**Capstone Report 1**

**Problem:** In mammals, neural development is highly sensitive to the fetal-maternal environment. For example, prenatal exposure of the brain to excess glucocorticoids, through maternal stress or exogenous administration of synthetic glucocorticoids (sGC) has been shown to have deleterious effects on development and function that persist in adolescence and adulthood [[1-3](#_ENREF_1)]. Glucocorticoids are a class of steroid hormone which provide developmental triggers during gestation [[4](#_ENREF_4)]. During gestation, there is a natural surge of glucocorticoids which occurs 10-15 days before delivery in most mammalian species. This surge is essential for normal maturation of multiple organ systems including the thyroid, kidney, lungs, and brain [[5](#_ENREF_5)]. Mothers that are at risk for preterm labor, however, are administered a dose of sGC to reduce the incidence of respiratory distress syndrome in the fetus. The effects of sGC on the developing brain have yet to be fully elucidated, while recent evidence suggests that multiple courses of sGC have lasting behavioral effects in children exposed in utero that were subsequently born at term [[3](#_ENREF_3)]. The hypothalamic paraventricular nucleus (PVN) plays a central-role in regulating stress response and behavior, as well as energy homeostasis. In this research, we investigate the effects of multiple courses of sGC exposure on the gene expression profiles of the PVN of juvenile female offspring. Furthermore, we employ predictive modelling techniques to further understand the relationship between gene expression and the behavioral profiles observed in these animals.

**Methods:** Pregnant guinea pigs received 3 courses of betamethasone (Beta;1mg/kg) or saline (C) in late gestation. Total locomotor activity in open-field (OFA) was measured in female offspring on postnatal day 24 and brains collected at day 40. PVN was micro-punched (C;n=5, Beta;n=5) and RNA was extracted. mRNA library preparation was performed using Illumina TruSeq V2 mRNA enrichment using standard protocols. High-throughput sequencing were performed on an Illumina HiSeq 2500 sequencing system using standard run, following the protocol recommended by Illumina for sequencing mRNA samples. Sequencing was done for each biological replicate at 1 × 51 bp by the Donnelly Centre for Cellular and Biomolecular Research. Sequence reads were aligned to the *Cavia porcellus* reference genome (cavPor3.83) with Tuxedo Suit tools [[6](#_ENREF_6)] accessed through The Galaxy Project [[7](#_ENREF_7)]. Sequencing depth for RNA-seq samples averaged 45 million reads per biological sample with >85% overall alignment rate. Subsequent analyses were performed in R (version 3.2.3). Gene read counts were determined with Genomic Alignments (version 1.6.3) as described by the authors [[8](#_ENREF_8)]. Outliers were removed using Cook’s distance with default cutoffs [[9](#_ENREF_9)], and data were normalized by residuals with RUVSeq (version 1.4.0) [[10](#_ENREF_10)]. Differential gene expression was assessed using EdgeR’s (version 3.12.1) [[11](#_ENREF_11), [12](#_ENREF_12)] general linear model likelihood ratio test and FDR-corrected *p* < 0.05 was considered significant. Principal component analysis (PCA) was carried out on normalized expression profiles of significantly up-regulated genes, and OFA scores and displayed on a circle of correlations. Relationships were analyzed by linear regression. Multiple regression combined gene profiles to predict OFA, linear regression determined the correlation of predicted and observed OFA.

**Results:** Differential expression analysis revealed that 597 genes were significantly upregulated (p < 0.05, FDR 5%) and 161 genes were significantly down-regulated (p < 0.05, FDR 5%) in the sGC exposed animals. PCA showed OFA is associated with expression of Greb1l (estrogen receptor signaling), Prlr (prolactin receptor), & Trim66 (transcriptional regulator). Linear regression revealed the correlation to be significant (Greb1l: R2=0.71, p=0.002, Prlr: R2=0.51, p=0.019, Trim66: R2=0.58, p=0.01), and significant correlation between predicted and observed OFA (R2=0.80, p=0.015). Leave One Out (LOO) cross-validation was employed to determine whether there was a possibility that this model would generalize to novel data. A significant correlation between predicted and observed OFA after LOO cross validation was observed (R2=0.51, p=0.03).

**Conclusions:** This is the first evidence of a correlation between stress-activated locomotor behavior and gene expression in the PVN following prenatal sGC. Interestingly, this association focused on a subset of genes that are co-expressed and involved in regulation of transcription and sex-hormone signaling. The major value of this analytical approach is that it enabled us to narrow the scope from over 700 genes that were significantly differentially expressed, to three genes that can be focused on in further investigations. These findings provide insight into the potential mechanisms of antenatal sGC and how these molecular events relate to behavior. Furthermore, we provide proof of principal for the use of gene expression modelling in disease prediction, detection, and prevention.

1. French, N.P., et al., *Repeated antenatal corticosteroids: effects on cerebral palsy and childhood behavior.* Am J Obstet Gynecol, 2004. **190**(3): p. 588-95.

2. Glover, V., *Annual Research Review: Prenatal stress and the origins of psychopathology: an evolutionary perspective.* J Child Psychol Psychiatry, 2011. **52**(4): p. 356-67.

3. Alexander, N., et al., *Impact of antenatal synthetic glucocorticoid exposure on endocrine stress reactivity in term-born children.* J Clin Endocrinol Metab, 2012. **97**(10): p. 3538-44.

4. Kapoor, A., S. Petropoulos, and S.G. Matthews, *Fetal programming of hypothalamic-pituitary-adrenal (HPA) axis function and behavior by synthetic glucocorticoids.* Brain Res Rev, 2008. **57**(2): p. 586-95.

5. Fowden, A.L., J. Li, and A.J. Forhead, *Glucocorticoids and the preparation for life after birth: are there long-term consequences of the life insurance?* Proc Nutr Soc, 1998. **57**(1): p. 113-22.

6. Trapnell, C., et al., *Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks.* Nat Protoc, 2012. **7**(3): p. 562-78.

7. Afgan, E., et al., *The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2016 update.* Nucleic Acids Research, 2016. **44**(W1): p. W3-W10.

8. Lawrence, M., et al., *Software for Computing and Annotating Genomic Ranges.* PLoS Comput Biol, 2013. **9**(8): p. e1003118.

9. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.

10. Risso, D., et al., *Normalization of RNA-seq data using factor analysis of control genes or samples.* Nat Biotechnol, 2014. **32**(9): p. 896-902.

11. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-40.

12. Zhou, X., H. Lindsay, and M.D. Robinson, *Robustly detecting differential expression in RNA sequencing data using observation weights.* Nucleic Acids Res, 2014. **42**(11): p. e91.