



Computational Pipeline for the PGV-001 Neoantigen Vaccine Trial

Alex Rubinsteyn¹, Julia Kodysh¹, Isaac Hodes¹, Sebastien Mondet¹, Bulent Arman Aksoy^{1,2}, John P. Finnigan^{1,3}, Nina Bhardwaj^{1,3} and Jeffrey Hammerbacher^{1,2*}

¹Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, United States,

²Department of Microbiology and Immunology, Medical University of South Carolina, Charleston, SC, United States,

³Icahn School of Medicine at Mount Sinai, Tisch Cancer Institute, New York, NY, United States

OPEN ACCESS

Edited by:

Mustafa Diken,
Translationale Onkologie an der
Universitätsmedizin der Johannes
Gutenberg-Universität Mainz,
Germany

Reviewed by:

Barbara Schrörs,
Translationale Onkologie an der
Universitätsmedizin der Johannes
Gutenberg-Universität Mainz,
Germany
Sine Reker Hadrup,
Technical University of Denmark,
Denmark

*Correspondence:

Jeffrey Hammerbacher
correspondence@hammerlab.org

Specialty section:

This article was submitted to
Cancer Immunity and
Immunotherapy,
a section of the journal
Frontiers in Immunology

Received: 06 August 2017

Accepted: 30 November 2017

Published: 18 January 2018

Citation:

Rubinsteyn A, Kodysh J, Hodes I,
Mondet S, Aksoy BA, Finnigan JP,
Bhardwaj N and Hammerbacher J
(2018) Computational Pipeline for the
PGV-001 Neoantigen Vaccine Trial.
Front. Immunol. 8:1807.
doi: 10.3389/fimmu.2017.01807

This paper describes the sequencing protocol and computational pipeline for the PGV-001 personalized vaccine trial. PGV-001 is a therapeutic peptide vaccine targeting neoantigens identified from patient tumor samples. Peptides are selected by a computational pipeline that identifies mutations from tumor/normal exome sequencing and ranks mutant sequences by a combination of predicted Class I MHC affinity and abundance estimated from tumor RNA. The personalized genomic vaccine (PGV) pipeline is modular and consists of independently usable tools and software libraries. We hope that the functionality of these tools may extend beyond the specifics of the PGV-001 trial and enable other research groups in their own neoantigen investigations.

Keywords: neoantigens, personalized vaccine, immunoinformatics, genomics, computational pipeline

INTRODUCTION

Cancer neoantigens are antigens presented on tumor cells which, due to either mutation or modification, are not presented on normal cells. Neoantigens generated by tumor DNA mutations have been shown to play a significant role in mediating the destruction of tumor cells by the adaptive immune system (1–3). Several groups have used therapeutic vaccines targeting neoantigens to clear tumors in murine models (4–6). Consequently, many human neoantigen vaccine trials are now under way and several have published promising early results (7, 8). Since very few cancer mutations are recurrent between patients, the identification of neoantigens requires a personalized genomic approach (9). We describe the sequencing protocol and immunogenomic pipeline of PGV-001, a neoantigen vaccine trial at the Mount Sinai Hospital (10).

The personalized genomic vaccine (PGV) computational pipeline takes tumor/normal sequencing data as an input and generates a ranked list of mutated peptide sequences. The steps along the way of determining a personalized vaccine's contents are implemented as configurable independent tools.

OVERVIEW OF THE PGV-001 PERSONALIZED VACCINE TRIAL

PGV-001 is a phase I clinical trial at Mount Sinai Hospital, studying the safety and immunogenicity of a multi-peptide personalized genomic vaccine for the treatment of cancers. A PGV dose consists of 10 synthetic long peptides (11), each containing a somatic mutation from the patient's tumor, as well as an immunostimulatory adjuvant: poly-ICLC (12). In the PGV-001 trial, the personalized vaccine is administered in the adjuvant setting, for patients who undergo a complete resection and have no evidence of residual disease.

When a new patient enrolls in the trial, their tumor and normal samples are collected and processed to isolate and sequence DNA and RNA. The computational pipeline of PGV-001 is then used to select the peptide contents of the vaccine. The major steps between surgery and vaccination are shown in **Figure 1**, whereas details of the computational pipeline are shown in **Figure 2**.

The candidate vaccine peptides generated by the computational pipeline are ranked by abundance and predicted MHC

affinity, which both contribute to immunogenicity. The manufacturer attempts to synthesize the top 15 ranked candidate peptide sequences and delivers 10 lyophilized peptides which they are able to purify to sufficient quality and quantity. The peptides are dissolved in DMSO and mixed with poly-ICLC immediately before use. The personalized vaccine is administered as an intracutaneous injection and is given to the patient 10 times over a span of 6 months.

