Meißner *et al. BMC Medical Genomics* (2015) 8:24
DOI 10.1186/s12920-015-0095-z

BMC Medical Genomics

SOFTWARE Open Access

OncoRep: an n-of-1 reporting tool to support genome-guided treatment for breast cancer patients using RNA-sequencing

Tobias Meißner*†, Kathleen M Fisch†, Louis Gioia and Andrew I Su

Abstract

Background: Breast cancer comprises multiple tumor entities associated with different biological features and clinical behaviors, making individualized medicine a powerful tool to bring the right drug to the right patient. Next generation sequencing of RNA (RNA-Seq) is a suitable method to detect targets for individualized treatment. Challenges that arise are i) preprocessing and analyzing RNA-Seq data in the n-of-1 setting, ii) extracting clinically relevant and actionable targets from complex data, iii) integrating drug databases, and iv) reporting results to clinicians in a timely and understandable manner.

Results: To address these challenges, we present OncoRep, an RNA-Seq based n-of-1 reporting tool for breast cancer patients. It reports molecular classification, altered genes and pathways, gene fusions, clinically actionable mutations and drug recommendations. It visualizes the data in an approachable html-based interactive report and a PDF clinical report, providing the clinician and tumor board with a tool to guide the treatment decision making process.

Conclusions: OncoRep is free and open-source (https://bitbucket.org/sulab/oncorep/), thereby offering a platform for future development and innovation by the community.

Keywords: Breast cancer, Individualized medicine, RNA-Seq, n-of-1 reporting

Background

Breast cancer is the leading cause of cancer among females making up 23 % of total cancer deaths [1]. It is a heterogenous disease comprising multiple tumor entities associated with distinctive histological patterns, different biological features and clinical behaviors [2, 3]. This is driven by the fact that different breast cancer subtypes are characterized by distinct molecular, genetic, epigenetic, and transcriptional patterns (e.g. gene amplifications, in-frame fusion genes or mutations, homozygous deletions, disrupting fusions and deleterious mutations) [4]. Five year survival rates from the time of diagnosis range from 98 % (localized cancer) to 24 % (metastatic cancer). Twenty percent of patients who completed either adjuvant or neoadjuvant systemic therapy had a recurrance of the disease within 10 years after treatment [5, 6]

Molecularly profiling breast cancer tumors takes advantage of the genomic characteristics of the tumor to improve the chances of patient response to targeted agents. This enables stratification of patients based on their molecular alterations. Therapies targeting specific genomic alterations have been shown to be effective in treating specific subgroups of breast cancer patients. Examples of targeted therapies include the efficacy of Trastuzumab in *HER2*-amplified breast cancers, the mTOR inhibitor Everolimus in hormone receptor positive, *HER2*-negative patients, and the PARP inhibitor Olaparib in patients whose tumors harbor *BRCA1/2* mutations [7–10]. However, the transition to an individualized medicine approach, in which one selects the optimal treatment for a patient based on genomic information remains challenging. One of the main challenges is the translation of tumor genome-based information into clinically actionable findings. This relies not only on the identification of biologically relevant alterations that can be used as therapeutic targets or predictive biomarkers [4], but also on the

*Correspondence: meissner.t@googlemail.com
†Equal contributors
Department of Molecular and Experimental Medicine, The Scripps Research Institute, 10550 North Torrey Pines Road, 92037 La Jolla, CA, USA

availability of appropriate reporting tools. These reporting tools need to integrate the wealth of genomic data and make it usable in a routine clinical setting. This will provide additional treatment options based on the genetic nature of the patient's tumor, enabling true individualized cancer medicine.

Gene expression profiling using RNA-sequencing (RNA-Seq) is an ideal tool to assess the molecular heterogeneity of breast cancer to inform individualized medicine. It enables the estimation of transcript abundance, the detection of altered genes and molecular pathways, the detection of fusion genes and the reliable identification of genomic variants [11–15]. RNA-Seq can be performed for nearly all breast cancer and metastatic breast cancer patients that require therapy using tissue collected during routine biopsy. The main difficulties remaining for prospective use of RNA-Seq in individualized breast cancer treatment are analyzing RNA-Seq data in the n-of-1 setting and the lack of an open source reporting tool providing clinically actionable information.

To address these challenges, we developed OncoRep, an open-source RNA-Seq based reporting framework for breast cancer individualized medicine. It can be used as part of the reproducible, automated next generation sequencing pipeline Omics Pipe [16], as a standalone reporting tool or it can be adapted to existing sequencing pipelines. OncoRep includes molecular classification, detection of altered genes, detection of altered pathways, identification of gene fusion events, identification of clinically actionable mutations (in coding regions) and identification of target genes. Furthermore, OncoRep reports drugs based on identified actionable targets, which can be incorporated into the treatment decision making process. To demonstrate the feasibility of OncoRep, we produced reports based on the mRNA profiles of 17 breast tumor samples of three different subtypes (TNBC, non-TNBC and HER2-positive) which have been previously analysed and described [17–19].

Implementation

OncoRep is developed within the open-source software environments R (v3.0.2) [20] and Bioconductor (v2.13) [21] using the knitr & knitr bootstrap packages for creating the patient report in HTML format and Sweave package for creating the PDF-based report. OncoRep is distibuted via Omics Pipe [16] which handles the processing of the raw RNA-Seq data using distributed computing either on a local high performance cluster or on Amazon EC2. Installation and setup are documented online at http://pythonhosted.org/omics_pipe/.

Reference cohort

The reference cohort incorporated into OncoRep (n = 1,057) consists of 947 breast cancer samples and 106 matched tumor normal tissue samples from The Cancer Genome Atlas (TCGA), one normal breast tissue sample from the Illumina body map project (ArrayExpress accession number E-MTAB-513) and 3 normal breast tissue samples from the Gene Expression Omnibus dataset GSE52194. Level 3 gene expression data (raw read counts) were downloaded as provided for the TCGA samples. The normal samples within E-MTAB-513 & GSE52194 have been downloaded as raw sequence data (.fastq files) and processed using STAR aligner [22] and htseq-count [23] (see alignment and gene expression quantification section). Finally, to create the reference cohort, count data from all samples were merged and normalized using the Bioconductor package DESeq2 [24]. Additionally, for use in predictor generation, the data were transformed into log2 scale after adding a constant +1.

n-of-1 add-on preprocessing

OncoRep proceses a single patient sample by applying a "documentation by value" strategy [25]. This uses preprocessing information gathered from the reference cohort generated from 1,057 breast cancer samples from TCGA. Generated thresholds can be applied to a subsequent RNA-Seq patient sample, which is a prerequisite for prospective use of transcriptomics data. Add-on preprocessing of a new patient sample was done utilizing the size factor method implemented in the DESeq2 Bioconductor package [24]. Raw read counts of a new patient sample were scaled using previously stored quantitative preprocessing information from the reference cohort, thus being the geometric mean of the counts from each gene across all samples in the reference cohort. To calculate the size factor (sequencing depth) of a new patient sample relative to the reference, the quotient of the counts in the sample divided by the counts of the reference was calculated. The median of the quotients was the scaling factor for the new patient sample. Additionally, scaled read counts were transformed to log2 scale after adding a constant +1.

Quality control

Quality control (QC) of raw RNA-Seq reads was implemented using FastQC. Basic QC statistics are listed tabularly and linked to the full report generated by FastQC. Post alignment QC included computation of insert size distribution and collecting basic RNA-Seq metrics using functionalities provided by Picard tools.

Alignment

RNA-Seq reads were aligned to the human genome (hg19) using STAR aligner [22]. Alignment statistics were reported in a table within the report.

Gene expression quantification and differential expression

Gene expression quantification was done using the htseq-count function within the Python HTSeq [23] analysis package, which counts all reads overlapping known exons using hg19 annotation from UCSC (v57). To reduce the number of genes that serve as input for differential expression calling and pathway analysis we introduced the measure of gene expression reliability. Instead of using a non specific filtering step, a gene was determined to be reliably expressed when its expression value succeeded an expression cutoff. The expression cutoff was calculated based on the background distribution of all genes that were not expressed (raw read count equals 0) in the reference cohort (n = 156 genes). This method has been described by Warren et al. [26] and adopted for our use case. Differential expression was calculated based on a model using the negative binomial distribution as implemented in the DESeq2 package [24].

Prediction of receptor status and molecular subtype

Using prediction analysis for microarrays [27], predictors for breast cancer receptor status (ER, PR, HER2) and molecular subtype (Luminal A, Luminal B, Her2, Basal) were implemented using samples and clinical data (was not available for every sample) provided by TCGA. TCGA samples were randomly split up into a training cohort, on which the predictors were trained, and a validation cohort, on which to validate the predictors:

ER+ Training n = 600; validation n = 305; number of genes: 26; overall error rate training: 0.065; overall error rate validation: 0.036
PR+ Training n = 600; validation n = 302; number of genes: 28; overall error rate training: 0.133; overall error rate validation: 0.099
HER2+ Training n = 136; number of genes: 12; overall error rate training: 0.139
Subtype Training n = 346; validation n = 100; number of genes: 254; overall error rate training: 0.248; overall error rate validation: 0.218

Pathway analysis

Pathway analysis was implemented using Signaling Pathway Impact Analysis (SPIA) on the list of differentially expressed genes and their log fold changes identified in the patient sample to identify significantly dysregulated pathways using the Bioconductor packages SPIA [13] and Graphite [28]. Graphite was used to create graph objects from pathway topologies derived from the Biocarta, KEGG, NCI and Reactome databases, which were then used with SPIA to run a topological pathway analysis.

Fusion gene identification

Fusion gene identification was implemented using FusionCatcher [14]. FusionCatcher searches for novel/known fusion genes, translocations, and chimeras in RNA-seq data from diseased samples. The oncogenic potential of the detected fusion genes was predicted using OncoFuse [29].

