Deep Insight Section

Common Chromosomal Fragile Sites: Breakages and Rearrangements in Somatic and Germline Cells

Jun Mitsui, Shoji Tsuji

Department of Neurology, Graduate School of Medicine, University of Tokyo, Tokyo 113-8655, Japan (JM, ST)

Published in Atlas Database: July 2011


DOI: 10.4267/2042/46078

This work is licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 2.0 France Licence.
© 2011 Atlas of Genetics and Cytogenetics in Oncology and Haematology

Introduction

Fragile sites have long been identified by cytogeneticists as specific chromosomal regions that exhibit an increased frequency of gaps or breaks when cells are exposed to a DNA replication stress in vitro (Sutherland, 1977). Fragile sites have been classified as rare or common on the basis of their expression frequency within a general population (Schwartz et al., 2006). Rare fragile sites (RFSs) are identifiable in less than 5% of the general population, whereas common fragile sites (CFSs) are intrinsic chromosomal regions that are present in all individuals. Most of the CFSs are induced by aphidicolin, whereas others are induced by bromodeoxyuridine or 5-azacytidine. Since the discovery of CFSs, it has been observed that many of them appear to coincide with aberrant chromosomal regions in various cancers, which has driven numerous studies to isolate genes at CFSs to better understand the molecular basis of their fragility and their implications in the mechanisms of tumorigenesis. To date, as many as 120 CFSs and RFSs have been identified in the human genome (Lukusa and Fryns, 2008). CFSs are a very broad topic that encompasses various fields including cytogenetics, epigenomics, and oncology. In this review, we will first overview the molecular characteristics of CFSs, their implications in tumorigenesis, and several representative genes located at CFSs. We will then focus on the molecular basis of the mechanisms underlying fragility and the repair processes of double-strand breaks involving CFSs based on the analyses of breakpoints. The findings indicate that chromosomal instability associated with CFSs plays an important role, not only in somatic rearrangements associated with cancers, but also in germline rearrangements leading to human hereditary diseases.

I. Molecular characteristics of CFSs

Several CFSs have been cloned and characterized at the molecular level using various methods. Although the mechanisms underlying CFS breakage are still unclear, several factors that may contribute to instability at CFSs have been suggested, including relatively AT-rich regions (Lukusa and Fryns, 2008), late-replicating regions (Handt et al., 2000; Hellman et al., 2000; Le Beau et al., 1998b; Palumbo et al., 2010), high-flexibility peaks (Mishmar et al., 1998; Zlotorynski et al., 2003), regions rich in nuclear matrix attachment regions (Mishmar et al., 1998; Morelli et al., 2002; Wang et al., 1997), and regions located at the interface of G- and R-bands (El Achkar et al., 2005). Notably, CFSs do not show any repeat motifs such as expanded trinucleotide or minisatellite repeats that could predispose them to fragility, as have been demonstrated in RFSs (Schwartz et al., 2006). A recent study on the replication dynamics of FRA3B in human lymphocytes has shown that, rather than breakage being due to replication stalling, the fragility of FRA3B depends on the paucity of initiation events. Even more surprisingly, the fragility of FRA3B is specific to cells showing this particular initiation pattern (Letessier et al., 2011). These findings suggest that the fragility of CFSs does not depend on nucleotide sequences alone, but that CFSs are epigenetically defined chromosomal loci that correspond to the initiation-sparse regions in a given cell type. For historical reasons, CFSs have usually been mapped in lymphocytes, and have been...
considered to be intrinsic characteristics of the chromosomes that are present regardless of the cell type. Therefore, the contribution of CFSs to chromosomal rearrangements should be thoroughly reassessed in cells derived from various tissue types.

II. Implications of CFSs
Since the discovery of CFSs, they have been increasingly recognized to be preferential sites for the integration of exogenous oncogenic DNA viruses (Popescu et al., 1990) and hotspots for chromosomal rearrangements (deletions, duplications, amplifications, or translocations) in various cancers (Lukusa and Fryns, 2008). Hence, it was hypothesized that CFSs can facilitate recombination events that result in clonally expanded cancer cell populations with specific chromosome alterations in specific cancer types. An extension of this hypothesis is that the clonal expansion is driven by damage to genes at CFSs and the consequent loss of expression these genes (Drusco et al., 2011). A counter argument is that, because of the frequent deletions within the CFSs, the loss of any associated gene expression is an unselected passenger event and does not drive the expansive growth of cancer cells (Bignell et al., 2010; Le Beau et al., 1998a). Although this issue should be considered on an individual CFS basis, a considerable amount of data supporting the tumor suppressor potential of FHIT at FRA3B, WWOX at FRA16D, and PARK2 at FRA6E has been accumulated. The biological effect of the loss of function of these genes has been evaluated from various view points.

