Supplemental Data

Expanded Version of Methods

STAT1 Chromatin Immunoprecipitations

HeLaS3 cells (ATCC) were cultured in suspension in spinner flasks in S-MEM (GIBCO-Invitrogen #11380-037) supplemented with 2 mM L-glutamine (GIBCO-Invitrogen #25030-081), 10% fetal bovine serum (GIBCO-Invitrogen #16140-071), and antibiotics (Antibiotic-Antimycotic, GIBCO-Invitrogen #15240-062) at 37°C and 5% CO_2. ChIP samples for a given biological replicate were prepared from distinct cell cultures grown, harvested and processed on separate days from all other biological replicates. For each biological replicate we grew on the order of 12 x10^8 cells which were split into IFN-γ treated and untreated halves for STAT1 ChIPs. These sample preparations yielded enough DNA to be distributed across many of the platforms and ChIP-PCR validations. STAT1 ChIP samples were prepared from IFN-γ stimulated HeLaS3 cells and ChIP DNA quality was verified as previously described (Hartman et al. 2005). HeLaS3 cultures were divided in half and were either induced with 5 ng/ml human recombinant IFN-γ (R&D Systems #285-IF), or left untreated, for 30 min at 37°C, 5% CO_2 and then fixed with 1% formaldehyde final concentration at room temperature for 10 min. Fixations were quenched by addition of glycine to 125 mM final concentration (from 2 M glycine stock in 1 x PBS) and cells were washed twice in cold 1 x Dulbecco's PBS (GIBCO-Invitrogen #14190-144). Cells were swelled for 10 min in hypotonic lysis buffer (20 mM Hepes, pH 7.9, 10 mM KCl, 1 mM EDTA, pH 8, 10% glycerol, 1 mM DTT, 0.5 mM PMSF, 0.1 mM sodium orthovanadate, and Roche protease inhibitors #11-697-498-001) and lysed by dounce homogenization (using pestle B). Nuclear pellets were collected and lysed in 1 x RIPA buffer (10 mM Tris-Cl, pH 8.0, 140 mM NaCl, 1% Triton X-100, 0.1% SDS, 1% deoxycholic acid, 0.5 mM PMSF, 1 mM DTT, 0.1 mM sodium orthovanadate, and Roche protease inhibitors). Nuclear lysates were sonicated with a Branson 250 Sonifier (Output 20%, 100% duty cycle) to shear the chromatin to approximately 1 kb in size. Clarified lysates were incubated overnight at 4°C with anti-STAT1 alpha p91 (C-24) rabbit polyclonal antibody (Santa Cruz Biotechnology #sc-345). Protein-DNA complexes were precipitated with RIPA-equilibrated protein A agarose beads (Upstate #16-156) and immunoprecipitates were washed three times in 1 x RIPA, once in 1 x PBS, and then eluted from the beads by addition of 1% SDS, 1 x TE (10 mM Tris-Cl at pH 7.6, 1 mM EDTA at pH 8), and incubation for 10 min at 65°C. Crosslinks were reversed overnight at 65°C. All samples were purified by treatment first with 200 μg/ml RNase A (Qiagen #19101) for 1 h at 37°C, then with 200 μg/ml Proteinase K (Ambion #2548) for 2 h at 45°C, followed by extraction with phenol:chloroform:isoamyl alcohol and precipitation at -70°C with 0.1 volume of 3 M sodium acetate, 2 volumes of 100% ethanol and 1.5 μl of pellet paint co-precipitant (Novagen #69049-3). ChIP DNA prepared from 1 x 10^8 cells was resuspended in 50 μl of ultrapure water (GIBCO-Invitrogen #10977-015).

ChIP sample preparation and labeling

Biological replicates are defined as STAT1 ChIP DNA prepared from distinct cell cultures grown, harvested and processed on separate days. ChIP DNA samples from individual biological replicates were labeled separately and hybridized separately (without pooling) as one biological replicate per array (Supplemental Table 1). In many cases the same biological replicates were hybridized to each of the array platforms. For the experiment comparing hybridizations in the presence and absence of Cot-1 DNA, 6 biological replicates were divided after labeling and hybridized over 12 arrays in plus and minus Cot sets.

