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

Giovanni Stracquadanio[1*], Xuting Wang[2*], Marsha 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[8], Pål Sætrom[7,8], Frederica Gemignani[9], Stefano Landi[9], Benjamin Schuster-Boeckler[1], Douglas A. Bell[2§], and Gareth L. Bond[1§]

[1]Ludwig Institute for Cancer Research, University of Oxford, Nuffield Department of Clinical Medicine, Old Road Campus Research Building, Oxford OX3 7DQ, United Kingdom
[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, The Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, United Kingdom
[5]Center for Computational Molecular Biology, Brown University, 115 Waterman Street, Providence, RI 02912, USA.
[6]Department of Molecular Biology, Cell Biology, and Biochemistry, Brown University, 70 Ship Street, Providence, RI 02903, USA
[7]Department of Computer and Information Science, Norwegian, University of Science and Technology, NO-7491 Trondheim, Norway
[8]Department of Cancer Research and Molecular Medicine, Norwegian, University of Science and Technology, NO-7491 Trondheim, Norway
[9]Genetics- Department of Biology, University of Pisa, Via Derna, 1, 56126 Pisa - Italy

*These authors contributed equally to this manuscript.

§Corresponding authors:
Gareth.Bond@ndm.ox.ac.uk. The Ludwig Institute for Cancer Research, The Nuffield Department of Clinical Medicine, The University of Oxford, Oxford, The United Kingdom.
Bell1@niehs.nih.gov. Environmental Genomics Group, Genomic Integrity and Structural Biology Laboratory, National Institute of Environmental Health Sciences, MD C3-03, NIEHS, PO Box 12233, Research Triangle Park, NC 27709, The United States of America

Preface

Decades of research have proven 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 p53 pathway cancer-associated single nucleotide polymorphisms (SNPs) have strikingly similar genetic characteristics [Au; I am not convinced that characteristics is the correct word here – do you mean that the genes for susceptibility and those that are somatically mutated are similar? If not, please ensure it is clear in the main text what you mean by characteristics. Otherwise, please clarify/rephrase.] to well-studied p53 pathway cancer-causing, somatic mutations [Au; OK?]. Our results enable insights into p53-mediated tumor suppression in humans and into p53 pathway-based cancer surveillance and treatment strategies.

There are substantial differences among individuals in their risk of developing cancer, of progressing to high-grade tumors and in response to therapies. This heterogeneity is a major obstacle to designing uniformly effective prevention, screening and treatment strategies and 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 single nucleotide polymorphisms (SNPs) have been shown to associate significantly with human cancer in genome-wide association studies (GWAS) [G] conducted in a broad spectrum of solid and hematological 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 to disease, has proven 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 cancer 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 (http://www.1000genomes.org) [Au; please provide url.] reveals the genetic diversity in individuals and populations and has 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, by 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) [G] 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 eQTL[9-11].

Analysis of data sets such as these allows rapid assignment of cancer-associated SNPs to well-studied signaling pathways that are known to be important in a broad range of cancers and for which somatic genetic variants are currently utilized as biomarkers or drug targets in the laboratory and clinic. Identifying which cancer-associated pathways frequently carry SNPs associated with differential cancer susceptibility could accelerate our biological understanding of the variants' influence on cancer and the potential clinical utility of SNP biomarkers[12,13]. One of the most well-studied and important cancer signaling pathways is the p53 tumor suppressor pathway. Decades of intensive research in mice and humans have shown that human genetic variants in the p53 stress response pathway can play key roles in the incidence and survival of many cancers. For example,

among individuals with the very rare Li-Fraumeni Syndrome (LFS) [G] - who carry inherited, heterozygous mutations in [Au; meaning the mutations occur rarely in cancer? They are rarely somatically acquired mutations? Please clarify as the meaning is unclear.] *TP53* (which encodes p53) [Au; are these mutations heterozygous? If so, I think that is worth mentioning here or in the glossary.] - the penetrance for cancer onset is 100% by the age of 70 years, with the cancers occurring in numerous tissues, such as bone, connective tissues, breast and brain[14,15]. Indeed, candidate SNP studies have clearly demonstrated that p53 signaling 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 over 50% of all cancer genomes[19,20] [Au; please cite reference for the origin of this statistic.], 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 hematological 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 critical biomarkers to inform therapy [Au; biomarkers of what? Please clarify.] in some patient-stratification strategies in the clinic[14,16,28-31]. 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 tumor suppression, regulating cell migration and invasion and mediating the cellular response to DNA damage-inducing cancer therapies, primarily mediated through the ability of p53 to regulate transcription[32-36].

These observations made in model systems, tumor studies and in a rare familial syndrome (LFS) suggest that genetic variants found in the general human population, such as SNPs, also would 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 last 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 cancers, 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 tumor 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 [Au; please reduce the heading length to 39 characters (including spaces).]

Somatic, causal mutations. Cancer driver genes, like *TP53*, when mutated, can promote tumorigenisis and have been identified through studies of inherited cancer predisposition syndromes, cancer genome sequencing [Au; OK?], and experimental models of cancer. While estimates regarding the exact number of these genes vary, one well-utilized curated list is the Cancer Gene Census, which utilizes sequencing data to identify genes, whose number and pattern of mutations are highly unlikely to be attributable to chance[37]. Their current list consists of 493 RefSeq autosomal genes that harbor somatic cancer-promoting mutations (The Cancer Gene Census, The Sanger Institute, http://cancer.sanger.ac.uk/census/. [Au; is it necessary to include this given that you comment on this being the current list (meaning at the time of publication)?]). Out of the 67 autosomal genes attributed to the p53 pathway (according to the Kyoto Encyclopedia of Genes and Genomes, KEGG, http://www.genome.jp/kegg/), 15 have been denoted as harboring somatic, causal mutations in at least one cancer type. This group includes the well-studied oncogenes and tumor suppressor genes: ataxia telangiectasia mutated (*ATM*), *MDM2*, cyclin-dependent kinase inhibitor 2A (*CDKN2A*; which encodes INK4A and ARF), *FAS*, and *MDM4* (Figure 1A). Thus, 22.38% of p53 pathway genes contain known causal mutations; a significant 11.15-fold enrichment over the 2.01% found in all 24,553 annotated autosomal genes (Figure 1B, RefSeq, p-value: 3.74e-12, adjusted for multiple hypothesis testing [G}[Au; please clarify what this is adjusted for.] p-value: 8.22e-10).

In order to assess further the importance of the 11.15-fold enrichment of cancer-associated causal mutations in p53 pathway genes, we compared it to 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 signaling pathways annotated by KEGG in the categories of metabolism, genetic information processing, environmental information processing, cellular processes and organismal systems. We found that 66 of the 220 cellular signaling pathways (30%), including the p53 pathway, demonstrated significant enrichment of causally mutated genes after correction for multiple hypothesis testing (Figure 1C; see Supplementary Information S1 (methods) and Supplementary Information S2 (table) for details [Au; I think it would be helpful to order the pathways in table 1 according to the fold enrichment values so that readers can easily see where the p53 pathway sits in this list according to this factor. Also, please would you ensure that the supplementary information is labeled properly and provided in one document in the correct order: all supplementary tables/figures/text should be labeled as: Supplementary Information Sx (table/figure/methods) and they should be numbered in the order in which

they appear in the main text.]). Thus, the enrichment noted in the p53 pathway is amongst the top 5% of all well-annotated pathways of the genome [Au; please also comment on the other pathways that were included in the top of the list.]. The other significantly enriched pathways also include frequently mutated tumor suppressor genes and oncogenes, such as *PIK3CA*, *APC* and *KRAS*[19] (Supplementary Table 1). It is important to note that the enrichment [Au; was this enrichment at the top of the list as well? Please give an indication about the extent of 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 different criteria by different researchers (Supplementary Information S3 (figure)[Au; supplementary figures have not been supplied, please clarify.])[1].

Cancer-associated SNPs. Together with the noted high cancer risk among *TP53* mutation carriers[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 signaling pathways. To begin to test this, we utilized the GWAS catalogue (download date: 12/10/2015, http://www.ebi.ac.uk/gwas/home) and the 10th revision of *International Classification of Diseases (ICD10, http://apps.who.int/classifications/icd10/browse/2016/en* [Au; please provide url.]*)* 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 preformed in European populations (n=756). [Au; please briefly comment on why only European populations have been assessed.]. Each of the 756 GWASs can be attributed to one of the major ICD10 disease categories, which includes 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, Neoplasms had the most studies attributed to it with 165. These 165 studies have been undertaken to assess differential susceptibility for 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 signaling 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 (MAF, minor allele frequency, [Au; please define MAF.] ≥ 0.001) from European populations within +/- 10Kb of the gene boundaries of the 24,553 autosomal RefSeq genes and found 7,106,459 SNPs in total [Au; is it correct that this analysis thus excludes distant regulatory sites? If so, please ensure this is explicitly stated

so that the reader can understand what the limitations and caveats are.]. Subsequently, we mined the GWAS catalogue, and extracted the 1,034 SNPs (750 unique loci) associated with susceptibility to approximately 17 different types of cancers in European populations, including epithelial, mesenchymal and hematological cancers (Figure 2). Next, we augmented this dataset with SNPs in linkage disequilibrium (LD) [G] (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 from 1 to 126 SNPs), and the linked SNPs spanned an average genomic region of 6,947.132 bp (range from 1 to 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 +/- 10Kb of an annotated gene body, we determined that 2,262 of 3,454 (65.5%) cancer GWAS SNPs mapped to 541 autosomal genes, which we refer to as cancer susceptibility genes (CSGs; Figure 3A, Supplementary Information S5 (table)). Interestingly, 10 of the 67 p53 pathway genes (14.93%) are CSGs, namely: *ATM*, *CHEK2*, caspase 8 (*CASP8*), cyclin D1 (*CCND1*), *CCND2*, *CCNE1*, *CDKN2A*, *FAS*, *MDM4* and *TP53*. [Au; for balance and to increase the critical insight provided in the discussion, please comment on whether these genes only have a role in the p53 pathway or if they can have p53-independent roles as well (and are therefore also included in other pathways).] This 14.93% of p53 pathway genes represents a significant 6.77-fold enrichment compared to the 2.2% of 24,553 annotated autosomal genes that are CSGs (p-value: 2.00e-06, adjusted p-value: 4.39e-04, Figure 3B). It is important to note that all but one p53 pathway CSG (MDM4) has 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 to the potential enrichment of CSGs in all 220 well-annotated pathways in the genome. Only 3 of these 220 cellular signaling pathways (1.36%), including the p53 pathway, demonstrated significant enrichments of CSGs after correction for multiple hypothesis testing (Figure 3C; Supplementary Information S6 (table)). The 2 other significant pathways are PI3K-AKT and Adherens Junction. Like the p53 pathway, these are also known to be important pan-cancer signaling pathways (KEGG Cancer Signaling Pathways). However, the enrichment of CSGs in the p53 pathway ranks highest for both the level of significance and the fold enrichment (Figure 3C). Specifically, the PI3K-AKT and Adherens Junction pathways are associated with fold enrichments of 2.73 (p-value: 5.42e-05) and 5.11 (p-value: 1.66e-04), respectively, compared to a fold enrichment of 6.77 (p-value: 2.00e-06) for the p53 pathway. Together, the results of these analyses thus far suggest that both p53 pathway somatic, causal mutations and inherited cancer-associated SNPs are found in a higher proportion of pathway genes compared to the proportions of somatic, causal mutations and inherited cancer-associated SNPs found in other annotated signaling pathways.