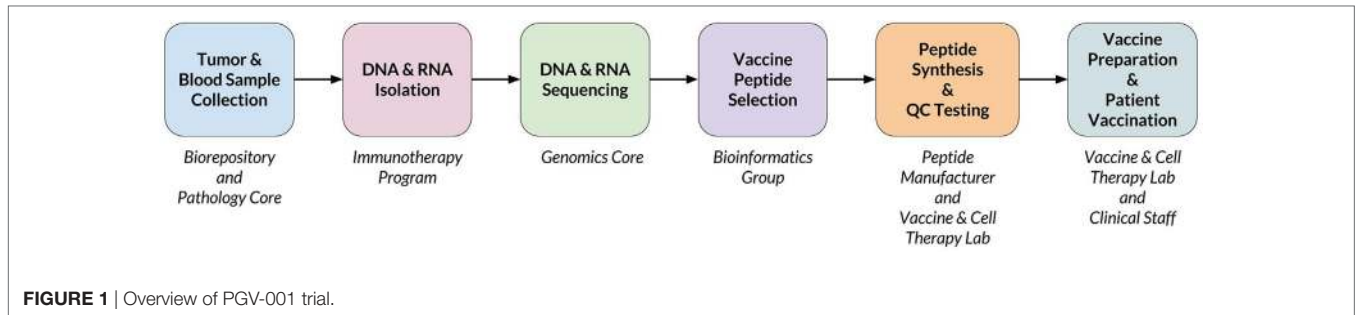


FIGURE 1 | Overview of PGV-001 trial.

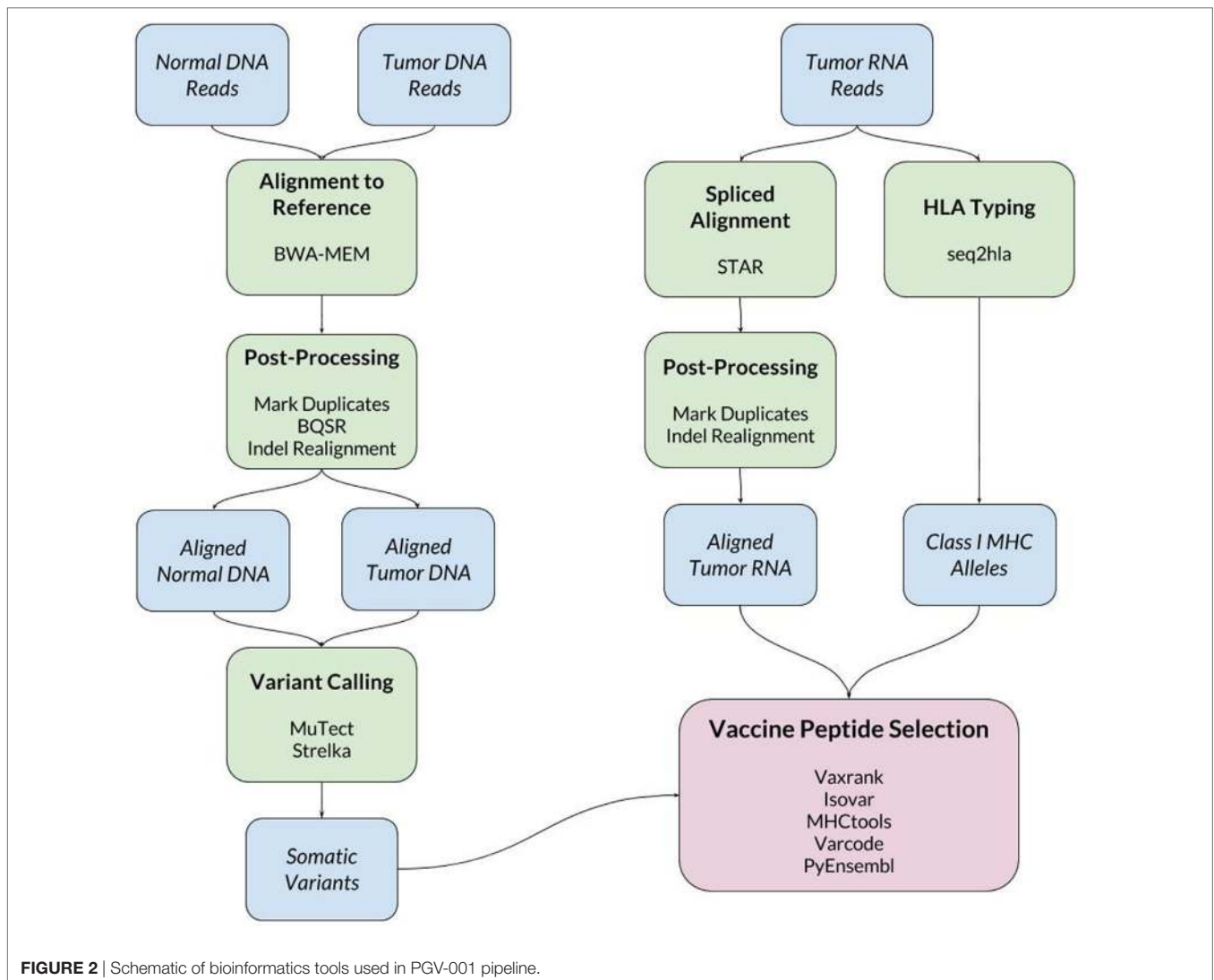


FIGURE 2 | Schematic of bioinformatics tools used in PGV-001 pipeline.

SEQUENCING PROTOCOL FOR DNA AND RNA

The sequencing protocols used for both DNA and RNA can dramatically affect the sensitivity of variant detection, and thus ultimately change the results of the vaccine pipeline. The largest determinants of sensitivity are the sample quality, method of sequencing library preparation, and quantity of sequenced reads. Whenever possible, PGV uses fresh frozen tumor tissue samples, which results in significantly improved variant detection accuracy as compared with sequencing of formalin-fixed (FFPE) samples (13). An additional benefit of using fresh frozen samples is that mRNA can be enriched using poly-A capture, whereas the fragmented RNA of FFPE samples can only be prepared with less efficient methods such as ribosomal depletion (14). For patients with solid tumors, normal DNA is extracted from peripheral blood rather than potentially contaminated adjacent tissue (15).

Fragmentation by sonication was preferred to transposase-based methods (16) due to significant sequence bias, leading to lost coverage after marking duplicate reads (17). Among the exome enrichment techniques which use sonication, we chose Agilent's SureSelect XT kit due to its efficient rate of capturing on-target reads (18).

We chose to target 150× mean coverage for the normal DNA (exome) sequencing since this was found to be the point of diminishing sensitivity across different variant calling pipelines (19). Several of the cancer types allowed in the PGV-001 trial (particularly lung and head/neck cancers) have been shown to result in systematically low purity samples (20). To accommodate a significant degree of non-cancerous DNA, we assume 50% tumor purity and consequently target 300× exome coverage for the tumor DNA sample.

A final consideration is the choice of read length, which does not significantly impact variant discovery from DNA but does impact variant phasing in RNA. Since a 25mer vaccine peptide is translated from 75 bp of coding sequence, PGV could theoretically use any read length above that minimum. To allow for many distinct aligned positions overlapping the same region of coding sequence, the PGV protocol uses 125-bp reads. These provide a good compromise between length and base quality on the HiSeq 2500 instrument.

OVERVIEW OF THE COMPUTATIONAL PIPELINE

The inputs to the computational pipeline are unmapped sequencing data from tumor DNA, tumor RNA, and normal patient DNA. The tumor and normal DNA samples are aligned against the human GRCh37 reference genome using BWA-MEM (21). The tumor RNA is aligned using STAR (22), which has been found to have particularly high sensitivity for detecting indel variants (23). Both DNA and RNA alignment files are processed according to GATK Best Practices (24). One noteworthy deviation from the standard GATK pipeline is our use of assembly based indel realignment on tumor RNA data (in addition to the DNA

samples). This is done to improve the sensitivity of detecting RNA reads which support indel somatic variants.

Somatic Variant Calling

Somatic variant calling is performed using MuTect (25) and Strelka (26), whose results are combined by taking a union of called variants. In cases where the final pipeline yields an insufficient number of vaccine peptides (fewer than 15), we rerun the pipeline including MuTect2 among the set of variant callers to increase sensitivity.

HLA Typing

To make predictions about epitope presentation to T-cells, it is necessary to know the patient's HLA type. This can be determined computationally from exome or bulk RNA sequencing or validated externally using HLA-specific amplicon sequencing (27). The PGV pipeline currently uses seq2hla (28) for HLA typing from tumor RNA while also using amplicon sequencing of normal DNA for validation. Across 10 patients, the two methods have only disagreed on a single allele, where HLA-C*07:02 was mistyped as HLA-C*07:01. This high degree of concordance matches our previous experience with HLA typing of fresh frozen tissue samples; formalin-fixed tissue is more likely to give discordant results between different sequencing methods.

Vaccine Peptide Selection

The bulk of the custom software developed for this trial is related to vaccine peptide selection. The results of the above steps are a set of somatic variants, aligned tumor RNA reads, and the patient's HLA type. These data are then used to determine mutant protein sequences, estimate mutation abundance, predict MHC ligands overlapping mutations, and finally to generate a ranked list of candidate vaccine peptides.

