

# Clariom D microarrays provide a deep view of our transcriptome

## The search for complete transcriptome analysis

The Human Genome Project was an enormous success. Not only did it promote our understanding of the content of genomes, but it also brought forth new technologies for collecting sequences and new algorithms for analyzing those sequences. But even as our tools for analyzing the human genome increased in power, the complexity of the genome continued to provide new surprises. For example, when the sequence of the human genome was first announced, the number of expressed protein-coding sequences was hypothesized to be around 22,000. As the analyses became more sophisticated, however, it was clear that the number of expressed sequences was much higher.

Further analyses revealed that these expressed sequences included microRNAs (miRNAs), long noncoding RNAs (lncRNAs), long intergenic noncoding RNAs (lincRNAs), splice isoforms, and circular RNA. An international Sequencing Quality Control (SEQC) consortium showed that in complex transcriptomes, more than 45,000 expressed sequences are detected when deep sequencing of up to a billion reads per sample is performed

### In this technical note, we describe how:

- A deep understanding of the transcriptome comes from incorporating multiple gene models
- Comprehensive analysis of the complete transcriptome allows for discovery of new biomarkers associated with disease or physiological states
- Direct measurement of transcript abundance removes read depth considerations in a gene expression experiment
- Applied Biosystems™ Clariom™ D microarrays facilitate complete transcriptome analysis, including coverage of novel noncoding transcripts and splice isoforms

[1]. More recently, the Genotype-Tissue Expression (GTEx) study analyzed the transcriptomes of 31 normal human tissues and found 43,126 genes and many novel isoforms of the genes [2]. Programs like the Human Cell Atlas aim to expand the human catalog of cells beyond the current ~300 defined types [3], adding to the list of what can be expressed in a transcriptome.

## Modeling gene structures

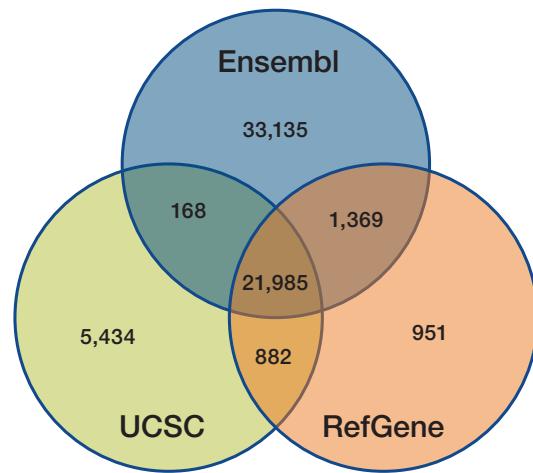
To better understand and manage all this information, several attempts have been made to catalog and model these expressed sequences. For example, the National Center for Biotechnology Information (NCBI) and the European Molecular Biology Laboratory (EMBL) are among the 16 public organizations that curate extensive databases of genes like RefSeq (NCBI) and Ensembl (EMBL).

Surprisingly, these databases do not overlap as extensively as one might expect, since each is a product of its own distinct curation algorithms (Figure 1). For example, the *SLC18A2* locus can be modeled as having different intron/exon structures, depending on the modeling database used (Figure 2). Thus, a view of a transcriptome can be very limited if we consider only one of these 16 databases. We can obtain a comprehensive map of the transcriptome only by aggregating data from multiple sources. While the validity of every sequence contained in all these databases may be open to interpretation, the totality of possible expressed sequences provides a more complete and reliable view of expressed loci in our chromosomes.

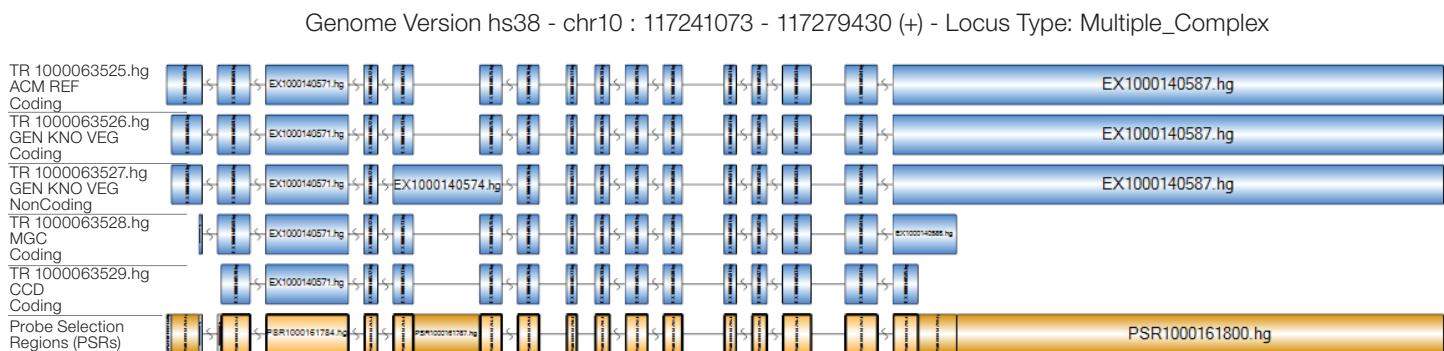
## Alternative splicing and a need for higher-resolution transcriptomics

Modern high-resolution transcriptomics aims to resolve transcripts that are up- or downregulated in various tissues or cellular states. This regulation can be measured and is thought of as comprising both gene-level and transcript-level events. When we use the term “gene level”, we typically are attempting to measure the combined transcript-level events along a given gene locus. In higher organisms, almost all coding genes have multiple exons, and depending on the exon usage, the genes have the potential to generate many alternative transcripts (Figure 2).

Simply trying to measure up- or downregulation at the gene level often obscures the changes that happen at the splicing or exon level. Multiple forces can simultaneously regulate gene expression and splicing, such that a ratio of isoforms can be altered with or without changes in the overall combined gene-level expression estimates. How technologies take this complexity into consideration greatly impacts the inherent assumptions behind measurements used for assessing transcriptome change. Modern high-resolution transcriptomics measures exons or sub-exonic events to help ascertain splicing and overall gene-level changes.



**Figure 1. Overlap and uniqueness of RefGene (a subset of the RefSeq database), UCSC, and Ensembl annotation models of genes.** In general, the different models have a high degree of overlap—there are 21,958 genes common to all three databases. However, note that more than half of the genes in the Ensembl model are unique.



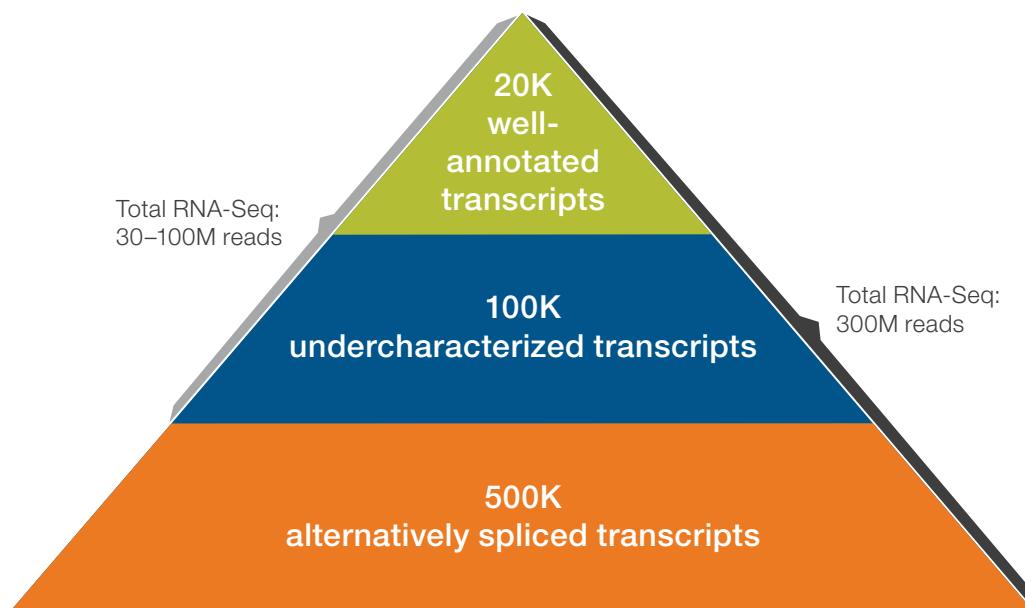
**Figure 2. Gene model of VMAT2, also known as SLC18A2, containing multiple isoforms seen in 6 different public databases.** Probe selection regions on the Clariom D array, shown in yellow, can be utilized to detect alternative splicing events and overall gene-level expression.

What has been increasingly clear as technologies mature is that alternative splicing is an important factor in many diseases. For example, cancer cells have been shown to increase proliferation, migration, and ultimately the rate of survival, through alternative splicing [4].

Most technologies used today are still focused on well-annotated genes. However, less well-known sequences, such as noncoding RNAs (ncRNAs), have been shown to regulate the expression of other genes and have been shown to be a rich source of biomarker discovery [5]. If we use technologies that survey uncharacterized gene loci or consider differential exon usage or splicing, the transcriptome comprises much more than the common set of ~20,000 well-annotated genes (Figure 3). A comprehensive biomarker survey should include exon resolution, noncoding RNAs, and undercharacterized genes.

## Methods for analyzing the transcriptome

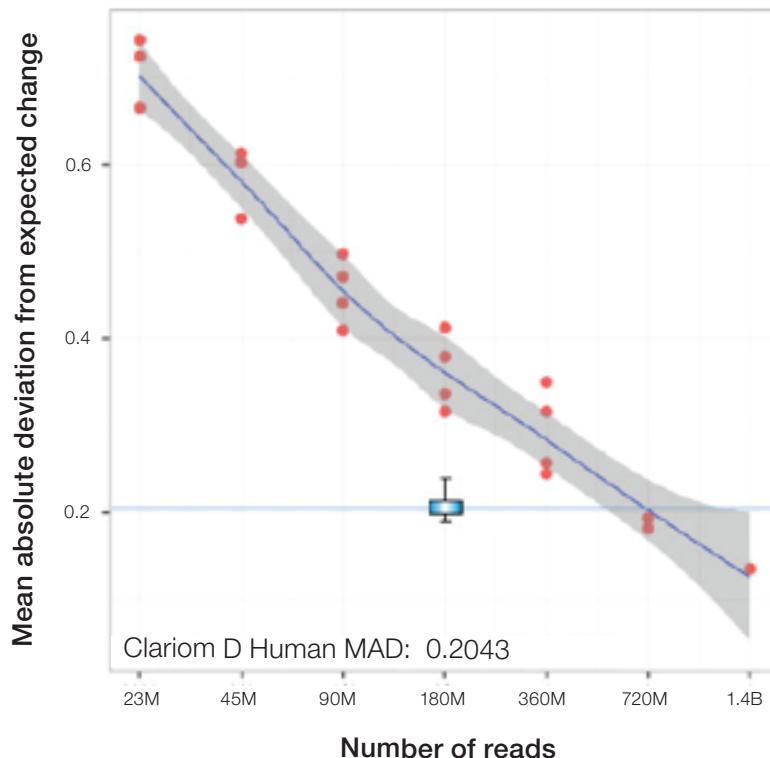
Two techniques are commonly used to analyze changes in gene expression across a transcriptome. One method, using next-generation RNA sequencing (RNA-Seq), uses massively parallel sequencing to sample RNA sequences present in a library. NGS reads are aligned to the genome in RNA-Seq. The read depth is required to find low-abundance transcripts, which can lead to higher cost per sample, depending on the platform [6].



**Figure 3. Signature search space can be thought of as the number of transcripts evaluated with different technologies.** Note that when sampling technologies such as Illumina™ NGS or Applied Biosystems™ SOLiD™ NGS RNA-Seq are used, very large numbers of events are needed to get deep information about the transcripts present in an experiment. In contrast, the Clariom D array is able to detect these events without sampling, and therefore similar depth can be achieved in one experiment. For details, see reference 1.

Moreover, as with any sampling technology, the more sampling events that are collected, the better the precision of the measurement. For example, a simple experiment can be designed where the values of fold changes in defined samples can be predicted [1]. In an RNA-Seq experiment, the mean absolute deviation from the predicted value drops with increasing sequencing depth (Figure 4). One corollary of this is that more sampling is required in order to measure rare events with statistical confidence. Results from the SEQC consortium showed that deeper sequencing (increased sampling) of the transcriptome is needed to reveal low-abundance transcripts and splice junctions. Thus, read depth should be a key consideration when experimental goals include rarely expressed or low-abundance transcripts, coverage of introns, and nonpolyadenylated (i.e., noncoding) targets [1]. In another example, complex samples such as tumor biopsies may contain multiple cell types. The transcriptome contained in this sample is more complex than that of a homogeneous cell line, and therefore to precisely measure the gene expression changes in the variety of cells present, including the rare ones, more reads are required.

Another strategy for analyzing gene expression changes involves using DNA oligonucleotide microarrays that have been used to analyze gene expression changes across a very large number of targets for close to 30 years. The methods used to generate measurements and analyze the resulting large datasets have been well established and are the gold-standard approaches. Originally, such microarrays only analyzed a few thousand targets. However, technology advances have increased the number of targets that can be analyzed; currently, over 6 million discrete sequences can be analyzed on a single chip. Queries for a very large and complex set of sequences are possible with a single well-established workflow. Moreover, since expression data are captured directly from the signal on the chip, sampling is not required to increase precision or detect rare sequences.



**Figure 4. Sequencing depth required for equivalent variation between the Clariom D array and RNA-Seq analysis.** By evaluating a tissue mixture model in which RNA from two samples is mixed in known proportions, the mean absolute deviation (MAD) of expression was evaluated across all measured exons. (Modified data from reference 13.)

## Clariom D applications: the answer to many research challenges

The Clariom D microarrays represent the latest advances in the family of transcriptome analysis solutions (Table 1). The sequences on the Clariom D arrays are based on coding and noncoding sequences culled from 16 different databases [7], representing over 540,000 transcripts in over 6.7 million probes. Included in these sequences are probes for different splice isoforms, noncoding RNAs (pre-miRNAs, lincRNAs, Piwi-interacting RNAs (piRNAs)), and circular RNAs, as well as annotated and speculative sequences. Querying the multiple transcript models on the Clariom D chip in a single experiment helps ensure that important biomarkers are not missed.

Applied Biosystems™ Transcriptome Analysis Console (TAC) Software was designed to quickly analyze the Clariom D array data. A report of all transcripts and the associated relative expression, using statistical and visualization tools, is provided from the data. Additionally, information regarding the role of these transcripts in biological pathways is provided. Together, the Clariom D solution of microarrays and software provides rapid, easy, and economical tools for obtaining meaningful gene expression data.

Microarrays continue to be an important tool for analyzing gene expression changes. For example, during the fiscal year of 2017–2018, the National Institute of Health (NIH) funded over 200 R01 grants that make use of microarrays to analyze functional changes from gene expression differences. In addition, the number of publications utilizing Clariom D technologies continues to grow. Some recent examples are highlighted below, and a comprehensive list is given in the reference section.

### Example 1: Gene expression variation in adipose tissue

Metabolic differences may influence weight gain, but the tissues and mechanisms involved are not known. As part of a study of two different cohorts of women, Arner et al. [8] examined gene expression differences in subcutaneous adipose tissue in baseline and follow-up samples, using the Clariom D microarrays. They found that a subset of previously described lipolysis gene transcripts was lower in the weight-gain samples versus the weight-stable samples. They hypothesized that these results implicate inefficient lipolysis in subcutaneous fat cells as a contributor to long-term weight gain.

### Example 2: Systemic dysregulation in subjects with Parkinson's disease

Miki et al. [9] explored the hypothesis that Parkinson's disease involves systemic dysregulation of the autophagic pathway. They collected peripheral blood mononuclear cells (PBMCs) from individuals with Parkinson's disease and unaffected individuals, and used the Clariom D arrays to generate transcriptomic profiles from each group. Using TAC Software to analyze the results, they confirmed that genes involved in autophagy were indeed dysregulated in individuals with Parkinson's disease. They also confirmed that genes in the lysosomal pathway, another catabolic mechanism, were also dysregulated. Together, these results suggest that systemic alteration of catabolic pathways may be a fundamental aspect of the disease.

**Table 1. Clariom D microarrays vs. RNA-Seq solutions.**

	Clariom D	RNA-Seq
Gene models	16 databases	One database at a time
Method of measuring transcript levels	Direct	Sampling
Data analysis	TAC Software	Varies

### **Example 3: Ex vivo breast cancer tumors**

Eckhardt et al. [10] were interested in modeling breast cancer tumors grown *ex vivo* in 3 dimensions as a more accurate representation of breast tumors than cultured cells. To show the validity of their model, they needed to show that the *ex vivo*-grown tumor mass had biological characteristics similar to those of the starting tumor. They used the Clariom D arrays to verify that the correlation between the gene expression patterns in the *ex vivo*-grown tumors and the original tumor was highly significant, even when noncoding RNAs were included in the analysis. Thus, they were able to rapidly verify that their *ex vivo* tumor model could be used as a surrogate for the original tumor.

### **Example 4: lncRNA analysis to examine olfactory function decline**

A common feature of aging is the decline in sensory function. To begin to understand olfactory decline, Wang et al. [11] examined expression profiles in young and aged mouse olfactory bulbs using mouse Clariom D microarrays. They obtained distinct sets of coding and noncoding RNAs that were differentially expressed in the two tissue samples from different sources. By performing pathway analysis of the differentially expressed genes, they hypothesized that a decline in olfactory function may be inversely correlated with the expression of some lncRNAs in the neuroactive ligand–receptor interaction pathway.

### **Summary**

Clariom D microarrays are a next-generation, cost-effective solution for comprehensive gene expression analysis. Leveraging RNA discoveries over several years and the information available in several public databases, the Clariom D microarrays enable investigators to find coding, noncoding, or as yet unannotated events with a simple and easy workflow and at exon resolution. Transcriptome complexity or sequencing depth considerations are not a concern with microarrays. Microarrays are built on trusted and well-understood chemistries and algorithms for analyzing gene expression differences. The Clariom D microarrays incorporate information from a large number of different transcript models, giving confidence that important transcript isoforms or noncoding sequences are analyzed. The TAC Software facilitates interpretation of the gene expression data generated by the Clariom D microarrays. Some researchers recently have stated that “modern microarrays can still outperform sequencing for standard analysis of gene expression in terms of reproducibility and costs” [12]. Together, these features make the Clariom D microarrays an ideal and economical choice for gene expression research.

### **Ordering information**

<b>Product</b>	<b>Cat. No.</b>
Clariom D Assay, human (includes the whole-transcriptome Clariom D microarray and the GeneChip WT Pico reagent kit)	902922
Transcriptome Analysis Console (TAC) Software 4.0	NA

## Cited references

1. SEQC/MAQC-III Consortium (2014) A comprehensive assessment of RNA-seq accuracy, reproducibility and information content by the Sequencing Quality Control consortium. *Nat Biotechnol* 32:903–914.
2. Shumate A et al. (2018) Thousands of large-scale RNA sequencing experiments yield a comprehensive new human gene list and reveal extensive transcriptional noise. doi.org/10.1101/332825.
3. Regev A et al. (2017) The Human Cell Atlas. *Elife* pii: e27041.
4. Urbanski LM et al. (2018) Alternative-splicing defects in cancer: Splicing regulators and their downstream targets, guiding the way to novel cancer therapeutics. *Wiley Interdiscip Rev RNA* 9:e1476.
5. Mattick JS (2018) The state of long non-coding RNA biology. *Noncoding RNA* 4: pii: E17.
6. Costa-Silva J et al. (2017) RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS One* 12:e0190152.
7. Public database sources for ClariomD Human: Refseq, ccdsGENE, mgcGenes, UCSC KnownGene, lincRNATranscripts, VEGA, GENCODE, NONCODE, MiTranscriptome, RNACentral, ENSEMBL, circBase, Guo CircRNAs, AceView, IncRNAWiki, InCRNADB.
8. Arner P et al. (2018) Weight gain and impaired glucose metabolism in women are predicted by inefficient subcutaneous fat lipolysis. *Cell Metab* 28:45–54.
9. Miki Y et al. (2018) Alteration of autophagy-related proteins in peripheral blood mononuclear cells of patients with Parkinson's disease. *Neurobiol Aging* 63:33–43.
10. Eckhardt B et al. (2018) Clinically relevant inflammatory breast cancer patient-derived xenograft-derived ex vivo model for evaluation of tumor-specific therapies. *PLoS One* 13:e0195932.
11. Wang M et al. (2017) Expression profiling of mRNAs and long non-coding RNAs in aged mouse olfactory bulb. *Sci Rep* 7:2079.
12. Nazarov PV et al. (2017) RNA sequencing and transcriptome arrays analyses show opposing results for alternative splicing in patient-derived samples. *BMC Genomics* 18:443.
13. Xu W et al. (2013) Cross-platform performance evaluation of RNA-Seq and microarray. researchgate.net/publication/266157821\_Cross-platform\_performance\_evaluation\_of\_RNA-Seq\_and\_microarray.
14. Fadden P et al. (2018) Immune checkpoint inhibitor induced tumor gene expression changes in murine syngeneic colon cancer models, AACR.
15. Frith TJ et al. (2018) Human axial progenitors generate trunk neural crest cells. *Elife* 7: pii: e35786
16. Gambardella V et al. (2018) PO-500 Nrf2 represents a convergent point of acquired resistance in her2 positive gastric cancer models. BMJ Publishing Group Limited.
17. Gebraad A et al. (2018) Monocyte-derived extracellular vesicles stimulate cytokine secretion and gene expression of matrix metalloproteinases by mesenchymal stem/stromal cells. *FEBS J* 285:2337–2359.
18. Gonzalo R et al. (2018) Introduction to microarrays technology and data analysis. *Comprehensive Analytical Chemistry* 82:37–69.
19. Halloran PF et al. (2018). The transcripts associated with organ allograft rejection. *Am J Transplant* 18:785–795.
20. He H et al. (2018) c-Jun/AP-1 overexpression reprograms ERα signaling related to tamoxifen response in ERα-positive breast cancer. *Oncogene* 37:2586.
21. Hitzemann R et al. (2018) Genes, Behavior, and Next-Generation Sequencing: The First 10 Years. Molecular-Genetic and Statistical Techniques for Behavioral and Neural Research, Elsevier: 289–308.
22. Jiménez-Sánchez A et al. (2018) Unraveling tumor-immune heterogeneity in advanced ovarian cancer uncovers immunogenic effect of chemotherapy. doi.org/10.1101/441428.
23. Kasendra M et al. (2018) Development of a primary human Small Intestine-on-a-Chip using biopsy-derived organoids. *Sci Rep* 8:2871.
24. Kodigepalli KM et al. (2018) SAMHD1 modulates in vitro proliferation of acute myeloid leukemia-derived THP-1 cells through the PI3K-Akt-p27 axis. *Cell Cycle* 17:1124–1137.
25. Koentges C et al. (2018) Gene expression analysis to identify mechanisms underlying heart failure susceptibility in mice and humans. *Basic Res Cardiol* 113:8.
26. Kore S et al. (2018) Transcriptomic profiling of tauopathy reveals gene populations responsive to tau expression and a subpopulation of therapeutically relevant genes. *Alzheimers Dement* 14(7): P1141.
27. Krementsov DN et al. (2018) Sex-specific gene-by-Vitamin D interactions regulate susceptibility to central nervous system autoimmunity. *Front Immunol* 9:1622.

## Clariom D references

28. Gökmen-Polar Y et al. (2019) Splicing factor *ESRP1* controls ER-positive breast cancer by altering metabolic pathways. *EMBO Rep* doi:10.15252/embr.201846078.
29. Kuasne H et al. (2018). Penile cancer in vitro models useful for the identification of targeted therapies, AACR.
30. Lavigne K et al. (2018) Effect of chemotherapy on immune infiltration status and immune pathway activation in high-grade serous ovarian cancer. *Gynecol Oncol* 149:53.
31. Liu K et al. (2018). Functional role of a long non-coding RNA LIFR-AS1/miR-29a/TNFAIP3 axis in colorectal cancer resistance to photodynamic therapy. *Biochim Biophys Acta Mol Basis Dis* 1864 (9 Pt B):2867-2880.
32. Machiela MJ et al. (2018) Genome-wide association study identifies multiple new loci associated with Ewing sarcoma susceptibility. *Nat Commun* 9:3184.
33. Martinez E et al. (2018) PO-326 Impact of mir-205–5 p and mir-425–5 p on Wnt and AR signalling pathways in castration resistant prostate cancer transition, BMJ Publishing Group Limited.
34. Minnier Jet al. (2018) RNA-Seq and Expression Arrays: Selection Guidelines for Genome-Wide Expression Profiling. *Gene Expression Analysis*, Springer: 7-33.
35. Morley JE et al. (2018) Metformin increases PKC and decreases APP and Tau in the SAMP8 mouse model of Alzheimer's disease. *Alzheimers Dement* 14(7)(Suppl):P1141.
36. Nelson AM et al. (2018). RNA splicing in the transition from B cells to antibody-secreting cells: the influences of ELL2, small nuclear RNA, and endoplasmic reticulum stress. *J Immunol* 201:3073-3083.
37. Nishida S et al. (2018) Collagen VI suppresses fibronectin-induced enteric neural crest cell migration by downregulation of focal adhesion proteins. *Biochem Biophys Res Commun* 495:1461-1467.
38. Oltra SS et al. (2018) Methylation deregulation of miRNA promoters identifies miR124-2 as a survival biomarker in Breast Cancer in very young women. *Sci Rep* 8:14373.
39. Ortiz IMDP et al. (2018) Loss of DNA methylation is related to increased expression of miR-21 and miR-146b in papillary thyroid carcinoma. *Clin Epigenetics* 10:144.
40. Prabhu AH et al. (2018) Integrative cross-platform analyses identify enhanced heterotrophy as a metabolic hallmark in glioblastoma. *Neuro Oncol* 10.1093/neuonc/noy185.
41. Romero JP et al. (2018) Comparison of RNA-seq and microarray platforms for splice event detection using a cross-platform algorithm. *BMC Genomics* 19:703.
42. Schäffler H et al. (2018) NOD2-and disease-specific gene expression profiles of peripheral blood mononuclear cells from Crohn's disease patients. *World J Gastroenterol* 24:1196.
43. Singhal G et al. (2018) Deficiency of fibroblast growth factor 21 (FGF21) promotes hepatocellular carcinoma (HCC) in mice on a long term obesogenic diet. *Mol Metab* 13:56-66.
44. Smith BR et al. (2018) A Tau cleavage product,  $\delta$ tau314, is increased in Alzheimer's disease and Parkinson's disease with dementia. *Alzheimers Dement* 14(7)(Suppl):P1141-P1142.
45. Spade D J et al. (2018) All-trans retinoic acid disrupts development in ex vivo cultured fetal rat testes. I: Altered seminiferous cord maturation and testicular cell fate. *Toxicol Sci* doi:10.1093/toxsci/kfy260.
46. Usenovic M et al. (2018) Novel targets for blocking the uptake of Tau oligomers in hiPSC neurons. *Alzheimers Dement* 14(7)(Suppl):P1140-P1141.
47. Vitali C et al. (2018) OP0202 gene expression profiles in primary Sjögren's syndrome with and without systemic manifestations, BMJ Publishing Group Ltd.
48. Wang K et al. (2018) Neat1-miRNA204-5p-PI3K-AKT axis as a potential mechanism for photodynamic therapy treated colitis in mice. *Photodiagnosis Photodyn Ther* 24:349-357.
49. Willebroards J et al. (2018) Protective effect of genetic deletion of pannexin1 in experimental mouse models of acute and chronic liver disease. *Biochim Biophys Acta Mol Basis Dis* 1864(3):819-830.
50. Yanguas SC et al. (2018) Genetic ablation of pannexin1 counteracts liver fibrosis in a chemical, but not in a surgical mouse model. *Arch Toxicol* 92:2607-2627.
51. Yao H et al. (2018) Leukaemia hijacks a neural mechanism to invade the central nervous system. *Nature* 560:55.
52. Yeung JC et al. (2018) Towards donor lung recovery—gene expression changes during ex vivo lung perfusion of human lungs. *Am J Transplant* 18:1518-1526.
53. Yu J et al. (2018) Host gene expression in nose and blood for the diagnosis of viral respiratory infection. *J Infect Dis* 10.1093/infdis/jiy608.
54. Zhang B et al. (2018) Pan-cancer analysis of RNA binding proteins (RBPs) reveals the involvement of the core pre-mRNA splicing and translation machinery in tumorigenesis. doi.org/10.1101/341263.
55. Zhang B et al. (2018) Genes related to the different stages of diabetic kidney disease: a Clariom™ D assay in patients with biopsy proven diabetic nephropathy. *Nephrology Dialysis Transplantation* 33(suppl\_1): i360-i360.
56. Zhang Q et al. (2018) CD90 determined two subpopulations of glioma-associated mesenchymal stem cells with different roles in tumour progression. *Cell Death Dis* 9:1101.

Find out more at [thermofisher.com/clariomassays](http://thermofisher.com/clariomassays)

**ThermoFisher**  
SCIENTIFIC