

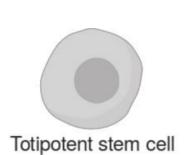




SIStemA: Gene expression database of human Stem Cell and their differentiated derivative.

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

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.





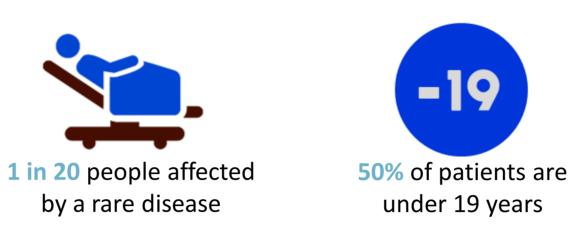


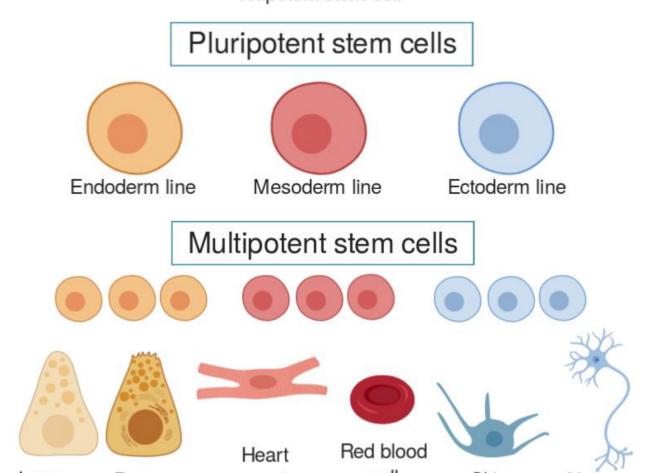


For 5 years, our platform has been sequencing a large number of samples of cells derived from hESC/hIPSC at different differentiation time points, in contact with drugs or at different degrees of rare diseases. The expression of genes from these samples was initially planned in the frame of projects 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.





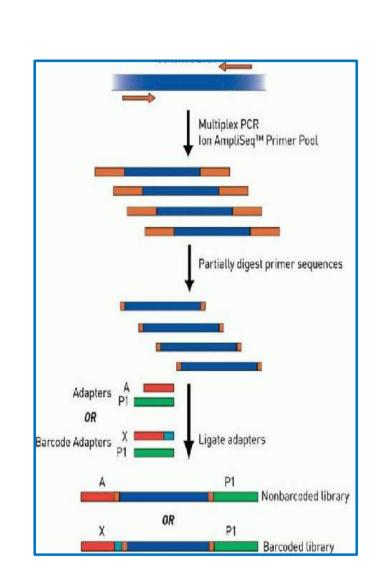


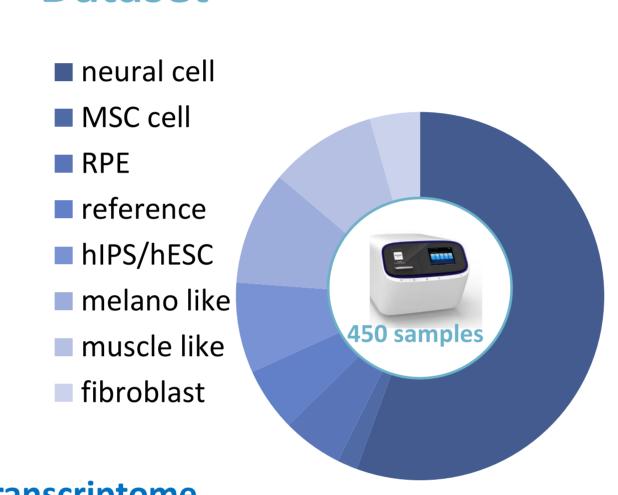


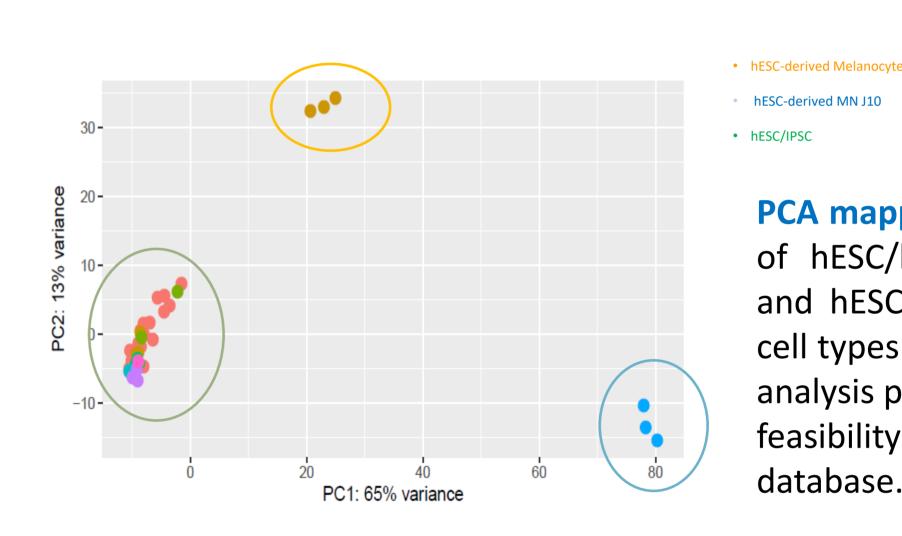
hESC-derived MN J10

Created with BioRender

I - Ion AmpliSeq Transcriptome™ Dataset







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.

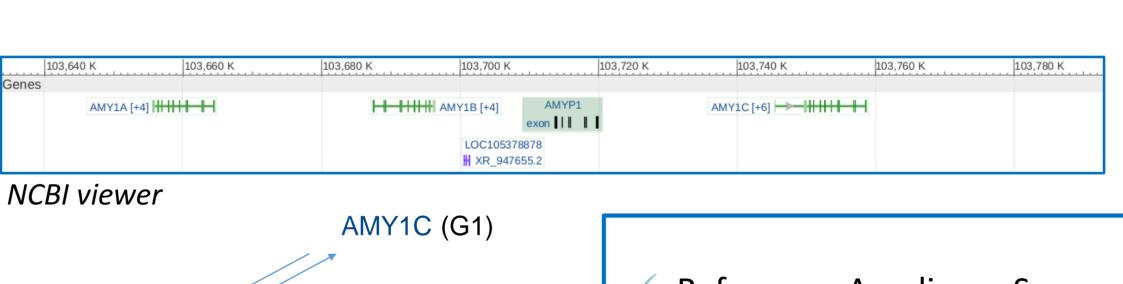
Ion AmpliSeq Transcriptome

human gene expression kit workflow and dataset from I-Stem

II - Bioinformatic processing

Query Sub		ct	percen of iden match	tical	jnment ength	mismatch	number gap opening	qst	art	qend	sstart	send	evalue	bitscore
GLMN_NM_053274 XM_017000138		8.1;GLMN	100.0	00	149	0	0	1	L	149	1003	1151	5,00E-073	276
GLMN_NM_053274 XM_01700013		7.1;GLMN	100.	00	149	0	0	1	1	149	1003	1151	5,00E-073	276
GLMN NM 052274 XM 01700014		4.1:GLMN	100.	00	112	0	0	1		112	624	726	5,00E-053	200
NANOG NM 024865 NM 001355281.1		1;NANOGP8	100.0	00	150	0	0	1		150	611	760	1,00E-073	278
NANOG NM 024865 NM 024865.3		NANOG	100.0	00	150	0	0	1		150	532	681	1,00E-073	278
NANOG NM 024865 NM 00129769		.1;NANOG	100.0	00	150	0	0	1		150	532	681	1,00E-073	278
NRG1 NM 013056 NM 00132220		6.1:NPG1	100.	00	150	0	0	1		150	1660	1919	1,00E 072	270
NRG1 NM 013956 NM 013957		4;NRG1	100.0	00	150	0	0	1		150	1035	1184	1,00E-073	278
NRG1 NM 0139	66 NM 00116000	4.2;NRG1	100.0	00	150 0		0	1		150	1035	1184	1,00E-073	278
gene_symbole	id_ampliseq	GeneID	chromosome	map_locatio	n	description		type_of_gene	Modificationdate		Synonyms		dbXrefs	
ABHD14A	ABHD14ANM_015407	25864	3	3p21.2	abhyd	rolase domain conta	aining 14A	protein-coding	20180805		DORZ1	HGNC:HGNC:24538 Ensembl:ENSG00000248487 Vega:OTTHUMG00000157818		00248487
ABHD14A-ACY1	ACY1_NM_001198895	100526760	3	3p21.2	AB	ABHD14A-ACY1 readthroug		protein-coding	20180329	e :			HGNC:HGNC:38856 Ensembl:ENSG00000114786 Vega:OTTHUMG00000166447	
ACY1	ACY1_NM_001198895	95	3	3p21.2		aminoacylase 1		protein-coding	20180805	ACY-1 ACY1D HEL-S-5			MIM:104620 HGNC:HGNC:177 Ensembl:ENSG00000243989 Vega:OTTHUMG00000157815	
NANOG	NANOGNM_024865	79923	12	12p13.31		Nanog homeobox		protein-coding	20180805	(L)			Ensembl:ENSG00000111704 Vega:OTTHUMG00000166019	
NANOGP8	NANOG_NM_024865	388112	15	15q14	Nan	Nanog homeobox retrogene P8		protein-coding	20180723	NANOGP1 PN8			HGNC:HGNC:23106 Ensembl:ENSG00000255192	
TTN	TTN NM 133378	7273	2	2q31.2	2 titin			protein-coding	20180805	CMD1G CMH9 CMPD4 EOMFQHMERF LGMD2JIMYLK5 SALMYITMD			Ensembl:ENSG00000155657	

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.



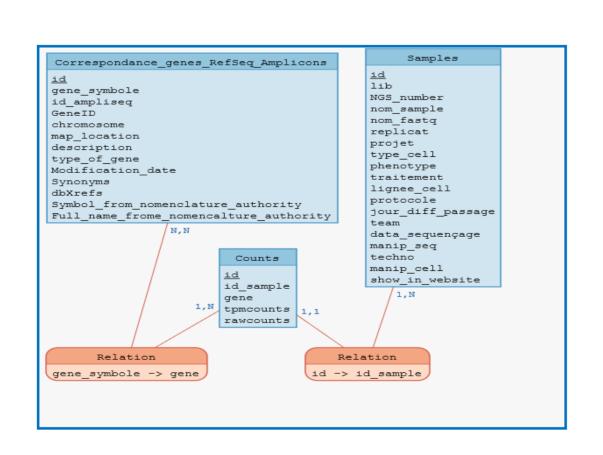
- Reference: Amplicons Sequences Alignment tool: Bowtie2 - Local
 - ✓ Counting: Assigning Amplicon' reads to a gene.
 - 21080 genes represented : 97% genes = 1 Amplicon

Example: AMY1C – AMY1A – AMY1B paralogs

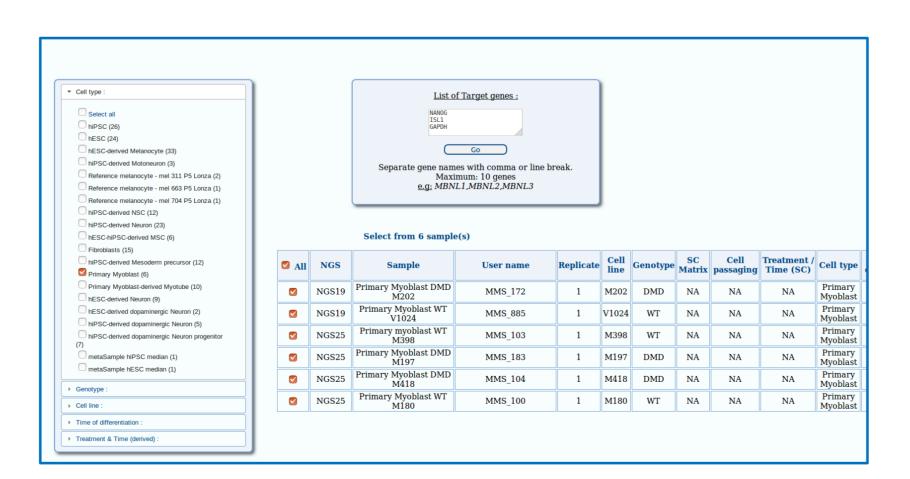
G2 AMY1A G3 AMY1B

G2 AMY1A G3 AMY1B

III - Database & Web interface



Ranking of the data. The database server is composed of three main tables (MySQL).

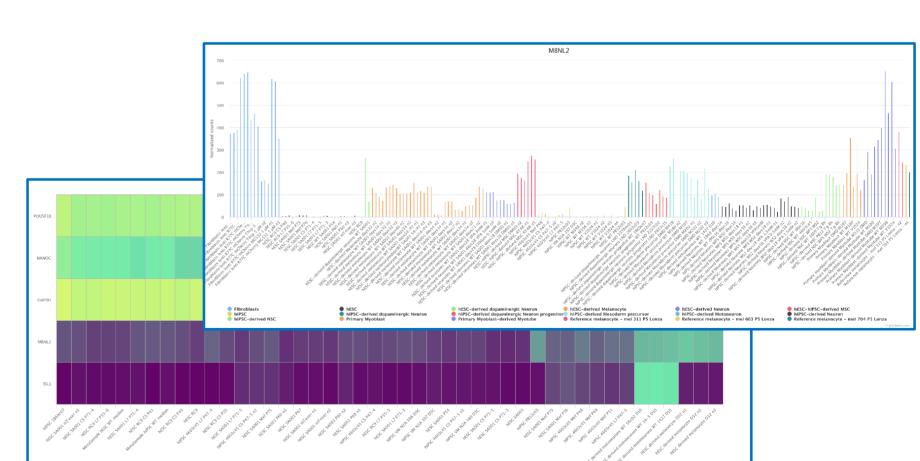


A1 AMY1C__NM_001008219

A2 AMY1A__NM_001008221

AMY1C = A1 + A2

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 tabdelimited 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.