Some of the tools used in vaccine peptide selection include:

- *Vaxrank* (29): overall vaccine selection tool with ranking logic.
- *Isovar* (30): determines mutant protein sequence from somatic variants and tumor RNA.
- *Varcode* (31): predicts variant effects for filtering out silent mutations.
- *PyEnsembl* (32): provides reference genome annotations that are used by Varcode to determine exon boundaries and transcript sequences.
- *MHCtools* (33): common interface to peptide-MHC-binding predictors.

Due to their importance, Isovar and Vaxrank are both described in greater detail in the following two sections.

ISOVAR: DETERMINING THE MUTANT PROTEIN SEQUENCE

There are several different software packages that predict the protein-level effect of a coding mutation (31, 34, 35). However, for the purposes of selecting a vaccine peptide's sequence, it is not sufficient to predict a DNA mutation's protein effect without considering the transcripts in which it occurs. A

somatic mutation can be associated with selective splicing of particular RNA isoforms (36) and can also cooccur with other genomic variants. Thus, in the PGV pipeline, the tumor RNA sequencing data are also used to determine a mutant coding sequence.

For each mutation, it is possible to infer multiple coding sequences from supporting RNA reads due to sequencing error, splicing diversity, and tumor heterogeneity. To account for these potentially complicating factors, we developed a tool called Isovar (30), which can be downloaded from <https://github.com/hammerlab/isovar>. Isovar uses RNA to assemble the most abundant coding sequence for each mutation. An overview of the algorithm is given in **Figure 3**.

One advantage of using RNA to determine the coding sequence is that it phases adjacent (germline or somatic) variants. Examples of the impact of adjacent variants on a coding sequence are shown in **Figures 4** and **5**. A further advantage is that Isovar, by using mutation-supporting RNA reads to determine each mutant protein sequence, naturally estimates allele-specific expression. If the PGV pipeline, on the other hand, used bulk expression it would potentially overestimate how much of a mutant protein is being made. In an extreme case, all of the RNA reads aligning to

a particular gene could be wild type, with none supporting the somatic variant of interest.

VAXRANK: VACCINE PEPTIDE SELECTION

Once we have determined the amino acid sequences containing somatic mutations and estimated their abundance in the tumor, the final step is to rank them according to desirability of inclusion in personalized vaccine.

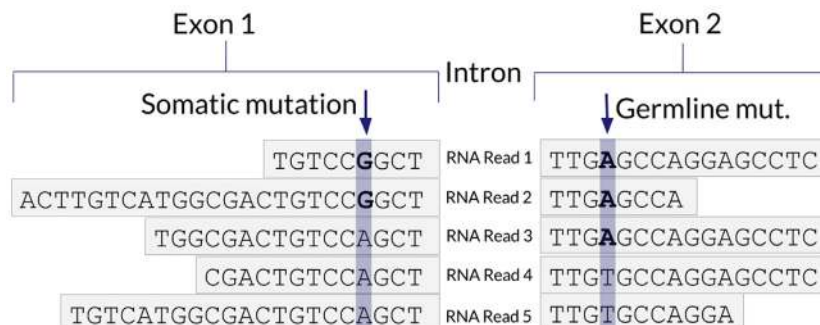
There are many potential correlates of immunogenicity that can be used to prioritize neoantigens, such as expression, MHC-binding affinity, peptide-MHC complex stability, proteasomal cleavage, and other antigen-processing steps. Of those, the PGV pipeline optimizes for high expression and predicted strong Class I MHC binding. There are several published computational predictors of Class I MHC-binding affinity which have demonstrated high accuracy (37–39). PGV uses NetMHCpan (37) due to its extensive coverage of patient alleles.

The final ranking of candidate vaccine peptides according to predicted MHC binding and expression is performed by a tool called Vaxrank (29). Vaxrank identifies high-affinity mutant MHC ligands within each peptide and combines these predictions into a single MHC-binding score. This score is then scaled according to the expression of that mutation in the tumor. The formula for computing these MHC and expression scores is given in **Figure 6**. The scale and offset for MHC affinity normalization was determined by a logistic fit of affinity versus immunogenicity from the dataset used to determine the classical 500-nM affinity threshold (40). There is no rigorous justification for the multiplicative scoring function, other than the intuition that epitope abundance and MHC affinity are independent prerequisites for immunogenicity.

