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 take a systematic approach to investigate imprinted gene over-representation at the organ, brain region, and cell-specific levels. We establish that imprinted genes are indeed over-represented in the adult brain, and in neurons particularly compared to other brain cell-types. We then examine brain-wide datasets to examine enrichment within distinct regions of the brain and demonstrate over-representation of imprinted genes in the hypothalamus, ventral midbrain, pons and medulla. Finally, using 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 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 genes converge. Our findings indicate the neuronal regulation of motivated behaviours such as feeding, parental behaviour and sleep as functional hotspots for imprinting, thus adding statistically 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 200 canonical (associated with a parental specific epigenetic mark) imprinted genes have been identified in the mouse (~150 in humans) (Ferguson-Smith 2011; Tucci et al. 2019). This epigenetic regulation makes genomic imprinting an evolutionary puzzle as 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 characterization 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 suggested by (Cattanach and Kirk 1985) and neurological phenotypes (Keverne et al. 1996a) observed in early imprinted gene mouse models, as well as the behavioural deficits seen in imprinting disorders such as Prader-Willi Syndrome and Angelman's syndrome (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 and 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).

In addition to studies on individual imprinted genes, there are a limited number of studies that take a systems level approach to characterizing the role of genomic imprinting in the brain. The early studies examining developing and adult chimeras of normal and parthenogenetic (two paternal genomes) or androgenetic/gynogenetic (two maternal genomes) cells indicated distinct regional distribution for maternal (cortex and hippocampus) and paternal (hypothalamus) expressed genes (Allen et al. 1995; Keverne et al. 1996a). More recently, Gregg et al. (2010) used the known imprinting status of 45 imprinted genes and the Allen Brain Atlas to track dichotomous expression of imprinted genes across 118 brain regions to identify brain-wide patterns of expression. Most imprinted genes were expressed in every brain region, but detectable expression of the largest number of imprinted genes was found in regions of the hypothalamus (MPOA, ARC), amygdala (CEA, MEA) and the monoaminergic nuclei, suggesting some form of specialisation. Although pioneering, this study, and others identifying novel imprinted genes and/or mapping allelic expression in the brain (DeVeale et al. 2012; Babak et al. 2015; Perez et al. 2015; Andergassen et al. 2017), did not test whether these genes were especially enriched in given brain regions but simply asked if they were expressed (at any level) or not.

Here we address the question of whether the brain and/or specific brain circuitry is a foci for genomic imprinting by exploiting the rapidly expanding number of single-cell RNA

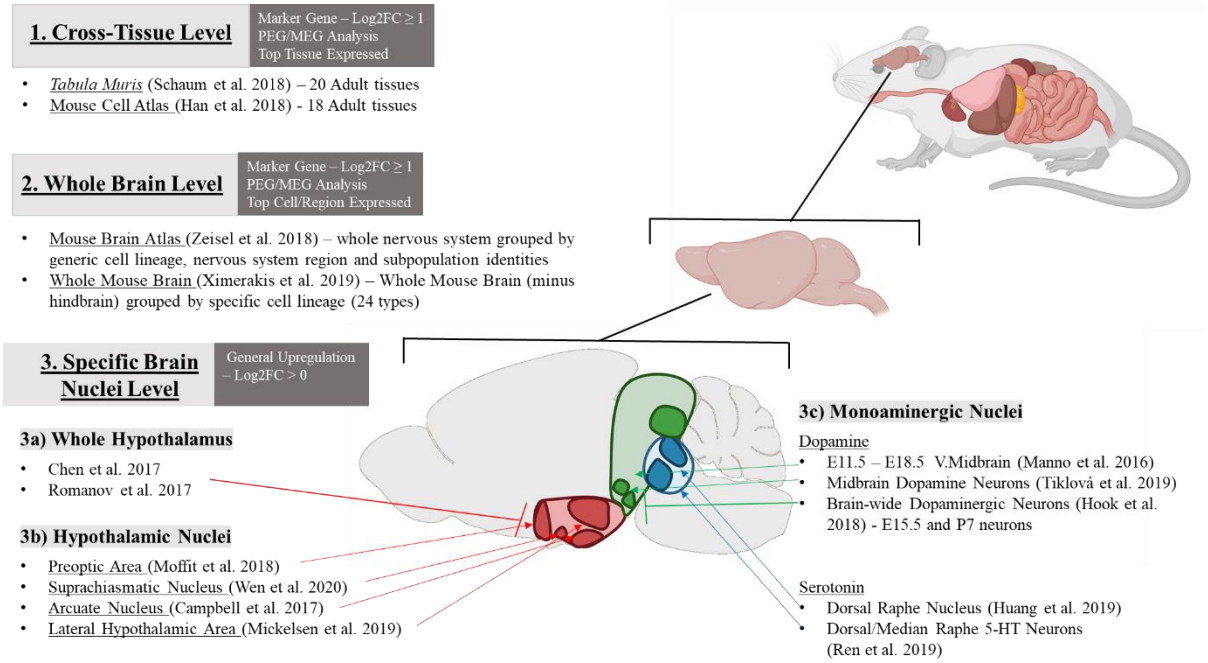
sequencing (scRNA-seq) datasets and systematically investigating imprinted gene enrichment and over-representation in the murine brain. We demonstrate that imprinted genes are indeed enriched in brain relative to other adult tissues, with over-representation in the neurons of the hypothalamus, ventral 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, statistically validating previous assumptions, identifying neuronal domains that have received less emphasis in earlier studies, 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.

Materials and Methods

Given the breadth of single-cell datasets that are publicly available, we sought to perform a hierarchical sequence of data analysis (from a cross-organ comparison to a specific brain-region comparison) with the outcome of each level informing the data selection for the next one (see Figure 1 for summary graphic). We began at level 1 with a cross-tissue level comparison to ask would imprinted genes emerge as a brain enriched group compared to other major organs. Level 2 (whole brain level) sought to clarify that enrichment by exploring whether imprinted genes would demonstrate a regional or cell lineage specific enrichment within the brain. Level 3 (brain-region level) was then guided by the regions/cell types we found as enriched in level 2 to investigate whether those regional enrichments would be further clarified at the cell subpopulation level. At each level of analysis, we aimed to be unbiased by using all the datasets that fitted the aim 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). Where possible two or more independent datasets were used at each level to strengthen any conclusions with

Figure 1. The hierarchical set of datasets in this analysis. The hierarchical set of datasets organised by cross-tissue, whole brain and brain region level analyses. Each analysis includes the original publication and specific tissue/s analysed. 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 were reported for imprinted genes.



convergent findings.

15 unique datasets were analysed across the three levels and analyses were conducted on each dataset independently. Data were acquired through publicly available resources or personal correspondence with the authors and each dataset was filtered and normalised according to the original published procedure. Table 1 details the basic parameters of each dataset. Due to the high variability in sequencing technology, mouse strain, sex and age, and pipeline we chose to perform our analyses independently and hence not perform direct comparisons between datasets but instead look for convergent patterns of imprinted gene enrichment between datasets on similar tissues/brain regions.

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 et al. 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)

Once processed, each dataset was run through the same basic workflow (see below and Figure 2), with minor adjustments laid out for each dataset detailed in the Supplemental Methods. The aim of the analysis differed slightly between the levels. Level 1 and level 2 analyses represented cells from a variety of organs, regions and cell-types and, in line with this cellular diversity, our analysis followed a procedure appropriate for identifying marker genes, meaning we were looking for distinct imprinted gene enrichment and the tissues they displayed ‘marker’ like expression for. Level 3 analysis featured many independent analyses with a reduced scope (i.e. one specific region or sub-region of the brain and often a single

cell-type). We did not anticipate imprinted genes to populate the marker genes at this level and hence our analyses focused on identifying patterns of upregulated genes looking for cell-types that imprinted genes are convergently upregulated in above background, using this relaxed criterion for upregulation.

Custom Imprinted Gene List

The gene list for the analysis was based on the list of murine imprinted genes recently published in Tucci et al. (2019). 119 imprinted genes were chosen for the analysis (Supplemental Table S1) consisting of canonical protein-coding and long noncoding RNA imprinted genes, the criteria for inclusion was those genes with at least two independent demonstrations of their POE status. The only exceptions to multiple independent demonstrations of a POE were four genes (*Bmf*, *B3gnt2*, *Ptk2*, *Gm16299*) identified by (Perez et al. 2015) where a 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. Another caveat with short-read RNA-seq libraries is that much of the expression data for a given transcription unit cannot discriminate differentially imprinted isoforms, such as the Gnas locus, as most of the reads mapping to the locus will map to shared exons. Indeed, only Gnas and Nespas are consistently distinguished from this locus.

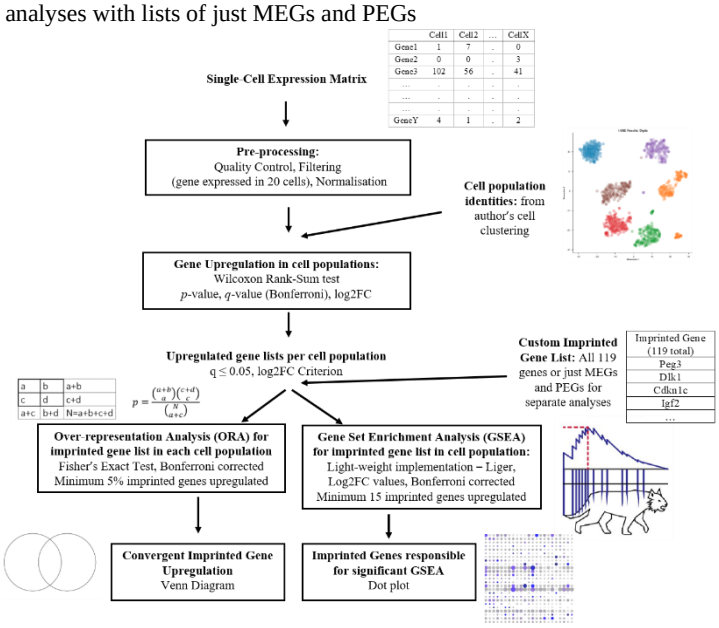
Basic Workflow

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 1. A consistent filter, to keep all genes expressed in at least 20 cells, was applied to remove genes unlikely to play a functional role due to being expressed in only

a few cells. Datasets from the brain were analysed both at the global cell level (neuronal and non-neuronal cells) and neuron specific level (only cells classified as neurons) with genes filtered for the ≥ 20 cell expression at each level before subsequent analysis. Cell identities were supplied using the outcome of cell clustering carried out by the original authors, so that each cell included in the analysis had a cell-type or tissue-type identity. This was acquired as metadata supplied with the dataset or as a separate file from personal correspondence with the author. Unless otherwise stated, all cells that were used to originally cluster the subpopulations were included in this analysis. In all cases this involved data from both sexes and from all post-natal ages (e.g. both young and old mice from Ximerakis et al. 2019). Embryonic and post-natal tissue was never pooled to contribute to the same cell populations. Positive differential expression between identity groups were carried out using one-sided Wilcoxon rank-sum tests (assuming the average expression of cells within the current identity group is 'greater' than the average of cells from all other groups). The test was performed independently for each gene and for each identity group vs. all other groups. The large number of p values were corrected for multiple comparisons using a horizontal Benjamini-Hochberg correction, creating q values. Fold-change (FC) values, percentage expression within the identity group and percentage expressed within the rest were also calculated. We considered genes to be significantly positively differentially expressed (significantly upregulated) in a group compared to background expression if it had a $q \leq 0.05$. In addition, for level 1 and level 2 analysis, the criteria for upregulated genes included demonstrating a Log2FC value of 1 or larger (i.e. 2-fold-change or larger), which selects for genes across a variety of tissue, cell-types and brain regions with distinctive upregulation, akin to a marker gene. Once the analysis was restricted to cell subpopulations within a specific region of the brain (i.e. level 3), the additional criteria for upregulation was relaxed to demonstrating just a positive Log2FC (i.e. the gene has a higher expression in this cell type than background).