III. Contributions of CFS genes to oncogenesis
(1). FHIT at FRA3B
FHIT (MIM 601153), located at 3p14.2, spans more than 1.5 Mb encompassing ten exons encoding an open reading frame of 444 bp that is translated into a 16.9 kDa protein (Ohta et al., 1996). FHIT is a histidine triad protein, which represents a small family of nucleotide-binding and hydrolyzing proteins (Hassan et al., 2010). Although the exact biological function of FHIT still remains unclear, it has received the attention of cancer biologists because of its potential role as a tumor suppressor gene. The injection of chromosome 3p14-p12 encompassing FHIT into a renal carcinoma cell line resulted in the partial suppression of tumor growth in nude mice (Sanchez et al., 1994). There have been many studies demonstrating that the overexpression of FHIT significantly inhibited cell growth in various FHIT-deficient cancer cell lines (Ji et al., 1999; Sard et al., 1999). Although spontaneous phenotypes are not observed in a FHIT-deficient mouse model, they showed vulnerability to high-dose-radiation-induced tumor development (Yu et al., 2009) and chemically induced bladder tumor development (Vecchione et al., 2004). FHIT-deficient mice were also used to produce mouse models deficient in multiple tumor suppressors and upregulated oncogenes, Vhl (Zanesi et al., 2005) and Her2 (Bianchi et al., 2007), respectively. These studies showed the tumorigenic effects of FHIT deficiency itself and of the simultaneous deregulation of another cancer-associated gene owing to FHIT deficiency.

(2). WWOX at FRA16D
WWOX (MIM 605131), located at 16q23, spans more than 1.1 Mb encompassing nine exons encoding an open reading frame of 1,245 bp that is translated into a 46.7 kDa protein. The WWOX protein includes two WW domains and a short-chain dehydrogenase/reductase domain homologous to 17β-hydroxysteroid reductase, which may be involved in sex-steroid metabolism (Aqeilan and Croce, 2007). WWOX overexpression induces the inhibition of tumorigenicity of breast cancer cells in nude mice (Bednarek et al., 2001). Several lines of evidence indicate that the overexpression of WWOX significantly inhibits cell growth in various WWOX-deficient cancer cell lines (Fabbri et al., 2005; Iliopoulos et al., 2007; Kuroki et al., 2004; Qin et al., 2006). Although Wwox null mice exhibit a metabolic disorder characterized by hypoglycemia and hypocalcemia, and die at 3-4 weeks of age, the spontaneous occurrence of osteosarcomas in Wwox null mice and lung papillary carcinoma in Wwox heterozygous mice were reported (Aqeilan et al., 2007b). In addition, Wwox heterozygous mice showed vulnerability to chemically induced lung tumors and lymphomas (Aqeilan et al., 2007a; Aqeilan et al., 2007b).

(3). PARK2 at FRA6E
PARK2 (MIM 602544), located at 6q26, spans approximately 1.4 Mb encompassing twelve exons encoding an open reading frame of 1,398 bp that is translated into a 51.6 kDa protein. The PARK2 protein is a ubiquitously expressed ubiquitin E3 ligase that is considered to target specific proteins for proteasomal degradation (Imai et al., 2000), and its mutations are responsible for autosomal recessive juvenile Parkinsonism (AR-JP [MIM 600116]) (Kitada et al., 1998). As FHIT and WWOX have been well established to be associated with tumor suppression genes, PARK2 was also considered to be one of the tumor suppressor genes. Indeed PARK2 overexpression induces the inhibition of tumorigenicity of lung cancer cells in nude mice (Picchio et al., 2004). PARK2 null mice do not develop any tumor spontaneously, but interbreeding of Park2 heterozygous mice with Apc mutant mice results in the acceleration of intestinal adenoma development and increases polyp multiplicity (Poulougiannis et al., 2010).

