The evolution of somatic tissues through the lens of somatic mutations in cancer

1. DNA damage and damage patterns[1]

Cells in an organism have different genomes. Although in molecular genetics we are said that mutations (spontaneous or induced) happen frequently, we are generally taught that cells of an organism have the same genome, which is inconsistent with the first statement. Cells have different machineries and pathways to repair DNA damage, however their function it's not perfect so some of the mutations are never repaired.

Somatic mutations can be caused by either DNA damage events or mismatches done during the replication because of polymerase mistakes. Factors that produce mutations can be endogenous (e.g., polymerase slippage or mismatch repair, UV, ROS) or exogenous processes (e.g., smoking).

What does it take to identify somatic variants in a tumour?

1. Sample from somatic tissue
2. Sequence and aligning with the reference genome
3. Workflows to:
    - Identify somatic vs germline differences
    - Detect somatic variants at a X% allele frequency
    - Remove germline variants and sequencing artifacts

It has been shown huge mutation differences between tissues, depending on cancer location. This variation is given both by the number of mutations (rate) and by the type of mutation. For instance, there's a signature for UV induced melanomas, the vast majority of mutations are C>T and they occur in specific regions (particular context). Those patterns of somatic mutations, found in cancer genomes through informatic and mathematical approaches, are called "mutational signatures".

It is remarkable that mutagenesis is also caused by altered chromatin conformation. Differential DNA damage and repair processes dependent on enzymes access to chromatin. Silenced genes have compact chromatin, so mismatch repair machinery has difficult access to DNA. This has been studied by SNP mapping.

In accordance with the above, Abel's research group, discovered a 10bp damage periodicity, this corresponds with the DNA length that involves a nucleosome. Regions corresponding to the DNA minor groove and are attached to histones have more mutations. It has a clear periodicity, just as nucleosome formation does. Another periodicity pattern is found between nucleosome linkers; repair activity is higher in linker regions. Some low frequency variants in germline turn out to have the same periodicity with the maximum mutation rate in the minor regions of nucleosomes.

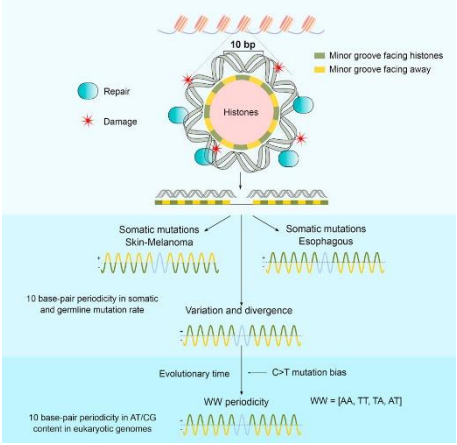

Figure 1: Graphical Abstract. From "Somatic and Germline Mutation Periodicity Follow the Orientation of the DNA Minor Groove around Nucleosomes" (Pich, O. et al.)

[1] Based on: O. Pich, F. Muiños, R. Sabarinathan, I. Reyes-Salazar, A. Gonzalez-Perez, and N. Lopez-Bigas, "Somatic and Germline Mutation Periodicity Follow the Orientation of the DNA Minor Groove around Nucleosomes," *Cell*, vol. 175, no. 4, pp. 1074-1087.e18, Nov. 2018

2. Identifying signatures associated to treatments[2]

As some cancer therapies agents are also carcinogenic agents, it is interesting to acknowledge if therapies leave a discernible mutational footprint or if there are specific treatment mutations. Is important to consider that chemo can cause thousands and thousands of mutations, so a frequency threshold needs to be established.

Some difficulties of this research are that treatments are chosen by tumour type, so some candidate mutations might be related to the tumour, not the treatment. In addition, usually cancers are treated with different medication, so distinguishing which treatment caused each mutation may be challenging. (Figure 2)Treatment-induced mutations occur independently across cells in a tissue after treatment. Therefore, they are private to each surviving cell; thus, their variant allele frequency (VAF) falls below the detection limit of bulk sequencing. However, some of these tumour cells exposed to treatment, experience clonal expansion; as a result, in the metastases treatment-induced mutations may be detectable through bulk sequencing.

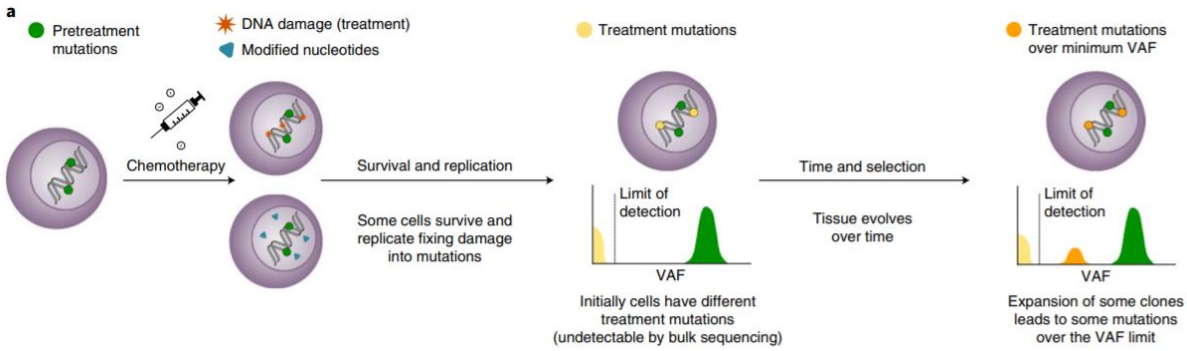

Figure 2: Mutational signatures active in metastatic tumors. From "The mutational footprints of cancer therapies." (Pich, O. et al).

Some cancer therapies damage DNA and cause mutations in both cancerous and healthy cells, so another interesting matter of study is to analyze long term effects in healthy cells. Although very rare, in some people it has been found tumour developing nonrelated with the previous tumours, but they were related to the treatment.

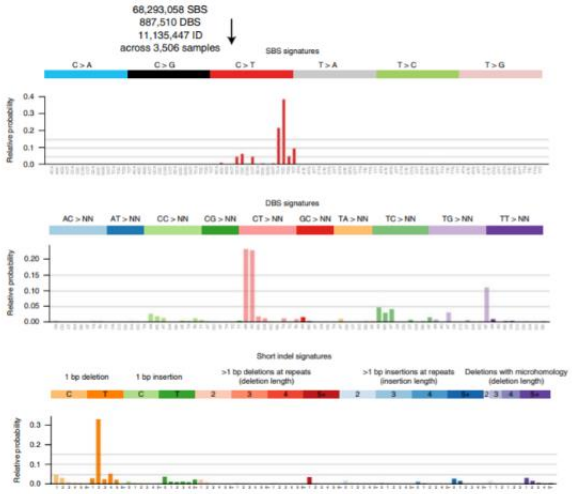

Figure 3: Mutational Signature Extraction; Example SBS, DBS and short-indel signatures extracted from the metastatic cohort using SignatureAnalyze. . From "The mutational footprints of cancer therapies." (Pich, O. et al).

[2] Based on: Pich, O., Muiños, F., Lolkema, M.P. et al. The mutational footprints of cancer therapies. Nat Genet 51, 1732–1740 (2019). https://doi.org/10.1038/s41588-019-0525-5

3. Tumorigenesis[3]

Tumours follow Darwinian evolution: mutations are necessary but no sufficient for tumour developing, they may be maintained in further generations and positively selected upon healthy cells.

Abel's research group (Biomedical genomics) has developed a pipeline with 7 different methods integrated (IntOGen) which apply to cohorts of different tumour types. Another purpose of Biomedical genomics group is to identify drive cancer mutations, as roughly 10% of mutations have been shown to have an oncogenic role. They built a classifier to distinguish driver mutations from passengers by computing the number of driver mutations detected in tumours and obtaining probability vectors.

Features such as signals of positive selection or the consequence type mutation (e.g., missense, nonsense etc.) are the clue to distinguish driver mutations. The actual aim of the group is to continue building gene-tissue specific models, as different genes produce different carcinogenesis development. They use a cross validation to assess the performance of the classifier.

BoostDM (IntOGen) is a method to score all possible point mutations in cancer genes for their potential to be involved in tumorigenesis and accurately recognizes rare oncogenic mutations. In this web you can query the predictions and explanations produced by BoostDM for the point mutations mapping the canonical transcripts in a collection of cancer genes and specific tumour-type contexts.

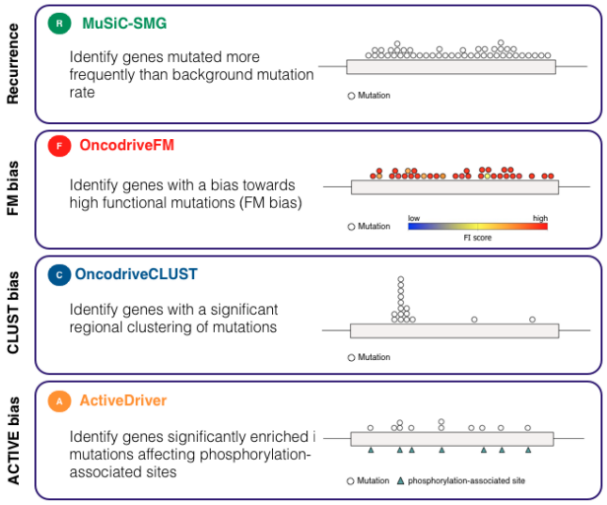

Figure 5: Signals of positive selection used to identify driver genes. From "Comprehensive identification of mutational cancer driver genes across 12 tumor types." (Tamborero, D et al)

Figure 4: QR to BoostDM web app.

It is to point out that a current limitation for the group to progress with more studies is that they have not access to all cancer genomes, so they need more data for better statistical approaches.

Abel ended his lecture mentioning that there are known treatments that don't leave any mutations and it may be due to a positive selection for the pre-existing mutations. So, treatment-mutations may occur, but they are not maintained because of a change in selective pressure. Thus, meaning we can use evolutionary biology to understand basis of tumour progression.

Altogether, although secondary somatic mutations may produce problems at long-term period, it is essential to balance mutational toxicity with the beneficial effects.

Opinion

I found these topics fascinating and the way the group approached them, an innovative perspective. Furthermore, this lecture made me realise that they would not have been able to obtain all these results, or at least not in such accurate way, without bioinformatics. I enjoyed the biological approach Abel gave to the speech but at the same time he was linking it to bioinformatics.

[3] Based on: Tamborero, D., Gonzalez-Perez, A., Perez-Llamas, C. et al. Comprehensive identification of mutational cancer driver genes across 12 tumor types. Sci Rep 3, 2650 (2013). https://doi.org/10.1038/srep02650