This was mainly because we were not expecting imprinted genes to be 'markers' of individual cell types at this level, but our aim was to identify enriched expression profiles for them. This additionally ensures consistent criteria for enrichment within levels, allowing meaningful comparison.

Figure 2. 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 custom imprinted gene list was used to select out upregulated genes that were imprinted 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,chttps://github.com/JEFworks/liger). Additionally, Level 1 and Level 2 Analyses repeated the enrichment analyses with lists of just MEGs and PEGs



The same custom list of imprinted genes with reliable parent-of-origin effects was used for all analyses, and all genes were included as long as the gene passed the 20-cell filter. The first statistical analysis for enrichment was an Over-Representation Analysis (ORA) using a one-way Fisher's Exact Test ('fisher.test' function in R core package 'stats v3.6.2'). The aim was to assess whether the number of imprinted genes considered to be upregulated as a proportion of the total number of imprinted genes in the dataset (passing the 20-cell filter) was statistically higher than would be expected by chance when compared to the total number of upregulated genes as a proportion of the overall number of genes in the dataset (passing the 20-cell filter). To limit finding over-represented identity groups with only a few upregulated imprinted genes, an identity group was required to have ≥ 5 % of the total number of imprinted genes upregulated for ORA to be conducted. Subsequent p-values for all eligible identity groups were corrected using a Bonferroni

correction. This provided a measure of whether imprinted genes are expressed above expectation (the expression pattern of a random gene list) in particular identity groups. Venn diagrams of the upregulated imprinted genes making up over-represented identity groups across datasets (within a level) are also reported. Full lists of upregulated imprinted genes can be found in the 'Upregulated_IGs.csv' file for each analysis in the repository (https://osf.io/jx7kr/).

To further examine the presence of imprinted genes within tissues, and to provide a different perspective to over-representation, we conducted a Gene-Set Enrichment Analysis (GSEA) of imprinted genes using a publicly available, light-weight implementation of the GSEA algorithm (Subramanian et al. 2005) in R (https://github.com/JEFworks/liger) in a manner similar to Moffitt et al. (2018). The GSEA was conducted using Log2FC values of statistically significant upregulated genes and is a more conservative measure since it tests whether imprinted genes are enriched in the stronger markers of a group (the genes with the highest fold change for a group vs. the rest). Groups analysed were selected as having an minimum number of upregulated imprinted genes to measure enrichment for (minimum of 15 as suggested by the GSEA user guide (https://www.gsea-msigdb.org/gsea/doc/GSEAUserGuideFrame.html) and having an average fold change of the upregulated imprinted genes greater than the average fold change of the rest of the upregulated genes for that tissue. Again, multiple p values generated from GSEA were corrected using a Bonferroni correction. Graphical representations of significant GSEA's (post-correction) are included in the main text or Appendix B, all other graphs can be found in the repository (https://osf.io/jx7kr/). To further elucidate the genes responsible for significant GSEA's, dot plots of the imprinted genes upregulated in that identity group were plotted across all identity groups with absolute expression and Log2FC mapped to size and colour of the dots respectively. If no cell populations met these criteria, GSEA was not run and were not included in the results.

The absolute number of IGs top-expressed in a tissue/cell-type were also reported for analyses in level 1 and level 2, since these analyses included a variety of cell-types and tissues which may demonstrate meaningful clustering of the highest normalised expression values.

For level 1 and level 2 analyses, we also carried out parent-of-origin specific analyses. Much interest exists concerning the different functions maternally expressed genes (MEGs) and paternally expressed genes (PEGs) plays and we thought it would be valuable to run the large-scale enrichment analyses (level 1 and level 2) using separate lists of PEGs and MEGs. The imprinted gene list was divided into MEGs and PEGs and the analyses detailed above were run separately for these two 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.

All graphical representations and statistical analyses were conducted using R 3.6.2 (Team 2013) in RStudio (Team 2015). 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 and Venn diagrams (Figures 1:4, 6:8, 10:11) were created with BioRender.com.

Results

The hierarchical order of enrichment testing, as demonstrated in Figure 1, began with comparisons of whole brain imprinted gene enrichment vs. other adult and developing tissues to ask whether imprinted genes were statistically enriched in the brain.

1) Whole tissue level analysis

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

The Mouse Cell Atlas (MCA) (Han et al. 2018) and the *Tabula Muris* (TM) (Schaum et al. 2018) are single cell compendiums of ~20, overlapping but not identical, mouse organs (the brain was included in both). These compendiums create a snapshot of gene expression across adult tissues to assess imprinted gene enrichment.

Analysis of the MCA and TM revealed convergence in the 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, bladder and uterus (Table 2).

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 with $q \leq 0.05$ and Log2FC $\geq$ 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 (20,547)	IG (97)	ORA *p*	ORA *q*	Mean FC IG	Mean FC Rest	Top Tissue IGs
Pancreas	2737	42	1.54E-13	**1.85E-12**	8.74	10.32	22
Brain	3401	34	4.37E-06	**5.24E-05**	8.76	125.00	19
Bladder	3184	29	0.000167	**0.002001**	4.45	8.51	8
Uterus	2568	22	0.002818	**0.033816**	4.66	8.46	7
Kidney	1714	10	0.267788	1	13.76	182.89	5
Liver	1739	8	0.559405	1	4.55	80.51	3
Lung	1204	8	0.192845	1	3.82	151.29	4
Ovary	2219	13	0.223000	1	7.46	11.27	5
Small Intestine	1726	5	0.909559	1	7.99	217.84	2
Stomach	1821	7	0.747992	1	4.24	88.60	3
Testis	5212	14	0.995849	1	27.04	5058.36	10
Thymus	1805	6	0.851156	1	2.78	6.76	2
Bone Marrow	1098	2	-	-	5.31	4.46	1
Mammary Gland Virgin	905	4	-	-	3.70	4.05	0
Muscle	1130	4	-	-	8.64	15.34	3
Peripheral Blood	1147	3	-	-	3.78	3.57	0
Prostate	369	0	-	-	0.00	478.10	0
Spleen	1507	1	-	-	4.90	4.83	1

In the *Tabula Muris* analysis (Table 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 $q \leq 0.05$ 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; *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

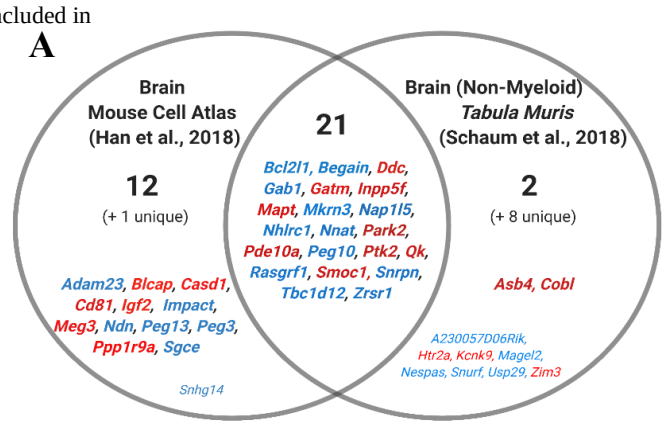
Separating the analysis for MEGs and PEGs (Supplemental Table S2a, S2b, S3a and S3b) revealed a similar pattern of enrichment in both the datasets (Figure 3). Focusing on the brain,

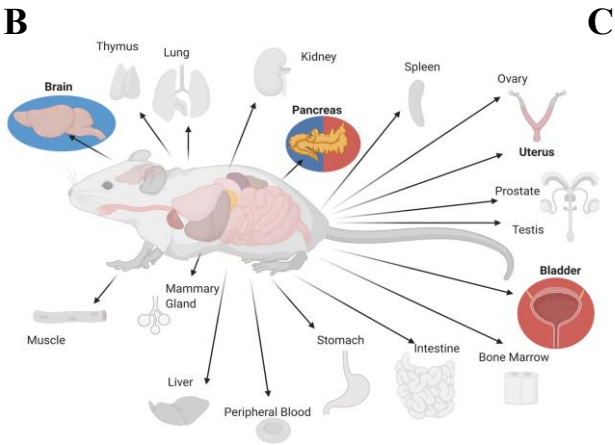
Figure 3. Level 1 cross-tissue comparison summary graphics.

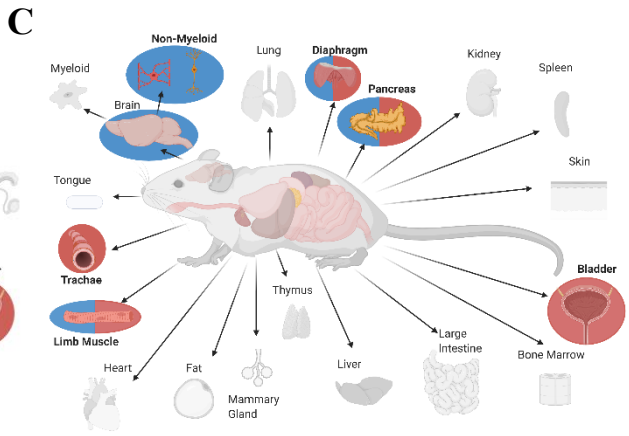
3a) Venn diagram of upregulated imprinted genes in the brain in Mouse Cell Atlas and in the brain (non-myeloid) in the Tabula Muris. Imprinted genes are listed which show significant upregulation ($q \leq 0.05$ and Log2FC ≥ 1) in the tissues. Although these tissues are not identical, these were the two 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.

3b) 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.

3c) 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.







PEGs were over-represented in both datasets (MCA – q = 4.51x10^{-6}, *TM* – q = 0.0002) and MEGs were not (MCA – q = 0.97, *TM* – q = 1). PEGs were also over-represented in the diaphragm (q = 0.0006), limb muscle (q = 9.63x10^{-5}) and pancreas (MCA – q =1.91x10^{-5}, *TM* – q = 0.0004), with a significant GSEA in the MCA pancreas (p = 0.02, Appendix B1). While MEGs were over-represented in the bladder (MCA – q = 0.002, *TM* – q = 0.021), the pancreas (MCA - q = 1.51x10^{-7}, *TM* – q = 0.03) and in the three muscular tissues of the *Tabula Muris* (diaphragm – q = 1.44x10^{-8}, limb muscle – q = 1.66x10^{-7}, trachea – q = 0.003).

In summary, across adult tissues, imprinted genes were convergently over-represented in the pancreas, bladder and the brain. The convergence of imprinted genes upregulated in the brain compared to other adult organs within the two compendiums is represented in Figure 3.

2. Whole Brain level analysis

Having found an over-representation in the brain at level 1, we next sought to test whether imprinted genes would demonstrate a cell-type or regional enrichment within the whole brain.

<u>2i) Imprinted genes are enriched in neurons and neuroendocrine cells</u>

We first analysed Ximerakis et al.'s (2019) dataset, in which cells were grouped from the whole mouse brain (minus the hindbrain) into major cell classes according to cell lineage. Imprinted genes were over-represented in neuroendocrine cells and mature neurons (Table 4).

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 $q \leq 0.05$ 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; *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 Cell IGs* – Sum of IGs with highest mean expression value for that identity group.

