Prediction of p53 target genes based on integrative analysis of chromatin-immunoprecipitated and sequenced tags, by using Galaxy, a web-based interactive platform for large-scale genome analysis

Hiroaki Mita[1)], Yasushi Sasaki[1)], Hiromu Suzuki[2)], Masashi Idogawa[1)], Lisa Kashima[1)], Naoki Anbo[1)], Hirofumi Akashi[3)], Haruyuki Tatsumi[4)], Minoru Toyota[1,5)] and Takashi Tokino[1)]

[1)]*Department of Molecular Biology, Cancer Research Institute,*
[2)]*First Department of Internal Medicine,*
[3)]*Scholarly Communication Center,*
[4)]*First Department of Anatomy, Sapporo Medical University, Sapporo 060-8556, Japan.*
[5)]*Present Address: Department of Biochemistry, Sapporo Medical University, Sapporo 060-8556, Japan.*

ABSTRACT

Chromatin immunoprecipitation (ChIP) followed by sequencing of immunoprecipitated DNA fragments is the high-throughput method for identifying transcription factor binding sites. In one such method, ChIP-PET, paired-end ditags (PETs) derived from both ends of the immunoprecipitated DNA fragments are sequenced and mapped to the genome. We report here the prediction of p53 target genes by meta-analyzing tags of p53 ChIP-PET and by combining with other genomic annotations, using Galaxy, a web-based platform for large-scale genome analysis. We found 327 of p53 binding sites on the genome of 5-fluorouracil (5-FU)-treated HCT116 colon cancer cells by searching the total 65,509 PETs for PET clusters. The search for p53 target gene, which focused on PET clusters with computationally-predicted p53 binding motif, identified 20 of putative p53 target genes as well as 11 of known p53 targets. Another search for p53 target genes, which focused on PET clusters located within 50-kb flanking regions of transcription start sites of genes, identified 278 of Refseq genes, 79 of non-coding RNAs and 5 of microRNAs as p53 targets which included lots of known validated targets. Our results indicate that sequencing-based ChIP analysis combined with the existing genome annotation is effective method to predict p53 binding loci and target genes, and also show that the Galaxy platform is well-suited for multiple-type analyses and visualization of ChIP data, leading to functional annotation of transcription factor binding sites.

Key words : Tags, Transcription factor binding site, Transcription start site, Genome database, Genome browser

---

CORRESPONDENCE TO :
Takashi Tokino, Professor
Department of Molecular Biology, Cancer Research Institute, Sapporo Medical University
South 1, West 17, Chuo-ku, Sapporo 060-8556, Japan.
E-mail : tokino@sapmed.ac.jp
TEL : 011-611-2111 (ext 2386)
FAX : 011-618-3313

INTRODUCTION

The p53 gene, encoding for sequence-specific nuclear transcription factors, is involved in the maintenance of genome integrity by transactivating genes controlling cell cycle (p21/CDKN2A, GADD45) and apoptosis (Bax, Fas, Puma and Noxa) in response to cellular stress signals[1, 2]. Many p53 target genes are currently known, e.g. identified with microarray expression profiling[3], and at the moment it is intensively studied how p53 determines which target genes to activate or repress in a certain stress response[1, 4]. The microRNAs miR-34a and miR-34b/c were also identified as direct, conserved p53 target genes that presumably mediate induction of apoptosis, cell cycle arrest, and senescence by p53[5-7]. Since microRNAs may regulate the levels of hundreds of different proteins[8], these findings add a new, challenging layer of complexity to the p53 network. Notably, cancer cells can escape the tumour suppression function of p53 through missense mutation of the p53 gene or deregulation of p53 activity[9]. Since p53 is the most frequently mutated tumor suppressor gene in malignant tumors[10, 11], identification of the transcriptional targets of p53 is a key to understanding functions of p53 and its signaling pathways in tumorigenesis, and to exploiting their potential molecular targets for cancer chemotherapeutic drugs.

The tetrameric p53 protein binds to two repeats of a consensus DNA sequence RRRCWWGYYY separated by a spacer of 0-13 bp, in which R = purine, W = A or T and Y = pyrimidine[12]. This motif is found in many identified p53 binding sites within a few thousand base pairs of the transcriptional start site (TSS) of p53 target genes, and the motif, which binds p53 that can regulate the transcription of the target genes, is called p53 response element (RE). The progression of the human genome project set a trend of computational approach to predict p53 binding sites, and several algorithms have been devised as follows: position-specific score matrix (PSSM)[13] which attempts to estimate the binding affinity of a putative site; profile hidden Markov models (PHMMs)[14] which has been trained on the existing data set of functional p53 REs, and can be used to score putative p53-binding sites; filtering by measurement of conservation of p53 REs among different species based on comparative genomics[15]. However, *in-silico* predictions do not necessarily reflect the actual target sites bound by p53 *in vivo*.

Chromatin immunoprecipitation (ChIP) has been widely used to map the localization of transcription factors on a specific gene locus. The combination of ChIP assays with DNA microarray (ChIP on chip) has in recent years enabled the profiling of occupancy sites of transcription factors including NF-κB, myc and p63[16-18]. p53 ChIP-on-chip data derived from ENCODE regions[19], from promoter regions[20-22], or from regions covered by genome-wide tilling array[23] suggested that there are between 300 and 3000 binding sites for p53 in the human genome. More recently, Wei *et al*. has developed new methods using sequencing instead of microarrays, termed ChIP-PET (ChIP and Paired-End diTag sequencing)[24]. In ChIP-PET analysis, chromatin immunoprecipitation with p53-specific antibodies is carried out to collect all of the tight binding sites for p53 in the genome, subsequently, paired-end ditags (PETs) derived from both 18-bp ends of the immunoprecipitated genomic DNA fragments are cloned, sequenced and then mapped to the genome. The ChIP-PET analysis of 5-fluorouracil (5-FU)-treated human colon cancer cell line HCT116, by sequencing approximately 66,000 PETs, identified 542 of p53 binding loci throughout the genome[24].

The combination of ChIP and fast-maturing next-generation sequencing technology has brought much excitement in the field of functional genomics. In the newer method, ChIP-sequencing (ChIP-seq)[25-27], millions of immunoprecipitated DNA fragments are directly sequenced at one end for ~30 bp, and the short sequence reads are then mapped to the reference genome. Comparing with ChIP on chip whose usability for large mammalian genomes

is limited by serious cross-hybridization, sequencing-based ChIP analyses including ChIP-PET offer not only direct whole-genome coverage but also high signal-to-noise ratio and sensitivity that increase with sequencing depth. On the other hand, these high-throughput sequencing analyses produce huge amount of tag data, and the data analysis requires researchers to have substantial programming experience and data management skills. In ChIP-PET data analysis, tag clusters that represent *in vivo* location of transcription factor binding sites, have to be identified after the calculation of tag count and tag positional distribution throughout the genome.

Galaxy, a web-based interactive platform for large-scale genome analysis, combines the power of existing genome annotation databases with a simple web portal to enable users to search remote resources, combine data from independent queries, and visualize the results[28]. To allow experimental biologists with no programming experience to easily and efficiently manipulate genomic data, Galaxy provides variety of integrated *Tools*, e.g., a tool to extract genomic data from popular sources of data like the UCSC Table Browser and a tool to search overlapping regions between two sets of genomic intervals. Subsequently, as shown in the Galaxy tutorial (http://screencast.g2.bx.psu.edu/galaxy/promoters_SNPs/), relatively complex analysis such as genome-wide search of promoters which have SNPs is easily implemented. Furthermore, every step of user's analyses is recorded in the Galaxy's history system, and those workflows are able to be shared with others.

We report here the prediction of p53 target genes by meta-analyzing raw tag data of p53 ChIP-PET and by combining with other genomic annotations, using the Galaxy platform. Our results indicate that sequencing-based ChIP analysis combined with the existing genome annotation is effective method to predict p53 binding loci and the candidate target genes, and also show that Galaxy platform provides the sophisticated methods of analysis and visualization of ChIP data.

MATERIALS AND METHODS

p53 related genomic data and databases

p53-related genomic data was retrieved from University of California Santa Cruz (UCSC) Genome Browser website (http://genome.ucsc.edu/, Mar. 2006 freeze, hg18), including: Gene Identification Signature (GIS) determined by chromatin immunoprecipitation (ChIP) and Paired-End diTag (PET) sequencing (GIS ChIP-PET)[24], which shows the starts and ends of genomic DNA fragments bound to p53 protein; Conserved Transcription Factor Binding Sites (TFBS) computed with the TRANSFAC[29] Matrix Database v7.0; the NCBI RNA reference sequences collection (RefSeq genes); UCSC genes[30]; microRNAs from miRBase[31]; cytosine-phosphate-guanine (CpG) islands.

GIS ChIP-PET data contains position of each PET on the genome, which was obtained by the previous p53-ChIP-sequencing analysis of HCT116 colon cancer cell line treated by 5-fluorouracil (5-FU) for 6 h [24], as Browser Extensible Data (BED)[32]-formatted data.

The TRANFAC database contains the location and score of binding sites for various transcription factors including p53, which was predicted by computational analysis of sequence motif and conservation in the human/mouse/rat alignment. The whole-genome localization of p53 transcription-factor binding sites (p53 TFBS), was extracted from the TRANFAC database using UCSC Table browser.

In the UCSC genes collection, transcripts are categorized in four group, coding, noncoding, antisense, or nearCoding as follows: a coding transcript is one where the evidence is relatively good that it produces a protein; the nearCoding transcripts overlap coding transcripts by at least 20 bases on the same strand, but themselves do not seem to produce protein products because they are splicing variants with introns after the stop codon, that therefore undergo nonsense mediated decay[33]; antisense

transcripts overlap coding transcripts by at least 20 bases on the opposite strand; the other transcripts, which are neither coding, nor overlapping coding, are categorized as noncoding (http://genome.ucsc.edu/cgi-bin/hgGene?hgg_do_txInfoDescription=1). In this study, we termed the three types of transcripts other than the coding transcripts in the UCSC data set, "non-coding" RNA. Genomic position of transcription start site (TSS) of each gene was extracted from Refseq or UCSC genes data sets.

From the miRBase database, we extracted microRNAs with "hsa-" names which signify the human microRNAs. Each entry in the miRBase Sequence database represents a predicted hairpin portion of a miRNA transcript, which are not strictly precursor miRNAs (pre-miRNAs), but include the pre-miRNA and some flanking sequence from the presumed primary transcript.

Gene expression data of HCT116 colon cancer cells

Gene expression data of HCT116 cells treated by hydroxyurea (GSM71424 and GSM 71436 from series GSE3176)[34] was retrieved from Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo/). HCT116 expressing wild-type p53 (WT) and its derived isogenic p53-/- (KO) cells were treated by 1.5 mM hydroxyurea (HU) for 24 h, and gene expression of each sample was analyzed by using NCI/ATC Hs-OperonV2 microarray (GPL1528) which contained 21,329 of 70-mer oligonucleotide probes[34]. Calibrated log2 expression ratio (treated / untreated) of each gene was coordinated to genomic position of each transcript, and recorded as score for each genomic region in Wiggle (WIG)[32]-formatted file by using a Perl (Active Perl 5.8.8) script. The data of GIS PET of polyA+ RNA (GIS PET-RNA)[35] for HCT116 cells treated by 5-FU for 6 h, which shows the starts and ends of full-length mRNA transcripts, was retrieved from the UCSC website.

Gene ontology data

Gene ontology (GO) terms for predicted p53 target genes were extracted by using High-Throughput GoMiner (http://discover.nci.nih.gov/gominer/) tool and AmiGO (http://amigo.geneontology.org/) database.

Galaxy

Genomic data was processed by Galaxy[28] (http://main.g2.bx.psu.edu/). Galaxy tools used in this study, whose names are shown in italic hereafter, are shown as follows: *Get Data (UCSC Main table browser)* was used to retrieve data from GIS ChIP-PET, TFBS (TRANSFAC), CpG island, Refseq genes, UCSC genes, microRNAs and GIS PET-RNA database and transfer to the Galaxy work space; *Operate on Genomic Intervals (Cluster the intervals of a query)* was used to extract PET clusters which consist of more than three PETs allowing 1 bp intervals between two PETs (termed PET3 clusters). *Intersect the intervals of two queries* was used to extract PET3 clusters with p53TFBS and/or CpG island, and was used to find Refseq genes and microRNAs which intersected with or were flanked by a PET cluster. *Get flanks* was used to obtain flanking genomic regions for PET cluster and TSS of each gene at defined distance); *Graph/Display Data (Build custom track for UCSC genome browser)* was used to display maps of PET clusters, p53TFBS, and expression graph of each gene on Custom Tracks of the UCSC genome browser in conjunction with existing browser data; *Workflow* was used to edit and manage workflows for repeating *Intersect* analyses by changing parameters. Self-explanatory description for tool usage is also shown on each tool 's web page.

