# illumına<sup>®</sup>

# Immuno-oncology Informatics Applications Using Next-Generation Sequencing

Improved informatics tools enable neoantigen discovery and tumor microenvironment analysis with standard Illumina sequencing applications.

## Introduction

Immuno-oncology is an emerging field of research that represents tremendous promise for both patients and those fighting the war on cancer. It is also an opportunity for the scientific community to gain a better understanding of the underlying biological mechanisms involved in the interactions between tumors and the immune system. New discoveries are leading to improved genomic solutions capable of driving development of novel therapeutics and predict patient response to immunotherapies. Next-generation sequencing (NGS) is a key technology being applied to these discoveries. Illumina has developed sequencing tools and applications that make it easy for any research group to begin their own immuno-oncology experiments.

Some forms of immunotherapy were developed to block the ability of cancers to avoid immune destruction, allowing researchers to target tumors in a personalized way.<sup>1</sup> Ideally, mutated proteins resulting from tumors with nonsynonymous somatic mutations are presented as antigens on the surface of cancer cells for recognition by the immune system. However, cancer cells have implemented various strategies to avoid recognition and eradication by the immune system. The first is to express components of the immune checkpoint pathway that inhibit T cell activation, such as PD-L1. A second strategy is to alter the tumor microenvironment by recruiting tumor associated macrophages and regulatory T cells that block infiltration of cytotoxic T cells.<sup>2</sup>

Checkpoint inhibitors, such as anti-CTLA4 and anti-PD1 antibodies, have recently received FDA approval and shown efficacy in a subset of patients with certain cancers such as melanoma and nonsmall cell lung cancer. Evidence also suggests that cancer patients with a higher number of nonsynonymous mutations have improved response to checkpoint inhibitor therapy.<sup>3-5</sup> A more personalized type of immunotherapy is the development of a cancer vaccine based on neoantigens created by tumor-specific nonsynonymous DNA alterations.<sup>6</sup> Neoantigens can persistently stimulate cytotoxic T lymphocytes, which may eradicate cancer cells in the presence of checkpoint inhibitors or a favorable immune microenvironment.

This application note describes 2 immuno-oncology workflows. The first workflow uses whole-exome sequencing (WES) and whole-transcriptome sequencing (WTS) to predict neoantigens that cancer cells present. Studies show that overall mutational load and specific neoantigen signatures may predict response to immuno-oncology therapies.<sup>3</sup> The second workflow uses WTS to characterize the tumor microenvironment in a broad survey of important expression markers (eg, immune modulatory genes, tumor-specific variants, and fusions). These tools enable detection of low-expressing targets with high confidence. In contrast, it is challenging for hybridization-based technologies to reach such breadth and

sensitivity. WES and WTS sequencing can be performed using the NextSeq<sup>®</sup> or HiSeq<sup>®</sup> Series of Illumina sequencing systems. Illumina sequencing kits are optimized for library preparation from FFPE or fresh/frozen samples. Illumina TruSeq<sup>®</sup> Exome and TruSeq RNA Access Library Preparation Kits use the same sequence-specific capture technologies to target exonic regions. The matched content enables researchers to target and compare equivalent regions of RNA and DNA. Both workflows use the BaseSpace<sup>®</sup> Informatics Suite, providing any researcher with rapid analysis of NGS data and flexible characterization of the immuno-oncology profile of a tumor sample.

**Isaac Enrichment RNA-Seq Alignment** Tumor/Normal WES FASTQ Tumor WTS FASTQ Variant Calling Assessment Tool Tumor Tumor ► WES VCF WTS VCF VCAT Intersected VCF **VariantStudio** Normal WES VCF Nonsynonymous Variants Gene Table Gene FPKM

#### BaseSpace Sequence Hub

**Figure 1: Neoantigen Discovery Workflow**—Using BaseSpace Sequence Hub apps, NGS analysis of immune pathways is efficient and accessible to researchers without informatics experience. VCF = variant call file. FPKM = fragments per kilobase of transcript per million mapped reads.

