RESEARCH ARTICLE

CANCER GENOMICS

Heritable defects in telomere and mitotic function selectively predispose to sarcomas

Mandy L. Ballinger[1,2]†, Swetansu Pattnaik[1,2]†, Piyushkumar A. Mundra[1,2], Milita Zaheed[3], Emma Rath[1], Peter Priestley[4,5], Jonathan Baber[4,5], Isabelle Ray-Coquard[6], Nicholas Isambert[7], Sylvain Causeret[7], Winette T. A. van der Graaf[8], Ajay Puri[9], Florence Duffaud[10], Axel Le Cesne[11], Beatrice Seddon[12], Coonoor Chandrasekar[13], Joshua D. Schiffman[14], Andrew S. Brohl[15], Paul A. James[16,17], Jean-Emmanuel Kurtz[18], Nicolas Penel[19], Ola Myklebost[20,21,22], Leonardo A. Meza-Zepeda[22], Hilda Pickett[23], Maya Kansara[1,2], Nicola Waddell[24], Olga Kondrashova[24], John V. Pearson[24], Andrew P. Barbour[25], Shuai Li[26,27,28,29], Tuong L. Nguyen[26], Diane Fatkin[2,30,31], Robert M. Graham[2,30], Eleni Giannoulatou[2,32], Melissa J. Green[33,34], Warren Kaplan[1,2], Shyamsundar Ravishankar[1], Joseph Copty[1], Joseph E. Powell[1,35], Edwin Cuppen[4], Kristel van Eijk[36], Jan Veldink[36], Jin-Hee Ahn[37], Jeong Eun Kim[37], R. Lor Randall[38], Kathy Tucker[3], Ian Judson[8], Rajiv Sarin[39], Thomas Ludwig[40], Emmanuelle Genin[40], Jean-Francois Deleuze[41], the French Exome Project Consortium‡, Michelle Haber[42], Glenn Marshall[42,43], Murray J. Cairns[44,45], Jean-Yves Blay[6], the International Sarcoma Kindred Study‡, David M. Thomas[1,2]*

Cancer genetics has to date focused on epithelial malignancies, identifying multiple histotype-specific pathways underlying cancer susceptibility. Sarcomas are rare malignancies predominantly derived from embryonic mesoderm. To identify pathways specific to mesenchymal cancers, we performed whole-genome germline sequencing on 1644 sporadic cases and 3205 matched healthy elderly controls. Using an extreme phenotype design, a combined rare-variant burden and ontologic analysis identified two sarcoma-specific pathways involved in mitotic and telomere functions. Variants in centrosome genes are linked to malignant peripheral nerve sheath and gastrointestinal stromal tumors, whereas heritable defects in the shelterin complex link susceptibility to sarcoma, melanoma, and thyroid cancers. These studies indicate a specific role for heritable defects in mitotic and telomere biology in risk of sarcomas.

Common and rare genetic variation currently explains less than 50% of the familial relative risk for cancer, leaving the majority of heritability unexplained (*1*, *2*). Nonetheless, genetic studies focused on common epithelial cancers have yielded major insights into the biological mechanisms underpinning specific cancer susceptibilities, exemplified by homologous recombination in breast cancer and mismatch repair in colorectal cancer (*3*, *4*). Studies of different cancer populations may therefore yield further insights into cancer biology. Differing fundamentally from epithelial cancers, sarcomas are rare connective tissue malignancies that arise predominantly from embryonic mesoderm and affect a younger population (*5*). Because of their rarity, they have been relatively understudied to date at the population level. For the most part, genetic studies into sarcomas have used either familial linkage approaches or genome-wide association studies (GWAS). Studies of rare sarcoma-associated syndromes have led to discovery of key cancer genes, such as the canonical tumor suppressor, *TP53* (*6*). Whole-exome sequencing (WES) or whole-genome sequencing (WGS) are now being used to catalog rare variants in known genes (*7*, *8*). In principle, combining both population- and family-based WGS approaches could uncover additional genes and pathways by integrating statistical methods with clinical information. In this study, we undertook a comprehensive, population-based, case-control study using WGS to identify penetrant genes and pathways that may explain sarcoma risk.

Results

Clinical findings

In total, 1644 sarcoma probands were recruited from sarcoma clinics in this international multi-institutional study, regardless of family history (tables S1 and S2). Subjects had a median age at first cancer diagnosis of 47 years, and 49 years at first sarcoma diagnosis. Soft-tissue sarcomas constituted 78.2% of diagnoses (Table 1 and table S3). Multiple primary cancers were common, including breast (n = 77), melanoma (n = 37), second connective tissue tumors (n = 37), nonmelanoma skin (n = 35), prostate (n = 23), colorectal (n = 21), and thyroid cancers (n = 17). Because radiation is a known risk factor for sarcomas (*9*), we determined the relationship with therapy for a prior cancer. In 110 (38%) of the 293 individuals with multiple primary cancers, the sarcoma was the first tumor diagnosed. In 183 individuals where sarcoma was the second tumor diagnosed, less than one-third (56) occurred within a prior radiation

[1]Garvan Institute of Medical Research, Sydney 2010, Australia. [2]St Vincent's Clinical School, University of New South Wales, Sydney 2010, Australia. [3]Hereditary Cancer Centre, Prince of Wales Hospital, Sydney 2031, Australia. [4]Hartwig Medical Foundation, 1098 XH Amsterdam, Netherlands. [5]Hartwig Medical Foundation Australia, Sydney 2000, Australia. [6]Department of Adult Medical Oncology, Centre Leon Berard, University Claude Bernard, 69373 Lyon, France. [7]Centre Georges Francois Leclerc, 21000 Dijon, France. [8]Sarcoma Unit, The Royal Marsden NHS Foundation Trust, London SW3 6JJ, UK. [9]Department of Orthopedic Oncology, Tata Memorial Hospital, Mumbai, Maharashtra 400012, India. [10]la Timone University Hospital, 13005 Marseille, France. [11]Gustave Roussy, 94805 Villejuif, France. [12]Sarcoma Unit, University College Hospital, London NW1 2BU, UK. [13]Liverpool University Hospitals NHS Foundation Trust, Liverpool L7 8XP, UK. [14]Division of Pediatric Hematology/Oncology, Department of Pediatrics, Huntsman Cancer Institute, University of Utah, Salt Lake City, UT 84112, USA. [15]Sarcoma Department, Moffitt Cancer Center, Tampa, FL 33612, USA. [16]The Sir Peter MacCallum Department of Oncology, University of Melbourne, Melbourne 3010, Australia. [17]Parkville Familial Cancer Centre, Peter MacCallum Cancer Centre and Royal Melbourne Hospital, Melbourne 3000, Australia. [18]Centre Hospitalier Universitaire Hautepierre, 67200 Strasbourg, France. [19]Centre Oscar Lambret, 59000 Lille, France. [20]Western Norway Familial Cancer Centre, Haukeland University Hospital, 5021 Bergen, Norway. [21]Department of Clinical Science, University of Bergen, 5007 Bergen, Norway. [22]Institute for Cancer Research, Oslo University Hospital, N-0424 Oslo, Norway. [23]Children's Medical Research Institute, The University of Sydney, Westmead 2145, Australia. [24]QIMR Berghofer Medical Research Institute, Brisbane 4006, Australia. [25]Faculty of Medicine. The University of Queensland, Brisbane 4072, Australia. [26]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, Faculty of Medicine, Dentistry and Health Sciences, University of Melbourne, Melbourne 3010, Australia. [27]Centre for Cancer Genetic Epidemiology, Department of Public Health and Primary Care, University of Cambridge, Cambridge CB1 8RN, UK. [28]Precision Medicine, School of Clinical Sciences at Monash Health, Monash University, Clayton 3800, Australia. [29]Murdoch Children's Research Institute, Royal Children's Hospital, Parkville 3051, Australia. [30]Molecular Cardiology Division, Victor Chang Cardiac Research Institute, Darlinghurst 2010, Australia. [31]Cardiology Department, St Vincent's Hospital, Sydney 2010, Australia. [32]Computational Genomics Division, Victor Chang Cardiac Research Institute, Sydney 2010, Australia. [33]School of Psychiatry, University of New South Wales, Sydney 2052, Australia. [34]Neuorscience Research Australia, Sydney 2031, Australia. [35]UNSW Cellular Genomics Futures Institute, University of New South Wales, Sydney 2052, Australia. [36]Department of Neurology, University Medical Centre Utrecht Brain Center, Utrecht University, 3584 CX Utrecht, Netherlands. [37]Department of Oncology, Asan Medical Centre, Seoul 05505, South Korea. [38]Department of Orthopaedic Surgery, University of California, Davis Health, Sacramento, CA 95817, USA. [39]Cancer Genetics Unit, ACTREC, Tata Memorial Centre, Mumbai, Maharashtra 410210, India. [40]Université de Brest, Inserm, EFS, UMR 1078, GGB, CHU de Brest, 29200 Brest, France. [41]Centre National de Recherche en Génomique Humaine, Institut de Génomique, 91057 Evry, France. [42]Children's Cancer Institute, Lowy Cancer Research Centre, University of New South Wales, Kensington 2033, Australia. [43]Kids Cancer Centre, Sydney Children's Hospital, Randwick 2031, Australia. [44]School of Biomedical Sciences and Pharmacy, The University of Newcastle, Callaghan 2308, Australia. [45]Centre for Brain and Mental Health Research, The Hunter Medical Research Institute, Newcastle 2305, Australia.

***Corresponding author. Email: d.thomas@garvan.org.au**

†These authors contributed equally to this work.

‡International Sarcoma Kindred Study and French Exome Project Consortium authors and affiliations are listed in the supplementary materials.

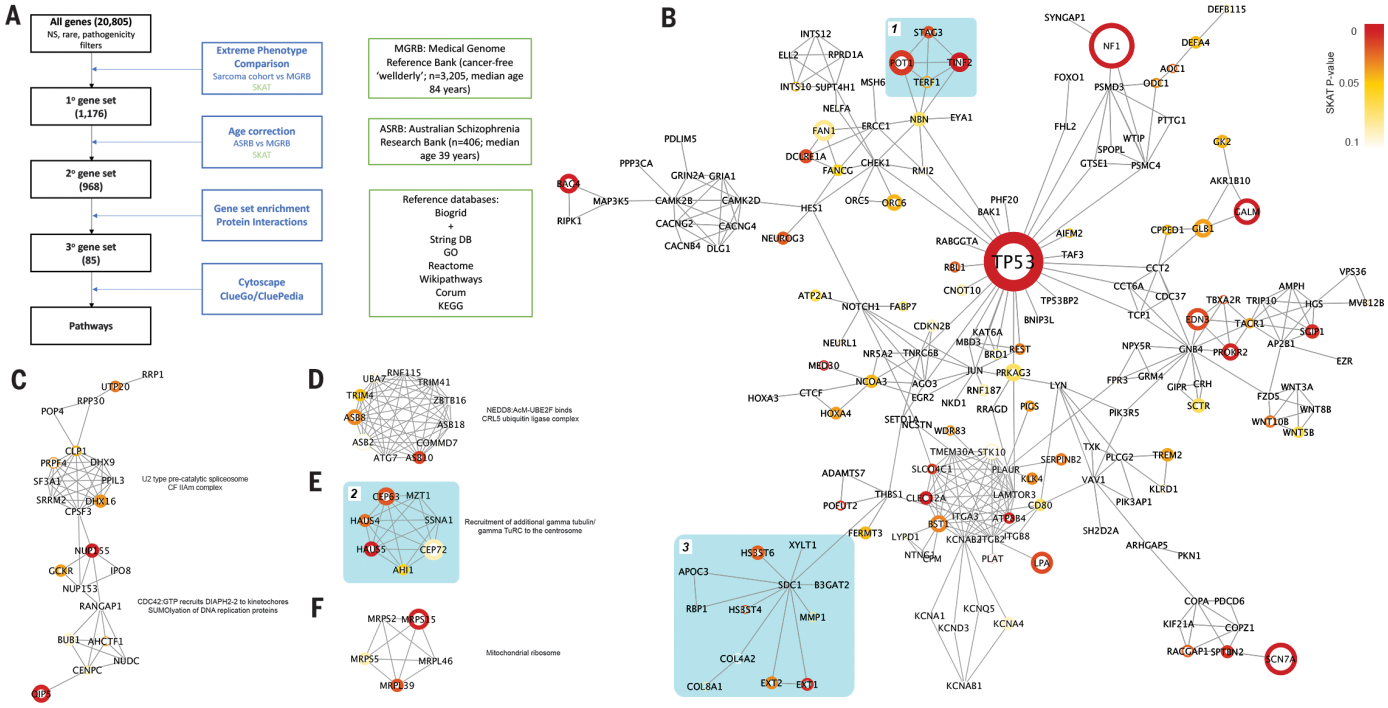

Fig. 1. A systematic analysis of genes and pathways implicated in sarcoma. (**A**) A schematic of the analytic approach of progressively enriching for genes and pathways carrying an excess burden of pathogenic variation in sarcoma probands compared to a universal control cohort (MGRB), including correction for age-depletion of nonsarcoma-related genes. Graph representations of the 968 genes (nodes) in the secondary gene set and their interactions (edges). The analysis was performed in Cytoscape. Node color represents the *P*-value of WRVBT enrichment for each gene in the sarcoma probands relative to the MGRB, the size of the node represents the excess weighted burden of pathogenic variants in the sarcoma probands relative to MGRB, and the width of the node border represents protein interaction and ontologic enrichment by a combined rankscore. Specific groups (1 to 3) relevant to subsequent analyses are highlighted (pale blue boxes). (**B** to **F**) Specific clusters are as follows: (B) Graph representation of the super cluster of 187 genes (nodes) in the secondary geneset centered on *TP53* and their interactions (edges); (C) A 22-gene cluster representing ontologies linked to RNA processing and activity in the M phase of the cell cycle. (D) Eleven genes implicated in antigen processing, ubiquitination, and neddylation. (E) Seven genes defined by centrosomes, spindle organization, and the G_2/M phase of the cell cycle. (F) Five-gene clique with a role in mitochondrial translation and ribosomes.

field. Of 60 women who developed a sarcoma after breast cancer, in more than half (32), the sarcoma arose outside the thorax. These data suggest that the majority of sarcomas in individuals with multiple primary tumors are not therapy related. Upon expert review of pedigrees, more than 20% of families met formal criteria for known hereditary syndromes, including Li Fraumeni Syndrome (LFS), hereditary breast and ovarian cancer (HBOC), and familial melanoma, or were considered to have unusual cancer patterns (Table 1) (*10*). Collectively, these data suggest a strong genetic basis for apparently sporadic sarcomas.

Spectrum of mutated cancer genes in the sarcoma population

After undertaking WGS, we identified 37,820 rare single-nucleotide variants and insertions or deletions (SNV/indels), of which 1033 were known pathogenic or likely pathogenic (C5), 10,702 were new loss-of-function or protein truncating (C4), and 26,085 were possibly pathogenic (C3) (fig. S1). One hundred and nine sarcoma probands (6.6%) carried 112 C4 or C5 variants in clinically relevant cancer genes based on an extended list from the American College of Medical Genetics (ACMG) (table S4) (*11*), including 13 with variants in *ATM* and *TP53*, 10 *NF1*, 9 *BRCA2*, 5 *CHEK2*, 4 *MSH6*, 4 *PALB2*, 4 *LZTR1*, 3 each in *BRIP1* and *EXT2*, 2 each in *APC*, *BRCA1*, *CDKN2B*, *ERCC2* (both compound heterozygotes), *MSH2*, *RHBDF2*, *SDHA*, *SDHB*, and *WT1*, and 1 each in *AIP*, *BAP1*, *BMPR1A*, *BUB1B*, *DICER1*, *EXT1*, *FH*, *FLCN*, *HNF1A*, *MLH1*, *MUTYH* (homozygote), *PMS2*, *POLD1*, *POLE*, *PTEN*, *PTCH1*, *PTPN11*, *RAD51C*, *RAD51D*, *RB1*, *SDHC*, and *SMARCB1* (data file S1). One individual carried C4 or C5 variants in both *SDHB* and *MSH6*. Carriers of C4 or C5 variants were younger at first cancer diagnosis (median age of 41 years) than the remainder of the cohort (median age of 48 years; $P = 0.0003$) and were more likely to develop multiple primary cancers {29 of 101 probands with sufficient clinical information compared with 264 of 1478; odds ratio of 1.85 [95% confidence interval (CI) 1.18–2.91]; $P = 0.011$}. Fewer than one in six of these probands had expected clinical phenotypes that were based on genotypes (*12*). Although 10 out of 13 (77%) carriers of C4 or C5 *TP53* SNV/indels met classic or Chompret criteria for LFS, only 10 out of 164 (6.1%) families meeting either LFS criteria carried a C4 or C5 *TP53* SNV/indel. Other genes in which C4 or C5 variants were detected in families meeting LFS criteria were *MSH6* (2), *ERCC2* (2), *BUB1B*, *LZTR1*, *CHEK2*, *HNF1A*, *PTEN*, *PMS2*, *BRCA2*, *MSH2*, and *SDHB*. Only 1 of 15 (6.7%) probands with C4 or C5 SNV/indels in *BRCA1*, *BRCA2*, or *PALB2* met criteria for HBOC (*13*), whereas only 1 of 27 (4%) families with clinical features of HBOC carried a cognate C4 or C5 variant (*PALB2*). Figure S2 illustrates a sarcoma family carrying a pathogenic *BRCA1* variant, which does not meet HBOC criteria. Twelve sarcoma families met criteria for both HBOC and classic or Chompret criteria for LFS, none of which were explained by C4 or C5 variants in *BRCA1*, *BRCA2*, *PALB2*, or *TP53*. None of 9 probands with pathogenic SNV/indels in mismatch repair genes met Amsterdam criteria for colorectal cancer (*14*), and none of 15 probands meeting familial melanoma criteria (*15*) carried pathogenic

SNV/indels in the gene most strongly associated with familial melanoma, *CDKN2A* (*16*). By contrast with SNV/indels, structural variants appeared to make little contribution to pathogenic burden. Potentially germline pathogenic rare structural variants were identified in *BRCA2* (proband with two breast cancers and an undifferentiated pleomorphic sarcoma), *SDHC* [gastrointestinal stromal tumor (GIST)], and *MSH6* (proband family met Amsterdam criteria for Lynch syndrome). All genes and clinical correlates are shown in data file S1.

Gene and pathway discovery analyses

To maximize power, we applied an extreme phenotype case: control design using 3205 healthy elderly controls [from the Medical Genome Reference Bank (MGRB)], which we previously showed are depleted in cancer-associated genetic variation relative to population-based controls (*17*). A weighted rare-variant burden test (WRVBT) compared the relative burden of C3 to C5 variants in sarcoma probands (discovery set) to the MGRB, after controlling for gender, relatedness, and population structure (Fig. 1A) (*18*). Sensitivity analyses were performed to control for critical parameters including population stratification (*18*). A relaxed *P* value of < 0.1 was applied to generate an inclusive primary gene set of 1176 genes (Fig. 1A). To control for noncancer, age-related genotypes depleted in MGRB (for example, cardiovascular disease or dementia), we used the Australian Schizophrenia Research Bank (ASRB) (*17*) as a second control set with a median age of 39 years. Genes also enriched in ASRB compared to MGRB were excluded as age- rather than sarcoma-related. A secondary gene set of the top-ranked 968 genes by pathogenic burden was subjected to protein interaction– and pathway-based analysis.

These analyses revealed gene clusters that were composed of highly interrelated subclusters, which we termed cliques. Of 277 (28.6%) genes with at least one first-degree neighbor, 224 were involved in clusters of five or more genes. A supercluster of 187 genes centered on *TP53* (Fig. 1B), the strongest known sarcoma risk gene. *TP53* not only had the greatest burden of C3 to C5 variants but also had the highest number of first-degree neighbors (n = 27). The second largest cluster (Fig. 1C) contained 22 genes, representing ontologies linked to RNA processing and M-phase of the cell cycle, kinetochores, and chromatid separation. A third cluster (Fig. 1D) comprised a clique of 11 genes implicated in antigen processing, ubiquitination, and neddylation. A fourth seven-gene clique (Fig. 1E) was defined by centrosomes, spindle organization, and the G_2/M phase of the cell cycle, and a five-gene clique (Fig. 1F) reflected a role in mitochondrial translation and ribosomes. The degree of interconnectedness within the secondary gene set suggests enrichment in distinct functional pathways.

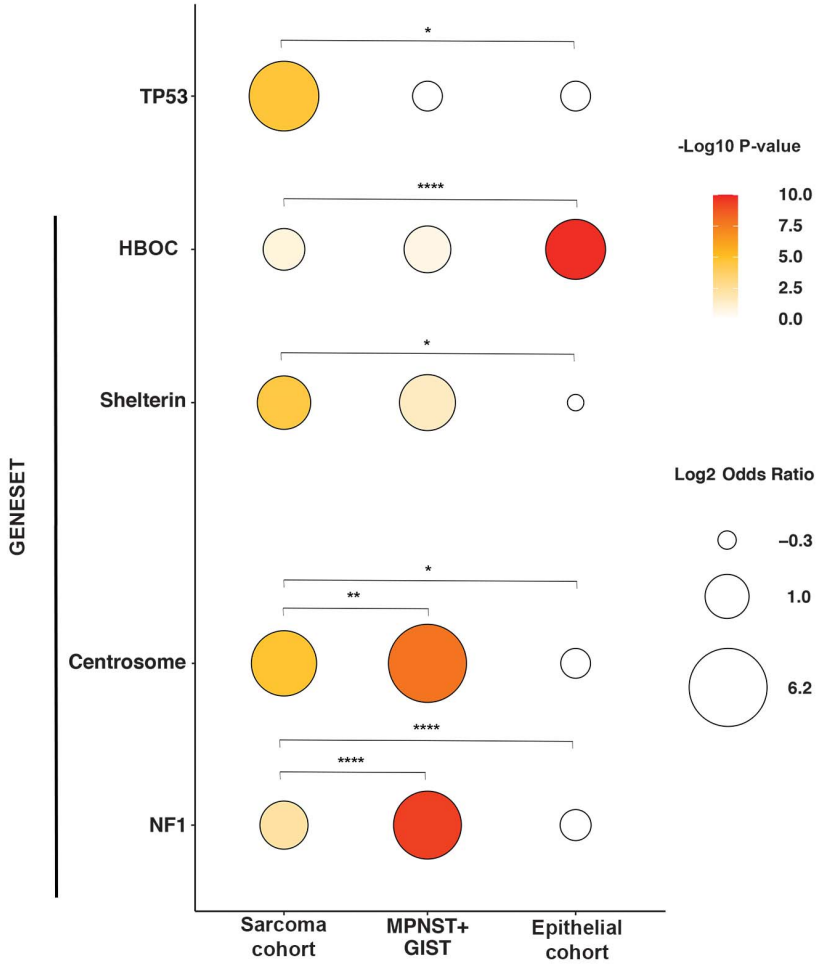

Fig. 2. Sarcoma-specific enrichment in rare pathogenic variants in the shelterin and centrosome pathways. Relative enrichment of pathogenic variants (Log2 Odds Ratio) in sarcoma probands, a subset of 157 cases with GIST or MPNST, and a nonsarcoma cancer population of 632 individuals with epithelial cancers. Gene sets include *TP53* alone (TP53); *BRCA1*, *BRCA2*, and *PALB2* (HBOC); *POT1*, *TINF2*, *TERF1*, *TERF2*, *TERF2IP*, *STAG3*, *SMARCAL1*, and *TIMELESS* (shelterin); *NF1*, *SDHA*, *SDHB*, *SDHC*, *SDHD*, and *LZTR1* (NF1); and *CEP63*, *CEP72*, *HAUS4*, *HAUS5*, *MZT1*, and *SSNA1* (centrosome). Circle size reflects the odds ratio, and the color represents the *P* value of the enrichment in each group relative to MGRB. Pairwise comparisons between groups are shown (*P < 0.05; **P < 0.01; ****P < 0.0001).

To further enrich for the most important pathways, we focused on the highest-ranked 85 genes with a *P* value < 0.05 by WRVBT and protein interactions (tertiary gene set) (Fig. 1A). Consistent with known roles in sarcoma susceptibility, the tertiary gene set included *TP53*, *NF1*, *EXT1*, and *EXT2* [UniProt Keywords KW-0043; false discovery rate (FDR) = 2.4×10^{-5}], as well as *PTCH1* (Table 2). Three ontological groups of interest are highlighted in blue (Fig. 1, B and E). Ontologic group 1 comprised the shelterin complex, represented by three genes: *POT1*, *TERF1*, and *TINF2* (GO Component GO.0070187; FDR = 2.8×10^{-5}) (Table 2). Group 2 comprised the mitotic spindle, including the centrosomal genes *CEP63*, *HAUS4*, and *HAUS5* (GO Component term GO.0005819; FDR = 4.5×10^{-5}) (Table 2). Group 3 includes *EXT1* and *EXT2*, linked to hereditary exostoses and bone sarcomas (UniProt Keywords KW-0361; FDR = 2.3×10^{-4}) (Table 2).

