SERVICE REPORT

mRNA NGS Profiling Data Report

SV00001

SERVICE REPORT

Customer: Dr John Doe
Company/Institute: ABC Company
Date: January 1, 2020

Performed by:
Norgen Biotek Corp.
3430 Schmon Pkwy, Thorold, ON
Canada L2V 4Y6

NORGEN BIOTEK CORP.

TABLE OF CONTENTS

Report Summary	3
RNA-Seq Data Analysis Workflow Used	4
Downloading Your Files	4
Structure of Files Provided	5
Sample Raw Read File Information	6
Quantification of RNA used for RNA-Seq	6
Sequencing Summary	6
Mapping Summary	7
Help and Technical Support	9

Report Summary

Client Name	Dr John Doe (ABC Company)
Norgen Order Number	SV00001
Number of Samples to Process	6 Samples
Sample Type	Plasma (Human)
Sequencing Service Provided	RNA-Seq
Sequencing Platform	Illumina NextSeq 500
Sequencing Platform Reagent	NextSeq 500/550 Mid Output Kit v2.5 (2 x 75 Cycles using a 150-Cycle Kit)
Product Used for Library Preparation	NEBNext® Poly(A) mRNA Magnetic Isolation Module (E7490) and NEBNext® Ultra™ II Directional RNA Library Prep Kit for Illumina (E7760)

Date and Time of Release of Report:		January 1, 2020	
Report Prepared by:	Dr. Mohamed El-Mogy	**Date:**	January 1, 2020
Report Approved by:	Dr. Yousef Haj-Ahmad	**Date:**	January 1, 2020

RNA-Seq Data Analysis Workflow Used

Cut-Adapt: Adapter trimming

Martin, M. (2011) . Cutadapt removed adapter sequences from high-throughput sequencing reads. EMBnet.journal, 10-12 DOI: 10.14806/ej.17.1.200.
https://journal.embnet.org/index.php/embnetjournal/article/view/200/479

STAR: Alignment

Dobin, A., Davis, C., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., et al. (2012). STAR: ultrafast universal RNA-seq aligner. Bioinformatics (Oxford, England), 15-21. DOI: 10.1093/bioinformatics/bts635.
http://bioinformatics.oxfordjournals.org/content/early/2012/10/25/bioinformatics.bts635

FeatureCounts: Annotation counting

Liao, Y., Smyth GK., Shi, W. (2014). FeatureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics (Oxford, England), 923-30. DOI: 10.1093/bioinformatics/btt656. http://academic.oup.com/bioinformatics/article/30/7/923/232889

Reference or Database Sequences Used

UCSC hg38 (Human)

Downloading Your Files

All files pertaining to this service have been uploaded to our servers. These files will be available for up to 1 month after the completion of the project. Please make sure to download all files before then.

To access your files, visit the link below and enter the user name and password provided. Click the Download button to start the download. If you have any issues accessing the files, please contact us for help.

Please note that all fields are case-sensitive.

Download Link:	https:\\services.norgenbiotek.com\sv00001
Username:	user-sv00001
Password:	AaBb1234
Expiration:	April 1, 2020

Structure of Files Provided

Inside the provided hard drive, you will find the following file structure that contains the analysis results.

- Raw Reads
  - Contains compressed original FASTQ file of each sample
- RNA Analysis
  - Mapping Files Per Sample
    - Contains alignment files (.BAM) and count files for individual samples
  - Post Processed Results
    - Contains summary of read counts to genes (*.txt format)
    - Mapping summary provided in the file: StatsSummary.txt

Sample Raw Read File Information

Norgen ID	Sample ID	Index 1	Associated Read 1 FATSQ File Name	Index 2	Associated Read 2 FATSQ File Name
SV00001-0001	1_SAMPLE1	I501	SV00001-0001_R1_S1_001	I701	SV00001-0001_R2_S1_001
SV00001-0002	2_SAMPLE2	I502	SV00001-0002_R1_S2_0001	I702	SV00001-0002_R2_S2_0001
SV00001-0003	3_SAMPLE3	I503	SV00001-0003_R1_S3_0001	I703	SV00001-0003_R2_S3_0001
SV00001-0004	4_SAMPLE4	I504	SV00001-0004_R1_S4_001	I704	SV00001-0004_R2_S4_001
SV00001-0005	5_SAMPLE5	I505	SV00001-0005_R1_S5_001	I705	SV00001-0005_R2_S5_001
SV00001-0006	6_SAMPLE6	I506	SV00001-0006_R1_S6_001	I706	SV00001-0006_R2_S6_001

Quantification of RNA used for RNA-Seq

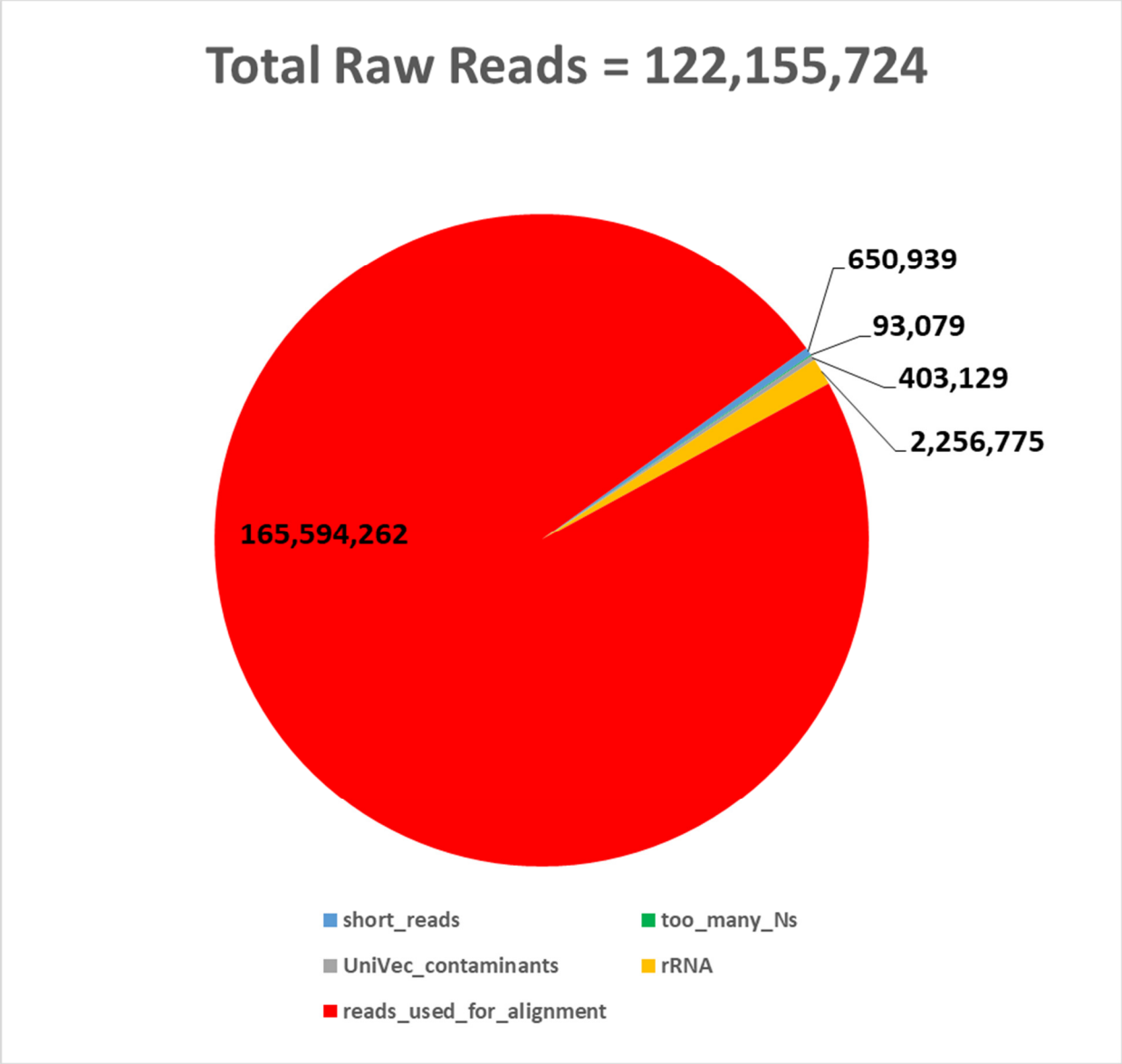
Sample ID	RNA ID	RNA Conc. ng/uL (Bioanalyzer)
SV00001-0001	SV00001-0001_0001_RNA	90
SV00001-0002	SV00001-0002_0002_RNA	80
SV00001-0003	SV00001-0003_0003_RNA	62
SV00001-0004	SV00001-0004_0004_RNA	58
SV00001-0005	SV00001-0005_0005_RNA	67
SV00001-0006	SV00001-0006_0006_RNA	79

