ACTIVE MOTIF®

LiGHTSW!TCH
Luciferase Assay System

TECHNICAL NOTE:

High-throughput functional analysis of 528 STAT binding sites in the human genome

Jie Wang[1], Katherine Harris[2], Troy Whitfield[1], Patrick J. Collins[2], Shelley Force Aldred[2], Nathan D. Trinklein[2], and Zhiping Weng[1]

[1]Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School
[2]Active Motif, Carlsbad, CA

ABSTRACT

The ENCODE consortium has generated binding maps for the STAT1 and STAT2 transcription factors during interferon stimulation using the ChIP-seq method. In this study we wished to functionally characterize and map the functional motifs of all of the STAT1 and STAT2 binding sites in the human genome identified by ChIP-seq. To accomplish this we used 528 reporter constructs representing all of the STAT binding sites. We identified nearly 200 DNA elements that were inducible by either interferon alpha or gamma. Furthermore we mutagenized nearly 300 5-10 bp motifs to map at high resolution the functional motifs involved in STAT signaling. The results show that there are STAT-bound elements that respond preferentially to either IFNa or IFNg. The results also show that the strongest interferon responses come from proximal promoter regions that are closest to transcription start sites of genes. Lastly, we used high-throughput site-directed mutagenesis to map the functional motifs with base pair resolution.

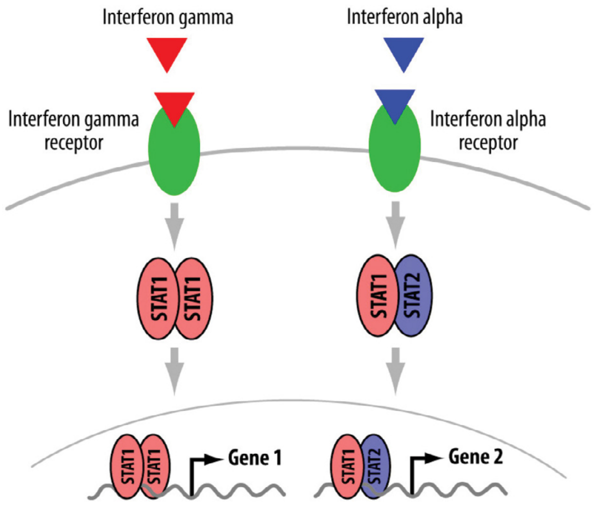


Figure 1: Model of STAT signaling.
Interferon gamma binds to its receptor and stimulates STAT1 homodimer formation, nuclear translocation, and binding to DNA elements in the genome. Interferon alpha binds to its receptor and stimulates STAT1/STAT2 heterodimer formation, nuclear translocation, and binding to DNA elements in the genome.

INTRODUCTION

To fully understand a genetic regulatory network, it is necessary to map and functionally characterize all of the DNA regulatory elements involved the network. Experimental approaches exist to map the genome-wide binding sites of transcription factors (i.e. ChIP-seq), however the functional consequences of these binding events still need to be experimentally determined. In this project, we chose to characterize the STAT-mediated regulatory network using a comprehensive experimental approach that measures the functional effects of STAT binding events in the human genome.

The STAT family of transcription factors are important master regulators of the immune response and mediators of cytokine signaling. There are 7 different STAT family members in the human genome, and in this study we focus on STAT1 and STAT2. Different interferons signal through different receptor proteins which in turn activate STAT proteins by specific phosphorylation events. As shown in Figure 1, in response to interferon alpha (IFNa) stimulation, STAT1 and STAT2 form a heterodimer that translocates to the nucleus and binds to its genomic targets. Similarly, in response to interferon gamma (IFNg) stimulation, STAT1 forms a homodimer that translocates to the nucleus and binds to its genomic targets.

