

Supplemental Data

Resource

A Global Map of p53 Transcription-Factor

Binding Sites in the Human Genome

Chia Lin Wei, Qiang Wu, Vinsensius Vega, Kuo Ping Chiu, Patrick Ng, Tao Zhang, Atif Shahab, How Choong Yong, YuTao Fu, Zhiping Weng, Jian Jun Liu, Xiao Dong Zhao, Joon-Lin Chew, Yen Ling Lee, Vladimir A. Kuznetsov, Wing-Kin Sung, Lance Miller, Bing Lim, Edison T. Liu, Qiang Yu, Huck Hui Ng, and Yijun Ruan

Table of Contents

I. Supplemental Experimental Procedures and Analysis

I-1 Cell Culture and Drug Treatments

I-2 ChIP Experiment

I-3 ChIP-PET Analysis

Construction of ChIP-PET Library

PET Extraction and Mapping to Genome

Localization of PET Clusters to Genes

I-4 Gene-Expression Microarray Experiment

Gene-Expression Analysis in HCT116 Cells

Gene-Expression Analysis in Breast Tumors

I-5 TaqMan Quantitative PCR Assay for Gene-Expression Analysis

I-6 Statistical Analyses

Saturation Kinetics of p53 ChIP-PET Library

Mapping Simulation of Overlapping PET Clusters

Goodness-of-Fit Analysis for Assessing the Reliability of PET Clusters

Enrichment of PET Clusters Associated with p53-Responsive Genes

Association of PET-Cluster Localization at the 5' of p53-Upregulated Genes

Significance of PET-Cluster Loci Not in Exons

I-7 Motif Analyses of p53 Binding Sites

De Novo Motif Discovery and Further Optimization

Performance Assessment

I-8 Comparison of PET and Monotag for ChIP DNA Analysis

II. Supplemental Figures

- S1 ChIP-PET Library Construction**
- S2 PET Sequence Processing and Mapping to the Human Genome**
- S3 Hill Function Curve for Estimating the Total PET Coverage**
- S4 Goodness-of-Fit Analysis for Assessing the Reliability of PET Clusters**
- S5 ChIP-qPCR Assay for p53 Binding Sites on BAX Promoter**
- S6 Additional Examples of PET Clusters to Known p53 Targets**
- S7 ChIP-qPCR Assay for Nine Previously Known p53 Targets**
- S8 Development of p53 Binding-Motif-Finding Model**
- S9 Performance Assessment of p53 Binding-Motif-Finding Models**
- S10 Detection of ChIP DNA Fragments by PET or Monotag**

III. Supplemental Tables

- S1 PET and PET-Cluster Distribution in the Human Genome**
- S2 PET Clusters Mapping to Known p53-Responsive Genes**
- S3 p53 Binding Loci and p53 Binding Motifs in Chromosome 21 and 22**
- S4 PET-Cluster-Identified p53 Binding Loci and Associated Genes**

I. Supplemental Experimental Procedures

I-1 Cell Culture and Drug Treatments

Human colon cancer cell line HCT116 and its derived isogenic p53 (-/-) cell line were kindly provided by Dr. Bert Vogelstein, Johns Hopkins University, Baltimore, MD. Cells were cultured in DMEM containing 10% FCS. All of the culture reagents and media were from Invitrogen. 5-Fluorouracil and Cycloheximide were purchased from Sigma.

I-2 ChIP Experiment

ChIP assays with HCT116 cells were carried out as described (Weinmann and Farnham, 2002; Wells and Farnham, 2002). Briefly, cells before and after 5-FU treatment was crosslinked with 1 % formaldehyde for 10 min at room temperature. Formaldehyde was inactivated by addition of 125 mM Glycine. Chromatin extracts containing DNA fragments of average size 500 bp were immunoprecipitated using anti-p53 DO1 monoclonal antibody (Santa Cruz). For all ChIP experiments, quantitative PCR analyses were performed in real time using 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. The control region is a 279 bp region on chromosome 22 and is amplified using the following primers: 5'-GGACTCGGAAGAGGTTACCTTCGG-3' and 5'-GTCGCCTCCGCTTGCTGAACTCAATGC-3'. The error for independent determinations is $\pm 10\%$. Primers for validation of ChIP-PET loci are available upon request. To ascertain the specificity of the antibody we used for p53 ChIP, we tested whether this antibody was able to enrich for p53 binding sites in a HCT116 cell-line with p53 gene deleted. Extracts were prepared from HCT116 cells, HCT116 p53 -/- cells, HCT116 cells treated with 5-FU for 6 hours, and HCT116 p53 null cells treated with 5-FU for 6 hours (Bunz et al, 1998). ChIP-qPCR analysis showed that DO1 enriched for p53 binding sites in HCT116 cells but not in a derivative cell without p53 (Figure S11A). Western blot analysis also showed that the DO1 antibody specifically recognized a band of the expected size in HCT116 extract. This antibody did not cross-react with the proteins derived from HCT116 cells without p53 (Figure S11B).

I-3 ChIP-PET Analysis

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 electrocompetent TOP10 bacterial cells to form the ChIP DNA library. Purified plasmid prepared from the ChIP DNA library was digested with MmeI, end-polished with T4 DNA polymerase to remove the 3'-dinucleotide overhangs, and the resulting plasmids containing a signature tag from each terminal of the original ChIP DNA insert were self-ligated to form single-ditag plasmids. These were then transformed into TOP10 cells to form a "single-ditag library". Plasmid DNA extracted from this library was digested with BamHI to release 50 bp paired end ditags. The PETs were PAGE-purified, then concatenated and separated on 4-20% gradient TBE-PAGE. An appropriate size fraction (1 kb-2 kb) of the concatenated DNA was excised, extracted and cloned into BamHI-cut pZErO-1 (Invitrogen) to form the final ChIP-PET library for sequencing.

PET Extraction and Mapping to Genome

PET sequences containing 18 bp from 5' and 18 bp from 3' ends of the original ChIP DNA fragments were extracted from the raw sequences obtained from the PET library, and mapped to human genome assembly (hg17). The process of PET extraction and mapping is essentially the same as previously described for cDNA analysis (Ng et al., 2005). The specific mapping criteria are that both the 5' and 3' signatures must be present on the same chromosome, on the same strand, in the correct orientation (5' \rightarrow 3'), with minimal 17 bp match, and within 4 kb of genomic distance.

Localization of PET Clusters to Genes

Based on the genomic coordinates, we took the center point of each PET cluster to measure the distances to the nearest genes on both sides of the cluster. The genes that had a distance from the nearest clusters of ≤ 100 kb were selected. We did not intend to absolutely associate the PET clusters with particular genes, but tried provide the distance of the clusters to the nearest genes along the chromosomes.

I-4 Gene-Expression Microarray Experiments

Gene-Expression Analysis in HCT116 Cells

Same as previously described (Kho et al., 2004)

Gene-Expression Analysis in Breast Tumors

Breast tumors (n=251) were profiled on Affymetrix U133 genechips as reported previously (Coates et al., 2003). The 122 high confidence p53 target genes were mapped to Affymetrix probesets via UniGene build 183. The expression values of the resulting probesets were mean-centered, and used to hierarchically cluster the tumors using Pearson correlation as the distance metric and average-linkage clustering (Cluster and TreeView software: <http://rana.lbl.gov/EisenSoftware.htm>). Note, the expression measurements of multiple probesets mapped to the same UniGene cluster were averaged to avoid clustering biases based on gene redundancy. Tumor and patient clinic-pathological data are published elsewhere (Bergh et al., 1995; Miller et al., 2005). The Kaplan Meier estimate was used to compute survival curves, and the p-value of the Likelihood-ratio test was used to assess statistical significance of the hazard ratios. All patients with contralateral or bilateral cancers and patients who died of their cancer 10 years post-diagnosis were censored.

I-5 TaqMan Quantitative PCR Assay for Gene-Expression Analysis

All quantitative real-time PCR analyses were performed by using the TaqMan gene-expression technology from Applied Biosystems. First, 200 ng RNA was reverse transcribed into 20 μ l cDNA using the SuperScript kit from Invitrogen. Subsequently, real-time PCR analysis was performed by using 2 μ l cDNA as template and gene-specific TaqMan probe from Applied Biosystem. All the real-time PCR analyses were done in ABI Prism 7700 sequence detection system by using TaqMan Universal PCR Master Mix and the following cycling conditions: 10 min at 95 °C and then 40 cycles of amplification (95°C for 15 s and 60°C for 60s). For normalization, the glyceraldehyde-3-phosphate dehydrogenase (GAPD) gene was analyzed as an endogenous control. Each real-time PCR analysis was done in triplicate.

All the Ct values of the real-time PCR analyses were analyzed using the comparative Ct ($\Delta\Delta$ Ct) method described by the manufacturer. Δ Ct values were calculated by first normalizing Ct values to the endogenous control (GAPD), and subsequently calculating $\Delta\Delta$ Ct values by using the Δ Ct value of 0 hr as a reference. Fold change in gene expression for a particular gene/probe between 0 hr and each of three time points (6hr, 10hr and 24hr) after p53 activation was calculated by using the formula of $2^{\Delta\Delta$ Ct} as described by the manufacturer. For each gene/probe, significance of the difference in $\Delta\Delta$ Ct value between the p53+ and p53- cell lines at each of three time points after p53 activation was investigated by the t-test. To account for the impact of multiple testing on statistical significance, the false discovery rate (FDR) was calculated for each test, and an FDR value of < 0.05 was considered to be statistically significant.

I-6 Statistical Analyses

Saturation Kinetics of p53 ChIP-PETLibrary

The unamplified p53 ChIP fragments were cloned as a PET library for sequencing analysis. To identify all unique ChIP DNA fragments captured in this library, we performed saturated sampling of sequencing. Of the 512,876 raw PET sequences, 382,742 PET sequences were mapped to single locations in the

human genome. The rest of the PETs were either not mapped or mapped multiple times in the human genome. Thereafter, only the 382,742 PETs were considered meaningful data for further analysis. The redundant PETs that mapped to the exact same location were condensed into 65,572 distinct PETs. The redundancy of these PETs is 87.2%. Assuming that as the number PET sequences increases, the discovery rate of distinct PET fragments decreases, then the total number of PET fragments identified should plateau when the sampling of PETs approaches saturation.

To calculate the level of saturation, we used the Hill function:

$$f(x) = \frac{ax^b}{c^b + x^b}$$

where x is the total number of PET matched to genome and $f(x)$ is the number of distinct PET fragments in the given sub-sample, respectively; the parameter a be the total number of ChIP DNA fragments identifiable by the PET approach, and b, c be the positive constants. We have shown that this model accurately fitted the growth kinetic curve for distinct SAGE tags (Kuznetsov et al., 2002). As shown in Figure S3, the growth function $f(x)$ fits to the PET experimental data very well. All of the three best-fit parameters are highly significant ($p < 0.0001$) ($a = 82,659 \pm 15$; $b = 0.9633 \pm 0.0003$; $c = 125,823 \pm 54$) and parameter a gives the estimation of total possible detectable distinct PETs. This model estimated a total of 78,778 detectable distinct PET fragments. Therefore, we estimated that 83% ($65,572/78,778$) of the original p53 ChIP DNA fragments captured in this library were sampled, or 13,206 ($78,778 - 65,572$) distinct PETs were missed.

Of the distinct 65,572 PETs, 61,270 (93.4%) were mapped to unique locations and thus classified as PET singletons, and 4,302 PETs were mapped in locations where other PETs also mapped, and thus defined as overlapping PETs, which were grouped into 1766 PET clusters. Together, the PET singletons and PET clusters represented 63,036 ($61,270 + 1766$) distinct PET loci. The majority (97.2%) of PET-defined loci was defined by PET singletons. 323 PET clusters had 3 or more overlapping members (PET-3+ clusters). When the sampling of PETs approaches saturation, the number of PET-3+ clusters also plateaus, as shown in the window plot in Figure Figure 4. Using the same Hill function, the PET-cluster data fitted well with this model. All of the three best-fit parameters were highly significant ($p < 0.0001$) ($a = 379 \pm 0.5$; $b = 1.511 \pm 0.0004$; $c = 108114 \pm 230.8$). This model estimated a total of 379 PET-3+ clusters if the PET sampling was saturated. Thus, we might identify 85.2% ($323/379$) of the original p53 ChIP overlapping clusters captured in this library, or we missed about 56 ($379 - 323$) p53 binding loci mapped by PET-3+ clusters.

Mapping Simulation of Overlapping PET Clusters

A Monte Carlo simulation was performed to assess the background level of overlapping PET sequences when mapped to the genome. In the simulation, we first randomly selected 65,572 genomic DNA segments (624 bp on average in size) that are unique sequences in the human genome assembly hg17, and then determined how many fragments overlapped with others. This process was repeated 100 times to compute the percentage of randomly selected DNA fragments that overlapped. The result is summarized in Table 1. Based on this simulation, we estimated that 771 PETs (1.17% of total) would result in two overlapping PETs (PET-2), 11 in PET-3, 0.0034 in PET-4, and so forth due to random chance. In contrast, the numbers of experimentally generated overlapping PETs are significantly higher than the estimated background. Therefore, it is highly likely that overlapping PETs resulted from the effect of immunoprecipitation rather than from random events.

Goodness-of-Fit Analysis for Assessing the Reliability of PET Clusters

To assess the reliability of using PET clusters as a readout for identifying p53 binding loci in the human genome we used the goodness-of-fit analysis to characterize the distribution of overlapping PET clusters based on cluster sizes, *i.e.*, the number of PET members per cluster, such as PET-2 clusters, PET-3 clusters, and so forth. The frequency of distribution of each class of PET clusters by cluster size was plotted as shown in Figure S4. The distribution curve was then decomposed into two standard Pareto

distribution functions, which allowed us to estimate the “noisy” subset of PET clusters and the “true” distribution of the PET clusters by cluster sizes (see Table 1). Additionally, an extrapolation of true distribution function to the PET-2 clusters and PET singletons (PET-1) allowed us to estimate that 3,742 PET singletons and 520 PET-2 clusters may represent real p53 binding loci in this study.

Enrichment of PET Clusters Associated to p53-Responsive Genes

Another potential piece of evidence for whether the sequenced PET fragments indeed represent p53 binding sites would be to see whether there is a strong association of p53 regulated genes with the PETs. For this purpose, 64 known p53 responsive genes identified in a previous study (Kho et al., 2004) (a review paper; JBC 279:21183) were further investigated. For 41 out of the 64 genes, PET-2+ cluster(s) were found within 100kbp of the genes. A similar scan of the 17206 annotated genes used in the expression microarray identified a total of 1428 genes having at least one PET cluster within 100kbp of the gene. The probability of picking up a gene with (at least) an associated PET cluster is thus 1428/17206 or 8.299%. Under the null hypothesis that the set of 64 genes was constructed randomly, the number of genes (out of 64 genes) found to be associated with PET clusters followed the Binomial distribution with number of samples n=64 and probability of success p=0.08299. We then computed the p-value of observing 41 (out of 64) genes with PET clusters within 100kbp, which is significant at p-value=9.1557e-14.

Association of PET-Cluster Localization at the 5’ of p53-Upregulated Genes

To simplify the visualization of the distribution of gene-associated PETs, we normalized their distances to the start of gene with respect to the flanking region size and the gene size. If a PET is D bp 5’ upstream of the gene, its normalized distance is $N=D/100000$. Should a PET fall within a gene and be D bp from the start of gene, its distance is normalized to be $N=D/\text{length_of_gene}$. For D nucleotides 3’ downstream PETs, their normalized distance is $N=1+D/\text{length_of_gene}$.

Visual inspection revealed that PETs associated with p53 up-regulated genes (PET-p53Up) tend to cluster around the start of gene. A Monte Carlo simulation and Wilcoxon rank sum test were employed to assess the statistical significance of this observation.

In the Monte Carlo simulation, with the null hypothesis that randomly picked PETs set is as close as PET-p53Up to the start of gene, we randomly picked 45 PETs and compute their average normalized distance. One million iterations of such random subset sampling were performed. The fraction of time the random subset average is less than or equal to the observed average of set PET-p53Up was reported as the p-value, which was estimated to be 3e-5.

Alternatively, the distance distribution problem could be reduced to two group problem. One group was the PET-p53Up, and the other group was the rest of the PETs. Testing for the null hypothesis that the two groups were randomly distributed, Wilcoxon rank-sum test was applied. The resultant p-value under this test was 4.934e-5.

Significance of High-Confidence PET-Cluster Loci Not in Exons

To examine whether the observation that all 474 high confidence loci are not present in exons is significant, we compared the PET singletons. In the total 20,563 PET singletons, 1,570 were exonic.

	exonic	others
PET-3+ clusters	0	474
singleton	1570	18993

The PET singletons represent random mapping noises. We performed Chi-square analysis, and the Chi-square p-value is 7.03991e-10. In addition, we modeled the problem as a binomial trial. The probability of finding non-exon hit by the 474 PET clusters is 1.447545e-8. Together, both statistical analyses

suggested that the observation of PET clusters not mapped to exon regions is significant, and is not due to random chance.

I-7 Motif Analyses of p53 Binding Sites

De Novo Motif Discovery and Further Optimization

Our initial empirical validation and manual data curation suggested that a PET cluster with 6 or more overlapping PETs (PET-6+) per cluster was highly likely to reflect actual p53 binding. Therefore, from 69 PET-6+ clusters, we picked 39 of the PET cluster defined regions as the initial seed set for the motif program training. We applied GLAM (Frith et al., 2004), a Gibbs-sampling-based algorithm for finding motif sites in unaligned sequences, to this training set and obtained 21bp long significant motif represented by a Position Specific Scoring Matrix (PSSM), depicted graphically in Figure S6A. The matrix in fact resembled the frame of the known p53 binding motif, which is composed of two repeats of the palindromic motif RRRCWWGYYY.

Two matrices were further derived from the PSSM. The *halfg1* matrix captures the first palindrome (i.e. the left half-site) and was constructed based on positions 2-11 of the discovered PSSM, while the second palindrome is characterized by the *halfg2* and was based on positions 12-21 (Figure S8 A). These seed matrices were further evolved and optimized. For this purpose, the PET-6+ cluster overlap set was supplemented with a negative-control sequence set C1, consisting of 1100 randomly picked coding sequences from the human genome. The size of the negative control was initially set to ensure that it was at least as large as the largest possible training set, i.e. PET-2+ clusters. Subsequent experiments showed that a negative set with more than 100 sequences gave almost identical results (data not shown). Under an EM (Expectation Maximization) type procedure, we first used ROVER (Haverty et al., 2004) with default settings and in conjunction with *halfg1* and *halfg2* independently to scan PET-6+ cluster overlaps (the positive set) and C1 (the control set) sequences. The matrices *halfg1* and *halfg2* were then rebuilt based on all corresponding non-overlapping hits that score above a rank-ratio threshold (Fu et al., 2004). This process was carried out iteratively until the overall degree of overabundance reported by ROVER could not increase further, i.e. the P-value stopped decreasing. Unlike other EM-based algorithms, this approach incorporates a seed PSSM and background sequences. The resultant matrix is shown in Figure S8 B. Overall, the optimized weight matrix (Figure S8 B) largely resembles that in the p53MH model (Figure S8 C). One notable difference is the fifth nucleotide of the right half-site, where the p53 PET-based model (p53PET) exhibits a stronger preference for adenine than in p53MH (Hoh et al., 2002). More importantly, the lengths of the spacers between the two half-sites in the motif sequences are considered in the p53PET model. The lengths of the spacers are predominantly zero, although some 1-bp-long and yet fewer longer spacers are also observed. This length distribution is much more specific than reported in previous studies, where spacers were simply said to vary between 0 and 14 bps. The optimized weight matrix and the highly peaked spacer length distribution make a powerful combination.

In making the predictions, the two half-sites were predicted independently and two adjacent half-sites were joined if they were the reciprocal nearest neighbors of each other and were at most 20bp apart (i.e. spacers). The score of a full-site is the sum of the logarithms of the half-site rank-ratio scores, and the strand of a full-site is determined by its highest scoring half-site. Using the PET-6+ cluster overlap and C1 sets, appropriate combinations of motif score thresholds for different spacer lengths were optimized to simultaneously achieve sensitive identification of p53 binding sequences (i.e. the PET-6+ cluster overlaps) and specific rejection of non-p53 binding sequences (i.e. the coding sequences in C1). In the 39 PET-6+ clusters tested, 30 of them contained the motif predicted by this p53 PET data derived prediction model (p53PET).

Performance Assessment

For testing the effectiveness of p53 TFBS predictions, another negative sequence set C2 was created by randomly picking a set of 1100 human coding sequences, while ensuring that C1 and C2 shared no common sequences. The 323 PET clusters with 3 or more overlapping members (PET-3+) are most likely

p53 binding as estimated (see the description on simulation of PET clusters, which indicate that it is extremely unlikely to obtain PET-3+ clusters by chance). After subtracting the above 39 PET-6+ clusters that were used for training our method, the remaining 284 clusters were used as the positive test set.

With this positive test set (PET-3+ clusters) and the negative sequence set (C2), we compared the performance of three prediction models: 1) the p53PET model as described above, 2) p53TRANSFAC, similar as the p53PET model but using the weight matrix taken from TRANSFAC (Wingender et al., 2000) (rather than using the matrix evolved using GLAM and ROVER), and 3) a much more complex model p53MH published previously (Hoh et al., 2002). In order to compare the sensitivities of the three methods across a range of specificities, we allowed each model to vary one cutoff value. Only the specificity range of 99.9-100% is shown in Figure 3B, because extremely high specificities are required to keep the number of genome-wide predictions within a reasonable bound. Within this range, the p53PET model is clearly the best among the three, as indicated by the areas under the Receiver Operating Characteristic (ROC) curve. Our model also has the highest sensitivity for the entire range of specificity as shown in Figure S8. Note that we only make predictions down to specificity of roughly 70%. Scanning the 323 PET-3+ clusters for p53 binding sites, we again observed the dominance of 0-bp spacer and the prevalence of 1-bp spacer (Figure S9), reiterating the importance of spacer length in predicting p53 binding sites.

I-8 Comparison of PET and Monotag Mappings of ChIP DNA Fragments

The ChIP DNA material fragmented by sonication and enriched by immunoprecipitation could be classified into five different categories based on their relative locations whether overlapping, adjacent, or far away from each other as illustrated in Figure S10. It is assumed that the sonication breaks were random. Therefore, the original ChIP DNA fragments were all considered to be unique with respect to each other. The original ChIP fragments were then amplified for detection by PCR or propagation during cloning. Since PET defines the two ends of each ChIP fragment, these fragments can be faithfully characterized by ChIP-PET analysis in terms of nucleotide sequences, fragment length, and the copy numbers of each fragment. Because of these features, ChIP-PET can clearly distinguish the five classes of ChIP DNA fragments. We have demonstrated in this study that the overlapping ChIP-PET fragments as PET clusters (mapping category I and II in Figure S10) represented the ChIP enrichment events and the PET singletons (mapping categories III, IV, and V) for experimental noise. In contrast, monotags (single tag per ChIP DNA fragment) generated by a monotag based approach (Impey et al., 2004) can only infer ChIP DNA fragments by marking one point of a line (while a PET demarcates the two end points of a line), and the tag count (copy number) cannot distinguish the copy number of an individual fragment from the sum of overlapping fragments. Because of these ambiguities, the monotag based approach would be incapable of differentiating the mapping categories I from III and II from IV (Figure S10). As observed, 73% of the mapping loci in I and II (PET-3+ clusters) contained a recognizable p53 binding motif, while only 2% or less of the mapping loci in category III, IV, and V contained a p53 motif. Therefore, if a monotag based approach had been applied in this study, we would either have only considered the type I and III categories as putative p53 binding loci, which would be too stringent and somewhat noisy (from type III), or consider all types except V, which would include substantial false positives.

Supplemental References

Bunz F., Dutriaux A., Lengauer C., Waldman T., Zhou S., Brown J. P., Sedivy J. M., Kinzler K. W., Vogelstein B. (1998). Requirement for p53 and p21 to sustain G2 arrest after DNA damage. *Science*. 282 , 1497-501.

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

Fu Y., Frith M. C., Haverty P. M., Weng Z. (2004). MotifViz: an analysis and visualization tool for motif discovery. *Nucleic Acids Res.* 32 (Web Server issue): p. W420-423.

Haverty P.M., Hansen U., 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. *et al.* (2002). The p53MH algorithm and its application in detecting p53-responsive genes. *Proc. Natl. Acad. Sci. U S A.* 99, 8467-8472.

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. (2004). Defining the CREB regulon: A genome-wide analysis of transcription factor regulatory regions. *Cell* 119, 1041-1054.

Kaaser, 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. U S A.* 99, 95-100.

Kho P. S. *et al.* (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., Iyer, V. R. (2005). Mapping DNA-protein interactions in large genomes by sequence tag analysis of genomic enrichment. *Nature Methods* 2, 47-53.

Kuznetsov V.A., Knott G.D., Bonner R.F. (2002). General statistics of stochastic process of gene expression in eukaryotic cells. *Genetics* 161, 1321-1322.

Ng H.H., Robert F., Young R. A., 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. *et al.*, (2005). Gene identification signature (GIS) analysis for transcriptome characterization and genome annotation. *Nature Methods* 2, 105-111.

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

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

Wingender E. *et al.* (2000). TRANSFAC: an integrated system for gene expression regulation. *Nucleic Acids Res.* 28, 316-319.

II. Supplemental Figures

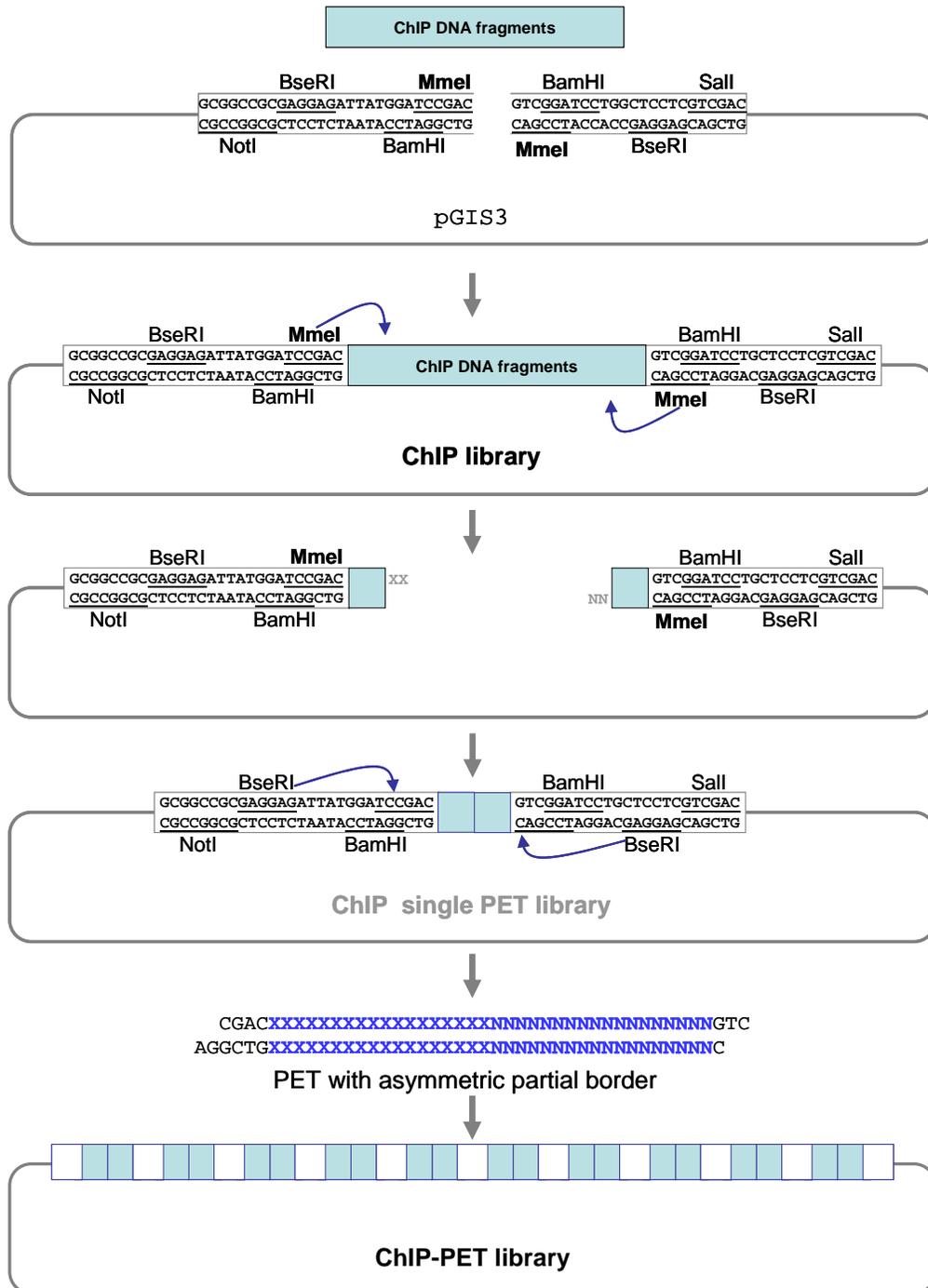


Figure S1. ChIP-PET Library Construction

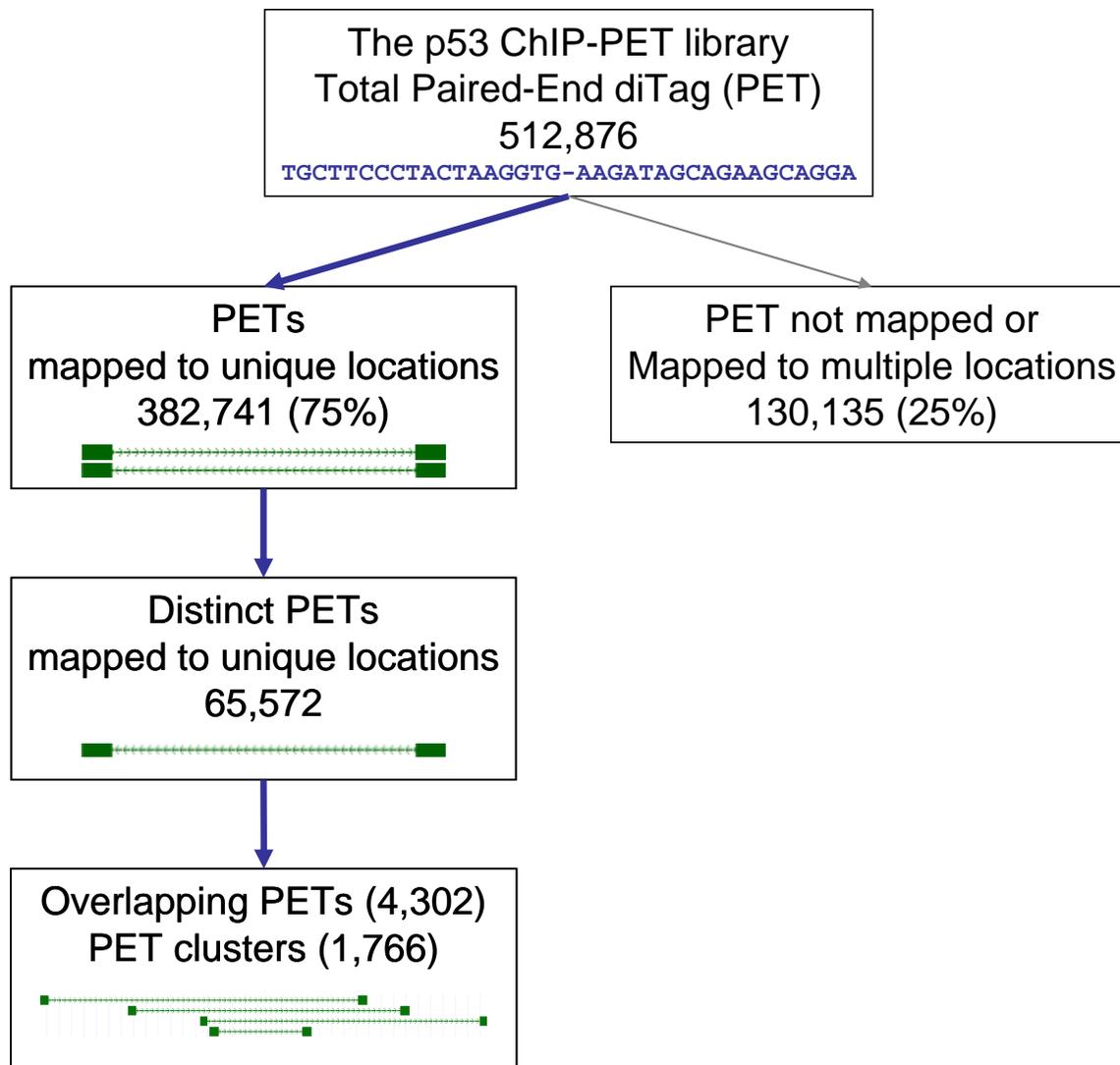


Figure S2. PET Sequence Processing and Mapping to the Human Genome

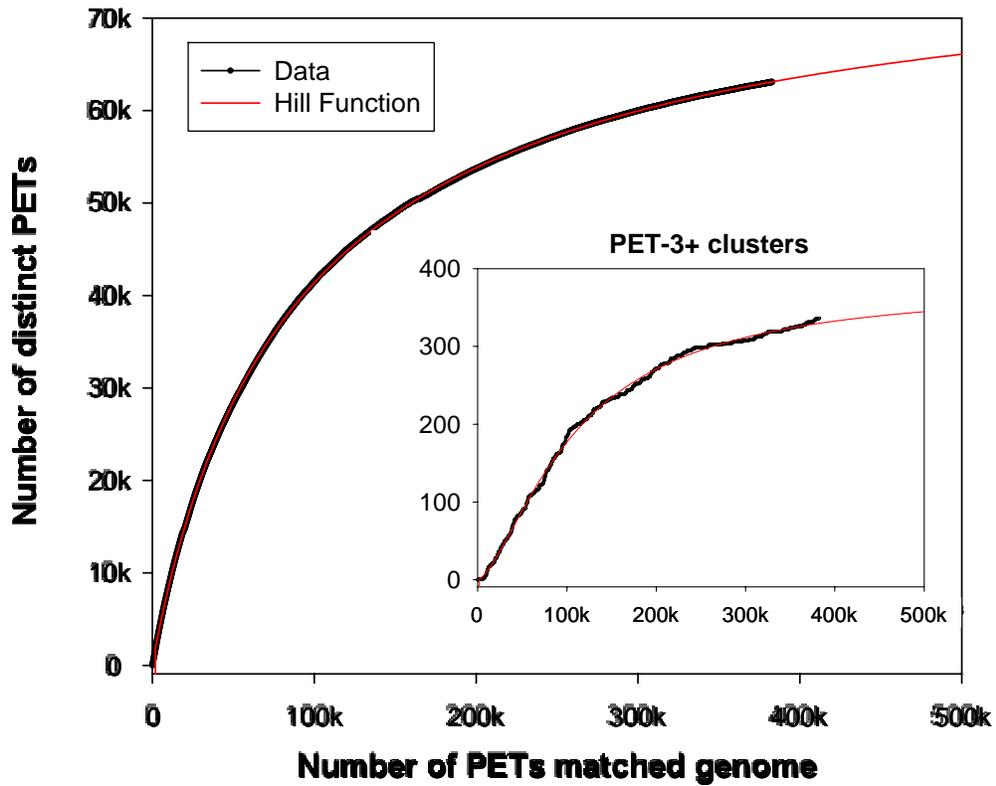


Figure S3. Hill Function Curve for Estimating the Total Distinct PET Coverage

The curve (black) of number of distinct PETs in total raw PET sequences fitted tightly with the Hill function model (red). The Hill function model estimated a total of 78,778 detectable distinct PET fragments when sampling was saturated. Inset: A similar Hill function curve for PET clusters with 3 or more PET members in each cluster (PET-3+ clusters). This PET cluster Hill model estimated a total of 379 PET-3+ clusters if the PET sampling was saturated.

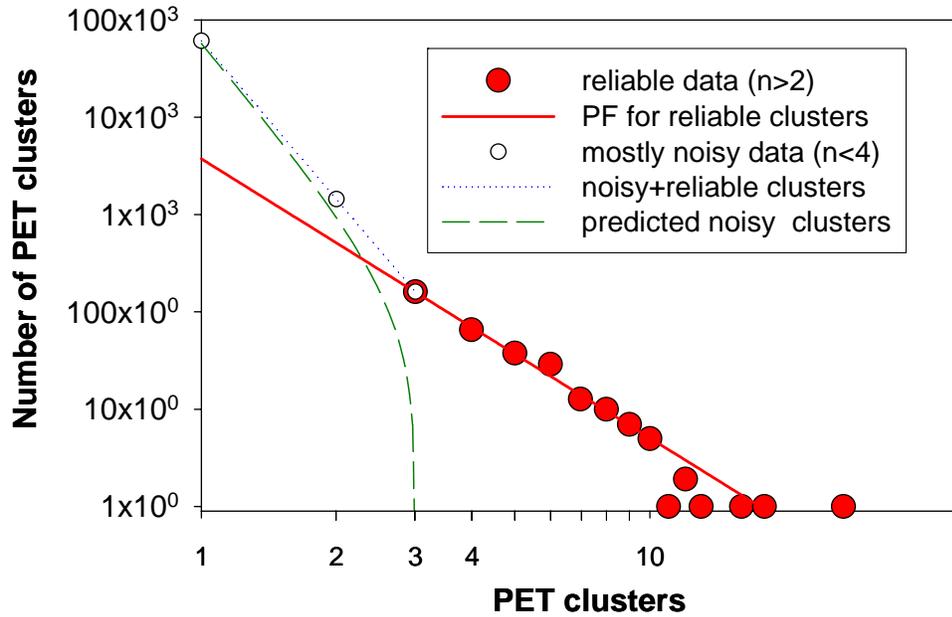
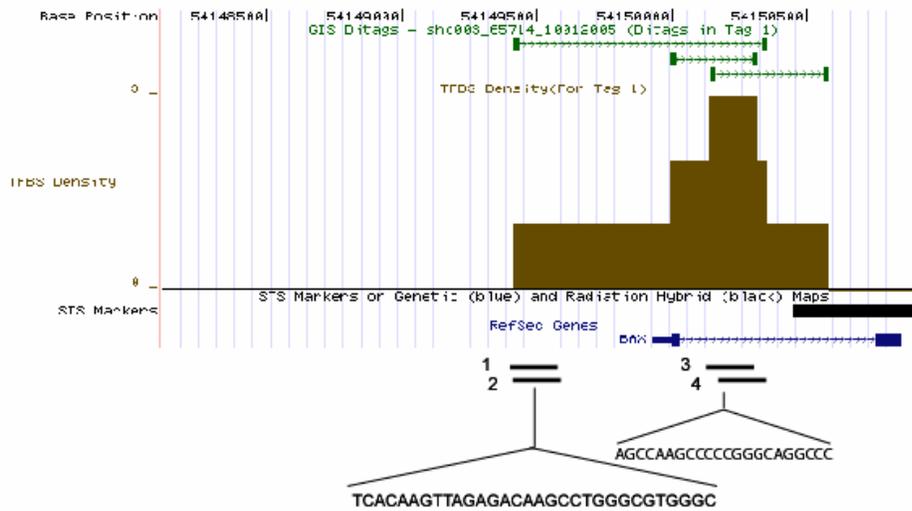


Figure S4. Goodness-of-Fit Analysis for Assessing the Reliability of PET Clusters

PET singletons and PET clusters were plotted based on the cluster size and the number of each cluster type. It is a log-log plot. Empty circles: PET singleton, PET-2 cluster, and PET-3 cluster. Red circle: high confident PET clusters (PET-3+ clusters that have 3 or more overlapping members). Blue dotted line: Standard Pareto distribution function (power law) with best-fit slope parameter $k=4.427$. Red line: best-fit standard Pareto distribution function with slope parameter $k=1.869$. This function fitted the “true” part of the observed distribution. Discontinuous green line: “noisy” PET cluster distribution results in the subtraction of the probability distribution function of the “true” clusters from the originally observed distribution function.

A



B

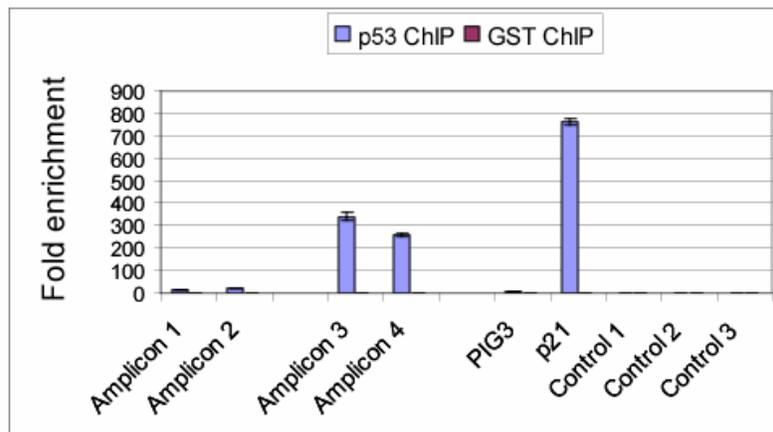


Figure S5. ChIP-qPCR Assay for p53 Binding Sites on the BAX Promoter

(A) A PET-3 cluster was mapped to the BAX promoter/first intron region, which covers the previously known site and identifies a new binding site. The two sites are 726 bp apart, and only the new site is in the PET cluster overlap sequence. We further validated these two sites by ChIP-qPCR analysis. The locations of the 4 amplicons were shown. Amplicons 1 and 2 interrogated the known site, and amplicons 3 and 4 interrogated the new site.

(B) A plot of ChIP enrichment of p53 binding to Bax, p21 and PIG3 promoter. Three other control regions showed only background enrichments of 1. In this experiment, the Amplicons 1 and 2 (known p53 site to BAX) showed an average 18-fold enrichment, which is at the same level as for PIG3, but much lower than p21. While the Amplicons 3 and 4 (new p53 site for BAX) showed an average enrichment of 300-fold.

