Common Fragile Site Profiling in Epithelial and Erythroid Cells Reveals that Most Recurrent Cancer Deletions Lie in Fragile Sites Hosting Large Genes

Benoît Le Tallec, Gaël armel Millot, Marion Esther Blin, Olivier Brison, Bernard Dutrillaux, Michelle Debatisse

Cell PRESS
Open ACCESS

Cell Reports
Report

Common Fragile Site Profiling in Epithelial and Erythroid Cells Reveals that Most Recurrent Cancer Deletions Lie in Fragile Sites Hosting Large Genes

Benoît Le Tallec,[1,2,3,5] Gaël Armel Millot,[1,2,3,5] Marion Esther Blin,[1,2,3] Olivier Brison,[1,2,3] Bernard Dutrillaux,[4] and Michelle Debatisse[1,2,3,*]

[1]Institut Curie, Centre de Recherche, 26 rue d'Ulm, 75248 Paris, France
[2]Pierre and Marie Curie University Paris 06, 75005 Paris, France
[3]CNRS UMR 3244, 75248 Paris, France
[4]CNRS UMR 7205, Muséum National d'Histoire Naturelle, 75005 Paris, France
[5]These authors contributed equally to this work
*Correspondence: michelle.debatisse@curie.fr
http://dx.doi.org/10.1016/j.celrep.2013.07.003

SUMMARY

Cancer genomes exhibit numerous deletions, some of which inactivate tumor suppressor genes and/or correspond to unstable genomic regions, notably common fragile sites (CFSs). However, 70%–80% of recurrent deletions cataloged in tumors remain unexplained. Recent findings that CFS setting is cell-type dependent prompted us to reevaluate the contribution of CFS to cancer deletions. By combining extensive CFS molecular mapping and a comprehensive analysis of CFS features, we show that the pool of CFSs for all human cell types consists of chromosome regions with genes over 300 kb long, and different subsets of these loci are committed to fragility in different cell types. Interestingly, we find that transcription of large genes does not dictate CFS fragility. We further demonstrate that, like CFSs, cancer deletions are significantly enriched in genes over 300 kb long. We now provide evidence that over 50% of recurrent cancer deletions originate from CFSs associated with large genes.

INTRODUCTION

Common fragile sites (CFSs) are megabase-long loci that recurrently exhibit instability, visible as breaks on mitotic chromosomes following perturbation of DNA replication (Durkin and Glover, 2007). CFSs drive chromosomal rearrangements in tumors (Beroukhim et al., 2010; Bignell et al., 2010), which may favor oncogenesis upon inactivation of tumor suppressor genes hosted by some of these sites (Iliopoulos et al., 2006; Saldivar et al., 2012) and/or amplification of some oncogenes (Coquelle et al., 1997).

It is largely agreed that CFSs remain incompletely replicated until mitotic onset upon replication stress, making them prone to breakage (Durkin and Glover, 2007). Recent studies of four major CFSs have shown that a specific replication program combining late replication with failure to activate origins along the core of the sites is responsible for their delayed replication completion (Le Tallec et al., 2011; Letessier et al., 2011). Because replication programs evolve along with cell differentiation (Méchali, 2010; Ryba et al., 2010), different chromosomal regions can be committed to fragility in different cell types, as illustrated by the different repertoires of CFSs found in human fibroblasts and lymphocytes (Debatisse et al., 2012).

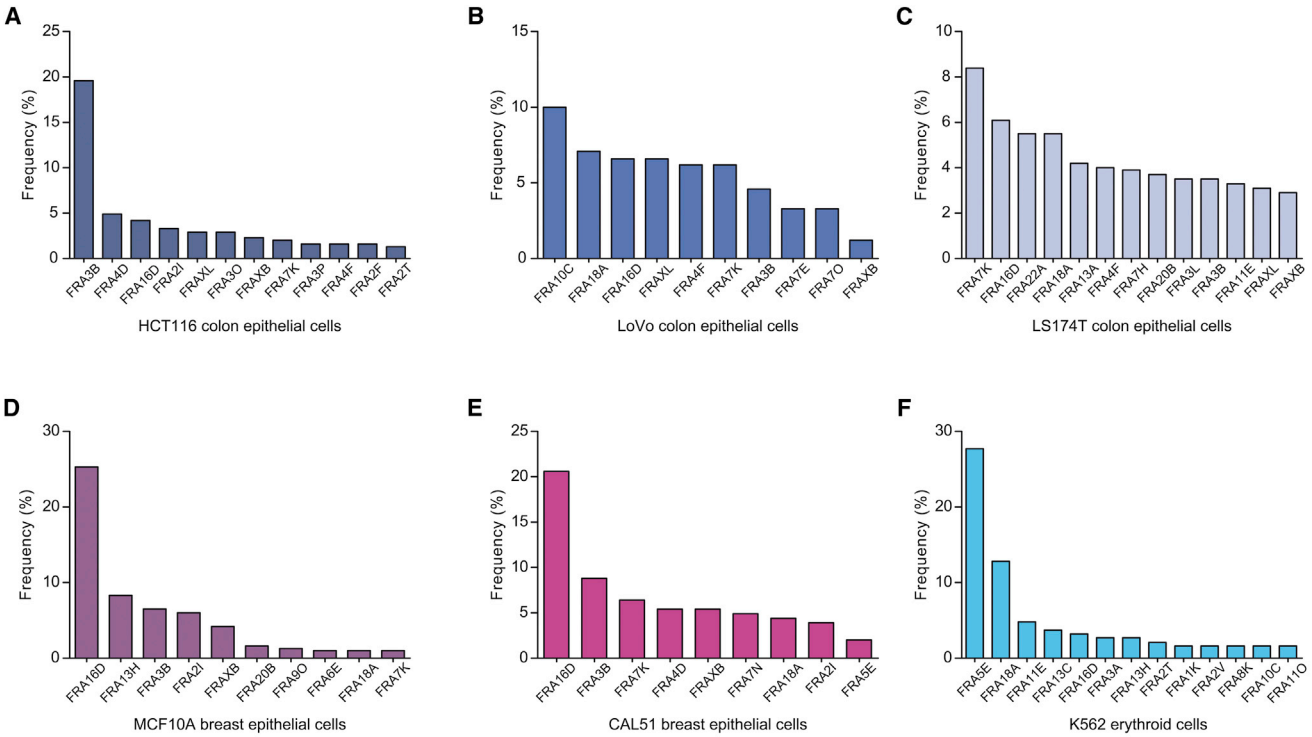
Human CFSs have only been localized in lymphocytes and fibroblasts thus far, which provides a restricted view of the CFS landscape. How many CFSs are present in the human genome and how many are shared by different cell types remain unknown. Answering these questions is required to reevaluate the importance of CFSs in rearrangements found in tumors originating from different cell types. Interestingly, pioneering work from D.I. Smith's group showed that several CFSs overlap genes spanning extremely large genomic regions (hereafter referred to as large genes) (Smith et al., 2007). Many recurrent focal deletions in cancer genomes also target large genes, some of which are associated with known CFSs (Dereli-Öz et al., 2011). The question therefore arises whether all the other unstable large genes are also associated with yet to be identified CFSs.

In this study, we have extended CFS mapping to a wide range of human cell lines including epithelial cells of breast and colon from which most human cancers originate. Mapping was performed at the molecular level and was combined with a comprehensive computational analysis to provide both a refined characterization of CFSs and a precise assessment of their contribution to cancer deletions.

RESULTS

Defining CFS Repertoires of Epithelial and Erythroid Cells

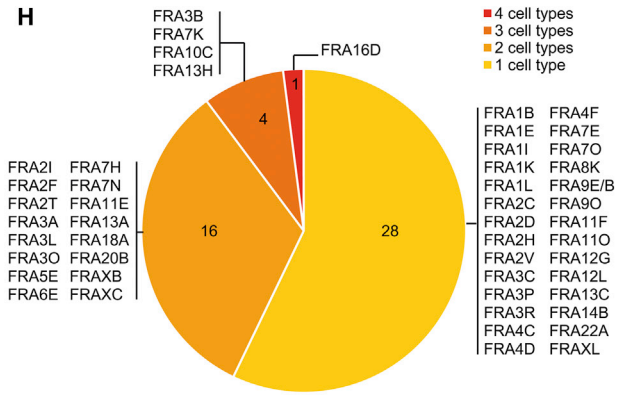
We localized CFSs in six human cell lines, namely three MSI colorectal cancer epithelial cell lines (LS174T, HCT116, and

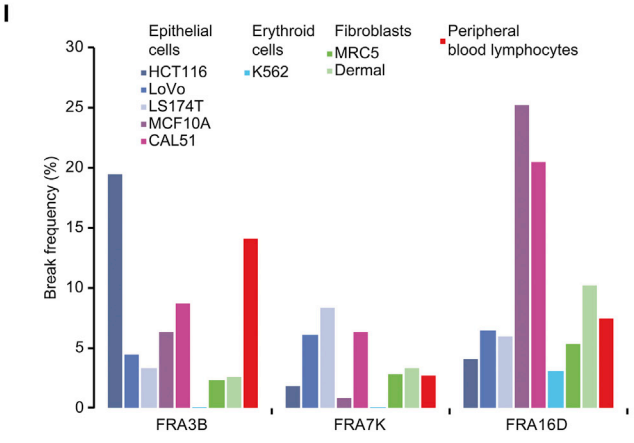


G

Percentage of CFSs shared by 2 cell lines

	HCT116	LoVo	LS174T	MCF10A	CAL51	K562	MRC5	Dm. fibro.	Lympho.
HCT116		38	32	29	40	9	26	22	18
LoVo			44	33	36	15	16	25	15
LS174T				35	29	13	19	21	21
MCF10A					46	15	22	33	24
CAL51						16	24	27	20
K562							9	15	6
MRC5								47	14
Dm. fibro.									11
Lympho.									





(legend on next page)

LoVo), nontumorigenic and cancer breast epithelial cells (MCF10A and CAL51, respectively), and leukemia-derived K562 erythroid cells. We conducted conventional cytogenetic analyses on R-banded metaphase chromosomes from cells treated with aphidicolin, an inhibitor of replicative DNA polymerases. A chromosomal locus was considered to be fragile if breaks at that site represented at least 1% of the total number of breaks. With this threshold, breaks clustered on a dozen sites in each cell line, representing approximately two-thirds of all lesions (Figures 1A–1F; Table S1). These results are in line with those reported for CFSs previously mapped in primary lymphocytes (Mrasek et al., 2010) and fibroblasts of fetal lung or dermal origin (Le Tallec et al., 2011; Murano et al., 1989).

We first compared CFSs newly mapped in the six cell lines (Figure 1G). We found that epithelial cells share, on average, 36% of their CFSs. In contrast, epithelial and erythroid cells have only 14% of their CFSs in common. Cell lines of the same cell type thus tend to share a larger part of their CFSs than cell lines of different cell types. Using the same conditions of cytogenetic analysis as those used above, we found that 45% of the CFSs are conserved between primary lymphocytes from different individuals, which is similar to the conservation of CFSs in primary dermal fibroblasts (48%; Table S2) calculated by reanalyzing the published data (Murano et al., 1989). The group of epithelial cells thus displays differences comparable to the interindividual variability measured in primary lymphocytes and dermal fibroblasts. This suggests that changes accompanying cancer development do not massively impact the CFS setting, although it is possible that mutations in DNA repair genes or deletions in unstable regions may affect the break frequency of some CFSs in those cell lines. Combining our results with data from primary lymphocytes (Mrasek et al., 2010) and fibroblasts (Le Tallec et al., 2011; Murano et al., 1989), we calculate that two different cell types share less than 20% of their CFSs (Figure 1G), which emphasizes that CFS setting is cell-type dependent, i.e., is defined epigenetically.

