Age at cancer onset in germline TP53 mutation carriers: association with polymorphisms in predicted G-quadruplex structures

Charlotte Sagne1,2, Virginie Marcel1,2,5, Maria Bota3, Ghyaslaine Martel-Planche4, Amanda Nobrega2,5,6, Edener Inéz Palmero7, Laury Perriaud1,2, Mathieu Bonioli3, Stephan Wagner2,5, David G.Cox7, Chang S.Chan10, Jean-Louis Mermey11,12, Magali Olivier4, Patricia Ashton-Prolla13,14, Janet Hall1,2, Pierre Hainaut3,6 and Maria Isabel Achatz5,6

1INSERM U612, Bât 110–112, Centre Universitaire, Orsay 91405, France, 2Institut Curie, Centre de Recherche, Bât 110–112, Centre Universitaire, Orsay 91405, France, 3International Prevention Research Institute, 15 chemin du Saquin, Ecully 69310, France, 4Molecular Carcinogenesis Group, International Agency for Research on Cancer, 150 cours Albert Thomas, 69372 Lyon Cedex 08, France, 5Department of Oncogenetics, A.C. Camargo Cancer Center, São Paulo 01508-010, Brazil, 6National Institute of Science and Technology in Populational Genetics Service, Hospital de Clínicas de Porto Alegre, Porto Alegre 90035-1900, Brazil, 7Molecular Oncology Research Center, Barretos Cancer Hospital, Av. Antenor Duarte Vilela, 1331, Barretos, São Paulo 01508-010, Brazil, 8CNRS UMR3348, Bât 110–112, Centre Universitaire, Orsay 91405, France, 9Cancer Research Center of Lyon, INSERM U1052 – CNRS UMR5286, Centre Leon Bérard, 28 rue Laennec, Lyon 69373, France, 10Department of Medicine and Center for Systems Biology, Cancer Institute of New Jersey, University of Dentistry and Medicine of New Jersey, Robert Wood Johnson Medical School, 195 Little Albany St, New Brunswick, NJ 08901, USA, 11INSERM U669, Institut Européen de Chimie Biologie, 2 rue Robert Escarpit, Pessac 33607, France, 12University of Bordeaux, Pessac 33607, France, 13Department of Genetics and Molecular Biology, Federal University of Rio Grande do Sul (UFRGS) and Medical Genetics Service, Hospital de Clínicas de Porto Alegre, Porto Alegre 90035-903, Brazil and 14National Institute of Science and Technology in Population Medical Genetics (INAGEMP), Porto Alegre 90035-903, Brazil

Present addresses:15Cancer Research Center of Lyon, UMR INSERM 1052 CNRS 5286, University Lyon 1, Centre Léon Bérard, 28 rue Laennec, 69373 Lyon Cedex 08, France and 16Institut Albert Bonniot, INSERM U823, University Grenoble-Alpes, Rond-Point de la Chantourne, La Tronche, Grenoble 38700, France

*To whom correspondence should be addressed. Tel: +33 4 72 17 11 87; Fax: +33 4 72 17 11 00; Email: pierre.hainaut@i-pri.org and pierre.hainaut@ujf-grenoble.fr

Germline TP53 mutations predispose to multiple cancers defining Li-Fraumeni/Li-Fraumeni-like syndrome (LFS/LFL), a disease with large individual disparities in cancer profiles and age of onset. G-quadruplexes (G4s) are secondary structural motifs occurring in guanine tracks, with regulatory effects on DNA and RNA. We analyzed 85 polymorphisms within or near five predicted G4s in TP53 in search of modifiers of penetrance of LFS/LFL in Brazilian cancer families with (n = 35) or without (n = 110) TP53 mutations. Statistical analyses stratified on family structure showed that cancer tended to occur ~15 years later in mutation carriers who also carried the variant alleles of two polymorphisms within predicted G4s in the 3′ flanking region remain to be elucidated. In conclusion, this study demonstrates that G4 polymorphisms in haplotypes of the WT TP53 allele have an impact on LFS/LFL penetrance in germline TP53 mutation carriers.

Introduction

Germline mutations in the TP53 tumor suppressor gene (17p13.1, OMIM #191170) predispose to a range of early-onset cancers that define the Li-Fraumeni syndrome (LFS) (1,2). The LFS tumor pattern is dominated by childhood adrenal cortical carcinoma, chordoid plexus carcinoma, medulloblastoma and rhabdomyosarcoma, followed by soft tissue sarcoma, osteosarcoma, premenopausal breast cancer and brain tumors in adolescents and young adults. Several definitions of ‘Li-Fraumeni-like’ (LFL) syndrome have been proposed in families showing only partial LFS traits (3–5). Currently, the only recurrent germline alterations found in patients from families with LFS/LFL traits are TP53 mutations. Depending upon the clinical definition, TP53 mutations are identified in 20–70% of the cases (6–9). In current practice, probands with suspected LFS/LFL are referred for TP53 mutation testing based on a set of individual and familial criteria known as the ‘modified Chompret criteria’ (10). Mutations are detected in ~21–29% of probands matching these criteria (9,11). In TP53 mutation carriers, the penetrance of the disease is ~50% by age 30–35 years and reaches 90% over lifetime. In addition, TP53 mutation carriers show a tendency to develop a wide range of cancers at an earlier age than in the general population, although there are large individual and familial variations in the age at diagnosis. Thus, it is extremely difficult to develop a protocol for prediction and surveillance of the disease. The current practice is to enroll subjects who tested positive for TP53 mutations in complex surveillance programs (12) (NCCN guidelines version 2.2013).