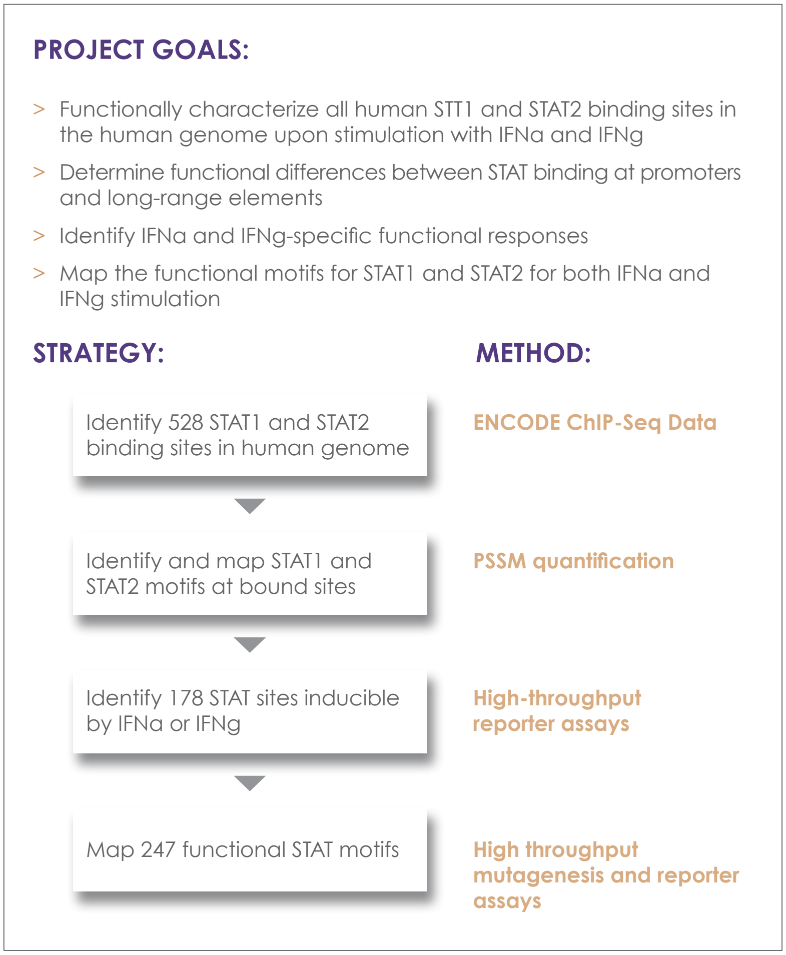
To better understand the response of the STAT1/2 regulatory network to interferon stimulation, we used a combination of bioinformatics and functional genomics approaches as outlined in Figure 2. First, we used ChIP-seq data generated by the ENCODE consortium to identify sites at which STAT1 or STAT2 is bound in the presence of IFNa or IFNg. We then used high-throughput cell-based reporter assays and pRT-PCR to determine which binding events were associated with functional changes in gene regulation. Finally, after identifying predicted STAT1/2 motifs in the bound and functional elements, we used high-throughput mutagenesis and reporter assays to map the functional motifs with base pair resolution.

METHODS

Initial Screen

The full datasets for STAT1 and STAT2 binding were downloaded from the ENCODE data portal at UCSC (genome.ucsc.edu). A total of 528 significant binding sites were identified for STAT1 and STAT2. Of the 528 total binding sites, 216 were located in proximal promoter regions (<2kb upstream from a TSS in the genome) and 312 were located in distal regions (>2kb from TSS in the genome). Figure 3 outlines the methods for using high-throughput reporter assays to functionally characterize these binding sites. For the proximal binding sites, 1- 3 kb promoter fragments were chosen from the LightSwitch GoClone Promoter Collection (SwitchGear Genomics). For the distal binding sites, 0.5-3kb fragments were cloned into the pLightSwitch Long Range Element Reporter Vector upstream of a 200 bp basal TK promoter. All of the pLightSwitch vectors utilize the RenSP luciferase reporter gene, which is an optimized luciferase gene designed specifically for use in induction and repression experiments. These 528 reporter vectors were then arrayed into a custom 96-well format for subsequent assays.

As shown in Figure 3, the panel of 528 reporter vectors were then individually transfected into K562 cells in 96-well format. A panel of 8 promoter controls was also used to normalize signals between plates, replicates, and treatments. Transfection complexes were formed by incubating 100 ng of each individual promoter construct with Lipofectamine LTX transfection reagent and Opti-MEM media in a total volume of 20 uL and incubated for 30 minutes. Transfection complexes were mixed with K562 cells such that 25,000 K562 cells were seeded in a volume of 100 uL in each well of a 96-well white tissue culture treated plate. Six replicate transfection wells for each promoter construct were performed representing duplicate assays in 3 different conditions: 1) no treatment, 2) 500 U/mL IFNa, and 3) 100 ng/mL IFNg.


PROJECT GOALS:

- > Functionally characterize all human STT1 and STAT2 binding sites in the human genome upon stimulation with IFNa and IFNg
- > Determine functional differences between STAT binding at promoters and long-range elements
- > Identify IFNa and IFNg-specific functional responses
- > Map the functional motifs for STAT1 and STAT2 for both IFNa and IFNg stimulation

STRATEGY:	**METHOD:**
Identify 528 STAT1 and STAT2 binding sites in human genome	ENCODE ChIP-Seq Data
Identify and map STAT1 and STAT2 motifs at bound sites	PSSM quantification
Identify 178 STAT sites inducible by IFNa or IFNg	High-throughput reporter assays
Map 247 functional STAT motifs	High throughput mutagenesis and reporter assays

Figure 2: Project goals and methods.
The overall goals of the project are summarized in the figure above. The strategies and associated methods are also summarized in the figure.

After incubating for 16 hours, the transfected cells were treated with IFNa, IFNg, or vehicle only for 8 hours. Plates were then frozen overnight at -80C.

To read the luminescent signal, plates were thawed for 30 minutes at room temperature. Then 100 uL of LightSwitch Assay Reagent (LS100, SwitchGear Genomics) was added and incubated for 30 minutes at room temperature. Then luminescence was read for 2 seconds per well on a 96-well compatible plate luminometer (Molecular Devices SpectraMax L).

The inducible activity of each promoter was measured by taking the average treated activity divided by the average untreated activity.

qRT-PCR Transcript Level Measurements

K562 cells were grown in 6-well format (1e6 cells/well). A total of 12 wells were treated as follows: 2 wells received IFNa stimulation (500 U/mL), 2 wells received IFNg stimulation (100 ng/mL), and 2 wells received no treatment. Total RNA was purified from each well using the RNeasy Kit (Qiagen). First strand cDNA synthesis was performed using Superscript III (Invitrogen). qPCR primers were designed to 96 total transcripts (amplicon sizes ranged from 60-100 bp). qPCR was performed using iQ SYBR Green Supermix (BioRad) according to the recommended standard protocol on a BioRad MyIQ real-time PCR thermal cycler. Relative transcript level inductions were calculated as: 2^(average threshold cycle IFN treated – average threshold cycle untreated).

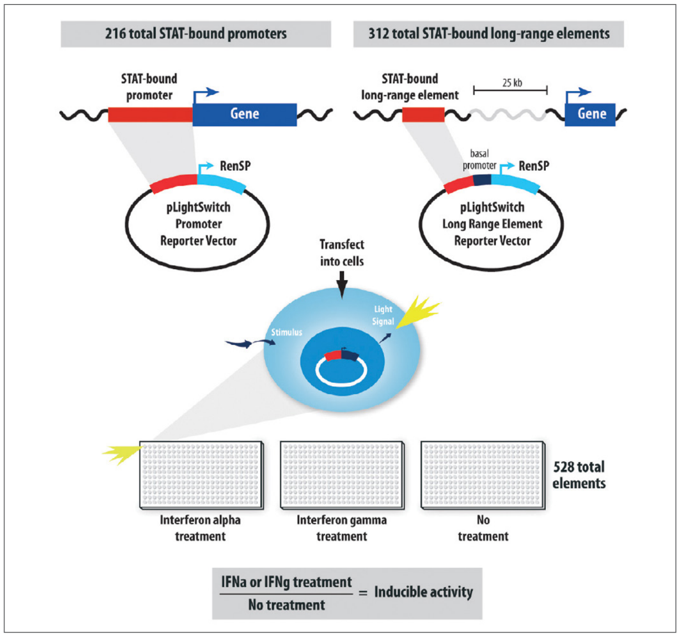

216 total STAT-bound promoters
312 total STAT-bound long-range elements
STAT-bound promoter
Gene
STAT-bound long-range element
25 kb
Gene
RenSP
basal promoter
RenSP
pLightSwitch Promoter Reporter Vector
pLightSwitch Long Range Element Reporter Vector
Transfect into cells
Stimulus
Light Signal
528 total elements
Interferon alpha treatment
Interferon gamma treatment
No treatment
IFNa or IFNg treatment
No treatment
= Inducible activity

Figure 3: High-throughput reporter assay experimental diagram. A total of 216 STAT-bound promoter elements and 312 distal elements were cloned into the pLightSwitch Promoter reporter vector and long-range reporter vector, respectively, as shown above. These vectors were individually transfected in duplicate into living cells. Reporter activity was measured after induction with IFNa, IFNg, and no treatment. The inducible activity of each STAT-bound element was calcuated as the ratio of treated activity divided by untreated activity.

High-throughput Mutagenesis

The results of the primary functional screen showed that 172 fragments were inducible by either IFNa or IFNg. The putative STAT1 or STAT2 binding sites in each of these fragments were identified and mutagenized using site-directed mutagenesis. Individual motifs were mutagenized by changing 5 bases in a 10 base window, and all mutant constructs were sequence validated. A total of 247 mutant fragments were generated and functionally tested along their wild-type partner using the functional assay described above.

RESULTS & DISCUSSION

One of the fundamental challenges in studying transcriptional networks is ascribing functional consequences to transcription factor binding events. One transcription factor may bind to hundreds of places in the genome. Some binding sites are in proximal promoter regions and many binding events are located far away from genes. A single transcription factor may also act either as an activator or repressor depending on the target site, so binding alone does not predict the functional outcome. Furthermore, many binding events have no functional effect and may represent evolutionary relics that have no deleterious effects.

With all of these considerations, a direct functional measure of transcription factor binding events is critical to understanding transcriptional networks. For this project, we wished to functionally characterize all of the known STAT1 and STAT2 binding events in the human genome during interferon stimulation. Furthermore, we wished to see which of these binding sites were associated with IFNa or IFNg-inducible genomic elements. Lastly, we used mutagenesis to map at high resolution the motifs necessary for the interferon response.

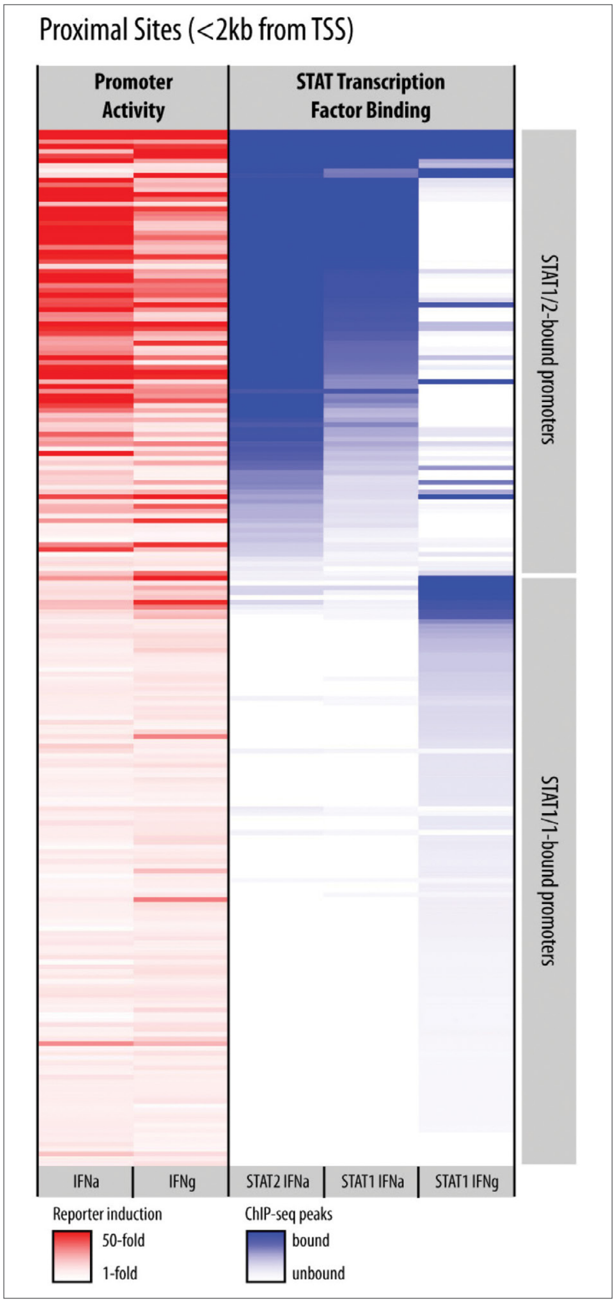


Figure 4: Promoter activity largely agrees with STAT binding. In this heatmap, each row is a single STAT-bound proximal promoter element. The intensity of red in the two left-hand columns indicate the degree of inducible promoter activity as measured by the reporter assay for IFNa and IFNg induction. The intensity of blue in the columns on the right indicate the ChIP-seq binding signal for IFNa and IFNg induction for STAT1 and STAT2. The promoters were manually clustered to approximately group the STAT1/1 bound promoters and the STAT1/2 bound promoters. Overall, the promoters with the strongest binding signal have the highest inducible activity.

Using the experimental approach outlined in Figure 3, we measured the activity of a total of 216 proximal promoter binding sites (within 2kb of a transcription start site) and 312 distal sites using reporter assays. Of the distal sites, 21% (65/312) showed a 2-fold or greater response to either IFNa or IFNg. Of the proximal sites, 35% (75/216) showed a 2-fold or greater response to either IFNa or IFNg. Not only is a larger proportion of the proximal sites inducible, they also are more strongly inducible. The top 5% of IFNa inducible proximal sites had a 28-fold average induction upon IFNa treatment, whereas the top 5% of IFNa inducible distal sites showed only a 6-fold average induction. Taken together, these data suggest that on a whole, binding events that are closer to transcription start sites are more likely to be functional and show stronger activity than binding sites that are located further away.

We also compared whether some inducible elements were more inducible by IFNa or IFNg and how this compared to STAT1 and STAT2 binding status. For the IFNa-inducible proximal sites, 35 of the fragments showed IFNa-inducible activity that was at least 2-fold greater that IFNg-inducible activity. For the IFNg-inducible proximal sites, 11 of the fragments showed IFNg-inducible activity that was at least 2-fold greater that IFNa-inducible activity. Figure 4 summarizes proximal promoter inducible activities and compares them with STAT1 and STAT2 binding status. Overall, there is significant agreement between IFNa inducible activity and STAT1:STAT2 binding under IFNa stimulation, as expected. Likewise, there is also significant agreement between IFNg inducible activity and STAT1: STAT1 binding under IFNg stimulation. However, the heatmap in Figure 4 also highlights the exceptions to the simple model outlined in Figure 1 and shows how the functional activity of these binding sites does not perfectly correlate to the binding status of the STAT factors.

Similarly, Figure 5 illustrates the distal element inducible activities and how they relate to STAT binding. Like with the proximal elements, there is significant agreement between IFNa and IFNg inducible activities and STAT1 and STAT2 binding under IFNa and IFNg induction, respectively. As with the proximal elements, the heatmap shows how the functional activity of these binding sites does not perfectly correlate to the binding signal of the STAT factors.

Comparison with endogenous transcript levels

The overall agreement between the ChIP-seq binding data and the reporter assay functional results provides strong support for the validity of both of these experimental approaches. To compliment these two types of data, we also wished to examine the endogenous transcript levels of STAT

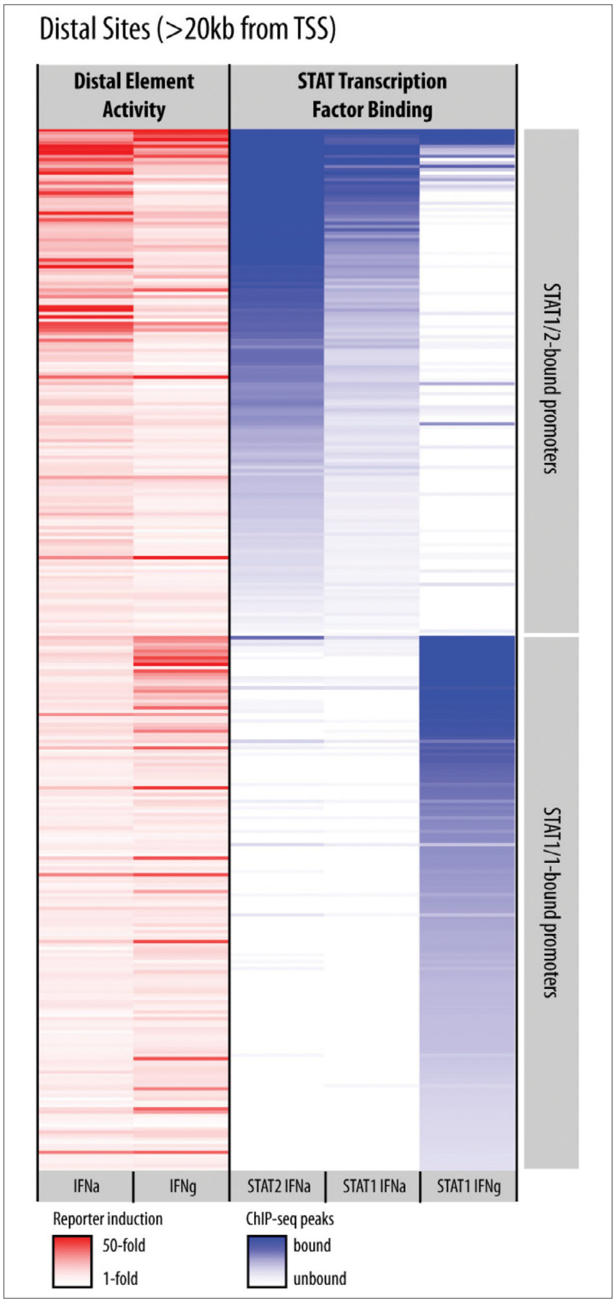


Figure 5: Distal element activity agrees with STAT binding.
In this heatmap, each row is a single STAT-bound distal element. The intensity of red in the two left-hand columns indicate the degree of inducible activity as measured by the reporter assay for IFNa and IFNg induction. The intensity of blue in the columns on the right indicate the ChIP-seq binding signal for IFNa and IFNg induction for STAT1 and STAT2. The distal elements were manually clustered to group the STAT1/1 bound promoters and the STAT1/2 bound promoters. Overall, the promoters with the strongest binding signal have the highest inducible activity.

target genes to see if this third type of data would explain any differences between the binding data and reporter assay results. We used qRT-PCR to measure transcript level changes due to IFNa and IFNg stimulation. We sampled a set of 29 genes with STAT binding sites in the proximal promoter representing a range of inductions and binding scores. Transcript level induction measurement by qRT-PCR was performed as described in the methods. The transcript level induction, promoter activity induction, and STAT binding are all represented in Figure 6 for the 29 genes tested. Overall, there is a very high correlation between IFNa-induced promoter activity and transcript levels (R= 0.64). The promoter activity correlation with transcript level for IFNg stimulation was even higher (R= 0.72). These results demonstrate that the activity of the STAT-bound promoters largely explains the transcript level changes under interferon stimulation even though the promoter region is taken out of its genomic context in the reporter assay. While changes in transcript levels represent the cumulative effect of promoter activity,

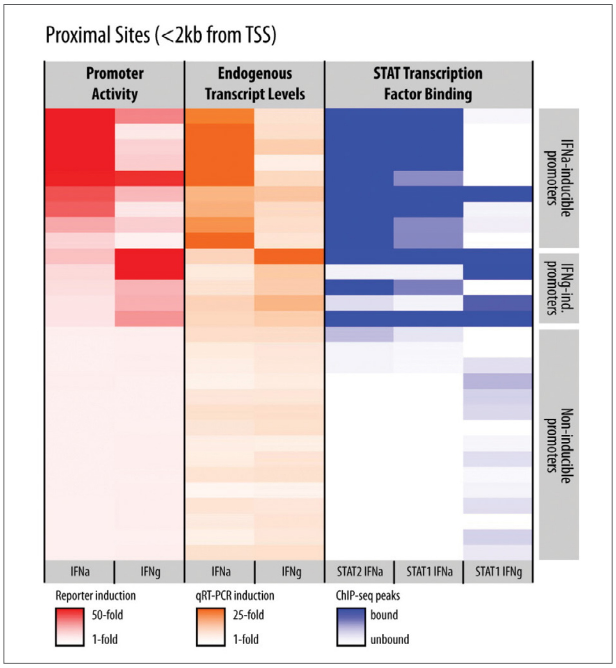


Figure 6: Promoter activity correlates with endogenous transcript levels.
In this heatmap, each row is a STAT-bound promoter. As in Figures 4 and 5, the intensity of red and blue indicate reporter activity and STAT binding, respectively. The intensity of orange in the middle two columns indicate the endogenous transcript level induction as measured by qRT-PCR. For the 29 promoters tested, there is a very strong correlation between induced promoter activity and induced transcript level indicating that much of the STAT regulation happens in the proximal promoter region.

long-range transcriptional elements, and post-transcriptional mechanisms, these results show that the regulation of STAT targets are dominated by the proximal promoter region.

As a further test of these observations we also measured transcript level changes of 30 genes that were located closest to distal STAT binding sites. The comparison of distal element functional activity with transcript level changes is considerably more difficult because the association between distal binding sites and genes is much less certain. Nevertheless, we made the simple assumption that the closest upstream or downstream gene was a potential target of a distal binding site, and there were several occasions where the distal site was in an intron of a gene. Figure 7 summarizes the relationship between transcript level induction, promoter activity induction, and STAT binding for the sample of 30 distal binding sites. There is a much lower correlation between distal element activity and transcript level induction for both IFNa and IFNg inductions (R= ~0.30). This is likely due to the difficulty in predicting distal binding site relationships to genes discussed above. Interestingly, the individual gene labeled 1 highlighted in Figure 7, does show strong agreement, and this distal site occurs in the intron of the transcript measured by qRT-PCR. This suggests that intronic STAT binding sites might be more often functional or at least easier to assign to a single gene. Example 2, highlighted in Figure 7, is also interesting. This gene shows strong transcript induction, strong binding, but weak distal element reporter activity. Interestingly, this gene also has a STAT-bound promoter that has highly inducible reporter activity. This is consistent with the previous observation that STAT-induced transcript induction is driven largely by the proximal promoter region. Lastly, examples 3 and 4 are distal elements with strongly inducible reporter activity and strong STAT binding but very little transcript induction. These may be cases where we did not choose the correct target gene, where taking the distal element out of its genomic context may have removed epigenetic repressors that act in the proper genomic context, or where binding has no functional effect. Interestingly many of the highly inducible distal elements are located in SINE repeat elements indicating that these binding sites may be by products of retro-transposon activity.

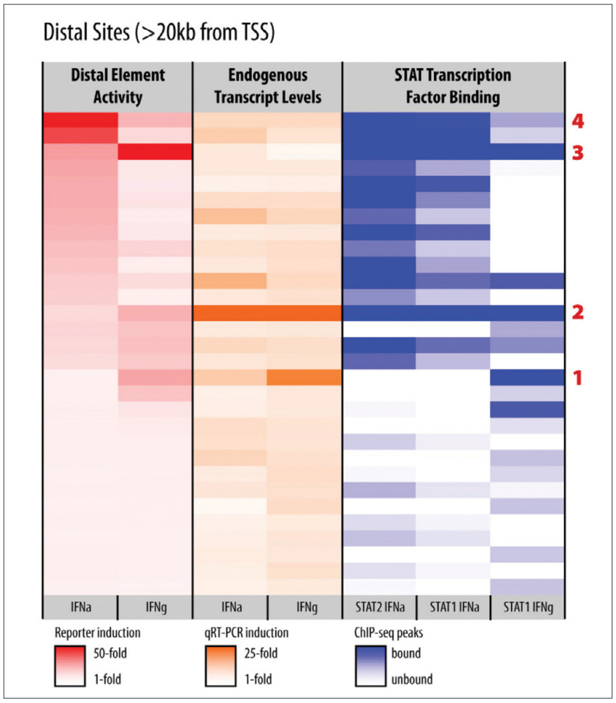


Figure 7: Distal element activity does not correlate as well with endogenous transcript levels.
In this heatmap, each row is a STAT-bound distal element. As in Figures 4 and 5, the intensity of red and blue indicate reporter activity and STAT binding, respectively. The intensity of orange in the middle two columns indicate the endogenous transcript level induction as measured by qRT-PCR. For the 30 distal elements tested, there is a weaker correlation between induced element activity and induced transcript level.

High-resolution mapping of functional motifs

The typical resolution for ChIP-seq binding data is usually a hundred bases or more. We wished to map the function binding sites within each inducible fragment at a resolution of 10 bases or less. To do this we used site-specific mutagenesis to mutate the predicted STAT1 and STAT2 motifs in the inducible fragments that we experimentally characterized. We generated a total of 247 mutant reporter constructs from a set of 172 wild-type proximal and distal binding sites. For some wild-type constructs we made up to 4 different mutant constructs to test 4 different motifs. After building and sequence validating the mutant constructs, we then measured the function of each mutant along-side its wild-type partner using the same high-throughput reporter assay approach described above. We tested each mutant and wild-type partner in triplicate in the same 3 conditions: 1) no treatment, 2) 500 U/mL IFNa, and 3) 100 ng/mL IFNg.

We tested the inducible activity of each fragment by dividing the average of the treated by the average of the non-treated activity. We then calculated the percent change between wild-type inducibility and mutant inducibility. The vast majority The vast majority of mutants showed significant decreases in inducible activity indicating that we were able to map a functional motif in the majority of cases. For example, of the 82 IFNa-inducible elements that we mutagenized, 76 (93%) showed more than a 50% drop in induction compared to wildtype and 43 (52%) showed more than a 75% drop in induction level. However, given that significant inducible activity remained in many of the mutated fragments tested, the mapped functional motifs are not the only contributors to the inducible activity of the fragment. This supports the notion that regulatory elements, especially extended promoter regions, are complex functional units and more than just a collection of independent motifs.

The functional results of our mutation analysis are shown in Figure 8. Mutating either motif in Promoter 1 eliminates virtually all of the inducible activity of the wild-type promoter. In contrast, mutating one of the predicted motifs in Promoter 2 has no functional effect, whereas mutating the other motif removes all inducible activity. Lastly, Promoter 3 is an example of motifs that have intermediate effects. One motif eliminates inducible activity for both IFNa and IFNg. The other motif, however, has no effect on IFNa induction, but decreases IFNg induction by more than 50%.

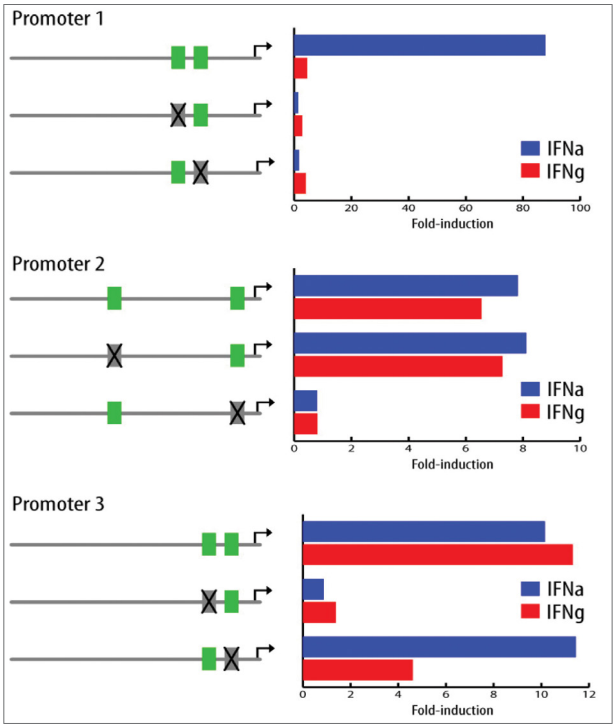


Figure 8: Mutating STAT motifs abolishes inducible activity. The figure above shows 3 examples of the functional effects of mutating STAT motifs in proximal promoter elements. Each promoter is approximately 1 kb in length. The 3 examples each have 2 predicted STAT motifs that are indicated as green boxes. The first construct is the wild-type sequence, and the second and third constructs show each motif mutated separately. The blue and red bars indicate the fold induction of each fragment upon stimulation with IFNa and IFNg, respectively.

CONCLUSIONS

The results of this study demonstrate the value of cell-based assays that measure the functional effects of transcription factor binding sites using the STAT signaling pathway as a model. By using promoter and long-range element reporter assays, we were able to identify the functional binding sites in a set of over 500 binding sites identified by ChIP-seq. We were also able to show which of these binding sites were inducible by IFNa or IFNg, and we were able to provide evidence that binding sites in close proximity to transcription start sites were more often functional and had stronger effects. By comparing the reporter assay data with endogenous transcript levels through qRT-PCR, we were able to show that a very large proportion of STAT-mediated regulation happens from the proximal promoter region. Lastly, we used site-directed mutagenesis to map the functional motifs within these larger regions and dissect the functional behavior of these elements in greater detail.

This two-phase approach, 1) TF target discover (ChIP-seq) followed by 2) functional screening (reporter assays) may also be generalized to study any transcription factor network in the human genome. In this way, a comprehensive functional profile can be generated for any pathway under thousands of different conditions. This approach would be very valuable in the pre-clinical development of new therapeutic compounds that seek to modulate certain biological pathways. The comprehensive pathway profiles can quantitatively identify selective modulators of a pathway and more quickly identify off-target effects.

Supporting on-line material

www.switchgeargenomics.com/STATproject