Using a multi-omics approach to study the immune evasion mechanisms of frameshift mutations in colorectal cancer

Minor internship final report

Gerard Llibre Palomar

Evolution and Cancer lab – Barts Cancer Institute

Supervisor: Eszter Lakatos, PhD

CSND Master - Utrecht University

Utrecht University

Barts Cancer Institute

Using a multi-omics approach to study the immune evasion mechanisms of frameshift mutations in colorectal cancer

Abstract

Through the generation of neoantigens, genetic mutations open the door for the immune system to recognize and target tumor cells. Frameshift insertions and deletions (indels) are a highly immunogenic mutation subtype because they alter the reading frame, generating completely novel peptide sequences. Although the nonsense-mediated decay (NMD) pathway degrades the majority of frameshift indels, it is known that some escape decay and elicit anti-tumor immune responses. Using a colorectal cancer multi-omics and multi-region sequencing dataset, we report that expressed frameshift indels are enriched in genomic positions estimated to avoid NMD. Furthermore, integration of chromatin accessibility profiling into the analysis suggested that the selective depletion of the antigenicity is not regulated by epigenetic mechanisms. Therefore, NMD-escape frameshift indels represent an interesting target for biomarker optimisation and immunotherapy outcome prediction.

Summary

- Because they alter the reading frame, frameshift mutations are known to be highly immunogenic
- The majority of frameshift mutations are degraded by an innate mechanism of our body named nonsense-mediated decay pathway
- This defense process is not fully efficient, leaving some mutations to be expressed depending on their position within the gene
- Expressed frameshift indels represent an attractive marker for anti-tumor immune responses

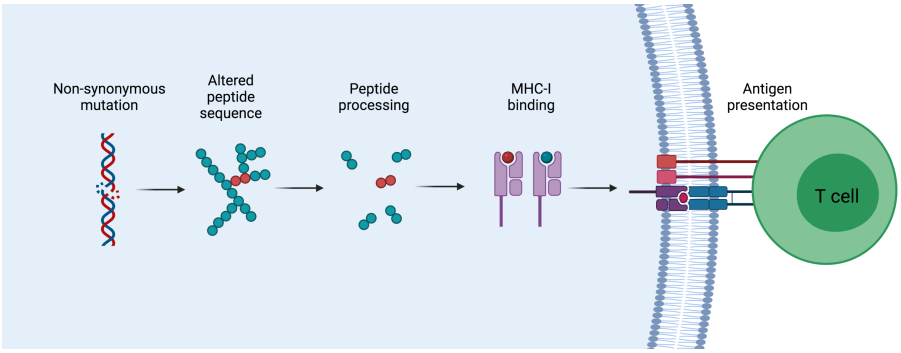
Introduction

The trade-off of genetic mutations

Throughout tumor progression, cancer cells face a crucial trade-off between the acquisition of genetic mutations, which confer them key capacities for a successful malignant transformation, and the fact that the same mutations generate neoantigens that will eventually elicit an immune response. By altering the nucleotide sequence, these genetic mutations generate novel peptides that will be recognized as non-self and bound to the major histocompatibility complexes (MHC) I or II, a process dependent on the patient specific human leukocyte antigen (HLA) alleles. Then,

the neoantigen-MHC complex will be presented in the membrane to finally be recognized by cytotoxic T-cell, triggering anti-tumor response[1] **[Figure 1A].** Recent computational tools such as NetMHCpan[2], which is built upon a neural network training, have allowed to predict the binding affinity between a novel peptide sequence produced by a single mutation and the patient specific HLA alleles. Particularly, frameshift insertions and deletions (indels) are expected to elicit a higher immune response since altering the reading frame leads to the generation of completely unique peptide sequences with greater MHC binding affinity[3] **[Figure 1B].**

A



B

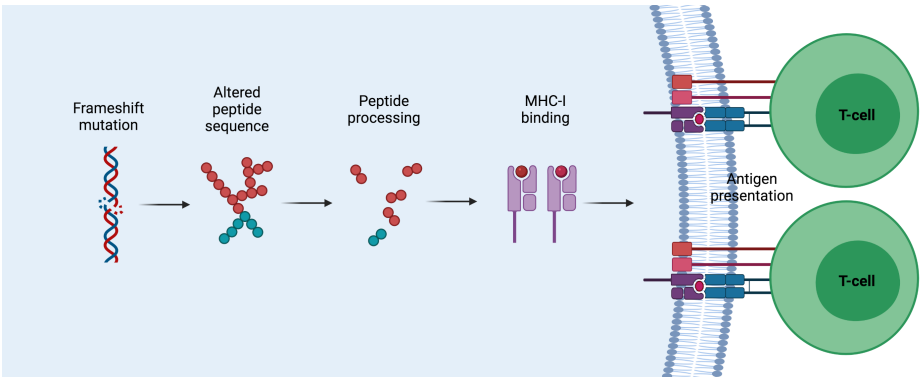


Figure 1. Schematic illustration of neoantigen presentation. *A) Non-synonymous mutations introduce a single nucleotide change, producing a peptide sequence with an altered aminoacid which will bind to the MHC complex as non-self and recognized by T-cells. B) In contrast to non-synonymous mutations, frameshift indels alter the reading frame, thus generating completely new peptide sequences that will elicit an increased immune response by binding with higher affinity to the MHC complex.*

Immune escape

Because of this targeted killing against antigenic represents a negative selective pressure against tumor cells referred as immunoediting, the spectrum of neoantigenic mutations is shaped by ecological and evolutionary interactions between a tumor and its microenvironment[4]. However,

tumor cells can also develop different tools, termed immune escape mechanisms, to reduce the immune system's ability to recognize or react to neoantigens[5]. These can be classified into two subgroups depending on whether they diminish the inherent anti-tumor capacity of the immune system (loss of immunogenicity) or reduce the probability of the tumor to elicit an immune response (loss of antigenicity). Examples of the first group include the production of immunosuppressive cytokines[6], the recruiting of immunosuppressive leukocytes or the stimulation of T-cell inhibitory checkpoint receptors such as CTLA-4[7] and PD-L1[8]. Alternatively, loss of antigenicity can occur through positively selecting cancer cells that lack or mutate immunogenic antigens as well as through the acquisition of flaws in the antigen presentation process (such as loss of MHC expression or dysregulation of the antigen processing machinery)[9]. Interestingly, recent studies have revealed that the nonsense-mediated mRNA decay pathway (NMD) also leads to the downregulation of neoantigen expression through the degradation of mRNAs containing premature termination codons (PTCs)[10,11]. However, it has been recently show that some mutations are able to escape this degradation, leading to higher RNA and protein levels and higher anti-tumor immunogenicity[12], thus pointing towards NMD-escape mutation burden as a promising biomarker for improved response to immunotherapy treatments.

Colorectal cancer

Colorectal cancer (CRC) is the third most frequent and the second most deadly type of cancer with more than 900,000 deaths worldwide in the last year[13]. Most CRC tumors develop from benign adenomas, with only small proportion progressing to malignant carcinoma stage through the sequential acquisition of genetic mutations in oncogenes and tumor suppressor genes[14,15]. CRC tumors are organized into glandular structures, resembling the crypts present in the normal intestinal epithelium, where cell proliferation is directed by a small number of stem cells located at the base of the crypt. Malignant glands follow a similar architecture with only a small subset of cells, termed cancer stem cells (CSC)[16,17], that have the capacity to give rise to the different types of cells that conform the tumor. Therefore cells within a single crypt share the same genetic heritage, making single crypt sequencing a widely used tool as a proxy for single-cell sequencing to study clonal dynamics.

CRC are typically classified into different subgroups according to the molecular features underlying the manifestation of the disease: chromosomal instability (CIN), CpG island methylator phenotype (CIMP) and microsatellite instability (MSI)[18]. Microsatellites, also referred as Short Tandem Repeats (STRs), are small (1-6 base pairs) repeating DNA sequences present in many regions of the genome comprising around 3% of it[19]. Because of their repeated sequences, microsatellites are predisposed to high mutation rate as a consequence of a defective DNA

mismatch repair system. MSI patients show a higher mutational burden and immune cell infiltration compared to microsatellite stable (MSS) or non-MSI patients, translated into greater responses to immunotherapy treatments[20] and a better overall prognosis.

Traditional tumor sequencing protocols have been constantly improving their coverage and depth in order to increase the sensibility to detect genetic alterations. However, single-region sequencing samples only a very small fraction of the total tumor mass, introducing a problematic under-sampling bias into the analysis. Besides, they also fail to incorporate a spatial component into the analysis, a key determinant when studying the clonal dynamics underlying the intra-tumor heterogeneity present in most neoplasms. For this reason, multi-region sequencing approaches are being exponentially used in oncological research as they provide a more representative sampling together with spatial information of the sequenced region in order to accurately capture the intra-tumor heterogeneity[21–23]. Furthermore, since carcinogenesis in general and immune escape in particular involve complex alterations at multiple biological levels, analyzing only mutations only at the DNA level may not allow to have a comprehensive analysis of the full process. Many recent studies have proved that the integration of genetic data with information from other levels - such as transcriptomics, epigenomics or metabolomics – known as multi-omics, permits to extract unique information that would not be found in a genetic-reductionistic analysis[24–26].

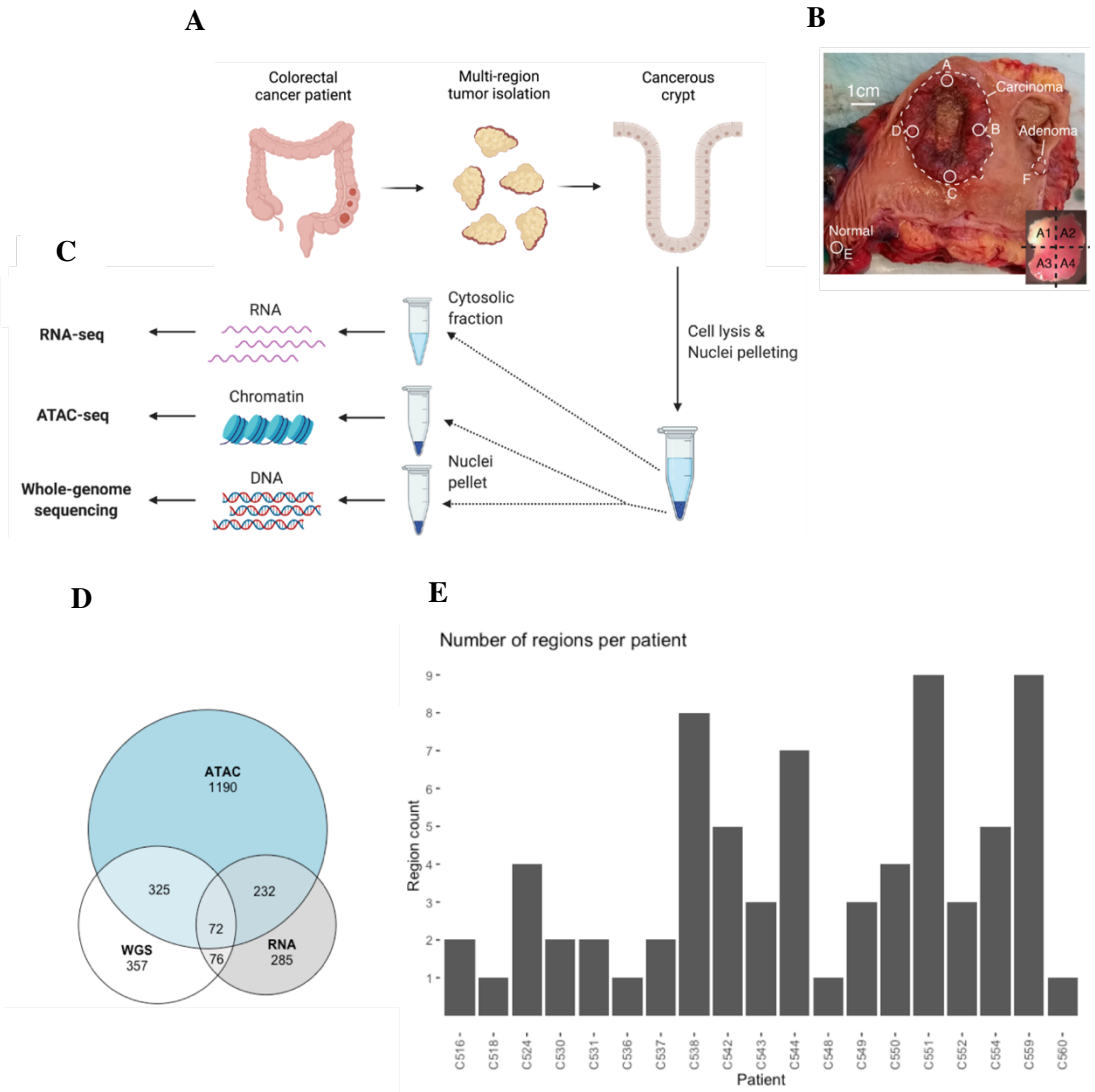
Still, it is unknown how frameshift mutations present in colorectal tumors can reduce their elevated antigenicity to evade the immune system. Here, we present the results of combined DNA, RNA and ATAC single gland profiling of 72 samples from 19 different colorectal cancer patients. Using an innovative multi-omics integrative approach, this study aims to shed light into post-genetic mechanisms, particularly NMD and chromatin organization, regulating the expression and the antigenicity of frameshift indels in colorectal tumors. Identifying those highly immunogenic mutations that escape such regulation could allow to improve the prognosis and immunotherapeutic strategies for colorectal cancer.

Results

For this study, the Evolutionary Predictions In Colorectal Cancer (EPICC) cohort was used. Samples were resected from 19 stage I-III primary colorectal patients from University College London Hospital and a multi-region sampling protocol was conducted by isolating spatially distant regions from every tumor **[Figure 2A-B].** Then, single glands from each region were isolated to lyse the cells to separate nuclear and cytosolic fragments. The nuclear fraction was used to perform whole-genome sequencing (WGS) and chromatin accessibility profiling with

ATAC-seq, while the cytosolic fraction was used to perform full transcriptome RNA-seq **[Figure 2C].** WGS was performed in a total of 830 samples, ATAC-seq was performed in 1.819 samples, and RNA-seq was performed in 595 samples, having a total of 72 samples from 19 patients (median=3 samples per patient) with shared information of the three levels **[Figure 2D-E].**

Genetic variants were called from WGS data using MuTect[27] and Platypus[28] (see Methods) and annotated using Annovar software[29], reporting the type of mutation (synonymous SNV, non-synonymous SNV, frameshift indels…) and the genomic and exonic localization of every variant. Five cases in the cohort were classified as MSI, and consistently showed a significantly higher mutational burden with an increased frameshift indel frequency compared to MSS patients **[Figure 2F-G].** Calculation of the number mutated peptides from each mutation that bind the MHC complex with high affinity (referred as Strong Binders, SB), allowed to confirm that frameshift indels generate higher numbers of SB than non-synonymous and synonymous SNV **[Figure 2H]**, thus reinforcing the aforementioned elevated immunogenicity of frameshift indels.



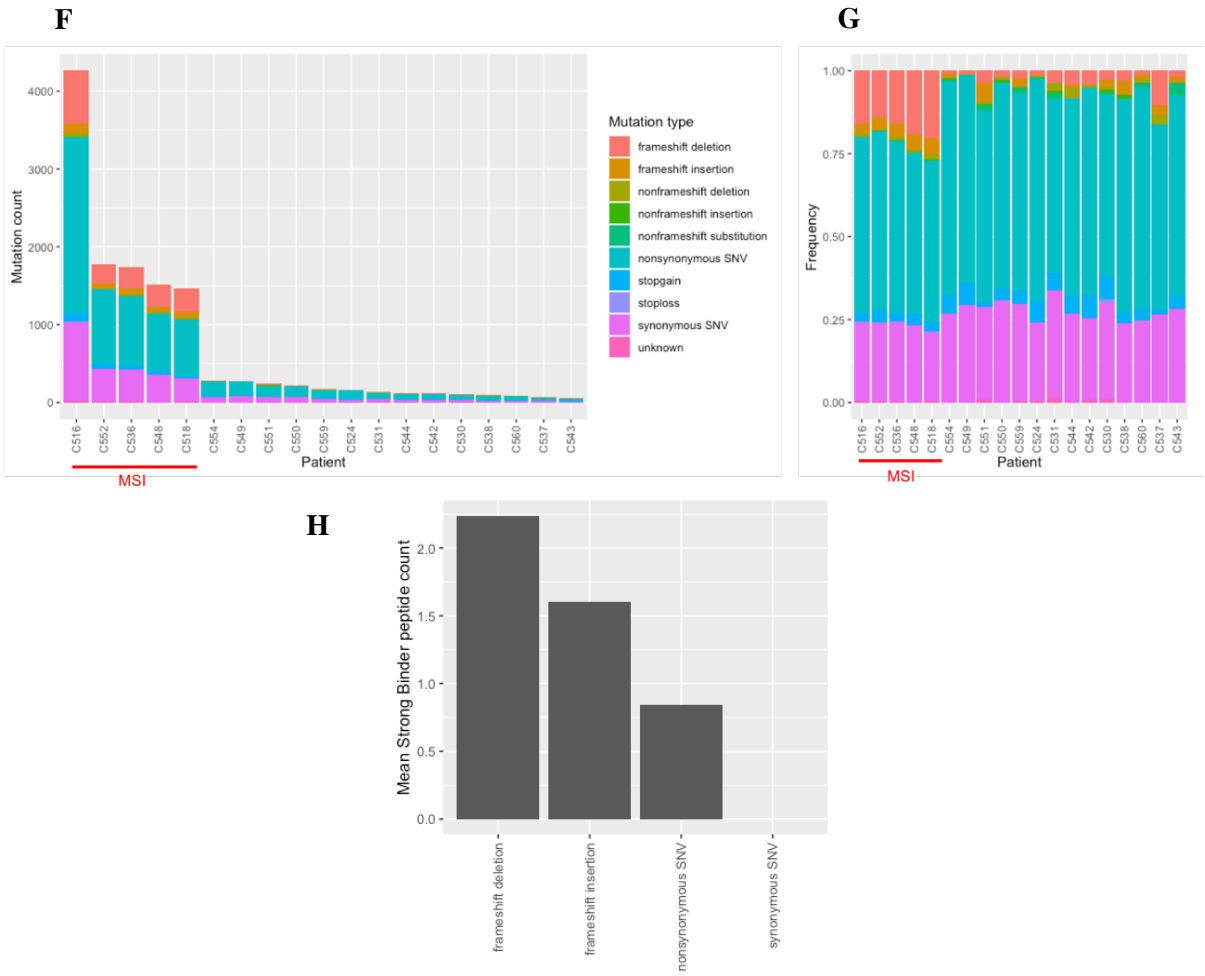
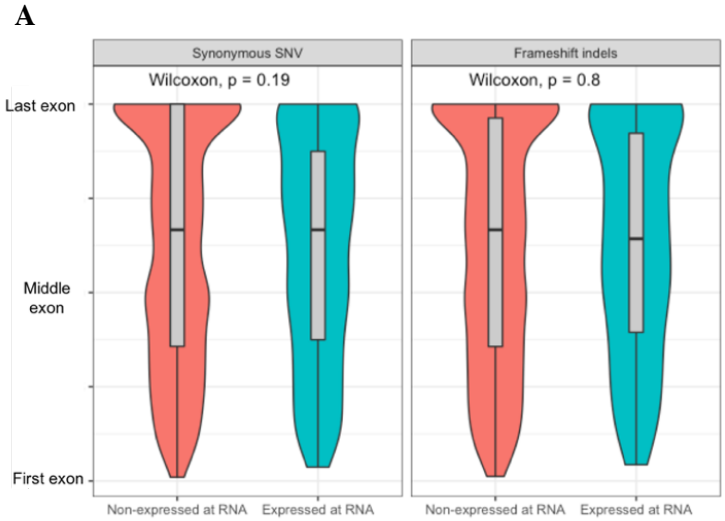


Figure 2. Sample collection protocol and data overview. *A) Fresh colectomy specimens from 19 stage I-III colorectal cancer patients were used to resect cancerous tissue. B) For each patient, different distant regions of the same tumor were collected (A, B, C, D) from which single crypts were isolated for sequencing. C) Lysing the cells allowed to obtain a nuclei pellet to separate nuclear and cytosolic fragments. The nuclear fraction was used to perform whole-genome sequencing (WGS) and chromatin accessibility profiling with ATAC-seq, while the cytosolic fraction was used to perform full transcriptome RNA-seq. Adapted from Heide et al.,*[30] *D) Venn diagram representing the overlapping samples across the three different sequencing methods. Only samples with shared information (n=72) were included in the analysis. E) Number of tumor regions with shared DNA/RNA/ATAC data per patient. F) Mutational burden across patients, with MSI patients separately. G) Frequency of each mutation type across patients. H) Bar plot of the mean number of strong binder peptides per mutation type*

We first evaluated if the expression pattern of frameshift mutations was influenced by nonsense-mediated decay (NMD). NMD is only partially efficient and the exonic positioning of the mutation is a crucial determinant to escape degradation: mutations on (i) last gene exon, (ii)

penultimate exon within 50 nucleotides of the 3′ exon junction, and (iii) first exon within the first 200 nucleotides of coding sequence (CDS) have been linked with a reduced NMD efficiency. Combining the exonic position of each variant with the total number of exons of the gene affected by the mutation, we were able to determine if the mutation fell in the first, middle, penultimate or last exon in consonance with the aforementioned NMD-escape rules. Furthermore, with the RNA-seq expression counts we were able to address the variant allele frequency (VAF) of each mutation at the transcriptomic level. In order to maintain a unified criteria similar to previous studies[12], if the variant had an RNA VAF ≥ 0.05 in at least one of the regions from that patient, we considered it to be an *expressed mutation*. Following this approach we observed that the exonic position of expressed frameshift indels was not significantly different than non-expressed ones and very similar to the pattern shown by synonymous SNVs **[Figure 3A]**, mutations that are expected to not provide any selective advantage to cancer cells and therefore represent a negative evolutionary control. However, when classifying all expressed frameshift indels by exon group, a significantly higher RNA VAF was observed on mutations falling in the first and last exons compared to the ones falling in middle and penultimate ones **[Figure 3B]**, consistent with NMD-escape rules that dictate that a higher proportion of first and last exon mutations are expressed. As for mutations on penultimate exons, a higher VAF was not observed but this could be explained by the fact that we considered all penultimate mutations (not just those within 50 nucleotides of the 3' exon junction) as the low number of mutations made further subsetting unfeasible. This trend seen at the RNA level is not present at the DNA VAF of the same mutations **[Figure 3C]**, confirming that the differences observed in RNA VAF are the result of post-genetic processing.



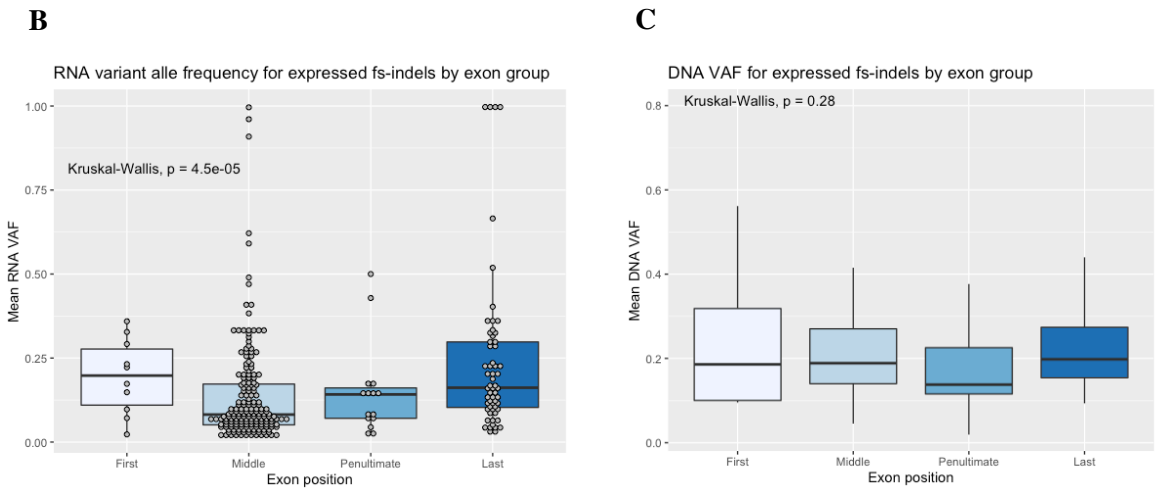


Figure 3. Expression of frameshift indels across exonic positions. *A) Violin plot comparing the exon position of synonymous mutations against frameshift indels. B) RNA variant allele frequency (VAF) of expressed frameshift indels by exon group. Each dot represents the averaged VAF of a single mutation across all regions of a patient. C) Similar to previous plot, DNA variant allele frequency (VAF) of expressed frameshift indels by exon group.*

We next examined inter- and intra-tumour heterogeneity of expression in frameshift indel mutations. Colorectal cancer is characterized by both an increased intra and inter-patient heterogeneity, exemplified by the differences observed between MSI and MSS patients. Taking advantage of the multi-region sequencing we were able to determine the intra-patient clonality of every mutation, that is whether the mutation is present at the DNA level in all the sequenced regions of a tumor (clonal) or restricted to specific regional areas (not clonal). Note that we could only include patients with n≥2 samples available, limiting the analysis to 16 patients and 73 samples. Interestingly, when combining this information with the expression at RNA level, we noticed that expressed frameshift indels are slightly less clonal at the DNA level than non-expressed ones **[Figure 4A],** suggesting that variants that are regionally restricted within the tumor are more likely to be expressed at the RNA level. On the other hand, a huge inter-patient heterogeneity was observed on the percentage of expressed frameshift indels across patients in a MSI-independent manner: while in some patients around 30% of indels annotated at the DNA level were expressed, other patients showed very low percentages **[Figure 4B]**. In order to decipher the reason behind these results, we then included into the analysis the expression of the gene affected by each mutation **[Figure 4C-D].** We defined a gene to be expressed if the Transcripts Per Million (TPM) ≥ 1. This allowed to determine that around 40% of frameshift indels were not expressed due to the whole gene not being expressed. However, 25-55% of the mutations were in expressed genes but still the mutation was not expressed, suggesting that there

may be additional epigenetic events, such as chromatin (re)organization, responsible for depletion of frameshift mutations at the RNA level.

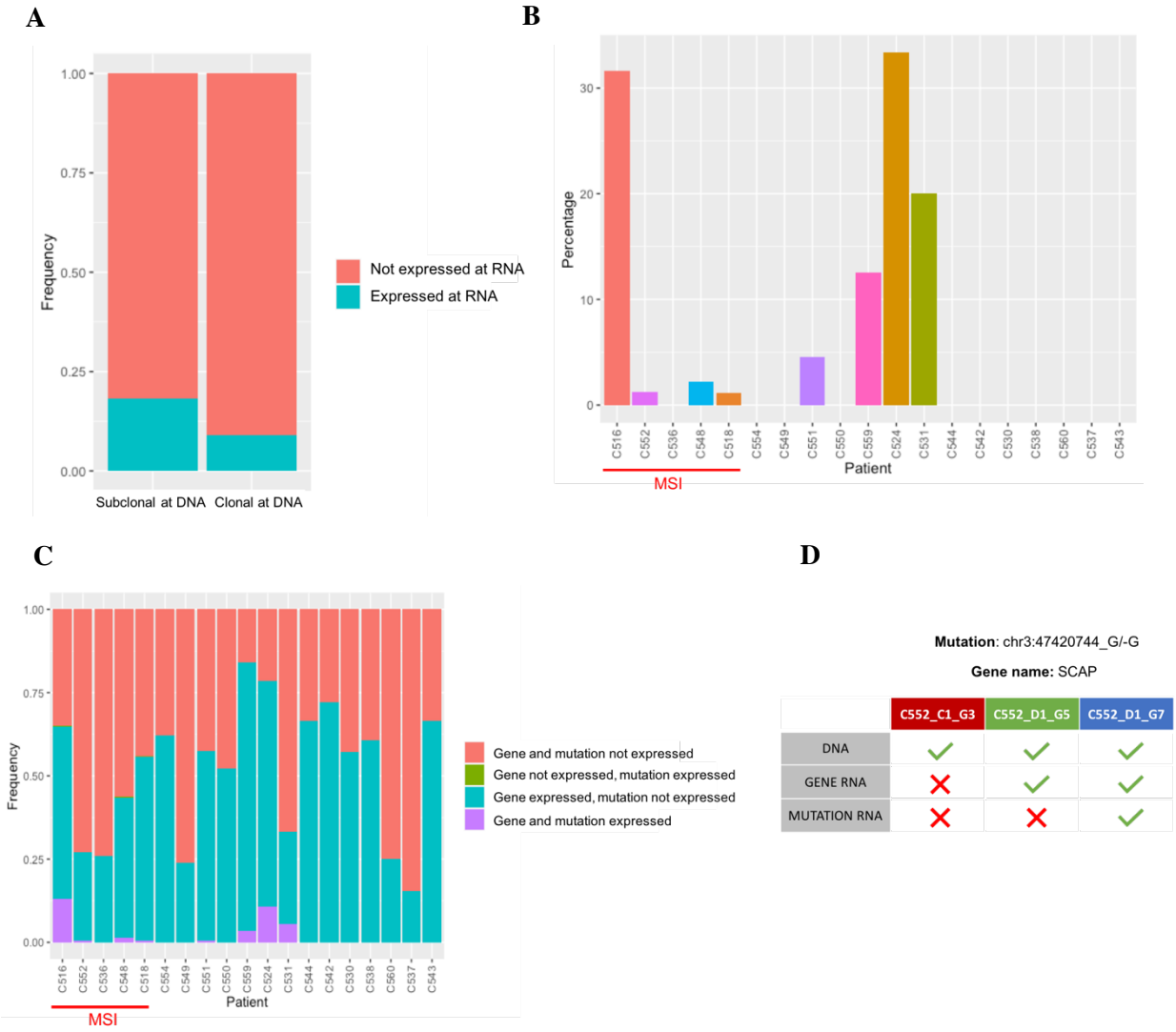
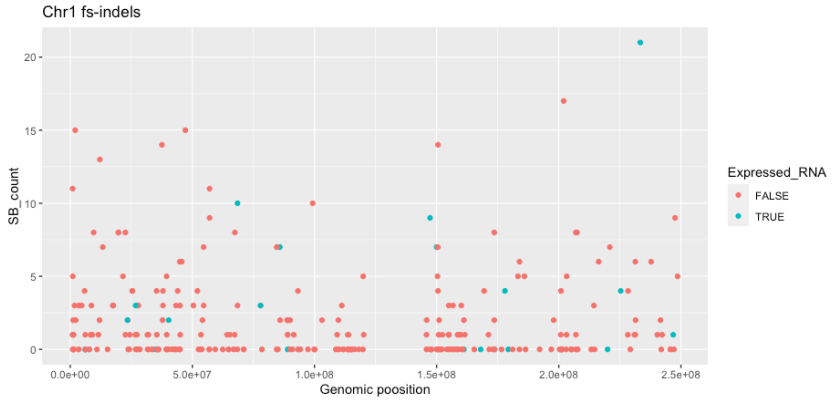


Figure 4. Intra and inter-patient heterogeneity of frameshift indel expression*. A) Frequency of expressed mutations within clonal and subclonal frameshift indels. B) Percentage of frameshift indels expressed. C) Frequency of frameshift mutations colored according to whether the gene and mutation are expressed. D) Schematic table focusing on a single frameshift indel heterogeneously expressed across three different regions of the same tumor: it is present at the DNA level in all regions but the affected gene, named SCAP, is only expressed in C552_D1_G5 and C552_D1_G7, while the mutation is only expressed in C552_D1_G7.*

Finally, in order to look for specific epigenetic events behind these differential expression of frameshift indels, we analyzed the chromatin accessibility profiling. Firstly, we plotted all frameshift indels of a single patient across all one chromosome, but we did not observe any trend between the genomic region and the expression at the RNA-level, such as clusters of expressed/non-expressed mutations or preference for certain genomic regions **[Figure 5A].** We

then analyzed chromatin organization using the ATAC-seq data. To eliminate inter-patient variability, we selected a single frameshift mutation in a patient where we observed high intra-tumor heterogeneity in the expression of this mutation **[Fig 4D]**. We focused on the enhancer and promoter regions of the gene SCAP affected by the mutation, as these regions could have a major epigenetic role regarding RNA expression. However, we observed no clear difference between samples in which the mutation was expressed and those were it was not expressed **[Figure 5B]**. In particular, we would expect sample *C552_C1_G3* to show lower ATAC-seq counts (closed chromatin) since neither the gene nor mutation are expressed in that sample, but there is not a significant difference compared to the other two samples. Similarly, we observed no significant difference between ATAC-seq peaks of other genes with intra-tumor heterogeneity either. However, our findings are limited by the quality of the ATAC-seq and the heterogenous number of enhancer regions for every gene, and therefore a more in-depth analysis should be carried out in the future. Nonetheless, despite the preliminary status of these results, they suggest that the difference on the expression levels observed from the RNA-seq data is not driven by epigenetic determinants, thus pointing towards exclusively transcriptomic events such as NMD to explain those differences.

A



B

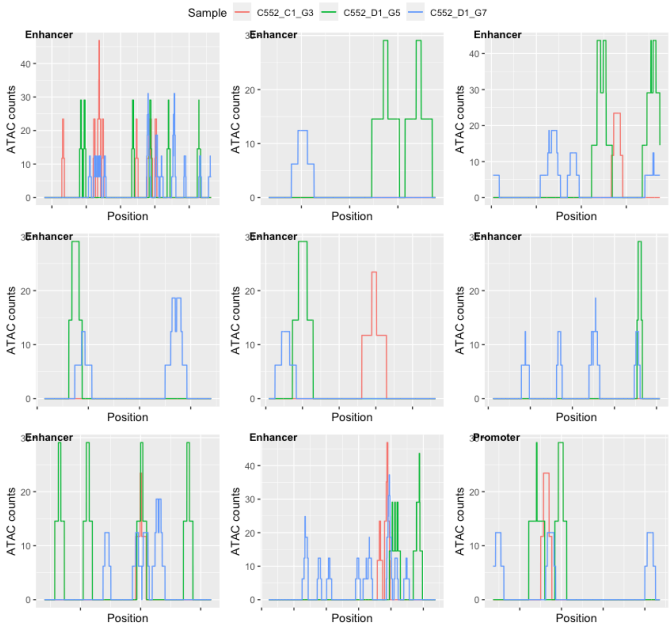


Figure 5. Chromatin organization and frameshift indel expression. *A) All expressed and non-expressed frameshift indels of chromosome 1 plotted by their genomic localization and number of SB. B) ATAC-seq peaks from three different tumor regions on enhancer and promoter regions of a single gene, SCAP, affected by a frameshift indel (see Fig. 4D).*

Conclusions

By adopting an integrative analysis of sequencing data from 19 colorectal cancer patients, a differential RNA expression of fs-indels is observed in accordance to NMD-escape rules: those indels falling in the first and last exons are more likely to avoid NMD mRNA degradation and show a significantly increased expression. This result confirms that this innate NMD degradation system is not fully efficient, thus opening a door for targeting those expressed mutations as long as they are expected to be prone for immune recognition. Furthermore, the EPICC cohort that was used for this study also showed an increased level of heterogeneity, which was translated into differences on the expression of frameshift indels: clonal indels appeared to be slightly less expressed at RNA level than non-clonal ones, while the percentage of expressed fs-indels also showed heterogeneous results across patients in a MSI-independent manner. Finally, the preliminary results obtained from the ATAC-seq suggest that there is no evidence for the aforementioned RNA expression changes to be regulated by chromatin organization events, pointing towards transcriptomic events as the main mechanism acting to selectively deplete the antigenicity of cancerous mutations.

Discussion

The results showed contribute to better depict the mechanistic processes behind differential expression and selective depletion of frameshift neoantigens, highlighting the role of NMD-escape and transcriptomic events over chromatin organization alterations. However, further studies integrating DNA, RNA and ATAC sequencing together with additional immunogenicity information (such as peptide-MHC binding affinity, immune infiltration levels, patient immunotherapy responses or co-cultures of cancer and immune cells) are required to confirm this hypothesis. Particularly, high-throughput ATAC-seq data analysis represents a major challenge because of its sensitivity to data purity and the fact that every gene can have different numbers of enhancers with an unbalanced - and most of the times unknown - responsibility regarding gene expression. Moreover, in this study RNA-seq constituted a limiting bottleneck in terms of sample size due to the quality of the data being affected by the rapid degradation of RNA molecules at the time of sequencing, requiring further work with an increased sample size to confirm the same tendencies. In order to contribute to additional studies deciphering the role of transcriptomic and

epigenomic events on immune evasion mechanisms, the code used for this project will be soon publicly available.

Methods

Study cohort. Samples with matched DNA/RNA/ATAC sequencing (n=72) from the EPICC colorectal cancer cohort were analyzed.

Sample processing and variant calling. Sequencing files were aligned to human genome version hg38 using bwa-mem, Bowtie and STAR for DNA, ATAC and RNA-seq, respectively. Mutations were called from whole-genome sequenced samples against matched blood or adjacent normal tissue using a composite pipeline of Mutect and Platypus. More details of bioinformatic processing are provided in https://www.biorxiv.org/content/10.1101/2021.07.12.451121v1.full and https://www.biorxiv.org/content/10.1101/2021.07.18.451272v1. Filtered mutations were annotated using Annovar. RNA-seq VAF were calculated using Bam-readcounts. Minimum thresholds were set to retain variants with ≥ 5 alternative reads, and VAF ≥0.05. Variants called at DNA were intersected with the ones called at RNA in order to address their expression. ATAC-seq counts were obtained from bed files using the rtracklayer and Gviz R packages.

Neoantigen calling. Human leukocyte antigen (HLA) haplotypes were called *in silico* for each patient by running Polysolver[31] on patient-specific normal samples (combining all available blood or normal tissue). Neoantigens were called with the multi-region mode of NeoPredPipe[32], using patient-specific HLA types and the variant calls derived from Platypus. Binding affinity between neoantigens and MHC was calculated using NetMHCpan[2].

Data processing and statistical analysis. All processing and statistical analysis was done in R version 3.6.1. Krustal-Wallis test was performed for a difference in distribution between three or more independent groups, Wilcoxon test was conducted for two independent groups. P=0,05 value was taken as a being significant.

References

1. Jiang T, Shi T, Zhang H, et al. Tumor neoantigens: from basic research to clinical applications. *J Hematol Oncol 2019 121*. 2019;12(1):1-13. doi:10.1186/S13045-019-0787-5
2. V J, S P, M A, P M, B P, M N. NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *J Immunol*. 2017;199(9):3360-3368. doi:10.4049/JIMMUNOL.1700893
3. S T, K L, H X, et al. Insertion-and-deletion-derived tumour-specific neoantigens and the

immunogenic phenotype: a pan-cancer analysis. *Lancet Oncol*. 2017;18(8):1009-1021. doi:10.1016/S1470-2045(17)30516-8
4. Schreiber RD, Old LJ, Smyth MJ. Cancer immunoediting: Integrating the role of immunity in cancer suppression and promotion. *Science (80-)*. 2011;331(March):78.
5. Beatty GL, Gladney WL. Immune Escape Mechanisms as a Guide for Cancer Immunotherapy. *Clin Cancer Res*. 2015;21(4):687-692. doi:10.1158/1078-0432.CCR-14-1860
6. Kim R, Emi M, Tanabe K, Arihiro K. Tumor-Driven Evolution of Immunosuppressive Networks during Malignant Progression. *Cancer Res*. 2006;66(11):5527-5536. doi:10.1158/0008-5472.CAN-05-4128
7. B R, N H, DM S. CTLA-4: a moving target in immunotherapy. *Blood*. 2018;131(1):58-67. doi:10.1182/BLOOD-2017-06-741033
8. X Z, J L. Soluble PD-1 and PD-L1: predictive and prognostic significance in cancer. *Oncotarget*. 2017;8(57):97671-97682. doi:10.18632/ONCOTARGET.18311
9. Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. *Nature*. 2017;541(7637):321-330. doi:10.1038/nature21349
10. RGH L, M V, B L, F S. The impact of nonsense-mediated mRNA decay on genetic disease, gene editing and cancer immunotherapy. *Nat Genet*. 2019;51(11):1645-1651. doi:10.1038/S41588-019-0517-5
11. RG L, F S, B L. The rules and impact of nonsense-mediated mRNA decay in human cancers. *Nat Genet*. 2016;48(10):1112-1118. doi:10.1038/NG.3664
12. Litchfield K, Reading JL, Lim EL, et al. Escape from nonsense-mediated decay associates with anti-tumor immunogenicity. *Nat Commun*. 2020;11(1):1-11. doi:10.1038/s41467-020-17526-5
13. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin*. 2021;71(3):209-249. doi:10.3322/caac.21660
14. J W, M K, J D, T H, PR G, MW B. Colorectal cancer. *Lancet (London, England)*. 2005;365(9454):153-165. doi:10.1016/S0140-6736(05)17706-X
15. Kuipers EJ, Grady WM, Lieberman D, et al. Colorectal cancer. *Nat Rev Dis Prim 2015 11*. 2015;1(1):1-25. doi:10.1038/nrdp.2015.65
16. J D, RH van J, B P, et al. Sequential cancer mutations in cultured human intestinal stem cells. *Nature*. 2015;521(7550):43-47. doi:10.1038/NATURE14415
17. A Z, M T, G S, R DM. Colorectal cancer stem cells: from the crypt to the clinic. *Cell Stem Cell*. 2014;15(6):692-705. doi:10.1016/J.STEM.2014.11.012
18. Mármol I, Sánchez-de-Diego C, Dieste AP, Cerrada E, Yoldi MJR. Colorectal Carcinoma: A General Overview and Future Perspectives in Colorectal Cancer. *Int J Mol Sci 2017, Vol 18, Page 197*. 2017;18(1):197. doi:10.3390/IJMS18010197
19. Nojadeh JN, Sharif SB, Sakhinia E. Microsatellite instability in colorectal cancer. *EXCLI J*. 2018;17:159. doi:10.17179/EXCLI2017-948
20. Ganesh K, Stadler ZK, Cercek A, et al. Immunotherapy in colorectal cancer: rationale, challenges and potential. *Nat Rev Gastroenterol Hepatol 2019 166*. 2019;16(6):361-375. doi:10.1038/S41575-019-0126-X
21. Turajlic S, Xu H, Litchfield K, et al. Deterministic Evolutionary Trajectories Influence Primary Tumor Growth: TRACERx Renal. *Cell*. 2018;173(3):595-610.e11. doi:10.1016/j.cell.2018.03.043
22. Cross W, Kovac M, Mustonen V, et al. The evolutionary landscape of colorectal tumorigenesis. *Nat Ecol Evol*. 2018;2(10):1661-1672. doi:10.1038/s41559-018-0642-z
23. Yan T, Cui H, Zhou Y, et al. Multi-region sequencing unveils novel actionable targets and spatial heterogeneity in esophageal squamous cell carcinoma. *Nat Commun*. 2019;10(1):1-15. doi:10.1038/s41467-019-09255-1
24. Ritchie MD, Holzinger ER, Li R, Pendergrass SA, Kim D. Methods of integrating data to uncover genotype-phenotype interactions. *Nat Rev Genet*. 2015;16(2):85-97. doi:10.1038/nrg3868
25. Subramanian I, Verma S, Kumar S, Jere A, Anamika K. Multi-omics Data Integration,

Interpretation, and Its Application. *Bioinform Biol Insights*. 2020;14:7-9. doi:10.1177/1177932219899051
26. Nam AS, Chaligne R, Landau DA. Integrating genetic and non-genetic determinants of cancer evolution by single-cell multi-omics. *Nat Rev Genet*. 2021;22(1):3-18. doi:10.1038/s41576-020-0265-5
27. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol 2013 313*. 2013;31(3):213-219. doi:10.1038/nbt.2514
28. Rimmer A, Phan H, Mathieson I, et al. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. *Nat Genet 2014 468*. 2014;46(8):912-918. doi:10.1038/ng.3036
29. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):1-7. doi:10.1093/nar/gkq603
30. Heide T, Househam J, Spiteri I, et al. The co-evolution of the genome and epigenome in colorectal cancer. *bioRxiv*. 2021:2021.07.12.451121.
31. Shukla SA, Rooney MS, Rajasagi M, et al. Comprehensive analysis of cancer-associated somatic mutations in class i HLA genes. *Nat Biotechnol*. 2015;33(11):1152-1158. doi:10.1038/nbt.3344
32. Schenck RO, Lakatos E, Gatenbee C, Graham TA, Anderson ARA. NeoPredPipe: high-throughput neoantigen prediction and recognition potential pipeline. *BMC Bioinforma 2019 201*. 2019;20(1):1-6. doi:10.1186/S12859-019-2876-4