (9q12) and *FRA19A* (19q13) are induced by 5-azacytidin, which is an analogue of cytisine and is incorporated into the DNA in substitution of cytosine during the replication (Sutherland *et al.*, 1985).

At present, 13 CFSs have been cloned and characterized at molecular level: *FRA2G* (2q31), *FRA3B* (3p14.2), *FRA4F* (4q22), *FRA6E* (6q26), *FRA6F* (6q21), *FRA7E* (7q21.2), *FRA7G* (7q31.2), *FRA7H* (7q32.3), *FRA7I* (7q36), *FRA8C* (8q24.1), *FRA9E* (9q32), *FRA16D* (16q23.2) and *FRAXB* (Xp22.31) (Schwartz *et al.*, 2006). All are APH sensitive, with genomic instability and breakage occurring along a very large genomic region extending over at least 500 kb. Examples are represented by *FRA3B*, which was found to span approximately 4 Mb (Becker *et al.*, 2002), FRA6E, which was reported to span 9 Mb (Russo *et al.*, 2006), *FRA9E*, which was found to span 10 Mb (Callahan *et al.*, 2003), *FRA4F*, which is a 7 Mb region (Rozier *et al.*, 2004) and *FRA7E*, in which the fragile region is at least of 5 Mb (Zlotorynski *et al.*, 2004). Sequence analysis did not revealed any expanded repeats, as demonstrated for rare fragile sites. Analysis have been focused on DNA torsional flexibility at the twist angle between consecutive base pairs along the DNA molecule backbone. Some CFSs were found to be enriched in sequences and clusters of sequences with high DNA flexibility, termed flexibility peaks (Mishmar *et al.*, 1998), composed by interrupted AT-dinucleotide-rich sequences of various length, which can form secondary structures and perturb the replication (Zlotorynski *et al.*, 2003). Moreover, many fragile regions were found to map in R-band, which generally are more G/C rich, gene rich and with a DNA replication

occurring in the early S phase, but presenting G-bands characteristics, high A/T content, gene poor, high flexibility clusters and undergo DNA replication late in the S phase. It has been supposed that this incongruity might affect the control of DNA replication and of the chromatin condensation, contributing to their fragility (Mishmar *et al.*, 1999).

Common fragile sites are typically stable in human cell cultures, under normal growth conditions, but when the DNA synthesis is perturbed, they become visible as cytogenetic lesions, such as single or double chromatid gap or breaks, translocations, deletion breakpoints, intrachromosomal gene amplification. However, not all CFSs form breaks at the same frequency, a small number of them in the genome are significantly more sensitive to lesions formation.

Common fragile sites have been associated with hotspots for translocations, deletions, intrachromosomal gene amplifications, integration of exogenous DNA and other rearrangements (Lukusa and Fryns, 2008).

Moreover, there is evidence that these *loci* colocalize to chromosomal rearrangements observed in tumors: Yunis and Soreng (1984) first reported this observation and proposed the possible involvement of CFS in cancer development.

The fragile site-specific rearrangements most frequently observed are one or more deletions of tens to hundred of kilobases directly within the CFS region, resulting in inactivation of the associated genes (Durkin and Glover, 2007). Some of the large genes, located at these sites, have been characterized and they were found to function as a tumor suppressor, such as *FHIT* at *FRA3B* and *WWOX* at *FRA16D* (Smith *et al.*, 2007).

1.3.1. *FRA3B*

FRA3B (3p14.2) is the most frequently expressed CFS in the human genome (Becker *et al.*, 2002) The cloning and the characterization of *FRA3B* revealed that it was constituted of a large region of chromosomal instability (Paradee *et al.*, 1996). It was initially suggested that this CFS was spanning a small region of 300

kb, but it was later demonstrated that the fragile site region extends over 4 Mb (Becker *et al.*, 2002).

Within *FRA3B*, a very large gene, the fragile histidine triad (*FHIT*), was found to span 1.5 Mb. This gene is composed by 10 exons and two large introns, spanning several hundreds of kilobases, where *FRA3B* is centrally located; (Otha *et al.*, 1996; Zimonjic *et al.*, 1997). The *FHIT* protein function has been linked to intracellular signaling and the DNA damage response (Pekarsky *et al.*, 2004).

Initially, numerous studies have shown that the *FHIT* protein was absent or reduced in most cancers, including many cancers of the alimentary tract, from the oral cavity and oesophagus to the colon, and organs such as the pancreas and the liver(Huebner *et al.*, 1999; Croce *et al.*, 1999). Evidence that *FHIT* is a tumour-suppressor gene came from different experiments. Of 14 genes analysed in tumours derived from hybrid cells that carried an apparently normal chromosome 3, 34 tumours showed physical or functional inactivation of the *FHIT* gene (Kholodnyuk *et al.*, 2000). It has been also hypothesised that loss of one *FHIT* allele can lead to development of benign tumours, such as is observed in mice, whereas inactivation of the second *FHIT* gene might occur with progression to malignancy (Huebner and Croce, 2001). Moreover, it has been demonstrated that *FHIT*-deficient mice increased susceptibility to NMBA-induced gastric tumours, rescued by introduction of a functional *FHIT* allele (Zanesi *et al.*, 2001).

FRA3B, as most of the common fragile regions, is late replicating and exposure to APH causes a further delay in replication, resulting in a failure to complete replication. Interestingly, a study of its replication timing showed that one allele replicated later than the other, with or without APH treatment, and that fragility was preferentially observed on the late replicating allele (Le Beau *et al.*, 1998; Wang *et al.*, 1999).

FHIT and *FRA3B* region are also found to be highly evolutionarily conserved, in fact the mouse *Fhit* gene is located in a common fragile region, known as *Fra14A2* in mice (Glover *et al.*, 1998) and sequence comparison of these two fragile sites reveals highly conserved regions (Shiraishi *et al.*, 2001).

In addition to *FHIT*, there is another large gene mapping at *FRA3B*, the *PTPRG* gene (732 kb). This gene encodes for a member of the protein tyrosine

phosphatase family, which are known to be signalling molecules that regulate cell growth, differentiation, mitotic cycle and oncogene transformation (Kaplan *et al.*, 1990).

1.3.2. *FRA6E*

FRA6E is the one of the most frequently expressed common fragile site in the human population. A first published study describing the molecular mapping of *FRA6E* reported an extension of 3.3 Mb (Denison *et al.*, 2003). Data obtained in our laboratory indicated however that *FRA6E* is a more wide region of instability, spanning approximately 9 Mb at 6q25.1-6q26, and presenting more than one single core of fragility. Based on sequence analysis, the presence of two regions of maximum flexibility, separated by a more stable sequence was reported (Russo *et al.*, 2006).

Many genes are located within *FRA6E* and for some of them a role in cancer development and progression have been hypothesised. In fact, it has been observed that this common fragile site is located in a susceptibility region frequently found rearranged in multiple tumors, including cancer of the ovary, breast, kidney and lung (Shridhar *et al.*, 1999; Bando *et al.*, 1998; Thrash-Bingham *et al.*, 1995; Luk *et al.*, 2001).

Cesari and colleagues firstly described a large gene located at the telomeric boundary of *FRA6E*, *PARK2* gene, which extends over 1.3 Mb (Cesari *et al.*, 2003). This gene consists of 12 exons, spaced by very large intronic sequences with an average size of 125 kb, similarly to other genes located at CFSs regions (*FHIT* at *FRA3B* and *WWOX* at *FRA16D*). Mutation analysis revealed the absence of missense substitutions in this gene, but its expression was found to be down-regulated or absent in different cancerous biopsies and tumour derived cell lines (Cesari *et al.*, 2003; Wang *et al.*, 2004) and its over-expression suppresses tumour growth *in vitro* (Denison *et al.*, 2003), revealing the putative role in cancer development and progression. Moreover, *PARK2* gene was found to be a mutational target in some patients with autosomal recessive juvenile Parkinsonism (Kitada *et al.*, 1998).

Considering the more recent mapping of *FRA6E* extension (Russo *et al.*, 2006), some other genes have been revealed to belong to this fragile region and to be candidates for a role in cancer development. Among them are *ARID1B*, a subunit of the human ATP-dependent chromatin remodelling complex SWI/SNF, which is involved in transcriptional activation by hormone receptors (Hurlstone *et al.*, 2002; Inoue *et al.*, 2002); *VIL2*, which encodes the Ezrin protein found to be over-expressed in several epithelial cancers (Pang *et al.*, 2004); *WTAP*, supposed to play a role in both transcriptional and posttranscriptional regulation of certain cellular genes and *TULP4*, a putative transcription factor. Expression analysis in primary tumours and tumours derived cell lines have been carried out for four of these genes (*ARID1B*, *TULP4*, *VIL2* and *WTAP*), showing a reduced expression in three melanoma tumours, leading the hypothesis that particularly *ARID1B* and *TULP4* can represent the target for down-regulation in this cancer development (Russo *et al.*, 2006).

1.4. Common fragile site instability in cancer

Many studies have shown that CFS are sites of frequent chromosome breakage and rearrangements in cancer cells. In most of the cases, one or more large deletions of hundreds of kilobases and translocations have been observed, with the consequent inactivation of the associated genes. Several fragile sites, in fact, are near or within genes that may play a role in cancer, as previously described for *FRA3B* and *FRA6E* (Sections 1.3.1. and 1.3.2.). Similarly, the second more expressed common fragile site *FRA16D* lies within a large intron (260 kb) of the *WWOX* tumor suppressor gene (Ried *et al.*, 2000; Bednarek *et al.*, 2001). *Wwox* is a pro-apoptotic WW domain-containing oxidoreductase that binds to p53 (Chang *et al.*, 2001). The presence of deletions in the gene and aberrant transcripts have been observed in different cancer cell lines and tumor cells, including carcinomas of the breast, ovary, colon, lung and stomach (O'Keefe and Richards, 2006).

Expression of fragile sites has also been shown to generate gene amplification by breakage-fusion-bridge (BFB) cycles, which can lead to amplification of

oncogenes during tumor progression (Coquelle *et al.*, 1997). Studies both in yeast and in mammalian cells have shown that gene amplifications can occur when a break is located near an inverted repeat. In fact, a broken single-stranded end containing the repeat can fold and form a hairpin, which then can join to the other DNA strand (Tanaka *et al.*, 2002). Therefore, breakage caused by a fork stalling at a secondary structure or a repeat sequence could explain the link between fragile sites and gene amplification (Freudenreich, 2007).

Another characteristic of common fragile sites is that they are preferential viral integration sites. Analysis of human papillomavirus (HPV) integration sites revealed that over half of them were in common fragile regions (Yu *et al.*, 2005). As this virus causes the cervical cancer, it has been hypothesized that the presence of a fragile region can favors the tumor development.

Breakage at rare fragile sites has not been yet associated with cancer, but it is a mechanism generating repeat expansions, which are the causes of several genetic diseases, as the fragile X syndrome, the myotonic dystrophy and Huntington's disease (Dick *et al.*, 2006).

1.5. Cell cycle checkpoints and fragile site stability

Cells have evolved a number of cell cycle checkpoint pathways that are activated in response to specific DNA damages, resulting in cell cycle delay to lead the repair, or the apoptosis induction. For example, mismatch repair proteins prevent repeat sequences from undergoing sequence expansion or contraction (Kolodner and Marsischky 1999). Telomeric proteins prevent telomeres from fusion and from the breakage-fusion-bridge (BFB) cycle that can lead to chromosomal rearrangement (van Steensel *et al.* 1998; de Lange 2002, 2005).

The DNA damage and replication checkpoint proteins represent a class of regulatory proteins that are highly conserved and preserve genome-wide and site-specific stability (Zhou and Elledge 2000; Nyberg *et al.* 2002; Kolodner *et al.* 2002; Kastan and Bartek 2004).

During the S and G_2 phases of the cell cycle, when DNA is replicated and chromosomes prepared for mitosis, the ATM and ATR kinases act in parallel, as

principal DNA damage checkpoint proteins in overlapping pathways (Harrison and Haber, 2006). These proteins may preserve stability by regulation of cell cycle progression or of DNA replication, the regulation of dNTP levels, origin firing after damage and replication fork stability (Lukas *et al.* 2004).

The most important DNA damage checkpoint protein, Ataxia telangiectasia and Rad3-related kinase (ATR), it has been demonstrated to be crucial in the maintenance of stability at fragile site regions (Casper *et al.*, 2002). The ATR-regulated pathway is important in the response to anomalous replication intermediates and single-stranded DNA, generated during the replication stalling (Abraham, 2001; Zou and Elledge, 2003). It has been observed that ATR-deficient cell lines showed an increased chromosome instability, particularly at fragile sites (Casper *et al.*, 2002, 2004). On the contrary, ATM-deficient cells did not exhibit spontaneous or APH-induced CFS breaks. This fact suggests that DSBs, the DNA lesion to which ATM responds, are not the initial or primary cause of cytogenetic CFS expression. On the other hand, the ATM pathway may be important in the regulation of the following events at CFSs, particularly in the DSBs repair that have to occur in these regions to form chromosomal rearrangements. The DSBs could be originated, for example, from a collapsed replication fork or by the conversion of an unreplicated gap to a DSB by nucleases or mechanical forces.

Also the ATR targets, BRCA1, FANCD2 and SMC1 have been shown to be involved in the fragile site stability regulation (Arlt *et al.*, 2004; Howlett *et al.*, 2005; Musio *et al.*, 2005), as shown in Figure 1. In fact, cells deficient of CHK1, BRCA1, HUS1, SMC1, FANCD2 are highly susceptible to the increased occurrence of gaps or breaks at CFSs (Arlt et al, 2006; Zhu and Weiss, 2007).

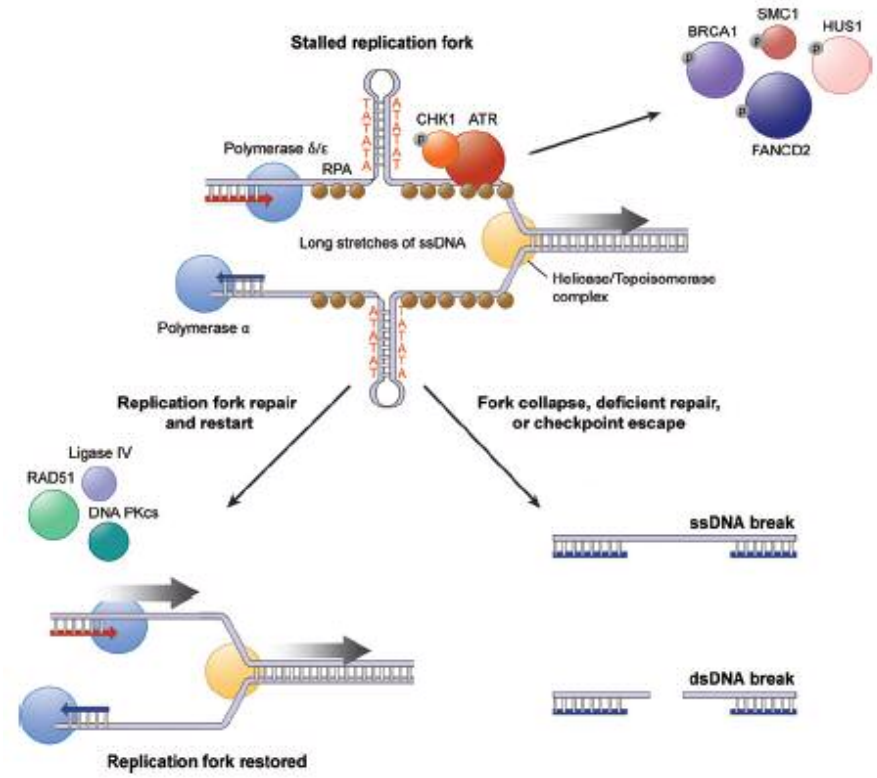


Figure 1. Model for common fragile site instability. This model predicts that fragile sites are derived from unreplicated ssDNA that are exposed when a replication fork is stalled or delayed, for example by treatment with APH. the DNA damage response checkpoint proteins are recruited, including ATR (*red*), which activate S-phase or G2/M checkpoints. Repair of these regions restores replication fork progression. However, sometimes these regions escape checkpoint activation or are left unrepaired, resulting in an unreplicated region that can appear as a fragile site on metaphase chromosomes or lead to a DSB (Adapted from Durkin and Glover, 2007).

ATR and its related DNA damage signal transducer, Ataxia telangiectasia mutated (ATM), regulate at least two major effector kinases: CHK1 and CHK2.

CHK1 is a checkpoint kinase, conserved in the evolution, that is the major target of ATR phosphorylation, and acts as a central regulator of cell cycle checkpoint delays in S- and G_2-phases (Liu *et al.*, 2000; Takai *et al.*, 2000; Bartek and Lukas, 2003). The CHK2 kinase is placed downstream of ATM, even if it can be activated by phosphorylation independently of ATM (Hirao *et al.*, 2002). CHK1 is activated in response of replication stalling, during the S and G_2/M cell cycle phases, in response to UV and hydroxyurea (HU), while CHK2 is activated by double strand breaks (DSBs), induced by ionizing radiations. Moreover, both kinases are phosphorylated and activated during S-phase in response to high dose

APH (Feijoo *et al.*, 2001). It was recently shown that CHK1 has a role in the stabilization of replication structures during S-phase in DT40 B-lymphoma cells (Zachos *et al.*, 2005). This evidence suggests that CHK1, and maybe CHK2, may respond to replication stress induced by low doses of APH, known to induce fragile site breaks, but not inhibit totally the replication progression, and that these proteins may regulate common fragile sites stability.

These two pathways can communicate and most DNA damaging agents can activate both: for example replication damage induced by topoisomerase inhibitors activates both ATM-CHK2 and ATR-CHK1 (Takemura *et al.*, 2006).

