Excessive genomic DNA copy number variation in the Li–Fraumeni cancer predisposition syndrome

Adam Shlien*†, Uri Tabori*‡§, Christian R. Marshall*¶‖, Malgorzata Pienkowska*, Lars Feuk*¶‖, Ana Novokmet*§, Sonia Nanda§, Harriet Druker§, Stephen W. Scherer*¶‖, and David Malkin*†‡§**

*Program in Genetics and Genome Biology, and Departments of †Medical Biophysics, ‡Pediatrics, and ‖Molecular Genetics, ¶The Centre for Applied Genomics, and §Division of Hematology/Oncology, Hospital for Sick Children, University of Toronto, Toronto, ON, Canada M5G 1X8

Edited by Joseph F. Fraumeni, Jr., National Institutes of Health, Bethesda, MD, and approved May 27, 2008 (received for review March 26, 2008)

DNA copy number variations (CNVs) are a significant and ubiquitous source of inherited human genetic variation. However, the importance of CNVs to cancer susceptibility and tumor progression has not yet been explored. Li–Fraumeni syndrome (LFS) is an autosomal dominantly inherited disorder characterized by a strikingly increased risk of early-onset breast cancer, sarcomas, brain tumors and other neoplasms in individuals harboring germline *TP53* mutations. Known genetic determinants of LFS do not fully explain the variable clinical phenotype in affected family members. As part of a wider study of CNVs and cancer, we conducted a genome-wide profile of germline CNVs in LFS families. Here, by examining DNA from a large healthy population and an LFS cohort using high-density oligonucleotide arrays, we show that the number of CNVs per genome is well conserved in the healthy population, but strikingly enriched in these cancer-prone individuals. We found a highly significant increase in CNVs among carriers of germline *TP53* mutations with a familial cancer history. Furthermore, we identified a remarkable number of genomic regions in which known cancer-related genes coincide with CNVs, in both LFS families and healthy individuals. Germline CNVs may provide a foundation that enables the more dramatic chromosomal changes characteristic of TP53-related tumors to be established. Our results suggest that screening families predisposed to cancer for CNVs may identify individuals with an abnormally high number of these events.

cancer genetics | p53 | genomic instability | microarray

Li–Fraumeni syndrome (LFS) is a clinically and genetically heterogeneous familial cancer syndrome associated with a diverse spectrum of germline *TP53* mutations (1, 2). In contrast to other familial cancer syndromes, LFS-affected families display a wide array of tumors, including sarcomas of the bone and soft tissue, carcinomas of the breast and adrenal cortex, brain tumors and acute leukemias, among others. The spectrum of reported germline *TP53* mutations is equally diverse and this has complicated efforts to derive a clear genotype-phenotype model for the syndrome. Indeed, even in the same LFS family, affected individuals sharing an identical germline *TP53* mutation develop tumors of varying severity, at different anatomical sites and at different ages (1). This heterogeneity is thought to be due in part to additional germline genetic variations present within and among LFS families. With this in mind, we undertook a genome-wide characterization study of the constitutional genetic variation of LFS family members.

A CNV is a segment of DNA 1 kb or larger that is present in variable copy number in the genomes of humans, primates and potentially many other species (3, 4). Despite efficient repair machinery, CNVs still occur 100 to 10,000 times more frequently than point mutations in the human genome (5). While the precise mechanisms that give rise to most human CNVs are not known, nonallelic homologous recombination (NAHR) and nonhomologous end joining (NHEJ) are thought to be involved (6). A first-generation map of CNVs in the human genome was recently completed, revealing 1,447 variable regions in 270 individuals from the HapMap collection (3). Knowledge of the frequency of CNVs per population is necessary for the characterization of rare disease-associated regions, while knowledge of the baseline number of CNVs per person will aid in identifying individuals with particularly unstable genomes.

The importance of acquired chromosomal changes in tumorigenesis has been established; for example, amplification of the *MYCN* oncogene and deletions of chromosome 1p are major prognostic indicators in neuroblastoma (7). Higher resolution analyses have recently provided clues into the etiology of lung adenocarcinoma and acute lymphoblastic leukemia (8, 9) and exciting new data from genome-wide association studies have implicated specific single nucleotide polymorphisms to susceptibility of many diseases, including prostate and breast cancer (10–14). However, the role of constitutional CNVs in cancer predisposition has not yet been explored. We set out to study the frequency of CNVs per person in apparently healthy individuals and in the LFS cancer-prone population. To our knowledge, this is the first reported genome-wide study of CNVs and genetic susceptibility to cancer.

Results

A cohort of individuals including 500 of European descent and the multiethnic 270 person HapMap collection has been previously assembled and used for studies of copy number variation (3, 15). We used this cohort to establish whether a baseline CNV frequency existed in a healthy population. In our independent analysis, we identified 3,884 CNVs in genomic DNA from these 770 reportedly healthy individuals using Affymetrix GeneChip 250K Nsp microarrays. The European cohort was analyzed on blood-derived DNA and the HapMap cohort on lymphoblastoid cell line derived DNA. Samples were grouped by microarray facility and normalized against members of their group to reduce batch effects. CNVs were then determined using dChip (16). To minimize false positives, we only counted CNVs on autosomal chromosomes comprised of 2 or more underlying single nucleotide polymorphism (SNP) probes.

Many CNVs were found in single individuals while others, such as the CNV at chromosome 10q11.22 identified in 63 persons, were found in numerous individuals, demonstrating the variability of the CNV population frequency. In contrast, the frequency of CNVs per genome appears to be highly conserved: the median number of CNVs detected per person was 3, with 75% of the population having 4 or fewer CNVs (Fig. 1*A*). Moreover, CNV frequency appeared to be independent of ethnicity, as a separate analysis of the Yorubans, Chinese, Japanese and individuals of European descent revealed a similar

Author contributions: A.S., U.T., M.P., S.W.S., and D.M. designed research; A.S. performed research; A.S., A.N., S.N., and H.D. contributed new reagents/analytic tools; A.S., U.T., C.R.M., L.F., A.N., and D.M. analyzed data; and A.S., S.W.S., and D.M. wrote the paper.

