Published online 8 February 2012 *Nucleic Acids Research, 2012, Vol. 40, No. 10* ***4461–4472***
doi:10.1093/nar/gks056

Genome-wide analyses reveal the extent of opportunistic STAT5 binding that does not yield transcriptional activation of neighboring genes

Bing-Mei Zhu[1,2], Keunsoo Kang[1,3], Ji Hoon Yu[1], Weiping Chen[4], Harold E. Smith[4], Daeyoup Lee[3], Hong-Wei Sun[5], Lai Wei[6] and Lothar Hennighausen[1,7,*]

[1]Laboratory of Genetics and Physiology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD, 20892, USA, [2]Key Laboratory of Acupuncture and Medicine Research of Ministry of Education, Nanjing University of Chinese Medicine, Nanjing, China, 210029, [3]Department of Biological Sciences, Korea Advanced Institute of Science and Technology, Daejeon, 305-701, Republic of Korea, [4]Microarray Core Facility with the Genomic Core Laboratory, National Institute of Diabetes and Digestive and Kidney Diseases, [5]Biodata Mining and Discovery Section, National Institute of Arthritis and Musculoskeletal and Skin Diseases [6]National Eye Institute, National Institutes of Health, Bethesda, MD 20892, USA and [7]National Department of Nanobiomedical Science and WCU Research Center of Nanobiomedical Science, Dankook University, Cheonan, Chungnam 330-714, Republic of Korea

Received November 11, 2011; Revised January 13, 2012; Accepted January 16, 2012

ABSTRACT

Signal Transducers and Activators of Transcription (STAT) 5A/B regulate cytokine-inducible genes upon binding to GAS motifs. It is not known what percentage of genes with GAS motifs bind to and are regulated by STAT5. Moreover, it is not clear whether genome-wide STAT5 binding is modulated by its concentration. To clarify these issues we established genome-wide STAT5 binding upon growth hormone (GH) stimulation of wild-type (WT) mouse embryonic fibroblasts (MEFs) and MEFs overexpressing STAT5A more than 20-fold. Upon GH stimulation, 23 827 and 111 939 STAT5A binding sites were detected in WT and STAT5A overexpressing MEFs, respectively. 13 278 and 71 561 peaks contained at least one GAS motif. 1586 and 8613 binding sites were located within 2.5 kb of promoter sequences, respectively. Stringent filtering revealed 78 genes in which the promoter/upstream region (−10 kb to +0.5 kb) was recognized by STAT5 both in WT and STAT5 overexpressing MEFs and 347 genes that bound STAT5 only in overexpressing cells. Genome-wide expression analyses identified that the majority of STAT5-bound genes was not under GH control. Up to 40% of STAT5-bound genes were not expressed. For the first time we demonstrate the magnitude of opportunistic genomic STAT5 binding that does not translate into transcriptional activation of neighboring genes.

INTRODUCTION

Signal transducers and activators of transcription (STAT) 5A and 5B (referred to as STAT5) are latent transcription factors that are activated by cytokines, such as interleukins, erythropoietin, thrombopoietin, growth hormone (GH) and prolactin, through their respective receptors (1). Mice, from which the *Stat5* genes had been deleted globally or in defined cell populations have shed light onto general and cell-specific functions. These include erythropoiesis, immune functions, liver metabolism, body growth and mammary gland development during pregnancy (2–10). In addition, humans carrying homozygous mutations in the *Stat5B* gene are characterized by growth defects, due to impaired IGF1 signaling and immune defects (11). STAT5 levels vary greatly between tissues and evidence is mounting that distinct levels are needed for different cellular functions (12). In turn, excessive activation of STAT5 has been linked to the progression of solid tumors and leukemia (13,14).

Genes containing GAS motifs (consensus sequence TTCnnnGAA) bound *in vivo* by STAT5 and whose expression is STAT5-dependent are considered direct STAT5 targets. Gene expression studies on cells ectopically overexpressing STAT5 (15) and siRNA-mediated reduction of STAT5 (16) have been employed in the identification of putative target genes. Last, comparative gene expression analyses of wild-type (WT) and

*To whom correspondence should be addressed. Tel: +1 301 496 217; Fax: +1 301 480 7312; Email: lotharh@mail.nih.gov

STAT5-null cells and tissues have provided additional candidate target genes (17). In many cases, GAS motifs were detected within regulatory regions of these putative STAT5 target genes. However, due to the experimental design of these studies, it is not necessarily possible to distinguish between *bona fide* STAT5 target genes and genes whose induction is secondary to the STAT5 signaling pathway. Excessive STAT5 activation has been observed in various tumor cells and a case has been made that this occurs through the aberrant activation of STAT5 target genes (18). However, there is little experimental evidence to suggest that elevated STAT5 levels and activity are able to control otherwise dormant STAT5 targets genes *in vivo*.

Here, we have addressed two questions pertaining to STAT5 as a cytokine-controlled DNA binding protein and transcriptional activator. First, we determined GH-induced genome-wide binding of STAT5 and asked to what extent a 20-fold elevated concentration of STAT5 affects its binding pattern. Second, we asked to what extent the genes that bind STAT5 are controlled by STAT5 and GH. For this purpose, we have established a well controlled experimental system that permitted qualitative and quantitative analyses of STAT5 binding and corresponding transcriptional consequences. We have used mouse embryonic fibroblasts (MEFs) with three distinct genotypes; WT, $Stat5^{-/-}$ and $Stat5^{-/-}$ ectopically expressing STAT5A at a more than 20-fold elevated concentration. ChIP-seq experiments were used to identify genomic loci that bind STAT5 at normal and elevated concentrations upon GH stimulation. Subsequently, we analyzed the corresponding gene expression. This approach allowed us to address the following questions: (i) How many genomic loci bind STAT5 in the presence of GH? (ii) Does a greatly elevated STAT5A concentration alter STAT5 binding quality and quantity? (iii) Does STAT5 binding correlate with GH-dependent and independent transcription of neighboring genes?

MATERIALS AND METHODS

Mice, isolation of cells and cell culture

Mice were handled and housed in accordance with the guidelines of National Institute of Diabetes, Digestive and Kidney Diseases (NIDDK) Animal Care and Use Committee. $Stat5^{+/-}$ mice (2) were bred into the C57BL/6 background. $Stat5^{+/-}$ mice were intercrossed and MEFs were isolated from 12.5–13.5-day fetuses. MEFs were expanded in DMEM with high glucose supplements and 15% fetal bovine serum (FBS). MEFs of passage 9–10 were used for experiments. After 5 h of starvation in serum-free medium with 0.1% of BSA, MEFs were treated with growth hormone for indicated time period.

Retroviral infection

The retroviral-expression vector carrying a WT *Stat5A* gene was based on an MSCV-IRES-GFP backbone (gift from Richard Moriggl, Ludwig-Boltzmann Institute, Vienna, Austria). 293 T cells were transfected with the plasmid using FuGENE (Roche, Indianapolis, IN, USA). Supernatants were collected for 48–72 h after transfection and passed through a 0.45-μm filter before freezing at −80°C. For the infection, 10^6 $Stat5^{-/-}$ MEFs were seeded on a 10-cm culture dish and infected the next day with retrovirus in the presence of 8 μg/ml polybrene. After infection, non-fluorescent cells and GFP-expressing cells were isolated using BD FACS Vantage Cell Sorter Flow Cytometer (Becton Dickinson, San Jose, CA, USA) and sorted directly into PBS. Sorted MEFs were maintained in Dulbecco's Modified Eagle medium (DMEM) supplemented as described above.

Western blots

After 5 h starvation followed by 45 min stimulation with GH, protein was extracted from primary $Stat5^{+/+}$ MEFs and $Stat5^{-/-}$ MEFs overexpressing *Stat5A*. In total, 1×10^6 cells were pelleted by centrifuging for 1 min at 1500*g* and then washed once with ice-cold PBS. Cells were lysed by adding lysis buffer containing: 50 mM Tris pH 7.4, 150 mM NaCl, 10% glycerol, 1% IGEPAL (NP-40), 2 mM EDTA, 100 mM NaF and protease inhibitor cocktail (Roche) and incubating for 30 min on ice. Lysates were cleared by centrifugation in at 4°C in a microfuge at maximum speed. Protein concentration was determined by Bradford reagent (Bio-Rad, Richmond, CA, USA). A quantity of 10 μg of total protein was electrophoretically separated and transferred to polyvinylidene fluoride (PVDF) membranes using the NuPAGE system (Invitrogen, Carlsbad, CA, USA). Membranes were probed with anti-STAT5 (C-17), anti-STAT3 (C-20) and anti-beta actin antibodies (C-4) (Santa Cruz Biotechnology, Santa Cruz, CA, USA) and phospho-STAT5 (Y694) antibody from Cell Signaling Technology (Danvers, MA, USA) and HRP-conjugated secondary antibodies (GE Healthcare, Wauwatosa, WI, USA). Proteins were detected using enhanced chemiluminescence substrate (Pierce, Rockford, IL, USA), Kodak MR film (Kodak, Rochester, NY, USA) and Amersham Hyperfilm ECL (GE Healthcare, Wauwatosa, WI, USA).

RNA isolation and quantitative real-time PCR analysis

After 5 h of starvation, MEFs were cultured for 2 h in the absence or presence of GH and total RNA was isolated using RNeasy mini kit (Qiagen, Valencia, CA, USA). A quantity of 1 μg amounts of RNA were reverse transcribed (cDNA reverse transcription kit; Applied Biosystems, Foster City, CA, USA). Real-time quantification of mRNA transcript levels was performed using the TaqMan gene Expression Master Mix (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions. Real-time PCR was carried out using an ABI Prism 7900HT (Applied Biosystems, Foster City, CA, USA). TaqMan probes for STAT5a (Mm00839861_m1), Socs1 (Mm00782550_s1), Socs2 (Mm00850544_g1), Bcl6 (Mm00477633_m1) and beta-actin (4352341 E) were used (Applied Biosystems, Foster City, CA, USA) for Real-time PCR.

Microarray data collection and analysis

After 5 h of starvation, cells were treated for 2 h with GH. Total cellular RNA from each group of the MEFs was extracted with TRIzol reagent (Invitrogen) according to the manufacturer's instructions. Microarray analyses were performed using Affymetrix Mouse Genome 430 2.0 GeneChips (Affymetrix, Santa Clara, CA, USA) (four groups, biological triplicates for each group). Expression values were determined with GeneChip Operating Software (GCOS) v1.1.1 software. Robust Multichip Average (RMA) signals were summarized using GeneSpring GX 10.0.1 (Agilent) and normalized by quantile normalization. All data analysis was performed with GeneSpring software GX 10.01.

Chromatin immunoprecipitation coupled by illumina sequencing

Chromatin immunoprecipitation coupled by illumina sequencing (ChIP-seq) experiments were performed as described previously (19,20). In brief, after 5-h starvation, WT, *Stat5*$^{-/-}$, or STAT5A-*Stat5*$^{-/-}$ MEFs were stimulated with or without 1 μg/ml GH for 45 min. MEFs were then cross-linked with 1% formaldehyde for 10 min. Chromatin from 5×10^6 cells was used for each ChIP experiment. Antibodies against STAT5 (sc-835, Santa Cruz, CA, USA) and IgG (AB-105C, R&D System) were used. The ChIP DNA fragments were blunt-ended, ligated to the Solexa paired-end adaptors and sequenced with the Illumina Hi-seq 2000 genome analyzer.

ChIP-seq data analysis

Experimental data sets came from WT (WT-GH), Stat5$^{-/-}$ (KO-GH) and Stat5$^{-/-}$ overexpressing STAT5A (OE-GH) MEFs upon GH stimulation. The mapped tags of STAT5 antibodies and IgG samples were analyzed using the MACS program (21) with the following parameters; -mfold = 10,20 (fold enrichment threshold for the detection of highly enriched regions) -g mm (mouse genome) –p 1.0 e−4 (*P*-value cut-off). The identified peaks were further split by using PeakSplitter (22). ChIP-seq data obtained from GH-treated Stat5$^{-/-}$ MEFs (Transfected with an empty vector, KO-empty-GH) served as negative controls and were independently processed with the same parameters. A total of 28 343 (WT-GH), 117,771 (OE-GH), 38,451 (KO-GH) and 23,404 (KO-empty -GH) peaks were identified. Since the identified STAT5 binding peaks obtained in Stat5$^{-/-}$ MEFs (KO-GH and KO-empty-GH) should be false-positives; we only used those peaks that did not overlap with those of the KO sets. With this approach, a total of 23 827 and 111 939 STAT5 peaks were identified in the WT-GH and OE-GH sets, respectively. STAT5 binds to DNA directly and its DNA binding motif (GAS motif) has been well defined. We utilized the GAS motif information to accurately identify *bona fide* STAT5 binding sites. The Find Individual Motif Occurrences (FIMO) DNA motif identification tool (23) was used to scan the peak regions with a STAT5 position-specific scoring matrix (*P*-value threshold: 1 e−3). Out of the 23 827 and 111 939 STAT5 peaks, 13 278 and 71 561 peaks contained at least one GAS motif within their peak regions. These final peaks were regarded as *bona fide* STAT5 binding sites and used in this study. For the comparison analysis (Figures 3 and 4), the total number of tags was normalized to 10 million.

Gene classification

Genes were classified based on the quantity of STAT5 binding within a 2.5 kb promoter region and/or adjacent 8 kb of upstream region. STAT5 binding within each promoter region was measured by means of the maximum peak height, as well as the sum of peak heights. The maximum peak height (max_peak_height) indicates STAT5 affinity to target sites, whereas the sum of peak heights represents the overall amount of STAT5 binding. The sum of peak heights was converted to *z*-score for comparison (*z*-score_sum). To identify *bona fide* STAT5 target genes, the following peaks were defined as strong (or weak) peaks; either max_peak_height ≥95th percentile (70th for weak) among the heights of identified peaks or *z*-score_sum ≥ 5. With the defined peaks, genes were classified as follows; class I—genes with STAT5 bindings in both WT-GH and OE-GH promoter regions, class II—genes with STAT5 bindings only in OE-GH promoter region, class III—genes with STAT5 bindings in both WT-GH and OE-GH upstream regions and class IV—genes with STAT5 bindings only in OE-GH upstream region. With the strong peaks, 29, 161, 49 and 186 genes were categorized into class I, II, III and IV, respectively.

RESULTS

STAT5A overexpression in MEFs

WT, *Stat5*$^{-/-}$ and *Stat5*$^{-/-}$ MEFs overexpressing mouse STAT5A (*Stat5*$^{-/-}$; *Stat5A*) MEFs were prepared (see 'Materials and Methods' section). The *Stat5a m*RNA concentration in *Stat5*$^{-/-}$; *Stat5A* MEFs was more than 100-fold higher than in control cells (Figure 1A). The STAT5A protein level in *Stat5*$^{-/-}$; *Stat5A* MEFs was at least 20-fold higher than in control MEFs (Figure 1B). Similarly, greatly elevated p-STAT5 levels were observed in *Stat5*$^{-/-}$; *Stat5A* MEFs (Figure 1B) upon GH stimulation. This demonstrates that neither the GH receptor nor JAK2 were limiting in this system and that the elevated STAT5 levels could be activated.

Genome-wide *in vivo* mapping of STAT5A binding sites

To identify genetic loci readily recognized by activated STAT5 *in vivo*, we used ChIP-seq to analyze WT *Stat5*$^{-/-}$ (WT-GH) and STAT5A overexpressing *Stat5*$^{-/-}$; *Stat5A* MEFs (OE-GH) upon GH stimulation. *Stat5*$^{-/-}$ MEFs and rabbit serum IgG served as negative controls. Control and experimental MEFs were grown to ~70% confluency, starved for 5 h, followed by GH stimulation for 45 min and ChIP-seq analysis (see 'Materials and Methods' section).

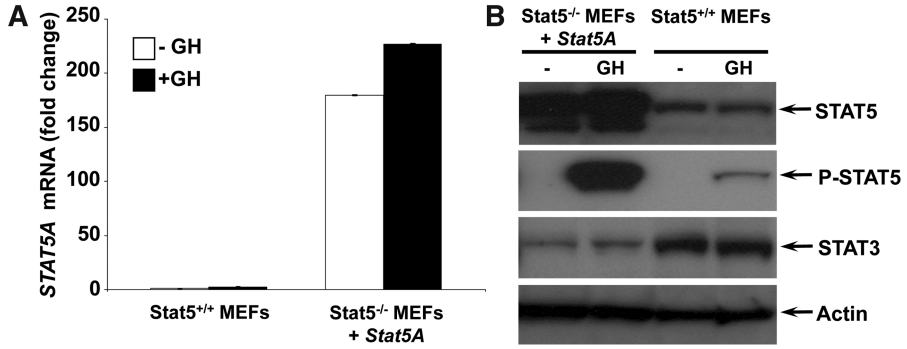

Figure 1. Analysis of WT and STAT5A overexpressing MEFs. $Stat5^{+/+}$ and $Stat5^{-/-}$ MEFs over expressing retrovirally transduced STAT5A ($Stat5^{-/-}$; Stat5A) were analyzed upon starvation and GH induction. (**A**) qPCR of mRNA isolated from control and transgenic MEFs in the absence and presence of GH. (**B**) western blot.

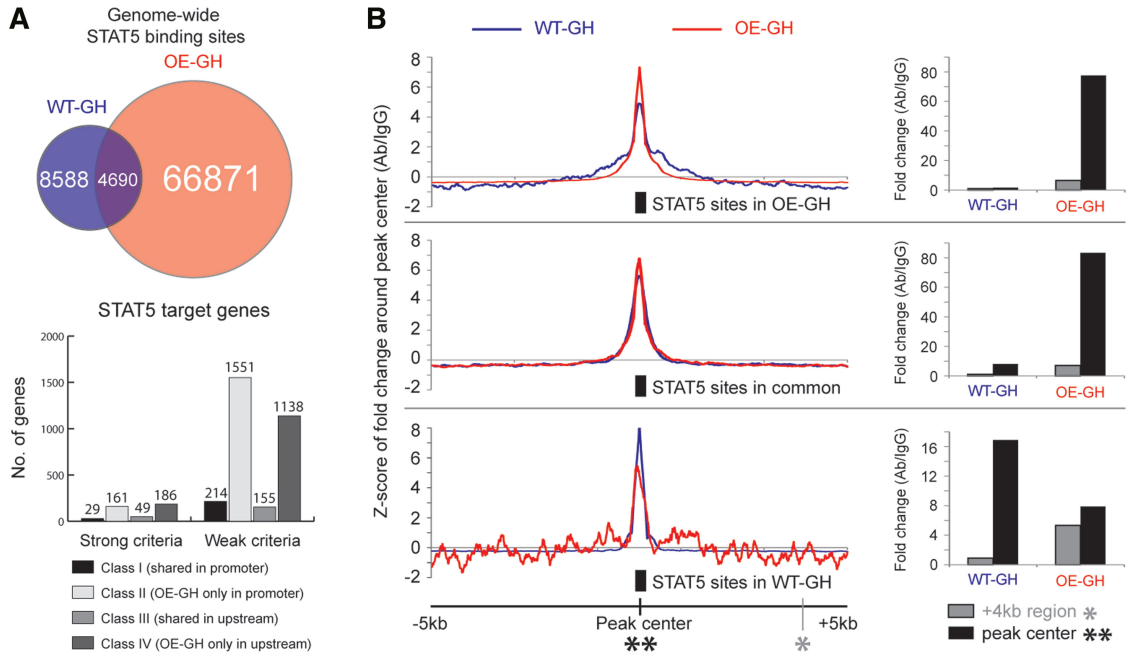

Figure 2. Comparison of genome-wide STAT5 binding in WT and STAT5A overexpressing MEFs. (**A**) Venn diagrams illustrate common and unique STAT5 binding sites between two different cell-contexts. The number of STAT5 target genes was calculated according to strong and weak criteria (see 'Materials and Methods' section). (**B**) STAT5 binding was quantified. Fold changes (antibody/IgG) around peak summits (±5 kb) were standardized in each group using *z*-score. Bar graphs in the right panel show actual fold changes at the peak summit and a region 4 kb away from the peak summit.

The results were analyzed as follows. Raw mapped tags of the WT-GH and OE-GH samples were processed with corresponding IgG controls by using a peak-calling program, MACS (see 'Materials and Methods' section for details) (21). A total of 28 343 and 117 771 regions were initially identified in the WT-GH and OE-GH sets, respectively. Subsequently, regions overlapping with STAT5 enriched sites in the *Stat5*$^{-/-}$ MEFs samples were discarded. We pinpointed GAS motifs within the binding regions using the FIMO motif identification tool (23) and defined peaks containing at least one GAS motif as *bona fide* STAT5 binding sites. These high quality peaks were used throughout this study. In sum, 13 278 and 71 561 STAT5 peaks were regarded as *bona fide* STAT5 binding sites in WT-GH and OE-GH MEFs, respectively (Figure 2). In total, 4690 peaks were shared between the two data sets and 66 871 were *de novo* sites (Figure 2A).

To further identify *bona fide* STAT5 target genes, we focused on peaks located within 2.5 kb of promoter sequences (−2 kb to +0.5 kb) or upstream sequences (−10 kb to −2 kb). Based on the max_peak_height we called strong (95th percentile) and weak (70th percentile) peaks (see 'Materials and Methods' section). Using these criteria we defined four classes of genes; class I—genes with STAT5 binding in their promoters in both WT-GH and OE-GH MEFs; class II—genes with STAT5 binding in their promoters only in OE-GH MEFs; class III—genes with STAT5 binding in upstream regions in both WT-GH and OE-GH MEFs; class IV—genes with STAT5 binding in upstream regions only in OE-GH MEFs. Based on

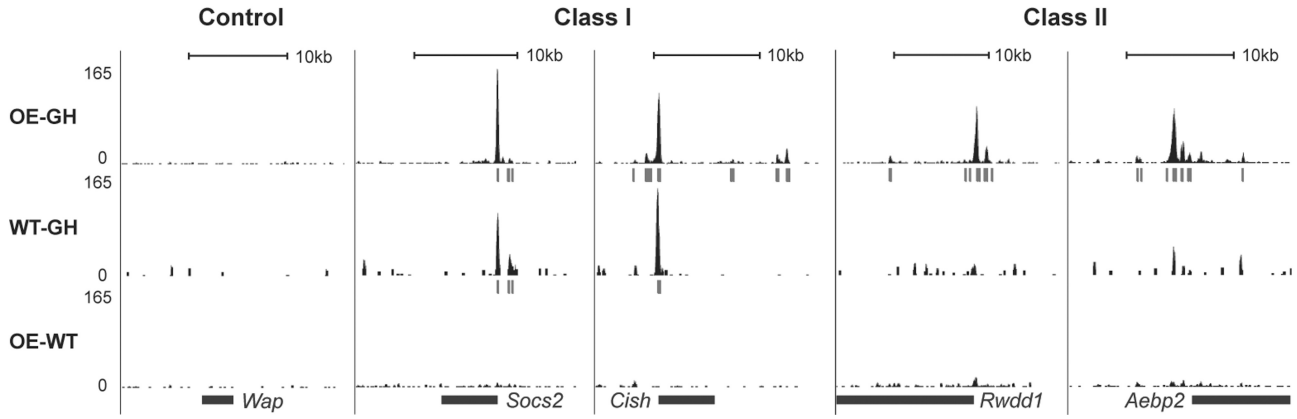

Figure 3. Examples of STAT5 target genes. ChIP-seq analysis pinpoints STAT5 binding sites within the 2.5 kb of promoter regions of the *Wap*, *Socs2*, *Cish*, *Rwdd1* and *Aebp2* genes in WT (WT-GH) and STAT5 overexpressing (OE-GH) MEFs stimulated with GH. STAT5 binding patterns in STAT5 overexpressing MEF (OE-WT, no GH) are shown as negative control. Horizontal bars indicate genes and small vertical bars show locations of GAS motifs.

strong peaks, 29, 161, 49 and 186 genes were categorized into class I, II, III and IV, respectively (Figure 2A). In general, STAT5 binding to promoter regions (−2 kb to +0.5 kb) was much stronger than binding to upstream sequences (−10 kb to −2 kb). In addition, we quantified the level of STAT5 bindings in each group using *z*-score. Comparison analysis reveals that the level of STAT5 binding is changed according to the context of cells. For example, 8588 STAT5 binding sites show very weak STAT5 binding (or not bound by STAT5) in the OE-GH context in terms of the fold change (Figure 2B) although these two cells were grown in very similar conditions except with the dosage of STAT5.

Class I: STAT5 binding to promoter regions independent of STAT5 concentration. Using strong and weak criteria, STAT5 bound to promoter sequences of 29 and 214 genes, respectively (Supplementary Table S1). The class of 29 genes encompassed *bona fide* STAT5 target genes, including *Socs2*, *Socs3*, *Cish* and *Bcl6* (Figure 3 and Table 1). Overall, both the maximum peak height and the peak sum were similar or slightly higher in STAT5A overexpressing cells compared with WT MEFs (Table 1). In general, STAT5 binding was observed only in the presence of GH.

Although higher STAT5A concentrations did not significantly alter binding at GAS motifs recognized already in WT MEFs, *de novo* binding to juxtaposed GAS motifs within a given gene was observed in several cases. For example, while promoter-bound GAS motifs in the *Bcl6* gene were recognized by STAT5 independent of its concentration, a subset of intronic GAS sites was recognized only upon STAT5A overexpression (Figure 4). The 12 kb *Bcl6* locus contains 20 GAS motifs within the STAT5 binding areas, 13 of which are conserved between mouse and opossum. Under physiological STAT5 concentrations, five conserved GAS motifs within the promoter region were occupied. *De novo* occupancy of a set of four highly conserved GAS motifs within the first intron but not other conserved GAS motifs was obtained at elevated STAT5 concentration (Figure 4). Within the *Cish* locus, de novo binding was observed over five conserved GAS motifs within the 3′ flanking region, although the degree of binding over promoter-bound GAS motifs remained unchanged (Figure 3). Similarly, elevated STAT5A levels led to the occupation of *de novo* binding sites within the promoter upstream region of the *Socs3* gene (Figure 4). The two distal ones are highly conserved between species.

Based on global gene expression analyses of these 29 genes, all 26 represented on the array were expressed. Expression ranged over 40-fold between the lowest (*Irf2*) and highest (*Ubc*) (Table 1). However, expression of only four genes (*Socs2*, *Socs3*, *Cish* and *Bcl6*) was under GH-STAT5 control in WT MEFs (Table 1). At 20-fold elevated STAT5 concentration, one additional gene (*Tle1*) entered this category (Table 1). Whereas GH-induced expression of the *Socs3* and *Bcl6* genes was not altered by extra STAT5A, *Socs2* and *Cish* mRNA levels increased (Table 1). Expression of selected genes was verified using RT–PCR (Supplementary Figure S1). These data demonstrate that in WT cells, GH-induced STAT5 binding to promoter sequences coincides only for some genes with GH-induced expression and STAT5 dependency.

Class II: STAT5 binding to promoter regions only at elevated STAT5A concentration. Using strong and weak criteria, STAT5 bound to promoter sequences of 161 and 1551 genes, respectively (Supplementary Table S1). Genes in this group, which includes *Esr1* and *Rwdd1*, exhibit STAT5 binding only in STAT5 overexpressing MEFs (OE-GH). Out of the 161 genes, 135 were represented by probes on the Affymetrix arrays. Out of the 135 genes, 18 were not expressed at detectable levels (Table 2). Among those that were expressed, mRNA levels ranged over 30-fold between genes. Using a 2-fold cut-off, only the *Esr1* gene had acquired STAT5-dependent GH-induction in STAT5A overexpressing MEFs (Table 2). These results demonstrate that elevated STAT5 levels produce *de novo* STAT5 binding to genes that are not targeted at lower

Table 1. STAT5 binding and expression profile in Class I genes (STAT5 binding in promoters regions in both WT-GH and OE-GH MEFs)

Name	Max_peak_height				Gene expression			
	WT-GH Upstream	WT-GH Promoter	OE-GH Upstream	OE-GH Promoter	WT	WT-GH	OE	OE-GH
Cish	0	149.3	27.2	120.7	148	695	<100	1217
Myd	0	39.9	0	119.1	1690	1866	1213	1664
Serpinh1	0	56.1	44.4	101.4	7166	6290	5498	5565
Plec	34.9	75.6	27.7	96.7	6450	7335	7300	7900
Socs3	0	92.8	71.6	83.6	738	1375	255	1097
Fn1	22.4	35.3	14.6	79.9	10 073	9805	10 399	10 814
Ubc	22.6	26.5	11.5	79.9	11 967	11 706	12 006	12 847
Psmd3	0	68.5	0	76.8	1179	1256	1264	1253
Nckap5	19.8	47.9	0	71.6	na			
Irf2	0	34.9	48.6	69.5	214	221	269	308
Slc25a37	0	45.3	0	66.4	625	739	817	1062
Bcl6	32.5	82	23.5	62.2	3032	1010	2463	1105
Ctgf	13	22.4	91.4	61.7	7518	7162	6891	8105
Rhoq	0	29.9	0	60.6	5172	4481	3701	3924
Gm8580	0	43.1	6.8	58	na			
Gm4262	0	40.1	0	57	na			
Gclm	0	24.1	0	57	473	559	414	509
Nacc2	0	32.1	15.7	55.9	1359	1285	2211	1926
Poli	0	41.9	7.3	53.3	232	172	425	473
Plekhg2	32.3	34.9	59.6	52.8	2484	3112	2950	4847
Ergic2	0	22.2	6.3	49.6	1018	1138	1170	1191
Ppm1k	0	57.5	0	49.1	425	235	314	412
Cyp1b1	0	46.7	0	48.6	9005	8340	5645	4382
Katnb1	0	59.5	0	47.5	133	186	229	344
Ugp2	0	50.5	10.5	44.9	3319	3138	4230	5644
Flot1	0	69.7	7.8	39.2	3591	3242	5209	5803
Sgk1	46.1	48.5	15.2	34.5	4446	4030	7093	8139
Tle1	0	39.3	0	34.5	192	223	338	703

The peak height numbers refer to sequence tags and the gene expression numbers are relative expression levels provided by the Affymetrix software. na, not available.

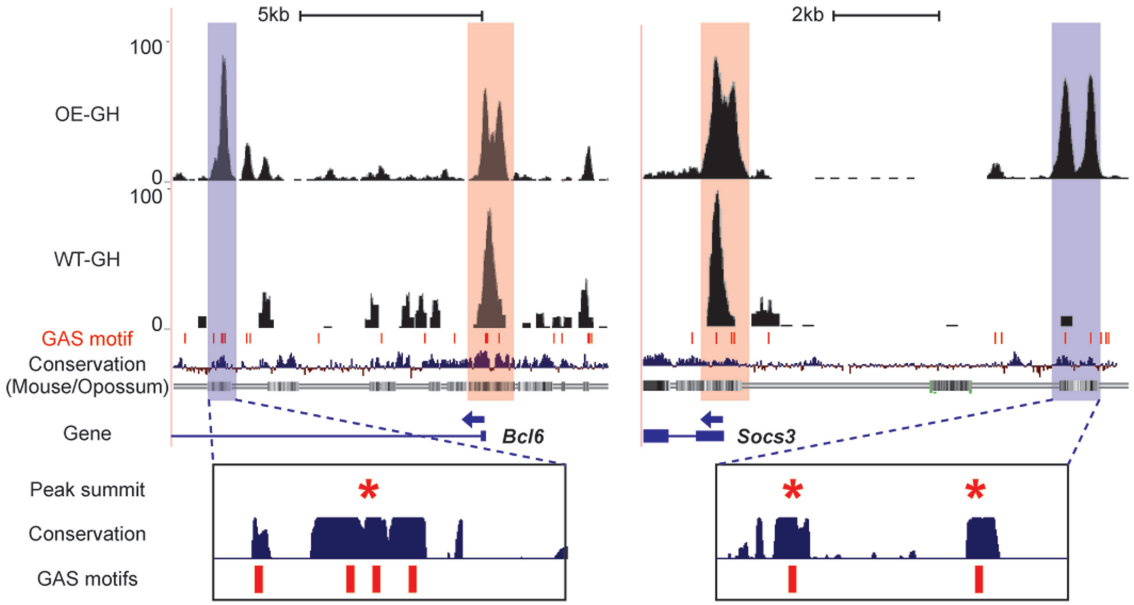

Figure 4. Browser images of *Bcl6* and *Socs3*. The STAT5A overexpression upon GH induction causes additional binding to GAS motifs near the promoter regions of *Bcl6* and *Socs3*. Light red vertical bars indicate the STAT binding regions shared between WT (WT-GH) and STAT5 overexpressing (OE-GH) MEFs. Light blue vertical bars represent unique STAT binding sites shown only in OE-GH, but not in WT-GH. The bottom panel depicts positions of GAS motifs, gene and peak summits, as well as conservation score (PhastCon score).

Table 2. STAT5 binding and expression profile in Class II genes (STAT5 binding in promoter regions only at elevated STAT5A concentration)

Name	Max_peak_height				Gene expression			
	WT-GH Upstream	WT-GH Promoter	OE-GH Upstream	OE-GH Promoter	WT	WT-GH	OE	OE-GH
Ncald	0	0	16.2	176.1	<100	<100	<100	130
Rbm25	0	0	0	117	na			
Cr1l	0	0	0	99.8	1979	2253	2161	2377
Emp1	0	0	0	97.7	3718	3276	2154	2434
Rwdd1	0	0	0	97.2	3063	3358	3980	4549
Fzd7	0	0	31.9	94.1	2645	2367	2374	1965
Kpna1	0	0	0	90.4	2987	3190	2568	3110
Nedd9	0	0	0	89.9	2382	1812	2023	1480
Pan2	0	0	0	88.3	407	390	591	500
Aebp2	0	0	11.5	86.7	3546	4429	3870	4164
Ccl2	33.7	0	35	84.6	3744	13568	171	3522
Olfr1423	0	0	10.5	84.6	na			
Opn5	11.8	0	5.7	84.6	na			
Dse	0	0	44.4	84.1	3272	2816	3427	4141
Mkl1	0	0	0	82	1091	1276	2155	2538
Rbms1	0	0	0	81	6571	6109	5408	6081
Shq1	0	0	16.7	79.9	191	265	138	225
Ppp1r12b	0	0	18.3	77.3	392	687	<100	145
Phlda1	0	2.4	3.1	75.8	922	1477	477	646
Zfp36l1	0	0	49.1	75.8	2867	3316	3397	3604
Psmd4	0	0	0	74.7	783	1027	920	1224
Ntn4	0	0	0	72.6	<100	<100	<100	<100
Tnc	0	0	0	69.5	7822	9009	3884	4716
Dapk3	0	0	28.7	68.5	405	391	783	870
Aldh18a1	0	0	0	66.4	4119	3250	2654	2851
Esr1	0	0	0	60.1	190	151	237	552
Snai3	0	0	8.9	54.3	<100	<100	<100	<100
Pxk	0	0	10.5	48.6	537	531	385	562
1110054M08Rik	0	0	9.4	38.7	na			
Mical1	0	0	0	28.2	2377	2538	4588	4862

The peak height numbers refer to sequence tags and the gene expression numbers are relative expression levels provided by the Affymetrix software. na, not available.

concentrations. However, this *de novo* binding does not place these cells under STAT5-dependent GH control. Moreover, 10% of the *de novo* recognized genes are silent.

Class III: STAT5 binding to gene upstream regions independent of STAT5 concentration. STAT5 bound to the upstream region (−10 kb to −2 kb) of 49 genes both in WT and STAT5A overexpressing cells. Out of the 39 genes with probes on the Affymetrix chips, 7 were not expressed at detectable levels (Table 3). Two genes, *Sbno2* and *Mapkapk3*, acquired modest GH-inducibility in GH overexpressing cells. *Sbno2* belongs to a family of helicase co-repressors and its expression is induced in macrophages by IL-10 through STAT5 (24).

Class IV: STAT5 binding to gene upstream regions only at elevated STAT5 concentration. Upon STAT5 overexpression, *de novo* binding to the upstream region (−10 kb to −2 kb) of 186 genes was observed (Table 4). Out of the 158 genes represented on the microarray, expression of 41 was undetectable. Out of the 117 expressed genes, 4 acquired a slightly <2-fold GH induction in STAT5 overexpressing MEFs (Table 4).

A biological replicate ChIP-seq experiment with STAT5A overexpressing MEFs in the absence and presence of GH validated and substantiated the above findings (Supplementary Figure S2). The pattern of STAT5 binding peaks is very similar to the original one, and we confirmed the unique STAT5 binding sites (Figure 4) which are only observed in the OE-GH cell-context. The MEME-ChIP program (25) successfully identified the canonical STAT5 binding motif (TTC…GAA) as the most frequently occurring sequence at the top 1000 STAT5 binding regions. Among the top 10 000 peaks identified in the replicate, 86% of the peaks overlapped with peaks in the original data set.

Motif analysis

Approximately 60% of STAT5 binding is located over GAS motifs, whereas in 40% of the binding areas no classical GAS motifs are found. The identified areas of STAT5 binding were separated into two sets according to the existence of GAS motif (canonical versus non-canonical). *De novo* motif prediction (25) that uses public motif databases for co-factor identification (26) established *bona fide* GAS motifs in all peaks in STAT5 overexpressing MEFs (Supplementary Figure S3A). Peak height in WT MEFs was in general lower and convincing GAS motifs were found only in strong peaks. In search for additional transcription factor binding motifs co-localizing with GAS motifs, we identified AP1 motifs in strong and intermediate peaks of STAT5 overexpressing MEF (Supplementary Figure S3). No clear

Table 3. STAT5 binding and expression profile in Class III genes (STAT5 binding in upstream regions independent of STAT5 concentration)

Name	Max_peak_height				Gene xpression			
	WT-GH Upstream	WT-GH Promoter	OE-GH Upstream	OE-GH Promoter	WT	WT-GH	OE	OE-GH
Hsp90aa1	43.9	0	156.8	0	2909	3218	2626	2978
Mapkapk3	149.3	0	120.7	27.2	892	1614	560	1317
Plec	75.6	63.9	96.7	0	6450	7335	7300	7900
Trim7	33.7	0	96.7	0	<100	<100	100	<100
Mtap4	45.9	0	87.8	16.2	4630	4536	5153	5344
Birc2	62.1	0	81	0	720	1324	879	931
Pds5a	37.7	0	81	0	142	160	185	298
0610040F04Rik	62.5	42.5	79.4	7.3	na			
Necap1	48.1	0	79.4	6.3	2449	2307	2175	1915
Ptrf	45.3	0	76.3	0	3027	2913	2742	2447
Nnat	36.7	0	75.8	0	331	361	<100	<100
Snrpf	36.7	0	73.2	0	na			
Pofut2	36.9	0	70	6.8	7222	6245	6491	7115
Sbno2	40.9	0	68.5	37.1	95	178	128	272
Stk11	40.9	23.4	68.5	12.5	345	445	580	764
Thy1	41.5	0	65.8	7.3	8465	7670	<100	<100
Ptgr2	46.9	0	65.8	0	1206	1050	1123	875
Lgals3	62.9	0	64.3	6.8	5029	5664	5246	6022
Zfp810	40.9	0	63.2	0	777	557	1209	1037
Atg16l2	51.5	0	61.7	18.3	149	177	368	484
Cln6	28.1	0	61.7	0	1616	1437	1946	1514
Dnaic2	21.8	0	61.1	0	<100	<100	<100	<100
Zfp36	32.3	0	59.6	20.4	954	1010	1228	620
Arhgap24	56.1	0	58.5	21.9	778	585	766	471
Afap1l2	39.7	0	58	12.5	217	245	<100	<100
Gja5	34.1	0	57.5	11	115	<100	<100	<100
Mpp1	75	0	56.4	18.8	1814	2076	2989	4150
Hcfc2	45.7	0	51.7	0	990	913	1173	1075
Trib3	95	0	50.7	31.9	1954	886	889	416

The peak height numbers refer to sequence tags and the gene expression numbers are relative expression levels provided by the Affymetrix software. na, not available.

consensus motifs were identified in the non-canonical STAT5 binding areas.

Cell-specific STAT5 target genes

A number of cell-specific STAT5 target genes have been reported. Among them are prolactin-induced milk protein genes in mammary tissue (27–29), *FoxP3* in Regulatory T cells (Tregs) (10), *Tbx21*, *Il4ra* and *Il17rb2* in T-cells (ref) and *IL17* in TH17 cells (30). Regardless of the STAT5 concentration, no binding was observed to GAS motifs in established regulatory regions of milk protein genes, the IL17a and IL17f genes and the Tbx21 and Il12rb2 genes that are activated by IL-2 through STAT5 in Th1 cells (30). Similarly, no binding was observed in the Il4ra gene that is induced by IL-2 through STAT5 in T-cells or the Foxp3 gene. These findings demonstrate that STAT5 is unable to access GAS motifs within genuine cell-specific STAT5 target genes even at very high concentrations and levels of activation.

Correlation analysis between gene expression levels and STAT5 binding

A total of nearly 900 genes in MEFs were induced more than 2-fold by GH (Supplementary Table S2). For each gene in the four classes, we established the correlation between the extent of STAT5 binding and GH-induced expression. Gene expression was estimated from microarray analysis (y-axis) and the degree of STAT5 bindings was calculated using maximum peak height in either promoter or upstream regions according to the classification (x-axis) (see 'Materials and Methods' section). Correlation was measured with Pearson's correlation coefficient. Each spot indicates a single gene. As shown in Figure 5, in none of the four classes was a direct correlation between the extent of STAT5 binding and the level of gene expression.

We extended data analysis and included STAT5 binding sites at genomic regions associated with gene body and downstream sequences (+0.5 kb ~ +9 kb) (Supplementary Figure S4). We were unable to establish a direct correlation between gene expression and STAT5 binding strength.

DISCUSSION

Although it is well established that many cytokines activate the transcription factor STAT5, evidence that STAT5 binding controls cytokine-induced gene expression is based on a reverse experimental designed. Uncovering that expression of a given gene is regulated by cytokines leads to the search for GAS motifs in the respective promoter region and establishing *in vitro* and *in vivo* STAT5 binding, followed by reporter gene experiments

Table 4. STAT5 binding and expression profile in Class IV genes (STAT5 binding in upstream regions only at elevated STAT5 concentration)

Name	Max_peak_height				Gene expression			
	WT-GH Upstream	WT-GH Promoter	OE-GH Upstream	OE-GH Promoter	WT	WT-GH	OE	OE-GH
Mir3091	0	0	144.2	0				
Rps21	0	0	144.2	0	15 214	15 729	15 247	14 743
C130039O16Rik0		0	127	15.7	na			
Zfp251	0	0	125.4	13.6	3423	3275	5425	5898
Zfp7	0	0	125.4	0	132	176	329	577
Adc	0	0	119.7	24	226	247	228	483
Eif2s2	0	0	119.7	8.9	8245	7973	6918	7112
Adamts5	0	0	118.6	4.2	5356	4007	5748	5027
St5	0	0	116	0	2865	2953	2493	2208
Themis	0	0	108.7	0	na			
Gpd2	0	0	107.6	21.4	1064	1283	896	1905
Sumo3	0	7	100.8	0	5796	5748	5579	5039
Slc18a1	0	0	100.3	7.3	<100	<100	<100	<100
Zfr	0	0	100.3	6.3	6698	6307	7507	6736
Pdp1	0	0	98.8	8.9	na			
2210012G02Rik	0	17.8	93	23	389	254	869	803
Speer7-ps1	0	0	89.3	14.6	<100	<100	603	584
Puf60	0	0	88.8	8.4	6338	6866	8408	8244
Il23a	0	0	88.3	0	<100	<100	<100	<100
Loxl2	0	0	87.3	0	6344	6737	<100	<100
Aebp2	0	0	86.7	16.2	3546	4429	3870	4164
Angptl2	0	0	86.7	0	4183	4202	5336	5405
Txnip	0	0	86.7	0	2182	1628	2912	1651
Chit1	0	0	86.2	0	<100	<100	<100	<100
Dock5	0	0	83.1	11	1452	1287	2106	2997
Zfp786	0	0	82	0	<100	<100	<100	<100
Snd1	0	19.2	81.5	19.3	2258	2411	2491	2180
Fam125a	0	0	81	0	1095	1274	2568	3098
Dennd1a	0	0	79.4	9.4	<100	<100	<100	<100
Cyb5d2	0	21	76.3	7.8	433	358	170	143

The peak height numbers refer to sequence tags and the gene expression numbers are relative expression levels provided by the Affymetrix software. na, not available.

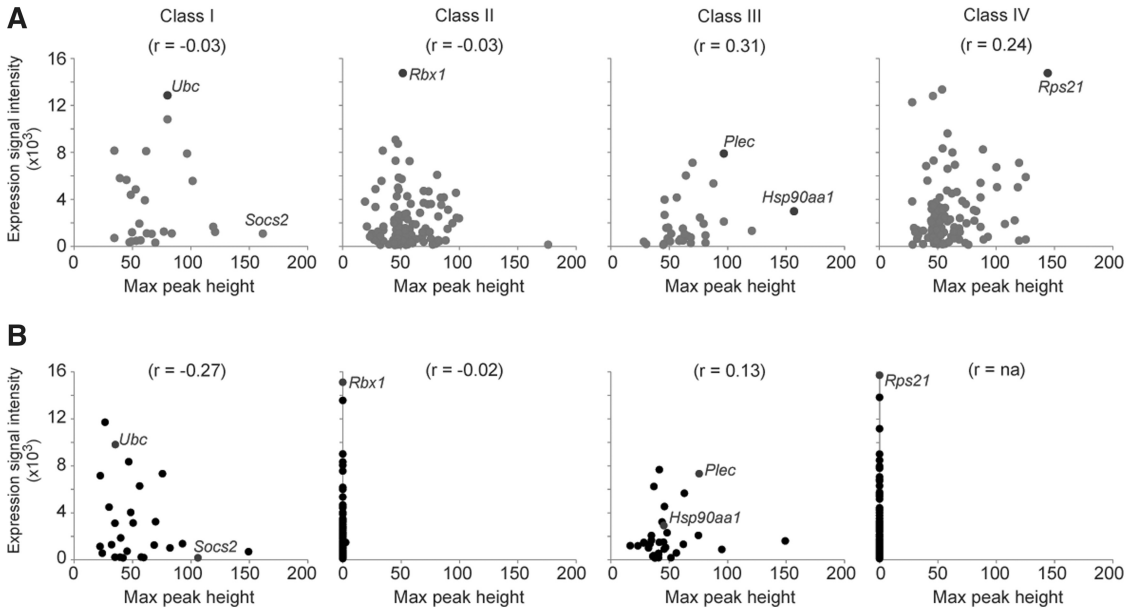

Figure 5. Correlation analysis between gene expression levels and STAT5 binding in MEFs upon GH induction. The level of gene expressions was estimated from microarray analysis (y-axis). The degree of STAT5 bindings was calculated using maximum peak height in either promoter or upstream regions according to the classification (x-axis) (see 'Materials and Methods' section). Correlation was measured with Pearson's correlation coefficient. Each spot indicates a single gene. (**A**) OE-GH. (**B**) WT-GH.

carrying these GAS motifs serves as a paradigm to demonstrate the link between cytokine stimulation, transcription factor binding and gene expression. However, it is not clear how many genes that are recognized by STAT5 are actually under STAT5 and cytokine control. We addressed this question by genome-wide identification of all genes that bind STAT5 and correlating their expression at different STAT5 concentrations.

The GAS motif sequence TTCnnnGAA occurs statistically once every 4096 bp and therefore can be found in the promoter/upstream sequence of virtually every single gene. At physiological concentrations of STAT5, GH-induced binding was observed at 13 278 sites with GAS motifs within the genome of MEFs. Using stringent conditions, 78 genes featured STAT5 binding within 10 kb of promoter/upstream sequences. However, only four of these genes (*Socs2, Socs3, Cish and Bcl6*) were soundly under the control of GH by STAT5. A >20-fold over-expression of STAT5A resulted in 71 561 GH-induced STAT5 binding peaks. Under stringent conditions, 347 genes featured *de novo* STAT5 binding within 10 kb of promoter/upstream sequences. Up to 15% of these genes did not show any detectable basal or GH-induced expression despite strong GH-induced STAT5 binding. Out of the 234 expressed genes, only 4 acquired a modest (<2-fold) GH-induction upon STAT5 overexpression.

Analyzing genome-wide GH-induced STAT5 binding and gene expression at different STAT5 concentrations revealed novel information about the role of STAT5 as a transcription factor. First, GH-induced STAT5 binding intensity to regulatory regions in *bona fide* STAT5 targets, such as *Socs2* and *Cish* is relatively concentration independent. Thus, in these cases STAT5 is not a limiting factor. However, GH-induced expression of both of these genes increased in STAT5 overexpression. It is possible, that the higher STAT5 concentration allows binding over longer time windows. In contrast, regulation of the *bona fide* STAT5 target *Bcl6* is independent of STAT5 levels although additional binding sites within the *Bcl6* gene were uncovered upon STAT5 overexpression. Second, the majority of GH-induced STAT5 binding in WT cells is associated with genes that are expressed over a wide range but are not under GH control. This finding demonstrates that STAT5 binding to promoter upstream sequences does not automatically convey STAT5 control over those genes. Third, increasing STAT5 concentrations by at last 20-fold, which is accompanied by GH-induced STAT5 phosphorylation, greatly increases the number of genes bound by STAT5. However, only a small fraction (18.6%, 30 out of 161) of these genes acquired GH-STAT5 control (>1.5 fold). Moreover, basal and GH-induced expression of a large number of genes that acquire STAT5 binding is undetectable. This finding demonstrates that at elevated concentrations, STAT5 can bind to otherwise dormant GAS motifs in promoter upstream sequences of both expressed and silent genes. However, STAT5 binding is not sufficient to convey transcriptional activation. Fourth, although STAT5 binding intensity to regulatory regions in *bona fide* target genes was mostly independent of the STAT5 concentration, STAT5 occupancy of a restricted set of neighboring GAS motifs was observed in several genes, including *Socs3* and *Bcl6*. This demonstrates that dependent on the concentration, STAT5 can also bind to dormant sites within STAT5 target genes. However, this binding appears to be of little or no consequence. A similar discrepancy between transcription factor binding and transcriptional regulation has been reported for the ligand-induced glucocorticoid receptor (GR) (31). More than 8200 GR binding sites were detected in a mammary cell line but there was no clear association with transcriptional regulation of nearby genes.

Of the STAT5 binding peaks, ~50% coincide with GAS motifs, confirming the concept that STAT5 binds to specific sequence motifs. However, the nature of STAT5 binding to sequences without *bona fide* GAS motif is not clear. We were unable to identify a unifying sequence motif in non-canonical STAT5 binding areas. In contrast, AP1 motifs were identified in canonical STAT5 binding areas at elevated STAT5 levels, suggesting cooperativity between STAT5 and AP1.

ChIP-seq analyses have shed light onto binding patterns of different STAT family members (8,9,30–33) in several cell types. Dependent on the specific STAT member and the investigating lab, each individual STAT member binds to between several hundreds and >4000 genes. In many cases there is substantial overlap of binding between different STATs members. For example, out of the >4000 binding sites for STAT4, 50% were also targeted by STAT6 and *vice versa* (33). The same seems to hold true for STAT3 and STAT5 (9). This observation is particularly relevant in light that different STATs not only serve similar functions but also can serve opposite functions. Although ChIP-seq studies provide a detailed view on the genome-wide occupation by a given transcription factor, the biological relevance of these findings remains to be determined. Specifically, it is not clear why only a small fraction of genes binding any given transcription factor appear to be under its control. This enigma cannot be explained by the lack of polymerase recruitment as many genes binding STAT5 display already high basal activity. Moreover, there were no apparent differences between GAS motifs and neighboring sequences between productive and non-productive sites. It is possible that STAT5 is a weak transcription factor that has considerable activity only in specific contexts.

ACCESSION NUMBERS

Raw microarray data can be accessed by GSE33688 and ChIP-seq data by GSE34896 from the NCBI Gene Expression Omnibus at http://www.ncbi.nlm.nih.gov/geo.

SUPPLEMENTARY DATA

Supplementary data are available at NAR Online: Supplementary Tables 1 and 2 and Supplementary Figures 1–4.

ACKNOWLEDGEMENTS

We thank G. Poy for help with microarray experiments, Y. Cui and W. Liao for technical help and Richard Moriggl for reagents.

FUNDING

The Intramural Research Programs (IRP) of NIDDK, NIAMS and NEI at the National Institutes of Health (NIH), USA; Epigenomic Research Program for Human Stem Cells (grant no. 2007-2004134); the World Class University Program, Ministry of Education, Science and Technology, through the National Research Foundation of Korea, South Korea (R31-10069); the National Basic Research Program of China (973 Program, grant no. 2012CB518501); Natural Science Fund for Colleges and Universities, Department of Education, Jiangsu, China (grant no. 11KJA360003); L.H. is an adjunct member of the Department of Nanobiomedical Science and WCU Research Center for Nanobiomedical Science, Dankook University, Chungnam, Korea. Funding for the open access charge: WCU Research Center, Dankook University National Institutes of Health.

Conflict of interest statement. None declared.

REFERENCES

1. Hennighausen,L. and Robinson,G.W. (2008) Interpretation of cytokine signaling through the transcription factors STAT5A and STAT5B. *Genes Dev.*, **22**, 711–721.
2. Cui,Y., Riedlinger,G., Miyoshi,K., Tang,W., Li,C., Deng,C.X., Robinson,G.W. and Hennighausen,L. (2004) Inactivation of Stat5 in mouse mammary epithelium during pregnancy reveals distinct functions in cell proliferation, survival, and differentiation. *Mol. Cell Biol.*, **24**, 8037–8047.
3. Cui,Y., Hosui,A., Sun,R., Shen,K., Gavrilova,O., Chen,W., Cam,M.C., Gao,B., Robinson,G.W. and Hennighausen,L. (2007) Loss of signal transducer and activator of transcription 5 leads to hepatosteatosis and impaired liver regeneration. *Hepatology*, **46**, 504–513.
4. Hoelbl,A., Schuster,C., Kovacic,B., Zhu,B., Wickre,M., Hoelzl,M.A., Fajmann,S., Grebien,F., Warsch,W., Stengl,G. *et al.* (2010) Stat5 is indispensable for the maintenance of bcr/abl-positive leukaemia. *EMBO Mol. Med.*, **2**, 98–110.
5. Hosui,A., Kimura,A., Yamaji,D., Zhu,B.M., Na,R. and Hennighausen,L. (2009) Loss of STAT5 causes liver fibrosis and cancer development through increased TGF-{beta} and STAT3 activation. *J. Exp. Med.*, **206**, 819–831.
6. Yao,Z., Cui,Y., Watford,W.T., Bream,J.H., Yamaoka,K., Hissong,B.D., Li,D., Durum,S.K., Jiang,Q., Bhandoola,A. *et al.* (2006) Stat5a/b are essential for normal lymphoid development and differentiation. *Proc. Natl Acad. Sci. USA*, **103**, 1000–1005.
7. Zhu,B.M., McLaughlin,S.K., Na,R., Liu,J., Cui,Y., Martin,C., Kimura,A., Robinson,G.W., Andrews,N.C. and Hennighausen,L. (2008) Hematopoietic-specific Stat5-null mice display microcytic hypochromic anemia associated with reduced transferrin receptor gene expression. *Blood*, **112**, 2071–2080.
8. Laurence,A., Tato,C.M., Davidson,T.S., Kanno,Y., Chen,Z., Yao,Z., Blank,R.B., Meylan,F., Siegel,R., Hennighausen,L. *et al.* (2007) Interleukin-2 signaling via STAT5 constrains T helper 17 cell generation. *Immunity*, **26**, 371–381.
9. Yang,X.P., Ghoreschi,K., Steward-Tharp,S.M., Rodriguez-Canales,J., Zhu,J., Grainger,J.R., Hirahara,K., Sun,H.W., Wei,L., Vahedi,G. *et al.* (2011) Opposing regulation of the locus encoding IL-17 through direct, reciprocal actions of STAT3 and STAT5. *Nat. Immunol.*, **12**, 247–254.
10. Yao,Z., Kanno,Y., Kerenyi,M., Stephens,G., Durant,L., Watford,W.T., Laurence,A., Robinson,G.W., Shevach,E.M., Moriggl,R. *et al.* (2007) Nonredundant roles for Stat5a/b in directly regulating Foxp3. *Blood*, **109**, 4368–4375.
11. Hwa,V., Nadeau,K., Wit,J.M. and Rosenfeld,R.G. (2011) STAT5b deficiency: lessons from STAT5b gene mutations. *Best Pract. Res. Clin. Endocrinol. Metab.*, **25**, 61–75.
12. Vafaizadeh,V., Klemmt,P., Brendel,C., Weber,K., Doebele,C., Britt,K., Grez,M., Fehse,B., Desrivieres,S. and Groner,B. (2011) Mammary epithelial reconstitution with gene-modified stem cells assigns roles to Stat5 in luminal alveolar cell fate decisions, differentiation, involution, and mammary tumor formation. *Stem Cells*, **28**, 928–938.
13. Moriggl,R., Sexl,V., Kenner,L., Duntsch,C., Stangl,K., Gingras,S., Hoffmeyer,A., Bauer,A., Piekorz,R., Wang,D. *et al.* (2005) Stat5 tetramer formation is associated with leukemogenesis. *Cancer Cell*, **7**, 87–99.
14. Warsch,W., Kollmann,K., Eckelhart,E., Fajmann,S., Cerny-Reiterer,S., Holbl,A., Gleixner,K.V., Dworzak,M., Mayerhofer,M., Hoermann,G. *et al.* (2011) High STAT5 levels mediate imatinib resistance and indicate disease progression in chronic myeloid leukemia. *Blood*, **117**, 3409–3420.
15. Wierenga,A.T., Vellenga,E. and Schuringa,J.J. (2008) Maximal STAT5-induced proliferation and self-renewal at intermediate STAT5 activity levels. *Mol. Cell Biol.*, **28**, 6668–6680.
16. Basham,B., Sathe,M., Grein,J., McClanahan,T., D'Andrea,A., Lees,E. and Rascle,A. (2008) In vivo identification of novel STAT5 target genes. *Nucleic Acids Res.*, **36**, 3802–3818.
17. Kimura,A., Martin,C., Robinson,G.W., Simone,J.M., Chen,W., Wickre,M.C., O'Shea,J.J. and Hennighausen,L. (2010) The gene encoding the hematopoietic stem cell regulator CCN3/NOV is under direct cytokine control through the transcription factors STAT5A/B. *J. Biol. Chem.*, **285**, 32704–32709.
18. Qazi,S. and Uckun,F.M. (2010) Gene expression profiles of infant acute lymphoblastic leukaemia and its prognostically distinct subsets. *Br. J. Haematol.*, **149**, 865–873.
19. Barski,A., Cuddapah,S., Cui,K., Roh,T.Y., Schones,D.E., Wang,Z., Wei,G., Chepelev,I. and Zhao,K. (2007) High-resolution profiling of histone methylations in the human genome. *Cell*, **129**, 823–837.
20. Wei,G., Wei,L., Zhu,J., Zang,C., Hu-Li,J., Yao,Z., Cui,K., Kanno,Y., Roh,T.Y., Watford,W.T. *et al.* (2009) Global mapping of H3K4me3 and H3K27me3 reveals specificity and plasticity in lineage fate determination of differentiating CD4+ T cells. *Immunity*, **30**, 155–167.
21. Zhang,Y., Liu,T., Meyer,C.A., Eeckhoute,J., Johnson,D.S., Bernstein,B.E., Nusbaum,C., Myers,R.M., Brown,M., Li,W. *et al.* (2008) Model-based analysis of ChIP-Seq (MACS). *Genome Biol.*, **9**, R137.
22. Salmon-Divon,M., Dvinge,H., Tammoja,K. and Bertone,P. (2010) PeakAnalyzer: genome-wide annotation of chromatin binding and modification loci. *BMC Bioinformatics*, **11**, 415.
23. Grant,C.E., Bailey,T.L. and Noble,W.S. (2011) FIMO: scanning for occurrences of a given motif. *Bioinformatics*, **27**, 1017–1018.
24. El Kasmi,K.C., Smith,A.M., Williams,L., Neale,G., Panopoulos,A.D., Watowich,S.S., Hacker,H., Foxwell,B.M. and Murray,P.J. (2007) Cutting edge: a transcriptional repressor and corepressor induced by the STAT3-regulated anti-inflammatory signaling pathway. *J. Immunol.*, **179**, 7215–7219.
25. Machanick,P. and Bailey,T.L. (2011) MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics*, **27**, 1696–1697.
26. Zhang,Z., Chang,C.W., Goh,W.L., Sung,W.K. and Cheung,E. (2011) CENTDIST: discovery of co-associated factors by motif distribution. *Nucleic Acids Res.*, **39**, W391–W399.
27. Hennighausen,L.G. and Sippel,A.E. (1982) Characterization and cloning of the mRNAs specific for the lactating mouse mammary gland. *Eur. J. Biochem.*, **125**, 131–141.
28. Hennighausen,L.G. and Sippel,A.E. (1982) Mouse whey acidic protein is a novel member of the family of 'four-disulfide core' proteins. *Nucleic Acids Res.*, **10**, 2677–2684.

29. Liu,X., Robinson,G.W., Wagner,K.U., Garrett,L., Wynshaw-Boris,A. and Hennighausen,L. (1997) Stat5a is mandatory for adult mammary gland development and lactogenesis. *Genes Dev.*, **11**, 179–186.
30. Liao,W., Lin,J.X., Wang,L., Li,P. and Leonard,W.J. (2011) Modulation of cytokine receptors by IL-2 broadly regulates differentiation into helper T cell lineages. *Nat. Immunol.*, **12**, 551–559.
31. John,S., Sabo,P.J., Thurman,R.E., Sung,M.H., Biddie,S.C., Johnson,T.A., Hager,G.L. and Stamatoyannopoulos,J.A. (2011) Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat. Genet.*, **43**, 264–268.
32. Wei,L., Vahedi,G., Sun,H.W., Watford,W.T., Takatori,H., Ramos,H.L., Takahashi,H., Liang,J., Gutierrez-Cruz,G., Zang,C. *et al.* (2010) Discrete roles of STAT4 and STAT6 transcription factors in tuning epigenetic modifications and transcription during T helper cell differentiation. *Immunity*, **32**, 840–851.
33. Durant,L., Watford,W.T., Ramos,H.L., Laurence,A., Vahedi,G., Wei,L., Takahashi,H., Sun,H.W., Kanno,Y., Powrie,F. *et al.* (2010) Diverse targets of the transcription factor STAT3 contribute to T cell pathogenicity and homeostasis. *Immunity*, **32**, 605–615.