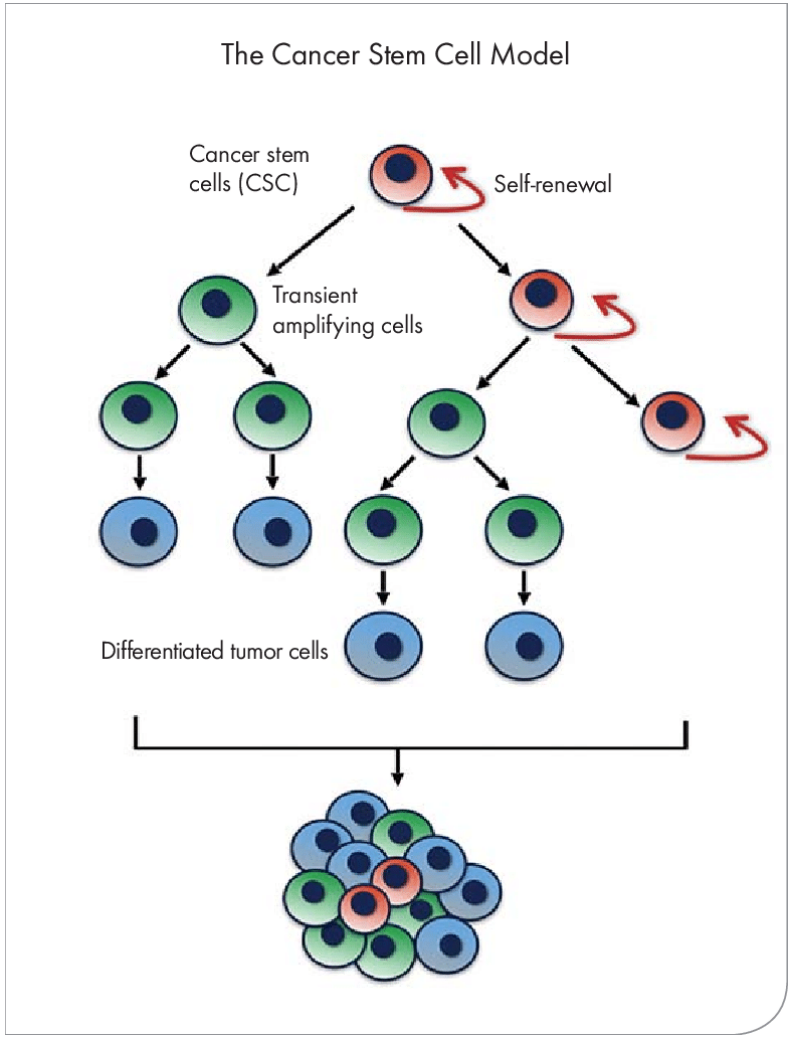
Utilizing Shared Big Data to Identify Liver Cancer Dedifferentiation Markers

Kirill Borziak and Joseph Finkelstein

Icahn School of Medicine at Mount Sinai

Cancer Stem Cells



- Growing evidence has implicated cancer stem cells for causing the therapeutic resistance, tumor recurrence, and metastasis.
- Cancer stem cells represent a key target for translational medicine in improving cancer treatment and outcomes.
- It is still not entirely understood how cancer stem cells are derived from adult fully differentiated cells with regards to expression of dedifferentiation factors.
- Here we present a meta analysis of adult liver and liver cancer single cell RNA-seq analyses.

Zhang, Kangyi. (2016). A Chemopreventive Nanodiamond Platform for Oral Cancer Treatment. Journal - California Dental Association. 44. 121-127.

Meta analysis

- The following studies were utilized for liver cancer cell profiles (GSE125449) and healthy liver cell profiles (GSE130473).
- The liver cancer study consists of 9946 single-cell RNA-seq profiles from 19 patients, totaling over 56 million reads and 4.2 billion base pairs.
- The adult liver study consists of 1467 single-cell RNA-seq profiles, totaling 283 million reads and 21 billion base pairs.
- We performed extensive pre-processing normalization, including filtering out low coverage samples (<1000 reads), low coverage genes (0 in all samples), and non-protein coding genes.
- To control for batch effects between samples, we utilized EdgeR to control for library size, calculate the common dispersion for all genes, and individual gene dispersion.
- Differential expression was assessed using the quasi-likelihood F-test after fitting the negative binomial GLM for each gene, using study type in the design matrix to further account for batch effect.
- False discovery rate was controlled using Bonferroni multiple testing correction.

Differential expression

- The final differential expression analysis consisted of 2434 single-cell samples across 18,263 protein-coding genes.
- We compared expression of two types of adult liver cells (CD235a-/EpCAM+/ASGPR1+ and CD235a-/EpCAM+) (444 single-cell samples) and the liver CSCs (1990 single-cell samples).
- We identified 519 genes that were differentially expressed between liver CSCs and adult liver cell types.
- 134 were significantly higher expressed in the liver CSCs.
- 385 protein coding genes were significantly higher expressed in the adult liver cell types.

Gene Ontology analysis

Gene categories enriched in liver CSCs

GO Term	GO ID	P-value
Structural constituent of ribosome	GO:0003735	8.9E-27
Translation initiation complex	GO:0070992	9.8E-24
rRNA processing	GO:0006364	1.1E-20
Mitochondrial respiratory chain complex I	GO:0005747	3.1E-5
NADH dehydrogenase (ubiquinone) activity	GO:0008137	3.2E-5
ATP biosynthetic process	GO:0006754	2.9E-3
Extracellular vesicle	GO:1903561	6.2E-12
ncRNA processing	GO:0034470	3.2E-14

Gene Ontology analysis

Gene categories enriched in adult liver cell types

GO Term	GO ID	P-value
organic acid metabolic process	GO:0006082	8.1E-18
carboxylic acid metabolic process	GO:0019752	1.2E-17
lipid metabolic process	GO:0006629	4.3E-7
drug metabolic process	GO:0006805	5.4E-5

Dedifferentiation factors

Factor	Gene ID	Fold change	P-value
Hepatocyte Nuclear Factor 4 Alpha	HNF4A	0.338X	4.43E-5
Transforming Growth Factor β1	TGFB1	4.74X	1.46E-104
Msh Homeobox 2	MSX-2	1.36X	1.99E-17

- HNF4A, the primary differentiation factor of liver cells, is significant downregulated in liver CSCs.
- TGFB1, driver of mesenchymal/stemness phenotype observed in hepatocellular carcinomas, is significantly upregulated in liver CSCs.
- Although Msx1 has been implicated in intestinal tumorigenesis, we report MSX-2 as a potential novel dedifferentiation factor involved in liver carcinoma development.

Thank you for your attention

Icahn School of Medicine at Mount Sinai