IV. Germline and somatic mutations in PARK2
As mentioned above, mutations of PARK2 are responsible for AR-JP. Among various causative germline mutations in PARK2, gross deletions account for 50 to 60% of these mutations (Periquet et al., 2001).
with the deletion hotspots clustering in exons 3 and 4 (Hedrich et al., 2004). PARK2 is also frequently targeted by deletions in various cancer cells (Denison et al., 2003b; Toma et al., 2008; Veeriah et al., 2010; Yin et al., 2009). Although it has not drawn much attention, the frequent occurrence of gross rearrangements in the genomic regions corresponding to CFSs in patients with AR-JP suggests that a common basis underlies the frequent occurrence of rearrangements in both germ cell and somatic cell lines. To explore why these particular genomic regions are prone to rearrangements in germ cells and cancer cells, it is essential to determine the precise positions of the regions with breakpoint clusters, and to analyze the junction sequence signatures in detail. The determination of junction sequences, however, has been extremely laborious by conventional methods such as the polymerase chain reaction (PCR)-based genome-walking method, particularly in the case of large rearrangements. To date, only a few breakpoints involving PARK2 have been determined at the nucleotide level in either germ cell or somatic cell mutations (Asakawa et al., 2009; Clarimon et al., 2005; Hedrich et al., 2004). To efficiently determine deletion/duplication breakpoints at the nucleotide level, we have recently applied a custom-designed high-density array comparative genomic hybridization (array CGH) system, which enabled us to determine approximately 300 breakpoints in patients with AR-JP as well as in cancer cell lines (Mitsui et al., 2010). The results of our recent study will be described below in detail.

(1). Rearrangement hot spots
Unrelated 206 patients with AR-JP were analyzed and 268 exonic rearrangements (243 deletions and 25 duplications) and 5 intronic deletions were detected. The nucleotide sequences of the 252 breakpoint junctions were determined, which included 235 deletions and 17 duplications. It was found that 140 of the 252 breakpoints in PARK2 in patients with AR-JP were distinct, indicating that multiple independent rearrangements frequently occurred in PARK2. For comparison of the breakpoints in the germline mutations in patients with AR-JP, we then conducted similar array CGH analyses of PARK2 in 125 cancer-derived cell lines and identified 42 rearrangements (39 deletions and 3 duplications) in 28 of the cancer cell lines (22.4%). Note that the frequencies of rearrangements in PARK2 in cancer cell lines were quite high, supporting the instability of the CFS-associated locus in cancer cell lines. The nucleotide sequences of the 41 breakpoint junctions, including 39 deletions and 2 duplications were determined. Because 10 deletions and 2 duplications were found among multiple cancer cell lines, 32 independent breakpoints (31 deletions and 1 duplication) were determined. Among 32 independent breakpoints, 2 (1 deletion and 1 duplication) were also found in patients with AR-JP, raising the possibility that they were derived from germ cell lines or that the identical rearrangements of germ cell lines independently occurred in somatic cell lines. Intriguingly, in one cancer cell line, 6 independent deletions were observed in PARK2 simultaneously.

We found that the breakpoints were obviously clustered at specific genomic regions in PARK2 and the regions with breakpoint clusters in patients with AR-JP closely coincided with the previously reported region in FRA6E prone to DNA double-strand breaks, which has been referred to as the center of FRA6E (Denison et al., 2003a) (Figure 1). The center of the breakpoint distribution in PARK2 may be similar in germ cell lines and cancer cell lines, but the variance of the distribution may be larger in cancer cell lines than that in germ cell lines. One possible explanation for the difference is that the sample selections for patients with AR-JP biased the breakpoint distributions and the cancer cell lines tended to generate larger rearrangements in that locus owing to increased genomic instability. The Database of Genomic Variants (accessed in March 2010) (Iafrate et al., 2004) included 48 copy number variations (CNVs: more than 1 kb in length) in the regions in PARK2. The distributions of these breakpoints in PARK2 showed similarities with those observed in patients with AR-JP (Figure 1).