### Neoantigen Discovery by Integrating WES and WTS Analysis

This protocol summarizes the steps for generating neoantigen candidates, using BaseSpace Sequence Hub apps to analyze data from a combined WES and WTS study (Figure 1). To begin with, WES tumor-normal and WTS tumor data are streamed from the sequencer directly to BaseSpace Sequence Hub. After sequencing is complete, the respective FASTQ files are made available in Sequence Hub. Data analysis can be performed in 4 main steps:

- Use the Isaac Enrichment v2.1 App in BaseSpace Sequence Hub for WES alignment and variant calling of matched tumor-normal samples. Inputs are tumor WES FASTQ and normal WES FASTQ files. The output of the Isaac Enrichment App is a fully annotated variant call file (VCF).
- Perform WTS alignment, gene expression estimation, and variant calling in parallel using the RNA-Seq Alignment v1.1 App in BaseSpace Sequence Hub. The inputs to the app are tumor WTS FASTQ files. The outputs from the app are a WTS VCF file and a gene expression count file (raw and FPKM). The gene expression quantification is used in the final step of neoantigen prediction.
- 3. Filter out false positive variants from the tumor data to focus on expressed variants from a pairwise intersection of the WES and WTS tumor variants using the Variant Calling Assessment Tool v2.3 App in BaseSpace Sequence Hub. Inputs are the tumor WES VCF file from step 1, and the tumor WTS VCF file from step 2. The output is an intersected VCF file, which contains only the variants present in both the WES and WTS tumor data.
- 4. Determine the variants relevant for neoantigen prediction, the WES/WTS-intersected tumor VCF file, and the WES normal VCF file are loaded into the VariantStudio App in BaseSpace Sequence Hub. With the click of a button, all intersected variants are annotated with transcript info, consequences, population frequencies, and materials extracted from other external databases (such as Cosmic, ClinVar, and dbSNP). The intersected variants are then filtered in VariantStudio by first subtracting out the variants from the WES normal samples. The remaining variants are filtered to retain nonsynonymous variants and variants that are present in less than 1% of the population. These variants are then exported and used for neoantigen prediction.





Ideal neoantigen candidates have high variant allele frequencies (VAF) in both tumor DNA and RNA samples. Additional prioritizations occur by variant type, such as frameshift, splice site mutations resulting in neo-open-reading-frames (neoORFs). Some of these variants are entirely specific to the tumor compared to missense mutations. Medium to high RNA expression can be selected to further filter candidate neoantigens. Based on the above criteria, final neoantigen candidates can be prioritized based on DNA and RNA VAFs, mutational type (missense vs neoORF), and gene expression levels (Figure 2).

Potential neoantigens can be further filtered according to their affinity to individual HLA class 1 types. Commonly used tools include mass spectrometry analysis and/or *in silico* algorithms for MHC-associated neoantigen prediction. Several *in silico* prediction algorithms are available online (Table 1).

5		
Application	Algorithm	URL
HLA Genotyping Using WES Data	Polysolver	www.broadinstitute.org/cancer/cga/polysolver7
	OptiType	github.com/FRED-2/OptiType <sup>8</sup>
HLA Class I Binding Affinity Prediction	netMHC	www.cbs.dtu.dk/services/NetMHC/9
	SYFPEITHI	www.syfpeithi.de/10

#### Table 1: In-silico Prediction Algorithms

### **Tumor Microenvironment Analysis**

To assess the tumor microenvironment, WTS samples are streamed from the sequencer directly to BaseSpace Sequence Hub. After sequencing is complete, the respective FASTQ files are made available in Sequence Hub. Data analysis can be performed in 2 steps (Figure 3A):

- Perform WTS alignment and gene expression estimation of the tumor-normal samples using the RNA-Seq Alignment v1.1 App in BaseSpace. To enable easy visualization of relevant tumor microenvironment genes, an immuno-oncology filter is available to display gene expression count information only for genes relevant to immune modulatory and cancer pathways (Figure 3B). Output files, including binary alignment map (BAM) files, are automatically recognized by the Differential Expression App in the next step. The same app also performs FPKM estimation, assembly of novel transcripts, variant calling, and optional fusion calling.
- 2. Perform differential expression analysis between a matched tumornormal sample using the Cufflinks Assemble and DE v2.1 or the DESeq v2.1 App in BaseSpace Sequence Hub. The Differential Expression App also provides the option to generate only content relevant to the tumor microenvironment, with the results from step 1 as input. Output includes a global analysis report summarizing statistics including sample correlations (Figure 3C), differential expression heat map incorporating differentially expressed immuno-oncology genes, and a gene browser with filters on various parameters (Figure 3D).