Amount of RNA is in line with observed yield of similar RNA samples. It meets the minimal requirement for library prep and sequencing.

Sequencing Summary

Norgen ID	Sample ID	Read 1 Length	Read 2 Length	Number of Reads
SV00001-0001	1_SAMPLE1	75	75	15,375,339
SV00001-0002	2_SAMPLE2	75	75	20,582,321
SV00001-0003	3_SAMPLE3	75	75	13,480,565
SV00001-0004	4_SAMPLE4	75	75	13,504,597
SV00001-0005	5_SAMPLE5	75	75	17,107,739
SV00001-0006	6_SAMPLE6	75	75	23,105,163
Average Read Per Sample (Paired-End Reads)				**17,192,621**

Mapping Summary



	Total Sample	SV00001-0001	SV00001-0002	SV00001-0003	SV00001-0004	SV00001-0005	SV00001-0006
Total Raw Read	**122,155,724**	20,375,339	24,582,321	17,480,565	16,504,597	20,107,739	23,105,163
short_reads	**650,939**	50,295	90,088	44,446	43,054	74,000	349,056
too_many_Ns	**93,079**	8,615	11,257	7,520	7,574	9,428	48,685
passed_initial_qc	**165,508,190**	15,316,429	20,480,976	13,428,599	13,453,969	17,024,311	85,803,906
rRNA	**2,256,775**	146,812	256,175	119,477	184,050	320,470	1,229,791
Total Mappable Reads	**165,594,262**	17,601,338	20,226,117	13,307,539	13,259,495	16,693,212	84,506,561
reads mapped to genome	**162,696,460**	15,161,975	20,227,165	13,299,786	13,205,368	16,578,243	84,223,923
not_mapped_to_genome_or_libs	**8,944,528**	131,001	1,199,636	1,121,770	1,165,792	1,272,890	4,053,439
Gene_Sense_Count	**126,133,578**	11,584,285	16,054,694	9,910,752	9,771,740	10,747,424	68,064,683
Gene_Antisense_Count	**11,917,009**	1,367,096	1,516,367	1,316,413	1,314,424	1,396,580	5,006,129
Other_Sense_Count	**7,667,472**	1,152,422	1,212,672	1,131,608	1,133,333	1,176,549	1,860,888
Other_Antisense_Count	**12,719,090**	1,497,535	1,669,194	1,434,712	1,367,976	1,448,154	5,301,519

Sample ID	SV00001-0001	SV00001-0002	SV00001-0003	SV00001-0004	SV00001-0005	SV00001-0006
Number of detected genes at > 5 raw reads	13,017	13,221	12,859	12,781	12,934	14,315
Number of detected genes at > 1 raw read	14,704	14,816	14,480	14,481	14,627	16,068

Help and Technical Support

For any questions or additional help regarding this report, please contact our Technical Support Team between the hours of 9:00 AM and 5:30 PM (Eastern Standard Time) at (905) 227-8848 or Toll Free at 1-866-667-4362. Technical support can also be obtained through email at techsupport@norgenbiotek.com.

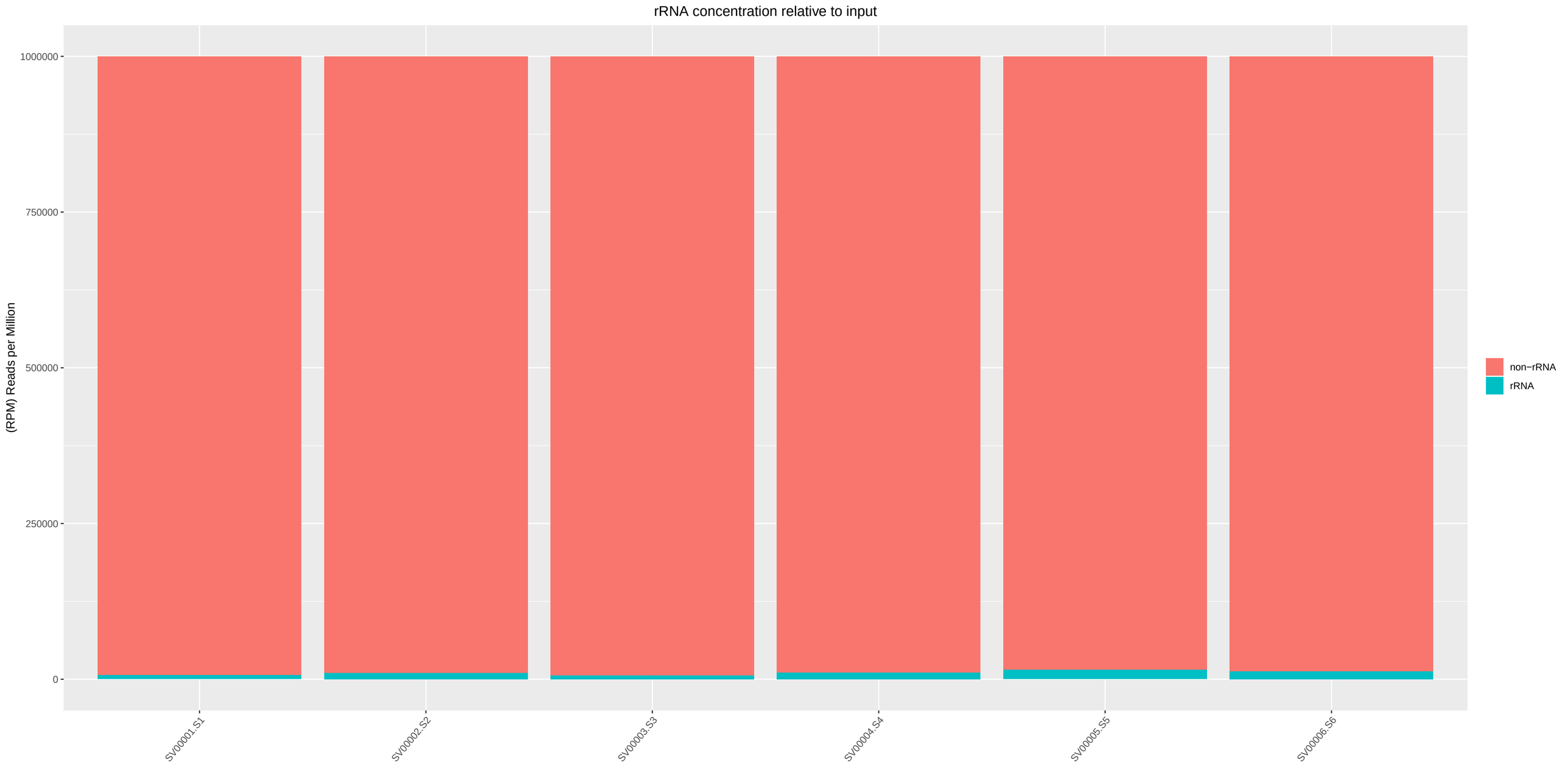
Norgen Biotek Corp.
3430 Schmon Parkway, Thorold, ON Canada L2V 4Y6
Copyright © 2020

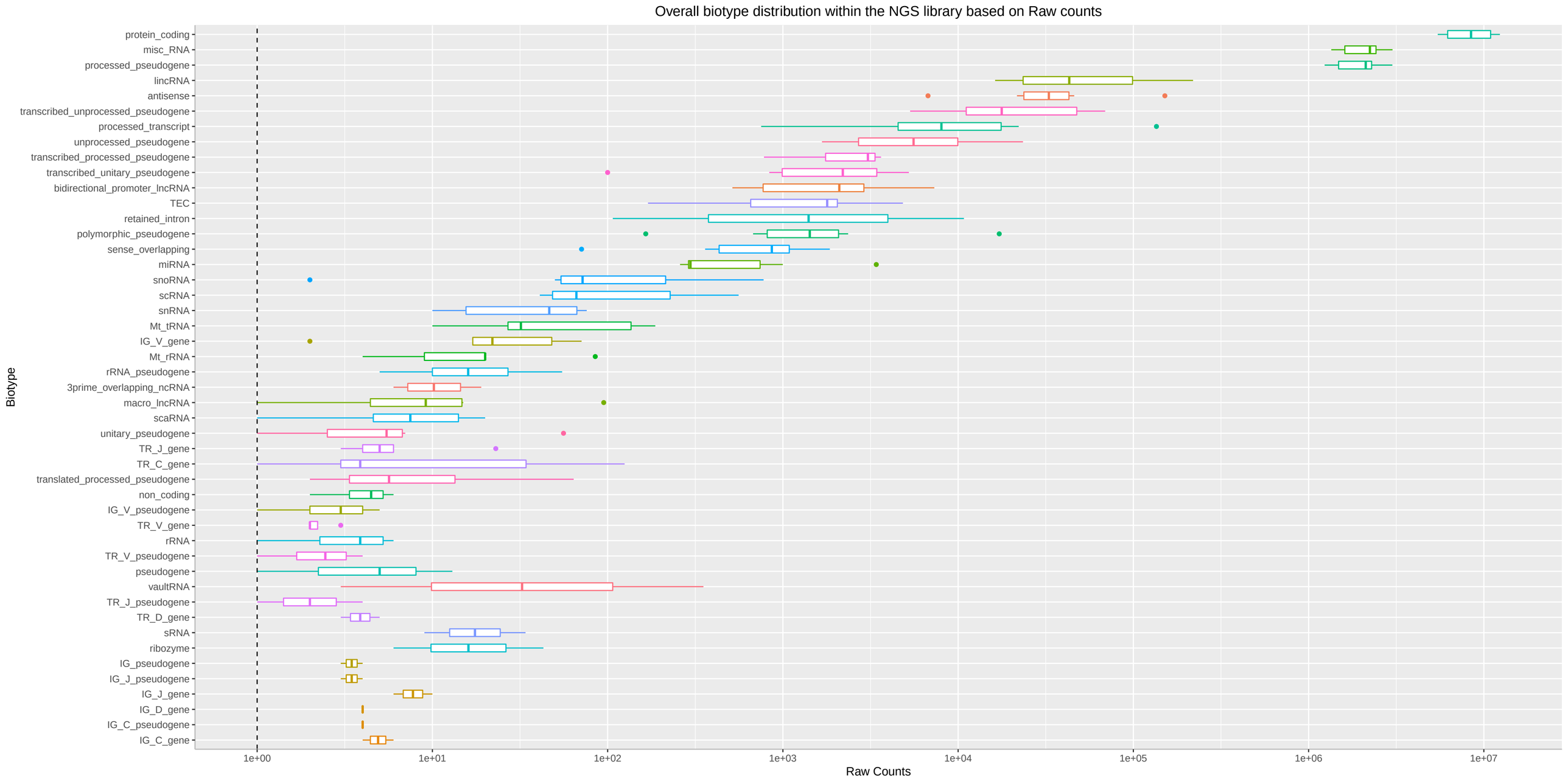
Fractioned aligned reads normalized by input

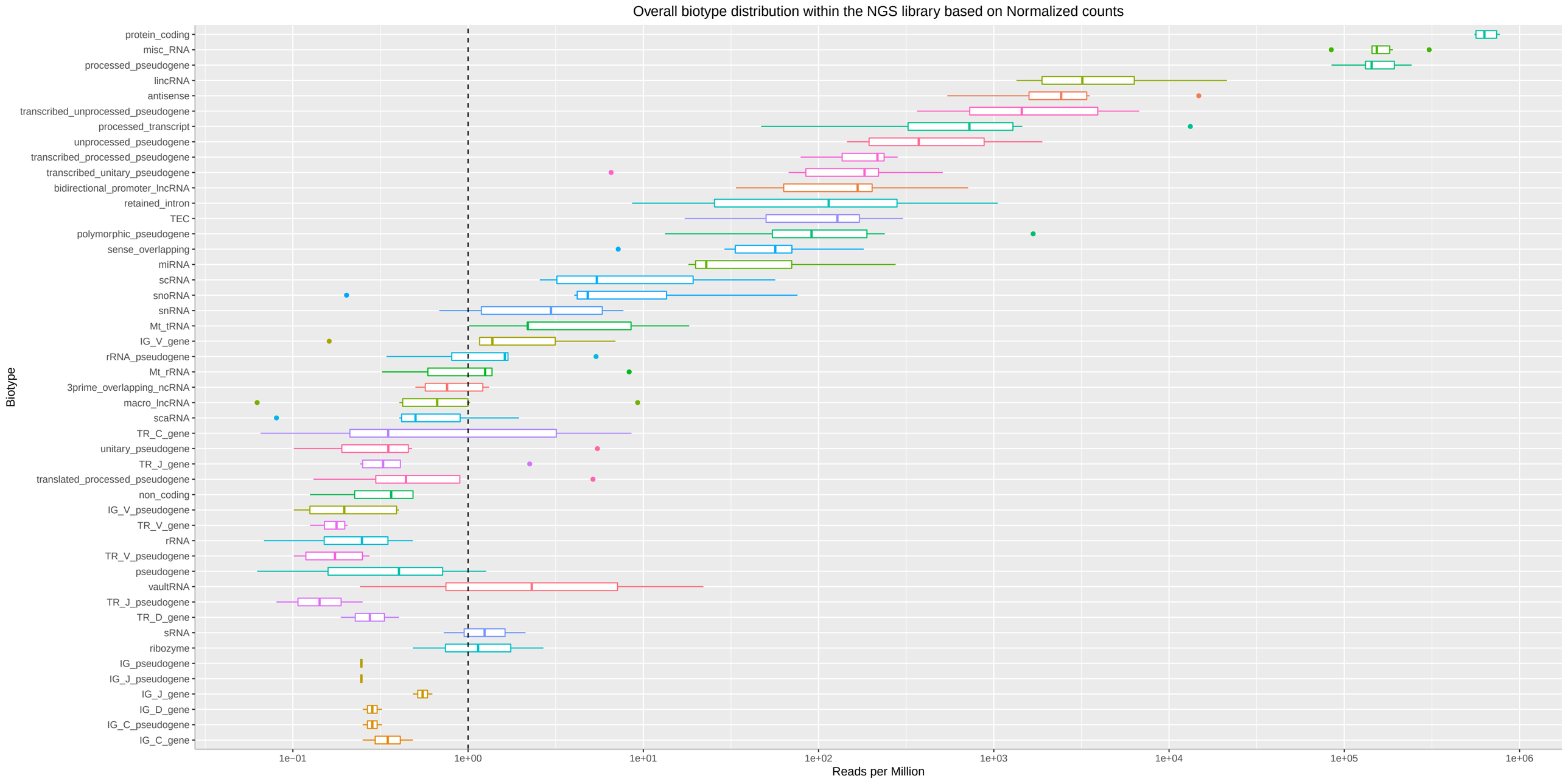
Stage	SV00001.S1	SV00002.S2	SV00003.S3	SV00004.S4	SV00005.S5	SV00006.S6
input	100	100	100	100	100	100
reads_without_adapter	0.29	0.41	0.3	0.31	0.41	0.44
short_reads	0.25	0.37	0.25	0.26	0.37	0.38
passed_initial_qc	75.17	83.32	76.82	81.52	84.67	94.08
UniVec_contaminants	0.13	0.16	0.13	0.21	0.23	0.26
rRNA	0.72	1.04	0.68	1.12	1.59	1.35
reads_used_for_alignment	86.39	82.28	76.13	80.34	83.02	92.66
genome	74.41	82.28	76.08	80.01	82.45	92.35
Gene_Sense_Count	56.85	65.31	56.7	59.21	53.45	74.63
Gene_Antisense_Count	6.71	6.17	7.53	7.96	6.95	5.49
Other_Sense_Count	5.66	4.93	6.47	6.87	5.85	2.04
Other_Antisense_Count	7.35	6.79	8.21	8.29	7.2	5.81
not_mapped_to_genome_or_transcriptome	0.64	4.88	6.42	7.06	6.33	4.44

