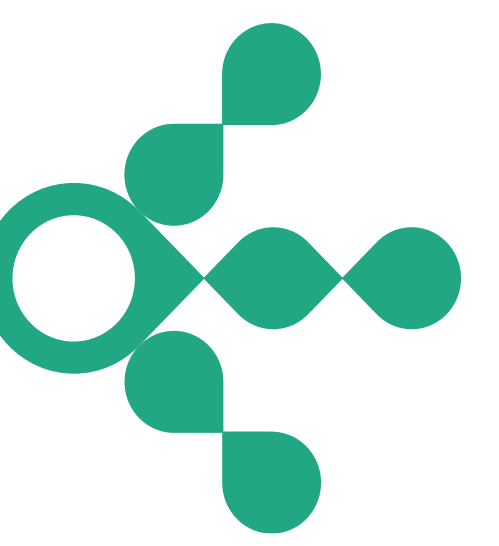




Linda D. Orzolek¹, Jonas Hansen², Sophie Quisling², Ryan Keivanfar², Sara Agee Le²
1 OMAPIx, Inc., Frederick, MD, USA, 2 Exponential Molecular, Mountain View, CA, USA

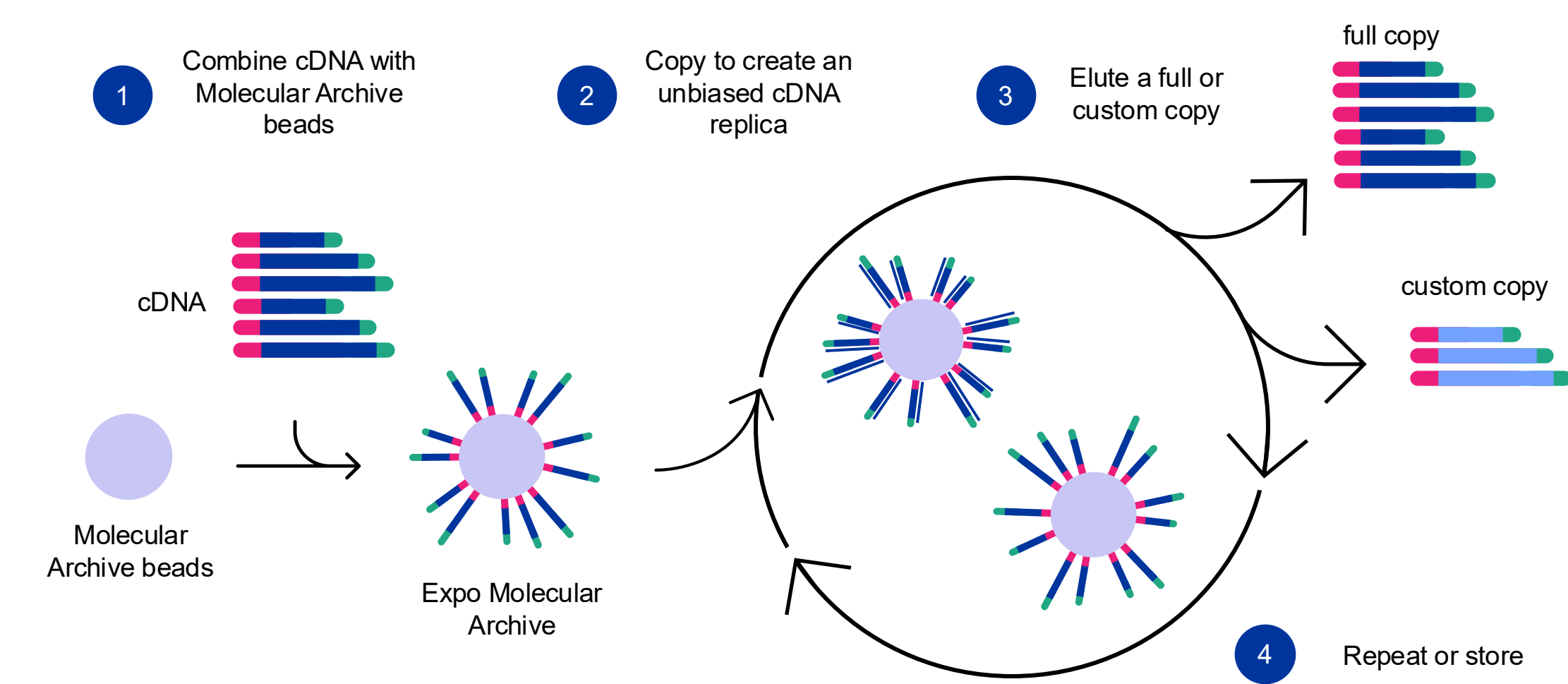
Abstract

Very recently, long-read sequencing (LRS) is gaining use to elucidate isoforms and other long-read information from fresh frozen tissue; however, short-read sequencing is still the norm. Although both types of sequencing can provide unique and potentially beneficial information, using combinatorial sequencing approaches is not often performed due to limited input material from a small number of samples.