Optionally, researchers can use the list of differentially expressed genes (their signature) for downstream pathway analysis or gene sets enrichment analysis. Commonly used algorithms are available online (Table 2).

#### Table 2: Algorithms for Pathway Analysis and Gene Set Enrichment

Application	Description	URL
BaseSpace Correlation Engine	Compares signatures to over 20,000 other studies and pathways	www.nextbio.com/b/nextbio.nb11
DAVID Bioinformatics Resources	Tests signatures for enrichment against pathways, gene ontology and more	david.ncifcrf.gov/12
Gene Set Enrichment Analysis (GSEA)	Tests signatures for enrichment against a database of signatures	$software.broad institute.org/gsea/index.jsp^{\rm 13}$



Figure 3: Tumor Microenvironment Analysis Workflow Using BaseSpace Sequence Hub Apps — A. BaseSpace workflow includes an alignment step followed by differential expression analysis. B. An immuno-oncology filter is available for both steps, enabling output that only displays genes relevant to the tumor microenvironment. C. Output from differential expression analysis, showing sample correlations for global gene expression analysis. D. Output from differential expression analysis, showing sample correlations for global gene expression analysis. D. Output from differential expression analysis, showing sample correlations for global gene expression analysis. D. Output from differential expression analysis, showing heat maps and scatterplots for immuno-oncology genes.

#### Conclusion

NGS technology from Illumina has contributed to a wealth of information about cancer genomics, including specialized applications for DNA and RNA profiling of tumors and the host environment. Using new informatics tools, proven NGS methods can be further customized for applications in immuno-oncology. The BaseSpace Sequence Hub can be integrated with the NextSeq and HiSeq Series of Sequencing Systems, using TruSeq Exome and RNA Access Library Prep Kits. The BaseSpace Informatics Suite provides flexibility and efficiency of analysis, making immuno-oncology research accessible to any research group.

#### Learn More

For more information on NGS applications for immuno-oncology, visit

www.illumina.com/immuno-oncology

#### References

- 1. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell.* 2011;144(5):646-674.
- Gajewski TF, Schreiber H, Fu YX. Innate and adaptive immune cells in the tumor microenvironment. Nat Immunol. 2013;14(10):1014-1022.
- Snyder A, Makarov V, Merghoub T, et al. Genetic basis for clinical response to CTLA-4 blockade in melanoma. N Engl J Med. 2014;371(23):2189-2199.
- Rizvi NA, Hellmann MD, Snyder A, et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in nonsmall cell lung cancer. *Science*. 2015;348(6230):124-128.
- Le DT, Uram JN, Wang H, et al. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. N Engl J Med. 2015;372(26):2509-2520.
- Castle JC, Kreiter S, Diekmann J, et al. Exploiting the mutanome for tumor vaccination. Cancer Res. 2012;72(5):1081-1091.
- 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.
- Szolek A, Schubert B, Mohr C, Sturm M, Feldhahn M, Kohlbacher O. OptiType: precision HLA typing from next-generation sequencing data. *Bioinformatics*. 2014;30(23):3310-3316.
- Andreatta M, Nielsen M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics*. 2016;32(4):511-517.
- Schuler MM, Nastke MD, Stevanovikc S. SYFPEITHI: database for searching and T-cell epitope prediction. Methods Mol Biol. 2007;409:75-93.
- Kupershmidt I, Su QJ, Grewal A, et al. Ontology-based metaanalysis of global collections of high-throughput public data. *PLoS One.* 2010;5(9). pii:e13066. doi: 10.1371/journal.pone.0013066.
- Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009;4(1):44-57.
- Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545-15550.

Illumina • 1.800.809.4566 toll-free (US) • +1.858.202.4566 tel • techsupport@illumina.com • www.illumina.com

For Research Use Only. Not for use in diagnostic procedures.

© 2016 Illumina, Inc. All rights reserved. Illumina, BaseSpace, HiSeq, NextSeq, TruSeq, and the pumpkin orange color are trademarks of Illumina, Inc. and/or its affiliate(s) in the U.S. and/or other countries. Pub. No. 1170-2016-013 Current as of 07 April 2016

