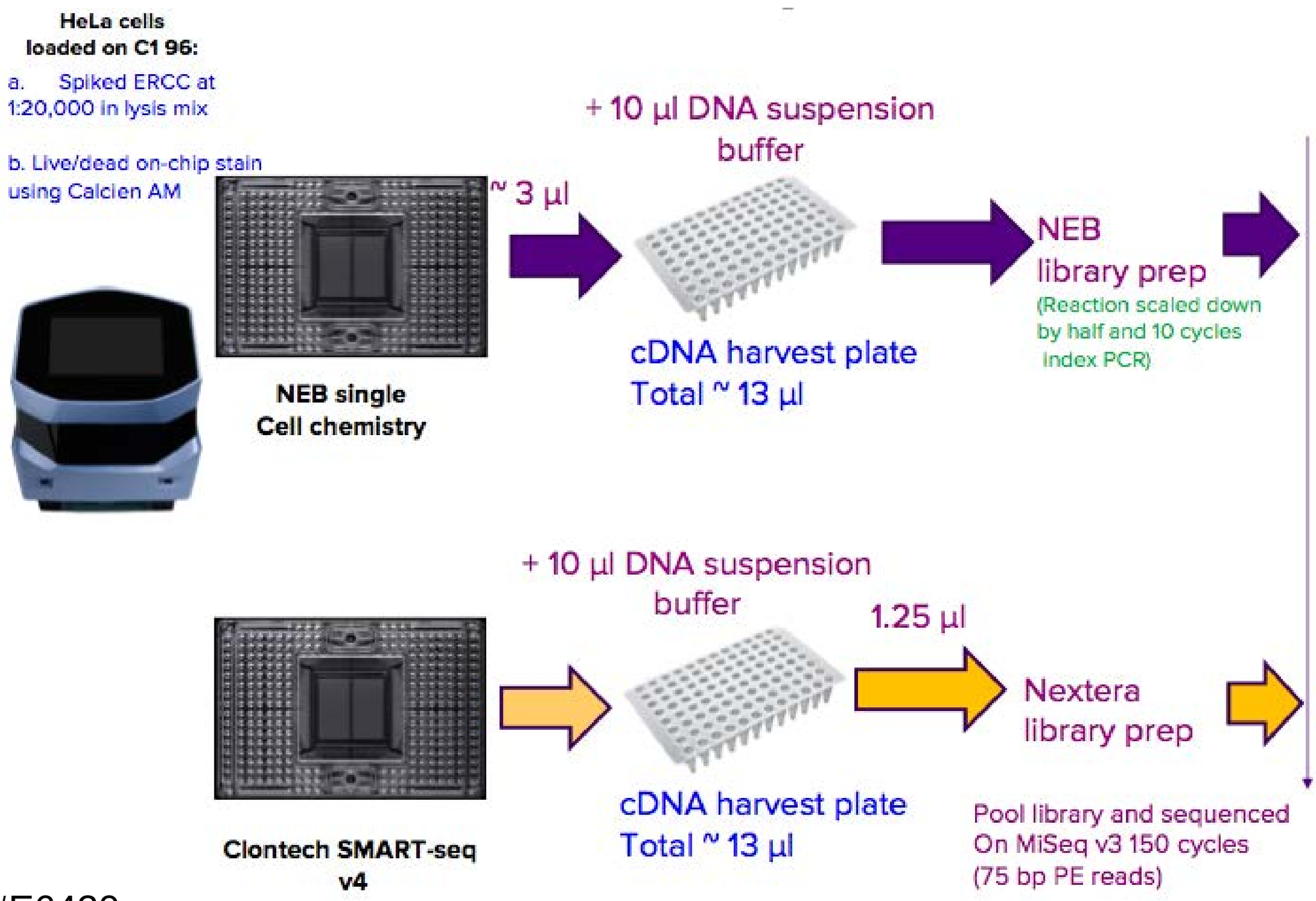
Summary of Single Cell Transcriptome Experiments Using HeLa and T Cells on Fluidigm C1