Expression quantitative trait loci (eQTL). The fact that in the above analyses we required that one or more of the cancer GWAS SNPs reside within 10Kb of gene boundaries to define CSGs increases the likelihood that a SNP will be *cis*-acting and functionally affect its proximal gene expression 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 eQTL studies (*cis*-acting, *cis*-eSNPs). To do this, we began by curating data from 14 publicly available eQTL studies performed in non-cancerous tissues and cells from European descendants[38-52] [Au; please cite references. Please also ensure that these are cited in the analytical procedures.]. These studies used 5 different cell types (lymphoblastoid cell lines (LCLs), $CD4^+$ T cells, primary monocyte samples, B cells, and peripheral blood cells), as well as cells from 6 primary tissue types (adipose, skin, liver, intestine, heart and lung). The median number of samples included in these studies is 659, (range from 129 to 1,490). We selected *cis*-eQTLs within +/- 10Kb 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, 75.06% were found to be *cis*-eSNPs in a single tissue, 16.05% in two tissues, 6% in three tissues, and 2.89% in 4 or more tissues.

In total, we identified 11,887 genes genome-wide that harbored *cis*-eSNPs. In 133 of these genes (1.12%), we observed that the eSNPs also overlap (i.e. in complete LD) with cancer GWAS SNPs. Thus, these genes harbor haplotype blocks [Au; correct wording here? yes] that associate with both differential cancer susceptibility and proximal gene expression in at least one tissue or cell type analyzed (eCSGs, Figure 4A, Supplementary Information S5 (table)). Interestingly, 6 of the eCSGs are among the 51 p53 pathway genes that harbor *cis*-eSNPs (11.76%): *FAS, MDM4, ATM, CCND1, CASP8,* and *CCNE1*. This represents a significant 10.51-fold enrichment compared to the 1.12% found in all 11,887 annotated genes that harbor at least one *cis*-eSNP (p= 2.09e-05, adjusted p= 4.59e-03, Figure 4B). Importantly, of the 220 annotated cellular signaling pathways, the p53 pathway is the only one that shows a significant enrichment of eCSGs (Figure 4C, Supplementary Information S7 (table)); in fact, 61.8% of the remaining pathways have no eCSGs, and 94.1% have no more than 2. 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 signaling pathways. [Au; what about p53 pathway-associated SNPs that do not affect expression, but effect the function of the encoded protein (eg/ GOF and LOF mutations) – can these be assessed in

this manner or are these excluded from these analyses? Please clarify so the caveats and limitations are clear.]

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 which have been studied in GWAS, including neurological [53-57], cardiovascular [58,59] and infectious diseases [60,61]. 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 attributed to it (Supplementary Information S4 (table) [Au; are the references to these 591 GWASs available via ICD10? If not, I think they should be included in Supplementary table 5 to ensure that the analysis is reproducible.]). In the same manner described above for our analysis of the 165 cancer GWASs (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 +/- 10Kb of an annotated gene body (which we term Susceptibility Genes, SGs). All 18 disease categories had at least 4 SGs (median: 88.5, ranging from 4 to 708 SGs). We then explored any potential enrichment of SGs for each disease category in all 220 signaling pathways in the genome, including the p53 pathway. For 8 out of the 18 disease categories, we were able to identify an average of 4.25 signaling pathways that were significantly enriched for non-cancer SGs after correction for multiple hypothesis testing (range from 1 to 8 pathways, Figure 5, Supplementary Information S8 (table)). However, in contrast to our findings for cancer (ICD10 Category Neoplasms), the p53 signaling 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.

The p53 pathway enrichment is consistent across pathway annotations. We have demonstrated that the autosomal genes (Figure 3) of the p53 pathway overlap with cancer GWAS loci to a greater extent than the genes of 219 other well-annotated signaling pathways, and that this enrichment is limited to cancer GWASs (Figure 5). Thus far, we have exclusively utilized KEGG pathway annotations for our analyses. In order to explore further the significance of our observations, we extended our analyses to pathway annotations from 2 different well-utilized, curated pathway databases, namely BioCarta (www.biocarta.com) and PANTHER (www.pantherdb.org/pathway/). Similar to the findings utilizing 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 (Figure 6A, Supplementary Information S9 (table)) or PANTHER (Figure 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 signaling pathway annotated by either databases (Figure 6A and 6B, additional panels).

P53 pathway variants among cancers [Au; please reduce the heading length to 39 characters (including spaces).]

Causal mutations. As mentioned above, the ability of p53 to suppress tumor formation in numerous tissues has been demonstrated in numerous mouse models[30,31,62]. Indeed, some of [Au; OK? Yes Because not all 67 are?] 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/lymphoma, and other (Cancer Gene Census, Sanger). In Figure 1, we have demonstrated that p53 pathway genes are enriched in genes known to harbor causal, somatic mutations [Au; meaning somatically-acquired cancer associated driver mutations? Please clarify.] in at least one of these four tissue types, whereby the enrichment noted places the p53 pathway amongst 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 different cancer types. We find that, in all four major cancer types, p53 pathway genes were enriched in causally mutated genes (Figure 7A).

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-value: 1.20e-10, adjusted p-value: 2.64e-08 , Figure 7A). 61 of the 220 cellular signaling pathways (27.7%), including the p53 pathway, demonstrated significant enrichments of causally mutated genes after correction for multiple hypothesis testing (Figure 7A; Supplementary Information S11 (table)), thereby placing the enrichment noted in the p53 pathway amongst the top 6.82% of all pathways [Au; please give an indication of the extent of this enrichment compared with other pathways – is this in the top 10/20/30…? As before, perhaps it would be helpful to organize the pathways in Supplementary table 9 according to fold enrichment?]. Similar significant enrichments were found in leukemia/lymphomas (fold-enrichment: 14.09, p-value: 2.18e-09, adjusted p-value 4.79e-07), mesenchymal cancers (fold-enrichment: 20.36, p-value: 4.77e-06, adjusted p-value: 1.05e-03) and cancers in the Other category (fold-enrichment: 45.81, p-value: 1.70e-10, adjusted p-value: 3.75E-8). As seen in Figure 7B, there are only a total of 13 signaling pathways (5.9%), including p53, significantly enriched in causally mutated genes shared by all four major cancer types.

These pan-cancer signaling pathways include many other well-studied oncogenic and tumor suppressor pathways, such as the PI3K-AKT, RAS and MAPK signaling pathways (Supplementary Information S11 (table)).

Cancer-associated SNPs. Given the observation that the p53 pathway belongs to the 5.9% of all 220 signaling pathways that are enriched in causally mutated genes in all four major cancer types, we next wanted to explore if similar observations can be found among the cancer-associated SNPs. In Figure 3, we have demonstrated 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 signaling 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-value: 9.12e-05, adjusted p-value: 2.01e-02, Figure 7C). Interestingly, only the p53 pathway demonstrated significant enrichments of CSGs after correction for multiple hypothesis testing (Figure 7C). A similar significant enrichment of CSGs in the p53 pathway was found in leukemias/lymphomas (Figure 7C, fold-enrichment: 12.64, p-value: 4.82e-05, adjusted p-value: 1.06e-02). No significant enrichments were noted for any of the 220 pathways in the mesenchymal cancers and cancers in the Other category. However, this is likely due to the relatively fewer studies performed in these cancer type categories (Figure 2A). Importantly, and as seen in Figure 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 signaling pathways.

SNPs and mutations in similar p53 genes [Au; please reduce the heading length to 39 characters (including spaces).]

As mentioned above, of the 67 autosomal genes attributed to the p53 pathway (KEGG), 15 genes have been denoted as harboring somatic, causal mutations in at least one cancer type (Cancer Gene Census, Sanger). In our analysis thus far, we determined that 9 of these genes (60%) are CSGs (Figure 8A): *ATM*, *CASP8*, *CCND1*, *CCND2*, *CCNE1*, *CDKN2A*, *FAS*, *MDM4*, and *TP53* (Figure 8B, Cancer Gene Census, Sanger) [1,31,63-80]. This represents a significant 4.02 fold-enrichment compared to the 10 CSGs (14.93%) found in all 67 p53 pathway genes (hyper-geometric test p-value: 1.06E-6, Figure 8C). This dramatic enrichment clearly demonstrates that genes in the p53 pathway known to harbor causal somatic mutations in cancer genomes are more likely to also harbor 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 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 do 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-value: 2.12E-08, Figure 8D).

P53 pathway SNPs and RNA processing [Au; please reduce the heading length to 39 characters (including spaces). Also, I think this part of the analysis should be mentioned in the abstract and introduction so as to entice readers further.]

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 studies (average of 5 SNPs per gene, ranging from 1 to 16, see Supplementary Information (table) S13-S14 for details). In contrast to causal somatic tumor mutations, cancer-associated SNPs are single nucleotide variations that, on average, cannot have negatively affected reproductive success, as 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, relative to somatic mutations. These lower penetrant effects represent a major ongoing challenge in determining the molecular underpinnings of significant SNP associations found in GWASs, along with the fact that the responsible, causal SNP(s) are often linked with many nonfunctional SNPs[2]. However, for two 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 (UTR) of the *TP53* gene (rs78378222, A/C) resides in a canonical polyadenylation signal sequence. This poly(A)-SNP was identified in a GWAS study 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[81]. 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 tumor suppression. Two recent studies have provided data supporting that the A to C change does result in aberrant 3'end processing[81,82]. 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 study, whereby the G-allele, which creates the stronger 5'splice site (CCGgtaagt compared to CCAgtaagt), was found to associate with increased risk[83]. Alternative splicing of this exon and the potential role of this SNP in affecting the 5' splice site was first reported over 20 years ago[84]. Indeed, multiple 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[83,85,86]. The functional influence of this variant on cancer risk, relative to the cyclin D1a variant, remains to be further elucidated. However, this heavily studied SNP has been frequently found in non-GWAS studies to associate with differential risk of many cancer types, such as such as basal cell carcinoma, bladder cancer, breast cancer, colorectal cancer, esophageal cancer, gastro-intestinal cancer, head and neck cancer, hepatoblastoma, leukemia, lung cancer, ovarian cancer, prostate cancer, and renal cell carcinoma. [Au; of all cancer types? Please clarify.] [87].

Together, these data clearly demonstrate that 2 out 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 GWASs. To test this, we determined the occurrence of similar poly(A) and 5' splice site SNPs among all 16,890 SNP in and around all 67 p53 pathway genes (KEGG). We identified only 2 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 2 SNPs reside in the *CCNB2* and thrombospondin (*THBS1*) pathway genes (Supplementary Information (table) S15). Thus, of the 3 SNPs in the 5'splice sites of p53 pathway genes, 1 (33.3%) is a cancer GWAS SNP (the above-mentioned CCND1 rs9344, 870G>A). This represents a significant enrichment compared to 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 4 SNPs in AAUAAA poly(A) sites in four different pathway genes (*TP53*, protein phosphatase, Mg2+/Mn2+ dependent 1D (*PPM1D*), *CCNG2*, ribonucleotide reductase regulatory TP53 inducible subunit M2B (*RRM2B*), Supplementary Information (table) S15). Of these 4 SNPs, only the one in *TP53* (25%, rs78378222) is a cancer GWAS SNP; this represents a significant enrichment compared to both the 50 (0.29%) cancer GWAS SNPs found among the total 16,890 p53 pathway SNPs (fold-enrichment: 84.45, p-value: 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-value: 0.021). It is important to note that no such significant enrichment was found among the missense coding SNPs, whereby out of the 143 missense SNPs identified amongst the p53 pathway SNPs, 2 are cancer GWAS SNPs (1.4%, hypergeometric test p-value: 0.07 when compared to all pathway SNPs). Together, these results support a causal relationship between these classes of RNA processing SNPs in p53 pathway genes and the noted differential cancer risk.

Discussion

Decades of research have clearly shown that genetic manipulation of p53 signaling can dramatically affect susceptibility to a broad range of cancers in mice and humans, and the topic is well-reviewed [14,15,21,29,31,88]. 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 (Figure 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 cancer susceptibility to a broad range of cancers. To do this, we utilized genome-wide datasets of genetic variation, cancer susceptibility loci derived from over 165 GWAS studies conducted in a broad range of cancers (Figure 2), and expression quantitative trait loci from 14 different eQTL databases from 11 different non-cancerous tissue/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 different tissues. Our results demonstrate that p53 pathway genes are more significantly enriched in cancer susceptibility loci compared to other signaling pathways, regardless of the pathway annotation database (Figures 3 and 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 (Figure 4). Moreover, we found that only the p53 pathway genes were significantly enriched in cancer susceptibility loci for different cancer types (Figure 7). We go on to show the p53-associated cancer susceptility 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 variants occur in a high proportion of p53 pathway genes relative to other pathways (Figures 1, 3 and 6), in multiple cancer types (Figure 7), and in similar pathway genes (Figure 8).

Our results enable insights into p53-mediated tumor suppression in humans and into p53 pathway-based surveillance and treatment strategies. First, the convergence of multiple lines of evidence, both genetic and functional, strongly suggests that the p53-dependent tumour supression 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) studied 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 programs 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 among cancer GWAS SNPs, as well as possible interactions of SNPs with the somatic mutations frequently found in the identical 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 clinical impact of p53 pathway SNP genotypes will require careful selection of large clinical populations with well-characterized tumor mutations, treatment protocols and follow-up. Stratifying patients based on somatic mutation signature or on SNP genotype are both possible strategies.

Another insight into p53 biology and its role in tumor suppression comes from our Analysis of susceptibility loci for major disease groupings other than cancer, for which sufficient GWAS studies were available to interrogate (Figures 5 and 6). The p53 stress response pathway has been implicated in the pathogenesis of many different diseases, including neurological [53-57], cardiovascular [58,59] and infectious diseases [60,61]. Indeed in our Analysis, we have found SNPs in p53 pathway genes that overlap susceptibility loci for other, non-cancerous diseases (Supplementary Information (table) S16). 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 and mice carrying highly penetrant *TP53* mutations, whereby carriers have a dramatically high risk of developing a broad range of cancers, but not other diseases [14,89]. 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[90-93]. However, the data generated in GWAS thus far suggest that genetic differences in the p53 pathway primarily affect susceptibility to cancer, rather than other major diseases such as Alzheimer's disease, multiple sclerosis, coronary heart disease, type 2 diabetes and schizophrenia.

The most unexpected insight into p53-dependent tumor suppression arose from one of the most striking findings of our Analyses, 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 variants 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 signaling are highly sensitive to both heritable and somatic genetic variants resulting in differential tumor suppression in

multiple tissue types. Certainly, for all of these genes their importance in cancer has been demonstrated in mouse models[94-114]. This group of genes encodes important known regulators and effectors of p53-dependent stress signaling[71,115-125]. 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[97,126-131]. We observe the occurrence of cancer-associated SNPs in somatically-mutated p53 pathway genes that are upstream regulators of p53 signaling (*TP53*, *ATM*, *MDM4*, *CDKN2A*) and key effector genes for both cell cycle arrest (*CCND1*, CCND2, *CCNE1*) and apoptosis (*FAS*, *CASP8*). This human genetic evidence for their central roles in regulating or affecting p53-dependent tumor suppression suggests that targeting these genes and their protein products could prove an efficient method to modulate p53 signaling in a clinical setting, compared to other pathway genes and proteins.

As agents that modulate the levels of p53 signaling are entering the clinic[29], 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[29,31]. For example, numerous compounds that aim to restore 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[29] and we hypothesize that p53 pathway SNPs may modulate some of this variation. The added information about the inherent differences in p53 signaling 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 [29].

References [Au; Please select the top 5–10% of references in your list that those entering the field must read and write a single sentence in bold beneath each one to explain the importance of the work.]

1 **Vogelstein, B. *et al.* Cancer genome landscapes. *Science* 339, 1546-1558, doi:10.1126/science.1235122 (2013).**

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, doi:10.1016/j.ajhg.2013.10.012 (2013).

3 **Manolio, T. A. Bringing genome-wide association findings into clinical use. *Nat Rev Genet* 14, 549-558, doi:10.1038/nrg3523 (2013).**

4 Genomes Project, C. *et al.* A global reference for human genetic variation. *Nature* **526**, 68-74, doi:10.1038/nature15393 (2015).
5 Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74, doi:10.1038/nature11247 (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, S421-422, doi:10.1038/ng.2204 (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, doi:10.1126/science.1228606 (2012).**
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, doi:10.1016/j.cell.2013.09.017 (2013).**
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, doi:10.1038/nrg2537 (2009).
10 Stranger, B. E. *et al.* Patterns of cis regulatory variation in diverse human populations. *PLoS Genet* **8**, e1002639, doi:10.1371/journal.pgen.1002639 (2012).
11 Veyrieras, J. B. *et al.* High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet* **4**, e1000214, doi:10.1371/journal.pgen.1000214 (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, doi:10.1371/journal.pgen.1000895 (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, doi:10.1371/journal.pgen.1000888 (2010).
14 **McBride, K. A. *et al.* Li-Fraumeni syndrome: cancer risk assessment and clinical management. *Nat Rev Clin Oncol* 11, 260-271, doi:10.1038/nrclinonc.2014.41 (2014).**
15 Merino, D. & Malkin, D. p53 and hereditary cancer. *Sub-cellular biochemistry* **85**, 1-16, doi:10.1007/978-94-017-9211-0_1 (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, doi:10.1038/nrd2656 (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, doi:cshperspect.a001032 [pii]
10.1101/cshperspect.a001032 (2010).
18 Whibley, C., Pharoah, P. D. & Hollstein, M. p53 polymorphisms: cancer implications. *Nature reviews* **9**, 95-107 (2009).
19 **Kandoth, C. *et al.* Mutational landscape and significance across 12 major cancer types. *Nature* 502, 333-339, doi:10.1038/nature12634 (2013).**

20 Soussi, T., Ishioka, C., Claustres, M. & Beroud, C. Locus-specific mutation databases: pitfalls and good practice based on the p53 experience. *Nature reviews* **6**, 83-90, doi:10.1038/nrc1783 (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**, D962-969, doi:10.1093/nar/gks1033 (2013).
22 Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061-1068, doi:10.1038/nature07385 (2008).
23 Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61-70, doi:10.1038/nature11412 (2012).
24 Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* **499**, 43-49, doi:10.1038/nature12222 (2013).
25 Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519-525, doi:10.1038/nature11404 (2012).
26 Kandoth, C. *et al.* Integrated genomic characterization of endometrial carcinoma. *Nature* **497**, 67-73, doi:10.1038/nature12113 (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, doi:10.1038/ng.3168 (2015).**
28 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, doi:10.1101/cshperspect.a001032 (2010).
29 **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, doi:10.1038/nrd4236 (2014).**
30 Muller, P. A. & Vousden, K. H. Mutant p53 in cancer: new functions and therapeutic opportunities. *Cancer Cell* **25**, 304-317, doi:10.1016/j.ccr.2014.01.021 (2014).
31 Wade, M., Li, Y. C. & Wahl, G. M. MDM2, MDMX and p53 in oncogenesis and cancer therapy. *Nature reviews. Cancer* **13**, 83-96, doi:10.1038/nrc3430 (2013).
32 **Ventura, A. *et al.* Restoration of p53 function leads to tumour regression in vivo. *Nature* 445, 661-665 (2007).**
33 **Xue, W. *et al.* Senescence and tumour clearance is triggered by p53 restoration in murine liver carcinomas. *Nature* 445, 656-660 (2007).**
34 Riley, T., Sontag, E., Chen, P. & Levine, A. Transcriptional control of human p53-regulated genes. *Nat Rev Mol Cell Biol* **9**, 402-412 (2008).
35 Muller, P. A., Vousden, K. H. & Norman, J. C. p53 and its mutants in tumor cell migration and invasion. *The Journal of cell biology* **192**, 209-218, doi:10.1083/jcb.201009059 (2011).
36 **Muller, P. A. & Vousden, K. H. p53 mutations in cancer. *Nature cell biology* 15, 2-8, doi:10.1038/ncb2641 (2013).**
37 Futreal, P. A. *et al.* A census of human cancer genes. *Nature reviews. Cancer* **4**, 177-183, doi:10.1038/nrc1299 (2004).

38 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, doi:10.1371/journal.pgen.1003649 (2013).
39 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, doi:10.1038/ng.2205 (2012).
40 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, doi:10.1371/journal.pgen.1002197 (2011).
41 Gaffney, D. J. *et al.* Dissecting the regulatory architecture of gene expression QTLs. *Genome biology* **13**, R7, doi:10.1186/gb-2012-13-1-r7 (2012).
42 Grundberg, E. *et al.* Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet* **44**, 1084-1089, doi:10.1038/ng.2394 (2012).
43 Hao, K. *et al.* Lung eQTLs to help reveal the molecular underpinnings of asthma. *PLoS Genet* **8**, e1003029, doi:10.1371/journal.pgen.1003029 (2012).
44 Innocenti, F. *et al.* Identification, replication, and functional fine-mapping of expression quantitative trait loci in primary human liver tissue. *PLoS Genet* **7**, e1002078, doi:10.1371/journal.pgen.1002078 (2011).
45 Liang, L. *et al.* A cross-platform analysis of 14,177 expression quantitative trait loci derived from lymphoblastoid cell lines. *Genome research* **23**, 716-726, doi:10.1101/gr.142521.112 (2013).
46 Schadt, E. E. *et al.* Mapping the genetic architecture of gene expression in human liver. *PLoS biology* **6**, e107, doi:10.1371/journal.pbio.0060107 (2008).
47 Zeller, T. *et al.* Genetics and beyond--the transcriptome of human monocytes and disease susceptibility. *PloS one* **5**, e10693, doi:10.1371/journal.pone.0010693 (2010).
48 Bryois, J. *et al.* Cis and trans effects of human genomic variants on gene expression. *PLoS Genet* **10**, e1004461, doi:10.1371/journal.pgen.1004461 (2014).
49 Fairfax, B. P. *et al.* Innate immune activity conditions the effect of regulatory variants upon monocyte gene expression. *Science* **343**, 1246949, doi:10.1126/science.1246949 (2014).
50 Kabakchiev, B. & Silverberg, M. S. Expression quantitative trait loci analysis identifies associations between genotype and gene expression in human intestine. *Gastroenterology* **144**, 1488-1496, 1496 e1481-1483, doi:10.1053/j.gastro.2013.03.001 (2013).
51 Koopmann, T. T. *et al.* Genome-wide identification of expression quantitative trait loci (eQTLs) in human heart. *PloS one* **9**, e97380, doi:10.1371/journal.pone.0097380 (2014).
52 Raj, T. *et al.* Polarization of the effects of autoimmune and neurodegenerative risk alleles in leukocytes. *Science* **344**, 519-523, doi:10.1126/science.1249547 (2014).

53 Alves da Costa, C. *et al.* Presenilin-dependent gamma-secretase-mediated control of p53-associated cell death in Alzheimer's disease. *J Neurosci* **26**, 6377-6385, doi:10.1523/JNEUROSCI.0651-06.2006 (2006).
54 Fogarty, M. P. *et al.* A role for p53 in the beta-amyloid-mediated regulation of the lysosomal system. *Neurobiol Aging* **31**, 1774-1786, doi:10.1016/j.neurobiolaging.2008.09.018 (2010).
55 Perier, C. *et al.* Two molecular pathways initiate mitochondria-dependent dopaminergic neurodegeneration in experimental Parkinson's disease. *Proc Natl Acad Sci U S A* **104**, 8161-8166, doi:10.1073/pnas.0609874104 (2007).
56 da Costa, C. A. *et al.* Transcriptional repression of p53 by parkin and impairment by mutations associated with autosomal recessive juvenile Parkinson's disease. *Nature cell biology* **11**, 1370-1375, doi:10.1038/ncb1981 (2009).
57 Feng, Z. *et al.* p53 tumor suppressor protein regulates the levels of huntingtin gene expression. *Oncogene* **25**, 1-7 (2006).
58 Matsumoto, S. *et al.* Circulating p53-responsive microRNAs are predictive indicators of heart failure after acute myocardial infarction. *Circ Res* **113**, 322-326, doi:10.1161/CIRCRESAHA.113.301209 (2013).
59 Sano, M. *et al.* p53-induced inhibition of Hif-1 causes cardiac dysfunction during pressure overload. *Nature* **446**, 444-448, doi:10.1038/nature05602 (2007).
60 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, doi:10.4049/jimmunol.1101459 (2011).
61 Takaoka, A. *et al.* Integration of interferon-alpha/beta signalling to p53 responses in tumour suppression and antiviral defence. *Nature* **424**, 516-523, doi:10.1038/nature01850 (2003).
62 Garcia, P. B. & Attardi, L. D. Illuminating p53 function in cancer with genetically engineered mouse models. *Semin Cell Dev Biol* **27**, 74-85, doi:10.1016/j.semcdb.2013.12.014 (2014).
63 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, doi:10.3324/haematol.2011.049270 (2012).
64 Renwick, A. *et al.* ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nat Genet* **38**, 873-875, doi:10.1038/ng1837 (2006).
65 Kim, H. S. *et al.* Inactivating mutations of caspase-8 gene in colorectal carcinomas. *Gastroenterology* **125**, 708-715 (2003).
66 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, doi:10.1038/sj.onc.1208244 (2005).
67 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, doi:10.1182/blood-2006-08-039859 (2007).

68 Gao, Y. B. *et al.* Genetic landscape of esophageal squamous cell carcinoma. *Nat Genet* **46**, 1097-1102, doi:10.1038/ng.3076 (2014).
69 Cancer Genome Atlas Research, N. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**, 1061-1068, doi:10.1038/nature07385 (2008).
70 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. *Clinical cancer research : an official journal of the American Association for Cancer Research* **12**, 3319-3328, doi:10.1158/1078-0432.CCR-06-0225 (2006).
71 Nakayama, N. *et al.* Gene amplification CCNE1 is related to poor survival and potential therapeutic target in ovarian cancer. *Cancer* **116**, 2621-2634, doi:10.1002/cncr.24987 (2010).
72 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).
73 Holzelova, E. *et al.* Autoimmune lymphoproliferative syndrome with somatic Fas mutations. *The New England journal of medicine* **351**, 1409-1418, doi:10.1056/NEJMoa040036 (2004).
74 Dowdell, K. C. *et al.* Somatic FAS mutations are common in patients with genetically undefined autoimmune lymphoproliferative syndrome. *Blood* **115**, 5164-5169, doi:10.1182/blood-2010-01-263145 (2010).
75 Park, W. S. *et al.* Somatic mutations in the death domain of the Fas (Apo-1/CD95) gene in gastric cancer. *The Journal of pathology* **193**, 162-168 (2001).
76 Gembarska, A. *et al.* MDM4 is a key therapeutic target in cutaneous melanoma. *Nature medicine* **18**, 1239-1247, doi:10.1038/nm.2863 (2012).
77 Forbes, S. A. *et al.* COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res* **39**, D945-950, doi:10.1093/nar/gkq929 (2011).
78 Dotsch, V., Bernassola, F., Coutandin, D., Candi, E. & Melino, G. p63 and p73, the ancestors of p53. *Cold Spring Harbor perspectives in biology* **2**, a004887, doi:10.1101/cshperspect.a004887 (2010).
79 Schuetz, J. M. *et al.* BCL2 mutations in diffuse large B-cell lymphoma. *Leukemia : official journal of the Leukemia Society of America, Leukemia Research Fund, U.K* **26**, 1383-1390, doi:10.1038/leu.2011.378 (2012).
80 Marchini, S. *et al.* DeltaNp63 expression is associated with poor survival in ovarian cancer. *Ann Oncol* **19**, 501-507, doi:10.1093/annonc/mdm519 (2008).
81 **Stacey, S. N. *et al.* A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. *Nat Genet* 43, 1098-1103, doi:10.1038/ng.926 (2011).**
82 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, doi:10.1002/humu.22799 (2015).

83 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, doi:10.1038/ng.2583 (2013).
84 Betticher, D. C. *et al.* Alternate splicing produces a novel cyclin D1 transcript. *Oncogene* **11**, 1005-1011 (1995).
85 Comstock, C. E. *et al.* Cyclin D1 splice variants: polymorphism, risk, and isoform-specific regulation in prostate cancer. *Clinical cancer research : an official journal of the American Association for Cancer Research* **15**, 5338-5349, doi:10.1158/1078-0432.CCR-08-2865 (2009).
86 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, doi:10.1158/0008-5472.CAN-09-3468 (2010).
87 Knudsen, K. E., Diehl, J. A., Haiman, C. A. & Knudsen, E. S. Cyclin D1: polymorphism, aberrant splicing and cancer risk. *Oncogene* **25**, 1620-1628, doi:10.1038/sj.onc.1209371 (2006).
88 Kruse, J. P. & Gu, W. Modes of p53 regulation. *Cell* **137**, 609-622, doi:10.1016/j.cell.2009.04.050 (2009).
89 Bieging, K. T. & Attardi, L. D. Deconstructing p53 transcriptional networks in tumor suppression. *Trends in cell biology* **22**, 97-106, doi:10.1016/j.tcb.2011.10.006 (2012).
90 Fumagalli, S. & Thomas, G. The role of p53 in ribosomopathies. *Semin Hematol* **48**, 97-105, doi:10.1053/j.seminhematol.2011.02.004 (2011).
91 McGowan, K. A. & Mason, P. J. Animal models of Diamond Blackfan anemia. *Semin Hematol* **48**, 106-116, doi:10.1053/j.seminhematol.2011.02.001 (2011).
92 Van Nostrand, J. L. & Attardi, L. D. Guilty as CHARGED: p53's expanding role in disease. *Cell Cycle* **13**, 3798-3807, doi:10.4161/15384101.2014.987627 (2014).
93 Van Nostrand, J. L. *et al.* Inappropriate p53 activation during development induces features of CHARGE syndrome. *Nature* **514**, 228-232, doi:10.1038/nature13585 (2014).
94 Adachi, M. *et al.* Targeted mutation in the Fas gene causes hyperplasia in peripheral lymphoid organs and liver. *Nat Genet* **11**, 294-300, doi:10.1038/ng1195-294 (1995).
95 Adachi, M. *et al.* Enhanced and accelerated lymphoproliferation in Fas-null mice. *Proc Natl Acad Sci U S A* **93**, 2131-2136 (1996).
96 Barlow, C. *et al.* Atm-deficient mice: a paradigm of ataxia telangiectasia. *Cell* **86**, 159-171 (1996).
97 Bieging, K. T., Mello, S. S. & Attardi, L. D. Unravelling mechanisms of p53-mediated tumour suppression. *Nature reviews* **14**, 359-370, doi:10.1038/nrc3711 (2014).
98 Chen, L. *et al.* CD95 promotes tumour growth. *Nature* **465**, 492-496, doi:10.1038/nature09075 (2010).
99 Deane, N. G. *et al.* Hepatocellular carcinoma results from chronic cyclin D1 overexpression in transgenic mice. *Cancer Res* **61**, 5389-5395 (2001).

100 Geng, Y. *et al.* Kinase-independent function of cyclin E. *Molecular cell* **25**, 127-139, doi:10.1016/j.molcel.2006.11.029 (2007).
101 Geng, Y. *et al.* Cyclin E ablation in the mouse. *Cell* **114**, 431-443 (2003).
102 Hakem, A. *et al.* Caspase-8 is essential for maintaining chromosomal stability and suppressing B-cell lymphomagenesis. *Blood* **119**, 3495-3502, doi:10.1182/blood-2011-07-367532 (2012).
103 Jiang, H. *et al.* The combined status of ATM and p53 link tumor development with therapeutic response. *Genes & development* **23**, 1895-1909, doi:10.1101/gad.1815309 (2009).
104 Kamijo, T. *et al.* Tumor suppression at the mouse INK4a locus mediated by the alternative reading frame product p19ARF. *Cell* **91**, 649-659 (1997).
105 Kwong, L. N., Weiss, K. R., Haigis, K. M. & Dove, W. F. Atm is a negative regulator of intestinal neoplasia. *Oncogene* **27**, 1013-1018, doi:10.1038/sj.onc.1210708 (2008).
106 Liu, S. C. *et al.* Overexpression of cyclin D2 is associated with increased in vivo invasiveness of human squamous carcinoma cells. *Molecular carcinogenesis* **34**, 131-139, doi:10.1002/mc.10057 (2002).
107 Salmena, L. *et al.* Essential role for caspase 8 in T-cell homeostasis and T-cell-mediated immunity. *Genes & development* **17**, 883-895, doi:10.1101/gad.1063703 (2003).
108 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).
109 Xiong, S. *et al.* Spontaneous tumorigenesis in mice overexpressing the p53-negative regulator Mdm4. *Cancer Res* **70**, 7148-7154, doi:10.1158/0008-5472.CAN-10-1457 (2010).
110 Bahassi el, M. *et al.* Mice with the CHEK2*1100delC SNP are predisposed to cancer with a strong gender bias. *Proc Natl Acad Sci U S A* **106**, 17111-17116, doi:10.1073/pnas.0909237106 (2009).
111 Hirao, A. *et al.* Chk2 is a tumor suppressor that regulates apoptosis in both an ataxia telangiectasia mutated (ATM)-dependent and an ATM-independent manner. *Molecular and cellular biology* **22**, 6521-6532 (2002).
112 Kwak, E. L. *et al.* Mammary tumorigenesis following transgenic expression of a dominant negative CHK2 mutant. *Cancer Res* **66**, 1923-1928, doi:10.1158/0008-5472.CAN-05-1237 (2006).
113 Takai, H. *et al.* Chk2-deficient mice exhibit radioresistance and defective p53-mediated transcription. *The EMBO journal* **21**, 5195-5205 (2002).
114 **Post, S. M. *et al.* A high-frequency regulatory polymorphism in the p53 pathway accelerates tumor development. *Cancer Cell* 18, 220-230, doi:10.1016/j.ccr.2010.07.010 (2010).**
115 Balint, E. E. & Vousden, K. H. Activation and activities of the p53 tumour suppressor protein. *Br J Cancer* **85**, 1813-1823, doi:10.1054/bjoc.2001.2128 (2001).
116 Feng, Z. *et al.* The regulation of AMPK beta1, 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, doi:10.1158/0008-5472.CAN-06-4149 (2007).
117 Harris, S. L. & Levine, A. J. The p53 pathway: positive and negative feedback loops. *Oncogene* **24**, 2899-2908, doi:10.1038/sj.onc.1208615 (2005).
118 Hermeking, H. & Benzinger, A. 14-3-3 proteins in cell cycle regulation. *Semin Cancer Biol* **16**, 183-192, doi:10.1016/j.semcancer.2006.03.002 (2006).
119 Hofseth, L. J., Hussain, S. P. & Harris, C. C. p53: 25 years after its discovery. *Trends Pharmacol Sci* **25**, 177-181, doi:10.1016/j.tips.2004.02.009 (2004).
120 Levine, A. J., Hu, W. & Feng, Z. The P53 pathway: what questions remain to be explored? *Cell Death Differ* **13**, 1027-1036, doi:10.1038/sj.cdd.4401910 (2006).
121 Pietenpol, J. A. & Stewart, Z. A. Cell cycle checkpoint signaling: cell cycle arrest versus apoptosis. *Toxicology* **181-182**, 475-481 (2002).
122 Sherr, C. J. Divorcing ARF and p53: an unsettled case. *Nature reviews. Cancer* **6**, 663-673, doi:10.1038/nrc1954 (2006).
123 Taylor, W. R. & Stark, G. R. Regulation of the G2/M transition by p53. *Oncogene* **20**, 1803-1815, doi:10.1038/sj.onc.1204252 (2001).
124 Tokino, T. & Nakamura, Y. The role of p53-target genes in human cancer. *Crit Rev Oncol Hematol* **33**, 1-6 (2000).
125 Hirao, A. *et al.* DNA damage-induced activation of p53 by the checkpoint kinase Chk2. *Science* **287**, 1824-1827 (2000).
126 Donehower, L. A. & Lozano, G. 20 years studying p53 functions in genetically engineered mice. *Nature reviews* **9**, 831-841 (2009).
127 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).
128 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, doi:10.1038/ng958 (2002).
129 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).
130 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, doi:10.1016/j.ccr.2009.05.008 (2009).
131 Yamamoto, K. *et al.* Kinase-dead ATM protein causes genomic instability and early embryonic lethality in mice. *The Journal of cell biology* **198**, 305-313, doi:10.1083/jcb.201204098 (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, and by the Intramural Research Program of National Institute of Environmental Health Sciences, National Institutes of Health (projects: Z01ES100475

and Z01ES46008). We would like to thank Joel S. Bader, Mary Muers, Micheal Resnick, and B.A. Merrick for their critical reading of the manuscript.

Conflict of interest

There is **NO** Competing Interest.

Figure Legends

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 KEGG. Genes for which mutations have been causally implicated in cancer appear darker and are outlined (Cancer Gene Census, Sanger). A blue outline indicates causally mutated genes in epithelial cancers, red in leukemia/lymphomas, purple in mesenchymal cancers and orange in other types of cancers. **(B)** A bar graph of the percent of genes in the p53 pathway with known causal mutations compared to all annotated autosomal genes of the genome. 15 out 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-value: 3.74e-12 is also depicted). **(C)** A scatter plot showing the fold enrichment of causally mutated genes on the x-axis (log-scale), and the p-value on the y-axis (-log10 scale). The horizontal line represents the 5% *Family Wise Error Rate* threshold (Bonferroni adjusted-p-value: 0.05). The enrichment of causal mutations in p53 pathway genes (in yellow) is the highest and the most significant compared to the other 220 annotated KEGG pathways (in blue). Overall, 30% pathways demonstrated significant enrichment of causally mutated genes.

Figure 2. **One hundred and sixty five GWAS studies of many cancers types have been performed in European populations.** A histo-pathological classification of all the cancers present in the NHGRI GWAS catalog (download date: 09/11/2015) and a bar graph illustrating the number of tag-SNPs which have been found to significantly associate with differential susceptibility to the particular cancer. Cancers are also classified as epithelial (blue), lymphoma/leukemia (red), mesenchymal (purple) and others (orange).

Figure 3. **Cancer-associated SNPs occur in a high proportion of p53 pathway genes.** (**A**) A karyogram of the 541 genes which harbor at least 1 cancer GWAS SNP within 10Kb from their boundaries (Cancer Susceptibility Genes, CSGs). CSGs are noted in red. **(B)** A bar graph of the percent of CSGs in the p53 pathway compared to all annotated genes of the genome. 10 CSGs are found among the 67 genes of the p53 pathway (14.93%), which represents a significant 6.77-fold enrichment compared to the rest of autosomal genes in the genome (p-value: 2.00e-06 depicted in the figure, adjusted p-value: 4.39e-04). **(C)** A scatter plot showing the fold enrichment of CSGs on the x-axis (log-scale), and the adjusted p-value on the y-axis (-log10 scale). The horizontal line represents the 5% *Family Wise Error Rate* threshold (Bonferroni adjusted-p-value: 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 to the other 220 annotated KEGG pathways (in blue). Overall, 1.36% pathways demonstrated significant enrichments of CSGs.

Figure 4. **Cancer-associated eQTL occur in a high proportion of p53 pathway genes.** **(A)** A karyogram of the 133 genes which harbor an eSNP within 10Kb from their boundaries, which overlaps at least 1 cancer GWAS SNP (*cis*-eCSG). *Cis*-eCSGs are shown in red. **(B)** A bar graph of the percent of e-CSGs in the p53 pathway compared to all annotated genes of the genome. Six *cis*-eCSG are found among the 67 genes of the p53 pathway, which represents a significant 10.51-fold enrichment compare to the 11,887 genes with at least 1 eSNP (p= 2.09e-05 denoted in the graph, adjusted p= 4.59e-03). **(C)** A scatter plot showing the fold enrichment of e-CSGs on the x-axis (log-scale), and the adjusted p-value on the y-axis (-log10 scale). The horizontal line represents the 5% *Family Wise Error Rate* threshold (Bonferroni adjusted-p-value: 0.05), which is the pre-fixed significance threshold. The enrichment of *cis*-eCSG in the genes of the p53 pathway (in yellow) is the highest and the most significant compared to all the other KEGG pathways (in blue). Overall, 0.45% pathways demonstrated significant enrichment of *cis*-eCSG.

Figure 5. CSGs are significantly enriched in the p53 pathway genes, but not susceptibility genes for other major disease groupings. Scatter plots showing fold enrichment of susceptibility genes (SGs) in KEGG pathways for the 9 ICD10 disease groups that had at least 1 pathway significantly enriched in SGs out of 19 ICD10 groups. The fold enrichment of SGs is on the x-axis (log-scale), and the adjusted p-value is on the y-axis (-log10 scale). The horizontal line represents the 5% *Family Wise Error Rate* threshold (Bonferroni adjusted p-value: 0.05), which is the pre-fixed significance threshold. The enrichment of p53 pathway SGs is shown in yellow; a significant enrichment is observed in cancer (Neoplasms), but not in any other disease groups. The y-axis represents the percentage of pathways that demonstrated significant enrichments of SGs for the given disease grouping.

Figure 6. CSG enrichment in p53 pathway genes is not limited to KEGG Pathway Annotation. Scatter plots showing the fold enrichment of susceptibility genes in Biocarta **(A)** and Panther **(B)** annotated pathways for ICD10 disease groups with at least 1 significant pathway. The fold enrichment of SGs is reported on the x-axis (log-scale), and the adjusted p-value on the y-axis (-log10 scale). The horizontal line represents the 5% *Family Wise Error Rate* threshold (Bonferroni adjusted-p-value: 0.05). The enrichment of SGs in the p53 pathway is observed in cancer (Neoplasms) with both pathway annotations, but never for other diseases. The y-axis represents the percentage of pathways that demonstrated significant enrichments of SGs for the given disease grouping.

Figure 7. p53 pathway mutations and cancer-associated SNPs both 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 KEGG pathways grouped by cancer type. Fold enrichment of causally mutated genes is reported on the x-axis (log-scale), and the adjusted p-value on the y-axis (-log10 scale). The horizontal line represents the 5% *Family Wise Error Rate* threshold (Bonferroni adjusted-p-value: 0.05). The y-axis represents the percentage of pathways that demonstrated significant enrichments of SGs for the given cancer type. (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 CSGs in KEGG pathways grouped by cancer type, and a Venn Diagram (D) showing the number of pathways with a significant enrichment of CSGs grouped by cancer type. For all scatter plots the p53 pathway is in yellow.

Figure 8. p53 pathway CSGs are frequently causally mutated in cancer. **(A)** A Venn diagram depicting the overlap of p53 pathway genes that harbor causal mutations and pathway genes that are CSGs. **(B)** A pathway diagram of the CSGs 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 those pathway genes without known mutations. **(D)** A bar-plot depicting the enrichment of CSGs in genes known harbor causal somatic mutations relative to non-causally mutated genes in non-p53 pathway genes.

Supplementary information

- **Supplementary Information S1 (methods).**
  Analytical procedures.
- **Supplementary Information S2 (table).**
  Somatically mutated genes enrichment analysis for KEGG pathways.
- **Supplementary Information S3 (figure).**
  Somatic, causal mutations occur in a high proportion of p53 pathway genes.
- **Supplementary Information S4 (table).**
  Diseases classification.
- **Supplementary Information S5 (table).**
  Autosomal Cancer Susceptibility Genes.
- **Supplementary Information S6 (table).**
  Cancer Susceptibility Gene (CSGs) Enrichment in KEGG pathways.
- **Supplementary Information S7 (table).**
  eQTL-Cancer Susceptibility Gene (eCSG) Enrichment in KEGG pathways
- **Supplementary Information S8 (table).**
  Disease Susceptibility Gene (SGs) Enrichment in KEGG pathways.
- **Supplementary Information S9 (table).**
  Susceptibility Genes (SGs) Enrichment in BioCarta pathways.

- **Supplementary Information S10 (table).**
  Susceptibility Genes (SGs) Enrichment in Panther pathways.
- **Supplementary Information S11 (table).**
  Somatically mutated gene enrichment analysis for KEGG pathways across different cancer types.
- **Supplementary Information S12 (table).**
  CSGs enrichment analysis for KEGG pathways across different cancer types.
- **Supplementary Information S13 (table).**
  The 12 CSGs annotated to the p53 pathway by at least one database.
- **Supplementary Information S14 (table).**
  The 7 eCSGs annotated to the p53 pathway by at least one database.
  Supplementary Information S15 (table).
  Functional SNPs in p53 pathway genes.
- **Supplementary Information S16 (table).**
  P53 non-cancer SGs annotated to the p53 pathway by at least one database.

Glossary

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

Multiple hypotheses testing correction. 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 hypotheses testing correction is applied. The correction tries to control the probability of finding false positives when performing multiple tests.

Linkage disequilibrium. Linkage disequilibrium (LD) is the nonrandom association of alleles of two or more SNPs. LD is influenced by many factors, including mutations rates, recombination, chromosomal distance, natural selection and genetic drift. It has been extensively utilized in designing and interpreting GWAS.

eQTL. Expression quantitative trait loci (eQTL) are genetic variants in the genome, typically SNPs or copy number variants, which are associated with differential expression of a gene. Typically, 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 eQTL.

Li-Fraumeni Syndrome. A familial cancer predisposition syndrome associated with certain cancers arising in multiple tissues, such as soft tissue sarcomas, breast cancer, leukemia, 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 and 90% by the age of 60.

Author biographies

1. Giovanni Stracquadanio, is a postdoctoral fellow at the Ludwig Institute for Cancer Research at the University of Oxford. His research focuses on computational methods for omic data analysis. He obtained a Ph.D. in computer science from the University of Catania, and was previously a postdoctoral fellow at Johns Hopkins University, working on the synthetic yeast genome.
2. Xuting Wang is a bioinformatics scientist at the National Institute of Environmental Health Sciences-NIH, who aims to develop bioinformatic approaches to identify human genetic variations that affect susceptibility to environmental exposure-induced diseases with a focus on polymorphisms in transcription factor binding sites (p53 and NRF2).
3. Marsha D. Wallace obtained a Ph.D. in genetics & genomics focusing on cancer drivers at Cornell University, USA. She is currently a postdoctoral researcher at the Ludwig Institute for Cancer Research in the University of Oxford, UK. Her research focuses on functional genomics for application in the prediction and treatment of human disease.
4. Anna M. Grawenda obtained a Ph.D. in clinical medicine focusing on molecular biomarkers in human malignancies at the University of Oxford. She currently works as a postdoctoral researcher at the Department of Oncology, University Oxford, where she continues her work on identification and characterization of novel biomarkers that could improve cancer prevention, diagnosis and treatment.
5. Ping Zhang obtained his Ph.D. in biology and genetics in 2014 at Heidelberg University, Germany. He is currently a postdoctoral scientist in the Ludwig Institute for Cancer Research at the University of Oxford, UK. His research focuses on understanding the effect of polymorphic cis-regulatory elements on cell survival and cancer risk.
6. Juliet Hewitt graduated from Oxford University in 2010 with BA in Biological Sciences. She joined Dr Bond's group as a postgraduate studying for a DPhil in Clinical Medicine. Her research is focused upon how inherited genetic variants impact upon cancer risk and progression.
7. Jorge Zeron-Medina qualified in medicine from the University La Salle in Mexico City. Subsequently, he completed his PhD at the Ludwig Institute for Cancer Research in the University of Oxford as a Clarendon fellow. He is currently a Medical Oncology Fellow at Vall d´Hebron University Hospital in Barcelona.

8. Francesc Castro-Giner joined the Wellcome Trust Centre for Human Genetics, University of Oxford, in January 2013 as a postdoctoral scientist. He earned his Ph.D. in 2009 from the Universitat Pompeu Fabra, Barcelona, Spain, and trained as a postdoctoral fellow at the Centre Nacional D'Anàlisi Genòmica, Barcelona. His interests are in cancer genomics, cancer evolution, complex disease genetics and computational biology.
9. Ian Tomlinson is Professor of Molecular and Population Genetics at the University of Oxford. His interests include identification of high- and low-penetrance tumour-predisposition genes and functional cancer genetics. He has a longstanding interest in cancer evolution, including simple mathematical and descriptive models, and experimental analysis of tumour genomes.
10. Colin Goding undertook a PhD in Virology at the NIMR, London, and post-doc in Pierre Chambon's lab, Strasbourg. He then established his own group at the Marie Curie Research Institute, Oxted, UK, and in 2008 moved to Oxford where his lab is deciphering the origins of phenotypic heterogeneity in melanoma.
11. Kamil Cygan is a Ph.D. candidate in Brown University, Department of Molecular Biology, Cell Biology, and Biochemisty (MCB) and the Center for Computational Molecular Biology (CCMB). Currently, he works in the field of RNA processing, detecting single nucleotide variants that cause aberrant splicing.
12. William Fairbrother received his BA in Chemistry from Oberlin College and his PhD from Columbia University in 2000. Dr. Fairbrother was a Post-doctoral Fellow at Massachusetts Institute of Technology under mentorship of Christopher Burge and Phillip Sharp. He is currently a tenured, Hassenfeld professor in the MCB Department at Brown University. His research has focused on precision medicine, splicing and RNA genomics.
13. Laurent F. Thomas received a PhD in bioinformatics and trained as a post-doctor at the Norwegian University of Science and Technology (NTNU). His research interests are focused on gene regulation, particularly on single nucleotide polymorphisms, alternative polyadenylation, and non-coding RNAs such as microRNAs and circular RNAs.
14. Pål Sætrom obtained a PhD in bioinformatics from the Norwegian University of Science and Technology (NTNU) and is currently a dual professor at departments of Computer and Information Science, and Cancer Research and Molecular Medicine at NTNU. His research focuses on non-coding RNAs and their roles in gene regulation and disease.
15. Federica Gemignani is an Associate Professor of Genetics at the University of Pisa since 2007. Her main research interest is to perform case-control association studies between genetic variants and predisposition to risk of cancer with different strategies: candidate gene, genome-wide association studies, and high-throughput sequencing.

16. Stefano Landi is a full professor of genetics at the University of Pisa with a research focus on the effects of single nucleotide polymorphisms on cancer risk. He also evaluates the functional role candidate SNPs with *in vitro* techniques. In particular, he has contributed one of the first studies evaluating the role of SNPs in microRNAs (miRSNPs).
17. Benjamin Schuster-Boeckler studied bioinformatics at the Freie Universität Berlin before pursuing a PhD at the Wellcome Trust Sanger Institute in Cambridge. Since 2013, he leads a research group at the Ludwig Institute in Oxford. His research focuses on the causes and consequences of genetic heterogeneity in cancer.
18. Douglas A. Bell is a molecular geneticist at the National Institute of Environmental Health Sciences-NIH who for 25 years has focused on identification and functional analysis of genetic and epigenetic factors in humans that affect carcinogen metabolism genes (particularly the NRF2 pathway) and the p53 pathway.
19. Gareth L. Bond is an Associate Professor of Human Cancer Genetics at the Ludwig Institute for Cancer Research in the University of Oxford. He received his PhD from Columbia University and was a Post-doctoral Fellow with Arnold Levine. His laboratory aims to improve the understanding and management of cancer heterogeneity with inherited genetic information.

Key points summary

1. In this Analysis, we explore the possibility that commonly inherited genetic variants in the p53 pathway play a significant role in cancer susceptibility to a broad range of cancers.
2. We utilize genome-wide datasets of genetic variation, cancer susceptibility loci derived from hundreds GWAS studies conducted in a broad range of cancers, and expression quantitative trait loci from eQTL databases from many different tissue types.
3. Our results demonstrate that p53 pathway genes are more significantly enriched in cancer susceptibility loci compared to other signaling pathways.
4. We did not find p53 pathway genes to be significantly enriched in susceptibility loci for any other major disease groupings.
5. We observe strong similarities between the causal, somatic mutations and the inherited, cancer-associated SNPs of the p53 pathway, wherein both classes of genetic variants are found to occur in a high proportion of p53 pathway genes in multiple cancer types, and in similar genes.
6. Our results enable insights into p53-mediated tumor suppression in humans and into p53 pathway-based cancer surveillance and treatment strategies.

ToC blurb [Au; please feel free to edit this blurb, without extending the length (maximum ~40 words).]

Using genomic data, this Analysis demonstrates that commonly inherited single nucleotide polymorphisms (SNPs) occurring in genes of the p53 pathway affect the incidence of a broad range of cancers, more so than SNPs in other pathways. This has implications for p53-mediated tumor suppression in humans.

Subject categories

Health sciences / Risk factors

[URI /692/499]

Biological sciences / Cancer / Tumour-suppressor proteins

[URI /631/67/1244]

Biological sciences / Genetics / Genotype / Genetic predisposition to disease

[URI /631/208/727/2000]

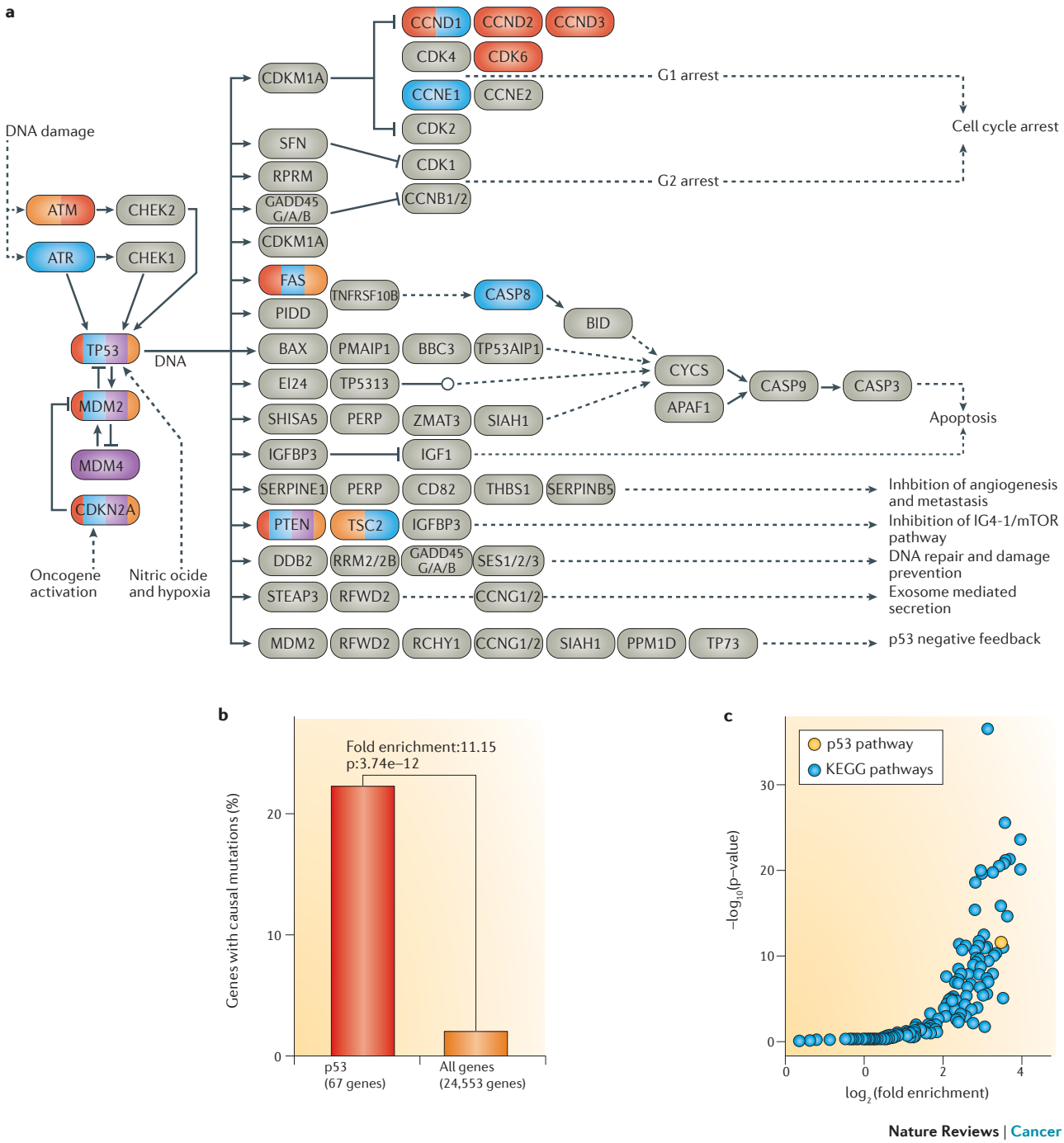
Biological sciences / Cancer / Cancer genomics

[URI /631/67/69]

Biological sciences / Cancer / Cancer genetics

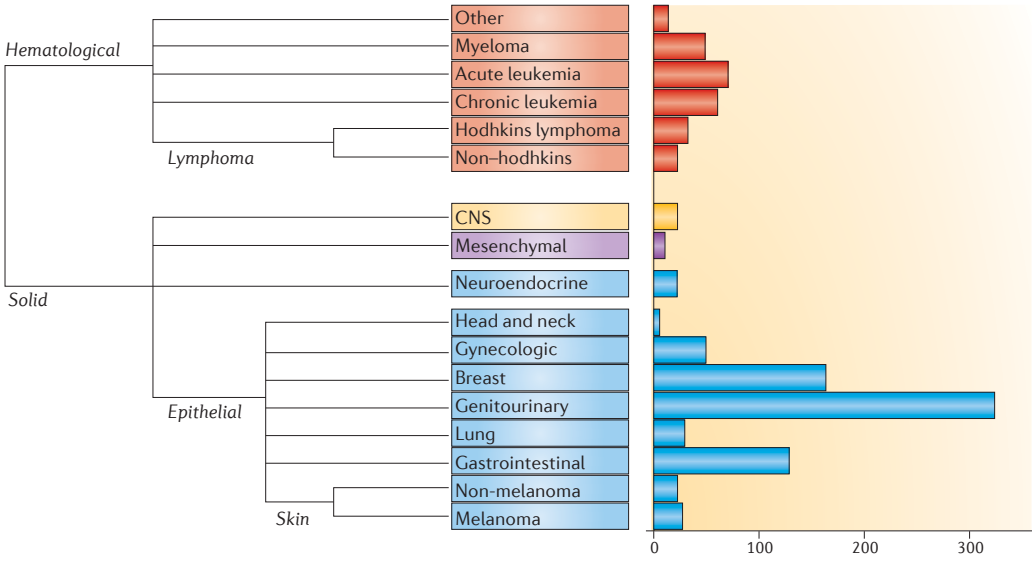
[URI /631/67/68]

Figure 1



Nature Reviews | Cancer

Figure 2



Nature Reviews | Cancer

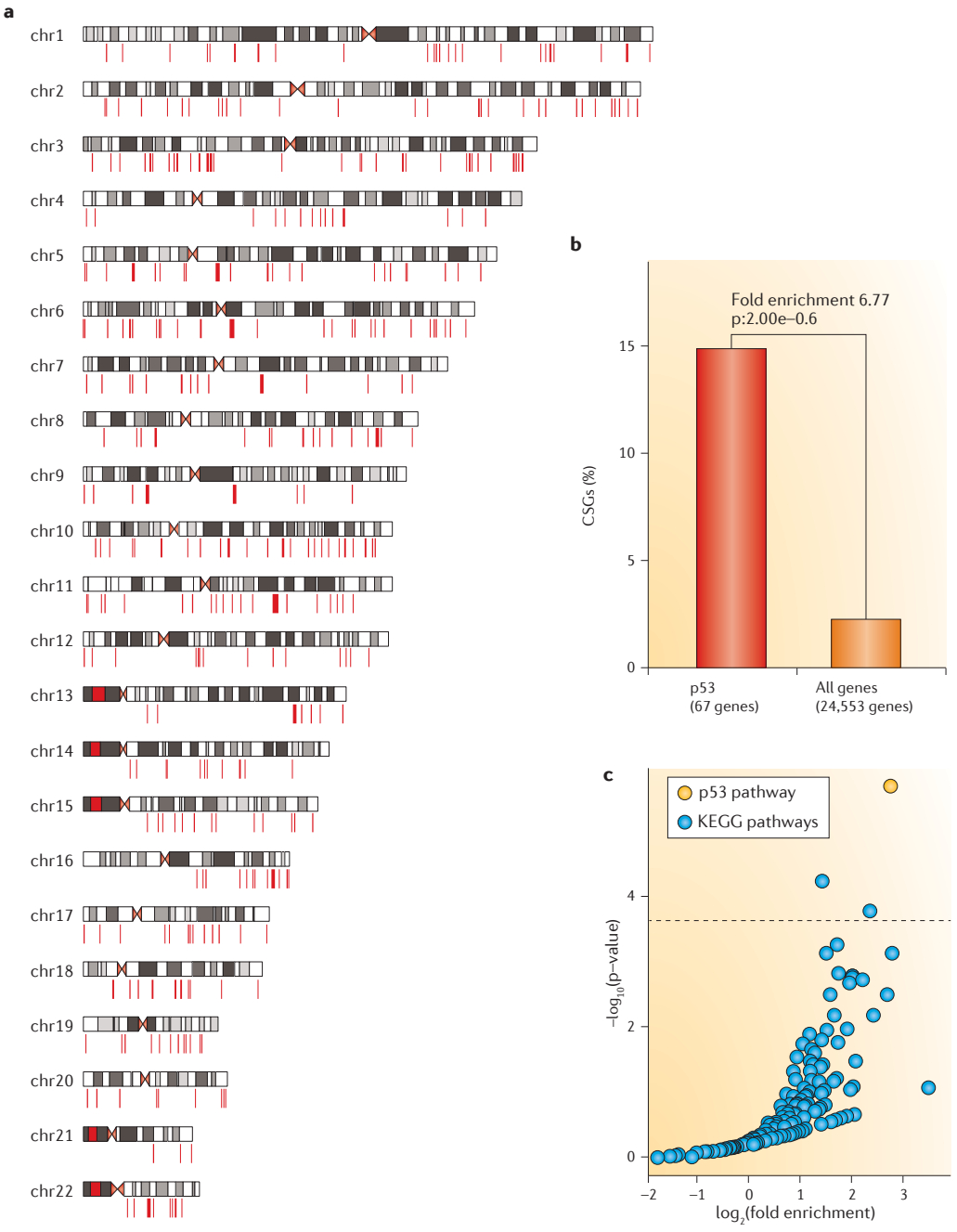

a
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
b
Fold enrichment 6.77
p:2.00e–0.6
CSGs (%)
15
10
5
0
p53
(67 genes)
All genes
(24,553 genes)
c
p53 pathway
KEGG pathways
–log10(p–value)
4
2
0
–2
–1
0
1
2
3
log2(fold enrichment)

Figure 4

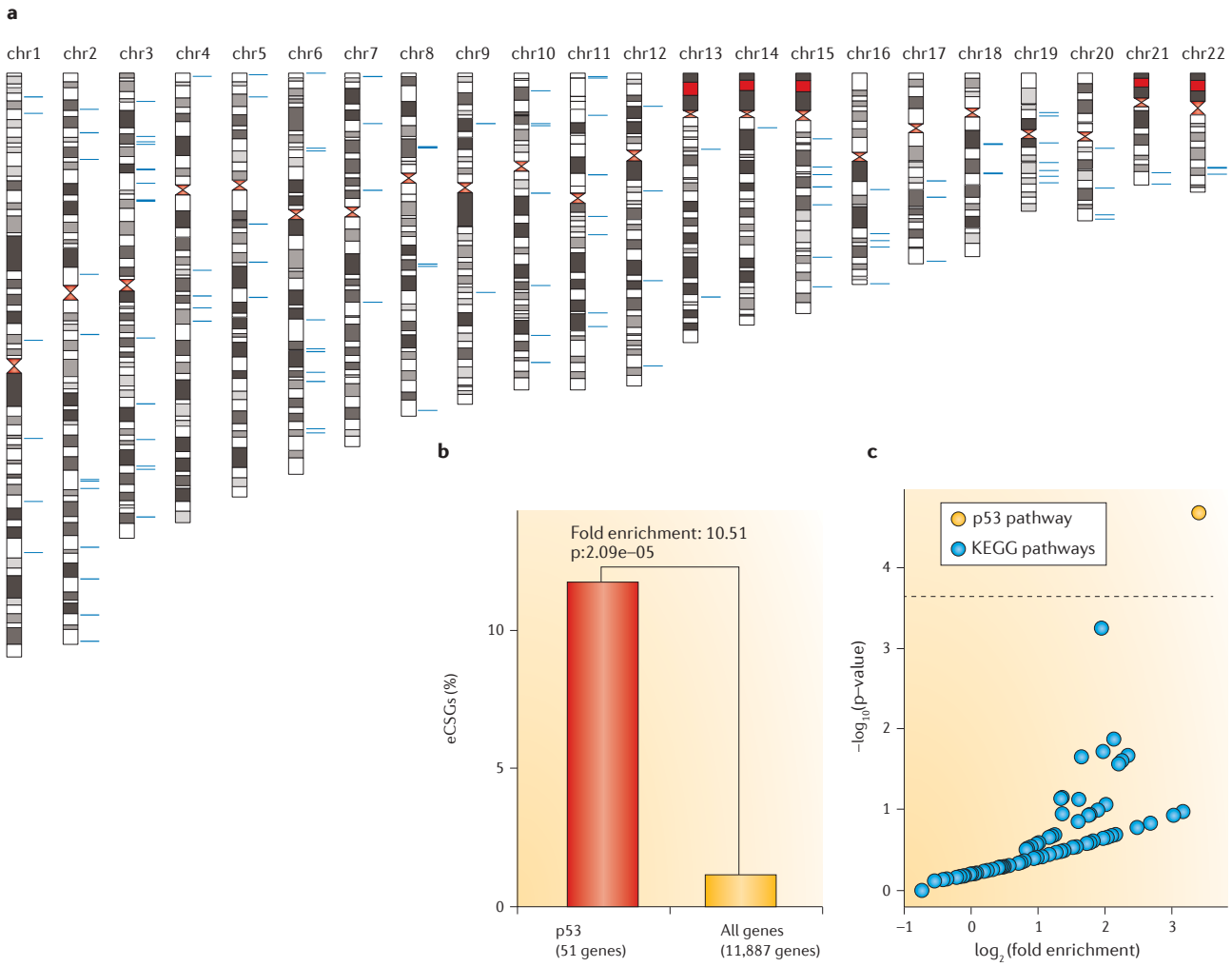
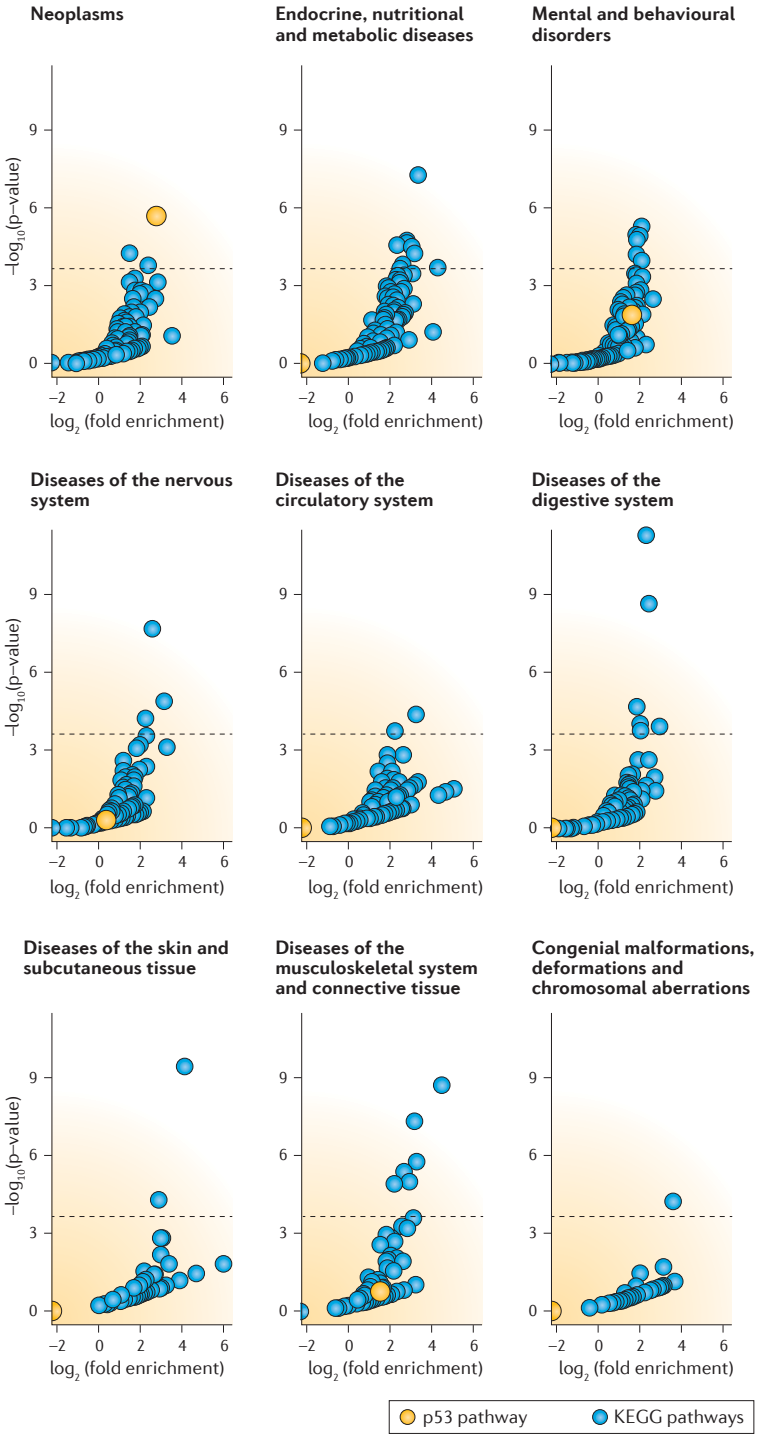

a
chr1
chr2
chr3
chr4
chr5
chr6
chr7
chr8
chr9
chr10
chr11
chr12
chr13
chr14
chr15
chr16
chr17
chr18
chr19
chr20
chr21
chr22
b
Fold enrichment: 10.51
p:2.09e–05
eCSGs (%)
10
5
0
p53
(51 genes)
All genes
(11,887 genes)
c
p53 pathway
KEGG pathways
4
3
2
1
0
−log10(p-value)
−1
0
1
2
3
log2(fold enrichment)

Figure 5



Nature Reviews | Cancer

Figure 6

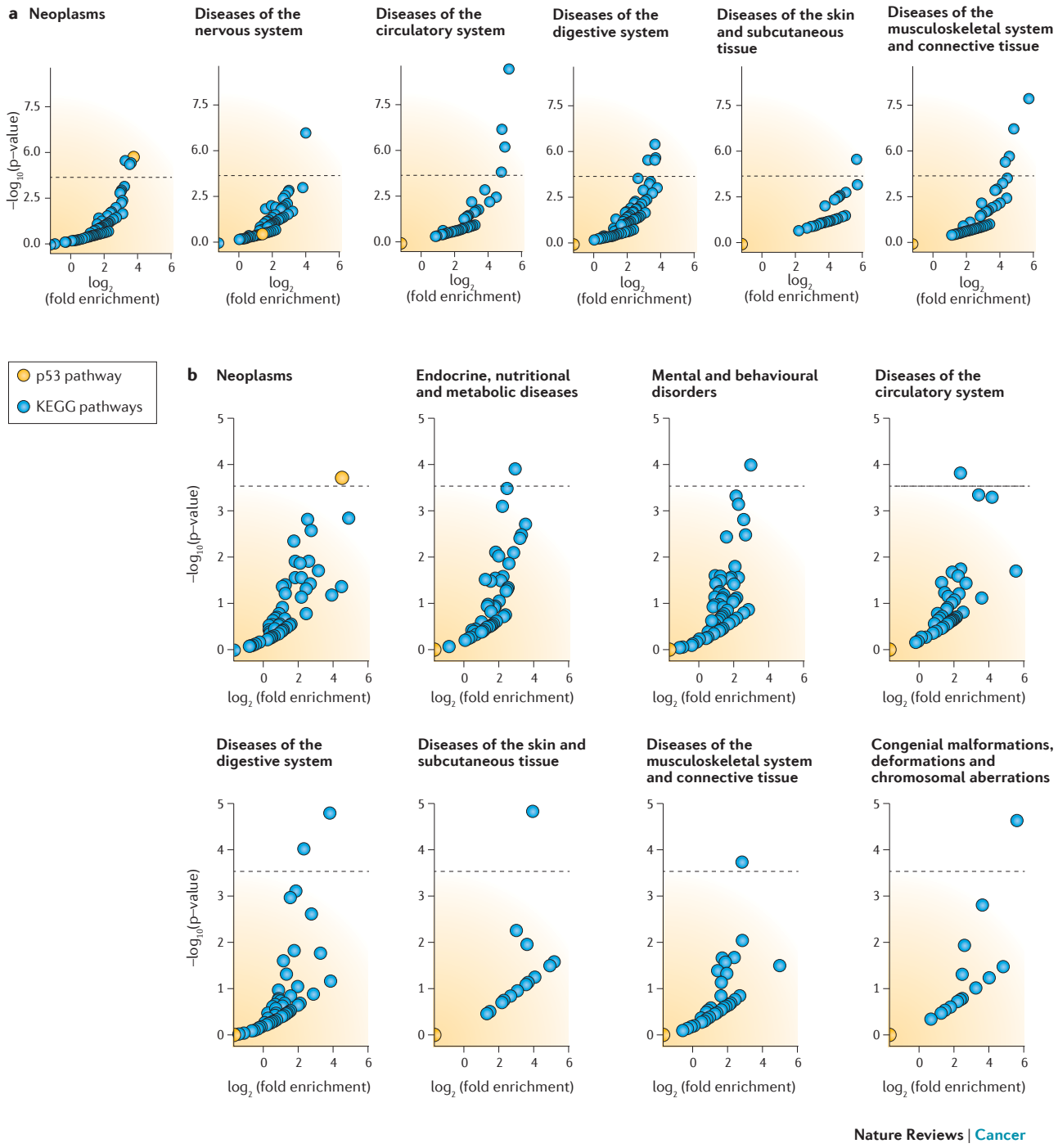

a Neoplasms
Diseases of the nervous system
Diseases of the circulatory system
Diseases of the digestive system
Diseases of the skin and subcutaneous tissue
Diseases of the musculoskeletal system and connective tissue
−log10(p–value)
log2 (fold enrichment)
p53 pathway
KEGG pathways
b Neoplasms
Endocrine, nutritional and metabolic diseases
Mental and behavioural disorders
Diseases of the circulatory system
Diseases of the digestive system
Diseases of the skin and subcutaneous tissue
Diseases of the musculoskeletal system and connective tissue
Congenial malformations, deformations and chromosomal aberrations

Figure 7

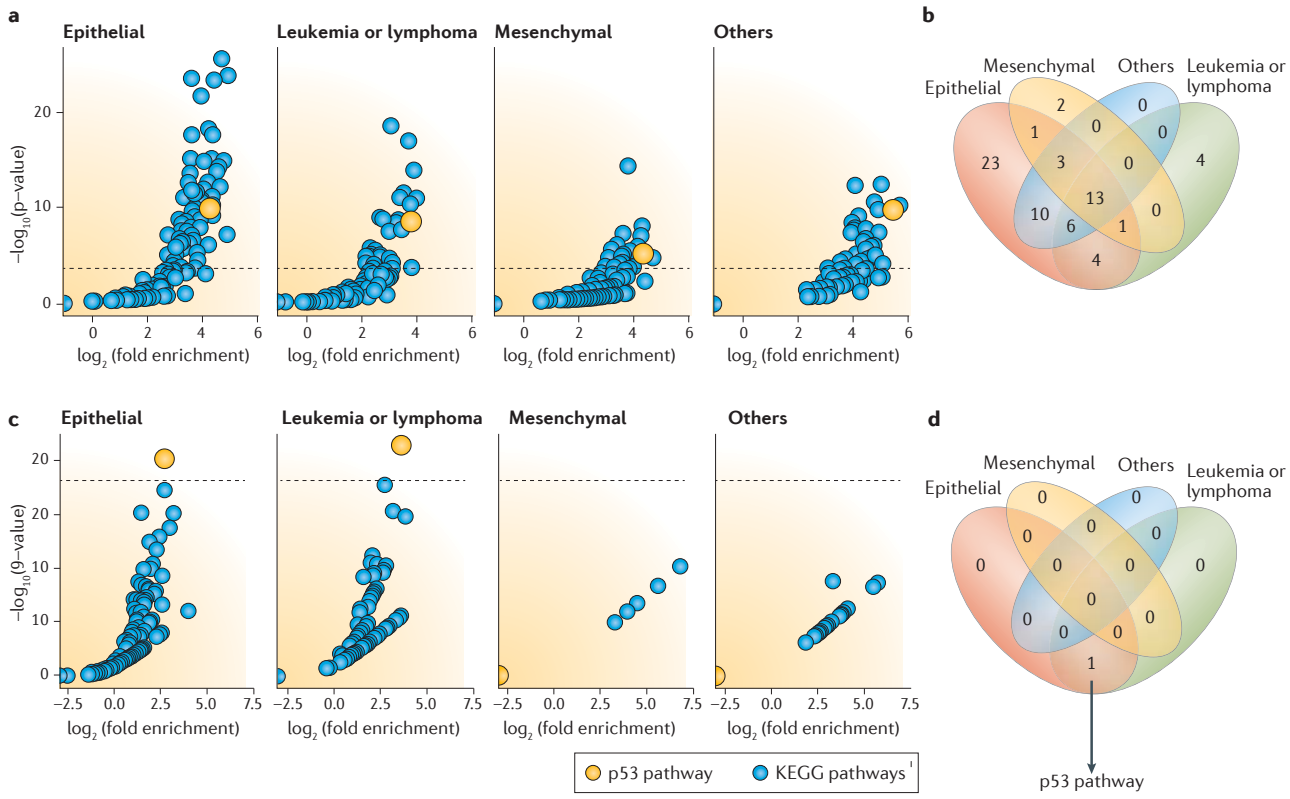
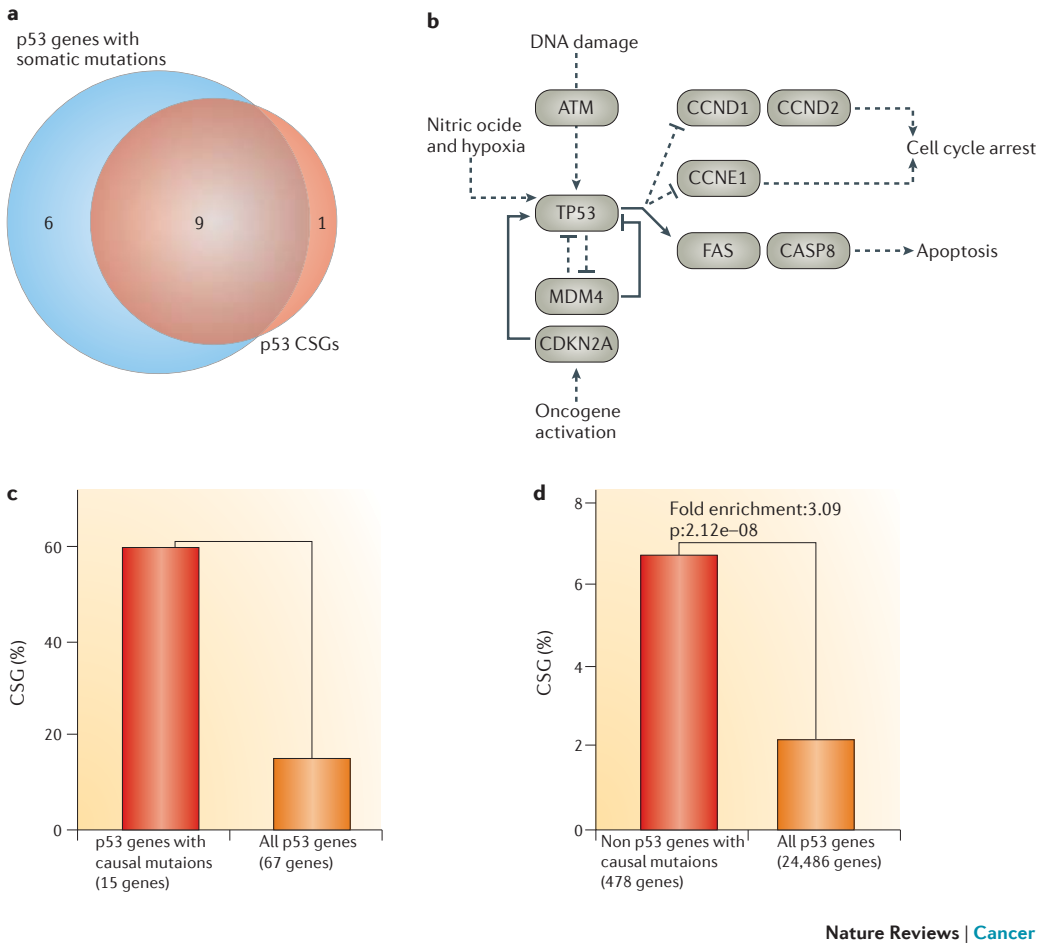

a
Epithelial
Leukemia or lymphoma
Mesenchymal
Others
−log10(p−value)
log2 (fold enrichment)
b
Mesenchymal
Others
Leukemia or lymphoma
Epithelial
2
0
1
0
0
23
3
0
4
13
10
6
1
0
4
c
Epithelial
Leukemia or lymphoma
Mesenchymal
Others
−log10(9−value)
log2 (fold enrichment)
p53 pathway
KEGG pathways
d
Mesenchymal
Others
Leukemia or lymphoma
Epithelial
p53 pathway

Figure 8


a
p53 genes with somatic mutations
6
9
1
p53 CSGs
b
DNA damage
ATM
Nitric ocide and hypoxia
TP53
MDM4
CDKN2A
Oncogene activation
CCND1
CCND2
Cell cycle arrest
CCNE1
FAS
CASP8
Apoptosis
c
CSG (%)
60
40
20
0
p53 genes with causal mutaions (15 genes)
All p53 genes (67 genes)
d
Fold enrichment:3.09
p:2.12e–08
CSG (%)
8
6
4
2
0
Non p53 genes with causal mutaions (478 genes)
All p53 genes (24,486 genes)