IMPACT: a whole-exome sequencing analysis pipeline for integrating molecular profiles with actionable therapeutics in clinical samples

RECEIVED 31 August 2015
REVISED 18 December 2015
ACCEPTED 1 February 2016
PUBLISHED ONLINE FIRST 28 March 2016

Jennifer Hintzsche,¹ Jihye Kim,^{1,7} Vinod Yadav,² Carol Amato,¹ Steven E Robinson,¹ Eric Seelenfreund,¹ Yiqun Shellman,^{3,7} Joshua Wisell,^{4,7} Allison Applegate,¹ Martin McCarter,^{5,7} Neil Box,^{3,7} John Tentler,^{1,7} Subhajyoti De,^{2,6,7} William A Robinson,^{1,7} Aik Choon Tan^{1,6,7}

ABSTRACT

Objective Currently, there is a disconnect between finding a patient's relevant molecular profile and predicting actionable therapeutics. Here we develop and implement the Integrating Molecular Profiles with Actionable Therapeutics (IMPACT) analysis pipeline, linking variants detected from whole-exome sequencing (WES) to actionable therapeutics.

Methods and materials The IMPACT pipeline contains 4 analytical modules: detecting somatic variants, calling copy number alterations, predicting drugs against deleterious variants, and analyzing tumor heterogeneity. We tested the IMPACT pipeline on whole-exome sequencing data in The Cancer Genome Atlas (TCGA) lung adenocarcinoma samples with known EGFR mutations. We also used IMPACT to analyze melanoma patient tumor samples before treatment, after BRAF-inhibitor treatment, and after BRAF- and MEK-inhibitor treatment.

Results IMPACT Food and Drug Administration (FDA) correctly identified known EGFR mutations in the TCGA lung adenocarcinoma samples. IMPACT linked these EGFR mutations to the appropriate FDA-approved EGFR inhibitors. For the melanoma patient samples, we identified NRAS p.Q61K as an acquired resistance mutation to BRAF-inhibitor treatment. We also identified CDKN2A deletion as a novel acquired resistance mutation to BRAF/MEK inhibition. The IMPACT analysis pipeline predicts these somatic variants to actionable therapeutics. We observed the clonal dynamic in the tumor samples after various treatments. We showed that IMPACT not only helped in successful prioritization of clinically relevant variants but also linked these variations to possible targeted therapies.

Conclusion IMPACT provides a new bioinformatics strategy to delineate candidate somatic variants and actionable therapies. This approach can be applied to other patient tumor samples to discover effective drug targets for personalized medicine.

IMPACT is publicly available at <http://tanlab.ucdenver.edu/IMPACT>.

Keywords: bioinformatics, whole exome sequencing, cancer, therapeutics, personalized medicine

BACKGROUND AND SIGNIFICANCE

The advancement of next-generation sequencing (NGS) technologies has revolutionized modern biomedical research, especially cancer research. It is now possible to obtain variants from whole-exome sequencing (WES) at single-base resolution from individual patients using NGS. In cancer genomics, recent studies using NGS have successfully identified novel somatic alterations such as single-nucleotide variants, insertions and deletions, copy number aberrations (CNAs), structural variants, and gene fusions. Determining the functional relevance of these identified somatic mutations in cancer is a prerequisite to performing precision oncology, which involves matching specific therapies to individual patients based on their genomics. Therefore, the challenge for cancer genomics has shifted from the bottleneck of data generation to data analysis and interpretation.

Several computational analysis pipelines have been developed, primarily focusing on identifying either single nucleotide variants or copy number alterations from WES data. For example, most of the data analysis pipelines incorporate variant callers such as SAMTools and GATK to obtain the Variant Calling Format (VCF) file. From the VCF

files, existing pipelines use a variety of variant annotation tools such as ANNOVAR to annotate the significance of the variants. These pipelines return the annotated files that users then have to narrow down manually to identify actionable targets. These WES data analysis pipelines include the Broad GDAC Firehose, STORMSeq,¹ SIMPLEX,² Atlas2 Suite,³ Fastq2vcf,⁴ WEP,⁵ and EVA.⁶ Other analysis tools were provided in an excellent review.⁷ In contrast, some data analysis pipelines focus solely on identifying copy number alterations from WES. These copy number analysis pipelines were recently reviewed.⁸ Only a limited number of pipelines were proposed to link candidate genes identified in WES with therapeutics.^{9–11}

Here, we introduce the Integrating Molecular Profiles with Actionable Therapeutics (IMPACT) analysis pipeline, which integrates single nucleotide variants and copy number alterations from WES data to identify a list of candidate genes for therapeutic targets. From the list of candidate genes, IMPACT returns a prioritized list of drugs predicted to target these cancer genes, using our recently published comprehensive drug-target database. The IMPACT analysis pipeline also allows users to explore the tumor heterogeneity of the sample. This

Correspondence to: Aik Choon Tan, PhD, Associate Professor of Bioinformatics, Division of Medical Oncology, Department of Medicine, University of Colorado Anschutz Medical Campus; Email: aikchoon.tan@ucdenver.edu; Tel: 303-724-2520; Fax: 303-724-3889 and William A. Robinson, MD, PhD, Professor of Medicine, Rella and Monroe Rifkin Endowed Chair in Cancer Research, Division of Medical Oncology, Department of Medicine, University of Colorado Anschutz Medical Campus; Email: William.Robinson@ucdenver.edu. For numbered affiliations see end of article.