RESULTS

p53 binding sites expected by search of PET3 clusters with p53 binding motif

To identify novel p53 target genes within the human genome, we predicted p53 binding sites by analyzing the existing data of chromatin immunoprecipitation (ChIP) and Paired-End diTag (PET) sequencing (GIS ChIP-PET)[24] of human colon cancer cell line HCT116 treated by

5-FU for 6 h, by using the Galaxy platform (Fig. 1). It has been shown that the PET clusters, which consist of three or more overlapping PETs, were highly specific for p53 ChIP enrichment according to a Monte Carlo simulation[24]. Therefore, we searched PET clusters which consists of three or more overlapping or adjacent clusters within 1-bp distance (termed PET3

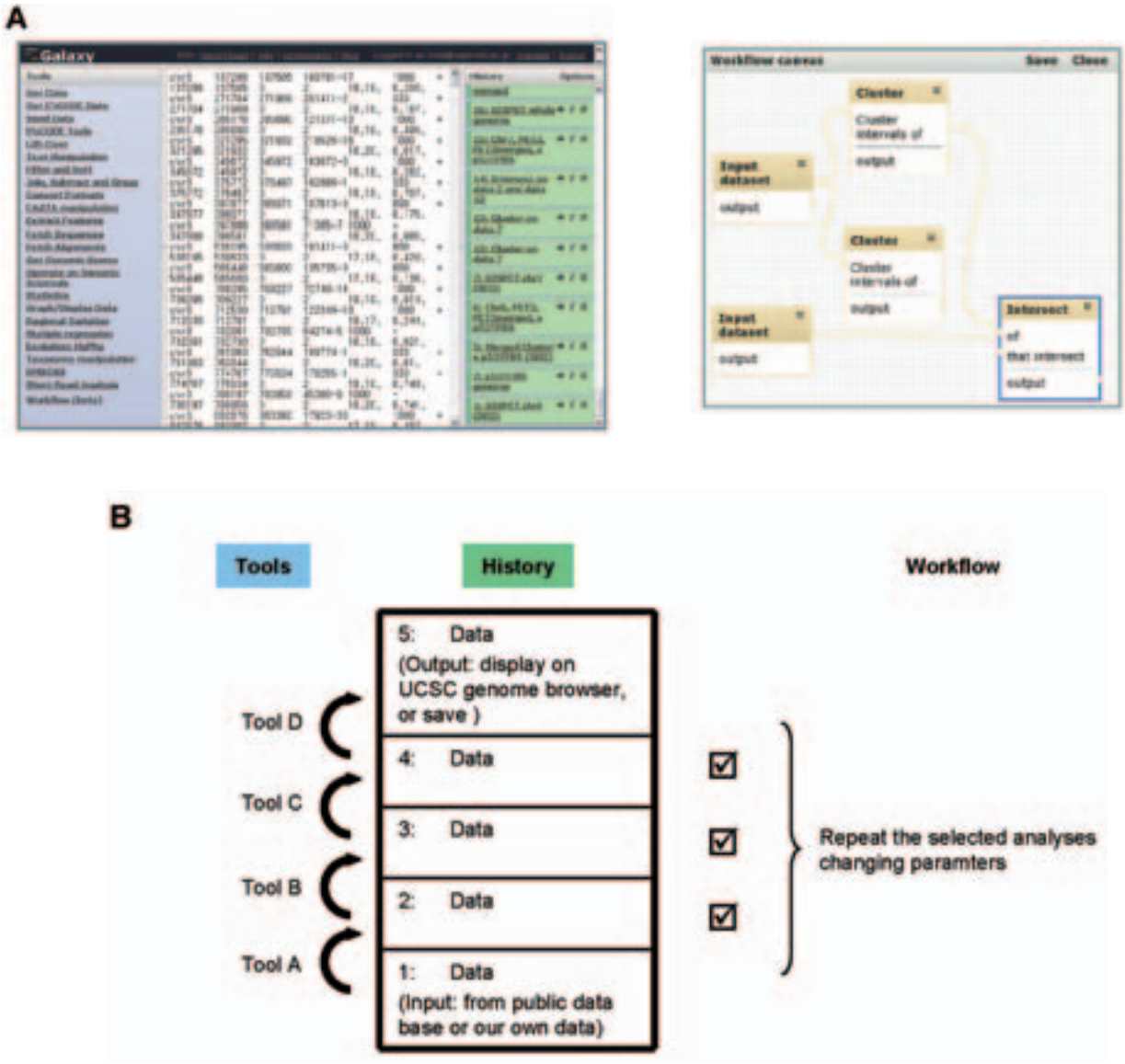

Fig. 1 Overview of Galaxy. **A**. The snap shots of the Galaxy main page (left) and the page of the workflow canvas (right). The Galaxy main page consists of 3 frames including the "Tools" menu, the "History" that stores the queries from each user, and the center page for data viewing. In the workflow canvas, selected items from the history system is connected according to the order of analysis, and shown as an editable flowchart. Some of parameters for Tools can be changed from the workflow canvas. **B**. The schematic of analysis using Galaxy. Each user inputs the initial genomic data for analysis into the history system, by retrieving from any public genome databases or by uploading their own data (History 1). The initial data is analyzed by Tool A and the result is stored in History 2. After a series of analyses, each user can visualize the result on the UCSC genome browser, download the result (History 5), or save selected histories as a workflow and repeat the analysis changing parameters.

clusters) in the human genome for p53 binding sites, by using the Galaxy tool *Cluster* (Fig. 2A, C). In the preliminary analysis of total 4,133 PETs on chromosome 6 (Fig. 3A, track "GIS-ChIP-PET"), 19 of PET3 clusters were detected (Fig. 3A, track "PET3_Cluster"). To predict the p53 binding sites with high confidence, we used the Galaxy tool *Intersect* (Fig. 2A,C) to search

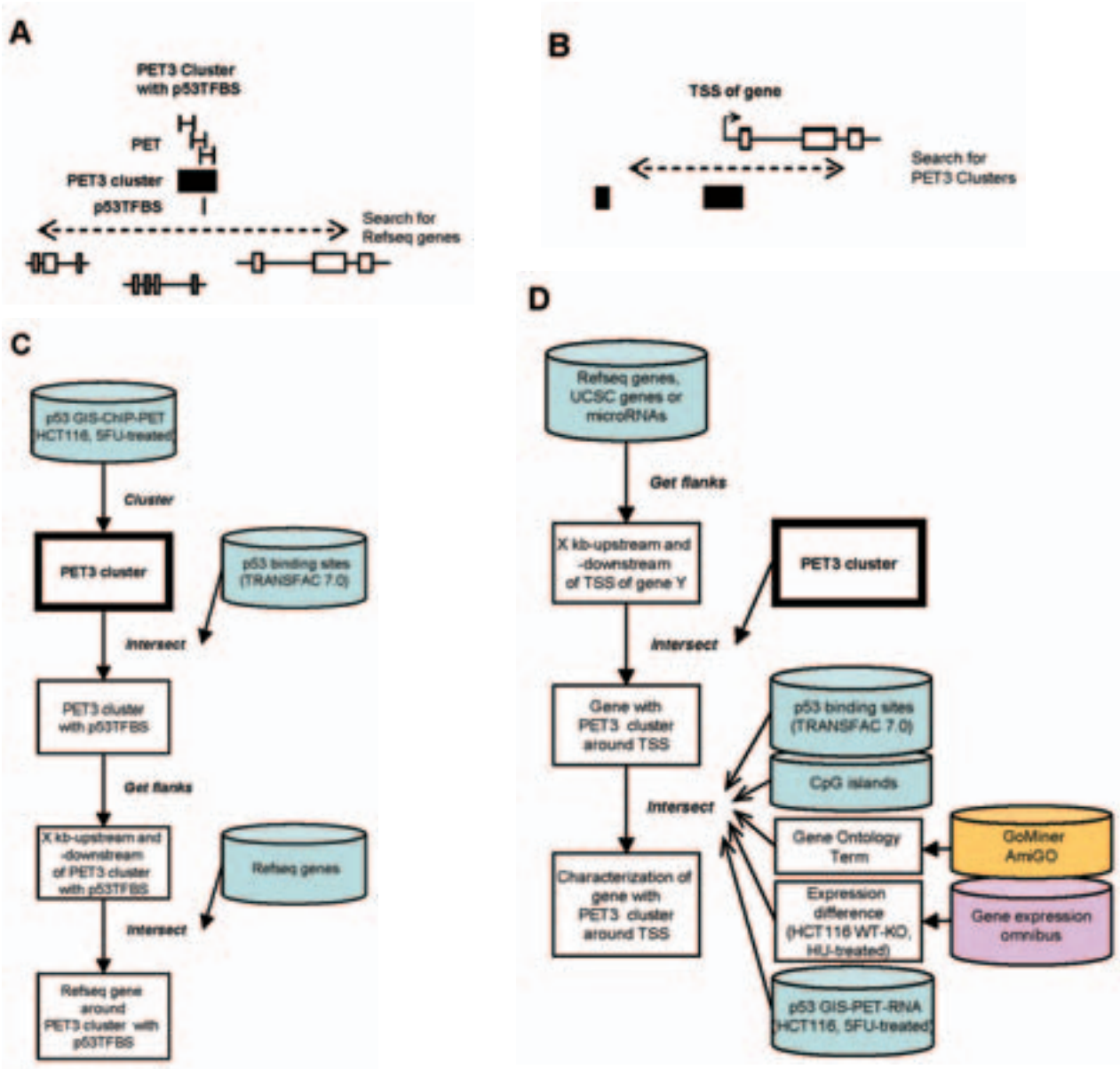

Fig. 2 Schematic strategies used to predict p53 target genes (**A**, **B**) and corresponding workflows (**C**,**D**). **A**. p53 target gene search focused on PET3 clusters with p53TFBS. PET3 cluster (black bar) is the genomic region defined by three or more overlapping or adjacent PETs (vertical lines at both ends, sequenced tags; connecting line, unsequenced interval). The flanking regions (dotted line) of PET3 clusters which have the p53TFBS (vertical line) was searched for putative p53 target genes (open boxes, exon; connecting lines, introns). **B**. p53 target gene search focused on flanking regions of TSS. The flanking regions of TSS of gene were searched for PET3 clusters. In the workflows (**C** for **A**, **D** for B), databases are shown in columns, and those available from the UCSC genome browser web site are light-blue-colored. The name of the Galaxy tool used in each step is shown in bold italic.

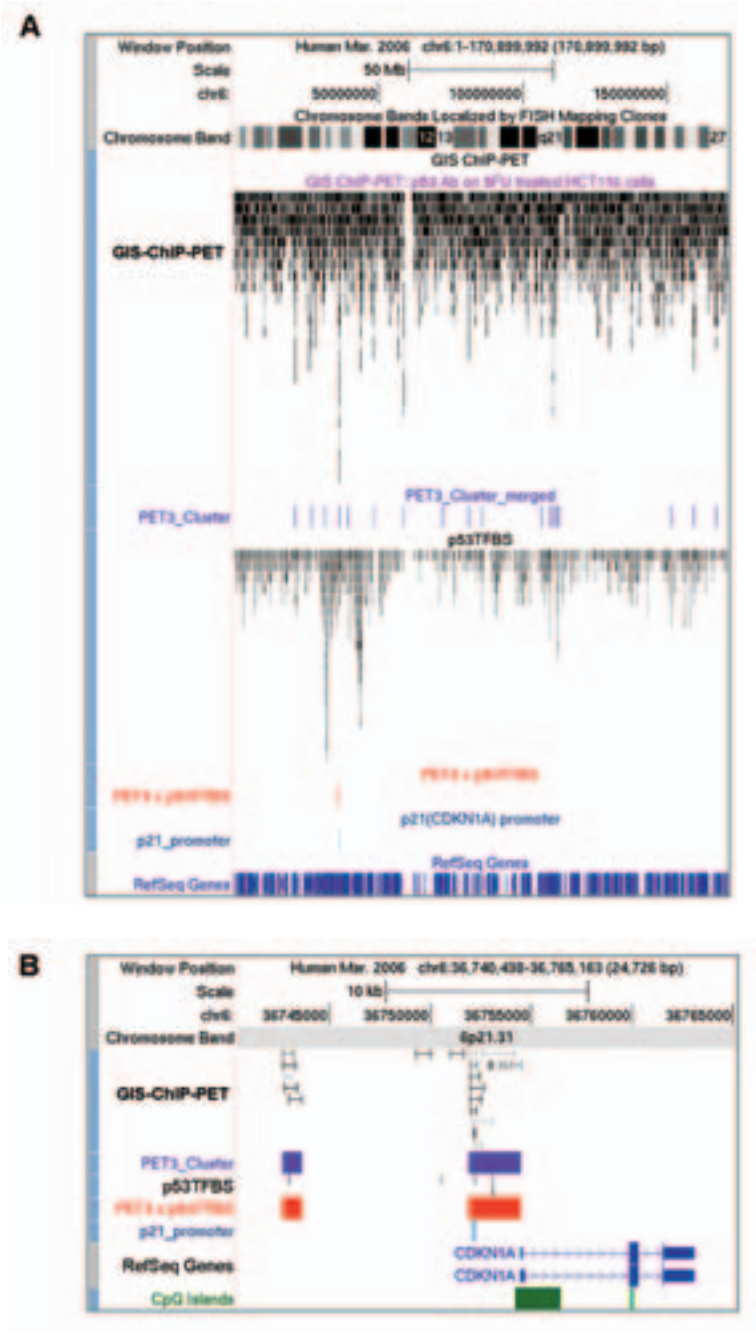

Fig. 3 p53 binding sites expected by search of PET3 clusters with p53 binding motif (p53TFBS). Map images are generated by the UCSC Genome Browser. PET3 clusters and those with p53TFBS were determined by using the Galaxy Tools *Cluster* and *Intersect*, respectively. Representative results of whole chromosome 6 (**A**), and the genomic region around p21/CDKN1A on chromosome 6 (**B**) are shown. The position of each PET is shown in the track labeled "GIS-ChIP-PET" as a vertical line (**A**), or horizontal line with vertical lines at both ends in the zoomed-in view (**B**). PET3 clusters (purple-colored) are shown in the "PET3_Cluster" track. The positions of computationally-predicted p53 binding sites are shown in the "p53TFBS" track. PET3 clusters with p53TFBS are shown in the "PET3 x p53TFBS" track, colored-red. The known p53 response element for p21/CDKN1A is shown in the "p21_promoter" track. Refseq genes are shown in dense mode (**A**), or labeled and have exons as boxes and introns as lines with arrowheads pointing in the direction of transcription (**B**).

for PET3 clusters that contain p53 binding sites (p53TFBS) which are predicted by computational motif search. We found that 2 out of 19 PET3 clusters on chromosome 6 had p53TFBS (Fig. 3A,B, track "PET3 x p53TFBS"). Importantly, both PET3 clusters were located upstream of p21/CDKN1A loci, which is one of the best characterized p53-target gene (Fig. 3B). Furthermore, the closest PET3 cluster with p53 TFBS was exactly overlapped with the known promoter region of p21 [36]. These data suggest that p53 binding sites are predicted with high confidence, by search for PET3 clusters with p53TFBS.

Table 1. PET3 clusters with p53TFBS

Chromosome	Start (bp)	End (bp)
chr1	9,163,561	9,165,385
chr1	117,223,139	117,223,856
chr1	179,334,953	179,335,808
chr1	179,370,171	179,371,557
chr2	70,676,714	70,678,517
chr3	180,269,948	180,271,345
chr3	195,203,195	195,204,970
chr4	157,911,377	157,912,872
chr4	188,078,585	188,079,681
chr5	57,793,143	57,794,331
chr5	118,687,055	118,688,091
chr5	173,688,069	173,689,158
chr6	36,742,674	36,743,642
chr6	36,751,901	36,754,502
chr7	123,673,332	123,674,414
chr8	128,875,603	128,877,901
chr8	143,893,707	143,894,655
chr9	117,745,208	117,746,886
chr9	138,563,812	138,565,132
chr10	67,044,279	67,045,604
chr11	34,663,116	34,664,263
chr12	19,475,469	19,477,289
chr12	50,855,975	50,856,990
chr13	109,572,384	109,573,846
chr13	113,572,577	113,573,996
chr14	36,359,058	36,360,419
chr15	61,235,855	61,237,660
chr15	86,964,785	86,966,080
chr18	22,294,390	22,295,582
chr19	47,055,921	47,057,051
chr19	54,149,415	54,150,580

p53 target genes predicted by search focused on PET3 clusters with p53TFBS

The whole-genome *Cluster* analysis of total 65,509 PETs identified 327 PET3 clusters on the genome, and 31 (9.5%) of those have the internal p53TFBS (Table 1). To search for putative p53 target genes, we examined total 27,090 of Refseq genes if they overlap with the region of PET3 clusters with p53TFBS or not, by using the Galaxy tool *Intersect*. We found that 12 genes overlapped with PET3 clusters with p53 TFBS (Table 2), of which 6 genes including AEN, BAX, KRT80, p21/CDKN1A, RPS27L and ZMAT3 were previously reported p53 targets[24]. To obtain positive evidence of the target genes, we retrieved microarray gene expression data of HCT116 expressing wild-type p53 (WT) and

Table 2. Refseq genes overlapping with PET3 clusters with p53TFBS, or located on flanking regions of PET3 clusters with p53TFBS

Position of Genes	Genes*	Number of genes
overlapping with PET3 clusters with p53TFBS	*AEN, BAX, CDKN1A,* GAS6, KCTD1, *KRT80,* PDGFC, RPS19, *RPS27L,* SLC25A21, TNFAIP8, *ZMAT3*	12
within ± 5 kb around PET3 clusters with p53TFBS	*NOTCH1, PLK2*	2
within ± 15 kb around PET3 clusters with p53TFBS	AEBP2, DHDH, DMRTC2, FTL, GYS1, *IER5,* LACTB	7
within ± 30 kb around PET3 clusters with p53TFBS	ARHGEF1, C5orf29, CD79A, *COL4A1,* EHF, FAM70B, *GML,* ISG20, LY6D, LYPD4	10
	Total	31

*Genes redundant in longer flanking regions are omitted. Previously known p53 targets are in italic.

p53-knocked-out (KO) cells treated by DNA-damaging reagent hydroxyurea (HU) from the GEO database[34], and visualized each expression value (log2 intensity ratio of treated / untreated) as a graph together with corresponding gene map on the UCSC genome browser. For example, RPS27L[37] and AEN[38] (Fig. 4A,B) showed increased expression in HCT116 WT cells but not in KO cells, similar to other known target genes BAX and ZMAT3 (data not shown). These data suggest that p53 target genes are accurately detected by the combined search for the position of PET3 clusters with p53TFBS and expression change of each gene specific to HCT116 WT cells under DNA-damaging stress.

We next searched Refseq genes located on

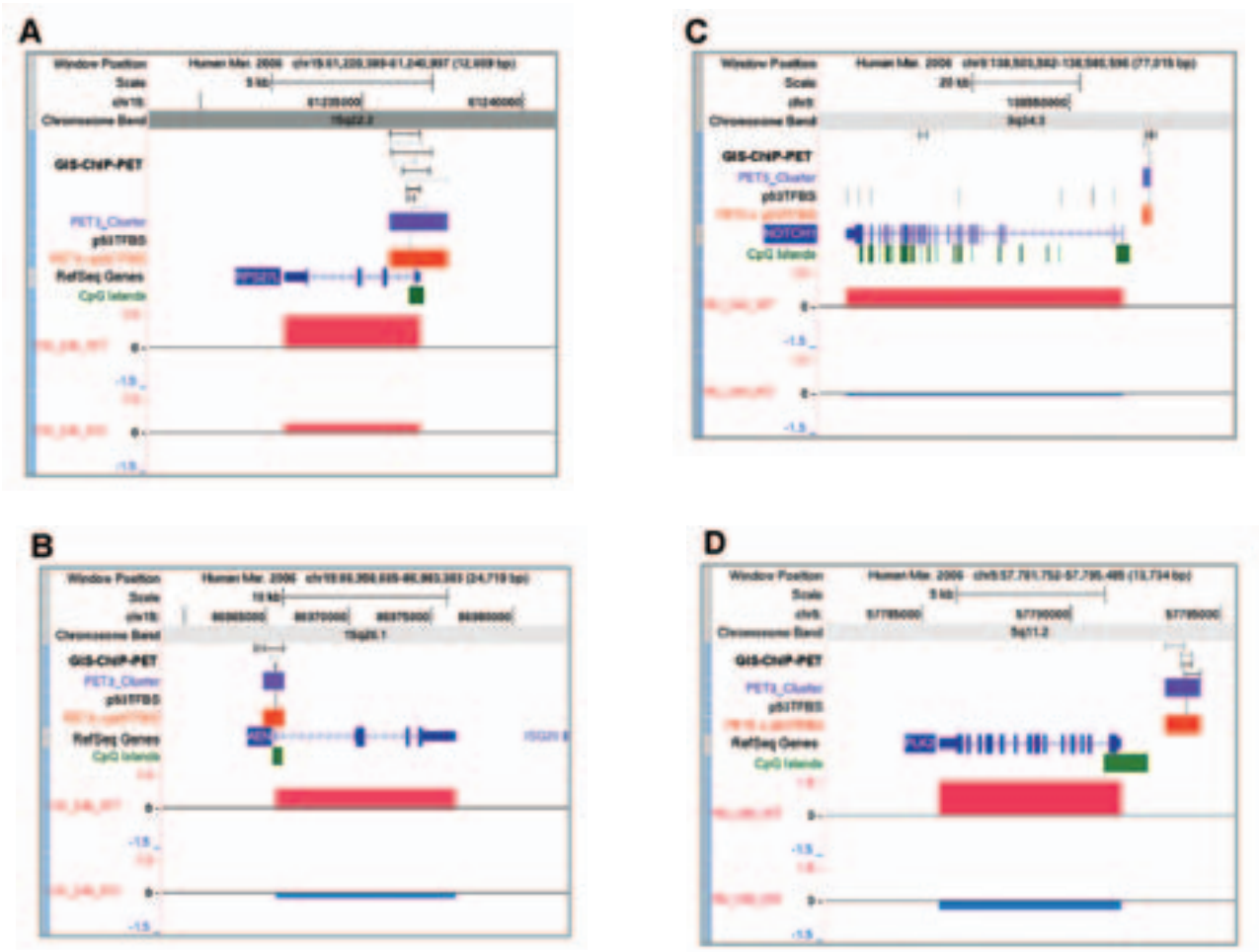

Fig. 4 p53 target genes identified by the search focused on PET3 clusters with p53TFBS. Map images are generated by the UCSC Genome Browser. Maps for genes which have PET3 clusters with p53TFBS on their TSS, RPS27L (**A**) and AEN (**B**), and for genes which have PET3 clusters with p53TFBS in upstream of their TSS, NOTCH1 (**C**) and PLK2 (**D**), are shown respectively. In each panel, calibrated log2 expression ratio (HU-treated / untreated) of each gene in HCT116 WT and KO cells is plotted as bar graph in the vertical direction ranging from −1.5 to +1.5, in the tracks labeled "HU_24h_WT"and "HU_24h_KO", respectively. Magenta-colored bar graph shows positive value, which means that the gene expression is increased in HU-treated cells. Blue-colored bar graph shows negative value, which means that the expression is decreased in HU-treated cells. Width of each bar graph is coordinated to genomic position of each transcript. Other map symbols are shown in the same manner as Figure 3.

flanking regions of PET3 clusters with p53TFBS since the p53 binding sites are often distant from the locus of the p53 target gene on the genome. Within 5-kb flanking region of PET3 clusters with p53TFBS, two known p53 target genes NOTCH1[39] and PLK2[40] were found, and those expressions were increased specifically in HU-treated HCT116 WT cells (Fig. 4C, D). By further expansion of search up to 30-kb flanking region of PET3 clusters with p53TFBS, additional 17 genes were found, including the reported p53 target genes COL4A1[24], GML[41] and IER5[42] (Table 2). These suggest that the p53 target genes are effectively detected by search for flanking regions of PET clusters with p53TFBS, and other genes picked up might be candidate p53 targets worth being examined by molecular biological experiment.

p53 target genes predicted by search focused on flanking regions of transcription start sites which have PET3 cluster

Although the target prediction focusing on PET3 clusters with p53TFBS picked up some putative and known p53 target genes, the number of PET3 clusters which have the internal p53TFBS was relatively less (9.5%, 31/327), in part due to incomplete list of p53TFBS in the TRANSFAC database which is based on purely computational prediction[29]. In order to screen genome-wide p53 target genes more comprehensively, we examined if Refseq genes have PET3 clusters around their TSS or not, irrespective of p53TFBS (Fig. 2B,D), by using the Galaxy tool *Intersect.* Out of total 27,090 Refseq genes, we found 278 (1.0%) genes (excluding isoforms) have PET3 clusters within 50-kb flanking regions of their TSS (Table 3.), including 61 genes reported previously[24, 43]. Classification of gene function based on gene ontology terms revealed that 33.1% (92/278) of putative and known p53 target genes were involved in biosynthesis or cellular metabolic process, in addition to known p53-related functions including apoptosis or cell cycle (9.7%, 27/278), chromatin modification or DNA repair (3.6%, 10/278), and cell growth or differentiation (15.5%, 43/278).

