GREAT, a functional enrichment approach and tool for interpretation of genome-wide *cis*-regulatory datasets

Cory Y. McLean[1], Dave Bristor[1,2], Michael Hiller[2], Shoa L. Clarke[3],
Bruce T. Schaar[2], Craig B. Lowe[4], Aaron M. Wenger[1], and Gill Bejerano[1,2]

[1]Department of Computer Science, Stanford University, Stanford, CA 94305, USA
[2]Department of Developmental Biology, Stanford University, Stanford, CA 94305, USA
[3]Department of Genetics, Stanford University, Stanford, CA 94305, USA
[4]Center for Biomolecular Science and Engineering,
University of California Santa Cruz, Santa Cruz, CA 95064, USA
Correspondence should be addressed to G.B. (bejerano@stanford.edu).

04/01/10

Contents

SUPPLEMENTARY NOTES 2
SUPPLEMENTARY REFERENCES 10
SUPPLEMENTARY FIGURES 13
SUPPLEMENTARY TABLES 17

SUPPLEMENTARY NOTES

Ontologies supported

GREAT assimilates knowledge from 20 separate ontologies containing biological knowledge about gene functions, phenotype and disease associations, biological pathways, gene expression data, presence of regulatory motifs, and gene families (**Supplementary Table 1**). Statistics for each ontology list the total number of terms in the ontology that are currently tested by GREAT, the number of genes annotated with one or more terms in the ontology, and the number of direct associations between ontology terms and genes (**Supplementary Tables 2** and **3**). Some ontologies contain parent/child relationships between terms expressed as a directed acyclic graph; general terms within these ontologies inherit genes that are only labeled with more specific child terms as indirect associations. To increase statistical power (by reducing the multiple hypothesis correction factor), GREAT does not test any general term whose associated gene list is identical to the associated gene list of a more specific child term. The following ontologies are currently used:

Gene Ontology

The *Gene Ontology* (GO; `http://www.geneontology.org/`) provides a controlled vocabulary to describe attributes of gene products[1]. GO contains three separate ontologies that describe molecular functions, biological processes, and cellular components of proteins.

Mouse Phenotype

The Mouse Genome Informatics (MGI) resource contains data about mouse genotype–phenotype associations primarily obtained via literature curation[2,3] (`http://www.informatics.jax.org/phenotypes.shtml`). Phenotypic terms are canonicalized and relationships between terms are enumerated in the Mammalian Phenotype Ontology[4].

MSigDB Ontologies

The Molecular Signatures Database (MSigDB; `http://www.broad.mit.edu/gsea/msigdb/`) contains a collection of gene sets[5]. The following description of the various ontologies within MSigDB is taken from `http://www.broad.mit.edu/gsea/msigdb/collections.jsp`.

- *MSigDB Cancer Neighborhood*
  Computational gene sets defined by mining large collections of cancer-oriented microarray data. Gene sets defined by expression neighborhoods centered on 380 cancer-associated genes[6]. This collection is identical to that previously reported in[5].

- *MSigDB Cancer Modules*
  Computational gene sets defined by mining large collections of cancer-oriented microarray data[7]. Briefly, the authors compiled gene sets ("modules") from a variety of resources such as KEGG, GO, and others. By mining a large compendium of cancer-related microarray data, they identified 456 such modules as significantly changed in a variety of cancer conditions.

- *MSigDB Pathway*
  Gene sets from pathway databases. Usually, these gene sets are canonical representations of a biological process compiled by domain experts.

- *MSigDB Perturbation*
  Gene sets that represent gene expression signatures of genetic and chemical perturbations.

- *MSigDB Predicted Promoter Motifs*
  Sets of genes that share a transcription factor binding site defined in the TRANSFAC (version 7.4, `http://www.gene-regulation.com/`) database. Each of these gene sets is annotated by a TRANSFAC record.

- *MSigDB miRNA Motifs*
  Sets of genes that share a 3'-UTR microRNA binding motif.

PANTHER Pathway

PANTHER Pathway (`http://www.pantherdb.org/pathway/`) contains information on biological pathways (primarily signaling pathways)[8]. PANTHER pathways are collections of biological molecules and the reactions in which they participate. Only well-documented reactions and relationships are listed.

Pathway Commons

Pathway Commons[9] contains a comprehensive collection of pathways from multiple sources listed at `http://www.pathwaycommons.org/pc/`. According to the website, "[p]athways include biochemical reactions, complex assembly, transport and catalysis events, and physical interactions involving proteins, DNA, RNA, small molecules and complexes."

BioCyc Pathway

BioCyc (`http://biocyc.org/`) contains information linking genes to the metabolic pathways in which they participate[10].

MGI Expression: Detected & Not Detected

The Gene Expression Database (GXD, `http://www.informatics.jax.org/expression.shtml`), a part of the Mouse Genome Informatics database, contains expression data with a focus on gene expression during mouse development[3,11]. The information is primarily obtained from the literature via manual curation. Each entry gives the expression in a specific anatomical structure during a specific developmental period or "Theiler stage"[12]. The anatomy for each developmental stage is represented by a directed acyclic graph that gives a hierarchy of anatomical terms and their relationships. The database contains information about which genes are expressed and which are not found to be expressed.

We represent the MGI Gene Expression Database by several sub-ontologies, where each sub-ontology is specific to a developmental stage. We then combine all sub-ontologies into one ontology so that all developmental stages are tested at once. *MGI Expression: Detected* contains data about genes that are expressed and *MGI Expression: Not Detected* contains data about genes whose expression is measured but not experimentally detected.

The human *MGI Expression* ontologies are derived by mapping expression information from all genes in the mouse ontologies to their human orthologs and assume large-scale conservation of expression patterns.

Transcription Factor Targets

The *Transcription Factor Targets* ontology contains transcription factor (TF) target sets for human and mouse collected from literature[13]. Most TF target genes were identified by ChIP-chip experiments (see `http://acgt.cs.tau.ac.il/amadeus/suppl/metazoan_compendium.htm`).

miRNA Targets

The *miRNA Targets* ontology contains miRNA target sets for human and mouse collected from literature[13]. miRNA target genes were identified as genes downregulated after miRNA overexpression (see `http://acgt.cs.tau.ac.il/amadeus/suppl/metazoan_compendium.htm`).

InterPro

InterPro (`http://www.ebi.ac.uk/interpro/`) is a database of protein domains, families and functional sites[14]. InterPro annotations give information about the function, structure and evolution of the domains. InterPro combines data from several other databases (PROSITE, PRINTS, Pfam, ProDom, SMART, TIGRFAMs, PIRSF, SUPERFAMILY, PANTHER and Gene3D).

TreeFam

The Tree families database (TreeFam, `http://www.treefam.org/`) contains information about the evolutionary history (both orthologs and paralogs) of gene families[15]. A gene family is defined as "a group of genes that evolved after the speciation of single-metazoan animals". We format this data into an ontology by creating an ontology term for each gene family and then associating each gene within the species (human or mouse) with its gene family.

HGNC Gene Families

The HUGO Gene Nomenclature Committee groups genes into gene families based on sequence similarity, data from the literature and other databases, and manual curation[16]. The groupings are listed at `http://www.genenames.org/genefamily.html`.

Website architecture

The GREAT website output is generated by way of server-side PHP code invoking a Python wrapper over a C program. The C code makes use of the UCSC Genome Browser[17] code libraries and performs the core calculations. Test results are formatted for web display using Python and PHP code. GREAT uses the DataTable control from the Yahoo! User Interface Library (`http://developer.yahoo.com/yui/`) and custom JavaScript code to allow many user operations without need for more server data. This allows rapid responses for all filtering requests once the initial results page has loaded locally. Operations which generate new pages, such as getting details for a single ontology term or the generation of publication quality table display do involve return trips to the server, which are handled by PHP code.

Graphical User Interface

The graphical user interface (GUI) of GREAT version 1.1.3 has the following components:

- *User Input*
  The user input page (**Supplementary Fig. 4a**) requires two inputs from the user: the organism genome assembly in which the analysis should be performed and the *cis*-regulatory regions to analyze in BED format[17]. Optionally, the user can upload a background set of genomic regions to test against rather than the whole genome. The user can also optionally alter the association rules between *cis*-regulatory regions and their putative target genes from the default *basal plus extension* rule via the Advanced options tab.

- *Global Output & Controls*
  Upon submission of a dataset, users are directed to the global output screen of GREAT (**Supplementary Fig. 4b**). By default, only ontology terms significant by both the binomial and hypergeometric tests using the multiple hypothesis correction false discovery rate (FDR)[18] ≤ 0.05 whose binomial fold enrichment is at least 2.0 are displayed. The extensive global controls at the top of the page allow users to change the ontologies shown, alter the data columns displayed for each result term, and change the multiple test correction type and threshold. The number of enriched terms shown for each ontology, the display of terms not significant by one or both of the enrichment tests, and the filtering of terms by their descriptions can all be changed via the global controls.

  Within each ontology table, terms can be sorted by any data column except FDR. The data for a single table can be downloaded either as HTML or as a tab-separated file. By clicking on a particular term, additional information is presented in an individual term page (described below).

- *Individual Term Page*
  Each ontology term tested for enrichment has an associated individual term page (**Supplementary Fig. 4c**). The page lists all *cis*-regulatory regions that reside within the regulatory domain of any gene annotated with the term, all the genes annotated with the term that possess one or more input regions within its regulatory domain, and the definition of the ontology term from the source website inset into

the page. By clicking on any *cis*-regulatory region listed, users can navigate to the UCSC Genome Browser with custom tracks of the entire input set and only the elements contributing to the specific term enrichment (**Supplementary Fig. 4d**).

- *Online Documentation*
  User help documentation is available at `http://great.stanford.edu/help` and includes information regarding all aspects of GREAT.

- *Demo Sets*
  The SRF[19] and limb p300 (ref. 20) datasets presented in the main text, as well as several additional sets, are available as demonstration input sets for GREAT. These sets can be tested by navigating to the "Demo" link.

Comparisons of GREAT to gene-based analyses of additional ChIP-Seq studies

To assess the ability of GREAT to improve upon existing gene-based analyses of ChIP-Seq datasets, we compared GREAT enrichments to gene-based tool enrichments for multiple datasets. Where they were not performed by the original authors, we performed gene-based enrichment analyses using the Database for Annotation, Visualization, and Integrated Discovery (DAVID)[21]. Many other gene list-based approaches available (see **Supplementary Table 4** for a partial list) assess statistical significance in generally similar manners (reviewed in ref. 22). We used DAVID as a representative tool for the gene-based enrichment methodology because it is web-based, it integrates many different ontologies including pathways and tissue expression data, and it is a popular choice among manuscript writers. To examine the relative contributions of the integration of distal binding events with the binomial test and the unique set of ontologies supported by GREAT, we also ran "gene-based GREAT" analyses that alter the regulatory domains and significance testing to exactly mimic existing gene-based tools. The analysis of next-generation binding data is not clearly mappable to more advanced techniques like the Gene Set Enrichment Analysis (GSEA)[5] which rank genes by the intensity difference of the different probes/genes and are typically applied to compare gene expression profiles from two classes (e.g. people with a disease vs. healthy controls).

For each ChIP-Seq dataset we analyzed using DAVID, we mapped the identified peaks that reside within 2 kb of the TSS of the nearest gene as identified by the UCSC Known Genes track[23] to the nearby gene and ran enrichments over the resulting gene list. The ten most enriched terms from each annotation cluster reported significant by DAVID at a threshold of 0.05 after an FDR multiple hypothesis correction are shown in each enrichment table. To run a "gene-based GREAT", we used the *basal plus extension* association rule with basal upstream and downstream parameters both set to 2 kb and an extension of 0 bp and excluded the set of curated regulatory domains. The ten most enriched terms significant by the hypergeometric test at a threshold of 0.05 after an FDR multiple hypothesis correction are shown in each enrichment table.

Since GREAT performs a *cis*-regulatory element-based test, no mappings from ChIP-Seq peaks to genes are required for preprocessing. For each ChIP-Seq dataset we analyzed using GREAT, we ran the identified peaks through GREAT using its default settings (the *basal plus extension* association rule with basal domains extending 5 kb upstream and 1 kb downstream of the TSS and extension to the basal domains of the nearest genes within 1 Mb). The top ten terms significant at a threshold of 0.05 by the binomial test after an FDR multiple hypothesis correction that have a fold enrichment of at least two and are also significant by the hypergeometric test are shown in each enrichment table.

P300 in mouse developing embryonic tissues

Recent tissue-specific ChIP-Seq experiments identified 2,453, 561, and 2,105 regions of the mouse genome bound by the transcriptional coactivator protein p300 at embryonic day 11.5 in forebrain, midbrain, and limb tissues, respectively[20]. Assays for enhancer activity in transgenic mice at embryonic day 11.5 showed that many p300-bound regions are reproducible enhancers with strong tissue specificity[20].

We ran a gene-based enrichment analysis of the p300 limb peaks using DAVID as described above, yielding the enrichments shown in **Supplementary Table 10a**. This gene-based analysis shows enrichment for p300 binding near transcription factors and hints at regulation of development and morphogenesis. However, no

enrichment for limb tissue expression or developmental stage of enhancer activity is featured, with the closest enrichments being the much broader terms "organ development" and "anatomical structure morphogenesis" (**Supplementary Table 10a**).

On the other hand, when we ran the 2,105 limb p300 ChIP-Seq peaks through GREAT using the 5+1 kb up to 1 Mb default settings, GO enrichments overwhelmingly emphasize limb morphogenesis and development as strongly enriched functions, and GREAT highlights specific functions of the set: while DAVID identifies "transcription" as a prominent term, GREAT finds significant enrichment for "transcription repressor activity", thus markedly narrowing the focus of a much broader term (**Supplementary Table 10b**). Similarly, the *Mouse Phenotype* ontology enrichments of GREAT all include an aspect of skeletal development, contrasting again to the broad morphogenesis terms enriched by DAVID (main text, **Supplementary Table 10**).

Notably, the GREAT enrichments all draw heavily from the appropriate integration of distal binding events; over 75% of the binding peaks that contribute to every GREAT enriched term occur further than 10 kb from the TSS of the nearest gene (**Figure 3**, **Supplementary Table 10b**). The importance of distal binding is implicitly shown by the results of a "gene-based GREAT" analysis that only associates peaks within 2 kb of the TSS of a gene (**Supplementary Table 11**). Only two limb-specific terms are identified as enriched, and each implicate 100 to 150 fewer genes than the standard GREAT analysis. The markedly improved enrichments reported by GREAT as compared to gene-based enrichment analyses are a testament to the importance of properly integrating distal regulators into analyses of vertebrate development, and the hypothesis that p300 limb peaks indeed play a role in large-scale regulation of key genes controlling embryonic limb development is strongly supported using GREAT.

We also ran both DAVID and GREAT on the forebrain p300 peaks. As before, the DAVID gene-based enrichments are much more general than GREAT enrichments, with the most prominent DAVID term being "transcription" (**Supplementary Table 15a**). Though DAVID does also highlight "forebrain development", GREAT analysis produces many more details from which to launch experiments to explore forebrain development (**Supplementary Table 15b**). GREAT highlights the regulation of transcription factors as well as mouse phenotypes in axonal tract formation that are affected by defects in early stages of neuronal differentiation (**Supplementary Table 15b**). In particular, GREAT offers enrichment for the basic helix-loop-helix family of transcription factors that are known to play a prominent role in cell fate at this stage of development[24]. Other highly relevant findings of GREAT include the *PANTHER Pathway* enrichment for the Notch signaling pathway and multiple enrichments for the Wnt signaling pathway (**Supplementary Table 15b**). At this stage of forebrain development, the production of postmitotic neurons and proliferative progenitors is indeed tightly regulated by both Notch and Wnt signaling[25,26].

A "gene-based GREAT" analysis identifies forebrain-related terms as enriched, though similarly to the limb enrichments the total number of genes identified as important to the processes is markedly reduced (**Supplementary Table 16**). The difference in enrichment specificity between proximal limb and forebrain peaks suggests that forebrain development may be more specialized than limb development.

When we ran GREAT and DAVID on the 561 midbrain p300 peaks, DAVID failed to yield any significant results. In contrast, GREAT analysis highlighted many terms related to embryonic brain development (**Supplementary Table 20**), including both terms also enriched in the forebrain set and enrichments unique to the midbrain (see main text). Nearly all of the 561 midbrain peaks lie distal to genes: only 28 genes have a midbrain peak within their proximal promoter. Consequently, "gene-based GREAT" identifies only three total enriched terms involving seven total genes (**Supplementary Table 21**).

All three datasets have the majority of their binding peaks occur over 50 kb from the TSS of any gene (**Figure 2a**). Consequently, while the limb, forebrain, and midbrain ontology terms are still enriched in GREAT analyses that only extend regulatory domains up to 50 kb (**Supplementary Tables 12**, **17**, and **22**, respectively), half of the genomic regions and associated genes are lost. The enrichment of limb (**Supplementary Tables 13**, **14**), forebrain (**Supplementary Tables 18**, **19**), and midbrain (**Supplementary Tables 23**, **24**) terms are all robust to variation of the distal association rule used.

P300 in mouse embryonic stem cells

A recent ChIP-Seq analysis of transcription factors involved in the maintenance of the self-renewal and pluripotency capabilities of embryonic stem cells (ESCs) assayed genome-wide binding of p300 within mouse ESCs[27]. The binding profile of p300 was noted to co-localize with the binding profiles of Nanog, Oct4, and Sox2 (ref. 27), which are known to be involved in stem cell maintenance[28].

When we ran DAVID on the associated gene set, no annotation terms were found to be statistically significant (**Supplementary Table 25a**). We then analyzed all 524 identified p300 ChIP-Seq peaks[27] with GREAT using the default settings (5+1 kb basal, up to 1 Mb extension). GO enrichments indicate that chromatin binding and transcriptional regulator proteins are targets of p300 in ESCs, with striking enrichments for genes involved in stem cell maintenance and stem cell differentiation (**Supplementary Table 25b**). The *MGI Expression: Detected* ontology shows enrichment for genes expressed during the very first stages of development[12], consistent with stem cell maintenance. The *Predicted Promoter Motifs* ontology shows enrichment for p300 binding near genes whose promoters contain binding sites for GTF3A and NHLH1. GTF3A, which helps to assemble active chromatin, is required for transcription of the 5S RNA genes that drive growth in early developing embryos[29]. While the significance of NHLH1 binding to stem cell maintenance is not known, NHLH1 is known to be expressed in the developing nervous system[30].

The "gene-based GREAT" results identify stem cell differentiation as an enriched term, but none of the other top enriched terms overlap the enrichments displayed by GREAT (**Supplementary Table 26**). Running GREAT with a more limited extension (5+1 kb basal, up to 50 kb extension) highlights more general terms as enriched and no longer emphasizes enrichment for genes expressed in early development (**Supplementary Table 27**). Distal binding appears to contribute greatly to the function of p300 in embryonic stem cells, as demonstrated by nearly 40% of all binding occurring outside of 50 kb from the TSS of any gene (**Figure 2a**). Variation in distal association rules leads to generally similar enrichments as the default GREAT, emphasizing genes involved in stem cell maintenance and expressed in early development (**Supplementary Tables 28**, **29**).

Signal transducer and activator of transcription 3 (Stat3) in mouse embryonic stem cells

The binding events of Signal transducer and activator of transcription 3 (Stat3) were also assayed within mouse ESCs using ChIP-Seq in the study mentioned above[27]. Stat3 is a transcriptional activator whose activity is sufficient to maintain an undifferentiated state of mouse ESCs[31], but whose constitutive expression has also been linked to various cancers[32]. Stat3 transduces signals from the IL-6 family of cytokine receptors[33]. One of these cytokines, Leukemia Inhibitory Factor (LIF), is a component of media used to culture ESCs in an undifferentiated state.

Gene-based DAVID enrichments for the Stat3 dataset were calculated in the manner described above. The enriched terms from the gene-based analysis are very general, hinting mainly at roles for Stat3 in metabolic processes and regulation (**Supplementary Table 30a**).

Supplementary Table 30b displays the *cis*-regulatory element-based enrichments produced by GREAT for the same set using the default settings (5+1 kb basal, up to 1 Mb extension). In contrast to the generality of DAVID's term enrichments, GREAT produces many highly specific and accurate enrichments and yields novel, testable hypotheses. *GO Biological Process* enrichments indicate Stat3 regulates genes involved in both stem cell maintenance and differentiation. The *Mouse Phenotype* ontology shows enrichment for genes whose alteration leads to embryonic lethality before somite formation and abnormal placental development (**Supplementary Table 30b**). Stat3 is indeed essential; Stat3 knockout mice are embryonic lethal at early stages of development[34]. The *PANTHER Pathway* ontology suggests that Stat3 modulates the Interferon-γ signaling pathway. Interestingly, though Stat3 mediation of the Interferon-γ pathway has yet to be shown, Interferon-γ has recently been shown to suppress Stat3 via dephosphorylation[35]. The *MSigDB Pathway* ontology shows enrichment for genes expressed in breast cancers, especially those involved in estrogen-receptor-dependent signal transduction. STAT3 expression is frequently detected in breast cancer tissues[36], though a clear link between Stat3 and estrogen receptor expression has yet to be shown. The *MSigDB Perturbation* ontology enrichment for genes upregulated after LIF treatment highlights the link between LIF and Stat3 (**Supplementary Table 30b**); LIF activates the JAK/STAT signaling pathway

of which Stat3 is a part[37]. LIF is also an important factor in uterine blastocyst implantation, though its expression is required in the utero-placental unit rather than the blastocyst (from which ESCs are derived)[38]. Thus it is striking that GREAT highlights the *Mouse Phenotype* "abnormal trophoblast layer morphology" and the *MGI Expression: Detected* enrichments for trophectoderm and extraembryonic component in Theiler stage 5 (**Supplementary Table 30b**). The binding of Stat3 to these regions may reflect the totipotent state of ESCs or an overlap between genes expressed in trophectoderm and the blastocyst. The *MSigDB Perturbation* ontology also shows enrichment for genes downregulated by expression of constitutively active JUN N-terminal kinase (JNK). Indeed, JNK is known to be a negative regulator of Stat3[39]. Furthermore, the strong enrichment for genes upregulated by insulin is explained by the ability of insulin to activate Stat3[40]. Finally, there is enrichment for transcriptional modulators present during myeloid differentiation, and indeed, Stat3 is an essential component for myeloid differentiation[41].

Results from a "gene-based GREAT" analysis recapitulate DAVID's generality in GO enrichments, with none of the top enrichments indicating the roles of Stat3 in stem cell maintenance and differentiation (**Supplementary Table 31**). The *Mouse Phenotype* ontology enrichments are also more general than in the standard GREAT analysis, with placenta morphology identified but the early embryonic lethality unidentified. Similarly to the p300 in embryonic stem cells example, the *basal plus extension* with a restricted enrichment domain highlights more general terms and appears more similar to the "gene-based GREAT" (**Supplementary Table 32**), while the *two nearest genes* and *single nearest gene* association rules lead to similar enrichments to the default GREAT (**Supplementary Tables 33** and **34**, respectively).

Overall, by coupling appropriate integration of distal binding events with data from many ontologies spanning a wide variety of biological phenomena, GREAT highlights many known functions of Stat3 in mouse ESCs that gene-based tools fail to feature prominently (**Supplementary Table 30**). Experimentally-validated links between Stat3 and its pathway involvements, its known cofactors, and its role in stem cell maintenance are all highlighted by various ontologies. In addition, novel hypotheses of Stat3 involvement in trophectoderm development and its additional cofactors can be studied by targeted future experimentation.

Neuron-restrictive silencer factor (NRSF) in human Jurkat cells

To assess the functional roles of Neuron-Restrictive Silencer Factor (NRSF, also known as RE1-Silencing Transcription Factor or REST), we used ChIP-Seq binding data from human Jurkat cells[19]. NRSF is a transcription factor involved in silencing neuron-specific genes[42–44].

A gene-based analysis of the dataset identified enrichment for genes "mostly involved in neuronal function"[19]. The top ten results of the analysis are reproduced in **Supplementary Table 35a**.

We ran GREAT on a dataset comprised of the most significant ChIP-Seq peaks of NRSF (QuEST score > 1; n = 1,712) using a whole genome background and default settings. Enriched terms overwhelmingly implicate NRSF as binding near genes involved in ion channel activity, neurotransmitter transport, and synaptic transmission (**Supplementary Table 35b**). Additionally, the *GO Cellular Component*, *Mouse Phenotype*, *InterPro*, and *HGNC Gene Families* ontologies indicate that NRSF binds near both calcium channel and potassium channel genes. NRSF has been shown to modulate aldosterone and cortisol production by regulating a calcium channel subunit[45]. NRSF has also been shown to regulate potassium channel expression, affecting the phenotype of human vascular smooth muscle cells[46]. The GREAT enrichments suggest that NRSF may play a role in regulating other calcium and potassium channel genes as well.

The enrichments are robust to all variations of association rule including a "gene-based GREAT" (**Supplementary Tables 36–39**). The ability of both the gene-based analyses and GREAT to identify the neuron-specific functions of NRSF binding data suggests that both proximal and distal binding events play a role in the transcriptional repression of neuron-specific genes[44]. Given the high information content of the 21 bp neuron-restrictive silencer element (NRSE) bound by NRSF[42,43], binding of NRSF to the NRSE may be predominantly functional regardless of the location of the binding area relative to nearby genes (**Figure 2a**).

GA-Binding Protein (GABP) in human Jurkat cells

The binding profile of GA-Binding Protein (GABP) was assayed in human Jurkat cells via ChIP-Seq[19]. GABP is a ubiquitous transcription factor that controls transcriptional regulation of genes involved in many

diverse functions including apoptosis, differentiation, cell cycle, and cellular energy metabolism[47].

A gene-based analysis of GABP-regulated genes showed enrichment for genes "involved in basic cellular processes, particularly those related to gene expression"[19]. The top ten results are reproduced in **Supplementary Table 40a**.

Due to the large number (6,442) of GABP binding peaks present in the dataset, more than 3,000 genes possess a GABP binding peak even within their proximal promoter. As a result the DAVID website cannot even be used to analyze this set, as it can only analyze datasets of 3,000 or fewer genes. In contrast, GREAT can handle datasets of hundreds of thousands of genomic peaks, and any number of resulting gene picks. We ran GREAT on the most significant ChIP-Seq peaks of GABP (QuEST score > 1, n = 3,585) using a whole genome background and default settings (**Supplementary Table 40b**).

Enrichments from the *Pathway Commons* ontology highlight the known functions of GABP as a transcriptional activator, as the strongest enrichment is for genes involved in transcription. Interestingly, there are also strong enrichments for genes involved in various aspects of mRNA processing, with the *MSigDB Pathway* ontology highlighting genes involved in mRNA splicing as its strongest enrichment. GABP has been shown to regulate Hepatocyte Growth Factor-Regulated Tyrosine Kinase Substrate, a protein that mediates alternative mRNA splicing during liver regeneration[47,48]. The *MSigDB Pathway* and *GO Molecular Function* ontologies also show strong enrichments for ribosomal proteins. Indeed, GABP is known to regulate multiple ribosomal proteins[49,50], though the extent of GABP binding suggests that many other ribosomal proteins may also be regulated by GABP. Furthermore, GABP regulates transcription of eIF6, an essential trans-acting factor in ribosome biogenesis[51]. The *MSigDB Pathway* enriched terms "oxidative phosphorylation" and "electron transport" also correspond to known functions of GABP; GABP regulates mtTFA, a mitochondrial transcription factor important in oxidative phosphorylation[52]. Finally, the *Transcription Factor Targets* ontology indicates that GABP binding peaks occur near genes that are regulated by ETS1 and YY1, suggesting a possible cooperative role between GABP and these factors. GABP itself is part of the ETS family, and ChIP-Seq experiments examining the binding of both GABP and ETS1 show that the proteins do bind many similar promoters, though GABP is in general a more ubiquitous factor[53]. Interactions between GABP and YY1 have also been experimentally shown[47,54].

As GABP binds predominantly near the promoter of its target genes (**Figure 2a**), the unique enrichments highlighted by GREAT ontologies are robust to both gene-based analysis (**Supplementary Table 41**) and for all tested variations of association rule (**Supplementary Tables 42–44**).

SUPPLEMENTARY REFERENCES

1. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25–29 (2000).

2. Blake, J. A. *et al.* The Mouse Genome Database genotypes::phenotypes. *Nucleic Acids Res.* **37**, D712–719 (2009).

3. Bult, C. J., Eppig, J. T., Kadin, J. A., Richardson, J. E. & Blake, J. A. The Mouse Genome Database (MGD): mouse biology and model systems. *Nucleic Acids Res.* **36**, D724–728 (2008).

4. Smith, C. L., Goldsmith, C. A. & Eppig, J. T. The Mammalian Phenotype Ontology as a tool for annotating, analyzing and comparing phenotypic information. *Genome Biol.* **6**, R7 (2005).

5. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 15545–15550 (2005).

6. Brentani, H. *et al.* The generation and utilization of a cancer-oriented representation of the human transcriptome by using expressed sequence tags. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 13418–13423 (2003).

7. Segal, E., Friedman, N., Koller, D. & Regev, A. A module map showing conditional activity of expression modules in cancer. *Nat. Genet.* **36**, 1090–1098 (2004).

8. Mi, H., Guo, N., Kejariwal, A. & Thomas, P. D. PANTHER version 6: protein sequence and function evolution data with expanded representation of biological pathways. *Nucleic Acids Res.* **35**, D247–252 (2007).

9. Cerami, E. G., Bader, G. D., Gross, B. E. & Sander, C. cPath: open source software for collecting, storing, and querying biological pathways. *BMC Bioinformatics* **7**, 497 (2006).

10. Caspi, R. *et al.* The MetaCyc Database of metabolic pathways and enzymes and the BioCyc collection of Pathway/Genome Databases. *Nucleic Acids Res.* **36**, D623–631 (2008).

11. Smith, C. M. *et al.* The mouse Gene Expression Database (GXD): 2007 update. *Nucleic Acids Res.* **35**, D618–623 (2007).

12. Theiler, K. *The House Mouse: Development and Normal Stages from Fertilization to 4 Weeks of Age* (Springer-Verlag, New York, 1989).

13. Linhart, C., Halperin, Y. & Shamir, R. Transcription factor and microRNA motif discovery: the Amadeus platform and a compendium of metazoan target sets. *Genome Res.* **18**, 1180–1189 (2008).

14. Hunter, S. *et al.* InterPro: the integrative protein signature database. *Nucleic Acids Res.* **37**, D211–215 (2009).

15. Ruan, J. *et al.* TreeFam: 2008 Update. *Nucleic Acids Res.* **36**, D735–740 (2008).

16. Bruford, E. A. *et al.* The HGNC Database in 2008: a resource for the human genome. *Nucleic Acids Res.* **36**, D445–448 (2008).

17. Kent, W. J. *et al.* The human genome browser at UCSC. *Genome Res.* **12**, 996–1006 (2002).

18. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate – a practical and powerful approach to multiple hypothesis testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).

19. Valouev, A. *et al.* Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. *Nat. Methods* **5**, 829–834 (2008).

20. Visel, A. *et al.* ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854–858 (2009).

21. Huang, d. a. W. *et al.* DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists. *Nucleic Acids Res.* **35**, W169–175 (2007).

22. Khatri, P. & Draghici, S. Ontological analysis of gene expression data: current tools, limitations, and open problems. *Bioinformatics* **21**, 3587–3595 (2005).

23. Hsu, F. *et al.* The UCSC Known Genes. *Bioinformatics* **22**, 1036–1046 (2006).

24. Powell, L. M. & Jarman, A. P. Context dependence of proneural bHLH proteins. *Curr. Opin. Genet. Dev.* **18**, 411–417 (2008).

25. Shen, Q., Zhong, W., Jan, Y. N. & Temple, S. Asymmetric Numb distribution is critical for asymmetric cell division of mouse cerebral cortical stem cells and neuroblasts. *Development* **129**, 4843–4853 (2002).

26. Zhou, C. J., Borello, U., Rubenstein, J. L. & Pleasure, S. J. Neuronal production and precursor proliferation defects in the neocortex of mice with loss of function in the canonical Wnt signaling pathway. *Neuroscience* **142**, 1119–1131 (2006).

27. Chen, X. *et al.* Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. *Cell* **133**, 1106–1117 (2008).

28. Ivanova, N. *et al.* Dissecting self-renewal in stem cells with RNA interference. *Nature* **442**, 533–538 (2006).

29. Drew, P. D. *et al.* Cloning and expression analysis of a human cDNA homologous to Xenopus TFIIIA. *Gene* **159**, 215–218 (1995).

30. Begley, C. G. *et al.* Molecular characterization of NSCL, a gene encoding a helix-loop-helix protein expressed in the developing nervous system. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 38–42 (1992).

31. Matsuda, T. *et al.* STAT3 activation is sufficient to maintain an undifferentiated state of mouse embryonic stem cells. *EMBO J.* **18**, 4261–4269 (1999).

32. Bromberg, J. F. *et al.* Stat3 as an oncogene. *Cell* **98**, 295–303 (1999).

33. Hirano, T., Ishihara, K. & Hibi, M. Roles of STAT3 in mediating the cell growth, differentiation and survival signals relayed through the IL-6 family of cytokine receptors. *Oncogene* **19**, 2548–2556 (2000).

34. Takeda, K. *et al.* Targeted disruption of the mouse Stat3 gene leads to early embryonic lethality. *Proc. Natl. Acad. Sci. U.S.A.* **94**, 3801–3804 (1997).

35. Fang, P., Hwa, V. & Rosenfeld, R. G. Interferon-gamma-induced dephosphorylation of STAT3 and apoptosis are dependent on the mTOR pathway. *Exp. Cell Res.* **312**, 1229–1239 (2006).

36. Hsieh, F. C., Cheng, G. & Lin, J. Evaluation of potential Stat3-regulated genes in human breast cancer. *Biochem. Biophys. Res. Commun.* **335**, 292–299 (2005).

37. Kisseleva, T., Bhattacharya, S., Braunstein, J. & Schindler, C. W. Signaling through the JAK/STAT pathway, recent advances and future challenges. *Gene* **285**, 1–24 (2002).

38. Auernhammer, C. J. & Melmed, S. Leukemia-inhibitory factor-neuroimmune modulator of endocrine function. *Endocr. Rev.* **21**, 313–345 (2000).

39. Lim, C. P. & Cao, X. Serine phosphorylation and negative regulation of Stat3 by JNK. *J. Biol. Chem.* **274**, 31055–31061 (1999).

40. Coffer, P. J. *et al.* Insulin activates Stat3 independently of p21ras-ERK and PI-3K signal transduction. *Oncogene* **15**, 2529–2539 (1997).

41. McLemore, M. L. *et al.* STAT-3 activation is required for normal G-CSF-dependent proliferation and granulocytic differentiation. *Immunity* **14**, 193–204 (2001).