Interestingly, although CFS repertoires differ extensively from one cell type to another, the comparison of CFSs with a break frequency over 1% in epithelial cells, erythroid cells, fibroblasts, and lymphocytes showed that 21 out of 49 CFSs (43%) are fragile in at least two cell types (Figure 1H). This result suggests that these loci share features predisposing them to fragility. However, in agreement with the epigenetic nature of CFSs, we find that the frequency of a given CFS can vary greatly across the cells in which it is fragile (Figure 1I). For example, FRA16D, which is fragile in all cell lines, accounts for more than 25% of all breaks in MCF10A cells but for only 3% in K562 cells.

CFSs Are Significantly Associated with Genes over 300 kb Long

To decipher which features contribute to fragility, it is necessary to map fragile regions at the molecular level by fluorescence in situ hybridization (FISH). We therefore hybridized over 20,000 metaphase spreads of aphidicolin-treated cells with BAC (bacterial artificial chromosome) probes delimitating candidate regions for 15 CFSs (Table S3). We found that 10 out of these 15 CFSs are associated with one or several large genes over 600 kb long (Table S4), which confirms and extends previous observations that many CFSs overlap genes ranging from 600 kb to more than 2 Mb (McAvoy et al., 2007a). In addition, four out of the five remaining sites are associated with genes from 366 to 582 kb in length (Table S4). This prompted us to determine what is a "large gene" in the context of CFSs and whether the association between CFSs and large genes is significant or occurs solely by chance. Indeed, given their size, large genes extend over a large proportion of the genome. For instance, genes over 600 kb represent only 0.8% of human genes (Figure 2A) but cover more than 5% of the human genome (data not shown). To address these issues, we took into account CFSs mapped molecularly in this study and in previous work (Table S4) and calculated whether the percentage of CFSs associated with genes of a given length could be explained by randomness. Strikingly, this analysis revealed a statistically significant association of CFSs with genes over 300 kb long (Figure 2B; $p = 0.017$). This length threshold is 15 times higher than the median length of human genes (20.9 kb), with genes over 300 kb long accounting for 3.4% of all human genes (Figure 2A).

All CFSs conserved between human and mouse described so far are associated with orthologous large genes, as exemplified by the human *FHIT* gene in FRA3B and the murine *Fhit* in Fra14A2 (Smith et al., 2007). To extend these results, we mapped CFSs by conventional cytogenetics in mouse embryonic fibroblasts (MEFs). We found that all these CFSs reside within chromosome bands hosting large genes (Table S1). This association was confirmed at the molecular level for the three major CFSs in MEFs (Table S4). Moreover, all large genes associated with MEF CFSs have human orthologs, five of which being now associated with human CFSs (Table S1; Smith et al., 2007).

CFSs have only been studied in mammals thus far (Durkin and Glover, 2007), but analysis of sequenced genomes has revealed that large genes associated with human CFSs are conserved in various vertebrates, notably birds. Interestingly, we detected recurrent breaks induced by aphidicolin in DT40 chicken lymphoid cells (data not shown). Strikingly, our molecular mapping revealed that the most fragile region in DT40 cells is

Figure 1. Defining CFS Repertoires of Epithelial and Erythroid Cells

(A–F) Aphidicolin-induced breaks in six human cell lines are shown: three MSI colorectal cancer epithelial cell lines (HCT116, LoVo, and LS174T); nontumorigenic breast epithelial cells (MCF10A); cancer breast epithelial cells (CAL51); and K562 erythroid cells. The names of human CFSs are given according to the nomenclature established in lymphocytes. Breaks at indicated CFSs are expressed relative to the total number of breaks. See Table S1 for details.

(G) The percentage of CFSs shared by different cell lines is presented. MRC5, fetal lung fibroblasts; Dm. fibro., primary dermal fibroblasts; lympho., peripheral blood lymphocytes.

(H) CFSs with a break frequency over 1% in epithelial cells (HCT116, LoVo, LS174T, CAL51, MCF10A), erythroid cells (K562), fibroblasts (MRC5 and dermal fibroblasts), and lymphocytes. The presence of each CFS among the four cell types was recorded. The name of the CFSs identified in one to four cell types is indicated. The total number of different CFSs is 49.

(I) Breakage frequencies of FRA3B, FRA7K, and FRA16D in different cell lines are shown.

See also Tables S1 and S2.