Since some peptides cannot be manufactured using solid-phase synthesis, our vaccine peptide selection algorithm includes manufacturability heuristics, such as minimization of cysteine content.

```
function MUTANTPROTEINSEQUENCES(v)
  Reads(v) = tumor RNA reads which contain somatic variant v
  Seqs(v) = longer sequences assembled from Reads(v)
  Transcripts(v) = reference transcripts overlapping variant v
  for s in Seqs(v) do
    for t in Transcripts(v) do
      if s matches t then
        i = offset to first codon in matching portion of t
        p = translate si into amino acid sequence
        yield p
```

FIGURE 3 | Overview of Isovar algorithm for determining mutant protein sequences.



Selected coding sequence includes germline mutation:

GGCGACTGTCCGGCTTTGAGCCAGGTGCCTC

FIGURE 4 | Schematic representation of a somatic mutation co-occurring with a germline mutation.

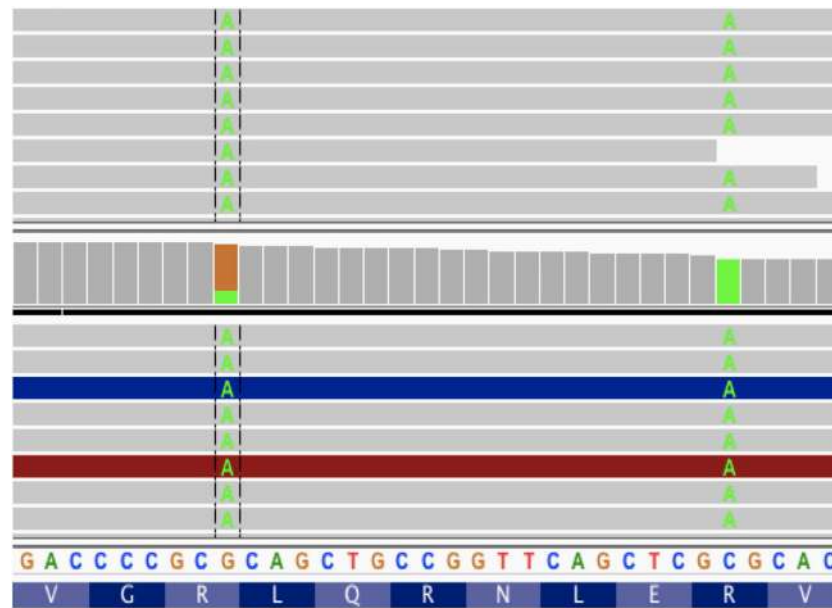


FIGURE 5 | Screenshot from IGV with tumor DNA on top and tumor RNA on bottom. The two somatic variants from patient data 7 amino acids apart. If these mutations were considered without phasing, we would get two different vaccine peptides, neither of which would match the protein sequence produced by tumor cells.

$$\text{TotalScore} = \text{ExpressionScore} \cdot \text{BindingScore}$$

$$\text{ExpressionScore} = \sqrt{\# \text{ supporting reads}}$$

$$\text{BindingScore} = \sum_p \sum_{mhc} \sigma(\text{IC50}(p, mhc))$$

$$\sigma(x) = \exp\left(-\frac{x - 150}{350}\right)$$

FIGURE 6 | TotalScore used to rank somatic variants in a way that attempts to balance predict MHC binding and abundance. ExpressionScore uses read count (instead of a normalized measure like FPKM) since these scoring criteria are not meant to be compared between patient samples. BindingScore sums normalized binding affinities of mutant peptides across all patient alleles and lengths between 8 and 11.

Vaxrank can be downloaded from <https://github.com/hammerlab/vaxrank>.

EPIDISCO: PARALLEL IMPLEMENTATION OF THE PGV PIPELINE

The workflow management tool that orchestrates the execution of the PGV pipeline is called Epidisco (41). Epidisco is used to set up a local or cluster compute environment, install all relevant bioinformatics tools (external software such as GATK, as well as our own tools including Isovar and Vaxrank), and coordinate the execution of these different tools on the input data.

Epidisco accelerates portions of the genomics pipeline on two levels. Independent computational tasks such as QC checks, the

processing of the RNA-sequencing data, and the joint analyses of the normal-tumor DNA-sequencing data are all run in parallel. Within the invocation of each tool, when possible, sequencing data are split into multiple segments, partial results are computed in parallel and then merged.

Epidisco supports local computation, traditional HPC schedulers such as LSF, and cloud-based resources from Google Cloud and AWS. On a typical machine, running the complete PGV pipeline for a single patient can take up to 4 days; but making use of five or more computers for parallelization reduces the overall running time down to a single day.

Epidisco also makes the PGV pipeline tolerant to failures of intermediate steps and allows resuming the pipeline from the point of failure with a simple restart request. By handling such failures in an automated way, carrying out cleaning procedures, and restarting only the tasks that need to be rerun, the workflow makes it easier for researchers to operate such complex computational tasks. Epidisco provides command line and web-based utilities to facilitate starting a new workflow, collecting the results, and troubleshooting specific parts of a pipeline.

The individual infrastructure tools used by the PGV pipeline are implemented as an OCaml stack and include:

- *Ketrew*: custom workflow manager who handles dependency management, parallelization, and smart restarts of failed tasks.
- *Biokepi*: wraps bioinformatics tools so they can be used with Ketrew and statically ensures the absence of common mistakes during pipeline construction.
- *Secotrec*: cluster management tool that allows deployment on cloud services such as the Google Cloud Platform and Amazon Web Services.
- *Epidisco*: the actual implementation of the PGV-001 pipeline.

DISCUSSION

The PGV pipeline is a modular, highly configurable, freely available method for selecting the contents of a therapeutic neoantigen vaccine. The PGV pipeline has been used to predict vaccine peptides for several mouse models (LLC, B16 F1/F10), five “dry run” patients whose samples were processed according to the PGV protocol but did not participate in the trial, and five patients who were being considered for enrollment in the trial. Of the patients eligible for enrollment, one has been treated so far and another enrolled. The remainder did not enroll due to progression of disease or low-quality tumor samples.

Several other groups have released pipelines for neoantigen vaccine prediction, most notably pVAC-seq (42) and MuPeXI (43). A deep comparison between neoantigen pipelines likely requires evaluating T-cell response and antitumor activity after vaccination, which is beyond the scope of this paper. There are, however, a few obvious differences between the PGV pipeline and others which have been published:

- **Modularity:** the PGV pipeline has been developed as a collection of flexible standalone tools, rather than a single monolithic script. These tools can be repurposed for other immunogenomics analyses and have already been used for retrospective analyses of checkpoint blockade clinical trials (44).
- **Inputs are FASTQ files:** MuPeXI and pVAC-seq both require the implementation of separate genomics pipeline to infer patient HLA type, call somatic variants, and quantify expression. The PGV pipeline, by contrast, is self-contained in the sense that its inputs are raw FASTQ files and its outputs are vaccine peptide predictions.
- **Dependence on tumor RNA:** the PGV pipeline relies on tumor RNA reads to determine the mutant protein coding sequence. MuPeXI and pVAC-seq, by contrast, only consider expression data after predicting a mutant protein sequence from a variant in isolation. PGV’s approach has potential benefits in capturing altered patterns of splicing and phasing somatic variants with other nearby variants. These potential benefits, however, have yet to be evaluated systematically.
- **Liberal software license:** all of the software components that comprise the PGV pipeline are freely available under the Apache software license. MuPeXI does not yet appear to have a fully open source license, while pVAC-seq uses the more restrictive non-profit open software license.
- **Optimization of peptide sequence for solid-phase synthesis:** PGV appears to be unique among freely available neoantigen pipelines in attempting to choose peptides whose sequence content is more likely to be successfully manufactured. We have found this to be an important step, especially when using longer peptides, due to the significant delays introduced by failed synthesis or purification attempts.

The PGV-001 trial is the first in a series of planned neoantigen vaccine investigations. Several improvements to the PGV pipeline are planned, including the use of genomic fusions and other structural variants as neoantigen sources, clonality as a consideration for variant prioritization, and additional immunological predictions such as proteasomal cleavage and Class II MHC binding. As immune response data from ongoing preclinical work and PGV-001 becomes available, our method for combining correlates of immunogenicity into a single ranking will require extensive evaluation.

AUTHOR CONTRIBUTIONS

The first draft of this paper was originally written by IH and substantially revised by AR and JK. Genomics and immunoinformatics tools were implemented by AR, JK, and other lab members. The workflow libraries were written mostly by SM. The actual pipeline (connecting all the individual tools and running them in parallel) was written mostly by IH with help from BA. The overall design of the software was determined through extensive conversation with JH. The PGV trial was designed by NB, JF, AR, and several others in the Bhardwaj lab.

FUNDING

This work is partially funded by the Parker Institute for Cancer Immunotherapy.

REFERENCES

1. Fritsch EF, Rajasagi M, Ott PA, Brusic V, Hacohen N, Wu CJ. HLA-binding properties of tumor neoepitopes in humans. *Cancer Immunol Res* (2014) 2:522–9. doi:10.1158/2326-6066.CIR-13-0227
2. Finnigan JP Jr, Rubinsteyn A, Hammerbacher J, Bhardwaj N. Mutation-derived tumor antigens: novel targets in cancer immunotherapy. *Oncology* (2015) 29:970–2, 974–5.
3. Ward JP, Gubin MM, Schreiber RD. The role of neoantigens in naturally occurring and therapeutically induced immune responses to cancer. *Adv Immunol* (2016) 130:25–74. doi:10.1016/bs.ai.2016.01.001
4. Gubin MM, Zhang X, Schuster H, Caron E, Ward JP, Noguchi T, et al. Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature* (2014) 515:577–81. doi:10.1038/nature13988
5. Castle JC, Kreiter S, Diekmann J, Löwer M, van de Roemer N, de Graaf J, et al. Exploiting the mutanome for tumor vaccination. *Cancer Res* (2012) 72:1081–91. doi:10.1158/0008-5472.CAN-11-3722
6. Yadav M, Jhunjhunwala S, Phung QT, Lupardus P, Tanguay J, Bumbaca S, et al. Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature* (2014) 515:572–6. doi:10.1038/nature14001
7. Sahin U, Derhovanessian E, Miller M, Kloke B-P, Simon P, Löwer M, et al. Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* (2017) 547:222–6. doi:10.1038/nature23003
8. Ott PA, Hu Z, Keskin DB, Shukla SA, Sun J, Bozym DJ, et al. An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* (2017) 547:217–21. doi:10.1038/nature22991
9. Hartmaier RJ, Charo J, Fabrizio D, Goldberg ME, Albacker LA, Pao W, et al. Genomic analysis of 63,220 tumors reveals insights into tumor uniqueness and targeted cancer immunotherapy strategies. *Genome Med* (2017) 9:16. doi:10.1186/s13073-017-0408-2
10. Kyi C, Sabado RL, Blazquez A, Posner MR, Genden EM, Miles BA, et al. A phase I study of the safety and immunogenicity of a multipptide personalized genomic vaccine in the adjuvant treatment of solid cancers. *J Clin Orthod* (2017) 35:TS3114–3114. doi:10.1200/JCO.2017.35.15_suppl.TPS3114

11. Rosalia RA, Quakkelaar ED, Redeker A, Khan S, Camps M, Drijfhout JW, et al. Dendritic cells process synthetic long peptides better than whole protein, improving antigen presentation and T-cell activation. *Eur J Immunol* (2013) 43:2554–65. doi:10.1002/eji.201343324
12. Ammi R, De Waele J, Willemsen Y, Van Brussel I, Schrijvers DM, Lion E, et al. Poly(I:C) as cancer vaccine adjuvant: knocking on the door of medical breakthroughs. *Pharmacol Ther* (2015) 146:120–31. doi:10.1016/j.pharmthera.2014.09.010
13. Oh E, Choi Y-L, Kwon MJ, Kim RN, Kim YJ, Song J-Y, et al. Comparison of accuracy of whole-exome sequencing with formalin-fixed paraffin-embedded and fresh frozen tissue samples. *PLoS One* (2015) 10:e0144162. doi:10.1371/journal.pone.0144162
14. Zhao W, He X, Hoadley KA, Parker JS, Hayes DN, Perou CM. Comparison of RNA-Seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC Genomics* (2014) 15:419. doi:10.1186/1471-2164-15-419
15. Wei L, Papanicolaou-Sengos A, Liu S, Wang J, Conroy JM, Glenn ST, et al. Pitfalls of improperly procured adjacent non-neoplastic tissue for somatic mutation analysis using next-generation sequencing. *BMC Med Genomics* (2016) 9:64. doi:10.1186/s12920-016-0226-1
16. Adey A, Morrison HG, Asan, Xun X, Kitzman JO, Turner EH, et al. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol* (2010) 11:R119. doi:10.1186/gb-2010-11-12-r119
17. García-García G, Baux D, Faugère V, Moclyn M, Koenig M, Claustres M, et al. Assessment of the latest NGS enrichment capture methods in clinical context. *Sci Rep* (2016) 6:20948. doi:10.1038/srep20948
18. Shigemizu D, Momozawa Y, Abe T, Morizono T, Boroevich KA, Takata S, et al. Performance comparison of four commercial human whole-exome capture platforms. *Sci Rep* (2015) 5:12742. doi:10.1038/srep12742
19. Cornish A, Guda C. A comparison of variant calling pipelines using genome in a bottle as a reference. *Biomed Res Int* (2015) 2015:456479. doi:10.1155/2015/456479
20. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. *Nat Commun* (2015) 6:8971. doi:10.1038/ncomms9971
21. Li H. Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. *arXiv [q-bio.GN]* (2013). Available from: <http://arxiv.org/abs/1303.3997>
22. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* (2013) 29:15–21. doi:10.1093/bioinformatics/bts635
23. Sun Z, Bhagwate A, Prodduturi N, Yang P, Kocher J-PA. Indel detection from RNA-seq data: tool evaluation and strategies for accurate detection of actionable mutations. *Brief Bioinform* (2016) 18:973–83. doi:10.1093/bib/bbw069
24. Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, Del Angel G, Levy-Moonshine A, et al. From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* (2013) 43:11.10.1–33. doi:10.1002/0471250953.bi1110s43
25. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* (2013) 31:213–9. doi:10.1038/nbt.2514
26. Saunders CT, Wong WSW, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* (2012) 28:1811–7. doi:10.1093/bioinformatics/bts271
27. Cereb N, Kim HR, Ryu J, Yang SY. Advances in DNA sequencing technologies for high resolution HLA typing. *Hum Immunol* (2015) 76:923–7. doi:10.1016/j.humimm.2015.09.015
28. Boegel S, Löwer M, Schäfer M, Bukur T, de Graaf J, Boisguérin V, et al. HLA typing from RNA-Seq sequence reads. *Genome Med* (2012) 4:102. doi:10.1186/gm403
29. Rubinsteyn A, Hodes I, Kodysh J, Hammerbacher J. Vaxrank: A Computational Tool for Designing Personalized Cancer Vaccines. *bioRxiv 142919* (2017) 142919. Available from: <http://biorxiv.org/content/early/2017/05/27/142919.abstract>
30. Rubinsteyn A, Kodysh J, Aksoy BA. *hammerlab/isovar: Version 0.7.0* (2017). doi:10.5281/zenodo.821224
31. Rubinsteyn A, O'Donnell T, Nathanson T, Ahuja A, Kodysh J, Aksoy BA, et al. *hammerlab/varcode: Version 0.5.15* (2017). doi:10.5281/zenodo.822503
32. Rubinsteyn A, Nathanson T, Kodysh J, O'Donnell T, Ahuja A, Hammerbacher J, et al. *hammerlab/pyensembl: Version 1.1.0* (2017). doi:10.5281/zenodo.822502
33. Rubinsteyn A, Nathanson T, Kodysh J, O'Donnell T, Ahuja A, Damaraju N. *hammerlab/mhctools: Version 1.5.0* (2017). doi:10.5281/zenodo.822504
34. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* (2012) 6:80–92. doi:10.4161/fly.19695
35. 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
36. Sveen A, Kilpinen S, Ruusulehto A, Lothe RA, Skotheim RI. Aberrant RNA splicing in cancer; expression changes and driver mutations of splicing factor genes. *Oncogene* (2016) 35:2413–27. doi:10.1038/ncr.2015.318
37. Hoof I, Peters B, Sidney J, Pedersen LE, Sette A, Lund O, et al. NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics* (2009) 61:1–13. doi:10.1007/s00251-008-0341-z
38. Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8–11. *Nucleic Acids Res* (2008) 36:W509–12. doi:10.1093/nar/gkn202
39. Karosiene E, Lundegaard C, Lund O, Nielsen M. NetMHCcons: a consensus method for the major histocompatibility complex class I predictions. *Immunogenetics* (2012) 64:177–86. doi:10.1007/s00251-011-0579-8
40. Sette A, Vitiello A, Reheman B, Fowler P, Nayarsina R, Kast WM, et al. The relationship between class I binding affinity and immunogenicity of potential cytotoxic T cell epitopes. *J Immunol* (1994) 153:5586–92.
41. Mondet S, Aksoy BA, Rozenberg L, Hodes I, Hammerbacher J. Bioinformatics Workflow Management with the Wobidisco Ecosystem. *bioRxiv 213884* (2017) 213884. doi:10.1101/213884
42. Hundal J, Carreno BM, Petti AA, Linette GP, Griffith OL, Mardis ER, et al. pVAC-Seq: a genome-guided in silico approach to identifying tumor neoantigens. *Genome Med* (2016) 8:11. doi:10.1186/s13073-016-0264-5
43. Bjerregaard A-M, Nielsen M, Hadrup SR, Szallasi Z, Eklund AC. MuPeXI: prediction of neo-epitopes from tumor sequencing data. *Cancer Immunol Immunother* (2017). doi:10.1007/s00262-017-2001-3
44. Snyder A, Nathanson T, Funt SA, Ahuja A, Buros Novik J, Hellmann MD, et al. Contribution of systemic and somatic factors to clinical response and resistance to PD-L1 blockade in urothelial cancer: an exploratory multi-omic analysis. *PLoS Med* (2017) 14:e1002309. doi:10.1371/journal.pmed.1002309

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer BS and handling editor declared their shared affiliation.

Copyright © 2018 Rubinsteyn, Kodysh, Hodes, Mondet, Aksoy, Finnigan, Bhardwaj and Hammerbacher. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.