Data in Brief

Data in brief: Transcriptome analysis of induced pluripotent stem cells from monozygotic twins discordant for trisomy 21

Youssef Hibaoui a,b, Iwona Grad a, Audrey Letourneau b, Federico A. Santoni b, Stylianos E. Antonarakis b,c,⁎ and Anis Feki a,d,⁎⁎

a Stem Cell Research Laboratory, Department of Obstetrics and Gynecology, Geneva University Hospitals, 30 bd de la Cluse, CH-1211 Geneva, Switzerland
b Department of Genetic Medicine and Development, University of Geneva Medical School and Geneva University Hospitals, 1 rue Michel-Servet, CH-1211 Geneva, Switzerland
c iGE3 Institute of Genetics and Genomics of Geneva, University of Geneva, Switzerland
d Department of Obstetrics and Gynecology, HFR Fribourg—Hôpital cantonal, Chemin des Pensionnats 2-6, Case postale 1708 Fribourg, Switzerland

A B S T R A C T

Down syndrome (DS, trisomy 21), is the most common viable chromosomal disorder, with an incidence of 1 in 800 live births. Its phenotypic characteristics include intellectual impairment and several other developmental abnormalities, for the majority of which the pathogenetic mechanisms remain unknown. In this "Data in Brief" paper, we sum up the whole genome analysis by mRNA sequencing of normal and DS induced pluripotent stem cells that was recently published by Hibaoui et al. in EMBO molecular medicine.

Direct link to deposited data

Deposited data can be found in the Gene Expression Omnibus (GEO) database: http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE52251

Experimental design, materials and methods

Cell culture and iPSC derivation

Primary fetal skin fibroblasts were isolated from monozygotic twins discordant for trisomy 21 [1]: Twin-N for the normal fibroblasts and Twin-DS for fibroblasts carrying trisomy 21 anomaly. These samples were obtained post mortem and the study was approved by the ethics committee of the Geneva University Hospital. Normal (Twin-N-iPSCs) and Down syndrome (Twin-DS-iPSCs) induced pluripotent stem cells (iPSCs) were established by transducing the parental fibroblasts (Twin-N and Twin-DS, respectively) with polyclonostic lentiviral vectors expressing OCT4, SOX2, KLF4 and c-MYC genes [2,3]. The generated iPSCs were cultured on human foreskin fibroblasts (ATCC, CCD 1112Sk, Manassas, USA) that were mitotically inactivated by irradiation at 35 Gy before seeding on a gelatin-coated 6-well plate at 3.5 × 10^5 cells/plate. iPSC colonies were maintained with daily changes in Knock-out
Dulbecco’s Minimal Essential Medium supplemented with 20% KO 
serum replacement, 1 mM L-glutamine, 100 μM non-essential 
amino acids, 100 μM 2-mercaptoethanol, 50 U/mL penicillin and 
50 mg/mL streptomycin (all from Gibco, Invitrogen, Basel, 
Switzerland) and 100 ng/mL human basic fibroblast growth factor 
(bFGF, Peprotech, London, UK). All iPSC lines were passaged by man-
ual dissection of cell clusters in the presence of 10 μM ROCK-inhibitor 
Y-27632 (Sigma-Aldrich, Buhs, Switzerland) [4].

RNA extraction and gene expression analysis by mRNA sequencing

Total RNA was extracted from the cell lines, using the QIAGEN 
(Hilden, Germany) RNeasy MiniKit according to the manufacturer’s protocol (Invitrogen). RNA integrity and quantity were assessed with an Agilent (Santa Clara, CA, USA) 2100 bioanalyzer, using RNA 6000 nanochips. mRNA-Seq libraries were prepared from 500 ng of total RNA using the Illumina TruSeq™ RNA Sample Preparation kit (Illumina RS-930-2001), following the manufacturer’s instructions. Libraries were sequenced on the Illumina HiSeq 2000 instrument to generate 100 bp paired-end reads. Those reads were mapped against the genome (hg 19) using the default param-
eters of the Burrows–Wheeler Aligner (BWA) [5]. An expression 
signal was detected in at least one sample for 20456 genes (com-
prising also non-coding RNA). For each gene, the level of expression 
was determined by calculating the exonic coverage (custom pipeline) 
and normalizing in Reads Per Kilobase per Million (RPKM). Twin-
DS-iPSCs and Twin-N-iPSCs samples were sequenced in three and 
four biological replicates, respectively, starting from different RNA 
preparations. Differential expression between Twin-N–iPSCs and 
Twin-DS–iPSCs was assessed using the default parameters of 
EdgeR (version 2.4.6) [6] and DESeq (version 1.6.1) [7] programs 
in R (version 2.14.0). Only the genes with more than 1 read per mil-
lion in at least 3 replicates were conserved for this analysis. 
Bonferroni correction was applied to adjust for multiple testing. 
A gene was considered differentially expressed if the Bonferroni-
corrected P-value was lower than 0.01 with both methods. Analysis 
of the functional annotations associated with the differentially 
expressed genes was performed using DAVID (Database for Annota-
tion, Visualization and Integrated Discovery) [8,9].

2. Discussion

Since the discovery that Down syndrome (DS) is caused by a tri-
somy of chromosome 21 (HSA21), a major challenge in DS research 
has been the recapitulation of DS phenotype and the identification 
of the mechanisms by which the extra copy of HSA21 leads to DS 
phenotypes [10]. In this respect, disease-specific iPSCs has opened 
up a new exciting avenue for the modeling and the correction of 
the phenotypes associated with several neurological diseases 
(reviewed in [11]). In our recent study, fetal skin fibroblasts isolat-
ed from monozygotic twins discordant for trisomy 21 (Twin-N for the normal fibroblasts 
and Twin-DS for the fibroblasts carrying the trisomy 21 anomaly) and used to estab-
lish normal and DS induced pluripotent stem cells (iPSCs) using OCT4, SOX2, KLF4 and 
c-myc genes; Twin-N-iPSCs for the normal iPSCs and Twin-DS-iPSCs for the iPSCs car-
rying trisomy 21 [4].

Then, whole transcriptome analysis was performed using 
mRNA-Sequencing and the normalized gene expression levels 
were compared between Twin-N–iPSCs and Twin-DS–iPSCs. Differ-
ential expression analysis was performed using the statistical R 
packages EdgeR and DESeq. Under these conditions, we obtained 
a list of 624 upregulated and 580 downregulated genes expressed 
in Twin-DS–iPSCs in comparison with Twin-N–iPSCs (listed in 
Table 1). The gene ontology (GO) analysis of the 624 genes upregu-
lated in Twin-DS–iPSCs using DAVID showed enrichment for func-
tions predominantly related to the regulation of RNA metabolic 
processes and regulation of transcription. The GO analysis of the 
580 genes downregulated in Twin-DS–iPSCs revealed significant enrich-
ment for genes involved in multiple developmental processes 
including embryonic development and morphogenesis, organ de-
velopment and morphogenesis, cellular adhesion and others [4]. 
Such alterations observed at the iPSC level illustrate the develop-
mental disease transcriptional signature of Down syndrome.

Considering our results and the recent studies recapitulating and 
rescuing the phenotypes associated with several neurological dis-
eases using pluripotent stem cell-based models (reviewed in [11]), we believe that a combination of the transcriptional profiling 
with the possibility to derive in principle all cell type relevant for 
the disease will contribute to a better understanding of the initia-
tion and the development of DS and ultimately to the design of 
new therapies.

Conflict of interest

The authors declare no competing interests.

Acknowledgments

This work was supported by grants from the Ernest Boninchi 
Fondation to A.F., Genico to A.F., the Swiss National Science Foundation 
(SNF-144082) and the European Research Council (ERC-249968) to 
S.E.A.
List of the differentially expressed genes between Twin-N-iPSCs and Twin-DS-iPSCs.

| DESeq (with a Bonferroni-corrected |
|-----------------------------------|
| S. Dahoun, et al., Monozygotic twins discordant for trisomy 21 and maternal 21q in- |
| K. Takahashi, et al., Induction of pluripotent stem cells from adult human fibroblasts |

Reference:
[1] S. Dahoun, et al., Monozygotic twins discordant for trisomy 21 and maternal 21q in- heritance: a complex series of events. Am. J. Med. Genet. 146A (2006) 2066–2070.

[2] K. Takahashi, et al., Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 153 (5) (2013) 861–872.
[6] M.D. Robinson, D.J. McCarthy, G.K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26 (1) (2010) 139–140.

[7] S. Anders, W. Huber, Differential expression analysis for sequence count data. Genome Biol. 11 (10) (2010) R106.

[8] D.W. Huang, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4 (1) (2008) 44–57.

[9] D.W. Huang, B.T. Sherman, R.A. Lempicki, Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 37 (1) (2009) 1–13.

[10] S.E. Antonarakis, et al., Chromosome 21 and Down syndrome: from genomics to pathophysiology. Nat. Rev. Genet. 5 (10) (2004) 725–738.

[11] Y. Hibaoui, A. Feki, Human pluripotent stem cells: applications and challenges in neurological diseases. Front. Physiol. 3 (2012) 267.