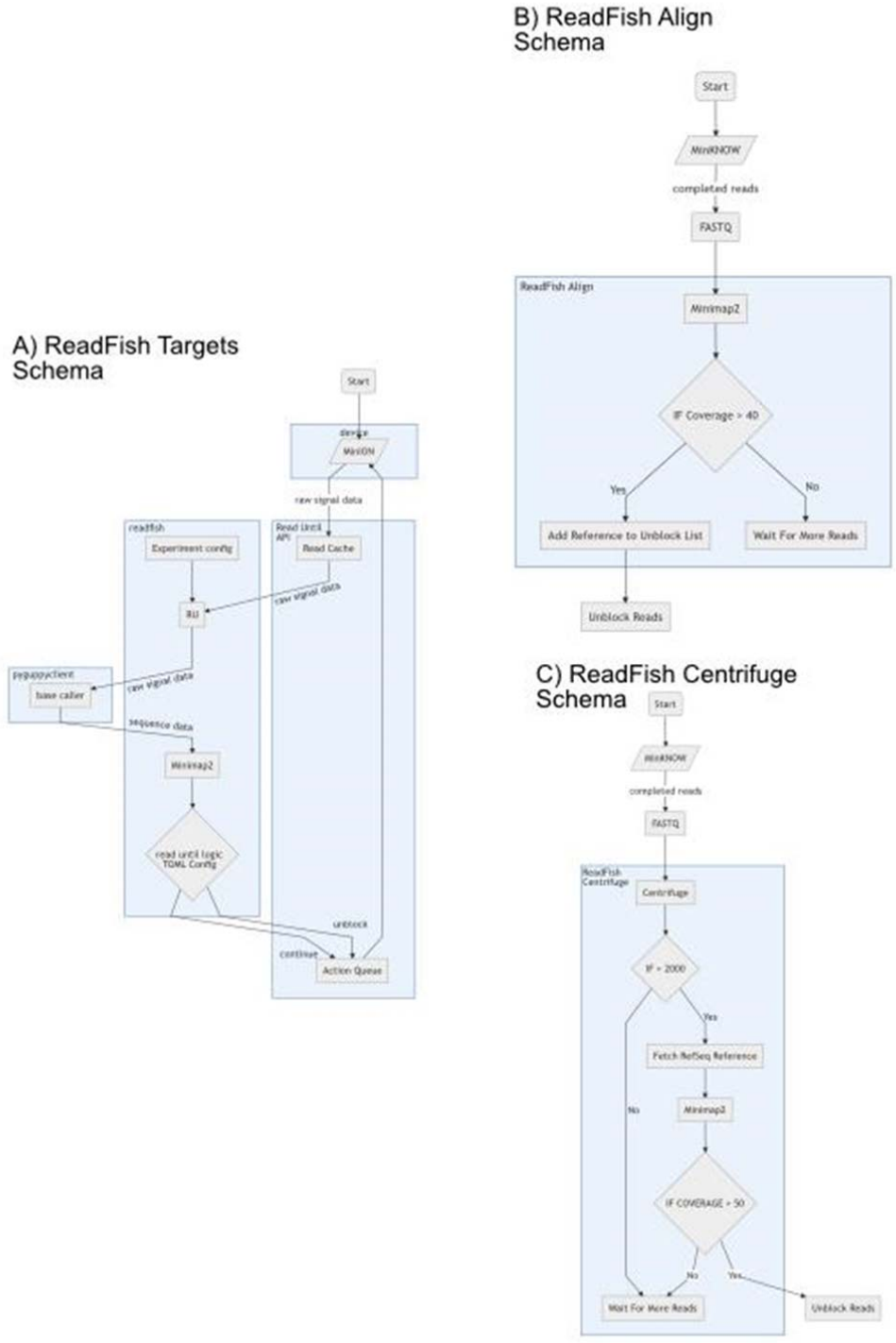
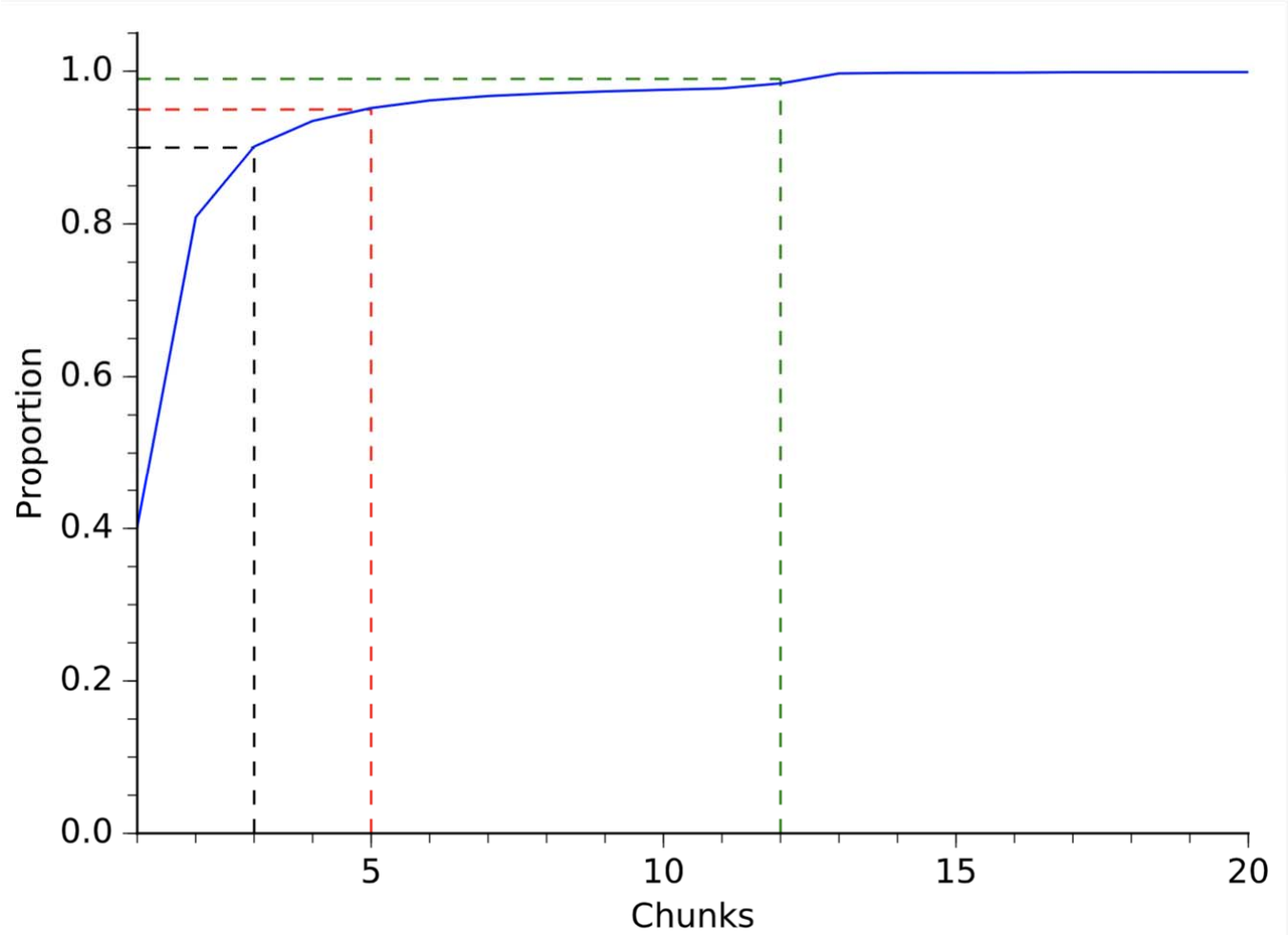
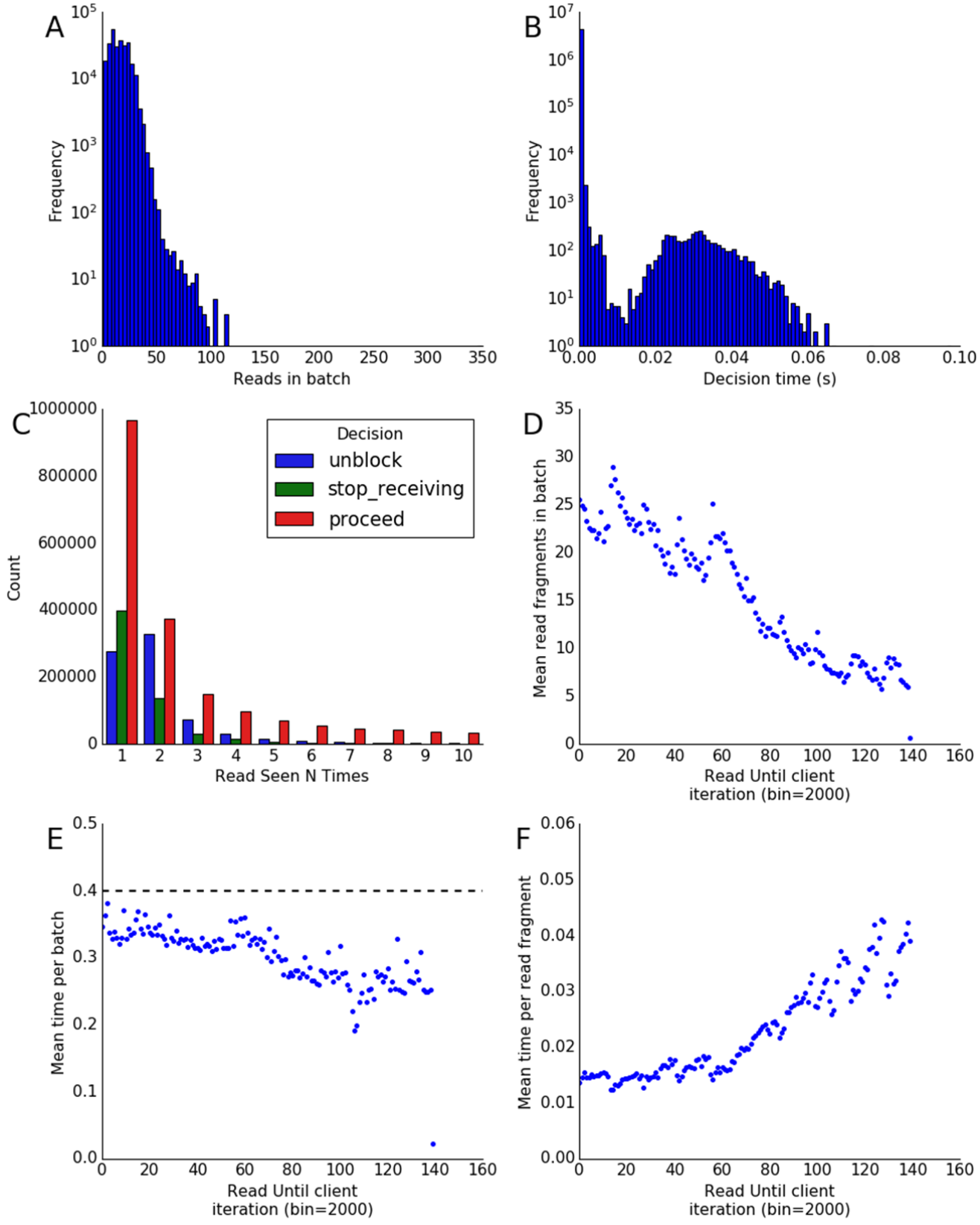
Supplementary Figures



