CHAPTER IV

IDENTIFICATION OF NUCLEAR PROTEINS THAT INTERACT WITH THE 5'-REGULATORY REGION OF THE MET RECEPTOR TYROSINE KINASE GENE

Mica Y. Bergman, Kian Kani, Parag Mallick, Pat Levitt

INTRODUCTION

The Met receptor tyrosine kinase is expressed in the developing and adult nervous systems and based *on in vitro* and *in vivo s*tudies, has the potential to mediate a variety of neurodevelopmental and neurophysiological processes, including peripheral and central neuron survival (Hamanoue et al., 1996; Zhang et al., 2000) and migration (Garzotto et al., 2008; Giacobini et al., 2007; Powell et al., 2001), axon guidance (Caton et al., 2000; Ebens et al., 1996; Powell et al., 2001), dendritic arborization(Gutierrez et al., 2004; Judson et al., 2010; Lim and Walikonis, 2008), and spine and synapse maturation and activity(Akimoto et al., 2004; Judson et al., 2010; Tyndall et al., 2007; Tyndall and Walikonis, 2006). Detailed mapping analysis of the *Met* transcript in the developing mouse forebrain reveals remarkably restricted patterns of expression in both the spatial and temporal domains (Judson et al., 2009), implying both tight regulation of transcription and specific function for the protein product. Temporally, *Met* levels peak dramatically during the first two postnatal weeks, suggesting a role for the protein in neurite outgrowth and synapse formation. Localization of the transcript to projection neurons within the neocortex and specific limbic structures suggests that it is involved in establishing connectivity between these structures. In fact, a recent study in which Met is deleted selectively form neocortical neurons reports selective effects on dendritic arborization and spine head volume of intrinsic pyramidal neurons and postsynaptic medium spiny striatal neurons(Judson et al., 2010). Consequently, deviations in the regulated expression of *Met* may predispose to alterations in the formation and functioning of

these circuits. Although several transcription factors have been identified that regulate Met *in vitro* (see Chapter I, Table 1), mostly using cancer cell lines, the complement of transcription factors controlling its expression in the nervous system is poorly understood. Yet knowledge of these regulatory factors is critical to understanding how Met expression patterns are established *in vivo*, and gives insight into sources of vulnerability to aberrant circuit patterning. This is particularly relevant in the context of recent studies of neurodevelopmental disorders. Three different alleles in the 5'-regulatory region of the human ortholog, *MET*, increase risk for Autism Spectrum Disorder (ASD) (Campbell et al., 2008; Campbell et al., 2006; Jackson et al., 2009; Sousa et al., 2009; Thanseem et al., 2010), and there is a reduction in MET protein expression in postmortem brain tissue from subjects with ASD (Campbell et al., 2007).

We sought a method for identifying a set of candidate DNA-binding proteins potentially involved in regulating *Met* transcription. One initial approach utilized commercially available transcription factor protein arrays in which transcription factors are spotted onto a PVDF membrane and incubated with oligonucleotides corresponding to the *Met* promoter sequence. This technique is limited by the number of transcription factors contained on the array. Also, given that, *in vivo*, many protein-DNA interactions require the presence of additional protein co-factors, there is a significant risk for false negative results. A second approach employed online databases of transcription factor binding sites, which scan the promoter of interest for similarity. This technique has successfully identified gene regulators (Liu et al., 2008), but it, too, carries disadvantages. As with the protein

array, it is limited by coverage, in this case to transcription factors with known binding sites. Moreover, this technique cannot account for co-factors required to promote binding and transcriptional control. This method, however, is more prone to false positive results due to co-factor stipulations (i.e. a protein is identified based on a predicted binding site match, but its requisite co-factor is not expressed *in vivo*). Additionally, the degree of discrepancy permitted between consensus binding site and actual sequence is not always well understood, rendering sequence-specific predictions more difficult.

Ultimately, we employed a combination DNA affinity column-mass spectrometry (MS) approach. Following a pull-down experiment in which nuclear extract harvested from postnatal day (P)7 cortical tissue was incubated with oligonucleotides corresponding to *Met* promoter sequence, candidate binding-proteins were collected and identified by MS. Using the pull-down-mass spectrometry combination, we identified 200 unique proteins, a subset of which may be part of a *Met* promoter-binding complex. We discuss here the validity of the approach, and the identified proteins we hypothesize may be veritable *Met* regulators.

MATERIALS AND METHODS

Mice and Tissue Collection

C57Bl/6J mice were purchased from Charles River Laboratories (Wilmington, MA). All research procedures using mice conformed to NIH guidelines

and were approved by the Institutional Animal Care and Use Committee at University of Southern California. Animals were housed on a 12 hour light/dark cycle with full access to food and water. Efforts were made to minimize animal suffering and to reduce the number of animals used.

Postnatal day (P) 7 mice were used for this study. At this time, *Met* gene expression is very high (Judson et al., 2009). To harvest brains, P7 mice were anesthetized with isofluorane prior to decapitation and brain removal. Harvested brains were immediately immersed in ice-cold Hanks' balanced salt solution (Sigma, St. Louis, MO) with the aid of an MZ-6 stereozoom microscope (Leica). The cerebral cortex from each hemisphere was rapidly dissected and weighed. Nuclear extracts were prepared immediately.

Nuclear extract preparation

The following protocol was modified from a number of previously published methods (Dignam et al., 1983; Dobi et al., 1997; Lewis and Konradi, 1996). Following dissection, tissue was immediately homogenized in 800 ul ice-cold PBS with fresh protease inhibitor cocktail (P8340; Sigma, St. Louis, MO) using a glass tissue homogenizer (Wheaton, Millvale, NJ). The homogenate was spun at 1500g for 5 minutes at 4°C, and supernatant was removed. The pellet was resuspended in Buffer A (10mM HEPES pH 7.9; 1.5 mM MgCl2; 10mM KCl with fresh 0.5 mM DTT and protease inhibitor cocktail). The volume of buffer A was approximately 10-fold over PCV. After buffer addition, samples were vortexed for 15 seconds and incubated on ice for 10 minutes. NP-40 was added to each sample for a final

concentration 0.55%, and samples were vortexed for 5 seconds, incubated on ice for 1 minute, vortexed an additional 5 seconds, and centrifuged at 16,000 g for 5 minutes at 4°C. The supernatant (cytoplasmic fraction) was discarded, and the nuclear pellet was resuspended in Buffer C (20 mM HEPES pH 7.9; 0.42 M NaCl; 1.5 mM MgCl2; 0.2 mM EDTA; 25% glycerol with fresh protease inhibitor cocktail), using a volume equal to approximately ¼ of the volume of buffer A. Samples were vortexed for 15 seconds and then incubated on ice for approximately 40 minutes, with vortexing every 10 minutes. Samples were then centrifuged at 16,000g for 10 minutes at 4°C. Supernatants were collected, DTT added for final concentration 0.5mM, and stored at -80°C.

