

RESEARCH ARTICLE

Open Access



Differential chromatin accessibility landscape of gain-of-function mutant p53 tumours

Bhavya Dhaka and Radhakrishnan Sabarinathan*

Abstract

Background: Mutations in TP53 not only affect its tumour suppressor activity but also exerts oncogenic gain-of-function activity. While the genome-wide mutant p53 binding sites have been identified in cancer cell lines, the chromatin accessibility landscape driven by mutant p53 in primary tumours is unknown. Here, we leveraged the chromatin accessibility data of primary tumours from The Cancer Genome Atlas (TCGA) to identify differentially accessible regions in mutant p53 tumours compared to wild-type p53 tumours, especially in breast and colon cancers.

Results: We identified 1587 lost and 984 gained accessible chromatin regions in breast, and 1143 lost and 640 gained regions in colon cancers. However, only less than half of those regions in both cancer types contain sequence motifs for wild-type or mutant p53 binding. Whereas, the remaining showed enrichment for master transcriptional regulators, such as FOX-Family TFs and NF- κ B in lost and SMAD and KLF TFs in gained regions of breast. In colon, ATF3 and FOS/JUN TFs were enriched in lost, and CDX family TFs and HNF4A in gained regions. By integrating the gene expression data, we identified known and novel target genes regulated by the mutant p53.

Conclusion: This study reveals the direct and indirect mechanisms by which gain-of-function mutant p53 targets the chromatin and subsequent gene expression patterns in a tumour-type specific manner. This furthers our understanding of the impact of mutant p53 in cancer development.

Keywords: TP53, Mutant p53, Chromatin accessibility, Gain-of-function, Transcription factors, Gene regulation

Background

The p53 protein, encoded by the TP53 gene, is a transcription factor that acts as a tumour suppressor. It is a multi-domain protein that consists of an N-terminal transactivation domain, a central DNA-binding domain (DBD), and an oligomerization domain that are required for the tetramer formation of p53, DNA binding and transactivation of target gene expression [1, 2]. The activity of p53 is induced by various stress signals like DNA damage, hypoxia, oxidative stress and nutrient

deficiency [3]. Under these conditions, the p53 binds to a specific DNA sequence (referred to as p53 response element) and regulates the target gene expression to prevent genomic instability, inhibit cell growth or trigger apoptosis [4]. In most cancers, the function of p53 is affected due to genetic aberrations (mutations, copy-number loss or/and epigenetic alterations) in TP53 or its regulators (such as MDM2) [5]. The mutations targeting different domains of p53 can exert distinct effects on the p53 function such as complete or partial loss of tumour suppressor activity, dominant-negative effects, and oncogenic gain-of-function (GOF) properties [6].

Missense mutations account for > 70% of all the somatic alterations observed in TP53 in human cancers,

* Correspondence: sabari@ncbs.res.in

National Centre for Biological Sciences, Tata Institute of Fundamental Research, Bengaluru 560065, India



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

most of these are clustered as hotspots in the DBD [6, 7]. DBD mutations can be classified into two main groups: *DNA-contact mutations*, which affect the amino acid residues involved in the DNA binding and *conformational mutations*, which alter the folding and structure of p53 [8]. In both cases, the mutations can directly or indirectly affect the DNA-binding ability of p53. Besides, in the early stages of tumorigenesis, some of these DBD mutations also show dominant-negative effects by inactivating the function of wild-type p53. However, in later stages, when the loss of wild-type TP53 allele has occurred, these DBD mutations exhibit oncogenic GOF activities that enhance tumorigenic potential. Eventually, the cancer cells become dependent on the mutant p53 for their growth and survival [6, 9].

Unlike the wild-type p53, the GOF mutant p53 does not bind to a specific DNA sequence motif [10]. Several mechanistic explanations for the GOF activity of p53 mutants have been proposed, including: a) Complex formation with cellular transcription factors, where mutant p53 contributes its active transactivation domain, leading to enhanced transactivation of their respective target genes. Examples of such TF partners include NF- κ B, SREBP-2, SP1/2, ETS1/2, PML and VDR; b) Interaction of mutant p53 with certain transcription factors (such as p63 and p73) leading to dysregulation of target gene expression; c) DNA structure-specific binding of mutant p53 in the promoter regions, resulting in transcriptional regulation of the relevant genes, and d) Direct recruitment of mutant p53 to the regulatory regions and/or chromatin changes through interaction with chromatin-modifying enzymes [2, 11, 12].

Although the potential targets of GOF mutant p53 have been discovered based on the analysis of gene expression [13], the transactivation mechanism is not fully understood (especially in the above-said context). Previous ChIP-seq experiments on GOF mutant p53 helped to identify the genomic locations bound by the mutant p53 directly or through its interacting partners [14, 15]. However, changes in the chromatin landscape of the target regions and nearby regulatory elements in the primary tumours remain elusive. In this study, we leveraged the chromatin accessibility data (ATAC-seq) generated in the primary tumours from The Cancer Genome Atlas (TCGA) (especially in breast and colon cancers) and identified differential chromatin accessible regions in GOF mutant p53 tumours compared to wild-type p53 tumours. Moreover, we identified transcription factor binding sites that were specifically enriched in the gained and lost peaks of chromatin accessibility in mutant p53 tumours, and their effect on the expression of the target genes. Overall, our data suggest the direct and indirect mechanisms by which GOF mutant p53 targets the

chromatin and subsequent gene expression patterns in a tumour-type specific manner.

Results

Identification of differential chromatin accessibility regions in gain-of-function mutant p53 tumours

From the TCGA cohort of 404 tumour samples (across 23 cancer types) with genome-wide chromatin accessibility data (ATAC-seq) [16], we identified 33 samples harbouring gain-of-function missense mutations (or hotspot mutations) in the DNA-binding domain of p53 (Additional file 1 Table S1). Most of which (28/33) exhibited biallelic alterations, that is, missense mutation on one allele and copy-number loss on another allele (consistent with Knudson's two-hit hypothesis [17]). On the other hand, we identified 64 samples as wild-type p53, which have neither protein-affecting mutations (non-synonymous or nonsense) nor copy-number alterations in TP53 (Additional file 1 Table S1, see Methods). To perform differential chromatin accessibility analysis between mutant and wild-type p53 tumours, we focused on two cancer subtypes, breast infiltrating ductal carcinoma and colon adenocarcinoma, which have sufficient numbers of samples in both groups. The breast infiltrating ductal carcinoma consists of six mutant samples (two R273C and one each of R273P, R273H, G245C, R175H) and five wild-type p53 samples; and colon adenocarcinoma with nine mutant samples (four R273H, two R175H, and one each of R282W, R248Q, R175C) and six wild-type p53 samples. All samples (except one) have two technical replicates each.

We performed differential chromatin accessibility analysis at the individual cancer subtype level to avoid confounding effects from tissue- and tumour-type variabilities (see Methods). With the $|\log\text{-fold change}| > 1$ and FDR (q-value) < 0.1 , we found 1587 lost and 984 gained peaks of chromatin accessibility in breast carcinoma (out of 209,835 peaks analysed), and 1143 lost and 640 gained peaks in colon adenocarcinoma (out of 119,925 peaks) (Fig. 1a, b; Additional file 1 Table S2). Based on the genomic annotations, we identified that the majority of gained and lost peaks were located in intronic and intergenic regions in both breast and colon adenocarcinoma (Fig. 1a, b), as expected. The unique number of protein-coding genes corresponds to 661 and 558, respectively, in lost and gained peaks of breast carcinoma; and 519 and 379, respectively, in lost and gained peaks of colon adenocarcinoma. Of these, 28 genes respectively in lost and gained peaks of breast; and 19 genes in lost and 24 genes in gained peaks in colon adenocarcinoma were known cancer drivers [18] (Fig. 1a, b).

Further to check if the gained and lost peaks are proximal to existing peaks (indicating the extension or reduction of already existing accessible chromatin region), we

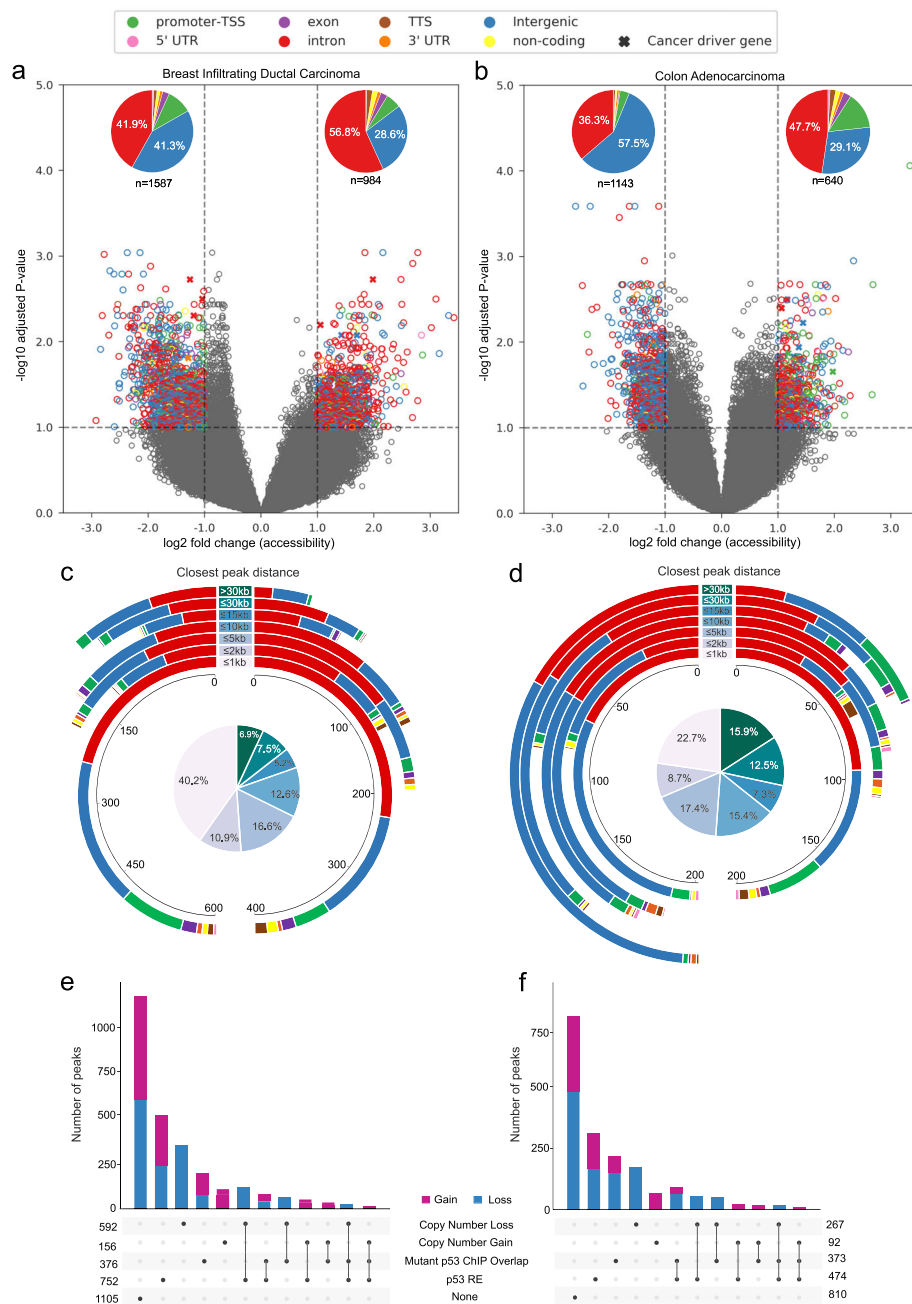


Fig. 1 Differential chromatin accessible regions in mutant p53 tumours. **a, b** Volcano plot shows the differential chromatin accessible peaks in the mutant p53 with respect to the wild-type p53 tumours in breast infiltrating ductal carcinoma (left) and colon adenocarcinoma (right). Each dot represents a peak region and the colour represents the genomic annotation (for significant candidates). The x-axis represents log2 fold change of accessible score in mutant versus wild-type p53 tumours, and the y-axis shows the $-\log_{10}$ adjusted p -value. The vertical and horizontal dashed lines indicate the threshold of FDR (q -value) < 0.1 and $|\log_2\text{FoldChange}(\text{Accessibility})|$ of 1. The symbol X indicates the peaks that overlapped cancer driver genes [18]. The pie chart on the top represents the distribution of each genomic feature in significant lost and gained peaks, where n represents the total number of significant peaks. **c, d** The donut plot shows the distribution of closest peak distance for significant lost (left) and gained (right) peaks compared to the non-significant peaks. Each circle represents a distance bin and the stacks within that represent the peaks with specific genomic annotation (as shown in panels a and b). The pie chart inside represents the overall proportion of peaks that fall under each distance group (highlighted with the same colour code mentioned at the center of the donut plot). **e, f** Upset R plot showing distribution and overlap within different sets of annotations of the significant lost and gained peaks

computed the distance of each of the gained/lost peaks against other peaks (which did not show the significant difference) in that respective cancer type. This showed that the majority of the gained and lost peaks (altogether 40%) in breast carcinoma were indeed proximal (within 1 kb distance) to the existing peaks (Fig. 1c). A similar trend was observed in colon adenocarcinoma for the gained peaks. However, for the lost peaks, besides the proximal peaks, a relatively higher number of peaks were located in distal regions (> 2 kb distance) as compared to breast adenocarcinoma (Fig. 1d). To check if this could be due to copy number alterations (CNA) in mutant or wild type p53 samples, we annotated each of the peaks having CNA gain or loss based on the CNA segmentation data from TCGA (see Methods). Overall, the lost peaks have higher overlap (23% in colon and 37% in breast) with the CNA alterations when compared to the gained peaks (15% in colon and 16% in breast), but these were not particularly enriched in the distal regions (> 2 kb distance bins, see Additional file 1 Table S3).

We then sought to identify if the gained and lost peaks overlap with the known binding sites of mutant p53 (inferred from previous ChIP-Seq analysis in cancer cell lines) [14, 15] or predicted to have wild-type p53 response elements (RE) [19], both canonical and non-canonical (see Methods) (Fig. 1e, f). Approximately one-fourth of the gained and lost peaks (29% in breast and 26% colon carcinomas) were predicted to have wild-type p53 RE. However, as compared to this, a relatively small proportion (15% in breast and 20% in colon carcinomas) overlapped with mutant p53 binding sites (inferred from previous ChIP-seq analysis). To check this further, we compared the chromatin accessibility score for all mutant p53 binding sites in wild-type and mutant p53 tumours (Additional file 2 Figure S1). This showed that overall there was not much difference between the wild-type and mutant p53 tumours (regardless of the genomic location, Additional file 2 Figure S2), with few exceptions detected from our differential analysis (Additional file 2 Figure S1). Taken together, these results suggest that the majority of the mutant p53 binding sites, inferred from the ChIP-seq analysis, were located in the existing open chromatin regions and that these regions did not exhibit any local chromatin structural changes when compared to wild-type p53 tumours.

In summary, we identified genomic regions that gained or lost chromatin accessibility in mutant p53 tumours compared to the wild-type p53 samples. However, only less than 50% of them have wild-type p53 response elements and/or overlap with mutant p53 binding sites.

Enrichment of transcription factors binding motifs in the differentially accessible regions

Mutant p53 is known to interact with other TFs (such as SP1/2, ETS1/2, E2F1, ETV1, NF-Y, NF-kB, SREBP, and SMAD1/2) and histone modifiers (MLL1/2) [15] that results in altered expression of target genes. To identify which TFs are bound to the gained and lost peaks, we performed enrichment analysis of known TF binding motifs, as well as de novo motif discovery (see Methods). The TF motif enrichment analysis (based on motif predictions from JASPAR [20]) exhibited that around 50–60% of gained or lost peaks overlapped with ZNF263, SP1, and SP2, and other known co-factors of p53 (as mentioned above) in both breast and colon cancers (Fig. 2a-d, Additional file 1 Table S4), but their enrichment was not significant ($q > 0.1$) as compared to non-significant peaks. This can be explained by the fact that some of these TFs are generally enriched in the promoter or enhancer regions. However, the TFs such as KLF5 (52%), KLF1 (39%), and SMAD2::SMAD3::SMAD4 (31%) showed significant enrichment ($q < 0.1$) in gained peaks and FOX family of TFs (FOXA1 28%, FOXA2 28% and FOXF2 17%) in lost peaks in breast carcinomas. In contrast, colon adenocarcinoma showed a significant enrichment of TFs such as CDX2 (26.5%), CDX1 (24%), HOXA13 (22%), HNF4G (37.2%), and TCF4 (36%) in gained peaks; and FOS::JUNB (45.4%), FOS::JUND (47.5%) and JUNB (46.3%) in lost peaks.

In addition, we performed de novo motif discovery analysis (using HOMER [21]) to check for the enrichment of sequence motifs in differentially accessible peaks, especially in the two distinct groups: (A) peaks with predicted p53 RE and/or overlap with mutant p53 ChIP-seq signal (discussed in the above section), and (B) the rest (see Methods, Additional file 2 Figure S3). The top k-mer motifs identified in each group, along with their similarity (alignment) score with known TF motifs, is shown in Fig. 2e-f (and the full list of motifs identified is shown in Additional file 3). In group A of breast carcinomas, we found a 10-mer motif in gained peaks (enriched in 37% of peaks) and 8-mer motif in lost peaks (62%) showing similarity with p53 motif and also partially resembling the p53 response element 5'-RRRC WWGYYY-3' (where R = purine, Y = pyrimidine and W is either A or T), as expected. Whereas in group B, we found an 8-mer motif in gained peaks (44%) showing high similarity with NF1-half site. On the other hand, we found one 10-mer motif in lost peaks (42%) showing high similarity with FOX family and another 12-mer motif showing high similarity with NF-kB and RELA (enriched in 7% of the peaks). In colon adenocarcinomas gained peaks, we found an 8-mer motif showing high similarity with CDX family TFs in both group A (36%) and B (38%). In group B, an additional 12-mer motif

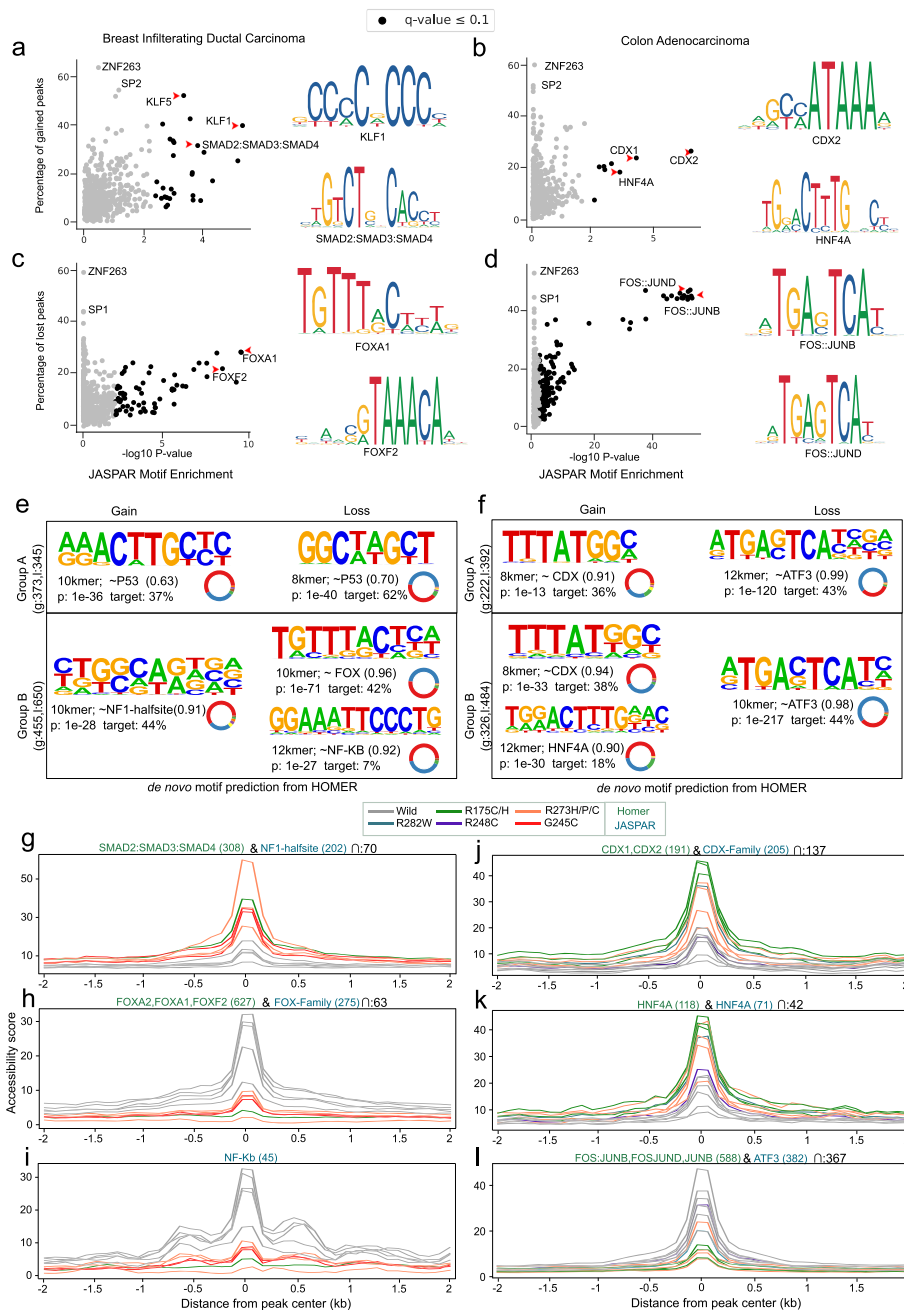


Fig. 2 Enrichment of transcription factor motifs in differential accessible regions. **a-d** Scatter plot shows the enrichment of predicted Transcription Factor (TF) motifs (from JASPAR) in gained and lost peaks of breast and colon cancers. Each dot represents a TF, the y-axis represents the proportion of peaks overlapped with the predicted motifs of that particular TF, and the x-axis represents the significance of enrichment compared to random peaks computed using one-sided Fisher's exact test. The TFs that were significantly enriched (q-value ≤ 0.1) are highlighted in black. Sequence logo of a few significant TF motifs is shown beside the scatter plots. **e, f** The panels show the logo of de novo sequence motifs identified by HOMER in the group A and B peaks (see methods) and their similarity with the known TF motifs (the value close to 1 indicates high similarity). The donut plot besides it shows the distribution of genomic features in peaks that contained the particular motif. The colour code corresponds to the genomic features shown in Fig. 1a-b. **g-l** The plot shows the accessibility score at the centre of peaks (containing specific TF motif) compared to its immediate flanking regions in mutant and wild-type p53 samples. The colour of the line indicates different mutant samples and the text colour of the motif name represents whether it was found from JASPAR or HOMER de novo motif analysis

showing similarity with HNF4A was found in 18% of gained peaks. Whereas in lost peaks, we found a 12-mer and 10-mer motif, respectively in group A (44%) and B (43%), showing high similarity with ATF3. Overall, the results obtained from this analysis is complementary to the above TF motif enrichment analysis and the motifs share the same genomic location (Additional file 2 Figure S4).

Interestingly, some of the TFs that showed high similarity with de novo motifs are the known master regulators involved in chromatin remodelling and gene regulation. For example, the FOX family of TFs, enriched in the lost peaks of breast carcinomas, is a pioneering factor which can bind to closed chromatin structure to facilitate DNA accessibility [22]. To check the extent of loss of chromatin accessibility in the neighbouring regions, we compared the accessibility score at the lost peaks (with FOX family TFs motifs) with the immediate flanking regions (± 2 kb) (Fig. 2h). In wild-type p53 samples, we observed an increase in accessibility at the centre of the peaks, and also in the surrounding regions (± 500 bp). However, in the mutant p53 samples, the accessibility score was lower at both the peak centre and flanking regions compared to the wild-type p53. This suggests that these regions are less accessible (or repressed) due to the loss of FOX family TF binding in mutant p53. Similarly, in the lost peaks enriched with NF- κ B and RELA motifs, we observed a decreased accessibility at the centre of the peaks and in the flanking regions in mutant p53 compared to the wild-type (Fig. 2i). In the case of colon adenocarcinoma, the CDX family TFs and HNF4A, enriched in the gained peaks, have previously been shown to bind enhancer regions and maintain chromatin accessibility [23]. Consistent with this, our accessibility profile also showed a sharp increase at the peak centre and in the surrounding regions (± 500 bp) (Fig. 2j, k). On the other hand, the TFs ATF3, FOS, JUNB and JUNC enriched in the lost peaks belong to members of the AP-1 complex. Depending on the complex formation between FOS, JUNB/C and ATF3, this can act as transcriptional activator or repressor [24]. In our case, we observed a strong decrease in the accessibility for mutant p53 (Fig. 2l), suggesting the repressive activity mediated by the loss of AP-1 complex at these sites.

Taken together, these results suggest that besides the enrichment of p53 response elements and mutant p53 binding sites, the master TFs (FOX, CDX and AP1) that could potentially alter the chromatin accessibility and gene regulation were enriched in the gained and lost peaks in a tissue- and tumour-specific manner.

Enrichment of enhancer marks and G-quadruplex structure in the differentially accessible regions

To check if the gained and lost peaks overlap with the regulatory regions (such as enhancer and histone marks), we overlapped these peaks with the enhancer predictions

(from GeneHancer [25]) as well as available histone marks (such as H3K27ac associated with enhancer regions) from cancer cell-lines harbouring gain-of-function mutant p53 (see Methods). In breast carcinoma, we observed that around 50% of the peaks, in both gained or lost, overlapped with the enhancer regions. Moreover, the peaks that overlapped both enhancer prediction and H3K27ac marks in breast cancer cell lines were higher in gained peaks compared to the lost peaks (Fig. 3a). Similar results of enhancer overlap were observed in colon adenocarcinoma (Fig. 3b), however, the regions that overlapped with both enhancer and H3K27ac were less (except in HT-29 cell line). Interestingly, the peaks that overlapped H3K27ac (including those that overlapped both H3K27ac and enhancers) showed higher accessibility compared to peaks that overlapped only the enhancers or none (Fig. 3c, d).

Previous studies have shown that the wild-type and mutant p53 protein can recognize specific non-B-form DNA structures [26]. To check this, we obtained the genome-wide prediction of different non-B-form structures and checked for their enrichment in the gained and lost peak regions (see Methods). This showed that only the G-quadruplex structure was significantly enriched in the gained and lost peaks, especially in the promoter regions, in breast carcinomas (and in gained peaks in colon adenocarcinoma) as compared to the random peaks in both tumour types (one-sided Fisher's exact test, $P < 0.001$, see Additional file 1 Table S5).

Impact of chromatin accessibility changes on gene expression

We then investigated if the differentially accessible peaks observed in the mutant p53 tumours lead to changes in the expression of the relevant genes. For this, we performed the differential RNA expression analysis between the wild-type and mutant p53 groups in breast and colon adenocarcinoma (see Methods) and compared the results with the genes annotated from the differential chromatin accessibility analysis. With the $|\log\text{-fold change}| > 1$ and FDR (q-value) < 0.1 , we identified 251 downregulated and 152 upregulated protein-coding genes in breast carcinoma (Fig. 4a, Additional file 1 Table S6). Of these, six downregulated genes (EGR2, D4S234E, POU3F1, IL24, PENK, DUSP13) and three upregulated genes (ME1, KRT16 and PI3) were targets previously known to be regulated by mutant p53 protein (based on the literature) [29]. In addition, 49 out of 251 downregulated and 19 out of 152 upregulated genes were also identified from the differential chromatin accessibility analysis (Fig. 4c, Additional file 1 Table S7), including the known target genes IL24 and ME1. To further extend this, we focused on the genes identified from differential chromatin accessibility analysis that was

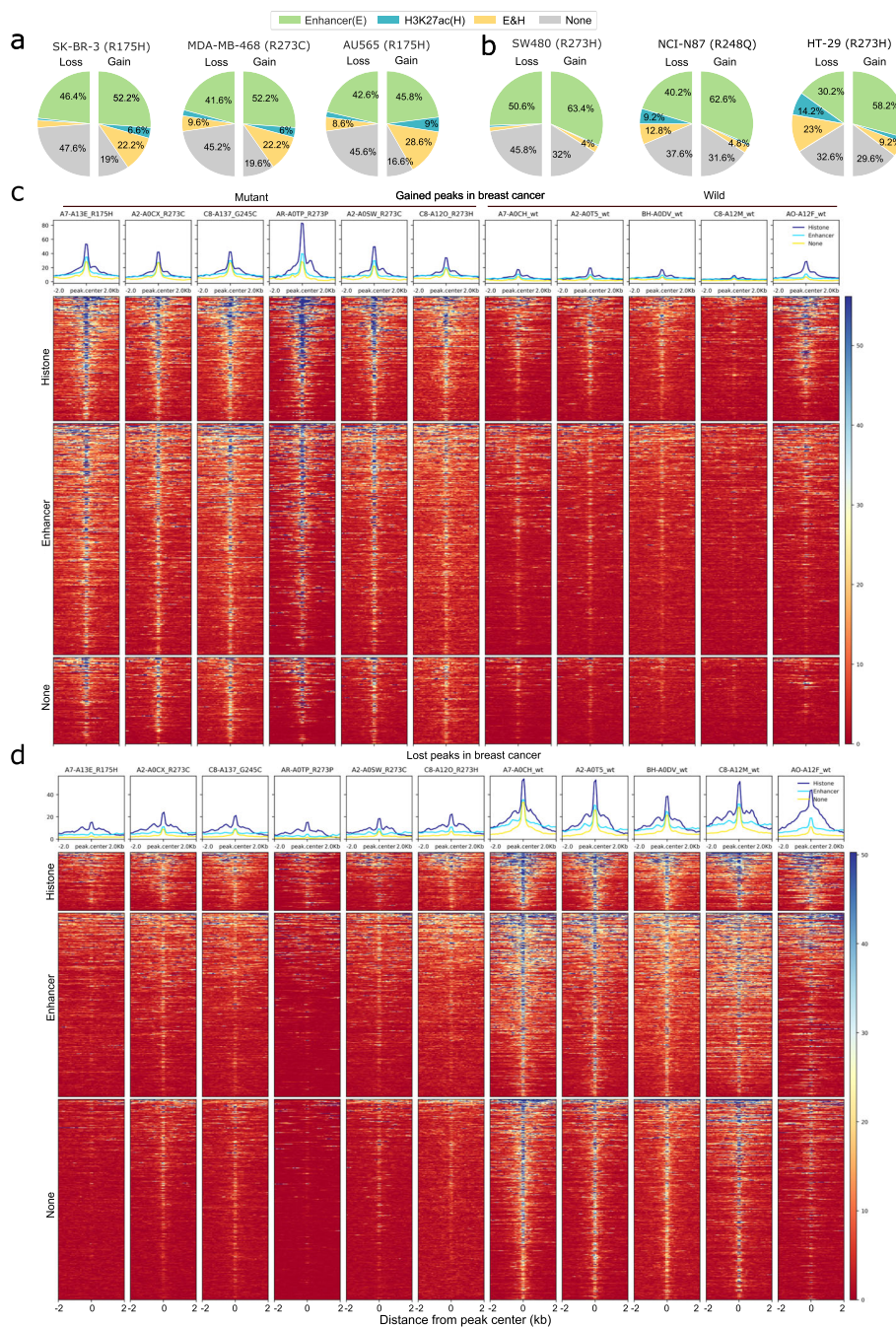


Fig. 3 Enhancer and histone marks overlap with the differential accessible peaks **a, b** The pie-chart shows the proportion of lost (left) and gained (right) peaks overlapped with the predicted enhancer regions (from GeneHancer) and H3K27ac mark from relevant cancer cell lines (harbouring gain-of-function mutations in p53) for breast and colon cancers. **c** Heatmap shows the chromatin accessibility score centred at the gained peaks (and immediate flanking regions) in breast cancer that overlapped both histone marks and enhancer predictions from MDA-MB-468 cell line (top), only the enhancer predictions (middle) or none (bottom). The column represents the individual sample that belongs to either mutant or wild-type p53 group. **d** Similar heatmap for the lost peaks in breast cancer

supported by expression change in a similar direction (that is, upregulated -- gain in accessibility with expression log-fold change > 1, and downregulated -- a loss in accessibility with expression log-fold change < -1), and were not affected by CNA events (Fig. 4c). Each gene

was annotated with the presence of de novo motifs identified, p53 response element and other known co-factors of p53 binding to the differentially accessible regions of the respective genes (as shown in Fig. 4c). Gene set enrichment analysis showed that some of the

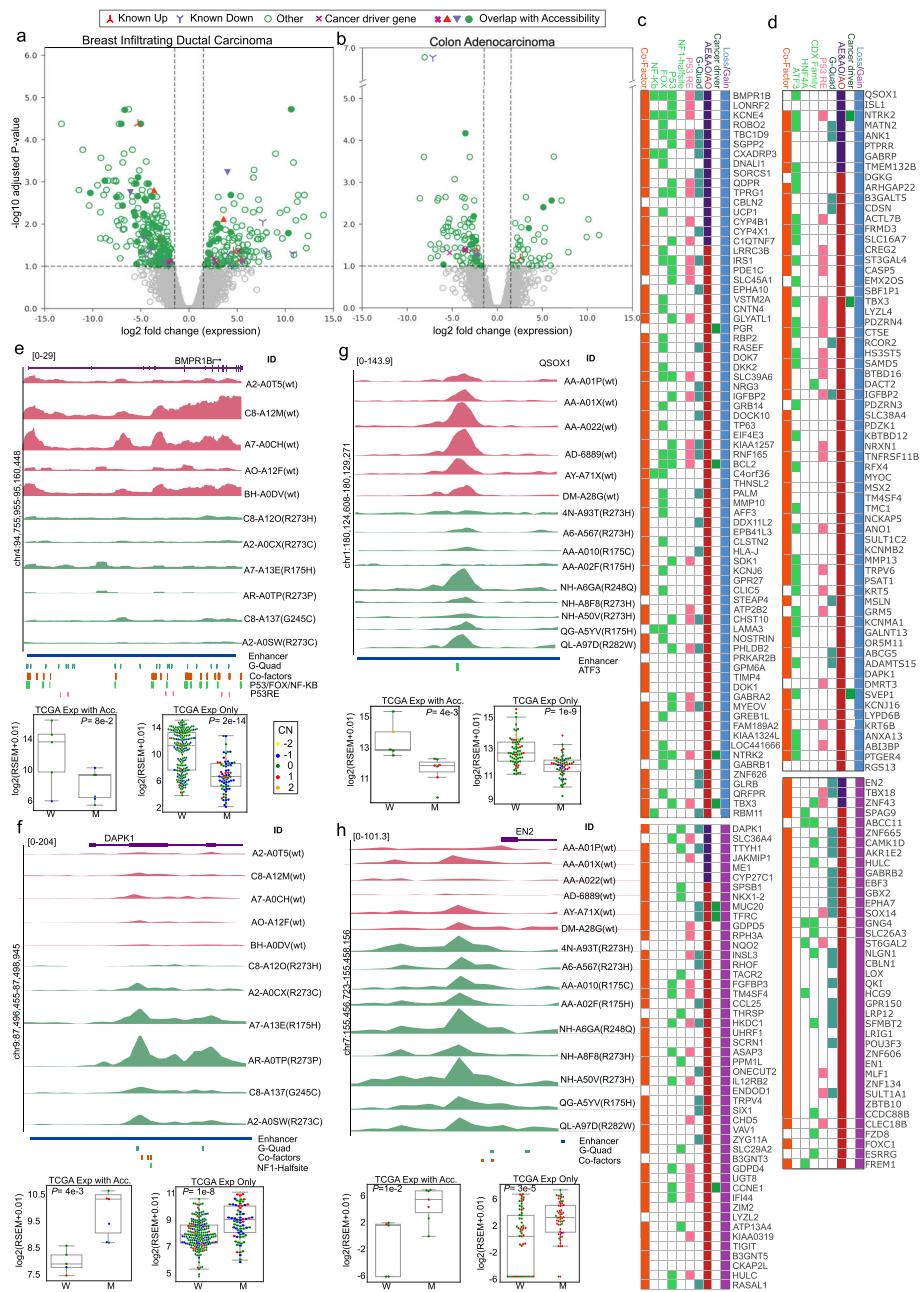


Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 Impact of differential chromatin accessibility on gene expression. **a** Volcano plot shows the differentially expressed genes in mutant p53 with respect to the wild-type p53 tumours. Each dot represents a gene, and the x-axis represents the log₂ fold change of expression value in mutant versus wild-type p53 tumours, and the y-axis shows the -log₁₀ adjusted p-value. The vertical and horizontal dashed lines indicate the threshold of FDR (q-value) < 0.1 and |log₂FoldChange(expression)| of 1. The triangle and inverted triangle symbol indicate the known target genes regulated by mutant p53, the symbol X indicates the cancer driver genes [18]. The filled symbols indicate the genes that appeared significant in the differential accessibility analysis as well. **c, d** The heatmap (generated using gitools [27]) shows the list of genes that have differential accessible peaks and have expression change in the same direction (gain in accessibility with increased expression, and loss in accessibility with reduced expression). Each gene is annotated whether it overlapped with known cancer driver genes [18], appeared in either differential expression analysis (AE), differential accessibility (AO) or both, presence of predicted G-quadruplex motif, p53 response element, de novo motifs and/or known mutant p53 co-factors in the differential accessibility region. **e-h** The chromatin accessibility profile in mutant and wild samples p53 samples for selected genes that showed significant differential gene expression and differential chromatin accessibility peaks. The plots were generated using SparK [28], each row represents a sample, and the tracks at the bottom represents annotation of predicted enhancer regions (from GeneHancer), G-quadruplex motifs, and binding sites of known mutant p53 co-factors (See methods). The boxplot shows the expression of the gene in wild-type (W) versus mutant p53 (M) tumours for the matched samples with chromatin accessibility data and in additional TCGA samples which do not have chromatin accessibility data. Each dot represents a sample and the colour indicates the relative copy number status of the gene (-2 deep deletion, -1 deletion, 1 amplification, 2 high amplification) in the respective samples. The p-value was computed using the Wilcoxon rank-sum test

downregulated genes (such as GCM1, TP63, SDK1, NRG3, BCL2, NTRK2, EPHA7, TBX3, CNTN4, ROBO2, BMPR1B, VSTM2A) were enriched in cell differentiation and proliferation regulation. Of these, BMPR1B (for example) was found with a high number ($n = 53$) of chromatin accessibility lost peaks across the gene body and enhancer regions (Fig. 4e). BMPR1B encodes a member of bone morphogenetic proteins (BMPs) that belong to the transforming growth factor- β (TGF- β) family. It has been reported that reduced expression of BMPR1B can increase the proliferation of breast cancer cells, and is also associated with poor prognosis and bone metastasis in breast cancer patients [30, 31]. Consistent with the loss of accessibility, we observed a lower expression of this gene in the mutant p53 samples as compared to the wild-type p53 samples (Wilcoxon rank-sum test, $P = 0.08$). We checked this association in additional TCGA breast cancer samples (which does not have chromatin accessible data) and found that the expression of BMPR1B was significantly lower in mutant p53 samples (Wilcoxon rank-sum test, $P = 2 \times 10^{-14}$) (Fig. 4e). This suggests that in mutant p53 tumours the lower expression of BMPR1B can be explained by the changes in the chromatin structure, mediated by the mutant p53 protein. On the other hand, among the up-regulated genes, we found DAPK1 (Death-associated protein kinase 1) as one of the interesting candidates (Fig. 4f). DAPK1 is a positive mediator of gamma-interferon induced programmed cell death. It has been shown that the high expression of DAPK1 can cause increased cell growth, and the depletion or inhibition of DAPK1 suppressed the cell growth, especially the mutant p53 but not wild-type p53 breast cancer cells [32]. We observed an increase in accessibility in the enhancer regions of DAPK1 (Fig. 4f), harbouring binding sites for the p53 response element, co-factors and G-quadruplex structure. The increase in DAPK1 expression was observed in these mutant p53

samples ($P = 0.004$), as well as in the additional TCGA breast cancer samples ($P = 1 \times 10^{-8}$). This suggests that the mutant p53 protein, together with other co-factors, could potentially transactivate the expression of DAPK1 in these tumours.

In colon adenocarcinomas, we found 95 downregulated and 44 upregulated genes from the differential RNA expression analysis (Fig. 4b, Additional file 1 Table S6). Of these, three downregulated genes (KIAA1324, ANK1, ABCA12) and one upregulated gene (ZNF415) were previously known targets of mutant p53 proteins [29]. In addition, 13 out of 95 downregulated (including ANK1) and 3 out of 44 upregulated genes were also identified from the differential chromatin accessibility analysis (Fig. 4d, Additional file 1 Table S7). The lesser number of overlap can be explained by the fact that four of the samples with chromatin accessibility data did not have matched gene expression data (see Methods). Further, we focused on the genes identified from differential chromatin accessibility analysis that showed expression change in a similar direction (as mentioned above) (Fig. 4d). Gene set enrichment analysis showed that the up-regulated genes, that has gain in the accessibility and increased expression (such as DMRT3, EBF3, EN1, EN2, ESRRG, FOXC1, GBX2, POU3F3, SFMBT2, SOX14, SPDEF, TBX18, TFAP2A, ZNF134, ZNF43, ZNF606, and ZNF665), were enriched in transcription regulatory activity and chromatin function. For example, the gene EN2 (Engrailed-2) is a member of the engrailed homeobox family. A recent study has shown the upregulation of EN2 in colorectal cancers associated with poor prognosis, and the knockdown of this in a colorectal cancer cell line, SW480 (harbouring GOF mutant p53), has inhibited proliferation and migration capacities of the cells [33]. We observed a gained peak in the promoter region of this gene. This peak harbours a G-quadruplex structure and overlaps with the mutant p53 ChIP-seq

derived from the SW480 cell line. Consistent with this, we observed a significant increase in expression of EN2 in mutant p53 samples in matched RNA seq data ($P = 0.01$) and in additional TCGA colorectal samples ($P = 3 \times 10^{-5}$) (Fig. 4h). This suggests that the expression of EN2 can be regulated by the mutant p53 through direct binding, or together with their co-factors, in the promoter region. Among the downregulated genes, with loss of chromatin accessibility and reduced gene expression, we found QSOX1 as one of the interesting candidates. The loss of chromatin accessibility was observed at QSOX1 gene-enhancer interaction region located 28 kb upstream of the TSS. This region also contains a predicted ATF3 motif (which was enriched in the lost peaks of colon adenocarcinoma, Fig. 4g). QSOX1 helps in protein folding by oxidizing protein thiols by reduction of oxygen to hydrogen peroxide. It has been shown that downregulation of QSOX1 reduces cell-cell adhesion, and thus increases tumour migration and metastasis in breast cancer cell lines [34]. QSOX1 expression has been shown to be a prognostic marker for breast and pancreatic cancer [35]. However, a study on breast cancer cell-line (of epithelial origin) showed an inverse relationship of QSOX1 expression and cell proliferation [36]. Consistent with this, we observed a significantly reduced expression of QSOX1 in matched samples ($P = 0.004$) and additional TCGA samples ($P = 1 \times 10^{-9}$) (Fig. 4g).

In sum, by integrating the gene expression data, we identified some of the known and novel targets regulated by the mutant p53 potentially through chromatin changes. Some of these genes are involved in transcriptional regulation and regulation of cell differentiation and proliferation.

Discussion

Several mechanisms have been proposed for the gain-of-function activity of mutant p53 under different cellular contexts and how that provides an advantage for the tumour cells to grow, metastasise and become resistant to treatments in different p53 tumour types [see review [37]]. Although the previous studies including the ChIP-seq of mutant p53 and gene expression-based analysis helped to identify genes (dys)regulated in the mutant p53 cell lines and tumours, the local chromatin changes induced by mutant p53 (either through direct DNA binding or by interacting with other TFs or chromatin-modifying enzymes) in primary tumours is poorly understood. In this study, we tried to address this limitation by analysing the available chromatin accessibility data (ATAC-seq) generated in patient tumours from TCGA cohort, and identified differential accessible regions (that is, gained or lost chromatin accessible regions) in mutant p53 tumours compared to the wild-type p53 tumours, especially in breast infiltrating ductal carcinoma and colon adenocarcinoma. We found that the number of lost peaks are higher compared to the gained peaks in both cancer types, and this can be explained by the loss of wild-type p53 activity. However, the overlap of gained or lost peaks with the mutant p53 binding regions (inferred from ChIP-seq) data was less. This suggests that the direct binding of mutant p53 occurs preferentially in existing open chromatin regions.

Moreover, we found that the gained and lost peaks were enriched for binding motifs for known co-factors of mutant p53 and other master TF regulators, enhancer marks and G-quadruplex structure (Fig. 5). Taken together, this result suggests that the majority of the

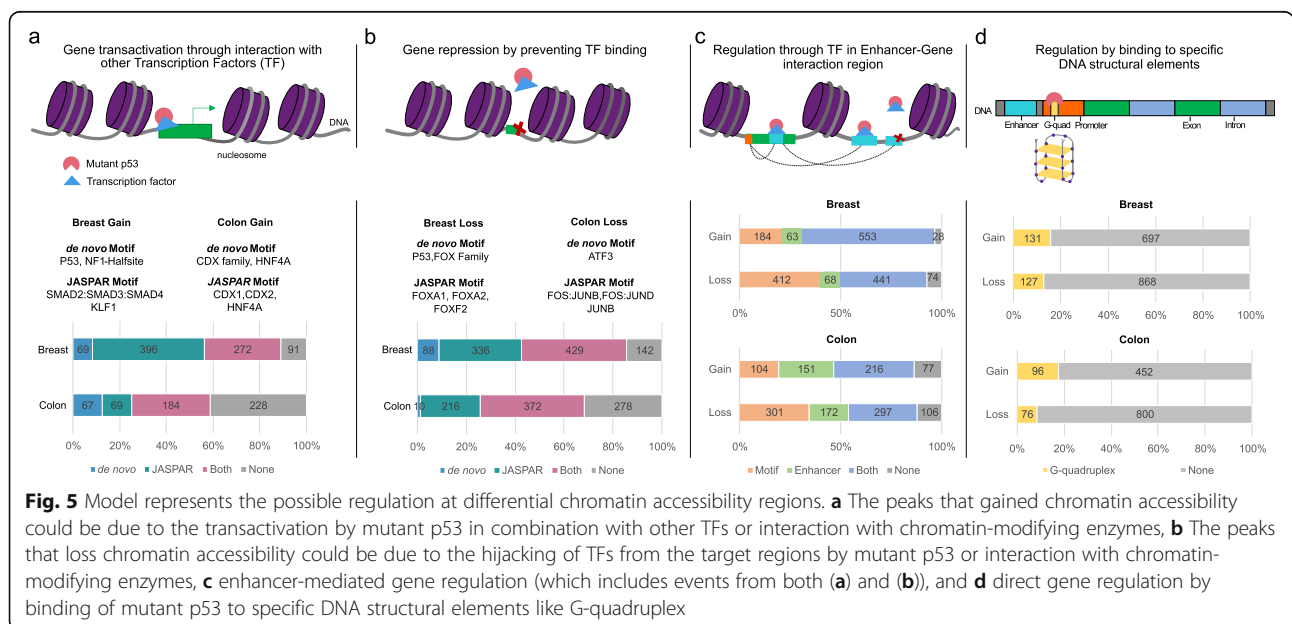


Fig. 5 Model represents the possible regulation at differential chromatin accessibility regions. **a** The peaks that gained chromatin accessibility could be due to the transactivation by mutant p53 in combination with other TFs or interaction with chromatin-modifying enzymes, **b** The peaks that loss chromatin accessibility could be due to the hijacking of TFs from the target regions by mutant p53 or interaction with chromatin-modifying enzymes, **c** enhancer-mediated gene regulation (which includes events from both (a) and (b)), and **d** direct gene regulation by binding of mutant p53 to specific DNA structural elements like G-quadruplex

gained and lost peaks were located in the gene regulatory regions. By integrating gene expression data, we found that some of the genes linked to the differential accessible peaks showed significant expression change in matched tumour samples. This includes some of the known genes regulated by mutant p53 and other novel candidates (shown in Fig. 4). Compared to the previous studies that revealed mutant p53 specific gene expression signatures (based on the gene expression data alone) [38], this study provides additional information for those genes that underwent both local chromatin changes in the genic or regulatory regions, as well as the gene expression changes.

The limitations of the study include (a) although we integrated the chromatin accessibility, gene expression and somatic mutations data from the matched patient samples available in TCGA cohort, these data were not generated from the same cells and thus we expect some cell-to-cell variability within the tumor, (b) we attempted to identify gained and lost peaks that potentially affected by the copy-number alteration. However, this can be further improved by incorporating allele-specific copy number data along with allele-specific chromatin accessibility data, and (c) given that many of the mutant p53 tumours analysed here have biallelic alteration in TP53 (that is, gain-of-function mutation together with loss of wild-type TP53 allelic), the differential chromatin accessibility peaks identified can be explained by the individual as well as combined effects of loss of wild-type p53 activity and GOF mutant p53 activity.

Overall, our analysis reveals tissue- and tumour-type specific chromatin changes in the mutant p53 tumours compared to the wild-type p53 tumours. Further studies are required to identify the mechanisms underlying the chromatin changes and the role of mutant p53 in it, and under different stress conditions.

Conclusion

Through this study, we uncover chromatin changes which take place due to direct/indirect mechanisms of mutant p53 regulation and therefore causing expression dysregulation of various target genes. Variability between the results from breast and colon cancers suggests the tissue- and tumour-type specific role of mutant p53 in regulating chromatin structure and gene expression. This study furthers our understanding of the impact of mutant p53 in cancer development.

Methods

Identification of mutant p53 samples

We obtained the sample details of TCGA primary tumours (404 unique samples across 23 cancer types) that have chromatin accessibility data (ATAC-seq) from Corces & Granja et al. [16]. For those samples, we

retrieved the somatic mutation calls (MC3) [39] and somatic copy number data from the TCGA pan-cancer atlas study (<https://gdc.cancer.gov/about-data/publications/pancanatlas>). The allele-specific copy number estimated from the ASCAT algorithm (using Affymetrix SNP6 arrays data) was obtained from Martincorena et al. [40]. Further, we searched for samples harbouring gain-of-function (GOF) missense mutations in the DNA-binding domain of p53 (also referred as hotspot mutants), such as R273C/H/L/P/S/W/X/Y, R175A/C/G/H/L, V157F, Y220C, G245A/C/D/F/R/S/V, R248E/G/L/P/Q/W, R249G/K/M/S/W, R282G/H/Q/W and classified them as GOF mutant p53 samples [7]. Samples that have both GOF missense mutation and minor allele copy number equals zero were classified as biallelic. Whereas, samples that do not carry any protein-affecting mutations (non-synonymous or non-sense) nor copy-number alterations in TP53 (that is, relative copy number value equals zero) were classified as wild-type p53 samples (see Additional File 1 Table S1).

Identification of differential chromatin accessibility regions

For the differential chromatin accessibility analysis, we focused on two specific cancer types (such as breast infiltrating ductal carcinoma and colon adenocarcinoma) that had sufficient sample size in the wild-type and mutant p53 groups (see Additional File 1 Table S1). Cancer type-specific raw counts of ATAC-Seq peaks (called at the individual sample-level) were obtained from Corces & Granja et al. [16] (<https://gdc.cancer.gov/about-data/publications/ATACseq-AWG>). This matrix consisted of peaks on the rows and the sample information on the columns. From this, we selected the columns that matched the wild-type and mutant p53 samples. All samples (except one) have two technical replicates and we considered both of them for our analysis. Further, to filter out peaks with low read counts, we converted the raw counts into CPM (counts per million) values to correct for the differences in library size between samples by using edgeR and removed peaks that do not have minimum 1 CPM in any of the samples considered (both wild-type and mutant). The remaining peaks (209,835 in breast infiltrating ductal carcinoma and 119,925 in colon adenocarcinoma) were subjected for differential analysis using the voom and limma eBayes packages [41]. These tools were selected based on the recent benchmark study for the differential accessibility analysis using ATAC-seq data [42] and also used in Corces & Granja et al. [16]. At first, the raw counts were transformed to log₂(CPM) score (with a prior count of 5), followed by quantile normalization using voom. The resultant count matrix, along with weights obtained from the mean-variance

relationship of each row (peak region), were fitted to a linear model using `lmFit` function. This model was then given to `limma` `eBayes` function, along with the TP53 mutation status for each sample, to obtain the differentially accessible peaks in mutant p53 samples compared to the wild-type p53 samples. We performed this analysis separately for breast and colon cancers to avoid any bias from the tumour- and tissue-specific differences.

Genomic annotation of peaks

The `annotatePeaks` function from HOMER [21] was used to annotate peaks to the closest gene possible (as well as obtained the gene-type information: protein-coding, ncRNA or pseudogene) and further categorize the location of the peaks as either promoter-TSS, 5' UTR, exon, intron, 3' UTR, TTS, non-coding or intergenic (based on the UCSC hg38 v6.4 annotations).

Closest peak analysis

To find if the differentially accessible peaks are located close to the existing peaks, we computed the closest peak distance for each significant differentially accessible peak against non-significant peaks (from the respective cancer type) using `bedtools` `closest` function (with default parameters) [43]. The frequency of distance values was then binned into different groups ≤ 1 kb, 1 to 2 kb, 2 to 5 kb, 5 to 10 kb, 10 to 15 kb, 15 to 30 kb, and > 30 kb (shown in Fig. 1c and d).

Copy number alteration annotation

The segmented copy number variation (SCNV) data from the Affymetrix SNP 6.0 array were downloaded from GDC portal (using TCGAbiolinks [44], especially the file type "grch38.seg.v2"). We intersected the genomic location of the differentially accessible peaks with the genomic location of segmented data (using `bedtools` [43]) and assigned the segment mean score to each peak. In cases where the peak overlapped with multiple segments, the median value of the overlapping segments was assigned. Then for each peak, we computed the median value of the segment mean score across wild-type and mutant p53 samples, respectively. Further, we annotated each peak that potentially has copy number gain if the median score > 0.3 in mutant or < -0.3 in wild-type, and copy number loss if median score < -0.3 in mutant or > 0.3 wild p53 samples. These annotations were shown in Fig. 1e and f.

The gene-level relative copy number data was used to calculate a median score for mutant and wild sample groups [45]. This was further used to filter out genes with copy number bias (median relative copy number was not equal to zero in both mutant and wild-type p53 samples) while selecting genes that showed both

differential chromatin accessibility and expression change (Fig. 4c, d).

p53 response element and mutant p53 ChIP data

To check whether any of the differentially accessible peaks contained a putative p53 response element, we used `p53 retriever` (with the default parameter) [19] to search for canonical (full site), as well as non-canonical (half-sites and 3/4 sites) p53 response elements.

Publicly available ChIP-seq data of mutant p53 from breast cancer cell lines HCC70 (harbouring GOF mutation R248Q) and MDA-MB-468 (R273C) [15], and colon cell line SW480 (R273H) [14], was obtained from ChIP-Atlas [46]. We obtained the data from all these cell-lines and intersected with the differentially accessible peaks (using `bedtools` [43]) to annotate peaks that overlap mutant p53 binding sites (shown in Fig. 1e and f). Further, to compare the accessibility score between wild-type and mutant p53 samples, we computed the median CPM (across samples) for the mutant and wild-type p53 samples. The distribution of median CPM was shown in Additional file 2 Figure S1, and the difference between the wild-type and mutant median values for each peak (separated based on their genomic annotation) was shown in Additional file 2 Figure S2.

Motif enrichment analysis

For TF motif enrichment analysis, we obtained the genome-wide transcription factor binding sites (TFBS) predictions based on curated TF motifs from JASPAR (with p -value ≤ 0.0001) [20]. The genomic coordinates of the gained and lost peaks were intersected with the TFBS to compute the percentage overlap (that is, the proportion of gained or lost peaks overlapping with the binding sites of a particular TF). Further, to check for the enrichment, we used one-sided Fisher's exact test to compare the proportion of peaks overlapping with the binding sites of particular TF in gained or lost peaks with the proportion in randomly selected non-significant peaks (with the size and genomic annotations distribution equal to that of gained or lost peaks). The p -values were subjected to multiple hypothesis testing correction using Benjamini-Hochberg's approach (Additional file 1 Table S4). Sites for previously known co-factors from JASPAR was also annotated in the significant peaks, these include: ELF1, ELF2, ELF4, GABPA, ERG, FLI1, FEV, ERF, ETV3, ELF3, ELF5, ESE3, ETS1, ETS2, SPDE F, ETV4, ETV5, ETV1, ETV2, SPI1, SPIB, SPIC, ELK1, ELK4, ELK3, ETV6, ETV7, SP1, SP5, SP2, AP1, HSF1, NFYA, NFYB, NFYC, TP63, TP73.

We also performed *de novo* motif enrichment analysis using `findMotifsGenome` (with default parameters) from HOMER to search for sequence motif enriched in gained and lost peaks. To identify the motifs that are specifically

enriched in the peaks that do not have predicted p53 response elements and/or overlap with the mutant p53 bindings, we split the gained and lost peaks based on the above criteria and performed the analysis separately in each group (for gain and lost peaks, Additional file 2 Figure S2). In both groups, we filtered out peaks that were annotated with copy-number alterations (gain or loss). The HOMER results for each group were shown in Additional file 3. The top k-mer motifs shown in Fig. 2e and f were selected based on the enrichment p-value and percentage overlap.

The plotheatmap and plotprofile function from DeepTools [47] package was used to plot the accessibility profile centred at the peaks enrichment with particular TF motif or de novo motifs (shown in Fig. 2g-l).

Enhancer and histone enrichment

Curated Double Elite Gene-Enhancer interaction regions were obtained from GeneHancer database [25] and overlapped with the significant differentially accessible peaks in both cancer types. Additional information on active histone marks (such as H3K27ac) associated with enhancer regions was obtained from the respective cancer cells (harbouring GOF p53 mutations). SK-BR-3 (R175H), MDA-MB-468 (R273C), and AU565 (R175H) were selected for breast cancer; and SW480 (R273H), HT-29 (R273H), and NCI-N87 (R248Q) for colon cancer. The plotheatmap from DeepTools tools was then used to compare the accessibility profile of peaks that overlapped with both enhancers and histone marks, only enhancers or none (as shown in Fig. 3c and d).

Non-B-form structure enrichment

A collection of predicted non-B DNA-forming motifs was obtained from non-B DB database [48]. This includes A-Phased Repeats, Direct Repeat, G-Quadruplex Motif, Inverted Repeat, Mirror Repeat, Short Tandem Repeat, and Z DNA Motif. The genomic locations of these predicted non-B-form structure regions were intersected (using bedtools [43]) with the gained or lost peaks to compute the percentage overlap and compared it with the randomly selected peaks (with size and genomic annotation ratio equal to that of gained or lost peaks) to compute the enrichment using one-sided Fisher's exact test.

Differential gene expression analysis

In breast, all samples with the chromatin accessibility data had matched gene expression data, but in colon four samples (one wild-type and three mutant p53) did not have the matched expression data (when aliquot identifier was taken into account). The raw expression counts (rsem.genes.results) of protein-coding genes for breast and colon cancer samples were obtained from the

GDC portal using TCGAAbiolinks [44]. Further, the count matrix was normalized (using TCGAanalyze_Normalization function) and filtered out genes with low expression count (using TCGAanalyze_Filtering with method = quantile, and qnt.cut = 0.25), leaving 14,892 genes in breast and 14,890 genes in colon. Further, the differential expression analysis was carried out using edgeR through TCGAanalyze_DEA function (with method = glmLRT). Genes with $|\log_2\text{foldchange}| > 1$ and FDR (q-value) < 0.1 were deemed significant.

The gene level normalized RSEM values was used to compute the median expression values for wild-type and mutant samples, and the expression fold change. The $|\log_2\text{foldchange}| > 1$ and neutral copy number in both mutant and wild p53 samples was used to select genes from differential accessibility (as shown in Fig. 4c and d). These genes were used for the gene set enrichment analysis using GSEA [49, 50]. For the selected candidates in breast and colon, the expression difference in the matched TCGA samples, as well as in additional TCGA samples, was computed using the Wilcoxon rank-sum test.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12885-021-08362-x>.

Additional file 1: Table S1. TCGA samples with TP53 mutation status.

Table S2. The annotated significant differentially accessibility peaks identified in breast and colon cancers. **Table S3.** Enrichment of CNA events in different distance bins. **Table S4.** Results of TF motif enrichment analysis (related to Fig. 2a-d). **Table S5.** Enrichment of non-B-form DNA structure in differentially accessible peaks. **Table S6.** Results from differential gene expression analysis between mutant and wild-type p53 tumours. **Table S7.** Summary gene list from the differential accessibility, differential expression, and overlap of both.

Additional file 2: Figure S1. The box-plot shows the distribution of accessibility scores (CPM) in wild-type and mutant p53 samples for peaks that overlapped the mutant p53 binding sites (inferred from ChIP-seq data). The peaks were split into three groups: non-significant, significant gained and significant lost, as shown on the x-axis. **Figure S2.** The box-plot shows the distribution of the difference between the accessibility scores (CPM) between wild-type and mutant p53 tumours in peaks (both significant and non-significant) that overlapped the mutant p53 binding sites (inferred from ChIP-seq data). The peaks were split based on the genomics annotations as shown on the x-axis. **Figure S3.** The significant gained and lost accessible peaks identified in breast and colon cancer have been categorized into those that overlapped regions with potentially copy number alterations, those that predicted to have p53 response element, overlap with mutant p53 binding sites (inferred from ChIP-seq data), both or none. **Figure S4.** The heatmap represents the overlap of differentially accessible peaks predicted with TF motifs from JASPAR and de novo motif analysis from HOMER. The colour scale represents the Jaccard index score. The higher the value the higher the overlap between the genomic regions.

Additional file 3. de novo motifs search result from HOMER analysis in the differentially accessible peaks of breast and colon cancers.

Acknowledgements

We thank Dimple Notani, Ramanathan Sowdhamini, S. Shivakumara, Nuria Lopez-Bigas, and members of RS lab for feedback and suggestions. We thank Eric Macwan for his help with the annotation of mutation status in patient

samples. The results shown here are in part based upon data generated by the TCGA Research Network: <https://www.cancer.gov/tcga>.

Code availability

The scripts used to run the analyses and generate figures are available at https://github.com/onkoslab/p53_mutant_atac.

Authors' contributions

R.S. conceived and supervised the study. B.D. performed all analyses. B.D. and R.S. designed analyses, interpreted the data and wrote the manuscript. All authors read and approved the final manuscript.

Funding

We acknowledge funding support from NCBS-TIFR and the Department of Atomic Energy, Government of India, under project no. 12-R&D-TFR-5.04-0800. R.S. acknowledges support from the Ramanujan fellowship (SERB, SB/S2/RJN-071/2018).

Availability of data and materials

The TCGA breast and colon cancers ATAC-seq peak raw counts, used for the differential accessibility analysis of mutant versus wild p53 samples, were obtained from <https://gdc.cancer.gov/about-data/publications/ATACseq-AWG>. The somatic mutations, copy number alterations and gene expression data were obtained from <https://portal.gdc.cancer.gov/>.

Declarations

Ethics approval and consent to participate

Not Applicable.

Consent for publication

Not Applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 13 December 2020 Accepted: 13 May 2021

Published online: 05 June 2021

References

- Harms KL, Chen X. The functional domains in p53 family proteins exhibit both common and distinct properties. *Cell Death Differ*. 2006;13(6):890–7. <https://doi.org/10.1038/sj.cdd.4401904>.
- Sullivan KD, Galbraith MD, Andryszk Z, Espinosa JM. Mechanisms of transcriptional regulation by p53. *Cell Death Differ*. 2018;25(1):133–43. <https://doi.org/10.1038/cdd.2017.174>.
- Levine AJ. p53: 800 million years of evolution and 40 years of discovery. *Nat Rev Cancer*. 2020;20(8):471–80. <https://doi.org/10.1038/s41568-020-0262-1>.
- Hafner A, Bulyk ML, Jambhekar A, Lahav G. The multiple mechanisms that regulate p53 activity and cell fate. *Nat Rev Mol Cell Biol*. 2019;20(4):199–210. <https://doi.org/10.1038/s41580-019-0110-x>.
- Integrated analysis of TP53 gene and pathway alterations in the cancer genome atlas. *Cell Rep*. 2019;28:1370–84.e5. <https://doi.org/10.1016/j.celrep.2019.07.001>.
- Sabapathy K, Lane DP. Therapeutic targeting of p53: all mutants are equal, but some mutants are more equal than others. *Nat Rev Clin Oncol*. 2018;15(1):13–30. <https://doi.org/10.1038/nrclinonc.2017.151>.
- Olivier M, Bouaoun L, Sonkin D, Ardin M, Hollstein M, Byrnes G, et al. TP53 variations in human cancers: new lessons from the IARC TP53 database and genomic studies. *Eur J Cancer*. 2016;61:S15. [https://doi.org/10.1016/s0959-8049\(16\)61042-0](https://doi.org/10.1016/s0959-8049(16)61042-0).
- Eldar A, Rozenberg H, Diskin-Posner Y, Rohs R, Shakked Z. Structural studies of p53 inactivation by DNA-contact mutations and its rescue by suppressor mutations via alternative protein–DNA interactions. *Nucleic Acids Res*. 2013;41(18):8748–59. <https://doi.org/10.1093/nar/gkt630>.
- Willis A, Jung EJ, Wakefield T, Chen X. Mutant p53 exerts a dominant negative effect by preventing wild-type p53 from binding to the promoter of its target genes. *Oncogene*. 2004;23(13):2330–8. <https://doi.org/10.1038/sj.onc.1207396>.
- Xu H, El-Gewely MR. Differentially expressed downstream genes in cells with normal or mutated p53. *Oncol Res*. 2003;13(6):429–36. <https://doi.org/10.3727/096504003108748456>.
- William A, Freed-Pastor CP. Mutant p53: one name, many proteins. *Genes Dev*. 2012;26(12):1268–86. <https://doi.org/10.1101/gad.190678.112>.
- Kim MP, Lozano G. Mutant p53 partners in crime. *Cell Death Differ*. 2018;25(1):161–8. <https://doi.org/10.1038/cdd.2017.185>.
- Bouaoun L, Sonkin D, Ardin M, Hollstein M, Byrnes G, Zavadii J, et al. TP53 variations in human cancers: new lessons from the IARC TP53 database and genomics data. *Hum Mutat*. 2016;37(9):865–76. <https://doi.org/10.1002/humu.23035>.
- Rahnamoun H, Hong J, Sun Z, Lee J, Lu H, Lauberth SM. Mutant p53 regulates enhancer-associated H3K4 monomethylation through interactions with the methyltransferase MLL4. *J Biol Chem*. 2018;293(34):13234–46. <https://doi.org/10.1074/jbc.RA118.003387>.
- Zhu J, Sammons MA, Donahue G, Dou Z, Vedadi M, Getlik M, et al. Prevalent p53 mutants co-opt chromatin pathways to drive cancer growth. *Nature*. 2015;525(7568):206–11. <https://doi.org/10.1038/nature15251>.
- Corces MR, Granja JM, Shams S, Louie BH, Seoane JA, Zhou W, et al. The chromatin accessibility landscape of primary human cancers. *Science*. 2018;362(6413):eaav1898. <https://doi.org/10.1126/science.aav1898>.
- Knudson AG. Mutation and cancer: statistical study of retinoblastoma. *Proc Natl Acad Sci U S A*. 1971;68(4):820–3. <https://doi.org/10.1073/pnas.68.4.820>.
- Tamborero D, Rubio-Perez C, Deu-Pons J, Schroeder MP, Vivancos A, Rovira A, et al. Cancer genome interpreter annotates the biological and clinical relevance of tumor alterations. *Genome Med*. 2018;10(1):25. <https://doi.org/10.1186/s13073-018-0531-8>.
- Tebaldi T, Zaccara S, Alessandrini F, Bisio A, Ciribilli Y, Inga A. Whole-genome cartography of p53 response elements ranked on transactivation potential. *BMC Genomics*. 2015;16(1):1–13. <https://doi.org/10.1186/s12864-015-1643-9>.
- Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA, et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res*. 2020;48:D87–92. <https://doi.org/10.1093/nar/gkz1001>.
- Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell*. 2010;38(4):576–89. <https://doi.org/10.1016/j.molcel.2010.05.004>.
- Maria L, Golson KHK. Fox transcription factors: from development to disease. *Development*. 2016;143(24):4558–70. <https://doi.org/10.1242/dev.112672>.
- Verzi MP, Shin H, Ak SR, Liu XS, Shivdasani RA. Intestinal master transcription factor CDX2 controls chromatin access for partner transcription factor binding. *Mol Cell Biol*. 2013;33(2):281–92. <https://doi.org/10.1128/MCB.01185-12>.
- Hess J, Angel P, Schorpp-Kistner M. AP-1 subunits: quarrel and harmony among siblings. *J Cell Sci*. 2004;117 Pt 25. <https://doi.org/10.1242/jcs.01589>.
- Fishilevich S, Nudel R, Rappaport N, Hadar R, Plaschkes I, Stein TI, et al. GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database*. 2017;2017. <https://doi.org/10.1093/database/bax028>.
- Petr M, Helma R, Polášková A, Krejčí A, Dvořáková Z, Kejnovská I, et al. Wild-type p53 binds to MYC promoter G-quadruplex. *Biosci Rep*. 2016;36(5). <https://doi.org/10.1042/BSR20160232>.
- Perez-Llamas C, Lopez-Bigas N. Gtools: analysis and visualisation of genomic data using interactive heat-maps. *PLoS One*. 2011;6(5):e19541. <https://doi.org/10.1371/journal.pone.0019541>.
- Kurtenbach S, William HJ. SparK: a publication-quality ngs visualization tool. *Cold Spring Harbor Lab*. 2019;845529. <https://doi.org/10.1101/845529>.
- Garritano S, Inga A, Gemignani F, Landi S. More targets, more pathways and more clues for mutant p53. *Oncogenesis*. 2013;2(7):e54. <https://doi.org/10.1038/oncogenesis.2013.15>.
- Dai K, Qin F, Zhang H, Liu X, Guo C, Zhang M, et al. Low expression of BMP1B indicates poor prognosis of breast cancer and is insensitive to taxane-anthracycline chemotherapy. *Oncotarget*. 2016;7. <https://doi.org/10.18632/oncotarget.6613>.
- Bokobza SM, Ye L, Kynaston HE, Mansel RE, Jiang WG. Reduced expression of BMP1B correlates with poor prognosis and increased proliferation of breast cancer cells. *Cancer Genomics Proteomics*. 2009;6. <https://pubmed.ncbi.nlm.nih.gov/19451094/>. Accessed 30 Oct 2020.
- Zhao J, Zhao D, Poage GM, Mazumdar A, Zhang Y, Hill JL, et al. Death-associated protein kinase 1 promotes growth of p53-mutant cancers. *J Clin Invest*. 2015;125(7):2707–20. <https://doi.org/10.1172/JCI70805>.

33. Li Y, Liu J, Xiao Q, Tian R, Zhou Z, Gan Y, et al. EN2 as an oncogene promotes tumor progression via regulating CCL20 in colorectal cancer. *Cell Death Dis.* 2020;11(7):1–11. <https://doi.org/10.1038/s41419-020-02804-3>.
34. Pernodet N, Hermetet F, Adami P, Vejux A, Descotes F, Borg C, et al. High expression of QSOX1 reduces tumorigenesis, and is associated with a better outcome for breast cancer patients. *Breast Cancer Res.* 2012;14(5):1–15. <https://doi.org/10.1186/bcr3341>.
35. Douglas F, Lake DOF. The emerging role of QSOX1 in Cancer. *Antioxid Redox Signal.* 2014;21(3):485–96. <https://doi.org/10.1089/ars.2013.5572>.
36. Involvement of sulfhydryl oxidase QSOX1 in the protection of cells against oxidative stress-induced apoptosis. *Exp Cell Res.* 2007;313(19):3971–82. <https://doi.org/10.1016/j.yexcr.2007.09.003>.
37. Kim MP, Zhang Y, Lozano G. Mutant p53: multiple mechanisms define biologic activity in Cancer. *Front Oncol.* 2015;5. <https://doi.org/10.3389/fonc.2015.00249>.
38. Miller LD, Smeds J, George J, Vega VB, Vergara L, Ploner A, et al. An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci U S A.* 2005;102(38):13550–5. <https://doi.org/10.1073/pnas.0506230102>.
39. Ellrott K, Bailey MH, Saksena G, Covington KR, Kandoth C, Stewart C, et al. Scalable Open Science approach for mutation calling of tumor exomes using multiple genomic pipelines. *Cell Syst.* 2018;6(3):271–281.e7. <https://doi.org/10.1016/j.cels.2018.03.002>.
40. Martincorena I, Raine KM, Gerstung M, Dawson KJ, Haase K, Van Loo P, et al. Universal patterns of selection in Cancer and somatic tissues. *Cell.* 2017;171(5):1029–1041.e21. <https://doi.org/10.1016/j.cell.2017.09.042>.
41. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res.* 2015;43:e47. <https://doi.org/10.1093/nar/gkv007>.
42. Reske JJ, Wilson MR, Chandler RL. ATAC-seq normalization method can significantly affect differential accessibility analysis and interpretation. *Epigenetics Chromatin.* 2020;13(1):1–17. <https://doi.org/10.1186/s13072-020-00342-y>.
43. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26(6):841–2. <https://doi.org/10.1093/bioinformatics/btq033>.
44. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, et al. TCGAAbiolinks : an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* 2015;44(8):e71. <https://doi.org/10.1093/nar/gkv1507>.
45. Taylor AM, Shih J, Ha G, Gao GF, Zhang X, Berger AC, et al. Genomic and functional approaches to understanding cancer aneuploidy. *Cancer Cell.* 2018;33:676–89.e3. <https://doi.org/10.1016/j.ccell.2018.03.007>.
46. Oki S, Ohta T, Shioi G, Hatanaka H, Ogasawara O, Okuda Y, et al. ChIP-Atlas: a data-mining suite powered by full integration of public ChIP-seq data. *EMBO Rep.* 2018;19. <https://doi.org/10.15252/embr.201846255>.
47. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* 2016;44. <https://doi.org/10.1093/nar/gkw257>.
48. Cer RZ, Donohue DE, Mudunuri US, Temiz NA, Loss MA, Starnier NJ, et al. Non-B DB v2.0: a database of predicted non-B DNA-forming motifs and its associated tools. *Nucleic Acids Res.* 2013;41(Database issue):D94–100. <https://doi.org/10.1093/nar/gks955>.
49. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci.* 2005;102(43):15545–50. <https://doi.org/10.1073/pnas.0506580102>.
50. Mootha VK, Lindgren CM, Eriksson K-F, Subramanian A, Sihag S, Lehar J, et al. PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet.* 2003;34(3):267–73. <https://doi.org/10.1038/ng1180>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

