# Title Page

# Gene expression analysis in peripheral blood of first episode psychosis patients

Short Title: Expression analysis in blood of psychosis patients

Daniel J Leirer1,2 , Conrad O Iyegbe1,3 , Marta Di Forti1,3 , Hamel Patel1,2 , Elena Carra3, Sara Fraietta3 , Marco Colizzi3, Valeria Mondelli4, Diego Quattrone1 , John Lally3, Olesya Ajnakina3, Sang Hyuck Lee1,2 , Charles J Curtis1,2 , Gerome Breen1,2 , Carmine Pariente4, Katherine Aitchison1, Paola Dazzan1,5 , Robin M Murray3, Richard J B Dobson1,2 and Stephen J Newhouse1,2 \*

Author details:

1 MRC Social, Genetic and Developmental Psychiatry (SGDP) Centre, Institute of Psychiatry, Psychology and Neuroscience, Box P080, De Crespigny Park, SE5 8AF London, UK. 2 NIHR Biomedical Research Centre for Mental Health and Biomedical Research Unit for Dementia and South London at Maudsley NHS Foundation Trust, Box P092, De Crespigny Park, SE5 8AF London, UK. 3 Department of Psychosis Studies, Institute of Psychiatry, Kings College London, Box P092, De Crespigny Park, SE5 8AF London, UK. 4 Department of Psychological Medicine, Institute of Psychiatry, Kings College London, Box P092, De Crespigny Park, SE5 8AF London, UK. 5 Departments of Psychiatry and Medical Genetics, University of Alberta, Alberta, Canada, 11361-87 Avenue, AB T6G 2E1, Edmonton, Canada.

\* Correspondence: stephen.newhouse@kcl.ac.uk 1 MRC Social, Genetic and Developmental Psychiatry (SGDP) Centre, Institute of Psychiatry, Psychology and Neuroscience, Box P080, De Crespigny Park, SE5 8AF London, UK

Transcriptomics; WGCNA; Psychosis; Schizophrenia; Blood; Neurogranin

# Abstract:

**Keywords:**

**Abbreviations:**

# 1.0 Introduction

Copy Introduction back in.

# 2.0 Methods and Materials

## 2.1 Ethics

The Study received ethical approval from the South London and Maudsley (SLaM), as well as from the Institute of Psychiatry Local Research Ethics Committee, IOP/SLAM research ethics approval number: 135/05. Informed written consent was obtained from all participants in the study prior to start.

## 2.2 Study design and participants

As part of the GAP study (Di Forti et al., 2009, 2015). We approached all patients aged 18–65 years who presented with first-episode psychosis at the inpatient units of SLaM were approached. We invited patients to participate if they met the International Classification of Diseases 10 criteria for a diagnosis of non-affective (F20–F29) or affective (F30–F33) psychosis, validated by administration of the Schedules for Clinical Assessment in Neuropsychiatry (SCAN). We excluded individuals who met the criteria for organic psychosis (F09). If patients were too unwell to cooperate, we re-contacted them after the start of treatment. Between May 1, 2005, and May 31, 2011, we recruited 461 patients with first-episode psychosis. The cohort consisted of a diverse multi-ethnic population. Further patient information, blood samples and genetic ancestry where acquired as described previously (Di Forti et al., 2012). During the same period, we recruited 389 control individuals, aged 18–65 years, who were similar to the local population in terms of gender, ethnic origin, education, and employment status, and socio-economic status. We recruited controls using internet and newspaper advertisements and by distributing leaflets at train stations, shops, and job centres. Volunteers were administered the Psychosis Screening Questionnaire (Bebbington & Nayani, 1996) and were excluded if they met the criteria for a psychotic disorder or if they reported a previous diagnosis of psychotic illness.

## 2.3 RNA processing and Quality Control

Whole blood samples were collected using PAXgene tubes for RNA, from a subset of GAP participants (227 cases and 168 controls). Psychotic patients were stabilized using anti-psychotics for a week. Samples were run at the NIHR Biomedical Research Centre for Mental Healthy (BRC-MH) microarray facility at the SGDP, Institute of Psychiatry, and King’s College London. Microarrays where run in accordance with the manufacturer’s protocol using Illumina HT-12 V4 beadchips (Illumina, USA).

All analysis was performed using R version 3.1.2 (Team, 2013). We performed rigorous quality control, by pre-processing the data using an adapted in-house developed pipeline (https://github.com/snewhouse/BRC\_MH\_Bioinformatics).

The pipeline takes raw gene expression data exported from Illumina’s Genomestudio, performs background correction (Xie, Wang, & Story, 2009) using negative bead expression levels in order to correct for noise. Lumi (version 2.22.1 (Du, Kibbe, & Lin, 2008)) was used to log base 2 transform the data followed by robust spline normalization (Du et al., 2008). Outlying samples were iteratively identified using fundamental network concepts and removed, following the methods described by Oldham et al. (Oldham, Langfelder, & Horvath, 2012).

In order to reduce the influence of batch effects we identified significant confounding variables by using the first principle component of housekeeping and undetected probes and regressing this against technical variables. In cases where the variables were significantly associated with the first principle component, they were regressed against expression for each probe, and the mean adjusted residuals were taken forward. The resulting adjusted expression matrix was subjected to surrogate variable analysis, using the SVA package (Leek, Johnson, Parker, Jaffe, & Storey, 2012), to identify potential unknown batch effects. Following this we compared recorded gender with gender determined by XIST and PRKY probes, and excluded samples that showed a mismatch. Finally, we excluded all probes that could not be reliably detected in 80% of the samples in at least one diagnostic group. We used the R package CellMix version 1.6 (Gaujoux & Seoighe, 2013), to test for potential significant differences in whole blood cell populations between cases and controls. Prior to further analysis we controlled for CellMix derived cell proportions, Age, Gender and Ethnicity using a linear model to create an adjusted expression matrix.

## 2.4 Differential gene expression analysis

Copy

## 2.5 Weighted Gene Co-expression Network Analysis

Copy

## 2.6 Gene enrichment analysis

## copy

## 2.7 Post hoc analysis of medication

## Describe stats, only have AF vs Control, Med vs Control, Med vs AF.

Supplementary, add Ola and Ris vs Control and AF (4 categories).

Also Describe Medication added in WGCNA.

# 3.0 Results

## 3.1 Sample Characteristics

Basically copy. Add med data

## 3.2 Differential Expression analysis

Copy and edit heavily

## 3.3 Weighted Gene Co-expression Network Analysis

Copy and edit Heavily

## 3.4 Functional Enrichment analysis

Copy and edit heavily

## 3.5 Post hoc analysis of medication

Describe results. Namely in reduced data, we see large statistically significant differences between AF and Control, and Med vs Control. Med vs Control, is more, AF has more logFC spread.

# 4.0 Discussion

## 4.1 Glutamate

Rewrite and purge

## 4.2 Platelets and DiGeorge syndrome

Rewrite and purge

## 4.3 Defensins

Rewrite and purge

## 4.4 WGCNA

Rewrite and purge

## 4.5 Enrichment Results

Rewrite and purge

## 4.6 Confounding by medication

Summary of WGCNA and volcanoplots. Note medication results are limited and cannot fully be addressed. Spread is more pronunced in AF than antipsychotics. May indicate effect of antipsychotics. Too noisy to be able to eliminate the possibility that medication has no effect. But the authors interpretation is that an effect is liekly which is backed by the literature., The extent of this is debatable.

Mention directionality in modules. BMI and Tobacco are included.

## 4.7 Conclusion and Limitations

Copy and edit

# Acknowledgements:

# Financial Disclosures:

# References:

# Table/Figure Legends: