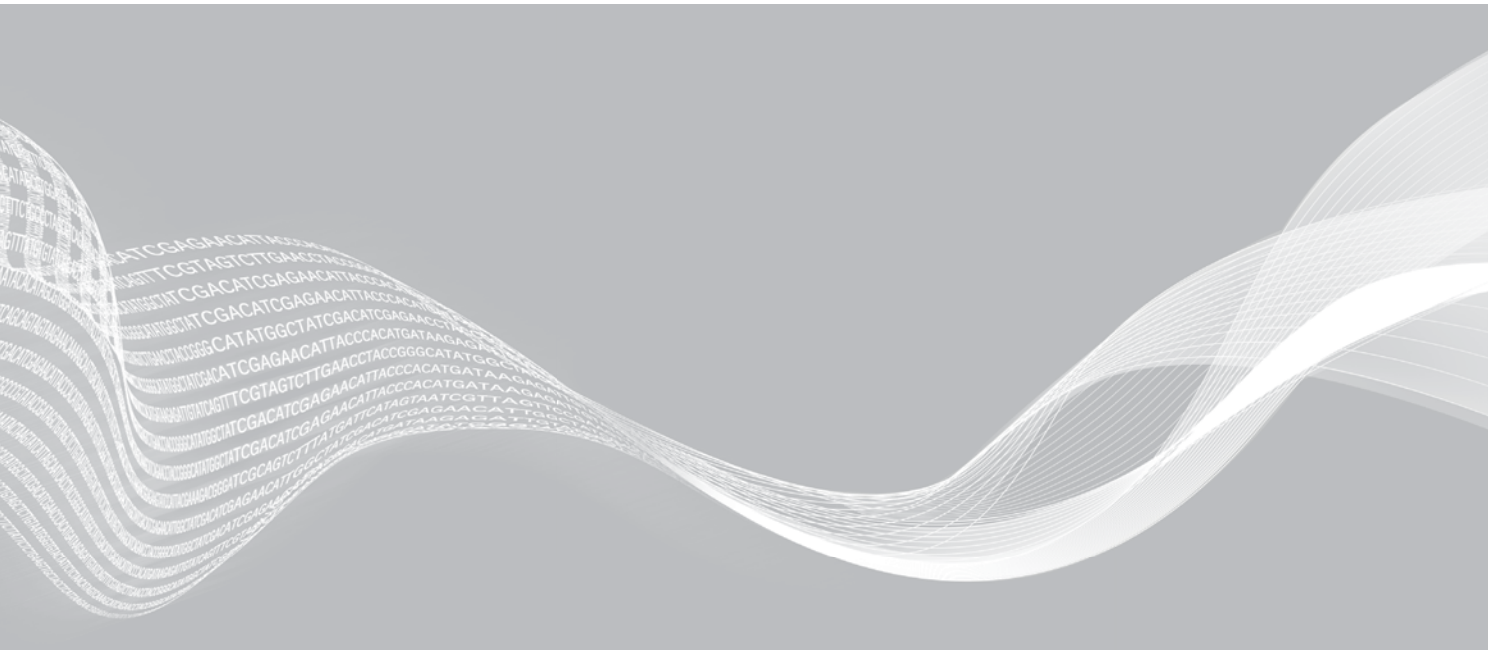
RNA-Based Drug Response Biomarker Discovery and Profiling

Section 1: Application Overview



Application solutions: RNA Drug Response Biomarker Discovery and Screening

RNA sequencing (RNA-Seq) is increasingly being utilized for the discovery of and profiling for RNA-based drug response biomarkers with the aim of improving the efficiency and success rate of the drug development process. While a number of technologies have been used for this application, the capabilities of RNA sequencing promise to be of particular benefit [1,3,4]. Consequently, there is a growing need to extend the accessibility of RNA sequencing-based workflow solutions for this application to a broader range of potential users, including those without prior experience with next-generation sequencing (NGS).

Towards that end, this document is designed to serve as a comprehensive resource for prospective users of any level of NGS experience who are considering adopting this application. It contains information that we have found to be particularly helpful to users across multiple stages of the process, from understanding the steps of an RNA sequencing workflow, to matching configuration options to specific program requirements, to preparing a plan for rapid navigation through the implementation process.

Application Overview	Workflow Introduction	Best Practices	Start-up Advice	Analysis Pipeline Review
An introduction to RNA-Seq drug response biomarker discovery and profiling	Key considerations, requirements and recommended components for multiple application use-cases	"How-to" guidance to facilitate workflow implementation	Tips from fellow application users and Illumina experts on how to get up and running quickly and smoothly	A screenshot-based walk-through from raw data through outputs needed to inform candidate assessment and prioritization

Section 1: Application overview – RNA-based drug response biomarker discovery and profiling

In the context of the historically low success rate of candidate compounds entering clinical trials, the discovery of and profiling for RNA-based biomarkers predictive of therapeutic response is increasingly becoming an integral part of the drug development process. Differential expression profiles and gene fusions detectible through RNA analysis have been shown to associate with a range of response characteristics, including efficacy, the incidence of adverse effects, pharmacodynamics, and other attributes.[2,5–7] Such biomarkers have consequently become an invaluable tool in multiple components of the development process, such as informing the interpretation of clinical trial data, allowing more efficient stratification of trial cohorts, and identifying neoantigen candidates for immunotherapy.[9–10] RNA-based biomarkers may also provide a foundation for the development of companion diagnostics, including for compounds that have previously failed clinical trials. As the breadth of biomarkers with established prognostic value continues to expand, it is critical that teams tasked with establishing and running a development pipeline are equipped to make well-informed decisions on how to integrate genome analysis into their particular programs.

Common methods

Historically, RNA-based drug response biomarker discovery and profiling has most often been performed using quantitative polymerase chain reaction (qPCR) and gene expression (GEX) arrays.

qPCR is a highly sensitive, reliable, and easy-to-use platform for focused analysis. It is commonly employed to query individual targets of interest or, in many cases, multi-assay panels. The latter often focus on functional pathways involved in drug metabolism and/or the mechanism of action of the class of compound being tested.

The primary limitation of qPCR as a biomarker discovery tool is the breadth of information that it can provide.[12] This is due in part to the practical limitations in the number of assays that can be run in parallel, and in part to the fact that assays must be pre-designed against specific targets and are therefore limited by prior knowledge of the transcriptome. In designing a biomarker discovery study, the project lead is forced to decide on what potential gene sets and functional pathways to prioritize, and thereby risks limiting the potential paths to a successful outcome.

GEX arrays have provided a complementary platform that enables transcriptome-scale analysis. The use of GEX arrays expands the scope of biomarker discovery beyond a defined set of functional pathways and provides a broader view of how compound administration impacts the transcriptome. The frequent use and applicability of GEX arrays are reflected by the extensive collection of datasets available in the public domain.

GEX array-based study designs are, however, also challenged by technical barriers.[1,13] For example, the hybridization-based chemistry on which arrays are typically based provides a relatively limited dynamic range. This is thought to be due to hybridization saturation on the high end of the expression spectrum, and by “background noise” and cross-hybridization on the low end. In the context of biomarker discovery, this limits the sensitivity of detection and visibility to candidates on both ends of the expression range, as well as the accuracy of measured differential expression between individual genes or gene sets across response categories. Further, as is the case with qPCR-based approaches, GEX array analysis is based on probes designed against known targets, so visibility to potential biomarkers requires that the features on which they are based were both known and selected for inclusion when the array was designed. This is particularly rate limiting with regard to visibility of biomarkers derived from alternative transcript isoforms and gene fusions.

Benefits of RNA-Seq-based approach

RNA-Seq has rapidly gained appeal both as a discovery and profiling tool for compound response biomarkers. As a discovery tool, it casts a wider net than other available methods. This is due to two complementary attributes: a high sensitivity of detection at both ends of the expression spectrum[1,13] and the capability to detect both known and novel features—including gene fusions and alternative transcripts—in a single assay.[1,3,4]

RNA-Seq has been shown to provide five logs of dynamic range compared to the three logs typical of GEX arrays.[1] The resulting 100x advantage in sensitivity benefits the capture of candidate biomarkers on the ends of the abundance spectrum and enables accuracy in the measurement of differential expression.

The value of the capability to detect biomarker candidates based on novel transcriptome features has become increasingly apparent. Previously unidentified transcript isoforms have been reported to associate with particular disease states and therapeutic response attributes.[14–16] Further, a growing body of literature supporting the utility of various forms of noncoding RNA as biomarkers for clinical outcomes has prompted the inclusion of whole-transcriptome analysis in more programs.[11, 17,18] In the context of cancer therapeutics, gene fusions have proven to be critical as biomarkers for response.[19–20] And while a growing number of common, well-characterized fusions have been cataloged, fusions often form as a result of spontaneous events. Pre-designed assays may be unable to detect even small variations of known targets, and cannot provide visibility to any novel events contributing to outcomes within a cohort. RNA-Seq ensures that both known and novel fusion events are captured.

Taken together, the more holistic, hypothesis-free approach that RNA-Seq enables has extended the scope of RNA-based biomarker discovery and the potential of this application to benefit the drug development process.

While the utility for biomarker discovery is well supported, it is also important to consider the value add of RNA-Seq in the framework of biomarker profiling with focused panels. Here, the features being a single platform for both the discovery and profiling phases of the process will benefit continuity and efficiency.

Barriers to adoption

While the advantages of RNA-Seq have become increasingly well-established in the industry, a number of barriers have historically challenged new users. Even though these barriers have largely become addressable as the core technology and available offerings across the workflow have evolved, it is important to assess potential gaps relative to the particular needs of your program and to confirm the solution under consideration is compatible.

One common barrier to adoption has been compatibility with particular study design requirements such as the available amounts and/or quality of input RNA. Workflows capable of addressing both requirements for the majority of users are now available (see Section 2: Workflow Introduction), though the compatibility of each option with additional requirements, such as the portion of the transcriptome being targeted and throughput capability, should also be confirmed.

Data analysis has also been a substantial barrier on multiple fronts. From the standpoint of ease of use, most pipelines that were initially developed to analyze RNA-Seq data are driven by command-line user interfaces that require informatics experts to run them. Also, the volume of data generated by RNA-Seq comes with hardware requirements to address storage needs, and a path must exist to retain the value of existing datasets by integrating RNA-Seq data with legacy qPCR, GEX array, or other data. In order to meet these requirements, pharmaceutical companies have usually either outsourced data analysis or invested in an internal bioinformatics core.

More recently available analysis solutions are able to address these requirements without the need for outsourcing or establishing a bioinformatics core via cloud-based tools. However, concerns regarding data security and, in some cases, company requirements, have frequently precluded their use. Still, the economic and workflow advantages of this path have prompted an increasing number of large pharmaceutical companies to conduct data security audits concerning cloud-based solutions and, ultimately, led to implementation in their programs (see Section 2: Workflow Introduction).

There are also more general concerns regarding potential downtime. Given the necessarily tight timeline requirements involved with drug development programs, transitioning from a well-established and productive platform to a new method that carries unique infrastructure, hardware, analysis, and training requirements introduces uncertainty and risk. For that reason, it is important that the particular needs of the program be considered at each step of a prospective workflow to ensure the right information and expertise is available to confirm a proposed solution and define a comprehensive implementation plan.

Adoption support resources: RNA-based drug response biomarker discovery and profiling

The documents described below are designed to assist users through the process of addressing the considerations outlined above, including the design and implementation of an RNA-Seq-based workflow solution for RNA drug response biomarker discovery and profiling.

Section 2: Workflow Introduction

This section introduces our recommended RNA-Seq workflows for drug response biomarker discovery and profiling, and outlines the process, from starting total RNA sample through analyzing data.

At each step, the following will be included:

- A high-level description of every step of the process
- Key points to consider when selecting a solution
- Outline of recommended solution(s)

Section 3: Best Practices

This section outlines sequencing-related design parameters that will need to be addressed ahead of planning your study. Included are considerations pertaining to read length, read depth, sequencer output modes, and other variables that should be considered to match the requirements of your program. Also captured are practical considerations related to how transitioning to RNA-Seq from platforms such as quantitative polymerase chain reaction (qPCR) and gene expression (GEX) arrays may affect day-to-day operations, and how you might best prepare.

Section 4: Start-up Advice

Experts across multiple functional areas, as well as users within the pharmaceutical industry currently running this application, offer advice to new users.

Section 5: Data Analysis Pipeline Review

Data analysis has historically been one of the most challenging barriers to the adoption of NGS workflows. This has been due, in part, to uncertainty about whether the desired endpoint for a particular application can be reached, what that process entails, and what level of expertise is required. This section provides a holistic view of our recommended analysis pipeline for this application, broken down into feature discovery, identification of biomarker candidates, and biomarker filtering and prioritization. For each work stream within the broader pipeline, a step-by-step, screenshot-based walk-through of the Illumina solution is provided.

References

1. Zhao S, Fung-Leung W-P, Bittner A, Ngo K, Liu X. Comparison of RNA-seq and microarray in transcriptome profiling of activated T cells. *PLoS ONE*. 2014;9(1):e78644. doi:10.1371/journal.pone.0078644.
2. Wang ZL, Zhang CB, Cai JQ, Li QB, Wang Z, Jiang T. Integrated analysis of genome-wide DNA methylation, gene expression and protein expression profiles in molecular subtypes of WHO II-IV gliomas. *J Exp Clin Cancer Res*. 2015;34:127. doi: 10.1186/s13046-015-0249-z.
3. Atak ZK, Gianfelici V, Hulselmans G, et al. Comprehensive analysis of transcriptome variation uncovers known and novel driver events in T-cell acute lymphoblastic leukemia. *PLoS Genet*. 2013;9(12):e1003997.
4. Kumar-Sinha C, Kalyana-Sundaram S, Chinnaiyan AM. Landscape of gene fusions in epithelial cancers: seq and ye shall find. *Genome Med*. 2015;7:129.
5. Ishikawa R, Amano Y, Kawakami M, et al. The chimeric transcript RUNX1–GLRX5: a biomarker for good postoperative prognosis in Stage IA non-small-cell lung cancer. *Jpn J Clin Oncol*. 2016;46(2):185-189.
6. Lu L, Zhang H, Pang J, Hou G, Lu M, Gao X. ERG rearrangement as a novel marker for predicting the extra-prostatic extension of clinically localised prostate cancer. *Oncol Lett*. 2016;11(4):2532-2538.
7. Perez-Gracia JL, Sanmamed MF, Bosch A, et al. Strategies to design clinical studies to identify predictive biomarkers in cancer research. *Cancer Treat Rev*. 2017;53:79-97.
8. Kantae V, Krekels EHJ, Esdonk MJV, et al. Integration of pharmacometabolomics with pharmacokinetics and pharmacodynamics: towards personalized drug therapy. *Metabolomics*. 2017;13(1):9.
9. Fang B, Mehran RJ, Heymach JV, Swisher SG. Predictive biomarkers in precision medicine and drug development against lung cancer. *Chin J Cancer*. 2015;34(7):295-309.
10. Zhao X, Modur V, Carayannopoulos LN, Laterza OF. Biomarkers in Pharmaceutical Research. *Clin Chem*. 2015;61(11):1343-1353.
11. Mishra PJ. Non-coding RNAs as clinical biomarkers for cancer diagnosis and prognosis. *Expert Rev Mol Diagn*. 2014;14(8):917-919.
12. Costa C, Giménez-Capitán A, Karachaliou N, Rosell R. Comprehensive molecular screening: from the RT-PCR to the RNA-seq. *Transl Lung Cancer Res*. 2013;2(2):87-91.
13. Perkins JR, Antunes-Martins A, Calvo M, et al. A comparison of RNA-seq and exon arrays for whole genome transcription profiling of the L5 spinal nerve transection model of neuropathic pain in the rat. *Molecular Pain*. 2014;10:7.
14. Brewer CT, Chen T. PXR variants: the impact on drug metabolism and therapeutic responses. *Acta Pharmaceutica Sinica B*. 2016.
15. Bracco L, Kearsey J. The relevance of alternative RNA splicing to pharmacogenomics. *Trends Biotechnol*. 2003;21(8):346-353.
16. Barrie ES, Smith RM, Sanford JC, Sadee W. mRNA Transcript diversity creates new opportunities for pharmacological intervention. *Mol Pharmacol*. 2012;81(5):620-630.
17. Ling H, Fabbri M, Calin GA. MicroRNAs and other non-coding RNAs as targets for anticancer drug development. *Nat Rev Drug Discov*. 2013;12(11):847-865.
18. Rönnau CG, Verhaegh GW, Luna-Velez MV, Schalken JA. Noncoding RNAs as novel biomarkers in prostate cancer. *Biomed Res Int.* 2014;2014:591703.
19. Moorman AV. New and emerging prognostic and predictive genetic biomarkers in B-cell precursor acute lymphoblastic leukemia. *Haematologica*. 2016;101:407-416.
20. Nalejska E, Mączyńska E, Lewandowska MA. Prognostic and predictive biomarkers: tools in personalized oncology. *Mol Diagn Ther*. 2014;18(3):273-284.
21. Shang C, Guo Y, Zhang H, Xue YX. Long noncoding RNA HOTAIR is a prognostic biomarker and inhibits chemosensitivity to doxorubicin in bladder transitional cell carcinoma. *Cancer Chemother Pharmacol*. 2016;77(3):507-513.
22. McCleland ML, Mesh K, Lorenzana E, et al. CCAT1 is an enhancer-templated RNA that predicts BET sensitivity in colorectal cancer. *J Clin Invest*. 2016;126(2):639-652. doi:10.1172/JCI83265.
23. Zhou M, Ye Z, Gu Y, Tian B, Wu B, Li J. Genomic analysis of drug resistant pancreatic cancer cell line by combining long non-coding RNA and mRNA expression profiling. *Int J Clin Exp Pathol*. 2015;8(1):38-52.
24. Zhao W, He X, Hoadley KA, Parker JS, Hayes DN, Perou CM. Comparison of RNA-seq by poly (A) capture, ribosomal RNA depletion, and DNA microarray for expression profiling. *BMC Genomics*. 2014;15:419.
25. SEQC/MAQC-III Consortium. A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control Consortium. *Nat Biotechnol*. 2014:32(9):903-914.

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.

illumina®