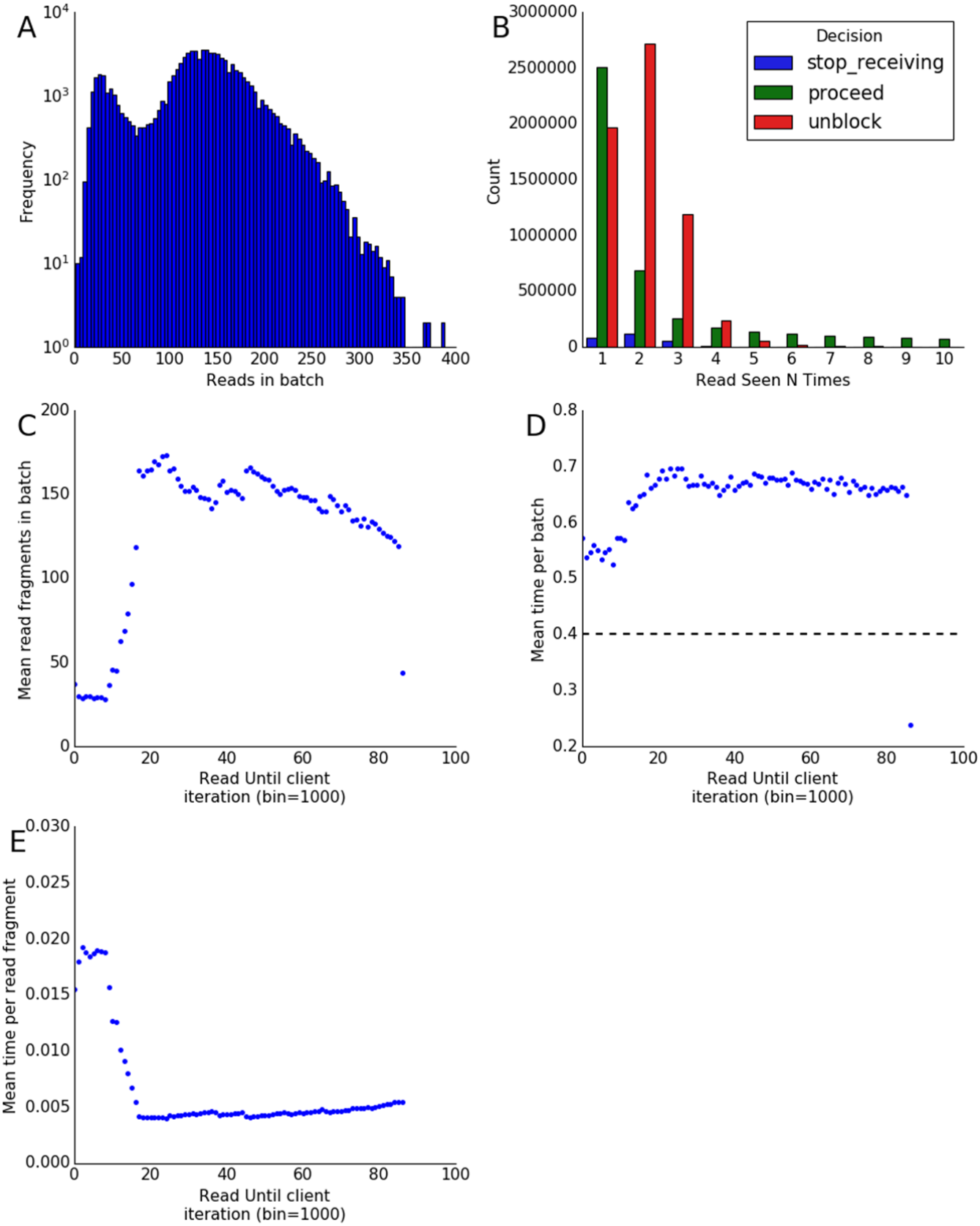
Supplementary Figure 1: Flow diagrams of ReadFish components. A) The core ReadFish targets program which manages base calling and aligning read fragments from the device. B) The ReadFish Align program monitors completed reads and computes coverage based on a known reference. C) ReadFish Centrifuge monitors completed reads as in ReadFish Align, but uses centrifuge to classify reads and then downloads a RefSeq genome once a defined threshold is reached.



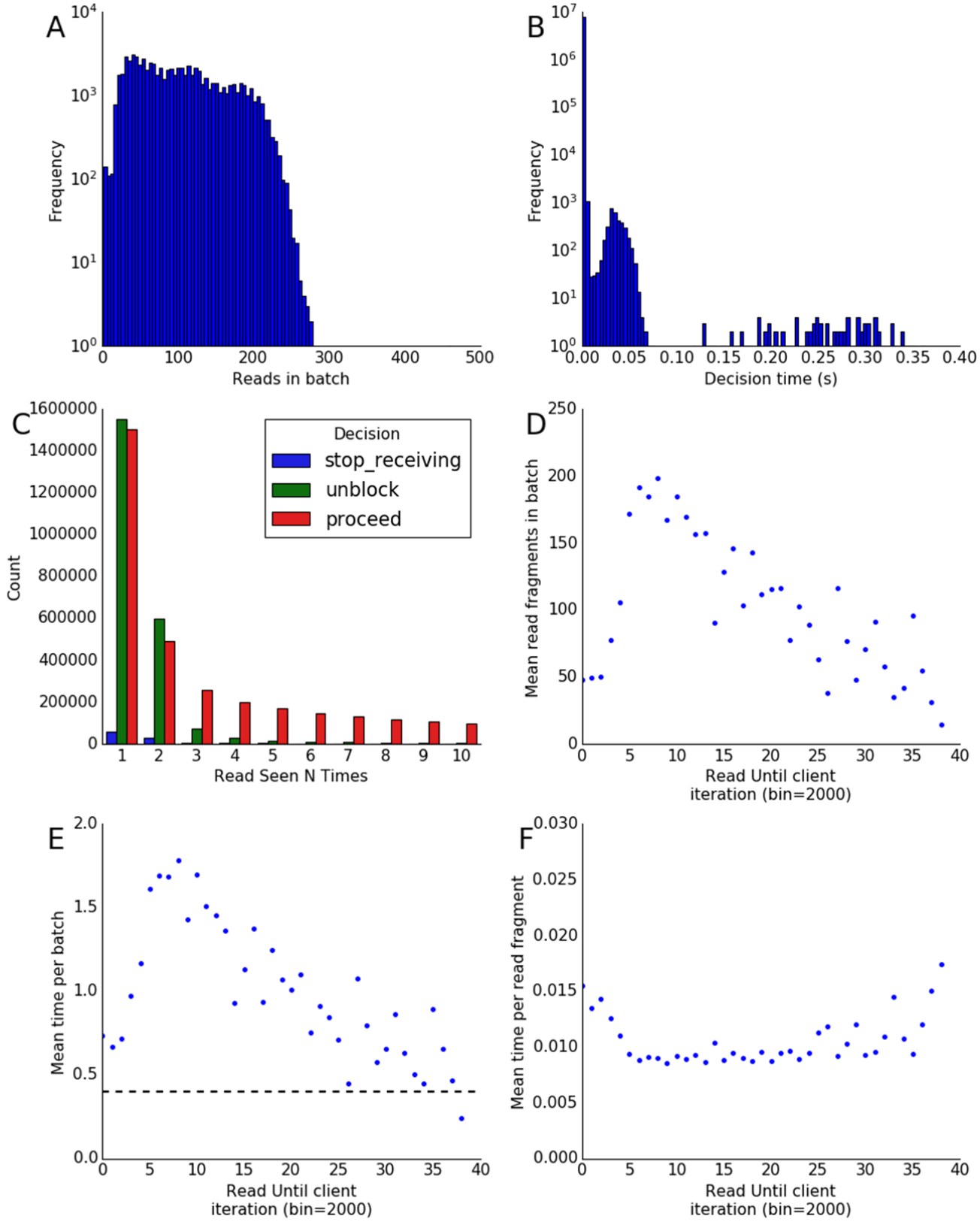
Supplementary Figure 2: Proportion of read fragments processed in a given number of chunks. 90% of reads (black dashed line) are processed in 3 chunks, 95% (red dashed line) in 5 chunks and 99% (green dashed line) in 12 chunks. This plot includes every sequencing run presented in this paper.



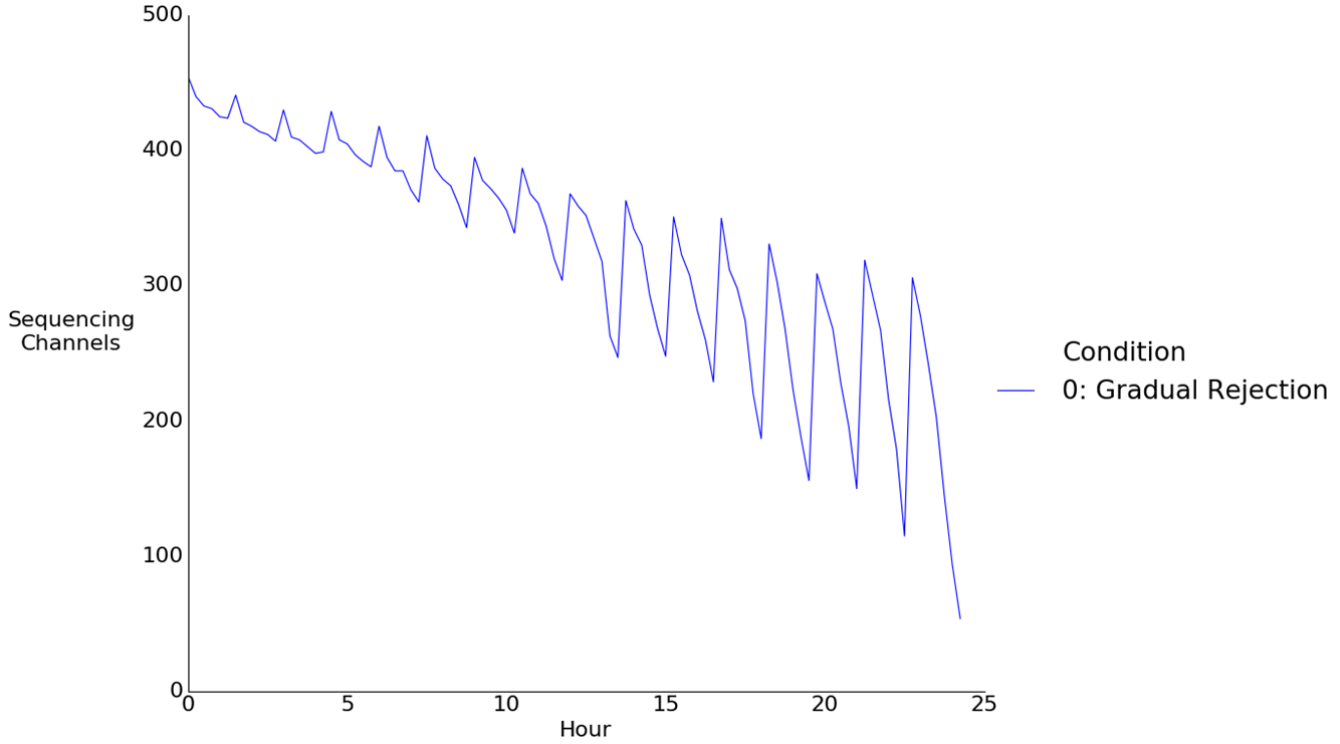
Supplementary Figure 3: Human Chromosome Enrichment. A) Histogram of read batch size throughout the selective sequencing program. B) Histogram of decision times (time to choose unblock, stop receiving, or proceed from an alignment). C) Counts of decision classifications for read fragments seen a given number of times. D) Mean batch size, in bins of 2000, seen throughout the selective sequencing program. E) Mean process time, in bins of 2000, for batches of read fragments throughout the run. Dashed line indicates the minimum data chunk size delivered by MinKNOW. F) Mean decision time per read fragment, in bins of 2000, throughout the run. As the number of reads in a batch reduces, the overhead time of calling becomes more apparent.



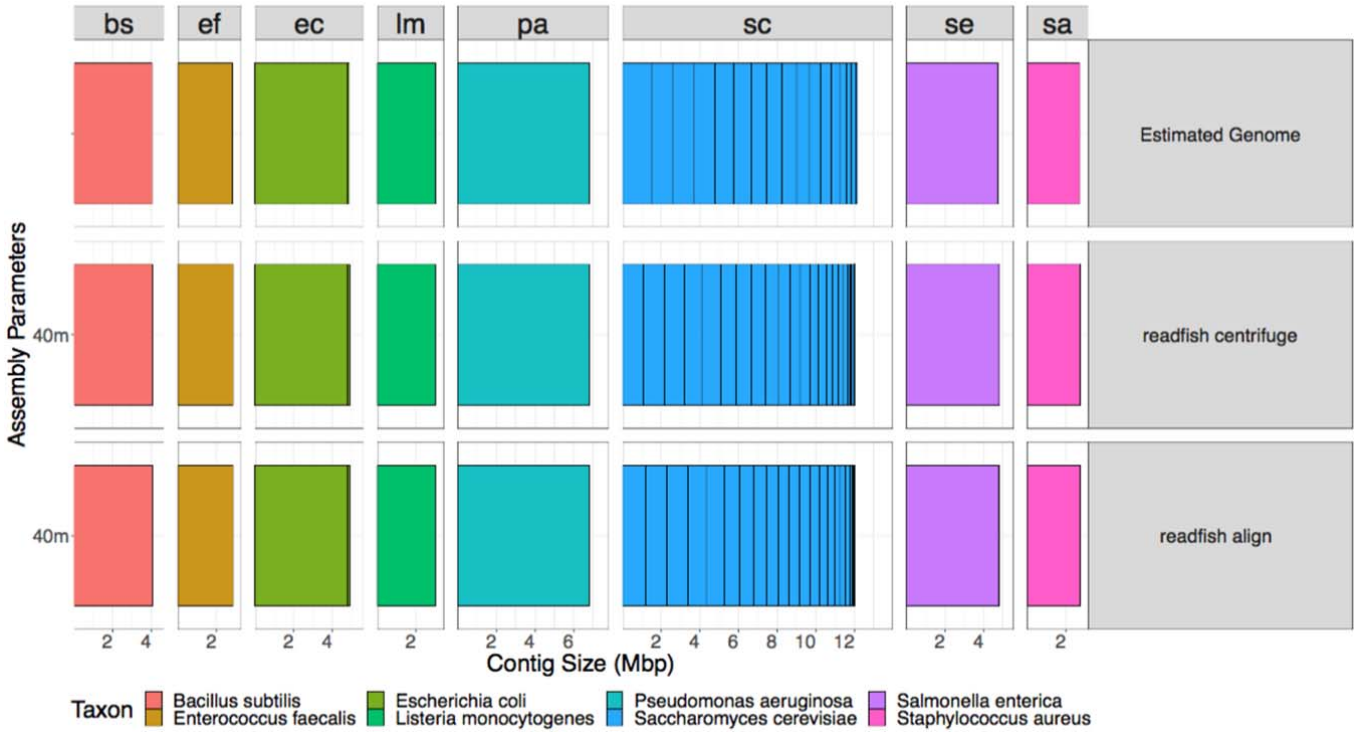
Supplementary Figure 4: ReadFish Align A) Histogram of read batch size throughout the selective sequencing program. B) Counts of decision classifications for read fragments seen a given number of times. C) Mean batch size, in bins of 1000, seen throughout the selective sequencing program. Dashed line indicates the minimum data chunk size delivered by MinKNOW. D) Mean process time, in bins of 1000, for batches of read fragments throughout the run. E) Mean decision time per read fragment, in bins of 1000, throughout the run. As the number of reads in a batch reduces, the overhead time of calling becomes more apparent.



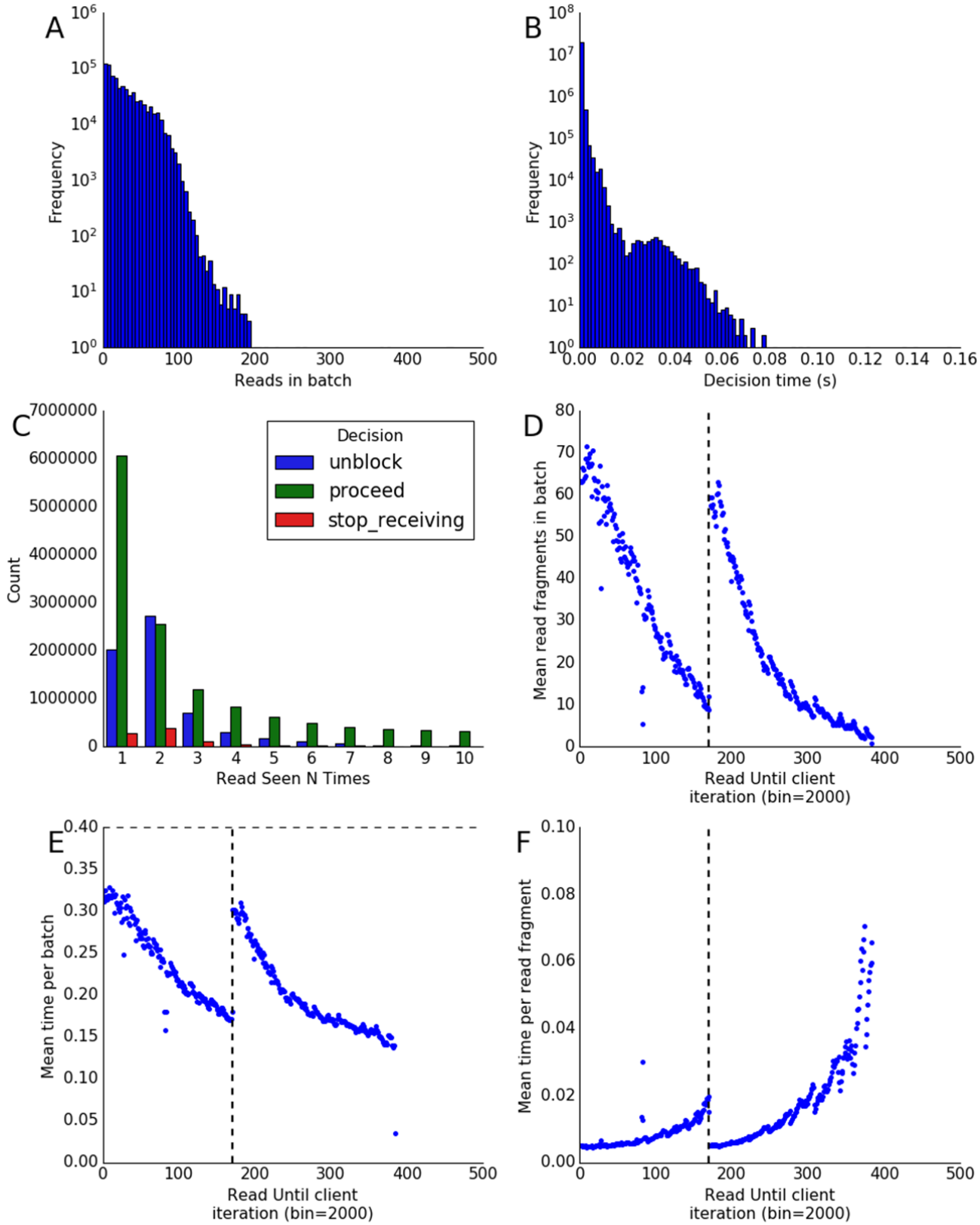
Supplementary Figure 5: ReadFish Centrifuge Performance. A) Histogram of read batch size throughout the selective sequencing program. B) Histogram of decision times (time to choose unblock, stop receiving, or proceed from an alignment). C) Counts of decision classifications for read fragments seen a given number of times. D) Mean batch size, in bins of 2000, seen throughout the selective sequencing program. E) Mean process time, in bins of 2000, for batches of read fragments throughout the run. Dashed line indicates the minimum data chunk size delivered by MinKNOW, the longer processing times here, reflect the addition of centrifuge to the pipeline. F) Mean decision time per read fragment, in bins of 2000, throughout the run. As the number of reads in a batch reduces, the overhead time of calling becomes more apparent.



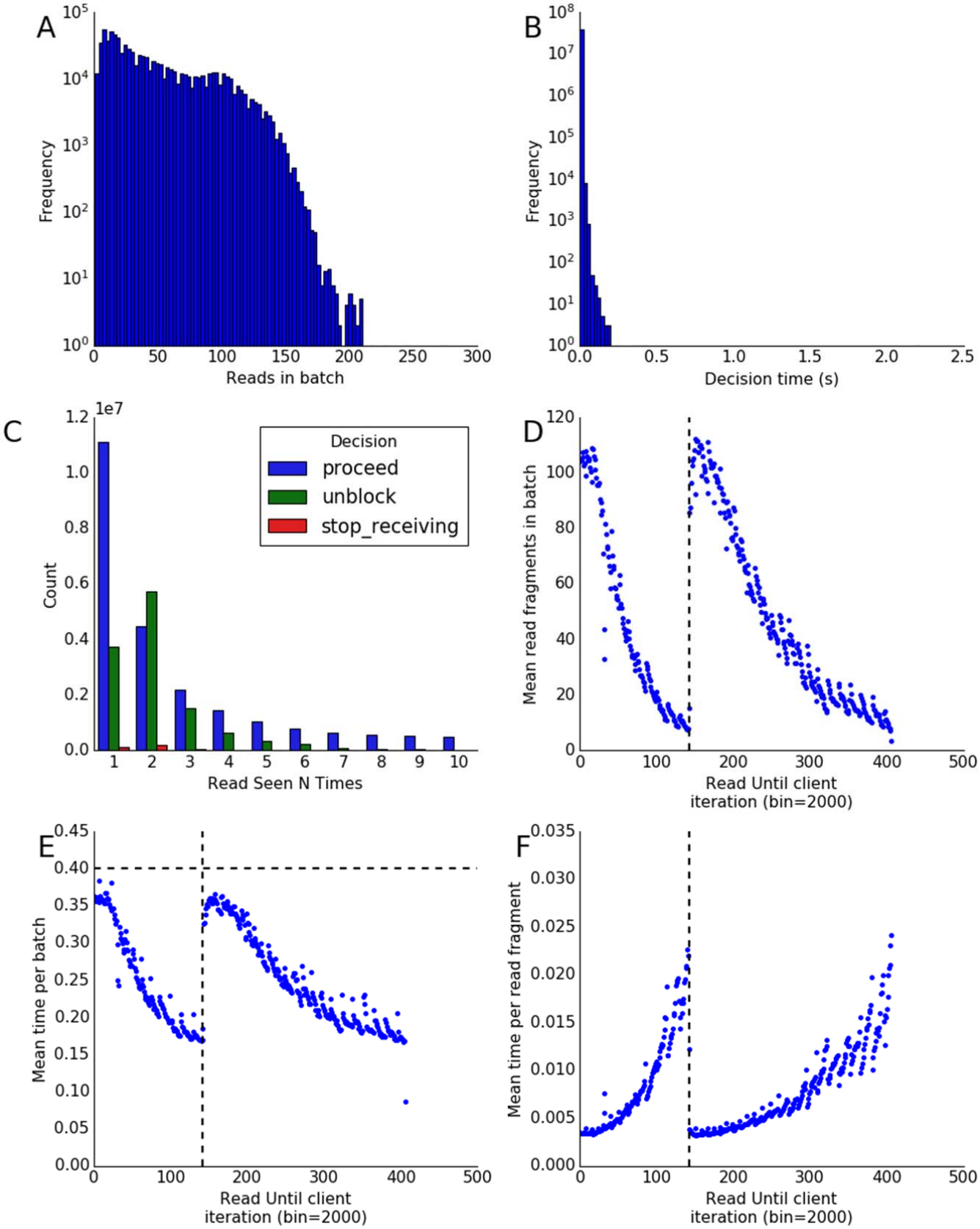
Supplementary Figure 6: Active sequencing channels plotted over time for a ReadFish Centrifuge sequencing run. This illustrates flow cell blocking over time.



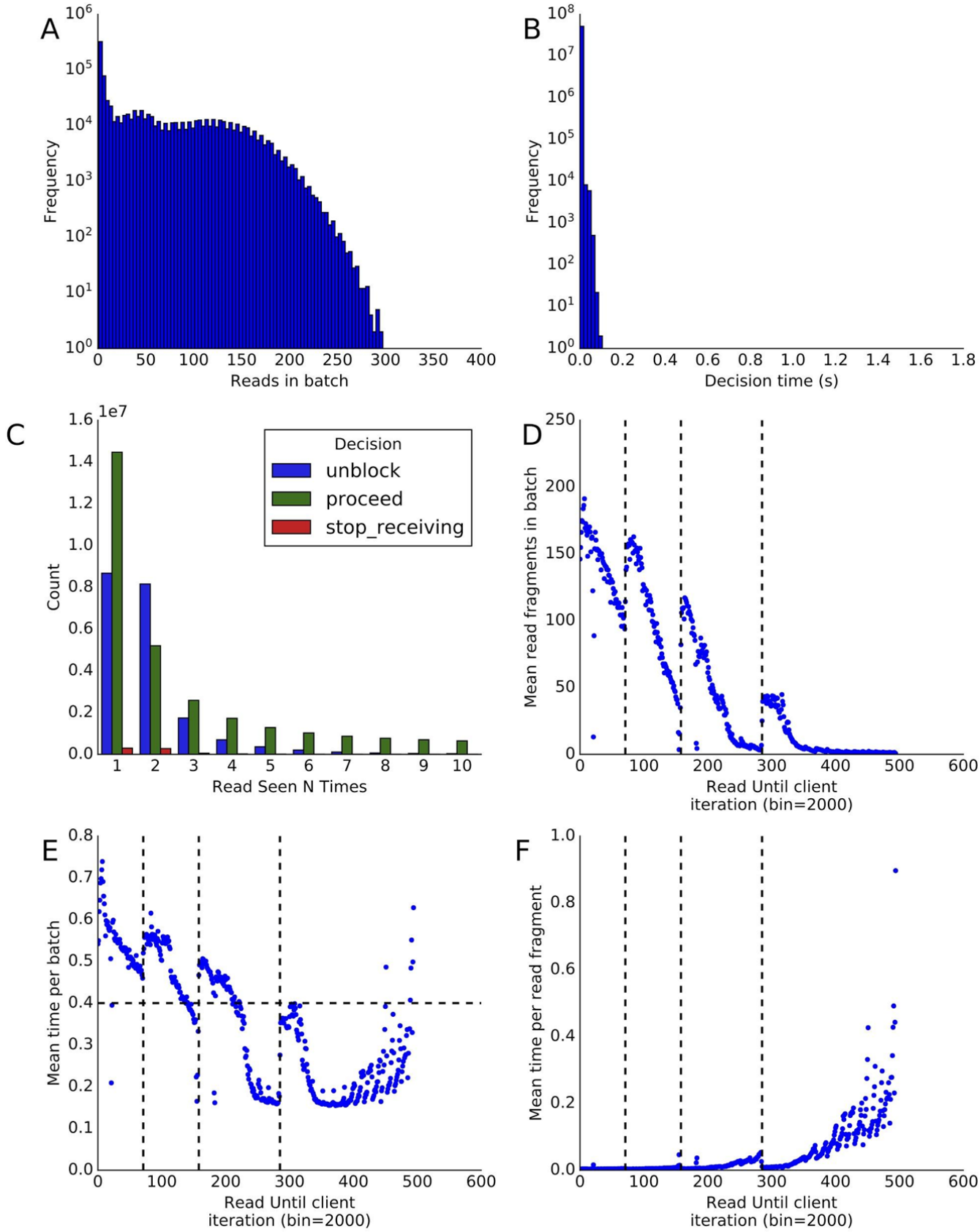
Supplementary Figure 7. Assemblies for ZymoBIOMICS readfish align/centrifuge enriched data. Data assembled using MetaFlye using an estimated genome size of 40 Mb. Species and composition are: bs - Bacillus subtilis (14%), ef - Enterococcus faecalis (14%), ec - Escherichia coli (14%), lm - Listeria monocytogenes (14%), pa - Pseudomonas aeruginosa (14%), sc - Saccharomyces cerevisiae (2%), se - Salmonella enterica (14%), sa - Staphylococcus aureus (14%). readfish centrifuge are the results of the reference unaware adaptive sampling presented in Fig. 3 of the main text. readfish align are the results of the reference aware adaptive sampling presented in Fig. 2 of the main text.



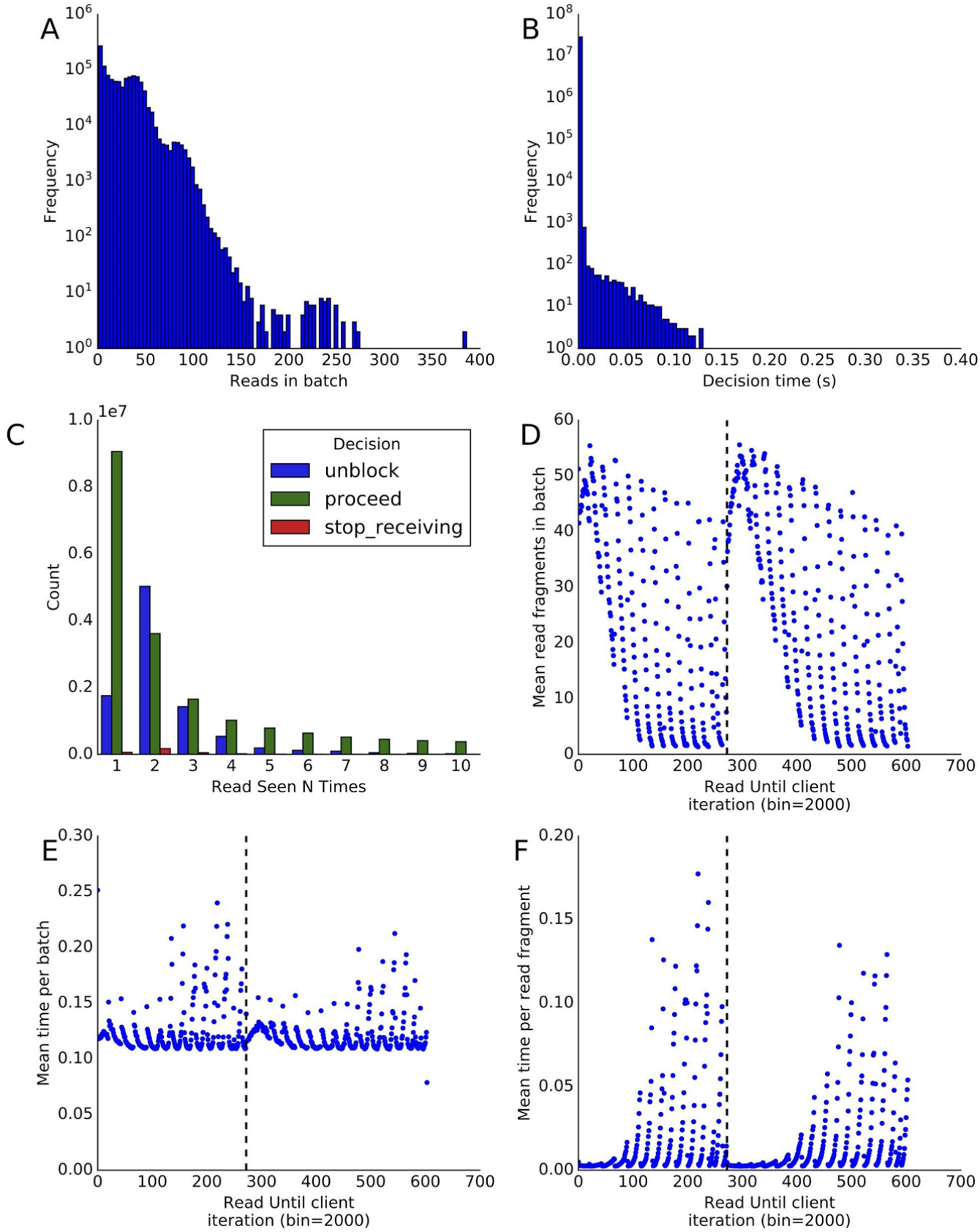
Supplementary Figure 8: Human Exome A) Histogram of read batch size throughout the selective sequencing program. B) Histogram of decision times (time to choose unblock, stop receiving, or proceed from an alignment). C) Counts of decision classifications for read fragments seen a given number of times. D) Mean batch size, in bins of 2000, seen throughout the selective sequencing program. E) Mean process time, in bins of 2000, for batches of read fragments throughout the run. F) Mean decision time per read fragment, in bins of 2000, throughout the run. As the number of reads in a batch reduces, the overhead time of calling becomes more apparent. The vertical dashed lines mark flushing and restart of the run and illustrate the benefits of flushing.