Conflict of interest statement: A.S. and D.M. have submitted a patent application on a portion of these findings.

This article is a PNAS Direct Submission.

**To whom correspondence should be addressed. E-mail: david.malkin@sickkids.ca.

This article contains supporting information online at www.pnas.org/cgi/content/full/0802970105/DCSupplemental.

© 2008 by The National Academy of Sciences of the USA

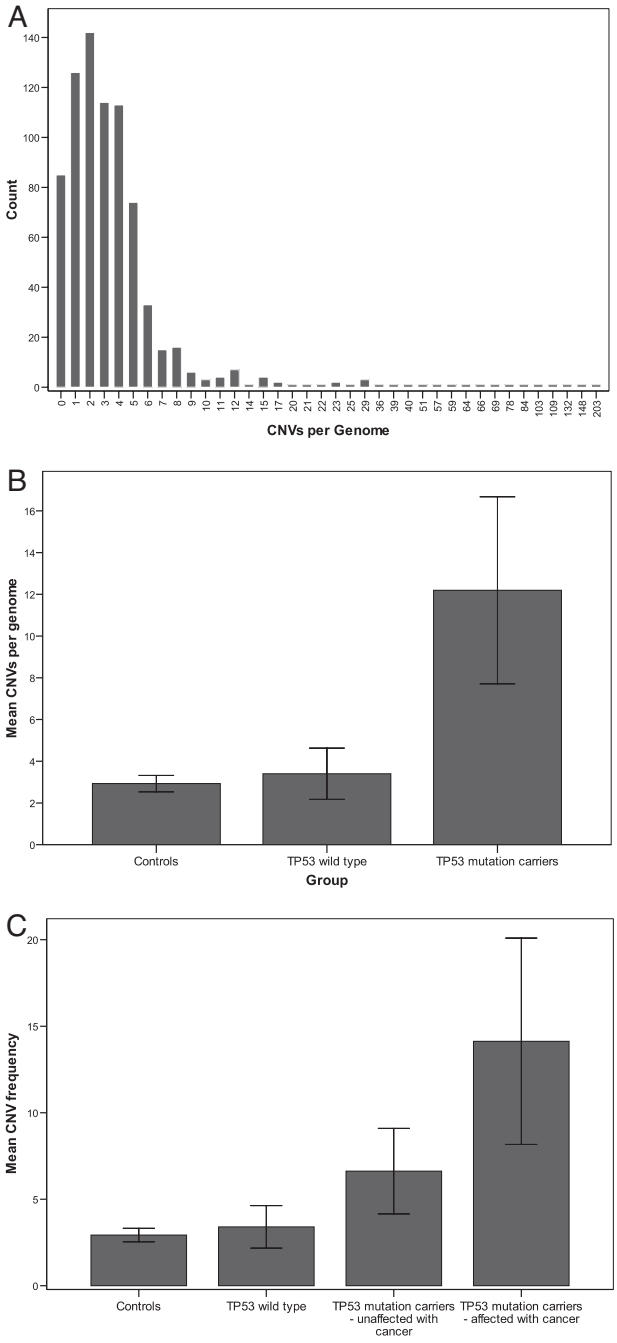


Fig. 1. Increased CNV frequency in LFS. (*A*) To establish a baseline CNV frequency (CNVs per genome), genomic DNA from a large healthy population (n = 770) was assessed for CNVs by using the Affymetrix Nsp SNP microarray. The distribution of CNV frequencies in the normal population is shown. Most individuals have few CNVs (median = 3). Seventy-five percent of the healthy population have four or fewer CNVs. (*B*) A significant increase in CNVs was observed in *TP53* mutation carriers as compared with controls (P = 0.01). The *TP53* wild-type group displayed no significant increase in CNV frequency (P = 0.994). As shown, the mean CNV frequencies are 2.93, 3.40, and 12.19 CNVs per genome in the control (n = 70), *TP53* wild type (n = 20), and *TP53* mutation carrier (n = 31) groups, respectively. Error bars represent SEM. (*C*) Bar graph of CNV frequency in controls (n = 70), *TP53* wild-type individuals (n = 20), *TP53* mutation carriers unaffected by cancer (n = 8), and *TP53* mutation carriers affected by cancer (n = 23). Both the unaffected and affected groups had significantly increased CNV frequencies as compared with controls (P = 0.009 and P = 0.046, respectively). There is also an increase in CNVs in the affected group as compared with the unaffected *TP53* mutation carriers, although not meeting statistical significance because of the loss of power caused by subdividing the group into small cohorts. Error bars represent SEM.

result [supporting information (SI) Fig. S1]. Despite conserved CNV frequencies, the varying size of these deletions and duplications could still result in individuals with different amounts of copy number-variable DNA. To investigate this real possibility we created a simple metric, termed total structural variation, defined as the CNV frequency multiplied by the individual's average CNV size (in bp). The median total structural variation showed a similar degree of conservation and was calculated to be 395 kb, with 75% of the healthy population having 1.1 Mb or less copy variable DNA (Fig. S1). Therefore, this is the first analysis establishing a baseline CNV frequency in the general population.

Having established the distribution and frequency of CNVs in a large reference population, we studied deviations from the global norm in 11 well characterized cancer predisposed LFS families. Inherited *TP53* mutations were observed in 9 families and *de novo* *TP53* mutations in the other two (Table S1). Forty-five family members were evaluated. Eight additional unrelated *TP53* mutation carriers were included for whom DNA samples were unavailable from other family members (Table S1). Of these 53 individuals, 33 were *TP53* mutation carriers and 20 harbored wild-type *TP53*. In addition, 70 unrelated healthy controls were evaluated for CNVs. Both Affymetrix GeneChip 250K Nsp and Sty microarrays were used for all analyses, and validation was performed using two additional CNV detecting algorithms and quantitative PCR (qPCR) (*SI Methods*).