We next determined whether these pathways were specific for sarcomas by comparing sarcoma probands to a population with predominantly epithelial cancers (n = 632) (table S5). As a positive control, an enrichment in pathogenic burden was seen for *TP53* in sarcoma probands compared with the epithelial cases. As a negative control, the reverse was observed with a HBOC gene set (*BRCA1*, *BRCA2*, and *PALB2*). Sarcoma-specific enrichment was observed in both the shelterin

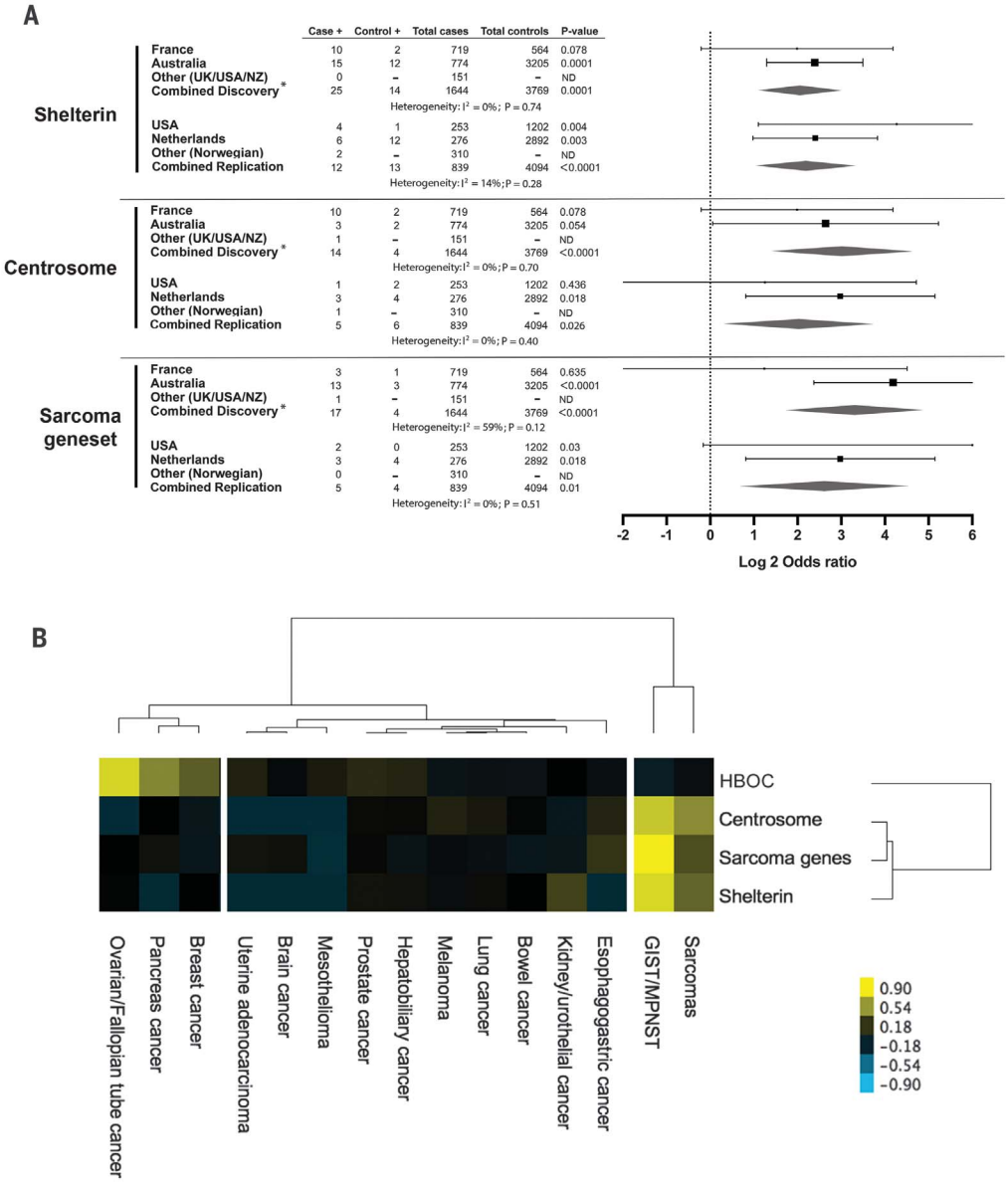

Fig. 3. Replication of enrichment in centrosome and shelterin-complex genes in sarcomas. (**A**) Enrichment of rare C4 or C5 pathogenic variants in sarcoma, shelterin, and centrosome gene sets were estimated for geographically matched discovery (1644 cases and 3769 controls) and replication (839 cases and 4094 controls) datasets. Odds ratios for each geographically matched set are shown with closed squares, and the corresponding horizontal line and whiskers represent 95% confidence intervals (CI). The pooled effect and 95% CI of combined datasets are shown with the diamond. Set heterogeneity was tested using the Cochran-Mantel-Haenszel test without continuity correction and heterogeneity estimates (I^2) and *P* values are presented. Sarcoma cases from the UK, USA, and NZ in the original discovery set and sarcoma cases from Norway in the replication set were excluded from individual analyses due to the lack of geographically matched controls. The combined odds ratio and 95% CI of geographically matched Discovery* and Replication datasets for shelterin, centrosome, and sarcoma gene sets are 5.58 [3.11, 9.64], 4.74 [1.90, 11.83], and 10.59 [4.36, 25.72], respectively. The corresponding I^2 values are 0% for all the gene sets. (**B**) This analysis used the Hartwig Foundation whole-genome germline dataset of 4178 cancers, including sarcomas (276, including 71 GIST or MPNST), breast cancer (801), bowel cancer (650), lung cancer (583), prostate cancer (412), kidney and urothelial cancers (312), melanoma (300), esophagastric cancer (193), ovarian and fallopian tube cancers (174), pancreatic cancer (149), hepatobiliary cancer (138), brain cancer (76), uterine adenocarcinoma (73), and mesothelioma (41). Gene sets were defined as for Fig. 2. Enrichment for each cancer class was determined by normalizing the frequency of variants in that class to the overall frequency in the entire dataset. These data were subject to unsupervised hierarchical clustering for both cancer class and gene sets.

and the centrosome gene sets (Fig. 2). For the centrosome set, further enrichment was specifically observed in malignant peripheral nerve sheath tumors (MPNSTs) and GISTs, which share neural origins and *NF1* as a susceptibility gene (*5*, *19*). The degree of enrichment in centrosome genes in MPNSTs or GISTs was comparable to that seen for *NF1* and related genes such as *LZTR1*, *SDHA*, *SDHB*, *SDHC*, and *SDHD*. A follow-up validation analysis of the discovery set with an iterative (1000-fold) resampling method also recapitulated the enrichment of rare variant burden in the shelterin and centrosome gene sets (fig. S3).

Independent replication of pathway enrichment

We next sought to validate these findings and to exclude population stratification. First, we compiled an independent WGS and WES replication set (n = 839 sarcoma cases from the US, Netherlands, and Norway) and geographically matched independent control set (n = 4094 cancer-free controls) (Fig. 3A and table S2). A WRVBT was performed using C4 and C5 variants in the shelterin, centrosome, and sarcoma gene sets, for each geographically matched case-control discovery and replication set, and after combining these for statistical power (Fig. 3A). Enrichment was confirmed in pathogenic variants in the sarcoma, shelterin, and centrosome gene sets within the combined geographically matched French and Australian discovery sets, as well as in the combined geographically matched replication sets from the Netherlands, US, and Norway, with no significant heterogeneity across these populations (Fig. 3A). Residual population stratification was excluded as a cause for signal enrichment using principal components ancestral matching within the geographically matched populations (figs. S4 and S5) (*18*). We next compared the relative burden of pathogenic variants in the sarcoma, shelterin, centrosome, and HBOC gene sets across a wide range of cancer types using the Hartwig Foundation dataset (table S6)

(*20*). We observed enrichment of HBOC genes in breast, pancreatic, ovarian, and fallopian tube cancers, but not in sarcomas (Fig. 3B). No enrichment was observed in nonsarcoma cancers for the sarcoma, shelterin, and centrosome gene sets. By contrast, a distinct relative enrichment in the shelterin and centrosome gene sets was again observed in the sarcoma cases, and specifically in GISTs or MPNSTs.