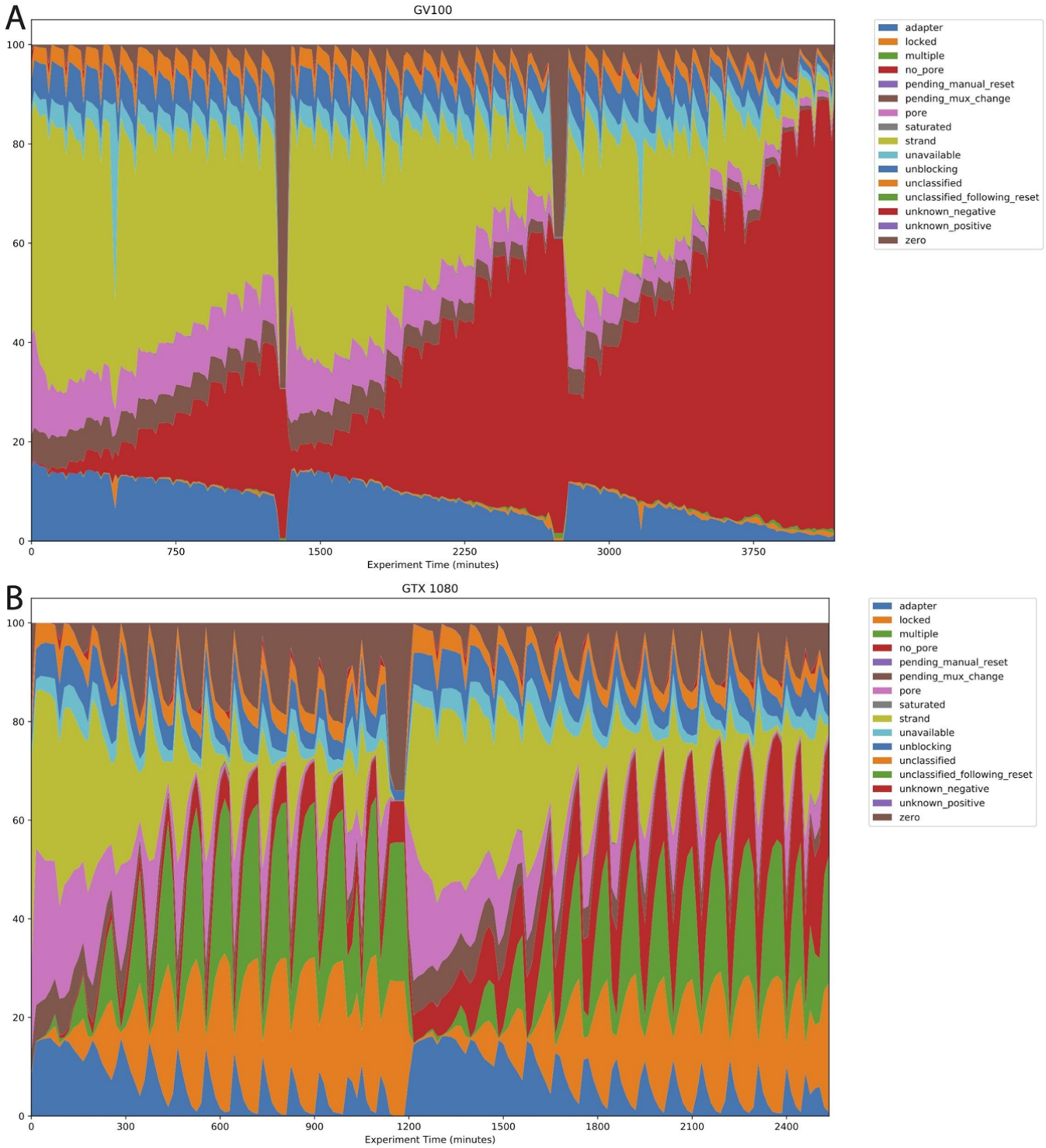
Supplementary Figure 9: COSMIC panel Run 1 A) Histogram of read batch size throughout the selective sequencing program. B) Histogram of decision times (time to choose unblock, stop receiving, or proceed from an alignment). C) Counts of decision classifications for read fragments seen a given number of times. D) Mean batch size, in bins of 2000, seen throughout the selective sequencing program. E) Mean process time, in bins of 2000, for batches of read fragments throughout the run. F) Mean decision time per read fragment, in bins of 2000, throughout the run. As the number of reads in a batch reduces, the overhead time of calling becomes more apparent.



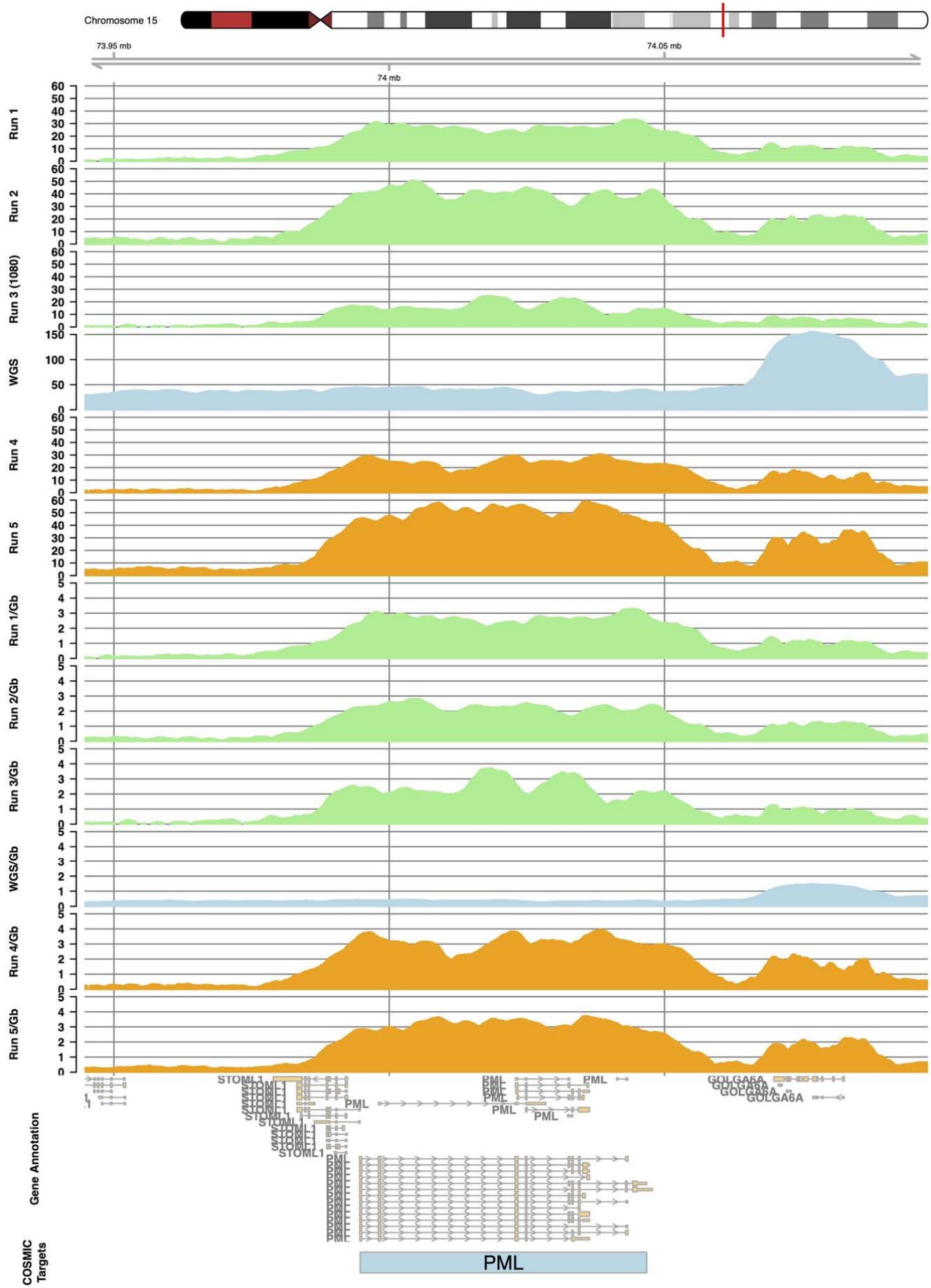
Supplementary Figure 10: COSMIC panel Run 2 A) Histogram of read batch size throughout the selective sequencing program. B) Histogram of decision times (time to choose unblock, stop receiving, or proceed from an alignment). C) Counts of decision classifications for read fragments seen a given number of times. D) Mean batch size, in bins of 2000, seen throughout the selective sequencing program. E) Mean process time, in bins of 2000, for batches of read fragments throughout the run. F) Mean decision time per read fragment, in bins of 2000, throughout the run. As the number of reads in a batch reduces, the overhead time of calling becomes more apparent.



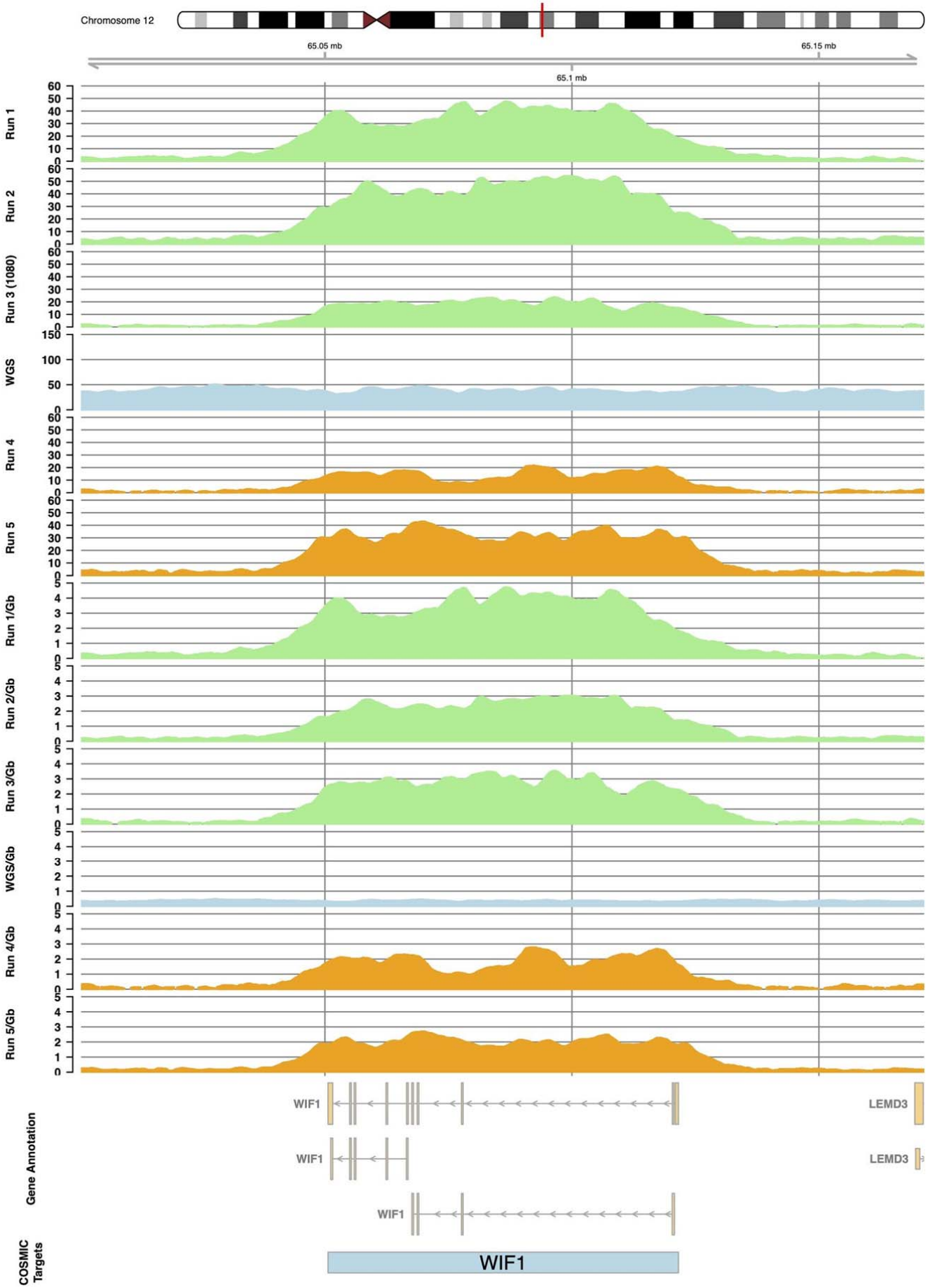
Supplementary Figure 11: COSMIC panel Run 3 - NVIDIA GeForce GTX 1080 A) Histogram of read batch size throughout the selective sequencing program. B) Histogram of decision times (time to choose unblock, stop receiving, or proceed from an alignment). C) Counts of decision classifications for read fragments seen a given number of times. D) Mean batch size, in bins of 2000, seen throughout the selective sequencing program. E) Mean process time, in bins of 2000, for batches of read fragments throughout the run. F) Mean decision time per read fragment, in bins of 2000, throughout the run. As the number of reads in a batch reduces, the overhead time of calling becomes more apparent.



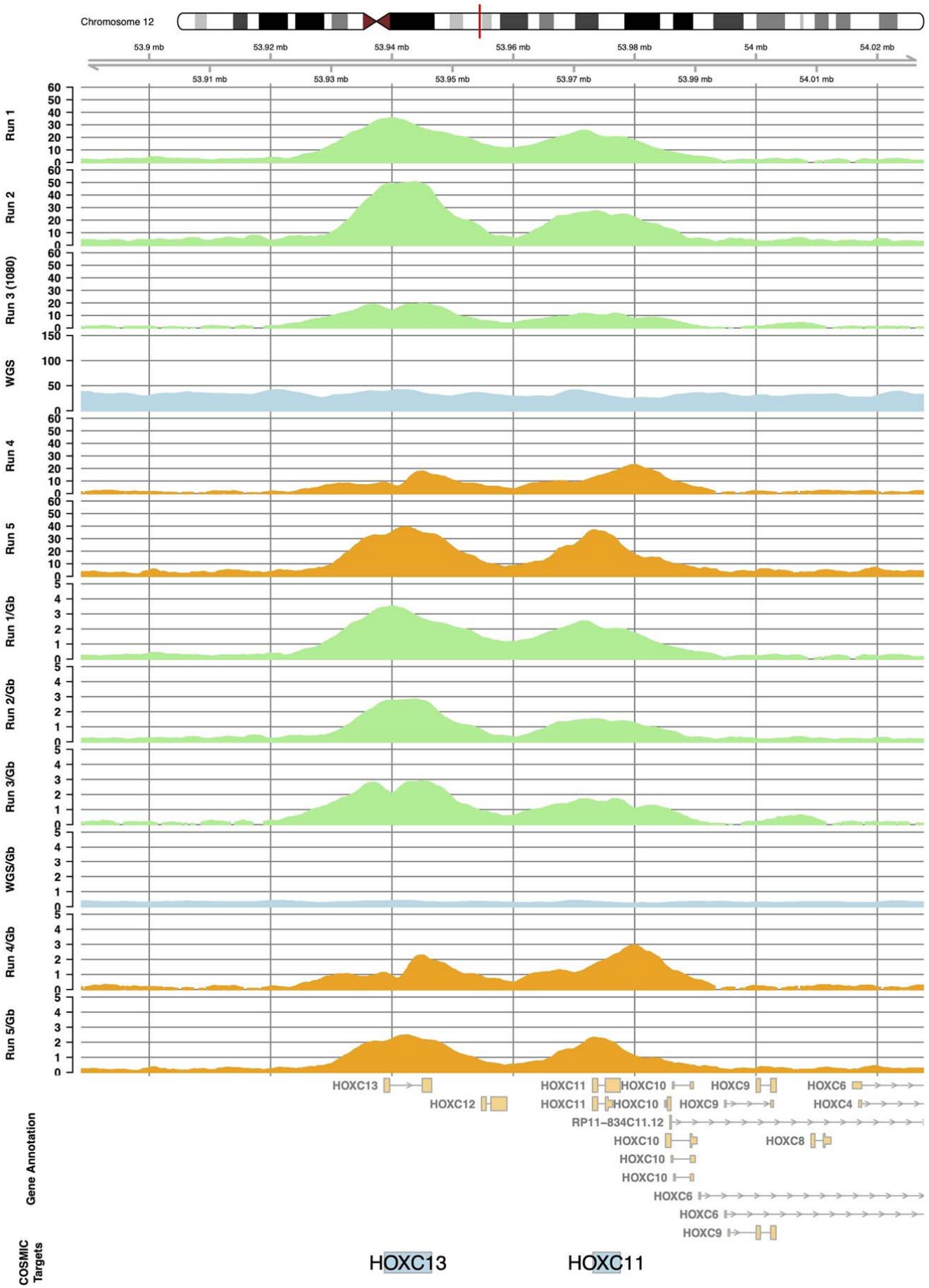
Supplementary Figure 12: Duty Time Plots. A) Indicative duty time plot for a flow cell performing well. The gradual accumulation of channels with no pore available for sequencing is clear. B) Unusual flowcell performance indicative of a flowcell issue where channels enter a multiple state. Even with this problem, this flowcell led to enrichment of targets when using readfish.



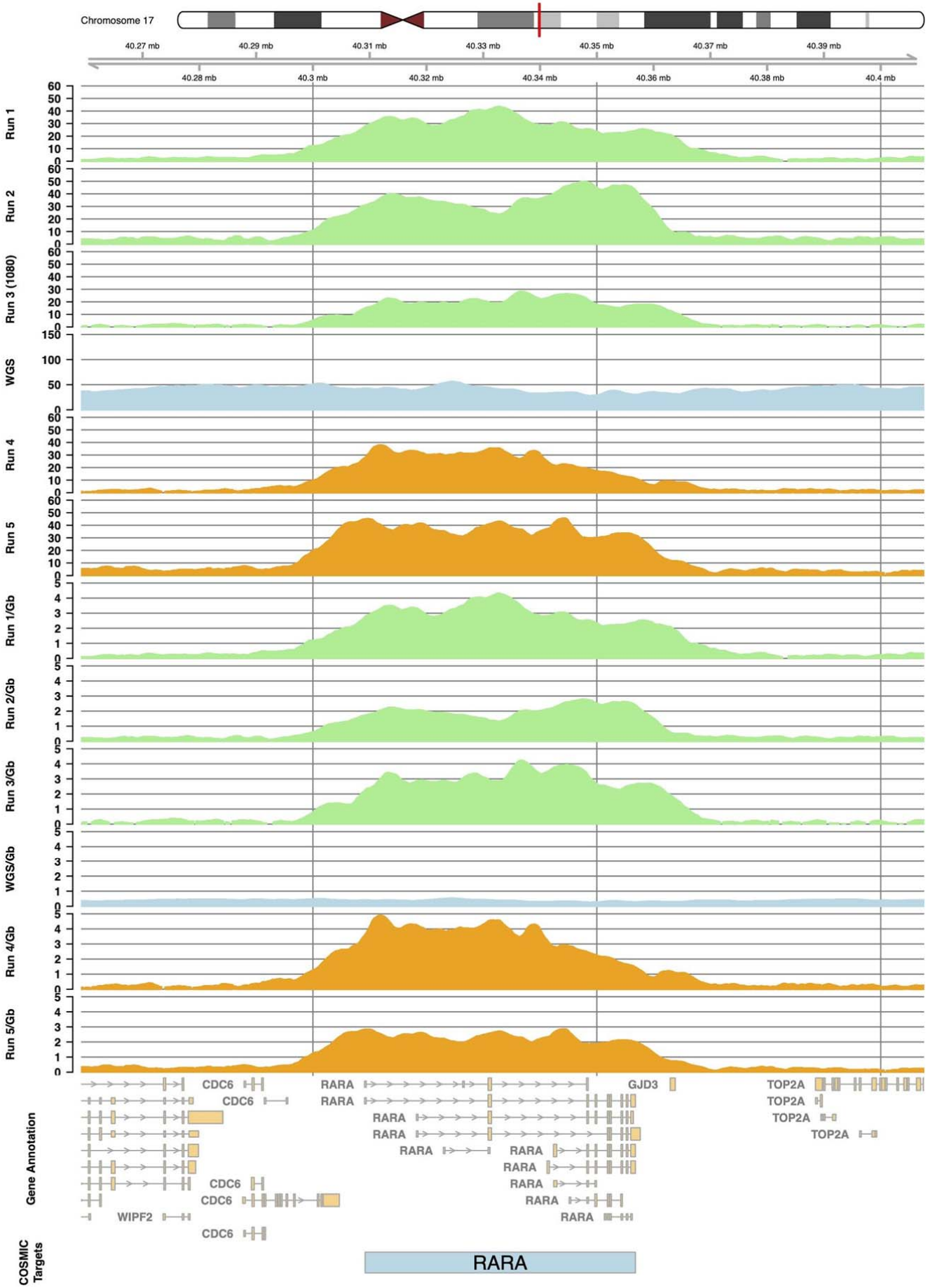
Supplementary Figure 13 - PML



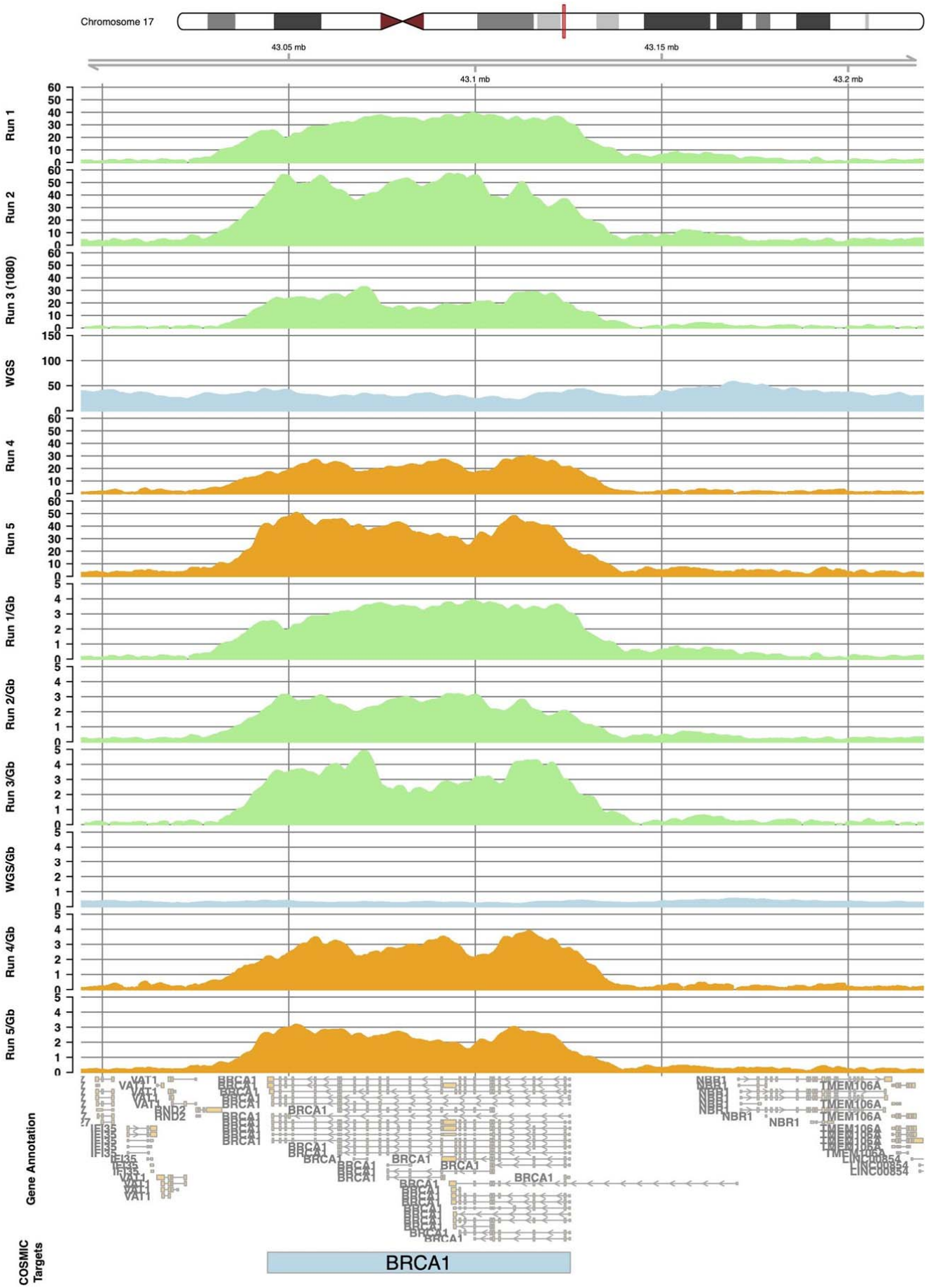
Supplementary Figure 14 - WIF1



Supplementary Figure 15 - HOXC11/C13

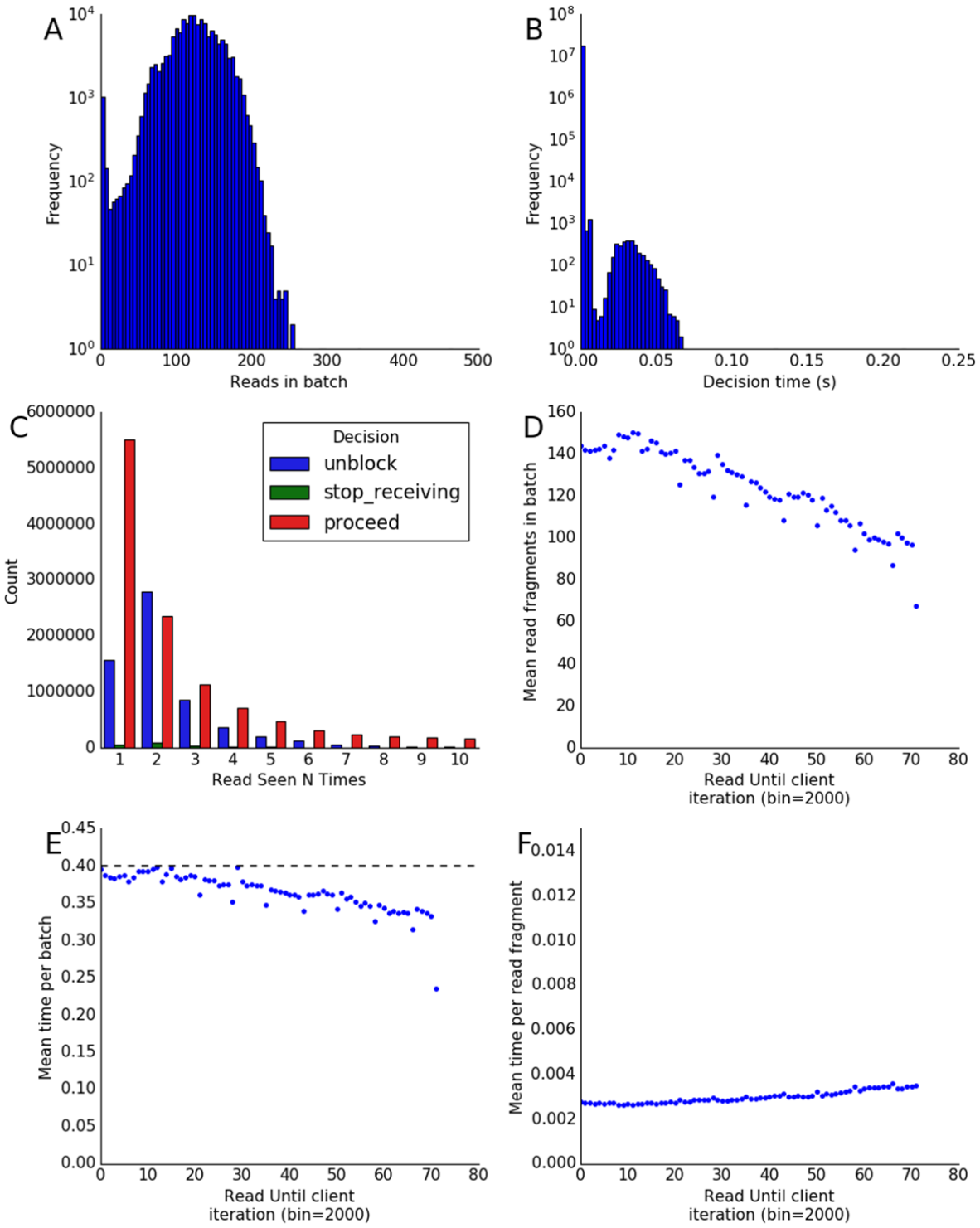


Supplementary Figure 16 - RARA

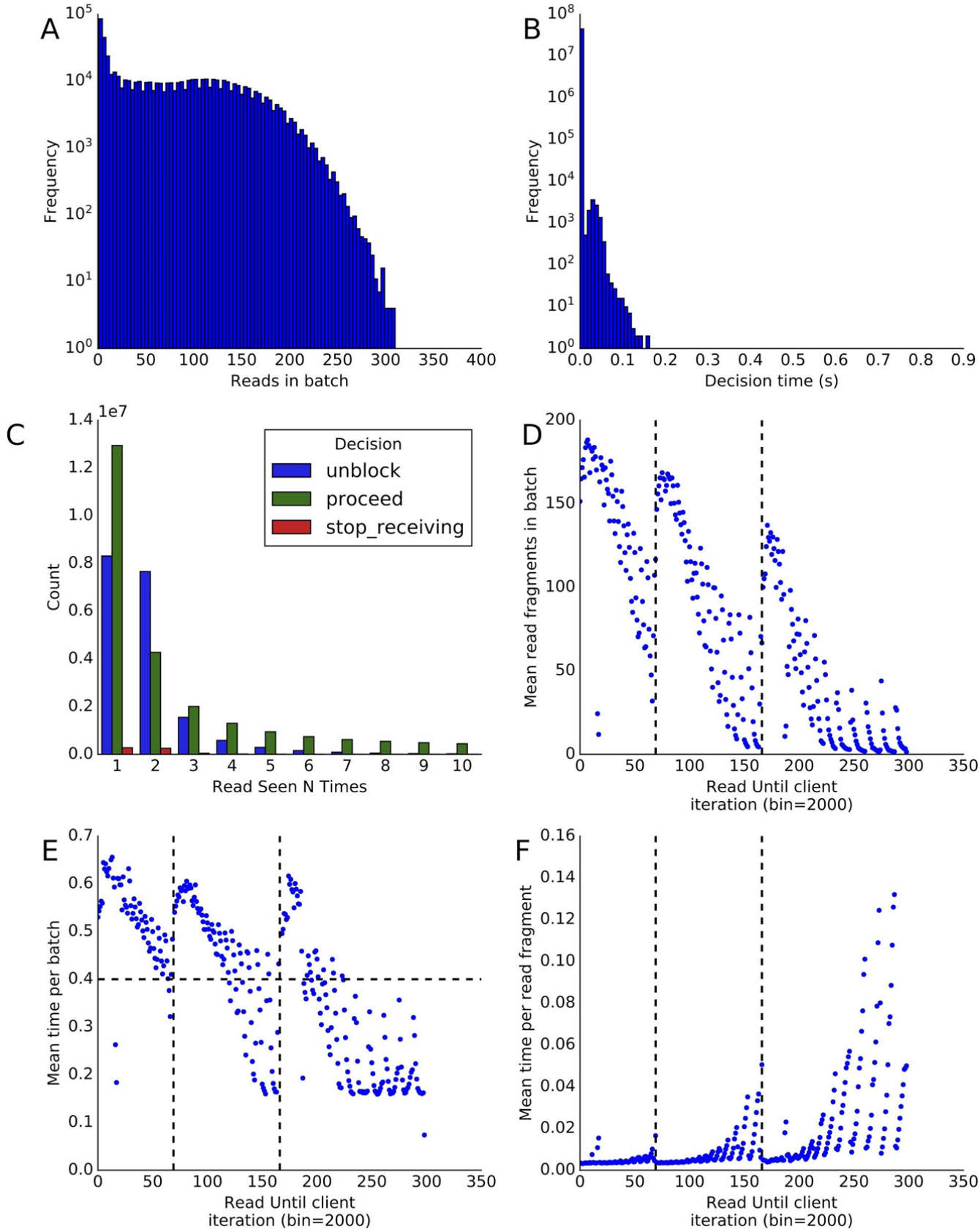


Supplementary Figure 17 - BRCA1

Supplementary Figures 13-17. Coverage plots for PML, WIF1, HOXC11/C13, RARA and BRCA1 in all 5 readfish samples presented here as well as a control 35x WGS sample from Jain et. al. This would require approximately 5 - 10 equivalent flow cells to generate. Green are NA12878, Orange are NB4, Blue is NA12878 without readfish applied. Samples are shown twice. Once with absolute values and once normalised per gigabase total yield from the flow cell. Run 4 == NB4 Run 1, Run 5 == NB4 Run 2.



Supplementary Figure 18: COSMIC panel against NB4 cells Run 1. A) Histogram of read batch size throughout the selective sequencing program. B) Histogram of decision times (time to choose unblock, stop receiving, or proceed from an alignment). C) Counts of decision classifications for read fragments seen a given number of times. D) Mean batch size, in bins of 2000, seen throughout the selective sequencing program. E) Mean process time, in bins of 2000, for batches of read fragments throughout the run. F) Mean decision time per read fragment, in bins of 2000, throughout the run. As the number of reads in a batch reduces, the overhead time of calling becomes more apparent.



Supplementary Figure 19: COSMIC panel against NB4 cells Run 2. A) Histogram of read batch size throughout the selective sequencing program. B) Histogram of decision times (time to choose unblock, stop receiving, or proceed from an alignment). C) Counts of decision classifications for read fragments seen a given number of times. D) Mean batch size, in bins of 2000, seen throughout the selective sequencing program. E) Mean process time, in bins of 2000, for batches of read fragments throughout the run. F) Mean decision time per read fragment, in bins of 2000, throughout the run. As the number of reads in a batch reduces, the overhead time of calling becomes more apparent.

Supplementary Tables

Duration(s)	Number of Aligned Reads	Proportion of Reads Aligning	Mean Read Length	Std Read Length	Species
all	44,930	99.83	36,651	67,827	human
0.4	5,813	12.92	143	19	human
0.5	16,249	36.1	171	22	human
0.6	25,972	57.71	198	28	human
0.8	37,026	82.27	260	40	human
1.0	40,494	89.97	326	52	human
1.2	42,045	93.42	392	67	human
1.4	42,951	95.43	457	84	human
1.5	43,106	95.78	490	92	human
1.6	43,338	96.29	522	102	human
1.8	43,782	97.28	586	122	human
2.0	43,951	97.66	649	143	human
all	34,106	99.74	12,463	14,746	zymo
0.4	14,138	41.35	127	25	zymo
0.5	20,905	61.13	154	33	zymo
0.6	24,328	71.14	185	40	zymo
0.8	28,535	83.45	250	55	zymo
1.0	30,517	89.24	317	69	zymo
1.2	30,999	90.65	387	81	zymo
1.4	31,476	92.05	456	96	zymo
1.5	31,686	92.66	490	103	zymo
1.6	31,973	93.5	523	112	zymo
1.8	32,448	94.89	590	129	zymo
2.0	32,725	95.7	656	147	zymo

Supplementary Table 1: Simulation of base calling performance from real read data sets, Progressively larger chunks of individual reads are base called and the resultant sequence mapped to the reference. By 0.8 seconds >80% of reads can be mapped. Given that a fixed chunk size is required during a run, 0.4 seconds allows us to map most reads within 1.6 seconds (see supplementary figure 2).

Input Sample		NA12878	Zymo	Zymo	NA12878	NA12878	NA12878	NA12878	NB4	NB4
Experiment		ReadFish Target	ReadFish Align	ReadFish Centrifuge	ReadFish Target	ReadFish Target	ReadFish Target	ReadFish Target	ReadFish Target	ReadFish Target
Goal		Chromosome Enrichment	40x Coverage	50x Coverage Species Blind	Exon Enrichment	COSMIC Panel	COSMIC Panel	COSMIC Panel	COSMIC Panel	COSMIC Panel
Platform		GridION MK1	GridION MK1	GridION MK1	GridION MK1	GridION MK1	GridION MK1	MinION NVIDIA GeForceGTX 180	GridION MK1	GridION MK1
Experiment ID		ml_014	ml007_zymo_gradual_reject_40x_hac	ml_019	ml_031	ml_032	ML_050	MHC_COSMIC_GUPPY_1080	ml_036	ML_048
Figure		1	2	3	4	5	5	5	6	6
All Data	**Read Count**	1,326,571	3,540,936	2,774,840	4,344,288	13,123,425	21,485,220	9,423,495	10,589,653	19,912,178
	Yield	9,580,744,050	4,415,735,206	5,995,439,489	6,635,492,653	10,034,236,814	18,269,535,376	6,704,644,049	7,696,318,846	16,289,418,025
	Mean Length	7,222	1,247	2,161	1,527	765	850	712	727	818
	N50	23,809	1,544	8,268	8,949	855	789	799	732	743
Accepted Reads	**Read Count**	591,458	495,164	478,349	607,000	636,138	1,102,859	491,653	717,017	973,488
	Yield	9,157,152,464	2,221,500,991	3,686,701,649	4,731,092,828	3,720,138,912	5,093,767,445	2,356,363,784	2,355,790,882	4,179,467,809
	Mean Length	15,482	4,486	7,707	7,794	5,848	4,619	4,793	3,286	4,293
	N50	24,691	21,408	22,704	11,641	11,191	8,044	8,180	7,747	7,117
Rejected Reads	**Read Count**	735,113	3,045,772	2,296,491	3,737,288	12,487,287	20,382,361	8,931,842	9,872,636	18,938,690
	Yield	423,591,586	2,194,234,215	2,308,737,840	1,904,399,825	6,314,097,902	13,175,767,931	4,348,280,265	5,340,527,964	12,109,950,216
	Mean Length	576	720	1,005	510	506	646	487	541	639
	N50	575	736	1,079	500	501	634	464	538	635
Total Unblock Events		85,999	3,123,817	2,314,851	3,857,980	12,879,558	20,712,117	9,626,831	10,151,972	19,187,172

Supplementary Table 2: Run statistics for all sequencing runs presented here. Statistics calculated using NanoStat. Run data split into accepted and rejected reads on the basis of unblocked read logs from ReadFish.

	Experiment	Basecaller	Coverage	Recall	Precision	F1
	WGS	Guppy 2.3.8	35 x	0.658011	0.289817	0.402399
	WGS	Guppy 3.6	35 x	0.729267	0.447407	0.554579
	Run 1 COSMIC	Guppy 3.4.5	30.7 x	0.770934	0.3633	0.493867
	Run 2 COSMIC	Guppy 3.4.5	40.5 x	0.779791	0.303602	0.437046
INDELs	Run 1 + 2	Guppy 3.4.5	71.2 x	0.83434	0.326508	0.469344
SNPs	WGS	Guppy 2.3.8	35 x	0.846758	0.919244	0.881513

WGS	Guppy 3.6	35 x	0.902159	0.935991	0.918764
Run 1 COSMIC	Guppy 3.4.5	30.7 x	0.949963	0.919445	0.934455
Run 2 COSMIC	Guppy 3.4.5	40.5 x	0.974226	0.928952	0.95105
Run 1 + 2	Guppy 3.4.5	71.2 x	0.985983	0.924854	0.954441

Supplementary Table 3: SNP concordance versus NA12878 truth set for Whole Genome Sequence (WGS) data at 35x from Jain et. al. 2018, Run 1 and Run 2 as well as both runs combined. WGS data were recalled using Guppy 3.6 for comparison.

				Deletions			Insertions		
Run	**Precision**	**Recall**	**F1**	**Mean Length**	**SD Length**	**Count**	**Mean Length**	**SD Length**	**Count**
Run1 COSMIC	0.87	0.89	0.88	326.09	646.86	138	189.59	127.89	189
Run2 COSMIC	0.86	0.92	0.89	276.88	425.29	136	182.67	121.3	204
WGS	1	1	1	319.21	653.7	135	185.24	119.11	186

Supplementary Table 4: SV concordance between readfish runs and a 35x Whole Genome Sequence from Jain et. al. 2018. WGS data were recalled using Guppy 3.6 for this comparison. Run1 had mean coverage of 30.7 x over the COSMIC targets. Run 2 had mean coverage of 40.5 over the COSMIC targets.

46
47

Panel	Sample	Chrom	Pos	Qual	Filter	Alt	Support	STD POS1	STD POS2
E	NA12878	chr1	16002843	7	PASS	]chr9:111642651]N	6	3	1
E	NA12878	chr3	151430755	5	PASS	[chr5:39787648[N	4	.	.
E	NA12878	chr5	39787648	5	PASS	[chr3:151430755[N	4	.	.
E	NA12878	chr6	382460	7	PASS	N]chr16:33626062]	6	.	.
E	NA12878	chr7	26213350	6	PASS	N]chr15:40561994]	5	.	.
E	NA12878	chr9	42900334	6	PASS	]chr9:64124004]N	5	5	19
E	NA12878	chr9	64124004	6	PASS	N[chr9:42900334[	5	19	5
E	NA12878	chr9	111642651	7	PASS	N[chr1:16002843[	6	1	3
E	NA12878	chr15	40561994	6	PASS	N]chr7:26213350]	5	.	.
E	NA12878	chr16	33626062	7	PASS	N]chr6:382460]	6	.	.
C	NA12878 Run 1	chr11	108715020	9	PASS	]13:21176520]N	8	.	2
C	NA12878 Run 2	chr11	108715020	8	PASS	]13:21176521]N	7	.	1
C	NB4 run 1 (15hours)	chr3	136559276	7	PASS	N[chr3:136714904[	6	.	9
C	NB4 run 1 (15hours	chr3	136714904	7	PASS	]chr3:136559276]N	6	9	.
C	NB4 run 1 (15hours	chr15	74034025	7	PASS	]chr17:40345927]N	6	2	5
C	NB4 run 1 (15hours	chr17	40345927	7	PASS	N[chr15:74034025[	6	5	2
C	**NB4 run 1 (end)**	**chr3**	**136559276**	**11**	**PASS**	**N[3:136714902[**	**9**	**.**	**8**
C	**NB4 run 1 (end)**	**chr3**	**136714902**	**11**	**PASS**	**]3:136559276]N**	**9**	**8**	**.**
C	NB4 run 1 (end)	chr3	60587561	8	PASS	N[3:60724671[	7	.	3
C	NB4 run 1 (end)	chr3	60724671	8	PASS	]3:60587561]N	7	3	.

C	NB4 run 1 (end)	chr3	60561654	6	PASS	N[3:60665735[	5	3	4
C	NB4 run 1 (end)	chr3	60665735	6	PASS	]3:60561654]N	5	4	3
C	NB4 run 1 (end)	chr8	113097106	5	PASS	N[9:76575900[	4	.	.
C	NB4 run 1 (end)	chr12	27614776	6	PASS	N[11:99323700[	5	2	2
C	**NB4 run 1 (end)**	**chr15**	**74034026**	**11**	**PASS**	**]17:40345928]N**	**9**	**1**	**4**
C	**NB4 run 1 (end)**	**chr17**	**40345928**	**11**	**PASS**	**N[15:74034026[**	**9**	**4**	**1**
C	NB4 run 1 (end)	chr19	8896333	6	PASS	N[19:46928346[	5	4	2
C	**NB4 run 2 (end)**	**chr3**	**136559275**	**17**	**PASS**	**N[3:136714905[**	**14**	**1**	**8**
C	**NB4 run 2 (end)**	**chr3**	**136714905**	**17**	**PASS**	**]3:136559275]N**	**14**	**8**	**1**
C	NB4 run 2 (end)	chr3	60587561	8	PASS	N[3:60724670[	7	1	1
C	NB4 run 2 (end)	chr3	60724670	8	PASS	]3:60587561]N	7	1	1
C	NB4 run 2 (end)	chr8	113097106	8	PASS	N[9:76575902[	7	1	2
C	NB4 run 2 (end)	chr12	27605316	8	PASS	]12:23602042]N	7	3	1
C	NB4 run 2 (end)	chr12	27598225	7	PASS	N[12:28445796[	6	4	4
C	NB4 run 2 (end)	chr12	27614777	7	PASS	N[11:99323699[	6	.	.
C	NB4 run 2 (end)	chr13	39367720	7	PASS	[13:43314627[N	6	1	.
C	**NB4 run 2 (end)**	**chr14**	**68295168**	**13**	**PASS**	**N[14:68479653[**	**11**	**.**	**3**
C	**NB4 run 2 (end)**	**chr14**	**68479653**	**13**	**PASS**	**]14:68295168]N**	**11**	**3**	**.**
C	**NB4 run 2 (end)**	**chr15**	**74034026**	**28**	**PASS**	**]17:40345928]N**	**23**	**1**	**2**
C	**NB4 run 2 (end)**	**chr17**	**40345928**	**28**	**PASS**	**N[15:74034026[**	**23**	**2**	**1**
C	NB4 run 2 (end)	chr19	8896334	9	PASS	N[19:46928346[	8	1	2

Supplementary Table 5: Structural variants as identified by svim. Variant calls were filtered with the default filter pass and non BND (Breakpoint End) structural variant types are not shown in this table. Panel: E - Exon Panel, C - COSMIC Cancer Panel, Sample: NA12878 reference cells or NB4 reference cells. Chrom: Starting chromosome of the read. POS: Position of the break point with respect to chromosome in Chrom. Qual: Quality as reported by SVIM. Filter: Default SVIM filter (>Q5). SVs with Q>10 indicated by bold text. Alt: Description of the second break point in the structural variant. Support: Number of reads supporting the observed structural variant. STD POS 1/2: The standard deviation in position of the breakpoint.

Supplementary Files

Supplementary File 1 - COSMIC_Coverage.csv - File containing coordinates used for selective sequencing of the COSMIC cancer panel along with the observed mean coverage for each run, gene name and span of the region being enriched for. All coordinates with respect to hg38.

Supplementary Notes

Supplementary Note 1 - ***Note on compute requirements***

Oxford Nanopore Technologies maintains an IT requirements document that outlines the minimal requirements for sequencing with a MinION, currently available at: https://community.nanoporetech.com/requirements_documents/minion-it-reqs.pdf. We aim to make our toolkit work with this configuration in mind. In addition, to make use of GPU accelerated base calling a CUDA enabled NVIDIA GPU is required with a compute capability of >6.1, more information can be found at https://developer.nvidia.com/cuda-gpus#compute. At the time of writing ONT supports the following GPUs: Tesla V100, Quadro GV100, GTX 1080 Ti, and Jetson TX2. Of these, only the Jetson TX2 is unable to perform ReadFish experiments with our toolkit.