Variant calling, filtering and annotation

Variant calling was implemented using SNPiR, a highly accurate approach to identify SNPs in RNA-seq data [15]. Basic genetic information was annotated using SnpEff [30] and information provided by dbNSFP [31]. Variants were further filtered based on being described as either common/no known medical impact in the NCBI variants database or having a MAF >0.1 in the 1000 genomes data. Identified variants were further annotated using information obtained from the following databases: the Sanger Institute's COSMIC (Catalogue of Somatic Mutations in Cancer) version 68 [32]; NCBI's ClinVar [33]; CADD (Combined Annotation Dependent Depletion) version 1.0 [34]; DrugBank version 4.0 [35]; and PharmGkb's Variant and Clinical Annotations Data [36]. Entries from these databases that exactly matched the mutated allele of a single nucleotide variant, which was called by the pipeline, were included as annotations. In addition, functional effect predictions (driver or passenger status and its likely implication in the cancer phenotype) were calculated by the IntOGen [37] pipeline and included for each variant.

Integrative drug matching

A list of all FDA approved compounds was extracted and integrated with information from DrugBank and KEGG Drug databases, which including meta information about gene targets, pathway involvements and type of drug (e.g. inhibitor, antibody, antagonist, agonist). Altered genes were matched against these data using the meta information to select appropriate drug-gene partners. Furthermore, variants were matched against SNP-drug relationships available from DrugBank and PharmGkb.

Results

OncoRep was integrated as an RNA-seq Cancer Report pipeline in Omics Pipe [16] which handles the processing of the raw RNA-seq data in an automated and parallel manner on a compute cluster. After the data were processed, the results files from each step and the patient specific meta data were automatically processed by OncoRep to produce a summary report for each patient. OncoRep performs the following analyses (Fig. 1): i) variant annotation; ii) gene expression estimation; iii) differential gene expression analysis; iv) pathway analysis; v) prediction of receptor status and molecular subtype; and vi) selection

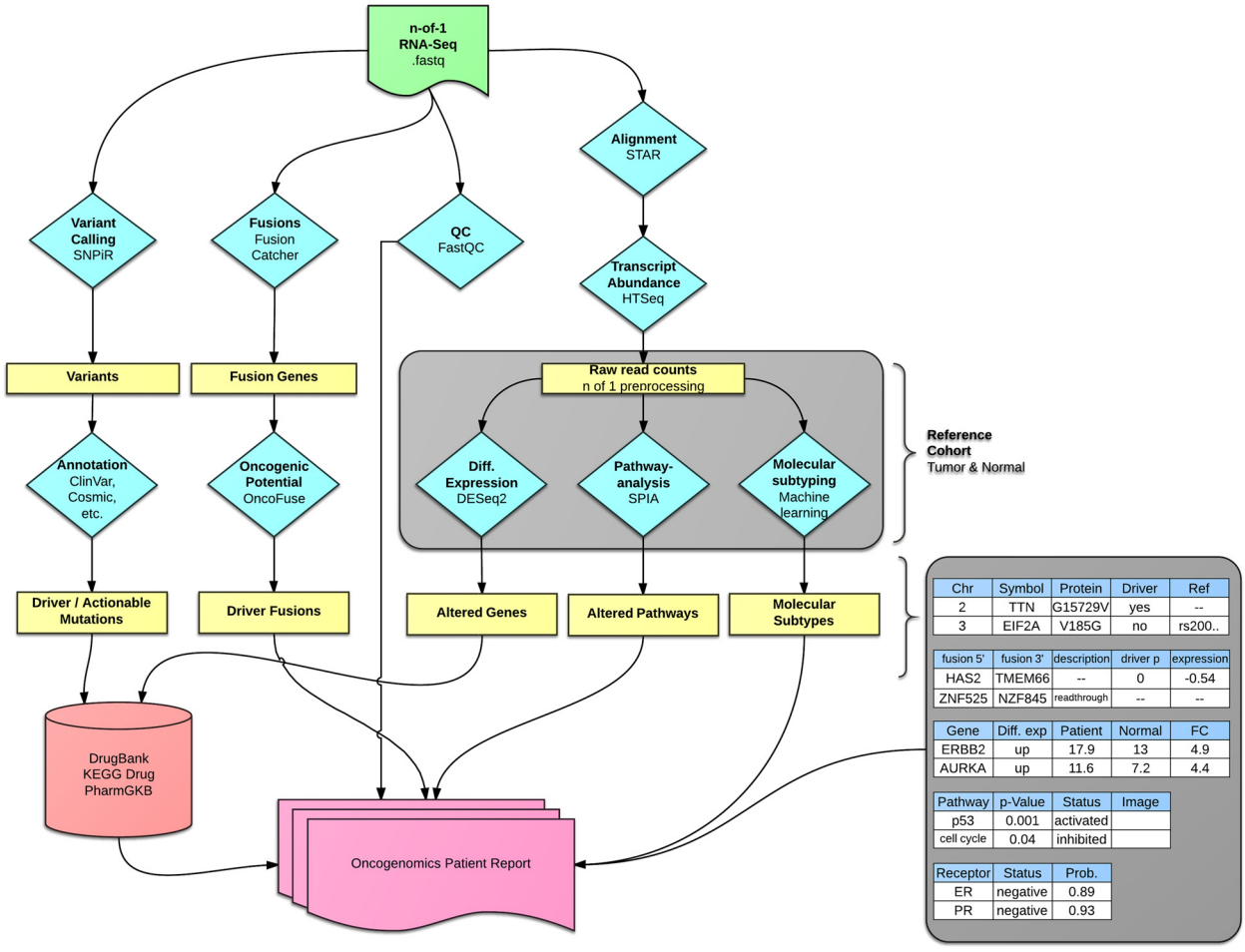


Fig. 1 Flowchart illustrating tools used and their interactions within OncoRep. The four main branches (left to right) are variant calling, fusion gene detection, quality control and gene expression quantification and analysis (for a detailed description of each step see materials and methods). Results from each branch are analyzed, annotated and integrated and an HTML report is created at the final stage of the pipeline

of drugs targeting dysregulated genes, variants and pathways. OncoRep displays these results along with the results from the quality control of the raw data and alignment, variant calling, fusion gene detection and estimation of oncogenic potential. The R package knitr is used to produce an interactive HTML report. A PDF file containing a final summary report is generated using the R package Sweave (Fig. 2). Analyzing a single patient sample (20-30 mio reads, 100bp, paired end) takes about one day in a cluster environment using four nodes.

Interactive report

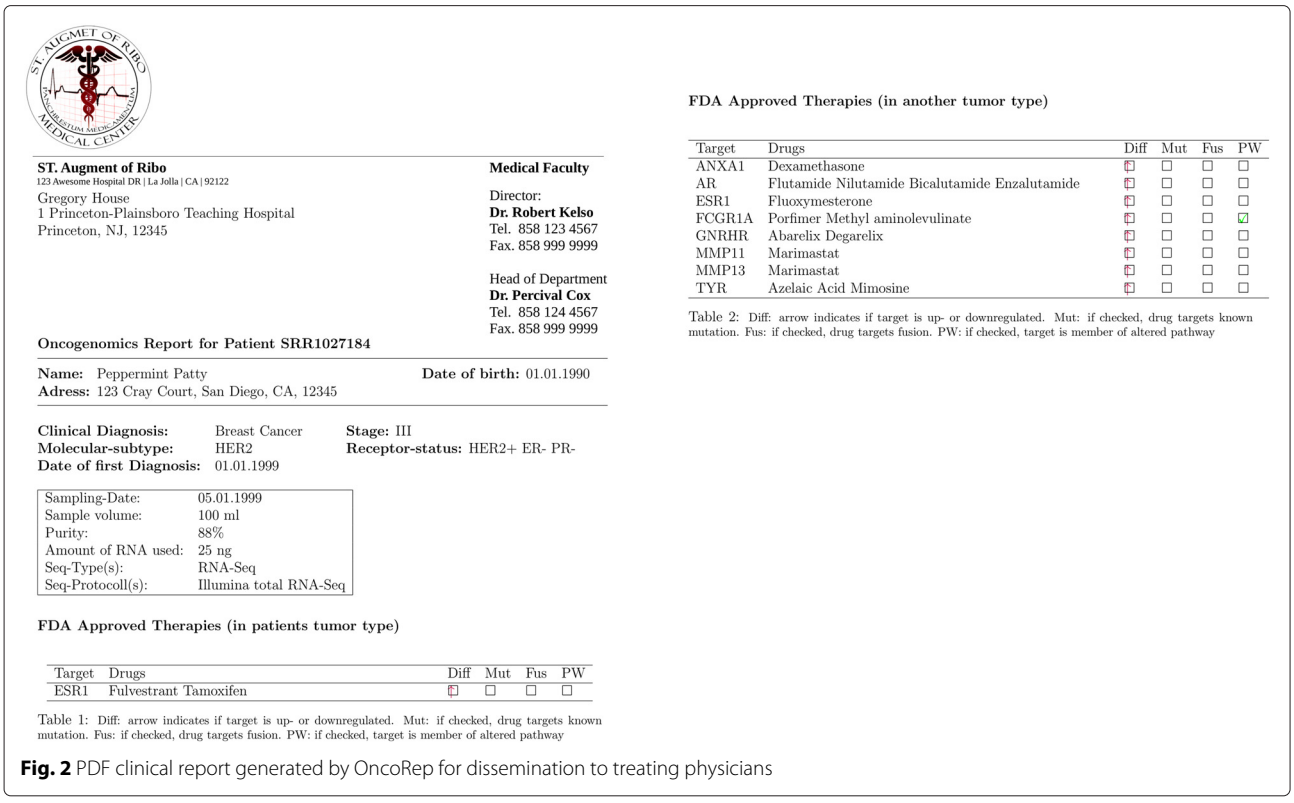
The HTML report produces interactive tables that are sortable and searchable. They can be exported as CSV files to be viewed in spreadsheet software. Gene descriptors and drugs are linked to the respective databases for easy access to further information. Pathways are visualized and they are annotated with differentially expressed genes. The interactive HTML reports for the 17 analyzed breast tumor samples can be viewed and browsed at http://sulab.org/tools/oncorep-oncogenomics-report/.

PDF Report

The PDF based report is generated in LaTeX, making it fully customizable (Fig. 2). The report, as displayed here, holds basic patient information, sample processing information and gives a list of FDA approved drugs recommended based on the altered variants, genes and pathways in a patient's tumor. An appendix holds all results from the various analysis steps in tabular form.

Quality control

OncoRep provides quality control of raw RNA-Seq reads using the FastQC tool. Basic QC results are displayed within the HTML report and linked to the detailed FastQC report for further inspection if needed (for details



ST. Augment of Ribo
123 Awesome Hospital DR | La Jolla | CA | 92122
Gregory House
1 Princeton-Plainsboro Teaching Hospital
Princeton, NJ, 12345

Medical Faculty
Director:
Dr. Robert Kelso
Tel. 858 123 4567
Fax. 858 999 9999

Head of Department
Dr. Percival Cox
Tel. 858 124 4567
Fax. 858 999 9999

Oncogenomics Report for Patient SRR1027184

Name: Peppermint Patty **Date of birth:** 01.01.1990
Adress: 123 Cray Court, San Diego, CA, 12345

Clinical Diagnosis: Breast Cancer **Stage:** III
Molecular-subtype: HER2 **Receptor-status:** HER2+ ER- PR-
Date of first Diagnosis: 01.01.1999

Sampling-Date:	05.01.1999
Sample volume:	100 ml
Purity:	88%
Amount of RNA used:	25 ng
Seq-Type(s):	RNA-Seq
Seq-Protocoll(s):	Illumina total RNA-Seq

FDA Approved Therapies (in patients tumor type)

Target	Drugs	Diff	Mut	Fus	PW
ESR1	Fulvestrant Tamoxifen	☐	☐	☐	☐

Table 1: Diff: arrow indicates if target is up- or downregulated. Mut: if checked, drug targets known mutation. Fus: if checked, drug targets fusion. PW: if checked, target is member of altered pathway

FDA Approved Therapies (in another tumor type)

Target	Drugs	Diff	Mut	Fus	PW
ANXA1	Dexamethasone	☐	☐	☐	☐
AR	Flutamide Nilutamide Bicalutamide Enzalutamide	☐	☐	☐	☐
ESR1	Fluoxymesterone	☐	☐	☐	☐
FCGR1A	Porfimer Methyl aminolevulinate	☐	☐	☐	☑
GNRHR	Abarelix Degarelix	☐	☐	☐	☐
MMP11	Marimastat	☐	☐	☐	☐
MMP13	Marimastat	☐	☐	☐	☐
TYR	Azelaic Acid Mimosine	☐	☐	☐	☐

Table 2: Diff: arrow indicates if target is up- or downregulated. Mut: if checked, drug targets known mutation. Fus: if checked, drug targets fusion. PW: if checked, target is member of altered pathway

Fig. 2 PDF clinical report generated by OncoRep for dissemination to treating physicians

see Implementation). Post alignment QC includes computation of insert size distribution and collecting basic RNA-Seq metrics using functionalities provided by Picard tools. The QC results and figures are presented within OncoRep.

Variant calling

Variants identified using the SNPiR pipeline [15] are provided in a tabular format in the HTML report. If available, the user is displayed with clinically relevant information on the variants (e.g. a matching drug or the NCBI ClinVar rating). The variants are annotated using information from SnpEff [30], dbNSFP [31], COSMIC [32], NCBI ClinVar [33], CADD [34], DrugBank [35], PharmGkb [36] and IntOGen [37] (for details see Implementation). Furthermore, variants are matched against SNP-drug relationships available from DrugBank and PharmGkb and possible hits are displayed in the table.

Fusion gene detection

Identified fusion gene candidates are provided in tabular manner in the HTML report. The information provided includes 5' and 3' fusion partners, fusion description (if available), and the the oncogenic potential prediction depicted as a p-value and expression gain/loss (for details see Implementation).

Differential gene expression

OncoRep filters out all genes estimated to have 'unreliable expression' based on the expression of a background gene set of 156 genes that are not expressed in any sample of the reference cohort (see Implementation). All remaining genes are further analyzed. Differentially expressed genes are detected by comparing the reliably expressed genes in the patient tumor to normal breast tissue samples. The results are presented in tabular format in the HTML report.

Pathway analysis

Pathway analysis is conducted based on the differential expressed genes. Altered pathways are presented in tabular form in the HTML report. Visualizations of the pathways are provided with the differentially expressed genes colored based on their log2FoldChange expression compared to normal tissue.

Receptor status

OncoRep includes predictors for the three receptors ER, PR and HER2 (see Implementation for details). A new patient sample is classified as being positive or negative for the expression of each receptor and the prediction probability is given. Results are presented in tabular format in the HTML report.

Molecular subtype

OncoRep includes a predictor for the molecular subtype of the sample (Basal, HER2, Luminal A and Luminal B). A new patient sample is classified into one of the groups and the prediction probability is given. Results are presented in tabular manner in the HTML report.

Drug matching

OncoRep reports FDA approved compounds that target the discovered differentially expressed genes, variants and pathways in the patient sample. Results are presented in tabular manner in the HTML report. Results are linked to their DrugBank and KEGG Drug entries for further investigation.

Discussion

In this article, we introduce OncoRep, a reporting tool that performs automated processing and interpretation of RNA-Seq raw data from breast cancer patients. Gene expression profiling using RNA-Seq generates vast amounts of data. This requires precise analyses and expert knowledge to generate clinically actionable information. Without expert knowledge, it remains challenging and time-consuming to do even simple data preprocessing and analysis. In a clinical setting, mosly clinically relevant data like actianable targets are needed from the RNA-Seq data. We address this problem by chaining software tools together to integrate them into a single analysis workflow that is able to deliver clinically digestable information within a short time span. OncoRep enables the prospective use of transcriptomic profiles within a clinical setting by performing molecular profiling, assessing altered genes and pathways, identifying mutations and fusion gene transcripts and by providing drug recommendations based on actionable targets to guide the treatment decision making process. This represents a critical first step towards individualized cancer treatment since it provides a reproducible approach in reporting actionable targets and allows for a quick turnaround time for real-time treatment of patients.

OncoRep detects altered genes, variants, fusions and dysregulated pathways in a patient's tumor. The challenge exists to distill this large amount of information into clinically actionable targets. OncoRep draws from several databases and employs several variant filtering and annotation steps to extract variants that are the most biologically meaningful. Integrating these databases and presenting them in a report provides the community with a valuable resource, as many databases are sparsely populated and information is distributed throughout many poorly curated databases and in the primary literature [38]. OncoRep also reports fusion genes annotated with their predicted oncogenic potential, as many fusion genes have been discovered in breast cancer that may make a substantial contribution to its development [14, 39, 40]. OncoRep uses several lines of molecular evidence to match drugs to altered drug targets in a patient's tumor by drawing on information provided by DrugBank, KEGG Drug and PharmGKB.

By distilling and reporting clinically actionable aberrations on an individual level, OncoRep provides researchers and clinicians with a powerful tool for implementing individualized medicine. For example, an OncoRep report for a patient may detect an aberration that is present in a small fraction of patients (e.g *ROS1* expression) for which targeted therapies exist. Since these are found in only a small fraction of patients, these treatments would not be used as standard of care, highlighting the importance of this method for identifying individualized treatments. In addition, OncoRep reports fusion genes and evidence exists that fusion genes may be suitable therapeutic targets. For example, Banerji *et al.* identified a recurrent *MAGI3-AKT3* fusion enriched in triple-negative breast cancer that leads to constitutive activation of AKT kinase, which can be targeted with an ATP-competitive AKT small-molecule inhibitor [39]. OncoRep advances individualized medicine by reporting all relevant information in a user-friendly way so that clinicians can access all of the results, as well as by extracting clinically actionable findings to aid in the treatment decision making process.

Conclusion

OncoRep addresses one of the main difficulties in bringing prospective use of transcriptome profiling into the clinics by creating reproducible and clinically digestible reports to guide clinical decision making. OncoRep is an open-source project, which increases the reproducibility and transparency of the analyses. A remaining problem in moving towards routine use in the clinical setting is the lack of consensus on the most accurate pipeline. OncoRep provides downstream next generation sequencing analysis and will work with any combination of aligners and variant callers. We invite researchers to use the code, refine it and provide further improvements, such as incorporating new methods and additional disease areas. We believe that offering this modular and extensible framework will provide a useful community platform for implementing individualized genomic medicine.

Availability and requirements

Project name: OncoRep

Project home page: http://sulab.org/tools/oncorep-oncogenomics-report/, https://bitbucket.org/sulab/onco rep

Operating system(s): Platform independent
Programming language: R
Other requirements: Omics Pipe (recommendet)
License: MIT

URLs

OncoRep: https://bitbucket.org/sulab/oncorep
Omics Pipe: https://bitbucket.org/sulab/omics_pipe
The R suite: http://www.r-project.org/
Bioconductor: http://bioconductor.org/
knitr: http://yihui.name/knitr/
knitr bootstrap: https://github.com/jimhester/knitrBootstrap
FastQC: http://www.bioinformatics.babraham.ac.uk/projects/fastqc
Picard tools: http://picard.sourceforge.net/
HTSeq: http://www-huber.embl.de/users/anders/HTSeq/doc/overview
FusionCatcher: https://code.google.com/p/fusioncatcher
OncoFuse: http://www.unav.es/genetica/oncofuse.html
SNPiR: http://lilab.stanford.edu/SNPiR
SnpEff: http://snpeff.sourceforge.net
Intogen: http://www.intogen.org
ClinVar: http://www.clinvar.com
DrugBank: http://www.drugbank.ca
Cosmic: http://cancer.sanger.ac.uk/cancergenome/projects/cosmic
PharmGKB: https://www.pharmgkb.org
The Cancer Genome Atlas Data Portal: http://tcga-data.nci.nih.gov/tcga

Abbreviations

RNA-Seq: Next generation sequencing of RNA; SPIA: Signaling pathway impact analysis; TCGA: The Cancer Genome Atlas; QC: Quality control.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

TM designed the research, developed OncoRep and wrote the manuscript. KF participated in designing and developing OncoRep and wrote the manuscript. LG coded the variant annotation part of OncoRep. AS designed and supervised the research and participated in writing the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by the National Center for Advancing Translational Sciences (Grant UL1TR001114). The authors thank Brian Leyland-Jones, Nicholas Schork, Casey Williams, Brandon Young, Tristan Carland and Ali Torkamani for comments and assistance.

Received: 31 December 2014 Accepted: 30 April 2015
Published online: 21 May 2015

References

1. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin;61(2):69–90. doi:10.3322/caac.20107.
2. Vargo-Gogola T, Rosen JM. Modelling breast cancer: one size does not fit all. Nat Rev Cancer. 2007;7(9):659–72. doi:10.1038/nrc2193.
3. Weigelt B, Reis-Filho JS. Histological and molecular types of breast cancer: is there a unifying taxonomy? Nat Rev Clin Oncol. 2009;6(12):718–30. doi:10.1038/nrclinonc.2009.166.
4. Natrajan R, Wilkerson P. From integrative genomics to therapeutic targets,. Cancer Res. 2013;73(12):3483–8. doi:10.1158/0008-5472.CAN-12-4717.
5. Howlader N, Noone AM, Krapcho M, Garshell J, Neyman N, Altekruse SF, et al. SEER Cancer Statistics Review, 1975-2010. 2013. http://seer.cancer.gov/csr/.
6. Brewster AM, Hortobagyi GN, Broglio KR, Kau SW, Santa-Maria CA, Arun B, et al. Residual risk of breast cancer recurrence 5 years after adjuvant therapy. J Natl Cancer Inst. 2008;100(16):1179–83. doi:10.1093/jnci/djn233.
7. Ross JS, Slodkowska EA, Symmans WF, Pusztai L, Ravdin PM, Hortobagyi GN. The HER-2 receptor and breast cancer: ten years of targeted anti-HER-2 therapy and personalized medicine. Oncologist. 2009;14(4):320–68. doi:10.1634/theoncologist.2008-0230.
8. Martin LA, André F, Campone M, Bachelot T, Jerusalem G. mTOR inhibitors in advanced breast cancer: ready for prime time? Cancer Treat Rev. 2013;39(7):742–52. doi:10.1016/j.ctrv.2013.02.005.
9. Fong PC, Boss DS, Yap TA, Tutt A, Wu P, Mergui-Roelvink M, et al. Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers. N Engl J Med. 2009;361(2):123–34. doi:10.1056/NEJMoa0900212.
10. De Mattos-Arruda L, Rodon J. Pilot studies for personalized cancer medicine: focusing on the patient for treatment selection. Oncologist. 2013;18(11):1180–1188. doi:10.1634/theoncologist.2013-0135.
11. Eswaran J, Cyanam D, Mudvari P, Reddy SDN, Pakala SB, Nair SS, et al. Transcriptomic landscape of breast cancers through mRNA sequencing. Sci Rep. 2012;2:264. doi:10.1038/srep00264.
12. Anders S, McCarthy DJ, Chen Y, Okoniewski M, Smyth GK, Huber W, et al. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. Nat Protoc. 2013;8(9):1765–86. doi:10.1038/nprot.2013.099.
13. Tarca AL, Draghici S, Khatri P, Hassan SS, Mittal P, Kim JS, et al. A novel signaling pathway impact analysis. Bioinformatics (Oxford, England). 2009;25(1):75–82. doi:10.1093/bioinformatics/btn577.
14. Edgren H, Murumagi A, Kangaspeska S, Nicorici D, Hongisto V, Kleivi K, et al. Identification of fusion genes in breast cancer by paired-end RNA-sequencing. Genome Biol. 2011;12(1):6. doi:10.1186/gb-2011-12-1-r6.
15. Piskol R, Ramaswami G, Li JB. Reliable identification of genomic variants from RNA-seq data. Am J Hum Genet. 2013;93(4):641–51. doi:10.1016/j.ajhg.2013.08.008.
16. Fisch KM, Meissner T, Gioia L, Ducom JC, Carland T, Loguercio S, Su AI. Omics Pipe: A Computational Framework for Reproducible Multi-Omics Data Analysis. Technical report August 2014. doi:10.1101/008383. http://biorxiv.org/content/early/2014/08/23/008383.abstract.
17. Eswaran J, Cyanam D, Mudvari P, Reddy SDN, Pakala SB, Nair SS, et al. Transcriptomic landscape of breast cancers through mRNA sequencing. Sci Rep. 2012;2:264. doi:10.1038/srep00264.
18. Eswaran J, Horvath A, Godbole S, Reddy SD, Mudvari P, Ohshiro K, et al. RNA sequencing of cancer reveals novel splicing alterations. Sci Rep. 2013;3:1689. doi:10.1038/srep01689.
19. Horvath A, Pakala SB, Mudvari P, Reddy SDN, Ohshiro K, Casimiro S, et al. Novel insights into breast cancer genetic variance through RNA sequencing. Sci Rep. 2013;3:2256. doi:10.1038/srep02256.
20. Team RDC. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2009. http://www.r-project.org.
21. Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004;5(10):80. doi:10.1186/gb-2004-5-10-r80.
22. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics (Oxford, England). 2013;29(1):15–21. doi:10.1093/bioinformatics/bts635.
23. Anders S, Pyl PT, Huber W. HTSeq A Python framework to work with high-throughput sequencing data. Technical report February 2014. doi:10.1101/002824. http://biorxiv.org/content/early/2014/08/19/002824.abstract.
24. Anders S, Huber W. Differential expression analysis for sequence count data,. Genome Biol. 2010;11(10):106. doi:10.1186/gb-2010-11-10-r106.

25. Kostka D, Spang R. Microarray based diagnosis profits from better documentation of gene expression signatures. PLoS Comput Biol. 2008;4(2):22. doi:10.1371/journal.pcbi.0040022.
26. Warren P, Taylor D, Martini PGV, Jackson J, Bienkowska J. {PANP} - a New Method of Gene Detection on Oligonucleotide Expression Arrays. In: Proc. 7th IEEE International Conference on Bioinformatics and Bioengineering BIBE 2007; 2007. p. 108–15. doi:10.1109/BIBE.2007.4375552.
27. Tibshirani R, Hastie T, Narasimhan B, Chu G. Diagnosis of multiple cancer types by shrunken centroids of gene expression. Proc Natl Acad Sci U S A. 2002;99(10):6567–572. doi:10.1073/pnas.082099299.
28. Sales G, Calura E, Cavalieri D, Romualdi C. graphite - a Bioconductor package to convert pathway topology to gene network. BMC Bioinformatics. 2012;13(1):20. doi:10.1186/1471-2105-13-20.
29. Shugay M, Ortiz de Mendíbil IN, Vizmanos JL, Novo FJ. Oncofuse: a computational framework for the prediction of the oncogenic potential of gene fusions. Bioinformatics (Oxford, England). 2013;29(20):2539–46. doi:10.1093/bioinformatics/btt445.
30. 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;6(2):80–92. doi:10.4161/fly.19695.
31. Liu X, Jian X, Boerwinkle E. dbNSFP v2.0: a database of human non-synonymous SNVs and their functional predictions and annotations. Hum Mutat. 2013;34(9):2393–402. doi:10.1002/humu.22376.
32. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res. 2011;39(Database issue):945–50. doi:10.1093/nar/gkq929.
33. Landrum MJ, Lee JM, Riley GR, Jang W, Rubinstein WS, Church DM, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. Nucleic Acids Res. 2014;42(Database issue):980–5. doi:10.1093/nar/gkt1113.
34. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46(3):310–5. doi:10.1038/ng.2892.
35. Wishart DS, Knox C, Guo AC, Shrivastava S, Hassanali M, Stothard P etal. DrugBank: a comprehensive resource for in silico drug discovery and exploration. Nucleic Acids Res. 2006;34(Database issue):668–72. doi:10.1093/nar/gkj067.
36. Whirl-Carrillo M, McDonagh EM, Hebert JM, Gong L, Sangkuhl K, Thorn CF, et al. Pharmacogenomics knowledge for personalized medicine. Clin Pharmacol Ther. 2012;92(4):414–7. doi:10.1038/clpt.2012.96.
37. Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, Tamborero D, Schroeder MP, Jene-Sanz A, et al. IntOGen-mutations identifies cancer drivers across tumor types. Nat Methods. 2013;10(11):1081–2. doi:10.1038/nmeth.2642.
38. Good BM, Ainscough BJ, McMichael JF, Su AI, Griffith OL. Organizing knowledge to enable personalization of medicine in cancer. Genome Biol. 2014;15(8):438. doi:10.1186/s13059-014-0438-7.
39. Banerji S, Cibulskis K, Rangel-Escareno C, Brown KK, Carter SL, Frederick AM, et al. Sequence analysis of mutations and translocations across breast cancer subtypes. Nature. 2012;486(7403):405–9. doi:10.1038/nature11154.
40. Edwards PAW, Howarth KD. Are breast cancers driven by fusion genes? Breast Cancer Res: BCR. 2012;14(2):303. doi:10.1186/bcr3122.