The intra-S phase checkpoint inhibits progression through S phase and initiation of later origins of replication (Petermann and Cadelcott, 2006). The signal for the intra-S phase checkpoint activation is created by stalled DNA polymerases associated with replicons, before the presence of the DNA damage.

MacDougall and colleagues recently showed that the single-stranded DNA generated is the signal that activates the checkpoint and that CHK1 is a component of this pathway. Once activated, the checkpoint stabilizes the forks that are stalled and prevents them from collapsing and undergoing repair by recombination, which can cause chromosomal rearrangements (MacDougall *et al.*, 2007). Moreover, recent evidence indicates that in mammalian cells the replication checkpoint includes also an "elongation checkpoint", defined as the ability of a cell to arrest the progression of normal forks after sensing DNA damage, but it can be restored by genetic inactivation of Chk1. (Seiler *et al.*, 2007). These two checkpoints stop replication not only at the damaged sites but also at distal ones from the DNA damage. This double mechanism allows the DNA repair at the damaged sites, and avoids further damage, by preventing normal replication forks from colliding into damaged DNA templates (Conti *et al.*, 2007b).

1.6. The replication dynamic and common fragile sites

1.6.1. The replication organization

A replicon is defined as a sequence of DNA that is replicated from a single origin and its size corresponds to the length of DNA replicated departing from the initiation site. The synthesis then proceed bidirectionally or unidirectionally, until ongoing replication forks from adjacent replicons merge and the replication terminates at random sites.

Nowadays, the application of new fluorescent methods, as the molecular combing, made possible to simultaneously visualize the DNA replication and specific DNA sequences on individual molecules to determine the location of replication origins. Moreover, through sequential pulse labels with two different halogenated nucleotides (IdU and CldU), it is possible to distinguish, with specific antibodies, the direction of the replication fork along the DNA molecule (Herrick and Bensimon, 1999).

Initial studies revealed that replication origins are stochastically and asynchronously activated at intervals of 5-20 kb and that the frequency of their activation increases during the S phase progression (Herrick *et al.*, 2000; Blow *et al.*, 2001; Marheineke and Hyrien, 2001), enlarging the probability of activate an origin in unreplicated regions.

During these years, many studies have been made to elucidate the regulation of origin density and activation. Results suggest that there is a well established spatio-temporal replication program, by which the excess of potential origins present across the genome are progressively redistributed in the course of the S-phase advance (Dimitrova and Gilbert, 1999). This dynamics is in agreement with other evidence that revealed the presence on initiation zones, in which replication can start anywhere within the region with the same probability (Dijkwel and Hamlin, 1995). In these regions, the origin which fires is stochastically selected and the fork progression can suppress origin firing of 1-2 flanking initiation zones, a mechanism called origin interference (Lebofsky *et al.*, 2006). This suppression may establish a hierarchical origin firing, which regulates the replication program.

Recent studies carried out on mammalian cells revealed that fork rates increased in proportion to the replicon size and that sister forks, in the same replicon, can change their rate progression simultaneously (Conti *et al.*, 2007). All these data suggest the existence of a homeostatic mechanism, by which frequency initiation and fork progression are strictly correlated. In fact, fork rates respond automatically to changes in origin firing frequency, which occurs during the S-phase, and origin density adjust to changes in fork rates (Herrick and Bensimon, 2008).

The presence of the homeostatic mechanism has been detected in numerous cell systems and associated to events of replication deregulation. In fact, when replication forks are stalling or slowing down, by the presence of replication fork inhibitors, new initiation sites are employed (Gilbert, 2007) and origins, usually inactive during the normal S-phase, can be activated (Anglana *et al.*, 2003), independently of the checkpoint response activation.

1.6.2. Late replicating regions and fragility

On 1987, Laird and colleagues have proposed a mechanism which correlates the common fragile sites expression with the replication process (Laird *et al.*, 1987). This model was then supported by further evidences proposing that APH-induced common fragile sites present unreplicated DNA, resulted from stalled replication forks (Wang *et al.*, 1999; Casper *et al.*, 2002). These regions replicate late during the S phase, also under normal culture conditions, indicating that their sequences have intrinsic features that may let to delay in replication.

APH induces a further delay in replication progression, that causes a significance portion of fragile regions to remain unreplicated in G_2 phase, as demonstrated for the common fragile site *FRA3B* (Le Beau *et al.*, 1998).

As for the rare fragile sites, the late replication of CFSs may represent the result of secondary structure formation, which inhibits replication fork progression or other factors affecting replication dynamics, such as organization and choice of replication origins (Durkin and Glover, 2007). Replication timing studies of the CFSs *FRA16D* and *FRA7H* highlighted that these sites may present difficulties in

replication fork progression (Hellman *et al.*, 2000; Palakodeti *et al.*, 2004). It has been demonstrated that in response to hydroxyurea (HU) treatment or DNA damage the intra-S phase checkpoint actively reduces the total rate of DNA replication, by inhibiting further initiation events and/or by slowing the progression of the replication forks (Seiler *et al.*, 2007). The contributions of the two mechanisms to replication reduction appears to vary from organism to organism and the choice of one of them can correlate with origins employ.

In fact, in higher eukaryotes and in fission yeast there is a large number of potential replication origins, but they are inefficient and only subsets are used by the cell in each cell cycle (Li *et al.*, 2003).

High-resolution assays of replication mechanics will be required to fully understand the details of replication origin firing and replication fork progression in CFS regions, because, as other regions in the genome are late replicating, this feature alone does not define a CFS. Nonetheless, the late-replicating DNA at CFSs is clearly particularly sensitive to further delay in response to replication inhibitors and it may play a role in CFS instability.

2. Material and Methods

2.1. Cell cultures

2.1.1. Cell lines

The following cell types were used in this study: the B-lymphoblastoid cell lines H691 and TK6; primary human peripheral T lymphocytes isolated from healthy donors. H691 are immortalised B-lymphoblasts derived in our laboratory from a healthy individual expressing at high frequency the common fragile site *FRA6E*. Cells have a normal chromosome set (46, XY). The TK6 cells is a well known and widely used cell line with chromosome number near to diploidy. Primary human peripheral lymphocytes have been extracted from different buffy coats, provided by the Hospital of Padova (Italy).

2.1.2. Cell culture procedures

All the cells were cultured in suspension in RPMI-1640 medium (Invitrogen, Italy) supplemented with Streptomcin/Penicillin mix (100 U/ml; Gibco, U. K.), 10% heat inactivated Fetal Bovine Serum (FBS, Invitrogen, Italy). Cells were grown at 37 °C and 5% CO_2. B-lymphoblastoid cell lines need to be sub-cultured three times a week to avoid confluence and the consequent growth arrest. For this purpose, cell concentration is assessed by an hemacytometer after Trypan blue staining. and counting. Then an aliquot of the cell suspension is diluted in fresh medium in a new bottle at the density of 0.3×10^6 cell /ml.

2.1.3. Primary human lymphocyte extraction from buffy coat

The buffy coat is the fraction of a blood sample which, after a density gradient centrifugation, contains most of white blood cells and platelets. The Padova Hospital (Italy) provided us with buffy coats from different healthy donors; the

lymphocyte subpopulations were extracted by the following steps. The blood was centrifuged 10 minutes, 1000 rpm and the superficial layer containing the platelets was removed. The blood was diluted by adding an equal volume of RPMI-1640 medium, then it was layered onto HISTOPAQUE 1077 (Sigma-Aldrich, Italy). HISTOPAQUE is an hydrophilic polysaccharide which allows separation of blood components (erythrocytes, lymphocytes, etc.). After centrifugation for 30 minutes at 1900 rpm, four different layers are visible: 1) the serum, 2) the lymphocyte ring, 3) the HISTOPAQUE 1077 and 4) the erythrocytes. Lymphocytes were collected, washed two times in PBS, pH 7.4 and centrifuged 10 minutes at 1900 rpm. Cells were resuspended in RPMI-1640 medium, and seeded in fresh medium at the concentration of 1 x 10^6/ml. In order to induce cell cycle entry, 1% phytohaemagglutinin (PHA, Remel, Diagnostic International Distribution, Italy) was added to the medium. Cultures were harvested after 42-72 h.

2.1.4. Cell synchronization

In order to obtain synchronous cell populations from heterogeneous cell cultures, we performed the Counterflow Centrifugal Elutriation (CCE). This method leads the isolation of cellular subpopulations on the basis of their sedimentation coefficient, which is a function of cell volume and density. Depending on their cycle phase and of their DNA content, cells can be fractioned and therefore synchronization does not need exposure to chemical agents. Moreover CCE permits the recovery of living cells, which can be re-cultured. The system requires a centrifugal rotor (JE-5.0, Beckman Coulter, U.S.A.), and pumps, which allow the formation of a flow through the chamber.

To obtain an efficient separation, a very large numbers of cells is requested. Cell cultures were expanded up to a total amount of 300-400 x 10^6 cells. Cells were centrifuged for 10 minutes at 1200 rpm, 4 °C, pellets were pooled together and resuspended in 8 ml of fresh medium, 10% FBS, under sterile conditions. Cells were injected into the centrifugal rotor and, while the pressure was slowly increased, cellular fractions were collected. All the fractions were then centrifuged 10 minutes at 1200 rpm, 4 °C, and resuspended appropriately to assess cell

number. Those fractions assumed to contain the cell cycle phase of interest were re-seeded in fresh medium at the concentration of 0.5 x 10^6/ml.

2.1.5. Flow cytometry: Flow-Activated Cell Sorting (FACS)

Flow cytometry (FACS) was applied to investigate cell proliferation. As it is known, this tool leads to discriminate, after excitation by a laser beam, different particles on the basis of the size and the fluorescent staining. In particular, the method enables to analyze different parameters: the cell size (FSC-H), the cell density (SSC-H), the cell fluorescence (FL1-H, FL2-H or FL3-H) as well as the area (FL2-A). The procedure is preceded by ethanol fixation and cell permeabilization. Finally, propidium iodide, a DNA intercalating dye detectable at 610 nm when excited at 540 nm, is used.

a. *Ethanol fixation:* cells were resuspended and counted. Then an aliquot containing 1.5-2 x 10^6 cells was centrifuged 10 minutes at 1200 rpm, 4 °C. The pellet was washed in PBS, pH 7.4 and then centrifuged again 10 minutes, at 1200 rpm. Cells were resuspended in 1 ml of PBS, pH 7.4 and 2 ml of cold absolute ethanol were added drop by drop. Cells can be stored in this solution at 4 °C before analysis.

b. *Propidium Iodide staining:* immediately before being analysed by FACscan (Beckton Dickinson, U.S.A.) preparations were centrifuged for 10 minutes at 1200 rpm, 4 °C. The pellet was then washed in distilled H_2O and centrifuged again. Finally cells were resuspended in 1 ml of a solution containing propidium iodide (50 µg/ml; Invitrogen, Italy) and RNase A (10 µg/ml; Sigma-Aldrich, Italy), and incubated for 1 hour at 37 °C in the dark. The results of the FACS analysis were elaborated by the Cell Quest software (Beckton Dickinson, U.S.A.).

2.1.6. Cytogenetic preparations

Cells were cultured in RPMI-1640, 10% FBS. During exponential cell growth, they were treated with aphidicolin (APH) 0.2 μM (Sigma-Aldrich, Italy) for 24 h or used as controls. In order to label nuclei in S-phase, cells were incubated with BrdU (Sigma-Aldrich, Italy) at a final concentration of 10 μM in the last 1 h and 30 minutes. At the same time, in order to increase the metaphase index, cells were exposed to Colcemid (Sigma-Aldrich, Italy) at the concentration of 0.1 μg/ml. After harvesting, cells were washed in PBS pH 7.4, resuspended drop by drop in 0.075 M KCl hypotonic solution and incubated for 15 minutes at 37 °C. A small volume of cold fixative (3:1 absolute ethanol:acetic acid) was added to the cell suspension, immediately before the next centrifugation step (10 minutes at 1200 rpm, 4 °C). Cells were resuspended in fresh fixative solution and incubated for 10 minutes on ice. This step was repeated three times. Finally, cells were resuspended in fresh fixative solution and they were ready for use or for prolonged storage at -20 °C. These preparations are equally useful for chromosome analysis or for interphase FISH.

2.2. The Molecular Combing technique

The molecular combing (Lebofsky *et al.*, 2006) was performed according to the protocol developed and set up from the research unit of Dr. Aaron Bensimon (Institute Pasteur, Paris, France) with modifications. The multiple step procedure will be fully described in the next sections.

2.2.1. Cell labelling and inclusion in agarose plugs

Exponentially growing cells were sequentially labelled with the halogenated nucleosides 5-iodo-2′-deoxyuridine (IdU) and 5-chloro-2′-deoxyuridine (CldU), both provided by Sigma-Aldrich (Italy). IdU at the final concentration of 100 μM was added, the culture was mixed very gently and incubated for 30 minutes at 37 °C. Then, without any washing step, CldU at the final concentration of 100 μM

was added and cells incubated for further 30 minutes. The cellular suspension was harvested, thoroughly resuspended and centrifuged for 10 minutes at 1500 rpm, 4 °C. After removing the supernatant, cells were washed in cold PBS pH 7.4, counted and resuspended appropriately in order to prepare agarose plugs (final cell number should be 100-200.000 cells *per* plug). Samples were incubated 5 minutes at 37 °C and an equal volume of molten LMP agarose (Sigma-Aldrich, Italy), 1.5% (w/v) in PBS pH 7.4, was added. After fast vortexing, 90 µl aliquots were immediately distributed into plug molds (BioRad, U.S.A.). To allow agarose solidification, molds were maintained at 4 °C for at least 1 h. The agarose blocks were then extracted from molds and incubated overnight at 53 °C in Proteinase K digestion solution, consisting in 1% N-lauroylsarcosine (Sigma-Aldrich, Italy); EDTA 0.5 M, pH 8.0; 2 mg/ml Proteinase K (Promega, Italy). The following day the plugs were washed twice, 1 h each, in TE buffer. Agarose plugs can be stored at 4 °C in 10 ml of EDTA 0.5 M, pH 8.0.

2.2.2. DNA extraction

Before DNA extraction, the agarose plugs were washed two or three times in TE buffer for 1 h, under slow speed rotation. Then, plugs were placed in 2 ml of MES 0.1 M, pH 6.5 and incubated for 30 min at 72 °C, to melt the agarose matrix. After a temperature equilibration step (20-30 minutes at 42 °C), 2 µl of β-Agarase I (BioLabs, U.S.A.) were added and the sample was incubated overnight at 42 °C. The following day, the sample was incubated 4 h at 50 °C, then left at room temperature for one week. This step favours unbending of high molecular weight genomic DNA. Finally DNA solution was gently poured into a teflon *reservoir* (height 22 mm, basis 4 x 30 mm), where the sample can be stored at room temperature.

2.2.3. Silanisation of glass surfaces

Starting from the protocol developed by Bensimon A. and colleagues (Lebofsky *et al.*, 2006), we applied the silanisation procedure in collaboration with Prof. Vito

Di Noto (Department of Chemical Sciences, University of Padova, Padova, Italy), who provided us with precious instruments and advices.
The procedure is composed by seven steps:

1) *The pre-Piraña washes*: perfectly clean high quality glass coverslips (22×22 mm, 0.13-0.16 mm thick, Menzel-Glaser, Germany), are placed on teflon racks, washed three times with MilliQ water (5 minutes each), then sonicated sequentially in absolute ethanol and in chloroform (50 seconds *per* sonication step). Coverslips are washed and sonicated in absolute ethanol three times, 5 minutes each;
2) *The Piraña solution*: racks with coverslips are transferred in the Piraña solution (NH_4OH, H_2O_2 33%, Sigma-Aldrich, Italy) and heated at 60 °C for at least 1 h and 30 minutes and then left at room temperature;
3) *The post-Piraña washes*: after five washes with MilliQ water (5 minutes each) coverslips are sonicated in 6% HCl for 75 seconds and then rinsed five times in MilliQ water;
4) *The pre-silanisation washes*: racks with coverslips are sonicated sequentially, 50 seconds each, in: absolute ethanol, absolute ethanol/toluene (1:1), toluene. Then, the glass surfaces are dried into a chemical hood;
5) *The silanisation process*: racks containing the coverslips are moved to a dry box, in the presence of dehydrated argon atmosphere. 300 µl of 7-octenyltrichlorosilane (Sigma-Aldrich, Italy) are injected in a small glass plate, placed near the coverslips. The silane is let evaporate overnight;
6) *The post-silanisation washes*: after racks have been taken out from the box, they are sequentially sonicated in toluene, in absolute ethanol/toluene (1:1) and in absolute ethanol for 50 seconds each. Finally coverslips are dried into the chemical hood.

2.2.4. DNA combing

After extraction, DNA must be tested for its quality (integrity, density, unbending of molecules). Silanised coverslips were placed in an appropriate coverslip holder of the combing apparatus, shown in Figure 2., (Genomic Vision, France) and dipped into the DNA solution placed in the teflon *reservoir*. After a 5 minute incubation, during which the binding through one or both extremities of DNA molecules occurs, the coverslips were removed, at a constant speed of 300 µm/s, automatically by the machinery. The coverslips with the irreversibly fixed molecules were stained 30 seconds with the fluorescent intercalating dye YOYO-1 (Molecular Probes, Invitrogen, Italy) diluted at 1 µM solution in MES 0.5 M, pH 5.5. The coverslip was then glued onto a glass slide with cyanoacrylate glue (Super Attak®, Henkel, Germany), mounted with a drop of Vectashield antifading medium (Vector, U.S.A.) and a non-silanised coverslip. The density, integrity and linearity of combed DNA molecules was checked by fluorescence microscopy (Axio Imager.M1, Zeiss, Germany).

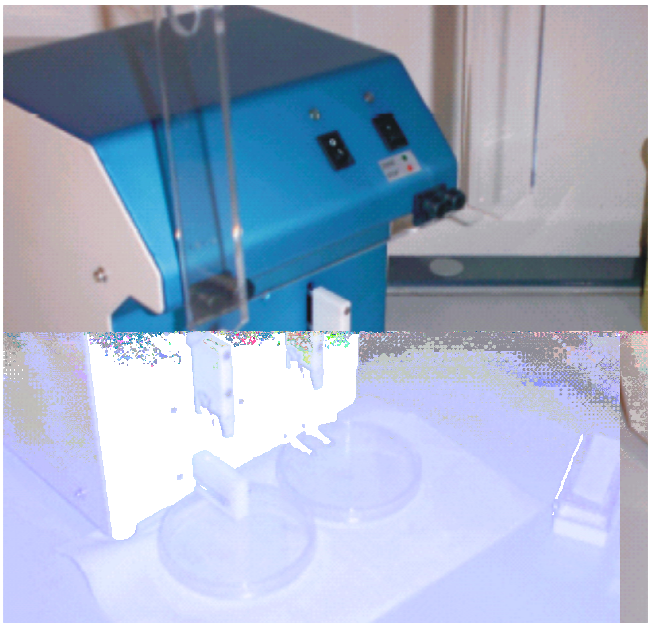


Figure 2. The molecular combing apparatus (Genomic Vision, France).

2.3. Analysis on combed DNA

After combing, the DNA is covalently bound to the silanised surface and it is the target for further cytogenetic analyses: coverslips are glued on glass slides and baked for 4 h at 60 °C (after this step they can be used immediately or stored at -20 °C).

Specific probes labelled with modified nucleotides can be hybridised to the combed genomic DNA. The hybridised regions are detected with several layers of affinity molecules and specific antibodies coupled to different fluorochromes; they are visualised as linear fluorescent signals by a fluorescence microscope equipped with a high resolution CCD camera. Additionally, the occurrence of replication along specific regions of interest can be studied, by detecting labelled probes together with the replication signals.

2.3.1. Genomic clone selection

Different sets of genomic clones were chosen to cover about 1 Mb for each region under analysis. Genomic clones have been selected after bioinformatic analyses carried out using the Human Genome Browser Ensembl (*www.ensembl.org*). The rationale for selecting the clones is to obtain a characteristic hybridization motif with three probes with different sizes and distances. These constraints are useful to detect the molecule of interest and to determine its orientation (telomere-centromere or centromere-telomere) by the univocal identification of the probes. This can be applied also in the presence of a partial hybridization pattern, as it occurs when the integrity of the molecule on the surface is not preserved.

All the different sets of genomic clones selected for each region under analysis are reported in Supplementary Data, Table I.

2.3.2. BAC and PAC DNA preparation

After the genomic clones have been selected, they were obtained by the *Children's Hospital Oakland Research Institute (CHORI)*, U.S.A. These clones are BAC

(Bacterial Artificial Chromosome) or PAC (P1 derived Artificial Chromosome) transformed in DH10 *E. coli* cells; in the inserted sequence a specific antibiotic resistance marker is present. After their arrival, bacteria were seeded on LB-agar solid plates supplemented with chloramphenicol (170 μg/ml), if BACs, or kanamycin (50 μg/ml), if PACs. Bacterial cells were incubated overnight at 37 °C. For each clone, single isolated colonies were picked up, seeded in LB liquid medium and incubated overnight at 37 °C under shaking. A fraction of the bacterial suspension was centrifuged 5 minutes at 6000 rpm, resuspended in equal volumes of glycerol/LB medium, and stored at -80 °C. The remaining fraction was employed for the DNA extraction.

The cell suspension was centrifuged 5 minutes at 6000 rpm and the pellet resuspended in the Solution I (glucose 50 mM, EDTA 10 mM and Tris-HCl 25 mM), supplemented with 4 mg/ml lysozime, an enzyme which digests the bacterial external wall. Then the Solution II (NaOH 0.2 N, SDS 1%) was added and samples were incubated for 4 minutes at room temperature, to allow the lysis of the cellular membrane. To stop the reaction, Solution III (CH_3COOK 3 M and CH_3COOH 11.5%), previously chilled, was added and the mix incubated for 5 minutes on ice. The suspension was centrifuged for 20 minutes at 14000 rpm, 4 °C, and DNA was extracted by the standard phenol:chlorophorm:isoamyl alcohol protocol. The pellet was finally resuspended in a proper volume of sterile TE buffer, and treated with RNAse A (Sigma-Aldrich, Italy) (40 μg/ml, 30 minutes at 37 °C). An aliquot of the DNA sample was checked for the quality by running it on 0.8% agarose gel.

2.3.3. Random priming

DNA clones are labelled with modified nucleotides in order to be employed as probes during the hybridization procedure. The random priming technique was performed employing the BioPrime DNA Labelling System (Invitrogen, Italy).

Briefly, 250-300 ng of double strand template DNA were mixed with octameric random primers at the appropriate concentration, denatured and incubated at 37 °C overnight with the Klenow fragment (40U) in the presence of a dNTPs mix,

containing the labelled nucleotide (Biotin-14-dCTP, provided by the manufacturer or Digoxygenin-11-dUTP, Roche, Germany). At these conditions, random incorporation of labelled nucleotides takes place; amplified fragments in a range of 100-1000 bp are produced. The following day, the reaction was stopped on ice and an aliquot of labelled probe was run on 1% agarose gel, to check if the reaction was successful and to determine the concentration of the labelled product. A direct control of the labelling quality (e.g. dot blot) was not performed.
Probes were stored at -20 °C until used.

2.3.4. Probe precipitation

Per each slide, 500 ng of each of three labelled probes mixed with a proper excess of Cot-1 DNA (range of 13X-17X; Invitrogen, Italy), 1 µg of Salmon sperm DNA and 20 µg of glycogen are precipitated by Na acetate/ethanol method. The resulting pellet is let dry, then resuspended in 20 µl of Hybridization Buffer (formamide 50%, SSC 2X, SDS 0.5%, N-lauroylsarcosine 1%, NaCl 10 mM and Block Aid 1X, Molecular Probes, Italy), which was pre-warmed at 37 °C. The solution was maintained for at least 30 minutes at 37 °C before denaturation.

2.3.5. Denaturation

Combed DNA was denatured by placing slides in NaOH 50 mM, NaCl 1 M for 15 minutes at room temperature, under gently shaking. Slides were rinsed in cold Tris-HCl 0.01 M, pH 7.6, dehydrated by sequential washes in 70%, 90%, and 100% ethanol (3 minutes each), air-dried at room temperature.
Probes were denatured at 80 °C, 10 minutes in a water bath, then maintained on ice to avoid the DNA renaturation.

2.3.6. Hybridization

20 µl *per* slide, containing 500 ng of each of the three probes, were loaded onto the slide carrying the denatured combed DNA, and covered with a 22x22 mm

coverslip. Slides were sealed with silicon glue and incubated overnight at 37 °C in a humidified box.

2.3.7. Stringency washes

The following day, samples were washed three times (5 minutes each) with formamide 50%/SSC 2X at room temperature and three times in SSC 2X (5 minutes each, room temperature), to remove aspecifically hybridized probes.

2.3.8. Signal detection

The affinity molecules used in three-probe, single-colour FISH and the panel of antibodies employed for FISH and replication detection (three fluorescence wavelengths) are listed in Table 1.

	I° Amplification Step	**II° Amplification Step**	**III° Amplification Step**
Probe Detection	SAV-488 (1:50)	Anti-SAV Biot (rabbit) (1:50)	SAV-488 (1:75)
IdU Detection	Anti-BrdU (mouse) (2:7)	Anti-mouse IgG-350 (goat) (1:50)	Anti-goat IgG-350 (donkey) (1:50)
CldU Detection	Anti-BrdU (rat) (1:40)	Anti-rat IgG-594 (donkey) (1:50)	

Table 1. The affinity molecules, used in three-probe single-colour FISH, and the antibodies employed for FISH and replication detection are reported. In brackets, the relative dilutions are indicated.

In both cases, biotin-labelled probes are detected by the Streptavidin (Molecular Probes, Invitrogen, Italy), an affinity molecule conjugated with a fluorochrome (488). By using a secondary specific biotinilated antibody, the fluorescence signal can be amplified with further affinity layers. The replication detection is performed by using two primary antibodies made to recognize the BrdU, but specifically cross-reacting with IdU and CldU, respectively. Also in this case,

fluorescence signals are amplified and visualized, by blue (350) and red (594) fluorochromes conjugated to secondary antibodies. The reaction mix was prepared immediately before each step; affinity molecules and antibodies were diluted in a blocking solution (Block Aid, Invitrogen, Italy). 30 µl were deposited on each slide, covered with a 22x22 mm coverslip, and incubated for 30 minutes at 37 °C in a humidified box. Then, the coverslip was gently removed and the slide washed three times (3 minutes each) with PBS, pH 7.4 under slow agitation. The following incubation steps were performed at the same conditions. Finally, slides were mounted in Vectashield medium (Vector, U.S.A.).

2.3.9. Immunodetection of replication

Molecular combing offers the possibility to visualize and study replication pattern on the DNA single molecule. For this purpose, growing cells are first differentially labelled with two pulses of halogenated nucleosides (IdU and CldU). Consequently, replicating DNA can be visualized by immunofluorescence techniques (Herrick *et al.*, 2000).

After the slide preparation (section 2.4.4), the DNA was denaturated and dehydrated, and the signal detection carried out as described previously (section 2.3.8.). The list of antibodies are reported in Table 2.

	I° Amplification Step	**II° Amplification Step**
IdU Detection	Anti-BrdU (mouse) (2:7)	Anti-mouse IgG-488 (goat) (1:50)
CldU Detection	Anti-BrdU (rat) (1:40)	Anti-rat IgG-594 (donkey) (1:50)

Table 2. Antibodies employed for replication detection are reported. In brackets, the relative dilutions are indicated.

Assuming that replication forks depart from the origin in a bidirectional way at the same rate, three types of signals are expected, as shown in Figure 3. In addition, a continuous red signal flanked by two green ones can be observed when merging of two forks from adjacent origins occurred (as shown by the star in Figure 3).

Therefore, replication origins can be mapped; inter-origin distances can be evaluated. The total space covered by the two arms of the proceeding fork emanating from the same origin is defined as replicon size.

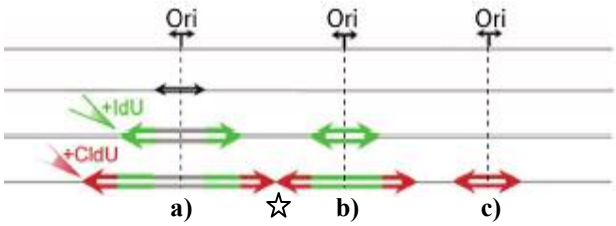


Figure 33. The three types of replication signals are represented: a) a bidirectional replication fork, with green and red segments, and a gap between the green ones, corresponding to a replication origin firing before the onset of the first pulse; b) a bicolour signal with a continuous green segment, corresponding to a replication origin firing during the first pulse; c) a single isolated red signal, which corresponds to origin firing during the second pulse. The merge of two ongoing forks from adjacent origins is indicated by the star (Adapted from Conti *et al.*, 2007).

Because in these analyses the combed genomic DNA is not counterstained, only a fraction of the total amount of the replication forks can be considered informative, as their pattern certainly corresponds to whole, uninterrupted molecules. In particular, only complete bidirectional forks were considered, observed either isolated or positioned on the same single molecule. Forks presenting only one arm were classified as unidirectional, only if more than one fork could be observed on the same DNA molecule. These constraints were adopted also for the classification of deregulation events. Different patterns of fork arrest events can be observed (Figure 4); forks showing a non-coordinated progression between the two arms were classified as asynchronous forks (Figure 4).

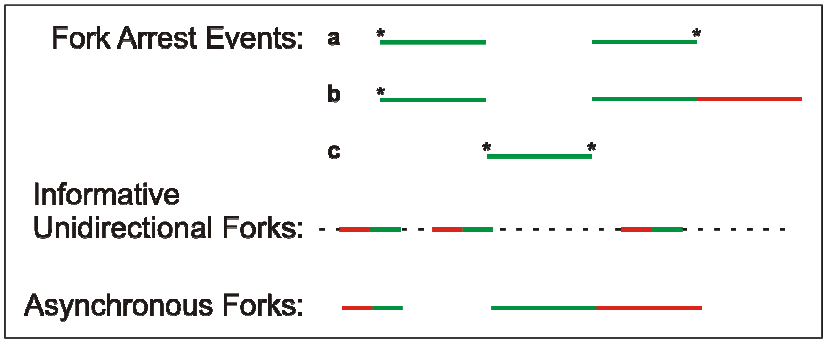


Figure 4. Stalled forks can be identified as in the three examples classified as fork arrest events (a) only two green signals can be detected; b) a bidirectional fork lacking of one red signal; c) a short

isolated green signal. The integrity of the molecule must in addition be demonstrated by upstream/downstream fluorescent signals. The asterisk indicates the fork arrest. Informative unidirectional forks are replication forks presenting only one arm and observed on the same DNA molecule. Asynchronous forks present the two arms running at different speeds.

In order to map origin positions, the origins observed within each significant informative molecule, at each *locus*, were reported onto a unique scheme, reproducing position and length of observed probes, and of the ongoing replication forks. Alternative classifications were reported, when additional hypotheses were possible.

2.4. Fluorescence *in situ* hybridization (FISH) on interphase nuclei and on metaphase chromosomes

Fluorescence *in situ* hybridization (FISH) is the canonical cytogenetic technique generally employed to study changes in chromosomal structure and number, either on metaphase chromosomes or in interphase nuclei. In this thesis, interphase FISH was used to investigate modification on the replication pattern timing. Chromosome analysis was applied to confirm the localization of probes at fragile sites of interest. Protocols are well established and therefore they will be described only briefly, in order to highlight the procedural differences with respect to the method used for combed DNA.

2.4.1. Probe labelling: the nick translation

Nick translation is the approach of election for probe labelling in conventional FISH method, because the small size of fragments deriving from this reaction (200-500 bp) is the optimal one when the DNA target is organised as chromatin. In contrast to random priming, this labelling procedure does not allow the amplification of the initial probe amount.

Nick translation was performed by using the Nick Translation Mix kit (Roche Biochemicals, Germany). Briefly, 1 μg of template DNA is mixed with a dNTPs cocktail (dATP, dCTP and dGTP 0.25 mM, dTTP 0.17 mM, Dig-11-dUTP or Bio-

16-dUTP 0.08 mM) and with the nick translation mix, which contains the two enzymes DNA polymerase I and DNAse I. The reaction was performed by incubating the sample 90 minutes at 16 °C, then stopped with 1 µl of EDTA 0.5 M, pH 8. The success of the labelling procedure can be assumed if the correct range of fragments is obtained. Therefore, an aliquot of reaction volume was checked on 1% agarose gel in TAE buffer. A direct control of the labelling quality (e.g. dot blot) was not performed.

2.4.2. Procedures

Glass slides must be carefully washed and kept in ice-cold water before preparing the cells or chromosome spreads. Few drops of cell suspension, prepared as described in section 2.1.6., were spotted on the wet, ice-cold slide from 30-40 cm of distance. Excess liquid was drained and the slide was placed on a 60 °C hot plate. Chromosome spreading and/or cellular density were checked. Before samples were used for hybridization, they were aged few days at room temperature.

Pre-hybridization treatments consisted in slide incubation with RNAse A (0.2 mg/ml).

100 ng of probe *per* slide were mixed with a proper excess of Cot-1 DNA (47X), Salmon sperm DNA (4.7X) 20 µg of glycogen, and were precipitated by Na acetate/ethanol method. The resulting pellet was let dry and resuspended in 10 µl/slide of Hybridization Mix (formamide 50%, SSC 2X, Dextran Sulphate 10%), which was pre-warmed at 40 °C.

Probes were denatured 10 minutes at 70 °C, then left at least 90 minutes in a pre-annealing condition (37 °C), during which the saturation of the repetitive sequences is favoured. The denaturation of chromosomes and nuclei, instead, was performed by placing the samples into a solution containing formamide 70%/SSC 2X, pH 7.0 for 4 minutes at 72 °C, which, decreasing the DNA denaturising temperature, preserves the chromosomal structure.

Samples were dehydrated by sequential washes (3 minutes each) in 70%, 90%, and 100% ethanol and then air-dried at room temperature.

10 µl *per* slide, with 100 ng of each of probe, were loaded onto the slide, and covered with a 22x22 mm coverslip. Slides were sealed with silicon glue and incubated overnight at 37 °C in a humidified box.

For stringency washes, samples were incubated three times with formamide 50%/SSC 2X solution at 42 °C (5 minutes each) and then with SSC 2X (5 minutes each, room temperature) to remove aspecifically hybridized probes.

Before the detection steps, samples were incubated with a Blocking Solution 30 minutes at 37 °C in a humidified box, in order to avoid unspecific antibody binding.

Also in this case, the detection of the fluorescent signal has to be amplified, in order to obtain more evident signal (Table 3).

	I° Amplification Steps	**II° Amplification Steps**	**III° Amplification Step**
Probe Detection (Biotin labelled)	SAV-594 (1:100)	Anti-SAV biot (rabbit) (1:100)	SAV-594 (1:100)
Probe Detection (Digoxygenin labelled)	Anti-Dig (mouse) (1:25)	Anti-mouse IgG-488 (goat) (1:100)	

Table 3. Affinity molecules and antibodies employed for FISH detection. In brackets, the relative dilutions are indicated.

30 µl were deposited on each slide, covered with a 22x22 mm coverslip, and incubated for 45 minutes at 37 °C in a humidified box. Then, the coverslip was gently removed and the slide washed three times (3 minutes each) with PBS pH 7.4, supplemented with the *Tween 20* 0.1% (Sigma-Aldrich, Italy) under slow agitation. The following incubation steps were performed at the same conditions. Finally, slides were mounted in 20 µl of Vectashield medium (Vector, U.S.A.) containing DAPI (4',6-diamidino-2-phenylindole) 2 µg/ml, a DNA intercalating dye detectable at 460 nm when excited at 358 nm.

2.5. Image analyses

A motorized fluorescence microscope Axio Imager.M1 (Zeiss, Germany), equipped with short pass filters (Omega Optical, U.S.A.) for the green (Green BP 450-490), red (RED TBP 400-495-570) and blue lights was used. Analysis of multicolour FISH images required isolation of the single fluorescent signals without any registration of shift. A fast, high resolution and ultra-low-noise CCD (Charge Coupled Device) monochromatic camera Coolsnap HQ^2 (Photometrics, Crisel Instruments s.r.l., Italy) was used to acquire digital images with the software MetaMorph (Version 7.1.3.0, Molecular Devices Corporation Analytical Technologies, U.S.A.). The objectives employed and the analysis procedures depended from the specimen evaluated:

a. For what it concerned the analyses on combed DNA a 40X oil immersion objective (N.A. = 1.30) was used. For each area of interest separate images, taken at necessary wavelengths, are recorded. To assure that the DNA molecule can be analysed along the whole area of interest, adjacent fields were recorded with the same approach. The acquired images were merged immediately and saved for further analysis. Adjacent images were overlapped with the aid of Adobe Photoshop software to reconstruct the DNA molecules. Finally, fluorescent signals (probes and replication forks) were measured by using the Metavue software. Thanks to the molecular combing calibrating factor (1 μm = 2 kb) and according to the magnification features of the objective and the CCD camera, 1 pixel = 0.16125 μm (0.3225 kb).
b. FISH signals on metaphase chromosomes were initially scored at low magnification (20X objectives), then the image acquisition was performed by using a 100X oil immersion objective (N.A. = 1.30). For interphase nuclei the analysis was carried out at 100X magnification. Images were captured and merged as described above.

3. Results

3.1. Preliminary experiments

The first phase of the project consisted in setting up the molecular combing and other related experimental procedures.

3.1.1. Setting up the protocol for Molecular Combing

Knowledge and skill on the molecular combing procedure was gained (by the supervisor of the thesis) at the laboratory directed by Dr. A. Bensimon at the Institute Pasteur of Paris (France). Some aspects of the procedure required however further improvement for the development of the present project: the silanisation of coverlips, the definition of the optimal conditions by which our cells are included in agarose plugs, the protocol for putting DNA into solution, the FISH protocol.

In the molecular combing procedure, the binding of DNA molecules onto the glass surface is very specific and it takes place because of precise chemical conditions on the glass surface, which is covered with silane, exposing vinylic (−*CH*=*CH*2) end groups. Working with the 7-octenyltrichlorosilane is very critical because of its reactivity with H_2O molecules and it is necessary to handle it under dehydrated controlled atmosphere. Home made silanised surfaces were prepared for this project and it was verified that the elongation properties were the same as previously published (Herrick and Bensimon, 1999).

To preserve as possible the integrity of the molecules, DNA must be put into solution without shaking or turning movements. This step appeared to be very critical: indeed, during the first phase of the project, a great variability was observed as far as the quality of the DNA solutions was concerned. In most cases, the solutions appeared not eligible to be used for combing procedure (Figure 5, panels A and B).

I decided, therefore, to determine the weight of the several factors contributing to the final quality of the DNA solution.

3.1.1.1. The number of cells *per* agarose plug

It was reasonable to retain the cell number *per* plug as an important parameter. When the plug was prepared with small amount of cells, the low density of DNA released after agarose digestion resulted not only in non analysable preparations, with rare fluorescent signals; also, many combed molecules appeared broken (Figure 5, panel B). On the other hand, when the starting cell number was too high, the DNA molecules could not be correctly combed, because they formed bundles, super-coiled fibres, or non-linear molecules (Figure 5, panel A).

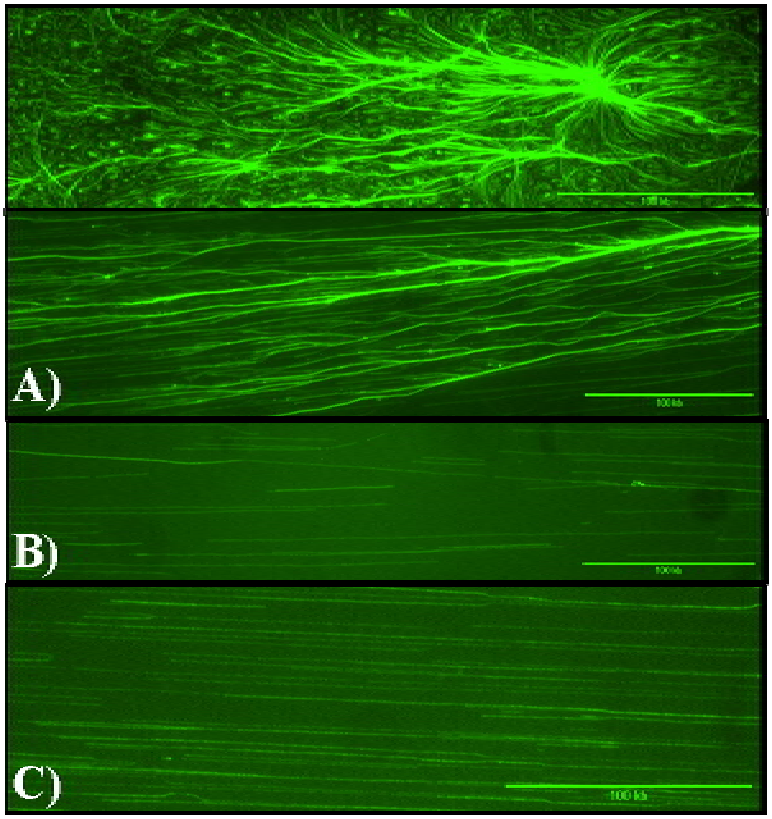


Figure 5. Combed DNA molecules, deriving from plugs containing different number of cells. A) Two examples of combed DNA molecules deriving from plugs with a cell number too high, where bundles, super-coiled fibres, or non-linear molecules can be visualised; B) As a result of a small amount of cells *per* plug the density and integrity of combed DNA molecules is not suitable for further analysis,; C) Combed DNA molecules well dissolved in the final solution and perfectly elongated. Calibration bar = 100 kb.

Clearly, the correct number of cells depends also on the genome size and the ploidy number, and therefore is a function of the cell type used. Looking at the optimal number of cells to be used in my experiments, I observed however that

additional parameters were crucial; for example the cell density originally used to form the plugs, and not simply the number of genomes put into solution seemed to be important. In other words, if X is the ideal number of genomes necessary to obtain good quality preparations, plugs containing 3X cells will not be useful even if 1/3 of the plug is used to prepare and comb DNA. According to this observation, it appeared that the original arrangement of the genomes entrapped in the agarose plug may play a role with respect to their ability to be perfectly dissolved in the final solution (Figure 5, panel C).

3.1.1.2. Hypotonic solution treatment

This treatment, consisting in a short incubation in Na citrate 0.9% (5 minutes at 37 °C), was introduced as a modification of the protocol provided by Bensimon's laboratory, because we aimed to analyse cells growing in suspension, which tend to form aggregates even after harvesting. The presence of cell clumps not only affects the accuracy of cell counting, but also can affect the preparation of agarose plugs, which should include homogeneous numbers of cells. I noticed however some general improvement on the DNA quality, when obtained from hypotonic treated cells. With the hypothesis that the nucleus swollen status, typically induced by a mild hypotonic treatment, could favour the subsequent DNA dissolution from their position in the agarose matrix, I tested the quality of DNA preparations, obtained from matched agarose plugs: lymphoblastoid cells coming from the same cell culture were exposed/not exposed to a hypotonic treatment before being included in agarose plugs at increasing concentrations. The results are summarised in Table 4 and the positive effect of the hypotonic treatment can be easily appreciated.

Cells *per* plug	Quality of DNA preparation as evaluated after combing	
	With hypotonic treatment	**Without hypotonic treatment**
200.000	Not dense enough	DNA bundles
350.000	Optimal density and elongation; used successfully in FISH experiments	DNA bundles
600.000	DNA bundles	DNA bundles

Table 4. Lymphoblastoid cells were exposed or not exposed to a hypotonic treatment before being included in agarose plugs at increasing concentrations. The DNA was dissolved in 0.1 M MES, pH 6.5.

3.1.1.3. The DNA solution

The agarose melting and the DNA extraction must occur in a proper solution, with defined chemical properties: the β-agarase enzyme acts more efficiently in the pH range of 5.0-8.5. Furthermore, the presence of a pH range (5.5-6.5) in the solution is essential to determine the specific binding of the DNA by its extremities on the hydrophobic silanised surface (Allemand *et al.*, 1997). Moreover, the solution molarity can influence directly the DNA stability, because a low ionic strength in the solution can cause its denaturation, with consequent molecule breaking. To optimize these parameters to our cell samples the quality of the DNA solution obtained by extracting and dissolving the genomic DNA at different conditions was compared: 0.5 M MES (pH 5.5), 0.1 M MES (pH 5.5), 0.1 M MES (pH 6.5). Table 5 summarizes the results of one representative among these trials.

DNA quality observed at	0.5 M MES (pH 5.5)	0.1 M MES (pH 6.5)
Day 1	Good density and elongation	DNA bundles
Day 7	Good density and elongation; used successfully in a replication assay	DNA bundles
Day 14	DNA degradation	Good elongation; persistence of DNA bundles

Table 5. Quality of DNA preparations obtained from plugs including 200.000 primary lymphocytes (in the presence of hypotonic treatment of the cells) and extracted as reported in columns. Preparations were re-evaluated at different times from first observation (rows).

In conclusion, although unbending of DNA seemed to be favoured at pH 5.5, the stability of the solution was low, and DNA underwent degradation in a short time. Best results, in term of DNA stability, were obtained by using 0.1 M MES at pH 6.5, however the DNA remained very often in bundles for a long time. The finding that the quality of the DNA seemed to improve along the time (Table 5) prompted us to introduce a further step, consisting in a very prolonged incubation of the extracted DNA solution (up to 7 days), before transferring it into the *reservoir* used for combing: as reported in Materials and Method (section 2.2.2.), at the end of the overnight digestion by β-agarase, the eppendorf vial was maintained at 50 °C step for 4 h and then at room temperature for at least 7 days. The comparison of matched DNA preparations confirmed that the process of DNA dissolution is slow, and it is not favoured into the thin *reservoir* well.

Combining the indications deriving from the above trials, it can be concluded that about 300.000 genomes are requested when primary lymphocytes or lymphoblastoid cells are considered. This number is in agreement with the indications coming from Bensimon's laboratory (personal communication). However, the hypotonic cell swelling is a prerequisite to achieve good quality DNA preparations from the cell types of our interest (Table 4.). This requirement, as well as the observation that most efficient swelling of the DNA can be obtained before, and not after moving it in the thin teflon *reservoir*, probably depend on the small size of the nucleus in lymphoblasts. These results suggest that the original

spatial organization of the high molecular weight genomes is retained after the purification step and it can influence the quality of the final preparation. This explanation is in agreement also with the previous set of observations concerning the number of cells *per* plug.

3.1.1.4. Setting up the FISH protocol for combed DNA

Setting up the FISH protocol required to check, by several trials, the optimal probe concentration in single and multiple probe experiments. The results consist in the detailed procedure described in Materials and Methods (section 2.3.).
On the basis of the already known probe size, it was possible to verify that the DNA stretching was uniform, leading to the very precise extension factor of 2 kb/µm, as expected (Herrick and Bensimon, 1999). Therefore, home made silanised surfaces have the same properties predicted by the laboratory in which the molecular combing has been developed.
Each of the several panels of probes used in this project cover on the average 1 Mb of DNA sequence. On the whole, 287 molecules were analysed and among them 197 showed double hybridization signals, whereas 45 carried triple signals (68.4% and 18.6% respectively). Based on these data it can be inferred that the average length of uninterrupted combed molecules should cover up to 600 kb-1 Mb.

3.1.2. Cell synchronization and Flow Cytometry

The separation of an enriched S-phase fraction from the other phases of the cell cycle was attempted by Counterflow Centrifugal Elutriation (CCE). This approach was chosen because it allows to obtain synchronous cell populations without any chemical or drug treatments, which reasonably may interfere with the fragile site expression. Instead, CCE is based on separation of the cells as a function of their size. Previous unpublished data obtained in our laboratory indicated that human primary lymphocytes are not suitable for CCE, due to the presence of many cell clumps in actively growing cultures, while lymphoblastoid

cell lines could be successfully separated in function of the cell cycle phase and further subcultured. A human lymphoblastoid cell line, H691, obtained in our laboratory after immortalization from B-lymphocytes of a healthy individual, is expressing at high frequency the common fragile site *FRA6E* (Russo *et al.*, 2006). For this reason, H691 line was considered for CCE experiments.

After separation of a total amount of 300-400 x 10^6 cells, ten cell fractions were collected (numbered from 1 to 5, each one sub-classified as A or B). An aliquot of each fraction was collected and analysed, together with the control population, by flow cytometry (FACscan, Beckton Dickinson, U.S.A.). In Figure 6. the results, elaborated by the Cell Quest software, are reported.

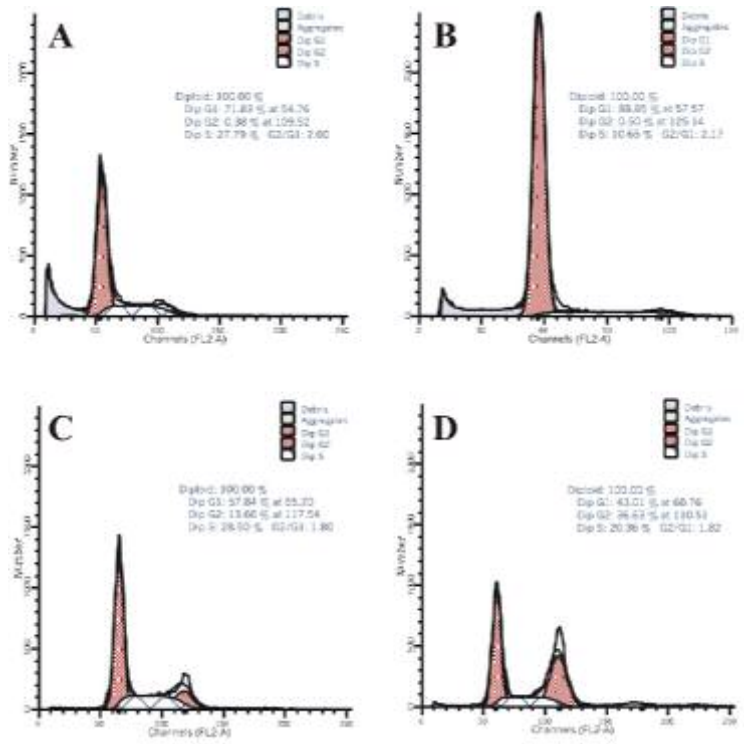


Figure 6. FACS analysis from elutriated H691 cells. Data were elaborated by Cell Quest software. Percentages of cells in different cell cycle phases are reported. A) The heterogeneous H691 control population; B) Results from the second fractioned subpopulation (1B fraction); C) Data from cell fraction 4A; D) The cellular fraction (5B).

In the heterogeneous H691 control population (panel A), G_1 cells were the more represented (71.8%), as expected for this rather long phase of the cell cycle, 27.8% of the cells were replicating (S-phase) and 0.4% of them were in the G_2 phase. In panel B, results from the second fractioned subpopulation (1B) are shown. The G_1

percentage resulted higher (88.8%) than in the controls, indicating an enrichment of cells in this phase, but we still observed a 10.7% of S-phase cells and a 0.5% of cells in G_2. In the third graph (panel C) representing the cell fraction 4A, where the S-phase was expected to be the more represented cell cycle phase, only 28.5 % of the cells were in S-phase, a value which is very near to the control one. With respect to control data, only a small decrease of the G_1 cells percentage (57.8%) was recorded indeed, while the G_2 fraction was 13.7%. In panel D, data from a late cellular fraction (5B) are presented. In this case a strong increase of the G_2 cells with respect to the control percentage was observed (36.6%), but the high percentages of G_1 (43%) and S (20.4%) cells still present indicate an unsuccessful separation among the different cell cycle phases.

Replication of this experiment as well as modulation of parameters, such as the total amount of the cell population, the centrifugal rotor speed and the flow rate, did not lead to improvement of the outcomes (data not shown). A different cell line (TK6) available in our laboratory was then used. In Figure 7. flow cytometry results are reported.

