Oral carcinoma biomarkers in Taiwanese population

Clarence Lei B. Bautista, Tanya Murakhovskaya, Elizabeth Kaminsky, Aleksei Efremov, Yaroslav Lozinsky, Katerina Nuzhdina

Summary

Problem

- Finding and studying biomarkers for OSCC (Oral Squamous Cell Carcinoma), concentrating on APOBEC group of genes with a frequent mutation profile.

Results

- Omics - analysis and integration
- Comparison with article's results
- Revealing differences and tech-issues, which affected the results
- Approbation of the biomarker

Background

Oral cancer is one of the most common forms of cancer found in humans where approximately 300,373 new cases were estimated in 2012 indicating its threat. Majority (>90%) of oral cancer cases can be classified under oral squamous cell carcinoma (OSCC). The fourth most common type of malignancy found in the Taiwan males is OSCC, owing the 6308 new cases during 2012. This can be attributed to risk behaviours such as cigarette and alcohol consumption as well as betel nut chewing.

One of the prominent mutational signatures observed for head and neck cancer are those found in the APOBEC (Apolipoprotein B mRNA Editing enzyme, Catalytic polypeptide) gene clusters. APOBEC genes are responsible for encoding a superfamily of enzymes responsible for attaching and mutating cytidines found in RNA to uracil (C->U). These genes are thought to be helpful in the epigenetic immune response of our body to viral infections and cancerous cells through transcriptional regulation. The human genome encodes for a total of 11 APOBEC genes namely: *APOBEC1*, *APOBEC2*, *APOBEC3A*, *APOBEC3B*, *APOBEC3C*, *APOBEC3D*, *APOBEC3E*, *APOBEC3F*, *APOBEC3H*, *APOBEC4*, and *AICDA*.

Recently, a paper explored the relation of *APOBEC* genes with the occurrences of OSCC in Taiwanese patients due to the observed common *APOBEC3B* (A3B) deletion polymorphism. This polymorphism produces a splicing variant of *APOBEC3A* (A3A) where the 3' UTR region of A3B attaches to it, named *APOBEC3A_B* (A3A_B). In our project, we explored the findings in this paper again through transcriptomic and exome analysis of the patients data to verify if same findings can be said through the use of alternative methods and algorithms.

Methods

Databases
NCBI, ENA, GeneCards, Uniprot

↓

Collecting data and running programmes
(Linux server, command line)

↓

Quality control
FastQC

↓

RNASeq: Genes expression
Kallisto

↓

WES: Reads mapping
Bwa-mem alignment (fastqc -> bam files)

↓

WES: Variant calling - somatic & germline mutations
STRELKA + Manta (insertion and deletion)

↓

WES: Copy number alterations (CNA)
Sequenza

↓

Visual analysis
Jupyter notebooks

Results

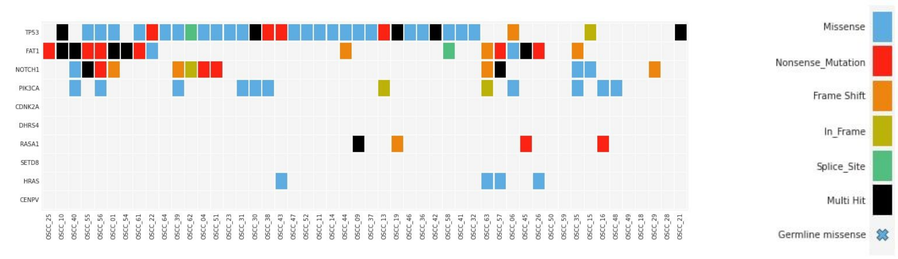
1) **FastQC analysis** of the read files showed that the quality of reads were of good quality meaning they are viable for NGS analysis.
2) **Gene deletions**
were not found as in the article (A3B). We believe that differences in the analysis algorithms used in the pipeline might be the cause for this.
3) **Mutations**
a. Despite numerous mutations being found, they are not expected to have clinical relevance as they are placed in introns, UTR and flanking regions.

```
In [54]: maf[maf.Hugo_Symbol=='APOBEC3A']
```

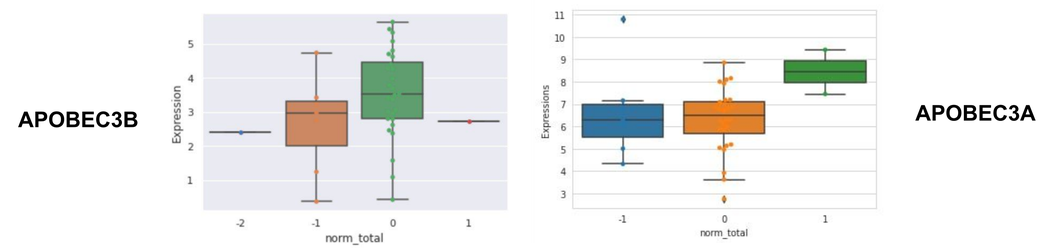
Out[54]:

n	Strand	Variant_Classification	Variant_Type	...	ba1	bs1	bp4	bp7	Tumor_VAF	Normal_VAF	Canonical_chr	chrom_n	HS_pos	Sample
6	+	Intron	SNP	...	False	False	False	False	0.400000	0.0	True	22.0	APOBEC3A_38968966	OSCC_54
8	+	3'UTR	SNP	...	False	False	False	False	0.236842	0.0	True	22.0	APOBEC3A_38962978	OSCC_39
5	+	5'Flank	INS	...	True	True	False	False	0.833333	0.0	True	22.0	APOBEC3A_38948644	OSCC_62
7	+	Intron	INS	...	True	True	False	False	0.523810	0.0	True	22.0	APOBEC3A_38958876	OSCC_52
5	+	3'UTR	SNP	...	False	False	False	False	0.400000	0.0	True	22.0	APOBEC3A_38962905	OSCC_36

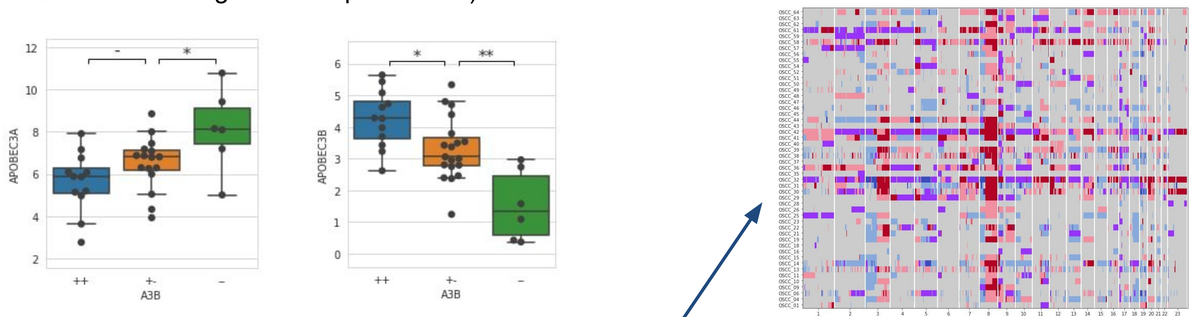
b. For mutation calling of several cancer gene biomarkers, it was observed that they correspond more or less to the mutations found in the previous paper.



4) **Fusions**
We found only several fusions, the majority of them was not annotated and the rest ones were considered unreliable. We also believe that is because of default settings and cohort data



5) **Expression of genes**
a. CNA_genes: Was reliable and corresponded to the article. (Evidence: the more reads are present in CNA file - the higher the expression is)



b. Correlation between APOBEC3A and APOBEC3B genes: corresponds to the article
6) **Interesting pattern** in CNA data for different patients (recurrent 8q amplification)

Conclusions

- Algorithms used for NGS analysis play a big role in the type of data that can be extracted from it. This means that it is important to specify the exact parameters of the processing pipeline used.
- Expression levels of A3A and A3B in relation to A3B deletion polymorphisms are aligned with the findings from the paper.
- Mutations found does not instantly translate to clinical relevance because they are present in non-coding regions of genes.

Future Directions

- *Pipeline parameters* can be changed to do alternative processing of the NGS data acquired. This can show different mutation patterns from the samples.
- Exploring *new* carcinoma *biomarkers* and applying the knowledge to *other types of cancer*.
- Future studies regarding the *clinical prognostic relevance* of other interesting genes in other types of cancer.

References

1. Chen, T.-W., Lee, C.-C., Liu, H., Wu, C.-S., Pickering, C. R., Huang, P.-J., Wang, J., Chang, I. Y.-F., Yeh, Y.-M., Chen, C.-D., Li, H.-P., Luo, J.-D., Tan, B. C.-M., Chan, T. E. H., Hsueh, C., Chu, L. J., Chen, Y.-T., Zhang, B., Yang, C.-Y., … Chang, Y.-S. (2017). APOBEC3A is an oral cancer prognostic biomarker in Taiwanese carriers of an APOBEC deletion polymorphism. *Nature Communications 2017 8:1*, *8*(1), 1–13. https://doi.org/10.1038/s41467-017-00493-9
2. AC Chi, T. D. B. N. (2015). Oral cavity and oropharyngeal squamous cell carcinoma--an update. *CA Cancer. J. Clin.*, *65*(5), 401–421. https://doi.org/10.3322/caac.21293