© The Author 2016. Published by Oxford University Press on behalf of the American Medical Informatics Association. All rights reserved. For Permissions, please email: journals.permissions@oup.com

allows users to understand the clonal dynamics from WES data during the course of treatment, or comparisons of clinically similar samples to each other. Taken together, IMPACT represents a novel analysis pipeline that links raw WES data with actionable therapeutics. To illustrate the utility of IMPACT, we analyzed 6 different tumor samples from 2 different types of cancer. Three of these samples were taken from 1 melanoma patient with a BRAF p.V600E mutation over the course of treatment. This demonstrates IMPACT's ability to provide actionable therapeutics even when the patient has become resistant to certain therapies. We also demonstrate IMPACT's ability to predict actionable therapeutics in other cancers by utilizing 3 samples from the TCGA lung adenocarcinoma cohort with EGFR mutations.¹²

MATERIALS AND METHODS

IMPACT Analysis Pipeline

The IMPACT analysis pipeline consists of 4 analytical modules: (1) single nucleotide variants and (2) copy number alterations analyses, with the candidate genes from these 2 analyses used to link with (3) drug prediction and (4) tumor heterogeneity analysis. IMPACT uses paired-end WES reads (fastq.gz files) as input, with the option to also include a matched normal sample. IMPACT returns 2 tiers of actionable therapeutics based on the molecular profile of the sample. The complete IMPACT analysis pipeline is illustrated in Figure 1. The following sections describe the IMPACT analytical components in detail. Specific parameters for each module are outlined in the IMPACT v1.0 user manual, which can be found at <http://tanlab.ucdenver.edu/IMPACT>, along with the melanoma patient data and software.

(1) Analysis of Single Nucleotide Variants. This module takes fastq.gz as input and outputs a VCF file containing predicted deleterious mutations. Sequencing reads generated from the Illumina next-generation sequencer were extracted and processed by the Illumina Bustard base calling and alignment program to convert the sequencing bases and quality scores to the FASTQ format. Using the converted FASTQ format as the input for this module, the sequencing reads were mapped to the human hg19 reference exome using the Burrows-Wheeler Aligner (v0.7.8-r455).^{13,14} The Burrows-Wheeler Aligner was run with default parameters for paired-end 125 bp-long read sequences. SAMtools (v1.1)¹⁵ was used to convert Sequence Alignment Map (SAM) files to sorted Binary Alignment Map (BAM) files. BCFtools (v1.1) was utilized to detect variants from the BAM file and outputs into a VCF file. In the IMPACT pipeline, we used ANNOVAR (v2014-11-12)¹⁶ to annotate the variants. Somatic mutations were determined by removing the normal variants found only in the tumor sample as compared to the matched normal sample. Synonymous and intronic

variants were removed, along with variants present in greater than 1 percent of the 1000 genome project¹⁷ and/or Exome Sequencing Project (ESP6500). Variants were then filtered based on the 10 deleterious predictors in ANNOVAR. The 2 most commonly known deleterious mutations in melanoma are found at BRAF^{V600} and NRAS^{Q61} and were thus used to test the accuracy of these deleterious predictors, as shown in Figure 2. Six prediction algorithms correctly predicted BRAF^{V600} and NRAS^{Q61} mutations as deleterious. Therefore, variants were further analyzed only if they were predicted deleterious by SIFT,¹⁸ PolyPhen2,¹⁹ FATHMM,²⁰ CADD,²¹ GERP,²² or MutationTaster.²³ To avoid filtering out potential actionable targets, we also included variants that were present in the COSMIC database and were predicted to be deleterious by SIFT or PolyPhen2.

(2) Copy Number Alterations. This module takes either fastq.gz files or BAM files as the inputs and returns copy number changes from a normal sample. We used VarScan2²⁴ as the tool for identifying CNAs from WES data in the IMPACT analysis pipeline. VarScan2 calculated the log2 score difference in sequencing depth between normal and cancer exomes in 100 base pair sliding segments. Fisher's exact test was used to determine whether the ratio of tumor and normal depth changed significantly. Circular binary segmentation was applied to each target region to merge adjacent small segments into large segments. Sample coverage for each segment was classified as being

Figure 2: Prediction algorithms on the deleterious variants

Prediction Methods	BRAF		NRAS	
	V600E	V600K	Q61K	Q61L
SIFT				
PolyPhen2				
FATHMM				
CADD				
GERP				
MutationTaster				
LRT				
RadialSVM				
MutationAssessor				
LR_PRED				

☒ Predicted Deleterious
 ☐ Predicted Benign

Figure 1: Overview of the IMPACT Analysis Pipeline

an amplification ($\log_2 > 2$), deletion ($\log_2 < -2$), or neutral ($-2 \leq \log_2 \leq 2$) event across each gene's total sample coverage. One of the caveats of VarScan2 is that it tends to call small segments either amplified or loss, and therefore it identified more CNAs.⁷ To overcome this, we have developed our own post-processing script to filter the output from VarScan2. First, the total coverage of each exon for each gene was calculated. Next, genes that had areas of no coverage were given a $\log_2$ score of -10 to represent a deletion. Last, genes that had an overall coverage of at least 5000 base pairs and consecutive segments constituting $>70\%$ of the total gene length were identified as amplified or deleted in each sample.

