The 4th conference on Cancer Genomics, organized by the European Association for Cancer Research, was held on 23-26 July 2019 in Cambridge, UK. The conference’s topics focused on the convergence of cancer biology, cancer medicine, new genomic technologies and computational methods with emphasis on statistics, data mining and machine learning. 260 participants from over 20 countries around the world in addition to 17 exhibitors took part in the conference, and 135 posters were displayed in the conference’s two poster sessions.

Attending the conference was an excellent opportunity to get up to date with cutting-edge technologies and computation methods that are relevant to cancer genomics, as well as interact with other Ph.D. students working on cancer genomics projects.

In general, I found the following topics novel and interesting:

- New and improved technologies provide new types of biological data that shed light on new levels of regulations. Examples: multi-omic single cell technologies (measure more than a single omic in each cell) and TCR-sequencing (mapping the repertoire of immune cells in a tumor).
- The need for integration of new combinations of data types – genomic data, health records, MRI scans and pathological images.
• Immuno-therapy is very promising in treating some cancers, but its effect is transient and is limited to some of the patients. This yields a great need for immune-therapy response biomarkers, a need that is being answered by biomarkers of different omics.

• New concepts are emerging on tumor diagnosis, prevention and treatment – for instance, the correlation of clonal complexity, intra-tumor variability and response to immune-therapy, and the prospects of cancer vaccinations (that will generate antibodies for the most frequent clonal antigens).

Here are some specific talks that I found most interesting, and they represent the topics I mentioned above:

**Interpreting the cancer genome through physical and functional models of the cancer cell**
Trey Ideker, UC San Diego, La Jolla, CA, USA

**Abstract:** Maps of gene and protein interactions play an important role in aggregating biological knowledge and are especially important for describing the main pathways involved in cancer. Together with other labs, we have launched the Cancer Cell Map Initiative (ccmi.org), whose goals is to generate a complete map of the gene and protein wiring diagram of a cancer cell. The initiative includes both data generation and also the development of bioinformatics algorithms for map generation and analysis. In our lab we develop a CRISP-R system for mapping gene-gene interactions, and generate maps connecting DNA mutations to downstream cancer events. In addition to methods for visualizing the generated maps, we developed a machine learning system for integrating the above data to create hierarchical models of cancer cells. In a recent paper by Ma et al., we have shown how a deep neural network can use these hierarchical maps for predicting the phenotype of mutated genes. The DCell system was able to predict cell proliferation in yeast after it was trained on millions of yeast genotypes. The vision is to create a similar system for predicting cancer phenotype on human cells.
My thoughts: It is admirable when someone takes the current flat 2D representation of biological pathways, turns them to a hierarchical model and uses it for training state of the art deep learning algorithms to predict phenotype. I also liked the multi-front approach – collecting data with other labs, developing tools for visualizing them and developing AI algorithms for analyzing them.

**Single cell mapping of plasticity in tumors, metastasis, and development**

Dana Pe’er, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, USA

**Abstract:** The single cell sequencing technology is so powerful because of the synchronization of cells comprising tissues. The data of single cell technologies is often noisy, but exhibits a spatial structure (nearby cells are similar) that can be harnessed for analysis. In a recent study, we collected gut cells from developing embryos, used single-cell sequencing to measure their gene expression, and showed that the strongest signal is determined by the location of the cells along the gut tube. In fact, the expression levels of only 20 genes are enough to predict the exact location of a cell along the tube. The surprising thing we saw in the data – two bridges of entropy (measuring probability of fate) – when cells move between locations, they start expressing genes of their new identity (Upper left image on Figure 2). This demonstrates the plasticity of cells along development.
Figure 2: The main signals in the single cell expression data confer to spatial deployment of the cells along the gut and to their development over time.

In a second study, we compared single cell sequencing of stage 1 and 2 Lung adenocarcinomas to metastasis samples. Interestingly, most cell types appearing in the samples were immune cells rather than lung cells (Figure 3). When we zoomed into the lung cells, we identified regeneration transcriptional programs within some of the tumor cells. These transcriptional programs and also wound healing programs are normally characteristic of embryo cells or response to injury, and appeared in a larger extent in metastasis samples compared with stage 1/2 samples (Figure 4). This example shows how tumor plasticity provides tumors with its ability to adapt to challenges, evade therapy and metastasize.
Figure 3: t-SNE analysis of single cell sequencing data from lung tumors shows that most cell types appearing in the tumor are immune cells.

Figure 4: Lung tumor cells show transcriptional programs related to regeneration and wound healing.

My thoughts: Interesting applications of the promising single cell sequencing technology, that tell us how normal and tumor cells "choose" their fate.
Multi-Dimensional Assays and Potential Biomarkers for Immunotherapy

Timothy Looney, Thermo Fisher Scientific, South San Francisco, CA, USA

Abstract: “Immune repertoire” represents the large collection of the highly variable B-cells and T-cells circulating the body. The enormous diversity of T and B cell receptors (TCRs and BCRs) lies at the heart of the adaptive immune system and allows it to identify an astronomical number of internal and external antigens. The TCR repertoire can change greatly with the emergence and during the progression of various conditions including auto-immune diseases and cancer. TCR sequencing is a technology that is based on next generation sequencing, which allows the reading in parallel of millions of TCR recognition sites. The speaker presented his company’s technology for TCR sequencing (a system called TCRB-SR), and presented several projects where the system was used to interrogate the immunological state of biological samples. In one of the projects, TCR sequencing was applied to 25 Non-small-cell lung carcinoma samples pre-labeled as immunotherapy responders (n=11) and non-responders (n=14). Analysis of this data showed that increased TCR convergence in those who benefited from anti-PD-1 therapy. T-cell convergence is the process whereby antigen-driven selection enriches for T-cell receptors having a shared antigen specificity but different amino acid or nucleotide sequence. The study demonstrates the utilization of a new type of biological information, probing the variability of immune cells within a biological system, to assess the level and complexity of the immune response and use it as a biomarker to immunotherapy effectiveness.

My thoughts: New type of biological data that can be used to teach us on tumor characteristics by mapping the specific immune cells taking part in the host's immune response.
A systems level view of breast cancer through the lens of artificial intelligence

Carlos Caldas, CRUK CI, UK

Abstract: In order to understand the effect of chemotherapy on breast tumors, and improve our ability to predict response, we have compared samples before and after chemo and attempted to characterize tumor changes. To this end, we have generated a versatile dataset composed of transcription, digital pathology and MRI images before, after 9w of treatments, and at the time of surgery. This dataset provides multiple levels of data (Figure 5), perturbation and opportunity to create a predictive model, but how can we integrate the three data types?

We started by identifying driver genes based on the transcription data. Pathology images were digitized, manually labeled by pathologists, and a classifier was developed to predict the pathological label based on this training set (Figure 6). Further, a deep learning classifier was developed to distinguish chemotherapy responders from non-responders based on pathological images (Figure 7). The analysis on the digitized pathological images suggests that chemotherapy response is correlated to clonal heterogeneity within the tumor, and to pre-chemotherapy immune response (hypothesis: tumors that response to treatment have already primed to fight the tumor).

We finally used the MRI data together with the expression data in order to identify expression biomarkers that are correlated with significant loss of tumor volume during chemotherapy treatment.

Figure 5: The three data types to be used for predicting chemotherapy response in Breast Cancer
Figure 6: A classifier was trained on 168 human-labeled pathological images for predicting chemotherapy response.

Figure 7: A deep learning algorithm was trained to predict chemotherapy response and identified informative features in the images.

My thoughts: I liked this study because it fuses together the known biology of breast cancer, with new results emerging from applying machine learning algorithms on new types of data combinations. Specifically, it seems like high intra-tumor heterogeneity (detected by features automatically selected by an AI algorithm) correlates with high immune response and with better response to treatment.