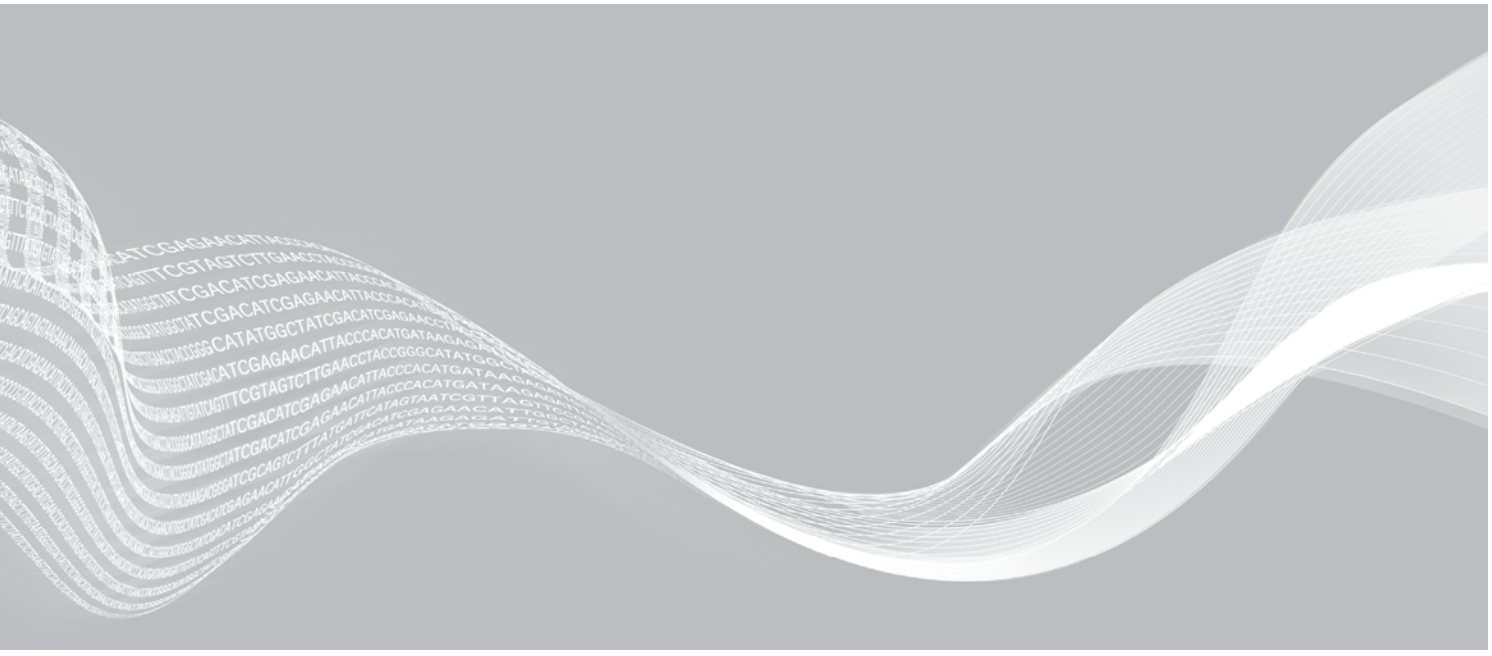
RNA-Based Drug Response Biomarker Discovery and Profiling

Section 2: Workflow Introduction



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 2: Workflow introduction

The workflows described below are designed for RNA-based drug response biomarker discovery and focused profiling. Each of the steps involved, from converting a starting RNA sample through data analysis and biological interpretation, is addressed. For each step, we provide the following:

- A description of what occurs at this step of the process
- An outline of key study design considerations to guide component selection
- The rationale for the recommended fit

In cases where needs may vary, depending on particular study design requirements, multiple component options are included.

Discovery

Step	Library prep	Sequencing	Feature detection	Biomarker candidate ID	Filtering/ prioritization
Requirements	• FFPE sample compatibility • Coding transcriptome coverage; option to capture coding and noncoding RNA • Minimal total RNA input requirement	• Solution compatible with 1s to mid 10s of samples/week • Solution compatible with 10s to low 100s of sample/week	• Measure gene, transcript expression • Detect known, novel gene fusions • Detect known, novel SNVs	• Identify expression/ response associations • Identify SNV, fusion/ response associations • Identify outliers within cohorts • Integrate RNA-Seq data w/array, quantitative PCR (qPCR)	• Identify correlations w/disease outcome (false positives) • Identify correlations w/ compounds, knock out (KO), tissue profiles • Identify known fusion, SNV gene locus associations
Component	• TruSeq® RNA Access • TruSeq Stranded Total RNA	• NextSeq® Series • HiSeq® Series • NovaSeq™ Series	• BaseSpace® Informatics Suite (RNA core apps) • Custom pipeline	• BaseSpace Informatics Suite (cohort analyzer) • Custom pipeline	• BaseSpace Informatics Suite (correlation engine) • Custom pipeline

Profiling

Step	Library prep	Sequencing	Biomarker detection
Requirements addressed	• FFPE sample compatibility • Develop custom panels for known targets • Detect novel fusions, transcripts, single nucleotide variants (SNVs) in focused regions • Minimal total RNA requirement	• Solution for known targets compatible with 10s to mid 1000s of samples/week • Solution for focused discovery compatible with 10s to mid 100s of samples/week	• Measure gene, transcript expression • Call known gene fusions, or novel gene fusions in focused regions • Call known SNVs, or novel SNVs in focused regions
Component	• TruSight® RNA Pan-Cancer Panel • TruSeq RNA Access Custom • TruSeq Targeted RNA	• NextSeq Series • MiSeq® Series • MiniSeq™ Platform	• BaseSpace Informatics Suite • On-instrument analysis • Custom/third-party pipeline

Library preparation

Description

Library prep is the process through which the total RNA input sample is converted to a form that is compatible with sequencing on an NGS instrument. In short, transcripts are fragmented, then ligated to oligonucleotides that function both to adhere to the surface of the instrument flow cell and to “tag” transcripts from individual sample sources for multiplexing.

What goes in: Total RNA sample recovered from tissue sample
What comes out: RNA-Seq “library” containing transcriptome sequencing captured within oligonucleotides compatible with the sequencing process

Key considerations

- Am I interested in coding transcriptome biomarkers only, or do I want to capture both coding and noncoding RNA (ncRNA) features?

  Coding RNA has historically been a major focus for compound response biomarker discovery, due both to the well-characterized biological function of coding RNA and the benefit to throughput scalability of directing sequencing output to a small fraction of the transcriptome. However, a growing body of literature supports the roles of multiple forms of noncoding RNA in a range of biological processes, and ncRNA-based biomarker candidates pertinent to attributes of drug response have been reported.[11, 17–18, 21–23] It is therefore important to consider the priorities of your particular program, as well as alternative options that are available, should your needs evolve.

- How many samples do I need to run on a weekly/monthly basis, and on what sequencer do I plan to run them?

  The decision of whether your studies will scope coding RNA only or coding and noncoding RNA will impact the amount of sequencing output required for each sample (*see Section 3: Best Practices*). The particular sequencer(s) that will be available to use, the volume of samples you intend to process in parallel, and your timeline for completing experiments are all important considerations.

- How much total RNA do I expect to obtain from samples in my study cohort?

  This is a critical factor in deciding on a library prep solution. As little as 20 ng of input RNA may be required, but this may vary depending on other study design needs. It is recommended that a selection be based on the anticipated lower range of available input amount across the targeted study cohorts, such that uniformity in the quality of output data can be ensured.

  It is also important to consider whether multiple methods (eg, other NGS methods, qPCR) may be run on a given sample. Selecting a solution that requires the least amount of input RNA will provide additional flexibility to run follow-up analysis.

- What is the quality (and range in quality) of samples that I expect to run?

  In many cases, particularly in studies focused on cancer-based therapeutics, formalin-fixed, paraffin-embedded (FFPE) compatibility will be a requirement. But even in cases for which the samples to be run are not FFPE-derived, use of an FFPE-compatible kit may still be advisable if a range in sample quality is expected, as this will promote consistency in the data output across the study cohort. As above, it is recommended that the "worst-case" samples from the anticipated cohort pool be considered in identifying requirements for a library prep solution.

Recommended solutions – Discovery

If coding RNA-based biomarkers will be the focus of your study, **TruSeq RNA Access** is recommended for the following reasons:

- It is FFPE-compatible. Even if you don't intend to run, or exclusively run, FFPE samples, the chemistry employed by TruSeq RNA Access will promote greater uniformity in the quality of data generated across samples of varying levels of input RNA quality.
- It requires only 10 ng of total RNA input (20 ng for FFPE samples). As above, even if a higher input amount is expected to be available for the majority of samples, this capability will broaden the range of addressable samples.
- It has a low requirement of greater than or equal to 25M reads of sequencing output per sample. Further information regarding throughput compatibility across different sequencers is included below in *Section 3: Best Practices*.
- At this level of output per sample, it is capable of detecting transcriptome features—both known and novel—across multiple biomarker categories:
  - SNVs
  - Gene fusions
  - Differential expression profiles (gene or transcript level)

If both coding and noncoding biomarkers will be scoped in your study, **TruSeq Stranded Total RNA** is recommended for the following reasons:

- It is FFPE-compatible
- It captures coding transcriptome and a variety of ncRNA species. The chemistry employed by this solution is able to output high-uniformity coverage of the coding transcriptome as well as a range of noncoding species greater than or equal to 200 nt in length, including lincRNA, which have been reported to be of particular value as biomarkers for drug response and disease states. Note, though, that this option will not capture microRNA, which can be addressed with an alternative kit and protocol.
- It maximizes efficiency of sequencing output by removing both cytoplasmic and mitochondrial rRNA and other forms of abundant RNA from the transcriptome prior to library capture.
- As stated above, it is capable of capturing multiple types of biomarkers without prior knowledge.
  - SNVs
  - Gene fusions
  - Differential expression profiles (gene or transcript level)
  - ncRNA species greater than or equal to 200 nt

Recommended solutions – Profiling

- If the biomarkers for which you wish to screen are based on gene or transcript expression, or on fusions with defined sequence, **TruSeq Targeted RNA** is the recommended solution for the following reasons:
  - It addresses required feature types with a pre-designed catalog of greater than 400,000 assays, with custom targets addressable through the **Illumina Concierge Service**.
  - Its output requirement is compatible with desktop sequencers with throughput capacity of up to 384 samples per run, meeting needs for mid- to large-scale studies.
  - Its high sensitivity and dynamic range enable reliable detection of features identified by discovery-phase studies.

- If you wish to[1]:
  - Screen for gene fusion-based biomarkers within a defined set of cancer-associated genes, but want the capability to detect novel as well as previously-observed fusions.
  - Detect SNVs and/or differential expression biomarkers within this set of genes.
  - Run FFPE-compatibility samples.
  - Run samples for which you have greater than or equal to 20 ng of total RNA.

  TruSight RNA Pan-Cancer Panel is recommended for the following reasons:
  - It includes 1385 genes that have been cited in public databases and implicated in cancer, including solid tumors, soft tissue cancers, and hematological malignancies.
  - It is compatible with desktop sequencers at the range of throughput requirements (see *Section 3: Best Practices*).

- If the above requirements apply, but you wish to target different or additional genes that are not included in the TruSight RNA Pan-Cancer Panel, the recommended solution is to design a custom **TruSeq RNA Access** panel through the Illumina Concierge Service.

Sequencing

Description

At this step, the RNA library generated from the starting sample is sequenced on an NGS sequencer. The process begins with cluster generation, in which the library is loaded into a flow cell. The flow cell is a removable glass slide with 1, 2, or 8 physically separated lanes, depending on the platform. It provides the surface on which fragments are captured on a lawn of surface-bound oligos complimentary to the library adapters. Each cDNA fragment is then amplified into distinct, clonal clusters through bridge amplification. When cluster generation is complete, the templates are ready for sequencing.

The sequencing step is based on sequencing by synthesis (SBS) technology, which utilizes a reversible terminator-based method that detects single bases as they are incorporated into DNA template strands. As all 4 reversible terminator-bound dNTPs are present during each sequencing cycle, natural competition minimizes incorporation bias and raw error rates.

What goes in: Pre-prepared RNA-Seq library
What comes out: Raw sequence (.bcl) files

Key considerations

- What volume of samples per week/per month do I expect to run?

  Anticipated project sample volumes are a key consideration. For example, you might consider whether you expect to run retrospective studies using smaller sample cohorts, or longitudinal studies scoping larger sample numbers. A mid-throughput instrument is likely to fit smaller, retrospective studies on clinical trial cohorts, whereas broader longitudinal studies of large sample populations may be a better fit for a high-throughput instrument. However, it is also important to think about per-week sample pull-through, as even a large study may fit a medium-output instrument if the samples will become available over a long period of time.

- Will I be running focused profiling studies, as well?

  If an instrument will be used for both discovery and profiling studies, it is important to assess flexibility and versatility. A high-output instrument may be a poor fit for focused profiling studies, whereas a mid-throughput instrument may be able to address discovery needs at the required sample volume, while also covering focused analysis on a lower output setting (see *Section 3: Best Practices*).

- What other methods do I, or others who will be using this sequencer, intend to run?

  Consider what additional methods a sequencer may be used to run, either in the near- or longer-term. Methods such as whole-genome sequencing or whole-genome bisulfite sequencing require substantially larger amounts of output per sample than RNA-Seq and may require a higher throughput sequencer.

Recommended solutions – Discovery

A **NextSeq Series** instrument is recommended if one or more of the following apply:

- Your throughput needs for discovery studies are in the low- to mid-range (such as for retrospective analysis of clinical trial cohorts, or larger-volume projects for which per-week pull-through is expected to be in single digits to the low 10s range).
- The primary methods that you intend to run do not include whole-genome sequencing or whole-genome bisulfite sequencing.
- You plan to run both discovery and profiling studies on the same instrument.

The fit is due in part to flexible output capabilities. The NextSeq Series enables the user to choose between 2 output modes—high-output (HO) and mid-output (MO)—that provide differing amounts of sequencing output per run (400M reads and 130M reads per run, respectively). This allows fine-tuning of instrument/facility management if, for example, the number of samples in the queue varies on a per-week basis. Additional information on sequencing modes and associated throughput capabilities are given in the *Section 3: Best Practices*.

HiSeq Series

A **HiSeq Series or NovaSeq™ Series** instrument is recommended if:

- You expect to run large sample volumes at a high, per-week pull-through rate (mid–high 10s to 100s).
- You intend to use TruSeq Stranded Total RNA (see *Library Prep* above).
- You plan to run additional methods that require high output, such as whole-genome sequencing or whole-genome bisulfite sequencing.

HiSeq and NovaSeq Series instruments are designed to serve as production level, high-throughput sequencers. While suitable for running coding RNA-focused RNA-Seq using TruSeq RNA Access for studies requiring large sample volumes, this series provides the additional flexibility to run whole-transcriptome RNA-Seq using TruSeq Stranded Total RNA—at the output requirement of less than or equal to 100M reads per sample—for mid- to large-sized studies as well. Additional details regarding models and associated output are provided in *Section 3: Best Practices*.

Recommended solutions – Profiling

MiniSeq/MiSeq Platform

If you wish to run TruSeq Targeted RNA, TruSight Pan-Cancer Panel, or a lower-complexity TruSeq RNA Access Custom panel (available through the Illumina Concierge Service), the **MiniSeq** or **MiSeq** Platform are recommended for the following reasons:

- Sequencing output per run fits requirements for each of the library prep options mentioned above (see also *Section 3: Best Practices*).
- Total run time is compatible with a wide range of project sizes and throughput requirements.
- Ease-of-use enables rapid ramp-up to proficiency for lab staff.

NextSeq Series

If you wish to run a TruSeq RNA Access Custom panel with a higher level of complexity,* the **NextSeq Series** at mid-output (MO) mode is recommended for the following reasons:

- It provides required sequencing output capacity with flexibility to accommodate mid- to large-sized study sizes (see *Section 3: Best Practices*).

*For guidance on per-sample output requirements vs panel complexity, please contact your Field Application Scientist (FAS) or Illumina Technical Support.

Data analysis

Description

This step encompasses the analysis process, starting from the time that files are generated by the sequencer through the interpretation of data. This process for RNA-based drug response biomarker discovery is broken down into 3 separate steps:

1. *Feature detection* includes the alignment of raw sequence data to the reference genome, the association of sequence data to specific genes and transcripts, the calling of SNVs, gene fusions, and other transcriptome features, as well as the quantification of gene and/or transcript abundance.
2. *Biomarker candidate identification* scopes the statistical analysis through which associations between detected transcriptome features and drug response attributes are identified. Integration of existing datasets generated by RNA-Seq, qPCR, GEX, or another method may also be performed at this step.
3. *Biomarker candidate filtering and prioritization* includes mining of public and internal datasets to identify correlations, by individual gene or gene panels, with transcriptome features observed in response to other drugs and candidate compounds, disease states, in gene knockout models, and under other conditions. Screens for enrichment of genes in known functional pathways may also be performed.

What goes in: Raw sequence files (.bcl)
What comes out: Prioritized biomarker candidates

Key considerations

- What types of biomarkers do you wish to identify?

  You should confirm that your analysis pipeline includes the functionality required to detect the full scope of feature types that you wish to consider as candidate biomarkers. A consultation with an Illumina bioinformatics support specialist is recommended early in the adoption process to ensure that your analysis plan is optimally tailored to your program needs.

- Who will perform data analysis?

  Pharmaceutical companies among the early adopters of NGS typically either set up and maintained cores staffed by bioinformatics experts or outsourced the management of data analysis needs. The advantages of this approach have been that analysis pipelines could be run, or even developed and tailored to suit the needs of individual programs. With respect to RNA-Seq, since the first available analysis tools were operated using command-line interfaces, access to in-house or external experts was often a requirement.

  The challenge with this approach has been that it introduces a gap within the workflow. The frontline personnel who design the study, perform the experiment, and ultimately interpret the data and determine the actions taken based on the results are required to hand off the experiment at a critical stage of the process. In many cases, data analysis may require manipulations that benefit from first-hand knowledge of the biology, experimental scope, and goals. For example, optimizing differential expression thresholds and mining genes, transcripts, or functional pathways that emerge in each instance as potential biomarker candidates may require multiple iterations of an analysis. Other practical barriers may exist as well, such as the availability of common resources for which multiple groups drive demand. Further, for a company that is just getting started with an NGS-based application, investing in an in-house bioinformatic core may not be a practical option. For these reasons, analysis options that are directly accessible to frontline, bench biologists and experimental teams may have appeal.

  If it is preferred that staff with no bioinformatic expertise are able to run the analysis from start to finish, or even that they are sufficiently familiar with the pipeline to be able to analyze and manipulate the data, you might consider graphic user interface (GUI)-based, click-and-go solutions that meet your functionality requirements.

- Where will your data be stored?

  Unlike qPCR in particular, the output files generated by RNA-Seq require considerable storage space, especially if large volumes of data will be processed by a particular lab or facility. For this reason, it is critical to assess your needs (see *Section 3: Best Practices*) and/or consult with an expert to ensure that your infrastructure is equipped to address your needs.

- Are data security concerns met?

  It is critical that your data analysis and storage solutions are properly vetted to ensure that they meet your institutional facility data security requirements. One key consideration is whether data will be analyzed and stored locally, or if a cloud-based system is a potential option. Company policy restrictions within the pharma industry have often precluded cloud-based solutions, and integrating with the onsite infrastructure has been a standard requirement.

  While this is still the case for some companies, a number of "big pharmas" have conducted data security audits and subsequently implemented cloud-based data analysis and storage. Examples of specific security concerns that have been addressed are listed in Transition Guidance in *Section 3: Best Practices*.

  If a cloud-based system is an option, there are a number of potential advantages, particularly in cases for which an NGS infrastructure is being built from scratch. In addition to bypassing the need to invest in hardware for computing and storage, cloud-based capability enables access to a number of GUI-based analysis options that address each of the steps in the biomarker discovery and profiling process (see *Section 5: Data Analysis Pipeline Review*).

Recommended solutions

BaseSpace Informatics Suite is recommended if:

- You do not have access to in-house bioinformatic expertise.
- You would prefer that personnel who design and run experiments are able to run your analysis pipeline as well as manipulate and analyze data.
- Your preferred data analysis software is either included in the BaseSpace Informatics Suite or can be enabled in BaseSpace Sequence Hub. (Please contact an Illumina bioinformatics support specialist for information.)
- You wish to implement a pipeline that covers feature detection/discovery, biomarker candidate identification, and candidate filtering/prioritization, as described above.
- (For cloud-based option only) Your data security requirements can be addressed.*

*See Data Security in *Section 3: Best Practices*.

Custom/third-party analysis pipelines are recommended if:

- You have an internal bioinformatic resource, and staff with no bioinformatic experience are not required to run your analysis pipeline.
- You have unique analysis needs and prefer to custom-design your own pipeline.
- Due to data security concerns, you are unable to use a cloud-based analysis solution, and BaseSpace Informatics Onsite Sequence Hub does not meet your needs. (For information regarding capabilities and program fit, please contact an Illumina bioinformatics support specialist.)

LIMS

Description

A laboratory information management system (LIMS) is a software-based lab and workflow support infrastructure that is frequently used within the pharma industry for sample tracking and data management. *(See Section 3: Best Practices)*

What goes in: (depending on the LIMS solution used)

- Source/sample identity
- Date and time performed
- Reagent lots used by sample/run
- Operator identity
- Results
- Other run/workflow information

What comes out: Results that include transcriptome features, and from that, identified biomarker candidates present in each, with the history of each sample tracked and stored to ensure data reliability and availability for future access and use.

Key considerations

- Do you need LIMS?

  Determining whether LIMS makes sense for your lab depends on a number of variables, and may be best approached by first identifying the challenges you face. A LIMS solution is recommended if one or more of the following are true:

  - The level of throughput/volume of sequencing data you generate is rapidly climbing and becoming difficult to manage.
  - You are using manual methods to organize and track samples and are concerned that errors and loss of data may result.
  - Workflow and data management are taking too much of your lab staff time.
  - Your data storage and retrieval infrastructure is inadequate or may become so as your operation scales up.

You can find additional information about LIMS at http://www.genologics.com/what-is-a-lims/

- Should you purchase a LIMS solution or build your own?

  There are advantages to building your own LIMS. Primarily, it enables full customization—your lab can build a LIMS that matches all of your needs. Furthermore, when you build your own system, you can integrate precisely with the instruments and tools in use at your lab.

  However, this path includes risks. Building a LIMS with such specificity can lead to rigid code structures that are difficult to modify if workflows or technologies change. Additionally, because internal staff have other priorities, the planning and implementation process can extend the start-up time. Since internal resources will have to maintain whatever LIMS is built, that may involve additional staffing and resourcing considerations. For example, it will be necessary to ensure that the knowledge and experience required to manage, troubleshoot, and update the LIMS are captured and easily transferable as staff evolves.

  You can find additional information at https://www.genologics.com/blog/lims-build-or-buy/

Recommended solutions

Clarity LIMS is recommended if one or more of the following apply:

- You want to start with core functionality, with options to scale as needs change.
- You want your lab staff to be able to manage your workflow from library prep through analysis, without assistance from bioinformatics experts.
- You want to integrate multiple components of your workflow.
- It is important to expedite your implementation process.

Functionality that suits labs of different sizes and operational requirements is built into pre-designed, modular solutions sets. User interface can be designed for a rapid learning curve and optimum ease-of-use. Full support is provided by specialized staff and infrastructure dedicated to this industry.

A **custom LIMS** is the recommended solution if the following conditions apply:

- You have a dedicated internal resource to develop and manage your LIMS
- Your lab has unique needs that you prefer to address with a custom solution
- You have an existing LIMS solution (eg, legacy systems, something has already been built)

No LIMS solution is needed if the following conditions apply:

- Low sample volume is manageable without sample tracking
- Microsoft Excel or alternative tracking option covers your needs

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®