(3) Drug Prediction. Candidate genes with variants identified from modules 1 and 2 were used as the inputs to link with therapeutics. We used a 2-tier system to connect genes with therapies. The first-tier system (Level 1) is to follow the National Cancer Institute (NCI)-Match (<http://www.cancer.gov/about-cancer/treatment/clinical-trials/nci-supported/nci-match>) and MD Anderson Personalized Cancer Therapy (PCT) (<https://pct.mdanderson.org>) criteria, where a gene with known variants with Food and Drug Administration (FDA)-approved drugs will be prioritized. First-tier therapeutics also included FDA-approved kinase inhibitors from DSigDB²⁵ (<http://tanlab.ucdenver.edu/DSigDB/>), a database of kinases and their targets manually curated from peer-reviewed literature. Second-tier therapeutics (Level 2) included all FDA-approved drugs and their gene targets found in DSigDB, a comprehensive drug-gene target database for ranking the potential therapeutics against the candidate genes. For each drug in the DSigDB database, we used a hypergeometric test and a permutation test to determine whether the target genes were enriched in the variants list. Both tests were done on the drugs that had at least 1 drug-target gene from the variants list. Since some drugs target only 1 or 2 genes and can be underestimated (ie, significance overestimation), we resampled target genes per drug 1000 times from the list of DSigDB FDA-approved drugs and compared common target genes (test statistic) from the IMPACT candidate genes. We assumed that every gene from the database had equal probability to be a target. Generally, both tests showed lower P -values for the drugs with smaller numbers of targets, which indicates that more target-specific drugs are considered as candidate therapeutics. Drugs with the lowest P -values will be considered as candidate therapeutics. The output from this step is a 2-tiered list of drugs with their gene targets and the source they were matched from (NCI Match Clinical Trial, MD Anderson PCT, or DSigDB). HTML output provides links to the DSigDB compound webpage for viewing the molecule information and drug-target genes.

(4) Tumor Heterogeneity Analysis. We measured tumor clonality and heterogeneity using mutant allele frequency estimates, similar to what has been published previously.^{26,27} We compared distributions of allele frequency estimates using the Kolmogorov-Smirnov test.²⁸

TCGA Lung Adenocarcinoma Samples

We downloaded 3 matched normal-tumor TCGA lung adenocarcinoma samples from the University of California Santa Cruz (UCSC) Cancer Genomics Hub (<https://cghub.ucsc.edu>): TCGA-49-4494, TCGA-50-5944, and TCGA-64-1681. These 3 TCGA samples were selected because all of them have EGFR p.L858R mutation. The TCGA-49-4494 sample has an additional EGFR p.T790M mutation. These EGFR mutations will serve as the positive controls for the variant detection module (Module 1) in the IMPACT analysis pipeline. EGFR p.L858R is a predictive biomarker for FDA-approved EGFR inhibitors such as gefitinib (first generation), erlotinib (first generation), afatinib (second generation), and, currently in clinical trial, a third-generation EGFR

inhibitor, AZD9291. However, EGFR p.T790M is a gatekeeper mutation that blocks the first-generation EGFR inhibitors and only responds to afatinib and AZD9291. These EGFR mutations and their linking with therapeutics will serve as positive controls for the drug prediction module (Module 3) of IMPACT.

WES sequencing

Genomic DNA was extracted and purified from tissue samples using the Clontech NucleoSpin® Blood QuickPure kit. Sequences of the exome were enriched using the Agilent SureSelect Exome enrichment kit. The estimated target size of the exome was about 50 Mb, covering $\sim 20\,000$ coding genes. Enriched exome sequences were sheared to 250–350 bp using the Covaris DNA shearing system. Illumina sequencing adapter and barcode (index) were added to the sheared DNA. Enriched samples (library) were sequenced on the Illumina HiSeq 2000 for 125 bp pair-end reads. We obtained an average 400X and 200X sequencing coverage for the cancer and normal exome(s), respectively.

Validation of the Identified Somatic Mutations

Specific mutations in *BRAF*, *CDKN2A*, *CTNNB1*, *GNAQ*, *HRAS*, *KIT*, *KRAS*, *NRAS*, *PIK3CA*, and *STK11* were validated using the qBiomarker Melanoma Somatic Mutation PCR array (Qiagen Inc., Valencia, CA, USA). This is an allele-specific and hydrolysis probe-based detection of somatic mutations commonly found in melanoma. One microgram of genomic DNA was assayed, and data was calculated using the delta CT method after 40 cycles of PCR.

Patient Recruitment

Tissue acquisition from consenting melanoma patients at the time of removal of a primary tumor or biopsy was conducted under a Colorado Multi-Institutional Review Board–approved protocol, COMIRB-05-0309.

RESULTS

IMPACT Pipeline for WES Analysis

We developed and implemented IMPACT as a novel WES analysis pipeline that integrates 4 analytical modules: (1) variant detection, (2) copy number estimation, (3) drug prediction, and (4) tumor heterogeneity analysis. The workflow of the IMPACT pipeline is illustrated in Figure 1. The IMPACT pipeline accepts WES FASTQ files (eg, tumor and normal samples) as the inputs, and returns the 4 analytical results from each module as the outputs. Users could also perform individual IMPACT module analysis on their samples. The IMPACT pipeline and the analytical modules were implemented as a script written in Perl (v5.10.1). See the IMPACT User Manual for parameters, dependencies of each module, and instructions on running the program. The IMPACT pipeline is freely available at <http://tanlab.ucdenver.edu/IMPACT>.

IMPACT Analysis of the TCGA Lung Adenocarcinoma Samples

To test the IMPACT pipeline in detecting variants and linking them to therapeutics, we first performed the analysis on 3 matched tumor-normal whole-exome sequences of TCGA lung adenocarcinoma samples (TCGA-49-4494, TCGA-50-5944, and TCGA-64-1681) with known EGFR p.L858R mutation. One of the samples, TCGA-49-4494, has an additional EGFR p.T790M mutation. These known EGFR mutations were used as the positive controls for testing the variants detection module (Module 1) of the IMPACT pipeline. Moreover, these EGFR mutations have been associated with different drugs: EGFR p.L858R is associated with responses to gefitinib, erlotinib, afatinib, and AZD9291, whereas EGFR p.T790M is associated with responses to afatinib and AZD9291. These gene-variant–drug associations were

used as the positive controls for testing the drug prediction module (Module 3) of the IMPACT pipeline.