The variable individual cancer patterns suggest that other genetic or epigenetic traits may act as modifiers. Studies by Bond et al. (13) observed that a single-nucleotide polymorphism (SNP) in MDM2, encoding a protein that regulates p53 protein stability, was associated with a reduction of 9 years in the age at diagnosis of first cancer. This effect was detected in several independent cohorts of LFS/LFL families (14–17). This SNP (rs2279744, SNP309) has been shown to alter the sequence-specific DNA binding of the transcription factor Sp1 to the MDM2 promoter, thus resulting in differences in levels of Mdm2 protein expression detected between individuals (13). There is evidence that the effect of rs2279744 might be modulated by another SNP in the MDM2 promoter (rs117039649, SNP285) (18). Borderline effects on age at diagnosis in TP53 mutation carriers have also been reported for a common non-synonymous SNP in exon 4 of TP53, R72P (rs1042522, TP53 PEX4) (15). In a previous analysis on a cohort of TP53 mutation carriers from Brazil, we have shown that a 16 bp duplication polymorphism in intron 3 (rs17878362, TP53 PIN3) was associated with a large difference in the age of first cancer diagnosis, with carriers of two alleles without the duplication (A1 allele) developing their first cancer on average 20 years earlier than heterozygote carriers (A1/A2) (15). However, this Brazilian cohort included a large proportion of subjects from apparently unrelated families who carried the same germline TP53 mutation, R337H, which is common in South-Eastern Brazil due to a widespread founder effect detected in 0.3% of the general population (19). So far, the effect of rs17878362 has not been replicated in an independent LFS/LFL cohort. Thus, it is possible that the effect of the A1 allele of rs17878362 may be due, at least in part, to specific effects on the penetrance of R337H or to other genetic factors in this particular Brazilian background.

Abbreviations: CI, confidence interval; G4, G-quadruplex; LFS/LFL, Li-Fraumeni/Li-Fraumeni-like syndrome; mRNA, messenger RNA; RFLP, restriction fragment length polymorphism; SNP, single-nucleotide polymorphism.
G-quadruplexes (G4s) consist of four-stranded structures occurring in guanine-rich sequences. Potential G4-forming structures are widespread in the genome and have been shown to have important regulatory effects on gene transcription (20), genomic stability (21–23) and DNA replication (24), messenger RNA (mRNA) splicing (25) and mRNA stability (25). Experimentally, drugs that modulate the stability of G4s exert significant effects on the expression (26) and splicing (27) of specific sets of genes. Two G4s have been experimentally demonstrated in the pre-mRNA of p53. A complex G4 structure is found in intron 3 in which the rs17878362 16 bp duplication is located, which regulates the splicing of p53’s pre-mRNA into either the fully spliced form (FSp53) generating the canonical p53 tumor suppressor protein, or into p53I2, a variant mRNA that retains intron 2 (28). This variant mRNA encodes a p53 protein isoform lacking the 39 N-terminal residues (29) that exerts regulatory effects on p53 suppressor activity (28,30–32). Another G4 has been identified downstream of the cleavage/ polyadenylation site of the p53 pre-mRNA. This G4 contributes to p53 pre-mRNA 39 end processing in response to DNA damage (33).

In this study, we have hypothesized that polymorphisms affecting G4 structures may exert modifier effects on the age at cancer onset in LFS/LFL subjects. We have identified, using bioinformatics approach, candidate G4 domains throughout the entire TP53 locus and analyzed the association between SNPs located within or in the close vicinity to these putative and demonstrated G4s and age at first cancer diagnosis in a series of 402 subjects from Brazilian families with LFS/LFL traits.

Materials and methods

LFS/LFL families

LFS/LFL Brazilian subjects were recruited from families attending the Cancer Risk Evaluation clinics in the Department of Oncogenetics of A.C. Camargo Cancer Center (São Paulo, Brazil) and Hospital de Clínicas, Universidade Federal do Rio Grande do Sul (Porto Alegre) (ethics statement number 0568/07). A total of 402 subjects from 145 families were included in the study (Supplementary Table 1, available at Carcinogenesis Online). Families matched at least one LFS/LFL criteria (classic LFS, Birch, Chompret, Eeles criteria), and their familial history was documented over several generations using the Progeny software (Progeny software, Wolfville, Nova Scotia, Canada). Of these families, 15 (10.3%) matched classic LFS criteria, whereas 66 (45.5%) matched Birch or Chompret criteria. Other families matched more relaxed criteria known as the Eeles criteria (8). The coding sequence (exons 2–11 including splice junctions) of TP53 was sequenced in DNA extracted from blood samples (see below). Mutations were found in 35 families (24.1%), including 19 families with the ‘Brazilian founder’ R337H mutation and 16 families with previously reported DNA-binding domain mutations (34) (http://p53.iarc.fr/TP53GermlineMutations.aspx). The institutional ethics committees of participating institutions approved the study and all patients provided informed consent.

Controls

Four groups of control subjects were selected. Group 1 consisted of 487 cancer-free subjects from the general population of Southeast Brazil (age 40–69 years) (34). Groups 2–4 consisted of individuals included as reference samples in the HapMap project. Group 2 (Caucasians) included 30 trios (two parents and a child) of Northern or Western European ancestry. Group 3 (Asians) comprised 90 East Asian subjects (including two groups of 45 subjects from the Tokyo and Beijing areas, respectively). Group 4 (Africans) consisted of 30 trios (two parents, one child) from the Yoruba people in Nigeria (35).

Prediction of G4 motifs and selection of polymorphisms

The TP53 sequence NC_000017.10 (ENSG00000141510) containing the entire TP53 gene sequence plus 500 bp upstream and downstream of the first and last exons, running from bp 1 to bp 15218 (for in silico prediction of G-rich regions likely to form G4 structures using the web-based prediction softwares Quadrifinder (http://miracle.igib.res.in/quadfinder/) and QGRS Mapper (http://bioinformatics.ramapo.edu/QGRS/index.php), each set at its own default parameters (Quadrifinder: G stretch = 3–5; loop size = 1–7; maximum length = 45; QGRS mapper: G stretch = minimum 2; loop size = 0–30; maximum length = 30 (36,37)). A total of five G4 regions were identified by the two softwares. Polymorphisms located in these regions or within 100 bp upstream or downstream of these regions were selected in dbSNP (http://www.ncbi.nlm.nih.gov/snp/). To this set, were added the well-defined TP53 SNPs rs1642785 (G/C, intron 2) and rs1045222 (G/C, exon 4). Overall, a total of 87 SNPs were selected (Supplementary Table 2, available at Carcinogenesis Online and Figure 1).

Analysis of mutations and polymorphisms

DNA was extracted from peripheral white blood cells obtained by venipuncture with the Qiagen DNA Extraction kit according to the manufacturer’s instructions (QiAamp DNA blood Maxi kit; Qiagen, Hilden, Germany). Purified DNA from HapMap subjects were obtained from the Coriell Institute Biorepository as described earlier (35). The TP53 coding sequence (exon 2–11 including flanking intronic regions containing splice sites) was analyzed according to the protocols of the International Agency for Research on Cancer (http://p53.iarc.fr/Download/TP53_DirectSequencing_IARC.pdf). The genotyping of TP53 polymorphisms that fell outside these regions was performed by direct sequencing using primers and PCR conditions described in Supplementary Table 2, available at Carcinogenesis Online. Genotyping for MDM2 rs2279744 was performed using a 59 exonuclease SNP genotyping assay (Proligo, St Louis, MO) (probes FAM 59-cceggecgea-39 and Hex 59-cceggecgea-39; primers 59-tctaggttaagtgacgg-39 and 59-tcaaacgcc-gcaagt-39) or by restriction fragment length polymorphism (RFLP) after PCR with primers and condition as described in Supplementary Table 2, available at Carcinogenesis Online. RFLP was performed using 59 of PCR product, digested with 1 unit of MspAI restriction enzyme (New England Biolabs, Ipswich, UK) during 4 h at 37°C. The results were analyzed on 3% agarose electrophoresis gels. The T allele of rs2279744 generated two RFLP fragments (53 and 158 bp), whereas the G allele generated three fragments (46, 53 and 112 bp). Comparison of the results obtained by the 59 exonuclease and RFLP assays showed complete concordance.

Haplotype reconstruction

The haplotypes carried by Brazilian subjects were reconstructed using pooled sequence variant data of individuals from the LFS/LFL families and the Brazilian control group (Group 1). Rare variants (minor allele frequency < 0.04) were excluded from the analysis. Haplotypes were constructed based on five SNPs (rs1642785, rs17878362, rs1042522, rs17880560 and rs1614984) and the TP53 germline mutation status, using R software haplo.stats package (http://mayoresearch.mayo.edu/mayo/research/schaid_lab/software.cfm).

Statistical analysis

Hardy–Weinberg equilibrium was assessed for each study group using a web-based calculator (http://www.tufs.edu/~mcourt1/Documents/Count%20lab%20-%20HW%20calculator.xls). Statistical analyses were performed using R software (weights package: http://cran.r-project.org/web/packages/ weights/weights.pdf) and running a hierarchical model that takes into account family size to compensate for family bias caused by multiple subjects belonging to the same large pedigree. Briefly, analyses were adjusted for the number of subjects in the family who were sequenced for TP53 mutation and for polymorphisms, thus taking into account possible bias due to specific characteristics of large families in the dataset. Comparisons between average ages at first cancer diagnostic were performed by t-test weighted by family size, or by taking into account family sizes either for the whole cohort (‘All’) or for a particular subgroup analyzed (e.g. only families with defined wild-type (WT) or mutant TP53 status; ‘Group’). Unadjusted Kaplan–Meier disease-free probability estimates were calculated using tools available at http://vassarstats.net. A ‘disease’ event was assigned as the age at first cancer diagnosis (subsequent diagnoses in the same subjects were not taken into account). For each year of age, subjects with no diagnosis were censored at their date of last follow-up.

Results

Distribution of G4s and polymorphisms in TP53

The presence of at least three consecutive guanines along a DNA or RNA strand is (as a first approximation) the classical minimal requirement for intramolecular G4 formation. Longer DNA/RNA sequences containing multiples of four G-tracts can, in principle, accommodate higher order structures defined by multiple G4 blocks, with wide topological and structural variations. The human genome has been predicted to contain up to 376 000 G4s (38). Using in silico prediction programs, we identified five regions predicted to contain G4 structures in a region of 20.2 kb encompassing the entire TP53 sequence (Figure 1A). The sequences and positions of these predicted G4s are given in Figure 1B and C and Supplementary Figure 1, available at Carcinogenesis Online. They are located in the proximal and distal part of intron 1, in intron 3, intron 6 and the 39 flanking region, respectively (Figure 1B and C). Of these predicted G4s, the one located...
rs17878362, rs1788056 and cancer risk

rs17878362, rs1788056 and cancer risk

in intron 3 (28) and another one (among several predicted domains) in the 3′ flank (33), have been previously identified by structural and molecular techniques. A total of 85 polymorphisms reported in dbSNP were located within or near these G4s and their distributions were analyzed in LFS/LFL subjects, Brazilian controls (Group 1) and in the HapMap series (Groups 2–4). Only 11 of these polymorphisms showed allelic variation in these series. Four had a minor allele frequency <4% including rs78378222 that has been associated with basal cell carcinoma (prostate cancer, glioma and colorectal adenoma) in an Iceland population (39). The allele distributions of other polymorphisms (polymorphisms A, B, C, E, F and G) and rs1042522 (TP53_PEX4, codon R72P) in LFS/LFL families and in the four control groups are given in Supplementary Table 3, available at Carcinogenesis Online. The distribution of these polymorphisms was in Hardy–Weinberg equilibrium except for SNPs D and E in Group 2 (HapMap Caucasian, P = 0.04 for both SNPs), SNP G in Group 1 (P < 0.01) and SNPs A and D in LFS/LFL group (P < 0.01 for both SNPs). When compared with Group 1, the deviation from Hardy–Weinberg equilibrium for SNP D (rs1042522) in LFS/LFL appeared to be due to low numbers of carriers of the C/C (72P/72P) allele combination, whereas numbers of G/G homozygotes and G/C heterozygotes were comparable in both groups. The distribution of polymorphisms was different among the three HapMap series (P < 0.01, t-test). The Brazilian LFS/LFL and the controls of Group 1 (Brazilian population) had a SNP profile similar to HapMap Caucasians (P = 0.47, Mann–Whitney t-test).

Association between polymorphisms in and/or near G4s and age at first diagnosis in LFS/LFL

The summary characteristics of subjects included in the LFS/LFL group are given in Supplementary Table 1, available at Carcinogenesis Online. Among the 145 families, 35 had a germline TP53 mutation (MUT group), 19 of which were carriers of the ‘Brazilian founder’ R337H mutation. The main types of cancer diagnosed in carriers of R337H and in carriers of other mutations, and their mean age at diagnosis, are summarized in Supplementary Table 4, available at Carcinogenesis Online. Overall, R337H carriers developed a range of cancer types that is typical of the LFS spectrum; however, with a non-significant tendency for diagnosis at a later age than carriers of other TP53 mutations. Within families with germline TP53
mutations, subjects who were not found to carry the familial mutation were identified as the WT1 group. Among the 110 families without mutations, 157 cases of cancer were diagnosed among 122 subjects. Subjects from these families without a TP53 mutation, either with or without cancer, were identified as the WT2 group. Types of cancer and ages at diagnosis in this group were compatible with LFS/LFL definitions (Supplementary Table 4, available at Carcinogenesis Online).

We next analyzed the mean age of first cancer diagnosis (± standard deviation) in relation to the carriage of the TP53 SNPs A-G and of rs2279744 in the following subgroups: the MUT group (including subjects with any TP53 germline mutation) and the WT2 group using an adjusted model weighted for family size to take into account a possible familial bias. P values (t-test) were adjusted on family size either for the whole LFS/LFL cohort (‘All’) or for the subgroup of patients considered (‘Group’). Effects on age at diagnosis were observed only with polymorphism C (rs17878362, 16 bp insertion/duplication in intron 3) and polymorphism F (rs17880560, also called rs79948390, rs72526905 or rs66470553, 6 bp insertion/duplication in 3’ flanking region). (Table I; data for other SNPs in Supplementary Table 5, available at Carcinogenesis Online). For each of these two polymorphisms, alleles were identified as A1 (non-duplicated) and A2 (duplicated), defining three genotypes, A1/A1, A1/A2 and A2/A2 at each polymorphism. No significant difference was found for the age at first cancer diagnosis within the WT2 group in relation to the status of these polymorphisms (the age at first diagnosis ranged between 33 ± 19 years in subjects with the rs17880560 A2/A2 allele status and 40 ± 16 in subjects with the rs17880560 A1/A1 allele status). In contrast, in the MUT group, there was a tendency for a difference of about 12.5 years between subjects with either A1/A1 or A1/A2 genotypes, the A1/A2 mutation carriers having their first cancer diagnosis later than the A1/A1 mutation carriers. The difference in age at diagnosis in relation to the genotype was significant in a model adjusted for family size in the whole LFS/LFL cohort (‘All’; P = 0.044 for rs17878362, P = 0.02 for rs17880560) but was at best borderline significance in a model adjusted for the size of families only in the MUT group (‘Group’; P = 0.081 for rs17878362, P = 0.068 for rs17880560). Of note, among subjects carrying a mutation, only one was an A2 homozygote carrier of rs17878362 (with a diagnosis of cancer at age 9), whereas one other subject was an A2 homozygote carrier of rs17880560 (with a diagnosis of cancer at age 14). Thus, it was not possible to evaluate separately the association between the homozygote A2 status of these polymorphisms and the age at first cancer onset. Mean ages at diagnosis in relation to the rs1042522 status, given as comparison in Table I, did not show any significant difference in either WT2 or MUT groups in any of the family weighted models. Results for the MDM2 rs2279744 show a non-significant tendency for later age at diagnosis in carriers of G/G as compared with T/T genotype, but no effect in T/G heterozygotes (Supplementary Table 6, available at Carcinogenesis Online).

We repeated these analyses after separating the MUT group in two subgroups, one with R337H carriers and the second with carriers of other germline TP53 mutations (Supplementary Table 7, available at Carcinogenesis Online). Compatible with Supplementary Table 2, available at Carcinogenesis Online, patients with R337H had their first diagnosis on average 10–12 years later than patients with other TP53 mutations. Nevertheless, an effect on age of cancer onset was seen in both mutation groups. In R337H mutation carriers, the difference in age at diagnosis was, on average, 15 years between rs17878362 A1/A1 and A1/A2 carriers, and 11 years between rs17880560 A1/A1 and A1/A2 carriers. In carriers of other mutations, these differences were of around 10 and 19 years, respectively. However, due to the small numbers in each of the groups and lack of power, these differences were not statistically significant.

The effects of rs17878362, rs17880560 and rs1042522 genotypes on Kaplan–Meier disease-free probability estimates are shown in Figure 2. No effect of any of the SNPs was detected on accrual of cancer in the WT2 group. In contrast, in the MUT group, the presence of one A2 allele of rs17878362 or of one A2 allele of rs17880560 was associated with a reduced childhood and adolescent cancer risk. Taking as the reference the estimated risk of cancer at age 25 years in A1/A2 carriers, the relative risk for cancer before or at 25 years is 4.0 [95% confidence interval (CI): 2.40–6.40] in rs17878362 A1/A1 carriers and 3.17 (95% CI: 1.83–5.17) in A1/A1 rs17880560 carriers. Again, no effect was seen for the rs1042522 genotypes. These observations suggest that variant alleles of both the rs17878362 and rs17880560 polymorphisms are associated with a substantial protection against early life cancers in carriers of germline TP53 mutations.

| Table I. Mean age at first cancer diagnosis depending on TP53 polymorphism genotypes and TP53 mutation family |
|---------------------------|---------------------|---------------------|---------------------|---------------------|
|                          | WT2                | MUT                |
|                          | n | Mean | SD | n | Mean | SD |
| rs17878362               |   |      |    |   |      |    |
| A1/A1                    | 90 | 37.91 | 17.06 | 40 | 30.65 | 19.84 |
| A1/A2                    | 29 | 39.14 | 15.22 | 11 | 43.36 | 19.00 |
| A2/A2                    | 2  | 35.50 | 12.02 | 1  | 9.00  | —   |
| P value                  | Group | 0.959 | —   | All  | 0.747 | 0.044 |
|                          | Without | 0.918 | —   |                  | 0.088 | —   |
| rs17880560               |   |      |    |   |      |    |
| A1/A1                    | 59 | 39.56 | 16.43 | 40 | 32.58 | 21.36 |
| A1/A2                    | 43 | 38.19 | 16.35 | 8  | 45.88 | 11.68 |
| A2/A2                    | 12 | 32.67 | 18.60 | 1  | 14.00 | —   |
| P value                  | Group | 0.410 | —   | All  | 0.442 | 0.020 |
|                          | Without | 0.426 | —   |                  | 0.152 | —   |
| rs1042522                |   |      |    |   |      |    |
| G/G                      | 68 | 37.96 | 18.09 | 35 | 33.29 | 21.11 |
| G/C                      | 46 | 37.76 | 14.87 | 16 | 35.63 | 19.89 |
| C/C                      | 7  | 42.86 | 9.63  | 2  | 19.50 | 14.85 |
| P value                  | Group | 0.569 | —   | All  | 0.609 | 0.545 |
|                          | Without | 0.742 | —   |                  | 0.582 | —   |

