Solving for X: evidence for sex-specific autism biomarkers across multiple transcriptomic studies

Samuel C. Lee1,+,\*, Thomas P. Quinn1,2,3,+,\*, Jerry Lai4, Sek Won Kong5, Irva Hertz-Picciotto6, Stephen J. Glatt7, Tamsyn M. Crowley2,3,8, Svetha Venkatesh1, and Thin Nguyen1

1Centre for Pattern Recognition and Data Analytics (PRaDA), Deakin University, Geelong, 3220, Australia 2Centre for Molecular and Medical Research, Deakin University, Geelong, 3220, Australia 3Bioinformatics Core Research Group, Deakin University, Geelong, 3220, Australia 4Deakin eResearch, Deakin University, Geelong, 3220, Australia | Intersect Australia, Sydney, 2000, Australia 5Computational Health Informatics Program, Boston Children’s Hospital, Boston, MA, USA | Department of Pediatrics, Harvard Medical School, Boston, MA, USA 6Department of Public Health Sciences and UC Davis MIND Institute, School of Medicine, Davis, California 7Psychiatric Genetic Epidemiology and Neurobiology Laboratory (PsychGENe Lab) | SUNY Upstate Medical University, Syracuse, NY, USA 8Poultry Hub Australia, University of New England, Armidale, New South Wales, 2351, Australia + contributed equally, \* *samleenz@me.com; contacttomquinn@gmail.com*

**Abstract**

Autism spectrum disorder (ASD) is a markedly heterogeneous condition with a varied phenotypic presentation. Its high concordance among siblings, as well as its clear association with specific genetic disorders, both point to a strong genetic etiology. However, the molecular basis of ASD is still poorly understood, although recent studies point to the existence of sex- specific ASD pathophysiologies and biomarkers. Despite this, little is known about how exactly sex influences the gene expression signatures of ASD probands. In an effort to identify sex- dependent biomarkers and characterise their function, we present an analysis of a single paired- end post-mortem brain RNA-Seq data set and a meta-analysis of six blood-based microarray data sets. Here, we identify several genes with sex-dependent dysregulation, and many more with sex-independent dysregulation. Moreover, through pathway analysis, we find that these sex-independent biomarkers have substantially different biological roles than the sex-dependent biomarkers, and that some of these pathways are ubiquitously dysregulated in both post- mortem brain and blood. We conclude by synthesizing the discovered biomarker profiles with the extant literature, by highlighting the advantage of studying sex-specific dysregulation directly, and by making a call for new transcriptomic data that comprise large female cohorts.

/Users/slee/Dropbox/Thin_Thom/solveForX/results/upset/gobp-upset.pdf

Figure 1: This UpSet plot shows set intersections (and their sizes) for a GSEA analysis against the Gene Ontology Biological Process pathways. For each of the four GSEA runs, all intersections of the significant results were calculated and totalled and are represented by (a) set identity by the joined lines and (b) set size by the top bar-chart. The bar-chart on the left shows the total set size for each individual GSEA run. Results are filtered using a liberal FDR threshold of FDR < 0.15 for the RNA-Seq data and FDR < 0.3 for the Meta-analysis data.

../../results/upset/reactome-upset.pdf

Figure 2: This UpSet plot shows set intersections (and their sizes) for a GSEA analysis against the Reactome pathways. For each of the four GSEA runs, all intersections of the significant results were calculated and totalled and are represented by (a) set identity by the joined lines and (b) set size by the top bar-chart. The bar-chart on the left shows the total set size for each individual GSEA run. Results are filtered using a liberal FDR threshold of FDR < 0.15 for the RNA-Seq data and FDR < 0.3 for the Meta-analysis data.

../../results/upset/hallmark-upset.pdf

Figure 3: This UpSet plot shows set intersections (and their sizes) for a GSEA analysis against the MSigDB Hallmark pathways. For each of the four GSEA runs, all intersections of the significant results were calculated and totalled and are represented by (a) set identity by the joined lines and (b) set size by the top bar-chart. The bar-chart on the left shows the total set size for each individual GSEA run. Results are filtered using a liberal FDR threshold of FDR < 0.15 for the RNA-Seq data and FDR < 0.3 for the Meta-analysis data.