MCB Accepted Manuscript Posted Online 13 July 2015
Mol. Cell. Biol. doi:10.1128/MCB.00060-15
Copyright © 2015, American Society for Microbiology. All Rights Reserved.

Accepted Manuscript Posted Online
Molecular and Cellular Biology
MCB
Molecular and Cellular Biology

Impact of the N-terminal domain of STAT3 in STAT3-dependent transcriptional activity

Tiancen Hu,[a,b] Jennifer E. Yeh,[c,d] Luca Pinello,[f,g] Jaison Jacob,[a] Srinivas Chakravarthy,[e] Guo-Cheng Yuan,[f,g] Rajiv Chopra,[a]# David A. Frank,[c-h]#

T.H. and J.E.Y. contributed equally to this work.

Center for Proteomic Chemistry[a] and Postdoctoral Program,[b] Novartis Institutes for BioMedical Research, Cambridge, MA; Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA[c]; Biophysical Collaborative Access Team) / Illinois Institute of Technology, Sector 18ID (Advanced Photon Source, Argonne National Laboratory), IL[e]; Departments of Medicine, Harvard Medical School[d] and Brigham and Women's Hospital,[h] Boston, MA; Department of Biostatistics and Computational Biology, Dana-Farber Cancer Institute, Boston, MA[f]; Department of Biostatistics, Harvard School of Public Health, Boston, MA[g]

Running Head: Structure and function of STAT3 N-terminal domain

#Address correspondence to Rajiv Chopra, rajiv.chopra@novartis.com, and David A. Frank, david_frank@dfci.harvard.edu.

Word count for Materials and Methods: 2201

Combined word count for Introduction, Results, and Discussion: 5056

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

Abstract

The transcription factor STAT3 is constitutively active in many cancers, where it mediates important biological effects including cell proliferation, differentiation, survival, and angiogenesis. The N-terminal domain (NTD) of STAT3 performs multiple functions such as cooperative DNA binding, nuclear translocation and protein-protein-interactions. However, it is unclear which subsets of STAT3 target genes depend on the NTD for transcriptional regulation. To identify such genes, we compared gene expression in *STAT3*-null mouse embryonic fibroblasts (MEFs) stably expressing wild-type or NTD-deleted STAT3. NTD deletion reduced cytokine-induced expression of specific STAT3 target genes by decreasing STAT3 binding to their regulatory regions. To better understand potential mechanisms of this effect, we determined the crystal structure of the STAT3 NTD and identified a dimer interface responsible for cooperative DNA binding *in vitro*. We also observed a Ni^{2+}-mediated oligomer with as yet unknown biological function. Mutations on both dimer and Ni^{2+}-mediated interfaces affected cytokine induction of STAT3 target genes. These studies shed light on the role of the NTD in the transcriptional regulation by STAT3 and provide a structural template to design STAT3 NTD inhibitors with potential therapeutic value.

Accepted Manuscript Posted Online
Molecular and Cellular Biology
MCB
Molecular and Cellular Biology

40 **Introduction**

41 Signal transducer and activator of transcription (STAT) proteins are a family of latent
42 transcription factors activated by cytokines and growth factors (1). Mammals contain seven
43 STAT proteins, each highly conserved across species. Of these, STAT3 is a multi-functional
44 member involved in acute-phase response, development, cell growth and differentiation,
45 immunity, hematopoiesis, and tumor survival. Upon stimulation by cytokines such as
46 interleukin-6 (IL-6), STAT3 is phosphorylated by Janus-family kinases (JAKs) at Tyr705,
47 dimerizes and translocates into the nucleus to regulate gene expression. While normal cells
48 display transient physiologic STAT3 activation due to tight regulation by inhibitory molecules
49 (1), many cancer cells depend on constitutive activation of STAT3 for survival; and ectopic
50 expression of STAT3 is sufficient for cell transformation (2-5). Given its necessity and
51 sufficiency for tumorigenesis, STAT3 represents a promising target for cancer therapy (6, 7).

52 STAT3 is composed of an N-terminal domain (NTD) and a "core domain" comprising a
53 coiled-coil domain for protein-protein interactions (PPI), a DNA-binding domain, a linker
54 region, an SH2 domain for dimerization, and a C-terminal domain for transactivation. Structures
55 of the "core domain" have already been determined (8-10). However, the structure of the NTD
56 is yet unknown. Its function can be summarized into three main categories (**Supplemental**
57 **Table ST1**). First, the NTD mediates tetramerization of two Y705-phosphorylated STAT3 (P-
58 STAT3) dimers to cooperatively bind closely-spaced STAT3 sites in gene promoters (11-13).
59 This cooperativity is critical for STAT3 to recognize weaker binding sites, potentially
60 broadening the pool of its target genes. However, the full complement of target genes dependent
61 on STAT3 NTD has not been determined. Second, the NTD mediates dimerization of
62 unphosphorylated STAT3 (U-STAT3) and is essential for its nuclear accumulation (14-16),

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

DNA binding (17), chromatin remodeling (17-19), and regulation of gene expression (18). A point mutation in the NTD (L78R) that disrupts U-STAT3 dimerization has been identified in inflammatory hepatocellular adenoma (IHCA) (20, 21). The NTD is also necessary for U-STAT3 to suppress pro-apoptotic genes, which drives the proliferation and survival of breast cancer cells (22). Finally, STAT3 NTD binds other proteins to form functional complexes in transcriptional regulation and anti-viral response. Many of these interactions require post-translational modifications of the STAT3 NTD (23-29). Besides these three main functions, other properties of STAT3 might also involve the NTD. These include the oligomerization of U-STAT3 in the cytosol (30), direct regulation of STAT3 by metal ions (31-33), and non-genomic functions of STAT3 in microtubule stabilization (34) and mitochondrial metabolism (35).

Despite the extensive studies of STAT3 NTD functions, a systematic understanding of the genes regulated by the STAT3 NTD is lacking, and the atomic structure of the STAT3 NTD remains unknown. In this study, we identified genes regulated by the STAT3 NTD by comparing gene induction in STAT3-null mouse embryonic fibroblasts (MEFs) stably expressing wild-type or NTD mutant STAT3. We also determined the crystal structure of the STAT3 NTD to elucidate functional interfaces. This study provides novel insight into transcriptional regulation by STAT3 and structural hints to design STAT3 NTD inhibitors.

Materials and Methods

Generation of stable cell lines

Wild-type mouse embryonic fibroblasts (MEFs) and STAT3-null MEFs (received from Valeria Poli, University of Turin, Italy) (36, 37) were maintained in Dulbecco's modified Eagle medium (DMEM) containing 10% fetal bovine serum. STAT3-null MEFs were stably

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

transfected with pCMV6-hSTAT3a using Lipofectamine 2000 (Invitrogen, Grand Island, NY) and selected for 2 weeks in 500 µg/mL G418 (Life Technologies, Grand Island, NY). Individual clones were picked with 5 mm cloning disks (Bel-Art, Wayne, NJ). All cells were maintained in a humidified incubator at 37 °C with 5% CO_2.

Cytokine stimulation

MEFs were stimulated with leukemia inhibitory factor (EMD Millipore Corporation, Billerica, MA), 10 ng/mL unless otherwise indicated, for 15 min for protein analyses, 30 min for mRNA analyses, and 15 min for chromatin immunoprecipitation (ChIP) analyses.

RNA isolation for RNA-Seq analysis

Total RNA was isolated from $5x10^5$ cells using TRIzol (Invitrogen) reagent as per the manufacturer's instructions. Quality, quantity, and integrity of total RNA were evaluated using a NanoDrop 8000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA) and a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). Library preparation (ribosomal depletion-RNASeq method) and sequencing were performed by the Dana-Farber Cancer Institute Center for Computational Biology. The cDNA library of good quality was PCR amplified and sequenced on an Illumina HiSeq 2000 System with a paired-end, 50-cycle flow cell.

RNA-Seq data analysis

Quality of reads was checked using FASTQC (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Available reads were aligned to the UCSC *Mus musculus* reference genome (mm9) using Tophat 2 (http://tophat.cbcb.umd.edu/)

Molecular and Cellular Biology

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

and gene expression levels in FPKM quantified using Cufflinks 2 (http://cufWTinks.cbcb.umd.edu/). All alignment statistics are reported in **Supplemental Table ST2**.

Htseq-count (http://www-huber.embl.de/users/anders/HTSeq/doc/count.html) was used to count the number of reads for each transcript followed by analysis using the Bioconductor package edgeR (http://www.bioconductor.org/packages/release/bioc/html/edgeR.html). This package, with default parameters, was used to filter out genes expressed at low levels, with a CPM (count per million) < 3 and to remove batch effects. Differentially expressed genes were calculated using the following groups:

- STAT3-null MEFs + WT unstimulated *vs.* STAT3-null MEFs + WT LIF-stimulated
- STAT3-null MEFs + WT, $Trp^{37}Phe$, or ΔNTD unstimulated *vs.* STAT3-null MEFs + WT, $Trp^{37}Phe$, or ΔNTD LIF-stimulated.

STAT3 binding site analysis

All available murine STAT3 ChIP-seq datasets were downloaded from the Gene Expression Omnibus (GEO): GSM288353, GSM494687, GSM494690, GSM494691, GSM494694, GSM580756, GSM686673. Each gene was checked for the presence of STAT3 peaks in a window of 5 kb centered at the transcription start site using Cistrome Finder (http://cistrome.org/finder/). Each gene was also checked for the presence of a STAT3 motif (MA0144.1) from JASPAR MOTIF database (http://jaspar.genereg.net/) in a window of 5 kb centered at the transcription start site using FIMO (38) software (http://meme.nbcr.net/meme/fimo-intro.html) with default parameters and p-value of 1×10^{-4}.

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

JASPAR contains transcription factor binding profiles derived from experimentally defined binding sites in eukaryotes presented as position-specific scoring matrices.

RT-PCR

Total RNA was extracted using RNeasy Mini kits (Qiagen, Valencia, CA) and reverse transcribed with TaqMan kits (Applied Biosystems, Foster City, CA). Quantitative PCR (qPCR) was performed in triplicate using SYBR select master mix (Applied Biosystems) on a 7300 or 7500 real-time PCR system (Applied Biosystems) with 40 cycles of: 95 °C for 30 sec, 55 °C for 30 sec, and 72 °C for 30 sec. Specificity of amplification was confirmed by melt curve analysis. Cycle threshold (C_T) values for target isoforms were normalized to an endogenous reference gene (HPRT). Primer sequences given in **Supplemental Table ST4** were designed from UCSC genome browser reference transcript sequences using Primer3 software.

Chromatin immunoprecipitation

ChIP was performed as previously described (39). Briefly, cells (1×10^7) were fixed in 1% formaldehyde for 10 min, sonicated using a Fisher Scientific Sonic Dismembranator Model 500 PDQ on setting 15 in 15 sec pulses, and lysates immunoprecipitated overnight at 4 °C with an antibody for STAT3 (sc-482) from Santa Cruz Biotechnology, Inc. (Dallas, TX). Quantitative PCR was performed using primers listed in **Supplemental Table ST4**. Results were expressed as % of input.

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

Immunoblot analyses

Cells were lysed on ice for 15 minutes in EBC lysis buffer [50 mM Tris (pH 8.0), 250 mM NaCl, 0.5% NP40] supplemented with phosphatase and complete protease inhibitors (Roche, Indianapolis, IN). Blots were probed with antibodies to STAT3 (sc-482) from Santa Cruz Biotechnology; phospho-STAT3 (Y705) (9131) from Cell Signaling Technology, Inc. (Danvers, MA); and tubulin (T5168) and actin (A5316) from Sigma-Aldrich Corp. (St. Louis, MO). Nuclear and cytoplasmic fractionations were performed according to the manufacturer's instructions (Active Motif, Carlsbad, CA).

Protein production

Various constructs including different truncations and solubilization tags of human STAT3 NTD (which has the same sequence as mouse) were tested for optimal expression and purification. We found that deleting the first two residues of STAT3 NTD (Met1 and Ala2) dramatically improved the soluble expression of the protein. Thus we cloned residues 3-120, 3-124, 3-126, 3-130, 3-135, and 3-138 into a custom vector derived from pET-series vectors (EMD Millipore) for recombinant expression in *E. coli*. Each insert was preceded by a 6xHis tag followed by a recognition site ("LEVLFQGP") of PreScission protease (GE Healthcare). The plasmids were transformed into BL21 (DE3) (Life Technologies). Single colonies were inoculated into 10 mL of Terrific Broth (TB) medium (Teknova) supplemented with 30 mg/L kanamycin and grown at 37°C overnight. The next day the overnight culture was amplified into 1 L of TB medium with 30 mg/ml kanamycin. The culture was grown at 37 °C for 3-4 hr until OD A600 reached 2.0, induced with isopropyl-beta-D-thiogalactoside (IPTG) to 0.4 mM to induce expression, and grown at 16 °C overnight. The cells were harvested by centrifugation at

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

177 5,000 g for 20 min. The pellet was resuspended in binding buffer (50 mM Tris·HCl (pH 8.0),
178 500 mM NaCl, 1 mM TECP, 10% glycerol) supplemented with complete protease inhibitors
179 (Roche, Indianapolis, IN) and 0.1 mg/mL deoxyribonuclease I from bovine pancreas (Sigma) per
180 50 mL of lysis buffer, then lysed by passing through a Microfluidizer (Microfluidics) two times.
181 The lysate was cleared by centrifugation at 25,000 g for 1 hr. Talon resin (Clontech) was then
182 added to the supernatant for batch binding overnight at 4 °C. The next day the resin was packed
183 into a column, washed with binding buffer plus 20 mM imidazole, and eluted with binding buffer
184 plus 300 mM imidazole. The eluted protein was supplemented with PreScission protease (100
185 μg per 10 mg of target protein) and dialyzed in binding buffer at 4 °C overnight. The mixture
186 was concentrated using Amicon Ultra (EMD Millipore) and loaded onto a HiLoad 16/60
187 Superdex 200 (GE Healthcare) size-exclusion column equilibrated in 20 mM Tris·HCl (pH 7.5),
188 150 mM NaCl, and 1 mM dithiothreitol (DTT). Peak fractions were analyzed by SDS-PAGE,
189 pooled, and concentrated for crystallization.

190 Wild-type STAT3 (mouse β isoform, residues 1-722, UniProt number P42227) was
191 cloned into pET-SUMO vector (Life Technologies) with a PreScission recognition site inserted
192 between the 6xHis-SUMO tag and STAT3. The plasmid was transformed into TKB1 (DE3)
193 (Agilent) or BL21 (DE3) to produce Y705-phosphorylated or non-phosphorylated STAT3,
194 respectively. Protein expression, affinity chromatography, and tag cleavage were the same as in
195 the production of STAT3 NTD. After tag cleavage, the mixture was bound back to 1 mL
196 HisTrap HP column (GE Healthcare) and eluted with an imidazole gradient (0 – 300 mM). Pure
197 full-length STAT3 (OD A260/A280 ratio ~ 0.55) was eluted at about 20 mM imidazole. The
198 protein was then concentrated and loaded onto a HiLoad 16/60 Superdex 200 size-exclusion
199 column equilibrated in 20 mM Tris·HCl (pH 8.0), 500 mM NaCl, 5% glycerol, and 1 mM DTT.

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

Peak fractions were analyzed by SDS-PAGE, pooled, and concentrated for assays. Mutant full-length STAT3 was produced in the same way as wild-type STAT3. The STAT3 core domain (residues 127-722 of mouse STAT3 β isoform) was also produced in the same way, except that the protein mixture after tag cleavage was loaded onto a HiTrap Q HP column (GE Healthcare) instead of the HisTrap column, and eluted with 20 mM – 1M NaCl gradient in 50 mM Tris·HCl pH 8.0, 10% glycerol and 1 mM TCEP. Size exclusion coupled with multi-angle static light scattering (SEC-MALS) of non-phosphorylated full-length STAT3 was performed on a DAWN^{R} HELEOS^{R} II multi-angle static light scattering detector (WYATT) connected to a Superdex 200 10/300 GL size-exclusion column (GE Healthcare). Data were analyzed using Dynamics 7 program (WYATT).

Crystallization and structure determination

All truncations of STAT3 NTD were screened for crystallization, and only residues 3-138 produced crystals. Crystals were grown by hanging drop vapor diffusion. One µL of STAT3 NTD at 8.5 mg/mL was mixed with 1 µL of reservoir solution containing 20% (w/v) PEG3350 and 0.2 M magnesium formate. The drop was immediately streaked with a needle touched to the small crystals grown from the same condition in the screening plate, and equilibrated against 500 µL of the same reservoir solution at 4 °C. The crystal grew to a size of 200 µm x 50 µm x 50 µm in 3 days. The crystals were then transferred to reservoir solution containing additional 20% ethylene glycol and flash cooled in liquid nitrogen. Diffraction data were collected at beamline 17-ID at the Advanced Photon Source (Argonne National Laboratory, USA), and processed and scaled using XDS (40). The structure of the STAT3 NTD was solved by molecular replacement using Phaser (41) with the structure of the STAT4 NTD (PDB ID: 1BGF) as a search model.

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

The final model was built in COOT (42) and refined with Phenix (43), CNS (44) and Buster (Global Phasing, LTD) (45). Statistics of the data and model are summarized in **Table 1**. The structure has been submitted to the Protein Data Bank (PDB ID: 4ZIA).

Size-Exclusion Chromatography – Small Angle X-ray Scattering (SEC-SAXS)

SEC-SAXS experiments were performed at BioCAT (beamline 18-ID, Advanced Photon Source at Argonne National Labs) (46). The set-up included a focused 12 KeV (1.03 Å) x-ray beam, a 1.5 mm quartz capillary sample cell, a sample to detector distance of ~2.5 m, and a Mar165 CCD detector. The q-range sampled was ~ 0.0065 – 0.3 $Å^{-1}$. In order to ensure sample monodispersity, we used an in-line SEC setup, which included an AKTA-pure FPLC unit and a Superdex-200 10/300 GL column (GE Healthcare Life Sciences). The column outlet was directly connected to the SAXS sample cell. One-second exposures were collected every 5 sec during the gel-filtration chromatography run. Exposures before and after the elution of the sample were averaged and used as the buffer curve, and the exposures during elution (co-incident with the UV peak on the chromatogram) were treated as protein+buffer curves. Data were corrected for background scattering by subtracting the buffer curve from protein+buffer curves. Data from the frame corresponding to the UV peak with S greater than 0.2 $Å^{-1}$ were analyzed using PRIMUS, GNOM, DAMMIF, DAMMIN, CRYSOL, DAMAVER and SUPCOMB of the ATSAS package (47) (**Supplemental Figure S4**). Molecular envelopes were generated by averaging 20 bead models generated by DAMMIF, with DAMAVER to use as a starting model for DAMMIN, applying 2-fold symmetry. The radii of gyration, R_g, determined by Guinier analysis or by GNOM were similar (21.20 ± 0.09 Å and 21.35 ± 0.09 Å respectively).

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

The maximum particle diameter D_m, as determined by PRIMUS, was 74.7Å. The crystal structure models were fitted to the envelope using SUPCOMB.

Electrophoretic mobility shift assay (EMSA)

DNA probes with 5-TAMRA™ (Azide) modifications at the 5' ends were purchased from Integrated DNA Technologies (DNA probe sequences listed in the corresponding figures). The protein-DNA mixture (10 µL) for EMSA consisted of 100 nM of DNA probe, STAT3 protein at the indicated concentrations, 20 mM Tris·HCl (pH 7.5), 150 mM NaCl, 1 mM DTT, 1 mg/mL bovine serum albumin (BSA, Roche, Indianapolis, IN), and 50 µg/mL salmon sperm DNA (R&D Systems). The sample was incubated on ice for 30 min, while 10-well DNA Retardation gels (6% polyacrylamide gel, Life Technologies) were pre-run in 0.5X TBE buffer at 200 V for 30 min at 4 °C. The sample was then added with 2 µL of 50% (v/v) glycerol to 10%, loaded onto the gel, and run in 0.5X TBE at 200 V for 45 min at 4 °C. The gel was scanned on a Typhoon 9410 fluorescence scanner at the TAMRA™ excitation and emission wavelengths.

Fluorescence polarization

Samples for fluorescence polarization (20 µL for 384-well plate) were prepared by mixing 2 nM of DNA probe (as described for EMSA) and STAT3 protein at a series of concentrations (serial dilution from 500 nM (STAT3 ΔNTD) or 150 nM (STAT3 FL and its mutants) by 1.5 fold for 23 iterations) in PBS supplemented with 1 mM DTT, 1 mg/mL BSA (Roche, Indianapolis, IN), and 50 µg/mL salmon sperm DNA (R&D Systems). The fluorescence polarization signal was recorded on an EnVision Multilabel plate reader (PerkinElmer) equipped with Optimized Bodipy TMR FP Dual Emission Label. Data were fit using GraphPad Prism

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

with the following "log(agonist) vs. response -- Variable slope" equation to generate K_D and Hill slope:

$$Y = Bottom + (Top-Bottom) / (1+10^{(LogKD-X) \cdot HillSlope})$$

Statistical analyses

Two-tailed T tests for paired samples were performed with Graphpad Prism 6 Software (La Jolla, CA). Data are presented as means ± SD for the indicated number of independent experiments (**Figure 5B**, **Supplemental Figure S1C**) or means ± SEM for one representative replicate (**Figure 2A&B, Figure 3C**, **Figure 4B**, **Figure 12B**).

Results

Identification of STAT3 NTD-dependent target genes

To identify transcripts whose expression is regulated by STAT3 NTD, we performed ribosomal depletion RNA-Seq on STAT3-null mouse embryonic fibroblasts (MEFs) stably expressing wild-type (WT) human STAT3 or a deletion mutant of the entire N-terminal domain (residues 1-126; ΔNTD), which were then stimulated with leukemia inhibitory factor (LIF) to induce STAT3 activation. Single clones expressing WT and ΔNTD STAT3 were chosen for having similar STAT3 protein levels as the levels of endogenous STAT3 in wild-type MEFs, and for showing comparable STAT3 phosphorylation in response to LIF stimulation (**Figure 1**). Stimulation conditions were optimized in wild-type MEFs and STAT3-null MEFs stably expressing WT STAT3. Relative mRNA levels of known STAT3 target genes, including *EGR1*, *SOCS3*, *JUNB*, *KLF4*, and *STAT3* (which positively regulates its own expression), were used to choose an induction time point of 30 minutes and a LIF concentration of 10 ng/mL (**Figure 2**).

Molecular and Cellular Biology

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

STAT3 mRNA is detectable in this system because these STAT3-null MEFs were generated via partial rather than complete deletion of STAT3 that produces a frame-shifted mRNA unable to encode functional protein (36, 37). These time point and cytokine concentration conditions were used for RNA-Seq analyses.

We initially focused on the 100 genes most upregulated by LIF in STAT3-null MEFs expressing WT STAT3 as detected by RNA-Seq. Many of these genes are known transcriptional targets of STAT3 (including *EGR1*, *STAT3*, *JUNB*, and *IER3* (48-51)), confirming that LIF induced STAT3 transcriptional activity in these cells. These genes are likely to be STAT3-dependent since LIF does not induce the transcription of known STAT3-regulated genes in STAT3-null MEFs (**Supplemental Figure S1**). Among the top 100 LIF-induced genes, the smallest gene induction observed was 1.48-fold for *PACS1* and the largest gene induction was 4.34-fold for *EGR1* (**Supplemental Table ST3**). Significantly, 83 of these 100 genes showed at least a twenty percent decrease in induction in MEFs carrying ΔNTD compared to WT STAT3 (**Figure 3A&B left**), indicating that they were likely directly regulated by STAT3 and that the NTD was responsible for important functional effects on gene induction. However, we considered the possibility that expression of ΔNTD STAT3 might decrease gene expression through a non-specific effect. Thus, we repeated this analysis for 100 genes selected from those that did not show any significant change with LIF stimulation in STAT3-null MEFs expressing WT STAT3. In cells expressing ΔNTD STAT3, only 12 of these genes showed a decrease of at least 20% compared to cells expressing WT STAT3 whereas 76 showed no change in induction and 12 showed an increase of at least 20% (**Figure 3A&B right**), indicating that ΔNTD STAT3 did not affect gene expression non-specifically. From the 20 most highly LIF-induced genes, we

Molecular and Cellular Biology

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

selected 4 for independent validation with qRT-PCR, and confirmed reduced induction of all of these genes with STAT3 containing deletion of the NTD (**Figure 3C**).

We next considered the hypothesis that the NTD might play an especially critical role in STAT3 transcriptional activity when Y705-phosphorylated STAT3 (P-STAT3) is present in lower amounts by helping it bind to weak sites through cooperativity. To test this hypothesis, we first transiently expressed WT or ΔNTD STAT3 in STAT3-null MEFs to achieve comparable levels of total STAT3 and then stimulated the cells with a range of LIF concentrations to determine the effect of NTD deletion on STAT3 tyrosine phosphorylation at low cytokine levels (**Figure 4A**). We selected a concentration of LIF (0.5 ng/mL) for mRNA analyses because this low dose induced observable PY-STAT3 levels in the MEFs. We then examined eight genes, three previously validated to be dependent on the STAT3 NTD by qRT-PCR (*EGR1*, *FOSB*, *ERRFI1*) and five others that had independent evidence of STAT3-dependency. In cells with ΔNTD STAT3, gene expression was not increased upon LIF induction but was in fact decreased in all eight of these genes (**Figure 4B**), suggesting ΔNTD STAT3 may act as a dominant inhibitory form under these low cytokine stimulation conditions. These analyses clearly indicate a role for the NTD in the expression of a subset of STAT3-dependent genes.

NTD deletion reduces STAT3 recruitment to target genes

To elucidate the mechanism by which NTD deletion reduces LIF-induced transcription of STAT3 target genes, we investigated STAT3 binding to NTD-dependent genes using chromatin immunoprecipitation (ChIP). If the NTD is essential for STAT3 cooperative DNA binding, then we predicted that genes whose regulatory regions have tandem STAT3 binding motifs might show less STAT3 DNA binding with the ΔNTD form of STAT3 compared to WT STAT3 (11,

Molecular and Cellular Biology

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

336 13). To identify candidates among the 100 most LIF-upregulated genes with regulatory regions
337 containing tandem STAT3 binding sites, we utilized published STAT3 ChIP-Seq data sets with
338 experimental evidence of STAT3 binding and also a sequence-based algorithm (52) to predict
339 STAT3 binding sites. Using both approaches, we searched the promoters of NTD-dependent
340 genes for the presence of at least two proximal STAT3 consensus motifs ($TTCN_3GAA$) and
341 identified 6 candidate genes (**Figure 5A**). NTD deletion reduced STAT3 recruitment to sites in
342 all of these genes (**Figure 5B**), directly correlating with the observed decreases in mRNA
343 transcript levels (**Figure 5C**) and supporting a role for the NTD in modulating STAT3
344 transcriptional activity at the level of DNA binding. The lack of full correlation between
345 reduction in STAT3 DNA binding measured with ChIP and gene expression measured with
346 RNA-Seq seen at this single time point is likely due to the rapid and transient nature of both
347 STAT3 chromatin recruitment and STAT3 target gene expression in response to LIF (**Figure**
348 **2A&B**, **Supplemental Figure S2**).
349 Based on recent evidence of the role of single-site cooperativity in activity of the related
350 transcription factor STAT1 (53), we considered the possibility that STAT3 NTD-mediated
351 cooperativity may not be restricted to genes with tandem binding sites. To test this hypothesis,
352 we analyzed the top 100 LIF-upregulated genes to identify genes that possess only a single
353 predicted STAT3 binding motif within a window 5 kb upstream and downstream of the
354 transcription start site. Of these 100 genes, only 4 had a single STAT site, which suggests that
355 binding to tandem sites may be generally important for STAT3-dependent gene regulation. We
356 then used ChIP to evaluate LIF-induced STAT3 DNA binding to these predicted motifs. Of the
357 4 genes with a single predicted STAT3 motif, only 2 (*SNORD87* and *ZFP184*) showed LIF-
358 induced STAT3 DNA binding in STAT3-null MEFs stably expressing WT STAT3. For these

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

two genes, NTD deletion did not significantly reduce LIF-induced STAT3 binding to the predicted motif (**Supplemental Figure S3**). Although this reflects the findings from only two genes, these results are consistent with the hypothesis that NTD-mediated cooperative binding of STAT3 occurs only with STAT3 binding to tandem sites.

Given the contribution of the NTD to STAT3 binding to tandem sites, we considered the hypothesis that NTD-dependent genes have more STAT3 binding motifs than NTD-independent genes. We compared the total number of STAT3 binding motifs in the regulatory regions of NTD-dependent vs. NTD-independent genes in a window 5 kb upstream and downstream of the transcription start site (TSS). We did not find any significant difference in the number of binding motifs, either upstream or downstream of the transcription start site, between NTD-dependent and NTD-independent genes (**Supplemental Figure S4**), suggesting that the spacing or relative binding strength of STAT3 sites, rather than the total number of sites, might be the key determinant of cooperativity.

Crystal structure of the STAT3 NTD

To better understand the structure-function relationship of the STAT3 NTD, we determined its crystal structure. The overall structure of the STAT3 NTD is similar to that of STAT1 (54) and STAT4 NTD (55, 56). Eight α-helices fold into a triangle "hook" (**Figure 6A**), starting with a ring of four small helices (α1-α4), linked by a short α5 to a coiled-coil (α6 and α7), and ending with a long helix (α8) perpendicular to the coiled-coil. Five copies of the STAT3 NTD constitute the asymmetric unit of the crystal (**Figure 7A**). There are two significant protein-protein interfaces: a "handshake" dimer interface and a Ni^{2+}-mediated tetramer interface (**Figure 6B**).

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

The "handshake" dimer buries ~960 Å^2 of surface area per monomer (~13% of the total surface area) and is formed by the tip of the coiled-coil (Val^{77} and Leu^{78}) inserting into the four-helix ring of another monomer (**Figure 6C**). The resulting interface is further stabilized by multiple hydrogen bonds (H-bonds). Across the three copies of "handshake" dimers in the asymmetric unit (one of them formed with a monomer from the neighboring asymmetric unit), the Val^{77}/Leu^{78}-ring interaction remains the same while several of the H-bonds are broken in some copies (*e.g.* Glu^{16}-Arg^{70}, His^{19}-His^{81}, not shown here), consistent with the essential role of Val^{77}/Leu^{78} in the dimerization of STAT NTDs (54, 57-61).

The Ni^{2+} interface is formed by four coiled-coil segments (from four monomers) centering a metal ion, designated Ni^{2+} (**Figure 6D**), since STAT3 NTD is purified by Ni^{2+}-NTA and Ni^{2+} shifts the STAT3 NTD from dimer to higher oligomer in size exclusion chromatography (**Supplemental Figure S5**). Although the Ni^{2+} is likely introduced in the purification process, it may reflect a physiologically relevant interaction at this site. The Ni^{2+} ion sits on a crystallographic 2-fold symmetry axis perpendicular to the non-crystallographic symmetry (NCS) 2-fold axis and is coordinated in a square planar geometry by four histidine residues (His^{58}). Therefore, the Ni^{2+} links four "handshake" NTD dimers into an octamer (**Figure 6D**). This interface buries ~730 Å^2 of surface area per monomer (~10% of the total surface area) and has a network of H-bonds between coiled-coil helices (**Figure 6D**). The Trp^{37} residue, previously reported to interfere with STAT tetramerization and cooperative DNA-binding (11, 55, 62, 63), happens to lie at the center of this interface and forms an H-bond to Glu^{63}. Interestingly, the X-ray scattering data and *ab initio* envelopes indicate good agreement with both "handshake" dimer and half of the Ni^{2+}-tetramer interface (chi values 0.77 and 1.79,

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

respectively; NSD values after SUPCOMB alignment 0.91 and 1.00, respectively) (**Supplemental Figure S6**).

To further assess the functional importance of both interfaces, we performed sequence alignment of the NTD for all STAT family members. The "handshake" dimerization interface of the STAT3 NTD, which has been shown to mediate tetramerization of STAT proteins (54, 57-61), is structurally conserved in STAT1 and STAT4 (**Figure 8A**). The "Ni^{2+}" interface also appears in the STAT4 NTD structure, with Trp^{37} in the middle of the interface. The Ni^{2+}-coordinating His^{58} in STAT3 overlays with Gln^{58} in the STAT4 interface, with the side-chain pointing in a similar orientation (**Figure 8A**).

We then compared the STAT3 NTD across species. Cross-species alignment of STAT3 NTD showed 90% sequence identity from zebrafish to human, precluding the functional evaluation of individual residues. STAT family alignment showed that the "handshake" interface is more conserved than the "Ni^{2+}" interface (**Figure 8B**), especially at the N-terminal four-helical ring region (**Figure 8C**). Several regions of interest are apparent in the sequence alignment: 1) In the Val^{77}/Leu^{78} tip of the "handshake" interface, Leu^{78} is completely conserved, and residues at the Val^{77} position are always hydrophobic; 2) In the N-terminal helical ring holding Val^{77}/Leu^{78}, both the hydrophobicity of the cavity (e.g. Trp^{4}, Leu^{18}, Leu^{15}, Met^{28}, Phe^{33}) and the ability to form H-bonds with the Val^{77}/Leu^{78} backbone (*e.g.* Gln^{8}, His^{19}, Gln^{32}) are conserved; 3) Trp^{37} in the middle of the Ni^{2+} interface is completely conserved, its H-bond acceptor Glu^{63} is Glu/Gln/His in the alignment; and 4) His^{58} coordinating the Ni^{2+} is only conserved in STAT1, STAT3 and maybe STAT2 (a Phe residue is at this position, but a His residue is next to it).

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

Role of STAT3 NTD in cooperative DNA binding in vitro

Although NTD interactions are thought to mediate cooperative binding of STAT3 to tandem sites in gene promoters (11-13), there has not been sufficient evidence of cooperativity using purified, full-length Y705-phosphorylated STAT3 (P-STAT3), probably due to difficulty generating this protein. We purified recombinant full-length P-STAT3 (mouse β isoform) with the help of an N-terminal SUMO fusion (cleaved subsequently) and co-expression with Elk1 tyrosine phosphatase (TKB1 (DE3) strain that has been used to produce P-STAT3 core domain (8)). The Tyr^{705} phosphorylation and dimerization of P-STAT3 were confirmed by intact mass LC-MS, peptide mapping, and size exclusion chromatography combined with multi-angle light scattering (SEC-MALS) (**Supplemental Figure S7**).

We then investigated STAT3 NTD-dependent cooperative DNA binding using electrophoretic mobility shift assay (EMSA) with a 39 bp dsDNA probe from the well-characterized α2-macrogobulin (α2M) promoter, which contains one "weak" and one "strong" STAT3 binding site (11, 12), each predicted to bind to a dimer of P-STAT3. We evaluated STAT3 cooperativity under conditions when STAT3 protein levels were greater than or less than that of the α2M probe. When STAT3 protein levels were less than that of the DNA probe, the majority of WT P-STAT3 was bound to both sites and migrated as a tetramer suggesting a cooperative interaction between DNA-bound P-STAT3 dimers (**Figure 9A**). Under the same binding conditions, ΔNTD P-STAT3 migrated mainly as a dimer and only as a tetramer when STAT3 protein levels were increased. This implies that cooperativity of P-STAT3 on tandem promoter sites is mediated by the NTD. We further confirmed that these complexes represented dimers and tetramers, respectively, by analytical size exclusion experiments (**Supplemental Table ST5**). This approach also excluded the possibility that STAT3 tetramers were forming on

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

450 a single strong site, as has been suggested for STAT1. To further understand the effect of the
451 NTD on cooperative DNA binding, we investigated the effect on cooperativity of mutations in
452 the "handshake" dimer interface of the STAT3 NTD domain. STAT3 with point mutations in
453 this interface (Val^{77}Ala or Leu^{78}Ala) did not form DNA-bound tetramers on the α2M probe until
454 levels of STAT3 protein were increased to stoichiometric excess (**Figure 9A**).

455 To assess the importance of the order of STAT3 binding sites on tetramer formation, we
456 swapped the positions of the "strong" and "weak" sites in the α2M probe and found that this did
457 not impact the cooperativity effect (**Figure 9B**). In fact, when a DNA probe with two "weak"
458 sites was used, the cooperativity effect appeared to be enhanced. Conversely, when two "strong"
459 sites were used, the cooperativity effect was slightly less evident. As a control, we used a DNA
460 probe containing a single "strong" site, which did not show DNA-bound tetramers on EMSA.

461 To quantify the cooperativity observed with EMSA experiments, we performed
462 fluorescence polarization (FP) assays in which the output is a Hill coefficient that describes the
463 degree of cooperativity. A Hill coefficient of 1 indicates completely independent binding
464 whereas values greater than 1 indicate positive cooperativity. Using the same α2M probe from
465 EMSA analyses, WT STAT3 showed cooperative DNA binding with a Hill coefficient of 2.1
466 while ΔNTD abolished cooperativity resulting in a Hill coefficient of 1.1 (**Figure 10A**). The
467 binding affinity of WT STAT3 for the DNA probe was also slightly stronger than that of ΔNTD
468 STAT3 (K_D = 8.2 nM and 25.2 nM, respectively). Point mutations in the "handshake" interface
469 also disrupted cooperativity, with Hill coefficients ~1, and decreased binding affinity (K_D) to a
470 similar extent as did the NTD deletion mutation. In contrast, a negative control mutation
471 Lys^{140}Met (64) and a mutation in the Ni^{2+}-interface (His^{58}Ala) which is not involved in
472 tetramerization had no effect on the Hill coefficient or K_D. It has been reported that STAT3

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

473 cooperativity is promoter-specific, as it is not necessary for transcription of all genes with
474 multiple STAT3 motifs. For example, the *SOCS3* promoter has tandem sites but does not require
475 STAT3 tetramerization on its promoter for transcription (65). Consistent with this, FP assays
476 performed using a probe from the *SOCS3* promoter showed a negligible effect of NTD deletion
477 on the Hill coefficient and K_D compared to WT STAT3 (**Figure 10A**). In addition, NTD
478 deletion did not abrogate LIF-induced transcription of *SOCS3* in our RNA-Seq experiment
479 (**Figure 10B**). Although LIF only slightly induced *SOCS3* mRNA at the time point used for
480 RNA-Seq, analysis by RT-PCR of *SOCS3* mRNA over a time course revealed that the
481 expression of *SOCS3* is robustly induced yet tightly regulated over a relatively short timeframe,
482 consistent with its biological function in negative feedback regulation of STAT3 signaling
483 (**Figure 2A&B**).
484
485 *Evaluation of the NTD interfaces in transcriptional activity of STAT3*
486 To understand the relevance of the observed NTD interfaces in STAT3 transcriptional
487 activity, we investigated the effect of point mutations in these interfaces on LIF-induced STAT3
488 transcriptional activity. Using information from our determination of the STAT3 NTD crystal
489 structure, we evaluated $Val^{77}Ala/Leu^{78}Ala$, a double point mutation in the NTD "handshake"
490 dimer interface, and $Trp^{37}Phe$, a point mutation in the Ni^{2+}-interface. The $Trp^{37}Phe$ point
491 mutation was chosen for its conservation in the structure of the STAT4 NTD and because it was
492 previously reported to play a role in STAT tetramerization and cooperative DNA binding (11,
493 55, 62, 63). We stably expressed these two point mutants of the NTD in STAT3-null MEFs and
494 performed RNA-Seq.

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

495 We compared the effect of these interface mutations and ΔNTD on induction of the top
496 100 LIF-upregulated genes in MEFs expressing WT STAT3 (**Figure 11A**). This global
497 induction profile analysis showed that NTD deletion reduced the induction of the largest number
498 of these genes (82%), followed closely by the Ni^{2+}-interface mutant $Trp^{37}Phe$ (76%), while the
499 "handshake" interface mutant $Val^{77}Ala/Leu^{78}Ala$ reduced a significantly smaller number of
500 genes (44%) (**Figure 11B**). Next we evaluated the overlap in STAT3 target genes showing
501 reduced induction with these different NTD mutations. Thirty-six genes were in common for all
502 3 mutations, with the deletion and $Trp^{37}Phe$ mutants showing extensive overlap (67 genes in
503 common) (**Figure 11C**). Although the $Val^{77}Ala/Leu^{78}Ala$ mutation resulted in a relatively
504 smaller proportion of genes with decreased induction compared to WT, the majority of those
505 genes (38 of 44) that did show reduction were found in common with the NTD deletion mutant.
506 Interestingly, the "handshake" interface mutant also enhanced a significant portion (17%) of
507 STAT3 regulated genes, suggesting a role for this surface in STAT3 mediated gene suppression.
508 Decreased STAT3 target gene induction was confirmed in a transient expression system
509 with low concentrations of LIF. We first confirmed that a low concentration of LIF (0.5 ng/mL)
510 was sufficient to induce STAT3 tyrosine phosphorylation in this system (**Figure 12A**), and then
511 examined 6 LIF-upregulated genes from RNA-Seq that were previously validated by qRT-PCR
512 or had independent evidence of STAT3-dependency. All of these genes showed suppressed
513 induction of mRNA expression in MEFs expressing NTD-mutant compared to WT STAT3
514 (**Figure 12B**).
515 Given that STAT3 can also mediate gene repression, we examined the top 100 genes
516 from RNA-Seq whose expression was downregulated following LIF treatment in STAT3-null
517 MEFs expressing WT STAT3. LIF repressed the mRNA levels of these 100 genes by 30-55%.

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

We then evaluated the effect of NTD mutations on the fold change of mRNA expression (LIF vs. unstimulated) of these genes. Both point and deletion mutations in the NTD relieved the suppression of a majority of these LIF-downregulated genes (**Figure 13**). These findings indicate that the STAT3 NTD may be equally important in mediating gene repression as well as induction.

Collectively, these studies show that deletion and point mutations in the STAT3 NTD known to disrupt cooperative DNA binding alter STAT3 regulation of its target genes, including genes whose expression is normally either upregulated or downregulated by STAT3. In our studies, the W37F point mutant was more similar to the NTD deletion mutant in the proportion of LIF-upregulated genes affected, whereas the V77A/L78A point mutant reduced the induction of a smaller percentage of LIF-upregulated genes. These data identify genes that are regulated by the NTD and specific protein-interacting surfaces and will enable gene-specific mechanisms of regulation to be elucidated.

Discussion

STAT3 is one of the most multi-faceted members of the STAT transcription factor family, involved in diverse biological processes from development to immune response (1). Importantly, it represents a promising target for anti-cancer therapy since it drives both the proliferation and immune evasion of tumor cells (66-70) and is required by more than a dozen types of human cancers for survival (2, 4). STAT3 performs the majority of its functions through protein-protein-interactions, many of which depend on the NTD (**Supplemental Table ST1**). Selective inhibition of the NTD has anti-tumor efficacy *in vivo* (7, 18, 71), further implicating the role of this domain in STAT3-driven oncogenesis. Despite these promising data

Molecular and Cellular Biology

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

541 supporting the STAT3 NTD as a therapeutic target in cancer, a comprehensive list of NTD-
542 dependent genes was lacking and the structure of the NTD remained unknown.

543 To identify STAT3-regulated genes dependent on the NTD, we stably expressed STAT3
544 constructs with or without the NTD in STAT3-null MEFs, selected clones with approximately
545 equal levels of total STAT3 and LIF-induced P-STAT3 protein levels, and used RNA-Seq to
546 investigate genome-wide changes in gene transcription. Although the ΔNTD mutant had similar
547 levels of STAT3 phosphorylation in response to LIF activation, it decreased induction of many
548 STAT3-regulated genes. The dependence of these genes on the NTD was further confirmed
549 using transient expression systems, in which WT and ΔNTD STAT3 protein levels were
550 stringently controlled to rule out potential differences arising during generation of the stable cell
551 lines. We found that the effect of NTD deletion was particularly prominent under low cytokine
552 concentrations when there are lower amounts of activated STAT3, consistent with a role of the
553 NTD in facilitating STAT3 binding to "weak" STAT3 sites (11-13). Furthermore, deletion of
554 the NTD reduced STAT3 occupancy at target genes with tandem STAT3 binding motifs, where
555 cooperative binding to DNA is thought to be most important. Together, these data indicate that
556 the NTD is necessary for maximal transcription of a subset of STAT3 target genes.

557 At a single time point, there was not always a full correlation between the reduction in
558 STAT3 DNA binding and STAT3 target gene expression upon NTD deletion. These findings
559 likely reflect the tightly-regulated kinetics of LIF-induced STAT3 activation, and thus can better
560 be appreciated in examining both of these processes over time (**Figure 2A&B**, **Supplemental**
561 **Figure S2**). The NTD could modulate the effect of LIF stimulation on STAT3 function via
562 several mechanisms including nucleo-cytoplasmic shuttling, chromatin remodeling or recruiting
563 transcriptional proteins that affect the basal levels of STAT3 DNA binding or gene expression.

Molecular and Cellular Biology

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

564 A subset of these affected genes may be regulated by cooperative STAT3 binding to tandem
565 DNA motifs. Alternative mechanisms are also possible. For example, it was recently reported
566 that the STAT1 NTD mediates single-site cooperativity (53), which has been attributed to
567 stronger protein-protein interactions within the STAT1 complex than between STAT1 and its
568 target gene regulatory regions. The STAT3 NTD shares a high degree of structural homology
569 with the STAT1 NTD (18), including the fact that amino acid 77 in the "handshake" interface
570 playing a critical role in tetramer formation. However, analytical size exclusion analyses found
571 that STAT3 tetramers did not form on single binding sites, only tandem sites (Supplemental
572 Table ST5). Furthermore, when we examined the effect of NTD deletion on STAT3 recruitment
573 to target genes with single STAT3 binding sites, we did not observe a significant reduction of
574 STAT3 DNA binding upon NTD deletion. Although our analysis was limited by the few genes
575 with single STAT3 sites among the top 100 LIF-upregulated genes, these findings suggest that
576 there may be differences in NTD-dependent interactions among STATs that contribute to
577 differential effects on cooperativity and transcriptional activity. Indeed, a small-molecule
578 inhibitor of the STAT3 NTD selectively binds the STAT3 NTD, but not the highly homologous
579 STAT1 NTD (18).

580 A global sequence analysis of the promoter regions of STAT3 NTD-dependent genes did
581 not reveal a significant dependence on the total number of STAT3 motifs, raising the possibility
582 that the STAT3 NTD can regulate STAT3 target gene expression via mechanisms such as
583 binding to other transcription factors. Indeed, the STAT3 NTD is known to modulate the ability
584 of STAT3 to form transcriptional enhanceosomes with other proteins (23-29). For example, the
585 α2M promoter has a canonical binding site for the transcription factor c-Jun and mutations in the
586 NTD reduce the interaction between STAT3 and c-Jun, thus preventing maximal cytokine-

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

587 induced transcription (11). The STAT3 NTD also enables STAT3 binding to the pioneer factor
588 FoxP3, which induces epigenetic modifications that increase the access of STAT3 to adjacent
589 gene promoters (23). Thus, the STAT3 NTD confers broader gene-specific effects on the
590 function of STAT3 via discreet gene-specific mechanisms involving protein-protein interactions.
591 In addition to its importance in gene induction, we found that the NTD also plays a role
592 in STAT3 gene repression, consistent with previous findings (18). Indeed, a pharmacological
593 inhibitor of the STAT3 NTD, ST3-H2A2, activated the expression of 147 genes normally
594 repressed by STAT3 and selectively induced the apoptotic death of cancer cells (18). Thus,
595 interactions mediated by the NTD can contribute to both increases and decreases in gene
596 expression mediated by STAT3.
597 In order to characterize, at a molecular level, how the NTD affects gene regulation, we
598 determined the crystal structure of STAT3 NTD, which revealed two interfaces critical for
599 oligomerization. The "handshake" dimer interface, believed to be the functional unit of the
600 NTD, is structurally similar to that of the STAT1 and STAT4 NTDs (54-56) and is mediated by
601 interactions between Val^{77}/Leu^{78} and the N-terminal four-helix ring (**Figure 6C**). The second
602 oligomerization in the STAT3 NTD crystal structure was formed around a Ni^{2+} ion linking four
603 "handshake" dimers into an octamer. Although the Ni^{2+} in the structure was likely introduced in
604 the purification process through the use of Ni^{2+}-NTA, it may reflect a physiologically important
605 mechanism. STAT3 forms higher order oligomers, such as para-crystals and nuclear bodies (72,
606 73), to serve as active reservoirs resistant to dephosphorylation. Ni^{2+} has been reported to
607 activate inflammatory transcription factors like NF-κB (74), while metal ions have been shown
608 to directly regulate STAT3 (31-33), another established pro-inflammation factor (75). Thus this
609 Ni^{2+}-oligomer hints at a possible role of STAT3 in Ni^{2+}-mediated inflammation. We further

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

610 showed that the STAT3 NTD mediates cooperativity via the conserved "handshake" dimer
611 interface (**Figure 8B&C**) rather than the "Ni^{2+}" interface, consistent with reported mutagenesis
612 data for STAT family members (13, 20, 54, 56-61). Data suggest a structural model of two
613 STAT3 dimers on tandem DNA sites, "holding hands" by NTD dimerization, in a *syn* geometric
614 arrangement (**Figure 7C**), which agrees with the near-one-turn inter-site spacing optimal for
615 cooperativity (11).

616 The crystal structure of STAT3 NTD may be used to provide novel drug design concepts.
617 For example, the helical ring encompassing Val^{77}/Leu^{78} (**Figure 6C**) is a good candidate small
618 molecule pocket. Indeed, a peptide mimetic of the α2 helix in the "handshake" interface induced
619 apoptosis of breast and prostate cancer cells but not normal cells (7, 18, 71). A somatic mutation
620 in this interface ($Leu^{78}Arg$) has been found in inflammatory hepatocellular adenoma, where it
621 disrupts homotypic interactions between unphosphorylated STAT3 dimers (20). This suggests
622 that targeting the NTD "handshake" interface may selectively inhibit the expression of a subset
623 of genes normally regulated by unphosphorylated STAT3. We compared the effect of mutations
624 on these NTD surfaces with an NTD-deleted protein on those genes most affected by LIF
625 induction. We found that the "Ni^{2+}" interface ($Trp^{37}Phe$) more closely reproduced the NTD
626 deletion mutant effect on LIF-upregulated genes than the "handshake" interface ($V^{77}A/L^{78}A$)
627 mutant. Importantly, our data show that the NTD interaction surfaces that mediate gene
628 regulation may have numerous, perhaps concomitant, functional roles and any putative
629 cooperative DNA binding may be difficult to characterize at the cellular level.

630 STAT3 cooperative binding to DNA has been proposed as an NTD-mediated regulatory
631 mechanism (11, 12, 65, 76). To fully characterize this effect *in vitro*, we performed EMSA and
632 fluorescence polarization assays on full-length purified P-STAT3 with DNA probes containing

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

633 tandem STAT3 motifs. Cooperative binding only occurred in the presence of the NTD and was
634 most pronounced with adjacent "weak" STAT3 motifs. Besides STAT3, STAT5 can also form
635 tetramers through its NTD. As with STAT3-regulated genes, some STAT5 target genes such as
636 *IL-2Rα* have pairs of "weak" binding sites in their promoter that require STAT5 cooperative
637 binding for transcriptional activation (77).
638
639 In summary, we identified STAT3 target genes that require the NTD for optimal
640 expression, which may include genes affected by cooperative DNA binding. We also
641 determined the crystal structure of the STAT3 NTD, and proposed the key residues in NTD
642 interfaces important for cooperative DNA binding, formation of higher order oligomers, and
643 perhaps other protein-protein interactions. These results provide insight into the mechanism of
644 NTD-mediated STAT3 functions, and may serve as a structural template for inhibitor design.
645

646 **Acknowledgements**

647 This work was supported by grants from the National Cancer Institute (R01-CA160979,
648 T32GM007226, F30 CA180340-01) and the Brent Leahey Fund. The authors thank Sarah
649 Walker and Kirk Clark for helpful discussions, the staff at IMCA-CAT for support and to
650 Stephen Harrison for his scientific mentorship.
651 This research used resources of the Advanced Photon Source, a U.S. Department of
652 Energy (DOE) Office of Science User Facility operated for the DOE Office of Science by
653 Argonne National Laboratory under Contract No. DE-AC02-06CH11357. BioCAT is supported
654 by grant P41 GM103622 from the National Institute of General Medical Sciences of the National
655 Institutes of Health.

Accepted Manuscript Posted Online
Molecular and Cellular Biology
MCB
Molecular and Cellular Biology

656 **References**

657 1. **Stark GR, Darnell JE, Jr.** 2012. The JAK-STAT pathway at twenty. Immunity **36:**503-514.

658
659 2. **Bromberg J.** 2002. Stat proteins and oncogenesis. J Clin Invest **109:**1139-1142.

660
661 3. **Bromberg JF, Wrzeszczynska MH, Devgan G, Zhao Y, Pestell RG, Albanese C, Darnell JE, Jr.** 1999.
662 Stat3 as an oncogene. Cell **98:**295-303.

663
664 4. **Buettner R, Mora LB, Jove R.** 2002. Activated STAT signaling in human tumors provides novel
665 molecular targets for therapeutic intervention. Clin Cancer Res **8:**945-954.

666
667 5. **Levy DE, Inghirami G.** 2006. STAT3: a multifaceted oncogene. Proc Natl Acad Sci U S A
668 **103:**10151-10152.

669
670 6. **Sansone P, Bromberg J.** 2012. Targeting the interleukin-6/Jak/stat pathway in human
671 malignancies. J Clin Oncol **30:**1005-1014.

672
673 7. **Timofeeva OA, Gaponenko V, Lockett SJ, Tarasov SG, Jiang S, Michejda CJ, Perantoni AO,**
674 **Tarasova NI.** 2007. Rationally designed inhibitors identify STAT3 N-domain as a promising
675 anticancer drug target. ACS Chem Biol **2:**799-809.

676
677 8. **Becker S, Groner B, Muller CW.** 1998. Three-dimensional structure of the Stat3beta homodimer
678 bound to DNA. Nature **394:**145-151.

679
680 9. **Nkansah E, Shah R, Collie GW, Parkinson GN, Palmer J, Rahman KM, Bui TT, Drake AF, Husby J,**
681 **Neidle S, Zinzalla G, Thurston DE, Wilderspin AF.** 2013. Observation of unphosphorylated STAT3
682 core protein binding to target dsDNA by PEMSA and X-ray crystallography. FEBS Lett **587:**833-
683 839.

684
685 10. **Ren Z, Mao X, Mertens C, Krishnaraj R, Qin J, Mandal PK, Romanowski MJ, McMurray JS, Chen**
686 **X.** 2008. Crystal structure of unphosphorylated STAT3 core fragment. Biochem Biophys Res
687 Commun **374:**1-5.

688
689 11. **Zhang X, Darnell JE, Jr.** 2001. Functional importance of Stat3 tetramerization in activation of the
690 alpha 2-macroglobulin gene. J Biol Chem **276:**33576-33581.

691
692 12. **Lerner L, Henriksen MA, Zhang X, Darnell JE, Jr.** 2003. STAT3-dependent enhanceosome
693 assembly and disassembly: synergy with GR for full transcriptional increase of the alpha 2-
694 macroglobulin gene. Genes Dev **17:**2564-2577.

Accepted Manuscript Posted Online
Molecular and Cellular Biology
MCB
Molecular and Cellular Biology

695
696 13. **Lin JX, Li P, Liu D, Jin HT, He J, Ata Ur Rasheed M, Rochman Y, Wang L, Cui K, Liu C, Kelsall BL,**
697 **Ahmed R, Leonard WJ.** 2012. Critical Role of STAT5 transcription factor tetramerization for
698 cytokine responses and normal immune function. Immunity **36:**586-599.

699
700 14. **Cimica V, Chen HC, Iyer JK, Reich NC.** 2011. Dynamics of the STAT3 transcription factor: nuclear
701 import dependent on Ran and importin-beta1. PLoS One **6:**e20188.

702
703 15. **Pranada AL, Metz S, Herrmann A, Heinrich PC, Muller-Newen G.** 2004. Real time analysis of
704 STAT3 nucleocytoplasmic shuttling. J Biol Chem **279:**15114-15123.

705
706 16. **Vogt M, Domoszlai T, Kleshchanok D, Lehmann S, Schmitt A, Poli V, Richtering W, Muller-**
707 **Newen G.** 2011. The role of the N-terminal domain in dimerization and nucleocytoplasmic
708 shuttling of latent STAT3. J Cell Sci **124:**900-909.

709
710 17. **Timofeeva OA, Chasovskikh S, Lonskaya I, Tarasova NI, Khavrutskii L, Tarasov SG, Zhang X,**
711 **Korostyshevskiy VR, Cheema A, Zhang L, Dakshanamurthy S, Brown ML, Dritschilo A.** 2012.
712 Mechanisms of unphosphorylated STAT3 transcription factor binding to DNA. J Biol Chem
713 **287:**14192-14200.

714
715 18. **Timofeeva OA, Tarasova NI, Zhang X, Chasovskikh S, Cheema AK, Wang H, Brown ML,**
716 **Dritschilo A.** 2013. STAT3 suppresses transcription of proapoptotic genes in cancer cells with the
717 involvement of its N-terminal domain. Proc Natl Acad Sci U S A **110:**1267-1272.

718
719 19. **Zhao Y, Zeng C, Tarasova NI, Chasovskikh S, Dritschilo A, Timofeeva OA.** 2013. A new role for
720 STAT3 as a regulator of chromatin topology. Transcription **4:**227-231.

721
722 20. **Domoszlai T, Martincuks A, Fahrenkamp D, Schmitz-Van de Leur H, Kuster A, Muller-Newen G.**
723 2014. Consequences of the disease-related L78R mutation for dimerization and activity of STAT3.
724 J Cell Sci **127:**1899-1910.

725
726 21. **Pilati C, Amessou M, Bihl MP, Balabaud C, Nhieu JT, Paradis V, Nault JC, Izard T, Bioulac-Sage P,**
727 **Couchy G, Poussin K, Zucman-Rossi J.** 2011. Somatic mutations activating STAT3 in human
728 inflammatory hepatocellular adenomas. J Exp Med **208:**1359-1366.

729
730 22. **Primiano T, Baig M, Maliyekkel A, Chang BD, Fellars S, Sadhu J, Axenovich SA, Holzmayer TA,**
731 **Roninson IB.** 2003. Identification of potential anticancer drug targets through the selection of
732 growth-inhibitory genetic suppressor elements. Cancer Cell **4:**41-53.

733

Accepted Manuscript Posted Online
Molecular and Cellular Biology
MCB
Molecular and Cellular Biology

734 23. **Hossain DM, Panda AK, Manna A, Mohanty S, Bhattacharjee P, Bhattacharyya S, Saha T,**
735 **Chakraborty S, Kar RK, Das T, Chatterjee S, Sa G.** 2013. FoxP3 acts as a cotranscription factor
736 with STAT3 in tumor-induced regulatory T cells. Immunity **39:**1057-1069.

737
738 24. **Hou T, Ray S, Brasier AR.** 2007. The functional role of an interleukin 6-inducible CDK9.STAT3
739 complex in human gamma-fibrinogen gene expression. The Journal of biological chemistry
740 **282:**37091-37102.

741
742 25. **Ray S, Boldogh I, Brasier AR.** 2005. STAT3 NH2-terminal acetylation is activated by the hepatic
743 acute-phase response and required for IL-6 induction of angiotensinogen. Gastroenterology
744 **129:**1616-1632.

745
746 26. **Ray S, Lee C, Hou T, Bhakat KK, Brasier AR.** 2010. Regulation of signal transducer and activator
747 of transcription 3 enhanceosome formation by apurinic/apyrimidinic endonuclease 1 in hepatic
748 acute phase response. Molecular endocrinology **24:**391-401.

749
750 27. **Ray S, Lee C, Hou T, Boldogh I, Brasier AR.** 2008. Requirement of histone deacetylase1 (HDAC1)
751 in signal transducer and activator of transcription 3 (STAT3) nucleocytoplasmic distribution.
752 Nucleic acids research **36:**4510-4520.

753
754 28. **Ray S, Zhao Y, Jamaluddin M, Edeh CB, Lee C, Brasier AR.** 2014. Inducible STAT3 NH2 terminal
755 mono-ubiquitination promotes BRD4 complex formation to regulate apoptosis. Cellular
756 signalling **26:**1445-1455.

757
758 29. **Wang WB, Levy DE, Lee CK.** 2011. STAT3 negatively regulates type I IFN-mediated antiviral
759 response. Journal of immunology **187:**2578-2585.

760
761 30. **Ndubuisi MI, Guo GG, Fried VA, Etlinger JD, Sehgal PB.** 1999. Cellular physiology of STAT3:
762 Where's the cytoplasmic monomer? J Biol Chem **274:**25499-25509.

763
764 31. **Chua AC, Klopcic BR, Ho DS, Fu SK, Forrest CH, Croft KD, Olynyk JK, Lawrance IC, Trinder D.**
765 2013. Dietary iron enhances colonic inflammation and IL-6/IL-11-Stat3 signaling promoting
766 colonic tumor development in mice. PLoS One **8:**e78850.

767
768 32. **Novak U, Ji H, Kanagasundaram V, Simpson R, Paradiso L.** 1998. STAT3 forms stable
769 homodimers in the presence of divalent cations prior to activation. Biochem Biophys Res
770 Commun **247:**558-563.

771

Accepted Manuscript Posted Online
Molecular and Cellular Biology
MCB
Molecular and Cellular Biology

772 33. **Kitabayashi C, Fukada T, Kanamoto M, Ohashi W, Hojyo S, Atsumi T, Ueda N, Azuma I, Hirota H,**
773 **Murakami M, Hirano T.** 2010. Zinc suppresses Th17 development via inhibition of STAT3
774 activation. Int Immunol **22:**375-386.

775
776 34. **Ng DC, Lin BH, Lim CP, Huang G, Zhang T, Poli V, Cao X.** 2006. Stat3 regulates microtubules by
777 antagonizing the depolymerization activity of stathmin. J Cell Biol **172:**245-257.

778
779 35. **Wegrzyn J, Potla R, Chwae YJ, Sepuri NB, Zhang Q, Koeck T, Derecka M, Szczepanek K, Szelag**
780 **M, Gornicka A, Moh A, Moghaddas S, Chen Q, Bobbili S, Cichy J, Dulak J, Baker DP, Wolfman A,**
781 **Stuehr D, Hassan MO, Fu XY, Avadhani N, Drake JI, Fawcett P, Lesnefsky EJ, Larner AC.** 2009.
782 Function of mitochondrial Stat3 in cellular respiration. Science **323:**793-797.

783
784 36. **Costa-Pereira AP, Tininini S, Strobl B, Alonzi T, Schlaak JF, Is'harc H, Gesualdo I, Newman SJ,**
785 **Kerr IM, Poli V.** 2002. Mutational switch of an IL-6 response to an interferon-gamma-like
786 response. Proceedings of the National Academy of Sciences of the United States of America
787 **99:**8043-8047.

788
789 37. **Alonzi T, Maritano D, Gorgoni B, Rizzuto G, Libert C, Poli V.** 2001. Essential role of STAT3 in the
790 control of the acute-phase response as revealed by inducible gene inactivation [correction of
791 activation] in the liver. Molecular and cellular biology **21:**1621-1632.

792
793 38. **Grant CE, Bailey TL, Noble WS.** 2011. FIMO: scanning for occurrences of a given motif.
794 Bioinformatics **27:**1017-1018.

795
796 39. **Nelson EA, Walker SR, Alvarez JV, Frank DA.** 2004. Isolation of unique STAT5 targets by
797 chromatin immunoprecipitation-based gene identification. The Journal of biological chemistry
798 **279:**54724-54730.

799
800 40. **Karplus PA, Diederichs K.** 2012. Linking crystallographic model and data quality. Science
801 **336:**1030-1033.

802
803 41. **McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ.** 2007. Phaser
804 crystallographic software. J Appl Crystallogr **40:**658-674.

805
806 42. **Emsley P, Lohkamp B, Scott WG, Cowtan K.** 2010. Features and development of Coot. Acta
807 Crystallogr D Biol Crystallogr **66:**486-501.

808
809 43. **Adams PD, Afonine PV, Bunkoczi G, Chen VB, Davis IW, Echols N, Headd JJ, Hung LW, Kapral GJ,**
810 **Grosse-Kunstleve RW, McCoy AJ, Moriarty NW, Oeffner R, Read RJ, Richardson DC, Richardson**

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

811 **JS, Terwilliger TC, Zwart PH.** 2010. PHENIX: a comprehensive Python-based system for
812 macromolecular structure solution. Acta Crystallogr D Biol Crystallogr **66:**213-221.

813
814 44. **Brunger AT.** 2007. Version 1.2 of the Crystallography and NMR system. Nat Protoc **2:**2728-2733.

815
816 45. **Blanc E, Roversi P, Vonrhein C, Flensburg C, Lea SM, Bricogne G.** 2004. Refinement of severely
817 incomplete structures with maximum likelihood in BUSTER-TNT. Acta Crystallogr D Biol
818 Crystallogr **60:**2210-2221.

819
820 46. **Mathew E, Mirza A, Menhart N.** 2004. Liquid-chromatography-coupled SAXS for accurate sizing
821 of aggregating proteins. Journal of synchrotron radiation **11:**314-318.

822
823 47. **Konarev P, Petoukhov M, Volkov V, Svergun D.** 2006. ATSAS 2.1, a program package for small-
824 angle scattering data analysis. J. Appl. Crystallogr. **39:**277-286.

825
826 48. **Bourillot PY, Aksoy I, Schreiber V, Wianny F, Schulz H, Hummel O, Hubner N, Savatier P.** 2009.
827 Novel STAT3 target genes exert distinct roles in the inhibition of mesoderm and endoderm
828 differentiation in cooperation with Nanog. Stem Cells **27:**1760-1771.

829
830 49. **Carpenter RL, Lo HW.** 2014. STAT3 Target Genes Relevant to Human Cancers. Cancers **6:**897-925.

831
832 50. **Snyder M, Huang XY, Zhang JJ.** 2008. Identification of novel direct Stat3 target genes for control
833 of growth and differentiation. The Journal of biological chemistry **283:**3791-3798.

834
835 51. **Alvarez JV, Febbo PG, Ramaswamy S, Loda M, Richardson A, Frank DA.** 2005. Identification of a
836 genetic signature of activated signal transducer and activator of transcription 3 in human tumors.
837 Cancer Res **65:**5054-5062.

838
839 52. **Sandelin A, Alkema W, Engstrom P, Wasserman WW, Lenhard B.** 2004. JASPAR: an open-access
840 database for eukaryotic transcription factor binding profiles. Nucleic acids research **32:**D91-94.

841
842 53. **Begitt A, Droescher M, Meyer T, Schmid CD, Baker M, Antunes F, Knobeloch KP, Owen MR,**
843 **Naumann R, Decker T, Vinkemeier U.** 2014. STAT1-cooperative DNA binding distinguishes type
844 1 from type 2 interferon signaling. Nature immunology **15:**168-176.

845
846 54. **Mao X, Ren Z, Parker GN, Sondermann H, Pastorello MA, Wang W, McMurray JS, Demeler B,**
847 **Darnell JE, Jr., Chen X.** 2005. Structural bases of unphosphorylated STAT1 association and
848 receptor binding. Mol Cell **17:**761-771.

849

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

850 55. **Vinkemeier U, Moarefi I, Darnell JE, Jr., Kuriyan J.** 1998. Structure of the amino-terminal
851 protein interaction domain of STAT-4. Science **279:**1048-1052.

852
853 56. **Chen X, Bhandari R, Vinkemeier U, Van Den Akker F, Darnell JE, Jr., Kuriyan J.** 2003. A
854 reinterpretation of the dimerization interface of the N-terminal domains of STATs. Protein Sci
855 **12:**361-365.

856
857 57. **Hou Z, Srivastava S, Mistry MJ, Herbst MP, Bailey JP, Horseman ND.** 2003. Two tandemly linked
858 interferon-gamma-activated sequence elements in the promoter of glycosylation-dependent cell
859 adhesion molecule 1 gene synergistically respond to prolactin in mouse mammary epithelial
860 cells. Mol Endocrinol **17:**1910-1920.

861
862 58. **Meyer T, Hendry L, Begitt A, John S, Vinkemeier U.** 2004. A single residue modulates tyrosine
863 dephosphorylation, oligomerization, and nuclear accumulation of stat transcription factors. J
864 Biol Chem **279:**18998-19007.

865
866 59. **Ota N, Brett TJ, Murphy TL, Fremont DH, Murphy KM.** 2004. N-domain-dependent
867 nonphosphorylated STAT4 dimers required for cytokine-driven activation. Nat Immunol **5:**208-
868 215.

869
870 60. **Mertens C, Zhong M, Krishnaraj R, Zou W, Chen X, Darnell JE, Jr.** 2006. Dephosphorylation of
871 phosphotyrosine on STAT1 dimers requires extensive spatial reorientation of the monomers
872 facilitated by the N-terminal domain. Genes Dev **20:**3372-3381.

873
874 61. **Wenta N, Strauss H, Meyer S, Vinkemeier U.** 2008. Tyrosine phosphorylation regulates the
875 partitioning of STAT1 between different dimer conformations. Proc Natl Acad Sci U S A
876 **105:**9238-9243.

877
878 62. **John S, Vinkemeier U, Soldaini E, Darnell JE, Jr., Leonard WJ.** 1999. The significance of
879 tetramerization in promoter recruitment by Stat5. Mol Cell Biol **19:**1910-1918.

880
881 63. **Murphy TL, Geissal ED, Farrar JD, Murphy KM.** 2000. Role of the Stat4 N domain in receptor
882 proximal tyrosine phosphorylation. Mol Cell Biol **20:**7121-7131.

883
884 64. **Yang J, Huang J, Dasgupta M, Sears N, Miyagi M, Wang B, Chance MR, Chen X, Du Y, Wang Y,**
885 **An L, Wang Q, Lu T, Zhang X, Wang Z, Stark GR.** 2010. Reversible methylation of promoter-
886 bound STAT3 by histone-modifying enzymes. Proc Natl Acad Sci U S A **107:**21499-21504.

887

Molecular and Cellular Biology

Accepted Manuscript Posted Online
Molecular and Cellular Biology
MCB
Molecular and Cellular Biology

888 65. **Zhang L, Badgwell DB, Bevers JJ, 3rd, Schlessinger K, Murray PJ, Levy DE, Watowich SS.** 2006.
889 IL-6 signaling via the STAT3/SOCS3 pathway: functional analysis of the conserved STAT3 N-
890 domain. Mol Cell Biochem **288:**179-189.

891
892 66. **Brantley EC, Benveniste EN.** 2008. Signal transducer and activator of transcription-3: a
893 molecular hub for signaling pathways in gliomas. Mol Cancer Res **6:**675-684.

894
895 67. **Hussain SF, Kong LY, Jordan J, Conrad C, Madden T, Fokt I, Priebe W, Heimberger AB.** 2007. A
896 novel small molecule inhibitor of signal transducers and activators of transcription 3 reverses
897 immune tolerance in malignant glioma patients. Cancer Res **67:**9630-9636.

898
899 68. **Kasprzycka M, Marzec M, Liu X, Zhang Q, Wasik MA.** 2006. Nucleophosmin/anaplastic
900 lymphoma kinase (NPM/ALK) oncoprotein induces the T regulatory cell phenotype by activating
901 STAT3. Proc Natl Acad Sci U S A **103:**9964-9969.

902
903 69. **Kortylewski M, Kujawski M, Wang T, Wei S, Zhang S, Pilon-Thomas S, Niu G, Kay H, Mule J,**
904 **Kerr WG, Jove R, Pardoll D, Yu H.** 2005. Inhibiting Stat3 signaling in the hematopoietic system
905 elicits multicomponent antitumor immunity. Nat Med **11:**1314-1321.

906
907 70. **Yu H, Kortylewski M, Pardoll D.** 2007. Crosstalk between cancer and immune cells: role of
908 STAT3 in the tumour microenvironment. Nat Rev Immunol **7:**41-51.

909
910 71. **Timofeeva OA, Tarasova NI.** 2012. Alternative ways of modulating JAK-STAT pathway: Looking
911 beyond phosphorylation. JAKSTAT **1:**274-284.

912
913 72. **Droescher M, Begitt A, Marg A, Zacharias M, Vinkemeier U.** 2011. Cytokine-induced
914 paracrystals prolong the activity of signal transducers and activators of transcription (STAT) and
915 provide a model for the regulation of protein solubility by small ubiquitin-like modifier (SUMO). J
916 Biol Chem **286:**18731-18746.

917
918 73. **Herrmann A, Sommer U, Pranada AL, Giese B, Kuster A, Haan S, Becker W, Heinrich PC, Muller-**
919 **Newen G.** 2004. STAT3 is enriched in nuclear bodies. J Cell Sci **117:**339-349.

920
921 74. **Freitas M, Fernandes E.** 2011. Zinc, cadmium and nickel increase the activation of NF-kappaB
922 and the release of cytokines from THP-1 monocytic cells. Metallomics **3:**1238-1243.

923
924 75. **Atsumi T, Singh R, Sabharwal L, Bando H, Meng J, Arima Y, Yamada M, Harada M, Jiang JJ,**
925 **Kamimura D, Ogura H, Hirano T, Murakami M.** 2014. Inflammation amplifier, a new paradigm in
926 cancer biology. Cancer research **74:**8-14.

927

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

928 76. **Bode JG, Fischer R, Haussinger D, Graeve L, Heinrich PC, Schaper F.** 2001. The inhibitory effect
929 of IL-1 beta on IL-6-induced alpha 2-macroglobulin expression is due to activation of NF-kappa B.
930 J Immunol **167:**1469-1481.

931

932 77. **Meyer WK, Reichenbach P, Schindler U, Soldaini E, Nabholz M.** 1997. Interaction of STAT5
933 dimers on two low affinity binding sites mediates interleukin 2 (IL-2) stimulation of IL-2 receptor
934 alpha gene transcription. J Biol Chem **272:**31821-31828.

935

936

937

Molecular and Cellular Biology

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

938 **Figure legends**

939 **Figure 1.** Characterization of STAT3-null mouse embryonic fibroblasts (MEFs) stably
940 expressing wild-type (WT) or NTD-mutant STAT3.
941 STAT3-null MEFs stably expressing WT, W37F, NTD-deleted (ΔNTD), and V77A/L78A
942 STAT3 were stimulated with LIF and analyzed by immunoblot. (-), negative control for STAT3
943 expression from parental STAT3-null MEFs. Tubulin serves as a loading control.
944

945 **Figure 2.** Optimization of time point and cytokine concentration for mRNA and ChIP analyses.
946 Wild-type MEFs and STAT3-null MEFs stably expressing wild-type (WT) STAT3 were
947 stimulated with **(A)** LIF (10 ng/mL) for a range of time points and **(B)** different concentrations of
948 LIF for 30 min then analyzed by qRT-PCR for expression of the indicated STAT3 target genes.
949 Data normalized to HPRT and then to mRNA expression in unstimulated cells.
950

951 **Figure 3.** NTD deletion reduces induction of LIF-upregulated genes.
952 **(A)** Induction of the top 100 LIF-upregulated genes and 100 non-LIF-regulated genes in STAT3-
953 null MEFs stably expressing wild-type (WT) or ΔNTD STAT3. **(B)** Percent of genes in (A) with
954 significantly altered induction (20% threshold) in ΔNTD relative to WT STAT3. **(C)** qRT-PCR
955 validation of select LIF-upregulated genes in STAT3-null MEFs stably expressing WT or ΔNTD
956 STAT3 (normalized to HPRT).
957

958 **Figure 4.** NTD deletion reduces induction of STAT3 target genes at low concentrations of LIF.
959 **(A)** STAT3-null MEFs transiently transfected with wild-type (WT) or ΔNTD STAT3 (0.2
960 μg/mL) for 24 h then stimulated with the indicated concentrations of LIF for 15 min were

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

961 analyzed by immunoblot (actin serves as a loading control). **(B)** STAT3-null MEFs transiently
962 transfected with WT or ΔNTD STAT3 then stimulated with LIF (0.5 ng/mL) for 30 min were
963 analyzed by qRT-PCR for expression of the indicated STAT3 target genes (normalized to HPRT;
964 representative of N = 3).
965
966 **Figure 5.** NTD deletion reduces STAT3 DNA binding to target genes.
967 **(A)** Regulatory regions of LIF-induced genes containing tandem STAT3 binding motifs (Bold
968 blue underline, "strong" sites; Non-bold red underline, "weak" sites). Chromosome locations
969 given based on mm9 assembly. **(B)** STAT3-null MEFs stably expressing wild-type (WT) or
970 NTD-deleted (ΔNTD) STAT3 were stimulated with LIF then analyzed by chromatin
971 immunoprecipitation with an antibody for STAT3 followed by qRT-PCR using primers flanking
972 the STAT3 binding sites indicated in (A). Data expressed as fold change of % input in LIF-
973 stimulated vs. unstimulated cells (N = 3). **(C)** RNA-Seq transcript levels in STAT3-null MEFs
974 stably expressing WT or ΔNTD STAT3. Data expressed as fold change of mRNA expression in
975 LIF-stimulated vs. unstimulated cells.
976
977 **Figure 6.** Crystal structure of the STAT3 NTD.
978 **(A)** Overall structure of the STAT3 NTD monomer in two view-angles. **(B)** Two interfaces are
979 observed in the crystal structure: a "handshake" dimer interface and a Ni^{2+}-mediated tetramer
980 interface. **(C)** The two molecules of the "handshake" dimer are related by a 2-fold non-
981 crystallographic symmetry (NCS) axis. V77 and L78 dock into the opposing molecule in a
982 cavity created mainly by the three N-terminal helices. Multiple hydrogen-bonds also form on the
983 dimer interface. **(D)** Another 2-fold NCS is observed in the crystal which involves multiple

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

hydrogen bonds between two long helices antiparallel to each other from two NTD molecules. A Ni^{2+} ion sits in the middle of the axis and is coordinated by four H58 residues, thus linking four "handshake" STAT3 dimers into an octamer. Also shown are the C-terminus of each NTD which links to the STAT3 core domain.

Figure 7. Model of cooperative binding between two STAT3 dimers.
(A) Asymmetric unit of STAT3 NTD crystal contains five copies of the molecules. Along with one copy from the neighboring unit, they form three "handshake" dimers with similar organization as shown from the superimposed image in **(B)**. **(C)** A model of two STAT3 dimers cooperatively binding to a tandem-site DNA with the help of NTD dimerization on each side.

Figure 8. Structural comparison and sequence alignment of the STAT3 NTD.
(A) STAT1 (PDB ID: 1YVL), STAT3, and STAT4 (1BGF) share similar "handshake" dimerization interfaces of their NTDs. The Ni^{2+}-interface observed in STAT3 is similar to a second dimer interface observed in the crystal structure of the STAT4 NTD, both of which contain W37 in the middle of the interface. The H58 that coordinates Ni^{2+} in STAT3 corresponds to an Asn residue in STAT4. **(B)** Sequence alignment of NTD across STAT proteins. **(C)** The STAT3 NTD surface colored by conservation scores calculated by ConSurf server based on the sequence alignment in (B). The "handshake" interface is more conserved than the Ni^{2+} interface.

Figure 9. NTD mutations disrupt STAT3 cooperative DNA binding *in vitro*.

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

1006 **(A)** Electrophoretic mobility shift assay (EMSA) of purified pY705 STAT3 (P-STAT3) binding
1007 to an α2M DNA probe containing two STAT3-binding sites (5' –
1008 AGCAGTAACTGGAAAGTCCTTAATCCTTCTGGGAATTCT - 3', STAT3 binding sites
1009 underlined). The 5' site is a "weak" binding site while the 3' site is a "strong" binding site. **(B)**
1010 EMSA of P-STAT3, NTD-deleted (ΔNTD) or wild-type (WT), on various DNA probes derived
1011 from the α2M promoter (sequences listed).
1012
1013 **Figure 10. (A)** Fluorescence polarization assay of wild-type (WT) and NTD-mutant P-STAT3
1014 binding to the indicated DNA probes. **(B)** Induction of SOCS3 mRNA levels (LIF vs.
1015 unstimulated) in STAT3-null MEFs stably expressing wild-type (WT) or ΔNTD STAT3 from
1016 RNA-Seq.
1017
1018 **Figure 11.** NTD interface mutations reduce induction of STAT3 target genes.
1019 **(A)** Induction of the top 100 LIF-upregulated genes in STAT3-null MEFs stably expressing
1020 wild-type (WT) or NTD-mutant STAT3. **(B)** Percent of genes with significantly altered
1021 induction (20% threshold) in NTD mutants relative to WT STAT3. **(C)** Overlap of genes
1022 affected by NTD point vs. deletion mutation.
1023
1024 **Figure 12.** NTD interface mutations reduce induction of LIF-upregulated genes at low
1025 concentrations of LIF.
1026 STAT3-null MEFs transiently transfected with WT or NTD-mutant STAT3 (0.2 µg/mL) for 24 h
1027 then stimulated with **(A)** LIF (0.5 ng/mL) were analyzed by immunoblot (tubulin serves as a

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

1028 loading control) and with **(B)** the indicated concentrations of LIF were analyzed by qRT-PCR for

1029 STAT3 target gene expression (normalized to HPRT; representative of N = 3).

1030

1031 **Figure 13.** Fold change of the top 100 LIF-downregulated genes in STAT3-null MEFs stably

1032 expressing wild-type (WT) or NTD-mutant STAT3.

Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

1033 **Table 1**. Crystallography statistics.

Data collection Statistics	
Space group	$P4_122$
Unit cell	
a, b, c (Å)	109.01, 109.01, 154.30
α, β, γ (°)	90, 90, 90
Wavelength (Å)	1.0
Resolution range (Å)	20 – 2.7
Completeness (%)	100.0 (99.6)
R_{sym} (%)	7.5 (38.8)
$\langle I \rangle/\sigma$	19.4 (5.5)
Redundancy	8.0 (8.2)
Wison B factor ($Å^2$)	54.0
Refinement Statistics	
Number of reflections	
Working set	24803
Test set	1329
Number of atoms	5149
Rmsd Bonds (Å)	0.009
Rmsd angles (°)	1.1
R_{work} (%)	23.5
R_{free} (%)	27.2
B factor ($Å^2$)	59.6
Ramachandran plot	
Most favored (%)	96.8
Additionally allowed (%)	3.2
Disallowed (%)	0

1034

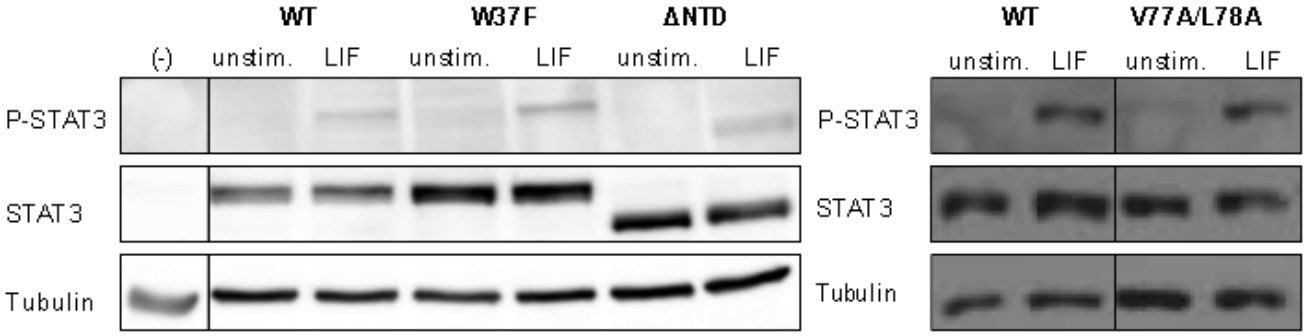
Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

Figure 1



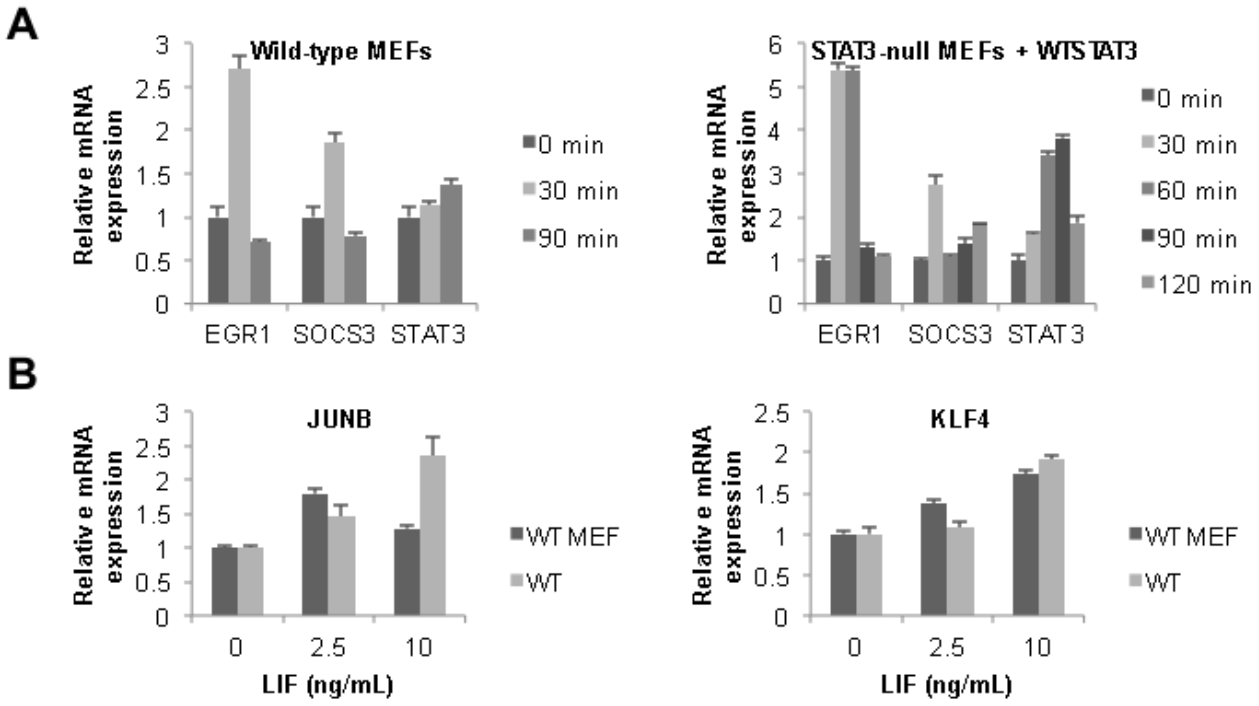
Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

Figure 2



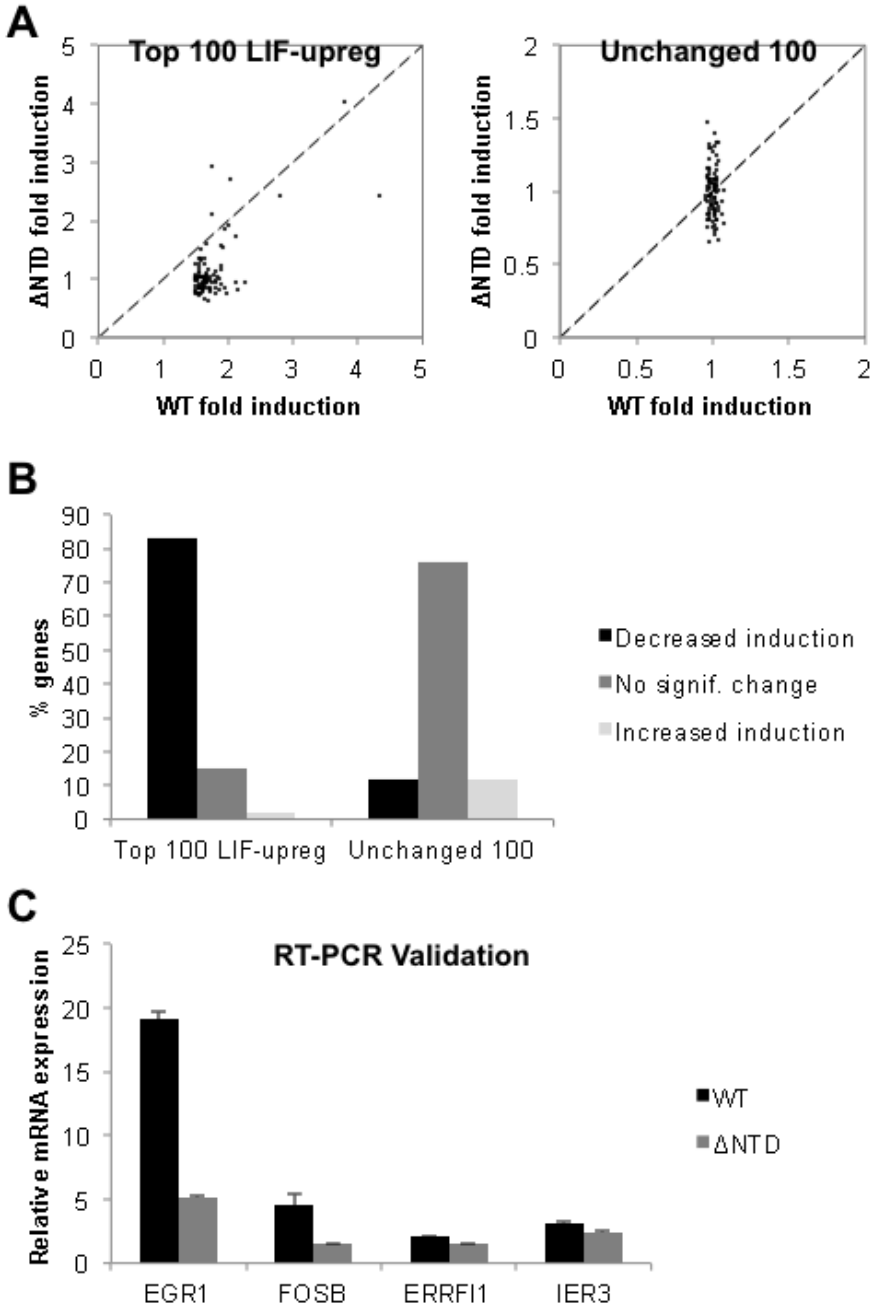
Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

Figure 3



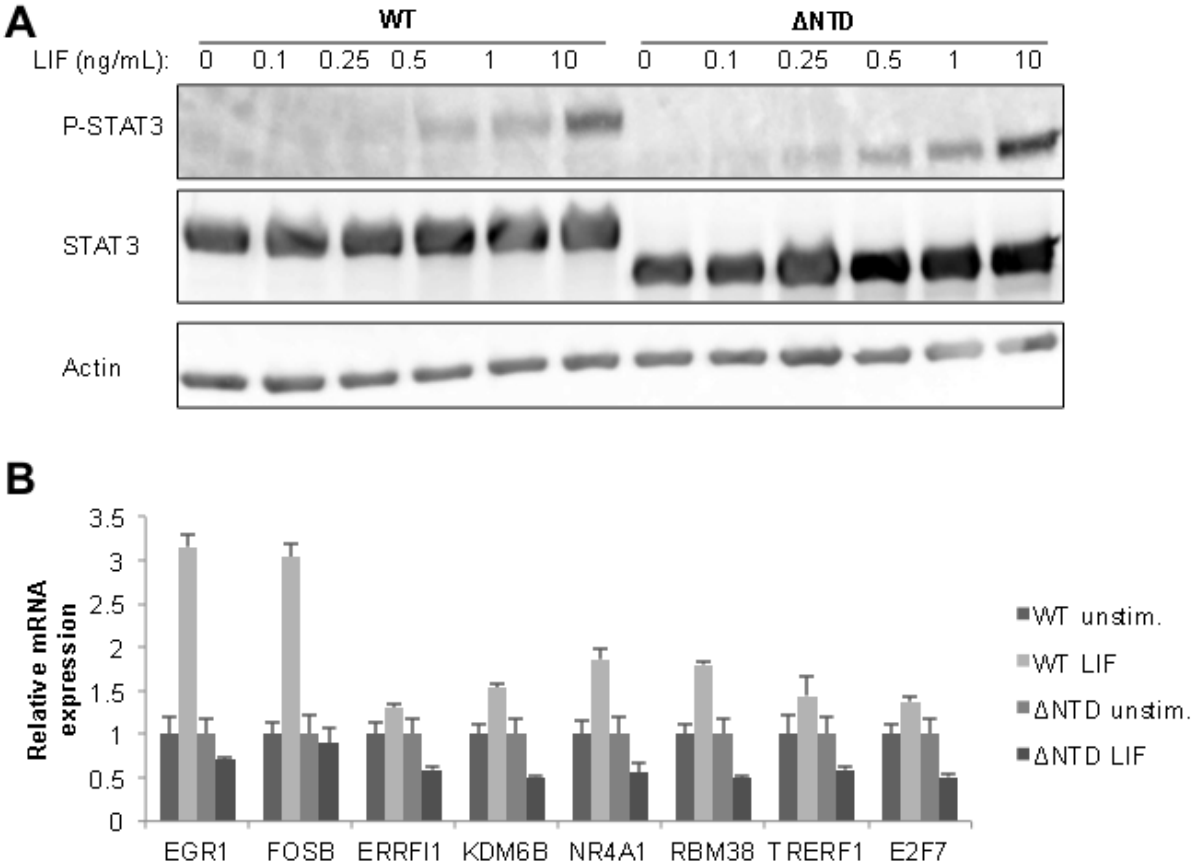
Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

Figure 4



Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

Figure 5

A

E2F7
chr10:110746119-110746318

TTTTCGTCTTCCCAGATTTCCAGTGAT
TCCTAAGTATATTTTAAGTGTTTGTCC
GTGTGAGTTTTAATGAGGAATGGAGGG
TGTCTTCATTCTTGGAAAATTCCAGGA
ATGCTGTAGGAAGGGAAAGTTGATGTT
CACAAGTGGCATTTTGTCGTTGTTGTT
GTAGGGGGAAATAAATGTAAAATCATT
TTCTGATATAG

GARNL3
chr2:33087180-3087319

CTGACTCTCTGCCAGGACACTGTTTGA
TAGGCTCCAAGAAAGACAGCATGAGAG
AAATGAGATTAGAGTCTAATGGTATTC
TGCCTCCACCATTTCACAGAAATGGCT
GAATTAATCTTATATGAGAAATTTAAC
CTTTT

NR4A1
chr15:101272824-101272949

AAAAAATCCCTGGCTTCATTGAGCTTT
GCCCAGGAGACCAAGACCTGTTGCTAG
AGTCTGCCTTCCTGGAACTCTTCATCC
TCCGCCTGGCATACCGGTAAGCTGCCC
ACCATCCTCCTAGCCCTG

RBM38
chr2: 173026640-173026841

TTGTGCCACACACAGGCACCGGTATAA
GCAAGTGCCATAGTGTCTAGGTCCCAT
AAGGGCCCCACACTGTACCAGACATTT
CCTGCACAGGCACATGGAGTGGGGTGA
GGCTGGGGGAAGGGCATTGTTTCTGTG
CCTTCAGAGAGGGAAGGTTGGGGGAGG
GACATTGCTCCTGTGCCTTCAGAGAGA
GTTACTAGTTTCA

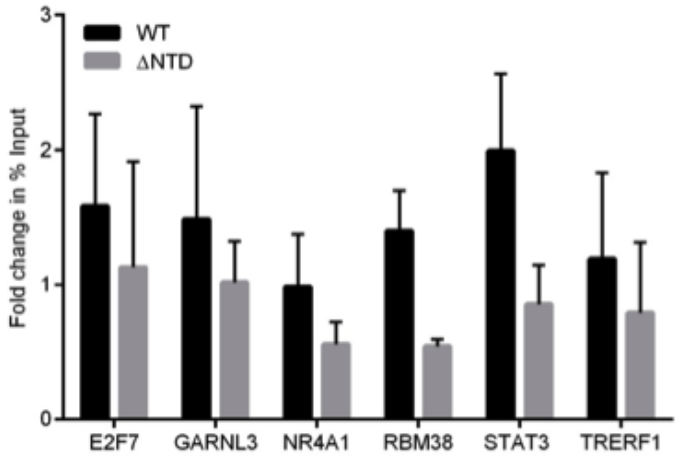
STAT3
chr 11: 100939668-100939869

CTCCCTGAGTTGGCTGTTCTGAGAGCT
TTGTGCTCCTCCCTCCGCGACCAAGGG
ACGCGCAGAGGCCTAGCTTACAAAAAA
GCCCGGGGGGAGGGAGGAGACATTAGC
GGAATGTCCTGCTGAAAACTCAGCTGA
GTTCCTGGCAGTGCGTGACGTCAAGAC
ACTTTAAATGCCCTGATACGGCTCGCT
TCTGCCCGCTCTC

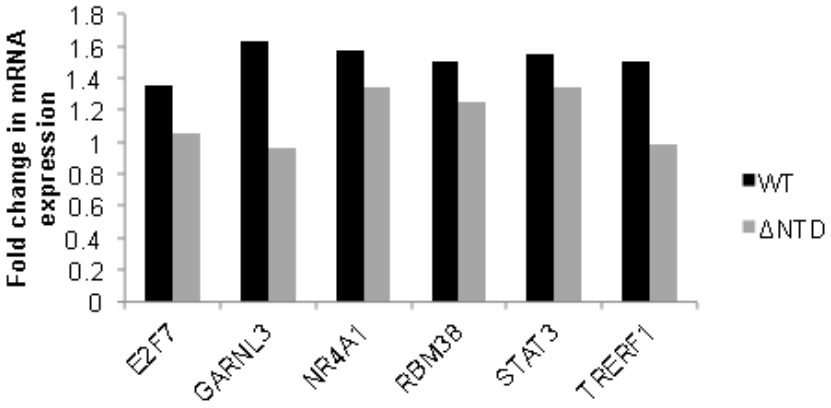
TRERF1
chr 17: 47225224-47225425

TATTGCTGTGTTGCAAAAACTTAATGG
GCTGGAGGCCTGGGGCCGTCTTTATCT
GCAATGAAAACCTAAGACATTGAGAGT
GGAGATGTTACTGGCAGTGTTTTTGCT
AGAAGAATTGTGTGCCCCGTCACTTCC
AAACTAGCATGGTCAGAGCGGGGAGGC
GGATCTGATAGCAGGTCTGCGGGAGTG
GGAGGTGTTTAAA

B



C



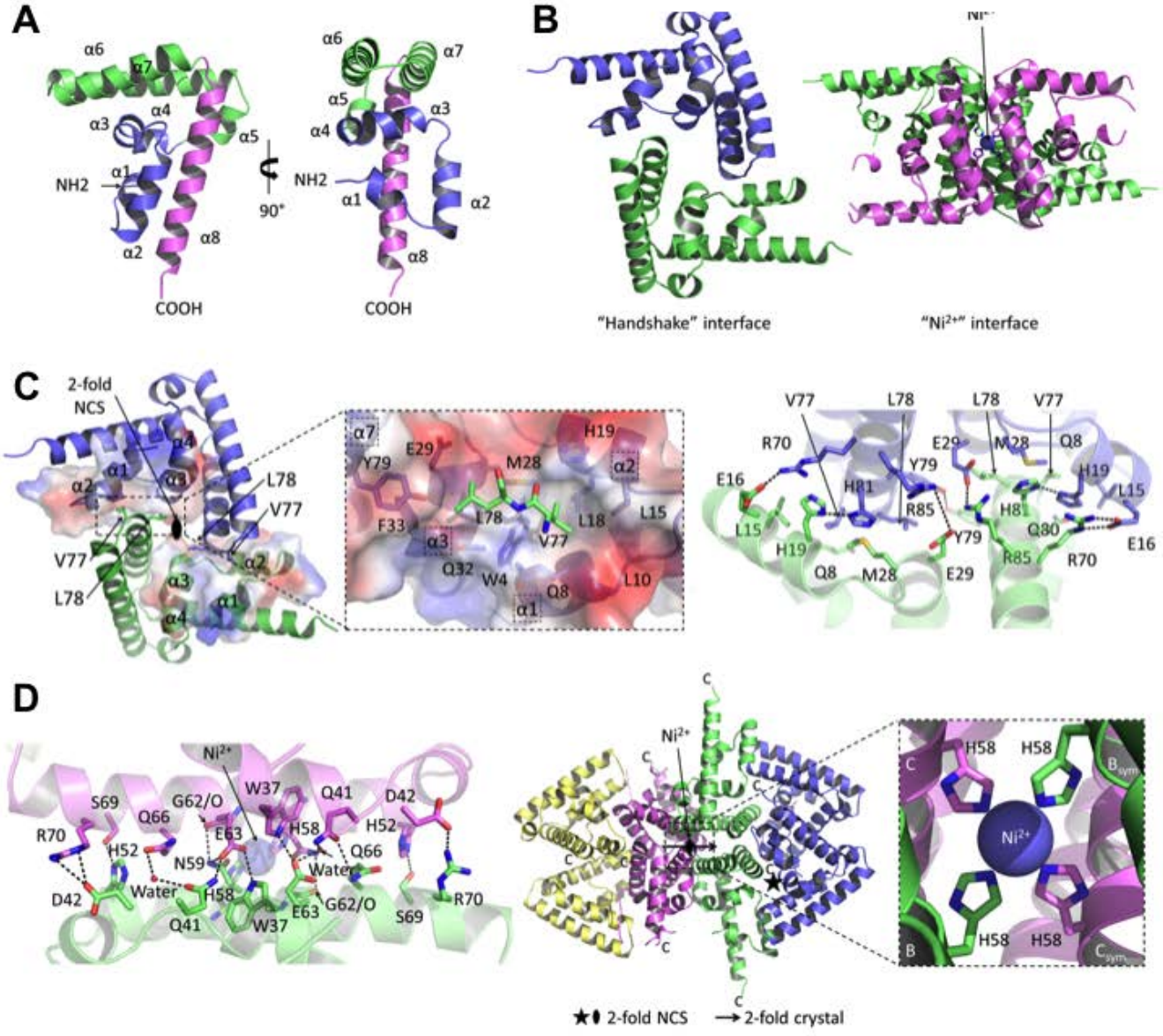
Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

Figure 6



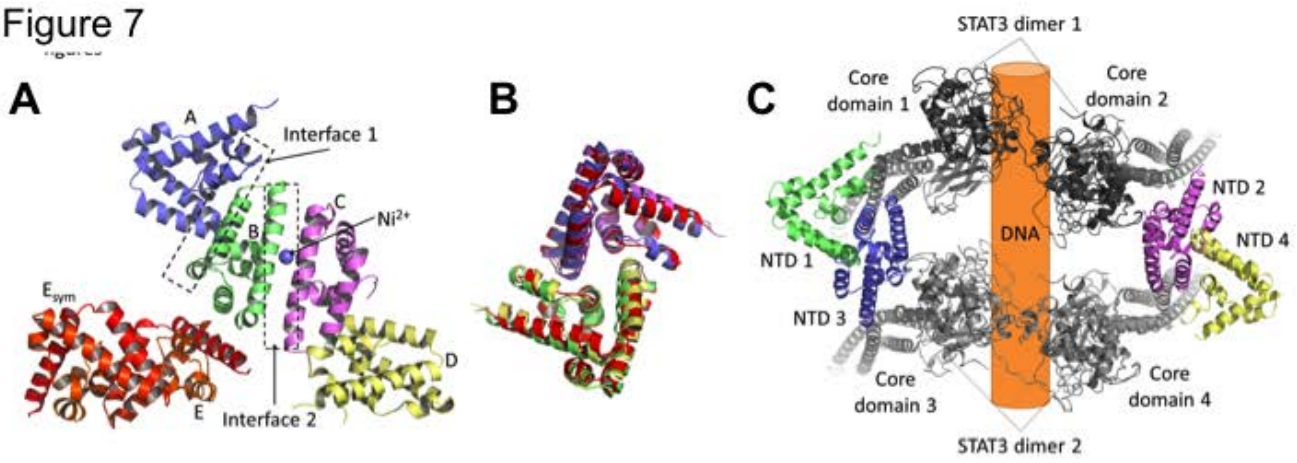
Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

Figure 7



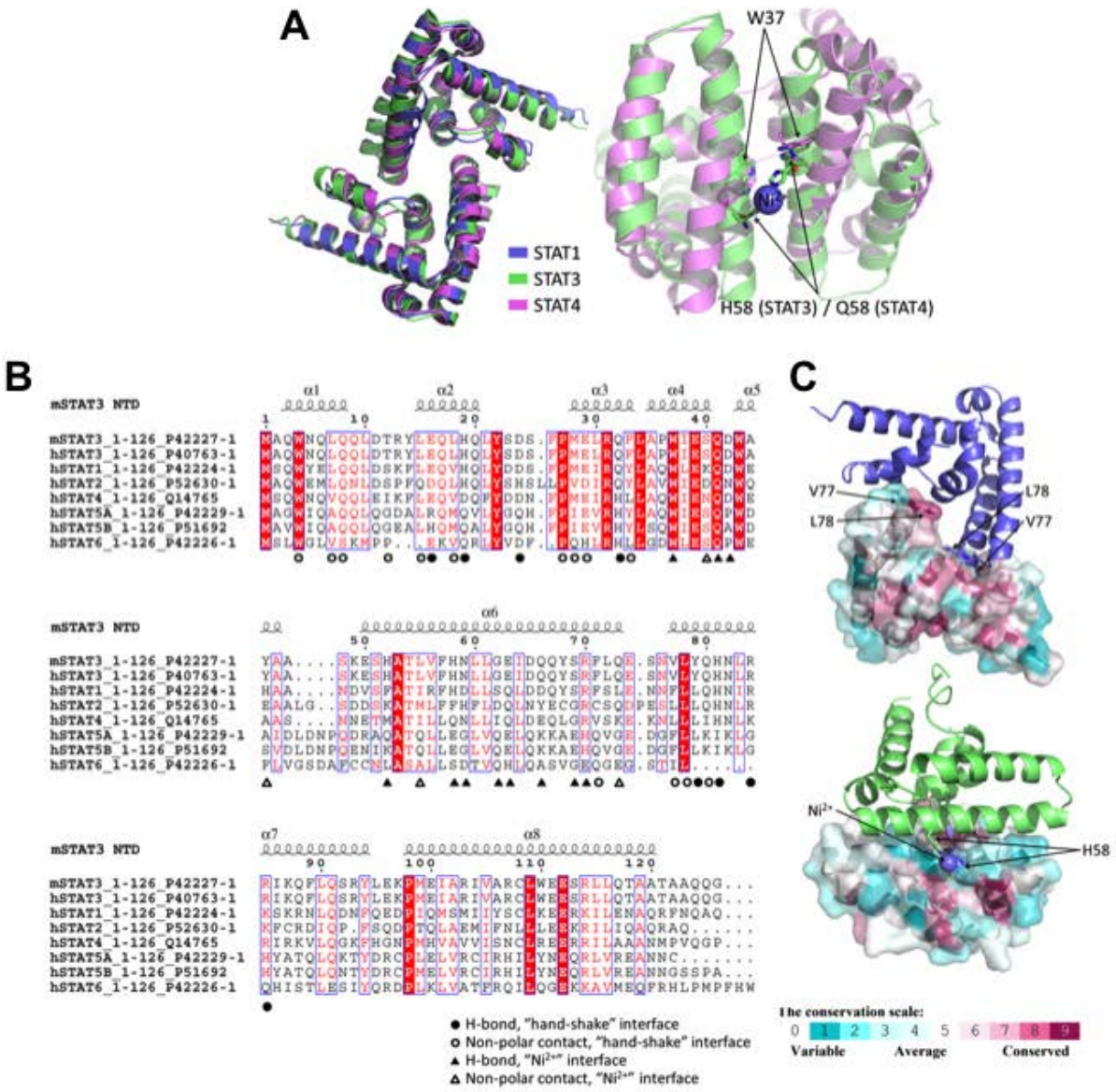
Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

Figure 8



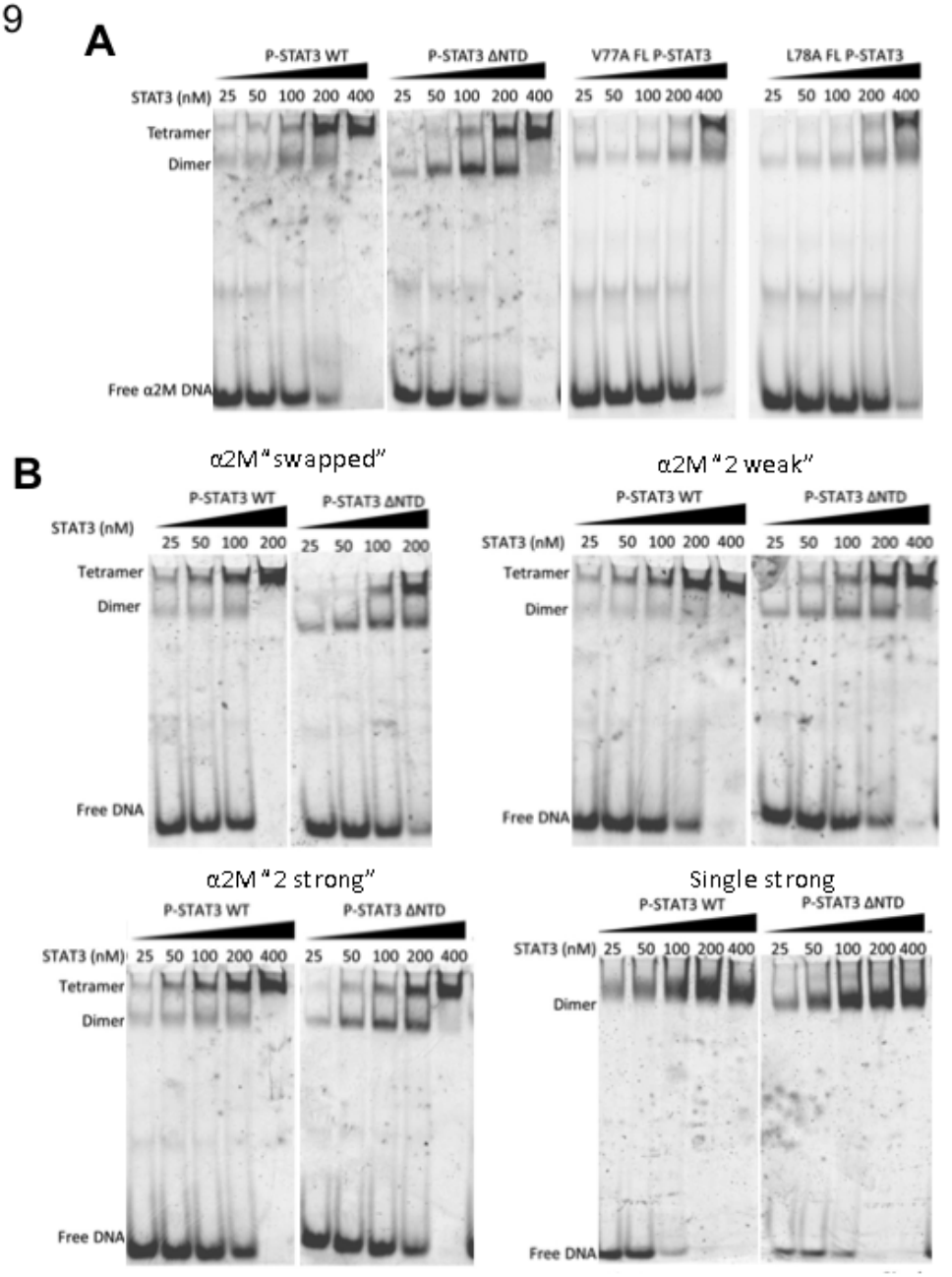
Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

Figure 9



DNA	Sequence
α2M	AGCAGTAACTGGAAAGTCCTTAATCCTTCTGGGAATTCT
α2M "swapped"	AGCAGTTCTGGGAAAGTCCTTAATCCTAACTGGAATTCT
α2M "2 weak"	AGCAGTAACTGGAAAGTCCTTAATCCTAACTGGAATTCT
α2M "2 strong"	AGCAGTTCTGGGAAAGTCCTTAATCCTTCTGGGAATTCT
Single strong	AGCAGTTCTGGGAAATCT

Accepted Manuscript Posted Online

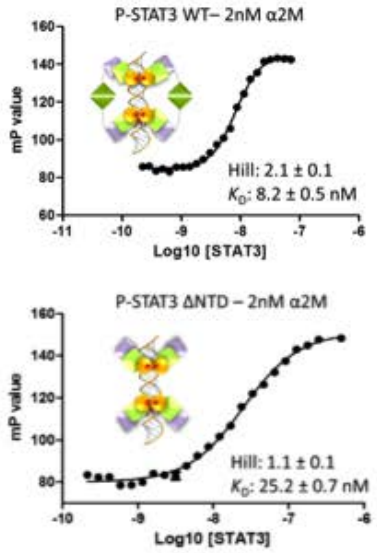
Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

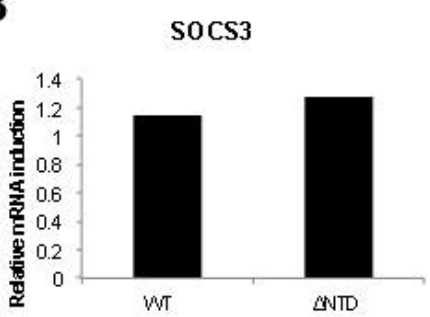
Figure 10

A



DNA probe	STAT3	Hill	K_D (nM)
α2M (AGCAGTAACTGGAAAGTCCTTAATCCTTCTGGGAATTCT)	WT	2.1 ± 0.1	8.2 ± 0.5
	ΔNTD	1.1 ± 0.1	25.2 ± 0.7
	V77A	1.2 ± 0.2	35.1 ± 8.8
	L78A	1.3 ± 0.1	25.3 ± 2.8
	H58A	2.2 ± 0.1	8.6 ± 0.2
	K140M	1.9 ± 0.1	9.0 ± 0.2
Single strong (AGCAGTTCTGGGAAATCT)	WT	1.3 ± 0.1	5.9 ± 0.1
	ΔNTD	1.2 ± 0.1	5.8 ± 0.1
SOCS3 (ACCCGATTCCTGGAACTGCGCGGCCGGCCTTCTTGTAATGTTT)	WT	1.7 ± 0.1	8.7 ± 0.2
	ΔNTD	1.3 ± 0.1	4.5 ± 0.1

B



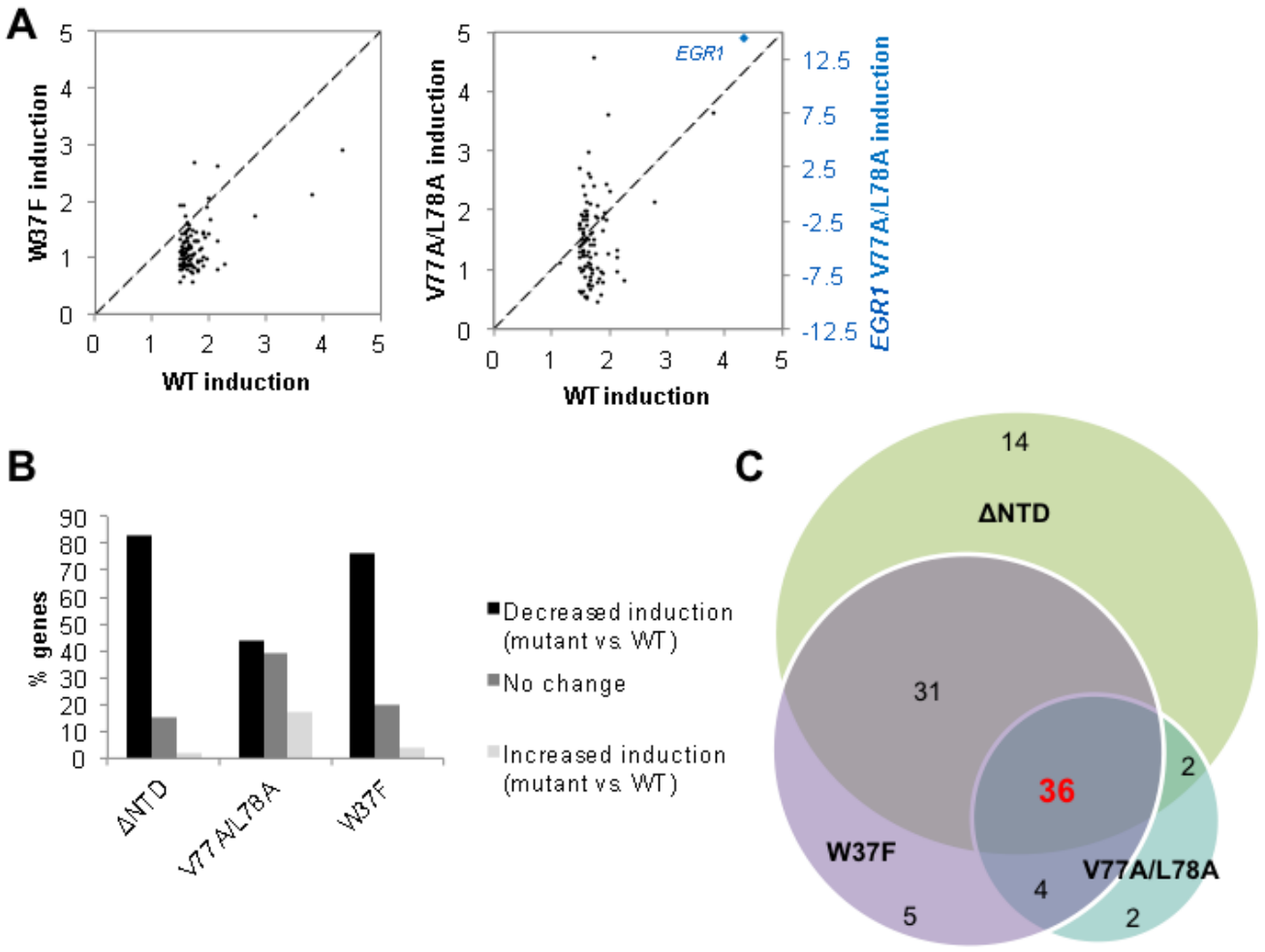
Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

Figure 11



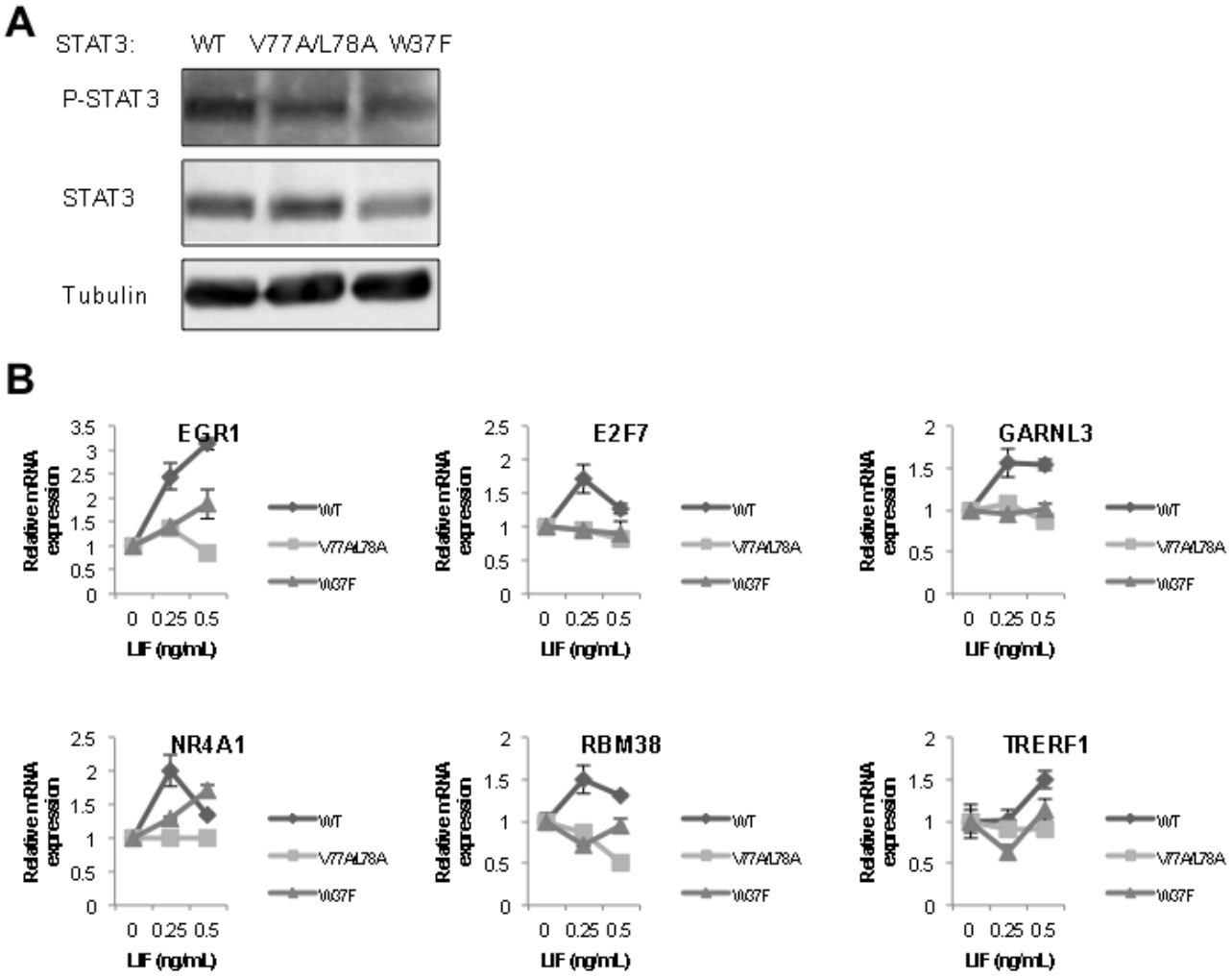
Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

Figure 12



Accepted Manuscript Posted Online

Molecular and Cellular Biology

MCB

Molecular and Cellular Biology

Figure 13