

1 **Evaluating the mouse neural precursor line, SN4741, as a suitable proxy for midbrain**
2 **dopaminergic neurons**

3

4 Rachel J. Boyd^{1,3}, Sarah A. McClymont^{1,3}, Nelson B. Barrientos¹, Paul W. Hook¹, William D.
5 Law¹, Rebecca J. Rose¹, Eric L. Waite,¹ Dimitrios Avramopoulos¹, and Andrew S. McCallion*^{1,2}

6

7 1. McKusick-Nathans Department of Genetic Medicine, Johns Hopkins University School of
8 Medicine, Baltimore, MD 21205, USA.

9 2. Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, MD
10 21287, USA.

11 3. These authors contributed equally.

12 *To whom correspondence should be addressed

13

14 Electronic addresses:

15 Rachel J. Boyd - rboyd25@jhmi.edu, Sarah A. McClymont - sarahmcclymont@gmail.com,
16 Nelson B. Barrientos - nbarrie1@jhu.edu, Paul W. Hook - phook2@jhmi.edu, William D. Law -
17 williamdlaw@gmail.com, Rebecca J. Rose – rrose10@jhmi.edu, Eric L. Waite –
18 eric.waite@pennmedicine.upenn.edu, Dimitrios Avramopoulos - adimitr1@jhmi.edu, Andrew S.
19 McCallion - andy@jhmi.edu

20

21

22 KEYWORDS:

23 Parkinson disease, mouse-derived cell lines, immortalized cell lines, chromatin accessibility,
24 RNA-seq, ATAC-seq, scRNA-seq, genomic characterization, disease-relevant model systems

25

26 ABSTRACT:

27 To overcome the ethical and technical limitations of *in vivo* human disease models, the broader
28 scientific community frequently employs model organism-derived cell lines to investigate of
29 disease mechanisms, pathways, and therapeutic strategies. Despite the widespread use of certain
30 *in vitro* models, many still lack contemporary genomic analysis supporting their use as a proxy
31 for the affected human cells and tissues. Consequently, it is imperative to determine how
32 accurately and effectively any proposed biological surrogate may reflect the biological processes
33 it is assumed to model. One such cellular surrogate of human disease is the established mouse
34 neural precursor cell line, SN4741, which has been used to elucidate mechanisms of
35 neurotoxicity in Parkinson disease for over 25 years. Here, we are using a combination of classic
36 and contemporary genomic techniques – karyotyping, RT-qPCR, single cell RNA-seq, bulk
37 RNA-seq, and ATAC-seq – to characterize the transcriptional landscape, chromatin landscape,
38 and genomic architecture of this cell line, and evaluate its suitability as a proxy for midbrain
39 dopaminergic neurons in the study of Parkinson disease. We find that SN4741 cells possess an
40 unstable triploidy and consistently exhibits low expression of dopaminergic neuron markers
41 across assays, even when the cell line is shifted to the non-permissive temperature that drives
42 differentiation. The transcriptional signatures of SN4741 cells suggest that they are maintained in
43 an undifferentiated state at the permissive temperature and differentiate into immature neurons at
44 the non-permissive temperature; however, they may not be dopaminergic neuron precursors, as

45 previously suggested. Additionally, the chromatin landscapes of SN4741 cells, in both the
46 differentiated and undifferentiated states, are not concordant with the open chromatin profiles of
47 *ex vivo*, mouse E15.5 forebrain- or midbrain-derived dopaminergic neurons. Overall, our data
48 suggest that SN4741 cells may reflect early aspects of neuronal differentiation but are likely not
49 a suitable proxy for dopaminergic neurons as previously thought. The implications of this study
50 extend broadly, illuminating the need for robust biological and genomic rationale underpinning
51 the use of *in vitro* models of molecular processes.

52

53 BACKGROUND:

54 *in vitro* cellular surrogates present an excellent opportunity for elucidating the molecular
55 mechanisms behind human disease without the ethical and technical limitations of *in vivo*
56 systems. As such, most studies of human disease that employ genomic or cellular manipulations
57 or assays that require high cell quantity and quality, are often conducted *in vitro* to ensure
58 biological and statistical robustness[1–3]. For example, *in vitro* models are frequently employed
59 in studies of the role of genomic regulation in human disease, identification of candidate genes
60 and regulatory elements and evaluation of their functional characteristics through genetic
61 manipulations and high-throughput assays[4–7]. As genome-wide association studies (GWASs)
62 continue to indict human disease-associated variants, it is becoming evident that most of them lie
63 within non-coding regions of the genome[8]. Such regions frequently represent cis-regulatory
64 elements (CREs), required for the transcriptional modulation of cognate genes. The assays
65 required to evaluate their function[4,9,10], or connect CREs with the promoters they
66 modulate[11], often require large cell numbers, making, *in vitro* cellular systems the preferred
67 strategy.

68 Prioritizing non-coding GWAS variants and disease-relevant sequences for extensive
69 investigation requires knowledge of their chromatin accessibility status. Open chromatin is prone
70 to harbor functional sequences; and since chromatin accessibility profiles vary across cell types
71 and developmental time, it is important to prioritize disease-associated variants that lie within
72 open chromatin regions in the disease-relevant cell type(s) [8,12,13]. It is also critical to
73 functionally evaluate the biological consequences of disease-associated variation, test the
74 efficacy of potential therapeutics, and observe the effects of disease-relevant insults in the
75 appropriate cellular context[8,14,15]. Therefore, when studying disease associated variation, the
76 most effective *in vitro* cellular surrogates should ideally mimic the chromatin architecture and
77 transcriptional profiles of the *in vivo* cell types affected by disease.

78 In Parkinson disease (PD), midbrain (MB) dopaminergic (DA) neurons in the substantia
79 nigra (SN) are the primary affected cell type[16]. Preferential degeneration of these neurons
80 elicits a progressive neurodegenerative disorder characterized by motor deficits[16]. As the
81 second most common neurodegenerative disorder, affecting approximately 1% of adults over 70
82 years old[17,18], PD is the focus of extensive research efforts. As such, various cell lines have
83 been used as *in vitro* proxies of MB DA neurons to study the cellular impacts of PD-relevant
84 insults, as well as candidate PD-associated sequences, their functions, and their potential as
85 therapeutic targets[19].

86 One such cell line, SN4741, is reported to be a clonal DA neuronal progenitor line that
87 was established in 1999 from mouse embryonic day 13.5 (E13.5) SN tissue[20]. The SN was
88 dissected from transgenic mice containing 9.0 kb of the 5' promoter region of tyrosine
89 hydroxylase (*TH*), fused to the temperature-sensitive mutant Simian Virus 40 T antigen
90 (SV40Tag-tsA58) oncogene[20]. The goal of this *TH* promoter transgene was to enable selective

91 acquisition of DA neurons, while the purpose of the SV40Tag oncogene was to facilitate
92 conditional immortalization of the cell line. The temperature sensitive mutant form of this
93 immortalizing gene (tsA58) should permit uncontrolled differentiation and proliferation at the
94 permissive temperature (33°C), maintain cells in an undifferentiated state at 37°C, and since
95 tsA58 displays diminished activity at 39°C, it should direct differentiation that more closely
96 resembles primary cells when the culture is shifted to this non-permissive temperature[20].

97 As an established mouse neural precursor line, SN4741 cells have since been used to
98 elucidate mechanisms of neurotoxicity in PD[21–25], test the efficacy of therapeutic targets
99 against PD relevant insults[26,27], and assay the impacts of PD-associated genetic
100 mutation[28,29] and transcriptional regulation[30–32]. Important technological advances have
101 also arisen since the genesis and implementation of the SN4741 cell line, including chromatin
102 conformation capture technologies[11,33,34], RNA-sequencing (RNA-seq)[35], and assay for
103 transposase-accessible chromatin using sequencing (ATAC-seq)[36]. In this study, we exploit
104 these modern approaches to assess the suitability of SN4741 as an *in vitro* proxy for DA neurons
105 and determine the extent to which this cell line is appropriate for prioritizing and investigating
106 the mechanisms by which PD-associated variation confers disease risk.

107 Through a combination of karyotyping, single-cell (sc)RNA-seq, and RT-qPCR, we
108 evaluate the genomic integrity of this immortalized cell line, determine how the transcriptional
109 profile and expression of DA neuron marker genes in this line changes between undifferentiated
110 (37°C) and differentiated (39°C) states, and evaluate whether these transcriptional changes are
111 consistent throughout the differentiation process. The data we collect suggests that while these
112 cells show evidence that they are exiting a proliferative state and entering a more differentiated
113 state, they are an unsuitable model of SN DA neurons, as they possess aneuploidy and structural

114 abnormalities, as well as consistently low expression of DA neuron markers upon differentiation.
115 We employ bulk RNA-seq to quantify transcriptional differences between differentiated and
116 undifferentiated SN4741 cells and determine that, while transcriptional profiles change to reflect
117 differentiation, they do not show strong evidence that these cells are entering a DA state. We
118 then compare chromatin accessibility profiles of undifferentiated and differentiated SN4741 cells
119 with those of *ex vivo* mouse E15.5 midbrain (MB) and forebrain (FB) neurons and determine that
120 the chromatin accessibility profiles of SN4741 cells do not reflect the cellular population from
121 which they were derived. Collectively, cytogenetic, chromatin, and transcriptional data suggest
122 that the SN4741 cell line is not as strong a cellular surrogate for DA neurons as previously
123 thought. Ultimately, this work underscores the importance of leveraging technological advances
124 in genomic and cellular analyses to evaluate, and re-evaluate, the suitability of established model
125 systems in disease biology.

126

127 RESULTS:

128 **SN4741 is an Unstable Polyploid Cell Line:**

129 G-band karyotyping was performed on 20 SN4741 metaphase spreads and a
130 representative karyogram (**Figure 1A**) was generated. The karyotype was interpreted as an
131 abnormal, polyploid, karyotype with complex numerical abnormalities and unbalanced,
132 structural abnormalities. While most, but not all, abnormalities were consistently present in these
133 cells; none of the 20 cells assessed had the same chromosome complement, and no normal cells
134 were observed. All cells possessed at least one copy of each mouse autosome (1 through 19) and
135 female sex chromosomes; however, most chromosomes were triploid in each cell (**Figure 1B**).
136 These karyotypic abnormalities already call into question the viability of these cells as a

137 surrogate for human neurodegenerative disease. Since these cells are genetically unstable, there
138 may be large experimental batch effects as the cell populations shift across divisions.
139 Furthermore, gene dosage effects that severely deviate from normal copy number in DA neurons
140 may lead to confounding and unreliable results.

141

142 **Undifferentiated and Differentiated SN4741 Cells Express Similar Levels of Dopaminergic**
143 **Neuron Marker Genes by RT-qPCR:**

144 Preliminary expression analysis by RT-qPCR confirmed expression of a variety of DA
145 neuron markers: forkhead box A2 (*Foxa2*), nuclear receptor subfamily 4 group A, member 2
146 (*Nr4a2*), solute carrier family 6 member 3 (*Slc6a3*), and tyrosine hydroxylase (*Th*). Compared to
147 the expression of these markers in the undifferentiated SN4741 cell culture (37°C), relative
148 expression of all markers remained at similar levels when the cells were shifted to the higher
149 temperature condition (39°C), except for an increase in *Th* expression (**Figure 1C**). While
150 elevated *Th* has been used as a marker of differentiation into DA neurons in previous work with
151 SN4741 cells[20,37,38], an increase in *Th* expression is not exclusively associated with DA
152 neurons. *Th* is a marker for all catecholaminergic neurons (dopaminergic and adrenergic)[39],
153 and evidence suggests that *Th* expression is transient in other neurons throughout embryonic
154 development[40–42]. These results indicate that at the non-permissive temperature, SN4741 cells
155 may not be fully differentiating into DA neuron progenitors.