To investigate the potential for combining both long- and short- read information from single cell and spatial transcriptomics analysis, we tested a series of samples using a novel click-chemistry-based method to construct an EXPO molecular archive from cDNA and tissue samples. This novel method preserves long-read molecules in a template that can be used for multiple generations of non-destructive downstream sequencing and molecular assays, including short-read sequencing (SRS), long-read sequencing (LRS), hybrid capture, targeted panel, or single gene amplification with PCR.

Using the Illumina iSeq 100 and the Element AVITI System for SRS and Oxford Nanopore for LRS, we show that an EXPO Molecular Archive can be prepared from pre-existing cDNA from commercially available single cell and spatial transcriptomics methods. Samples with intact full-length cDNA can be archived to preserve long molecules for LRS, while also allowing SRS to occur from the same template.

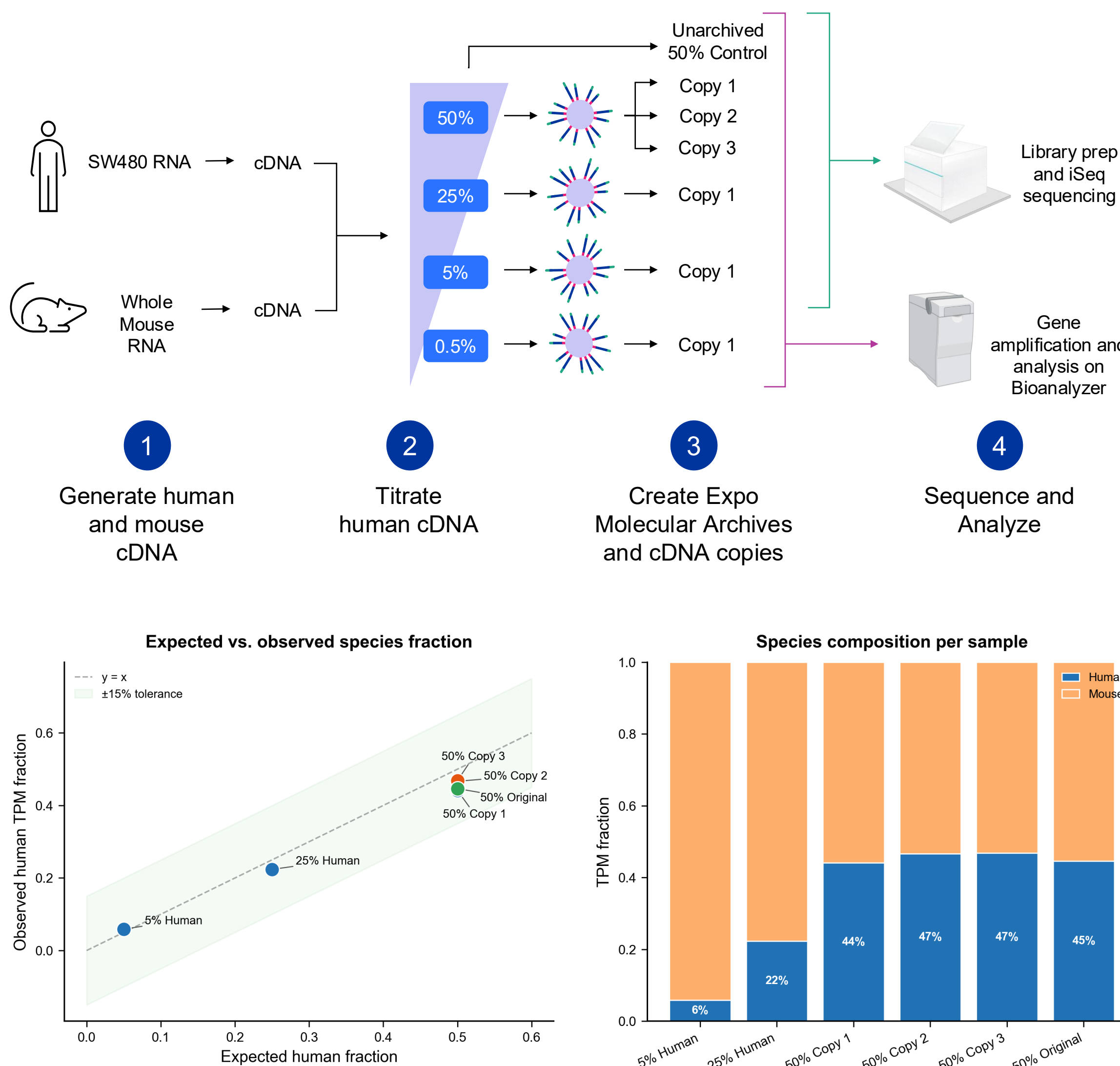
Exponential Molecular visual workflow:



An Expo Molecular Archive is constructed by covalently conjugating double stranded DNA to the surface of specialized magnetic beads. The version of the Expo Molecular Archive presented here is optimized for cDNA. After conjugating to the beads, the cDNA can be iterative copied and eluted using PCR primers that either amplify the entire cDNA pool or are designed for custom partial copying. After eluting the copy, the Expo Molecular Archive is enzymatically cleaned before storing or making further copies.

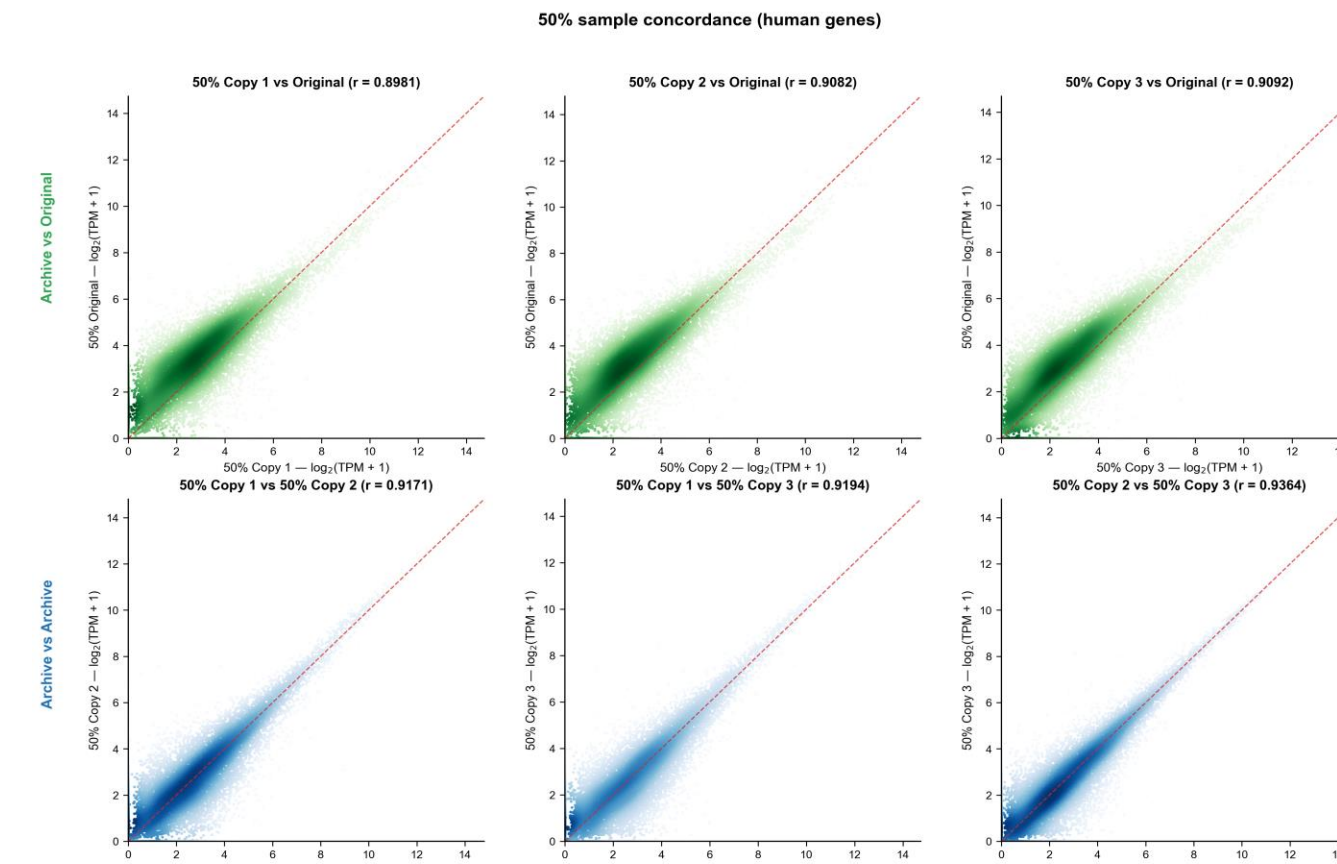
Assay Validation: Methodology

The Exponential Molecular Archive was validated through a titration experiment of cDNA from human and mouse sources at known proportions (0.5/99.5, 5/95, 25/75, 50/50) with 3 iterative copies at the 50/50 level. Following sequencing, pseudoalignment and quantification were performed with Kallisto against a combined human and mouse reference transcriptome. The analysis addresses four primary objectives: (1) consistency of top-expressed gene identification, (2) reproducibility across replicate copies of archived samples, (3) detection sensitivity across the titration gradient, and (4) stability of gene expression rankings.



Expected vs. observed species fraction (left) and species composition per sample (right). The observed fractions are highly correlated with expected values with small deviations.

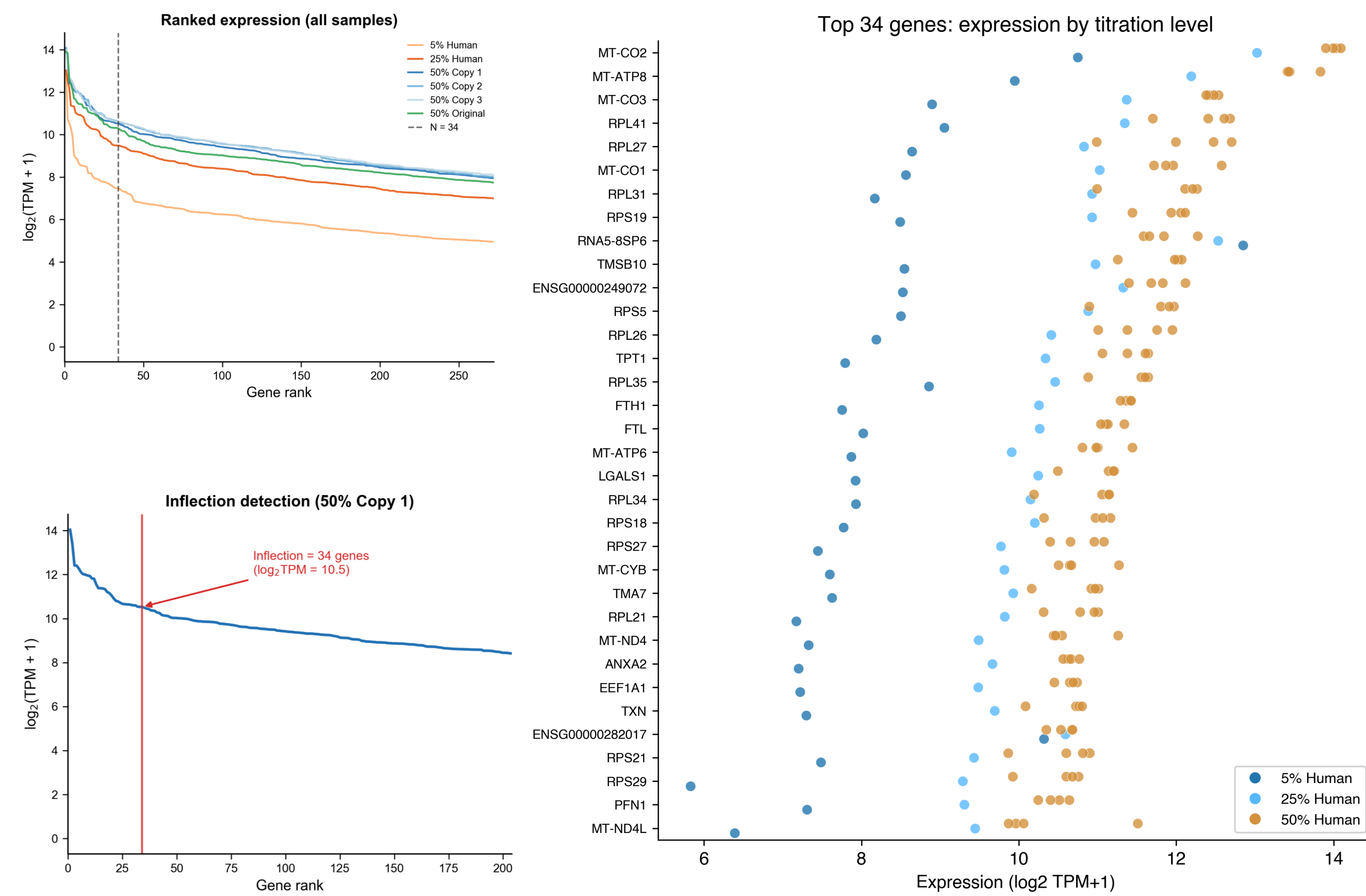
Assay Validation: Template Copy Reproducibility



Combined 6-panel concordance scatter: top row compares each 50% human and mouse cDNA Molecular Archive copy to the original (non-archived) 50% human sample; while the bottom row compares 50% human Molecular Archive copies to each other. Points are colored by local density (Gaussian KDE). All comparisons $\log_2(\text{TPM}+1)$ for the top 1,736 genes.

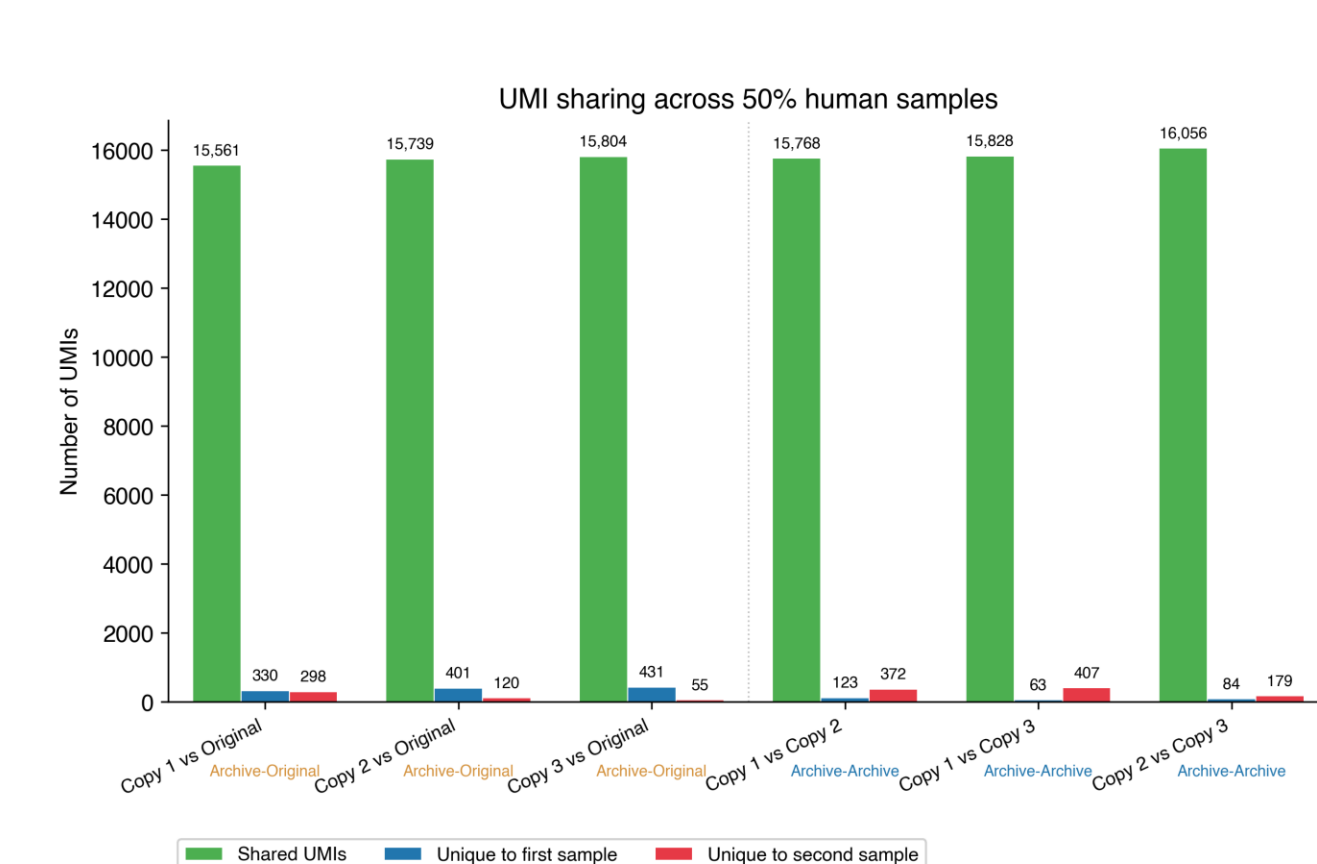
The replicate scatter plots show strong concordance (Pearson R = 0.92 - 0.94). The comparison of the Molecular Archive copies vs. original shows similarly high agreement (R ~ 0.90), indicating that use of the Molecular Archive does not introduce measurable distortion. This plot directly contrasts the two comparison types, confirming comparable concordance in both cases.

Assay Validation: Titration Sensitivity



Ranked $\log_2(\text{TPM})$ curve with inflection point at N = 34. The top left panel shows all samples; the bottom left panel shows just the reference sample. N = 34 marks where expression transitions from a high plateau to a steep decline. These first 34 most common genes were used to generate the plot on the right, showing the core set (N=34) of per-gene expression at each titration level. Each point represents an individual sample, with expression showing high similarity between the 50% human/mouse original sample and three copies. Expression decreases monotonically with the human input fraction, as expected for human-origin transcripts.

