Custom Publishing From:

TheScientist
EXPLORING LIFE, INSPIRING INNOVATION

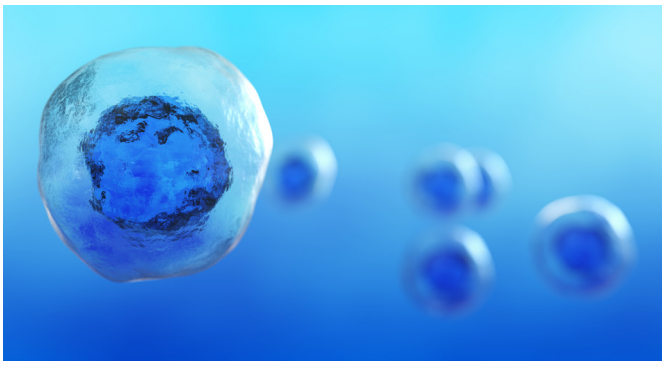
Sponsored By:

BD

HITTING A BIOLOGICAL BULLSEYE: PERFECT YOUR AIM WITH TARGETED SEQUENCE ANALYSIS

Standing Out in a Crowd: Single-Cell Targeted Analysis

"...the ability to quickly identify an individual patient's disease genotype would be invaluable to developing a personalized treatment course."



Same Type Doesn't Mean Exactly Alike

Genomic analysis is widely used in research to search for underlying mutations that contribute to disease. However, many analysis techniques derive their findings from multicellular samples, which means that key data from individual cells or rare subsets can be buried and lost amongst a summarization of the mean.[1] Cellular heterogeneity plays a critical role in many systems and pathologies, including immune system function and oncopathogenesis.[2,3] Detecting and understanding cell-to-cell differences is absolutely essential to not only identifying key mechanisms of pathogenesis, but also to developing therapeutic strategies with an eye towards personalized medicine.

The Challenges of Looking at Individual Cells

Early attempts to isolate, detect, and analyze genetic differences on the individual-cell level using techniques such as fluorescence in situ hybridization yielded promising findings, but these techniques were limited in the number of genes that could be simultaneously probed.[4] Single-cell whole transcriptome amplification analysis (WTA) techniques can overcome this limitation, and are more accessible now with the advent of next-generation sequencing (NGS) techniques.[4,5]

However, cost effectiveness per-run presents an obstacle for researchers using NGS-based single-cell WTA techniques.[6] Furthermore, WTA techniques generate a tremendous volume of data, with not all of it directly relevant to a researcher's specific needs and aims. Analyzing the sequencing information and then parsing through the analysis report to obtain the relevant portions can take significant amounts of time.

The Solution

The solution to these challenges lies in utilizing a more focused "targeted" analysis strategy. Targeted analysis is the active prioritization of selected genomic regions (e.g., specific genes or intragene regions) which are associated with a scientist's research aims–such as a specific cell type or a certain pathology.[6-8] In this way, it gives the researcher a smaller, more manageable, but much more relevant dataset to work with, in shorter amounts of time while consuming fewer resources and leaving a larger budget for downstream analyses. The approach sequences key regions of interest to greater depths, resulting in greater sensitivity and allowing for the identification of novel or rare genetic variants, as well as their linkage to pathogenesis.[9,10]

Best Practices

Targeted analysis is typically performed using designed gene panels that generally include between several dozen to a few hundred select genes linked with a particular cell, system, or disease of interest. Pre-assembled gene panels are available commercially, with many companies also offering custom gene panel construction services.

Single-cell targeted analysis allows for the efficient delineation and characterization of cellular heterogeneity, thus potentiating the revelation of deeper insights into physiological and pathological mechanisms.[5] Targeted analysis offers attractive potential as a clinical diagnostic tool[11,12]–the ability to quickly identify an individual patient's disease genotype would be invaluable to developing a personalized treatment course. The sensitivity and accessibility of single-cell targeted analysis is opening new windows and offering new promise to basic, translational, and clinical researchers alike.

References

1. S.J. Altschuler et al., "Cellular Heterogeneity: Do Differences Make a Difference?" *Cell*, 141(4): 559-563, 2010.
2. N.E. Navin, "Cancer genomics: one cell at a time," *Genome Biol*, 15: 452, 2014.
3. E. Papalexi and R. Satija, "Single-cell RNA sequencing to explore immune cell heterogeneity," *Nat Rev Immunol*, doi:10.1038/nri.2017.76, 2017.
4. F. Buettner et al., "Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells," *Nat Biotechnol*, 33:155-160, 2015.
5. C. Korfhage et al., "Whole-Transcriptome Amplification of Single Cells for Next-Generation Sequencing," *Curr Protoc Mol Biol*, 111:7.20.1-19, 2015.
6. J.R. Golbus et al., "Targeted Analysis of Whole Genome Sequence Data to Diagnose Genetic Cardiomyopathy," *Circ Cardiovasc Genet*, 7(6): 751-759, 2014.
7. H.C. Howard et al., "Whole-genome sequencing in newborn screening? A statement on the continued importance of targeted approaches in newborn screening programmes," *Eur J Hum Genet*, 23(12):1593-1600, 2015.
8. J. Kammermeier et al., "Targeted gene panel sequencing in children with very early onset inflammatory bowel disease--evaluation and prospective analysis," *J Med Genet*, 51(11):748-755, 2014.
9. F. Xu et al., "Whole-exome and targeted sequencing identify ROBO1 and ROBO2 mutations as progression-related drivers in myelodysplastic syndromes," *Nat Commun*, 6:8806, 2015.
10. T. Mantere et al., "Targeted Next-Generation Sequencing Identifies a Recurrent Mutation in MCPH1 Associating with Hereditary Breast Cancer Susceptibility," *PLoS Genet*, 12(1):e1005816, 2016.
11. J.R. Lemke et al., "Targeted next generation sequencing as a diagnostic tool in epileptic disorders," *Epilepsia*, 53(8):1387-1398, 2012.
12. Y. Liu et al., "Targeted Next-Generation Sequencing for Clinical Diagnosis of 561 Mendelian Diseases," *PLoS One*, 10(8): e0133636, 2015.

OPTIMIZING SINGLE-CELL GENE EXPRESSION ANALYSIS

Sequencing data has proven invaluable for finding genetic variations, but using aggregated RNA samples can mask individual cell-level variations. Single-cell sequencing is the solution, and targeted analysis helps bridge the logistical gap between bulk and single-cell sequencing. Explore how to undertake and optimize a targeted single-cell analysis workflow!

ENRICH YOUR CELL POPULATION FOR MORE ACCURATE RESULTS

Cells of interest can be very few in number and hard to detect. Screen and sort your cells (e.g., using fluorescence-activated cell sorting) to enrich your starting sample.

TOP-QUALITY RESULTS START WITH TOP-QUALITY SAMPLE

Limit manipulations as much as possible to maintain cellular phenotypic integrity and generate a sample better representative of cells in vivo.

PICKING THE RIGHT AMPLIFICATION METHOD

Proper amplification is absolutely critical for data integrity. Multiple methods exist, each with their own pros and cons. Pick a method that suits your needs, limits amplification bias, and ensures product specificity.

MOLECULAR INDEXING EXPOSES SINGLE-CELL HETEROGENEITY

A genetic variation within a single cell can easily get lost. Molecular indexing labels mRNA molecules to trace mutations/variations and measure amplification bias.

TARGETED ANALYSIS PROVIDES GREATER RUN EFFICIENCY

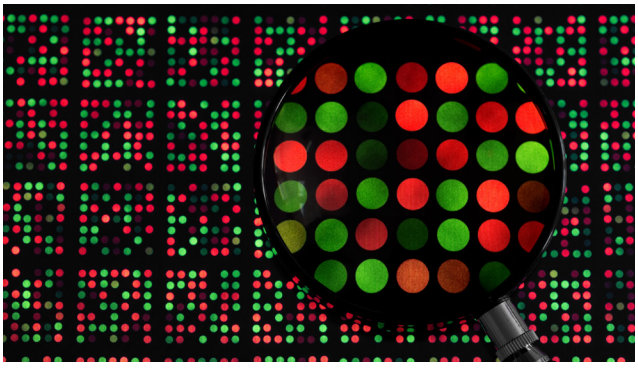
Not all sequencing data may be immediately useful. Targeted analysis prioritizes limited panels of relevant genes, resulting in more robust datasets, conserving sample, and reducing run frequency.

SEQUENCING, VISUALIZATION, AND THE RIGHT SOFTWARE

DNA sequences are read and assembled after amplification. High read counts per cell are important for ensuring that the constructed assemblies are representative. Top-end software can present researchers with cell phenotype profiles and expression variability ranges.

TARGETING GREAT DISCOVERIES

Building Your Experiment to Focus on Your Target Biology



"Targeted analysis using gene panels gives researchers the option to cast a more stringent net and make optimal use of their time and energy."

The Problem with Whole Transcriptome Analysis

Whole transcriptome analysis (WTA) has become increasingly popular for basic and translational research, as scientists have flocked to the increased resolution and higher reproducibility offered by WTA compared to older microarray-based approaches.[1] However, WTA is not always the most efficient option available. The majority of reads can be devoted to highly expressed genes exhibiting low variance, and the amount of generated data can easily exceed researcher requirements. Dealing with these issues is time, resource, and labor intensive, and does not contribute to our understanding of the impact of cellular heterogeneity in physiology and in disease.

Targeted analysis using gene panels gives researchers the option to cast a more stringent net and make optimal use of their time and energy. While WTA may detect variations in thousands, if not tens of thousands, of genes, research indicates that the bulk of transcription at the tissue level is dominated by only a few hundred genes.[2] Similarly, the number of genes conclusively demonstrated to be linked with a specific condition/pathology or affecting the function of a given cell type typically numbers in the hundreds at most.[3] Gene panel-based NGS methods dedicate analytical priority upon these smaller, more focused sets of genes, resulting in a more sensitive, relevant, and accessible output.

The Gene Panel-based NGS Solution

Gene panels are particularly useful for delineating cell-to-cell heterogeneity and identifying rare cell subsets.[4] They are also applicable in clinical contexts where multiple diseases present overlapping phenotypes or are associated with genes possessing a common pathway,[3] and can be used to ascertain patient sensitivity or resistance to specific treatment options (e.g., checkpoint or growth-factor inhibitors).[5]

During the infancy of targeted analysis, research teams undertook the painstaking process of constructing their own gene panels. Today, commercial ready-to-use gene panels offer options for researchers looking for specific mutation types (e.g., single nucleotide variants, fusions) and/or specific fields (e.g., cancer, cardiovascular disease, inherited conditions). Gene panels can also be created for non-human species, making them particularly useful for basic and translational researchers working with animal models.

Are Gene Panels for You?

However, pre-designed gene panels are assembled based on established research, and while they're specialized to a degree, they cannot meet the exact needs of every researcher.[7] Custom gene panel construction services are available for those at the forefront of their fields seeking tools tailored precisely to their aims. Most manufacturers also offer consultation during the gene selection process to meet researcher demands.

Gene panels can vary in size from several dozen to several hundred genes,[3] depending on researcher needs. Larger gene panels are more resource intensive, but can potentially yield more comprehensive data and the possibility of discovering new variations. They are therefore more suitable for researchers seeking novel genetic markers or attempting to identify mechanistic causes of pathogenesis. Conversely, smaller gene panels typically encompass genes with stronger links to the condition of interest,[3] resulting in more relevant and rigorous datasets. This more selective approach is useful to establish a firm causal link between a variation and a phenotype, and is the recommended course of action for gene panel-based NGS methods used for clinical diagnostic purposes.[6]

References

1. I.S. Yang and S. Kim, "Analysis of Whole Transcriptome Sequencing Data: Workflow and Software," *Genomics Inform*, 13(4): 119-125, 2015.
2. M. Melé et al., "The human transcriptome across tissues and individuals," *Science*, 348(6235): 660-665, 2015.
3. Y. Xue et al., "Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-gene, gene panel, or exome/genome sequencing," *Genet Med*, 17(6):444-451, 2015.
4. C. Blassl et al., "Gene expression profiling of single circulating tumor cells in ovarian cancer - Establishment of a multi-marker gene panel," *Mol Oncol*, 10(7):1030-1042, 2016.
5. C.A. Cummings et al., "The Role of Next-Generation Sequencing in Enabling Personalized Oncology Therapy," *Clin Transl Sci*, 9(6):283-292, 2016.
6. G. Matthijs et al., "Guidelines for diagnostic next-generation sequencing," *Eur J Hum Genet*, 24(1): 2-5, 2016.
7. C. Rubio-Perez et al., "Rational design of cancer gene panels with OncoPaD," *Genome Med*, 8:98, 2016.

Filtering Out the Noise: Maximizing Data Analysis Accuracy

"The integrity of the extracted nucleic acid sample is critical to the ultimate integrity of the sequencing data..."

Data Quality Relies on Sample Integrity

Single-cell targeted analysis offers a more streamlined and precise examination of genomic variation, but generates a significant amount of information that must be analyzed in a legitimate and proper manner. The quality of targeted analysis-generated sequencing data is affected by many of the same variables that impact WTA, including sample preparation, amplification strategy, and how fragments are read and sequenced.

Targeted analysis is subject to the same core tenet as any experimental workflow: input begets output. The integrity of the extracted nucleic acid sample is critical to the ultimate integrity of the sequencing data, and care must be taken to avoid degradation and contamination. Samples should not be subjected to excess heat (> 60 °C) more than necessary, and freeze/thaw cycles should be avoided. In addition, reagents used during the extraction process (e.g., alcohols, salts, chelators, or detergents) that would detrimentally affect RNA integrity post-extraction should be removed as thoroughly as possible. Finally, steps should be taken to limit or inhibit any potential residual nuclease activity.[1]

Analyzing single-cell transcriptomes adds a layer of complexity to the NGS workflow. In addition to avoiding external contamination from handling or reagents, care must also be taken to avoid the presence of cellular aggregates, fragments, or debris. The vast majority of the cells should be viable throughout the extraction process.[2,3]

Overcoming Challenges with Fluorescence-activated Cell Sorting

Fluorescence-activated cell sorting (FACS)-based methodologies can precisely characterize individual cells within a heterogeneous sample, making it easy to isolate the exact cell of interest. FACS's capacity for automation and multiplexing also makes it ideal for high-throughput NGS workflows.[4] FACS does require larger sample volumes (i.e., more cells), and additionally does not facilitate the direct visualization of cells post-sorting, so it is essential to identify and exclude doublets or multiplets by looking at light scatter measurements prior to NGS.[4-6]

Recent advances have brought microfluidics-based techniques to the fore for single-cell omics analysis. Here, proprietary technologies are used to segregate and sort single cells into individual chambers for further processing and/or nucleic acid amplification.[4,6] Microfluidics drastically lowers required sample volumes down to nanoliter-ranges, and some microfluidics-based approaches allow cells to be visually monitored post-segregation.[4]

When interpreting NGS data, make sure that an over- or under-expressed gene is a true phenomenon and not the result of amplification bias—where amplification extent across the genome is uneven. Amplification bias is a challenge for all NGS experiments, with primer design/amplicon complexity, GC content, gene/fragment length, and gene expression levels/read depth all potentially affecting amplification extent.[7-9] Single-cell analysis is particularly susceptible to this issue, since lower levels of starting materials necessitate more extensive amounts of amplification.[7]

Best Practices

One tactic to detect and compensate for amplification bias is the use of unique molecular identifiers (UMIs), which are short sequences (sometimes referred to as "barcodes"[4]) ligated onto the ends of nucleic acid fragments prior to amplification.[7,10] Counting the number of UMIs rather than the number of reads thus mitigates amplification bias.

Given the vulnerabilities of single cell-based techniques, it is critically important to have a firm understanding of the steps necessary for input RNA to become output data. This will give researchers the ability to not only recognize when variabilities, biases, and errors have crept into the final results, but also to take precautions during the workflow to guard against such events.

References

1. C. Endrullat et al., "Standardization and quality management in next-generation sequencing," *Appl Transl Genom*, 10: 2-9, 2016.
2. A. Haque et al., "A practical guide to single-cell RNA-sequencing for biomedical research and clinical applications," *Genome Med*, 9(1):75, 2017.
3. A. Guillaumet-Adkins et al., "Single-cell transcriptome conservation in cryopreserved cells and tissues," *Genome Biol*, 18(1):45, 2017.
4. D. Grün and A. van Oudenaarden, "Design and Analysis of Single-Cell Sequencing Experiments," *Cell*, 163(4):799-810, 2015.
5. R.P. Wersto et al., "Doublet discrimination in DNA cell-cycle analysis," *Cytometry*, 46(5):296-306, 2011.
6. I.C. Macaulay et al., "Single-Cell Multiomics: Multiple Measurements from Single Cells," *Trends Genet*, 33(2): 155-168, 2017.
7. B. Phipson et al., "Gene length and detection bias in single cell RNA sequencing protocols," *F1000Res*, 6:595, 2017.
8. V. Bhargava et al., "Technical variations in low-input RNA-seq methodologies," *Sci Rep*, 4:3678, 2014.
9. M.G. Ross et al., "Characterizing and measuring bias in sequence data," *Genome Biol*, 14(5):R51, 2013.
10. T. Kivioja et al., "Counting absolute numbers of molecules using unique molecular identifiers," *Nat Methods*, 9(1):72-74, 2011.