Supplemental information
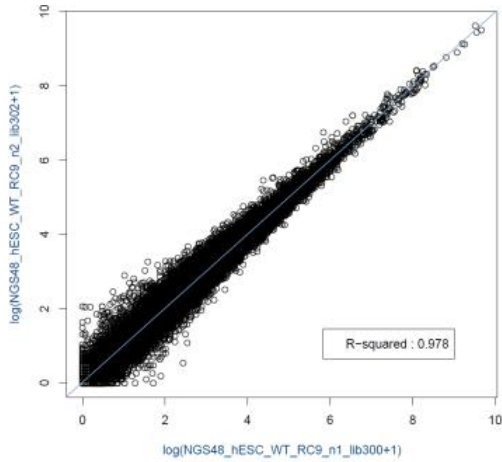
SISTEMA: A large and standardized

collection of transcriptome data sets for human

pluripotent stem cell research

Margot Jarrige, Hélène Polvèche, Alexandre Carteron, Stéphane Janczarski, Marc Peschanski, Didier Auboeuf, and Cécile Martinat
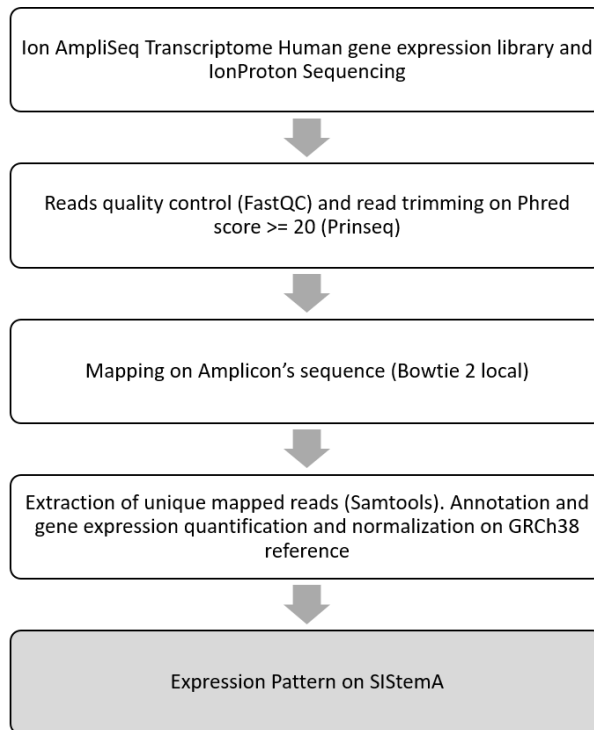
A

hESC RC9 (lib300) vs hESC RC9 (lib 302) :
R²=97,8

B

Frequency	# genes	% genes
1	20700	98.20
2	348	1.65
3	30	0.14
4	1	0.005
5	1	0.005

Frequency	# amplicons	% amplicons
1	20248	97.44
2	454	2.18
3	36	0.17
4	23	0.11
5	7	0.03
6	3	0.01
7	2	0.009
8	3	0.01
9	1	0.005
10	2	0.01
11	1	0.005
12	1	0.005

Figure S1: *Validation of Ampliseq Technical and informatic treatment workflow (related to Figure 1 and 2)*
a) Pearson correlation plot of the same RNA sample processing and sequencing independently: R²=97.8
b) Frequence of genes with 1 or more amplicons (21080 genes). Frequence of amplicons with 1 or more genes (20781 amplicons)

A

Ion AmpliSeq Transcriptome Human gene expression library and IonProton Sequencing

↓

Reads quality control (FastQC) and read trimming on Phred score >= 20 (Prinseq)

↓

Mapping on Amplicon's sequence (Bowtie 2 local)

↓

Extraction of unique mapped reads (Samtools). Annotation and gene expression quantification and normalization on GRCh38 reference

↓

Expression Pattern on SIStemA

B

sample	Starting Material	deep coverage	% alignement
mean sample	50ng	10156774 +/- 28%	97.71

Figure S2: _Sistema processing and analysis workflow (related to Figure 1 and 2)._
a) Bio-informatic pipeline used for SIStemA
b) Summary of sequencing parameters