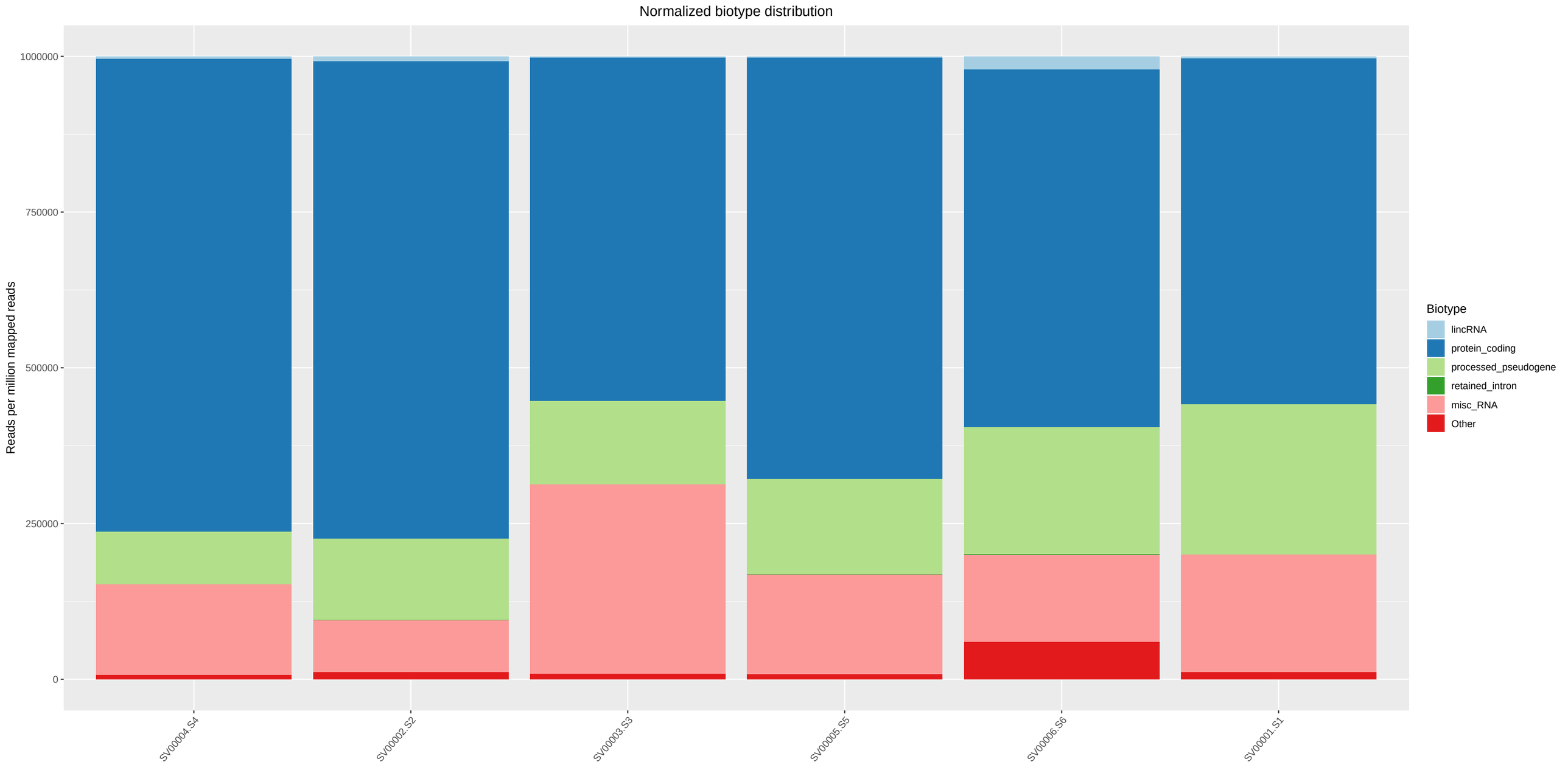
Sample

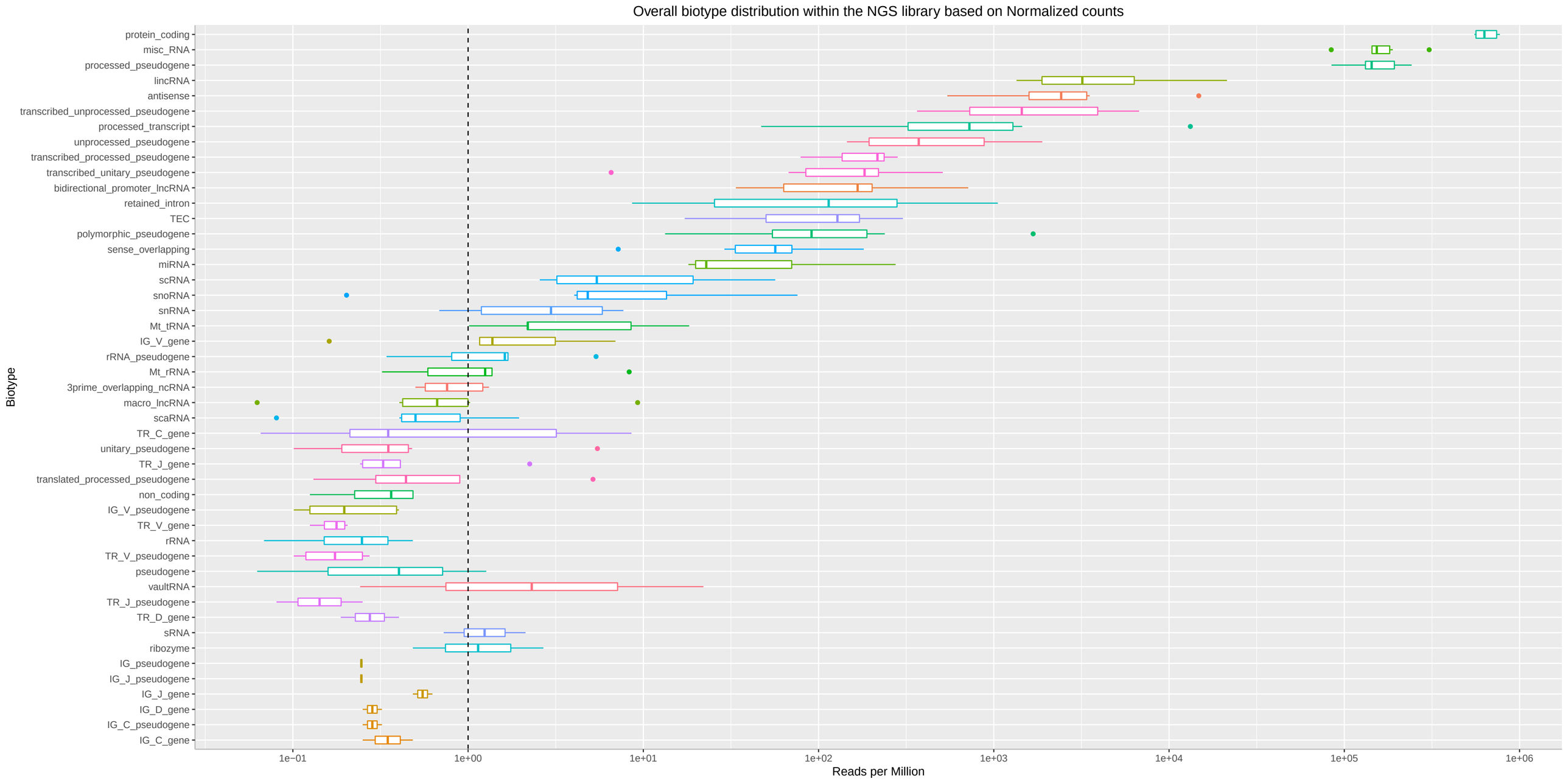
Fraction
100
75
50
25

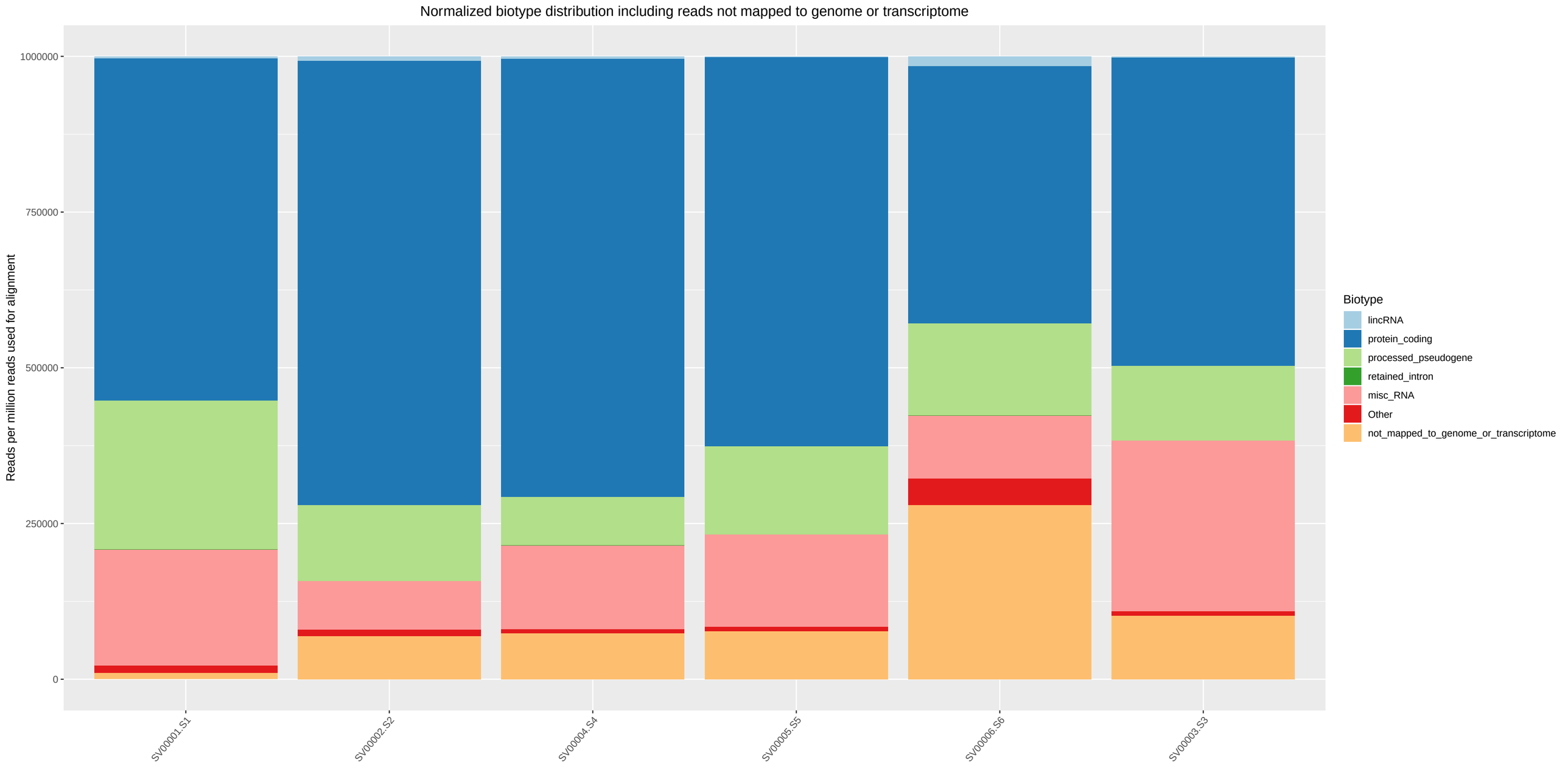

rRNA concentration relative to input
(RPM) Reads per Million
1000000
750000
500000
250000
0
SV00001.S1
SV00002.S2
SV00003.S3
SV00004.S4
SV00005.S5
SV00006.S6
non−rRNA
rRNA

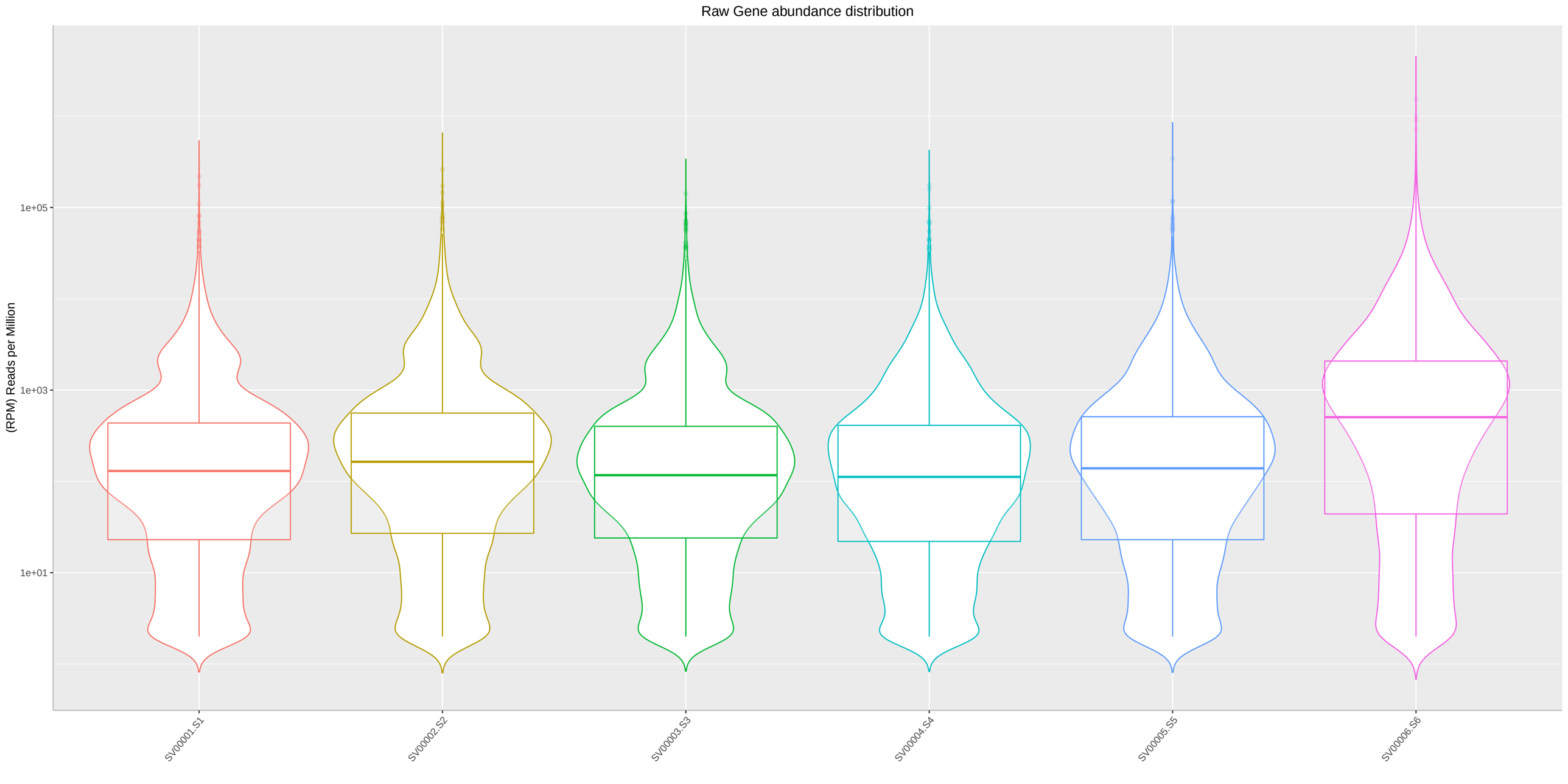

Overall biotype distribution within the NGS library based on Raw counts
Biotype
protein_coding
misc_RNA
processed_pseudogene
lincRNA
antisense
transcribed_unprocessed_pseudogene
processed_transcript
unprocessed_pseudogene
transcribed_processed_pseudogene
transcribed_unitary_pseudogene
bidirectional_promoter_lncRNA
TEC
retained_intron
polymorphic_pseudogene
sense_overlapping
miRNA
snoRNA
scRNA
snRNA
Mt_tRNA
IG_V_gene
Mt_rRNA
rRNA_pseudogene
3prime_overlapping_ncRNA
macro_lncRNA
scaRNA
unitary_pseudogene
TR_J_gene
TR_C_gene
translated_processed_pseudogene
non_coding
IG_V_pseudogene
TR_V_gene
rRNA
TR_V_pseudogene
pseudogene
vaultRNA
TR_J_pseudogene
TR_D_gene
sRNA
ribozyme
IG_pseudogene
IG_J_pseudogene
IG_J_gene
IG_D_gene
IG_C_pseudogene
IG_C_gene
1e+00
1e+01
1e+02
1e+03
1e+04
1e+05
1e+06
1e+07
Raw Counts