HeLa Experiment

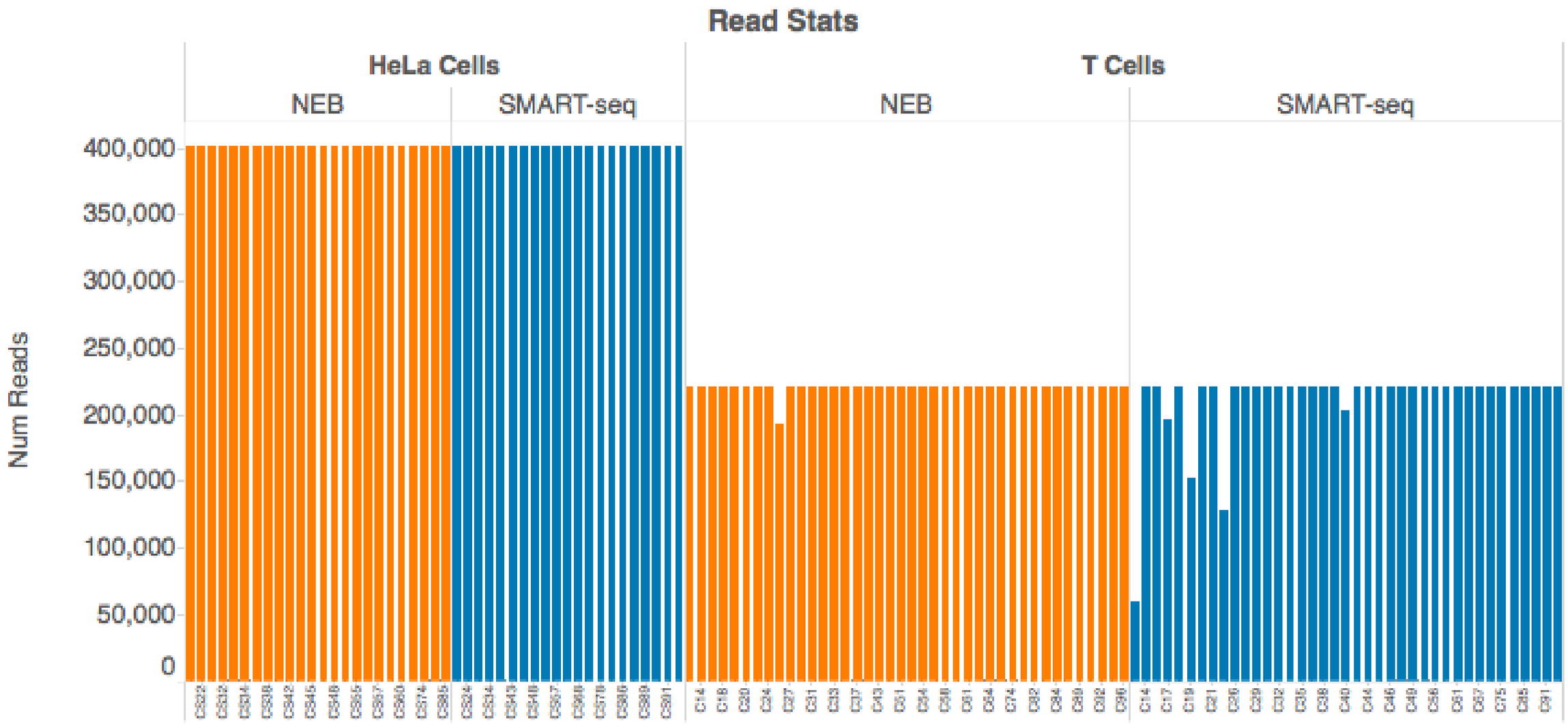


NEB #E6420

Takara #63489

T cell Experiment

- Activated frozen T cells using Dynabeads for 48h
- Removed cells from beads and ran two C1 96 IFCs:
  - NEBNext Single Cell (NEB #E6420)
  - SMART-seq v4 (Takara #63489)
- Included Live/Dead stain and imaged IFCs prior to chemistry
- Selected 48 live single cells from each IFC for cDNA synthesis and Nextera XT library prep
- NEB Libraries were amplified with 17 PCR cycles
- SMART-seq libraries were amplified with 21 PCR cycles
- Sequencing on MiSeq v3, paired end 75 bp

