Hiding in Plain Sight: How Stromal Signatures May Inform Cancer Prognosis

Researchers use RNA-Seq and gene expression arrays to identify stromal contributions to the colorectal cancer transcriptome.

Introduction

Colorectal cancer (CRC) is the fourth leading cause of cancer-related deaths in the world.[1] By 2035, an estimated 2.4 million cases of CRC will be diagnosed annually worldwide.[2] Though 70% of people diagnosed with stage III CRC are alive 5 years after diagnosis, the average survival rate for people with metastatic cancer is 3 years.[3] To date, it has been hard to determine which individuals will fall prey to the disease despite therapeutic intervention.

Research by Enzo Medico, MD, PhD, an associate professor at the University of Torino, School of Medicine, could change the CRC diagnostic paradigm. Dr. Medico leads an oncogenomics laboratory at Italy's Candiolo Cancer Institute (IRCCS) that is performing integrative genomic and molecular analysis of CRC tumor samples to subtype them into actionable clinical subgroups. Their research has uncovered fascinating new evidence that gene signatures defining the poor prognosis stem/serrated/mesenchymal (SSM) CRC subtype reflect a strong contribution of cancer-associated fibroblasts (CAF)—the stromal or connective tissue cells of the colon and rectum.

iCommunity spoke with Dr. Medico about the value of patient-derived xenografts, how CRC subtyping could aid treatment, and how the power of integrative genomic techniques is transforming the way we study and treat cancers.

Q: What sparked your interest in studying cancer?
Enzo Medico (EM): While I was studying medicine, I started doing some lab work in a research laboratory at the University of Torino. I worked on experimental cancer research, and essentially, I've been studying cancer ever since.

Cancer is a terrible disease. It always finds a way to keep spreading, despite treatment. I became interested in cancer because I believed that there had to be a way to understand how it grows and metastasizes. Cancer affects so many people. I wanted to do research that could further our knowledge of the disease and improve treatment.

Q: When did you begin using microarrays in your research studies?
EM: I began working with microarrays in the late 1990s. We were studying cancer metastasis in cellular models. We figured out that by exploring the transcriptional changes during the induction of invasiveness, we could identify genetic programs for metastasis. I published my first paper using microarrays in 2001. At the time, we did not have a microarray scanner in our laboratory. We conceived the experiments and then had the expression profiling performed by a service provider. Then we analyzed the data and validated the findings with functional studies.

I became more of an expert in using microarrays when I worked in Dr. George Church's laboratory at Harvard Medical School during a short-term collaborative project on transcriptional analysis. In 2005, we acquired an Illumina BeadArray Reader at IRCCS, becoming one of the first customers of the system.

Q: How can integrative genomics enable us to understand cancer progression and metastasis?
EM: Looking only at a single molecular dimension limits you. When multiple molecular analyses on the same sample became feasible, it was a huge advance. Before, you needed a much larger amount of material and it was cost-prohibitive to perform many procedures. Then next-generation sequencing (NGS) systems were introduced, like the Illumina Genome Analyzer. Over the last 10 years, the cost of NGS technology has decreased, enabling us to perform mutational analysis, copy number analysis, RNA profiling, microRNA profiling, and methylation analysis on the same tumor sample. After it became evident that integrative molecular profiling was feasible, several possibilities opened up for our research. We can now compare different molecular dimensions to understand what is happening in tumors.

Q: How does looking at different dimensions of molecular activity improve our understanding of cancer progression and metastasis?
EM: By combining DNA and RNA-level data, we can learn whether a target might have therapeutic value. If there is a mutation, we can see whether the mutated allele is expressed or not, and then determine if it could be a therapeutic target. We can also study gene

Dr. Enzo Medico is an associate professor at the University of Torino, School of Medicine.

amplification. Normally, when a gene is amplified, it means that there was selection for a tumor cell to have multiple copies of a particular gene. That would qualify the amplified gene as a possible driver of that tumor—and then, perhaps, a therapeutic target. However, this amplification is not always reflected in the overexpression of the transcript. It might not be clear what is driving overexpression. When you combine DNA and RNA-level information, and focus on amplified and overexpressed genes, it is easier to see the genes that might be driving the cancer.

Q: What triggered your focus on the role that SSM transcriptional subtypes might play in CRC?
EM: We were performing gene expression profile research, concentrating more on tumor models than on using human tumor samples. We became interested in subtyping CRC to improve characterization of the CRC cell line collections that we were assembling at IRCCS. We studied recent CRC gene expression profile papers, and saw that all had one subtype which was considered to have a very poor prognosis. It featured low cell differentiation of epithelial to mesenchymal transition. Although other subtypes were more variable among the various studies, there was always this one aggressive, mesenchymal, stem-like type of tumor. When we assembled the CRC gene expression data set from the Cancer Genome Atlas (TCGA), we were able to define it as a strong homogenous subtype in human CRC. We processed RNA-Seq data from 450 samples profiled using the Illumina Genome Analyzer and HiSeq® System, concentrating on the 3 classifiers we thought were the most representative of what was found.[4] We thought if we tried to apply the classifiers to this independent data set, then every sample could be assigned to a subtype by any classifier and we could see the extent of the overlaps. There was a consensus assignment where one sample, if assigned to one subtype of classifier A, was also a given subtype of classifier B. We found a significant, but not absolute, consensus in the SSM subtype. This was the most concordant subtype found by all the classifiers we explored.

> "When you combine DNA and RNA-level information, and focus on amplified and overexpressed genes, it is easier to see the genes that might be driving the cancer."

Q: What prompted you to consider that this SSM subtype could be derived from stromal rather than epithelial cancer cells?
EM: When we tried to classify our collection of expression data from patient-derived xenografts, we saw that the subtype was almost disappearing. It looked like we didn't have any SSM subtype in our samples. We had 2 hypotheses for why this was happening. Perhaps we lost the features of the epithelial to mesenchymal transition by transplanting the human tumor in the mouse tissue. This was a reasonable explanation, although the capability of originating a new cancer is still maintained in this tumor that grows and propagates in mice.

The other explanation was that the subtype was derived from the stromal cells. This came out of the fact that our patient-derived xenograft profiles were obtained with Illumina HumanHT12 Arrays. HumanHT12 Arrays have 50-mer probes that are long and specific, so they hybridize much better on human transcript than on mouse transcripts. When a human tumor is transplanted into an immunocompromised mouse, the human stroma doesn't really grow in the mouse. It is substituted by mouse stroma, so the transcripts originating from the stroma in the xenografts are from the mouse. The human array will not detect the signal of a mouse stromal transcript. We were able to confirm this hypothesis, finding that the SSM subtype is not derived from epithelial cells, but from stromal cells.

> "We can use each gene's expression as a proxy of the stromal abundance and stromal function in a human sample."

Q: How did you distinguish between stromal and tumor cells in CRC?
EM: When we sequenced mRNA from xenograft samples, we sequenced together both human and mouse transcripts. By mapping the reads, we could distinguish which were mouse and which were from the human cells. That allowed us to measure precisely the expression level of both the human and mouse transcripts so we could calculate the fraction of transcript derived from stromal cells and the fraction from cancer cells. Using this technique, we identified genes that were expressed exclusively by stromal cells because we couldn't find human transcripts for those genes.

The technique has allowed us to move away from tumor models and move to human samples. We now have a list of genes that we know are not expressed by epithelial cancer cells—but rather, the stromal cells. We can use each gene's expression as a proxy of the stromal abundance and stromal function in a human sample.

Q: Why was a human tumor xenograft necessary?
EM: Previous microarray approaches sorted out the cells so they could obtain pure populations of epithelial cells, leukocytes, endothelial cells, and fibroblasts—and then perform expression profiling. In this way, one could have specific expression profiles for the different components to help define which genes were involved. But with a xenograft, you perform the profiling of the tumor as such—and now all the components are present in the appropriate fraction within the tumor. It allows you to calculate what fraction of the gene is human and what fraction is mouse, and therefore stromal. This could not be done by expression profiling of individual populations. For example, if a particular transcript is expressed 10 times more highly in a fibroblast than a cancer cell, you can see this expression profiling in a sorted population. However, if there are only 10% of fibroblasts in a tumor, 50% of the expression of the gene is coming from cancer cells and 50% coming from fibroblasts. So that gene, even if it is expressed 10 times more by the fibroblasts, cannot be considered fibroblast-specific. When you look at real tumors growing as xenografts, you can

calculate the fraction of the transcript that originated from the stroma. In this way you are calculating stromal contribution to gene expression and to the disease. It's much more precise and, therefore it's easier to select genes that are specific to either cancer or stromal cells.

Q: How did you identify the cancer associated fibroblast (CAF) specific stromal signature?
EM: We built upon the work in which single cell types were sorted for CRC. We had the expression profiles for different cell populations and removed all the genes we found expressed by human cancer cells. It allowed us to filter out the genes for which epithelial contribution could confound our measurements. From there, we built transcriptional signatures specific to each cell type. We then mapped them to the human tumors in such a way that we could build a leukocyte score, an endothelial cell score, and a fibroblast score. We saw they were overlapping, but not identical. There were tumors richer in fibroblasts, others that were rich in endothelial cells, and others in leukocytes.

Q: Why would SSM genes be expressed at a higher level by these CAFs?
EM: We're not sure SSM genes are expressed at higher levels by CAFs. It's possible that there are more CAFs in a given tumor than in another. We don't really know whether a higher signature means higher expression by the stromal cells or a higher number of stromal cells. It's likely a combination of both.

> "We are now using the NextSeq® 500 System to study the interaction between cancer cells and stroma by exploring ligand receptor couples in which the ligand is of stromal origin and the receptor is of cancer cell origin."

Q: How does this CAF-specific stromal signature relate to treatment?
EM: We have seen that the poor prognosis associated with the CAF signature is significant in tumors that do not undergo treatment. With tumors at a lower stage, not treated with chemotherapy after surgery, we saw that the CAF signature has a very strong prognostic value. In the tumors that were treated with chemotherapy, our data suggests that CAF-rich tumors might be more responsive to chemotherapy.

With CRC, it is common to treat the tumor with radiotherapy before removal to reduce its size and support less invasive surgery. However, we saw that stromal scores were correlated with resistance to radiotherapy. A future challenge is to assess precisely the therapies where stroma confers resistance, and where stroma confers sensitivity. It's currently not clear.

Q: Are tumor subtypes with high stromal gene expression unique to CRC?
EM: We have been focusing on CRC, but many of these genes are generic. We have been communicating with other research groups who are performing similar analyses on other types of cancer xenografts and their findings are similar. I anticipate this is a generic feature of tumors and by exploiting patient-derived xenografts, it will be possible to determine precisely which genes are contributed by the stroma and which are contributed by cancer cells.

> "To perform transcription analysis, we prefer RNA-Seq because we can assess fusion transcripts and alternative splicing. The advantage of NGS is enormous."

Q: What have you learned about the role that stromal cells play in tumorigenesis?
EM: When we first found this contribution, I went straight to my pathologist and asked him whether anyone had found a prognostic correlation between stromal tumors and CRC prognosis. He said that there had been efforts to find a correlation, but that nothing conclusive had come out of them. My interpretation is that when you look at the tumor section on the microscope you can see the stromal abundance. However, you don't have any real measurement of stromal activity. When we measure gene expression, we are measuring the amount of RNA obtained from the stroma, reflecting the activity of the stroma. A tumor with a strong stromal signature is not just a tumor full of inactive scars, but a tumor where the stroma plays an active role metabolically and transcriptionally. Ideally, we would like to use our findings to create a simpler tumor analysis method with RNA-Seq or microarrays by looking at specific genes to find a prognostic correlate that can be seen with conventional pathology.

Q: Does the CRC classification system need to be updated to reflect your discoveries?
EM: Yes, the CRC classification system is being updated now. We are building a new classifier, exploiting expression data from patient-derived xenografts to perform class discovery in the absence of the confounding stromal gene expression. We are identifying subtypes that reflect the intrinsic genetic and functional features of CRC cells. We think it is more precise to evaluate these features of cancer cells and then evaluate the stromal activity and components. This will allow us to be more precise in understanding how aggressive a particular tumor is—and how it might respond to treatment.

Q: When did you start using Illumina sequencing systems?
EM: We became familiar with Illumina systems in 2011. The quality and speed of Illumina NGS systems is ever improving. I was very satisfied with our first experiments and with the improvements being made in the technology to make it more affordable and efficient. The quality is quite good, especially for quantitative analysis where single-base sequencing errors do not count very much.

Q: What impact has NGS had on your research?
EM: NGS has been fundamental to our work. It has enabled us to discriminate stromal and cancer cell contribution to the transcriptome precisely. We could not have performed these studies with microarrays alone. We are now using the NextSeq 500 System to study the interaction between cancer cells and stroma by exploring ligand receptor couples in which the ligand is of stromal origin and the receptor is of cancer cell origin. NGS allows us to identify precisely those ligand receptor couples to study the ongoing interaction. This could have very important therapeutic implications.

> "NGS is changing the field of cancer. It enables us to explore many genes at one time, finding mutations that individually are rare, but still very relevant."

Q: When do you use RNA-Seq vs. microarrays for gene expression studies?
EM: You can have a microarray data set generated in one week, while RNA-Seq requires a lot more time because analysis is still a significant issue. It depends on the result we are trying to obtain. When we want to perform functional analysis to find signatures, microarrays are more than enough. To perform transcription analysis, we prefer RNA-Seq because we can assess fusion transcripts and alternative splicing. The advantage of NGS is enormous.

Q: What are the next steps in your research?
EM: First, we want to review the molecular taxonomy of CRC and add in the new "stroma-aware" taxonomy. Our goal is to explore the clinical, pharmacological, and biological correlates of these various subtypes. We have more than 150 colorectal cell lines and 500 patient-derived xenographs to classify. We'd like to extend this kind of analysis in the future to test core subtype sample models with drugs and other interventions to understand dependencies and drug sensitivities. We also want to explore tumor stromal interaction by species-specific sequencing through integrative techniques.

Q: How do you see NGS being implemented in the future?
EM: NGS is changing the field of cancer. It enables us to explore many genes at one time, finding mutations that individually are rare, but still very relevant. Instead of covering 1% of cases by testing a single rare genetic event, we can put more genes together and end up finding a potential therapeutic target for 10–20% of cases. NGS is essential for this demanding research.

References

1. Global Cancer Facts & Figures, 3rd Edition. American Cancer Society. www.cancer.org/acs/groups/content/@research/documents/document/acspc-044738.pdf. Accessed September 28, 2015.
2. Colorectal Cancer Statistics. World Cancer Research Fund International. www.wcrf.org/int/cancer-facts-figures/data-specific-cancers/colorectal-cancer-statistics. Accessed September 28, 2015.
3. Your Guide in the Fight. Stage III and Stage IV Colorectal Cancer. Fight Colorectal Cancer. http://fightcolorectalcancer.org/fight-it/survival-statistics/. Accessed September 28, 2015.
4. Isella C, Terrasi A, Petti C, et al. Stromal contribution to the colorectal cancer transcriptome. Nat Genet. 2015; 47(4): 312–319.
5. Chang HY, Nuyten DS, Sneddon JB, et al. Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. Proc Natl Acad Sci U S A. 2005; 102(10): 3738–3743.

Learn more about the Illumina products and systems mentioned in this article:

HiSeq Systems
www.illumina.com/systems/hiseq_2500_1500.html

NextSeq 500 System
www.illumina.com/systems/nextseq-sequencer.html

Illumina • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

For Research Use Only. Not for use in diagnostic procedures.

illumina®