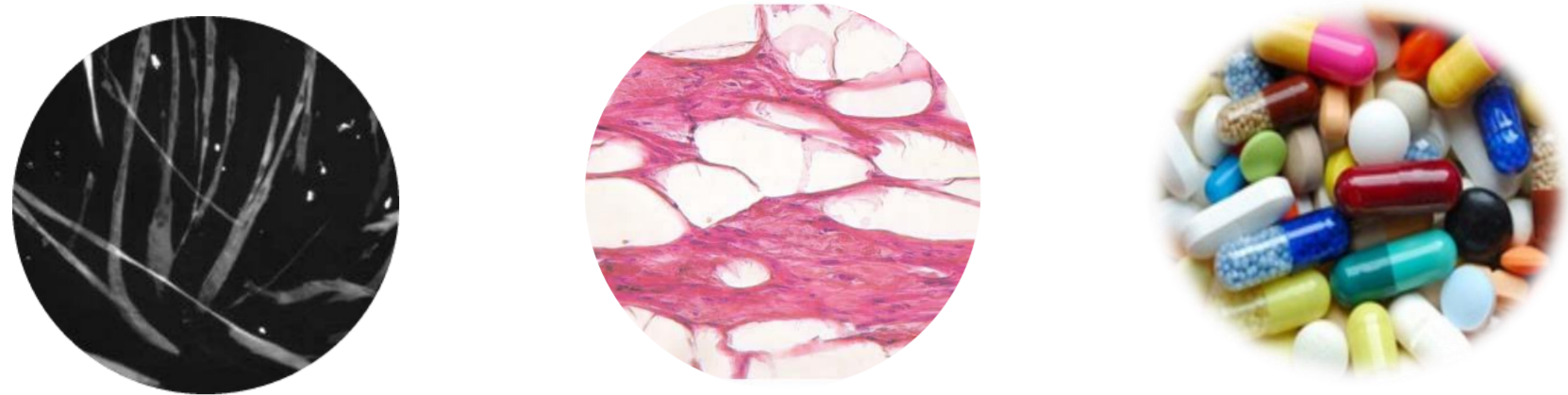# SIStemA : Gene expression database of human Stem Cell and their differentiated derivative.

Margot Jarrige & Hélène Polvèche   CECS / I-Stem, plateforme NGS, Corbeil-Essonnes, 91100, France

I-Stem is a French laboratory for research and development on rare monogenic diseases using human pluripotent stem cells (hESC/hIPSC). These cells, which are able to self-renewal and to differentiate into any cell type, have emerged as a powerful tool for disease modelling, drug screening and cell therapies.

