



UvA-DARE (Digital Academic Repository)

Comparative genomic analysis of oral versus laryngeal and pharyngeal cancer

Vossen, D.M.; Verhagen, C.V.M.; Verheij, M.; Wessels, L.F.A.; Vens, C.; van den Brekel, M.W.M.

DOI

[10.1016/j.oraloncology.2018.04.006](https://doi.org/10.1016/j.oraloncology.2018.04.006)

Publication date

2018

Document Version

Other version

Published in

Oral Oncology

License

Article 25fa Dutch Copyright Act (<https://www.openaccess.nl/en/in-the-netherlands/you-share-we-take-care>)

[Link to publication](#)

Citation for published version (APA):

Vossen, D. M., Verhagen, C. V. M., Verheij, M., Wessels, L. F. A., Vens, C., & van den Brekel, M. W. M. (2018). Comparative genomic analysis of oral versus laryngeal and pharyngeal cancer. *Oral Oncology*, 81, 35-44.
<https://doi.org/10.1016/j.oraloncology.2018.04.006>

General rights

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: <https://uba.uva.nl/en/contact>, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (<https://dare.uva.nl>)

Supplementary Methods

Sequencing and bioinformatics protocol of NKI dataset

Genomic DNA was isolated with the AllPrep DNA Mini Kit (Qiagen). Paired-end libraries were prepared and subsequently target enriched with the TruSeq DNA library preparation kit (Illumina) and a SureSelect custom-based bait library (Agilent) targeting 556 genes (Supplementary Table 1). HPV gene baits were included in order to determine the HPV status of tumor samples. The enriched libraries were sequenced on the Illumina HiSeq 2000 using a 2x75bp paired-end protocol. The reads were aligned to the NCBI build 37 (hg19) human reference genome using the Burrows-Wheeler Aligner backtrack algorithm [1]. Potential PCR duplicates were removed with Picard Tools (<http://picard.sourceforge.net>). The average read coverage was 352.

We called single nucleotide variants and indels with VarScan 2.3.9 [2] in conjunction with Samtools mpileup 0.1.19 [3]. We annotated these variants with Annovar version date 11-05-2016 [4]. The bash script that was used to call and annotate the variants is available at <https://github.com/dvossen/OSCC-versus-LPSCC>. Variants were classified as germline variants or SPMs using public SNP databases because a matched normal sample was unavailable for most tumor samples. Variants were classified as SPMs if they (1) were absent from the 1000 Genomes Project [5], Exome Sequencing Project (Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, WA (URL: <http://evs.gs.washington.edu/EVS/>) (Accessed 22 December 2014)) and Exome Aggregation Consortium [6] databases, or (2) occurred in more than 20 samples of the Catalogue Of Somatic Mutations In Cancer (COSMIC), using an R script available at <https://github.com/dvossen/OSCC-versus-LPSCC>. All analyses were based on these SPMs; listed in Supplementary Table 2.

Copy number and ploidy were estimated from DNA sequencing data with PureCN [7], following the protocol in its manual. We used thresholds recommended by PureCN for calling homozygous deletions (copy number ≤ 0.5) and focal amplifications (copy number ≥ 7).

Mutational signatures

Different mutational processes (e.g. DNA repair defects or tobacco exposure) cause different patterns of somatic point mutations (SPMs) throughout the genome. Such characteristic patterns are often referred to as 'mutational signatures' and COSMIC currently lists 30 curated mutational signatures (<http://cancer.sanger.ac.uk/cosmic/signatures>, 18 Dec 2017). We used the R package deconstructSigs [8] to determine which of these mutational signatures contributed to the SPMs found in each sample. A signature was classified as contributing to a tumor sample's SPMs when its weight was at least 0.06, following deconstructSigs' recommendations [8]. Following these, we only included samples with at least 50 SPMs in this analysis and ran deconstructSigs with default normalization.

We then focused on signatures that made a contribution to at least 25% of OSCC or L/P-SCC samples (signatures 1, 2, 3, 4, 6, 7, 13, 15, 24). For each signature, we compared the

proportion of OSCC and L/P-SCC tumor samples to which the signature made a contribution, using Fisher's exact test. We controlled the false discovery rate at 0.10 by correcting for multiple hypothesis testing with the Benjamini and Hochberg method. The corrected P-values are reported as Q-values.

References

- [1] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–60. doi:10.1093/bioinformatics/btp324.
- [2] Koboldt DC, Zhang Q, Larson DE, Shen D, Mclellan MD, Lin L, et al. VarScan 2 : Somatic mutation and copy number alteration discovery in cancer by exome sequencing VarScan 2 : Somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res* 2012;22:568–76. doi:10.1101/gr.129684.111.
- [3] Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078–9. doi:10.1093/bioinformatics/btp352.
- [4] Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38:e164. doi:10.1093/nar/gkq603.
- [5] Auton A, Abecasis GR, Altshuler DM, Durbin RM, Bentley DR, Chakravarti A, et al. A global reference for human genetic variation. *Nature* 2015;526:68–74. doi:10.1038/nature15393.
- [6] Lek M, Karczewski KJ, Minikel E V., Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016;536:285–91. doi:10.1038/nature19057.
- [7] Riester M, Singh AP, Brannon AR, Yu K, Campbell CD, Chiang DY, et al. PureCN: copy number calling and SNV classification using targeted short read sequencing. *Source Code Biol Med* 2016;11:13. doi:10.1186/s13029-016-0060-z.
- [8] Rosenthal R, McGranahan N, Herrero J, Taylor BS, Swanton C. deconstructSigs: delineating mutational processes in single tumors distinguishes DNA repair deficiencies and patterns of carcinoma evolution. *Genome Biol* 2016;17:31. doi:10.1186/s13059-016-0893-4.