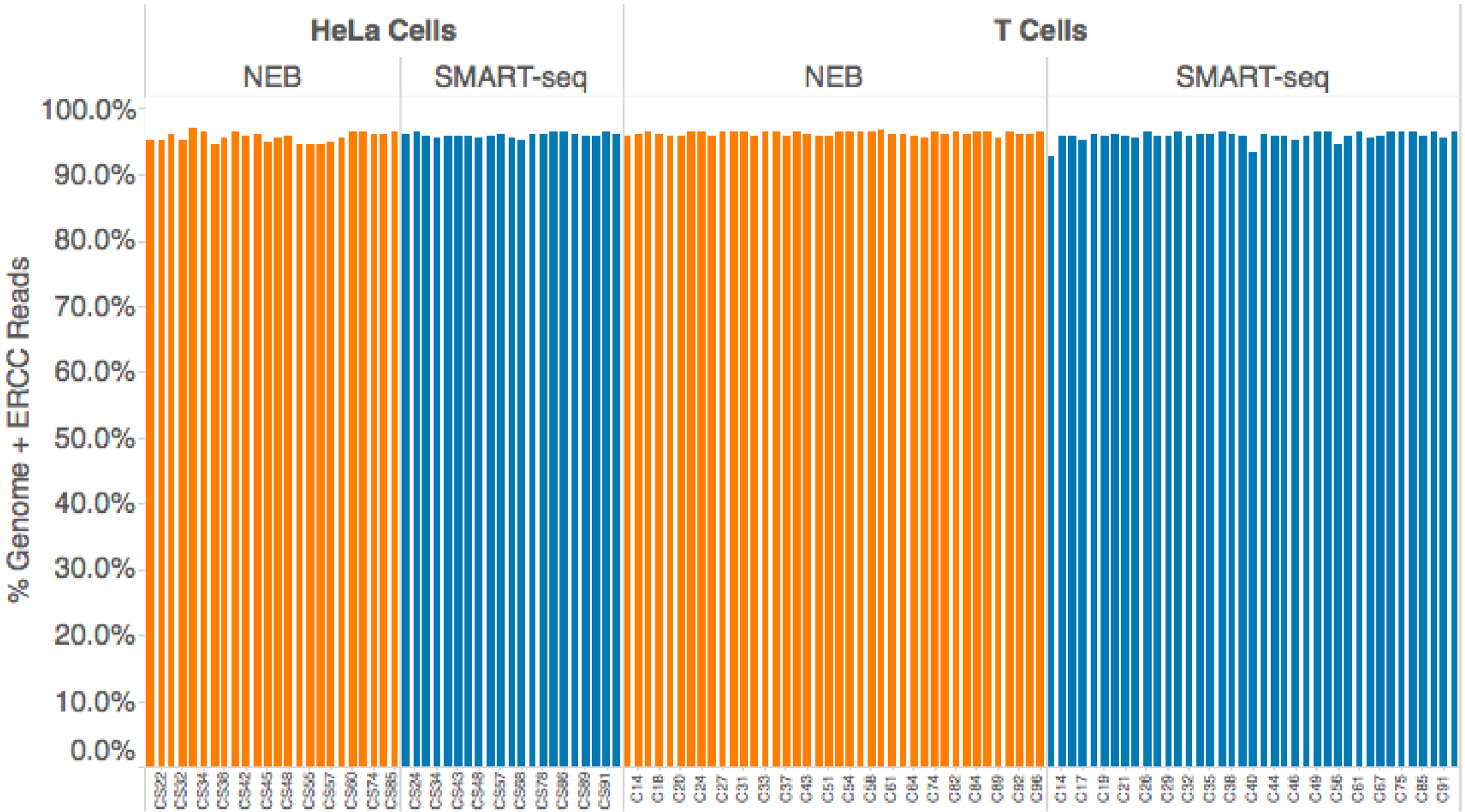


All HeLa libraries were pooled and sequenced on an Illumina MiSeq (2x75 bp) before sampling 400,000 reads per library for analysis (seqtk sample[1]). All T cell libraries were sequenced together on an Illumina MiSeq (2x75 bp) before sampling to a maximum of 220,000 reads for analysis (some libraries did not have enough reads).

Each bar represents a library derived from a single cell.