156

157 **scRNA-seq Reveals that SN4741 Cells Differentiate at the Non-Permissive Temperature,**
158 **but Lack Expression of DA Neuron Marker Genes:**

159 To assess the consistency of the differentiation protocol, transcriptomes were generated
160 from $\geq 17,000$ cells across four replicates cultured at the permissive temperature (37°C) and four
161 replicates cultured at the non-permissive temperature (39°C). Analysis of the single-cell
162 transcriptomes reveal that the cells cluster by growth temperature (**Figure 1D**). This separation
163 of cells by temperature is accompanied by changes to the cell cycle, with cells at the permissive
164 (37°C) temperature mostly in either G2M or S phase, while cells at the non-permissive
165 temperature (39°C) are mostly in G1 phase (**Figure 1E**), indicating that they may be
166 differentiated. In expression analysis, markers of proliferation that are expressed in G2M phase,
167 like Marker of Proliferation Ki-67 (*Mki67*), are predominantly expressed in cells at the
168 permissive temperature (**Figure 1F**), corroborating the cell cycle analysis. When shifted to the
169 non-permissive temperature, SN4741 cells appear to robustly differentiate, exemplified by a
170 decrease in the expression of Nestin (*Nes*), a neural stem cell marker (**Figure 1F**). Additional
171 transcriptional changes at this non-permissive temperature include an increase in the expression
172 of a neural marker CUGBP Elav-Like Family Member 5 (*Celf5*)[43], as well as genes that have
173 been found to regulate neural stem cell self-renewal (Inhibitor of DNA Binding 2, *Id2*; High
174 Mobility Group AT-Hook 2, *Hmga2*)[44,45], neurogenesis (Iroquois Homeobox 3, *Irx3*)[46],
175 and arborization of neurons (Sodium Voltage-Gated Channel Beta Subunit 1, *Scn1b*)[47],
176 indicating that these cells may be differentiating into neural precursor cells (**Figure 1G**).
177 Furthermore, Cadherin 13 (*Cdh13*), a modulator of GABAergic neurons, is significantly
178 upregulated at this non-permissive temperature (**Figure 1G**), while the expression of a variety of
179 DA neuron markers fail to be detected in either the permissive or non-permissive temperatures.
180 Markers, including Aldehyde Dehydrogenase 1 Family Member A1 (*Aldh1a1*), *Foxa2*, LIM
181 Homeobox Transcription Factor 1 Beta (*Lmx1b*), *Nr4a2*, Paired-like homeodomain 3 (*Pitx3*),

182 *Slc6a3*, and *Th* have few to no reads assigned to them (**Figure 1H**). Collectively, these results
183 suggest that while SN4741 cells are differentiating towards a neuronal fate when shifted to the
184 nonpermissive temperature, they may not be entering a clear DA trajectory under these
185 conditions.

186

187 **ATAC-seq Identifies Differential Open Chromatin Profiles in SN4741 Cells at the** 188 **Permissive and Non-Permissive Temperatures:**

189 To consider how chromatin accessibility changes between the two temperatures, we
190 performed ATAC-seq on SN4741 cells in both the undifferentiated and differentiated states.
191 Libraries were confirmed to be technically and biologically relevant (**Supplemental Figure 1**),
192 and well correlated between replicates (**Supplemental Figure 2; Figure 2A-B**).

193 A total of 83,778 consensus open chromatin regions were identified, with 70% of peaks
194 shared between the two temperatures (**Figure 2C**). Principal component analysis of these
195 consensus regions suggests a clear separation in the chromatin state between the two
196 temperatures (**Figure 2D**). To explore these differences, we performed differential accessibility
197 analysis with DiffBind[48], to find a total of 5,055 differentially accessible regions: 2,654
198 enriched in the permissive temperature and 2,401 enriched at the non-permissive temperature
199 ($\log_2FC > 1$, $FDR < 0.05$; **Figure 2E**).

200 Gene ontology of genes adjacent to differentially accessible regions largely recapitulate
201 the scRNA-seq analysis; functions associated with regions preferentially open at the permissive
202 temperature suggest the maintenance of the undifferentiated, cell-cycling state (**Figure 2F**). The
203 gene ontology of genes adjacent to those regions preferentially accessible at the non-permissive
204 temperature is less coherent and suggest cell differentiation towards several fates (blood vessels,

205 cartilage, tooth), none of which are neuronal and, perhaps unsurprisingly, demonstrate evidence
206 of response to temperature stress[49] (**Figure 2G**).

207 Overall, there is a shift in the chromatin accessibility between the two temperatures that
208 indicate the cells transition from an undifferentiated to differentiated state as the cells move from
209 the permissive to non-permissive temperature. The differences in chromatin accessibility further
210 confirm that SN4741 cells are not differentiating towards a neuronal lineage.

211

212 **Comparison of chromatin accessibility in SN4741 cells fails to recapitulate the chromatin** 213 **landscape of *ex vivo* mouse DA neurons:**

214 To evaluate the potential relationship between SN4741 cells and DA neurons they are
215 presumably modelling, we compared the chromatin accessibility between the SN4741 cells at
216 both temperatures to previously generated *ex vivo* mouse embryonic DA neuron chromatin
217 accessibility profiles (NCBI GEO: GSE122450;[50]).

218 Considering the consensus peak set of 165,334 regions generated from all *in vivo* and *ex*
219 *vivo* samples and their normalized read counts, we observe a clear separation between the
220 SN4741 cell culture model and the *ex vivo* DA neurons by correlation and principal component
221 analysis (**Figure 3A, B**). Examining the raw overlap of peaks between the SN4741 cells and *ex*
222 *vivo* neurons, just 12.5% (20,667) are present in all four cell types/conditions (**Figure 3C**). The
223 chromatin profiles are largely exclusive between the SN4741 cell culture model and the *ex vivo*
224 DA neurons: 41.3% (68,304) of regions are accessible solely in the *ex vivo* neuron populations
225 and 40% (65,857) are exclusively accessible in the SN4741 cell culture models. There is little
226 overlap between the *ex vivo* and cultured samples. In comparison of the *ex vivo* midbrain DA

227 neurons to the non-permissive, differentiated temperature, only 183 peaks are restricted to these
228 populations.

229 The chromatin profiles between *ex vivo* embryonic DA neurons and their prospective *in*
230 *vitro* cell culture surrogate are virtually independent. They exhibit scant overlap in their global
231 chromatin profiles and bear little resemblance to each other at regulatory regions of key DA
232 neuron genes (**Figure 3D**). Neither the analysis of the SN4741 chromatin accessibility profiles in
233 isolation or in comparison with *ex vivo* neurons would suggest these cells to be appropriate
234 models of embryonic DA neurons.

235

236 **Transcriptional Changes in SN4741 Cells Indicate Differentiation from Pluripotent Stem** 237 **Cells into Brain Cells that do not Fully Resemble MB DA Neurons:**

238 Bulk RNA-seq data were also generated for SN4741 cells, at both the permissive and
239 non-permissive temperatures, to determine whether transcriptome changes reflect differentiation
240 into DA neurons, or other neural cell types. To evaluate the RNA-seq libraries, quality-control
241 measures were performed *in silico* (**Supplemental Figure 3**). PCA (**Supplemental Figure 3B**),
242 and sample-sample distances (**Supplemental Figure 3C**) reaffirmed that samples cultured at the
243 same temperature are more like one-another than samples cultured at the alternate temperature.

244 We found that 735 genes were upregulated at the non-permissive temperature (adjusted
245 p-value < 0.01 and \log_2 FC > 1.5), and 954 genes were downregulated (adjusted p-value < 0.01
246 and \log_2 FC < -1.5) at the non-permissive temperature. The list of genes significantly
247 downregulated at the non-permissive temperature was submitted to Enrichr
248 (<https://maayanlab.cloud/Enrichr/>) [51–53] for gene ontology (GO) and analysis of cell type
249 markers. Consistent with the observation that cells at the non-permissive temperature are

250 differentiated and in G1 phase of the cell cycle, downregulated genes resulted in GO terms
251 strongly enriched for mitotic and DNA replication processes (**Figure 4A**). Additionally,
252 significantly downregulated genes at the non-permissive temperature overlap with subsets of
253 PanglaoDB[54] cell type marker genes, suggesting that these cells are shifting away from a state
254 that resembles neural stem cells (**Figure 4B**).

255 Similarly, the list of significantly upregulated genes was submitted to Enrichr for GO and
256 analysis of cell type markers. As expected, upregulated genes resulted in GO terms for biological
257 processes that indicate a more terminally differentiated cell type (**Figure 4C**): “synaptic vesicle
258 docking,” “negative regulation of osteoblast proliferation,” “lens fiber cell differentiation,”
259 “regulation of osteoblast proliferation,” and “forebrain regionalization”. While not included in
260 the top 10 terms by combined score ranking, “neuron remodeling,” “synaptic transmission,
261 glutamatergic,” “neuron maturation,” and “synaptic transmission, cholinergic” were also
262 identified as significantly associated terms. Notably, “synaptic transmission, dopaminergic” and
263 “dopaminergic neuron differentiation” were also listed as insignificant terms (**Figure 4C**), as *Th*
264 was the lone overlapping marker gene for these terms.

265 In line with GO terms enriched for biological processes involving differentiation,
266 possibly in neuronal cells, overlapping PanglaoDB[54] cell type marker genes suggest that
267 SN4741 cells at the non-permissive temperature most significantly resemble immature neurons
268 (**Figure 4D**). “Oligodendrocytes,” “retinal progenitor cells,” “satellite glial cells,” “dopaminergic
269 neurons,” “adrenergic neurons,” “GABAergic neurons,” and “glutamatergic neurons” were also
270 listed as cell types with significant marker gene overlap.

271 The distribution of various cell type marker genes on a volcano plot, indicating the
272 \log_2 FC in expression and $-\log_{10}$ adjusted p-values of DE genes, reveals that the specific genes

273 overlapping “pluripotent stem cell” markers (26/112), cluster as the most highly significantly
274 downregulated genes (**Figure 4E**). In contrast, only two of the upregulated marker genes
275 overlapping “immature neurons” (16/136, **Figure 4F**) and “oligodendrocytes” (17/178, **Figure**
276 **4G**) cluster in a similarly strong way. Plotting the 11/119 overlapping upregulated genes for
277 “dopaminergic neurons” (**Figure 4H**) reveals that 7/11 overlapping genes (*Celf5*, *Dpys15*,
278 *Cacna1b*, *Tmem179*, *Nova2*, *Nrx1*, and *Cntn2*) are also marker genes for immature neurons.
279 Plotting the DA neuron markers also assayed by RT-qPCR validates that the relative expression
280 of these markers is consistent between these highly sensitive assays. At 39°C, *Th* expression
281 increases ($\log_2FC = 1.552651$); *Nr4a2* expression decreases ($\log_2FC = -1.042794$), and
282 expression of *Slc6a3* and *Foxa2* is not significantly different between conditions: (*Slc6a3* was
283 filtered out due to low read counts across both temperature conditions and *Foxa2* $\log_2FC = -$
284 0.4142541).

285 To confirm the GO-indicted cell types, normalized read counts for select marker genes
286 were plotted for each temperature replicate: *Celf5*[43], *Nrxn1*[55], *Ntrk1*[56], and *Unc13a*[57],
287 for “immature neurons” were upregulated at 39°C relative to cells at 37°C (**Figure 5A**);
288 *Olig3*[58], *Il33*[59], *Hdac11*[60], and *Ptgds*[61] for “oligodendrocytes” were upregulated at
289 39°C relative to cells at 37°C (**Figure 5B**); and *Ccna2*[62], *Cdc6*[63], *Cenpf*[64], and *Gins1*[65]
290 for “pluripotent stem cells” were downregulated at 39°C relative to cells at 37°C (**Figure 5C**).

291 The differentially expressed gene sets were then analyzed using STRING ([https://string-](https://string-db.org/)
292 [db.org/](https://string-db.org/))[66]. The set of upregulated genes was enriched for protein-protein interactions (number
293 of edges = 705; expected number of edges = 438; PPI enrichment p-value = $< 1.0e-16$) and GO
294 terms such as “Neuron differentiation” (GO:0030182; FDR = 0.0016), “Neuron development”
295 (GO:0048666; FDR = 0.0070), and “Neurogenesis” (GO:0022008; FDR = 0.0085) supporting

296 that this upregulated gene set is a meaningful group likely belonging to a network involved in
297 neuronal maturation. The set of downregulated genes was also enriched for protein-protein
298 interactions (number of edges = 9251; expected number of edges = 1910; PPI enrichment p-
299 value = $< 1.0e-16$) and GO terms such as “Cell cycle” (GO:0007049; FDR = $5.90e-36$), “Mitotic
300 cell cycle” (GO:0000278; FDR = $6.97e-34$), and “Cell division” (GO:0051301; $6.35e-25$) further
301 confirming that these cells are no longer undergoing cell division.

302 Finally, previously generated reads per kilobase of exon per million reads mapped
303 (RPKM) from *ex vivo* E15.5 mouse embryonic DA neuron bulk RNA-seq[50] were used to
304 compare how closely the SN4741 transcriptome resembles the neuronal populations they are
305 expected to model. Similar to our results comparing chromatin accessibility between these two
306 datasets, correlation of RPKM shows a clear separation between the SN4741 cell culture model
307 and the *ex vivo* DA neurons (**Figure 5D**). Collectively, these results confirm that at the non-
308 permissive temperature, SN4741 cells are no longer rapidly dividing, neural stem cells.
309 However, while the transcriptional profile of these cells indicates that they are differentiating
310 towards cell types present in the brain, these cells do not fully possess characteristics of the MB
311 DA neurons they are meant to model.

312

313 DISCUSSION:

314 It is critically important that studies of human disease generate biologically accurate data,
315 whether aimed at elucidating molecular mechanisms, onset and progression, or management and
316 therapeutics. In the context of discovery biology or the illumination of human health and disease
317 mechanisms, misattribution of cellular identity, or other deviations from biological accuracy,
318 may result the misinterpretation of biological findings or misdirected research efforts. When

319 studying human disease, cellular surrogates are often used to overcome the ethical and technical
320 limitations of employing animal models. Therefore, it is imperative that disease-relevant insights
321 are predicated on robust data generated from model systems representing human biology as
322 accurately as possible.

323 Here, we demonstrate the importance of assessing *in vitro* models of disease to determine
324 the extent to which they can yield biologically accurate data that can be used to inform aspects of
325 human disease. The SN4741 cell line has been used to study neurotoxicity and therapeutic
326 interventions [21–27], PD-associated genetic mutation[28,29] and cell signaling and
327 transcriptional regulation[30–32,37], since it was initially characterized as an immortal, mouse
328 MB-derived cell line that differentiates into DA neurons at a non-permissive temperature[20].
329 However, contemporary genomic analyses have not been leveraged to characterize and evaluate
330 the SN4741 cell line as a suitable proxy for DA neurons in PD, until now.

331 We employed karyotyping, RT-qPCR, and scRNA-seq to assess the genomic stability of
332 these cells and determine how consistently they differentiate into DA neurons at the non-
333 permissive temperature. We generated bulk RNA-seq and ATAC-seq data from this cell line at
334 both the permissive and non-permissive temperatures, to extensively characterize this cell line
335 and document how transcriptional landscapes and chromatin accessibility profiles shift in
336 response to temperature-induced differentiation and compare to known profiles of *ex vivo* DA
337 neurons. Our results suggest that SN4741 is an unstable, polyploid cell line that is unlikely to be
338 a viable differentiation model of DA neurons; and thus, is likely not a robust proxy by which to
339 study MB DA neurons in the context of human phenotypes, including PD, schizophrenia,
340 addiction, memory, or movement disorders.

341 The results of karyotyping alone indicate that any data generated using SN4741 cells may
342 be biologically inaccurate due to extreme variability in chromosome complement and therefore,
343 copy number variation, between individual cells. Consequently, the results of previous studies
344 evaluating neurotoxicity[21,23,24,26,27,30,32,38,67], cellular signaling
345 pathways[22,25,28,68,69], and transcriptional profiling[37] in these cells may have been unduly
346 influenced by the extreme imbalance in gene dosage that we found to vary from cell to cell. For
347 example, alpha-synuclein (*SNCA*) has been consistently implicated in PD risk[70–73],
348 particularly due to variants that promote α -synuclein misfolding[74] and overexpression[50] or
349 events that result in gene amplification [75,76]. *Snca* is present on mouse chromosome 6 and the
350 karyotypes generated for SN4741 cells show that chromosome 6 is triploid in most assayed cells
351 (**Figure 1B**). Therefore, using the SN4741 cell line to model neurodegeneration in PD may result
352 in inaccurate data due to an exaggerated vulnerability towards degeneration imposed by elevated
353 *Snca* copy number, by gene dosage effects of other interacting gene products in relevant
354 pathways, or by the structural instability of this line.

355 Even if this cell line could be adopted to study *Snca* overexpression/amplification,
356 ATAC-seq profiling of open chromatin regions in this cell line at the permissive and non-
357 permissive temperatures indicates that these cells do not possess chromatin accessibility profiles
358 similar to those of *ex vivo*, mouse E15.5 MB neurons. In PD, disease is characterized by the
359 degeneration of MB DA neurons, while DA neurons of the FB are spared. Therefore, the
360 chromatin profiles of MB DA neurons, as well as the differentially accessible regions of the
361 genome between MB and FB neurons, may influence the preferential vulnerability of MB
362 neurons in PD[50]. In the context of exploiting these chromatin profiles to study PD-associated

363 variability and neurotoxicity, SN4741 cells are likely a poor model, as the open chromatin
364 regions of these cells are not a reliable proxy for mouse E15.5 MB or FB DA neurons.

365 The chromatin accessibility profiles of SN4741 cells not only fail to cluster with *ex vivo*
366 populations of mouse MB neurons, but the transcriptional landscapes of these cells suggest that
367 these cells have shifted towards a more differentiated state that may be less DA than previously
368 thought. Examination of cell cycle markers by scRNA-seq demonstrates that SN4741 cells at the
369 non-permissive temperature are more differentiated than cells at the permissive temperature, as
370 expected[20]. GO terms for genes that are significantly downregulated at the non-permissive
371 temperature reinforces that these cells are no longer rapidly dividing, pluripotent stem cells.
372 However, RT-qPCR, scRNA-seq, and bulk RNA-seq in these cells fail to detect significant
373 upregulation of most key DA neuron markers in the differentiated cells, except for *Th*. *Th* is not
374 exclusively expressed by DA neurons at embryonic timepoints[40–42]. In fact, significantly
375 upregulated genes in SN4741 cells at the non-permissive temperature that overlap with GO terms
376 and cell cycle marker genes suggests that *Th* is the only significantly upregulated gene
377 overlapping with biological processes involving DA neurons. Rather, additional overlapping cell
378 type marker genes suggest that these cells more closely resemble immature neurons.

379 In parallel, we generated promoter capture (pc)Hi-C data at the non-permissive
380 temperature with the intention of exploring how non-coding disease-relevant variants interact
381 with promoters and potentially regulate gene expression in MD DA neurons. As our group is
382 focused on PD-associated variation, which is unlikely to act broadly in immature neurons, our
383 group has not analyzed the resulting data, beyond basic quality control (**Supplemental Figure**
384 **4**). While the SN4741 cells at the non-permissive temperature fail to recapitulate the
385 transcriptomic or chromatin state of DA neurons, it is of potential interest for follow-up studies

386 that they do resemble some immature neuron types. Although not analyzed by our group, we
387 generated output files for interaction detection, and this data has been made publicly available
388 for others to explore (accessible through: https://github.com/rachelboyd/SN4741_pcHiC), as it
389 may be useful to study genomic interactions at promoters driving an immature neuronal state.
390 However, the cell type best represented by SN4741 cells at the non-permissive temperature still
391 requires deeper characterization.

392 Any data generated using SN4741 cells in the context of DA neuron modeling and/or PD
393 must be interpreted with caution and in light of the appropriate caveats. Due to the instability and
394 polyploidy of this cell line, we recommend that the use of SN4741 cells for PD- related research
395 be re-evaluated. Future studies designed to fine-tune the classification of these cells may support
396 the use of SN4741 cells as a model of other neuronal or non-neuronal cells. Additionally, the
397 differentiation trajectory of these cells may be amenable to intervention(s) that could drive their
398 molecular state towards one that resembles DA neurons more closely.

399

400 CONCLUSIONS:

401 This study establishes a valuable precedent with broad implications across biological and
402 disease-related research. Prior to using SN4741 cells to study non-coding regulatory variation in
403 PD, we characterized this cell line to determine its suitability as a model of DA neurons in PD,
404 and found that these cells are unstable, polyploid cells that do not demonstrate strong molecular
405 characteristics of MB DA neurons. These cells express low levels of DA neuron markers, and
406 chromatin landscapes in differentiated SN4741 cells scarcely overlap open chromatin regions in
407 *ex vivo* mouse E15.5 midbrain neurons. We demonstrate the importance of genomic
408 characterization of *in vitro* model systems prior to generating data and valuable resources that

409 may be used to inform aspects of human disease. In future studies that utilize *in vitro* models of
410 any human disease, due diligence to confirm their suitability as surrogates could save time,
411 resources, and possibly lives, by avoiding misdirection and advancing successful therapeutic
412 development.

413

414 METHODS:

415 **Cell Culture**

416 SN4741 cells were obtained from the Ernest Arenas group at the Karolinska Institutet.
417 SN4741 cells were confirmed to be mycoplasma free using a MycoAlert® Mycoplasma
418 Detection Assay (Lonza) and were maintained in high glucose Dulbecco's Modified Eagle
419 Medium (DMEM; Gibco 1196502), supplemented with 1% penicillin–streptomycin and 10%
420 fetal bovine serum (FBS) in a humidified 5% CO₂ incubator at 37°C. Cells at 80% confluence
421 were passaged by trypsinization approximately every 2-3 days. To induce differentiation, 24
422 hours after the cells were passaged, media was replaced by DMEM supplemented with 1%
423 penicillin–streptomycin and 0.5% FBS at 39°C. Cells were allowed to grow and differentiate in
424 these conditions for 48 hours before harvesting for experimentation.

425

426 **G-Band Karyotyping**

427 At passage 21, undifferentiated SN4741 cells were sent to the WiCell Research Institute
428 (Madison, Wisconsin), at 40-60% confluency, for chromosomal G-band analyses. Karyotyping
429 was conducted on 20 metaphase spreads, at a band resolution of >230, according to the
430 International System for Human Cytogenetic Nomenclature.

431

432 **cDNA Synthesis and RT-qPCR for DA neuron markers**

433 RNA was extracted from both differentiated and undifferentiated cells by following the
434 RNeasy Mini Kit (QIAGEN) protocol, as written. 1 µg of each RNA sample underwent first-
435 strand cDNA synthesis using the SuperScript III First-Strand Synthesis System for RT-PCR
436 (Invitrogen) according to the oligo(dT) method. qPCR was performed with Power SYBR Green
437 Master Mix (Applied Biosystems), using primers for β -actin (*Actb*), *Foxa2*, *Nr4a2*, *Slc6a3*, and
438 *Th* (**Table S1**). Reactions were run in triplicate under default SYBR Green Standard cycle
439 specifications on the Vii7 Real-Time PCR System (Applied Biosystems). Normalized relative
440 quantification and error propagation followed the data analysis and associated calculations
441 proposed by Taylor *et al.* (2019)[77], with results normalized to *Actb*.

442

443 **Single cell RNA-seq Library Preparation, Sequencing, and Alignment**

444 Both differentiated (39°C) and undifferentiated (37°C) cells were trypsinized, and
445 scRNA-seq libraries were generated following the Chromium 10X pipeline[78]. Four replicates
446 at each temperature across >17,000 cells were assayed. Cell capture, cDNA generation, and
447 library preparation were performed with the standard protocol for the Chromium Single Cell 3'
448 V3 reagent kit. Libraries were quantified with the Qubit dsDNA High Sensitivity Assay
449 (Invitrogen) in combination with the High Sensitivity DNA Assay (Agilent) on the Agilent 2100
450 Bioanalyzer. Single-cell RNA-sequencing libraries were pooled and sequenced on an Illumina
451 NovaSeq 6000 (SP flow cells), using 2x50 bp reads per library, to a combined depth of 1.6
452 billion reads. The quality of sequencing was evaluated via FastQC. Paired-end reads were
453 aligned to the mouse reference genome (mm10) using the CellRanger v3.0.1 pipeline. Unique

454 molecular identifier (UMI) counts were quantified per gene per cell (“cellranger count”) and
455 aggregated (“cellranger aggr”) across samples with no normalization.

456 **Single Cell RNA-seq Analysis**

457 Using Seurat[79](v4.2.0), cells were filtered to remove stressed/dying cells (% of reads
458 mapping to the mitochondria > 15%) and empty droplets or doublets (number of unique genes
459 detected <200 or >6,000). Cells were scored for their stage in the cell cycle using
460 “CellCycleScoring()” on cell cycle genes provided by Seurat (“cc.genes”). Cells were then
461 normalized using “SCTransform” (vst.flavor = “v2”) and corrected for percent mitochondrial
462 reads and sequence depth. Principal component (PC) analysis was performed and a PC cut-off
463 was identified using “ElbowPlot()”. Using this PC cutoff and a minimum distance of 0.001,
464 UMAP clustering was used for dimensionality reduction. Expression was plotted on a log scale
465 with “VlnPlot()” for a variety of proliferation and DA neuron markers.

466

467 **ATAC-seq Library Preparation and Quantification**

468 ATAC-seq libraries were generated for four replicates of undifferentiated (37°C) and
469 differentiated (39°C) SN4741 cells, according to the Omni-ATAC protocol[80], with minor
470 modifications. Aliquots of 50,000 cells were centrifuged at 2000 x g for 20 mins at 4°C, and the
471 resulting pellets were resuspended in 50µL of resuspension buffer. Cells were left to lyse for 3
472 minutes on ice before being centrifuged again at 2000 x g for 20 minutes at 4°C. The resulting
473 nuclei pellets were then tagmented, as written, using 50µL of transposition mixture and then
474 incubated at 37°C for 30 mins in a 1000 RPM thermomixer. After transposition, DNA was
475 purified with the Zymo DNA Clean and Concentrator -5 Kit and eluted in 21µL of elution buffer.

476 Pre-amplification of the transposed fragments was performed according to the conditions
477 outlined in the Omni-ATAC protocol[80]; however, 12 pre-amplification cycles were run in lieu
478 of qPCR amplification to determine additional cycles. The amplified libraries were prepared
479 according to the Nextera DNA Library Prep Protocol Guide, except that libraries were purified
480 with 40.5 μ L AMPure XP beads (Beckman Coulter), and 27.5 μ L of resuspension buffer was
481 added to each sample. All libraries were quantified with the Qubit dsDNA High Sensitivity
482 Assay (Invitrogen) in combination with the High Sensitivity DNA Assay (Agilent) on the
483 Agilent 2100 Bioanalyzer.

484 **ATAC-seq Sequencing, Alignment, and Peak Calling**

485 Libraries were sequenced on Illumina NovaSeq 6000 (SP flow cells), using 2x50 bp reads
486 per library, to a total combined depth of 1.6 billion reads. The quality of sequencing was
487 evaluated with FastQC (v0.11.9)[81] and summarized with MultiQC (v1.13)[82]. Reads were
488 aligned to the mouse reference genome (mm10) in local mode with Bowtie2[83](v2.4.1), using –
489 X 1000 to specify an increased pair distance to 1000bp. Samtools (v1.15.1)[84] and Picard
490 (v2.26.11; <http://broadinstitute.github.io/picard/>) were used to sort, deduplicate and index reads.
491 Peaks were called with MACS3 (v3.0.0a7; <https://github.com/macs3-project/MACS>)[85] and
492 specifying --nomodel and --nolambda for the `callpeaks()` command. Peaks overlapping mm10
493 blacklisted/block listed regions called by ENCODE[86,87] and in the original ATAC-seq
494 paper[36] were also removed with BEDTools (v2.30.0)[88].

495 For visualization with IGV, IGVTools (v2.15.2) was used to convert read pileups to
496 TDFs. The fraction of reads in peaks was calculated with DeepTools (v3.5.1)[89] using the
497 plotEnrichment command. The average mapping distance flag was extracted from the SAM files
498 with a custom script available at our GitHub repo

499 (https://github.com/sarahmcclymont/SN4741_ATAC/) to generate the fragment length plot.
500 Mouse (mm10) transcriptional start site (TSS) coordinates were downloaded from the UCSC
501 Genome Browser[90] (Mouse genome; mm10 assembly; Genes and Gene Predictions; RefSeq
502 Genes track using the table refGene), and DeepTools (v3.5.1)[89] was used to plot the pileup of
503 reads overtop of these TSSs. Conservation under peaks (phastCons)[91] and the genomic
504 distribution of peaks were calculated using the Cis-regulatory Element Annotation System
505 (CEAS)[92] and conservation tool of the Cistrome[93] pipeline. Analysis can be found at
506 <http://cistrome.org/ap/u/smcclymont/h/sn4741-atac-seq-ceas-and-conservation>.

507 **ATAC-seq Normalization and Differential Peak Analysis**

508 Each sample's peak file and BAMs were read into and analysed with DiffBind
509 (v3.8.1)[48]. Peaks present in two or more libraries were considered in the consensus peakset.
510 Reads overlapping these consensus peaks were counted with ``dba.count()`` specifying `summits =`
511 `100`, `bRemoveDuplicates = TRUE`. These read counts were normalized with ``dba.normalize()`` on
512 the full library size using the RLE normalization method as it is native to the DESeq2 analysis
513 we employed in the following ``dba.analyze()`` step. The volcano plot was generated using a
514 custom script using the output of the ``dba.report()`` command, where `th = 1` and `fold = 0` and
515 `bCounts = T`, to output all peaks regardless of their foldchange or significance. Significantly
516 differentially accessible regions (filtered for `abs(Fold)>1` & `FDR < 0.05`) were submitted to
517 GREAT (v4.0.4)[94] and the gene ontology of the nearest gene, as identified with the
518 basal+extension method (where proximal was considered to be ± 5 kb) was assessed and plotted.

519 **ATAC-seq comparison to *ex vivo* MB and FB DA neurons**

520 Previously generated ATAC-seq libraries from *ex vivo* E15.5 mouse embryonic DA
521 neurons[50] were re-analyzed in parallel, following all the above alignment and filtering steps.

522 DiffBind (v3.8.1) was used to compare the samples, as above, and the R package UpSetR
523 (v1.4.0)[95] was used to plot the overlap of peaks between conditions.

524

525 **Bulk RNA-seq Library Preparation, Sequencing, and Alignment**

526 Cells were run through QIAshredder (Qiagen) and total RNA was extracted using the
527 RNeasy Mini Kit (Qiagen) according to the manufacturer's recommendations, except that RNA
528 was eluted twice in 50 μ L of water. Total RNA integrity was determined with the RNA Pico Kit
529 (Agilent) on the Agilent 2100 Bioanalyzer. RNA samples were sent to the Johns Hopkins
530 University Genetic Resources Core Facility (GRCF) for library prep (NEBNext Ultra II
531 directional library prep kit with poly-A selection) and sequencing. The libraries were pooled and
532 sequenced on an Illumina NovaSeq 6000 (SP flow cells), using 2x50 bp reads per library, to a
533 combined depth of 1.6 billion reads. The quality of sequencing was evaluated via FastQC.
534 FASTQ files were aligned to the mouse reference genome (mm10) with HISAT2[96] (v2.0.5)
535 and sample reads from different lanes were merged using samtools[84](v.1.10) function
536 "merge.". Aligned reads from individual samples were quantified against the mm10 reference
537 transcriptome with the subread[97–99](v1.6.1) function "featureCounts" [100], using -t exon and
538 -g gene_id, (**Supplemental Figure 3A**).

539 **Bulk RNA-seq Analysis**

540 The DESeq2 (v3.15) package was used for data quality assessment and analyses. A
541 DESeqDataSet of count data was generated using "DESeqDataSetFromMatrix" (design = ~
542 temp). The data underwent variance stabilizing transformation (vst) prior to using "plotPCA" to
543 visualize experimental covariates/batch effects (**Supplemental Figure 3B**) and R package

544 “pheatmap” (v1.0.12; <https://CRAN.R-project.org/package=pheatmap>) to visualize the sample-
545 to-sample distances (**Supplemental Figure 3C**).

546 Genes with an average of at least 1 read for each sample were analyzed to identify
547 differentially expressed (DE) genes between temperature conditions, using the function
548 “DESeq.” P-value distribution after differential expression (DE) analysis (**Supplemental Figure**
549 **3D**) verified that the majority of called DE genes are significant. Results ($\alpha = 0.01$) were
550 generated and subjected to log fold change shrinkage using the function “lfcShrink” (type =
551 “apeglm”)[101] for subsequent visualization and ranking. The function “plotMA” was used to
552 generate MA plots, both before and after LFC shrinkage, to visualize the \log_2 fold changes
553 attributable to the non-permissive temperature shift over the mean of normalized counts for all
554 the samples in the DESeqDataSet (**Supplemental Figure 3E-F**). MA plots demonstrated that log
555 fold change shrinkage of the data successfully diminished the effect size of lowly expressed
556 transcripts with relatively high levels of variability.

557 Volcano plots were generated using a custom function to visualize \log_2 fold changes of
558 specific genes in the dataset. A gene was considered significantly differentially expressed if it
559 demonstrated an adjusted p-value < 0.01 and $|\log_2 \text{FC}| > 1.5$. These significantly differentially
560 expressed genes were submitted to Enrichr[51–53] for analyses within the “ontologies” and “cell
561 types” categories. The upregulated and downregulated gene sets were passed to STRING [66] for
562 analysis of protein-protein interactions and network relationships.

563 **Bulk RNA-seq comparison to *ex vivo* MB and FB DA neurons**

564 Read counts from the SN4741 bulk RNA-seq dataset were converted to RPKM and
565 compared to bulk RNA-seq data from previously generated *ex vivo* mouse embryonic DA

566 neurons (NCBI GEO: GSE122450;[50]). A Pearson correlation heatmap was generated using
567 ggplot2[102].

568

569 **Promoter capture HiC library generation**

570 PcHiC was performed as previously described[103], with minor modifications. Briefly,
571 SN4741 cells were cultured at the non-permissive temperature and plated at five million cells per
572 10cm dish. The cells were crosslinked using 1% formaldehyde, snap frozen using liquid
573 nitrogen, and stored at -80°C. The cells were dounce homogenized and restriction enzyme
574 digestion, using 400 units *HindIII-HF* overnight at 37°C. The total volume was maintained at
575 500µL, through addition of 1X NEBuffer 2.1. Heat inactivation was performed at 80°C for 20
576 minutes, and biotinylated-dCTP was used for biotin fill-in reaction. Blunt-end ligation was
577 performed using Thermo T4 DNA ligase, with cohesive end units maintained at 15,000 and
578 buffer and water volumes adjusted to ensure a total volume of 665µL was added to each Hi-C
579 tube. Cross-linking was performed overnight, with additional (50µL) proteinase K added for two
580 hours the following day. DNA purification was split across two reactions using 2mL PhaseLock
581 tubes, and volumes were adjusted accordingly. Each PhaseLock reaction was split again into two
582 vials for ethanol purification, and centrifugation at step 6.3.8 was performed at room
583 temperature. The pellets were dissolved in 450µL 1X TLE and transferred to a 0.5mL 30kD
584 Amicon Column. After washing, the column was inverted into a new container, and no
585 additional liquid was added to raise the volume to 100µL. All four reactions were combined, the
586 total volume determined, and RNaseA (1mg/mL) equal to 1% of the total volume was added for
587 30 minutes at 37°C.

588 The libraries were assessed for successful blunt-end ligation by a ClaI restriction enzyme
589 digest of PCR products, as previously described[103]. Biotin was removed from un-ligated ends
590 and DNA was sheared to a size of 200-300bp using the Covaris M220 (High setting, 35 cycles of
591 30s “on” and 90s “off”; vortexing/spinning down samples and changing sonicator water every 5
592 cycles). Size selection was performed using AMPure XP magnetic beads, as previously
593 described[103] except that all resuspension steps were increased by 5 μ L, so that 5 μ L could be
594 used for QC with the High Sensitivity DNA Assay (Agilent) on the Agilent 2100 Bioanalyzer (at
595 three stages: post-sonication, post-0.8x size-selection, and post-1.1x size-selection). The
596 remaining protocol was performed as described. Capture probes (Arbor Biosciences;
597 https://github.com/nbbarrientos/SN4741_pcHiC) were designed against mouse (mm10) RefSeq
598 transcription start sites, filtering out “XM” and “XR” annotated genes. The remaining promoters
599 were intersected with the *in silico* digested *HindIII* mouse genome, to retain all *HindIII*
600 fragments containing a promoter. Potential probes sites were assessed \pm 330bp of the *HindIII* cut
601 site on either end of the fragment and finalized probe sets were filtered using no repeats and
602 “strict” criteria, as defined by Arbor Biosciences. After generating a uniquely indexed HiC
603 library with complete Illumina adapters, probes targeting promoter containing fragments were
604 hybridized following Arbor Biosciences capture protocol (v4) at 65°C, 1 μ g DNA, and one round
605 of capture. The library was PCR amplified before sequencing on an Illumina NovaSeq 6000 (SP
606 flow cells), using 2x50 bp reads per library, to a combined depth of 1.6 billion reads.

607 **Promoter capture HiC data analysis**

608 Raw pcHiC reads for each replicate (n=4) were evaluated for quality via FastQC. FASTQ
609 files were mapped to mm10 using Bowtie2[83] (v.2.4.1) and filtered using HiCUP[104](v. 0.8).
610 The HiCUP pipeline was configured with the following parameters: FASTQ format (Format:

611 Sanger), maximum di-tag length (Longest: 700), minimum di-tag length (shortest: 50), and
612 filtering and alignment statistics were reported (**Supplemental Figure 4A-C**). BAM files were
613 generated for each replicate using samtools[84](v.1.10). DeepTools[89](v.3.5.1) before read
614 coverage similarities and replicate correlation was assessed using the function
615 “multiBamSummary” (in *bins* mode) to analyze the entire genome. A Pearson correlation
616 heatmap was generated using the function “plotCorrelation” (**Supplemental Figure 4D**). As a
617 result of high Pearson correlation coefficient among replicates ($r > 0.93$), library replicates were
618 combined. The CHiCAGO[105](v. 1.18.0) pipeline was used to convert the merged BAM file
619 into CHiCAGO format. The digested mm10 reference genome was used to generate a restriction
620 map file, a baited restriction map file, and the rest of required input files (.npb, .nbpb, and .poe)
621 required to run the CHiCAGO pipeline.

622

623 AVAILABILITY OF DATA AND MATERIALS:

624 All data and analysis pipelines are available at <https://github.com/rachelboyd>. ATAC-
625 sequencing, RNA-sequencing, and promoter-capture Hi-C data will be available at the Gene
626 Expression Omnibus (GEO) under the accession number GSEXXXXXX upon manuscript
627 acceptance.

628

629 ABBREVIATIONS:

<i>ActB:</i>	β -actin
<i>Aldh1a1:</i>	Aldehyde Dehydrogenase 1 Family Member A1
<i>ATAC-seq:</i>	Assay for Transposase-Accessible Chromatin using Sequencing
<i>BAM:</i>	Binary Alignment and Map
<i>Cacna1b:</i>	Calcium channel, voltage-dependent, N type, alpha 1B subunit
<i>Ccna2:</i>	Cyclin A2
<i>Cdc6:</i>	Cell division cycle 6

<i>Cdh13:</i>	Cadherin 13
<i>CEAS:</i>	Cis-Regulatory Element Annotation System
<i>Celf5:</i>	CUGBP Elav-Like Family Member 5
<i>Cenpf:</i>	Centromere protein F
<i>Cntn2:</i>	Contactin 2
<i>CO2:</i>	Carbon Dioxide
<i>CRE:</i>	Cis Regulatory Element
<i>DA:</i>	Dopaminergic
<i>DMEM:</i>	Dulbecco's Modified Eagle Medium
<i>Dpysl5:</i>	Dihydropyrimidinase-like 5
<i>E13.5/15.5:</i>	Embryonic Day 13.5/15.5
<i>ENCODE:</i>	Encyclopedia of DNA Elements
<i>FB:</i>	Forebrain
<i>FBS:</i>	Fetal Bovine Serum
<i>FDR:</i>	False Discovery Rate
<i>Foxa2:</i>	Forkhead Box A2
<i>Gins1:</i>	GINS complex subunit 1 (Psf1 homolog)
<i>GO:</i>	Gene Ontology
<i>GRCF:</i>	Genetics Core Research Facility
<i>GWAS:</i>	Genome-Wide Association study
<i>Hdac11:</i>	Histone deacetylase 11
<i>Hmga2:</i>	High Mobility Group AT-Hook 2
<i>Id2:</i>	Inhibitor Of DNA Binding 2
<i>IGV:</i>	Integrative Genomics Viewer
<i>Il33:</i>	Interleukin 33
<i>Irx3:</i>	Iroquois Homeobox 3
<i>KEGG:</i>	Kyoto Encyclopedia of Genes and Genomes
<i>LFC:</i>	Log Fold-Change
<i>Lmx1b:</i>	LIM Homeobox Transcription Factor 1 Beta
<i>MB:</i>	Midbrain
<i>Mki67:</i>	Marker of Proliferation Ki-67
<i>Nes:</i>	Nestin
<i>Nova2:</i>	NOVA alternative splicing regulator 2
<i>Nr4a2:</i>	Nuclear Receptor Subfamily 4 Group A, Member 2
<i>Nrx1:</i>	Neurexin 1
<i>Ntrk1:</i>	Neurotrophic receptor tyrosine kinase 1
<i>OCR:</i>	Open Chromatin Region
<i>Olig3:</i>	Oligodendrocyte transcription factor 3
<i>PC(A):</i>	Principal Component (Analysis)
<i>pcHi-C:</i>	Promoter-Capture Hi-C
<i>PD:</i>	Parkinson Disease
<i>Pitx3:</i>	Paired-like homeodomain 3
<i>Ptgds :</i>	Prostaglandin D2 synthase
<i>(q)PCR:</i>	(Quantitative) Polymerase Chain Reaction

<i>QC:</i>	Quality Control
<i>RNA:</i>	Ribonucleic Acid
<i>RPKM:</i>	Reads per kilobase of exon per million reads mapped
<i>RT:</i>	Reverse Transcriptase
<i>Scn1b:</i>	Sodium Voltage-Gated Channel Beta Subunit 1
<i>scRNA-seq:</i>	Single Cell RNA sequencing
<i>Slc6a3:</i>	Solute Carrier Family 6 Member 3
<i>SN:</i>	Substantia Nigra
<i>SNCA/Snca:</i>	Alpha-synuclein
<i>SV40Tag:</i>	Simian Virus 40 T antigen
<i>TH/Th:</i>	Tyrosine Hydroxylase
<i>Tmem179:</i>	Transmembrane protein 179
<i>ts:</i>	Temperature-Sensitive
<i>TSS:</i>	Transcriptional Start Site
<i>Unc13a:</i>	Unc-13 homolog A
<i>vst:</i>	Variance Stabilizing Transformation

630

631 **ACKNOWLEDGEMENTS:**

632 The authors would like to acknowledge Ernest Arenas (Karolinska Institutet), for providing
633 SN4741 cells, as well as the Johns Hopkins Genomics Core Research Facility (GCRF) and the
634 WiCell Research Institute, for providing technical services.

635

636 **FUNDING:**

637 This research, undertaken at Johns Hopkins University School of Medicine, was supported in
638 part by awards from the National Institutes of Health (NS62972 and MH106522) to A.S.M., by
639 T32 GM007814-40 to R.J.B. and N.B.B., and by the Canadian Institutes of Health Research
640 (DFD-181599) to R.J.B.

641

642 **AUTHOR INFORMATION:**

643 **Authors and Affiliations:**

644 *McKusick-Nathans Department of Genetic Medicine, Johns Hopkins University School of*
645 *Medicine, Baltimore, Maryland, USA.*

646 Rachel J. Boyd, Sarah A. McClymont, Nelson B. Barrientos, Paul W. Hook, William D. Law,
647 Rebecca J. Rose, Eric L. Waite, Dimitrios Avramopoulos, & Andrew S. McCallion
648 *Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland,*
649 *USA.*

650 Andrew S. McCallion

651 **Contributions:**

652 New data was generated by S.A.M., P.W.H., W.D.L., and E.W.L.; analyzed by R.J.B., S.A.M.,
653 P.W.H., N.B.B., W.D.L., and A.S.M. The manuscript was written by R.J.B., S.A.M., N.B.B.,
654 W.D.L., and A.S.M. Figures were created by R.J.B., S.A.M., and N.B.B. All authors reviewed
655 and approved the manuscript.

656

657 ETHICS DECLARATIONS:

658 **Ethics approval and consent to participate:**

659 Not applicable.

660 **Consent for publication:**

661 Not applicable.

662 **Competing interests:**

663 The authors declare no competing interests.

664

665 REFERENCES:

666 1. Ormond KE, Mortlock DP, Scholes DT, Bombard Y, Brody LC, Faucett WA, et al.
667 Human germline genome editing. *Am J Hum Genet.* 2017;101(2):167–76.

- 668 2. Barbosa DJ, Capela JP, de Lourdes Bastos M, Carvalho F. In vitro models for
669 neurotoxicology research. *Toxicol Res.* 2015;4(4):801–42.
- 670 3. Hirsch C, Schildknecht S. In vitro research reproducibility: Keeping up high standards.
671 *Front Pharmacol.* 2019;10.
- 672 4. Fisher S, Grice EA, Vinton RM, Bessling SL, Urasaki A, Kawakami K, et al. Evaluating
673 the biological relevance of putative enhancers using Tol2 transposon-mediated
674 transgenesis in zebrafish. *Nat Protoc.* 2006;1(3):1297–305.
- 675 5. Gorkin DU, Lee D, Reed X, Fletez-Brant C, Bessling SL, Loftus SK, et al. Integration of
676 ChIP-seq and machine learning reveals enhancers and a predictive regulatory sequence
677 vocabulary in melanocytes. *Genome Res.* 2012;22(11):2290–301.
- 678 6. Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: From properties to genome-
679 wide predictions. *Nat Rev Genet.* 2014;15(4):272–86.
- 680 7. Gasperini M, Findlay GM, McKenna A, Milbank JH, Lee C, Zhang MD, et al.
681 CRISPR/Cas9-mediated scanning for regulatory elements required for HPRT1 expression
682 via thousands of large, programmed genomic deletions. *Am J Hum Genet.*
683 2017;101(2):192–205.
- 684 8. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic
685 localization of common disease-associated variation in regulatory DNA. *Science (1979).*
686 2012;337(6099):1190–5.
- 687 9. Kheradpour P, Ernst J, Melnikov A, Rogov P, Wang L, Zhang X, et al. Systematic
688 dissection of regulatory motifs in 2000 predicted human enhancers using a massively
689 parallel reporter assay. *Genome Res.* 2013;23(5):800–11.
- 690 10. Shim S, Kwan KY, Li M, Lefebvre V, Nenad &, Estan S. Cis-regulatory control of
691 corticospinal system development and evolution. *Nature.* 2012;486:74–9.
- 692 11. Schoenfelder S, Javierre BM, Furlan-Magaril M, Wingett SW, Fraser P. Promoter Capture
693 Hi-C: High-resolution, genome-wide profiling of promoter interactions. *JoVE (Journal of
694 Visualized Experiments).* 2018;2018(136):e57320.
- 695 12. Schaub MA, Boyle AP, Kundaje A, Batzoglou S, Snyder M. Linking disease associations
696 with regulatory information in the human genome. *Genome Res.* 2012;22(9):1748–59.
- 697 13. Ernst J, Kheradpour P, Mikkelsen TS, Shores N, Ward LD, Epstein CB, et al. Mapping
698 and analysis of chromatin state dynamics in nine human cell types. *Nature.*
699 2011;473(7345):43–9.
- 700 14. Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: From polygenic to
701 omnigenic. *Cell.* 2017;169(7):1177–86.

- 702 15. Lee D, Gorkin DU, Baker M, Strober BJ, Asoni AL, McCallion AS, et al. A method to
703 predict the impact of regulatory variants from DNA sequence. *Nat Genet.*
704 2015;47(8):955–61.
- 705 16. Fearnley JM, Lees AJ. Ageing and Parkinson's disease: Substantia nigra regional
706 selectivity. *Brain.* 1991;114:2283–301.
- 707 17. Marras C, Beck JC, Bower JH, Roberts E, Ritz B, Ross GW, et al. Prevalence of
708 Parkinson's disease across North America. *NPJ Parkinsons Dis.* 2018;4(1):21.
- 709 18. Dorsey ER, Bloem BR. The Parkinson pandemic - A call to action. *JAMA Neurol.*
710 2018;75(1):9–10.
- 711 19. Ferrari E, Cardinale A, Picconi B, Gardoni F. From cell lines to pluripotent stem cells for
712 modelling Parkinson's Disease. *J Neurosci Methods.* 2020;340:108741.
- 713 20. Son JH, Chun HS, Joh TH, Cho S, Conti B, Lee JW. Neuroprotection and neuronal
714 differentiation studies using substantia nigra dopaminergic cells derived from transgenic
715 mouse embryos. *J Neurosci.* 1999;19(1):10.
- 716 21. Chang J, Zhang X Le, Yu H, Chen J. Downregulation of RTN1-C attenuates MPP+-
717 induced neuronal injury through inhibition of mGluR5 pathway in SN4741 cells. *Brain*
718 *Res Bull.* 2019;146:1–6.
- 719 22. Chen J, Li M, Zhou X, Xie A, Cai Z, Fu C, et al. Rotenone-induced neurodegeneration is
720 enabled by a p38-Parkin-ROS signaling feedback loop. *J Agric Food Chem.*
721 2021;69(46):13942–52.
- 722 23. Guiney SJ, Adlard PA, Lei P, Mawal CH, Bush AI, Finkelstein DI, et al. Fibrillar α -
723 synuclein toxicity depends on functional lysosomes. *J Biol Chem.* 2020;295(51):17497–
724 513.
- 725 24. Chun HS, Gibson GE, Degiorgio LA, Zhang H, Kidd VJ, Son JH. Dopaminergic cell
726 death induced by MPP+, oxidant and specific neurotoxicants shares the common
727 molecular mechanism. *J Neurochem.* 2001;76(4):1010–21.
- 728 25. Chun HS, Lee H, Son JH. Manganese induces endoplasmic reticulum (ER) stress and
729 activates multiple caspases in nigral dopaminergic neuronal cells, SN4741. *Neurosci Lett.*
730 2001;316(1):5–8.
- 731 26. Zeng W, Zhang W, Lu F, Gao L, Gao G. Resveratrol attenuates MPP+-induced
732 mitochondrial dysfunction and cell apoptosis via AKT/GSK-3 β pathway in SN4741 cells.
733 *Neurosci Lett.* 2017;637:50–6.
- 734 27. Cai Z, Zeng W, Tao K, Lu F, Gao G, Yang Q. Myricitrin alleviates MPP+-induced
735 mitochondrial dysfunction in a DJ-1-dependent manner in SN4741 cells. *Biochem*
736 *Biophys Res Commun.* 2015;458(2):227–33.

- 737 28. Mao K, Chen J, Yu H, Li H, Ren Y, Wu X, et al. Poly (ADP-ribose) polymerase 1
738 inhibition prevents neurodegeneration and promotes α -synuclein degradation via
739 transcription factor EB-dependent autophagy in mutant α -synucleinA53T model of
740 Parkinson's disease. *Aging Cell*. 2020;19(6).
- 741 29. Gui C, Ren Y, Chen J, Wu X, Mao K, Li H, et al. p38 MAPK-DRP1 signaling is involved
742 in mitochondrial dysfunction and cell death in mutant A53T α -synuclein model of
743 Parkinson's disease. *Toxicol Appl Pharmacol*. 2020;388:114874.
- 744 30. Dong Y, Xiong J, Ji L, Xue X. MiR-421 aggravates neurotoxicity and promotes cell death
745 in Parkinson's disease models by directly targeting MEF2D. *Neurochem Res*.
746 2021;46(2):299–308.
- 747 31. Yoo MS, Chun HS, Son JJ, DeGiorgio LA, Kim DJ, Peng C, et al. Oxidative stress
748 regulated genes in nigral dopaminergic neuronal cells: correlation with the known
749 pathology in Parkinson's disease. *Mol Brain Res*. 2003;110(1):76–84.
- 750 32. Wang B, Cai Z, Lu F, Li C, Zhu X, Su L, et al. Destabilization of survival factor MEF2D
751 mRNA by neurotoxin in models of Parkinson's disease. *J Neurochem*. 2014;130(5):720–
752 8.
- 753 33. Dekker J, Rippe K, Dekker M, Kleckner N. Capturing chromosome conformation.
754 *Science*. 2002;295(5558):1306–11.
- 755 34. Lieberman-Aiden E, Van Berkum NL, Williams L, Imakaev M, Ragooczy T, Telling A, et
756 al. Comprehensive mapping of long-range interactions reveals folding principles of the
757 human genome. *Science*. 2009;326(5950):289–93.
- 758 35. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying
759 mammalian transcriptomes by RNA-Seq. *Nat Methods*. 2008;5(7):621–8.
- 760 36. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native
761 chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding
762 proteins and nucleosome position. *Nat Methods*. 2013;10(12):1213–8.
- 763 37. Nissim-Eliraz E, Zisman S, Schatz O, Ben-Arie N. Noto3 integrates with the Shh-Foxa2
764 transcriptional network regulating the differentiation of midbrain dopaminergic neurons. *J*
765 *Mol Neurosci*. 2013;51:13–27.
- 766 38. Fishman-Jacob T, Reznichenko L, Youdim MBH, Mandel SA. A sporadic Parkinson
767 disease model via silencing of the ubiquitin-proteasome/E3 ligase component SKP1A. *J*
768 *Biol Chem*. 2009;284(47):32835–46.
- 769 39. Weihe E, Depboylu C, Schu□z B, Schäfer MKH, Eiden LE. Three types of tyrosine
770 hydroxylase-positive CNS neurons distinguished by dopa decarboxylase and VMAT2 co-
771 expression. *Cell Mol Neurobiol*. 2006;26(4–6):659–78.

- 772 40. Jonakait GM, Markey KA, Goldstein M, Dreyfus CF, Black IB. Selective expression of
773 high-affinity uptake of catecholamines by transiently catecholaminergic cells of the rat
774 embryo: studies in vivo and in vitro. *Dev Biol.* 1985;108(1):6–17.
- 775 41. Cochard P, Goldstein M, Black IB. Ontogenetic appearance and disappearance of tyrosine
776 hydroxylase and catecholamines in the rat embryo. *Proc Natl Acad Sci U S A.*
777 1978;75(6):2986–90.
- 778 42. Asmus SE, Parsons S, Landis SC. Developmental changes in the transmitter properties of
779 sympathetic neurons that innervate the periosteum. *J Neurosci.* 2000;20(4):1495–504.
- 780 43. Ladd AN, Charlet-B. N, Cooper TA. The CELF family of RNA binding proteins is
781 implicated in cell-specific and developmentally regulated alternative splicing. *Mol Cell*
782 *Biol.* 2001;21(4):1285–96.
- 783 44. Nishino J, Kim I, Chada K, Morrison SJ. Hmga2 promotes neural stem cell self-renewal in
784 young, but not old, mice by reducing p16Ink4a and p19Arf expression. *Cell.*
785 2008;135(2):227.
- 786 45. Park HJ, Hong M, Bronson RT, Israel MA, Frankel WN, Yun K. Elevated Id2 expression
787 results in precocious neural stem cell depletion and abnormal brain development. *Stem*
788 *Cells.* 2013;31(5):1010.
- 789 46. Dou Z, Son JE, Hui CC. Irx3 and Irx5 - Novel regulatory factors of postnatal
790 hypothalamic neurogenesis. *Front Neurosci.* 2021;15:1447.
- 791 47. Reid CA, Leaw B, Richards KL, Richardson R, Wimmer V, Yu C, et al. Reduced
792 dendritic arborization and hyperexcitability of pyramidal neurons in a Scn1b-based model
793 of Dravet syndrome. *Brain.* 2014;137:1701–15.
- 794 48. Stark R, Brown G. DiffBind: Differential binding analysis of ChIP-Seq peak data.
795 <http://bioconductor.org/packages/devel/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf>
- 796 49. Li H, Liu Y, Gu Z, Li L, Liu Y, Wang L, et al. p38 MAPK-MK2 pathway regulates the
797 heat-stress-induced accumulation of reactive oxygen species that mediates apoptotic cell
798 death in glial cells. *Oncol Lett.* 2018;15(1):775.
- 799 50. McClymont SA, Hook PW, Soto AI, Reed X, Law WD, Kerans SJ, et al. Parkinson-
800 associated SNCA enhancer variants revealed by open chromatin in mouse dopamine
801 neurons. *Am J Hum Genet.* 2018;103(6):874–92.
- 802 51. Xie Z, Bailey A, Kuleshov M v., Clarke DJB, Evangelista JE, Jenkins SL, et al. Gene set
803 knowledge discovery with Enrichr. *Curr Protoc.* 2021;1(3):e90.
- 804 52. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles G v., et al. Enrichr: Interactive
805 and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics.*
806 2013;14.

- 807 53. Kuleshov M v., Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr:
808 A comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids*
809 *Res.* 2016;44(W1):W90–7.
- 810 54. Franzén O, Gan LM, Björkegren JLM. PanglaoDB: a web server for exploration of mouse
811 and human single-cell RNA sequencing data. *Database.* 2019;2019(1):46.
- 812 55. Zeng L, Zhang P, Shi L, Yamamoto V, Lu W, Wang K. Functional impacts of NRXN1
813 knockdown on neurodevelopment in stem cell models. *PLoS One.* 2013;8(3).
- 814 56. Kaplan DR, Miller FD. Neurotrophin signal transduction in the nervous system. *Curr Opin*
815 *Neurobiol.* 2000;10(3):381–91.
- 816 57. Reddy-Alla S, Böhme MA, Reynolds E, Beis C, Grasskamp AT, Mampell MM, et al.
817 Stable positioning of Unc13 restricts synaptic vesicle fusion to defined release sites to
818 promote synchronous neurotransmission. *Neuron.* 2017;95(6):1350-1364.e12.
- 819 58. Storm R, Cholewa-Waclaw J, Reuter K, Bröhl D, Sieber M, Treier M, et al. The bHLH
820 transcription factor Olig3 marks the dorsal neuroepithelium of the hindbrain and is
821 essential for the development of brainstem nuclei. *Development.* 2009;136(2):295–305.
- 822 59. Sung HY, Chen WY, Huang HT, Wang CY, Chang S bin, Tzeng SF. Down-regulation of
823 interleukin-33 expression in oligodendrocyte precursor cells impairs oligodendrocyte
824 lineage progression. *J Neurochem.* 2019;150(6):691–708.
- 825 60. Liu H, Hu Q, D’Ercole AJ, Ye P. Histone Deacetylase 11 regulates oligodendrocyte-
826 specific gene expression and cell development in OL-1 oligodendroglia cells. *Glia.*
827 2009;57(1):1.
- 828 61. Sakry D, Yigit H, Dimou L, Trotter J. Oligodendrocyte precursor cells synthesize
829 neuromodulatory factors. *PLoS One.* 2015;10(5).
- 830 62. Yam CH, Fung TK, Poon RYC. Cyclin A in cell cycle control and cancer. *Cellular and*
831 *Mol Life Sci.* 2002;59(8):1317–26.
- 832 63. Borlado LR, Méndez J. CDC6: from DNA replication to cell cycle checkpoints and
833 oncogenesis. *Carcinogenesis.* 2008;29(2):237–43.
- 834 64. Ma L, Zhao X, Zhu X. Mitosin/CENP-F in mitosis, transcriptional control, and
835 differentiation. *J Biomed Sci.* 2006;13(2):205–13.
- 836 65. Nagahama Y, Ueno M, Miyamoto S, Morii E, Minami T, Mochizuki N, et al. PSF1, a
837 DNA replication factor expressed widely in stem and progenitor cells, drives tumorigenic
838 and metastatic properties. *Cancer Res.* 2010;70(3):1215–24.
- 839 66. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING
840 database in 2021: customizable protein-protein networks, and functional characterization
841 of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 2021;49(D1):D605–12.

- 842 67. Chun HS, Lee H, Son JH. Manganese induces endoplasmic reticulum (ER) stress and
843 activates multiple caspases in nigral dopaminergic neuronal cells, SN4741. *Neurosci Lett.*
844 2001;316(1):5–8.
- 845 68. Choi KC, Kim SH, Ha JY, Kim ST, Son JH. A novel mTOR activating protein protects
846 dopamine neurons against oxidative stress by repressing autophagy related cell death. *J*
847 *Neurochem.* 2010;112(2):366–76.
- 848 69. Bryja V, Čajánek L, Grahn A, Schulte G. Inhibition of endocytosis blocks Wnt signalling
849 to β -catenin by promoting dishevelled degradation. *Acta Physiologica.* 2007;190(1):55–
850 61.
- 851 70. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, et al.
852 Identification of novel risk loci, causal insights, and heritable risk for Parkinson’s disease:
853 a meta-analysis of genome-wide association studies. *Lancet Neurol.* 2019;18(12):1091–
854 102.
- 855 71. Blauwendraat C, Nalls MA, Singleton AB, Blauwendraat C, Singleton AB, Nalls MA.
856 The genetic architecture of Parkinson’s disease. *Lancet Neurol.* 2020;19:170–8.
- 857 72. Chang D, Nalls MA, Hallgrímsdóttir IB, Hunkapiller J, Brug M van der, Cai F, et al. A
858 meta-analysis of genome-wide association studies identifies 17 new Parkinson’s disease
859 risk loci. *Nat Genet.* 2017;49(10):1511–6.
- 860 73. Blauwendraat C, Heilbron K, Vallerga CL, Bandres Ciga S, von Coelln R, Pihlstrøm L,
861 et al. Parkinson’s disease age at onset genome-wide association study: Defining
862 heritability, genetic loci, and α -synuclein mechanisms. *Movement Disorders.*
863 2019;34(6):866–75.
- 864 74. Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, Dutra A, et al. Mutation in
865 the α -synuclein gene identified in families with Parkinson’s disease. *Science.*
866 1997;276(5321):2045–7.
- 867 75. Singleton AB, Farrer M, Johnson J, Singleton A, Hague S, Kachergus J, et al. Alpha-
868 synuclein locus triplication causes Parkinson’s disease. *Science.* 2003;302(5646):841.
- 869 76. Ibáñez P, Bonnet AM, Débarges B, Lohmann E, Tison F, Pollak P, et al. Causal relation
870 between alpha-synuclein gene duplication and familial Parkinson’s disease. *The Lancet.*
871 2004;364(9440):1169–71.
- 872 77. Taylor SC, Nadeau K, Abbasi M, Lachance C, Nguyen M, Fenrich J. The ultimate qPCR
873 experiment: Producing publication quality, reproducible data the first time. *Trends*
874 *Biotechnol.* 2019;37(7):761–74.
- 875 78. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively
876 parallel digital transcriptional profiling of single cells. *Nat Commun.* 2017;8(1):1–12.

- 877 79. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell
878 transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol.*
879 2018;36(5):411–20.
- 880 80. Corces MR, Trevino AE, Hamilton EG, Greenside PG, Sinnott-Armstrong NA, Vesuna S,
881 et al. An improved ATAC-seq protocol reduces background and enables interrogation of
882 frozen tissues. *Nat Methods.* 2017;14(10):959–62.
- 883 81. Andrews S. FASTQC: A quality control tool for high throughput sequence data. 2010.
884 <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- 885 82. Ewels P, Magnusson M, Lundin S, Källér M. MultiQC: summarize analysis results for
886 multiple tools and samples in a single report. *Bioinformatics.* 2016;32(19):3047–8.
- 887 83. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods.*
888 2012;9(4):357.
- 889 84. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence
890 Alignment/Map format and SAMtools. *Bioinformatics.* 2009;25(16):2078.
- 891 85. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, et al. Model-based
892 analysis of ChIP-Seq (MACS). *Genome Biol.* 2008;9(9):1–9.
- 893 86. Amemiya HM, Kundaje A, Boyle AP. The ENCODE blacklist: Identification of
894 problematic regions of the genome. *Sci Rep.* 2019;9(1).
- 895 87. Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, et al. An integrated
896 encyclopedia of DNA elements in the human genome. *Nature.* 2012;489(7414):57–74.
- 897 88. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic
898 features. *Bioinformatics.* 2010;26(6):841–2.
- 899 89. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: A
900 next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.*
901 2016;44(W1):W160–5.
- 902 90. Karolchik D, Hinricks AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, et al. The
903 UCSC Table Browser data retrieval tool. *Nucleic Acids Res.* 2004;32(Database
904 issue):D493.
- 905 91. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, et al.
906 Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes.
907 *Genome Res.* 2005;15(8):1034–50.
- 908 92. Shin H, Liu T, Manrai AK, Liu SX. CEAS: cis-regulatory element annotation system.
909 *Bioinformatics.* 2009;25(19):2605–6.
- 910 93. Liu T, Ortiz JA, Taing L, Meyer CA, Lee B, Zhang Y, et al. Cistrome: An integrative
911 platform for transcriptional regulation studies. *Genome Biol.* 2011;12(8):1–10.

- 912 94. McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, et al. GREAT
913 improves functional interpretation of cis-regulatory regions. *Nat Biotechnol.*
914 2010;28(5):495–501.
- 915 95. Conway JR, Lex A, Gehlenborg N. UpSetR: an R package for the visualization of
916 intersecting sets and their properties. *Bioinformatics.* 2017;33(18):2938–40.
- 917 96. Kim D, Langmead B, Salzberg SL. HISAT: A fast spliced aligner with low memory
918 requirements. *Nat Methods.* 2015;12(4):357–60.
- 919 97. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al.
920 Bioconductor: Open software development for computational biology and bioinformatics.
921 *Genome Biol.* 2004;5(10).
- 922 98. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, et al.
923 Orchestrating high-throughput genomic analysis with Bioconductor. *Nat Methods.*
924 2015;12(2):115–21.
- 925 99. Liao Y, Smyth GK, Shi W. The Subread aligner: fast, accurate and scalable read mapping
926 by seed-and-vote. *Nucleic Acids Res.* 2013;41(10):e108–e108.
- 927 100. Liao Y, Smyth GK, Shi W. Sequence analysis featureCounts: an efficient general purpose
928 program for assigning sequence reads to genomic features. *Bioinformatics.*
929 2014;30(7):923–30.
- 930 101. Zhu A, Ibrahim JG, Love MI. Heavy-tailed prior distributions for sequence count data:
931 removing the noise and preserving large differences. *Bioinformatics.* 2019;35(12):2084–
932 92.
- 933 102. Wickham H. *ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York;
934 2016. <https://ggplot2.tidyverse.org>.
- 935 103. Belaghzal H, Dekker J, Gibcus JH. Hi-C 2.0: An optimized Hi-C procedure for high-
936 resolution genome-wide mapping of chromosome conformation. *Methods.* 2017;123:56–
937 65.
- 938 104. Wingett S, Ewels P, Furlan-Magaril M, Nagano T, Schoenfelder S, Fraser P, et al. HiCUP:
939 Pipeline for mapping and processing Hi-C data. *F1000Res.* 2015;4.
- 940 105. Cairns J, Freire-Pritchett P, Wingett SW, Várnai C, Dimond A, Plagnol V, et al.
941 CHiCAGO: Robust detection of DNA looping interactions in Capture Hi-C data. *Genome*
942 *Biol.* 2016;17(1):1–17.
- 943

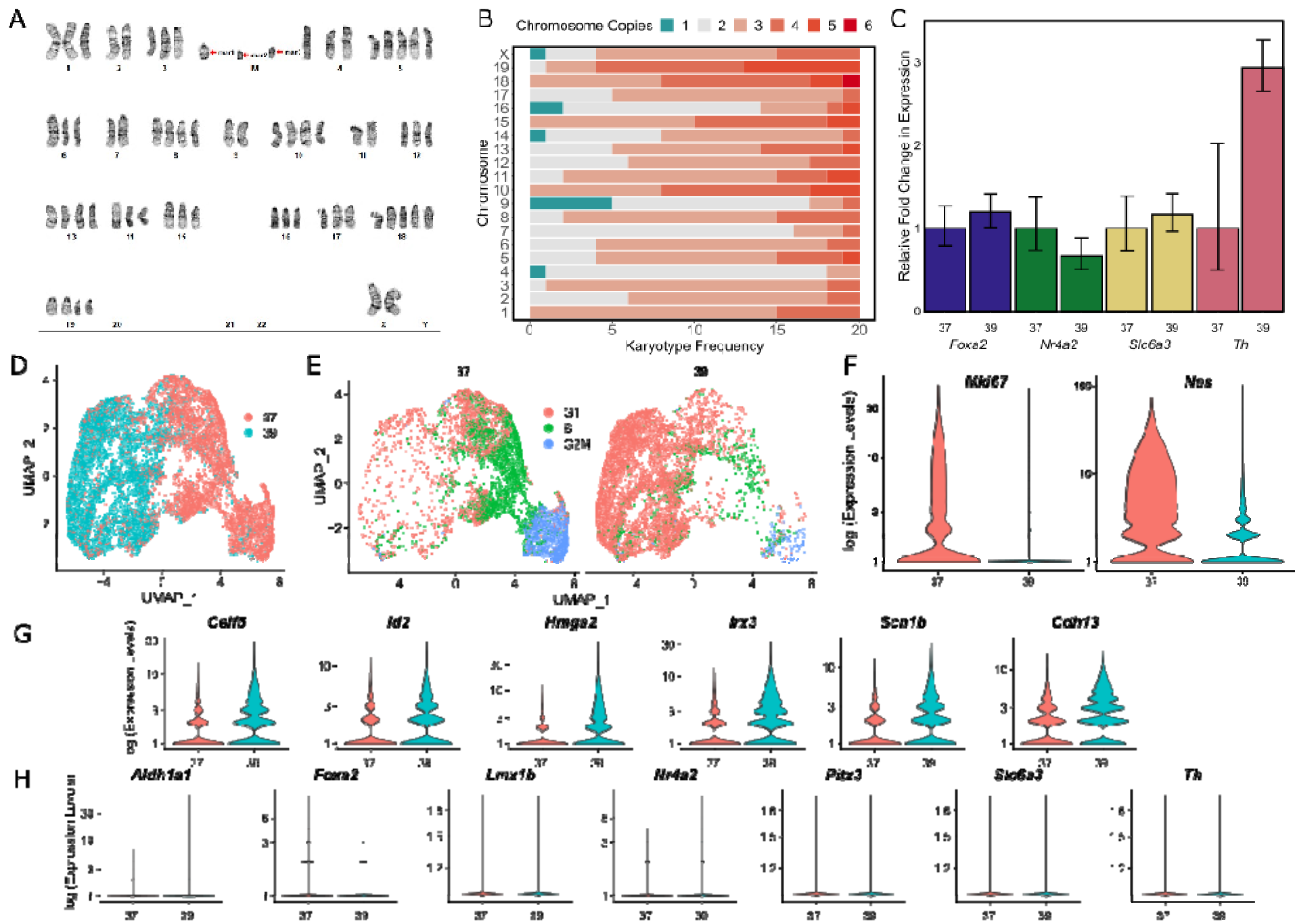


Figure 1. Characterizing the genomic stability and differentiation consistency of the temperature sensitive SN4741 cell line. **A)** A representative karyogram of SN4741 cells, indicating structural instability (M; marker chromosomes) and unstable triploidy. **B)** A stacked bar plot summarizing the aneuploidy frequency of each chromosome over 20 SN4741 karyotypes. **C)** Assaying expression of dopaminergic neuron markers by RT-qPCR indicates that *Foxa2*, *Nr4a2*, and *Slc6a3* remain at similar expression levels, but *Th* expression increases, when SN4741 cells are shifted from the permissive temperature (37°C) to the non-permissive temperature (39°C). **D)** UMAP plot of scRNA-seq at the permissive and non-permissive temperatures indicates that cells at each temperature are transcriptionally distinct. **E)** Analysis of scRNA-seq data demonstrates that shifting the cells to the non-permissive temperature is accompanied by a shift in cell cycle stage from G2M and S phases to primarily G1 phase. **F)** Violin plots generated with scRNA-seq data show that *Mki67*, a marker of cellular proliferation, and *Nes*, a neural stem cell marker, are both expressed at the permissive temperature (37°C), with little to no expression at the non-permissive temperature (39°C). **G)** Violin plots generated with scRNA-seq data show that transcripts associated with immature neurons are upregulated when SN4741 cells are shifted to the non-permissive temperature. **H)** Violin plots generated with scRNA-seq data show that expression of DA neural markers, *Aldh1a1*, *Foxa2*, *Lmx1b*, *Nr4a2*, *Pitx3*, *Slc6a3*, and *Th*, remain at similar levels when SN4741 cells are shifted to the non-permissive temperature.

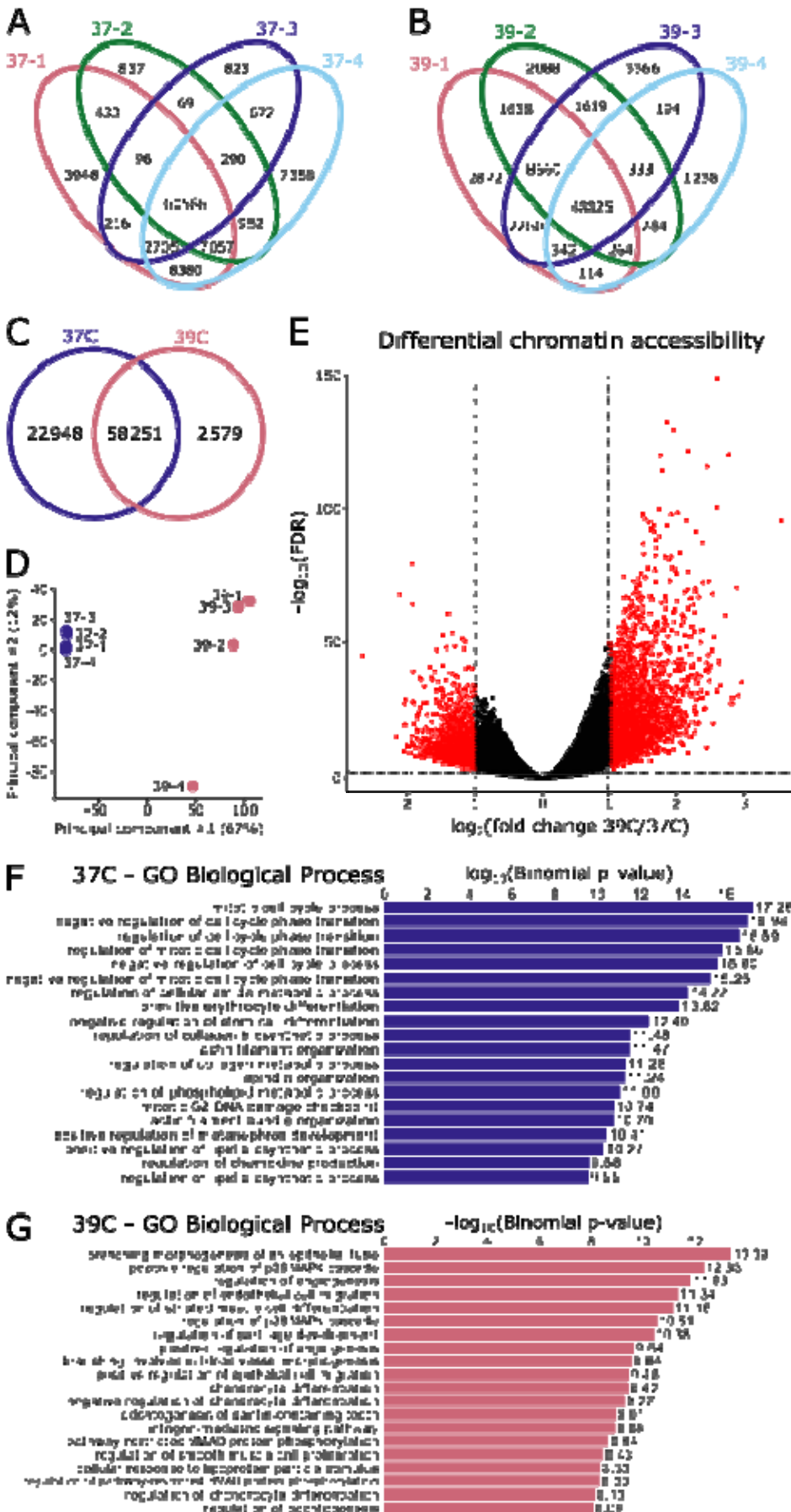


Figure 2: Changes in chromatin accessibility suggest a reduction in potency at the non-permissive temperature. A, B) Replicates are highly similar within temperature conditions, with the majority of peaks present in all four replicates. C) The two temperatures share 58,251 regions of open chromatin but do not overlap completely. D) Principal component analysis resolves the two temperatures on the first principal component. E) Differential accessibility analysis identifies 5,055 differentially accessible regions, with 2,654 preferentially open in the permissive temperature (37°C) and 2,401 preferentially accessible at the non-permissive temperature (39°C). F) Gene ontology (GO) of genes adjacent regions that are preferentially open at the permissive temperature are associated with regulation of the cell cycle and negative regulation of differentiation, as is appropriate for this temperature. G) Gene ontology of genes adjacent regions that are preferentially open at the non-permissive temperature are associated with a variety of differentiation fates (blood vessels, muscle cells, cartilage/chondrocytes). Additionally, two of the top gene ontology terms relate to the p38 MAPK cascade, which has been found to be activated as a cellular response to heat stress.

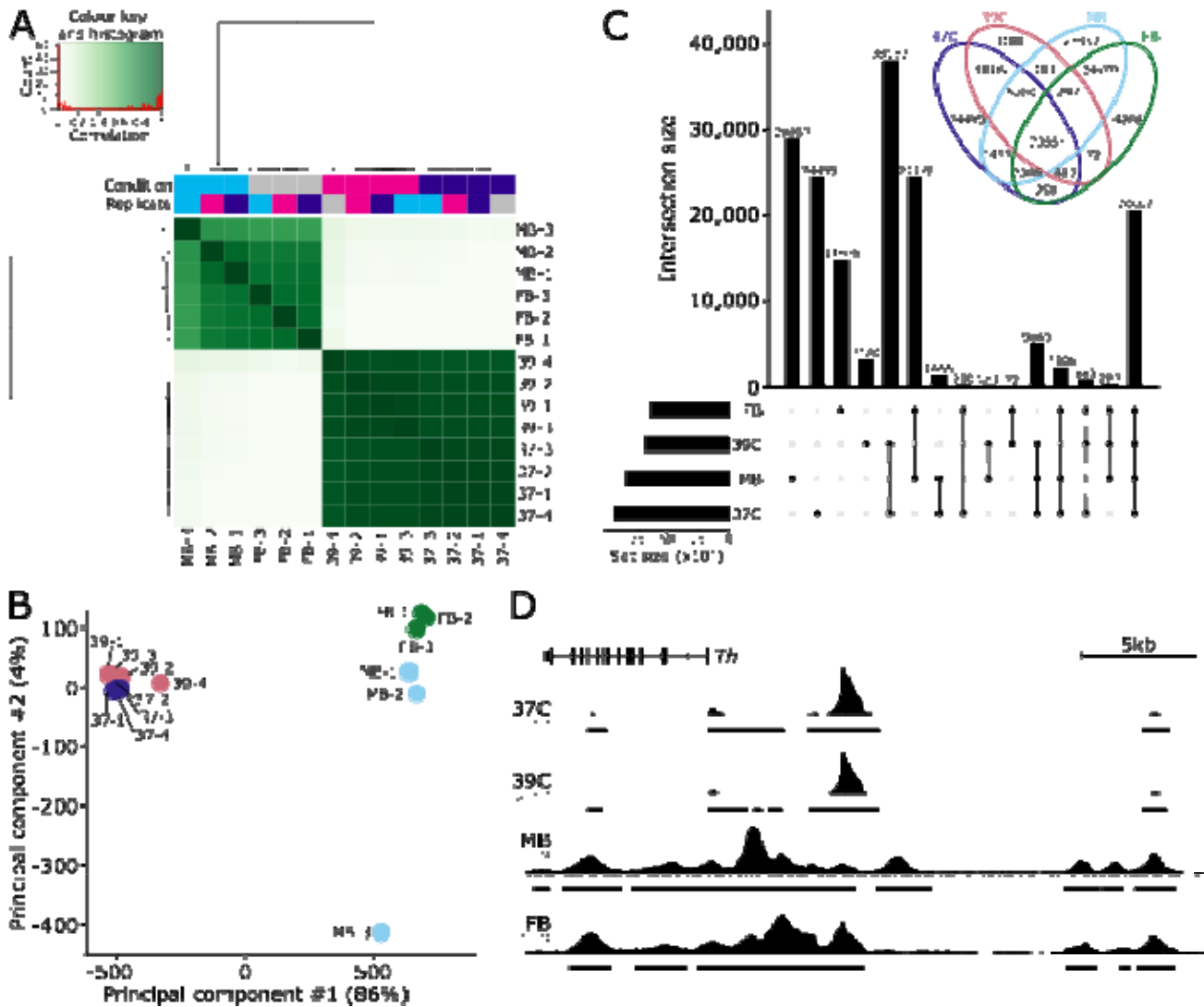


Figure 3: Chromatin accessibility of SN4741 cells do not resemble *ex vivo* dopaminergic neurons. **A)** SN4741 samples are highly correlated with each other but very poorly correlate with the open chromatin landscape of either midbrain (MB) or forebrain (FB) embryonic mouse dopaminergic neurons. **B)** Principal component analysis shows a clear separation between the *ex vivo* and *in vitro* samples along PC1, representing 86% of the variance. **C)** An upset plot and associated Venn diagram quantify the overlap of peaks between the four conditions and show the poor relationship between the SN4741 cells and the *ex vivo* mouse dopaminergic neurons. Most peaks are specific to a single cell type/temperature or are restricted to either the *ex vivo* or *in vitro* samples. Few peaks are specifically shared between the non-permissive temperature and the *ex vivo* samples; for example, there are just 183 peaks that are shared exclusively by the MB dopaminergic neurons and the SN4741 cells at the non-permissive temperature. **D)** A genome track showing the normalized read pile up and called consensus peaks in each of the cell types/temperatures at the key dopaminergic neuron specification gene, *Th*. The chromatin accessibility is largely similar within *ex vivo* or *in vitro* cells but bear little resemblance to each other.

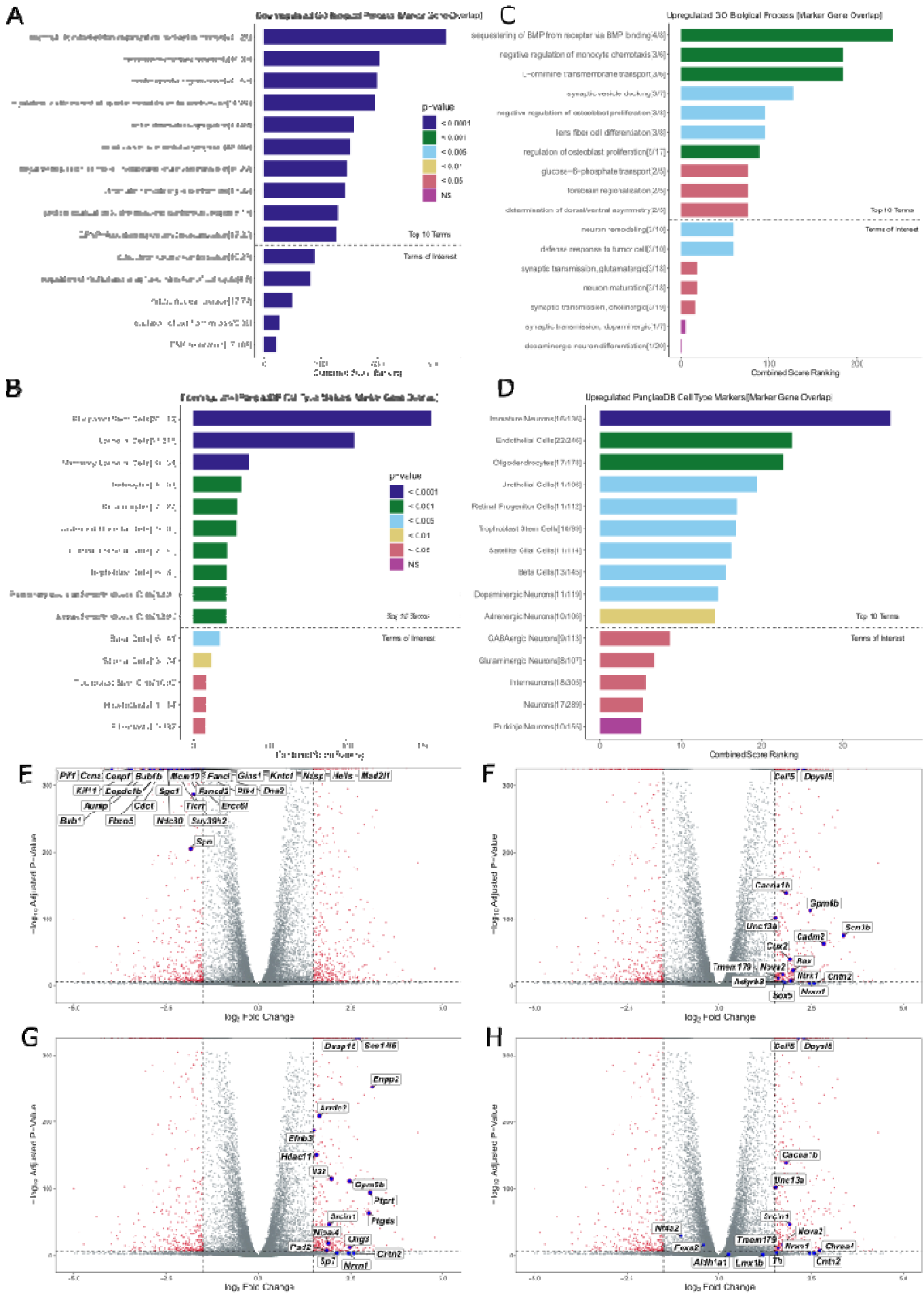


Figure 4. Gene Ontology and Differential Expression Analysis of Bulk RNA-seq Data: **A)** Top 10 GO terms for downregulated DE genes in SN4741 cells at the non-permissive temperature, followed by GO terms of interest (below dotted line). Terms were evaluated using Combined Score Ranking = (p-value computed using the Fisher exact test)*(z-score computed by assessing the deviation from the expected rank), based on enrichment of DE genes that overlap with Enrichr input genes for each term (the end of each bar). **B)** Top 10 predicted cell types based on downregulated DE genes in SN4741 cells at the non-permissive temperature, followed by predicted cell types of interest (below dotted line). Terms were evaluated using Combined Score Ranking = (p-value computed using the Fisher exact test)*(z-score computed by assessing the deviation from the expected rank), based on enrichment of DE genes that overlap with PanglaoDB input genes for each term (the end of each bar). **C)** Top 10 GO terms for downregulated DE genes in SN4741 cells at the non-permissive temperature. **D)** Top 10 predicted cell types based on upregulated DE genes in SN4741 cells at the non-permissive temperature. **E)** Volcano plot of $-\log_{10}$ adjusted p-value versus \log_2 fold change with DESeq2 after lfc shrinkage, contrasting the fold change in expression of SN4741 cells at 39°C, using SN4741 cells at 37°C as reference. Red points = genes that are statistically differentially expressed (adjusted p-value < 0.01, $|\log_2FC| > 1.5$). Blue points = Overlapping immature neuron marker genes. **F)** Blue points = Overlapping immature neuron marker genes **G)** Blue points = Overlapping oligodendrocyte marker genes. **H)** Blue points = Overlapping DA neuron marker genes.

Figure 5. Validation of gene ontology and comparison to bulk-RNA-seq from ex vivo DA neurons. **A)** A bar chart showing normalized bulk RNA-seq read counts from genes upregulated at 39°C that overlap the “immature neurons” predicted cell type. **B)** A bar chart showing normalized bulk RNA-seq read counts from genes upregulated at 39°C that overlap the “oligodendrocyte” predicted cell type. **C)** A bar chart showing normalized bulk RNA-seq read counts from genes downregulated at 39°C that overlap the “pluripotent stem cell” predicted cell type. **D)** A Pearson correlation heatmap comparing the transcriptomes of SN4741 cells at 37°C and 39°C to midbrain (MB) or forebrain (FB) embryonic mouse dopaminergic neurons.