Disease Modelling   **Cell therapies**   **Drug screening**

6 000 to 8 000 rare monogenic diseases

3 million people concerned in France & 30 million in Europe

1 in 20 people affected by a rare disease

-19

50% of patients are under 19 years

Totipotent stem cell

Pluripotent stem cells

Endoderm line   Mesoderm line   Ectoderm line

Multipotent stem cells

Lung   Pancreas   Heart muscle   Red blood cell   Skin   Neuron

Created with BioRender

For 5 years, our platform has been sequencing a large number of samples of cells derived from hESC/hIPSC at different differentiation time points, in contact with drugs or at different degrees of rare diseases. The expression of genes from these samples was initially planned in the frame of projects of the Institute's teams. This information is now compiled in the same database accessible through a web interface called SIStemA. This tool allows anyone at I-Stem to easily access this knowledge.
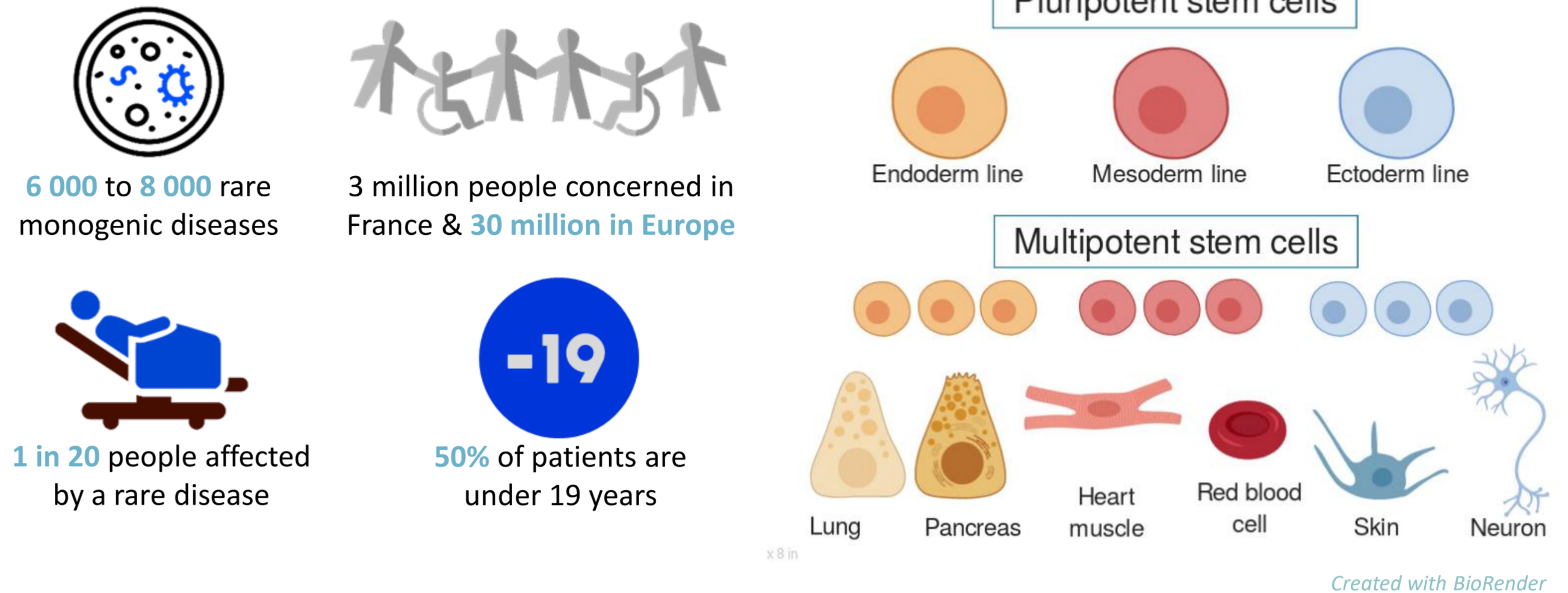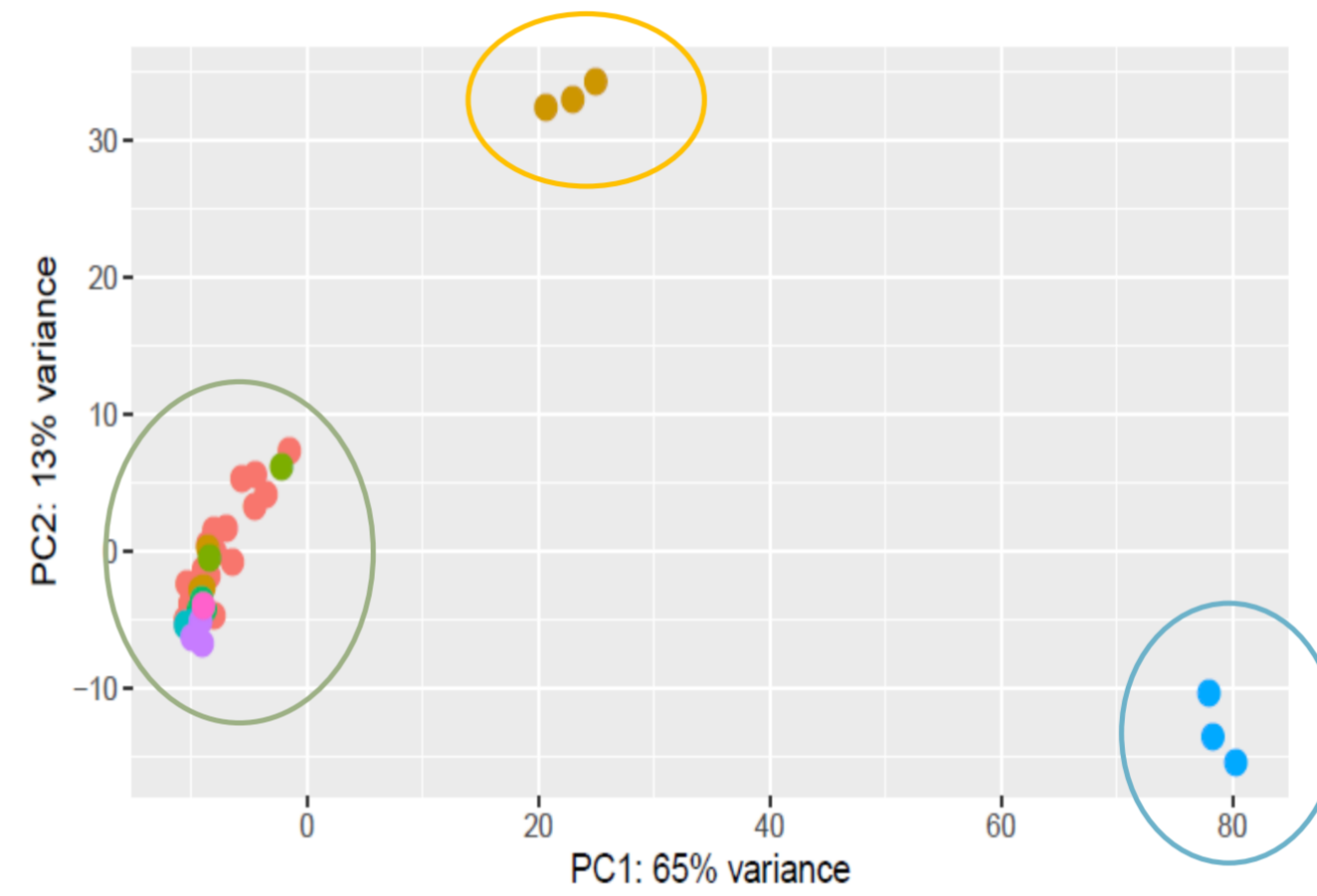
I - Ion AmpliSeq Transcriptome™ Dataset
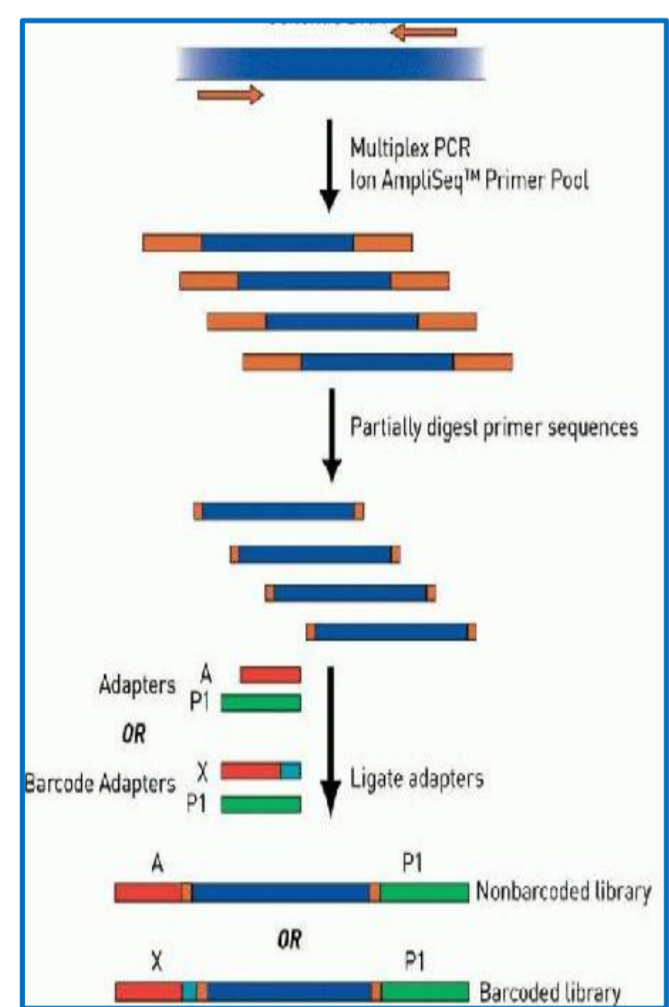
- neural cell
- MSC cell
- RPE
- reference
- hIPS/hESC
- melano like
- muscle like
- fibroblast

450 samples
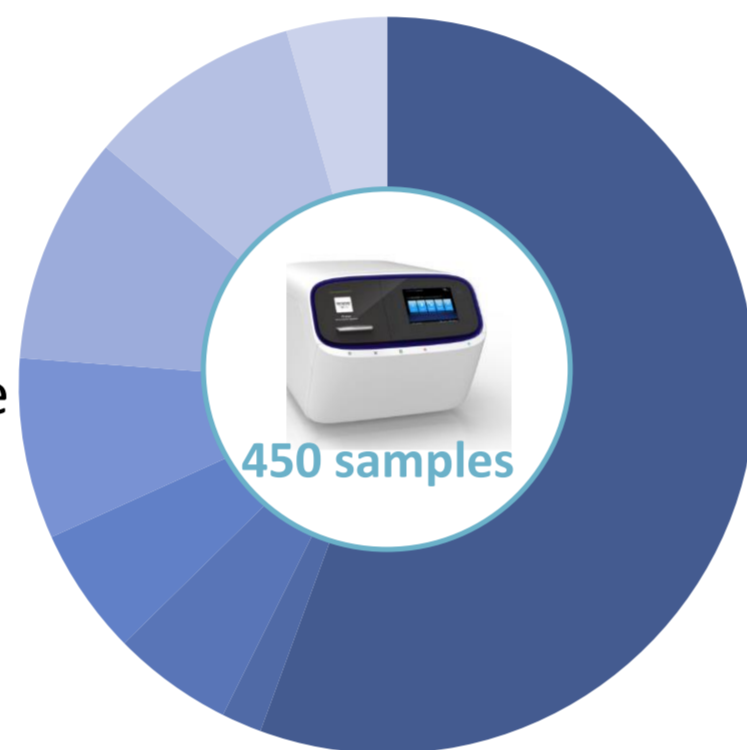
Ion AmpliSeq Transcriptome human gene expression kit workflow and dataset from I-Stem

- hESC-derived Melanocyte
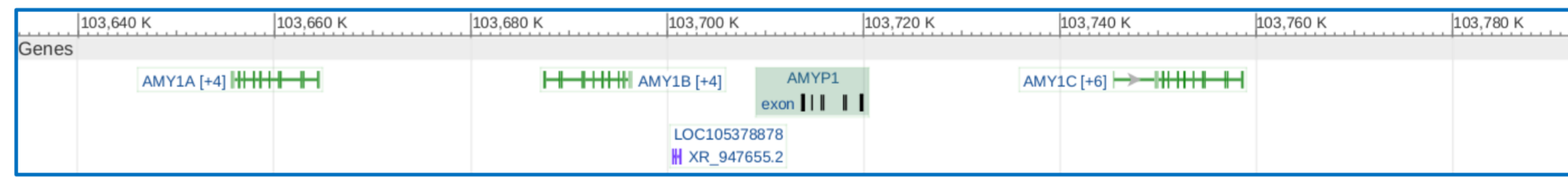- hESC-derived MN J10
- hESC/IPSC

PCA mapping of subsets of hESC/hIPSC samples and hESC/hIPSC-derived cell types to validate the analysis pipeline and the feasibility of the database.
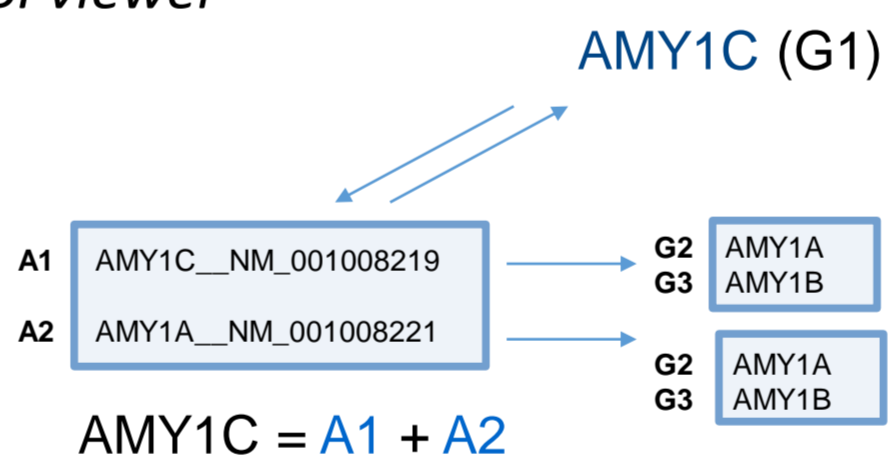
II – Bioinformatic processing

NCBI viewer

AMY1C (G1)

AMY1C = A1 + A2

Example : AMY1C – AMY1A – AMY1B paralogs

✓ Reference: Amplicons Sequences
✓ Alignment tool: Bowtie2 - Local
✓ Counting: Assigning Amplicon' reads to a gene.
✓ 21080 genes represented : 97% genes = 1 Amplicon

Verification that an amplicon corresponds to a single gene (Overlapping genes, Pseudogenes,...). Alignment of Amplicons sequences (primers + inserts ~150 bp) to GRCh38 RefSeq transcriptome using BlastN.
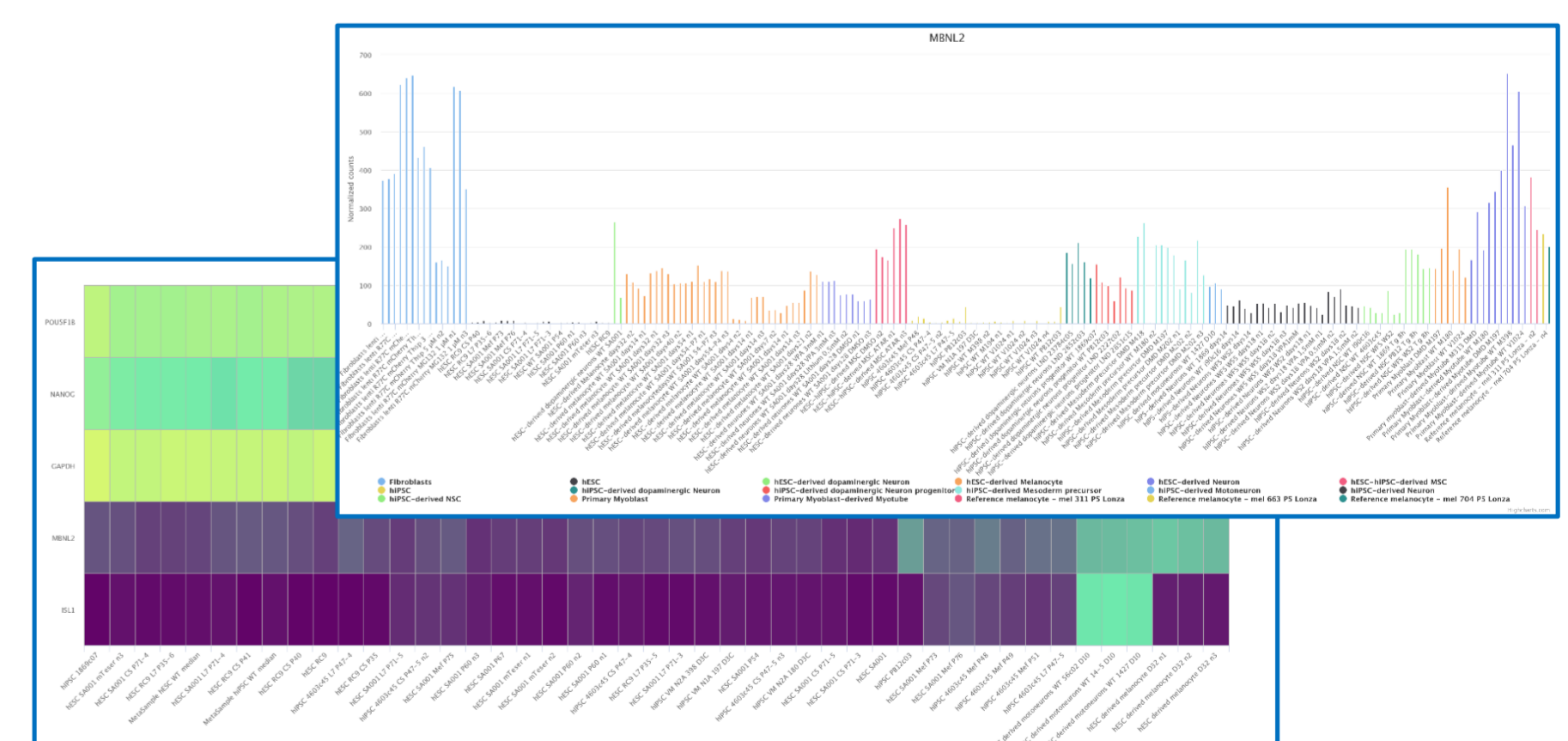
III – Database & Web interface
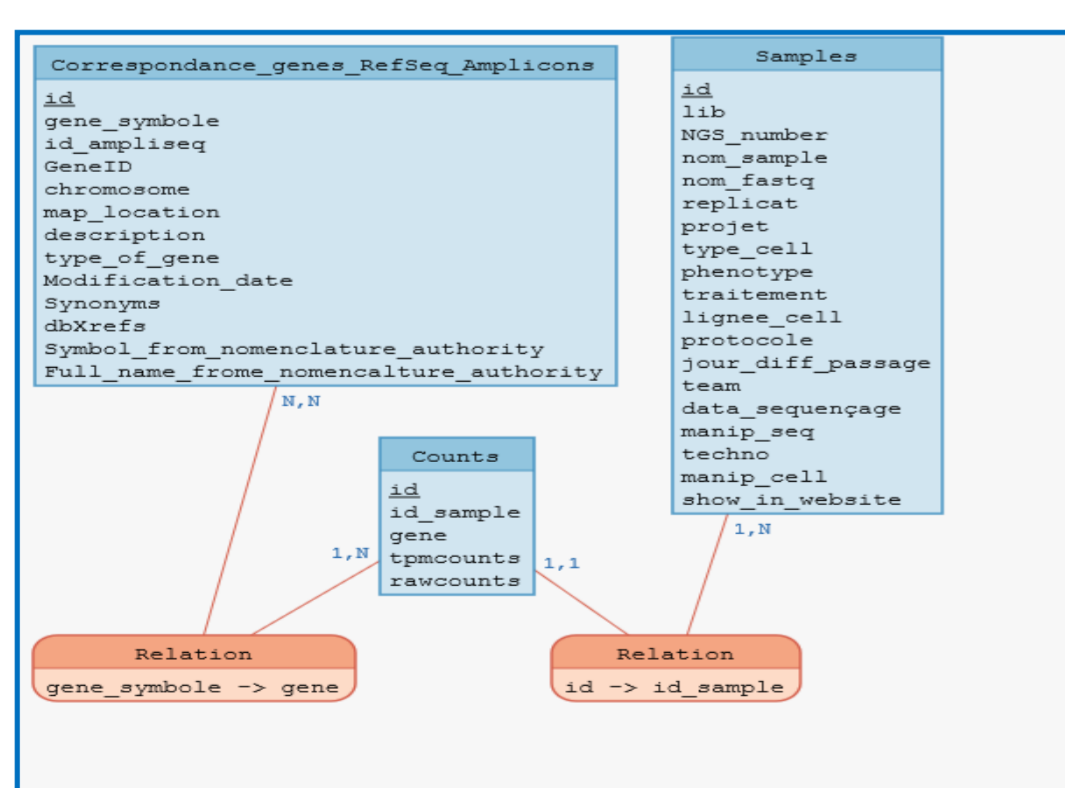
Ranking of the data. The database server is composed of three main tables (MySQL).

Users can filter the sample by cell type, genotype or project and access to all information corresponding to the sample. They can examine one or more genes at the same time.

Visualization of the data and direct download in tab-delimited format and graphics.

Programs & API : - Php, Java-script, jQuery, Ajax
- Hightcharts, NCBI Viewer, StringDB

Conclusions

This database will allow users to compare and study the expression of one or several genes in the catalog of I-Stem cells. Future database addition could include other technologies such as miRseq or available public data. Expanding the database will increase the knowledge and help the researcher to answer biological questions and develop new therapeutic strategies.

B. Langmead, S.L. Salzberg. Fast gapped-read alignment with Bowtie 2. *Nature Methods* volume 9, pages 357–359 (2012)
C. Camacho , et al . BLAST+: architecture and applications. *BMC Bioinformatics* 10:421 (2008)