[1] http://broadinstitute.github.io/picard/

Mapping Rate

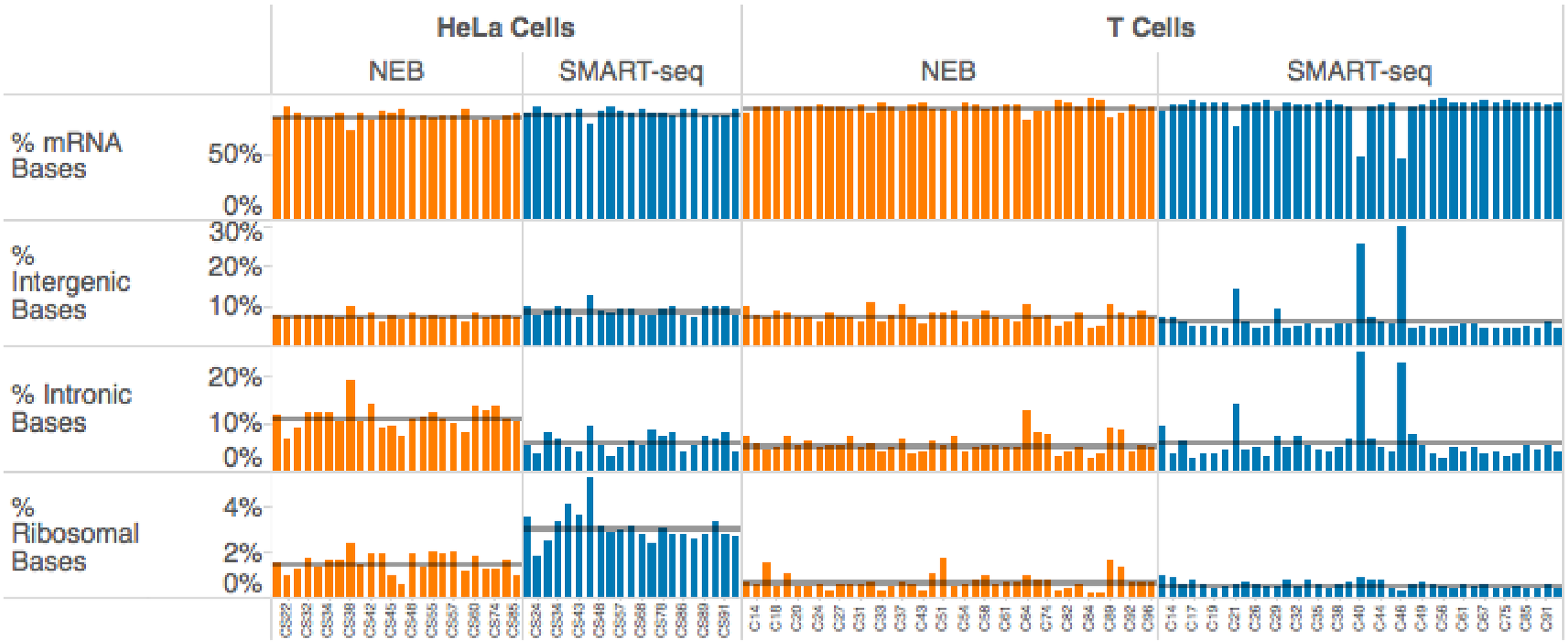


Reads were adapter trimmed[1], then aligned to the GRCh37 reference genome using Hisat 2.0.7[2]. Reads were also aligned to Gencode v28 and ERCC transcripts (added to HeLa libraries) using Salmon 0.9.1 [3].

Greater than 95% of reads aligned for all NEB libraries and nearly all SMART-seq libraries.

[1] Jiang et al., "Skewer: A Fast and Accurate Adapter Trimmer for next-Generation Sequencing Paired-End Reads."
[2] Kim, Langmead, and Salzberg, "HISAT: A Fast Spliced Aligner with Low Memory Requirements."
[3] Patro et al., "Salmon Provides Accurate, Fast, and Bias-Aware Transcript Expression Estimates Using Dual-Phase Inference."

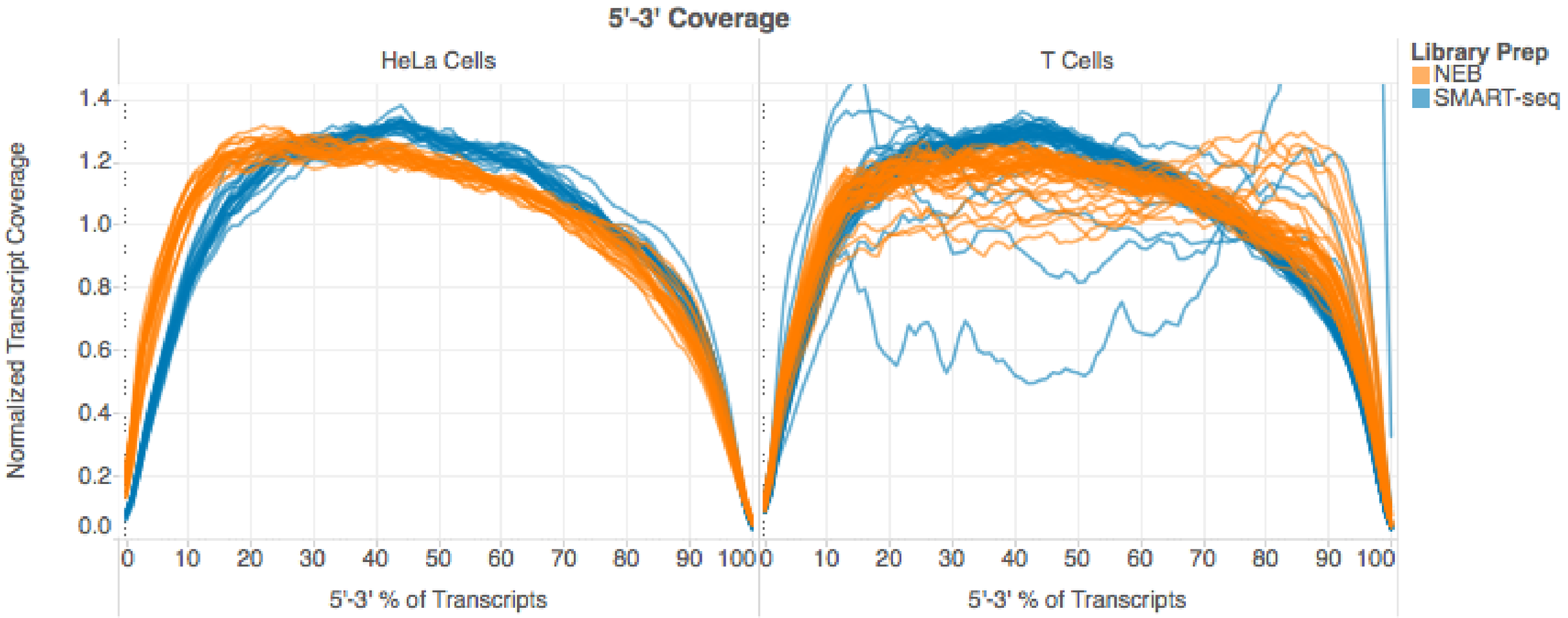
Genome Feature Distribution



Reads were adapter trimmed[1], then aligned to the GRCh37 reference genome using Hisat 2.0.7[2]. Coverage of mRNA, intronic and ribosomal regions was assessed using Picard's RNA-seq Metrics 1.56[3] and NCBI's RefSeq transcript list. Intronic bases may be due to nascent transcripts that have not yet been spliced. Bases annotated as Ribosomal include 18S, 28S, 5.8S, 5S, 16S, and 12S species. Horizontal gray lines represent an average value for all libraries.

All NEB and most SMART-seq libraries show consistent mRNA and intergenic amounts. Aside from outliers, Clontech libraries had fewer Intronic bases and more rRNA bases.

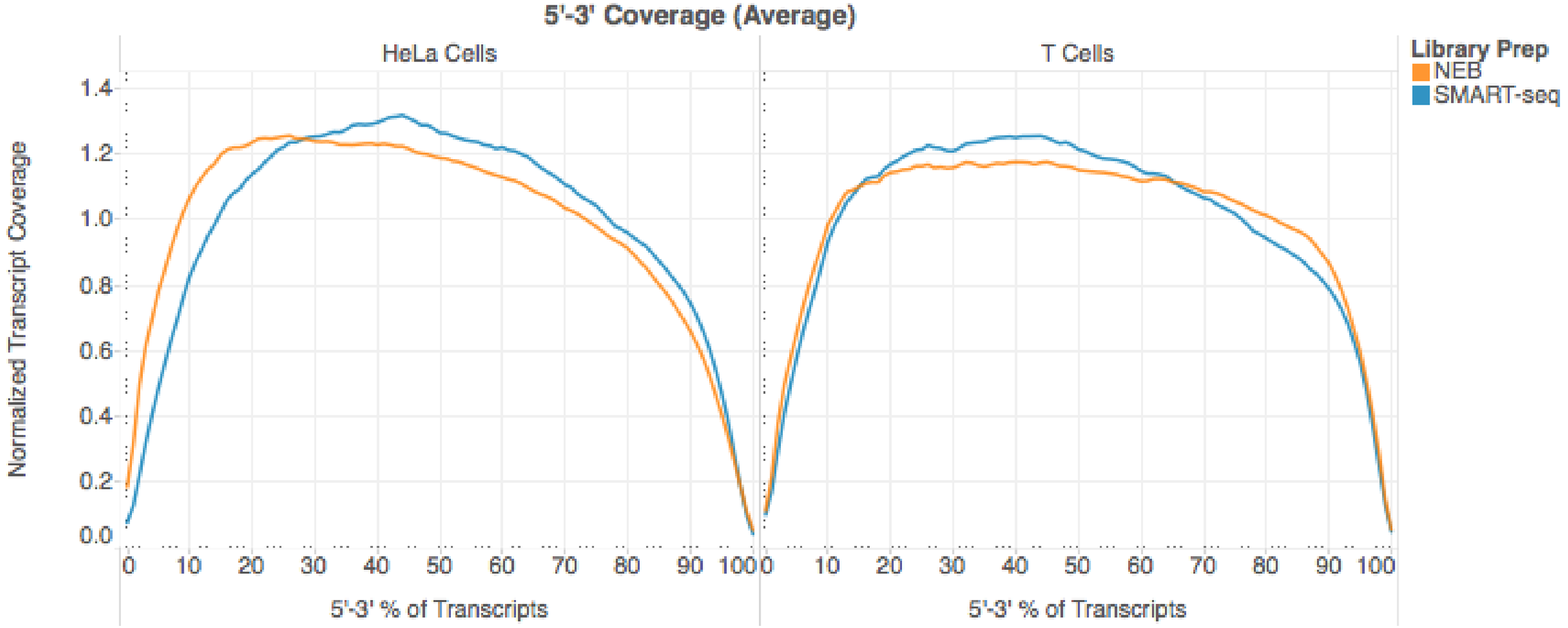
[1] Jiang et al., "Skewer: A Fast and Accurate Adapter Trimmer for next-Generation Sequencing Paired-End Reads."
[2] Kim, Langmead, and Salzberg, "HISAT: A Fast Spliced Aligner with Low Memory Requirements."
[3] http://broadinstitute.github.io/picard/



Reads were adapter trimmed[1], then aligned to the GRCh37 reference genome using Hisat 2.0.7[2]. Normalized 5'-3' coverage of RefSeq transcripts was assessed using Picard's RNA-seq Metrics 1.56[3]. T cell libraries show more variation.

Most libraries show a consitent pattern of coverage across transcripts. NEB libraries show more coverage of transcript ends, resulting in more even coverage.

