Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press



Multiplex enhancer-reporter assays uncover unsophisticated TP53 enhancer logic

Annelien Verfaillie, Dmitry Svetlichnyy, Hana Imrichova, et al.

Genome Res. published online May 18, 2016
Access the most recent version at doi:10.1101/gr.204149.116

P<P Published online May 18, 2016 in advance of the print journal.

Accepted Manuscript Peer-reviewed and accepted for publication but not copyedited or typeset; accepted manuscript is likely to differ from the final, published version.

Open Access Freely available online through the *Genome Research* Open Access option.

Creative Commons License This manuscript is Open Access.This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International license), as described at http://creativecommons.org/licenses/by/4.0/.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**



To subscribe to *Genome Research* go to:
https://genome.cshlp.org/subscriptions

Published by Cold Spring Harbor Laboratory Press

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

Multiplex enhancer-reporter assays uncover unsophisticated TP53 enhancer logic

Annelien Verfaillie[1], Dmitry Svetlichnyy[1], Hana Imrichova[1], Kristofer Davie[1], Mark Fiers[2], Zeynep Kalender Atak[1], Gert Hulselmans[1], Valerie Christiaens[1], and Stein Aerts[1#]

[1]Laboratory of Computational Biology, Center for Human Genetics, University of Leuven, Leuven, Belgium

[2]VIB Center for the Biology of Disease, Leuven, Belgium

correspondence to: stein.aerts@med.kuleuven.be

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

Abstract

Transcription factors regulate their target genes by binding to regulatory regions in the genome. Although the binding preferences of TP53 are known, it remains unclear what distinguishes functional enhancers from non-functional binding. In addition, the genome is scattered with recognition sequences that remain unoccupied. Using two complementary techniques of multiplex enhancer-reporter assays, we discovered that functional enhancers could be discriminated from non-functional binding events by the occurrence of a single TP53 canonical motif. By combining machine learning with a meta-analysis of TP53 ChIP-seq data sets we identified a core set of more than one thousand responsive enhancers in the human genome. This TP53 *cistrome* is invariably used between cell types and experimental conditions, while differences between experiments can be attributed to indirect non-functional binding events. Our data suggest that TP53 enhancers represent a class of unsophisticated cell-autonomous enhancers containing a single TP53 binding site, distinct from complex developmental enhancers that integrate signals from multiple transcription factors.

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

Introduction

Enhancers are essential regulatory elements that are bound by transcription factors (TFs) to shape the gene expression network underlying a cellular phenotype. Understanding the enhancer code is crucial to achieve a functional regulatory annotation of the human genome, which is ultimately required to understand developmental processes and disease-related variation in the non-coding part of the genome. However, the complexity of the enhancer logic, their sparse distribution, context-specificity and distal location from genes in the genome make it challenging to identify and validate enhancers. During recent years, high-throughput sequencing efforts like ENCODE and the Epigenomics Roadmap have yielded an enormous wealth of regulatory data (Gerstein et al. 2012; Roadmap Epigenomics Consortium et al. 2015). Some of the most commonly used approaches are chromatin immunoprecipitation (ChIP-seq) to localize regions bound by a certain TF or a modified histone, and various assays aimed at localizing accessible and free regions within the genome (e.g. DNase hypersensitivity sites sequencing (DNase-seq) or ATAC-seq (Assay for Transposase-Accesible Chromatin sequencing)). But while these methods generate genome-wide regulatory profiles and chromatin states (Ernst et al. 2011), they have not lead to sequence-based enhancer models and usually contain high levels of non-functional noise.

A key feature of TF binding is the presence of the TF's DNA recognition sequence. Interestingly, sequence analysis on ChIP-seq data in yeast showed that not all identified peaks are directly bound by their TFs (Gordân et al. 2009). Similarly, amongst the hundreds of TF ChIP-seq data of ENCODE only a fraction of peaks were found to contain the canonical recognition sequence of their respective TF (Wang et

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

al. 2012). But while a considerable amount of ChIP-seq peaks seem to be independent of the TF motif, it is not clear whether this indirect binding plays a functional role, i.e. whether it is involved in the regulation of a target gene. In Drosophila it has been suggested that indirect binding could be functional and contribute to gene regulation via transcription factor cooperative activity i.e. the tethering of a TF to an enhancer by other, directly bound factors (Junion et al. 2012). On the other hand, indirect binding could also reflect a technical aspect of ChIP and represent for instance fixation artifacts (Keren and Segal 2013; Waldminghaus and Skarstad 2010; Baranello et al. 2015).

Using the motif of the bound factor may be a good guide to identify functional binding, but for most – if not all - TFs, tens of thousands of recognition sequences, present throughout the genome, remain unoccupied or unbound. The current paradigm of how transcription factors discriminate between functional and non-functional locations is based on the combinatorial action of transcription factors. Here, binding specificity is achieved through clusters of binding motifs co-occurring within an enhancer (Stampfel et al. 2015; Maston et al. 2012; Shlyueva et al. 2014; Lee and Young 2013; Panne 2008), or by generating new recognition sequences for pairs of TFs . Such enhancers are generally classified under the enhanceosome or billboard models, depending on whether the order and spacing of motifs is important or not (Arnosti and Kulkarni 2005). Other local features in the DNA sequence of an enhancer have also been shown to contribute to the discrimination of bound versus unbound sites, such as GC content (White et al. 2013), preferential sequences for nucleosome positioning (Lidor Nili et al. 2010), DNA shape features (Chiu et al. 2015) and dinucleotide repeat motifs (Yáñez-Cuna et al. 2014). However, to gain

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

further insight into the *cis*-regulatory enhancer code and identify those that are truly bound and functional, validation assays are needed.

The recent development of multiplex enhancer-reporter assays has been invaluable in procuring such information on a large scale. Whereas classically enhancer-reporter assays consist of cloning each enhancer one by one, first *in vitro*, later *in vivo* (Banerji et al. 1981; Chiocchetti et al. 1997; O'Kane and Gehring 1987; Dailey 2015), now hundreds to thousands of enhancers can be tested in parallel (Patwardhan et al. 2009; Arnold et al. 2013; White et al. 2013; Vanhille et al. 2015; Patwardhan et al. 2012; Kheradpour et al. 2013; Smith et al. 2013; Melnikov et al. 2012; Kwasnieski et al. 2012). These methods can be broadly categorized in two groups, namely massively parallel reporter assays (MPRA) utilizing barcodes as a measure of activity of synthesized enhancer fragments (Kheradpour et al. 2013; Patwardhan et al. 2009, 2012; Smith et al. 2013; Melnikov et al. 2012; Kwasnieski et al. 2012; White et al. 2013)), and self-transcribing active regulatory region sequencing (STARR-seq) (Arnold et al. 2013; Vanhille et al. 2015).

In this work we unravel the genome-wide binding profile of TP53 (protein product of *TP53*, also known as p53), addressing these two questions simultaneously by investigating (i) the differences between direct and indirect binding and (ii) the differences between bound and unbound recognition sequences in the genome. TP53 is a tumor-suppressor that regulates its target genes in response to different stimuli like DNA damage or cellular stress, eliciting functions like growth arrest or apoptosis (Vousden and Prives 2009). Its importance is moreover reflected by the fact that *TP53* is the most commonly mutated gene found in cancer (Muller and Vousden 2014). Although much research has gone in understanding how and where TP53 interacts

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

with DNA, many questions and contradictions remain. For instance, it is unclear what the functional role is of indirect binding of TP53 to the DNA (Kirschner et al. 2015), whether TP53 also directly represses genes (Ho and Benchimol 2003; Johnson et al. 2001; Rinn and Huarte 2011) and how TP53 elicits different responses like apoptosis and growth arrest, activating different genes for each process (Smeenk et al. 2008; Menendez et al. 2013). Several elements of the TP53 binding site architecture have been proposed to contribute to the functional properties of the enhancer. These include variations of the spacer between two TP53 half-sites (Godar et al. 2008; Hoffman et al. 2002), or variations in binding sites for co-regulatory factors (Thornborrow and Manfredi 2001; Koutsodontis et al. 2001). However, these properties have not been evaluated on a global scale. To resolve these uncertainties and to generate a global TP53 enhancer model, we have combined two types of multiplex enhancer-reporter assays followed by machine learning. Our results yield a new unsophisticated model for TP53-mediated transcriptional regulation and allow us to create a ranked list of all potentially functional TP53 binding sites in the human genome.

Results

Quantitative enhancer-reporter activity for hundreds of enhancers in parallel

We developed a new method, called CHEQ-seq (Captured High-Throughput Enhancer testing by Quantitative sequencing) to test enhancer-reporter activities for hundreds of enhancers simultaneously (**Fig. 1a**). Although conceptually CHEQ-seq is similar to the recently published capture-and-clone variant of CRE-seq (Shen et al. 2015; Kwasnieski et al. 2012), CHEQ-seq uses a different cloning and sequencing strategy (**Supplemental protocol**). Firstly, candidate enhancers from sheared

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

genomic DNA are captured using custom designed baits **(Fig. 1a, Supplemental Fig. S1a)**. These enriched DNA fragments, averaging 500 base pairs, are then cloned upstream of a fluorescent protein reporter preceded by a minimal promoter and a synthetic intron (Arnold et al. 2013) and followed by a 17 base pair random barcode allowing for $17x10^9$ possible barcodes (see Methods). Using CHEQ-seq we tested the enhancer activity of 1,526 TP53 ChIP-seq peaks obtained in MCF7 breast cancer cells treated with Nutlin-3a (Janky et al. 2014). Additionally, 94 promoters of housekeeping genes (HKG) were selected as control regions assuming that they drive stable reporter gene expression independent of any perturbation (Eisenberg and Levanon 2013). Also, 66 negative control regions in the genome were selected (see Methods, (Yip et al. 2012)). Unique to CHEQ-seq is long-read sequencing of the entire library to resolve the randomly formed enhancer-barcode pairs which after processing and removing redundancy yielded 20,751 distinct genomic fragments linked to 24,906 different barcodes (Methods and **Supplemental Fig. S1b, S2**).

To test which candidate enhancers are TP53-responsive, we transfected the enhancer-reporter library into MCF7 cells treated with Nutlin-3a, activating TP53 (p53-high) or containing a stable shRNA knock down for *TP53* (p53-off) (Phillips et al. 2010) **(Supplemental Fig. S1a)**. Reporter activity levels were determined by counting sequencing reads of barcoded cDNAs and were normalized both by the input library and by the re-extracted plasmid DNA (Methods and **Supplemental Fig. S3**). Of all the barcodes linked to a genomic fragment, we could measure 21,182 distinct barcodes representing 18,399 unique genomic fragments**,** of which 22.8% overlapped with the designed regions (81428-fold enrichment) **(Supplemental Table S1, Supplemental Fig. S1c)**.

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

We performed several quality control steps to assess whether the barcode expression represent accurate enhancer-reporter levels. Firstly, independent biological replicates show a very high correlation amongst barcodes within the p53-high condition (r^2 = 0.99). The correlation of the p53-off condition is much lower, as expected, since the majority of enhancers are inactive in the p53-off condition (r^2 = 0.56). The induced activity, measured as the fold-change between p53-high and p53-off also correlates between the two biological replicates (r^2 = 0.72) (**Supplemental Fig. S4**). Secondly, the expression levels of two different barcodes linked to the same enhancer region showed strong correlation of the fold-changes between p53-high and p53-off (r^2 = 0.73). Thirdly, different regions that overlap with the same designed region also show strong correlation (r^2 = 0.72) (**Supplemental Fig. S5**). Finally, the CHEQ-seq enhancer-reporter differential expression values are recapitulated when performing classical enhancer-luciferase assays **(Fig. S1b-f)**. These validation experiments confirm that the expression measured by CHEQ-seq are reliable and reproducible. Overall, CHEQ-seq allows cloning of hundreds of predefined enhancers into a complex barcoded library and generates reliable and reproducible reporter expression representing the functionality of a subset of these enhancers.

Only 40% of TP53 ChIP-seq peaks are functional enhancers

Having established CHEQ-seq as an accurate multiplex enhancer-reporter assay we turned to all the tested regions covering TP53 ChIP-seq peaks. Of the 1,526 targeted peaks, 1010 (66%) were represented by at least one captured sequence and non-ambiguous barcode. An additional 49 peaks were covered by randomly captured sequences. Of the 859 peaks that were covered sufficiently **(Fig. S2a)** (60% overlap, see **Supplemental Fig. S6, S7**), only 350 (40.7%) are significantly higher in p53-high

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

compared to p53-off conditions (called “positives”, adj. p-val < 0.05 and log2FC >= 1.5, (see Methods)). Interestingly, TP53 does not seem to directly repress enhancers since only ten peaks show significant down-regulation upon TP53 activation (down, adj. p-val < 0.05 and log2FC =< -1.5). Of the remaining ChIP-seq peaks, 337 are TP53-unresponsive (negatives), while another 162 show borderline expression patterns (greyzone) (**Fig. 2a**, **Supplemental Table S2**). Note that within the promoters of HKG, negative regions or non-specifically captured regions almost no TP53 inducible enhancers are found (0.95%, of HKG and negative controls and 1.6% of non-specific regions) (**Supplemental Fig. S6, S8**).

Next we compared the barcode reporter activity levels measured by CHEQ-seq to the existing multiplex enhancer-reporter method STARR-seq (Arnold et al. 2013), using the same captured fragments and the same transfection conditions (**Supplemental Fig. S9-S10, Supplemental Table S3**). The main difference in STARR-seq is that the tested regions are inserted at the 3’ of the transcription start site, causing the enhancers to transcribe themselves rather than relying on barcodes. STARR-seq yields reproducible expression values for 975 peaks, of which 242 peaks are TP53 responsive, 463 peaks as non-responsive and 216 peaks as greyzone **(Supplemental Fig. S11, Supplemental Table S4)**. The slightly smaller number of TP53 responsive elements compared to CHEQ-seq may possibly be due to differences in location of the enhancer regions within the reporter construct. STARR-seq also confirms that only very few ChIP-peaks (45 peaks) are repressed, and that the control groups do not show TP53-dependent changes **(Supplemental Fig. S12).** In total 600 ChIP-seq peaks have reporter activity data from both the CHEQ-seq and STARR-seq method. When looking at the subsets created independently for each method we see a highly significant overlap, with 190 out of 231 CHEQ-seq positives being labeled as

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

positives or greyzone in STARR-Seq (82.3%, chi-square p-value = 1.24 10^{-26}) (**Fig. 2b-c**). Interestingly, the small number of down-regulated ChIP-seq peaks identified by either method was not validated by the other method, suggesting that these were mainly false positives and that TP53 exclusively activates gene expression. In conclusion, multiplex enhancer reporter assays provide reproducible sets of direct TP53 enhancers and reveal that a relatively small subset of ChIP-seq peaks act as enhancers.

Unsophisticated TP53 enhancer logic

Using CHEQ-seq we discriminated true positive TP53 target enhancers from unresponsive yet TP53-bound regions. As these regions are all tested in the same episomal reporter environment, the enhancer-determining information should be contained within the DNA sequence. To compare the sequences between both sets we employed three motif discovery tools, namely i-cisTarget (Imrichová et al. 2015), RSAT peak motifs (Thomas-Chollier et al. 2011) and HOMER (Heinz et al. 2010), allowing for *de novo* motif discovery as well as enrichment of known motifs, using libraries of position weight matrices (PWM). All tools identified the TP53 motif as most over-represented in the TP53-responsive sequences, with highly significant p-values (HOMER p-values <= 10^{-322}, RSAT significance value = 72.30, i-cisTarget NES > 28). Note that, while some other motifs were found marginally overrepresented, these findings were not consistent across tools, were enriched at much lower p-values and occurred in a limited number of the positive regions (maximally 25%) **(Supplemental Fig. S13-S15)**. This suggests that TP53 mainly functions alone, without other regulatory factors co-binding at the DNA level. This is surprising as one of the proposed mechanisms for TP53 target specificity is through

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

the recruitment of co-regulatory transcription factors (Thornborrow and Manfredi 2001; Koutsodontis et al. 2001).

Previous reports indicate that small differences in the motif composition can influence binding affinity and determine target specificity (Godar et al. 2008; Inga et al. 2002; Wei et al. 2006; Smeenk et al. 2008). We therefore tested whether individual motifs, differing slightly from one another, perform differently in identifying true targets. We selected the ten best TP53 motifs based on their significance and low occurrence in the negative set. These ten motifs differ in length and composition but all retain the essential double C/G core of the TP53 binding site **(Fig. 3a).** When plotting the maximum score of each motif for both the positive and negative sets, the TP53 motif strongly distinguishes the sets with a marked absence of motifs amongst the non-responsive regions **(Fig. 3b)**. Only 45 out of 687 peaks were misclassified by CHEQ-seq, likely representing technical limitations of the experimental method itself. Indeed, the CHEQ-seq negatives that do have a TP53 motif are often identified as positive in STARR-seq (4 out of the 6). Vice versa, CHEQ-seq positives without TP53 binding site are often not identified as positive by STARR-seq (only 6 of the 17). Note that almost no down-regulated enhancers, identified by CHEQ-seq score for a TP53 motif **(Supplemental Fig. S16)**.

The classification performance of these motifs, and combinations thereof can be assessed using a Receiver Operating Characteristic (ROC) **(Fig. 3c-e)**. Overall the performance for the different motifs is very comparable, with multiple motifs yielding an area under the ROC curve (AUC) above 0.95. A 2-fold cross validation at the level of feature selection (i.e. *de novo* motifs) ensured that these models are not over-fitted (see Methods, **Supplemental Fig. S17**). Interestingly, accurate TP53 binding site prediction requires the full PWM of the TP53 tetramer, as one half-site alone has a

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

poor predictive performance (AUC = 0.81). In addition, and contrary to previous reports (Riley et al. 2008; Cook et al. 1995; Tokino et al. 1994), all TP53 motifs, both *de novo* and known, have no gap between the two half sites. As each of the individual motifs has a slightly different nucleotide composition flanking the C/G quadruple, we wondered whether scoring with a combination of all motifs would improve the predictive power (**Fig. 3f**). Surprisingly this was not the case as the combination of all motifs had the same predictive power as the best-scoring motif (AUC = 0.98). This suggests that the nucleotide composition of the TP53 motif can be largely captured by a single optimal PWM. Next, we tested whether homotypic clusters of TP53 sites were characteristic for TP53 responsiveness. Although previously suggested (Bourdon et al. 1997), our data suggests that a single TP53 site is sufficient to distinguish TP53-responsive from non-responsive enhancers (**Fig. 3f-g**). Although the presence of a TP53 binding site is predictive of enhancer activity, the strength of the binding site is not indicative of the quantitative levels of reporter, suggesting a binary on/off state of a TP53 enhancer (**Fig. 3h**). This is further confirmed by comparing the quantitative enhancer-reporter levels with the height of the ChIP-seq peak score, which also does not correlate beyond the on/off categories (**Fig. 3i, Supplemental Fig. S16, S18**). In conclusion, TP53 enhancers are unsophisticated in their architecture with the presence of a single TP53 binding site containing a double C/G core that is both necessary and sufficient to actively drive expression in a binary fashion.

Indirect ChIP-seq peaks have no regulatory function

The ChIP-seq peaks that were not directly bound by TP53 through a TP53 binding site showed no increased enhancer-reporter activity, nor any basal level of enhancer

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

activity in the reporter assay (**Fig. 4a**). To test whether indirect peaks may have another regulatory function, we first predicted which ChIP-seq peaks within the full set of 3634 TP53 ChIP-seq peaks (Janky et al. 2014) are likely directly bound based on the presence of a TP53 motif. To this end, we used a Random Forest model trained on the CHEQ-seq positive set, with the nine TP53 PWMs identified above (**Supplemental Fig. S19**). This classifier predicted 671 direct and 2963 indirect peaks. To investigate whether the indirect peaks could work as enhancers in their endogenous genomic context, we performed ChIP-seq against H3K27ac under the same conditions. These data confirm that only the directly bound peaks are enhancers with H3K27ac marks (**Fig. 4b**). From this figure, it can also be seen that the direct peaks are overall higher and wider than the indirect peaks, although the distributions overlap with each other (**Fig. 4b, Supplemental Fig. S20**). Additionally, while the directly bound regions show increased chromatin accessibility upon TP53 activation, this is not the case for indirectly bound regions (DNase-seq data under similar conditions (The ENCODE Project Consortium 2012; Thurman et al. 2012)) (**Supplemental Fig. S21**). Furthermore, whereas the direct peaks are located near TP53 target genes, as determined by gene annotation or by up-regulated gene expression, the indirect peaks are not enriched near putative target genes, and are often found to overlap coding exons (**Fig. 4c, Supplemental Fig. S22, S23 and Supplemental Table S5, S6**). Finally, direct peaks significantly overlap with long terminal repeats (LTRs), indirect peaks do not (**Supplemental Fig. S24**).

None of the above tested features suggest that the negative indirect peaks have a regulatory function. We therefore considered the possibility that they could represent crosslinking artifacts (Keren and Segal 2013; Waldminghaus and Skarstad 2010; Baranello et al. 2015). If so, then these artifacts should not be reproducible across

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

different experimental conditions. To test this, we turned to other publicly available datasets and collected 15 different TP53 ChIP-seq experiments performed in 7 different cell lines under different TP53-stimulating conditions (**Supplemental Table S7**) (Desantis et al. 2015; Botcheva and McCorkle 2014; Smeenk et al. 2011; Nikulenkov et al. 2012; Sammons et al. 2015; McDade et al. 2014; Hünten et al. 2015; Zeron-Medina et al. 2013; Sánchez et al. 2014). After calling peaks for each experiment, we categorized all peaks in each dataset as direct or indirect using the Random Forest model. Whereas the total number of peaks differs greatly between data sets, this difference can be mainly attributed to differences in the number of indirect peaks (**Fig. 4d**). Remarkably, in contrast to the direct peaks that are strongly conserved between experiments, the indirect peaks are largely unique to each data set (**Fig. S4e, Supplemental Fig. S25**), strongly suggesting a non-functional role. In conclusion, the presence of an unsophisticated TP53 binding model is predictive of functional enhancers amongst ChIP peaks, while remaining peaks have no obvious regulatory function and may represent crosslinking artifacts.

Genome-wide TP53 responsive enhancers are invariably used across cell types

Unsophisticated enhancer logic, with only a single high-scoring TP53 binding site being necessary and sufficient for a TP53 responsive enhancer, would predict invariable genomic binding across different cell types and experimental conditions. This simple model would thus contradict previous reports that proposed a direct role for TP53 binding sites and enhancers in differentially regulating cell-type dependent activation of its targets. To address this, we decided to test our simple model on a genomic scale and across cell types and treatment conditions. We first applied our previously trained Random Forest model on the entire human genome and predicted

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

21659 potential TP53 responsive enhancers. We then plotted the ChIP-seq signal across all publicly available datasets for all these binding sites. After testing several different clustering parameters, we found that the average coverage across binding sites within each generated cluster always converged onto three robust clusters (see Methods**, Supplemental Fig. S26**). A first cluster (strongly bound) with 1148 sites is preferentially bound by TP53 across all datasets. A second cluster (weakly bound), with 3147 sites is also shared across data sets, but with significantly lower binding signal. Finally, a third cluster with the remaining 17364 sites show no binding across the datasets (**Fig. 5**). This finding suggest that TP53 binds only a limited number of sites throughout the genome and that, importantly, in contrast to earlier reports, these sites are highly conserved across different experimental conditions and cell types **(Supplemental Fig. S27)**.

Sequence context and DNA shape of TP53 responsive enhancers are different, but not predictive

To explain why TP53 preferentially binds to only a small subset (19.8%) of all the genome-wide predicted binding sites, and to improve our predictive model, we investigated several local characteristics of the enhancer itself, including the sequence constraint across species, the strength of the TP53 motif, dinucleotide composition and the DNA shape flanking the motif. Additionally we investigated several characteristics of the genomic locus outside the enhancers, such as the presence of a nearby TATA or CpG promoter. Including these properties outside the enhancer is inspired by the possibility that binding of TP53 could perhaps be stabilized when it results in an effective target gene regulation, resulting in longer and thus higher frequency binding across cells in culture. In support of the three clusters (strong-

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

weak-unbound), we found that sequence constraint across vertebrate genomes is much higher for the strongly bound sites than for the weakly and unbound sites, corroborating their functional role (**Fig. 6a**). Note that we decided not to incorporate this feature into our predictive model because it is not a primary feature of the genome. We compared all other features, both local and global, by including them one by one, and in combinations, as features into our Random Forest classifier. Surprisingly, only the TP53 PWMs contributed significantly to the performance. In other words, the TP53 motif not only allows distinguishing direct from indirect ChIP-seq peaks (see above), but also further discriminates strongly bound sites from unbound sites in the genome (AUC=0.87, blue curve **Fig. 6b**). When investigating the feature importance in the Random Forest model, we found the TRANSFAC motifs M01655 and M01656 to have the highest weights (**Supplemental Fig. S28**). The poor predictive performance of the other sequence features is surprising because on average many of these features show distinct patterns between the strongly bound and unbound sites. For example, functional TP53 sites show a drop in the occurrence of A/T dinucleotide sequences (TT, AA, TA, and AT), around 100 bp on each side of the binding site (**Fig. 6c-d)**. Secondly, the strongly bound sites differ from unbound sites in DNA shape properties like propeller twist, helical twist and GC content, again around 100 bp each side of the binding site (**Fig. 6f-h**). Note that the weakly bound sites have an intermediate profile, having values for these features halfway between the strong and unbound site. This suggests that this cluster should indeed be considered as a separate group within the TP53 binding sites. To test whether additional unknown sequence features could play a role in determining strongly bound TP53 enhancers, we also trained a deep learning model directly on the bound versus unbound sequences, which automatically learns discriminative features (see

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

Methods). The classification performance is comparable to the random forest model using TP53 PWMs, which suggests that no additional features could be identified (**Fig. 6b**, yellow curve). In conclusion, a key role is played by the TP53 motif in defining functional binding, while other features contribute only marginally to the binding specificity.

Strength of binding site predicts quantitative TP53 binding

While the strongly bound and unbound clusters are clearly separated in their ability to bind TP53, the weakly bound sites exhibit intermediary characteristics. To avoid the arbitrary cut-offs of the clustering we instead ranked all the 21659 sites using the ChIP-seq coverage across each experiment, followed by a rank aggregation step yielding a final meta-ranking (see Methods). This meta-ranking strongly recapitulates the three clusters from above (**Fig. 7**). Interestingly, the ChIP-seq signals are quantitatively comparable between different experiments. Furthermore, decreasing ChIP-seq signals are strongly correlated with decreasing probability of TP53 binding both by the Random Forest and the Deep Learning models (**Fig. 7a,** black and purple curves). The ranked list is also correlated with the functionality, as shown by correlating H3K27ac signal, DNase-seq and GRO-seq data. This observation provides additional confirmation that the regions are ranked in decreasing functionality **(Fig. 7a).** Note that many of the weakly bound sites also show increased chromatin accessibility, which is expected when TP53 binds and displaces nucleosomes, but they have significantly less H3K27ac and GRO-seq marks (Su et al. 2015; Sammons et al. 2015). This functional difference is furthermore corroborated by the fact that genes located near highly ranked sites are located near TP53 related genes, while lower ranked sites show no Gene Ontology or pathway enrichment for TP53-related

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

processes **(Supplemental Table S8-S9)**. Nevertheless, weakly bound sites can function as enhancers, since 49% of weakly bound sites is positive in the CHEQ-seq or STARR-seq assay, compared to nearly 80% of the strong sites driving gene expression in the context of a reporter assay (**Fig. 7c** and **Supplemental Fig. S29, S30**). Taken together, these results suggest that functionality is predictable by the sequence, and decreases gradually by the strength of the TP53 binding site.

Discussion

The functional annotation of all regulatory elements in the non-coding part of the human genome is a key challenge in genome biology. Although biochemical events such as protein binding to the DNA occur very frequently, and such events cover more than 80% of the genome (The ENCODE Project Consortium 2012), only a fraction of these binding events is expected to be functional (Carvunis et al. 2015). In the context of the presented work, we consider a genomic region to be functional if it contributes in a deterministic and observable fashion to the regulation of gene expression. In our study, we found that among the 95077 observed binding events for TP53, across sixteen data sets covering diverse cell types and conditions, and among the more than 20,000 possible TP53 recognition sites in the genome, only a small fraction are bona fide TP53 responsive enhancers. Rather than deciding on a cut-off, we generated a ranking of all candidate sites based on the combination of the aggregated binding data and found that this meta-ranking correlates strongly with the strength of the TP53 binding site. This correlation can already be observed with a high-quality position weight matrix (PWM), but is even stronger for a trained classifier based on multiple PWMs or a deep learning model trained directly on strongly bound enhancer sequences. We provide a hub at the UCSC Genome

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

Browser, containing genome-wide binding across experiments and the scores for the model predictions (see Methods). This TP53 hub can serve as a reference for TP53-related future studies. Indeed, we argue that the observed consistency of direct TP53 binding across experimental conditions is so high that our ranking can be a good guide for the putative functionality of a given TP53 site, regardless of the experimental conditions. Note that the models presented here are derived from data sets after inducing TP53 in cell culture, both in cancer and normal cell lines. The question whether this unsophisticated enhancer model for TP53 also applies to other TP53-related functions, for example during *in vivo* development, remains to be investigated.

Only a small, yet predictable subset of experimentally determined binding events represent TP53 responsive elements. The other binding events (up to >90%, depending on the data set, see **Fig. 4f**) are presumable the result of cross-linking artifacts or other technical aspects of the ChIP method or its analysis. Similar observations have been made before, where binding of a transcription factor to the DNA can occur either due to the presence of its motif, when the binding is functional, or independently of the motif, when the binding is not functional. For example, Kvon *et al.* found that Twist is often bound to HOT regions in Drosophila, but only the binding in the mesoderm at the right time point, when Twist is actually expressed, corresponded to motif-dependent direct binding sites (Kvon et al. 2012).

In this study we have learned several new things about TP53. Firstly, our data supports the most common model for TP53 binding, namely that TP53 binds the DNA strictly as a tetramer, to a duplicate of the consensus palindromic responsive

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

element RRRCWWGYYY (R = purine; W = adenine/thymine; Y = pyrimidine), separated by a spacer of length N = 0, and not to single half-sites. Whereas previous reports argued that the sequence composition of the TP53 binding site may play a role in explaining context-dependent activation (Beckerman and Prives 2010; Weinberg et al. 2005; Szak et al. 2001), our results are consistent with other studies that reject this hypothesis (Smeenk et al. 2008; Wei et al. 2006). Another clear result from our experiments is that TP53 can only directly activate enhancers and in contrast to previous reports (Godar et al. 2008; Hoffman et al. 2002), we find no evidence of direct repression. Another intriguing finding is that TP53 seems to bind to chromatin independently of pre-existing nucleosome accessibility. For this we confirm earlier studies (Lidor Nili et al. 2010; Cui et al. 2011; Su et al. 2015). This finding supports the clutch-like model whereby transcription factors are able to displace nucleosomes, and are in competition with nucleosome binding. Finally we find that TP53 mostly acts alone as a TF bound to its target enhancers, although previous studies had suggested co-factorship at the DNA level (Thornborrow and Manfredi 2001; Koutsodontis et al. 2001). Note that non sequence-specific co-factors (e.g. P300) are likely to interact with TP53 at the protein level, independent of the DNA sequence, to recruit RNA polymerase and activate target gene transcription.

Enhancer sequences have recently been shown to share characteristic features beyond transcription factor binding sites, such as a particular distribution of Cs, Gs, and CpGs (Kwasnieski et al. 2012) or the presence of dinucleotide repeat motifs, such as CA, GA or CG (Yáñez-Cuna et al. 2014). The flanking nucleotides of our TP53 responsive elements also show an intriguing bias in nucleotide composition, which likely

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

represent preferential nucleosome binding positions. Although these patterns are clearly visible at the global level, these features do not have predictive power.

Among the ~21,000 potential TP53 binding sites, we find ~1000 regions that are strongly bound and represent *bona fide* responsive elements, regulating target gene expression, as effectors of the TP53 response. In addition, another subset of around ~3,000 sites show weak TP53 binding. This set is much less related to enhancer activity, gene expression, or nearby gene function. On the other hand, the recognition sites are more conserved in evolution and they have stronger PWM matches than the unbound. It is intriguing to speculate what the function of these sites could be. One possibility is that these sites could be preferentially bound when there is an excess of TP53 protein. Indeed, from a statistical point of view, given that TFs need "time" to find their functional binding sites, a cell needs to produce an excess of functional transcription factors to ensure that the entire functional *cistrome* remains occupied. Thus, at all times a fraction of molecules is undergoing fast binding turnover at non-functional sites (also called treadmilling (Lickwar et al. 2012)) where they become fixated during the ChIP protocol. Interestingly, we find that a quantitative relationship between prolonged residence and the strength of the TP53 binding site, rather than other additional sequence features. This idea adheres to the recently proposed clutch-like model of transcription factor binding (Lickwar et al. 2012).

Achieving high-confidence predictions at the genome-wide scale requires machine learning classifiers that are trained on large sets of positive and negative enhancer sequences. We, and others before us, have shown that massively parallel enhancer reporter assays can relatively quickly lead to such training sets, and usually lead to

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

exciting new insight into the *cis*-regulatory logic of enhancers (Arnold et al. 2013; Kwasnieski et al. 2012; White et al. 2013; Kheradpour et al. 2013; Melnikov et al. 2012). In our study we have tested long enhancer sequences, of several hundreds of base pairs. Earlier methods for massively parallel enhancer reporter assays (MPRA) often relied on oligonucleotide synthesis to generate sequences, thereby limiting the fragment size considerably below that of an average metazoan enhancer (~200 bp (LeProust et al. 2010) as compared to ~500-800 bp). This issue has recently been overcome by cloning captured fragments into a barcoded reporter assay (Shen et al. 2015). In addition, the STARR-seq method bypasses the issue of short input fragments by inserting randomly fragmented regions of the genome straight into a library downstream of a reporter gene rather than upstream (Arnold et al. 2013). STARR-seq was originally developed for Drosophila, where the genome size is manageable to clone in its entirety. However, considering that the human genome is 25 times larger, it presents a considerable challenge towards genome-wide assays. A solution to this is to preselect the input, as has been done in our study here, as well as recently in mice, using a STARR-seq like approach, called CapStarr-seq (Vanhille et al. 2015). The possibilities that these methodologies provide with regards to enhancer validation, both *in vitro* and *in vivo* (White et al. 2013; Shen et al. 2015), underscores the value and the need for such approaches in the field of regulatory genomics today.

Our case study of TP53 may seem peculiar in the sense that TP53 acts in isolation, without co-regulatory transcription factors that bind to the same enhancer. Although this was quite an unexpected finding, in retrospect this observation fits well within the cell-autonomous function of TP53. Indeed, upon DNA damage TP53 activates the appropriate target genes to either repair the damage, or launch the apoptotic program

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

(Beckerman and Prives 2010). Although we now have a better understanding of the TP53 "*cistrome*", it remains to be discovered how generic activation of the same set of enhancers is differentially steered in the currently operational gene regulatory network, and how the cellular context further contributes to the resulting responses like apoptosis or growth arrest.

The unsophisticated enhancer model, with only a single TP53 binding site, is to our knowledge the first report of such a new class of enhancers in the human genome. Previously, such enhancer models were used for compact genomes from bacteria, or sometimes yeast, but not for multicellular organisms. Plant and metazoan enhancers are usually classified as enhanceosome or billboard model (Arnosti and Kulkarni 2005), but these types of models are based on the combinatorial action of multiple transcription factors to reach specificity. Whereas developmental enhancers need to integrate multiple signals and environmental queues (e.g., signaling gradients), each cell can be considered independently responsible for its own genome integrity. Such a model may be valid for additional cell-autonomous factors, such as the cell-cycle related factors of the E2F family. Thus, we consider the TP53 enhancer model to represent a new class of unsophisticated, single-factor metazoan enhancers.

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

Material and methods

CHEQ-seq plasmid and bait design

A super core promoter, a synthetic intron (Arnold et al. 2013) and a venus reporter gene (Roure et al. 2007) were inserted between the in the KpnI and XbaI restriction sites of the pGL4.23 plasmid (Promega cat No E8411). Barcodes were incorporated using a inverse PCR into the modified pGL4.23 backbone with AscI (see supplemental materials). 120 ng of plasmid was electroporated per 20 µl of electrocompetent cells (Invitrogen, cat No C6400-03) and extracted using a Giga prep (Qiagen no. 12191). ChIP-seq peaks for TP53 were called against input (GSE47043) as described before (Janky et al. 2014) and were filtered for centromere and telomere regions and ranked based on their peak score. The top 1700 regions were selected for bait design as described in supplemental Materials and Methods.

Generating the CHEQ-seq and STARR-seq libraries

Genomic DNA was extracted, adapter ligated and amplified. Targeted regions were captured with the MYbaits protocol (Custom bait libraries, MYcroarray). At least three captures were performed and pooled after purification. The CHEQ-seq plasmid containing the barcode pool was linearized and combined with ~250 ng input DNA in a total of 4 infusion reactions (Clontech). The recombined library was precipitated overnight and transformed at 100 ng per 20 µl electrocompetent cells. See supplemental Materials and Methods for additional details.

Cell work and extractions

MCF7 cells (*TP53* wild type or *TP53* knock down) were cultured and transfected as described before (Janky et al. 2014). DNA and RNA was extracted and cDNA prepared according to the manufacturer's guidelines. See Supplemental Materials and Methods for more details.

Library preparations

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

cDNA or (extracted) plasmid DNA was amplified with two rounds of PCR using the Phusion High-fidelity PCR master mix (cat no M0532S, PCR details in Supplemental Materials and Methods). The library was purified using the AMPure XP Beads (Beckman Coulter, cat no A63880) and sequenced on the Illumina HiSeq2500 platform. For STARR-seq the DNA and cDNA libraries were created as described before (Arnold et al. 2013). For PacBio sequencing**,** the input library was amplified using the Phusion High-fidelity PCR master mix (cat no M0532S, PCR details in supplemental Materials and Methods). The linear fragments of around 2-3 kb were purified using AMPure XP Beads (Beckman Coulter, cat no A63880). Libraries were subsequently sent for SMRT PacBio sequencing (Pacific Biosciences).

Data processing and subset determination for both CHEQ-seq and STARR-seq libraries are described in Supplemental Materials and Methods.

Motif discovery and motif scoring

HOMER (Heinz et al. 2010) and RSAT peak motifs (Thomas-Chollier et al. 2011) were run on the positive set with the negative set as background. i-cisTarget (Imrichová et al. 2015) was run on the positive set. For HOMER, length of the motif was set at length 19 or 20 (-len) to allow for discovery of the consensus TP53 motif in the *de novo* option. Motifs were selected based on their overall performance and low occurrence in the negative set (see supplemental Materials and Methods). For each obtained motif or the combination of all 10 motifs Cluster-Buster (Frith et al. 2003) was used to score the positive and negative sets using –c 0 and –m 0. The highest motif score (or CRM score) for each region was obtained and used to determine the predictive value of each motif to classify regions into positives or negatives. In short, the sensitivity and specificity for each motif was calculated and the area under the receiver operating characteristic curve (AUC) determined.

ChIP-seq, RNA-seq and public data

ChIP-seq against H3K27ac was performed and analyzed as described before (Verfaillie et al. 2015). RNA-seq for MCF7 *TP53* knock down was extracted and performed as described

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

previously (Janky et al. 2014). The collected public ChIP-seq data against TP53 are summarized in Supplemental Table S7. See supplemental Materials and Methods for a more extended methodology and processing of the data.

Random Forest model and Feature-vector representation

Different Random Forest models were generated (see supplemental Materials and Methods). As Random Forest implementation we used the scikit-learn Python package. Each classifier uses an ensemble of 151 decision trees. The parameter max_features (responsible for number of features to consider when looking for the best split) was set to sqrt (number of features). To calculate the feature importance we used the Gini impurity criterion averaged across trees, using the whole training data. The quality of each model was estimated in 5-fold cross-validations. For each PWM the motif score was calculated employing a Hidden Markov Model as implemented in Cluster-Buster (Frith et al. 2003). Number of coding genes and lncRNAs was calculated using bedTools and a custom bash script. The file with TSSs of genes was downloaded from the UCSC Genome Browser. High confident subsets of lncRNAs were downloaded from LNCipedia (Volders et al. 2015). Files with the positions of promoters with TATA-box and/or GpC islands have been downloaded from the FANTOM5 resource (Lizio et al. 2015, 5).

seqMINER, clustering and DNA shape

BAM files of all public data (15 samples) and in-house data (see GSE47043) were loaded into seqMINER (Ye et al. 2011). A BED file with all predicted TP53 binding sites was loaded. Alternatively in-house TP53 ChIP-seq data, H3K27ac or DNase-seq BAM files were loaded and compared across the positive and negative CHEQ-seq peaks. The flanking area was set at 2000 bp around the binding site. Heatmaps show the raw tag count coverage from each BAM file for each input site or peak. Determination of clustering is described in supplemental Materials and Methods (**Supplemental Fig. S26**). DNA shape data indicating Helix Twist (HelT) and Propellor Twist (ProT) for HG19 were downloaded from

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

ftp://rohslab.usc.edu/hg19/ in bigWig format (Chiu et al. 2015) and analyzed as described in supplemental Materials and Methods.

Prediction of TP53 binding using Deep Learning

The network was trained using the RMSprop algorithm for Stochastic Gradient Descent with 100 training samples in each mini-batch and binary cross-entropy loss function for minimization. the Keras 0.2.0 library (https://github.com/fchollet/keras) with the Theano 0.7.1 backend was used for implementation. Calculations have been performed with NVIDIA K40c accelerator. The regularization parameters are: dropout proportion (fraction of outputs randomly set to 0) for layer 2: 10%; layer 3: 10%; layer 6: 50%; all other layers: 0%. The details of the CNN model architecture are listed in Supplemental Materials and Methods.

Additional information

Data access

Data generated for this study have been submitted to NCBI Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/) under accession number GSE76657.

The predicted genome-wide TP53 binding sites and combined ChIP-seq data from all experiments used in this work are available as a track hub (http://ucsctracks.aertslab.org/p53/hub.txt). To activate this track hub, go to My Data, Track hubs in the UCSC Genome Browser menu, and provide this URL in “My Hubs”.

Acknowledgments

We thank Alexander Stark for providing the STARR-seq plasmid, for advice on the execution of the STARR-seq protocol, and for helpful discussions. We also thank Jean-Christophe Marine and Jonas Demeulemeester for helpful discussions and advice on the manuscript.

Funding

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

This work is funded by The Research Foundation - Flanders (FWO, www.fwo.be) (grants G.0640.13 and G.0791.14 to SA), Special Research Fund (BOF) KU Leuven (http://www.kuleuven.be/research/funding/bof/) (grant PF/10/016 and OT/13/103 to SA), Foundation Against Cancer (http://www.cancer.be) (grants 2010-154 and 2012-F2 to SA). AV is funded by a research grant from FWO and Kom op tegen kanker. ZKA is funded by a research grant from Kom of tegen kanker. KD and HI have PhD Fellowships from the agency for Innovation by Science and Technology (IWT, www.iwt.be). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Contributions

SA obtained funding, conceived and supervised the study. AV developed the CHEQ-seq methodology and performed all experiments with the help of VC. AV processed the data with the help of KD, MF, HI, and ZKA. DS performed the analyses related to machine learning. MF performed the DNA shape analysis. SA and AV wrote the manuscript.

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

Figure legends

Figure 1 – Overview of the CHEQ-seq reporter assay. (a) (I) Genomic DNA is sheared and custom baits are used to capture the regions of interest (ROI). (II) Captured ROIs are cloned into a reporter library consisting of a GFP-based reporter linked to a pool of 17 x10^9 barcodes. (III) The reporter library is transfected under various conditions after which the RNA of transcribed barcodes is extracted. (IV) Randomly coupled ROI-barcode couples are identified using PacBio sequencing and barcode expression is measured using Illumina short-read sequencing. **(b-e)** Four TP53 ChIP-seq peaks comparing the CHEQ-seq barcode level with luciferase activity of a manually cloned fragment. **(f)** CHEQ-seq and luciferase induction only agree when they both overlap with the ChIP-seq peak summit.

Figure 2 – CHEQ-seq identifies TP53 responsive enhancers. (a) MAplot showing the distribution of CHEQ-seq barcode expression levels versus the fold induction, showing a large number of activation (green), and almost no repression (red). Negatives and greyzone are defined by thresholds on fold change and significance (see Methods). **(b)** CHEQ-seq positives are mostly positive or greyzone in STARR-seq (first bar), while down-regulated regions could not be confirmed by STARR-seq (second bar). * = p-value < 0.001 as determined by Chi-square. **(c)** Positives from CHEQ-seq are also mostly (91%) positives in STARR-seq.

Figure 3 – TP53 enhancer model (a) 10 motifs were selected from motif discovery tools i-cisTarget, HOMER and RSAT peak motif. **(b)** Heatmap showing the best Cluster-Buster score for each peak of the non-responsive (negatives) and TP53-responsive (positives) subset. Discordance between PWM scores and CHEQ-seq are indicated with a yellow square. **(c-e)** Classification accuracy for each PWM shown as ROC curves. **(f)** Comparison between a single maximum score (red) versus a homotypic cluster of motifs (black dashed). **(g)** All positive regions with 1kb flanking sequence centered on the best scoring motif illustrating the absence of binding site clusters. **(h-i)** TP53 enhancers are binary on/off enhancers as shown

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

by the lack of correlation between the motif score and barcode expression (h), as well as the lack of correlation between the peak score and the barcode expression (i).

Figure 4 – Only directly bound peaks behave as enhancers (a) The average basemean expression values indicate that the direct and indirect peaks show significantly different reporter activity levels (p-val = $3.28x10^{-51}$). **(b)** Comparison of TP53 ChIP-seq signal and H3K27ac ChIP-seq signal between positives and negatives. Peaks are extended to 2000bp each side. Heatmaps show the raw tag count coverage per peak. **(c)** The average differential expression of genes near (<20kb) peaks (asterisk indicates p-val of $4.73x10^{-8}$). **(d)** Comparison of the proportion of direct versus indirect ChIP-seq peaks across 15 publicly available TP53 ChIP-seq datasets. Experiments are ordered along the x-axis based on total number of peaks called. **(e)** Directly bound peaks agree, but indirectly bound do not, between in-house ChIP-seq peaks and other data sets. The percentage overlap is compared to the in-house peaks.

Figure 5 – TP53 binding is conserved across data sets. 15 public data sets containing ChIP-seq against TP53 under various conditions were collected and remapped (also see **Supplemental Table S7**). 21649 predicted TP53 binding sites throughout the genome are clustered based on their coverage across all data sets and can be subdivided into shared strong (green), shared weak (yellow) and shared unbound (red) regions. On the bottom, individual aggregation plots show the coverage for each cluster per sample.

Figure 6 –DNA features of TP53 responsive enhancers. (a) Sequence constraint (phastCons) of the DNA sequence around the predicted TP53 binding sites for the three classes (strongly bound, weakly bound, unbound). Inset: zoom in of the TP53 binding sites shows the highest conservation around the core C and G nucleotides. **(b)** Different features and different machine learning methods were tested individually and in combination for there ability to discriminate strongly bound from unbound binding sites. **(c-d)** Dinucleotide composition of the 800 bp sequence around the binding sites. Bound sequences (c) show

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

depletion of TT and AA (blue lines) and AT and TA (grey-black lines) at 100 bp flanking the binding site compared to unbound sequences (d). **(e-g)** DNA shape features within 700 bp sequences around the predicted binding sites. Grey region ~100 bp away from the binding site shows the strongest differences between bound and unbound sites.

Figure 7 – Quantitative prediction of functional TP53 sites. (a) Heatmap showing the TP53 ChIP-seq coverage across all 16 datasets for the ~21000 predicted sites, after rank aggregation, alongside H3K27ac, DNase and GRO-seq status in untreated and TP53-stabilizing conditions. The color gradient on the left indicates the original clusters from Figure 5. Smoothed scores on the right (DL = deep learning; RF = Random Forest; PWM score = position weight matrix score) show gradual decline with the meta-ranking. **(b)** For each binding site within the clusters the closest gene within 20kb was assigned and the average differential expression calculated using RNA-seq data. Only strongly bound sites associate with genes that are up-regulated upon TP53 stimulation compared to random control (p-val * = 2.45×10^{-10}). **(c)** CHEQ-seq barcode reporter increase is also correlated with the level of binding , as compared to random regions as control (p-val * = 9.83×10^{-49}, ** = 4.06×10^{-27}, *** = 3.55×10^{-5}).

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

References

Arnold CD, Gerlach D, Stelzer C, Boryń ŁM, Rath M, Stark A. 2013. Genome-Wide Quantitative Enhancer Activity Maps Identified by STARR-seq. *Science* **339**: 1074–1077.

Arnosti DN, Kulkarni MM. 2005. Transcriptional enhancers: Intelligent enhanceosomes or flexible billboards? *J Cell Biochem* **94**: 890–898.

Banerji J, Rusconi S, Schaffner W. 1981. Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. *Cell* **27**: 299–308.

Baranello L, Kouzine F, Sanford S, Levens D. 2015. ChIP bias as a function of cross-linking time. *Chromosome Res Int J Mol Supramol Evol Asp Chromosome Biol*.

Beckerman R, Prives C. 2010. Transcriptional Regulation by P53. *Cold Spring Harb Perspect Biol* **2**: a000935.

Botcheva K, McCorkle SR. 2014. Cell context dependent p53 genome-wide binding patterns and enrichment at repeats. *PloS One* **9**: e113492.

Bourdon JC, Deguin-Chambon V, Lelong JC, Dessen P, May P, Debuire B, May E. 1997. Further characterisation of the p53 responsive element--identification of new candidate genes for trans-activation by p53. *Oncogene* **14**: 85–94.

Carvunis A-R, Wang T, Skola D, Yu A, Chen J, Kreisberg JF, Ideker T. 2015. Evidence for a common evolutionary rate in metazoan transcriptional networks. *eLife* **4**.

Chiocchetti A, Tolosano E, Hirsch E, Silengo L, Altruda F. 1997. Green fluorescent protein as a reporter of gene expression in transgenic mice. *Biochim Biophys Acta BBA - Gene Struct Expr* **1352**: 193–202.

Chiu T-P, Yang L, Zhou T, Main BJ, Parker SCJ, Nuzhdin SV, Tullius TD, Rohs R. 2015. GBshape: a genome browser database for DNA shape annotations. *Nucleic Acids Res* **43**: D103–D109.

Cook JL, Ré RN, Giardina JF, Fontenot FE, Cheng DY, Alam J. 1995. Distance constraints and stereospecific alignment requirements characteristic of p53 DNA-binding consensus sequence homologies. *Oncogene* **11**: 723–733.

Cui F, Sirotin MV, Zhurkin VB. 2011. Impact of Alu repeats on the evolution of human p53 binding sites. *Biol Direct* **6**: 2.

Dailey L. 2015. High throughput technologies for the functional discovery of mammalian enhancers: New approaches for understanding transcriptional regulatory network dynamics. *Genomics*.

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

http://linkinghub.elsevier.com/retrieve/pii/S0888754315300070 (Accessed July 15, 2015).

Desantis A, Bruno T, Catena V, De Nicola F, Goeman F, Iezzi S, Sorino C, Gentileschi MP, Germoni S, Monteleone V, et al. 2015. Che-1 modulates the decision between cell cycle arrest and apoptosis by its binding to p53. *Cell Death Dis* **6**: e1764.

Eisenberg E, Levanon EY. 2013. Human housekeeping genes, revisited. *Trends Genet* **29**: 569–574.

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, et al. 2011. Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473**: 43–49.

Frith MC, Li MC, Weng Z. 2003. Cluster-Buster: finding dense clusters of motifs in DNA sequences. *Nucleic Acids Res* **31**: 3666–3668.

Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan K-K, Cheng C, Mu XJ, Khurana E, Rozowsky J, Alexander R, et al. 2012. Architecture of the human regulatory network derived from ENCODE data. *Nature* **489**: 91–100.

Godar S, Ince TA, Bell GW, Feldser D, Donaher JL, Bergh J, Liu A, Miu K, Watnick RS, Reinhardt F, et al. 2008. Growth-Inhibitory and Tumor- Suppressive Functions of p53 Depend on Its Repression of CD44 Expression. *Cell* **134**: 62–73.

Gordân R, Hartemink AJ, Bulyk ML. 2009. Distinguishing direct versus indirect transcription factor-DNA interactions. *Genome Res* **19**: 2090–2100.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol Cell* **38**: 576–589.

Hoffman WH, Biade S, Zilfou JT, Chen J, Murphy M. 2002. Transcriptional Repression of the Anti-apoptoticsurvivin Gene by Wild Type p53. *J Biol Chem* **277**: 3247–3257.

Ho J, Benchimol S. 2003. Transcriptional repression mediated by the p53 tumour suppressor. *Cell Death Differ* **10**: 404–408.

Hünten S, Kaller M, Drepper F, Oeljeklaus S, Bonfert T, Erhard F, Dueck A, Eichner N, Friedel CC, Meister G, et al. 2015. p53-Regulated Networks of Protein, mRNA, miRNA, and lncRNA Expression Revealed by Integrated Pulsed Stable Isotope Labeling With Amino Acids in Cell Culture (pSILAC) and Next Generation Sequencing (NGS) Analyses. *Mol Cell Proteomics* **14**: 2609–2629.

Imrichová H, Hulselmans G, Kalender Atak Z, Potier D, Aerts S. 2015. i-cisTarget 2015 update: generalized cis-regulatory enrichment analysis in human, mouse and fly. *Nucleic Acids Res* gkv395.

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

Inga A, Storici F, Darden TA, Resnick MA. 2002. Differential transactivation by the p53 transcription factor is highly dependent on p53 level and promoter target sequence. *Mol Cell Biol* **22**: 8612–8625.

Janky R 's, Verfaillie A, Imrichová H, Van de Sande B, Standaert L, Christiaens V, Hulselmans G, Herten K, Naval Sanchez M, Potier D, et al. 2014. iRegulon: From a Gene List to a Gene Regulatory Network Using Large Motif and Track Collections. *PLoS Comput Biol* **10**: e1003731.

Johnson RA, Ince TA, Scotto KW. 2001. Transcriptional repression by p53 through direct binding to a novel DNA element. *J Biol Chem* **276**: 27716–27720.

Jolma A, Yin Y, Nitta KR, Dave K, Popov A, Taipale M, Enge M, Kivioja T, Morgunova E, Taipale J. 2015. DNA-dependent formation of transcription factor pairs alters their binding specificity. *Nature* **527**: 384–388.

Junion G, Spivakov M, Girardot C, Braun M, Gustafson EH, Birney E, Furlong EEM. 2012. A transcription factor collective defines cardiac cell fate and reflects lineage history. *Cell* **148**: 473–486.

Keren L, Segal E. 2013. Fixated on fixation: using ChIP to interrogate the dynamics of chromatin interactions. *Genome Biol* **14**: 138.

Kheradpour P, Ernst J, Melnikov A, Rogov P, Wang L, Zhang X, Alston J, Mikkelsen TS, Kellis M. 2013. Systematic dissection of regulatory motifs in 2000 predicted human enhancers using a massively parallel reporter assay. *Genome Res* **23**: 800–811.

Kirschner K, Samarajiwa SA, Cairns JM, Menon S, Pérez-Mancera PA, Tomimatsu K, Bermejo-Rodriguez C, Ito Y, Chandra T, Narita M, et al. 2015. Phenotype specific analyses reveal distinct regulatory mechanism for chronically activated p53. *PLoS Genet* **11**: e1005053.

Koutsodontis G, Tentes I, Papakosta P, Moustakas A, Kardassis D. 2001. Sp1 plays a critical role in the transcriptional activation of the human cyclin-dependent kinase inhibitor p21(WAF1/Cip1) gene by the p53 tumor suppressor protein. *J Biol Chem* **276**: 29116–29125.

Kvon EZ, Stampfel G, Yáñez-Cuna JO, Dickson BJ, Stark A. 2012. HOT regions function as patterned developmental enhancers and have a distinct cis-regulatory signature. *Genes Dev* **26**: 908–913.

Kwasnieski JC, Mogno I, Myers CA, Corbo JC, Cohen BA. 2012. Complex effects of nucleotide variants in a mammalian cis-regulatory element. *Proc Natl Acad Sci U S A* **109**: 19498–19503.

Lee TI, Young RA. 2013. Transcriptional regulation and its misregulation in disease. *Cell* **152**: 1237–1251.

LeProust EM, Peck BJ, Spirin K, McCuen HB, Moore B, Namsaraev E, Caruthers MH. 2010. Synthesis of high-quality libraries of long (150mer)

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

oligonucleotides by a novel depurination controlled process. *Nucleic Acids Res* **38**: 2522–2540.

Lickwar CR, Mueller F, Hanlon SE, McNally JG, Lieb JD. 2012. Genome-wide protein-DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature* **484**: 251–255.

Lidor Nili E, Field Y, Lubling Y, Widom J, Oren M, Segal E. 2010. p53 binds preferentially to genomic regions with high DNA-encoded nucleosome occupancy. *Genome Res* **20**: 1361–1368.

Lizio M, Harshbarger J, Shimoji H, Severin J, Kasukawa T, Sahin S, Abugessaisa I, Fukuda S, Hori F, Ishikawa-Kato S, et al. 2015. Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome Biol* **16**. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4310165/ (Accessed December 22, 2015).

Maston GA, Landt SG, Snyder M, Green MR. 2012. Characterization of enhancer function from genome-wide analyses. *Annu Rev Genomics Hum Genet* **13**: 29–57.

McDade SS, Patel D, Moran M, Campbell J, Fenwick K, Kozarewa I, Orr NJ, Lord CJ, Ashworth AA, McCance DJ. 2014. Genome-wide characterization reveals complex interplay between TP53 and TP63 in response to genotoxic stress. *Nucleic Acids Res* **42**: 6270–6285.

Melnikov A, Murugan A, Zhang X, Tesileanu T, Wang L, Rogov P, Feizi S, Gnirke A, Callan CG, Kinney JB, et al. 2012. Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. *Nat Biotechnol* **30**: 271–277.

Menendez D, Nguyen T-A, Freudenberg JM, Mathew VJ, Anderson CW, Jothi R, Resnick MA. 2013. Diverse stresses dramatically alter genome-wide p53 binding and transactivation landscape in human cancer cells. *Nucleic Acids Res* **41**: 7286–7301.

Muller PAJ, Vousden KH. 2014. Mutant p53 in Cancer: New Functions and Therapeutic Opportunities. *Cancer Cell* **25**: 304–317.

Nikulenkov F, Spinnler C, Li H, Tonelli C, Shi Y, Turunen M, Kivioja T, Ignatiev I, Kel A, Taipale J, et al. 2012. Insights into p53 transcriptional function via genome-wide chromatin occupancy and gene expression analysis. *Cell Death Differ* **19**: 1992–2002.

O'Kane CJ, Gehring WJ. 1987. Detection in situ of genomic regulatory elements in Drosophila. *Proc Natl Acad Sci U S A* **84**: 9123–9127.

Panne D. 2008. The enhanceosome. *Curr Opin Struct Biol* **18**: 236–242.

Patwardhan RP, Hiatt JB, Witten DM, Kim MJ, Smith RP, May D, Lee C, Andrie JM, Lee S-I, Cooper GM, et al. 2012. Massively parallel functional dissection of mammalian enhancers in vivo. *Nat Biotechnol* **30**: 265–270.

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

Patwardhan RP, Lee C, Litvin O, Young DL, Pe'er D, Shendure J. 2009. High-resolution analysis of DNA regulatory elements by synthetic saturation mutagenesis. *Nat Biotechnol* **27**: 1173–1175.

Phillips A, Teunisse A, Lam S, Lodder K, Darley M, Emaduddin M, Wolf A, Richter J, Lange J de, Vries MV, et al. 2010. HDMX-L Is Expressed from a Functional p53-responsive Promoter in the First Intron of the HDMX Gene and Participates in an Autoregulatory Feedback Loop to Control p53 Activity. *J Biol Chem* **285**: 29111–29127.

Riley T, Sontag E, Chen P, Levine A. 2008. Transcriptional control of human p53-regulated genes. *Nat Rev Mol Cell Biol* **9**: 402–412.

Rinn JL, Huarte M. 2011. To repress or not to repress: This is the guardian's question. *Trends Cell Biol* **21**: 344–353.

Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, Ernst J, Bilenky M, Yen A, Heravi-Moussavi A, Kheradpour P, Zhang Z, Wang J, et al. 2015. Integrative analysis of 111 reference human epigenomes. *Nature* **518**: 317–330.

Roure A, Rothbächer U, Robin F, Kalmar E, Ferone G, Lamy C, Missero C, Mueller F, Lemaire P. 2007. A Multicassette Gateway Vector Set for High Throughput and Comparative Analyses in Ciona and Vertebrate Embryos ed. J.-N. Volff. *PLoS ONE* **2**: e916.

Sammons MA, Zhu J, Drake AM, Berger SL. 2015. TP53 engagement with the genome occurs in distinct local chromatin environments via pioneer factor activity. *Genome Res* **25**: 179–188.

Sánchez Y, Segura V, Marín-Béjar O, Athie A, Marchese FP, González J, Bujanda L, Guo S, Matheu A, Huarte M. 2014. Genome-wide analysis of the human p53 transcriptional network unveils a lncRNA tumour suppressor signature. *Nat Commun* **5**: 5812.

Shen SQ, Myers CA, Hughes AE, Byrne LC, Flannery JG, Corbo JC. 2015. Massively parallel cis-regulatory analysis in the mammalian central nervous system. *Genome Res* gr.193789.115.

Shlyueva D, Stampfel G, Stark A. 2014. Transcriptional enhancers: from properties to genome-wide predictions. *Nat Rev Genet* **15**: 272–286.

Smeenk L, van Heeringen SJ, Koeppel M, Gilbert B, Janssen-Megens E, Stunnenberg HG, Lohrum M. 2011. Role of p53 serine 46 in p53 target gene regulation. *PloS One* **6**: e17574.

Smeenk L, van Heeringen SJ, Koeppel M, van Driel MA, Bartels SJJ, Akkers RC, Denissov S, Stunnenberg HG, Lohrum M. 2008. Characterization of genome-wide p53-binding sites upon stress response. *Nucleic Acids Res* **36**: 3639–3654.

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

Smith RP, Taher L, Patwardhan RP, Kim MJ, Inoue F, Shendure J, Ovcharenko I, Ahituv N. 2013. Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. *Nat Genet* **45**: 1021–1028.

Stampfel G, Kazmar T, Frank O, Wienerroither S, Reiter F, Stark A. 2015. Transcriptional regulators form diverse groups with context-dependent regulatory functions. *Nature* **528**: 147–151.

Su D, Wang X, Campbell MR, Song L, Safi A, Crawford GE, Bell DA. 2015. Interactions of chromatin context, binding site sequence content, and sequence evolution in stress-induced p53 occupancy and transactivation. *PLoS Genet* **11**: e1004885.

Szak ST, Mays D, Pietenpol JA. 2001. Kinetics of p53 Binding to Promoter Sites In Vivo. *Mol Cell Biol* **21**: 3375–3386.

The ENCODE Project Consortium. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**: 57–74.

Thomas-Chollier M, Herrmann C, Defrance M, Sand O, Thieffry D, Helden J van. 2011. RSAT peak-motifs: motif analysis in full-size ChIP-seq datasets. *Nucleic Acids Res* gkr1104.

Thornborrow EC, Manfredi JJ. 2001. The tumor suppressor protein p53 requires a cofactor to activate transcriptionally the human BAX promoter. *J Biol Chem* **276**: 15598–15608.

Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, et al. 2012. The accessible chromatin landscape of the human genome. *Nature* **489**: 75–82.

Tokino T, Thiagalingam S, el-Deiry WS, Waldman T, Kinzler KW, Vogelstein B. 1994. p53 tagged sites from human genomic DNA. *Hum Mol Genet* **3**: 1537–1542.

Vanhille L, Griffon A, Maqbool MA, Zacarias-Cabeza J, Dao LTM, Fernandez N, Ballester B, Andrau JC, Spicuglia S. 2015. High-throughput and quantitative assessment of enhancer activity in mammals by CapStarr-seq. *Nat Commun* **6**. http://www.nature.com/ncomms/2015/150415/ncomms7905/abs/ncomms7905.html (Accessed August 6, 2015).

Verfaillie A, Imrichova H, Atak ZK, Dewaele M, Rambow F, Hulselmans G, Christiaens V, Svetlichnyy D, Luciani F, Van den Mooter L, et al. 2015. Decoding the regulatory landscape of melanoma reveals TEADS as regulators of the invasive cell state. *Nat Commun* **6**. http://www.nature.com/ncomms/2015/150409/ncomms7683/full/ncomms7683.html (Accessed April 25, 2015).

Volders P-J, Verheggen K, Menschaert G, Vandepoele K, Martens L, Vandesompele J, Mestdagh P. 2015. An update on LNCipedia: a database for annotated human lncRNA sequences. *Nucleic Acids Res* **43**: D174–180.

Downloaded from genome.cshlp.org on November 8, 2026 . Published by Cold Spring Harbor Laboratory Press

Vousden KH, Prives C. 2009. Blinded by the Light: The Growing Complexity of p53. *Cell* **137**: 413–431.

Waldminghaus T, Skarstad K. 2010. ChIP on Chip: surprising results are often artifacts. *BMC Genomics* **11**: 414.

Wang J, Zhuang J, Iyer S, Lin X, Whitfield TW, Greven MC, Pierce BG, Dong X, Kundaje A, Cheng Y, et al. 2012. Sequence features and chromatin structure around the genomic regions bound by 119 human transcription factors. *Genome Res* **22**: 1798–1812.

Wei C-L, Wu Q, Vega VB, Chiu KP, Ng P, Zhang T, Shahab A, Yong HC, Fu Y, Weng Z, et al. 2006. A Global Map of p53 Transcription-Factor Binding Sites in the Human Genome. *Cell* **124**: 207–219.

Weinberg RL, Veprintsev DB, Bycroft M, Fersht AR. 2005. Comparative Binding of p53 to its Promoter and DNA Recognition Elements. *J Mol Biol* **348**: 589–596.

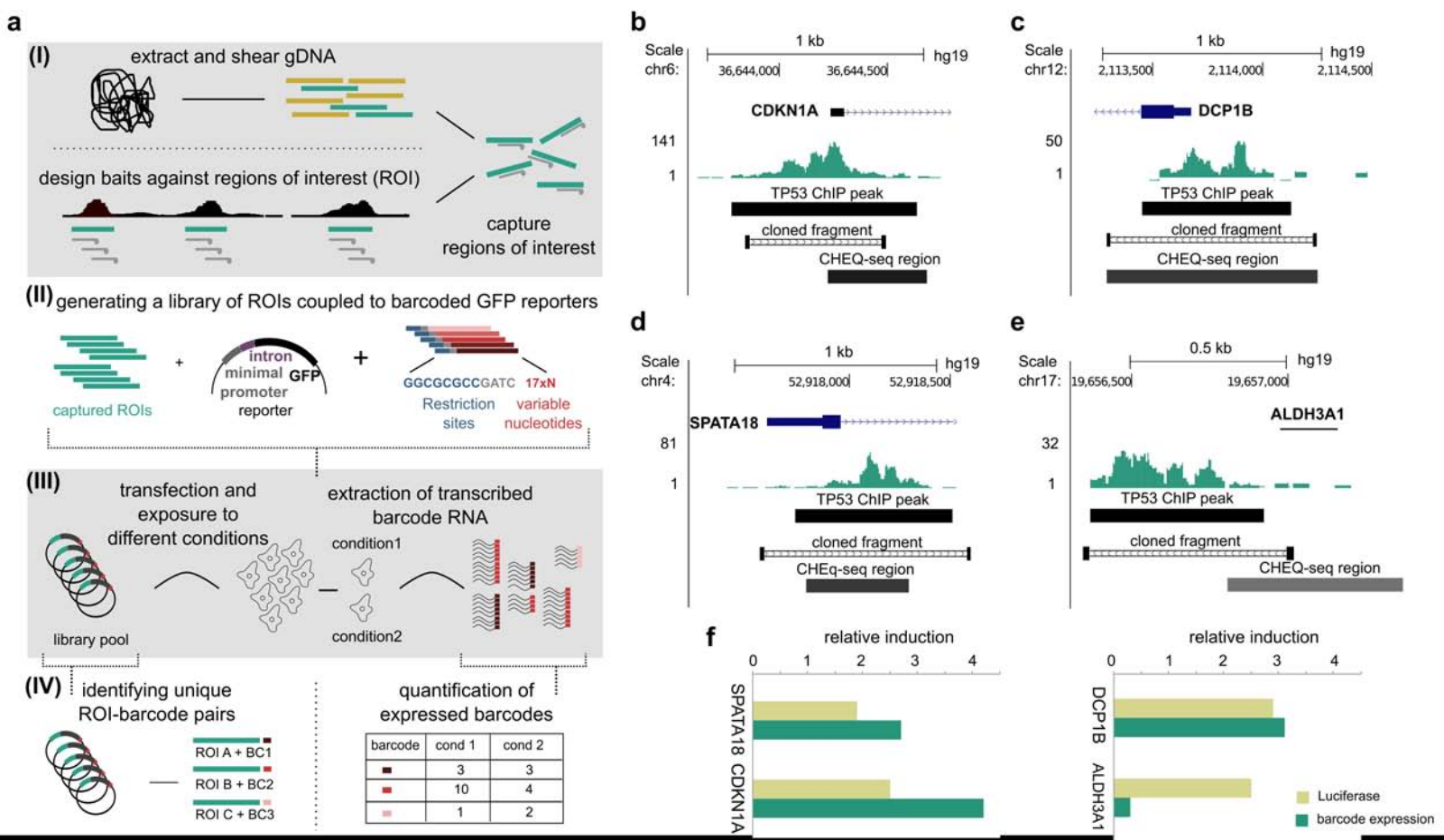
White MA, Myers CA, Corbo JC, Cohen BA. 2013. Massively parallel in vivo enhancer assay reveals that highly local features determine the cis-regulatory function of ChIP-seq peaks. *Proc Natl Acad Sci* **110**: 11952–11957.

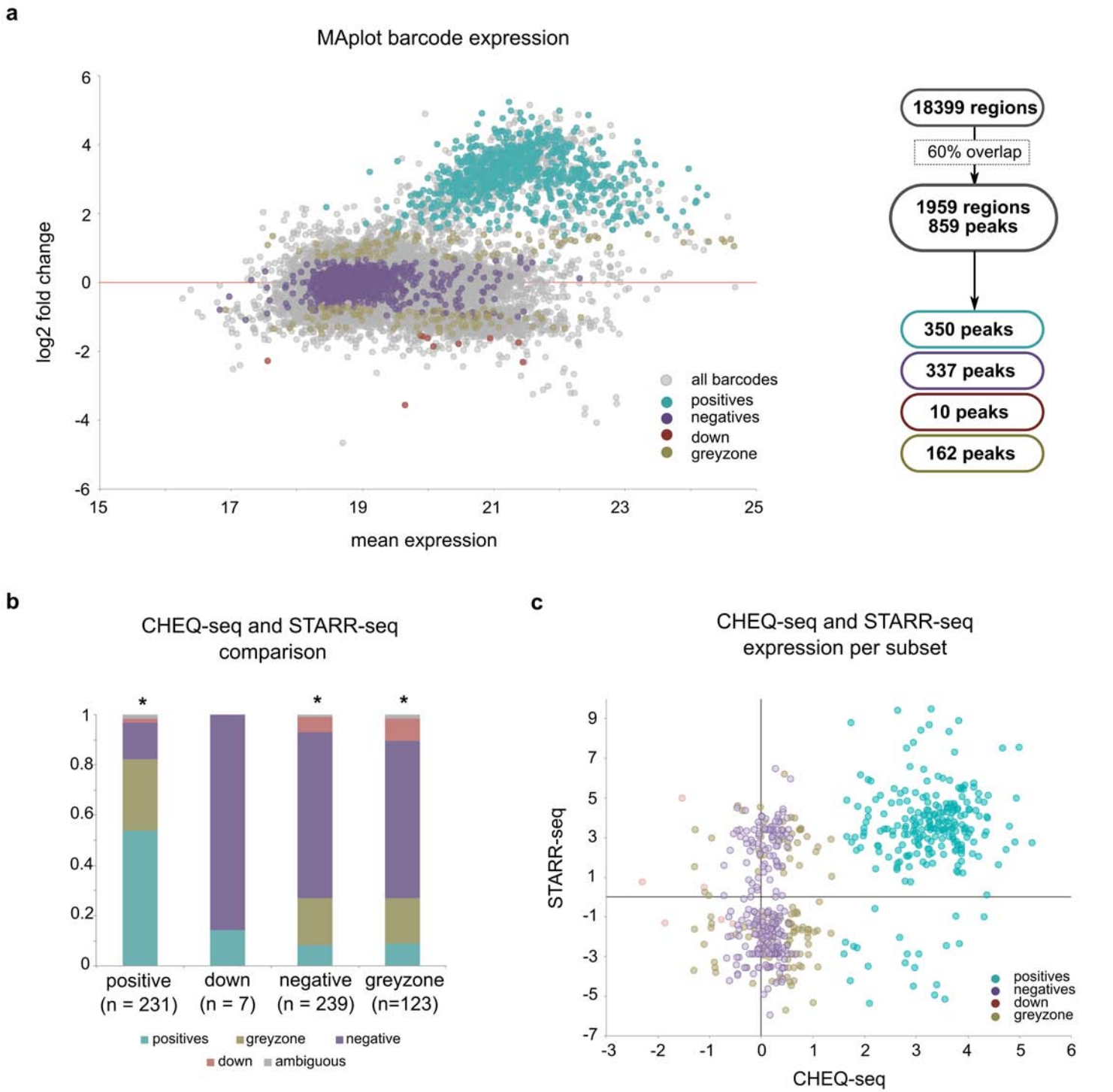
Yáñez-Cuna JO, Arnold CD, Stampfel G, Boryń ŁM, Gerlach D, Rath M, Stark A. 2014. Dissection of thousands of cell type-specific enhancers identifies dinucleotide repeat motifs as general enhancer features. *Genome Res* **24**: 1147–1156.

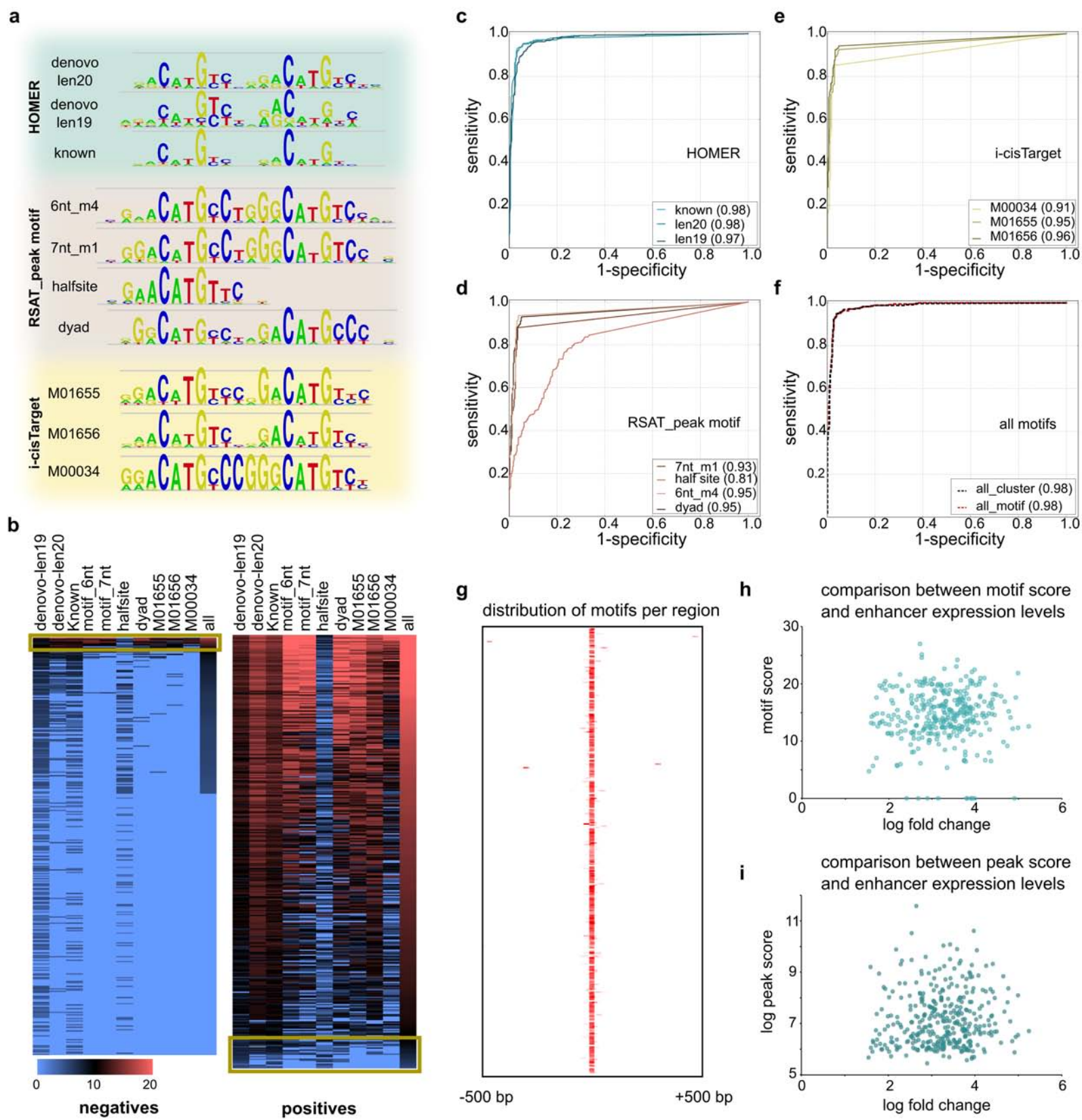
Ye T, Krebs AR, Choukrallah M-A, Keime C, Plewniak F, Davidson I, Tora L. 2011. seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic Acids Res* **39**: e35.

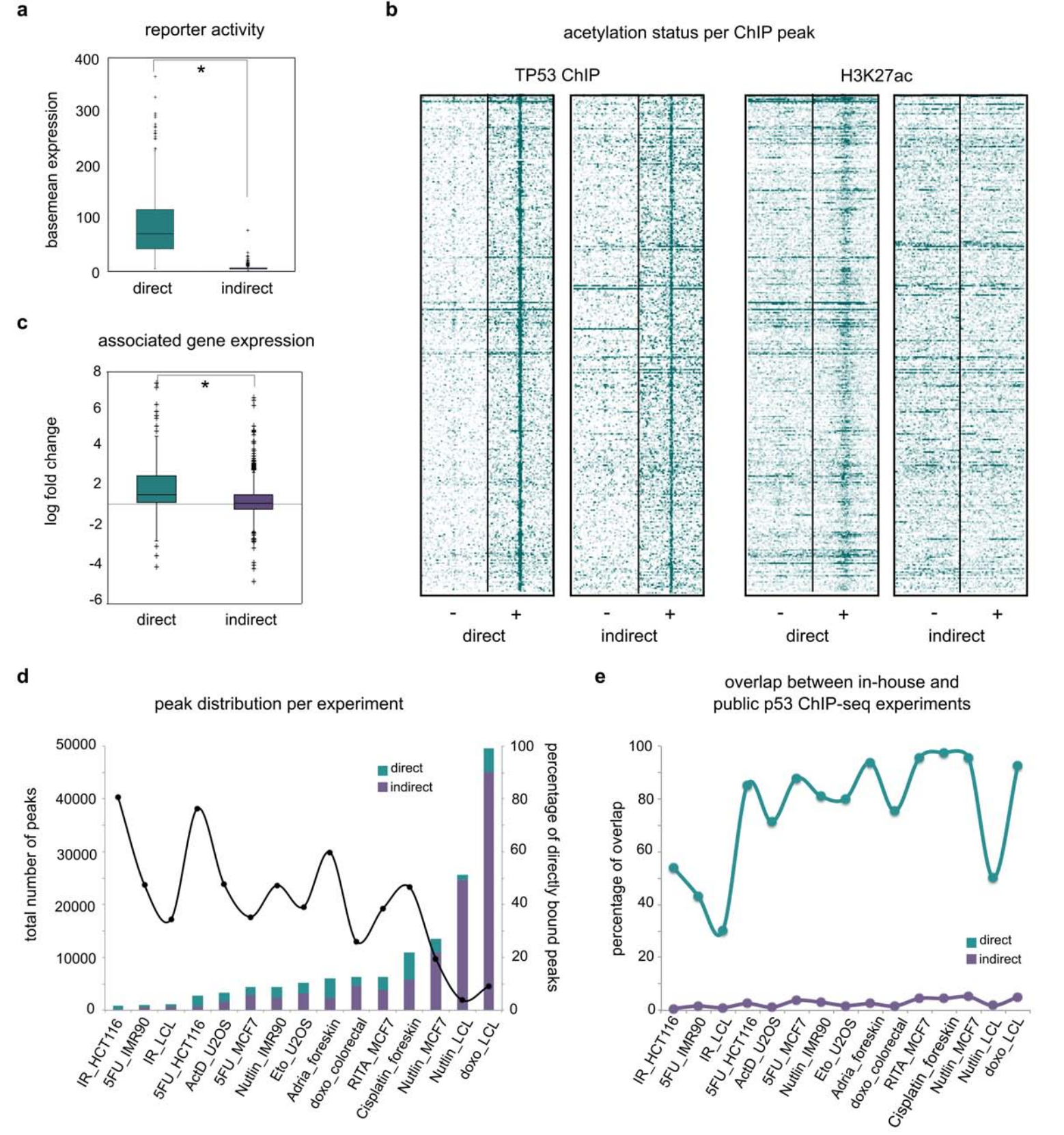
Yip KY, Cheng C, Bhardwaj N, Brown JB, Leng J, Kundaje A, Rozowsky J, Birney E, Bickel P, Snyder M, et al. 2012. Classification of human genomic regions based on experimentally determined binding sites of more than 100 transcription-related factors. *Genome Biol* **13**: R48.

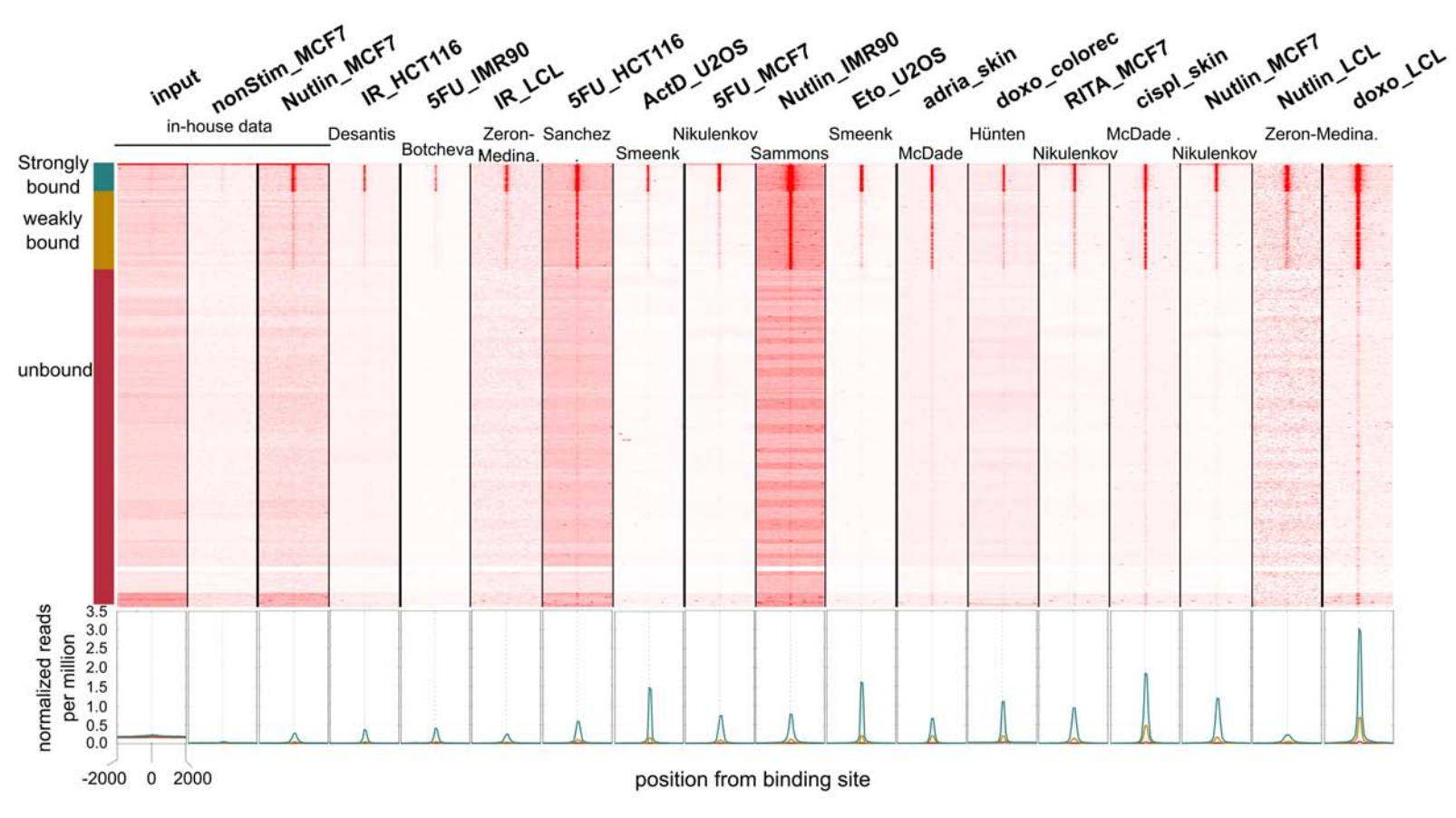
Zeron-Medina J, Wang X, Repapi E, Campbell MR, Su D, Castro-Giner F, Davies B, Peterse EFP, Sacilotto N, Walker GJ, et al. 2013. A polymorphic p53 response element in KIT ligand influences cancer risk and has undergone natural selection. *Cell* **155**: 410–422.

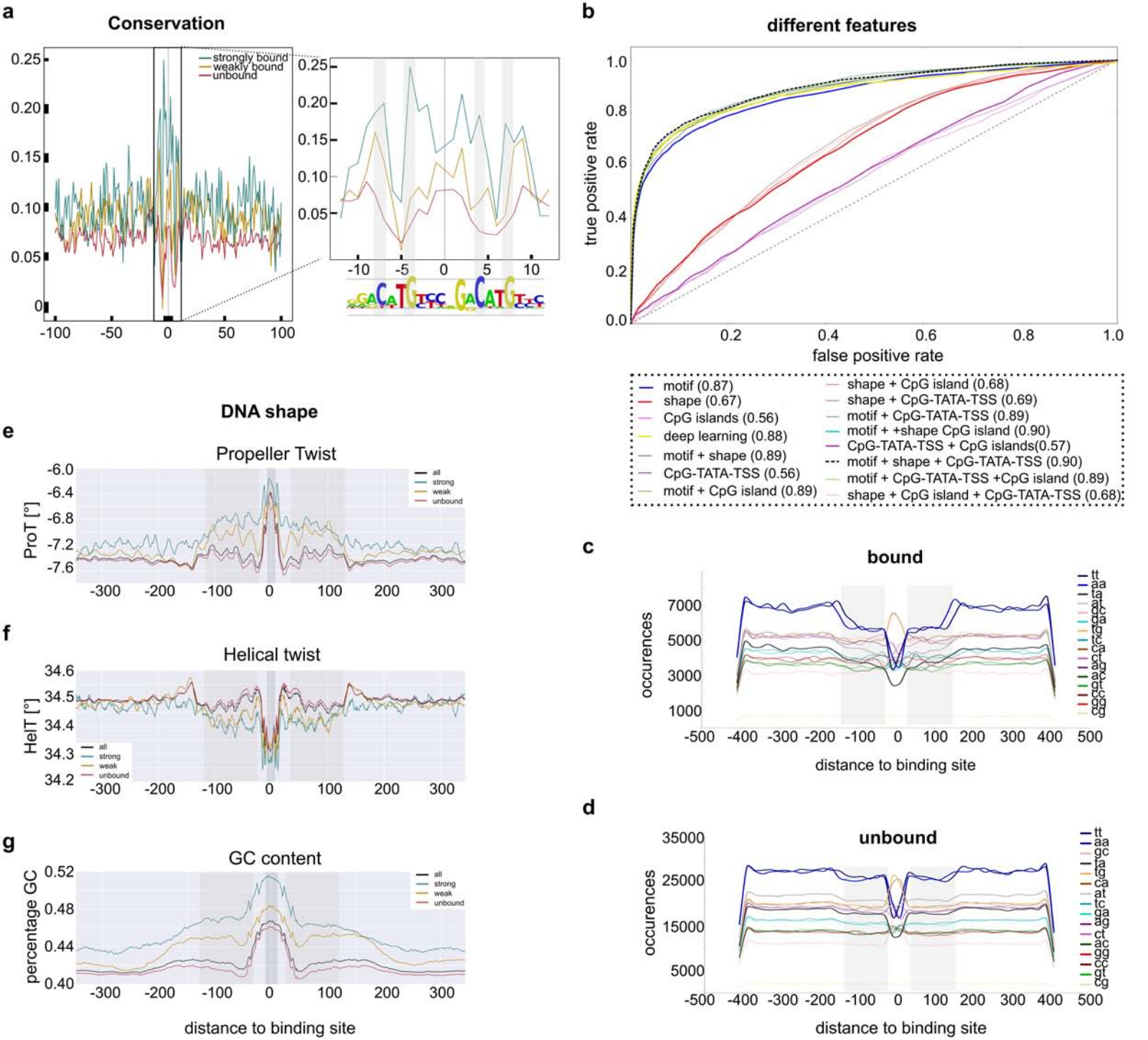

a
(I)
extract and shear gDNA
design baits against regions of interest (ROI)
capture
regions of interest
(II) generating a library of ROIs coupled to barcoded GFP reporters
captured ROIs
intron
minimal
promoter
GFP
reporter
GGCGCGCCGATC
17xN
Restriction
sites
variable
nucleotides
(III)
transfection and
exposure to
different conditions
extraction of transcribed
barcode RNA
condition1
condition2
library pool
(IV)
identifying unique
ROI-barcode pairs
ROI A + BC1
ROI B + BC2
ROI C + BC3
quantification of
expressed barcodes
barcode
cond 1
cond 2
3
3
10
4
1
2
b
Scale
chr6:
1 kb
hg19
36,644,000
36,644,500
CDKN1A
141
1
TP53 ChIP peak
cloned fragment
CHEQ-seq region
c
Scale
chr12:
1 kb
hg19
2,113,500
2,114,000
2,114,500
DCP1B
50
1
TP53 ChIP peak
cloned fragment
CHEQ-seq region
d
Scale
chr4:
1 kb
hg19
52,918,000
52,918,500
SPATA18
81
1
TP53 ChIP peak
cloned fragment
CHEq-seq region
e
Scale
chr17:
0.5 kb
hg19
19,656,500
19,657,000
ALDH3A1
32
1
TP53 ChIP peak
cloned fragment
CHEQ-seq region
f
relative induction
0
1
2
3
4
SPATA18
CDKN1A
DCP1B
ALDH3A1
Luciferase
barcode expression

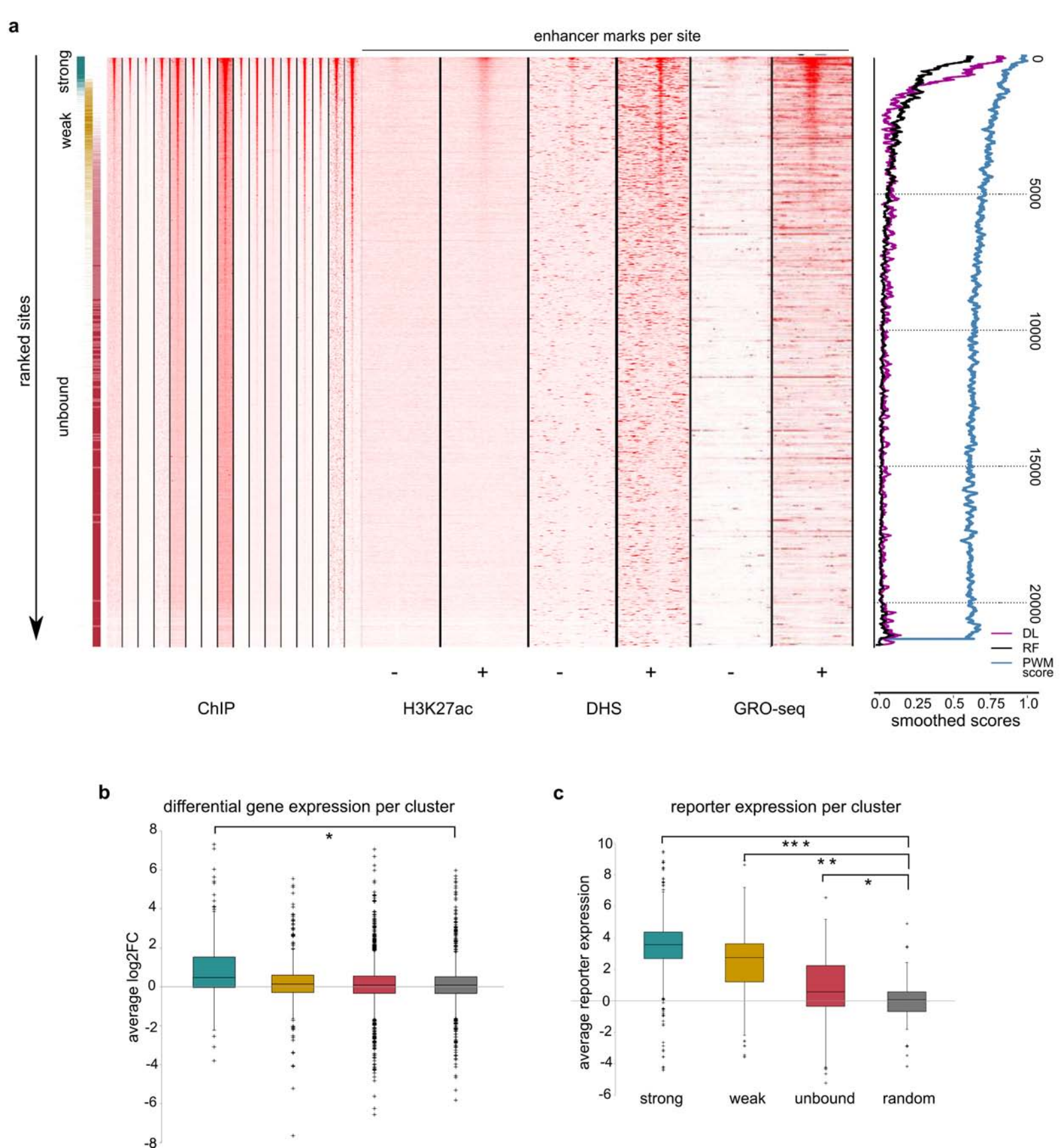

a
MAplot barcode expression
log2 fold change
mean expression
all barcodes
positives
negatives
down
greyzone
18399 regions
60% overlap
1959 regions
859 peaks
350 peaks
337 peaks
10 peaks
162 peaks
b
CHEQ-seq and STARR-seq
comparison
positive
(n = 231)
down
(n = 7)
negative
(n = 239)
greyzone
(n=123)
positives
greyzone
negative
down
ambiguous
c
CHEQ-seq and STARR-seq
expression per subset
STARR-seq
CHEQ-seq
positives
negatives
down
greyzone


a
HOMER
denovo len20
denovo len19
known
RSAT_peak motif
6nt_m4
7nt_m1
halfsite
dyad
i-cisTarget
M01655
M01656
M00034
b
denovo-len19
denovo-len20
Known
motif_6nt
motif_7nt
halfsite
dyad
M01655
M01656
M00034
all
0
10
20
negatives
positives
c
HOMER
sensitivity
1-specificity
known (0.98)
len20 (0.98)
len19 (0.97)
d
RSAT_peak motif
7nt_m1 (0.93)
half site (0.81)
6nt_m4 (0.95)
dyad (0.95)
e
i-cisTarget
M00034 (0.91)
M01655 (0.95)
M01656 (0.96)
f
all motifs
all_cluster (0.98)
all_motif (0.98)
g
distribution of motifs per region
-500 bp
+500 bp
h
comparison between motif score and enhancer expression levels
motif score
log fold change
i
comparison between peak score and enhancer expression levels
log peak score
log fold change


a
reporter activity
basemean expression
400
300
200
100
0
*
direct
indirect
b
acetylation status per ChIP peak
TP53 ChIP
H3K27ac
- +
direct
indirect
c
associated gene expression
log fold change
8
6
4
2
-2
-4
-6
*
direct
indirect
d
peak distribution per experiment
total number of peaks
percentage of directly bound peaks
direct
indirect
50000
40000
30000
20000
10000
0
100
80
60
40
20
IR_HCT116
5FU_IMR90
IR_LCL
5FU_HCT116
ActD_U2OS
5FU_MCF7
Nutlin_IMR90
Eto_U2OS
Adria_foreskin
doxo_colorectal
RITA_MCF7
Cisplatin_foreskin
Nutlin_MCF7
Nutlin_LCL
doxo_LCL
e
overlap between in-house and
public p53 ChIP-seq experiments
percentage of overlap
direct
indirect
100
80
60
40
20
0
IR_HCT116
5FU_IMR90
IR_LCL
5FU_HCT116
ActD_U2OS
5FU_MCF7
Nutlin_IMR90
Eto_U2OS
Adria_foreskin
doxo_colorectal
RITA_MCF7
Cisplatin_foreskin
Nutlin_MCF7
Nutlin_LCL
doxo_LCL


input
nonStim_MCF7
Nutlin_MCF7
IR_HCT116
5FU_IMR90
IR_LCL
5FU_HCT116
ActD_U2OS
5FU_MCF7
Nutlin_IMR90
Eto_U2OS
adria_skin
doxo_colorec
RITA_MCF7
cispl_skin
Nutlin_MCF7
Nutlin_LCL
doxo_LCL
in-house data
Desantis
Botcheva
Zeron-Medina.
Sanchez
Smeenk
Nikulenkov
Sammons
Smeenk
McDade
Hünten
Nikulenkov
McDade
Nikulenkov
Zeron-Medina.
Strongly bound
weakly bound
unbound
normalized reads per million
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
-2000
0
2000
position from binding site


a
Conservation
strongly bound
weakly bound
unbound
b
different features
true positive rate
false positive rate
motif (0.87)
shape (0.67)
CpG islands (0.56)
deep learning (0.88)
motif + shape (0.89)
CpG-TATA-TSS (0.56)
motif + CpG island (0.89)
shape + CpG island (0.68)
shape + CpG-TATA-TSS (0.69)
motif + CpG-TATA-TSS (0.89)
motif + +shape CpG island (0.90)
CpG-TATA-TSS + CpG islands(0.57)
motif + shape + CpG-TATA-TSS (0.90)
motif + CpG-TATA-TSS +CpG island (0.89)
shape + CpG island + CpG-TATA-TSS (0.68)
DNA shape
e
Propeller Twist
ProT [°]
all
strong
weak
unbound
f
Helical twist
HelT [°]
g
GC content
percentage GC
distance to binding site
c
bound
occurences
distance to binding site
d
unbound
occurences
distance to binding site


a
enhancer marks per site
strong
weak
unbound
ranked sites
0
5000
10000
15000
20000
DL
RF
PWM score
0.0 0.25 0.5 0.75 1.0
smoothed scores
- + - + - +
ChIP
H3K27ac
DHS
GRO-seq
b
differential gene expression per cluster
average log2FC
8
6
4
2
0
-2
-4
-6
-8
*
c
reporter expression per cluster
average reporter expression
10
8
6
4
2
0
-2
-4
-6
***
**
*
strong
weak
unbound
random