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

Although a number of imprinted genes are known to be highly expressed in the brain, and in certain brain regions in particular, whether they are truly over-represented in the brain has never been formally tested. Using fifteen single-cell RNA sequencing datasets we systematically investigated imprinted gene over-representation at the organ, brain region, and cell-specific levels. We established that imprinted genes are indeed over-represented in the adult brain, and in neurons particularly compared to other brain cell-types. We then examined brain-wide datasets to test enrichment within distinct brain regions and demonstrated over-representation of imprinted genes in the hypothalamus, ventral midbrain, pons and medulla. Finally, using datasets focusing on these regions of enrichment, we identified hypothalamic neuroendocrine populations and the monoaminergic hindbrain neurons as specific hotspots of imprinted gene expression. These analyses provide the first robust assessment of the neural systems on which imprinted genes converge. Moreover, the unbiased approach, with each analysis informed by the findings of the previous level, permits highly informed inferences about the functions on which imprinted gene expression converges. Our findings indicate the neuronal regulation of motivated behaviours such as feeding, parental behaviour and sleep as functional hotspots for imprinting, thus adding statistical rigour to prior assumptions and providing testable predictions for novel neural and behavioural phenotypes associated with specific genes and imprinted gene networks. In turn, this work sheds further light on the potential evolutionary drivers of genomic imprinting in the brain.

INTRODUCTION

Imprinted genes demonstrate a preferential or exclusively monoallelic expression from either the maternal or paternal allele in an epigenetically predetermined manner (a parent-of-origin effect, POE), and to date approximately 260 imprinted genes (demonstrating biased allelic expression and/or associated with a parental-specific epigenetic mark) have been identified in the mouse (~230 in humans) (Ferguson-Smith 2011; Tucci et al. 2019). This epigenetic regulation makes genomic imprinting an evolutionary puzzle as many of these genes are effectively haploid and thereby negate many of the benefits of diploidy (Orr 1995). Studying the patterns of expression and function of imprinted genes may therefore shed light on the drivers leading to the evolution of genomic imprinting. For instance, functional characterisation of a number of imprinted genes points to convergence on placental function (Peters 2014), in line with the predictions of early theoretical ideas (Moore and Haig 1991). Outside of the placenta, the brain consistently emerges as an adult tissue with a large number of expressed imprinted genes (Gregg et al. 2010; Babak et al. 2015; Andergassen et al. 2017). However, given that it is estimated that ~80% of all genes in the genome are expressed in the brain (Lein et al. 2007; Negi and Guda 2017), the question remains, *is* imprinted gene expression actually enriched in the brain compared to other adult tissues? To date this has never been formally tested.

A role for imprinted genes in the brain was initially suggested by Cattanach and Kirk (1985) and neurological phenotypes have been observed in early imprinted gene mouse models (Keverne et al. 1996a), as well as the behavioural deficits seen in imprinting disorders such as Prader-Willi and Angelman syndromes (Nicholls et al. 1989; Angulo et al. 2015). Subsequent studies have revealed diverse roles for imprinted genes in the brain. During development, several imprinted genes are involved in the processes of neural differentiation, migration, axonal outgrowth and apoptosis (Perez et al. 2016). In the adult brain, studies of mice carrying manipulations of individual imprinted genes have suggested a wide range of behavioural roles including maternal care (Li et al. 1999), feeding (Davies et al. 2015), social behaviour (Garfield et al. 2011; McNamara et al. 2018), learning/memory (Jiang et al. 1998), cognition (Relkovic et al. 2010; Dent et al. 2018), and more recently, sleep and circadian activity (Lassi et al. 2012).

59 In addition to studies on individual imprinted genes, there are a limited number of studies that take a
60 systems level approach to characterizing the role of genomic imprinting in the brain. Early studies
61 examining developing and adult chimeras of normal and parthenogenetic/gynogenetic (Pg/Gg - two
62 maternal genomes) or androgenetic (Ag - two paternal genomes) cells indicated distinct regional
63 distribution for maternally (cortex and hippocampus) and paternally (hypothalamus) expressed genes
64 (Allen et al. 1995; Keverne et al. 1996a). More recently, Gregg et al. (2010) used the known imprinting
65 status of 45 imprinted genes and the Allen Brain Atlas to track dichotomous expression of imprinted
66 genes across 118 brain regions to identify brain-wide patterns of expression. Most imprinted genes were
67 expressed in every brain region, but detectable expression of the largest number of imprinted genes was
68 found in regions of the hypothalamus (medial preoptic area, arcuate nucleus), central amygdala, basal
69 nuclei of the stria terminalis and the monoaminergic nuclei, suggesting some form of specialisation.
70 Although pioneering, this study, and others identifying novel imprinted genes and/or mapping allelic
71 expression in the brain (DeVeale et al. 2012; Babak et al. 2015; Perez et al. 2015; Andergassen et al.
72 2017), did not test whether the expression of these genes was especially enriched in given brain regions
73 but simply asked if they were expressed, at any level, or not.
74 Here we address the question of whether the brain and/or specific brain circuitry is a foci for genomic
75 imprinting by exploiting the rapidly expanding number of single-cell RNA sequencing (scRNA-seq)
76 datasets and systematically investigating imprinted gene enrichment and over-representation in the
77 murine brain. We performed this by a hierarchical sequence of data analysis, using datasets that allowed
78 a multi-organ comparison first, before proceeding to brain-specific and brain region-specific
79 comparisons with the outcome of each level informing the data selection for the next one, to identify a
80 consistent pattern of enrichment (Figure 1). We sought to provide a robust assessment of the neural
81 systems on which imprinted genes converge, statistically validating previous assumptions, identifying
82 neuronal domains that have received less emphasis in earlier studies, and providing testable predictions
83 for novel neural and behavioural phenotypes associated with specific genes and imprinted gene
84 networks.
85

86 **RESULTS**

87 **Imprinted gene expression is enriched in the brain in a multi-organ analysis**

88 The Mouse Cell Atlas (MCA) (Han et al. 2018) and the *Tabula Muris* (TM) (Schaum et al. 2018) are
89 single cell compendiums containing ~20 overlapping, but not identical, adult mouse organs. Key
90 overlapping organs include the bladder, brain, kidney, lung, limb muscle, and pancreas while organs
91 included in only one dataset include the ovary, testes, uterus, stomach within the MCA and the heart,
92 fat, skin, trachea and diaphragm within the TM. These compendiums create a snapshot of gene
93 expression across adult tissues to assess imprinted gene enrichment. Since this study focused on the
94 adult body and brain, fetal tissues (including the placenta) were not assessed.

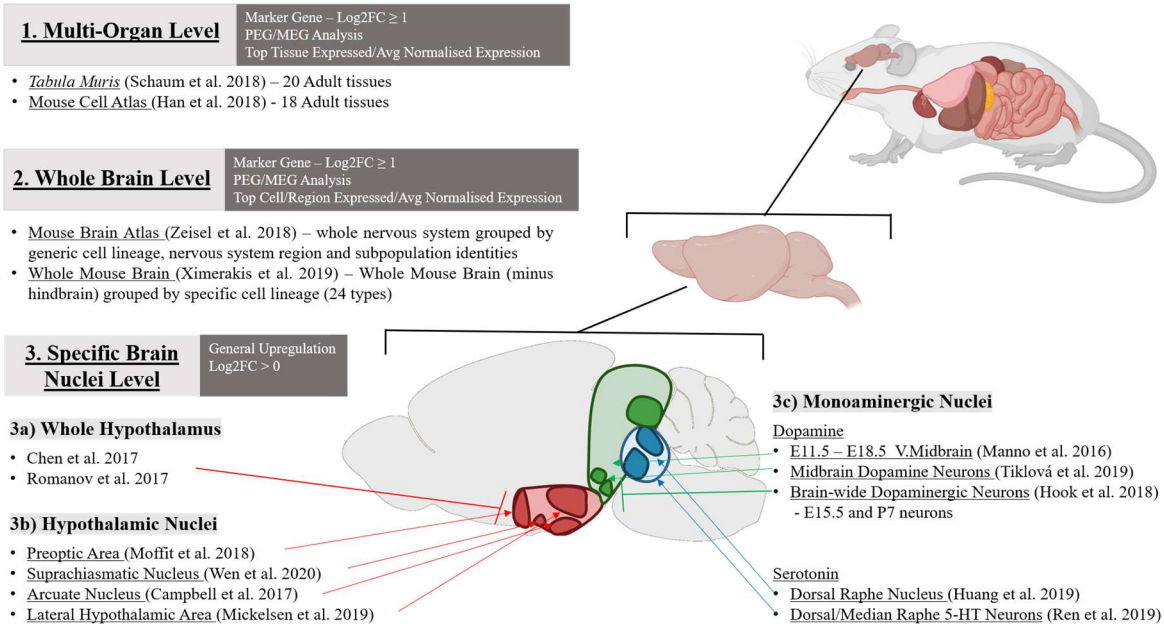


Figure 1. The hierarchical set of datasets in this analysis. The datasets are sorted into multi-organ, whole brain and specific brain nuclei level analyses. The original publication and specific tissue/s analysed are provided for each analysis. White text in dark grey box indicates specifics to the analysis at that level – whether the analysis used the 'marker gene' Log2FC criteria or the relaxed Log2FC > 0 criterion, whether paternally and maternally expressed gene (PEG/MEG) analysis was carried out and whether the top tissue/cell-types and the average normalised expression were reported for imprinted genes.

95 An over-representation analysis (ORA) was performed on both datasets. All data were processed
96 according to the original published procedure, a list of upregulated genes was produced for each
97 tissue/identity group (vs. all other tissue/identity groups) and a one-sided Fisher's Exact test was
98 performed using a custom list of imprinted genes (Supplemental Table S1) to identify tissues in which
99 imprinted genes were over-represented amongst the upregulated genes for that tissue. Each dataset in
100 this study was analysed independently which allowed us to look for convergent patterns of enrichment
101 between datasets of similar tissues/cell-types. Across only adult tissues, imprinted genes were

102 convergently over-represented in the pancreas, bladder and the brain in both datasets (Figure 2A). In
103 addition, in the MCA adult tissue dataset there was a significant over-representation in the uterus (Table
104 1), and in the *Tabula Muris* analysis (Table 2), there was a significant over-representation in the muscle-
105 based tissues - diaphragm, trachea, and limb muscles. In addition to the ORA, to identify situations in
106 which imprinted genes were in fact enriched amongst the stronger markers of a tissue/cell-type, we
107 performed a Gene-Set Enrichment Analysis (GSEA) on tissues meeting minimum criteria (see Materials
108 and Methods), which assessed whether imprinted genes were enriched within the top ranked
109 upregulated genes for that tissue (ranked by Log2 Fold Change). No tissue at this level showed a
110 significant GSEA for imprinted genes. Mean normalised expression of imprinted genes across identity
111 groups (Supplemental Table S2) was the highest for Brain in the MCA and highest for Pancreas in the
112 TM (Brain Non-Myeloid was the fourth highest).

Table 1. 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 with $q \leq 0.05$ and Log2FC ≥ 1 (total number of genes in the dataset in brackets); *IG* – number of imprinted genes upregulated with $q \leq 0.05$ and Log2FC ≥ 1 (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 (20,534)	IG (95)	ORA *p*	ORA *q*	Mean FC IG	Mean FC Rest	Top Tissue IGs
Pancreas	2737	42	1.57E-13	**1.89E-12**	8.74	10.32	22
Brain	3401	34	4.43E-06	**5.31E-05**	8.76	125.00	19
Bladder	3183	29	0.000168	**0.002012**	4.45	8.51	8
Uterus	2567	22	0.002827	**0.033919**	4.66	8.46	7
Lung	1203	8	0.192705	1	3.82	151.41	4
Ovary	2219	13	0.223666	1	7.46	11.27	5
Kidney	1714	10	0.268425	1	13.76	182.89	5
Liver	1739	8	0.560145	1	4.55	80.51	3
Stomach	1821	7	0.748590	1	4.24	88.60	3
Thymus	1805	6	0.851579	1	2.78	6.76	2
Small Intestine	1719	5	0.908008	1	7.99	218.64	2
Testis	5212	14	0.995891	1	27.04	5058.36	10
Bone Marrow	1095	2	-	-	5.31	4.43	1
Mammary Gland Virgin	902	4	-	-	3.70	4.03	0
Muscle	1127	4	-	-	8.64	15.05	3
Peripheral Blood	1146	3	-	-	3.78	3.57	0
Prostate	369	0	-	-	0.00	478.10	0
Spleen	1501	1	-	-	4.90	4.77	1

113

Table 2. Imprinted gene over-representation in *Tabula Muris* adult tissues (Schaum et al. 2018). *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. All other column descriptions can be found in the legend of Table 1.

Identity	Up Reg (20,839)	IG (107)	ORA *p*	ORA *q*	Mean FC IG	Mean FC Rest	GSEA *p*	GSEA *q*	Top Tissue IGs
Diaphragm	416	19	3.66E-13	**4.75E-12**	6.49	4.83	0.1898	0.3796	4
Limb Muscle	761	24	6.32E-13	**8.22E-12**	9.02	5.09	0.0552	0.1104	8
Pancreas	4104	43	8.31E-07	**1.08E-05**	12.52	12.60	-	-	29
Trachea	1979	25	1.78E-05	**0.0002**	3.81	4.57	-	-	5
Brain (Non-Myeloid)	3081	31	0.0001	**0.0016**	12.16	14.17	-	-	14
Bladder	3338	31	0.0005	**0.0068**	3.30	5.30	-	-	16

Fat	1263	12	0.0286	0.3713	3.46	3.68	-	-	1
Heart	1108	10	0.0585	0.7601	2.87	5.14	-	-	0
Mammary Gland	1826	12	0.2264	1	3.52	5.24	-	-	3
Liver	1808	7	0.8307	1	6.19	54.93	-	-	3
Aorta	3515	14	0.8832	1	7.47	16.08	-	-	2
Tongue	4295	15	0.9696	1	4.15	7.16	-	-	8
Large Intestine	4758	11	0.9998	1	5.95	12.22	-	-	5
Brain (Myeloid)	1024	5	-	-	3.39	6.80	-	-	2
Kidney	584	3	-	-	24.94	22.90	-	-	1
Lung	914	2	-	-	2.73	5.41	-	-	0
Marrow	1957	5	-	-	7.65	5.25	-	-	4
Skin	1612	4	-	-	4.15	8.36	-	-	1
Spleen	625	1	-	-	4.63	4.28	-	-	0
Thymus	678	4	-	-	3.46	7.45	-	-	1

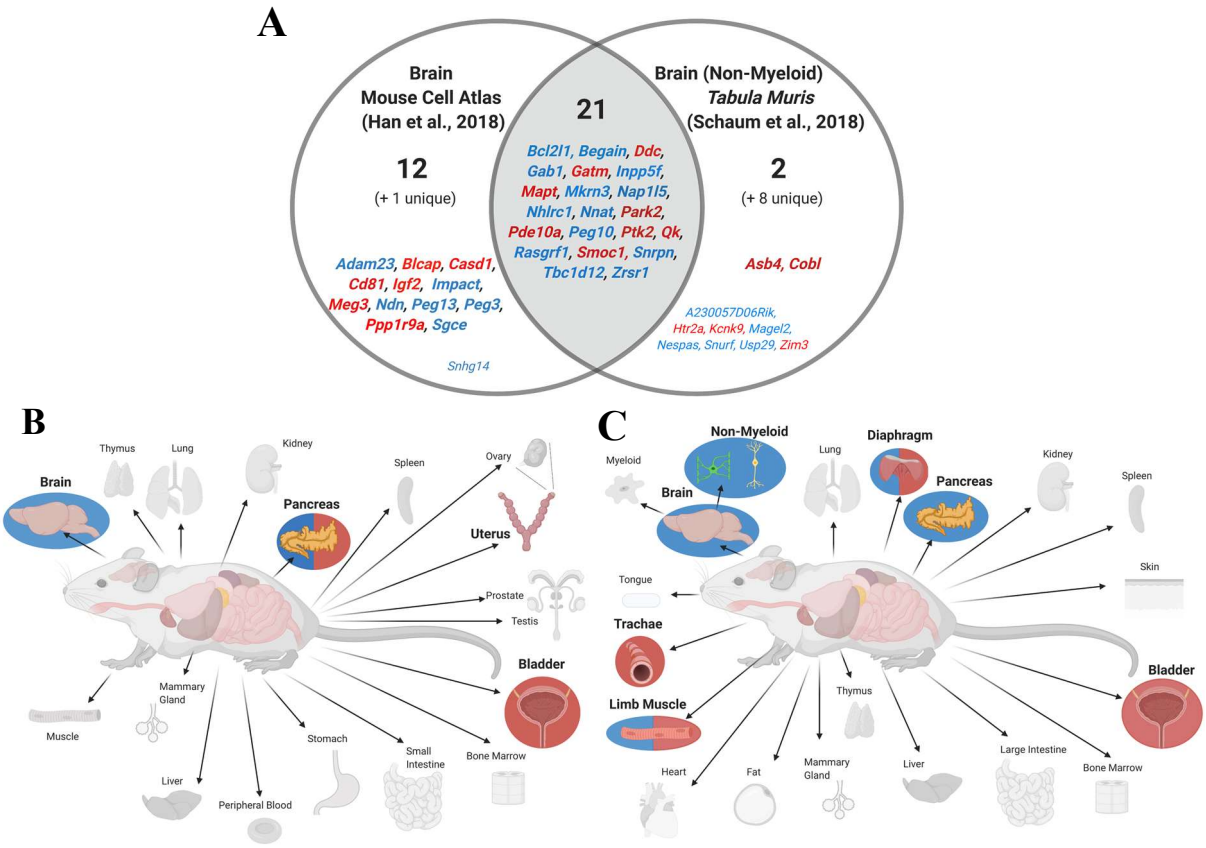


Figure 2. Level 1 multi-organ comparison summary graphics. (*A*) Venn diagram of upregulated imprinted genes in the brain in Mouse Cell Atlas and in the brain (non-myeloid) in the Tabula Muris. Imprinted genes are listed which show significant upregulation ($q \leq 0.05$ and Log2FC ≥ 1) in the tissues. Although these tissues are not identical, these were the two brain associated over-representations in the enrichment analysis. Parental-bias is indicated by colour (MEG - red, PEG - blue). From the 119 imprinted genes in the gene list, only 92 were common to both analyses (i.e., successfully sequenced and passed gene quality control filters). 34 imprinted genes were upregulated in the brain in the MCA and 31 genes in the *TM*. Genes in common from the two analyses are presented in bold and totalled in each section of the Venn Diagram, while genes found upregulated in one analysis but not available in the other analysis are included in small font and the number indicated in brackets. (*B*) 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, a red circle represent the same for MEGs, and a red/blue split circle were over-represented for both PEGs and MEGs. (*C*) Tissues with over-representation in *Tabula Muris* (Schaum et al. 2018). See 2b description for details.

Given the interest in the different functions of maternally expressed genes (MEGs) and paternally expressed genes (PEGs), we additionally ran the large-scale enrichment analyses (Levels 1 and 2) using separate lists of PEGs and MEGs. At Level 1, MEGs and PEGs (Supplemental Table S3A, S3B, S4A and S4B) revealed a similar pattern of enrichment in both datasets (Fig. 2). PEGs were over-represented

118 in the brain in both datasets (MCA - $q = 4.56 \times 10^{-6}$, *TM* - $q = 0.0005$) while MEGs were over-represented

119 in neither. PEGs were also over-represented in the diaphragm ($q = 0.0007$), limb muscle ($q = 0.0001$)

120 and pancreas (MCA - $q = 1.93 \times 10^{-5}$, *TM* - $q = 0.0002$), with a significant GSEA in the MCA pancreas

121 ($p = 0.02$, Supplemental Fig. S1). While MEGs were over-represented in the bladder (MCA - $q = 0.002$,

122 *TM* - $q = 0.020$), the pancreas (MCA - $q = 1.53 \times 10^{-7}$) and in the three muscular tissues of the *Tabula*

123 *Muris* (diaphragm - $q = 2.13 \times 10^{-8}$, limb muscle - $q = 2.43 \times 10^{-7}$, trachea - $q = 0.004$).

124 **Imprinted gene expression is enriched in neurons and neuroendocrine cells of the brain**

125 We next analysed cells from the whole mouse brain (Level 2), first analysing Ximerakis et al.'s (2019)

126 dataset, in which cells were grouped from the whole mouse brain (minus the hindbrain) into major cell

127 classes according to cell lineage. Imprinted genes were over-represented in neuroendocrine cells and

128 mature neurons (Table 3).

Table 3. Imprinted gene over-representation in neural lineage types (Ximerakis et al. 2019)**.** *Identity* – Cell lineage identities for the cells used in analysis. All other column descriptions can be found in the legend of Tables 1 & 2.

Identity (Abbr.)	Up Reg (14,498)	IG (85)	ORA *p*	ORA *q*	Mean FC IG	Mean FC Rest	GSEA *p*	GSEA *q*	Top Cell IGs
Neuroendocrine cells (NendC)	3868	47	2.12E-08	**3.82E-07**	11.88	5.42	0.0017	**0.0051**	26
Mature Neurons (all types) (mNEUR)	2968	32	0.0002	**0.0035**	8.80	9.28	-	-	2
Arachnoid barrier cells (ABC)	2287	20	0.0396	0.7120	16.84	22.63	-	-	7
Tanycytes (TNC)	1279	12	0.0692	1	6.64	12.01	-	-	8
Vascular and leptomeningeal cells (VLMC)	1714	15	0.0724	1	15.06	13.03	0.0468	0.1404	4
Oligodendrocyte precursor cells (OPC)	1524	13	0.1067	1	3.03	7.17	-	-	1
Pericytes (PC)	1801	14	0.1649	1	8.20	8.22	-	-	2
Olfactory ensheathing glia (OEG)	1086	9	0.1848	1	7.95	26.03	-	-	1
Oligodendrocytes (OLG)	1183	9	0.2561	1	3.73	12.91	-	-	5
Choroid plexus epithelial cells (CPC)	2602	17	0.3524	1	7.43	19.34	-	-	5
Hemoglobin-expressing vascular cells (Hb_VC)	1798	11	0.4889	1	5.25	6.33	-	-	3
Vascular smooth muscle cells (VSMC)	3006	17	0.6093	1	8.94	6.71	0.1376	0.4128	5
Astrocyte-restricted precursors (ARP)	1445	8	0.6214	1	4.50	5.09	-	-	1
Neural stem cells (NSC)	1009	5	0.7138	1	4.00	4.09	-	-	0
Ependymocytes (EPC)	3233	17	0.7346	1	15.04	53.27	-	-	4
Endothelial cells (EC)	1455	7	0.7619	1	5.80	8.54	-	-	0
Hypendymal cells (HypEPC)	1525	6	0.8946	1	17.24	20.80	-	-	5
Neuronal-restricted precursor (NRP)	2339	10	0.8979	1	3.07	10.20	-	-	1
Astrocytes (ASC)	1384	4	-	-	2.22	6.04	-	-	0
Dendritic cells (DC)	1209	1	-	-	3.50	16.02	-	-	1
Immature Neurons (ImmN)	652	4	-	-	3.37	5.79	-	-	0
Macrophages (MAC)	1222	2	-	-	3.47	21.56	-	-	0
Microglia (MG)	1342	3	-	-	19.28	19.22	-	-	3
Monocytes (MNC)	947	2	-	-	16.49	19.13	-	-	1
Neutrophils (NEUT)	519	2	-	-	9.18	62.13	-	-	0

129 Neuroendocrine cells were defined as a heterogeneous cluster, containing peptidergic neurons and

130 neurosecretory cells expressing neuronal marker genes (e.g., *Syt1* and *Snap25*) alongside neuropeptide

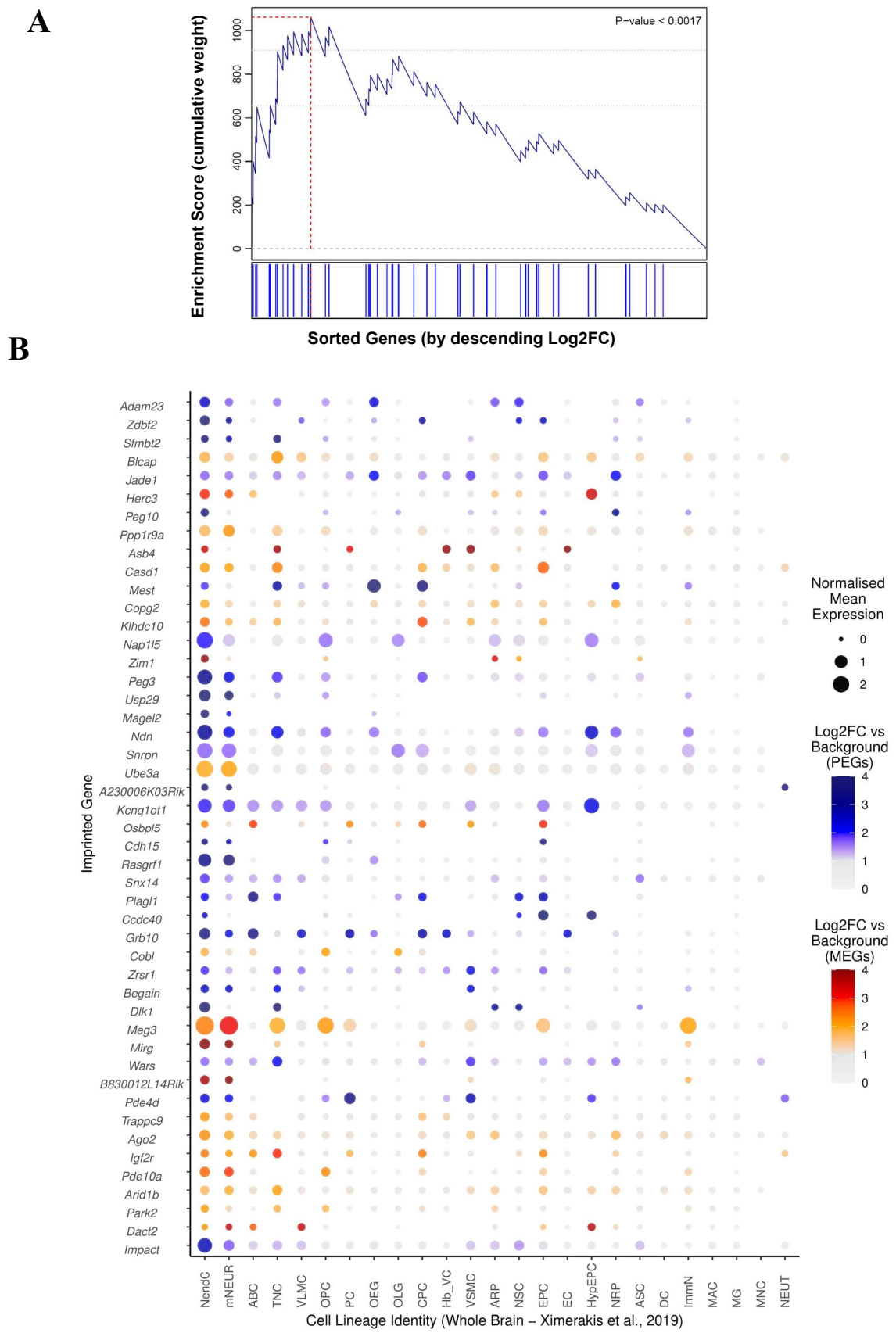


Figure 3. GSEA and dot plots for imprinted genes upregulated in neuroendocrine cells in the Ximerakis et al. (2019) whole mouse brain dataset. (*A*) GSEA for imprinted genes upregulated in the neuroendocrine cells. In the analysis, genes are sorted by strength by which they mark this neuronal cluster (sorted by Log2FC 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. The *q*-value for this analysis was significant at 0.0036. (*B*) Dot plot of imprinted genes upregulated in the 'Neuroendocrine cells' plotted across all identified cell types (Abbr. in Table 3). Imprinted genes were plotted in chromosomal order. Size of points represented absolute mean expression; colour represented the size of the Log2FC value for the cell identity group (e.g., neuroendocrine cells) vs. all other cells. Unique colour scales are used for MEGs (red/orange) and PEGs (blue).

genes (e.g., *Oxt*, *Avp*, *Gal*, *Agrp* and *Sst*) but distinguished by Ximerakis et al. (2019) by the unique expression of *Baiap3* which plays an important role in the regulation of exocytosis in neuroendocrine cells (Zhang et al. 2017). GSEA additionally showed that the imprinted genes were enriched in the genes with the highest fold change values for neuroendocrine cells (Fig. 3). 26 imprinted genes had their highest expression in the neuroendocrine cells and the mean normalised expression of imprinted genes was almost twice as high for neuroendocrine cells as the next highest identity group (Supplemental Table S2). The MEG/PEG analysis (Supplemental Table S5A and S5B) for this dataset found that PEGs were over-represented in mature neurons (q = 0.027) and neuroendocrine cells (q = 8.97 x 10^{-6}). MEGs were also over-represented in neuroendocrine cells (q = 0.047) and uniquely over-represented in Arachnoid barrier cells (q = 0.014). Only PEGs replicated the significant GSEA in neuroendocrine cells (p = 4x10^{-4}, Supplemental Fig. 2)

The second dataset at this level was Zeisel et al.'s (2018) Mouse Brain Atlas (MBA) and it allowed a much deeper investigation of nervous system enrichment with sequencing of the entire murine nervous system and identifying cells by both brain region and cell type. Concordant with the previous findings, primary analysis separating cells by lineage revealed over-representation of imprinted genes in neurons only (Table 4). The overlap between the upregulated imprinted genes for the over-represented neural-lineage cells from the Level 2 datasets are displayed in Figure 4. Additionally, PEGs alone demonstrated no significant over-representations in cell lineage types while MEGs demonstrated over-representation in vascular cells only (q = 0.0004) (Supplemental Table S6A and S6B).

Table 4. Imprinted gene over-representation in nervous system cell types (Zeisel et al. 2018). *Identity* – Cell identities for the cells used in analysis. All other column descriptions can be found in the legend of Table 1.

Identity	Up Reg (19,547)	IG (108)	ORA *p*	ORA *q*	Mean FC IG	Mean FC Rest	Top Cell IGs
Neurons	5710	44	0.0081	**0.0487**	11.73	24.97	45
Vascular	2473	22	0.0171	0.1029	17.91	26.64	16
Oligos	1587	11	0.2701	1	4.64	11.48	12
Peripheral Glia	2820	16	0.5117	1	5.42	12.64	12

Ependymal	3683	20	0.5912	1	24.52	66.97	15
Immune	1564	7	0.7787	1	13.42	93.05	5
Astrocytes	1539	4	-	-	2.88	10.73	3

150 **The hypothalamus, ventral midbrain, pons and medulla are enriched for imprinted gene**
151 **expression**

152 After confirming neuron-specific enrichment of imprinted genes in the MBA dataset, further MBA
153 analysis was performed on cells classified as neurons and then grouped by brain/nervous system
154 regions. Significant over-representation was seen in neurons of the hypothalamus, ventral midbrain,
155 medulla, and pons (Table 5). The pons and medulla had the largest number, 45 and 44 respectively, of
156 imprinted genes upregulated (Figure 5A).

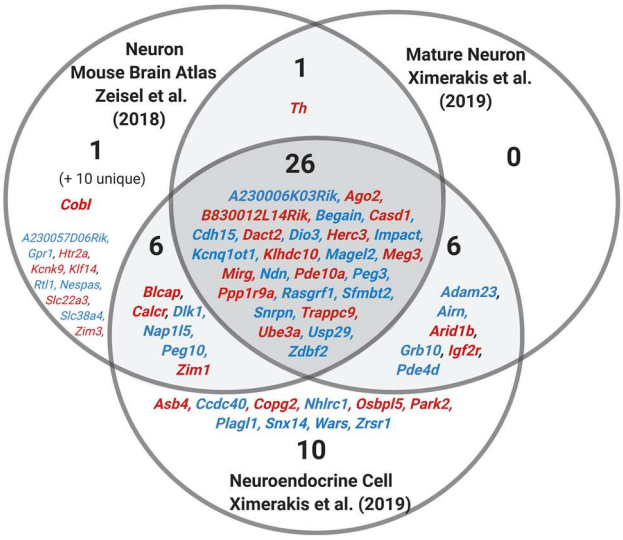
Table 5. Imprinted gene over-representation in nervous system regions (Zeisel et al. 2018)**.** *Identity* – Nervous system regional identities for the cells used in analysis genes. All other column descriptions can be found in the legend of Tables 1 & 2.

Identity	Up Reg (18,335)	IG (105)	ORA *p*	ORA *q*	Mean FC IG	Mean FC Rest	GSEA *p*	GSEA *q*	Top Region IGs
Medulla	3147	45	8.38E-10	**1.26E-08**	4.79	4.01	0.1	0.2	15
Hypothalamus	1040	22	9.81E-08	**1.47E-06**	4.92	5.84	-	-	8
Pons	3581	44	1.62E-07	**2.43E-06**	4.20	3.91	0.1169	0.2338	22
Vent. Midbrain	1228	18	0.0002	**0.0034**	4.90	4.99	-	-	3
Vent. Striatum	689	8	0.0463	0.6941	3.92	4.92	-	-	0
Posterior Cortex	1090	9	0.1788	1	2.64	3.20	-	-	2
Enteric Nervous System	3885	26	0.2311	1	8.98	121.04	-	-	11
Sympathetic Nervous System	2804	18	0.3535	1	11.37	57.96	-	-	9
Anterior Cortex	979	6	0.5016	1	2.72	3.30	-	-	1
Dors. Midbrain	1045	6	0.5663	1	2.20	4.85	-	-	3
Thalamus	1441	8	0.6000	1	2.90	6.36	-	-	0
Hippocampus - CA1	1082	6	0.6008	1	3.01	4.02	-	-	2
Somatosensory Cortex	2121	11	0.6943	1	4.09	3.70	-	-	8
Dors. Striatum	1196	6	0.6974	1	4.03	5.43	-	-	2
Dorsal Root Ganglion	3607	16	0.9088	1	11.56	75.89	-	-	9
Middle Cortex	623	5	-	-	3.29	3.24	-	-	0
Spinal Cord	972	5	-	-	4.57	12.36	-	-	1
Amygdala	452	4	-	-	4.65	4.11	-	-	2
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
Antero-Middle Cortex	646	3	-	-	4.59	4.31	-	-	1
Cerebellum	240	0	-	-	0.00	32.30	-	-	0

157 Regional analysis for MEGs and PEGs separately (Supplemental Table S7A and S7B), revealed that
158 PEGs were over-represented in hypothalamus ($q = 6.53\text{x}10^{-7}$), ventral midbrain ($q = 0.018$), the pons (q
159 $= 4.65\text{x}10^{-5}$) and the medulla ($q = 4.10\text{x}10^{-6}$); while MEGs were only over-represented in the medulla
160 ($q = 0.002$) but had a significant GSEA for the pons ($q = 0.027$, Supplemental Fig. S3); see Figure 5B.

Neurons were then recategorized into the individual subtypes (Zeisel et al. 2018) to uncover the specific neural populations underlying the enrichment seen in the hypothalamus, pons and medulla, and midbrain (Fig. 6; Supplemental Table S8). Each neural population is identified by its distinct gene expression and suspected location within the brain (see http://mousebrain.org/ for an online resource with detailed information on each cluster).

Figure 4. Venn diagram of upregulated imprinted genes in the mature neuronal cells in the whole brain datasets of Zeisel et al. (2018) and Ximerakis et al. (2019). Imprinted genes are listed which show significant upregulation ($q \leq 0.05$ and Log2FC ≥ 1) in the cells. Although these cell types are not identical, these were all mature neural lineage cells with over-representations in the enrichment analysis. Parental-bias is indicated by colour (MEG - red, PEG - blue. From the 119 imprinted genes in the gene list, only 88 were common to both analyses (i.e., successfully sequenced and passed gene quality control filters). 45 imprinted genes were upregulated in neurons in the MBA, and in Ximerakis et al. (2019), 33 imprinted genes were upregulated in neurons and 48 genes in neuroendocrine cells. Genes in common from the two analyses are presented in bold and totalled in each section of the Venn Diagram, while genes found upregulated in one analysis but not available in the other analysis are included in small font and the number indicated in brackets.



The hypothalamus was represented by a selection of inhibitory and peptidergic neurons. Inhibitory neurons with over-representation of imprinted genes included: Deinh4 (notable genes *Lhx8, Gabrq*) representing a Subthalamic Nucleus population, Deinh5 (*Nts*, *Dlk1, Esr1, Prlr*) representing a Preoptic Area population, Deinh6 (*Agrp, Otp*) representing an Arcuate nucleus population, and Deinh7 (*Avp, Rgs16, Nms*) and Deinh8 (*Six6*, *Nms, Vip*) suspected to be Suprachiasmatic nucleus populations. For peptidergic neurons, over-representation was seen in HYPEP3 (*Gpr101, Tac1,*

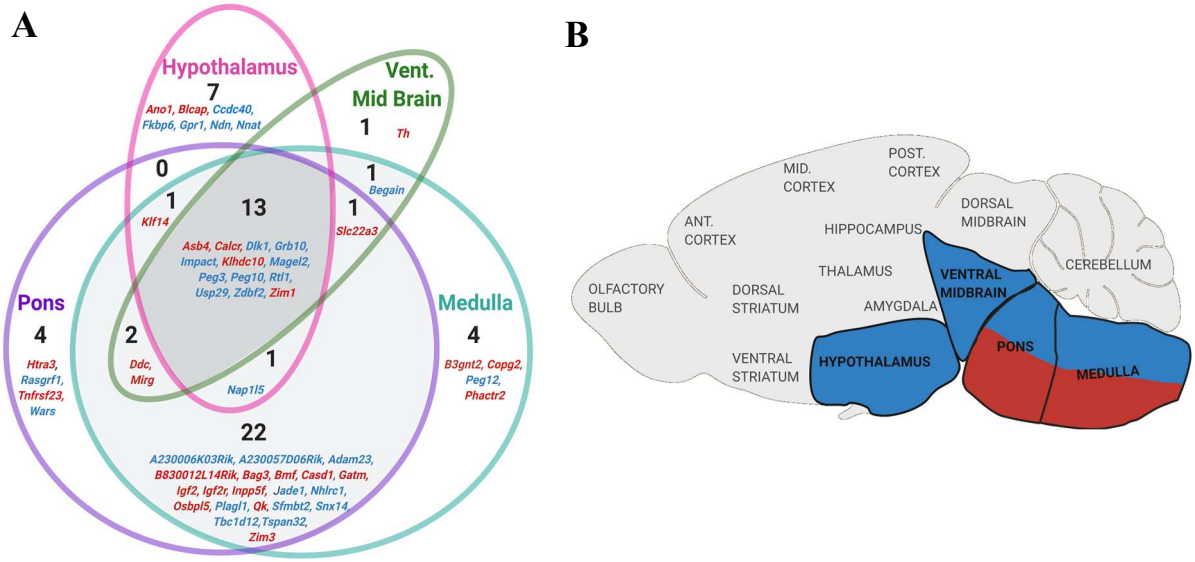


Figure 5. Level 2ii) brain region analysis summary figures. (*A*) Venn diagram of upregulated imprinted genes in the neurons of enriched nervous system regions from the Mouse Brain Atlas (Zeisel et al. 2018). Imprinted genes are listed which show significant upregulation ($q \leq 0.05$ and Log2FC ≥ 1) in the regions specified. The number of imprinted genes in each region of the Venn diagram are specified. Parental-bias of imprinted genes is indicated by colour (MEG - red, PEG - blue). (*B*) Brain regions enriched for imprinted gene expression via ORA or GSEA in the MBA (Zeisel et al. 2018). Regions over-represented for all imprinted genes are bolded. Regions over-represented for PEG expression alone are coloured blue while regions enriched for MEG expression alone are coloured red.

Baiap3) 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, and HYPEP6 (*Hcrt, Pdyn, Trhr*) an orexin producing population of the dorsomedial/lateral hypothalamus.

The midbrain, medulla and pons were represented by a number of cell groups, with over-representation seen in the medulla-based adrenergic (HBAR) and noradrenergic (HBNOR) groups and the dopaminergic neurons of the midbrain in the Periaqueductal Gray (PAG) (MBDOP1) and the Ventral Tegmental Area (VTA) and Substantia Nigra (SNc) (MBDOP2). There were also several inhibitory (MEINH, HBIN) and excitatory neuron (MEGLU, HBGLU) types spread across the nuclei from the three regions (Fig. 6). The serotonergic populations of the raphe nuclei of these regions (HBSER) are particularly prominent since the pons and medulla base serotonin neuron populations (HBSER2, HBSER4 and HBSER5) were the only neuron subpopulations out of the 214 total to have a significant GSEA for imprinted genes after correction (Supplemental Fig. S4).

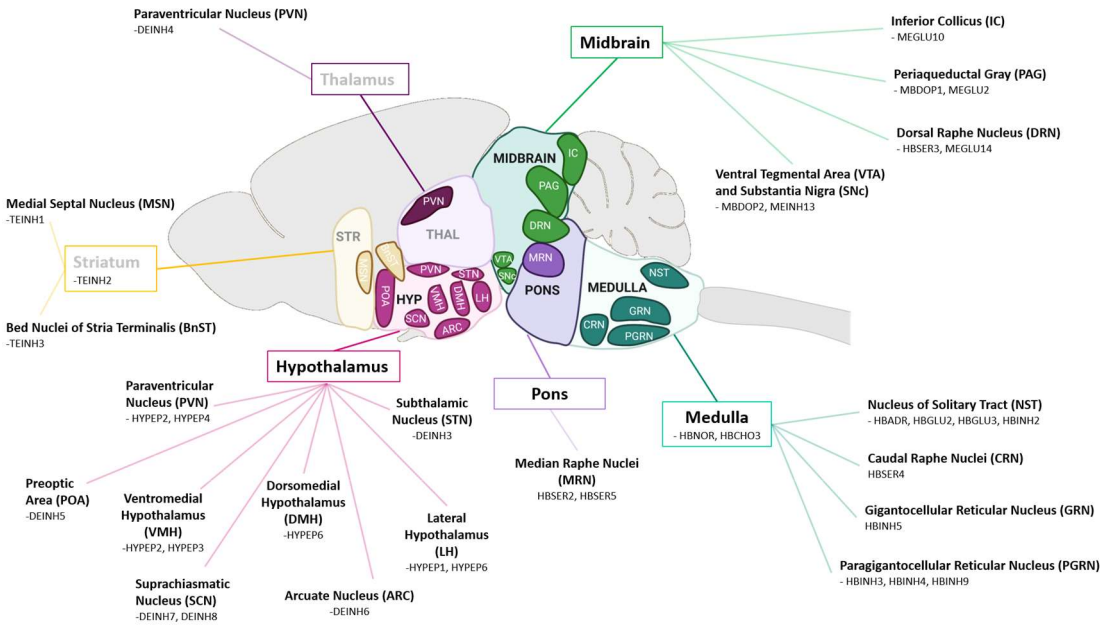


Figure 6. Anatomical labelling of all the neural subpopulations with a significant over-representation of imprinted genes ($q \leq 0.05$ and Log2FC ≥ 1) in the Mouse Brain Atlas (Zeisel et al. 2018). The predicted brain nuclei localisation of the 32neuronal subpopulations(out of 214 populations identified across the nervous system) 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 Supplemental Table S8.

185 Additional regions of over-representation included TEINH1-3 representing neurons in the pallidum and

186 striatum and DEINH4 representing PVN neurons from the thalamus. In this comparison of 214 neuron

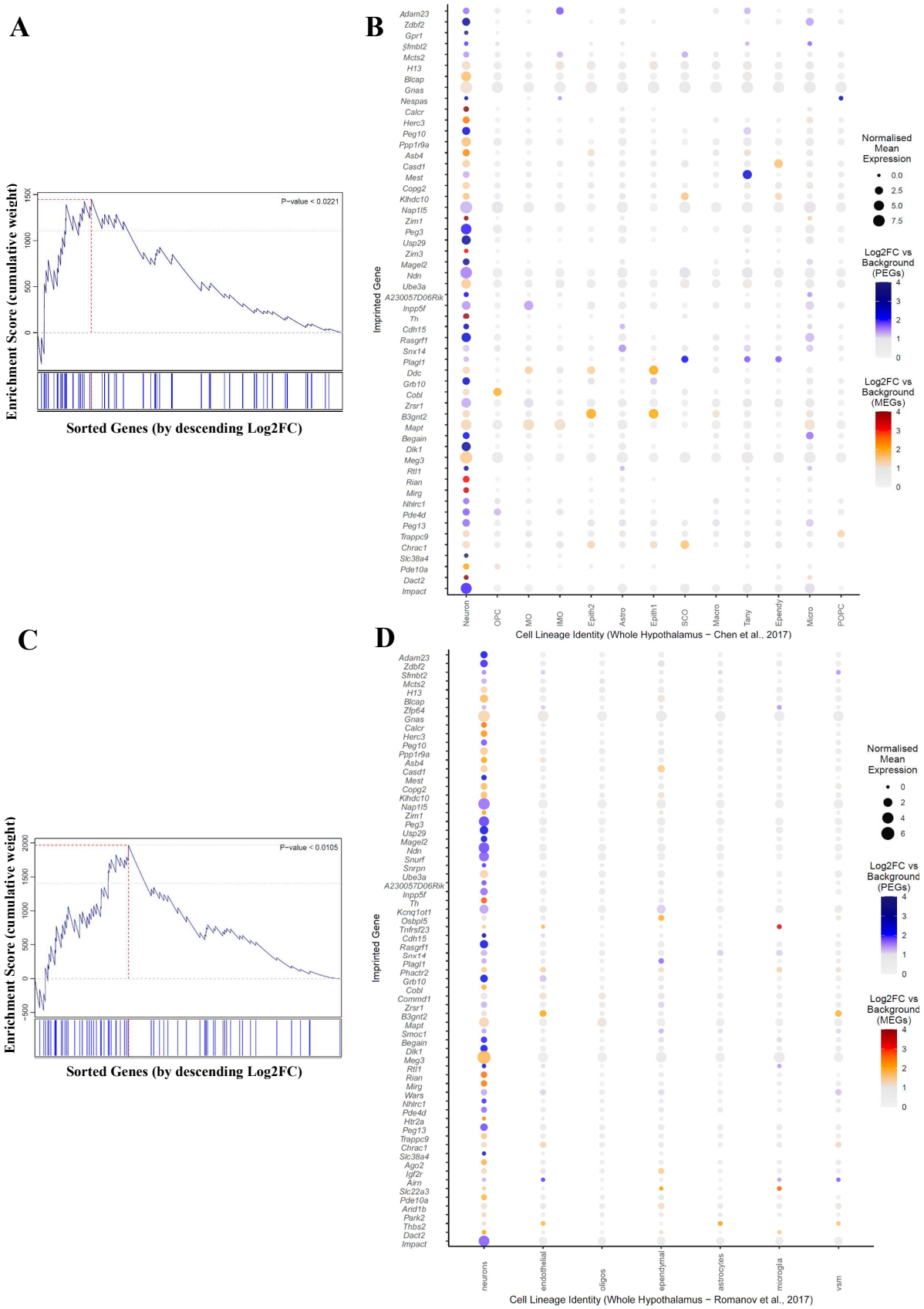


Figure 7. GSEA and Dot plots for imprinted genes upregulated in neurons across the whole hypothalamus. (*A*) GSEA for imprinted genes upregulated in the 'Neuron' cell type in the whole hypothalamic dataset of Chen et al. (2017). See legend of Figure 3A for a description of how to interpret the plot. (*B*) Dot plot of imprinted genes upregulated in the 'Neuron' cell type plotted across all identified cell types in the Chen et al. (2017) whole hypothalamic dataset. See legend of Figure 3B for a description of how to interpret the plot. Abbr: OPC = Oligodendrocyte Precursor Cell, MG = Myelinating Oligodendrocyte, IMG = Immature Oligodendrocyte, Astro = Astrocyte, Epith = Epithelial, Macro = Macrophage, Tany = Tanycyte, Ependy = Ependymocyte, Micro = Microglia, POPC = Proliferating Oligodendrocyte Progenitor Cell. (*C*) GSEA for imprinted genes upregulated in 'neurons' in the whole hypothalamic dataset of Romanov et al. (2017). See legend of Figure 3A for a description of how to interpret the plot. (*D*) Dot plot of imprinted genes upregulated in 'neurons' plotted across all identified cell types in the Romanov et al. (2017) whole hypothalamic dataset. See legend of Figure 3B for a description of how to interpret the plot.

populations, no neurons from areas such as the cortex, cerebellum or peripheral nervous system were enriched, and neither were they over-represented in the previous regional analysis. Hence, further analysis focused on those brain regions enriched in this whole brain level analysis.

Imprinted genes are over-represented in specific hypothalamic neuron subtypes

We next sought to investigate whether those regional neuron enrichments found within the whole brain comparisons would be further clarified with enriched expression in specific neuronal subpopulations within those regions. Namely within the hypothalamus (to identify neural populations enriched across the whole hypothalamus and those enriched within specific hypothalamic nuclei) and whether imprinted gene expression was enriched in the dopaminergic and serotonergic populations, located in the ventral midbrain and hindbrain, which were found as key subpopulations in the whole brain analysis.

Two datasets with single cell sequencing data for the adult hypothalamus existed (Chen et al. 2017; Romanov et al. 2017). Both 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 (Supplemental Table S9A and S10A). Within the Romanov et al. (2017) data, there was a significant over-representation of imprinted genes in neurons ($q = 0.02$) and a similar observation was seen in the Chen et al. (2017) data ($q = 0.001$), and both also demonstrated a significant GSEA in neurons (Fig. 7A-D, Romanov et al. 2017 – $p = 0.011$, Chen et al. 2017 - $p = 0.022$).

Within the Chen et al. (2017) dataset, 4/33 hypothalamic neuronal subtypes had a significant over-representation of imprinted genes (Supplemental Table S9B). The four subtypes were all GABAergic neurons, specifically: GABA13:*Slc18a2/Gal* ($q = 0.0079$), with high expression of *Gal, Esr1, Prlr* and *Th,* present in a few hypothalamic regions (but concentrated in the POA); GABA17:*Slc6a3* ($q = 0.0001$)

209 a dopaminergic neuron type with high expression of *Th* and *Prlr*, which most likely corresponds to the
210 TIDA neurons of the arcuate nucleus; GABA8:*Vipr2* (q = 0.0071) with very high *Avp* and *Nms*
211 expression suggesting these are neurons from the SCN; and GABA15:*Agrp* (q = 0.034) feeding
212 promoting neurons of the Arcuate Nucleus. Within the Romanov et al. (2017) dataset, 3/62 subtypes
213 had significant over-representation of imprinted gene expression (Supplemental Table S10B):
214 GABA5:*Calcr/Lhx1* ($q = 1.63 \times 10^{-6}$) was the top hit but was a poorly segregated population likely due
215 to a deeper inner cluster heterogeneity. This cluster was interesting since the imprinted genes *Calcr* and
216 *Asb4* were amongst its most significant marker genes, and it was notably the only cluster with high
217 expression of all three of *Th*, *Slc6a3* and *Prlr.* Romanov et al. (2017) did not identify any of their
218 populations as the TIDA neurons, but the above pattern of gene expression suggests that GABA5 may
219 contain these neurons. Furthermore, GABA17:*Slc6a3* from the Chen et al. (2017) dataset shared 21/40
220 upregulated genes of GABA5 (see Supplemental Table S11 for full comparison); GABA14:*Agrp*/*Npy*
221 (q = 0.013) which were the Arcuate Nucleus feeding neurons also reported in Chen et al. (2017); and a
222 *Ghrh* neuronal type (q = 0.032), also strongly expressing *Th*, which was not over-represented in Chen
223 et al. (2017).

224 ***Arcuate nucleus (ARC)***

225 The first nuclei investigated was the ARC sequenced by Campbell et al. (2017). Imprinted gene over-
226 representation was found in 8/24 arcuate neuron types (Supplemental Table S12). These included the
227 *Agrp*/*Sst* neuron type (with high expression of *Npy,* q = 0.003) and two *Pomc* neuron types
228 (*Pomc/Anxa2*, q = 0.004; *Pomc/Glipr1*, q = 0.03) known to be feeding suppressants (Rau and Hentges
229 2017). Additional significant over-representation was found in the *Ghrh* neuron type (q = 0.009), which
230 was also enriched in Galanin and *Th*. Finally, a highly significant over-representation of imprinted genes
231 was found in the *Th*/*Slc6a3* neuron type ($q = 1.72 \times 10^{-8}$) identified by the authors as one of the most
232 likely candidates for the TIDA dopaminergic neuron population. Marker genes for this identity group
233 overlapped with the TIDA candidates from the previous two datasets (e.g., *Slc6a3*, *Th*, *Lhx1*, *Calcr*).
234 *Agrp* neurons and these TIDA candidate neurons were identified in both whole hypothalamic datasets
235 and at the nuclei level.

236 ***Preoptic Area (POA)***

237 Imprinted gene over-representation was found in 14/66 of the neuron clusters identified by Moffitt et
238 al. (2018) in the preoptic area (Supplemental Table S13A). Moffitt et al. (2018) stated that 9 of the 66
239 clusters most likely originate from outside the POA including 3/14 of the imprinted gene enriched
240 neuron clusters (one of which was the *Avp*/*Nms* neurons (q = 0.0003) likely to be from the SCN). To
241 focus our analysis on the POA alone, we removed the non-POA neuron clusters and ran the analysis on
242 the remaining neural groups (Supplemental Table S13B). Imprinted genes were found to be over-
243 represented in 15 neuron types including two Galanin enriched clusters (i22:*Gal/Pmaip1*, $q = 2.41 \times 10^{-}$
244 ^{5}; i8:*Gal/Amigo2*, q = 0.0003), the major *Th* enriched cluster (i17:*Th/Nos1*, q = 0.003)), e15:*Ucn3*/*Brs3*
245 (q = 0.0015, a cluster activated by parenting behaviour), a *Crh* cluster (i35:*Crh/Tac2*, q = 0.003), a *Trh*
246 cluster (e4:*Trh/Angpt1*, q = 0.003), and both major glutamatergic *Ghrh* clusters (e13:*Ghrh/C1ql1*, q =
247 0.02; e19:*Ghrh/Trh*, q = 0.005). Since a considerable number of neuron types were over-represented,
248 the above analysis was run again with the Log2FC ≥ 1 criteria to find out if imprinted genes were
249 enriched in the more distinct markers (Supplemental Table S13C). Only two neuron types were over-
250 represented: i35:*Crh*/*Tac2* (q = 0.0126) and i16:*Gal*/*Th* (q = 0.0026). i16:*Gal*/*Th* is a prominent
251 behavioural neuron group (parenting and mating behaviour) which was not identified in the previous
252 analysis. Only three of Moffitt et al.'s (2018) neuron groups were upregulated in *Slc18a2, Gal, Prlr, Th*
253 and *Esr1* (i8, i16, i38) like GABA13 in Chen et al. (2017) and all of them are represented above.

254 ***Lateral Hypothalamic Area (LHA)***

255 Over-representation of imprinted genes in the LHA neurons was analysed for Glutamatergic and
256 GABAergic neurons separately since they were sequenced separately by Mickelsen et al. (2019). For
257 Glutamatergic neurons, imprinted gene over-representation was found in 6/15 clusters (Supplemental
258 Table S14A). These included: Glut 6:*Hcrt* (q = 0.008) identified as the Orexin cluster, a classic
259 motivated behaviour neuronal subset identified in the MBA as HYPEP6; and Glut 9:*Synpr*/*Gad1* (q =
260 0.0032)) with a significant GSEA (q = 0.0254, Supplemental Fig. S5) and synaptoporin as a marker
261 gene; it is a cell type found scattered across the LHA and DMH. The four others were Glut
262 14:*Otp/Gpr101* (q = 0.0005), Glut 7:*Gpr101/Tcf4* (q = 0.0001), Glut 5:*Ebf3/Otp* (q = 0.0045) and Glut
263 11:*Calca/Col27a1* ($q = 3.01 \times 10^{-5}$) which were all enriched in similar genes such as *Otp, Tcf4* and
264 *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 (Level 2). For GABAergic neurons imprinted gene over-representation was found in 2/15 clusters: Gaba 1:*Gal*/*Dlk1* (q = 0.0004), a galanin enriched cluster, previously shown to modulate food reward (Qualls-Creekmore et al. 2017) and Gaba 5:*Meis2*/*Calb2* (q = 0.015), another sub-type lacking functional classification (Supplemental Table S14B).

Suprachiasmatic Nucleus (SCN)

Analysis of the 10x chromium data of SCN neurons (Supplemental Table S15) revealed a significant over-representation (q = 1.51 x 10^{-8}) and GSEA (p = 0.004, Supplemental Fig. S6) in the *Avp*/*Nms* neuronal cluster (out of 5 neuronal clusters). 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 likely corresponds to 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. Figure 8 presents the overlapping upregulated imprinted genes from the convergently upregulated neuron subtypes in the hypothalamic analysis of Level 3a and 3b.

Imprinted gene expression is over-represented in monoaminergic nuclei of the mid- and hindbrain

In the MBA, Whole Hypothalamus and Arcuate Nucleus analyses, dopaminergic clusters were consistently enriched and, to explore this further, 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

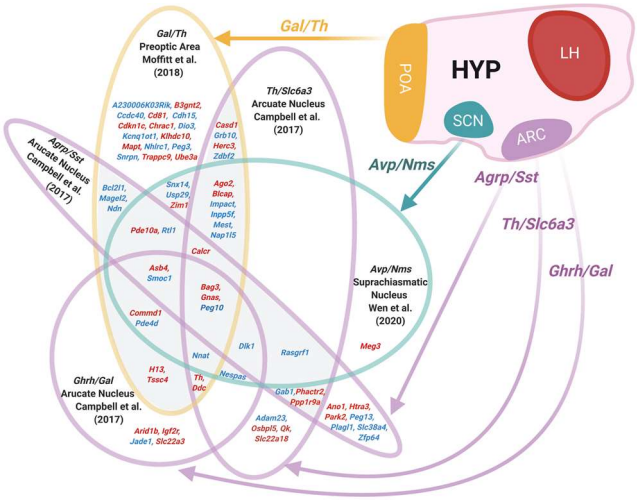


Figure 8. Venn diagrams of upregulated imprinted genes in the neuronal subpopulations from level 3b that were also identified in level 3a. Imprinted gene overlap was contrasted from *Th/Slc6a3*, *Ghrh* and *Agrp/Sst* neuronal populations of the Arcuate Nucleus (Campbell et al. 2017) , *Avp/Nms* neurons from the Suprachiasmatic Nucleus (Wen et al. 2020) and i16 (*Gal/Th*) neurons from the Preoptic Area (Moffitt et al. 2018). These populations were represented at level 3a and level 3b analyses. Imprinted genes are listed which show significant upregulation ($q \leq 0.05$ and Log2FC > 0) in the subpopulation. Parental-bias is indicated by colour (MEG - red, PEG - blue).

292 dopamine neurons emerged as the clearest over-represented subgroups (Supplemental Table S16). This
293 included the *Th*/*Slc6a3* neurons (q = 1.15x10^{-8}) and the *Gal/Ghrh* cluster (q = 4.79x10^{-5}) the latter of
294 which were referred to as 'neuroendocrine' cells by Hook et al. (2018), and the former a mixture of
295 arcuate nucleus populations with *Prlr* was one of the marker genes, suggesting this includes the TIDA
296 neurons. Additionally, P7 midbrain neurons were the other group with significant over-representation
297 (specifically from the PAG and VTA) as well as the neuroblasts at this time point.
298 Although no specific adult mouse midbrain datasets exist, La Manno et al.'s (2016) ventral midbrain
299 sequencing at E11.5 - E18.5 allowed us to identify imprinted enrichment within the midbrain at a
300 timepoint when the major neuronal populations are differentiating but still identifiable (Supplemental
301 Table S17). As anticipated, we found significant over-representation in both mature (DA1; high *Th* and
302 *Slc6a3,* q = 0.0103), and developing (DA0, q = 0.0129) dopaminergic neurons, as well as the
303 serotonergic neurons (q = 3.09x10^{-7}) likely from the midbrain raphe nuclei.
304 As a final investigation, we used Tiklová et al.'s (2019) data in which purely dopamine neurons (*Pitx3*+)
305 of the ventral midbrain were sequenced. 2/7 dopaminergic subpopulations identified in this study had
306 significant over-representation of imprinted genes (Supplemental Table S18). The *Slc6a3*+ neurons (q
307 = 0.0032) from the VTA and *Vip*+ neurons (q = 0.045) from the PAG. The seven subpopulations were
308 found to belong to two major subtypes of neurons by Tiklová et al. (2019), Dat^{High} neurons with high
309 expression of *Th* and *Slc6a3* that made up 89% of *Pitx3* positive neurons identified in the ventral
310 midbrain, and Dat^{Low} neurons, the rarer *Pitx3* positive neurons lacking high expression of the typical
311 dopamine markers (*Th/Slc6a3*). The two neuronal subpopulations found enriched here were both
312 members of the Dat^{High} subtype, with *Aldh1a1*+ neurons (SNc - not enriched here) completing this
313 subtype. The PAG and VTA over-representation mirrors the same over-representation pattern found in
314 Hook et al.'s (2018) midbrain dopaminergic neurons and Figure 9A presents the overlap in imprinted
315 genes upregulated in these midbrain dopaminergic neurons.
316 To examine enrichment of imprinted genes in serotonergic neurons, analyses were performed on data
317 for the Raphe nuclei of the midbrain/hindbrain which are the key serotonergic regions of the brain.
318 Analysis of all cell types in Huang et al.'s (2019) sequencing of the Dorsal Raphe Nucleus (DRN)
319 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 DRN (Supplemental Table S19A). When compared to all other cell populations, significant ORA was seen for Dopaminergic (q = 0.009), Serotonergic (q = 0.012) and Peptidergic neurons (q = 0.0008), however, a significant GSEA was found for all five neuronal populations (Supplemental Fig. S7). When compared against each other (i.e., serotonergic upregulation vs. the other neurons), only the serotonergic neurons of the DRN (q = 0.0019) were found to have a significant over- representation of imprinted genes (Supplemental Table S19B). GSEA's were non-significant but the mean fold change for imprinted genes was markedly higher in both serotoninergic (52% higher) and dopaminergic neurons (68% higher). 2/5 serotonin subpopulations had significant over-representation of imprinted genes: *Hcrtr1/Asb4* (q = 0.0014) and *Prkcq/Trh* (q = 0.007) (Supplemental Table S19C). 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).

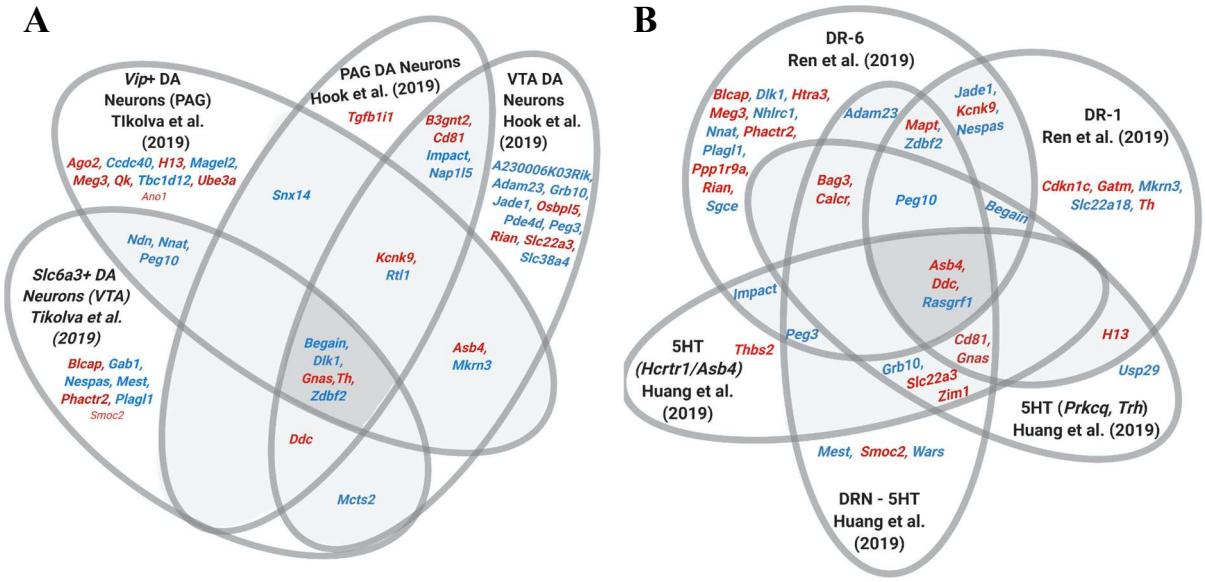


Figure 9. Venn Diagrams of upregulated imprinted genes in the adult dopaminergic and serotonergic populations. (*A*) Dopaminergic neurons subtypes from the PAG and VTA from Hook et al. 2019 and Tiklová et al. 2019. (*B*) Serotonin neurons and subtypes from Huang et al. 2019 and Ren et al. 2019 - Imprinted genes are listed which show significant upregulation ($q \leq 0.05$ and Log2FC > 0) in that subpopulation. Parental-bias is indicated by colour (MEG - red, PEG - blue). Genes upregulated in one analysis but not detected in the other are presented in small, non-bold, font.

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 (Supplemental

337 Table S20A). Since the majority of subtypes were considered over-represented, the above analysis was
338 run again with the Log2FC ≥ 1 criteria to find out if imprinted genes were enriched in the more distinct
339 markers of any subtypes (Supplemental Table S20B). The only significantly over-represented subtype
340 in this more conservative analysis was DR-1 (q = 0.005), a dorsal/lateral DR population with the largest
341 enrichment of *Trh*, the marker for serotonin neurons that project to subcortical areas, namely the
342 hypothalamus and thalamus (See Figure 9B for the overlap in imprinted genes upregulated in these
343 serotonergic neurons).

344 **DISCUSSION**

345 Using publicly available single cell transcriptomics data, we apply an unbiased systems biology
346 approach to examine the enrichment of imprinted genes at the level of the brain in comparison to other
347 adult tissues, refining this analysis to specific brain regions and then to specific neuronal populations.
348 We confirm a significant over-representation in the brain, with imprinted genes over-represented in
349 neurons at every level tested and a marked enrichment in neuroendocrine cells lineages. Within-brain
350 analyses revealed that the hypothalamus and the monoaminergic system of the mid- and hindbrain were
351 foci for imprinted gene enrichment. While not all imprinted genes follow these patterns of expression,
352 these findings highlight collective gene expression which is non-random in nature. As such, these
353 analyses identify 'expression hotspots', which in turn suggest 'functional hotspots'. Specifically, our
354 results at the systems and cellular level highlight a major role for imprinted genes in the neuronal
355 regulation of feeding, parental behaviour and sleep.

356 Some of the earliest studies of genomic imprinting identified the brain as a key area for imprinted gene
357 expression (Allen et al. 1995; Keverne et al. 1996). However, it is estimated that ~80% of the genome
358 is expressed in the brain and consequently, imprinted gene expression here may not be a purposeful
359 phenomenon. Our current analysis definitively show that imprinted genes were significantly over-
360 represented in the brain as a whole. This over-representation was found again with PEGs alone, but not
361 MEGs. Within specific brain regions, imprinted genes were over-represented in the hypothalamus,
362 ventral midbrain, pons and medulla. This confirms some previous findings from studies of Pg/Gg and
363 Ag chimera studies (Allen et al. 1995; Keverne et al. 1996) and summaries of imprinted gene expression

364 (Gregg et al. 2010). However, unlike these earlier studies, our analyses do not simply ask if imprinted
365 genes are expressed (at any level) or not, but robustly test whether this expression is meaningful, and
366 the expression of these genes are especially enriched in any given brain region. Additionally, in the
367 chimera studies, Pg/Gg cells with two maternal genomes preferentially allocated to the developing adult
368 cortex and hippocampus, and Ag cells with two paternal genomes preferentially allocated to the
369 developing hypothalamus and midbrain. Our analysis does not reproduce this distinct pattern of MEG
370 and PEG expression in the brain, and indeed we find no specific enrichment of imprinted genes in cortex
371 or hippocampus. Although the pattern of regional enrichment seen with all imprinted genes is replicated
372 when analysing PEGs alone, separate analysis of MEGs only shows over-representation in the pons and
373 medulla. This difference between our analysis of enrichment and the Pg/Gg and Ag chimeras studies
374 indicate that the distribution of Pg/Gg and Ag cells in the brain is not driven by, or indeed reflective of
375 adult PEG and MEG expression, but instead is probably determined by expression of specific imprinted
376 genes during brain development (Davies et al. 2005).
377 At the whole brain level, mature neurons and, in particular, neural-lineage neuroendocrine cells had
378 disproportionately higher numbers of imprinted genes expressed, and high levels of imprinted gene
379 expression. It is likely that this neural-lineage neuroendocrine population comprises members of the
380 key hypothalamic populations in which the expression of imprinted genes are enriched and, when
381 treated as their own cluster, demonstrate strong imprinted gene enrichment compared to other cell
382 lineages of the brain, even other mature neurons.
383 Within the hypothalamus, a selection of informative neuronal subpopulations were over-represented.
384 Strikingly, and suggestive of meaningful enrichment, we saw convergence across our different levels
385 of analysis with several key neuronal types identified in the whole hypothalamus and/or hypothalamic-
386 region-level analysis, already having been identified against the background of general imprinted gene
387 expression in the whole-brain-level analysis. These subpopulations are collectively associated with a
388 few fundamental motivated behaviours. Feeding behaviour is directly controlled by two Arcuate
389 Nucleus populations. The *Agrp* neurons, as feeding promotors, were enriched for imprinted gene
390 expression at the whole brain level, in both whole hypothalamic level datasets and within the Arcuate
391 nucleus; and the *Pomc* neurons as feedings suppressants (enriched with the ARC dataset only) (Aponte

392 et al. 2011). These neurons have already been associated with a few imprinted genes previously (*Asb4,*
393 *Magel2, Snord116*) (Cassidy and Charalambous 2018) but never as an enriched population. Circadian
394 processes are controlled principally by the Suprachiasmatic Nucleus and here we find strong imprinted
395 gene enrichment in the circadian regulatory population with Avp/Nms neurons enriched at the whole
396 brain level, the whole hypothalamic level and as the only enriched neuronal type within the SCN. This
397 population is of interest given the growing appreciation of the role imprinted genes play in circadian
398 processes and the SCN suggested by studies of individual imprinted genes (Tucci 2016). Pituitary
399 endocrine regulation also emerged as a key function, considering the over-representation in the
400 dopaminergic: *Th/Slc6a3/Prlr* neuron type (top hit in the ARC and across dopaminergic neurons of the
401 brain) and the *Ghrh/Gal* subpopulation. These neurons both exist in the arcuate nucleus and regulate
402 prolactin and growth hormone release, respectively. Finally, neuronal control of parenting behaviour
403 was also suggested from the enrichment we saw in the *Gal/Esr1* neuronal population of the Preoptic
404 Area (enriched at whole brain, whole hypothalamic and at POA level). This neural population is known
405 to be associated with feeding and, principally, parenting behaviour (Moffitt et al. 2018). In addition
406 prolactin (regulated by the TIDA neurons of the ARC) is well-known as a mediator of maternal-infant
407 interaction priming the maternal brain during pregnancy for the nurturing of pups (Grattan et al. 2008).
408 In this analysis the hypothalamus was a clear hot spot for imprinted gene expression, in line with the
409 prevailing view of imprinted gene and hypothalamic function (Ivanova and Kelsey 2011; Pulix and
410 Plagge 2020). However, outside of the hypothalamus other distinct hotspot emerged from our whole
411 brain analysis including the monoaminergic system of the midbrain/hindbrain. Analysing the dorsal
412 raphe nucleus and the ventral midbrain revealed the serotonergic neurons to be a foci of imprinted gene
413 expression within this region; more specifically the serotonergic neurons that project to the subcortical
414 regions of the brain known to be associated with feeding and other motivated behaviours (Donovan and
415 Tecott 2013), providing function convergence with the functional hotspots seen in the hypothalamus.
416 In contrast to the specificity of the imprinted gene enrichment in serotonergic neurons, the enrichment
417 seen for adult midbrain dopamine neurons was more generalised. Midbrain dopamine neurons were
418 enriched in adult and developmental ventral midbrain analyses, and sub-type specific enrichment
419 centred on the DAT^{high} lineage of midbrain DA neurons. These DAT^{high} neurons amount to the standard

420 dopamine neurons (*Th/Slc6a3* high) and are found throughout the SNc, VTA and PAG (Tiklová et al.
421 2019), although enrichment was only found in neurons of the latter two nuclei.
422 Analyses of these kind are always bound by the available data and therefore there are notable limitations
423 and caveats to this study. The aim of this study was to generate information about 'hotspots' of
424 imprinted gene expression. This approach, and the use of over-representation analysis and GSEA,
425 therefore do not provide an exhaustive list of sites of expression. For example, oxytocin neurons of the
426 hypothalamus were not significantly enriched in any of the hypothalamic datasets despite a clear
427 oxytocin neuron phenotype reported in a handful of imprinted gene models (Li et al. 1999; Dombret et
428 al. 2012). This may be an example of a functional effect occurring below the level of over-
429 representation, or that imprinted genes act during development and are not functionally enriched in
430 adult oxytocin neurons, or simply that compared to other hypothalamic neuronal populations, oxytocin
431 neurons are not a 'hotspot' of imprinted expression. Specific sequencing of oxytocinergic brain regions
432 will be required to distinguish between these possibilities. A second caveat is that, due to the nature of
433 the datasets used, not all imprinted genes were included, and our analysis was missing a significant
434 subset of imprinted genes encoding small RNAs or isoforms from the same transcription unit. A third
435 caveat is that we did not assess parent-of-origin expression for the 119 imprinted genes we included in
436 the analysis. Previous expression profiling of imprinted genes have also not measured the POEs
437 (Steinhoff et al. 2009; Gregg et al. 2010) but have restricted their gene selection to genes with reliable
438 imprinting status. Consequently, we only included the canonical imprinted genes and genes with more
439 than one demonstration of a POE when looking for enrichment. Furthermore, for the vast majority of
440 these genes, a brain-based POE effect has also already been reported (Supplemental Table S1).
441 Although this does not replace validating the imprinting status of all 119 in the tissues and subregions
442 examined, it does provide justification for looking at imprinted gene over-representation. To resolve
443 this issue, scRNA-seq using hybrid tissues will be key; for example, the recent work of Laukoter et al.
444 (2020) with cortical cell types provides an example of the allelic specific single-cell expression
445 measurements necessary to confirm the enrichments found in this study.
446 By exploiting scRNA-seq data we have asked whether imprinted genes as a group are disproportionately
447 represented in the brain, in specific brain regions, and in certain neuronal cell-types. In the adult brain

448 imprinted genes were over-represented in neurons, and particularly the hypothalamic neuroendocrine
449 populations and the monoaminergic hindbrain neurons, with the serotonergic neurons demonstrating
450 the clearest signal. Interestingly, PEGs, but not MEGs recreate this signal at Levels 1 and 2 - most
451 notably only PEGs display the hypothalamic neuronal enrichment. By extension, these data also identify
452 behaviours that are foci for the action of imprinted genes. Although there are high profile examples of
453 individual imprinted genes expressed in the key brain regions we highlight and that have roles in feeding
454 (*Magel2*) (Schaller et al. 2010), parental behaviour (*Mest)* (Lefebvre et al. 1998) and sleep (*Snord116*)
455 (Lassi et al. 2016), our analyses indicate that imprinted genes *as a group* are strongly linked to these
456 behaviours and also identify other individual genes that should be explored in these domains.
457 Conversely, there are high-profile examples of imprinted genes involved in hippocampus related
458 learning and memory (*Ube3a*) (Jiang et al. 1998), but we did not find enrichment for cell types related
459 to this brain function. The idea that imprinted genes converge on specific physiological or behavioural
460 processes is not unprecedented. Specialisation of function is predicted when considering why genomic
461 imprinting evolved at all (Moore and Haig 1991; Keverne et al. 1996b; Trivers and Burt 1999; Keverne
462 2014). Moreover, there is increasing evidence that the imprinted genes themselves appear to be co-
463 expressed in an imprinted gene network (IGN) and have confirmed regulatory links between each other
464 (Varrault et al. 2006; Gabory et al. 2009; Al Adhami et al. 2015). The idea of an IGN or, at the very
465 least, heavily correlated and coordinated expression between imprinted genes adds further support to
466 the idea that imprinted genes work in concert to influence processes, rather than in isolation, and that
467 perturbating one may influence many others (Patten et al. 2016). Our findings add substance to these
468 general ideas and highlight the neuronal regulation of feeding, parental behaviour and sleep as being
469 key functional hotspots on which imprinted genes converge which probably provides the best current
470 basis for discerning evolutionary drivers of genomic imprinting in the brain.

471 **MATERIALS AND METHODS**

472 **Data Processing**

473 15 unique datasets were analysed across the three levels of analysis (see Fig.1) and analyses were
474 conducted on each dataset independently. At each level of analysis, we aimed to be unbiased by using

all the datasets that fitted the scope of that level, but the availability of public scRNA-seq datasets was limited, which prevented us from exploring all avenues (for example, a direct comparison of enrichment between hypothalamic nuclei). All sequencing data were acquired through publicly available resources and each dataset was filtered and normalised according to the original published procedure. Supplemental Table S21 details the basic parameters of each dataset. Once processed, each dataset was run through the same basic workflow (see below and Fig. 10), with minor adjustments laid out for each dataset detailed in the Supplemental Methods

Due to the high variability in sequencing technology, mouse strain, sex and age, and processing pipeline, we have avoided doing analysis on combined datasets. Rather we chose to perform our analyses independently for each dataset and look for convergent patterns of imprinted gene enrichment between datasets on similar tissues/brain regions. As with any single-cell experiment, the identification of upregulation or over-representation of genes in a cell-type depends heavily on which other cells are included in the analysis to make up the 'background'. Analysing separate datasets (with overlapping cell-types alongside distinct ones) and looking for convergent patterns of enrichment is one way of counteracting this limitation.

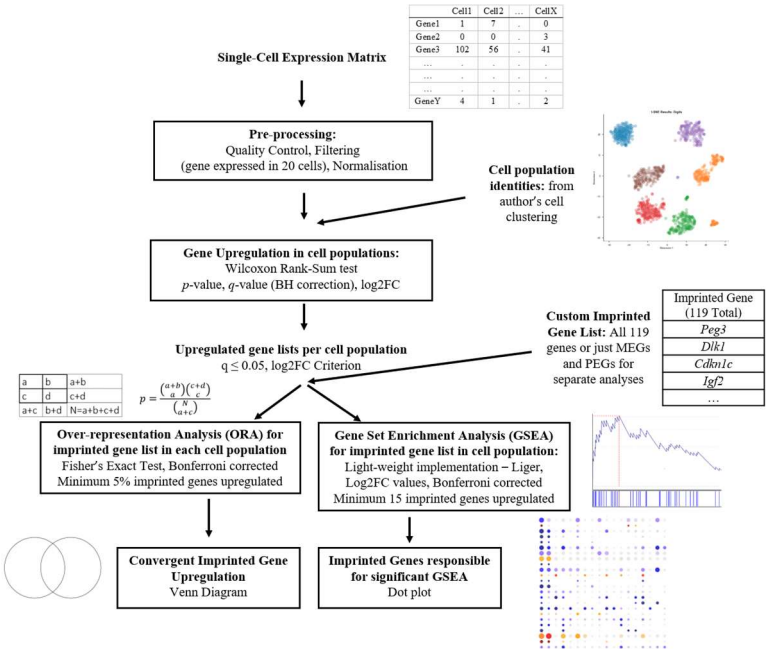


Figure 10. Basic workflow schematic. Single Cell Expression Matrices were acquired through publicly available depositories. Data were processed according to the author's original specifications and all genes were required to be expressed in 20 or more cells. Cell population identities were acquired from the author's original clustering. Positive differential gene expression was calculated via Wilcoxon Rank-Sum Test. Upregulated genes were considered as those with $q \leq 0.05$ and a Log2FC ≥ 1 for analysis levels 1 and 2, while this criterion was relaxed to Log2FC > 0 for level 3. Our imprinted gene list was used to filter upregulated genes and two different enrichment analyses were carried out, over-representation analysis via Fisher's Exact Test and Gene Set Enrichment Analysis via Liger algorithm (Subramanian et al.2005, https://github.com/JEFworks/liger). Venn diagrams and dot plots were utilised for visualisation.

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

503 Supplemental Table S21. A consistent filter, to remove all genes expressed in fewer than 20 cells, was
504 applied to remove genes unlikely to play a functional role due to being sparsely expressed. Datasets of
505 the whole brain/hypothalamus were analysed both at the global cell level (neuronal and non-neuronal
506 cells) and neuron specific level (only neurons) with genes filtered for the ≥ 20 cell expression at each
507 level before subsequent analysis. Cell identities were supplied using the outcome of cell clustering
508 carried out by the original authors, so that each cell included in the analysis had a cell-type or tissue-
509 type identity. This was acquired as metadata supplied with the dataset or as a separate file primarily
510 from the same depository as the data but occasionally acquired from personal correspondence with the
511 authors. Cells were used from mice of both sexes when provided and all mice were aged 15 weeks or
512 younger across all datasets. Although our focus was the adult mouse brain, embryonic data were
513 included in some comparisons or when no alternatives were present. However, embryonic and post-
514 natal cells were never pooled to contribute to the same cell populations.
515 Positive differential expression between identity groups were carried out using one-sided Wilcoxon
516 rank-sum tests (assuming the average expression of cells within the current identity group is 'greater'
517 than the average of cells from all other groups). The test was performed independently for each gene
518 and for each identity group vs. all other groups. The large number of p values were corrected for
519 multiple comparisons using a horizontal Benjamini-Hochberg correction, creating q values. Fold-
520 change (FC) values, percentage expression within the identity group and percentage expressed within
521 the rest were also calculated. We considered genes to be significantly positively differentially
522 expressed (significantly upregulated) in a group compared to background expression if it had a $q \leq$
523 0.05. In addition, for Level 1 and Level 2 analyses, the criteria for upregulated genes included
524 demonstrating a Log2FC value of 1 or larger (i.e., 2-fold-change or larger). The datasets at these
525 levels represented cells from a variety of organs, regions and cell-types, and in line with this cellular
526 diversity, the aim of these analyses was to look for distinctive upregulation, akin to a marker gene.
527 Once the analysis was restricted to cell subpopulations within a specific region of the brain (i.e., Level
528 3), the additional criteria for upregulation was relaxed to demonstrating just a positive Log2FC (i.e.,
529 the gene has a higher expression in this cell type than background). This was mainly because we were
530 not expecting imprinted genes to be 'markers' of individual subpopulations at this level, but our aim

531 was to identify enriched expression profiles for them. This additionally ensures consistent criteria for
532 enrichment within levels, allowing meaningful comparison.
533 The same custom list of imprinted genes with reliable parent-of-origin effects (see below) was used
534 for all analyses, and all genes were included as long as the gene passed the 20-cell filter. The first
535 statistical analysis for enrichment was an Over-Representation Analysis (ORA) using a one-sided
536 Fisher's Exact Test ('fisher.test' function in R core package 'stats v3.6.2'). The aim was to assess
537 whether the number of imprinted genes considered to be upregulated as a proportion of the total
538 number of imprinted genes in the dataset (passing the 20-cell filter) was statistically higher than
539 would be expected by chance when compared to the total number of upregulated genes as a
540 proportion of the overall number of genes in the dataset (passing the 20-cell filter). To limit finding
541 over-represented identity groups with only a few upregulated imprinted genes, an identity group was
542 required to have ≥ 5 % of the total number of imprinted genes upregulated for ORA to be conducted.
543 Subsequent p-values for all eligible identity groups were corrected using a Bonferroni correction. This
544 provided a measure of whether imprinted genes are expressed above expectation (as opposed to the
545 expression pattern of any random gene selection) in particular identity groups.
546 Venn diagrams of the upregulated imprinted genes making up over-represented identity groups across
547 datasets (within a level) were also reported. Full lists of upregulated imprinted genes can be found in
548 the 'Upregulated_IGs.csv' file for each analysis in the Supplemental Data.
549 To further examine the presence of imprinted genes within tissues/cell types, and to provide a
550 different perspective to over-representation, we conducted a Gene-Set Enrichment Analysis (GSEA)
551 for imprinted genes amongst the upregulated genes of an identity group using a publicly available,
552 light-weight implementation of the GSEA algorithm (Subramanian et al. 2005) in R
553 (https://github.com/JEFworks/liger). This was done in a manner similar to Moffitt et al. (2018) since
554 we were similarly using this computational method to identify enrichment of our gene sets *within the*
555 *upregulated genes* of the different identity groups. Here, the GSEA was conducted for each individual
556 identity group using Log2FC values to rank the upregulated genes. The GSEA acts as a more
557 conservative measure than the ORA since it tests whether imprinted genes are enriched in the stronger
558 markers of a group (the genes with the highest fold change for a group vs. the rest) and hence whether

559 the imprinted genes are enriched in those genes with a high specificity to that tissue/cell type. To
560 prevent significant results being generated from just 2 or 3 genes, identity group to be analysed were
561 selected as having a minimum of 15 upregulated imprinted genes (i.e. the custom gene set) to measure
562 enrichment for (a value suggested by the GSEA user guide (https://www.gsea-
563 msigdb.org/gsea/doc/GSEAUserGuide Frame.html)) and to prevent significant results in which
564 imprinted genes cluster at the tail, identity groups were selected as having an average fold change of
565 the upregulated imprinted genes greater than the average fold change of the rest of the upregulated
566 genes for that group. Again, multiple p values generated from GSEA were corrected using a
567 Bonferroni correction. To further elucidate the genes responsible for significant GSEA's, dot plots of
568 the imprinted genes upregulated in that identity group were plotted across all identity groups with
569 absolute expression and Log2FC mapped to size and colour of the dots, respectively. Graphical
570 representations of significant GSEA's (post-correction) are included in the main text or as
571 supplemental figures, all other graphs, including additional dot plots not discussed in this study, can
572 be found in the repository (https://osf.io/jx7kr/) and Supplemental Data. If no cell populations met
573 these criteria, GSEA was not run and not included for that analysis.
574 For Level 1 and Level 2 analyses, we also carried out parent-of-origin specific analyses. The
575 imprinted gene list was divided into MEGs and PEGs and the analyses detailed above were run
576 separately for these two gene groups. For imprinted genes with known parent-of-origin variability
577 based on tissue type (*Igf2* and *Grb10*), the parent-of-origin characterisation of these genes was
578 changed accordingly. The absolute number of imprinted genes top-expressed in a tissue/cell-type were
579 also reported for analyses in Level 1 and Level 2 in the tables, since these analyses included a variety
580 of cell-types and tissues which may demonstrate meaningful clustering of the highest normalised
581 expression values. The mean normalised expression for all imprinted genes across the series of
582 identity groups in the datasets in Level 1 and Level 2 was also calculated alongside the mean
583 normalised expression for the rest of the genes (Supplemental Table S2).
584 All graphical representations and statistical analyses were conducted using R 3.6.2 (Team 2013) in
585 RStudio (Team 2015). Diagrams in Figures 1:2, 4:6 and 8:10 were created with BioRender.com.

586 **Custom Imprinted Gene List**

587 The gene list for the analysis was based on the list of murine imprinted genes recently published in
588 Tucci et al. (2019). Although the original list of imprinted genes was 260 genes long, only 163 genes
589 were identified in the most comprehensive of the datasets. We further refined this list to 119 imprinted
590 genes (Supplemental Table S1) which excluded the X-linked genes, consisting of mostly the canonical
591 protein-coding and long noncoding RNA imprinted genes, but the criteria for inclusion was those genes
592 with at least two independent demonstrations of their POE status. The only exceptions to multiple
593 independent demonstrations of a POE were four genes (*Bmf*, *B3gnt2*, *Ptk2*, *Gm16299*) identified by
594 (Perez et al. 2015) where a POE was validated for 16 brain regions and 7 adult tissues. For Level 2, the
595 MEG/PEG status of a gene was primarily based on reported allelic expression within the brain. Small
596 non-coding RNAs such as micro-RNAs (miRs) and small nucleolar RNAs (snoRNAs), which represent
597 ~10% of identified imprinted genes, were excluded from the analysis as their sequences were not
598 detected/subsumed by larger transcripts in the majority of the datasets. Another caveat with short-read
599 RNA-seq libraries is that much of the expression data for a given transcription unit cannot discriminate
600 differentially imprinted isoforms nor do some of the technologies (e.g., Smart-Seq2) possess stranded
601 libraries to distinguish antisense transcripts. For complex imprinting loci such as the *Gnas* locus, most
602 reads as result map to only *Gnas* and *Nespas* ignoring several overlapping and antisense genes.

603 **Data Availability**

604 All sequencing data used in this project were acquired through publicly available resources
605 (Databases, Accession Numbers and Depositories for each dataset can be found in the Supplemental
606 Methods). All data generated in this experiment is provided as Supplemental Data and in the
607 following Open Science Framework repository (https://osf.io/jx7kr/). Custom R scripts to analyse
608 each dataset are provided as Supplemental Code and are available at https://github.com/MJHiggs/IG-
609 Single-Cell-Enrichment.

610 **Competing interests**

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

612 **Acknowledgments**

This work was supported by a Wellcome Trust PhD studentship (220090/Z/20/Z). We would like to thank all the research groups that carried out the single-cell RNA sequencing that made this study possible and to particularly acknowledge Dr. Paul Hook, Dr. Alina Isakova, Dr. Alexander Jackson, Dr. Katarína Tiklová, and Dr. Shao'Ang Wen for help accessing the cell metadata for their associated studies. *Author contributions*: M.Higgs performed bioinformatic analysis, with input from M.Hill; M.Higgs., and A.Isles contributed to project design, data interpretation, and wrote the manuscript; all co-authors reviewed and edited the manuscript.

References

Al Adhami H, Evano B, Le Digarcher A, Gueydan C, Dubois E, Parrinello H, Dantec C, Bouschet T, Varrault A, 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**: 353-367.

Allen ND, Logan K, Lally G, Drage DJ, Norris ML, Keverne EB. 1995. Distribution of parthenogenetic cells in the mouse brain and their influence on brain development and behavior. *Proceedings of the National Academy of Sciences* **92**: 10782-10786.

Andergassen D, Dotter CP, Wenzel D, Sigl V, Bammer PC, Muckenhuber M, Mayer D, Kulinski TM, Theussl H-C, Penninger JM. 2017. Mapping the mouse Allelome reveals tissue-specific regulation of allelic expression. *Elife* **6**: e25125.

Angulo M, Butler M, Cataletto M. 2015. Prader-Willi syndrome: a review of clinical, genetic, and endocrine findings. *Journal of endocrinological investigation* **38**: 1249-1263.

Aponte Y, Atasoy D, Sternson SM. 2011. AGRP neurons are sufficient to orchestrate feeding behavior rapidly and without training. *Nature neuroscience* **14**: 351.

Babak T, DeVeale B, Tsang EK, Zhou Y, Li X, Smith KS, Kukurba KR, Zhang R, Li JB, van der Kooy D. 2015. Genetic conflict reflected in tissue-specific maps of genomic imprinting in human and mouse. *Nature genetics* **47**: 544-549.

Campbell JN, Macosko EZ, Fenselau H, Pers TH, Lyubetskaya A, Tenen D, Goldman M, Verstegen AM, Resch JM, McCarroll SA. 2017. A molecular census of arcuate hypothalamus and median eminence cell types. *Nature neuroscience* **20**: 484-496.

Cassidy FC, Charalambous M. 2018. Genomic imprinting, growth and maternal–fetal interactions. *Journal of Experimental Biology* **221**.

Cattanach BM, Kirk M. 1985. Differential activity of maternally and paternally derived chromosome regions in mice. *Nature* **315**: 496-498.

Chen R, Wu X, Jiang L, Zhang Y. 2017. Single-cell RNA-seq reveals hypothalamic cell diversity. *Cell reports* **18**: 3227-3241.

Davies JR, Humby T, Dwyer DM, Garfield AS, Furby H, Wilkinson LS, Wells T, Isles AR. 2015. Calorie seeking, but not hedonic response, contributes to hyperphagia in a mouse model for Prader–Willi syndrome. *European Journal of Neuroscience* **42**: 2105-2113.

Davies W, Isles AR, Wilkinson LS. 2005. Imprinted gene expression in the brain. *Neuroscience & Biobehavioral Reviews* **29**: 421-430.

Dent CL, Humby T, Lewis K, Ward A, Fischer-Colbrie R, Wilkinson LS, Wilkins JF, Isles AR. 2018. Impulsive choice in mice lacking paternal expression of Grb10 suggests intragenomic conflict in behavior. *Genetics* **209**: 233-239.

DeVeale B, Van Der Kooy D, Babak T. 2012. Critical evaluation of imprinted gene expression by RNA–Seq: a new perspective. *PLoS Genet* **8**: e1002600.

Dombret C, Nguyen T, Schakman O, Michaud JL, Hardin-Pouzet H, Bertrand MJ, 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**: 4703-4717.

Donovan MH, Tecott LH. 2013. Serotonin and the regulation of mammalian energy balance. *Frontiers in neuroscience* **7**: 36.

Ferguson-Smith AC. 2011. Genomic imprinting: the emergence of an epigenetic paradigm. *Nature Reviews Genetics* **12**: 565-575.

Gabory A, Ripoche M-A, Le Digarcher A, Watrin F, Ziyyat A, Forné T, Jammes H, Ainscough JF, Surani MA, Journot L. 2009. H19 acts as a trans regulator of the imprinted gene network controlling growth in mice. *Development* **136**: 3413-3421.

Garfield AS, Cowley M, Smith FM, Moorwood K, Stewart-Cox JE, Gilroy K, Baker S, Xia J, Dalley JW, Hurst LD. 2011. Distinct physiological and behavioural functions for parental alleles of imprinted Grb10. *Nature* **469**: 534-538.

Grattan DR, Steyn FJ, Kokay IC, Anderson GM, Bunn SJ. 2008. Pregnancy-induced adaptation in the neuroendocrine control of prolactin secretion. *Journal of neuroendocrinology* **20**: 497-507.

Gregg C, Zhang J, Weissbourd B, Luo S, Schroth GP, Haig D, Dulac C. 2010. High-resolution analysis of parent-of-origin allelic expression in the mouse brain. *science* **329**: 643-648.

Han X, Wang R, Zhou Y, Fei L, Sun H, Lai S, Saadatpour A, Zhou Z, Chen H, Ye F. 2018. Mapping the mouse cell atlas by microwell-seq. *Cell* **172**: 1091-1107. e1017.

Hook PW, McClymont SA, Cannon GH, Law WD, Morton AJ, Goff LA, McCallion AS. 2018. Single-cell RNA-seq of mouse dopaminergic neurons informs candidate gene selection for sporadic Parkinson disease. *The American Journal of Human Genetics* **102**: 427-446.

Huang KW, Ochandarena NE, Philson AC, Hyun M, Birnbaum JE, Cicconet M, Sabatini BL. 2019. Molecular and anatomical organization of the dorsal raphe nucleus. *Elife* **8**: e46464.

Ivanova E, Kelsey G. 2011. Imprinted genes and hypothalamic function. *Journal of molecular endocrinology* **47**: R67-R74.

Jiang Y-h, Armstrong D, Albrecht U, Atkins CM, Noebels JL, Eichele G, Sweatt JD, Beaudet AL. 1998. Mutation of the Angelman ubiquitin ligase in mice causes increased cytoplasmic p53 and deficits of contextual learning and long-term potentiation. *Neuron* **21**: 799-811.

Keverne E. 2014. Significance of epigenetics for understanding brain development, brain evolution and behaviour. *Neuroscience* **264**: 207-217.

Keverne EB, Fundele R, Narasimha M, Barton SC, Surani MA. 1996a. Genomic imprinting and the differential roles of parental genomes in brain development. *Developmental Brain Research* **92**: 91-100.

Keverne EB, Martel FL, Nevison CM. 1996b. Primate brain evolution: genetic and functional considerations. *Proceedings of the Royal Society of London Series B: Biological Sciences* **263**: 689-696.

La Manno G, Gyllborg D, Codeluppi S, Nishimura K, Salto C, Zeisel A, Borm LE, Stott SR, Toledo EM, Villaescusa JC. 2016. Molecular diversity of midbrain development in mouse, human, and stem cells. *Cell* **167**: 566-580. e519.

Lassi G, Ball ST, Maggi S, Colonna G, Nieus T, Cero C, Bartolomucci A, Peters J, Tucci V. 2012. Loss of Gnas imprinting differentially affects REM/NREM sleep and cognition in mice. *PLoS Genet* **8**: e1002706.

Lassi G, Priano L, Maggi S, Garcia-Garcia C, Balzani E, El-Assawy N, Pagani M, Tinarelli F, Giardino D, Mauro A. 2016. Deletion of the Snord116/SNORD116 alters sleep in mice and patients with Prader-Willi syndrome. *Sleep* **39**: 637-644.

Laukoter S, Pauler FM, Beattie R, Amberg N, Hansen AH, Streicher C, Penz T, Bock C, Hippenmeyer S. 2020. Cell-type specificity of genomic imprinting in cerebral cortex. *Neuron* **107**: 1160-1179. e1169.

Lefebvre L, Viville S, Barton SC, Ishino F, Keverne EB, Surani MA. 1998. Abnormal maternal behaviour and growth retardation associated with loss of the imprinted gene Mest. *Nature genetics* **20**: 163-169.

Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A, Boe AF, Boguski MS, Brockway KS, Byrnes EJ. 2007. Genome-wide atlas of gene expression in the adult mouse brain. *Nature* **445**: 168-176.

Li L-L, Keverne E, Aparicio S, Ishino F, Barton S, Surani M. 1999. Regulation of maternal behavior and offspring growth by paternally expressed Peg3. *Science* **284**: 330-334.

McNamara GI, John RM, Isles AR. 2018. Territorial behavior and social stability in the mouse require correct expression of imprinted Cdkn1c. *Frontiers in behavioral neuroscience* **12**: 28.

Mickelsen LE, Bolisetty M, Chimileski BR, Fujita A, Beltrami EJ, Costanzo JT, Naparstek JR, Robson P, Jackson AC. 2019. Single-cell transcriptomic analysis of the lateral hypothalamic area reveals molecularly distinct populations of inhibitory and excitatory neurons. *Nature neuroscience* **22**: 642-656.

Moffitt JR, Bambah-Mukku D, Eichhorn SW, Vaughn E, Shekhar K, Perez JD, Rubinstein ND, Hao J, Regev A, Dulac C. 2018. Molecular, spatial, and functional single-cell profiling of the hypothalamic preoptic region. *Science* **362**: eaau5324.

Moore T, Haig D. 1991. Genomic imprinting in mammalian development: a parental tug-of-war. *Trends in genetics* **7**: 45-49.

Negi SK, Guda C. 2017. Global gene expression profiling of healthy human brain and its application in studying neurological disorders. *Scientific reports* **7**: 1-12.

Nicholls RD, Knoll JH, Butler MG, Karam S, Lalande M. 1989. Genetic imprinting suggested by maternal heterodisomy in non-deletion Prader-Willi syndrome. *Nature* **342**: 281-285.

724 Orr HA. 1995. Somatic mutation favors the evolution of diploidy. *Genetics* **139**: 1441-1447.
725 Patten MM, Cowley M, Oakey RJ, Feil R. 2016. Regulatory links between imprinted genes: evolutionary
726 predictions and consequences. *Proceedings of the Royal Society B: Biological Sciences* **283**: 20152760.
727 Perez JD, Rubinstein ND, Dulac C. 2016. New Perspectives on Genomic Imprinting, an Essential and
728 Multifaceted Mode of Epigenetic Control in the Developing and Adult Brain. *Annual Review of*
729 *Neuroscience* **39**: 347-384.
730 Perez JD, Rubinstein ND, Fernandez DE, Santoro SW, Needleman LA, Ho-Shing O, Choi JJ, Zirlinger M, Chen
731 S-K, Liu JS. 2015. Quantitative and functional interrogation of parent-of-origin allelic expression
732 biases in the brain. *Elife* **4**: e07860.
733 Peters J. 2014. The role of genomic imprinting in biology and disease: an expanding view. *Nature Reviews*
734 *Genetics* **15**: 517-530.
735 Pulix M, Plagge A. 2020. Imprinted Genes and Hypothalamic Function. In *Developmental Neuroendocrinology*,
736 pp. 265-294. Springer.
737 Qualls-Creekmore E, Yu S, Francois M, Hoang J, Huesing C, Bruce-Keller A, Burk D, Berthoud H-R, Morrison
738 CD, Münzberg H. 2017. Galanin-expressing GABA neurons in the lateral hypothalamus modulate food
739 reward and noncompulsive locomotion. *Journal of Neuroscience* **37**: 6053-6065.
740 Rau AR, Hentges ST. 2017. The relevance of AgRP neuron-derived GABA inputs to POMC neurons differs for
741 spontaneous and evoked release. *Journal of Neuroscience* **37**: 7362-7372.
742 Relkovic D, Doe CM, Humby T, Johnstone KA, Resnick JL, Holland AJ, Hagan JJ, Wilkinson LS, Isles AR.
743 2010. Behavioural and cognitive abnormalities in an imprinting centre deletion mouse model for
744 Prader–Willi syndrome. *European journal of neuroscience* **31**: 156-164.
745 Ren J, Isakova A, Friedmann D, Zeng J, Grutzner SM, Pun A, Zhao GQ, Kolluru SS, Wang R, Lin R. 2019.
746 Single-cell transcriptomes and whole-brain projections of serotonin neurons in the mouse dorsal and
747 median raphe nuclei. *Elife* **8**: e49424.
748 Romanov RA, Zeisel A, Bakker J, Girach F, Hellysaz A, Tomer R, Alpár A, Mulder J, Clotman F, Keimpema
749 E. 2017. Molecular interrogation of hypothalamic organization reveals distinct dopamine neuronal
750 subtypes. *Nature neuroscience* **20**: 176-188.
751 Schaller F, Watrin F, Sturny R, Massacrier A, Szepetowski P, Muscatelli F. 2010. A single postnatal injection of
752 oxytocin rescues the lethal feeding behaviour in mouse newborns deficient for the imprinted Magel2
753 gene. *Human Molecular Genetics* **19**: 4895-4905.
754 Schaum N, Karkanias J, Neff NF, May AP, Quake SR, Wyss-Coray T, Darmanis S, Batson J, Botvinnik O,
755 Chen MB. 2018. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris: The Tabula
756 Muris Consortium. *Nature* **562**: 367.
757 Steinhoff C, Paulsen M, Kielbasa S, Walter J, Vingron M. 2009. Expression profile and transcription factor
758 binding site exploration of imprinted genes in human and mouse. *BMC Genomics* **10**: 144.
759 Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL,
760 Golub TR, Lander ES. 2005. Gene set enrichment analysis: a knowledge-based approach for
761 interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**:
762 15545-15550.
763 Team R. 2015. RStudio: integrated development for R. *RStudio, Inc, Boston, MA URL http://wwwrstudio.com*
764 **42**: 14.
765 Team RC. 2013. R: A language and environment for statistical computing. Vienna, Austria.
766 Tiklová K, Björklund ÅK, Lahti L, Fiorenzano A, Nolbrant S, Gillberg L, Volakakis N, Yokota C, Hilscher
767 MM, Hauling T. 2019. Single-cell RNA sequencing reveals midbrain dopamine neuron diversity
768 emerging during mouse brain development. *Nature communications* **10**: 1-12.
769 Trivers R, Burt A. 1999. Kinship and genomic imprinting. In *Genomic imprinting*, pp. 1-21. Springer.
770 Tucci V. 2016. Genomic imprinting: a new epigenetic perspective of sleep regulation. *PLoS genetics* **12**:
771 e1006004.
772 Tucci V, Isles AR, Kelsey G, Ferguson-Smith AC, Bartolomei MS, Benvenisty N, Bourc'his D, Charalambous
773 M, Dulac C, Feil R. 2019. Genomic imprinting and physiological processes in mammals. *Cell* **176**:
774 952-965.
775 Varrault A, Gueydan C, Delalbre A, Bellmann A, Houssami S, Aknin C, Severac D, Chotard L, Kahli M, Le
776 Digarcher A. 2006. Zac1 regulates an imprinted gene network critically involved in the control of
777 embryonic growth. *Developmental cell* **11**: 711-722.
778 Wen S, Ma D, Zhao M, Xie L, Wu Q, Gou L, Zhu C, Fan Y, Wang H, Yan J. 2020. Spatiotemporal single-cell
779 analysis of gene expression in the mouse suprachiasmatic nucleus. *Nature neuroscience* **23**: 456.
780 Ximerakis M, Lipnick SL, Innes BT, Simmons SK, Adiconis X, Dionne D, Mayweather BA, Nguyen L,
781 Niziolek Z, Ozek C. 2019. Single-cell transcriptomic profiling of the aging mouse brain. *Nature*
782 *neuroscience* **22**: 1696-1708.

783 Zeisel A, Hochgerner H, Lönnerberg P, Johnsson A, Memic F, Van Der Zwan J, Häring M, Braun E, Borm LE,
784 La Manno G. 2018. Molecular architecture of the mouse nervous system. *Cell* **174**: 999-1014. e1022.
785 Zhang X, Jiang S, Mitok KA, Li L, Attie AD, Martin TFJ. 2017. BAIAP3, a C2 domain–containing Munc13
786 protein, controls the fate of dense-core vesicles in neuroendocrine cells. *Journal of Cell Biology* **216**:
787 2151-2166.