Clinical, familial, and molecular genotype–phenotype patterns

Of 19 individuals carrying C4, C5, or essential splice-site variants in the centrosome core genes *CEP63*, *CEP72*, *HAUS4*, and *HAUS5*, as well as in the related genes *CEP89*, *SSNA1*, and *PCM1* (table S7), 8 (42%) developed GIST (7) or MPNST (1). Notably, one individual carried variants in both *HAUS4* and *HAUS5*. An additional 20 individuals carried one C3 variant each in *CEP57*, *CEP63*, *CEP89*, *HAUS4*, *HAUS5*, *PCM1*, *SSNA1*, and *MZT1*. Consistent with the observations in individuals with C4 or C5 variants, 6 of these 20 individuals (30%) with C3 variants developed either GIST (3) or MPNST (3). Altogether, 14 of 39 (36%) carriers of C3 to C5 variants in centrosome genes developed MPNST or GIST, compared with 143 of 1462 (9.8%) in the remaining cohort (RR 4.029 [95% CI 2.372-6.142]; $P < 0.0001$). Of five carriers with GIST, four had somatic c-KIT variants, and one was wild-type. The median age at first cancer diagnosis in C4 or C5 variant carriers (43 years) was comparable to that of the cohort as a whole. Two of the centrosome pedigrees met clinical criteria for LFS, and another two were deemed clinically suspicious. Among individuals with MPNST or GIST, centrosome variants were mutually exclusive with C4 or C5 variants in genes known to cause these tumors: *NF1* (9), *LZTR1*, *SDHA*, and *SDHB*. An analysis of 4179 cancers, including 277 sarcomas, found that 2 of 24 carriers of germline pathogenic variants in centrosome genes showed somatic loss-of-heterozygosity, comparable to 5 of 26 cancers in individuals with pathogenic germline variants in *NF1*, *SDHA*, *SDHB*, and *SDHD*. These data suggest that defects in centrosome pathways appear to confer increased risk of MPNST or GIST.

There were 13 carriers of C4 or C5 variants in four of the six canonical genes in the shelterin complex (Table 3), *POT1* (6), *TERF1* (2), *TINF2* (3), *TERF2IP* (2), and an additional 12 carriers of C4 or C5 variants in genes selected on the basis of their roles in telomere biology, *TIMELESS* (3), *STAG3* (3), and *SMARCAL1* (6) (*21*, *22*). *POT1*, the only established hereditary cancer gene in this group, has been most strongly linked to melanoma, but also has been reported in thyroid cancer and sarcoma (*23*). A loss-of-heterozygosity analysis performed as described above identified 9 somatic loss-of-function events in 55 tumors from carriers of pathogenic germline variants in the shelterin genes, including 2 of 8 in *POT1* and 5 of 19 in *SMARCAL1*.

Amongst carriers of shelterin C4 or C5 variants, formal review of pedigree patterns identified four families meeting GenoMEL criteria for familial melanoma, and four families that met Chompret or classic LFS criteria. Two additional families meeting GenoMEL criteria carried C3 variants in *POT1* (G176R and D224N; fig. S6) (*24*). Notably, three individuals also had thyroid cancer. Several lines of evidence suggest the association between

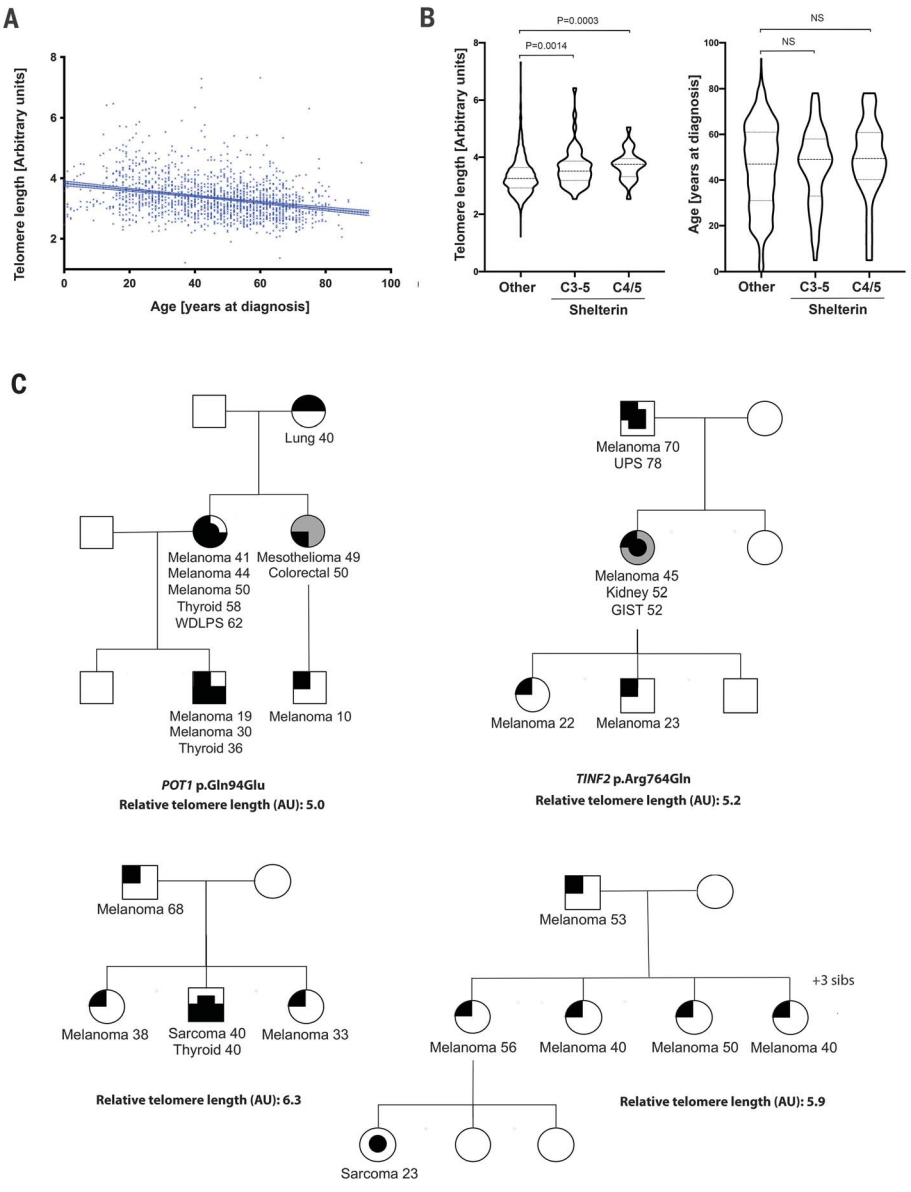

Fig. 4. Leukocyte telomere length in carriers of shelterin-complex variants. (**A**) Relative telomere lengths (RTL) were derived from telomere analysis (TelSeq) of whole-genome sequences on peripheral blood DNA for the sarcoma probands. The *x* axis indicates the age for each individual in years, and the *y* axis indicates RTL in arbitrary units. (**B**) Leukocyte RTL is longer in carriers of variants in the shelterin complex. Left panel: RTL in probands with shelterin complex C3 to C5 ($n = 35$) or C4 or C5 variants ($n = 25$) compared to the remainder of the sarcoma probands. Right panel: Age distribution for these groups. (**C**) Representative pedigrees showing autosomal dominant cancer patterns with an excess of melanoma. Top left panel: family with a *POT1* pathogenic variant; top right panel: family with a *TINF2* pathogenic variant; bottom panels: representative pedigrees of probands with long telomeres and excess of melanoma with no pathogenic variants identified.

Table 1. Clinical and pedigree characteristics of sarcoma probands. HNPCC, Hereditary nonpolyposis colorectal cancer; HBOC, Hereditary breast or ovarian cancer.

Characteristic		N (%)
Gender	Female	855 (52%)
	Male	789 (48%)
	Total	1644
Age at diagnosis, years (mean, SD)	First cancer	45.4, 18.3
	Sarcoma	46.7, 18.7
Number with multiple primary cancers		211
	three primary cancers	63
	≥ four primary cancers	19
	Total	293 (18%)
Pedigree classification	No syndrome	1159
	Classic or Chompret LFS	164
	HNPCC	2
	Familial melanoma	15
	HBOC	27
	Clinically suspicious*	110
	Other**	14
	Uninformative	153
Pathology subtype	Bone	358
	Soft tissue	1286
Genomic class	Complex	836
	Simple	372
	Translocation associated	362
	Unknown	74

*Clinically suspicious: >0.5 cancers or FDR with ≥3 FDR reported and ≥1 FDR <50 years at diagnosis; three or more cancers per proband at any age; 1 or more sarcomas in an FDR; >1 sarcoma or connective tissue tumor at any age per proband; two cancers under 50 years of age per proband. **Neurofibromatosis type 1 (5), retinoblastoma (2), multiple endocrine neoplasia type 1 (2), familial papillary thyroid (1), familial paraganglioma (1), schwannomatosis (1), Gorlin syndrome (1), and McCune Albright syndrome (1).

sarcoma, melanoma, and thyroid cancers is not therapy related. Both melanoma and thyroid cancer are curatively treated by surgery without genotoxic therapy; moreover, relatives of shelterin probands carrying C4 or C5 variants also had an increased incidence of cancer {standardized incidence ratio (SIR) 2.06 [1.50–2.82]; $P = 1.84 \times 10^{-4}$}, comparable to that of *TP53* {SIR 2.59 [1.53–4.37]; $P = 3.69 \times 10^{-4}$}. The risk of melanoma was markedly increased in families carrying C4 or C5 variants in the shelterin complex {SIR 5.60 [3.25–9.65]; $P = 5.61 \times 10^{-10}$}, as was the risk of thyroid cancer {SIR 19.74 [8.22–47.43]; $P = 2.56 \times 10^{-11}$}. An additional individual carried a C4 or C5 variant in *CTC1* and another individual carried a C4 or C5 variant in *NHP2*—, genes that are both implicated in telomere maintenance through the CST and telomerase complexes (*25*, *26*). Taken together, these data suggest that heritable defects in telomere function predispose to a syndrome of sarcoma, melanoma, and thyroid cancers.

We next analyzed relative leukocyte telomere lengths (RLTL) derived from WGS. As expected, RLTL progressively shortened over life span in sarcoma probands (Fig. 4A), unrelated to chemotherapy or radiotherapy exposure (*18*). Despite a similar age at first cancer diagnosis, shelterin variant carriers had longer RLTL than the remainder of the cohort (Fig. 4B). No carriers had clinical or genomic evidence of clonal hemopoiesis, dysplasia, or leukemia (*17*). Independent of genotype, relatives of probands with long telomeres (>2 SD above the mean for age) had an increased risk of melanoma {SIR 3.58 [2.19–5.84]; $P = 3.36 \times 10^{-7}$} and thyroid cancer {SIR 8.43 [3.16–22.45]; $P = 5.96 \times 10^{-9}$}. Representative melanoma-associated pedigree patterns associated with *POT1*, *TINF2*, and *SMARCAL1* variants are shown in Fig. 4C and fig. S6. For two families of probands with long RLTL and cognate tumor patterns in which no causal coding variants were found in shelterin complex genes (Fig. 4C, lower panels), we did not find evidence for causal noncoding variants.

Discussion

Together, these data suggest that heritable defects in telomere and mitotic function increase the risk of sarcoma, in contrast to most epithelial cancers. Telomere maintenance and mitosis are fundamental to chromosome integrity. Sarcomas are predominantly genomically unstable (*27*) and use the alternate mechanism of telomere maintenance (ALT) (*28*). Through replication stress, ALT generates genomic instability that signals to *TP53*, which is strongly linked to sarcoma susceptibility (*29*–*33*). Fifty-two (3.2%) sarcoma probands carried C3 to C5 variants in the shelterin complex and related telomere genes, compared to 13 for *TP53* (0.8%). Pathogenic variation in these genes is rare in the healthy individuals, with the MGRB containing only one C3 to C5 variant in *TP53* (0.03%), and six in shelterin complex genes (0.2%).

Shelterin is a six-subunit protein complex which protects the ends of telomeres (*34*) and is associated with melanoma risk (*35*, *36*). Shelterin variants occurred in 6 out of 15 (40%) sarcoma families with familial melanoma, in contrast to the gene most commonly mutated in familial melanoma (*CDKN2A*), in which no variants were found (*16*). Like *TP53*, germline shelterin pathogenic variants exhibit a dominant pattern of inheritance and cancer risk. Unlike *TP53*, cancer onset occurs at an older age, perhaps the result of an interaction with age-dependent telomere shortening. In addition to shelterin, three genes with roles in telomere maintenance were identified with multiple C4 or C5 variants: *SMARCAL1*, *TIMELESS*, and *STAG3*. *SMARCAL1* and *TIMELESS* regulate telomere stability in ALT cells (*21*, *22*, *37*–*40*), and *STAG3* plays roles in both centromere and telomere–sister chromatid cohesion (*41*, *42*). Shelterin complex pathogenic variant carriers had long leukocyte telomeres, recently associated with increased risk for multiple cancers (*32*). Sarcomas, melanomas, and thyroid cancers may share an association with long-telomere syndrome (*36*). Notably, we did not observe any variants in components of the telomerase complex, the dominant mechanism of telomere maintenance in epithelial cancers.

Overall, 2.3% of sarcoma probands (comprising 10% of MPNSTs and 8.4% of GISTs) carry C3 to C5 variants in centrosome genes. Centrosome variants were mutually exclusive with variants in known MPNST or GIST genes (*NF1*, *LZTR1*, *SDHA*, and *SDHB*), which themselves account for 15% of MPNSTs and 5% of GISTs in our study. Chromosome segregation during mitosis begins with centrosome assembly (*43*). Although somatic centrosome abnormalities are common in cancer (*44*), a germline role for this pathway is limited to *CEP57*. *CEP57* has been associated with mosaic variegated aneuploidy (*45*), a recessive condition associated with chromosomal instability and increased risk of cancer, including sarcomas (*46*). *CEP63* interacts with *CEP72* and *CEP152* to mediate formation of the centrioles, centrosomes, and bipolar spindle

Table 2. Biological pathways enriched in pathogenic variants in the top-ranked 85 sarcoma-specific genes by WRVBT. The top 85 genes by WRVBT and protein interactions with at least one other gene were analyzed in Cytoscape using StringApp. Ontologic groups refer to relevant groups in the secondary gene set (Fig. 1, B and E). Term names are derived from relevant databases (Category). Genes refers to the those found related to the term name, and % term refers to the percentage of total genes in the term represented. The uncorrected *P* values and FDR values are provided.

	Group 1	Group 2	Group 3	Other
Category	GO Component	GO Component	UniProt Keywords	UniProt Keywords
Term name	GO.0070187	GO.0005819	KW-0361	KW-0043
Description	Shelterin complex	Spindle	Hereditary multiple exostoses	Tumor suppressor
Genes	*TINF2 TERF1 POT1*	*CEP63 HAUS4 HAUS5 MAD2L2 TERF1 MYH10 FAM161A HEPACAM RANGAP1 RACGAP1*	*EXT1 EXT2*	*TP53 PTCH1 STARD13 EPB4IL3 NF1 RBL1 EXT1 EXT2*
% Term	42.9	3.1	100	4.4
***P* value**	0.0000277	0.0000452	0.00023	0.000024
FDR	0.0016	0.0021	0.0118	0.0029

Table 3. Clinical, molecular, and familial features of probands carrying C4 or C5 variants in the shelterin complex and associated genes. AS, angiosarcoma; CS, chondrosarcoma; FMS, fibromyxosarcoma; GIST, gastrointestinal stromal tumor; LPSWD, Well- or dedifferentiated liposarcoma; LPSM, myxoid liposarcoma; MPNST, malignant peripheral nerve-sheath tumor; OS, osteosarcoma; PNET, primitive neuroectodermal tumor; RMS, rhabdomyosarcoma, not otherwise specified; SS, synovial sarcoma; UPS, undifferentiated pleomorphic sarcoma; LFS, Li Fraumeni Syndrome; NS, no syndrome; U, uninformative.

ID	Gene	Protein change	Gender	Sarcoma	Age at diagnosis	Second malignancies (ages at diagnoses)	Pedigree pattern
017.468.1	*POT1*	p.Gln94Glu	F	LPSWD	62	Thyroid (58); melanoma (41, 44, 50)	Familial melanoma
041.1570.1	*POT1*	p.Tyr501Ter	M	UPS	60	NA	NS
041.1598.1	*POT1*	p.Gln358SerfsTer13	F	AS	50	NA	NS
041.2877.1	*POT1*	p.Tyr89Ter	F	FMS	46	NA	NS
041.2496.1	*POT1*	p.Tyr89Ter	M	MPNST	50	NA	Familial melanoma
017.761.1	*POT1*	p.Ala532HisfsTer13	M	UPS	74	NA	NS
012.3145.1	*SMARCAL1*	p.Arg645ProfsTer46	M	OS	33	NA	NS
041.1055.1	*SMARCAL1*	p.Trp843Ter	M	PNET	22	NA	Chompret
041.3022.1	*SMARCAL1*	p.Lys534Ter	F	UPS	61	NA	NS
012.492.1	*SMARCAL1*	p.Glu848Ter	F	OS	14	NA	NS
045.2520.1	*SMARCAL1*	p.Arg764Gln	M	UPS	52	Thyroid (49)	NS
012.2401.1	*SMARCAL1*	p.Arg764Gln	F	GIST	52	Kidney (52); melanoma (45)	Familial melanoma
013.347.1	*STAG3*	p.Arg1018AspfsTer14	F	OS	52	NHL (50)	NS
L-P001402	*STAG3*	p.Cys141Ter	M	RMS	5	NA	U
013.3129.1	*STAG3*	p.Cys141Ter	F	UPS	49	NA	NS
013.329.1	*TERF1*	p.Arg147Ter	F	UPS	61	Melanoma (64)	Familial melanoma
012.45.1	*TERF1*	p.Arg147Ter	M	CS	74	NA	NS
043.1874.1	*TERF2IP*	p.Arg153GlnfsTer6	F	GIST	43	NA	NS
016.558.1	*TERF2IP*	p.Arg364Ter	F	LPSWD	64	Thyroid (42)	NS
013.297.1	*TIMELESS*	p.Leu867ThrfsTer8	F	CS	78	NA	NS
012.235.1	*TIMELESS*	p.Ala383GlyfsTer19	M	SS	40	NA	Chompret
013.746.1	*TIMELESS*	p.Tyr25Ter	M	CS	58	NA	NS
041.1280.1	*TINF2*	p.Val67TrpfsTer3	F	LPS	66	NA	NS
043.2441.1	*TINF2*	p.Arg256Ter	F	LPSM	37	NA	Chompret
015.1183.1	*TINF2*	p.Arg265Ter	F	UPS	58	NA	Classic LFS

