SOFTWARE   **Open Access**

NeoPredPipe: high-throughput neoantigen prediction and recognition potential pipeline

Ryan O. Schenck[1,2]* , Eszter Lakatos[3], Chandler Gatenbee[1], Trevor A. Graham[3] and Alexander R.A. Anderson[1]

Abstract

Background: Next generation sequencing has yielded an unparalleled means of quickly determining the molecular make-up of patient tumors. In conjunction with emerging, effective immunotherapeutics for a number of cancers, this rapid data generation necessitates a paired high-throughput means of predicting and assessing neoantigens from tumor variants that may stimulate immune response.

Results: Here we offer NeoPredPipe (Neoantigen Prediction Pipeline) as a contiguous means of predicting putative neoantigens and their corresponding recognition potentials for both single and multi-region tumor samples. NeoPredPipe is able to quickly provide summary information for researchers, and clinicians alike, on predicted neoantigen burdens while providing high-level insights into tumor heterogeneity given somatic mutation calls and, optionally, patient HLA haplotypes. Given an example dataset we show how NeoPredPipe is able to rapidly provide insights into neoantigen heterogeneity, burden, and immune stimulation potential.

Conclusions: Through the integration of widely adopted tools for neoantigen discovery NeoPredPipe offers a contiguous means of processing single and multi-region sequence data. NeoPredPipe is user-friendly and adaptable for high-throughput performance. NeoPredPipe is freely available at https://github.com/MathOnco/NeoPredPipe.

Keywords: Neoantigens, Cancer, Evolution, Heterogeneity, Next-generation sequencing

Background

Cancer cells are fraught with genomic variants in all regions of the genome with high degrees of heterogeneity in a spatially complex tumor. This intra-tumor hetero-geneity (ITH) realizes a fitness landscape upon which natural selection can act (reviewed by [1]). Neoanti-gens, epitopes derived from proteins translated from non-synonymous variants, are able to make their way to the cell surface in the hopes of stimulating an immune response after a number of cellular processing steps have occurred, primarily proteosomal cleavage and binding with major histocompatibility complexes (MHC) I or II.

This binding depends upon the patient specific human leukocyte antigen (HLA) alleles. From here, the bound neoantigen with its MHC-Class I complex makes its way to the cell surface where it may bind with cytotoxic T-cell receptors thereby eliciting infiltration of cytotoxic T-cells capable of detecting and eliminating cells carry-ing the neoantigen in the absence of immune evading tactics. The immune response is strongly influenced by the total number of neoantigens within a tumor, espe-cially in hyper-mutated cancers ([2]), as well as the ITH of antigenic mutations ([3]). Recent advances in sequencing techniques allow for multi-region sequencing approaches whereby adjacent regions of the same tumor or tissue are able to provide greater insights into variant clon-ality (i.e. truly clonal, subclonal, or shared). There is increasing evidence that the neoantigen landscape of tumours can be highly heterogeneous, containing regions of subclonal immune escape and significantly different

*Correspondence: ryan.schenck@univ.ox.ac.uk
[1]Integrated Mathematical Oncology, Moffitt Cancer Center, 33612 Tampa, FL, USA
[2]Wellcome Centre for Human Genetics, University of Oxford, OX3 7BN Oxford, UK
Full list of author information is available at the end of the article

neoantigen load that can influence a patient's response to immunotherapy [4–6].

A number of tools are available that provide mutated peptide annotation, binding affinity prediction, wild-type and mutant peptide comparison, and neoantigen ranking based on these measures [7–10]. Their input varies from raw sequencing files (e.g. fastq) [7, 8, 10] to highly annotated vcf files [9]; some provide HLA-typing as part of their pipeline [7, 10], but require further dependencies for HLAtyping software. Most rely on a version of netMHC or netMHCpan for binding prediction, but [9] offers a choice of additional software. For an in-depth comparison of available pipelines for neoantigen calling, we refer the reader to the recent review of Lancaster et al. [11].

