The importance of p53 pathway genetics in inherited and somatic cancer genomes

Giovanni Stracquadanio[1], Xuting Wang[2]*, Marsha D. Wallace[1], Anna M. Grawenda[1], Ping Zhang[1], Juliet Hewitt[1], Jorge Zeron-Medina[3], Francesc Castro-Giner[4], Ian P. Tomlinson[4], Colin R. Goding[1], Kamil J. Cygan[5,6], William G. Fairbrother[5,6], Laurent F. Thomas[7], Pål Sætrom[7,8], Federica Gemignani[9], Stefano Landi[9], Benjamin Schuster-Böckler[1], Douglas A. Bell[2] and Gareth L. Bond[1]*

Decades of research have shown that mutations in the p53 stress response pathway affect the incidence of diverse cancers more than mutations in other pathways. However, most evidence is limited to somatic mutations and rare inherited mutations. Using newly abundant genomic data, we demonstrate that commonly inherited genetic variants in the p53 pathway also affect the incidence of a broad range of cancers more than variants in other pathways. The cancer-associated single nucleotide polymorphisms (SNPs) of the p53 pathway have strikingly similar genetic characteristics to well-studied p53 pathway cancer-causing somatic mutations. Our results enable insights into p53-mediated tumour suppression in humans and into p53 pathway-based cancer surveillance and treatment strategies.

There are substantial differences between individuals with regard to their risk of developing cancer, of progressing to high-grade tumours and in response to therapies. This heterogeneity is a major obstacle to designing uniformly effective prevention, screening and treatment strategies and it motivates the large effort to personalize these strategies using biomarkers[1]. Commonly inherited genetic variants, such as single nucleotide polymorphisms (SNPs) hold great promise as easily obtainable and measurable biomarkers. More than 1,000 SNPs have been shown to associate significantly with human cancer in genome-wide association studies (GWAS) conducted in a broad range of solid and haematological malignancies. However, despite these findings, major challenges remain in translating these associations into clinical applications[2,3]. For example, discerning the functional consequences of the variant, as well as the genes and molecular pathways connecting the variant with disease, has proved extremely challenging. This limited understanding of the biology behind these significant associations has clearly constrained our ability to integrate SNP biomarkers into the proper genetic, cellular and clinical context to maximize their effective use in the clinic.

In recent years, the field of human genetics has made great strides in generating datasets that are crucial for revealing the mechanistic relationship between SNPs and tumorigenesis. For example, data generated by the 1000 Genomes Project reveal the genetic diversity in individuals and populations and have been crucial for identifying haplotypes that are linked to diseases studied with GWAS[4]. Moreover, functional genomic scans of gene regulatory features, such as transcription factor binding or specific histone modifications, using chromatin immunoprecipitation coupled with sequencing (ChIP-seq) can connect these SNPs with functional biological differences[5–8]. Additionally, the advent of expression quantitative trait loci (eQTL) mapping approaches that measure gene expression levels for tens of thousands of transcripts in genotyped samples has provided an intermediate biological phenotype that is useful for interpreting many GWAS associations. In these studies, global gene expression measurements and whole-genome SNP genotypes are correlated in order to connect the abundance of a specific gene transcript with an allelic variant and define eQTLs[9–11].

Analysis of datasets such as these enables rapid assignment of cancer-associated SNPs to well-studied signalling pathways that are known to be important in a broad range of cancers and for which somatic genetic variants are currently used as biomarkers or drug targets in the laboratory and clinic. Identifying which cancer-associated pathways frequently carry SNPs associated

Genome-wide association studies (GWAS). Analysis of the association of genetic variants, typically single nucleotide polymorphisms (SNPs), with a specific trait or disease. They are often very large case–control studies in which SNPs throughout the whole genome are examined for differences in allele frequencies between the two different populations.

**These authors contributed equally to this work. Correspondence to D.A.B. and G.L.B. bell1@niehs.nih.gov; gareth.bond@ndm.ox.ac.uk*

doi:10.1038/nrc.2016.15
Published online 24 Mar 2016

Expression quantitative trait loci
(eQTLs). Genetic variants in the genome, typically single nucleotide polymorphisms (SNPs) or copy number variants, that are associated with differential expression of a gene. Typically, global gene expression measurements and whole-genome SNP genotypes are correlated to connect the abundance of a specific gene transcript with an allelic variant and define eQTLs.

Li–Fraumeni syndrome
(LFS). A familial cancer predisposition syndrome associated with certain cancers arising in multiple tissues, such as soft tissue sarcomas, breast cancer, leukaemia and osteosarcomas; 50% of patients are heterozygous for cancer-causing mutations in *TP53*. Increased cancer risk is extremely high and has been estimated to be 50% by the age of 40 years and 90% by the age of 60 years.

Multiple hypothesis testing
When testing many hypotheses simultaneously, the likelihood of one test reaching a significance threshold of $P < 0.05$ increases. Thus, to reduce the likelihood of false positives, a multiple hypothesis testing correction is applied.

with differential cancer susceptibility could accelerate our biological understanding of the influence of the variants on cancer and the potential clinical utility of SNP biomarkers[12,13]. One of the best studied and important cancer signalling pathways is the p53 tumour suppressor pathway. Decades of intensive research in mice and humans have shown that human genetic variants in the p53 stress response pathway can have key roles in the incidence and survival of many cancers. For example, among individuals with the very rare Li–Fraumeni syndrome (LFS) — who carry inherited, heterozygous mutations in *TP53* (which encodes p53) — the penetrance for cancer onset is 100% by the age of 70 years, with the cancers occurring in numerous tissues, including bone, connective tissues, breast and brain[14,15]. Indeed, candidate SNP studies have clearly demonstrated that p53 signalling can be affected by functional SNPs that in some cases result in differential cancer susceptibility[16–18]. Moreover, very similar somatic mutations in the *TP53* gene are found in more than 50% of all cancer genomes[19,20], making it the most frequently mutated gene that is causally implicated in cancer and the gene mutated in the broadest range of cancer types, including epithelial, mesenchymal and haematological cancers[19,21]. Importantly, recent network analyses that have taken advantage of advances in high-throughput exome sequencing of cancer genomes have shown that the p53 pathway represents the largest, most frequently identified network of genes carrying mutations in the broadest spectrum of cancers identified thus far[22–27]. In fact, common somatic mutations in many genes of the pathway are known to directly affect susceptibility to a broad range of cancer types and are being developed as crucial biomarkers to inform therapy in some patient-stratification strategies in the clinic[14,16,17,28–30]. Moreover, these somatic mutations and low-frequency inherited mutations have been shown to affect cancer risk, progression and response to therapies for many cancers in humans and many mouse models. Such findings are consistent with the well-defined roles of the p53 stress response pathway in tumour suppression, regulating cell migration and invasion and mediating the cellular response to DNA damage-inducing cancer therapies, primarily through the ability of p53 to regulate transcription[31–35].

These observations made in model systems, tumour studies and in a rare familial syndrome (LFS) suggest that genetic variants found in the general human population, such as SNPs, would also be more likely to affect susceptibility to a broader range of cancer types than functional SNPs in other genes and pathways. In this Analysis, we demonstrate that the abundant genomic data generated in the past decade support this prediction. We go on to show that these common SNP variants are similar to well-studied inherited and somatic p53 pathway disease-associated mutations, in that they are frequently found in a high proportion of p53 pathway genes and they associate with multiple types of cancer, but not other diseases. Moreover, they are found almost exclusively in p53 pathway genes that can carry cancer-causing mutations in cancer genomes, thereby suggesting that particular p53 pathway genes are highly sensitive to heritable and somatic genetic variants, resulting in altered tumour suppression in many tissue types. Our results also support a causal relationship between certain classes of RNA-processing SNPs in p53 pathway genes and the noted differential cancer risk.

Cancer variants in p53 pathway genes

Somatic, causal mutations. Cancer driver genes, such as *TP53*, when mutated, can promote tumorigenesis and have been identified through studies of inherited cancer predisposition syndromes, cancer genome sequencing and experimental models of cancer. Although estimates regarding the exact number of these genes vary, one well-used curated list is the COSMIC Cancer Gene Census, which uses sequencing data to identify genes for which the number and pattern of mutations are highly unlikely to be attributable to chance[36]. The current list consists of 493 RefSeq autosomal genes that harbour somatic cancer-promoting mutations. Of the 67 autosomal genes attributed to the p53 pathway (according to the Kyoto Encyclopedia of Genes and Genomes (KEGG)), 15 have been denoted as harbouring somatic, causal mutations in at least one cancer type. This group includes the well-studied oncogenes and tumour suppressor genes ataxia telangiectasia mutated (*ATM*), *MDM2*, cyclin-dependent kinase inhibitor 2A (*CDKN2A*; which encodes INK4A and ARF), *FAS* and *MDM4* (FIG. 1a). Thus, 22.38% of genes of the p53 pathway contain known causal mutations; a significant 11.15-fold enrichment over the 2.01% found in all 24,553 annotated autosomal genes (FIG. 1b; RefSeq, $P = 3.74e^{-12}$, adjusted for multiple hypothesis testing $P = 8.22e^{-10}$).

In order to assess further the importance of the 11.15-fold enrichment of cancer-associated causal mutations in genes of the p53 pathway, we compared it with potential fold enrichments of causally mutated genes in all well-annotated pathways in the genome. To do this, we determined potential enrichments in all 220 signalling pathways annotated by KEGG in the categories of metabolism, genetic information processing, environmental information processing, cellular processes and

Author addresses

[1]Ludwig Institute for Cancer Research, University of Oxford, Nuffield Department of Clinical Medicine, Old Road Campus Research Building, Oxford OX3 7DQ, UK.
[2]Environmental Genomics Group, Genome Integrity and Structural Biology, National Institute of Environmental Health Sciences, National Institutes of Health, Research Triangle Park, North Carolina 27709, USA.
[3]Vall d'Hebron University Hospital, Oncology Department, Passeig de la Vall D'Hebron 119, 08035 Barcelona, Spain.
[4]Molecular and Population Genetics Laboratory, Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK.
[5]Center for Computational Molecular Biology, Brown University, 115 Waterman Street, Providence, Rhode Island 02912, USA.
[6]Department of Molecular Biology, Cell Biology, and Biochemistry, Brown University, 70 Ship Street, Providence, Rhode Island 02903, USA.
[7]Department of Cancer Research and Molecular Medicine, Norwegian University of Science and Technology.
[8]Department of Computer and Information Science, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway
[9]Genetics, Department of Biology, University of Pisa, Via Derna 1, 56126 Pisa, Italy.

