Cell

A Global Map of p53 Transcription-Factor Binding Sites in the Human Genome

Chia-Lin Wei,[1] Qiang Wu,[1] Vinsensius B. Vega,[1] Kuo Ping Chiu,[1] Patrick Ng,[1] Tao Zhang,[1] Atif Shahab,[2] How Choong Yong,[2] YuTao Fu,[3] Zhiping Weng,[3,4] JianJun Liu,[1] Xiao Dong Zhao,[1] Joon-Lin Chew,[1,6] Yen Ling Lee,[1] Vladimir A. Kuznetsov,[1] Wing-Kin Sung,[1] Lance D. Miller,[1] Bing Lim,[1,5] Edison T. Liu,[1] Qiang Yu,[1] Huck-Hui Ng,[1,6,*] and Yijun Ruan[1,*]

[1] Genome Institute of Singapore, Singapore 138672
[2] Bioinformatics Institute, Singapore 138671
[3] Bioinformatics Program, Boston University, Boston, MA 02215, USA
[4] Biomedical Engineering Department, Boston University, Boston, MA 02215, USA
[5] Harvard Institutes of Medicine, Harvard Medical School, Boston, MA 02115, USA
[6] Department of Biological Sciences, National University of Singapore, Singapore 117543
*Contact: nghh@gis.a-star.edu.sg (H.-H.N.); ruanyj@gis.a-star.edu.sg (Y.R.)
DOI 10.1016/j.cell.2005.10.043

SUMMARY

The ability to derive a whole-genome map of transcription-factor binding sites (TFBS) is crucial for elucidating gene regulatory networks. Herein, we describe a robust approach that couples chromatin immunoprecipitation (ChIP) with the paired-end ditag (PET) sequencing strategy for unbiased and precise global localization of TFBS. We have applied this strategy to map p53 targets in the human genome. From a saturated sampling of over half a million PET sequences, we characterized 65,572 unique p53 ChIP DNA fragments and established overlapping PET clusters as a readout to define p53 binding loci with remarkable specificity. Based on this information, we refined the consensus p53 binding motif, identified at least 542 binding loci with high confidence, discovered 98 previously unidentified p53 target genes that were implicated in novel aspects of p53 functions, and showed their clinical relevance to p53-dependent tumorigenesis in primary cancer samples.

INTRODUCTION

The recent completion of human genome sequencing (International Human Genome Sequencing Consortium, 2004) marked a major milestone in modern biology. The focus now has turned to the annotation of genomes for functional content, including gene-coding units and *cis*-acting regulatory elements that modulate gene expression (ENCODE Project Consortium, 2004). Gene expression in eukaryotic cells is controlled by regulatory elements that recruit transcription factors with specific DNA recognition properties. Thus, the identification of functional elements such as transcription-factor binding sites (TFBS) on a whole-genome level is the next challenge for genome sciences and gene-regulation studies.

Chromatin immunoprecipitation (ChIP) is a powerful technique for analyzing TFBS in living cells. The technology most commonly employed to map TFBS in a high-throughput manner is ChIP-on-CHIP. This strategy has been successfully applied for whole-genome localization analysis of TFBS in yeast (Ren et al., 2000). However, it has not been readily applicable for comprehensive survey of TFBS in human and other mammals due to the large size and complexity of these genomes. Recently, substantial progress has been reported (Kim et al., 2005b), in which high-density-tiling oligo arrays that cover 25% of the sequenced human genome were used to map active promoters. Nevertheless, ChIP-on-CHIP technology for mammalian systems has been developed on a limited scale. Most applications are so far restricted to promoter microarrays containing CpG islands or flanking sequences around transcription start sites and specific chromosome arrays (Horak et al., 2002; Weinmann et al., 2002; Cawley et al., 2004; Boyer et al., 2005). Despite considerable success, these partial genomic arrays have provided limited information.

Alternatively, immunoprecipitated DNA fragments from ChIP experiments can be cloned and sequenced (Weinmann et al., 2001; Hug et al., 2004). Although ChIP can enrich for TFBS-containing DNA fragments, a significant amount of background DNA will still be present in the immunoprecipitated DNA material. With a limited survey of the cloned ChIP DNA fragment pool, it is difficult to distinguish between genuine binding sites and noise without further molecular validation. However, with a larger sampling of the DNA pool, the sequencing-based approach has the potential to identify the DNA segments enriched by ChIP. The limitation of standard sequencing is the time and cost of

sequencing to achieve information saturation. The depth of coverage can be enhanced by alternative sequencing strategies such as serial analysis of gene expression (SAGE), which was originally developed for counting transcripts and was also recently applied to genome scanning for TFBS and histone modification (Impey et al., 2004; Kim et al., 2005a; Chen and Sadowski, 2005; Roh et al., 2005). Yet this monotagging approach suffers from the inherent ambiguity of mapping short monotags to the genome and the inability to distinguish true ChIP enrichment from amplified noise generated during molecular cloning.

To exploit the efficiency of sequencing short tags, to increase the information content, and to enhance the accuracy in mapping to the genome, we have developed a *p*aired-*e*nd di*t*ag (PET) method that extracts 36 bp signatures with 18 bp from the 5′ end and another 18 bp from the 3′ end of each cDNA clone, concatenates the PETs for efficient sequencing, and maps the PET sequences to the genome to demarcate gene-transcription boundaries (Ng et al., 2005). Conceptually, this strategy has the advantage over SAGE of higher information content, permitting the definitive mapping of the majority of tags to the genome, and is at least 30-fold more efficient than standard cloning and sequencing approaches. To develop and validate this strategy for applications in ChIP (which we now call ChIP-PET), we chose the p53 tumor suppressor, a sequence-specific DNA binding transcription factor.

As a transcription factor, p53 regulates the expression of genes involved in a variety of cellular functions, including cell-cycle arrest, DNA repair, and apoptosis (Vogelstein et al., 2000). In the past decade, numerous efforts have been made to identify p53-targeted genes through various gene-expression techniques, including microarray and SAGE combined with bioinformatics tools (Yu et al., 1999; Zhao et al., 2000; Kannan et al., 2001; Kho et al., 2004). To date, a large number of p53-responsive genes have been identified, mostly based on gene-expression data, yet only a small subset of these genes had direct binding evidence (el-Deiry et al., 1992; Yoon et al., 2002; Yin et al., 2003; Chen and Sadowski, 2005). Individual ChIP assay has been applied to validate p53 direct targets (Mirza et al., 2003) but is highly inefficient for identifying p53 targets on a global scale. Recently, the ChIP-on-CHIP approach was used to identify p53 binding sites in chromosomes 21 and 22 (Cawley et al., 2004). However, this experiment represented only 2% of the human genome and therefore provided only limited coverage for p53 binding. Herein, we describe the whole-genome localization of p53 TFBS, show the effectiveness of ChIP-PET for the identification of novel p53 target genes, and demonstrate the clinical relevance of a subset of these genes in tumorigenesis in vivo.

RESULTS

Mapping of p53 ChIP DNA Fragments by Paired-End diTags

The underlying concept of ChIP-PET analysis is to clone the immunoprecipitated chromatin fragments into a DNA library that captures the original representation of ChIP DNA fragments. The clones of the library are then converted into paired-end ditags that are concatenated and cloned as the final ChIP-PET library for sequencing analysis (10–15 PETs per sequence read). The PET sequences are then mapped to the genome to define the boundaries of the cloned ChIP fragments. We expect that the PETs derived from nonspecific fragments will be randomly distributed along the genome as background (PET singletons), while the PETs originating from the same locus containing the target binding site will overlap with each other in a PET cluster when mapped to the genome (Figure 1; see also Figure S1 in the Supplemental Data available with this article online).

With this efficient and specific readout, we characterized the p53 ChIP fragments generated from human HCT116 colorectal cancer cells treated with 5-fluorouracil (5-FU) for 6 hr, a condition known to activate p53 expression and transcription of its downstream targets (Kho et al., 2004). From approximately 40,000 sequencing reads, we produced 512,876 PET sequences, of which 75% (382,741) were mapped to single locations in the human genome. The rest of the PETs were either mapped to multiple locations, as they might derive from repetitive sequences, or were not mapped at all to the genome. These 382,741 PETs were further grouped as distinct PETs (Figure S2), representing 65,572 PET-identified ChIP DNA fragments ranging from 100–4,000 bp with an average length of 624 bp. Since the probability of generating identical DNA fragments by sonication during the ChIP procedure is assumed to be extremely rare, the original ChIP DNA fragments prior to cloning are most likely distinct from each other. Therefore, the redundant PETs were regarded as copies amplified from the original ChIP fragments during the cloning process, and the nonredundant distinct PETs were considered to represent the original ChIP DNA fragments. Based on the degree of PET redundancy, we estimated the total number of identifiable PET fragments in the original CHIP DNA material to be 82,659 by the Hill function (Kuznetsov, 2005) (Supplemental Data I-6). Therefore, by extrapolation, the 65,572 distinct PETs cover ~80% of the entire ChIP DNA-fragment pool captured in this library (Figure S3).

While the majority (61,270) of the distinct PET fragments were located in the genome discretely (classified as PET singletons), 4,302 (7%) PETs were found overlapping with others and were grouped into 1,766 PET clusters (Table 1). These PET-cluster-defined genomic loci represent potential p53 interaction sites in the genome. To assess the probability that PET overlapping was due to random chance, we performed a Monte Carlo simulation (Supplemental Data I-6). We estimated that 27% of the PET clusters with two overlapping members (hereafter referred to as PET-2 for PET clusters with two overlapping members, PET-3 for clusters with three overlapping members, and so forth), 2.3% of the PET-3 clusters, and 0.001% of the PET-4 clusters could result from random sampling (Table 1). This suggested that about 73% of PET-2 clusters and over 97% of the PET clusters with three or more overlapping members (PET-3+ clusters) most likely represent the real ChIP enrichment events.

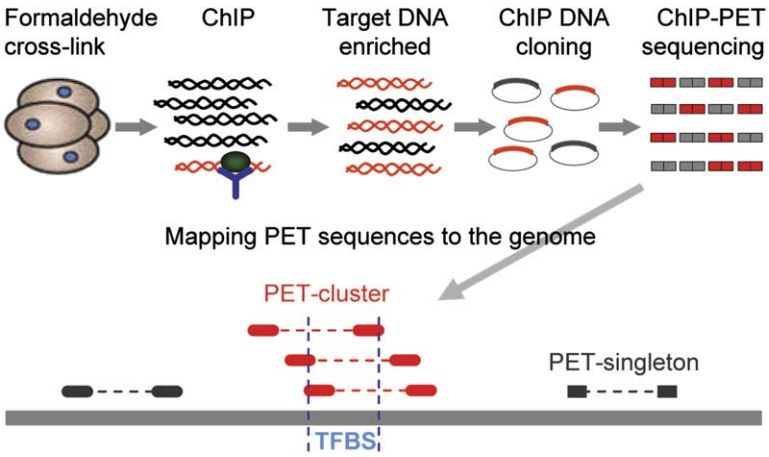

Figure 1. Schematic View of ChIP-PET Analysis

The ChIP DNA fragments were cloned into a plasmid vector. Plasmids were then converted into PETs for concatenation, cloning, and sequencing. The PET sequences were mapped to genome to demarcate the boundaries of DNA fragments. PET singletons were considered background, while overlapping PETs (PET clusters) were regarded as enrichment by the same immunoprecipitation events. The overlapping regions (PET overlap) in a PET cluster may, therefore, contain TFBS.

Furthermore, based on the frequency distribution of PET clusters by size (number of members in each PET cluster), we were able to establish a true PET-cluster curve that is distinctive from the potential noise curve and estimate the level of nonspecific PET clustering events (Supplemental Data I-6). By extrapolating the true PET-cluster curve, we projected that less than 36% (520 of 1443) of the PET-2 clusters but over 99% of the PET-3+ clusters might represent true enrichment by p53 ChIP (Table 1; Figure S4). Thus, using two statistical analyses, one based on simulation of randomness and the other based on the data-distribution curve, we concluded that PET singletons were most likely background,

Table 1. Enrichment of p53 Binding Loci by PET Clusters

PET Mapping and Statistic Estimation

	Total PETs	PET Singletons	PET Clusters							PET-3+ Clusters
			PET-2	PET-3	PET-4	PET-5	PET-6	PET-7	PET-8+	
Distinct PETs	65,572	61,270	2,886	471	252	175	168	91	259	1,416
Monte Carlo simulation	65,572	64,790	770.9	11.1	0.0034	<0.0001	<0.00001	<0.00001	<0.00001	
Random probability (%)			27	2.3	<0.001	<0.0001	<0.0001	<0.0001	<0.0001	
PET-defined loci		61,270	1,443	157	63	35	28	13	27	323
Goodness-of-fitting analysis		3,742	520	159	69	38	22	14	33	335

Enrichment of p53 Binding Motifs in PET Clusters

	Genome Background	PET Singletons	PET Clusters							PET-3+ Clusters
			PET-2	PET-3	PET-4	PET-5	PET-6	PET-7	PET-8+	
PET-defined loci	63,036[a]	61,270	1,443	157	63	35	28	13	27	323
p53PET prediction	430	968	219	96	53	25	26	11	24	235
p53PET prediction %	0.68	1.58	15.18	61.15	84.13	71.43	92.86	84.62	88.89	72.76
p53MH prediction	1,117	1,541	196	69	45	23	21	11	21	190
p53MH prediction %	1.77	2.51	13.5	43.95	71.43	65.71	75	84.62	77.78	58.82

The random probability of PET overlapping was calculated based on the simulated numbers versus the observed numbers in each category of PET clusters.

[a] The same number (61,270 + 1,766 = 63,036) and sizes (average 630 bp) of genomic DNA segments as the PET-defined loci were randomly extracted from the human genome assembly (hg17) as background.

the PET-2 clusters were enriched for p53 ChIP DNA fragments but included substantial noise, and the PET-3+ clusters were highly specific for p53 ChIP enrichment.

Verification of PET-Cluster-Identified p53 Binding Loci

To verify whether the genomic loci determined by PET clusters are associated with p53 interactions, we examined a list of 66 known p53-responsive genes for the localization of PET clusters (±100 kb around each curated gene in the human genome). These genes had been demonstrated to be activated by genotoxic treatment in HCT116 cells (Kho et al., 2004) or are well-known p53 targets (Polyak et al., 1997; Vogelstein et al., 2000). It is expected that some of these genes would be directly targeted by p53 binding and some secondary effectors. Forty-one of these sixty-six genes were localized by PET clusters, including twenty-three genes by PET-3+ clusters, eighteen by PET-2 clusters (Table S2), and three by multiple PET clusters. For instance, *CDKN1A* is a well-characterized p53 target gene encoding a cyclin-dependent kinase inhibitor (Kaeser and Iggo, 2002) with a confirmed p53 binding site in its promoter region. We found a PET-13 cluster within the first 2,600 bp of the promoter region, identifying 97 bp of overlap that coincided with the previously characterized p53 binding site (el-Deiry et al., 1993) (Figure 2A). Unexpectedly, we also found a PET-5 cluster located 11,447 bp further upstream of the *CDKN1A* transcription start site. The overlapping segment (153 bp) in the PET-5 cluster also contained a recognizable p53 binding motif. To specifically validate the localization of PET clusters in the 5′ region of *CDKN1A*, we scanned the entire 12,000 bp genomic span using the conventional ChIP quantitative PCR (ChIP-qPCR) assay. As illustrated in Figure 2B, both of the p53 binding loci were confirmed, and the genomic segments showing peak ChIP enrichment were superimposable on the PET overlapping regions (Figures 2A and 2B). More examples of PET clusters mapped to known p53 targets are shown in Figures S5 and S6.

The remaining 25 genes in this list either were not hit by any PETs or were hit only by PET singletons. Hence, over 62% (41 of 66) of known p53-responsive genes in this list were localized by PET clusters. This high matching rate of PET clusters to known p53-responsive genes is statistically significant (p value = $9e^{-14}$), suggesting that genomic loci determined by PET clusters are substantially enriched with reliable p53 binding sites. Furthermore, 16 out of the 25 p53-responsive genes not associated with PET clusters had no binding data in previous studies, suggesting that these genes are not p53 direct targets but secondary effectors in p53 regulation pathways. For the nine genes that had previous binding data but were missed by PET clusters, we conducted ChIP-qPCR assay for the previously known binding regions and found that the binding loci of three genes were significantly enriched by p53 ChIP, including one gene (*TRAF4*) hit by a PET singleton covering an authentic p53 consensus motif. The other six were marginally enriched and not statistically significant above background (Figure S7), including *PIG3* and *p53AIP1*, known for their low binding affinity for p53 protein (Kaeser and Iggo, 2002), and were therefore not easily detected by PET sequencing or other measurement. These results indicate that more than 93% (41 of 44) of p53 targets enriched by ChIP procedure in this study were identified.

For further validation, we randomly selected 40 genomic loci defined by PET-3+ clusters as target segments for ChIP-qPCR assay. All 40 loci (100%) showed significant enrichment (Figure 2C), indicating that these regions are true p53 binding targets.

Together, based on the high percentage of PET-cluster hits to known p53 targets, the precise localization of many previously known p53 binding sites by PET overlapping regions, and the 100% confirmation by ChIP-qPCR assays of the 40 binding loci identified by PET clusters, we have convincingly established the use of PET clusters as an efficient and accurate readout for identifying p53 binding loci. We therefore believe with high confidence that the 323 genomic loci determined by PET-3+ clusters in this study embrace true p53 protein binding sites.

Characterization of the p53 Binding Motif Using the PET-Cluster-Defined Loci

The currently known p53 binding motif is loosely defined (el-Deiry et al., 1992). Although the degenerate nature of the p53 DNA binding element may reflect the diversity and flexibility of p53-mediated responses to numerous cellular stress signals, this degeneracy complicates the detection and prediction of p53 binding sites in the whole genome.

The genome-wide identification of p53 binding loci as represented by the large number of PET clusters in this study provided an unprecedented opportunity for delving deeper into the nature of DNA binding by p53. To ask whether a key motif (or motifs) existed among the PET clusters, we first randomly picked 39 binding loci as the initial seed set for motif discovery followed by program training from the 68 PET-6+ cluster sequences. After applying a de novo motif-discovery algorithm, GLAM (Frith et al., 2004), a single prominent motif was identified, which undisputedly resembled the known consensus of p53 binding sites (Supplemental Data I-7). After further expectation-maximization-type optimization employing ROVER (Haverty et al., 2004), we established a highly effective model (hereafter referred to as the p53PET model) (Figure 3A). The effectiveness of the p53PET model for prediction of p53 binding sites was tested using the remaining 284 binding loci localized by PET-3+ clusters, and the performance of p53PET was evaluated in comparison with the previously reported p53MH model (Hoh et al., 2002) and the p53PET model with its weight matrix replaced by the one in the TRANSFAC database (Wingender et al., 2000). As shown in Figure 3B using receiver operating characteristic (ROC) curves, it is evident that the p53PET model achieved much higher sensitivity for detecting p53 binding motifs than the other two models at all specificity levels. More importantly, the lengths of the spacers between the two half-sites in these 284 motif sequences are predominantly zero, although a few are 1 bp, and longer spacers are also observed (Figure 3C). This length distribution

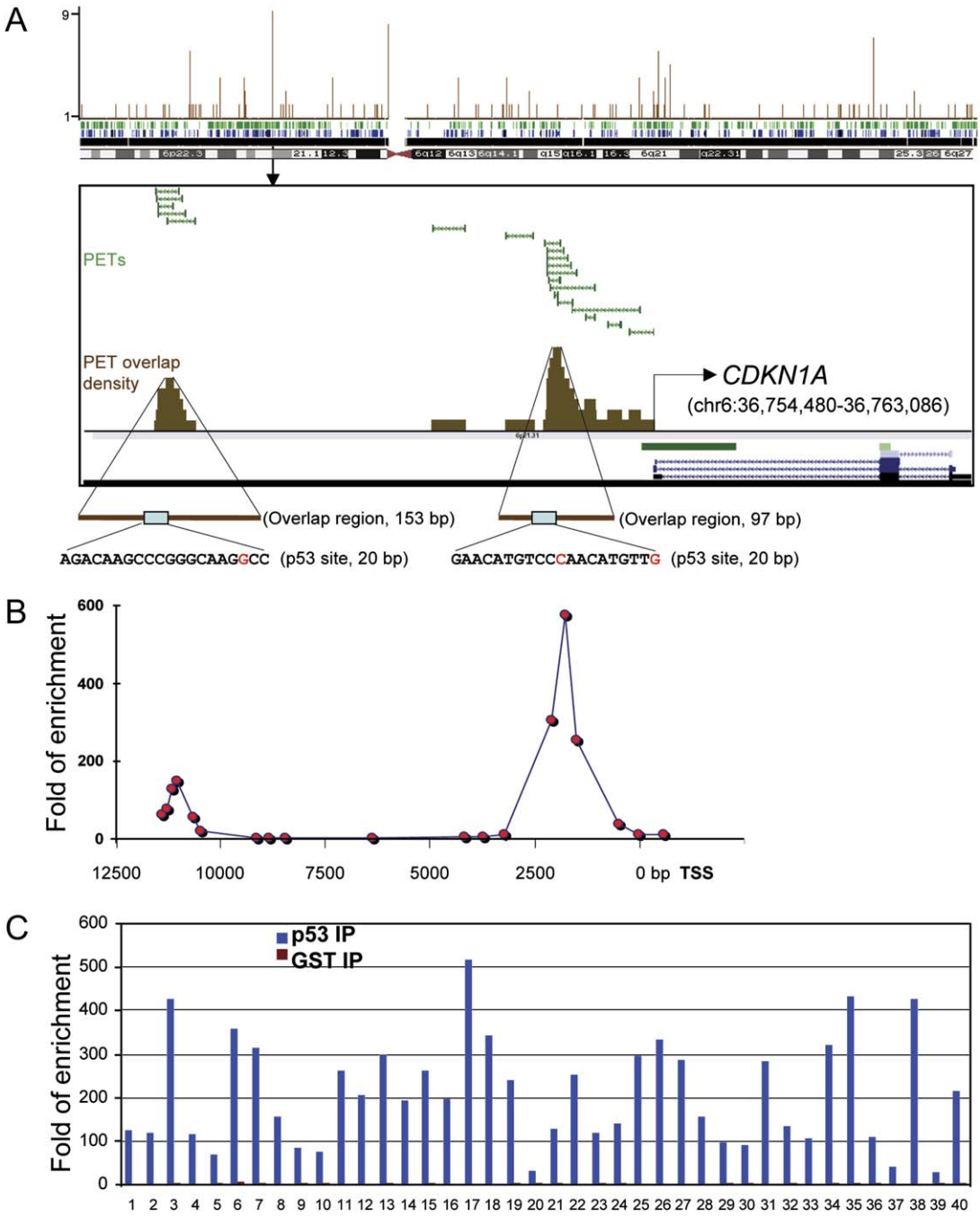

Figure 2. Validation of PET-Cluster-Identified p53 Binding Loci

(A) The whole-chromosome view of p53 ChIP-PETs mapping to chr6. A genomic span of 23 kb that contains the *CDKN1A* gene and its 5′ region is enlarged. *CDKN1A* was localized by two PET clusters; one contained 5 PETs, and the other contained 13 PETs. The two PET overlaps were 153 bp and 97 bp and were located in chr6:36742675–36743642 and chr6:36751902–36754502, respectively. Both PET overlaps contained recognizable p53 binding motifs.

(B) ChIP-qPCR validation in the 5′ upstream region of *CDKN1A*.

(C) p53 ChIP DNA (blue) and control GST ChIP DNA (red) were subjected to ChIP-qPCR analyses to determine the relative enrichment of candidate regions identified by ChIP clusters.

is much more specific than reported in previous studies, where spacers were simply said to vary between 0 and 14 bp.

Using the p53PET prediction model, we then analyzed all PET-localized regions for p53 motif finding. As summarized in Table 1, the percentages of the predicted p53 binding

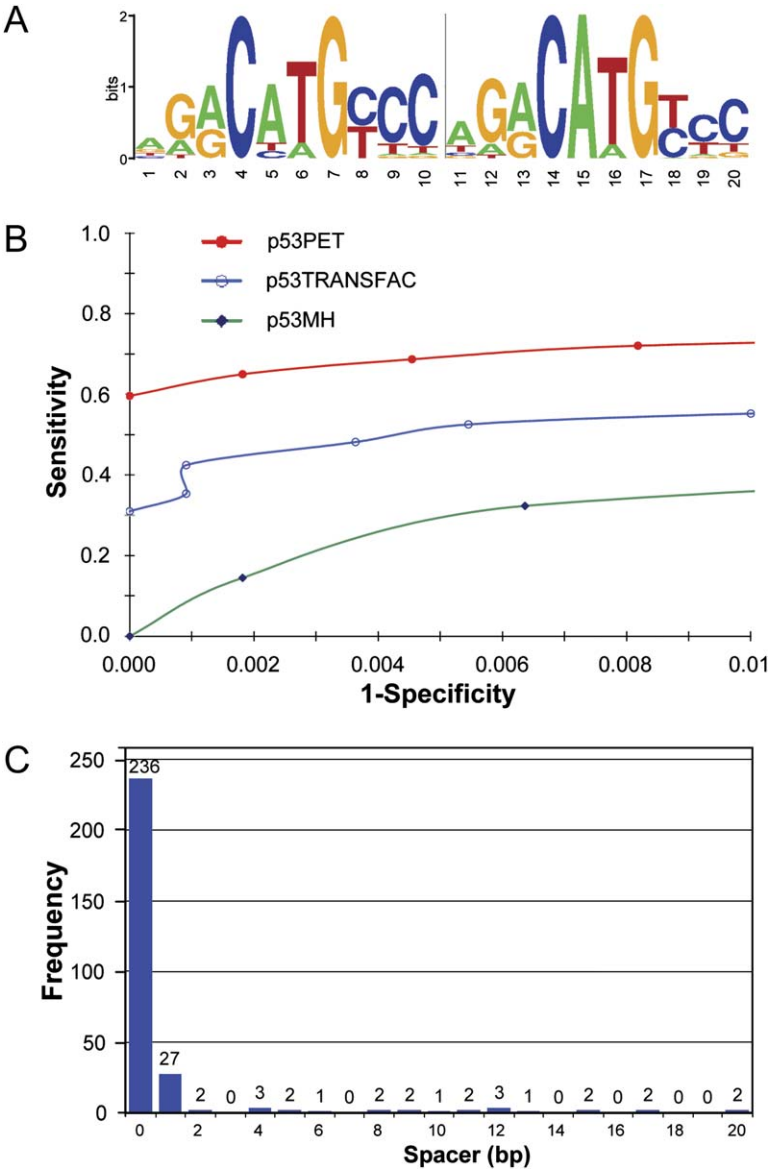

Figure 3. Motif Analysis of p53 Binding Sites

(A) Sequence logos depicting nucleotide distributions for the two p53 half-sites based on the p53PET model.

(B) ROC curve comparison between p53PET, p53TRANSFAC, and p53MH.

(C) The spacer lengths between the two halves of p53 binding motifs in PET-3+ clusters.

motifs were very low (0.68%) in the randomly selected genomic segments taken to represent background noise and similarly low (1.58%) in the PET singletons, reiterating the fact that most of the PET singletons are experimental noise, but significantly higher in PET clusters. We also observed a sharp increase in the p53 motif-containing rate, from 15.18% in PET-2 clusters to 61.15% in PET-3 clusters, and the escalation continued. This is consistent with our early estimates by statistical analysis that, although PET-2 clusters are enriched for p53 response elements, they also contain substantial noise, while the PET-3+ clusters are highly reliable. Overall, 73% of the PET-3+ clusters possessed recognizable p53 binding sites, which is a significant enrichment (up to 107-fold) as compared to background, suggesting again that the specific p53 interaction with the genome is predominantly through direct binding to a single binding motif. We suspect that the 27% non-motif-containing binding loci identified by PET-3+ clusters might be due to recruitment of p53 to genomic locations through indirect DNA binding as has been found for the estrogen receptor (Carroll et al., 2005). Again, compared with p53MH, the p53PET model showed better prediction results for p53 motif finding, with greater distinction between background and high-probability PET clusters (i.e., fewer hits in the background set of random genomic DNA segments and PET singletons and greater detection in PET-cluster sequences) (Table 1). Despite the relative nonspecificity of PET-2 clusters, using the new p53PET motif-finding model, we were able to identify 219 PET-2 clusters with high likelihood of p53 interaction regions containing p53 binding motifs. Thus, including the 323 binding loci identified by PET-3+ clusters, we have established a total of 542 high-probability p53 binding loci.

While our total number (1,766) of PET clusters is in good agreement with the 1,600 binding sites as extrapolated from the p53 localization analysis for chromosomes 21 and 22 using ChIP-on-CHIP (Cawley et al., 2004), the specific binding sites on these chromosomes had significant nonoverlap between the two experiments. In the two chromosomes, 48 loci based on hybridization peaks were identified, while in this study we had 55 PET clusters. Within these PET clusters, 5 were PET-3+ clusters (3 of them contain the p53 motif), and 8 were PET-2 clusters that contain idealized p53 binding motifs (Table S3). By our earlier validation results, these 13 loci identified by PET clusters (11 containing p53 motifs) were considered high confidence with regards to p53 binding, including one that was mapped in the first intron of a known p53-responsive gene (*PRODH*/*PIG6*) (Polyak et al., 1997). Three of the thirteen PET-cluster-determined loci were also identified by the p53 ChIP-on-CHIP analysis. One of the common loci was in a gene desert region with the nearest gene model (*C21orf116*) 112 kb away from its 5′ side, one was localized in an internal intron region of *SMARCB1*, and the other was in the first intron of *AB051436*. We further applied our optimized p53PET motif-finding model to the 48 loci derived from ChIP-on-CHIP analysis and found that only 5 of them had the requisite p53 binding motif. We observed that the PET-derived loci were significantly more likely to contain a p53 motif (11 of 13, or 85%) than loci identified by ChIP-on-CHIP (5 of 48, or 10%). The most interesting discrepancy in this group is the binding locus localized by a PET-8 cluster on chromosome 21 (chr21:33660665–33662530) but missed in the ChIP-on-CHIP study. This locus is 6,672 bp downstream of the 3′ side of *IFNAR1*, which is involved in stress response to viral infection. Our ChIP-qPCR analysis indeed confirmed that this locus is a genuine in vivo binding site for p53 under 5-FU induction conditions in HCT116 cells. Similarly, the binding locus on chromosome 22 (chr22:27702966–27705354) localized by a PET-5 cluster was also validated. The discrepancy between the two studies could be attributed to different chemical treatments (5-FU versus bleomycin) and possibly different stringencies used for determining binding loci.

Using the optimized p53PET motif-finding model, we scanned the entire human genome and identified 13,885 ab initio p53 binding sites. Although with increased stringency the p53 binding sites predicted by p53PET could be reduced to a few thousand, the number is still significantly larger than that experimentally identified. Besides a certain level of false positives, it is possible that the predicted p53 binding sites represent the total capacity of p53 targeting in the genome, while the experimentally identified loci in each study may reflect only a subset of functional p53 sites in that particular biological condition in a specific cell line.

Identification of Novel p53 Target Genes

Having established that the PET-cluster loci were highly associated with p53 interactions, the 542 loci determined by a combined PET-clustering and motif analysis represent a rich resource for the identification of novel p53 target genes. Based on their location within 100 kb of transcription units, we assigned 474 such clusters to 458 known genes (Table S4). One hundred and fifty-six of the clusters were 5′ upstream, forty-six were in the first introns, one hundred and fifty-two were in internal introns, and one hundred and twenty were in 3′ downstream regions of the genes (Figure 4A). Significantly, none were found in exonic regions (p value = $7e^{-10}$, Supplemental Data I-6). Over 90% of the binding sites were within 60 kb of the target genes, with the highest density of binding sites (338 of 474; 71%) located within approximately 20 kb of the 5′ and 3′ flanking regions.

To validate and further characterize these candidates for p53 direct target genes, we obtained gene-expression data for the same cell line (HCT116) treated under the same condition (5-FU for 6 hr) using oligonucleotide microarrays containing 20,000 gene probes (Kho et al., 2004). Out of the 458 PET-cluster-associated genes, 275 have corresponding expression data, in which 65 were upregulated and 57 downregulated in response to 5-FU in p53 wild-type (+/+) versus p53 mutant (−/−) cells. We therefore consider these 122 genes, characterized by both PET binding data and expression data, as direct p53 target genes (Table 2). We asked whether upregulated genes had different binding characteristics from downregulated genes and observed that a statistically significant proportion of upregulated genes have their binding loci at 5′ proximity and first introns (38 of 65 upregulated genes, p = $7.4e^{-5}$; Supplemental Data I-6). This suggests a potential difference between genes upregulated and genes downregulated by p53 based on binding-site location (Figure 4A).

The 122 direct targets identified by p53 binding compiled in Table 2 include 24 known p53-responsive genes, while the other 98 were not previously associated with p53 response. Functional categorization of these genes revealed a broad spectrum of p53 functions, including cell motility and migration and receptor-tyrosine-kinase signaling cascades (RTK/PTPase), in addition to well-characterized p53 functions. Strikingly, 20 novel p53 target-gene candidates are associated with the regulation of cell motility and adhesion. p53 has been implicated in regulation of tumor invasion and metastasis (Singh et al., 2002). However, it was not clear which p53 target genes were involved in this cellular process. To explore the possibility that p53 regulates metastasis through transcriptional regulation of cell adhesion and motility genes, 18 targets in this category were selected to measure their expression levels in 5-FU-treated cells using real-time qPCR. Of the tested genes, 15 were indeed modulated (7 were up- and 8 were downregulated) by p53 activation, and 3 were not affected. *PCDH7* and *VIM*, which are involved in cell adhesion and cytoskeleton structure, were both downregulated, whereas *ITGAM* and *Col4A1* were upregulated (Figure 4B). Our results point to the possibility that p53 can suppress metastasis through direct transcriptional regulation of a new category of molecular targets.

Clinical Relevance of p53 Direct Targets in Primary Cancer Tissues

It is known that transcriptional regulation in cultured cells might reflect in vitro artifacts, and tissue-dependent p53

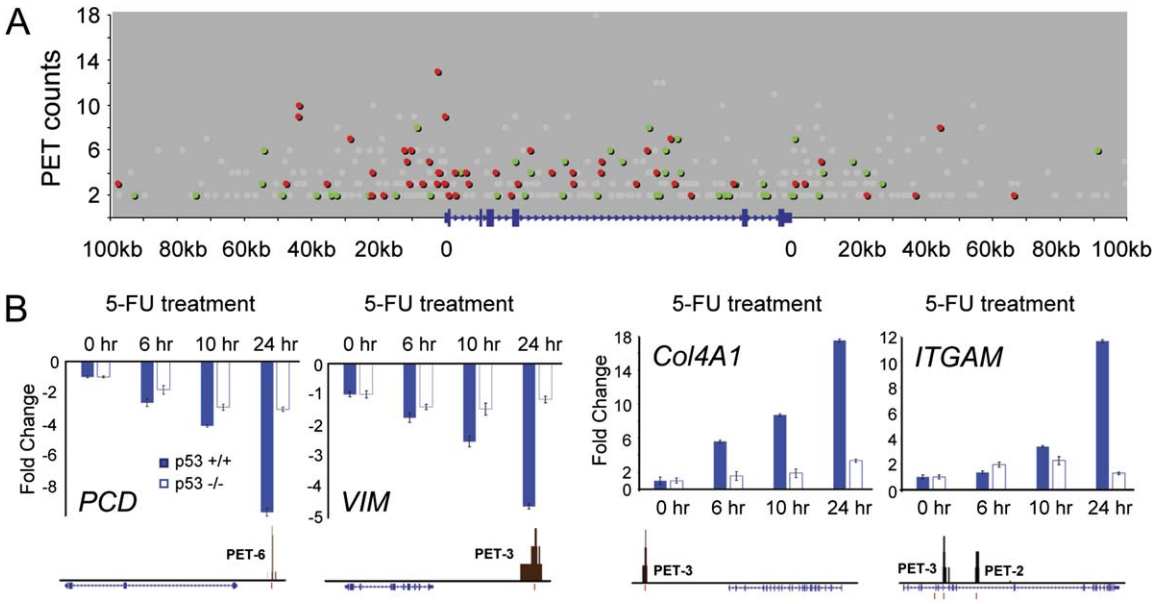

Figure 4. Location of p53 Binding Loci around Target Genes and Validation by Gene Expression

(A) Four hundred and seventy-four PET clusters were plotted against PET counts of each PET cluster (y axis) and locations (x axis) of corresponding genes represented by a gene model based on *BAX*. Locations in 5′ and 3′ regions are indicated in kilobases, while locations in introns were plotted in proportion to the gene length of that intron. The gray dots indicate PET clusters mapped to genes that either did not have expression data or showed no change in expression levels.

(B) Four novel p53 target genes (*PCDH7*, *VIM*, *Col4A1*, and *ITGAM*) were validated using real-time PCR for expression in 5-FU-treated HCT116 cells. Fold changes relative to time 0 at indicated time points are plotted with HCT116 as solid blue bars and HCT116 $p53^{-/-}$ as hollow bars. The error bars represent 95% confidence intervals. The locations of PET clusters with respect to their corresponding genes and the motifs (red bars) identified by p53PET are shown.

transcriptional activity has been previously described (Coates et al., 2003). To further validate the genes identified by ChIP-PET as bona fide p53 targets, and to determine the extent of their response to p53 in primary tumors, we studied their expression patterns in a collection of 251 primary breast tumors profiled using the Affymetrix U133A and B microarrays (Miller et al., 2005). In this set of tumors, the p53 cDNA had been previously sequenced, leading to the identification of 58 p53 mutant tumors and 193 tumors with p53 wild-types (Bergh et al., 1995). All except one of the 122 p53 direct target genes were represented by probes on the Affymetrix array. Using expression data derived from the 251 breast tumors for 65 p53-activated genes and 56 p53-repressed genes, respectively, we performed unsupervised hierarchical clustering, which resulted in two primary tumor clusters significantly associated with the p53 mutation status (Figures 5A and 5B). A number of p53-upregulated genes showed higher expression levels in most of the p53 wild-type tumors relative to the p53 mutant tumors, consistent with their transcriptional dependence on p53. Similarly, a number of p53-downregulated genes were expressed at lower levels in the p53 wild-type tumors relative to the mutants, consistent with their transcriptional repression by p53. Furthermore, dysregulation of these target genes (i.e., lower expression of p53-activated genes and higher expression of p53-repressed genes) was, in each case, significantly linked to the development of distant metastasis within 5 years of diagnosis. Pathologically, tumors associated with this dysregulation appeared to be more aggressive as evidenced by their higher tumor grades and the observation that patients with these tumors had a significantly lower probability of surviving their cancer (Figures 5C and 5D). Interestingly, two of the p53-repressed genes known for their antiapoptotic functions, *BCL2A1* and *TNFAIP8*, showed the highest correlations with both p53 mutation status and high tumor grades. Although p53 is known to repress antiapoptotic genes, such as *BCL2*, to regulate apoptosis, to our knowledge this is the first report that *BCL2A1* and *TNFAIP8* are transcriptionally silenced by p53. The observation that their expression patterns in the breast tumors correlate highly with p53 status and clinical behavior (Figure 5B) suggests they may be powerful new biomarkers for patient prognosis.

Taken together, these findings strongly argue that most of the novel p53 direct target genes identified by PET clusters are bona fide p53 direct targets, are regulated by p53 in different cell types, and are functional in p53-mediated tumorigenesis. Furthermore, their expression characteristics in vivo can potentially be used as molecular gauges of tumor aggressiveness and clinical outcome.

DISCUSSION

The ChIP-PET strategy demonstrated in this study represents a substantial advance in our ability to identify *cis*-regulatory elements, notably transcription-factor binding sites, on

Table 2. Categories of p53 Target Genes Identified by ChIP-PET Analysis

Apoptosis Cell Cycle	DNA Repair Chr. Modifier	Cell Growth Differentiation	Transcription Regulation	Protein Catabolism	Signal Transduction	Cell Adhesion Mobility	Biosynthesis Metabolism	Transport and Ion Channel	Unknown
BAX	*PCNA*	*IER5*	*ATF3*	TRIM22	*RRAD*	*GPC3*	*RPS27L*	*STAU*	*FLJ11259*
GADD45A	*RRM2B*	*TGFA*	*MYBL1*	CPN1	*SNX5*	*S100A2*	TPO	STARD4	*WIG1*
CCNG2	*TP53AP1*	BCAS3	ADORA2B	NEDD4L	DKK3	ANK1	ASTN2	EEA1	AB011136
CDKN1A	*XPC*	C2orf29	ADRB1	USP34	EIF2AK3	ARHGAP5	CHST12	KCNMA1	AK055226
GML	CHD2	FGF2	CBLC	USP9X	ERBB4	BICD2	CYP4F3	OSBP	ANKRD10
PIG6	DDB2	GSPT1	GNAQ		GNAI1	CALD1	FTHFSDC1	SLC4A10	ANP32D
SNK	HDAC9	KITLG	GPR39		NCK2	CDC42EP3	LOC144501	TRPM1	B1
TNFRSF10B	MLH1	MDM4	NAB1		NMU	COL4A1	NAV3		BC004942
BCL2A1	MSH6		NR6A1		NOTCH1	CTNNA3	PCCA		C2orf25
CNAP1			PCAF		PPM2C	FAT	PRKAG2		CDKAL1
RBL1			PRDM1		PTPRE	FLJ20972			DDIT4
SMARCB1			TIF1		PTPRM	FRMD4A			FLJ12484
TNFAIP8			UBP1		PTPRO	ITGAM			FLJ20045
						LTBP1			FLJ22457
						MYO1A			HIG1
						NEO1			LATS2
						NID2			MDS009
						NLGN1			PHF14
						PCDH7			PIAS2
						PPFIBP1			PSTPIP2
						PTK2			ShrmL
						VIM			SPAG9

Previously known p53 targets are in italic; novel p53 targets are in roman.

a whole-genome level. Unlike array-based approaches, ChIP-PET is an open system for identifying any regulatory binding loci that can be enriched by ChIP and requires only standard sequencing capacity. The method is therefore readily applicable for global localization analyses of TFBS in any genome as long as the whole-genome sequence assembly is available. ChIP-PET is also more precise for TFBS mapping than the current approaches. We have demonstrated that >80% of known and new p53 binding sites identified in this study resided in the overlapping regions of PET clusters, providing a way to narrow the TFBS down to less than 100 bp. This is made possible by the unique feature that characterizes the termini of individual PET-identified fragments. As a result, we can unambiguously distinguish the original ChIP DNA fragments (distinct PETs) from the amplified noise (redundant PETs with multiple copies) regardless of how much the amplification might be.

This feature of paired-end ditagging also sets the PET strategy apart from the recently reported method using SAGE-like monotags to map TFBS (Impey et al., 2004; Kim et al., 2005a; Chen and Sadowski, 2005; Roh et al., 2005). In the monotag approach, each ChIP DNA fragment is represented by a single tag of 20 bp, and tag counts (copy number) are used to measure ChIP enrichment; this approach cannot distinguish overlapped different ChIP DNA fragments from redundant tags due to amplification and therefore would significantly increase false positives as we simulated with the data generated in this study (Figure S10). In contrast, the PET-cluster readout scheme is more accurate in identifying binding loci and more specific in narrowly defining binding sites.

Although the amount of sequencing required (~40,000 sequencing reads) for a comprehensive ChIP-PET experiment is miniscule for most sequencing centers and within the reach of core facilities in university laboratories, the cost for each ChIP-PET experiment is substantial. One approach to increase efficiency is to develop an effective subtraction scheme (Chen and Sadowski, 2005) to reduce the level of background noise so as to decrease the number of sequencing reads required. Ultimately, the ChIP-PET

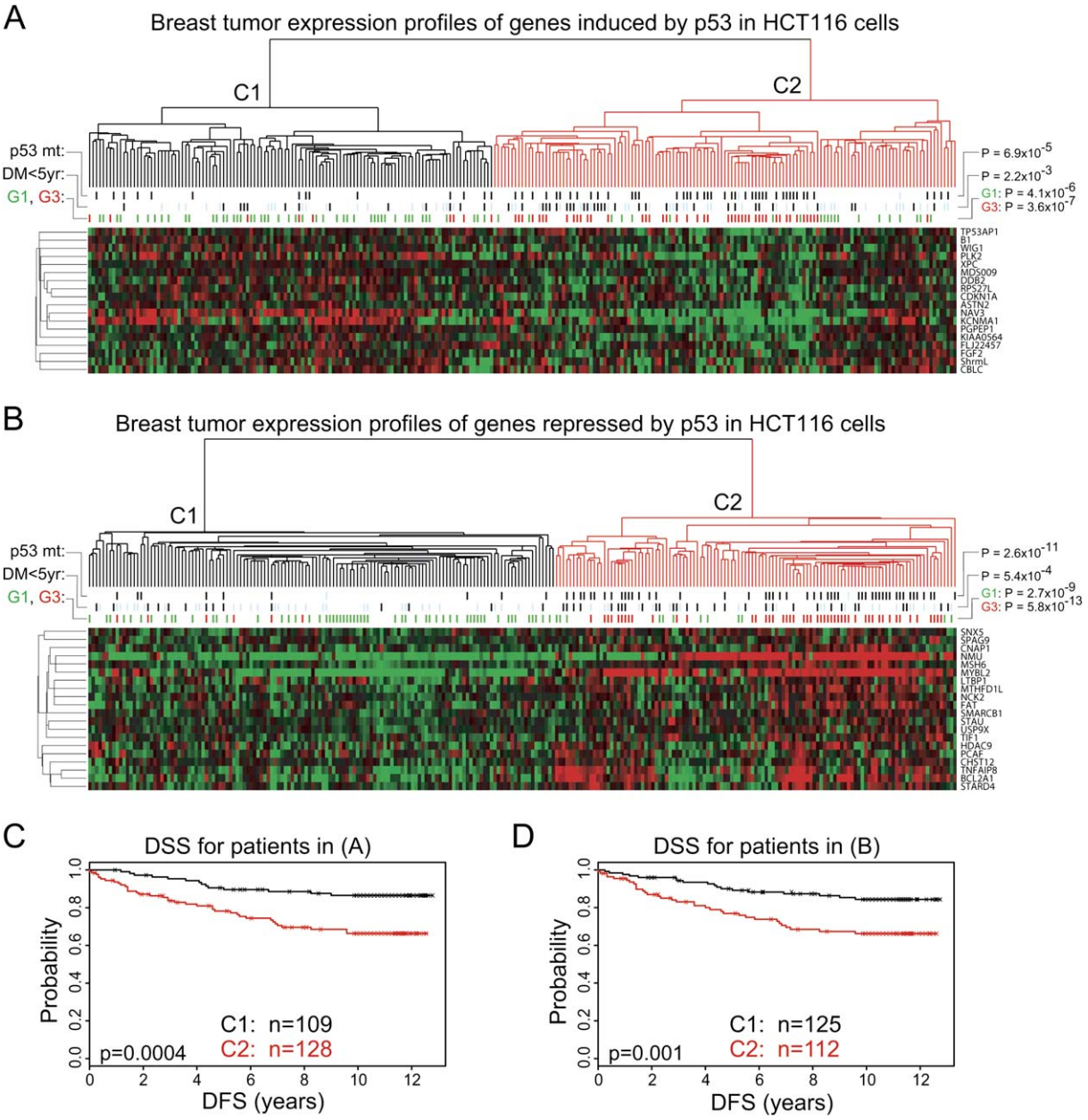

Figure 5. In Vivo Analysis of p53 Target-Gene Expression

Unsupervised hierarchical cluster analysis of 251 breast tumors was performed using the 65 upregulated genes (A) or 56 downregulated genes (B) by p53 in 5-FU-treated HCT116 cells. The formation of two tumor clusters (C1 and C2) and the major tumor branch points are shown in the colored heat map. Red indicates above-mean expression; green denotes below-mean levels. The degree of color saturation reflects the magnitude of expression value. Black vertical bars represent p53 mutant tumors (p53 mt) or those that gave rise to a distant metastasis within 5 years of diagnosis (DM < 5 yr). Pale blue bars in the rows of "p53 mt" and "DM < 5 yr" reflect missing data. Green and red bars reflect histological grade I and grade III tumors, respectively. Kaplan-Meier disease-specific survival (DSS) plots are shown for the two major cluster branches formed in (A) (C) and (B) (D). p values were calculated by the chi-square test.

approach will be further empowered by new cost-effective sequencing technologies under rapid development (Margulies et al., 2005; Shendure et al., 2005). In particular, we have adapted the multiplex sequencing method (Margulies et al., 2005) for PET-based sequencing analysis to characterize mammalian transcriptomes and interrogate complex genomes (P.N., J.J.S. Tan, K.P.C., H.S. Ooi, Y.L L., M.J. Fullwoods, L. Du, W.-K.S., C.-L.W., and Y.R., unpublished data) and expect to reduce the sequencing cost to under $5,000 per ChIP-PET experiment within a year.

After a saturated sampling, we scanned the entire human genome for p53 TFBS under a given cellular condition, established a comprehensive map of p53 binding, and identified 542 loci with high confidence of p53 interaction. This number is different from what was extrapolated (1,600) from the p53 ChIP-on-CHIP analysis (Cawley et al., 2004), and p53 loci on chromosome 21 and 22 localized by these two different experiments shared minimal overlap. Comparison between the two data sets is difficult because these two chromosomes are poorly populated with well-characterized p53 target genes, with the exception of *PIG6*/*PRODH* (22q11.21), which was previously reported to be responsive to p53 activation (Polyak et al., 1997). In our study, a PET-2 cluster was mapped to the first intron of this gene (Table S3) and covered a p53 binding motif.

Given that p53 binding sites on chromosomes 21 and 22 identified by the ChIP-PET approach are significantly enriched for p53 binding motifs (85% versus 10%) as opposed to sites assigned by ChIP-on-CHIP, it is likely that the ChIP-PET methodology is more specific than ChIP-on-CHIP. In addition, besides potential experimental variations and different techniques used in the two studies, one possible explanation for such a discrepancy between the two results is that, although the two experiments were done using the same cells (HCT116), different induction treatments and time points were employed (5-FU for 6 hr and bleomycin for 12 hr). Time-course studies have shown clearly that the binding of p53 to target sites is dynamic and changes significantly in the first 6 to 12 hr after any exposure to conditions that induce DNA damage (Crosby et al., 2004). The differences between the effects of 5-FU and bleomycin could also be substantial. 5-FU is a nucleoside analog and replaces replicating nucleic acid with fluorinated uracil, resulting in wide-scale coding and structural alterations. Bleomycin nicks DNA and introduces generalized strand breaks. In fact, these two experiments may merely represent two snapshots of potentially very large and fluid bodies of transcriptional networks in response to different p53-activating signals. Thus, it is perhaps only through sampling of many combinations of biological settings and p53-activating signals that we can obtain a truly comprehensive and complete atlas of p53 genomic activity.

In addition to presenting a global view of p53 TFBS in the human genome for the first time, this study also provided a comprehensive list of p53 target genes and their responses to p53 activation in colorectal cancer cells. Through crossvalidation in clinical breast tumors, we identified a comprehensive panel of likely direct targets of p53, many of which appear to play a role in p53-dependent tumorigenesis in primary cancer tissues. Functional analysis of these genes revealed a broad spectrum of novel p53 functions, including cell adhesion and migration and involvement in receptor-tyrosine-kinase signaling cascades (RTK/PTPase). Recently, p53 has been implicated in the regulation of tumor invasion and metastasis (Tlsty, 1998). Our discovery of 20 previously unidentified targets involved in cell motility, adhesion, and migration suggests that a large number of novel p53 targets could be involved in p53-mediated suppression of tumor metastasis. For example, PTK2 (known as focal adhesion kinase, FAK) is known to promote cell invasion and metastasis through integrin-mediated signaling (Lin et al., 2004) and is overexpressed in invasive breast and colon cancers. It is very likely that p53 suppresses metastasis through down-regulation of *PTK2*. Furthermore, we identified *VIM*, the expression of which has been closely correlated with prostate and breast cancer metastasis (Thompson et al., 1992; Singh et al., 2002). The biological roles of these promising new candidates in p53-regulated suppression of tumor migration and metastasis warrant further investigation.

This study also raised new questions concerning p53 DNA binding dynamics. Of the 542 high-confidence binding loci, we have observed that many were either far away from proximal promoter of genes or inactive in inducing adjacent gene expression. Are these binding sites functional? If so, how do they operate? It is possible that many of the p53 binding sites function through long-distance interactions as enhancers or locus control regions (LCRs) to modulate gene expressions (West and Fraser, 2005), which can be investigated using the 3C approach (Dekker et al., 2002). It is also possible that a number of the p53 binding sites not associated with array-detected transcriptional activity identified in this study indicate that array probe analysis is blind to alterations in splicing and alternative transcriptional start and end sites. In addition, these sites might not permit recruitment of required cofactors to trigger transcriptional activity. p53 is known to interact with coactivators or corepressors under various conditions. For example, *BNIP3L* is a p53 binding target, and its induction requires the simultaneous activation of both p53 and hypoxia-inducible factor 1 (HIF-1) under hypoxic conditions (Fei et al., 2004). The data generated in this study potentially provide the initial framework for higher-level interactions of p53 regulation.

In summary, we have developed an unbiased, highly precise, and efficient mapping methodology to allow the whole-genome survey of TFBS with an unprecedented resolution. The application of this approach to the discovery of p53 binding sites has enabled us to identify many new in vivo targets of the p53 tumor-suppressor protein. The characterization of these targets by expression profiling in cultured cancer cells and primary tumors uncovered potentially important pathological and clinical roles. Our findings expand the current knowledge base surrounding p53 function and implicate p53 in a greater diversity of biological activities than previously suspected.

EXPERIMENTAL PROCEDURES

Cell Culture and Drug Treatments

Human colon cancer cell line HCT116 and its derived isogenic *p53*$^{-/-}$ cells (provided by Dr. Bert Vogelstein) were cultured in DMEM containing 10% FCS and treated with 5-fluorouracil and cycloheximide.

ChIP Experiment

ChIP assays with HCT116 cells were carried out as described (Weinmann and Farnham, 2002; Wells and Farnham, 2002). For all ChIP experiments, DO1 monoclonal antibody was used for immunoprecipitation, and quantitative PCR analyses were performed in real time using the ABI PRISM

7900 Sequence Detection System and SYBR Green Master Mix as described (Ng et al., 2003). Relative occupancy values were calculated by determining the apparent immunoprecipitation efficiency (ratios of the amount of immunoprecipitated DNA over that of the input sample) and normalized to the level observed at a control region, which was defined as 1.0.

Construction of ChIP-PET Library

The end-polished ChIP DNA fragments were ligated to the cloning vector pGIS3, which contains two MmeI recognition sites (Figure S1). The ligations were transformed into TOP10 cells (Invitrogen) to form the ChIP DNA library. Plasmid of the ChIP DNA library was digested with MmeI and end polished with T4 DNA polymerase. The resulting vector containing a signature tag from each terminal of the ChIP DNA insert was self-ligated and then transformed into TOP10 cells to form the "single-PET" library. The plasmids from this library were digested with BamHI to release 50 bp PETs, which were concatenated into long fragments (1–2 kb) and cloned into pZErO-1 (Invitrogen) as the final ChIP-PET library for sequencing.

PET Extraction and Mapping to Genome

PET sequences were extracted from the raw sequence reads obtained from the ChIP-PET library and were mapped to the human genome assembly (hg17). The process of PET extraction and mapping is essentially same as previously described for cDNA analysis (Ng et al., 2005).

Additional Methods

The Supplemental Data contain the above methods and additional methods in more detail, statistical analyses, and p53 motif analyses.

Supplemental Data

Supplemental Data include Supplemental Experimental Procedures, Supplemental References, 11 figures, and 4 tables and can be found with this article online at http://www.cell.com/cgi/content/full/124/1/207/DC1/.

ACKNOWLEDGMENTS

The authors acknowledge Mr. H. Thoreau, Mr. L. Lim, and the Cloning and Sequencing Group at the Genome Institute of Singapore for technical support; Mr. Au Yong Wing Yau, Mr. Choo Siew Woh, and Mr. Wong Chee Hong of the Bioinformatics Institute of Singapore for bioinformatics support; and Ms. Melissa Jane Fullwood for manuscript proofreading. This work was supported by A*STAR of Singapore and NIH ENCODE grant 1R01HG003521-01 to Y.R. and C.L.W. B.L. is supported partially by a grant from the NIH.

Received: May 4, 2005
Revised: September 13, 2005
Accepted: October 25, 2005
Published: January 12, 2006

REFERENCES

Bergh, J., Norberg, T., Sjogren, S., Lindgren, A., and Holmberg, L. (1995). Complete sequencing of the p53 gene provides prognostic information in breast cancer patients, particularly in relation to adjuvant systemic therapy and radiotherapy. Nat. Med. *10*, 1029–1034.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. Cell *122*, 1–10.

Carroll, J.S., Liu, X.S., Brodsky, A.S., Li, W., Meyer, C.A., Szary, A.J., Eeckhoute, J., Shao, W., Hestermann, E.V., Geistlinger, T.R., et al. (2005). Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell *122*, 33–43.

Cawley, S., Bekiranov, S., Ng, H.H., Kapranov, P., Sekinger, E.A., Kampa, D., Piccolboni, A., Sementchenko, V., Cheng, J., Williams, A.J., et al. (2004). Unbiased mapping of transcription factor binding sites along human chromosome 21 and 22 points to widespread regulation of noncoding RNAs. Cell *116*, 499–509.

Chen, J., and Sadowski, I. (2005). Identification of the mismatch repair genes PMS2 and MLH1 as p53 parget genes by using serial analysis of binding elements. Proc. Natl. Acad. Sci. USA *102*, 4813–4818.

Coates, P.J., Lorimore, S.A., Lindsay, K.J., and Wright, E.G. (2003). Tissue-specific p53 responses to ionizing radiation and their genetic modification: the key to tissue-specific tumour susceptibility? J. Pathol. *201*, 377–388.

Crosby, M.E., Oancea, M., and Almasan, A. (2004). p53 binding to target sites is dynamically regulated before and after ionizing radiation-mediated DNA damage. J. Environ. Pathol. Toxicol. Oncol. *23*, 67–79.

Dekker, J., Rippe, K., Dekker, M., and Kleckner, N. (2002). Capturing chromosome conformation. Science *295*, 1306–1311.

el-Deiry, W.S., Kern, S.E., Pietenpol, J.A., Kinzler, K.W., and Vogelstein, B. (1992). Definition of a consensus binding site for p53. Nat. Genet. *1*, 45–49.

el-Deiry, W.S., Tokino, T., Velculescu, V.E., Levy, D.B., Parsons, R., Trent, J.M., Lin, D., Mercer, W.E., Kinzler, K.W., and Vogelstein, B. (1993). WAF1, a potential mediator of p53 tumor suppression. Cell *75*, 817–825.

ENCODE Project Consortium (2004). The ENCODE (ENCyclopedia Of DNA Elements) Project. Science *306*, 636–640.

Fei, P., Wang, W., Kim, S.H., Wang, S., Burns, T.F., Sax, J.K., Buzzai, M., Dicker, D.T., McKenna, W.G., Bernhard, E.J., and el-Deiry, W.S. (2004). Bnip3L is induced by p53 under hypoxia, and its knockdown promotes tumor growth. Cancer Cell *6*, 597–609.

Frith, M.C., Hansen, U., Spouge, J.L., and Weng, Z. (2004). Finding functional sequence elements by multiple local alignment. Nucleic Acids Res. *32*, 189–200.

Haverty, P.M., Hansen, U., and Weng, Z. (2004). Computational inference of transcriptional regulatory networks from expression profiling and transcription factor binding site identification. Nucleic Acids Res. *32*, 179–188.

Hoh, J., Jin, S., Parrado, T., Edington, J., Levine, A.J., and Ott, J. (2002). The p53MH algorithm and its application in detecting p5-responsive genes. Proc. Natl. Acad. Sci. USA *99*, 8467–8472.

Horak, C.E., Mahajan, M.C., Luscombe, N.M., Gerstein, M., Weissman, S.M., and Synder, M. (2002). GATA-1 binding sites mapped in the β-globin locus by using mammalian chIp-chip analysis. Proc. Natl. Acad. Sci. USA *99*, 2924–2929.

Hug, B.A., Ahmed, N., Robbins, J.A., and Lazar, M.A. (2004). A chromatin immunoprecipitation screen reveals protein kinase cβ as a direct RUNX1 target gene. J. Biol. Chem. *279*, 825–830.

Impey, S., McCorkle, S.R., Cha-Molstad, H., Dwyer, J.M., Yochum, G.S., Boss, J.M., McWeeney, S., Dunn, J.J., Mandel, G., and Goodman, R.H. (2004). Defining the CREB regulon: a genome-wide analysis of transcription factor regulatory regions. Cell *119*, 1041–1054.

International Human Genome Sequencing Consortium (2004). Finishing the euchromatic sequence of the human genome. Nature *431*, 931–945.

Kaeser, M.D., and Iggo, R.D. (2002). Chromatin immunoprecipitation analysis fails to support the latency model for regulation of p53 DNA binding activity in vivo. Proc. Natl. Acad. Sci. USA *99*, 95–100. Published online December 26, 2001. 10.1073/pnas.012283399.

Kannan, K., Amariglio, N., Rechavi, G., Jakob-Hirsch, J., Kela, I., Kaminski, N., Getz, G., Domany, E., and Givol, D. (2001). DNA microarray identification of primary and secondary target genes regulated by p53. Oncogene *20*, 2225–2234.

Kho, P.S., Wang, Z., Zhuang, L., Li, Y., Chew, J.L., Ng, H.H., Liu, E.T., and Yu, Q. (2004). p53 regulated transcriptional program associated with genotoxic stress-induced apoptosis. J. Biol. Chem. *279*, 21183–21192.

Kim, J., Bhinge, A.A., Morgan, X.C., and Iyer, V.R. (2005a). Mapping DNA-protein interactions in large genomes by sequence tag analysis of genomic enrichment. Nat. Methods *2*, 47–53.

Kim, T.H., Barrera, L.O., Zheng, M., Qu, C., Singer, M.A., Richmond, T.A., Wu, Y., Green, R.D., and Ren, B. (2005b). A high-resolution map of active promoters in the human genome. Nature *436*, 876–880.

Kuznetsov, V.A. (2005). Mathematical Analysis and Modeling of SAGE Transcriptome. In SAGE: Current Technologies and Applications, S.M. Wang, ed. (Norwich, United Kingdom: Horizon BioScience), pp. 139–180.

Lin, Y.H., Park, Z.Y., Lin, D., Brahmbhatt, A.A., Rio, M.C., Yates, J.R., and Klemke, R.L. (2004). Regulation of cell migration and survival by focal adhesion targeting of Lasp-1. J. Cell Biol. *165*, 421–432.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., et al. (2005). Genome sequencing in microfabricated high-density picolitre reactors. Nature *437*, 376–380.

Miller, L.D., Smeds, J., George, J., Vega, V.B., Vergara, L., Ploner, A., Pawitan, Y., Hall, P., Klaar, S., Liu, E.T., et al. (2005). An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects and patient survival. Proc. Natl. Acad. Sci. USA *102*, 13550–13555.

Mirza, A., Wu, Q., Wang, L., McClanahan, T., Bishop, W.R., Gheyas, F., Ding, W., Hutchins, B., Hockenberry, T., Kirschmeier, P., et al. (2003). Global transcriptional program of p53 target genes during the process of apoptosis and cell cycle progression. Oncogene *22*, 3645–3654.

Ng, H.H., Robert, F., Young, R.A., and Struhl, K. (2003). Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Mol. Cell *11*, 709–719.

Ng, P., Wei, C.L., Sung, W.K., Chiu, K.P., Lipovich, L., Ang, C.C., Gupta, S., Shahab, A., Ridwan, A., Wong, C.H., et al. (2005). Gene identification signature (GIS) analysis for transcriptome characterization and genome annotation. Nat. Methods *2*, 105–111.

Polyak, K., Xia, Y., Zweuer, J.L., Kinzler, K.W., and Vogelstein, B. (1997). A model for p53-induced apoptosis. Nature *389*, 300–305.

Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., et al. (2000). Genome wide location and function of DNA binding proteins. Science *290*, 2306–2309.

Roh, T.Y., Cuddapah, S., and Zhao, K. (2005). Active chromatin domains are defined by acetylation islands revealed by genome-wide mapping. Genes Dev. *19*, 542–552. Published online February 10, 2005. 10.1101/gad.1272505.

Shendure, J., Porreca, G.J., Reppas, N.B., Lin, X., McCutcheon, J.P., Rosenbaum, A.M., Wang, M.D., Zhang, K., Mitra, R.D., and Church, G.M. (2005). Accurate multiplex polony sequencing of an evolved bacterial genome. Science *309*, 1728–1732. Published online August 4, 2005. 10.1126/science.1117389.

Singh, B., Reddy, P.G., Goberdhan, A., Walsh, C., Dao, S., Ngai, I., Chou, T.C., O-Charoenrat, P., Levine, A.J., Rao, P.H., and Stoffel, A. (2002). p53 regulates cell survival by inhibiting PIK3CA in squamous cell carcinomas. Genes Dev. *16*, 984–993.

Thompson, E.W., Paik, S., Brunner, N., Sommers, C.L., Zugmaier, G., Clarke, R., Shima, T.B., Torri, J., Donahue, S., Lippman, M.E., et al. (1992). Association of increased basement membrane invasiveness with absence of estrogen receptor and expression of vimentin in human breast cancer cell lines. J. Cell. Physiol. *150*, 534–544.

Tlsty, T.D. (1998). Cell-adhesion-dependent influences on genomic instability and carcinogenesis. Curr. Opin. Cell Biol. *10*, 647–653.

Vogelstein, B., Lane, D., and Levine, A. (2000). Surfing the p53 network. Nature *408*, 307–310.

Weinmann, A.S., Bartley, S.M., Zhang, T., Zhang, M.Q., and Farnham, P.J. (2001). Use of chromatin immunoprecipitation to clone novel E2F target promoters. Mol. Cell. Biol. *21*, 6820–6832.

Weinmann, A.S., and Farnham, P.J. (2002). Identification of unknown target genes of human transcription factors using chromatin immunoprecipitation. Methods *26*, 37–47.

Weinmann, A.S., Pearlly, S.Y., Oberley, M.J., Huang, T.H.-M., and Farnham, P.J. (2002). Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis. Genes Dev. *16*, 235–244.

Wells, J., and Farnham, P.J. (2002). Characterizing transcription factor binding sites using formaldehyde crosslinking and immunoprecipitation. Methods *26*, 48–56.

West, A.G., and Fraser, P. (2005). Remote control of gene transcription. Hum. Mol. Genet. *15*, R101–R111.

Wingender, E., Chen, X., Hehl, R., Karas, H., Liebich, I., Matys, V., Meinhardt, T., Prüß, M., Reuter, I., and Schacherer, F. (2000). TRANSFAC: an integrated system for gene expression regulation. Nucleic Acids Res. *28*, 316–319.

Yin, Y., Liu, Y., Jin, Y.J., Hall, E.J., and Barrett, J.C. (2003). PAC1 phosphatase is a transcription target of p53 in signalling apoptosis and growth suppression. Nature *422*, 527–531.

Yoon, H., Liyanarachchi, S., Wright, F.A., Davuluri, R., Lockman, J.C., Chapelle, A., and Pellegata, N.S. (2002). Gene expression profiling of isogenic cells with different *TP53* gene dosage reveals numerous genes that are affected by *TP53* dosage and identifies *CSPG2* as a direct target of p53. Proc. Natl. Acad. Sci. USA *99*, 15632–15637.

Yu, J., Zhang, L., Hwang, P.M., Rago, C., Kinzler, K.W., and Vogelstein, B. (1999). Identification and classification of p53-regulated genes. Proc. Natl. Acad. Sci. USA *96*, 14517–14522.

Zhao, R., Gish, K., Murphy, M., Yin, Y., Notterman, D., Hoffman, W.H., Tom, E., Mack, D.H., and Levine, A.J. (2000). Analysis of p53-regulated gene expression patterns using oligonucleotide arrays. Genes Dev. *14*, 981–993.

Accession Numbers

The PET DNA sequences reported in this study have been deposited in the ENCODE tracks of the UCSC genome browser under the p53 ChIP-PET analysis (http://genome.ucsc.edu/).