GLOBAL EFFECTS OF A PD RISK-SNP AT THE ALPHA-SYNUCLEIN LOCUS

Jordan Prahl*, Steven E. Pierce*, Edwin JC van der Schans, Gerhard A. Coetzee**, & Trevor Tyson**

* Co-first authors, ** Co-last authors

Department of Neurodegenerative Research, Van Andel Institute, Grand Rapid, MI, USA

*Correspondence

jordan.prahl@vai.org

SUMMARY

One of the most significant Parkinson's disease (PD) risk variants, rs356182, is located at the PD-associated locus near the alpha-synuclein encoding gene, *SNCA*. *SNCA*-proximal variants, including rs356182, are thought to function in PD via allele-specific regulatory effects on SNCA expression. However, this interpretation discounts the complex activity of genetic enhancers and possible nonconical effects of alpha-synuclein. Here we investigate a novel risk mechanism for rs356182. We use CRISPR-Cas9 in LUHMES cells, a model for dopaminergic neurons, to generate precise hemizygous lesions at rs356182. The PD-protective (A/-), PD-risk (G/-), and WT (A/G) strains are differentiated into dopaminergic neurons then compared transcriptionally and morphologically. We observe effects not typically ascribed to *SNCA*; hundreds of differentially expressed genes associated with neuronal differentiation and axonogenesis. Together, the data implicate a risk mechanism for rs356182 in which the risk-allele (G) is associated with abnormal neuronal differentiation. We speculate the disease-relevant effect originates as a diminished population of DA neurons leading to the predisposition for PD later in life.

Keywords: Parkinson's, rs356182, SNCA, differentiation, LUHMES, epigenetics, GWAS, neurogenesis, morphology

INTRODUCTION

With each consecutive genome-wide association study (GWAS), Parkinson's disease (PD) is associated with an ever-increasing number of single-nucleotide polymorphisms (SNPs) [1, 2]. Because these associations are based on large epidemiological studies, descriptions of risk mechanisms require follow-up experiments. For most PD risk loci the identification of the causal SNP(s) has not been definitive, with few notable exceptions [3-6]. Most GWAS-identified risk loci are located in non-coding regions of the genome, further complicating this process [7]. One of the the top GWAS-identified PD risk-SNP, rs356182, is an example of a prominent SNP lacking a confirmed mechanism.

Rs356182 has a meta p-value of 1.85×10^{-82}, making it the most significant association among non-coding PD risk-SNP [8]. The risk allele (G) is robustly represented in the population (frequency=40-50%) but also has one of the highest odds ratios (1.34) of any PD-associated SNP [8, 9]. This means the proportion of G alleles in the PD population compared to healthy controls is unusually high even compared to statistically significant risk signals. Enticingly, this SNP is within a genetic enhancer and located close to *SNCA*, the gene encoding alpha-synuclein, which is one of the first proteins linked to familial PD as well as neuropathologically present in Lewy bodies (Figure 1A) [7, 10, 11]. These data have led to the assumption that rs356182 confers risk for PD by directly and even exclusively regulating the expression of *SNCA*. However, this narrow view of the rs356182 mechanism neglects the potential for additional enhancer-promotor interactions, as well as secondary and tertiary effects of transcription factor (TF) misregulation.

Unlike genes and promotors, enhancers are highly dynamic genetic elements, in the sense of functioning both cryptically via multiple TFs and with different gene targets in different cell types and stages. Enhancer variants have continuous effects (graded responses on gene expression) as apposed the binary effects of gene triplet code variation. Furthermore, enhancers can interact with the closest gene promotor on linear DNA (*SNCA* in this case). However, up to two-thirds of enhancers skip the nearest gene entirely and interact exclusively with distal promotors [12, 13]. Additionally, enhancers may interact with multiple promoters, and a single promoter often interacts with numerous enhancers in the three-dimensional space of chromatin [14, 15]. Enhancers regulate the expression of target genes by recruiting TFs to

specific regulatory binding motifs. In this way, SNPs may modulate downstream gene expression by changing the binding affinity of TFs to a particular allele and alter the strength of the enhancer-promotor interactions.

Lund Human Mesencephalic (LUHMES; a.k.a. MESC2.10) cells offer a unique platform for studying immature human midbrain neurons and the epigenetic architecture present during neuronal development, due to their isolation from 8-week old fetal mesencephalic tissue [16]. LUHMES cells are immortalized in a stem-like state with the v-MYC oncogene, which, when deactivated allows the cells to enter the differentiation pathway and reach their terminal state within 6 days. In their differentiated state, LUHMES cells have a neuronal phenotype, displaying long neuronal projections (Figure 1B), and express the rate-limiting enzyme in dopamine production, tyrosine hydroxylase (TH) (Figure 1C). Beneficial to this research, LUHMES cells are heterozygous for rs356182 with one chromosomal homolog possessing the PD risk-conferring allele, guanine (G), and the other homolog possessing the protective allele, adenosine (A), on the Watson strand. Interestingly, when differentiated, these cells also contain an enhancer surrounding the rs356182 SNP [7]. This makes LUHMES cells an ideal model for investigating PD pathology or predisposition originating very early in development, including neuronal differentiation-based mechanisms.

This study seeks to elucidate the mechanism surrounding rs356182 by examining allele-specific gene regulation and changes to differentiated morphology associated with this SNP. To that end, we used CRISPR-Cas9 to generate mono-allelic lesions of rs356182 in LUHMES cells (Figure 2A), and examined the changes in gene expression and morphology therein. We observed a novel rs356182 risk mechanism not previously attributed to *SNCA*, pertaining to neuronal differentiation processes.
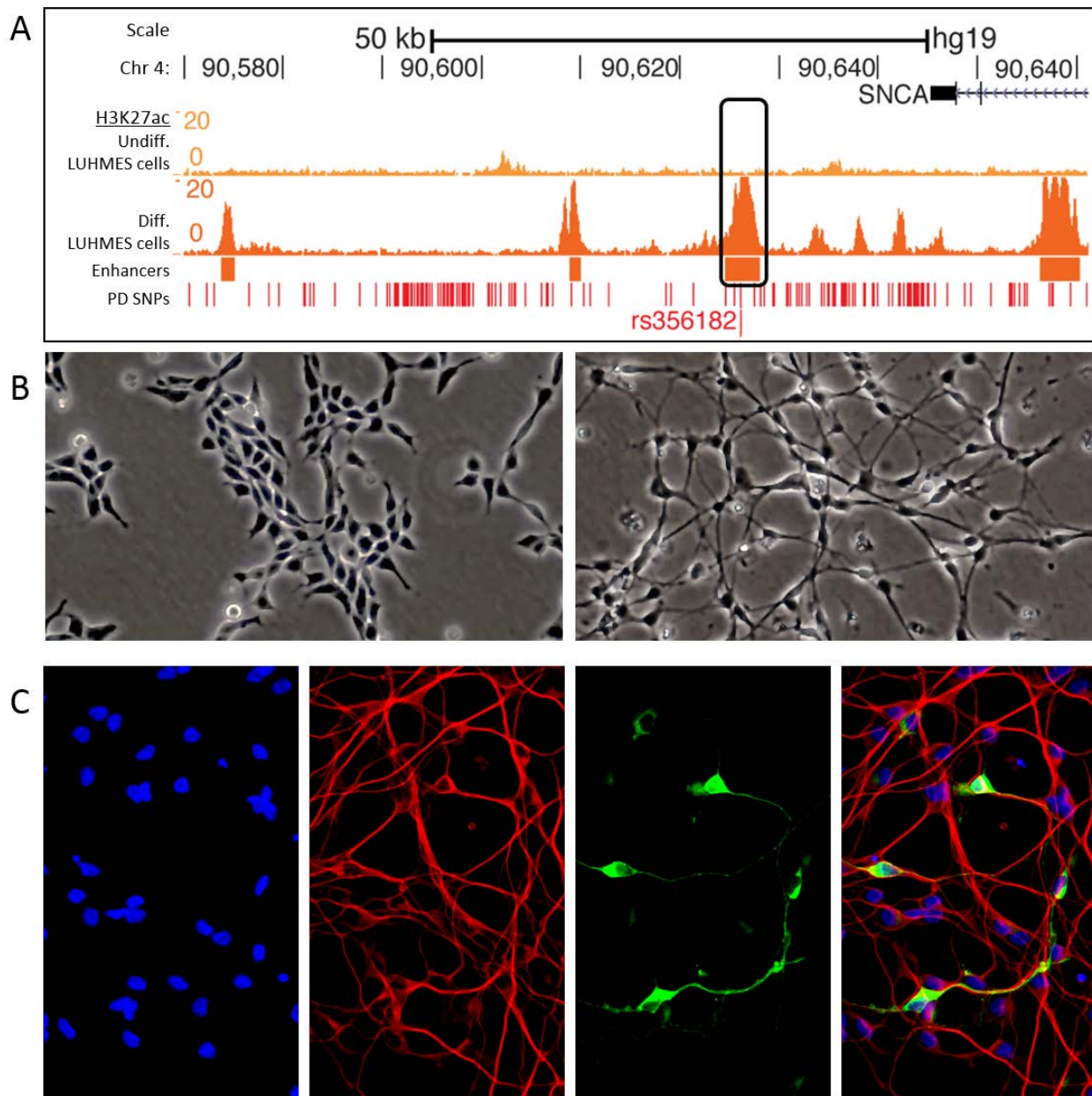
Figure 1: An enhancer encompassing rs356182 is active in differentiated LUHMES cells. (A) Histone H3K27ac track for undifferentiated (top) and differentiated (bottom) LUHMES cells. (B) Bright-field images showing the morphology of wildtype undifferentiated (left) and differentiated (right) LUHMES cells. (C) Immunofluorescent (TUJ1= red, and TH= green) and stained images (nuclei DAPI= blue) of differentiated LUHMES cells (day 6 of differentiation).
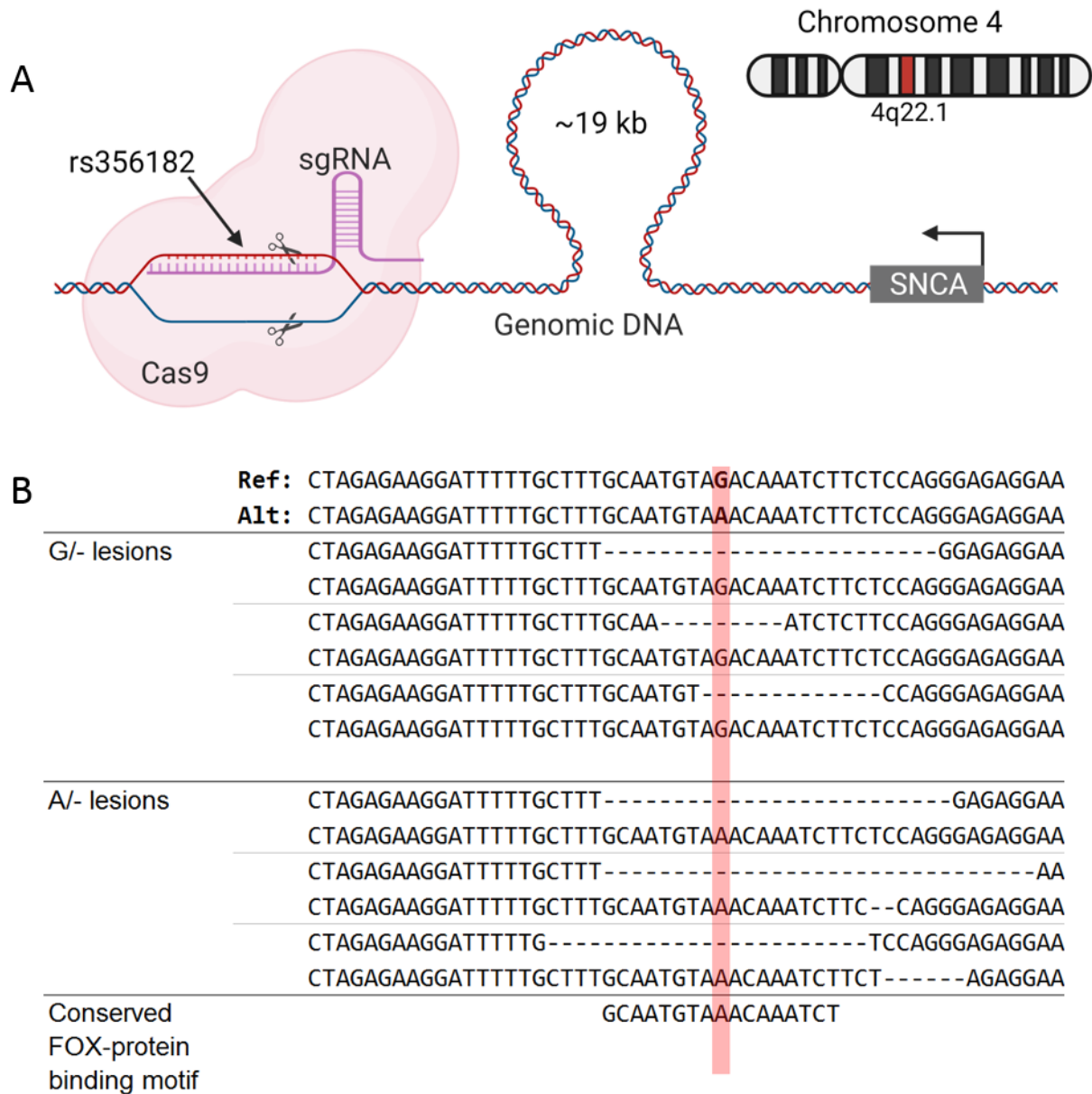
Figure 2: Excised rs356182 by CRISPR-Cas9. (A) Schematic depicting the position of rs356182 relative to *SNCA* on chromosome 4 and the guide-RNA targeting that locus for a double stranded DNA break. (B) The resulting deletions (deleted bases depicted as dashes) from targeting rs356182 with CRISPR-Cas9; position of rs356182 highlighted in red.

METHODS

LUHMES cell model: LUHMES cells, obtained from ATCC (CRL-2927), were cultured essentially as described by Scholz et al. [17]. As previously described [7], the cells were incubated in a humidified 37°C, 5% CO2 incubator in flasks pre-coated with 50 mg/mL poly-L-ornithine (Sigma, Cat # P3655) and 1 mg/mL fibronectin (Sigma, Cat # F114) in water. The coated flasks were incubated at 37°C overnight, rinsed with water, and allowed to dry before seeding cells. Cells were cultured in complete growth medium containing Advanced DMEM:F12 (Thermo Fisher, Cat # 12634-010) with 2mM L-glutamine (Thermo Fisher, Cat # 25030081), 1X N-2 supplement (Thermo Fisher, Cat # 17502-048), and 0.04 mg/mL bFGF (Stemgent, Cat # 03-0002). Cells were allowed to reach 80% confluency before passaging with 0.025% trypsin/EDTA. Before differentiation, cells were seeded at 3.5×10^6 per T75 flask containing complete growth medium and incubated at 37 °C for 24 h (day -1). For induction of differentiation, culture medium was changed to freshly prepared DMEM:F12 with 2mM L-glutamine, 1X N-2, 1mM cAMP (Carbosynth, Cat # ND07996), 1 mg/mL tetracycline (Sigma, Cat # T7660), and 2 ng/mL glial cell line-derived neurotrophic factor (GDNF) (Sigma, Cat # G1777) (day 0). LUHMES cells then grow in the differentiation media for 2 days before being passaged, again into differentiation media (day 2). Finally, cells were harvested on day 6 after the initial introduction of differentiation media (day 6).

CRISPR-Cas9 Editing: For each target sequence, double-stranded DNA sequences complementary to the target sequences were generated by PCR and then cloned into the pSpCas9(BB)-2A-GFP (PX458) vector (AddGene, cat. # 48138) (Table S1). Ligated plasmids were delivered into Stbl3 chemically competent E. coli cells and selected from ampicillin-treated auger plates. Transformed E. coli colonies were isolated and expanded in LB media. Expanded plasmids were purified using the QIAprep Spin MiniPrep Kit (Qiagen, cat. # 27106). LUHMES cells were electroporated and transfected with 2 μg of plasmid using the Amaxa™ Basic Primary Neurons Nucleofector™ Kit and protocol (Lonza, cat. # VPI-1003). Transfected cells were selected by flow-sorting for DAPI negative, GFP positive singlets into pre-coated 96-well plates (Figure S1). Sorted cells were clonally expanded and screened via Sanger DNA sequencing from Genewiz. Individual allele sequencing was achieved by TOPO-TA cloning (Thermo Fisher, cat. #

K4575J10), followed again by Sanger sequencing. Clones with a confirmed edit disrupting rs356182 were used in the edited conditions, while sorted clones which did not have a confirmed edit at rs356182 were used as the A/G lesion controls (Figure 2B).

RNAseq: Three clones of each of the G/- lesion model and A/- lesion model and parent (wild-type) G/A clones were selected for RNAseq analysis (Table S2). The three wildtype control clones (described in the CRISPR-Cas9 editing protocol above) were also selected for RNAseq, as well as 9 true wildtype clones (unexposed to CRISPR and dilution sorted) in both their undifferentiated and differentiated states. Clones were differentiated following standard LUHMES cell differentiation protocol. Clonal RNA was isolated and purified using the QIAGEN QIAshredder (Cat. #: 79654) and RNeasy isolation kit (Cat. #: 74104). Total RNA was submitted to the Van Andel Research Institute's Genomics Core for QC, library preparation, and paired-end sequencing using the Illumina NovaSeq 6000 in split-lane SP Flowcells. Sequencing results were saved as zipped FASTQ files.

Differential gene expression: First, Illumina sequencing adapters were trimmed from the raw count files using the TrimGalore software package (version- 0.6.0). Next, trimmed reads were aligned to the GRCh37- hg19 reference genome using the STAR software (Spliced Transcripts Aligned to a Reference; version- 2.7.8a). After the alignment, sub-threshold count reads were filtered out and raw counts were normalized to count per million (CPM) and trimmed mean of M values (TMM normalization) using EdgeR (version-3.32.1). Samples were submitted with a minimum biological triplicate; therefore, genes expressed in fewer than 3 samples were excluded. A z-score was determined for each gene per sample and fold-change and adjusted P-values were calculated using the limma-voom method (version- limma-3.46.0. Detailed scripts on Github.

Gene Set Enrichment Analysis (GSEA): GSEA for biological processes was performed using ClusterProfiler (version- 3.18.1). The background list of genes in the universe was described as the full list of genes expressed in our LUHMES models. In each case, an adjusted P-value < 0.05 was the maximum threshold for modulated genes. Gene sets were separated by condition and increased/decreased gene expression. An absolute fold-change greater than 2 was used as the minimum threshold for modulated genes. To reduce redundancy of ontological terms and the

bias introduced by over-represented genes, terms that have greater than 70% of shared gene annotations were collapsed into one group.

Morphology: For morphological analysis, clones were differentiated following the standard LUHMES cell differentiation protocol. On day-2 of differentiation, clones were passaged to the standard density of 1.5×10^5 cells/cm^2 into pre-coated 24-well plates. On day 6 of differentiation, clones are fixed by incubating in 2% formaldehyde solution for 5 minutes, followed by incubation in 4% formaldehyde solution for 15 minutes, then stored in DPBS at 4°C for up to 2 weeks. Immunofluorescence was conducted as described by the Invitrogen Human Dopaminergic Neuron Immunocytochemistry Kit (Catalog no. A29515). In brief, cells were incubated in perm/block buffer (1% BSA, 0.3% Triton X-100, and DPBS) for 30 minutes at room temp. Primary antibody was added directly to the perm/block buffer in the wells and incubated overnight at 4°C. Wells were washed 3x in DPBS for 2 minutes at RT, and then secondary Ab was added to the cells for 1h at RT (Table S3). Wells again were washed 3x in DPBS. On the third wash, 1-2 drops/ml of DAPI was added to the wash buffer. Cells are then stored in DPBS for <2 weeks.

After immunohistochemical and DAPI staining, samples were imaged on the Zeiss Celldiscover7 (CD7) at 10x magnification (20 Z-stacks) by the VARI Optical Imaging Core. Composite images were prepared using the ZEN 3.2 (blue edition) software by using automated threshold settings for the DAPI and TUJ1 channels and manually setting the TH channel and z-stack. The TH channel must be manually set due to the skewed intensity distribution disrupting the automated threshold settings. Split fluorescent channels were converted to binary images in ImageJ and analyzed for pixel coverage and cell count. Measurement of cell density was calculated as the percent of the well covered by DAPI stain (binary images; Black pixels / total pixels). Neurite development was calculated as the ratio of cell body stain (TUJ1) and nuclei stain (DAPI) (binary images: black pixels / total pixels, Red channel / blue channel).

$$Cell\ density\ (\%) = \frac{Pixel_{(blue)}}{Total\ Pixels} \qquad Neurite\ outgrowth = \frac{Pixels_{(red)}/Total\ Pixels}{Pixels_{(blue)}/Total\ Pixels}$$

RESULTS

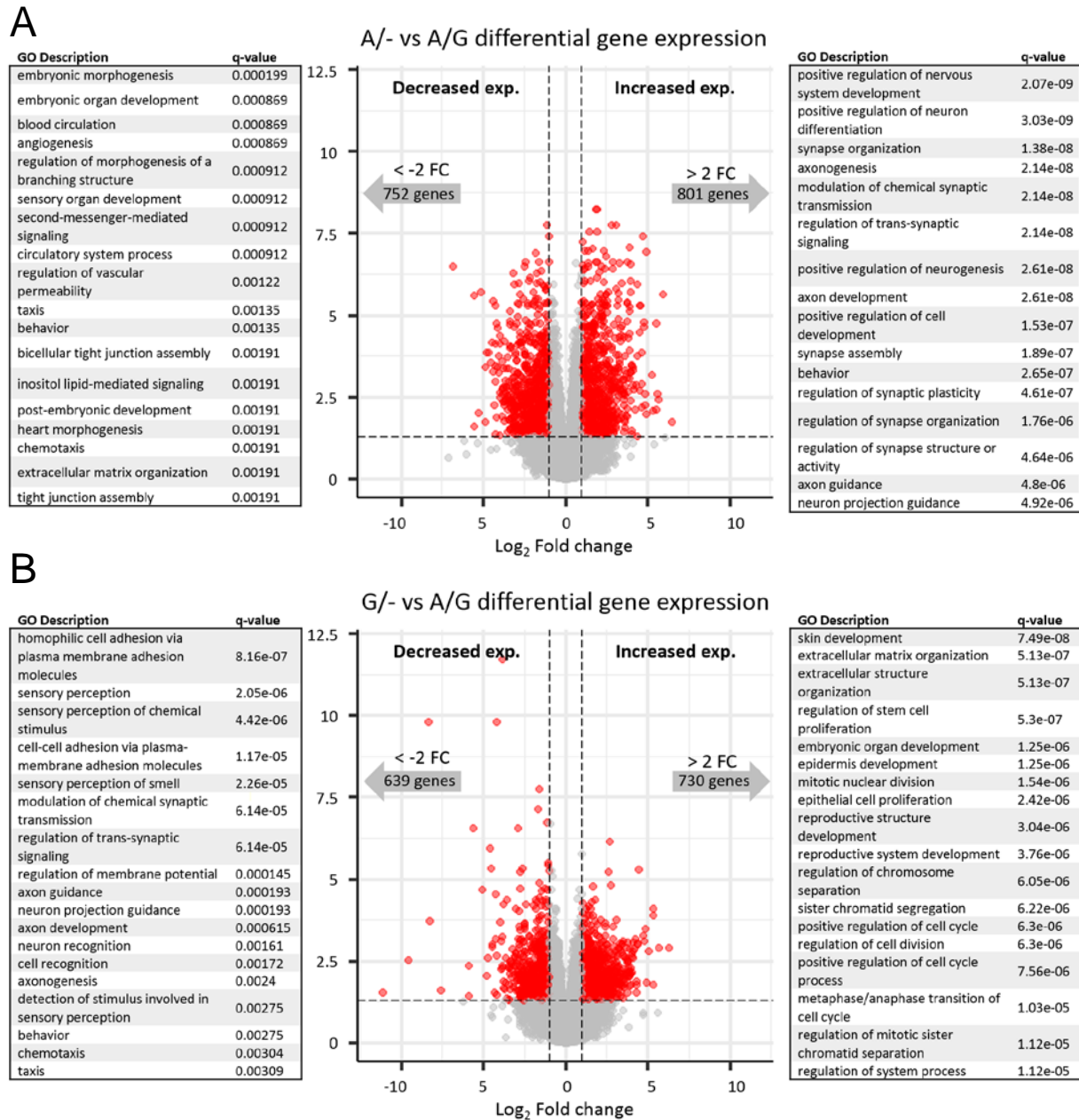**Disruption of rs356182 has global effects and allele-specific activity.** We sought to determine if rs356182 is indeed functional and identify the potential allele-specific activity of this SNP. To that end, we created hemizygous clones at this site and therefore possessing a lesion (i.e. short deletion) on one chromosome while maintaining the original major (G) or minor (A) alleles on the other. To achieve this, we used CRISPR-Cas9 to target the PAM site nearest to rs356182 (Figure 2A & Table S1). In individual cells, lesions were randomly formed at either the A-allele or G-allele and were approximately 20 ± 8 bp following DNA repair (Figure 2B). After screening, 3 clones for each of 3 conditions were generated; the PD-protective condition which maintains only the protective allele (A/-), the PD-risk condition which maintains only the risk allele (G/-), or the wildtype control condition in which clones show no edits and maintain both rs356182 alleles (A/G). Statistical analysis comparing the length of lesions replicates in the protective and risk conditions indicated no significant difference in the average deletion size (mean leasion 20 bp, G/- vs A/-, Student's T-test: P=0.1176, Wilcox test: P=0.20). Clonally derived hemizygous (single allele) strains and wildtype (both alleles) controls were differentiated and allele-specific gene expression was examined using RNA-seq. As the deletion method provided slightly different-sized lesions, expression results were analyzed for relation to lesion size but showed no correlation. We found many and widespread differentially expressed genes, contrary to the expectation that *SNCA* would be the primary gene target. Whereas *SNCA* was significantly modulated in both hemizygous conditions compared to WT (Figure 3A), neither *SNCA* nor any other genes near rs356182 were the most significant or highest fold-changed genes affected (Figure 3B). Our results here corroborate previous results which showed a higher *SNCA* expression in the A/A genotype and lowest expression in the G/G genotype [18]. We observed allele-specific differential gene expression spanning the entire genome (Figure 3C). We used principle component analysis (PCA) of gene expression for the 9 samples in order to visualize the systematic variance in transcription. The results show that the protective, A/- and A/G conditions were closely space along principle component 1 (accounting for 34.4% of the sample variance), and imply that the PD-protective condition is more similar to the wild-type condition

than it is to the risk condition (Figure 3D). Ultimately, both conditions showed several hundred genes significantly altered compared to the wildtype controls, including both up and down-regulated genes with absolute fold-change greater than two (Table S5).
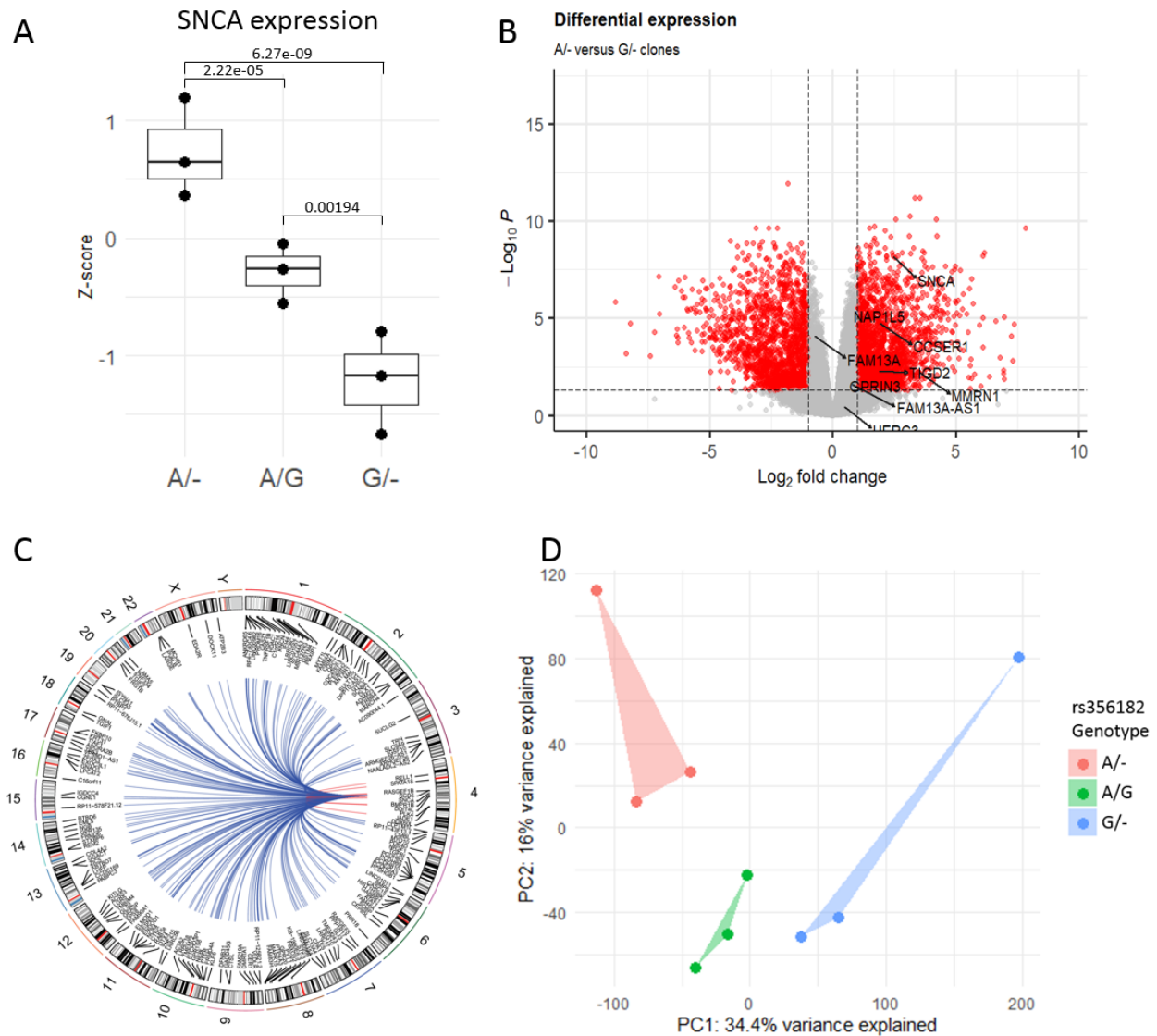
Figure 3: Disruption of rs356182 has global effects and allele-specific activity. (A) Box-and-whicker plot of z-scores for SNCA gene expression; adjusted P-value displayed above each group comparison. (B) Volcano plot showing the differentially expressed genes between the A/- and G/- clones with the genes proximal to rs356182 annotated; Log_2 fold change cutoff, 1; p-value cutoff, 0.05. (C) Circos plot depicting the significantly modulated genes between the A/- and G/- clones across the entire genome. (D) PCA plot showing principle components 1 (X-axis) and 2 (Y-axis).

CRISPR-mediated lesions at rs356182 modulate the capacity for differentiation. Based on the specific activation of the rs356182-containing enhancer during the differentiation of DA neurons (Figure 1A) [7], we hypothesized that the elements at this locus are involved in the differentiation process. Neuronal differentiation is a complex biological process and determines cell proliferation, morphology, metabolism, and communication (among many other processes). We have previously shown that the differentiation of LUHMES cells has a profound impact on gene expression, with thousands of genes significantly modulated between states of differentiation [7]. As expected, gene set enrichment analysis (GSEA) of those impacted genes showed enrichment in GO terms associated with neuronal differentiation and neurogenesis (upregulated genes), and cell cycle (down-regulated genes) [7]. Relative to the extreme comparison of neuronal vs proliferating LUHMES, the lesions strains we generated here had more modest effects on differentiated gene expression (Figure 4 & Table S5). Unexpectedly though, our lesion conditionss also resulted in differential expression for gene sets enriched for several of the same neuronal developmental ontological branches (Figure 4 & Table S6). Specifically, genes upregulated in the protective condition (A/-) participate in the positive regulation of neurogenesis, axonogenesis, and neuronal differentiation. Meanwhile, genes down-regulated in the risk condition (G/-) are enriched in terms associated with axonogenesis, neuron projection guidance, and synaptic function. The conclusion is that differentiation and neuronal morphogenesis are enhanced in the protective condition and inhibited in the PD risk condition.

A

A/- vs A/G differential gene expression

GO Description	q-value
embryonic morphogenesis	0.000199
embryonic organ development	0.000869
blood circulation	0.000869
angiogenesis	0.000869
regulation of morphogenesis of a branching structure	0.000912
sensory organ development	0.000912
second-messenger-mediated signaling	0.000912
circulatory system process	0.000912
regulation of vascular permeability	0.00122
taxis	0.00135
behavior	0.00135
bicellular tight junction assembly	0.00191
inositol lipid-mediated signaling	0.00191
post-embryonic development	0.00191
heart morphogenesis	0.00191
chemotaxis	0.00191
extracellular matrix organization	0.00191
tight junction assembly	0.00191

Decreased exp.   Increased exp.

< -2 FC   752 genes   > 2 FC   801 genes

GO Description	q-value
positive regulation of nervous system development	2.07e-09
positive regulation of neuron differentiation	3.03e-09
synapse organization	1.38e-08
axonogenesis	2.14e-08
modulation of chemical synaptic transmission	2.14e-08
regulation of trans-synaptic signaling	2.14e-08
positive regulation of neurogenesis	2.61e-08
axon development	2.61e-08
positive regulation of cell development	1.53e-07
synapse assembly	1.89e-07
behavior	2.65e-07
regulation of synaptic plasticity	4.61e-07
regulation of synapse organization	1.76e-06
regulation of synapse structure or activity	4.64e-06
axon guidance	4.8e-06
neuron projection guidance	4.92e-06

B

G/- vs A/G differential gene expression

GO Description	q-value
homophilic cell adhesion via plasma membrane adhesion molecules	8.16e-07
sensory perception	2.05e-06
sensory perception of chemical stimulus	4.42e-06
cell-cell adhesion via plasma-membrane adhesion molecules	1.17e-05
sensory perception of smell	2.26e-05
modulation of chemical synaptic transmission	6.14e-05
regulation of trans-synaptic signaling	6.14e-05
regulation of membrane potential	0.000145
axon guidance	0.000193
neuron projection guidance	0.000193
axon development	0.000615
neuron recognition	0.00161
cell recognition	0.00172
axonogenesis	0.0024
detection of stimulus involved in sensory perception	0.00275
behavior	0.00275
chemotaxis	0.00304
taxis	0.00309

Decreased exp.   Increased exp.

< -2 FC   639 genes   > 2 FC   730 genes

| GO Description	q-value
skin development	7.49e-08
extracellular matrix organization	5.13e-07
extracellular structure organization	5.13e-07
regulation of stem cell proliferation	5.3e-07
embryonic organ development	1.25e-06
epidermis development	1.25e-06
mitotic nuclear division	1.54e-06
epithelial cell proliferation	2.42e-06
reproductive structure development	3.04e-06
reproductive system development	3.76e-06
regulation of chromosome separation	6.05e-06
sister chromatid segregation	6.22e-06
positive regulation of cell cycle	6.3e-06
regulation of cell division	6.3e-06
positive regulation of cell cycle process	7.56e-06
metaphase/anaphase transition of cell cycle	1.03e-05
regulation of mitotic sister chromatid separation	1.12e-05
regulation of system process	1.12e-05

Figure 4: CRISPR-mediated lesions at rs356182 modulate the capacity for differentiation. Volcano plot of RNA-seq data depicting the differentially expressed genes between the lesion clones and wildtype clones. The grey dots represent genes that fail to meet minimum thresholds for adjust p-value (<0.05) and/or $\log_2$ fold change (>1). To the left and right, the top GO terms and associated P-values for each differentially expressed (up or down) gene set. (A) A/- versus A/G. (B) G/- versus A/G.

The rs356182 risk-allele mediates a pathological morphology change. Following GSEA, it was clear that the neurodevelopment was generally altered in our lesion conditions. To determine whether those genetic changes resulted in obvious morphological differences or other differences in cell growth we differentiated and stained cells with DAPI, as a nuclear marker, and treated with anti-TUJ1 (neuron-specific class III beta-tubulin) which marks neuronal soma and neurites, and anti-TH (tyrosine hydroxylase) as an indicator of dopamine production. The G/- clones had significantly more cells per well on day 6 of differentiation than their wildtype counterparts (seeded at equal densities on day 2 of LUHMES cell differentiation) (Figure 5A & 5B). We believe this is due to a differentiation-specific modulation of proliferative activity because edited clones did not show different growth rates in their undifferentiated state (Figure S2). Additionally, the cell density within each well was more variable in the risk clone samples than in the protective clones or wild-type clones (Figure 5C). These wells more commonly had clusters of densely packed cells and then sections of sparsely populated cells (Figure 5A, middle). The TUJ1/DAPI stain ratio is also smaller in the risk (G/-) condition indicating reduced neurite growth per cell, in line with results from GSEA (Figure 5D). Finally, the protective clones had significantly more TH-expressing cells on day 6 of differentiation than either the wildtype controls or the risk clones (Figure 5E).
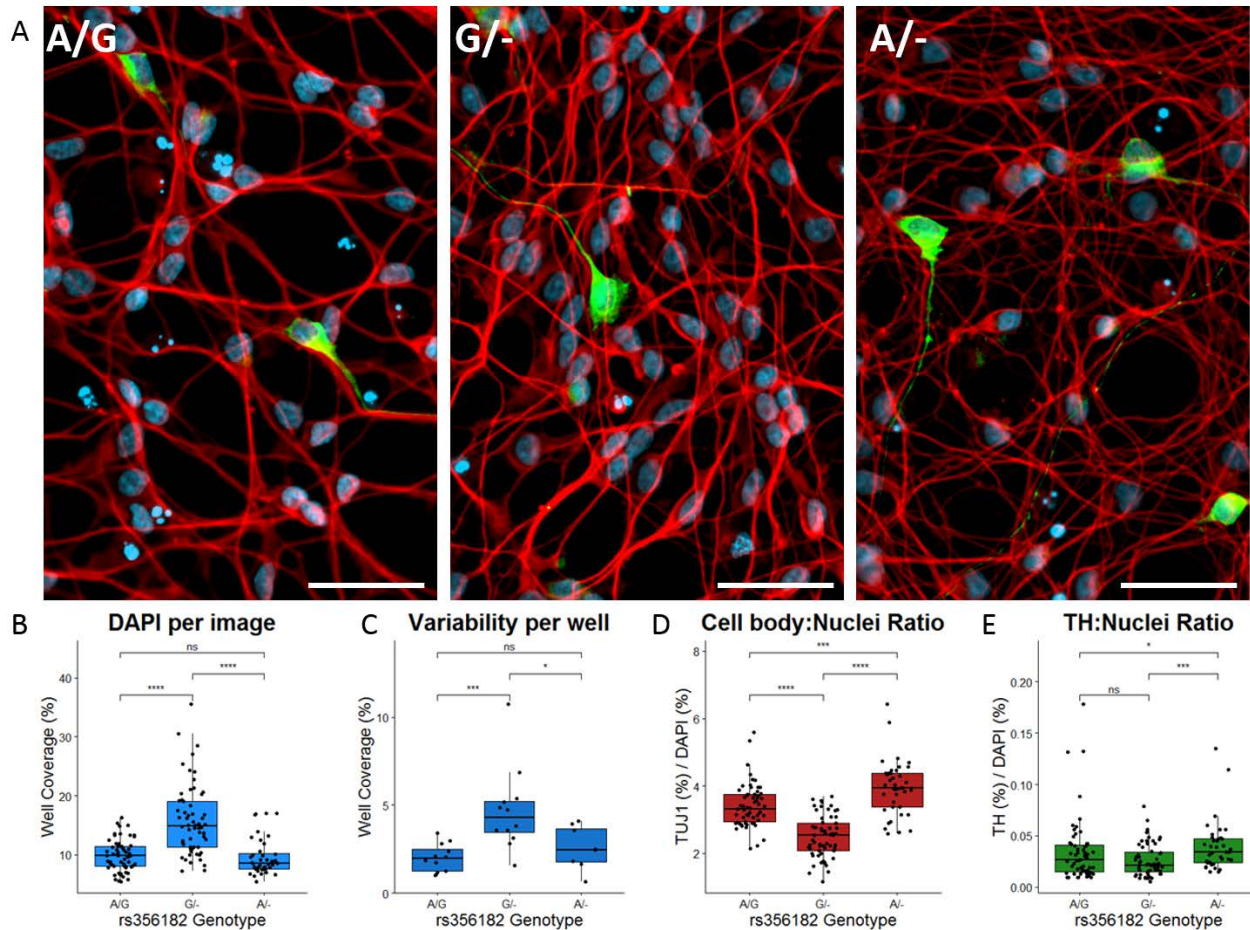
Figure 5: The rs356182 risk-allele mediates a pathological morphology change. (A) Representative immunofluorescent images from wildtype A/G clones (left), risk G/- clones (middle), and protective A/- clones (right); white bar= 50 μm, blue= DAPI, red= TUJ1, green= TH. (B-E) Quantification of immunofluorescent images by pixel coverage represented in box-and-whisker plots; Wilcoxon comparisons of nonparametric groups, ns: p>0.05, *: p<=0.05, **: p<=0.01. ***: p<=0.001, ****: p<=0.0001. (B) DAPI coverage per image as a measure of cell density. (C) Standard deviation of inter-well DAPI coverage. (D) Ratio of TUJ1/DAPI as a relative measure of cell body to nucleus. (E) Ratio of TH/DAPI as a normalized measure of TH-expressing cells within the cell population.

rs356182 modulates transcription factor binding affinity. In situ analysis of the TF binding-motifs at the rs356182 region using HaploReg (Table S3) and MotifbreakR (Table S4) revealed a preferential binding affinity for the protective A-allele over the risk G-allele for the strongest interacting TFs at this locus (Tables S3 & S4) [19-21]. Notably, the Fox family of proteins was identified as strong candidate binders to the A-containing motif (Figure 2B & 6). Fox proteins

regulate the expression of genes involved in branching morphogenesis (FOXA1/2), brain development (FOXA2/C1), and axon guidance (FOXD1), among a host of other biological processes [22]. The FOXA1 motif has previously been shown to be particularly vulnerable to variants that disrupt the 3 consecutive adenines, as the MotifbreakR results would suggest, and which is true of rs356182 (Figure 6) [23]. Additionally, the FOX family of proteins are known to engage in pioneer factor activity [24]. This data suggest that the rs356182 risk-allele (G) disrupts a particularly vital position within the putative binding motif, resulting in diminished binding affinity for FOX TFs, with the potential to reduce enhancer activity entirely.
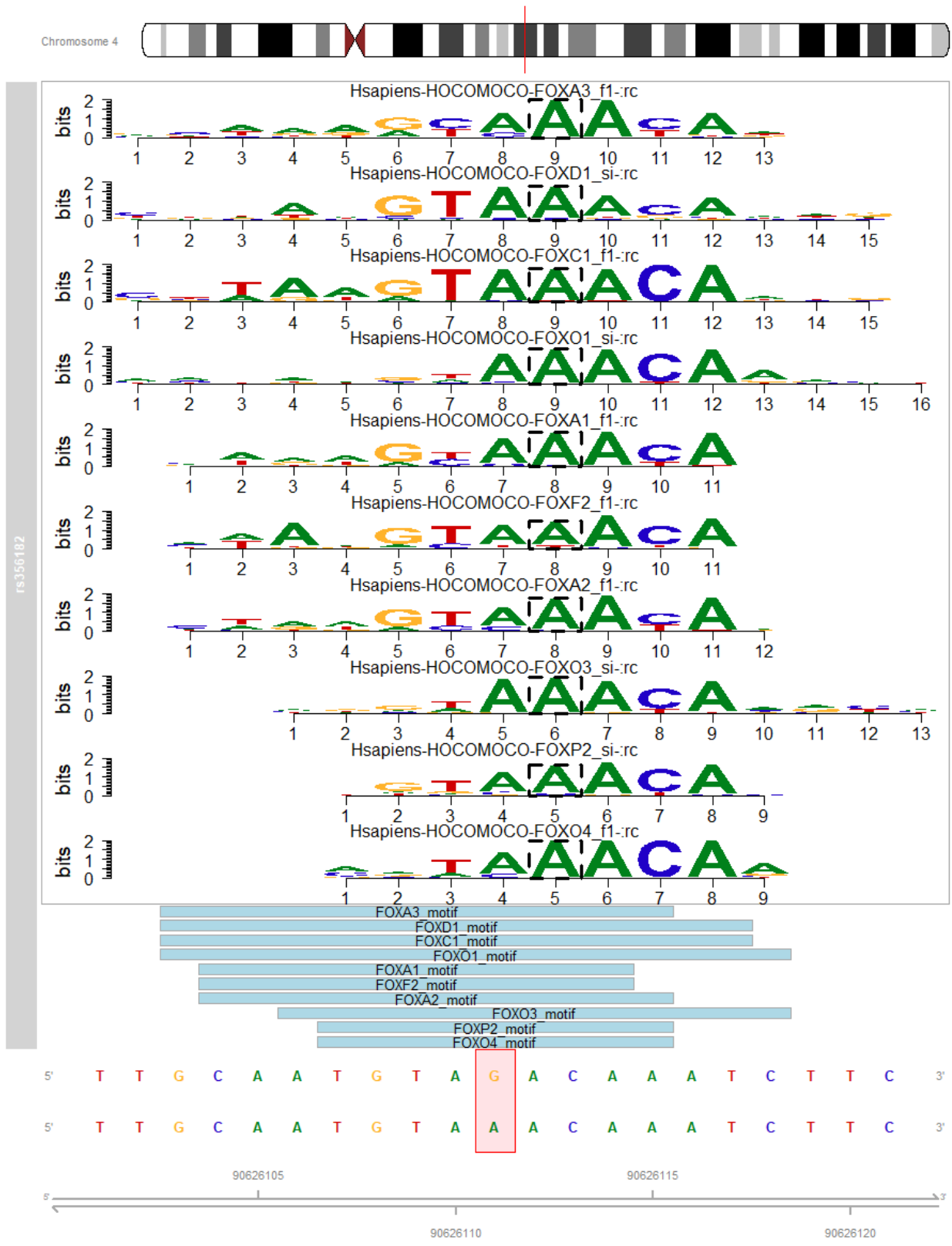
Figure 6: rs356182 modulates transcription factor binding affinity. (A) Results from MotifBreakR depicting the sequence and importance (letter height) of each base within the motif sequence for the "Strong interaction" proteins at this locus.

DISCUSSION

There are more than a million genetic variants in the human genome, occurring on average once every six base-pairs depending on the minor allele frequency in the population. These variants, the majority of which are SNPs, have been used in GWASs to identify loci associated with a variety of traits and diseases, including PD. The value of the results obtained from these associations has been widely debated [25, 26]. The reason for this debate is that the mechanisms underpinning the associations are not always clear and most (>90%) of the risk SNPs don't reside in protein coding regions, but instead in presumptive regulatory elements, which affect the expression of often unknown genes. Furthermore, many SNPs at a given locus are in linkage disequilibrium with each other, making causality assignment unclear. As of September 2018, the NHGRI-EBI GWAS catalog contains 5,687 GWASs of 71,673 variant-trait associations [27]. The vast majority of them provide a landscape view of possible functional involvements, leaving the majority with unknown functional significance. A major gap in our ability to understand and utilize GWAS results is thus the lack of in-depth mechanistic insights.

A rare example of mechanistic detailing of risk variant/gene association involves the variant within an intron of *FTO* and associated with increased risk for obesity and type-2 diabetes (T2D) [28]. Since epidemiological associations and data from mouse models indicate that *FTO* expression levels influence body mass, it seemed likely that the *FTO* intronic variants impose T2D risk via *FTO* expression regulation affecting enhancer activity. However, contrary to expectations, it was shown that that the *FTO* intronic variants are functionally and structurally connected with the homeobox gene *IRX3*, some mega-bases distant on linear DNA. It was demonstrated that IRX3 has a roll in controlling body mass composition and metabolism and represents a novel therapeutic target for diabetes.

Here we present evidence regarding the mechanism(s) surrounding rs356182 that contradicts the assumption that *SNCA* is the only affected gene. This assumption is that

rs356182 directly and exclusively regulates the expression of *SNCA*, and in this way confers risk for PD. We show by the mono-allelic deletion of rs356182 that hundreds of genes, spanning the entire genome, are affected by rs356182. However, it remains unclear if these downstream gene changes are primary, secondary, or tertiary events. Furthermore, these affected genes appear to be enriched in GO terms related to the differentiation-proliferation processes. Our results suggest that: (i) *SNCA* has an important role as a regulator of DA neuron differentiation and that by manipulating rs356182, a cascade of genes is impinged by the resulting *SNCA* misregulation, or (ii) rs356182 does not exclusively regulate *SNCA* and is independently relevant to DA neuronal differentiation via the regulation of many genes. One, or both, of these scenarios, is likely true of the mechanism underlying the PD risk associated with rs356182.

To function as a transcription factor, α-SYN would need to be present at the site of transcription-- the nucleus. α-SYN was observed in both the nucleus and cytosol of neurons as early as the 1980's, hence the name, but the implications of this localization have since been debated [29-31]. Under pathological conditions, such as in A53T *SNCA*-mutant containing neurons, nuclear localization of α-SYN is increased [32-34]. The α-SYN translocation is mediated by retinoic acid [35], and the accumulation and neurotoxicity are regulated by TRIM28 [36]. Once inside the nucleus, the mode of α-SYN associated neurotoxicity is less clear. α-SYN has been reported to bind to histones and increase fibrillation [32], bind to DNA directly to mediate DNA repair [37, 38], and interact with retinoic-acid response elements to regulate associated gene transcription which is linked to PD [35]. The DNA-repair mechanism associated with α-SYN has been debated and may be a result of normal cell response to DNA damage induced by α-SYN toxicity [38-40]. While it is clear that nuclear localization is linked to neurotoxicity [33], evidence for *SNCA* behaving as a TF is speculative.

The results from the motif analysis of rs356182 (Figure 1C, Table S3, & Table S4) imply that the A-allele is the primary functional allele and that the FOX family of proteins are the primary TFs binding to the A-containing motif. rs356182 is in the middle of the strong triple-A motif, which has previously been shown to be particularly important for TF binding, disruption of which having pathological consequences [23, 41]. Since FOX proteins have pioneer TF abilities, they are likely partially or wholly responsible for activation of the encompassing

enhancer, resulting in the regulation of many genes by both primary- and secondary-/tertiary-interactions. What remains less clear is what is happening in the presence of the G-allele.

Based on the mechanism proposed above, deletion of the G-allele should have few, or no, consequences on gene expression, but this is not what we observed. In fact, deletion of the G-allele in the PD-protective clones promoted neuronal differentiation and neurogenesis (Figure 3). The fact that losing one allele increases the expression of certain genes and processes indicates that, contrary to what we predicted from binding-motif analysis, the G-allele is not simply a less functional allele, but is counteractively working against the A-allele. Further evidence that this is not a simple dose-response is that ~21% of the nearly 25,000 genes in our dataset have opposite expression changes between experimental conditions (i.e., one edited condition increases while the other edited condition decreases in expression compared to the wildtype), an example of which is *SNCA* itself (Figure 2A). Additionally, morphological analysis of body-nucleus ratio and the ratio of TH-expressing cells per condition showed the A/-clones seemingly differentiating more effectively than the risk clones, and even the wild-type A/G clones (i.e., A/- clones grew more/larger neurites per cell and more cells reached dopaminergic maturity within the timeframe). In other words, the A- and G-alleles impinge on the differentiation-proliferation mechanism by actively working against each other; the A-allele pushing the cell towards differentiation, and the G-allele holding it back in a more stem-like state. The dual functionality of the rs356182 alleles could explain why this locus has such a strong association with PD.

The PD field has known about the significant association of rs356182 to PD since the earliest GWASs due to its strong association with risk. Until now, the mechanism surrounding this risk was attributed to allele-specific expression of *SNCA*, variation of which would presumably predispose individuals to PD in their later years. However, we have identified a novel mechanism in which rs356182 impinges on the differentiation/proliferation mechanism during development. These results contribute to the growing body of evidence that PD is, at least in part, a developmental disorder [42, 43]. We speculate that rs356182 regulates neuronal differentiation and a cascade of related processes, leading to a diminished population of healthy dopaminergic (DA) neurons in individuals with the risk allele, making the individual

more sensitive to subsequent insults to the DA cell population, and in this way confers risk for PD later in life [44].

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS

CRediT Taxonomy: Conceptualization, JP, SEP, GAC, and TT; Methodology, JP and TT; Software, JP and SEP; Formal Analysis, JP and SEP; Investigation, JP; Resources, JP, EJCV, and TT; Writing-original draft, JP; Writing- Review & editing, JP, TT, SEP, and GAC; Visualization, JP and SEP; Supervision, GAC and TT; Funding Acquisition, GAC.

DECLARATIONS OF INTERESTS

We have no competing interests.

REFERENCES

1.   Genomes Project, C., et al., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.
2.   Consortium, E.P., *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.
3.   Soldner, F., et al., *Parkinson-associated risk variant in distal enhancer of alpha-synuclein modulates target gene expression.* Nature, 2016. **533**(7601): p. 95-9.
4.   Van Heyningen, V. and P.L. Yeyati, *Mechanisms of non-Mendelian inheritance in genetic disease.* Hum Mol Genet, 2004. **13 Spec No 2**: p. R225-33.
5.   Schaffer, A.A., *Digenic inheritance in medical genetics.* J Med Genet, 2013. **50**(10): p. 641-52.
6.   Slatkin, M., *Linkage disequilibrium--understanding the evolutionary past and mapping the medical future.* Nat Rev Genet, 2008. **9**(6): p. 477-85.
7.   Pierce, S.E., et al., *Parkinson's disease genetic risk in a midbrain neuronal cell line.* Neurobiol Dis, 2018. **114**: p. 53-64.
8.   Nalls, M.A., et al., *Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies.* Lancet Neurol, 2019. **18**(12): p. 1091-1102.

9.   Nalls, M.A., et al., *Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease.* Nat Genet, 2014. **46**(9): p. 989-93.

10.   Pihlstrom, L., et al., *A comprehensive analysis of SNCA-related genetic risk in sporadic parkinson disease.* Ann Neurol, 2018. **84**(1): p. 117-129.

11.   Nott, A., et al., *Brain cell type-specific enhancer-promoter interactome maps and disease-risk association.* Science, 2019. **366**(6469): p. 1134-1139.

12.   Gusev, A., et al., *Integrative approaches for large-scale transcriptome-wide association studies.* Nat Genet, 2016. **48**(3): p. 245-52.

13.   Zhu, Z., et al., *Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets.* Nat Genet, 2016. **48**(5): p. 481-7.

14.   Chepelev, I., et al., *Characterization of genome-wide enhancer-promoter interactions reveals co-expression of interacting genes and modes of higher order chromatin organization.* Cell Res, 2012. **22**(3): p. 490-503.

15.   Ernst, J., *Mapping enhancer and promoter interactions.* Cell Res, 2012. **22**(5): p. 789-90.

16.   Lotharius, J., et al., *Effect of mutant alpha-synuclein on dopamine homeostasis in a new human mesencephalic cell line.* J Biol Chem, 2002. **277**(41): p. 38884-94.

17.   Scholz, D., et al., *Rapid, complete and large-scale generation of post-mitotic neurons from the human LUHMES cell line.* J Neurochem, 2011. **119**(5): p. 957-71.

18.   Cooper, C.A., et al., *Common variant rs356182 near SNCA defines a Parkinson's disease endophenotype.* Ann Clin Transl Neurol, 2017. **4**(1): p. 15-25.

19.   Kheradpour, P. and M. Kellis, *Systematic discovery and characterization of regulatory motifs in ENCODE TF binding experiments.* Nucleic Acids Res, 2014. **42**(5): p. 2976-87.

20.   Ward, L.D. and M. Kellis, *HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants.* Nucleic Acids Res, 2012. **40**(Database issue): p. D930-4.

21.   Coetzee, S.G., G.A. Coetzee, and D.J. Hazelett, *motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites.* Bioinformatics, 2015. **31**(23): p. 3847-9.

22.   Tuteja, G. and K.H. Kaestner, *SnapShot: forkhead transcription factors I.* Cell, 2007. **130**(6): p. 1160.

23.   Jia, L., et al., *Functional Enhancers at the Gene-Poor 8q24 Cancer-Linked Locus.* PLoS Genetics, 2009. **5**(8): p. e1000597.

24.   Zaret, K.S. and J.S. Carroll, *Pioneer transcription factors: establishing competence for gene expression.* Genes Dev, 2011. **25**(21): p. 2227-41.

25.   Bailey, K.R. and C. Cheng, *Conference Scene: The great debate: genome-wide association studies in pharmacogenetics research, good or bad?* Pharmacogenomics, 2010. **11**(3): p. 305-8.

26.   Ioannidis, J.P., *Non-replication and inconsistency in the genome-wide association setting.* Hum Hered, 2007. **64**(4): p. 203-13.

27.   Buniello, A., et al., *The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019.* Nucleic Acids Res, 2019. **47**(D1): p. D1005-D1012.

28.   Smemo, S., et al., *Obesity-associated variants within FTO form long-range functional connections with IRX3.* Nature, 2014. **507**(7492): p. 371-5.

29.   Surguchov, A., *Intracellular Dynamics of Synucleins: "Here, There and Everywhere".* Int Rev Cell Mol Biol, 2015. **320**: p. 103-69.

30.   Maroteaux, L., J.T. Campanelli, and R.H. Scheller, *Synuclein: a neuron-specific protein localized to the nucleus and presynaptic nerve terminal.* J Neurosci, 1988. **8**(8): p. 2804-15.

31.   Yu, S., et al., *Extensive nuclear localization of alpha-synuclein in normal rat brain neurons revealed by a novel monoclonal antibody.* Neuroscience, 2007. **145**(2): p. 539-55.

32. Goers, J., et al., *Nuclear localization of alpha-synuclein and its interaction with histones.* Biochemistry, 2003. **42**(28): p. 8465-71.

33. Kontopoulos, E., J.D. Parvin, and M.B. Feany, *Alpha-synuclein acts in the nucleus to inhibit histone acetylation and promote neurotoxicity.* Hum Mol Genet, 2006. **15**(20): p. 3012-23.

34. Fares, M.B., et al., *The novel Parkinson's disease linked mutation G51D attenuates in vitro aggregation and membrane binding of alpha-synuclein, and enhances its secretion and nuclear localization in cells.* Hum Mol Genet, 2014. **23**(17): p. 4491-509.

35. Davidi, D., et al., *alpha-Synuclein Translocates to the Nucleus to Activate Retinoic-Acid-Dependent Gene Transcription.* iScience, 2020. **23**(3): p. 100910.

36. Rousseaux, M.W., et al., *TRIM28 regulates the nuclear accumulation and toxicity of both alpha-synuclein and tau.* Elife, 2016. **5**.

37. Schaser, A.J., et al., *Alpha-synuclein is a DNA binding protein that modulates DNA repair with implications for Lewy body disorders.* Sci Rep, 2019. **9**(1): p. 10919.

38. Paiva, I., et al., *Sodium butyrate rescues dopaminergic cells from alpha-synuclein-induced transcriptional deregulation and DNA damage.* Hum Mol Genet, 2017. **26**(12): p. 2231-2246.

39. Vasquez, V., et al., *Chromatin-Bound Oxidized alpha-Synuclein Causes Strand Breaks in Neuronal Genomes in in vitro Models of Parkinson's Disease.* J Alzheimers Dis, 2017. **60**(s1): p. S133-S150.

40. Milanese, C., et al., *Activation of the DNA damage response in vivo in synucleinopathy models of Parkinson's disease.* Cell Death Dis, 2018. **9**(8): p. 818.

41. Hazelett, D.J., S.G. Coetzee, and G.A. Coetzee, *A rare variant, which destroys a FoxA1 site at 8q24, is associated with prostate cancer risk.* Cell Cycle, 2013. **12**(2): p. 379-80.

42. Barlow, B., et al., *The gestational environment and Parkinson's disease: Evidence for neurodevelopmental origins of a neurodegenerative disorder.* Reproductive Toxicology, 2007. **23**(3): p. 457-470.

43. Schwamborn, J.C., *Is Parkinson's Disease a Neurodevelopmental Disorder and Will Brain Organoids Help Us to Understand It?* Stem Cells Dev, 2018. **27**(14): p. 968-975.

44. von Linstow, C.U., et al., *Does Developmental Variability in the Number of Midbrain Dopamine Neurons Affect Individual Risk for Sporadic Parkinson's Disease?* J Parkinsons Dis, 2020.