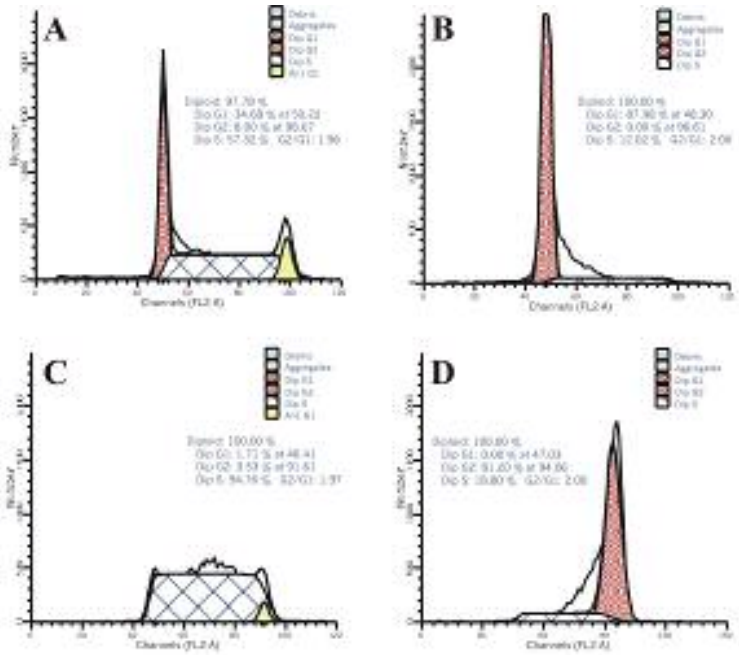


Figure 7. FACS analysis from TK6 cells. Data were elaborated by Cell Quest software. Percentages of cells in different cell cycle phases are reported. A) The heterogeneous TK6 control population; B) Results from the second fractioned subpopulation (1B fraction); C) Data from cell fraction 5B; D) The late cellular fraction (8B).

As shown in panel A, in the heterogeneous TK6 control population the G_1 phase fraction represented the 34.7%, 57.3% of the cells were in S-phase and 8% in G_2. In panel B, the results from the second subpopulation collected (1B) are shown. The G_1 percentage resulted much higher (88%) than in the controls, indicating a conspicuous cells enrichment in this phase, also remarked by the small amount of S-phase cells (12%) and by the absence of G_2 cells. In panel C, representing fraction 5B, the S-phase sub-population was the most represented one (94.8%), as expected, and with respect to control data a strong decrease of the G_1 cells percentage (1.7%) was recorded, while the G_2 fraction constituted the 3.5%. In panel D, data from fraction 8B are presented. In this case a strong increase of the G_2 cells (81.2%) was obtained with respect to the control, together with a small percentage of S-phase cells (18.8%) and the total absence of G_1 cells, indicating a very precise cell-cycle dependent cell separation.

3.2. Replication analysis at whole genome level

Primary human peripheral lymphocytes, isolated from two donors (one male and one female), were used to investigate the pattern of DNA replication at whole genome level.

PHA-stimulated lymphocytes were labelled, as described in Materials and Methods (section 2.2.1.). For each donor, parallel cell cultures were started in order to evaluate the effect of aphidicolin (APH), at different doses (0.02 μM, 0.04 μM, 0.4 μM) and times (2 h, 24 h), with respect to unperturbed (control) condition. Flow cytometry analysis and DNA combing was carried out on male donor samples. Figure 8. shows the results obtained after flow cytometry analysis of the harvested cells.

After 2 h of APH treatment, the frequencies of each cell cycle phase were stably represented, with respect to the control population, at each treatment condition. After 24 h, the S phase percentage resulted two fold higher than in the untreated cells, when cells were exposed to 0.4 μM APH. Treatments at low APH concentrations gave no effect, as above.

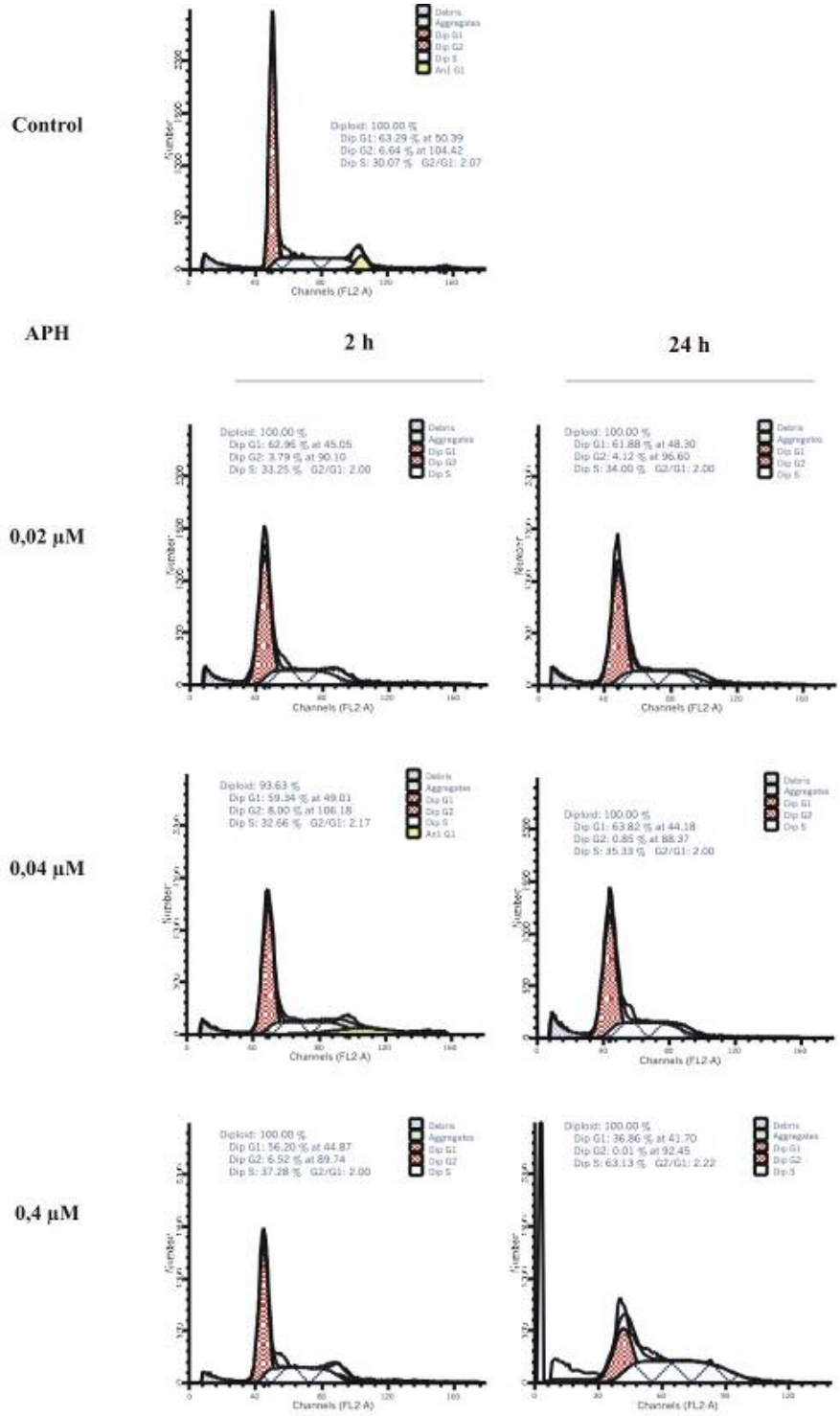


Figure 8. Flow cytometry analysis of APH treated lymphocytes. In each panel, corresponding to the different doses and times of APH tested, the relative percentages of the cell cycle phases are reported

In order to investigate the APH effect on the replication process at high resolution level, two main parameters were analysed to describe origin regulation and fork progression in our cell samples: the mean fork speed and the inter-origin distance.

The analysis was focused on the high APH concentration (0.4 µM) and the short time interval (2 h). The results are shown in Table 6.

Measurements of bidirectional complete forks (according to Material and Methods, section 2.3.9.) showed that control data deriving from the two donors were comparable; therefore they have been pooled together (Table 6). In APH treated cells isolated from the female donor, a slowing effect on the replication process was visible after 2 h of treatment, when a strong decrease of the mean fork speed value was found with respect to the pooled control data ($P < 0.001$). Also the average replicon size was strongly decreased, confirming that replication forks proceed slowly ($P < 0.001$). Figure 9 shows an example of a slowed replication fork, as demonstrated by the complete double labelling in spite of the reduced arm size, as typically found in response to APH.

	Donor	**Mean Fork speed (kb/min) average** ± SE	**Mean Replicon size (kb) average** ± SE
CTR	Female	1.97 ± 0.14	318,1± 33.2
	Male	2.17 ± 0.10	323.3 ± 16.2
	Pooled data	2.09 ± 0.08	321.3 ± 15.9
APH 0,4 µM 2 h	Female	0.24 ± 0.02***	54.3 ± 8.1***

Table 6. Measurements (kb) of bidirectional complete forks observed in control lymphocytes in comparison to APH treated cultures. *** = $P < 0.001$.

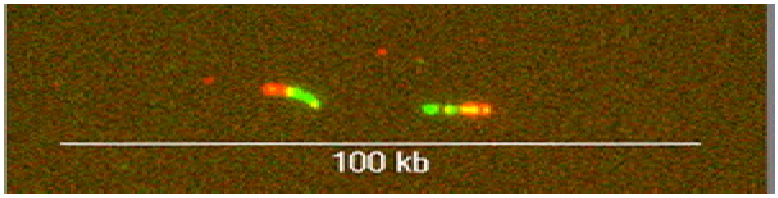


Figure 9. Digital image of a complete bidirectional fork detected in APH treated cells (0.4 µM, 2 h), showing a decreased replicon size. Calibration bar = 100 kb.

3.3. Single *locus* replication analysis

3.3.1. Probe selection

The single-*locus* analyses were focused on specific sub-regions within two CFS (*FRA6E* and *FRA3B*): in *FRA3B*, we investigated the region of *FHIT* gene, which corresponds to the core of fragility. In *FRA6E*, the centromeric and telomeric boundaries of the fragile region were considered, where *ARID1B* and *PARK2* genes are located respectively (Russo *et al.*, 2006). We have chosen different sets of genomic clones to cover at least 1 Mb sequence for each region under analysis.
By performing single colour FISH on combed DNA, we tested preliminarily the hybridization pattern of each set of genomic clones at each *locus* under analysis. Measurements of lengths and distances of all probes were performed, in order to verify if values were comparable with ones provided by the Human Genome Browser Ensembl database. In most of the cases, the measurements we obtained were not in agreement with size reported, in agreement with the fact that the database reports only the non-overlapping sequences of genomic clones, in which form the contig. Only those sets which allowed us to an unequivocal identification and orientation centromere-telomere of the sequence, were selected and used for the further FISH experiments. In Table 7. the selected set employed in the analysis for each region are reported.

Locus	Genomic Clones	Ensembl Reported Size (kb)	Measured Size (kb)
HPRT	**RP11-355K23**	4	207.0
	D1	249	97.1 ± 1.7
	RP11-674A04	233	273.5
	D2	68	72.8 ± 0.7
	RP11-746F03	169	192.2
FRA6E-ARID1B	**RP11-230C9**	170	193.5
	D1	115	124.3 ± 22.0
	RP1-80E10	111	184.5
	D2	250	310.2 ± 67.7
	RP11-96F3	200	193.8
FRA6E-PARK2	**RP3-473J16**	171	194.5
	D1	204	216.2 ± 1.3
	RP1-45F6	135	151.9
	D2	231	240.6 ± 1.7
	RP11-735H10	151	179.3
FRA3B-FHIT	**RP11-137N22**	162	184.1
	D1	198	172.0 ± 10.1
	RP11-468L11	150	199.3
	D2	209	201.0 ± 16.1
	RP11-48E21	32	188.3

Table 7. The set of three genomic clones selected for each *locus* under analysis. The Ensembl reported lengths as well as the distances between pairs of probes (D1/D2), are reported in comparison to measurements performed after DNA combing .

3.3.2. DNA replication pattern at *HPRT locus*

This *locus* (Xq26.1) has been chosen as a control region, because it has been demonstrated that in normal lymphoblastoid cells it is an early replicating region, with a high degree of asynchronous replication (Subramanian and Chinault, 1997). In Figure 10. the *HPRT locus* is depicted, with the set of genomic probes employed for the analysis.

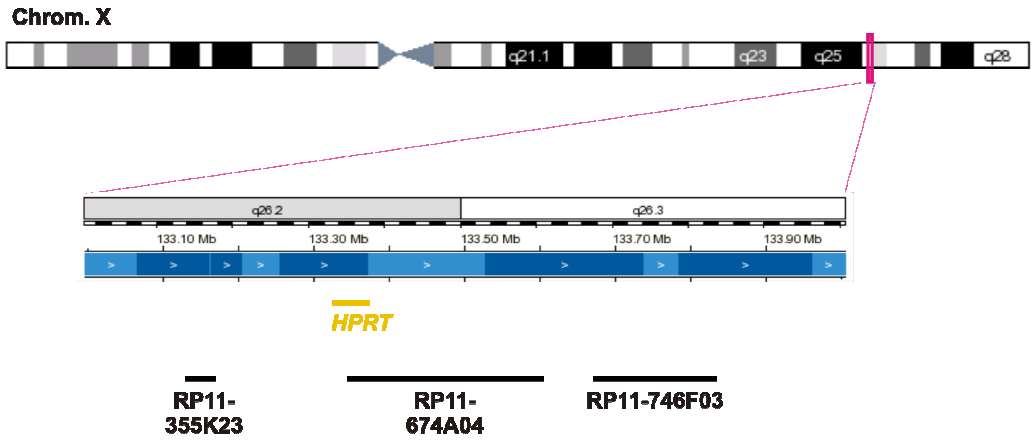


Figure 10. The set of genomic probes employed for the analysis at *HPRT locus.* Top: the position of the whole region along the chromosome X is reported. Bottom: The detailed location of the selected set of clones is represented.

31 informative molecules, presenting two or three probe signals were found, and in 14 of them (45%) the replication of at least one segment was visible. This percentage represents the replicating DNA fraction. Measurements of informative IdU and CldU fluorescent signals revealed a mean fork speed of 1.5 ± 0.21 kb/min (N = 18).

In the pool of the 45% replicating molecules, we detected 27 replication forks. The mean inter-origin distance resulted 124.4 ± 36.6 kb (N = 6; Min-Max values of 24.7-288.3 kb). Some unexpected fork types (Materials and Methods, section 3.3.9, Figure.3.) were recorded: 19% were unidirectional forks (5/27), 11% corresponded to fork arrest events (3/27) and 22% to asynchronous forks (6/27).

Concerning the origin mapping, a more active initiation zone was observed in the region spanning the 355K23 and 674A04 clones, where the *HPRT* gene is located, than in the downstream region. 10 origins were positioned very closed to each other (Supplementary Data, Figure I). Although it has been demonstrated to show a normal early replication pattern, unidirectional forks (Supplementary Data, Figure I, M10 and M14) and several fork arrest events were detected, as presented in Figure I (Supplementary Data). Instead, only 3 origins were mapped between the 746F03 clone and the extreme field of the molecule, suggesting a different regulation of the replication process in this region than in the upstream sequence.

3.3.3. DNA replication pattern at *FRA3B*

In Figure 11. the *FRA3B locus* is depicted, with the set of genomic probes employed for the analysis.

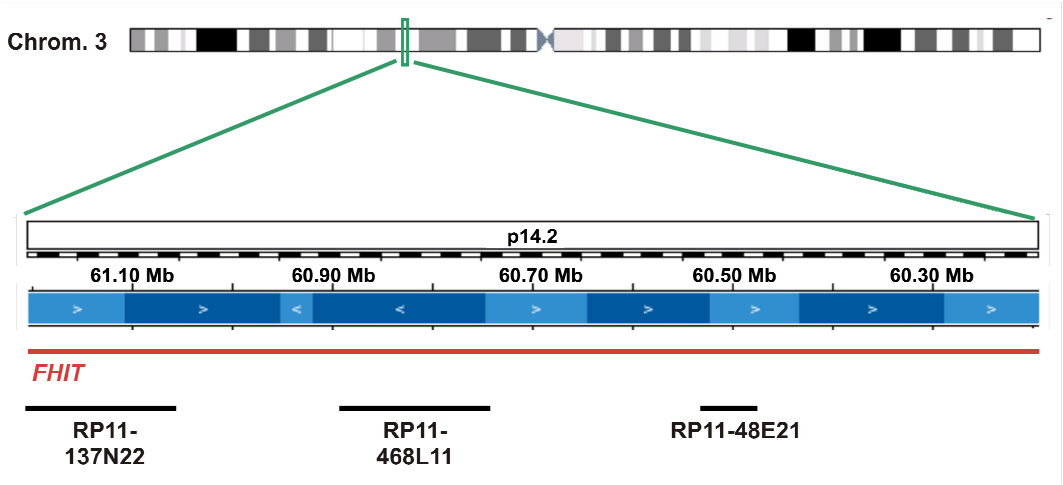


Figure 11. The set of genomic probes employed for the analysis at *FRA3B locus*. Top: the position of the whole region along the chromosome 3 is reported. Bottom: The detailed location of the selected set of clones is represented.

27 informative molecules have been detected into the *FHIT locus*, and in 18 of them the replication was visualised, therefore the replicating DNA fraction was 67%. Measurements of 19 forks revealed a mean fork speed of 1.4 ± 0.26 kb/min.
In the pool of the 18 replicating molecules, we observed 26 replication forks. Among them: 31% were unidirectional forks (8/26), 11.5% fork arrest events (3/26) and 15.5% asynchronous forks (4/26).
Concerning the origin mapping, it was not possible to define accurately the position of any replication origin, because of the low replication rate observed. In fact, only unidirectional forks were detected (Supplementary Data, Figure II).

3.3.4. DNA replication pattern at *FRA6E*

Because of its fragility pattern, *FRA6E* can be divided in three sub-regions: the central sub-region is less fragile than the proximal and distal ones (Russo *et al.*, 2006). In my analysis, the centromeric and telomeric boundaries of the fragile

region were considered, where *ARID1B* and *PARK2* genes are located respectively (Figure 12.).

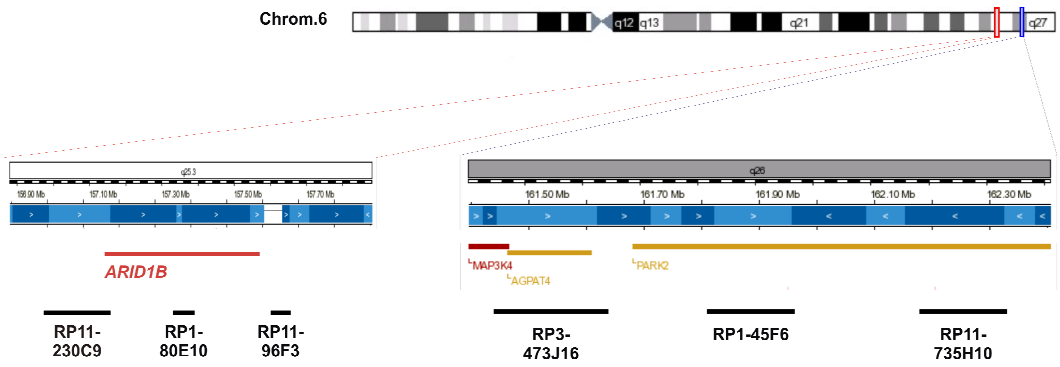
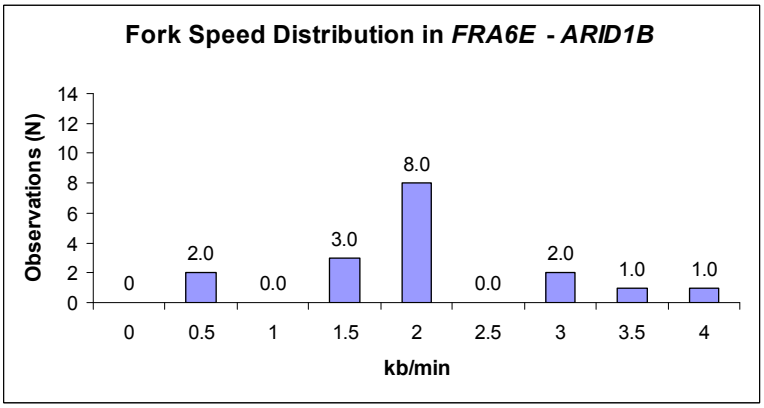


Figure 12. The set of genomic probes employed for the analysis at *FRA6E-ARID1B* and *FRA6E-PARK2 loci.* Top: the position of the two regions along the chromosome 6 is reported. Bottom: The detailed location of the selected clones is represented.

In *ARID1B* region, 48 informative molecules were found and 22 of them presented replication signals (46%). In *PARK2*, 58 informative molecules have been considered, with replicating DNA fraction equal to 57% (N = 33).

Concerning the mean fork speed, this value is higher in *ARID1B* (2.0 ± 0.27 kb/min, N = 17) than in *PARK2* region (1.2 ± 0.12 kb/min, N = 48) (0,001 < P < 0,01).The distribution of fork rates in *FRA6E-ARID1B* region was remarkably less homogeneous, if compared with *FRA6E-PARK2*. In fact, in this region fork speed values are spread in a large range (0.5-4 kb/min), while in *FRA6E-ARID1B* the more frequently observed values were in a narrow range of 1.5-2 kb/min, as shown in Figure 13.



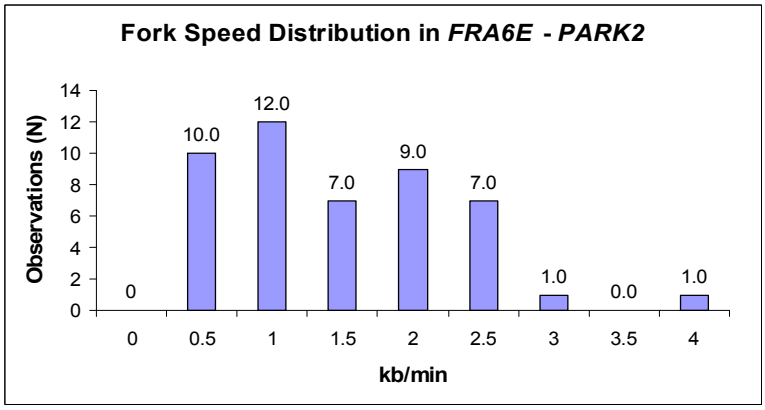


Figure 13. Fork rate distributions in *FRA6E-ARID1B* and *FRA6E-PARK2* regions.

Notably, these forks have been detected in correspondence of a specific 200 kb sub-region.

Within the fully informative forks analysed, we found several unidirectional, asynchronous and fork arrest events. In particular, concerning the *ARID1B* region, in the pool of the 22 replicating molecules, we detected a total of 32 forks. 16% of them were unidirectional forks (5/32), 9.5% were fork arrest events (3/32) and 3% of them were asynchronous forks (1/32). Only one measurement was possible concerning the inter-origin distance, which is in the normal range: 87.7 kb (N = 1).

In *PARK2*, among 33 replicating molecules it was possible to analyse 61 forks. Unidirectional forks were 31% of the events (19/61), fork arrest occurred in 5% of the cases (3/61) and asynchronous forks constituted 18% of the events (11/61). The inter-origin distance estimated on 8 observations was found to be 54.9 ± 14.4 kb (Min-Max values of 8.8-129.4 kb).

In *ARID1B* the origins were not mapped, because the number of observations was not sufficient.

Concerning the *PARK2* sequence three different sub-regions were identified, on the basis of the origin firing pattern. 20 origins were positioned and 9 of them were mapping in a range of 10-100 kb, in correspondence of clone 473J16, therefore upstream the *PARK2* gene sequence. Moreover, an high number of origins and also many fork arrest events were detected in this region (Supplementary Data, Figure III, M6, M11, M15, M16 and M28). Only 3 origins were observed within a 30 kb sub-region corresponding to clone 45F6, but in that

location a wide replicon was detected, which is probably able to complete the replication of the entire sub-region. Concerning the region spanning clone 735H10 and the extreme field of the molecule, 8 origins were mapped, even if 3 of them need to be confirmed. Unidirectional forks were found to be widely distributed in all the sub-regions considered and several fork arrest events were detected (Supplementary Data, Figure III).

3.4. Single *locus* DNA replication analysis in APH treated cells

Considering all the parameters described in section 3.2., we performed FISH on APH-treated combed DNA, in order to evaluate the replication pattern of *FRA6E-PARK2* and *HPRT locus*, used as a control region.

Concerning the *HPRT locus*, 38 informative molecules have been recorded and 16 of them presented replication signals (42.1%). Fork measurements (N = 34) revealed a mean fork speed of 0.3 ± 0.03 kb/min, which is highly significantly different for the control value (1.5 ± 0.21 kb/min, $P < 0.001$, section 3.3.2). In the pool of the 16 replicating molecules, we observed 34 replication forks. Among them, 29.4% were unidirectional forks (10/34), while the fork arrest events and asynchronous forks frequencies were not estimated because of the strong slowing effect of the APH on the replication. In the region spanning clones 355K23 and 674A04, 14 origins were mapped. Although origin positions appeared very similar to those detected in the control, the decreased average inter-origin distance suggests that APH treatment lead to the firing of usually inactive origins. Interestingly, the replication pattern consisted of many forks activated during the second labelling pulse (visible as red signals), indicating that after APH exposure the replication was arrested, then resumed in the last 30 minutes before harvesting, although with rather low speed. Based on 14 observations, the inter-origin distance was found to be 44.8 ± 11.5 kb, which is statistically ($P < 0.05$) different from the control value (124.4 ± 36.6 kb).

Concerning the *FRA6E-PARK2* region, 41 informative molecules were recorded and in 13 of them the replication was visible (replicating DNA fraction = 31.7%).

Fork measurements (N = 32) revealed a mean fork speed of 0.28 ± 0.04 kb/min, highly significantly different for the control value of 1.2 ± 0.12 kb/min ($P < 0.001$, section 3.3.4.). In the pool of the 13 replicating molecules, we observed 32 replication forks. Among them, 53.1% were unidirectional forks (17/32), 6.3% corresponded to fork arrest (2/32) and 3.1% were asynchronous forks (1/32). As shown in Supplementary Data, Figure V, the replication process was strongly impaired by the APH treatment, and for this reason it was not possible to determine any origin position. In most of the cases, forks activated during the second labelling pulse (only red signals) or unidirectional forks were detected, as shown in M1, M2, M4, M5, M7 (Supplementary Data, Figure V). Due to the uncertain origin positioning, and on the basis of only 4 observations, at the present we can assume an inter-origin distance in a range of 73.6-95.5 ± 24.5-16.0 kb.

3.5. Evaluation of the replication timing by FISH on interphase nuclei

The replication timing of the regions under analysis was evaluated also by the canonical cytogenetic approach based on FISH analysis in interphase nuclei. With this approach, DNA sequences which have not yet replicated must show a single hybridization spot, while the sequences which have been replicated appear as doublets. For each probe under analysis, 250 nuclei were scored and, in order to specifically visualise nuclei in S-phase, BrdU was immunodetected and only positive nuclei were counted. Different probes were employed: to characterize the replication features of *FRA6E*, the genomic clone RP11-306O13, located immediately outside the centromeric boundary of the fragility region at 6q25 (Russo *et al.*, 2006) was used as an internal control sequence; the clone RP11-211O7, was located at *FRA6E*, downstream *PARK2*. The two sequences are separated by 9 Mb; unfortunately, probes mapping at the identical location studied by molecular combing gave interphase fluorescent spots of low quality, which were not useful for this approach. For the other two *loci*, the probes were respectively: RP11-468L11, located at *FRA3B-FHIT* region, i.e. in the core of fragility, RP11-674A04, mapping at the *HPRT locus*. Initially we focused the

study on lymphocyte control population to determine the replication timing of the sequences in normal conditions. Concerning *HPRT* and *FRA3B* regions, data we obtained highlighted an high degree of asynchronous replication (52.5%) correlated to an early replication pattern (21.3% of double signals) for what concerned *HPRT*, and a late replicating pattern (33.7% of singlet signals), associated to asynchronous replication between the two alleles (50.0%), for *FRA3B*. The results are summarised in Figure 14 (panels A and B) and can be considered in agreement with already published studies (Subramanian and Chinault, 1997; Le Beau *et al.*, 1998). Panels C and D of Figure 14 show the replication modality of *FRA6E*; in particular data from clone 211O7 clone suggested asynchronous replication of the two alleles (53.1%), while the external probe 306O13 revealed an early replication pattern, underlined by the high percentage of double signals (49.5%).

In order to determine in which way stress conditions can influence the replication process into these regions leading eventually to the fragile site expression, cultures from the female donor lymphocytes treated for 24 hours with APH (0.2 μM) were analysed by interphase FISH. In *HPRT* and *FRA3B* regions, the effect induced by APH on replication was clearly visible as an increase of nuclei showing two singlets or one singlet and one doublet signal, as shown in Figure 14, panels A and B.

Concerning *FRA6E*, and its control sequence, detected by the probe 306O13, the results suggested that APH was differently effective in delaying the replication along this *locus*. In fact, looking at the early replicating probe 306O13, we observed a strong delay on its replication progression (48.9% of singlet signals vs. 13.1% in the matched controls, $P < 0.001$, Figure 14 panel C). Downstream the *PARK2* gene, at the telomeric end of *FRA6E*, the 211O7 probe still highlighted a strong asynchrony of the replication, as in controls, but further slow down. In fact, the two-singlet percentage increases from 14.1% to 30.2% ($P < 0.05$, Figure 14, panel D).

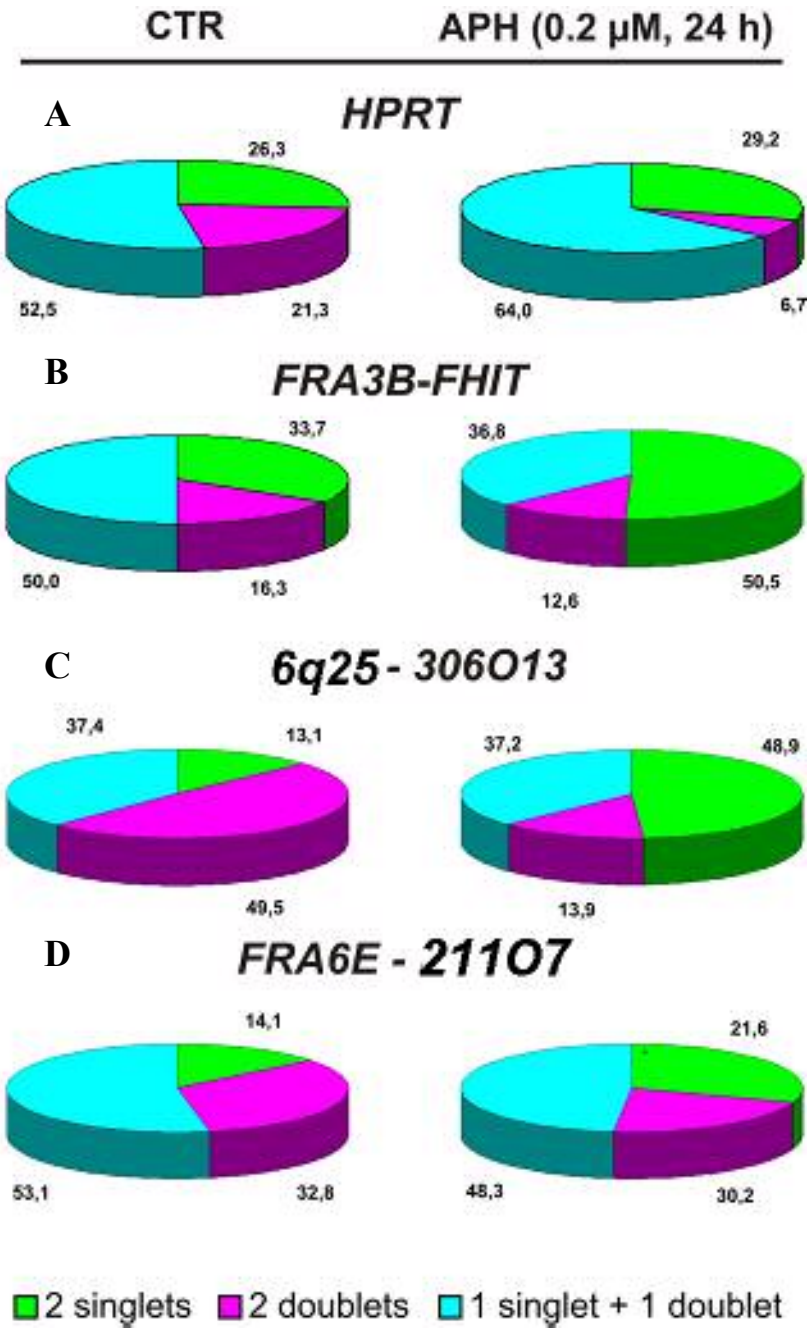


Figure 14. Replication pattern analysis at *HPRT*, *FRA3B-FHIT*, *6q25-306O13* and *FRA6E-PARK2 loci* evaluated by FISH on interphase nuclei of APH treated and control (CTR) lymphocytes. The clones used for the four *loci* are respectively: RP11-674A04, RP11-468L11 (panel A), RP11-306O13, which is located immediately outside the centromeric boundary of *FRA6E* fragility region and RP11-211O7 which is mapping downstream *PARK2* gene. Single hybridization spots correspond to unreplicated regions while double spots indicate replicated sequences.

4. Discussion

In higher eukaryotes, some features of the replication mechanisms are still not well defined. It has been proposed that alterations of these processes could be involved in the genomic instability. In particular, stalled or collapsed replication forks or unreplicated DNA may be one of the causes of fragile site expression. When replication forks are stalled or blocked, the DNA synthesis is arrested or strongly delayed, cell cycle progression is stopped and the intra-S checkpoints response is activated (MacDougall et al., 2007).

The main focus of the present project was to investigate if the instability of common fragile site *FRA6E* could be related to the perturbation of its replication. Using a single molecule, high-resolution approach based on molecular combing, we focused the analyses in order to map replication origins, to determine fork density and replication rates within specific regions. The replication dynamics was compared in control conditions and after aphidicolin-induced replication stress. The study was carried out in primary human lymphocytes, and it considered as control regions the early replicating *HPRT locus*, representative of a normal genomic region, and the well characterized common fragile site *FRA3B*, for which a late replication behaviour has been clearly demonstrated (Le Beau *et al.*, 1998).

In heterogeneous untreated primary lymphocytes, our results suggested that the mean fork speeds were similar in all regions analysed; the data were also in agreement with published data obtained at whole genome level on primary keratinocytes (Conti *et al.*, 2007).

Concerning *FRA6E*, which is a large region spanning about 9 Mb, the two boundaries of the fragile region, where *ARID1B* and *PARK2* genes are located, were analysed. Differences in the mean fork speed seem to exist, with higher values in correspondence of *ARID1B* than in *PARK2* region ($0{,}001 < P < 0{,}01$). An interesting observation concerns the fork rate distribution, because lower variability was found in *FRA6E-ARID1B* region than in the other *loci* analysed. Indeed, in most of the regions analysed, fork speed values spread in a wide range (0.5-4.0 kb/min), according to the fact that fork rates can be adjusted dynamically (Conti *et al.*, 2007). In *FRA6E-ARID1B* fork rates were mainly in the range 1.5-

2.0 kb/min. Notably, all these forks were detected in a specific 200 kb sub-region. It has been proposed (Takebayashi *et al.*, 2001) that there is a progressive increase in fork rate as the S phase advances. Accordingly, *FRA6E-ARID1B* could be a late replicating region, and the opportunity to observe replication events along this region could be affected by the short temporal window involved compared to the whole cell cycle duration. S-phase enriched cell fractions obtained in the course of the project will be useful for better understanding the replication modality of this region.

In order to well characterise the entire replication process of the region, we considered the possible deregulation events. While forks with unexpected pattern (e.g. fork arrest events and asynchronous forks) were observed at comparable frequencies in all the regions investigated, unidirectional forks seem to occur at higher frequency in correspondence of the two common fragile sites *FRA6E-PARK2* and *FRA3B-FHIT* regions than at the centromeric boundary of *FRA6E* (*ARID1B*) and at the *HPRT locus*. This observation could reflect an intrinsic feature of the replication mechanism within these fragile regions, in agreement with the hypotheses that the fragility at CFS could be consequent to deregulation of the replication dynamics (Hellman *et al.*, 2000; Palakodeti *et al.*, 2004).

However, preliminary analysis on the replication dynamics of another early replicating *locus*, *LAMINB2*, carried out in the same laboratory where this study has been developed, suggested that the presence of unidirectional forks could represent a normal feature of the replication process in human cells.

The total number of active origins and their initiation timing are two important parameters for the complete genome replication (Machida *et al.*, 2005; Shechter and Gautier, 2005). They seem to be under the control of the intra-S checkpoint, which can affect the origin firing and can regulate the rate of fork progression after DNA damage (Grallert and Boye, 2008). Moreover, it has been recently proposed the existence of an homeostatic regulation, by which fork rates adapt to changes in initiation frequency during the S phase and origin densities adjust spontaneously to accommodate changes in fork speed (Herrick and Bensimon, 2008). For these reasons, inter-origin distances were evaluated in these *loci*; the

records obtained at the different regions are in agreement with published data concerning the whole genome (Conti *et al.*, 2007).

In order to better understand in which way stress conditions can influence the replication process into the fragile regions, lymphocytes derived from two donors were exposed to aphidicolin (APH), which is a well known inhibitor of the replication process and also it induces common fragile sites expression. The effect of APH was evaluated at different doses (0.02 μM, 0.04 μM, 0.4 μM) and times (2 h, 24 h) with respect to unperturbed (control) condition. By FACS analysis it was found that the high concentration of APH (0.4 μM) resulted in accumulation of S phase cells after 24 h of treatment, whereas the distribution of cells into the different cell cycle phases did not appear significantly modified, at the other tested conditions. Chromosome breakages, observed at *FRA3B-FHIT* and at *FRA6E-PARK2* by FISH analysis with specific probes on metaphase chromosome spreads, was used to confirm the effectiveness of APH treatment (data not shown).

Interestingly, the analysis performed by molecular combing at whole genome level highlighted that already after 2 h of APH treatment a strong slowing statistically significant ($P < 0.001$) effect can be observed on the fork rate, with consequent reduction of the observed replicon size ($P < 0.001$). Preliminary analysis suggests in addition a partial rescue of fork progression after 24 h of APH treatment (data not shown).

The replication timing at *FRA3B*, *FRA6E* and *HPRT loci* was elucidated by FISH on interphase nuclei of APH-treated and control cells. As observed previously by molecular combing at the whole genome level, APH treatment was effective in slowing replication process, but with different extent with respect to the *locus* considered. In the control population, data obtained at *HPRT* and *FRA3B loci* highlighted a high degree of replication asynchrony; this was correlated to an early replication pattern as far as *HPRT locus* is concerned, and a late replicating pattern associated to the asynchronous replication of the two alleles, in *FRA3B*. After APH treatment, the effect on replication was clearly visible in both regions and these results can be considered in agreement with already published studies (Subramanian and Chinault, 1997; Le Beau *et al.*, 1998). Concerning *FRA6E*, and its control sequence mapping at the centromeric boundary, the results suggested

that APH was differently effective along the region. In fact, looking at the control early replicating region, APH induced a strong slow of replication, while into the fragile site the replication was characterised by strong asynchrony between the two alleles, as in control cells, but it appeared further delayed. These data are in agreement with the view that common fragile sites may represent sequences replicating very late, which may not be able to recover from a further delay in DNA synthesis (Palakodeti *et al.*, 2004) and it may induce the formation of chromosomes breaks derived from regions of unreplicated DNA (Durkin and Glover, 2007).

Single *locus* analysis on APH treated combed DNA, carried out to evaluate the replication pattern of *FRA6E-PARK2* and *HPRT locus*, showed as the mean fork speed in both regions strongly decreased, if compared to the controls, in agreement with the data obtained at whole genome level (Results, section 4.2., Table 5). Because of the strong general effect of APH on the replication process, fork arrest events and asynchronous forks frequencies could not be estimated. However, at both *loci*, unidirectional forks seem to be more frequent than in the controls. Interestingly, although different proportions of unidirectional events characterised the two regions in unperturbed conditions, in both of them a 1.5-fold increase of the unidirectional forks was suggested after APH treatment.

According to the *locus* considered, the inter-origin distances were found to be differently affected from APH treatment: in *HPRT* region a significant difference was observed after APH with respect to the control (44.8 ± 11.5 kb vs.124.4 ± 36.6 kb, $P < 0.05$), while in *PARK2* no significant differences were found. Based on our observations by combing analysis, in the early and normally replicating *HPRT* region the APH treatment lead to the firing of usually inactive origins, and this response of is in agreement with the homeostatic mechanism which controls fork rates and replicon size during the proceeding of the S phase (Gilbert, 2007; Herrick and Bensimon, 2008; Conti *et al.*, 2007). It has been demonstrated that new replication initiation sites appear when replication forks are stalled or blocked by perturbed conditions, with a consequent decreasing of the inter-origin distance (Gilbert, 2007; Herrick and Bensimon, 2008; Conti *et al.*, 2007). Moreover, from our data it appears that during the 2 h APH exposure most of origins fired during

the last 30 minutes, corresponding to the second labelling pulse. This indicates that in this region the initial response to APH was replication arrest, followed by the slow progression of the replication forks which eventually fired.

On the contrary, in *FRA6E-PARK2* this fine regulation seems to be less efficient, resulting in the inhibition of normally active origins. It can be speculated that as a consequence of the block induced by the APH treatment and of the persistence of unreplicated DNA, the formation of chromosomes breaks is possible.

The results obtained in this thesis highlight the advantage to adopt, together with the classical cytogenetic techniques, novel strategies overcoming the limited spatial resolution of 1-5 Mb typical of FISH analysis. FISH onto single combed molecules is characterised by high resolution (1-5 kb) which is necessary for understanding the replication dynamics in single *loci* of the human genome.

The results collected up to now provide a view of the replication dynamics in primary cells; a next step will be to take advantage from the existence of DNA preparations enriched in late S-phase (deriving from TK6 elutriated cells, section 3.1.2.), to study in more details the response of late replicating fragile sites to stress conditions.

A further development of this study will consist in evaluating the role of the cell pathways involved in DNA damage response, in determining the instability of common fragile site *FRA6E*, in particular looking at the identification of the proteins which associate to fragile regions, when replication stress is induced.

5. References

Abraham R. T. (2001) Cell cycle checkpoint signaling through the ATM and ATR kinases. *Genes & Dev.* 15, 2177-2196.

Aguilera A. and Gomez-Gonzales B. (2008) Genome instability: a mechanistic view of its causes and consequences. *Nature* 9, 204-217.

Allemand J. F., Bensimon D., Jullien L., Bensimon A., Croquette V. (1997) pH-dependent specific binding and combing of DNA. *Biophys*. *J.* 73, 2064-2070.

Anglana M., Apiou F., Bensimon A., Debatisse M. (2003) Dynamics of DNA replication in mammalian somatic cells: nucleotide pool modulates origin choice and interorigin spacing. *Cell* 114, 385-394.

Arlt M. F., Casper A. M., Glover T. W. (2003) Common fragile sites. *Cytogenet. Genome Res.* 100, 92-100.

Arlt M. F., Xu B., Durkin S. G., Casper A. M., Kastan M. B., Glover T. W (2004) BRCA1 is required for common-fragilesite stability via its G2/M checkpoint function. *Mol. Cell. Biol.* 24, 6701-6709.

Arlt M. F., Durkin S. G., Ragland R. L., Glover T. W. (2006) Common fragile sites as targets for chromosome rearrangements. *DNA Repair (Amst).* 5, 1126-1135.

Bando K., Matsumoto S., Onda M., Akiyama F., Sakamoto G., Yoshimoto M., Emi M., Jordan V. C. (1998) Frequent Allelic Loss at 6q26-27 in Breast Carcinomas of the Solid-tubular Histologic Type. *Breast Cancer.* 5, 127-130.

Bartek J. and Lukas J. (2003) Chk1 and Chk2 kinases in checkpoint control and cancer. *Cancer Cell.* 3, 421-429.

Becker N. A., Thorland E. C., Denison S. R., Phillips L. A., Smith D. I. (2002) Evidence that instability within the *FRA3B* region extends four megabases. *Oncogene* 21, 8713-8722.

Bednarek A., Keck-Waggoner C. L., Daniel R. L., Laflin K. J., Bergsagel P. L., Kiguchi K., Brenner A. J., Aldaz C. M. (2001) *WWOX*, the *FRA16D* gene, behaves as a suppressor of tumor growth. *Cancer Res.* 61, 8068-8073.

Blow JJ G. P., Francis D., Jackson D. A. (2001) Replication origins in *Xenopus* Egg extract are 5-15 kilobases apart and are activated in clusters that fire at different times. *JCB* 152, 15-25.

Callahan G., Denison S. R., Phillips L. A., Shridhar V., Smith D. I. (2003) Characterization of the common fragile site *FRA9E* and its potential role in ovarian cancer. *Oncogene* 22, 590-601.

Casper A. M., Nghiem P., Arlt M. F., Glover T. W. (2002) ATR regulates fragile site stability. *Cell* 111, 779-789.

Casper A. M., Durkin S. G., Arlt M. F., Glover T. W. (2004) Chromosomal instability at common fragile sites in seckel syndrome. *Am. J. Hum. Genet.* 75, 65-660.

Chang N. S., Pratt N., Heath J., Schultz L., Sleve D., Carey G. B., Zevotek N. (2001) Hyaluronidase induction of a WWdomain-containing oxidoreductase that enhances tumor necrosis factor cytotoxicity. *J. Biol. Chem.* 276, 3361-3370.

Cesari R., Martin E. S., Calin G. A., Pentimalli F., Bichi R., McAdams H., Trapasso F., Drusco A., Shimizu M., Masciullo V., D'Andrilli G., Scambia G., Picchio M. C., Alder H., Godwin A. K., Croce C. M. (2003) Parkin, a gene implicated in autosomal recessive juvenile parkinsonism, is a candidate tumor

suppressor gene on chromosome 6q25-q27. *Proc. Natl. Acad. Sci.* 100, 5956-5961.

Conti C., Sacca B., Herrick J., Lalou C., Pommier Y., Bensimon A. (2007) Replication fork velocities at adjacent replication origins are coordinately modified during DNA replication in human cells. *Mol. Biol. Cell.* 18, 3059-3067.

Conti C., Seiler J. A., Pommier Y. (2007b) The mammalian DNA replication elongation checkpoint: implication of Chk1 and relationship with origin firing as determined by single DNA molecule and single cell analyses. *Cell Cycle* 6, 2760-2767.

Coquelle A., Pipiras E., Toledo F., Buttin G., Debatisse M. (1997) Expression of fragile sites triggers intrachromosomal mammalian gene amplification and sets boundaries to early amplicons. *Cell* 89, 215-225.

Croce C. M., Sozzi G., Huebner K. (1999) Role of *FHIT* in human cancer. *J. Clin. Oncol.* 17, 1618-1624.

Debacker K., Winnepenninckx B., Longman C., Colgan J., Tolmie J., Murray R., van Luijk R., Scheers S., Fitzpatrick D., Kooy F. (2007) The molecular basis of the folate-sensitive fragile site *FRA11A* at 11q13. *Cytogenet. Genome Res.* 119, 9-14.

Debacker K. and Kooy F. (2007) Fragile sites and human disease. *Hum. Mol. Genet.* 16, 150-158.

Dekaban A. (1965) Persisting clone of cells with an abnormal chromosome in a woman previously irradiated. *J. Nucl. Med.* 6, 740-746.

De Lange T. (2002) Protection of mammalian telomeres. *Oncogene* 21, 532-540.

De Lange T. (2005) Telomere-related genome instability in cancer. *Cold Spring Harb. Symp. Quant. Biol.* 70, 197-204.

Denison S. R., Callahan G., Becker N. A., Phillips L. A., Smith D. I. (2003) Characterization of *FRA6E* and its potential role in autosomal recessive juvenile parkinsonism and ovarian cancer. *Genes Chromosomes Cancer* 38, 40-52.

Dick K. A., Margolis J. M., Day J. W., Ranum L. P. W. (2006) Dominant non-coding repeat expansion in human disease. *Genome Dyn.* 1, 67-83.

Durkin S. G. and Glover T. W. (2007) Chromosomes Fragile Sites. *Annu. Rev. Genet.* 41, 169-192.

Denison S. R., Wang F., Becker N. A., Schüle B., Kock N., Phillips L. A., Klein C., Smith D. I. (2003) Alterations in the common fragile site gene *Parkin* in ovarian and other cancers. *Oncogene* 22, 8370-8378.

Feijoo C., Hall-Jackson C., Wu R., Jenkins D., Leitch J., Gilbert D. M., Smythe C. (2001) Activation of mammalian Chk1 during DNA replication arrest: a role for Chk1 in the intra-S phase checkpoint monitoring replication origin firing. *J. Cell. Biol.* 154, 913-923.

Freudenreich C. H. (2007) Chromosome fragility: molecular mechanisms and cellular consequences. *Front Biosci.* 12, 4911-4924.

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

Glover T. W., Hoge A. W., Miller D. E., Ascara-Wilke J. E., Adam A. N., Dagenais S. L., Wilke C. M., Dierick H. A., Beer D. G. (1998) The murine *Fhit* gene is highly similar to its human orthologue and maps to a common fragile site region. *Cancer Res.* 58, 3409-3414.

Grallert B. and Boye E. (2008) The multiple facets of the intra-S checkpoint. *Cell Cycle* 7, 2315-2320.

Harrison J. C. and Haber J. E. (2006) Surviving the breakup: the DNA damage checkpoint. *Annu. Rev. Genet.* 40, 209-235.

Hellman A., Rahat A., Scherer S. W., Darvasi A., Tsui L.-C., Kerem B. (2000) Replication Delay along *FRA7H*, a Common Fragile Site on Human Chromosome 7, Leads to Chromosomal Instability. *Mol. Cell. Biol.* 20, 4420-4427.

Herrick J. and Bensimon A. (1999) Single molecule analysis of DNA replication. *Biochimie* 81, 859-871.

Herrick J., Stanislawski P., Hyrien O., Bensimon A. (2000). Replication fork density increases during DNA synthesis in *Xenopus laevis* egg extracts. *J. Mol. Biol.* 300, 1133-1142.

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

Hirao A., Cheung A., Duncan G., Girard P. M., Elia A. J., Wakeham A., Okada H., Sarkissian T., Wong J. A., Sakai T., De Stanchina E., Bristow R. G., Suda T., Lowe S. W., Jeggo P. A., Elledge S. J., Mak T. W. (2002) Chk2 is a tumor suppressor that regulates apoptosis in both an ataxia telangiectasia mutated (ATM)-dependent and an ATM-independent manner. *Mol. Cell. Biol.* 22, 6521-6532.

Howlett N. G., Taniguchi T., Durkin S. G., D'Andrea A. D., Glover T. W. (2005) The Fanconi anemia pathway is required for the DNA replication stress response and for the regulation of common fragile site stability. *Hum. Mol. Genet.* 14, 693-701.

Huebner K., Garrison P. N., Barnes L. D., Croce C. M. (1998) The role of the *FHIT/FRA3B locus* in cancer. *Annu. Rev. Genet.* 32, 7-31.

Huebner K. and Croce C. M. (2001) *FRA3B* and other common fragile sites: the weakest links. *Nat. Rev. Cancer* 1, 214-221.

Hurlstone A. F., Olave I. A., Barker N., van Noort M., Clevers H. (2002) Cloning an characterization of hELD/OSA 1, a novel BRG1 interacting protein. *Biochem. J.* 364, 255-264.

Inoue H., Furukawa T., Giannakopoulos S., Zhou S., King D. S., Tanese N. (2002) Largest subunits of the human SWI/SNF chromatin-remodeling complex promote transcriptional activation by steroid hormone receptors. *J. Biol. Chem.* 277, 41674-41685.

Jones C., Penny L., Mattina T., Yu S., Baker E., Voullaire L., Langdon W. Y., Sutherland G. R., Richards R. I., Tunnacliffe A. (1995) Association of a chromosome deletion syndrome with a fragile site within the proto-oncogene CBL2. *Nature* 376, 145-149.

Kaplan R., Morse B., Huebner K., Croce M. C., Howk R., Ravaera M. (1990) Cloning of three human tyrosine phosphatases reveals a multigene family of receptorlinked protein tyrosine-phosphatases expressed in brain. *Proc. Natl. Acad. Sci.* 87, 7000-7004.

Kastan M. B. and Bartek J. (2004) Cell-cycle checkpoints and cancer. *Nature* 432, 316-323.

Kholodnyuk I. D., Szeles A., Yang Y., Klein G., Imreh S. (2000) Inactivation of the human fragile histidine triad gene at 3p14.2 in monochromosomal human/mouse microcell hybrid-derived severe combined immunodeficient mouse tumors. *Cancer Res.* 60, 7119-7125.

Kitada T., Asakawa S., Hattori N., Matsumine H., Yamamura Y., Minoshima S., Yokochi M., Mizuno Y., Shimizu N. (1998) Mutations in the *parkin* gene cause autosomal recessive juvenile parkinsonism. *Nature* 392, 605-608.

Knight S. J., Flannery A. V., Hirst M. C., Campbell L., Christodoulou Z., Phelps S. R., Pointon J., Middleton-Price H. R., Barnicoat A., Pembrey M. E., Holland J., Oostra B. A., Bobrow M., Davies K. E. (1993) Trinucleotide repeat amplification and hypermethylation of a CpG island in *FRAXE* mental retardation. *Cell* 74, 127-134.

Kolodner R. D. and Marsischky G. T. (1999) Eukaryotic DNA mismatch repair. *Curr. Opin. Genet. Dev.* 9, 89-96.

Kolodner R. D., Putnam C. D., Myung K. (2002) Maintenance of genome stability in *Saccharomyces cerevisiae*. *Science* 297, 552-557.

Laird C.D. (1987) Proposed mechanism of inheritance and expression of the human fragile-X syndrome of mental retardation. *Genetics* 117, 587-599.

Lebofsky R., Heilig R., Sonnleitner M., Weissenbach J., Bensimon A. (2006) DNA replication origin interference increases the spacing between initiation events in human cells. *Mol. Biol. Cell.* 17, 5337-5345.

Le Beau M. M., Rassool F. V., Neilly M. E., Espinosa R. III, Glover T. W., Smith D. I., McKeithan T. W. (1998) Replication of a common fragile site, *FRA3B*, occurs late in S phase and is delayed further upon induction: implications for the mechanism of fragile site induction. *Hum. Mol. Genet.* 7, 755-761.

Lemoine F. J., Degtyareva N. P., Lobcchev K., Petes T. D. (2005) Chromosomal translocations in yeast induced by low levels of DNA polymerase a model for chromosome fragile sites. *Cell* 120, 587–598.

Li F., Chen J., Solessio E., Gilbert D. M. (2003) Spatial distribution and specification of mammalian replication origins during G1 phase. *J. Cell. Biol.* 161, 257-266.

Liu Q., Guntuku S., Cui X. S., Matsuoka S., Cortez D., Tamai K., Luo G., Carattini-Rivera S., DeMayo F., Bradley A., Donehower L. A., Elledge S. J. (2000) Chk1 is an essential kinase that is regulated by Atr and required for the G(2)/M DNA damage checkpoint. *Genes Dev.* 14, 1448-1459.

Luk C., Tsao M. S., Bayani J., Shepherd F., Squire J. A. (2001) Molecular cytogenetic analysis of non-small cell lung carcinoma by spectral karyotyping and comparative genomic hybridization. *Cancer. Genet. Cytogenet.* 125, 87-99.

Lukas C., Melander F., Stucki M., Falck J., Bekker-Jensen S., Goldberg M., Lerenthal Y., Jackson S. P., Bartek J., Lukas J. (2004) Mdc1 couples DNA double-strand break recognition by Nbs1 with its H2AX-dependent chromatin retention. *EMBO J.* 23, 2674-2683.

Lukusa T. and Fryns J. P. (2008) Human chromosome fragility. *Biochim. Biophys. Acta* 1779, 3-16.

MacDougall C. A., Byun T. S., Van C., Yee M. C., Cimprich K. A. (2007) The structural determinants of checkpoint activation. *Genes Dev.* 21, 898-903.

Machida Y. J., Hamlin J. L., Dutta A. (2005) Right place, right time, and only once: Replication initiation in metazoans. *Cell* 123, 13-24.

Magenis R. E., Hecht F., Lovrien E. W. (1970) Heritable fragile site on chromosome 16: probable localization of haptoglobin *locus* in man. *Science* 170, 85-86.

Marheineke K. and Hyrien O. (2001) Aphidicolin triggers a block to replication origin firing in *Xenopus* egg extracts. *J. Biol. Chem.* 276, 17092-17100.

Michalet X., Ekong R., Fougerousse F., Rousseaux S., Schurra C., Hornigold N., van Slegtenhorst M., Wolfe J., Povey S., Beckmann J. S., Bensimon A. (1997) Dynamic molecular combing: stretching the whole human genome for high-resolution studies. *Science* 277, 1518-1523.

Mishmar D., Rahat A., Scherer S. W., Nyakatura G., Hinzmann B., Kohwi Y., Mandel-Gutfroind Y., Lee J. R., Drescher B., Sas D. E., Margalit H., Platzer M., Weiss A., Tsui L. C., Rosenthal A., Kerem B. (1998) Molecular characterization of a common fragile site (*FRA7H*) on human chromosome 7 by the cloning of a simian virus 40 integration site. *Proc. Natl. Acad. Sci.* 95, 8141-8146.

Mishmar D., Mandel-Gutfreund Y., Margalit H., Rahat A., Kerem B. (1999) Common fragile sites: G-band characteristics within an R-band. *Am. J. Hum. Genet.* 64, 908-910.

Musio A., Montagna C., Mariani T., Tilenni M., Focarelli M. L., Brait L., Indino E., Benedetti P. A., Chessa L., Albertini A., Ried T., Vezzoni P. (2005) Smc1 involvement in fragile site expression. *Hum. Mol. Genet.* 14, 525-533.

Nancarrow J. K., Kremer E., Holman K., Eyre H., Doggett N. A., Le Paslier D., Callen D. F., Sutherland G. R., Richards R. I. (1994) Implications of *FRA16A* structure for the mechanism of chromosomal fragile site genesis. *Science* 264, 1938-1941.

Nyberg K. A., Michelson R. J., Putnam C. W., Weinert T. A. (2002) Toward maintaining the genome: DNA damage and replication checkpoints. *Annu. Rev. Genet.* 36, 617-656.

O'Keefe L. V. and Richards R. I. (2006) Common chromosomal fragile sites and cancer: focus on *FRA16D*. *Cancer Lett.* 232, 37-47.

Ohta M., Inoue H., Cotticelli M. G., Kastury K., Baffa R., Palazzo J., Siprashvili Z., Mori M., McCue P., Druck T., Croce C. M., Huebner K. (1996) The *FHIT* gene, spanning the chromosome 3p14.2 fragile site and renal carcinoma-associated t(3;8) breakpoint, is abnormal in digestive tract cancers. *Cell* 84, 587-597.

Palakodeti A., Han Y., Jiang Y., Le Beau M. M. (2004) The role of late/slow replication of the *FRA16D* in common fragile site induction. *Genes Chromosomes Cancer* 39, 71-6.

Pang S. T., Fang X., Valdman A., Norstedt G., Pousette A., Egevad L., Ekman P. (2004) Expression of ezrin in prostatic intraepithelial neoplasia. *Urology* 63, 609-612.

Paradee W., Wilke C. M., Wang L., Shridhar R., Mullins C. M., Hoge A., Glover T. W., Smith D. I. (1996) A 350-kb cosmid contig in 3p14.2 that crosses the t(3;8) hereditary renal cell carcinoma translocation breakpoint and 17 aphidicolin-induced *FRA3B* breakpoints. *Genomics* 35, 87-93.

Parrish J. E., Oostra B. A., Verkerk A. J., Richards C. S., Reynolds J., Spikes A. S., Shaffer L. G., Nelson D. L. (1994) Isolation of a GCC repeat showing expansion in *FRAXF*, a fragile site distal to *FRAXA* and *FRAXE*. *Nat. Genet.* 8, 229-235.

Pearson C. E., Nichol Edamura K., Cleary J. D. (2005) Repeat instability: mechanisms of dynamic mutations. *Nat. Rev. Genet.* 6, 729-742.

Pekarsky Y., Garrison P. N., Palamarchuk A., Zanesi N., Ageilan R. I., Huebner K., Barnes L. D., Croce C. M. (2004) Fhit is a physiological target of the protein kinase Src. *Proc. Natl. Acad. Sci.* 101, 3775-3779.

Petermann E. and Caldecott K. W. (2006) Evidence that the ATR/Chk1 pathway maintains normal replication fork progression during unperturbed S phase. *Cell Cycle* 5, 2203-2209.

Ried K., Finnis M., Hobson L., Mangelsdorf M., Dayan S., Nancarrow J. K., Woollatt E., Kremmidiotis G., Gardner A., Venter D., Baker E., Richards R. I. (2000) Common chromosomal fragile site *FRA16D* sequence: identification of the *FOR* gene spanning *FRA16D* and homozygous deletions and translocation breakpoints in cancer cells. *Hum. Mol. Genet.* 9,1651-1663.

Rozier L., El-Achkar E., Apiou F., Debatisse M. (2004) Characterization of a conserved aphidicolin-sensitive common fragile site at human 4q22 and mouse 6C1: possible association with an inherited disease and cancer. *Oncogene* 23, 6872-6880.

Russo A., Acquati F., Rampin M., Monti L., Palumbo E., Graziotto R., Taramelli R. (2006) Molecular characterisation of human common fragile site *FRA6E*, a large genomic region spanning 9 Mb, in: Trends in Genome Research, Clyde R. Williams (Ed.), Nova Science Publishers, Hauppauge NY, USA, pp 155-172.

Ruiz-Herrera A., Castresana J., Robinson T. J. (2006) Is mammalian chromosomal evolution driven by regions of genome fragility?. *Genome Biology* 7, R115.

Sarafidou T., Kahl C., Martinez-Garay I., Mangelsdorf M., Gesk S., Baker E., Kokkinaki M., Talley P., Maltby E. L., French L., Harder L., Hinzmann B., Nobile C., Richkind K., Finnis M., Deloukas P., Sutherland G. R., Kutsche K., Moschonas N. K., Siebert R., Gécz J.; European Collaborative Consortium for the Study of ADLTE. (2004) Folate-sensitive fragile site *FRA10A* is due to an expansion of a CGG repeat in a novel gene, *FRA10AC1*, encoding a nuclear protein. *Genomics* 84, 69-81.

Schwartz M., Zlotorynski E., Kerem B. (2006) The molecular basis of common and rare fragile sites. *Cancer Lett.* 232, 13-26.

Seiler J. A., Conti C., Syed A., Aladjem M. I., Pommier Y. (2007) The intra-S-phase checkpoint affects both DNA replication initiation and elongation: single-cell and -DNA fiber analyses. *Mol. Cell. Biol.* 27, 5806-5818.

Shechter D. and Gautier J. (2005) ATM and ATR check in on origins: a dynamic model for origin selection and activation. *Cell Cycle* 4, 235-238.

Shiraishi T., Druck T., Mimori K., Flomenberg J., Berk L., Alder H., Miller W., Huebner K., Croce C. M. (2001) Sequence conservation at human and mouse orthologous common fragile regions, *FRA3B/FHIT* and *Fra14A2/Fhit*. *Proc. Natl. Acad. Sci.* 98, 5722-5727.

Shridhar V., Staub J., Huntley B., Cliby W., Jenkins R., Pass H. I., Hartmann L., Smith D.I. (1999) A novel region of deletion on chromosome 6q23.3 spanning less than 500 Kb in high grade invasive epithelial ovarian cancer. *Oncogene* 18, 3913-3918.

Smith D. I., McAvoy S., Zhu Y., Perez D. S. (2007) Large common fragile site genes and cancer. *Seminars in Cancer Biology* 17, 31-41.

Subramanian P. S. and Chinault A. C. (1997) Replication timing properties of the human *HPRT locus* on active, inactive and reactivated X chromosomes. *Somat. Cell. Mol. Genet.* 23, 97-109.

Sutherland G. R., Parslow M. I., Baker E. (1985) New classes of fragile sites induced by 5-azacytidine and BrdU. *Hum. Mol. Genet.* 69, 233-237.

Sutherland G. R. and Richards R. I. (1995) The molecular basis of fragile sites in human chromosomes. *Curr. Opin. Genet. Dev.* 5, 323-327.

Tanaka H., Tapscott S. J., Trask B. J., Yao M. C. (2002) Short inverted repeats initiate gene amplification through the formation of a large DNA palindrome in mammalian cells. *Proc. Natl. Acad. Sci.* 99, 8772-8777.

Takai H., Tominaga K., Motoyama N., Minamishima Y. A., Nagahama H., Tsukiyama T., Ikeda K., Nakayama K., Nakanishi M., Nakayama K. (2000) Aberrant cell cycle checkpoint function and early embryonic death in Chk1(-/-) mice. *Genes Dev.* 14, 1439-1447.

Takebayashi S. I., Manders E. M., Kimura H., Taguchi H., Okumura K. (2001) Mapping sites where replication initiates in mammalian cells using DNA fibers. *Exp. Cell. Res.* 271, 263-268.

Takemura H., Rao V. A., Sordet O., Furuta T., Miao Z. H., Meng L., Zhang H., Pommier Y. (2006) Defective Mre11-dependent activation of Chk2 by ataxia telangiectasia mutated in colorectal carcinoma cells in response to replication-dependent DNA double strand breaks. *J. Biol. Chem.* 281, 30814-30823.

Thrash-Bingham C. A., Salazar H., Freed J. J., Greenberg R. E., Tartof K. D. (1995) Genomic alterations and instabilities in renal cell carcinomas and their relationship to tumor pathology. *Cancer Res.* 55, 6189-6195.

van Steensel B., Smogorzewska A., de Lange T. (1998) TRF2 protects human telomeres from end-to-end fusions. *Cell* 92, 401-413.

Verkerk A. J. M., Pieretti M., Sutcliffe J. S., Fu Y. H., Kuhl D. P., Pizzuti A., Reiner O., Richards S., Victoria M. F., Zhang F. P., Eussen B. E., van Ommen G.-J. B., Blonden L. A. J., Riggins G. J., Chastain J. L., Kunst C. B., Galjaard H., Caskey C. T., Nelson D. L., Oostraa B. A., Warren S. T. (1991) Identification of a gene (*FMR-1*) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* 65, 905-914.

Wang Y.-H., Gellibolian R., Shimizu M., Wells R. D., Griffith J. (1996) Long CCG triplet repeat blocks exclude nucleosomes: a possible mechanism for the nature of fragile sites in chromosomes. *J. Mol. Biol.* 263, 511-516.

Wang L., Darling J., Zhang J. S., Huang H., Liu W., Smith D. I. (1999) Allele-specific late replication and fragility of the most active common fragile site, *FRA3B*. *Hum. Mol. Genet.* 8, 431-437.

Wang F., Denison S., Lai J. P., Philips L. A., Montoya D., Kock N., Schüle B., Klein C., Shridhar V., Roberts L. R., Smith D. I. (2004) *Parkin* gene alterations in hepatocellular carcinoma. *Genes Chromosomes Cancer.* 40, 85-96.

Winnepenninckx B., Debacker K., Ramsay J., Smeets D., Smits A., FitzPatrick D. R., Kooy R. F. (2007) CGG-repeat expansion in the *DIP2B* gene is associated with the fragile site *FRA12A* on chromosome 12q13.1. *Am. J. Hum. Genet.* 80, 221–231.

Yu T., Ferber M. J., Cheung T. H., Chung T. K, Wong Y. F., Smith D. I. (2005) The role of viral integration in the development of cervical cancer. *Cancer Genet. Cytogenet.* 158, 27-34.

Yunis J. J. and Soreng A. L. (1984) Constitutive fragile sites and cancer. *Science* 226, 1199-1204.

Zachos G., Rainey M. D., Gillespie D. A. (2005) Chk1-dependent S-M checkpoint delay in vertebrate cells is linked to maintenance of viable replication structures. *Mol. Cell. Biol.* 25, 563-574.

Zanesi N., Fidanza V., Fong L. Y., Mancini R., Druck T., Valtieri M., Rüdiger T., McCue P. A., Croce C. M., Huebner K. (2001) The tumour spectrum in *Fhit*-deficient mice. *Proc. Natl. Acad. Sci.* 98, 10250-10255.

Zhu M and Weiss R. S. (2007) Increased common fragile site expression, cell proliferation defects, and apoptosis following conditional inactivation of mouse Hus1 in primary cultured cells. *Mol. Biol. Cell.* 18, 1044-55.

Zimonjic D. B., Druck T., Ohta M., Kastury K., Croce C. M., Popescu N. C., Huebner K. (1997) Positions of chromosome 3p14.2 fragile sites (*FRA3B*) within the FHIT gene. *Cancer Res.* 57, 1166-1170.

Zlotorynski E., Rahat A., Skaug J., Ben-Porat N., Ozeri E., Hershberg R., Levi A., Scherer S. W., Margalit H., Kerem B. (2003) Molecular basis for expression of common and rare fragile sites. *Mol. Cell. Biol.* 23, 7143-7151.

Zou L. and Elledge S. J. (2003) Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science* 300, 1542-1548.

6. Supplementary Data

Locus	Genomic Clones
HPRT	RP11-355K23
	D
	RP11-674A04
	D
	RP11-746F03
FRA3B	RP11-164G20
	D
	RP11-137N22
	D
	RP11-468L11
	D
	RP11-48E21
	D
	RP11-354I3
FRA6E-ARID1B	RP11-230C9
	D
	RP1-80E10
	D2
	RP11-96F3
FRA6E-PARK2	RP3-473J16
	D1
	RP1-45F6
	RP1-119H20
	D
	RP11-735H10
	D
	RP11-168A05
	RP1-292F10
Outside-PARK2	RP11-307K1
	RP11-211O7
	RP11-621H02
	RP11-257A15

Table I. The complete list of genomic clones selected for each *locus*.