We used the IMPACT pipeline to analyze the somatic variants in the 3 TCGA lung adenocarcinoma samples. IMPACT identified 21 111, 27 348, and 18 163 somatic variants in TCGA-64-1681, TCGA-49-4494, and TCGA-50-5944, respectively. Among the somatic variants, 1 923,

1909, and 1399 were exonic and non-synonymous in TCGA-64-1681, TCGA-49-4494, and TCGA-50-5944, respectively. Among these non-synonymous exonic variants, 181, 126, and 72 somatic variants were predicted as deleterious in TCGA-64-1681, TCGA-49-4494, and TCGA-50-5944, respectively. Importantly, IMPACT correctly identified the EGFR p.L858R mutation in all 3 samples, as well as the EGFR p.T790M mutation in TCGA-49-4494. As these EGFR mutations have been validated in the TCGA studies and used as the positive controls, this supports the idea that IMPACT could identify actionable variants from WES data. For the copy number alterations, IMPACT identified 77, 76, and 82 deletions in TCGA-64-1681, TCGA-49-4494, and TCGA-50-5944, respectively. [Table 1](#) summarizes the number of variants and copy number alterations detected in these samples. The list of somatic variants and copy number alterations detected in these lung samples is available in [Supplementary Tables 1–3](#).

Next, we examined the drugs predicted by IMPACT on these 3 TCGA lung adenocarcinoma samples. For TCGA-64-1681 and TCGA-50-5944, where both samples harbored EGFR, p.L858R was predicted to link with afatinib, cetuximab, erlotinib, gefitinib, ibrutinib, lapatinib, panitumumab, and vandertanib as Level 1 actionable therapeutics. All these drugs except ibrutinib target EGFR; ibrutinib targets BTK. As described, EGFR p.L858R is a biomarker for FDA-approved drugs afatinib, erlotinib, and gefitinib in non-small-cell lung cancer. For TCGA-49-4494, afatinib and AZD9291 were predicted as therapeutic for EGFR p.T790M. Both drugs are second- and third-generation EGFR kinase inhibitors that overcome the gatekeeper mutation EGFR p.T790M. AZD9291 is not an FDA-approved drug; however, this novel compound is a potent, orally administered, selective third-generation irreversible inhibitor for EGFR mutant sensitizing (eg, p.L858R) and p.T790M-resistant mutants.²⁹ AZD9291 is currently being tested in a Phase 3 clinical trial for non-small-cell lung cancer (ClinicalTrials.gov; ID: NCT02296125). IMPACT also predicts cetuximab, erlotinib, gefitinib, ibrutinib, lapatinib, panitumumab, and vandertanib as Level 1 actionable therapeutics for this sample. [Table 2](#) lists the Level 1 actionable therapeutics for these 3 samples. [Supplementary](#)

Table 1: Somatic mutations identified by the IMPACT analysis pipeline

Type of Variants	TCGA Lung Adenocarcinoma Samples		
	TCGA-64-1681	TCGA-49-4494	TCGA-50-5944
Total Somatic Variants	21 131	27 348	18 163
Exonic	3112	3044	2350
Total SNVs	2800	2704	2118
Non-synonymous	1920	1909	1399
Frameshifts	120	176	85
Insertion	52	120	41
Deletion	68	56	43
Stop gain	124	112	91
Stop loss	2	2	7
Predicted deleterious	104	126	72
Total Copy Number Alterations	77	76	82
Copy number gain	0	0	0
Copy number loss	77	76	82

Table 2: Actionable therapeutics (Level 1) predicted by the IMPACT analysis pipeline in the 3 TCGA lung adenocarcinoma samples

Level 1 Actionable Therapeutics	TCGA-64-1681		TCGA-49-4494		TCGA-50-5944	
	Target	Drug(s)	Target	Drug(s)	Target	Drug(s)
NCI MATCH Clinical Trials	EGFR p.L858R	Afatinib	EGFR p.L858R, p.T790M	Afatinib AZD9291	EGFR p.L858R	Afatinib
MD Anderson PCT	EGFR p.L858R	Afatinib Cetuximab Erlotinib Gefitinib Ibrutinib Lapatinib Panitumumab Vandertanib	EGFR p.L858R, p.T790M	Afatinib Cetuximab Erlotinib Gefitinib Ibrutinib Lapatinib Panitumumab Vandertanib	EGFR p.L858R	Afatinib Cetuximab Erlotinib Gefitinib Ibrutinib Lapatinib Panitumumab Vandertanib
DSigDB FDA-Approved Kinase Inhibitors	EGFR p.L858R	Gefitinib Erlotinib Lapatinib Vandertanib Afatinib	EGFR p.L858R, p.T790M	Gefitinib Erlotinib Lapatinib Vandertanib Afatinib	EGFR p.L858R	Gefitinib Erlotinib Lapatinib Vandertanib Afatinib
	BTK p.Y510N	Ibrutinib	NTRK1 p.L644M	Imatinib		

Tables 4–6 provide details for Level 2 actionable therapeutics predicted for these 3 samples.

For tumor heterogeneity analysis, IMPACT estimated 62%, 53%, and 90% of tumor purity based on EGFR p.L858R allele frequency in TCGA-64-1681, TCGA-49-4494, and TCGA-50-5944, respectively. From TCGA biopsy information, these tumors have 70–80% tumor nuclei context, implying the tumor purity. The difference between IMPACT tumor purity and the TCGA biopsy is about 10%. Taken together, these results suggest that the IMPACT pipeline is capable of detecting somatic variants and connecting these variants with therapeutics in WES data.

Case Study: IMPACT Analysis of the Melanoma Patient

To illustrate the utility of the IMPACT analysis pipeline in clinical samples, we performed WES on a series of tumors obtained from a melanoma patient harboring BRAF p.V600E mutation. Three subcutaneous tissue metastases and peripheral blood were collected from the patient and WES was performed on these samples. The first sample was taken at the time of diagnosis, before the patient underwent BRAF inhibitor (BRAFi) therapy (vemurafenib). After 2 years, the patient relapsed while on vemurafenib, and another subcutaneous sample was taken at the time of relapse. The patient was then put on a BRAFi and MEKi combination study (dabrafenib and trametinib). The patient eventually relapsed while on BRAFi/MEKi combination therapy after 2 years, and a multidrug-resistant tumor sample was obtained post-BRAF/MEK combination treatment.

Single Nucleotide Variants Detection

From the IMPACT analysis pipeline, we identified 31 793, 72 408, and 33 732 somatic variants in the pre-treatment, post-BRAFi treatment, and post-BRAFi/MEKi combination treatment melanoma samples, respectively. Among these somatic variants, 2606, 6378, and 2511 non-synonymous exonic variants were identified in the pre-treatment,

post-BRAFi treatment, and post-BRAFi/MEKi treatment samples, respectively. We also identified 137, 166, and 107 frameshift variations in the pre-treatment, post-BRAFi treatment, and post-BRAFi/MEKi combination treatment samples, respectively. Table 3 summarizes the number of variations detected in all 3 samples.

In this analysis, we focused on identifying exonic somatic variants, non-synonymous somatic variants, and somatic variants predicted to be deleterious in IMPACT, as these are the candidates for cancer pathogenesis and drug targets. The numbers of somatic variants found in exonic, non-synonymous, and deleterious were 92, 306, and 39 in the pre-treatment, post-BRAFi treatment, and post-BRAFi/MEKi combination treatment samples, respectively (Figure 3 and Supplementary Tables 7–9). Interestingly, there were only 21 somatic variants found in all 3 melanoma samples (Figure 3 and Supplementary Table 10), and BRAF p.V600E is one of the common somatic variants. Only 1 common somatic variant (NRAS p.Q61K) was found in the post-BRAFi and post-BRAFi/MEKi combination treatments. NRAS p.Q61K has been previously identified as an acquired resistance mutation in vemurafenib treatment.³⁰ This suggests that NRAS p.Q61K is an acquired resistance mutation to vemurafenib treatment and persists to the post-BRAFi/MEKi combination treatment.

Copy Number Alterations

From the IMPACT analysis pipeline, we identified 473 genes that had copy number alterations, 70 of which were copy number losses in the pre-treatment sample. Both post-BRAFi and post-BRAFi/MEKi treatment samples had 67 copy number alterations, with only 1 gene (*TCF3*) amplified in post-BRAFi treatment, and no genes amplified in post-BRAFi/MEKi treatment. One notable deletion in the post-BRAFi/MEKi treated sample was *CDKN2A*, a known tumor suppressor. Figure 4 illustrates the *CDKN2A* deletion in post-BRAFi/MEKi treatment in the Integrative Genome Viewer. This is an interesting finding, as it indicates that *CDKN2A* deletion might be a novel acquired mutation for BRAFi/MEKi treatment.

Prediction of Actionable Therapeutics

Next, we used the IMPACT analysis pipeline to predict actionable therapeutics for the individual somatic mutations identified in each of the 3 samples. The presence of the BRAF p.V600E mutation in all melanoma samples linked to the combination of dabrafenib and trametinib in Level 1 actionable therapeutics. This drug combination was recently approved by the FDA as the treatment for melanoma patients with

Table 3: Somatic mutations identified by the IMPACT analysis pipeline

Type of Variants	Melanoma Samples		
	Pre-treatment	Post-BRAFi treatment	Post-BRAFi/MEKi treatment
Total Somatic Variants	31 793	72 408	33 732
Exonic	4580	11 252	9969
Total SNVs	3991	9380	9677
Non-synonymous	2606	6378	2511
Frameshifts	137	166	107
Insertion	55	86	50
Deletion	82	80	57
Stop gain	282	1455	50
Stop loss	7	53	6
Predicted deleterious	92	306	39
Total Copy Number Alterations	473	67	67
Copy number gain	403	1	1
Copy number loss	70	66	66

Figure 3: Venn diagram of the somatic variants identified by IMPACT in the 3 melanoma samples

Figure 4: CDKN2A loss detection in the melanoma samples

BRAF p.V600E mutation. Other Level 1 drugs predicted in the pre-treatment sample were dasatinib, regorafenib, sorafenib, and vemurafenib. In the post-BRAFi sample, NRAS p.Q61K (trametinib), SMO p.R703L (vismodegib), MET p.H676N (cabozantinib or crizotinib), and YES p.G294V (dasatinib) were predicted as actionable therapeutic targets by IMPACT. In the post-BRAFi/MEKi, BRAF p.V600E and NRAS p.Q61K were the only Level 1 therapeutic targets. The list of drugs for Levels 1 and 2 for the melanoma samples are summarized in Table 4 and Supplementary Tables 11–13.

Clonal dynamics and Tumor heterogeneity

Next, we explored the tumor heterogeneity and clonal dynamics of this melanoma patient under different treatments using the IMPACT analysis pipeline. We focused the analysis on the 3 gene variants: BRAF p.V600E, NRAS p.Q61K, and *CDKN2A* deletion. Allele frequency estimates indicated that BRAF mutation was present in all 3 time points, ie, even after BRAFi and BRAFi/MEKi treatment. In fact, allele frequency of BRAF p.V600E heterozygous mutation was ~0.5 in the later 2 time points, indicating that most if not all tumor cells were carrying this mutation at those stages. The NRAS mutation was detectable only after BRAFi treatment, initially with low allele frequency, indicating that it was probably present as a subclonal mutation. After BRAFi/MEKi treatment, NRAS p.Q61K showed an increase in allele frequency, indicating that the clone carrying this NRAS mutation was becoming dominant. The total loss of *CDKN2A* detectable at the third time point implied that BRAF p.V600E, NRAS p.Q61K, and *CDKN2A* deletion constituted the dominant tumor cell population after BRAFi/MEKi (Figure 5A).