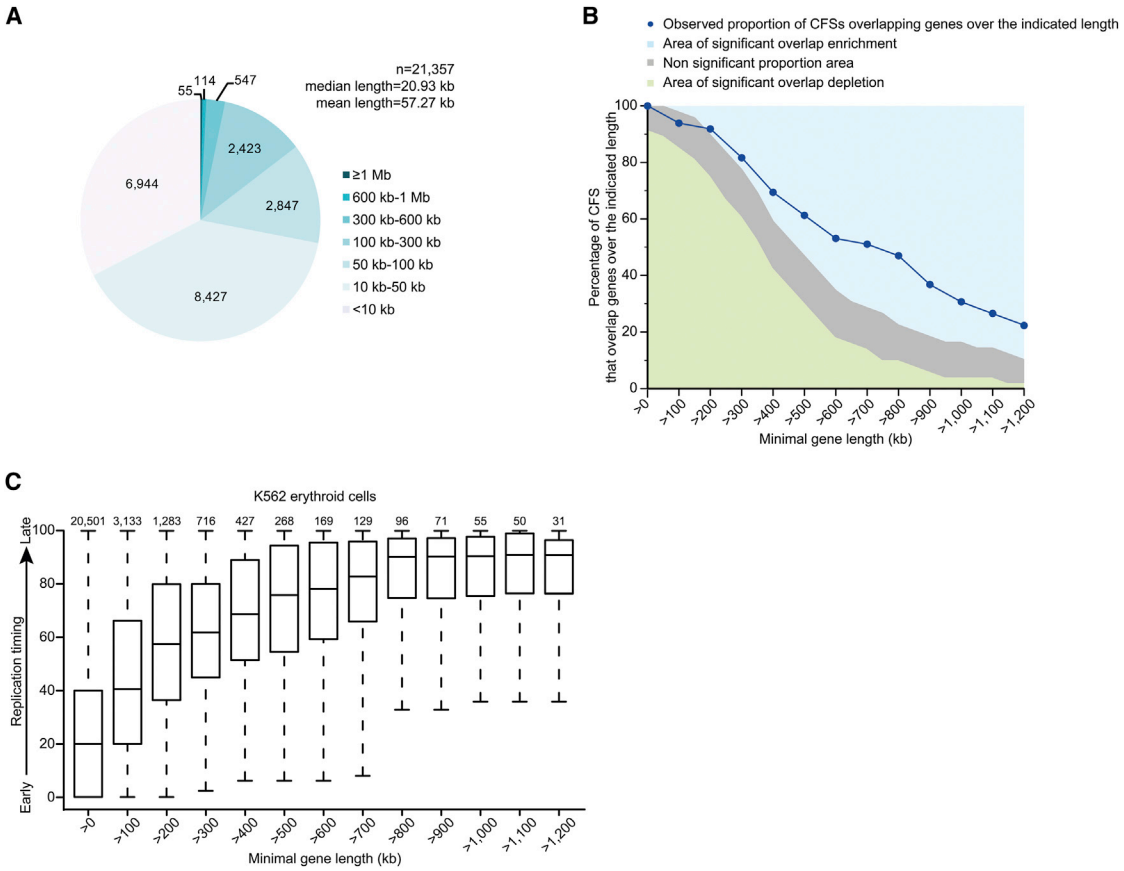


Figure 2. CFSs Are Significantly Associated with Genes over 300 kb Long

(A) The number of human genes according to their length is shown. A total of 169 out of 21,357 genes (0.8%) and 716 out of 21,357 genes (3.4%) extend over 600 and 300 kb, respectively.

(B) Nonrandom association between genes over 300 kb long and the 49 CFSs mapped at the molecular level is presented. Dark-blue dots correspond to the observed percentage of CFSs that overlap genes over the indicated length (genes ">0" correspond to all genes of the genome). The gray area delimits percentages of overlap between CFSs and genes resulting from a random positioning of the 49 CFSs in the genome (for example, CFSs randomly positioned in the genome can overlap between 1.9% and 11.1% of the genes over 1,200 kb long). Thus, blue dots above, in, or below the gray area are the result of a respective significant excess, random number, or significant paucity of CFSs overlapping genes. See the Extended Experimental Procedures for details.

(C) Distribution of replication timing of human genes in K562 cells with respect to a minimal gene length is shown. Replication timing of each gene (vertical axis) was assessed by a score reflecting its average timing of replication along the cell cycle, with 0, 20, 40, 60, 80, and 100 corresponding to G1, S1, S2, S3, S4, and G2, respectively. See the Extended Experimental Procedures for details. Boxes extend from the 25th to the 75th percentiles. Whiskers extend down to the minimum and up to the maximum values. Horizontal black lines represent the median. The number of genes is indicated above each box. Repli-Seq data are not available for some chromosome segments, which explains why the total number of genes is 20,501 instead of 21,357.

See also Figure S1 and Tables S3 and S4.

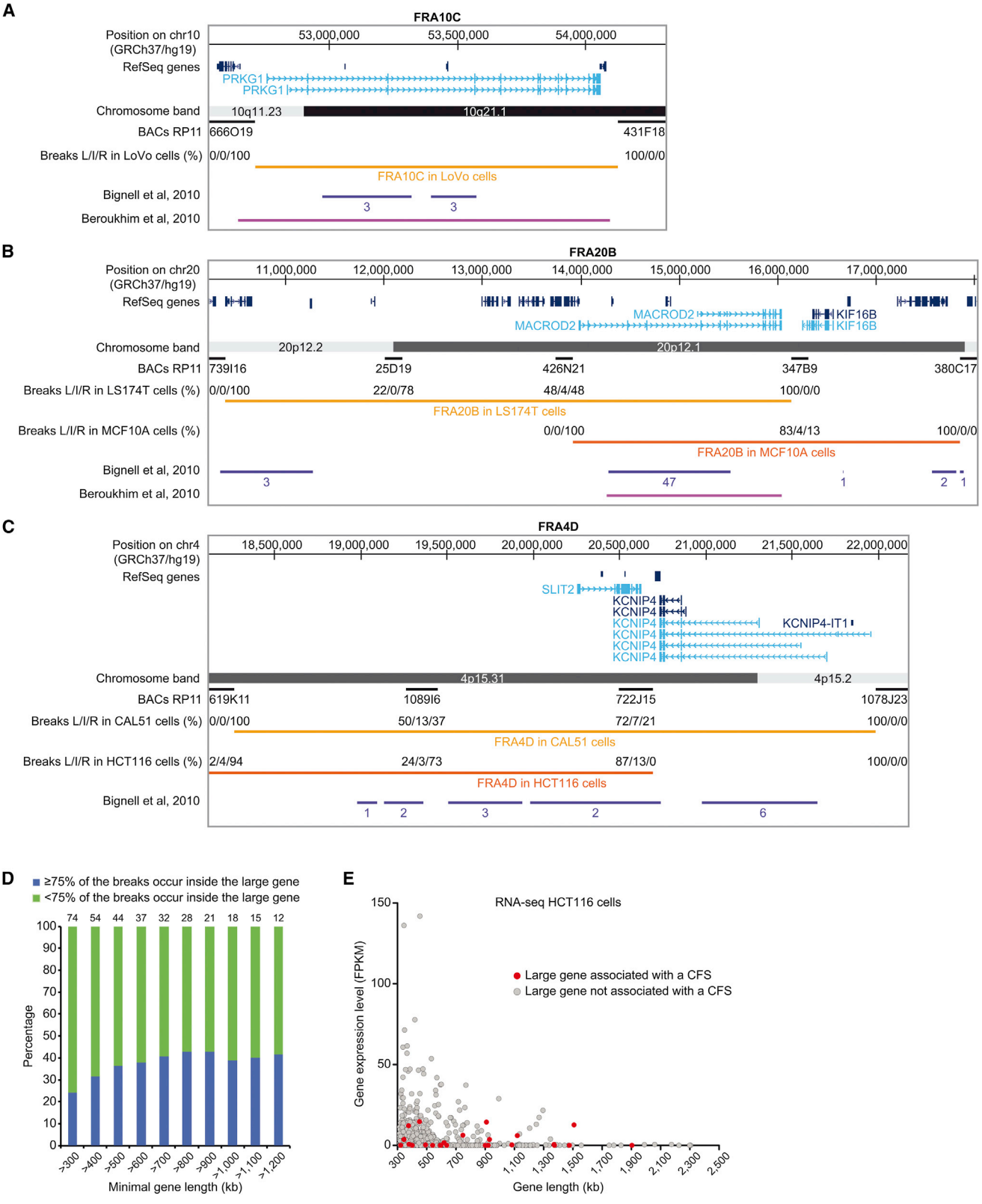
orthologous to human FRA4F and murine Fra6C1, this CFS being associated in the three species with the large genes *CCSER1* and *GRID2* (Table S4; Durkin and Glover, 2007). Our data thus highlight the conservation of human, murine, and avian CFSs and suggest a causal role of large genes in this conservation.

We next asked if large genes have specific characteristics that may predispose them to fragility. Because late replication is a key feature of CFS instability (Debatisse et al., 2012), we first analyzed the replication timing of human genes in three cell types, namely erythroid cells, lymphocytes, and fibroblasts, using available genome-wide timing profiles established by the Repli-Seq technique (Hansen et al., 2010). Strikingly, we observe a strong association between replication timing and gene length, showing that large genes tend to lie in late-replicating domains (Figures 2C, S1A, and S1B). Importantly, such an association is not found in a randomized data set (Figure S1C). We also find that large genes show a high AT content (Figure S1D), a characteristic displayed by most CFSs (Durkin and Glover, 2007).

In conclusion, our results demonstrate that the vast majority of CFSs relate to the presence of genes over 300 kb long. Correlatively, they suggest that all genomic regions containing such genes may be potential CFSs.

Transcription of Large Genes Does Not Dictate Fragility

A recent study of five CFSs associated with large genes has suggested that collisions between replication forks and transcribing



(legend on next page)

RNA polymerase II are responsible for their instability (Helmrich et al., 2011). Surprisingly, whereas this model implies that breaks should be confined to large genes, our molecular mapping of CFSs revealed that fragile regions could be either precisely or only partly nested within the cognate large gene(s) (Figures 3A–3C). For instance, all breaks at FRA10C occur within the 1.3-Mb-long *PRKG1* gene in LoVo cells (Figure 3A), whereas only 48% of the breaks at FRA20B localize within the 2-Mb-long *MACROD2* gene in LS174T cells, the other half being scattered over a 3.5-Mb-long region extending 5′ of the gene (Figure 3B). These two situations were observed independently of the length of the large genes associated with CFSs (Figure 3D). These results suggest that there may be two distinct types of CFSs. However, we repeatedly found that the molecular location of a given CFS could vary between cell types or even between cell lines originating from the same tissue (Figures 3B, 3C, and S2A–S2C). For example, breaks at FRA20B occur almost exclusively within *MACROD2* in MCF10A cells, unlike what is observed in LS174T (Figure 3B). This plasticity in break localization seems hardly compatible with two classes of CFSs and, rather, suggests that transcription units per se do not set the borders of CFSs. To determine whether the collision mechanism could at least account for breaks occurring inside large genes, we compared the frequency of breaks inside the 24 genes over 300 kb long associated with CFSs in HCT116 cells with their expression levels measured by quantitative RT-PCR (qRT-PCR). In contrast to the results from Helmrich et al. (2011), we find no correlation between the mRNA levels of large genes and their instability (Figure S2E). Moreover, the analysis of RNA-seq data produced by the ENCODE Project Consortium (2011) revealed that the vast majority of large genes expressed in HCT116 cells are not associated with CFSs (Figure 3E). Together, our results show that transcription of large genes does not dictate the instability of cognate CFSs.

The Majority of Recurrent Cancer Deletions Originate from CFSs

Two studies have cataloged recurrent focal deletions in large cohorts of human tumors and cancer cell lines (Beroukhim et al., 2010; Bignell et al., 2010). Although a portion of the recurrent deletions has been correlated with the presence of known tumor suppressor genes or with CFSs, 70%–80% of them remain unexplained. We find that the 15 CFSs mapped at the molecular level in our study account for an additional 10% of recurrent cancer deletions (Table S5). Notably, FRA20B overlaps one of the most prevalent clusters of unexplained deletions in human cancers (Figure 3B). These deletions precisely map inside the 2-Mb-long *MACROD2* gene, namely the subregion of FRA20B that we found unstable in different cell types. Other deletions identified by Bignell et al. (2010) lie in regions flanking *MACROD2*, which supports our finding that the fragile region is not confined to the large gene. Although molecular mapping of additional CFSs will undoubtedly extend the number of deletions attributable to CFSs, we reasoned that the association between CFSs and chromosome regions hosting large genes might be reflected in deletions mapped in tumors. We therefore reanalyzed the data provided by Bignell et al. (2010) (Figures 4A, 4C, and 4E) and Beroukhim et al. (2010) (Figures 4B, 4D, and 4F), which first revealed that the proportion of genes overlapping recurrent cancer deletions increases with gene length (Figures 4A and 4B). As illustrated in Figure 4C, 21 out of 28 genes (75%) over 1,200 kb long overlap recurrent deletions mapped by Bignell et al. (2010). Importantly, we observed an extensive overlap between large genes associated with cancer deletions and large genes associated with CFSs molecularly mapped thus far (Figures 4A and 4B, dashed lines, and Figures 4C and 4D). In addition, it has been reported that eight of the ten most frequent focal deletions in human cancers target large genes (Dereli-Öz et al., 2011). Not surprisingly, all but one of these large genes have now been assigned to a CFS (Table S4; Dereli-Öz et al., 2011). Our results therefore suggest that all large genes overlapping cancer deletions are associated with CFSs.

We next determined the minimal length of genes associated with cancer deletions. We calculated that recurrent deletions mapped by Bignell et al. (2010) nonrandomly occur in genomic regions containing genes over 300 kb long (Figure 4E; $p = 4.6 \times 10^{-4}$), which is reminiscent of the results obtained for CFSs (Figure 2B). Strikingly, 56.4% of recurrent cancer deletions take place in regions hosting such large genes (Figure 4E, dark-blue line). Analysis of recurrent cancer deletions identified by Beroukhim et al. (2010) gave consistent results (nonrandom overlap of genes over 500 kb long, accounting for 51.4% of recurrent cancer deletions, $p = 0.017$; Figure 4F). Together, our results thus suggest that the majority of recurrent focal deletions found in human cancers originate from loci hosting large genes that are fragile in the cell types from which the cancer cells derive.

DISCUSSION

Our mapping of CFSs in epithelial and erythroid cells illustrates the diversity of CFS repertoires found in different cell types and in different isolates of the same tissue, which confirms results we obtained previously in lymphocytes and fibroblasts (Le Tallec et al., 2011; Letessier et al., 2011), emphasizing the epigenetic

Figure 3. Transcription of Large Genes Does Not Dictate Fragility

(A–C) Schematic representation of FRA10C (A), FRA20B (B), and FRA4D (C) is shown. From top to bottom: RefSeq genes (dark and light blue indicate genes smaller or larger than 300 kb, respectively); chromosome band; BACs used in FISH experiments (black line) with the frequency of metaphases where the break appeared left (L), inside (I), or right (R) to the hybridization signal; CFS localization (yellow or orange lines); focal deletions in cancer cells (blue line shows singletons and deletion clusters [from Bignell et al., 2010], with the total number of deletions within the cluster indicated; violet line shows recurrent focal deletions [from Beroukhim et al., 2010]). Of note, chromosomal rearrangements have been reported for some cell lines used in this study, for instance HCT116 cells (Alsop et al., 2008). Only loci displaying the expected localization of FISH probes were analyzed.

(D) Break localization relative to genes over 300 kb long associated with CFSs is shown. Genes were divided into two categories depending on whether more or less than three-quarters of the breaks occur inside the gene. The number of genes for each minimal gene length is indicated.

(E) Expression in fragments per kilobase per million reads (FPKM) of genes over 300 kb long in HCT116 cells as measured by RNA-seq is presented.

See also Figure S2.

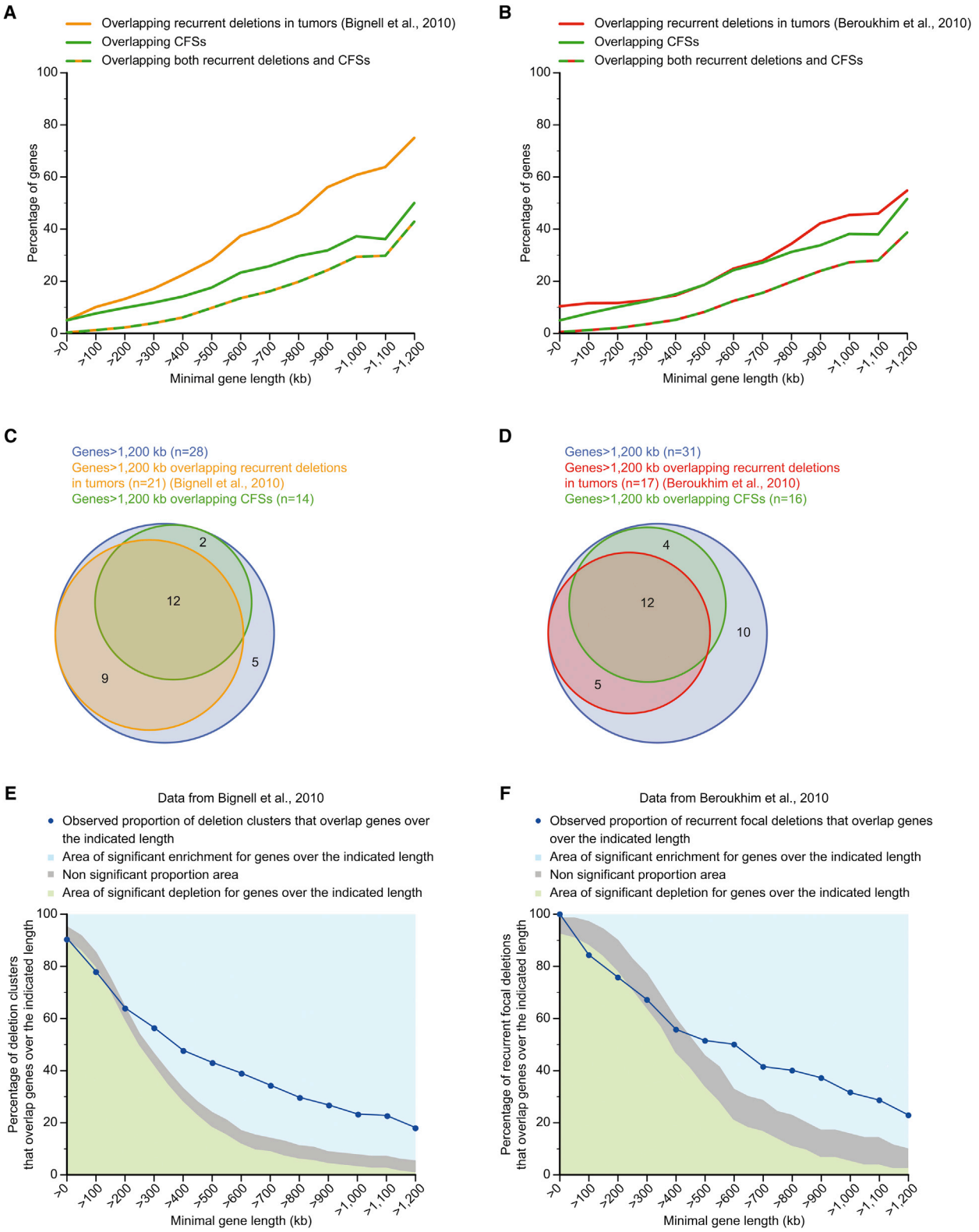


Figure 4. Analysis of the Association between CFSs, Large Genes and Recurrent Deletions Identified in Tumors

(A, C, and E) Analysis of recurrent cancer deletions identified by Bignell et al. (2010) on autosomal chromosomes is presented.

(B, D, and F) Analysis of recurrent cancer deletions identified by Beroukhim et al. (2010) is shown.

(legend continued on next page)

nature of CFSs. In addition, our results show that virtually all CFSs molecularly mapped thus far relate to the presence of large genes, and strongly suggest that CFS conservation in vertebrates relies on the conservation of large genes. We demonstrate here that this association, which has long been described for extremely large genes (Smith et al., 2007), actually occurs nonrandomly for genes over 300 kb long in humans and that CFSs display features that are characteristic of large genes. Our study also provides an example of a CFS associated with a large nonprotein-coding gene (the 545-kb-long *LINC00669* gene in FRA18A in LS174T cells, Figure S2D). Because an increasing number of previously nonannotated RNAs are being cataloged (Djebali et al., 2012), it is possible that the 15%–20% of CFSs devoid of large genes host yet to be identified large transcription units. In conclusion, we propose that chromosome regions hosting genes over 300 kb long constitute the pool of CFSs for all human cell types, a specific subset of these loci being committed to fragility in a given cell type. The human genome contains approximately 700 such genes (Figure 2A). However, we observe that CFSs overlap, on average, 1.5 large genes, which may theoretically decrease the size of the human pool to approximately 450 loci.

The striking association of CFSs with large genes advocates a causal role of those genes in fragility. Surprisingly, we find that the proportion of breaks affecting the large gene itself or its flanking regions is extremely variable from CFS to CFS. We also observe some plasticity in the molecular localization of a given CFS across cell types. Interestingly, deletions in cancer cells mirror the localization of breaks inside or outside the large genes, as exemplified by FRA20B or FRA4D (Figures 3B and 3C). These results indicate that the large gene itself does not set the boundaries of a given CFS. Moreover, in contrast with a previous report analyzing a small number of CFSs (Helmrich et al., 2011), we show that the transcription status of large genes does not dictate the fragility of cognate CFSs. Importantly, we find that even genes larger than 800 kb that require more than one complete cell cycle to be transcribed (Helmrich et al., 2011) are not inevitably committed to fragility when they are expressed. Therefore, although transcription might contribute to the instability of certain CFSs, we propose that fragility is primarily related to chromosomal organization rather than transcription per se. Interestingly, a strong link between replication timing and chromatin domains identified by genome-wide chromatin interaction studies has been observed by Ryba et al. (2010). Large genes, via the association of their transcription regulatory elements, may contribute to organize flexible chromatin domains that govern local replication timing and origin density in a given cell type, two parameters controlling CFS stability (Letessier et al., 2011).

The prevalence of certain CFSs is expected to impact the number of deletions found at the cognate loci in tumors from various origins. Indeed, the observation that FRA16D and FRA3B are fragile in every or virtually every cell type where CFSs have been mapped agrees with studies showing that they are among the top regions of the human genome affected by deletions in cancer cells (Dereli-Öz et al., 2011). The fact that both FRA3B and FRA16D overlap tumor suppressor genes likely also contributes to the selection of cells with deletions in these sites (Iliopoulos et al., 2006; Saldivar et al., 2012). Conversely, CFSs that are found in a limited number of cell types will unlikely show up as major rearranged regions in global analyses of large cohorts of many different cancer classes. However, they can be detected in studies focusing on specific types of tumors as shown recently for FRA1F in bladder cancer (Schepeler et al., 2012). Importantly, like CFSs, cancer deletions are significantly enriched in genes over 300 kb long, and we have demonstrated that the extensive overlap between CFSs and cancer deletions relies on their mutual association with large genes. We have also shown that late replication, one of the most documented features of CFSs, is a characteristic of large genes. Accordingly, it has recently been shown that late-replicating regions are enriched in cancer deletions (De and Michor, 2011). Together, these results strongly suggest that recurrent cancer deletions overlapping large genes originate from CFSs, which explains more than half of recurrent deletions found in tumors. Recently identified early-replicating fragile sites (Barlow et al., 2013) or regions encompassing high densities of negative regulators of cell proliferation (Solimini et al., 2012) could explain, at least in part, the remaining recurrent deletions.

EXPERIMENTAL PROCEDURES

Experimental Procedures and any associated references are available in the Extended Experimental Procedures. See also Figure S3.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, three figures, and five tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2013.07.003.

