Stem Cell Hierarichies within Tumor Genetic Subclones

SIMULTANEOUS MUTATION DETECTION AND RNA-SEQ IN SINGLE CANCER CELLS

by

Jonathan Fisher

Harvard-MIT Division of Health Sciences and Technology

Submitted in Partial Fulfillment of
The Requirements for the M.D. Degree

February 2019

Area of Concentration:	Computational Biology/Neuro-oncology
Project Advisor:	Dr. Aviv Regev
Prior Degrees:	B.S. (Biochemistry)

I have reviewed this thesis. It represents work done by the author under my guidance/supervision.

Thesis Advisor

Table of Contents

Abstract

Intratumoral heterogeneity has been appreciated at the genetic and transcriptional levels, yet it remains unclear how they relate to one another. My thesis project applies a novel qPCR technique to combine single-cell RNA-seq with same-cell cDNA genotyping at single-nucleotide resolution. Using whole-exome sequencing in a human oligodendroglioma tumor, we identified a mutation in *CIC* that is present in only a subset of cancer cells. A previous analysis of single-cell RNA-seq in this tumor showed a substantial minority of cancer cells that display a stemness phenotype distinct from differentiated lineage phenotypes previously known in low-grade glioma. Results of parallel qPCR genotyping demonstrate that both *CIC*-wild-type and *CIC*-mutant cells are found across the spectrum of stem-differentiated phenotypes. This suggests that the stemness phenotype exists independently of mutation status and that it is instead more likely epigenetically driven.

Introduction

Targeting the subset of cells with tumor-regenerative potential is key to cancer therapy.(1) These cancer stem cells may follow a hierarchical model wherein cancer cells are confined to a differentiation hierarchy (Figure 1). Alternatively, the stochastic model claims that stemness can be acquired by plastic malignant cells. The existence of widespread genetic heterogeneity within tumors suggests that mutation status could be the plasticity factor driving cell differentiation state in the stochastic model. Genetically-driven differentiation state is an appealing model which ties together observed heterogeneity at the DNA and RNA levels and indicates a treatment strategy of targeting “stem cell” genetic subclones.

Hierarchical structures mimicking normal developmental hierarchies are seen in several cancers including high grade glioma.(2) However, the complex genetics of this disease present a

major challenge in assessing the relationship between genetic and hierarchical heterogeneities. In contrast, grade II oligodendroglioma is characterized by a relatively simpler genetic landscape, yet functional assessment of developmental hierarchy has failed due to lack of *in vitro* and *in vivo* models. Data from our group revealed a striking tumor hierarchy in oligodendroglioma reminiscent of a developmental hierarchy.(3) The authors performed single-cell RNA-seq on 3828 tumor cells from four primary oligodendrogliomas. Principle component analysis after normalization to data quality showed first three principle components corresponding to groups of genes known to be involved with lineage (PC1) and self-renewal/differentiation (PC2, PC3) (Figure 2). These gene lists were honed by comparison to previous functional studies and used to define lineage and stemness scores by averaging expression values of the genes on each respective stringent list. Plotting the stemness score of each cell against lineage score revealed a tumor structure typical of stemness and differentiation into two lineages.

How does the tumor's genetic heterogeneity overlay with the hierarchical structure evident in Figure 2? The aim of my thesis project is the application of a novel technique to acquire concurrent genetic information from cells of known hierarchical phenotype. At the time of this study, probing mutation status of single cells remains technically challenging due to low genetic input material. In contrast, single-cell RNA-seq is in part facilitated by the existence of many copies of a wide variety of transcripts in each cell. My project took advantage of the relative abundance of cellular transcripts by developing a methodology to simultaneously read-out mutation status from cDNA alongside RNA-seq in single cells. I hypothesized that the hierarchy observed in this patient's tumor is not accounted for by genetic heterogeneity but rather is caused by epigenetic differences between cancer cells.

Methods

Oligodendroglioma tissue from a Massachusetts General Hospital patient was collected at the time of resection. The fresh tissue was minced with a scalpel and enzymatically dissociated using a brain tumor dissociation kit. Live cells were collected by centrifugation and sorted via fluorescence-activated cell sorting to identify viable cancer cells. Whole exome sequencing of oligodendroglioma tumor cells was analyzed as follows. Output from Illumina software was processed by the Picard pipeline to yield BAM files containing aligned reads. Sample contamination was estimated using ContEst. Somatic single nucleotide variants were identified using MuTect and annotated using Oncotator. Variant clonality estimation via ABSOLUTE detected subclonal mutations within the tumor along with estimated mutation abundance.(4)

To detect mutations in single tumor cells, qPCR was performed using SuperSelective PCR primers which contain a loop-out sequence adjacent to the mutant base endowing them with high specificity to a single base change.(5) Experimental feasibility was first validated in single cells from oligodendroglioma cell lines of known mutation status. SuperSelective PCR primers were then designed to target a tumor suppressor gene identified by exome sequencing to be heterogeneously mutated in the patient's tumor. Specificity of these primers for the mutation of interest was validated on artificial templates synthesized to match the mutant and wild-type sequences for the gene of interest.

SupersSelective qPCR was subsequently performed on cDNA from 467 cells from the patient's tumor, all of which had previously been subjected to single-cell RNA-seq via Smartseq2 protocol.(6) cDNA was taken from frozen stocks of product from the pre-amplification reaction of the Smartseq2 protocol. 1 µl from each well of cDNA was used as template for a second round of Smartseq2 pre-amplification and bead purification in order to

increase overall signal downstream. qPCR was performed with the Fast Plus EvaGreen qPCR Master Mix Low Rox (Biotium 31014-1) according to the manufacturer's instructions with the sole modification of adding EDTA to a final reaction concentration of 1.6 mM to enhance primer specificity. Ct ≥ 33 were considered negative signal; Ct <33 was considered positive signal.

Results

Validation SuperSelective qPCR reactions were performed by myself and Christopher Rodman from our group. SuperSelective qPCR targeting the IDH1 wild-type sequence and the c.641A>G (p.R132H) mutation was validated in single cells from BT88 and BT54 oligodendroglioma cell lines which are IDH1 homozygous wild-type and heterozygous mutant/wild-type, respectfully (Figure 3). Reactions from a minority of cells from both cell lines showed negative signal. All positive signals were consistent with known cell line mutation status.

Genomic material from the patient's tumor was prepared for whole-exome sequencing by me, and sequencing results were analyzed by Keren Yizhak from our group. Although most subclonal mutations were of unknown functional relevance, Keren identified a c.4543C>T (p.R1515C) mutation in *CIC*—a known tumor suppressor in oligodendroglioma(7)—that was present in approximately 30% of cells from the patient's tumor. Single-cell RNA-seq reads detected the *CIC* mutation only in 7 cells from the patient's tumor. SuperSelective primers were therefore designed in collaboration with Kenneth Livak from Fluidigm to target the mutant region of *CIC* in order to identify mutation status in a greater number of cells. In positive control reactions using artificial *CIC* templates that I performed, qPCR signal was consistent between duplicate reactions. Reactions yielding positive signal showed comparable signal between duplicates. Reactions using wild-type *CIC* primers yielded positive signal only when combined

with wild-type artificial template, and reactions using mutant primers yielded signal only when combined with mutant artificial template.

In analyzing results of SuperSelective qPCR on single cell cDNA from the patient's tumor, we took advantage of the fact that *CIC* is located on chromosome 19q for which the tumor has loss of heterozygosity; thus, *CIC* is either homozygous wild-type or homozygous mutant in a given cancer cell. Of 467 single cells from the patient's tumor on which Christopher Rodman and I performed SuperSelective qPCR, 61 cells had signal in both replicates for either mutant or wild type *CIC* primers, but never for both. These were used to define 28 *CIC*-mutant cells and 27 *CIC*-wild-type cells for which single-cell RNA-seq data was available. Six cells that yielded qPCR signal did not pass single-cell RNA-seq quality control filters and were therefore excluded from analysis. When superimposed on a mapping of the stem-differentiation signature in each cell, both *CIC*-wild-type and *CIC*-mutant cells were found to span the differentiation hierarchy (Figure 4). We found no evidence that genotypic heterogeneity is associated with the cancer stem cell phenotype, and conversely, genetic and transcriptional-program heterogeneity appeared to be independent of one another.

Discussion

The results of my validation experiments demonstrate a highly specific system to detect mutations in single cells from cDNA that can be subjected to whole-transcriptome RNA-seq in parallel. The limited sensitivity of our qPCR approach as seen in Figure 3 is likely due to the susceptibility of single-cell RNA analysis to transcript dropout in early stages of processing (6). However, even in validation experiments using artificial template, although all positive signal results were consistent with template status, a substantial fraction of reactions nonetheless

yielded negative signal, suggesting a decreased reaction sensitivity inherent to the loop-out design of our qPCR primers.

The results of our qPCR performed alongside RNA-seq in single cells from the patient's tumor confirm the coexistence of heterogeneity at the levels of both genetic sequence and gene expression. Furthermore, this was the first study to demonstrate a subclonal *CIC* mutation in oligodendroglioma as identified by whole-exome sequencing and confirmed in single-cell qPCR analyses. *CIC* is a tumor suppressor with inactivating mutations in up to 70% of oligodendrogliomas.(7) Oligodendroglioma mutations in *EGFR*, *TP53*, *ATRX*, and *CDKN2A* are mutually exclusive with those in *CIC*, suggesting a shared tumorigenesis pathway.(8,9) Mammalian studies have shown that *CIC* functions as a transcriptional repressor, with important targets including the PEA3 gene subfamily.(10,11) *CIC* is activated by receptor tyrosine kinase signaling, and its repressor activity in mice is partially dependent on binding to Ataxin.(12-15) A recent study using CRISPR-Cas9 knockout of Cic in mice showed that Cic binds to MAPK effector genes and represses their transcription.(16)

In light of the above evidence, *CIC* inactivation is a very likely cancer driver in a subset of cells within our patient's tumor. In general, although our method only allowed us to directly observe mutations in a single gene, tumor subclonal mutations tend to co-segregate during the process of tumor evolution. Additionally, since between patients *CIC* mutations are mutually exclusive with other mutations, this phenomenon could occur within an individual tumor as well. Thus, *CIC* wild-type cells that we observed may harbor one or several cancer driver mutations in other genes that are absent in the *CIC*-mutant cells.

Since both *CIC*-mutant and wild-type cancer cells were found to span the complete differentiation hierarchy as inferred from single-cell RNA-seq, it is likely that any co-segregating

mutations also span the hierarchy. This line of reasoning supports a model of differentiation hierarchy within which subclonal mutations accrue, as opposed to a hierarchical phenotype driven by genetic events. Moreover, the hierarchical states are likely epigenetically driven rather than genetically driven. We note that without a comprehensive phylogenetic reconstruction, we cannot categorically rule out a genetic influence on differentiation hierarchy.

This study possessed multiple limitations. As discussed above, the existence of both mutant and wild-type *CIC* cells across the differentiation hierarchy suggest that the hierarchy is independent of subclonal genetic events, but validation in additional genes would strengthen this conclusion. Furthermore, our study was limited to cells from a single tumor, and the results will need to be validated in additional glioma samples. Because our observed differentiation hierarchy is unique to glioma, it is difficult to predict whether this phenomenon is generalizable to other organs or tumor types. Another caveat of my project is that my results are observational and have not been functionally validated through experimentation. Because grade II oligodendroglioma cells do not grow in xenotransplantation, we relied on inference from mutation and sequence analysis in a human tumor.

Conclusions

Our study has demonstrated that RNA-seq combined with genetic readout from single cells allows unbiased analysis of genetic contribution to cellular expression programs. We have shown that a known driver mutation in tumor suppressor *CIC* exists subclonally in a human oligodendroglioma tumor. Both *CIC*-mutant and wild-type cancer cells span the tumor differentiation hierarchy suggesting that the tumor hierarchy is driven by epigenetic rather than genetic events.

Summary

The nature of cancer stem cells remains unclear. In the hierarchical model, cancer cells are confined to a differentiation hierarchy. An alternative stochastic model proposes acquisition of stemness by plastic malignant cells. Evidence in favor of the stochastic model is the existence of widespread genetic heterogeneity within tumors; acquired genetic alterations may thus represent the plasticity factor necessary for cancer stemness. In collaboration with others in our group and at Fluidigm, I developed a platform to combine single-cell RNA-seq with same-cell cDNA qPCR genotyping in order to test these stem cell models. Our technique utilized qPCR primers containing a novel loop structure, endowing it biochemical specificity at single-nucleotide resolution. This system distinguishes between wild-type sequence and single-nucleotide mutations in PCR-amplified cDNA from single cells that in parallel is subjected to whole-transcriptome RNA-seq. In this system, the expression profiling from single-cell RNA-seq illuminated the cancer stem cell phenotype while qPCR enabled simultaneous genotyping in those same cells.

Using whole-exome sequencing in a human oligodendroglioma tumor, our group identified a mutation in tumor suppressor *CIC* that is present in only a subset of cancer cells. Inference from previous single-cell RNA-seq in glioma identified differentiation along two specialized glial programs as well as an undifferentiated subpopulation expressing neural stem cell markers. Simultaneous single-cell RNA-seq with qPCR genotyping confirmed genetic heterogeneity of *CIC* at cellular resolution. When superimposed on a mapping of the stem-differentiation signature in each cell, both *CIC*-wild-type and *CIC*-mutant cells were found to span the differentiation hierarchy. We found no evidence that genotypic heterogeneity is associated with the cancer stem cell phenotype, and conversely, genetic and transcriptional-program

heterogeneity appear to be independent of one another. These results therefore favor a hierarchical interpretation of oligodendroglioma in which a relatively fixed subpopulation of cancer stem cells are the main source of malignant proliferation. The resulting tumor differentiation hierarchy appears to exist independently of subclonal mutations, and it instead serves as a scaffold upon which the tumor acquires subclonal mutations.

References

1. Kreso A, Dick JE. Evolution of the cancer stem cell model. *Cell Stem Cell*. 2014;14(3):275-91.
2. Lathia JD, Mack SC, Mulkearns-hubert EE, Valentim CL, Rich JN. Cancer stem cells in glioblastoma. *Genes Dev.* 2015;29(12):1203-17.
3. Tirosh I, Venteicher AS, Hebert C, et al. Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature*. 2016;539(7628):309-313.
4. Carter SL, Cibulskis K, Helman E, et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol*. 2012;30(5):413-21.
5. Vargas DY, Kramer FR, Tyagi S, Marras SA. Multiplex Real-Time PCR Assays that Measure the Abundance of Extremely Rare Mutations Associated with Cancer. *PLoS ONE*. 2016;11(5):e0156546.
6. Picelli S, Faridani OR, Björklund AK, Winberg G, Sagasser S, Sandberg R. Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc.* 2014;9(1):171-81.
7. Bettegowda, C. *et al.* Mutations in *CIC* and *FUBP1* contribute to human oligodendroglioma. *Science* 333, 1453–1455 (2011).
8. Brat DJ, Verhaak RG, Aldape KD, et al. Comprehensive, Integrative Genomic Analysis of Diffuse Lower-Grade Gliomas. *N Engl J Med*. 2015;372(26):2481-98.
9. Suzuki H, Aoki K, Chiba K, et al. Mutational landscape and clonal architecture in grade II and III gliomas. *Nat Genet*. 2015;47(5):458-68.
10. Kawamura-saito M, Yamazaki Y, Kaneko K, et al. Fusion between CIC and DUX4 up-regulates PEA3 family genes in Ewing-like sarcomas with t(4;19)(q35;q13) translocation. *Hum Mol Genet.* 2006;15(13):2125-37.

11. Okimoto RA, Breitenbuecher F, Olivas VR, et al. Inactivation of Capicua drives cancer metastasis. *Nat Genet* 2017;49: 87–96.
12. Dissanayake K, Toth R, Blakey J, et al. ERK/p90(RSK)/14-3-3 signalling has an impact on expression of PEA3 Ets transcription factors via the transcriptional repressor capicúa. *Biochem J*. 2011;433(3):515-25.
13. Lam YC, Bowman AB, Jafar-nejad P, et al. ATAXIN-1 interacts with the repressor Capicua in its native complex to cause SCA1 neuropathology. *Cell*. 2006;127(7):1335-47.
14. Fryer JD, Yu P, Kang H, et al. Exercise and genetic rescue of SCA1 via the transcriptional repressor Capicua. *Science*. 2011;334(6056):690-3.
15. Lee Y, Fryer JD, Kang H, et al. ATXN1 protein family and CIC regulate extracellular matrix remodeling and lung alveolarization. *Dev Cell*. 2011;21:746–57.
16. Weissmann S, Cloos PA, Sidoli S, et al. The Tumor Suppressor CIC Directly Regulates MAPK Pathway Genes via Histone Deacetylation. *Cancer Res*. 2018;78(15):4114-4125.

Figures

Figure 1: Schematic of traditional genetic versus hierarchical models of cancer stem cells. In the genetic model, cancer cells become stem-like via acquisition of new mutations. In the hierarchical model, epigenetically defined cancer stem cells give rise to differentiated progeny. Subclonal mutations in cancer stem cells are passed on to cancer progenitor cells.

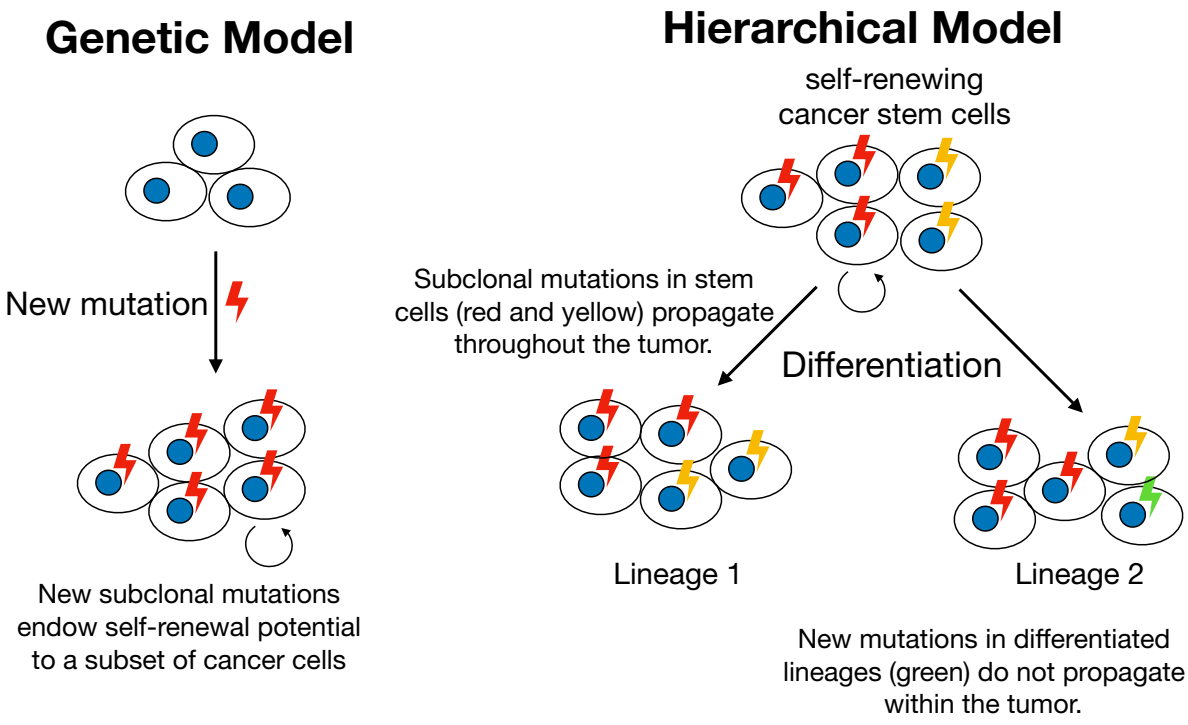

Figure 2: Principle Component Analysis of single-cell RNA-seq data from the patient's oligodendroglioma where each dot represents the scores in principle component 1 versus the sum of principle components 2 and 3 for the expression profile of a single cell. Genes whose expression levels correlated with PC1 scores included many glial lineage-related genes such as *ALDOC*, *APOE*, *ID1*, *ID4*, *SOX9*, *OLIG1*, *OLIG2*, *OMG*, and *SOX8*. Genes whose expression levels correlated with PC2 and PC3 scores included neurodevelopmental stemness factors *SOX2*, *SOX4*, and *SOX11*.

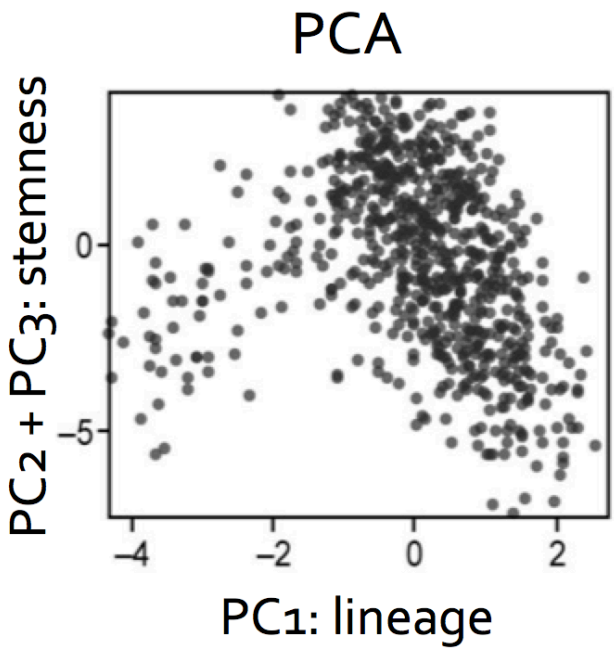

Figure 3: Results of proof-of-principle qPCR with SuperSelective primers designed against *IDH1* mutation present (heterozygous) in BT54 and absent in BT88. Each dot represents qPCR result from a single cell. Amplification Ct cutoff of 33 was chosen to distinguish between negative or positive signal. SuperSelective qPCR demonstrated our system's ability to selectively detect single base changes in cDNA from single cells.

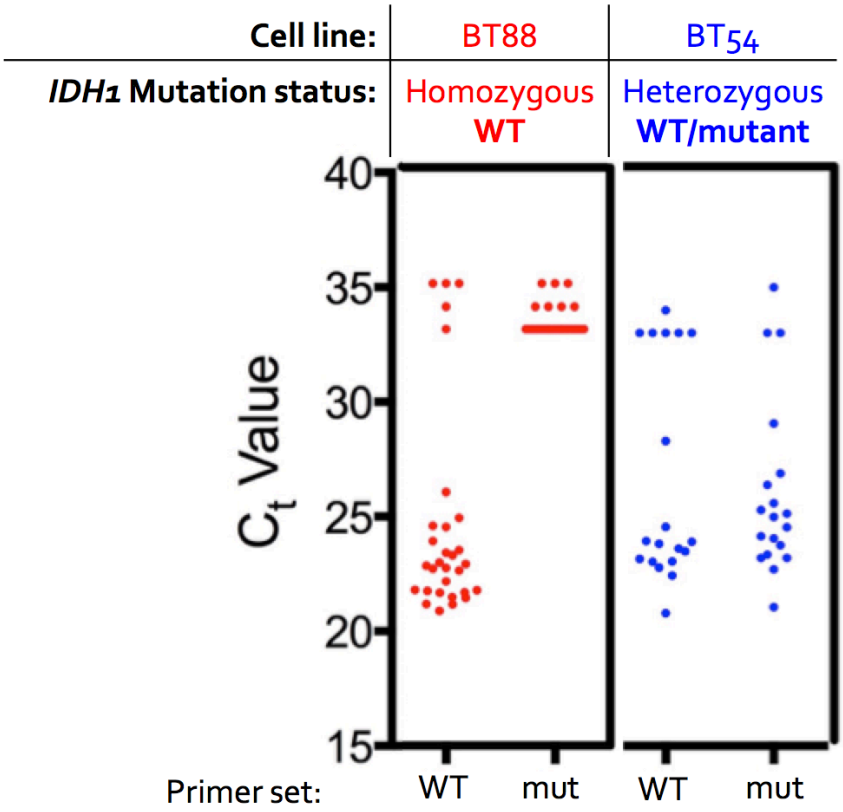

Figure 4: 28 *CIC*-mutant (orange) and 27 wild-type (green) cells among 467 single cells (black) from the patient's tumor plotted along a developmental hierarchy. Stemness score and lineage score were previously defined according to combined expression profiles using genes identified from principle component analysis(3). AC = astrocyte-like, OC = oligodendrocyte-like.