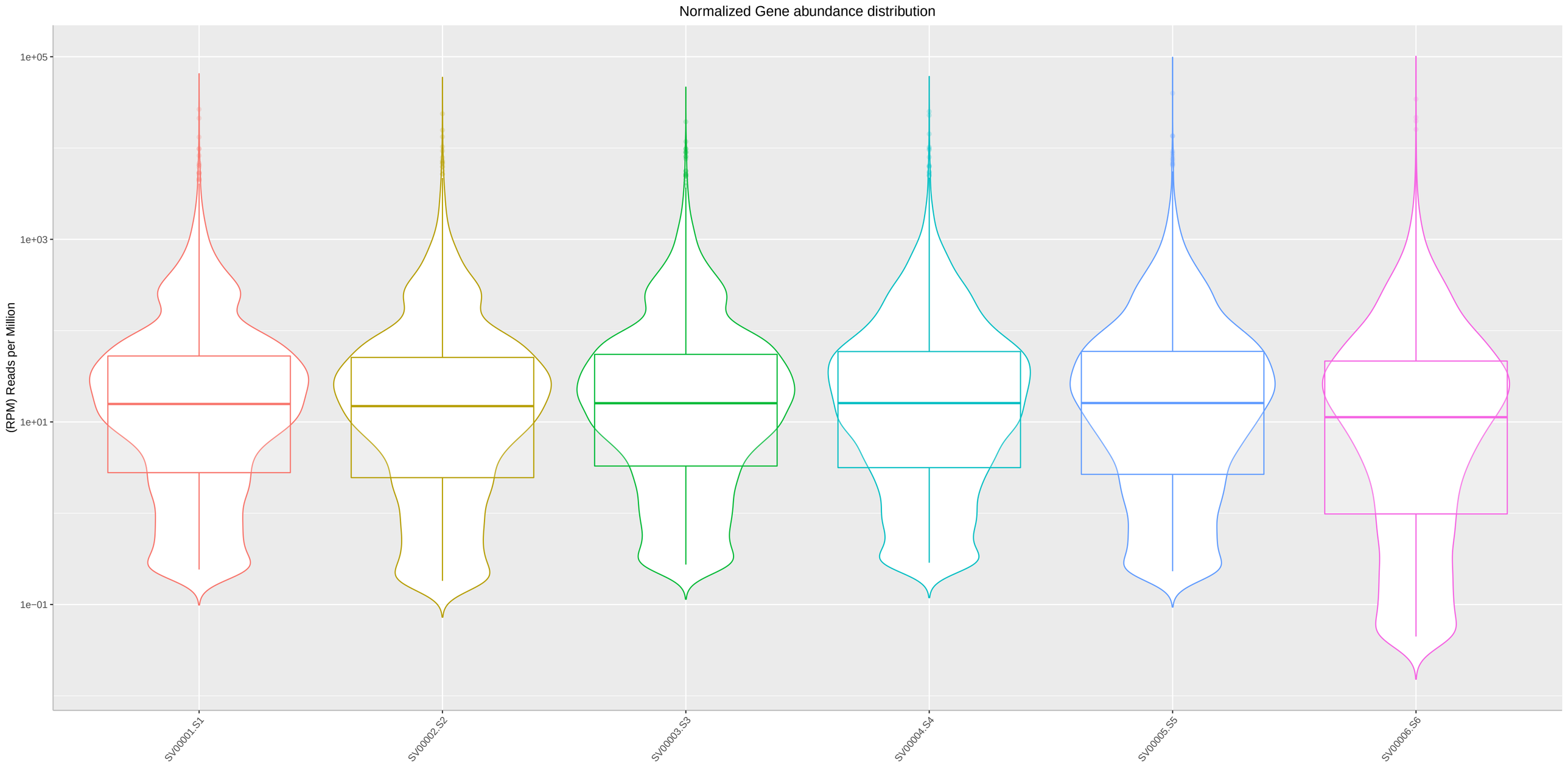

Overall biotype distribution within the NGS library based on Normalized counts
Biotype
protein_coding
misc_RNA
processed_pseudogene
lincRNA
antisense
transcribed_unprocessed_pseudogene
processed_transcript
unprocessed_pseudogene
transcribed_processed_pseudogene
transcribed_unitary_pseudogene
bidirectional_promoter_lncRNA
retained_intron
TEC
polymorphic_pseudogene
sense_overlapping
miRNA
scRNA
snoRNA
snRNA
Mt_tRNA
IG_V_gene
rRNA_pseudogene
Mt_rRNA
3prime_overlapping_ncRNA
macro_lncRNA
scaRNA
TR_C_gene
unitary_pseudogene
TR_J_gene
translated_processed_pseudogene
non_coding
IG_V_pseudogene
TR_V_gene
rRNA
TR_V_pseudogene
pseudogene
vaultRNA
TR_J_pseudogene
TR_D_gene
sRNA
ribozyme
IG_pseudogene
IG_J_pseudogene
IG_J_gene
IG_D_gene
IG_C_pseudogene
IG_C_gene
1e−01
1e+00
1e+01
1e+02
1e+03
1e+04
1e+05
1e+06
Reads per Million

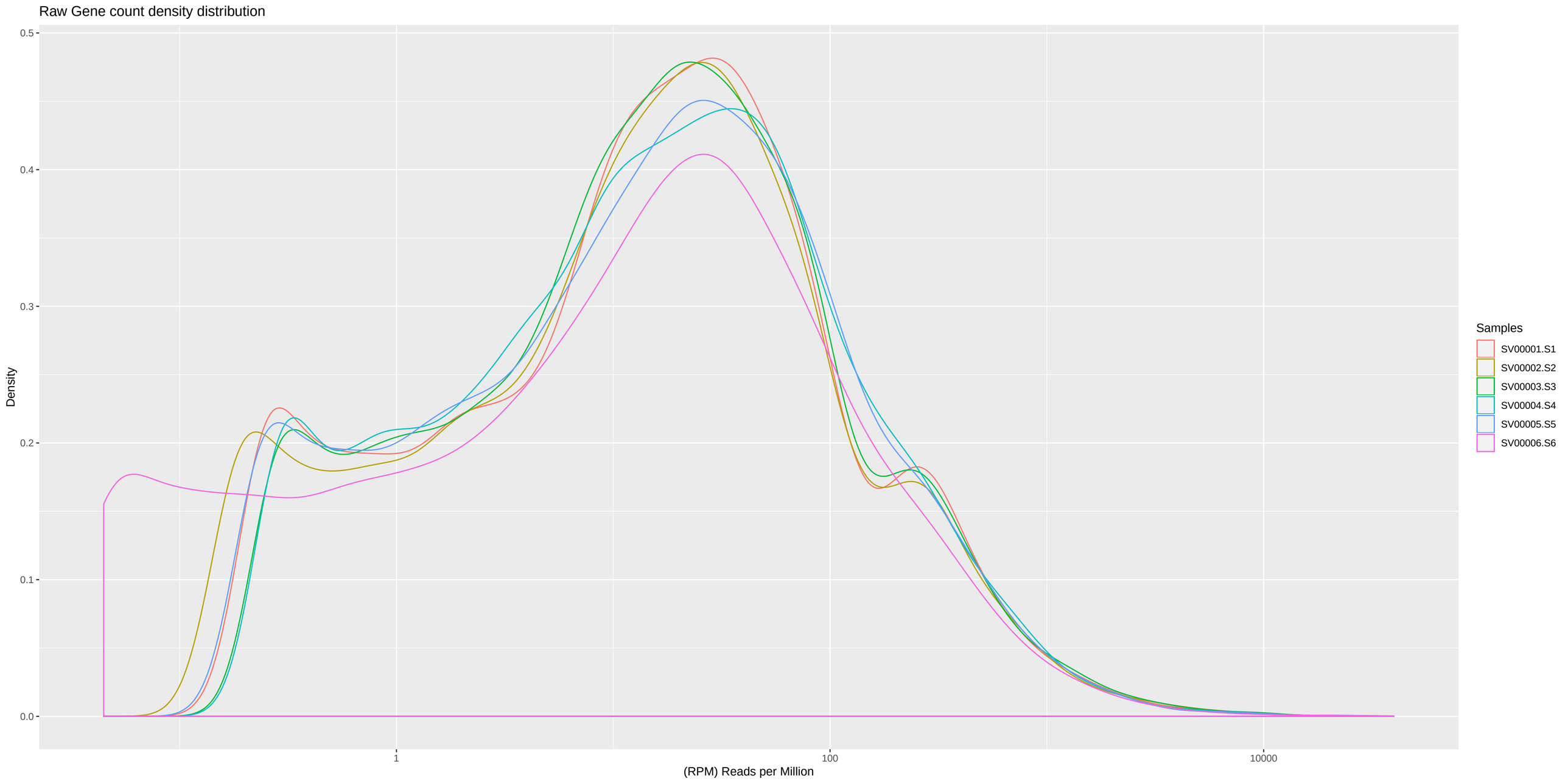

Normalized biotype distribution
Reads per million mapped reads
1000000
750000
500000
250000
0
SV00004.S4
SV00002.S2
SV00003.S3
SV00005.S5
SV00006.S6
SV00001.S1
Biotype
lincRNA
protein_coding
processed_pseudogene
retained_intron
misc_RNA
Other

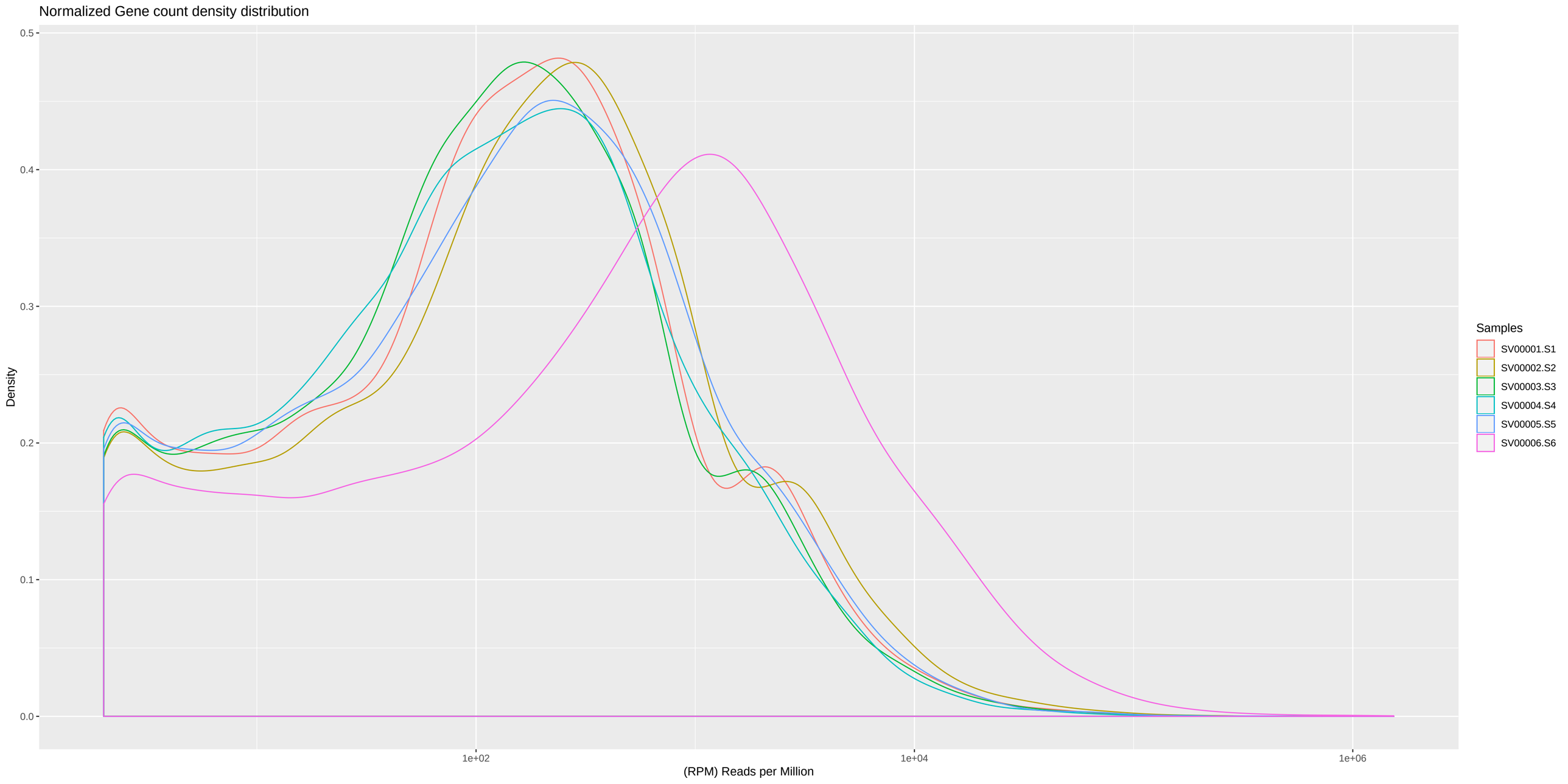

Overall biotype distribution within the NGS library based on Normalized counts
Biotype
protein_coding
misc_RNA
processed_pseudogene
lincRNA
antisense
transcribed_unprocessed_pseudogene
processed_transcript
unprocessed_pseudogene
transcribed_processed_pseudogene
transcribed_unitary_pseudogene
bidirectional_promoter_lncRNA
retained_intron
TEC
polymorphic_pseudogene
sense_overlapping
miRNA
scRNA
snoRNA
snRNA
Mt_tRNA
IG_V_gene
rRNA_pseudogene
Mt_rRNA
3prime_overlapping_ncRNA
macro_lncRNA
scaRNA
TR_C_gene
unitary_pseudogene
TR_J_gene
translated_processed_pseudogene
non_coding
IG_V_pseudogene
TR_V_gene
rRNA
TR_V_pseudogene
pseudogene
vaultRNA
TR_J_pseudogene
TR_D_gene
sRNA
ribozyme
IG_pseudogene
IG_J_pseudogene
IG_J_gene
IG_D_gene
IG_C_pseudogene
IG_C_gene
1e−01
1e+00
1e+01
1e+02
1e+03
1e+04
1e+05
1e+06
Reads per Million


Normalized biotype distribution including reads not mapped to genome or transcriptome
Reads per million reads used for alignment
1000000
750000
500000
250000
0
SV00001.S1
SV00002.S2
SV00004.S4
SV00005.S5
SV00006.S6
SV00003.S3
Biotype
lincRNA
protein_coding
processed_pseudogene
retained_intron
misc_RNA
Other
not_mapped_to_genome_or_transcriptome


Raw Gene abundance distribution
(RPM) Reads per Million
1e+05
1e+03
1e+01
SV00001.S1
SV00002.S2
SV00003.S3
SV00004.S4
SV00005.S5
SV00006.S6


Normalized Gene abundance distribution
(RPM) Reads per Million
1e+05
1e+03
1e+01
1e−01
SV00001.S1
SV00002.S2
SV00003.S3
SV00004.S4
SV00005.S5
SV00006.S6


Raw Gene count density distribution
Density
0.5
0.4
0.3
0.2
0.1
0.0
1
100
10000
(RPM) Reads per Million
Samples
SV00001.S1
SV00002.S2
SV00003.S3
SV00004.S4
SV00005.S5
SV00006.S6


Normalized Gene count density distribution
Density
0.5
0.4
0.3
0.2
0.1
0.0
1e+02
1e+04
1e+06
(RPM) Reads per Million
Samples
SV00001.S1
SV00002.S2
SV00003.S3
SV00004.S4
SV00005.S5
SV00006.S6