PUBLICATION NOTE

Illuminating transcriptional activities of lowly-expressed genes using myBaits® RNA-Seq Capture

INTRODUCTION

Protocadherin genes, which exhibit stochastic transcriptional activation, are essential for establishing a neuronal surface identity code involved in neural circuit assembly during brain development. Researchers from Columbia University, Yale University, the Whitehead Institute for Biomedical Research, New York Genome Center and UCSF have teamed up to address the mechanism behind the stochastic promoter choice of Pcdh (1). However, the low level of expression of Pcdh genes provided significant cost and bioinformatic challenges to detect them in standard RNA-seq libraries. To overcome this challenge, the researchers used a custom-designed myBaits® target capture kit from Arbor Biosciences as an efficient and cost-effective solution to enrich the Pcdh transcripts from RNA-Seq NGS libraries prior to sequencing. This solution allowed the researchers to achieve an over 45-fold enrichment of Pcdh transcripts using only 10% of the sequencing throughput of standard RNA-Seq, and thus dramatically increased sensitivity in detecting Pcdh expression.

THE IMPORTANCE OF TRANSCRIPTIONAL REGULATION OF Pcdh GENES

During brain development, the ability of neurites of individual neurons to selectively avoid each other plays a key role in neural circuit assembly. This process, known as self-avoidance, is mediated by a unique combination of cell surface homophilic recognition molecules that function as a molecular identity code (2). In mammals, this identity code is generated by the clustered protocadherin (Pcdh) genes, which consist of three tandemly arrayed α, β, and γ subclusters that together span nearly 1 Mb of the genome, via a poorly understood mechanism of stochastic and combinatorial promoter choice (3, 4). These stochastic enhancer/promoter interactions of Pcdhα genes were previously observed to require binding of the CTCF/cohesin protein complex to two binding sites (CBS) on alternate exons (5). DNA methylation of the CBSs also likely plays an important role in the mechanism of stochastic Pcdhα promoter choice (5). However, the temporal relationship between Pcdh gene promoter DNA methylation and promoter choice remains unclear.

TECHNICAL CHALLENGE

To decipher the mechanism of stochastic promoter choice, the study published in Cell by Canzio and colleagues (1) investigated the methylation state and the corresponding transcriptional state of Pcdhα promoter DNA under various experimental conditions. Due to their low level of expression, one of the major technical challenges of this study was to accurately detect expression and evaluate the transcriptional state changes of Pcdh genes. Dr. Daniele Canzio, Assistant Professor from Neurology UCSF Weill Institute for Neurosciences, University of California, San Francisco, explains:

"The expression level of Pcdh genes is generally very low that using whole transcriptome RNA-Seq would be cost-prohibitive and its sensitivity of detecting rare transcripts is low. Also, since the Pcdh locus spans 1 Mb of the genome, using legacy technologies such as RT-qPCR was impractical. Therefore, we needed an efficient, reliable yet cost-effective way to detect and quantify Pcdh transcripts."
– Dr. Canzio

THE SOLUTION

To affordably sequence lowly-expressed Pcdh transcripts in neuron cells, the research team employed an RNA probe-based enrichment strategy to capture the cDNA from RNA precursors (pre-mRNA) and messenger RNAs (mRNA). They selected Arbor Biosciences to produce a custom-designed myBaits® target capture kit comprising a total of 16,357 biotinylated RNA probes covering about 90% of the Pcdh α and γ clusters, and a positive control CBX5 locus.

RESULTS

Following myBaits® capture library protocol (Figure 1), Canzio and colleagues achieved a greater than 45-fold enrichment of Pcdh RNA transcript sequencing reads using 10% of the sequencing throughput of their direct shotgun sequencing (an over 470-fold enrichment after throughput normalization, Figure 3). Remarkably, this enrichment revealed a high level of antisense RNA transcription of the Pcdhα alternate exons, which contain dual CBSs in SK-N-SH cells (Figure 2).

"Arbor Biosciences provided exceptional customer service and professional scientific support that greatly helped our project move forward rapidly and smoothly." – Dr. Canzio

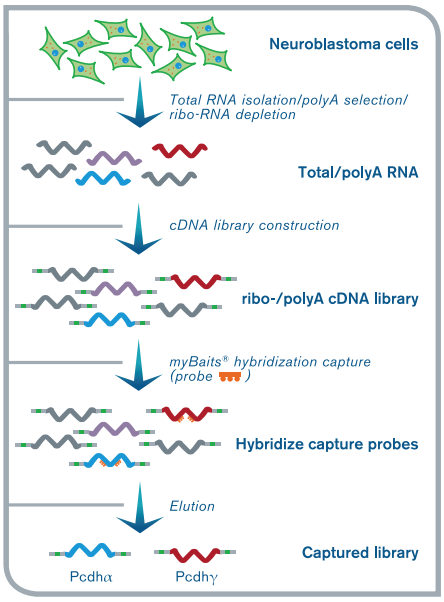

Figure 1. A schematic overview of the myBaits® RNA-Seq capture workflow *from total RNA isolation to sequencing. The blue, red and purple samples indicate RNA from the Pcdh α, β and γ gene clusters, respectively. The gray indicate RNA from the rest of the genome.*

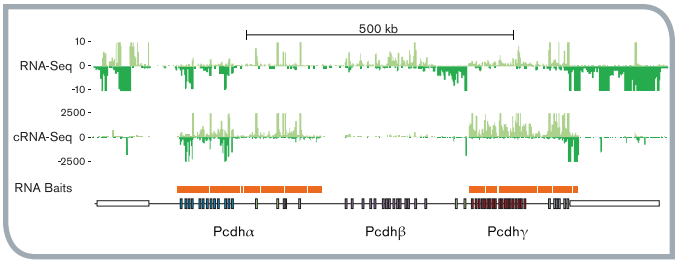

Figure 2. A high level of antisense RNA transcription *of the Pcdhα alternate exons was revealed by myBaits® RNA-Seq capture with high confidence from increased coverage. Orange bar: myBaits for Pcdh α and γ clusters.*

Dr. Canzio says that "myBaits® was our preferred choice due to its flexibility and reliability in custom design, excellent capture performance, and intelligent cost-savings."

	%Pcdhα Mapped Reads	%Pcdhγ Mapped Reads	%CBX5 Mapped Reads
Standard RNA-Seq	0.01	0.02	0.005
myBaits® RNA-Seq Capture	4.7	10.9	2.8
Fold Enrichment	470	545	560

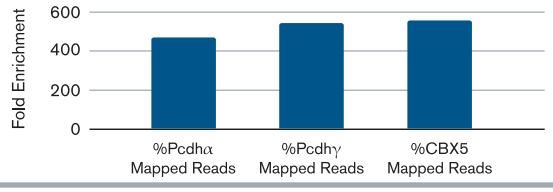

Figure 3. Comparison of mapped read counts between standard RNA-Seq and myBaits® RNA-Seq capture *showing a three order of magnitude enrichment of Pcdh transcripts using myBaits® RNA-Seq capture.*

CONCLUSION

Canzio and colleagues have revealed that coupling transcription of a long noncoding RNA to DNA demethylation ensures stochastic promoter choice for clustered Pcdhα genes. This work illuminated for the first time a general mechanism for stochastic promoter activation, thereby laying the foundation for further studies of other clustered gene families where stochastic gene expression occurs.

"The advancement in RNA-Seq capture has allowed us to design inexpensive, custom assays that can target RNA-Seq to only genomic regions of interest, thereby enriching lowly expressed Pcdhα transcripts. myBaits® RNA-Seq capture has been a valuable tool in reaching those remarkable discoveries and we are confident that it will continue to support us in better understanding of transcriptional regulation of Pcdh genes." – Dr. Canzio

REFERENCES

1. Canzio, D. *et al.* (2019) **Antisense lncRNA Transcription Mediates DNA Demethylation to Drive Stochastic Protocadherin α Promoter Choice.** *Cell.*
2. Zipursky S.L. *et al.* (2013) The molecular basis of self-avoidance. *Annu. Rev. Neurosci.*
3. Lefebvre J.L. *et al.* (2015) Development of dendritic form and function. *Annu. Rev. Cell Dev. Biol.*
4. Mountoufaris, G. *et al.* (2018) Writing, reading, and translating the clustered protocadherin cell surface recognition code for neural circuit assembly. *Annu. Rev. Cell Dev. Biol.*
5. Guo, Y. *et al.* (2012) CTCF/cohesin-mediated DNA looping is required for protocadherin a promoter choice. *Proc. Natl. Acad. Sci. USA.*

web www.arborbiosci.com
email info@arborbiosci.com
phone 1-734-998-0751
twitter @ArborBio

RNA-Seq Capture
www.arborbiosci.com/myBaits