Streptavidin-pulldown assays

Dynabeads MyOne Streptavidin T1 superparamagnetic beads with a surface streptavidin monolayer (Invitrogen, Carlsbad, CA) were used. Beads (375 ug) were combined in 1.5 ml tubes with 100ul 1X binding and washing (B&W) buffer (5 mM Tris-HCl (pH 7.5), 0.5 mM EDTA, 1M NaCl, 0.1% Tween-20). Tubes were placed on a magnetic strip (New England Biolabs, Ipswich, MA), and supernatant was removed. Beads were then washed three times with 1X B&W buffer and resuspended in 200 ul 2X B&W buffer. 5'-biotinylated probes (10-50pmol in 200ul water) were added to the beads and incubated at room temperature for 30 minutes with end-over-end rotation. After rotation, tubes were placed on a magnetic strip to allow removal of unbound probe. Beads were then washed twice with 1X B&W buffer and once with Buffer C-150 (20 mM HEPES pH 7.9; 0.15 M NaCl; 1.5 mM MgCl2; 0.2 mM EDTA;

25% glycerol) with 0.1% Tween-20, followed by resuspension in 400ug nuclear extract in Buffer C-150 with 0.1%tween. Beads and protein were then incubated at room temperature for 30 minutes with end-over-end rotation. After rotation, tubes were placed on a magnetic strip to facilitate removal of unbound protein. Beads were then watched 4 times with Buffer C-150 with 0.1% Tween-20. Contents were transferred to clean tubes at the fourth wash to reduce protein leeching. Bound proteins were eluted in 10 mM EDTA with 95% formamide at 90°C for 10 minutes and collected off of the beads using the magnetic strip.

Silver staining

Protein eluates were fractionated by SDS-PAGE, and silver staining was performed using the ProteoSilver™ Plus Silver Stain Kit (Sigma Aldrich, St. Louis, MO) according to the manufacturer's instructions.

Trypsin Digestion

Following removal of glycerol or formamide from the sample, the protein pellet to be analyzed was resuspended in 3M guanidine hydrochloride (GuHCl) in 100mM ammonium bicarbonate, 1% m/v n-octyl glucoside. Proteins were reduced by addition of DTT to final concentration 25 mM, and incubation at 65°C for one hour and subsequently alkylated by addition of iodoacetamide to final concentration 64 mM with incubation at room temperature for 45 minutes, without light. GuHCl concentration was reduced to 1M by addition of 100mM ammonium bicarbonate. Samples were digested by addition of sequencing grade trypsin (Promega, Madison,

WI) with a 50:1 (m/m) ratio and incubation at 37°C with rotation for 16 hours. The digestion was quenched by addition of 10% (v/v) glacial acetic acid. Digested samples were dried down in vacuum centrifuge overnight, and resuspended in mass spectrometry grade water.

Mass Spectrometry

Following trypsin digestion, each pulldown's eluted fraction was individually analyzed by LC-MS/MS (liquid chromatography MS/MS) in a LTQ-FTICR or LTQ-ORBITRAP mass spectrometer (Thermo-Finnigan, Waltham, MA) coupled to a nanoflow chromatography system (Eksigent, Dublin, CA) using a 25-cm column (Picofrit 75 μm ID, New Objectives, Woburn, MA) packed in-house with MagicC18 resin (Michrome Bioresources, Auburn, CA) over a 90 minute linear gradient. Acquired data was automatically processed by the Computational Proteomics Analysis System V8.2 – CPAS (33). The tandem mass spectra were searched against version 3.62 of the mouse IPI database (47,692 protein entries) with X!Tandem (2005.12.01). The mass tolerance for precursor ions was set to 1.5 Daltons. The mass tolerance for fragment ions was set to 0.5 Daltons. All identifications with a PeptideProphet probability greater than 0.75 were submitted to ProteinProphet and the subsequent protein identifications were filtered at a 1% error rate with tryptic fragments (1 missed cleavage), with allowance for fixed modification on C = 57.021 and variable modifications on C = -17.027, E = -18.011, K = 6.020, M= 15.995, and Q = -17.027

RESULTS

DNA oligonucleotide probe design

To identify proteins binding either directly or in a complex to the 5'-*Met* promoter, we employed streptavidin-pulldown assays in which nuclear protein is incubated with 5'-biotinylated, double-stranded DNA tethered to streptavidin-coated magnetic beads. We generated two sets of *Met* promoter probes to use as bait in this experiment: one set was created by PCR-mediated amplification of genomic DNA, and the other set was synthetic. Because *cis*-regulatory regions have been identified up to tens of kilobases upstream of (Clark et al., 2006; Galli et al., 2008) and within introns downstream of the transcription start site (TSS) (Kleinjan and van Heyningen, 2005; Lettice et al., 2003), it is not possible to scan the entire putative regulatory region for transcription factor binding with this technique. We narrowed our search based on the assumption that approximately 60-75% of mammalian gene promoters are located within CpG islands (Bird, 2002; Sharma et al., 2010). We used the UCSC Genome Browser (http://ucsc.genome.edu) to identify CpG islands proximal to the *Met* TSS, and were certain to include the island extending from -568 to +206 in our analysis.

We used Primer3 (http://frodo.wi.mit.edu/primer3/) to design primers for amplifying four overlapping amplicons that, in sum, extended 771 bp upstream and 338 bp downstream of the TSS. We constrained the size of our DNA probes to 400bp each, based on previous reports using this technique (Deng et al., 2003;

Stauff et al., 2007; Szemes et al., 2006). Because of the high GC content in this genomic region, there were a limited number of sites that were amenable to priming for PCR. As a result, the probes we designed were of non-uniform length. The oligonucleotides synthesized by the company were limited to 200bp, but were less restricted in location, and as a result, each was 200bp in length, and overlapped 25bp each with the upstream and downstream synthetic probes. Five probes, extending from -695bp upstream of the TSS through 225bp downstream of it were designed, but due to the guanine tetraplex "knot" often formed when synthesizing strings of greater than four G's, (Poon and Macgregor, 1998), one of the probes was not synthesizable. Details of the probes are found in Figure 1 and Table 1.

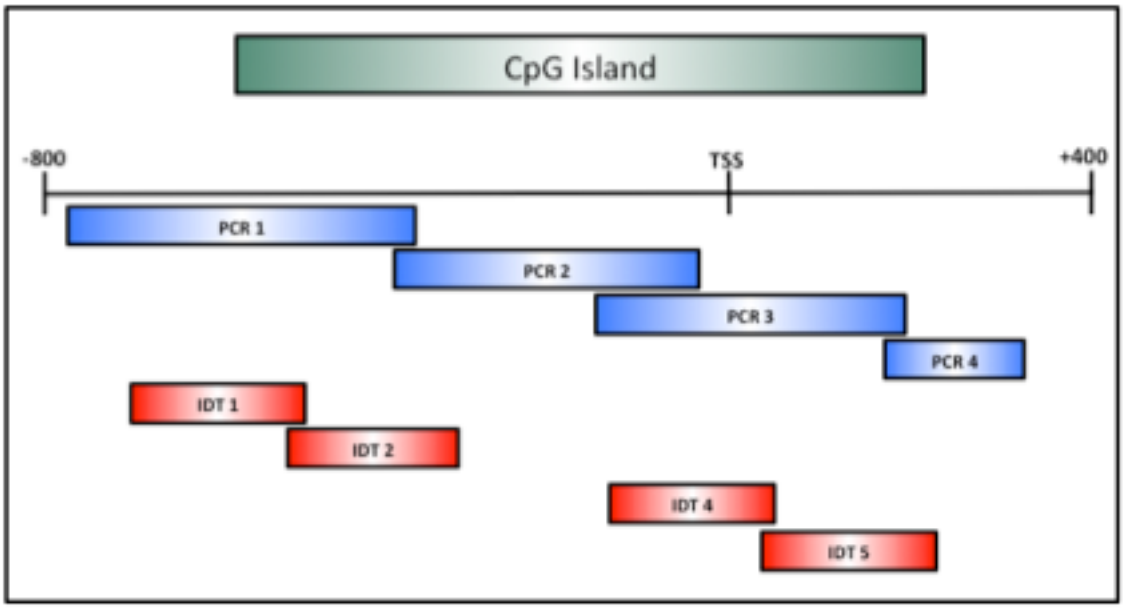

Figure 1. Location of oligonucleotide probes used in streptavidin-pulldown assays. The black line represents mouse *Met* DNA sequence, and the transcription start site is indicated by the TSS designation. The probes designated "PCR" represent amplicons generated by PCR, whereas the probes labeled, "IDT" represent synthetic ultramers.

Probe Name	Start Relative to TSS	Stop Relative to TSS	Length bp
PCR 1	-771	-372	400
PCR 2	-395	-43	353
PCR 3	-162	+200	362
PCR 4	+177	+338	162
IDT 1	-695	-496	200
IDT 2	-515	-316	200
IDT 4	-155	+45	200
IDT 5	+26	+225	200

Table 1: Positional location of oligonucleotide probes used in streptavidin-pulldown experiments. Positions are based on the mouse *Met* promoter sequence, relative to the transcription start site (TSS).

Streptavidin-pulldown experiments are compatible with mass spectrometric identification approaches

Before pursuing a large-scale mass spectrometric (MS) analysis, we aimed to establish that our experimental paradigm was suited for such a discovery approach. First, we demonstrated that the streptavidin-pulldown experiments could differentiate proteins binding to individual regions of the *Met* promoter. Using nuclear extract obtained from P7 mouse cortex as prey and DNA oligonucleotides representing the mouse *Met* promoter as bait, we allowed *in vitro* binding as described in Materials and Methods. We ran the eluted fractions on polyacrylamide gels and used silver staining to visualize nonspecifically the eluted proteins, which had formed DNA-binding complexes with the *Met* probes. The silver stains showed that each probe was capable of capturing multiple proteins from within the nuclear extract (Fig. 2). Additionally, at a qualitative level, protein binding to the probes

appeared to be sequence-specific, as the banding patterns and band intensities differed depending on the probe that was used (Fig. 2).

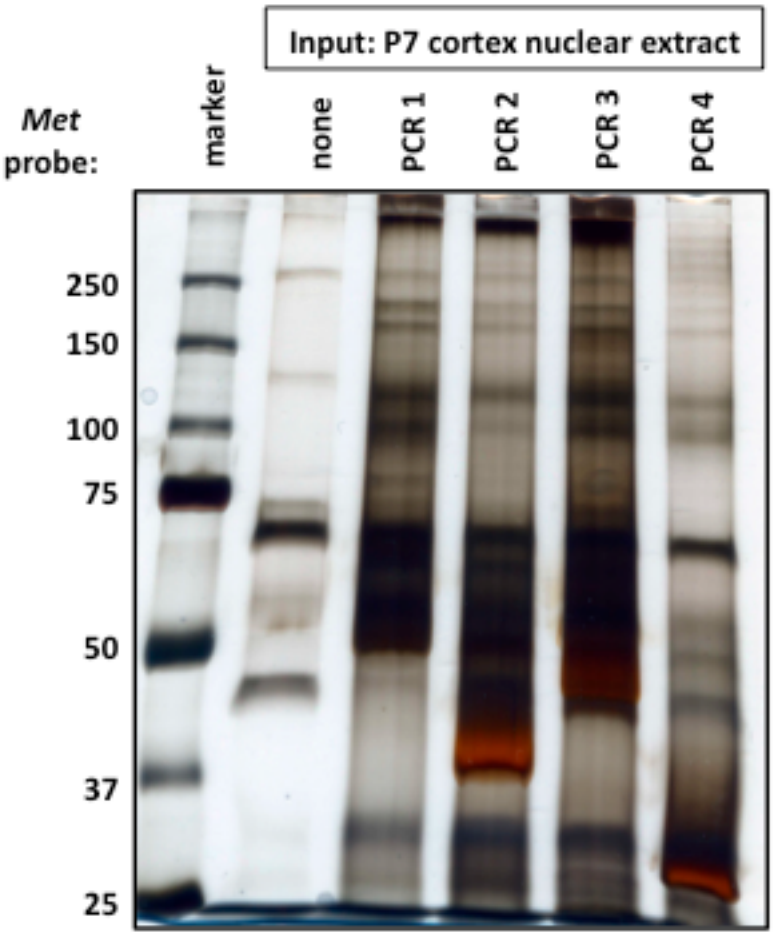

Figure 2: Silver stain of protein eluted following streptavidin-pulldown experiments with indicated oligonucleotide probe. Incubation with each of the probes results in the capture of multiple proteins from the input nuclear extract. Qualitatively, the banding patterns and band intensities differ depending on the probe that was used.

MS analysis requires that the proteins being identified first undergo tryptic digestion into their component peptides. In the final step of our pulldown approach, the proteins of interest were eluted with 95% formamide in 10mM EDTA.

Formamide is not a standard buffer for trypsinization, and we demonstrated by silver stain that it inhibits the digestion process (Fig. 3; compare lanes 1 and 2). Removal of formamide by acetone precipitation was successful and allowed subsequent tryptic digestion (Fig. 3; compare lanes 2 and 4).

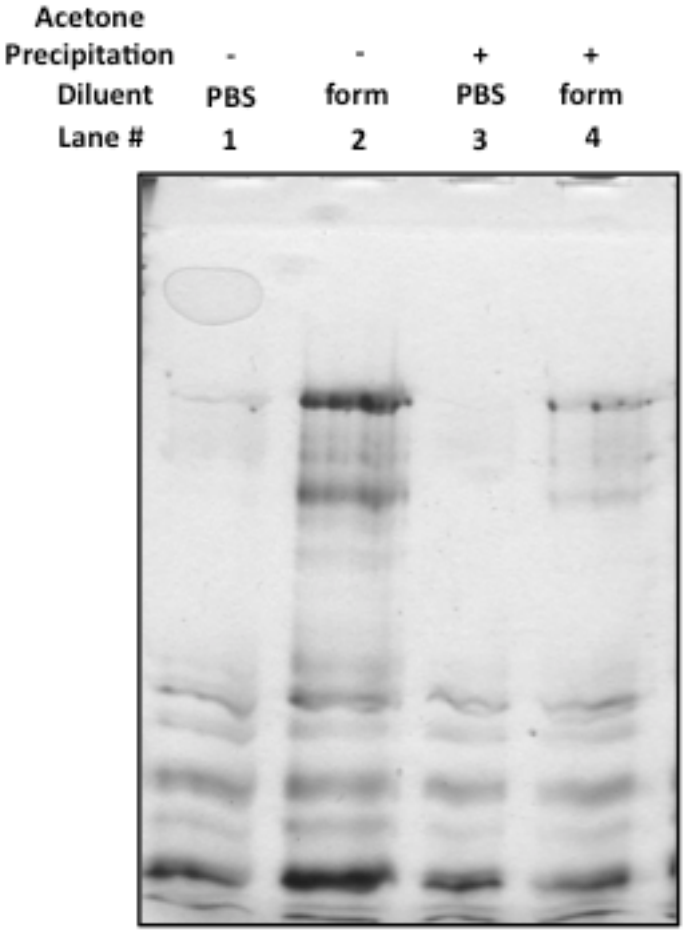

Figure 3. Silver stain to assess tryptic digestion. Bovine serum albumin was resuspended in indicated diluent. Samples were precipitated with acetone and resuspended in PBS. All samples were reduced, alkylated, and digested with trypsin. The presence of higher order species (lane 2) indicates incomplete digestion in the presence of formamide. Removal of the formamide by acetone precipitation (lane 4) significantly improves the extent of digestion.

Summary of Mass Spectrometric Results

Mass spectrometric analysis of the tryptic fragments of the eluted proteins yielded 200 unique identifications. We utilized the Database for Annotation, Visualization and Integrated Discovery (DAVID) to assign Gene Ontology (GO) "Biological Process" categories to each of these proteins. Forty three proteins were absent from the DAVID database, and therefore excluded from further analysis. These proteins were all either predicted genes or genes that do not yet have an official symbol, with the exception of Cnbp2. We searched individually for GO terms associated with Cnbp2 and did not find any. Among the remaining 157 proteins, 112 GO-Biological Process categories were represented, and proteins ranged from having 0 associations (10 proteins) to 45 associations.

Prioritization of Mass Spectrometric Results

We used the GO classifications to prioritize the MS results for further study of *Met* regulation. Proteins classified as relevant to transcriptional regulation were clear choices for continued analysis. Additionally, we selected proteins with classifications relevant to central nervous system (CNS) development or physiology. We reasoned that proteins with known roles in these domains that also showed *in vitro* binding to the *Met* promoter were strong candidates for follow-up study. Proteins with classifications relevant to both transcriptional regulation and the CNS were grouped separately and considered highest priority.

We defined which GO terms were included in these broad categorical sets, and these terms are listed in Table 2. GO terms relevant to chromatin alteration

were considered to be a part of the transcriptional regulation category, as modifying chromatin structure is an important mechanism for permitting or restricting transcription factor access to gene promoters, thus influencing transcription. GO terms pertaining to establishing or maintaining chromatin structure were not included in this category.

Proteins were assigned to a broad category -- Transcriptional Regulation, CNS Development/Physiology, or both -- if they were associated with one or more of the GO terms included in that category. The redundancy and nested nature of GO terms within each category was not problematic because of the qualitative nature of these assignments, and because proteins were not ranked within the category. The proteins found in each of these categories are listed in Table 3.

TRANSCRIPTIONAL REGULATION	BRAIN DEVELOPMENT/ PHYSIOLOGY
Positive regulation of gene expression	Central nervous system neuron differentiation
Negative regulation of gene expression	Forebrain development
Transcription	Telencephalon development
Regulation of transcription	Telencephalon regionalization
Positive regulation of transcription	Cerebral cortex regionalization
Negative regulation of transcription	Neuron development
Regulation of transcription, DNA-dependent	Neuron differentiation
Positive regulation of transcription, DNA-dependent	Neuron migration
Negative regulation of transcription, DNA-dependent	Neuron projection development
Regulation of transcription from RNA Polymerase II Promoter	Neuron projection morphogenesis
Positive regulation of transcription from RNA Polymerase II promoter	Dendrite development
Negative regulation of transcription from RNA Polymerase II promoter	Axonogenesis
Chromatin Modification	Axon extension
Chromatin remodeling	Learning or memory
Positive regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	
Negative regulation of nucleobase, nucleoside nucleotide and nucleic acid metabolic process	

Table 2: Gene Ontology (GO) terms used to define broad categories for prioritizing proteins identified by mass spectrometry

Transcriptional Regulation and CNS Development/Physiology

Gene Symbol	Gene Name	*Met* Probes
Bcl11b	B-cell CLL/lymphoma 11b	All 8
Emx2	Empty spiracles homolog 2	PCR1
Mecp2	Methyl CpG binding protein 2	All 8
Nfib	Nuclear factor I/b	IDT1,2; PCR1
Nr2f1	Nuclear receptor subfamily 2, group F, member 1	IDT5; PCR3
Satb2	Special AT-rich sequence binding protein 2	PCR1
Zeb2	Zinc finger E-box binding homeobox	IDT4,5; PCR1,2,3,4

CNS Development/Physiology

Gene Symbol	Gene Name	*Met* Probes
Capza2	Capping protein (actin filament) muscle Z-line, alpha 2	IDT2,5
Dclk1	Doublecortin-like kinase 1	IDT1,2,5; PCR1,2,3,4
Dcx	Doublecortin	IDT5; PCR1,2
Ppp1r9a	Protein phosphatase 1, regulatory (inhibitor) subunit 9a	IDT1,2; PCR3
Ptn	Pleiotrophin	IDT1,2,4,5; PCR1,3
Tmod2	Tropomodulin 2	IDT2,5; PCR2
Top2b	Topoisomerase (DNA) II beta	All 8
Vdac1	Voltage-dependent anion channel	PCR2,4

Transcriptional Regulation

Gene Symbol	Gene Name	*Met* Probes
Bcl11a	B-cell leukemia/lymphoma 11a	All8
Bclaf1	BCL2-associated transcription factor 1	IDT1,2,5
Cenpv	centromere protein V	IDT4
Cnbp	cellular nucleic acid binding protein	IDT1
Ctcf	CCCTC-binding factor	IDT1,5; PCR1,2,3,4
Ddx5	DEAD (Asp-Glu-Ala-Asp) box polypeptide 5; predicted gene 12183	IDT1,2,4; PCR1,2,3
Foxk2	forkhead box K2	PCR3
Gatad2b	GATA zinc finger domain containing 2B	PCR3
Hmga1	high mobility group AT-hook I, related sequence 1; high mobility group AT-hook 1	IDT1
Hnrpdl	Heterogeneous nuclear riboprotein D1	PCR1
Khdrbs3	KH domain containing, RNA binding, signal transduction associated 3	PCR1
Maz	MYC-associated zinc finger protein (purine-binding transcription factor)	IDT2,5
Mta1	metastasis-associated gene family, member 1	PCR1,2
Mta2	metastasis-associated gene family, member 2	IDT1,2; PCR1,2,4
Myef2	myelin expression factor 2	PCR1,2,3,4
Nfic	nuclear factor I/C (CCAAT-binding transcription factor)	IDT1; PCR1
Nfix	nuclear factor I/X (CCAAT-binding transcription factor)	All 8

Transcriptional Regulation Continued

Gene Symbol	Gene Name	*Met* Probes
Psip1	PC4 and SFRS1 interacting protein 1	All 8
Purb	purine rich element binding protein B	All 8
Rbbp4	retinoblastoma binding protein 4	IDT1,2; PCR2,4
Rbm14	RNA binding motif protein 14	All 8
Ruvbl1	RuvB-like protein 1	PCR1
Ruvbl2	RuvB-like protein 2	All 8
Sltm	SAFB-like, transcription modulator	IDT 4
Tcof1	Treacher Collins Franceschetti syndrome 1	IDT1,2,5; PCR1,3
Tfam	Transcription factor A	IDT2,5; PCR1,2,5
Thrap3	Thyroid hormone receptor associated portein 3	IDT1,2,4,5; PCR1,2,5
Tmpo	Thymopoietin	PCR1
Trim28	Tripartite motif-containing 28	PCR1,2,3
Ubtf	Upstream binding transcrpition factor, RNA polymerase 1	All 8
Usf1	Upstream transcription factor 1	PCR3
Xrn2	5'-3' exoribonuclease 2	IDT4,5; PCR1
Zfp42	Zinc finger protein 42	IDT1
Zfp512	Zinc finger protein 512	IDT1,2,4,5; PCR1,2,3

Table 3: Proteins selected due to membership in one or both of the prioritized categories of GO terms. For each protein, the gene symbol and gene name are given, along with the *Met* probes they were incubated with prior to elution and identification.

DISCUSSION

This study identified a large number of proteins that bind, either directly or indirectly, to the *Met* promoter *in vitro*. Given the limited number of transcriptional regulators reported in the literature, this discovery-based approach provides a new

set of proteins for more detailed analyses. However, although mass spectrometry provides a rich source of candidate regulators in the context of cortical tissue, the relevance of these proteins to *Met* regulation must be validated with additional *in vitro* and *in vivo* studies, such as luciferase assays and chromatin immunoprecipitation experiments. We discuss here two newly-identified candidate *Met*-binding proteins that we hypothesize are likely to have a role in regulating *Met* expression in the developing mouse forebrain.

B-cell CLL/lymphoma 11a (Bcl11a)

The B-cell CLL/lymphoma 11a (Bcl11a, also known as CTIP1 and Evi9) gene product is a zinc finger transcription factor that is expressed primarily in the hematopoietic/immune and nervous systems (Leid et al., 2004; Nakamura et al., 2000), and has documented roles in neurodevelopment (Kuo et al., 2010; Kuo et al., 2009). We highlight Bcl11a for validation due to the molecular mechanisms by which it regulates gene transcription, its roles in neurite outgrowth and branching, and its expression patterns in the developing rodent forebrain.

Bcl11a has a nine base pair consensus binding site with a core six base pair motif of 5'-GGCCGG-3', which is similar to the canonical GC box (Avram et al., 2002). Electromobility shift assays using various competitor oligonucleotides indicate that the five prime GGCC sequence is most critical to binding. Although the nine base pair consensus site is not found within the region of the *Met* promoter contained in our oligonucleotide probes, the core six base pair motif is found once, and is located in probe IDT 4 and probe PCR 2, and the critical GGCC sequence is found seven

times, and is present in IDT probes 1, 2, and 4, and PCR probes 1, 2, and 3. Bcl11a was identified in the eluted fractions collected with each of the eight *Met* promoter probes. Further tests will be necessary to determine which of these identifications represents a biologically relevant finding. It is possible that all of the identifications are true positives; Bcl11a may exhibit more promiscuous binding than is currently appreciated, or Bcl11a may bind some of the probes directly and others indirectly, as part of a complex. It is also possible that some, or all, of these results are false positives. The presence of the Bcl11a canonical binding site within the *Met* promoter, however, supports the prediction that this transcription factor is a true positive result in at least some cases.

Functionally, Bcl11a has demonstrated roles in both axon and dendrite morphogenesis. In hippocampal culture systems, overexpression of Bcl11a results in a reduction of total axon and total dendrite length, and a reduction in axon branch number. Conversely, knockdown by RNA interference results in increased neurite complexity; both axons and dendrites display increased total length and branching. Together, these data suggest that endogenous Bcl11a plays a role in limiting neurite growth and complexity. Met has been shown to be involved in promoting dendritic outgrowth in both *in vitro* and *in vivo* conditions (Gutierrez et al., 2004; Judson et al., 2010; Lim and Walikonis, 2008). A negative regulatory relationship between Bcl11a and Met could explain the opposing effects that the two proteins have on dendritic properties. To this end, it is intriguing that Bcl11a has been shown to function as a transcriptional repressor in several *in vitro* systems (Durum, 2003; Kuo and Hsueh, 2007). The nature of the relationship between Bcl11a and *Met* must be clarified in

the context of the developing mouse forebrain, as the possibility of context-dependent transcriptional activation has been proposed, and examples of other zinc finger proteins capable of both activating and repressing transcription exist in the literature (Durum, 2003; Kallio et al., 1995; Yet et al., 1998).

Bcl11a expression in the developing rodent forebrain has been characterized (Kuo and Hsueh, 2007; Leid et al., 2004) and reveals patterning that is consistent with a *Met*-regulating role. The transcription factor neither overlaps entirely with nor is restricted entirely from areas of *Met* expression, suggesting that a one-to-one relationship between transcription factor and transcript does not exist, but rather that the nature of the relationship is influenced by the complement of additional regulatory proteins expressed in particular areas of the developing brain. For example, early postnatally, Bcl11a and *Met* show similar layer-specific patterning in the cortex. Both are abundantly expressed in layers II, III, V, and VI, but are absent from layer IV (Judson et al., 2009; Kuo and Hsueh, 2007), suggesting a possible activating role. Conversely, Bcl11a is expressed in both the cortex and the striatum, whereas *Met* is expressed solely in the cortex. It is possible that the protein milieu of the striatum results in Bcl11a acting as a repressor on the *Met* promoter in this structure, whereas in the cortex, the set of co-expressed proteins results in Bcl11a-mediated activation of the *Met* promoter.

The streptavidin-pulldown technique does not differentiate proteins that bind directly to DNA from those that bind DNA indirectly vis a vis another protein intermediary. Intriguingly, a known Bcl11a binding partner, Nr2f1 (Coup-TF1), was also identified in our pulldown experiments. In fact, early identification of Bcl11a

was as CTIP1, or Coup-TF1 interacting protein 1 (Avram et al., 2000). Coup-TF1 is an orphan nuclear receptor implicated in transcriptional regulation during nervous system development. Its identification in the eluted fractions lends support to the idea that Bcl11a may be a veritable *Met* regulator, and suggests that Coup-TF1, too, may be a player in controlling *Met* expression.

Special AT-rich sequence-binding protein 2 (SATB2)

The special AT-rich sequence-binding protein 2 (Satb2) is another top candidate for validation, based on its DNA binding and regulatory properties and its localization in the developing brain.

Satb2 was identified initially in a DNA affinity purification screen designed to isolate proteins binding the enkephalin gene promoter from neonatal rat cerebral cortex. Like its ortholog, Satb1, Satb2 binds preferentially to AT-rich DNA which the authors define as a stretch of double-stranded DNA at least 36bp in length, that is composed of at least 70% adenine and thymine nucleotides(Szemes et al., 2006). Although the *Met* promoter is GC rich overall, it contains a stretch of 56 nucleotides, of which 71.4% are AT, beginning 761 base pairs upstream of the TSS. This sequence is found in PCR probe 2, which is the *Met* probe that precipitated Satb2 in our study. Satb2 regulates gene expression indirectly, but definitively, by modifying chromatin structure (Leone et al., 2008; Szemes et al., 2006), rendering the promoter more or less accessible to direct DNA-binding regulators. By this mechanism, Satb2 has been shown to repress expression of some genes (Alcamo et al., 2008), but to activate expression of others (Dobreva et al., 2006).

In the cortex, Satb2 is most abundantly expressed perinatally and early postnatally (Britanova et al., 2005; Cahoy et al., 2008; Szemes et al., 2006), the time when *Met* expression peaks and *Met* laminar patterning becomes most apparent (Judson et al., 2009). Satb2 protein and *Met* transcript show overlapping, but non-identical radial distribution. Satb2 is found primarily in layers II, III, and IV, but is also expressed in layers V and VI, albeit at much lower levels. *Met* shows high expression in layers II,III, V, and VI, but is virtually absent from layer IV (Judson et al., 2009). While the expression patterns suggest that Satb2 is not a blanket regulator of *Met* in the cortex, the patterns of expression are consistent with a regulatory relationship that depends on the various combinations of transcription factors and chromatin modifiers expressed within each cortical layer.

Interestingly, as with Bcl11a, a known Satb2 interacting partner was pulled down by the *Met* promoter oligonucleotide. Metastasis-associated protein 2 (MTA2) was identified using the same *Met* probe that identified Satb2, in addition to two other probes. MTA2 is co-expressed with SATB2 in early postnatal cortical cells, and is a part of the same chromatin remodeling complex (Gyorgy et al., 2008). Identification of the pair of proteins by MS provides further evidence that Satb2 represents a true positive result, and reconfirms its selection as a top choice for validation.

Assessment of technique and results

The combination DNA-affinity column-MS method employed in this study has several advantages over approaches employed previously to identify *Met*-binding

transcriptional regulators. First, it is unbiased and does not rely on pre-selected putative regulatory proteins. Given that there is a paucity of known *Met*-regulators, and that they were identified mostly in cancer cells, the unbiased screen provides a rich list of neurodevelopmentally-pertinent candidates upon which to follow up. We note that the protein pool from which candidates are identified is biologically relevant, having been harvested directly from the tissue of interest. This property permits interactions between necessary cofactors and restricts the search to proteins that are expressed at the time and place of interest. Despite its advantages, however, the DNA-affinity column-MS approach is not flawless, and both pieces of the technique carry technical caveats. DNA-protein binding in an *in vitro* environment does not necessarily reflect binding in an *in vivo* environment due to differences in physical, spatial, and kinetic parameters. Additionally, the long oligonucleotides may predispose to false positives resulting from nonspecific DNA binding. On the MS side, protein quantity is a factor in identification, so important, but low-abundance proteins can be missed. Additionally, some proteins are inherently more suited to digestion and analysis than others, introducing another potential layer of bias. For a candidate approach, however, an exhaustive list is not necessary, and this technique is well suited for generating an initial set of candidate proteins.

In this study, we identified 200 proteins from P7 mouse cortex that bind to various segments of the *Met* promoter in an *in vitro* binding assay. Of this group, seven proteins are particularly noteworthy as they have GO annotations related to both transcriptional regulation and CNS development/physiology. An additional 42

proteins are also highlighted as strong candidates due to their membership in one of these categories. A risk of prioritizing candidates based on GO category membership is that the strategy relies on pre-existing classifications that may be incomplete. It is possible that an important transcriptional regulator has never before shown DNA-binding or transcriptional effects and is, therefore, excluded from consideration. There is a particular risk for errors of this type when considering complexed proteins that affect gene regulation via interaction with another protein. For example, a particular kinase may not be assigned a transcriptionally relevant GO term, but its role in phosphorylating a transcription factor may be critical to the downstream protein's regulatory function, and thus the role of the kinase in regulating transcription is also critical.

Despite the associated caveats, the approach demonstrated here is an effective method for identifying candidate proteins involved in controlling *Met* expression by binding to its 5'-regulatory region. Candidates chosen for further analysis can be validated subsequently using *in vitro* and *in vivo* techniques. Our laboratory is currently pursuing these studies.

Literature Cited

Akimoto, M., Baba, A., Ikeda-Matsuo, Y., Yamada, M.K., Itamura, R., Nishiyama, N., Ikegaya, Y., and Matsuki, N. (2004). Hepatocyte growth factor as an enhancer of nmda currents and synaptic plasticity in the hippocampus. Neuroscience *128*, 155-162.

Alcamo, E.A., Chirivella, L., Dautzenberg, M., Dobreva, G., Fariñas, I., Grosschedl, R., and McConnell, S.K. (2008). Satb2 regulates callosal projection neuron identity in the developing cerebral cortex. Neuron *57*, 364-377.

Avram, D., Fields, A., Pretty On Top, K., Nevrivy, D.J., Ishmael, J.E., and Leid, M. (2000). Isolation of a novel family of C(2)H(2) zinc finger proteins implicated in transcriptional repression mediated by chicken ovalbumin upstream promoter transcription factor (COUP-TF) orphan nuclear receptors. J Biol Chem *275*, 10315-10322.

Avram, D., Fields, A., Senawong, T., Topark-Ngarm, A., and Leid, M. (2002). COUP-TF (chicken ovalbumin upstream promoter transcription factor)-interacting protein 1 (CTIP1) is a sequence-specific DNA binding protein. Biochem J *368*, 555-563.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev *16*, 6-21.

Britanova, O., Akopov, S., Lukyanov, S., Gruss, P., and Tarabykin, V. (2005). Novel transcription factor Satb2 interacts with matrix attachment region DNA elements in a tissue-specific manner and demonstrates cell-type-dependent expression in the developing mouse CNS. Eur J Neurosci *21*, 658-668.

Cahoy, J.D., Emery, B., Kaushal, A., Foo, L.C., Zamanian, J.L., Christopherson, K.S., Xing, Y., Lubischer, J.L., Krieg, P.A., Krupenko, S.A., *et al.* (2008). A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. J Neurosci *28*, 264-278.

Campbell, D.B., D'Oronzio, R., Garbett, K., Ebert, P.J., Mirnics, K., Levitt, P., and Persico, A.M. (2007). Disruption of cerebral cortex MET signaling in autism spectrum disorder. Ann Neurol *62*, 243-250.

Campbell, D.B., Li, C., Sutcliffe, J.S., Persico, A.M., and Levitt, P. (2008). Genetic evidence implicating multiple genes in the MET receptor tyrosine kinase pathway in autism spectrum disorder. Autism Res *1*, 159-168.

Campbell, D.B., Sutcliffe, J.S., Ebert, P.J., Militerni, R., Bravaccio, C., Trillo, S., Elia, M., Schneider, C., Melmed, R., Sacco, R.*, et al.* (2006). A genetic variant that disrupts MET transcription is associated with autism. Proc Natl Acad Sci U S A *103*, 16834-16839.
Caton, A., Hacker, A., Naeem, A., Livet, J., Maina, F., Bladt, F., Klein, R., Birchmeier, C., and Guthrie, S. (2000). The branchial arches and HGF are growth-promoting and chemoattractant for cranial motor axons. Development *127*, 1751-1766.

Clark, R.M., Wagler, T.N., Quijada, P., and Doebley, J. (2006). A distant upstream enhancer at the maize domestication gene tb1 has pleiotropic effects on plant and inflorescent architecture. Nat Genet *38*, 594-597.

Deng, W.-G., Zhu, Y., Montero, A., and Wu, K.K. (2003). Quantitative analysis of binding of transcription factor complex to biotinylated DNA probe by a streptavidin-agarose pulldown assay. Anal Biochem *323*, 12-18.

Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. Nucleic Acids Res *11*, 1475-1489.

Dobi, A., Palkovits, M., Ring, M.A., Eitel, A., Palkovits, C.G., Lim, F., and Agoston, D.V. (1997). Sample and probe: a novel approach for identifying development-specific cis-elements of the enkephalin gene. Brain Res Mol Brain Res *52*, 98-111.

Dobreva, G., Chahrour, M., Dautzenberg, M., Chirivella, L., Kanzler, B., Fariñas, I., Karsenty, G., and Grosschedl, R. (2006). SATB2 is a multifunctional determinant of craniofacial patterning and osteoblast differentiation. Cell *125*, 971-986.

Durum, S.K. (2003). Bcl11: sibling rivalry in lymphoid development. Nat Immunol *4*, 512-514.

Ebens, A., Brose, K., Leonardo, E.D., Hanson, M.G., Jr., Bladt, F., Birchmeier, C., Barres, B.A., and Tessier-Lavigne, M. (1996). Hepatocyte growth factor/scatter factor is an axonal chemoattractant and a neurotrophic factor for spinal motor neurons. Neuron *17*, 1157-1172.

Galli, C., Zella, L.A., Fretz, J.A., Fu, Q., Pike, J.W., Weinstein, R.S., Manolagas, S.C., and O'Brien, C.A. (2008). Targeted deletion of a distant transcriptional enhancer of the receptor activator of nuclear factor-kappaB ligand gene reduces bone remodeling and increases bone mass. Endocrinology *149*, 146-153.

Garzotto, D., Giacobini, P., Crepaldi, T., Fasolo, A., and De Marchis, S. (2008). Hepatocyte growth factor regulates migration of olfactory interneuron precursors in the rostral migratory stream through Met-Grb2 coupling. J Neurosci *28*, 5901-5909.

Giacobini, P., Messina, A., Wray, S., Giampietro, C., Crepaldi, T., Carmeliet, P., and Fasolo, A. (2007). Hepatocyte growth factor acts as a motogen and guidance signal

for gonadotropin hormone-releasing hormone-1 neuronal migration. J Neurosci *27*, 431-445.

Gutierrez, H., Dolcet, X., Tolcos, M., and Davies, A. (2004). HGF regulates the development of cortical pyramidal dendrites. Development *131*, 3717-3726.

Gyorgy, A.B., Szemes, M., de Juan Romero, C., Tarabykin, V., and Agoston, D.V. (2008). SATB2 interacts with chromatin-remodeling molecules in differentiating cortical neurons. Eur J Neurosci *27*, 865-873.

Hamanoue, M., Takemoto, N., Matsumoto, K., Nakamura, T., Nakajima, K., and Kohsaka, S. (1996). Neurotrophic effect of hepatocyte growth factor on central nervous system neurons in vitro. J Neurosci Res *43*, 554-564.

Jackson, P.B., Boccuto, L., Skinner, C., Collins, J.S., Neri, G., Gurrieri, F., and Schwartz, C.E. (2009). Further evidence that the rs1858830 C variant in the promoter region of the MET gene is associated with autistic disorder. Autism Res *2*, 232-236.

Judson, M.C., Bergman, M.Y., Campbell, D.B., Eagleson, K.L., and Levitt, P. (2009). Dynamic gene and protein expression patterns of the autism-associated met receptor tyrosine kinase in the developing mouse forebrain. J Comp Neurol *513*, 511-531.

Judson, M.C., Eagleson, K.L., Wang, L., and Levitt, P. (2010). Evidence of cell-nonautonomous changes in dendrite and dendritic spine morphology in the Met-signaling-deficient mouse forebrain. J Comp Neurol.

Kallio, P.J., Poukka, H., Moilanen, A., Jänne, O.A., and Palvimo, J.J. (1995). Androgen receptor-mediated transcriptional regulation in the absence of direct interaction with a specific DNA element. Mol Endocrinol *9*, 1017-1028.

Kleinjan, D.A., and van Heyningen, V. (2005). Long-range control of gene expression: emerging mechanisms and disruption in disease. Am J Hum Genet *76*, 8-32.

Kuo, T.-Y., Chen, C.-Y., and Hsueh, Y.-P. (2010). Bcl11A/CTIP1 mediates the effect of the glutamate receptor on axon branching and dendrite outgrowth. J Neurochem.

Kuo, T.-Y., Hong, C.-J., and Hsueh, Y.-P. (2009). Bcl11A/CTIP1 regulates expression of DCC and MAP1b in control of axon branching and dendrite outgrowth. Mol Cell Neurosci *42*, 195-207.

Kuo, T.-Y., and Hsueh, Y.-P. (2007). Expression of zinc finger transcription factor Bcl11A/Evi9/CTIP1 in rat brain. J Neurosci Res *85*, 1628-1636.

Leid, M., Ishmael, J.E., Avram, D., Shepherd, D., Fraulob, V., and Dollé, P. (2004). CTIP1 and CTIP2 are differentially expressed during mouse embryogenesis. Gene Expr Patterns *4*, 733-739.

Leone, D.P., Srinivasan, K., Chen, B., Alcamo, E., and McConnell, S.K. (2008). The determination of projection neuron identity in the developing cerebral cortex. Curr Opin Neurobiol *18*, 28-35.

Lettice, L.A., Heaney, S.J.H., Purdie, L.A., Li, L., de Beer, P., Oostra, B.A., Goode, D., Elgar, G., Hill, R.E., and de Graaff, E. (2003). A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Human molecular genetics *12*, 1725-1735.

Lewis, S., and Konradi, C. (1996). Analysis of DNA-Protein Interactions in the Nervous System Using the Electrophoretic Mobility Shift Assay. Methods *10*, 301-311.

Lim, C.S., and Walikonis, R.S. (2008). Hepatocyte growth factor and c-Met promote dendritic maturation during hippocampal neuron differentiation via the Akt pathway. Cell Signal *20*, 825-835.

Liu, H., Zhou, M., Luo, X., Zhang, L., Niu, Z., Peng, C., Ma, J., Peng, S., Zhou, H., Xiang, B., *et al.* (2008). Transcriptional regulation of BRD7 expression by Sp1 and c-Myc. BMC Mol Biol *9*, 111.

Nakamura, T., Yamazaki, Y., Saiki, Y., Moriyama, M., Largaespada, D.A., Jenkins, N.A., and Copeland, N.G. (2000). Evi9 encodes a novel zinc finger protein that physically interacts with BCL6, a known human B-cell proto-oncogene product. Molecular and Cellular Biology *20*, 3178-3186.

Poon, K., and Macgregor, R.B., Jr. (1998). Unusual behavior exhibited by multistranded guanine-rich DNA complexes. Biopolymers *45*, 427-434.

Powell, E.M., Mars, W.M., and Levitt, P. (2001). Hepatocyte growth factor/scatter factor is a motogen for interneurons migrating from the ventral to dorsal telencephalon. Neuron *30*, 79-89.

Sharma, R.P., Gavin, D.P., and Grayson, D.R. (2010). CpG Methylation in Neurons: Message, Memory, or Mask? Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology.

Sousa, I., Clark, T.G., Toma, C., Kobayashi, K., Choma, M., Holt, R., Sykes, N.H., Lamb, J.A., Bailey, A.J., Battaglia, A., *et al.* (2009). MET and autism susceptibility: family and case-control studies. Eur J Hum Genet *17*, 749-758.

Stauff, D.L., Torres, V.J., and Skaar, E.P. (2007). Signaling and DNA-binding activities of the Staphylococcus aureus HssR-HssS two-component system required for heme sensing. J Biol Chem *282*, 26111-26121.

Szemes, M., Gyorgy, A., Paweletz, C., Dobi, A., and Agoston, D.V. (2006). Isolation and characterization of SATB2, a novel AT-rich DNA binding protein expressed in development- and cell-specific manner in the rat brain. Neurochem Res *31*, 237-246.

Thanseem, I., Nakamura, K., Miyachi, T., Toyota, T., Yamada, S., Tsujii, M., Tsuchiya, K.J., Anitha, A., Iwayama, Y., Yamada, K.*, et al.* (2010). Further evidence for the role of MET in autism susceptibility. Neuroscience Research.

Tyndall, S.J., Patel, S.J., and Walikonis, R.S. (2007). Hepatocyte growth factor-induced enhancement of dendritic branching is blocked by inhibitors of N-methyl-D-aspartate receptors and calcium/calmodulin-dependent kinases. J Neurosci Res *85*, 2343-2351.

Tyndall, S.J., and Walikonis, R.S. (2006). The receptor tyrosine kinase Met and its ligand hepatocyte growth factor are clustered at excitatory synapses and can enhance clustering of synaptic proteins. Cell Cycle *5*, 1560-1568.

Yet, S.F., McA'Nulty, M.M., Folta, S.C., Yen, H.W., Yoshizumi, M., Hsieh, C.M., Layne, M.D., Chin, M.T., Wang, H., Perrella, M.A.*, et al.* (1998). Human EZF, a Krüppel-like zinc finger protein, is expressed in vascular endothelial cells and contains transcriptional activation and repression domains. J Biol Chem *273*, 1026-1031.

Zhang, L., Himi, T., Morita, I., and Murota, S. (2000). Hepatocyte growth factor protects cultured rat cerebellar granule neurons from apoptosis via the phosphatidylinositol-3 kinase/Akt pathway. J Neurosci Res *59*, 489-496.