For PCR product arrays (gift of Bing Ren, UCSD) and maskless arrays with 50 b every 50 b and 36 b every 36 b spacings (both oligo length arrays manufactured by NASA Ames Research Center), ChIP DNA from 1 x 10^8 cells was random primed with Klenow (enzyme and primers from BioPrime DNA Labeling System, Invitrogen #18094-011) and

Aminoallyl-dUTP (Sigma #A0410) was incorporated. Next Alexa Fluor dyes (Invitrogen #A32755; Alexa647 for ChIP DNA isolated from IFN γ-stimulated cells and Alexa555 for ChIP DNA isolated from unstimulated cells) were coupled to the Aminoallyl-dUTP. Coupling reactions were terminated with hydroxylamine. Alexa555- and Alexa647-coupled ChIP DNA samples were combined and recovered using a CyScribe GFX Purification Kit (Amersham #27-9606-02) according to the manufacturer's protocol. The recovered probe was further purified by ethanol precipitation with 0.1 volume of 3M sodium acetate (pH 5.2).

During the course of our studies we tested a number of different labeling technologies including the MICROMAX tyramide signal amplification method (NEN Life Science Products), 3DNA dendrimer technology (Genisphere) and anti-biotin and anti-fluorescein coated Resonance Light Scattering (RLS) particles (Genicon Sciences). These were tested primarily using PCR product arrays. For detection of STAT1 targets, the labeling methods reported here were the most consistently positive in terms of signal, array uniformity, reproducibility, time efficiency and cost effectiveness.

For maskless arrays (Nuwaysir et al. 2002) with 50 b every 38 b spacing (NimbleGen Systems of Iceland, LLC) ChIP DNA from 1 x 10^8 cells was directly labeled (per manufacturer's protocol) by Klenow random priming with Cy5 nonamers (ChIP DNA isolated from IFN γ-stimulated cells) or Cy3 nonamers (ChIP DNA isolated from unstimulated cells).

Microarray hybridizations

All arrays were hybridized with mixing in MAUI hybridization stations from BioMicro Systems (Salt Lake City, UT) for 16–18 h at 42°C. Before deciding on the hybridization protocols described below we tested a number of experimental parameters. The oligonucleotide arrays were hybridized at temperatures ranging from 14 to 33°C below their estimated melting temperatures (T_m), using the formula described in (Sambrook et al. 1989) and assuming 44% GC content for the ENCODE tiling arrays. Hybridization buffers varied from 0.825 to 1.0 M [Na^+] and from 0–40% formamide, final concentrations. Note that optimal hybridizations are performed at ~25°C below the estimated melting temperatures although hybridization rates are only modestly affected by conditions 15–30°C below the T_m (Wetmur and Davidson 1968).

PCR product arrays were prehybridized in 5x SSC/ 25% formamide/ 0.05% SDS/1% BSA for 1 h at 42°C. Labeled ChIP DNA was precipitated and resuspended in 60 μl of 5x SSC/ 25% formamide/ 0.05% SDS with 5 μg of human Cot-1 DNA (Invitrogen #15279-011) per array. The PCR product arrays were washed in 42°C 2x SSC/0.1% SDS, room temperature 0.1x SSC/0.1% SDS, and 0.1x SSC.

Labeled ChIP DNA for maskless arrays (Nuwaysir et al. 2002) with 50 b every 50 b and 36 b every 36 b spacings (both oligo length arrays manufactured by NASA Ames Research Center) was precipitated with 30 μg of human Cot-1 DNA (Invitrogen #15279-011) per array and pellets were resuspended in 45 μl of hybridization buffer (final concentrations: 40% formamide, 5x SSC, 0.1% SDS, and 0.2x TE). Arrays were washed once with 42°C 0.2% SDS/0.2x SSC, once with room temperature NSWB (6x SSPE, 0.01% Tween-20, 1 mM DTT), twice with 0.2x SSC, and twice with 0.05x SSC.

