Report for dataset 5prime_v4_GEX_hPBMCs_20k_Rep1-Cell7 (all samples)

Dataset 7aefee39-0ce4-4640-b0fe-f554dc69a979

Summary

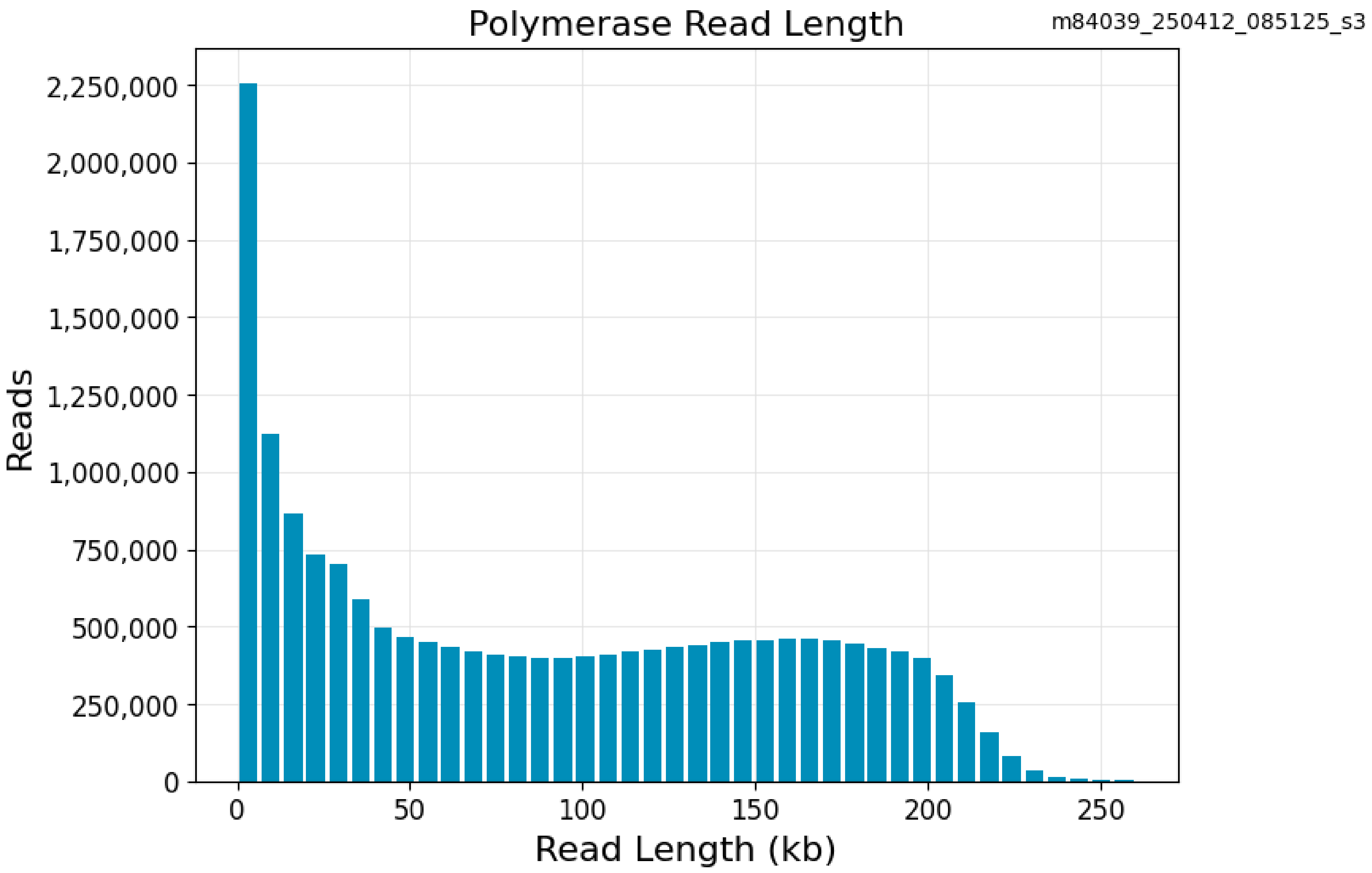
Name	5prime_v4_GEX_hPBMCs_20k_Rep1-Cell7 (all samples)
Created At	2025-04-14 06:04:57.453
Number of Records	9,358,946
Total Length	142,529,153,666
Movie Name	m84039_250412_085125_s3
ICS Version	13.3.0.253824
Well Sample	5prime_v4_GEX_hPBMCs_20k_Rep1
Biological Sample	5prime_v4_GEX_hPBMCs_20k_Rep1
Barcode Name	bcM0004--bcM0004

Raw Data Report

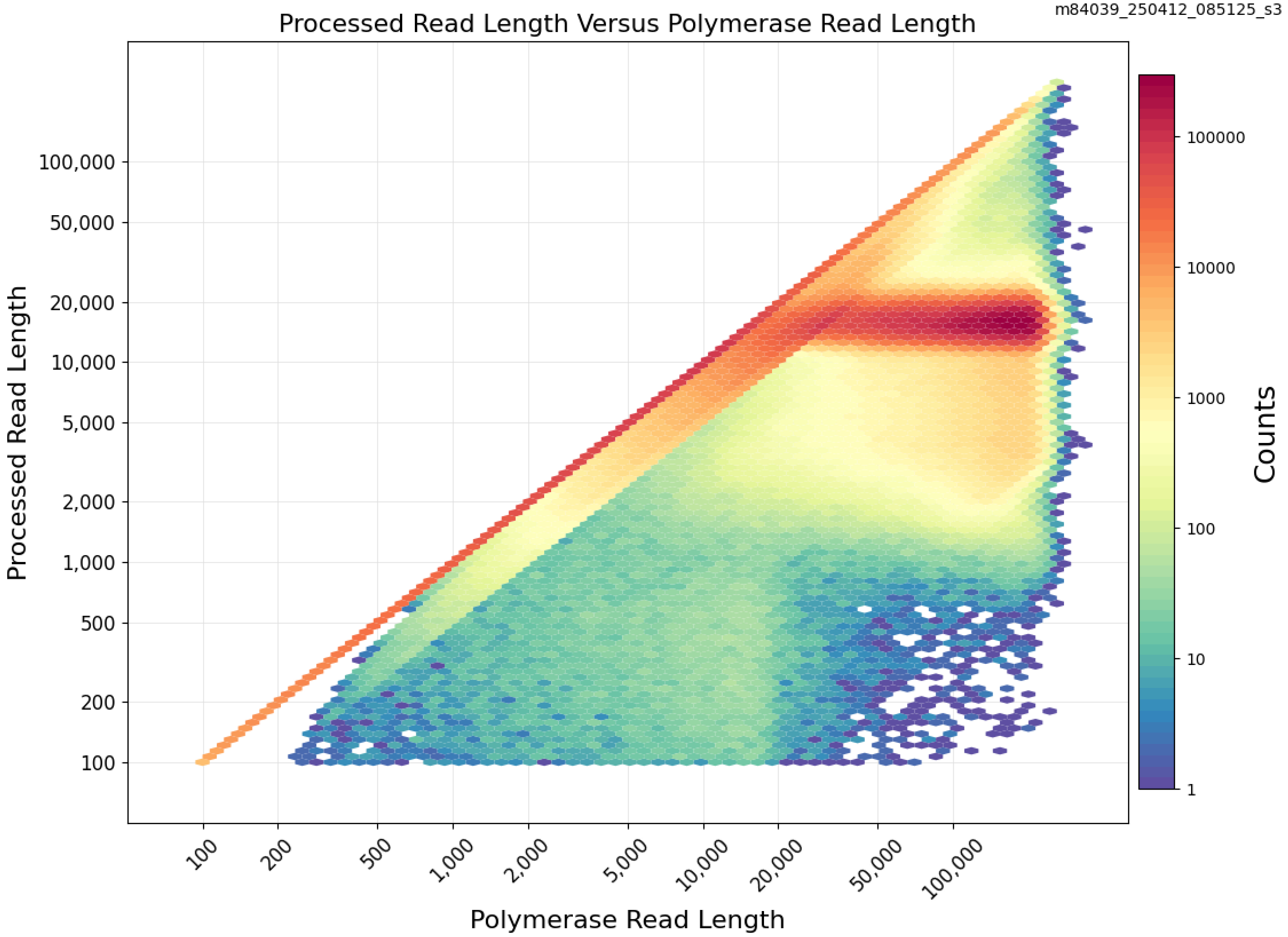
Summary

Polymerase Read Bases	1,525,952,325,356
Polymerase Reads	17,984,011
Polymerase Read Length (mean)	84.85 kb
Polymerase read length (N50)	151.25 kb
Polymerase read length longest subread length (mean)	15.78 kb
Polymerase read length longest subread length (N50)	17.25 kb
Unique Molecular Yield	271,435,792,384
Local Base Rate	2.44

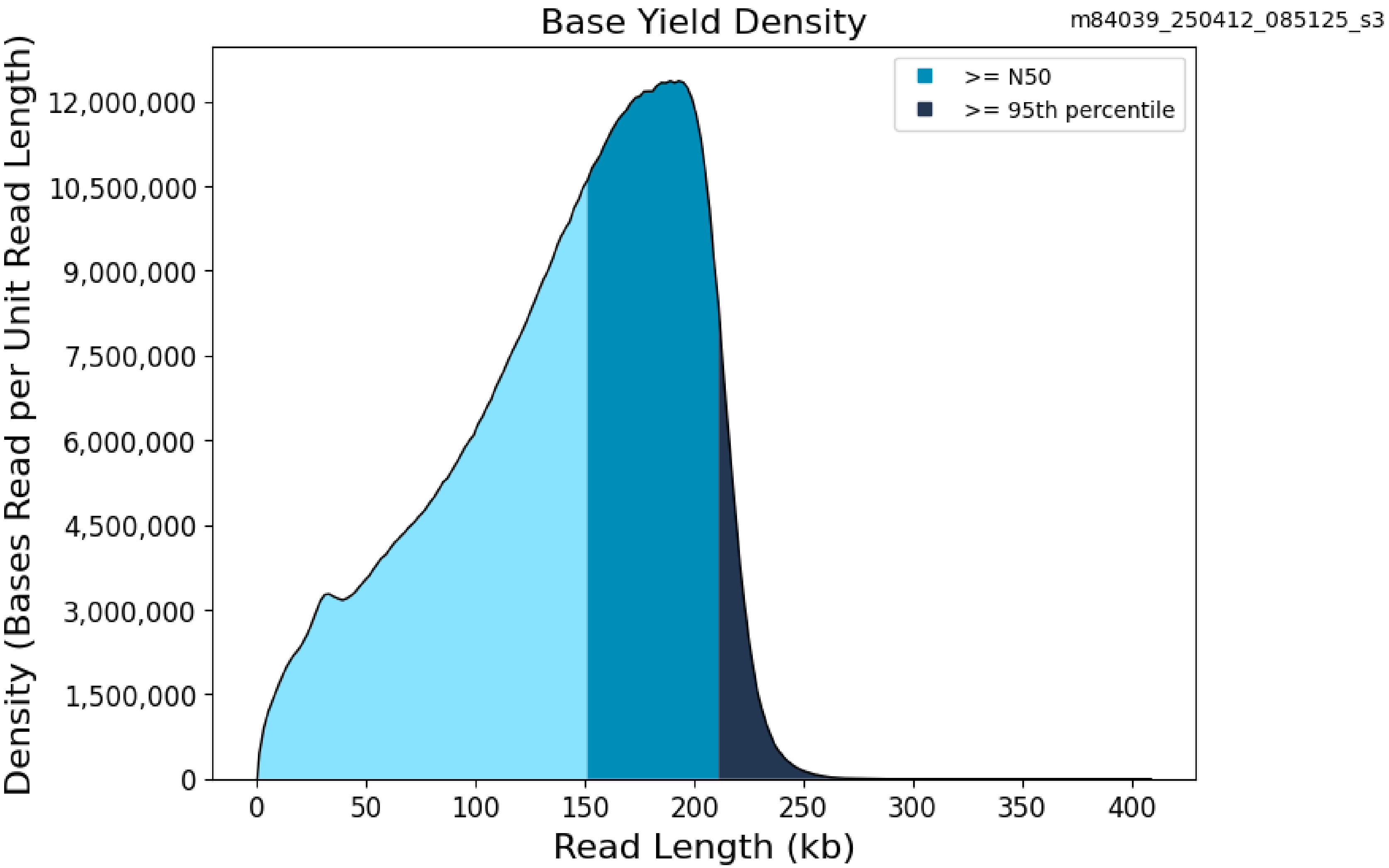
Polymerase Read Length



Longest Subread Length Versus Polymerase Read Length



Base Yield Density



Barcodes

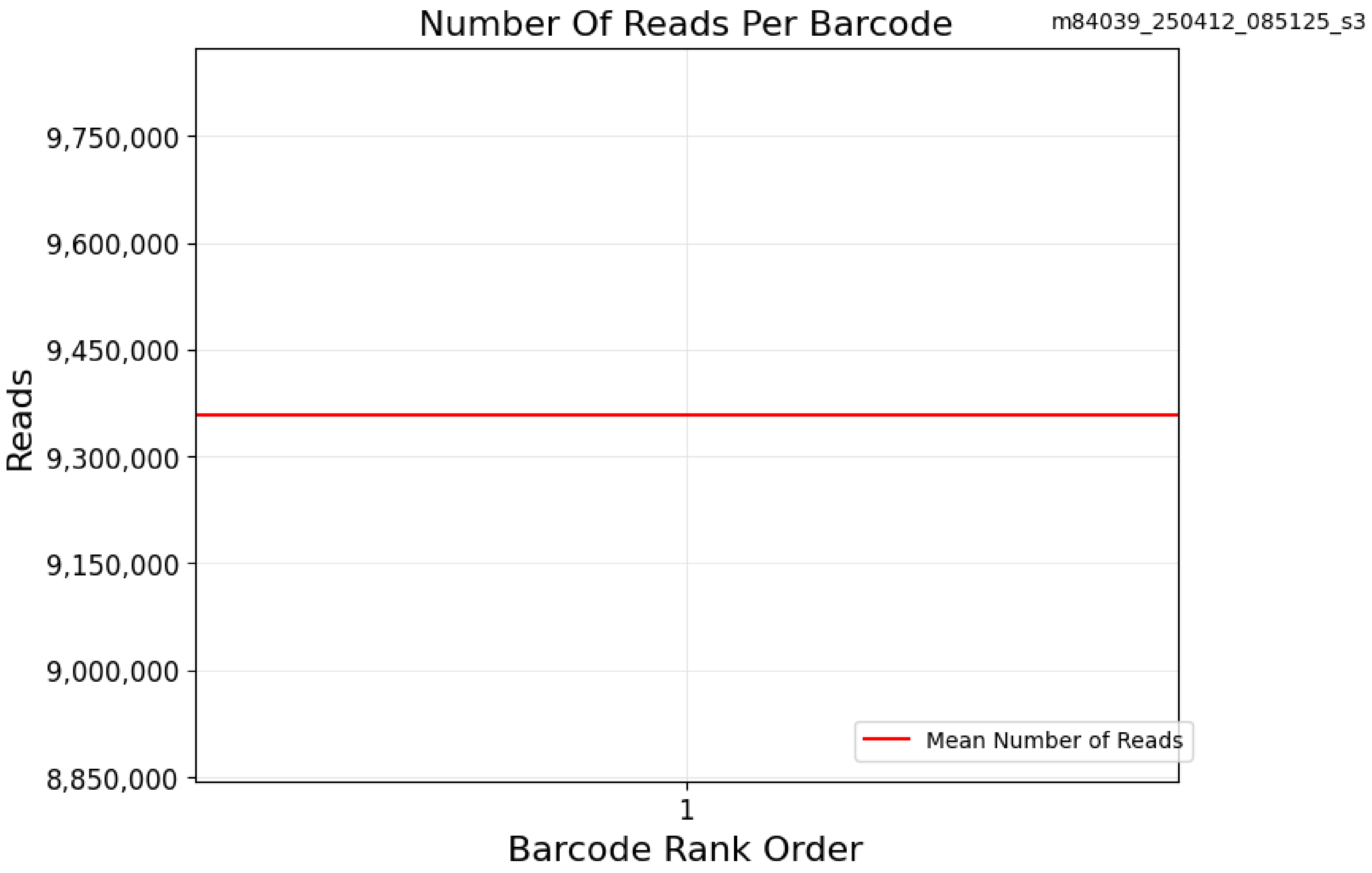
Summary

Unique Barcodes	1
Barcoded HiFi Reads	9,358,946
Unbarcoded HiFi Reads	20,508
Barcoded HiFi Reads (%)	99.78 %
Barcoded HiFi yield (Gb)	142.53 Gb
Unbarcoded HiFi yield (Gb)	0.22 Gb
Barcoded HiFi Yield (%)	99.85 %
Mean HiFi Reads per Barcode	9,358,946
Max. HiFi Reads per Barcode	9,358,946
Min. HiFi Reads per Barcode	9,358,946
Barcoded HiFi read length (mean, kb)	15.23 kb
Unbarcoded HiFi read length (mean, kb)	10.76 kb

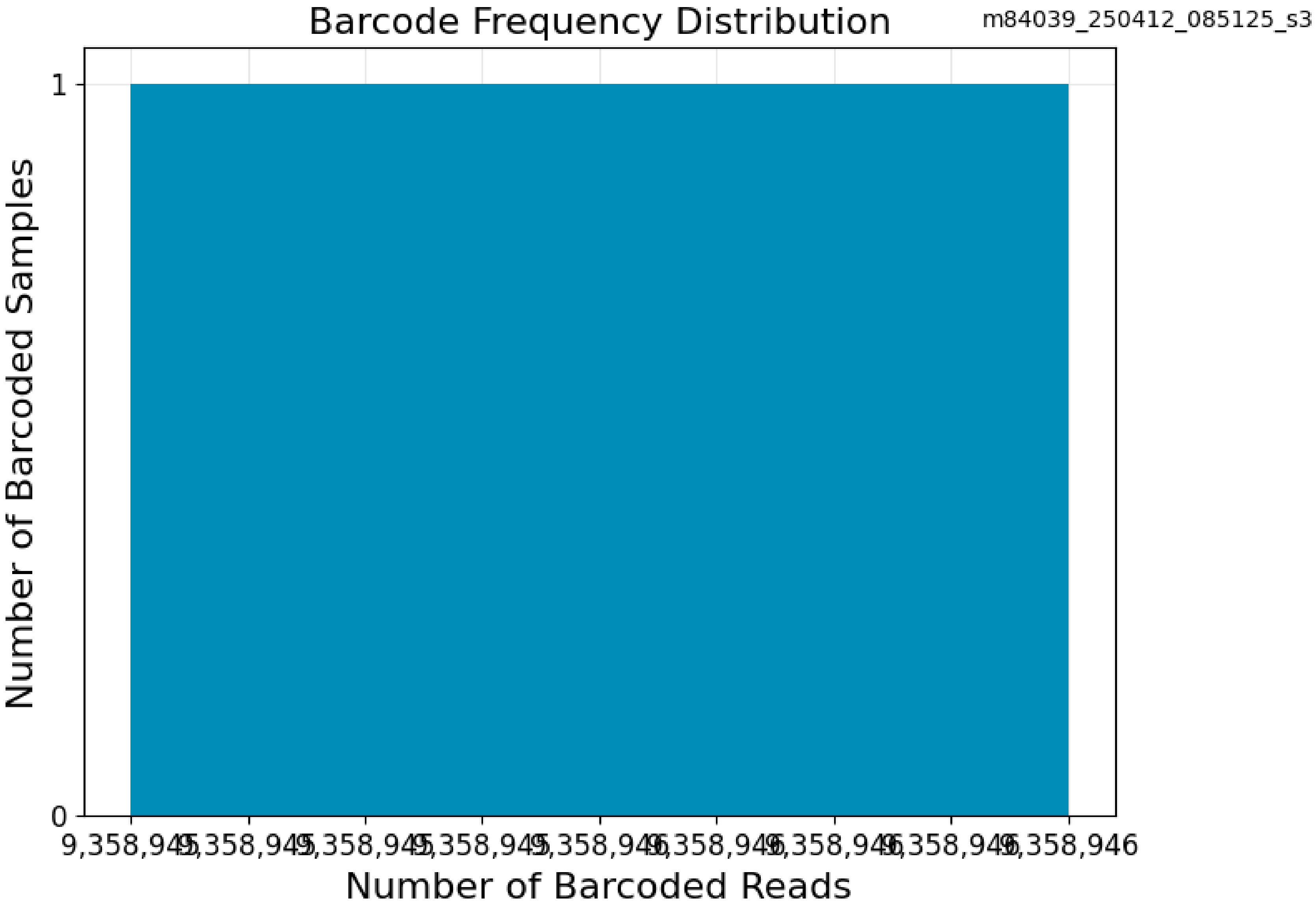
Barcode Data

Sample Name	Barcode	Barcode Quality	HiFi Reads	HiFi Read Length (mean, bp)	HiFi Read Quality (median, QV)	HiFi Yield (bp)	Polymerase Read Length (mean, bp)
5prime_v4_GEX_hPBMCs_20k_Rep1	bcM0004--bcM0004	97.3	9358946	15229	Q29	142529153666	141245
No Name	Not Barcoded	0.0	20508	10756	Q27	220601474	112259

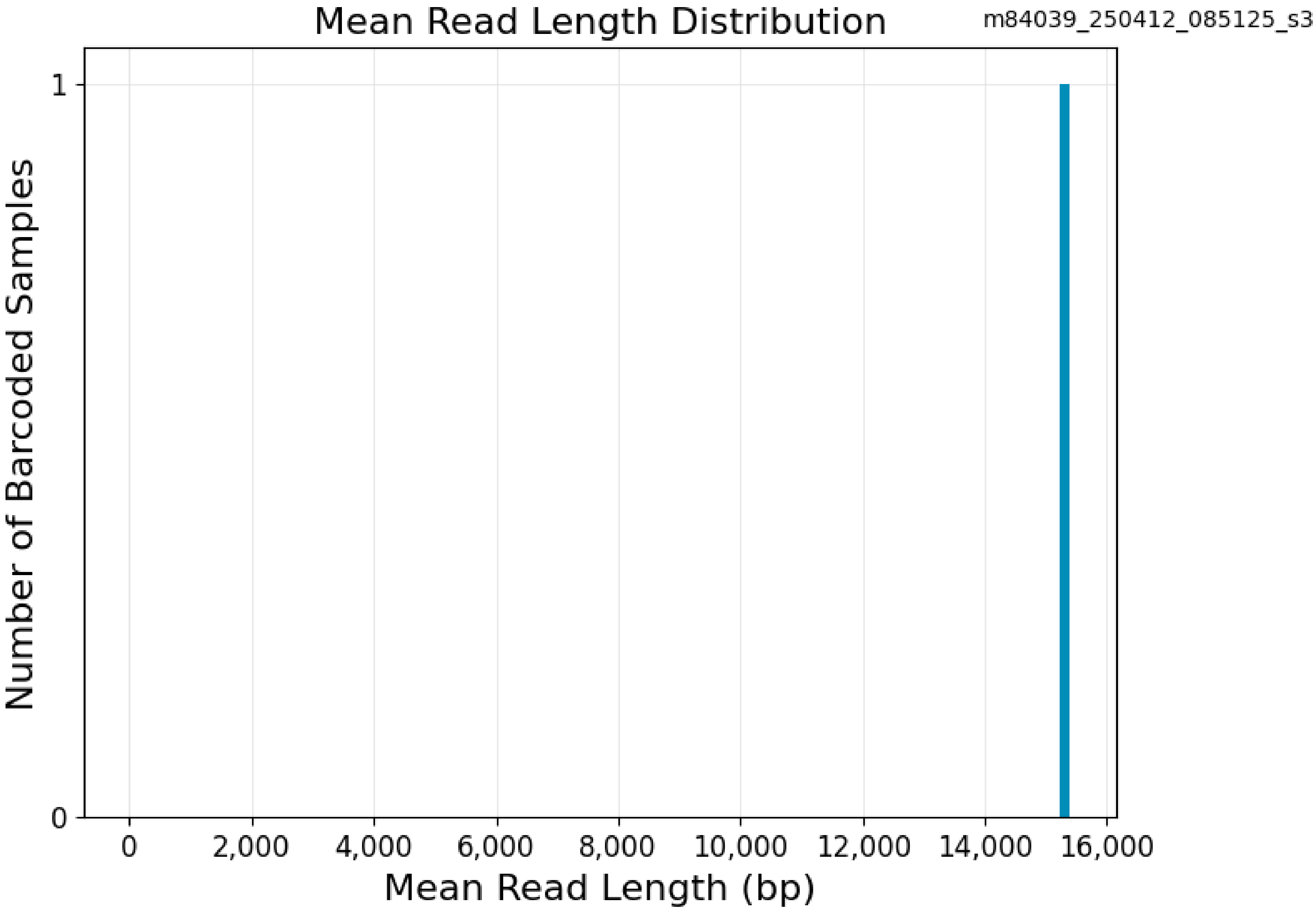
Number Of Reads Per Barcode



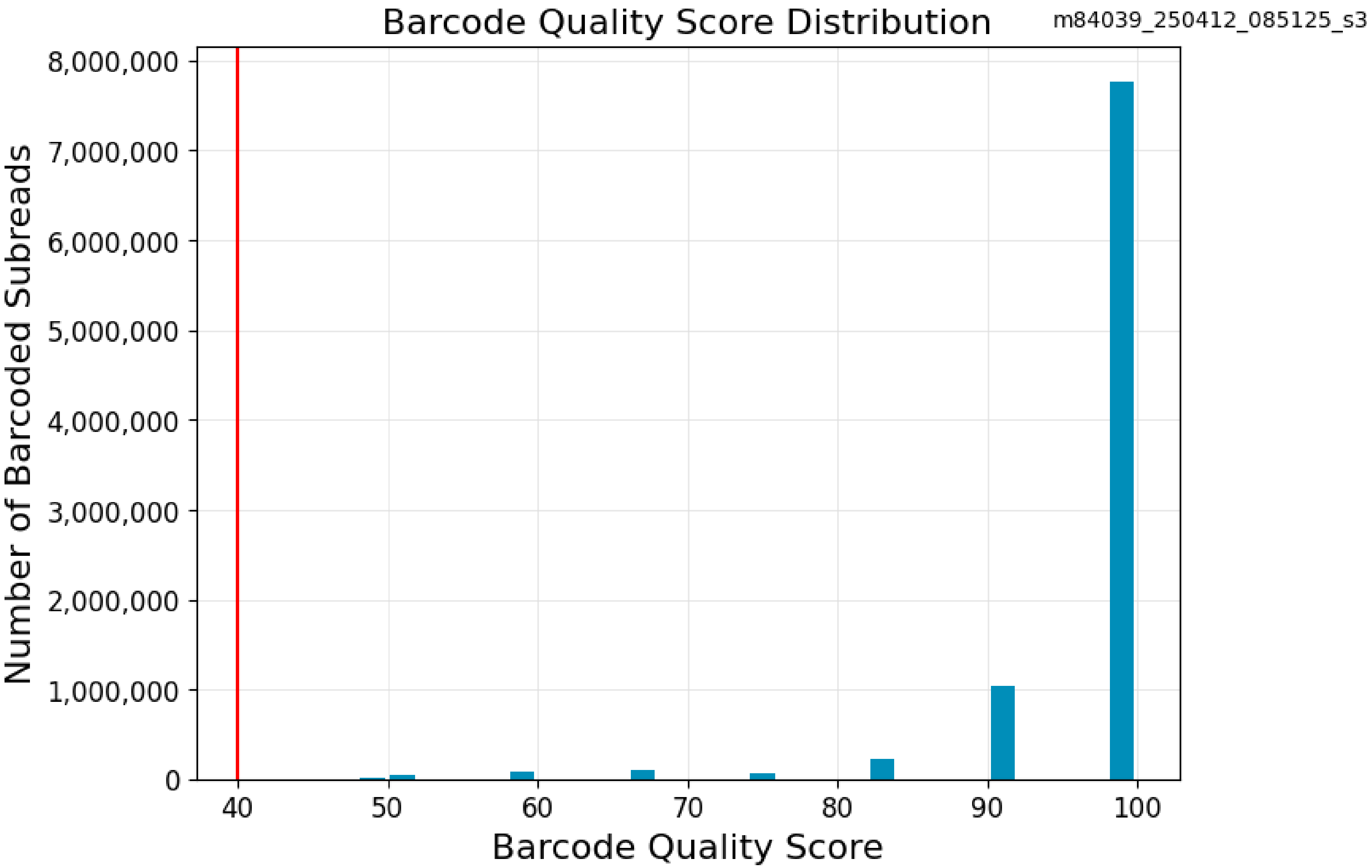
Barcode Frequency Distribution



Mean Read Length Distribution



Barcode Quality Score Distribution

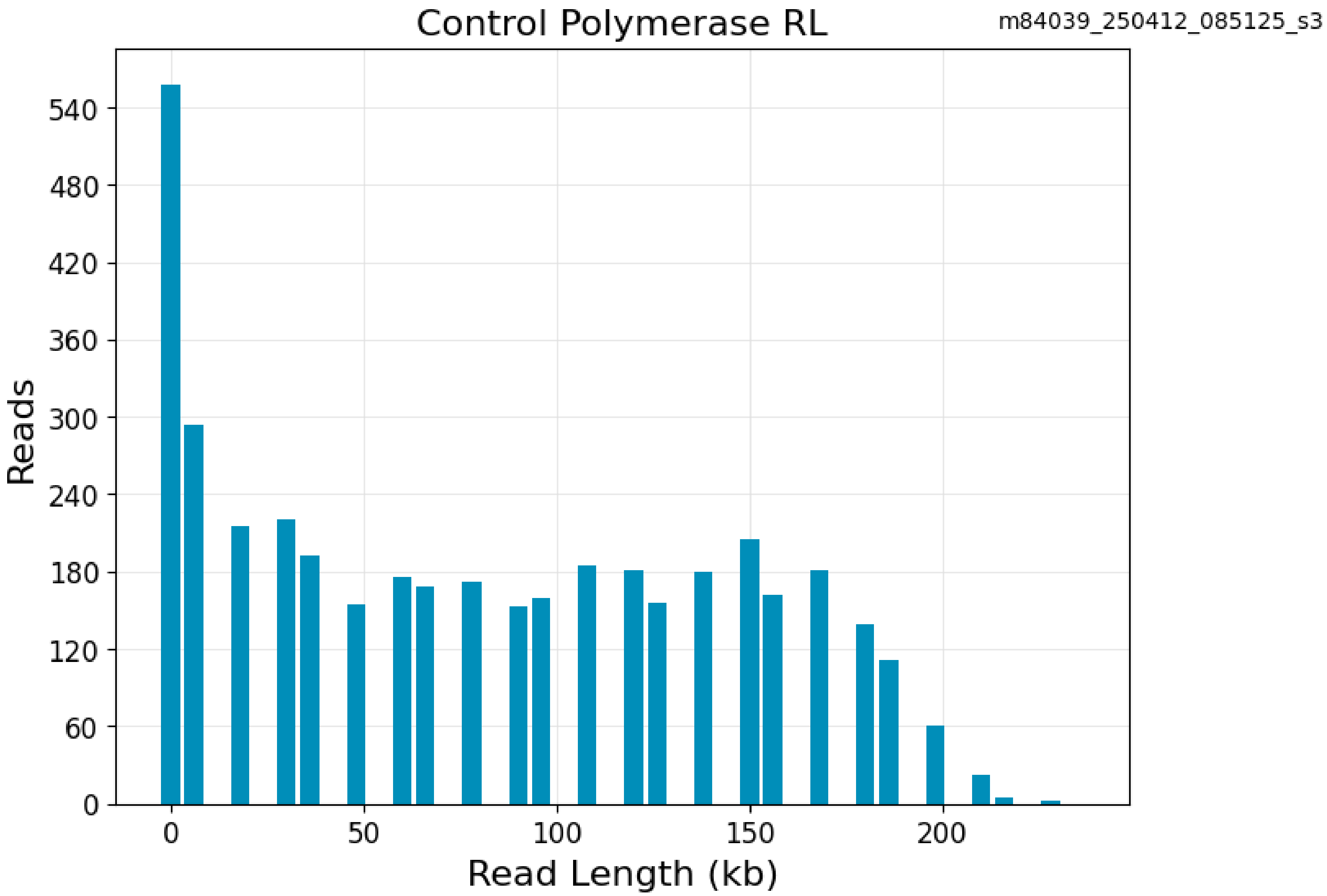


Control Report

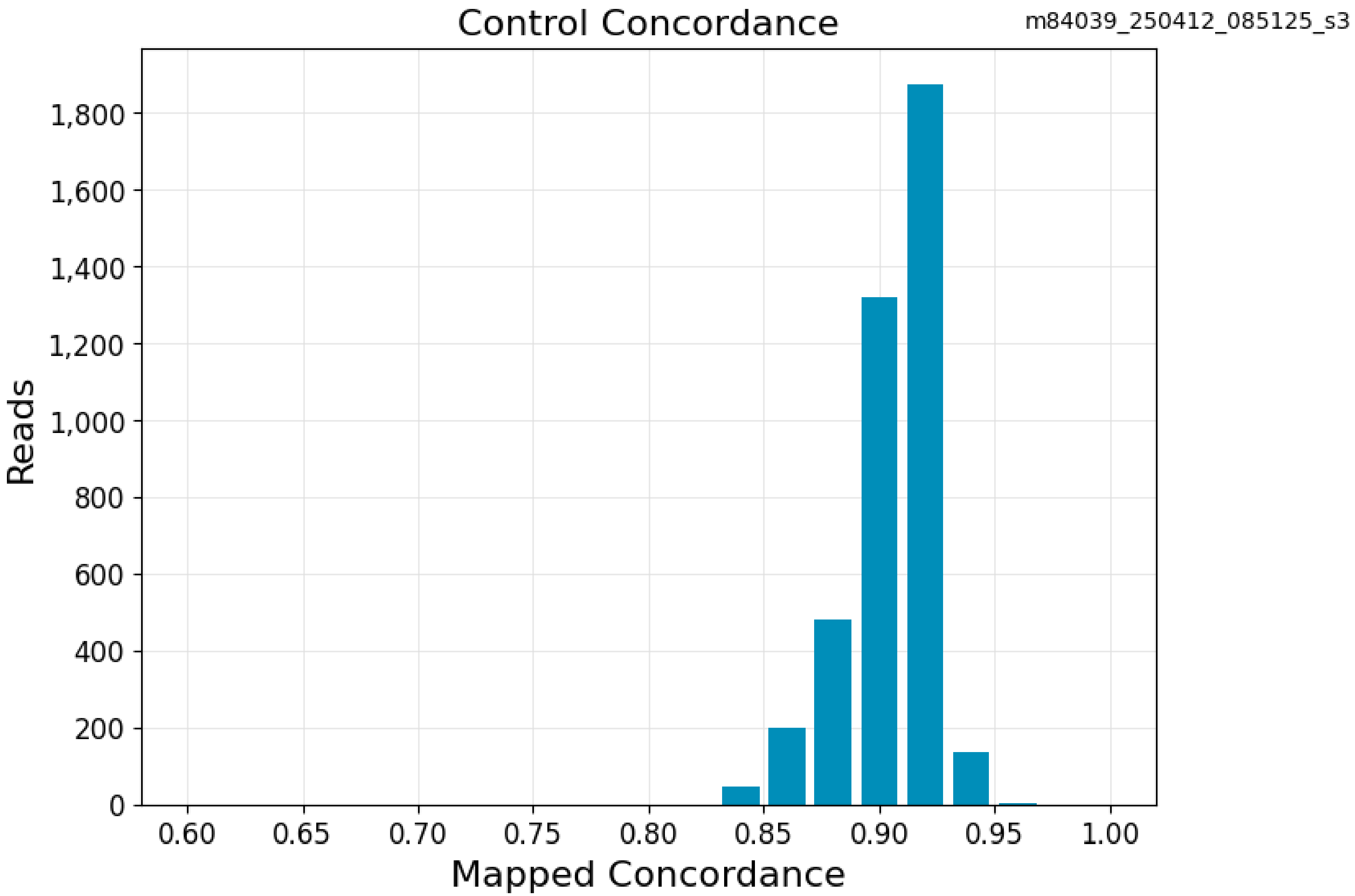
Summary

Number of Control Reads	4,053
Control Read Length Mean	82,979
Control Read Concordance Mean	0.92
Control Read Concordance Mode	0.93

Control Polymerase RL



Control Concordance



CCS Analysis Report

Summary

HiFi reads	9.4 M
HiFi reads yield	142.75 Gb
HiFi reads length (mean)	15.22 kb
HiFi reads length (median, bp)	15,478
HiFi Read Length N50 (bp)	15,825
HiFi Read Quality (median)	Q29
HiFi Read Quality (median)	29
Base Quality Q30 (%)	95.14%
HiFi Number of Passes (mean)	9
Missing adapters (%)	5.18%

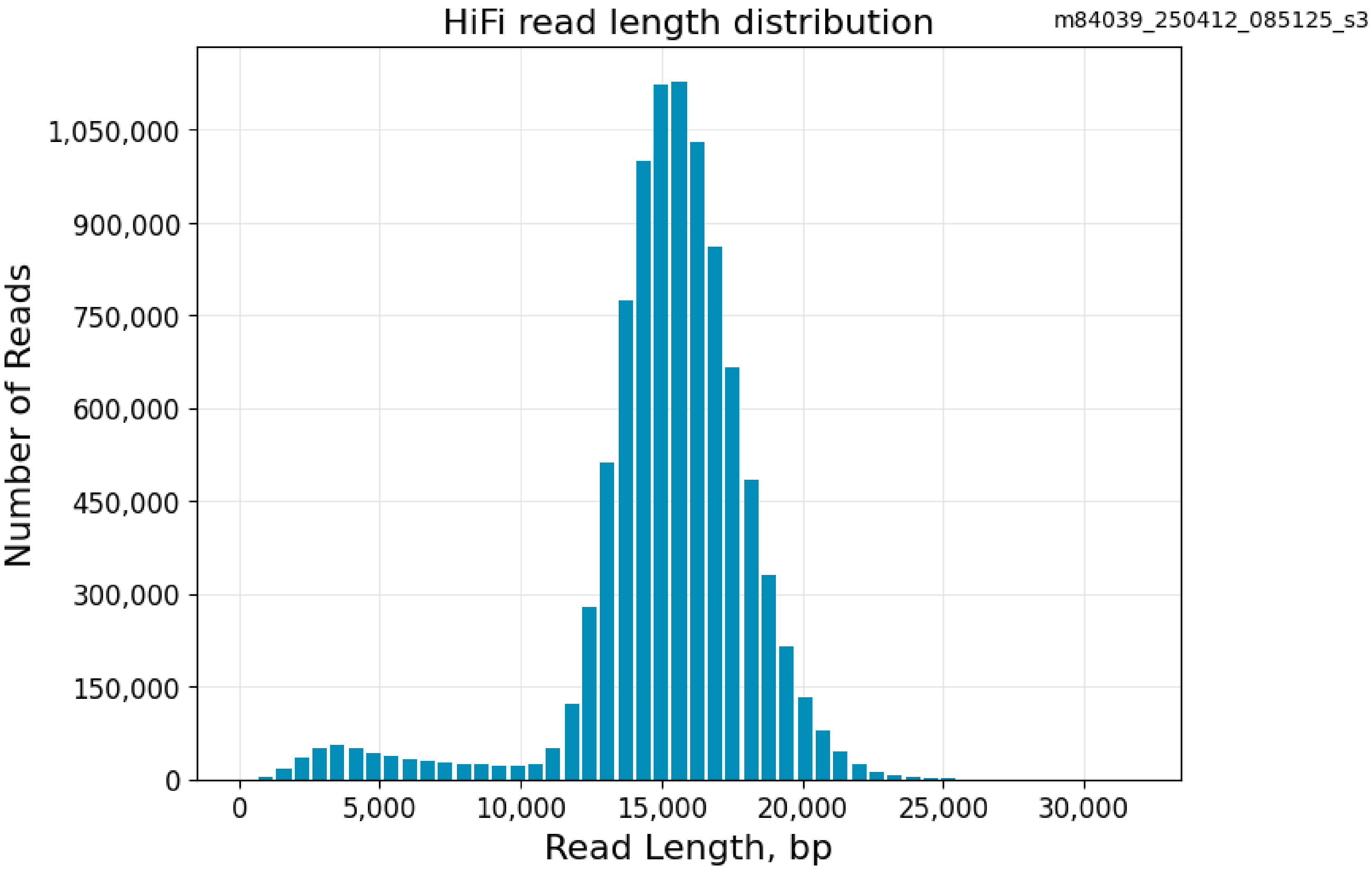
HiFi Read Length Summary

Read Length (kb)	Reads	Reads (%)	Yield (Gb)	Yield (%)
0	9379454	100	142.75 Gb	100
5,000	9127803	97	141.91 Gb	99
10,000	8908548	95	140.32 Gb	98
15,000	5545634	59	94.03 Gb	66
20,000	248722	3	5.24 Gb	4
25,000	2261	0	0.06 Gb	0
30,000	315	0	0.01 Gb	0
35,000	28	0	0.00 Gb	0
40,000	1	0	0.00 Gb	0

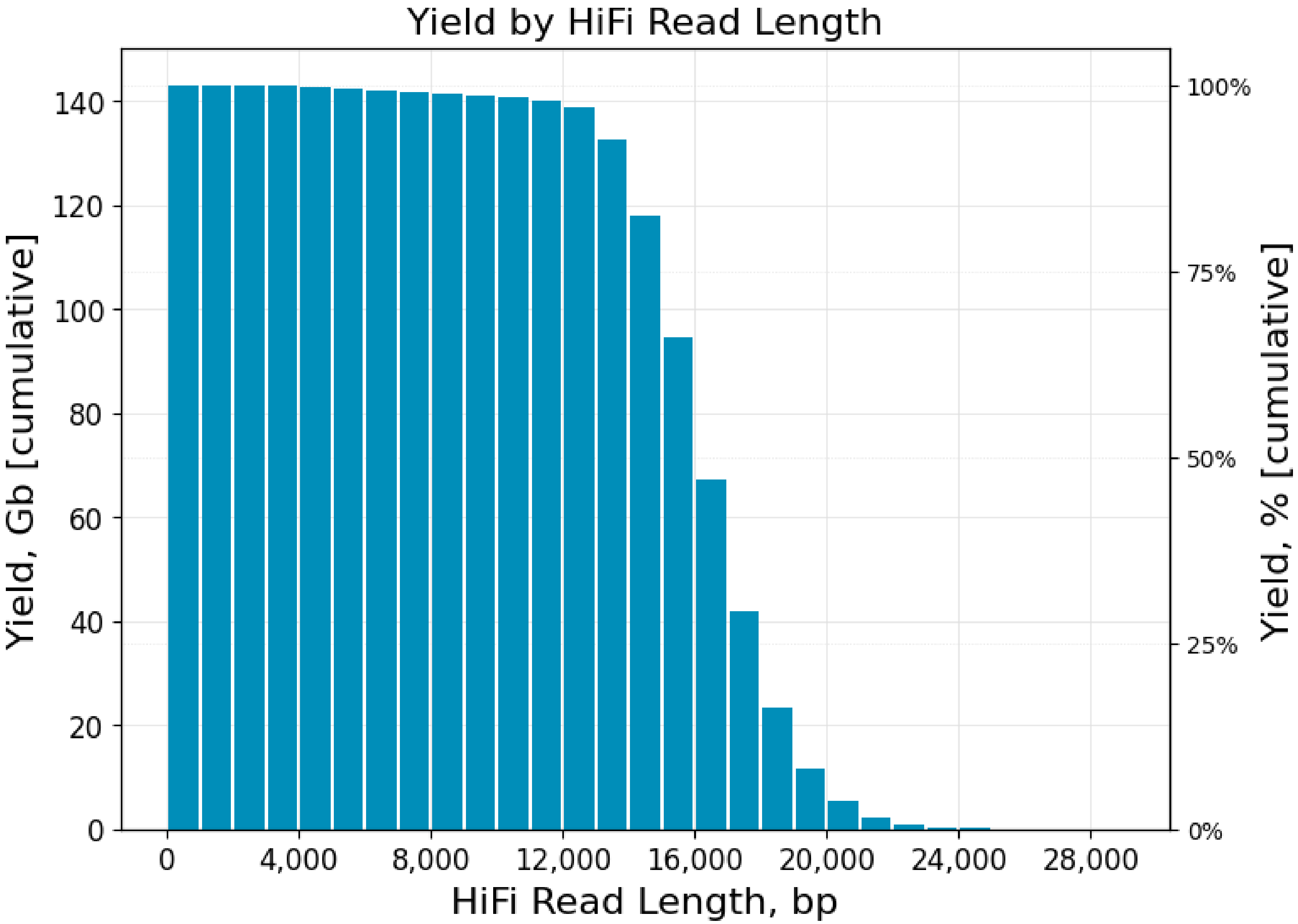
HiFi Read Quality Summary

Read Quality (Phred)	Reads	Reads (%)	Yield (Gb)	Yield (%)
Q20	9379454	100	142.75 Gb	100
Q30	3576949	38	52.41 Gb	37
Q40	147766	2	0.73 Gb	1
Q50	41070	0	0.12 Gb	0

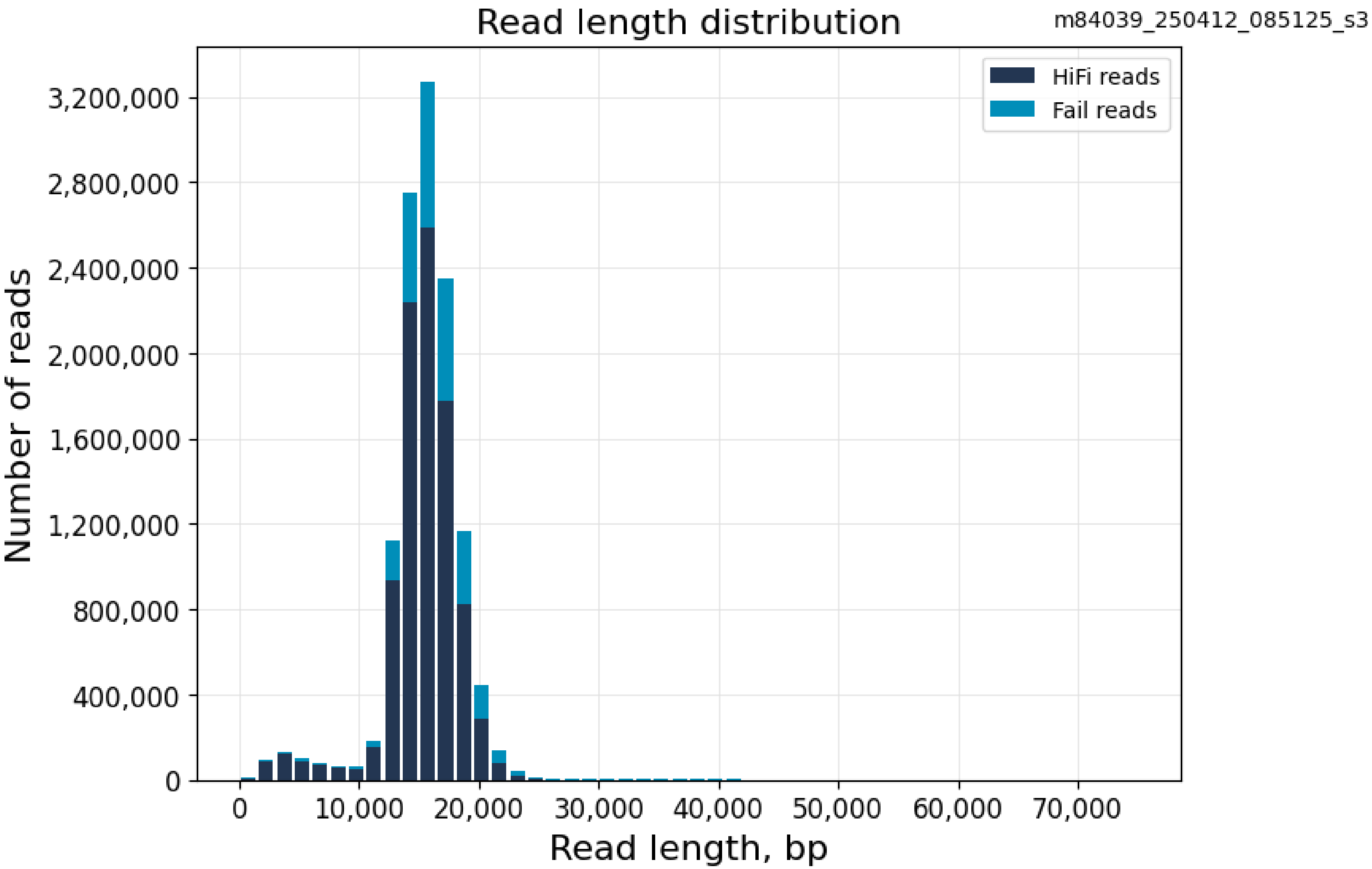
HiFi read length distribution



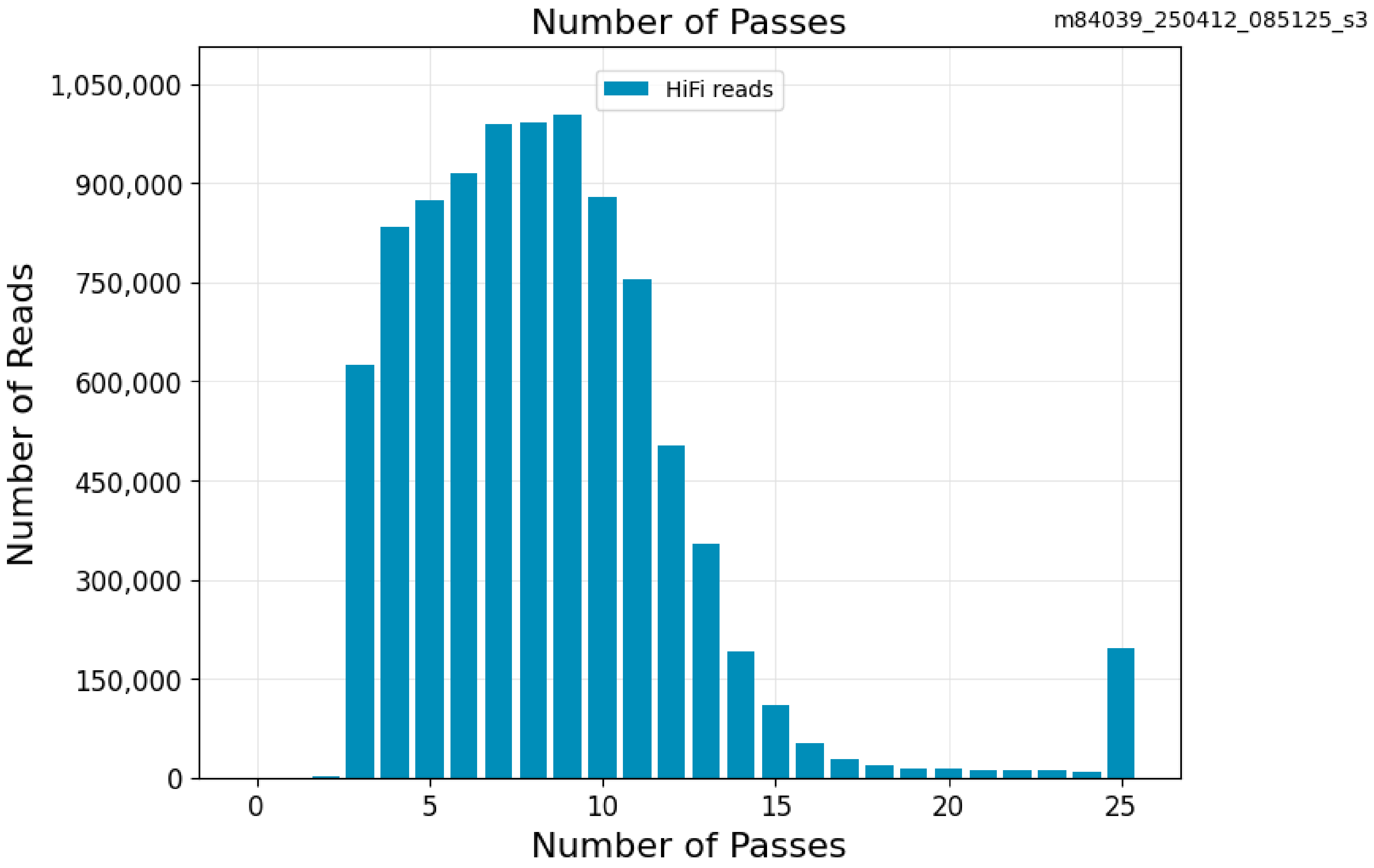
Yield by HiFi Read Length



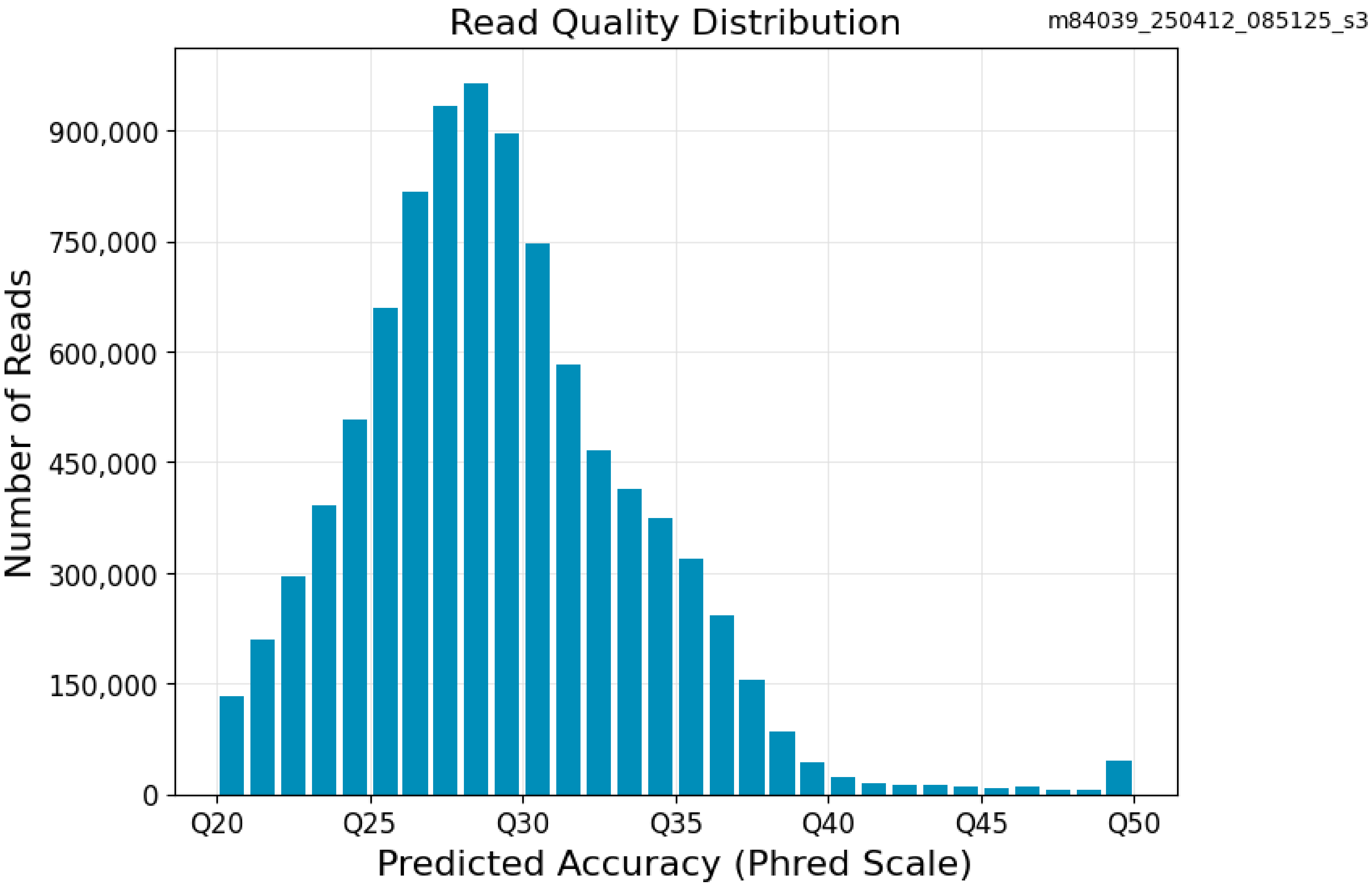
Read length distribution



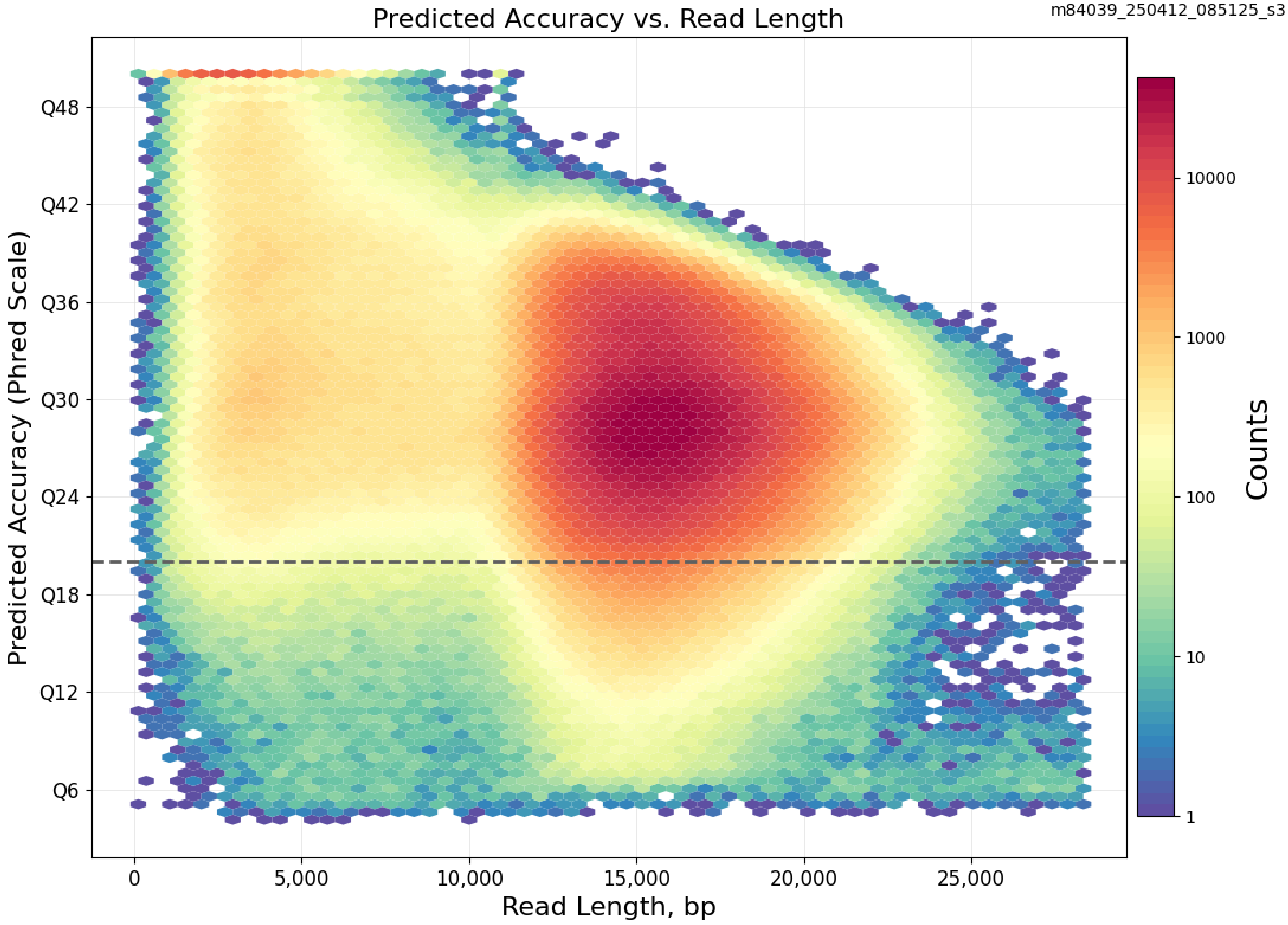
Number of Passes



Read Quality Distribution



Predicted Accuracy vs. Read Length



Loading Report

Summary

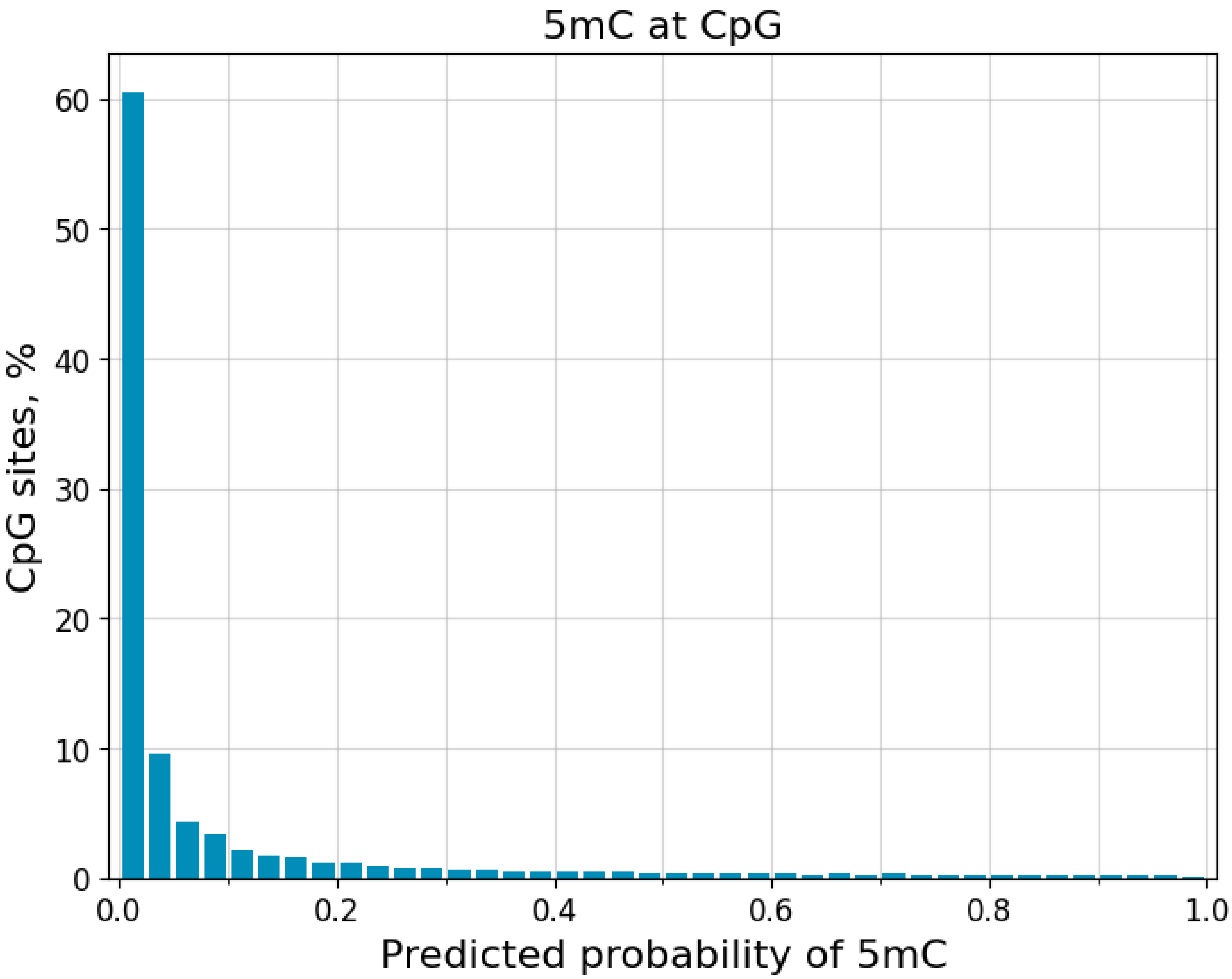
Productive ZMWs	25,165,824
Productivity 0	7,094,537
Productivity 1	17,988,064
Productivity 2	83,223

Methylation

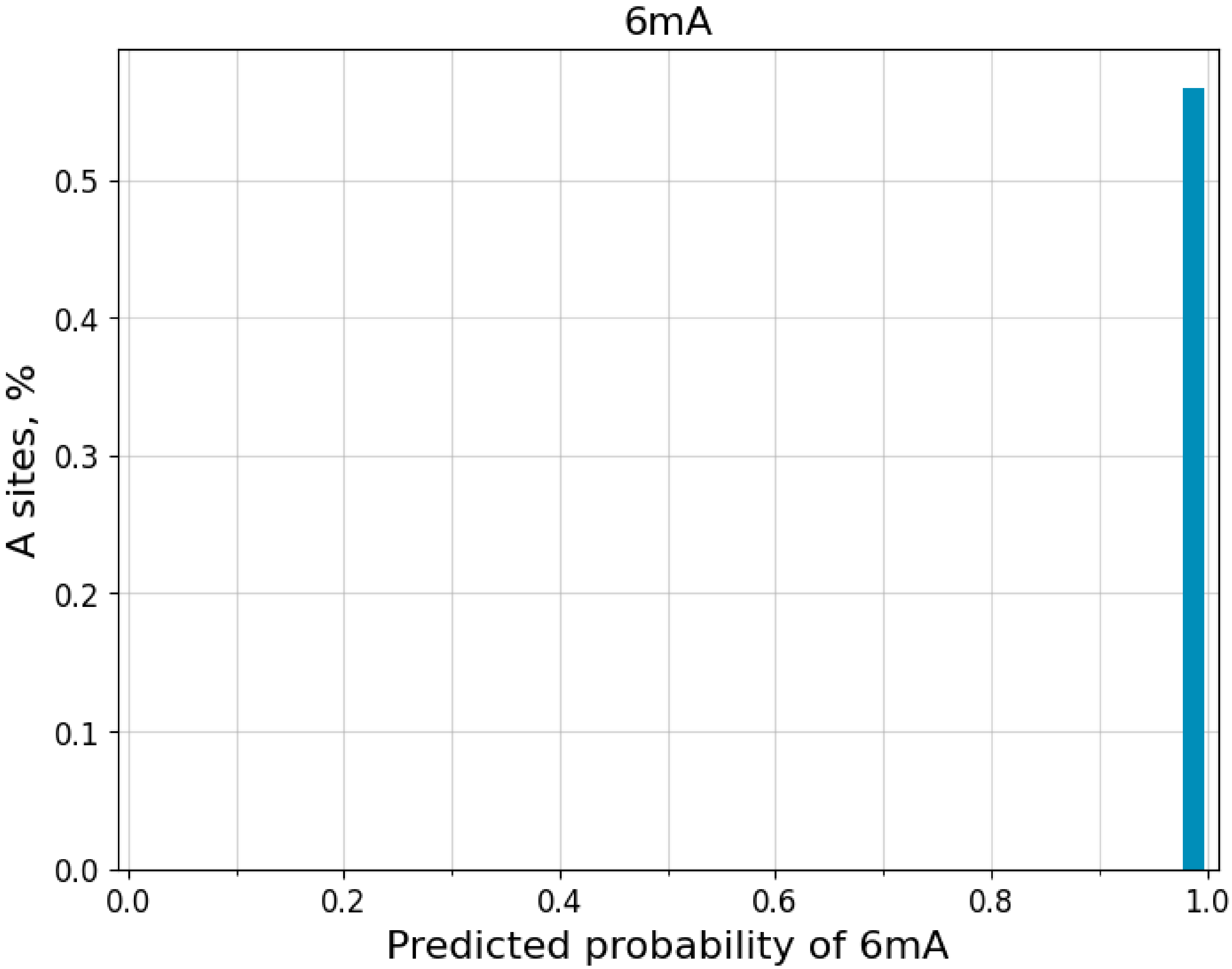
Summary

Modification	Motif	Scored sites	Modified sites (Pr > 0.5)
5mC	CpG	98.4%	5.8%
6mA	A	0.6%	0.6%

5mC methylation at CpG in reads


5mC at CpG
CpG sites, %
0
10
20
30
40
50
60
0.0
0.2
0.4
0.6
0.8
1.0
Predicted probability of 5mC

6mA methylation in reads


6mA
A sites, %
0.0
0.1
0.2
0.3
0.4
0.5
0.0
0.2
0.4
0.6
0.8
1.0
Predicted probability of 6mA

No Sample Setup found

No instrument run found

Parent jobs (1)

Job 118846

Summary

Job Type	import-dataset
Name	import-dataset
Comments	Description for job Import PacBio DataSet
Created At	2025-04-24 15:59:27.198
SMRT Link Version	25.3.0.SNAPSHOT269086

Child jobs (1)

Job 124627

Summary

Job Type	analysis
Pipeline	cromwell.workflows.pb_segment_reads_and_sc_isoseq
Name	KinnexRelease-10x-5pGEX-20k-Revio2cell-largeMem
Comments	Description for job Run Analysis Application
Created At	2025-08-02 13:32:20.076
SMRT Link Version	25.3.0.273777

Other inputs

File Type	Name	Dataset ID	Number of Records
ReferenceSet	Human Genome hg38, with Gencode v39 annotations	ba3866bf-2aba-7c99-0570-0d6709174e4a	195
BarcodeSet	MAS-Seq Adapter v1 (MAS16)	10ce52cd-91c5-4bb9-4cac-2233832ecb12	17
BarcodeSet	10X Chromium single cell 5' cDNA primers	5d9656f9-bbcc-6eca-1a46-56a2d31db49e	2
ConsensusReadSet	5prime_v4_GEX_hPBMCs_20k_Rep1-Cell6 (all samples)	a941bc6d-dc66-473a-8d34-a05a23126997	9,408,444

Output files

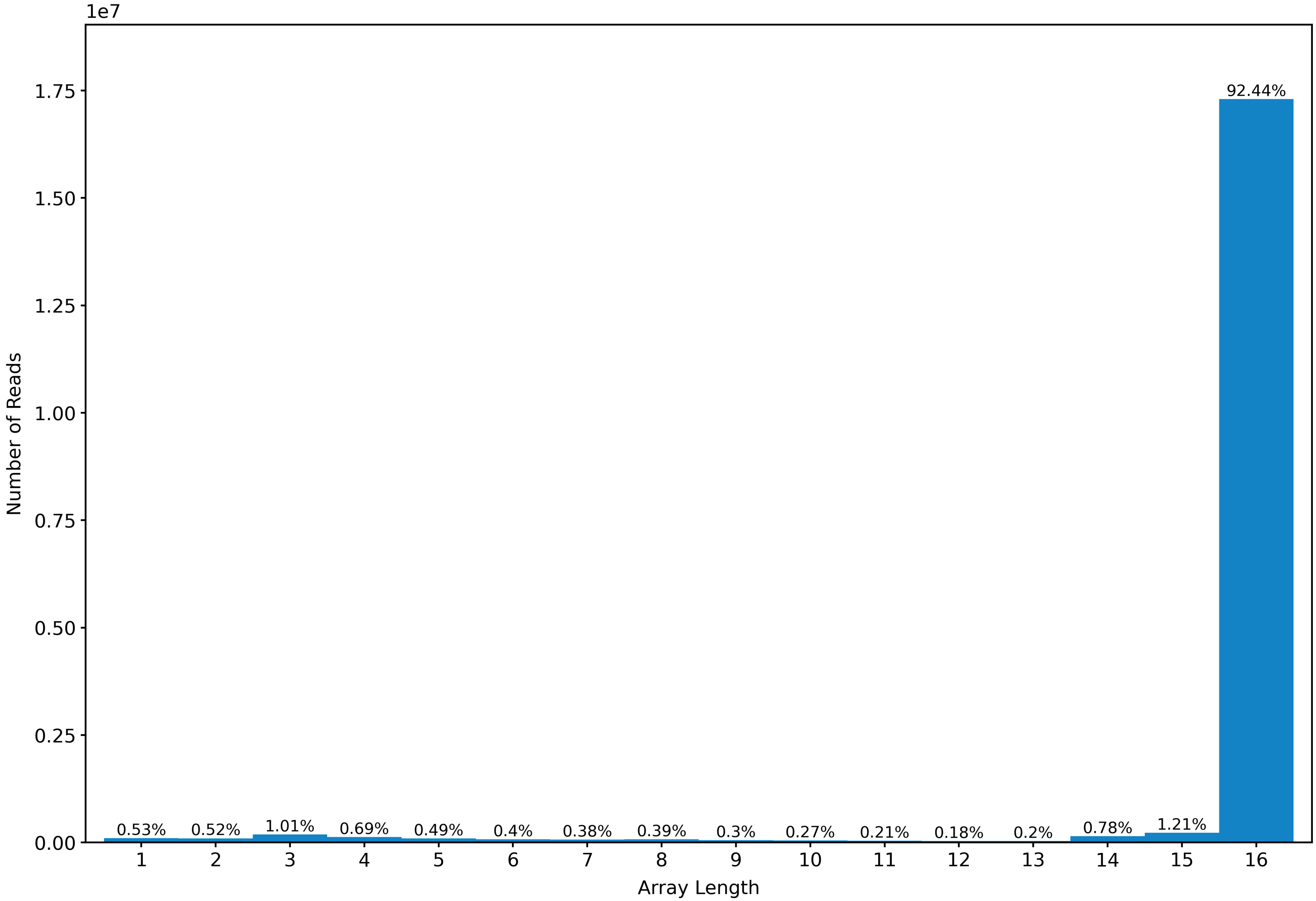
File name	Size	Type
KinnexRelease-10x-5pGEX-20k-Revio2cell-largeMem Segmented Reads	47,582	ConsensusReadSet
Deduplicated reads after cell barcode correction, unmapped, BAM	58,630,025,312	bam
Single-cell isoform and gene matrix, tar-gzipped	5,229,835,901	tgz
Segmented Reads, passing, unaligned	93,966,703,973	bam
Segmented Reads, passing, unaligned	93,478,605,474	bam
Non-passing reads, unaligned	3,598,645,913	bam
Non-passing reads, unaligned	3,359,211,767	bam
Unique mapped transcripts, GFF	2,866,215,348	gff
Unique mapped transcripts, filtered, GFF	1,241,683,689	gff
Deduplicated reads after cell barcode correction, mapped, BAM	29,712,220,750	bam
Deduplicated reads after cell barcode correction, mapped, BAM index	37,932,976	bam_bai
Unique mapped transcripts, junctions TXT	2,207,733,257	txt
Unique mapped transcripts, filtered, junctions TXT	1,173,175,043	txt
Deduplicated transcripts	42,745,109,308	Fasta
Collapsed transcript groups	2,977,338,014	txt
Unique mapped transcripts, classification TXT	1,972,267,425	txt
Unique mapped transcripts, filtered, classification TXT	704,402,006	txt
bcstats_report.tsv.gz	18,086,654	tsv
Analysis Log	101,851	log
3M_5pgex_jan_2023.txt.gz	6,945,252	txt
3M_february_2018_REVERSE_COMPLEMENTED.txt.gz	17,812,866	txt
Analysis Log	4,083	log
SMRT Link Log	20,183	log

Read Segmentation

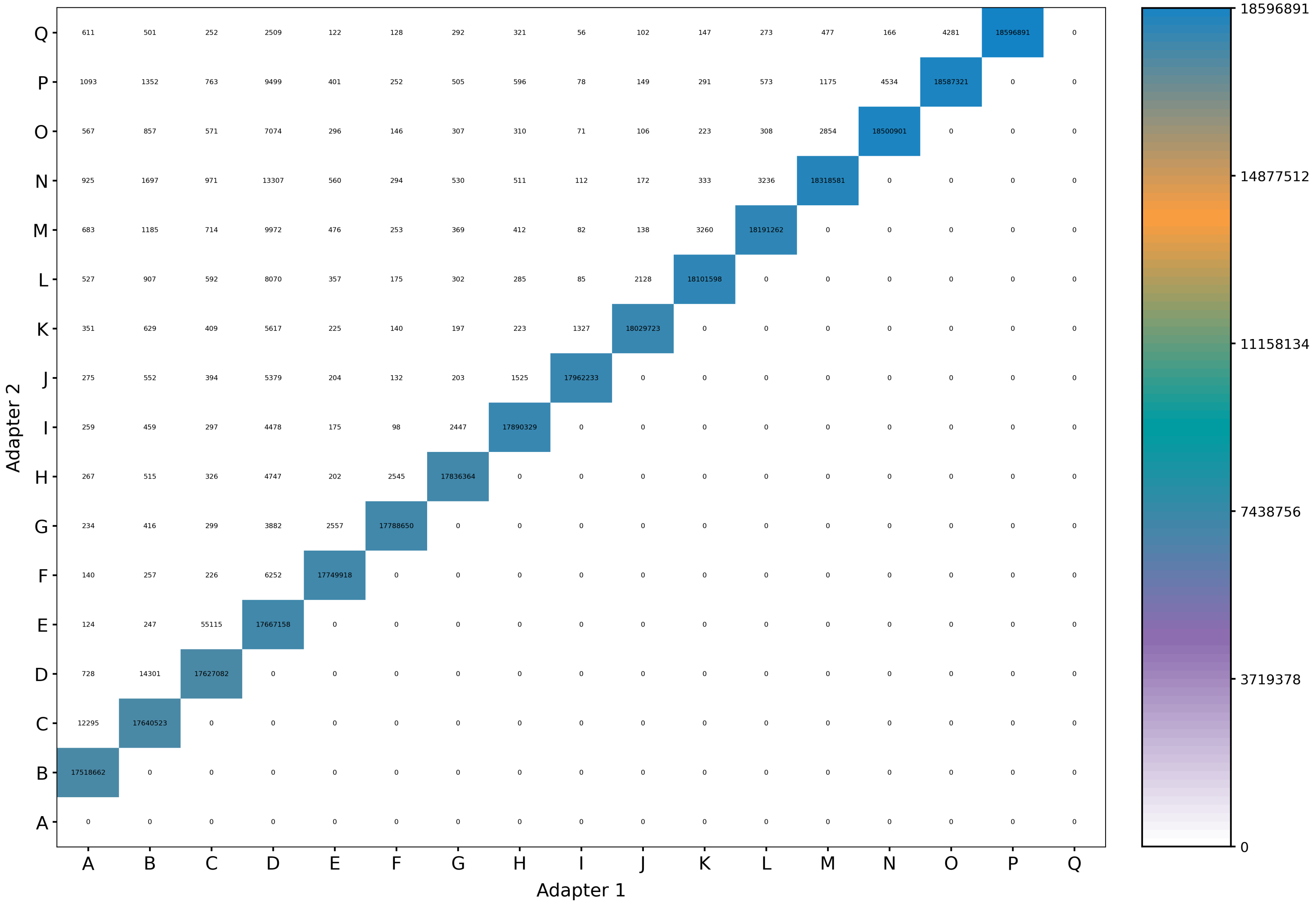
Summary

Reads	18,767,390
Segmented reads (S-reads)	288,007,196
Mean length of S-reads	970.0055384102278
Percent of reads with full arrays	92.20 %
Mean array size (concatenation factor)	15.35

Histogram of concatemers

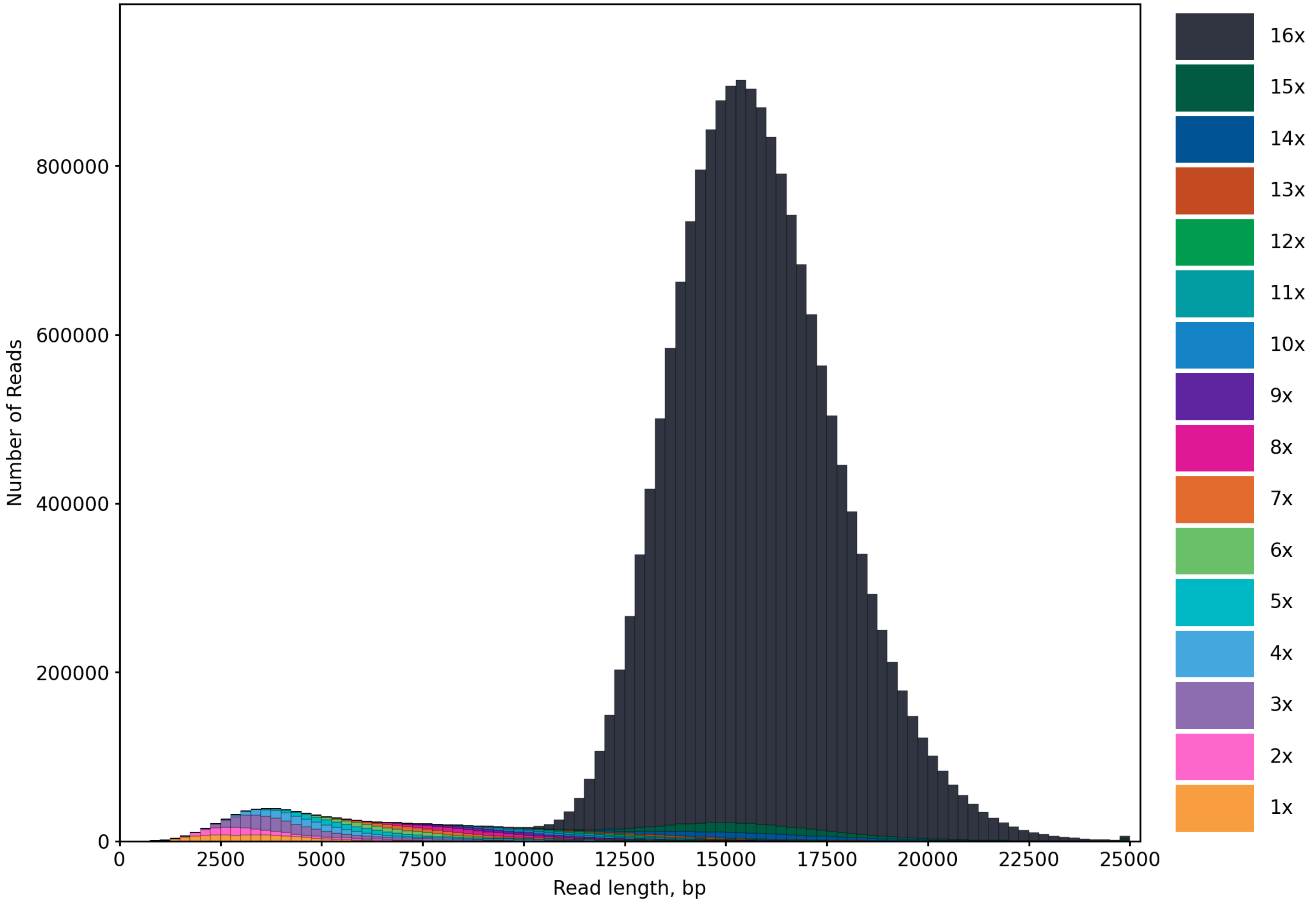


Heatmap of concatemer ligations

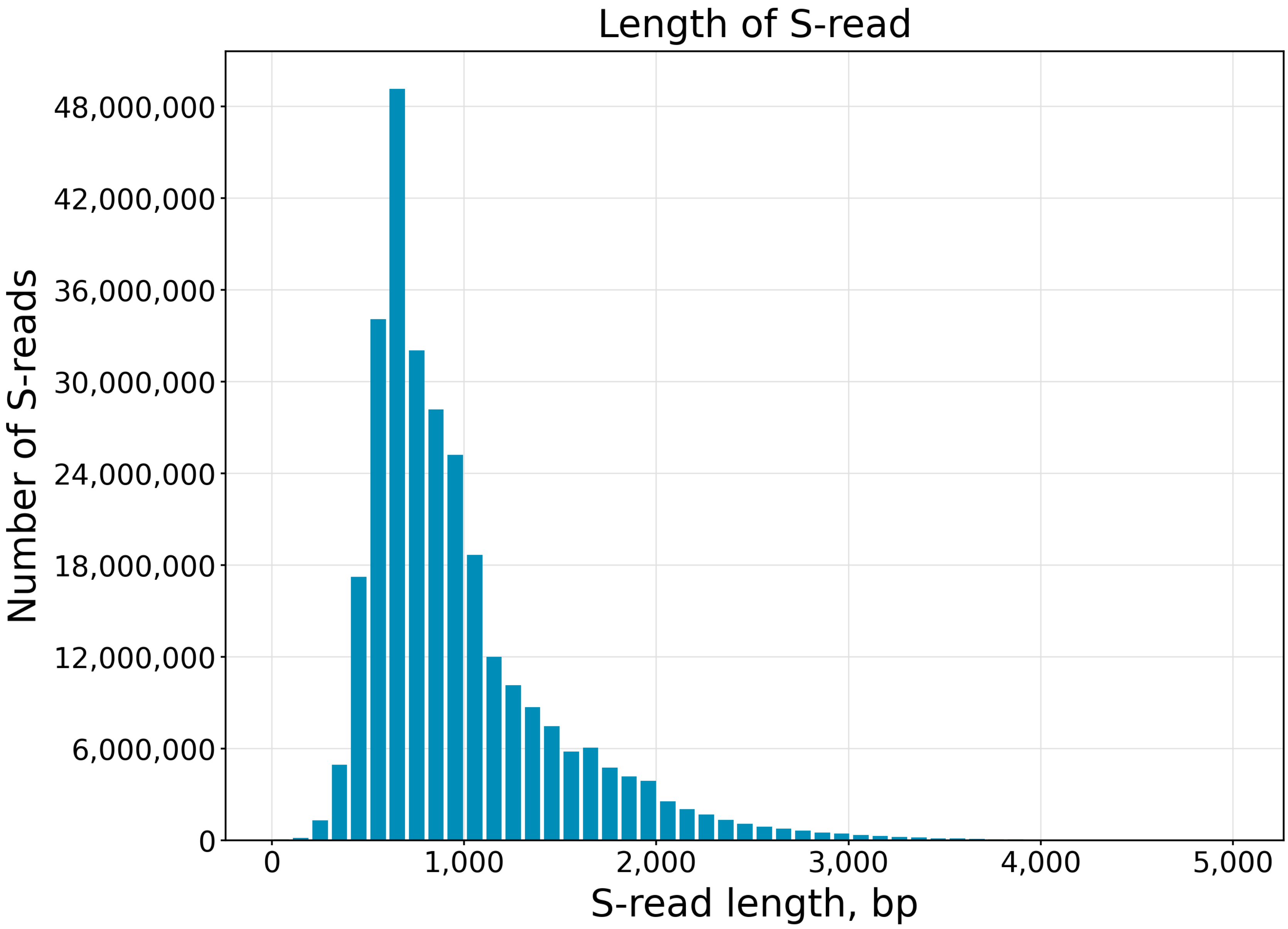


	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
Q	611	501	252	2509	122	128	292	321	56	102	147	273	477	166	4281	18596891	0
P	1093	1352	763	9499	401	252	505	596	78	149	291	573	1175	4534	18587321	0	0
O	567	857	571	7074	296	146	307	310	71	106	223	308	2854	18500901	0	0	0
N	925	1697	971	13307	560	294	530	511	112	172	333	3236	18318581	0	0	0	0
M	683	1185	714	9972	476	253	369	412	82	138	3260	18191262	0	0	0	0	0
L	527	907	592	8070	357	175	302	285	85	2128	18101598	0	0	0	0	0	0
K	351	629	409	5617	225	140	197	223	1327	18029723	0	0	0	0	0	0	0
J	275	552	394	5379	204	132	203	1525	17962233	0	0	0	0	0	0	0	0
I	259	459	297	4478	175	98	2447	17890329	0	0	0	0	0	0	0	0	0
H	267	515	326	4747	202	2545	17836364	0	0	0	0	0	0	0	0	0	0
G	234	416	299	3882	2557	17788650	0	0	0	0	0	0	0	0	0	0	0
F	140	257	226	6252	17749918	0	0	0	0	0	0	0	0	0	0	0	0
E	124	247	55115	17667158	0	0	0	0	0	0	0	0	0	0	0	0	0
D	728	14301	17627082	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C	12295	17640523	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
B	17518662	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Length of Reads



Length of S-reads

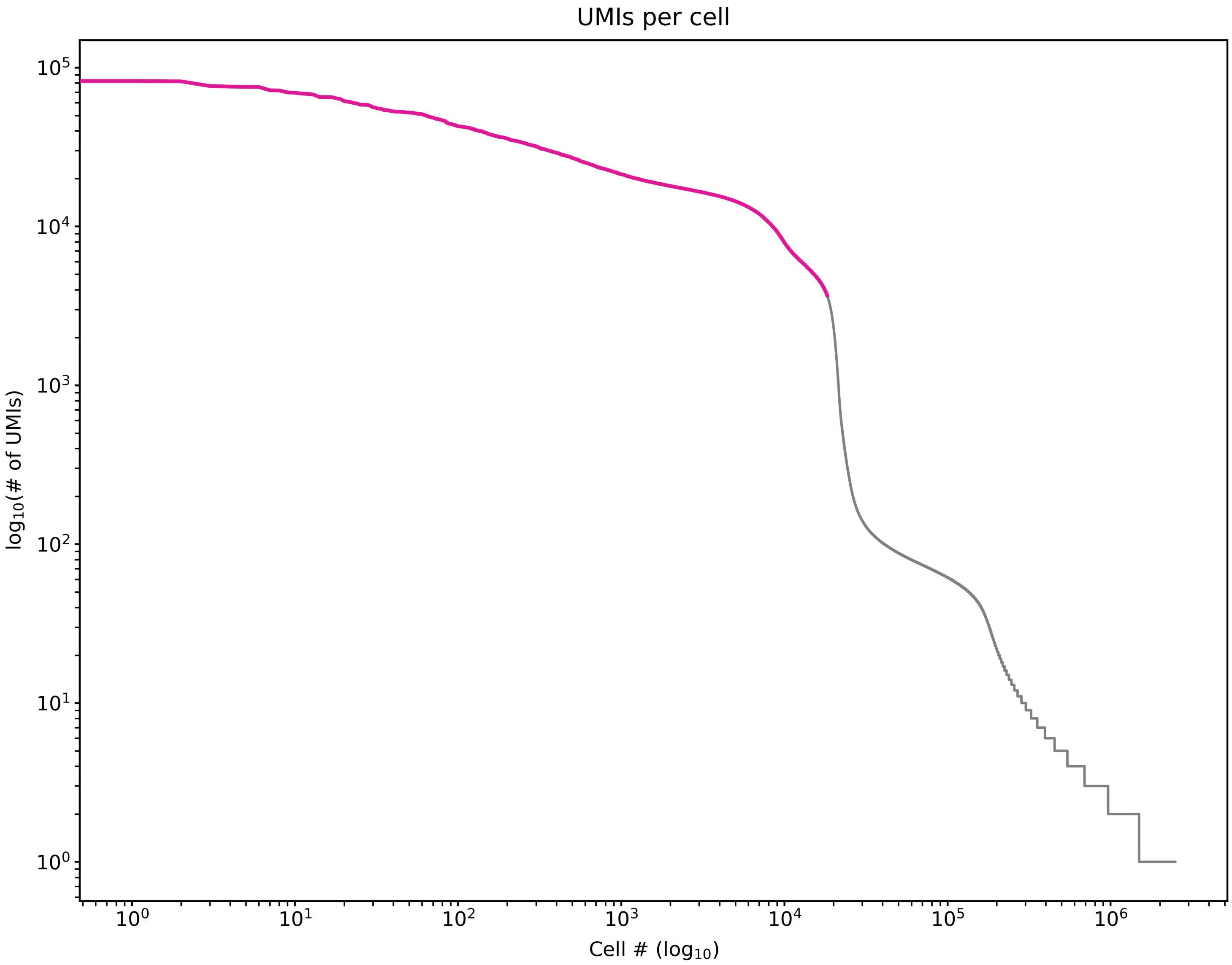


Cell Statistics

Summary

Estimated Number of Cells	17,654
Reads in Cells	87.13%
Mean Reads per Cell	14,011
Median UMIs per Cell	9,524

Barcode Rank Plot



Read Statistics

Summary

Reads	288,007,196
Read Type	SEGMENT
Reads with 5' and 3' Primers with extracted UMIs and Barcodes	285,630,076
Non-Concatamer Reads with 5' and 3' Primers and Poly-A Tail (FLNC reads)	283,891,388
FLNC Reads with Valid Barcodes	271,375,072
FLNC Reads with Valid Barcodes, corrected	280,263,363
Reads after Barcode Correction and UMI Deduplication	197,544,312

Transcript Statistics

Summary

FLNC Reads Mapped Confidently to Genome	191,189,865
FLNC Reads Mapped Confidently to Transcriptome	119,177,062
Total Unique Genes	2,160,147
Total Unique Genes, filtered	135,011
Total Unique Genes, known genes only	39,976
Total Unique Genes, filtered, known genes only	31,071
Total Unique Transcripts	5,637,804
Total Unique Transcripts, filtered	1,384,709
Total Unique Transcripts, known transcripts only	108,978
Total Unique Transcripts, filtered, known transcripts only	99,281

Transcript Summary

Value	Analysis Metric
3,154	Median Genes per Cell
2,407	Median Genes per Cell, known genes only
4,137	Median Transcripts per Cell
1,950	Median Transcripts per Cell, known transcripts only
2,160,147	Total Unique Genes
39,976	Total Unique Genes, known genes only
5,637,804	Total Unique Transcripts
108,978	Total Unique Transcripts, known transcripts only

Transcript Summary, filtered

Value	Analysis Metric
1,772	Median Genes per Cell
1,743	Median Genes per Cell, known genes only
2,255	Median Transcripts per Cell
1,670	Median Transcripts per Cell, known transcripts only
135,011	Total Unique Genes
31,071	Total Unique Genes, known genes only
1,384,709	Total Unique Transcripts
99,281	Total Unique Transcripts, known transcripts only

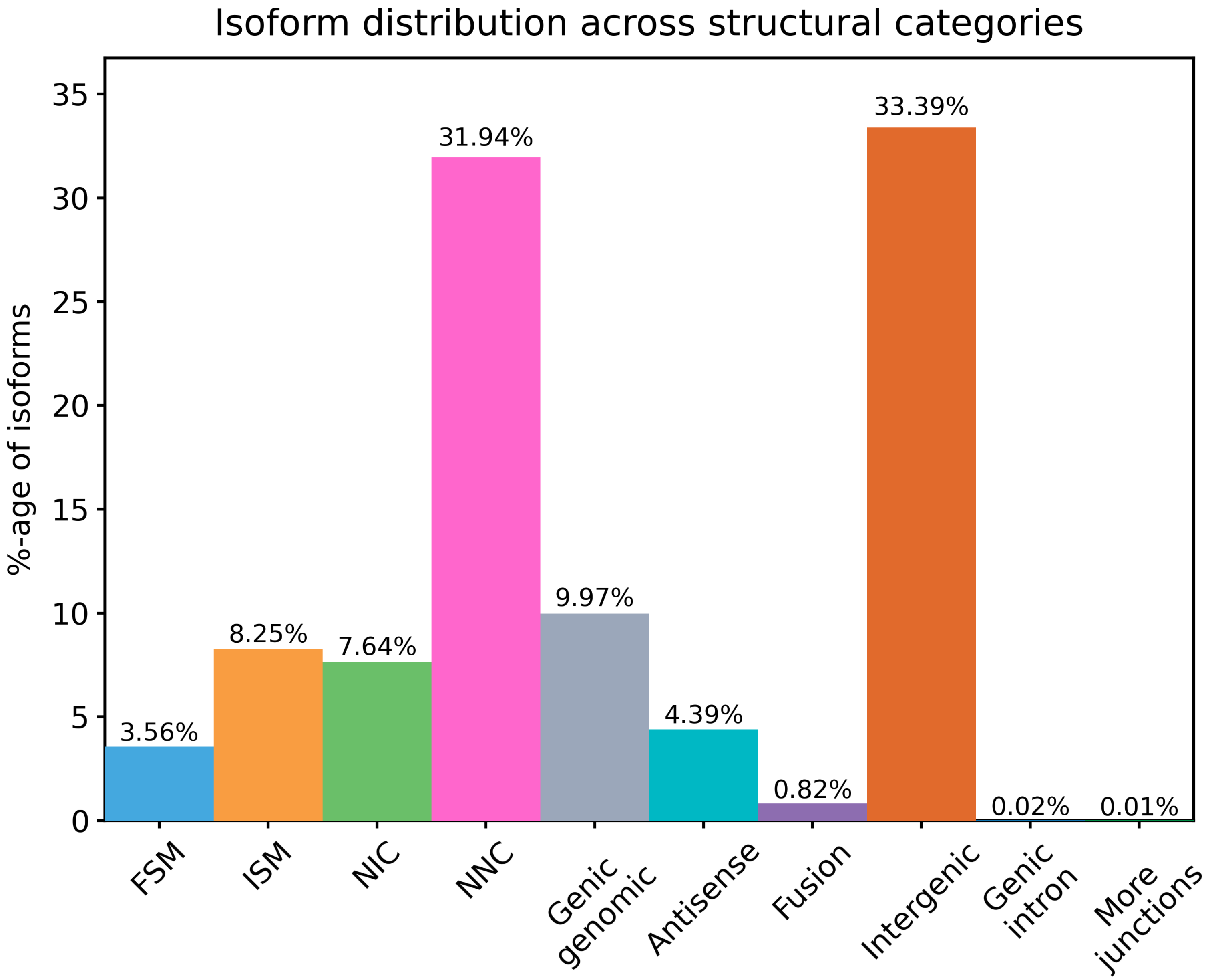
Transcript Classification

Category	Count	CAGE Detected	CAGE Detected (%)	polyA Motif Detected	polyA Motif Detected (%)
FSM	222891	128967	57.86%	67652	30.35%
ISM	517523	212697	41.10%	156434	30.23%
NIC	478879	308184	64.36%	161479	33.72%
NNC	2002321	1231439	61.50%	757232	37.82%
Antisense	275513	12153	4.41%	82855	30.07%
Fusion	51581	28536	55.32%	20474	39.69%
More junctions	857	418	48.77%	258	30.11%
Genic intron	1099	6	0.55%	362	32.94%
Genic genomic	625205	149505	23.91%	189730	30.35%
Intergenic	2093371	21491	0.999	712321	34.03%

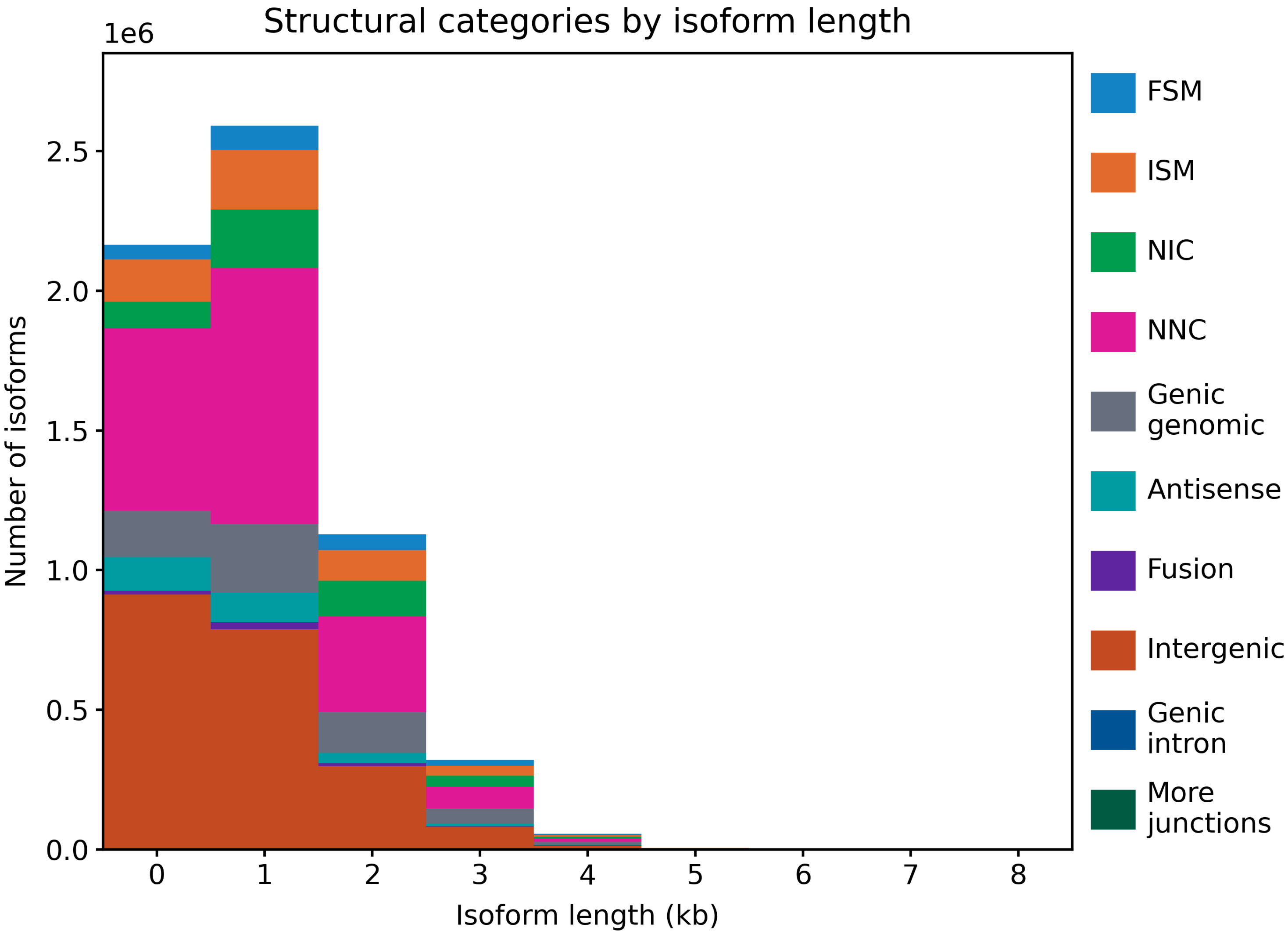
Transcript Classification, filtered

Category	Count	CAGE Detected	CAGE Detected, (%)	polyA Detected	polyA Detected, (%)
FSM	187860	111558	59.38%	67652	36.01%
ISM	356165	161074	45.22%	138556	38.90%
NIC	375188	243279	64.84%	160170	42.69%
NNC	739730	466058	63.00%	351468	47.51%
Antisense	23799	2438	10.24%	14541	61.10%
Fusion	24723	14126	57.14%	12404	50.17%
More junctions	490	252	51.43%	218	44.49%
Genic intron	0	0	0.00%	0	0.00%
Genic genomic	24507	10487	42.79%	12053	49.18%
Intergenic	96991	2584	2.66%	70613	72.80%

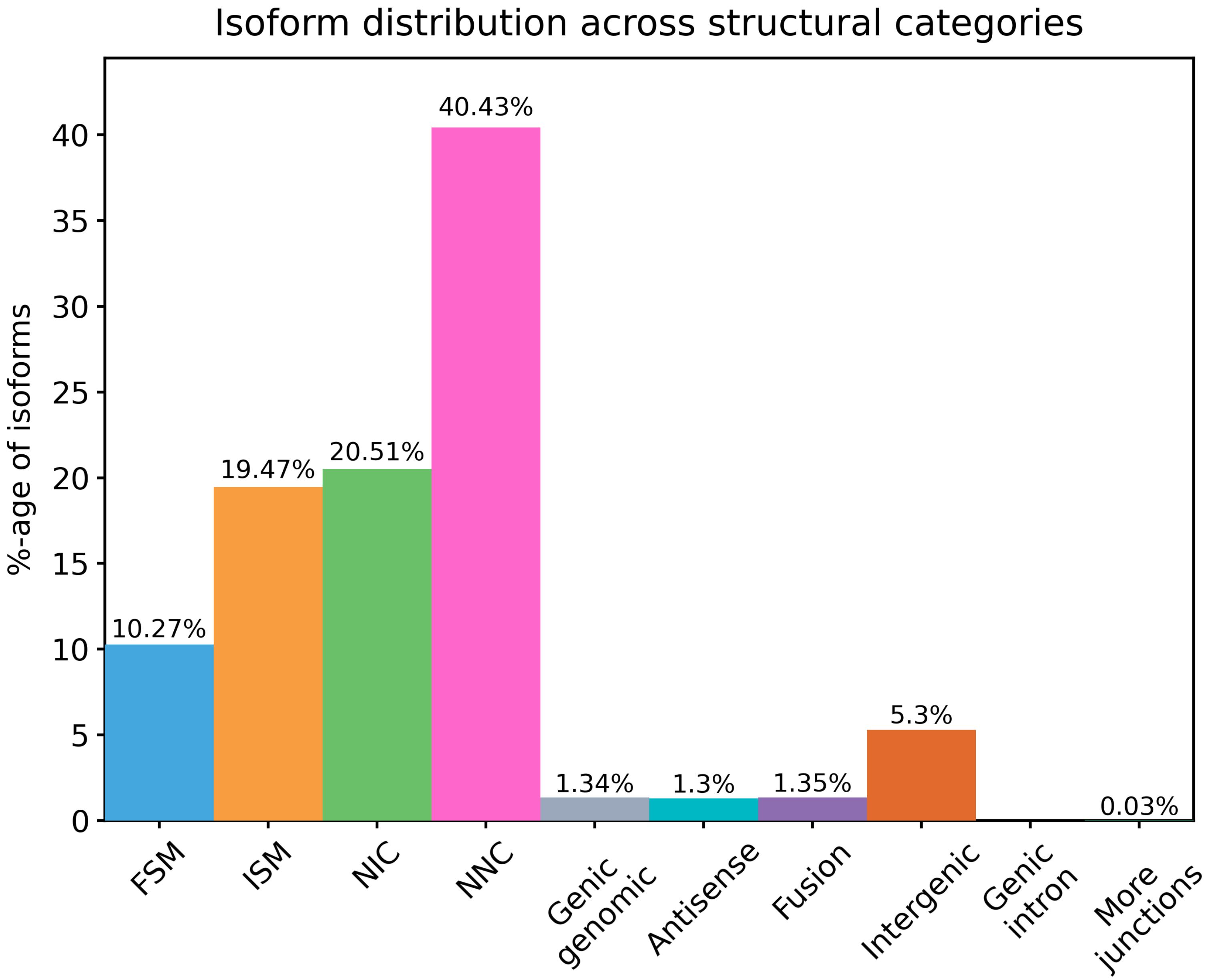
Isoform Distribution Across Categories


Isoform distribution across structural categories
%-age of isoforms
0
5
10
15
20
25
30
35
3.56%
8.25%
7.64%
31.94%
9.97%
4.39%
0.82%
33.39%
0.02%
0.01%
FSM
ISM
NIC
NNC
Genic genomic
Antisense
Fusion
Intergenic
Genic intron
More junctions

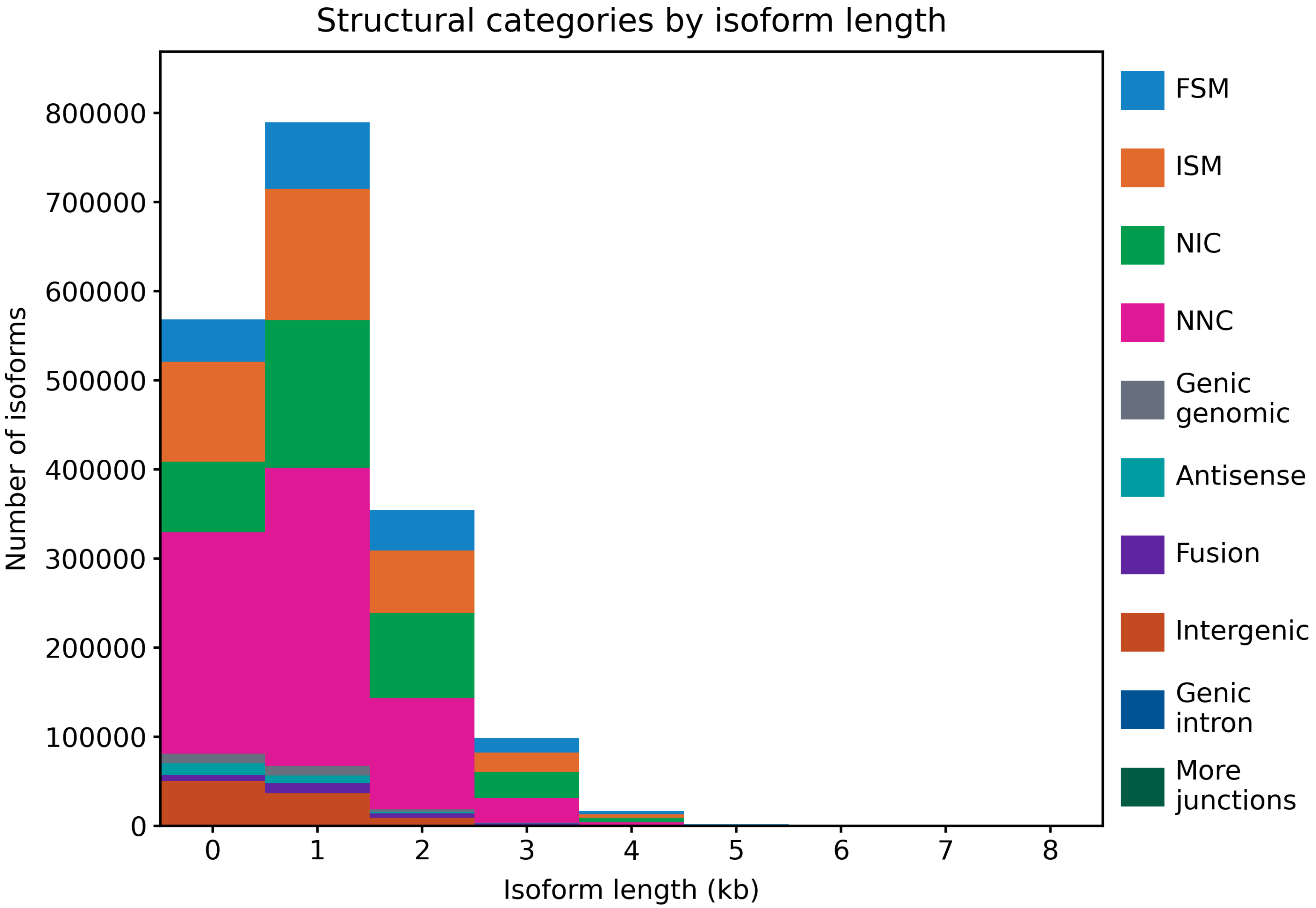
Structural Categories by Transcript Length



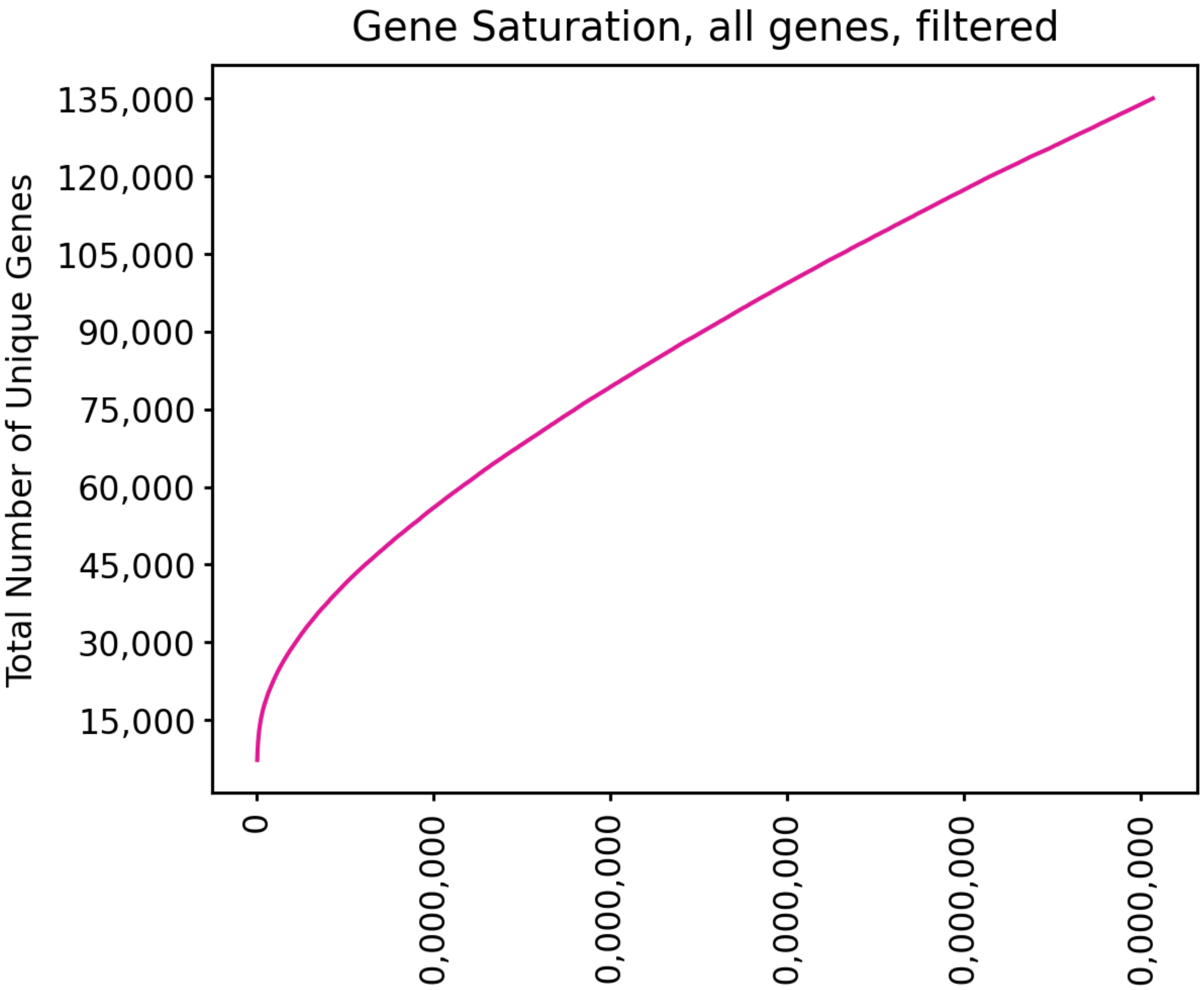
Isoform Distribution Across Categories

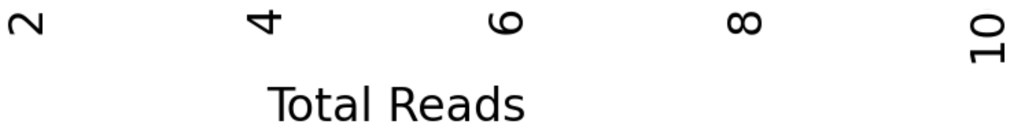

Isoform distribution across structural categories
%-age of isoforms
0
5
10
15
20
25
30
35
40
10.27%
19.47%
20.51%
40.43%
1.34%
1.3%
1.35%
5.3%
0.03%
FSM
ISM
NIC
NNC
Genic genomic
Antisense
Fusion
Intergenic
Genic intron
More junctions

Structural Categories by Transcript Length

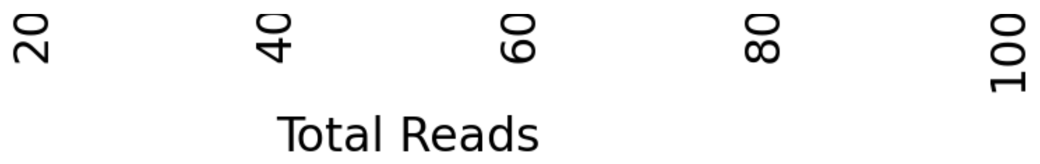


sc_isoseq_transcript_statistics.gene_saturation_plotgroup.gene_saturation_plot_0




2
4
6
8
10
Total Reads

sc_isoseq_transcript_statistics.gene_saturation_plotgroup.gene_saturation_plot_1


20
40
60
80
100
Total Reads