Epigenetic Determinants of Gender in Rats

Tony Hui, Rashedul Hoque, Emma Laks, Emma Titmuss, David Rattray

University of British Columbia, Statistical Methods for High Dimensional Biology

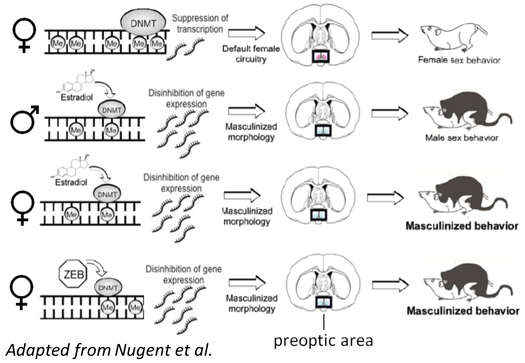
INTRODUCTION

Genetic sex (XX vs. XY) has been held as the dominant sexual differentiation model, causing differentiation of the gonads, which secrete sex hormones such as estradiol, to masculinize the brain. However, recent evidence suggests that epigenetic influences also contribute to sex differences.[1,2,3] Enzymes such as methyltransferases influence the epigenome via the methylation of the genetic code, and male rats are known to have lower DNA (cytosine-5)-methyltransferase 3A (DNMT3a) activity and DNA methylation than females. Nugent et al demonstrated using DNMT inhibitors (zebularine) or estradiol treatment that female rats display masculinized behaviour.[1]

Table 1. Rat experimental conditions for data from Nugent et al.[1]

Sample	RNASeq (Day 2)	WGBS (Day 4)
Male	3 replicates	3 replicates, merged into 1
Male treated with zebularine	3 replicates	-
Female	3 replicates	3 replicates, merged into 1
Female treated with zebularine	3 replicates	-
Female treated with estradiol	-	3 replicates, merged into 1

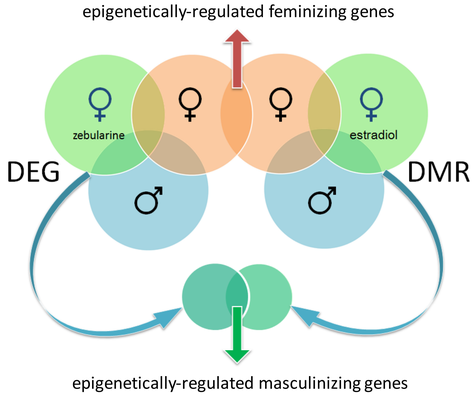
In their study, whole genome bisulfite sequencing (WGBS) was carried out on DNA from the preoptic area of the rat brain (table 1), generating ~270 million raw single-end bisulfite reads. WGBS is a method that detects methylated cytosines in DNA by converting unmethylated cytosine residues to uracil and leaving the methylated cytosine residues intact. RNASeq was also performed on the preoptic area of the rat brain, resulting in an average of 30 million paired-end reads per sample. We used the raw reads generated from Nugent et al to find differentially expressed genes (DEGs) and differentially methylated regions (DMRs) between all conditions to identify gender-related genes that are epigenetically regulated.



Adapted from Nugent et al.

OBJECTIVE

Find overlaps between DEGs and DMRs to reveal potential epigenetically-regulated genes involved in masculinization and feminization of the rat brain.



METHODS

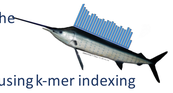
Whole genome bisulphite sequencing analysis:

- Aligned and called CpG methylation using **Bismark.**
- Methylation calls are smoothed across the genome using **BSmooth** - a local likelihood estimator conceptually similar to LOESS smoothing or running average.
- Differentially methylated regions between samples were determined and the nearest gene to the DMR was determined using **HOMER**.[4]

Sanity checking RNASeq and WGBS data:

- Pearson correlated each sample's normalized read counts to determine if any of the samples appeared to be outliers (figure 2). Correlations were visualized using a heatmap.
- A density plot was made of the samples to ensure there were no unexpected spikes in expression or methylation coverage that could have resulted from a technical error and affect the DEGs / DMRs.
  - This led to the merging of female replicate WGBS samples (figure 1A).

RNASeq analysis:

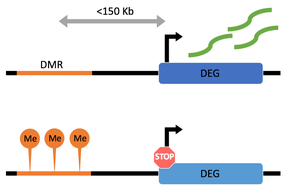


- **SAILFISH** was used to estimate isoform abundances from the RNASeq reads and reference sequences.[5]
- **SAILFISH** uses an alignment free algorithm to estimate the abundances without complexity of read mapping (instead using k-mer indexing and counting), making this a very fast and reliable tool.

Find differentially expressed genes between male vs. female:

- A number of R packages were used to find DEGs between male and female; **edgeR**, **limma**, **DESeq** and **NOISeq** (figure 3A).
- The main approach was the use of **glmQLFit** in **edgeR** to address two types of dispersion; the gene specific dispersion modelled by a quasi-likelihood parameter, and the global negative binomial parameter over all of the genes.

DMR and DEG overlap analysis:



- When overlapping the DMRs and DEGs, the distance was limited to 150kbs as this has been described as the distance chromatin loops occur most frequently.[6]
- Each gene with higher expression in female, and the associated DMR is less methylated in female, was considered to be epigenetically regulated.

WGBS RESULTS

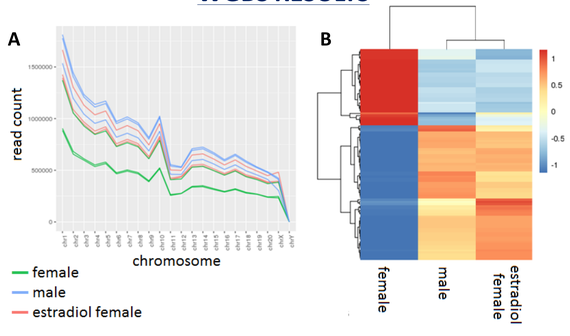


Figure 1. (A) Low coverage of methylation data revealed in female control samples. Each line represents one replicate, colored by the sample. In addition to the low coverage in two of three female replaces, overall the libraries had between 1.4 - 3x coverage per CpG. These results motivated us to merge replicates together in order to improve sensitivity.[7] **(B) Heatmap demonstrating the correlation of methylation regions between estradiol-treated females and males.** Each row represents one of 100 DMRs between female and estradiol samples that overlap the 263 DMRs between female and male that are within 150Kb of a male/female DEG. The colours represent row wise z-score normalized fractional methylation values. Hierarchical clustering was performed on rows and columns using Euclidian distance and complete linkage. In all of the 100 overlapping DMRs, the methylation of male and estradiol females were concordantly higher or lower than female.

RNASEQ RESULTS

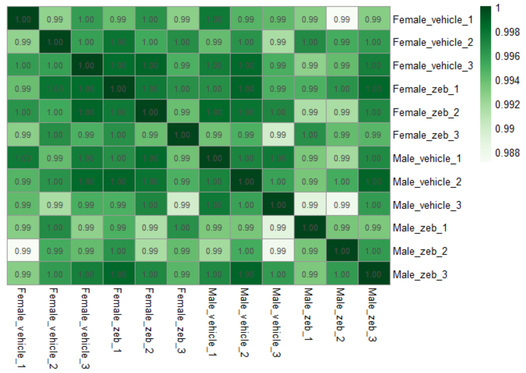


Figure 2. Pearson correlation heatmap of SAILFISH analysed RNASeq data. Following SAILFISH analysis on the RNASeq samples to estimate isoform abundances, the samples were correlated using the Pearson method and displayed in a heatmap. Female_zeb and Male_zeb refer to the samples treated with zebularine, a DNMT inhibitor. Samples correlated very well to others in the same group and so there was no concern with outliers, or a need to merge samples.

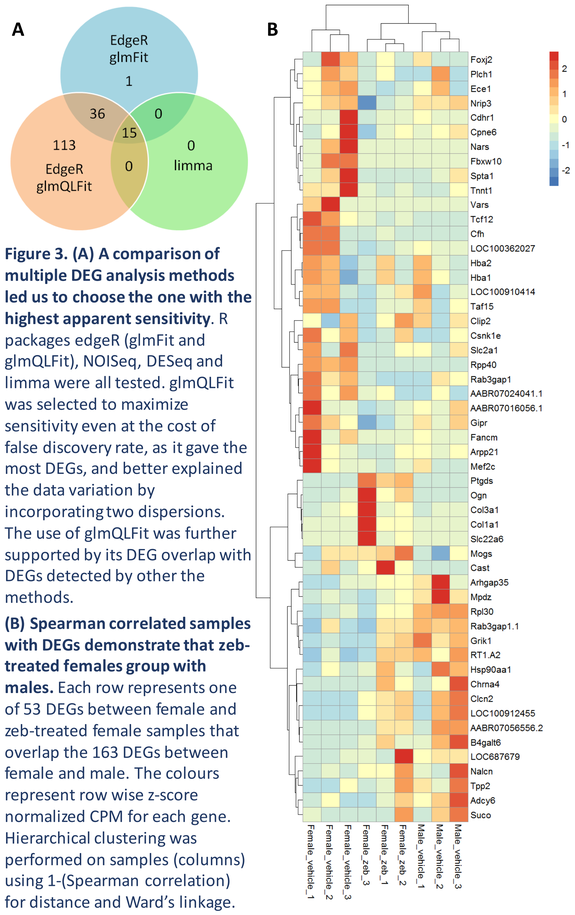


Figure 3. (A) A comparison of multiple DEG analysis methods led us to choose the one with the highest apparent sensitivity. R packages edgeR (glmFit and glmQLFit), NOISeq, DESeq and limma were all tested. glmQLFit was selected to maximize sensitivity even at the cost of false discovery rate, as it gave the most DEGs, and better explained the data variation by incorporating two dispersions. The use of glmQLFit was further supported by its DEG overlap with DEGs detected by other the methods.

(B) Spearman correlated samples with DEGs demonstrate that zeb-treated females group with males. Each row represents one of 53 DEGs between female and zeb-treated female samples that overlap the 163 DEGs between female and male. The colours represent row wise z-score normalized CPM for each gene. Hierarchical clustering was performed on samples (columns) using 1-(Spearman correlation) for distance and Ward's linkage. Row wise clustering was done using Euclidian distance and Ward's linkage. 43 of these DEGs are concordant in expression with DEGs between female and male, and were considered for further analysis.

DMR/DEG OVERLAP RESULTS

Table 2. The final gene list showing feminising and masculinising genes following comparison of differentially methylated regions aligned with their nearest gene to the differentially expressed genes.

DEG	Description	DMR annotation	dist.to.tss	influence
Gipr	gastric inhibitory polypeptide receptor	promoter-TSS	456	feminizing
LOC100362027	ribosomal protein L30-like	intron	-116711	feminizing
Fbxw10	F-box and WD repeat domain containing 10	exon	-110780	feminizing
Tcf12	transcription factor 12	intergenic	120886	feminizing
Plch1	phospholipase C, eta 1	intergenic	107433	feminizing
		intergenic	106442	feminizing
		intergenic	-143383	feminizing
Vars	valyl-tRNA synthetase	intron	-93035	feminizing
Foxj2	forkhead box J2	intron	-54560	feminizing
Hsp90aa1	heat shock protein 90aa1	intergenic	103431	masculinizing
		intron	80776	masculinizing
Adcy6	adenylate cyclase 6	intron	-138546	masculinizing
Tpp2	tripeptidyl peptidase II	intron	-66157	masculinizing

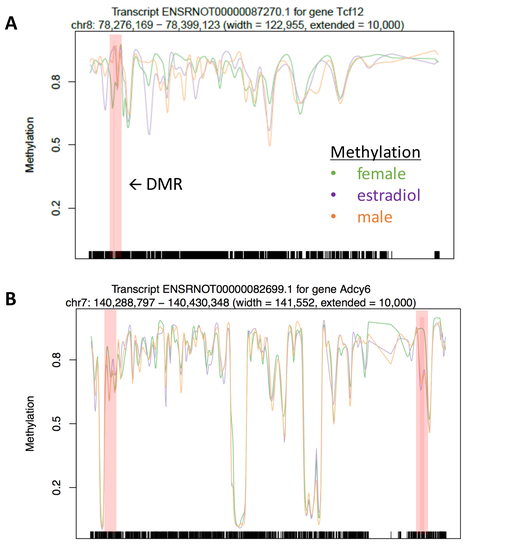


Figure 4. DMR/DEG overlap analysis reveals biologically relevant genes. (A) TCF12, a transcriptional regulator involved in the initiation of neuronal differentiation was found to be both hypomethylated and upregulated in females relative to to males and masculinized females, suggesting it could be a epigenetically regulated feminizing gene.[8] **(B) ADCY6** was both upregulated and associated with a hypomethylated DMR in males and masculinized females compared to vehicle females, making it a putative epigenetically regulated masculinizing gene. ADCY6 is an adenylate cyclase; this family of proteins catalyzes the formation of the signaling molecule cAMP and is known to be involved in intracellular signalling,[9] and in developmentally and neurophysiologically relevant signaling pathways, including neural plasticity.[10]

DISCUSSION

Our study presents a method for analysis of WGBS and RNASeq data to integrate DMRs and DEGs, and uses it to identify ten putative genes (table 2) that determine sexual differentiation and behaviour and can be epigenetically regulated. Two of these genes, TCF12 and ADCY6, are already known to have biological significance to neurology. Our findings provide insight into the mechanisms of gender, and may help explain non-heterosexual gender identities in mammals.

REFERENCES

1. Nugent et al. Brain feminization requires active repression of masculinization via DNA methylation, Nature Neuroscience (2015) 18:690–697. doi:10.1038/nn.3882
2. Matsuda et al. Epigenetic mechanisms are involved in sexual differentiation of the brain. Reviews in Endocrine and Metabolic Disorders (2012) 13(3):163-171. doi:10.1007/s11154-012-9202-z
3. McCarthy et al. Reframing sexual differentiation of the brain. Nature Neuroscience (2011) 14:677–683. doi:10.1038/nn.2834
4. Heinz et al. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol Cell (2010) 28.38(4):576-589. doi:10.1038/nn3949
5. Patro et al. Sailfish enables alignment-free isoform quantification from RNA-seq reads using lightweight algorithms. Nature Biotechnology (2014) 32:462–464. doi:10.1038/nbt.2862
6. van Arensbergen et al. In search of the determinants of enhancer–promoter interaction specificity. Trends in Cell Biology (2014) 24(11):695–702. doi:10.1016/j.tcb.2014.07.004
7. Zilber et al. Coverage recommendations for methylation analysis by whole-genome bisulfite sequencing. Nature Methods (2015) 12:230–232. doi:10.1038/nmeth.3152
8. Uittenbogaard et al. Expression of the bHLH transcription factor Tcf12 (ME1) gene is linked to the expansion of precursor cell populations during neurogenesis. Gene Expression Patterns (2002) 1(2):115–121. doi:10.1016/S1567-133X(01)00025-9
9. Sunabra et al. Neuroanatomical distribution and neurochemical characterization of cells expressing adenylyl cyclase isoforms in mouse and rat brain. Journal of Chemical Neuroanatomy (2011) 41(3):143–54. doi:10.1016/j.jchemneu.2010.11.001
10. Yael et al. Comprehensive analysis of the expression patterns of the adenylate cyclase gene family in the developing and adult mouse brain. Journal of Comparative Neurology (2006) 496(5):684–697. doi:10.1002/cne.20953

CONTACT

GitHub IDs: hui-tony-zk, RashedHUBC, eclaks, emminic93, david-rattray

Project repository containing data, methods, analysis, and research: https://github.com/STAT540-UBC/team_treed_rats-DNA-methylation