Analysis of positional distribution of PET3 clusters (Table 3, Fig. 5A) revealed that 22.3% (62/278) of genes have PET3 clusters within 10-kb flanking regions of TSS, which included 27 of reported p53 target genes. However, remaining 216 genes including 34 of reported p53 target genes had PET3 clusters 10 to 50-kb away from TSS. These results suggest that p53 often bind to regulatory region which is more than 10-kb distant from TSS, and regulate gene expression.

To gain insight into the putative p53 target genes, we calculated expression difference between HCT116 WT and KO cells treated by HU for each gene according to the microarray data[34], and combined with the data of position of PET3 clusters. By the combined plot analysis (Fig. 5A), we found that a novel p53 target candidate DKFZP564O0823 that has PET3 cluster 46-kb upstream of TSS was upregulated specifically in HCT116 WT cells (Fig. 6B). We also found that a novel target candidate DDX60L (DEAD (Asp-Glu-Ala-Asp) box polypeptide 60-like) which has PET3 cluster in an intron 14-kb downstream of TSS was downregulated specifically in HCT116 WT cells (Fig. 6C). Furthermore, this analysis illustrated that the PET3 cluster overlapping with the TSS of the reported p53 target TNFRSF10B (tumor necrosis factor receptor superfamily, member 10b)[44] is shared by adjacent family genes, TNFRSF10C and TNFRSF10D because their expressions were similarly induced in HCT116 WT cells (Fig. 6A,D). Taken together, these results indicate that the combined analysis of TSS, p53 binding regions deduced by PET3 cluster, and gene expression would provide comprehensive prediction of p53 target genes, and also give information on common promoters shared by several p53 target genes around a p53 binding site.

PET3 clusters which were not located within 50-kb flanking regions of TSS of Refseq genes

We next examined how many PET3 clus-

ters were associated with the 278 of p53 target genes which were picked up by the TSS-focused analysis. Out of total 327 PET3 clusters on the genome, 157 (48%) of PET3 clusters were associated with the Refseq genes, that is, located within 50-kb flanking regions of TSS of Refseq genes, while other 170 (52%) of PET3 clusters were not. By using the Galaxy Tool *Intersect,* 55 out of the remaining 170 PET3 clusters were shown to be located more than 50-kb away from TSS but within the loci of Refseq genes, which included 19 of reported p53 targets (ASTN2, CDKAL1, CTNNA3, ERBB4, FRMD4A, GPR39, KIAA0564, NAV3, NEO1, NLGN1, NR6A1, PHF14, PRKAG2, PTPRM, SHROOM3, SLC4A10, TPO, USP34 and

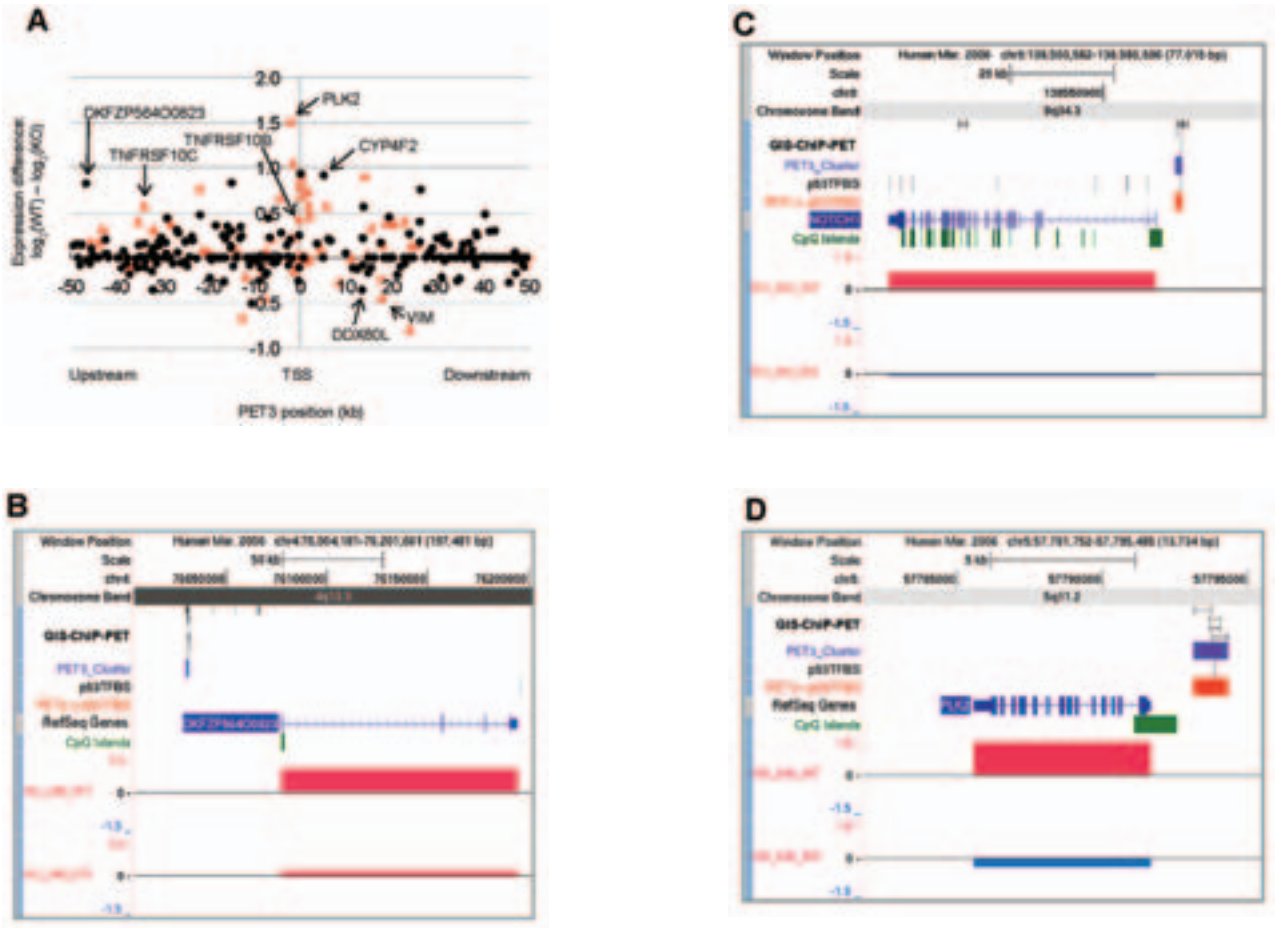

Fig. 5 p53 target genes predicted by the search focused on flanking regions of transcriptional start sites which have PET3 cluster. **A**. Two hundred and seventy-eight PET3 clusters, which were found within 50-kb flanking regions of TSS of genes, are plotted against positions from TSS (x axis) and gene-expression difference for corresponding genes between HU-treated HCT116 WT and KO cells (y axis). PET3 clusters located in the flanking regions of TSS of known p53 target genes are plotted as red triangles, and other PET3 clusters for putative p53 target genes are plotted as black dots. Some plots are labeled by the gene names which are closest to the corresponding PET3 clusters. **B and C**. Maps for putative p53 target genes identified by the TSS-focused search combined with the analysis of expression difference. DKFZP564O0823 (**B**) and DDX60L (**C**) have PET3 clusters upstream and downstream of their TSS, respectively. **D**. Maps for TNFRSF10 family genes which are possibly upregulated by one p53 binding site.

USP9X)[24] (data not shown). Furthermore, by the analysis combined with gene expression data, we found that novel p53 target candidates ANKS1B which has the PET3 cluster 280-kb away from TSS and were upregulated in HCT116 WT cells (Fig. 6A), and SBF2 which has the PET3 cluster 250-kb away from TSS and were downregulated in HCT116 WT cells (Fig. 6B). The remaining 115 PET3 clusters were not located on any regions within 50-kb franking regions of TSS, or on gene loci of Refseq genes. These findings suggest that the expression of part of huge p53 target genes which consist of many exons and long introns might be regulated by p53 binding site distant from TSS but within the genes loci.

p53-target non-coding RNAs predicted by search focused on flanking regions of transcriptional start sites which have PET3 cluster

Recent analyses of the mammalian transcriptome have shown that the transcript-abundant regions occupy more than half of the genomic sequence[45]. To investigate the possibility that p53 regulate the expression of various transcripts other than well-annotated mRNAs which code proteins, we sought to look for PET3 clusters around TSS of RNAs other than coding RNAs. The data set of UCSC genes[30], which is a moderately conservative set of gene predictions and is based on RefSeq, Genbank, Consensus Coding Sequences (CCDS) and UniProt databases, has approximately five times as many putative non-coding genes as the RefSeq gene collection. We searched within 10-kb flanking regions of TSS of non-coding RNAs (13,762 of total 66,803 transcripts, including 6,686 narrowly-defined noncoding transcripts, 6,295 nearCoding transcripts and 781 antisense transcripts: see the definition of each transcript in Materials and Methods) in the UCSC genes collection for PET3 clusters, by using Galaxy tool *Intersect.* Out of total 327 PET3 clusters on the genome, 31 (9.5%) of PET3 clusters were found to be located within 10-kb flanking regions of TSS of non-coding RNAs. On the other hand, out of the 13,762 non-coding RNAs, 79 (0.6%) transcripts, which include 44 of narrowly-defined noncoding transcripts, 35 of nearCoding transcripts, and none of antisense transcripts, had PET3 clusters within 10-kb flanking region of their TSS. For example, PVT1[46-48], which is also registered as a non-coding RNA in the Refseq genes collection and listed in Table 3, had the TSS overlapped with PET3 cluster with p53TFBS (Fig. 7A). In

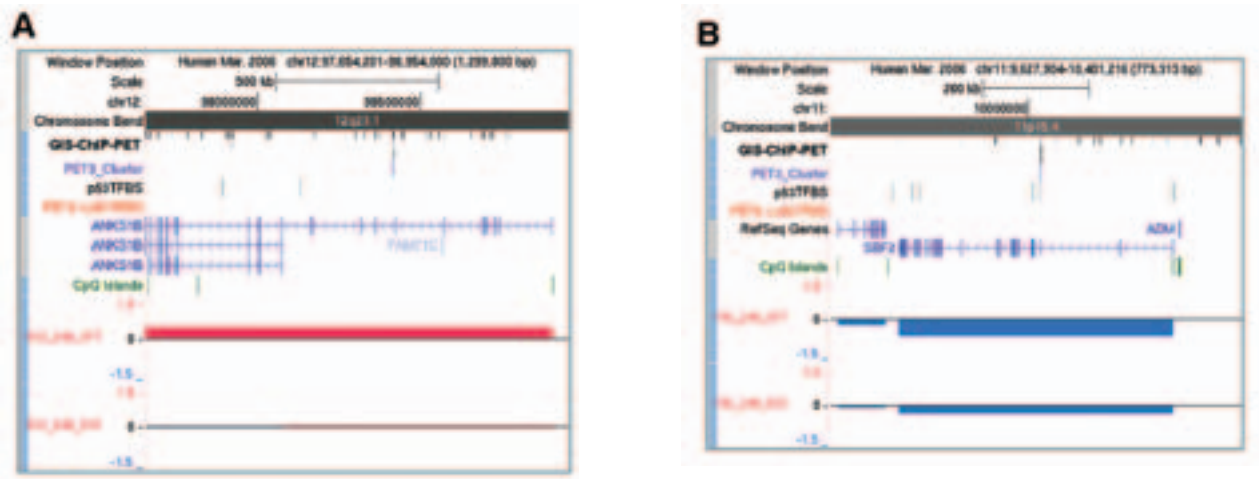

Fig. 6 Gene maps of putative p53 targets ANKS1B **(A)** and SBF2 **(B)**, which have PET3 clusters more than 50-kb distant from their TSS. See text for details.

