Which are the variants behind colorectal cancer?

Besides the research of each SNP that we performed in the overleaf discussion section, we have also looked for characteristic signatures of this type of cancer. This further analysis was performed in order to determinate if the mutations we found through this GWAS analysis can also correlate with the most frequent signatures in colorectal cancer.

The most frequent signatures in colorectal cancer are signature 1, signature 5, signature 6 and signature 7, as we can observe in *Figure 1* if we look in the colorectum column.

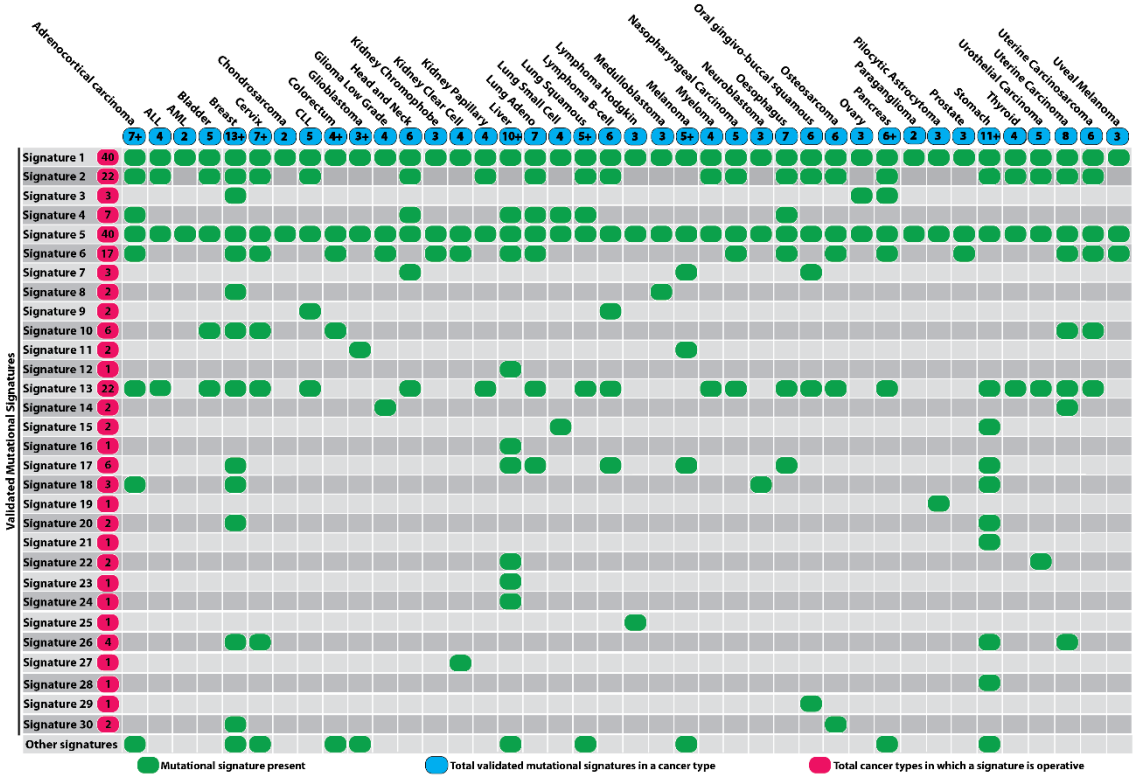

Figure 1. Matrix that correlates diferent types of cancer with signatures. Extracted from COSMIC.

Next step is looking into these specific signatures and determine if there is a correlation with our findings. The first signature is Signature 1 (*Figure 2*). This signature has been found in all cancer types, so it would not be a colorectal cancer specific signature. As we can see in *Figure 2,* the most common substitution here is C>T. If we take a look at our GWAS results, there are four SNPs that have this type of substitution C>T: rs10112382, rs12912791 (*MEF2A* gene), rs280768 and rs10519732 (*CSNK163* gene). So, if 4 out of 11 SNPs found in GWAS are C>T, we can consider it as a frequent substitution also in our analysis.

This signature is thought to be a result of an endogenous mutational process initiated by spontaneous deamination of 5-methylcytosine. Furthermore, it correlates with age of cancer diagnosis, but with out data we can't check if this is found as well in our analysis.

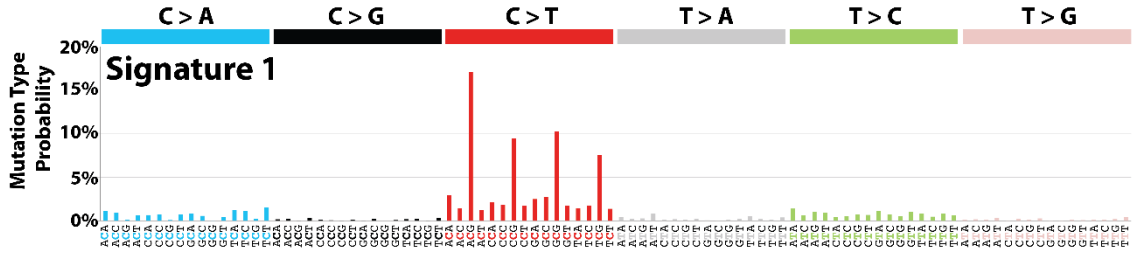

Figure 2. Signature 1. Extracted from COSMIC.

The next signature that can be found in colorectal cancer is Signature 5 (*Figure 3*). Like Signature 1, this signature has been found in all cancer types, so it would not be a colorectal cancer specific signature. As we can see in *Figure 3*, this signature shows that many different substitutions can be found, so we can't focus in any in particular. Its aetiology is unknown.

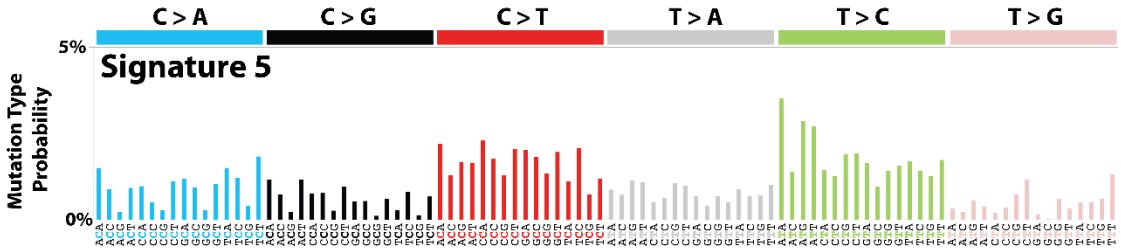

Figure 3. Signature 5. Extracted from COSMIC.

Signature 6 has been found in 17 cancer types, which means it is more specific than the two we have mentioned before but, even so, it still is represented in a large number of cancer types. Nevertheless, it is more frequent in colorectal and uterine cancer. In *Figure 4* we can observe that the most frequent substitutions of this signature are C>T and C>A, so we can assume the same conclusion as in Signature 1 because in our analysis we didn't find any SNP C>A.

The aetiology of this signature is known. It is associated with defective DNA mismatch repair and is found in microsatellite unstable tumours.

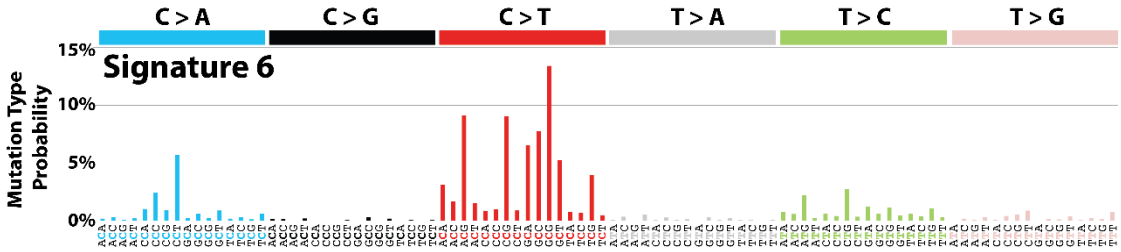

Figure 4. Signature 6. Extracted from COSMIC.

The last signature that correlates with colorectal cancer is Signature 10. This signature is similar to Signature 6, as we can see in *Figure 5*, so we extracted the same conclusions.

The hypothesis proposed is that the mutational process underlying this signature is altered activity of the error-prone polymerase (POLE).

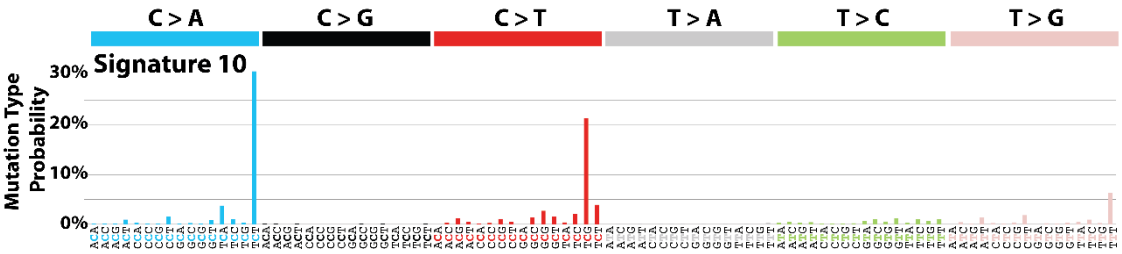

Figure 5. Signature 10. Extracted from COSMIC.

To sum up, the most frequent substitution in cancer types in C>T substitution. Looking into our data, we found 4 SNPs with this substitution, so the frequency is pretty high, but this would not be a colorectal cancer specific signature because this substitution is the most characteristic of many of the signatures.

Our conclusion is that we should look up the nucleotides next to the nucleotide of the SNP in order to be able to classify more specifically our SNPs into this different signatures.