42. Chong, J. A. *et al.* REST: a mammalian silencer protein that restricts sodium channel gene expression to neurons. *Cell* **80**, 949–957 (1995).

43. Schoenherr, C. J. & Anderson, D. J. The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes. *Science* **267**, 1360–1363 (1995).

44. Chen, Z. F., Paquette, A. J. & Anderson, D. J. NRSF/REST is required in vivo for repression of multiple neuronal target genes during embryogenesis. *Nat. Genet.* **20**, 136–142 (1998).

45. Somekawa, S. *et al.* Regulation of aldosterone and cortisol production by the transcriptional repressor neuron restrictive silencer factor. *Endocrinology* **150**, 3110–3117 (2009).

46. Cheong, A. *et al.* Downregulated REST transcription factor is a switch enabling critical potassium channel expression and cell proliferation. *Mol. Cell* **20**, 45–52 (2005).

47. Rosmarin, A. G., Resendes, K. K., Yang, Z., McMillan, J. N. & Fleming, S. L. GA-binding protein transcription factor: a review of GABP as an integrator of intracellular signaling and protein-protein interactions. *Blood Cells Mol. Dis.* **32**, 143–154 (2004).

48. Du, K., Leu, J. I., Peng, Y. & Taub, R. Transcriptional up-regulation of the delayed early gene HRS/SRp40 during liver regeneration. Interactions among YY1, GA-binding proteins, and mitogenic signals. *J. Biol. Chem.* **273**, 35208–35215 (1998).

49. Curcić, D., Glibetić, M., Larson, D. E. & Sells, B. H. GA-binding protein is involved in altered expression of ribosomal protein L32 gene. *J. Cell. Biochem.* **65**, 287–307 (1997).

50. Genuario, R. R., Kelley, D. E. & Perry, R. P. Comparative utilization of transcription factor GABP by the promoters of ribosomal protein genes rpL30 and rpL32. *Gene Expr.* **3**, 279–288 (1993).

51. Donadini, A. *et al.* GABP complex regulates transcription of eIF6 (p27BBP), an essential trans-acting factor in ribosome biogenesis. *FEBS Lett.* **580**, 1983–1987 (2006).

52. Chinenov, Y., Coombs, C. & Martin, M. E. Isolation of a bi-directional promoter directing expression of the mouse GABPalpha and ATP synthase coupling factor 6 genes. *Gene* **261**, 311–320 (2000).

53. Collins, P. J., Kobayashi, Y., Nguyen, L., Trinklein, N. D. & Myers, R. M. The ets-related transcription factor GABP directs bidirectional transcription. *PLoS Genet.* **3**, e208 (2007).

54. Deléhouzée, S. *et al.* GABP, HCF-1 and YY1 are involved in Rb gene expression during myogenesis. *Genes Cells* **10**, 717–731 (2005).

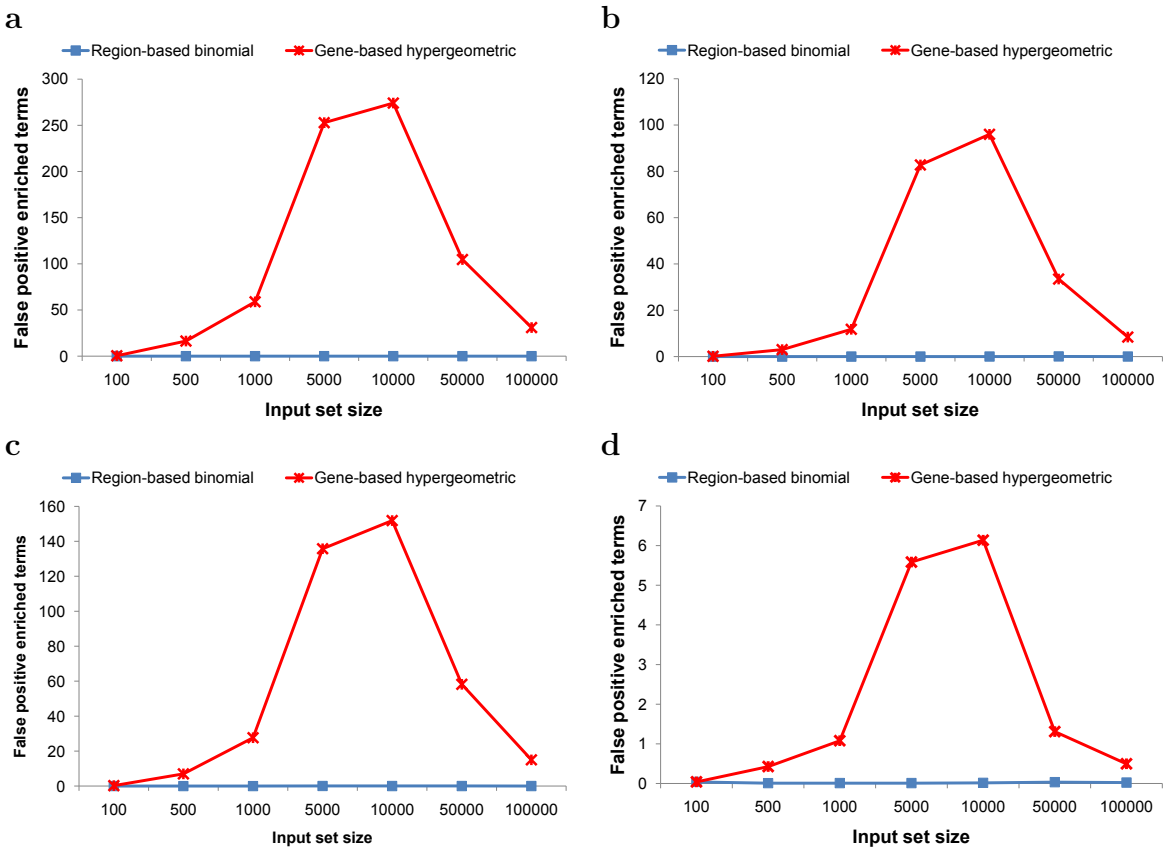
55. Al-Shahrour, F. *et al.* Babelomics: advanced functional profiling of transcriptomics, proteomics and genomics experiments. *Nucleic Acids Res.* **36**, W341–346 (2008).

56. Khatri, P. *et al.* Onto-Tools: new additions and improvements in 2006. *Nucleic Acids Res.* **35**, W206–211 (2007).

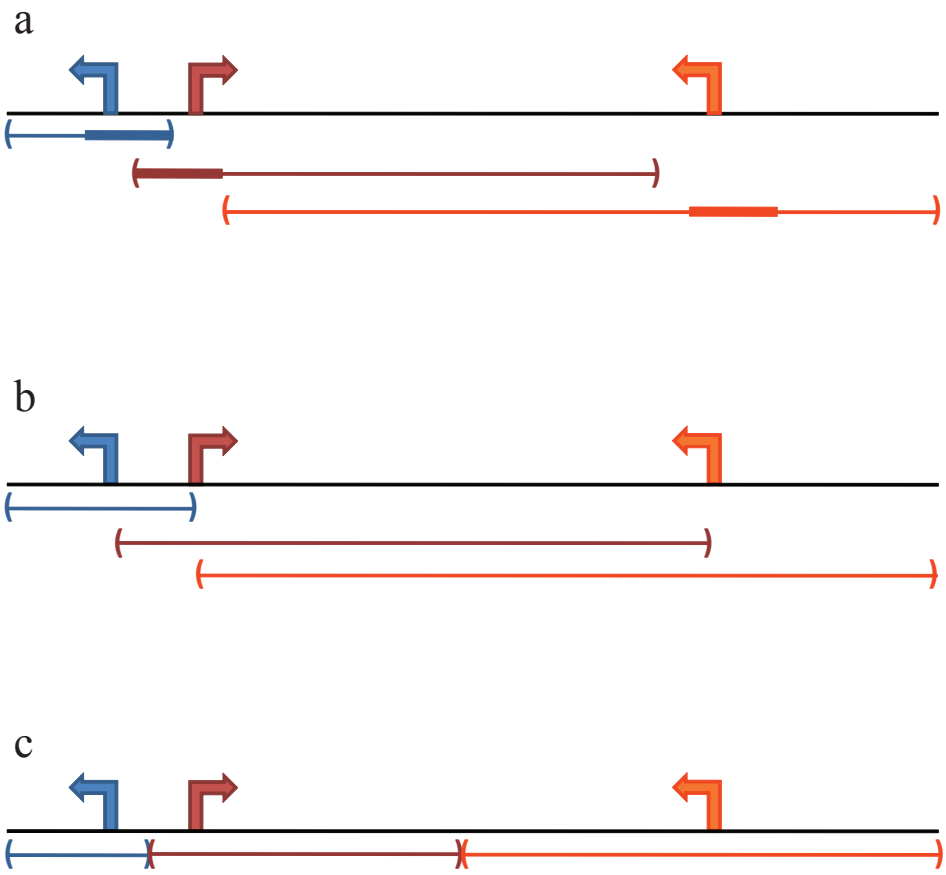
57. Beissbarth, T. & Speed, T. P. GOstat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* **20**, 1464–1465 (2004).

58. Zeeberg, B. R. *et al.* GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol.* **4**, R28 (2003).

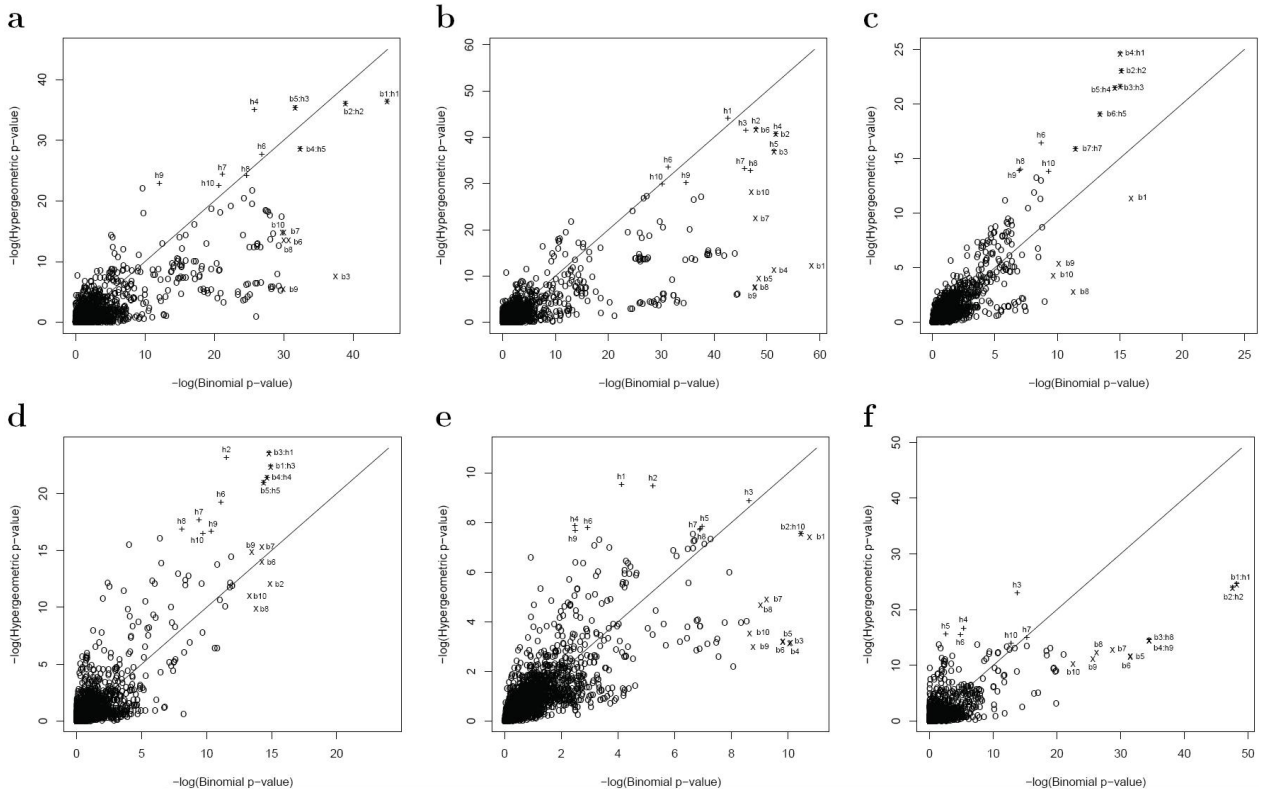
SUPPLEMENTARY FIGURES



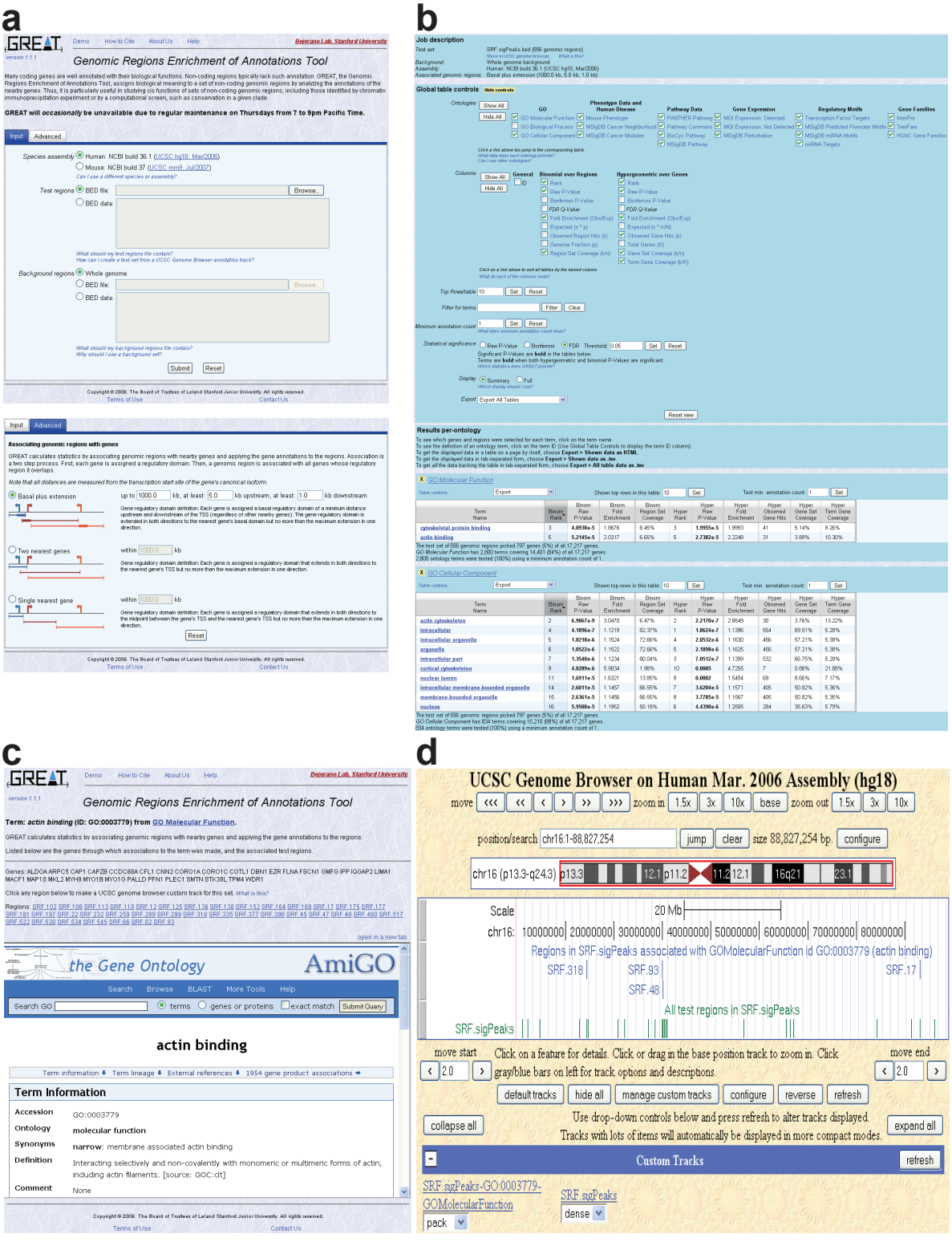
Supplementary Figure 1: The gene-based hypergeometric test generates false positive enriched terms in many ontologies when not restricted to only proximal binding events. The average number of false positive enriched terms for the region-based binomial (blue) and gene-based hypergeometric (red) test over 1,000 random input sets in which each base pair in the human genome (excluding assembly gaps) is equally likely to be included in the set is shown for **(a)** *MGI Expression: Detected*, **(b)** *MGI Expression: Not Detected*, **(c)** *Mouse Phenotype*, and **(d)** *InterPro*. Each test associates genomic regions to genes using GREAT's default *basal plus extension* association rule with 5+1 kb basal domain and extension up to 1 Mb (**Supplementary Methods**). False positive enriched terms are defined as those significant at a threshold of 0.05 after applying the conservative Bonferroni correction. Though the total number of average false positive terms varies across ontology (note scale changes on y-axis), the qualitative shape of the graphs is similar with the majority of false positive enriched terms occurring for input sets containing 1-50k elements.



Supplementary Figure 2: Computationally-defined regulatory domains. The transcription start site (TSS) of each gene is shown as an arrow. The corresponding regulatory domain for each gene is shown in matching color as a bracketed line. The association rule and relevant parameters used in a run of GREAT can be altered via the web interface prior to execution. **(a)** The *basal plus extension* association rule assigns a basal regulatory domain to each gene regardless of genes nearby (thick line). The domain is then extended to the basal regulatory domain of the nearest upstream and downstream genes. **(b)** The *two nearest genes* association rule extends the regulatory domain to the TSS of the nearest upstream and downstream genes. **(c)** The *single nearest gene* association rule extends the regulatory domain to the midpoint between this gene's TSS and the nearest gene's TSS both upstream and downstream. All regulatory domain extension rules limit extension to a user-defined maximum distance for genes that have no other genes nearby.



Supplementary Figure 3: Binomial and hypergeometric p-value differences for several different datasets. Each set uses GREAT's default *basal plus extension* association rule with 5+1 kb basal domain and extension up to 1 Mb (**Supplementary Methods**). "x" denotes a top ten most enriched term using the binomial test, with "b1" the top ranking, etc. "+" denotes a top ten most enriched term using the hypergeometric test, with "h1" denoting the top ranking, etc. **(a)** p300 mouse embryonic limb data set [20]. **(b)** p300 mouse embryonic forebrain data set [20]. **(c)** p300 mouse embryonic midbrain data set [20]. **(d)** NRSF human Jurkat data set [19]. **(e)** p300 mouse embryonic stem cell data set [27]. **(f)** Stat3 mouse embryonic stem cell data set [27].



Supplementary Figure 4: Screenshots of the GREAT version 1.1.3 workflow. **(a)** The input screen where the user chooses an organism, inputs a set of *cis*-regulatory regions, and optionally alters the mapping of *cis*-regulatory regions to genes. **(b)** The main output screen displays enriched terms from twenty different ontologies. Global and per-table controls allow the user to set significance criteria and level of detail. Publication-grade HTML tables and tab-separated files for independent analysis or formatting are provided on the fly. **(c)** The individual term details screen, available by clicking on any term in any ontology, displays information related to the enrichment of the term including the *cis*-regulatory regions and genes that make this term enriched and an inset definition of the ontology term. **(d)** Clicking any listed *cis*-regulatory region in an individual term details screen opens a UCSC Genome Browser display focused on that region that includes custom tracks for the entire set of input *cis*-regulatory regions and for the subset that contributes to that particular term.

SUPPLEMENTARY TABLES

Supplementary Table 1: Human and mouse ontologies currently supported by GREAT.

Ontology	References
Gene Ontology	
GO Molecular Function	ref. 1
GO Biological Process	ref. 1
GO Cellular Component	ref. 1
Phenotype Data and Human Disease	
Mouse Phenotype	refs. 2–4
MSigDB Cancer Neighborhood*	refs. 5,6
MSigDB Cancer Modules*	refs. 5,7
Pathway Data	
PANTHER Pathway	ref. 8
Pathway Commons	ref. 9
BioCyc Pathway	ref. 10
MSigDB Pathway	ref. 5
Gene Expression Data	
MGI Expression: Detected	refs. 3,11
MGI Expression: Not Detected	refs. 3,11
MSigDB Perturbation	ref. 5
Regulatory Motifs	
MSigDB Predicted Promoter Motifs	ref. 5
Transcription Factor Targets	ref. 13
MSigDB miRNA Motifs	ref. 5
miRNA Targets	ref. 13
Gene Families	
InterPro	ref. 14
TreeFam	ref. 15
HGNC Gene Families*	ref. 16

* Ontology only supported in human.

Supplementary Table 2: Ontology contents for human.

Ontology	Terms	Genes	Direct associations	Download date
GO Molecular Function	2,800	14,401	43,207	March 5, 2009
GO Biological Process	5,215	13,293	47,287	March 5, 2009
GO Cellular Component	834	15,210	39,984	March 5, 2009
Mouse Phenotype	5,781	5,377	96,704	April 22, 2009
MSigDB Cancer Neighborhood	427	4,717	41,713	March 11, 2009
MSigDB Cancer Modules	456	7,918	47,511	March 11, 2009
PANTHER Pathway	150	1,983	4,676	March 9, 2009
Pathway Commons	1,253	3,921	52,505	July 20, 2009
BioCyc Pathway	288	693	1,860	March 13, 2009
MSigDB Pathway	706	6,473	25,280	March 11, 2009
MGI Expression: Detected	6,700	7,330	190,463	March 23, 2009
MGI Expression: Not Detected	3,079	4,812	82,621	March 23, 2009
MSigDB Perturbation	911	11,189	67,052	March 11, 2009
Transcription Factor Targets	19	5,375	9,980	March 12, 2009
MSigDB Predicted Promoter Motifs	615	11,777	154,911	March 11, 2009
MSigDB miRNA Motifs	222	6,896	32,101	March 11, 2009
miRNA Targets	9	1,095	1,199	March 12, 2009
InterPro	6,587	15,228	53,630	March 3, 2009
TreeFam	8,272	16,684	16,812	March 3, 2009
HGNC Gene Families	238	4,616	5,021	March 6, 2009

Supplementary Table 3: Ontology contents for mouse.

Ontology	Terms	Genes	Direct associations	Download date
GO Molecular Function	2,380	13,932	47,595	March 23, 2009
GO Biological Process	4,539	12,805	45,480	March 23, 2009
GO Cellular Component	686	14,548	34,827	March 23, 2009
Mouse Phenotype	5,798	5,536	98,514	April 22, 2009
PANTHER Pathway	149	1,759	4,023	March 9, 2009
Pathway Commons*	83	116	206	July 20, 2009
BioCyc Pathway	275	888	2,109	March 13, 2009
MSigDB Pathway	456	3,479	10,778	March 11, 2009
MGI Expression: Detected	6,701	7,730	194,545	March 23, 2009
MGI Expression: Not Detected	3,119	5,121	87,257	March 23, 2009
MSigDB Perturbation	248	7,419	23,073	March 11, 2009
Transcription Factor Targets	6	1,194	1,377	March 12, 2009
MSigDB Predicted Promoter Motifs	615	9,117	127,459	March 11, 2009
MSigDB miRNA Motifs	222	5,948	28,369	March 11, 2009
miRNA Targets	1	97	97	March 12, 2009
InterPro	6,281	16,096	51,140	March 3, 2009
TreeFam	7,953	16,974	17,040	March 3, 2009

* The mouse Pathway Commons ontology contains considerably less data than its human counterpart because many input databases in Pathway Commons are specific to human pathways.

Supplementary Table 4: Comparison of several enrichment tools.

	Primary use	Ontologies supported	Web based	Real-time response	Reference
GREAT	***Cis*-regulatory regions**	**Many**	**Yes**	**Yes**	*This publication*
DAVID	Gene sets	**Many**	**Yes**	**Yes**	ref. 21
Babelomics	Gene sets	**Many**	**Yes**	No	ref. 55
OntoTools	Gene sets	GO, chromosome	**Yes**	**Yes**	ref. 56
GOstat	Gene sets	Only GO	**Yes**	**Yes**	ref. 57
GoMiner	Gene sets	Only GO	No	N/A	ref. 58

Supplementary Table 5: Datasets analyzed by GREAT.

Dataset	Species	Tissue	References
Serum Response Factor	*Homo sapiens*	Jurkat cells	ref. 19
Neuron-Restrictive Silencer Factor	*Homo sapiens*	Jurkat cells	ref. 19
GA-Binding Protein	*Homo sapiens*	Jurkat cells	ref. 19
p300	*Mus musculus*	Embryonic limb	ref. 20
p300	*Mus musculus*	Embryonic forebrain	ref. 20
p300	*Mus musculus*	Embryonic midbrain	ref. 20
p300	*Mus musculus*	Embryonic stem cells	ref. 27
Signal transducer and activator of transcription 3	*Mus musculus*	Embryonic stem cells	ref. 27

Supplementary Table 6: "Gene-based GREAT" enrichments of all genes that possess an SRF binding peak within 2 kb of its transcription start site. Shown are the top ten hypergeometric enriched terms at a false discovery rate of 0.05.

Ontology	Term	Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Observed Gene Hits
GO Molecular Function	nucleic acid binding	1.5983e-15	4.4752e-12	323
	protein binding	7.1549e-15	1.0017e-11	644
	RNA binding	1.5366e-12	1.4342e-9	97
	binding	1.4983e-8	1.0488e-5	918
	structural constituent of ribosome	6.9336e-7	0.0004	28
	transcription repressor activity	9.4062e-7	0.0004	39
	transcription factor binding	1.6303e-6	0.0007	59
	transcription regulator activity	4.7408e-5	0.0166	132
	transcription cofactor activity	4.7863e-5	0.0149	42
	transcription corepressor activity	0.0001	0.0379	21
GO Biological Process	cellular macromolecule metabolic process	5.5441e-28	2.8912e-24	518
	macromolecule metabolic process	1.6637e-27	4.3380e-24	520
	cellular biopolymer metabolic process	3.7180e-27	6.4632e-24	508
	biopolymer metabolic process	8.8278e-27	1.1509e-23	509
	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	3.4330e-22	3.5807e-19	323
	gene expression	2.5137e-21	2.1848e-18	281
	cellular metabolic process	7.4676e-21	5.5634e-18	593
	primary metabolic process	1.2033e-19	7.8442e-17	576
	RNA metabolic process	2.6534e-18	1.5375e-15	129
	metabolic process	4.5534e-18	2.3746e-15	621
GO Cellular Component	intracellular	1.0175e-35	8.4863e-33	939
	intracellular part	3.6376e-34	1.5169e-31	908
	intracellular organelle	9.1566e-30	2.5455e-27	787
	organelle	1.1070e-29	2.3081e-27	787
	intracellular membrane-bounded organelle	1.6916e-28	2.8216e-26	719
	membrane-bounded organelle	1.9496e-28	2.7099e-26	719
	nucleus	9.1791e-22	1.0936e-19	496
	intracellular organelle part	2.6415e-21	2.7538e-19	413
	organelle part	3.3485e-21	3.1029e-19	414
	nuclear part	1.3251e-20	1.1052e-18	186
Pathway Commons	RNA Polymerase II Transcription Initiation And Promoter Clearance	2.3080e-14	2.8919e-11	60
	Formation and Maturation of mRNA Transcript	2.3080e-14	2.8919e-11	60
	RNA Polymerase II Promoter Escape	2.3080e-14	2.8919e-11	60
	RNA Polymerase II Transcription Initiation	2.3080e-14	2.8919e-11	60
	RNA Polymerase II Transcription Pre-Initiation	2.3080e-14	2.8919e-11	60
	RNA Polymerase II Transcription	2.3080e-14	2.8919e-11	60
	Gene Expression	2.7639e-14	4.9474e-12	68
	Elongation of Intron-Containing Transcripts and co-transcriptional mRNA splicing	2.1048e-13	3.2966e-11	56
	Elongation and Processing of Capped Transcripts	2.1048e-13	3.2966e-11	56
	mRNA Capping	2.9039e-13	3.6385e-11	56
MSigDB Pathway	RIBOSOMAL_PROTEINS	1.2397e-7	8.7523e-5	24
	PGC related genes	9.7838e-7	0.0003	57
	Genes involved in ribosome	1.1507e-6	0.0003	18
	Genes involved in mRNA splicing	0.0001	0.0195	13
	Transcription factors enriched in fetal liver	0.0001	0.0178	16
	Genes highly expressed in hepatocellular carcinoma with poor survival.	0.0002	0.0240	26
	Mitochondrial genes	0.0002	0.0223	52
	Tricarboxylic acid related genes	0.0004	0.0311	6

Ontology	Term	Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Observed Gene Hits
Transcription Factor Targets	Targets of SRF, identified by ChIP-chip in different cell lines: Jurkat, T/G HA-VSMC, and Be(2)-C cells.	2.8894e-75	5.4898e-74	102
	Targets of CREB, identified by ChIP-chip in HEK293T cells in three different time points after forskolin stimulation.	3.4466e-64	3.2743e-63	372
	Targets of ETS1, identified by ChIP-chip in Jurkat T-cells.	5.0666e-39	3.2088e-38	206
	Targets of YY1 identified by ChIP-chip.	5.6200e-18	2.6695e-17	110
	Targets of HNF4alpha, identified by ChIP-chip in hepatocytes.	4.9191e-16	1.8692e-15	181
	Targets of NRF1, identified by ChIP-chip in quiescent T98G cells.	1.3109e-7	4.1512e-7	82
	Genes that are bound by both E2F4 and p130 in three different growth arrest condtions, identified by ChIP-chip in T98G and U2OS cells under growth arrest.	2.9623e-7	8.0405e-7	34
	Targets of estrogen receptor alpha, identified by ChIP-DSL in MCF-7 cells.	5.6159e-6	1.3338e-5	58
	Targets of Nanog, identifed by ChIP-chip in embryonic stem cells.	0.0001	0.0003	75
	Targets of Sox2, identifed by ChIP-chip in embryonic stem cells.	0.0002	0.0003	62
Predicted Promoter Motifs	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif GGGCGGR which matches annotation for SP1: Sp1 transcription factor	4.4111e-29	2.7128e-26	346
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif SCGGAAGY which matches annotation for ELK1: ELK1, member of ETS oncogene family	5.8988e-28	1.8139e-25	183
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif DCCWTATATGGNCWN which matches annotation for SRF: serum response factor (c-fos serum response element-binding transcription factor)	8.2882e-28	1.6991e-25	67
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif ATGCCCATATATGGWNNT which matches annotation for SRF: serum response factor (c-fos serum response element-binding transcription factor)	4.0786e-17	6.2708e-15	26
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif CCAWATAWGGMNMNG which matches annotation for SRF: serum response factor (c-fos serum response element-binding transcription factor)	7.3770e-17	9.0737e-15	54
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif GNCCAWATAWGGMN which matches annotation for SRF: serum response factor (c-fos serum response element-binding transcription factor)	6.9116e-16	7.0844e-14	56
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif SCCAWATAWGGMNMNNNN which matches annotation for SRF: serum response factor (c-fos serum response element-binding transcription factor)	1.5217e-15	1.3369e-13	52
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif VCCGGAAGNGCR which matches annotation for GABPA: GA binding protein transcription factor, alpha subunit 60kDa GABPB2: GA binding protein transcription factor, beta subunit 2	1.4177e-14	1.0899e-12	54
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif RCGCANGCGY which matches annotation for NRF1: nuclear respiratory factor 1	2.0319e-13	1.3885e-11	120
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif MGGAAGTG which matches annotation for GABPA: GA binding protein transcription factor, alpha subunit 60kDa GABPB2: GA binding protein transcription factor, beta subunit 2	4.7599e-13	2.9273e-11	105

Supplementary Table 7: GREAT enrichments of SRF using the *basal plus extension* association rule with a basal regulatory region extending 5 kb upstream and 1 kb downstream of the transcription start site and a maximum extension of 50 kb. Shown are the top ten binomial enriched terms at a false discovery rate of 0.05 with a fold enrichment of at least two that are also significant by the hypergeometric test, using the highest-scoring SRF peaks anywhere in the genome (QuEST score > 1; n = 556).

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Cellular Component	intracellular	4.6225e-94	1.2651e-91	2.7438	1.5475e-5	1.1840	390
	intracellular part	9.4136e-90	1.9627e-87	2.7556	1.8326e-5	1.1891	376
	intracellular organelle	2.3290e-79	3.8646e-77	2.8723	6.6753e-5	1.2159	323
	organelle	2.5932e-79	3.6046e-77	2.8711	6.8851e-5	1.2153	323
	intracellular membrane-bounded organelle	2.5494e-69	3.0374e-67	2.8753	0.0023	1.2018	285
	membrane-bounded organelle	2.7954e-69	2.9142e-67	2.8741	0.0020	1.2013	295
	nucleus	7.1757e-52	6.6495e-50	3.1226	0.0028	1.2803	197
	intracellular non-membrane-bounded organelle	2.7688e-26	1.6494e-24	3.6146	0.0442	1.4123	87
	nuclear lumen	1.4675e-20	7.6496e-19	4.3983	0.0292	1.6571	50
	actin cytoskeleton	1.8915e-13	6.5729e-12	7.1454	0.0005	3.0900	22
Pathway Commons	Class I PI3K signaling events	1.1681e-16	1.4636e-13	7.9903	0.0404	2.6793	20
	TRAIL signaling pathway	1.7751e-16	1.1121e-13	6.8445	0.0362	2.3809	23
	Role of Calcineurin-dependent NFAT signaling in lymphocytes	6.5036e-16	2.7163e-13	13.9023	0.0393	3.9854	11
	Further platelet releasate	6.6573e-8	1.8958e-6	20.9506	0.0564	8.6955	6
MSigDB Pathway	Mouse genes associated with signal transduction through calcium, calcineurin, and NF-AT.	6.1451e-20	4.3384e-17	15.4156	0.0141	4.0241	13
	Transcription factors enriched in fetal liver	5.2110e-17	1.8395e-14	16.0163	0.0136	4.3840	11
Transcription Factor Targets	Targets of SRF, identified by ChIP-chip in different cell lines: Jurkat, T/G HA-VSMC, and Be(2)-C cells.	1.1407e-110	2.1674e-109	40.6389	6.8064e-75	15.4586	80
	Targets of CREB, identified by ChIP-chip in HEK293T cells in three different time points after forskolin stimulation.	5.3922e-38	5.1226e-37	4.0982	6.2784e-8	1.6796	117
	Targets of YY1 identified by ChIP-chip.	1.3220e-19	8.3726e-19	5.5997	0.0003	1.9501	40
	Targets of ETS1, identified by ChIP-chip in Jurkat T-cells.	1.9745e-17	9.3789e-17	3.9291	0.0035	1.5546	55
	Targets of HNF4alpha, identified by ChIP-chip in hepatocytes.	4.6146e-13	1.7535e-12	3.0283	0.0385	1.3106	58
	Targets of NRF1, identified by ChIP-chip in quiescent T98G cells.	1.0197e-11	3.2291e-11	4.1047	0.0086	1.6807	34
	Genes that are bound by both E2F4 and p130 in three different growth arrest condtions, identified by ChIP-chip in T98G and U2OS cells under growth arrest.	1.9151e-6	4.0430e-6	5.2615	0.0168	2.2165	13
	Targets of HSF1, identified by ChIP-chip in HeLa cells under heat shock.	3.4585e-6	5.9738e-6	3.8622	0.0213	1.8573	18
	Targets of E2F4 that are expressed during cell cycle entry, identified by ChIP-chip in quiescent WI-38 cells.	2.3408e-5	3.4211e-5	7.0802	0.0096	3.2701	8
	Genes whose expression peaks periodically in the G1/S cell cycle phase.	5.4927e-5	7.4543e-5	3.8085	0.0412	1.8599	14
Predicted Promoter Motifs	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif DCCWTATATGGNCWN which matches annotation for SRF: serum response factor (c-fos serum response element-binding transcription factor)	2.3053e-60	1.4178e-57	20.7384	8.5263e-26	7.5020	48
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif GNCCAWATAWGGMN which matches annotation for SRF: serum response factor (c-fos serum response element-binding transcription factor)	2.9199e-48	8.9788e-46	15.8078	6.4411e-17	5.4695	41
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif CCAWATAWGGMNMNG which matches annotation for SRF: serum response factor (c-fos serum response element-binding transcription factor)	4.8422e-48	9.5164e-46	16.8717	1.0734e-16	5.8105	39
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif SCCAWATAWGGMNMNNNN which matches annotation for SRF: serum response factor (c-fos serum response element-binding transcription factor)	8.9504e-48	1.3761e-45	16.6604	1.0734e-16	5.8105	39
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif GGGCGGR which matches annotation for SP1: Sp1 transcription factor	1.4863e-40	1.8281e-38	3.6955	5.5235e-6	1.5411	135
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif ATGCCCATATATGGWNNT which matches annotation for SRF: serum response factor (c-fos serum response element-binding transcription factor)	7.0794e-39	7.2564e-37	44.4133	2.3170e-15	12.5033	20
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif SCGGAAGY which matches annotation for ELK1: ELK1, member of ETS oncogene family	3.6138e-29	3.1750e-27	5.1858	5.1486e-6	1.9573	69
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif CCAWWNAAGG which matches annotation for SRF: serum response factor (c-fos serum response element-binding transcription factor)	3.6746e-26	2.8249e-24	20.1797	1.4221e-7	6.2301	17
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif MGGAAGTG which matches annotation for GABPA: GA binding protein transcription factor, alpha subunit 60kDa GABPB2: GA binding protein transcription factor, beta subunit 2	6.1914e-20	3.8077e-18	5.1783	0.0041	1.8650	42
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif GTGACGY which matches annotation for E4F1: E4F transcription factor 1	1.4545e-16	7.4545e-15	4.8373	0.0226	1.7798	35
TreeFam	Early growth response protein	4.2561e-17	3.5207e-13	141.1322	0.0386	25.5066	4

Supplementary Table 8: GREAT enrichments of SRF using the *two nearest genes* association rule with a maximum extension of 1 Mb. Shown are the top ten binomial enriched terms at a false discovery rate of 0.05 with a fold enrichment of at least two that are also significant by the hypergeometric test, using the highest-scoring SRF peaks anywhere in the genome (QuEST score > 1; n = 556).

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Molecular Function	actin filament binding	5.7178e-13	1.6010e-9	8.7231	0.0078	4.7469	11
	cytoskeletal protein binding	1.1226e-11	1.5717e-8	2.5415	0.0057	1.9755	52
	actin binding	4.7472e-11	4.4307e-8	2.8115	0.0053	2.1806	39
GO Cellular Component	actin cytoskeleton	3.0321e-16	2.5288e-13	4.0765	3.3562e-6	2.7432	37
Pathway Commons	TRAIL signaling pathway	8.4074e-12	1.0534e-8	2.9916	0.0002	2.3496	43
	Class I PI3K signaling events	6.3320e-11	3.9670e-8	3.1776	0.0001	2.5457	36
	TNF receptor signaling pathway	1.5415e-8	4.8288e-6	2.6731	0.0497	1.9913	31
	TCR signaling in na&#65533;ve CD8+ T cells	1.3913e-7	6.4568e-6	3.7078	0.0501	2.5036	18
	Glypican 1 network	1.3941e-7	6.2385e-6	2.1469	0.0135	1.8477	46
	Glypican pathway	1.8834e-7	6.9410e-6	2.0445	0.0150	1.8205	49
	Further platelet releasate	1.7813e-6	4.6500e-5	10.1373	0.0414	5.3550	7
	IRS-mediated signalling	0.0034	0.0342	3.3112	0.0476	3.9860	9
MSigDB Pathway	Transcription factors enriched in fetal liver	3.4283e-9	8.0679e-7	4.5621	0.0450	3.1556	15
Transcription Factor Targets	Targets of SRF, identified by ChIP-chip in different cell lines: Jurkat, T/G HA-VSMC, and Be(2)-C cells.	1.8284e-76	3.4740e-75	13.1009	3.8646e-56	8.4660	83
	Targets of NRF1, identified by ChIP-chip in quiescent T98G cells.	2.3209e-7	8.8194e-7	2.0783	0.0177	1.4612	56
	Genes that are bound by both E2F4 and p130 in three different growth arrest condtions, identified by ChIP-chip in T98G and U2OS cells under growth arrest.	0.0025	0.0079	2.0895	0.0497	1.7100	19
	Targets of E2F4 that are expressed during cell cycle entry, identified by ChIP-chip in quiescent WI-38 cells.	0.0211	0.0334	2.0499	0.0456	2.1577	10
Predicted Promoter Motifs	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif ATGCCCATATATGGWNNT which matches annotation for SRF: serum response factor (c-fos serum response element-binding transcription factor)	5.3652e-29	3.2996e-26	15.5920	2.4284e-13	7.5900	23
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif DCCWTATATGGNCWN which matches annotation for SRF: serum response factor (c-fos serum response element-binding transcription factor)	1.1270e-28	3.4656e-26	5.0865	3.0902e-19	4.5375	55
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif SCCAWATAWGGMNMNNNN which matches annotation for SRF: serum response factor (c-fos serum response element-binding transcription factor)	4.5660e-23	9.3603e-21	4.2901	4.3472e-14	3.8536	49
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif CCAWATAWGGMNMNG which matches annotation for SRF: serum response factor (c-fos serum response element-binding transcription factor)	1.0032e-22	1.5424e-20	4.1672	4.3472e-14	3.8536	49
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif GNCCAWATAWGGMN which matches annotation for SRF: serum response factor (c-fos serum response element-binding transcription factor)	5.9134e-20	7.2734e-18	3.6911	1.1667e-13	3.5913	51
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif SCGGAAGY which matches annotation for ELK1: ELK1, member of ETS oncogene family	1.4762e-13	1.5131e-11	2.1970	0.0175	1.4225	95
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif CCAWWNAAGG which matches annotation for SRF: serum response factor (c-fos serum response element-binding transcription factor)	2.0656e-13	1.8148e-11	5.0535	2.6439e-6	4.0624	21
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif VCCGGAAGNGCR which matches annotation for GABPA: GA binding protein transcription factor, alpha subunit 60kDa GABPB2: GA binding protein transcription factor, beta subunit 2	4.6931e-10	3.2069e-8	3.8196	0.0278	1.9635	28
	Genes with promoter regions [-2kb,2kb] around transcription start site containing motif NGGGACTTTCCA. Motif does not match any known transcription factor	1.6595e-6	8.5051e-5	2.4198	0.0107	2.0441	30
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif WTGCGTGGGCGK which matches annotation for EGR1: early growth response 1	3.4186e-6	0.0002	2.3093	0.0341	1.8925	28
TreeFam	FOSL2/JDP2/FOS/FOSL1/FOSB./ATF3	1.1291e-8	3.1134e-5	27.2523	0.0346	14.0249	5

Supplementary Table 9: GREAT enrichments of SRF using the *single nearest gene* association rule with a maximum extension of 1 Mb. Shown are the top ten binomial enriched terms at a false discovery rate of 0.05 with a fold enrichment of at least two that are also significant by the hypergeometric test, using the highest-scoring SRF peaks anywhere in the genome (QuEST score > 1; n = 556).

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Cellular Component	nuclear lumen	2.3328e-8	1.9455e-6	2.1917	0.0147	1.6980	50
	actin cytoskeleton	3.1098e-6	0.0001	3.2884	0.0078	2.7345	19
	cytosolic large ribosomal subunit	1.0174e-5	0.0004	9.7434	0.0104	6.1808	7
	cortical cytoskeleton	1.3415e-5	0.0005	7.6613	0.0279	6.1256	6
Pathway Commons	Further platelet releasate	3.6552e-6	0.0001	15.2060	0.0492	8.9100	6
MSigDB Pathway	Transcription factors enriched in fetal liver	8.6945e-10	3.0691e-7	6.4954	0.0218	4.4921	11
Transcription Factor Targets	Targets of SRF, identified by ChIP-chip in different cell lines: Jurkat, T/G HA-VSMC, and Be(2)-C cells.	6.5241e-79	1.2396e-77	19.2536	1.0002e-66	14.6519	74
	Targets of YY1 identified by ChIP-chip.	6.2083e-9	3.9319e-8	2.9484	0.0032	1.7983	36
	Targets of ETS1, identified by ChIP-chip in Jurkat T-cells.	5.7312e-7	2.7223e-6	2.1033	0.0078	1.5061	52
	Targets of NRF1, identified by ChIP-chip in quiescent T98G cells.	1.1505e-5	4.3720e-5	2.2756	0.0140	1.6208	32
	Genes that are bound by both E2F4 and p130 in three different growth arrest condtions, identified by ChIP-chip in T98G and U2OS cells under growth arrest.	0.0005	0.0016	2.7446	0.0031	2.6206	15
	Targets of HSF1, identified by ChIP-chip in HeLa cells under heat shock.	0.0021	0.0050	2.1214	0.0168	1.9031	18
Predicted Promoter Motifs	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif DCCWTATATGGNCWN which matches annotation for SRF: serum response factor (c-fos serum response element-binding transcription factor)	7.1743e-33	4.4122e-30	7.4699	3.0452e-23	7.2066	45
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif ATGCCCATATATGGWNNT which matches annotation for SRF: serum response factor (c-fos serum response element-binding transcription factor)	1.3569e-29	4.1725e-27	25.0321	5.6216e-13	11.5305	18
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif SCCAWATAWGGMNMNNNN which matches annotation for SRF: serum response factor (c-fos serum response element-binding transcription factor)	1.0567e-24	2.1663e-22	5.9846	2.5024e-15	5.6485	37
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif CCAWATAWGGMNMNG which matches annotation for SRF: serum response factor (c-fos serum response element-binding transcription factor)	1.5687e-22	2.4120e-20	5.4667	8.7953e-15	5.4959	36
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif GNCCAWATAWGGMN which matches annotation for SRF: serum response factor (c-fos serum response element-binding transcription factor)	1.2218e-20	1.5028e-18	4.8411	1.6842e-15	5.3311	39
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif SCGGAAGY which matches annotation for ELK1: ELK1, member of ETS oncogene family	9.2830e-17	9.5151e-15	2.9964	7.4345e-5	1.8602	64
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif VCCGGAAGNGCR which matches annotation for GABPA: GA binding protein transcription factor, alpha subunit 60kDa GABPB2: GA binding protein transcription factor, beta subunit 2	2.0867e-14	1.8333e-12	6.4613	2.9210e-5	3.2670	24
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif CCAWWNAAGG which matches annotation for SRF: serum response factor (c-fos serum response element-binding transcription factor)	5.0566e-14	3.8873e-12	7.1981	9.8285e-8	6.3838	17
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif MGGAAGTG which matches annotation for GABPA: GA binding protein transcription factor, alpha subunit 60kDa GABPB2: GA binding protein transcription factor, beta subunit 2	2.4939e-10	1.7042e-8	2.6804	0.0006	2.0021	44
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif ACCGGAAGNG which matches annotation for NFE2L2: nuclear factor (erythroid-derived 2)-like 2	1.1150e-6	5.7143e-5	3.9979	0.0260	2.3522	18
TreeFam	Early growth response protein	6.4549e-15	5.3395e-11	80.3988	0.0350	26.1359	4

Supplementary Table 10: Enrichments for regions bound by p300 in mouse limb. **(a)** DAVID gene-based enrichments of genes with proximal p300 binding events. **(b)** GREAT *cis*-regulatory element enrichments for all regions bound by p300.

a

DAVID Gene-based Enrichments of p300 Binding Peaks in Mouse Limb

Annotation Cluster 1	Enrichment Score: 6.31			Count	P_Value	Benjamini
SP_PIR_KEYWORDS	Transcription regulation	RT		43	4.6E-10	4.0E-7
SP_PIR_KEYWORDS	dna-binding	RT		44	6.2E-10	2.7E-7
SP_PIR_KEYWORDS	Transcription	RT		43	8.0E-10	2.3E-7
SP_PIR_KEYWORDS	nucleus	RT		74	1.5E-9	3.3E-7
GOTERM_CC_ALL	nucleus	RT		86	3.7E-9	2.9E-6
GOTERM_MF_ALL	transcription regulator activity	RT		41	6.4E-9	1.7E-5
GOTERM_BP_ALL	regulation of cellular process	RT		80	1.6E-8	8.2E-5
GOTERM_BP_ALL	regulation of biological process	RT		84	4.9E-8	1.3E-4
GOTERM_BP_ALL	regulation of transcription, DNA-dependent	RT		55	5.0E-8	8.7E-5
GOTERM_MF_ALL	DNA binding	RT		52	5.4E-8	7.3E-5
Annotation Cluster 2	**Enrichment Score: 3.4**			**Count**	**P_Value**	**Benjamini**
GOTERM_MF_ALL	transcription regulator activity	RT		41	6.4E-9	1.7E-5
GOTERM_MF_ALL	transcription factor activity	RT		30	9.2E-7	8.3E-4
GOTERM_BP_ALL	pattern specification process	RT		12	1.2E-4	2.6E-2
GOTERM_CC_ALL	nucleoplasm part	RT		19	1.3E-4	1.2E-2
GOTERM_MF_ALL	sequence-specific DNA binding	RT		17	1.9E-4	7.2E-2
GOTERM_CC_ALL	transcription factor complex	RT		16	2.0E-4	1.7E-2
GOTERM_CC_ALL	protein complex	RT		39	2.1E-4	1.7E-2
SP_PIR_KEYWORDS	Developmental protein	RT		20	2.7E-4	3.3E-2
GOTERM_CC_ALL	nucleoplasm	RT		19	2.8E-4	2.0E-2
Annotation Cluster 3	**Enrichment Score: 3.08**			**Count**	**P_Value**	**Benjamini**
GOTERM_BP_ALL	organ morphogenesis	RT		21	4.1E-6	1.1E-3
GOTERM_BP_ALL	anatomical structure morphogenesis	RT		31	5.4E-5	1.3E-2
GOTERM_BP_ALL	organ development	RT		34	9.9E-5	2.2E-2

b

GREAT Enrichments of p300 Binding Peaks in Mouse Limb

Ontology	Term	Binomial Results			Hypergeometric Results			Distal Binding% (> 10 kb from nearest TSS)
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits	
GO Molecular Function	transcription repressor activity	5.1566e-14	2.0455e-11	2.2015	0.0014	2.0428	41	87.4%
	extracellular matrix structural constituent	0.0004	0.0240	2.6247	0.0133	3.6537	11	94.1%
GO Biological Process	embryonic limb morphogenesis	1.9792e-30	1.1230e-27	3.8271	9.8517e-16	4.0786	44	79.2%
	limb morphogenesis	9.6264e-29	2.9129e-26	3.5370	5.6001e-16	3.8651	48	78.2%
	embryonic morphogenesis	1.7374e-28	4.9287e-26	2.2764	1.6596e-16	2.5054	99	85.3%
	limb development	3.0309e-28	8.0925e-26	3.3920	1.3037e-16	3.8698	50	78.9%
	skeletal system development	3.4240e-26	5.5506e-24	2.4141	8.4605e-20	2.8874	92	80.4%
	negative regulation of transcription, DNA-dependent	1.8616e-21	1.9651e-19	2.2922	1.4514e-7	2.1258	64	82.6%
	negative regulation of RNA metabolic process	2.3064e-21	2.3792e-19	2.2870	1.9380e-7	2.1082	64	82.6%
	negative regulation of transcription from RNA polymerase II promoter	5.0111e-21	4.8394e-19	2.4287	1.8092e-8	2.3666	57	80.9%
	negative regulation of gene expression	6.3047e-19	5.1102e-17	2.0679	4.8988e-8	2.0258	77	83.0%
	negative regulation of transcription	7.2652e-19	5.8013e-17	2.0911	1.3078e-7	2.0206	73	82.5%
Mouse Phenotype	abnormal craniofacial morphology	2.8489e-52	8.2588e-49	2.0985	3.0174e-32	2.4142	202	90.2%
	abnormal axial skeleton morphology	2.7970e-48	3.2434e-45	2.1069	3.4770e-32	2.4596	195	86.0%
	abnormal limbs/digits/tail morphology	3.3965e-46	2.8133e-43	2.1821	3.2913e-30	2.5099	176	86.8%
	abnormal skull morphology	7.5980e-44	5.5066e-41	2.3680	2.0707e-29	2.9252	131	88.4%
	abnormal craniofacial bone morphology	8.2628e-43	5.3360e-40	2.3048	6.0949e-30	2.8911	136	88.3%
	abnormal limb morphology	1.2698e-41	7.4781e-39	2.3530	2.9027e-26	2.7644	129	87.2%
	abnormal paw/hand/foot morphology	1.7751e-38	8.5767e-36	2.8485	7.1707e-22	3.2194	84	88.0%
	abnormal digit morphology	6.3264e-37	2.4454e-34	2.9851	5.4602e-20	3.3565	72	87.2%
	abnormal appendicular skeleton morphology	1.7529e-36	6.3520e-34	2.3454	6.1182e-25	2.8150	119	84.0%
	abnormal head morphology	7.1868e-36	2.4511e-33	2.1002	3.8552e-23	2.4321	144	89.1%
PANTHER pathway	TGF-beta signaling pathway	4.6289e-5	0.0034	2.0042	0.0164	2.2456	20	85.0%
MGI Expression: Detected	TS19_limb	3.3410e-52	7.4628e-49	2.7930	1.4719e-36	3.6576	117	86.3%
	TS19_forelimb bud	6.8017e-41	5.0643e-38	2.9673	4.6454e-32	4.1550	86	83.5%
	TS22_upper jaw	1.6641e-38	7.9650e-36	2.1906	5.7007e-17	2.1567	135	87.7%
	TS20_limb	2.4317e-38	1.0863e-35	2.3883	2.0866e-30	3.1727	119	87.6%
	TS19_hindlimb bud	6.5820e-37	2.4503e-34	3.1744	5.2268e-23	4.0831	63	87.2%
	TS20_forelimb	2.1505e-34	4.9691e-32	2.7208	6.9945e-24	3.4903	81	84.9%
	TS17_limb	3.7256e-33	7.8017e-31	2.2171	1.8401e-26	2.6509	139	86.1%
	TS22_palatal shelf	3.5386e-32	6.9701e-30	2.1655	4.3955e-12	2.0107	113	87.1%
	TS17_hindlimb bud	2.0087e-31	3.5421e-29	3.1294	1.9178e-22	4.3275	57	86.6%
	TS21_limb	9.0320e-31	1.5131e-28	2.5063	9.2670e-27	3.4323	93	84.3%
InterPro	MAD homology 1, Dwarfin-type	1.7109e-9	1.1940e-6	4.5958	0.0295	5.3145	8	95.8%

Supplementary Table 11: "Gene-based GREAT" enrichments of all genes that possess a p300 limb binding peak within 2 kb of its transcription start site. Shown are the top ten hypergeometric enriched terms at a false discovery rate of 0.05.

Ontology	Term	Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Observed Gene Hits
GO Molecular Function	transcription regulator activity	3.6557e-9	8.7007e-6	35
	DNA binding	3.6993e-9	4.4022e-6	47
	transcription factor activity	5.3544e-8	4.2478e-5	26
	nucleic acid binding	4.3081e-6	0.0026	51
	sequence-specific DNA binding	8.0480e-6	0.0038	18
	protein binding	6.0880e-5	0.0241	84
GO Biological Process	skeletal system development	9.2697e-10	4.2075e-6	18
	regulation of RNA metabolic process	1.4155e-9	3.2125e-6	52
	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	7.9548e-9	1.2036e-5	53
	regulation of transcription, DNA-dependent	9.3464e-9	1.0606e-5	50
	regionalization	1.4001e-8	1.2710e-5	15
	anatomical structure development	2.5013e-8	1.8923e-5	49
	regulation of metabolic process	3.2642e-8	2.1166e-5	58
	regulation of transcription	3.3637e-8	1.9085e-5	50
	regulation of cellular metabolic process	3.5466e-8	1.7887e-5	56
	regulation of gene expression	4.2431e-8	1.9259e-5	52
GO Cellular Component	nucleus	1.3518e-8	9.2736e-6	82
	membrane-bounded organelle	1.9918e-5	0.0068	103
	intracellular membrane-bounded organelle	3.4720e-5	0.0079	102
	organelle	3.6713e-5	0.0063	112
	intracellular organelle	6.3686e-5	0.0087	111
	intracellular	6.3946e-5	0.0073	130
	intracellular part	0.0002	0.0201	125
Mouse Phenotype	abnormal skeleton morphology	7.4292e-10	4.3074e-6	34
	abnormal axial skeleton morphology	1.0089e-9	2.9249e-6	28
	skeleton phenotype	3.2505e-9	6.2821e-6	34
	abnormal appendicular skeleton morphology	7.6078e-8	0.0001	18
	abnormal limbs/digits/tail morphology	1.3570e-7	0.0002	23
	abnormal skull morphology	1.8021e-7	0.0002	18
	abnormal skeleton extremities morphology	2.2887e-7	0.0002	17
	growth/size phenotype	3.0783e-7	0.0002	52
	abnormal craniofacial bone morphology	3.7211e-7	0.0002	18
	lethality-prenatal/perinatal	8.7640e-7	0.0005	47
MSigDB Pathway	The attachment of a cell, either to another cell or to the extracellular matrix, via cell adhesion molecules.	4.0477e-5	0.0185	10
MGI Expression: Detected	TS21_metanephros;excretory component;cortex;nephrons	6.0936e-9	4.0833e-5	6
	TS21_metanephros;excretory component;cortex	1.0059e-8	3.3701e-5	6
	TS28_tooth	1.1303e-8	2.5248e-5	25
	TS17_central nervous system	1.7965e-8	3.0096e-5	46
	TS21_organ system	1.8311e-8	2.4540e-5	52
	TS20_visceral organ	2.1460e-8	2.3968e-5	30
	TS17_nervous system	2.1561e-8	2.0640e-5	46
	TS21_metanephros;excretory component	3.6246e-8	3.0360e-5	6
	TS19_limb	4.5927e-8	3.4195e-5	16
	TS17_organ system	5.3151e-8	3.5616e-5	50
InterPro	Sequence-specific single-strand DNA-binding protein	1.5139e-6	0.0095	3
	Single-stranded DNA-binding protein, SSDP	1.5139e-6	0.0095	3
TreeFam	Homeobox protein	5.8622e-7	0.0047	4
	Single-stranded DNA-binding protein	1.5139e-6	0.0060	3
	DNA-binding protein inhibitor	6.0039e-6	0.0159	3

Supplementary Table 12: GREAT enrichments of all p300 limb peaks using the *basal plus extension* association rule with a basal regulatory region extending 5 kb upstream and 1 kb downstream of the transcription start site and a maximum extension of 50 kb. Shown are the top ten binomial enriched terms at a false discovery rate of 0.05 with a fold enrichment of at least two that are also significant by the hypergeometric test.

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Molecular Function	transcription regulator activity	5.6169e-13	1.3368e-9	2.0203	1.2322e-7	1.9388	98
	sequence-specific DNA binding	8.1588e-7	0.0003	2.0215	6.6705e-5	2.1673	51
	protein tyrosine kinase activity	0.0004	0.0465	2.3137	0.0489	2.5252	19
GO Biological Process	embryonic limb morphogenesis	1.6288e-20	7.3931e-17	6.4212	5.0357e-10	5.7567	25
	limb morphogenesis	1.1569e-19	2.6256e-16	5.9118	1.8612e-9	5.2008	26
	limb development	4.9014e-19	7.4159e-16	5.6788	4.1765e-9	4.9989	26
	embryonic morphogenesis	1.7101e-16	1.9406e-13	3.2874	2.8018e-7	2.7662	44
	skeletal system development	7.3647e-15	4.7755e-12	3.4967	2.0494e-10	3.5084	45
	embryonic development	8.5880e-12	4.3312e-9	2.2938	9.4936e-6	2.0685	61
	embryonic skeletal system development	4.0128e-11	1.3010e-8	5.3263	0.0004	4.0142	15
	organ morphogenesis	1.0976e-10	3.3213e-8	2.1752	1.3843e-7	2.2527	67
	negative regulation of transcription from RNA polymerase II promoter	1.2511e-10	3.5492e-8	3.3375	5.6948e-5	2.8880	28
	mammary gland development	2.7168e-9	4.2523e-7	7.0476	0.0113	4.8008	8
Mouse Phenotype	abnormal skeleton morphology	7.5433e-28	4.3736e-24	2.7202	3.5401e-17	2.5474	115
	skeleton phenotype	8.6486e-28	2.5072e-24	2.6645	2.3731e-17	2.4885	119
	abnormal appendicular skeleton morphology	9.4039e-26	1.8175e-22	3.8649	1.6816e-13	3.4083	58
	abnormal limbs/digits/tail morphology	5.9854e-25	8.6759e-22	3.0870	8.0748e-15	2.8695	81
	abnormal axial skeleton morphology	3.0148e-24	3.4959e-21	2.9220	1.4403e-15	2.7887	89
	abnormal long bone morphology	1.5957e-22	1.5420e-19	4.0174	3.1734e-11	3.5172	46
	abnormal skull morphology	3.8286e-22	3.1711e-19	3.4599	3.5024e-13	3.2728	59
	abnormal craniofacial bone morphology	4.8977e-22	3.5496e-19	3.3865	2.6738e-13	3.2213	61
	abnormal limb morphology	4.9646e-22	3.1983e-19	3.4152	2.4143e-13	3.2473	61
	abnormal skeleton extremities morphology	7.3329e-22	4.2516e-19	3.6231	1.5524e-11	3.2394	53
PANTHER Pathway	Angiogenesis	0.0002	0.0235	2.4564	0.0406	2.4573	17
	PDGF signaling pathway	0.0004	0.0222	2.6000	0.0656	2.8004	14
MGI Expression: Detected	TS20_visceral organ	1.0389e-19	6.9614e-16	2.4665	3.5592e-12	2.3610	98
	TS21_visceral organ	5.3113e-19	1.7796e-15	2.4464	7.6081e-11	2.2542	93
	TS19_embryo	4.0952e-18	9.1473e-15	2.0961	3.2546e-12	2.0625	128
	TS19_forelimb bud	4.7047e-18	7.8815e-15	4.4011	3.0380e-12	4.5607	38
	TS19_limb	3.3198e-17	4.4492e-14	3.5236	2.5667e-12	3.8053	49
	Theiler_stage_19	4.6587e-17	4.4597e-14	2.0391	2.6771e-11	1.9881	128
	TS17_embryo;mesenchyme	6.8990e-17	5.7787e-14	3.5109	7.8631e-11	3.4637	46
	TS20_forelimb	8.2565e-17	6.1474e-14	4.0308	7.6413e-11	4.0677	38
	TS20_limb	1.4352e-16	9.6170e-14	3.2641	5.8255e-11	3.3116	50
	TS19_hindlimb bud	5.6127e-15	2.2124e-12	4.5367	1.9081e-10	4.8300	30
InterPro	EGF-like calcium-binding, conserved site	1.2223e-5	0.0070	3.7003	0.1353	3.9006	13
	EGF calcium-binding	2.4595e-5	0.0119	3.4893	0.0945	3.7859	13
	EGF-like calcium-binding	3.8742e-5	0.0135	2.9752	0.0473	3.2332	16
	Calponin-like actin-binding	7.6706e-5	0.0219	3.3134	0.0945	3.7859	13
TreeFam	DNA-binding protein inhibitor	2.9770e-16	2.3676e-12	43.5910	0.0257	19.8032	4
	Homeobox protein	1.0887e-12	4.3291e-9	33.7101	0.0498	14.1451	5

Supplementary Table 13: GREAT enrichments of all p300 limb peaks using the *two nearest genes* association rule with a maximum extension of 1 Mb. Shown are the top ten binomial enriched terms at a false discovery rate of 0.05 with a fold enrichment of at least two that are also significant by the hypergeometric test.

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Molecular Function	transcription repressor activity	2.2361e-14	7.6028e-12	2.2121	0.0032	1.9555	42
	extracellular matrix structural constituent	0.0005	0.0231	2.5811	0.0161	3.4143	11
GO Biological Process	regulation of transcription from RNA polymerase II promoter	1.5185e-32	4.9231e-30	2.0043	2.6681e-13	1.9634	136
	embryonic limb morphogenesis	8.2964e-32	2.2151e-29	3.8863	1.9695e-15	3.8979	45
	embryonic morphogenesis	3.7150e-30	7.3315e-28	2.3116	5.8877e-17	2.4595	104
	limb morphogenesis	4.8919e-30	8.5401e-28	3.5904	1.5079e-15	3.6871	49
	limb development	1.6779e-29	2.6262e-27	3.4425	3.2959e-16	3.6885	51
	skeletal system development	1.1826e-25	1.4126e-23	2.3868	2.3382e-19	2.7862	95
	negative regulation of transcription, DNA-dependent	2.0548e-22	2.0276e-20	2.3164	1.3373e-7	2.0796	67
	negative regulation of RNA metabolic process	2.5791e-22	2.3413e-20	2.3110	1.7852e-7	2.0624	67
	negative regulation of transcription from RNA polymerase II promoter	1.0005e-21	8.5681e-20	2.4462	2.8140e-8	2.2891	59
	negative regulation of gene expression	1.0549e-19	8.1155e-18	2.0856	3.5211e-8	1.9914	81
Mouse Phenotype	abnormal craniofacial morphology	5.4778e-56	1.5880e-52	2.1345	6.2099e-34	2.3901	214
	abnormal axial skeleton morphology	1.0466e-49	1.0114e-46	2.1174	8.9529e-33	2.4045	204
	abnormal limbs/digits/tail morphology	1.6304e-48	1.3504e-45	2.2068	3.9785e-30	2.4387	183
	abnormal skull morphology	6.4357e-46	4.6643e-43	2.3867	1.8358e-29	2.8379	136
	abnormal craniofacial bone morphology	3.2635e-45	1.8922e-42	2.3389	2.0334e-30	2.8208	142
	abnormal limb morphology	2.2391e-42	1.1802e-39	2.3581	2.9425e-26	2.6834	134
	abnormal paw/hand/foot morphology	8.2520e-39	3.4175e-36	2.8459	1.1918e-20	3.0442	86
	abnormal appendicular skeleton morphology	1.5614e-38	6.0352e-36	2.3801	3.3973e-25	2.7410	124
	abnormal head morphology	3.6176e-37	1.3109e-34	2.1157	3.6570e-23	2.3674	150
	abnormal vertebrae morphology	5.3876e-37	1.7354e-34	2.7506	2.2193e-14	2.4930	84
PANTHER Pathway	TGF-beta signaling pathway	2.9388e-5	0.0015	2.0268	0.0298	2.0984	20
MGI Expression: Detected	TS19_limb	2.2356e-53	4.9936e-50	2.8063	1.5992e-35	3.4764	119
	TS19_forelimb bud	7.1220e-42	3.6711e-39	2.9852	8.3868e-31	3.9278	87
	TS22_upper jaw	1.7967e-40	7.5247e-38	2.2157	8.9634e-18	2.1348	143
	TS20_limb	7.4747e-39	2.2767e-36	2.3906	7.8311e-28	2.9648	119
	TS19_hindlimb bud	1.7814e-36	4.1162e-34	3.1471	2.8176e-22	3.8761	64
	TS20_forelimb	3.1187e-35	6.3329e-33	2.7356	5.3056e-22	3.2616	81
	TS17_limb	5.3195e-34	9.9017e-32	2.2279	1.2394e-24	2.5128	141
	TS22_palatal shelf	8.0176e-34	1.4521e-31	2.1902	1.0523e-12	1.9954	120
	TS22_tooth#	3.5421e-33	5.9340e-31	2.0962	2.7174e-12	1.9224	128
	TS22_jaw#	4.0153e-33	6.5626e-31	2.0946	3.1462e-12	1.9185	128
InterPro	High mobility group, HMG1/HMG2	8.5018e-11	5.9333e-8	2.7325	0.0358	2.7365	18
	MAD homology 1, Dwarfin-type	1.8831e-9	1.0752e-6	4.5728	0.0392	4.9662	8

Supplementary Table 14: GREAT enrichments of all p300 limb peaks using the *single nearest gene* association rule with a maximum extension of 1 Mb. Shown are the top ten binomial enriched terms at a false discovery rate of 0.05 with a fold enrichment of at least two that are also significant by the hypergeometric test.

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Molecular Function	transcription repressor activity	2.3908e-12	8.1288e-10	2.5356	0.0016	2.3669	31
GO Biological Process	embryonic limb morphogenesis	2.5708e-29	1.1669e-25	5.0461	1.2157e-16	5.3979	38
	limb morphogenesis	4.7533e-28	1.0788e-24	4.6297	5.2672e-16	4.9359	40
	limb development	1.1141e-27	1.6856e-24	4.3571	7.1912e-17	4.9814	42
	regulation of transcription from RNA polymerase II promoter	1.1096e-25	5.0364e-23	2.2123	3.2387e-13	2.3202	98
	negative regulation of transcription, DNA-dependent	2.5099e-19	4.0688e-17	2.6709	4.4165e-8	2.5451	50
	negative regulation of RNA metabolic process	2.8937e-19	4.5291e-17	2.6654	5.7807e-8	2.5240	50
	negative regulation of transcription from RNA polymerase II promoter	4.7733e-19	6.7706e-17	2.8608	2.1719e-8	2.7996	44
	embryonic morphogenesis	3.2235e-17	3.1131e-15	2.3069	3.6864e-13	2.7535	71
	regulation of cell proliferation	4.5127e-17	4.2673e-15	2.1152	1.3884e-11	2.3844	81
	skeletal system development	7.5513e-17	6.7206e-15	2.5210	1.3323e-12	2.9338	61
GO Cellular Component	transcription factor complex	4.0469e-8	2.1355e-6	2.0335	0.0031	2.0926	37
Mouse Phenotype	abnormal limbs/digits/tail morphology	2.3266e-37	1.3490e-33	2.4595	5.6301e-28	2.9284	134
	abnormal limb morphology	1.3929e-36	4.0380e-33	2.7658	4.0750e-26	3.3496	102
	abnormal appendicular skeleton morphology	4.4198e-35	6.4065e-32	2.8516	2.8255e-25	3.4438	95
	abnormal skeleton morphology	5.6527e-35	6.5549e-32	2.0918	2.1384e-28	2.4870	182
	abnormal craniofacial morphology	2.5246e-34	2.4396e-31	2.1960	4.0889e-28	2.7473	150
	skeleton phenotype	3.8589e-34	3.1962e-31	2.0487	4.3114e-28	2.4123	187
	abnormal paw/hand/foot morphology	1.6653e-32	1.2069e-29	3.4082	8.2344e-22	3.9938	68
	abnormal skeleton extremities morphology	2.1662e-32	1.3955e-29	2.7866	2.0898e-23	3.3934	90
	abnormal digit morphology	9.8712e-32	5.7233e-29	3.6051	6.0554e-21	4.2864	60
	abnormal axial skeleton morphology	3.9868e-31	2.1014e-28	2.1828	1.8916e-28	2.8221	146
MGI Expression: Detected	TS19_limb	2.2483e-31	1.5066e-27	2.8239	2.9298e-27	4.0242	84
	TS22_upper jaw	1.2669e-28	8.4893e-26	2.4002	1.7504e-15	2.4482	100
	TS19_hindlimb bud	4.3317e-27	2.2328e-24	3.4949	4.4759e-19	4.7673	48
	TS28_tooth	1.2030e-26	4.4784e-24	2.3475	1.0977e-13	2.2425	105
	TS21_embryo;head#	3.0887e-26	1.0894e-23	2.0030	1.5722e-17	2.1507	147
	TS20_limb	3.2784e-26	1.0984e-23	2.5442	3.5913e-26	3.6772	90
	TS19_forelimb bud	4.1487e-26	1.2637e-23	3.0787	1.9024e-23	4.5904	62
	TS17_limb	1.7165e-25	4.6009e-23	2.3786	2.1016e-20	2.8933	99
	TS22_lung	2.6541e-25	6.8405e-23	2.0191	6.0779e-15	2.1175	130
	TS22_metanephros	9.3375e-24	2.1576e-21	2.0035	1.6062e-13	2.0754	123
InterPro	High mobility group, HMG1/HMG2	1.2523e-8	1.5732e-5	3.1425	0.0158	3.7397	15

Supplementary Table 15: Enrichments for regions bound by p300 in mouse forebrain. **(a)** DAVID gene-based enrichments of genes with proximal p300 binding events. **(b)** GREAT *cis*-regulatory element enrichments for all regions bound by p300.

a

DAVID Gene-based Enrichments of p300 Binding Peaks in Mouse Forebrain

	Annotation Cluster 1	Enrichment Score: 5.03		Count	P_Value	Benjamini
	SP_PIR_KEYWORDS	Transcription	RT	33	4.5E-9	3.9E-6
	SP_PIR_KEYWORDS	Transcription regulation	RT	31	4.4E-8	1.9E-5
	GOTERM_CC_ALL	nucleus	RT	61	1.0E-7	8.1E-5
	GOTERM_BP_ALL	transcription	RT	42	3.5E-7	1.8E-3
	SP_PIR_KEYWORDS	nucleus	RT	51	4.4E-7	1.3E-4
	GOTERM_BP_ALL	regulation of gene expression	RT	42	8.4E-7	2.2E-3
	GOTERM_BP_ALL	RNA metabolic process	RT	44	9.4E-7	1.6E-3
	GOTERM_BP_ALL	regulation of transcription, DNA-dependent	RT	39	9.8E-7	1.3E-3
	GOTERM_BP_ALL	regulation of cellular process	RT	55	9.8E-7	1.0E-3
	GOTERM_BP_ALL	transcription, DNA-dependent	RT	39	1.3E-6	9.6E-4
	Annotation Cluster 2	**Enrichment Score: 4.29**		**Count**	**P_Value**	**Benjamini**
	GOTERM_CC_ALL	nucleus	RT	61	1.0E-7	8.1E-5
	SP_PIR_KEYWORDS	nucleus	RT	51	4.4E-7	1.3E-4
	GOTERM_CC_ALL	intracellular organelle	RT	86	9.6E-7	3.8E-4
	GOTERM_CC_ALL	organelle	RT	86	1.0E-6	2.6E-4
	GOTERM_CC_ALL	intracellular membrane-bound organelle	RT	74	7.1E-5	1.4E-2
	GOTERM_CC_ALL	membrane-bound organelle	RT	74	7.4E-5	1.2E-2
	GOTERM_CC_ALL	intracellular	RT	95	8.1E-5	1.0E-2
	GOTERM_CC_ALL	intracellular part	RT	91	8.8E-5	9.8E-3
	Annotation Cluster 3	**Enrichment Score: 3.72**		**Count**	**P_Value**	**Benjamini**
	GOTERM_BP_ALL	nervous system development	RT	20	1.1E-6	9.5E-4
	GOTERM_BP_ALL	developmental process	RT	47	1.5E-6	8.9E-4
	GOTERM_BP_ALL	multicellular organismal development	RT	38	5.4E-6	1.8E-3
	GOTERM_BP_ALL	central nervous system development	RT	11	3.3E-5	8.6E-3
	Annotation Cluster 4	**Enrichment Score: 2.51**		**Count**	**P_Value**	**Benjamini**
	GOTERM_BP_ALL	nervous system development	RT	20	1.1E-6	9.5E-4
	GOTERM_BP_ALL	forebrain development	RT	8	1.6E-5	4.3E-3
	GOTERM_BP_ALL	central nervous system development	RT	11	3.3E-5	8.6E-3
	GOTERM_BP_ALL	brain development	RT	9	1.7E-4	4.0E-2

b

GREAT Enrichments of p300 Binding Peaks in Mouse Forebrain

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Molecular Function	transcription activator activity	9.8138e-17	3.3367e-14	2.1671	0.0054	1.7854	50
	RNA polymerase II transcription factor activity, enhancer binding	2.3185e-10	5.5180e-8	2.8890	0.0009	3.5351	16
	chromatin binding	6.3828e-9	1.3810e-6	2.0208	0.0006	2.2133	36
	RNA polymerase II transcription factor activity	1.5151e-8	3.0049e-6	2.2535	0.0053	2.3257	25
	non-G-protein coupled 7TM receptor activity	2.3884e-6	0.0004	4.2285	0.0144	4.9492	7
GO Biological Process	central nervous system development	2.9203e-38	8.2845e-36	2.3002	3.1913e-25	2.8137	117
	brain development	5.2647e-31	7.7085e-29	2.3400	8.9757e-19	2.7614	91
	forebrain development	2.4494e-27	2.4169e-25	2.6771	1.7658e-12	2.8087	58
	positive regulation of transcription, DNA-dependent	1.1227e-26	9.9921e-25	2.0015	6.5891e-13	2.1333	105
	positive regulation of RNA metabolic process	1.1562e-26	1.0092e-24	2.0011	7.9606e-13	2.1272	105
	positive regulation of transcription from RNA polymerase II promoter	1.7927e-26	1.4530e-24	2.0408	2.0528e-13	2.2292	99
	cell fate commitment	2.3145e-24	1.6675e-22	2.4560	8.3241e-18	3.2826	65
	Wnt receptor signaling pathway	6.1087e-21	4.2658e-19	2.8889	0.0013	1.9942	33
	pallium development	3.0346e-19	2.0558e-17	3.6197	0.0002	2.9203	19
	telencephalon development	2.4607e-18	1.6187e-16	3.2370	1.0366e-5	2.8977	25
Mouse Phenotype	abnormal neurogenesis	2.3609e-38	2.7377e-35	2.4125	1.2103e-29	3.2517	109
	abnormal brain white matter morphology	4.0745e-36	3.3748e-33	2.8781	2.0405e-20	3.7503	61
	abnormal tract	1.3838e-34	8.9150e-32	2.8782	3.8695e-21	3.8919	60
	abnormal brain commissure morphology	7.3498e-33	3.5512e-30	2.9718	3.5567e-18	3.8700	52
	abnormal corpus callosum morphology	5.2543e-30	2.1760e-27	3.2697	9.2958e-14	3.8837	39
	abnormal dorsal telencephalic commissure morphology	5.7327e-30	2.2159e-27	3.2480	8.7070e-14	3.8218	40
	abnormal cerebrum morphology	1.5317e-29	5.5506e-27	2.0824	4.1948e-21	2.5342	119
	abnormal diencephalon morphology	3.1726e-29	1.0820e-26	2.4724	7.6801e-17	2.7666	81
	abnormal cerebral cortex morphology	1.3363e-26	3.8740e-24	2.2773	2.4136e-16	2.6632	84
	abnormal brain ventricle/choroid plexus morphology	1.2743e-24	3.3585e-22	2.4304	1.5390e-12	2.6028	67
PANTHER pathway	Notch signaling pathway	8.0267e-6	0.0012	2.8632	0.0094	2.7853	13
MSigDB pathway	Wnt signaling genes	6.7099e-6	0.0010	2.1432	0.0180	2.4380	20
MGI Expression: Detected	TS21_thalamus	6.8849e-52	5.7670e-49	2.4054	1.8271e-32	2.8281	146
	TS17_brain	2.6467e-49	1.2668e-46	2.0581	6.1376e-38	2.5065	212
	TS21_midbrain	4.5075e-48	1.7767e-45	2.3606	3.6552e-28	2.6490	142
	TS21_hypothalamus	1.8893e-47	6.0287e-45	2.4920	9.0526e-28	2.8513	123
	TS21_cerebral cortex	4.8221e-46	1.3464e-43	2.0738	3.9298e-19	1.9652	184
	TS21_diencephalon	2.6050e-44	6.7140e-42	2.1540	7.0221e-32	2.6639	159
	TS17_forebrain	2.7451e-43	6.3431e-41	2.2704	2.1686e-39	3.2979	139
	TS15_central nervous system	7.2934e-40	1.3576e-37	2.1206	6.6358e-32	2.7547	150
	TS15_nervous system	2.5730e-39	4.6600e-37	2.1073	6.2093e-31	2.7055	150
	TS22_nasal cavity	1.3011e-38	2.2356e-36	2.0188	7.0372e-17	1.9377	168
InterPro	Basic helix-loop-helix dimerisation region bHLH	2.1957e-12	2.7582e-9	2.3208	0.0011	2.2525	36
	Helix-loop-helix DNA-binding	1.0584e-7	3.6934e-5	2.1762	0.0197	2.1627	26
	Frizzled related	8.6113e-7	0.0003	3.7554	0.0027	4.7135	10
	Frizzled protein	2.3884e-6	0.0007	4.2285	0.0228	4.9492	7

Supplementary Table 16: "Gene-based GREAT" enrichments of all genes that possess a p300 forebrain binding peak within 2 kb of its transcription start site. Shown are the top ten hypergeometric enriched terms at a false discovery rate of 0.05.

Ontology	Term	Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Observed Gene Hits
GO Biological Process	forebrain development	5.3984e-8	0.0002	10
	brain development	5.7312e-8	0.0001	12
	regulation of RNA metabolic process	5.6528e-7	0.0009	32
	developmental process	5.6682e-7	0.0006	40
	multicellular organismal development	5.9016e-7	0.0005	35
	central nervous system development	7.0350e-7	0.0005	12
	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	1.0877e-6	0.0007	33
	transcription	1.3867e-6	0.0008	29
	nervous system development	1.4906e-6	0.0008	18
	regulation of transcription, DNA-dependent	1.5149e-6	0.0007	31
GO Cellular Component	nucleus	1.2067e-5	0.0083	48
Mouse Phenotype	abnormal neurogenesis	6.9075e-8	0.0004	12
	abnormal nervous system development	1.4332e-7	0.0004	20
	abnormal nervous system morphology	1.7663e-6	0.0034	28
	abnormal neuron number	2.0286e-6	0.0029	10
	abnormal telencephalon morphology	2.3922e-6	0.0028	14
	abnormal telencephalon development	2.4874e-6	0.0024	7
	increased cochlear inner hair cell number	2.5283e-6	0.0021	4
	abnormal forebrain morphology	2.8269e-6	0.0020	16
	increased cochlear outer hair cell number	4.3625e-6	0.0028	4
	nervous system phenotype	5.1339e-6	0.0030	31
PANTHER Pathway	Notch signaling pathway	6.9011e-5	0.0103	4
MGI Expression: Detected	TS22_telencephalon	2.3933e-11	1.6038e-7	27
	TS22_eye	5.1719e-11	1.7328e-7	29
	TS22_sensory organ	9.8128e-11	2.1918e-7	31
	TS22_forebrain	1.6560e-10	2.7742e-7	28
	TS22_retina	8.6109e-10	1.1540e-6	25
	TS17_telencephalon	1.2685e-9	1.4167e-6	12
	TS17_forebrain	1.5400e-9	1.4743e-6	15
	TS17_brain	3.0625e-9	2.5652e-6	20
	TS22_organ system	3.1382e-9	2.3365e-6	45
	TS22_embryo	4.8295e-9	3.2363e-6	48

Supplementary Table 17: GREAT enrichments of all p300 forebrain peaks using the *basal plus extension* association rule with a basal regulatory region extending 5 kb upstream and 1 kb downstream of the transcription start site and a maximum extension of 50 kb. Shown are the top ten binomial enriched terms at a false discovery rate of 0.05 with a fold enrichment of at least two that are also significant by the hypergeometric test.

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Molecular Function	sequence-specific DNA binding	5.6365e-11	6.7074e-8	2.3030	3.6427e-11	2.7938	65
	transcription factor activity	7.0053e-11	5.5575e-8	2.0469	2.7930e-11	2.4698	82
GO Biological Process	nervous system development	9.3953e-26	4.2645e-22	2.6618	3.8931e-24	3.1201	112
	generation of neurons	6.5209e-20	1.4799e-16	2.8982	1.2514e-16	3.3149	71
	neurogenesis	1.1914e-19	1.8026e-16	2.8182	7.5612e-17	3.2504	74
	brain development	2.6667e-18	3.0260e-15	3.5475	2.4341e-13	3.9544	46
	neuron differentiation	2.9767e-18	2.7022e-15	3.0122	3.9547e-14	3.3958	59
	forebrain development	6.2971e-17	4.0832e-14	4.1323	1.2948e-12	4.8017	35
	cell development	1.8253e-16	9.2055e-14	2.2851	3.5652e-16	2.6936	94
	central nervous system development	2.6074e-16	1.1835e-13	3.0445	3.3369e-13	3.5427	52
	positive regulation of Notch signaling pathway	1.2509e-11	3.5488e-9	34.3459	0.0049	20.0298	3
	neuron migration	5.8788e-11	1.4824e-8	5.0207	1.8633e-6	5.4318	16
Mouse Phenotype	abnormal neurogenesis	9.2653e-27	5.3720e-23	4.1798	1.3047e-17	4.4792	53
	abnormal brain commissure morphology	5.2512e-21	1.5223e-17	5.5353	7.9443e-15	6.5360	31
	abnormal nervous system development	1.6112e-20	3.1140e-17	2.4118	1.7936e-20	2.8690	108
	abnormal tract	8.1442e-19	1.1805e-15	4.8786	3.8195e-13	5.6965	31
	abnormal brain white matter morphology	1.0514e-18	1.0160e-15	4.7502	2.8238e-13	5.5735	32
	abnormal telencephalon morphology	2.6908e-18	2.2287e-15	2.7547	1.6810e-14	3.0747	70
	abnormal forebrain morphology	5.5993e-18	4.0581e-15	2.4731	1.6478e-16	2.8947	87
	abnormal corpus callosum morphology	1.1435e-17	7.3665e-15	5.7913	8.2744e-13	7.0527	25
	abnormal thalamus morphology	1.8060e-17	1.0471e-14	7.9364	5.3193e-9	7.4023	17
	abnormal neuron morphology	1.8713e-17	9.8636e-15	2.2283	2.0600e-15	2.5098	103
PANTHER Pathway	Angiogenesis	0.0002	0.0320	2.2995	0.0023	2.9241	20
MGI Expression: Detected	TS17_brain	3.9148e-27	2.6233e-23	2.9737	4.9790e-26	3.4834	104
	TS17_forebrain	3.0409e-26	1.0188e-22	3.7605	1.2149e-25	4.7722	71
	TS12_embryo;ectoderm;neural ectoderm	3.7467e-25	8.3689e-22	5.1954	2.6554e-16	5.3504	39
	TS12_embryo;ectoderm	5.2856e-25	8.8546e-22	5.0760	3.5752e-16	5.1690	40
	TS17_hindbrain	9.1765e-25	1.2298e-21	3.8903	1.4691e-22	4.7763	62
	TS17_telencephalon	1.2028e-24	1.3434e-21	4.6996	4.7914e-22	5.9124	49
	TS20_central nervous system	2.8007e-24	2.6811e-21	2.6746	8.9760e-23	3.1040	106
	TS20_brain	5.1619e-24	4.3237e-21	2.8496	4.4517e-22	3.2801	94
	TS20_nervous system	8.3027e-24	6.1818e-21	2.6202	7.5020e-23	3.0511	108
	TS22_telencephalon	8.3940e-23	5.6248e-20	2.3754	8.4823e-23	2.7964	123
InterPro	Basic helix-loop-helix dimerisation region bHLH	1.4452e-6	0.0030	2.9976	0.0011	3.5451	20
	Homeobox	5.7503e-6	0.0072	2.2878	6.4240e-6	3.0751	37
	Homeodomain-related	2.6498e-5	0.0166	2.0180	3.0618e-5	2.6663	41
	Helix-turn-helix motif, lambda-like repressor	5.3286e-5	0.0279	3.2664	0.0008	4.4511	16

Supplementary Table 18: GREAT enrichments of all p300 forebrain peaks using the *two nearest genes* association rule with a maximum extension of 1 Mb. Shown are the top ten binomial enriched terms at a false discovery rate of 0.05 with a fold enrichment of at least two that are also significant by the hypergeometric test.

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Molecular Function	transcription activator activity	1.3202e-16	4.4886e-14	2.1520	0.0110	1.7262	50
	RNA polymerase II transcription factor activity, enhancer binding	3.0994e-10	7.3766e-8	2.8629	0.0013	3.4178	15
	chromatin binding	2.6117e-9	5.6508e-7	2.0425	0.0012	2.1398	36
	RNA polymerase II transcription factor activity	9.8882e-9	1.9612e-6	2.2634	0.0037	2.3385	26
	non-G-protein coupled 7TM receptor activity	2.8037e-6	0.0004	4.1745	0.0180	4.7849	7
GO Biological Process	central nervous system development	7.4503e-38	1.9892e-35	2.2837	6.1973e-24	2.7203	117
	brain development	9.4748e-31	1.2267e-28	2.3246	8.7016e-18	2.6697	91
	forebrain development	2.7816e-27	2.5252e-25	2.6636	7.1774e-12	2.7155	58
	positive regulation of transcription, DNA-dependent	6.0624e-27	5.1919e-25	2.0006	2.1403e-12	2.0821	106
	positive regulation of RNA metabolic process	6.3071e-27	5.3015e-25	2.0001	2.5879e-12	2.0761	106
	positive regulation of transcription from RNA polymerase II promoter	8.2969e-27	6.4930e-25	2.0417	5.8283e-13	2.1769	100
	cell fate commitment	9.5530e-25	6.8827e-23	2.4613	5.0614e-17	3.1737	65
	Wnt receptor signaling pathway	1.6230e-20	1.1334e-18	2.8488	0.0023	1.9280	33
	pallium development	5.2235e-19	3.4867e-17	3.5816	0.0003	2.8234	19
	telencephalon development	4.6088e-18	2.9885e-16	3.1996	1.8938e-5	2.8015	25
Mouse Phenotype	abnormal neurogenesis	1.8436e-38	2.1378e-35	2.4062	2.5321e-28	3.1438	109
	abnormal brain white matter morphology	3.6478e-36	2.6437e-33	2.8706	1.2752e-19	3.6258	61
	abnormal tract	1.1887e-34	6.8920e-32	2.8713	2.5569e-20	3.7627	60
	abnormal brain commissure morphology	5.4024e-33	2.8476e-30	2.9678	1.6510e-17	3.7416	52
	abnormal corpus callosum morphology	3.1362e-30	1.2988e-27	3.2695	2.9945e-13	3.7548	39
	abnormal dorsal telencephalic commissure morphology	3.5282e-30	1.3638e-27	3.2469	2.8164e-13	3.6949	40
	abnormal cerebrum morphology	1.6071e-29	5.8237e-27	2.0750	2.5126e-20	2.4707	120
	abnormal diencephalon morphology	1.2133e-28	4.1382e-26	2.4447	1.4691e-16	2.7078	82
	abnormal cerebral cortex morphology	2.0049e-26	5.8122e-24	2.2641	1.9301e-15	2.5748	84
	abnormal brain ventricle/choroid plexus morphology	2.2779e-25	6.0033e-23	2.4477	2.4143e-12	2.5540	68
PANTHER Pathway	Notch signaling pathway	1.1541e-6	0.0002	3.0355	0.0131	2.6928	13
MSigDB Pathway	Wnt signaling genes	4.2290e-6	0.0010	2.1627	0.0290	2.3571	20
MGI Expression: Detected	TS21_thalamus	1.2947e-51	9.6398e-49	2.3911	5.1098e-32	2.7717	148
	TS17_brain	9.9512e-50	4.7631e-47	2.0573	2.8939e-37	2.4576	215
	TS21_midbrain	8.6056e-48	3.2037e-45	2.3361	9.6675e-28	2.5972	144
	TS21_hypothalamus	2.3402e-47	7.8408e-45	2.4800	1.3592e-27	2.8015	125
	TS21_cerebral cortex	3.5600e-47	1.0843e-44	2.0799	1.1938e-18	1.9309	187
	TS21_diencephalon	5.5268e-44	1.3717e-41	2.1411	2.8697e-31	2.6079	161
	TS17_forebrain	1.0799e-43	2.4954e-41	2.2704	4.0758e-39	3.2343	141
	TS15_central nervous system	1.9176e-40	3.6714e-38	2.1232	2.0708e-31	2.6987	152
	TS15_nervous system	7.1368e-40	1.2585e-37	2.1095	1.9353e-30	2.6605	152
	TS22_pallidum	2.4780e-38	3.3887e-36	2.2584	3.8420e-18	2.2160	130
InterPro	Basic helix-loop-helix dimerisation region bHLH	8.3285e-13	1.0462e-9	2.3381	0.0008	2.2382	37
	Helix-loop-helix DNA-binding	3.4567e-8	1.2772e-5	2.2168	0.0137	2.1713	27
	Frizzled related	1.0401e-6	0.0003	3.7084	0.0034	4.5571	10
	Frizzled protein	2.8037e-6	0.0008	4.1745	0.0274	4.7849	7

Supplementary Table 19: GREAT enrichments of all p300 forebrain peaks using the *single nearest gene* association rule with a maximum extension of 1 Mb. Shown are the top ten binomial enriched terms at a false discovery rate of 0.05 with a fold enrichment of at least two that are also significant by the hypergeometric test.

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Molecular Function	transcription regulator activity	8.6250e-42	2.0528e-38	2.0767	5.4690e-26	2.2401	202
	transcription factor activity	1.0425e-29	6.2030e-27	2.0642	1.8859e-22	2.4539	147
	sequence-specific DNA binding	2.6933e-25	1.2820e-22	2.1660	2.4230e-18	2.5965	109
	RNA polymerase II transcription factor activity, enhancer binding	2.2185e-10	7.5429e-8	3.6222	0.0066	4.0703	11
	RNA polymerase II transcription factor activity	1.0935e-8	2.8916e-6	2.7570	0.0460	2.4831	17
	transcription activator activity	2.3165e-8	5.5132e-6	2.0201	0.0274	1.9062	34
	transcription corepressor activity	7.8935e-8	1.7079e-5	2.7571	0.0062	3.1160	16
	transcription repressor activity	9.1033e-8	1.8055e-5	2.0269	0.0013	2.2895	33
	chromatin binding	2.1428e-7	3.6427e-5	2.2380	0.0063	2.4132	25
	transcription cofactor activity	4.2886e-6	0.0006	2.0110	0.0162	2.2032	26
GO Biological Process	nervous system development	6.6255e-37	5.0122e-34	2.0018	3.9404e-42	2.9643	192
	central nervous system development	2.2803e-32	7.3930e-30	2.6161	1.7035e-22	3.3227	88
	forebrain development	1.3275e-30	3.1714e-28	3.4644	1.3022e-14	3.7256	49
	regulation of transcription from RNA polymerase II promoter	1.7584e-29	3.6278e-27	2.2088	5.1864e-19	2.5170	117
	brain development	5.9863e-29	1.1814e-26	2.7760	1.5386e-18	3.3827	71
	neurogenesis	6.3038e-27	1.1445e-24	2.0814	8.7502e-28	3.0186	124
	generation of neurons	4.4489e-26	6.7311e-24	2.0854	1.7130e-27	3.0792	119
	neuron differentiation	7.5793e-24	9.8292e-22	2.1530	8.1487e-25	3.2218	101
	negative regulation of transcription	4.6702e-20	5.1702e-18	2.4274	3.0146e-8	2.3127	60
	negative regulation of gene expression	1.5699e-19	1.6572e-17	2.3645	1.2799e-8	2.3081	63
GO Cellular Component	transcription factor complex	2.7781e-10	1.9058e-8	2.1137	1.0270e-5	2.3127	45
Mouse Phenotype	abnormal nervous system development	8.3961e-44	4.8681e-40	2.1738	1.2352e-38	2.8120	191
	abnormal neurogenesis	7.4807e-37	2.1686e-33	2.8765	2.0472e-28	4.0281	86
	abnormal forebrain morphology	9.5574e-32	9.2356e-29	2.1376	1.3564e-28	2.7475	149
	abnormal telencephalon morphology	3.7471e-30	2.7157e-27	2.2711	1.0295e-24	2.8969	119
	abnormal brain white matter morphology	1.3452e-29	8.6660e-27	3.2734	4.2628e-21	4.9230	51
	abnormal tract	8.6801e-29	5.0327e-26	3.3016	1.3477e-20	4.9903	49
	abnormal brain commissure morphology	1.0931e-28	5.7619e-26	3.4674	3.4747e-19	5.1414	44
	abnormal brain development	1.3153e-28	6.3552e-26	2.2505	5.8928e-26	2.9377	122
	abnormal cerebrum morphology	5.0801e-27	2.1039e-24	2.3944	2.2651e-20	3.0427	91
	abnormal dorsal telencephalic commissure morphology	6.6515e-26	2.5710e-23	3.8206	1.3998e-15	5.2504	35
MGI Expression: Detected	TS22_telencephalon	1.5017e-46	3.3544e-43	2.1486	1.6396e-37	2.6083	207
	TS22_forebrain	1.1983e-44	2.0074e-41	2.0055	8.3887e-42	2.5683	239
	TS22_cerebral cortex	2.1689e-42	2.4223e-39	2.2038	1.3275e-28	2.4716	175
	TS21_thalamus	1.6637e-40	1.2387e-37	2.6833	2.0805e-30	3.4063	112
	TS17_brain	5.0178e-39	2.4017e-36	2.2089	3.9075e-39	3.0815	166
	TS15_organ system	7.2861e-39	2.8720e-36	2.1076	4.3027e-37	2.8306	179
	TS21_embryo;head#	9.3468e-38	3.4796e-35	2.1420	2.2480e-22	2.2335	168
	TS15_central nervous system	1.1730e-36	4.1368e-34	2.4359	1.2084e-34	3.5177	122
	TS17_forebrain	2.8214e-36	9.0028e-34	2.5454	3.6953e-40	4.2094	113
	TS15_nervous system	5.4168e-36	1.6499e-33	2.4117	8.1279e-34	3.4548	122
InterPro	High mobility group, HMG1/HMG2	6.1670e-28	3.8735e-24	5.3934	0.0019	3.6248	16
	Basic helix-loop-helix dimerisation region bHLH	9.4465e-13	1.4833e-9	2.9176	0.0005	2.7506	28
	Homeobox	5.7525e-10	4.5164e-7	2.0233	4.7506e-8	2.6255	57
	Winged helix repressor DNA-binding	1.4045e-7	7.3514e-5	2.0658	0.0220	2.0239	33
	Helix-loop-helix DNA-binding	5.5638e-7	0.0002	2.5390	0.0376	2.4814	19
	Transcription factor, fork head	1.4963e-5	0.0049	2.6182	0.0253	3.2798	13
	Helix-turn-helix motif, lambda-like repressor	0.0001	0.0233	2.3513	0.0062	2.9294	19

Supplementary Table 20: Enrichments for regions bound by p300 in mouse midbrain. **(a)** DAVID gene-based enrichments of genes with proximal p300 binding events. **(b)** GREAT *cis*-regulatory element enrichments for all regions bound by p300.

a

No terms were found significant after multiple hypothesis correction in DAVID's gene-based test.

b

GREAT Enrichments of p300 Binding Peaks in Mouse Midbrain

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Biological Process	nervous system development	9.2743e-16	1.4032e-12	2.0118	1.1714e-21	3.0999	103
	compartment specification	7.5434e-11	4.8913e-8	56.6618	0.0003	21.6390	4
	embryonic morphogenesis	3.6243e-9	1.2654e-6	2.3118	1.8393e-5	2.5417	37
	neuron differentiation	4.7647e-9	1.5448e-6	2.0688	2.3965e-11	3.2334	52
	central nervous system development	7.4432e-9	2.2523e-6	2.1925	4.4324e-10	3.3121	45
	neural tube development	6.8274e-8	1.8229e-5	3.7266	5.7239e-5	4.5716	15
	negative regulation of cell differentiation	4.5097e-7	6.8232e-5	2.5632	8.6419e-6	3.4024	25
	central nervous system neuron axonogenesis	6.7001e-7	8.9446e-5	6.3792	0.0012	9.2739	6
	neuron fate commitment	6.8025e-7	8.8218e-5	3.7782	6.5742e-5	5.6450	12
	Wnt receptor signaling pathway	1.7775e-6	0.0002	3.0076	0.0126	2.7742	15
Mouse Phenotype	abnormal brain white matter morphology	3.1332e-15	1.8167e-11	3.5860	2.8509e-15	6.3976	34
	abnormal tract	3.3884e-14	9.8231e-11	3.5351	1.3403e-13	6.1542	31
	abnormal corpus callosum morphology	1.0113e-12	1.1727e-9	4.1327	1.9290e-9	6.4003	21
	abnormal brain commissure morphology	1.0792e-12	1.0429e-9	3.5513	9.2371e-12	6.1500	27
	abnormal dorsal telencephalic commissure morphology	1.5312e-12	1.2682e-9	4.0734	4.0557e-9	6.1408	21
	abnormal telencephalon morphology	1.9685e-10	1.2681e-7	2.1070	1.1697e-11	2.9422	62
	abnormal neural tube morphology/development	7.1653e-10	3.7768e-7	2.2260	4.2634e-7	2.5710	48
	abnormal brain ventricle/choroid plexus morphology	1.7941e-9	8.6684e-7	2.7999	1.8626e-8	3.8047	32
	abnormal brain ventricle morphology	2.4459e-9	1.0909e-6	3.0550	1.1960e-7	4.1675	26
	abnormal brain development	2.8577e-9	1.1835e-6	2.0208	4.0864e-9	2.6755	57
PANTHER Pathway	Notch signaling pathway	1.2115e-7	1.8052e-5	6.7578	0.0010	5.9016	9
	Cadherin signaling pathway	3.2652e-5	0.0024	2.8019	0.0002	4.1217	16
	Angiogenesis	0.0002	0.0098	2.2468	0.0221	2.5272	16
	Alzheimer disease-presenilin pathway	0.0004	0.0167	2.5829	0.0191	2.9847	12
MGI Expression: Detected	TS15_central nervous system	1.5649e-18	1.0487e-14	2.6193	3.1569e-20	3.9906	71
	TS15_nervous system	2.3785e-18	7.9692e-15	2.6029	4.7789e-20	3.9193	71
	TS24_sensory organ	4.1014e-18	9.1611e-15	2.4085	2.5112e-11	2.4010	80
	TS21_diencephalon	1.7600e-17	2.9485e-14	2.5099	1.6111e-17	3.5381	69
	TS22_retina	4.2770e-17	5.7320e-14	2.1061	8.3204e-15	2.4640	101
	TS21_thalamus	5.2927e-17	5.9110e-14	2.6515	3.7795e-17	3.7350	63
	TS15_future brain	8.4557e-17	8.0945e-14	2.7516	6.3490e-20	4.4617	60
	TS24_neural retina	1.3797e-16	1.1557e-13	2.7280	2.6578e-10	2.7049	59
	TS24_eye	6.5700e-16	4.0023e-13	2.3875	8.7481e-11	2.4567	73
	TS24_retina	4.2469e-15	1.6740e-12	2.4456	4.0265e-10	2.4868	67
InterPro	High mobility group, HMG1/HMG2	7.2904e-7	0.0046	3.7610	0.0421	4.4161	10

Supplementary Table 21: "Gene-based GREAT" enrichments of all genes that possess a p300 midbrain binding peak within 2 kb of its transcription start site. Shown are the top ten hypergeometric enriched terms at a false discovery rate of 0.05.

Ontology	Term	Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Observed Gene Hits
Pathway Commons	NOTCH	0.0005	0.0422	2
MGI Expression: Detected	TS17_rhombomere 04	7.2518e-6	0.0486	3
	TS19_future spinal cord	7.9060e-6	0.0265	5

Supplementary Table 22: GREAT enrichments of all p300 midbrain peaks using the *basal plus extension* association rule with a basal regulatory region extending 5 kb upstream and 1 kb downstream of the transcription start site and a maximum extension of 50 kb. Shown are the top ten binomial enriched terms at a false discovery rate of 0.05 with a fold enrichment of at least two that are also significant by the hypergeometric test.

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Biological Process	compartment specification	3.4224e-13	1.5534e-9	123.4873	3.2236e-5	69.1934	4
	nervous system development	4.8803e-10	1.1076e-6	3.1246	1.3251e-6	3.3683	35
	pattern specification process	6.7342e-9	1.0189e-5	4.8099	0.0006	4.2418	16
	regionalization	1.2986e-8	1.4736e-5	5.4395	0.0025	4.4311	13
	neuron differentiation	1.7491e-8	1.5878e-5	3.8738	0.0001	3.9766	20
	neurogenesis	4.1338e-8	2.6805e-5	3.3827	0.0001	3.4900	23
	generation of neurons	4.7179e-8	2.6768e-5	3.4562	0.0002	3.5484	22
	system development	5.2215e-8	2.6334e-5	2.1244	1.0637e-5	2.2281	55
	negative regulation of neuron differentiation	1.0694e-7	4.4127e-5	14.8047	0.0127	11.5323	5
	anterior/posterior pattern formation	1.5514e-7	5.8681e-5	6.0353	0.0112	4.8051	10
Mouse Phenotype	abnormal nervous system development	1.0022e-11	5.8108e-8	3.3143	2.2649e-8	3.5790	39
	abnormal brain morphology	2.6041e-10	7.5493e-7	2.8170	6.2619e-7	2.9624	42
	abnormal neurogenesis	6.1625e-10	1.1910e-6	5.0043	1.6068e-5	5.2552	18
	increased cochlear outer hair cell number	5.3865e-8	7.8078e-5	21.6228	0.0203	16.2609	4
	abnormal nervous system morphology	5.9841e-8	6.9391e-5	2.1497	1.1652e-5	2.3036	53
	abnormal neural tube morphology/development	8.5420e-8	8.2545e-5	3.5607	0.0001	3.7680	22
	abnormal brain development	1.2262e-7	0.0001	3.2921	0.0002	3.4522	23
	abnormal cochlear inner hair cell number	1.4001e-7	0.0001	18.7499	0.0277	13.8387	4
	increased cochlear hair cell number	2.7189e-7	0.0002	16.9722	0.0475	11.5323	4
	fused dorsal root ganglion	6.2119e-7	0.0003	32.5026	0.0332	23.0645	3
PANTHER Pathway	Notch signaling pathway	9.7994e-8	1.4601e-5	14.9774	0.0154	10.4839	5
MGI Expression: Detected	TS19_nervous system	5.4084e-15	3.6241e-11	5.2381	4.2400e-11	5.8474	30
	TS19_central nervous system	5.6423e-15	1.8905e-11	5.4007	3.0586e-11	6.0078	29
	TS19_future spinal cord	6.3614e-15	1.4209e-11	7.4921	2.3349e-11	8.6492	22
	TS15_future spinal cord	2.0744e-13	3.4752e-10	7.1151	3.9904e-9	7.3221	20
	TS15_future spinal cord;neural tube	2.6184e-13	3.5091e-10	7.9228	6.3456e-9	8.0876	18
	TS19_future spinal cord;neural tube	3.6644e-13	4.0925e-10	8.3203	3.1185e-9	9.5634	17
	TS13_embryo;ectoderm	4.5789e-13	4.3833e-10	5.6744	5.0055e-9	5.9829	23
	TS13_embryo;ectoderm;neural ectoderm	7.3579e-13	6.1631e-10	5.7982	9.1665e-9	6.0648	22
	TS20_spinal cord	1.5525e-12	1.1559e-9	5.6102	3.7434e-8	5.5965	22
	TS15_central nervous system	1.7154e-12	1.1495e-9	4.6585	3.6061e-9	5.0323	28

Supplementary Table 23: GREAT enrichments of all p300 midbrain peaks using the *two nearest genes* association rule with a maximum extension of 1 Mb. Shown are the top ten binomial enriched terms at a false discovery rate of 0.05 with a fold enrichment of at least two that are also significant by the hypergeometric test.

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Biological Process	nervous system development	8.9282e-16	1.3508e-12	2.0057	4.2823e-21	3.0289	104
	compartment specification	1.1038e-10	6.2625e-8	53.6168	0.0003	20.9402	4
	embryonic morphogenesis	2.2034e-9	1.0001e-6	2.3233	1.5125e-5	2.5261	38
	neuron differentiation	7.1913e-9	2.3315e-6	2.0481	8.7222e-11	3.1290	52
	central nervous system development	1.1010e-8	3.3317e-6	2.1693	1.3647e-9	3.2051	45
	neural tube development	8.3610e-8	1.6500e-5	3.6850	8.4892e-5	4.4240	15
	negative regulation of cell differentiation	2.1678e-7	3.0750e-5	2.6025	5.0347e-6	3.4242	26
	central nervous system neuron axonogenesis	7.2944e-7	8.0755e-5	6.3268	0.0014	8.9744	6
	neuron fate commitment	7.9073e-7	8.1571e-5	3.7408	9.2964e-5	5.4627	12
	Wnt receptor signaling pathway	2.2642e-6	0.0002	2.9659	0.0173	2.6846	15
GO Cellular Component	axon	0.0013	0.0469	2.2162	0.0441	3.3198	13
Mouse Phenotype	abnormal brain white matter morphology	4.2983e-15	2.4922e-11	3.5574	7.8448e-15	6.1910	34
	abnormal tract	4.5586e-14	1.3215e-10	3.5070	4.1825e-13	5.9555	31
	abnormal corpus callosum morphology	1.2667e-12	1.4689e-9	4.1004	3.5900e-9	6.1936	21
	abnormal brain commissure morphology	1.3814e-12	1.3349e-9	3.5245	1.7899e-11	5.9514	27
	abnormal dorsal telencephalic commissure morphology	1.9284e-12	1.5973e-9	4.0408	7.0441e-9	5.9425	21
	abnormal telencephalon morphology	1.4640e-10	9.4317e-8	2.1081	1.9582e-11	2.8931	63
	abnormal brain ventricle/choroid plexus morphology	8.1425e-10	4.2918e-7	2.8331	9.9391e-9	3.7968	33
	abnormal brain ventricle morphology	9.7071e-10	4.6901e-7	3.1040	4.8994e-8	4.1880	27
	abnormal neural tube morphology/development	1.2455e-9	5.5549e-7	2.1960	1.1836e-6	2.4879	48
	abnormal brain development	2.1990e-9	9.1072e-7	2.0215	5.6241e-9	2.6346	58
PANTHER Pathway	Notch signaling pathway	1.4854e-7	2.2132e-5	6.6365	0.0008	5.7110	9
	Cadherin signaling pathway	1.1926e-5	0.0009	2.9131	5.5533e-5	4.2379	17
	Alzheimer disease-presenilin pathway	0.0002	0.0090	2.7017	0.0088	3.1290	13
	Angiogenesis	0.0002	0.0091	2.2151	0.0251	2.4456	16
MGI Expression: Detected	TS15_central nervous system	1.0221e-18	6.8491e-15	2.6191	4.4268e-20	3.9161	72
	TS15_nervous system	1.5796e-18	5.2926e-15	2.6023	6.7759e-20	3.8462	72
	TS24_sensory organ	1.1608e-17	2.5928e-14	2.3738	4.9774e-11	2.3525	81
	TS21_diencephalon	1.2949e-17	2.1693e-14	2.5063	1.7240e-18	3.5727	72
	TS21_thalamus	3.5516e-17	4.7599e-14	2.6502	3.0790e-18	3.7864	66
	TS15_future brain	4.7966e-17	5.3570e-14	2.7571	1.0055e-19	4.3895	61
	TS22_retina	5.0541e-17	4.8382e-14	2.0938	2.0399e-14	2.4080	102
	TS24_neural retina	3.1346e-16	2.3339e-13	2.6888	3.1730e-10	2.6619	60
	TS21_hindbrain	1.3880e-15	8.4552e-13	2.1289	1.5626e-14	2.5523	93
	TS24_eye	1.6077e-15	8.2868e-13	2.3539	1.4690e-10	2.4099	74

Supplementary Table 24: GREAT enrichments of all p300 midbrain peaks using the *single nearest gene* association rule with a maximum extension of 1 Mb. Shown are the top ten binomial enriched terms at a false discovery rate of 0.05 with a fold enrichment of at least two that are also significant by the hypergeometric test.

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Biological Process	nervous system development	8.1472e-12	3.6980e-8	2.1354	1.6451e-17	3.6263	70
	compartment specification	9.8991e-11	1.4977e-7	91.9146	0.0037	27.9350	3
	neuron fate commitment	2.3139e-9	1.5004e-6	6.0958	5.2404e-6	8.9068	11
	embryonic morphogenesis	1.6608e-8	9.4230e-6	2.7479	1.6886e-5	3.1926	27
	regionalization	2.0593e-8	1.0386e-5	3.4995	5.9249e-5	3.6696	20
	cell fate commitment	2.0981e-8	9.5233e-6	3.4021	1.6037e-6	5.0549	19
	central nervous system development	4.5272e-8	1.8681e-5	2.5354	8.3451e-9	4.0541	32
	negative regulation of neuron differentiation	9.9073e-8	3.7474e-5	6.2662	8.2796e-7	12.4156	10
	neurogenesis	1.2379e-7	4.0134e-5	2.1030	2.0751e-9	3.4306	42
	anterior/posterior pattern formation	1.4783e-7	4.4733e-5	3.9935	0.0004	3.8799	15
Mouse Phenotype	abnormal nervous system development	2.8409e-13	1.6471e-9	2.2987	1.2389e-12	3.1615	64
	abnormal neural tube morphology/development	9.0074e-12	2.6112e-8	2.9090	1.9678e-7	3.3190	36
	abnormal brain white matter morphology	5.0400e-11	9.7406e-8	3.8622	3.8731e-9	6.8016	21
	abnormal neural tube closure	2.3581e-10	2.7344e-7	3.4734	4.0347e-5	3.6419	22
	abnormal tract	2.4894e-10	2.4056e-7	3.8179	6.9999e-8	6.4926	19
	abnormal brain development	1.4844e-9	1.2295e-6	2.4384	1.8183e-7	3.1510	39
	abnormal corpus callosum morphology	1.9512e-9	1.4141e-6	4.5107	1.2394e-5	6.8198	13
	abnormal dorsal telencephalic commissure morphology	2.4743e-9	1.5940e-6	4.4549	1.8942e-5	6.5434	13
	abnormal brain commissure morphology	4.6744e-9	2.7102e-6	3.7568	3.1765e-7	6.6652	17
	abnormal neurogenesis	5.1836e-9	2.7322e-6	2.7803	9.1146e-9	4.5576	29
PANTHER Pathway	Angiogenesis	0.0006	0.0476	2.5886	0.0041	3.5344	13
MSigDB Pathway	Presenilin is required for gamma-secretase activity to activate Notch signaling; presenilin also inhibits beta-catenin in the Wnt/Frizzled pathway.	3.6052e-8	1.6440e-5	17.0943	0.1379	11.4606	4
MGI Expression: Detected	TS22_retina	4.3275e-16	2.8998e-12	2.4928	6.2062e-13	2.8554	68
	TS19_central nervous system	7.9398e-15	2.6602e-11	2.9376	2.1571e-20	5.6874	51
	TS19_nervous system	9.8208e-15	2.1936e-11	2.8658	2.7272e-20	5.4559	52
	TS15_future spinal cord	2.3211e-14	3.8885e-11	3.9851	1.9413e-14	6.3063	32
	TS15_future brain	6.4296e-14	8.6170e-11	3.1026	8.6337e-18	5.6318	44
	TS15_central nervous system	7.8037e-14	8.7154e-11	2.8232	6.9668e-17	4.7405	49
	TS15_nervous system	1.2209e-13	1.1688e-10	2.7952	1.3408e-16	4.6559	49
	TS21_thalamus	2.3139e-13	1.7228e-10	2.9710	7.8886e-18	5.0003	49
	TS22_eye	2.8938e-13	1.7629e-10	2.1732	3.0439e-12	2.6126	74
	TS24_sensory organ	2.9306e-13	1.6365e-10	2.6174	2.5565e-11	2.9446	57
InterPro	High mobility group, HMG1/HMG2	4.6040e-5	0.0482	4.1617	0.0018	7.6014	10

Supplementary Table 25: GREAT enrichments for regions bound by p300 in mouse embryonic stem cells. **(a)** DAVID gene-based enrichments of genes with proximal p300 binding events. **(b)** GREAT *cis*-regulatory element enrichments for all regions bound by p300.

a

No terms were found significant after multiple hypothesis correction in DAVID's gene-based test.

b

GREAT Enrichments of p300 Binding Peaks in Mouse Embryonic Stem Cells

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Molecular Function	chromatin binding	8.9388e-6	0.0053	2.8029	0.0430	2.7963	16
	N-acetylglucosamine-6-sulfatase activity	9.4269e-5	0.0321	11.3550	0.0365	20.0988	3
	arylsulfatase activity	0.0001	0.0438	10.3091	0.0365	20.0988	3
GO Biological Process	stem cell differentiation	5.7696e-10	2.6188e-6	8.5555	0.0014	6.6996	8
	stem cell maintenance	9.4722e-10	2.1497e-6	9.1341	0.0020	7.4048	7
	regulation of transport	4.7020e-6	0.0006	2.6468	0.0030	2.4864	24
	anatomical structure homeostasis	5.9911e-6	0.0008	3.1457	0.0184	2.8712	14
	negative regulation of transcription, DNA-dependent	2.2051e-5	0.0027	2.1161	0.0002	2.5961	31
	negative regulation of RNA metabolic process	2.3123e-5	0.0028	2.1114	0.0002	2.5746	31
	lens morphogenesis in camera-type eye	0.0002	0.0141	4.5447	0.0106	8.3745	5
	regulation of exocytosis	0.0003	0.0197	5.5958	0.0279	5.4815	6
	salivary gland development	0.0004	0.0233	5.4270	0.0018	9.2763	6
	regulation of cellular localization	0.0004	0.0254	2.8053	0.0188	3.0033	13
MGI Expression: Detected	Theiler_stage_4	9.7259e-14	6.5173e-10	2.0894	2.3351e-6	1.9010	96
	TS4_embryo	1.2640e-13	4.2351e-10	2.0903	3.3447e-6	1.8780	94
	TS4_extraembryonic component	2.7494e-13	6.1412e-10	2.0934	3.8689e-6	1.8875	91
	TS4_inner cell mass	6.8093e-13	1.1407e-9	2.1335	3.6068e-6	1.9428	84
	TS4_compacted morula	1.6680e-12	2.2355e-9	2.3158	0.0001	1.8651	67
	Theiler_stage_2	4.9740e-12	5.5551e-9	2.0815	8.3133e-6	1.8850	83
	TS4_zona pellucida	7.4298e-12	6.2234e-9	2.2925	0.0002	1.8402	64
	TS4_second polar body	7.4298e-12	6.2234e-9	2.2925	0.0002	1.8402	64
	TS28_oocyte	3.3623e-11	2.2531e-8	2.0108	2.7026e-5	1.8418	80
	TS5_inner cell mass	3.8743e-11	2.3601e-8	2.1867	3.1103e-5	1.9594	66
Predicted Promoter Motifs	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif SNNNCCNCAGGCN which matches annotation for GTF3A: general transcription factor IIIA	0.0001	0.0294	2.2094	0.0088	2.0662	22
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif NNGGGNCGCAGCTGCGNCCCNN which matches annotation for NHLH1: nescient helix loop helix 1	0.0004	0.0440	2.2420	0.0006	2.8712	21

Supplementary Table 26: "Gene-based GREAT" enrichments of all genes that possess a p300 binding peak in mouse embryonic stem cells within 2 kb of its transcription start site. Shown are the top ten hypergeometric enriched terms at a false discovery rate of 0.05.

Ontology	Term	Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Observed Gene Hits
GO Biological Process	stem cell differentiation	9.9234e-6	0.0450	3
Mouse Phenotype	gastrointestinal ulcer	4.0296e-6	0.0234	3
	decreased embryo size	8.7452e-6	0.0254	7
	abnormal embryo size	9.5616e-6	0.0185	7
	digestive/alimentary phenotype	1.9548e-5	0.0283	9
	abnormal large intestine morphology	3.3942e-5	0.0394	4
MGI Expression: Detected	TS21_ovary;primordial germ cells	2.7721e-6	0.0186	3
	TS21_testis;primordial germ cells	7.5685e-6	0.0254	3
MSigDB Perturbation	Downregulated in MES cells from elongin-A knockout mice	8.1473e-6	0.0020	5
Transcription Factor Targets	Targets of Foxp3, identified by ChIP-chip in Foxp3-CD4+ T-cell hybridomas that were transduced with FLAG-tagged Foxp3 and stimulated by phorbol myristate acetate/ionomycin.	0.0060	0.0360	6

Supplementary Table 27: GREAT enrichments of all p300 binding peaks in mouse embryonic stem cells using the *basal plus extension* association rule with a basal regulatory region extending 5 kb upstream and 1 kb downstream of the transcription start site and a maximum extension of 50 kb. Shown are the top ten binomial enriched terms at a false discovery rate of 0.05 with a fold enrichment of at least two that are also significant by the hypergeometric test.

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Molecular Function	protein binding	3.0335e-20	3.6098e-17	2.1427	0.0026	1.3996	149
	nucleic acid binding	9.2792e-15	7.3615e-12	2.5578	0.0280	1.5509	77
	DNA binding	1.4548e-14	8.6560e-12	2.8909	0.0079	1.7534	61
GO Biological Process	regulation of biological process	1.7971e-19	4.0786e-16	2.0383	0.0356	1.2585	165
	regulation of cellular process	3.7840e-19	4.2939e-16	2.0565	0.0327	1.2723	160
	developmental process	2.1366e-18	1.9396e-15	2.5470	0.0005	1.6736	96
	multicellular organismal development	3.1273e-15	2.3658e-12	2.5521	0.0012	1.7283	80
	regulation of cellular metabolic process	1.1133e-14	7.2192e-12	2.5121	0.0078	1.6301	80
	regulation of metabolic process	5.6911e-14	3.2290e-11	2.4208	0.0172	1.5681	81
	regulation of cellular biosynthetic process	4.3273e-13	2.1824e-10	2.4841	0.0160	1.6008	72
	regulation of biosynthetic process	4.7411e-13	2.1520e-10	2.4795	0.0154	1.5978	72
	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	1.5776e-12	5.5083e-10	2.4755	0.0292	1.5756	68
	regulation of macromolecule metabolic process	6.0619e-12	1.7197e-9	2.3181	0.0383	1.4950	74
MGI Expression: Detected	TS17_organ system	6.4094e-20	4.2949e-16	2.9919	0.0002	1.8456	77
	Theiler_stage_17	1.1317e-19	3.7917e-16	2.7292	0.0002	1.7034	89
	TS17_embryo	1.3942e-19	3.1143e-16	2.7396	0.0002	1.7155	88
	TS28_visceral organ	2.2937e-17	2.5617e-14	2.2499	0.0024	1.4217	114
	TS21_embryo	1.4383e-16	1.3769e-13	2.5277	0.0008	1.6138	85
	TS4_inner cell mass	1.5124e-16	1.2868e-13	3.9928	0.0002	2.4501	45
	TS21_organ system	2.5475e-16	1.8967e-13	2.7098	0.0005	1.7239	74
	Theiler_stage_21	2.7409e-16	1.8367e-13	2.5013	0.0010	1.5908	85
	TS15_embryo	4.4056e-16	2.6838e-13	3.1488	0.0004	1.9700	57
	TS17_nervous system	1.2210e-15	6.8182e-13	2.8960	0.0006	1.8003	64
MSigDB Perturbation	Downregulated in MES cells from elongin-A knockout mice	2.4658e-15	6.1151e-13	9.3446	2.8464e-7	5.6538	19
	Enriched in mouse embryonic stem cells, compared to differentiated brain and bone marrow cells	8.4647e-10	1.0496e-7	2.7101	0.0181	1.7640	45
	Trans-regulated hematopoietic stem cell (HSC) transcripts detected in bone marrow tissue (high likelihood ratio statistic (LRS) value and genome-wide linkage P < 0.005)	1.5634e-8	1.2924e-6	2.9450	0.0224	1.8754	35
	Down-regulated in brown preadipocytes from Irs1-knockout mice, which display severe defects in adipocyte differentiation, versus wild-type controls	5.9293e-5	0.0015	5.3965	0.0222	4.0944	9
Predicted Promoter Motifs	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif CTTTGT which matches annotation for LEF1: lymphoid enhancer-binding factor 1	1.7633e-12	1.0844e-9	2.6744	8.0264e-5	1.9714	63
	Genes with promoter regions [-2kb,2kb] around transcription start site containing motif CANCCNNWGGGTGDGG. Motif does not match any known transcription factor	7.0762e-6	0.0004	4.3290	0.0375	3.1098	14
	Genes with promoter regions [-2kb,2kb] around transcription start site containing motif ACTWSNACTNY. Motif does not match any known transcription factor	2.0446e-5	0.0008	8.7267	0.0403	5.5199	7
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif NNNTTCYN which matches annotation for STAT5A: signal transducer and activator of transcription 5A	2.6474e-5	0.0009	4.4089	0.0423	3.5934	12

Supplementary Table 28: GREAT enrichments of all p300 binding peaks in mouse embryonic stem cells using the *two nearest genes* association rule with a maximum extension of 1 Mb. Shown are the top ten binomial enriched terms at a false discovery rate of 0.05 with a fold enrichment of at least two that are also significant by the hypergeometric test.

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Molecular Function	N-acetylglucosamine-6-sulfatase activity	9.6114e-5	0.0286	11.3076	0.0421	19.1741	3
	arylsulfatase activity	0.0001	0.0397	10.2700	0.0421	19.1741	3
GO Biological Process	stem cell differentiation	6.9247e-10	3.9289e-7	8.4398	0.0019	6.3914	8
	stem cell maintenance	1.1038e-9	5.5669e-7	9.0239	0.0028	7.0642	7
	regulation of transport	2.2954e-6	0.0003	2.6928	0.0026	2.4709	25
	anatomical structure homeostasis	7.3617e-6	0.0009	3.1024	0.0275	2.7392	14
	negative regulation of transcription, DNA-dependent	2.9547e-5	0.0031	2.0866	0.0005	2.4767	31
	negative regulation of RNA metabolic process	3.1006e-5	0.0032	2.0818	0.0005	2.4562	31
	regulation of cellular localization	0.0002	0.0109	2.9474	0.0109	3.0855	14
	lens morphogenesis in camera-type eye	0.0002	0.0143	4.5125	0.0136	7.9892	5
	regulation of secretion	0.0003	0.0160	2.9337	0.0453	2.6803	13
	regulation of exocytosis	0.0003	0.0193	5.5041	0.0340	5.2293	6
MGI Expression: Detected	Theiler_stage_4	1.2304e-13	8.2447e-10	2.0725	5.2306e-6	1.8324	97
	TS4_embryo	1.5836e-13	5.3057e-10	2.0736	9.9722e-6	1.8107	95
	TS4_extraembryonic component	3.3193e-13	7.4143e-10	2.0775	1.0436e-5	1.8205	92
	TS4_compacted morula	1.6333e-12	2.7362e-9	2.3015	0.0002	1.8059	68
	TS4_inner cell mass	1.8117e-12	2.4280e-9	2.0978	1.3487e-5	1.8534	84
	Theiler_stage_2	5.3625e-12	5.9891e-9	2.0685	2.0237e-5	1.8199	84
	TS4_zona pellucida	7.0420e-12	5.8986e-9	2.2797	0.0004	1.7830	65
	TS4_second polar body	7.0420e-12	5.8986e-9	2.2797	0.0004	1.7830	65
	TS3_zona pellucida	7.2609e-11	4.8655e-8	2.0516	0.0007	1.6743	73
	TS5_inner cell mass	8.5820e-11	4.7923e-8	2.1506	0.0001	1.8693	66
Predicted Promoter Motifs	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif NNGGGNCGCAGCTGCGNCCCNN which matches annotation for NHLH1: nescient helix loop helix 1	0.0002	0.0278	2.3044	0.0005	2.8696	22
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif SNNNCCNCAGGCN which matches annotation for GTF3A: general transcription factor IIIA	0.0002	0.0229	2.1730	0.0136	1.9712	22
	Genes with promoter regions [-2kb,2kb] around transcription start site containing motif NRCCACGTGASN. Motif does not match any known transcription factor	0.0004	0.0322	2.3173	0.0431	1.8312	17
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif NANCACGTGNNW which matches annotation for MYC: v-myc myelocytomatosis viral oncogene homolog (avian) MAX: MYC associated factor X	0.0004	0.0301	2.0545	0.0035	2.2339	24
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif NNANCACGTGNTNN which matches annotation for MAX: MYC associated factor X	0.0007	0.0437	2.1740	0.0325	1.8399	19
	Genes with promoter regions [-2kb,2kb] around transcription start site containing motif CTCTAAAAATAACYCY. Motif does not match any known transcription factor	0.0008	0.0442	2.1134	0.0013	2.9499	18
	Genes with promoter regions [-2kb,2kb] around transcription start site containing motif NNWWWWWNGMCACGTCATYNYWNNN. Motif does not match any known transcription factor	0.0009	0.0453	2.8381	0.0216	2.8290	9
	Genes with promoter regions [-2kb,2kb] around transcription start site containing motif NNNNRRCCAATSR. Motif does not match any known transcription factor	0.0010	0.0452	2.0707	0.0184	1.9565	20

Supplementary Table 29: GREAT enrichments of all p300 binding peaks in mouse embryonic stem cells using the *single nearest gene* association rule with a maximum extension of 1 Mb. Shown are the top ten binomial enriched terms at a false discovery rate of 0.05 with a fold enrichment of at least two that are also significant by the hypergeometric test.

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Molecular Function	chromatin binding	7.9333e-7	0.0005	3.9031	0.0051	4.0972	13
GO Biological Process	stem cell differentiation	7.8285e-14	3.5533e-10	16.2918	2.6561e-5	12.0814	8
	stem cell maintenance	2.3293e-13	5.2863e-10	17.3315	6.0582e-5	13.3532	7
	anatomical structure homeostasis	6.2044e-8	1.4822e-5	4.8903	0.0050	4.0682	11
	negative regulation of transcription, DNA-dependent	7.3758e-8	1.6739e-5	3.0515	3.1432e-6	3.7754	25
	negative regulation of RNA metabolic process	7.7077e-8	1.6660e-5	3.0453	3.5304e-6	3.7442	25
	regulation of transcription from RNA polymerase II promoter	5.8763e-7	9.5258e-5	2.1364	1.4552e-7	2.9501	42
	negative regulation of cellular biosynthetic process	1.2693e-6	0.0002	2.5687	5.7895e-5	3.0018	27
	negative regulation of biosynthetic process	1.3101e-6	0.0002	2.5649	5.8430e-5	2.9745	27
	negative regulation of cellular metabolic process	1.4930e-6	0.0002	2.3929	5.7751e-5	2.7539	31
	negative regulation of macromolecule biosynthetic process	1.5441e-6	0.0002	2.5906	0.0001	2.9541	26
GO Cellular Component	nucleoplasm part	2.0530e-5	0.0016	2.3924	0.0134	2.3818	23
	nucleoplasm	6.7684e-5	0.0046	2.2348	0.0364	2.1596	23
MGI Expression: Detected	TS4_inner cell mass	2.0937e-17	1.4030e-13	2.9581	9.5169e-9	2.5859	62
	Theiler_stage_4	2.2247e-17	7.4539e-14	2.8057	3.1541e-8	2.3925	67
	TS4_embryo	3.9337e-17	8.7865e-14	2.8008	3.6959e-8	2.3779	66
	TS4_extraembryonic component	7.3534e-17	1.2319e-13	2.8164	5.0816e-8	2.3938	64
	Theiler_stage_5	7.9283e-17	1.0626e-13	2.7444	4.1493e-8	2.3873	66
	TS5_inner cell mass	5.6549e-16	6.3156e-13	3.1322	4.5785e-8	2.7304	51
	TS5_embryo	1.0499e-15	1.0050e-12	3.0575	5.6397e-8	2.6696	52
	TS3_zona pellucida	5.2571e-15	4.4035e-12	2.8663	2.5469e-6	2.2978	53
	Theiler_stage_2	1.0294e-14	7.6647e-12	2.7486	1.5723e-6	2.2934	56
	TS3_second polar body	1.1695e-14	7.8371e-12	2.8190	3.1502e-6	2.2814	53
MSigDB Perturbation	Downregulated in MES cells from elongin-A knockout mice	1.6281e-13	4.0378e-11	5.6821	7.1557e-7	4.7870	21
	Down-regulated by stable RNAi knock-down of PRMT5 in NIH 3T3 cells	2.4935e-6	0.0002	9.6774	0.0059	7.4988	6
	Genes upregulated in Egr2Lo/Lo mice (who bear mutations in the transcription factor Egr2 and in which peripheral nerve myelination is disrupted) whose expression is significantly altered after sciatic nerve injury.	0.0001	0.0045	3.2558	0.0066	3.9474	11
	Up-regulated by PDGF in mouse embryonic stem cells, via microarray-coupled gene-trap mutagenesis	0.0009	0.0235	5.5731	0.0319	8.0543	4
	Genes up-regulated in anergic mouse T helper cells (A.E7), versus non-anergic stimulated controls	0.0010	0.0245	3.3630	0.0056	4.5305	10
InterPro	Homeodomain-related	5.8593e-6	0.0368	2.5135	0.0029	3.0596	26

Supplementary Table 30: Enrichments for regions bound by Stat3 in mouse embryonic stem cells. **(a)** DAVID gene-based enrichments of genes with proximal Stat3 binding events. **(b)** GREAT *cis*-regulatory element enrichments for all regions bound by Stat3.

a

DAVID Gene-based Enrichments of Stat3 Binding Peaks in Mouse Embryonic Stem Cells

	Annotation Cluster 1	Enrichment Score: 11.94			Count	P_Value	Benjamini
	GOTERM_CC_ALL	intracellular	RT		320	1.6E-20	1.2E-17
	GOTERM_CC_ALL	intracellular part	RT		307	6.2E-20	2.4E-17
	GOTERM_CC_ALL	intracellular membrane-bound organelle	RT		245	2.3E-16	5.8E-14
	GOTERM_CC_ALL	membrane-bound organelle	RT		245	2.5E-16	4.4E-14
	GOTERM_CC_ALL	intracellular organelle	RT		263	4.4E-15	7.0E-13
	GOTERM_CC_ALL	organelle	RT		263	5.1E-15	6.7E-13
	GOTERM_CC_ALL	nucleus	RT		175	1.0E-14	1.2E-12
	SP_PIR_KEYWORDS	nucleus	RT		137	5.5E-9	2.4E-6
	GOTERM_CC_ALL	cytoplasm	RT		185	2.9E-6	2.9E-4
	GOTERM_CC_ALL	cell part	RT		379	4.7E-3	1.7E-1
	Annotation Cluster 2	Enrichment Score: 5.72			Count	P_Value	Benjamini
	GOTERM_CC_ALL	nucleus	RT		175	1.0E-14	1.2E-12
	GOTERM_BP_ALL	cellular metabolic process	RT		245	1.6E-11	8.2E-8
	GOTERM_BP_ALL	primary metabolic process	RT		242	1.3E-10	3.4E-7
	GOTERM_BP_ALL	metabolic process	RT		259	1.0E-9	1.8E-6
	SP_PIR_KEYWORDS	nucleus	RT		137	5.5E-9	2.4E-6
	GOTERM_BP_ALL	biopolymer metabolic process	RT		170	2.8E-8	3.7E-5
	GOTERM_BP_ALL	regulation of cellular process	RT		142	1.1E-7	9.2E-5
	GOTERM_BP_ALL	macromolecule metabolic process	RT		208	1.5E-7	1.1E-4
	GOTERM_BP_ALL	biological regulation	RT		163	1.8E-7	1.2E-4
	GOTERM_BP_ALL	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	RT		131	3.7E-7	2.1E-4
	Annotation Cluster 3	Enrichment Score: 3.28			Count	P_Value	Benjamini
	GOTERM_BP_ALL	organelle organization and biogenesis	RT		53	1.6E-4	3.5E-2
	GOTERM_BP_ALL	establishment and/or maintenance of chromatin architecture	RT		19	1.9E-4	3.9E-2
	GOTERM_BP_ALL	DNA packaging	RT		19	2.6E-4	4.5E-2
	Annotation Cluster 4	Enrichment Score: 3.17			Count	P_Value	Benjamini
	GOTERM_BP_ALL	post-translational protein modification	RT		63	1.3E-5	3.7E-3
	GOTERM_BP_ALL	cellular protein metabolic process	RT		108	2.2E-4	4.2E-2
	GOTERM_BP_ALL	protein modification process	RT		65	2.3E-4	4.3E-2
	GOTERM_BP_ALL	cellular macromolecule metabolic process	RT		109	2.3E-4	4.2E-2
	GOTERM_BP_ALL	biopolymer modification	RT		67	2.4E-4	4.2E-2
	Annotation Cluster 5	Enrichment Score: 2.91			Count	P_Value	Benjamini
	GOTERM_BP_ALL	positive regulation of cellular process	RT		43	6.6E-5	1.6E-2
	GOTERM_BP_ALL	positive regulation of biological process	RT		47	8.1E-5	1.9E-2

b

GREAT Enrichments of Stat3 Binding Peaks in Mouse Embryonic Stem Cells

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Biological Process	stem cell maintenance	8.7375e-8	9.4428e-6	3.4913	0.0001	3.7396	13
	stem cell differentiation	1.0488e-6	7.8038e-5	3.0521	0.0035	2.9605	13
	response to retinoic acid	9.3973e-6	0.0005	3.1690	0.0151	3.0364	10
	response to nutrient	1.0882e-5	0.0005	3.1561	0.0395	2.7328	10
	trophectodermal cell differentiation	2.7987e-5	0.0012	3.1670	0.0088	3.2150	10
	ER-nuclear signaling pathway	0.0008	0.0182	3.0507	0.0395	2.7328	10
Mouse Phenotype	embryonic lethality before somite formation	1.8975e-12	5.2389e-10	2.0213	2.3511e-6	1.8142	79
	abnormal placenta development	2.1831e-11	5.7535e-9	2.1411	0.0004	1.8218	50
	abnormal placenta labyrinth morphology	2.7099e-8	3.2734e-6	2.0236	0.0020	1.8375	39
	abnormal trophoblast layer morphology	8.1698e-8	9.1094e-6	2.2123	0.0037	1.9010	32
	abnormal energy balance	2.4084e-7	2.3667e-5	2.1543	0.0103	1.7847	32
	abnormal embryonic hematopoiesis	2.4978e-7	2.4137e-5	2.6279	0.0245	2.1329	16
	abnormal wound healing	7.8197e-7	6.7670e-5	2.5022	0.0248	1.9129	21
	abnormal pro-B cell morphology	8.2604e-6	0.0005	2.0095	0.0117	1.8947	26
	echinocytosis	3.1297e-5	0.0015	6.1677	0.0192	4.5546	5
	abnormal neuronal precursor proliferation	6.2130e-5	0.0024	2.3991	0.0445	2.5889	9
PANTHER pathway	Interferon-gamma signaling pathway	9.6342e-10	1.4355e-7	4.3884	0.0498	2.6234	12
MSigDB Pathway	Genes preferentially expressed in breast cancers, especially those involved in estrogen-receptor-dependent signal transduction.	4.0498e-9	9.2395e-7	2.2355	2.7260e-5	2.3013	40
MGI Expression: Detected	TS5_trophectoderm	1.7029e-11	3.5661e-9	2.5054	0.0002	1.9737	39
	TS5_extraembryonic component	3.9589e-10	6.6321e-8	2.2903	0.0004	1.8831	41
	TS12_future midbrain	1.0814e-6	0.0001	2.1152	2.3384e-5	2.6770	24
	TS15_primordial germ cells	3.4369e-5	0.0022	4.3170	0.0211	3.6437	6
	TS25_ovary	5.4173e-5	0.0032	4.1136	0.0018	3.4159	10
	TS25_reproductive system;female	6.1079e-5	0.0035	4.0611	0.0034	3.2150	10
	TS8_epiblast	0.0001	0.0071	2.5415	2.2095e-5	3.5645	15
	TS21_ovary;primordial germ cells	0.0001	0.0072	2.9928	0.0096	3.0743	9
	TS21_testis;primordial germ cells	0.0002	0.0089	2.8111	0.0325	2.4843	10
	TS23_sciatic nerve	0.0004	0.0166	3.1263	0.0128	4.5546	5
MSigDB Perturbation	Downregulated in MES cells from elongin-A knockout mice	1.1427e-26	2.8339e-24	3.0013	9.2930e-9	2.2000	64
	Up-regulated by insulin in murine adipocytes, but response is blunted following induction of insulin-resistance with TNFalpha treatment	8.9619e-17	7.4085e-15	11.9927	0.0498	2.9812	6
	Genes that increased after LIF treatment (10 ng/ml, overnight) in AtT20 cells	1.2669e-14	6.2840e-13	4.3311	0.0002	2.6692	21
	Downregulated by expression of constitutively active JNK in 3T3 cells	3.3296e-10	1.3762e-8	3.5661	0.0216	2.2920	13
	Genes identified as time indicators in mouse liver.	6.7468e-10	2.3903e-8	2.0644	0.0018	1.6869	50
	Up-regulated by PDGF in mouse embryonic stem cells, via microarray-coupled gene-trap mutagenesis	1.7617e-8	5.4613e-7	3.3094	4.2516e-5	3.9473	13
	Downregulated following iNOS induction in hepatocytes (Tables 3-17)	5.9465e-7	1.3407e-5	2.3412	0.0058	1.9129	28
	Upregulated by nickel(II) in sensitive AJ mouse lung tissue	9.6776e-5	0.0014	2.2867	0.0064	2.3228	17
	Up-regulated by insulin in murine adipocytes, and continue to respond following induction of insulin-resistance with TNFalpha treatment	0.0001	0.0013	2.6576	0.0385	2.4695	9
	Transcription modulators presented during myeloid differentiation	0.0013	0.0110	2.0860	0.0388	2.0698	13

Supplementary Table 31: "Gene-based GREAT" enrichments of all genes that possess a Stat3 binding peak in mouse embryonic stem cells within 2 kb of its transcription start site. Shown are the top ten hypergeometric enriched terms at a false discovery rate of 0.05.

Ontology	Term	Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Observed Gene Hits
GO Molecular Function	binding	3.6186e-7	0.0009	311
	nucleic acid binding	1.1622e-6	0.0014	97
GO Biological Process	cellular metabolic process	1.5460e-11	7.0172e-8	215
	metabolic process	1.0073e-9	2.2861e-6	224
	primary metabolic process	2.9303e-9	4.4336e-6	202
	cellular macromolecule metabolic process	7.4631e-8	8.4688e-5	168
	macromolecule metabolic process	2.0175e-7	0.0002	168
	cellular biopolymer metabolic process	2.9796e-7	0.0002	163
	biopolymer metabolic process	5.5103e-7	0.0004	163
	positive regulation of cellular process	2.5304e-6	0.0014	51
	DNA damage response, signal transduction by p53 class mediator resulting in induction of apoptosis	4.5115e-6	0.0023	5
	cellular catabolic process	5.4379e-6	0.0025	43
GO Cellular Component	intracellular membrane-bounded organelle	1.1571e-9	7.9379e-7	227
	membrane-bounded organelle	1.2939e-9	4.4379e-7	227
	nucleus	2.3517e-9	5.3776e-7	158
	intracellular organelle	4.5871e-9	7.8669e-7	247
	organelle	4.9418e-9	6.7802e-7	247
	intracellular part	1.3180e-8	1.5069e-6	261
	intracellular	1.7920e-8	1.7569e-6	267
	nuclear part	7.0866e-7	6.7628e-5	45
	nuclear lumen	1.9187e-5	0.0015	29
	intracellular organelle lumen	3.5717e-5	0.0025	31
Mouse Phenotype	prenatal lethality	1.1941e-7	0.0007	71
	lethality-prenatal/perinatal	2.1215e-6	0.0062	84
	embryonic lethality	2.7530e-6	0.0053	53
	abnormal cell physiology	3.6327e-6	0.0053	49
	abnormal placenta labyrinth morphology	9.1559e-6	0.0106	13
	abnormal dendritic cell number	1.1978e-5	0.0116	6
	cellular phenotype	1.2178e-5	0.0101	51
	abnormal extraembryonic tissue morphology	1.6228e-5	0.0118	31
	abnormal placenta morphology	2.2914e-5	0.0148	21
	abnormal homeostasis	3.7834e-5	0.0219	75

Ontology	Term	Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Observed Gene Hits
MGI Expression: Detected	TS28_midbrain	1.7371e-6	0.0116	49
	TS28_corpus striatum	1.8633e-6	0.0062	48
	TS28_hand	2.0215e-6	0.0045	27
	TS28_forelimb	2.3468e-6	0.0039	27
	TS5_inner cell mass	2.4488e-6	0.0033	39
	TS5_embryo	2.7203e-6	0.0030	40
	Theiler_stage_5	5.7046e-6	0.0055	50
	TS28_hippocampus	5.8132e-6	0.0049	58
	TS4_inner cell mass	6.8008e-6	0.0051	45
	TS28_uterine cervix	7.2730e-6	0.0049	29
MSigDB Perturbation	Enriched in mouse embryonic stem cells, compared to differentiated brain and bone marrow cells	1.3825e-7	3.4286e-5	62
	Downregulated in MES cells from elongin-A knockout mice	1.5875e-5	0.0020	15
	Enriched in mouse hematopoietic stem cells, compared to differentiated brain and bone marrow cells	2.8197e-5	0.0023	57
	Genes that increased after LIF treatment (10 ng/ml, overnight) in AtT20 cells	0.0001	0.0064	7
	Up-regulated by insulin in murine adipocytes, but response is blunted following induction of insulin-resistance with TNFalpha treatment	0.0001	0.0062	4
	Up-regulated in the cerebral cortex of aged (22 months) BALB/c mice, compared to young (2 months) controls	0.0001	0.0060	6
	Genes upregulated in Egr2Lo/Lo mice (who bear mutations in the transcription factor Egr2 and in which peripheral nerve myelination is disrupted) whose expression is significantly altered after sciatic nerve injury.	0.0003	0.0098	10
	Down-regulated at 48-96 hours during differentiation of 3T3-L1 fibroblasts into adipocytes with IDX (insulin, dexamethasone and isobutylxanthine), vs. fibroblasts treated with IDX + TSA to prevent differentiation (cluster 1)	0.0004	0.0139	6
	Enriched in mouse neural stem cells, compared to differentiated brain and bone marrow cells	0.0005	0.0137	66
	Genes identified as time indicators in mouse liver.	0.0010	0.0253	12

Supplementary Table 32: GREAT enrichments of all Stat3 binding peaks in mouse embryonic stem cells using the *basal plus extension* association rule with a basal regulatory region extending 5 kb upstream and 1 kb downstream of the transcription start site and a maximum extension of 50 kb. Shown are the top ten binomial enriched terms at a false discovery rate of 0.05 with a fold enrichment of at least two that are also significant by the hypergeometric test.

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Biological Process	binding	2.5949e-173	6.1758e-170	2.0476	6.1338e-13	1.1504	1246
	protein binding	9.7876e-95	1.1647e-91	2.1671	1.7561e-6	1.2023	650
	nucleic acid binding	7.2358e-83	5.7404e-80	2.7991	1.8306e-10	1.4239	359
	catalytic activity	2.2357e-71	1.3303e-68	2.0824	0.0018	1.1549	579
	DNA binding	3.0532e-67	1.4533e-64	2.9482	1.8398e-8	1.4716	260
	transcription regulator activity	1.5822e-46	6.2760e-44	3.0174	2.5617e-8	1.6195	174
	nucleotide binding	1.9174e-44	6.5191e-42	2.4296	1.1206e-5	1.3656	265
	transferase activity	2.3052e-40	6.8581e-38	2.4533	0.0003	1.3409	230
	transition metal ion binding	2.5393e-38	5.0363e-36	2.1414	0.0287	1.1988	298
	zinc ion binding	2.5524e-37	4.6728e-35	2.2866	0.0021	1.2736	254
GO Biological Process	cellular process	1.3765e-166	6.2482e-163	2.0431	0.0006	1.0769	1218
	metabolic process	3.5271e-128	8.0048e-125	2.2512	4.8966e-10	1.2103	817
	cellular metabolic process	1.6127e-124	2.4400e-121	2.3078	4.2053e-10	1.2297	755
	primary metabolic process	8.4370e-119	9.5739e-116	2.2960	1.7510e-9	1.2239	730
	macromolecule metabolic process	2.6026e-107	2.3626e-104	2.3762	1.7879e-8	1.2397	615
	cellular macromolecule metabolic process	5.0435e-107	3.8154e-104	2.3855	1.1814e-8	1.2448	609
	biopolymer metabolic process	7.7545e-105	5.0283e-102	2.3763	9.3911e-8	1.2322	597
	cellular biopolymer metabolic process	1.3745e-104	7.7983e-102	2.3819	8.1644e-8	1.2351	593
	regulation of macromolecule metabolic process	5.4955e-84	2.2676e-81	2.7586	1.0188e-10	1.4401	362
	biosynthetic process	1.0399e-82	3.6307e-80	2.6258	1.6191e-9	1.3666	400
GO Cellular Component	intracellular part	9.8828e-179	6.7796e-176	2.2256	2.6149e-17	1.1999	1112
	intracellular	2.6843e-178	9.2071e-176	2.1978	1.7433e-17	1.1963	1142
	intracellular organelle	1.1842e-149	2.0310e-147	2.2552	5.7583e-12	1.1972	931
	organelle	2.2422e-149	3.0763e-147	2.2530	5.3039e-12	1.1964	931
	membrane-bounded organelle	3.3219e-136	3.7980e-134	2.2907	1.8192e-10	1.2025	828
	intracellular membrane-bounded organelle	5.4510e-136	5.3420e-134	2.2905	1.6864e-10	1.2024	827
	nucleus	5.0998e-110	4.3731e-108	2.5308	2.3283e-11	1.2998	560
	cytoplasm	4.7498e-104	3.6204e-102	2.1907	2.4124e-7	1.1875	718
	intracellular organelle part	7.2907e-54	5.0014e-52	2.5001	0.0003	1.2632	301
	organelle part	5.7597e-53	3.5920e-51	2.4767	0.0006	1.2530	301
Mouse Phenotype	mammalian phenotype	1.5076e-126	8.7410e-123	2.2739	3.5926e-14	1.2757	758
	lethality-prenatal/perinatal	2.0111e-82	5.8303e-79	2.8578	1.2157e-14	1.5599	330
	prenatal lethality	3.9576e-78	7.6488e-75	3.1910	1.4756e-14	1.6753	262
	growth/size phenotype	2.4755e-69	3.5883e-66	2.5544	1.6161e-9	1.4175	337
	embryonic lethality	2.6401e-61	3.0615e-58	3.2927	2.8154e-11	1.7051	194
	abnormal embryogenesis/ development	4.4065e-55	4.2581e-52	3.0984	6.0791e-11	1.6830	196
	abnormal postnatal growth/weight/body size	2.2374e-51	1.8532e-48	2.4768	1.8541e-6	1.3848	268
	abnormal blood cell morphology/development	8.4753e-49	6.1425e-46	2.8459	9.2167e-7	1.4869	203
	abnormal hematopoietic system morphology/development	6.9927e-48	4.5049e-45	2.6981	7.9659e-6	1.4154	218
	decreased body size	1.2252e-47	7.1037e-45	2.6249	3.5735e-6	1.4293	220
MGI Expression: Detected	TS28_organ system	3.7281e-133	2.4982e-129	2.3329	4.6085e-15	1.2841	763
	Theiler_stage_28	8.9991e-133	3.0152e-129	2.3189	1.0998e-14	1.2757	769
	TS28_visceral organ	8.9115e-103	1.9905e-99	2.4696	7.2024e-9	1.2794	521
	TS28_central nervous system	3.6460e-98	6.1079e-95	2.4392	1.2372e-13	1.3558	538
	TS28_nervous system	5.4726e-98	7.3344e-95	2.4344	1.3785e-13	1.3539	539
	TS28_brain	2.6555e-93	2.9658e-90	2.4666	4.3629e-13	1.3673	500
	TS28_reproductive system	3.2029e-89	3.0661e-86	2.5440	1.2198e-9	1.3340	443
	Theiler_stage_4	3.7746e-74	3.1617e-71	3.7374	1.4092e-15	1.8633	203
	TS4_embryo	8.4167e-72	6.2667e-69	3.6969	2.6965e-15	1.8522	200
	TS21_embryo	2.0583e-71	1.3793e-68	2.5015	2.9468e-10	1.3945	373
MSigDB Perturbation	Enriched in mouse hematopoietic stem cells, compared to differentiated brain and bone marrow cells	6.1060e-61	1.5143e-58	3.1865	8.9217e-8	1.4965	204
	Enriched in mouse embryonic stem cells, compared to differentiated brain and bone marrow cells	3.7324e-52	4.6281e-50	2.9748	2.4581e-9	1.5669	203
	Downregulated in MES cells from elongin-A knockout mice	4.7725e-38	3.9453e-36	6.6059	2.3489e-10	2.9298	50
	Enriched in mouse neural stem cells, compared to differentiated brain and bone marrow cells	2.7866e-31	1.7277e-29	2.2188	0.0022	1.2656	227
	Trans-regulated hematopoietic stem cell (HSC) transcripts detected in bone marrow tissue (high likelihood ratio statistic (LRS) value and genome-wide linkage P < 0.005)	3.8027e-24	1.8862e-22	2.5423	0.0252	1.2872	122
	Genes that increased after LIF treatment (10 ng/ml, overnight) in AtT20 cells	4.4037e-21	1.8202e-19	10.2546	0.0002	3.4667	16
	Genes identified as time indicators in mouse liver.	1.6406e-17	5.8125e-16	4.2744	0.0001	2.1854	38
	Up-regulated by insulin in murine adipocytes, but response is blunted following induction of insulin-resistance with TNFalpha treatment	1.2277e-16	3.8058e-15	22.2054	0.0388	4.2348	5
	Genes with at least five fold change in expression between Pre-BI and Large Pre-BII cells	2.1701e-16	5.9797e-15	3.7099	0.0034	1.7594	44
	Up-regulated by PDGF in mouse embryonic stem cells, via microarray-coupled gene-trap mutagenesis	1.9551e-15	4.8487e-14	12.0666	2.0394e-5	5.6935	11

Supplementary Table 33: GREAT enrichments of all Stat3 binding peaks in mouse embryonic stem cells using the *two nearest genes* association rule with a maximum extension of 1 Mb. Shown are the top ten binomial enriched terms at a false discovery rate of 0.05 with a fold enrichment of at least two that are also significant by the hypergeometric test.

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Biological Process	stem cell maintenance	7.9378e-9	6.4339e-7	3.7145	4.3980e-5	3.6718	14
	stem cell differentiation	1.2632e-7	7.6450e-6	3.2424	0.0020	2.9069	14
	trophectodermal cell differentiation	7.7369e-7	3.7761e-5	3.6356	0.0168	2.9313	10
	microtubule-based movement	1.1064e-6	5.1774e-5	2.2130	0.0185	1.8064	29
	cytoskeleton-dependent intracellular transport	2.5472e-6	0.0001	2.0851	0.0385	1.6786	32
	response to retinoic acid	3.2378e-6	0.0001	3.3118	0.0065	3.0453	11
	response to nutrient	3.8858e-6	0.0002	3.2719	0.0190	2.7408	11
	N-acetylglucosamine metabolic process	9.3244e-5	0.0023	3.6312	0.0463	2.6227	10
	amino sugar metabolic process	0.0007	0.0127	2.6868	0.0313	2.4916	12
	peptidyl-citrulline biosynthetic process from peptidyl-arginine	0.0024	0.0311	11.6060	0.0116	4.9832	5
Mouse Phenotype	abnormal placenta morphology	1.2745e-18	4.1052e-16	2.0709	2.0526e-5	1.6268	95
	embryonic lethality before somite formation	1.3055e-13	2.6100e-11	2.0710	5.9979e-6	1.7378	83
	abnormal placenta development	9.2992e-13	1.5858e-10	2.2172	0.0005	1.7607	53
	increased resistance to diet-induced obesity	8.0304e-11	1.0582e-8	3.2740	0.0096	2.0302	22
	increased energy expenditure	1.0444e-10	1.3457e-8	3.2451	0.0119	1.9933	22
	improved glucose tolerance	1.9379e-10	2.2931e-8	2.8319	0.0071	1.9221	27
	abnormal placenta labyrinth morphology	1.1467e-9	1.1873e-7	2.1262	0.0055	1.7183	40
	abnormal energy balance	7.0476e-9	6.6987e-7	2.3042	0.0030	1.8306	36
	abnormal trophoblast layer morphology	1.0955e-8	9.9246e-7	2.2937	0.0140	1.7333	32
	decreased circulating free fatty acid level	1.4528e-8	1.2762e-6	3.2622	0.0473	1.9447	16
MSigDB Pathway	Genes preferentially expressed in breast cancers, especially those involved in estrogen-receptor-dependent signal transduction.	1.4531e-9	3.3131e-7	2.2638	3.6361e-5	2.2031	42
MGI Expression: Detected	TS5_inner cell mass	2.2945e-38	5.1252e-35	2.0383	1.2992e-13	1.6488	224
	TS4_compacted morula	2.6378e-37	4.4190e-34	2.0488	1.7481e-10	1.5391	223
	TS4_zona pellucida	3.4259e-37	4.5914e-34	2.0645	3.3937e-10	1.5399	216
	TS4_second polar body	3.4259e-37	4.5914e-34	2.0645	3.3937e-10	1.5399	216
	TS3_4-cell stage	1.1807e-33	5.2748e-31	2.0602	2.4545e-10	1.5745	200
	TS5_trophectoderm	6.2396e-12	9.5027e-10	2.5246	0.0006	1.8456	40
	TS5_extraembryonic component	1.6344e-10	2.1062e-8	2.3077	0.0013	1.7588	42
	TS12_future midbrain	3.0219e-7	2.2753e-5	2.1738	9.3258e-6	2.6441	26
	TS19_primordial germ cells	2.1114e-6	0.0001	7.1837	0.0484	3.5594	5
	TS25_ovary	1.5408e-5	0.0007	4.3455	0.0040	3.1145	10
MSigDB Perturbation	Downregulated in MES cells from elongin-A knockout mice	1.5715e-30	1.9486e-28	3.1559	3.6518e-11	2.2879	73
	Genes that increased after LIF treatment (10 ng/ml, overnight) in AtT20 cells	6.1771e-15	3.0639e-13	4.3311	0.0007	2.4337	21
	Downregulated by expression of constitutively active JNK in 3T3 cells	9.5081e-12	3.9300e-10	3.8068	0.0371	2.0897	13
	Genes identified as time indicators in mouse liver.	3.7588e-11	1.3317e-9	2.1388	0.0067	1.5688	51
	Up-regulated by PDGF in mouse embryonic stem cells, via microarray-coupled gene-trap mutagenesis	2.2140e-8	5.4908e-7	3.2753	0.0001	3.5990	13
	Downregulated following iNOS induction in hepatocytes (Tables 3-17)	2.6681e-8	6.0153e-7	2.5053	0.0022	1.9310	31
	Down-regulated during the TGFbeta-induced epithelial-to-mesenchymal transition (EMT) of Ras-transformed mouse mammary epithelial (EpH4) cells (EMT is representative of late-stage tumor progression and metastasis)	1.6996e-5	0.0002	2.1744	0.0063	2.0929	21
	Upregulated by nickel(II) in sensitive A/J mouse lung tissue	5.4458e-5	0.0007	2.3299	0.0065	2.2424	18
	Down-regulated at 48-96 hours during differentiation of 3T3-L1 fibroblasts into adipocytes with IDX (insulin, dexamethasone and isobutylxanthine), vs. fibroblasts treated with IDX + TSA to prevent differentiation (cluster 1)	0.0003	0.0025	2.4007	0.0465	1.9166	15
	Transcription modulators presented during myeloid differentiation	0.0007	0.0045	2.1452	0.0350	2.0519	14

Supplementary Table 34: GREAT enrichments of all Stat3 binding peaks in mouse embryonic stem cells using the *single nearest gene* association rule with a maximum extension of 1 Mb. Shown are the top ten binomial enriched terms at a false discovery rate of 0.05 with a fold enrichment of at least two that are also significant by the hypergeometric test.

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Biological Process	stem cell maintenance	2.3042e-10	3.3737e-8	5.6054	0.0007	4.5612	10
	stem cell differentiation	2.4462e-9	3.3646e-7	4.9178	0.0069	3.6110	10
	response to retinoic acid	2.6410e-9	3.5257e-7	5.7320	0.0028	4.3332	9
	response to nutrient	3.1752e-9	4.1178e-7	5.6659	0.0071	3.8999	9
	trophectodermal cell differentiation	1.1398e-7	1.2318e-5	5.0212	0.0002	5.0978	10
	blastocyst formation	8.0351e-7	5.7891e-5	4.3506	0.0019	4.1268	10
	T cell activation during immune response	1.5567e-6	0.0001	10.3712	0.0446	4.8146	5
	lymphocyte activation during immune response	2.1885e-6	0.0001	8.2767	0.0327	4.3332	6
	DNA damage response, signal transduction by p53 class mediator	9.3056e-6	0.0005	6.8978	0.0493	3.9998	6
	placenta development	0.0001	0.0034	2.7570	0.0067	2.7659	15
Mouse Phenotype	abnormal placenta morphology	2.9517e-15	1.5558e-12	2.3511	3.6222e-7	2.0549	69
	embryonic lethality before somite formation	6.2400e-14	2.4120e-11	2.5172	2.5507e-8	2.2940	63
	abnormal placenta development	2.1501e-12	6.2332e-10	2.7221	6.5187e-6	2.3688	41
	abnormal myeloid leukocyte morphology	2.9815e-12	8.2318e-10	2.0653	0.0004	1.6396	77
	abnormal phagocyte morphology	4.4649e-12	1.0787e-9	2.1653	0.0004	1.7074	66
	abnormal placenta labyrinth morphology	7.9322e-11	1.5859e-8	2.7417	3.4253e-5	2.4654	33
	abnormal embryonic erythropoiesis	3.1651e-9	5.2432e-7	4.8470	0.0090	3.5453	9
	abnormal erythrocyte morphology	1.1657e-8	1.8266e-6	2.2712	0.0019	1.8383	42
	abnormal macrophage morphology	1.3275e-8	2.0254e-6	2.2450	0.0005	1.9259	44
	abnormal mononuclear phagocyte morphology	2.8303e-8	4.1026e-6	2.0441	0.0050	1.6602	50
PANTHER Pathway	Interferon-gamma signaling pathway	7.0009e-9	1.0431e-6	5.3891	0.0033	3.8132	11
MSigDB Pathway	Genes preferentially expressed in breast cancers, especially those involved in estrogen-receptor-dependent signal transduction.	4.5686e-7	0.0001	2.4552	0.0008	2.5543	28
MGI Expression: Detected	Theiler_stage_4	1.8078e-36	1.2114e-32	2.1586	2.2710e-19	1.8955	222
	TS4_embryo	1.9877e-35	6.6598e-32	2.1460	1.5290e-18	1.8780	218
	TS5_inner cell mass	2.0051e-34	4.4787e-31	2.3833	4.5551e-18	2.0866	163
	TS4_zona pellucida	1.0520e-33	1.7623e-30	2.4327	3.8498e-13	1.8845	152
	TS4_second polar body	1.0520e-33	1.7623e-30	2.4327	3.8498e-13	1.8845	152
	TS4_compacted morula	1.0784e-33	1.2044e-30	2.4096	3.0386e-13	1.8725	156
	TS4_extraembryonic component	2.5590e-33	2.4497e-30	2.1254	5.0500e-17	1.8513	207
	TS4_inner cell mass	2.5941e-33	2.1729e-30	2.1855	6.6822e-18	1.9347	194
	TS5_embryo	4.1652e-33	3.1012e-30	2.3199	1.8090e-17	2.0377	166
	TS3_4-cell stage	1.1865e-32	7.9510e-30	2.4810	2.4053e-14	1.9852	145
MSigDB Perturbation	Downregulated in MES cells from elongin-A knockout mice	1.0742e-22	1.3320e-20	3.5084	2.4974e-10	2.8343	52
	Up-regulated by insulin in murine adipocytes, but response is blunted following induction of insulin-resistance with TNFalpha treatment	4.8992e-16	3.0375e-14	17.8189	0.0094	4.7271	6
	Genes that increased after LIF treatment (10 ng/ml, overnight) in AtT20 cells	1.0313e-13	5.1153e-12	5.2830	8.2833e-5	3.4262	17
	Up-regulated by PDGF in mouse embryonic stem cells, via microarray-coupled gene-trap mutagenesis	1.7918e-9	7.4060e-8	4.5881	2.5481e-6	5.7776	12
	Genes identified as time indicators in mouse liver.	2.1430e-9	7.5925e-8	2.4471	0.0002	2.0863	39
	Down-regulated during the TGFbeta-induced epithelial-to-mesenchymal transition (EMT) of Ras-transformed mouse mammary epithelial (EpH4) cells (EMT is representative of late-stage tumor progression and metastasis)	2.2124e-7	5.4866e-6	3.1516	0.0006	2.9466	17
	Downregulated following iNOS induction in hepatocytes (Tables 3-17)	6.8287e-7	1.5396e-5	2.9065	0.0015	2.3832	22
	Genes with at least five fold change in expression between Pre-BI and Large Pre-BII cells	2.2090e-6	4.2141e-5	2.0613	0.0151	1.5994	43
	Genes upregulated in Egr2Lo/Lo mice (who bear mutations in the transcription factor Egr2 and in which peripheral nerve myelination is disrupted) whose expression is significantly altered after sciatic nerve injury.	6.3201e-6	0.0001	2.1060	8.9015e-5	2.4884	29
	Fifty genes most strongly up-regulated in liver tissue from mice deficient in the lamin-protease Zmpste24/Face1, versus wild-type controls	3.1745e-5	0.0005	4.0477	0.0395	2.6896	9

Supplementary Table 35: Enrichment for regions bound by NRSF in human Jurkat cells. **(a)** The top ten proximal binding gene-based enrichments (reproduced from ref. 19). **(b)** GREAT *cis*-regulatory element enrichments for all regions bound by NRSF.

a **Gene-based GO Enrichments of NRSF Promoter Binding Peaks**

Term	p-value
membrane	1.33×10^{-43}
ion transport	9.35×10^{-37}
calcium ion binding	4.27×10^{-31}
synaptic transmission	4.45×10^{-29}
integral to membrane	2.67×10^{-28}
ion channel activity	5.82×10^{-21}
nervous system development	2.19×10^{-20}
potassium ion binding	7.49×10^{-20}
potassium ion transport	1.87×10^{-19}
protein binding	1.14×10^{-18}

b **GREAT Enrichments of NRSF Binding Peaks in Human Jurkat Cells**

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Molecular Function	gated channel activity	1.5998e-15	1.4931e-12	2.0570	8.0373e-21	2.8621	102
	cation channel activity	7.9288e-12	2.0182e-9	2.0126	4.4314e-17	2.8284	84
	voltage-gated ion channel activity	1.5717e-11	3.6673e-9	2.1760	6.3725e-12	2.7647	62
	voltage-gated cation channel activity	1.5840e-8	2.9567e-6	2.1059	3.3303e-11	3.0330	50
	potassium ion binding	5.4853e-8	9.0346e-6	2.0688	1.6418e-8	2.8154	43
	extracellular ligand-gated ion channel activity	2.6791e-6	0.0004	2.2487	8.5382e-7	3.2082	28
	excitatory extracellular ligand-gated ion channel activity	3.1568e-6	0.0004	2.4910	3.5373e-6	3.6860	21
	voltage-gated calcium channel activity	4.1770e-5	0.0045	2.9562	0.0018	3.8075	12
	neurotransmitter transporter activity	5.1973e-5	0.0050	4.1286	0.0429	3.3749	9
	glutamate receptor activity	8.0650e-5	0.0066	2.1672	1.0008e-7	4.9498	18
GO Biological Process	neurotransmitter transport	1.9784e-9	1.1463e-6	2.9707	3.9691e-5	2.9790	26
	potassium ion transport	4.4179e-9	2.3040e-6	2.0745	7.3303e-10	2.8428	51
	secretion by cell	4.8275e-6	0.0011	2.0303	0.0097	1.9980	31
	acid secretion	0.0003	0.0269	3.6490	0.0319	4.9498	6
GO Cellular Component	ion channel complex	2.4766e-9	3.4425e-7	2.2568	2.0454e-11	3.0393	49
	cation channel complex	4.3799e-9	5.2184e-7	2.2729	4.1686e-11	3.1461	45
	voltage-gated potassium channel complex	3.6769e-6	0.0002	2.1709	3.7475e-7	3.0554	30
	synaptic vesicle	6.6753e-6	0.0003	2.5735	0.0001	3.1349	19
	voltage-gated calcium channel complex	5.2759e-5	0.0022	3.2663	0.0039	3.7498	10
	calcium channel complex	0.0003	0.0100	2.6667	0.0027	3.6298	11
	synaptosome	0.0011	0.0257	2.0738	0.0094	2.5384	16
Mouse Phenotype	abnormal synaptic transmission	8.3480e-26	4.8260e-22	2.0782	6.8005e-35	3.0600	148
	abnormal CNS synaptic transmission	2.0501e-22	5.9257e-19	2.0843	1.2993e-31	3.1427	128
	abnormal miniature excitatory postsynaptic currents	8.7322e-13	1.0096e-9	4.5723	3.7452e-8	4.8982	19
	abnormal excitatory postsynaptic currents	8.0395e-12	7.7461e-9	2.7837	5.8037e-11	3.8957	34
	abnormal neurotransmitter level	7.9516e-11	6.5669e-8	3.0199	2.6786e-6	3.2737	25
	abnormal touch/ nociception	1.5816e-10	1.1429e-7	2.1530	2.6007e-12	2.9355	58
	abnormal inhibitory postsynaptic currents	7.7688e-10	4.0828e-7	3.0215	2.7203e-9	4.2057	26
	abnormal neurotransmitter secretion	1.0699e-9	4.7578e-7	3.0365	0.0001	3.1131	20
	abnormal circulating potassium level	3.2989e-8	1.1218e-5	4.4168	0.0026	3.9284	10
	abnormal pain threshold	3.5390e-8	1.1366e-5	2.1291	1.8127e-8	2.8400	42
PANTHER pathway	Ionotropic glutamate receptor pathway	4.0644e-5	0.0020	2.2878	2.3182e-6	3.9284	20
	Synaptic vesicle trafficking	0.0002	0.0056	3.5920	0.0080	3.7123	9
Predicted Promoter Motifs	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif TTCAGCACCACGGACAGMGCC which matches annotation for REST: RE1-silencing transcription factor	6.8383e-47	4.2055e-44	5.5282	9.6830e-55	6.9332	79
	Genes with promoter regions [-2kb,2kb] around transcription start site containing motif CAGNWMCNNNGAC. Motif does not match any known transcription factor	8.0892e-30	2.4874e-27	4.9207	6.4693e-30	5.5321	57
InterPro	Ion transport	6.8644e-7	0.0009	2.0443	4.7124e-10	3.3784	43
	Extracellular ligand-binding receptor	2.3723e-5	0.0112	2.2839	6.5158e-6	4.3540	19
	Neurotransmitter-gated ion-channel, conserved site	0.0001	0.0339	2.6041	0.0072	3.0696	16
	Neurotransmitter-gated ion-channel ligand-binding	0.0001	0.0339	2.6041	0.0072	3.0696	16
	Neurotransmitter-gated ion-channel	0.0001	0.0339	2.6041	0.0072	3.0696	16
	Neurotransmitter-gated ion-channel transmembrane region	0.0001	0.0339	2.6041	0.0072	3.0696	16
	Voltage-dependent potassium channel	0.0002	0.0314	2.3935	0.0003	3.9998	16
HGNC Gene Families	CACN	1.0532e-7	1.2533e-5	3.9509	0.0001	4.2898	13
	KCN	3.7073e-6	0.0003	2.0175	1.3878e-8	3.2442	35
	KV	0.0003	0.0114	2.1680	3.5454e-7	4.2306	20

Supplementary Table 36: "Gene-based GREAT" enrichments of all genes that possess an NRSF binding peak within 2 kb of its transcription start site. Shown are the top ten hypergeometric enriched terms at a false discovery rate of 0.05.

Ontology	Term	Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Observed Gene Hits
GO Molecular Function	gated channel activity	7.0232e-10	1.9665e-6	29
	channel activity	2.2608e-8	3.1511e-5	31
	ion channel activity	2.7122e-8	2.5314e-5	30
	substrate specific channel activity	4.6749e-8	3.2724e-5	30
	transporter activity	9.4958e-8	5.3177e-5	60
	extracellular ligand-gated ion channel activity	2.8926e-7	0.0001	12
	glutamate receptor activity	6.6412e-7	0.0003	8
	metal ion transmembrane transporter activity	8.2976e-7	0.0003	24
	ion transmembrane transporter activity	1.0130e-6	0.0003	41
	transmembrane transporter activity	1.0398e-6	0.0003	47
GO Biological Process	synaptic transmission	4.5482e-12	2.3719e-8	29
	transmission of nerve impulse	1.1148e-11	2.9069e-8	31
	cell-cell signaling	4.8939e-8	8.5073e-5	38
	secretion	6.4895e-8	8.4607e-5	22
	ion transport	3.2498e-7	0.0003	42
	transport	5.3108e-7	0.0005	102
	localization	6.3871e-7	0.0005	114
	establishment of localization	7.2216e-7	0.0005	102
	secretion by cell	1.1111e-6	0.0006	15
	generation of a signal involved in cell-cell signaling	1.2422e-6	0.0006	10
GO Cellular Component	synapse	3.5143e-14	2.9310e-11	34
	cell junction	4.3565e-10	1.8166e-7	38
	synapse part	1.7587e-9	4.8892e-7	22
	plasma membrane part	8.6004e-9	1.7932e-6	87
	neuron projection	3.0639e-7	5.1107e-5	16
	membrane part	3.7971e-7	5.2780e-5	195
	synaptic vesicle	4.8967e-7	5.8341e-5	10
	membrane	7.3669e-7	7.6800e-5	225
	intrinsic to membrane	4.1092e-6	0.0004	173
	integral to membrane	4.6276e-6	0.0004	170
Mouse Phenotype	abnormal synaptic transmission	1.1494e-24	6.6450e-21	55
	abnormal CNS synaptic transmission	3.8354e-23	1.1086e-19	49
	abnormal nervous system physiology	8.2955e-22	1.5985e-18	86
	abnormal behavior	9.9315e-18	1.4354e-14	102
	nervous system phenotype	1.3544e-15	1.5659e-12	110
	abnormal nervous system electrophysiology	1.9843e-13	1.9119e-10	27
	abnormal motor capabilities/coordination/movement	1.4200e-12	1.1727e-9	71
	abnormal excitatory postsynaptic currents	3.3137e-11	2.3946e-8	16
	abnormal anxiety-related response	1.5644e-10	1.0049e-7	21
	seizures	2.3286e-10	1.3462e-7	27
PANTHER Pathway	Ionotropic glutamate receptor pathway	9.0651e-5	0.0136	7
	Synaptic vesicle trafficking	0.0001	0.0095	5
	Metabotropic glutamate receptor group III pathway	0.0002	0.0104	8
	Opioid proenkephalin pathway	0.0011	0.0413	5
	Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway	0.0012	0.0364	9
	Endogenous cannabinoid signaling	0.0013	0.0313	4
	Opioid proopiomelanocortin pathway	0.0013	0.0273	5
MSigDB Pathway	Genes involved in neuroactive ligand-receptor interaction	5.7412e-10	4.0533e-7	27

Ontology	Term	Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Observed Gene Hits
MSigDB Perturbation	Downregulated in correlation with overt Alzheimer's Disease, in the CA1 region of the hippocampus	1.3391e-9	1.2199e-6	64
	The 30 genes showing the greatest decrease in expression in NBa Ews/Fli-1 infectants	4.2760e-5	0.0195	6
	Genes highly asscciated with favorable response to treatment for medulloblastoma	0.0001	0.0448	6
Predicted Promoter Motifs	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif TTCAGCACCACGGACAGMGCC which matches annotation for REST: RE1-silencing transcription factor	9.1016e-108	5.5975e-105	77
	Genes with promoter regions [-2kb,2kb] around transcription start site containing motif CAGNWMCNNNGAC. Motif does not match any known transcription factor	5.2045e-65	1.6004e-62	54
	Genes with promoter regions [-2kb,2kb] around transcription start site containing motif GGARNTKYCCA. Motif does not match any known transcription factor	8.9388e-12	1.8325e-9	17
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif GGGAGGRR which matches annotation for MAZ: MYC-associated zinc finger protein (purine-binding transcription factor)	6.0069e-9	9.2356e-7	99
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif CYYTGACGTCA which matches annotation for ATF1: activating transcription factor 1	2.7615e-8	3.3967e-6	23
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif GGGCGGR which matches annotation for SP1: Sp1 transcription factor	1.9629e-7	2.0120e-5	114
	Genes with promoter regions [-2kb,2kb] around transcription start site containing motif NTGACGTCANYS. Motif does not match any known transcription factor	9.4521e-7	8.3044e-5	18
	Genes with promoter regions [-2kb,2kb] around transcription start site containing motif AACTTT. Motif does not match any known transcription factor	2.0885e-6	0.0002	79
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif VGTGACGTMACN which matches annotation for ATF2: activating transcription factor 2	2.6193e-6	0.0002	21
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif TGAYRTCA which matches annotation for ATF3: activating transcription factor 3	3.0665e-6	0.0002	32
InterPro	Potassium channel, voltage-dependent, EAG/ELK/ERG	2.3347e-7	0.0015	6
	Gamma-aminobutyric acid A receptor	5.8761e-6	0.0194	6
	Cyclic nucleotide-binding, conserved site	8.9400e-6	0.0196	7
	Cyclic nucleotide-binding	1.4135e-5	0.0233	7
	Ion transport	1.7025e-5	0.0224	12
	Chromogranin/secretogranin	1.7153e-5	0.0188	3
	Cyclic nucleotide-binding-like	2.6410e-5	0.0249	7
	Extracellular ligand-binding receptor	3.2061e-5	0.0264	7
	RmlC-like jelly roll fold	3.8670e-5	0.0283	7
	Chromogranin, conserved site	6.7292e-5	0.0443	3
HGNC Gene Families	KV	5.5253e-5	0.0132	7
	CNG	8.1707e-5	0.0097	4
	KCN	0.0005	0.0385	9

Supplementary Table 37: GREAT enrichments of NRSF using the *basal plus extension* association rule with a maximum a basal regulatory region extending 5 kb upstream and 1 kb downstream of the transcription start site and extension of 50 kb. Shown are the top ten binomial enriched terms at a false discovery rate of 0.05 with a fold enrichment of at least two that are also significant by the hypergeometric test, using the highest-scoring NRSF peaks anywhere in the genome (QuEST score > 1; n = 1,712).

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Molecular Function	transporter activity	6.0427e-37	8.4598e-34	3.2142	9.7745e-15	2.0837	139
	ion transmembrane transporter activity	1.4346e-35	1.3398e-32	3.8760	1.5131e-16	2.5472	105
	ion channel activity	7.3009e-35	5.1107e-32	5.2112	4.0315e-22	3.7223	80
	gated channel activity	1.2489e-34	6.9936e-32	5.7506	2.6124e-23	4.2295	73
	channel activity	2.7445e-34	1.2808e-31	5.0561	1.2034e-21	3.5922	81
	substrate specific channel activity	2.9608e-34	1.1843e-31	5.1106	1.4460e-21	3.6330	80
	substrate-specific transmembrane transporter activity	6.2047e-33	2.1716e-30	3.5565	1.2436e-14	2.3429	108
	substrate-specific transporter activity	1.4575e-31	4.6346e-29	3.2575	1.3710e-12	2.1126	116
	transmembrane transporter activity	1.5942e-31	4.4839e-29	3.3491	1.4798e-13	2.2178	112
	cation transmembrane transporter activity	1.4411e-30	3.6682e-28	4.1763	1.2832e-13	2.6661	80
GO Biological Process	ion transport	2.9784e-33	1.6532e-29	3.7590	3.0373e-16	2.5410	104
	system process	5.3360e-32	1.3914e-28	2.9392	6.8089e-9	1.7673	140
	localization	8.0045e-32	1.3814e-28	2.1506	4.9031e-6	1.4008	235
	transport	1.4300e-30	1.8843e-27	2.2388	4.8083e-6	1.4393	208
	establishment of localization	3.7760e-30	3.9384e-27	2.2232	7.7261e-6	1.4300	206
	neurological system process	4.7935e-29	4.1663e-26	3.0960	9.4937e-8	1.8081	117
	cell-cell signaling	5.5295e-29	3.6045e-26	3.8369	2.9861e-12	2.5468	83
	synaptic transmission	1.2998e-28	7.5316e-26	5.8079	2.2588e-15	3.9190	55
	transmission of nerve impulse	9.3418e-28	4.4269e-25	5.2437	1.0442e-14	3.6629	59
	cation transport	1.7340e-25	7.5358e-23	3.6527	2.1494e-10	2.5114	73
GO Cellular Component	membrane part	6.9561e-56	5.8014e-53	2.0736	2.3508e-14	1.3737	453
	intrinsic to membrane	6.7802e-51	1.8849e-48	2.0981	9.8117e-13	1.3791	405
	integral to membrane	4.1574e-50	8.6882e-48	2.1010	4.5660e-12	1.3720	395
	plasma membrane	6.6802e-41	7.8398e-39	2.2132	4.5703e-12	1.4846	295
	synapse	4.0590e-34	5.0655e-32	5.8006	1.7895e-20	4.0971	66
	plasma membrane part	1.6870e-32	1.5633e-30	2.4886	2.0720e-13	1.7704	169
	cell junction	9.2206e-23	7.6900e-21	3.6648	1.4328e-12	2.6620	73
	synapse part	1.3216e-22	1.0020e-20	5.7613	1.5971e-14	4.2339	45
	intrinsic to plasma membrane	1.4695e-17	1.0213e-15	2.3314	8.9125e-7	1.6896	114
	integral to plasma membrane	2.3769e-17	1.6248e-15	2.3365	1.1971e-6	1.6879	112
Mouse Phenotype	abnormal synaptic transmission	6.2853e-66	3.6395e-62	6.8925	6.0945e-37	4.4815	105
	abnormal nervous system physiology	1.1236e-63	3.2446e-60	4.0620	1.5474e-36	2.6365	168
	abnormal CNS synaptic transmission	2.3940e-57	4.6132e-54	7.0194	2.4808e-31	4.5108	89
	abnormal behavior	8.7362e-55	1.2626e-51	3.1418	3.1607e-24	2.1484	205
	nervous system phenotype	9.0571e-54	1.0472e-50	2.8731	1.5233e-20	1.9946	223
	abnormal motor capabilities/coordination/movement	4.6455e-35	3.6374e-32	3.0524	2.0501e-14	2.1002	136
	abnormal nervous system electrophysiology	3.4066e-30	2.4817e-27	6.7148	5.2105e-14	4.2106	45
	abnormal sensory capabilities/reflexes/nociception	3.6630e-28	2.3526e-25	4.3667	1.7760e-13	2.9596	69
	abnormal locomotor activity	4.3811e-28	2.5327e-25	3.1585	1.1111e-11	2.1838	104
	abnormal voluntary movement	1.9898e-27	1.0510e-24	3.0521	4.3739e-11	2.1017	107
PANTHER Pathway	Ionotropic glutamate receptor pathway	9.4313e-14	1.4147e-11	9.7430	1.0155e-6	6.0620	15
	Metabotropic glutamate receptor group III pathway	1.0737e-8	8.0531e-7	5.9293	0.0130	3.2437	12
	Heterotrimeric G-protein signaling pathway-Gq alpha and Go alpha mediated pathway	1.7135e-7	8.5674e-6	4.1667	0.0297	2.5292	15
	Synaptic vesicle trafficking	3.2057e-7	1.2021e-5	10.4726	0.0070	5.9934	7
MSigDB Pathway	Genes involved in neuroactive ligand-receptor interaction	3.9134e-14	2.7628e-11	3.7849	8.0367e-7	2.8049	42
Predicted Promoter Motifs	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif TTCAGCACCACGGACAGMGCC which matches annotation for REST: RE1-silencing transcription factor	1.0180e-94	6.2610e-92	22.3065	4.8419e-76	13.9498	77
	Genes with promoter regions [-2kb,2kb] around transcription start site containing motif CAGNMMCNNNGAC. Motif does not match any known transcription factor	3.8168e-62	1.1736e-59	18.2589	7.0480e-46	11.2195	56
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif GGGAGGRR which matches annotation for MAZ: MYC-associated zinc finger protein (purine-binding transcription factor)	9.6955e-31	2.0269e-28	2.3624	1.1363e-6	1.4894	189
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif CAGSTG which matches annotation for TCF3: transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)	2.1521e-25	3.3088e-23	2.1228	9.3055e-5	1.3942	191
	Genes with promoter regions [-2kb,2kb] around transcription start site containing motif CTGCAGY. Motif does not match any known transcription factor	7.1449e-16	7.3235e-14	2.6830	0.0012	1.8715	69
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif GTGGGSGCRRS which matches annotation for EGR1: early growth response 1 EGR2: early growth response 2 (Krox-20 homolog, Drosophila) EGR3: early growth response 3	6.2202e-15	5.4649e-13	4.0619	0.0009	2.2270	34
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif CACCTG which matches annotation for REPIN1: replication initiator 1	4.6152e-14	3.5479e-12	2.0612	0.0242	1.3323	110
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif GCAMCTGNY which matches annotation for MYOD1: myogenic differentiation 1	4.3084e-13	2.6603e-11	2.3302	0.0113	1.4911	76
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif GGGTGGRR which matches annotation for PAX4: paired box gene 4	1.3061e-12	7.3021e-11	2.1073	0.0396	1.3337	95
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif WTGCGTGGSCGK which matches annotation for EGR1: early growth response 1	5.7309e-11	2.9371e-9	3.6624	0.0122	1.9634	29
InterPro	Ion transport	2.5481e-15	1.6771e-11	6.0564	7.6889e-9	4.7034	29
	Immunoglobulin subtype 2	7.0949e-11	2.3367e-7	3.7763	0.0007	2.6935	31
	Immunoglobulin	1.9610e-10	3.4275e-7	3.9469	0.0007	2.7917	30
	Immunoglobulin-like	1.6570e-10	2.7286e-7	2.7684	0.0023	1.9635	52
	Fibronectin, type III	5.6745e-9	7.4756e-6	3.4746	0.0169	2.3392	25
	Potassium channel, voltage-dependent, EAG/ELK/ERG	7.3799e-9	8.1019e-6	16.4212	4.9491e-6	12.7723	9
	Immunoglobulin-like fold	9.9242e-9	9.3387e-6	2.6538	0.0101	1.8683	40
	Extracellular ligand-binding receptor	1.0545e-8	8.6827e-6	7.5866	0.0007	5.6766	12
	Immunoglobulin subtype	2.2621e-8	1.6565e-5	2.9468	0.0064	2.1359	37
	Fibronectin, type III-like fold	3.0759e-8	2.0261e-5	3.2203	0.0497	2.1602	25
TreeFam	Voltage-dependent calcium channel subunit	6.9969e-7	0.0014	20.3694	0.0327	14.1914	5
	Carcinoembryonic antigen-related cell adhesion molecule precursor	1.4360e-6	0.0020	13.2316	0.0232	8.5148	7
	Potassium voltage-gated channel subfamily C member	5.4018e-5	0.0194	20.2503	0.0326	17.0297	4
HGNC Gene Families	CACN	7.2158e-12	1.7174e-9	11.8880	9.5893e-7	8.1742	12
	KCN	6.1191e-11	7.2817e-9	5.4475	7.1956e-7	4.2036	22
	KV	2.3813e-7	1.8892e-5	6.4011	1.4080e-5	5.5766	13
	CAV	5.8313e-7	3.4686e-5	15.1762	0.0042	9.4609	5
	IG	7.2558e-6	0.0003	8.3848	0.0055	5.4185	7
	SIGLEC	1.3748e-5	0.0006	17.1394	0.0054	8.5148	5
	CNG	0.0001	0.0031	11.1712	0.0054	8.5148	5
	LILR	0.0013	0.0238	14.3456	0.0310	7.5667	4

Supplementary Table 38: GREAT enrichments of NRSF using the *two nearest genes* association rule with a maximum extension of 1 Mb. Shown are the top ten binomial enriched terms at a false discovery rate of 0.05 with a fold enrichment of at least two that are also significant by the hypergeometric test, using the highest-scoring NRSF peaks anywhere in the genome (QuEST score > 1; n = 1,712).

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Molecular Function	ion channel activity	1.5008e-16	1.4008e-13	2.0023	6.1720e-19	2.4501	120
	gated channel activity	3.3229e-16	1.8608e-13	2.0795	8.4824e-20	2.6942	106
	cation channel activity	3.2797e-12	8.3482e-10	2.0275	1.6745e-15	2.6231	86
	voltage-gated ion channel activity	5.0087e-12	1.1687e-9	2.2021	4.7134e-11	2.5851	64
	voltage-gated cation channel activity	4.5224e-9	9.0448e-7	2.1481	9.2151e-11	2.8572	52
	potassium ion binding	3.5695e-8	5.8791e-6	2.1002	2.9342e-8	2.6688	45
	extracellular ligand-gated ion channel activity	1.4269e-6	0.0002	2.2818	4.1278e-7	3.1136	30
	excitatory extracellular ligand-gated ion channel activity	1.4400e-6	0.0002	2.5443	6.4384e-7	3.6568	23
	voltage-gated calcium channel activity	1.4865e-5	0.0015	3.0831	0.0008	3.7363	13
	glutamate receptor activity	3.8823e-5	0.0037	2.2243	4.8598e-8	4.7327	19
GO Biological Process	neurotransmitter transport	9.5847e-10	4.5440e-7	2.9995	7.1955e-5	2.8022	27
	potassium ion transport	2.9877e-9	1.2984e-6	2.0824	9.8793e-9	2.6255	52
	secretion	8.0528e-9	2.4703e-6	2.0025	0.0010	1.8342	54
	secretion by cell	7.1850e-7	0.0002	2.1225	0.0060	1.9849	34
	arachidonic acid secretion	3.6236e-5	0.0042	4.6627	0.0123	5.6045	6
	acid secretion	8.1250e-5	0.0078	3.9353	0.0069	5.2309	7
GO Cellular Component	ion channel complex	6.6417e-10	6.1546e-8	2.3007	5.2239e-11	2.8655	51
	cation channel complex	1.1091e-9	9.2498e-8	2.3220	7.4053e-11	2.9764	47
	synaptic vesicle	3.1677e-6	0.0002	2.6264	0.0005	2.8396	19
	voltage-gated potassium channel complex	4.5769e-6	0.0002	2.1513	3.6780e-6	2.7677	30
	voltage-gated calcium channel complex	4.6569e-6	0.0002	3.6281	0.0002	4.0760	12
	calcium channel complex	4.3094e-5	0.0016	2.9492	0.0002	3.8858	13
	synaptosome	0.0003	0.0072	2.2240	0.0090	2.4430	17
Mouse Phenotype	abnormal synaptic transmission	1.0072e-27	5.8228e-24	2.1183	5.6900e-33	2.8655	153
	abnormal CNS synaptic transmission	1.0852e-23	2.0911e-20	2.1144	8.8928e-30	2.9357	132
	abnormal miniature excitatory postsynaptic currents	1.2321e-12	1.4245e-9	4.5142	2.0726e-7	4.4369	19
	abnormal excitatory postsynaptic currents	4.1101e-12	3.3944e-9	2.8025	8.8127e-10	3.5288	34
	abnormal neurotransmitter level	1.1920e-11	8.6137e-9	3.1088	4.2907e-6	3.0840	26
	abnormal touch/ nociception	5.4198e-11	3.1332e-8	2.1785	4.6203e-11	2.7048	59
	abnormal nervous system electrophysiology	7.3810e-11	3.5558e-8	2.0356	1.9714e-12	2.7099	66
	abnormal neurotransmitter secretion	4.6761e-10	1.8022e-7	3.0751	0.0001	2.9609	21
	abnormal inhibitory postsynaptic currents	1.0258e-9	3.7062e-7	2.9924	2.4110e-8	3.8096	26
	abnormal Purkinje cell innervation	5.1857e-9	1.7635e-6	4.5071	0.0015	4.9818	8
PANTHER Pathway	Ionotropic glutamate receptor pathway	1.9678e-5	0.0010	2.3431	1.9833e-6	3.7363	21
	Synaptic vesicle trafficking	0.0002	0.0056	3.5223	0.0141	3.3627	9
	Endogenous cannabinoid signaling	0.0006	0.0131	3.1383	0.0364	3.1464	8
	Thyrotropin-releasing hormone receptor signaling pathway	0.0009	0.0146	2.1123	0.0151	2.3444	16
	Adrenaline and noradrenaline biosynthesis	0.0017	0.0216	2.9317	0.0444	2.8022	9
Predicted Promoter Motifs	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif TTCAGCACCACGGACAGMGCC which matches annotation for REST: RE1-silencing transcription factor	1.0197e-48	6.2714e-46	5.6078	2.2333e-51	6.2802	79
	Genes with promoter regions [-2kb,2kb] around transcription start site containing motif CAGNWMCNNNGAC. Motif does not match any known transcription factor	7.0810e-31	2.1774e-28	4.9788	1.3161e-27	5.0111	57
InterPro	Ion transport	9.4755e-7	0.0012	2.0227	1.4492e-8	3.0602	43
	Extracellular ligand-binding receptor	1.1280e-5	0.0057	2.3382	5.8512e-6	4.1515	20
	Neurotransmitter-gated ion-channel, conserved site	6.1580e-5	0.0184	2.6979	0.0058	2.9543	17
	Neurotransmitter-gated ion-channel ligand-binding	6.1580e-5	0.0184	2.6979	0.0058	2.9543	17
	Neurotransmitter-gated ion-channel	6.1580e-5	0.0184	2.6979	0.0058	2.9543	17
	Neurotransmitter-gated ion-channel transmembrane region	6.1580e-5	0.0184	2.6979	0.0058	2.9543	17
	Voltage-dependent potassium channel	0.0002	0.0293	2.3732	0.0010	3.6231	16
HGNC Gene Families	CACN	3.0307e-8	7.2130e-6	4.0854	4.7129e-5	4.1847	14
	KCN	2.2116e-6	0.0002	2.0396	4.8527e-8	3.0226	36
	SIGLEC	2.3323e-5	0.0014	15.3330	0.0440	4.4836	6
	KV	0.0003	0.0129	2.1501	1.9980e-6	3.8321	20

Supplementary Table 39: GREAT enrichments of NRSF using the *single nearest gene* association rule with a maximum extension of 1 Mb. Shown are the top ten binomial enriched terms at a false discovery rate of 0.05 with a fold enrichment of at least two that are also significant by the hypergeometric test, using the highest-scoring NRSF peaks anywhere in the genome (QuEST score > 1; n = 1,712).

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Molecular Function	ion transmembrane transporter activity	1.0686e-17	1.4964e-14	2.0805	4.4417e-16	2.2958	123
	gated channel activity	2.5929e-17	2.4201e-14	2.5132	1.8263e-22	3.6545	82
	ion channel activity	7.2864e-17	5.1005e-14	2.3698	6.7660e-21	3.2220	90
	substrate specific channel activity	1.1789e-16	6.6019e-14	2.3535	2.8181e-20	3.1447	90
	channel activity	1.4524e-16	6.7781e-14	2.3361	3.1087e-20	3.1051	91
	cation transmembrane transporter activity	2.1382e-15	7.4837e-13	2.2371	1.3490e-12	2.3590	92
	cation channel activity	1.0594e-13	2.6967e-11	2.5123	4.0842e-16	3.6832	67
	metal ion transmembrane transporter activity	5.0180e-13	1.1709e-10	2.3499	5.8951e-16	3.1429	71
	voltage-gated ion channel activity	1.2784e-11	2.7535e-9	2.6978	2.2137e-11	3.3288	47
	voltage-gated cation channel activity	2.4222e-10	4.8444e-8	2.6767	1.6915e-12	3.9501	41
GO Biological Process	cation transport	8.0678e-13	2.1037e-9	2.0744	5.8735e-10	2.2500	85
	metal ion transport	4.3665e-12	7.5905e-9	2.0812	3.9167e-11	2.4704	79
	neurotransmitter transport	5.3493e-11	5.5793e-8	4.1234	6.7797e-7	4.1858	23
	transmission of nerve impulse	8.8821e-10	5.7900e-7	2.0635	7.5431e-15	3.2060	69
	synaptic transmission	1.3571e-9	7.8638e-7	2.1587	4.3188e-14	3.3442	61
	potassium ion transport	4.1852e-8	1.8188e-5	2.3352	6.1261e-10	3.5413	40
	calcium ion transport	4.6190e-8	1.8529e-5	2.7238	6.3506e-5	3.0148	26
	monovalent inorganic cation transport	9.8806e-8	3.4351e-5	2.0042	2.9855e-7	2.4183	55
	cellular ion homeostasis	1.3864e-7	4.2593e-5	2.0746	0.0004	2.1041	44
	cellular chemical homeostasis	1.5008e-7	4.3480e-5	2.0559	0.0003	2.1133	45
GO Cellular Component	synapse	6.8638e-13	1.1449e-10	2.2089	3.2465e-20	3.5735	75
	ion channel complex	1.5078e-10	2.0959e-8	2.6446	4.2744e-11	3.7436	38
	cation channel complex	2.0295e-9	2.4180e-7	2.7595	4.5533e-10	3.7754	34
	synapse part	7.5768e-8	6.3207e-6	2.0834	1.6318e-14	3.6919	51
	synaptic vesicle	3.1314e-7	1.8654e-5	3.7002	2.2917e-5	4.1929	16
	voltage-gated calcium channel complex	4.1361e-7	2.2997e-5	5.6900	0.0050	4.7646	8
	calcium channel complex	7.8999e-6	0.0004	4.3143	0.0131	4.1929	8
	clathrin-coated vesicle	1.9303e-5	0.0009	2.6331	0.0050	2.6205	18
	coated vesicle	3.7866e-5	0.0015	2.4718	0.0146	2.3267	19
	voltage-gated potassium channel complex	3.8177e-5	0.0014	2.3345	2.1491e-7	3.6823	24
Mouse Phenotype	abnormal synaptic transmission	4.8662e-31	2.8132e-27	2.6663	1.0292e-36	3.9078	119
	abnormal CNS synaptic transmission	2.0201e-27	5.8391e-24	2.6894	3.2614e-32	3.9776	102
	abnormal miniature excitatory postsynaptic currents	2.2017e-15	3.1820e-12	6.5854	3.4255e-9	6.9608	17
	abnormal nervous system electrophysiology	8.4892e-14	8.1793e-11	2.7178	1.6394e-14	3.7436	52
	abnormal inhibitory postsynaptic currents	7.2252e-13	5.2211e-10	4.4781	1.2073e-12	6.4229	25
	abnormal excitatory postsynaptic currents	7.2783e-13	4.6751e-10	3.6642	2.3580e-10	4.9135	27
	abnormal neurotransmitter level	3.6408e-12	2.1048e-9	4.0176	7.4554e-8	4.5756	22
	abnormal touch/ nociception	5.7838e-12	3.0397e-9	2.7504	4.3928e-9	3.2154	40
	seizures	3.4626e-11	1.5398e-8	2.2757	4.2019e-13	3.1554	59
	abnormal neurotransmitter secretion	9.9134e-11	4.0935e-8	4.0211	1.5953e-5	4.2028	17
PANTHER Pathway	Ionotropic glutamate receptor pathway	4.1905e-6	0.0006	3.0509	6.2355e-8	5.6155	18
	Opioid proopiomelanocortin pathway	5.1855e-6	0.0004	4.4811	0.0120	3.6851	9
	Endogenous cannabinoid signaling	5.3608e-6	0.0003	5.7536	0.0099	4.6273	7
	5HT4 type receptor mediated signaling pathway	3.0659e-5	0.0011	4.3657	0.0216	3.4941	8
	Opioid proenkephalin pathway	8.6826e-5	0.0022	3.9072	0.0227	3.3814	8
	Thyrotropin-releasing hormone receptor signaling pathway	0.0003	0.0055	2.6365	0.0227	2.6261	11
	Corticotropin releasing factor receptor signaling pathway	0.0004	0.0070	4.2148	0.0165	3.7436	8
	Dopamine receptor mediated signaling pathway	0.0006	0.0093	2.6369	0.0135	2.9117	12
	Metabotropic glutamate receptor group III pathway	0.0008	0.0126	2.1199	4.8674e-5	3.7436	18
	Adrenaline and noradrenaline biosynthesis	0.0019	0.0233	3.6653	0.0220	3.8216	7
Predicted Promoter Motifs	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif TTCAGCACCACGGACAGMGCC which matches annotation for REST: RE1-silencing transcription factor	2.3531e-62	1.4472e-59	9.0171	7.0169e-59	10.6725	78
	Genes with promoter regions [-2kb,2kb] around transcription start site containing motif CAGNWMCNNNGAC. Motif does not match any known transcription factor	1.0441e-39	3.2106e-37	7.7188	1.4051e-39	8.6324	56
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif NNGGGGCGGGGNN which matches annotation for SP1: Sp1 transcription factor	3.1347e-8	6.4262e-6	2.3571	0.0056	1.8183	34
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif GGGSTCWR which matches annotation for ITGAL: integrin, alpha L (antigen CD11A (p180), lymphocyte function-associated antigen 1, alpha polypeptide)	2.4746e-7	3.8048e-5	2.2639	0.0031	1.9038	34
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif NGGGGGCGGGGYN which matches annotation for SP1: Sp1 transcription factor	2.0156e-6	0.0002	2.1173	0.0186	1.6647	31
	Genes with promoter regions [-2kb,2kb] around transcription start site containing the motif CSGCGCGR which matches annotation for ZFP161: zinc finger protein 161 homolog (mouse)	4.0319e-6	0.0003	2.0112	0.0043	1.8972	32
	Genes with promoter regions [-2kb,2kb] around transcription start site containing motif YRCCAKNNGNCGC. Motif does not match any known transcription factor	0.0003	0.0108	2.6195	0.0106	2.5423	13
InterPro	Ion transport	7.4080e-8	0.0005	2.5623	9.4341e-10	4.2428	34
	Extracellular ligand-binding receptor	3.1935e-5	0.0351	2.6108	6.4086e-7	6.1874	17
	Potassium channel, voltage-dependent, EAG/ELK/ERG	8.6197e-5	0.0437	4.2245	8.1606e-7	10.9189	10
HGNC Gene Families	CACN	1.8861e-9	4.4890e-7	6.2309	0.0006	5.2411	10
	KCN	7.0177e-5	0.0056	2.1243	5.6682e-8	3.9750	27
	KV	0.0007	0.0323	2.4263	1.0080e-6	5.3755	16

Supplementary Table 40: Enrichment for regions bound by GABP in human Jurkat cells. **(a)** The top ten proximal binding gene-based enrichments (reproduced from ref. 19). **(b)** GREAT *cis*-regulatory element enrichments for all regions bound by GABP.

a

Gene-based GO Enrichments of GABP Promoter Binding Peaks

Term	p-value
nucleus	1.21×10^{-240}
protein binding	1.87×10^{-120}
transcription	2.03×10^{-80}
nucleotide binding	1.13×10^{-77}
metal ion binding	9.06×10^{-73}
RNA binding	4.19×10^{-68}
intracellular	1.60×10^{-67}
DNA binding	2.35×10^{-65}
zinc ion binding	5.12×10^{-64}
mitochondrion	2.98×10^{-62}

b

GREAT Enrichments of GABP Binding Peaks in Human Jurkat Cells

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Molecular Function	structural constituent of ribosome	4.5489e-25	1.2737e-21	3.3100	1.2151e-22	2.7230	93
	translation regulator activity	2.9929e-10	2.0950e-7	2.4762	1.7722e-7	2.1309	57
	nuclease activity	2.7452e-9	1.5373e-6	2.2633	0.0107	1.5666	55
	translation factor activity, nucleic acid binding	1.7071e-8	6.8284e-6	2.3911	2.4775e-7	2.2271	50
	RNA splicing factor activity, transesterification mechanism	7.4957e-8	2.6235e-5	4.8808	3.1700e-5	3.1403	18
	unfolded protein binding	3.0537e-7	9.5004e-5	2.2564	0.0007	1.8472	45
	transcription elongation regulator activity	5.4679e-7	0.0002	5.0761	0.0321	2.5122	12
	DNA-directed RNA polymerase activity	7.6845e-7	0.0002	3.3933	0.0176	2.0935	20
	retinoid-X receptor activity	1.0121e-6	0.0002	7.8129	0.0434	3.0451	8
	translation initiation factor activity	1.4156e-6	0.0003	2.6428	0.0019	2.0935	28
GO Biological Process	RNA processing	6.6273e-37	4.9373e-34	2.2316	4.9809e-41	2.1813	261
	translation	8.2428e-34	4.7762e-31	2.7888	2.9650e-32	2.4179	164
	nuclear mRNA splicing, via spliceosome	3.1698e-26	1.5028e-23	3.4054	2.0656e-18	2.5065	88
	RNA splicing	2.4400e-25	9.7880e-23	2.3894	5.7152e-27	2.3143	152
	ribonucleoprotein complex biogenesis	5.0060e-23	1.8647e-20	2.8001	8.9944e-24	2.5496	109
	ncRNA metabolic process	1.1075e-22	3.8505e-20	2.7865	4.6052e-18	2.2742	107
	translational elongation	1.1143e-21	3.2285e-19	3.9453	8.6555e-17	2.7771	65
	mRNA processing	1.5375e-18	3.8182e-16	2.0388	1.3605e-22	2.1422	154
	ribosome biogenesis	3.3727e-18	7.9947e-16	3.1998	2.0573e-15	2.6218	67
	ncRNA processing	8.5011e-18	1.9275e-15	2.7329	6.7274e-14	2.2545	84
GO Cellular Component	ribonucleoprotein complex	2.3165e-59	1.9320e-56	3.0144	8.3195e-57	2.5342	253
	ribosome	1.4526e-29	1.5143e-27	3.2160	6.9804e-27	2.5835	116
	spliceosome	2.0579e-23	1.7163e-21	3.1854	2.8336e-21	2.6673	86
	ribosomal subunit	2.3524e-20	1.4014e-18	3.4328	1.3885e-17	2.6889	70
	cytosolic ribosome	1.3981e-14	5.3001e-13	3.4075	6.6730e-14	2.8291	50
	small nuclear ribonucleoprotein complex	4.6792e-14	1.6967e-12	6.6480	3.1038e-8	3.4892	20
	large ribosomal subunit	4.0734e-13	1.4155e-11	3.6805	3.0246e-13	3.1015	40
	proteasome complex	6.5892e-13	2.1962e-11	3.9146	1.4457e-9	2.8786	33
	cytosolic part	3.0242e-12	9.7005e-11	2.5147	6.6710e-10	2.1257	66
	mitochondrial membrane part	7.6827e-12	2.2883e-10	2.7469	7.8343e-6	1.9066	51
Pathway Commons	Transcription	1.0524e-65	1.3187e-62	3.6609	3.3267e-50	2.6438	209
	Elongation of Intron-Containing Transcripts and co-transcriptional mRNA splicing	2.3747e-65	1.4877e-62	4.2681	1.7373e-44	2.7038	175
	Elongation and Processing of Capped Transcripts	2.3747e-65	1.4877e-62	4.2681	1.7373e-44	2.7038	175
	mRNA Processing	8.0356e-65	2.5172e-62	4.1802	3.1846e-45	2.7200	178
	mRNA Capping	8.1633e-65	2.0457e-62	4.1981	2.6755e-45	2.7147	177
	RNA Pol II CTD phosphorylation and interaction with CE	8.1633e-65	2.0457e-62	4.1981	2.6755e-45	2.7147	177
	RNA Polymerase II Transcription Initiation And Promoter Clearance	1.0264e-64	1.8373e-62	4.0355	5.0973e-45	2.6658	184
	Formation and Maturation of mRNA Transcript	1.0264e-64	1.8373e-62	4.0355	5.0973e-45	2.6658	184
	RNA Polymerase II Promoter Escape	1.0264e-64	1.8373e-62	4.0355	5.0973e-45	2.6658	184
	RNA Polymerase II Transcription Initiation	1.0264e-64	1.8373e-62	4.0355	5.0973e-45	2.6658	184
MSigDB Pathway	Genes involved in mRNA splicing	1.9443e-16	1.3727e-13	4.4098	6.8199e-10	2.8412	38
	RIBOSOMAL_PROTEINS	1.5552e-14	5.4897e-12	2.9125	2.4091e-14	2.6271	64
	Genes involved in ribosome	4.0239e-14	9.4697e-12	3.3626	3.7411e-14	2.9907	50
	MRNA_PROCESSING_REACTOME	1.0913e-10	1.5408e-8	2.2348	3.8779e-12	2.3774	67
	Genes involved in mRNA processing	5.2325e-8	4.6177e-6	3.0845	8.3483e-6	2.5486	28
	Genes involved in proteasome	5.9977e-8	4.7049e-6	4.6941	0.0045	2.6645	14
	Genes involved in oxidative phosphorylation	7.4103e-8	5.2317e-6	2.1875	2.2890e-5	1.9099	52
	Genes involved in electron transport	2.0313e-7	1.3038e-5	2.2461	0.0001	1.9070	46
	RNA_TRANSCRIPTION_REACTOME	9.9760e-7	5.8692e-5	3.4532	0.0349	1.9888	19
	Genes involved in glycosylphosphatidylinositol(GPI)-anchor biosynthesis	8.1779e-6	0.0004	3.8295	0.0304	2.3666	13
Transcription Factor Targets	Targets of ETS1, identified by ChIP-chip in Jurkat T-cells.	1.8169e-218	3.4521e-217	3.2266	3.2230e-212	2.7616	744
	Targets of YY1 identified by ChIP-chip.	3.1323e-74	1.9838e-73	2.7464	2.6395e-56	2.1831	341

Supplementary Table 41: "Gene-based GREAT" enrichments of all genes that possess an GABP binding peak within 2 kb of its transcription start site. Shown are the top ten hypergeometric enriched terms at a false discovery rate of 0.05.

Ontology	Term Name	Hypergeometric Results: Raw P-Value	FDR Q-Val	Observed Gene Hits
GO Molecular Function	nucleic acid binding	1.6855e-56	4.7194e-53	1067
	RNA binding	4.3765e-46	6.1271e-43	317
	structural constituent of ribosome	4.6834e-32	4.3712e-29	100
	DNA binding	3.7326e-16	2.6128e-13	670
	nucleotide binding	1.0885e-15	6.0962e-13	608
	binding	1.6987e-13	7.9271e-11	2956
	catalytic activity	2.4196e-13	9.5785e-11	1331
	protein binding	2.4689e-13	8.6341e-11	1912
	zinc ion binding	4.3667e-13	1.3585e-10	646
	hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	4.7127e-11	1.3196e-8	237
GO Biological Process	biopolymer metabolic process	2.7091e-125	1.4128e-121	1748
	cellular biopolymer metabolic process	3.3099e-125	8.8089e-122	1740
	cellular macromolecule metabolic process	3.5956e-121	6.2503e-118	1754
	macromolecule metabolic process	1.5286e-119	1.9930e-116	1762
	gene expression	8.4272e-110	8.7895e-107	994
	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	3.2682e-103	2.8580e-100	1114
	cellular metabolic process	1.2047e-100	8.9750e-98	2039
	primary metabolic process	8.4220e-92	5.4901e-89	1972
	metabolic process	9.3241e-87	5.4028e-84	2130
	RNA metabolic process	2.1869e-82	1.1405e-79	442
GO Cellular Component	intracellular	1.1749e-141	9.7978e-139	3135
	intracellular membrane-bounded organelle	7.0432e-138	2.9370e-135	2467
	membrane-bounded organelle	1.3493e-137	3.7510e-135	2467
	intracellular part	4.3790e-136	9.1303e-134	3033
	intracellular organelle	1.5523e-125	2.5892e-123	2648
	organelle	3.4694e-125	4.8224e-123	2648
	nucleus	5.3472e-90	6.3708e-88	1669
	ribonucleoprotein complex	2.5339e-77	2.6416e-75	275
	intracellular organelle part	1.8714e-65	1.7342e-63	1321
	organelle part	2.7699e-65	2.3101e-63	1325
Mouse Phenotype	abnormal cell content/ morphology	8.9635e-9	5.1818e-5	87
	embryonic lethality	1.0011e-6	0.0029	313
	abnormal inner cell mass	1.8654e-6	0.0036	35
	embryonic lethality before implantation	3.2302e-6	0.0047	48
	decreased cell proliferation	5.1983e-6	0.0060	73
	prenatal lethality	7.7317e-6	0.0074	406
	embryonic lethality before somite formation	8.1200e-6	0.0067	86
	cellular phenotype	1.0200e-5	0.0074	303
	abnormal cell physiology	1.9253e-5	0.0124	275
	abnormal blastocyst morphology	3.3417e-5	0.0193	39

Ontology	Term Name	Hypergeometric Results: Raw P-Value	FDR Q-Val	Observed Gene Hits
PANTHER Pathway	General transcription regulation	9.3231e-5	0.0140	17
	Transcription regulation by bZIP transcription factor	0.0001	0.0079	21
	mRNA splicing	0.0002	0.0084	6
	Tetrahydrofolate biosynthesis	0.0007	0.0268	5
	Parkinson disease	0.0015	0.0451	32
	General transcription by RNA polymerase I	0.0018	0.0450	8
Pathway Commons	Gene Expression	2.2038e-75	2.7614e-72	245
	Transcription	2.7287e-75	1.7095e-72	234
	RNA Polymerase II Transcription Initiation And Promoter Clearance	2.1512e-66	8.9849e-64	205
	Formation and Maturation of mRNA Transcript	2.1512e-66	8.9849e-64	205
	RNA Polymerase II Promoter Escape	2.1512e-66	8.9849e-64	205
	RNA Polymerase II Transcription Initiation	2.1512e-66	8.9849e-64	205
	RNA Polymerase II Transcription Pre-Initiation	2.1512e-66	8.9849e-64	205
	RNA Polymerase II Transcription	2.1512e-66	8.9849e-64	205
	mRNA Capping	4.6077e-63	6.4150e-61	194
	RNA Pol II CTD phosphorylation and interaction with CE	4.6077e-63	6.4150e-61	194
MSigDB Pathway	MRNA_PROCESSING_REACTOME	7.5237e-23	5.3117e-20	78
	RIBOSOMAL_PROTEINS	2.0811e-20	7.3463e-18	68
	Genes involved in ribosome	3.3584e-20	7.9034e-18	53
	Genes involved in mRNA splicing	4.6592e-15	8.2235e-13	41
	Mitochondrial genes	2.6609e-13	3.7572e-11	172
	PGC related genes	2.9192e-13	3.4349e-11	163
	Mitochondrial genes	7.0362e-13	7.0965e-11	164
	Genes involved in mRNA processing	3.1617e-10	2.7902e-8	31
	Genes involved in electron transport	3.1462e-9	2.4680e-7	51
	RNA_TRANSCRIPTION_REACTOME	4.2302e-9	2.9885e-7	27
Transcription Factor Targets	Targets of ETS1, identified by ChIP-chip in Jurkat T-cells.	5.3746e-299	1.0212e-297	828
	Targets of CREB, identified by ChIP-chip in HEK293T cells in three different time points after forskolin stimulation.	3.3370e-250	3.1701e-249	1201
	Targets of YY1 identified by ChIP-chip.	8.6583e-86	5.4838e-85	383
	Targets of HNF4alpha, identified by ChIP-chip in hepatocytes.	1.0137e-50	4.8149e-50	573
	Targets of NRF1, identified by ChIP-chip in quiescent T98G cells.	5.6202e-34	2.1357e-33	289
	Targets of estrogen receptor alpha, identified by ChIP-DSL in MCF-7 cells.	1.2541e-19	3.9713e-19	191
	Genes that are bound by both E2F4 and p130 in three different growth arrest conditions, identified by ChIP-chip in T98G and U2OS cells under growth arrest.	5.5259e-19	1.4999e-18	100
	Targets of Nanog, identifed by ChIP-chip in embryonic stem cells.	1.5190e-14	3.6077e-14	250
	Genes whose expression peaks periodically in the G1/S cell cycle phase.	2.6230e-10	5.5374e-10	100
	Genes bound by one of the five NF-kB subunits in U937 cells before or 1 hour after lipopolysaccharide stimulation.	1.0425e-6	1.9808e-6	98

Supplementary Table 42: GREAT enrichments of GABP using the *basal plus extension* association rule with a maximum a basal regulatory region extending 5 kb upstream and 1 kb downstream of the transcription start site and extension of 50 kb. Shown are the top ten binomial enriched terms at a false discovery rate of 0.05 with a fold enrichment of at least two that are also significant by the hypergeometric test, using the highest-scoring GABP peaks anywhere in the genome (QuEST score > 1; n = 3,585).

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Molecular Function	nucleic acid binding	0	0	4.0909	5.0778e-43	1.4722	943
	protein binding	0	0	2.9122	1.6051e-9	1.1193	1708
	binding	0	0	2.7646	2.5905e-8	1.0636	2625
	catalytic activity	3.1573e-269	2.2101e-266	2.8910	5.6031e-5	1.1178	1143
	DNA binding	2.4794e-190	1.3885e-187	3.6411	4.3195e-11	1.2960	596
	transition metal ion binding	1.5612e-174	7.2855e-172	3.2149	5.8631e-7	1.2094	662
	zinc ion binding	1.0497e-169	4.1989e-167	3.4600	1.3524e-9	1.2773	581
	nucleotide binding	4.5295e-126	1.1530e-123	3.1341	3.3032e-5	1.2137	501
	RNA binding	2.7196e-125	6.3456e-123	5.4805	5.4821e-30	1.9679	269
	purine nucleotide binding	1.5810e-97	3.1620e-95	3.0421	0.0047	1.1791	408
GO Biological Process	biopolymer metabolic process	0	0	4.0340	2.1096e-92	1.5065	1526
	cellular biopolymer metabolic process	0	0	4.0408	2.7142e-92	1.5076	1518
	cellular macromolecule metabolic process	0	0	4.0017	4.1312e-89	1.4914	1530
	macromolecule metabolic process	0	0	3.9819	2.9677e-88	1.4852	1538
	gene expression	0	0	5.0169	1.3366e-84	1.7889	871
	nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	0	0	4.5240	1.1987e-73	1.6573	962
	cellular metabolic process	0	0	3.6113	6.9341e-70	1.3594	1772
	primary metabolic process	0	0	3.5896	7.1651e-63	1.3483	1712
	metabolic process	0	0	3.4776	3.1213e-59	1.3081	1853
	cellular process	0	0	2.8221	6.3702e-8	1.0675	2438
GO Cellular Component	intracellular membrane-bounded organelle	0	0	3.5994	3.8315e-100	1.3616	2155
	membrane-bounded organelle	0	0	3.5979	3.2193e-100	1.3610	2155
	intracellular	0	0	3.2754	1.7053e-99	1.2473	2742
	intracellular part	0	0	3.2986	2.1265e-93	1.2533	2645
	intracellular organelle	0	0	3.4639	3.6595e-89	1.3034	2311
	organelle	0	0	3.4624	5.7550e-89	1.3028	2311
	nucleus	0	0	3.6242	2.3436e-70	1.4324	1471
	intracellular organelle part	0	0	3.9100	6.2587e-47	1.4141	1150
	organelle part	0	0	3.9036	1.0881e-46	1.4116	1153
	cytoplasmic part	0	0	3.5096	2.8658e-30	1.3126	1154
Mouse Phenotype	abnormal cell content/ morphology	4.7188e-28	1.4357e-25	4.3937	0.0009	1.7351	77
	abnormal nucleus morphology	7.9129e-17	7.6241e-15	5.4661	0.0216	2.0560	34
	abnormal inner cell mass	1.2982e-15	1.1201e-13	5.6211	0.0253	2.0858	31
Pathway Commons	Transcription	1.0444e-144	1.3086e-141	9.3182	6.5253e-57	2.9587	205
	Gene Expression	1.3150e-143	8.2384e-141	8.8684	7.0113e-57	2.8798	214
	RNA Polymerase II Transcription Initiation And Promoter Clearance	1.4482e-130	6.0485e-128	9.6388	1.7959e-51	2.9919	181
	Formation and Maturation of mRNA Transcript	1.4482e-130	6.0485e-128	9.6388	1.7959e-51	2.9919	181
	RNA Polymerase II Promoter Escape	1.4482e-130	6.0485e-128	9.6388	1.7959e-51	2.9919	181
	RNA Polymerase II Transcription Initiation	1.4482e-130	6.0485e-128	9.6388	1.7959e-51	2.9919	181
	RNA Polymerase II Transcription Pre-Initiation	1.4482e-130	6.0485e-128	9.6388	1.7959e-51	2.9919	181
	RNA Polymerase II Transcription	1.4482e-130	6.0485e-128	9.6388	1.7959e-51	2.9919	181
	mRNA Capping	2.4184e-126	3.3670e-124	9.8009	1.6931e-51	3.0448	174
	RNA Pol II CTD phosphorylation and interaction with CE	2.4184e-126	3.3670e-124	9.8009	1.6931e-51	3.0448	174
MSigDB Pathway	PGC related genes	1.1742e-60	8.2902e-58	4.8612	2.4445e-9	1.6691	145
	Mitochondrial genes	2.8360e-56	1.0011e-53	4.5873	1.5542e-8	1.6139	150
	Mitochondrial genes	1.2315e-55	2.8982e-53	4.6481	1.1113e-7	1.5962	141
	RIBOSOMAL_PROTEINS	3.2406e-43	5.7196e-41	9.6198	1.2654e-16	2.9506	63
	MRNA_PROCESSING_REACTOME	1.0950e-39	1.5462e-37	7.5078	3.8894e-13	2.5910	64
	Genes involved in ribosome	4.4364e-39	5.2202e-37	11.7343	1.0293e-16	3.4123	50
	Genes involved in mRNA splicing	3.1605e-29	3.1876e-27	9.8927	6.6155e-11	3.1564	37
	Genes involved in oxidative phosphorylation	1.2418e-26	1.0959e-24	6.6753	7.4546e-6	2.0534	49
	Human CD34 enriched transcription factors	2.3174e-25	1.8179e-23	4.2628	0.0078	1.5287	64
	Genes involved in electron transport	6.3985e-24	4.5173e-22	6.7021	1.6861e-5	2.0812	44
Transcription Factor Targets	Targets of ETS1, identified by ChIP-chip in Jurkat T-cells.	0	0	9.6392	6.4083e-241	3.0959	731
	Targets of CREB, identified by ChIP-chip in HEK293T cells in three different time points after forskolin stimulation.	0	0	6.1625	1.2602e-185	2.2154	1030
	Targets of YY1 identified by ChIP-chip.	1.0698e-211	6.7755e-211	7.7766	5.4823e-64	2.4032	329
	Targets of HNF4alpha, identified by ChIP-chip in hepatocytes.	1.1979e-177	5.6902e-177	4.4366	1.1713e-32	1.6285	481
	Targets of NRF1, identified by ChIP-chip in quiescent T98G cells.	1.3109e-94	4.9813e-94	4.8868	3.9021e-18	1.7035	230
	Targets of Nanog, identifed by ChIP-chip in embryonic stem cells.	2.4853e-57	7.8702e-57	3.6112	5.9997e-7	1.3807	200
	Targets of estrogen receptor alpha, identified by ChIP-DSL in MCF-7 cells.	7.6028e-48	2.0636e-47	4.0712	4.5207e-10	1.6066	151
	Genes that are bound by both E2F4 and p130 in three different growth arrest conditions, identified by ChIP-chip in T98G and U2OS cells under growth arrest.	6.1638e-45	1.4639e-44	6.2143	1.1301e-14	2.2225	87
	Targets of Sox2, identifed by ChIP-chip in embryonic stem cells.	5.5284e-38	1.1671e-37	3.3727	0.0021	1.2508	144
	Genes whose expression peaks periodically in the G1/S cell cycle phase.	2.5639e-30	4.8714e-30	4.2710	9.8875e-6	1.5924	80

Supplementary Table 43: GREAT enrichments of GABP using the *two nearest genes* association rule with a maximum extension of 1 Mb. Shown are the top ten binomial enriched terms at a false discovery rate of 0.05 with a fold enrichment of at least two that are also significant by the hypergeometric test, using the highest-scoring GABP peaks anywhere in the genome (QuEST score > 1; n = 3,585).

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Molecular Function	RNA binding	2.8477e-39	2.6579e-36	2.0040	2.0193e-12	1.4125	327
	structural constituent of ribosome	1.8690e-37	1.3083e-34	3.9493	5.6407e-14	1.9725	100
	translation regulator activity	5.4800e-17	1.1803e-14	3.0382	3.0414e-5	1.6873	67
	translation factor activity, nucleic acid binding	1.0638e-14	1.8616e-12	2.9923	4.0063e-5	1.7404	58
	ATP-dependent helicase activity	3.3682e-11	3.4929e-9	2.4342	0.0056	1.5846	50
	helicase activity	1.7966e-10	1.7347e-8	2.0599	0.0021	1.5072	70
	damaged DNA binding	2.4357e-10	2.2000e-8	3.2689	0.0116	1.7949	28
	translation initiation factor activity	3.6057e-10	3.1549e-8	3.2237	0.0232	1.6621	33
	aminoacyl-tRNA ligase activity	1.3738e-8	1.0123e-6	3.0756	0.0066	1.8004	30
	ribosome binding	4.5980e-8	3.1401e-6	6.7343	0.0198	2.5642	10
GO Biological Process	RNA metabolic process	4.2382e-65	2.2102e-62	2.2094	1.2631e-28	1.5457	451
	translation	1.7417e-55	6.0552e-53	3.3977	1.2160e-21	1.8274	184
	RNA processing	9.3906e-55	3.0607e-52	2.5397	3.4616e-23	1.6383	291
	RNA splicing	7.9247e-36	1.7968e-33	2.6997	8.6801e-14	1.6719	163
	translational elongation	1.9584e-35	4.0853e-33	5.0046	5.0700e-11	2.0147	70
	nuclear mRNA splicing, via spliceosome	5.6968e-34	1.1003e-31	3.8015	5.9817e-10	1.7845	93
	ncRNA metabolic process	1.1302e-30	1.9013e-28	3.1360	3.2646e-11	1.7181	120
	mRNA processing	1.4702e-29	2.2551e-27	2.3674	5.8427e-13	1.6211	173
	mRNA metabolic process	2.8978e-29	4.0843e-27	2.2480	9.8524e-14	1.6022	192
	ribonucleoprotein complex biogenesis	1.5300e-28	1.9948e-26	3.0379	1.7360e-13	1.8279	116
GO Cellular Component	ribonucleoprotein complex	2.3452e-85	2.1732e-83	3.4787	3.6903e-34	1.8422	273
	ribosome	1.0257e-43	4.2773e-42	3.8010	5.8068e-16	1.8604	124
	spliceosome	1.6355e-32	5.2462e-31	3.6619	8.6851e-14	1.9431	93
	ribosomal subunit	5.3162e-29	1.6421e-27	4.0207	4.2256e-11	1.9408	75
	cytosolic part	5.0204e-24	1.4954e-22	3.3189	4.4737e-6	1.6273	75
	cytosolic ribosome	1.0812e-21	2.8179e-20	4.0959	3.1202e-8	1.9821	52
	large ribosomal subunit	3.0685e-20	7.3118e-19	4.5569	7.7293e-9	2.1938	42
	mitochondrial envelope	2.7070e-19	6.2712e-18	2.0567	0.0017	1.2655	166
	organelle inner membrane	3.9079e-19	8.8085e-18	2.2564	0.0001	1.3648	135
	mitochondrial membrane	1.9503e-18	4.1705e-17	2.0527	0.0039	1.2545	157
Pathway Commons	Gene Expression	3.9265e-89	4.9200e-86	4.0080	3.0220e-31	1.8910	238
	Transcription	5.9934e-87	3.7549e-84	4.1185	1.5935e-30	1.9088	224
	Elongation of Intron-Containing Transcripts and co-transcriptional mRNA splicing	2.9275e-86	1.2227e-83	4.8261	1.1830e-27	1.9463	187
	Elongation and Processing of Capped Transcripts	2.9275e-86	1.2227e-83	4.8261	1.1830e-27	1.9463	187
	mRNA Processing	3.7889e-86	9.4949e-84	4.7393	6.0385e-28	1.9559	190
	RNA Polymerase II Transcription Initiation And Promoter Clearance	1.5086e-85	3.1504e-83	4.5575	9.1246e-28	1.9227	197
	Formation and Maturation of mRNA Transcript	1.5086e-85	3.1504e-83	4.5575	9.1246e-28	1.9227	197
	RNA Polymerase II Promoter Escape	1.5086e-85	3.1504e-83	4.5575	9.1246e-28	1.9227	197
	RNA Polymerase II Transcription Initiation	1.5086e-85	3.1504e-83	4.5575	9.1246e-28	1.9227	197
	RNA Polymerase II Transcription Pre-Initiation	1.5086e-85	3.1504e-83	4.5575	9.1246e-28	1.9227	197
MSigDB Pathway	PGC related genes	9.6755e-28	6.8309e-25	2.1772	2.7293e-7	1.4069	207
	Mitochondrial genes	1.3383e-23	4.7240e-21	2.0340	5.5695e-6	1.3468	212
	RIBOSOMAL_PROTEINS	2.0663e-23	4.8627e-21	3.5812	6.0776e-9	1.9357	70
	Mitochondrial genes	2.2399e-22	3.9533e-20	2.0127	2.2529e-5	1.3368	200
	Genes involved in ribosome	1.2110e-20	1.7100e-18	4.0108	2.1374e-7	2.0147	50
	Genes involved in mRNA splicing	5.0332e-20	5.9224e-18	4.8383	6.2215e-6	2.0147	40
	MRNA_PROCESSING_REACTOME	1.1964e-16	1.2067e-14	2.6343	2.4596e-7	1.7689	74
	Genes involved in oxidative phosphorylation	1.2690e-12	1.1199e-10	2.6232	0.0002	1.6082	65
	Genes involved in electron transport	3.6622e-11	2.8728e-9	2.6395	0.0009	1.5918	57
	Genes involved in mRNA processing	8.5381e-10	6.0279e-8	3.3697	0.0024	1.8395	30
Transcription Factor Targets	Targets of ETS1, identified by ChIP-chip in Jurkat T-cells.	3.4132e-270	6.4851e-269	3.4879	2.1488e-143	2.0204	808
	Targets of CREB, identified by ChIP-chip in HEK293T cells in three different time points after forskolin stimulation.	3.5031e-183	3.3279e-182	2.1772	5.3658e-94	1.5646	1232
	Targets of YY1 identified by ChIP-chip.	1.1973e-103	7.5828e-103	3.1056	9.8671e-32	1.6303	378
	Targets of NRF1, identified by ChIP-chip in quiescent T98G cells.	1.2152e-39	4.6177e-39	2.0980	2.7527e-10	1.3425	307
	Genes that are bound by both E2F4 and p130 in three different growth arrest condtions, identified by ChIP-chip in T98G and U2OS cells under growth arrest.	3.7910e-12	1.0290e-11	2.0126	0.0002	1.3877	92

Supplementary Table 44: GREAT enrichments of GABP using the *single nearest gene* association rule with a maximum extension of 1 Mb. Shown are the top ten binomial enriched terms at a false discovery rate of 0.05 with a fold enrichment of at least two that are also significant by the hypergeometric test, using the highest-scoring GABP peaks anywhere in the genome (QuEST score > 1; n = 3,585).

Ontology	Term	Binomial Results			Hypergeometric Results		
	Name	Raw P-Value	FDR Q-Val	Fold Enrichment	FDR Q-Val	Fold Enrichment	Observed Gene Hits
GO Molecular Function	structural constituent of ribosome	2.1218e-36	2.9705e-33	5.2719	1.5868e-25	3.2101	87
	RNA binding	2.5194e-34	2.3514e-31	2.2915	1.0909e-26	1.9797	245
	translation regulator activity	3.6836e-16	1.7190e-13	3.8984	4.4586e-7	2.3084	49
	translation factor activity, nucleic acid binding	2.0086e-14	8.0345e-12	3.8774	2.5104e-7	2.4698	44
	nuclease activity	1.4380e-12	5.0329e-10	3.2284	0.0043	1.7229	48
	unfolded protein binding	3.7230e-11	1.0424e-8	3.4905	0.0012	1.9657	38
	translation initiation factor activity	1.9623e-10	4.9948e-8	4.4195	0.0012	2.3556	25
	endonuclease activity	3.9776e-10	9.2810e-8	3.8104	0.0145	1.8844	30
	ribonucleoprotein binding	1.5622e-9	3.3647e-7	8.8610	6.8637e-5	3.8879	14
	RNA splicing factor activity, transesterification mechanism	1.9253e-9	3.3892e-7	7.8946	7.7540e-5	3.5176	16
GO Biological Process	gene expression	3.3917e-106	3.5375e-103	2.1177	2.3665e-83	1.8397	811
	RNA metabolic process	1.0834e-66	5.6497e-64	2.7423	2.7702e-52	2.2119	345
	RNA processing	3.1977e-58	1.1911e-55	3.2788	5.6219e-43	2.4434	232
	translation	2.8058e-50	9.1451e-48	4.3120	9.0183e-35	2.7683	149
	RNA splicing	8.1540e-39	2.3624e-36	3.4811	8.4019e-31	2.6862	140
	nuclear mRNA splicing, via spliceosome	4.1184e-35	1.1304e-32	5.0670	2.3206e-20	2.9074	81
	protein transport	3.1541e-32	8.2244e-30	2.2169	6.5829e-21	1.8253	238
	establishment of protein localization	1.8205e-31	4.5208e-29	2.1916	1.5340e-20	1.8147	238
	mRNA processing	1.1170e-30	2.6478e-28	2.9530	1.9632e-24	2.4191	138
	ribonucleoprotein complex biogenesis	1.5682e-30	3.5557e-28	4.0674	1.0442e-23	2.8593	97
GO Cellular Component	ribonucleoprotein complex	2.6615e-84	3.1710e-82	4.6002	4.0476e-58	2.8654	227
	nuclear part	4.9215e-58	3.4204e-56	2.1425	8.4057e-45	1.8410	463
	intracellular organelle lumen	2.9592e-47	1.7629e-45	2.0643	7.3359e-38	1.8000	422
	membrane-enclosed lumen	8.5657e-46	4.7625e-44	2.0099	1.0067e-36	1.7642	435
	ribosome	1.2868e-40	6.3128e-39	4.8366	1.6749e-29	3.0031	107
	nuclear lumen	6.7757e-40	3.1394e-38	2.1144	1.0124e-32	1.8539	338
	nucleoplasm	7.3763e-38	3.2378e-36	2.3100	2.4876e-30	1.9632	269
	mitochondrion	4.6332e-36	1.8400e-34	2.1445	4.1703e-16	1.5682	299
	spliceosome	1.8773e-33	7.1168e-32	4.9201	8.2011e-23	3.0877	79
	cytosol	9.3974e-32	3.4076e-30	2.0378	1.4327e-18	1.6387	291
Pathway Commons	Transcription	2.8978e-90	3.6310e-87	5.8057	8.5929e-56	3.0925	194
	Gene Expression	3.4127e-89	2.1381e-86	5.5758	1.0478e-55	3.0172	203
	Elongation of Intron-Containing Transcripts and co-transcriptional mRNA splicing	3.3726e-88	1.4086e-85	6.8918	2.8220e-50	3.1931	164
	Elongation and Processing of Capped Transcripts	3.3726e-88	1.4086e-85	6.8918	2.8220e-50	3.1931	164
	RNA Polymerase II Transcription Initiation And Promoter Clearance	5.0913e-88	1.2759e-85	6.5059	1.5533e-51	3.1586	173
	Formation and Maturation of mRNA Transcript	5.0913e-88	1.2759e-85	6.5059	1.5533e-51	3.1586	173
	RNA Polymerase II Promoter Escape	5.0913e-88	1.2759e-85	6.5059	1.5533e-51	3.1586	173
	RNA Polymerase II Transcription Initiation	5.0913e-88	1.2759e-85	6.5059	1.5533e-51	3.1586	173
	RNA Polymerase II Transcription Pre-Initiation	5.0913e-88	1.2759e-85	6.5059	1.5533e-51	3.1586	173
	RNA Polymerase II Transcription	5.0913e-88	1.2759e-85	6.5059	1.5533e-51	3.1586	173
MSigDB Pathway	RIBOSOMAL_PROTEINS	1.7957e-21	1.2677e-18	4.5398	1.0374e-15	3.0521	59
	PGC related genes	4.8434e-21	1.7097e-18	2.3807	2.3458e-8	1.6783	132
	Genes involved in ribosome	1.1448e-20	2.6940e-18	5.3257	1.1287e-16	3.6181	48
	Genes involved in mRNA splicing	1.2956e-20	2.2867e-18	7.1205	1.5591e-10	3.2978	35
	Mitochondrial genes	5.6209e-18	7.9368e-16	2.2070	3.3073e-7	1.6043	135
	Mitochondrial genes	2.7105e-17	3.1893e-15	2.1924	1.6410e-6	1.5879	127
	MRNA_PROCESSING_REACTOME	3.7798e-17	3.8122e-15	3.3351	2.7966e-13	2.7276	61
	Genes involved in mRNA processing	4.2519e-12	3.7523e-10	5.2418	7.5923e-6	2.8676	25
	RNA_TRANSCRIPTION_REACTOME	5.1228e-10	4.0185e-8	5.9776	0.0023	2.5063	19
	Genes involved in oxidative phosphorylation	6.3831e-10	4.5065e-8	2.9198	0.0004	1.9439	42
Transcription Factor Targets	Targets of ETS1, identified by ChIP-chip in Jurkat T-cells.	1.3317e-297	2.5302e-296	4.8992	1.5743e-226	3.1949	683
	Targets of CREB, identified by ChIP-chip in HEK293T cells in three different time points after forskolin stimulation.	1.1324e-205	1.0758e-204	2.8108	2.2906e-164	2.2237	936
	Targets of YY1 identified by ChIP-chip.	3.6189e-106	2.2920e-105	4.1515	2.8140e-61	2.4769	307
	Targets of NRF1, identified by ChIP-chip in quiescent T98G cells.	2.8400e-33	1.0792e-32	2.4186	2.5183e-15	1.6852	206
	Genes that are bound by both E2F4 and p130 in three different growth arrest condtions, identified by ChIP-chip in T98G and U2OS cells under growth arrest.	9.5204e-12	2.5841e-11	2.3270	9.6987e-10	2.0316	72

Supplementary Table 45: GREAT analysis table locations.

Dataset	Test type					
	Author/DAVID Proximal promoter	GREAT *basal+extension** Up to 1,000 kb	"Gene-based GREAT" Proximal promoter 2 kb	GREAT *basal+extension** Up to 50 kb	GREAT *two nearest genes* Up to 1,000 kb	GREAT *single nearest gene* Up to 1,000 kb
SRF	Table 2	Table 3	Sup. Table 6	Sup. Table 7	Sup. Table 8	Sup. Table 9
p300 Limb	Sup. Table 10a	Sup. Table 10b	Sup. Table 11	Sup. Table 12	Sup. Table 13	Sup. Table 14
p300 Forebrain	Sup. Table 15a	Sup. Table 15b	Sup. Table 16	Sup. Table 17	Sup. Table 18	Sup. Table 19
p300 Midbrain	Sup. Table 20a	Sup. Table 20b	Sup. Table 21	Sup. Table 22	Sup. Table 23	Sup. Table 24
p300 mESC	Sup. Table 25a	Sup. Table 25b	Sup. Table 26	Sup. Table 27	Sup. Table 28	Sup. Table 29
Stat3	Sup. Table 30a	Sup. Table 30b	Sup. Table 31	Sup. Table 32	Sup. Table 33	Sup. Table 34
NRSF	Sup. Table 35a	Sup. Table 35b	Sup. Table 36	Sup. Table 37	Sup. Table 38	Sup. Table 39
GABP	Sup. Table 40a	Sup. Table 40b	Sup. Table 41	Sup. Table 42	Sup. Table 43	Sup. Table 44

* The *basal plus extension* rules both define basal regulatory domains to extend 5 kb upstream and 1 kb downstream from the transcription start site of each gene.

Supplementary Table 46: Analysis of SRF GO term enrichments. **(a)** Terms significant by both the binomial and hypergeometric tests highlight many genes involved in the process with many genomic regions implicating the genes as well. Skews between the fraction of genes annotated with the term and the fraction of the genome that maps to one or more genes annotated with the term are generally modest. **(b)** Terms significant by the hypergeometric test but not the binomial test arise either due to large differences between the fraction of genes annotated with the term and the fraction of the genome that maps to one or more genes annotated with the term or the association of a single genomic region to multiple genes annotated with the term. **(c)** Terms significant by the binomial test but not the hypergeometric test arise when many genomic regions cluster near one or few genes annotated with the term, and indicate gene-specific enrichments rather than broad term-based enrichment.

a

Terms significant by both binomial and hypergeometric tests (B ∩ H, listed in Table 1b)

GO ID	Description	Genes Hit	SRF Peaks	Fraction of Genes	Fraction of Genome
GO:0015629	actin cytoskeleton	30	36	0.013185	0.021250
GO:0030863	cortical cytoskeleton	11	7	0.001859	0.003351
GO:0003779	actin binding	31	37	0.017483	0.032754

b

Terms significant by the hypergeometric test but not the binomial test (H\B)

GO ID	Description	Genes Hit	SRF Peaks	Fraction of Genes	Fraction of Genome
GO:0010604	positive regulation of macromolecule metabolic process	53	61	0.033862	0.073249
GO:0005634	nucleus	284	279	0.284951	0.419860
GO:0009893	positive regulation of metabolic process	54	62	0.036011	0.077215
GO:0005515	protein binding	397	351	0.423419	0.604715
GO:0019899	enzyme binding	33	37	0.018702	0.037255

c

Terms significant by the binomial test but not the hypergeometric test (B\H)

GO ID	Description	Genes Hit	SRF Peaks	Fraction of Genes	Fraction of Genome
GO:0032796	uropod organization	2	5	0.000116	0.000100
GO:0035267	NuA4 histone acetyltransferase complex	2	6	0.000348	0.000231
GO:0043189	H4/H2A histone acetyltransferase complex	2	6	0.000407	0.000262
GO:0043534	blood vessel endothelial cell migration	2	6	0.000290	0.000309
GO:0000212	meiotic spindle organization	1	4	0.000116	0.000092