addition, non-coding RNAs ANKRD19, CR605611 and AX721264, which are not registered in the Refseq collection, had PET3 clusters within 10-kb flanking regions of their TSS, respectively (Fig. 7B,C,D). Since non-coding RNAs often have no corresponding probes on standard microarrays and therefore lack information on their expression, we retrieved the existing data of GIS PET-RNA, which shows the starts and ends of full-length mRNA transcripts, and whose number is rough indication of gene expression level. PET-RNAs of 5-FU-treated HCT116 cells were observed in the genomic regions corresponding to PVT1 and ANKRD19 loci (Fig. 7A,B, track

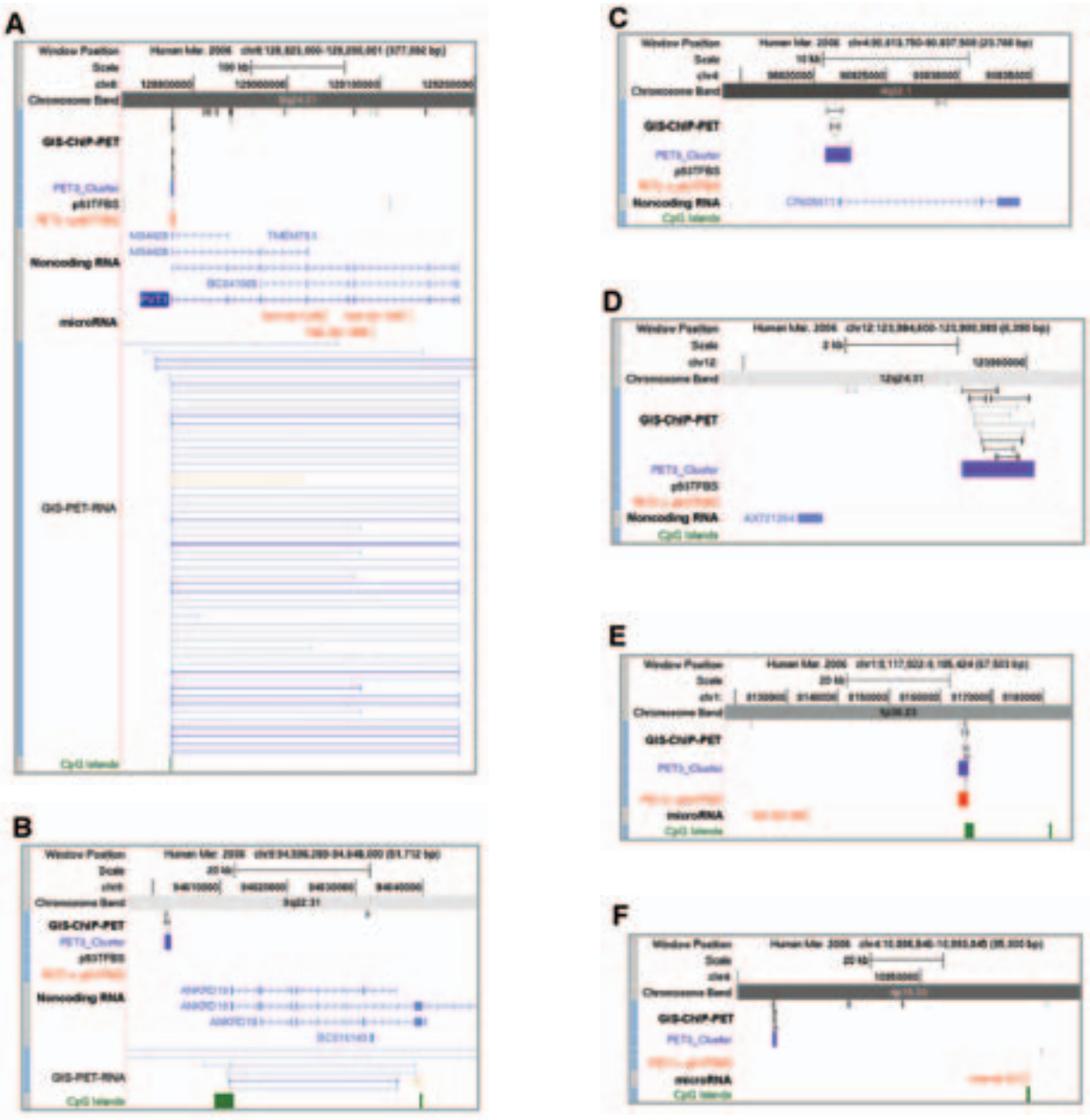

Fig. 7 Gene maps for putative p53-target non-coding RNAs PVT1 **(A)**, ANKRD19 **(B)**, CR605611 **(C)** and AX721264 **(D)** and for the known p53-target microRNA hsa-mir-34a **(E)** and putative target hsa-mir-572 **(F)**, which have PET3 clusters on TSS or in their upstream regions, respectively. See text for details.

Table 3. Refseq genes which have PET3 cluster within 50-kb flanking regions of their TSSs

Name	Accession Number	Chr.	Strand	Distance from TSS to the closest PET3 cluster[2] (bp)	p53TFBS[3]	CpG[4] (>100bp)	Gene ontology[5]								Reported[6]
							Apoptosis・Cell cycle	Chromatin modification・DNA repair	Cell growth・Differentiation	Transcriptional regulation	Signal Transduction	Cell Adhesion	Biosynthesis・Metabolism	Transport・Ion channel	
ABCB9	NM_019624	12	-	38,336										■	
ACP2	NM_001610	11	-	33,602		■									
ADAMTS3	NM_014243	4	-	-2,152									■		
ADCY9	NM_001116	16	-	-24,898							■		■		
ADIPOR1	NM_001127687	1	-	-43,846							■		■		
ADORA2B	NM_000676	17	+	18,849							■		■		■
AEBP2	NM_153207	12	+	-7,495	■										
AEN	NM_022767	15	+	-98	■	■									■
AK3	NM_016282	9	-	-47,595									■		
AKAP10	NM_007202	17	-	-9,297							■				
AMZ1	NM_133463	7	+	16,179									■		
ANK1	NM_000037	8	-	-21,105							■			■	■
ANKRD10	NM_017664	13	-	-37,178											■
ANKRD45	NM_198493	1	-	27,408											
AP3B2	NM_004644	15	-	-32,373									■	■	
ARFGAP2	NM_032389	11	-	-38,437		■					■			■	
ARHGEF1	NM_004706	19	+	-22,620	■	■			■		■				
ARMC2	NM_032131	6	+	48,515											
ART1	NM_004314	11	+	-29,033									■		
ART5	NM_053017	11	-	26,158									■		
ATAD1	NM_032810	10	-	-25,080											
ATF3	NM_004024	1	+	-11,836						■			■		■
ATL1	NM_015915	14	+	40,649											
ATP5A1	NM_004046	18	-	27,864									■	■	
BAX	NM_004324	19	+	69	■	■	■		■		■				■
BAZ1B	NM_032408	7	-	-11,520				■		■			■		
BCAM	NM_005581	19	+	-37,702							■	■			
BCL2A1	NM_004049	15	-	-31,988			■		■						■
BCL3	NM_005178	19	+	22,659			■		■	■	■		■	■	
BCL7B	NM_001707	7	-	23,897											
BEST1	NM_004183	11	+	49,418										■	
BICD2	NM_015250	9	-	-35,356											■
BRP44L	NM_016098	6	-	34,841											
C15orf33	NM_152647	15	-	-4,607											
C17orf62	NM_001033046	17	-	-45,469											
C17orf98	NM_001080465	17	-	-28,345		■									
C1orf88	NM_181643	1	+	18,332											
C20orf30	NM_014145	20	-	-7,140		■									
C2orf51	NM_152670	2	+	13,819											
C6orf138	NM_001013732	6	-	-42,301											
C9orf114	NM_016390	9	-	46,946											
C9orf27	NR_024032	9	-	-18,850	■										
CACNA1E	NM_000721	1	+	38,496										■	
CALD1	NM_033139	7	+	-45,485											■
CAPN12	NM_144691	19	-	-30,978									■		
CBLC	NM_012116	19	+	-6,490							■		■		■
CCDC49	NM_017748	17	-	-44,398		■									
CCDC5	NM_138443	18	+	-33,901			■								
CCDC54	NM_032600	3	+	-14,987											
CCDC58	NM_001017928	3	-	-10,969											
CCL14	NM_004166	17	-	-28,786					■		■				
CCL15	NM_004167	17	-	-28,786							■				
CCL16	NM_004590	17	-	-49,275											
CCL18	NM_002988	17	+	-33,845							■				
CCL23	NM_005064	17	-	-12,793					■		■				
CD70	NM_001252	19	-	13,679			■		■		■				
CD79A	NM_001783	19	+	-16,543	■	■			■		■				
CDC42EP3	NM_006449	2	-	17,903							■				■
CDH16	NM_004062	16	-	-8,779								■			
CDKN1A	NM_000389	6	+	-1,235	■	■	■		■				■		■
CDS2	NM_003818	20	+	-6,609		■							■		
CEACAM21	NM_033543	19	+	-47,948											
CES2	NM_003869	16	+	-6,802									■		
CES3	NM_024922	16	+	-33,593											
CHD2	NM_001271	15	+	49,820				■		■			■		■
CHI3L2	NM_004000	1	+	-13,015									■		
CHP2	NM_022097	16	+	-44,840											
CLEC6A	NM_001007033	12	+	-15,100											
CLRN3	NM_152311	10	-	29,300											
CROT	NM_021151	7	+	-14,811									■	■	
CSRP1	NM_004078	1	-	37,772		■									
CYB5R1	NM_016243	1	-	-35,142									■		
CYP3A4	NM_017460	7	-	46,289									■		
CYP3A7	NM_000765	7	-	-2,699											
CYP4F11	NM_021187	19	-	42,075											
CYP4F12	NM_023944	19	+	-26,541											
CYP4F2	NM_001082	19	-	5,283									■		
CYP4F3	NM_000896	19	+	5,639									■		■
CYP4F8	NM_007253	19	+	31,317									■		
DCUN1D3	NM_173475	16	-	34,246											
DDB2	NM_000107	11	+	363		■		■					■		■
DDIT4	NM_019058	10	+	23,874		■	■		■		■				■

