Unraveling the Transcriptional Heterogeneity in Human Colon Cancer: Potential New Insights regarding Pathobiology and Treatment

Mats Nilsson

In a recent hallmark report in *Nature Biotechnology*, Dalerba and colleagues from Stephen Quake’s and Michael Clarke’s laboratories at Stanford demonstrated the power of single-cell expression analysis for the elucidation of human tissue function and pathology (1). They made a detailed molecular dissection of healthy and colon cancer tissue to address fundamental questions about tumor heterogeneity and clonality. This pioneering and impressive work shows that much is to be gained by resolving the cellular heterogeneity of tissues. This more detailed characterization of normal and diseased tissue can provide valuable insights into cancer pathobiology that can be used for improving diagnosis and stratification of patients, and it has the potential to lead to new insights regarding treatment.

Dalerba and colleagues first made a single-cell suspension of healthy colon tissue. They used fluorescence-activated cell sorting with established differentiation markers to isolate and separate well-differentiated epithelial cells at the top of colon crypts from the more immature stem cell–like cells at the bottom of colon crypts. They then performed single-cell quantitative PCR analyses of 57 genes in 336 cells of each isolated cell population from each tissue sample to generate single-cell expression profiles. They used a microfluidics Fluidigm PCR instrument to generate the profiles, which were used to cluster individual cells into groups of cell types according to their molecular profiles by means of sophisticated computational methods and prior knowledge about cell-specific expression. Cells from the tops of colon crypts showed profiles compatible with mature enterocytes, whereas cell groups from the bottoms of crypts showed profiles compatible with immature cells, goblet cells, and stem cells/progenitor cells.

These investigators then used the markers for immature epithelial cells to isolate single cells from a colonic adenoma and performed high-throughput single-cell gene expression profiling. They found that they could identify cells from 2 of the cell types they had found in healthy colon, namely the goblet and progenitor cells, whereas the other molecularly defined cell types were greatly depleted. As a next step, they created a human colon cancer mouse xenograft model derived from a single cancer cell and performed single-cell expression profiling of the resulting tumor. They observed largely the same population of molecularly defined cell types as they found in the tumor from which the progenitor cell was isolated. This result indicates that much of the transcriptional heterogeneity found in tumors is caused by a differentiation process very similar to the one seen in healthy tissue. It has not been possible to study the origin of this heterogeneity by analyzing RNA prepared in bulk from tumors, because it is not possible to establish whether the differences in the RNA composition between one tissue and another are due to variations in cell type composition or in the expression profile of these cells.

From the knowledge of the expression profile of cell types constituting colon tumors, Dalerba and colleagues were able to estimate the cell composition of 1568 tumors on the basis of their bulk expression profiles measured with traditional microarrays, and they grouped the tumors into 3 classes (groups 1–3). Once they had defined these groups, they were able to pinpoint 2 genes that would accurately define them: \( KRT20^{+}/top-crypt^{\text{high}} \) (group 1), \( KRT20^{+}/top-crypt^{\text{low}} \) (group 2), and \( KRT20^{-}/top-crypt^{\text{low}} \) (group 3), where \( KRT20 \) is the keratin 20 gene and \( top-crypt \) is any one of the following genes: \( CA1 \) (carbonic anhydrase I), \( MS4A12 \) (membrane-spanning 4-domains, subfamily A, member 12), \( CD177 \) (CD177 molecule), and \( SLC26A3 \) (solute carrier...
family 26, member 3)]. When a set of 299 patients at different clinical stages and with known clinical outcomes was analyzed, the groups were shown to reflect tumors from more differentiated cells (top-crypt\textsuperscript{high}) to tumors from more immature cells (KRT20'/top-crypt\textsuperscript{low}). The tumors with immature cell expression profiles correlated with high-grade tumors and a poor prognosis. This 2-gene predictor even performed better than the pathology grading that is routinely used for the selection of therapy in colon cancer.

This pioneering study may pave the way for a new research approach, in which new technology is used to link detailed knowledge gained from cell biology studies using cell lines—which has been the dominant approach in biomedicine for decades—with traditional bulk tissue-level expression analysis and morphologic information. Such technology may allow us to gain a higher level of understanding of both normal and disease physiology in humans. This study also shows that such understanding can be of immediate utility, in this case for improved cancer prognostics.

Author Contributions: All authors confirmed they have contributed to the intellectual content of this paper and have met the following 3 requirements: (a) significant contributions to the conception and design, acquisition of data, or analysis and interpretation of data; (b) drafting or revising the article for intellectual content; and (c) final approval of the published article.

Authors’ Disclosures or Potential Conflicts of Interest: No authors declared any potential conflicts of interest.

Reference