Down syndrome and congenital heart disease: RNA-seq reveals differentially expressed cardiac tissue related genes

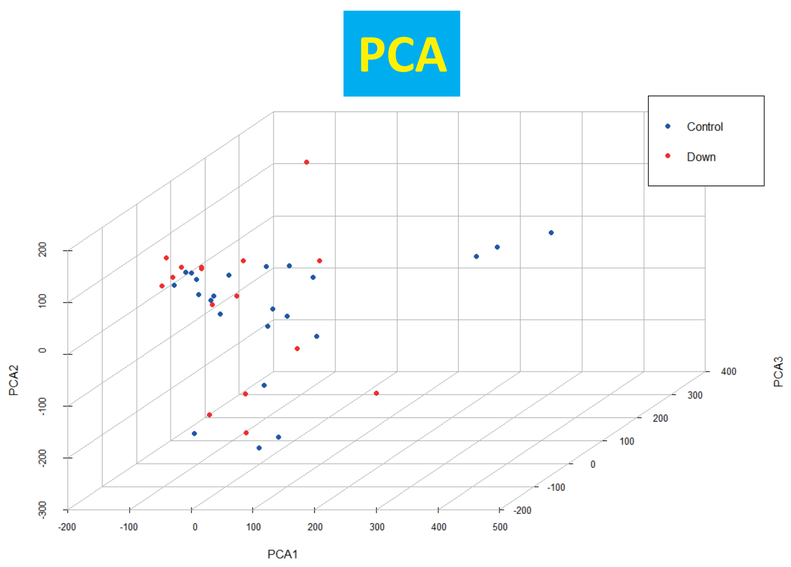
Adriano Voltolini[1], Mario Lauria[2]
[1]Department of Cellular, Computational and Integrative Biology, University of Trento, Povo (TN), Italy
[2]The Microsoft Research - University of Trento Centre for Computational and Systems Biology, Rovereto (TN), Italy

Introduction

Down syndrome (DS), also known as trisomy 21, is the most common chromosome anomaly in humans. About 44.3% of people with DS are diagnosed with a type of Congenital Heart Disease (CHD), which is the main cause of death in this population during the first two years of life.
Three of the most common heart conditions seen in children with DS are atrioventricular septal defect, patent ductus arteriosus, and tetralogy of Fallot. The molecular causes for these cardiac defects are not yet entirely understood.

The objective of this study is to identify these molecular causes using a RNA-seq data analysis approach.

PCA



The first 3 principal components explain only 16.8% of the total variance.

Materials and Methods

Dataset

E-MTAB-10604 is a publicly available RNA-seq dataset that describes RNAs from peripheral blood mononuclear cells (PBMCs) of people with DS, and of controls.
PBMCs separation was performed using Ficoll-Paque and the RNA was extracted using the TRIzol reagent.
Each whole blood sample (approx. 10 ml) was processed within 2 hours of collection and the RNA was stored at -80 °C.
Indexed libraries were prepared from 1 µg/ea purified RNA with TruSeq Stranded mRNA (Illumina) Library Prep Kit.
Libraries were sequenced using an Illumina NextSeq 550 Dx System (Illumina) in a 2x75 paired-end format.
The read-count for the genes of interest was normalized with the R package DESeq2 using the median of ratio.

Principal Component Analysis (PCA)

Done using the built-in R function prcomp.
Plot was obtained using the R package scatterplot3d.

Linear Discriminant Analysis (LDA)

Initial gene filtering done with the R package genefilter.
RNAs with no statistically significant mean difference between the two groups have been ignored (pvalue > 0.1).
LDA was done using the caret package. The accuracy of the fitted model was evaluated through 13-fold cross validation repeated 3 times.
The heatmap was built using the R packages pheatmap, grid and RColorBrewer.
The AUC-ROC curve plot was built using the pROC package.
Gene importance for sample classification was measured by conducting ROC curve analysis on each predictor.

DAVID

Used for initial functional enrichment analysis on the top 200 RNAs for LDA classification.
187 genes have been identified from the input.
The default annotation categories plus UP_TISSUE and CHROMOSOME have been used.

STRING

Used to build a protein-protein interaction network on the peptides encoded by the top 200 RNAs for LDA classification. 168 Proteins have been identified from the input.
Confidence was set to 0.5 and no more than 10 interactors have been added. The obtained network was exported to Cytoscape and the results of the DAVID analysis have been color-coded onto it.

EnrichNet

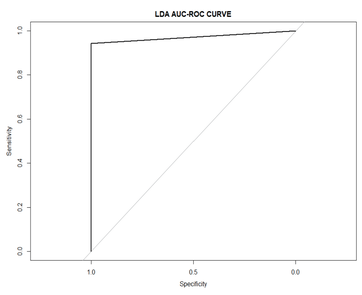
Used for tissue-specific network-based enrichment analysis on the top 200 RNAs, in conjunction with the Gene Ontology database. 150 genes have been identified from the input.
Results have been exported in csv format and filtered for pathways with cardiac tissue specific XD-scores above 2.5 (only AV-node and heart tissues have been considered).

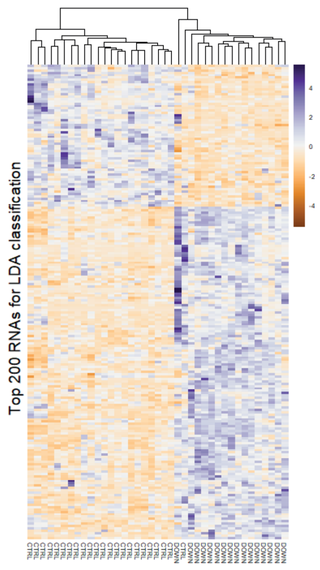
PathfindR

Used for analyzing the entire dataset, in conjunction with the KEGG database. 168 RNAs out of 26201 had low enough adjusted p-value and known interactions.

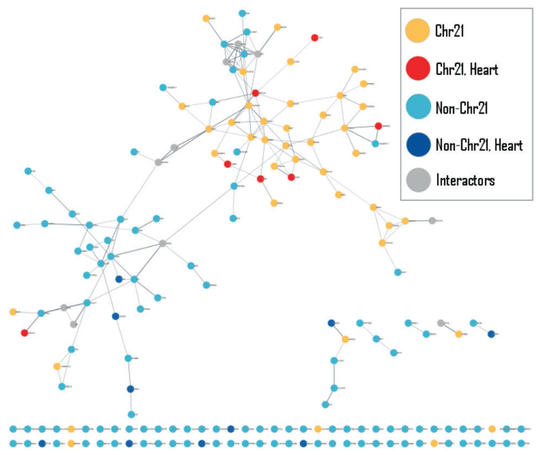
LDA

The fitted LDA classification model had an accuracy of 97%
and kappa of 94%.
The 200 most important RNAs for sample classification have been used in the following data analyses.





DAVID and STRING



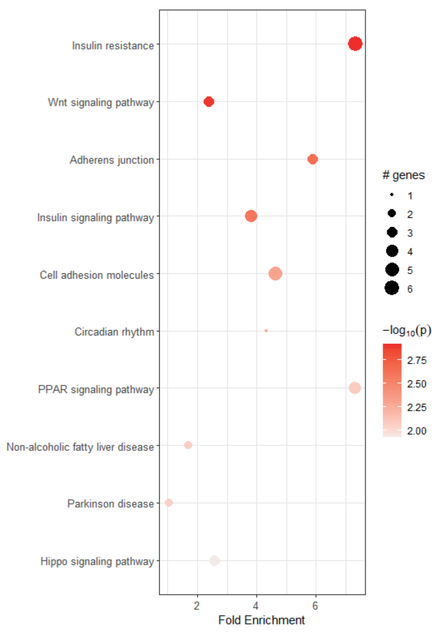
DAVID identified 17 genes tipically expressed in heart tissue (FDR = 0.02).
Only 7 of them are located on chromosome 21.

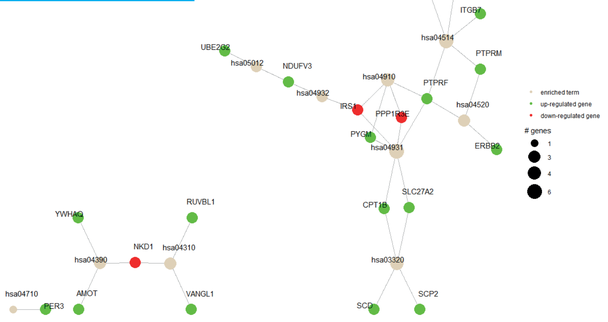
EnrichNet

Annotation	AV-node XD-score	Heart XD-score
heart trabecula formation	2.94	-0.07
calcium-dependent cell-cell adhesion	0.94	4.43
striated muscle cell differentiation	1.44	4.43
porphyrin-containing compound metabolic process	4.44	-0.07
positive regulation of NF-kappaB import into nucleus	2.94	-1.07
blood vessel morphogenesis	0.94	2.93
heterophilic cell-cell adhesion	0.63	4.43
positive regulation of muscle cell differentiation	0.58	8.93
positive regulation of MAPK cascade	0.26	8.93
regulation of protein localization	0.69	2.93
muscle cell differentiation	0.44	8.93
regulation of Rho protein signal transduction	0.24	4.43
Cell-cell adhesion	0.20	2.93

Enrichnet shows that both muscle cell differentiation pathways and cell-cell adhesion processes have been affected in cardiac tissues of people with DS.

PathfindR





The Wnt signaling pathway (hsa04310) has been proven to be important for early precardiac mesoderm differentiation.
Both Enrichnet and PathfindR show that cell adhesion processes and cell differentiation pathways have been affected.

Conclusions

In this study we attempted to understand the molecular causes of CHD in people with DS through a RNA-seq data analysis approach. We first fitted our data using LDA and we identified the 200 most important RNAs for correct sample classification. We analyzed them using STRING and DAVID, through which we discovered a set of 17 differentially expressed cardiac tissue specific genes and highlighted how they interact with one another.
We analyzed the top 200 RNAs also using EnrichNet together with the Gene Ontology database, and discovered that muscle cell differentiation and cell-cell adhesion processes have been affected in cardiac tissues of people with DS. We confirmed this hypothesis by doing another analysis using PathfindR in conjunction with the KEGG database, which showed similar results of affected cell adhesion processes and differentiation pathways.
We are aware that the importance of this study is limited by the use of data coming from PBMCs rather than from actual cardiac tissue. But nonetheless we managed to obtain potentially useful information regarding CHD in people with DS while avoiding important ethical issues.
Additional studies regarding cell adhesion and muscle cell differentiation during cardiac tissue development are needed in order to confirm our findings.

Acknowledgement: Genomix4Life S.r.l., for publishing the dataset used in this study.