Table 3. Continued

Name	Accession Number	Chr.	Strand	Distance from TSS to the closest PET3 cluster[2] (bp)	p53TFBS[3]	CpG[4] (>100bp)	Gene ontology[5]								Reported[6]
							Apoptosis・Cell cycle	Chromatin modification・DNA repair	Cell growth・Differentiation	Transcriptional regulation	Signal Transduction	Cell Adhesion	Biosynthesis・Metabolism	Transport・Ion channel	
DDX60L	NM_001012967	4	-	13,639											
DENND2D	NM_024901	1	-	-13,985											■
DERA	NM_015954	12	+	26,163									■		
DHDH	NM_014475	19	+	21,248	■	■							■		
DHRS2	NM_005794	14	+	34,382					■				■		
DKFZP564O0823	NM_015393	4	+	-46,565											
DMRTC2	NM_001040283	19	+	15,562	■	■			■	■			■		
DRAM	NM_018370	12	+	-21,929											■
DTWD1	NM_020234	15	+	4,430											■
DYNLT1	NM_006519	6	-	11,567											
EBNA1BP2	NM_006824	1	-	19,218											
EDN2	NM_001956	1	-	2,091							■				■
EFCAB3	NM_173503	17	+	35,569											
EIF6	NM_002212	20	-	-23,477									■		
ENDOG	NM_004435	9	+	-35,639									■		
ERN2	NM_033266	16	-	3,714			■		■	■			■		
FAM12A	NM_006683	14	+	6,258											
FAM12B	NM_022360	14	+	-16,230					■						
FAM162A	NM_014367	3	+	10,021											
FAM183A	NM_001101376	1	+	5,121											
FAM71A	NM_153606	1	+	-27,658											
FAM75B	NM_001098807	9	-	40,843											
FAM83C	NM_178468	20	-	-15,846											
FAM96B	NM_016062	16	-	6,759											
FCHO2	NM_138782	5	+	-234		■									
FER1L6	NM_001039112	8	+	-6,999											
FGF2	NM_002006	4	+	-28,254					■		■		■		■
FLJ46321	NM_001001670	9	+	32,476											
FOXK2	NM_004514	17	+	-23,431						■			■		
FTH1	NM_002032	11	-	-31,641					■					■	
FTL	NM_000146	19	+	-10,380	■	■								■	
GADD45A	NM_001924	1	+	1,431		■	■	■	■				■		■
GALNT11	NM_022087	7	+	46,854											
GAPDH	NM_002046	12	+	-26,641									■		
GAPT	NM_152687	5	+	-29,349	■										
GARNL3	NM_032293	9	+	30,884							■				
GAS6	NM_000820	13	+	26,375	■				■		■				
GDF15	NM_004864	19	+	-12,631							■				■
GDNF	NM_000514	5	-	-8,496			■		■		■			■	
GML	NM_002066	8	+	-19,037	■		■		■		■				■
GNPTAB	NM_024312	12	-	-24,544					■	■					
GPM6A	NM_005277	4	-	-12,831											
GPR182	NM_007264	12	+	36,913							■				
GSN	NM_000177	9	+	22,919									■		
GYS1	NM_002103	19	-	38,364	■	■							■		
HELLS	NM_018063	10	+	40,440			■	■	■	■			■		
HLA-B	NM_005514	6	-	19,639											
HNRNPA3	NM_194247	2	+	-42,401											
HSD17B3	NM_000197	9	-	35,252					■				■		
HSN2	NM_213655	12	+	-19,598					■						
IER5	NM_016545	1	+	11,121	■										■
IFFO1	NM_080731	12	-	41,259											
IFNAR1	NM_000629	21	+	42,515							■		■		
IFNGR2	NM_005534	21	+	-35,474							■				
IGFL2	NM_001002915	19	+	-13,917											
IGFL3	NM_207393	19	-	-9,191											
IL13RA1	NM_001560	X	+	12,594							■				
INSIG2	NM_016133	2	+	27,709									■		
IRF2BP2	NM_182972	1	-	-9,472											
ISG20	NM_002201	15	+	-17,610	■	■			■				■		
ITGAM	NM_000632	16	+	14,081							■	■			■
KCNC2	NM_139136	12	-	31,537										■	
KCNK16	NM_032115	6	-	39,442										■	
KCNK17	NM_031460	6	-	31,348										■	
KIAA0574	NM_015307	15	-	-31,827											
KIAA1370	NM_019600	15	-	14,875											
KILLIN	NM_001126049	10	-	20,198											
KITLG	NM_000899	12	-	20,469					■		■	■			■
KLHDC4	NM_017566	16	-	-1,698											
KLHL38	NM_001081675	8	-	40,688											
KRT80	NM_182507	12	-	15,569	■										■
LACTB	NM_032857	15	+	35,674	■	■							■		
LASP1	NM_006148	17	+	-381		■					■			■	
LGALS4	NM_006149	19	-	37,649								■			
LGALS7	NM_002307	19	-	-1,935			■		■			■			
LGALS7B	NM_001042507	19	+	-13,758											
LOC392979	NM_001085391	7	-	18,545											
LOC728621	NM_001080850	1	+	-1,268											■
LRRC25	NM_145256	19	-	24,079											
LSM3	NM_014463	3	+	-105		■							■		
LSM4	NM_012321	19	-	-41,675									■		
LY6D	NM_003695	8	-	-29,172	■							■			

Table 3. Continued

Name	Accession Number	Chr.	Strand	Distance from TSS to the closest PET3 cluster[2] (bp)	p53TFBS[3]	CpG[4] (>100bp)	Gene ontology[5]								Reported[6]
							Apoptosis・Cell cycle	Chromatin modification・DNA repair	Cell growth・Differentiation	Transcriptional regulation	Signal Transduction	Cell Adhesion	Biosynthesis・Metabolism	Transport・Ion channel	
LYNX1	NM_177457	8	-	-37,540	■										
LYPD4	NM_173506	19	-	-16,139	■	■									
LYRM1	NM_020424	16	+	-34,241											
MDM4	NM_002393	1	+	8,560			■		■	■			■		■
MED4	NM_014166	13	-	28,596						■	■		■		
METTL2A	NM_181725	17	+	-7,764											
MRPL51	NM_016497	12	-	-14,545									■		
MS4A13	NM_001012417	11	+	48,430							■				
MTHFD1L	NM_015440	6	+	-8,158									■		■
MYBL1	NM_001080416	8	-	22,692						■			■		■
MYO1A	NM_005379	12	-	18,627											■
MYO3B	NM_138995	2	+	43,950									■		
NAB1	NM_005966	2	+	-37,073						■			■		■
NARF	NM_012336	17	+	37,624											
NBPF10	NM_001039703	1	-	32,316											
NCAPD2	NM_014865	12	+	13,719			■								
NCK2	NM_001004720	2	+	-6,283					■		■		■		■
NID2	NM_007361	14	-	-7,387								■	■		■
NOTCH1	NM_017617	9	-	-4,414	■				■	■	■		■		■
NPL	NM_030769	1	+	-30,730									■		
NR1H3	NM_001130102	11	+	-33,593		■				■	■		■		
NUDT15	NM_018283	13	+	28,943											
OC90	NM_001080399	8	-	24,633									■		
OGFOD2	NM_024623	12	+	-46,633									■		
OR10A7	NM_001005280	12	+	-48,345							■				
OR9K2	NM_001005243	12	+	42,912											
PACSIN3	NM_016223	11	-	-28,898		■								■	
PALLD	NM_016081	4	+	-30,217											
PASD1	NM_173493	X	+	-19,609							■				
PCNA	NM_182649	20	-	226		■		■	■		■		■	■	■
PCNX	NM_014982	14	+	-42,006											
PDP2	NM_020786	16	+	47,110											
PGPEP1	NM_017712	19	+	24,269									■		■
PHLDA3	NM_012396	1	-	180		■									
PIK3C2B	NM_002646	1	-	-34,596											
PLK1	NM_005030	16	+	30,908			■		■				■		
PLK2	NM_006622	5	-	-2,068	■		■				■		■		■
PPFIA4	NM_015053	1	+	-48,765											
PPFIBP1	NM_003622	12	+	25,606								■			■
PRAGMIN	NM_001080826	8	-	18,972											
PRDM1	NM_182907	6	+	-501						■			■		■
PRR18	NM_175922	6	-	-39,775											
PSTPIP2	NM_024430	18	-	1,827											■
PTEN	NM_000314	10	+	-20,198			■		■		■	■	■		■
PTGFRN	NM_020440	1	+	-30,714	■										
PTK2	NM_005607	8	-	-7,819							■	■	■		■
PTPRE	NM_006504	10	+	-43,413							■		■		■
PTPRO	NM_030668	12	+	-10,652							■		■		■
PTPRV	NR_015444	9	+	14,792											
PVT1	NR_003367	8	+	792	■	■									
RAB8B	NM_016530	15	+	-32,023	■	■					■			■	
RCL1	NM_005772	9	+	-4,012											
RECK	NM_021111	9	+	37,945			■								
RNASE6	NM_005615	14	+	-28,854									■		
RPL23	NM_000978	17	-	-15,934		■							■	■	
RPS19	NM_001022	19	+	660	■	■			■				■	■	
RPS27L	NM_015920	15	-	-36	■	■							■		■
RRAD	NM_004165	16	-	-2,106							■				■
RRM2B	NM_015713	8	-	2,165				■					■		■
RUVBL2	NM_006666	19	+	-38,970	■	■		■	■	■			■		
S100A14	NM_020672	1	-	48,790											
S100A16	NM_080388	1	-	45,514											
S100A2	NM_005978	1	-	-1,695											■
S100A3	NM_002960	1	-	-18,267											
S100A4	NM_002961	1	-	-21,719					■						
S100A5	NM_002962	1	-	-25,760											
S100A6	NM_014624	1	-	-31,284			■		■		■				
SCARNA10	NR_004387	12	+	-2,372											
SCARNA15	NR_003011	15	+	-13,689											
SFT2D1	NM_145169	6	-	-5,655										■	
SHANK2	NM_012309	11	-	17,611							■				
SLC16A10	NM_018593	6	+	19,519										■	
SLC9A11	NM_178527	1	-	-39,414										■	
SNORA21	NR_002576	17	-	-16,739		■									
SNX16	NM_022133	8	-	-6,254										■	
SPTLC1	NM_006415	9	-	-32,217									■		
SSBP4	NM_032627	19	+	-45,799						■					
STARD4	NM_139164	5	-	17,695									■	■	■
STAU1	NM_004602	20	-	49,159											■
STOM	NM_004099	9	-	47,549									■		
SUSD1	NM_022486	9	-	29,921											
SYTL3	NM_001009991	6	+	-16,872										■	

Table 3. Continued

							Gene ontology[5]								
Name	Accession Number	Chr.	Strand	Distance from TSS to the closest PET3 cluster[2] (bp)	p53TFBS[3]	CpG[4] (>100bp)	Apoptosis・Cell cycle	Chromatin modification・DNA repair	Cell growth・Differentiation	Transcriptional regulation	Signal Transduction	Cell Adhesion	Biosynthesis・Metabolism	Transport・Ion channel	Reported[6]
TAC3	NM_013251	12	-	-14,942							■				
TBC1D13	NM_018201	9	+	-4,370							■				
TBL2	NM_012453	7	-	44,886											
TGFA	NM_003236	2	-	-43,003	■		■		■		■		■		■
TMEM183A	NM_138391	1	+	-4,988											
TMEM183B	NM_001079809	1	+	-4,990											
TMEM194A	NM_015257	12	-	47,308											
TNFAIP8	NM_014350	5	+	-31,921	■		■		■						■
TNFRSF10B	NM_003842	8	-	-184		■	■		■		■				■
TNFRSF10C	NM_003841	8	+	-33,926		■	■		■		■				■
TNFSF9	NM_003811	19	+	46,476			■		■		■				
TNNI1	NM_003281	1	-	-47,605		■									
TP53TG1	NR_015381	7	-	14,609											■
TRDMT1	NM_004412	10	-	-45,441									■		
TRIM55	NM_033058	8	+	-2,607							■				
TRPC2	NR_002720	11	+	-10,386											
TTC19	NM_017775	17	+	-35,703											
UBC	NM_021009	12	-	-23,969									■		
UBE2F	NM_080678	2	+	-26,795									■		
VAMP1	NM_014231	12	-	-37,173										■	
VIM	NM_003380	10	+	17,825											■
VPS37B	NM_024667	12	-	-32,009										■	
WDR5B	NM_019069	3	-	21,840									■		
WDR65	NM_152498	1	+	-19,312											
WSB1	NM_015626	17	+	37,319							■		■		
XPC	NM_004628	3	-	93		■	■	■			■		■		■
ZBTB39	NM_014830	12	-	-25,037						■			■		
ZER1	NM_006336	9	-	-10,942									■		
ZMAT3	NM_022470	3	-	1,632	■		■		■					■	■
ZNF322A	NM_024639	6	-	-45,033						■			■		
ZNF81	NM_007137	X	+	-17,740						■			■		
ZSWIM7	NM_001042697	17	-	35,928				■					■		
Total (genes)	278				32	40	27	10	43	21	59	11	92	30	61

[1]Accession numbers of isoforms are omitted.
[2]Distance between TSS and the midpoint of the region of the closest PET3 cluster is shown. Positive numbers are used in the direction of gene transcription, indicating that the location of PET3 cluster is downstream of the TSS. Negative numbers are used in the opposite direction of gene transcription, indicating that the location of PET3 cluster is upstream of the TSS.
[3]Gray solid boxes indicate that the PET3 cluster closest to each gene TSS has the internal p53TFBS.
[4]Gray solid boxes indicate that the PET3 cluster closest to each gene TSS overlaps with CpG island.
[5]Gray solid boxes indicate that the term shown in top column, which describes molecular function or biological process, is included in gene ontology terms of each gene.
[6]Gray solid boxes indicate that the gene is reported as a p53 target previously in the references 24) and 43).

"GIS-PET-RNA"), indicating that those 2 non-coding RNAs were expressed in 5-FU-treated HCT116 cells while no tags were observed in CR605611 and AX721264. These data suggest that expression of part of non-coding RNAs is also regulated by p53 and induced by DNA damage.

Putative p53-target microRNAs which have PET3 clusters in the upstream regions

MicroRNAs are essential post-transcriptional regulators that determine cell identity and fate, via mechanisms of cleavage-dependent RNA degradation of the transcript or miRNA-mediated translational repression of the target transcript[8]. Primary transcripts of microRNAs form a long hairpin loop, pri-miRNA, which is processed by Drosha to form pre-miRNA. Pre-miRNA is exported to the cytoplasm where Dicer cleaves off the hairpin loop to form a duplex that contains the mature 21-24 nucleotide microRNA. The mature miRNA is then incorporated into the RNA-induced silencing complex to target the 3' untranslated region of the target mRNA[49, 50]. Since aberrant expression of miRNAs can lead to diseases including cancer[51-53], we sought to examine if microRNAs have PET3 clusters near their loci and are under p53 regulation or not. Although microRNA belongs to non-coding RNA, we noticed that UCSC genes collection contains only one third of microRNAs (approximately 200 microRNAs, data not shown). Therefore, we analyzed the data of total 685 of human microRNAs retrieved from the miRbase

database[31], which have information on predicted hairpin portions of miRNA transcripts, but have no information on exact transcription start sites of primary transcripts.

Our search using the Galaxy tool *Intersect* revealed that only one microRNA, has-mir-34a, which has been reported as the first p53-target microRNA[5-7], has the PET3 cluster within 30-kb-upstream flanking regions of the microRNA (Fig. 7E). Expanded search revealed that two microRNAs, hsa-mir-572 (Fig. 7F) and hsa-mir-101-2 (data not shown) have PET3 clusters in 60 to 70-kb-upstream flanking regions of the microRNAs, and additional two microRNAs hsa-mir-181a-2 and hsa-mir-181b-2 have PET3 clusters in 90 to 100-kb-upstream flanking regions of the microRNAs, respectively (data not shown). We found only 5 (0.7%) of microRNAs related to PET3 clusters while future studies on exact TSS for each microRNA may change the result because primary transcripts of non-coding RNAs including microRNAs, which are marked by trimethylated histone H3 lysine 36 (H3K36me3)[54], seem to be quite huge and harbor clusters of microRNAs. Actually, PVT1 locus contained 3 microRNAs (has-mir-1205, -1206, and -1207) while they are > 100-kb distant from the TSS of PVT1 and the PET3 cluster with p53TFBS (Fig. 7A), suggesting that those 3 microRNAs could also be p53 targets.

DISCUSSION

In this report, we described the prediction of p53 target genes based on the data of sequencing analysis of p53 ChIP-PET, combined with existing substantial genomic data by using the Galaxy platform. We deduced total 327 of p53 binding sites on the genome of 5-FU treated HCT116 cells by picking up PET3 clusters, and subsequently predicted 278 of Refseq genes, 79 of non-coding RNAs and 5 of microRNAs as p53 targets which included lots of known validated targets, suggesting that our analysis worked effectively. Although there are approximately 130 of well-validated protein-coding p53 target genes[24, 43], it is not fully understood how many non-coding RNAs, which are presumably much more than coding RNAs, are regulated by p53. Since p53 ChIP-PET and ChIP-seq analysis shows unbiased genome-wide information on p53 binding sites, accumulation of p53 ChIP data under various conditions (cell and tissue type, and the stress) would provide more comprehensive understanding of p53 targets including non-coding RNAs.

While ChIP-PET has been a useful approach for identifying candidates, we first extracted PET3 clusters which contain computationally-expected p53 binding sites (p53TFBS) in their regions, in order to increase the probability of prediction. Unexpectedly, PET3 clusters which have p53TFBS was less (9.5%). PET3 clusters associated with validated targets often had no internal p53TFBS (Table 3), implying the problem of computational expectation of p53 binding sites in the TRANSFAC database, that is, the strength and predictive power of expectation model is dependent on the sampling size and quality of the training set. On the other hand, the recent study using p53 ChIP-on-chip employing high-resolution tiling arrays with an average probe spacing 100bp, reported high-confidence motif that was contained in 83% of all binding sites[23]. Future studies such as p53 ChIP-seq would provide higher-resolution data which serve as a valuable knowledgebase for p53 binding sites to clarify the p53 binding motif.

Our combined analysis of positional distribution of PET3 clusters with change of gene expression after DNA damage (Fig. 6A) showed that 216 of genes including 34 of reported p53 target genes had PET3 clusters 10 to 50-kb away from TSS, and in part, expression of those is likely to be regulated by p53. Generally, the p53 RE in a gene is most commonly located in the 5' promoter-enhancer region of the gene (~50%) or in intron 1 (~25%)[43]. More rarely it is located in introns 2 or 3 of a gene. Therefore, in addition to the distance between TSS and p53 binding sites, if PET3 cluster is found downstream of TSS and within the gene locus, taking account of the exon-intron structure of each

gene would provide more sensitive prediction of p53 targets. On the other hand, it is also known that eukaryotic cells contain transcription-factor-binding proteins that bind together sticky transcription factors and can mediate DNA looping. This process can bring distal transcription-factor-bound binding sites close to the TATA box, and can confer regulation. In case of p53, in the absence of a proximal p53 RE, the p53-cofactor Sp1 has been reported to serve as a surrogate, provided that the RE is present close to the TSS and the distal p53 RE. An example of Sp1-mediated DNA looping may be found in MDM2, where a functional single-nucleotide polymorphism (SNP), SNP309 T/G, within a cluster of Sp1-binding sites affects the level of regulation of nearby oestrogen and p53 REs, and has been associated with an early onset of breast cancer in pre-menopausal women[55]. Further investigation will be necessary to determine exactly how and when distant p53 REs regulate gene expression.

By using the Galaxy platform, we carried out almost all of analyses in this study, except for very limited data arrangement for gene expression and ontology. The whole or part of data from the seven databases (GIS ChIP-PET, TRANSFAC, Refseq genes, UCSC genes, microRNAs, CpG islands and GIS ChIP-RNA) was easily retrieved by using the Galaxy tool *Get Data,* and transferred into the Galaxy history system without consuming any hard disk space of our own computers. Furthermore, the results of analyses were outputted to the UCSC genome browser by a few clicks, and immediately, we were able to watch and explore the scalable and intuitive genome map that illustrates the positional relation among PETs, PET3 clusters, p53TFBS, CpG islands and genes. The Galaxy tool *Intersect* enabled us to found the genomic regions which meet multiple conditions, for example, the PET3 clusters which have an internal p53TFBS and located on TSS of a gene. Subsequently, we were able to predict p53 target genes by some different characterizations of PET3 clusters. Surprisingly, the Galaxy tools for the analysis of genomic fragments returned the results very quickly, for example, the Galaxy tool *Cluster* extracted 327 of PET3 clusters from the total 65,509 PETs in just 15 seconds. In recent studies of transcription factors by ChIP-seq analyses using next-generation sequencers, various types of software (e.g. PeakFinder[25], SISSRs[56] and FindPeaks 3.1[57]) have been developed to indentify tag-enriched regions as transcription factor binding sites. Although those programs are equipped with refined statistical algorithms for tag-clustering, it seems to be too hard for biologists to setup the environment, install and run the software by giving appropriate command-line parameters. We think that the Galaxy platform is a biologist-friendly tool of first choice to analyze and overlook large-scale ChIP data.

In conclusion, we showed that the Galaxy platform fits the analysis of ChIP-PET data combined with other genomic data. Progress of ChIP-seq studies will promote the evolution of the Galaxy platform as a versatile genome analysis tool and vice versa. Further studies are warranted to validate the putative p53 target genes experimentally.

ACKNOWLEDGEMENT

This study was supported in part by Grants-in-Aid for Scientific Research on Priority Areas from the Ministry of Education, Culture, Sports, Science, and Technology (to H.M., and T.T.).

REFERENCES

1. Vogelstein B, Lane D, Levine AJ. Surfing the p53 network. Nature 2000; 408: 307-310.
2. Vousden KH, Lu X. Live or let die: the cell's response to p53. Nat Rev Cancer 2002; 2: 594-604.
3. Sbisa E, Catalano D, Grillo G, Licciulli F, Turi A, Liuni S, Pesole G, De Grassi A, Caratozzolo MF, D'Erchia AM, Navarro B, Tullo A, Saccone C, Gisel A. p53FamTaG: a database resource of human p53, p63 and p73 direct target genes combining in silico pre-

diction and microarray data. BMC Bioinformatics 2007; 8 (Suppl 1): S20.
4. Horn HF, Vousden KH. Coping with stress: multiple ways to activate p53. Oncogene 2007; 26: 1306–1316.
5. Chang TC, Wentzel EA, Kent OA, Ramachandran K, Mullendore M, Lee KH, Feldmann G, Yamakuchi M, Ferlito M, Lowenstein CJ, Arking DE, Beer MA, Maitra A, Mendell JT. Transactivation of miR-34a by p53 broadly influences gene expression and promotes apoptosis. Mol Cell 2007; 26: 745–752.
6. He L, He X, Lim LP, de Stanchina E, Xuan Z, Liang Y, Xue W, Zender L, Magnus J, Ridzon D, Jackson AL, Linsley PS, Chen C, Lowe SW, Cleary MA, Hannon GJ. A microRNA component of the p53 tumour suppressor network. Nature 2007; 447: 1130–1134.
7. Raver-Shapira N, Marciano E, Meiri E, Spector Y, Rosenfeld N, Moskovits N, Bentwich Z, Oren M. Transcriptional activation of miR-34a contributes to p53-mediated apoptosis. Mol Cell 2007; 26: 731–743.
8. Flynt AS, Lai EC. Biological principles of microRNA-mediated regulation: shared themes amid diversity. Nat Rev Genet 2008; 9: 831–842.
9. Toledo F, Wahl GM. Regulating the p53 pathway: in vitro hypotheses, in vivo veritas. Nat Rev Cancer 2006; 6: 909–923.
10. Hollstein M, Sidransky D, Vogelstein B, Harris CC. p53 mutations in human cancers. Science 1991; 253: 49–53.
11. Leary RJ, Lin JC, Cummins J, Boca S, Wood LD, Parsons DW, Jones S, Sjoblom T, Park BH, Parsons R, Willis J, Dawson D, Willson JK, Nikolskaya T, Nikolsky Y, Kopelovich L, Papadopoulos N, Pennacchio LA, Wang TL, Markowitz SD, Parmigiani G, Kinzler KW, Vogelstein B, Velculescu VE. Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers. Proc Natl Acad Sci U S A 2008; 105: 16224–16229.
12. el-Deiry WS, Kern SE, Pietenpol JA, Kinzler KW, Vogelstein B. Definition of a consensus binding site for p53. Nat Genet 1992; 1: 45–49.
13. Stormo GD, Fields DS. Specificity, free energy and information content in protein-DNA interactions. Trends Biochem Sci 1998; 23: 109–113.
14. Eddy SR. Profile hidden Markov models. Bioinformatics 1998; 14: 755–763.
15. Maruyama R, Aoki F, Toyota M, Sasaki Y, Akashi H, Mita H, Suzuki H, Akino K, Ohe-Toyota M, Maruyama Y, Tatsumi H, Imai K, Shinomura Y, Tokino T. Comparative genome analysis identifies the vitamin D receptor gene as a direct target of p53-mediated transcriptional activation. Cancer Res 2006; 66: 4574–4583.
16. Martone R, Euskirchen G, Bertone P, Hartman S, Royce TE, Luscombe NM, Rinn JL, Nelson FK, Miller P, Gerstein M, Weissman S, Snyder M. Distribution of NF-kappaB-binding sites across human chromosome 22. Proc Natl Acad Sci U S A 2003; 100: 12247–12252.
17. Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, Kampa D, Piccolboni A, Sementchenko V, Cheng J, Williams AJ, Wheeler R, Wong B, Drenkow J, Yamanaka M, Patel S, Brubaker S, Tammana H, Helt G, Struhl K, Gingeras TR. Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of noncoding RNAs. Cell 2004; 116: 499–509.
18. Yang A, Zhu Z, Kapranov P, McKeon F, Church GM, Gingeras TR, Struhl K. Relationships between p63 binding, DNA sequence, transcription activity, and biological function in human cells. Mol Cell 2006; 24: 593–602.
19. Kaneshiro K, Tsutsumi S, Tsuji S, Shirahige K, Aburatani H. An integrated map of p53-binding sites and histone modification in the human ENCODE regions. Genomics 2007; 89: 178–188.

20. Jen KY, Cheung VG. Identification of novel p53 target genes in ionizing radiation response. Cancer Res 2005; 65: 7666-7673.
21. Ceribelli M, Alcalay M, Vigano MA, Mantovani R. Repression of new p53 targets revealed by ChIP on chip experiments. Cell Cycle 2006; 5: 1102-1110.
22. Shaked H, Shiff I, Kott-Gutkowski M, Siegfried Z, Haupt Y, Simon I. Chromatin immunoprecipitation-on-chip reveals stress-dependent p53 occupancy in primary normal cells but not in established cell lines. Cancer Res 2008; 68: 9671-9677.
23. Smeenk L, van Heeringen SJ, Koeppel M, van Driel MA, Bartels SJ, Akkers RC, Denissov S, Stunnenberg HG, Lohrum M. Characterization of genome-wide p53-binding sites upon stress response. Nucleic Acids Res 2008; 36: 3639-3654.
24. Wei CL, Wu Q, Vega VB, Chiu KP, Ng P, Zhang T, Shahab A, Yong HC, Fu Y, Weng Z, Liu J, Zhao XD, Chew JL, Lee YL, Kuznetsov VA, Sung WK, Miller LD, Lim B, Liu ET, Yu Q, Ng HH, Ruan Y. A global map of p53 transcription-factor binding sites in the human genome. Cell 2006; 124: 207-219.
25. Johnson DS, Mortazavi A, Myers RM, Wold B. Genome-wide mapping of in vivo protein-DNA interactions. Science 2007, 316: 1497-1502.
26. Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, Zeng T, Euskirchen G, Bernier B, Varhol R, Delaney A, Thiessen N, Griffith OL, He A, Marra M, Snyder M, Jones S. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods 2007; 4: 651-657.
27. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K. High-resolution profiling of histone methylations in the human genome. Cell 2007; 129: 823-837.
28. Giardine B, Riemer C, Hardison RC, Burhans R, Elnitski L, Shah P, Zhang Y, Blankenberg D, Albert I, Taylor J, Miller W, Kent WJ, Nekrutenko A. Galaxy: a platform for interactive large-scale genome analysis. Genome Res 2005; 15: 1451-1455.
29. Matys V, Fricke E, Geffers R, Gossling E, Haubrock M, Hehl R, Hornischer K, Karas D, Kel AE, Kel-Margoulis OV, Kloos DU, Land S, Lewicki-Potapov B, Michael H, Munch R, Reuter I, Rotert S, Saxel H, Scheer M, Thiele S, Wingender E. TRANSFAC: transcriptional regulation, from patterns to profiles. Nucleic Acids Res 2003; 31: 374-378.
30. Hsu F, Kent WJ, Clawson H, Kuhn RM, Diekhans M, Haussler D. The UCSC Known Genes. Bioinformatics 2006; 22: 1036-1046.
31. Griffiths-Jones S. miRBase: the microRNA sequence database. Methods Mol Biol 2006; 342: 129-138.
32. Zweig AS, Karolchik D, Kuhn RM, Haussler D, Kent WJ. UCSC genome browser tutorial. Genomics 2008; 92: 75-84.
33. McGlincy NJ, Smith CW. Alternative splicing resulting in nonsense-mediated mRNA decay: what is the meaning of nonsense? Trends Biochem Sci 2008; 33: 385-393.
34. Staib F, Robles AI, Varticovski L, Wang XW, Zeeberg BR, Sirotin M, Zhurkin VB, Hofseth LJ, Hussain SP, Weinstein JN, Galle PR, Harris CC. The p53 tumor suppressor network is a key responder to microenvironmental components of chronic inflammatory stress. Cancer Res 2005; 65: 10255-10264.
35. Ng P, Wei CL, Sung WK, Chiu KP, Lipovich L, Ang CC, Gupta S, Shahab A, Ridwan A, Wong CH, Liu ET, Ruan Y. Gene identification signature (GIS) analysis for transcriptome characterization and genome annotation. Nat Methods 2005; 2: 105-111.
36. el-Deiry WS, Tokino T, Waldman T, Oliner JD, Velculescu VE, Burrell M, Hill DE, Healy E, Rees JL, Hamilton SR, et al. Topological control of p21WAF1/CIP1 expression in normal and neoplastic tissues. Cancer Res 1995; 55: 2910-2919.

37. Li J, Tan J, Zhuang L, Banerjee B, Yang X, Chau JF, Lee PL, Hande MP, Li B, Yu Q. Ribosomal protein S27-like, a p53-inducible modulator of cell fate in response to genotoxic stress. Cancer Res 2007; 67: 11317-11326.
38. Kawase T, Ichikawa H, Ohta T, Nozaki N, Tashiro F, Ohki R, Taya Y. p53 target gene AEN is a nuclear exonuclease required for p53-dependent apoptosis. Oncogene 2008; 27: 3797-3810.
39. Lefort K, Mandinova A, Ostano P, Kolev V, Calpini V, Kolfschoten I, Devgan V, Lieb J, Raffoul W, Hohl D, Neel V, Garlick J, Chiorino G, Dotto GP. Notch1 is a p53 target gene involved in human keratinocyte tumor suppression through negative regulation of ROCK1/2 and MRCKalpha kinases. Genes Dev 2007; 21: 562-577.
40. Burns TF, Fei P, Scata KA, Dicker DT, El-Deiry WS. Silencing of the novel p53 target gene Snk/Plk2 leads to mitotic catastrophe in paclitaxel (taxol)-exposed cells. Mol Cell Biol 2003; 23: 5556-5571.
41. Kimura Y, Furuhata T, Urano T, Hirata K, Nakamura Y, Tokino T. Genomic structure and chromosomal localization of GML (GPI-anchored molecule-like protein), a gene induced by p53. Genomics 1997; 41: 477-480.
42. Kis E, Szatmari T, Keszei M, Farkas R, Esik O, Lumniczky K, Falus A, Safrany G. Microarray analysis of radiation response genes in primary human fibroblasts. Int J Radiat Oncol Biol Phys 2006; 66: 1506-1514.
43. Riley T, Sontag E, Chen P, Levine A. Transcriptional control of human p53-regulated genes. Nat Rev Mol Cell Biol 2008; 9: 402-412.
44. Takimoto R, El-Deiry WS. Wild-type p53 transactivates the KILLER/DR5 gene through an intronic sequence-specific DNA-binding site. Oncogene 2000; 19: 1735-1743.
45. Carninci P, Kasukawa T, Katayama S, Gough J, Frith MC, Maeda N, Oyama R, Ravasi T, Lenhard B, Wells C, Kodzius R, Shimokawa K, Bajic VB, Brenner SE, Batalov S, Forrest AR, Zavolan M, Davis MJ, Wilming LG, Aidinis V, Allen JE, Ambesi-Impiombato A, Apweiler R, Aturaliya RN, Bailey TL, Bansal M, Baxter L, Beisel KW, Bersano T, Bono H, Chalk AM, Chiu KP, Choudhary V, Christoffels A, Clutterbuck DR, Crowe ML, Dalla E, Dalrymple BP, de Bono B, Della Gatta G, di Bernardo D, Down T, Engstrom P, Fagiolini M, Faulkner G, Fletcher CF, Fukushima T, Furuno M, Futaki S, Gariboldi M, Georgii-Hemming P, Gingeras TR, Gojobori T, Green RE, Gustincich S, Harbers M, Hayashi Y, Hensch TK, Hirokawa N, Hill D, Huminiecki L, Iacono M, Ikeo K, Iwama A, Ishikawa T, Jakt M, Kanapin A, Katoh M, Kawasawa Y, Kelso J, Kitamura H, Kitano H, Kollias G, Krishnan SP, Kruger A, Kummerfeld SK, Kurochkin IV, Lareau LF, Lazarevic D, Lipovich L, Liu J, Liuni S, McWilliam S, Madan Babu M, Madera M, Marchionni L, Matsuda H, Matsuzawa S, Miki H, Mignone F, Miyake S, Morris K, Mottagui-Tabar S, Mulder N, Nakano N, Nakauchi H, Ng P, Nilsson R, Nishiguchi S, Nishikawa S, Nori F, Ohara O, Okazaki Y, Orlando V, Pang KC, Pavan WJ, Pavesi G, Pesole G, Petrovsky N, Piazza S, Reed J, Reid JF, Ring BZ, Ringwald M, Rost B, Ruan Y, Salzberg SL, Sandelin A, Schneider C, Schonbach C, Sekiguchi K, Semple CA, Seno S, Sessa L, Sheng Y, Shibata Y, Shimada H, Shimada K, Silva D, Sinclair B, Sperling S, Stupka E, Sugiura K, Sultana R, Takenaka Y, Taki K, Tammoja K, Tan SL, Tang S, Taylor MS, Tegner J, Teichmann SA, Ueda HR, van Nimwegen E, Verardo R, Wei CL, Yagi K, Yamanishi H, Zabarovsky E, Zhu S, Zimmer A, Hide W, Bult C, Grimmond SM, Teasdale RD, Liu ET, Brusic V, Quackenbush J, Wahlestedt C, Mattick JS, Hume DA, Kai C, Sasaki D, Tomaru Y, Fukuda S, Kanamori-Katayama M, Suzuki M, Aoki J, Arakawa T, Iida J, Imamura K, Itoh M, Kato T, Kawaji H, Kawagashira N, Kawashima T, Kojima M,

Kondo S, Konno H, Nakano K, Ninomiya N, Nishio T, Okada M, Plessy C, Shibata K, Shiraki T, Suzuki S, Tagami M, Waki K, Watahiki A, Okamura-Oho Y, Suzuki H, Kawai J, Hayashizaki Y. The transcriptional landscape of the mammalian genome. Science 2005; 309: 1559-1563.

46. Huppi K, Volfovsky N, Runfola T, Jones TL, Mackiewicz M, Martin SE, Mushinski JF, Stephens R, Caplen NJ. The identification of microRNAs in a genomically unstable region of human chromosome 8q24. Mol Cancer Res 2008; 6: 212-221.
47. Guan Y, Kuo WL, Stilwell JL, Takano H, Lapuk AV, Fridlyand J, Mao JH, Yu M, Miller MA, Santos JL, Kalloger SE, Carlson JW, Ginzinger DG, Celniker SE, Mills GB, Huntsman DG, Gray JW. Amplification of PVT1 contributes to the pathophysiology of ovarian and breast cancer. Clin Cancer Res 2007; 13: 5745-5755.
48. Beck-Engeser GB, Lum AM, Huppi K, Caplen NJ, Wang BB, Wabl M. Pvt1-encoded microRNAs in oncogenesis. Retrovirology 2008; 5: 4.
49. Filipowicz W, Bhattacharyya SN, Sonenberg N. Mechanisms of post-transcriptional regulation by microRNAs: are the answers in sight? Nat Rev Genet 2008; 9: 102-114.
50. Papagiannakopoulos T, Kosik KS. MicroRNAs: regulators of oncogenesis and stemness. BMC Med 2008; 6: 15.
51. Mallardo M, Poltronieri P, D'Urso OF. Non-protein coding RNA biomarkers and differential expression in cancers: a review. J Exp Clin Cancer Res 2008; 27: 19.
52. Schickel R, Boyerinas B, Park SM, Peter ME. MicroRNAs: key players in the immune system, differentiation, tumorigenesis and cell death. Oncogene 2008; 27: 5959-5974.
53. Slack FJ, Weidhaas JB. MicroRNA in Cancer Prognosis. N Engl J Med 2008; 359: 2720-2722.
54. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP, Lee W, Mendenhall E, O'Donovan A, Presser A, Russ C, Xie X, Meissner A, Wernig M, Jaenisch R, Nusbaum C, Lander ES, Bernstein BE. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 2007; 448: 553-560.
55. Bond GL, Hu W, Bond EE, Robins H, Lutzker SG, Arva NC, Bargonetti J, Bartel F, Taubert H, Wuerl P, Onel K, Yip L, Hwang SJ, Strong LC, Lozano G, Levine AJ. A single nucleotide polymorphism in the MDM2 promoter attenuates the p53 tumor suppressor pathway and accelerates tumor formation in humans. Cell 2004; 119: 591-602.
56. Jothi R, Cuddapah S, Barski A, Cui K, Zhao K. Genome-wide identification of in vivo protein-DNA binding sites from ChIP-Seq data. Nucleic Acids Res 2008; 36: 5221-5231.
57. Fejes AP, Robertson G, Bilenky M, Varhol R, Bainbridge M, Jones SJ. FindPeaks 3.1: a tool for identifying areas of enrichment from massively parallel short-read sequencing technology. Bioinformatics 2008; 24: 1729-1730.

(Accepted for publication, Dec. 24, 2008)