WT2: WT subjects in WT families; MUT: mutant subjects in mutant families; Group: P value calculated with weight in the analyzed group. In bold, significant P value; All: P value calculated with weight in the all cohort. In bold, significant P value; Without: P value without weight. In bold, significant P value.

G corresponds to the arginine and C to the proline variants.

810
Fig. 2. Genotype-dependent Kaplan–Meier disease-free probability estimates in LFL/LFS family members with or without TP53 mutations. Kaplan–Meier probability is shown for rs17878362 (A), rs17880560 (B) and rs1042522 (C). In each panel, the left panel corresponds to subjects of the MUT group (TP53 mutation carriers) and the right panel to subjects of the WT2 group (families with no mutation detected). The tables under the graphs show disease-free probability estimates at different ages (10, 30, 50 and 65 years) according to genotype. Only probabilities up to 65 years are shown. *: G corresponds to the arginine and C to the proline variants of rs1042522.
Linkage disequilibrium studies show that rs17878362 and rs17880560 are in strong linkage disequilibrium in all groups analyzed (P = 0.024). To determine whether the two alleles may exert effects either alone or in combination, we reconstructed TP53 haplotypes in subjects of the MUT and WT2 groups using data for the five of the polymorphisms genotyped in this study, plus, in the MUT group, the germline TP53 mutation (Supplementary Table 7, available at Carcinogenesis Online). We next used the genotypes at rs17878362 and rs17880560 to define eight haplotypes (with and without a TP53 mutation). Of these haplotypes, all were represented in the WT2 or MUT series, namely A1-A1 (carrying non-duplicated forms of rs17878362 and rs17880560), A1-A2 (non-duplicated for rs17878362, duplicated for rs17880560), A2-A1 (duplicated for rs17878362 and non-duplicated rs17880560) and A2-A2 (duplicated for both rs17878362 and rs17880560). The ‘Brazilian founder’ R337H mutation was exclusively carried on the A1-A1 haplotype. Of the 16 families with other TP53 germline mutations, 12 carried the mutation on the same, A1-A1 haplotype. Given the predominance of this haplotype as the carrier of the germline mutation, we reasoned that the modifier effect of TP53 polymorphisms might be due to variations in the haplotype of the remaining WT allele, which shows much larger genetic diversity than the mutant haplotype. Figure 3 shows Kaplan–Meier disease-free probability estimates in MUT and WT2 groups, given one ‘fixed’ A1-A1 TP53 haplotype (carrying a germline mutation in subjects of the MUT group and a WT TP53 in subjects of the WT2 group) and considering the effect of the other TP53 allele. This analysis showed that, in patients where the haplotype carrying the germline mutation was A1-A1, the presence of either A1-A2 or A2-A1 on the remaining WT allele was associated with a delay in the age at first cancer diagnosis (Table II and Figure 3). A significant difference was observed in mean age at first diagnosis in the MUT group in relation to the haplotype status of the WT allele after weighting for family size (‘Group’: P = 0.019; ‘All’: P = 0.035) (Table II). Importantly, none of the individuals with a WT A1-A2 haplotype, and only one with a WT A2-A1 haplotype, had developed cancer by age 25 years among carriers of a germline mutation, whereas 25% (95% CI: 15–38) of subjects with a WT A1-A1 haplotype had developed cancer by that age. Thus, the presence of a WT allele carrying a duplication of either rs17878362 or rs17880560 appears to exert a strong protective effect toward early cancer occurrence in this cohort of patients with germline TP53 mutations.

Discussion

LFS/LFL is a heterogeneous familial predisposition syndrome with risk of multiple cancers and ~90% penetrance over lifetime in carriers of a germline TP53 mutation. However, there are considerable interindividual and familial variations in cancer patterns and age at disease onset, raising the hypothesis that a number of other genetic, epigenetic or lifestyle factors may affect the course of the disease. Guanine-rich tracts involved in G4 motifs may represent important targets for genetic polymorphisms with significant functional effects. The enrichment of G4 motifs in key chromosomal regions has suggested a functional role for these motifs in genomic regulation. Data collected from human SNP databases indicated that guanine triplets involved in G4s are more conserved and less polymorphic than their neutral counterparts (40). However, a recent analysis of SNPs occurring within predicted G4 motifs in the promoter of 48 genes has demonstrated a strong correlation between G4 SNPs and expression of the corresponding gene at the individual level (41). These observations support the hypothesis that G4 SNPs may play significant roles in the mechanisms of variations in gene expression among individuals. In this study, we have systematically searched the sequence of the TP53 locus for areas that may form G4s and we have analyzed polymorphisms in these areas in search of genetic variations that may modulate disease onset in TP53 mutation carriers and therefore account for at least part of individual and familial variations in LFS/LFL.

Among five regions predicted to form G4s in TP53, our results identified two polymorphisms, rs17878362 and rs17880560, with a frequency of above 4% that appeared to modulate the age at first cancer onset in TP53 mutation carriers, but not in subjects with LFS/LFL traits that do not carry a mutation in TP53. rs17878362 occurs within a previously demonstrated G4 located in intron 3 (28,33) and...
rs17878362, rs1788056 and cancer risk

---

**Table II.** Mean age at first cancer diagnosis depending on TP53 haplotypes and TP53 mutation family status

<table>
<thead>
<tr>
<th>rs17878362</th>
<th>rs17880560</th>
<th>WT2</th>
<th>MUT</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1</td>
<td>A1</td>
<td>42</td>
<td>35</td>
</tr>
<tr>
<td>A1</td>
<td>A2</td>
<td>35</td>
<td>18</td>
</tr>
<tr>
<td>A2</td>
<td>A1</td>
<td>18</td>
<td>3</td>
</tr>
<tr>
<td>A2</td>
<td>A2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Group</th>
<th>Mean</th>
<th>SD</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>0.869</td>
<td>0.035</td>
<td>0.019</td>
</tr>
<tr>
<td>Without</td>
<td>0.800</td>
<td>0.138</td>
<td></td>
</tr>
</tbody>
</table>

One allele, corresponding to the frequent allele (rs17878362-A1 and rs17880560-A1), is fixed and the haplotype frequencies correspond to the second allele. WT2: WT subjects in WT families; MUT: mutant subjects in mutant families; Group: P value calculated with weight in the analyzed group. In bold, significative P value; All: P value calculated with weight in the all cohort. In bold, significative P value; Without: P value without weight. In bold, significative P value.

---

rs17880560 overlaps with a predicted G4 in the 3’ flanking region of the pre-mRNA of p53. Interestingly, the 3’ flanking region contains a succession of several G4s. Of these, one has been experimentally demonstrated (33) but is distinct from the predicted G4 structure overlapping with rs17880560. However, the presence of a G4 overlapping with the sequence of rs17880560 remains to be demonstrated using biophysical or molecular biological methods.

Both the rs17878362 and rs17880560 polymorphisms consist of duplications introducing additional runs of guanines, which may change the composition, position, structure or stability of the G4s. For both G4 regions, effects were seen in association with the duplicated A2 alleles. Compared with the non-duplicated A1 alleles, the presence of one A2 allele of either rs17878362 or rs17880560 retarded by ~15 years (range: 12.5–18.0 years) the age at first cancer diagnosis in TP53 mutation carriers. As a comparison, we used a large group of subjects from families with cancer patterns corresponding to LFS/ LFL traits but in whom no familial TP53 mutation had been identified.

Neither of the two polymorphisms had a detectable impact on age at first cancer diagnosis or on cancer-free survival in subjects without mutations. To assess the statistical significance of the effects observed in mutation carriers, we performed a t-test on average age at diagnosis, using a model weighted for family structure and number of members in each family, in order to rule out bias due to family-specific traits that may influence the clinical presentation of LFS/LFL. Although this approach decreases statistical power, it allows the clarification of effects that are solely due to the presence of particular allele combinations. Using this method, we detected that each of the polymorphisms had a borderline significant effect on age at first cancer diagnosis ($P = 0.082$ for rs17878362; $P = 0.067$ for rs17880560). However, the magnitude of the effect was large (12.5–18.0 years) and separated subjects with cancer diagnosis in childhood or adolescence from subjects with diagnosis in adulthood. Taking as reference the risk of developing cancer before the age 25 years in carriers of one A2 allele of either polymorphism, the relative risk of early cancer in mutation carriers was 4.0 (95% CI: 2.40–6.40) for rs17878362 A1/A1 and 3.17 (95% CI: 1.83–5.17) for rs17880560 A1/A1. By comparison, the common non-synonymous SNP at codon 72 (rs1042522, TP53 PEX4, R72P) had no effect on age at first cancer diagnosis ($P = 0.864$).

Given that rs17878362 and rs17880560 are in linkage disequilibrium, we reconstructed TP53 haplotypes to estimate the joint effects of these polymorphisms. We made the assumption that the effects of these polymorphisms were associated with the haplotype of the WT allele, rather than the mutant allele. This assumption is based on the observation that, in our series of Brazilian families with TP53 mutation, the mutation is almost systematically located on a haplotype carrying the A1 alleles of both rs17878362 and rs17880560. Thus, the effect of alleles carrying A2 haplotypes, if any, is associated with the properties of the residual WT allele and with its capacity to compensate the functional defect caused by the mutant allele. Using this framework, we found that presence of haplotypes defined by either A1 rs17878362–A2 rs17880560 or A2 rs17878362–A1 rs17880560 had a significant effect on delaying the age at first cancer diagnosis. Among 91 documented subjects carrying a germline mutation, only 1 of 29 having one of these two haplotypes developed cancer before 25 years of age (3.4%), whereas 15 of 57 having the WT A1-A1 haplotype developed cancer by this age (26.3%). We therefore conclude that, when the mutation is present on the A1-A1 haplotype (the most common haplotype in all populations, in particular Caucasians and Asians), the presence of one A2 variant of either rs17878362 or rs17880560 on the residual WT allele is protective against the risk of cancer in childhood and adolescence. This effect is compatible with a simple model in which, in the presence of a mutant allele with partial or total loss of function, TP53 suppressor activity is contributed by the residual WT allele (Figure 4). Thus, WT haplotypes carrying one A2 allele of either rs17878362 or rs17880560 would define ‘strong’ TP53 haplotypes, contributing WT TP53 activity sufficient to suppress cancer until at least early adulthood, whereas A1-A1 WT haplotypes would be ‘weak’ TP53 haplotypes, unable to compensate the loss of function of the mutant haplotype and thus being permissive for early cancer onset.

The present study differs from previous ones on modifier polymorphisms in TP53 by several characteristics. First, we have used a weighted model to take into account family size and control for
possible bias cause by specific familial traits. Second, we have used as reference a group of LFS/LFL families with no mutation (WT2) rather than a heterogeneous group of ‘wild type’ subjects pooling the members of TP53 families who have not inherited the mutation (and therefore should be considered as low risk for cancer) and members of families with no mutations (some of them likely to be at high risk due to germline mutation in yet unidentified gene(s) other than TP53). Third, our series is constituted of LFS/LFL families recruited in Brazil, whereas previous studies were based on Western European or North American cohorts (9,13,14). A common mutant haplotype, R337H, is present in 0.3% of the population of South-Eastern Brazil and represents over half of the germline TP53 mutations in Brazilian families with LFS/LFL traits (19,35). This mutation is carried by the same haplotype in all families and is considered to result from a widespread founder effect. Since its initial identification in the germline of children with adrenal cortical carcinoma (42), there has been debate as whether R337H should be considered as predisposing to LFS/LFL or only to specific tumor types. Our results on LFS/LFL families have shown that carriers of R337H are prone to develop a spectrum of cancers that covers the whole LFS/LFL spectrum, although the average age at first diagnosis for adolescent and adult cancers tend to be up to 10 years later than in carriers of classic, DNA-binding domain TP53 mutations commonly associated with LFS (19). In the present study, we can rule out that the modifier effects of rs17878362 and rs17880560 polymorphisms are restricted to this ‘Brazilian founder’ allele because we observed effects of similar type and amplitude in Brazilian carriers of “classic” TP53 mutations, who have cancer patterns and age at onset identical to Caucasian TP53 mutation carriers from North America or Western Europe (data not shown).

The molecular mechanisms responsible for the association of the A2 allele of rs17878362 or rs17880560 with late age at first cancer onset are a matter of conjecture. The G4s containing rs17878362 (intron 3) has been shown to exert a regulatory effect on the alternative splicing of intron 2 and on the generation of the predicted polymorphic G4 in the 3’ flanking region remain to be demonstrated.

To conclude, this study has identified polymorphisms associated with G4 structures, which modify the penetrance of germline TP53 mutation in LFS/LFL. They have a general effect on the mean age at first cancer diagnosis but, more importantly, they distinguish patients who are at high risk of developing childhood/adolescent cancers from those who develop mainly, if not exclusively, cancers during adulthood. These findings require evaluation and replication in independent LFS/LFL cohorts but may provide an important marker for developing personalized strategies for surveillance of cancer risk in carriers of germline TP53 mutations.

Supplementary material

Supplementary Figure 1 and Tables 1–7 can be found at http://carcin.oxfordjournals.org/

Funding

French National Cancer Institute (INCa 2009–192 ‘TP53 intron3’ to J.H., P.H. and J.-L.M.); research in Inserm U612 is also supported by funding from Institut Curie and Inserm; C.S. has a PhD fellowship from the French Ministry of Research and VM. and L.P. were supported by funding from EU FP7 (249689) for the network of excellence DoReMi (low dose research toward multidisciplinary integration); INSERM (D.G.C.); Conselho Nacional de Pesquisa (CNPq 40.0949/2005–9 to P.P.) and Susan G Komen for the Cure (POP 0403033 to P.P.). Research on G-quadruplex is supported by Fondation ARC, Région Aquitaine and ANR program “Quarpediem” (J.-L.M.).

Conflict of Interest Statement: None declared.

References

rs17878362, rs1788056 and cancer risk


Received August, 27, 2013; revised November 13, 2013; accepted November 18, 2013