We also detected changes in clonal architecture with treatment. Allele frequency distribution post-BRAFi treatment was significantly lower compared to that prior to treatment (Figure 5B, $P < 5 \times 10^{-3}$; Kolmogorov Smirnov test). It is likely that treatment-induced selection pressure had a role in shaping clonal architecture. Interestingly, both BRAFi and BRAFi/MEKi treatments led to a gain of new somatic mutations. However, the burden of newly acquired somatic mutations was higher post-BRAFi treatment (relative to pre-treatment) compared to post-BRAFi/MEKi treatment (relative to post-treatment). This could be due to changes in genome maintenance, tumor growth rate, or change in clonal makeup due to treatment-induced selection pressure. We were unable to distinguish between the possibilities. Nevertheless, our results indicate that targeted therapy has an underappreciated role in influencing clonal makeup and genetic heterogeneity within a tumor.

Validation of the identified somatic mutations

Using the qBiomarker Melanoma Somatic Mutation PCR array, we validated the mutations of BRAF p.V600E present in the pre-treatment and post-BRAFi/MEKi samples (Figure 6A). In the post-BRAFi/MEKi treatment sample, BRAF p.V600E, NRAS p.Q61K, and *CDKN2A* deletion (Figure 6B) was detected. However, NRAS p.Q61K and *CDKN2A* deletion was not detected in the pre-treatment sample. This validates that the IMPACT analysis pipeline could identify actionable mutations from WES data.

DISCUSSION

We described IMPACT, a novel whole-exome sequencing analysis pipeline that integrates the analysis of single nucleotide and copy

Table 4: Actionable therapeutics (Level 1) predicted by the IMPACT analysis pipeline in the melanoma patient over the course of treatment

Level 1 Actionable Therapeutics	Pre-treatment		Post-BRAFi		Post-BRAFi/MEKi	
	Target	Drug(s)	Target	Drug(s)	Target	Drug(s)
NCI MATCH Clinical Trials	BRAF p.V600E	Dabrafenib Trametinib	BRAF p.V600E	Dabrafenib Trametinib	BRAF p.V600E	Dabrafenib Trametinib
MD Anderson PCT	BRAF p.V600E	Dabrafenib Dasatinib Regorafenib Sorafenib Trametinib Vemurafenib	BRAF p.V600E p.P75Q	Dabrafenib Dasatinib Regorafenib Sorafenib Trametinib Vemurafenib	BRAF p.V600E p.P75Q	Dabrafenib Dasatinib Regorafenib Sorafenib Trametinib Vemurafenib
			NRAS p.Q61K	Trametinib	NRAS p.Q61K	Trametinib
			SMO p.R703L	Vismodegib		
			MET p.H676N	Cabozatinib Crizotinib		
DSigDB FDA-Approved Kinase Inhibitors	BRAF p.V600E	Dabrafenib Sorafenib Vemurafenib	BRAF p.V600E	Dabrafenib Sorafenib Vemurafenib	BRAF p.V600E	Dabrafenib Sorafenib Vemurafenib
			MET p.H676N	Cabozatinib Crizotinib		
			YES1 p.G294V	Dasatinib		

Figure 5: IMPACT tumor heterogeneity and clonal dynamic analyses. **(A)** In the melanoma patient, allele frequency changes for BRAF (p.V600E) and NRAS (p.Q61K), and CDKN2A gene copy number changes between the 3 time points. **(B)** Allele frequency distributions of somatic mutations in pre-treatment, post-BRAFi-treatment, and post-BRAFi/MEKi treatment in the melanoma patient

Figure 6: Validation of the BRAF, NRAS, and CDKN2A mutations. (A) By using qBiomarker Melanoma Somatic Mutation PCR array, we validated BRAF p.V600E is identified in the pre-treatment sample. NRAS p.Q61K was not detected and CDKN2A was detected in the pre-treatment sample. (B) In post-BRAF/MEKi treatment, BRAF p.V600E and NRAS p.Q61K were detected, and CDKN2A was not detected. This confirms the IMPACT variants and copy number detection in these 3 genes.

number variations from cancer samples. The IMPACT analysis pipeline represents a unique computational tool to analyze WES clinical samples where the somatic variants identified could be linked to therapeutic options. The allele frequency analysis in the IMPACT pipeline provides an option for users to explore tumor heterogeneity in WES samples. Here, IMPACT was utilized on 3 samples from 2 different cancers. First, IMPACT was used to analyze 3 TCGA lung adenocarcinomas with known EGFR mutations and successfully detected them. Next, IMPACT was utilized on longitudinal samples from 1 melanoma patient to explore resistance mechanisms to targeted therapy. We demonstrated the ability of IMPACT to distinguish between clinically similar tumors and offer insights into potentially new actionable therapeutics and targets.

Early last year, the FDA approved BRAF (dabrafenib) and MEK (trametinib) inhibitors for use in combination for treating BRAF mutation (V600E or V600K) metastatic melanoma patients.³¹ The BRAF/MEK combination has demonstrated a prolonged response duration as compared to single BRAF inhibitor treatment (10.5 months versus 5.6 months).³² Similar to other targeted therapy, patients initially showed promising response; however, almost all patients eventually acquired resistance to this treatment. Resistance mechanisms for the BRAF/MEK combination are currently unclear. The current understanding of resistance mechanisms in oncogene-addiction tumors such as BRAF p.V600E melanoma can occur through mutation of the drug target or activation of an alternate signaling pathway (escape pathway).

Using a melanoma patient who received treatments specific to the particular mutations present, we illustrate the utility of the IMPACT analysis pipeline in identifying the actionable mutations and the therapeutics that could be used for treatment. The post-BRAF sample

carried the highest mutational burden, of which 306 were predicted to be deleterious by all 6 tested prediction algorithms. The pre-treatment sample contained 92 variants predicted to be deleterious by all 6 tested prediction algorithms. The post-BRAF/MEKi treatment sample only contained 39 deleterious variants. The difference in the number of somatic deleterious variants between these samples follows a similar pattern to the tumor heterogeneity pattern of these samples shown in Figure 5B. The pre-treatment sample contained a biclonal architecture with a median number of variants. However, the post-BRAF treatment sample contained a greater allele frequency, correlating with one of the biclonal peaks from the pre-treatment sample along with the largest number of somatic deleterious variants. On the other hand, the post-BRAF/MEKi sample contained a more homogeneous tumor architecture, corresponding to the other clonal peak in the pre-treatment sample, with respect to the post-BRAF sample. This sample also contained the fewest somatic variants. Use of a BRAF inhibitor and clinical response suggest that a majority of the tumor cell population was killed off during this treatment. However, some tumor cells developed a subsequent NRAS p.Q61K mutation that allowed them to become resistant to BRAF inhibition. Only these cells survived, and had a subsequent lower mutational burden than the cells with a single BRAF p.V600E mutation, allowing the initial BRAF p.V600E mutation to be maintained in the tumor population. Therefore, after BRAF inhibition therapy, we hypothesized decreased tumor heterogeneity and thus a smaller, more homogenous tumor population was able to acquire the NRAS p.Q61K mutation.

Through the use of the IMPACT analysis pipeline, we identified NRAS p.Q61K as one of the acquired mutations in BRAFi treatment. This acquired resistance mutation was collaborated by previous

published papers. However, we identified CDKN2A loss as an acquired resistance mutation in the post-BRAFi/MEKi combination treatment, which has not been previously reported. CDKN2A is a known tumor suppressor commonly deleted in metastatic melanoma. In this particular sample, we believe that the CDKN2A loss is a consequence of therapy, as opposed to a driving force of tumor progression. This suggests that the post-BRAFi/MEKi combination treatment sample might use acquired mutation as the survival advantage to resist the combination treatment. Pre- and post-BRAFi/MEK treatment samples identified by IMPACT were validated by the qBiomarker Melanoma Somatic Mutation PCR array.

We suggest that IMPACT be included in the variety of analytical pipelines with which WES data can be analyzed. IMPACT is unique in allowing all 4 of the modules to be customized as well as to be run individually. Studies focused only on variant detection, copy number alterations, tumor heterogeneity, or drug predictions could be analyzed using the individual modules of the IMPACT pipeline. Linking to actionable therapeutics module is particularly powerful, as it only requires a list of variants and/or genes with copy number alterations to be used. The drug prediction module could be used as a general tool to link actionable therapeutics for any type of cancer-related NGS and “omics” data.

CONCLUSIONS

In summary, we have developed IMPACT, a novel computational analysis pipeline tool for identifying and linking actionable mutations from whole-exome sequencing of clinical samples. IMPACT is designed to be run in 4 separate modules, with an optional place for a matched normal sample in 3 modules. The first module takes in WES reads and outputs a list of predicted deleterious, exonic, and non-synonymous variants. The second module requires a matched normal sample to determine the copy number variation between the normal and tumor sample. The third module explores the tumor allele frequency and subsequent clonal architecture of the tumor. The last module takes the output from the molecular profile determined in modules 1 and 2 and links these to 2 tiers of actionable therapeutics.

We tested the IMPACT analysis pipeline on TCGA lung adenocarcinoma samples and positively identified the reported EGFR mutations in all 3 samples. Mutations identified by IMPACT could be linked to FDA-approved drugs and compounds currently tested in clinical trials as actionable therapeutics. Using IMPACT, we identified novel acquired resistance mutations to post-BRAFi and post-BRAFi/MEKi combination treatment in a melanoma patient. The mutations of BRAF, NRAS, and CDKN2A in the melanoma patient samples predicted by IMPACT were experimentally validated. Our analysis revealed loss of CDKN2A as a novel resistance mechanism to BRAFi/MEKi combination treatment and predicted potential drugs for further pharmacological and biological studies.

We believe that IMPACT, the WES analysis pipeline reported and demonstrated in this study, provides a new bioinformatics strategy to delineate candidate somatic variants and actionable therapies. As our use of TCGA data has demonstrated, this approach can be applied to other patient tumor samples to discover effective drug targets for personalized medicine.

CONTRIBUTORS

A.C.T. and W.A.R. conceived and supervised the project. A.C.T. oversaw the overall project and wrote the final manuscript. J.H. and J.K. implemented the program. J.H. performed the data analysis and interpretation and drafted the manuscript. V.Y. and S.D. performed data analysis. C.A., S.E.R., and E.S. collected the data and performed the experimental validation. A.J.A. collected the

human subject. N.B., J.W., Y.S., J.J.T., and M.M. performed data interpretations and offered suggestions for refining the analysis pipeline. W.A.R. oversaw the clinical study of the project. All authors have read and contributed to the drafting of the manuscript and approved the final version.

FUNDING

This work is partly supported by the National Institutes of Health P30CA046934 Genomics and Microarray Shared Resource (Dr Bifeng Gao, Katrina Diener, Todd Woessner, and Ted Shade); P50CA058187; the Cancer League of Colorado; the David F. and Margaret T. Grohne Family Foundation; the Rifkin Endowed Chair (W.A.R.); and the Moore Family Foundation.

COMPETING INTERESTS

None.

REFERENCES

- Karczewski KJ, Fernald GH, Martin AR, Snyder M, Tatonetti NP, Dudley JT. STORMSeq: an open-source, user-friendly pipeline for processing personal genomics data in the cloud. *PLoS ONE*. 2014;9:e84860.
- Fischer M, Snajder R, Pabinger S, *et al.* SIMPLEX: cloud-enabled pipeline for the comprehensive analysis of exome sequencing data. *PLoS ONE*. 2012;7:e41948.
- Challis D, Yu J, Evani US, *et al.* An integrative variant analysis suite for whole exome next-generation sequencing data. *BMC Bioinformatics*. 2012;13:8.
- Gao X, Xu J, Starmer J. Fastq2vcf: a concise and transparent pipeline for whole-exome sequencing data analyses. *BMC Res Notes*. 2015;8:72.
- D'Antonio M, D'Onorio De Meo P, Paoletti D, *et al.* WEP: a high-performance analysis pipeline for whole-exome data. *BMC Bioinformatics*. 2013;14(Suppl 7):S11.
- Coutant S, Cabot C, Lefebvre A, *et al.* EVA: Exome Variation Analyzer, an efficient and versatile tool for filtering strategies in medical genomics. *BMC Bioinformatics*. 2012;13(Suppl 14):S9.
- Pabinger S, Dander A, Fischer M, *et al.* A survey of tools for variant analysis of next-generation genome sequencing data. *Brief Bioinformatics*. 2014;15:256–278.
- Nam J-Y, Kim NKD, Kim SC, *et al.* Evaluation of somatic copy number estimation tools for whole-exome sequencing data. *Brief Bioinformatics*. 2015. doi: 10.1093/bib/bbv055.
- Daneshjou R, Zappala Z, Kukurba K, *et al.* PATH-SCAN: a reporting tool for identifying clinically actionable variants. *Pac Symp Biocomput*. 2014: 229–240.
- Dorschner MO, Amendola LM, Turner EH, *et al.* Actionable, pathogenic incidental findings in 1,000 participants' exomes. *Am J Hum Genet*. 2013;93:631–640.
- Amendola LM, Dorschner MO, Robertson PD, *et al.* Actionable exomic incidental findings in 6503 participants: challenges of variant classification. *Genome Res*. 2015;25:305–315.
- Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature*. 2014;511:543–550.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25:1754–1760.
- Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26:589–595.
- Li H, Handsaker B, Wysoker A, *et al.*; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078–2079.
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
- Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, *et al.* A map of human genome variation from population-scale sequencing. *Nature*. 2010;467:1061–1073.
- Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4:1073–1081.

19. Adzhubei IA, Schmidt S, Peshkin L, *et al.* A method and server for predicting damaging missense mutations. *Nature Methods*. 2010;7:248–249.
20. Shihab HA, Gough J, Cooper DN, *et al.* Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum Mutat*. 2013;34:57–65.
21. Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genetics*. 2014;46:310–315.
22. Cooper GM, Stone EA, Asimenos G, *et al.*; NISC Comparative Sequencing Program. Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res*. 2005;15:901–913.
23. Schwarz JM, Rödelberger C, Schuelke M, Seelow D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods*. 2010;7:575–576.
24. Koboldt DC, Zhang Q, Larson DE, *et al.* VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012;22:568–576.
25. Yoo M, Shin J, Kim J, *et al.* DSigDB: drug signatures database for gene set analysis. *Bioinformatics*. 2015;31(18):3069–3071.
26. Mroz EA, Tward AD, Hammon RJ, Ren Y, Rocco JW. Intra-tumor genetic heterogeneity and mortality in head and neck cancer: analysis of data from the Cancer Genome Atlas. *PLoS Med*. 2015;12(2):e1001786.
27. Yadav VK, De S. An assessment of computational methods for estimating purity and clonality using genomic data derived from heterogeneous tumor tissue samples. *Brief Bioinformatics*. 2015;16:232–241.
28. Smirnov N. Table for estimating the goodness of fit of empirical distributions. *Ann Mathematical Stat*. 1948;19(2):279–281.
29. Cross DAE, Ashton SE, Ghiorghiu S, *et al.* AZD9291, an irreversible EGFR TKI, overcomes T790M-mediated resistance to EGFR inhibitors in lung cancer. *Cancer Discov*. 2014;4:1046–1061.
30. Nazarian R, Shi H, Wang Q, *et al.* Melanomas acquire resistance to B-RAF(V600E) inhibition by RTK or N-RAS upregulation. *Nature*. 2010;468:973–977.
31. Hauschild A, Grob JJ, Demidov LV, *et al.* Dabrafenib in BRAF-mutated metastatic melanoma: a multicentre, open-label, phase 3 randomised controlled trial. *Lancet*. 2012;380:358–365.
32. Menzies AM, Long GV. Dabrafenib and trametinib, alone and in combination for BRAF-mutant metastatic melanoma. *Clin Cancer Res*. 2014;20:2035–2043.

AUTHOR AFFILIATIONS

¹Division of Medical Oncology, Department of Medicine, School of Medicine

²Division of Biomedical Informatics and Personalized Medicine, Department of Medicine, School of Medicine

³Department of Dermatology, School of Medicine

⁴Department of Pathology, School of Medicine

⁵Department of Surgery, School of Medicine

⁶Department of Biostatistics and Informatics, Colorado School of Public Health

⁷University of Colorado Cancer Center

All: University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA