Neurobiology of Disease 114 (2018) 53–64

Contents lists available at ScienceDirect

Neurobiology of Disease

journal homepage: www.elsevier.com/locate/ynbdi

Parkinson's disease genetic risk in a midbrain neuronal cell line

Steven E. Pierce[1], Trevor Tyson[1], Alix Booms, Jordan Prahl, Gerhard A. Coetzee*

Center for Neurodegenerative Science, Van Andel Research Institute, Grand Rapids, MI, United States

ARTICLE INFO

Keywords:
Parkinson's disease
LUHMES
Genetic enhancers
Vesicle transport
Neurogenesis
Differentiation
GWAS
Differential gene expression
RNA-Seq
CHIP-Seq
Risk SNP

ABSTRACT

In genome-wide association studies of complex diseases, many risk polymorphisms are found to lie in non-coding DNA and likely confer risk through allele-dependent differences in gene regulatory elements. However, because distal regulatory elements can alter gene expression at various distances on linear DNA, the identity of relevant genes is unknown for most risk loci. In Parkinson's disease, at least some genetic risk is likely intrinsic to a neuronal subpopulation of cells in the brain regions affected. In order to compare neuron-relevant methods of pairing risk polymorphisms to target genes as well as to further characterize a single-cell model of a neurodegenerative disease, we used the portionally-dopaminergic, neuronal, mesencephalic-derived cell line LUHMES to dissect differentiation-specific mechanisms of gene expression. We compared genome-wide gene expression in undifferentiated and differentiated cells with genome-wide histone H3K27ac and CTCF-bound regions. Whereas promoters and CTCF binding were largely consistent between differentiated and undifferentiated cells, enhancers were mostly unique. We matched the differentiation-specific appearance or disappearance of enhancers with changes in gene expression and identified 22,057 enhancers paired with 6388 differentially expressed genes by proximity. These enhancers are enriched with at least 13 transcription factor response elements, driving a cluster of genes involved in neurogenesis. We show that differentiated LUHMES cells, but not undifferentiated cells, show enrichment for PD-risk SNPs. Candidate genes for these loci are largely unrelated, though a subset is linked to synaptic vesicle cycling and transport, implying that PD-related disruption of these pathways is intrinsic to dopaminergic neurons.

1. Introduction

Parkinson's disease (PD) is a neurological disorder characterized by the selective degeneration of dopaminergic neurons within the substantia nigra (Poewe et al., 2017). The resulting loss of dopamine signaling from the substantia nigra to the striatum is the primary cause of the distinctive motor symptoms associated with PD, such as rigidity, slowness in movement (bradykinesia), postural instability, and a resting tremor (Fahn, 2003). This damage is associated with the formation of intraneuronal inclusions, Lewy bodies, and Lewy neurites that are composed primarily of α-synuclein (α-syn) (Spillantini et al., 1997). The misfolding of α-syn is a key step in the development of PD pathology: the results from numerous cell culture and animal experiments indicate that oligomeric and/or fibrillar forms of α-syn are cytotoxic (Lee et al., 2014). However, the events that result in α-syn misfolding are not completely understood, even though multiple molecular pathways and cellular mechanisms that have been identified suggest that regulation of cellular homeostasis is crucial (Lopes da Fonseca et al., 2015).

The heritability of PD is complex. For a succinct overview of what is currently believed we recommend, from *Cell, SnapShot: Genetics of Parkinson's Disease* (Bras et al., 2015). Rare Mendelian-inherited cases of PD contribute about 10% of disease variability, whereas at least 41 common single nucleotide polymorphism (SNP) variants contribute about 30%, with each of the latter imposing low but significant risk (Chang et al., 2017; Nalls et al., 2014; Verstraeten et al., 2015). > 90% of the identified risk SNPs are located in non-coding DNA, making the assignment of potential functionality or causality difficult.

We (Coetzee et al., 2016; Pierce and Coetzee, 2017) and others (Lebouvier et al., 2009; Tyson et al., 2016) have reported that diverse tissues are implicated in PD predisposition, but substantia nigra neurons are clearly most directly involved. Most recently, based on a meta-analysis of PD GWAS studies, 71 candidate loci and putative target genes were reported (Chang et al., 2017). The likely functional processes include inflammation, energy metabolism, protein degradation, vesicle trafficking, small-molecule clearing, cellular import, and oxidative stress responses, and it is clear that several of these processes are not intrinsic to the ultimately affected substantia nigra neuronal cells.

* Corresponding author at: Van Andel Research Institute, 333 Bostwick Ave., N.E., Grand Rapids, MI 49503, United States.
E-mail address: gerry.coetzee@vai.org (G.A. Coetzee).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.nbd.2018.02.007
Received 16 October 2017; Received in revised form 30 January 2018; Accepted 21 February 2018
Available online 24 February 2018

Therefore, the relevant tissue and cell types in which these processes, when impaired, lead to PD is a crucial question.

The unique properties of substantia nigra dopaminergic neurons may contribute to their vulnerability. Each human dopaminergic neuron contains up to one million axon terminals, many at vast distances from the soma, presenting a serious challenge to maintaining cellular homeostasis (Pissadaki and Bolam, 2013). In comparison, neocortical neurons are estimated to have from a few thousand to a few tens-of-thousands of synapses (DeFelipe et al., 2002). Aging also exacerbates this stress, as neurons do not renew themselves through cellular division and toxic products and other damage can accumulate within a cell. Furthermore, the metabolism of dopamine results in reactive oxidative species that add extra stress to the system (Ryan et al., 2013). However, not all non-dividing neurons or highly arborized neurons are subject to neurodegeneration in PD brains, suggesting that a combination of factors may be responsible for pathogenesis.

In the present study, we used an established cell line called Lund human mesencephalic cells (LUHMES) as a model of human substantia nigra neurons (Scholz et al., 2011). LUHMES cells differentiate into post-mitotic neurons that are electrically active and produce both dopamine and wild-type human α-syn (Lotharius et al., 2002). Further post-differentiation analysis of these cells revealed phenotypic markers of neuronal maturation (Scholz et al., 2011). These cells have been used previously to study neurodegeneration caused by dopamine-induced oxidative stress (Lotharius et al., 2005), α-syn-induced toxicity (Hollerhage et al., 2017), α-syn-induced transcriptional deregulation (Paiva et al., 2017), and other neurotoxic effects (Smirnova et al., 2016). The line is diploid, making genetic manipulation relatively easy, so it is an appropriate model for substantia nigra differentiation and mechanistic PD genetic risk assessments.

We report here genome-wide histone H3K27ac, CTCF occupancy, and RNA expression in differentiated (6–7 days) and undifferentiated LUHMES cells. Acetylation of H3K27 marks the location of active promoters and enhancers, while CTCF is a transcriptional regulatory protein that is required for three dimensional nuclear and demarcates topologically-associated domains (TADs). We related the differentiation-dependent appearance of the chromatin features to differentially altered expression of nearby genes. This allowed detailed annotation of gene expression and enhancer locations that determine the process of differentiation as well as of processes that characterize terminally differentiated neurons. We found that the differentiated, but not the undifferentiated, cell enhancer profile was enriched for PD-risk SNPs, suggesting that relevant PD processes are active in the differentiated condition only. In turn, we defined risk enhancers as H3K27ac peaks containing PD-risk SNPs. These represent targets for later experimental validation. They point to a subset of genes and accompanying neuronal processes that are active in differentiated LUHMES cells and are likely impaired during the etiology of PD. In this way, we defined PD-risk gene networks and possible mechanisms involved in PD onset and progression.

2. Results

2.1. The LUHMES cell model

We cultured and propagated LUHMES cells in their undifferentiated state (Fig. 1A) to allow for genetic manipulation and cloning (see below). After tetracycline addition, they differentiated into neurons as assessed by morphology after 6 days (Fig. 1B) and by ßIII tubulin (*TUBB3*), at 100% of cells, and tyrosine hydroxylase (TH), at 10–20% of cells, which has been previously reported for this cell line (Ghosh et al., 2016) (Fig. 1C–E). The mRNA of both of these genes (*TH* and *TUBB3*) were also upregulated in differentiated LUHMES cells as shown in our RNA-seq data set. Our data also show that other key phenotypic markers of dopaminergic biology are expressed in differentiated LUHMES cells (and at statistically significant lower levels in undifferentiated cells) such as: dopamine receptors (*DRD2* and *DRD4*), Dopa Decarboxylase (*DDC*), vesicular monoamine transporter (*SLC18A2*), α-synuclein (*SNCA*), and many others (Fig. S1). However, because TH is only visible, following immunocytochemistry, in a subset of the cells to be precise we refer to the cells as 'portionally-dopaminergic'. Undifferentiated and differentiated cells were also clearly distinguishable by their RNA expression profiles (Fig. S2), and can be visualized as the change in transcript abundance for each gene (Fig. 2). Between the two conditions, a total of 14,603 transcripts are expressed at a normalized level of at least one count per million (CPM) on average between biological replicates. The volcano plot in Fig. 2 depicts differences in gene expression between differentiated and undifferentiated cells. Color represents the different levels in mean expression for each gene in the undifferentiated cells. For instance, the red dots on the right represent transcripts with low expression in undifferentiated cells that increase dramatically after differentiation of LUHMES cells. On the left of the plot are genes that show decreased expression after differentiation. In total, out of 57,905 genes mapped (including non-coding genes and pseudogenes and not restricted by overall expression level), 6147 genes were significantly down-regulated (2449 ≥ 4-fold) and 7621 genes were significantly up-regulated (3939 ≥ 4-fold) following the differentiation of LUHMES into neurons. Overall, the expression changes are as expected, with replication and cell cycle genes turning off and neuronal processes like axon growth and synaptic signaling turning on, as revealed by gene ontology analyses (Fig. 2). Considering only the 14,603 most highly expressed genes: the majority (10,230) were expressed at similar levels in each condition (Fig. 3A).

We next annotated genomic features in the two LUHMES states by histone H3K27ac ChIP-seq (promoters and enhancers) and CTCF ChIP-seq (insulators and TAD borders) (Fig. 3). Between the two LUHMES conditions we recorded about 25,000 active enhancers, 12,000 active promoters, and 40,000 occupied CTCF sites. There was a high degree of overlap between differentiation conditions for CTCF and for promoters, whereas the enhancer activation was much more distinct. This is consistent with previous work establishing that enhancers play a key role in modulating differences in gene expression between cell types and cell stages (Parker et al., 2013; Rubin et al., 2017; Shen et al., 2012). An example of differentiation-mediated increase in gene expression of *SNCA* and the dramatic formation of H3K27ac peaks at the *SNCA* locus, along with PD risk SNPs in the latter, is shown in Fig. 3E.

