Journal of
Emerging Investigators

Transcriptional Regulators are Upregulated in the Substantia Nigra of Parkinson's Disease Patients

Marianne Cowherd[1] and Inhan Lee[2]

[1]Community High School, Ann Arbor, MI

[2]miRcore, Ann Arbor, MI

Summary

Parkinson's disease (PD) affects approximately 10 million people worldwide with tremors, bradykinesia, apathy, memory loss, and language issues. Though such symptoms are due to the loss of the substantia nigra (SN) brain region, the ultimate causes and complete pathology are unknown. To understand the global gene expression changes in SN, microarray expression data from the SN tissue of 9 controls and 16 PD patients were compared, and significantly upregulated and downregulated genes were identified. Among the upregulated genes, a network of 33 interacting genes centered around the cAMP-response element binding protein (*CREBBP*) was found. The downstream effects of increased CREBBP-related transcription and the resulting protein levels may result in PD symptoms, making CREBBP a potential therapeutic target due to its central role in the interactive gene network affected in PD SN.

Received: June 9, 2014; **Accepted:** September 28, 2015; **Published:** November 15, 2015

Introduction

Parkinson's disease (PD) is a progressive neurodegenerative disease that affects approximately 10 million people worldwide. Symptoms involve movement and cognition and commonly include tremors, bradykinesia, apathy, memory loss, and language issues. Depression is common among patients and caretakers. Treatments such as levodopa and deep-brain stimulation (DBS) are effective, but the progressive deterioration of treatment responses after 5 to 10 years is an unresolved limitation of these treatments (1). Abnormal levels of dopamine, a motor control and reward pathway neurotransmitter, beginning in the SN is a proximate cause of PD symptoms (2). The SN, in the basal ganglia of the brain, is a major dopaminergic center.

Gene expression as a marker for PD and other neurological conditions is an established practice (3). Significant gene expression dysregulation in the SN and in the striatum has been described, particularly decreased expression in PD synapses. Protein degradation has been found to be upregulated (4). Mutations in *SNCA* (5), *LRRK2* (6), and *GBA* (6) have also been identified as familial markers of PD. *SNCA* encodes alpha-synuclein, a protein found in presynaptic terminals that may regulate vesicle presence and dopamine release. Eighteen *SNCA* mutations have been associated with PD and, although the exact pathogenic mechanism is not confirmed, mutated alpha-synuclein is the major component of protein aggregates, called Lewy bodies, that are often found in PD brains and may contribute to cell death. Abnormal *SNCA* may also impair protein degradation (7). *LRRK2* encodes the dardarin protein, which has functions related to protein phosphorylation and GTPase activity. More than 100 mutations to *LRRK2* have been identified as markers of late-onset familial PD, but a mechanism of action is not clear. *GBA* encodes an enzyme that is active in lysosomes, and while the details on how it contributes to PD remain unclear, it is speculated that impaired lysosome function prevents nerve-damaging proteins from being degraded. These proteins then go on to damage dopaminergic neurons. None of these genes was identified in this analysis, which is not unusual because their associations with PD are mutation-based and this study looked only at expression.

This study identifies *CREBBP*, an upregulated gene, as interacting with most other upregulated genes. *CREBBP* encodes CREB-binding protein, which binds to and enhances the activity of phosphorylated cAMP-response element binding protein (CREB), a phosphorylation-dependent transcriptional regulator. CREBBP further increases transcription via CREB by acting as a scaffold for other proteins to interact with CREB. CREBBP also acetylates histone proteins, removing the positive charge that attracts negatively-charged DNA, making the DNA region more accessible to transcription. In this analysis, *CREBBP* forms the center of a large web of interacting upregulated genes. It is present in a majority of gene ontologies (GO) (biological processes, molecular functions, and cellular components) that are overrepresented among the upregulated and among the upregulated genes

constituting the *CREBBP*-centered interaction network. Many of the GO terms overrepresented in upregulated genes are related to transcriptional regulation, including chromatin organization and modification, chromosome organization, and histone modification. An increase in transcriptional activity, specifically by the cAMP-CREBBP-CREB pathway, may be an important pathological process in PD and may be a possible therapeutic target.

Results

We analyzed publicly available microarray expression data (GSE7621) from the postmortem SN tissues of 9 controls and 16 Parkinson's disease patients (8). **Table 1** shows the sample information. A moderated t-test within the GEO2R platform provided the top 250 significantly differentially expressed genes (ranked by *p*-value) between control and Parkinson cases (top 25 microarray probe IDs in **Table 2**). A moderated t-test compares two sets of data and returns information on the significance of the differences between those data. The t-test was used to identify which patterns of differential expression between disease and control samples were unlikely due to chance alone and are therefore possibly relevant to disease state. Of those, 88 genes were downregulated in PD and 162 genes were upregulated in PD. **Figure 1** shows the expression profiles of 25 samples for six upregulated genes in PD. Genes that are downregulated in PD are less expressed in the PD samples as a group when compared to the control samples as a group. Upregulated genes have a higher expression in PD samples than in control samples.

CREBBP regulates transcription factor access to DNA by modifying histones and by binding to CREB, which is involved in chromatin remodeling. XPO1 mediates the transport of specific snRNAs and proteins

Sample Name	GEO Accession	Disease State	Sex
Substantia nigra normal rep1	GSM184354	Control	female
Substantia nigra normal rep2	GSM184355	Control	male
Substantia nigra normal rep3	GSM184356	Control	female
Substantia nigra normal rep4	GSM184357	Control	female
Substantia nigra normal rep5	GSM184358	Control	female
Substantia nigra normal rep6	GSM184359	Control	female
Substantia nigra normal rep7	GSM184360	Control	male
Substantia nigra normal rep8	GSM184361	Control	male
Substantia nigra normal rep9	GSM184362	Control	male
Substantia nigra PD rep1	GSM184363	Parkinson's Disease	male
Substantia nigra PD rep2	GSM184364	Parkinson's Disease	male
Substantia nigra PD rep3	GSM184365	Parkinson's Disease	male
Substantia nigra PD rep4	GSM184366	Parkinson's Disease	male
Substantia nigra PD rep5	GSM184367	Parkinson's Disease	male
Substantia nigra PD rep6	GSM184368	Parkinson's Disease	female
Substantia nigra PD rep7	GSM184369	Parkinson's Disease	male
Substantia nigra PD rep8	GSM184370	Parkinson's Disease	male
Substantia nigra PD rep9	GSM184371	Parkinson's Disease	female
Substantia nigra PD rep10	GSM184372	Parkinson's Disease	male
Substantia nigra PD rep11	GSM184373	Parkinson's Disease	male
Substantia nigra PD rep12	GSM184374	Parkinson's Disease	female
Substantia nigra PD rep13	GSM184375	Parkinson's Disease	male
Substantia nigra PD rep14	GSM184376	Parkinson's Disease	male
Substantia nigra PD rep15	GSM184377	Parkinson's Disease	male
Substantia nigra PD rep16	GSM184378	Parkinson's Disease	male

Table 1: Gene expression data was extracted from 25 postmortem SN tissue samples from 9 control (5 female, 4 male) and 16 PD (3 female, 13 male) cases.

with leucine-rich nuclear export signals out of the nuclear membrane. SETD2 methylates histones and may also have a function related to the phosphorylation of RNA polymerase II. NCOR1 effects transcription inhibition by condensing the chromatin around specific thyroid hormone and retinoic acid receptors. ANAPC16 is part of a complex that controls cell cycle progression through ubiquitination mediation. AXIN1 is a member of several different complexes with functions including the negative regulation of Wnt signaling, apoptosis induction, and phosphorylation/ubiquitination regulation.

We separately analyzed gene-gene interactions among 88 downregulated and 162 upregulated genes. Viewing interactions in String showed few groups of interacting downregulated genes. String's database contains known and predicted protein-protein interactions identified from published articles, experimental data, and co-expression based computational prediction methods. When a group of genes are input together in

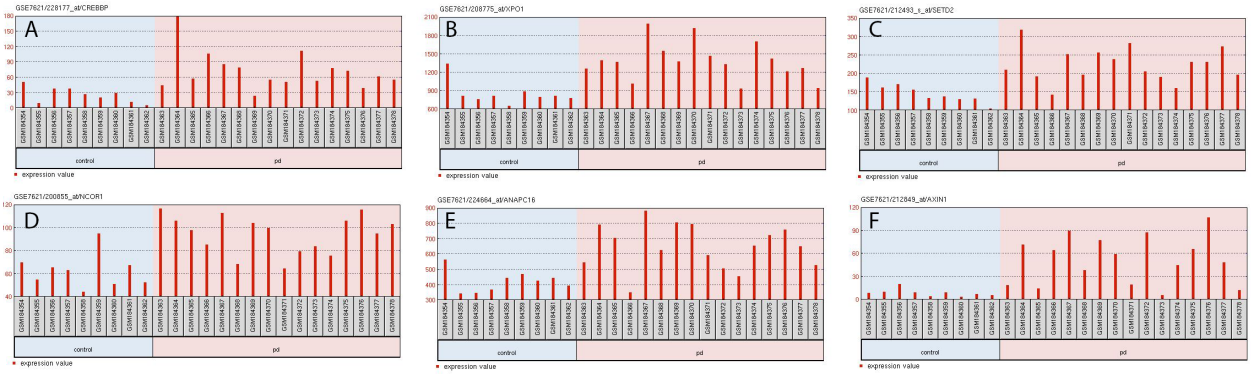

Figure 1: Expression Profiles for Six Genes Significantly Upregulated in PD. Control sample expression in blue and PD sample expression in pink. Genes shown are (A) *CREBBP*, (B) *XPO1*, (C) *SETD2*, (D) *NCOR1*, (E) *ANAPC16*, and (F) *AXIN1*. Expression is from publicly available dataset GSE7621 and was measured using Affymetrix Human Genome U133 Plus microarrays. Expression values are normalized to an arbitrarily-set trimmed mean of 100 using global scaling, which is a standard normalization technique for this type of data.

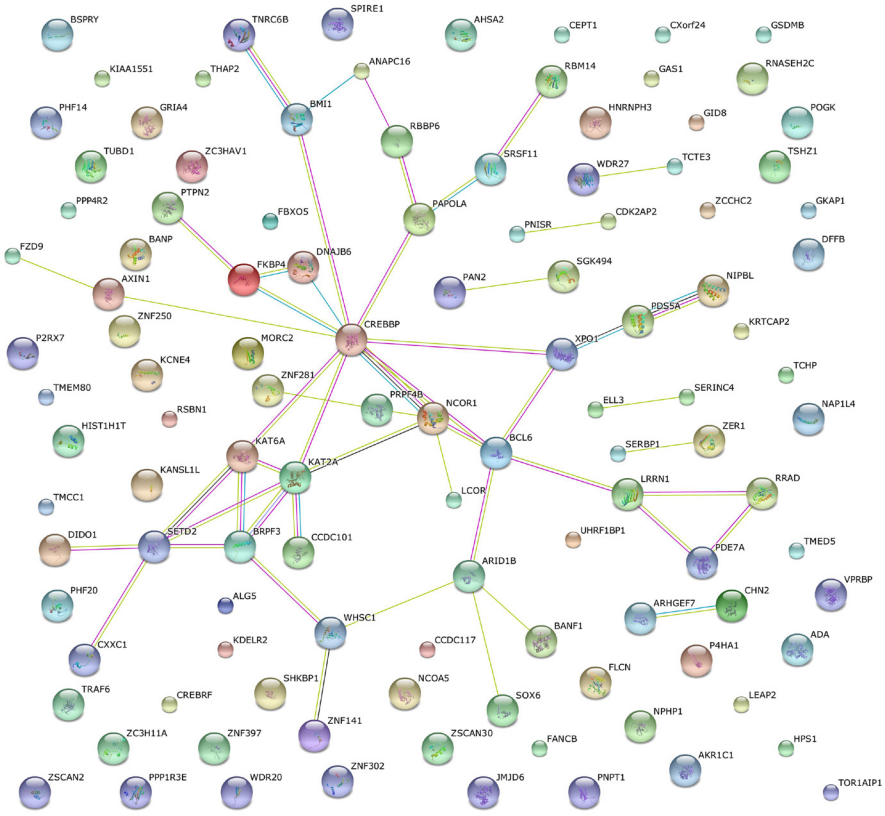

Figure 2: Interactions Between Upregulated Genes. A 33-gene interaction network with *CREBBP* as a central hub, in addition to 3 groups of 2-gene interactions, are represented with a circle for each gene and a line for each type of interaction.

String, any interactions in the database between any of the listed genes are identified and displayed in a web. Among the upregulated genes, a group of 33 were connected to each other, centered on CREBBP (**Figure 2**).

BiNGO identified as overrepresented 8 biological processes, 0 cellular components, and 1 molecular function among the downregulated genes, of which many systems were related to catecholamine processing and amine transport. BiNGO is a plugin for Cytoscape, a software platform for visualizing computational biology data. BiNGO identifies statistically overrepresented GO categories by comparing the prevalence of genes in each ontology between those input and all those within the human genome. BiNGO's human genome contains 14,303 genes, despite there being 20,000 to 25,000 genes in the actual genome, because only genes with well-defined ontologies are included.

Among the upregulated genes, BiNGO identified 44 biological processes, 15 molecular functions, and 12 cellular components as overrepresented (**Table 3**). The group of 33 String-identified interacting upregulated genes had 76 overrepresented biological processes, 53 overrepresented molecular functions, and 24 overrepresented cellular components (**Table 4**).

The GO categories in the upregulated genes and in the group of 33 connected upregulated genes contain functions of regulating expressions at the transcriptional level, including chromatin organization and modification, chromosome organization, and histone modification. *CREBBP*, the gene at the center of the interaction networks, is within 54.5% of biological processes, 80% of the molecular functions, and 100% of the cellular components pathways that are among those that the upregulated genes function in. Among overrepresented ontologies in the group of 33, CREBBP is found in 53.9% of biological pathways, 74.0% of molecular functions, and 91.7% of cellular components.

Discussion

The upregulation of these transcriptional regulators in PD SN characterizes a difference in the molecular environment of PD and control tissues. CREBBP's role as a common regulatory factor in many upregulated genes suggests an upstream cause of the upregulation. Both male and female samples were present with distributions of 5 female and 4 male, and 3 female and 13 male, in the control and PD groups, respectively. Of the genes studied in depth, none were found on the Y chromosome. The upregulated and downregulated genes are potential surrogate biomarkers for PD, as they correlate with disease and may or may not be related to the pathology.

In the few small groups of interacting genes in the downregulated list, one gene may be causally related to PD and have a regulatory effect on the others. Given the lack of interaction between them, it is unlikely that a single pathway or regulator is affecting the expression of all of the downregulated genes. It is also less likely that targeting one of the downregulated genes would have far-reaching therapeutic effects. Testing for overrepresented

GO terms among the downregulated genes indicated systems related to catecholamine processing and amine transport. Dopamine, a catecholamine, is affected by both of these processes, suggesting a relationship between the downregulation of these genes and the dopamine pathology in PD. The downregulation of catecholamine processing in amine transport aligns with the low levels of dopamine in PD patients. These genes relevant to dopamine production that are affected in PD may be therapeutic targets or relevant markers for the efficacy of other natural dopamine production–increasing treatments.