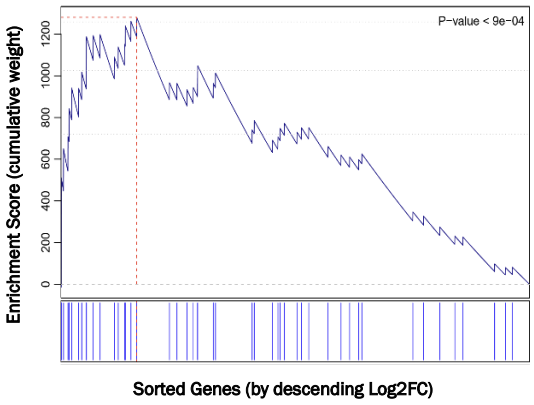
Identity (Abbr.)	Up Reg (14,661)	IG (88)	ORA *p*	ORA *q*	Mean FC IG	Mean FC Rest	GSEA *p*	GSEA *q*	Top Cell IGs
Neuroendocrine cells (NendC)	4234	48	3.84E-07	**6.92E-06**	16.34	5.49	9.00E-04	**0.0036**	32
Mature Neurons (all types) (mNEUR)	2989	33	0.000156	**0.002800**	7.62	8.76	-	-	1
Vascular and leptomeningeal cells (VLMC)	1735	19	0.006444	0.115995	85.82	12.22	0.0946	0.3784	4
Tanycytes (TNC)	1698	18	0.011174	0.201139	6.55	8.67	-	-	7
Arachnoid barrier cells (ABC)	2275	20	0.047213	0.849842	15.31	23.21	-	-	7
Oligodendrocyte precursor cells (OPC)	1549	12	0.216078	1	3.06	7.43	-	-	2
Pericytes (PC)	1764	13	0.256801	1	7.95	7.59	-	-	3

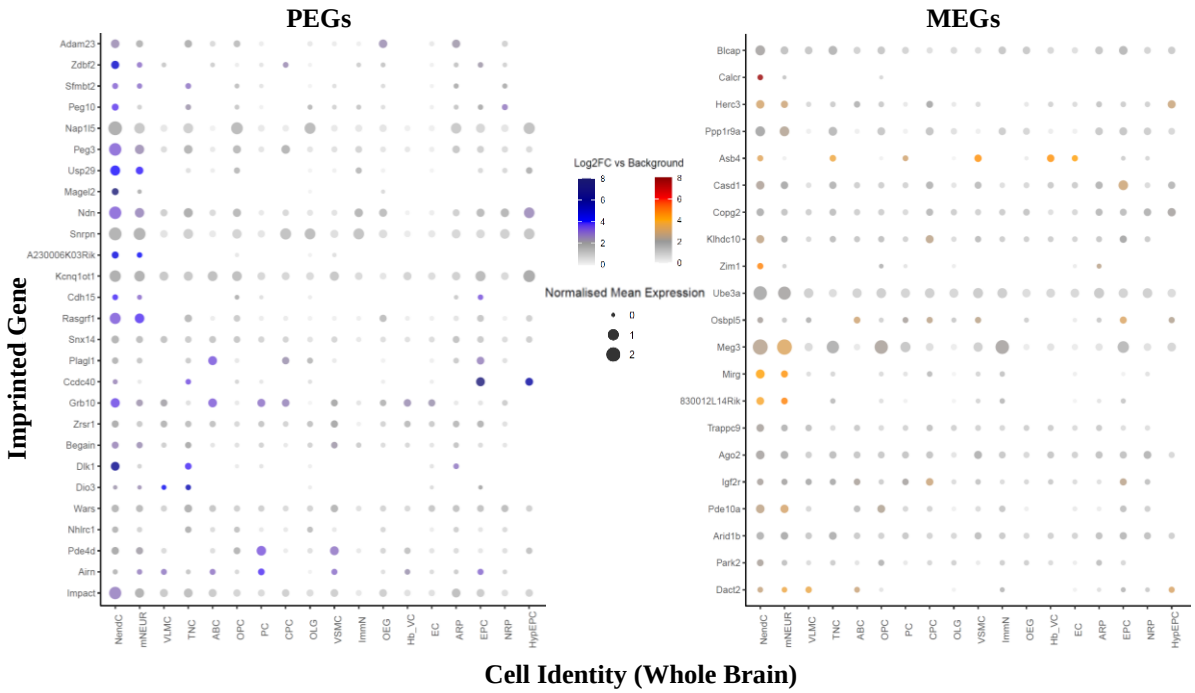
Choroid plexus epithelial cells (CPC)	2873	20	0.266209	1	79.24	23.94	0.1717	0.6868	9
Oligodendrocytes (OLG)	1124	8	0.361824	1	3.95	10.45	-	-	5
Vascular smooth muscle cells (VSMC)	3133	20	0.418819	1	7.95	6.85	0.1622	0.6488	2
Immature Neurons (immN)	713	5	0.427092	1	3.34	5.90	-	-	0
Olfactory ensheathing glia (OEG)	1195	8	0.428082	1	8.42	23.85	-	-	2
Hemoglobin-expressing vascular cells (Hb_VC)	1930	11	0.620935	1	5.07	6.01	-	-	3
Endothelial cells (EC)	1496	8	0.686994	1	5.19	8.25	-	-	0
Astrocyte-restricted precursors (ARP)	1853	10	0.688635	1	3.88	4.58	-	-	1
Ependymocytes (EPC)	3429	17	0.849270	1	15.45	54.05	-	-	3
Neuronal-restricted precursor (NRP)	2458	9	0.969946	1	2.75	13.07	-	-	0
Hypendymal cells (HypEPC)	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

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* and *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 (i.e. amongst the genes with the highest fold change values) 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. The MEG/PEG analysis (Supplemental Table S4a and S4b) for this dataset found that PEGs were novelly over-represented in tanycytes (q = 0.009) while also demonstrating over-representation in mature neurons (q = 0.01) and neuroendocrine cells (q = 1.86×10^{-5}) and a significant GSEA in the latter also ($p = 9 \times 10^{-4}$, Appendix B2). In contrast, MEGs were novelly over-represented in Arachnoid barrier cells (q = 0.019) and Vascular and Leptomeningeal cells (q = 0.037) with no significant ORA or GSEA in the neurons or neuroendocrine cells.

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

Figure 4. Imprinted gene GSEA for genes upregulated in neuroendocrine cells. Right GSEA graph - 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. **Bottom Dot Plot** – Upregulated imprinted genes in neuroendocrine cells, separated by MEGs and PEGs, were plotted across all cell types with 5 or more imprinted genes upregulated in Ximerakis et al. 2019 (Abbr. in Table 4). Size of points represented absolute mean expression; colour represented the size of the Log2FC value for neuroendocrine cells vs. all other cells.





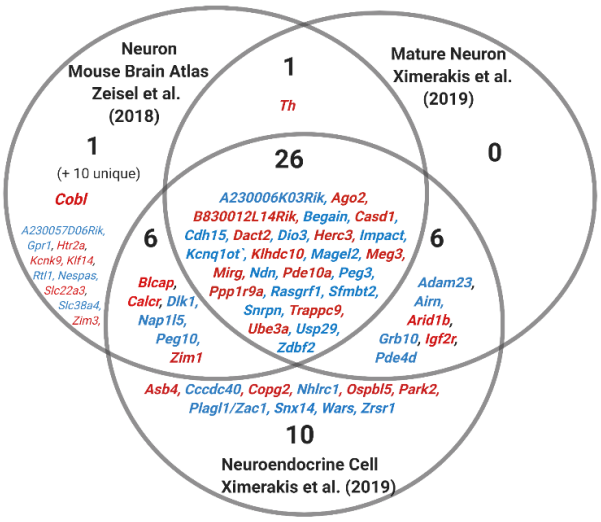
entire murine nervous system and identifying cells by both brain region and cell type. Concordant with the previous findings, primary analysis separating cells by lineage revealed over-representation of imprinted genes in neurons only (Table 5). The overlap between the upregulated imprinted genes for the over-represented neural-lineage cells from both the previous datasets are displayed in Figure 5. Additionally, PEGs alone demonstrated no significant over-representations in global cell types while MEGs demonstrated over-representation in vascular cells (q = 0.001) (Supplemental Table S5a and S5b).

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 $q \leq 0.05$ 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 Cell IGs* – Sum of IGs with highest mean expression value for that identity group.

Identity	Up Reg (19,550)	IG (107)	ORA *p*	ORA *q*	Mean FC IG	Mean FC Rest	Top Cell 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

Figure 5. 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.



2ii) 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). The imprinted gene overlap for the over-represented brain regions is shown in the Venn Diagram in Figure 6a.

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 $q \leq 0.05$ 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; *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 Region 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 Region 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
Vent. Midbrain	1229	18	0.000255	**0.003818**	4.90	5.55	-	-	3
Vent. 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
Dors. 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
Dors. 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 (Supplemental Table S6a and S6b), revealed that PEGs were over-represented in hypothalamus ($q = 1.21\text{x}10^{-7}$), 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 ($q = 0.028$, Appendix B3); see Figure 6b.

We then recategorized the neurons into individual neuronal subtypes to uncover the specific neural population underlying the enrichment seen in the hypothalamus, pons and medulla, and midbrain (Figure 7; Supplemental Table S7). The hypothalamus was represented by a selection of inhibitory and peptedergic neurons. Inhibitory neurons with over-representation of imprinted genes included: Deinh4 (*Syt2, Smim17*) representing a Subthalamic Nucleus population, Deinh5 (*Nts*, *Gal 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

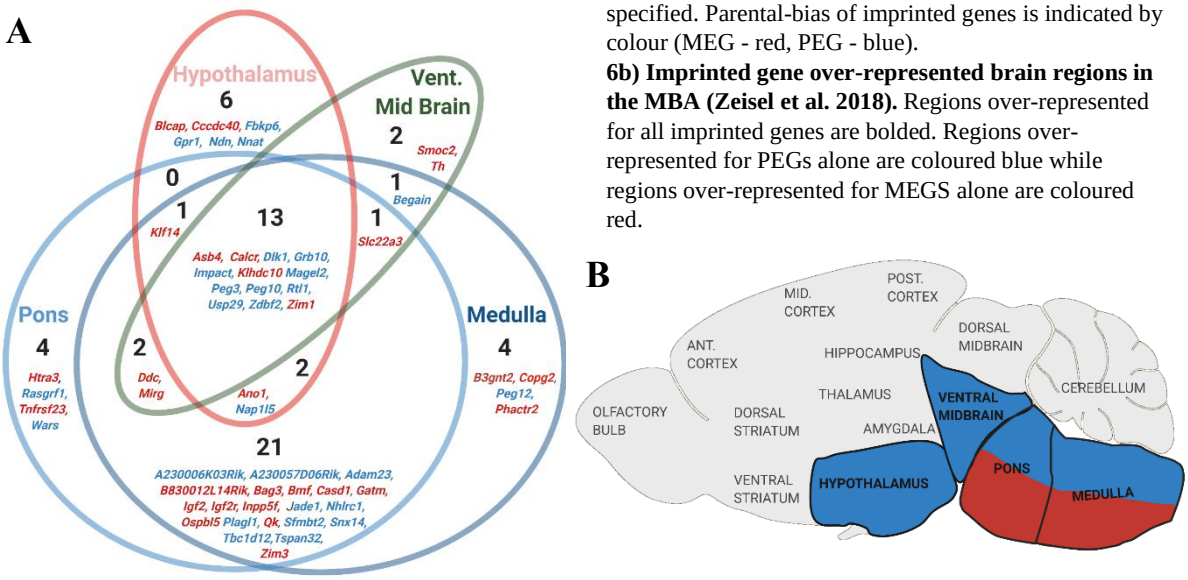
neurons, over-representation was seen in HYPEP3 (*Gpr101*, *Tac1*) a ventromedial population, HYPEP2 (*Otp*, *Trh*, *Ucn3*) a ventromedial/paraventricular population, HYPEPE1 (*Trh*, *Otp*, *Ngb*) a lateral hypothalamic population, HYPEP4 (*Oxt*, *Otp*) an oxytocin magnocellular population of the paraventricular and supraoptic nuclei, 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/P AG (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

Figure 6. Level 2ii) brain region analysis summary figures

6a) Venn diagram of upregulated imprinted genes in the neurons of over-represented 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).

6b) Imprinted gene over-represented brain regions in the MBA (Zeisel et al. 2018). Regions over-represented for all imprinted genes are bolded. Regions over-represented for PEGs alone are coloured blue while regions over-represented for MEGS alone are coloured red.