For maskless arrays (Nuwaysir et al. 2002) with 50 b every 38 b spacing (NimbleGen Systems of Iceland, LLC) labeled ChIP DNA was hybridized in buffer containing 20% formamide, 1.2 M Betaine, and 0.1 μg/μl herring sperm DNA per manufacturer's protocol. The plus Cot-1 experiments included 10 μg of human Cot-1 DNA (Invitrogen

#15279-011) per array. Arrays were washed in 42°C 0.2% SDS/0.2x SSC, room temperature 0.2x SSC, and 0.05x SSC.

ChIP-PET experiment

The STAT1 ChIP-PET library was constructed as previously described (Wei et al. 2006). Briefly, the ChIP enriched DNA fragments were cloned into the cloning vector pGIS3 to generate the ChIP DNA library. Purified plasmid from the ChIP DNA library was digested with MmeI to release the internal fragments and a signature tag from each terminal of the original ChIP DNA insert were self-ligated to form a 'single-ditag library'. 50 bp paired-end-ditags (PETs) were released by BamHI, PAGE-purified and concatenated to clone into pZErO-1 to form the final ChIP-PET library for sequencing.

PET sequences were extracted from the raw reads and mapped to human genome sequence assembly [hg17]. The process of PET extraction and mapping is essentially the same as previously described for cDNA analysis (Ng et al. 2005). The mapping criteria are that both the 5′ and 3′ signatures must have a minimal 17 bp match, be present on the same chromosome and same strand, in the correct orientation (5′→3′), and within 6 kb of genomic distance.

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 4007 unique genomic DNA segments from 44 ENCODE regions (similar to the average fragment size from STAT1 ChIP DNA) and then determined how many fragments overlapped with others. This process was repeated 10,000 times to compute the percentage of randomly selected DNA fragments that overlapped. The results are summarized in Table 4. Based on this simulation, we estimated that 463 PETs (97% of total) would result in two overlapping PETs (PET-2), 47 in PET-3, 3 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 immunoprecipitation of STAT1-associated DNA rather than from random events.

STAT1 Target Validations

Primers were designed to amplify 200-350 bp fragments from regions throughout the rank ordered target lists as well as regions where array signals were below cut-off values. ChIP DNA from either 4 x10^6 IFN γ-stimulated or unstimulated cells was amplified in 40 μl reactions with 1 μM of target specific primer pairs and 1 x Qiagen Master Mix (Qiagen # 201203). For each primer pair parallel reactions were run with 0.2 μg HeLaS3 genomic DNA to ensure that a sample set would yield a single band of the expected size. Some primer pairs required addition of PCR additives, either Betaine or Qiagen Q solution at varying concentrations. Cycling conditions were as follows: 5 min at 94°C, 29 cycles of 30 sec at 94°C, 30 sec at 52°C, 30 sec at 72°C, and a final extension period of 10 min at 72°C. The entire completed PCR reactions were loaded on 1.5% agarose gels and only those primer sets in which entire sample volumes were loaded were analyzed further. Each plate of PCR reactions included positive and negative controls, and all reactions from a plate were loaded on the same gel. Densitometric analyses were made using ImageJ software <http://rsb.info.nih.gov/ij/>. For each primer pair, enrichments were calculated for yield from IFN γ-stimulated cells relative to yield from unstimulated cells. To qualify as a validated region, enrichments had to be consistently greater than 2-fold from each of two or more biological replicates. In many cases more than two biological replicates were tested and for some regions validation results were quantified from multiple primer pairs (in separate reactions) to eliminate any primer artifacts. In total 280 regions were tested for validation. Primer sequences used in the ChIP-PCR assays are available at <http://encode.gersteinlab.org/data/Euskirchen_etal/>.

Design of Genomic Tiling Microarrays

All tiling arrays were designed using the sequence from the ENCODE regions based on human genome build NCBIv34 [hg16]. For analysis coordinates from all array designs were remapped on to human genome build NCBIv35 [hg17] using liftOver from the UCSC Genome Browser (Hinrichs et al. 2006). The 50 b every 50 b tiling array was custom designed with 192,040 50 b oligoneucleotides tiling the forward strand approximately every 50 b (end-to-end) across the following ENCODE regions; ENm001-014, ENr114, ENr132, ENr233, ENr321, ENr331 and ENr333. The 36 b every 36 b tiling array was also custom designed using 382,454 36 b oligonucleotides tiling one strand approximately every 36 b (end-to-end) across all the ENCODE regions excluding ENr112, ENr121, ENr131, ENr211, ENr222, ENr313, ENr324, ENr334. Both of these arrays were designed using the tiling array design tool http://tiling.gersteinlab.org (Bertone et al. 2006). The 50 b every 38 b array uses 382,885 50 b oligonucleotides to tile the forward strand of all ENCODE regions with average overlap between oligonucleotides of 38 b. The PCR product array (supplied by Bing Ren, UCSD) uses 24,341 PCR amplicons of average size 620 bp to tile all of the nonrepetitive ENCODE regions. Relative coverage of the ENCODE regions for each of the array formats is shown in Supplemental Table 1.

Analysis of Microarray Data

For each hybridization the files (in .pair file format) to the two channels corresponding to the ChIP DNA and reference DNA, were uploaded to the TileScope pipeline for high-density tiling array data analysis <http://tilescope.gersteinlab.org> for normalization and scoring (Zhang et al. submitted 2006). The pipeline first performs intra- and inter-slide scaling (between biological replicates) using quantile normalization, the results of which are then integrated using a sliding window approach (a window of size 1000 bp in genomic space was used to integrate neighboring probes from replicate arrays). For each window centered at the genomic coordinate of each oligonucleotide probe, the pseudomedian signal (median of pairwise averages of the $\log_2$ ratio of test to reference signals for all oligonucleotide probes within the window), as well as a p-value measuring the likelihood that the region is bound by the transcription factor (using a Wilcoxon paired signed rank test comparing test signal against reference signal for all oligonucleotide probes in the window) are computed. In an iterative fashion regions with the highest signal (and p-values less than 10^{-4}) were selected. In order to ensure that the shoulders of other target regions are not identified as distinct binding regions we required that the centers of target regions be spaced at least 1300 bp apart. This procedure generated a ranked list of non-overlapping target regions of size 1300bp. For each set of arrays the top 200 regions were scored in this fashion where possible. (For the 36 b array dataset we were only able to extend the rank list to the top 39 targets, beyond which the regions did not show enriched signal at the statistically significant cutoff used). The data from the PCR product arrays were analyzed with the on-line microarray processing tool ExpressYourself (Luscombe et al. 2003). This microarray data series is available at the Gene Expression Omnibus <http://www.ncbi.nlm.nih.gov/geo/> with accession number GSE2714. The ranked target lists are available at both <http://dart.gersteinlab.org/> as well as at <http://encode.gersteinlab.org/data/Euskirchen_etal/>.

Comparison of Target Lists

As described above target lists are rank lists of non-overlapping target regions of uniform size 1300 bp. In order to fairly compare the ChIP-chip data against the ChIP-PET experiment, the ChIP-PET targets were likewise converted into 1300 bp regions centered on the ChIP-PET cluster. Also, comparisons were done for targets identified in regions common to both platforms because the 50 b every 50 b array does not tile all of the

ENCODE regions (Supplemental Table 1). When computing the overlap between any two lists of regions (whether the data are from ChIP-chip or ChIP-PET), the number of entries in the first list intersecting the second is not necessarily the same as the number of the second list intersecting the first (this discrepancy typically happens in loci where multiple target sites are located in a short genomic span). In order to avoid this ambiguity we chose to first merge the two lists under comparison to form a list of union regions comprising the union of targets from both lists. Then using the union set of regions as a basis, one can compute the number of regions belonging to only one of the two original lists, or union regions that came from both lists. One important note is that some union target regions occurring in more complicated loci tend to be longer and might only contribute one joint region to the counts of number of union regions shared by both lists, even though the region might correspond to multiple entries on each of the original two lists. Regions that have been tested for validation can also be compared against these union target regions to assess validation rates for union regions that were detected on only one of the two lists or by both datasets. This is how the data in Tables 1, 2, 3, 6, S2 and S3, were generated.

References for Supplemental Methods

Hinrichs, A.S., Karolchik, D., Baertsch, R., Barber, G.P., Bejerano, G., Clawson, H., Diekhans,M., Furey, T.S., Harte, R.A., Hsu, F., et al. 2006. The UCSC Genome Browser database: update 2006. Nucleic Acids Res. 34 (Database issue): D590-D598.

Luscombe, N.M., Royce, T.E., Bertone, P., Echols, N., Horak, C.E., Chang, J.T., Snyder, M., and Gerstein, M. 2003. ExpressYourself: A modular platform for processing and visualizing microarray data. Nucleic Acids Res. 31: 3477–3482.

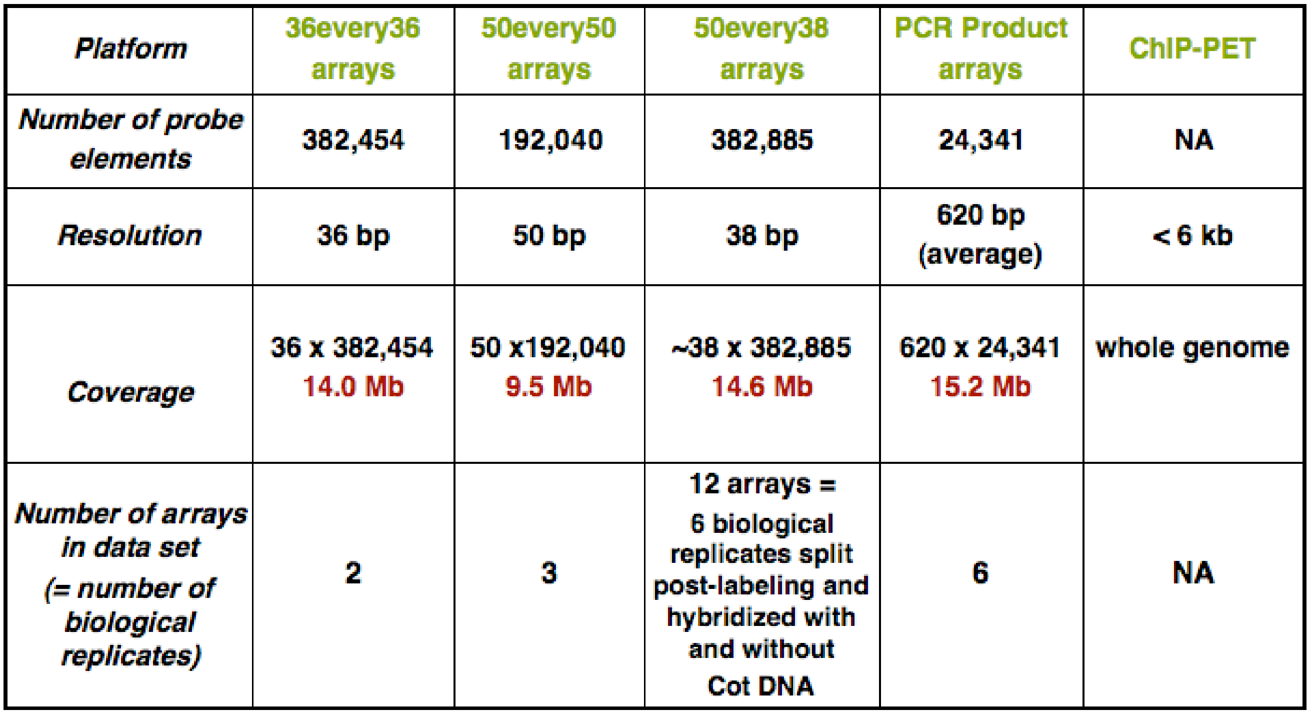
Nuwaysir, E.F., Huang, W., Albert, T.J., Singh, J., Nuwaysir, K., Pitas, A., Richmond, T., Gorski, T., Berg, J.P., Ballin, J., et al. 2002. Gene expression analysis using oligonucleotide arrays produced by maskless photolithography. Genome Res. 12: 1749–1755.

Sambrook, J., Fritsch, E.F., and Maniatis, T. 1989. *Molecular cloning: A laboratory manual,* 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Wetmur, J.G., and Davidson, N. 1968. Kinetics of renaturation of DNA. J. Mol. Biol. 31: 349-370.

Zhang, Z., Rozowsky, J., Lam, H., Snyder, M., and Gerstein, M. 2006. Tilescope: online analysis pipeline for high-density tiling microarray data. Submitted.

Supplemental Table 1
Summary of data sets used for the analyses presented in platform comparisons



Platform	36every36 arrays	50every50 arrays	50every38 arrays	PCR Product arrays	ChIP-PET
Number of probe elements	382,454	192,040	382,885	24,341	NA
Resolution	36 bp	50 bp	38 bp	620 bp (average)	< 6 kb
Coverage	36 x 382,454 14.0 Mb	50 x192,040 9.5 Mb	~38 x 382,885 14.6 Mb	620 x 24,341 15.2 Mb	whole genome
Number of arrays in data set *(= number of biological replicates)*	2	3	12 arrays = 6 biological replicates split post-labeling and hybridized with and without Cot DNA	6	NA

Supplemental Table 2 Comparison of ranked target lists between the 50 b every 50 b and the 36 b every 36 b array platforms

A. False Positive Rates of the ranked target lists considered separately

	50 every 50 dataset	36 every 36 dataset	Union
Count	Top 75	Top 25	74
FPR	0.26	0.38	

B. False Positive Rates of the merged ranked target lists from Suppl. Table 2A

	Specific to 50 every 50 set	Specific to 36 every 36 set	Common to both datasets
Count	55	5	14
Positives (ChIP-PCR validation)	22	0	6
Negatives (ChIP-PCR validation)	8	2	2
FPR	0.27 (= 8/30)	1.00 (= 2/2)	0.25 (= 2/8)

Supplemental Table 3 Comparison of ranked target lists between the 50 b every 50 b and the PCR product array platforms

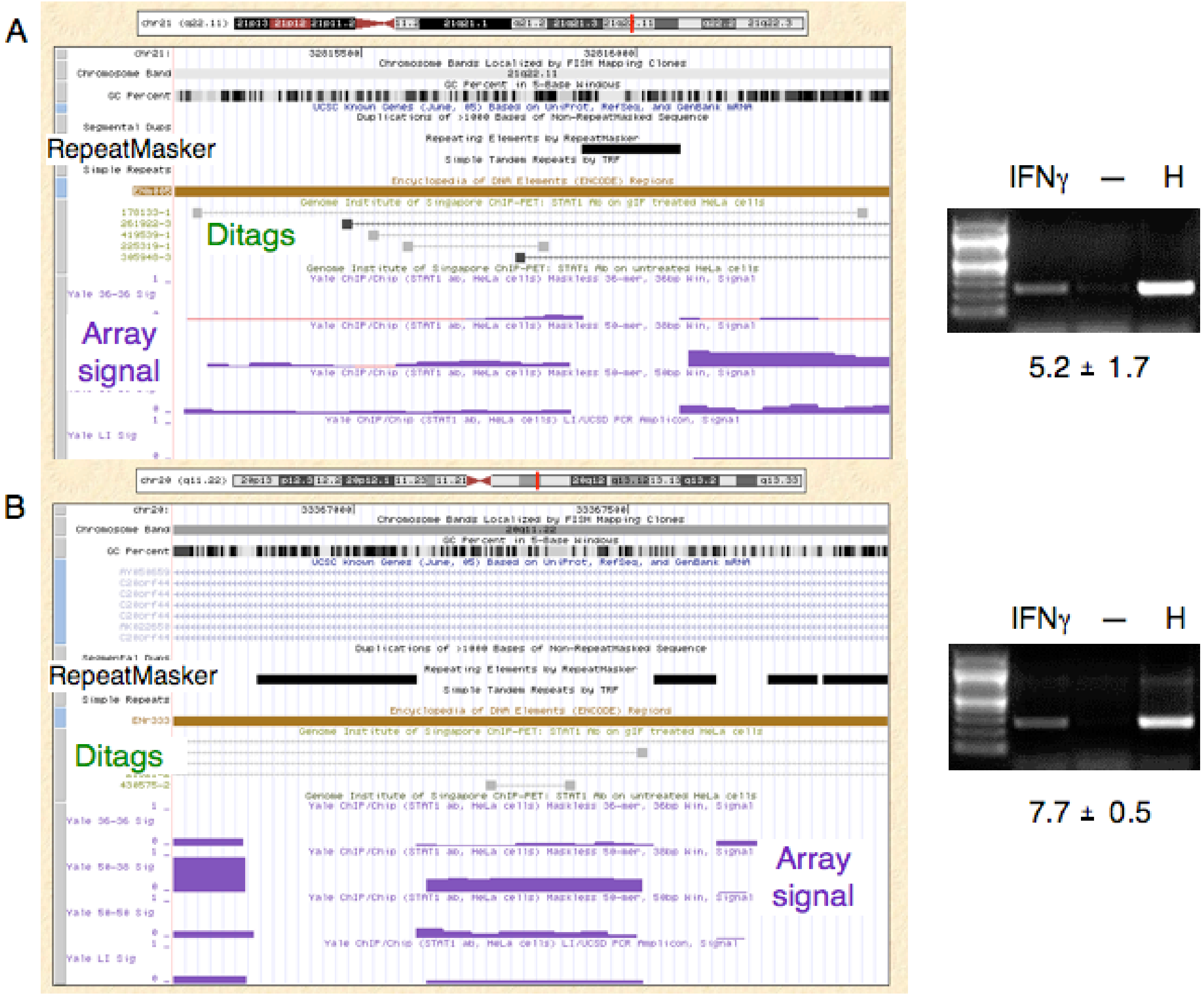
A. False Positive Rates of the ranked target lists considered separately

	50 every 50 dataset	PCR product dataset	Union
Count	75	33	93
FPR	0.26	0.40	

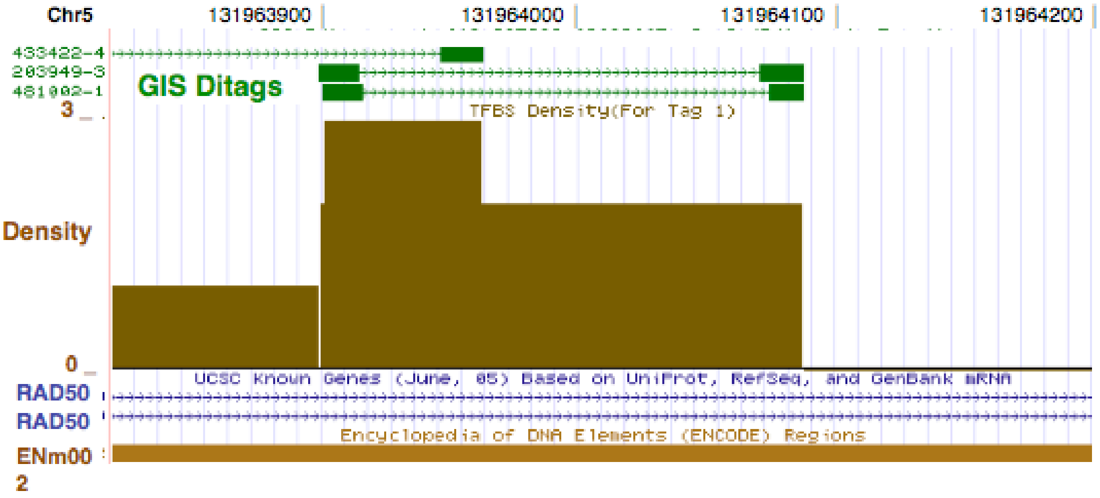
B. False Positive Rates of the merged ranked target lists from Suppl. Table 3A

	Specific to 50 every 50 set	Specific to PCR product set	Common to both datasets
Count	65	22	6
Positives (ChIP-PCR validation)	21	7	6
Negatives (ChIP-PCR validation)	11	12	0
FPR	0.34 (= 11/32)	0.63 (= 12/19)	0.00 (= 0/6)

Supplemental Figure 1 The two PET-5 regions detected by ChIP-PET and validated by ChIP-PCR analysis that were undetected by ChIP-chip with the 50 b every 50 b platform



Supplemental Figure 2
Example of a ChIP-PET-3 cluster that was reassigned to a ChIP-PET-2 cluster



Supplemental Table 1

Summary of datasets used for the analyses presented in platform comparisons. For each array platform the number of features, oligonucleotide length, genomic coverage and number of biological replicates performed is listed. Biological replicates are defined as STAT1 ChIP DNA prepared from distinct cell cultures grown, harvested and processed on separate days. ChIP DNA samples from individual biological replicates were labeled separately and hybridized separately (without pooling) as one biological replicate per array. In many cases the same biological replicates were hybridized to each of the array platforms.

Supplemental Table 2

Similarly to Table 1 the target lists from the 50 b every 50 b and the 36 b every 36 b array datasets are compared, but with the target list restricted to only the top 25 targets for the 36 b arrays in order to compare lists of higher accuracy. Again the upper panel displays the false positive rates (FPR) calculated for each list considered separately. The lower panel displays the results after merging the list of the top 75 targets from the 50 b arrays with the list of the top 25 targets from the 36 b arrays.

Supplemental Table 3

Similarly to Table 1 the target lists from the 50 b every 50 b and the PCR product array datasets are compared, but with the target list restricted to only the top 33 targets for the PCR product arrays in order to compare lists of higher accuracy. Again the upper panel displays the false positive rates (FPR) calculated for each list considered separately. The lower panel displays the results after merging the list of the top 75 targets from the 50 b arrays with the list of top 33 targets from the PCR product arrays.

Supplemental Table 4

Primer pairs used to validate regions sampled across the various ranked targets lists are available at <http://encode.gersteinlab.org/data/Euskirchen_etal/>. The coordinates listed are based on human genome build NCBIv35 [hg17]. The PCR product sizes shown are the results of In Silico PCR run at <http://genome.ucsc.edu/cgi-bin/hgPcr>.

Supplemental Figure 1

Genomic features of the two PET-5 regions that were not detected in the 50 b every 50 b ChIP-chip dataset (Table 5). These regions validated by ChIP-PCR (gel images shown on the right). **A.** Chromosome 21 between coordinates 32,815,161 and 32,816,460 [hg17]. **B**. Chromosome 20 between coordinates 33,366,674 and 33,367,973 [hg17]. For both chromosomal regions, sizable lengths of repetitive sequence (as identified by RepeatMasker) coincided with the PET overlap spans, thus likely impairing the ability of the arrays to detect these targets due to decreased probe density. The ChIP-PCR lanes are labeled for ChIP DNA from IFN-γ stimulated cells, ChIP DNA from unstimulated cells and for HeLaS3 genomic DNA. Fold enrichments as calculated for several biological replicates (see Materials and Methods) are indicated.

Supplemental Figure 2

Example of a ChIP-PET-3 cluster that was reassigned to a ChIP-PET-2 cluster. This target in the region chr5:131963298-131964597 [hg17] could not be confirmed by ChIP-PCR analysis. Closer inspection revealed an unusual instance of 2 overlapping PETs that have an almost identical mapping (with 2 bp difference) and were likely derived from the same ChIP fragment.