2.2. Enhancer–gene relationships

To further examine the relationship between active enhancer locations and gene expression changes, we related differentially expressed genes to differentially activated enhancers (Fig. 4A–D). Enhancers can regulate genes at over 1 Mb distant and also occur at high density near groups of active genes. The large number of expressed genes near active enhancers and the converse, enhancers near active genes, makes establishing relationships difficult. For instance, in the 100 kb window shown in Fig. 3E, there are 4 differentiated-specific enhancers (those active in differentiated but not undifferentiated cells) near the 3′ end of the upregulated gene *SNCA*. Within 1 Mb of all differentially upregulated genes in LUHMES are an average of approximately 26 active enhancers across conditions (12.4 differentiated-specific enhancers, 7.3 bi-conditionally active enhancers, and 6.4 undifferentiated-specific enhancers). However, correlative relationships can be clearly seen through filtering, such as by considering only the most significantly altered genes and enhancers in close proximity. By comparing the 300 genes with the greatest increase in expression and the 300 with the greatest decrease, it is apparent that the transcription start sites of the majority (about 66%) of the upregulated genes were within 100 kb of a differentiated-specific enhancer, while a minority were within 100 kb of an undifferentiated-specific enhancer (Fig. 4A). A similar ratio of genes with decreased expression were within 100 kb of an undifferentiated-specific enhancer and not near a differentiated-specific enhancer. Based

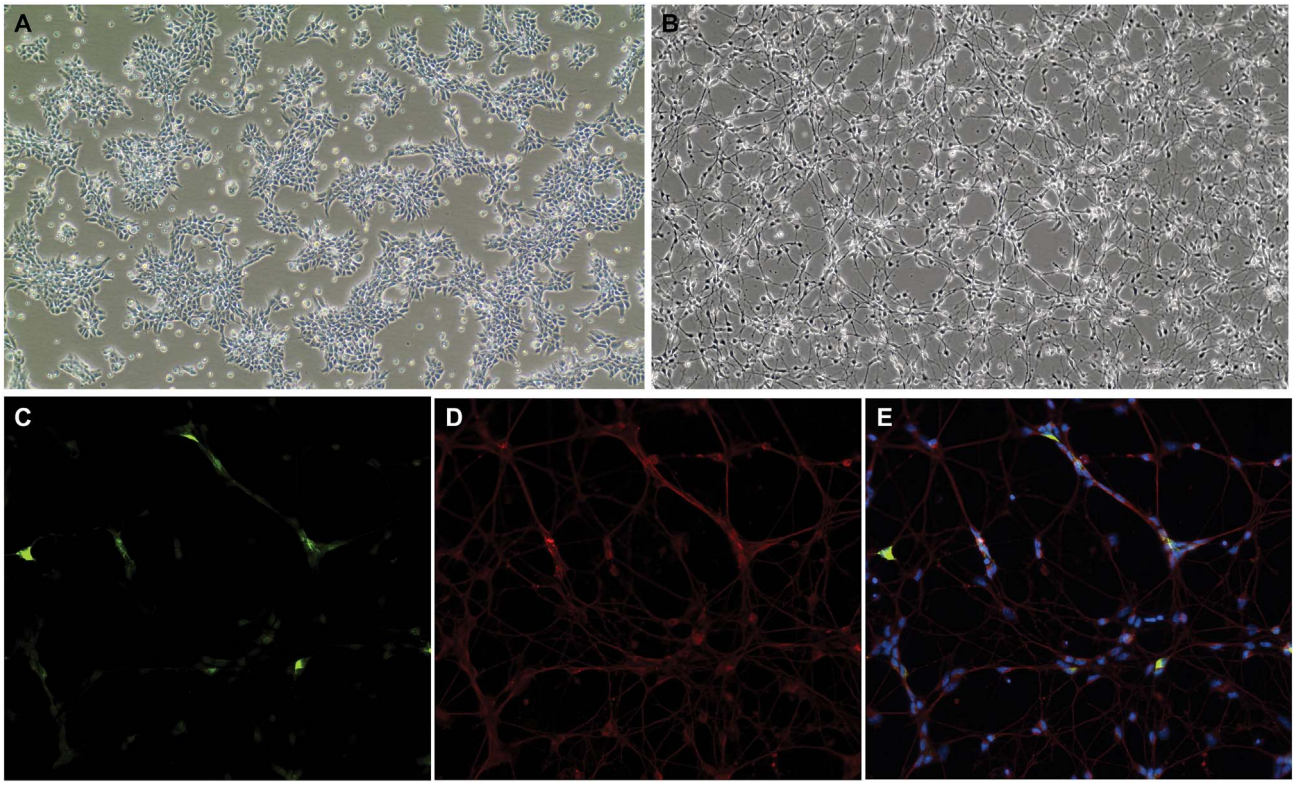
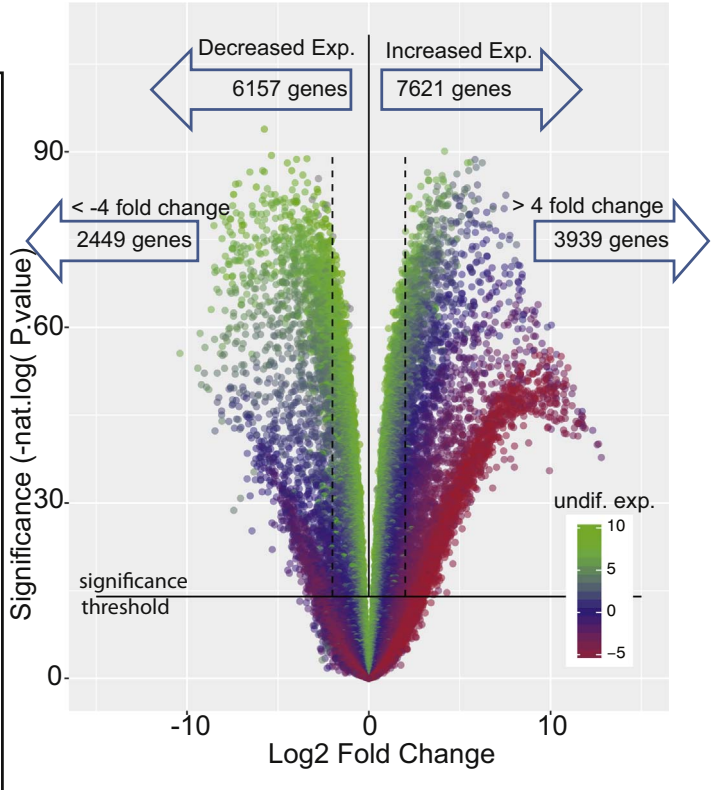


Fig. 1. Differentiation of LUHMES. DIC micrograph (60×) of LUHMES cells, (A) cycling, (B) differentiated for 6 days. Fluorescent micrographs (600×) of differentiated cells, (C) green, tyrosine hydroxylase (D) red, ßIII tubulin; (E) blue, DAPI; green, tyrosine hydroxylase; red, ßIII tubulin.

GO Category	P-value
cell cycle	1.2E-34
cell cycle process	8.1E-33
mitotic cell cycle	9.7E-33
mitotic cell cycle process	3.5E-29
chromosome segregation	4.8E-26
cell division	6.4E-24
ribosome biogenesis	1.1E-23
nuclear chromosome segregation	1.1E-22
rRNA processing	1.3E-22
rRNA metabolic process	4.3E-22
chromosome organization	6.9E-20
ribonucleoprotein complex biogenesis	3.0E-18
DNA replication	3.9E-18
sister chromatid segregation	6.4E-18
nuclear division	1.9E-17
ncRNA processing	2.5E-17
cell cycle phase transition	5.4E-16
organelle fission	7.7E-16
mitotic cell cycle phase transition	1.2E-15
DNA metabolic process	3.9E-15



GO Category	P-value
nervous system development	6.6E-53
trans-synaptic signaling	3.3E-34
synaptic signaling	3.3E-34
chemical synaptic transmission	2.2E-33
anterograde trans-synaptic signaling	2.2E-33
generation of neurons	2.5E-33
neurogenesis	1.0E-32
cell-cell signaling	4.2E-30
neuron differentiation	1.9E-26
cell-cell adhesion via plasma-membrane adhesion molecules	4.0E-25
homophilic cell adhesion via plasma membrane adhesion molecules	1.7E-24
regulation of membrane potential	1.5E-22
neuron development	3.2E-22
behavior	1.7E-20
cell-cell adhesion	2.7E-20
modulation of synaptic transmission	3.1E-20
cell adhesion	8.3E-20
biological adhesion	1.9E-19
regulation of nervous system development	2.2E-19
neuron projection dev	3.2E-18

Fig. 2. Differential gene expression. Volcano plot of RNA-seq. expression data pre- and post-differentiation. The horizontal line corresponds to a Bonferroni-corrected significance value of < 0.05. The dotted vertical lines bound the minimal fold-change for the most-differentially-expressed genes. To right and left: the top 20 GO functional categories for each (up or down) most-differentially-expressed gene set.

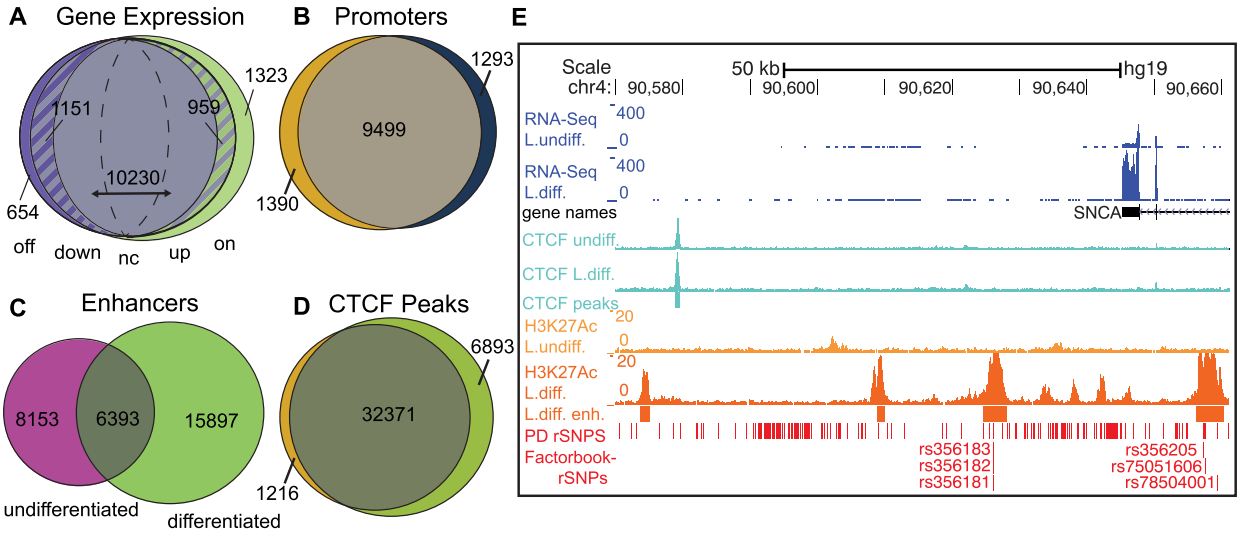


Fig. 3. Comparison of response elements and gene expression changes. Euler diagrams of overlap between LUHMES genomic elements under undifferentiated (left) and differentiated (right) conditions. (A) Gene transcripts with expression > 1 CPM. Solid colors indicate significant DE genes with < 1 CPM in the off condition. The striped area represents significant DE genes with > +/−4-fold change and that are expressed under both conditions. The dotted lines represent significant DE genes but with low fold change. (B) Promoters (+/−1 kb TSS) (C) enhancers, (D) CTCF Peaks. (E) Genomic browser view downstream of the SNCA locus. CTCF peaks are condition-independent but the increase in enhancer activity corresponds to greater exon fragment mapping in the differentiated cells. In red: PD GWAS risk SNPs (set: pd_all) and rSNPs overlapping enhancer region and also predicted (using the software, MotifBreakR) to show allele dependent disruption of TF binding motifs (Factorbook rSNPs).

on the most differentially expressed genes then, following differentiation into neurons, in general, genes that showed increased expression were near enhancers that become active, whereas genes that showed lower expression were near enhancers that became inactive. These enhancer-gene relationships represent the most common type of near-cis interaction but do not exclude interactions at greater distances and even those that occur in trans (between chromosomes).

We also noted nearby gene expression from an enhancer-centered perspective. Instead of counting enhancers within a window around gene start sites, we counted expressed genes within a window around the center of enhancers. At varying distances from condition-specific enhancers, we compared the ratio of differentially expressed genes that went up or down after differentiation into neurons. We found that differentiated-specific enhancers had more up-regulated genes nearby, while cycling-cell (undifferentiated-specific) enhancers had more down-regulated genes nearby. This relationship gradually came to match overall ratio of differentially expressed genes around 400 kb from active enhancers. This suggests that at closer than 400 kb there is a strong average causal relationship between the activation of an enhancer and the change in expression of nearby genes (Fig. S3).

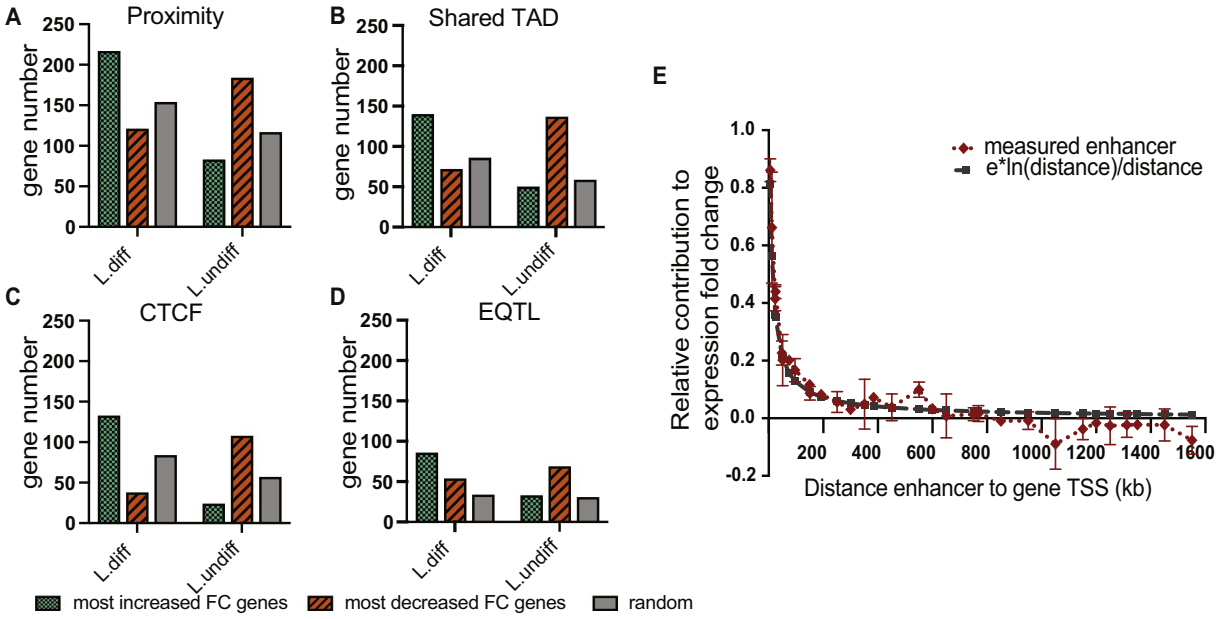


Fig. 4. Relationship between enhancers and target genes. In panels A–D: the top 300 (largest fold change) up-regulated, 300 down-regulated genes, and 300 random genes were counted based on whether they could be paired with at least one condition-specific enhancer, those that are uniquely present in undifferentiated (L.undiff) or differentiated (L.diff) LUHMES cells. The y-axis represents the number of genes (out of 300) which can define such a gene-enhancer pair. (A) Genes that are < 100 kb from one or more condition-specific enhancers. (B) Gene–enhancer pairs in close proximity and also entirely within a single GM12878-defined TAD. (C) Gene–enhancer pairs < 100 kb apart and with no intervening LUHMES CTCF peak. (D) Gene–enhancer pairs within 100 kb and corresponding to a GTEX-defined eQTL relationship. (E) Linear modeling coefficient values relating the number of condition-specific enhancers within bins at varying distances to the differential expression fold-change of the most-differentially-expressed genes (up or down).

We next compared the frequency of differentially expressed (DE) genes near active enhancers for which there are no intervening CTCF peaks, using the same 600 highest fold-change (FC) genes. One of the functions of CTCF is to act as an insulator, and it can be used to predict topologically associated domain (TAD) boundaries (Defossez and Gilson, 2002; Lu et al., 2016). As a comparison, we counted enhancer–gene associations that were located in a GM12878-defined TADs (Fig. 4B). However, in order to investigate TAD information specific to LUHMES, we examined whether adding CTCF binding site information increased the association of enhancers to the most over-expressed genes relative to proximity alone (Fig. 4C). We found that the total number of gene–enhancer associations was reduced, but the ratio of high-fold-change up-regulated genes near differentiated-specific enhancers (true positives) relative to the most down-regulated genes near the same enhancers (false positives) was increased. The same relationship was true for undifferentiated enhancers. In other words, removing gene-enhancer pairs with an intervening CTCF peak decreased the sensitivity but increased the specificity in pairing active enhancers to nearby high DE genes compared to proximity alone. This confirms that CTCF binding is significantly related to the interaction between active enhancers and genes and that this information can be used to improve enhancer–gene pairing.

Finally, we assessed whether existing expression-type quantitative-trait-loci (eQTL) data could be used to accurately associate conditionally specific active enhancers with the 600 highest fold-change genes (Fig. 4D). eQTL data relate the allele identity of common variants to nearby changes in gene expression, and is the basis of the gene names for the PD risk loci in the most recent PD meta-analysis (Chang et al., 2017). By considering all variants which overlap active enhancers in LUHMES cells, we queried GTEx (Consortium, 2013) data for any genes which have associated expression changes in either brain tissue or any tissue with our enhancers of interest. Interestingly, both cases produced a lower specificity and sensitivity of enhancer–gene association than proximity alone. This suggests that eQTL data is currently suboptimal for identifying regulatory relationships of the type we are studying here.

After comparing different methods that link enhancers and the most differentially expressed genes within 100 kb we next sought to define the relationships more generally. We constructed a linear model relating enhancer activity at varying distances to the fold-change in the differential expression of a larger set of DE genes. This model ultimately can be used to predict the likely relevant target genes to a regulatory element, for instance one which coincides with a GWAS-risk-associated SNP. The large number of enhancers relative to genes means that nearly all DE genes are within 1 Mb of both a cycling- and a differentiated-specific active enhancer. We performed multiple linear regression analyses to determine the predictive value of enhancer presence at different distances or with a CTCF peak in relation to measured RNA fold-change. We defined several parameters consisting of the count of active enhancers in bins of varying distances from genes of interest (see Materials and methods). Gene expression fold-change was significantly related to the number of differentiated-specific active enhancers nearby and inversely related to the number of undifferentiated-specific active enhancers nearby. In a model including the initial expression level as well as enhancer count variables for test genes, $r^2 = 0.453$ and the p-value was $< 2.2E-16$. The magnitudes of the coefficients and significance of associations for enhancer counts fell off according to the distance from enhancer to gene start site. At distances, greater than about 500 kb, only enhancers that had no intervening CTCF peak were significantly predictive. Predictive value was greatest for enhancers within 50 kb of a gene of interest.

The relationship between enhancer distance and gene expression fold-change followed an exponential decay function. We found that the relationship is well described by the function of enhancer to gene-start-site distance: $F(dist) = e * \ln(dist) / dist$ (Fig. 4E). Variables representing the sum of all enhancers within 1.6 Mb, weighted by this distance function and signed according to undifferentiated or differentiated status, explained approximately 10% of the variability in fold-change. These two variables, encompassing nearby enhancer burden, are more useful for predicting a binary description of differential expression rather than the continuous variable FC and are also further informed by including the number of active enhancers with no intervening CTCF site. This is shown in Fig. S4 where we assessed the use of the linear model in predicting differentially expressed genes based on a $\log_2$(fold change) of 2.

2.3. Role of transcription factors

Enhancers function through the recruitment of DNA-binding proteins combined with interaction at nearby gene promoters and with the influence of CTCF demarcation (Ren et al., 2017). Thus, it is likely that the activity of specific transcription factors (TFs) further contributes to the differential expression of genes and this explains why enhancer activation alone is not more predictive of gene expression levels. Indeed, a recent paper reports how information about both TF expression and enhancer activity can predict enhancer targets (Duren et al., 2017).

In order to examine different transcriptional regulation profiles, we obtained a list of 4182 human genes based on their annotation as “transcription factor” or “regulation of transcription” from the Gene Ontology Consortium and Uniprot. A total of 2869 of these are expressed in LUHMES cells, with 90.6% expressed under both conditions. Based on a minimal expression measurement of 1 CPM and at least a 4-fold change, statistically significant (< 0.05 adjusted p-value) undifferentiated LUHMES TFs numbered 332, whereas 299 TFs showed increased expression after differentiation. GO analysis of the differentiation TFs revealed enrichment for multiple ontologies, including several related to neurogenesis and axonogenesis.

To further examine which transcription factors may be instrumental in driving differentiation, a more specific subset of transcription factors with defined binding motifs was separately examined, based on data from ENCODE, Hocomoco, and Homer. This consisted of 755 transcription factors, of which 416 were expressed in LUHMES (82% under both conditions). Of expressed transcription factors with statistically significant and strong differential expression, 66 were more expressed in undifferentiated and 61 in differentiated cells (Table S3). Again, the set of differentiated TFs was enriched for known neurogenesis ontologies. We then checked for enrichment of known binding motifs within a subset of differentially active enhancers within 400 kb of the most (> 4 FC) differentially expressed genes (7588 differentiated and 3954 undifferentiated enhancers). Of the 416 expressed TFs with known motifs, 13 TFs were at least twice as enriched (had an odds ratio > 2) in differentiated enhancers relative to undifferentiated enhancers, and those 13—*CUX1*, *CUX2*, *LHX2*, *ONECUT1*, *ONECUT2*, *POU3F1*, *POU4F1*, *POU6F1*, *RFX3*, *RFX5*, *VAX2*, *ZFHX3*, and *ZNF740*—were also present in at least 1% of differentiated enhancers (Table 1, Table S3).

Finally, we looked for differentiation TFs in a third way by examining DNA binding sites based on ChIP-seq data sets collected by ReMap (Griffon et al., 2015). Binding sites for a total of 497 TFs across a variety of tissue types were available, giving a total of 2825 datasets. We counted the number of times that TF peak locations overlapped LUHMES enhancer locations for each TF. The binding peak locations for six TFs (*DUX4*, *EP300*, *KAT2B*, *ONECUT1*, *REST*, and *TAF1*) coincided twice as often, proportionally, with differentiated-specific enhancers than with undifferentiated-specific enhancers and also were present in at least 1% of enhancers (Table 2, Table S3). Interestingly, *REST*, a repressor of neuronal genes in non-neuronal tissues, showed statistically significant overlap but is starkly downregulated in differentiated LUHMES. This suggests that the same enhancer regions can be active but lead to opposite effects, and via different TFs, in different tissues. In contrast, ONECUT1, a transcriptional activator, overlaps differentiated-specific enhancers frequently and is upregulated in differentiated LUHMES cells.

Table 2
LUHMES differentiated-specific enhancers coincide with PD risk SNPs at 11 loci. Differentiated-specific enhancers that were within 400 kb of highly expressed and upregulated LUHMES genes were intersected with 6869 minimally significant (p-value < 0.000001) PD risk SNPs. Multiple enhancers at individual risk loci overlap PD rSNPS. The coordinates of enhancers, identifiers (rs#) for PD rSNPs, and the PD GWAS significance of the most significant index SNP (associated by LD to overlapping rSNPs) is listed. Genes which are upregulated, highly expressed, and near each enhancer (center of enhancer to TSS) are shown, those genes annotated with enriched GO categories are emboldened.

Locus count	Enh. chr.	Enh. start	Enh. stop	PD risk SNPs that overlap LUHMES enhancers (rs#)	SNP p-value	Nearby upregulated LUHMES genes
1	chr1	54555307	54558907	1981039, 4926619	6.00E−08	FAM159A, RP4-758J24.5
	chr1	54560654	54564069	7555099, 6588502	6.00E−08	FAM159A, RP4-758J24.5
2	chr1	205651656	205652285	61824663	7.00E−08	**CNTN2**, KLHDC8A, LEMD1, NFASC, PLEKHA6, RASSF5
3	chr2	135340364	135342182	842361	1.53E−07	CXCR4, **TMEM163**
	chr2	135367018	135368615	7573390, 35215000	1.15E−09	CXCR4, **TMEM163**
	chr2	135396134	135398670	10928507, 10803548, 6705916	1.48E−07	CXCR4, **TMEM163**
	chr2	135404783	135407527	13424016, 1568121, 1568120, 6724774, 6724777, 6724866, 6752634, 6742638, 6739706	2.66E−13	CXCR4, **TMEM163**
	chr2	135427298	135429266	7571113, 11898084	1.53E−07	CXCR4, **TMEM163**
	chr2	135430881	135436846	1123184, 11898465, 4954160, 4954161, 55920206, 35570087, 6430529, 6713239, 883964, 1104802, 1104801	1.15E−09	CXCR4, **TMEM163**
	chr2	135460532	135465801	28788021, 730946	6.17E−07	CXCR4, **TMEM163**
	chr2	135603195	135604690	10928520	8.98E−07	CXCR4, **TMEM163**
	chr2	135622243	135625233	6430552	8.98E−07	CXCR4, **TMEM163**
	chr2	135626435	135628962	7580655	8.98E−07	CXCR4, **TMEM163**
	chr2	135648859	135651589	55865348	8.98E−07	CXCR4, **TMEM163**
	chr2	135786533	135788334	16831264, 1551497	5.69E−08	CXCR4, **TMEM163**
	chr2	135789382	135791186	6709763, 6737635, 56369660, 60350840	2.98E−07	CXCR4, **TMEM163**
	chr2	136706117	136708430	6754311	8.96E−09	CXCR4, **TMEM163**
	chr2	136782731	136783764	6714750	2.48E−07	CXCR4, **TMEM163**
4	chr4	827328	828816	11727899	6.87E−10	**CPLX1**, RP11-1263C18.1, TMEM175
5	chr4	90412135	90413937	112744012	2.34E−07	NAP1L5, **SNCA**, TMSB4XP8
	chr4	90624633	90628074	356183, 356182, 356181	1.85E−82	NAP1L5, **SNCA**, TMSB4XP8
	chr4	90656356	90660382	356205, 3775423, 75051606, 78504001	3.48E−30	NAP1L5, **SNCA**, TMSB4XP8
	chr4	90834286	90838722	76707913, 17806425, 78586832, 10516850, 3775471, 3775473, 10516851, 12644375, 3775474	5.10E−18	NAP1L5, **SNCA**, TMSB4XP8
	chr4	91109824	91111760	9307081, 17016622	5.68E−07	NAP1L5, **SNCA**, TMSB4XP8
6	chr7	23135953	23137039	6967419, 10266123	2.24E−11	CCDC126, KLHL7, STK31
7	chr8	92045275	92047872	34250051	1.14E−07	**CALB1**, NECAB1, RUNX1T1, TMEM55A
8	chr11	77174125	77176121	12271542	9.03E−07	AQP11, KCTD21, RP11-111M22.4, RP11-111M22.5, USP35
9	chr16	31092740	31095147	750952	3.53E−07	ASPHD1, FAM57B, GDPD3, ORAI3, **PRRT2**, RP11-455F5.3, SEZ6L2, **STX1B**, TGFB1I1, YPEL3, ZNF843
10	chr17	17690230	17697458	3803763, 11649804, 11078398	6.00E−08	FAM211A, MYO15A, TOM1L2, TRPV2
11	chr17	43818184	43818636	17563827, 80072429	6.11E−49	AC003102.3, CRHR1, CTD-2020K17.3, FAM171A2, FMNL1, MAP3K14-AS1, **MAPT**, RUNDC3A
	chr17	43847744	43850000	56070245, 56387266, 34303488, 62055932, 62055933, 7225082, 62055934, 62055935, 62055936, 62055937, 76294809, 75916678, 79730878, 62055938, 62055939, 62055940	6.11E−49	AC003102.3, CRHR1, CTD-2020K17.3, FAM171A2, FMNL1, MAP3K14-AS1, **MAPT**, RUNDC3A
	chr17	44012628	44013985	77924424, 62061719, 17650381, 17650417, 62061720, 62061721, 79857651, 113756354, 12150111	6.11E−49	CRHR1, CTD-2020K17.3, FAM171A2, FMNL1, MAP3K14-AS1, **MAPT**
	chr17	44343836	44344244	2532343	6.11E−49	CRHR1, CTD-2020K17.3, FMNL1, MAP3K14-AS1, **MAPT**

2.4. PD risk loci

The LUHMES cell line is reported to be useful for the study of Parkinson's disease (Schule et al., 2009). We speculate that at least some PD relevance is specific to the differentiated cells and may not be present in the undifferentiated cells, or in other cell types. Thus, we determined enrichment of GWAS-linked Parkinson's risk SNPs in regulatory elements in LUHMES cells. We considered three partially overlapping sets of PD risk SNPs of different sizes and origins, the largest (pd_all) consisting of significant SNPs from three sources and expanded to include SNPs in linkage disequilibrium ($r^2 > 0.8$) (see Materials and methods). SNPs can only impose risk by functioning in active genomic regions, so by examining in which tissues or cell lines SNPs overlap active regions more than expected by chance, we can infer which cell types and cellular functions are relevant (schematically depicted in Fig. 5A). PD-risk SNPs are not enriched in active CTCF peaks or undifferentiated enhancers, but they are enriched in differentiated enhancers in LUHMES cells with a p-value = 1E−4 (Fig. 5B). This enrichment corresponds to 1.4% of PD risk SNPs (out of 23,918) overlapping differentiated-active enhancers (including those also present in undifferentiated cells) compared to a background level of 1.2%, describing the proportion of all SNPs (135,276,726) that overlap the same enhancers.

As an examination of the significance of this PD risk SNP enrichment, we compared the risk SNPs for other diseases and enrichment in other tissues (Fig. 5C). Based on the largest set of PD risk SNPs (pd_all), PD risk loci are enriched for the enhancers of, and so likely active in, blood or immunological cells and brain tissue. PD rSNPs are even more enriched in differentiated-specific enhancers in LUHMES (L.diff.excl) and most enriched in the smaller subset of differentiated-specific enhancers that are very near highly upregulated genes (L.diff.up.genes). This enrichment indicates that PD risk SNPs are non-randomly located with respect to differentiated-specific LUHMES enhancers and, so, the identity of these enhancers and target genes is likely informative in understanding PD etiology.

Because > 90% of PD-risk SNPs (i.e., GWAS index SNPs plus

Table 1
Transcription factor binding sites are enriched in differentiated LUHMES cell enhancers. The top portion of the table lists the identity and consensus motif for TFs which were present in differentiated-specific (Diff.) enhancer (Enh.) regions at a rate at least 2-fold higher than in undifferentiated-specific (Undiff.) enhancer regions. The bottom portion of the table identifies the binding site (ChIP-seq) datasets of TFs that overlap differentiated-specific enhancer locations. TF expression and DE fold change (FC) in LUHMES is displayed, as is the number of LUHMES enhancers that overlap TF peaks or TF motif locations (total enhancer number in parenthesis). Consensus motifs are in standard IUPAC nucleotide code. Odds ratio (OR) represent the ratio of frequencies (Freq.) in differentiated to undifferentiated. Significance is given for differentiated-specific enhancer-overlap compared to any-LUHMES-enhancer-overlap, as well as differentiated-specific compared to undifferentiated-specific overlap.

Protein	Undiff. Exp. log2 (CPM)	Diff. Exp. log2 (CPM)	DE log2 (FC)	Data source	TF consensus binding motif	All Enh. overlap (30,443)	Diff. Enh. overlap (7588)	Undiff. Enh. overlap (3954)	Freq. in Diff.	Freq. in Undiff.	OR	Signif. Diff. vs. all	Signif. Diff. vs. Undiff.
CUX2	−4.58	4.27	9.08	ENCODE	TRATCRATAHz	2103	634	99	0.08	0.03	3.34	7.8E−09	2.3E−35
ONECUT1	−6.29	2.88	9.15	HOMER	NTATYGATCH	2723	832	141	0.11	0.04	3.07	1.3E−12	3.6E−41
POU6F1	1.61	1.96	0.38	ENCODE	VHNVWTAATKAGSWDDH	942	251	49	0.03	0.01	2.67	1.0E−01	8.8E−12
ONECUT1	−6.29	2.88	9.15	ENCODE	VRAWAATCRATAHH	453	106	21	0.01	0.01	2.63	7.6E−01	1.2E−05
ONECUT2	5.44	9.45	4.06	ENCODE	VRAWAATCRATAHH	453	106	21	0.01	0.01	2.63	7.6E−01	1.2E−05
CUX1	6.00	6.00	0.04	ENCODE	TRATCRATMH	2909	851	170	0.11	0.04	2.61	1.0E−08	1.1E−33
POU4F1	8.26	9.22	0.97	ENCODE	NTRMATWWTTWATK	1745	443	94	0.06	0.02	2.46	3.1E−01	3.8E−17
POU3F1	5.62	6.04	0.45	HOCOMOCO	VRTKSTWATGCWWD	421	99	22	0.01	0.01	2.34	7.3E−01	1.5E−04
RFX5	4.08	3.99	−0.06	ENCODE	NDRKHRCBMWRGYAACVD	296	88	20	0.01	0.01	2.29	2.5E−02	4.7E−04
LHX2	−1.92	7.05	9.00	ENCODE	WVHAYYAATTRRYKNNN	1043	258	60	0.03	0.02	2.24	5.4E−01	2.8E−09
POU6F1	1.61	1.96	0.38	ENCODE	RCATAAWTWAT	1059	257	63	0.03	0.02	2.13	6.8E−01	2.0E−08
VAX2	3.47	4.70	1.23	ENCODE	BDNNRYTAATTAVBVS	813	221	56	0.03	0.01	2.06	6.2E−02	4.6E−07
ONECUT1	−6.29	2.88	9.15	HOCOMOCO	DWYATTGATTWHDH	1644	361	92	0.05	0.02	2.04	1.0E+00	1.8E−10
ZNF740	5.85	5.48	−0.33	ENCODE	MCCCCCCCAY	1275	413	106	0.05	0.03	2.03	3.6E−10	1.6E−11
ZFHX3	4.15	5.91	1.81	HOCOMOCO	RTTAATWATTW	1514	346	89	0.05	0.02	2.03	9.7E−01	6.1E−10
RFX3	5.96	6.83	0.90	HOCOMOCO	BGTTRCCATGGHRN	341	97	25	0.01	0.01	2.02	5.9E−02	1.1E−03

Protein	Undiff. Exp. log2 (CPM)	Diff. Exp. log2 (CPM)	DE log2 (FC)	ChIP tissue source	TF peak number total	All Enh. overlap (30,443)	Diff. Enh. overlap (7588)	Undiff. Enh. overlap (3954)	Freq. in Diff.	Freq. in Undiff.	OR	Signif. Diff. vs. all	Signif. Diff. vs. Undiff.
EP300	6.56	6.70	0.18	Neural	47,966	4399	1134	213	0.15	0.05	2.77	7.6E−02	2.2E−47
REST	5.85	1.26	−4.54	Neural	57,775	5840	1566	312	0.21	0.08	2.62	1.0E−04	1.7E−57
ONECUT1	−6.29	2.88	9.15	ESC H9	50,519	3151	839	173	0.11	0.04	2.53	9.6E−03	1.1E−31
KAT2B	2.98	5.45	2.51	HepG2	2776	255	90	21	0.01	0.01	2.23	7.9E−05	5.6E−04
REST	5.85	1.26	−4.54	HCT166	4423	192	81	19	0.01	0.00	2.22	5.0E−08	9.7E−04
REST	5.85	1.26	−4.54	HepG2	5375	248	98	23	0.01	0.01	2.22	1.3E−07	3.2E−04
TAF1	0.12	5.67	5.74	Neural	22,267	1692	340	85	0.04	0.02	2.08	1.0E+00	2.7E−10
REST	5.85	1.26	−4.54	K562	7909	419	150	38	0.02	0.01	2.06	2.3E−07	3.5E−05
DUX4	NA	NA	NA	HEK293	9020	355	96	25	0.01	0.01	2.00	1.6E−01	1.4E−03

surrogates in linkage disequilibrium) occur in non-coding DNA, we explored the possibility that many of them act at enhancers, which can be associated with gene regulation the LUHMES cells. In this paper, we have described several indirect ways to match differentially expressed genes to enhancers in LUHMES cells (e.g. proximity vs. proximity combined with CTCF binding information) and have categorized enhancers in multiple ways (e.g. differentiated-specific vs. differentiated-specific and also close to highly upregulated genes). There are also multiple potential sets of PD risk SNPs that can be considered. Due to this complexity, there is no single appropriate way to overlap risk SNPs with enhancers and then obtain a single linked set of genes. However, we compared multiple methods and found similar results. We will describe three methods and also provide the lists of risk SNPs, enhancer locations, and DE gene locations as supplemental material.

Firstly, we matched the risk SNPs locations with the presence of active enhancers as annotated by H3K27ac specifically in differentiated or undifferentiated LUHMES cells or both. We used the largest set of SNPs coming from three data-bases (pd_all) and filtered these for a minimum significance (p-value < 0.000001). Thus, we identified 203 PD-risk SNPs that intersect 73 enhancers (in differentiated LUHMES, 41 out of 15,897; shared, 15 out of 6393; in undifferentiated LUHMES, 17 out of 8153). Using the most specific gene pairing method, those 73 enhancers are associated with 87 genes by proximity (< 1.6 Mb) and not interrupted by a CTCF peak. Twenty-eight of these genes are expressed in differentiated LUHMES cells (> 1 CPM). Among this set, 4 genes (GAK, *SNCA*, *STX1B*, *STX4*) are sufficient to give statistical enrichment for "synaptic vesicle transport" (p-value = 0.0000055).

We also determined the overlap of risk SNPs with only those differentiated-specific enhancers that are also near genes which are both highly expressed and up-regulated in differentiated LUHMES cells. This represents a subset of enhancers which we think are most influential in differentiation, and so are particularly good experimental targets. In this case, 33 enhancers overlapped risk SNPs corresponding to 11 loci (Table 2). These were associated (solely by proximity, the most sensitive pairing method) with 52 genes that are within 400 kb, are highly expressed, and showed increased expression after differentiation. Eight of these loci gave genes (*CALB1*, *CNTN2*, *CPLX1*, *MAPT*, *PRRT2*, *SNCA*, *STX1B*, *TMEM163*) that lead to statistical GO enrichment of "synapse", "axon part", and "locomotory behavior" (respectively, FDR = 0.025, 0.002, 0.0039).

Finally, we considered only the PD-risk loci reported in the most recent meta-analysis of PD GWAS results (Chang et al., 2017), labeled "pd_2017" in Fig. 5c. We considered both the reported highest risk index SNP and SNPs in linkage disequilibrium (LD), as well as the putative causal genes for each of the 41 loci. This is the smallest set of risk SNPs and does not include the majority of significant SNPs at a given locus (only the most significant), but represents the most recent definitive list. Unlike the other two rSNP-sets it is not enriched for colocalization with LUHMES enhancers. Adding LD SNPs ($r^2 > 0.8$) to the index risk SNPs produced a set of 4514 SNPs, 3232 of which were associated with the single large LD block of the MAPT locus at chromosome 17. A total of 16 LUHMES enhancers overlapped these risk SNPs which corresponded to seven general risk loci. Nine of these enhancers are differentiated-specific LUHMES enhancers, corresponding to four

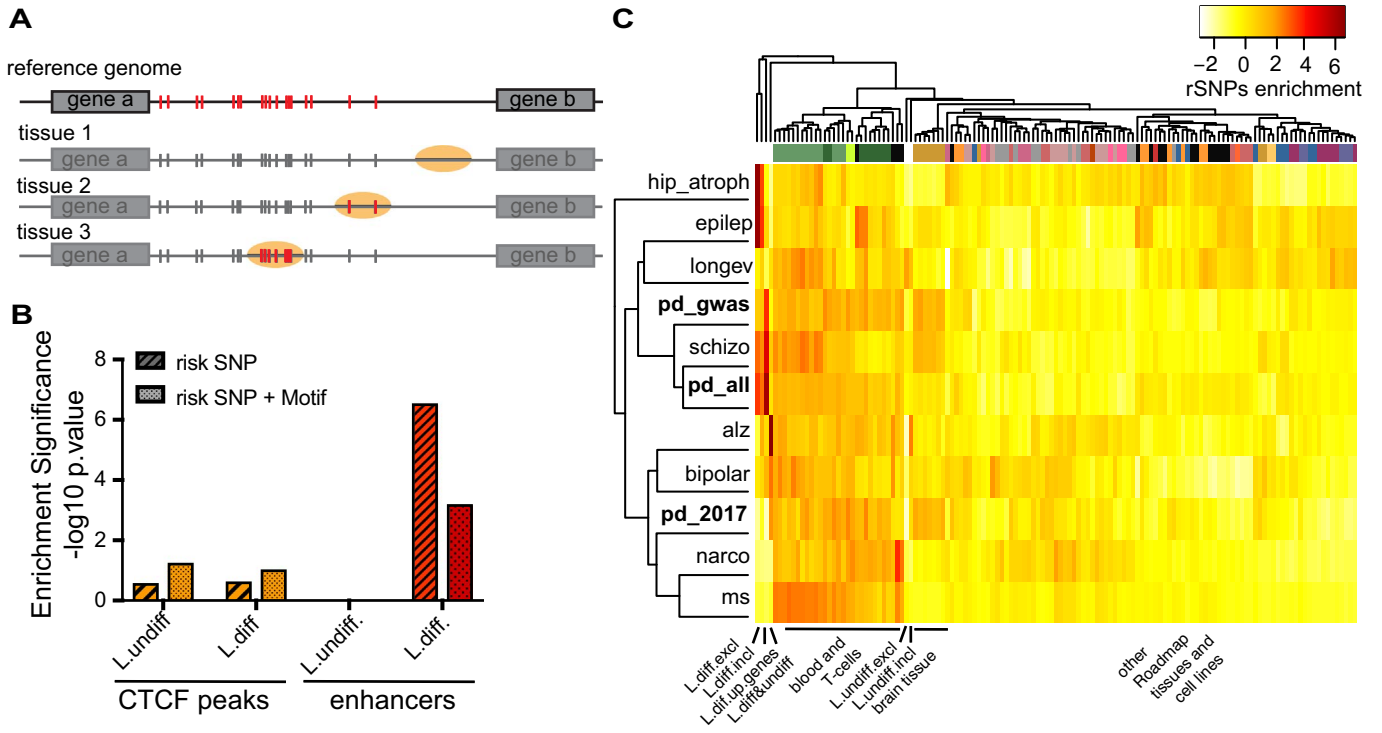


Fig. 5. Enrichment of PD rSNPs in enhancer elements. (A) Conceptual schematic conveying the intuition for comparing tissue-dependent regulatory element activity or accessibility (shown as orange ovals) with rSNP genomic coordinates. Statistical enrichment (greater overlap than expected by chance) implies that SNPs are capable of imposing risk in enriched tissue via specifically active regulatory elements and the given tissue may, therefore, be disease relevant. (B) Enrichment of all PD rSNPs or PD rSNPs with significant predicted TF motif disruption (+Motif) in CTCF peaks or enhancer peaks present in all undifferentiated (L.undiff) or differentiated (L.diff) LUHMES cells. (C) Enrichment of rSNPs for multiple GWAS disorders in LUHMES enhancers or Epigenomics Roadmap tissue and cell-type enhancers. Values are Z-score normalized odds ratios comparing the tissue specific enhancer overlap for rSNPs vs. all 135 million dB144 SNPs for each disorder. Red indicates that the proportion of rSNP-enhancer overlap compared to all SNP-enhancer overlap is particularly high for that disease. L.diff.incl: all enhancers in differentiated; L.diff.excl: exclusive enhancers, in differentiated but not undifferentiated; L.diff.up.genes: enhancers in differentiated but not undifferentiated and close to the most unregulated genes; L.diff&undiff: shared enhancers present in both undifferentiated and differentiated LUHMES. Colors on top are standard for tissue categorization, the full set of values is included as Table S4.

PD risk loci and are associated (by proximity) with only five additional genes– *SLC25A20*, *KLHDC8B*, *COL7A1*, *C3orf62*, and *CELSR3*—which were not already implicated previously (Table S5).

In the paper, 71 candidate risk genes were reported, which had been associated by eQTL. Of these, 15 are not expressed in either LUHMES condition, 49 are clearly expressed under both differentiation conditions, and 7 are only expressed (> 1CPM) in a single condition. Of the 56 genes expressed in at least one LUHMES condition, 12 are more highly expressed (> 4-fold) in differentiated cells, and 3 are more highly expressed in undifferentiated cells. Whereas the 56 genes as a set are not enriched for any gene ontology annotations, the 12 genes that are up-regulated in differentiated LUHMES (*ANK2*, *ATP6V0A1*, *CRHR1*, *GALC*, *KLHL7*, *MAPT*, *SCN3A*, *SNCA*, *SYT4*, *TMEM163*, *TMEM175*, and *TMEM229B*) are enriched for the cell component categories of “synapse part”, “synapse”, “synaptic vesicle”, “presynapse”, and “cytoplasmic vesicle part”.

3. Discussion

In this study, we sought to determine both which gene regulatory networks are activated during the differentiation of mesencephalic cells into functional dopaminergic neurons and which regulatory elements and target genes likely impinge on PD risk. We found that differentiation of LUHMES cells resulted in statistically-significant differential expression of 11,157 out of 14,317 (78%) high-expression (> 1 CPM) genes. Of these, 4087 (29%) were strongly altered by at least 4-fold (Fig. 3A). These gene expression profiles corresponded to (and were almost certainly due to) a set of approximately 30,000 enhancers, of which nearly 80% were unique to one of the two differentiation conditions.

The spatial relationship between enhancers and gene expression changes is clear for the gene set on average, with a halving of effect size for every doubling of genomic distance between enhancers and gene transcription start sites. This is similar to the relationships found by other groups (Cao et al., 2017). The differentially expressed genes had a mean of approximately 10–15 active enhancers within 1 Mb (depending on condition), and in our data, we cannot distinguish between a model in which more-distant enhancer–gene pairs have increasingly weak interactions with distance vs. a model in which distant enhancer–gene pairs have increasingly less probable interactions of a large fixed effect size. A comprehensive comparison of enhancer deletions would distinguish these possibilities. In the first model, most distant enhancer deletions would lead to a small change in target gene expression. In the second model, most distant deletions would result in no change to a target gene but a small number of deletions would result in a very large change. Our additive models relating expression fold-change to enhancer location explained only 10% of variability, and specific enhancer deletions by us (data not shown) and by others have led to vary large changes irrespective of distance (Yao et al., 2015). However, both models are relevant to understanding regulatory networks and to predicting likely outcomes of enhancer sequence polymorphisms.

The correlation between enhancer activation and gene regulation provides circumstantial evidence to link specific enhancers to specific nearby genes within a neuronal context. Our working hypothesis is that enhancers do not control the expression of only single genes, as most GWAS analyses imply, but that many genes are affected by a particular locus, thus orchestrating overall disease risk. This is consistent with ENCODE data which estimates that enhancers regulate an average of 2.5 genes (Mumbach et al., 2017). Parkinson's disease (like most GWAS studied complex traits) has associated polymorphisms which are

primarily intergenic or intronic and non-coding. Therefore, the underlying causal gene expression changes, which mechanistically link risk loci to PD, are likely due to differences in enhancer activity of the sort which we are dissecting here. Enhancer activation profiles are characteristic of, and drive the differences in, different tissues, and so statistically significant differences in risk SNP/enhancer overlap between tissues is informative and can confirm that a tissue or cell type is disease relevant (Corradin et al., 2016). Here we found that differentiated LUHMES have significant enrichment for PD-risk SNPs in enhancers and undifferentiated LUHMES cells do not. It is perhaps unsurprising that undifferentiated cells do not show risk SNP enrichment; because PD is accompanied by specific loss of a single type of neuronal tissue it seems unlikely that undifferentiated neuronal precursor cells are relevant. However, the finding that differentiated LUHMES cells do show more PD-risk SNPs in enhancers than expected by chance indicates that PD related processes are active in this neural model.

Further dissection of risk enhancers is important, though, and there remain two basic ways that neuronal risk factors could be causal in PD. Firstly, some loci may impair the ability of cells like LUHMES to *become* fully differentiated. The effect of this for PD patients may be a reduced basal number of functional, healthy dopaminergic neurons in the substantia nigra. There is evidence that a smaller starting population of neurons predisposes a brain to the earlier display of Parkinson-like symptoms from neuronal death (Pakkenberg et al., 1991; Rubin et al., 2017). Interestingly, some genes including, *CBFA2T2*, *NEUROG2*, *NEUROD4*, and *TCF3*, which are thought to be temporarily active during the transition into fully differentiated neurons, are still expressed in our differentiated LUHMES cells (Aaker et al., 2009). Hence, we are likely measuring both transitional as well as fully differentiated cell signature activity, and as such, we may be capturing signals for both the functioning of dopaminergic neurons and the differentiation process itself.

The second explanation is that some risk loci alter processes that are required for neuronal function but which are not active in precursor cells. The enrichment of vesicle trafficking and synapse-related gene pathways indicates that some risk loci are of this type. However, these processes are also present in other neuronal types. This raises the common question, then, of what makes PD cell death generally so anatomically specific. Is there PD biology (perhaps involving the loci not associated with vesicle transport) which is unique to dopaminergic neurons such as LUHMES, or is disease due to confluence of multiple factors that merely coincide frequently in cells of this kind?

LUHMES is known to be a relevant PD model but may be relevant for other diseases as well. Comparing the overall enrichment of PD-risk SNPs in enhancers in LUHMES cells to all Roadmap tissue enhancers and to other neurodegenerative diseases indicated that LUHMES cells are more enriched for risk SNPs in enhancers for multiple diseases than any bulk Roadmap tissue alone (Fig. 5C). PD-risk SNPs are enriched in enhancers of blood and brain tissues generally but are most enriched in LUHMES. LUHMES cells also show greater PD rSNP enrichment in a focused subsets of active differentiated enhancers relative to hippocampal atrophy or epilepsy, for instance. This suggests that the differential effects we have measured are more relevant to PD than these other diseases. Interestingly, Alzheimer risk SNPs are more enriched in LUHMES cells than in other Roadmap tissues, but they are most enriched in the category of enhancers that are active in *both* undifferentiated LUHMES and differentiated LUHMES. One explanation may be that, unlike in PD, genes common to multiple neuronal cell types are affected by Alzheimer risk loci and so these are accompanied by enhancers that remain stable after LUHMES differentiation. Interestingly, LUHMES cells, in either differentiation state, use these risk enhancer sets more clearly than Roadmap bulk tissue.

Finally, in this study we hoped to have identified target genes and processes for PD relevant experiments in LUHMES cells. In theory, the most direct way to identify SNP-associated gene targets is through eQTL studies, which relate allele identity to tissue-specific gene expression differences that are apparent in large populations of donors (Chang et al., 2017). Unfortunately, well-known biases in these studies include donor demographics, sample size, heterogeneous cell populations in tissue samples, and statistical constraints that prevent discrimination of effects in genes with sub-optimal baseline expression or variability levels. Such biases currently cause this technique to be insufficient. The problem is even more pressing for neurodegenerative diseases such as Parkinson's and Alzheimer's, where at least some genetic risk is likely intrinsic to a neuronal subpopulation of cells in the brain regions affected. Noise from heterogeneous tissue is likely similar to what we saw here in comparing Roadmap tissue enhancer/SNP enrichment. Likewise, we used our neuronal model to evaluate SNP–gene pairing methods in which differentiated-specific enhancers were associated with differentially expressed genes. We found that eQTL-based pairing using GTEx data was less effective in capturing regulatory element/gene-expression relationships. Instead, mere proximity or proximity combined with CTCF binding location was more informative. As such, we don't believe eQTL studies have currently captured reliable relationships between risk loci and nearby genes that describe the activity specific to neurons like LUHMES cells. By either proximity or proximity combined with CTCF binding information, we predict up to 80 genes (Table S5) are affected by the risk alleles of PD-risk SNPs which are located within differentiated-specific enhancers. These genes are largely unrelated, leaving open the potential identification of new disease-relevant processes. On the other hand, the enrichment for synapse and vesicle trafficking categories indicates both that the risk-gene association described here is meaningful and highlights the importance of these categories to PD.

Our study validates the use of LUHMES as an in vitro model for the study of mechanisms involved in PD risk. By replacing PD risk alleles at enhancers identified here by using CRISPR/cas9, future insights in PD biology can be gained.

4. Materials and methods

4.1. LUHMES culture

LUHMES cells were cultured essentially as done by (Scholz et al., 2011). Briefly, the cells were incubated in a humidified 37 °C, 5% CO_2 incubator on 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 2 mM 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. Prior to differentiation, cells were seeded at 3.5×10^6 per T75 flask containing complete growth medium and incubated at 37 °C for 24 h. For induction of differentiation, culture medium was changed to freshly prepare DMEM:F12 with 2 mM L-glutamine, 1X N-2, 1 mM 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).

4.2. Immunocytochemistry

LUHMES cells were grown as above and switched to differentiation medium 48 h prior to being trypsinized and re-plated on coverslips coated with PLO and fibronectin (as above) in 24 well plates at a density of 50,000 cells per well. Cells were allowed to differentiate up to a total of 6 days. Cells were fixed with 4% paraformaldehyde in 1 × PBS for 20 min and processed for immunocytochemistry. First, non-specific sites were blocked with 0.2% bovine serum albumin, 0.5% Triton X-100, and 0.05% Tween 20 in PBS for 1 h at room temperature.

Cells were then incubated with primary antibodies: TH (1:1000, Pel Freez P40101-0); Tuj-1 (8 μg/mL, R&D Systems MAB1195) at 4 °C overnight. Appropriate secondary antibodies (Alexa Fluor 488 or 594, Invitrogen, Carlsbad, CA, USA) were used followed by incubation with DAPI to stain the nucleus. The coverslip-containing stained cells were washed twice with PBS and mounted on slides. Cells were viewed under a NIKON Eclipse Ni-U fluorescence microscope (Nikon, Melville, NY, USA); images were captured with a Retiga Exi digital camera using NIS Elements AR 4.00.08 software (Nikon).

4.3. ChIP-seq

For ChIP experiments, we used previously published protocols (Rhie et al., 2014). Briefly, about 1×10^7 cells were fixed by adding fresh formaldehyde directly to the culture medium at a final concentration of 1%. The reaction was quenched with 10× (1.15 M) glycine for 5 min at room temperature. Chromatin from fixed cells was sonicated using a Bioruptor Pico (Diagenode, Cat # B01060001) with 30 s on and 30 s off cycles to produce fragments between 200 and 500 base pairs. For immunoprecipitation, 100 μg of sonicated chromatin was used and 10 μg (10%) was saved as an input control. To probe for active enhancers in both undifferentiated and differentiated cells, samples were incubated at 4 °C overnight with an H3K27ac primary antibody (Active Motif, Cat # 39133) or an IgG control (Sigma, Cat # R9133). For CTCF sites, CTCF (D31H2) XP (Cell Signaling Technology, cat # 3418) was used as primary antibody. For secondary antibody, A/G magnetic beads (Pierce, Cat # 88802) were added to the samples prior to an additional incubation for 2 h at 4 °C. The beads were then washed with a series of salt buffers before elution. The immunoprecipitated and input control DNA was purified using A QIAprep Spin Miniprep Kit (Qiagen, Cat # 27104).

4.4. Construction and sequencing of directional mRNA-seq libraries

Libraries were prepared by the Van Andel Research Institute Genomics Core from 1 μg of material using the KAPA Stranded mRNA-seq Kit (v4.16) (Kapa Biosystems, Wilmington, MA USA). RNA was sheared to 250–300 bp. Prior to PCR amplification, cDNA fragments were ligated to Bio Scientific NEXTflex Adapters (Bioo Scientific, Austin, TX, USA). The quality and quantity of the finished libraries were assessed using a combination of Agilent DNA High Sensitivity chip (Agilent Technologies, Inc.), QuantiFluor dsDNA System (Promega Corp., Madison, WI, USA), and Kapa Illumina Library Quantification qPCR assays (Kapa Biosystems).

4.5. Differential gene expression analysis

In total, 12 biological replicates were sequenced: 6 were sequenced using single-end libraries and 6 were sequenced with both single end and paired-end sequencing libraries, generating a total of 18 RNA-seq datasets. Briefly, a parental LUHMES culture was split into 6 cultures and grown to 80% confluence, from which RNA was isolated, and each culture was also passaged to new flasks. These 6 new cultures were grown 24 h then switched to differentiation media, from which RNA isolated 6 days later. These 12 samples were used to produce single-end RNA libraries and 6 of the 12 (3 from undifferentiated and 3 from differentiated (sample numbers 1–3 and 10–12)) were also used to produce paired-end libraries. Following sequencing, fastq files were aligned to HG19 using STAR v2.5 (Dobin and Gingeras, 2015). Alignments (bam files) were converted to feature counts using HTSeq v0.6.0 referenced against the ENSEMBLE annotation of HG19: Homo_sapiens.GRCh37.87.gtf counting against the feature "exon", grouped by "gene_id", and using the strand parameter "reverse". This set included exon locations for 57,905 genomic entities including pseudogenes, lncRNAs, and 20,356 protein coding genes. The resulting gene_id map counts were normalized using edgeR (TMM) and tested for significant differential expression with Limma and Voom, in R (v3.3.1) (Law et al., 2014; Ritchie et al., 2015; Robinson and Oshlack, 2010). The normalized count data for all 18 datasets revealed, through principle component analysis (PCA), that there was very high similarity between paired-end and single-end mapping relative to differentiation status (Fig. S2). Therefore, the datasets were combined for analysis. However, Voom weighted the single end sequencing counts more highly than paired-end data – meaning statistical significance is largely independent of the paired-end data sets. More complicated modeling designs, which treated the paired-end and single-end libraries as separate design variables, produced roughly the same end results. It should also be noted that Voom recommends an initial filtering out of low count genes prior to statistical modeling. We found that the significant DE results differed little between different pre-filtering conditions, only serving to alter the marginal DE cases and to reduce the gene set size, and so elected to jointly evaluate all 57,905 genes without a filtering step. The log2 FC and p-value for DE, as well as normalized log2CPM were derived for each gene for each replicate and presented in Fig. S1. In total 57,905 genes were mapped and a conservative Bonferroni adjusted p-value of 0.05 was used to define the significance threshold. An additional cutoff of at least $+/-2$ log2FC was used to define the "most" differentially altered genes. In addition, for each gene a *t*-test was used to determine whether expression was likely less than or equal to 0 log2(CPM) on average.

4.6. Construction and sequencing of ChIP-seq libraries

Libraries for Input and IP samples were prepared by the Van Andel Genomics Core from 10 ng of input material and all available IP material using the KAPA Hyper Prep Kit (v5.16) (Kapa Biosystems, Wilmington, MA USA). Prior to PCR amplification, end-repaired and A-tailed DNA fragments were ligated to Bio Scientific NEXTflex Adapters (Bioo Scientific, Austin, TX, USA). The quality and quantity of the finished libraries were assessed using a combination of Agilent DNA High Sensitivity chip (Agilent Technologies, Inc.), QuantiFluor dsDNA System (Promega Corp., Madison, WI, USA), and Kapa Illumina Library Quantification qPCR assays (Kapa Biosystems). Sequencing (75 bp, single end) was performed on an Illumina NextSeq 500 sequencer using a 75-bp sequencing kit (v2) (Illumina Inc., San Diego, CA, USA). Base calling used Illumina NextSeq Control Software (NCS) v2.0, and the output of NCS was demultiplexed and converted to FastQ format with Illumina Bcl2fastq v1.9.0.

4.7. Identification of ChIP-seq peaks

Two biological replicates with 2 technical replicates each were used for input and ChIP for both H3K27Ac and CTCF. Following sequencing, fastq files were aligned to the HG19 genome assembly using default setting for BWA v0.7.15 (Li and Durbin, 2009). Aligned reads were merged with Samtools and peaks were called using MACS2 v2.1 using a liberal FDR cutoff of 0.1 (Zhang et al., 2008). Narrowpeaks were aligned and an irreproducible discovery rate (IDR) of < 0.01 was used for filtering (Li et al., 2011). The R package ChIPseeker was used with TxDb·Hsapiens.UCSC.hg19.knownGene to define H3K27Ac peaks as proximal promoters if within 100 bp of a gene TSS or as enhancers if further than 2 kb from a TSS.

4.8. Shared enhancer or CTCF peak loci identification

Bedtools v2.26.0 was used to intersect CTCF and H3K27ac peaks between differentiation conditions based on a default overlap of 1 bp overlap (Quinlan and Hall, 2010). H3k27ac peaks that were present in both conditions were merged to form the shared peak set. CTCF peaks showed high overlap and so, except when otherwise indicated, the undifferentiated and differentiated CTCF peak files were merged and combined to generate the set of LUHMES CTCF peak locations.

4.9. Enhancer–gene association

All highly differentially expressed genes (Bonferroni adjusted p-value < 0.05, abs(log2FC) > 2) were paired with all LUHMES enhancers within 1.6 Mb. These enhancer–gene pairs where then filtered according to the parameters specified in the text, i.e. removing no pairs (distance only), removing pairs with a CTCF peak within (CTCF), removing pairs not within predefined TAD region, or removing those not independently associated by eQTL. Except for analysis shown in Fig. 5c, enhancer peaks that were present in both differentiation conditions (intersecting by at least 1 bp) were excluded from the further analysis. R v3.3.1 based linear modeling and robust linear modeling was used to construct a model relating enhancer count in different distance, as well as initial gene expression to expression fold-change. Multiple bin sizes and number were related to FC, and the derived coefficients were aggregated and plotted by mid-bin distance to estimate the relative predictive contribution for an enhancer at varying distance. Because undifferentiated enhancers reduced FC and differentiated enhancers increased FC, the undifferentiated enhancer coefficients were inverted and combined to produce 2 (similar) coefficients for each distance as seen in Fig. 5c. The theoretical function fits the data with a residual standard error of 0.028. This function relating gene fold change to enhancer count was trained with gene-enhancer pairing based on proximity only.

4.10. Gene ontology analysis

Gene ontology enrichment analysis was done using Panther and String against the gene sets specified in the text (Franceschini et al., 2013; Mi et al., 2013).

4.11. Transcription factor enrichment

DNA binding proteins with defined binding motifs were obtained from MotifbreakR v 3.5 (Coetzee et al., 2015) by selecting those sequences which were annotated as Hocomoco, ENCODE, and Homer. Uniprot IDs were converted to ensemble ID and those TFs which were also expressed in LUHMES were examined further using the summarizePatternInPeaks function in the R package, ChIPpeakAnno V3.10.2. For the TF binding data analysis, ChIP-seq binding peaks were obtained from ReMap 2018 (Griffon et al., 2015) and all peaks were intersected with LUHMES enhancers using with Bedtools v2.26.0. Intersection counts were consolidated for each motif or for each ChIP-seq experiment according to the specific-type of LUHMES enhancer set and R v3.3.1 was used to determine the hypergeometric distribution p-value comparing LUHMES enhancer subsets against all LUHMES enhancers and fisher p-value comparing differentiated- vs. undifferentiated LUHMES enhancers. In the text, TF enrichment is based on fisher p-values filtered for a minimal odds ratio of 2.

4.12. GWAS risk SNPs

Three different overlapping sets of PD-risk SNPs were used. The largest set (PD_all) consisted of combined risk SNPs obtained from pdgene.org, NHGRI-EBI, and NCBI; were combined with LD SNPs ($r^2 > 0.8$) in Europeans using HapMap phase 3 data in RaggR (Barrett et al., 2005) as described in (Pierce and Coetzee, 2017); and then were filtered using Bedtools with the NCBI dbSNP build 144 generating 23,918 SNPs. This set was used to define the LUHMES enhancer and CTCF peak enrichment of PD risk SNPs by intersecting SNPs with enhancers via Bedtools. This set was then reduced to the most significant 6869 SNPs by p-value for use in identifying specific LUHMES enhancers and genes, again using Bedtools. The next set of PD-risk SNPs (as well as all other disease-risk SNPs used in Fig. 5) were downloaded from the GWAS catalog 04/2017, enlarged with LD SNPs as above, and filtered with dbSNP144. This set was used for fair comparison between GWAS conditions. The last set of PD-risk SNPs used was based on the newest PD meta-analysis (Chang et al., 2017). Here the highest-significance reported risk SNPs were combined with high-LD SNPs to make the smallest of the three PD rSNP sets. For the enhancer–gene relationships described in the text, this third set consisted of all SNPs, whereas for the enrichment analysis in Fig. 5c, this set was first filtered by intersection with dbSNP144.

4.13. Risk SNP enrichment

Enrichment of SNPs in regulatory elements was based on a hypergeometric probability distribution comparing the ratio of all dbSNP144 SNPs that overlap a set of coordinates which define enhancers or CTCF peaks vs. the subset of rSNPs that overlap that same set of peaks. A random sample set of SNPs from the background set will produce a similar ratio of overlap, while nonrandom set, with respect to LUHMES enhancers, will be statistically significant. The heatmap in Fig. 5C displays the relative relationship of these two ratios directly (not the p-value) and has been normalized to z-scores for each disorder separately. Table S4 contains the raw count data, ORs, and full name for each cell type. For each of the PD risk SNP lists, rSNPs at the *MAPT* locus were removed from analysis due to the large number of LD SNPs. Roadmap enhancer regions are based on 12-mark 25-type segmentation and SNP overlap was counted for segment types, “E13”, “E14”, “E15”, “E16”, “E17”, “E18”. For Fig. 5B, SNPs were also queried for likely TF motif interruption using MotifbreakR (Coetzee et al., 2015).

4.14. Sequence data

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (Edgar et al., 2002) and are accessible through GEO Series accession number GSE109706.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.nbd.2018.02.007.

Acknowledgements

We acknowledge the Center for Neurodegenerative Science at the Van Andel Research Institute for financial support. We thank JC VanderSchans for technical assistance and thank Megan Bowman, Ben Johnson, and Zachary Madaj of the Van Andel Research Institute Bioinformatics and Biostatistics Core for technical assistance, helpful discussion, and editing. We thank Marie Adams of the Van Andel Research Institute Genomics Core for providing Next Generation Sequencing facilities and services. We acknowledge technical editing by David Nadziejka (VARI).

Conflict of interest statement

The authors declare no conflict of interest.

References

Aaker, J.D., et al., 2009. Feedback regulation of NEUROG2 activity by MTGR1 is required for progression of neurogenesis. Mol. Cell. Neurosci. 42, 267–277.
Barrett, J.C., et al., 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21, 263–265.
Bras, J., et al., 2015. SnapShot: genetics of Parkinson's disease. Cell. 160 e1, 570.
Cao, Q., et al., 2017. Reconstruction of enhancer-target networks in 935 samples of human primary cells, tissues and cell lines. Nat. Genet. 49, 1428–1436.
Chang, D., et al., 2017. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. Nat. Genet. 49, 1511–1516.
Coetzee, S.G., et al., 2015. motifbreakR: an R/Bioconductor package for predicting variant effects at transcription factor binding sites. Bioinformatics 31, 3847–3849.
Coetzee, S.G., et al., 2016. Enrichment of risk SNPs in regulatory regions implicate diverse tissues in Parkinson's disease etiology. Sci. Rep. 6, 30509.
Consortium, G.T., 2013. The genotype-tissue expression (GTEx) project. Nat. Genet. 45, 580–585.
Corradin, O., et al., 2016. Modeling disease risk through analysis of physical interactions

between genetic variants within chromatin regulatory circuitry. Nat. Genet. 48, 1313–1320.
DeFelipe, J., et al., 2002. Microstructure of the neocortex: comparative aspects. J. Neurocytol. 31, 299–316.
Defossez, P.A., Gilson, E., 2002. The vertebrate protein CTCF functions as an insulator in *Saccharomyces cerevisiae*. Nucleic Acids Res. 30, 5136–5141.
Dobin, A., Gingeras, T.R., 2015. Mapping RNA-seq reads with STAR. Curr. Protoc. Bioinformatics 51 (11), 14 (1–19).
Duren, Z., et al., 2017. Modeling gene regulation from paired expression and chromatin accessibility data. Proc. Natl. Acad. Sci. U. S. A. 114, E4914–E4923.
Edgar, R., et al., 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Res. 30, 207–210.
Fahn, S., 2003. Description of Parkinson's disease as a clinical syndrome. Ann. N. Y. Acad. Sci. 991, 1–14.
Franceschini, A., et al., 2013. STRING v9.1: protein-protein interaction networks, with increased coverage and integration. Nucleic Acids Res. 41, D808–15.
Ghosh, A., Tyson, T., George, S., Hildebrandt, E.N, Steiner, J.A, Madaj, Z., et al., 2016. Mitochondrial pyruvate carrier regulates autophagy, inflammation, and neurodegeneration in experimental models of Parkinson's disease. Sci. Transl. Med. 8 (368) 368ra174.
Griffon, A., et al., 2015. Integrative analysis of public ChIP-seq experiments reveals a complex multi-cell regulatory landscape. Nucleic Acids Res. 43, e27.
Hollerhage, M., et al., 2017. Protective efficacy of phosphodiesterase-1 inhibition against alpha-synuclein toxicity revealed by compound screening in LUHMES cells. Sci. Rep. 7, 11469.
Law, C.W., et al., 2014. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. 15, R29.
Lebouvier, T., et al., 2009. The second brain and Parkinson's disease. Eur. J. Neurosci. 30, 735–741.
Lee, H.J., et al., 2014. Extracellular alpha—synuclein-a novel and crucial factor in Lewy body diseases. Nat. Rev. Neurol. 10, 92–98.
Li, H., Durbin, R., 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25, 1754–1760.
Li, Q.H., et al., 2011. Measuring reproducibility of high-throughput experiments. Ann. Appl. Stat. 5, 1752–1779.
Lopes da Fonseca, T., et al., 2015. The interplay between alpha-synuclein clearance and spreading. Biomol. Ther. 5, 435–471.
Lotharius, J., et al., 2002. Effect of mutant alpha-synuclein on dopamine homeostasis in a new human mesencephalic cell line. J. Biol. Chem. 277, 38884–38894.
Lotharius, J., et al., 2005. Progressive degeneration of human mesencephalic neuron-derived cells triggered by dopamine-dependent oxidative stress is dependent on the mixed-lineage kinase pathway. J. Neurosci. 25, 6329–6342.
Lu, Y., et al., 2016. Defining the multivalent functions of CTCF from chromatin state and three-dimensional chromatin interactions. Nucleic Acids Res. 44, 6200–6212.
Mi, H., et al., 2013. Large-scale gene function analysis with the PANTHER classification system. Nat. Protoc. 8, 1551–1566.
Mumbach, M.R., et al., 2017. Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. Nat. Genet. 49, 1602–1612.
Nalls, M.A., et al., 2014. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. Nat. Genet. 46, 989–993.
Paiva, I., et al., 2017. Sodium butyrate rescues dopaminergic cells from alpha-synuclein-induced transcriptional deregulation and DNA damage. Hum. Mol. Genet. 26, 2231–2246.
Pakkenberg, B., et al., 1991. The absolute number of nerve cells in substantia nigra in normal subjects and in patients with Parkinson's disease estimated with an unbiased stereological method. J. Neurol. Neurosurg. Psychiatry 54, 30–33.
Parker, S.C., et al., 2013. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Proc. Natl. Acad. Sci. U. S. A. 110, 17921–17926.
Pierce, S., Coetzee, G.A., 2017. Parkinson's disease-associated genetic variation is linked to quantitative expression of inflammatory genes. PLoS One 12, e0175882.
Pissadaki, E.K., Bolam, J.P., 2013. The energy cost of action potential propagation in dopamine neurons: clues to susceptibility in Parkinson's disease. Front. Comput. Neurosci. 7, 13.
Poewe, W., et al., 2017. Parkinson disease. Nat. Rev. Dis. Primers 3, 17013.
Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842.
Ren, G., et al., 2017. CTCF-mediated enhancer-promoter interaction is a critical regulator of cell-to-cell variation of gene expression. Mol. Cell 67, 1049–1058 (e6).
Rhie, S.K., et al., 2014. Nucleosome positioning and histone modifications define relationships between regulatory elements and nearby gene expression in breast epithelial cells. BMC Genomics 15, 331.
Ritchie, M.E., et al., 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 43, e47.
Robinson, M.D., Oshlack, A., 2010. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 11, R25.
Rubin, A.J., et al., 2017. Lineage-specific dynamic and pre-established enhancer-promoter contacts cooperate in terminal differentiation. Nat. Genet. 49, 1522–1528.
Ryan, S.D., et al., 2013. Isogenic human iPSC Parkinson's model shows nitrosative stress-induced dysfunction in MEF2-PGC1alpha transcription. Cell 155, 1351–1364.
Scholz, D., et al., 2011. Rapid, complete and large-scale generation of post-mitotic neurons from the human LUHMES cell line. J. Neurochem. 119, 957–971.
Schule, B., et al., 2009. Can cellular models revolutionize drug discovery in Parkinson's disease? Biochim. Biophys. Acta 1792, 1043–1051.
Shen, Y., et al., 2012. A map of the cis-regulatory sequences in the mouse genome. Nature 488, 116–120.
Smirnova, L., et al., 2016. A LUHMES 3D dopaminergic neuronal model for neurotoxicity testing allowing long-term exposure and cellular resilience analysis. Arch. Toxicol. 90, 2725–2743.
Spillantini, M.G., et al., 1997. Alpha-synuclein in Lewy bodies. Nature 388, 839–840.
Tyson, T., et al., 2016. Sorting out release, uptake and processing of alpha-synuclein during prion-like spread of pathology. J. Neurochem. 139 (Suppl. 1), 275–289.
Verstraeten, A., et al., 2015. Progress in unraveling the genetic etiology of Parkinson disease in a genomic era. Trends Genet. 31, 140–149.
Yao, L., et al., 2015. Demystifying the secret mission of enhancers: linking distal regulatory elements to target genes. Crit. Rev. Biochem. Mol. Biol. 50, 550–573.
Zhang, Y., et al., 2008. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 9, R137.