Comparative Analysis of Public Data Sets to Identify Stemness Markers That Differentiate Liver Cancer Stem Cells

Kirill BORZIAK[a,1] and Joseph FINKELSTEIN[a]

[a] *Icahn School of Medicine at Mount Sinai, New York, New York, USA*

Abstract. Cancer stem cells (CSCs) represent an important field in translational medicine for generating novel cancer treatments. To identify important stemness markers in liver CSCs that potentially explain their resistance to treatment, we analyzed 10865 single-cell RNA-seq samples across 42684 coding and non-coding genes. Our results show that CSCs have significantly increased expression of two Yamanaka factors (Oct4, 2.14X and SOX2, 1.13X) and three stemness factors (CD44, 3.25X; KRT7, 2.2X; SOX9, 1.71X), relative to liver progenitor cells. Our study demonstrates the potential power of harnessing shared big data for driving translational medicine for novel hypothesis generation.

Keywords. Liver cancer, single-cell RNA-seq, stemness factors, cancer stem cells

1. Introduction

Investigations of cancer stem cells (CSCs) generate voluminous amount of unique data, however the value of comparative analysis of cancer and healthy stem cells has not yet been systematically demonstrated. Growing evidence has implicated cancer stem cells for causing the therapeutic resistance, tumor recurrence, and metastasis [1]. Since these cells are responsible for perpetuating the tumor, even in post-therapy patients, they represent a key target for translational medicine in improving cancer treatment and outcomes. However, it is still not entirely understood how CSCs compare to naturally occurring progenitor cells with regards to expression of stemness factors. In order to address this question, we have undertaken a reanalysis of publically available single-cell RNA-seq datasets of primary liver cancer samples [2] and healthy liver samples [3].

2. Methods

Expression data was obtained from Ma et al (GSE125449) [2] and Segal et al (GSE130473) [3]. We performed stringent filtering and normalization to account for sample-specific effects. Differential expression was performed using edgeR [4], with Bonferroni multiple testing correction. Gene Ontology analysis was performed using DAVID 6.8 Functional Annotation Tool [5], with Benjamini multiple testing correction.

---

[1] Corresponding Author, Kirill Borziak, PhD, Icahn School of Medicine at Mount Sinai, 1 Gustave L. Levy Pl, New York, NY 10029, USA; E-mail: Kirill.Borziak@mountsinai.org.

3. Results

Using previously published single-cell RNA-seq data for liver cancer [2] and fetal and adult healthy liver [3], we reanalyzed 10865 samples across 42684 genes. Our aim was to identify expression changes responsible for pluripotency in malignant cells, by comparing liver CSCs to fetal progenitor cell types. We find that 248 genes are at least 5 fold overexpressed in liver CSCs (p-value <0.001). Liver CSCs had a significant enrichment of overexpressed genes functioning in SRP-dependent cotranslational protein targeting to membrane (GO:0006614, p-value =5.5E-14), structural constituent of ribosome (GO:0003735, p-value =4.0E-12) and translational initiation (GO:0006413, p-value =1.9E-11). Two Yamanaka factors were significantly overexpressed in liver CSCs relative to fetal progenitor cells: Oct4/POU5F1 (2.14 fold, p-value =8.28E-48) and SOX2 (1.13 fold, p-value =0.0392). Liver CSCs also had significantly higher expression of three additional stemness factors: CD44 (3.25 fold, p-value =4.24E-21), KRT7 (2.2 fold, p-value =1.27E-15) and SOX9 (1.71 fold, p-value =1.08E-7).

4. Discussion

With the diversity of cancer studies now being carried out with single-cell next generation sequencing, the abundance of data allows us to begin asking additional questions beyond the original scope of the researchers. We aimed to understand how the expression profiles of CSCs compare to progenitor cells in order to better understand the self-renewal and differentiation capabilities of CSCs. Our results highlight significant differences in the expression of stemness factors between liver CSCs and fetal progenitor cell types that potentially help explain the self-renewal and treatment resistance phenotypes. These results provide a novel insight into cancer biology that was made possible by harmoniously utilizing publically available datasets.

5. Conclusion

We show that specific stemness factors are overexpressed in CSCs, relative to other normal progenitor cell types. Our results have the potential to advance the knowledge of the important genes relevant to CSCs' increased treatment resistance ability, and to create potential new avenues for application of translational medicine to cancer treatment by decreasing their potential for self-renewal.

References

[1] Beck B, Blanpain C. Unravelling cancer stem cell potential. Nat Rev Cancer. 2013 Oct;13(10):727-38.
[2] Ma L, Hernandez MO, Zhao Y, et al. Tumor Cell Biodiversity Drives Microenvironmental Reprogramming in Liver Cancer. Cancer Cell. 2019 Oct 14;36(4):418-430.
[3] Segal JM, Kent D, Wesche DJ, et al. Single cell analysis of human foetal liver captures the transcriptional profile of hepatobiliary hybrid progenitors. Nat Commun. 2019 Jul 26;10(1):3350.
[4] Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010 Jan 1;26(1):139-40.
[5] Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009;4(1):44-57.