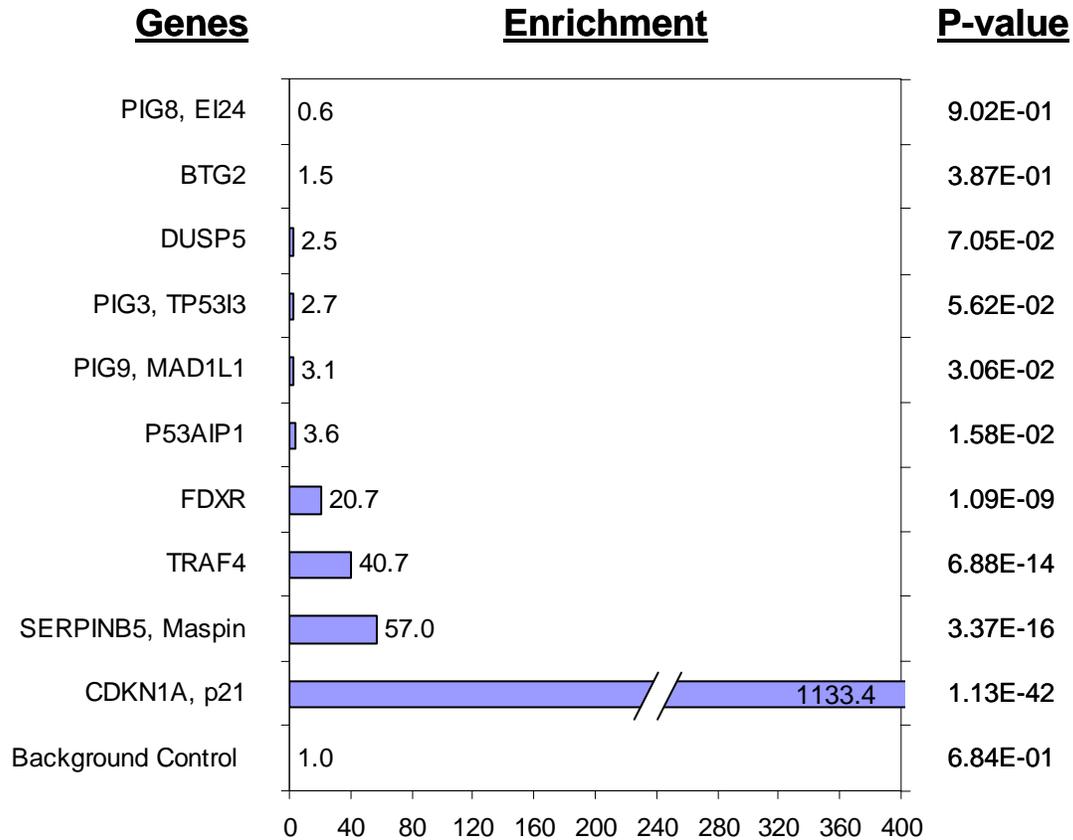


Figure S7. ChIP-PCR Assay of Nine Previously Known p53 Targets

The 9 previously known p53 targets were not detectable by ChIP-PET analysis in HCT116 cells. To investigate the cause, we conducted ChIP-qPCR assay of the 9 loci using p53 ChIP enriched DNA fragments as template. The p53 binding locus on the promoter of CDKN1A (p21) was included as a positive control along with GST pull-down as background control. Three (SERPINB5, TRAF4, FDXR) of the 9 p53 binding loci showed significant enrichment, while the others (6) had only marginal or below background signal. The p-values were calculated based on the background as a Gaussian distribution and then calculate the p-values for each of the enrichment level.

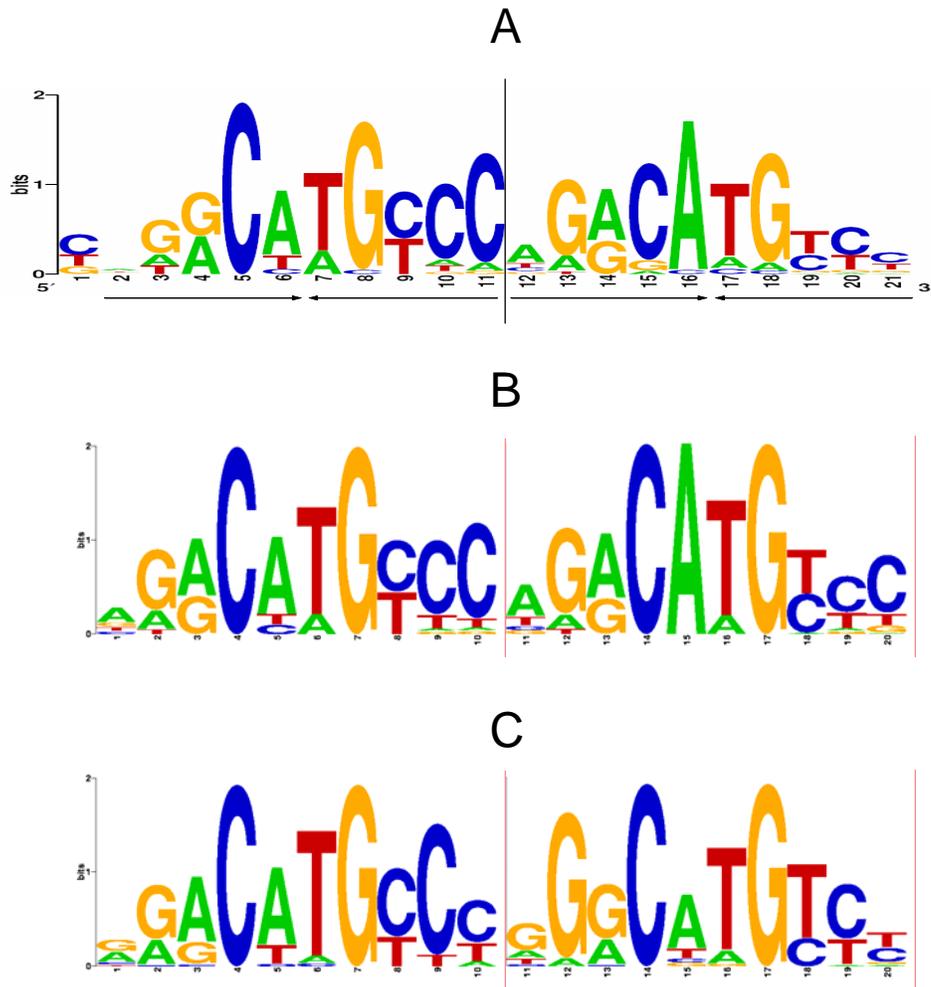


Figure S8. Development of p53 Binding-Motif-Finding Model

(A) Sequence logo of p53 motif derived by GLAM from the initial training set of PET-6+ cluster sequences.

(B) Sequence logo derived from the optimized p53 PET model.

(C) Sequence logo derived from the p53MH model (Hoh et al., 2002).

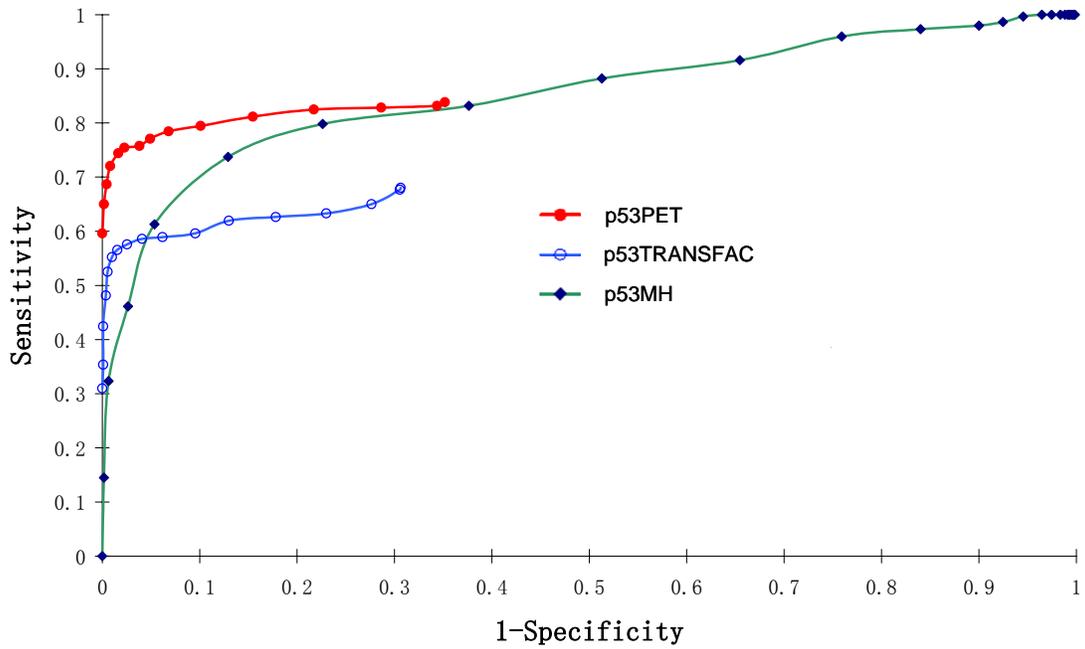


Figure S9. Performance Assessment of p53 Binding-Motif-Finding Models

Receiver Operating Characteristic (ROC) curves comparison among p53PET, p53MH, and p53TRANSFAC. Evidently, the proposed model p53PET outperformed the existing models.

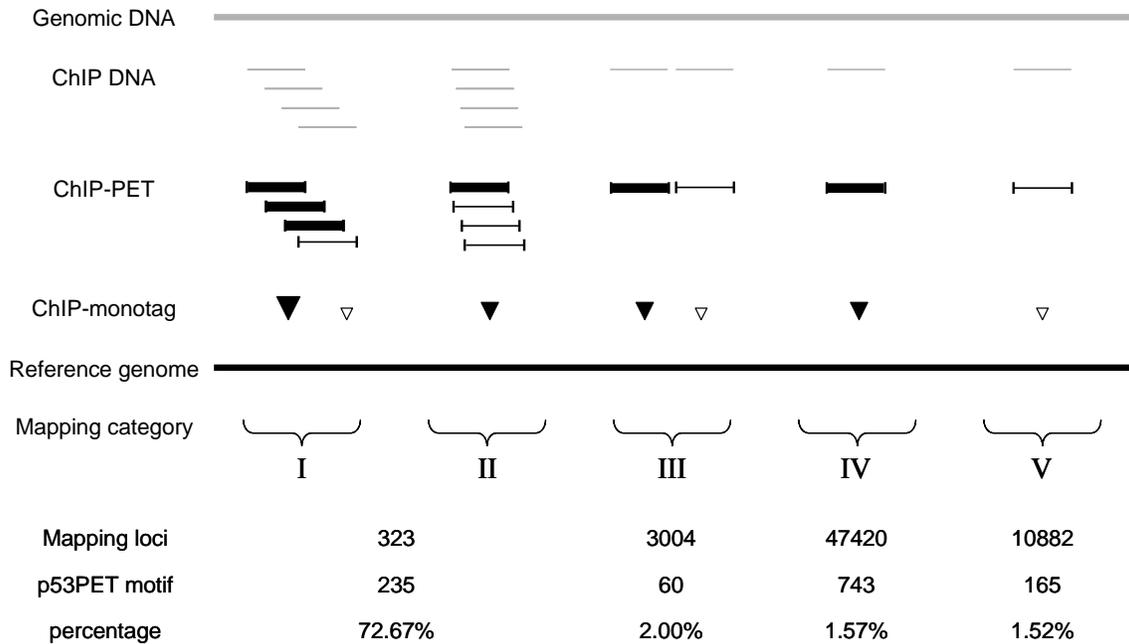
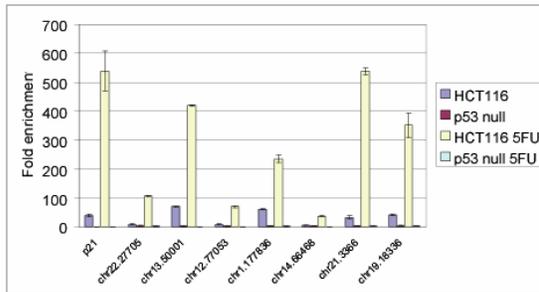


Figure S10. Detection of ChIP DNA Fragments by PET or Monotag

Genomic DNA (long gray line) was fragmented by sonication. It is assumed that all ChIP DNA fragments (short gray bar) are distinctive from each other and occurs as single copies. During the molecular cloning of tag-based analysis, the original ChIP DNA or tags were amplified (plasmid propagation or PCR). It is known that such amplifications are not always proportional and could introduce bias. The amplification effect could be reflected in copy numbers as detected either by PETs (heavy bars for PETs with multiple copies, light bars for single copy PETs) or by monotags (solid triangles for tags with multiple copies, open triangles for single copy tags). Based on mapping to the reference genome, the ChIP DNA fragments were classified into 5 categories: I) a number of overlapping fragments but spread out, detected as a PET cluster, or as a number of monotags (with single or multiple copies) clustered near each other; II) a number of tightly overlapping fragments, detected as a PET cluster, or as a monotag with multiple copies; III) a number of non-overlapping adjacent fragments detected as distinct PET singletons (with single or multiple copies), or as cluster of monotags (with single or multiple copies); IV) a fragment was located alone, and could be detected as a PET singleton with multiple copies, or a monotag with multiple copies; V) a single fragment detected as a single-copy PET singleton, or a single-copy monotag. The numbers provided here are for the mapping loci of each category (I and II are combined for PET-3+ clusters), the occurrence of the p53 motifs predicted by the p53PET program, and the percentage of motif-containing loci in each category.

A



B

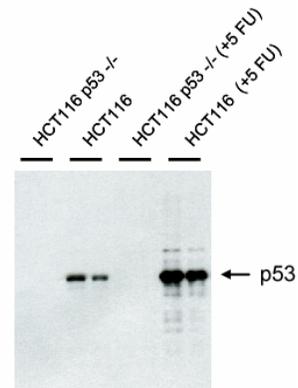


Figure S11. Specificity of DO1 Monoclonal Antibody against p53

(A) Extracts prepared from HCT116 cells, HCT116 p53 $-/-$ cells, HCT116 cells treated with 5 FU for 6 hours, HCT116 p53 null cells treated with 5 FU for 6 hours were immunoprecipitated with DO1 monoclonal antibody. Standard deviation is shown.

(B) Western blot of extracts from (A) using DO1 monoclonal antibody. Two different amounts of extract were loaded for each sample.

III. Supplemental Tables

Table S1. PET and PET-Cluster Distribution in the Human Genome

Chr.	Chr. Size (mb)	Chr. Size (mb) [Masked]	PET	PET Density (PET/mb)	PET Density (PET/mb) [Masked]	PET Clusters	Cluster Density (Cluster/mb)	Cluster Density (Cluster/mb) [Masked]
1	222.828	114.617	5337	23.95	46.56	149	0.67	1.30
2	237.503	128.039	5583	23.51	43.60	145	0.61	1.13
3	194.636	101.464	4543	23.34	44.77	118	0.61	1.16
4	187.161	96.080	4285	22.89	44.60	104	0.56	1.08
5	177.703	92.653	4059	22.84	43.81	92	0.52	0.99
6	167.318	88.726	4140	24.74	46.66	106	0.63	1.19
7	154.759	80.619	3454	22.32	42.84	109	0.70	1.35
8	142.613	74.057	4174	29.27	56.36	124	0.87	1.67
9	117.781	61.084	2661	22.59	43.56	75	0.64	1.23
10	131.614	70.302	3324	25.26	47.28	99	0.75	1.41
11	131.131	67.074	3068	23.40	45.74	72	0.55	1.07
12	130.259	65.815	3212	24.66	48.80	105	0.81	1.60
13	95.56	52.019	2127	22.26	40.89	47	0.49	0.90
14	88.291	45.956	2142	24.26	46.61	49	0.55	1.07
15	81.342	43.139	1867	22.95	43.28	56	0.69	1.30
16	78.885	40.353	1847	23.41	45.77	67	0.85	1.66
17	77.8	40.979	2235	28.73	54.54	74	0.95	1.81
18	74.656	41.253	1738	23.28	42.13	35	0.47	0.85
19	55.786	23.963	1260	22.59	52.58	39	0.70	1.63
20	59.505	30.618	1392	23.39	45.46	30	0.50	0.98
21	34.17	18.502	819	23.97	44.27	28	0.82	1.51
22	34.765	18.135	787	22.64	43.40	27	0.78	1.49
X	150.394	61.745	1601	10.64 (21.28)	25.93 (51.86)	31	0.206 (0.41)	0.50 (1.00)
Y	24.872	9.537	34					
Total	2851.332	1466.729	65689		44.79	1781		1.21

Table S2. PET Clusters Mapping to Known p53-Responsive Genes

RefSeq Accession	Gene Name	PET-Cluster Location	Cluster Size	PET-to-Gene Location*	p53 Motif	Previous Binding Evidence**
NM_006142	14-3-3-sigma; SFN	chr1:26877722-26878743	1	3' (2633)	yes	None
NM_004024	ATF3	chr1:209157999-209159107	6	5' (12172)	yes	Mobility shift
NM_138761	BAX	chr19:54149416-54150580	3	5' (581)	yes	ChIP-PCR in multiple cell types
NM_000623	BDKRB2	chr14:95736756-95737306	2	5' (3644)		Mobility shift
NM_006763	BTG2					Mobility shift
NM_020375	C12orf5					None
NM_004354	CCNG2	chr4:78611167-78612302	6	intron10	yes	None
	CCNG2	chr4:78620134-78621688	7	intron10	yes	
	CCNG2	chr4:78622099-78622692	2	intron10	yes	
NM_000389	CDKN1A, p21	chr6:36742675-36743642	5	5' (10838)	yes	
	CDKN1A, p21	chr6:36751902-36754502	13	5' (2225)	yes	ChIP-PCR in multiple cell types
NM_022457	Cop-1; Cop1; RFWD2	chr1:172954068-172954909	2	5' (46297)	yes	None
NM_000107	DDB2 (Reed1)	chr11:47192744-47194120	3	5' (496)	yes	Mobility shift
NM_007026	DUSP14					None
NM_004419	DUSP5					ChIP-PCR (HCT115 treated with adriamycin)
NM_004431	EPHA2	chr1:16255985-16256808	2	5' (28115)		None
NM_004110	FDXR					ChIP-PCR in H1299 cells overexpressing p53
NM_002023	FMOD					None
NM_001924	GADD45	chr1:67863807-67864861	4	intron3	yes	Mobility shift; DNaseI footprinting

NM_004864	GDF15; PLAB; MIC-1	chr19:18344715- 18345958	3	5' (12009)	yes	None
NM_005497	GJA7					None
NM_002066	GML	chr8:143893708- 143894655	7	5' (18563)	yes	None
M96843	ID2B					None
NM_016545	IER5	chr1:177799988- 177800842	5	3' (8357)		None
NM_019099	LOC55924					None
NM_002392	MDM2	chr12:67488794- 67489505	2	5' (464)		ChIP-PCR in multiple cell types
NM_000249	MLH1	chr3:36949079- 36949987	3	5' (59995)	yes	ChIP-PCR in human fibroblast
NM_004148	NINJ1					None
NM_022112	P53AIP1					ChIP-PCR in HCT116 5FU for 12 hours
NM_014454	PA26; SESN1	chr6:109466894- 109469143	2	intron1		None
NM_022470	PAG608; WIG1	chr3:180269957- 180271353	4	intron1	yes	Mobility shift
NM_003884	PCAF	chr3:20094710- 20095306	2	intron2	yes	ChIP-PCR in human breast cancer cell
NM_182649	PCNA	chr20:5048561- 5049184	3	5' (2633)		ChIP-PCR in TR9-7 cell
NM_002307	PIG1 / LGALS7	chr19:43957618- 43958245	3	5' (13445)		None
NM_199161	PIG4; SAA1					None
NM_001012302	PIG5 / TP53I5					None
NM_016335	PIG6 / PRODH	chr22:17291232- 17291990	2	intron1	yes	None
NM_004862	PIG7 / LITAF	chr16:11563564- 11564336	2	intron1		None
NM_004879	PIG8 / EI24					Mobility shift
NM_003550	PIG9 / MAD1L1					ChIP-PCR (Hep3B transfected with p53)
NM_003633	PIG10 / ENC1	chr5:73978546- 73979692	2	5' (6273)		None

NM_006034	PIG11 / TP53I11					None
NM_198797	PIG12 / PTGES					None
NM_004073	PLK3 / CNK	chr1:44934161-44934929	2	5' (200)		None
D90070	PMAIP1 / NOXA					None
NM_003620	PPM1D / Wip-1					None
NM_006404	PROCR / EPCR	chr20:33254851-33257922	2	3' (26025)		None
NM_000314	PTEN	chr10:89592571-89593382	3	5' (19792)	yes	Mobility shift
NM_014417	PUMA / BBC3	chr19:52425976-52427186	2	5' (895)		ChIP-PCR in multiple cell types
NM_015920	RPS27L	chr15:61235856-61237660	9	5' (931)	yes	None
AB036063	RRM2B / p53R2	chr8:103317718-103318997	4	intron1	yes	None
NM_006096	RTP/rit42 / NDRG1	chr8:134472408-134474124	3	5' (93731)		ChIP-PCR in HCT116 treated with Dox
NM_005978	S100A2	chr1:150352200-150353947	4	5' (821)		None
NM_013376	SEI1 / SERTAD1					None
NM_002639	SERPINB5 / Maspin					ChIP-PCR in breast cancer MCF10A, MDA-MB, BT549 and HS578T cells
NM_031459	SES2 / Hi95	chr1:28302743-28303867	2	3' (9600)		None
NM_006622	SNK / PLK2	chr5:57793144-57794331	4	5' (1300)		ChIP-PCR in U2OS osteosarcoma cells
NM_003236	TGFA	chr2:70734862-70736664	10	5' (42277)	yes	Mobility shift
NM_003246	THBS1 / TSP-1	chr15:37645510-37647421	2	5' (13151)		None
NM_003842	TNFRSF10B / Killer/DR5	chr8:22981574-22983332	3	5' (848)		ChIP-PCR in H1299 cells
NM_003840	TNFRSF10D					None

NM_000043	TNFRSF6 / Fas/APO1	chr10:90850518-90851917	2	3' (84997)		None
	TNFRSF6 / Fas/APO1	chr10:90852441-90853476	2	3' (86920)		
NM_005749	TOB1	chr17:46335537-46336323	2	5' (39153)		None
NM_007233	TP53AP1 / TP53TG1	chr7:86604121-86605580	3	intron2		None
NM_004881	TP53I3 / PIG3					ChIP-PCR in HCT116 cells
AL133074	TP53INP1 / P53DINP1	chr8:96029784-96031833	2	5' (1066)		None
NM_005427	TP73	chr1:3650125-3652312	2	intron2		ChIP-PCR in human U2OS cells
NM_004295	TRAF4	chr17:24094143-24094535	1	5' (630)	yes	Mobility shift
NM_004628	XPC	chr3:14194514-14195958	3	5' (455)		Mobility shift

Genes with multiple names are included and separated by “/”. Genes not hit by PET clusters are in red.

*Numbers in parentheses are distance of base pairs from the ends (5' or 3') of RefSeq genes.

**Evidence is compiled from literature.

Table S3. p53 Binding Loci and p53 Binding Motifs in Chromosome 21 and 22

PET cluster-defined loci location	PET cluster size	PET cluster length (bp)	Gene id	Location to gene	Distance (bp) to	Affy p53 loci location	p53PET motif
* chr21:33660665-33662530	8	1866	IFNAR1	3'	6672 bp		yes
* chr22:27702966-27705354	5	2389	AB051436	5'	98358	chr22:27,703,126-27,704,528	yes
chr21:18002576-18003298	4	723	AF304443	3'	932		yes
chr21:15458636-15459767	3	1132	C21orf116	3'	112073	chr21:15,458,831-15,459,120	
chr22:27036061-27037690	3	1630	KIAA1043	5'	208350		yes
chr21:19888722-19890405	2	1684	PRSS7	5'	1190878		yes
chr21:29697827-29698558	2	732	BACH1	3'	41741		yes
chr21:44842261-44843024	2	764	KRTAP10-7	5'	1900		yes
chr22:17291232-17291990	2	759	** PRODH	intron1	-6370		yes
chr22:17937314-17938068	2	755	CLDN5	5'	50837		yes
chr22:22492800-22493381	2	582	SMARCB1	intron		chr22:22,492,363-22,493,807	yes
chr22:44185994-44186654	2	661	FBLN1	5'	32642		yes
chr22:48474602-48475733	2	1132	BRD1	3'	12059		yes
						chr21:26,684,125-26,684,805	yes
						chr22:25,814,326-25,814,690	yes
						chr22:29,897,780-29,898,704	yes
						chr22:36,619,194-36,619,712	yes

* Validated by ChIP-PCR assay

** Previously known p53-responsive gene; also known as PIG6

Table S4. PET-Cluster-Identified p53 Binding Loci and Associated Genes

Cluster Size	PET Cluster Location	Known Gene	Distance (bp)	Location	p53-Induced Expression	Gene Function (GO)
2	chr14:68523977-68525535	ACTN1	8334	5'		actin cytoskeleton (0015629)
3	chr9:121164065-121165037	GSN	0.68	internal intron		actin cytoskeleton (0015629)
3	chr12:55710965-55712102	MYO1A	0.84	internal intron	Up	actin filament-based movement (0030048)
2	chr1:177216737-177217336	ACBD6	13601	5'		acyl-CoA binding (0000062)
2	chr16:76583084-76584178	KIAA1576	11580	3'		alcohol dehydrogenase activity (0004024)
2	chr17:19596922-19598200	ALDH3A1	4584	5'		aldehyde metabolism (0006081)
3	chr2:162379239-162382441	SLC4A10	0.2	first intron	Down	anion transport (0006820)
3	chr5:118687056-118688091	TNFAIP8	31357	5'	Down	anti-apoptosis (0009616)
6	chr4:90958946-90960704	SNCA	43178	3'		anti-apoptosis (0009616)
3	chr19:54149416-54150580	BAX	0.03	first intron	Up	apoptosis (0006915)
4	chr1:67863807-67864861	GADD45A	0.29	internal intron	Up	apoptosis (0006915)
7	chr8:143893708-143894655	GML	18563	5'	Up	apoptosis (0006915)
3	chr16:23628038-23629180	ERN2	0.13	internal intron		apoptosis (0006915)
2	chr1:50780893-50781817	FAF1	0.69	internal intron		apoptosis (0006915)
7	chr1:241224465-241225547	PNAS-4	28866	3'		apoptosis (0006915)
4	chr19:6527895-6529074	TNFSF7	7776	3'		apoptosis (0006915)
2	chr6:110034808-110035812	AK124171	0.16	internal intron		ATP binding (0005524)
10	chr12:12394778-8-123949064	DHX37	7186	3'		ATP-dependent helicase activity (0008026)
3	chr10:96335617-96336388	HELLS	0.72	internal intron		ATP-dependent helicase activity (0008026)
4	chr7:137306143-137307138	AKR1D1	45838	3'		bile acid catabolism (0030573)
4	chr13:11040193-3-110403257	ANKRD10	36516	5'	Down	biological process unknown
2	chr12:47144559-47145569	ANP32D	7145	5'	Down	biological process unknown
2	chr2:150292543-150293901	C2orf25	22762	5'	Down	biological process unknown
6	chr6:20974586-20976134	CDKAL1	0.48	internal intron	Down	biological process unknown
4	chr10:73726812-73728300	DDIT4	21013	3'	Up	biological process unknown
2	chr3:42800284-42800713	HIG1	85	3'	Down	biological process unknown
2	chr2:226326587-226328415	AK055226	0.71	internal intron	Up	biological process unknown
2	chr7:33164389-33164898	B1	0.03	internal intron	Up	biological process unknown
4	chr7:33100235-33101679	B1	0.3	internal intron	Up	biological process unknown
3	chr19:18344715-18345958	BC004942	2956	3'	Up	biological process unknown
2	chr1:28383535-28384557	FLJ20045	36918	3'	Up	biological process unknown
6	chr1:111468427-111470189	FLJ22457	9354	5'	Up	biological process unknown

2	chr13:20422679-20424088	LATS2	22490	3'	Up	biological process unknown
5	chr1:141466453-141468029	AB007925	1922	5'		biological process unknown
2	chr7:130296155-130297251	AB029551 (mRNA)	44062	3'		biological process unknown
2	chr10:24617458-24618206	AB033043	0.7	internal intron		biological process unknown
3	chr15:50742608-50743868	AB037791	0.15	first intron		biological process unknown
5	chr17:58612743-58614058	AB046856	17821	5'		biological process unknown
5	chr22:27702966-27705354	AB051436	98358	5'		biological process unknown
3	chr7:2508083-2509081	AB075830	0.44	first intron		biological process unknown
5	chr12:19475470-19477289	AEBP2	6610	5'		biological process unknown
2	chr1:98642523-98643036	AF005043 (mRNA)	47714	3'		biological process unknown
2	chr1:98693292-98694324	AF005043 (mRNA)	2511	5'		biological process unknown
3	chr22:27036061-27037690	AF119850	18726	3'		biological process unknown
2	chr9:84356926-84357776	AF174394	90262	3'		biological process unknown
4	chr21:18002576-18003298	AF304443	932	3'		biological process unknown
4	chr16:73458784-73460362	AF370390	4609	3'		biological process unknown
3	chr1:16631449-16631910	AF379627	2311	3'		biological process unknown
2	chr5:129514886-129515524	AJ578034	0.87	internal intron		biological process unknown
2	chr2:12599036-12600002	AK001558 (mRNA)	0.88	internal intron		biological process unknown
3	chr4:169762141-169763319	AK055595	24883	5'		biological process unknown
2	chr12:10241994-7-102420629	AK058052	27734	5'		biological process unknown
3	chr12:8484013-8485503	AK074886	43347	5'		biological process unknown
4	chr18:10199133-10200034	AK091542	33250	3'		biological process unknown
2	chr8:126868504-126869122	AK093407 (mRNA)	0.34	internal intron		biological process unknown
5	chr11:61522493-61524204	AK093779	28955	3'		biological process unknown
3	chr15:40750917-40751625	AK122666	1377	3'		biological process unknown
2	chr2:208598880-208599522	AK123204	0.54	internal intron		biological process unknown
4	chr6:48186032-48187338	AK123835	41684	5'		biological process unknown
2	chr1:219366949-219368121	AK125494	54407	3'		biological process unknown
3	chr19:51328391-51329532	AK125754	0.66	first intron		biological process unknown
5	chr17:34456386-34457200	AK125814	0.72	internal intron		biological process unknown
3	chr16:44958822-44962053	AK125968	98698	3'		biological process unknown
4	chr16:44980671-44986383	AK125968	74368	3'		biological process unknown
2	chr5:99545387-99546467	AK126705 (mRNA)	89596	5'		biological process unknown
3	chr10:61917221-61918368	AK126851	0.16	first intron		biological process unknown

2	chr14:88153029-88154009	AK127024	0.94	internal intron	biological process unknown
3	chr1:170342552-170343949	AK127170	0.43	internal intron	biological process unknown
2	chr20:46055158-46055622	AK127532	71005	5'	biological process unknown
2	chr4:55126825-55128330	AK127618 (mRNA)	65811	3'	biological process unknown
2	chr10:13345977-3-133460132	AK128177	0.7	internal intron	biological process unknown
6	chr9:81865320-81866112	AK128831	25600	3'	biological process unknown
2	chr16:74494786-74495821	AK129593 (mRNA)	21401	5'	biological process unknown
2	chr19:55246645-55247401	AK130360	0.06	first intron	biological process unknown
2	chr3:25308734-25309397	AK131021 (mRNA)	90205	3'	biological process unknown
3	chr12:97287654-97289608	AL833235	60480	3'	biological process unknown
2	chr16:88035835-88036929	ANKRD11	0.21	first intron	biological process unknown
2	chr16:88051454-88052975	ANKRD11	0.14	first intron	biological process unknown
3	chr6:109323787-109325865	ARMC2	0.38	internal intron	biological process unknown
3	chr16:27912374-27913306	AY358206	0.25	first intron	biological process unknown
2	chr8:53621206-53621781	AY358213	0.6	internal intron	biological process unknown
2	chr15:73388842-73389865	BC006530	6483	5'	biological process unknown
2	chr13:75994012-75995491	BC012127 (mRNA)	94183	3'	biological process unknown
5	chr5:72286527-72288315	BC014311	500	5'	biological process unknown
2	chr4:119764568-119765029	BC014590	30068	5'	biological process unknown
3	chr11:11174917-6-111750529	BC029570	18521	5'	biological process unknown
5	chr5:27505201-27508642	BC039509	260	5'	biological process unknown
2	chr20:4379823-4380481	BC039673 (mRNA)	19466	5'	biological process unknown
2	chr10:16277178-16279385	BC040585 (mRNA)	33709	3'	biological process unknown
7	chr1:184338383-184340304	BC040856	1710	5'	biological process unknown
2	chr4:165999595-166000793	BC062791 (mRNA)	32094	5'	biological process unknown
4	chr14:22075307-22076384	BC063432	0.02	internal intron	biological process unknown
2	chr16:51968563-51970133	BX537874	6180	5'	biological process unknown
4	chr1:199702497-199703908	C1orf37	4272	5'	biological process unknown
3	chr20:33358933-33360037	C20orf44	657	5'	biological process unknown
2	chr4:113550507-113551321	C4orf16	1700	3'	biological process unknown
2	chr6:119001605-119003029	C6orf204	7433	5'	biological process unknown
2	chr6:119076858-119077690	C6orf204	0.02	first intron	biological process unknown
3	chr6:166731532-166732582	C6orf83	5093	5'	biological process unknown
6	chr9:115784942-115786619	C9orf27	17902	5'	biological process unknown

2	chr1:6858299-6859653	CAMTA1	0.08	internal intron	biological process unknown
2	chr1:158795674-158796416	CAPON	0.08	first intron	biological process unknown
2	chr9:26966133-26967862	CCDC2	0.19	internal intron	biological process unknown
3	chr10:89592571-89593382	CFLP1	0.9	internal intron	biological process unknown
2	chr8:57329451-57330247	CHCHD7	35723	3'	biological process unknown
2	chr10:125287767-125288788	CR607950 (mRNA)	19794	3'	biological process unknown
2	chr1:216147723-216149091	CR609066 (mRNA)	20606	3'	biological process unknown
2	chr2:68882195-68884048	D29642	29722	5'	biological process unknown
2	chr3:101112955-101114217	DOC1	35287	5'	biological process unknown
2	chr17:34738689-34739376	FBXL20	0.51	internal intron	biological process unknown
2	chr1:229621902-229622662	FLJ11383	0.29	internal intron	biological process unknown
2	chr3:48448871-48449598	FLJ12436	0.87	internal intron	biological process unknown
2	chr4:38393481-38394661	FLJ13197	42226	3'	biological process unknown
2	chr2:172112599-172113283	FLJ13984	0.03	first intron	biological process unknown
2	chr15:42724512-42726099	FLJ21439	0.41	internal intron	biological process unknown
2	chr8:143407398-143407877	FLJ31164	0.39	internal intron	biological process unknown
2	chr3:196389663-196390223	FLJ35155	0.41	internal intron	biological process unknown
2	chr2:108519802-108520625	GCC2	3588	5'	biological process unknown
2	chr10:118527092-118527865	HSPA12A	35017	5'	biological process unknown
2	chr10:86274468-86275311	KIAA1128	6216	3'	biological process unknown
2	chr1:142475316-142476424	KIAA1245	0.32	internal intron	biological process unknown
2	chr9:112486398-112487823	KIAA1958	0.91	internal intron	biological process unknown
6	chr1:111618845-111620291	LOC128344	11175	3'	biological process unknown
2	chr12:2676616-2677718	LOC283439	586	3'	biological process unknown
2	chr19:4575513-4576350	MGC17791	14179	5'	biological process unknown
2	chr6:169626932-169628334	MGC43690	46604	3'	biological process unknown
3	chr1:120553086-120555329	MGC57827	0.1	first intron	biological process unknown
2	chrX:17325543-17326285	NHS	0.48	first intron	biological process unknown
2	chr10:13210351-13211978	OPTN	0.76	internal intron	biological process unknown
2	chr14:70612561-70613759	PCNX	0.83	internal intron	biological process unknown
2	chr1:171298284-171299542	RABGAP1L	0.53	internal intron	biological process unknown
2	chr11:1808137-1808988	SYT8	2609	5'	biological process unknown
2	chr6:76038723-76040001	TMEM30A	0.36	internal intron	biological process unknown
3	chr7:72392433-72393124	BCL7B	2212	3'	biological_process unknown (0000004)

6	chr9:91988780-91990143	SPTLC1	31560	5'		biosynthesis (0009058)
2	chr15:29229665-29230722	TRPM1	48444	5'	Down	calcium channel activity (0005262)
4	chr1:150352200-150353947	S100A2	821	5'	Up	calcium ion binding (0005509)
3	chr17:57846357-57848070	FLJ25818	0.97	internal intron		calcium ion binding (0005509)
6	chr9:111986264-111988117	SUSD1	0.22	internal intron		calcium ion binding (0005509)
3	chr12:73857967-73858516	KCNC2	0.18	first intron		cation transport (0006812)
6	chr14:31559311-31560651	ARHGAP5	53735	5'	Down	cell adhesion (0007155)
2	chr4:188034775-188036171	FAT	14639	5'	Down	cell adhesion (0007155)
3	chr3:175329355-175330050	NLGN1	0.78	internal intron	Down	cell adhesion (0007155)
6	chr4:30911633-30912963	PCDH7	90636	3'	Down	cell adhesion (0007155)
3	chr10:68646653-68648526	CTNNA3	0.27	internal intron	Up	cell adhesion (0007155)
5	chr15:71244564-71245730	NEO1	0.45	internal intron	Down	cell adhesion (0007155)
3	chr14:51611868-51614297	NID2	6352	5'	Up	cell adhesion (0007155)
4	chr12:27593036-27594797	PPFIBP1	0.14	first intron	Up	cell adhesion (0007155)
12	chr13:11357257-8-113573996	GAS6	0.59	internal intron		cell adhesion (0007155)
2	chr4:78622099-78622692	CCNG2	0.68	internal intron	Down	cell cycle (0007049)
6	chr4:78611167-78612302	CCNG2	0.64	internal intron	Down	cell cycle (0007049)
7	chr4:78620134-78621688	CCNG2	0.67	internal intron	Down	cell cycle (0007049)
5	chr6:36742675-36743642	CDKN1A	10838	5'	Up	Cell cycle (0007049)
13	chr6:36751902-36754502	CDKN1A	3	5'	Up	Cell cycle (0007049)
3	chr12:6486806-6487748	CNAP1	0.36	internal intron	Down	cell cycle (0007049)
4	chr5:57793144-57794331	SNK	1300	5'	Up	cell cycle (0007049)
2	chr2:205138572-205138959	ALS2CR19	97269	5'		cell cycle (0007049)
3	chr10:12764556-5-127646819	FANK1	0.62	first intron		Cell cycle (0007049)
2	chr11:11088640-3-110887294	BTG4	0.06	first intron		cell cycle arrest (0007050)
2	chr17:10049719-10050050	GAS7	7126	5'		cell cycle arrest (0007050)
18	chr17:65114449-65116192	MAP2K6	63382	3'		cell cycle arrest (0007050)
2	chr17:56277707-56278256	BCAS3	0.23	internal intron	Down	cell growth and/or maintenance (0008151)
5	chr1:177799988-177800842	IER5	8359	3'	Up	cell growth and/or maintenance (0008151)
8	chr1:177835206-177836591	IER5	43577	3'	Up	cell growth and/or maintenance (0008151)
2	chr20:56941551-56942313	GNAS	21911	3'		cell growth and/or maintenance (0008151)
2	chr8:20219256-20220193	LZTS1	62173	5'		cell growth and/or maintenance (0008151)
11	chr4:158049533-158051027	PDGFC	0.95	internal intron		cell growth and/or maintenance (0008151)
2	chr1:183440255-183440759	PTGS2	59056	5'		cell motility (0006928)

2	chr2:101234532-101235091	C2orf29	92804	5'	Down	cell proliferation (0008283)
2	chr16:11881529-11882129	GSPT1	0.81	internal intron	Down	cell proliferation (0008283)
7	chr4:124076519-124077909	FGF2	27558	5'	Up	cell proliferation (0008283)
6	chr12:87455649-87456827	KITLG	0.24	first intron	Up	cell proliferation (0008283)
10	chr2:70734862-70736664	TGFA	42249	5'	Up	cell proliferation (0008283)
3	chr5:141927955-141929665	FGF1	23640	3'		cell proliferation (0008283)
10	chr8:128875604-128877901	MYC	52751	3'		cell proliferation (0008283)
3	chrX:117655729-117656338	IL13RA1	0.18	first intron		cell surface receptor linked signal transduction (0007166)
2	chr11:11914513-5-119146475	PVRL1	40506	5'		cell-cell adhesion (0016337)
4	chr1:41616611-41618083	EDN2	0.22	internal intron		cell-cell signaling (0007267)
2	chr1:150311630-150312331	S100A6	7817	3'		cell-cell signaling (0007267)
5	chr6:112414685-112416243	WISP3	65911	5'		cell-cell signaling (0007267)
2	chr16:31203300-31204614	ITGAM	0.34	internal intron	Up	cell-matrix adhesion (0007160)
3	chr16:31192294-31193444	ITGAM	0.18	internal intron	Up	cell-matrix adhesion (0007160)
8	chr7:150822098-150823498	PRKAG2	0.59	internal intron	Up	cholesterol biosynthesis (0006695)
2	chr1:151688943-151690013	PMVK	20267	3'		cholesterol biosynthesis (0006695)
3	chr15:91146087-91147084	CHD2	97338	5'	Up	chromatin assembly or disassembly (0006333)
3	chr15:91293728-91295019	CHD2	265	3'	Up	chromatin assembly or disassembly (0006333)
2	chr7:18233502-18234210	HDAC9	74398	5'	Down	chromatin modification (0016568)
2	chr5:98202843-98203731	CHD1	15617	3'		chromosome organization and biogenesis (0007001)
7	chr2:37792420-37793729	CDC42EP3	0.63	first intron	Up	cytoskeletal regulatory protein binding (0005519)
5	chr7:133986832-133989025	CALD1	0.34	internal intron	Up	cytoskeleton (0005856)
3	chr8:142088029-142088637	PTK2	7515	5'	Down	cytoskeleton (0005856)
3	chr9:92641535-92642452	BICD2	34977	5'	Up	cytoskeleton (0005856)
3	chr10:13839987-13841321	FRMD4A	0.83	internal intron	Up	cytoskeleton (0005856)
12	chr14:66468089-66470675	GPHN	0.63	internal intron		cytoskeleton (0005856)
10	chr8:41794748-41796055	ANK1	20451	5'	Down	cytoskeleton organization and biogenesis (0007010)
4	chr1:42667726-42669044	FLJ20972	0.91	internal intron	Down	cytoskeleton organization and biogenesis (0007010)
2	chr7:2209382-2211990	CHST12	4473	5'	Down	dermatan sulfate biosynthesis (0030208)
2	chr11:12043756-12044667	DKK3	56551	5'	Down	development (0007275)
2	chr5:52822200-52822745	FST	4513	3'		development (0007275)
2	chr18:42701671-42702587	PIAS2	0.61	internal intron	Down	DNA binding (0003677)
2	chr5:70892094-70893102	BDP1	0.93	internal intron		DNA binding (0003677)
3	chr11:47192744-47194120	DDB2	342	5'	Up	DNA repair (0006281)

3	chr20:5048561-5049184	PCNA	200	5'	Up	DNA repair (0006281)
4	chr8:103317718-103318997	RRM2B	0.04	first intron	Up	DNA repair (0006281)
3	chr3:14194514-14195958	XPC	0	5'	Up	DNA repair (0006281)
2	chr11:107710961-107711632	ATM	0.79	internal intron		DNA repair (0006281)
2	chr6:57573724-57574265	PRIM2A	0.86	internal intron		DNA replication
7	chr8:124692871-124694497	FBXO32	70244	5'		electron transport (0006118)
3	chr17:78047056-78047847	NARF	7628	3'		electron transport (0006118)
4	chr9:116287827-116290043	ASTN2	0.98	internal intron	Up	energy pathways (0006091)
3	chr15:86964786-86966080	FLJ12484	4300	5'	Up	exonuclease activity (0004527)
7	chr16:20783998-20785636	MGC48972	0.8	first intron		exonuclease activity (0004527)
2	chrX:132645449-132646433	GPC3	0.44	internal intron	Down	extracellular matrix (0005578)
2	chr2:33150960-33151544	LTBP1	0.15	internal intron	Down	extracellular matrix (0005578)
3	chr4:73801010-73802395	ADAMTS3	1507	5'		extracellular matrix (0005578)
2	chr22:44185994-44186654	FBLN1	32642	5'		extracellular matrix (0005578)
3	chr13:109572385-109573846	COL4A1	26606	3'	Up	extracellular matrix structural constituent (0005201)
2	chr13:99500023-99501047	PCCA	38290	5'	Down	fatty acid metabolism (0006631)
8	chr6:151269949-151271345	FTHFSDC1	7620	5'	Down	folic acid and derivative biosynthesis (0009396)
5	chr6:81302271-81303465	AK126122	66756	3'		glycolysis (0006096)
6	chr1:9175241-9177064	ENO1	0.12	internal intron		glycolysis (0006096)
2	chr9:77910011-77911059	GNAQ	34086	5'	Down	G-protein coupled receptor protein signaling pathway (0007186)
4	chr17:15807293-15808314	ADORA2B	0.6	first intron	Up	G-protein coupled receptor protein signaling pathway (0007186)
2	chr10:115774642-115775397	ADRB1	18407	5'	Up	G-protein coupled receptor protein signaling pathway (0007186)
3	chr2:133133837-133134898	GPR39	0.55	first intron	Up	G-protein coupled receptor protein signaling pathway (0007186)
2	chr12:61899590-61900899	AVPR1A	68657	5'		G-protein coupled receptor protein signaling pathway (0007186)
3	chr11:87927175-87929030	GRM5	0.91	internal intron		G-protein coupled receptor protein signaling pathway (0007186)
2	chr11:4441824-4442933	OR52K3P	10338	5'		G-protein coupled receptor protein signaling pathway (0007186)
3	chr12:53852376-53853086	OR9K2 (refseq)	41549	3'		G-protein coupled receptor protein signaling pathway (0007186)
6	chr5:152171146-152172529	AK123816	0.45	Internal intron		G-protein coupled receptor protein signaling pathway [0007186]
6	chr2:48979734-48981172	LHCGR	85266	5'		G-protein signaling (0007187)
2	chr7:80897005-80897475	HGF	78620	3'		growth factor activity (0008083)
2	chr3:17610395-17612103	TBC1D5	0.25	internal intron		GTPase activator activity (0005096)
2	chr19:17313844-17314438	GTPBP3	21	3'		GTPase activity (0003924)
2	chr9:95778398-95779577	AL389952	0.4	internal intron		helicase activity (0004386)
3	chr19:43957618-43958245	LGALS7	1630	5'		heterophilic cell adhesion (0007157)

5	chr9:136719829-136721148	NOTCH1	3754	5'	Up	immune response (0006955)
3	chr2:204716756-204717594	ICOS	64954	3'		immune response (0006955)
4	chr6:31412635-31414025	M12678	15604	3'		immune response (0006955)
3	chr6:31465968-31467646	MICA	11703	5'		immune response (0006955)
3	chr6:31562637-31563757	MICB	10084	5'		immune response (0006955)
5	chr14:50136549-50137733	SPG3A	0.55	internal intron		immune response (0006955)
3	chr8:22981574-22983332	TNFRSF10B	100	5'	Up	induction of apoptosis (0006917)
2	chr22:17291232-17291990	PIG6	0.28	first intron	Up	induction of apoptosis by oxidative stress (0008631)
5	chr10:45253603-45254649	ALOX5	0.89	internal intron		inflammatory response (0006954)
2	chr8:57981054-57981801	FLJ20421	53117	3'		inositol or phosphatidylinositol phosphatase activity (0004437)
2	chr1:227069711-227070623	COG2	14349	5'		intracellular protein transport (0006886)
4	chr10:10655889-6-106560109	SORCS3	0.27	first intron		intracellular protein transport (0006886)
2	chr20:17778498-17779090	SNX5	91155	3'	Down	intracellular signaling cascade (0007242)
6	chr16:4130436-4131734	ADCY9	25412	5'		intracellular signaling cascade (0007242)
2	chr15:99019902-99020324	ASB7	13138	3'		intracellular signaling cascade (0007242)
2	chr7:29141547-29142007	CHN2	0.42	first intron		intracellular signaling cascade (0007242)
3	chr2:46310853-46311743	PRKCE	0.98	internal intron		intracellular signaling cascade (0007242)
3	chr8:82922424-82924235	SNX16	5440	5'		intracellular signaling cascade (0007242)
3	chr17:22682037-22683065	WSB1	17251	3'		intracellular signaling cascade (0007242)
3	chr20:47188589-47189717	STAU	0.65	internal intron	Down	intracellular transporter activity (0005478)
3	chr7:116864069-116865036	CFTR	0.8	internal intron		ion channel activity (0005216)
3	chr1:230150728-230151420	KCNK1	35972	3'		ion transport (0006811)
3	chr6:39358686-39359048	KCNK17	15710	3'		ion transport (0006811)
2	chr1:211562905-211564072	KCNK2	3518	5'		ion transport (0006811)
2	chr21:44842261-44843024	KRTAP10-7	1900	5'		keratin filament (0045095)
4	chr19:15617927-15618763	CYP4F3	0.24	internal intron	Up	leukotriene metabolism (0006691)
3	chr19:15863930-15865273	CYP4F2	0.23	internal intron		leukotriene metabolism (0006691)
2	chr2:178164773-178164967	AGPS	0.56	internal intron		lipid biosynthesis (0008610)
2	chr11:59096869-59097884	OSBP	562	3'	Down	lipid transport (0006869)
7	chr5:110857599-110859125	STARD4	508	3'	Down	lipid transport (0006869)
10	chr7:40529658-40531147	C7orf10	0.8	internal intron		metabolism (0008152)
4	chr14:23209479-23210108	DHRS2	24793	3'		metabolism (0008152)
3	chr15:47345137-47346358	GALK2	0.64	internal intron		metabolism (0008152)
3	chr9:96108151-96109324	HSD17B3	0.52	internal intron		metabolism (0008152)

3	chr2:118589205-118590770	INSIG2	5381	3'		metabolism (0008152)
2	chr16:87721146-87721433	LOC197322	0.54	internal intron		metabolism (0008152)
2	chr7:21706371-21708009	DNAH11	0.98	internal intron		microtubule motor activity (0003777)
3	chr3:36949079-36949987	MLH1	54245	5'	Down	mismatch repair (0006298)
2	chr2:47889038-47889689	MSH6	32282	5'	Down	mismatch repair (0006298)
5	chr1:198169831-198170441	PHLDA3	2868	5'		morphogenesis (0009653)
2	chr20:35050354-35051735	RBL1	8160	3'	Down	negative regulation of cell cycle (0045786)
2	chr22:22492800-22493381	SMARCB1	0.82	internal intron	Down	negative regulation of cell cycle (0045786)
4	chr13:50001247-50002206	DLEU1	1046	3'		negative regulation of cell cycle (0045786)
3	chr9:36064280-36065428	RECK	0.43	internal intron		negative regulation of cell cycle (0045786)
4	chr1:201225177-201226276	MDM4	0.04	first intron	Down	negative regulation of cell proliferation (0008285)
3	chr19:49965946-49967006	CBLC	5959	5'	Up	negative regulation of epidermal growth factor receptor activity (0007175)
4	chr1:117133659-117134375	PTGFRN	30355	5'		negative regulation of protein biosynthesis (0017148)
2	chr3:33484389-33484666	UBP1	27515	5'	Down	negative regulation of transcription (0016481)
3	chr12:98392890-98393747	EB-1	0.39	internal intron		neurogenesis (0007399)
6	chr5:37883475-37884595	GDNF	7936	5'		neurogenesis (0007399)
2	chr1:151599121-151599699	KCNN3	0.34	internal intron		neurogenesis (0007399)
2	chr6:36683759-36684225	SFRS3	4572	3'		nuclear mRNA splicing (0000398)
2	chr6:152635159-152635700	SYNE1	1000	5'		nuclear organization and biogenesis (0006997)
4	chr3:180269957-180271353	WIG1	0.02	first intron	Up	nucleic acid binding (0003676)
2	chr9:16627209-16627911	BNC2	0.29	internal intron		nucleic acid binding (0003676)
2	chr18:58390748-58391432	ZCCHC2	0.77	internal intron		nucleic acid binding (0003676)
4	chr6:26812561-26813388	ZNF322A	44819	5'		nucleic acid binding (0003676)
2	chr10:63766158-63766966	ZNF365	36990	5'		nucleic acid binding (0003676)
3	chr13:41274727-41275802	ABO11136	0.21	internal intron	Up	nucleotide binding (0000166)
5	chr12:77053081-77054242	NAV3	0.41	internal intron	Up	nucleotide binding (0000166)
2	chr12:12195584-121956603	ABCB9	0.61	internal intron		oligopeptide transport (0006857)
3	chr12:12193649-121938707	ABCB9	2045	3'		oligopeptide transport (0006857)
2	chr15:61507445-61508569	CA12	46347	5'		one-carbon compound metabolism (0006730)
2	chr2:42934765-42935017	HAAO	3399	5'		oxidoreductase activity (0016491)
2	chr19:46275427-46276304	CYP2A13	9912	5'		oxidoreductase activity (0016712)
3	chr7:98979755-98980587	CYP3A7	2385	5'		oxidoreductase activity (0016712)
3	chr7:151206867-151207691	GALNT11	0.48	first intron		polypeptide N-acetylgalactosaminyltransferase activity (0004653)

3	chr2:154826593-154828585	GALNT13	0.4	internal intron		polypeptide N-acetylgalactosaminyltransferase activity (0004653)
3	chr11:34663117-34664263	EHF	24005	3'		positive regulation of transcription (0045893)
3	chr18:22294391-22295582	KCTD1	0.96	internal intron		potassium ion transport (0006813)
2	chr8:95072056-95073257	PPM2C	66279	3'	Up	protein amino acid dephosphorylation (0006470)
3	chr8:134472408-134474124	SIAT4A	66202	3'		protein amino acid glycosylation (0006486)
5	chr2:88676930-88677568	EIF2AK3	8566	3'	Down	protein amino acid phosphorylation (0006468)
2	chr8:8302075-8302982	AK122582	25476	5'		protein amino acid phosphorylation (0006468)
8	chr8:8257062-8258330	AK122582	0.29	internal intron		protein amino acid phosphorylation (0006468)
3	chr4:114839516-114840722	CAMK2D	0.66	internal intron		protein amino acid phosphorylation (0006468)
2	chr1:203208778-203209454	DYRK3	0.84	internal intron		protein amino acid phosphorylation (0006468)
2	chr5:179616260-179617779	MAPK9	0.5	internal intron		protein amino acid phosphorylation (0006468)
3	chr1:46134298-46136015	MAST2	0.66	internal intron		protein amino acid phosphorylation (0006468)
6	chr12:826876-828190	PRKWNK1	0.6	internal intron		protein amino acid phosphorylation (0006468)
3	chr5:171467474-171468762	STK10	0.55	internal intron		protein amino acid phosphorylation (0006468)
8	chr4:78059015-78060785	ShrmL	894	3'	Up	protein binding (0005515)
2	chr10:25085197-25086813	ARHGAP21	32647	5'		protein binding (0005515)
2	chr16:21916662-21917972	MGC50721	0.12	first intron		protein binding (0005515)
9	chr15:61235856-61237660	RPS27L	83	5'	Up	protein biosynthesis (0006412)
4	chr19:47055922-47057051	RPS19	0.01	first intron		protein biosynthesis (0006412)
2	chr4:12405485-124056557	FLJ35630	31783	3'		protein folding (0006457)
3	chr15:81207295-81208829	AP3B2	31752	5'		protein transport (0015031)
8	chr4:40986864-40989274	APBB2	777	5'		protein transport (0015031)
2	chr1:232159128-232160368	CHS1	0.95	internal intron		protein transport (0015031)
2	chr3:123679743-123680672	KPNA1	0.4	internal intron		protein transporter activity (0008565)
2	chr4:7821881-7822865	SORCS2	0.61	internal intron		protein transporter activity (0008565)
3	chr10:11446559-114466275	VTI1A	0.92	internal intron		protein transporter activity (0008565)
3	chr2:212165550-212166599	ERBB4	0.92	internal intron	Down	protein tyrosine kinase [0004713]
9	chr10:12955111-129552684	PTPRE	43715	5'	Up	protein tyrosine phosphatase activity (0004725)
7	chr1:211000966-211002000	PTPN14	0.67	internal intron		protein tyrosine phosphatase activity (0004725)
2	chr11:5667802-5668510	TRIM22	0.01	first intron	Up	protein ubiquitination (0016567)
2	chr19:58490202-58490660	BIRC8	3511	5'		protein ubiquitination (0016567)
2	chr1:172954068-172954909	COP1	46297	5'		protein ubiquitination (0016567)
2	chr17:58162840-58164902	FLJ35757	0.7	internal intron		protein ubiquitination (0016567)

2	chr10:92988050-92989198	RNF159	10354	3'		protein ubiquitination (0016567)
2	chr10:101793183-101794775	CPN1	0.93	internal intron	Down	proteolysis and peptidolysis (0006508)
6	chr19:18335537-18337814	PGPEP1	2000	3'		proteolysis and peptidolysis (0006508)
2	chr3:114101822-114103174	CD200R1	21052	3'		receptor activity (0004872)
2	chr2:119547367-119548295	MARCO	78907	3'		receptor activity (0004872)
2	chr6:129509627-129510918	LAMA2	0.42	internal intron		regulation of cell adhesion (0030155)
4	chr2:105919948-105921131	NCK2	0.37	first intron	Down	regulation of epidermal growth factor receptor activity (0007176)
2	chrX:69113561-69114296	IGBP1	22042	5'		regulation of signal transduction (0009966)
3	chr8:67664389-67665687	MYBL1	13313	5'	Down	regulation of transcription (0006355)
2	chr3:20094710-20095306	PCAF	0.34	internal intron	Down	regulation of transcription (0006355)
2	chr7:137701994-137704423	TIF1	0.79	internal intron	Down	regulation of transcription (0006355)
6	chr1:209157999-209159107	ATF3	11399	5'	Up	regulation of transcription (0006355)
3	chr2:191301555-191303006	NAB1	46383	5'	Up	regulation of transcription (0006355)
3	chr9:124445453-124446468	NR6A1	0.66	internal intron	Up	regulation of transcription (0006355)
5	chr6:106652399-106653614	PRDM1	0.49	internal intron	Up	regulation of transcription (0006355)
2	chr15:78564721-78565132	ARNT2	0.42	internal intron		regulation of transcription (0006355)
2	chr21:29697827-29698558	BACH1	41741	3'		regulation of transcription (0006355)
2	chr22:48474602-48475733	BRD1	12059	3'		regulation of transcription (0006355)
2	chr16:3829126-3829829	CREBBP	0.27	internal intron		regulation of transcription (0006355)
2	chr2:100052349-100053864	LAF4	0.3	internal intron		regulation of transcription (0006355)
2	chr2:169134238-169135249	LASS6	3092	5'		regulation of transcription (0006355)
2	chr3:120994904-120996070	NR1I2	0.3	first intron		regulation of transcription (0006355)
6	chr1:196649507-196650735	NR5A2	77691	5'		regulation of transcription (0006355)
5	chr4:26084221-26085683	RBPSUH	0.77	internal intron		regulation of transcription (0006355)
3	chr9:3202391-3203845	RFX3	10803	3'		regulation of transcription (0006355)
3	chr6:85602044-85603706	TBX18	72154	5'		regulation of transcription (0006355)
2	chr10:114816891-114817646	TCF7L2	0.54	internal intron		regulation of transcription (0006355)
3	chr9:81435988-81437072	TLE1	0.92	internal intron		regulation of transcription (0006355)
4	chr13:47628530-47629500	VDRIP	61289	5'		regulation of transcription (0006355)
7	chr13:47538049-47539243	VDRIP	8849	3'		regulation of transcription (0006355)
3	chr8:145896298-145898453	ZNF251	18650	3'		regulation of transcription (0006355)
2	chr19:9011211-9011684	ZNF560	26870	3'		regulation of transcription (0006355)
2	chr1:31189338-31190186	PUM1	0.13	internal intron		regulation of translation (0006445)
3	chr7:86604121-86605580	TP53AP1	0.69	internal intron	Up	response to DNA damage stimulus (0006974)

4	chr5:41157842-41159114	C6	18981	3'		response to pathogenic bacteria (0009618)
8	chr21:33660665-33662530	IFNAR1	6672	3'		response to virus (0009615)
4	chr19:13624625-13626341	MGC10471	93411	5'		response to virus (0009615)
3	chr1:43286807-43288807	EBNA1BP2	10267	3'		ribosome biogenesis (0007046)
3	chr2:177859894-177861162	HNRPA3	41842	5'		RNA binding (0003723)
4	chr17:52956395-52957111	MSI2	0.63	internal intron		RNA binding (0003723)
3	chr17:34278957-34280068	LASP1	100	5'		SH3/SH2 adaptor protein activity (0005070)
2	chr1:237426033-237427272	RGS7	0.72	internal intron		signal transducer activity (0004871)
2	chr4:56317130-56317964	NMU	0.62	internal intron	Down	signal transduction (0007165)
2	chr7:79425001-79425827	GNAI1	0.19	internal intron	Up	signal transduction (0007165)
3	chr17:19830631-19831404	AKAP10	8910	5'		signal transduction (0007165)
2	chr5:141073540-141074067	ARAP3	31606	5'		signal transduction (0007165)
2	chr18:55007293-55007857	GRP	30522	5'		signal transduction (0007165)
5	chr11:60087306-60088475	MS4A12	55831	3'		signal transduction (0007165)
3	chrX:150382746-150383185	PASD1	19476	5'		signal transduction (0007165)
2	chr5:149124037-149125083	PPARGC1 B	0.29	first intron		signal transduction (0007165)
3	chr8:67198890-67199560	RNF29	2271	5'		signal transduction (0007165)
2	chr4:187363156-187363995	TLR3	1464	5'		signal transduction (0007165)
3	chr14:101726033-101727328	WDR20	0.67	first intron		signal transduction (0007165)
3	chr16:65518223-65519868	RRAD	1284	5'	Up	small GTPase mediated signal transduction (0007264)
3	chr9:127136622-127137764	GARNL3	0.23	internal intron		small GTPase regulatory/interacting protein activity (0005083)
6	chr9:4778277-4779366	RCL1	3570	5'		snRNP recycling (0000244)
6	chr14:20289485-20290907	FAM12A	3110	3'		sperm displacement (0007321)
4	chr10:17328119-17330136	VIM	8521	3'	Down	structural constituent of cytoskeleton (0005200)
2	chr18:3063649-3063837	MYOM1	0.95	internal intron		structural constituent of muscle (0008307)
3	chr12:50855976-50856990	LOC144501	0.65	internal intron	Up	structural molecule activity (0005198)
2	chr8:8520497-8521395	AK123547	75680	5'		structural molecule activity (0005198)
2	chr22:17937314-17938068	CLDN5	50837	5'		structural molecule activity (0005198)
2	chr17:23633681-23634161	FLJ40504	0.78	internal intron		structural molecule activity (0005198)
2	chr12:9955086-9956444	AY359126	0.59	internal intron		sugar binding (0005529)
2	chr8:42725978-42727080	CHRNA6	3065	3'		synaptic transmission (0007268)
6	chr2:1470401-1472500	TPO	0.57	internal intron	Up	thyroid hormone generation (0006590)
2	chr11:119400457-119401122	TRIM29	86082	3'		transcription factor activity (0003700)
2	chr16:14401388-14402374	PARN	34685	3'		transcriptional repressor activity (0016564)

5	chr4:10976711-10977770	HS3ST1	98918	3'		transferase activity (0016740)
2	chrX:131737795-131738539	HS6ST2	0.24	internal intron		transferase activity (0016740)
3	chr11:3593372-3594436	ART5	21875	3'		transferase activity (0016757)
3	chr12:15579421-15580381	PTPRO	10171	5'	Up	transmembrane receptor protein tyrosine phosphatase activity (0005001)
2	chr18:7723666-7724030	PTPRM	20940	5'	Down	transmembrane receptor protein tyrosine phosphatase activity (0005001)
3	chr18:7718071-7719064	PTPRM	25906	5'	Down	transmembrane receptor protein tyrosine phosphatase activity (0005001)
4	chr6:111534382-111535658	SLC16A10	0.14	first intron		transport (0006810)
3	chr14:36359059-36360419	SLC25A21	0.71	internal intron		transport (0006810)
7	chr9:84083088-84084177	SLC28A3	38468	3'		transport (0006810)
2	chr16:66557346-66558162	SLC12A4	0.08	first intron		transporter activity (0005215)
4	chr18:53798740-53799589	NEDD4L	63027	5'	Down	ubiquitin cycle (0006512)
5	chr2:238630189-238631620	NCE2	26079	5'		ubiquitin cycle (0006512)
4	chr2:61415888-61417156	USP34	0.68	internal intron	Down	ubiquitin-dependent protein catabolism (0006511)
4	chrX:40807995-40809364	USP9X	0.63	internal intron	Down	ubiquitin-dependent protein catabolism (0006511)
2	chr8:103508334-103509177	EDD	14663	5'		ubiquitin-dependent protein catabolism (0006511)
2	chr12:91680696-91681867	EEA1	0.93	internal intron	Down	vesicle fusion (0006906)
4	chr1:178222231-178223504	CACNA1E	0.12	internal intron		voltage-gated calcium channel activity (0005245)
2	chr10:79089142-79089778	KCNMA1	21845	5'	Up	voltage-gated potassium channel activity (0005249)
2	chr9:92830162-92831561	FGD3	0.22	internal intron		zinc ion binding (0008270)
8	chr15:78081823-78083548	BCL2A1	31125	5'	Down	
3	chr7:10895281-10896157	PHF14	0.84	internal intron	Down	
4	chr17:46385667-46387669	SPAG9	8612	3'	Down	
4	chr12:100751190-100752098	FLJ11259	21498	5'	Up	
5	chr15:47704355-47705679	MDS009	0.17	internal intron	Up	
3	chr18:41903484-41905359	PSTPIP2	0.01	first intron	Up	
2	chr7:1115351-1116123					4.7 kb from CpG
4	chr1:179461712-179462929	AF336876	19349	3'		
9	chr4:84448849-84450177	AF338197	12901	5'		
3	chr2:186257853-186258613	AF424542	0.31	Internal intron		
2	chr10:29420510-29422132	AK023966	22661	3'		
3	chr15:60341378-60342654	AK126077	472	3'		
3	chrX:47434292-47435337	AK126949	17217	5'		
3	chr2:51394754-51396351	AK127244	0.16	Internal intron		
4	chr6:72133713-72135052	AK127397	4624	3'		

5	chr4:174278188-174279299	AK128523	3778	3'
3	chr18:66260749-66262447	BC001368	5228	3'
4	chr12:64318800-64320324	BC001368	0.26	Internal intron
6	chr8:29683561-29684618	BC003524	22017	5'
3	chr3:195203204-195204978	BC027871	500	5'
8	chr7:123480048-123481129	BC031318	90255	3'
5	chr12:12590463-8-125906869	BC032874	17784	3'
3	chr15:27681234-27682857	BC034566	6314	3'
2	chr5:57166888-57167291	BC041836 (mRNA)	54623	3'
3	chr4:133089219-133089940	BC041865	0.33	first intron
2	chr6:147406014-147407278	BC044307 (mRNA)	0.44	internal intron
4	chr17:31381324-31382497	CCL23	12206	5'
9	chr12:15980549-15982680	CGI-26	0.2	first intron
3	chr11:10022164-10023259	CMT4B2	0.48	internal intron
3	chr2:62098415-62099097	COMMD1	0.23	first intron
3	chr8:133106297-133107034	D63477	11346	3'
6	chr4:76168203-76169620	DKFZP564O0823	45898	5'
4	chr3:123595148-123596318	E2IG5	0.36	first intron
3	chr17:64204479-64206055	FAM20A	95789	5'
9	chr8:124925928-124926890	FLJ23790	29070	3'
9	chr8:126270209-126271296	FLJ32440	0.35	internal intron
3	chr11:11448765-5-114488749	IGSF4	56416	3'
9	chr1:231060802-231062153	IRF2BP2	8796	5'
8	chr6:110309889-110311139	KIAA0274	56565	3'
3	chr6:159023724-159025441	KIAA1423	894	3'
4	chr16:86357183-86360298	KLHDC4	100	5'
4	chr1:75549433-75550502	MGC34032	0.58	internal intron
3	chr14:37019275-37020476	MIPOL1	0.8	internal intron
5	chr2:170903601-170904621	MYO3B	0.09	internal intron
3	chr3:108563329-108564453	NYD-SP17	14424	5'
3	chr14:70401330-70402407	PCNX	41467	5'
5	chr19:46725987-46726858	R29124_1	47512	5'
6	chr14:71310476-71311948	SIPA1L1	34605	3'
4	chr13:83413163-83414947	SLITRK1	58634	5'

3	chr9:128624334-128625054	TBC1D13	4127	5'
3	chr2:99115322-99116193	TSGA10	0.7	internal intron
3	chr4:177311146-177312109	WDR17	50171	5'
2	chr11:28499144-28499729			
2	chr11:97119470-97119979			
2	chr12:68771023-68772035			
2	chr13:103455021-103455830			
2	chr13:71569716-71570279			
2	chr18:21366958-21367718			
2	chr2:103712507-103713424			
2	chr2:164634644-164635869			
2	chr2:21965746-21967214			
2	chr20:7700458-7701452			
2	chr21:19888722-19890405			
2	chr3:103894239-103894961			
2	chr4:129964821-129965240			
2	chr4:169192724-169193699			
2	chr5:56399106-56400000			
2	chr5:98813448-98813761			
2	chr7:108383554-108385256			
2	chr7:120744726-120745778			
2	chr7:52025652-52026598			
2	chr8:20329730-20330217			
2	chr8:5165758-5166971			
2	chr9:73528539-73529367			
2	chrX:113120101-113120884			
3	chr1:69232796-69233680			
3	chr10:16246681-16248118			
3	chr10:67044280-67045604			
3	chr11:48766160-48772288			
3	chr11:97066965-97068506			
3	chr12:22875285-22876047			
3	chr12:87205972-87207147			
3	chr12:89129714-89131284			

3	chr12:90798732-90799810
3	chr13:26386739-26387340
3	chr14:27936394-27937799
3	chr14:33836568-33838083
3	chr16:44947116-44949762
3	chr16:53238996-53240473
3	chr17:48158254-48160111
3	chr17:52149611-52150336
3	chr2:136966740-136967531
3	chr2:226753583-226754432
3	chr21:15458636-15459767
3	chr4:108657743-108658946
3	chr4:134668632-134669736
3	chr4:163616396-163617779
3	chr5:123605810-123606498
3	chr5:174563720-174564238
3	chr5:81865175-81867348
3	chr9:78870565-78872193
4	chr1:184108916-184109905
4	chr11:70167437-70168383
4	chr12:77641940-77643468
4	chr17:57576098-57577369
4	chr2:169265298-169266349
4	chr4:117247922-117249179
4	chr4:188216741-188217836
4	chr5:173688070-173689158
4	chr5:36477866-36479004
4	chr5:67386-68792
4	chr7:120739160-120740669
5	chr1:215162337-215163779
5	chr16:25472258-25473207
5	chr19:32423843-32425732
5	chr3:154814334-154815818
5	chr7:90879021-90880289

6	chr7:152123892- 152125163
7	chr19:32428502- 32432178