(2). Junction sequences
On the basis of the sequences flanking the breakpoints of PARK2, junction sequence signatures were analyzed and then classified into three groups: 1. junctions with extended homologies, 2. junctions with microhomoologies, and 3. junctions without extended homologies or microhomoologies (Figure 2). We refer to short stretches of identical sequences (≤ 8 bp) at breakpoint junctions as microhomologies. A search for extended homologies revealed that 7 of the 162 junctions (4.3%) in patients with AR-JP and 1 of the 32 junctions (3.1%) in cancer cell lines had extended homologies, all of which were embedded in the same repetitive sequences (Alu/Alu).
Figure 1. Cumulative frequency distributions and histograms of breakpoint positions in PARK2.
A. Cumulative frequency distributions of breakpoint positions in PARK2 in patients with AR-JP, cancer cell lines, and control subjects. The horizontal axis represents the nucleotide positions of breakpoints. The vertical axis represents the cumulative frequencies of breakpoints. The upstream breakpoints are shown in white, while the downstream breakpoints are shown in black. Physical maps of PARK2 along with schematic representations of the center of FRA6E are shown above.
B. Histograms of breakpoint positions in PARK2 in AR-JP patients and cancer cell lines. The horizontal axis represents nucleotide positions and the vertical axis represents the number of breakpoints. The numbers of the positions of the upstream (toward the transcriptional initiation site) breakpoints are shown in white, while those of the downstream breakpoints are shown in black. Physical maps of PARK2 along with schematic representations of the center of FRA6E are shown above.
The majority of the junctions were frequently associated with microhomologies; 97 of the 162 junctions (59.9%) in patients with AR-JP and 19 of the 32 junctions (59.4%) in cancer cell lines had microhomologies. Regarding the junctions without extended homologies or microhomologies, it was revealed that 58 of the 162 junctions (35.8%) in patients with AR-JP and 12 of the 32 junctions (37.5%) in cancer cell lines were without extended homologies or identical sequences. Among these, 51 of the 162 junctions in patients with AR-JP and 8 of the 32 junctions in cancer cell lines had inserted sequences. We found that 4 junctions in patients with AR-JP and 2 junctions in cancer cell lines had inserted sequences of more than 19 bp, whose origins were searched for using the BLAST program. It was revealed that 2 inserted sequences in cancer cell lines originated from repetitive sequences (1 Alu and 1 THE1B). The origins of the other inserted sequences remained undetermined.

Notably, there are similarly high frequencies of microhomologies in germ cell and cancer cell lines, which support the notion that a common mechanism underlies the generation of rearrangements in germ cell and cancer cell lines. Consistent with our findings, microhomologies at junctions have recently been observed in the rearrangements in human culture cells experimentally induced using aphidicolin (Arlt et al., 2009). In contrast, rearrangements that can be explained by the homology-dependent nonallelic homologous recombination (NAHR) are relatively rare, because there are only a limited number of rearrangements whose junctions show extended homologies. Considering the observation that multiple independent rearrangements frequently occurred in PARK2 in germ cells, it is in striking contrast to other common genomic disorders such as Charcot-Marie-Tooth disease type 1A (Lupski, 1998) or Smith-Magenis syndrome (Chen et al., 1997) whose recurrent mutations are characterized by homologous recombination and unequal crossing over between the flanking repeat elements. Various mechanisms of rearrangement processes that can result in microhomologies at junctions have been proposed, which include nonhomologous end joining (NHEJ), microhomology-mediated end joining (MMEJ), microhomology-mediated break-induced replication (MMBIR), and/or fork stalling and template switching (FoSTeS) (Figure 2). In eukaryotes, NHEJ is the major repair pathway of DNA double-strand breaks, and functions by ligating their two ends together (Lieber, 2008). It has the potential to ligate any type of double-strand break end without the requirement for an extended homology. Even when starting with two identical DNA ends, NHEJ is a highly flexible process accounting for the diverse breakpoint junctions, with some ends showing short microhomologies (usually 1 to 4 bp) and some ends showing inserted sequences without microhomologies (Lieber, 2008). In addition, it was shown that replication stress leads to the focus formation of key components of the NHEJ pathway (Rad51 and DNA-PKcs) colocalized with markers of DNA double-strand breaks (MDC1 and gamma H2AX), and the down-regulation of the component of the NHEJ pathway (Rad 51, DNA-PKcs, or DNA ligase 4) leads to a significant increase in gaps and breaks at CFSs (Schwartz et al., 2005). MMEJ is another distinctive pathway of end-joining repair, which requires microhomologies of terminal ends in contrast to NHEJ. High frequencies of microhomologies at junctions observed in this study would favor the involvement of MMEJ at CFSs. Recently, the MMBIR and/or FoSTeS model with emphasis on replication fork collapse and/or stalling has also been proposed to explain the origin of rearrangements on the basis of the findings of complex rearrangements and junction sequences showing microhomologies of 2 to 5 bp (Zhang et al., 2009). Because replication mechanisms at CFSs have been implicated to underlie the rearrangements involving CFSs, MMBIR/FoSTeS deserves serious consideration as a possible mechanism underlying the rearrangements at CFSs.

3. Genomic and epigenomic characteristics

On the basis of a recent study of a replication timing map (Woodfine et al., 2005), it was found that one of the latest-replication regions coincided with the regions with breakpoint clusters in PARK2. To investigate flexibility peaks, the regions with breakpoint clusters and the neighboring regions were analyzed. Although flexibility peaks in the breakpoint clustering regions in PARK2 were not overrepresented compared with those in their neighboring regions, there were regions with high AT content (AT repeats) near the regions with breakpoint, and the highest-flexibility peaks flanked the regions with breakpoint clusters. On a high-resolution map of the interaction sites of the human genome with nuclear lamina components, it was revealed that PARK2 was embedded in large lamina associated domains (LADs) (Guelen et al., 2008). This prompted us to investigate the relationships of LADs with other CFS genes including FHIT, WWOX, DMD, GRID2, LARGE, CTNNA3, NBEA, and CTNAP2. Intriguingly, all the CFS genes were embedded in large LADs spanning several Mb. Using the deCODE map, the meiotic recombination rate of the breakpoint clustering regions in PARK2 was found to be high, as previously reported (Asakawa et al., 2009).
Common Chromosomal Fragile Sites: Breakages and Rearrangements in Somatic and Germline Cells

Mitsui J, Tsuji S

Atlas Genet Cytogenet Oncol Haematol. 2010; 14(10)

Figure 2. Schematic representations of rearrangement mechanisms.

A. Junctions with extended homologies are usually explained by nonallelic homologous recombination (NAHR), a form of homologous recombination that occurs between two lengths of DNA sequences (red and pink regions) that have high sequence homologies, but are not alleles.

B. Junctions with microhomologies can be explained by nonhomologous end joining (NHEJ), microhomology-mediated end joining (MMEJ), or microhomology-mediated break-induced replication/fork stalling and template switching (MMBIR/FoSTeS). Schematic representations of MMEJ and MMBIR/FoSTeS are shown.

C. Junctions without extended homologies or microhomologies can be explained by NHEJ.

(4). Involvement of CFSs with rearrangements in germlines leading to human diseases

Although several lines of evidence have demonstrated that somatic rearrangements that occur within CFSs are associated with cancer development, CFSs have rarely drawn attention as genomic structures associated with germline rearrangements. Our study suggests that chromosomal instability associated with CFSs plays an important role in gross deletions and duplications in germ cell lines leading to human diseases. Supporting this hypothesis, we also found that, similarly to PARK2, DMD (MIM 300377), which is embedded in a CFS (FRAXC) (McAvoy et al., 2007), is frequently targeted by gross deletions in patients with Duchenne and Becker muscular dystrophy (MIM 310200 and 300376, respectively) and in those with various cancers (Mitsui et al., 2010). Recently, CNVs in the human genome have been identified in control subjects by various methodologies, including array CGH, single-nucleotide polymorphism genotyping, and massively parallel sequencing (Kidd et al., 2008; Korbel et al., 2007; Mills et al., 2011; Perry et al., 2006; Redon et al., 2006; Sudmant et al., 2011). Because sample selection bias inevitably affects the distributions of germline rearrangements, unbiased knowledge about CNVs distributions in the human genome will also be required to determine whether a common mechanism can underlie CFSs. Such investigations will certainly be essential for a better understanding of the molecular basis of CFSs and human diseases associated with instabilities in the human genome.

External Links

The NCBI Database of Genomic Structural Variation
References

Sutherland GR. Fragile sites on human chromosomes: demonstration of their dependence on the type of tissue culture medium. Science. 1977 Jul 15;197(4300):265-6


McAvoy S, Ganapathiraju S, Perez DS, James CD, Smith D, DMD and IILRPL1: two large adjacent genes localized within a common fragile site (FRA3B) have reduced expression in cultured brain tumors. Cytogenet Genome Res. 2007;119(3-4):196-203


This article should be referenced as such: