README\_diff\_expression\_Project

DJL

15/09/2016

# Project Overview

This Project is part of Daniel Leirer's PhD. I work primarily with the Genes and Psychosis data from the IoPPN. The aim is to identify differentially expressed genes and networks.

# Folder Structure

proj/  
├── P0\_Characterise/  
│├── figs/  
│└── output/  
├┬─ P1\_Diff\_Ex/  
│├── figs/  
│└── output/  
├┬─ P2\_WGCNA/  
│├── figs/  
│└── output/  
├┬─ P3\_Enrichment/  
│├── figs/  
│└── output/  
├┬─ P4\_Posthoc/  
│├── figs/  
│└── output/  
├┬─ P5\_Additional\_dataset\_Posthoc/  
│├── figs/  
│└── output/  
├── R/  
├── data/  
└── doc/

**Directory Function:**  
root directory = contains README and Project files for version control.  
P0-PXX = subproject directories  
figs = contains figures for subproject  
output = contains output files from subproject.  
R = Contains Scripts  
data = contains all data (with exception of large data frames. Uploaded on Google Drive.)  
doc = contains paper, labbook and other documentation.

# Subproject 0: Characterise

**Aim**:  
To make sure data is suitable for future steps, by giving overview of demographics and data available. Also Cell Mix is performed to make sure everything is okay down the line.

**Script\_Strategy:**

1. *p0\_CellMix\_GAP\_FEP\_data\_03\_11\_2016.Rmd*

* apply cellmix to identify poetential confounders from cell proportions.

1. *p0\_characterise\_GAP\_FEP\_data\_03\_11\_2016.Rmd*

* Characterise cohort. Plot demographics, make tables, do stats.

**Output:**

# Subproject 1: Diff Expression

**Aims:**  
Differential Expression adjusting for Ethnciity Age and Gender using Limma.

**Input:**  
LumiBatch object.

**Script\_Strategy:**

1. *p1\_limma\_03\_11\_2016.Rmd*

* Script that does limma.

**Output:**  
List of expressed probes no duplicates for all samples.  
List of unexpressed probes no duplicates.  
Differentially Expressed Genes. Figures Volcano Plot.

# Subproject 2: WGCNA

**Aims:**  
WGCNA to identify modules.

**Input:**  
Lumibatch Object

**Script\_Strategy:**

1. *p2\_WGCNA\_03\_11\_16.Rmd*

* Script for WGCNA.

**Output:**  
Figures .  
Files .

# Subproject 3: Enrichment

**Aims:**  
Enrichment of Diff Ex and WGCNA results.

**Input:**  
Lumibatch Object

**Script\_Strategy:**

1. *p3\_ENRICHMENT\_05\_11\_16.Rmd*

* Script for userlistenrichment.

**Output:**  
Files.  
Enrichment files.

# Subproject 4: P4\_Posthoc

**Aims:**  
Find what the impact of Medication and Tobacco and BMI is.  
What genes are affected? Do Limma Ideally.

**Input:**  
Lumibatch Object

**Script\_Strategy:**

1. *p4\_Posthoc\_med\_21\_11\_16.Rmd*

* Script for userlistenrichment.

**Output:**  
Files.  
Diff expression:  
AF\_vs\_Con Ris\_vs\_Con Ola\_vs\_AF Ris\_vs\_AF Ris\_vs\_Ola Med\_vs\_AF\_Con

Enrichment files.

## Genes and Psychosis (GAP)

The Main Dataset is the Genes and Psychosis Data internal to the IoPPN.  
The following files are associated to this Data.

File: **Full Gene Expression Object**  
*GAP\_FEP\_Full\_Gene\_Expression\_Data\_Linear.RData*  
Description: This file contains a lumibatch object with all probes deemed expressed. It is from the gene expression pre processing pipeline. Most probes here are defined as not expressed. It is First Episode Gap Samples, processed using background correction, log 2 transformed, robust spline normalisation. The following tech variables have been regressed out using a linear model:  
ConcNanodrop, Dateout, concentrationoflabelledcRNA, DatecRNApurification

Source file name:  
*FINAL\_GAP\_DL\_FEP.eset\_bg\_log2\_rsn\_SVA\_Good.RData*  
Source:  
Daniel Leirer  
[Daniel.Leirer@kcl.ac.uk](mailto:Daniel.Leirer@kcl.ac.uk)

## The following files are sources for various parts of the data in the Lumibatch Object

File: **Demographic Data GAP**  
*Basic\_Demographics\_GAP.csv*

Description:  
Data approved by Marta Di Forti. Contains gene expression data, demographics etc. Age, Sex, Phenotype etc.

Source file name: *GAP\_full\_final\_expression\_database\_22\_04\_2015\_Dan\_Marta\_consent.csv*

Source:  
Daniel Leirer created this document. Information from various sources within GAP.

File: **GAP Master Database**  
*GAP\_large\_demographic\_database\_16\_Oct\_2014.sav*

Description:  
Huge database containing a lot of information including PANSS data. This is a secordary database.

Source file name: *Master\_database\_GAP\_UPDATE\_16\_Oct\_2014.sav*

Source:  
GAP team. Contact Robin Murray, Marta Di Forti, or people working in the Psychosis department.

File: **Medication Data**  
*Daniel\_RNA\_DQ.csv*

Description:  
Medication Data, Weight, Smoking, Some Demographics.

Source:  
Diego Quattrone compiled this file. [diego.quattrone@kcl.ac.uk](mailto:diego.quattrone@kcl.ac.uk)

# Acknoledgements

The Following people are involved in this project

Name: Daniel Leirer Role: PhD Student Email: [daniel.leirer@kcl.ac.uk](mailto:daniel.leirer@kcl.ac.uk)

Name: Dr. Stephen Newhouse Role: Main Supervisor Email: [stephen.j.newhouse@gmail.com](mailto:stephen.j.newhouse@gmail.com)

Name: Professor Richard Dobson Role: Primary Supervisor Email: [richard.j.dobson@kcl.ac.uk](mailto:richard.j.dobson@kcl.ac.uk)

Name: Sir Professor Robin Murray Role: Clinical Supervisor Email: N/A