Despite the increasing number and diversity of neoantigen-prediction tools, none of them possess the capability of providing predicitions on multi-region sequence data and assessing ITH of the antigenic landscape of tumours. Here, we present NeoPredPipe, a pipeline connecting commonly used bioinformatic software via custom python scripts to allow for the processing of single and multi-region variant call format (VCF) files, variant annotations, neoantigen predictions, cross-referencing with known epitopes, and performing in silico TCR recognition potential predictions in a single, clear, and proficient workflow (Fig. 1).

Implementation

The first stage in neoantigen identification from a VCF file is the proper annotation of variants to identify nonsynonymous variants. To this end, NeoPredPipe employs the widely used and efficient genomics tool, ANNOVAR ([12]). Specifically, ANNOVAR processes samples in a way that prioritizes exonic variants, this step provides a useful means for quickly partitioning variant calls for downstream applications. The user is able to specify the genome build that they would like to use, provided it is compatible with ANNOVAR. Finally, using the *coding_change* function of ANNOVAR and custom code, the mutated amino acid sequence is predicted from annotated nonsynonymous variant calls, and the peptide sequence surrounding the newly introduced amino acid is extracted for epitope prediction. From this step, mutations that give rise to a single amino acid change, and mutations that mutate a larger peptide segment (e.g. indels and stop-losses) are handled separately and reported in separate files to help further assessment.

Once the VCF files have been annotated and partitioned with ANNOVAR, the program determines if HLA haplotypes have been provided by the user containing the HLA-A, -B, and -C haplotypes. NeoPredPipe does not include HLA allele identification as this step in the pipeline is highly dependent upon the source of the data (WES, WGS, targeted gene panels, transcriptome data, or
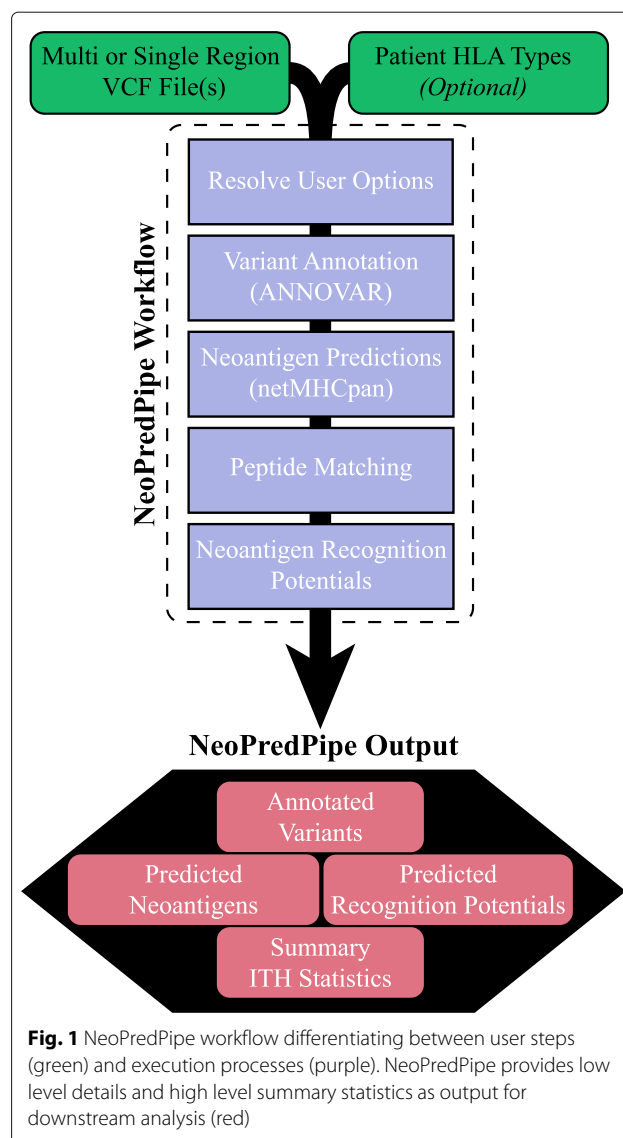

Fig. 1 NeoPredPipe workflow differentiating between user steps (green) and execution processes (purple). NeoPredPipe provides low level details and high level summary statistics as output for downstream analysis (red)

conducted via experimental methods), but the pipeline's github page provides detailed advice on haplotyping from WES/WGS data using the popular tool POLYSOLVER [13], and the output of POLYSOLVER is automatically processed in NeoPredPipe. In cases where no HLA haplotype information is available the most common alleles of each haplotype are assessed; while in cases where the HLA haplotypes are homozygous only that HLA haplotype is used for prediction. HLA haplotypes are cross-referenced with available HLA haplotypes prior to executing netMHCpan ([14]) for the primary neoantigen predictions. As with the primary tool, the user is able to specify the epitope lengths to conduct predictions for (typically epitopes of 8-, 9-, or 10-mers). The output from this process yields a single file containing either filtered or unfiltered (dependent on user options)

neoantigen predictions with information on the sample possessing the neoantigen and, in the case of multi-region variant calling, a presence/absence indicator for each of the sequenced regions. These predicted neoantigens are then, optionally, cross-referenced with normal peptides utilizing PeptideMatch ([15]), whereby the candidate epitopes are assessed for novelty against a reference proteome that can be supplied by the user as a fasta file (e.g. from Ensembl or UniProt). When available, users may also provide expression data as a tsv file specific to each sample (or a single reference file) to quickly assess expression levels of the gene carrying a predicted neoantigen. This information is included in the final output table.

The steps outlined above deliver candidate information for neoantigens from provided variant calls that may be presented to cytotoxic T-cells, however, this does not inform the likelihood of a neoantigen eliciting an immune response (i.e. being recognised by a TCR). In order to predict the recognition potential we employ the algorithms and process utilized by [16]. The recognition potential is defined as the product of A and R, where A is the amplitude of the ratio of the relative probabilities of binding for the wild-type and mutant epitopes to the MHC-class I molecules; and R is a measure of similarity to pathogenic peptides, meant to represent the probability that the neoantigen in question is recognised by a TCR clone already present in the tissue/blood. To define A it is necessary to perform neoantigen predictions for the wild-type and mutant epitope: this is not performed by default by NeoPredPipe, but is supplied as an option to employ as a contiguous pipeline. To define R, NeoPredPipe utilizes the multistate thermodynamic model employed by [16], which requires alignment scores for each epitope to a curated Immune Epitope Database list of known epitopes (can be refined and updated by the user, but is provided).

In order to incorporate the ability to assess ITH in regards to both effective mutations (non-synonymous variants and indels) and neoantigen burdens, NeoPredPipe is capable of handling multi-region VCF files; further these files can be multi-region in only a select number of samples and differ in the number of regions. Similarly, NeoPredPipe can process multi-region expression data for samples where information on regions are compiled into separate columns. Thus NeoPredPipe is able to efficiently handle various, potentially multi-region experimental designs for neoantigen prediction and assessments providing a summary table and an optional web-based visualization tool for downstream statistical and in-depth analysis.

Results
The output of the pipeline depends largely on the options set by the user, but at the very least, NeoPredPipe provides two tables of putative neoantigens and their predicted binding affinities, one for single nucleotide/amino acid, and one for indel(-type) variants. With additional options selected it is possible to include, within a single output, whether an epitope matches a reference proteome, its expression on the RNA level and the neoantigen's recognition potential. In additon, for rapid assessment, NeoPredPipe yields summary statistics on the neoantigen burden for each sample, a rapidly executed web-based visualization, as well as information to assess ITH by reporting neoantigen burdens for clonal, subclonal, and shared variants for multi-region samples. A detailed description of NeoPredPipe's output tables and each field in these can be found at https://github.com/MathOnco/NeoPredPipe.

Use Case
While a small, two sample, multi-region example dataset is provided with the source code for users, we demonstrate the usefulness of NeoPredPipe by applying it to a previously published dataset examining the evolutionary landscape of colorectal tumors [17]. We select two exemplary patient samples (Adenoma 3 and Carcinoma 7 in the original paper) from the dataset, and apply our pipeline using default parameters to evaluate neoantigens in each sample. Figure 2 illustrates the information included in the standard output of NeoPredPipe and potential analysis that can be performed if NeoPredPipe is combined with the output of other standard bioinformatic methods.

Figure 2a provides a summary of the complex interactions between different regions of Adenoma 3, and highlights both Region 4, which harbours the highest amount of subclonal (only present in a single region) neoantigens, and the overall clonality of the sample, with 72 neoantigens detected in all regions. For quick analysis, NeoPredPipe directly outputs a summary of the clonality of neoantigens, also divided into categories of strong and weak binders (peptides with a netMHCpan percentile rank $\leq$ 0.5 and $\leq$ 2, respectively, as recommended in [14]). Figure 2b visualizes this summary on two bar-charts for Adenoma 3 and Carcinoma 7. We find that whilst the number of shared neoantigens (present in more than one, but not all regions) is highly similar between the two samples, Carcinoma 7 harbours both more clonal (present in all regions) and subclonal neoantigens; and in total 26% of the neoantigens are clonal, compared to 16% of Adenoma 3. Figure 2c shows the recognition potential value for all neoantigens in the two samples. NeoPredPipe identified 10 peptides in Adenoma 3 and 9 in Carcinoma 7 with a recognition potential value above 1. In Fig. 2d, we provide an example of integrating NeoPredPipe outputs with downstream multi-region variant analysis. By inferring phylogenetic trees of each tumor, constructed using all exonic mutations with a variant

allele frequency above 0.05 (see [17] for full methods), we find that neoantigen distributions across regions can reflect the phylogenetic distance of regions and clonal structure of samples. 31% and 23.5% of total exonic mutations are clonal in Carcinoma 7 and Adenoma 3, similarly to the clonality of neoantigens shown in Panel B. This approach also highlights regions with neoantigen loads different from their closest neighbors, such as Region61 and Region62 of Carcinoma 7. Therefore the analysis can inform future experimental and bioinformatic investigations of samples allowing for new evolutionary and mechanistic insights into tumor development, evolution, and progression.

Conclusions

We present NeoPredPipe, an efficient, high-throughput, and user-friendly pipeline for neoantigen prediction and interrogation for single and multi-region tumor VCF files. By tying together commonly utilized bioinformatics toolsets and integrating recent advances in neoantigen assessment, NeoPredPipe yields concise information typically required by researchers and clinicians. Through user options, based on the individuals own computational

limitations, the pipeline is scalable for a high performance computing (HPC) cluster environment and customizable for individual research questions. Furthermore, unlike existing methods[7–10], NeoPredPipe can process a directory containing numerous samples in a single command; therefore provides a user-friendly way for not computer-proficient users to analyse the output of large studies or compare against reference datasets. All source code and an extensive read me for each component of NeoPredPipe with all pipeline options are available at https://github.com/MathOnco/NeoPredPipe.

Availability and requirements

Project name: NeoPredPipe
  Project home page: https://github.com/MathOnco/NeoPredPipe
  Operating system: Unix-based operating system
  Programming languages: Python and Bash
  Other requirements: Python 2.7, ANNOVAR, netMHCpan, PeptideMatch, and (optionally) NCBI BlastX+.
  License: GNU GPLv3
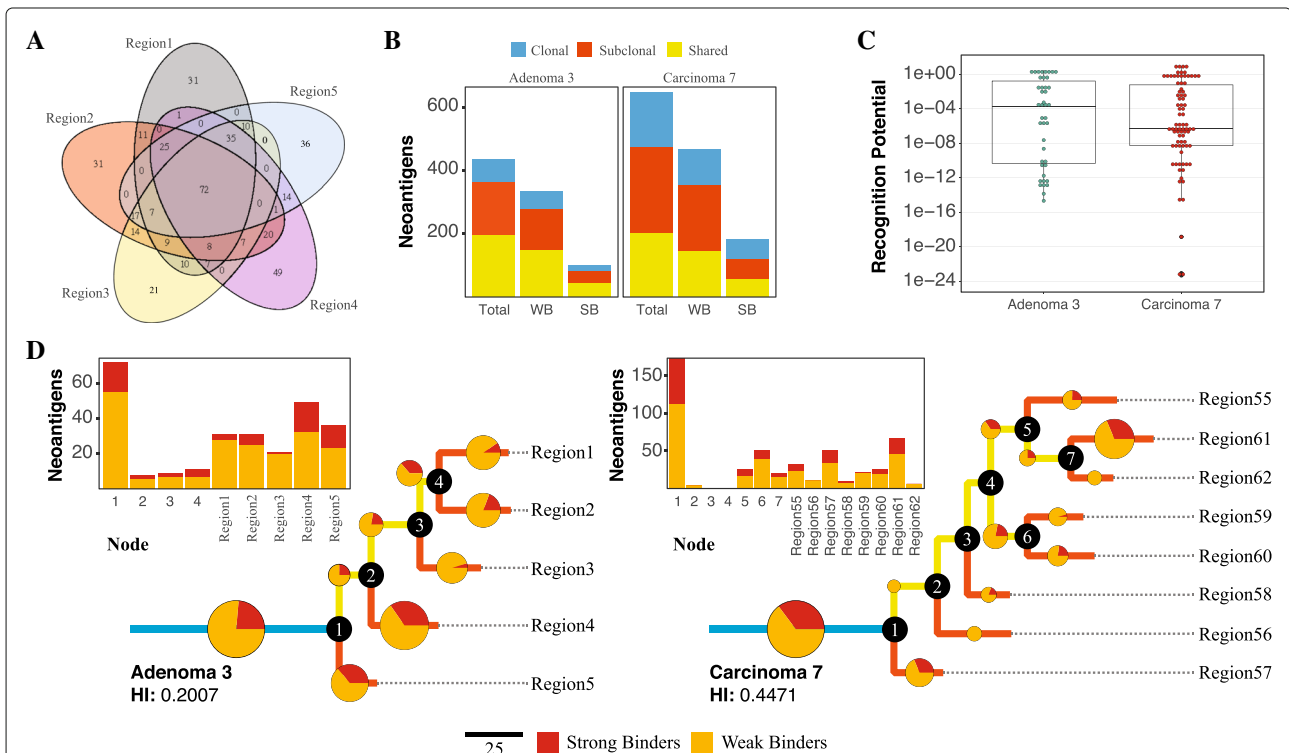 **Any restrictions to use by non-academics:** None

Fig. 2 Analysis of neoantigens in two colorectal tumors using NeoPredPipe. **a** Venn diagram of all neoantigens in the five regions of Adenoma 3. **b** Number of neoantigens in the two samples that are clonal (present in all regions, shown in blue), shared (present in at least two regions, in yellow) or subclonal (present in a single region, red). Separate counts of weak and strong MHC-binding neoantigens (WB and SB, respectively) are also shown. **c** Distribution of recognition potential values of neoantigens present in Adenoma 3 (green) and Carcinoma 7 (red). The boxplots represent the median and upper and lower 25 percentile. Only neoantigens with recognition potential higher than zero are shown. **d** Phylogenetic tree reconstructed from all exonic mutations for Adenoma 3 (left) and Carcinoma 7 (right). Pie-charts and the bar-charts represent the number of weak (orange) and strong (red) binder neoantigens assigned to each branch. The size of each circle is proportional to the percentage of total neoantigens on that branch

Schenck *et al. BMC Bioinformatics*   (2019) 20:264

Page 5 of 6

Availability of data and materials

All raw data (bam files, and processed data including vcf files) are deposited in the EGA archive under accession number EGAS00001003066.

Authors' contributions

ROS conceived NeoPredPipe, wrote all scripts, and prepared the manuscript. EL co-wrote the final code base, led debugging efforts, conceived the use case example and wrote the revised manuscript. CG provided insights into NeoPredPipe's necessary outputs. TAG and ARAA provided guidance on code development and oversaw all work efforts. All authors read, edited, and approved the manuscript.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details

[1]Integrated Mathematical Oncology, Moffitt Cancer Center, 33612 Tampa, FL, USA. [2]Wellcome Centre for Human Genetics, University of Oxford, OX3 7BN Oxford, UK. [3]Evolution and Cancer Laboratory, Barts Cancer Institute, Queen Mary University of London, EC1M London, UK.

References

1. McGranahan N, Swanton C. Clonal heterogeneity and tumor evolution: Past, present, and the future. Cell. 2017;168(4):613–628. 2018/08/27.
2. Schumacher TN, Schreiber RD. Neoantigens in cancer immunotherapy. Science. 2015;348(6230):69–74.
3. McGranahan N, Furness AJS, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, Jamal-Hanjani M, Wilson GA, Birkbak NJ, Hiley CT, Watkins TBK, Shafi S, Murugaesu N, Mitter R, Akarca AU, Linares J, Marafioti T, Henry JY, Allen EMV, Miao D, Schilling B, Schadendorf D, Garraway LA, Makarov V, Rizvi NA, Snyder A, Hellmann MD, Wolchok JD, Shukla SA, Wu CJ, Peggs KS, Chan TA, Hadrup SR, Quezada SA, Swanton C. Clonal neoantigens elicit t cell immunoreactivity and sensitivity to immune checkpoint blockade. Science. 2016;351(6280):1463–9.
4. Lakatos E, Williams MJ, Schenck RO, Cross WCH, Househam J, Werner B, Gatenbee C, Robertson-Tessi M, Barnes CP, Anderson ARA, Sottoriva A, Graham TA. Evolutionary dynamics of neoantigens in growing tumours. bioRxiv. 2019536433.
5. McGranahan N, Rosenthal R, Hiley CT, Rowan AJ, Watkins TBK, Wilson GA, Birkbak NJ, Veeriah S, Loo PV, Herrero J, Swanton C, Jamal-Hanjani M, Shafi S, Czyzewska-Khan J, Johnson D, Laycock J, Bosshard-Carter L, Gorman P, Hynds RE, Wilson G, Birkbak NJ, Watkins TBK, Horswell S, Mitter R, Escudero M, Stewart A, Rowan A, Xu H, Turajlic S, Hiley C, Abbosh C, Goldman J, Stone RK, Denner T, Matthews N, Elgar G, Ward S,

Costa M, Begum S, Phillimore B, Chambers T, Nye E, Graca S, Bakir MA, Joshi K, Furness A, Aissa AB, Wong YNS, Georgiou A, Quezada S, Hartley JA, Lowe HL, Lawrence D, Hayward M, Panagiotopoulos N, Kolvekar S, Falzon M, Borg E, Marafioti T, Simeon C, Hector G, Smith A, Aranda M, Novelli M, Oukrif D, Janes SM, Thakrar R, Forster M, Ahmad T, Lee SM, Papadatos-Pastos D, Carnell D, Mendes R, George J, Navani N, Ahmed A, Taylor M, Choudhary J, Summers Y, Califano R, Taylor P, Shah R, Krysiak P, Rammohan K, Fontaine E, Booton R, Evison M, Crosbie P, Moss S, Idries F, Joseph L, Bishop P, Chaturved A, Quinn AM, Doran H, Leek A, Harrison P, Moore K, Waddington R, Novasio J, Blackhall F, Rogan J, Smith E, Dive C, Tugwood J, Brady G, Rothwell DG, Chemi F, Pierce J, Gulati S, Naidu B, Langman G, Trotter S, Bellamy M, Bancroft H, Kerr A, Kadiri S, Webb J, Middleton G, Djearaman M, Fennell D, Shaw JA, Quesne JL, Moore D, Nakas A, Rathinam S, Monteiro W, Marshall H, Nelson L, Bennett J, Riley J, Primrose L, Martinson L, Anand G, Khan S, Amadi A, Nicolson M, Kerr K, Palmer S, Remmen H, Miller J, Buchan K, Chetty M, Gomersall L, Lester J, Edwards A, Morgan F, Adams H, Davies H, Kornaszewska M, Attanoos R, Lock S, Verjee A, MacKenzie M, Wilcox M, Bell H, Hackshaw A, Ngai Y, Smith S, Gower N, Ottensmeier C, Chee S, Johnson B, Alzetani A, Shaw E, Lim E, De Sousa P, Barbosa MT, Bowman A, Jordan S, Rice A, Raubenheimer H, Proli C, Cufari ME, Ronquillo JC, Kwayie A, Bhayani H, Hamilton M, Bakar Y, Mensah N, Ambrose L, Devaraj A, Buderi S, Finch J, Azcarate L, Chavan H, Green S, Mashinga H, Nicholson AG, Lau K, Sheaff M, Schmid P, Conibear J, Ezhil V, Ismail B, Irvin-sellers M, Prakash V, Russell P, Light T, Horey T, Danson S, Bury J, Edwards J, Hill J, Matthews S, Kitsanta Y, Suvarna K, Fisher P, Keerio AD, Shackcloth M, Gosney J, Postmus P, Feeney S, Asante-Siaw J, Aerts HJWL, Dentro S, Dessimoz C. Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. Cell. 2017;171(6):1259–1271.e11.

6. Rosenthal Rachel, Cadieux EL, Salgado R, Bakir MA, Moore DA, Hiley CT, Lund T, Tanić M, Reading JL, Joshi K, Henry JY, Ghorani E, Wilson GA, Birkbak NJ, Jamal-Hanjani M, Veeriah S, Szallasi Z, Loi S, Hellmann MD, Feber A, Chain B, Herrero J, Quezada SA, Demeulemeester J, Loo PV, Beck S, McGranahan N, Swanton C, Czyzewska-Khan J, Johnson D, Laycock J, Gorman P, Hynds RE, Wilson G, Birkbak NJ, Watkins TBK, Escudero M, Stewart A, Rowan A, Hiley C, Abbosh C, Goldman J, Stone RK, Denner T, Ward S, Nye E, Aissa AB, Wong YNS, Georgiou A, Quezada S, Hartley JA, Lowe HL, Lawrence D, Hayward M, Panagiotopoulos N, Falzon M, Borg E, Marafioti T, Janes SM, Forster M, Ahmad T, Lee SM, Papadatos-Pastos D, Carnell D, Mendes R, George J, Ahmed A, Taylor M, Choudhary J, Summers Y, Califano R, Taylor P, Shah R, Krysiak P, Rammohan K, Fontaine E, Booton R, Evison M, Crosbie P, Moss S, Joseph L, Bishop P, Quinn AM, Doran H, Leek A, Harrison P, Moore K, Waddington R, Novasio J, Blackhall F, Rogan J, Smith E, Dive C, Tugwood J, Brady G, Rothwell DG, Pierce J, Gulati S, Naidu B, Langman G, Trotter S, Bancroft H, Kerr A, Kadiri S, Middleton G, Djearaman M, Fennell D, Shaw JA, Quesne JL, Moore DA, Nakas A, Rathinam S, Monteiro W, Marshall H, Nelson L, Riley J, Primrose L, Martinson L, Anand G, Khan S, Nicolson M, Kerr K, Palmer S, Remmen H, Miller J, Buchan K, Chetty M, Gomersall L, Lester J, Morgan F, Adams H, Davies H, Kornaszewska M, Attanoos R, Lock S, MacKenzie M, Wilcox M, Bell H, Hackshaw A, Ngai Y, Smith S, Gower N, Ottensmeier C, Chee S, Johnson B, Alzetani A, Shaw E, Lim E, De Sousa P, Barbosa MT, Bowman A, Jordan S, Rice A, Raubenheimer H, Bhayani H, Hamilton M, Mensah N, Ambrose L, Devaraj A, Chavan H, Nicholson AG, Lau K, Sheaff M, Schmid P, Conibear J, Ezhil V, Prakash V, Russell P, Light T, Horey T, Danson S, Bury J, Edwards J, Hill J, Matthews S, Kitsanta Y, Suvarna K, Fisher P, Shackcloth M, Gosney J, Feeney S, Asante-Siaw J, Ryanna K, Dawson A, Tuffail M, Bajaj A, Brozik J, Walter H, Carey N, Price G, Gilbert K, Webb J, Patel A, Chaturvedi A, Granato F, Baker K, Carter M, Priest L, Krebs MG, Lindsay C, Gomes F, Chemie F, George R, Patrini D, Khiroya R, Shaw P, Skrzypski M, Sunderland MW, Reading JL, Beastall C, Mangal N, Peggs K, Lim E, Al-Bakir M, Navani N, Scarci M, Ensell L, Biswas D, Razaq M, Nicod J, Lopez S, Huebner A, Dietzen M, Mourikis T, Adefila-Ideozu T, Begum S, Klein H, Mani A, Carvalho S, Kaniu D, Realingo C, Malima M, Booth S, Lim L, Rao J, Tenconi S, Socci J, Kibutu F, Agyemang M, Young R, Blyth KG, Dick C, Kirk A, Kidd A, The TRACERx consortium. Neoantigen-directed immune escape in lung cancer evolution. Nature. 2019;567:479–485.

7. Bais P, Namburi S, Zhang X, Chuang JH, Gatti DM. CloudNeo: a cloud pipeline for identifying patient-specific tumor neoantigens. Bioinformatics. 2017;33(19):3110–2.

8. Bjerregaard AM, Nielsen M, Hadrup SR, Szallasi Z, Eklund AC. Mupexi: prediction of neo-epitopes from tumor sequencing data. Cancer Immunol Immunother. 2017;66(9):1123–30.

9. Hundal J, Carreno BM, Petti AA, Linette GP, Griffith OL, Mardis ER, Griffith M. pvac-seq: A genome-guided in silico approach to identifying tumor neoantigens. Genome Med. 2016;8(1):11.

10. Zhou Z, Lyu X, Wu J, Yang X, Wu S, Zhou J, Gu X, Su Z, Chen S. Tsnad: an integrated software for cancer somatic mutation and tumour-specific neoantigen detection. Roy Soc Open Sc. 2017;4(4):170050.

11. Lancaster EM, Jablons D, Kratz JR. Applications of next-generation sequencing in neoantigen prediction and cancer vaccine development. Genet Test Mol Biomarkers. 2019;00:1–8. 2019/04/04.

12. Wang K, Li M, Hakonarson H. Annovar: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.

13. Shukla SA, Rooney MS, Rajasagi M, Tiao G, Dixon PM, Lawrence MS, Stevens J, Lane WJ, Dellagatta JL, Steelman S, Sougnez C, Cibulskis K, Kiezun A, Brusic V, Wu CJ, Getz G. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. Nat Biotechnol. 2015;33(11):1152–8.

14. Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Morten N. Netmhcpan-4.0: Improved peptide–mhc class i interaction predictions integrating eluted ligand and peptide binding affinity data. J Immunol. 2017;199(9):3360–8.

15. Chen C, Li Z, Huang H, Suzek BE, Wu CH, UniProt Consortium. A fast peptide match service for uniprot knowledgebase. Bioinformatics. 2013;29(21):2808–9.

16. Łuksza M, Riaz N, Makarov V, Balachandran VP, Hellmann MD, Solovyov A, Rizvi NA, Merghoub T, Levine AJ, Chan TA, Wolchok JD, Greenbaum BD. A neoantigen fitness model predicts tumour response to checkpoint blockade immunotherapy. Nature. 2017;551:517–20. 11.

17. Cross W, Kovac M, Mustonen V, Temko D, Davis H, Baker AM, Biswas S, Arnold R, Chegwidden L, Gatenbee C, Anderson AR, Koelzer VH, Martinez P, Jiang X, Domingo E, Woodcock DJ, Feng Y, Kovacova M, Maughan T, Adams R, Bach S, Beggs A, Brown L, Buffa F, Cazier JB, Blake A, Wu C-H, Chatzpili E, Richman S, Dunne P, Harkin P, Higgins G, Hill J, Holmes C, Horgan D, Kaplan R, Kennedy R, Lawler M, Leedham S, McDermott U, McKenna G, Middleton G, Morton D, Murray G, Quirke P, Salto-Tellez M, Samuel L, Schuh A, Sebag-Montefiore D, Seymour M, Sharma R, Sullivan R, Tomlinson I, West N, Wilson R, Jansen M, Rodriguez-Justo M, Ashraf S, Guy R, Cunningham C, East JE, Wedge DC, Wang LM, Palles C, Heinimann K, Sottoriva A, Leedham SJ, Graham TA, Tomlinson IPM, The S:CORT Consortium. The evolutionary landscape of colorectal tumorigenesis. Nat Ecol Evol. 2018;2(10):1661–72.