Supplementary Information Liu et al:

Supplementary Table S1: Gene expression studies in human DS cells and tissues

ArrayExpress accession ID	GEO accession ID	Tissue	Number of samples	Array design ID	Reference
E-GEOD-47014	GSE47014	iPS cell line	27	A-GEOD-15207	Jiang et al, 2013
E-MTAB-1238	NA	lymphoblasts	12	A-AFFY-44	Granese et al, 2013
E-GEOD-38931	GSE38931	iPSC from fb	8	A-GEOD-10558	Li et al, 2012
E-GEOD-10758	GSE10758	amniotic fluid	20	A-AGIL-9	Chou et al, 2008
E-GEOD-16176	GSE16176	amniotic fluid	14	A-AFFY-44	Slonim et al, 2009
E-GEOD-48051	GSE48051	amniotic fluid	38	A-AGIL-28	Volk et al, 2013
E-GEOD-6283	GSE6283	amniotic fluid	21	A-AFFY-44	Altug-Teber et al, 2007
E-GEOD-35665	GSE35665	PBC, PBMC	37	A-AFFY-143	Li et al, 2012
E-GEOD-35561	GSE35561	CD41+/235+ iPSC-derived progenitor	6	A-AFFY-141	Chou et al, 2012
E-GEOD-20910	GSE20910	DS-ALL-BM	49	A-AFFY-44	Loudin et al, 2011
E-GEOD-17459	GSE17459	DS-ALL-WBC	119	A-AFFY-33/A-AFFY-44	Hertzberg et al, 2010
E-GEOD-16677	GSE16677	DS-AMKL-BM, PBC	18	A-AFFY-44	Klusmann et al, 2010a
E-GEOD-4119	GSE4119	DS-AMKL-BM, PBMC	80	A-AFFY-33	Bourquin et al, 2006
E-MEXP-72	NA	DS-AMKL-, TAM-PBMC	20	A-AFFY-33	McElwaine et al, 2004
E-GEOD-32388	GSE32388	CMK cell line	4	A-AGIL-28	Xavier et al, 2011
E-GEOD-56332	GSE56332	*Note 1*	9	A-AGIL-28	Maroz et al, 2014
E-GEOD-1789	GSE1789	heart	15	A-AFFY-33	Conti et al, 2007
E-GEOD-1397	GSE1397	heart, astrocyte, cerebrum, cellebrum.	28	A-AFFY-33	Mao et al, 2005
E-GEOD-42956	GSE42956	fb and fb-derived iPSC.	54	A-GEOD-10558	Briggs et al, 2013 *
E-GEOD-42772	GSE42772	astrocyte	16	A-MEXP-931	Helguera et al, 2013
E-GEOD-5390	GSE5390	brain	15	A-AFFY-33	Lockstone et al, 2007 *
E-GEOD-48611	GSE48611	iPSC, neuron from iPSC	18	A-AFFY-44	Weick et al, 2013 *
E-GEOD-52249	GSE52249	iPSC	7	A-GEOD-11154	Hibaoui et al, 2014

E-GEOD-55504	GSE55504	iPSC	30	A-GEOD-11154	Letourneau et al, 2014
E-GEOD-42142	GSE42142	2nd tri placenta	9	A-GEOD-10999	Jin et al, 2013
EGAS00001000546		TAM, DS-AMKL BM, PB	66	A-GEOD-11154	Yoshida et al, 2013

Note 1: normal CD34+ HSPC lentivirally transduced to ectopically express Gata1s or Gata1.

* Datasets were used in differential expression analysis across T21 tissues (Figure 3).

Abbreviations:
fb - fibroblast; PBC – peripheral blood cell; PBMC – peripheral blood mononuclear cell; BM – bone marrow
WBC – white blood cell; TAM/TMD – transient abnormal myelopoiesis/transient myeloproliferative disorder ; DS-ALL – Down-syndrome acute lymphoblastic leukemia; DS-AMKL – Down-syndrome acute megakaryocyte leukemia (also known as ML-DS)

A keywords search performed on databases ArrayExpress, Gene Expression Omnibus (GEO) and European Genome-phenome Archive (EGA) was used to extract human GEP datasets from public repositories. Keywords included "trisomy 21", "Down syndrome", "Downs syndrome" and "iAMP". Gene expression profiles include those obtained by microarray or RNA-seq experiments. Differential expression analysis was performed using the *limma* algorithm. Studies with small sample numbers and leukemia were excluded. Therefore, datasets investigated were: adult DS brain, DS neurons differentiated from iPSCs and DS fibroblast. An FDR cutoff of 0.15 was used. These data were compared with unpublished *Affymetrix HuEX_1-0_st array* data from our lab of trisomy 21 human second trimester fetal liver CD34+ cells (n=4) compared with normal (disomic) second trimester fetal liver CD34+ cells (n=3). These data were analysed as follows:

Preprocessing: Expression measurement – antigenomic background correction, quantile normalization, and RMA summarization at natural scale using *xps v1.22*.

Filtering: Based on the design of the exon array probe sets on Affymetrix Exon HT array are grouped into "core", "extended" and "full". Several filtering steps were applied at the probe set and transcript levels prior to analysis:

- Restrict analysis to core probe sets;
- Remove probe sets with low detection above background call (DABG) with p value>=0.05;
- Remove cross-hybridisation probes;
- Remove transcripts that are not detected in either sample group.

Differential expression analysis was carried out using 2-sample student t-test integrated in genefilter and a linear model method, *limma* v3.18.13.