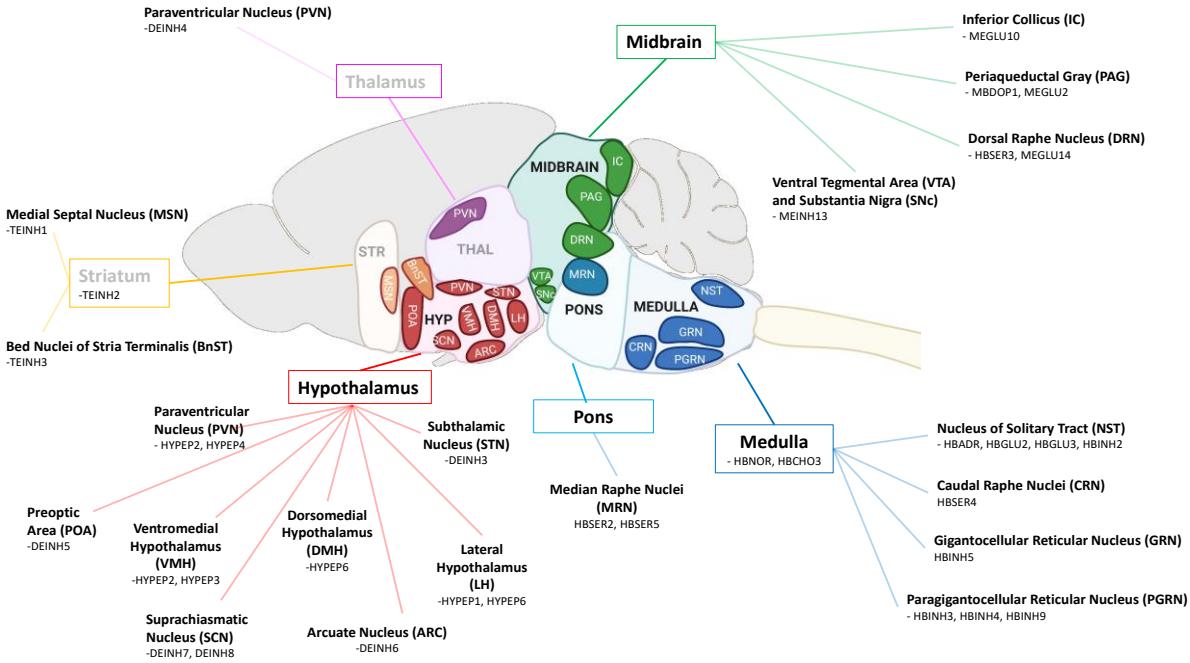
(HBSER) are particularly prominent since the pons and medulla base serotonin neuron populations (HBSER2, HBSER4 and HBSER5) were the only neuron subpopulations out of the 214 total to have a significant GSEA for imprinted genes after correction (Appendix B4). Additional regions of over-representation included TEINH1-3 representing neurons in the pallidum and striatum and DEINH4 representing PVN neurons from the thalamus. In this

comparison of 214 neuron 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 cross-brain level analysis.

3) Brain Region level Analysis

We next sought to identify enriched neuronal subpopulations within the brain regions

Figure 7. 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 32 neuronal 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 Supplementary Table 6.



identified above. Namely within the hypothalamus (to identify populations enriched across the whole hypothalamus and those enriched within specific hypothalamic nuclei) and whether imprinted genes were enriched in the dopaminergic and serotonergic populations spread across the ventral midbrain and hindbrain and key subpopulations in the whole brain analysis

<u>3a) Whole Hypothalamus</u>

<u>3ai) Imprinted genes are over-represented in hypothalamic neurons</u>

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 (Supplemental Table S8a and S9a). 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 (Figure 8a-d, Romanov et al. 2017 – p = 0.012, Chen et al. 2017 - p = 0.019).

3aii) 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; Supplemental table S8b). 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 concentrated in the POA); GABA17 (*Slc6a3*) a dopaminergic neuron type with high expression of *Th* and *Prlr*, which most likely corresponds to the 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; Supplemental table S9b): GABA5 (*Calcr, Lhx1*) whose largest upregulated gene is *Slc6a3* and which likely

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 number of upregulated genes with $q \leq 0.05$ and positive Log2FC (total number of genes in the dataset in brackets); *IG* – number of imprinted genes upregulated with $q \leq 0.05$ and positive Log2FC (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

Identity	Up Reg	IG	ORA *p*	ORA *q*	Mean FC IG	Mean FC Rest
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
GABA8 (*Vipr2*)	478	11	0.000247	**0.005935**	1.96	2.72
GABA13 (*Slc18a2, Gal*)	568	12	0.000279	**0.006687**	1.89	2.01
GABA15 (*Agrp*)	764	13	0.001190	**0.028553**	2.40	2.78
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
GABA 14 (*Npy, Agrp*)	193	6	0.0018726	**0.013108**	2.94	6.12
Ghrh	235	6	0.004954	**0.034677**	18.31	17.19

corresponds 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 A for full comparison); GABA14 (*Agrp*/*Npy*) which are the Arcuate Nucleus feeding neurons also reported in Chen et al. (2017); and a *Ghrh* neuronal type, also strongly upregulated for *Th*, which was not over-represented in Chen et al. (2017).

3b) Specific Hypothalamic Nuclei

3bi) Whole Hypothalamic imprinted gene enrichment mirrored in specific hypothalamic nuclei

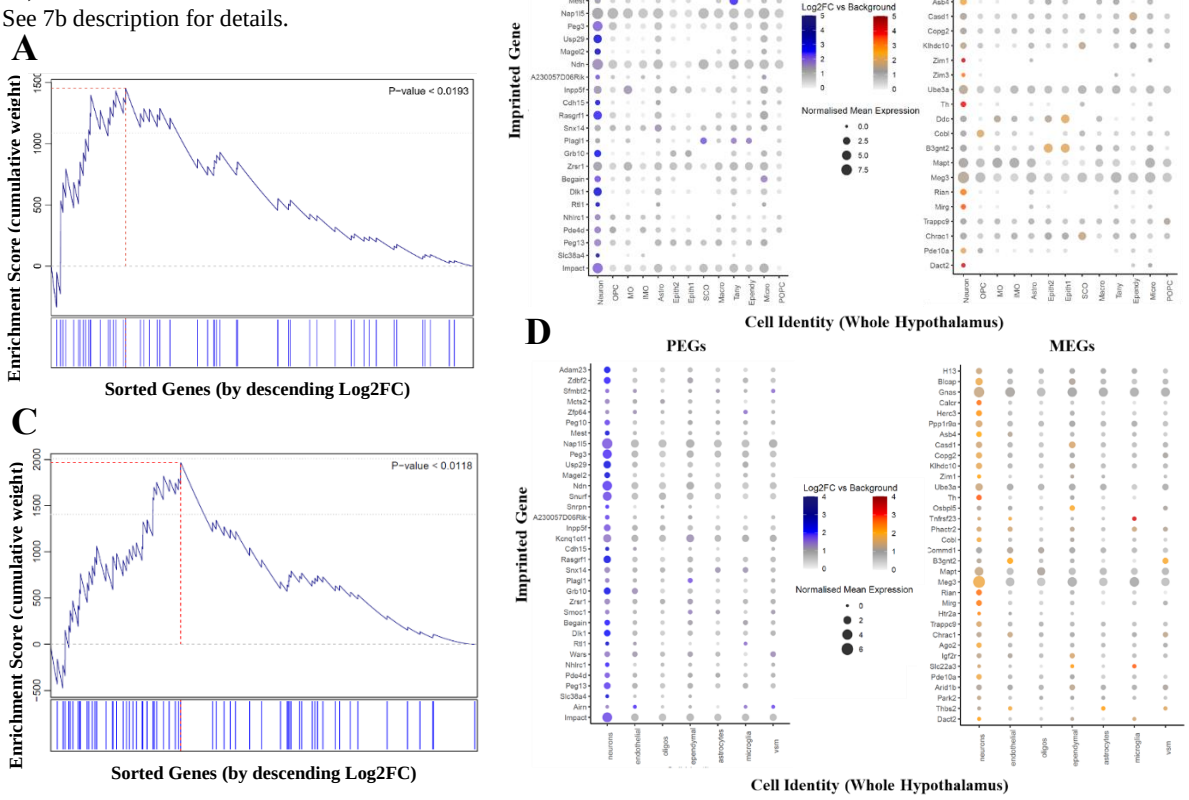
Figure 8. Imprinted gene GSEA for genes upregulated in neurons across the whole hypothalamus.

8a) Chen et al. 2017 - GSEA graph - 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. 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.

8b) Chen et al. 2017 - Dot Plot – Upregulated imprinted genes in neurons, separated by MEGs and PEGs, were plotted across all cell types in Ximerakis et al. 2019. Size of points represented absolute mean expression; colour represented the size of the Log2FC value for neurons vs. all other cells across the hypothalamus.

8c) Romanov et al. 2017 - GSEA Graph – See 7a description for details.

8d) Romanov et al. 2017 - Dot Plot – See 7b description for details.



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; Supplemental Table S10). These included the major *Agrp*/*Npy* neuron type (*Agrp*/*Sst*) and both major *Pomc* neuron types known to be feeding suppressants (Rau and 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 group overlapped with the TIDA candidates from the previous two datasets (e.g. *Slc6a3*, *Th*, *Lhx1*, *Calcr*). *Agrp* neurons and these TIDA candidate neurons were identified in both whole hypothalamic datasets and at the nuclei level.

Preoptic Area - POA (Moffitt et al. 2018)

Imprinted gene over-representation was found in 14/66 of the neuron clusters identified by Moffitt et al. (2018) in the preoptic area (Supplemental Table S11a). Moffitt et al. (2018) stated that 9 of the 66 clusters most likely originate from outside the POA including 3/14 of the imprinted gene enriched neuron clusters (one of which was the *Avp*/*Nms* neurons (q = 0.0003) likely to be from the SCN). To focus our analysis on the POA alone, we removed the non-POA neuron groups and ran the analysis on the remaining neural groups restricted to POA (Table 8b; Supplemental Table S11b). Imprinted genes were found to be over-represented in 15 neuron types including two Galanin enriched clusters (i22, i8), the major *Th* enriched cluster (i17), e15:*Ucn3*/*Brs3* (a cluster activated by parenting), a *Crh* cluster (i35) and a *Trh* cluster (e4), and both major glutamatergic *Ghrh* clusters (e13, e19). Since a considerable number of neuron types were over-represented, the above analysis was run again with the Log2FC $\geq$ 1 criteria to find out if imprinted genes were enriched in the more distinct markers (Supplemental

Table S11c). Only two neuron types were over-represented: i35:*Crh*/*Tac2* (q = 0.0126) and i16:*Gal*/*Th* (q = 0.0026). i16:*Gal*/*Th* is a prominent behavioural neuron group (parenting and mating behaviour) which was not identified in the previous analysis. Only three of Moffitt et al.'s (2018) neuron groups were upregulated in *Slc18a2, Gal, Prlr, Th* and *Esr1* (i8, i16, i38) like GABA13 in Chen et al. (2017) and all of them are represented above.

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

Over-representation of imprinted genes in the Lateral Hypothalamic Area neurons was analysed for Glutamatergic and GABAergic neurons separately since they were sequenced separately by Mickelsen et al. (2019). For Glutamatergic neurons, imprinted gene over-representation was found in 6/15 clusters (Table 8ci, Supplemental Table S12a). These included: Glut 6 (*Hcrt*) identified as the Orexin cluster, a classic motivated behaviour neuronal subset identified in the MBA as HYPEP6; and Glut 9 (*Synpr*/*Gad1*) with a significant GSEA (Appendix B5) 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; Supplemental Table S12b).

Suprachiasmatic Nucleus – SCN (Wen et al. 2020)

Analysis of the 10x chromium data of only SCN neurons (Table 8d, Supplemental Table S13) revealed a significant over-representation and GSEA (Appendix B6) in the *Avp*/*Nms* neuronal cluster only. This cluster shows the strongest expression for *Oxt*, *Avp*, *Avpr1a* and *Prlr* and is

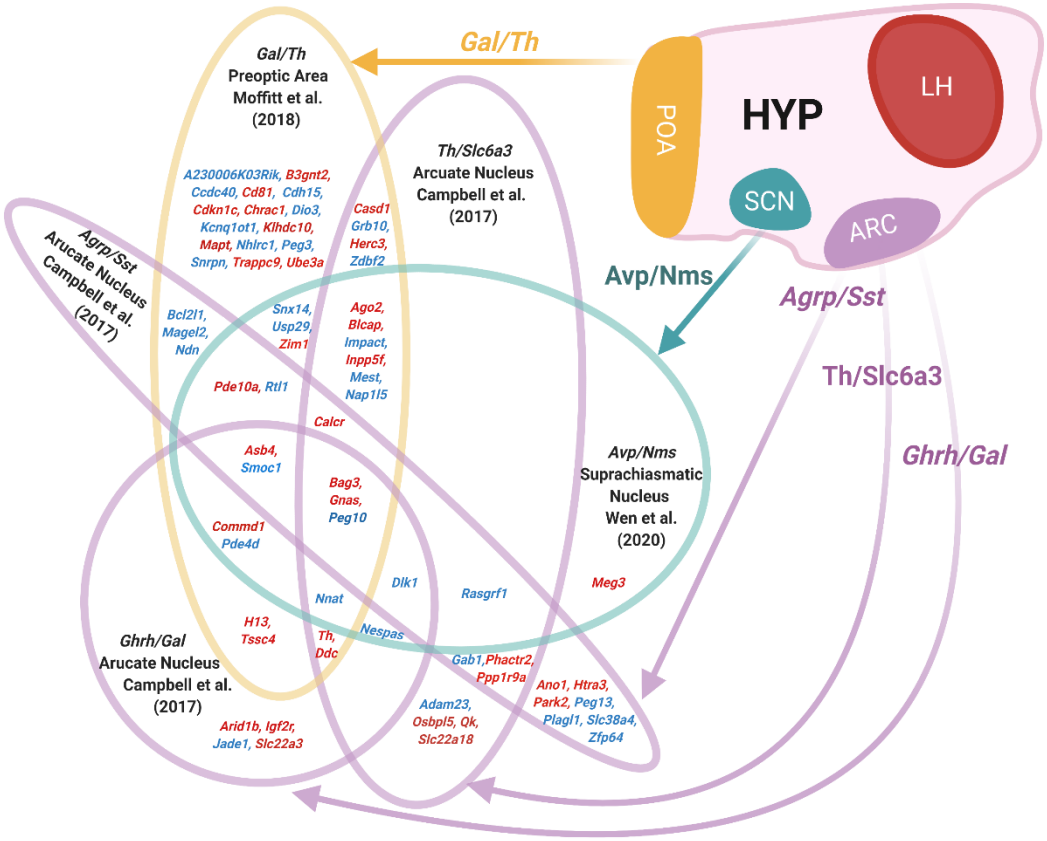
one of the three neural group that Wen et al. (2020) found had robust circadian gene expression, and the only subtype with notable phase differences in circadian gene expression in the dorsal SCN). This cluster likely corresponds to the GABA8 cluster found enriched in the Chen et al. (2017) dataset and the *Avp*/*Nms* group identified by Moffitt et al. (2018) as an off-target neuronal group.

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 with $q \leq 0.05$ and positive Log2FC (total number of genes in the dataset in brackets); *IG* – number of imprinted genes upregulated with $q \leq 0.05$ and positive Log2FC (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

Identity	Up Reg	IG	ORA *p*	ORA *q*	Mean FC IG	Mean FC Rest	GSEA *p*	GSEA *q*
a) ARC	**(16,098)**	**(94)**	***(24 neuron types)***					
Th/Slc6a3	1113	26	7.49E-10	**1.72E-08**	2.14	4.04	-	-
Arx/Nr5a2	599	16	3.98E-07	**9.16E-06**	1.33	3.15	-	-
Agrp/Sst	1704	23	0.000111	**0.002552**	1.61	1.78	-	-
Pomc/Anxa2	3162	34	0.000168	**0.003866**	1.70	1.82	-	-
Ghrh	1265	18	0.000382	**0.008793**	2.02	2.82	-	-
Gpr50	578	11	0.000591	**0.013595**	2.52	5.02	-	-
Pomc/Glipr1	1536	19	0.001426	**0.032792**	1.91	1.90	0.2851	0.8553
Slc17a6/Trhr	1547	19	0.001550	**0.035651**	3.22	2.35	0.1143	0.3429
b) POA	**(16,402)**	**(100)**	***(57 neuron types)***					
i22:*Gal/Pmaip1*	1569	27	4.64E-07	**2.41E-05**	2.02	2.14	-	-
i8:*Gal/Amigo2*	3625	42	6.29E-06	**0.000327**	1.78	1.53	0.1628	1
e15:*Ucn3/Brs3*	996	18	2.90E-05	**0.001510**	6.15	3.43	0.04	0.32
i12:Gaba	1229	20	4.40E-05	**0.002286**	1.76	1.68	0.2236	1
i35:*Crh/Tac2*	1464	22	5.69E-05	**0.002961**	1.98	2.44	-	-
i17:*Th/Nos1*	1358	21	5.76E-05	**0.002998**	2.02	2.00	0.4519	1
e4:*Trh/Angpt1*	2770	33	6.13E-05	**0.003188**	1.90	2.28	-	-
e19:*Ghrh/Trh*	1974	26	9.75E-05	**0.005068**	2.14	3.88	-	-
e1:Glut	1518	22	9.76E-05	**0.005073**	1.58	1.87	-	-
e12:*Nos1/Foxp2*	1635	22	0.000284	**0.014790**	1.81	2.64	-	-
i1:Gaba	602	12	0.000295	**0.015355**	1.20	1.55	-	-
i24:*Nmu*	3020	33	0.000334	**0.017349**	1.71	2.02	-	-
e13:*Ghrh/C1ql1*	3588	37	0.000398	**0.020691**	2.58	2.81	-	-
i37:*Bdnf/Chrm2*	2078	25	0.000561	**0.029168**	2.80	3.17	-	-
i38:*Kiss1/Th*	2078	25	0.000561	**0.029168**	2.09	5.11	-	-
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	-	-
Glut_7 (*Gpr101, Tcf4*)	253	10	7.00E-06	**9.80E-05**	1.55	2.59	-	-
Glut_14 (*Otp/Gpr101*)	369	11	3.36E-05	**0.000470**	1.38	2.08	-	-
Glut_9 (*Synpr/Gad1*)	784	15	0.000196	**0.002743**	2.17	2.02	0.0099	**0.0198**
Glut_5 (*Ebf3/Otp*)	720	14	0.000280	**0.003915**	1.74	1.64	-	-
Glut_6 (*Hcrt*)	3097	35	0.000423	**0.005924**	3.15	2.29	0.1385	0.277
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
Gaba_5 (*Meis2, Calb2*)	646	12	0.001076	**0.015062**	1.26	1.82	-	-
d) SCN	**(11,391)**	**(77)**	***(5 neuron types)***					
Avp/Nms	909	24	3.01E-09	**1.51E-08**	3.12	1.94	**0.004**	-

Figure 9 presents the overlapping upregulated imprinted genes from the convergently upregulated neuron subtypes in the hypothalamic analysis of level 3a and 3b.

Figure 9. 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).



3c) Monoaminergic Nuclei

3ci) Imprinted genes are over-represented in *Th* +ve mature 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 with $q \leq 0.05$ and positive Log2FC (total number of genes in the dataset in brackets); *IG* – number of imprinted genes upregulated with $q \leq 0.05$ and positive Log2FC (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

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

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 with $q \leq 0.05$ and positive Log2FC (total number of genes in the dataset in brackets); *IG* – number of imprinted genes upregulated with $q \leq 0.05$ and positive Log2FC (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.

Identity	Up Reg (14,027)	IG (93)	ORA *p*	ORA *q*	Mean FC IG	Mean FC Rest	GSEA *p*	GSEA *q*
Serotonergic Neurons	1781	33	1.42E-08	**2.99E-07**	4.78	2.36	0.0272	0.0544
Pericytes	1409	22	0.000103	**0.002161**	8.57	8.04	0.2209	0.4418
Dopaminergic 1 (*Th*)	347	9	0.000484	**0.010164**	3.17	2.84	-	-
Dopaminergic 0 (Immature)	219	7	0.000624	**0.013114**	4.03	2.91	-	-
Lateral Neuroblasts 2	2158	26	0.001353	**0.028403**	2.11	2.48	-	-
Oculomotor and trochlear nucleus	1994	22	0.009922	0.208356	1.83	2.28	-	-
Mediolateral Neuroblasts 3	707	10	0.018500	0.388509	2.25	2.27	-	-
GABAergic neurons 1b	943	12	0.021628	0.454193	1.84	2.03	-	-
GABAergic neurons 1a	743	10	0.025112	0.527346	2.05	2.12	-	-
Dopaminergic 2 (*Aldh1a1*)	464	7	0.034351	0.721369	3.22	3.43	-	-
Endothelial cell	2555	24	0.042792	0.898639	7.38	8.31	-	-
Mediolateral Neuroblasts 1	302	5	0.050300	1	1.76	2.24	-	-
GABAergic neurons 2	5711	46	0.053803	1	1.89	1.95	-	-

Radial Glia-like Cells 3	1142	9	0.344985	1	3.43	5.45	-	-
Red Nucleus	1021	8	0.365408	1	2.21	2.01	-	-
Mediolateral Neuroblasts 4	811	6	0.452000	1	1.54	2.42	-	-
Microglia	1293	7	0.765321	1	6.06	25.18	-	-
Radial Glia-like Cells 1	1309	7	0.776475	1	3.08	2.79	-	-
Ependymal	1391	7	0.827678	1	3.49	13.28	-	-
Neuronal Progenitor	1775	9	0.847792	1	2.32	2.76	-	-
Radial Glia-like Cells 2	1726	8	0.899405	1	2.89	3.41	-	-
Lateral Neuroblasts 1	529	4	-	-	1.76	2.93	-	-
Mediolateral Neuroblasts 5	125	4	-	-	1.86	1.92	-	-
Neuroblast Dopaminergic	196	3	-	-	3.55	2.14	-	-
Medial Neuroblast	313	2	-	-	1.48	2.34	-	-
Mediolateral Neuroblasts 2	206	1	-	-	1.13	1.82	-	-

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 with $q \leq 0.05$ and positive Log2FC (total number of genes in the dataset in brackets); *IG* – number of imprinted genes upregulated with $q \leq 0.05$ and positive Log2FC (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.

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

3cii) Imprinted genes are enriched in serotonergic neurons, specifically those projecting sub-cortically

To examine enrichment of imprinted genes in serotonergic neurons further, analyses were performed on data for the Raphe nuclei of the midbrain/hindbrain which are the key serotonergic regions of the brain. Analysis of all cell types in 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 (Supplemental Table S14a). 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 (Appendix B7). When compared against each other (i.e. serotonergic upregulation vs. the other neurons), only the serotonergic neurons of the DRN were found to have a significant over- representation of imprinted genes (Table 12). GSEA's 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 with $q \leq 0.05$ and positive Log2FC (total number of genes in the dataset in brackets); *IG* – number of imprinted genes upregulated with $q \leq 0.05$ and positive Log2FC (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.

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

The analysis was further broken down into neural subpopulations of all neuron types (Supplemental Table S14b). 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).

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 with $q \leq 0.05$ and positive Log2FC (total number of genes in the dataset in brackets); *IG* – number of imprinted genes upregulated with $q \leq 0.05$ and positive Log2FC (total number of IGs in the dataset in brackets); *ORA p* – *p* value from

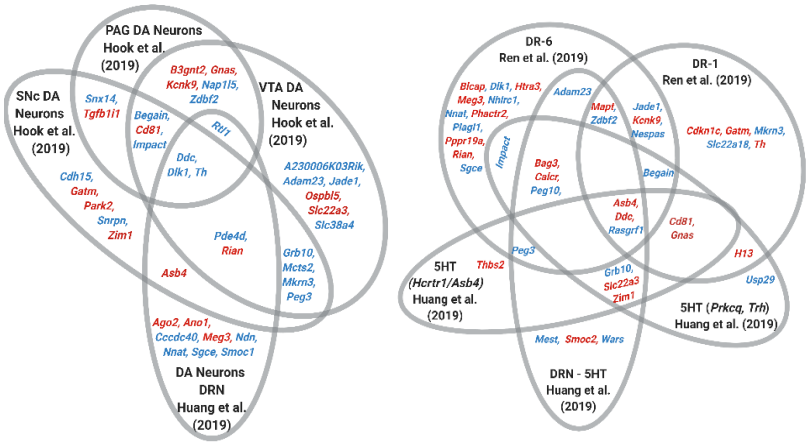
over representation analysis on groups with minimum 5% of total IGs; *ORA q* – Bonferroni corrected *p* value from ORA; *Mean FC IG* – mean fold change for upregulated imprinted genes; *Mean FC Rest* – mean fold change for all other upregulated genes; *GSEA p* – *p* value from Gene Set Enrichment Analysis for identity groups with 15+ IGs and Mean FC IG > Mean FC Rest; *GSEA q* – Bonferroni corrected *p* values from GSEA.

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

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 above analysis was run again with the Log2FC ≥ 1 criteria to find out if imprinted genes were enriched in the more distinct markers (Supplemental Table S15). 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).

Figure 10 presents the overlapping upregulated imprinted genes from the convergently upregulated neuron subtypes in the monoaminergic analysis of level 3c.

Figure 10. Venn Diagrams of Upregulated Imprinted Genes in the adult dopaminergic and serotonergic populations. Left – Global Dopaminergic Neurons from Hook et al. 2019 and Huang et al. 2019. Right – 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 the subpopulation. Parental-bias is indicated by colour (MEG - red, PEG - blue).



Discussion

Here we apply an unbiased 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 confirm 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 suggest 'functional hotspots'. Specifically, our findings at the systems and cellular level highlight a major role for imprinted genes in the neuronal regulation of feeding, parental behaviour and sleep.

Some of the earliest studies of genomic imprinting identified the brain as a key area for imprinted gene expression (Allen et al. 1995; Keverne et al. 1996). However, it is estimated that ~80% of the genome is expressed in the brain and consequently, imprinted gene expression in the brain might not be a purposeful phenomenon. Our current analysis revealed that imprinted genes were statistically significantly over-represented in the brain as a whole. This over-representation was mostly driven by paternally expressed genes (PEGs). Within specific brain regions, imprinted genes were over-represented in the hypothalamus, ventral midbrain, pons and medulla. This confirms some previous findings from studies of Pg/Gg and Ag chimera studies (Allen et al. 1995; Keverne et al. 1996) and summaries of imprinted gene expression (Gregg et al. 2010). However, unlike these earlier studies, our analyses do not simply ask if imprinted genes are expressed (at any level) or not, but robustly test whether this expression is meaningful, and these genes are especially enriched in any given brain region. Additionally, in the chimera studies, Pg/Gg cells with two maternal genomes preferentially allocated to the

developing adult cortex and hippocampus, and Ag cells with two paternal genomes preferentially allocated to the developing hypothalamus and midbrain. Our analysis does not reproduce this distinct pattern of maternally expressed gene (MEG) and PEG expression in the brain, and indeed we find no specific enrichment of imprinted genes in cortex or hippocampus. Although the pattern of regional enrichment seen with all imprinted genes is replicated when analysing PEGs alone, separate analysis of MEGs only shows over-representation in the pons and medulla. This difference between our analysis of enrichment and the Pg/Gg and Ag chimeras studies indicate that the distribution of Pg/Gg and Ag cells in the brain is not driven by, or indeed reflected by adult PEG and MEG expression, but instead is probably determined by developmental expression of key imprinted genes.

By taking advantage of scRNA-seq data we were able to examine imprinted gene enrichment in specific cell-types. At the whole brain level, mature neurons and, in particular, neural-lineage neuroendocrine cells had disproportionately higher numbers of imprinted genes expressed and high levels of imprinted gene expression. It is likely that this neural-lineage neuroendocrine population comprises members of the key hypothalamic populations in which imprinted genes are enriched and, when treated as their own cluster, demonstrate strong imprinted gene enrichment compared to other cell lineages of the brain, even other mature neurons.

Within the hypothalamus, a selection of informative neuronal subpopulations were over-represented. Strikingly, and suggestive of meaningful enrichment, we saw convergence across our different levels of analysis with several key neuronal types identified in the whole hypothalamus and/or hypothalamic-region-level analysis, already having been identified against the background of general imprinted gene expression in the whole-brain-level analysis. These subpopulations are collectively associated with a few fundamental motivated behaviours. Feeding behaviour is directly controlled by two Arcuate Nucleus populations: the *Agrp* neurons, as feeding promotors, (which were enriched for imprinted genes at the whole

brain level, in both whole hypothalamic level datasets and within the Arcuate nucleus) and the *Pomc* neurons as feedings suppressants (enriched with the ARC dataset only) (Aponte et al. 2011). These neurons have already been associated with a few imprinted genes previously (*Asb4, Magel2, Snord116*) (Cassidy and Charalambous 2018) but never as an enriched population. Circadian process are controlled principally by the Suprachiasmatic Nucleus and here we find strong imprinted gene enrichment in the circadian regulatory population – Avp/Nms neurons (enriched at the whole brain level, the whole hypothalamic level and as the only enriched neuronal type within the SCN). This population is of interest given the growing appreciation of the role imprinted genes play in circadian processes and the SCN suggested by studies of individual imprinted genes (Tucci 2016). Pituitary endocrine regulation also emerged as an enriched function, considering the over-representation in the dopaminergic: *Th/Slc6a3/Prlr* neuron type (top hit in the ARC and across dopaminergic neurons of the brain) and the *Ghrh/Gal* subpopulation. These neurons both exist in the arcuate nucleus and regulate prolactin and growth hormone release, respectively. Finally, neuronal control of parenting behaviour was also suggested from the enrichment we saw in the *Gal/Esr1* neuronal population of the Preoptic Area (enriched at whole brain, whole hypothalamic and at POA level). This neural population is known to be associated with feeding and, principally, parenting behaviour (Moffitt et al. 2018). In addition prolactin (regulated from the ARC) is a well-known as a mediator of maternal-infant interaction priming the maternal brain during pregnancy for the nurturing of pups (Grattan et al. 2008).

In this analysis the hypothalamus was a clear hot spot for imprinted gene expression, in line with the prevailing view of imprinted gene and hypothalamic function (Ivanova and Kelsey 2011; Pulix and Plagge 2020). However, outside of the hypothalamus other distinct hotspot emerged from our whole brain analysis including the monoaminergic system of the midbrain/hindbrain. Analysing the dorsal raphe nucleus and the ventral midbrain revealed the

serotonergic neurons to be a foci of imprinted gene expression within this region; more specifically the serotonergic neurons that project to the subcortical regions of the brain known to be associated with feeding and other motivated behaviours (Donovan and Tecott 2013), mirroring the functional hotspots seen in the hypothalamus. In contrast to the specificity of the imprinted gene enrichment in serotonergic neurons, the enrichment seen for adult midbrain dopamine neurons was more generalised. Midbrain dopamine neurons were enriched in adult whole brain and developmental ventral midbrain analyses, and sub-type specific enrichment centred on the DAT^{high} lineage of midbrain DA neurons. These DAT^{high} neurons amount to the standard dopamine neurons (*Th/Slc6a3* high) and are found throughout the VTA, SNc and PAG (Tiklová et al. 2019) leading to little subdivision of the dopaminergic signal.

Analyses of these kind are always bound by the available data and therefore there are notable limitations and caveats to this study. The aim of this study was to generate information about 'hotspots' of imprinted gene expression. This approach, and the use over-representation analysis and GSEA, therefore do not provide an exhaustive list of sites of expression. For example, oxytocin neurons of the hypothalamus were not significantly enriched in any of the hypothalamic datasets despite a clear oxytocin neuron phenotype reported in a handful of imprinted gene models (Li et al. 1999; Dombret et al. 2012). This may be an example of a functional effect occurring below the level of over-representation, or that imprinted genes act during development and are not functionally enriched in adult oxytocin neurons, or simply that compared to other hypothalamic neuronal populations, oxytocin neurons are not a 'hotspot' of imprinted expression. Specific sequencing of oxytocinergic brain regions will be required to distinguish between these possibilities. A second caveat is that we did not assess parent-of-origin expression for the 119 imprinted genes we included in the analysis. Previous expression profiling of imprinted genes have also not measured the POEs (Steinhoff et al. 2009; Gregg et al. 2010) but have restricted their gene selection to genes with reliable imprinting status.

Consequently, we only included the canonical imprinted genes and genes with more than one demonstration of a POE when looking for enrichment. Furthermore, for the vast majority of these genes, a brain-based POE effect has also already been reported (Supplemental Table S1). Although this does not replace validating the imprinting status of all 119 in the tissues and subregions examined, it does provide justification for looking at imprinted gene over-representation. To resolve this issue, scRNA-seq using hybrid tissues will be key; for example, the recent work of Laukoter et al. (2020) with cortical cell types provides an example of the allelic specific single-cell expression measurements necessary to confirm the enrichments found in this study.

By exploiting scRNA-seq data we have asked whether imprinted genes as a group are disproportionately represented in the brain, in specific brain regions, and in certain neuronal cell-types. In the adult brain imprinted genes were over-represented in neurons, and particularly the hypothalamic neuroendocrine populations and the monoaminergic hindbrain neurons, with the serotonergic neurons demonstrating the clearest signal. Interestingly, PEGs appear to be the major component of these enrichments, particularly the hypothalamic signal. By extension, these data also identify behaviours that are foci for the action of imprinted genes. Although there are high profile examples of individual imprinted genes that have a role in feeding (*Magel2*) (Schaller et al. 2010), parental behaviour (*Mest)* (Lefebvre et al. 1998) and sleep (*Snord116*) (Lassi et al. 2016), our analyses indicate that imprinted genes *as a group* are strongly linked to these behaviours and also identify other individual genes that should be explored in these domains. Conversely, there are high-profile examples of imprinted genes involved in hippocampus related learning and memory (*Ube3a*) (Jiang et al. 1998), but we did not find enrichment for cell types related to this brain function. The idea that imprinted genes converge of specific physiological or behavioural processes is not unprecedented. Specialisation of function is predicted when considering why genomic imprinting evolved at

all (Moore and Haig 1991; Keverne et al. 1996b; Trivers and Burt 1999; Keverne 2014). Moreover, there is increasing evidence that the imprinted genes themselves appear to be co-expressed in an imprinted gene network (IGN) and have confirmed regulatory links between each other (Varrault et al. 2006; Gabory et al. 2009; Al Adhami et al. 2015). The idea of the IGN or, at the very least, heavily correlated and coordinated expression between imprinted genes adds further support to the idea that imprinted genes work in concert to influence processes, rather than in isolation, and that perturbating one may influence many others (Patten et al. 2016). Our findings add substance to these general ideas and highlight the neuronal regulation of feeding, parental behaviour and sleep as being key functional hotspots on which imprinted genes converge which probably provides the best (but not perfect) basis for discerning evolutionary drivers of genomic imprinting in the brain.

Acknowledgments

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

Competing interests

All authors declare no financial and non-financial competing interests

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.

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**.

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.

Orr HA. 1995. Somatic mutation favors the evolution of diploidy. *Genetics* **139**: 1441-1447.

Patten MM, Cowley M, Oakey RJ, Feil R. 2016. Regulatory links between imprinted genes: evolutionary predictions and consequences. *Proceedings of the Royal Society B: Biological Sciences* **283**: 20152760.

Perez JD, Rubinstein ND, 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* **39**: 347-384.

Perez JD, Rubinstein ND, Fernandez DE, Santoro SW, Needleman LA, Ho-Shing O, Choi JJ, Zirlinger M, Chen S-K, Liu JS. 2015. Quantitative and functional interrogation of parent-of-origin allelic expression biases in the brain. *Elife* **4**: e07860.

Peters J. 2014. The role of genomic imprinting in biology and disease: an expanding view. *Nature Reviews Genetics* **15**: 517-530.

Pulix M, Plagge A. 2020. Imprinted Genes and Hypothalamic Function. In *Developmental Neuroendocrinology*, pp. 265-294. Springer.

Qualls-Creekmore E, Yu S, Francois M, Hoang J, Huesing C, Bruce-Keller A, Burk D, Berthoud H-R, Morrison CD, Münzberg H. 2017. Galanin-expressing GABA neurons in the lateral hypothalamus modulate food reward and noncompulsive locomotion. *Journal of Neuroscience* **37**: 6053-6065.

Rau AR, Hentges ST. 2017. The relevance of AgRP neuron-derived GABA inputs to POMC neurons differs for spontaneous and evoked release. *Journal of Neuroscience* **37**: 7362-7372.

Relkovic D, Doe CM, Humby T, Johnstone KA, Resnick JL, Holland AJ, Hagan JJ, Wilkinson LS, Isles AR. 2010. Behavioural and cognitive abnormalities in an imprinting centre deletion mouse model for Prader–Willi syndrome. *European journal of neuroscience* **31**: 156-164.

Ren J, Isakova A, Friedmann D, Zeng J, Grutzner SM, Pun A, Zhao GQ, Kolluru SS, Wang R, Lin R. 2019. Single-cell transcriptomes and whole-brain projections of serotonin neurons in the mouse dorsal and median raphe nuclei. *Elife* **8**: e49424.

Romanov RA, Zeisel A, Bakker J, Girach F, Hellysaz A, Tomer R, Alpár A, Mulder J, Clotman F, Keimpema E. 2017. Molecular interrogation of hypothalamic organization reveals distinct dopamine neuronal subtypes. *Nature neuroscience* **20**: 176-188.

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**: 4895-4905.

Schaum N, Karkanias J, Neff NF, May AP, Quake SR, Wyss-Coray T, Darmanis S, Batson J, Botvinnik O, Chen MB. 2018. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris: The Tabula Muris Consortium. *Nature* **562**: 367.

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**: 144.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES. 2005. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences* **102**: 15545-15550.

Team R. 2015. RStudio: integrated development for R. *RStudio, Inc, Boston, MA URL http://wwwrstudio.com* **42**: 14.

Team RC. 2013. R: A language and environment for statistical computing. Vienna, Austria.

Tiklová K, Björklund ÅK, Lahti L, Fiorenzano A, Nolbrant S, Gillberg L, Volakakis N, Yokota C, Hilscher MM, Hauling T. 2019. Single-cell RNA sequencing reveals midbrain dopamine neuron diversity emerging during mouse brain development. *Nature communications* **10**: 1-12.

Trivers R, Burt A. 1999. Kinship and genomic imprinting. In *Genomic imprinting*, pp. 1-21. Springer.

Tucci V. 2016. Genomic imprinting: a new epigenetic perspective of sleep regulation. *PLoS genetics* **12**: e1006004.

Tucci V, Isles AR, Kelsey G, Ferguson-Smith AC, Bartolomei MS, Benvenisty N, Bourc'his D, Charalambous M, Dulac C, Feil R. 2019. Genomic imprinting and physiological processes in mammals. *Cell* **176**: 952-965.

Varrault A, Gueydan C, Delalbre A, Bellmann A, Houssami S, Aknin C, Severac D, Chotard L, Kahli M, Le Digarcher A. 2006. Zac1 regulates an imprinted gene network critically involved in the control of embryonic growth. *Developmental cell* **11**: 711-722.

Wen S, Ma D, Zhao M, Xie L, Wu Q, Gou L, Zhu C, Fan Y, Wang H, Yan J. 2020. Spatiotemporal single-cell analysis of gene expression in the mouse suprachiasmatic nucleus. *Nature neuroscience* **23**: 456.

Ximerakis M, Lipnick SL, Innes BT, Simmons SK, Adiconis X, Dionne D, Mayweather BA, Nguyen L, Niziolek Z, Ozek C. 2019. Single-cell transcriptomic profiling of the aging mouse brain. *Nature neuroscience* **22**: 1696-1708.

Zeisel A, Hochgerner H, Lönnerberg P, Johnsson A, Memic F, Van Der Zwan J, Häring M, Braun E, Borm LE, La Manno G. 2018. Molecular architecture of the mouse nervous system. *Cell* **174**: 999-1014. e1022.

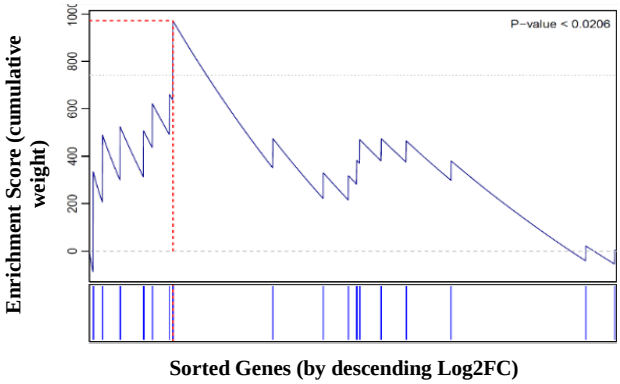
Appendix A – Comparison of Upregulated Genes for GABA5 in Romanov et al. (2017) and GABA17 in Chen et al. (2017).

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.

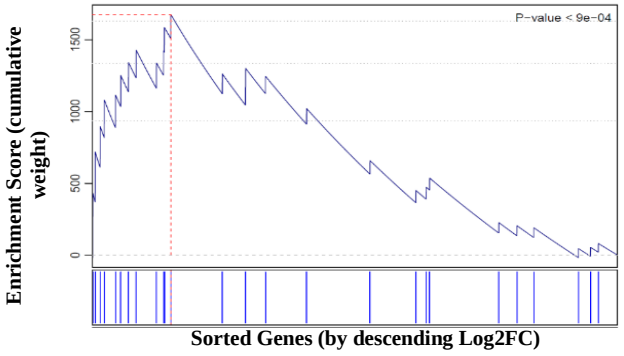
Gene	Position in GABA5 sorted by *p* (41 upreg genes)	GABA5 *p*	GABA5 *q*	GABA5 FC	Position in GABA17 sorted by *p* (823 upreg genes)	GABA17 *p*	GABA17 *q*	GABA17 FC
Calcr	1	2.09E-38	1.30E-36	8.13	41	4.12E-13	4.67E-12	4.96
Lhx1	2	1.82E-10	1.13E-08	9.56	29	1.35E-16	1.53E-15	4.74
Slc6a3	3	6.44E-09	1.33E-07	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 B – 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. For some analyses, Dot plots of Upregulated imprinted genes in that cell type, separated by MEGs and PEGs, were plotted across all cell types in that analysis. Size of points represented absolute mean expression; colour represented the size of the Log2FC value for neuroendocrine cells vs. all other cells.

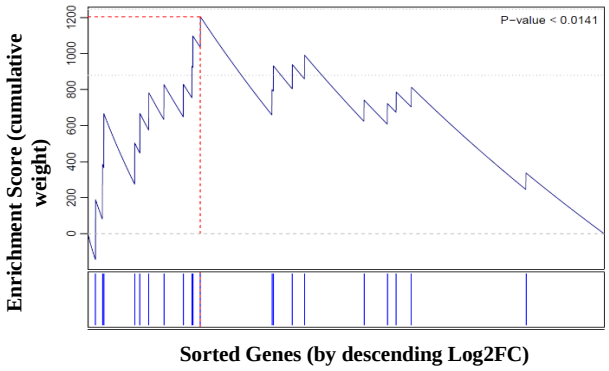
Appendix B1 - Pancreas GSEA of PEGs vin Mouse Cell Atlas (Han et al. 2018)



Appendix B2 - Neuroendocrine Cell GSEA for PEGs in Whole Brain Analysis (Ximerakis et a., 2019)

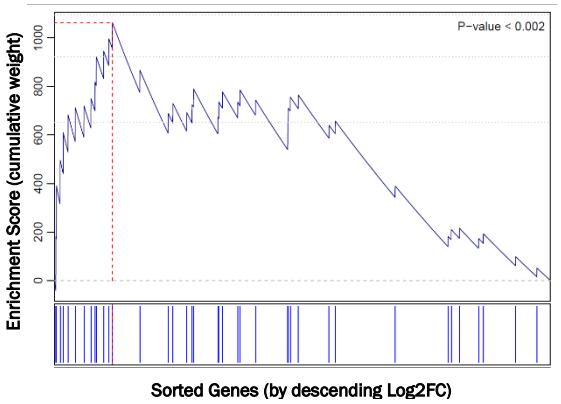


Appendix B3 - Pons GSEA for MEGS in Mouse Brain Atlas Analysis (Zeisel et al. 2018), *q* = 0.028

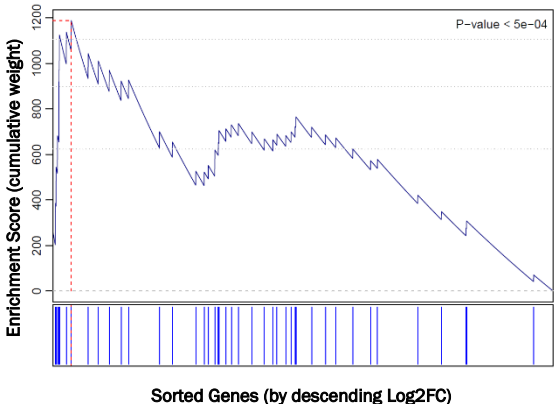


Appendix B4 – Hindbrain Serotonin neurons (left to right, top to bottom - HBSER2, HBSER4, HBSER5) GSEA in Mouse Brain Atlas (Zeisel et al. 2020). Left – GSEA plots for the three enriched serotonin populations after Bonferroni Correction. Right – Dot Plot of upregulated imprinted genes across the 3 enriched serotonin neuron populations plotted across all over-represented neuron subpopulations.

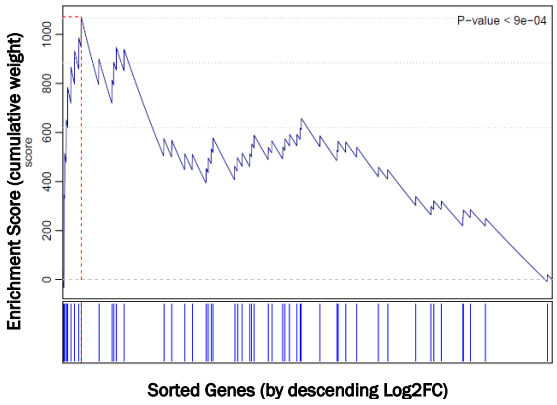
HBSER2 (q = 0.044)

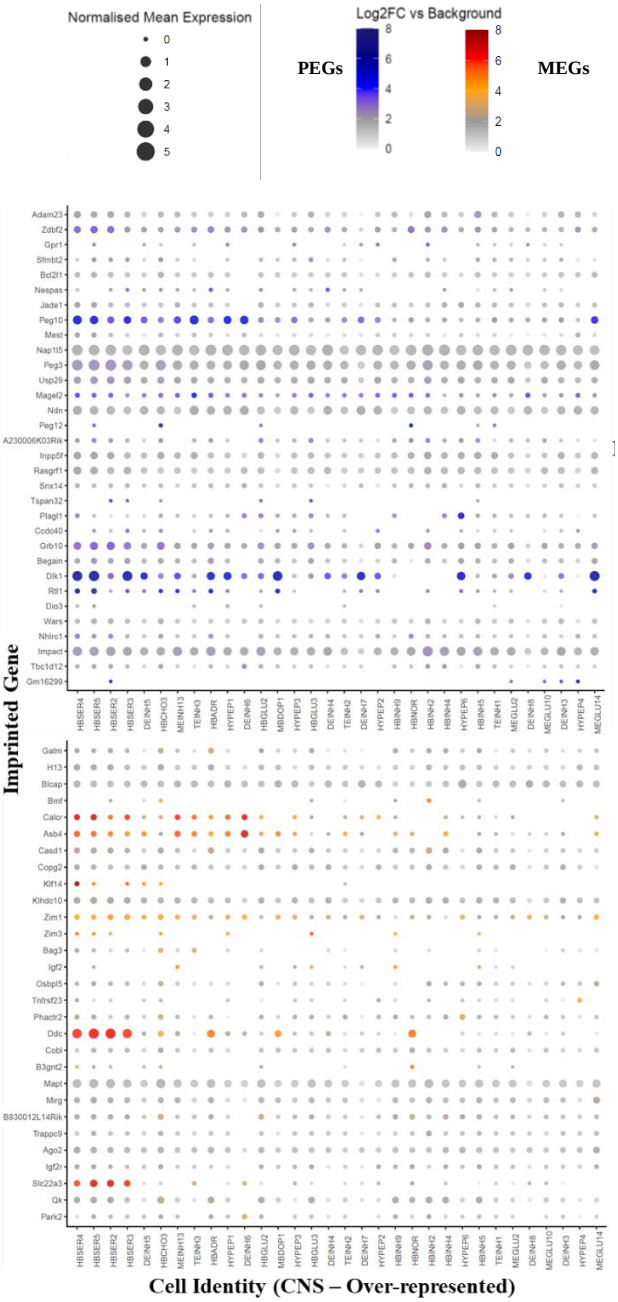


HBSER4 (q = 0.011)

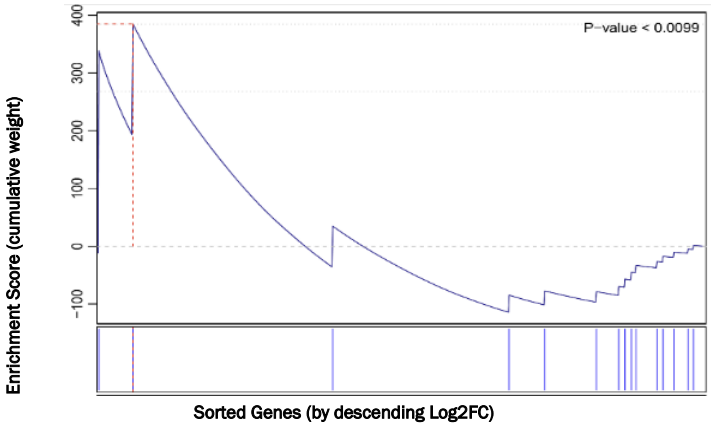


HBSER5 (q = 0.012)

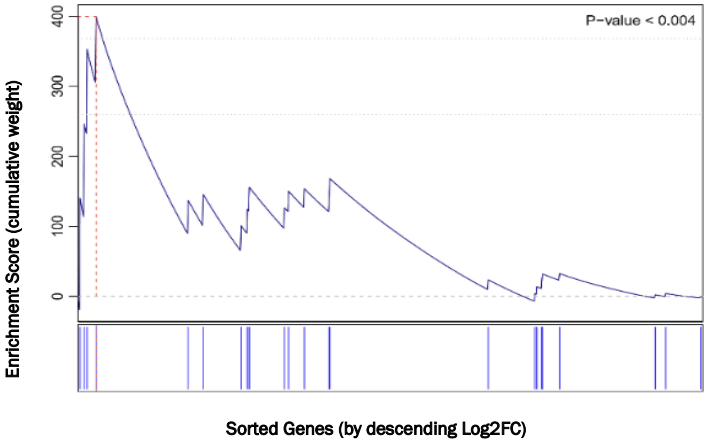


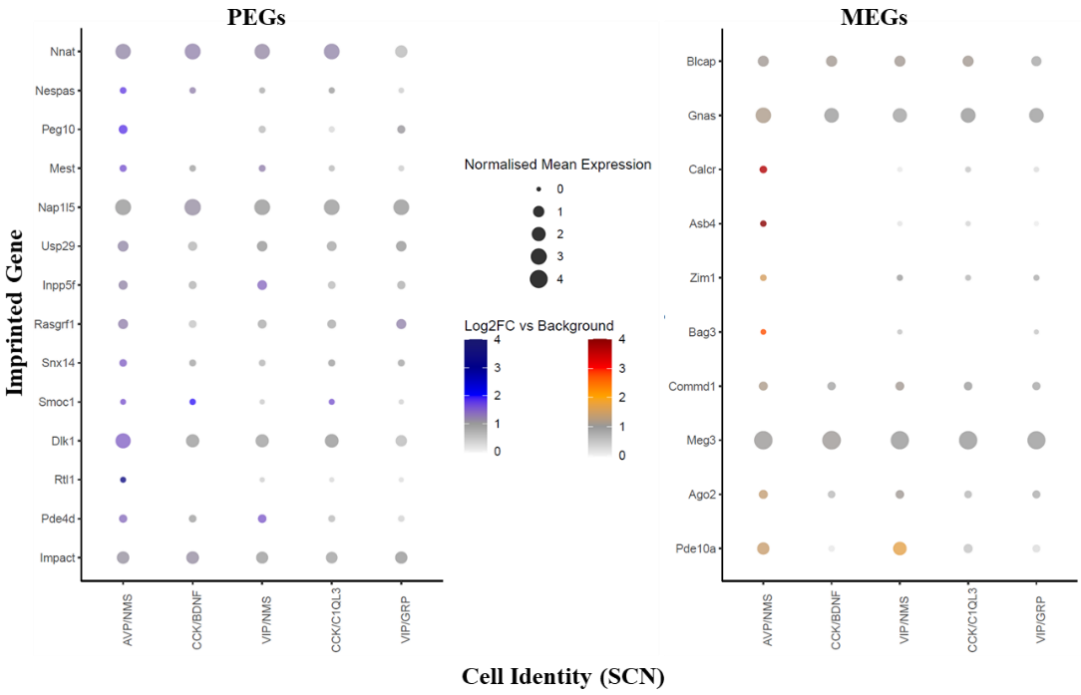


Appendix B5 – Glut_9 (*Synpr/Gad1*) neurons in LHA (Mickelsen et al. 2019), *q* = 0.0198.



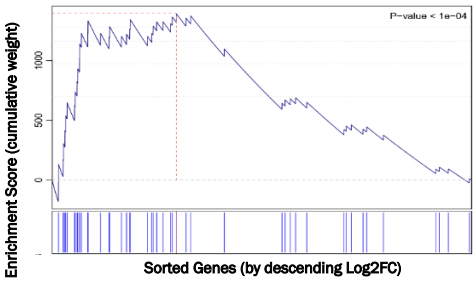
Appendix B6 – *Avp/Nms* neurons in SCN (Wen et al. 2020). Top – GSEA plot for AVP/NMS neurons in SCN, Bottom – Dot Plot of expression of upregulated imprinted genes in Avp/Nms Neurons plotted across all SCN neuron types.





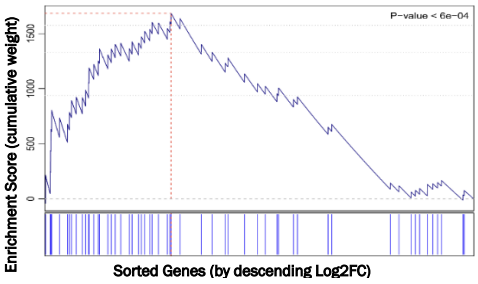
Appendix B7 - Dorsal Raphe Nucleus (Huang et al. 2019) (left, top to bottom – Serotonin neurons, Dopamine neurons, GABA neurons, Glutamatergic neurons, Peptidergic neurons). Right – Dot plots of the expression of all imprinted genes, upregulated across the 5 neuron types, plotted across all cells in the DRN.

Serotonergic Neurons (q = 0.0005)

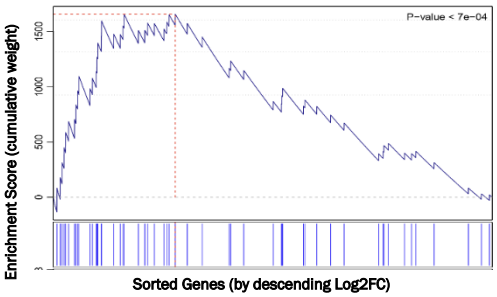
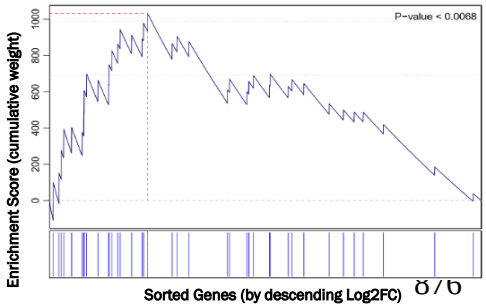
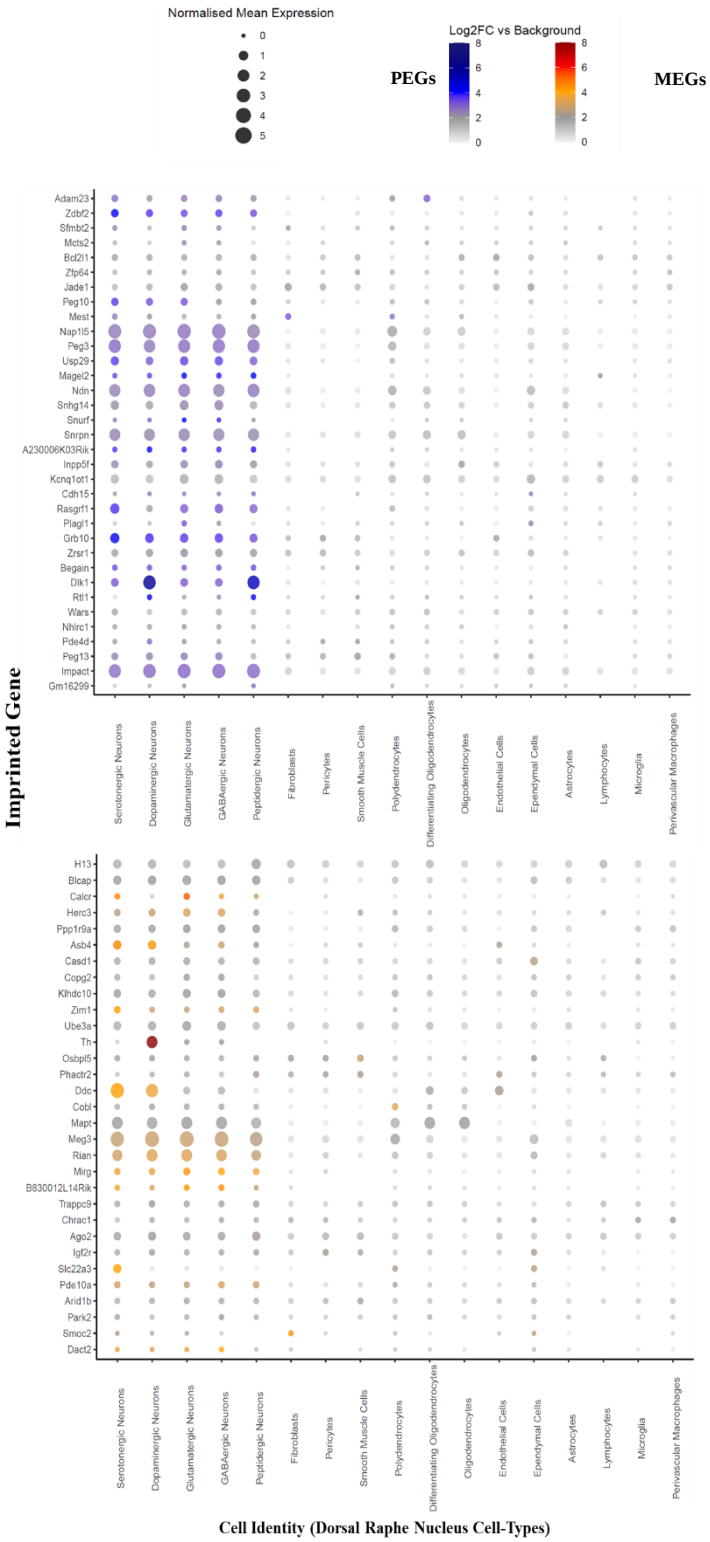


Dopaminergic Neurons (*q* = 0.0005)

GABAergic Neurons (*q* = 0.003)



Glutamatergic Neurons (*q* = 0.0035)



Peptidergic Neurons (*q* = 0.034)