Similar to the large reference population, our controls displayed a median of 2 CNVs per genome, with 75% of the population having 4 or fewer CNVs (mean = 2.93). Additionally, we saw no significant difference in CNV frequency between controls and the *TP53* wild-type group (median = 2, 75th percentile = 3, mean = 3.4). In contrast, the *TP53* mutation carriers displayed a significant increase in CNVs (P = 0.01). This cancer-prone group displayed a mean of 12.19 CNVs per genome with 75% having 10 or fewer CNVs (median = 3; Fig. 1*B*). Of the 33 carriers, 17 exhibited more alterations than the baseline (>3). Remarkably, every LFS family with an inherited *TP53* mutation, except one, contained individuals with CNV counts above the global norm of 3. The majority of specific CNVs in LFS family trios were acquired and not found in either parent (on average twice as common than inherited CNVs) and, among families with a history of cancer, offspring were significantly more likely to have an increase in CNVs when compared with their mutation carrier parent (P = 0.015 by Fisher's exact test, observed/expected ratios: 2.0 for carriers and 0.0 for their wild-type siblings).

Eight of the 11 families studied had histories of cancer. The only families that did not have high CNV frequencies were those that did not have a family history of cancer (3 of 11 families). Of these, two had a single affected proband with a *de novo* *TP53* mutation (Tyr163Cys and His193Pro). The other family had a single affected child who harbored an extremely rare paternally inherited *TP53* mutation (Phe134Tyr).

Many of the *TP53* mutation carriers also had higher total structural variation scores than *TP53* wild-type individuals, which is as one would expect given their numerous CNVs. Less anticipated were individuals found to have few CNVs but high total structural variation scores, as a consequence of exceptionally large deletions or duplications. The most dramatic example found was a paternally inherited 6.1-Mb deletion encompassing 13% of chromosome 21 (21q21.1-q21.2) in an LFS family (shown below the pedigree of family 1 in Fig. 2 as a contiguous faintly-colored vertical bar). The deletion was confirmed by qPCR of DNA derived from blood or normal paraffin-embedded tissue in the absence of available blood ($P < 0.01$ in all cases; Fig. S2). Furthermore, we examined the SNP genotypes in the same region and identified a 6-Mb stretch of homozygosity, which is as expected since the individual has only one allele at this locus. Despite the presence of a germline *TP53* mutation (codon R273S), the hallmarks of the syndrome (strong

GENETICS

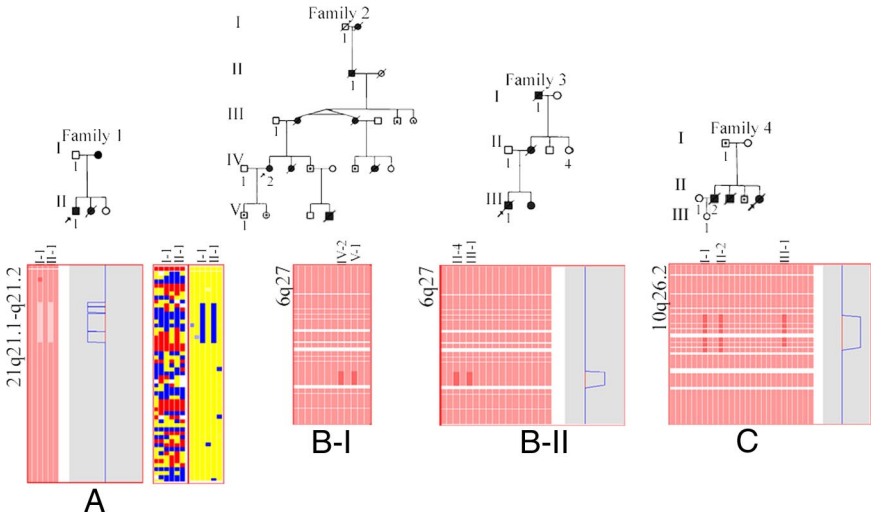


Fig. 2. Inherited deletions and duplications in four LFS families. Three examples of CNVs found in LFS families are shown. The upper portion shows pedigrees for four LFS families, and the lower portion shows the chromosomal size and relative microarray hybridization intensity of each CNV and the family member in whom that CNV was identified. It was not possible to evaluate all members in every pedigree. However, in each of the four families, an affected member, usually designated as the proband (arrow), harbored both the displayed CNV and a *TP53* mutation. In these pedigrees: open circles and squares, healthy females and males; black circles and squares, females or males affected with cancer, respectively; dotted circles or squares, *TP53* mutation carriers who have not yet developed cancer. Oblique lines indicate that the person is deceased. Arrows point to the proband in each family. The lower portion of the figure highlights chromosomal regions of interest undergoing copy number alteration in these four families. In copy number analysis, faint coloring indicates deletions, while duplications are colored more intensely. Vertical columns represent a single individuals' copy number for the region. Individuals of interest from the pedigree are numbered. (*A*) Paternally inherited 6.1-Mb deletion on chromosome 21q21.1-q21.2 (13% of the chromosome), the largest deletion seen in the 893 genomes assessed in this study. Of the three children, two inherited both the deletion and *TP53* mutation (II-1 and II-2). Each developed two neoplasms, first diagnosed at ages 6 and 7. The remaining child (II-3) harbored neither the deletion nor the mutation and is unaffected. The mother (I-2), carrying only the *TP53* mutation, developed a single tumor (fibrous histiocytoma) at age 27. This exceptional deletion is inherited from an as-yet-unaffected individual and is associated with a worsening of the clinical phenotype between generation I and II: the second generation displays the hallmarks of the syndrome (multiple cancers in multiple offspring), whereas the first generation does not. From left to right: the copy number deletion (red), the SNP genotype calls (red, blue, and yellow squares), and loss of heterozygosity (LOH) analysis (blue and yellow) in the same region. There is a concomitant region of homozygous genotype calls for individuals I-1 and II-1. For the same individuals, SNP genotypes are colored red or blue if homozygous, yellow if heterozygous, or white if not called. An extended region of LOH is shown in blue at the right. (*B*) An inherited 240-kb duplication at 6q27, overlapping the leukemia gene *MLLT4*, in four individuals from two LFS families: in B-I, the proband transmitted the duplication of *MLLT4* to her son (V-1, as-yet unaffected carrier), and in B-II the proband inherited the same duplication of *MLLT4* from his mother (II-2, affected and carrier), although this inheritance is presumptive because it could not be ascertained directly in the mother but was found in her sister (II-4, unaffected and noncarrier). The frequency of this CNV is significantly enriched in LFS probands ($P = 0.006$, Fisher's exact test; *SI Discussion*). (*C*) A 574-kb duplication, overlapping the cancer-related gene *ADAM12*, inherited through three generations of an LFS family (*SI Discussion*). The proband's brother (II-2, affected) harbored the paternally inherited duplication at *ADAM12* (I-1, as-yet unaffected carrier), which he transmitted to his daughter (III-1, unaffected and noncarrier).

family history, multiple early onset tumors) are conspicuously absent in the first generation. However, the second generation of the family prominently displays these hallmarks. The full presentation of the syndrome is therefore associated with an increase in copy number variable DNA, although in this instance from an apparently healthy individual (the father I-1). While it is possible that the accelerated clinical phenotype in the affected mutant *TP53* carrier children (II-1 and II-2) and the presence of the 6.1 Mb deletion inherited from the *TP53* wild-type father may be coincidental, an alternative explanation is that the phenotype may have resulted from the effect of an additional genetic modifier effect conferred by the presence of this exceptionally large deletion. The confluence of these two genetic events, high total structural variation and a germline *TP53* mutation, thus correlates with the increase in cancer incidence observed in the family.

Increased CNV frequency was found by comparing individuals at elevated risk for cancer to those at normal risk (*TP53* mutation carriers versus *TP53* wild type). We found no increase in cancer in those individuals that are *TP53* wild type. Although nearly all mutant *TP53* carriers will develop cancer in their lifetime (17), we sought to determine whether CNV frequency may also explain the clinical variability within the *TP53* mutant (at-risk) group. We examined the CNV frequency of *TP53* mutation carriers affected by cancer separately from the unaffected carriers. The unaffected and affected groups each had significantly increased CNV frequencies as compared with controls ($P = 0.009$ and 0.046, respectively). Of particular interest is the presence of an even greater number of CNVs present in those affected by cancer, when compared with those who have not as yet developed cancer. Although not meeting the threshold for statistical significance because of the loss of power caused by splitting this group into small cohorts, this trend suggests a dose-response relationship between CNV frequency and severity of the LFS phenotype (Fig. 1*C*). Whether exposure to chemotherapy influences accumulation of germline structural alterations is not known. However, the fact that blood was drawn before starting therapy in almost all of the patients in this study, and the observation of increased germline CNVs even in those mutant *TP53* carriers who are not yet affected with cancer, suggest that therapy does not contribute to accumulation of germline DNA structural variations (Fig. 1*C*).

We next examined the effect of germline CNVs on the development of somatic chromosomal alterations in paired tumor tissue (Fig. 3*A*). In a separate analysis, DNA was extracted from four frozen tumor samples, taken from individuals whose constitutional CNVs were known, and hybridized on the same platform. Choroid plexus tumors were selected since they frequently occur in the context of LFS. As expected, the tumor DNA contained many structural changes, however, we focused only on those changes that were previously found in normal tissue from the same person. Three of four tumors had loci where germline hemizygous deletions progressed into homozygous deletions in the tumor or where germline duplications became larger in the tumor. Fifteen of 21 loci overlapping germline CNVs became substantially larger (>50%) in paired tumors and in all cases the new somatic alteration was of the same orientation as the germline CNV (i.e., a deletion became a larger deletion in the tumor and an amplification, a yet larger amplification). Because the presence of gross tumor chromosome changes could artificially inflate the observed number of such events, we only selected regions undergoing discrete changes localized to the underlying CNV. This phenomenon was also validated by comparing SNP genotype homozygosity between blood and tumor at these loci (*SI Methods*). One such CNV, a loss at 22q11.23, underwent an additional somatic deletion while the rest of the chromosome maintained diploidy (Fig. 3*B*). Paired blood tumor analysis also revealed a deletion in the tumor sample, indicating that the deletion is located at the same locus and is

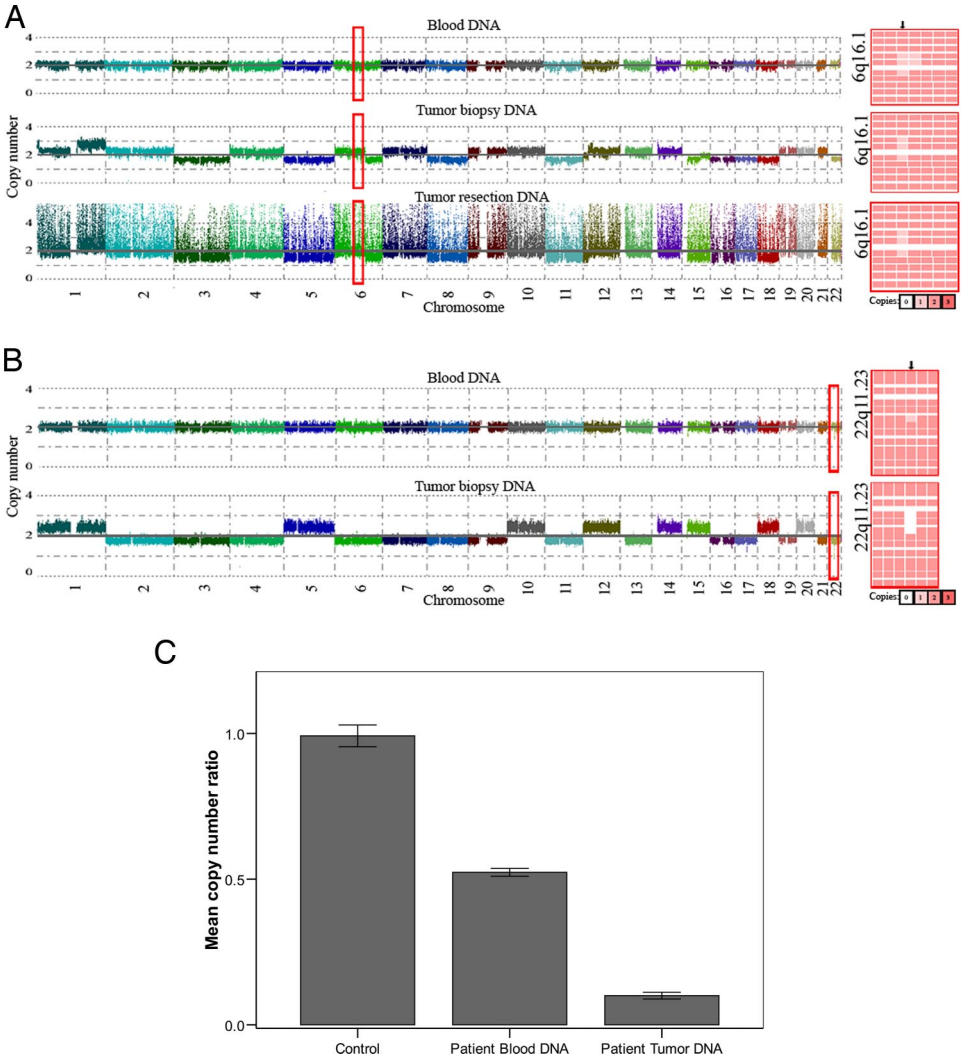


Fig. 3. Progression of germline chromosomal alterations in paired tumor DNA. Shown are copy number alterations in blood DNA and in paired tumor DNA for two individuals (*A* and *B*). Germline CNVs are displayed plotted across all autosomal chromosomes. Immediately below is the copy number of the tumor, which was biopsied or resected from the same person. (*A*) A CNV region (deletion) at 6q16.1 is highlighted on chromosome 6 and enlarged at the right. An arrow points to the patient of interest, and their deletion is indicated by a fainter color. The neighboring column represents the patient's sister, who also has this CNV. Although the patient's tumor genome displays a high level of instability, a deletion identical to the germline CNV was found in both the tumor biopsy and resection as shown below. (*B*) In this person, a 70-kb hemizygous deletion in blood DNA at 22q11.23 is highlighted and displayed enlarged at the right. The same CNV was found to be further deleted in the patient's paired tumor. That is, the remaining allele was lost, as indicated by the yet fainter color, which represents a reduction in the array's signal intensity in the same region. (*C*) The copy number of genomic DNA at 22q11.23 was confirmed by qPCR to be both specific to the underlying CNV and complete. From left: a diploid control, patient blood DNA with a hemizygous deletion, and tumor DNA from the same patient showing a further deletion. The difference in mean copy number between reference, blood, and tumor DNA are highly significant ($P < 0.01$). Error bars represent $\pm$ 2 SEM. qPCR shows that the relative copy number in tumor DNA is 0.1, meaning that >80% of the tumor specimen is homozygous for the deletion (zero copies). We can approximate that only 20% of remaining cells have the germline hemizygous deletion (one copy).

deleted beyond that observed in the patient's blood. qPCR confirmed a one copy loss in the germline as compared with a diploid reference, and at the same locus, a one copy loss in tumor DNA as compared with the germline (Fig. 3*C*). It therefore appears that germline CNVs can act as a basis for more dramatic tumor-specific changes.

DNA rearrangements, such as CNVs, can predispose to or cause disease when they encompass, overlap or disrupt dosage-sensitive genes (6, 18). CNVs can also unmask recessive mutations at dosage-insensitive loci (18). We sought to determine which cancer-related genes fall within copy number variable regions. In both the large reference population as well as the LFS cohort, we observed copy number variability in cancer-related genes. We observed inherited duplications at *MLLT4* and *ADAM12* in LFS families (Fig. 2 *B* and *C*, and *SI Discussion*). *MLLT4* is a target of *Ras* and is fused with *MLL* in the common leukemia translocation t(6, 11)(q27;q23) and the frequency of this CNV is significantly enriched in LFS probands ($P = 0.006$, Fisher's exact test; see *SI Discussion*). *ADAM12* is disintegrin-metalloproteinase, whose dysregulation has been reported in brain, breast, liver, stomach, and colon cancer. The contribution of these CNVs to tumor predisposition, initiation or progression will require further investigation (*SI Discussion* and Table S2).