Assay Validation: Ultra-sensitivity



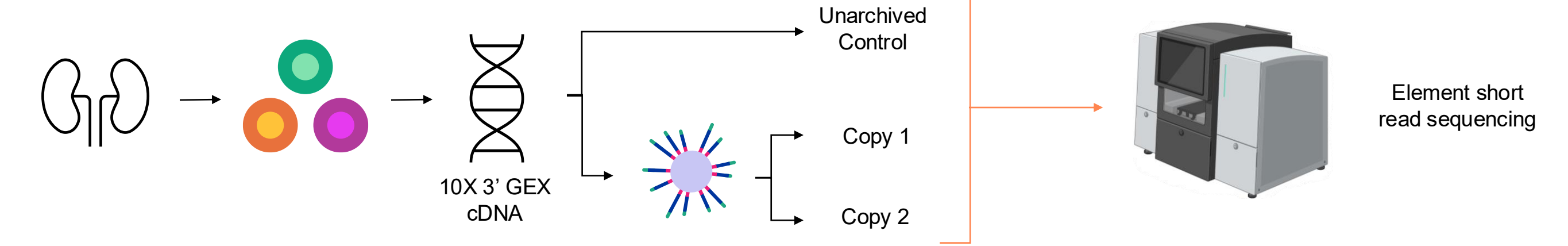
UMI sharing across 50% human samples. Each group of three bars shows the number of shared UMIs (green), UMIs unique to the first sample (blue), and UMIs unique to the second sample (red). For each sample, the number of shared unique UMI sequences and UMIs unique to each sample. All sample pairs share the vast majority of their 7-bp UMI sequences (~15,500 - 16,000) out of 16,384 possible, with only a few hundred UMIs unique to each sample.

PCR based gene detection was performed on the titration and iterative copies. For the most dilute Expo Molecular Archive with just 0.5% human input, we show clear detection of 6 genes chosen from the list of expressed genes identified using Bioanalyzer traces.

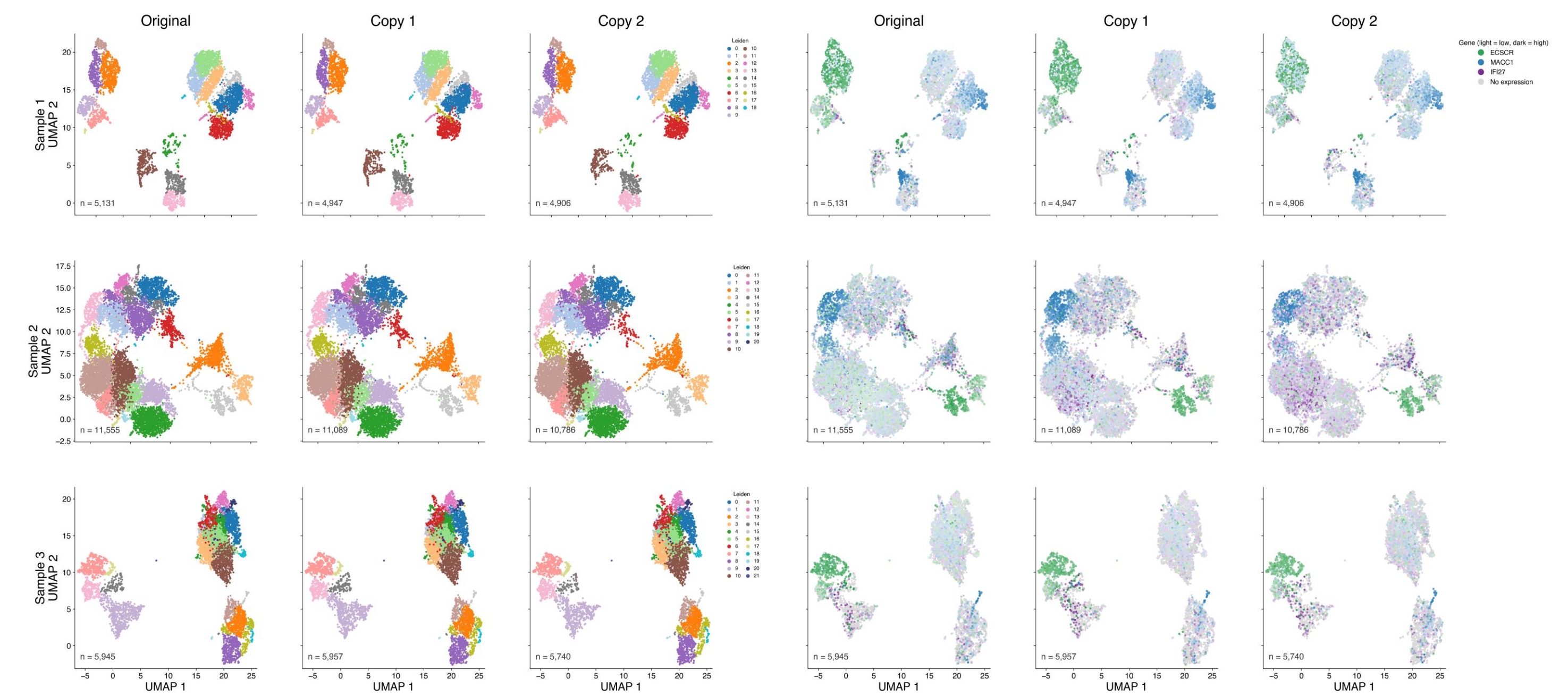
Conclusions

In conclusion, we show the analytical validation of the Expo Molecular Archive technology including reproducibility and sensitivity is sufficient for taking combined analytical approaches from a single input sample. Sequencing analysis and PCR show high sensitivity and concordance across copies, and the sequencing data also show concordance amongst top-gene expression from multiple Expo copies. In more complex kidney tissue samples, the Expo Molecular Archive technology shows high concordance across multiple copies compared to the parent full length cDNA library, retaining cell type clustering and differential gene expression from the original single cell libraries. This retention rate opens the door for more targeted applications. Proof of concept for integrating long read sequencing was shown from a kidney cell line from which an Expo Molecular Archive copy was generated and showed presence of isoform expression in the OGG1 gene. This shows potential for integrating sequencing approaches from the same originating sample by using each Expo copy as a new sequencing input, enabling a multitude of analytical modalities from a single sample.

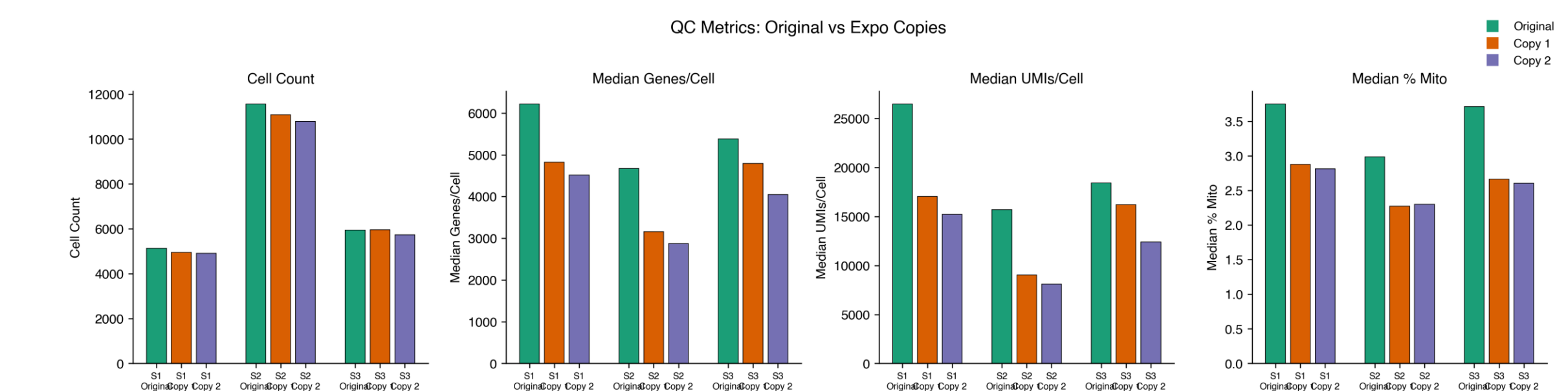
Proof of Concept: Short Read Biological Complexity



To demonstrate use of the Expo Molecular Archive in a biologically complex context, gene expression cDNA from three human kidney single cell libraries were archived. Two iterative copies of each Molecular Archive were created and compared to an unarchived control using Element short read sequencing, showing concordant gene expression and preservation of cell populations across the original samples and Molecular Archive copies.

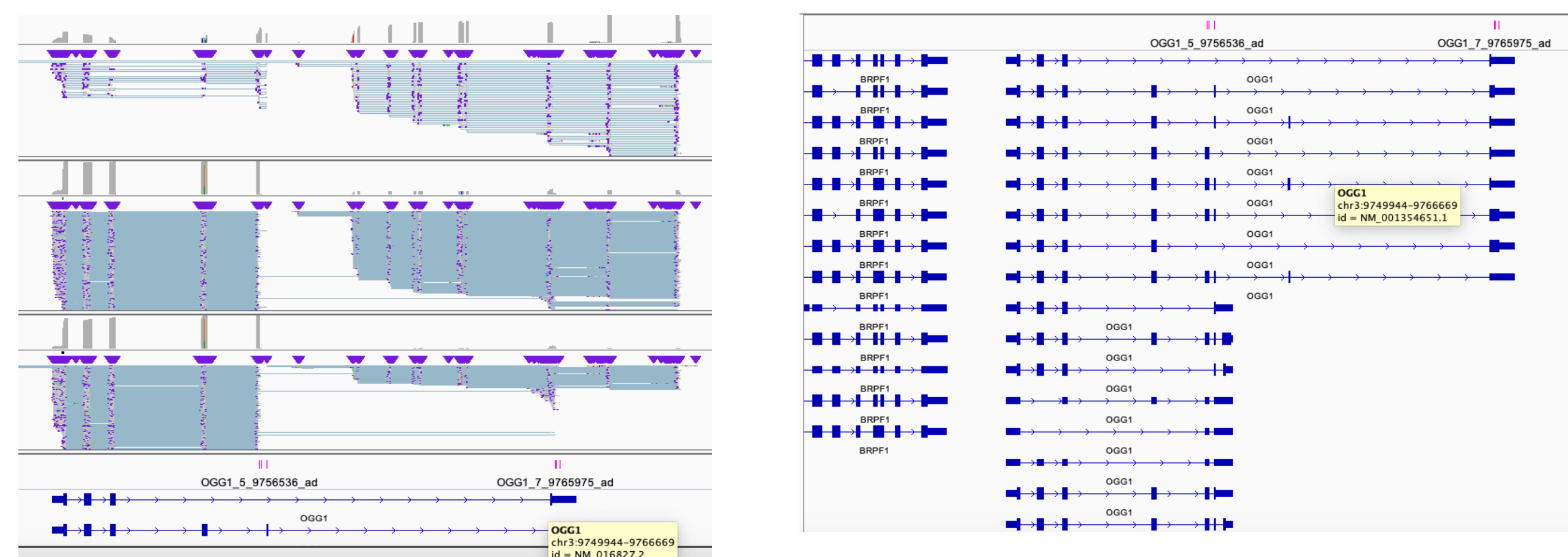


Leiden cluster and marker gene expression across original and Expo copy libraries. Three human kidney samples (rows) were each archived using the Expo platform, and two copies were generated per archive. All libraries were sequenced on Element, processed through Cell Ranger, and integrated using scVI to correct for technical differences between library versions. Left three columns: Cells colored by Leiden cluster identity. Each row uses a consistent color scheme across Original, Copy 1, and Copy 2, with per-sample cluster legends shown at right. Right three columns: Per-cell expression intensity for three genes: ECSCR (green), MACC1 (blue), and IFI27 (purple). Color intensity scales from light (low expression) to dark (high expression), with non-expressing cells in gray. In both representations, cells from Copy 1 and Copy 2 recapitulate the spatial organization and expression patterns of the original library, with no version-specific clusters or preferential loss of any cell population. Cell counts per panel are shown in the lower left.



Quality control metrics for original and Expo copy libraries. Four QC metrics are shown for three human kidney samples (S1, S2, S3), each with an original library (green), Copy 1 (orange), and Copy 2 (purple). Cell count (number of barcodes passing Cell Ranger filtering) is closely matched across versions within each sample. Median genes per cell and median UMIs per cell show a consistent reduction in copies relative to the original, as expected from an additional round of molecular capture. Median percent mitochondrial reads remains low across all libraries.

Proof of Concept: Long Read Isoform Detection



In order to demonstrate proof of concept for multiplex PCR gene panel amplification and long read isoform detection from Expo Molecular Archives, we used PCR to amplify a panel of over 20 genes from an Expo molecular archived copy of a single cell RNAseq library of wild type HEK293T-cells, which are of kidney origin and of similar tissue classification to the biological kidney tissue sample shown above. Zooming in on the OGG1 gene specific locus, and showing sequencing data from ONT long read sequencing, resulting analysis shows multiple OGG1 isoforms, which would not be detectable using short read sequencing alone.

Future Directions

Future projects will utilize amplification primers from specific cells, using the known cellular barcode, to enrich gene capture from low frequency cell populations and improve confidence in associated cell type subclassification. The technology will also be applied to spatial applications, including 10x Genomics Visium 3' HD and Takara Trekker pipelines which provide full length spatial based gene expression, providing researchers with a novel approach to multi-omic analysis of single cell and spatial libraries.

Acknowledgements

We would like to acknowledge founding partners of Exponential Molecular, including the Hanlee Ji Lab at the Stanford School of Medicine, and General Inception Partners. We would also like to acknowledge the support of the OMAPIX team, including Sam Thompson, Hyungwoo Nam and Shawn Clairmont. We would also like to thank those who contributed to planning and supporting this project, including David Symer and Chris Barbazette at Exponential Molecular and Christina Chiu from Element Biosciences.