# Stereotyped p53 binding tuned by chromatin accessibility

**Antonina Hafner[1,*,2], Galit Lahav[1,3] and Jacob Stewart-Ornstein[1*]**

[1]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA

*These authors contributed equally to this work

Corresponding author: jacob_stewart-ornstein@hms.harvard.edu

[2]Antonina Hafner: antonina@hms.harvard.edu

[3]Galit Lahav: galit_lahav@hms.harvard.edu

Keywords: p53, ChIPseq, Chromatin, DNA damage, Gene Expression

**Abstract**

Background

Transcription factors orchestrate the cellular response to stimuli by binding specifically to DNA and activating associated genes. In the DNA damage response, the tumor suppressor p53 regulates much of the transcriptional response and has been suggested to selectively regulate gene expression in different contexts. However, comparison between genome-wide studies shows a large overlap between p53 bound loci and thus questions this selectivity.

Results

To systematically assess the cell type specificity of p53, we directly measured its association with DNA in 12 p53 wild-type cell lines in response to ionizing radiation. We found that the vast majority of bound sites are occupied across all cells lines uniformly. Gene expression, on the other hand showed substantial variability between cell lines. The coherence of our dataset, allowed us to identify a small subset of binding sites that appeared in just one or a few cell lines. We found that intrinsic chromatin accessibility of a cell line explained these differential p53 binding preferences. Moreover, we were able to manipulate p53 binding by altering chromatin state.

Conclusion

Our results show the limited contribution of genomic sequence to p53 binding and suggest that in vivo factors including chromatin accessibility largely regulate its binding.

## Background

Transcription factors (TF) are a large family of proteins that can bind to DNA and induce or repress transcription of genes [1]. Substantial effort, notably through the ENCODE project, has gone into generating genome wide maps of TF-DNA associations, gene expression and chromatin states. These genome wide datasets have been used to identify gene regulatory networks that allow cell type and stimulus-specific gene expression and led to characterization of a handful of networks, as for example in the embryonic stem cells system [2, 3]. However, these studies have also revealed the complexity of gene regulation and the large space of combinatorial interactions between TFs [4, 5]. Identifying the mechanisms that allow tissue and stimulus-specific gene expression remains a challenge and an active field of research.

In this work, we explored cell type and stimulus specificity of the tumor suppressing transcription factor p53 at the level of DNA binding and regulation of gene expression. The major DNA damage response (DDR) regulator in mammals, p53 is necessary and sufficient to impose terminal cell fates on irradiated cells. Though ubiquitously expressed across human tissues, it remains unclear if p53 DNA binding is universal, which could be advantageous for a critical stress response system, or shaped by the chromatin and cofactor environment of different cell types and lineages. Context specific regulation of gene expression by p53 has been a long standing hypothesis in the p53 field, and implies that p53 can integrate information about cellular context and the type of stress to selectively activate some target genes versus others [6-9]. However, studies querying p53 binding genome wide in different human cell lines and upon different treatments have variously found strong agreement in p53 binding locations [10, 11] and activation of a core set of target genes [12]. Other studies have argued for unique cell type or transformation status specific p53 activity [13-15]. These studies compared pairs of cell lines or supplemented single cell line data with meta-analysis of published datasets, an approach that is powerful for identifying universal p53

binding sites, but has limits for detection of cell line specific binding patterns due to divergent experimental conditions across datasets.

To study how p53 binding varies across cell lines, we measured p53 DNA binding in 12 cell lines from different tissue types in response to a single treatment. Indeed, we took advantage of our previous characterization of these cell lines which showed a comparable acute response of p53 [16] in response to ionizing radiation (IR). By treating this panel of epithelial cell lines with a dose of IR sufficient to induce uniform p53 activation across cell lines, and measuring p53 binding at an early (2 hours) time-point we minimized secondary effects and focused on measuring the rapid and direct binding of p53. Our approach differs from the majority of p53 datasets in the literature, which use chemotherapy agents such as doxorubicin or the p53 activator Nutlin3A at later time-points of 6-12 hours. This coherent set of samples allowed us to rigorously explore the heterogeneity of p53 binding and identify the influence of universal genomic and cell line specific chromatin factors on p53 binding and gene expression.

## Results

### p53 binding across the genome is stereotyped between cell lines and treatments

To study how p53 binding varies across cell lines we treated a set of 12 cell lines expressing wild type p53 with ionizing radiation (IR; X-Ray 4Gy) to induce p53 activity and 2 hours later cross-linked and harvested each cell line for ChIP-Seq analysis. Well established p53 target genes showed robust binding in all 12 cell lines (Fig. 1A). Overall, by pooling data from all cell lines we confidently called ~9000 p53 ChIP peaks. *De novo* motif analysis identified a motif bearing strong similarity to previously identified p53 binding motifs as the strongest single motif and also found it to be centrally enriched within peaks (Fig. 1B).

The quantitative strength of p53 binding at each genomic locus was highly conserved across the 12 cell lines (Fig. 1C). Though no strong groups of cell lines appeared by eye, hierarchical clustering correctly sorted the cell lines by tissue of origin, with pairs of lung and kidney lines, and all five melanoma lines

clustering together (Fig. 1C). These p53 bound regions were also similar to other published datasets (average within dataset correlation 0.53, average correlation to external datasets 0.49; Fig. S1) and consistent with previous work on p53 binding locations [10]. It was previously suggested that cancer cell lines show a different p53 binding profile from non-cancerous cells [13]. We therefore compared the 12 cancer cell lines to an identically treated non-transformed line, RPE1. We found that p53 binding at identified sites in RPE1 cells in response to IR was highly correlated with our dataset in transformed cell lines (Fig. 1D; average R= 0.48 for correlation (RPE, Cancer Lines) vs 0.53 for correlation (Cancer, Cancer)).

To further explore if the apparent uniformity of p53 binding is specific to IR, we compared p53 ChIP peaks in MCF7 and UACC257 cell lines treated with either IR or the p53 activator Nutlin3A. We observed condition-condition correlations within each line that were stronger than any line- line correlations (Fig. 1E; R=0.87 or 0.88 for MCF7 and UACC257, respectively, vs R=0.73 for the maximum line-line). Thus in addition to the uniformity we observe between cell lines, IR induced and pharmacologically induced p53 do not lead to distinct p53 function as measured by acute p53 DNA binding.

**Genomic DNA sequence has limited predictive power for p53 binding**

Given the strong conservation of p53 binding across cell lines we hypothesized that p53 binding might be largely determined by genomic DNA sequences. We tested this by comparing motif scores (calculated from the position weight matrix) with p53 ChIP-Seq signal intensity. The extent of the correlation between p53 binding and PWM score was highly cell line dependent (Fig. 2A), ranging from no correlation to correlation of 0.22 in a single cell line. Averaging p53 binding over increasing numbers of cell lines resulted in better agreement between genomic motif score and p53 binding, with the highest correlation of 0.26 averaged across all datasets (Fig. 2 A,B). Therefore, although the motif score significantly correlates with p53 DNA binding (p=1.98e-132), it only allows to account for slightly more than 5% of the variance.

One possible explanation for why the p53 motif only had modest predictive power was the presence of binding elements in the DNA that were not well accounted for by a PWM. It has been proposed, for example, that flanking nucleotide identity and also general 'DNA shape' features of the binding site might influence binding site choice [17, 18]. To explore if our motif analysis was simply a poor model of p53 binding we performed an *in vitro* ChIP experiment where recombinant p53 was incubated with fragmented genomic, followed by immunoprecipitation and deep sequencing similarly to a recently published protocol [19]. As this assay uses sheared genomic DNA with a size of ~300-600bp factors such as flanking nucleotides and DNA shape should be appropriately measured. Preforming this assay, we obtained a strong signal of p53 binding, recovering a consensus p53 motif (1e-2422, Fig 2C). We observed p53 binding sites, such as the one proximal to the CDKN1A/p21 promoter, that showed strong *in vivo* binding, a strong motif, and substantial *in vitro* p53 binding signal (Fig. 2D). Whereas, other binding sites such as the one contained in the first intron of MDM2 showed substantial *in vivo* binding, but little *in vitro* binding (and no strong motif). Other sites such as one in the MDM4 gene showed strong *in vitro* binding (and a strong motif), but little *in vivo* binding. Overall, the *in vitro* p53 binding signal did not show a better correlation (R=0.252, P= 3.10e-127, Fig. 2E) with *in vivo* p53 binding than the motif score These results suggest that factors other than just DNA sequence determine p53 binding *in vivo*.

**A subset of p53 binding sites is highly variable**

To see if we could identify what factors regulate p53 binding *in vivo* we looked for binding sites that were variably occupied across cell lines, with the reasoning that perhaps by understanding cell line-cell line variation in binding would give a clue to *in vivo* regulators of p53 binding. We compared the cell line to cell line variability in p53 peak height, correcting for the average peak height (which contributes shot noise to our analysis), and identified about 5% of binding sites (494 peaks) that showed high variation between cell lines relative to their average peak strength (Fig. 3A, B). For example, p53 peaks nearby the inflammatory associated genes IL1A and CXCL1 showed clear p53 binding in the LOXIMVI line, weaker

association in the UO31 and H460 lines, and no binding in other cell lines (Fig. 3B). We also found variability in p53 binding at the promoters of previously reported p53 target genes, ALDH3A1and EPHA2, ranging from no binding in some cell lines to strong peaks in others (Fig. 3B). *De novo* motif search on this set of variable peaks identified the p53 binding site as the most significantly enriched motif (HOMER, p=1.0e-46), suggesting that these sites represent direct p53 binding events.

To determine if these highly variable binding sites had novel cell line specific functions, we assigned each peak to its closest gene (with a 10 kb cutoff) and clustered the resulting 218 peaks on their p53 binding. We found that most cell lines showed a few unique p53 binding sites, but without strong clustering which would sort the lines into groups (Fig. 3C) as in Fig. 1C. Enrichment analysis identified inflammatory/chemotaxis associated genes as being enriched in these highly variable p53 bound genes. The cell line LOXIMVI showed particularly strong enrichment for p53 binding to inflammatory genes including IL1A, IL1B, CLL20, and CXCL1 genes, although UO31 also showed substantial binding for many of these targets. We also observed, that in the estrogen receptor (ER) positive MCF7 breast cancer cell line, several unique binding sites overlapped with ESR1 (estrogen receptor) binding sites, including TFF1, IGFBP4, and PRLH. These results suggest that the non-uniformities in p53 binding we observed may be linked to cell line specific regulatory programs.

**p53 binding correlates with basal gene expression**

To explore the connection between cell line specific p53 binding and gene expression, we collected RNA sequencing data from each cell line in the basal, untreated state and 3 hours after IR. In contrast to the uniformity of p53 binding, we saw substantially different gene expression programs in response to IR across cell lines. Focusing on established p53 target genes [12] (list of target genes provided in Supplement), we found that while some canonical p53 target genes were universally activated, others showed more variable activation after IR. Notably, while CDKN1A/p21 showed fairly uniform induction in

all cell lines, MDM2 and BBC3/Puma showed variable activation despite having conserved p53 binding (Fig. 4A; Fig. 1A). More generally, there was substantial variation in the activation of p53 targets after IR (Fig. 4B, 77 genes).However we found that, cell lines originating from the same tissue were highly correlated in their gene expression response to IR. For example, the two lung cancer cell lines (A549 and H460) showed a correlation of R=0.90 in their p53 target gene response to IR (Fig. 4C) compared to low or even negative correlation, between other cell line pairs as for example UO31 and H460 (R=-0.49, Fig. 4B,C).

We next asked if differential p53 binding could explain this variation in IR induced gene expression in different cell lines. Quantitative variation in p53 binding was essentially uncorrelated (median R=0.01) to the induction of p53 target genes after IR (Fig. 4D). However, for the highly variable p53 bound genes (Fig. 2C), p53 binding was correlated with basal gene expression (p=1.9e-31, t-test; Fig. 4D, E). Taken together, these results suggest that while the transcriptional response of cell lines to IR is diverse, relatively little of this diversity is explained by p53 binding site occupancy. Conversely, basal gene expression appears to correlate with the ability of p53 to bind to a subset of genes.

**Cell line specific chromatin accessibility accounts for variability in p53 binding sites**

Given our observation that basal gene expression correlated with cell line specific p53 binding events, we wondered if the p53 peaks specific to some cell lines can be attributed to increased chromatin accessibility in these lines. It has been suggested that p53 can act as a pioneer factor with a high affinity for histone occupied regions [20-22], whereas others have shown that p53 binds readily in open regions [23, 24]. Our results linking basal expression of nearby genes to p53 binding suggest that the 'openness' of the genomic region might influence p53 binding. To study this we focused on the LOXIMVI cell line, which showed strong, and unique binding of p53 nearby inflammatory related genes and the MCF7 cell line, which showed p53 binding at estrogen receptor associated genes, and for which DNase accessibility data was

179 available. We performed a modified ATAC-Seq protocol using the MuA transposase to generate genome
180 wide maps of accessible regions in the MCF7 and LOXIMVI cell lines. We note that in this context the MuA
181 transposition occurs in a 5min 30C step, a potential advantage over the more conventional TN5 approach
182 that requires a 30min 37C incubation. Our ATAC-Seq data and ENCODE produced DNase sensitivity data
183 from MCF7 showed strong overlap with greater than 90% of ATAC-Seq peaks being DNase accessible [25].
184 We compared our ATAC-Seq data to the p53 ChIP-Seq signal for the inflammatory genes that showed p53
185 binding in LOXIMVI but not in MCF7, and observed strong ATAC-Seq signal only in the LOXIMVI cell line
186 (Fig. 5A), consistent with increased accessibility at these loci leading to stronger p53 binding. Conversely,
187 GREB1, a breast cancer associated gene showed only p53 binding and ATAC-Seq sensitivity in the MCF7
188 cell line (Fig. 5A). Genome wide the difference in ATAC-Seq signal between the two lines accounted for
189 just over 22% of the variance in p53 binding between the two datasets ($R^2$=0.225; Fig 5B). More generally,
190 as has been observed for other transcription factors [26], combining accessibility and motif scoring allows
191 for improved prediction of DNA binding. Indeed, accessibility and motif score accounted for 13.8% and
192 20.9% of the variance in the log2(p53 ChIP-Seq peak signal) for MCF7 and LOXIMVI respectively, compared
193 to ~5% with the motif alone. We therefore conclude that chromatin accessibility favors p53 binding and
194 accounts for a substantial fraction of the cell line specific gain of p53 DNA binding sites between MCF7
195 and LOXIMVI cells. Interestingly, we also found that genome wide chromatin accessibility was negatively
196 correlated with *in vitro* p53 binding (R=-0.2, p=2.1e-80, MCF7 ATAC-Seq vs. *in vitro* binding), suggesting
197 that many strong p53 binding sites are obscured by local chromatin context.

198 To explore if, within a single cell line, perturbations to accessibility could alter p53 binding we treated
199 MCF7 cells with decitabine, a demethylase inhibitor that has been shown to broadly alter chromatin
200 structure [27]. We then treated these cells with IR and preformed p53 ChIP-Seq and ATAC-Seq on the
201 samples. Comparing p53 binding between the decitabine treated and untreated cells, we observed
202 minimal alterations in p53 binding (Fig. 5C). Indeed, we found only one binding site, adjacent to the

DLGAP5 gene, that showed a substantial change in p53 binding (Fig. 5D). This increase in p53 binding was accompanied by increased accessibility (Fig. 5D). The DLGAP5 binding site has a consensus p53 motif and showed occupancy in other cell lines such as UACC62 (Fig. 5D), consistent with this binding site being obscured by chromatin in wild type MCF7 cells. Computing across p53 binding sites genome-wide, there was a modest but significant correlation between change in chromatin accessibility and change in p53 DNA binding between decitabine treated and untreated samples (R=0.16, p=3.99e-13).

**Discussion**

The transcription factor p53 regulates the cellular response to DNA damage, including up regulating repair, cell cycle arrest and apoptotic proteins. The nature, strength, and balance between the DNA repair and cell death arms of p53 signaling varies across tissues in the body [6, 16, 28], and can be modified by drug treatment [29, 30] and genetic perturbation [31]. The role of the p53 itself in this decision making is controversial, with arguments for p53 behaving as a smart 'signal integrator' (reviewed in [8]) or a simple effector [10]. We sought to understand the role of p53 in diverse cell lines by focusing on p53 DNA binding and gene expression in response to ionizing radiation.

As other studies have suggested, p53 DNA binding does not greatly vary across cell lines or treatments. However, we did find that p53 binding could group cell lines by their tissue of origin, suggesting some degree of tissue specificity. Further, we noted a modest, but significant correlation between the strength of p53 binding (measured by ChIP-Seq) and the predicted strength of p53 association (using the PWM model). This correlation varied across cell lines and was strongest in the pooled dataset containing all cell lines. More strikingly, we observed a similar correlation when comparing genome wide *in vitro* association of p53 with *in vivo* p53 binding. These results suggest that the PWM motif is a relatively accurate measure of p53 binding affinity, and that flanking regions or other DNA features do not greatly improve p53 binding prediction. In general, p53 binding at any given location in the genome was relatively poorly predicted by

226 either *in vitro* binding or motif analysis suggesting that *in vivo* factors greatly contribute to p53 binding
227 specificity.

228 Taking advantage of the coherence of our dataset we identified p53 binding sites that were variably
229 occupied across cell lines. This subset of peaks were nearby genes enriched for specific cellular programs,
230 most notably the inflammatory response in the melanoma LOXIMVI cell line and ER specific response in
231 the MCF7 cell line. Our ATAC-Seq data showed that this differential p53 binding could be substantially
232 explained by chromatin accessibility in the MCF7 and LOXIMVI cell lines we examined. Globally, our data
233 showed that a higher degree of chromatin accessibility favored p53 binding and is consistent with a
234 previous report showing that open chromatin can provide a permissive environment for p53 [21] as well
235 as other transcription factors such as GR [32] for example. We tested this prediction by using a
236 pharmacological agent to modify p53 chromatin state and observed increased accessibility around the
237 DLGAP5 gene that correlated with increased p53 binding.

238 Looking at gene expression of p53 bound genes, we observed no direct correlation between variation in
239 p53 binding and induction of gene expression. This suggests that although p53 binding is required for the
240 DNA damage response, the extent and mode of the IR gene expression response, is dictated by other
241 factors than p53 binding itself. We recently showed that mRNA stability of direct p53 target genes can
242 lead do differences in the timing and level of expression [33]. Further studies coupling chromatin
243 accessibility, p53 binding, post-translational modifications, and measurements of RNA synthesis and
244 degradation rates will be required to reconcile these observations and identify what features tune the
245 cellular response to DNA damage in different cellular backgrounds.

246 **Conclusions**

247 We found that the vast majority of p53 binding events as universal across cancerous and non-cancerous
248 cell lines, with strong quantitative agreement in binding magnitude. We further found that Nutlin3A

249 treatment resulted in a nearly identical set of p53 binding events as IR, emphasizing the universality of
250 these binding sites. These binding sites were not, however, well predicted by the local genomic sequence.
251 We identified a set of variable p53 binding events (~5%) present in only one or a handful of cell lines.
252 These binding events were often in the vicinity of transcriptionally active genes and correlated strongly
253 with cell line specific chromatin accessibility. Consistent with this, pharmacological modification of
254 chromatin state could modify p53 binding. Interestingly, we did not find a strong correlation of p53 DNA
255 binding to the transcriptional response to IR, suggesting additional layers of transcriptional control.
256 Overall we found that the acute activation and binding of p53 to DNA in response to damage is highly
257 stereotyped, with modest tuning by the chromatin environment.

258 **Methods**

259 General genomic analysis

260 All DNA reads were single end Illumina reads and were aligned to HG19 genome build using bowtie [34].
261 Aligned reads were further processed using HOMER (V4.6, [35]) to assemble tag files and call peaks. Peak
262 locations and tag files were read by custom Matlab code (Mathworks) which was used to integrate the
263 different datasets/types. For each ChIP-Seq dataset the number of reads in p53 peaks were normalized
264 to the average of all cell lines, and for subsequent analyses and comparisons, peaks with less than 2
265 normalized counts were discarded. Clustering and comparison was based on Pearson correlations
266 between p53 binding signals. RNA data was aligned to hg19 and the Refseq HG19 transcriptome using
267 Tophat, CuffQuant, and CuffMerg [36]. Most computation was done on the Harvard Medical School
268 cluster (Orchestra). Genomic binding and signals were visualized using the UCSC genome browser [37].
269 Motif analysis was performed in Matlab on the Hg19 genome using a ChIP-Seq derived PWM adjusted to
270 have a minimum probability of each nucleotides occurrence.

271 ATAC-Seq

ATAC seq was performed as described [38], with the major exception of the use of a MuA transposase (Thermo) rather than the TN5 transposase. Briefly, MCF7 or LOXIMVI cells were trypsinized and 50K cells spun down, washed once with PBS, and lysed with a hypotonic buffer containing 0.1% NP-40, and spun down to generate a crude nuclei pellet. This pellet was transposed in a 30μl volume using MuA (0.7μl), MuA buffer (10μl), and H2O (19μl) for 5min at 30C. The sample was treated with 3μl stop solution, and incubated at 30C for a further minute. The sample was then collected and purified by addition of 45μl of SPRI beads (Aline Biosciences). The purified sample was PCR amplified in two steps to add barcoded adaptors suitable for Illumina sequencing. Samples were sequenced with single end 75bp reads on an Illumina NextSeq. Reads (>30M) were trimmed to remove adaptors with cutadapt [39], aligned to the genome with Bowtie, and analyzed with Matlab. Genomic DNA (50ng) from MCF7 and LOXIMVI was transposed, amplified and sequenced in parallel to estimate background.

ChIP-Seq

Briefly, 10M cells were treated with 4Gy IR (RS-2000, RadSource) and 2hrs later were fixed by addition of 1% paraformeldahyde (Alfa Aesar) at room temperature for 10 minutes with agitation. Fixation was stopped by addition of 250mM glycine. Cells were scraped and flash frozen. Cell pellets were thawed in hypotonic lysis buffer and spun to generate a crude nuclei prep. These nuclei were lysed in a SDS buffer and sonicated (Bioruptor) to fragment DNA. Fragmented DNA was diluted in IP buffer and agitated overnight with 2mg/ml DO-1 (anti-p53, Santa Cruz). 20μl of protein A magnetic beads (Invitrogen) were used to isolate the p53 associated fragments and samples were washed with low salt, high salt, and LiCl buffers. DNA was eluted from beads with an SDS/NaCO3 buffer and was decross-linked at 65C for 6hrs in a high salt buffer. ChIP libraries were constructed with the commercial NEBnext kit (NEB) and associated protocols, although reaction volumes were reduced by 4 fold and custom adaptors and barcodes were employed. Libraries were sequenced with single end 75bp reads on Illumina NextSeq.

Reads were aligned to the genome with Bowtie1.1 [34], and analyzed with HOMER [35], MACS2 [40] and custom Matlab scripts. Peak calling was done after pooling reads from ChIP-Seq experiments in all cell lines. The final set of peaks represented the consensus of HOMER (default settings) and MACS2 (using the q 0.01 threshold) identified peaks, and was filtered to remove ENCODE black-list locations. The number of reads within each peak region was computed from HOMER tag files using custom Matlab scripts. Background regions around each peak were subtracted from peak scores to correct for high background regions. The HOMER package [35] was used for *de novo* motif discovery. WebLogo was used to generate the motif plot [41] in (Figs. 1B, 2C) for the top enriched motif. The top enriched motif (Fig. 1B) was then used to re-scan and score all peaks and background regions. Background regions were generated by selecting 500bp regions adjacent to either side of the peak and excluding regions that overlap with p53 peak regions. Clustering of peaks was accomplished using a correlation distance metric and average linkage in Matlab.

*In vitro* ChIP-Seq

To generate recombinant p53 we in vitro transcribed/translated human p53 with a c-terminal HA tag using a rabbit reticulocyte system (Promega). To generate fragmented genomic DNA we tagmented 50ng of human genomic DNA from MCF7 cells using the MuSeq kit (Thermo) and amplified it using PCR and custom adaptor primers for 8 cycles. DNA was cleaned up on SPRI beads (Aline Biosciences) and quantified. At room temperature 20ng of DNA and recombinant p53 (0.1uM final) were combined in a binding buffer (10mM TRIS, 5mM MgCl2, 10% glycerol, 1mM DTT) and incubated at room temperature for 30 min. The mixture was diluted 2 fold (to 20ul) and 1.5ul of anti-HA antibody was added (Rockland) and the sample incubated at 4C overnight with shaking. A 1:1 mixture of magnetic proteinA/proteinG beads was added (sigma) and incubated at 4C for 1hr with shaking. The beads were then washed 3x with washing buffer (10mM Tris, 5mM HCL, 0.1% triton, 150mM NaCl) and DNA eluted with elution buffer (1%SDS, 100mM Na2CO3) at 37C for 15 minutes. Samples were cleaned up, and adaptors and barcodes added by PCR.

Reads (>30M) were trimmed to remove adaptors with cutadapt [39], aligned to the genome with Bowtie, and analyzed with Matlab.

RNA-Seq

For each cell line 50,000 cells were plated in 35 mm dishes, 24hrs later cells were treated (or not) with 4Gy IR (RS-2000, RadSource), 3hrs after that cells were lysed with Trizol (Ambion). RNA was purified on affinity columns and DNAse treated (Zymo). Purified RNA (500ng) was polyA purified using magnetic beads (NEB), fragmented and reverse transcribed using protoscript RT (NEB), second strand synthesized (NEB), and then assembled into libraries with the commercial NEBnext kit (NEB) and associated protocols, although reaction volumes were reduced by 4 fold and custom adaptors and barcodes were employed. Libraries were sequenced with single end 75bp reads on a NextSeq. Reads were aligned to the genome/transcriptome with Bowtie/Tophat and analyzed with custom Matlab scripts.

Cell Culture and Cell Treatment

Parental cell lines were obtained from ATTC, with the exception of the RPE cells which were a gift from Prof. Steve Elledge (Harvard Medical School), and were thawed and propagated in RPMI (GIBCO) with 5% FBS. All experiments were performed in this media. All media was supplemented with 1% antibiotic and antimycotic (Corning). Treatment with Nutlin3A (sigma) was at 5uM. X-ray induced DNA damage was generated with a RS-2000 source (RadSource, 160KeV). MCF7 cells were treated with 2uM (5-AZA-2'-deoxcytidine, MP Biomedicals) for 5 days, cells were split on day 2, replated in decitabine containing media. Treated and untreated cells were then further treated with IR or not as with other samples.

Public datasets

Raw fastq datasets were downloaded from the Sequence Read Archive (see supplement). These datasets were all single end Illumina reads and were aligned to the HG19 genome with using the same pipeline as

341 described above for our ChIP-Seq samples, and further analyzed with HOMER to generate tag files.
342 Custom Matlab code was used to compare these datasets to our ChIP-seq data.

343 **Declarations**

344 Ethics approval and consent to participate

345 Not Applicable.

346 Consent for publication

347 Not Applicable

348 Availability of data and material

349 All sequencing data has been deposited in NCBI's Gene Expression Omnibus under accession number
350 GSE100292. Data is also available as UCSC tracks as a custom session accessible at:
351 https://genome.ucsc.edu/cgi-
352 bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=jacobso&hgS_otherUserSessionName=H
353 afner_et_al_hg19

354 Competing interests

355 The authors declare no conflicts of interest.

356 Funding

357 AH was supported by a Boehringer Ingelheim Fonds PhD fellowship. This study was supported by NIH
358 grants GM083303 and CA207727 (J.S-O).

359 Authors' contributions

JSO and AH initiated and conceived the study; JSO and AH collected and analyzed the data; AH, JSO and GL wrote the paper, JSO and GL supervised the study.

Acknowledgements

We thank Megha Padi and all members of the Lahav lab for feedback. We thank the Bauer sequencing core and Christian Daly for sequencing support and HMS research computing for maintaining the Orchestra cluster.

**References**

1. Vaquerizas, J.M., et al., *A census of human transcription factors: function, expression and evolution.* Nat Rev Genet, 2009. **10**(4): p. 252-63.

2. Chen, X., et al., *Integration of external signaling pathways with the core transcriptional network in embryonic stem cells.* Cell, 2008. **133**(6): p. 1106-17.

3. Boyer, L.A., et al., *Core transcriptional regulatory circuitry in human embryonic stem cells.* Cell, 2005. **122**(6): p. 947-56.

4. Gerstein, M.B., et al., *Architecture of the human regulatory network derived from ENCODE data.* Nature, 2012. **489**(7414): p. 91-100.

5. Garber, M., et al., *A high-throughput chromatin immunoprecipitation approach reveals principles of dynamic gene regulation in mammals.* Mol Cell, 2012. **47**(5): p. 810-22.

6. Fei, P., E.J. Bernhard, and W.S. El-Deiry, *Tissue-specific induction of p53 targets in vivo.* Cancer Res, 2002. **62**(24): p. 7316-27.

7. Brady, C.A. and L.D. Attardi, *p53 at a glance.* J Cell Sci, 2010. **123**(Pt 15): p. 2527-32.

8. Kruiswijk, F., C.F. Labuschagne, and K.H. Vousden, *p53 in survival, death and metabolic health: a lifeguard with a licence to kill.* Nat Rev Mol Cell Biol, 2015. **16**(7): p. 393-405.

9. Aylon, Y. and M. Oren, *The Paradox of p53: What, How, and Why?* Cold Spring Harb Perspect Med, 2016. **6**(10).

10. Verfaillie, A., et al., *Multiplex enhancer-reporter assays uncover unsophisticated TP53 enhancer logic.* Genome Res, 2016. **26**(7): p. 882-95.

11. Nikulenkov, F., et al., *Insights into p53 transcriptional function via genome-wide chromatin occupancy and gene expression analysis.* Cell Death Differ, 2012. **19**(12): p. 1992-2002.

12. Fischer, M., *Census and evaluation of p53 target genes.* Oncogene, 2017.

13. Botcheva, K., et al., *Distinct p53 genomic binding patterns in normal and cancer-derived human cells.* Cell Cycle, 2011. **10**(24): p. 4237-49.

14. Smeenk, L., et al., *Role of p53 serine 46 in p53 target gene regulation.* PLoS One, 2011. **6**(3): p. e17574.

15. Akdemir, K.C., et al., *Genome-wide profiling reveals stimulus-specific functions of p53 during differentiation and DNA damage of human embryonic stem cells.* Nucleic Acids Res, 2014. **42**(1): p. 205-23.

16. Stewart-Ornstein, J. and G. Lahav, *p53 dynamics in response to DNA damage vary across cell lines and are shaped by efficiency of DNA repair and activity of the kinase ATM.* Sci Signal, 2017. **10**(476).

17. Gordan, R., et al., *Genomic regions flanking E-box binding sites influence DNA binding specificity of bHLH transcription factors through DNA shape.* Cell Rep, 2013. **3**(4): p. 1093-104.

18. Rohs, R., et al., *The role of DNA shape in protein-DNA recognition.* Nature, 2009. **461**(7268): p. 1248-53.

19. Bartlett, A., et al., *Mapping genome-wide transcription-factor binding sites using DAP-seq.* Nat Protoc, 2017. **12**(8): p. 1659-1672.

20. Lidor Nili, E., et al., *p53 binds preferentially to genomic regions with high DNA-encoded nucleosome occupancy.* Genome Res, 2010. **20**(10): p. 1361-8.

21. Su, D., et al., *Interactions of chromatin context, binding site sequence content, and sequence evolution in stress-induced p53 occupancy and transactivation.* PLoS Genet, 2015. **11**(1): p. e1004885.

22. Sammons, M.A., et al., *TP53 engagement with the genome occurs in distinct local chromatin environments via pioneer factor activity.* Genome Res, 2015. **25**(2): p. 179-88.

23. Tebaldi, T., et al., *Whole-genome cartography of p53 response elements ranked on transactivation potential.* BMC Genomics, 2015. **16**: p. 464.

24. Allen, M.A., et al., *Global analysis of p53-regulated transcription identifies its direct targets and unexpected regulatory mechanisms.* Elife, 2014. **3**: p. e02200.

425 25. ENCODE, *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012.
426 **489**(7414): p. 57-74.
427 26. Zhong, S., X. He, and Z. Bar-Joseph, *Predicting tissue specific transcription factor binding sites.*
428 BMC Genomics, 2013. **14**: p. 796.
429 27. Ponnaluri, V.K.C., et al., *NicE-seq: high resolution open chromatin profiling.* Genome Biol, 2017.
430 **18**(1): p. 122.
431 28. Gudkov, A.V. and E.A. Komarova, *The role of p53 in determining sensitivity to radiotherapy.* Nat
432 Rev Cancer, 2003. **3**(2): p. 117-29.
433 29. Purvis, J.E., et al., *p53 dynamics control cell fate.* Science, 2012. **336**(6087): p. 1440-4.
434 30. Paek, A.L., et al., *Cell-to-Cell Variation in p53 Dynamics Leads to Fractional Killing.* Cell, 2016.
435 **165**(3): p. 631-42.
436 31. Sullivan, K.D., et al., *ATM and MET kinases are synthetic lethal with nongenotoxic activation of*
437 *p53.* Nat Chem Biol, 2012. **8**(7): p. 646-54.
438 32. John, S., et al., *Chromatin accessibility pre-determines glucocorticoid receptor binding patterns.*
439 Nat Genet, 2011. **43**(3): p. 264-8.
440 33. Hafner, A., Stewart-Ornstein, J., Purvis, J.E., Forrester, W.C., Bulyk, M.L., Lahav, G, *p53 pulses lead*
441 *to distinct patterns of gene expression albeit similar DNA binding dynamics.* Nature Structural &
442 Molecular biology, 2017. **In press**.
443 34. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the*
444 *human genome.* Genome Biol, 2009. **10**(3): p. R25.
445 35. Heinz, S., et al., *Simple combinations of lineage-determining transcription factors prime cis-*
446 *regulatory elements required for macrophage and B cell identities.* Mol Cell, 2010. **38**(4): p. 576-
447 89.

448 36. Trapnell, C., et al., *Differential gene and transcript expression analysis of RNA-seq experiments*
449 *with TopHat and Cufflinks.* Nat Protoc, 2012. **7**(3): p. 562-78.
450 37. Kent, W.J., et al., *The human genome browser at UCSC.* Genome Res, 2002. **12**(6): p. 996-1006.
451 38. Buenrostro, J.D., et al., *Transposition of native chromatin for fast and sensitive epigenomic*
452 *profiling of open chromatin, DNA-binding proteins and nucleosome position.* Nat Methods, 2013.
453 **10**(12): p. 1213-8.
454 39. Martin, M., *Cutadapt removes adapter sequences from high-throughput sequencing reads.*
455 EMBnet.journal, 2011. **17**(1): p. 10-12.
456 40. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome Biol, 2008. **9**(9): p. R137.
457 41. Crooks, G.E., et al., *WebLogo: a sequence logo generator.* Genome Res, 2004. **14**(6): p. 1188-90.
458 42. Zeron-Medina, J., et al., *A polymorphic p53 response element in KIT ligand influences cancer risk*
459 *and has undergone natural selection.* Cell, 2013. **155**(2): p. 410-22.

460

461

462

463

464

465

466

467 **Figure Legends**

468 **Fig. 1: Stereotyped p53 binding across twelve cell lines.** (A) ChIP-Seq for p53 in 12 p53 wild-type cell
469 lines. UCSC screen shots of p53 binding sites for three canonical p53 target genes are shown. (B) Motif
470 analysis recovered a p53 motif as the single most enriched sequence. Centrality analysis showed that this
471 motif was also centrally enriched within peaks. (C) Heatmap showing p53 binding intensity in 8742
472 locations in the genome. Cell lines were clustered on p53 binding and resulted in lineages clustering
473 together. (D) Comparison of p53 binding in two cancer cell lines (UACC62 and UACC257) as well as
474 between one cancer (UACC257) and one non-cancerous cell line (RPE1). (E) Comparison of p53 binding
475 between Nutlin3A and IR treated samples in MCF7 or UACC257 cells.

476 **Fig. 2: Genomic sequence is weakly predictive of p53 binding**. (A) The correlation between motif strength
477 and p53 binding is shown as a function of the number of cell lines across which the peak height was
478 averaged, box plot represent the distribution of correlations across all possible cell line combinations. (B)
479 The degree to which p53 binding motif predicts the strength of p53 binding is shown in a box plot, with
480 p53 binding sites binned by their motif strength. (C) The top enriched motif identified by *in vitro* ChIP.
481 (D) UCSC screen shots of p53 binding sites, motif score, and *in vitro* p53 binding signal are shown for
482 CDKN1A/p21, MDM2, and MDM4. (E) I*n vivo* p53 binding strength is shown in a box plot, binned by *in*
483 *vitro* p53 binding signal at each genomic site.

484 **Fig. 3: Variable p53 binding sites show cell type specific functional enrichment**. (A) Scatterplot of all
485 8742 p53 binding sites by their average ChIP signal and coefficient of variation (CV). Highlighted in red
486 are 'highly variable' peaks defined as having higher than expected CV relative to the peak height. Example
487 binding sites are labeled with the associated gene names. (B) UCSC screen shots of four example 'variable'
488 peaks. (C) Heatmap of 'variable' p53 peaks that are also nearby (<10kb) transcription start sites of genes.
489 The intensity of each peak is normalized to the average across 12 cell lines. Cell lines and peaks were
490 hierarchically clustered, with no grouping by lineage observed for cell lines. Groups of inflammatory and
491 the ER associated are highlighted.

**Fig. 4: Basal gene expression, but not DNA damage induced gene expression, correlates with p53 binding.** (A) Box plots of the fold change of three canonical p53 target genes 3h after IR. Each dot represents a cell line. CDKN1A/p21 is induced in all cell lines, while MDM2 and BBC3/Puma is cell line dependent. (B) Correlation of 12 cell lines by p53 target gene (77 genes) induction after DNA damage. (C) Scatter plots of p53 target genes comparing the highly correlated lung cancer lines (A549 and H460, upper) and negatively correlated lung and kidney lines (H460 and UO31, lower). (D) Boxplots showing the correlation of either basal or DNA damage induced fold change of gene expression with p53 binding at the p53 target or variable p53 binding gene sets (Fig. 3). (E) Scatterplots illustrating the relationship between basal gene expression and p53 binding across the 12 cell lines for four p53 peaks/genes (note that in many cases multiple cell lines show little gene expression or p53 binding and therefore cluster near the origin).

**Fig. 5 Chromatin accessibility contributes to variable p53 binding.** (A) UCSC screen shots of two p53 binding sites; p53 binding in the proximity of GREB1 is found in MCF7 treated with IR, while IL1A was bound in IR treated LOXIMVI cells. ATAC-Seq data and published DNase hypersensitivity data (for MCF7, untreated) showing that IR induced p53 binding correlates with basal DNA accessibility in each cell line. (B) Scatter plot of p53 binding post IR in MCF7 compared to LOXIMVI, colored by the difference in ATAC-Seq signal log2(LOXIMVI-MCF7) between the two cell lines. (C) Scatter plot of p53 binding in IR treated MCF7 cells compared to MCF7 treated with IR and decitabine, colored by the difference in ATAC-Seq signal between decitabine treated and untreated cells: log2(decitabine-not treated). (D) UCSC screen shot of region around the gene DLGAP5, showing changes in p53 binding and accessibility in the decitabine treated MCF7 cells (the new peak is indicated by an arrow). Binding of p53 in IR treated UACC62 cells at the DLGAP5 locus without decitabine treatment is also shown.

**Supplement**

515 Public p53 ChIP-Seq datasets

516 SRR048928, SRR048929 – U2OS cells: Actinomycin D or Etoposide treated [14]

517 SRR287798, SRR287799, SRR287800 – MCF7: Nutlin3A, RITA, 5-FU [11]

518 SRR575903, SRR575904, SRR575905 – hESC : RA or Doxorubicin [15]

519 SRR851807, SRR851807, SRR851807 -- lymphoblastoid cell lines: doxorubicin [42]

520 p53 Target Genes used in this analysis

521 This gene list is based on the list of [12] and removing genes with low read coverage in our dataset.

| | | |
|---|---|---|
| CDKN1A | HSPA4L | ZNF337 |
| RRM2B | ISCU | ZNF79 |
| MDM2 | PHLDA3 | CERS5 |
| GDF15 | SLC12A4 | CSF1 |
| SUSD6 | TRAF4 | DUSP14 |
| BTG2 | CCDC90B | EPS8L2 |
| DDB2 | CES2 | FAM210B |
| GADD45A | DYRK3 | FUCA1 |
| PLK3 | KITLG | IER5 |
| RPS27L | NADSYN1 | IKBIP |
| TNFRSF10B | NTPCR | MICALL1 |
| TRIAP1 | ORAI3 | PMAIP1 |
| BAX | SESN2 | RAP2B |
| BLOC1S2 | SLC30A1 | RNF19B |
| POLH | TM7SF3 | SAC3D1 |

| | | |
|---|---|---|
| PPM1D | TMEM68 | TNFRSF10D |
| SULF2 | ANXA4 | |
| XPC | APOBEC3C | |
| AEN | ASCC3 | |
| ANKRA2 | BBC3 | |
| NINJ1 | DCP1B | |
| PLK2 | EPHA2 | |
| SERTAD1 | FDXR | |
| SESN1 | FOSL1 | |
| TP53I3 | LIF | |
| CCNG1 | PGPEP1 | |
| CMBL | PRKAB1 | |
| CYFIP2 | PTP4A1 | |
| DRAM1 | TGFA | |
| FBXO22 | ZNF219 | |

## Supplementary Figure Legends

**Fig. S1: Comparison of p53 DNA binding with published p53 ChIP-Seq datasets.** Heatmap showing p53 binding intensity in 8742 locations in the genome in 12 IR treated cell lines (same as Fig. 1C) as well as published datasets.

**Figure 1**

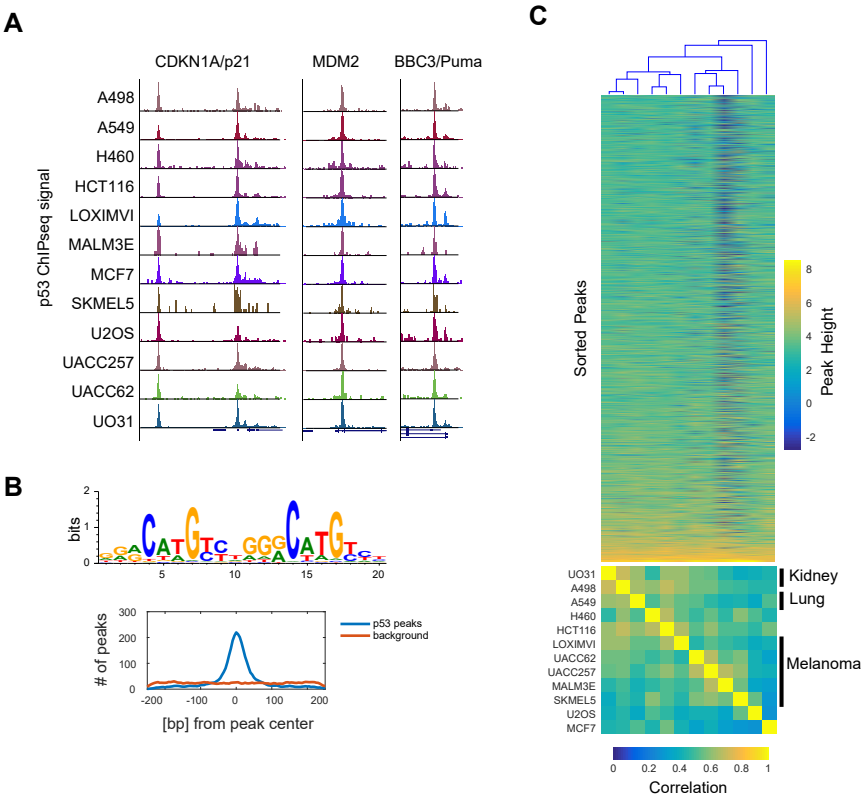

**Figure 2**

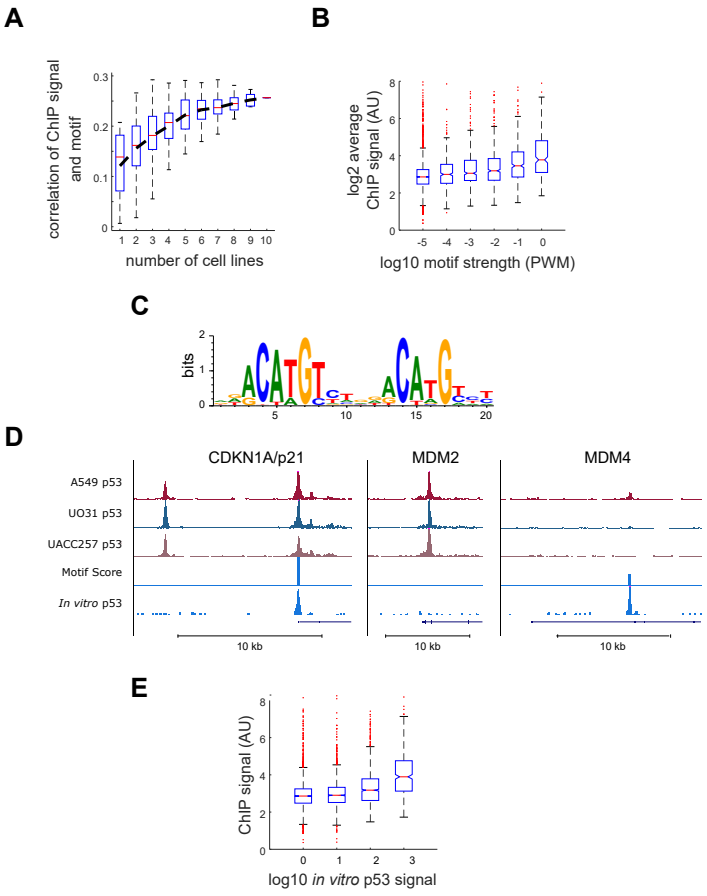

**Figure 3**

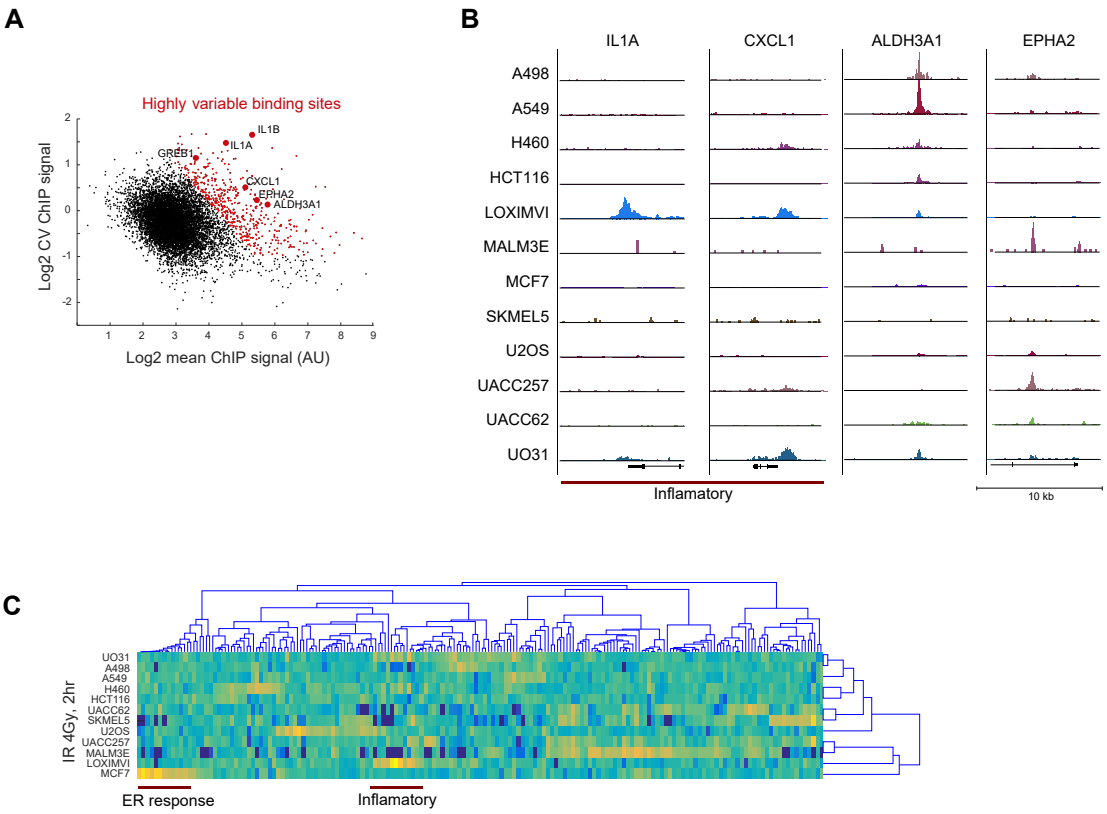

**Figure 4**

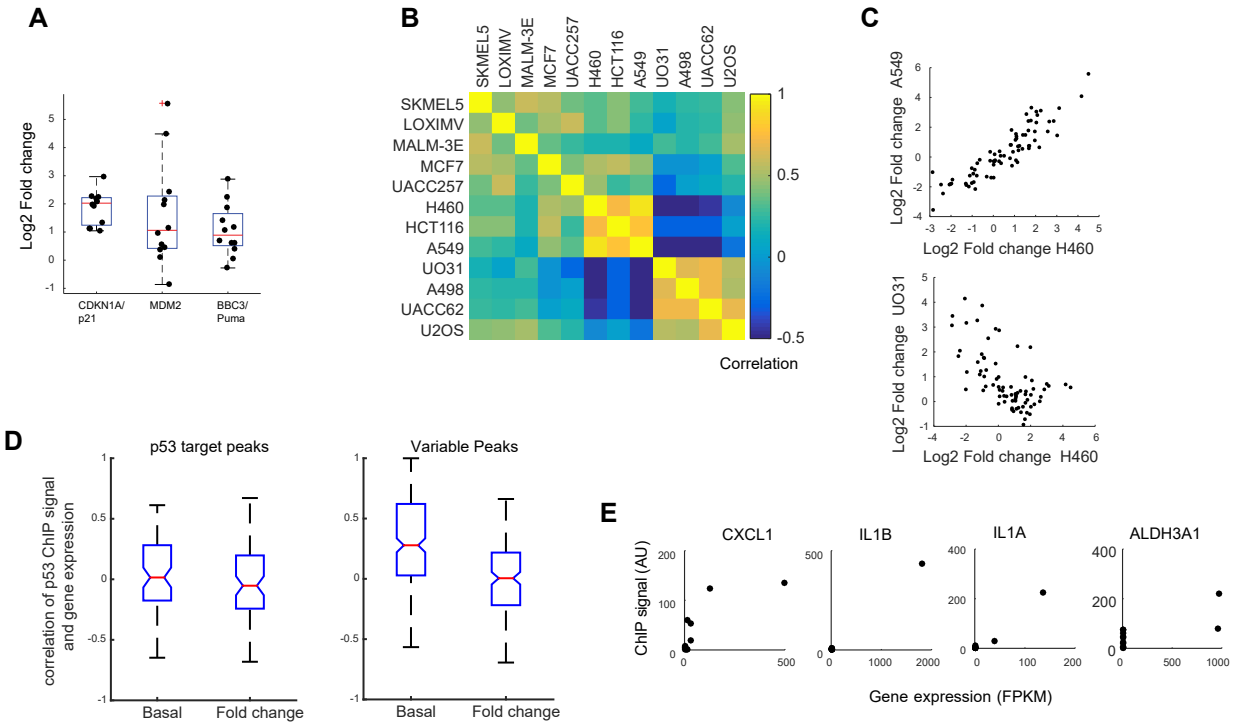

**Figure 5**

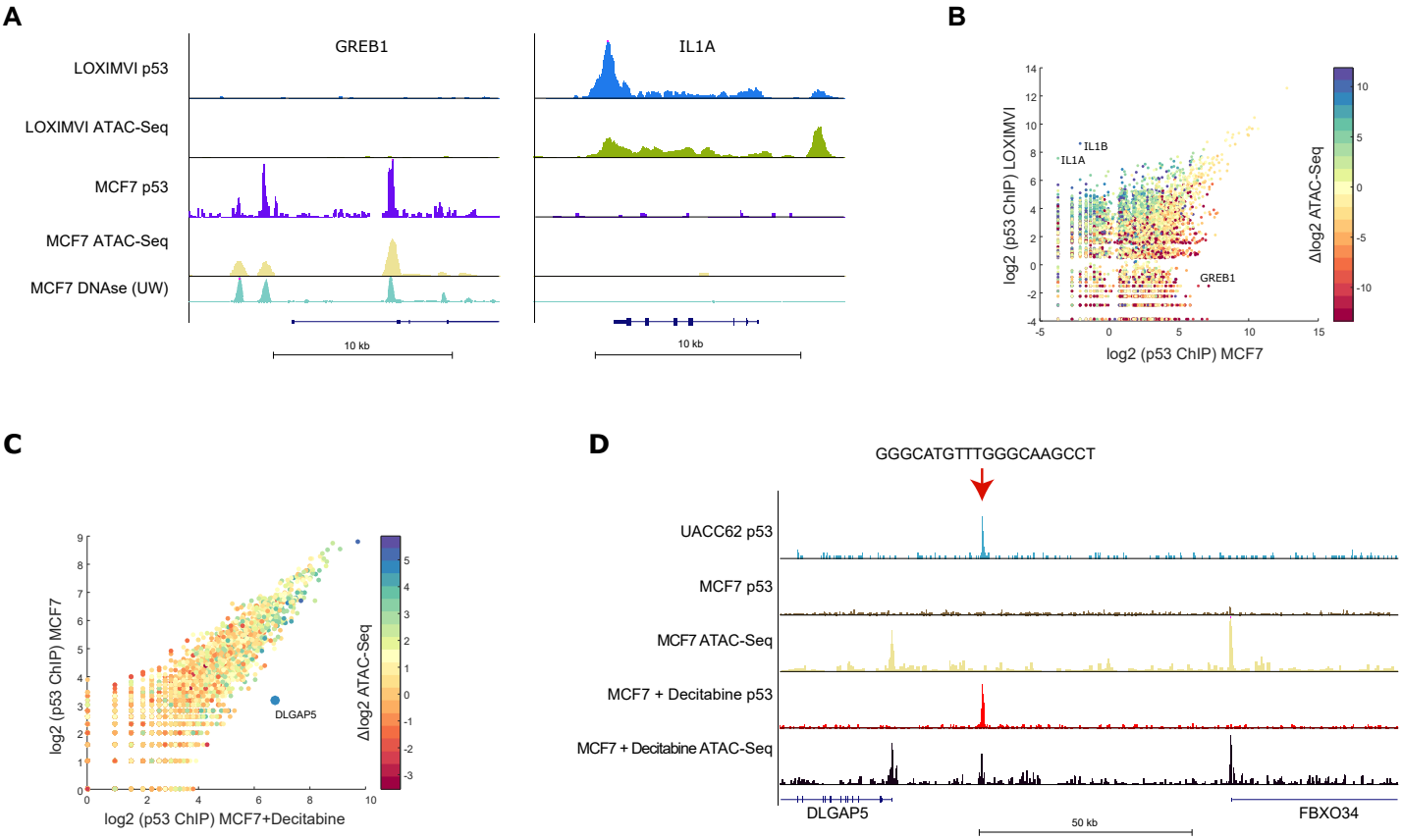

**Figure S1**