Discussion

LFS is an ideal model for the discovery of genetic modifiers of cancer and research on these rare families has had a disproportionately large impact on our understanding of cancer biology in general (1, 19–21). The primary reason for this is that defects of *TP53*, the most frequent genetic alteration in LFS, are the most commonly acquired genetic alteration in sporadic human cancer. Non-transformed fibroblasts and lymphocytes from *TP53* mutation carriers display aberrant growth characteristics when passaged in culture, spontaneously acquire properties of a transformed and ultimately immortalized 'tumor cell', and ultimately display mass chromosomal aneuploidy (22). While defective TP53 function is known to cause increased copy number variation and instability in tumors (23–25), our observations in primary non-cultured lymphocytes of *TP53* mutation carriers suggest a new model of carcinogenesis wherein the existence of excessive submicroscopic copy number alterations represent early germline events that may inform the progressive changes required for neoplastic transformation (Fig. 4). These subtle changes are likely the earliest manifestations of instability conferred by the constitutional *TP*53 mutation, which then progress in complexity into events that can be seen by conventional cytogenetic techniques. TP53, as guardian of the genome, actively suppresses cell cycle advance and DNA replication after dsDNA damage, and is involved in the very processes known to give rise to CNVs, including suppression of homologous recombination (26). While the ubiquity and non-random distribution of CNVs in humans highlights genomic regions that are intrinsically unstable, their increased abundance in LFS *TP53* mutation carriers can be explained by germline *TP53* haploinsufficiency. Genomic

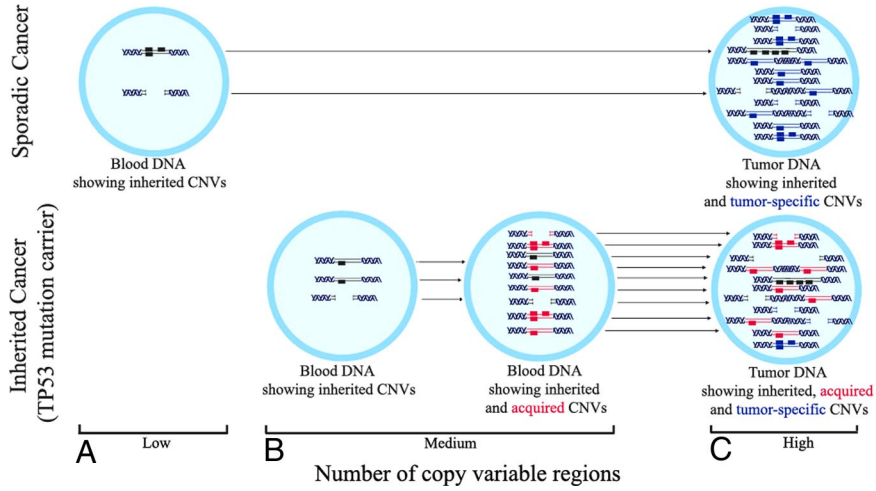
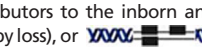
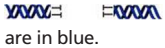


Fig. 4. Proposed model for the progression of copy number variable DNA regions in the Li–Fraumeni cancer predisposition syndrome. Shown is a model of copy number variable DNA regions in patients with sporadic (top row) or inherited cancer (bottom row). (*A*) The total number of CNVs in the genomes of healthy individuals is similar. Non-cancer predisposed individuals have intact DNA repair mechanisms that maintain the number of CNVs close to this baseline (Fig. 1*A*). Despite efficient repair machinery, CNVs still occur 100–10,000 times more frequently than point mutations in the human genome (5). This is largely facilitated by the genomic sequence architecture. The precise mechanisms that give rise to most human CNVs are not known; however, nonallelic homologous recombination (NAHR) and nonhomologous end joining (NHEJ) are thought to be involved (6). Both NHEJ and NAHR are processes by which double-strand (ds) DNA breaks are repaired. The ubiquity and nonrandom distribution of CNVs in humans highlights genomic regions that are intrinsically unstable. (*B*) CNVs are more abundant in Li–Fraumeni cancer predisposed *TP53* mutation carriers because of germline *TP53* haploinsufficiency. TP53, as the ''guardian of the genome'' (26), suppresses cell cycle advance and DNA replication after dsDNA damage. Furthermore, TP53 is involved in the very processes known to give rise to CNVs, including suppressing the level of homologous recombination. Although defective TP53 is known to cause increased copy number variation and instability in tumors (23–25), our data suggests a new model wherein these alterations arise much earlier in cancer-prone individuals. We have observed this increase of CNVs in primary LFS lymphocyte DNA, but this effect may be more dramatic in other cells undergoing rapid remodeling, replicative stress, or in the normal tissue of patients with other cancer predisposition disorders. (*C*) CNVs become fertile ground for changes in cancer. Genomic instability may be preferentially directed toward CNV regions that are hotspots for recombination as suggested by our observation (Fig. 3) that CNVs can act as the genetic foundation on which larger somatic chromosomal deletions and duplications develop in tumors (shown here as arrows from CNVs in blood to those in tumor DNA). Tumor changes may be secondary to an underlying nontumor CNV or arise *de novo* at the same locus. It is likely that the sequence architecture of genomic regions that gives rise to CNVs also facilitates large somatic alterations. In this model, CNVs are seen as crucial regions in both sporadic and inherited tumors. Furthermore, the early age of onset of inherited tumors might be explained by the patient's increased CNV frequency. CNVs should therefore be viewed as important contributors to the inborn and acquired genetic changes that give rise to cancer. CNVs are shown as  (one copy loss),  (two copy loss), or  (one copy gain). Inherited CNVs are represented in black, acquired CNVs are in red, and tumor-specific CNVs are in blue.

instability is a feature of all cancers (27) but it may be preferentially directed toward CNV regions that are hotspots for recombination. As we have observed, this can be accomplished in the tumor genome by expansion of the linear extent of the same allele or by loss of the opposite allele (Fig. 3 *B* and *C*). Our observation that CNVs can act as the genetic foundation on which larger somatic chromosomal deletions and duplications develop in tumors, suggest that CNVs are fertile ground for subsequently acquired changes in cancer cells. Thus, the sequence architecture of genomic regions that give rise to CNVs may facilitate both acquired constitutional as well as tumor-specific genetic changes.

In this study, mutation carriers found to not have high CNV frequencies were also those who did not have family histories of cancer, either because their cancers arose as a consequence of *de novo TP53* mutations or a low-penetrant mutation. It therefore appears that CNV frequency, or another high-resolution measure of instability, may help to define the nature and severity of the germline *TP53* mutations found in LFS families. It will be important in the future to determine the exact patterns of Mendelian inheritance of CNVs in a larger cohort of complete LFS families. It appears that the reason that LFS offspring have a greater CNV frequency is because not only do they inherit CNVs from their parents, but they also acquire *de novo* CNVs. Therefore, the total number of CNVs can be greater than in either parent. Because mutation carriers affected with cancer had more CNVs than unaffected carriers (while both groups harbored more CNVs than individuals carrying wild-type *TP53*), it is tempting to speculate that CNV frequency might also help to categorize *TP53* mutation carriers into "risk groups" and provide a more rational basis for screening and genetic counseling.

With respect to the role of CNVs and sporadic cancer, while similar genome-wide analyses are now routinely performed on tumor samples (8, 9), the frequent lack of matched constitutional DNA means that the germline contribution to the detected somatic alteration cannot be known. Our observation of a surprising number of genomic regions where cancer-related genes coincide with CNVs suggesting that germline CNVs can provide the foundation for somatic chromosomal changes, in both LFS families and healthy individuals, highlights the need for matched analyses in cancer studies and for the establishment of a baseline for structural variation in healthy human genomes.

Our data demonstrate that the CNV frequency is remarkably similar among healthy individuals, but significantly increased in individuals with germline *TP53* mutations. In addition, LFS family members can contain exceptionally large deletions or duplications, as identified by their total structural variation scores. This constitutional structural dynamism may act as the genetic foundation on which larger somatic chromosomal deletions and duplications build, leading to the development of cancer. These findings also establish a method for identifying individuals with constitutional chromosomal instability and inherent susceptibility to cancer.

Materials and Methods

Subject Recruitment. After obtaining written informed consent, DNA was extracted from peripheral blood leukocytes of 53 individuals from families with a germline *TP53* mutation and from 70 unrelated controls. These included 20 *TP53* wild type and 33 *TP53* mutation carriers. Of these, one individual had been diagnosed as a *TP53* mosaic and was grouped with the *TP53* mutation carriers in the CNV analysis. In addition, genomic DNA from five frozen choroid plexus tumors was extracted. DNA was quantified by using a NanoDrop spectrophotometer, and quality was assessed by agarose gel electrophoresis. This study was approved by the Research Ethics Board at the Hospital for Sick Children in Toronto. Subject recruitment for the 500 individuals of European descent and the 270 individuals from the HapMap collection are described elsewhere (28, 29).

DNA Microarray Analysis. Genomic DNA was genotyped with Affymetrix GeneChip Human Mapping 250K arrays (Fig. S3) (30). Samples were restriction enzyme digested, amplified, purified, labeled, fragmented, and hybridized according to the manufacturer's protocol. For the reference samples (n = 770), DNA copy number analysis was performed with dChip (16) by using Affymetrix Nsp CEL files. The LFS case-control cohort (n = 123) was assessed with dChip, CNAG (31), and GEMCA (32) by using Affymetrix Nsp and Sty CEL files. The average call rate in the

LFS hybridizations, which is an indicator of the overall performance of the assay, was 97.5% (SD = 1.6). Two blood DNA samples with >150 CNVs were excluded from the *TP53* mutation carrier group to avoid calling a high number of false positives. The corresponding paired tumor for one these samples was therefore also excluded. The characterization of copy number variation is described in more detail in the *SI Methods*.

Quantitative PCR Validation. Quantitative PCR of genomic DNA copy number was performed by relative quantification on a LightCycler 480 (Roche Applied Science) instrument, using the Roche SYBR green kit. Primers were designed by using Primer3 and the human genome reference assembly (UCSC version hg17, based on National Center for Biotechnology Information build 35). All samples were run in triplicate. Copy number alterations were assessed by relative quantification methods that compensate for differences in target and reference amplification efficiencies. Primer sequences and PCR cycling conditions are detailed in Table S3.

Statistical Analyses. Data were analyzed by using SPSS versions 14.0 and 15.0 (SPSS, Chicago). CNV frequencies were natural-logarithm-transformed and compared by two-tailed, independent-samples *t* tests after assessing for normality by using stem and leaf plots and histograms. A *P*-value of <0.05 was considered to be significant. Levene's test for equality of variances was used to determine when to assume equal variances. To compare the frequency of the cancer-related CNV overlapping MLLT4 (*SI Discussion*), Fisher's exact test was used. Unrelated probands in the LFS cohort ($n = 19$) were evaluated for the CNV and contrasted to unrelated individuals in the reference population ($n = 710$; all children from the Centre d'Etude du Polymorphisme Humain and Yoruban trios were excluded to ensure independent observations).

Computational Assessment of Cancer-Related Genes. Cancer-related genes were selected from the CancerGenes database (33). Genes with zero sources were excluded, yielding a final list of ≈400 known cancer-related genes. Genomic coordinates of CNVs and genes were based on the National Center for Biotechnology Information build 35 reference human genome sequence (Fig. S4). Custom software (available upon request) was used to determine CNVs encompassing or overlapping genes in more than one individual.

***TP53* Mutation Screening.** *TP53* mutations were detected by direct sequencing of exons 2–11 and intron–exon boundaries of PCR products from blood-derived DNA, using an ABI automated sequencer. The primer sequences have been published elsewhere (19).

ACKNOWLEDGMENTS. We thank Kerri Joffe for assistance and proofreading, and Dr. Johanna Rommens for critical review of the manuscript. The Centre for Applied Genomics is acknowledged for technical support. This work was supported by the Genome Canada/Ontario Genomics Institute (S.W.S.), the National Cancer Institute of Canada with funds from the Canadian Cancer Society, the SickKids Foundation, and the Canada Foundation for Innovation (D.M.). A.S. is supported by a Scholarship from the Canadian Institutes for Health Research and a Restracomp Scholarship from the SickKids Foundation. C.R.M. is supported by the National Alliance for Research on Schizophrenia and Depression Mental Health Research Association. L.F. is supported by the Canadian Institutes of Health Research. S.W.S. is supported by a GlaxoSmithKline–Canadian Institutes of Health Research Pathfinder Chair.

1. Malkin D (2004) Predictive genetic testing for childhood cancer: Taking the road less traveled by. *J Pediatr Hematol Oncol* 26:546–548.
2. Malkin D, *et al.* (1990) Germ line p53 mutations in a familial syndrome of breast cancer, sarcomas, and other neoplasms. *Science* 250:1233–1238.
3. Redon R, *et al.* (2006) Global variation in copy number in the human genome. *Nature* 444:444–454.
4. Perry GH, *et al.* (2006) Hotspots for copy number variation in chimpanzees and humans. *Proc Natl Acad Sci USA* 103:8006–8011.
5. Lupski JR (2007) Genomic rearrangements and sporadic disease. *Nat Genet* 39:S43–S47.
6. Inoue K, Lupski JR (2002) Molecular mechanisms for genomic disorders. *Annu Rev Genomics Hum Genet* 3:199–242.
7. Vasudevan SA, Nuchtern JG, Shohet JM (2005) Gene profiling of high risk neuroblastoma. *World J Surg* 29:317–324.
8. Mullighan CG, *et al.* (2007) Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* 446:758–764.
9. Weir BA, *et al.* (2007) Characterizing the cancer genome in lung adenocarcinoma. *Nature* 450:893–898.
10. Camp NJ, *et al.* (2007) Compelling evidence for a prostate cancer gene at 22q12.3 by the International Consortium for Prostate Cancer Genetics. *Hum Mol Genet* 16:1271–1278.
11. Easton DF, *et al.* (2007) Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature* 447:1087–1093.
12. Gudmundsson J, *et al.* (2007) Genome-wide association study identifies a second prostate cancer susceptibility variant at 8q24. *Nat Genet* 39:631–637.
13. Hunter DJ, *et al.* (2007) A genome-wide association study identifies alleles in FGFR2 associated with risk of sporadic postmenopausal breast cancer. *Nat Genet* 39:870–874.
14. Stacey SN, *et al.* (2007) Common variants on chromosomes 2q35 and 16q12 confer susceptibility to estrogen receptor-positive breast cancer. *Nat Genet* 39:865–869.
15. Pinto D, Marshall C, Feuk L, Scherer SW (2007) Copy-number variation in control population cohorts. *Hum Mol Genet* 16(Spec No 2):R168–R173.
16. Lin M, *et al.* (2004) dChipSNP: Significance curve and clustering of SNP-array-based loss-of-heterozygosity data. *Bioinformatics* 20:1233–1240.
17. Nichols KE, Malkin D, Garber JE, Fraumeni JF, Jr, Li FP (2001) Germ-line p53 mutations predispose to a wide spectrum of early-onset cancers. *Cancer Epidemiol Biomarkers Prev* 10:83–87.
18. Feuk L, Carson AR, Scherer SW (2006) Structural variation in the human genome. *Nat Rev Genet* 7:85–97.
19. Tabori U, Nanda S, Druker H, Lees J, Malkin D (2007) Younger age of cancer initiation is associated with shorter telomere length in Li-Fraumeni syndrome. *Cancer Res* 67:1415–1418.
20. Bond GL, *et al.* (2004) A single nucleotide polymorphism in the MDM2 promoter attenuates the p53 tumor suppressor pathway and accelerates tumor formation in humans. *Cell* 119:591–602.
21. Bond GL, Hu W, Levine A (2005) A single nucleotide polymorphism in the MDM2 gene: From a molecular and cellular explanation to clinical effect. *Cancer Res* 65:5481–5484.
22. Bischoff FZ, *et al.* (1990) Spontaneous abnormalities in normal fibroblasts from patients with Li-Fraumeni cancer syndrome: Aneuploidy and immortalization. *Cancer Res* 50:7979–7984.
23. Eyfjord JE, *et al.* (1995) TP53 abnormalities and genetic instability in breast cancer. *Acta Oncol* 34:663–667.
24. Georgiades IB, Curtis LJ, Morris RM, Bird CC, Wyllie AH (1999) Heterogeneity studies identify a subset of sporadic colorectal cancers without evidence for chromosomal or microsatellite instability. *Oncogene* 18:7933–7940.
25. Primdahl H, *et al.* (2002) Allelic imbalances in human bladder cancer: Genome-wide detection with high-density single-nucleotide polymorphism arrays. *J Natl Cancer Inst* 94:216–223.
26. Lane DP (1992) Cancer p53, guardian of the genome. *Nature* 358:15–16.
27. Hanahan D, Weinberg RA (2000) The hallmarks of cancer. *Cell* 100:57–70.
28. Krawczak M, *et al.* (2006) PopGen: Population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet* 9:55–61.
29. International HapMap Consortium (2005) A haplotype map of the human genome. *Nature* 437:1299–1320.
30. Matsuzaki H, *et al.* (2004) Genotyping over 100,000 SNPs on a pair of oligonucleotide arrays. *Nat Methods* 1:109–111.
31. Nannya Y, *et al.* (2005) A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. *Cancer Res* 65:6071–6079.
32. Komura D, *et al.* (2006) Genome-wide detection of human copy number variations using high-density DNA oligonucleotide arrays. *Genome Res* 16:1575–1584.
33. Higgins ME, Claremont M, Major JE, Sander C, Lash AE (2007) CancerGenes: A gene selection resource for cancer genome projects. *Nucleic Acids Res* 35:D721–D726.