

p53motifDB: integration of genomic information and tumour suppressor p53 binding motifs

Gabriele Baniulyte, Sawyer M. Hicks, Morgan A. Sammons *

Department of Biological Sciences and the RNA Institute, University at Albany, SUNY, 1400 Washington Ave, Life Sciences Building Room 2078, Albany, NY 12222, United States

*Corresponding author. Department of Biological Sciences and the RNA Institute, University at Albany, SUNY, 1400 Washington Ave, Life Sciences Building Room 2078, Albany, NY 12222, United States. E-mail: masammons@albany.edu

Citation details: Baniulyte, G., Hicks, S.M., Sammons, M.A. p53motifDB: integration of genomic information and tumour suppressor p53 binding motifs. Database (2025) Vol. 2025: article ID baaf053; DOI: <https://doi.org/10.1093/database/baaf053>

Abstract

The tumour suppressor gene *TP53* encodes the DNA binding transcription factor p53 and is one of the most mutated genes in human cancer. Tumour suppressor activity requires binding of p53 to its DNA response elements and subsequent transcriptional activation of a diverse set of target genes. Despite decades of close study, the logic underlying p53 interactions with its numerous potential genomic binding sites and target genes is not yet fully understood. Here, we present a database of DNA and chromatin-based information focused on putative p53 binding sites in the human genome to allow users to generate and test new hypotheses related to p53 activity in the genome. Users can query genomic locations based on experimentally observed p53 binding, regulatory element activity, genetic variation, evolutionary conservation, chromatin modification state, and chromatin structure. We present multiple use cases demonstrating the utility of this database for generating novel biological hypotheses, such as chromatin-based determinants of p53 binding and potential cell type-specific p53 activity. All database information is also available as a precompiled SQLite database for use in local analysis or as a Shiny web application.

Database URL: <https://p53motifDB.its.albany.edu>

Introduction

Sequence-specific transcription factors (TFs) are key regulators of cellular and developmental processes [1]. Understanding the mechanisms by which TFs accurately recognize and bind to cognate DNA motifs and discriminate against other DNA sequences has been a central question in molecular and developmental biology for decades [2]. The number of potential TF motifs in the genome far outweighs the number of observed binding events [3], suggesting that sequence alone does not fully dictate binding. TFs must also integrate multiple other types of DNA and chromatin-embedded information during binding site selection *in vivo*. Nucleosome positioning, histone and DNA modifications, and chromatin conformation join DNA sequence and shape as factors regulating TF binding to DNA [4–6]. For example, most TFs cannot bind to their cognate motif when the sequence is engaged with a nucleosome, providing one molecular mechanism reducing the ratio of observed to potential binding events [5, 7].

Areas of open chromatin flanked by nucleosomes with histone modification patterns can be used as indirect evidence for regulatory region activity [8–10]. Massive efforts in mapping genomic locations of open chromatin and histone modification localization revealed clear patterns of cell type specificity in transcriptional regulation [11–13]. These data have been successfully integrated with TF binding information and other data sources, such as genetic variation and evolutionary conservation, to provide insight into the biological function and mechanisms of transcriptional regulatory regions [6, 14].

The tumour suppressor gene *TP53* encodes a DNA binding TF called p53. p53 controls a wide-ranging gene regulatory network that dictates a series of cellular behaviours, including cell cycle arrest, apoptosis, DNA repair, and metabolic control [15, 16]. Germline or somatic loss of p53 activity leads to a predisposition for tumorigenesis and cancer. Increased p53 activity can manifest in phenotypes like decreased fertility and germline dysfunction and in premature ageing [17–22]. The well-characterized tumour suppressor activity of p53 requires two distinct but related functions: DNA binding and the activation of transcription [23]. DNA binding is central to the activity of p53, as the majority of cancer-associated *TP53* mutations are found in the DNA binding domain and disrupt interactions between p53 and its cognate DNA response element [24]. The nucleotide preferences within a p53 response element/motif (p53RE) have been rigorously validated using multiple low- and high-throughput methodologies, including EMSA, SELEX, and ChIP-seq [2, 25, 26]. DNA sequence variation across p53REs only partially explains observations of differential p53 binding and transcriptional activity *in vivo* [27, 28]. Post-translational modification of p53 can serve as a key regulator of DNA binding affinity, through modulation of DNA-binding cooperativity, p53 stability, and cofactor recruitment, which then influences the selectivity of p53 target gene regulation [29–34]. The kinetics and stoichiometry of p53: DNA binding are also important for target gene transcription and plays crucial roles in p53-dependent cell fate [35–38]. Balancing p53 activities through direct regulation of

the p53 protein is critical in organismal-level phenotypes like tumour suppression and ageing [17, 18, 39, 40]. Beyond DNA sequence and p53 protein dynamics, other genomic features can play a key role in p53 DNA binding and p53-dependent activities. Recent work suggests that p53: DNA interactions differ between cell types [41], and are influenced by features such as local and long-distance chromatin structure [42], histone modifications [43], and DNA methylation [44]. Differential p53 binding is itself linked to differential p53 transcriptional targets across cell types, suggesting that p53 activity can be modulated by differences in cell- or condition-specific chromatin structure.

Unsurprisingly, data resources examining various aspects of p53 regulation have been published given the importance of p53 in human biology and cancer genetics. The *TP53* website and the National Cancer Institute *TP53* Database are invaluable resources describing the history of p53 research, available cell lines used to study p53 biology and *TP53* gene status, and clinically observed *TP53* mutations [45, 46]. The focus of these databases is on the genetics of the *TP53* gene and putative mutations that are linked to cancer or other human disorders. The TargetGeneRegulation Database links p53-regulated gene expression from microarray and RNA-seq-style experiments to experimentally observed p53 genomic occupancy data [47]. Using a vote-counting meta-analysis approach [48], TargetGeneRegulation Database is a strong resource for those interested in asking whether their gene of interest may be under control of p53 or regulated by the cell cycle. Resources such as the Cistrome Data Browser or the ReMap Project reanalyse publicly available ChIP-seq experiments using standardized data processing approaches to define the breadth of TF genome binding [49, 50], including for p53. Relatedly, p53 itself has been the target of more in-depth meta-analyses of genomic occupancy and potential gene regulation [41, 51], focusing primarily on the analysis of ChIP-seq data and the relation of observed p53 binding sites to gene regulation. Each of these resources begins with either experimentally observed p53 binding or p53-dependent gene expression changes, which are limited to certain cell lines and treatment conditions. Thus, available resources for p53 biology are highly valuable for the community, but are limited to documenting *TP53* mutation status or creating meta-analyses allowing users to validate potential p53-mediated regulation of their favourite gene or pathway.

In this manuscript, we present a novel data resource for exploring p53 biology. The p53motifDB integrates predictions of p53RE motifs within the human genome with multiple genetic, epigenetic, and functional datasets. The key difference between this data resource and other p53-focused databases or large-scale analyses of ChIP-seq data is the focus on *potential* p53 binding sites. The goals of the p53motifDB are (1) to act as a comprehensive resource for quickly obtaining key information about both putative and validated p53 binding events in the human genome and (2) to serve as a tool to generate novel hypotheses about p53-dependent transcriptional regulation. The entire database is available as standalone tables for integration into machine learning paradigms or as a precompiled SQLite database. Users can also access the p53motifDB via a local or web-based Shiny app. We provide multiple examples of the utility of integrating these datasets by confirming previous observations in the field and by generating and testing novel hypotheses.

Materials and methods

Data sources and processing

Table 1 contains the file name, data type, source/download location, and relevant publication (if applicable) for each dataset used to construct the p53motifDB. Putative p53 motifs in the GRCh38/hg38 genome assembly were identified using two separate methodologies. We first used p53 motifs defined by JASPAR (matrix ID MA0106.3) from an HT-SELEX-derived position weight matrix (PWM) [2]. We then used a PWM derived from experimental ChIP-seq data and the GRCh38/hg38 assembly as the inputs for running the scanMotifsGenomewide.pl script from HOMER (v.5) [52]. The output of the two p53 motif datasets was merged using bedTools (v.2.29.2) to create a non-redundant master list of 412 586 individual motifs [53]. The liftOver tool and corresponding chain files were used to identify corresponding loci in the hg19 and T2T/hs1 human genome assemblies and to identify syntenic locations in the mm10 and MM39 mouse genome assemblies (ucsc-liftover, bioconda, v.469) [54]. Unless otherwise stated, p53 motif locations were integrated with datasets available in interval formats using bedTools. Data extraction from bigWig file types was performed using deepTools (v.3.4.1) [55]. The dbSNP156 dataset in VCF file format was queried using bcftools/samtools (v.1.7) [56, 57]. Data were parsed from program-specific output files and joined into database-compatible tables using dplyr (v.1.1.4) and tidyr (v.1.3.1) from the tidyverse package (v.2.0.0) implemented in R (version 3.6.0) [58, 59].

Data access, tutorial development, and Shiny web app construction

Processed datasets were added into an SQLite3 relational database built using the DBI (v.1.2.3) and RSQLite (v.2.3.7) packages within R (v.3.6.0). The database primary key is the hg38 genome coordinate for each motif in the 'chr_start_stop' format, called the 'unique_id' in each table. SQLite3 database files are available to download from Zenodo (10.5281/zenodo.13351805). We also built a Shiny app under R (v.3.6.0) that allows users to access the database via an online portal (<https://p53motifdb.its.albany.edu/>). The Shiny app was built using shinythemes (v.1.2.0), shinyBS (v.0.61.0), and shinyjs (v.2.1.0). The app and code can be downloaded for offline use from Zenodo or can be deployed locally via a pre-built Docker image [60, 61]. All raw data tables used to construct the SQLite3 database and the Shiny app can also be downloaded directly from Zenodo under CC-BY data restrictions. A website containing a tutorial and database summary statistics was built using the bookdown package (v.0.42) and is available at <https://masammons.github.io/p53motifDB/>.

Results

Characteristics and selection of p53REs in the human genome

The content of the p53motifDB is focused on providing key genetic and regulatory information on non-redundant genomic locations that contain putative p53 binding locations based on experimentally validated p53 binding preferences. We thus began our investigation by identifying potential p53RE motifs in the human genome. The canonical p53RE motif contains two half-sites separated by a six-nucleotide

Table 1. Location of Data Sources

Name	File type	Location	Publication DOI (if applicable)
ABC	TSV	ftp://ftp.broadinstitute.org/outgoing/linRNA/ABC/AllPredictions.AvgHiC.ABC0.015.minus150.ForABCPaperV3.txt.gz	10.1038/s41586-021-03446-x
Aerts ChIP-seq Metaanalysis	BED	Supplemental Data from Verfaillie <i>et al.</i>	10.1101/gr.204149.116
BlackList	BED	https://github.com/Boyle-Lab/BlackList/blob/master/lists/hg38-blacklist.v2.bed.gz	
Capture MicroC	TSV	This paper (GEO:GSE275042)	
chromHMM fullstack	BED/TSV	https://public.hoffman2.idre.ucla.edu/ernst/UUKP7/hg38lift_genome_100_browser.bed.gz	10.1186/s13059-021-02572-z
ClinVar	VCF	https://ftp.ncbi.nlm.nih.gov/pub/clinvar/vcf_GRCh38/clinvar.vcf.gz	
dbSNP156	VCF	https://ftp.ncbi.nlm.nih.gov/snp/archive/b156/VCF/GCF_000001405.40.gz	
ENCODE cCRE	BED	https://downloads.wenglab.org/V3/GRCh38-cCREs.bed	
ENCODE DHS	BED	https://hgdownload.soe.ucsc.edu/goldenPath/hg38/database/wgEncodeRegDnaseClustered.txt.gz	
ENSEMBL v.111	GFF	Accessed via biomaRt package in R	
Fischer ChIP-seq Metaanalysis	BED	Supplemental Data from Riege <i>et al.</i>	10.7554/eLife.63266
GeneHancer v.5.17	TSV	Proprietary academic license (contact GeneCards.org for non-commercial access)	
hg38	FASTA	https://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/hg38.fa.gz	
hg38 to hg19	chain File	https://hgdownload.soe.ucsc.edu/goldenPath/hg38/liftOver/hg38ToHg19.over.chain.gz	
hg38 to hs1/T2T	chain File	https://hgdownload.gi.ucsc.edu/hubs/GCA/009/914/755/GCA_009914755.4/liftOver/hg38-chm13v2.over.chain.gz	
hg38 to mm10	chain File	https://hgdownload.soe.ucsc.edu/goldenPath/hg38/liftOver/hg38ToMm10.over.chain.gz	
hg38 to mm39	chain File	https://hgdownload.soe.ucsc.edu/goldenPath/hg38/liftOver/hg38ToMm39.over.chain.gz	
hs1/T2T	FASTA	https://hgdownload.soe.ucsc.edu/goldenPath/hs1/bigZips/hs1.fa.gz	
JASPAR p53 motif locations	BED/TSV	http://expdata.cmmi.ubc.ca/JASPAR/downloads/UCSC_tracks/2022/hg38/MAA0106.3.tsv.gz	10.1093/nar/gkab1113
mm10	FASTA	https://hgdownload.soe.ucsc.edu/goldenPath/mm10/bigZips/mm10.fa.gz	
MM39	FASTA	https://hgdownload.soe.ucsc.edu/goldenPath/mm39/bigZips/mm39.fa.gz	
Nguyen ChIP-seq Metaanalysis	BED	Supplemental Data from Nguyen <i>et al.</i>	10.1093/nar/gky720
phastCons	bigwig	http://hgdownload.cse.ucsc.edu/goldenpath/hg38/phastCons100way/hg38.phastCons100way.bw	
PhyloP	bigwig	http://hgdownload.cse.ucsc.edu/goldenpath/hg38/phyloP100way/hg38.phyloP100way.bw	
Promoter-Capture HiC	TSV	Supplemental Data from Serra <i>et al.</i>	10.1038/s41467-024-46666-1
ReMap 2022 ChIP-seq	BED	https://remap.univ-amu.fr/storage/remap2022/hg38/MACS2/TF/TP53/remap2022_TP53_nr_macs2_hg38_v1.0.bed.gz	10.1093/nar/gkab996
Metaanalysis		https://hgdownload.soe.ucsc.edu/goldenPath/hg38/database/rmsk.txt.gz	
RMSK	TSV		

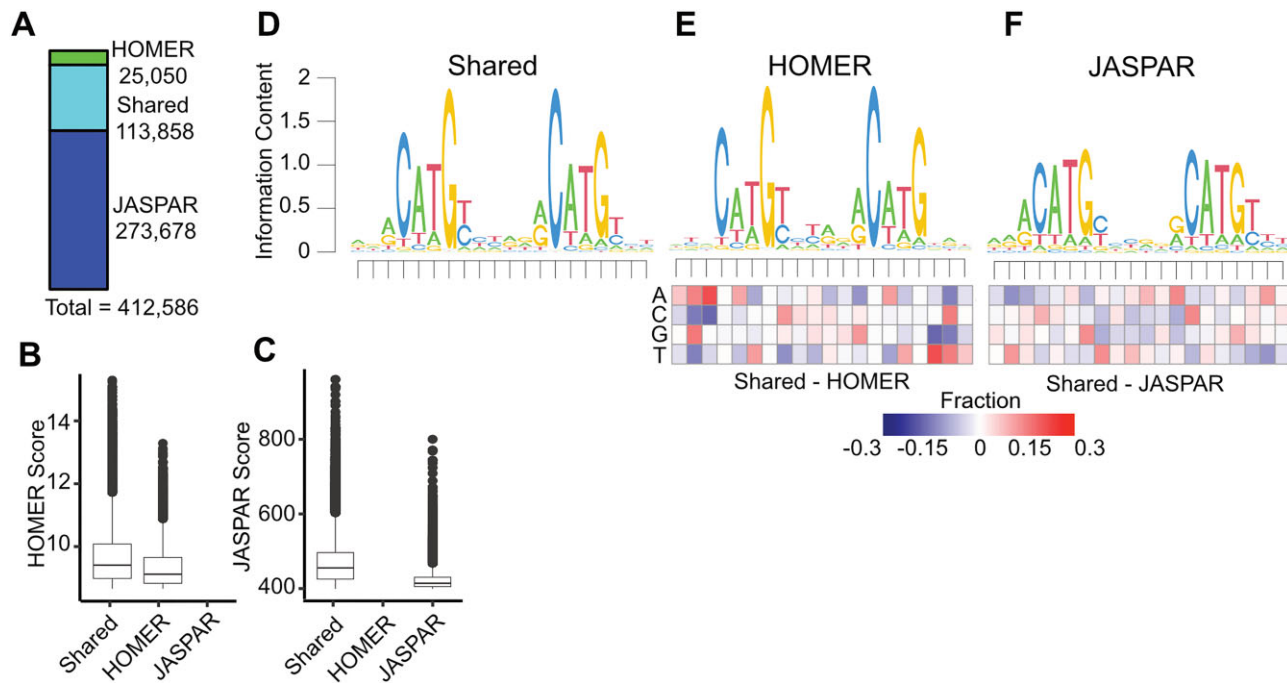


Figure 1. Selection and characteristics of p53 response elements (p53RE) in the hg38 human genome assembly. (A) The number of p53RE identified using nucleotide binding preferences from HOMER (ChIP-seq-derived) compared to JASPAR (*in vitro* SELEX-derived). The distribution of method-specific scoring for p53RE either found with both methods or with only a single method for (B) HOMER or (C) JASPAR. (D) A seqLogo representation of nucleotide preferences for p53RE identified from both HOMER and JASPAR binding preferences. seqLogo representation of nucleotide preferences for (E) HOMER-specific or (F) JASPAR-specific p53RE. Each heatmap represents the nucleotide frequency between the shared p53RE and either the HOMER- or JASPAR-specific p53RE.

spacer [25]. This concept of identifying potential p53RE in the genome is not new, and multiple algorithms and approaches have been previously used to identify genomic p53RE [26, 62–64]. All approaches used either *in vitro* or ChIP-style data to derive motif likelihood scores based on p53 affinity for DNA sequences. Recent meta-analyses of dozens of ChIP-seq datasets suggest that most *in vivo* p53 binding occurs at ‘canonical’ motifs [51, 41]. Therefore, we did not consider p53 $\frac{1}{2}$ or $\frac{3}{4}$ sites or potential motifs with spacers, which would greatly increase the number of potential p53RE with only a marginal gain in *bona fide* binding events [25]. We allowed for overlapping p53RE, with the 3′ half-site of one p53RE serving as the 5′ half-site of another. Our approach depended on updated p53RE PWM from high-throughput *in vitro* (SELEX) and *in vivo* (ChIP-seq) and two separate software approaches. We first extracted pre-compiled p53 motif locations from hg38 in the JASPAR 2022 database [65], which uses SELEX data to define the PWM. These motifs were merged with putative p53RE identified using scanMotifGenomeWide.pl from HOMER (v.4.10.4) [52] and its built-in p53 position weight matrix (p53.motif) derived from ChIP-seq data.

We then standardized the length of all p53RE from HOMER and JASPAR to 20 nucleotides to account for the different lengths of the underlying PWM used to call motif locations, and converted all motif locations to the plus-strand. This resulted in a total of 412 586 non-redundant p53RE motifs on the canonical somatic [1–22], sex (X/Y), and mitochondrial (MT) chromosomes (Fig. 1A). Motifs present on unassigned scaffolds were removed from further analysis. Interestingly, only 113 858 motifs were identified by both JASPAR and HOMER (Fig. 1A). The JASPAR database contained substantially more unique p53RE (273 678) than those identi-

fied by HOMER (25 050) (Fig. 1A). p53REs identified in both datasets are scored higher (i.e. more closely aligned to the consensus) than those identified via HOMER (Fig. 1B) or JASPAR (Fig. 1C) alone. p53RE identified uniquely by HOMER had nearly identical preferences for the key C/G residues within the half-sites to shared sites (Fig. 1D versus Fig. 1E), whereas unique JASPAR elements had reduced prevalence for these crucial nucleotides (Fig. 1D versus Fig. 1F). HOMER-specific motifs were depleted for purines at the 5′ end and pyrimidines at the 3′ end relative to p53RE shared across methods (Fig. 1E). The nucleotide frequency differences between HOMER and JASPAR (Fig. 1E versus Fig. 1F) may reflect either differences in the statistical methods used to call motifs or differences in experimental approaches used to derive the underlying PWM (*in vitro* SELEX versus chromatin immunoprecipitation). We also marked p53RE that contains potential cytosine methylation sites (6.9%, 28 485/412 586). Biochemical and structural evidence indicates that DNA methylation within a p53RE can strongly influence p53 binding and transcriptional regulatory activity [44]. Methylation at CG dinucleotides within a p53RE will be cell- and context-dependent and should be considered as part of any comprehensive analysis of p53 biochemical activity on DNA.

Integration with other reference genome assemblies and repetitive elements

To extend the utility of these datasets, we examined whether the standardized 412 586 p53RE motifs and their locations were also present in two additional human reference genome assemblies and in two commonly used mouse reference genomes using the UCSC liftOver tool [54, 66]. Pro-

Table 2. p53 Motifs Across Human and Mouse Genome Assemblies

Date of release	Assembly name	Common name	Present	Absent
February 2009	GRCh19	hg19	411 273	1313
January 2022	T2T CHM13v2.0	hs1	409 542	3044
December 2011	GRCm38	mm10	88 624	323 962
June 2020	GRCm39	mm39	89 074	323 512

viding information from additional mouse and human reference genomes will allow users to more quickly integrate p53-centric information with the wealth of pre-parsed data from prior assemblies (hg19 and mm10). Further, the inclusion of genomic locations from the recently completed full telomere-to-telomere human hs1 genome will provide some level of futureproofing as new datasets are provided with updated genomic coordinates [67]. As expected, nearly all hg38 p53RE locations are present in the hg19 genome assembly and are preserved in the most recent hs1/T2T complete human genome assembly (>99%, Table 2).

We also report p53RE intersections with two additional key pieces of genome assembly data. First, we incorporated genome blacklist information into p53motifDB, which covers 6163 locations in the hg38 genome. These locations are generally excluded from most genomic analyses due to pervasive issues in mapping next-generation sequencing reads from experiments like ChIP-seq and RNA-seq to a reference genome assembly [68]. We also provide information on the location of p53RE relative to repetitive DNA elements. Repetitive DNA elements, like those derived from transposable elements and viral DNA, significantly contribute to difficulty in genome assembly and next-generation sequencing-based analyses of genome function [69]. Repetitive elements also influence gene regulatory networks [70, 71], serve as platforms for TF binding [72], and can have *cis*-regulatory element activity [73]. p53 is known to bind to and regulate activity of repetitive elements [74–76]. We provide information regarding the presence of p53RE within repetitive DNA elements using the RepeatMasker compendium [77, 78]. Over 60% of identified p53RE (254 075/412 586) are found in repetitive elements. Our database also includes information about the repeat DNA itself, including the repeat class and family.

Mouse models have been extensively used to identify foundational tenets of p53 biology. Tumour suppressor activity is conserved across vertebrates, but p53 biochemical activity and specific genes regulated by p53 differ between mouse and human [79, 80]. Functional TF binding events controlling gene expression often vary even between closely related organisms [81]. Prior work identified a limited number of p53RE with conserved sequence and binding across a range of vertebrates with a focus on whether these p53RE might be functionally linked to gene expression [82, 83]. We extend this work here by examining synteny between p53RE locations in the human genome and two commonly used mouse genome assemblies. As expected, the majority of p53RE from hg38 are unique to human assemblies and are not syntenic within the mouse genome. We found ~21% of p53RE locations are syntenic in the MM10 (88 624) or MM39 (89 074) mouse reference genomes. Within the database, we provide the genomic coordinates from the hg38 assembly and, if applicable, the corresponding coordinates in the hg19, hs1/T2T, MM10, or MM39 genomes. We have also provided average phyloP and phastCon vertebrate conservation scores for the 20 bp p53RE [84,

85] allowing the end user to rapidly identify human p53 motifs with different evolutionary conservation constraints. Ultimately, the conservation and synteny data presented here are not meant to replace a more comprehensive and focused evolutionary analysis of p53RE, but can be used in combination with other datasets to quickly generate and test hypotheses related to the p53 gene regulatory network.

Integration and analysis of p53 genomic binding datasets

The p53motifDB also contains information on whether individual p53 motifs are locations for experimentally observed p53 binding as determined by four recent meta-analyses of dozens of p53 chromatin immunoprecipitation-coupled sequencing (ChIP-seq) datasets [50, 51, 41, 86]. Overall, 37 628 p53RE have evidence of p53 binding from ChIP-seq data in at least one cell line or condition (Fig. 2). There are 3833 p53RE with evidence of experimental p53 binding across all four meta-analyses (Fig. 2). Each meta-analysis used a different combination of p53 ChIP-seq datasets, different genome mapping methods, and statistical approaches for calling positive binding events, resulting in a range of binding events across meta-analyses. Each meta-analysis only considered p53 binding events that met some statistical or enrichment cut-off based on the peak calling tool of choice. Except for the ReMAP dataset, each meta-analysis also only considered a p53 binding event as legitimate if observed across multiple experiments or cell lines. The Verfaillie dataset characterized both ‘Strong’ and ‘Weak’ p53 binding sites from 16 individual p53 ChIP-seq datasets from a total of seven different cell types based on inferred affinity from read pileup data. We combined both Strong and Weak categories into one group to simplify our analysis, yielding 4922 p53RE that overlap p53 binding events. Of note, this analysis considered the smallest number of datasets. Nguyen *et al.* used two different metrics for determining p53 binding based on the number of datasets in which a given putative binding site reached a predetermined cutoff value [41]. We use the nomenclature ‘ubiquitous’ and ‘validated’ from the original publication to maintain consistency across analyses. These analyses yielded a set of highly stringent, ‘ubiquitous’ p53 binding events (1288 sites, identified in ≥ 20 independent datasets) and a less-stringent set of ‘validated’ (12 048 sites, ≥ 2 independent datasets) p53 binding sites. We used only the less-stringent set of p53 binding events in this analysis. Riege *et al.* used an experimental cut-off of five independent datasets, resulting in 7804 p53RE with observed p53 occupancy. Finally, the ReMap2022 dataset considered 126 p53 ChIP-seq datasets, and included all binding events that occurred in any individual experiment. In total, 3833 p53RE are occupied by p53 across the Nguyen (validated), Verfaillie, Riege, and Remap datasets. An additional 3654 p53RE have p53 binding in at least three of these datasets (Fig. 2). The Remap dataset contains 24 322 p53 ChIP-seq binding events not found in any other meta-analysis, likely

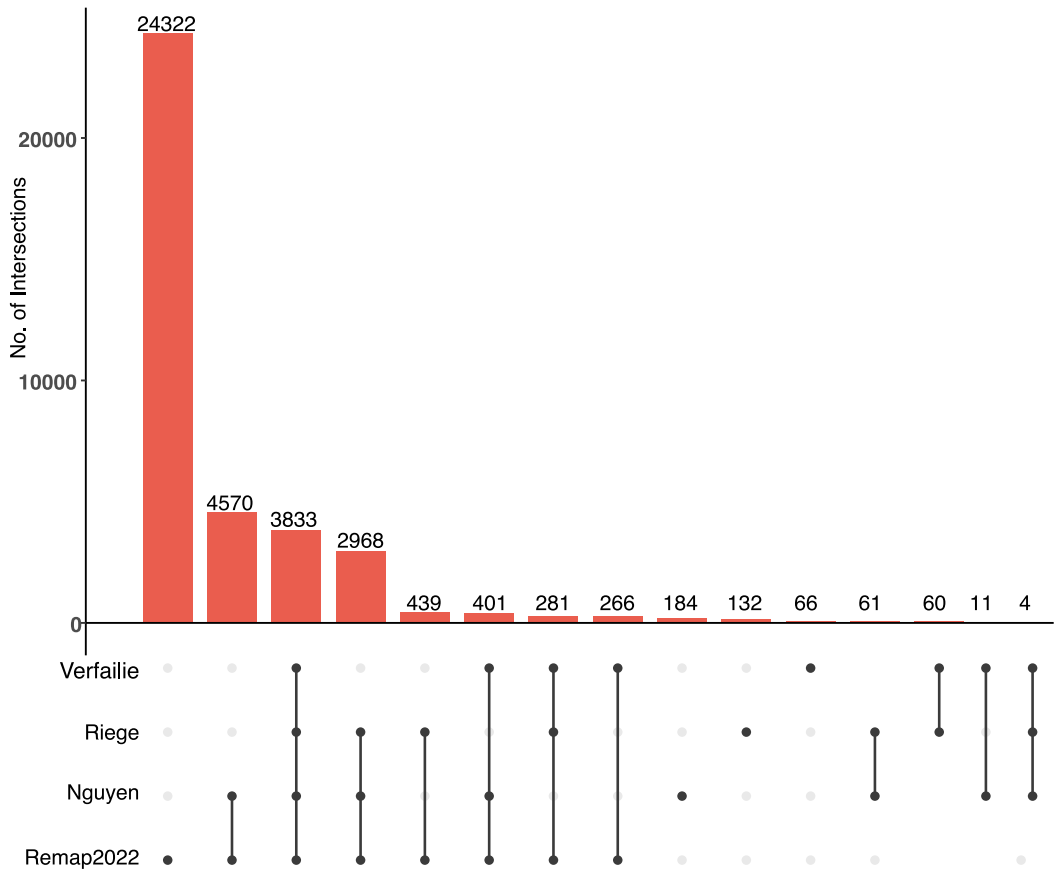


Figure 2. Overlap between identified p53RE and experimentally validated p53 binding events across four comprehensive meta-analyses. Upset-style plot of overlap between p53RE and experimentally derived p53 binding events (via ChIP-seq) for four separate meta-analyses.

due to the combination of a larger number of datasets considered and a reduced threshold for calling a binding event relative to other datasets. Differences between datasets also represent tissue and cell-specific binding events, including the Nguyen dataset, which identified cell lineage-specificity to p53 binding [41]. Cell type-specificity for p53 binding has been further confirmed in additional studies [41, 87, 88]. The majority of human p53 genomic binding data comes from transformed cell lines; thus, we expect that the number of p53RE with observed p53 binding events will likely increase as additional cell lineages, primary cells and tissues, and new stimulus paradigms are considered. The full dataset includes genomic coordinates for each observed ChIP-seq peak and experimental observation frequency data in supplemental database tables for ReMap and Riege meta-analyses, and included information about the cell lines used for analysis from the ReMap2022 dataset.

Integration of gene features and gene expression changes with p53RE locations

p53 is a TF regulating a diverse set of target genes under multiple physiological conditions, including in response to DNA damage. We included multiple datasets to help researchers explore relationships between the location of p53REs and gene expression. Most p53REs are found within ENSEMBL gene models (intragenic, 61%, 253 727/412 586) compared to outside genes (intergenic, 39%, 158 859/412 586). When applicable, the HGNC gene symbol is also provided (i.e. *TP53*). The dataset can be filtered by HGNC symbol, ENSEMBL gene ID

(ENSG*), or ENSEMBL transcript ID (ENST*) when those values are available. The frequency of all intragenic p53RE is nearly identical to that of bound p53RE (61% versus 63%) when considering the 3833 p53RE that are occupied by p53 across all four ChIP-seq meta-analyses (Fig. 2). These data suggest that there is no innate preference for actual p53 binding within genes other than the natural distribution of p53RE across the genome. We also report distance to the nearest transcriptional start site (TSS) both up and downstream of the p53RE, and include ENSEMBL gene ID, transcript ID, and HGNC gene symbols associated with those TSS. We also integrated gene expression fold-change data and gene expression ‘scores’ from a prior meta-analysis of p53’s effect on transcription [47]. Thus, researchers can rapidly search the database for genes and transcripts whose activity is known to be influenced by p53 and identify the nearest set of p53RE and p53 binding events. We note that linking TF binding to gene expression changes is a complex process that requires more than correlation between proximity of two elements, but that this information can be used to generate hypotheses for downstream functional validation.

Integration of local chromatin and regulatory element information with p53RE

Local chromatin structure strongly influences TF binding and activity [4, 5]. Multiple publications have examined how these features can influence p53 interactions with the genome and subsequent transcriptional activation [86, 89, 90]. We thus incorporated multiple summary datasets describing chro-

matin modification states and chromatin accessibility into the database to allow further exploration of these potential linkages. DNase I hypersensitive sites (DHSs) are genomic regions susceptible to enzymatic cleavage, thus reflecting ‘open’, nucleosome-free locations [91]. We incorporated ENCODE DHS cluster data, which integrates millions of DHS locations across 125 cell types and conditions [92, 93]. These data include the cell type where the DHS was observed and an accessibility score based on the maximum observed DNase signal. Overall, 86 851 p53REs (21%) are found within DHS. The ENCODE Project also produced a series of analyses focused on predicting regions of the genome that function as candidate *cis*-regulatory elements (cCREs) based on local chromatin structure and gene distance [93]. The analysis produced five cCRE classes: promoter (high DNase and H3K4me3 signal, <200 bp from TSS), promoter-like (high DNase and H3K4me3 signal, >200 bp from TSS), proximal enhancer (high DNase and H3K27ac, low H3K4me3, <2 kb from TSS), distal enhancer (high DNase and H3K27ac, low H3K4me3, >2 kb from TSS), and CTCF/CCCTC-binding factor occupancy (high CTCF and DNase, low H3K4me3 and H3K27ac). A total of 48 177 p53RE (11.67% of total) overlap an ENCODE cCRE. Within that 48 177, 2.5% are within the promoter group, 2.0% are promoter-like, 12.8% are at proximal enhancer elements, 79.6% at distal enhancers, and 3.0% overlap CTCF-binding elements.

Chromatin accessibility (DNase hypersensitivity/ATAC-seq) can be combined with multiple chromatin modification states to predict transcription regulatory activity. The ENCODE and Epigenome Roadmap projects produced thousands of histone modification datasets across a vast array of cell and tissue types [13, 94]. While these data provided incredible insight into connections between the epigenome, TFs, and gene expression, the breadth and scope of these datasets are often unwieldy for non-computational biologists. ChromHMM uses hidden Markov models to incorporate multiple types of chromatin modification and accessibility data to summarize the likely functional chromatin state of a given genomic location across different cell types [95–97]. An updated fullstack chromHMM segmentation was designed to create a ‘single universal’ chromatin-based annotation for different segments of the genome [98]. Thus, we integrated the ‘fullstack’ chromHMM genome segmentation dataset in order to simplify analysis of local chromatin context surrounding p53RE. The dataset includes 100 detailed chromatin-state annotations, which can be collapsed into 16 broader chromatin state ‘groups’ as previously defined [98]. Nearly all p53REs are located within a chromHMM genome annotation segment (99%, 409 061/412 586), providing researchers with a simple description of the most likely local chromatin environment for a given p53RE.

Incorporation of three-dimensional chromatin interactions with p53RE locations

Assigning individual TF binding events to specific gene expression changes can be difficult without additional genetic evidence, such as the deletion of a specific TF binding site or regulatory element. Connecting TF binding and gene activity is difficult even when it occurs at a gene promoter, and is even more difficult for distal TF binding events [99]. Chromatin looping has emerged as a key driver of transcription and partially explains how distal regulatory elements can control gene

expression over long distances [100–102]. Here, we integrate four datasets interrogating three-dimensional chromatin interactions with p53RE locations. The GeneHancer and Activity-By-Contact (ABC) datasets incorporate chromatin conformation assays with gene expression and regulatory element-associated activity data, such as chromatin modifications, to call enhancer–promoter interactions across a range of cell types and conditions [103, 104]. Promoter capture approaches use *in situ* proximity-based ligation approaches to identify distal regions interacting with targeted promoters. This database incorporates recently published promoter-capture Hi-C focused on p53-mediated gene regulation from HCT116 cells [42] and newly generated promoter-capture Micro-C (GEO GSE275042) from MCF10A mammary epithelial cell lines treated with either DMSO or the p53 activating drug etoposide. Two important caveats should be considered when assessing Hi-C-based approaches and their application to p53 biology. First, the experimental and statistical bias in Hi-C analysis is biased towards identification of distal interactions. Since many p53REs are within or near promoters, we expect reduced sensitivity in detecting TSS-proximal interactions. Second, the resolution of the technique (normally 1kb+) means it can be difficult to determine specific p53RE or p53 binding events actually participating in a looping event. To streamline the analysis and standardize the data for use in this database, biological and technical replicates, treatment conditions, and time points for the promoter-capture experiments were merged to create a single ‘snapshot’ of potential 3D interactions involving p53RE. Across all four datasets, a total of 206 575/412 586 p53RE (50.06%) are found within at least one chromatin looping interaction. Localization within a chromatin loop anchor does not necessarily imply transcriptional regulatory potential, just as the absence from a loop does not mean a given p53RE is not functional. These data allow users of the p53motifDB to quickly identify and generate hypotheses regarding experimentally observed chromatin loops containing potential or validated p53 binding events without having to parse or re-analyse large chromatin conformation capture datasets.

Human genetic variation at p53RE

Reference genomes do not represent the vast diversity of genomic space in the human population. Genetic variation in cCRE and TF binding sites can directly influence biochemical activities on DNA [105]. Genetic variation in TF binding sites can also lead to a range of additional biological effects, ranging from changes in gene expression up to observable phenotypic traits with the potential to alter health and lifespan. The DNA sequence-based determinants of p53 binding and activity are well studied [25, 27], but are still under active investigation [106]. Variation in p53RE motifs can have strong functional effects *in vivo*. Users may be interested in whether p53 binding and gene regulation might be affected by natural human genetic variation in p53REs. We therefore incorporated genetic variation data from the dbSNP156 build, which includes millions of single nucleotide polymorphisms and small insertion and deletion variants [57]. Over 99% of p53REs have at least one reported genetic variant from dbSNP156 (408 872/412 586), with a median of nine variants per p53RE. p53RE almost universally contain single nucleotide variants, but deletions (33%, 136 826/412 586) or insertions (7.8%, 32 261/412 586) of at least one nucleotide

are also relatively common. We have also included variation reported in the ClinVar database, which contains variants with known or predicted clinical significance [107, 108]. For example, rs4590952 (G > A) is a common single nucleotide polymorphism (SNP) in a p53RE that reduces p53 binding, alters transcription of *KITLG*, and is associated with increased cancer risk [109]. The master p53motifDB table contains information only about whether a given p53RE contains a ClinVar or other SNP, but that information can be used to query a secondary table containing information about the specific reference and alternate allele for each SNP in the database.

The data integrated into this current database specifically focus on variation within the canonical 20 bp p53 motif. Variation outside of this core motif may influence p53 binding and activity, such as variation within other TF binding motifs required for regulatory element activity [110, 111]. Further, our analysis does not consider genetic variation that could result in *de novo* p53 motifs and gene regulation as has been previously observed [112, 113]. The rapid increase in genome resequencing projects focused on human genetic diversity, personal genomes, and cancer genomics should allow these and other types of advanced analyses in the future.

Accessing and querying the p53motifDB

We have provided multiple methods for users to interface with the data found in the p53motifDB. Our goal was not to produce a comprehensive analysis platform; rather, we aimed to provide end users with multiple, flexible methods of interacting with this dataset. The tables were built and analysed via R using *tidyverse*-style data methods, but users can easily import the underlying data into their preferred data analytics tools and pipelines. All processed datasets are available in tabular format via Zenodo. A pre-compiled *SQLite* relational database is also fully available for download via Zenodo, which can be analysed offline using standard query methods. Power users, or those who wish to perform more advanced relational queries, are encouraged to download either the SQL database or the tabular-format datasets for use in custom data analysis pipelines.

We also integrated the datasets into a Shiny app, which can be queried online (<https://p53motifDB.its.albany.edu>) or downloaded and used locally from our Zenodo repository (10.5281/zenodo.13351805). Users can find tutorial information on the use of the Shiny app at <https://masammons.github.io/p53motifDB/>. This website also contains database summary statistics in Chapter 7, many of which are described above. The interface for the Shiny app allows users to initially filter the p53motifDB based on categorical information in the main table. Drop-down boxes with predefined choices are available for data types with a limited number of choices, such as when searching by chromosome or searching for p53RE with experimentally observed p53 binding. Input boxes can be used for other types of data, such as when querying by gene names. By default, p53REs are filtered by whether there is experimental evidence of p53 binding in any of the four p53 ChIP-seq meta-analyses representing hundreds of assays. Data can be further filtered, and results can be exported to a local tab-delimited file for offline analysis. Users can export either the primary table information or can query additional data sources using the filtered p53RE locations. As an example, a user might filter p53RE that are found in an ENCODE DHS cluster and where there is experimental p53 binding ev-

idence, but then want to know more information about the cell types where those DHS are found. Users can quickly export this advanced information from accessory tables, such as DHS clusters or genetic variation, based on filtered p53RE locations by selecting one of a series of buttons in the Shiny app. All underlying database information and code for building and deploying the Shiny app are available on the Zenodo repository (10.5281/zenodo.13351805).

Use cases for the p53motifDB

Enrichment of p53RE and p53 binding in repetitive elements

Repetitive genomic elements encompass satellite and microsatellite DNA as well as those derived from viral and mobile genetic elements. Repetitive elements are key regulators of gene expression and genome stability, and their misregulation can lead to increased DNA damage, inflammation, and ultimately an increase in cancer and age-related diseases [114]. p53 inhibits specific repetitive elements through direct and indirect methods [115–119]. Further, the contribution of repetitive and mobile genetic elements to the distribution of p53 motifs in the genome is well documented [76]. We identified 254 075 p53RE motifs (61.6%) within repetitive DNA elements using the RepeatMasker UCSC Genome Browser [77, 78]. Almost half of the p53RE within repetitive elements are contained in LINE elements (43.08%) (Fig. 3A). SINE (32.97%), LTR (13.29%), DNA (5.31%), and simple (4.73%) repeat elements are also substantial contributors of p53RE motifs (Fig. 3A). All other repetitive element classes add up to ~10% of the total. Consistent with the primate specificity of SINE and LINE expansion [120], p53REs found within the MM39 mouse reference genome are considerably less likely than expected to be found in repetitive elements compared to regions lacking synteny (Fig. 3B, $P < 2.2e-16$, *Pearson's Chi-squared with Yates' continuity correction*).

We then asked whether experimentally determined p53 binding sites are enriched in different repeat element classes, as has been previously demonstrated [74, 76]. We focused on the p53 binding events from the Riege meta-analysis, as it included a large number of datasets and the threshold for calling a p53 binding event was stringent (i.e. p53 binding in at least five separate datasets) [51]. A slight majority of p53 binding events occur within repetitive elements (53.9%), but this is less than expected based on the frequency of repeat-associated p53RE motifs in the genome (61.6%, $P < 2.2e-16$, *Pearson's Chi-squared with Yates' continuity correction*). This suggests actual p53 binding is slightly biased against p53RE within repetitive elements. Compared to p53RE distribution in the genome, p53 binding is significantly enriched within LTR elements (Fig. 3D). ERV1 and ERV1-MaLR elements are preferentially bound relative to the distribution of p53RE genomewide (Fig. 3E). Consistent with prior observations on a more limited set of p53 binding sites, LTR-associated p53 binding frequently occurs within MLT1H, LTR10C/E, and MER61C/E elements [90]. In contrast, p53 binding to LINEs is less frequent than expected (Fig. 3D), with L1 LINEs particularly depleted for p53 occupancy and primarily contributing to this observation (Fig. 3E).

Local chromatin states at p53RE

The local chromatin environment at a given p53 binding site can provide context clues as to whether a particular binding site may function as an enhancer or promoter [42, 89, 90].

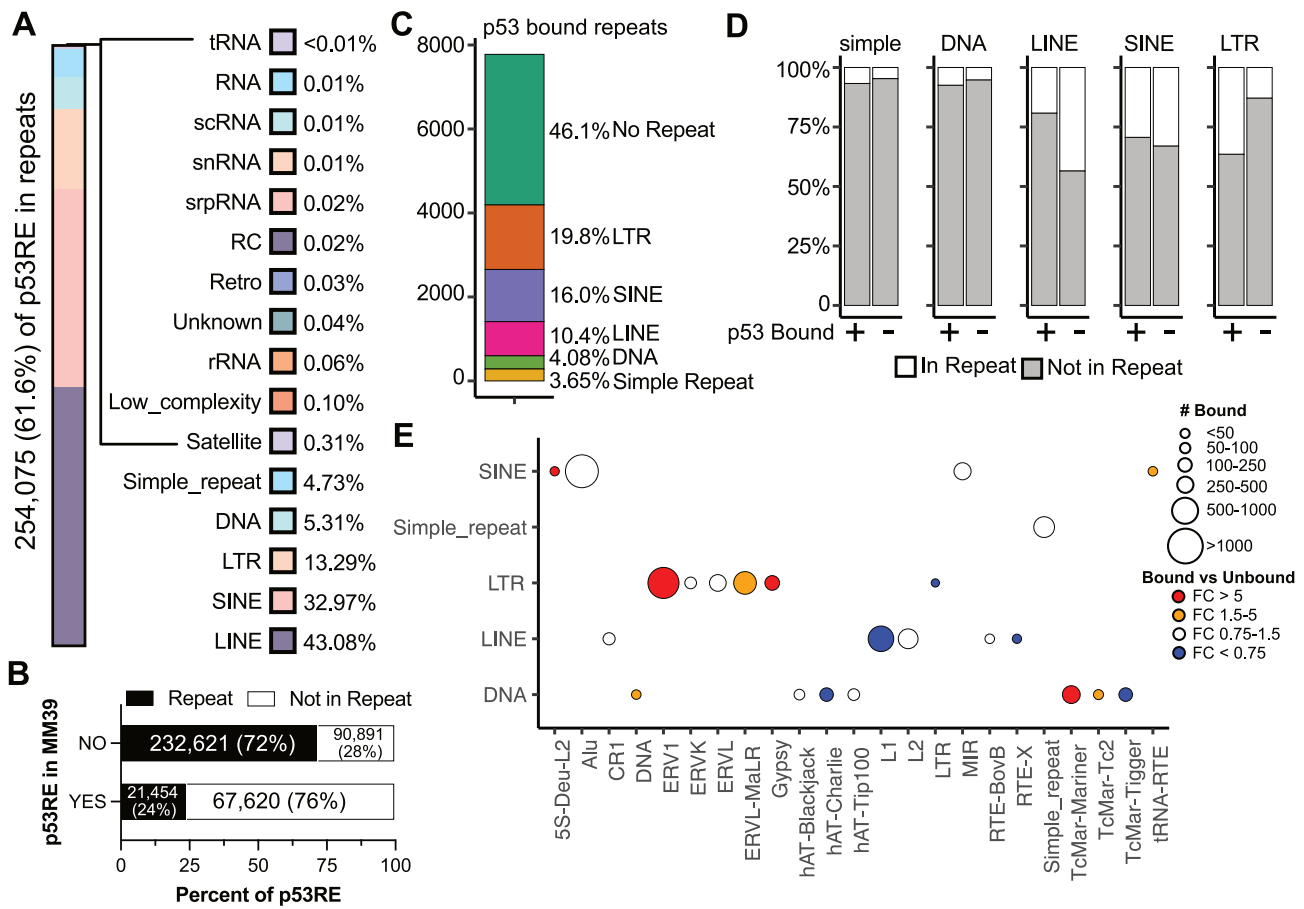


Figure 3. Characterization of p53RE and their localization within repetitive genomic elements. (A) The percentage of p53RE found within each class of repeat element. A total of 254,075/412,586 p53RE (61.6%) are found within repeat elements. (B) The distribution of p53RE with synteny to MM39 and found in repeat elements. (C) The percentage of p53RE with experimentally validated p53 binding [51] and their distribution within repeat elements. (D) The distribution of p53-bound versus p53-unbound p53RE and their localization within LINE, SINE, LTR, DNA, and simple repeats, which represent the five most common repeat types with p53RE. (E) Number of p53-bound p53RE within each class and type of repeat element for the five most common repeat types and their enrichment versus p53-unbound p53RE.

Chromatin structure and histone modification patterns at p53 binding sites can differ across cell types, which may inform differential activity [5, 41]. We assessed the average chromatin status of experimentally determined p53 binding sites from the Riege meta-analysis to determine whether p53 is enriched in any specific chromatin locations. First, we assessed the genome-wide enrichment of p53RE in each of the 16 chromHMM summary groups. The distribution of p53RE almost perfectly mirrors the percentage of the genome covered by each chromHMM segment (Fig. 4A).

We then examined the relationship between chromHMM summary states and p53RE that were bound by p53 versus those that remained unbound. Our analysis suggests that p53 binding is enriched in chromatin contexts reflecting gene regulatory elements, like promoters, enhancers, and other open chromatin regions (Fig. 4B). Nearly 40% of all p53 binding events are found in regions defined as active enhancers, an enrichment of 3.75-fold versus the distribution of unbound p53RE. Promoter and TSS-associated binding was enriched nearly seven-fold. p53 binding is also depleted in regions with H3K27me3/Polycomb-associated heterochromatin, but not in heterochromatin regions enriched with H3K9me3 (Fig. 4B). p53 binding is also depleted strongly from quiescent regions of the genome, which lack chromatin modifications indicative

of regulated activity. Combined with the presence of p53 in H3K9me3-marked heterochromatin and the high enrichment in accessible, gene regulatory regions, the lack of p53 binding in quiescent chromatin may be a function of p53's documented pioneer TF activity [89, 121–125].

We wanted to further demonstrate the utility of integrating p53RE data with local chromatin context by assessing enrichment of p53 binding within sub-categories of chromHMM segments. The fullstack chromHMM dataset contains 100 different chromatin states categorized primarily by enrichment of specific chromatin modifications, cell type, and other genomic features, like gene distance. We graphed these data by the % of total p53 binding sites in that chromatin segment in rank order and then colour-coded each based on enrichment or depletion status (p53-bound/unbound ratio) (Fig. 4C). Almost 20% of p53 binding events are located in EnhA12 and EnhA13 chromHMM states, and p53 binding is enriched over two-fold relative to unbound p53RE (Fig. 4C and D). EnhA12/EnhA13 represent epithelial-specific enhancer regions, consistent with expanded p53 binding and activity in epithelial cell types [51, 87, 126]. The PromF4 subclass, representing a chromatin state found downstream of transcriptional start sites, is the most enriched (Fig. 4D). This may reflect the well-documented preference of p53 binding within the first in-

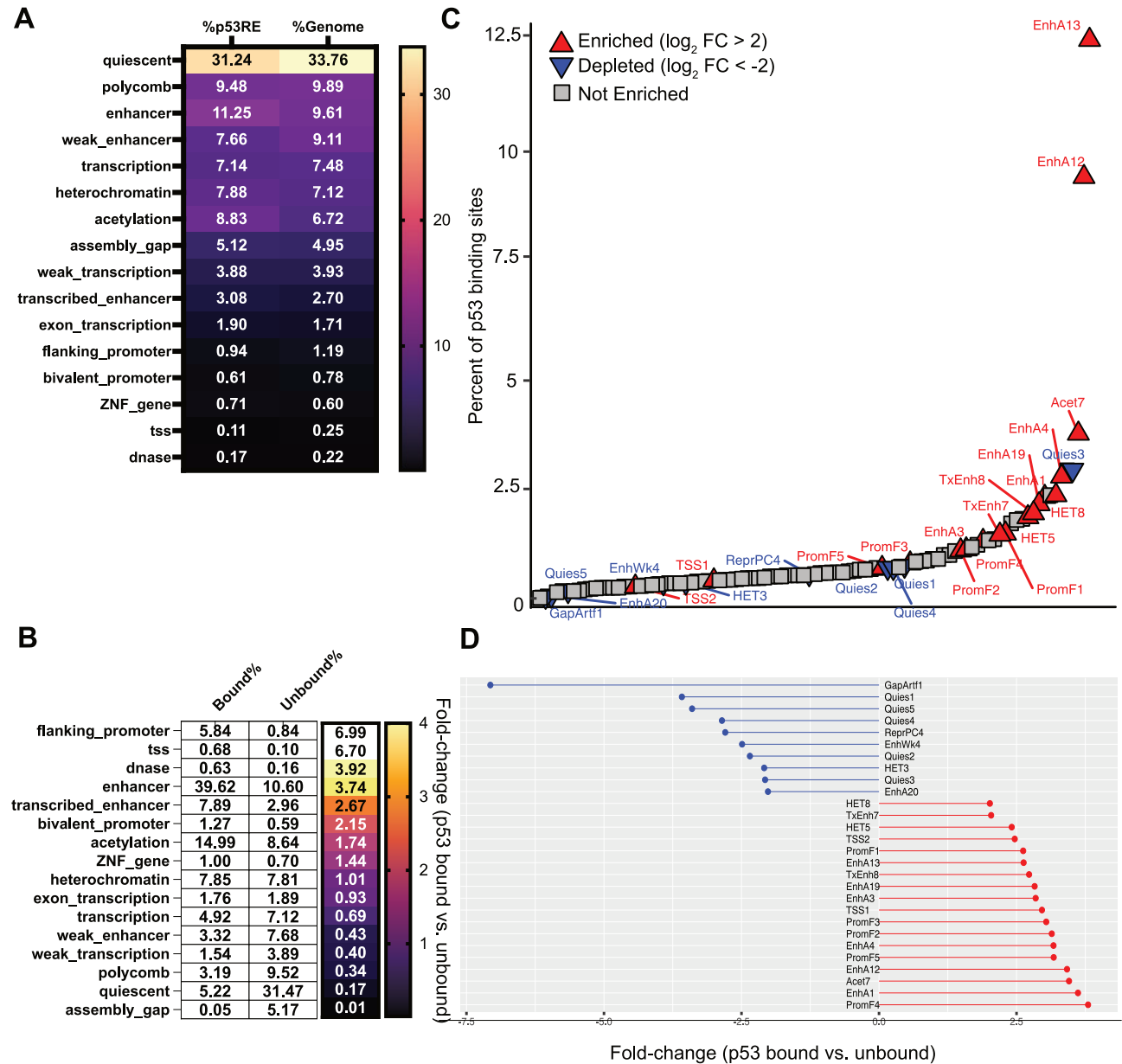


Figure 4. Analysis of p53RE and their localization within chromHMM genome segments. (A) The percentage of p53RE found within each of the 16 chromHMM fullstack summary groups compared to the percentage of the human genome covered by that chromHMM feature [98]. (B) The percentage (white boxes) and fold-change enrichment (colour scale) of bound p53RE versus unbound p53RE found within each of the 16 chromHMM fullstack summary groups. p53 binding data are from the Riege *et al.* meta-analysis [51]. (C) The rank order of p53-bound p53RE (x-axis) versus the % of p53-bound p53RE that are found within each of the 100 detailed chromHMM fullstack genomic segments. Arrows represent an enrichment/fold-change of greater than 2 (red upward arrows) or less than -2 (blue downward arrows) for p53-bound versus p53-unbound p53RE. (D) The actual enrichment/fold-change for p53-bound versus unbound p53RE for each chromHMM fullstack segment labelled in (C).

tron of target genes [15, 127, 128–130]. We also observe counterintuitive enrichment of p53 within the HET5 and HET8 subclasses of H3K9me3-heterochromatin (Fig. 4C and D). Although p53 binding was not depleted at H3K9me3-enriched regions like we observed at Polycomb-regulated H3K27me3-enriched regions (Fig. 4B), we did not expect specific enrichment of any subclass of heterochromatin. Interestingly, HET5 and HET8 represent chromatin states found at LTR repetitive elements, which we previously demonstrated support higher-than-expected p53 binding (Fig. 3D). p53 binding is also depleted from the HET3 subclass (Fig. 4C and D), which represents LINE-associated chromatin, consistent

with our prior observation of LINE-mediated depletion of p53 binding (Fig. 3D). Taken together, analysis of p53RE and p53 binding enrichment across chromHMM summary and subclass datasets provides further support for prior observations in literature, but also can be used to generate novel hypotheses about p53 activity relative to cell type and chromatin state.

Discussion

In summary, we constructed a database resource containing local genetic, regulatory, chromatin, and variation informa-

tion at putative motifs for the p53 TF in the human genome. These data are accessible in a web-facing application where the end user can query the database without prior knowledge of structured query language. The entire dataset is also available for download in multiple offline formats allowing more advanced users to analyse the data using tools of their choice. This resource provides users a simple, yet powerful, method for retrieving information on validated p53 binding locations or new putative sites identified in their own laboratories. Future development of this database can incorporate additional genetic and epigenetic data sources as they become available. For example, the rapid rate of genome sequencing brought forth through more accessible and inexpensive long-read technologies will assuredly grow the availability of genome variation data. Improved functionality, including built-in data analysis and graphing tools and tighter integration with genome browsers or other datasets, can also be added as the needs of the end-user change over time. This type of motif-centric database can also be extended to other TF motifs. We envision this database as a tool that can be used to generate new hypotheses about how chromatin structure, genetic variation, and evolutionary constraint might affect p53 activity across cell types and across organisms with implications in cancer, ageing, stem cell, and developmental biology.

Acknowledgements

We would like to thank all of the labs and research programs that generated the raw and analysed datasets and metadata used in the construction of this database. We would also like to thank the University at Albany Information Technology Services (ITS) and the Advanced Research Computing Cluster (ARCC) for computing resources, storage, and hosting of the Shiny app.

Conflict of interest: None declared.

Funding

Funding for this publication was provided by the National Institutes of Health (R35 GM138120 to M.A.S.).

References

- Lambert SA, Jolma A, Campitelli LF *et al.* The human transcription factors. *Cell* 2018;172:650–65. <https://doi.org/10.1016/j.cell.2018.01.029>
- Jolma A, Yan J, Whittington T *et al.* DNA-binding specificities of human transcription factors. *Cell* 2013;152:327–39. <https://doi.org/10.1016/j.cell.2012.12.009>
- Inukai S, Kock KH, Bulyk ML. Transcription factor-DNA binding: beyond binding site motifs. *Curr Opin Genet Dev* 2017;43:110–19. <https://doi.org/10.1016/j.gde.2017.02.007>
- Neph S, Vierstra J, Stergachis AB *et al.* An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* 2012;489:83–90. <https://doi.org/10.1038/nature11212>
- Thurman RE, Rynes E, Humbert R *et al.* The accessible chromatin landscape of the human genome. *Nature* 2012;489:75–82. <https://doi.org/10.1038/nature11232>
- Wang J, Zhuang J, Iyer S *et al.* Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res* 2012;22:1798–812. <https://doi.org/10.1101/gr.139105.112>
- Zaret KS, Carroll JS. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev* 2011;25:2227–41.
- Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. *Cell Res* 2011;21:381–95. <https://doi.org/10.1038/cr.2011.22>
- Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. *Nat Rev Genet* 2019;20:207–20. <https://doi.org/10.1038/s41576-018-0089-8>
- Millán-Zambrano G, Burton A, Bannister AJ *et al.* Histone post-translational modifications—cause and consequence of genome function. *Nat Rev Genet* 2022;23:563–80. <https://doi.org/10.1038/s41576-022-00468-7>
- Dorschner MO, Hawrylycz M, Humbert R *et al.* High-throughput localization of functional elements by quantitative chromatin profiling. *Nat Methods* 2004;1:219–25. <https://doi.org/10.1038/nmeth721>
- Bernstein BE, Kamal M, Lindblad-Toh K *et al.* Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* 2005;120:169–81. <https://doi.org/10.1016/j.cell.2005.01.001>
- Kundaje A, Meuleman W, Ernst J *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* 2015;518:317–30. <https://doi.org/10.1038/nature14248>
- Vierstra J, Lazar J, Sandstrom R *et al.* Global reference mapping of human transcription factor footprints. *Nature* 2020;583:729–36. <https://doi.org/10.1038/s41586-020-2528-x>
- Riley T, Sontag E, Chen P *et al.* Transcriptional control of human p53-regulated genes. *Nat Rev Mol Cell Biol* 2008;9:402–12. <https://doi.org/10.1038/nrm2395>
- Sullivan KD, Galbraith MD, Andrysik Z *et al.* Mechanisms of transcriptional regulation by p53. *Cell Death Differ* 2018;25:133–43. <https://doi.org/10.1038/cdd.2017.174>
- Wang M, Attardi LD. A balancing act: p53 activity from tumor suppression to pathology and therapeutic implications. *Annu Rev Pathol* 2022;17:205–26. <https://doi.org/10.1146/annurev-pathol-042320-025840>
- Horikawa I. Balancing and differentiating p53 activities toward longevity and no cancer? *Cancer Res* 2020;80:5164–65. <https://doi.org/10.1158/0008-5472.CAN-20-3080>
- Rodier F, Campisi J, Bhaumik D. Two faces of p53: aging and tumor suppression. *Nucleic Acids Res* 2007;35:7475–84. <https://doi.org/10.1093/nar/gkm744>
- Biteau B, Jasper H. It's all about balance: p53 and aging. *Aging* 2009;1:884–86. <https://doi.org/10.18632/aging.100102>
- Chakravarti A, Thirumanne HN, Brown S *et al.* *Drosophila* p53 isoforms have overlapping and distinct functions in germline genome integrity and oocyte quality control. *eLife* 2022;11:e61389. <https://doi.org/10.7554/eLife.61389>
- Tasnim S, Kelleher ES. p53 is required for female germline stem cell maintenance in P-element hybrid dysgenesis. *Dev Biol* 2018;434:215–20. <https://doi.org/10.1016/j.ydbio.2017.12.021>
- Kastenhuber ER, Lowe SW. Putting p53 in context. *Cell* 2017;170:1062–78. <https://doi.org/10.1016/j.cell.2017.08.028>
- Hollstein M, Sidransky D, Vogelstein B *et al.* p53 mutations in human cancers. *Science* 1991;253:49–53. <https://doi.org/10.1126/science.1905840>
- el-Deiry WS, Kern SE, Pietenpol JA *et al.* Definition of a consensus binding site for p53. *Nat Genet* 1992;1:45–49. <https://doi.org/10.1038/ng0492-45>
- Wei C-L, Wu Q, Vega VB *et al.* A global map of p53 transcription-factor binding sites in the human genome. *Cell* 2006;124:207–19. <https://doi.org/10.1016/j.cell.2005.10.043>
- Wang B, Xiao Z, Ren EC. Redefining the p53 response element. *Proc Natl Acad Sci* 2009;106:14373–78. <https://doi.org/10.1073/pnas.0903284106>
- Vyas P, Beno I, Xi Z *et al.* Diverse p53/DNA binding modes expand the repertoire of p53 response elements. *Proc Natl Acad Sci* 2017;114:10624–29. <https://doi.org/10.1073/pnas.1618005114>
- Laptenko O, Shiff I, Freed-Pastor W *et al.* The p53 C terminus controls site-specific DNA binding and promotes structural changes within the central DNA binding domain.

- Mol Cell* 2015;57:1034–46. <https://doi.org/10.1016/j.molcel.2015.02.015>
30. Hamard P-J, Lukin DJ, Manfredi JJ. p53 basic C terminus regulates p53 functions through DNA binding modulation of subset of target genes*. *J Biol Chem* 2012;287:22397–407. <https://doi.org/10.1074/jbc.M111.331298>
 31. Timofeev O, Koch L, Niederau C et al. Phosphorylation control of p53 DNA-binding cooperativity balances tumorigenesis and aging. *Cancer Res* 2020;80:5231–44. <https://doi.org/10.1158/0008-5472.CAN-20-2002>
 32. Xia Z, Kon N, Gu AP et al. Deciphering the acetylation code of p53 in transcription regulation and tumor suppression. *Oncogene* 2022;41:3039–50. <https://doi.org/10.1038/s41388-022-02331-9>
 33. Kruse J-P, Gu W. SnapShot: p53 posttranslational modifications. *Cell* 2008;133:930–930.e1. <https://doi.org/10.1016/j.cell.2008.05.020>
 34. Timofeev O, Schlereth K, Wanzel M et al. p53 DNA binding cooperativity is essential for apoptosis and tumor suppression *in vivo*. *Cell Rep* 2013;3:1512–25. <https://doi.org/10.1016/j.celrep.2013.04.008>
 35. Lu D, Faizi M, Drown B et al. Temporal regulation of gene expression through integration of p53 dynamics and modifications. *Sci Adv* 2024;10:eadp2229. <https://doi.org/10.1126/sciadv.adp2229>
 36. Stewart-Ornstein J, Iwamoto Y, Miller MA et al. p53 dynamics vary between tissues and are linked with radiation sensitivity. *Nat Commun* 2021;12:898. <https://doi.org/10.1038/s41467-021-21145-z>
 37. Hafner A, Bulyk ML, Jambhekar A et al. The multiple mechanisms that regulate p53 activity and cell fate. *Nat Rev Mol Cell Biol* 2019;20:199–210. <https://doi.org/10.1038/s41580-019-0110-x>
 38. Purvis JE, Karhohs KW, Mock C et al. p53 dynamics control cell fate. *Science* 2012;336:1440–44. <https://doi.org/10.1126/science.1218351>
 39. Liu Y, Stockwell BR, Jiang X et al. p53-regulated non-apoptotic cell death pathways and their relevance in cancer and other diseases. *Nat Rev Mol Cell Biol* 2025;26:600–14.
 40. Liu Y, Su Z, Tavana O et al. Understanding the complexity of p53 in a new era of tumor suppression. *Cancer Cell* 2024;42:946–67.
 41. Nguyen T-AT, Grimm SA, Bushel PR et al. Revealing a human p53 universe. *Nucleic Acids Res* 2018;46:8153–67. <https://doi.org/10.1093/nar/gky720>
 42. Serra F, Nieto-Aliseda A, Fanlo-Escudero L et al. p53 rapidly restructures 3D chromatin organization to trigger a transcriptional response. *Nat Commun* 2024;15:2821. <https://doi.org/10.1038/s41467-024-46666-1>
 43. Isbel L, Iskar M, Durdu S et al. Readout of histone methylation by Trim24 locally restricts chromatin opening by p53. *Nat Struct Mol Biol* 2023;30:948–57.
 44. Kribelbauer JF, Laptenko O, Chen S et al. Quantitative analysis of the DNA methylation sensitivity of transcription factor complexes. *Cell Rep* 2017;19:2383–95. <https://doi.org/10.1016/j.celrep.2017.05.069>
 45. Leroy B, Fournier JL, Ishioka C et al. The TP53 website: an integrative resource centre for the TP53 mutation database and TP53 mutant analysis. *Nucleic Acids Res* 2013;41:D962–69. <https://doi.org/10.1093/nar/gks1033>
 46. de Andrade KC, Lee EE, Tookmanian EM et al. The TP53 Database: transition from the International Agency for Research on Cancer to the US National Cancer Institute. *Cell Death Differ* 2022;29:1071–73. <https://doi.org/10.1038/s41418-022-00976-3>
 47. Fischer M, Schwarz R, Riege K et al. TargetGeneReg 2.0: a comprehensive web-atlas for p53, p63, and cell cycle-dependent gene regulation. *NAR Cancer* 2022;4:zcac009. <https://doi.org/10.1093/narcan/zcac009>
 48. Fischer M, Hoffmann S. Synthesizing genome regulation data with vote-counting. *Trends Genet* 2022;38:1208–16. <https://doi.org/10.1016/j.tig.2022.06.012>
 49. Mei S, Qin Q, Wu Q et al. Cistrome Data Browser: a data portal for ChIP-Seq and chromatin accessibility data in human and mouse. *Nucleic Acids Res* 2017;45:D658–62. <https://doi.org/10.1093/nar/gkw983>
 50. Hammal F, de Langen P, Bergon A et al. ReMap 2022: a database of human, mouse, *Drosophila* and arabidopsis regulatory regions from an integrative analysis of DNA-binding sequencing experiments. *Nucleic Acids Res* 2022;50:D316–25. <https://doi.org/10.1093/nar/gkab996>
 51. Riege K, Kretzmer H, Sahm A et al. Dissecting the DNA binding landscape and gene regulatory network of p63 and p53. *eLife* 2020;9:e63266. <https://doi.org/10.7554/eLife.63266>
 52. Heinz S, Benner C, Spann N et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 2010;38:576–89. <https://doi.org/10.1016/j.molcel.2010.05.004>
 53. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 2010;26:841–42. <https://doi.org/10.1093/bioinformatics/btq033>
 54. Hinrichs AS, Karolchik D, Baertsch R et al. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res* 2006;34:D590–98. <https://doi.org/10.1093/nar/gkj144>
 55. Ramírez F, Ryan DP, Grüning B et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 2016;44:W160–65.
 56. Sherry ST, Ward M, Sirotkin K. dbSNP—database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res* 1999;9:677–79. <https://doi.org/10.1101/gr.9.8.677>
 57. Sherry ST, Ward M-H, Kholodov M et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001;29:308–11. <https://doi.org/10.1093/nar/29.1.308>
 58. Wickham H, Vaughan D, Girlich M. tidy: Tidy Messy Data. R. package version 1.3.1. 2025. <https://tidyverse.org>
 59. Wickham H, François R, Henry L et al. dplyr: A Grammar of Data Manipulation. R package version 1.1.4. 2018. <https://dplyr.tidyverse.org>
 60. Merkel D. Docker: lightweight linux containers for consistent development and deployment. *Linux J* 2014;2014:2.
 61. Chang W, Cheng J, Allaire JJ et al. shiny: Web Application Framework for R. 2024. <https://shiny.posit.co>
 62. Smeenk L, van Heeringen SJ, Koeppl M et al. Characterization of genome-wide p53-binding sites upon stress response. *Nucleic Acids Res* 2008;36:3639–54. <https://doi.org/10.1093/nar/gkn232>
 63. Veprintsev DB, Fersht AR. Algorithm for prediction of tumour suppressor p53 affinity for binding sites in DNA. *Nucleic Acids Res* 2008;36:1589–98. <https://doi.org/10.1093/nar/gkm1040>
 64. Cui F, Sirotin MV, Zhurkin VB. Impact of Alu repeats on the evolution of human p53 binding sites. *Biol Direct* 2011;6:2. <https://doi.org/10.1186/1745-6150-6-2>
 65. Castro-Mondragon JA, Riudavets-Puig R, Rauluseviciute I et al. JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. *Nucleic Acids Res* 2022;50:D165–73. <https://doi.org/10.1093/nar/gkab1113>
 66. Kuhn RM, Haussler D, Kent WJ. The UCSC genome browser and associated tools. *Briefings Bioinf* 2013;14:144–61. <https://doi.org/10.1093/bib/bbs038>
 67. Nurk S, Koren S, Rhie A et al. The complete sequence of a human genome. *Science* 2022;376:44–53. <https://doi.org/10.1126/science.abcj6987>
 68. Amemiya HM, Kundaje A, Boyle AP. The ENCODE blacklist: identification of problematic regions of the genome. *Sci Rep* 2019;9:9354. <https://doi.org/10.1038/s41598-019-45839-z>

69. Treangen TJ, Salzberg SL. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet* 2012;13:36–46. <https://doi.org/10.1038/nrg3117>
70. Feschotte C. Transposable elements and the evolution of regulatory networks. *Nat Rev Genet* 2008;9:397–405. <https://doi.org/10.1038/nrg2337>
71. Sundaram V, Cheng Y, Ma Z *et al.* Widespread contribution of transposable elements to the innovation of gene regulatory networks. *Genome Res* 2014;24:1963–76. <https://doi.org/10.1101/gr.168872.113>
72. Bourque G, Leong B, Vega VB *et al.* Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res* 2008;18:1752–62. <https://doi.org/10.1101/gr.080663.108>
73. Chuong EB, Elde NC, Feschotte C. Regulatory activities of transposable elements: from conflicts to benefits. *Nat Rev Genet* 2017;18:71–86. <https://doi.org/10.1038/nrg.2016.139>
74. Wang T, Zeng J, Lowe CB *et al.* Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proc Natl Acad Sci* 2007;104:18613–18. <https://doi.org/10.1073/pnas.0703637104>
75. Harris CR, DeWan A, Zupnick A *et al.* p53 responsive elements in human retrotransposons. *Oncogene* 2009;28:3857–65. <https://doi.org/10.1038/onc.2009.246>
76. Tiwari B, Jones AE, Abrams JM. Transposons, p53 and genome security. *Trends Genet* 2018;34:846–55. <https://doi.org/10.1016/j.tig.2018.08.003>
77. Jurka J. Repbase Update: a database and an electronic journal of repetitive elements. *Trends Genet* 2000;16:418–20. [https://doi.org/10.1016/S0168-9525\(00\)02093-X](https://doi.org/10.1016/S0168-9525(00)02093-X)
78. Smit A, Hubley R, Green P. *RepeatMasker Open-4.0*. 2013. <http://www.repeatmasker.org>
79. Stewart-Ornstein J, Cheng HW(J), Lahav G. Conservation and divergence of p53 oscillation dynamics across species. *Cell Syst* 2017;5:410–417.e4. <https://doi.org/10.1016/j.cels.2017.09.012>
80. Fischer M. Conservation and divergence of the p53 gene regulatory network between mice and humans. *Oncogene* 2019;38:4095–109. <https://doi.org/10.1038/s41388-019-0706-9>
81. Tanay A, Regev A, Shamir R. Conservation and evolvability in regulatory networks: the evolution of ribosomal regulation in yeast. *Proc Natl Acad Sci* 2005;102:7203–7208. <https://doi.org/10.1073/pnas.0502521102>
82. Horvath MM, Wang X, Resnick MA *et al.* Divergent evolution of human p53 binding sites: cell cycle versus apoptosis. *PLoS Genet* 2007;3:e127. <https://doi.org/10.1371/journal.pgen.0030127>
83. Jegga AG, Inga A, Menendez D *et al.* Functional evolution of the p53 regulatory network through its target response elements. *Proc Natl Acad Sci* 2008;105:944–49. <https://doi.org/10.1073/pnas.0704694105>
84. Pollard KS, Hubisz MJ, Rosenbloom KR *et al.* Detection of non-neutral substitution rates on mammalian phylogenies. *Genome Res* 2010;20:110–21. <https://doi.org/10.1101/gr.097857.109>
85. Siepel A, Bejerano G, Pedersen JS *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res* 2005;15:1034–50. <https://doi.org/10.1101/gr.3715005>
86. Verfaillie A, Svetlichnyy D, Imrichova H *et al.* Multiplex enhancer-reporter assays uncover unsophisticated TP53 enhancer logic. *Genome Res* 2016;26:882–95. <https://doi.org/10.1101/gr.204149.116>
87. Karsli Uzunbas G, Ahmed F, Sammons MA. Control of p53-dependent transcription and enhancer activity by the p53 family member p63. *J Biol Chem* 2019;294:10720–36. <https://doi.org/10.1074/jbc.RA119.007965>
88. Hafner A, Kublo L, Tsabar M *et al.* Identification of universal and cell-type specific p53 DNA binding. *BMC Molecular and Cell Biology* 2020;21:5. <https://doi.org/10.1186/s12860-020-00251-8>
89. Sammons MA, Zhu J, Drake AM *et al.* TP53 engagement with the genome occurs in distinct local chromatin environments via pioneer factor activity. *Genome Res* 2015;25:179–88. <https://doi.org/10.1101/gr.181883.114>
90. Su D, Wang X, Campbell MR *et al.* Interactions of chromatin context, binding site sequence content, and sequence evolution in stress-induced p53 occupancy and transactivation. *PLoS Genet* 2015;11:e1004885. <https://doi.org/10.1371/journal.pgen.1004885>
91. Vierstra J, Stamatoyannopoulos JA. Genomic footprinting. *Nat Methods* 2016;13:213–21. <https://doi.org/10.1038/nmeth.3768>
92. Sheffield NC, Thurman RE, Song L *et al.* Patterns of regulatory activity across diverse human cell types predict tissue identity, transcription factor binding, and long-range interactions. *Genome Res* 2013;23:777–88. <https://doi.org/10.1101/gr.152140.112>
93. Moore JE, Purcaro MJ, Pratt HE *et al.* Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature* 2020;583:699–710.
94. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489:57–74. <https://doi.org/10.1038/nature11247>
95. Ernst J, Kheradpour P, Mikkelsen TS *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* 2011;473:43–49. <https://doi.org/10.1038/nature09906>
96. Ernst J, Kellis M. Large-scale imputation of epigenomic datasets for systematic annotation of diverse human tissues. *Nat Biotechnol* 2015;33:364–76. <https://doi.org/10.1038/nbt.3157>
97. Ernst J, Kellis M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc* 2017;12:2478–92. <https://doi.org/10.1038/nprot.2017.124>
98. Vu H, Ernst J. Universal annotation of the human genome through integration of over a thousand epigenomic datasets. *Genome Biol* 2022;23:9. <https://doi.org/10.1186/s13059-021-02572-z>
99. Chen C-H, Zheng R, Tokheim C *et al.* Determinants of transcription factor regulatory range. *Nat Commun* 2020;11:2472. <https://doi.org/10.1038/s41467-020-16106-x>
100. Krivega I, Dean A. Enhancer and promoter interactions—long distance calls. *Curr Opin Genet Dev* 2012;22:79–85.
101. Kulaeva OI, Nizovtseva EV, Polikanov YS *et al.* Distant activation of transcription: mechanisms of enhancer action. *Mol Cell Biol* 2012;32:4892–97. <https://doi.org/10.1128/MCB.01127-12>
102. Dekker J, Marti-Renom MA, Mirny LA. Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. *Nat Rev Genet* 2013;14:390–403. <https://doi.org/10.1038/nrg3454>
103. Fishilevich S, Nudel R, Rappaport N *et al.* GeneHancer: genome-wide integration of enhancers and target genes in GeneCards. *Database* 2017;2017:bax028. <https://doi.org/10.1093/database/bax028>
104. Fulco CP, Nasser J, Jones TR *et al.* Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nat Genet* 2019;51:1664–69. <https://doi.org/10.1038/s41588-019-0538-0>
105. Spivakov M, Akhtar J, Kheradpour P *et al.* Analysis of variation at transcription factor binding sites in *Drosophila* and humans. *Genome Biol* 2012;13:R49. <https://doi.org/10.1186/gb-2012-13-r49>
106. Fischer M, Sammons MA. Determinants of p53 DNA binding, gene regulation, and cell fate decisions. *Cell Death Differ* 2024;31:836–43.
107. Landrum MJ, Lee JM, Benson M *et al.* ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* 2018;46:D1062–67. <https://doi.org/10.1093/nar/gkx1153>
108. Landrum MJ, Chitipiralla S, Brown GR *et al.* ClinVar: improvements to accessing data. *Nucleic Acids Res* 2020;48:D835–44. <https://doi.org/10.1093/nar/gkz972>
109. Zeron-Medina J, Wang X, Repapi E *et al.* A polymorphic p53 response element in KIT ligand influences cancer risk and has undergone natural selection. *Cell* 2013;155:410–22. <https://doi.org/10.1016/j.cell.2013.09.017>

110. Korkmaz G, Lopes R, Ugalde AP *et al.* Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat Biotechnol* 2016;34:192–98. <https://doi.org/10.1038/nbt.3450>
111. Catizone AN, Uzunbas GK, Celadova P *et al.* Locally acting transcription factors regulate p53-dependent *cis*-regulatory element activity. *Nucleic Acids Res* 2020;48:4195–213. <https://doi.org/10.1093/nar/gkaa147>
112. Menendez D, Inga A, Snipe J *et al.* A single-nucleotide polymorphism in a half-binding site creates p53 and estrogen receptor control of vascular endothelial growth factor receptor 1. *Mol Cell Biol* 2007;27:2590–600. <https://doi.org/10.1128/MCB.01742-06>
113. Menendez D, Snipe J, Marzec J *et al.* p53-responsive TLR8 SNP enhances human innate immune response to respiratory syncytial virus. *J Clin Invest* 2019;129:4875–84. <https://doi.org/10.1172/JCI128626>
114. Copley KE, Shorter J. Repetitive elements in aging and neurodegeneration. *Trends Genet* 2023;39:381–400. <https://doi.org/10.1016/j.tig.2023.02.008>
115. Ishak CA, Marhon SA, Tchakian N *et al.* Chronic viral mimicry induction following p53 loss promotes immune evasion. *Cancer Discov* 2025;15:793–817. <https://doi.org/10.1158/2159-8290.CD-24-0094>
116. Levine AJ, Ting DT, Greenbaum BD. P53 and the defenses against genome instability caused by transposons and repetitive elements. *Bioessays* 2016;38:508–13. <https://doi.org/10.1002/bies.201600031>
117. Wylie A, Jones AE, Abrams JM. p53 in the Game of Transposons. *Bioessays* 2016;38:1111–16. <https://doi.org/10.1002/bies.201600115>
118. Wylie A, Jones AE, D'Brot A., *et al.* p53 genes function to restrain mobile elements. *Genes Dev* 2016;30:64–77. <https://doi.org/10.1101/gad.266098.115>
119. Tutton S, Azzam GA, Stong N *et al.* Subtelomeric p53 binding prevents accumulation of DNA damage at human telomeres. *EMBO J* 2016;35:193–207. <https://doi.org/10.15252/embj.201490880>
120. Konkel MK, Walker JA, Batzer MA. LINEs and SINEs of primate evolution. *Evol Anthropol* 2010;19:236–49. <https://doi.org/10.1002/evan.20283>
121. Espinosa JM, Emerson BM. Transcriptional regulation by p53 through intrinsic DNA/chromatin binding and site-directed co-factor recruitment. *Mol Cell* 2001;8:57–69. [https://doi.org/10.1016/S1097-2765\(01\)00283-0](https://doi.org/10.1016/S1097-2765(01)00283-0)
122. Nili EL, Field Y, Lubling Y *et al.* p53 binds preferentially to genomic regions with high DNA-encoded nucleosome occupancy. *Genome Res* 2010;20:1361–68. <https://doi.org/10.1101/gr.103945.109>
123. Laptenko O, Beckerman R, Freulich E *et al.* p53 binding to nucleosomes within the p21 promoter *in vivo* leads to nucleosome loss and transcriptional activation. *Proc Natl Acad Sci* 2011;108:10385–90. <https://doi.org/10.1073/pnas.1105680108>
124. Younger ST, Rinn JL. p53 regulates enhancer accessibility and activity in response to DNA damage. *Nucleic Acids Res* 2017;45:9889–900. <https://doi.org/10.1093/nar/gkx577>
125. Yu X, Buck MJ. Defining TP53 pioneering capabilities with competitive nucleosome binding assays. *Genome Res* 2019;29:107–15. <https://doi.org/10.1101/gr.234104.117>
126. McDade SS, Patel D, Moran M *et al.* Genome-wide characterization reveals complex interplay between TP53 and TP63 in response to genotoxic stress. *Nucleic Acids Res* 2014;42:6270–85. <https://doi.org/10.1093/nar/gku299>
127. Zauberman A, Flusberg D, Haupt Y *et al.* A functional p53-responsive intronic promoter is contained within the human mdm2 gene. *Nucleic Acids Res* 1995;23:2584–92. <https://doi.org/10.1093/nar/23.14.2584>
128. Takimoto R, El-Deiry WS. Wild-type p53 transactivates the KILLER/DR5 gene through an intronic sequence-specific DNA-binding site. *Oncogene* 2000;19:1735–43. <https://doi.org/10.1038/sj.onc.1203489>
129. Thornborrow EC, Patel S, Mastropietro AE *et al.* A conserved intronic response element mediates direct p53-dependent transcriptional activation of both the human and murine bax genes. *Oncogene* 2002;21:990–99. <https://doi.org/10.1038/sj.onc.1205069>
130. Liu X, Yue P, Khuri FR *et al.* p53 Upregulates death receptor 4 expression through an intronic p53 binding site. *Cancer Res* 2004;64:5078–83. <https://doi.org/10.1158/0008-5472.CAN-04-1195>