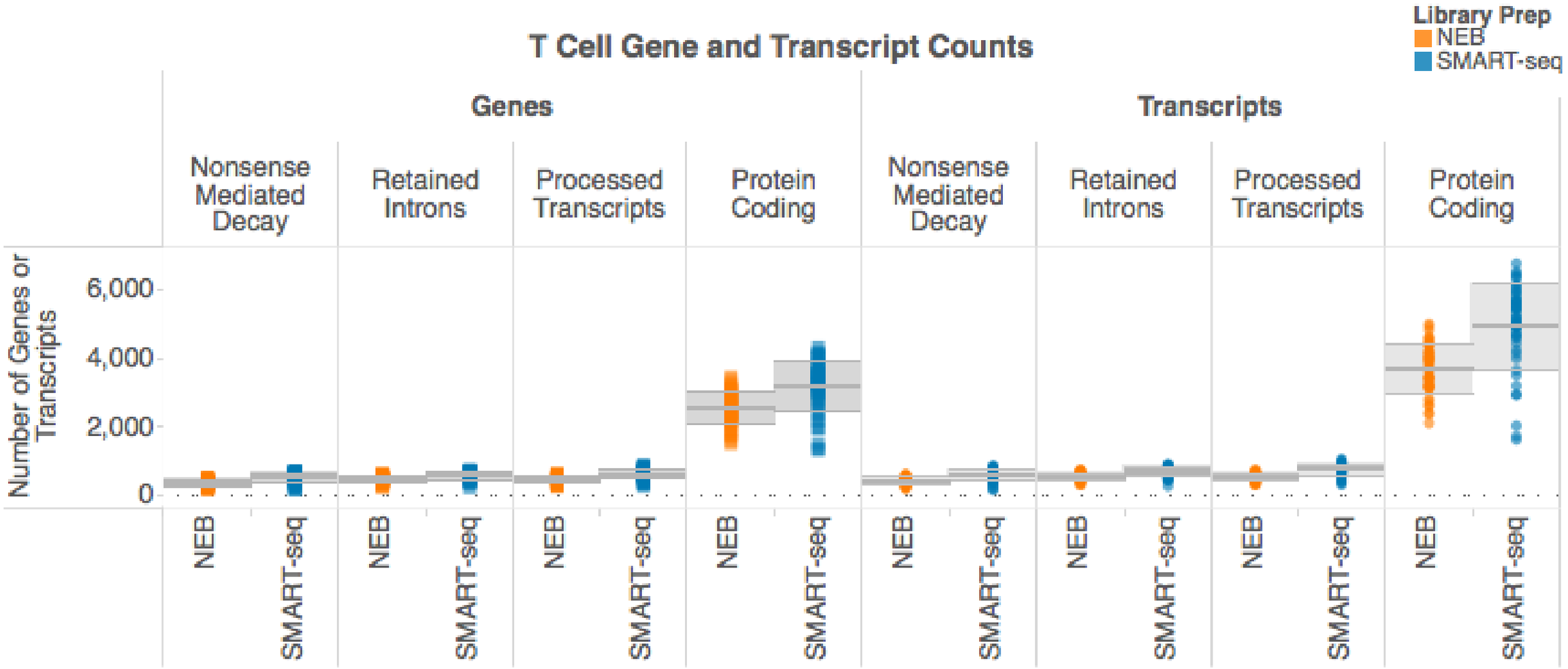
[1] Jiang et al., "Skewer: A Fast and Accurate Adapter Trimmer for next-Generation Sequencing Paired-End Reads."
[2] Kim, Langmead, and Salzberg, "HISAT: A Fast Spliced Aligner with Low Memory Requirements."
[3] http://broadinstitute.github.io/picard/



Reads were adapter trimmed[1], then aligned to the GRCh37 reference genome using Hisat 2.0.7[2]. Normalized 5'-3' coverage of RefSeq transcripts was assessed using Picard's RNA-seq Metrics 1.56[3]. Coverage is averaged for all cells in each library prep method and cell type for display simplicity.

On average, NEB libraries show more coverage of transcript ends, resulting in more even coverage.

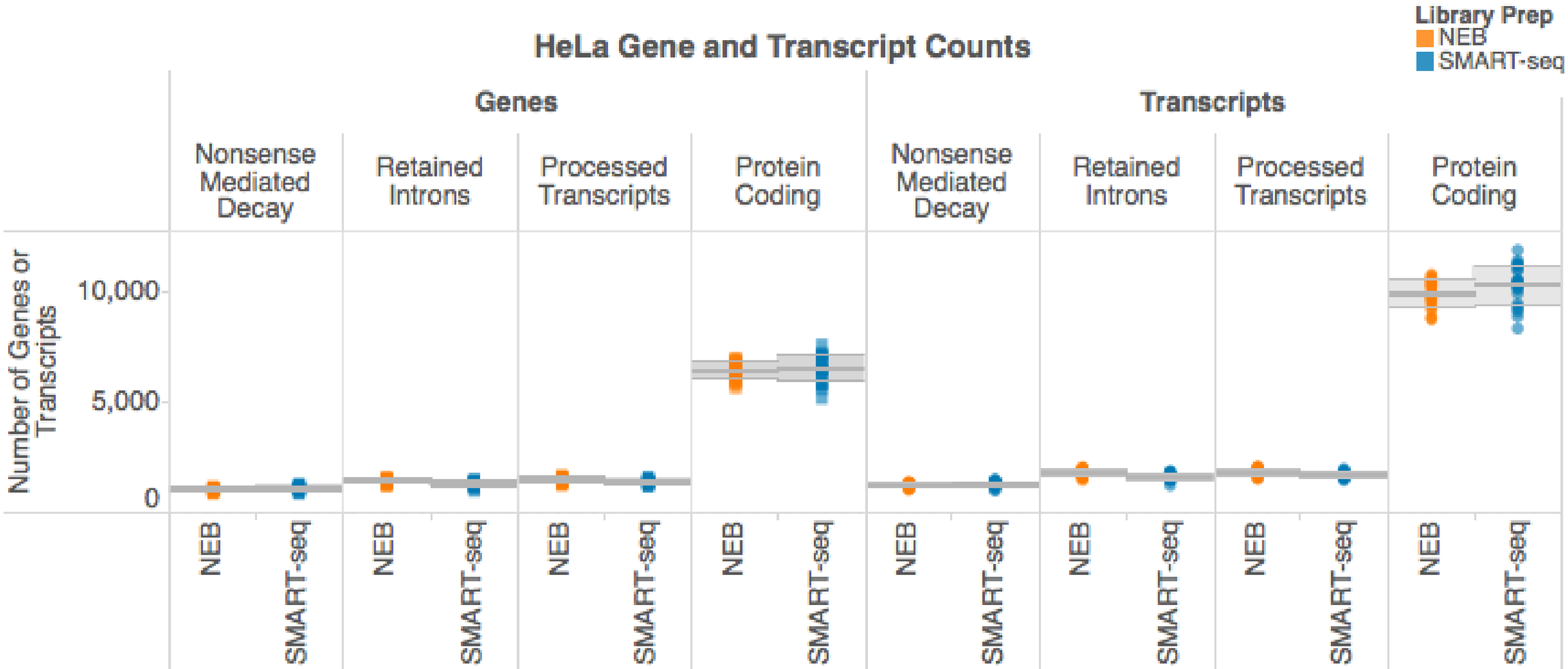
[1] Jiang et al., "Skewer: A Fast and Accurate Adapter Trimmer for next-Generation Sequencing Paired-End Reads."
[2] Kim, Langmead, and Salzberg, "HISAT: A Fast Spliced Aligner with Low Memory Requirements."
[3] http://broadinstitute.github.io/picard/



Reads were adapter trimmed[1] before mapping to Gencode v28 and ERCC transcripts using Salmon 0.9.1 [2]. At this sequencing depth, low abundance transcripts are only detected stochastically, so genes names and transcript IDs with TPM >=1 were counted. The four most abundant transcript categories (Protein coding, Processed transcripts, Retained introns and Nonsense mediated decay) are shown. Center lines indicate average counts by library type. Shaded areas indicate +/- 1 standard deviation. NEB libraries were amplified with 17 PCR cycles. SMART-seq libraries were amplified with 21 PCR cycles.

NEB and SMART-seq libraries produced numbers of transcripts and genes within experimental error of each other. NEB libraries produced more consistent counts.

[1] Jiang et al., "Skewer: A Fast and Accurate Adapter Trimmer for next-Generation Sequencing Paired-End Reads."
[2] Patro et al., "Salmon Provides Accurate, Fast, and Bias-Aware Transcript Expression Estimates Using Dual-Phase Inference."



Reads were adapter trimmed[1] before mapping to Gencode v28 and ERCC transcripts using Salmon 0.9.1 [2]. At this sequencing depth, low abundance transcripts are only detected stochastically, so genes names and transcript IDs with TPM >=1 were counted. The four most abundant transcript categories (Protein coding, Processed transcripts, Retained introns and Nonsense mediated decay) are shown. Center lines indicate average counts by library type. Shaded areas indicate +/- 1 standard deviation.

NEB and SMART-seq libraries produced numbers of transcripts and genes within experimental error of each other. NEB libraries produced more consistent counts.

[1] Jiang et al., "Skewer: A Fast and Accurate Adapter Trimmer for next-Generation Sequencing Paired-End Reads."
[2] Patro et al., "Salmon Provides Accurate, Fast, and Bias-Aware Transcript Expression Estimates Using Dual-Phase Inference."

Results Summary

- Fluidigm C1 produces consistent results.
- NEBNext Single Cell or SMART-seq V4 chemistry can be used for cDNA production and showed comparable performance.
- NEB libraries show greater consistency between cells across all metrics studied (number of reads, mapping rate, 5’-3’ coverage, genome features, and transcript detection).