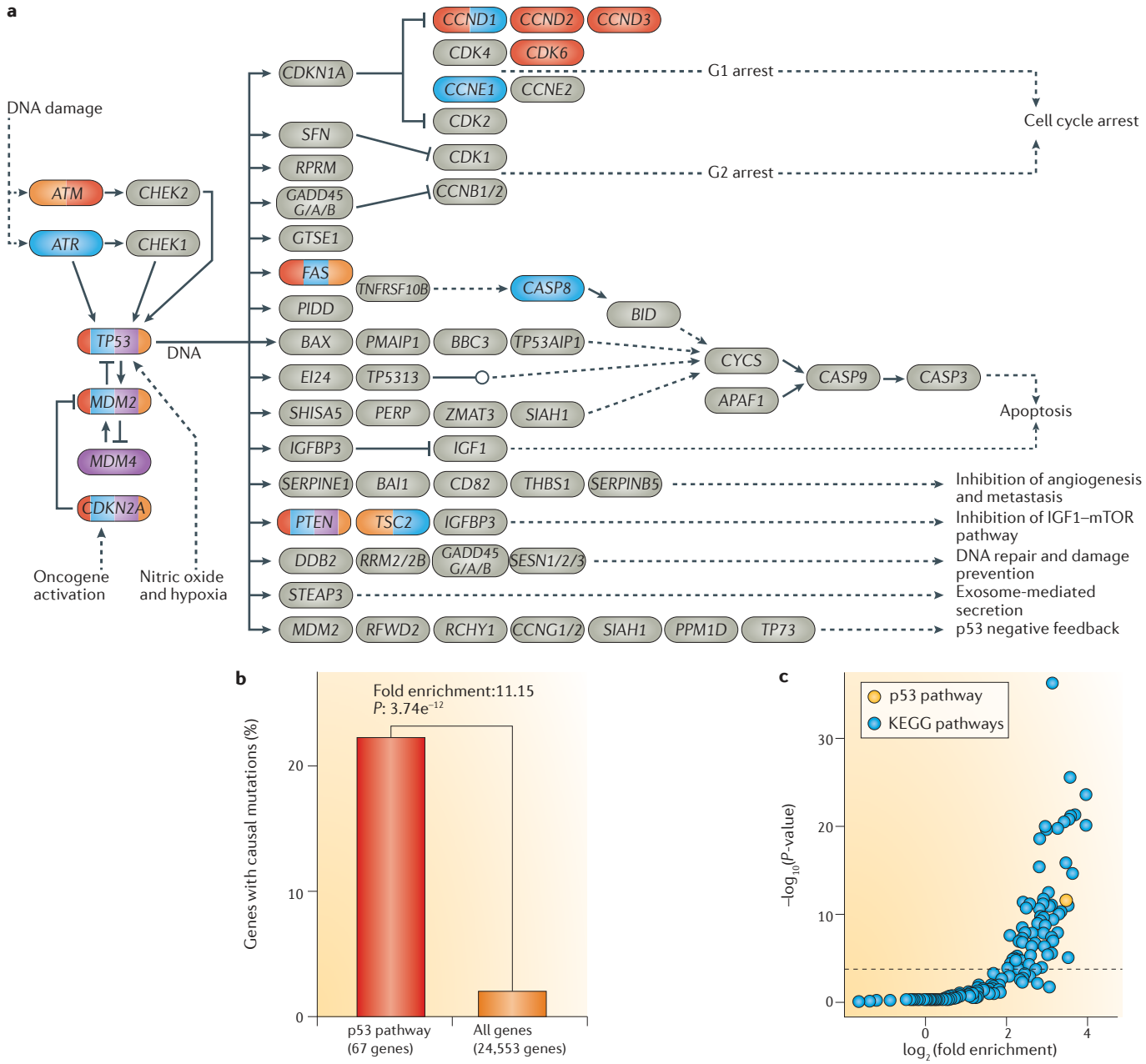

Figure 1 | **Somatic, causal mutations occur in a high proportion of p53 pathway genes. a** | A pathway diagram of the p53 pathway as annotated by Kyoto Encyclopedia of Genes and Genomes (KEGG). Genes for which mutations have been causally implicated in cancer are coloured (COSMIC). Colours indicate causally mutated genes in different cancers: blue, epithelial cancers; red, leukaemia or lymphoma; purple, mesenchymal cancers and orange, other types of cancer. **b** | A bar graph showing the percentage of genes in the p53 pathway with known causal mutations compared with all annotated autosomal genes of the genome. Fifteen of 67 genes in the p53 pathway (22.38%) are known to be causally mutated, which represents a significant 11.15-fold enrichment over the rest of the genes in the genome ($P = 3.74e^{-12}$ is also depicted). **c** | A scatter plot showing the fold enrichment of causally mutated genes on the *x*-axis ($\log_2$ scale) and the *P*-value on the *y*-axis ($-\log_{10}$ scale). The horizontal dashed line represents the 5% family-wise error rate threshold (Bonferroni-adjusted *P*-value: 0.05). The enrichment of causal mutations in p53 pathway genes is marked in yellow and the other 220 annotated KEGG pathways are marked in blue. Overall, 30% of pathways demonstrated significant enrichment of causally mutated genes. *APAF1*, apoptotic peptidase activating factor 1; *ATM*, ataxia telangiectasia mutated; *ATR*, ataxia telangiectasia and Rad3 related; *BAI1*, adhesion G protein-coupled receptor B1; *BAX*, BCL-2-associated X protein; *BBC3*, BCL-2 binding component 3; *BID*, BH3 interacting domain agonist; *CASP3*, caspase 3; *CCND1*, cyclin D1; *CDKN1A*, cyclin-dependent kinase inhibitor 1A; *CYCS*, cytochrome *c*, somatic; *DDB2*, damage specific DNA binding protein 2; *GADD45G/A/B*, growth arrest and DNA damage inducible gamma, alpha or beta; *GTSE1*, G2 and S phase expressed 1; *IGF1*, insulin like growth factor 1; *IGFBP3*, IGF binding protein 3; *PERP*, TP53 apoptosis effector; *PIDD*, p53-induced death domain protein 1; *PMAIP1*, phorbol-12-myristate-13-acetate-induced protein 1; *PPM1D*, protein phosphatase, Mg^{2+}/Mn^{2+} dependent 1D; *RCHY1*, ring finger and CHY zinc finger domain containing 1; *RFWD2*, ring finger and WD repeat domain 2, E3 ubiquitin protein ligase; *RPRM*, reprimo; *RRM2*, ribonucleotide reductase regulatory subunit M2; *SERPINE1*, serpin peptidase inhibitor, clade E; *SESN2*, sestrin 2; *SFN*, stratifin; *THBS1*, thrombospondin 1; *TSC2*, tuberous sclerosis 2; *ZMAT3*, zinc finger matrin-type 3.

Linkage disequilibrium (LD). The non-random association of alleles of two or more single nucleotide polymorphisms (SNPs). LD is influenced by many factors, including mutation rate, recombination, chromosomal distance, natural selection and genetic drift. It has been extensively used in the design and interpretation of genome-wide association studies (GWAS). A commonly used measure of linkage disequilibrium between two loci is the squared correlation or r^2.

organismal systems. We found that 66 of the 220 cellular signalling pathways (30%), including the p53 pathway, demonstrated significant enrichment of causally mutated genes after correction for multiple hypothesis testing (FIG. 1c; see Supplementary information S1 (methods) and Supplementary information S2 (table) for details). Thus, the enrichment noted in the p53 pathway places it among the top 5% of all well-annotated pathways of the genome. The other significantly enriched pathways also include frequently mutated tumour suppressor genes and oncogenes, such as PI3K catalytic subunit-α (*PIK3CA*), adenomatous polyposis coli (*APC*) and *KRAS*[19] (Supplementary information S2 (table)). It is important to note that the enrichment of causally mutated genes in the p53 pathway was also found to be in the top 5% of all pathways when we used a list of causally mutated genes generated using various criteria by different researchers (Supplementary information S3 (figure))[1].

Cancer-associated SNPs. Together with the noted high cancer risk among carriers of *TP53* mutations[15], the significant enrichment of causally mutated genes in the p53 pathway suggests that inherited SNPs in p53 pathway genes could affect cancer susceptibility to a greater extent than SNPs in other annotated signalling pathways. To begin to test this, we used the GWAS catalog (download date: 12 October 2015) and the 10th revision of *International Classification of Diseases* (ICD10) to identify all cancer susceptibility GWAS studies that have been undertaken to date (Supplementary information S1 (methods)). We first identified all GWAS studies that were designed to study disease susceptibility and noted that the vast majority of studies have been performed in European populations ($n = 756$). Each of the 756 GWAS can be attributed to one of the major ICD10 disease categories, which include Neoplasms (Supplementary information S4 (table)). We found that 19 different ICD10 disease categories have at least one GWAS study (average: 39.8 studies per disease category). Importantly for this Analysis article, Neoplasms had the most studies attributed to it with 165. These 165 studies were undertaken to assess differential susceptibility to a broad range of cancers (ICD10 subcategory, Supplementary information S4 (table)) with a median of 11,647.5 individuals with cancer per study. If our hypothesis is correct, we would expect that genes of the p53 pathway would overlap with the cancer susceptibility loci (CSLs) identified in these GWAS studies to a greater extent than the genes of other well-defined signalling pathways.

To test this, we first determined which CSLs mapped to the 24,553 autosomal RefSeq genes in the genome. We began by using the 1000 Genomes Phase 3 dataset to identify all known SNPs (minor allele frequency (MAF) ≥0.001) from European populations within ±10 kb of the gene boundaries of the 24,553 autosomal RefSeq genes and found 7,106,459 SNPs in total. Subsequently, we mined the GWAS catalogue, and extracted the 1,034 SNPs (750 unique loci) associated with susceptibility to approximately 17 different types of cancer in European populations, including epithelial, mesenchymal and haematological cancers (FIG. 2). Next, we augmented this dataset with SNPs in linkage disequilibrium (LD) ($r^2 = 1.0$, MAF ≥0.001) in European populations using data from the 1000 Genomes Phase 3 dataset. On average, a cancer GWAS SNP was in perfect LD with 4.926 SNPs (range 1–126 SNPs), and the linked SNPs spanned an average genomic region of 6,947.132 bp (range 1–201,267 bp). Our final dataset consists of a total of 3,454 unique cancer GWAS SNPs.

Using our parameters, which require at least one cancer GWAS SNP to reside within ±10 kb of an annotated gene body, we determined that 2,262 (65.5%) of 3,454 cancer GWAS SNPs mapped to 541 autosomal genes, which we refer to as cancer susceptibility genes (CSGs; FIG. 3a; Supplementary information S5 (table)). Interestingly, 10 of the 67 p53 pathway genes (14.93%) are CSGs, namely: *ATM*, checkpoint kinase 2 (*CHEK2*), caspase 8 (*CASP8*), cyclin D1 (*CCND1*), *CCND2*, *CCNE1*, *CDKN2A*, *FAS*, *MDM4* and *TP53*. This 14.93% of p53 pathway genes represents a significant 6.77-fold enrichment compared with the 2.2% of 24,553 annotated autosomal genes that are CSGs ($P = 2.00e^{-06}$, adjusted $P = 4.39e^{-04}$; FIG. 3b). It is important to note that all but one p53 pathway CSG (*MDM4*) have been attributed to at least one other KEGG annotated pathway. Thus, in order to assess further the importance of the 6.77-fold enrichment of CSGs in the p53 pathway, we compared it with the potential enrichment of CSGs in all 220 well-annotated pathways in the genome. Only 3 of these 220 cellular signalling pathways (1.36%), including the p53 pathway, demonstrated significant enrichments of CSGs after correction for multiple hypothesis testing (FIG. 3c; Supplementary information S6 (table)). The two other significant pathways are PI3K-AKT and Adherens Junction. Like the p53 pathway, these are also known to be important pan-cancer signalling pathways (KEGG cancer signalling pathways). However, the enrichment of CSGs in the p53 pathway ranks highest for both the level of significance and the fold enrichment (FIG. 3c). Specifically, the PI3K-AKT and Adherens Junction pathways are associated with fold enrichments of 2.73 ($P = 5.42e^{-05}$) and 5.11 ($P = 1.66e^{-04}$), respectively, compared with a fold enrichment of 6.77 ($P = 2.00e^{-06}$) for the p53 pathway. Together, the results of these analyses thus far suggest that in the p53 pathway somatic, causal mutations and inherited cancer-associated SNPs occur in a high proportion of p53 pathway genes relative to other pathways.

Expression quantitative trait loci. The fact that in the above analyses we required that one or more of the cancer GWAS SNPs reside within 10 kb of gene boundaries to define CSGs increases the likelihood that a SNP will be *cis*-acting and will functionally affect expression of its proximal gene or the protein it encodes. However, in order to gain more certainty that the SNPs can functionally affect the genes in which they reside, we further restricted our analyses to those SNPs, in and around genes, with genotypes that also associate with mRNA levels of their proximal genes in expression quantitative trait loci (eQTL) studies (*cis*-acting, *cis*-eSNPs). To do this, we began by curating data from 14 publicly available

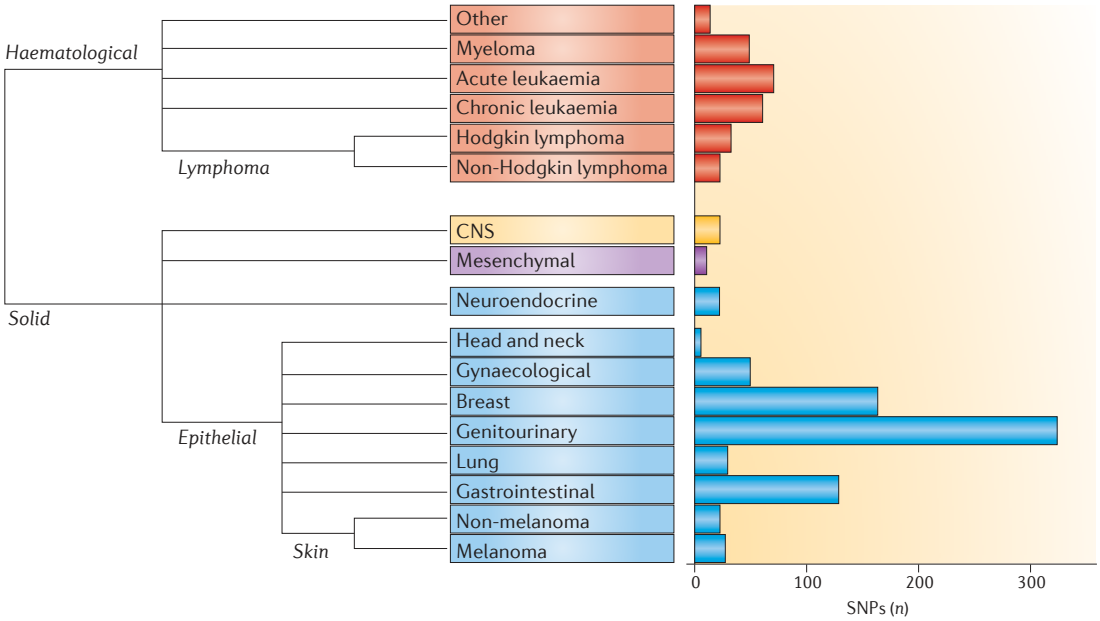

Figure 2 | **One hundred and sixty-five genome-wide association studies (GWAS) of many types of cancer have been carried out in European populations.** A histopathological classification of all the cancers present in the US National Human Genome Research Institute (NHGRI) GWAS catalogue (download date: 12 October 2015) and a bar graph illustrating the number of tag single nucleotide polymorphisms (SNPs) that have been found to significantly associate with differential susceptibility to a particular cancer. Cancers are also classified as epithelial (blue), lymphoma/leukaemia (red), mesenchymal (purple) and other (orange). CNS, central nervous system.

eQTL studies performed in non-cancerous tissues and cells from people of primarily European descent[37–51]. These studies used five different cell types (lymphoblastoid cell lines (LCLs), $CD4^+$ T cells, primary monocyte samples, B cells and peripheral blood cells), as well as cells from six primary tissue types (adipose, skin, liver, intestine, heart and lung). The median number of samples included in these studies is 659 (range 129–1,490). We selected *cis*-eQTLs within ±10 kb of the gene boundaries of all 24,553 RefSeq genes and identified a total of 412,962 unique *cis*-eSNPs, including 8,891 in B cells, 71,242 in $CD4^+$ T cells, 2,130 in cardiac tissue, 4,844 in intestine, 33,256 in adipose tissue, 265,671 in LCLs, 2,664 in liver, 6,163 in lung, 133,425 in monocytes, 17,345 in peripheral blood cells and 26,417 in skin. Of these *cis*-eSNPs, 75.06% were found to be in a single tissue, 16.05% in two tissues, 6% in three tissues and 2.89% in four or more tissues.

In total, we identified 11,887 genes genome-wide that harboured *cis*-eSNPs. In 133 of these genes (1.12%), we observed that the eSNPs also overlap (that is, are in complete LD) with cancer GWAS SNPs. Thus, these genes harbour haplotype blocks that associate with both differential cancer susceptibility and proximal gene expression in at least one tissue or cell type analysed (eCSGs; FIG. 4a; Supplementary information S5 (table)). Interestingly, 6 of the eCSGs are among the 51 p53 pathway genes that harbour *cis*-eSNPs (11.76%): *FAS*, *MDM4*, *ATM*, *CCND1*, *CASP8* and *CCNE1*. This represents a significant 10.51-fold enrichment compared with the 1.12% found in all 11,887 annotated genes that harbour at least one *cis*-eSNP ($P = 2.09e^{-05}$, adjusted $P = 4.59e^{-03}$; FIG. 4b). Importantly, of the 220 annotated cellular signalling pathways, the p53 pathway is the only one that shows a significant enrichment of eCSGs (FIG. 4c; Supplementary information S7 (table)); in fact, 61.8% of pathways do not have any eCSGs. These results demonstrate that even when we restrict our analyses to SNPs that can associate with differential gene expression of their proximal genes, cancer-associated SNPs still occur in a high proportion of pathway genes relative to all annotated signalling pathways.

p53 genes are not enriched in SNPs associated with other diseases. These results clearly lend support to the hypothesis that commonly inherited genetic variation in p53 pathway genes will affect cancer susceptibility to a greater extent than the variation found in genes of other pathways. However, the p53 stress response pathway has also been implicated in the pathogenesis of many other diseases that have been studied in GWAS, including neurological[52–56], cardiovascular[57,58] and infectious[59,60] diseases. Therefore, we next explored the potential impact of p53 pathway SNPs on susceptibility to non-cancerous disease. To do this, we took advantage of the 591 GWAS studies that have been carried out to measure the genetic basis of differences in susceptibility to other non-cancer diseases in Europeans. As mentioned above, 18 ICD10 disease categories, other than Neoplasms, had at least one susceptibility GWAS study attributed to it (Supplementary information S4 (table)). In the same manner as described above for our analysis of the 165 cancer GWAS (ICD10 category Neoplasms), we identified a set of genes for each of the other 18 disease categories in which at least one GWAS SNP was found to reside within ±10 kb of

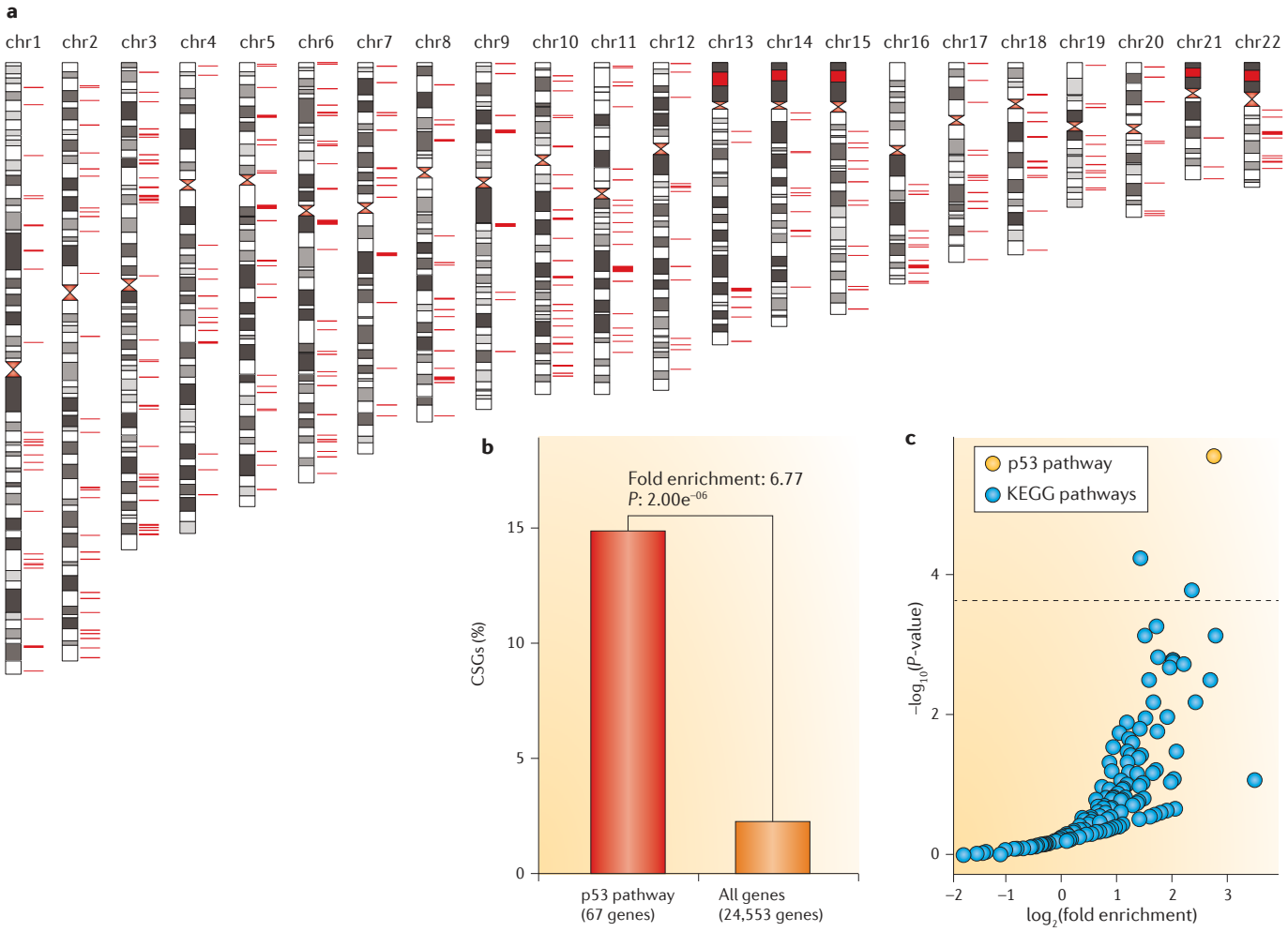

Figure 3 | **Cancer-associated single nucleotide polymorphisms (SNPs) occur in a high proportion of p53 pathway genes.** **a** | A karyogram of the 541 genes that harbour at least one SNP within 10 kb of their boundaries (cancer susceptibility genes; CSGs) as determined by cancer genome-wide association studies (GWAS). CSGs are shown in red. **b** | A bar graph of the percentage of CSGs in the p53 pathway compared with all annotated genes of the genome. Ten CSGs are found among the 67 genes of the p53 pathway (14.93%), which represents a significant 6.77-fold enrichment compared with the rest of the autosomal genes in the genome ($P = 2.00e^{-06}$ depicted in the figure, adjusted $P = 4.39e^{-04}$). **c** | A scatter plot showing the fold enrichment of CSGs on the *x*-axis ($\log_2$ scale), and the adjusted *P*-value on the *y*-axis ($-\log_{10}$ scale). The horizontal dashed line represents the 5% family-wise error rate threshold (Bonferroni-adjusted $P = 0.05$), which is the pre-fixed significance threshold. The enrichment of CSGs in p53 pathway genes (in yellow) is the highest and the most significant compared with the other 220 annotated Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (in blue). Overall, 1.36% of pathways demonstrated significant enrichments of CSGs. chr, chromosome.

an annotated gene body (which we term susceptibility genes; SGs). All 18 disease categories had at least 4 SGs (median 88.5, range 4–708). We then explored any potential enrichment of SGs for each disease category in all 220 signalling pathways in the genome, including the p53 pathway. For 8 of the 18 disease categories, we were able to identify an average of 4.25 signalling pathways that were significantly enriched for non-cancer SGs after correction for multiple hypothesis testing (range 1–8 pathways; FIG. 5; Supplementary information S8 (table)). However, in contrast to our findings for cancer (ICD10 category Neoplasms), the p53 signalling pathway was not significantly enriched in any of these 8 non-cancerous disease categories, which include diseases of the nervous, circulatory, digestive and musculoskeletal systems.

p53 pathway enrichment is consistent across pathway annotations. We have demonstrated that the autosomal genes (FIG. 3) of the p53 pathway overlap with cancer GWAS loci to a greater extent than the genes of 219 other well-annotated signalling pathways, and that this enrichment is limited to cancer GWAS (FIG. 5). Thus far, we have exclusively used KEGG pathway annotations for our analyses. In order to explore further the significance of our observations, we extended our analyses to pathway annotations from two different well-used, curated pathway databases, namely BioCarta and PANTHER. Similar to the findings using KEGG pathway annotation, the enrichment of CSGs in the p53 pathway ranks highest for the level of significance relative to all other pathways annotated by either BioCarta (FIG. 6a;

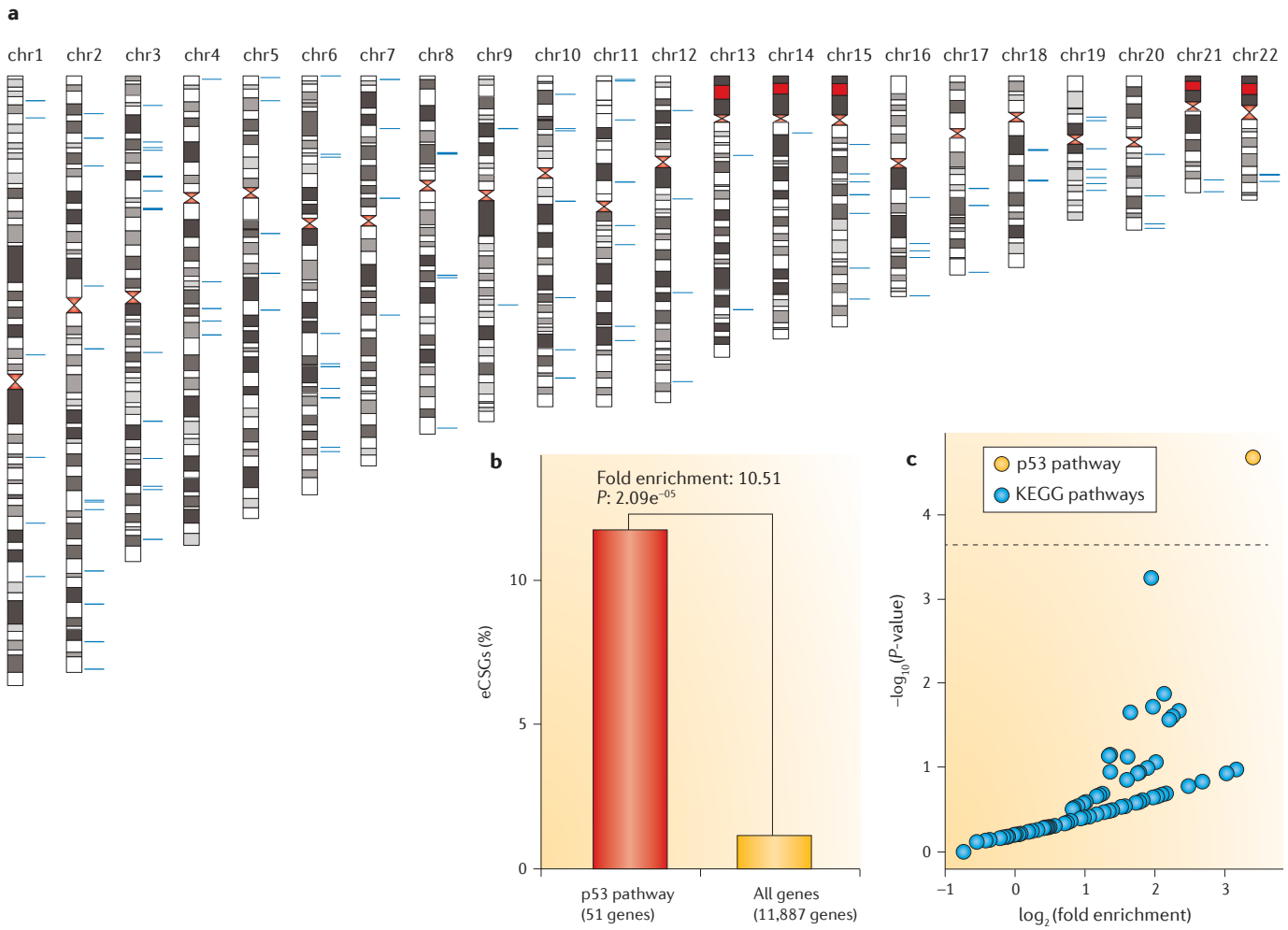

Figure 4 | **Cancer-associated expression quantitative trait loci (eQTLs) occur in a high proportion of p53 pathway genes. a** | A karyogram of the 133 genes that harbour a single nucleotide polymorphism (SNP) that also associates with mRNA levels of their proximal genes in eQTL studies within 10 kb of their boundaries (*cis*-eSNP), which overlaps at least one cancer genome-wide association study SNP (*cis*-e-cancer susceptibility gene (CSG)). *Cis*-eCSGs are shown in red. **b** | A bar graph of the percentage of eCSGs in the p53 pathway compared with all annotated genes of the genome. Six *cis*-eCSGs are found among the 67 genes of the p53 pathway, which represents a significant 10.51-fold enrichment compared with the 11,887 genes with at least one eSNP ($P = 2.09e^{-05}$ denoted in the graph, adjusted $P = 4.59e^{-03}$). **c** | A scatter plot showing the fold enrichment of eCSGs on the *x*-axis ($\log_2$ scale), and the adjusted *P*-value on the *y*-axis ($-\log_{10}$ scale). The horizontal dashed line represents the 5% family-wise error rate threshold (Bonferroni-adjusted $P = 0.05$), which is the pre-fixed significance threshold. The enrichment of *cis*-eCSGs in the genes of the p53 pathway (in yellow) is the highest and the most significant compared with all the other Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways (in blue). Overall, 0.45% of pathways demonstrated significant enrichment of *cis*-eCSGs. chr, chromosome.

Supplementary information S9 (table)) or PANTHER (FIG. 6b; Supplementary information S10 (table)). Also similar to the analyses conducted with KEGG annotation, when we explored the potential enrichment of p53 pathway genes among the susceptibility loci of the other 18 disease groupings defined above, we found no significant enrichment for the p53 signalling pathway annotated by either database (FIG. 6a,b, additional panels).

p53 pathway variants among cancers

Causal mutations. As mentioned above, the ability of p53 to suppress tumour formation in numerous tissues has been demonstrated in several mouse models[29,30,61]. Indeed, some of the genes of the p53 pathway have been found to be causally mutated in cancer genomes from all four major annotated tissue type groupings: epithelial, mesenchymal, leukaemia or lymphoma and other (COSMIC). In FIG. 1, we demonstrate that p53 pathway genes are enriched in genes known to harbour causal, somatic mutations in at least one of these four tissue types, whereby the enrichment noted places the p53 pathway among the top 5% of all annotated pathways of the genome (KEGG). Interestingly, we find similar significant enrichments when we restrict our analyses to causal, somatic mutations found in the various cancer types. We find that, in all four major cancer types, p53 pathway genes were enriched in causally mutated genes (FIG. 7a).

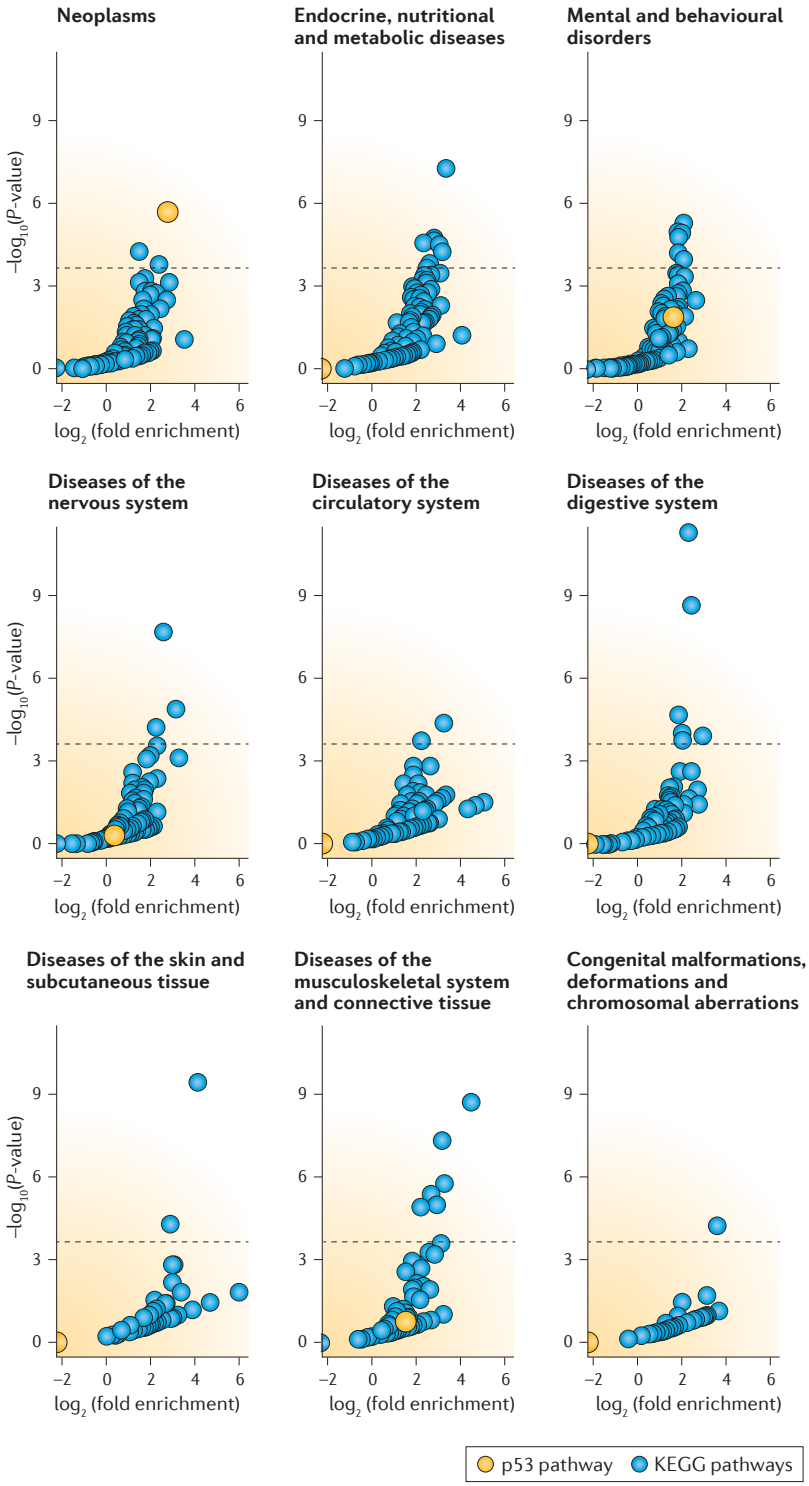

Figure 5 | **p53 pathway genes are significantly enriched in cancer susceptibility genes, but not susceptibility genes for other major disease groupings.** Scatter plots showing fold enrichment of susceptibility genes (SGs) in Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways for the 9 of 19 International Classification of Diseases revision 10 (ICD10) disease groups that had at least one pathway significantly enriched in SGs. The fold enrichment of SGs is on the *x*-axis ($\log_2$ scale) and the adjusted *P*-value is on the *y*-axis ($-\log_{10}$ scale). The horizontal dashed line represents the 5% family-wise error rate threshold (Bonferroni-adjusted $P = 0.05$), which is the pre-fixed significance threshold. The enrichment of p53 pathway SGs are shown in yellow; significant enrichment is observed in cancer (Neoplasms), but not in any other disease groups.

In epithelial cancers, 14.93% of p53 pathway genes can be causally mutated, which represents an 18.99-fold enrichment over the 0.79% causally mutated genes found in all 24,553 annotated autosomal genes ($P = 1.20e^{-10}$, adjusted $P = 2.64e^{-08}$; FIG. 7a). Of the 220 cellular signalling pathways 61 (27.7%), including the p53 pathway, demonstrated significant enrichment of causally mutated genes after correction for multiple hypothesis testing (FIG. 7a; Supplementary information S11 (table)), thereby placing the enrichment noted in the p53 pathway among the top 6.82% of all pathways. Similar significant enrichments were found in leukaemias/lymphomas (fold-enrichment 14.09; $P = 2.18e^{-09}$, adjusted $P = 4.79e^{-07}$), mesenchymal cancers (fold-enrichment 20.36; $P = 4.77e^{-06}$, adjusted $P = 1.05e^{-03}$) and cancers in the Other category (fold-enrichment 45.81; $P = 1.70e^{-10}$, adjusted $P = 3.75e^{-8}$). As seen in FIG. 7b, there are only 13 signalling pathways (5.9%), including p53, that are significantly enriched in causally mutated genes shared by all four major cancer types. These pan-cancer signalling pathways include many other well-studied oncogenic and tumour suppressor pathways, such as the PI3K-AKT, RAS and MAPK signalling pathways (Supplementary information S11 (table)).

Cancer-associated SNPs. Given the observation that the p53 pathway belongs to the 5.9% of all 220 signalling pathways that are enriched in causally mutated genes in all four major cancer types, we next wanted to explore whether similar observations can be found among the cancer-associated SNPs. In FIG. 3, we demonstrate that p53 pathway genes are enriched in genes overlapping CSLs (CSGs), whereby the noted 6.77-fold enrichment places the p53 pathway at the top of all 220 annotated signalling pathways for CSG enrichment (KEGG). However, we also find similar significant enrichments when we restrict our analyses to CSLs found in the individual cancer types.

In epithelial cancers, 10.45% of the 67 p53 pathway genes are CSGs, which represents a 6.59-fold enrichment over the 1.58% CSGs found in all 24,553 annotated autosomal genes ($P = 9.12e^{-05}$, adjusted $P = 2.01e^{-02}$; FIG. 7c). Interestingly, only the p53 pathway demonstrated significant enrichments of CSGs after correction for multiple hypothesis testing (FIG. 7c). A similar significant enrichment of CSGs in the p53 pathway was found in leukaemias/lymphomas (FIG. 7c; Supplementary information S12 (table); fold-enrichment 12.64; $P = 4.82e^{-05}$, adjusted $P = 1.06e^{-02}$). No significant enrichments were noted for any of the 220 pathways in the mesenchymal cancers or cancers in the Other category. However, this is probably due to the relatively fewer number of studies that have been carried out in these categories of cancer (FIG. 2a). Importantly, and as seen in FIG. 7d, only the genes of the p53 pathway are significantly enriched in CSGs in more than one cancer type. Together, these results clearly demonstrate that p53 pathway mutations and cancer-associated SNPs both occur in a high proportion of p53 pathway genes in multiple cancer types relative to other cellular signalling pathways.

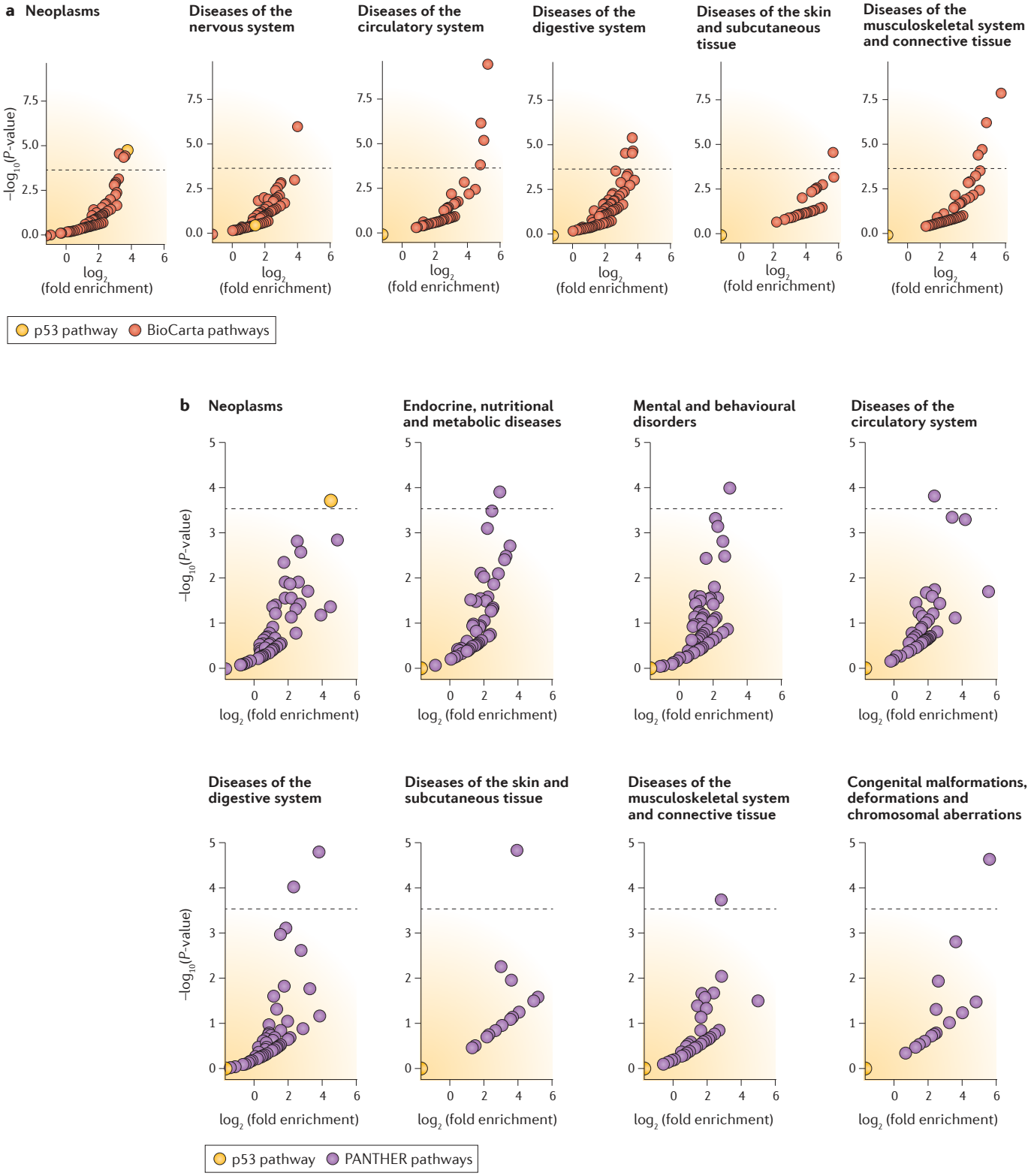

Figure 6 | **Cancer susceptibility gene enrichment in p53 pathway genes is not limited to Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway annotation.** Scatter plots showing the fold enrichment of susceptibility genes (SGs) in BioCarta (panel **a**) and PANTHER (panel **b**) annotated pathways for Interntional Classification of Diseases revision 10 (ICD10) disease groups with at least one significant pathway. The fold enrichment of SGs is reported on the *x*-axis ($\log_2$ scale), and the adjusted *P*-value on the *y*-axis ($-\log_{10}$ scale). The horizontal dashed line represents the 5% family-wise error rate threshold (Bonferroni-adjusted $P = 0.05$). The enrichment of SGs in the p53 pathway is observed in cancer (Neoplasms) with both pathway annotations, but never for other diseases.

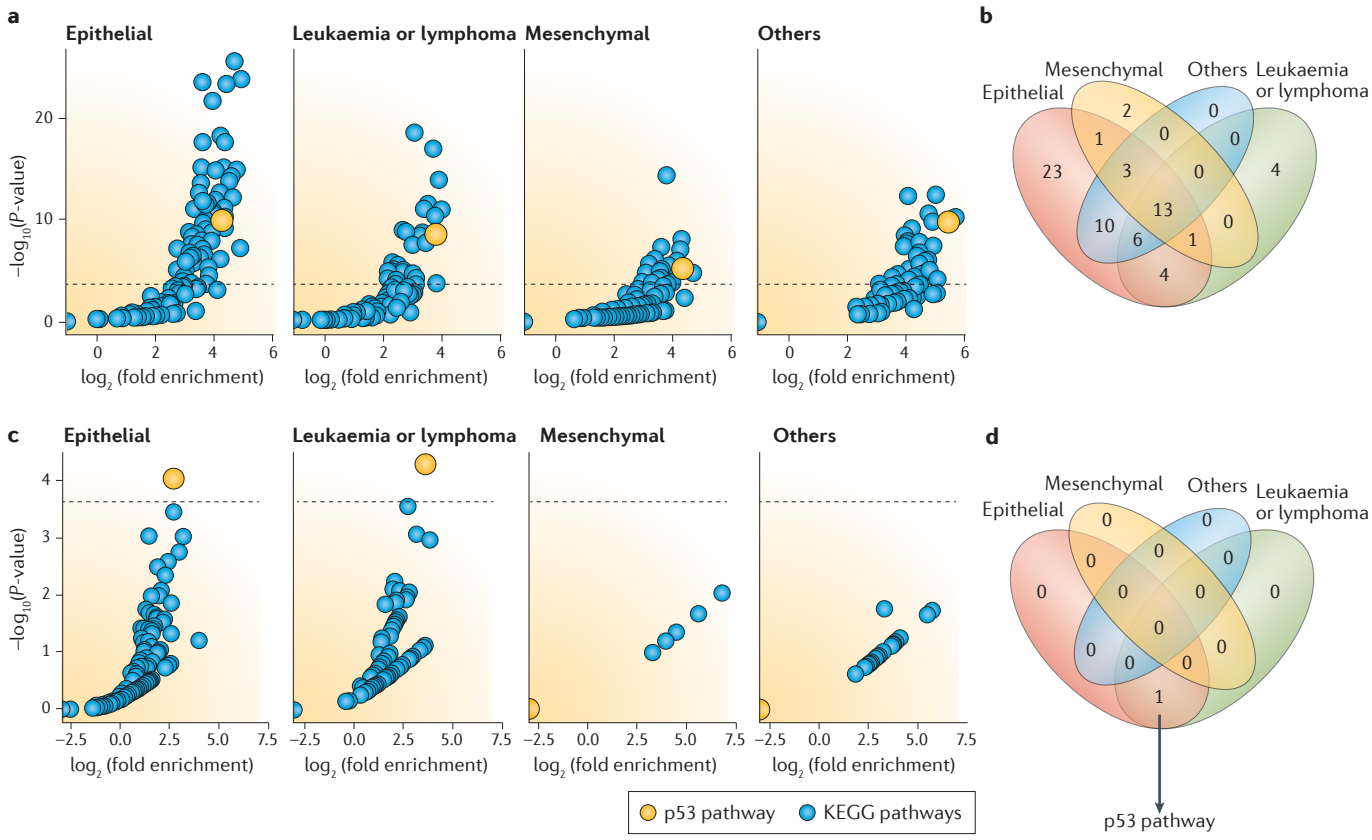

Figure 7 | **Both p53 pathway mutations and cancer-associated single nucleotide polymorphisms (SNPs) occur in a high proportion of pathway genes in multiple cancer types. a** | Scatter plots show the enrichment of genes with causal, somatic mutations in Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways grouped by cancer type. Fold enrichment of causally mutated genes is reported on the *x*-axis ($\log_2$ scale), and the adjusted *P*-value on the *y*-axis ($-\log_{10}$ scale). The horizontal dashed line represents the 5% family-wise error rate threshold (Bonferroni-adjusted $P = 0.05$). **b** | A Venn diagram showing the number of pathways with a significant enrichment of causally mutated genes across the four different types of cancer considered. **c** | Analogously, a scatter plot shows the enrichment of cancer susceptibility genes (CSGs) in KEGG pathways grouped by cancer type **d** | A Venn diagram showing the number of pathways with significant enrichment of CSGs grouped by cancer type. For all scatter plots the p53 pathway is in yellow.

SNPs and mutations in similar p53 genes

As mentioned above, of the 67 autosomal genes attributed to the p53 pathway (KEGG), 15 have been denoted as harbouring somatic, causal mutations in at least one cancer type (COSMIC). In our analysis thus far, we determined that 9 of these genes (60%) are CSGs (FIG. 8a): *ATM*, *CASP8*, *CCND1*, *CCND2*, *CCNE1*, *CDKN2A*, *FAS*, *MDM4* and *TP53* (FIG. 8b; COSMIC)[1,30,62–76]. This represents a significant 4.02-fold enrichment compared with the 10 CSGs (14.93%) found in all 67 p53 pathway genes (hypergeometric test, $P = 1.06e^{-6}$; FIG. 8c). This dramatic enrichment clearly demonstrates that genes in the p53 pathway known to harbour causal somatic mutations in cancer genomes are more likely to also harbour SNPs associated with differential cancer risk as measured in GWAS. However, it is important to note that this is not limited to the genes of the p53 pathway. Specifically, similar enrichments, albeit smaller, can be found among causally mutated genes that are not in the p53 pathway. For example, when we restrict our analysis to the 478 causally mutated genes not attributed to the p53 pathway (KEGG), we observe a significant enrichment of CSGs relative to the 24,486 non-p53 pathway genes of the genome, but to a lesser degree (fold enrichment: 3.09; hypergeometric test, $P = 2.12e^{-08}$; FIG. 8d).

p53 pathway SNPs and RNA processing

In this study, we have determined that 50 SNPs in 10 p53 pathway genes (KEGG) have been either directly or indirectly found to associate with differential cancer risk in GWAS (average of 5 SNPs per gene, range 1–16; see Supplementary information S13–S14 (tables) for details). In contrast to causal somatic tumour mutations, cancer-associated SNPs are single nucleotide variations that, on average, cannot have negatively affected reproductive success, as they occur at relatively high frequencies in the human population. This obvious difference between inherited and somatic genetic variation results in the vast majority of SNPs having weaker net effects on protein activity, and ultimately cancer development, than somatic mutations. These lower penetrant effects represent a major ongoing challenge in determining the molecular underpinnings of significant SNP associations found in GWAS, along with the fact that the responsible,

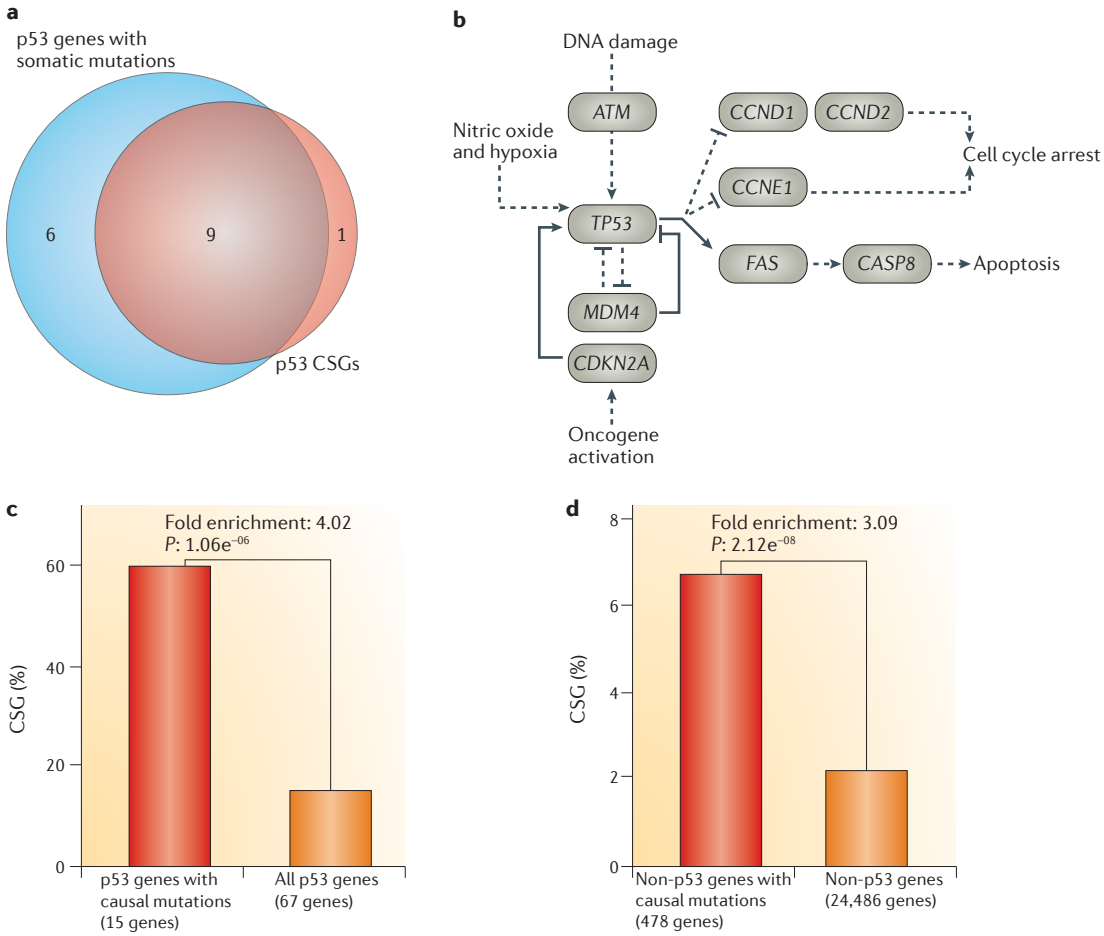

Figure 8 | **p53 pathway cancer susceptibility genes (CSGs) are causally mutated in cancer. a** | A Venn diagram depicting the overlap of p53 pathway genes that harbour causal mutations and pathway genes that are CSGs. **b** | A pathway diagram of the CSGs with causal somatic mutations annotated to the p53 pathway. **c** | A bar graph depicting the percentage of CSGs found among those p53 pathway genes with known causal mutations in cancers and all p53 pathway genes. **d** | A bar plot depicting the enrichment of CSGs among non-p53 pathway genes known to harbour causal somatic mutations and all non-p53 pathway genes. *ATM*, ataxia telangiectasia mutated; *CASP8*, caspase 8; *CCND1*, cyclin D1; *CDKN2A*, cyclin dependent kinase inhibitor 2A.

causal SNP(s) are often linked with many non-functional SNPs[2]. However, for 2 of the 50 cancer-associated SNPs there is mounting experimental evidence that they reside in RNA-processing regulatory elements.

Specifically, one SNP, in the 3′ untranslated region (3′ UTR) of the *TP53* gene (rs78378222, A/C) resides in a canonical polyadenylation signal sequence. This poly(A) SNP was identified in a GWAS for basal cell carcinoma, whereby the risk allele (C) is predicted to disrupt the poly(A) signal in the gene by changing AATAAA to AATACA[77]. Such a disruption of the poly(A) signal could impede cleavage of the nascent mRNA and addition of the poly(A) tail, ultimately resulting in less cellular p53 and potentially less p53-mediated tumour suppression. Two recent studies have provided data that support the proposal that the A to C change does result in aberrant 3′ end processing[77,78]. The other regulatory SNP resides in a 5′ splice site (donor) at the exon 4–intron 4 boundary in the *CCND1* gene (rs9344, 870G→A). This 5′ splice site SNP was found to associate with differential risk for t(11;14)(q13;q32) multiple myeloma in a GWAS, whereby the G-allele, which creates the stronger 5′ splice site (CCGgtaagt compared with CCAgtaagt), was found to associate with increased risk[79]. Alternative splicing of this exon and the potential role of this SNP in affecting the 5′ splice site was first reported more than 20 years ago[80]. Indeed, many subsequent studies conducted in various cell types have demonstrated an association of the A-allele with less exon 4 splicing, resulting in the production of the cyclin D1b variant[79,81,82]. The functional influence of this variant on cancer risk, relative to the cyclin D1a variant, remains to be further elucidated. However, in non-GWAS this heavily studied SNP has frequently been found to associate with differential risk of many cancer types, such as basal cell carcinoma and renal cell carcinoma, bladder, breast, colorectal, oesophageal, gastrointestinal, head and neck, lung, ovarian and prostate cancer, and hepatoblastoma and leukaemia[83].

Together, these data clearly demonstrate that 2 of the 50 p53 pathway cancer GWAS SNPs reside in regulatory elements that affect differential RNA processing. If there is a causal relationship between the RNA processing

SNPs and the noted differential cancer risk, we could expect similar RNA processing SNPs among the 16,890 p53 pathway SNPs to be significantly enriched in cancer GWAS. To test this, we determined the occurrence of similar poly(A) and 5′ splice site SNPs among all 16,890 SNPs in and around all 67 p53 pathway genes (KEGG). We identified only two additional SNPs that, like *CCND1* rs9344, also reside in the exonic region of the 9-mer 5′ splice site sequences and are predicted to demonstrate similar allelic differences in splice site recognition. The additional two SNPs reside in the *CCNB2* and thrombospondin 1 (*THBS1*) genes (Supplementary information S15 (table)). Thus, of the three SNPs in the 5′ splice sites of p53 pathway genes, one (33.3%) is a cancer GWAS SNP (the above-mentioned *CCND1* rs9344, 870G→A). This represents a significant enrichment compared with both the 50 (0.29%) cancer GWAS SNPs found among the total 16,890 p53 pathway SNPs (fold-enrichment 112.6; hypergeometric test *P*-value 0.0088) and the 3 (1.03%) cancer GWAS SNPs found among the 290 SNPs in coding exons (fold-enrichment: 32.22; hypergeometric test *P*-value 0.030). We identified four SNPs in AAUAAA poly(A) sites in four different pathway genes (*TP53*, protein phosphatase Mg^{2+}/Mn^{2+} dependent 1D (*PPM1D*), *CCNG2*, ribonucleotide reductase regulatory TP53 inducible subunit M2B (*RRM2B*); Supplementary information S15 (table)). Of these 4 SNPs, only the one in *TP53* (25%, rs78378222) is a cancer GWAS SNP; this represents a significant enrichment compared with both the 50 (0.29%) cancer GWAS SNPs found among the total 16,890 p53 pathway SNPs (fold-enrichment 84.45; $P = 0.011$) and the 3 (0.54%) cancer GWAS SNPs found among the 555 SNPs that occur in 3′ UTRs (fold-enrichment 46.25; $P = 0.021$). It is important to note that no such significant enrichment was found among the missense coding SNPs, whereby of the 143 missense SNPs identified among the p53 pathway SNPs, 2 are cancer GWAS SNPs (1.4%; hypergeometric test *P*-value 0.07 when compared with all pathway SNPs). Together, these results support a causal relationship between these classes of RNA processing SNP in p53 pathway genes and the noted differential cancer risk.

Discussion

Decades of research have clearly shown that genetic manipulation of p53 signalling can dramatically affect susceptibility to a broad range of cancers in mice and humans, and the topic has been well reviewed[14,15,21,28,30,84]. However, most evidence has been restricted to the characterization of rare inherited mutations found in families with LFS and common somatic mutations found in cancer genomes (FIG. 1). In this Analysis, we aimed to explore the possibility that commonly inherited genetic variants in the p53 pathway also have a significant role in susceptibility to a broad range of cancers. To do this, we used genome-wide datasets of genetic variation, CSLs derived from more than 165 GWAS conducted in a broad range of cancers (FIG. 2) and eQTLs from 14 different eQTL databases from 11 different non-cancerous tissue or cell types. Specifically, we took an integrated bioinformatics approach that linked SNPs and haplotypes from the 1000 Genomes Project to cancer GWAS SNPs and eQTLs in genes expressed in many tissues. Our results demonstrate that p53 pathway genes are more significantly enriched in CSLs compared with other signalling pathways, regardless of the pathway annotation database (FIGS 3,6). Indeed, when we restricted our analyses to SNPs that reside in known *cis*-eQTLs, only the enrichment of the p53 pathway genes remained significant after multiple hypothesis correction (FIG. 4). Moreover, we found that only the p53 pathway genes were significantly enriched in CSLs for different cancer types (FIG. 7). We go on to show that the p53-associated cancer susceptibility loci are enriched in polymorphic regulatory elements for RNA processing. One of the most striking findings of our analyses are the strong similarities between the causal, somatic mutations and the inherited, cancer-associated SNPs of the p53 pathway. We have found that both classes of genetic variant occur in a high proportion of p53 pathway genes relative to other pathways (FIGS 1,3,6), in multiple cancer types (FIG. 7) and in similar pathway genes (FIG. 8).

Our results enable insights into p53-mediated tumour suppression in humans and p53 pathway-based surveillance and treatment strategies. First, the convergence of multiple lines of evidence, both genetic and functional, strongly suggests that p53-dependent tumour suppression is highly sensitive to inherited genetic variation, whether it is rare, highly penetrant mutations (as occurs in patients with LFS)[15] or common, lower-penetrant variants (SNPs) reported in this Analysis, and that this sensitivity can contribute to the observed heterogeneity of cancer risk in the broader population. It is intriguing to speculate that the identified cancer GWAS SNPs in p53 pathway genes could aid in risk assessment for a broad range of cancers, potentially informing asymptomatic screening for early-stage cancer diagnosis, when curative interventions are possible. Indeed, such p53 pathway risk biomarkers are needed to define the heterogeneity of age-dependent and organ-dependent cancer risk seen among families with LFS, which remains a major hurdle in designing effective screening programmes to reduce the substantial morbidity and mortality associated with a genetically weakened p53 pathway[14]. Moreover, in order to maximize the prognostic value of SNPs in the broader population, the known interactions of the p53 pathway members can serve as starting points to explore possible interactions between cancer GWAS SNPs, as well as possible interactions of SNPs with the somatic mutations frequently found in the same genes. However, one caveat is that GWAS SNPs identified in large heterogeneous population studies have relatively small effect sizes, and/or may be at low population frequency, and therefore such SNPs cannot easily be tested in small studies with mixed clinical phenotypes. Thus, assessing the clinical impact of p53 pathway SNP genotypes will require careful selection of large clinical populations with well-characterized tumour mutations, treatment protocols and follow-up. Stratification of patients based on somatic mutation signature or on SNP genotype are both possible strategies.

Another insight into p53 biology and its role in tumour suppression comes from our analysis of susceptibility loci for major disease groupings other than cancer, for which sufficient GWAS data were available to interrogate (FIGS 5,6). The p53 stress response pathway has been implicated in the pathogenesis of many diseases, including neurological[52–56], cardiovascular[57,58] and infectious[59,60] diseases. Indeed in our analysis, we have found SNPs in p53 pathway genes that overlap susceptibility loci for other, non-cancerous diseases (Supplementary information S16 (table)). However, we did not find p53 pathway genes to be significantly enriched in susceptibility loci for any other major disease groupings. These included diseases of the nervous, circulatory, digestive and musculoskeletal systems. Similar observations have been made in LFS families carrying highly penetrant *TP53* mutations and in mice carrying *Trp53* mutations, whereby carriers have a dramatically high risk of developing a broad range of cancers, but not other diseases[14,85]. A clearer picture will emerge with the accrual of more data on the genetic basis of human disease susceptibility. Indeed, a challenge of the GWAS design is the necessity for large patient cohorts to compensate for the much needed multiple hypothesis testing correction. Thus, a limitation to this Analysis is that for rarer diseases and syndromes in which p53 (mis-)activity is implicated, sufficiently large studies have yet to be completed and thus cannot be thoroughly examined[86–89]. However, the data generated in GWAS thus far indicate that genetic differences in the p53 pathway primarily affect susceptibility to cancer, rather than other major diseases such as Alzheimer disease, multiple sclerosis, coronary heart disease, type 2 diabetes or schizophrenia.

The most unexpected insight into p53-dependent tumour suppression arose from one of the most striking findings of our Analysis, namely the strong similarities between the causal, somatic mutations and the inherited, cancer-associated SNPs of the p53 pathway. Specifically, we found that both classes of genetic variant occur in a high proportion of p53 pathway genes in multiple cancer types, and in similar genes. Together, these observations suggest that certain genes in p53 signalling are highly sensitive to both heritable and somatic genetic variants, resulting in differential tumour suppression in multiple tissue types. The importance of all these genes in cancer has been demonstrated in mouse models[90–106]. This group of genes encodes important known regulators and effectors of p53-dependent stress signalling[70,107–113]. Indeed, it has been conclusively demonstrated that moderate alterations in expression levels of these genes through genome engineering of mice can lead to differences in cancer risk and progression[93,114–119]. We observe the occurrence of cancer-associated SNPs in somatically mutated p53 pathway genes that are upstream regulators of p53 signalling (*TP53*, *ATM*, *MDM4* and *CDKN2A*) and key effector genes for both cell cycle arrest (*CCND1*, *CCND2* and *CCNE1*) and apoptosis (*FAS* and *CASP8*). This human genetic evidence for their central roles in regulating or affecting p53-dependent tumour suppression suggests that targeting these genes and their protein products could prove an efficient method of modulating p53 signalling in a clinical setting, compared with other pathway genes and proteins.

As agents that modulate the levels of p53 signalling are entering the clinic[28], our observations also suggest that high-frequency inherited p53 pathway variants should be considered when designing and testing patient stratification strategies. Such information could prove useful for explaining and predicting potential side effects, as well as in understanding responsiveness to conventional and targeted cancer therapies[28,30]. For example, numerous compounds that aim to restore the function of specific components of the p53 pathway, such as upstream regulators or downstream targets, are being developed or are currently in clinical trials. These therapeutics may vary in efficacy and also in the on-target or off-target side effects that they display[28] and we hypothesize that p53 pathway SNPs may modulate some of this variation. The added information about the inherent differences in p53 signalling gained by these easily accessible and measurable biomarkers could contribute to improving the efficacy of p53 pathway-based surveillance and treatment strategies, which has been lacking in the vast majority of cases[28].

1. Vogelstein, B. *et al.* Cancer genome landscapes. *Science* **339**, 1546–1558 (2013).
This paper takes advantage of the ever-increasing knowledge of cancer genome sequences to define and classify cancer driver gene mutations and begins to place them into signalling pathways.
2. Edwards, S. L., Beesley, J., French, J. D. & Dunning, A. M. Beyond GWASs: illuminating the dark road from association to function. *Am. J. Hum. Genet.* **93**, 779–797 (2013).
3. Manolio, T. A. Bringing genome-wide association findings into clinical use. *Nat. Rev. Genet.* **14**, 549–558 (2013).
This paper summarizes the impact of GWAS on the clinic and laboratory and discusses their future impact.
4. The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
5. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
6. Schodel, J. *et al.* Common genetic variants at the 11q13.3 renal cancer susceptibility locus influence binding of HIF to an enhancer of cyclin D1 expression. *Nat. Genet.* **44**, 420–425 (2012).
7. Sur, I. K. *et al.* Mice lacking a Myc enhancer that includes human SNP rs6983267 are resistant to intestinal tumors. *Science* **338**, 1360–1363 (2012).
This paper describes a polymorphic transcriptional regulatory element in *MYC* that is able to affect cancer susceptibility in a mouse model of intestinal cancer.
8. Zeron-Medina, J. *et al.* A polymorphic p53 response element in KIT ligand influences cancer risk and has undergone natural selection. *Cell* **155**, 410–422 (2013).
This paper describes a SNP in a functional p53 response element that has undergone positive selection and influences testicular cancer risk.
9. Cookson, W., Liang, L., Abecasis, G., Moffatt, M. & Lathrop, M. Mapping complex disease traits with global gene expression. *Nat. Rev. Genet.* **10**, 184–194 (2009).
10. Stranger, B. E. *et al.* Patterns of *cis* regulatory variation in diverse human populations. *PLoS Genet.* **8**, e1002639 (2012).
11. Veyrieras, J. B. *et al.* High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet.* **4**, e1000214 (2008).
12. Nica, A. C. *et al.* Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet.* **6**, e1000895 (2010).
13. Nicolae, D. L. *et al.* Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet.* **6**, e1000888 (2010).
14. McBride, K. A. *et al.* Li-Fraumeni syndrome: cancer risk assessment and clinical management. *Nat. Rev. Clin. Oncol.* **11**, 260–271 (2014).
This paper reviews the effect of *TP53* germline mutations on a heritable cancer syndrome and the clinical implications of heritable *TP53* mutations for cancer diagnosis and prevention.
15. Merino, D. & Malkin, D. p53 and hereditary cancer. *Subcell. Biochem.* **85**, 1–16 (2014).
16. Vazquez, A., Bond, E. E., Levine, A. J. & Bond, G. L. The genetics of the p53 pathway, apoptosis and cancer therapy. *Nat. Rev. Drug Discov.* **7**, 979–987 (2008).

17. Grochola, L. F., Zeron-Medina, J., Meriaux, S. & Bond, G. L. Single-nucleotide polymorphisms in the p53 signaling pathway. *Cold Spring Harb. Perspect. Biol.* **2**, a001032 (2010).
18. Whibley, C. & Pharoah, P. D. & Hollstein, M. p53 polymorphisms: cancer implications. *Nat. Rev. Cancer* **9**, 95–107 (2009).
19. Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* **502**, 333–339 (2013).
This paper analyses cancer genome sequencing data to describe the distributions of somatic mutations across tumour types, wherein the *TP53* gene is found to be the most frequently mutated gene.
20. Soussi, T., Ishioka, C., Claustres, M. & Beroud, C. Locus-specific mutation databases: pitfalls and good practice based on the p53 experience. *Nat. Rev. Cancer* **6**, 83–90 (2006).
21. Leroy, B. *et al.* The TP53 website: an integrative resource centre for the TP53 mutation database and TP53 mutant analysis. *Nucleic Acids Res.* **41**, D962–D969 (2013).
22. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).
23. The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
24. The Cancer Genome Atlas Research Network. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**, 43–49 (2013).
25. The Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519–525 (2012).
26. Kandoth, C. *et al.* Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67–73 (2013).
27. Leiserson, M. D. *et al.* Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat. Genet.* **47**, 106–114 (2015).
This paper develops and uses a novel algorithm to define cancer driver genes and integrate them into interacting networks. The authors clearly note that *TP53* belongs to the largest mutated subnetwork found in the broadest spectrum of cancers.
28. Khoo, K. H., Verma, C. S. & Lane, D. P. Drugging the p53 pathway: understanding the route to clinical efficacy. *Nat. Rev. Drug Discov.* **13**, 217–236 (2014).
This paper comprehensively describes the development of druggable targets in the p53 pathway and their clinical impact.
29. Muller, P. A. & Vousden, K. H. Mutant p53 in cancer: new functions and therapeutic opportunities. *Cancer Cell* **25**, 304–317 (2014).
30. Wade, M., Li, Y. C. & Wahl, G. M. MDM2, MDMX and p53 in oncogenesis and cancer therapy. *Nat. Rev. Cancer* **13**, 83–96 (2013).
31. Ventura, A. *et al.* Restoration of p53 function leads to tumour regression *in vivo*. *Nature* **445**, 661–665 (2007).
32. Xue, W. *et al.* Senescence and tumour clearance is triggered by p53 restoration in murine liver carcinomas. *Nature* **445**, 656–660 (2007).
References 31 and 32 demonstrate that restoration of p53 activity inhibits tumorigenesis in mice by activating apoptosis and/or senescence.
33. Riley, T., Sontag, E., Chen, P. & Levine, A. Transcriptional control of human p53-regulated genes. *Nat. Rev. Mol. Cell Biol.* **9**, 402–412 (2008).
34. Muller, P. A., Vousden, K. H. & Norman, J. C. p53 and its mutants in tumor cell migration and invasion. *J. Cell Biol.* **192**, 209–218 (2011).
35. Muller, P. A. & Vousden, K. H. p53 mutations in cancer. *Nat. Cell Biol.* **15**, 2–8 (2013).
This paper reviews the oncogenic properties and mechanisms of the gain-of-function p53 mutants.
36. Futreal, P. A. *et al.* A census of human cancer genes. *Nat. Rev. Cancer* **4**, 177–183 (2004).
37. Brown, C. D., Mangravite, L. M. & Engelhardt, B. E. Integrative modeling of eQTLs and cis-regulatory elements suggests mechanisms underlying cell type specificity of eQTLs. *PLoS Genet.* **9**, e1003649 (2013).
38. Fairfax, B. P. *et al.* Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat. Genet.* **44**, 502–510 (2012).
39. Fehrmann, R. S. *et al.* Trans-eQTLs reveal that independent genetic variants associated with a complex phenotype converge on intermediate genes, with a major role for the HLA. *PLoS Genet.* **7**, e1002197 (2011).
40. Gaffney, D. J. *et al.* Dissecting the regulatory architecture of gene expression QTLs. *Genome Biol.* **13**, R7 (2012).
41. Grundberg, E. *et al.* Mapping *cis*- and *trans*-regulatory effects across multiple tissues in twins. *Nat. Genet.* **44**, 1084–1089 (2012).
42. Hao, K. *et al.* Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet.* **8**, e1003029 (2012).
43. Innocenti, F. *et al.* Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. *PLoS Genet.* **7**, e1002078 (2011).
44. Liang, L. *et al.* A cross-platform analysis of 14,177 expression quantitative trait loci derived from lymphoblastoid cell lines. *Genome Res.* **23**, 716–726 (2013).
45. Schadt, E. E. *et al.* Mapping the genetic architecture of gene expression in human liver. *PLoS Biol.* **6**, e107 (2008).
46. Zeller, T. *et al.* Genetics and beyond—the transcriptome of human monocytes and disease susceptibility. *PLoS ONE* **5**, e10693 (2010).
47. Bryois, J. *et al.* Cis and *trans* effects of human genomic variants on gene expression. *PLoS Genet.* **10**, e1004461 (2014).
48. Fairfax, B. P. *et al.* Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* **343**, 1246949 (2014).
49. Kabakchiev, B. & Silverberg, M. S. Expression quantitative trait loci analysis identifies associations between genotype and gene expression in human intestine. *Gastroenterology* **144**, 1488–1496 (2013).
50. Koopmann, T. T. *et al.* Genome-wide identification of expression quantitative trait loci (eQTLs) in human heart. *PLoS ONE* **9**, e97380 (2014).
51. Raj, T. *et al.* Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science* **344**, 519–523 (2014).
52. Alves da Costa, C. *et al.* Presenilin-dependent γ-secretase-mediated control of p53-associated cell death in Alzheimer's disease. *J. Neurosci.* **26**, 6377–6385 (2006).
53. Fogarty, M. P. *et al.* A role for p53 in the β-amyloid-mediated regulation of the lysosomal system. *Neurobiol. Aging* **31**, 1774–1786 (2010).
54. Perier, C. *et al.* Two molecular pathways initiate mitochondria-dependent dopaminergic neurodegeneration in experimental Parkinson's disease. *Proc. Natl Acad. Sci. USA* **104**, 8161–8166 (2007).
55. Alves da Costa, C. *et al.* Transcriptional repression of p53 by parkin and impairment by mutations associated with autosomal recessive juvenile Parkinson's disease. *Nat. Cell Biol.* **11**, 1370–1375 (2009).
56. Feng, Z. *et al.* p53 tumor suppressor protein regulates the levels of huntingtin gene expression. *Oncogene* **25**, 1–7 (2006).
57. Matsumoto, S. *et al.* Circulating p53-responsive microRNAs are predictive indicators of heart failure after acute myocardial infarction. *Circ. Res.* **113**, 322–326 (2013).
58. Sano, M. *et al.* p53-induced inhibition of Hif-1 causes cardiac dysfunction during pressure overload. *Nature* **446**, 444–448 (2007).
59. Munoz-Fontela, C. *et al.* p53 serves as a host antiviral factor that enhances innate and adaptive immune responses to influenza A virus. *J. Immunol.* **187**, 6428–6436 (2011).
60. Takaoka, A. *et al.* Integration of interferon-α/β signalling to p53 responses in tumour suppression and antiviral defence. *Nature* **424**, 516–523 (2003).
61. Garcia, P. B. & Attardi, L. D. Illuminating p53 function in cancer with genetically engineered mouse models. *Semin. Cell Dev. Biol.* **27**, 74–85 (2014).
62. Guarini, A. *et al. ATM* gene alterations in chronic lymphocytic leukemia patients induce a distinct gene expression profile and predict disease progression. *Haematologica* **97**, 47–55 (2012).
63. Renwick, A. *et al. ATM* mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nat. Genet.* **38**, 873–875 (2006).
64. Kim, H. S. *et al.* Inactivating mutations of caspase-8 gene in colorectal carcinomas. *Gastroenterology* **125**, 708–715 (2003).
65. Soung, Y. H. *et al.* Caspase-8 gene is frequently inactivated by the frameshift somatic mutation 1225_1226delTG in hepatocellular carcinomas. *Oncogene* **24**, 141–147 (2005).
66. Wiestner, A. *et al.* Point mutations and genomic deletions in *CCND1* create stable truncated cyclin D1 mRNAs that are associated with increased proliferation rate and shorter survival. *Blood* **109**, 4599–4606 (2007).
67. Gao, Y. B. *et al.* Genetic landscape of esophageal squamous cell carcinoma. *Nat. Genet.* **46**, 1097–1102 (2014).
68. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061–1068 (2008).
69. Sieuwerts, A. M. *et al.* Which cyclin E prevails as prognostic marker for breast cancer? Results from a retrospective study involving 635 lymph node-negative breast cancer patients. *Clin. Cancer Res.* **12**, 3319–3328 (2006).
70. Nakayama, N. *et al.* Gene amplification CCNE1 is related to poor survival and potential therapeutic target in ovarian cancer. *Cancer* **116**, 2621–2634 (2010).
71. Gronbaek, K. *et al.* Concurrent disruption of p16INK4a and the ARF-p53 pathway predicts poor prognosis in aggressive non-Hodgkin's lymphoma. *Leukemia* **14**, 1727–1735 (2000).
72. Holzelova, E. *et al.* Autoimmune lymphoproliferative syndrome with somatic Fas mutations. *N. Engl. J. Med.* **351**, 1409–1418 (2004).
73. Dowdell, K. C. *et al.* Somatic FAS mutations are common in patients with genetically undefined autoimmune lymphoproliferative syndrome. *Blood* **115**, 5164–5169 (2010).
74. Park, W. S. *et al.* Somatic mutations in the death domain of the Fas (Apo-1/CD95) gene in gastric cancer. *J. Pathol.* **193**, 162–168 (2001).
75. Gembarska, A. *et al.* MDM4 is a key therapeutic target in cutaneous melanoma. *Nat. Med.* **18**, 1239–1247 (2012).
76. Forbes, S. A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res.* **39**, D945–D950 (2011).
77. Stacey, S. N. *et al.* A germline variant in the *TP53* polyadenylation signal confers cancer susceptibility. *Nat. Genet.* **43**, 1098–1103 (2011).
This paper describes a SNP in the *TP53* gene that affects the poly(A) signal sequence, thus affecting cancer risk.
78. Wang, Z. *et al.* Further confirmation of germline glioma risk variant rs78378222 in *TP53* and its implication in tumor tissues via integrative analysis of TCGA data. *Hum. Mutat.* **36**, 684–688 (2015).
79. Weinhold, N. *et al.* The *CCND1* c.870G>A polymorphism is a risk factor for t(11;14)(q13;q32) multiple myeloma. *Nat. Genet.* **45**, 522–525 (2013).
80. Betticher, D. C. *et al.* Alternate splicing produces a novel cyclin D1 transcript. *Oncogene* **11**, 1005–1011 (1995).
81. Comstock, C. E. *et al.* Cyclin D1 splice variants: polymorphism, risk, and isoform-specific regulation in prostate cancer. *Clin. Cancer Res.* **15**, 5338–5349 (2009).
82. Olshavsky, N. A. *et al.* Identification of ASF/SF2 as a critical, allele-specific effector of the cyclin D1b oncogene. *Cancer Res.* **70**, 3975–3984 (2010).
83. Knudsen, K. E., Diehl, J. A., Haiman, C. A. & Knudsen, E. S. Cyclin D1: polymorphism, aberrant splicing and cancer risk. *Oncogene* **25**, 1620–1628 (2006).
84. Kruse, J. P. & Gu, W. Modes of p53 regulation. *Cell* **137**, 609–622 (2009).
85. Bieging, K. T. & Attardi, L. D. Deconstructing p53 transcriptional networks in tumor suppression. *Trends Cell Biol.* **22**, 97–106 (2012).
86. Fumagalli, S. & Thomas, G. The role of p53 in ribosomopathies. *Semin. Hematol.* **48**, 97–105 (2011).
87. McGowan, K. A. & Mason, P. J. Animal models of Diamond Blackfan anemia. *Semin. Hematol.* **48**, 106–116 (2011).
88. Van Nostrand, J. L. & Attardi, L. D. Guilty as CHARGED: 53's expanding role in disease. *Cell Cycle* **13**, 3798–3807 (2014).
89. Van Nostrand, J. L. *et al.* Inappropriate p53 activation during development induces features of CHARGE syndrome. *Nature* **514**, 228–232 (2014).
90. Adachi, M. *et al.* Targeted mutation in the *Fas* gene causes hyperplasia in peripheral lymphoid organs and liver. *Nat. Genet.* **11**, 294–300 (1995).

91. Adachi, M. *et al.* Enhanced and accelerated lymphoproliferation in Fas-null mice. *Proc. Natl Acad. Sci. USA* **93**, 2131–2136 (1996).
92. Barlow, C. *et al.* Atm-deficient mice: a paradigm of ataxia telangiectasia. *Cell* **86**, 159–171 (1996).
93. Bieging, K. T., Mello, S. S. & Attardi, L. D. Unravelling mechanisms of p53-mediated tumour suppression. *Nat. Rev. Cancer* **14**, 359–370 (2014).
94. Chen, L. *et al.* CD95 promotes tumour growth. *Nature* **465**, 492–496 (2010).
95. Deane, N. G. *et al.* Hepatocellular carcinoma results from chronic cyclin D1 overexpression in transgenic mice. *Cancer Res.* **61**, 5389–5395 (2001).
96. Geng, Y. *et al.* Kinase-independent function of cyclin E. *Mol. Cell* **25**, 127–139 (2007).
97. Geng, Y. *et al.* Cyclin E ablation in the mouse. *Cell* **114**, 431–443 (2003).
98. Hakem, A. *et al.* Caspase-8 is essential for maintaining chromosomal stability and suppressing B-cell lymphomagenesis. *Blood* **119**, 3495–3502 (2012).
99. Jiang, H. *et al.* The combined status of ATM and p53 link tumor development with therapeutic response. *Genes Dev.* **23**, 1895–1909 (2009).
100. Kamijo, T. *et al.* Tumor suppression at the mouse *INK4a* locus mediated by the alternative reading frame product p19ARF. *Cell* **91**, 649–659 (1997).
101. Kwong, L. N., Weiss, K. R., Haigis, K. M. & Dove, W. F. Atm is a negative regulator of intestinal neoplasia. *Oncogene* **27**, 1013–1018 (2008).
102. Liu, S. C. *et al.* Overexpression of cyclin D2 is associated with increased *in vivo* invasiveness of human squamous carcinoma cells. *Mol. Carcinogen.* **34**, 131–139 (2002).
103. Salmena, L. *et al.* Essential role for caspase 8 in T-cell homeostasis and T-cell-mediated immunity. *Genes Dev.* **17**, 883–895 (2003).
104. Varfolomeev, E. E. *et al.* Targeted disruption of the mouse caspase 8 gene ablates cell death induction by the TNF receptors, Fas/Apo1, and DR3 and is lethal prenatally. *Immunity* **9**, 267–276 (1998).
105. Xiong, S. *et al.* Spontaneous tumorigenesis in mice overexpressing the p53-negative regulator Mdm4. *Cancer Res.* **70**, 7148–7154 (2010).
106. Post, S. M. *et al.* A high-frequency regulatory polymorphism in the p53 pathway accelerates tumor development. *Cancer Cell* **18**, 220–230 (2010). **This is the first description of a human regulatory SNP affecting tumorigenesis in a mouse model.**
107. Balint, E. E. & Vousden, K. H. Activation and activities of the p53 tumour suppressor protein. *Br. J. Cancer* **85**, 1813–1823 (2001).
108. Feng, Z. *et al.* The regulation of AMPK β1, TSC2, and PTEN expression by p53: stress, cell and tissue specificity, and the role of these gene products in modulating the IGF-1–AKT–mTOR pathways. *Cancer Res.* **67**, 3043–3053 (2007).
109. Harris, S. L. & Levine, A. J. The p53 pathway: positive and negative feedback loops. *Oncogene* **24**, 2899–2908 (2005).
110. Hofseth, L. J., Hussain, S. P. & Harris, C. C. p53: 25 years after its discovery. *Trends Pharmacol. Sci.* **25**, 177–181 (2004).
111. Levine, A. J., Hu, W. & Feng, Z. The p53 pathway: what questions remain to be explored? *Cell Death Differ.* **13**, 1027–1036 (2006).
112. Sherr, C. J. Divorcing ARF and p53: an unsettled case. *Nat. Reviews Cancer* **6**, 663–673 (2006).
113. Tokino, T. & Nakamura, Y. The role of p53-target genes in human cancer. *Crit. Rev. Oncol. Hematol.* **33**, 1–6 (2000).
114. Donehower, L. A. & Lozano, G. 20 years studying p53 functions in genetically engineered mice. *Nat. Rev. Cancer* **9**, 831–841 (2009).
115. Kamijo, T., Bodner, S., van de Kamp, E., Randle, D. H. & Sherr, C. J. Tumor spectrum in ARF-deficient mice. *Cancer Res.* **59**, 2217–2222 (1999).
116. Spring, K. *et al.* Mice heterozygous for mutation in *Atm*, the gene involved in ataxia-telangiectasia, have heightened susceptibility to cancer. *Nat. Genet.* **32**, 185–190 (2002).
117. Spring, K. *et al. Atm* knock-in mice harboring an in-frame deletion corresponding to the human ATM 7636del9 common mutation exhibit a variant phenotype. *Cancer Res.* **61**, 4561–4568 (2001).
118. Wang, Y. V., Leblanc, M., Wade, M., Jochemsen, A. G. & Wahl, G. M. Increased radioresistance and accelerated B cell lymphomas in mice with Mdmx mutations that prevent modifications by DNA-damage-activated kinases. *Cancer Cell* **16**, 33–43 (2009).
119. Yamamoto, K. *et al.* Kinase-dead ATM protein causes genomic instability and early embryonic lethality in mice. *J. Cell Biol.* **198**, 305–313 (2012).

Acknowledgements
This work was funded in part by the Ludwig Institute for Cancer Research, the Nuffield Department of Medicine, the Development Fund, Oxford Cancer Research Centre, University of Oxford, UK, and by the Intramural Research Program of the National Institute of Environmental Health Sciences, US National Institutes of Health (projects: Z01ES100475 and Z01ES46008). We thank J. S. Bader, M. Muers, M. Resnick and B. A. Merrick for their critical reading of the manuscript.

Competing interests statement
The authors declare no competing interests.

DATABASES
BioCarta: www.biocarta.com
COSMIC Cancer Gene Census: http://cancer.sanger.ac.uk/census/
GWAS catalog: http://www.ebi.ac.uk/gwas/home
Kyoto Encyclopedia of Genes and Genomes (KEGG): http://www.genome.jp/kegg/
PANTHER: www.pantherdb.org/pathway/
1000 Genomes Project: http://www.1000genomes.org

FURTHER INFORMATION
International Classification of Diseases (ICD10): http://apps.who.int/classifications/icd10/browse/2016/en

SUPPLEMENTARY INFORMATION
See online article: S1 (methods) | S2 (table) | S3 (figure) | S4 (table) | S5 (table) | S6 (table) | S7 (table) | S8 (table) | S9 (table) | S10 (table) | S11 (table) | S12 (table) | S13 (table) | S14 (table) | S15 (table) | S16 (table)

ALL LINKS ARE ACTIVE IN THE ONLINE PDF