ACKNOWLEDGMENTS

We thank H. Técher, R. Rothstein, and M. Schertzer for critical reading of the manuscript. We thank the imaging facility PICTIBiSA@BDD for technical assistance. The M.D. team is supported by Institut National du Cancer (INCa) (2009-1-PLBIO-10-IC-1), by Agence Nationale de la Recherche (ANR-09-GENO-000/repinsCFS), and by Association pour la Recherche sur le Cancer (Subvention Libre no. SL220100601348 and Equipements mi-lourds no. 8514). G.A.M. research work is supported by the Département Transfert of the Institut Curie. B.L.T. is supported by a fellowship from INCa.

Received: April 17, 2013
Revised: June 5, 2013
Accepted: July 2, 2013
Published: August 1, 2013

(A and B) The percentages of genes overlapping recurrent deletions in tumors, CFSs mapped at the molecular level, and both with respect to a minimal gene length are illustrated.
(C and D) Venn diagrams of overlap between genes over 1,200 kb long, recurrent deletions in tumors, and CFSs mapped at the molecular level are presented. The number of genes in each class is indicated.
(E and F) Nonrandom association between large genes and recurrent deletions in tumors is shown (see Figure 2B for details).
See also Table S5.

REFERENCES

Alsop, A.E., Taylor, K., Zhang, J., Gabra, H., Paige, A.J., and Edwards, P.A. (2008). Homozygous deletions may be markers of nearby heterozygous mutations: The complex deletion at FRA16D in the HCT116 colon cancer cell line removes exons of WWOX. Genes Chromosomes Cancer *47*, 437–447.

Barlow, J.H., Faryabi, R.B., Callén, E., Wong, N., Malhowski, A., Chen, H.T., Gutierrez-Cruz, G., Sun, H.W., McKinnon, P., Wright, G., et al. (2013). Identification of early replicating fragile sites that contribute to genome instability. Cell *152*, 620–632.

Beroukhim, R., Mermel, C.H., Porter, D., Wei, G., Raychaudhuri, S., Donovan, J., Barretina, J., Boehm, J.S., Dobson, J., Urashima, M., et al. (2010). The landscape of somatic copy-number alteration across human cancers. Nature *463*, 899–905.

Bignell, G.R., Greenman, C.D., Davies, H., Butler, A.P., Edkins, S., Andrews, J.M., Buck, G., Chen, L., Beare, D., Latimer, C., et al. (2010). Signatures of mutation and selection in the cancer genome. Nature *463*, 893–898.

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

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

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

Dereli-Öz, A., Versini, G., and Halazonetis, T.D. (2011). Studies of genomic copy number changes in human cancers reveal signatures of DNA replication stress. Mol. Oncol. *5*, 308–314.

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

Durkin, S.G., and Glover, T.W. (2007). Chromosome fragile sites. Annu. Rev. Genet. *41*, 169–192.

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

Helmrich, A., Ballarino, M., and Tora, L. (2011). Collisions between replication and transcription complexes cause common fragile site instability at the longest human genes. Mol. Cell *44*, 966–977.

Iliopoulos, D., Guler, G., Han, S.Y., Druck, T., Ottey, M., McCorkell, K.A., and Huebner, K. (2006). Roles of FHIT and WWOX fragile genes in cancer. Cancer Lett. *232*, 27–36.

Le Tallec, B., Dutrillaux, B., Lachages, A.M., Millot, G.A., Brison, O., and Debatisse, M. (2011). Molecular profiling of common fragile sites in human fibroblasts. Nat. Struct. Mol. Biol. *18*, 1421–1423.

Letessier, A., Millot, G.A., Koundrioukoff, S., Lachagès, A.M., Vogt, N., Hansen, R.S., Malfoy, B., Brison, O., and Debatisse, M. (2011). Cell-type-specific replication initiation programs set fragility of the FRA3B fragile site. Nature *470*, 120–123.

McAvoy, S., Ganapathiraju, S.C., Ducharme-Smith, A.L., Pritchett, J.R., Kosari, F., Perez, D.S., Zhu, Y., James, C.D., and Smith, D.I. (2007a). Nonrandom inactivation of large common fragile site genes in different cancers. Cytogenet. Genome Res. *118*, 260–269.

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

Mrasek, K., Schoder, C., Teichmann, A.C., Behr, K., Franze, B., Wilhelm, K., Blaurock, N., Claussen, U., Liehr, T., and Weise, A. (2010). Global screening and extended nomenclature for 230 aphidicolin-inducible fragile sites, including 61 yet unreported ones. Int. J. Oncol. *36*, 929–940.

Murano, I., Kuwano, A., and Kajii, T. (1989). Fibroblast-specific common fragile sites induced by aphidicolin. Hum. Genet. *83*, 45–48.

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

Saldivar, J.C., Miuma, S., Bene, J., Hosseini, S.A., Shibata, H., Sun, J., Wheeler, L.J., Mathews, C.K., and Huebner, K. (2012). Initiation of genome instability and preneoplastic processes through loss of Fhit expression. PLoS Genet. *8*, e1003077.

Schepeler, T., Lamy, P., Laurberg, J.R., Fristrup, N., Reinert, T., Bartkova, J., Tropia, L., Bartek, J., Halazonetis, T.D., Pan, C.C., et al. (2012). A high resolution genomic portrait of bladder cancer: correlation between genomic aberrations and the DNA damage response. Oncogene. Published online August 27, 2012. http://dx.doi.org/10.1038/onc.2012.381.

Smith, D.I., McAvoy, S., Zhu, Y., and Perez, D.S. (2007). Large common fragile site genes and cancer. Semin. Cancer Biol. *17*, 31–41.

Solimini, N.L., Xu, Q., Mermel, C.H., Liang, A.C., Schlabach, M.R., Luo, J., Burrows, A.E., Anselmo, A.N., Bredemeyer, A.L., Li, M.Z., et al. (2012). Recurrent hemizygous deletions in cancers may optimize proliferative potential. Science *337*, 104–109.