(*47*). *CEP63* and *CEP89* also interact with *HAUS4* and *HAUS5*, augmin-like complex members that regulate centrosome integrity (*48*). Notably, this centrosome group adds to the known genes linked to MPNST and GIST.

The methodologies applied here have broad applications for disease-focused pathway discovery in the emerging era of WGS, allowing mapping of rare SNV/indels and structural variation (SV) in large populations. In our study, SNV/indels (rather than SVs) appear the dominant source of excess pathogenic burden. In contrast to population-based controls, controls of healthy elderly individuals increase statistical power by stripping out pathogenic variation–causing phenotypes (such as cancer) that emerge later in life. Notably, the combination of population-scale and familial cohort designs, historically distinct from each other, allows correlation of genotypes to clinical and familial patterns. The risk-to-relatives analyses

demonstrated an increased cancer incidence in family members of probands with shelterin-complex variants comparable with that seen for carriers of variants in *TP53*. Statistical approaches that combine gene-level pathogenic enrichment with pathway-based approaches based on protein interactions (*49*) and ontologies enhance the power for biological insights into cancer pathogenesis (*50*, *51*). These approaches have identified genes and biologic pathways that appear specific to mesenchymal malignancies, as well as mapping the contributions of known cancer genes. Amongst known and actionable genes, the lack of enrichment in *BRCA1 or BRCA2* is notable, although some sarcoma families meet both HBOC and LFS criteria. We add 14 candidates (including *TERF1*, *TINF2*, *TERF2IP*, *SMARCAL1*, *TIMELESS*, *STAG3*, *CEP63*, *CEP72*, *HAUS4*, and *HAUS5*) to more than 100 known sarcoma-associated genes (*7*). Biologically, these data suggest that the telomeric and mitotic pathways may play specific roles in sarcoma susceptibility, analogous to homologous recombination and mismatch repair in susceptibility to breast and colorectal cancers.

REFERENCES AND NOTES

1. O. G. Bahcall, *Nat. Genet.* **45**, 343 (2013).
2. P. Lichtenstein *et al*., *N. Engl. J. Med.* **343**, 78–85 (2000).
3. A. Paul, S. Paul, *Front. Biosci. (Landmark Ed.)* **19**, 605–618 (2014).
4. C. Wang, Y. Wang, K. S. Hughes, G. Parmigiani, D. Braun, *JNCI Cancer Spectr.* **4**, pkaa027 (2020).
5. World Health Organisation, *WHO Classification of Tumours: Soft Tissue and Bone Tumours*. V. W. Dilani Lokuhetty, Ian Cree, Ed., WHO Classification of Tumours (International Agency for Research on Cancer, Lyon, France., ed. 5, 2020).
6. M. L. Ballinger, M. Pinese, D. M. Thomas, *Genes Chromosomes Cancer* **58**, 130–136 (2019).
7. L. Mirabello *et al*., *JAMA Oncol.* **6**, 724–734 (2020).
8. J. Zhang, K. E. Nichols, J. R. Downing, *N. Engl. J. Med.* **374**, 1391 (2016).
9. C. M. Mery, S. George, M. M. Bertagnolli, C. P. Raut, *Cancer* **115**, 4055–4063 (2009).
10. M. L. Ballinger *et al*., *Lancet Oncol.* **17**, 1261–1271 (2016).
11. S. S. Kalia *et al*., *Genet. Med.* **19**, 249–255 (2017).
12. D. C. Chung, D. A. Haber, Massachusetts General Hospital, *Principles of clinical cancer genetics: a handbook from the Massachusetts General Hospital*. (Springer, New York, 2010), pp. xi, 227 p.
13. D. G. Evans *et al*., *J. Med. Genet.* **46**, 811–817 (2009).
14. H. F. Vasen, P. Watson, J. P. Mecklin, H. T. Lynch, *Gastroenterology* **116**, 1453–1456 (1999).
15. GenoMEL, the Melanoma Genetics Consortium, What is the chance that you have inherited a mutation in CDKN2A? Available from https://genomel.moffitt.org/GenoMELPREDICT. Accessed October 2020.
16. N. J. Taylor *et al*., *J. Am. Acad. Dermatol.* **81**, 386–394 (2019).
17. M. Pinese *et al*., *Nat. Commun.* **11**, 435 (2020).
18. See supplementary materials.
19. J. Andersson *et al*., *Am. J. Surg. Pathol.* **29**, 1170–1176 (2005).
20. P. Priestley *et al*., *Nature* **575**, 210–216 (2019).
21. K. E. Cox, A. Maréchal, R. L. Flynn, *Cell Rep.* **14**, 1032–1040 (2016).
22. M. C. Gadaleta, A. González-Medina, E. Noguchi, *Curr. Genet.* **62**, 725–730 (2016).
23. Y. Wu, R. C. Poulos, R. R. Reddel, *Cancers* **12**, 2739 (2020).
24. C. D. Robles-Espinoza, M del. Castillo Velasco-Herrera, N. K. Hayward, D. J. Adams, *Mol. Cancer Res.* **13**, 211–*222* (2015).
25. C. Huang, P. Jia, M. Chastain, O. Shiva, W. Chai, *Exp. Cell Res.* **355**, 95–104 (2017).
26. T. Vulliamy *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **105**, 8073–8078 (2008).
27. F. Chibon *et al*., *Nat. Med.* **16**, 781–787 (2010).
28. R. L. Dilley, R. A. Greenberg, *Trends Cancer* **1**, 145–156 (2015).
29. E. A. Akbay *et al*., *Oncogene* **32**, 2211–2219 (2013).
30. Y. J. Chen *et al*., *Cancer Res.* **66**, 6473–6476 (2006).
31. M. Lee *et al*., *Nucleic Acids Res.* **46**, 4903–4918 (2018).
32. E. J. McNally, P. J. Luncsford, M. Armanios, *J. Clin. Invest.* **129**, 3474–3481 (2019).
33. A. M. Pinzaru *et al*., *Genes Dev.* **34**, 1619–1636 (2020).
34. M. J. Giraud-Panis, M. T. Teixeira, V. Géli, E. Gilson, *Mol. Cell* **39**, 665–676 (2010).
35. J. Shi *et al*., *Nat. Genet.* **46**, 482–486 (2014).
36. L. G. Aoude *et al*., *J. Natl. Cancer Inst.* **107,** (2014).
37. A. R. Leman *et al*., *Cell Cycle* **11**, 2337–2347 (2012).
38. L. A. Poole *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **112**, 14864–14869 (2015).
39. J. Min, W. E. Wright, J. W. Shay, *Mol. Cell. Biol.* **37**, e00226-17 (2017).
40. E. Feng *et al*., *J. Cell Sci.* **133**, jcs.234914 (2020).
41. B. Liebe, M. Alsheimer, C. Höög, R. Benavente, H. Scherthan, *Mol. Biol. Cell* **15**, 827–837 (2004).
42. T. Winters, F. McNicoll, R. Jessberger, *EMBO J.* **33**, 1256–1270 (2014).
43. P. T. Conduit, A. Wainman, J. W. Raff, *Nat. Rev. Mol. Cell Biol.* **16**, 611–624 (2015).
44. S. A. Godinho, D. Pellman, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, 20130467 (2014).
45. K. Snape *et al*., *Nat. Genet.* **43**, 527–529 (2011).
46. S. Hanks *et al*., *Nat. Genet.* **36**, 1159–1161 (2004).
47. Z. Wei *et al*., *Mol. Cell. Biol.* **40**, e00535-19 (2020).
48. S. Lawo *et al*., *Curr. Biol.* **19**, 816–826 (2009).
49. R. Oughtred *et al*., *Nucleic Acids Res.* **47** (D1), D529–D541 (2019).
50. D. Szklarczyk *et al*., *Nucleic Acids Res.* **47** (D1), D607–D613 (2019).
51. The Gene Ontology Consortium, *Nucleic Acids Res.* **47**, D330–D338 (2019).

ACKNOWLEDGMENTS

We acknowledge and thank the ISKS families and ISKS investigators. We thank J. Hopper, J. Desai, P. Stalley, T. Corpuz, I. Han, C. Napier, R. Reddel, and the Kinghorn Centre for Clinical Genomics for helpful discussions, advice, and technical assistance. We gratefully acknowledge access to the Medical Genome Reference Bank, the Australian Schizophrenia Research Bank, the Zero Childhood Cancer Program, the Hartwig Foundation dataset, the cardiovascular disease and spontaneous coronary artery dissection cohorts, the French Exome Project (FREX) consortium, and the Amyotrophic Lateral Sclerosis dataset. We thank O. Holmes and S. Wood from Genome Informatics, QIMR Berghofer for their assistance in public data processing. The results shown here are in part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga. The NoSarC samples were exome sequenced by the Genomics Facilities at Oslo University Hospital and University of Bergen. **Funding:** Australian National Health and Medical Research Council APP1004017, APP1081477 (D.M.T., M.L.B.), Bioplatforms Australia (D.M.T.), NSW Office of Health and Medical Research Cancer Moonshot Program (D.M.T.), Cancer Australia Priority–driven Collaborative Cancer Research Scheme APP1067094 (D.M.T., M.L.B.), Australian Genomics Health Alliance (D.M.T.), Rainbows for Kate Foundation (D.M.T., M.L.B.), Liddy Shriver Sarcoma Initiative (D.M.T., M.L.B., J.D.S., I.J., A.P.), Johanna Sewell Research Foundation (D.M.T., M.L.B.), Victorian Cancer Agency BFRes11_03 (M.L.B., D.M.T.), CanTeen (D.M.T., M.L.B.), NSW Office of Health and Medical Research (D.M.T.), Sydney Genomics Collaborative Grants Scheme (D.M.T.), Lions Club Kids Cancer Genome Project (D.M.T.), Sarcoma UK (I.J., B.S.), National Health Service funding to the National Institute for Health Research Biomedical Research Centre, UK (I.J.), Norwegian Cancer Society 144385 (O.M.), Norwegian Radium Hospital Legacy, 170502 (O.M.), Trond Mohn Foundation (O.M.), Norwegian Research Council 218241 and 221580 (O.M.), Bergen Research Foundation BFS2017TMT04 (O.M.), Australian National Health and Medical Research Council Principal Research Fellowship APP1104364 (D.M.T.), Cancer Institute NSW Career Development Fellowship CDF171109 (M.L.B.), Cure Cancer Australia Foundation through Cancer Australia Priority-driven Collaborative Cancer Research Scheme (APP1159399) (T.L.N.), Victorian Cancer Council Post-Doctoral Fellowship (CCV1139144) (T.L.N.), Picchi Foundation (T.L.N.), Victorian Comprehensive Cancer Centre (T.L.N.), Victorian Cancer Agency Early Career Research Fellowship ECRF19020 (S.L.), National Health and Medical Research Council of Australia Senior Research Fellowship APP1139071 (N.W.), NetSARC+ INCA and DGOS (J.-Y.B., I.R.-C., N.P., N.I., F.D., J.-E.K.), EURACAN EC 739521 (J.-Y.B. and I.R.-C.) LYRICAN INCA-DGOS-INSERM 12563 (J.-Y.B. and I.R.-C.), LabEx DEvweCAN ANR-10-LABX-0061 (J.-Y.B. and I.R.-C.), Institut Convergence PLASCAN 17-CONV-0002 (J.-Y.B. and I.R.-C.), RHU4 DEPGYN (ANR-18-RHUS-0009 (J.-Y.B. and I.R.-C.), TMRG (INCA and DGOS) (J.-Y.B. and I.R.-C.), InterSARC (INCA and DGOS) (J.-Y.B. and I.R.-C.), Foundation ARC Ligue Contre Le Cancer (J.-Y.B. and I.R.-C.), Cancer Institute NSW Cancer Genetics Training Grant (M.Z.), Hyundai Hope on Wheels (J.D.S.), Soccer for Hope Foundation (J.D.S.), Li-Fraumeni Syndrome Association (J.D.S.), Kneaders Bakery & Café Hope Campaign (J.D.S.), 5 For The Fight (Qualtrics) (J.D.S.), Elephant p53 (EP53) Program funded through Huntsman Cancer Institute by the State of Utah (J.D.S.), Australian National Health and Medical Research Council Senior Research Fellowship APP1121474 (M.J.C.), and University of Newcastle Faculty of Health and Medicine Gladys M Brawn Senior Fellowship (M.J.C.). The FREX consortium was supported by France Génomique National infrastructure, funded as part of "Investissement d'avenir" program managed by Agence Nationale pour la Recherche (contrat ANR-10-INBS-09), https://www.france-genomique.org/. **Author contributions:** Conceptualization: D.M.T. Methodology: D.M.T., M.L.B., S.P., and P.A.M. Software: S.P. and P.A.M. Validation: D.M.T., M.L.B., S.P., P.A.M., P.P., J.B., D.F., R.M.G., E.G., E.C., K.v.E., J.V., T.L., E.G., and J.-F.D. Formal analysis: D.M.T., M.L.B., S.P., P.A.M., E.R., M.Z., K.T., H.P., M.K., S.L., T.L.N., W.K., S.R., J.C., and J.E.P. Resources: I.R.C., N.I., S.C., W.T.A.v.d.G., A.P., F.D., A.L.C., B.S., C.C., J.D.S., A.S.B., P.A.J., J.E.-K., N.P., O.M., L.A.M.-Z., N.W., O.K., J.V.P., A.P.B., M.J.G., J.-H.A., J.E.-K., R.L.R., I.J., R.S., M.H., G.M., M.J.C., J.-Y.B., and J.V. Data Curation: M.L.B., S.P., and P.A.M. Writing – Original Draft: D.M.T., M.L.B., S.P., and P.A.M. Writing – Review and Editing: All authors. Visualization: D.M.T., M.L.B., S.P., and P.A.M. Project administration: D.M.T. and M.L.B. Funding acquisition: D.M.T. **Competing interests:** J.V.P. and N.W. are founders and shareholders of genomiQa Pty Ltd, and members of its board. J.D.S. is employee, stock option holder, and board member of PEEL Therapeutics, Inc. J.V. has sponsored research agreements with Biogen and Astra Zeneca. O.K. has consulted for XING Technologies on development of diagnostic assays for HR deficiency. Other authors declare they have no competing interests. **Data and materials availability:** ISKS, RisC, and ZCCP summary variant frequency data and phenotypic data are available upon application to the respective data access committees at isks@garvan.org.au and agcmp@garvan.org.au and subject to MTA. Summary variant frequency data for the MGRB cohort are available at the web portal: https://sgc.garvan.org.au/explore. Raw genomic data have been deposited at the European Genome-Phenome Archive under study ID EGAS00001003511. Phenotype data are available upon application to the MGRB Data Access Committee at mgrb@garvan.org.au. ASRB data are available upon application at https://www.neura.edu.au/discovery-portal/asrb/. TCGA data are available from the Genomic Data Commons at https://gdc.cancer.gov. NoSarc dataset is available from https://sarcoma.no/en/english-data-access/ upon application and subject to MTA. Hartwig Foundation dataset is available upon application at https://www.hartwigmedicalfoundation.nl/en/. The cardiovascular disease (CVD) and spontaneous coronary artery dissection (SCAD) cohort data will be made available upon reasonable request and in accordance with other publication embargoes. The Wellderly cohort data are available upon application to atorkama@scripps.edu. The FREX consortium data are available upon application at french.exome@gmail.com. The Amyotrophic Lateral Sclerosis dataset is the Dutch cohort of Project MinE. Data access can be requested here: https://www.projectmine.com/research/download-data/.

SUPPLEMENTARY MATERIALS

science.org/doi/10.1126/science.abj4784
Collaborators
Materials and Methods
Text
Tables S1 to S13
Figs. S1 to S11
References (*52*–*113*)
MDAR Reproducibility Checklist
Data Files S1 to S3

View/request a protocol for this paper from *Bio-protocol*.

Submitted 8 June 2021; resubmitted 25 November 2021
Accepted 16 November 2022
10.1126/science.abj4784