Systematic investigation of imprinted gene expression and enrichment in the mouse brain explored at single-cell resolution

M. J. Higgs[1], M. J. Hill[2,3], R. M. John[4], and A. R. Isles[1,*]

1 Behavioural Genetics Group, MRC Centre for Neuropsychiatric, Genetics and Genomics, Neuroscience and Mental Health Research Institute, Cardiff University, Cardiff, UK

2 Division of Psychological Medicine and Clinical Neurosciences, MRC Centre for Neuropsychiatric Genetics and Genomics, School of Medicine, Cardiff University, Cardiff, UK

3 UK Dementia Research Institute, School of Medicine, Cardiff University, Cardiff, UK

4 School of Biosciences, Cardiff University, Cardiff, UK

* Corresponding author: A. R. Isles, Behavioural Genetics Group, MRC Centre for Neuropsychiatric Genetics and Genomics, Neuroscience and Mental Health Research Institute, Cardiff University, Cardiff CF24 4HQ, UK. E-mail: islesar1@cardiff.ac.uk

Abstract

Over 200 murine imprinted genes have currently been identified, many are expressed in the brain, and a subset have been associated with neurological and behavioural phenotypes. These studies have led to the suggestion that imprinted genes show enriched expression in certain key brain functions; however, this has never been formally tested. Here we use fifteen available single-cell RNA sequencing datasets to investigate imprinted gene over-representation in the brain, with the aim of identifying key hotspots of expression and, by extension, function. We establish that imprinted genes are indeed over-represented in the adult brain compared to other adult tissues, and in neurons specifically compared to other cell-types. Brain wide datasets allowed us to demonstrate enrichment of imprinted genes in the hypothalamus, ventral midbrain, pons and medulla. Finally, using scRNA-seq datasets focusing on these regions of enrichment, we were able to identify hypothalamic neuroendocrine populations and the monoaminergic hindbrain neurons as specific hotspots of imprinted gene expression. These analyses provide a robust assessment of the neural systems on which imprinted genes converge and in turn suggest the neuronal regulation of motivated behaviours such as feeding, parental behaviour and sleep as functional hotspots for imprinting.

28 **INTRODUCTION**

29 Imprinted genes demonstrate a preferential or exclusively monoallelic expression from either the
30 maternal or paternal allele in an epigenetically predetermined manner (a parent-of-origin effect, POE),
31 and to date approximately 200 canonical (associated with a parental specific epigenetic mark)
32 imprinted genes have been identified in the mouse (~150 in humans) (Ferguson-Smith, 2011; Tucci et
33 al., 2019). This epigenetic regulation makes genomic imprinting an evolutionary puzzle as these genes
34 are effectively haploid and thereby negate many of the benefits of diploidy (Orr, 1995). Studying the
35 patterns of expression and function of imprinted genes may therefore shed light on the drivers leading
36 to the evolution of genomic imprinting. For instance, characterization of a number of imprinted genes
37 points to convergence on placental function (Peters, 2014), in line with the predictions of early
38 theoretical ideas (Moore & Haig, 1991). Other studies of expression and/or function suggest that the
39 brain and other major endocrine tissues are also foci for genomic imprinting; with the brain
40 consistently emerging as the adult tissue with the largest number of expressed imprinted genes (e.g.
41 Andergassen et al., 2017).

42 A role for imprinted genes in the brain was suggested by behavioural (Cattanach & Kirk, 1985) and
43 neurological phenotypes (Keverne, Fundele, Narasimha, Barton, & Surani, 1996) observed in early
44 imprinted gene mouse models, as well as the behavioural deficits seen in imprinting disorders such as
45 Prader-Willi Syndrome and Angelman's syndrome (Angulo, Butler, & Cataletto, 2015; Nicholls, Knoll,
46 Butler, Karam, & Lalande, 1989). Subsequent studies have revealed diverse roles for imprinted genes
47 in the brain. During development, several imprinted genes are involved in the processes of neural
48 differentiation, migration, axonal outgrowth and apoptosis (Perez, Rubinstein, & Dulac, 2016). In the
49 adult brain, studies of mice carrying manipulations of individual imprinted genes have suggested a
50 wide range of behavioural roles including maternal care (Li et al., 1999), feeding (Davies et al., 2015)
51 social behaviour (Garfield et al., 2011; McNamara, John, & Isles, 2018), learning and memory (Jiang et

52 al., 1998), cognition (Dent et al., 2018; Relkovic et al., 2010), and more recently, sleep and circadian
53 activity (Lassi et al., 2012).

54 In addition to studies on individual imprinted genes, there are a limited number of studies that take a
55 systems level approach to characterizing the role of genomic imprinting in the brain. The early studies
56 examining developing and adult chimeras of normal and parthenogenetic (two paternal genomes) or
57 androgenetic/gynogenetic (two maternal genomes) cells indicated distinct regional distribution for
58 maternal (cortex and hippocampus) and paternal (hypothalamus) expressed genes (Allen et al., 1995;
59 Keverne et al., 1996). More recently, Gregg et al. (2010) used the known imprinting status of 45
60 imprinted genes and the Allen Brain Atlas to track dichotomous expression of imprinted genes across
61 118 brain regions to identify brain-wide patterns of expression. A number of imprinted genes were
62 expressed in every brain region but, in agreement with the previous work, there were key regions with
63 the largest number of imprinted genes expressed including the hypothalamus (MPOA & ARC), the
64 monoaminergic nuclei (Dorsal Raphe (5-HT), Locus Coeruleus (NA) and PAG & VTA (DA)), alongside
65 two amygdala nuclei (CEA, MEA).

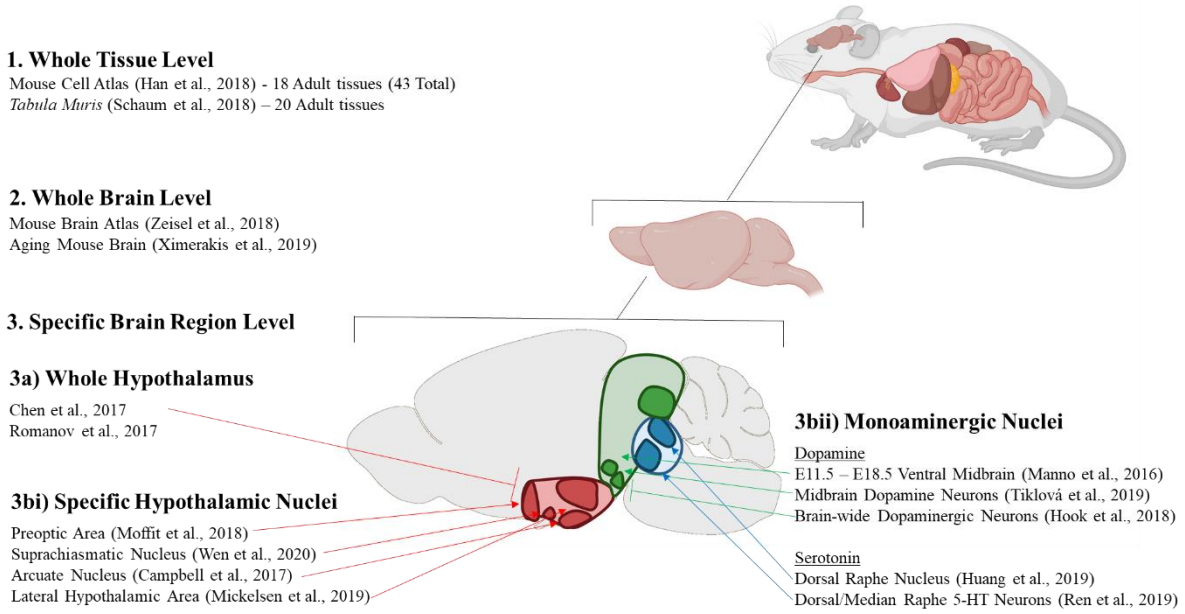
66 This previous work demonstrates that there are a number of imprinted genes expressed in the brain
67 and when expression is disrupted there are resultant brain and/or behavioural changes. However, it
68 is estimated ~80% of all genes in the genome are expressed in the brain (Lein et al., 2007; Negi &
69 Guda, 2017) so the question remains, *is* imprinted gene expression actually enriched in the brain
70 compared to other adult tissues? Similarly, many of the previous studies identifying novel imprinted
71 genes and/or mapping expression of imprinted genes in the brain did not test whether these genes
72 were especially enriched in given brain regions but simply asked if they were expressed (at any level)
73 or not (Andergassen et al., 2017; Babak et al., 2015; DeVeale, van der Kooy, & Babak, 2012; Gregg et
74 al., 2010; Perez et al., 2015). Here we address these questions by systematically investigating
75 imprinted gene enrichment and over-representation in the murine brain by exploiting the rapidly
76 expanding number of single-cell RNA sequencing (scRNA-seq) datasets. We demonstrate that

imprinted genes are indeed enriched in brain relative to other adult tissues, with over-representation in the hypothalamus, ventral midbrain, medulla, and pons. Furthermore, within these regions, imprinted genes are enriched in key, behaviourally associated, hypothalamic neurons and the midbrain/hindbrain raphe serotonin neurons. These analyses provide a robust assessment of the neural systems on which imprinted genes converge, provide testable predictions for the role of specific genes, and putative imprinted gene networks, in neural and behavioural phenotypes and, in turn, shed light on the putative evolutionary drivers of genomic imprinting in the brain.

Materials and Methods

All data were acquired through publicly available resources or personal correspondence with the authors. Datasets were chosen based on data availability, but also to create a hierarchical sequence of data analysis (from whole tissue to specific brain region); see Figure 1 for a summary graphic of the datasets used in this analysis. Each dataset was filtered and normalised according to the original published procedure and then run through the same basic workflow as specified below, with minor adjustments laid out for each dataset collated in Appendix A. Table 1 lays out the basic parameters of each dataset included in the analyses. The levels of analysis included: 1) Whole-tissue level – would imprinted genes emerge as a brain enriched group compared to other major organs? 2) Whole-brain level – would imprinted genes demonstrate a regional or cell lineage specific enrichment? 3) Brain-region level – if level 2 demonstrated a regional specificity could imprinted gene enrichment be observed at the neuronal subpopulation level? Where possible two or more independent datasets were used at each level of analysis to strengthen any conclusions with convergent findings.

Figure 1. Summary of datasets used in this analysis. The tissue sequenced and original publication are included for each dataset. Datasets are further organised hierarchically by whole tissue, whole brain, and brain region specific level.



All graphical representations and statistical analyses were conducted using R 3.6.2 (R Core Team, 2019) in RStudio (RStudio Team, 2020). All processing scripts are available at https://github.com/MJHiggs/IG-Single-Cell-Enrichment. All data files are available at the following Open Science Framework repository (https://osf.io/jx7kr/) All summary diagrams were created with BioRender.com.

Table 1. Dataset specific sequencing and processing information for all datasets analysed. Datasets are organised by level of analysis and compared for single-cell sequencing protocol, Animal and Tissue processing, Cell Quality Filters used, No. of Cells in final dataset and Data Normalisation procedure followed.

Dataset	Level	Protocol	Animal/Tissue	Cell Filter	Cell No.	Normalisation
Mouse Cell Atlas (Han et al., 2018)	Whole Tissue	Microwell -seq	C57BL/6J , M+F, 6-10 weeks, E14.5 and neonatal	1500 highest quality cells per tissue	61,637	100,000 transcripts log transformed
Tabula Muris (Schaum et al., 2018)	Whole Tissue	Smart-seq2 (FACs)	C57BL/6J 7 mice (4F/3M), 10-15 weeks, virgin	reads>50,000 genes>500	44,879	ln(CPM+1)
Mouse Brain Atlas (Zeisel et al., 2018)	Whole Brain	10X chromium	CD1, M+F, p12 – 30, week 6 and week 8,	600<UMI 1.2 UMI:gene	160,796	10,000 UMI per cell log transformed
Whole Brain (Ximerakis et al., 2019)	Whole Brain	10X chromium	C57BL/6J, 16 mice (8 mice 2-3 months, 8 mice 21-22 months), whole brain minus hindbrain	200<UMI<30,000 250<gene<6,000	37,089	10,000 UMI per cell log transformed
Whole Hypothalamus (Romanov et al., 2017)	Whole Hypothalamus	Fluidigm C1	C57BL/6J, M+F, 14-28 days	UMI>1500	2,882	10,000 UMI per cell log transformed
Whole Hypothalamus (Chen, Wu, Jiang, & Zhang, 2017)	Whole Hypothalamus	Drop-seq	B6D2F1 mice (C57B6 female × DBA2 male) - 7 Female, 8-10 weeks	Genes>2000	3319	10,000 UMI per cell log transformed

Preoptic Area (POA) (Moffitt et al., 2018)	Specific Hypothalamic Nucleus	10x chromium	C57BL/6J -6 mice (3M, 3F) - 7-8 weeks, tissue block spanning POA and surrounding	none stated	31,299	By total UMI per cell, log10
Arcuate Nucleus (ARC) (Campbell et al., 2017)	Specific Hypothalamic Nucleus	Drop-seq	C57BL/6J – 53mice - 4-12 weeks, virgin, M+F	genes>800	20,921	ln(CPM+1)
Lateral Hypothalamic Area (LHA) (Mickelsen et al., 2019)	Specific Hypothalamic Nucleus	Drop-seq (FACs)	C57BL/6J M+F p25-p32 – LH dissected with minor contamination from DMH, VZI	UMI≥500 mito<0.4	6,944	By total UMI per cell log transformed
Suprachiasmatic Nucleus (SCN) (Wen et al., 2020)	Specific Hypothalamic Nucleus	10x chromium	C57BL/6J at timepoint ZT8	genes>200	8,679	By total UMI per cell, log transformed
Dorsal Raphe Nucleus (DRN) (Huang et al., 2019)	Monoaminergic System	Drop-seq	C57BL/6J -8 mice (4M, 4F), 8-10 weeks	18,000>UMI>500 6000>gene>200 mito<0.1	39,411	10,000 UMI per cell log transformed
Serotonin neurons of DRN and MRN (Ren et al., 2019)	Monoaminergic System	Smart-seq2	C57BL/6J - 14 mice (8M, 6F) 40 – 45 days	none stated	999	ln(CPM+1)
E11.5 - E18.5 ventral midbrain (La Manno et al., 2016)	Monoaminergic System	Fluidigm C1	CD1, E11.5 - E18.5 - 271 embryos	2000>UMI>26,000	1,907	By total UMI per cell, log transformed
Midbrain Dopamine Neurons (Tiklová et al., 2019)	Monoaminergic System	Smart-seq2	C57BL/6NRj mice, M+F, Pitx3 expressing cells from E13.5, E15.5, E18.5, P1, P7, P90	Norm reads >100,000 >15% of all genes	1,106	RPKM
Whole Brain Dopamine (Hook et al., 2018)	Monoaminergic System	Smart-seq2	C57BL/6J at E15.5 and P7	2000< genes <10,000 1000> RNA >40,000	396	log(FPKM+1)

Custom Imprinted Gene List

Based on the list of murine imprinted genes recently published in Tucci et al. (2019), 119 imprinted genes were chosen for analysis (Supplemental Table 1) consisting of canonical protein-coding and long noncoding RNA imprinted genes (with at least two independent demonstrations of their POE status). The only exceptions to multiple independent demonstrations of POE were four genes (*Bmf*, *B3gnt2*, *Ptk2*, *Gm16299*) identified by Perez et al. (2015)where POE was validated for 16 brain regions and 7 adult tissues. Small non-coding RNAs such as micro-RNAs (miRs) and small nucleolar RNAs (snoRNAs), which represent ~10% of identified imprinted genes, were excluded from the analysis as their sequences were not detected/subsumed by larger transcripts in the majority of the datasets.

Basic Workflow

Data were downloaded in the available form provided by the original authors (either raw or processed) and, where necessary, were processed (filtered, batch-corrected and normalized) to match the author's original procedure. Cell quality filters were specific to each dataset and summarised in Table

115 1. A consistent filter, to keep all genes expressed in at least 20 cells or (when possible) with at least 50
116 reads, was applied to remove genes unlikely to play a functional role due to low expression and being
117 expressed in few cells. Datasets from the brain were analysed both at the global cell level (neuronal
118 and non-neuronal cells) and neuron specific level (only cells classified as neurons) with genes filtered
119 for the ≥20 cell expression at each level before subsequent analysis. Cell identities were supplied using
120 the outcome of cell clustering carried out by the original authors, so that each cell included in the
121 analysis had a cell-type or tissue-type identity. This was acquired as metadata supplied with the
122 dataset or as a separate file from personal correspondence with the author.

123 Positive differential expression between identity groups were carried out using one-sided Wilcoxon
124 rank-sum tests (assuming the average expression of cells within the current identity group is 'greater'
125 than the average of cells from all other groups). The test was performed independently for each gene
126 and for each identity group vs. all other groups. The large number of *p* values were corrected for
127 multiple comparisons using a horizontal Benjamini-Hochberg correction, creating *q* values. Fold-
128 change (FC) values, percentage expression within the identity group and percentage expressed within
129 the rest were also calculated. We considered genes to be significantly positively differentially
130 expressed (significantly upregulated) in a group compared to background expression if it had a $q \leq$
131 0.05 and a positive Log2 fold change compared to background. Occasionally, when identity groups
132 within a dataset indicated a disproportionately large number of upregulated genes (in our case >50%
133 of all genes upregulated), a more conservative enrichment analysis was necessary and so a Log2 fold
134 change cut off of 1 (i.e. 2-fold-change or larger) was used to restrict upregulated genes to stronger
135 markers of their identity groups (analyses in which this occurred are described within Appendix A, and
136 indicated in the table legends of the relevant analyses).

137 The same custom list of imprinted genes with reliable parent-of-origin effects was used for all
138 analyses, and all genes were included as long as the gene passed the 20-cell filter. The first statistical
139 analysis for enrichment was an Over-Representation Analysis (ORA) using a one-way Fisher's Exact

140 Test. The aim was to assess whether the number of imprinted genes considered to be upregulated as
141 a proportion of the total number of imprinted genes in the dataset (passing the 20-cell filter) was
142 statistically higher than would be expected by chance when compared to the total number of
143 upregulated genes as a proportion of the overall number of genes in the dataset (passing the 20-cell
144 filter). To limit finding over-represented identity groups with only a few upregulated imprinted genes,
145 an identity group was required to have ≥5% of the imprinted genes upregulated for ORA to be
146 conducted. Subsequent p-values for all eligible identity groups were corrected using a Bonferroni
147 correction. This provided a measure of whether imprinted genes are expressed above expectation
148 (the expression pattern of a random gene list) in particular identity groups.

149 To further examine the presence of imprinted genes within tissues, and to provide a different
150 perspective to over-representation, we conducted a Gene-Set Enrichment Analysis (GSEA) of
151 imprinted genes using a publicly available, light-weight implementation of the GSEA algorithm in R
152 (Subramanian et al., 2005, https://github.com/JEFworks/liger). The GSEA was conducted using log-
153 fold-change values of statistically significant upregulated genes and is a more conservative measure
154 since it tests whether imprinted genes enriched in the stronger markers of a group (the genes with
155 the highest fold change for a group vs. the rest). Groups analysed were selected as having an minimum
156 number of upregulated imprinted genes to measure enrichment for (minimum of 15 as suggested by
157 the GSEA user guide (https://www.gsea-msigdb.org/gsea/doc/GSEAUserGuideFrame.html) and
158 having an average fold change of the upregulated imprinted genes greater than the average fold
159 change of the rest of the upregulated genes for that tissue. Again, multiple p values generated from
160 GSEA were corrected using a Bonferroni correction. The absolute number of IGs top-expressed in a
161 tissue was also reported for every analysis.

162 For whole tissue level and whole brain level datasets, additional parent-of-origin specific analyses
163 were carried out. The imprinted gene list was divided into maternally expressed genes (MEGs) and
164 paternally expressed genes (PEGs) and the analyses detailed above were run separately for these two

new gene groups. For imprinted genes with known parent-of-origin variability based on tissue type (*Igf2* and *Grb10*), the parent-of-origin characterisation of these genes was changed accordingly.

The specific imprinted genes making up the enrichments can be found in the 'Upregulated_IGs.csv' and 'IG_Subpopulation_list.csv' files for each analysis in the repository (https://osf.io/jx7kr/).

Results

1) Whole tissue level analysis

1.1 Imprinted gene are over-represented in the brain at whole level analysis

The Mouse Cell Atlas (MCA) (Han et al., 2018) and the *Tabula Muris* (TM) (Schaum et al., 2018) provide single cell compendiums of mouse organs with overlap of a number of tissues including the brain providing a snapshot of imprinted gene expression in adult tissues. The MCA additionally sequenced a selection of organs at a foetal and embryonic timepoint including the embryonic day (E)14.5 placenta.

Comparing imprinted gene expression in adult tissues from both datasets revealed the convergence of tissues in which imprinted genes are over-represented. In the MCA adult tissue dataset, there was an over-representation of imprinted genes in the pancreas, brain and bladder (Table 2, Figure 2). Including all the tissues from the three stages of development in the analysis (Supp Table 2), we demonstrated that imprinted genes were further over-represented in neonatal tissues (5/6 tissues), the placenta and embryonic mesenchyme. The only adult tissue to be significantly over-represented in this comparison was the pancreas.

Table 2. Imprinted gene over-representation in MCA adult tissues (Han et al., 2018). *Identity* – Tissue identities for the cells used in analysis; *Up Reg* – number of upregulated genes (total number of genes in the dataset in brackets); *IG* – number of imprinted genes upregulated (total number of IGs in the dataset in brackets); *ORA p* – *p* value from over representation analysis on groups with minimum 5% of total IGs; *ORA q* – Bonferroni corrected *p* value from ORA; *Mean FC IG* – mean fold change for upregulated imprinted genes; *Mean FC Rest* – mean fold change for all other upregulated genes; *GSEA p* – *p* value from Gene Set Enrichment Analysis for identity groups with 15+ IGs and Mean FC IG > Mean FC Rest; *Top Tissue IGs* – Sum of IGs with highest mean expression value for that identity group.

Identity	Up Reg (20,547)	IG (97)	ORA *p*	ORA *q*	Mean FC IG	Mean FC Rest	GSEA *p*	Top Tissue IGs
Pancreas	4338	53	1.39E-13	**2.22E-12**	7.27	7.09	0.0985	22

Brain	4878	42	8.91E-06	**0.000143**	7.42	87.52	-	19
Bladder	7612	54	6.66E-05	**0.001066**	3.09	4.42	-	8
Uterus	4597	31	0.013730	0.219678	3.76	5.40	-	7
Kidney	2728	15	0.275439	1	9.70	115.45	-	5
Ovary	5791	29	0.341656	1	4.14	5.25	-	5
Lung	2330	12	0.391714	1	3.07	78.81	-	4
Liver	2692	11	0.715070	1	3.78	52.54	-	3
Mammary Gland Virgin	2004	7	0.830750	1	2.82	2.66	-	0
Stomach	4307	14	0.952432	1	2.85	38.29	-	3
Small Intestine	4553	14	0.973867	1	3.78	83.51	-	2
Thymus	3574	10	0.977553	1	2.29	4.18	-	2
Bone Marrow	2615	6	0.986131	1	2.83	2.74	-	1
Peripheral Blood	2620	6	0.986360	1	2.56	2.41	-	0
Muscle	2689	6	0.989175	1	6.28	7.33	-	3
Testis	6086	15	0.999428	1	25.32	4331.20	-	10
Spleen	2246	1	-	-	4.90	3.75	-	1
Prostate	539	0	-	-	0.00	327.78	-	0

In the *Tabula Muris* dataset analysis (Table 3, Figure 3), there was a significant over-representation in the brain (non-myeloid) cells, pancreas, bladder and the muscle-based tissues - diaphragm, trachea, and limb muscles. No tissue at this level showed a significant GSEA.

Table 3. Imprinted gene over-representation in *Tabula Muris* adult tissues (Schaum et al., 2018). *Identity* – Tissue identities for the cells used in analysis; *Up Reg* – number of upregulated genes with 2FC (total number of genes in the dataset in brackets); *IG* – number of imprinted genes upregulated with 2FC (total number of IGs in the dataset in brackets); *ORA p* – *p* value from over representation analysis on groups with minimum 5% of total IGs; *ORA q* – Bonferroni corrected *p* value from ORA; *Mean FC IG* – mean fold change for upregulated imprinted genes; *Mean FC Rest* – mean fold change for all other upregulated genes; *GSEA p* – *p* value from Gene Set Enrichment Analysis for identity groups with 15+ IGs and Mean FC IG > Mean FC Rest; *GSEA q* – Bonferroni corrected *p* values from GSEA; *Top Tissue IGs* – Sum of IGs with highest mean expression value for that identity group.

Identity	Up Reg (22,112)	IG (107)	ORA *p*	ORA *q*	Mean FC IG	Mean FC Rest	GSEA *p*	GSEA *q*	Top Tissue IGs
Diaphragm	426	19	1.99E-13	**2.58E-12**	6.49	4.90	0.2159	0.4318	4
Limb Muscle	783	24	3.32E-13	**4.32E-12**	9.02	5.34	0.0661	0.1322	8
Pancreas	4154	43	2.17E-07	**2.82E-06**	12.52	12.88	-	-	29
Trachea	2024	25	9.49E-06	**0.000123**	3.81	4.82	-	-	5
Brain (Non-Myeloid)	3122	31	5.09E-05	**0.000661**	12.16	14.14	-	-	14
Bladder	3354	31	0.000196	**0.002549**	3.30	5.34	-	-	16
Fat	1273	12	0.020121	0.261578	3.46	3.72	-	-	1
Heart	1113	10	0.043228	0.561958	2.87	5.16	-	-	0
Mammary Gland	1836	12	0.176572	1	3.52	5.27	-	-	3
Liver	1819	7	0.786891	1	6.19	54.80	-	-	3
Tongue	4341	15	0.948735	1	4.15	7.40	-	-	8
Aorta	4444	14	0.978213	1	7.47	26.91	-	-	2
Large Intestine	5399	11	0.999941	1	5.95	13.00	-	-	5

Brain (Myeloid)	1032	5	-	-	3.39	6.80	-	-	2
Kidney	604	3	-	-	24.94	22.84	-	-	1
Lung	924	2	-	-	2.73	5.65	-	-	0
Marrow	1969	5	-	-	7.65	5.26	-	-	4
Skin	1625	4	-	-	4.15	8.42	-	-	1
Spleen	627	1	-	-	4.63	4.32	-	-	0
Thymus	700	4	-	-	3.46	7.56	-	-	1

Separating the analysis for MEGs and PEGs (Supp Table 3a, 3b, 4a, 4b) revealed a similar pattern of enrichment in both the datasets (Figure 2 and Figure 3). Focusing on the brain, PEGs were over-represented in both datasets (MCA – q = 0.0002, *TM* – q = 0.0002) and MEGs were not (MCA – q = 0.51, *TM* – q = 1). PEGS were also over-represented in the pancreas (MCA – q = 0.0004, *TM* – q = 0.0004), diaphragm (q = 0.0006) and limb muscle ($q = 9.63 \times 10^{-5}$) while MEGs were over-represented in the bladder (MCA – q = 0.003, *TM* – q = 0.021), the pancreas (MCA - $q = 9.59 \times 10^{-9}$, *TM* – q = 0.03) and in all three muscular tissues of the *Tabula Muris* (diaphragm – $q = 1.44 \times 10^{-8}$, limb muscle – $q = 1.66 \times 10^{-7}$, trachea – q = 0.003)

In summary, at this level, imprinted genes were enriched in embryonic and neonatal tissues compared to adult. However, across the adult tissues imprinted genes were convergently over-represented in the pancreas, bladder and the brain.

Figure 2. Tissues with over-representation in MCA (Han et al., 2018). Coloured tissues with bold labels were over-represented tissues using all imprinted genes, tissues with a blue circle behind were over-represented for PEGS alone and red circles represent the same for MEGs.

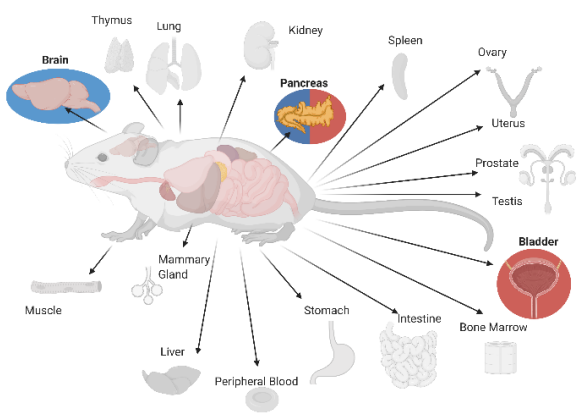
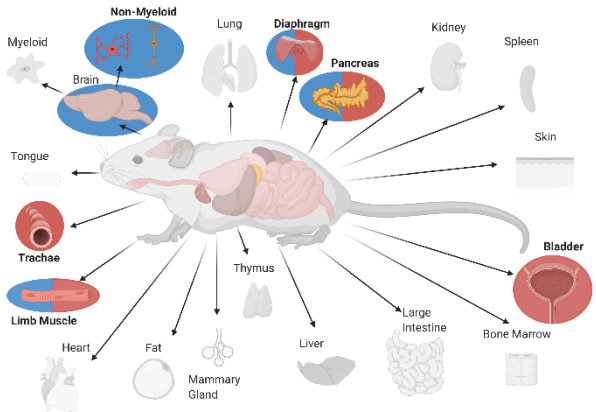


Figure 3. Tissues with over-representation in *Tabula Muris* (Schaum et al., 2018). Coloured tissues with bold labels were over-represented tissues using all imprinted genes, tissues with a blue circle behind were over-represented for PEGS alone and red circles represent the same for MEGs.



2. Whole Brain level analysis

2.1 Imprinted genes are enriched in neurons and neuroendocrine cells

For the brain enrichment, we first analysed Ximerakis et al.'s (2019) dataset, where cells were grouped from the whole mouse brain (minus the hindbrain) according to cell lineage. Imprinted genes were over-represented in neuroendocrine cells and mature neurons (Table 4). Neuroendocrine cells were defined as a heterogeneous cluster, containing peptedergic neurons and neurosecretory cells (e.g. *Avp* and *Oxt* expressing), expressing neuronal marker genes (e.g. *Syt1* & *Snap25*) but critically having unique expression of *Baiap3* which plays an important role in the regulation of exocytosis in neuroendocrine cells. GSEA additionally showed that the imprinted genes were enriched in the strong markers for neuroendocrine cells only (Figure 4). 32 imprinted genes had their highest expression in the neuroendocrine cells as compared to one imprinted gene in the other mature neurons.

Figure 4. Gene Set Enrichment Analysis of imprinted Genes for genes upregulated in neuroendocrine cells. In the analysis, genes are sorted by strength by which they mark this neuronal cluster (sorted by fold change values) indicated by the bar (bottom). The genes are arrayed left (strongest marker) to right and blue lines mark where imprinted genes fall on this array. The vertical axis indicates an accumulating weight, progressing from left to right and increasing or decreasing depending on whether the next gene is an imprinted gene or not. The *p*-value represents the probability of observing the maximum value of the score (red dashed line) if the imprinted genes are distributed randomly along the horizontal axis.

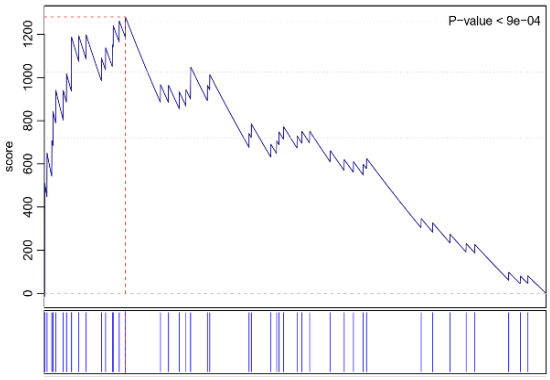


Table 4. Imprinted gene over-representation in neural lineage types (Ximerakis et al., 2019). *Identity* – Cell Lineage identities for the cells used in analysis; *Up Reg* – number of upregulated genes with 2FC (total number of genes in the dataset in brackets); *IG* – number of imprinted genes upregulated with 2FC (total number of IGs in the dataset in brackets); *ORA p* – *p* value from over representation analysis on groups with minimum 5% of total IGs; *ORA q* – Bonferroni corrected *p* value from ORA; *Mean FC IG* – mean fold change for upregulated imprinted genes; *Mean FC Rest* – mean fold change for all other upregulated genes; *GSEA p* – *p* value from Gene Set Enrichment Analysis for identity groups with 15+ IGs and Mean FC IG > Mean FC Rest; *GSEA q* – Bonferroni corrected *p* values from GSEA; *Top Tissue IGs* – Sum of IGs with highest mean expression value for that identity group.

Identity	Up Reg (14,661)	IG (88)	ORA *p*	ORA *q*	Mean FC IG	Mean FC Rest	GSEA *p*	GSEA *q*	Top Tissue IGs
Neuroendocrine cells	4234	48	3.84E-07	**6.92E-06**	16.34	5.49	9.00E-04	**0.0036**	32
Mature Neurons (all types)	2989	33	0.000156	**0.002800**	7.62	8.76	-	-	1

Vascular and leptomeningeal cells	1735	19	0.006444	0.115995	85.82	12.22	0.0946	0.3784	4
Tanycytes	1698	18	0.011174	0.201139	6.55	8.67	-	-	7
Arachnoid barrier cells	2275	20	0.047213	0.849842	15.31	23.21	-	-	7
Oligodendrocyte precursor cells	1549	12	0.216078	1	3.06	7.43	-	-	2
Pericytes	1764	13	0.256801	1	7.95	7.59	-	-	3
Choroid plexus epithelial cells	2873	20	0.266209	1	79.24	23.94	0.1717	0.6868	9
Oligodendrocytes	1124	8	0.361824	1	3.95	10.45	-	-	5
Vascular smooth muscle cells	3133	20	0.418819	1	7.95	6.85	0.1622	0.6488	2
Immature Neurons	713	5	0.427092	1	3.34	5.90	-	-	0
Olfactory ensheathing glia	1195	8	0.428082	1	8.42	23.85	-	-	2
Hemoglobin-expressing vascular cells	1930	11	0.620935	1	5.07	6.01	-	-	3
Endothelial cells	1496	8	0.686994	1	5.19	8.25	-	-	0
Astrocyte-restricted precursors	1853	10	0.688635	1	3.88	4.58	-	-	1
Ependymocytes	3429	17	0.849270	1	15.45	54.05	-	-	3
Neuronal-restricted precursor	2458	9	0.969946	1	2.75	13.07	-	-	0
Hypendymal cells	1843	6	0.972273	1	12.15	21.45	-	-	3
Astrocytes	1430	4	-	-	2.57	6.20	-	-	0
Neural stem cells	1202	4	-	-	8.29	4.02	-	-	0
Dendritic cells	1493	3	-	-	3.61	14.66	-	-	0
Microglia	1452	3	-	-	18.96	17.95	-	-	3
Macrophages	1275	2	-	-	2.96	19.89	-	-	0
Monocytes	1150	2	-	-	10.53	25.45	-	-	1
Neutrophils	607	1	-	-	6.48	89.29	-	-	0

Zeisel et al.'s (2018) Mouse Brain Atlas (MBA) allowed a much deeper investigation of nervous system enrichment with sequence of the entire murine nervous system and identifying cells by both brain regions and cell type. Concordant with the previous findings, primary analysis separating cells by lineage and using a 2-fold cut off revealed over-representation of imprinted genes in neurons only (Table 5).

Table 5. Imprinted gene over-representation in nervous system cell types (Zeisel et al., 2018). *Identity* – Cell identities for the cells used in analysis; *Up Reg* – number of upregulated genes with 2FC (total number of genes in the dataset in brackets); *IG* – number of imprinted genes upregulated with 2FC (total number of IGs in the dataset in brackets); *ORA p* – *p* value from over representation analysis on groups with minimum 5% of total IGs; *ORA q* – Bonferroni corrected *p* value from ORA; *Mean FC IG* – mean fold change for upregulated imprinted genes; *Mean FC Rest* – mean fold change for all other upregulated genes; *Top Tissue IGs* – Sum of IGs with highest mean expression value for that identity group.

Identity	Up Reg	IG	ORA *p*	ORA *q*	Mean FC	Mean FC	Top Tissue

	(19,550)	(107)			IG	Rest	IGs
Neurons	5710	45	0.005619	**0.033711**	11.69	24.97	45
Vascular	2474	23	0.009964	0.059782	17.24	26.64	16
Oligos	1587	11	0.279767	1	4.13	11.48	11
Peripheral Glia	2821	16	0.527651	1	5.42	12.64	12
Ependymal	3684	20	0.609092	1	24.52	66.95	15
Immune	1564	7	0.786456	1	13.42	93.05	5
Astrocytes	1539	4	-	-	2.88	10.73	3

2.2 Imprinted genes are enriched in the hypothalamus, ventral midbrain, pons, and medulla.

After confirming neuron-specific enrichment of imprinted genes in the MBA dataset, further MBA analysis was performed on cells classified as neurons. These neuronal cells were then grouped by brain/nervous system regions. Significant over-representation was seen in neurons of the hypothalamus, ventral midbrain, medulla, and pons (Table 6). The pons and medulla had the largest number of imprinted genes upregulated (44 for both).

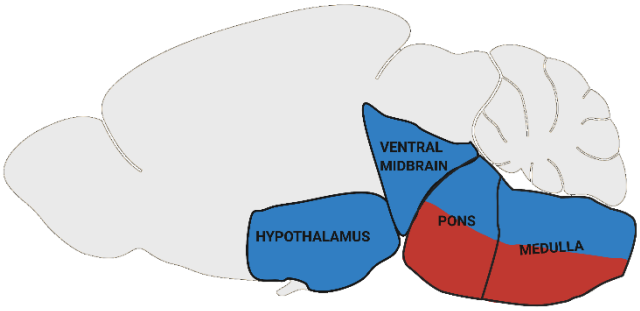
Table 6. Imprinted gene over-representation in nervous system regions (Zeisel et al., 2018). *Identity* – Nervous System Regional identities for the cells used in analysis; *Up Reg* – number of upregulated genes with 2FC (total number of genes in the dataset in brackets); *IG* – number of imprinted genes upregulated with 2FC (total number of IGs in the dataset in brackets); *ORA p – p* value from over representation analysis on groups with minimum 5% of total IGs; *ORA q* – Bonferroni corrected *p* value from ORA; *Mean FC IG* – mean fold change for upregulated imprinted genes; *Mean FC Rest* – mean fold change for all other upregulated genes; *GSEA p – p* value from Gene Set Enrichment Analysis for identity groups with 15+ IGs and Mean FC IG > Mean FC Rest; *GSEA q* – Bonferroni corrected *p* values from GSEA; *Top Tissue IGs* – Sum of IGs with highest mean expression value for that identity group.

Identity	Up Reg (18,338)	IG (104)	ORA *p*	ORA *q*	Mean FC IG	Mean FC Rest	GSEA *p*	GSEA *q*	Top Tissue IGs
Medulla	3148	44	4.28E-09	**6.43E-08**	4.79	4.01	0.0838	0.1676	15
Hypothalamus	1040	22	1.17E-07	**1.75E-06**	4.92	5.84	-	-	8
Pons	3583	44	2.26E-07	**3.40E-06**	4.15	3.91	0.0862	0.1724	22
Ventral Midbrain	1229	18	0.000255	**0.003818**	4.90	5.55	-	-	3
Ventral Striatum	689	8	0.048406	0.726091	3.92	4.92	-	-	0
Posterior Cortex	1090	9	0.185621	1	2.64	3.20	-	-	2
Enteric Nervous System	3885	25	0.325423	1	9.07	121.01	-	-	11
Sympathetic Nervous System	2805	18	0.369582	1	11.37	57.95	-	-	9
Anterior Cortex	979	6	0.510693	1	2.72	3.30	-	-	1
Dorsal Midbrain	1045	6	0.575441	1	2.20	4.85	-	-	3
Hippocampus - CA1	1082	6	0.609895	1	3.01	4.02	-	-	2
Thalamus	1441	8	0.610565	1	2.90	6.36	-	-	0
Dorsal Striatum	1196	6	0.705864	1	4.03	5.43	-	-	2

Somatosensory Cortex	2121	11	0.705992	1	4.09	3.70	-	-	8
Dorsal Root Ganglion	3607	15	0.949802	1	11.55	75.87	-	-	8
Amygdala	452	4	-	-	4.65	4.11	-	-	2
Cerebellum	240	0	-	-	0.00	32.30	-	-	0
Antero-Central Cortex	646	3	-	-	4.59	4.31	-	-	1
Middle Cortex	623	5	-	-	3.29	3.24	-	-	0
Dentate Gyrus	796	4	-	-	3.79	4.16	-	-	2
Hippocampus	631	4	-	-	4.86	3.82	-	-	2
Olfactory Bulb	445	4	-	-	4.02	8.27	-	-	2

Regional analysis for MEGs and PEGs separately (Supp Table 5a, 5b), revealed that PEGs were over-represented in hypothalamus (q = 1.21x10^{-7}), ventral midbrain (q = 0.023), the pons (q = 0.001) and the medulla (q = 0.0001); while MEGs were only over-represented in the medulla (q = 0.002) and pons (q = 0.011) with a significant GSEA for the pons (q = 0.028), (Figure 5).

Figure 5. Imprinted gene over-represented brain regions in the MBA (Zeisel et al., 2018). Regions over-represented for PEGs alone are coloured blue while regions over-represented for MEGS alone are coloured red.



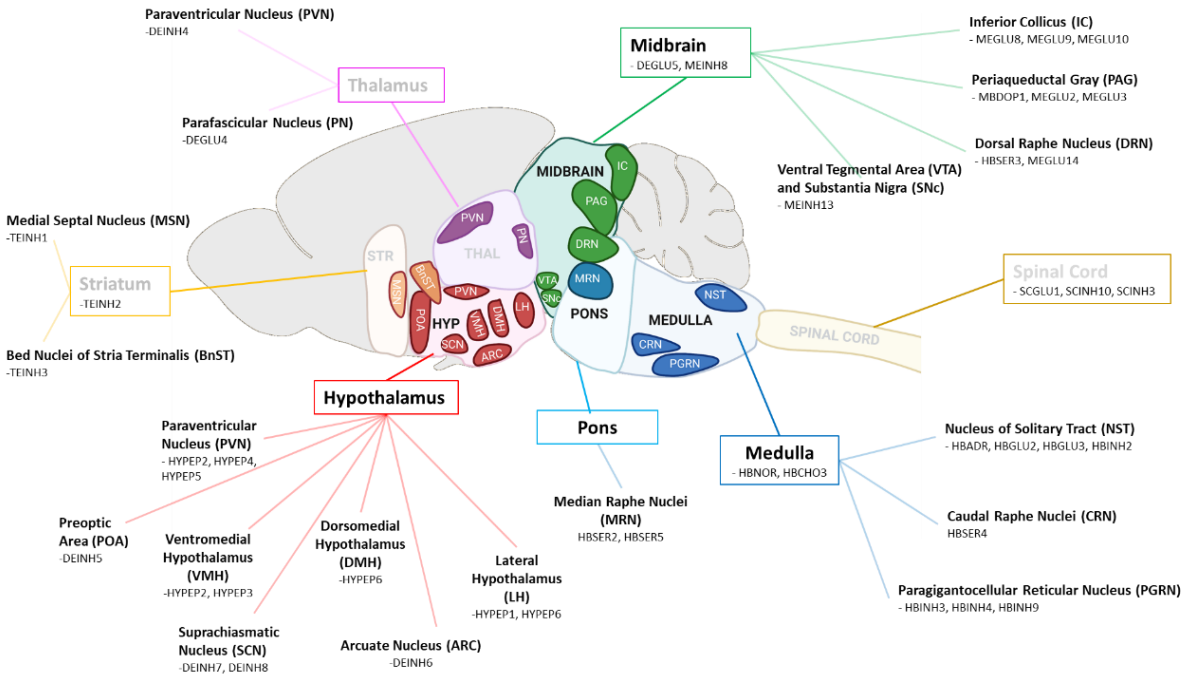
We then recategorized the neurons into individual neuronal subtypes to uncover the specific neural population underlying the enrichment seen in the hypothalamus, pons & medulla, and midbrain (Figure 6; Supp Table 6). The hypothalamus was represented by a selection of inhibitory and peptedergic neurons. Inhibitory neurons with over-representation of imprinted genes included: Deinh5 (*Esr1, Prlr*) representing a Preoptic Area population, Deinh6 (*Agrp, Npy*) representing an Arcuate nucleus population, and Deinh7 (*Avp, Rgs16*) and Deinh8 (*Nms, Vip*) suspected to be Suprachiasmatic nucleus populations. For peptedergic neurons, over-representation was seen in HYPEP3 (*Gpr101, Tac1*) a ventromedial population, HYPEP2 (*Otp, Trh, Ucn3*) a ventromedial/paraventricular population, HYPEPE1 (*Trh, Otp, Ngb*) a lateral hypothalamic population, HYPEP4 (*Oxt, Otp*) an oxytocin magnocellular population of the paraventricular and supraoptic nuclei,

HYPEP5 (*Avp, Otp, Pdyn*) a vasopressin paraventricular nucleus population and HYPEP6 (*Hcrt, Trhr*) an orexin producing population of the dorsomedial/lateral hypothalamus.

The midbrain, medulla and pons were represented by a number of cell groups. Of interest, over-representation was seen in the medulla-based adrenergic (HBAR) and noradrenergic (HBNOR) groups and dopaminergic neurons of the midbrain/PAG (MBDOP1). There were also several inhibitory (MEINH, HBIN) and excitatory neuron (MEGLU, HBGLU) types spread across the 3 regions in nuclei (Figure 6). The serotonergic populations of the raphe nuclei of these regions (HBSER) are particularly prominent since HBSER2-5 all have an imprinted mean fold change over twice as large as the rest (and are the only neuron groups to do so) and HBSER4 was the only neuron subpopulation out of the 214 to have a significant GSEA for imprinted genes after correction (Figure 7).

Additional regions of over-representation included TEINH1-3 representing neurons in the pallium and striatum, DEGLU4 and DEINH4 representing neurons from the thalamus and a selection of neurons from the spinal cord (SC).

Figure 6. Anatomical labelling of all the neural subpopulations with a significant over-representation of imprinted genes ($q \leq 0.05$) in the Mouse Brain Atlas (Zeisel et al., 2018). The predicted brain nuclei localisation of neuronal subpopulations specified in the MBA and enriched for imprinted genes. Brain regions that were not found to be enriched for imprinted genes are greyed out. The full Enrichment Analysis is available in Supplementary Table 6.



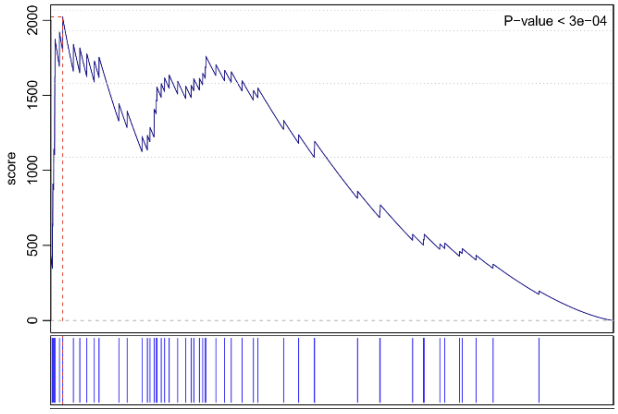
3) Brain Region level Analysis

3a) Whole Hypothalamus

3a.1) Imprinted genes are over-represented in hypothalamic neurons

Two datasets with single cell sequencing data for the adult hypothalamus exist (Chen et al., 2017; Romanov et al., 2017). Both datasets clustered their data into neuronal subpopulations allowing us to look for convergent imprinted enrichment across major hypothalamic neuronal subtypes. Analysis revealed a clear neuronal bias in expression for imprinted genes (Supp Table 7a, 8a). Within the Romanov et al. (2017) data, there was a significant over-representation of imprinted genes in neurons (q = 3.01x10^{-6}). A similar pattern was present in the Chen et al. (2017) data (q = 0.001), with the latter also demonstrating a significant GSEA in neurons (Appendix C1, p = 0.019).

Figure 7. Gene Set Enrichment Analysis of Imprinted Genes for genes upregulated in HBSER4 neurons in the Mouse Brain Atlas (Zeisel et al., 2018). In the analysis, genes are sorted by strength by which they mark this neuronal cluster (sorted by fold change values) indicated by the bar (bottom). The genes are arrayed left (strongest marker) to right and blue lines mark where imprinted genes fall on this array. The vertical axis indicates an accumulating weight, progressing from left to right and increasing or decreasing depending on whether the next gene is an imprinted gene or not. The p-value represents the probability of observing the maximum value of the score (red dashed line) if the imprinted genes are distributed randomly along the horizontal axis.



3a.2) Imprinted genes are over-represented in specific hypothalamic neuronal sub-types

Within the Chen et al. (2017) dataset, 4/33 hypothalamic neuronal subtypes had a significant over-representation of imprinted genes (Table 7a; Supp table 7b). The four subtypes were all GABAergic neurons, specifically: GABA13 (*Slc18a2, Gal*), with high expression of *Gal, Esr1, Prlr* and *Th,* present in a few hypothalamic regions (but most likely the POA); GABA17 (*Slc6a3*) a dopaminergic neuron type with high expression of *Th* and *Prlr* suggesting its identity as TIDA neurons of the arcuate nucleus; GABA8 (*Vipr2*) with very high *Avp* and *Nms* expression suggesting these are neurons from the SCN; and GABA15 (*Agrp*) feeding promoting neurons of the Arcuate Nucleus. Within the Romanov et al. (2017) dataset, 3/62 subtypes had significant over-representation of imprinted genes (Table 7b; Supp

table 8b): GABA5 (*Calcr, Lhx1*) whose largest upregulated gene is *Slc6a3* and which may correspond to GABA17 from the Chen et al. (2017) dataset as 21 out of the 41 upregulated genes were shared, including the top 5 marker genes - *Calcr, Lhx1, Slc6a3, Asb4, 1500016L03Rik* (Appendix B for full comparison); GABA14 (*Agrp*/*Npy*) which are likely to be the Arcuate Nucleus feeding neuron reported in Chen et al. (2017); and a *Ghrh* neuronal type which was not over-represented in Chen et al. (2017)

Table 7a-b. Imprinted gene over-representation in Whole Hypothalamic Datasets: 7a. Chen et al. (2017) and 7b. Romanov et al. (2017). *Identity* – Neuronal Subpopulations for the cells used in analysis; *Up Reg* – number of upregulated genes -2FC limit used for Romanov only (total number of genes in the dataset in brackets); *IG* – number of imprinted genes upregulated - 2FC limit used for Romanov only (total number of IGs in the dataset in brackets); *ORA p* – *p* value from over representation analysis on groups with minimum 5% of total IGs; *ORA q* – Bonferroni corrected *p* value from ORA; *Mean FC IG* – mean fold change for upregulated imprinted genes; *Mean FC Rest* – mean fold change for all other upregulated genes; *Top Tissue IGs* – Sum of IGs with highest mean expression value for that identity group.

Identity	Up Reg	IG	ORA *p*	ORA *q*	Mean FC IG	Mean FC Rest	Top Tissue IGs
7a) **Whole Hypothalamus**	**(12,167)**	**(79)**	**Chen et al. *(33 neuron types)***				
GABA17 (*Slc6a3*)	823	18	3.96E-06	**9.50E-05**	2.73	3.28	6
GABA8 (*Vipr2*)	478	11	0.000247	**0.005935**	1.96	2.72	2
GABA13 (*Slc18a2, Gal*)	568	12	0.000279	**0.006687**	1.89	2.01	0
GABA15 (*Agrp*)	764	13	0.001190	**0.028553**	2.40	2.78	2
7b) **Whole Hypothalamus**	**(12,812)**	**(86)**	**Romanov et al. *(62 neuron types)***				
GABA 5 (*Calcr, Lhx1*)	41	6	2.86E-07	**2.00E-06**	3.47	5.10	1
GABA 14 (*Npy, Agrp*)	193	6	0.0018726	**0.013108**	2.94	6.12	2
Ghrh	235	6	0.004954	**0.034677**	18.31	17.19	21

3b) Specific Brain Nuclei/Neuron types

3b.1) Imprinted gene enrichment confirmed in specific hypothalamic nuclei

Arcuate Nucleus - ARC (Campbell et al., 2017)

The first nuclei investigated was the Arcuate Nucleus sequenced by Campbell et al. (2017). Imprinted gene over-representation was found in 8/24 arcuate neuron types (Table 8a; Supp Table 9). These included the major *Agrp*/*Npy* neuron type (*Agrp*/*Sst*) and both major *Pomc* neuron types known to be feeding suppressants (Rau & Hentges, 2017). Additional significant over-representation was found in the *Ghrh* neuron type, which was also enriched in Galanin and *Th*. Finally, a highly significant over-representation of IGs was found in the *Th*/*Slc6a3* neuron type identified by the authors as one of the most likely candidates for the TIDA dopaminergic neuron population. Marker genes for this identity

298 group overlapped with the TIDA candidate from the previous two datasets (e.g. *Slc6a3*, *Th*, *Lhx1*,
299 *Calcr*). *Agrp* neurons and these TIDA candidate neurons were identified in both whole hypothalamic
300 datasets and at the nuclei level validating the association.

301 Preoptic Area - POA (Moffitt et al., 2018)

302 Imprinted gene over-representation was found in 14/66 of the neuron clusters identified by Moffitt
303 et al. (2018) in the preoptic area (Supp Table 10a). Moffitt et al. (2018) stated that 9 of the 66 clusters
304 most likely originate from outside the POA including 3/14 of the imprinted gene enriched neuron
305 clusters (one of which was the *Avp*/*Nms* neurons (q = 0.0003) likely to be from the SCN). To focus our
306 analysis on the POA alone, we removed the non-POA neuron groups and ran the analysis on the
307 remaining neural groups restricted to POA (Table 8b; Supp Table 10b). Imprinted genes were found to
308 be over-represented in 15 neuron types including two Galanin enriched clusters (i22, i8), the major *Th*
309 enriched cluster (i17), e15:*Ucn3*/*Brs3* (a cluster activated by parenting), a *Crh* cluster (i35) and a *Trh*
310 cluster (e4), and both major *Ghrh* clusters (e13, e19). Since a considerable number of neuron types
311 were over-represented, the above analysis was run again with a 2-fold change cut off to find out if
312 imprinted genes were enriched in the more distinct markers (Supp Table 10c). Only two neuron types
313 were over-represented: i35:*Crh*/*Tac2* (q = 0.0126) and i16:*Gal*/*Th* (q = 0.0026). i16:*Gal*/*Th* is a
314 prominent behavioural neuron group (parenting and mating behaviour) which was not identified in
315 the previous analysis. Only three POA neuron groups are upregulated in *Slc18a2, Gal, Prlr, Th* and *Esr1*
316 (i8, i16, i38) like GABA13 in Chen et al. (2017) and all of them are represented above.

317 Lateral Hypothalamic Area- LHA (Mickelsen et al., 2019)

318 Over-representation of imprinted genes in the Lateral Hypothalamic Area neurons was analysed for
319 Glutamatergic and GABAergic neurons separately since they were sequenced separately by Mickelsen
320 et al. (2019). For Glutamatergic neurons, imprinted gene over-representation was found in 6/15
321 clusters (Table 8ci, Supp Table 11a). These included: Glut 6 (*Hcrt*) identified as the Orexin cluster, a
322 classic motivated behaviour neuronal subset identified in the MBA as HYPEP6; and Glut 9 (*Synpr*/*Gad1*)

with a significant GSEA (Appendix C2) and synaptoporin as a marker gene an; it is a cell type scattered across the LHA and DMH. The four others were Glut 14 (*Otp/Gpr101*), Glut 7 (*Gpr101/Tcf4*), Glut 5 (*Ebf3/Otp*) and Glut 11 (*Calca/Col27a1*) which were all enriched in similar genes such as *Otp, Tcf4* and *Gpr101*. The authors did not describe a clear functional role associated with these neurons, but their gene expression (namely *Gpr101* and *Otp*) likens these neurons to HYPEP1 and HYPEP3 from the MBA. For GABAergic neurons imprinted gene over-representation was found in 2/15 clusters: Gaba 1 (*Gal*/*Dlk1*), a galanin enriched cluster, previously shown to modulate food reward (Qualls-Creekmore et al., 2017) and Gaba 5 (*Meis2*/*Calb2*), a sub-type also lacking functional classification (Table 8cii; Supp Table 11b).

Suprachiasmatic Nucleus - SCN (Wen et al., 2020)

Analysis of the 10x chromium data of only SCN neurons (Table 8d, Supp Table 12) revealed a significant over-representation and GSEA (Appendix C3) in the *Avp*/*Nms* neuronal cluster only. This cluster shows the strongest expression for *Oxt, Avp*, *Avpr1a* and *Prlr* and is one of the three neural group that Wen et al. (2020) found had robust circadian gene expression, and the only subtype with notable phase differences in circadian gene expression in the dorsal SCN). This cluster may be the GABA8 cluster found enriched in the Chen et al. (2017) dataset and the *Avp*/*Nms* group identified by Moffitt et al. (2018) as an off-target neuronal group.

Table 8a-d. Imprinted gene over-representation in individual hypothalamic datasets: 8a. Arcuate Nucleus (ARC) – Campbell et al. (2017), 8b. Preoptic Area (POA) – Moffitt et al. (2018), 8c. Lateral Hypothalamic Area (LHA) – Mickelsen et al. (2019) and, 8d. Suprachiasmatic Nucleus (SCN) – Wen et al. (2020). *Identity* – Nervous System Regional identities for the cells used in analysis, total number of neuronal types in brackets; *Up Reg* – number of upregulated genes (total number of genes in the dataset in brackets); *IG* – number of imprinted genes upregulated (total number of IGs in the dataset in brackets); *ORA p* – *p* value from over representation analysis on groups with minimum 5% of total IGs; *ORA q* – Bonferroni corrected *p* value from ORA; *Mean FC IG* – mean fold change for upregulated imprinted genes; *Mean FC Rest* – mean fold change for all other upregulated genes; *GSEA p* – *p* value from Gene Set Enrichment Analysis for identity groups with 15+ IGs and Mean FC IG > Mean FC Rest; *GSEA q* – Bonferroni corrected *p* values from GSEA; *Top Tissue IGs* – Sum of IGs with highest mean expression value for that identity group.

Identity	Up Reg	IG	ORA *p*	ORA *q*	Mean FC IG	Mean FC Rest	GSEA *p*	GSEA *q*	Top Tissue IGs
a) ARC	**(16,098)**	**(94)**	***(24 neuron types)***						
Th/Slc6a3	1113	26	7.49E-10	**1.72E-08**	2.14	4.04	-	-	4
Arx/Nr5a2	599	16	3.98E-07	**9.16E-06**	1.33	3.15	-	-	3
Agrp/Sst	1704	23	0.000111	**0.002552**	1.61	1.78	-	-	1

Pomc/Anxa2	3162	34	0.000168	**0.003866**	1.70	1.82	-	-	1
Ghrh	1265	18	0.000382	**0.008793**	2.02	2.82	-	-	0
Gpr50	578	11	0.000591	**0.013595**	2.52	5.02	-	-	8
Pomc/Glipr1	1536	19	0.001426	**0.032792**	1.91	1.90	0.2851	0.8553	5
Slc17a6/Trhr	1547	19	0.001550	**0.035651**	3.22	2.35	0.1143	0.3429	1
b) POA	**(16,402)**	**(100)**	***(57 neuron types)***						
i22:*Gal/Pmaip1*	1569	27	4.64E-07	**2.41E-05**	2.02	2.14	-	-	0
i8:*Gal/Amigo2*	3625	42	6.29E-06	**0.000327**	1.78	1.53	0.1628	1	0
e15:*Ucn3/Brs3*	996	18	2.90E-05	**0.001510**	6.15	3.43	0.04	0.32	1
i12:Gaba	1229	20	4.40E-05	**0.002286**	1.76	1.68	0.2236	1	0
i35:*Crh/Tac2*	1464	22	5.69E-05	**0.002961**	1.98	2.44	-	-	3
i17:*Th/Nos1*	1358	21	5.76E-05	**0.002998**	2.02	2.00	0.4519	1	0
e4:*Trh/Angpt1*	2770	33	6.13E-05	**0.003188**	1.90	2.28	-	-	0
e19:*Ghrh/Trh*	1974	26	9.75E-05	**0.005068**	2.14	3.88	-	-	4
e1:Glut	1518	22	9.76E-05	**0.005073**	1.58	1.87	-	-	0
e12:*Nos1/Foxp2*	1635	22	0.000284	**0.014790**	1.81	2.64	-	-	2
i1:Gaba	602	12	0.000295	**0.015355**	1.20	1.55	-	-	0
i24:*Nmu*	3020	33	0.000334	**0.017349**	1.71	2.02	-	-	0
e13:*Ghrh/C1ql1*	3588	37	0.000398	**0.020691**	2.58	2.81	-	-	5
i37:*Bdnf/Chrm2*	2078	25	0.000561	**0.029168**	2.80	3.17	-	-	3
i38:*Kiss1/Th*	2078	25	0.000561	**0.029168**	2.09	5.11	-	-	1
ci) LHA GLUT	**(11,909)**	**(78)**	***(15 neuron types)***						
Glut_11 (*Calca, Col27a1*)	392	13	1.86E-06	**2.61E-05**	1.89	3.47	-	-	5
Glut_7 (*Gpr101, Tcf4*)	253	10	7.00E-06	**9.80E-05**	1.55	2.59	-	-	6
Glut_14 (*Otp/Gpr101*)	369	11	3.36E-05	**0.000470**	1.38	2.08	-	-	4
Glut_9 (*Synpr/Gad1*)	784	15	0.000196	**0.002743**	2.17	2.02	0.0099	**0.0198**	3
Glut_5 (*Ebf3/Otp*)	720	14	0.000280	**0.003915**	1.74	1.64	-	-	0
Glut_6 (*Hcrt*)	3097	35	0.000423	**0.005924**	3.15	2.29	0.1385	0.277	13
cii) LHA GABA	**(11,572)**	**(75)**	***(15 neuron types)***						
Gaba_1 (*Gal, Dlk1*)	1021	19	2.82E-05	**0.000395**	1.62	1.50	0.0601	0.1202	1
Gaba_5 (*Meis2, Calb2*)	646	12	0.001076	**0.015062**	1.26	1.82	-	-	2
d) SCN	**(11,391)**	**(77)**	***(5 neuron types)***						
Avp/Nms	909	24	3.01E-09	**1.51E-08**	3.12	1.94	**0.004**	-	17

3bii) Imprinted gene enrichment confirmed in serotonergic and dopaminergic neurons of the midbrain/hindbrain

Dopaminergic Neurons

In the MBA, Whole Hypothalamus and Arcuate Nucleus analyses, dopaminergic clusters were consistently enriched and, to further explore dopaminergic enrichment, analysis of Hook et al.'s (2018)

data allowed comparison for dopamine neurons across the brain (olfactory bulb, arcuate nucleus and midbrain) at two developmental timepoints (E15.5 and Post-natal day (P) 7). The arcuate nucleus P7 dopamine neurons emerged as the clearest over-represented subgroups (Table 9). This included the *Th*/*Slc6a3* neurons and the *Gal/Ghrh* cluster considered neuroendocrine cells by Hook et al. (2018). Additionally, P7 midbrain neurons were the other group with significant over-representation (PAG and VTA) as well as the neuroblasts at this time point.

Table 9. Imprinted gene over-representation in Brain-wide dopaminergic neurons (Hook et al., 2018). *Identity* – Cell identities for the cells used in analysis; *Up Reg* – number of upregulated genes (total number of genes in the dataset in brackets); *IG* – number of imprinted genes upregulated (total number of IGs in the dataset in brackets); *ORA p* – *p* value from over representation analysis on groups with minimum 5% of total IGs; *ORA q* – Bonferroni corrected *p* value from ORA; *Mean FC IG* – mean fold change for upregulated imprinted genes; *Mean FC Rest* – mean fold change for all other upregulated genes; *Top Tissue IGs* – Sum of IGs with highest mean expression value for that identity group.

Identity	Up Reg (13,095)	IG (86)	ORA *p*	ORA *q*	Mean FC IG	Mean FC Rest	Top Tissue IGs
Arcuate Nucleus (*Th, Slc6a3, Prlr*) (P7)	912	26	9.90E-11	**1.09E-09**	2.10	2.10	3
Arcuate Nucleus (*Th, Ghrh, Gal*) (P7)	552	14	1.50E-05	**0.000165**	3.34	4.05	9
Periaqueductal Gray (P7)	656	14	9.90E-05	**0.001089**	2.95	4.48	12
Midbrain Neuroblast (P7)	1092	18	0.000251	**0.002757**	2.36	2.81	13
Ventral Tegmental Area (P7)	1855	24	0.000775	**0.008526**	2.01	2.41	3
Substantia Nigra (P7)	2532	21	0.157969	1	2.37	2.59	7
Olfactory Bulb Mature *Th*+ (P7)	678	7	0.163674	1	3.93	4.33	4
Post-mitotic midbrain neuron (E15.5)	1362	11	0.292853	1	3.41	2.32	13
Forebrain neuroblast (E15.5)	3207	20	0.668062	1	1.92	1.97	8
Midbrain Neuroblast (E15.5)	5117	29	0.888038	1	1.94	2.63	6
Post-mitotic forebrain *Th*+ (E15.5)	1174	5	0.900062	1	2.05	2.76	3
Olfactory Bulb maturing *Th*+ (P7)	350	4	-	-	4.47	5.35	2
Olfactory Bulb immature *Th*+ (P7)	127	2	-	-	3.74	5.46	3

Although no detailed adult mouse midbrain datasets exist, La Manno et al.'s (2016) ventral midbrain sequencing at E11.5-E18.5 allowed us to identify imprinted enrichment within the midbrain at a timepoint when the major neuronal populations are differentiating but still identifiable (Table 10). As anticipated, we found significant over-representation in dopaminergic neurons, both mature (DA1; high *Th* and *Slc6a3*) and developing (DA0) as well as the serotonergic neurons likely from the midbrain raphe nuclei.

Table 10. Imprinted gene over-representation in Developing Midbrain cell types (La Manno et al., 2016). *Identity* – Nervous System Regional identities for the cells used in analysis; *Up Reg* – number of upregulated genes (total

number of genes in the dataset in brackets); *IG* – number of imprinted genes upregulated (total number of IGs in the dataset in brackets); *ORA p* – p value from over representation analysis on groups with minimum 5% of total IGs; *ORA q* – Bonferroni corrected p value from ORA; *Mean FC IG* – mean fold change for upregulated imprinted genes; *Mean FC Rest* – mean fold change for all other upregulated genes; *GSEA p* – p value from Gene Set Enrichment Analysis for identity groups with 15+ IGs and Mean FC IG > Mean FC Rest; *GSEA q* – Bonferroni corrected p values from GSEA; *Top Tissue IGs* – Sum of IGs with highest mean expression value for that identity group.

Identity	**Up Reg (14,027)**	**IG (93)**	**ORA *p***	**ORA *q***	**Mean FC IG**	**Mean FC Rest**	**GSEA *p***	**GSEA *q***	**Top Tissue IGs**
Serotonergic	1781	33	1.42E-08	**2.99E-07**	4.78	2.36	0.0272	0.0544	9
Pericytes	1409	22	0.000103	**0.002161**	8.57	8.04	0.2209	0.4418	12
Dopaminergic 1 (*Th/Slc6a3*)	347	9	0.000484	**0.010164**	3.17	2.84	-	-	4
Dopaminergic 0 (Immature)	219	7	0.000624	**0.013114**	4.03	2.91	-	-	3
Lateral Neuroblasts 2	2158	26	0.001353	**0.028403**	2.11	2.48	-	-	13
Oculomotor & trochlear nucleus	1994	22	0.009922	0.208356	1.83	2.28	-	-	5
Mediolateral Neuroblasts 3	707	10	0.018500	0.388509	2.25	2.27	-	-	1
GABAergic neurons 1b	943	12	0.021628	0.454193	1.84	2.03	-	-	2
GABAergic neurons 1a	743	10	0.025112	0.527346	2.05	2.12	-	-	2
Dopaminergic 2 (*Aldh1a1*)	464	7	0.034351	0.721369	3.22	3.43	-	-	2
Endothelial cell	2555	24	0.042792	0.898639	7.38	8.31	-	-	5
Mediolateral Neuroblasts 1	302	5	0.050300	1	1.76	2.24	-	-	1
GABAergic neurons 2	5711	46	0.053803	1	1.89	1.95	-	-	8
Radial Glia-like Cells 3	1142	9	0.344985	1	3.43	5.45	-	-	4
Red Nucleus	1021	8	0.365408	1	2.21	2.01	-	-	2
Mediolateral Neuroblasts 4	811	6	0.452000	1	1.54	2.42	-	-	0
Microglia	1293	7	0.765321	1	6.06	25.18	-	-	3
Radial Glia-like Cells 1	1309	7	0.776475	1	3.08	2.79	-	-	3
Ependymal	1391	7	0.827678	1	3.49	13.28	-	-	7
Neuronal Progenitor	1775	9	0.847792	1	2.32	2.76	-	-	2
Radial Glia-like Cells 2	1726	8	0.899405	1	2.89	3.41	-	-	1
Lateral Neuroblasts 1	529	4	-	-	1.76	2.93	-	-	3
Mediolateral Neuroblasts 5	125	4	-	-	1.86	1.92	-	-	1
Neuroblast Dopaminergic	196	3	-	-	3.55	2.14	-	-	0
Medial Neuroblast	313	2	-	-	1.48	2.34	-	-	0
Mediolateral Neuroblasts 2	206	1	-	-	1.13	1.82	-	-	0

As a final investigation, we used Tiklová et al.'s (2019) data in which purely dopamine neurons (*Pitx3*+) of the ventral midbrain were sequenced. 3/7 dopaminergic subpopulations identified in this study had significant over-representation of imprinted genes (Table 11). The *Aldh1a1*+ neurons were localised

to the SNc, *Slc6a3*+ neurons to the VTA and *Vip*+ neurons to the PAG. The seven subpopulations were found to belong to two major subtypes of neurons by Tiklová et al. (2019), Dat^{High} neurons with high expression of *Th* and *Slc6a3* that made up 89% of *Pitx3* positive neurons identified in the ventral midbrain, and Dat^{Low} neurons, the rarer *Pitx3* positive neurons lacking high expression of the typical dopamine markers (*Th/Slc6a3*). The 3 neuronal subpopulations enriched here formed the entirety of the Dat^{High} subtype.

Table 11. Imprinted gene over-representation in Midbrain Dopamine Neuron Subtypes (Tiklová et al., 2019). *Identity* – Cell identities for the cells used in analysis; *Up Reg* – number of upregulated genes (total number of genes in the dataset in brackets); *IG* – number of imprinted genes upregulated (total number of IGs in the dataset in brackets); *ORA p* – *p* value from over representation analysis on groups with minimum 5% of total IGs; *ORA q* – Bonferroni corrected *p* value from ORA; *Mean FC IG* – mean fold change for upregulated imprinted genes; *Mean FC Rest* – mean fold change for all other upregulated genes; *Top Tissue IGs* – Sum of IGs with highest mean expression value for that identity group.

Identity	Up Reg (24,079)	IG (104)	ORA *p*	ORA *q*	Mean FC IG	Mean FC Rest	Top Tissue IGs
***Vip*+**	3534	29	0.000706	**0.004942**	3.45	61.85	32
***Aldh1a1*+**	1215	14	0.001069	**0.007481**	2.99	26.25	13
***Slc6a3*+**	1571	16	0.001696	**0.011869**	2.06	32.78	11
Gad2+Th	3961	24	0.071815	0.502704	2.01	17.58	22
Nxph4+Th	4902	27	0.140455	0.983185	1.87	15.69	11
Gad2+	1105	7	0.227457	1	1.46	101.14	8
Nxph4+	2148	6	0.927658	1	1.65	76.49	7

Serotonergic Neurons

To examine enrichment of imprinted genes in serotonergic neurons further, analysis was performed on data for the Raphe nuclei of the midbrain/hindbrain which are the key serotonergic regions of the brain. Analysis of Huang et al.'s (2019) sequencing of the Dorsal Raphe Nucleus (DRN) revealed a clear enrichment of imprinted genes in the neuronal populations of the DRN as compared to the other non-neuronal cell populations of the brainstem (Supp Table 13a). When compared to all other cell populations, significant GSEA after Bonferroni correction was found for all five neuronal populations (Appendix C4). When compared against each other (i.e. serotonergic upregulation vs. the other neurons), only the serotonergic neurons of the DRN was found to have a significant over-representation of imprinted genes (Table 12). Over-representation in the dopaminergic neurons was significant but not after correction. GSEA was also non-significant but the mean fold change for

imprinted genes was markedly higher in both serotoninergic (52% higher) and dopaminergic neurons (68% higher).

Table 12. Imprinted gene over-representation in Dorsal Raphe Neuron types (Huang et al., 2019).
Identity – Nervous System Regional identities for the cells used in analysis; *Up Reg* – number of upregulated genes (total number of genes in the dataset in brackets); *IG* – number of imprinted genes upregulated (total number of IGs in the dataset in brackets); *ORA p* – *p* value from over representation analysis on groups with minimum 5% of total IGs; *ORA q* – Bonferroni corrected *p* value from ORA; *Mean FC IG* – mean fold change for upregulated imprinted genes; *Mean FC Rest* – mean fold change for all other upregulated genes; *GSEA p* – *p* value from Gene Set Enrichment Analysis for identity groups with 15+ IGs and Mean FC IG > Mean FC Rest; *GSEA q* – Bonferroni corrected *p* values from GSEA; *Top Tissue IGs* – Sum of IGs with highest mean expression value for that identity group.

Identity	Up Reg (14,397)	IG (85)	ORA *p*	ORA *q*	Mean FC IG	Mean FC Rest	GSEA *p*	GSEA *q*	Top Tissue IGs
Serotonergic Neurons	1275	18	0.000389	**0.001943**	4.03	2.66	0.1601	0.3202	15
Dopaminergic Neurons	1520	15	0.031646	0.158232	4.93	2.93	0.1045	0.209	14
Glutamatergic Neurons	3459	28	0.039052	0.195259	1.67	1.82	-	-	22
Peptidergic Neurons	1298	11	0.141362	0.706809	1.83	6.42	-	-	17
GABAergic Neurons	2796	19	0.285350	1	1.34	1.76	-	-	17

The analysis was further broken down into neural subpopulations of all neuron types (Supp Table 13b). 2/5 serotonin subpopulations had significant over-representation of imprinted genes: *Hcrtr1/Asb4* – *q* = 0.0014 and *Prkcq/Trh* – *q* = 0.007. These clusters were identified by Huang et al. (2019) as the only clusters localised in the dorsal/lateral DRN and the serotonin clusters enriched in *Trh*. Huang et al. (2019) hypothesised that these were the serotonin neurons that project to hypothalamic nuclei, and motor nuclei in the brainstem (as opposed to cortical/striatal projection).

Ren et al. (2019) sequenced serotonin neurons across the midbrain and pons identifying thirteen subpopulations including six principle dorsal raphe population (DR), one caudal raphe population (cDR) and four median raphe populations (MR). Imprinted gene over-representation was found across numerous serotonergic subtypes including 4/6 DR populations and 2/4 MR populations (Table 13). Since the majority of subtypes were considered over-represented, the analysis was run again with the 2-fold change upregulation cut off to find out if imprinted genes were enriched in the more distinct markers (Supp Table 14.). The only significantly over-represented subpopulation in this more conservative analysis was DR-1 (*q* = 0.005), a dorsal/lateral DR population with the largest enrichment

of *Trh* (the marker for serotonin neurons that project to subcortical areas – namely the hypothalamus and thalamus).

Table 13. Imprinted gene over-representation in Raphe Serotonin neuron sub-types (Ren et al., 2019). ***Identity*** – Nervous System Regional identities for the cells used in analysis; *Up Reg* – number of upregulated genes (total number of genes in the dataset in brackets); *IG* – number of imprinted genes upregulated (total number of IGs in the dataset in brackets); *ORA p* – *p* value from over representation analysis on groups with minimum 5% of total IGs; *ORA q* – Bonferroni corrected *p* value from ORA; *Mean FC IG* – mean fold change for upregulated imprinted genes; *Mean FC Rest* – mean fold change for all other upregulated genes; *GSEA p* – *p* value from Gene Set Enrichment Analysis for identity groups with 15+ IGs and Mean FC IG > Mean FC Rest; *GSEA q* – Bonferroni corrected *p* values from GSEA; *Top Tissue IGs* – Sum of IGs with highest mean expression value for that identity group.

Identity	Up Reg (18,919)	IG (96)	ORA *p*	ORA *q*	Mean FC IG	Mean FC Rest	GSEA *p*	GSEA *q*	Top Tissue IGs
DR-6	1197	26	1.46E-10	**1.46E-09**	1.54	1.68	-	-	4
DR-1	1199	18	2.91E-05	**0.000291**	2.15	2.07	0.0472	0.1416	17
MR-2	1925	21	0.000555	**0.005554**	1.62	1.6	0.4339	1	8
MR-1	2002	21	0.000927	**0.00927**	1.44	1.55	-	-	4
DR-5	1826	19	0.001862	**0.018619**	1.16	1.66	-	-	5
DR-2	3908	31	0.004993	**0.049925**	1.62	1.8	-	-	10
cDR	1339	12	0.038173	0.381725	3.12	2.28	-	-	18
DR-3	2577	19	0.058054	0.580543	2.03	1.92	0.1602	0.4806	8
MR-3	1238	8	0.29177	1	1.18	2.03	-	-	7
MR-4	3475	15	0.793673	1	1.82	2.23	-	-	11
DR-4	1270	3	-	-	1.04	2.78	-	-	4

Discussion

Previous studies have suggested that imprinted genes are preferentially expressed in the adult brain relative to other adult tissues (Andergassen et al., 2017). The current investigation applied a systems biology approach to examine the enrichment of imprinted genes at the level of the brain in comparison to other adult tissues, refining this analysis to specific brain regions and then to single cell populations. By analysing two independent adult tissue datasets, we discovered a significant over-representation of imprinted genes in the brain. Furthermore, imprinted genes were over-represented in neurons at every level tested with marked enrichment in the neuroendocrine cells lineages. Within-brain analyses revealed that the hypothalamus and the monoaminergic system of the midbrain/hindbrain were foci for imprinted gene enrichment. While not all imprinted genes follow these patterns of expression, this over-representation analysis highlights collective gene expression which is non-random in nature. As such, these analyses identify ‘expression hotspots’, which in turn

407 suggest 'functional hotspots'. Specifically, our findings at the systems and cellular level highlight a
408 major role for imprinted genes in the neuronal regulation of feeding, parental behaviour and sleep.

409 Some of the earliest studies of genomic imprinting identified the brain as a key area for imprinted
410 gene expression (Allen et al., 1995; Keverne et al., 1996). However, it is estimated that ~80% of the
411 genome is expressed in the brain and consequently imprinted gene expression in the brain might not
412 be a purposeful phenomenon. Our current analysis revealed that imprinted genes were statistically
413 significantly over-represented in the brain as a whole. This over-representation was mostly driven by
414 paternally expressed genes (PEGs). Within specific brain regions, imprinted genes were over-
415 represented in the hypothalamus, ventral midbrain, pons and medulla. Our formal analysis
416 strengthens the idea that these brain regions are hotspots of imprinted gene expression, as suggested
417 in earlier summaries of imprinted gene expression (Gregg et al., 2010) and the Pg/Gg and Ag chimera
418 studies (Allen et al., 1995; Keverne et al., 1996). In these chimera studies, Pg/Gg cells with two
419 maternal genomes preferentially allocated to the developing adult cortex and hippocampus, and Ag
420 cells with two paternal genomes preferentially allocated to the developing hypothalamus and
421 midbrain. In our analyses, PEGs appeared to be driving regional enrichment, with MEGs only over-
422 represented in the pons and medulla with no specific enrichment in cortex or hippocampus. This
423 difference between our analysis of enrichment and the Pg/Gg and Ag chimeras studies indicate that
424 the distribution of Pg/Gg and Ag cells in the brain is not driven by, or indeed a reflected of adult PEG
425 and MEG expression but instead is determined by developmental expression of key imprinted genes.

426 By taking advantage of scRNA-seq data we were able to examine imprinted gene enrichment in
427 specific cell-types. At the whole brain level mature neurons and, in particular, neural-lineage
428 neuroendocrine cells had disproportionately higher numbers of imprinted genes expressed and high
429 levels of imprinted gene expression. Neuroendocrine cells are uniquely distinguished by the
430 expression of *Baiap3,* the synaptic regulator of neurotransmitter exocytosis, and higher expression of
431 *Oxt*, *Avp* and other key hypothalamic genes including *Agrp, Pomc, Gal* and *Sst*. In the Mouse Brain

432 Atlas, *Baiap3* was highly expressed in the peptedergic neurons of the hypothalamus (and a top 5
433 marker gene for Hypep2 & Hypep3). It is likely that this neural-lineage neuroendocrine population
434 comprises members of the key hypothalamic populations in which imprinted genes are enriched (see
435 below) and, when treated as their own cluster, demonstrate strong imprinted gene enrichment
436 compared to other cell lineages of the brain, even other mature neurons.

437 Within the hypothalamus, a selection of informative neuronal subpopulations were over-represented.
438 Strikingly, and suggestive of meaningful enrichment, we saw convergence across our different levels
439 of analysis with many of the key neuronal types identified in the whole hypothalamus and/or
440 hypothalamic-region-level analysis, already having been identified against the background of general
441 imprinted gene expression in the whole-brain-level analysis. The first key population were the
442 *Avp/Nms* SCN population identified at the whole brain level, the whole hypothalamic level and as the
443 only enriched neuronal type within the SCN. This population is of interest given the growing
444 appreciation of the role IGs play in circadian processes and the SCN suggested by studies of individual
445 imprinted genes (Tucci, 2016). The *Agrp* neurons were identified at the whole brain level, in both
446 whole hypothalamic level datasets and within the Arcuate nucleus. These neurons have already been
447 associated with individual imprinted genes (Cassidy & Charalambous, 2018) and play a crucial role in
448 promoting feeding behaviour (Aponte, Atasoy, & Sternson, 2011). The *Th/Slc6a3/Prlr* neuron type,
449 proposed to represent the dopamine neurons in the arcuate nucleus regulating prolactin, was again
450 identified in both hypothalamic datasets as the top hit, in the arcuate nucleus as the top hit and as the
451 top hit in the dopamine neurons across the brain. Prolactin is a well-known as a mediator of maternal-
452 infant interaction priming the maternal brain during pregnancy for the nurturing of pups (Grattan,
453 Steyn, Kokay, Anderson, & Bunn, 2008), and modulating prolactin is well within the imprinted gene
454 phenotype remit. Additionally, the *Gal/Esr1* neuronal population, identified as Deinh5 at the whole
455 brain level, as a population in Chen et al.'s (2017) whole hypothalamic analysis and again as a few
456 populations (i8 and i16) in the POA. This neural population is known to be associated with feeding and
457 parenting behaviour (Moffitt et al., 2018). *Ghrh* neuron clusters were isolated at the whole-brain level

458 and whole-hypothalamic level and was identified in the POA/ARC but significantly the dopaminergic
459 *Ghrh/Gal* cluster was identified in both the dopamine-wide and arcuate nucleus datasets. Finally,
460 other neural populations were identified which have previously been associated with imprinted genes
461 including the *Hcrt* population of the LHA (Seifinejad et al., 2019) identified at the whole brain and LHA
462 level, and the *Pomc* neurons of the ARC (Cassidy & Charalambous, 2018) identified at the arcuate
463 nucleus level.

464 In this analysis the hypothalamus was a clear hot spot for imprinted gene expression, in line with the
465 prevailing view of imprinted gene and hypothalamic function (Ivanova & Kelsey, 2011; Pulix & Plagge,
466 2020). However, outside of the hypothalamus other distinct hotspot emerged from our whole brain
467 analysis including the monoaminergic system of the midbrain/hindbrain. Analysing the dorsal raphe
468 nucleus and the ventral midbrain revealed the serotonergic neurons to be a foci of imprinted gene
469 expression within this region; more specifically the serotonergic neurons that project to the
470 subcortical regions of the brain known to be associated with feeding and other motivated behaviours
471 (Donovan & Tecott, 2013). These neurons are recognised by higher expression of *Trh*. Imprinted genes
472 were over-represented in specific serotonergic neuron groups from the DRN in both datasets (Huang
473 et al., 2019; Ren et al., 2019) that were all enriched for *Trh*. The caudal population of serotonin
474 neurons (HBSER4 in MBA) could not be investigated further due to lack of single cell RNA-seq data. In
475 contrast to the specificity of the imprinted gene enrichment in serotonergic neurons, the enrichment
476 seen for adult midbrain dopamine neurons was more generalised. Midbrain dopamine neurons were
477 enriched in adult whole brain and developmental ventral midbrain analyses, and sub-type specific
478 enrichment centred on the DAT^{high} lineage of midbrain DA neurons. These DAT^{high} neurons amount to
479 the standard dopamine neurons (*Th/Slc6a3* high) and are found throughout the VTA, SNc and PAG
480 (Tiklová et al., 2019) leading to little subdivision of the dopaminergic signal.

481 Analyses of these kind are always bound by the available data and therefore there are notable
482 limitations and caveats to this study. The aim of this study was to generate information about

483 ‘hotspots’ of imprinted gene expression. This approach, and the use over-representation analysis and
484 GSEA, therefore do not provide an exhaustive list of sites of expression. For example, oxytocin neurons
485 of the hypothalamus were not significantly enriched in any of the hypothalamic datasets despite a
486 clear oxytocin neuron phenotype reported with imprinted gene models (Dombret et al., 2012; Li et
487 al., 1999). This may be an example of a functional effect occurring below the level of over-
488 representation, or that imprinted genes act during development and are not functionally enriched in
489 adult oxytocin neurons, or simply that compared to other hypothalamic neuronal populations,
490 oxytocin neurons are not a ‘hotspot’ of imprinted expression. Specific sequencing of oxytocinergic
491 brain regions will be required to distinguish between these possibilities. A second caveat is that we
492 did not assess parent-of-origin expression for the 119 imprinted genes we included in the analysis.
493 Previous expression profiling of imprinted genes have also not measured the POEs (Gregg et al., 2010;
494 Steinhoff, Paulsen, Kielbasa, Walter, & Vingron, 2009) but have restricted their gene selection to genes
495 with reliable imprinting status. Consequently, we only included the canonical imprinted genes and
496 genes with more than one demonstration of a POE when looking for enrichment. Furthermore, for the
497 vast majority of these genes, a brain-based POE effect has also already been reported (Supp Table 1).
498 Although this does not replace validating the imprinting status of all 119 in the tissues and subregions
499 examined, it does provide justification for looking at imprinted gene over-representation.

500 By exploiting scRNA-seq data we have asked whether imprinted genes as a group are
501 disproportionately represented in the brain, in specific brain regions, and in certain neuronal cell-
502 types. In the adult brain imprinted genes were over-represented in neurons, and particularly the
503 hypothalamic neuroendocrine populations and the monoaminergic hindbrain neurons, with the
504 serotonergic neurons demonstrating the clearest signal. Interestingly, PEGs appear to be the major
505 component of these enrichments, particularly the hypothalamic signal. By extension, these data also
506 identify behaviours that are foci for the action of imprinted genes. Although there are high profile
507 examples of individual imprinted genes that have a role in feeding (*Magel2*) (Schaller et al., 2010),
508 parental behaviour (*Peg3)* (Li et al., 1999) and sleep (*Snord116*) (Lassi et al., 2016), our analyses

509 indicate that imprinted genes *as a group* are strongly linked to these behaviours and also identify
510 other individual genes that should be explored in these domains. Conversely, there are high-profile
511 examples of imprinted genes involved in hippocampus related learning and memory (*Ube3a*) (Jiang et
512 al., 1998), but we did not find enrichment for cell types related to this brain function. The idea that
513 imprinted genes converge of specific physiological or behavioural processes is not unprecedented.
514 Specialisation of function is predicted when considering why genomic imprinting evolved at all
515 (Keverne, 2014; Keverne, Martel, & Nevison, 1996; Moore & Haig, 1991; Trivers & Burt, 1999).
516 Moreover, there is increasing evidence that the imprinted genes themselves appear to be co-
517 expressed in an imprinted gene network (IGN) and have confirmed regulatory links between each
518 other (Al Adhami et al., 2015; Gabory et al., 2009; Varrault et al., 2006). The idea of the IGN or, at the
519 very least, heavily correlated and coordinated expression between imprinted genes adds further
520 support to the idea that imprinted genes work in concert to influence processes, rather than in
521 isolation, and that perturbating one may influence many others (Patten, Cowley, Oakey, & Feil, 2016).
522 Our findings add substance to these general ideas and highlight the neuronal regulation of feeding,
523 parental behaviour and sleep as being key functional hotspots on which imprinted genes converge
524 and indicate possible evolutionary drivers of genomic imprinting in the brain.

525 ## Acknowledgments

526 This work was supported by a Wellcome Trust PhD studentship (220090/Z/20/Z).

527 ## Competing interests

528 All authors declare no financial and non-financial competing interests

529

530

531

References

Al Adhami, H., Evano, B., Le Digarcher, A., Gueydan, C., Dubois, E., Parrinello, H., … Journot, L. (2015). A systems-level approach to parental genomic imprinting: The imprinted gene network includes extracellular matrix genes and regulates cell cycle exit and differentiation. *Genome Research*, *25*(3), 353–367. https://doi.org/10.1101/gr.175919.114

Allen, N. D., Logan, K., Lally, G., Drage, D. J., Norris, M. L., & Keyerne, E. B. (1995). Distribution of parthenogenetic cells in the mouse brain and their influence on brain development and behavior. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(23), 10782–10786. https://doi.org/10.1073/pnas.92.23.10782

Andergassen, D., Dotter, C. P., Wenzel, D., Sigl, V., Bammer, P. C., Muckenhuber, M., … Hudson, Q. J. (2017). Mapping the mouse Allelome reveals tissue-specific regulation of allelic expression. *ELife*, *6*. https://doi.org/10.7554/eLife.25125

Angulo, M. A., Butler, M. G., & Cataletto, M. E. (2015). Prader-Willi syndrome: A review of clinical, genetic, and endocrine findings. *Journal of Endocrinological Investigation*, *38*(12), 1249–1263. https://doi.org/10.1007/s40618-015-0312-9

Aponte, Y., Atasoy, D., & Sternson, S. M. (2011). AGRP neurons are sufficient to orchestrate feeding behavior rapidly and without training. *Nature Neuroscience*, *14*(3), 351–355. https://doi.org/10.1038/nn.2739

Babak, T., DeVeale, B., Tsang, E. K., Zhou, Y., Li, X., Smith, K. S., … Fraser, H. B. (2015). Genetic conflict reflected in tissue-specific maps of genomic imprinting in human and mouse. *Nature Genetics*, *47*(5), 544–549. https://doi.org/10.1038/ng.3274

Campbell, J. N., Macosko, E. Z., Fenselau, H., Pers, T. H., Lyubetskaya, A., Tenen, D., … Tsai, L. T. (2017). A molecular census of arcuate hypothalamus and median eminence cell types. *Nature Neuroscience*, *20*(3), 484–496. https://doi.org/10.1038/nn.4495

Cassidy, F. C., & Charalambous, M. (2018). Genomic imprinting, growth and maternal–fetal interactions. *The Journal of Experimental Biology*, *221*(Suppl 1), jeb164517. https://doi.org/10.1242/jeb.164517

Cattanach, B. M., & Kirk, M. (1985). Differential activity of maternally and paternally derived chromosome regions in mice. *Nature*, *315*(6019), 496–498. https://doi.org/10.1038/315496a0

Chen, R., Wu, X., Jiang, L., & Zhang, Y. (2017). Single-Cell RNA-Seq Reveals Hypothalamic Cell Diversity. *Cell Reports*, *18*(13), 3227–3241. https://doi.org/10.1016/j.celrep.2017.03.004

Davies, J. R., Humby, T., Dwyer, D. M., Garfield, A. S., Furby, H., Wilkinson, L. S., … Isles, A. R. (2015). Calorie seeking, but not hedonic response, contributes to hyperphagia in a mouse model for Prader-Willi syndrome. *European Journal of Neuroscience*, *42*(4), 2105–2113. https://doi.org/10.1111/ejn.12972

Dent, C. L., Humby, T., Lewis, K., Ward, A., Fischer-Colbrie, R., Wilkinson, L. S., … Isles, A. R. (2018). Impulsive choice in mice lacking paternal expression of Grb10 suggests intragenomic conflict in behavior. *Genetics*, *209*(1), 233–239. https://doi.org/10.1534/genetics.118.300898

DeVeale, B., van der Kooy, D., & Babak, T. (2012). Critical Evaluation of Imprinted Gene Expression by RNA–Seq: A New Perspective. *PLoS Genetics*, *8*(3), e1002600. https://doi.org/10.1371/journal.pgen.1002600

Dombret, C., Nguyen, T., Schakman, O., Michaud, J. L., Hardin-Pouzet, H., Bertrand, M. J. M., & De Backer, O. (2012). Loss of Maged1 results in obesity, deficits of social interactions, impaired sexual behavior and severe alteration of mature oxytocin production in the hypothalamus. *Human Molecular Genetics*, *21*(21), 4703–4717. https://doi.org/10.1093/hmg/dds310

Donovan, M. H., & Tecott, L. H. (2013). Serotonin and the regulation of mammalian energy balance. *Frontiers in Neuroscience*, *7*. https://doi.org/10.3389/fnins.2013.00036

Ferguson-Smith, A. C. (2011). Genomic imprinting: The emergence of an epigenetic paradigm. *Nature Reviews Genetics*, *12*(8), 565–575. https://doi.org/10.1038/nrg3032

Gabory, A., Ripoche, M. A., Le Digarcher, A., Watrin, F., Ziyyat, A., Forné, T., … Dandolo, L. (2009). H19 acts as a trans regulator of the imprinted gene network controlling growth in mice. *Development*, *136*(20), 3413–3421. https://doi.org/10.1242/dev.036061

Garfield, A. S., Cowley, M., Smith, F. M., Moorwood, K., Stewart-cox, J. E., Gilroy, K., … Ward, A. (2011). Distinct physiological and behavioural functions for parental alleles of imprinted Grb10. *Nature*, *469*(7331), 534–538. https://doi.org/10.1038/nature09651

Grattan, D. R., Steyn, F. J., Kokay, I. C., Anderson, G. M., & Bunn, S. J. (2008). Pregnancy-induced adaptation in the neuroendocrine control of prolactin secretion. *Journal of Neuroendocrinology*, *20*(4), 497–507. https://doi.org/10.1111/j.1365-2826.2008.01661.x

Gregg, C., Zhang, J., Weissbourd, B., Luo, S., Schroth, G. P., Haig, D., & Dulac, C. (2010). High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *Science*, *329*(5992), 643–648. https://doi.org/10.1126/science.1190830

Han, X., Wang, R., Zhou, Y., Fei, L., Sun, H., Lai, S., … Guo, G. (2018). Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell*, *172*(5), 1091-1107.e17. https://doi.org/10.1016/j.cell.2018.02.001

Hook, P. W., McClymont, S. A., Cannon, G. H., Law, W. D., Morton, A. J., Goff, L. A., & McCallion, A. S. (2018). Single-Cell RNA-Seq of Mouse Dopaminergic Neurons Informs Candidate Gene Selection for Sporadic Parkinson Disease. *American Journal of Human Genetics*, *102*(3), 427–446. https://doi.org/10.1016/j.ajhg.2018.02.001

Huang, K. W., Ochandarena, N. E., Philson, A. C., Hyun, M., Birnbaum, J. E., Cicconet, M., & Sabatini, B. L. (2019). Molecular and anatomical organization of the dorsal raphe nucleus. *ELife*, *8*. https://doi.org/10.7554/eLife.46464

Ivanova, E., & Kelsey, G. (2011). Imprinted genes and hypothalamic function. *Journal of Molecular Endocrinology*, *47*(2), R67–R74. https://doi.org/10.1530/jme-11-0065

Jiang, Y. hui, Armstrong, D., Albrecht, U., Atkins, C. M., Noebels, J. L., Eichele, G., … Beaudet, A. L. (1998). Mutation of the Angelman ubiquitin ligase in mice causes increased cytoplasmic p53 and deficits of contextual learning and long-term potentiation. *Neuron*, *21*(4), 799–811. https://doi.org/10.1016/S0896-6273(00)80596-6

Keverne, E. B. (2014). Significance of epigenetics for understanding brain development, brain evolution and behaviour. *Neuroscience*, *264*, 207–217. https://doi.org/10.1016/j.neuroscience.2012.11.030

Keverne, E. B., Martel, F. L., & Nevison, C. M. (1996). Primate brain evolution: Genetic and functional considerations. *Proceedings of the Royal Society B: Biological Sciences*, *263*(1371), 689–696. https://doi.org/10.1098/rspb.1996.0103

Keverne, Eric B., Fundele, R., Narasimha, M., Barton, S. C., & Surani, M. A. (1996). Genomic imprinting and the differential roles of parental genomes in brain development. *Developmental Brain Research*, *92*(1), 91–100. https://doi.org/10.1016/0165-3806(95)00209-X

La Manno, G., Gyllborg, D., Codeluppi, S., Nishimura, K., Salto, C., Zeisel, A., … Linnarsson, S. (2016). Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. *Cell*, *167*(2), 566-580.e19. https://doi.org/10.1016/j.cell.2016.09.027

Lassi, G., Ball, S. T., Maggi, S., Colonna, G., Nieus, T., Cero, C., … Tucci, V. (2012). Loss of Gnas imprinting differentially affects REM/NREM sleep and cognition in mice. *PLoS Genetics*, *8*(5), e1002706. https://doi.org/10.1371/journal.pgen.1002706

Lassi, G., Priano, L., Maggi, S., Garcia-Garcia, C., Balzani, E., El-Assawy, N., … Tucci, V. (2016). Deletion of the Snord116/SNORD116 Alters Sleep in Mice and Patients with Prader-Willi Syndrome . *Sleep*, *39*(3), 637–644. https://doi.org/10.5665/sleep.5542

Lein, E. S., Hawrylycz, M. J., Ao, N., Ayres, M., Bensinger, A., Bernard, A., … Jones, A. R. (2007). Genome-wide atlas of gene expression in the adult mouse brain. *Nature*, *445*(7124), 168–176. https://doi.org/10.1038/nature05453

Li, L., Keverne, E. B., Aparicio, S. ., Ishino, F., Barton, S. ., & Surani, M. . (1999). Regulation of Maternal Behavior and Offspring Growth by Paternally Expressed Peg3. *Science*, *284*(5412), 330–334. https://doi.org/10.1126/science.284.5412.330

McNamara, G. I., John, R. M., & Isles, A. R. (2018). Territorial behavior and social stability in the mouse require correct expression of imprinted cdkn1c. *Frontiers in Behavioral Neuroscience*, *12*, 28. https://doi.org/10.3389/fnbeh.2018.00028

Mickelsen, L. E., Bolisetty, M., Chimileski, B. R., Fujita, A., Beltrami, E. J., Costanzo, J. T., … Jackson, A. C. (2019). Single-cell transcriptomic analysis of the lateral hypothalamic area reveals molecularly distinct populations of inhibitory and excitatory neurons. *Nature Neuroscience*, *22*(4), 642–656. https://doi.org/10.1038/s41593-019-0349-8

Moffitt, J. R., Bambah-Mukku, D., Eichhorn, S. W., Vaughn, E., Shekhar, K., Perez, J. D., … Zhuang, X. (2018). Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science*, *362*(6416). https://doi.org/10.1126/science.aau5324

Moore, T., & Haig, D. (1991). Genomic imprinting in mammalian development: a parental tug-of-war. *Trends in Genetics*, *7*(2), 45–49. https://doi.org/10.1016/0168-9525(91)90230-N

Negi, S. K., & Guda, C. (2017). Global gene expression profiling of healthy human brain and its application in studying neurological disorders. *Scientific Reports*, *7*(1), 1–12. https://doi.org/10.1038/s41598-017-00952-9

Nicholls, R. D., Knoll, J. H. M., Butler, M. G., Karam, S., & Lalande, M. (1989). Genetic imprinting suggested by maternal heterodisomy in non-deletion Prader-Willi syndrome. *Nature*, *342*(6247), 281–285. https://doi.org/10.1038/342281a0

Orr, H. A. (1995). Somatic mutation favors the evolution of diploidy. *Genetics*, *139*(3), 1441–1447. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7768451

Patten, M. M., Cowley, M., Oakey, R. J., & Feil, R. (2016). Regulatory links between imprinted genes: Evolutionary predictions and consequences. *Proceedings of the Royal Society B: Biological Sciences*. https://doi.org/10.1098/rspb.2015.2760

Perez, J. D., Rubinstein, N. D., & Dulac, C. (2016). *New Perspectives on Genomic Imprinting, an Essential and Multifaceted Mode of Epigenetic Control in the Developing and Adult Brain*. *Annual Review of Neuroscience* (Vol. 39). https://doi.org/10.1146/annurev-neuro-061010-113708

Perez, J. D., Rubinstein, N. D., Fernandez, D. E., Santoro, S. W., Needleman, L. A., Ho-Shing, O., … Dulac, C. (2015). Quantitative and functional interrogation of parent-of-origin allelic expression biases in the brain. *ELife*, *4*, 1–41. https://doi.org/10.7554/elife.07860

Peters, J. (2014, June 24). The role of genomic imprinting in biology and disease: An expanding view. *Nature Reviews Genetics*. Nature Publishing Group. https://doi.org/10.1038/nrg3766

Pulix, M., & Plagge, A. (2020). Imprinted Genes and Hypothalamic Function (pp. 265–294). Springer, Cham. https://doi.org/10.1007/978-3-030-40002-6_10

Qualls-Creekmore, E., Yu, S., Francois, M., Hoang, J., Huesing, C., Bruce-Keller, A., … Münzberg, H. (2017). Galanin-expressing GABA neurons in the lateral hypothalamus modulate food reward and noncompulsive locomotion. *Journal of Neuroscience*, *37*(25), 6053–6065. https://doi.org/10.1523/JNEUROSCI.0155-17.2017

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Rau, A. R., & Hentges, S. T. (2017). The relevance of AgRP neuron-derived GABA inputs to POMC neurons differs for spontaneous and evoked release. *Journal of Neuroscience*, *37*(31), 7362–7372. https://doi.org/10.1523/JNEUROSCI.0647-17.2017

Relkovic, D., Doe, C. M., Humby, T., Johnstone, K. A., Resnick, J. L., Holland, A. J., … Isles, A. R. (2010). Behavioural and cognitive abnormalities in an imprinting centre deletion mouse model for Prader-Willi syndrome. *European Journal of Neuroscience*, *31*(1), 156–164. https://doi.org/10.1111/j.1460-9568.2009.07048.x

Ren, J., Isakova, A., Friedmann, D., Zeng, J., Grutzner, S. M., Pun, A., … Luo, L. (2019). Single-cell transcriptomes and whole-brain projections of serotonin neurons in the mouse dorsal and median raphe nuclei. *ELife*, *8*. https://doi.org/10.7554/eLife.49424

Romanov, R. A., Zeisel, A., Bakker, J., Girach, F., Hellysaz, A., Tomer, R., … Harkany, T. (2017). Molecular interrogation of hypothalamic organization reveals distinct dopamine neuronal subtypes. *Nature Neuroscience*, *20*(2), 176–188. https://doi.org/10.1038/nn.4462

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA

URL http://www.rstudio.com/.188.

Schaller, F., Watrin, F., Sturny, R., Massacrier, A., Szepetowski, P., & Muscatelli, F. (2010). A single postnatal injection of oxytocin rescues the lethal feeding behaviour in mouse newborns deficient for the imprinted Magel2 gene. *Human Molecular Genetics*, *19*(24), 4895–4905. https://doi.org/10.1093/hmg/ddq424

Schaum, N., Karkanias, J., Neff, N. F., May, A. P., Quake, S. R., Wyss-Coray, T., … Weissman, I. L. (2018). Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature*, *562*(7727), 367–372. https://doi.org/10.1038/s41586-018-0590-4

Seifinejad, A., Li, S., Mikhail, C., Vassalli, A., Pradervand, S., Arribat, Y., … Tafti, M. (2019). Molecular codes and in vitro generation of hypocretin and melanin concentrating hormone neurons. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(34), 17061–17070. https://doi.org/10.1073/pnas.1902148116

Steinhoff, C., Paulsen, M., Kielbasa, S., Walter, J., & Vingron, M. (2009). Expression profile and transcription factor binding site exploration of imprinted genes in human and mouse. *BMC Genomics*, *10*(1), 144. https://doi.org/10.1186/1471-2164-10-144

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., … Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(43), 15545–15550. https://doi.org/10.1073/pnas.0506580102

Tiklová, K., Björklund, Å. K., Lahti, L., Fiorenzano, A., Nolbrant, S., Gillberg, L., … Perlmann, T. (2019). Single-cell RNA sequencing reveals midbrain dopamine neuron diversity emerging during mouse brain development. *Nature Communications*, *10*(1), 1–12. https://doi.org/10.1038/s41467-019-08453-1

Timshel, P. N., Thompson, J. J., Pers, T. H., Timshel, P. N., Thompson, J. J., & Pers, T. H. (2020). Mapping heritability of obesity by brain cell types. *BioRxiv*, 2020.01.27.920033. https://doi.org/10.1101/2020.01.27.920033

Trivers, R., & Burt, A. (1999). Kinship and genomic imprinting. *Results and Problems in Cell Differentiation*. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-69111-2_1

Tucci, V. (2016). Genomic Imprinting: A New Epigenetic Perspective of Sleep Regulation. *PLoS Genetics*, *12*(5), e1006004. https://doi.org/10.1371/journal.pgen.1006004

Tucci, V., Isles, A. R., Kelsey, G., Ferguson-Smith, A. C., Bartolomei, M. S., Benvenisty, N., … Wilkins, J. (2019). Genomic Imprinting and Physiological Processes in Mammals. *Cell*, *176*(5), 952–965. https://doi.org/10.1016/j.cell.2019.01.043

Varrault, A., Gueydan, C., Delalbre, A., Bellmann, A., Houssami, S., Aknin, C., … Journot, L. (2006). Zac1 Regulates an Imprinted Gene Network Critically Involved in the Control of Embryonic Growth. *Developmental Cell*, *11*(5), 711–722. https://doi.org/10.1016/j.devcel.2006.09.003

Wen, S., Ma, D., Zhao, M., Xie, L., Wu, Q., Gou, L., … Yan, J. (2020). Spatiotemporal single-cell analysis of gene expression in the mouse suprachiasmatic nucleus. *Nature Neuroscience*, *23*(3), 456–467. https://doi.org/10.1038/s41593-020-0586-x

Ximerakis, M., Lipnick, S. L., Innes, B. T., Simmons, S. K., Adiconis, X., Dionne, D., … Rubin, L. L. (2019). Single-cell transcriptomic profiling of the aging mouse brain. *Nature Neuroscience*, *22*(10), 1696–1708. https://doi.org/10.1038/s41593-019-0491-3

Zeisel, A., Hochgerner, H., Lönnerberg, P., Johnsson, A., Memic, F., van der Zwan, J., … Linnarsson, S. (2018). Molecular Architecture of the Mouse Nervous System. *Cell*, *174*(4), 999-1014.e22. https://doi.org/10.1016/j.cell.2018.06.021

726 **Appendix A : Specific Data set methods**

727 **1) Whole tissue level**

728 Mouse cell atlas (adult and developing mouse – 43 tissues) (Han et al., 2018)

729 The Mouse Cell Atlas (MCA) involved the collection and sequencing of 50+ tissues in an attempt to
730 distinguish genetic signatures of tissues across a mouse's life, of which 43 passed quality control
731 measures. These tissues included: 22 adult tissues (6-10 weeks old), 11 fetal tissues (E14.5), 6 neonatal
732 tissues and 4 stem cell derived tissues. An Expression matrix ('MCA_Figure2-batch-
733 removed.txt.tar.gz') of 60,000 cells of high quality (~1500 cells from the 43 tissues) was downloaded
734 from Figshare from (https://figshare.com/articles/MCA_DGE_Data/5435866) alongside the cell
735 metadata ('MCA_Figure2_Cell.info.xlsx'). Data were then scaled to 100,000 reads and log normalised.
736 Tissue cell identities were acquired from the 'Tissue' annotations. Data were run through the core
737 workflow twice, once for all adult tissues and once using all 43 tissues including those from the
738 developmental timepoints.

739 Tabula Muris (adult mouse – 20 tissues) (Schaum et al., 2018)

740 The Tabula Muris Consortium released the Tabula Muris soon after the MCA. The Tabula muris
741 comprises 20 adult mouse organs with a partial overlap with the MCA but sequenced these tissues to
742 a much greater depth and using two sequencing methods. Droplet sequencing allowed the surveying
743 of a large number of cells at relatively low coverage and was completed for 13 organs. FACs sorting
744 allowed characterisation of fewer cells from the organ but at much higher sensitivity and depth and
745 was completed for all 20 organs. This was completed using 3 female mice and 4 male mice, all 10-15
746 weeks old and on a C57BL/6JN background. The FACs dataset was selected for analysis as it contained
747 the largest number of tissues and the only one to include the brain (broken down as myeloid brain
748 cells (neurons, glia, endothelial etc) and non-myeloid brain cells (microglia and macrophages)).
749 Clustering for cells in this dataset allowed cell identification at the Tissue and cell identity level. All

750 data were downloaded through Figshare as Robjects
751 (https://figshare.com/articles/Robject_files_for_tissues_processed_by_Seurat/5821263/1). Raw
752 data and processed data were accessible from these files and cell metadata were acquired as
753 annotation csv's from the respective sequencing technology Figshare page (FACS -
754 https://figshare.com/articles/Single-cell_RNA-seq_data_from_Smart-
755 seq2_sequencing_of_FACS_sorted_cells_v2_/5829687). Tissue cell identities were acquired from the
756 'tissue' annotations and run through the core workflow once for all adult tissues. Multiple tissues had
757 >50% genes upregulated and hence the 2-fold-change upregulation cut off was used.

758 **2) Whole Brain level**

759 Mouse Brain Atlas (Zeisel et al., 2018)

760 Zeisel et al. (2018) created the Mouse Brain Atlas in 2018 – a comprehensive sequencing of the
761 adolescent mouse nervous system. This involved sequencing the entire Central and the Peripheral
762 nervous system, including the enteric nervous system. Cells were clustered into class (e.g. neuron,
763 oligodendrocyte), tissue of origin (e.g. hypothalamus, pons) and specific cell type by neurotransmitter
764 (e.g. serotonergic, dopaminergic) and location. Data were downloaded from mousebrain.org
765 (http://mousebrain.org/downloads.html) and the Level 5 data were downloaded as a one cell per
766 column loom file which also included the cell metadata as column attributes ("l5.all.loom"). Data were
767 then log normalised and scaled to 10,000 UMI. Cell lineage labels were acquired from 'Class'
768 annotations, cell tissue labels were acquired from 'tissue' annotations and cell identity labels were
769 acquired from 'ClusterName' annotations. The core workflow as run three times, once for cell lineage
770 with all cells, before restricting the analysis to neurons only and running it once for nervous system
771 regions and once for the specific neuron identities across the nervous system. For the cell lineage and
772 nervous system region analysis, neuronal lineage and a selection of brain regions had >50% genes
773 upregulated and hence the 2-fold-change upregulation cut off was used for both analyses.

774 Single-Cell Transcriptome profiling of the aging mouse brain (Ximerakis et al., 2019)

775 To complement the analysis of the mouse brain atlas, we also analysed Ximerakis et al.'s (2019) single
776 cell profile of the young and old mouse brains. Although the original aim of the study was to identify
777 aging-related genes/pathways, the data could also be used to investigate differential expression
778 between the different cell types within the brain, grouped by cell lineage. 16 C57BL/6 mice (8 at 2/3
779 months and 8 at 21/22 months) had whole brains dissected and hindbrain removed before tissue was
780 dissociated into single cells. Expression data and meta data were acquired from Gene Expression
781 Omnibus through accession no. GSE129788, scaled to 10,000 UMI and log normalised. Data were
782 run through the workflow once comparing all major CNS (minus hindbrain) cell types. Multiple cell
783 types had >50% genes upregulated and hence the 2-fold-change upregulation cut off was used.

784 **3) Brain Region level**

785 **3a) Whole Hypothalamus**

786 Informed by the above analysis and current theoretical understanding, the hypothalamus was
787 selected as a brain region of specific interest for imprinted genes and two datasets exist sequencing
788 large portions of the hypothalamus at once.

789 Chen et al. (2017)

790 Chen et al. (2017) sequenced adult (8-10 weeks) whole hypothalamus of female B6D2F1 mice (C57B6
791 female × DBA2 male). Processed/normalised data were downloaded as an R object from Gene
792 Expression Omnibus through accession no. GSE87544 alongside count data and a metadata file
793 with the cell identities. The original 14,437 cells were filtered to 3,319 high quality cells by the
794 authors (containing > 2000 unique genes expressed) and this filtered number was maintained for
795 this analysis. Cell identities included 11 non-neuronal and 34 neuronal identities identified as
796 Snap25/Syt1-high (and split into GABA and Glut populations alongside a Hista cluster). Cells
797 without a cluster (labelled as 'zothers') were discarded from the analysis leaving 2275 neurons.

798 Cell identities for all cells were taken from 'SVM_clusterID' annotations from the metadata file.
799 The core workflow as then run twice, once for all cells and once for just the neurons.

800 Romanov et al. (2017)

801 To complement the previous study, the data from Romanov et al. (2017) were also analysed. The
802 authors sequenced cells from "a central column" of the mouse hypothalamus which spanned from
803 the posterior POA to the ARC (rostro-caudal), PVN to the VLH (dorsal-lateral). Mice were on C57BL6/N
804 background, of both sexes and aged 14-28 days. The expression matrix and meta data was acquired
805 from a single file available from Gene Expression Omnibus through accession no. GSE74672. The
806 expression matrix was prefiltered by the authors to 2,882 high quality cells (>1500 UMI) and was
807 scaled to 10,000 UMI and log2 normalised. Cell identities included 6 non-neuronal identities and
808 62 neuronal subtypes (identified from 898 neurons) identified by expression of key neuropeptides
809 (e.g. *Avp, Oxt*, DA). The 'level1 class' annotations were used for non-neuronal populations and the
810 'level2 class (neurons only)" annotations were used for the 62 neuronal identities. The core
811 workflow was then run twice, once for all cells and once for just the neurons. Neurons had >50%
812 genes upregulated and hence the 2-fold-change upregulation cut off was used for the all cells run.

813 **3b) Specific Brain Nuclei/Neuron types**

814 ***3bi) Specific Hypothalamic Nuclei***

815 Several datasets focusing on individual regions of the hypothalamus were in existence prior to this
816 investigation and each was investigated individually for imprinted enrichment.

817 Arcuate Nucleus (ARC) (Campbell et al., 2017)

818 Campbell et al. (2017) sequenced the arcuate-median eminence complex of adult C57BL/6J mice
819 under different feeding conditions. They identified 65 distinct cellular identities of which 34 were
820 unique neuronal identities defined by key neuropeptides and novel markers (and 24 originated from

821 the arcuate nucleus). Raw and processed/filtered expression matrices alongside the cell metadata
822 (including cluster annotations) were acquired from Gene Expression Omnibus through accession no.
823 GSE93374. Data were scaled to 10,000 UMI (following Timshel et al., 2020) and log normalised.
824 Cell labels were acquired from the "Subcluster" annotations from the metadata. Data were run
825 through the workflow twice, once with all neurons and a second time with only ARC neurons
826 (removing the 10 suspected non-ARC subpopulations).

827 Preoptic Area (POA) (Moffitt et al., 2018)

828 The Preoptic area is of particular interest to parenting behaviour and feeding behaviour containing
829 neural populations that contribute to a number of innate hypothalamic behaviour outputs. Moffitt et
830 al. (2018) conducted a molecular, spatial, and functional characterisation of the murine POA and as
831 part of this investigation they profiled 31,299 cells from male and female adult mouse POA. Alongside
832 21 major cell classes, 18,553 neurons were clustered into 66 distinct neuronal populations (of which
833 56 originated from the POA). Barcodes, Features and Matrix files were acquired from Gene Expression
834 Omnibus through accession no. GSE113576 and used to create a Seurat object for the analysis. Data
835 were scaled to 10,000 UMI and log10-normalised. Cluster identities were acquired from
836 Supplementary Table S1. accompanying the original publication
837 (https://science.sciencemag.org/content/362/6416/eaau5324). Neuronal cluster identities were
838 taken from 'Neuronal cluster (determined from clustering of inhibitory or excitatory neurons)'
839 annotations. Data were run through the workflow thrice, once with all neurons and twice with
840 only POA neurons (removing the 9 suspected extra-POA groups) with upregulated genes capped
841 at the standard 1FC cap which generated a large number of over-represented sub populations
842 and so again at the more conservative 2FC cap.

843 Lateral Hypothalamic Area (LHA) (Mickelsen et al., 2019)

Mickelsen et al. (2019) sequenced male and female C57BL/6J LHAs to characterise the suspected diverse neuronal populations relating to the variety of motivated behaviours the LHA is implicated in. Clustering of 6,944 cells produced 20 major clusters, and clustering of the 3,589 neurons produced 30 distinct neuronal clusters (15 GABA, 15 Glut). Normalised expression data and cell metadata for Glut and GABA cells separately were acquired upon personal correspondence with the authors, from the publicly accessible box folder (https://thejacksonlaboratory.ent.box.com/v/mickelsen-nat-neuro-2019-data). For each file, the cluster mappings were acquired from the 'dbCluster' annotation and prepended with the name of the particular cell e.g. "Gaba1" or "Glut13". Data were run through the workflow twice, for the GABA and Glut neurons separately.

Suprachiasmatic Nucleus (SCN) (Wen et al., 2020)

In a spatio-temporal manner, Wen et al. (2020) used single-cell RNA sequencing to investigate the basic and circadian expression of various cell types in the SCN by micro dissecting the SCN and surrounding areas of 55 virgin 7-8 week old male C57BL/6J mice and sequencing them at 12 circadian time points (5 per timepoint). For the main experiment, conducted via Drop-seq, 8 major cell types (incl. neurons) were identified and within the neuronal cluster, 5 SCN neuronal subtypes were identified (ignoring neuronal subtypes likely from other neighbouring brain regions). Neuron identities were then reconfirmed with 10x Genomics, processing an additional 8,679 cells including 1,251 SCN neurons. These neurons were classified into the same 5 neuronal clusters as drop-seq and found a high convergence between the groups. Since the 10x data was more comprehensive for neurons, this data were chosen to look for neural specific enrichment only. The 10X Seurat files was acquired from Gene Expression Omnibus from accession no. GSE132608 and a Seurat object was created. Cell metadata were acquired through personal communication with the authors. SCN neuron identities were acquired for the 10X from the 'cluster' annotation. The Data were scaled by UMI per cell and log-normalised and run through the workflow once.

3bii) Monoaminergic System

869 In addition to the specific hypothalamic nuclei, we decided to investigate imprinted gene expression
870 in other brain areas of interest, namely the dopamine neurons of the midbrain and the various
871 neurons of the Dorsal Raphe Nucleus, both of which had datasets available and were regions of over-
872 representation in the analysis of the Mouse Brain Atlas.

873 Dorsal Raphe Nucleus (Huang et al., 2019)

874 The Dorsal Raphe Nucleus is a critical nucleus in the CNS responsible for a significant proportion of
875 neuromodulators, and most significantly, approximately a third of all serotonergic neurons. Huang et
876 al. (2019) sequenced the dorsal raphe nucleus and surrounding areas of 8-10-week-old C57BL/6J mice.
877 They discovered 17 major clusters (of which 5 were major neuron classes e.g. serotonergic,
878 dopaminergic) and the neuron classes were found to have an additional 17 sub clusters. Data were
879 downloaded as processed and annotated Seurat objects (log normalised, scaled to 10,000 UMI) for
880 each of the major cell types from Harvard Dataverse (https://doi.org/10.7910/DVN/QB5CC8). The
881 workflow was run three times, once using major cell cluster identities acquired from the
882 'Curated_cellTypeLabels' annotations and once using just neurons with the major cell clusters and a
883 final time using just the neurons with the sub cluster identities acquired from the
884 'Curated_subtypeLabels'. Multiple major cell clusters had >50% genes upregulated and hence the 2-
885 fold-change upregulation cut off was used for that analysis.

886 Serotonin Neurons of the Raphe Nuclei (Ren et al., 2019)

887 To further investigate the presence of imprinted genes in the serotonergic system, we also analysed
888 Ren et al.'s (2019) more in-depth single cell RNA seq data set of individual serotonin neurons in the
889 dorsal/median raphe nuclei. Using FACS and Smart-seq2, the authors sequenced 999 serotonin
890 neurons from the brainstem region. They characterised 13 distinct serotonergic neuron clusters from
891 the 999 neurons with distinct marker gene profiles. The authors then proceeded to anatomically
892 organise these clusters into distinct regions of the brainstem alongside the neuronal projection
893 destinations. The expression matrix (CPM) was acquired from Gene Expression Omnibus using

894 accession no. GSE135132 and cluster identities for the 999 neurons were acquired through
895 personal communication with the authors. Data were run through the workflow twice, once with
896 the normal FC upregulation cap which revealed over-representation in over half of the subgroups
897 so was run again at the more conservative 2FC cap.

898 Brain-wide Dopamine Neurons (Hook et al., 2018)

899 The other key neuron type that suggested a presence of imprinted genes was the dopamine neuron
900 population. Hook et al. (2018) sequenced 473 dopamine neurons in the C57BL/6J mouse brain at
901 embryonic (E15.5) and early postnatal (P7). Midbrain and Forebrain dopamine neurons were
902 microdissected from E15.5 mouse brains and olfactory bulb, midbrain and forebrain (arcuate nucleus)
903 dopamine neurons from P7 mice, all identified using FACS. A total of 13 cell populations were found
904 (4 for E15.5 and 9 for P7) which included clustering of dopamine neurons from the arcuate nucleus,
905 VTA, PAG and SN individually. The raw and normalised/pre-filtered expression matrix for the 396
906 dopamine neurons was acquired from github (https://github.com/pwh124/sc-da-parkinsons) along
907 with accompanying cell metadata (cell cluster identities). Data were run through the workflow once.

908 Midbrain Dopamine Neurons (Tiklová et al., 2019)

909 Tiklová et al. (2019) focused on midbrain dopamine neurons only as a heterogenous group of cells in
910 their own right. Using Pitx3 as a marker, the authors sequenced individual neurons (isolated using
911 FACS) from the ventral midbrain of mice at 6 timepoints (E13.5, E15.5, E18.5, P1, P7, P90) and
912 identified 7 distinct neuron subgroups based on expression of key genes: Slc6a3 (Dat), Nxph4,
913 Aldh1a1, Vip, and Gad2. In addition, the subgroups were localised back to their region of the midbrain.
914 Raw and normalised gene expression matrices were downloaded from Gene Expression Omnibus
915 using accession no. GSE116138 and cell metadata (cell cluster identities) were acquired through
916 personal correspondence with the authors. Data were run through the workflow once.

917 Developing Midbrain (La Manno et al., 2016)

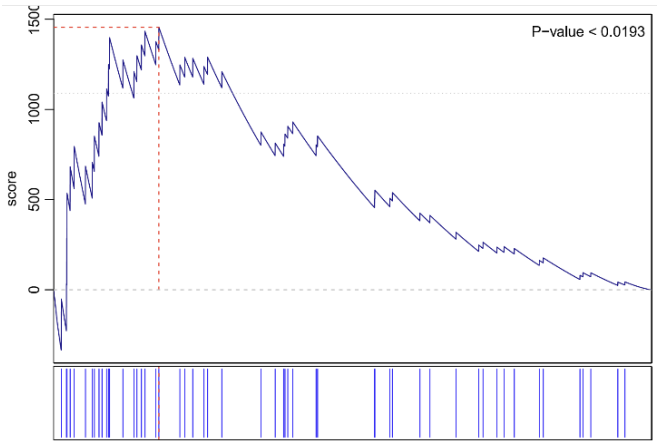
As a final investigation into the presence of imprinted genes in the midbrain, the data from La Manno et al. (2016) were analysed. The authors sequenced single cells from the CD-1 mouse ventral midbrain at time points between e11.5 – e18.5 covering dopaminergic progenitor specification and differentiation. Over this time period, 26 distinct cellular identities were identified including a variety of unique neuronal identities (e.g. serotonergic, dopaminergic, GABAergic). Data were downloaded from Gene Expression Omnibus using accession no. GSE76381. The expression matrix was scaled to 10,000 UMI and log normalised. Metadata was included as a row within the gene expression matrix and cell identities were prescribed using the 'cell-type' annotations. Data were run through the workflow once using all cell identities.

Appendix B – Comparison of Upregulated Genes for GABA5 in Romanov et al. (2017) and GABA17 in Chen et al. (2017). 21 genes were upregulated in both subgroups which is over 50% of the upregulated genes found in GABA5 including the top 5 marker genes. Position of genes organised by *p* value is included for both subtypes alongside the *p* values, Bonferroni corrected *q* values, and the fold change (FC) values for expression in this neuronal subpopulation vs. background. *Prlr* is a notable absentee; it is an upregulated gene in GABA17 but was not significant among the more restrictive markers of GABA 5 despite having a fold change of 1.86 vs. background.

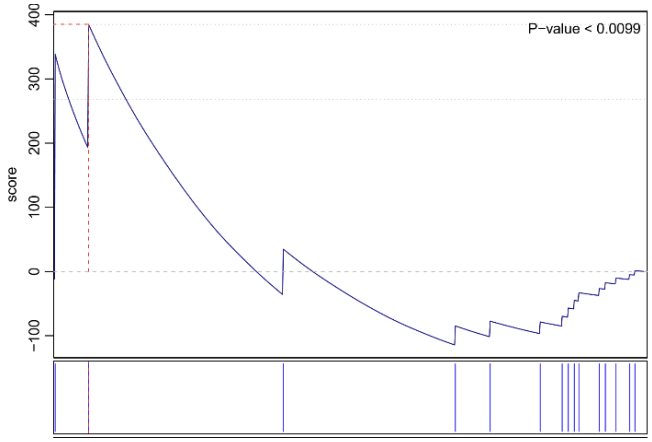
Gene	Position in GABA5 sorted by *p* (41 upreg genes)	GABA5 *p*	GABA5 *q*	GABA5 FC	Position in GABA17 sorted by *p* (823 upreg genes)	GABA17 *p*	GABA17 *q*	GABA17 FC
Calcr	1	2.09E-38	1.30E-36	8.13	41	4.12E-13	4.67E-12	4.96
Lhx1	2	1.82E-10	1.13E-08	9.56	29	1.35E-16	1.53E-15	4.74
Slc6a3	3	6.44E-09	1.33E-07	15.71	1	2.20E-111	7.47E-110	69.29
Asb4	4	6.76E-08	4.19E-06	2.82	443	0.000408	0.002311	1.99
1500016L03Rik	5	1.13E-07	3.50E-06	4.74	23	1.25E-19	1.42E-18	4.61
Npbwr1	7	8.22E-06	0.000510	9.02	817	0.008319	0.047138	3.55
Diap3	9	4.79E-05	0.002967	7.51	12	1.73E-27	5.87E-26	9.97
Gpr83	11	0.000118	0.003456	4.27	14	3.49E-25	1.19E-23	6.27
Slc18a2	13	0.000158	0.001961	3.24	3	9.52E-44	1.62E-42	9.43
Peg10	16	0.000240	0.004955	2.25	40	9.95E-14	3.38E-12	2.46
Dlk1	20	0.000324	0.009853	1.89	123	1.24E-07	1.05E-06	1.93
Fam159b	21	0.000371	0.007659	4.98	71	5.46E-10	6.19E-09	4.19
Rab3b	25	0.000462	0.009550	1.85	30	1.39E-16	4.72E-15	1.96
Six6	26	0.000526	0.008157	4.35	11	1.01E-27	3.45E-26	6.41
AA388235	31	0.000863	0.036304	3.79	82	2.36E-09	8.01E-08	10.61
Hnrnpul1	34	0.001399	0.043371	3.87	600	0.001031	0.035041	2.20
Arhgap36	37	0.001821	0.028222	2.19	46	2.61E-12	4.44E-11	3.50
Th	38	0.002050	0.014126	2.24	8	1.45E-35	2.46E-34	8.29
Dlx2	39	0.002801	0.034730	4.09	138	4.42E-07	3.01E-06	3.23
Dlx6	40	0.003047	0.047231	9.07	192	5.23E-06	5.93E-05	3.53
Dlx1	41	0.006400	0.036074	2.07	34	1.23E-15	2.08E-14	3.60

Appendix C – Significant GSEA not included in the main text. In the analysis, genes are sorted by strength by which they mark this neuronal cluster (sorted by fold change values) indicated by the bar (bottom). The genes are arrayed left (strongest marker) to right and blue lines mark where imprinted genes fall on this array. The vertical axis indicates an accumulating weight, progressing from left to right and increasing or decreasing depending on whether the next gene is an imprinted gene or not. The *p*-value represents the probability of observing the maximum value of the score (red dashed line) if the imprinted genes are distributed randomly along the horizontal axis.

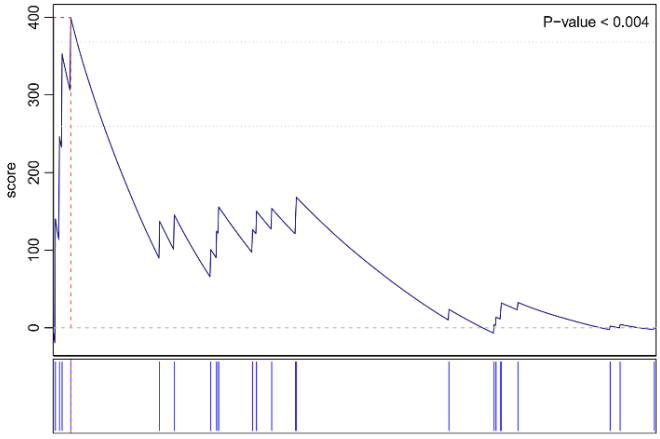
Appendix C1 – Neurons GSEA in Whole Hypothalamus (Chen et al., 2017)



Appendix C2 – Glut_9 (*Synpr/Gad1*) neurons in Lateral Hypothalamic Area (Mickelsen et al., 2019)

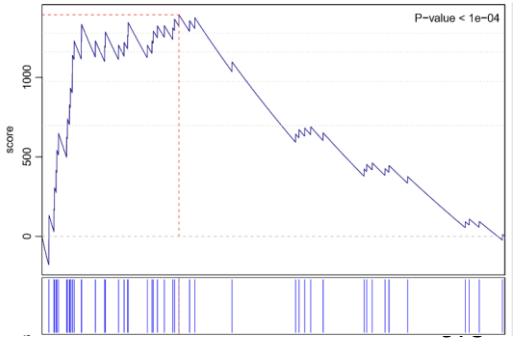


Appendix C3 – AVP/NMS neurons in Suprachiasmatic Nucleus (Wen et al., 2020)

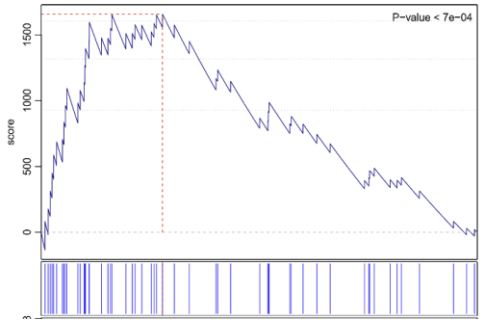


Appendix C4 - Dorsal Raphe Nucleus (Huang et al., 2019) (left to right, top to bottom – Serotonin neurons, Dopamine neurons, GABA neurons, Glutamatergic neurons, Peptidergic neurons)

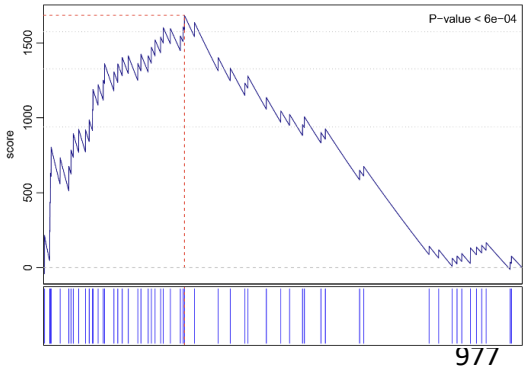
Serotonergic Neurons



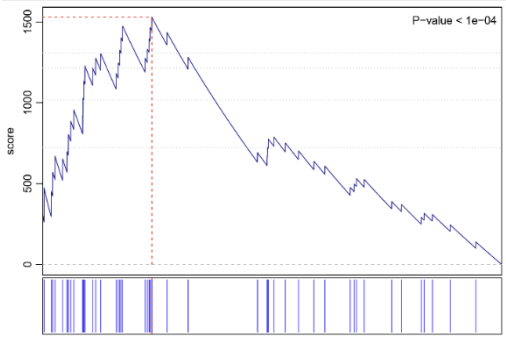
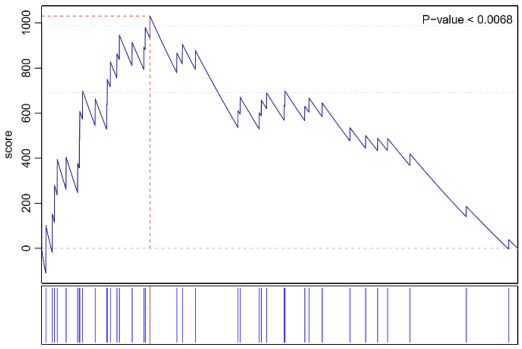
Dopaminergic Neurons



GABAergic Neurons



Glutamatergic Neurons



Peptidergic Neurons