The upregulated genes, however, interacted with each other to a much greater degree. Interactions between different branches of the 33-member, CREBBP-centered group were common. This may be partly due to CREBBP's function as a transcriptional regulator, but the presence of so many interacting genes in the web points to the significance of the interaction. Testing for overrepresented GO terms among all of the upregulated genes and among the group of 33 indicated many transcription regulation–related pathways. CREBBP was present in the majority of these pathways. Upregulation of CREBBP increases transcription via CREB, as well as by increasing DNA availability to transcription factors by histone acetylation. CREBBP activity in the brain has downstream effects on a wide variety of proteins, but it may increase the levels of proteins found in Lewy bodies or proteins with other effects on the function of SN neurons. An increase in CREBBP activity, the expression of interacting genes with functions relevant to cAMP-mediated transcriptional regulation, and histone-DNA interactions all present in PD patients suggest molecular markers or pathologies of PD in the SN.

Further functional and biological studies are necessary to confirm the relevance of these statistical results in human brain tissue. Disrupting the expression or functionality of CREBBP and measuring the resulting expression of the identified transcriptional regulators would test the relationship between CREBBP activity and transcriptional regulator activity. Inducing upregulation of these transcriptional regulators in mouse models and monitoring for Parkinson's disease symptoms would test if this expression pattern is a functional cause of the disease. Inducing PD by other methods (MPTP) and measuring the expression levels of these genes help determine whether the upregulation of the genes is a result of PD or a result of other ultimate causes of PD, with the qualification that induced PD in models cannot be assumed to mimic the human disease in all ways. The causes of PD are not well understood and exploring transcription regulation as a possible cause or reaction may help elucidate the underlying mechanisms of PD.

Methods

Dataset GSE7621 (8) contains gene expression data from 9 control and 16 Parkinson (**Table 1**) postmortem SN tissues collected by hybridization on Affymetrix Human Genome U133 Plus 2.0 arrays and made public in 2007. Expression is represented numerically, based on analysis in Microarray Suite 5.0 with Affymetrix default settings and global scaling normalization with a trimmed target mean of 100. Global scaling uses regression analysis to transform sets of data onto the same frame of reference so that the raw numbers can be reasonably compared to each other, in this case to a mean of 100 with outliers excluded (trimmed) in mean calculation. Of the control samples, 5 were female and 4 were male; of the PD samples, 3 were female and 13 were male. Microarrays use oligonucleotide probes to measure amounts of labeled cDNA reverse-transcribed from sample mRNA to provide mRNA amount-based gene expression data. Significant differentially expressed genes were obtained based on a t-test run using GEO2R, an online interactive data comparison tool from the NCBI (9, 10).

The 250 most significantly differentially expressed (as sorted by *p*-value) genes when comparing control and Parkinson samples were separated into upregulated or downregulated in Parkinson groups as indicated by fold change sign. String, an online gene interaction viewer, was used to map the interactions between genes within each list (11).

BiNGO, an application within the open-source genetic information software program Cytoscape (12,13), was used to run a hypergeometric test with *p*-value 0.05 and FDR (false discovery rate) correction to identify overrepresented GO categories within the downregulated and upregulated groups of genes. Large interaction groups identified in String were also run through BiNGO to identify overrepresented ontologies. Connections and actions of genes were explored and confirmed using GeneCards (14-18), UCSC Genome Browser (19), and the EpiTect ChIP qPCR Primers transcription factors search tool from QIAGEN (20).

Acknowledgments

We appreciate the support of University of Michigan WISE (Women in Science and Engineering), the Prechter Family Fund, and Zingerman's Community of Business. We also thank miRcore volunteers for voting on Parkinson's disease as the year's disease of focus and other GIDAS (Genes in Disease and Symptoms) club members for encouraging this research throughout the year.

References

1. Marsden CD, Parkes JD. (1977). Success and

problems of long-term levodopa therapy in Parkinson's disease. *Lancet* 1(8007): 345-49.
2. Lotharius J, Brundin P. (2002). Pathogenesis of Parkinson's disease: dopamine, vesicles and alpha-synuclein. *Nat Rev Neurosci* 3(12): 932-42.
3. Papapetropoulos S., Shehadeh L, McCorquodale D (2007). Optimizing human post-mortem brain tissue gene expression profiling in Parkinson's disease and other neurodegenerative disorders: From target "fishing" to translational breakthroughs. *J. Neurosci. Res.* 85: 3013–24.
4. Miller RM, *et al*. (2006). Robust dysregulation of gene expression in substantia nigra and striatum in Parkinson's disease. *Neurobiol Dis* 21(2): 305-13.
5. Polymeropoulos MH, *et al*. (1997). Mutation in the *a*-synuclein gene identified in families with Parkinson's disease. *Science* 276(5321): 2045-47.
6. Lees AJ, et al.. (2009). Parkinson's disease. *Lancet* 373(9680): 2055-2066.
7. Spillantini MG, *et al*. (1997). [alpha]-Synuclein in Lewy bodies. *Nature* 388(6645): 839-840.
8. Lesnick TG, *et al*. (2007). A genomic pathway approach to a complex disease: axon guidance and Parkinson disease. *PLoS Genet* 3(6):e98.
9. Edgar R, Domrachev M, Lash AE. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 30.1: 207-10.
10. Barrett T, *et al*. (2013). NCBI GEO: archive for functional genomics data sets--update. *Nucleic Acids Res.* 41:D991-5.
11. Szklarczyk D, *et al*. The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic Acids Res.* 39: D561–D568.
12. Shannon P, *et al*. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* 13(11): 2498-504.
13. Maere S, *et al*. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* 21(16): 3448-9.
14. Rebhan M., Chalifa-Caspi V., Prilusky J, Lancet D. (1997) GeneCards: integrating information about genes, proteins and diseases. *Trends in Genetics* 13: 163.
15. Safran M *et al*. (2010). GeneCards Version 3: the human gene integrator. *Database* doi: 10.1093/database/baq020.
16. Lancet D, Safran M, Olender T, Dalah I, Iny-Stein T, Inger A, Harel A, Stelzer G. "GeneCards tools for combinatorial annotation and dissemination of human genome information." *GIACS Conference on Data in Complex Systems*. April, (2008).
17. Stelzer G, Harel A, Dalah A, Rosen N, Shmoish M, Iny Stein T, Sirota A, Madi A, Safran M, Lancet D. "GeneCards: One stop site for human gene research." *FISEB (ILANIT)*, January, (2008).
18. Safran M, *et al*. (2002). GeneCards 2002: towards a complete, object-oriented, human gene compendium. *Bioinformatics* 18(11): 1542-3.
19. Karolchik D, *et al*. (2014). The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res.* 42(1): D764-70.
20. Qiagen. "EpiTect ChIP qPCR Primers Transcription Factor." SABiosciences.

ID	*P*-value	t	logFC	Gene	Gene Name
1557104_at	9.51E-08	7.34	3.096	ZSCAN30	zinc finger and SCAN domain containing 30
228845_at	4.15E-06	-5.83	-1.394	PLBD2	phospholipase B domain containing 2
235282_at	4.65E-06	-5.78	-1.036	LOC100506295	uncharacterized LOC100506295
1553549_at	5.90E-06	-5.69	-2.131	VN1R2	vomeronasal 1 receptor 2
209797_at	7.70E-06	-5.59	-0.865	CNPY2	canopy FGF signaling regulator 2
243806_at	1.04E-05	-5.47	-2.452	----	
212849_at	1.25E-05	5.4	2.315	AXIN1	axin 1
205357_s_at	1.56E-05	-5.31	-2.18	AGTR1	angiotensin II receptor, type 1
1553328_a_at	1.73E-05	-5.27	-3.15	SLC18A2	solute carrier family 18 (vesicular monoamine transporter), member 2
223588_at	2.00E-05	5.22	1.279	THAP2	THAP domain containing, apoptosis associated protein 2
223766_at	2.01E-05	-5.22	-0.891	LOC100133130	PRO1102
201436_at	2.27E-05	-5.17	-0.666	EIF4E	eukaryotic translation initiation factor 4E
1565567_at	2.49E-05	5.13	0.747	----	
227321_at	2.58E-05	-5.12	-1.027	GATS	GATS, stromal antigen 3 opposite strand
1557882_at	2.73E-05	5.1	1.978	----	
222315_at	3.05E-05	5.06	2.199	LOC100510224	hypothetical LOC100510224
242263_at	3.30E-05	5.03	0.918	TMED5	transmembrane emp24 protein transportdomain contaiing 5
202127_at	3.72E-05	4.98	1.082	PRPF4B	pre-mRNA processing factor 4B
213789_at	3.76E-05	-4.98	-1.891	EBP	emopamil binding protein (sterol isomerase)
205857_at	3.92E-05	-4.96	-3.055	SLC18A2	solute carrier family 18 (vesicular monoamine transporter), member 2
208319_s_at	4.19E-05	-4.93	-2.445	RBM3	RNA binding motif (RNP1, RRM) protein 3
240310_at	4.33E-05	4.92	1.012	TOR1AIP1	torsin A interacting protein 1
230741_at	4.44E-05	4.91	1.028	----	
233186_s_at	4.55E-05	4.9	0.925	BANP	BTG3 associated nuclear protein
1569700_s_at	4.65E-05	-4.9	-1.839	AK7	adenylate kinase 7

Table 2: The most significant differentially expressed genes, identified by a moderated t-test ranked by *p*-value.

DOWNREGULATED			
Biological Process	**Sample Presence**	**Genome Presence**	**Genes**
catecholamine biosynthesis	3/54, 5.5%	12/14302, 0.0%	DDC SLC6A3 TH
dopamine biosynthesis	2/54, 3.7%	7/14302, 0.0%	SLC6A3 TH
amine biosynthesis	4/54, 7.4%	86/14302, 0.6%	DDC SLC6A3 TH ASL
UPREGULATED			
Biological Process	**Sample Presence**	**Genome Presence**	**Genes**
chromosome organization	19/88, 21.5%	525/14303, 3.6%	KAT2A BMI1 DFFB CREBBP BANP WHSC1 ARID1B SOX6 NAP1L4 CXXC1 HIST1H1T NIPBL BRPF3 JMJD6 RBM14 SETD2 NCOR1 FANCB KAT6A
chromatin organization	16/88, 18.1%	413/14303, 2.8%	KAT2A BMI1 CREBBP WHSC1 BANP ARID1B SOX6 NAP1L4 CXXC1 HIST1H1T BRPF3 JMJD6 RBM14 SETD2 NCOR1 KAT6A
chromatin modification	13/88, 14.7%	311/14303, 2.1%	BMI1 KAT2A CREBBP WHSC1 BANP ARID1B CXXC1 BRPF3 JMJD6 RBM14 SETD2 NCOR1 KAT6A
histone modification	8/88, 9.0%	143/14303, 0.9%	KAT2A BRPF3 JMJD6 CREBBP RBM14 SETD2 KAT6A CXXC1
covalent chromatin modification	8/88, 9.0%	146/14303, 1.0%	KAT2A BRPF3 JMJD6 CREBBP RBM14 SETD2 KAT6A CXXC1
regulation of transcription	32/88, 36.3%	2621/14303, 18.3%	BMI1 TSHZ1 ZNF250 PHF20 SOX6 CXXC1 NIPBL CCDC101 ZNF302 ZNF397 BCL6 TRAF6 FANCB KAT2A ZNF281 CREBBP ZNF141 WHSC1 BANP ARID1B ZSCAN2 UHRF1 POGK JMJD6 ZSCAN30 NCOA5 RBM14 SETD2 LCOR NCOR1 DNAJB6 KAT6A
Molecular Function	**Sample Presence**	**Genome Presence**	**Genes**
nucleic acid binding	40/92, 43.4%	3251/15440, 21.0%	XPO1 TSHZ1 ZC3HAV1 PNPT1 SRSF11 ZNF250 PHF20 SOX6 BANF1 CXXC1 ZNF302 ZNF397 BCL6 THAP2 TNRC6B FANCB ZNF281 PAN2 ZCCHC2 CREBBP ZNF141 BANP WHSC1 ARID1B RBBP6 ZSCAN2 PAPOLA HIST1H1T UHRF1 HNRNPH3 POGK ZSCAN30 SERBP1 ZC3H11A RBM14 SETD2 LCOR NCOR1 DNAJB6 KAT6A
transcription factor activity	22/92, 23.9%	1506/15440, 9.7%	KAT2A ZNF281 TSHZ1 CREBBP ZNF141 ARID1B SOX6 ZSCAN2 CXXC1 UHRF1 NIPBL ZSCAN30 LRRN1 ZNF397 BCL6 RBM14 LCOR TRAF6 NCOR1 FANCB DNAJB6 KAT6A
histone acetyltransferase activity	4/92, 4.3%	41/15440, 0.2%	KAT2A NIPBL TRAF6 NCOR1

Table 3: Overrepresented GO terms (molecular functions, biological processes, and cellular components) among all downregulated (top section) and all upregulated (middle section) genes.

Biological Process	Sample Presence	Genome Presence	Genes
chromatin modification	11/34, 32.3%	311/14306, 2.1%	KAT2A BMI1 BRPF3 CREBBP WHSC1 ARID1B SETD2 RBM14 NCOR1 KAT6A CXXC1
chromatin organization	12/34, 35.2%	413/14306, 2.8%	KAT2A BMI1 BRPF3 CREBBP WHSC1 ARID1B SOX6 SETD2 RBM14 NCOR1 KAT6A CXXC1
chromosome organization	12/34, 35.2%	525/14306, 3.6%	KAT2A BMI1 BRPF3 CREBBP WHSC1 ARID1B SOX6 SETD2 RBM14 NCOR1 KAT6A CXXC1
histone modification	7/34, 20.5%	143/14306, 0.9%	KAT2A BRPF3 CREBBP SETD2 RBM14 KAT6A CXXC1
covalent chromatin modification	7/34, 20.5%	146/14306, 1.0%	KAT2A BRPF3 CREBBP SETD2 RBM14 KAT6A CXXC1
histone acetylation	4/34, 11.7%	50/14306, 0.3%	KAT2A BRPF3 CREBBP KAT6A
protein amino acid acetylation	4/34, 11.7%	55/14306, 0.3%	KAT2A BRPF3 CREBBP KAT6A
protein amino acid acylation	4/34, 11.7%	65/14306, 0.4%	KAT2A BRPF3 CREBBP KAT6A
regulation of gene expression	18/34, 52.9%	2928/14306, 20.4%	BMI1 KAT2A ZNF281 CREBBP ZNF141 WHSC1 SOX6 ARID1B CXXC1 CCDC101 BCL6 LCOR SETD2 RBM14 TNRC6B NCOR1 DNAJB6 KAT6A
Molecular Function	**Sample Presence**	**Genome Presence**	**Genes**
nucleic acid binding	20/33, 60.6%	3252/15443, 21.0%	ZNF281 XPO1 SRSF11 CREBBP ZNF141 WHSC1 SOX6 ARID1B RBBP6 BANF1 CXXC1 PAPOLA BCL6 LCOR SETD2 RBM14 TNRC6B NCOR1 DNAJB6 KAT6A
transcription regulator activity	14/33, 42.4%	1507/15443, 9.7%	KAT2A ZNF281 CREBBP ZNF141 SOX6 ARID1B CXXC1 LRRN1 BCL6 RBM14 LCOR NCOR1 DNAJB6 KAT6A
histone acetyltransferase activity	3/33, 9.0%	20/15443, 0.1%	KAT2A CREBBP KAT6A
transcription activator activity	3/33, 9.0%	20/15443, 0.1%	KAT2A CREBBP KAT6A

Table 4: Overrepresented GO terms (molecular functions, biological processes, and cellular components) among 33 interacting upregulated genes.