UMI downsampling QC reports

First, a qualitative view… UMI does work!

263R

263R experimental details

Sample				Pre-PCR			Post-PCR		Pooling prior to hybridization	
Name	size	ng/uL	Fetal fraction	WellId	Input	Theoretical max coverage	Avg. Size [bp]	Conc. [ng/µl]	Mass added to pre-hyb pool (ng)	Moles added to pre-hyb pool
266R01				A01						
266R02				B01						
266R03				C01						
266R04				D01						
266R05				E01						
266R06				F01						
266R07				G01						
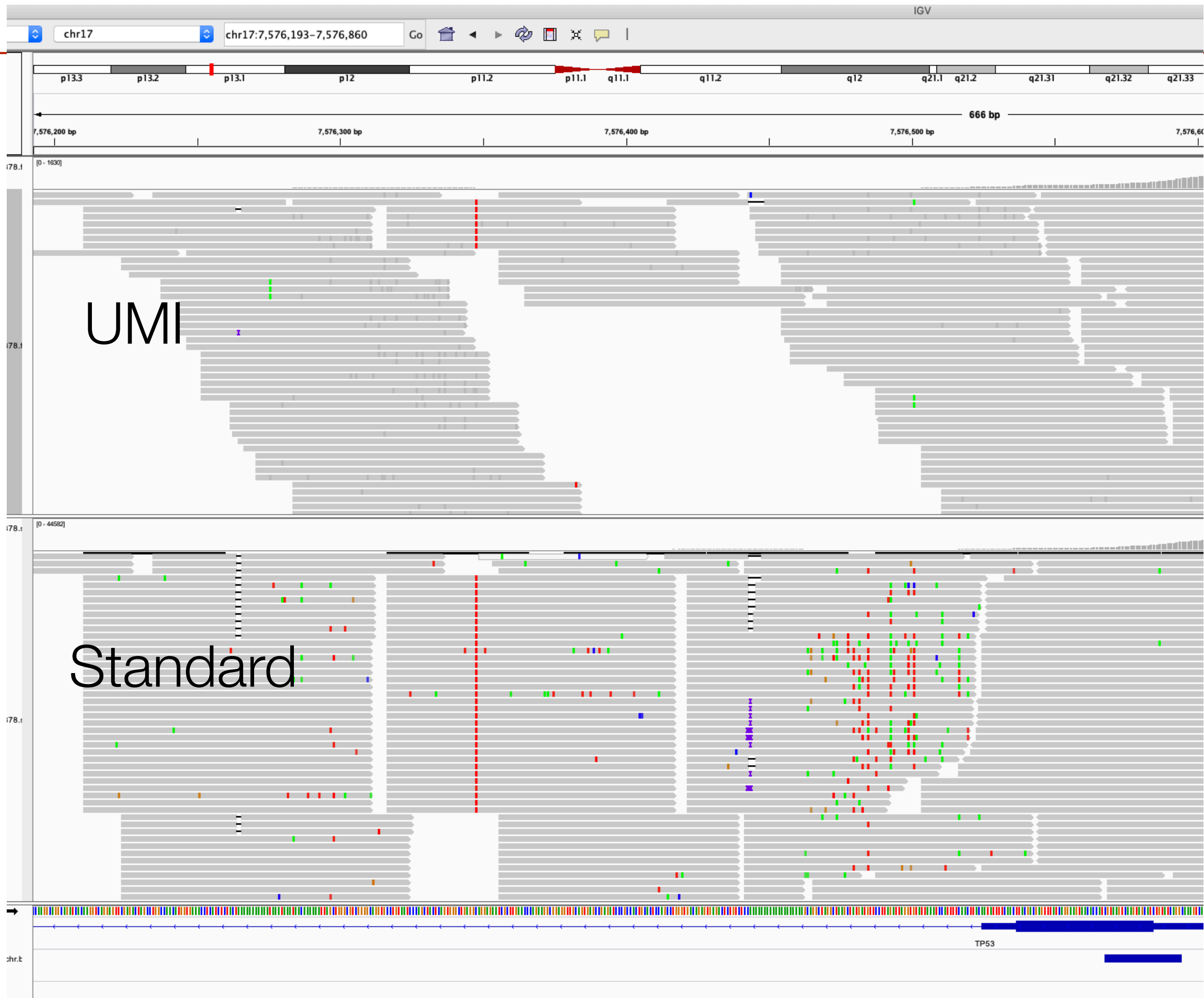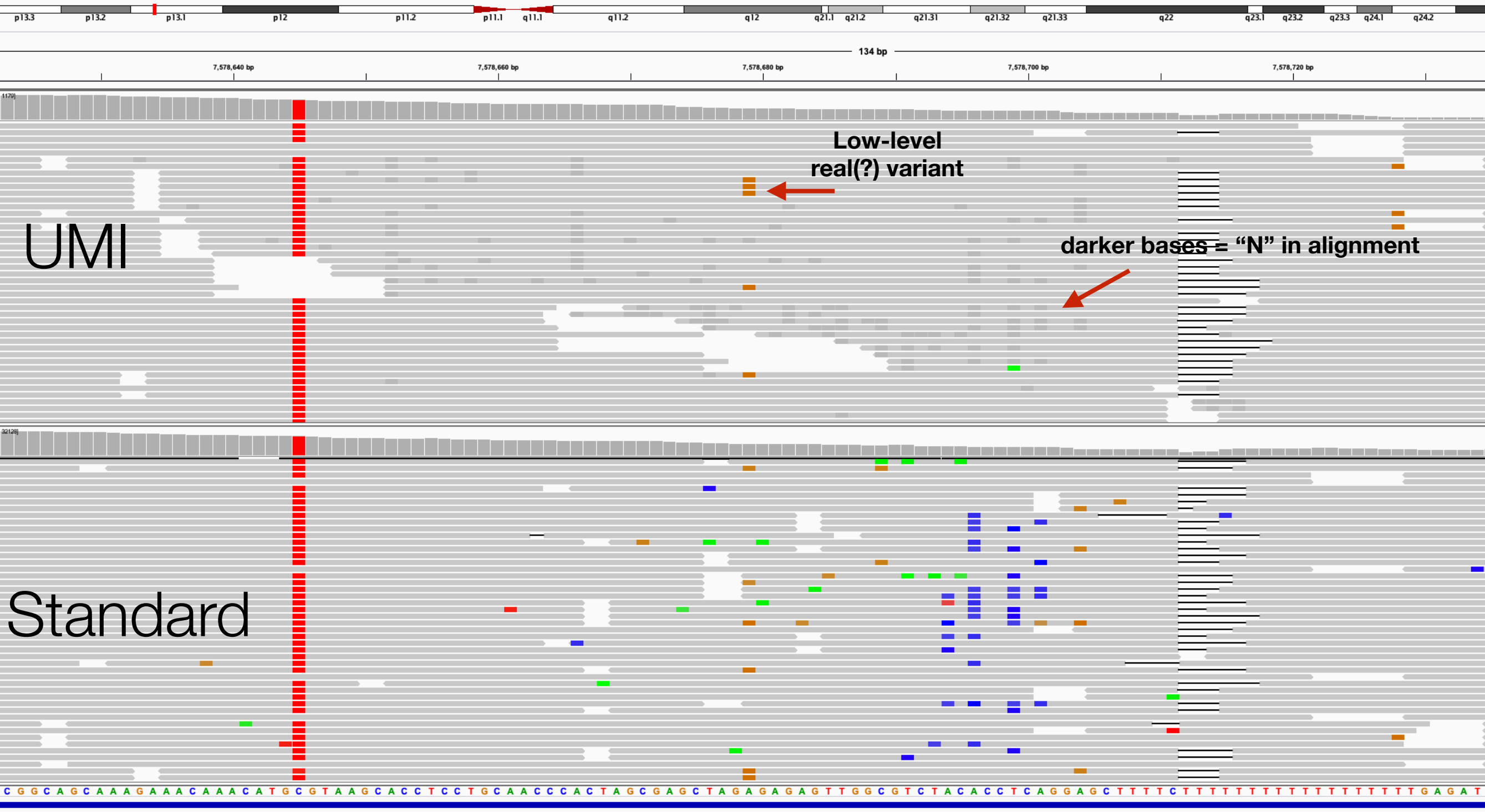
263R stats

Table 1

Sample Name		Total reads		Median Target Coverage		% on bait bases		% on target bases		% Bases >100x		% UMI ≥2
		standard	final	standard	final	standard	final	standard	final	standard	final	
263R01		50,000,000	7,926,802	12,513	2,454	79%	68%	41%	33%	100%	99.9%	25%
263R02		50,000,000	10,050,168	12,482	3,029	79%	69%	39%	33%	100%	99.9%	39%
263R03		50,000,000	7,223,314	12,390	2,116	77%	69%	38%	33%	100%	99.8%	21%
263R04		50,000,000	7,900,734	12,330	2,472	79%	67%	41%	33%	100%	99.9%	27%
263R05		50,000,000	8,276,458	11,871	2,432	77%	66%	39%	32%	100%	100%	28%
263R06		50,000,000	7,083,300	12,104	2,184	77%	64%	41%	33%	100%	100%	24%
263R07		26,327,000	5,655,734	6,392	1,571	75%	68%	37%	33%	100%	100%	42%

Table 1-1

Means (w/o R07)		50,000,000	8,076,796	12,282	2,448	78%	67%	40%	33%	100%	100%	27%

▸ R07 failed — note poor insert size distribution + low read count

263 UMI family size

fgbio: Family size count

263R insert sizes

Picard: Insert Size
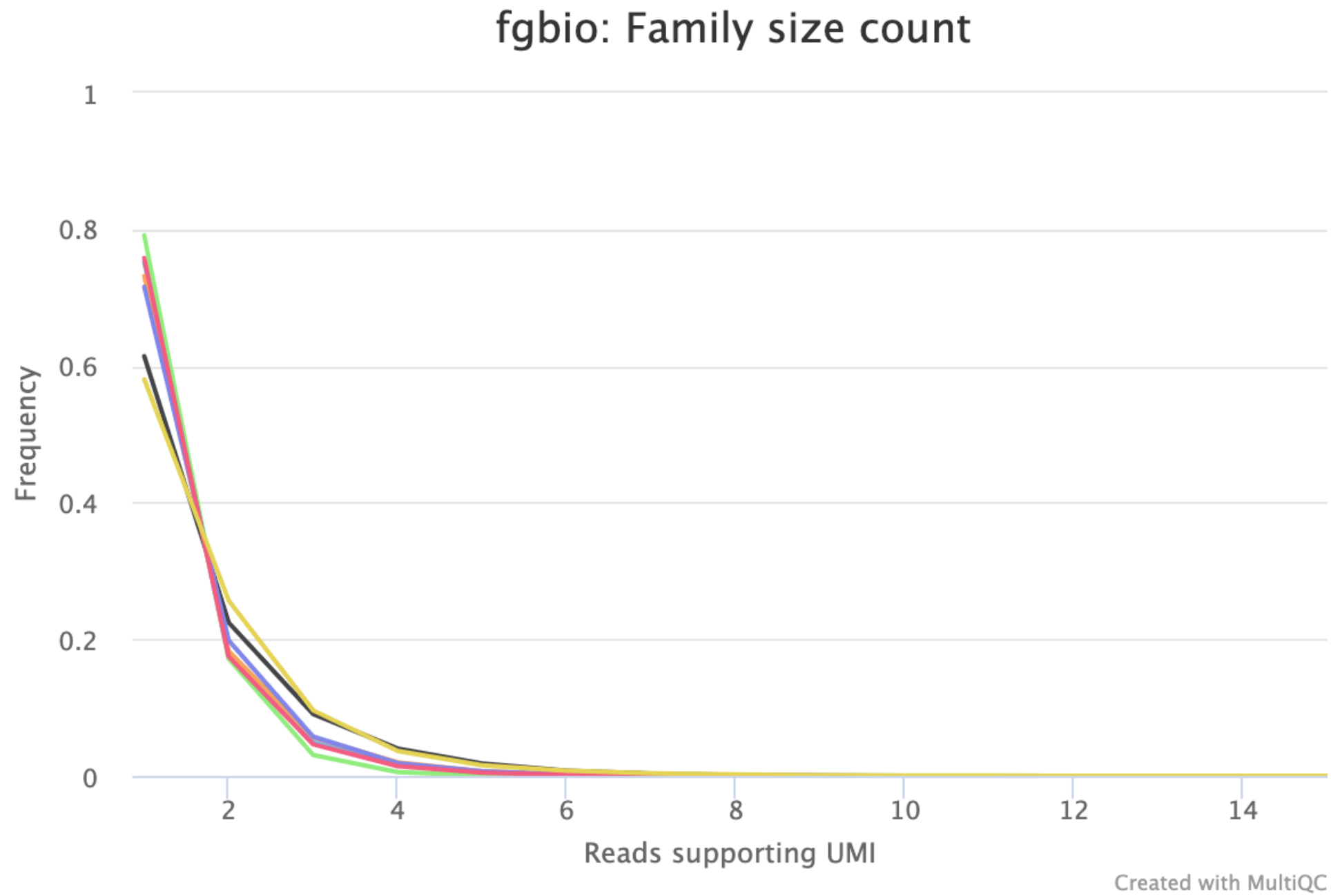
Coverage of standard + final bam
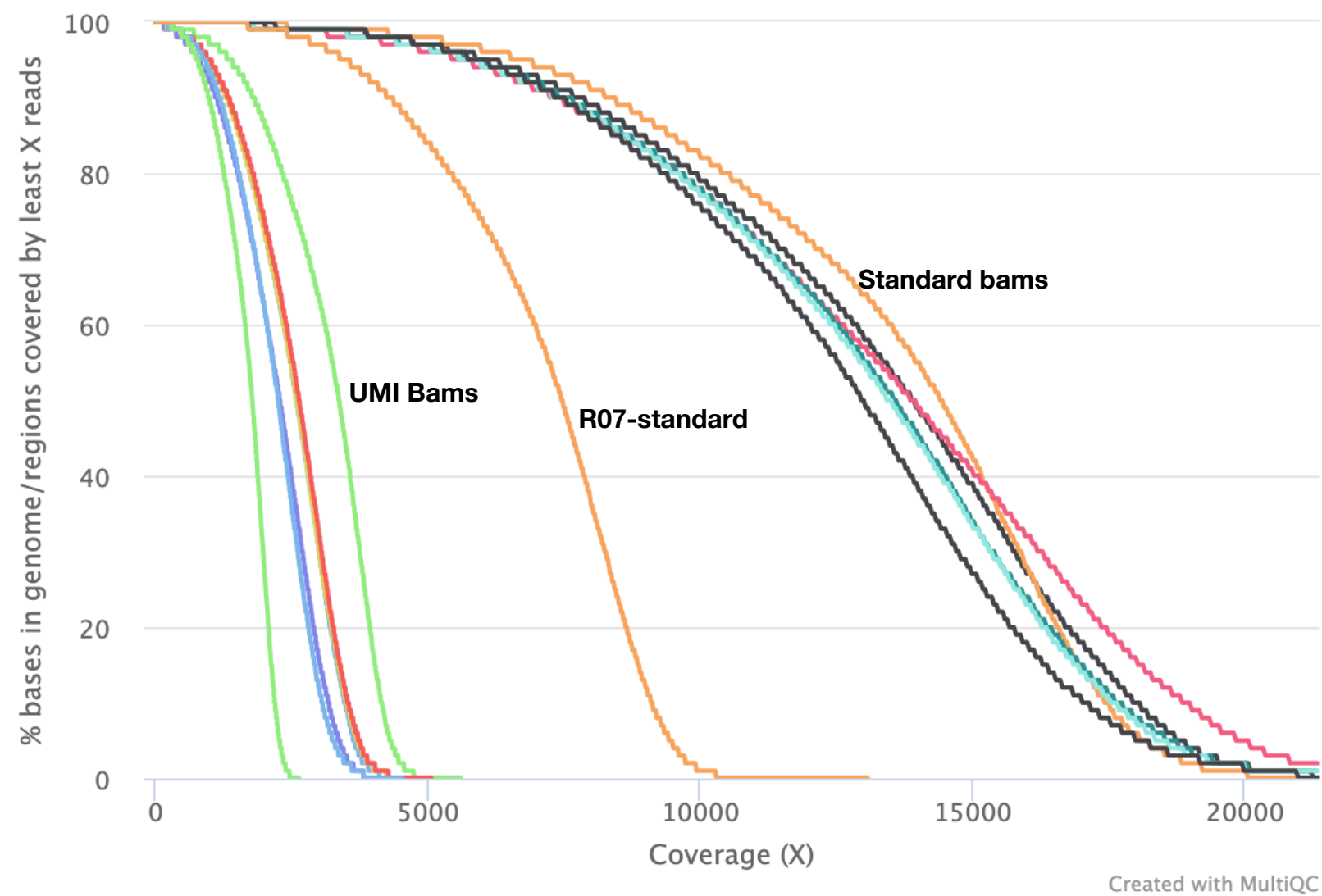
Important!

263 vs 266

Table 1

Sample		Total reads		Median Target Coverage		% on bait bases		% on target bases		% Bases >100x		% UMI ≥2	# singleton reads	# reads in 2+ Families
		standard	final	standard	final	standard	final	standard	final	standard	final			
263R01	1:1	50,000,000	7,926,802	12,513	2,454	79%	68%	41%	33%	100%	99.9%	25%	5,945,102	1,981,701
263R02	1:1	50,000,000	10,050,168	12,482	3,029	79%	69%	39%	33%	100%	99.9%	39%	6,130,602	3,919,566
263R03	1:1	50,000,000	7,223,314	12,390	2,116	77%	69%	38%	33%	100%	99.8%	21%	5,706,418	1,516,896
263R04	1:1	50,000,000	7,900,734	12,330	2,472	79%	67%	41%	33%	100%	99.9%	27%	5,767,536	2,133,198
263R05	1:1	50,000,000	8,276,458	11,871	2,432	77%	66%	39%	32%	100%	100%	28%	5,959,050	2,317,408
263R06	1:1	50,000,000	7,083,300	12,104	2,184	77%	64%	41%	33%	100%	100%	24%	5,383,308	1,699,992
266R01	1:5	50,000,000	2,021,550	13,477	469	73%	63%	36%	31%	100%	99%	57%	869,267	1,152,284
266R02	1:10	50,000,000	1,291,048	13,597	251	74%	53%	37%	26%	100%	98%	73%	348,583	942,465
266R03	1:15	50,000,000	1,072,306	13,481	192	74%	49%	37%	24%	100%	96%	79%	225,184	847,122
266R04	1:20	50,000,000	720,616	13,537	98	74%	38%	37%	18%	100%	47%	71%	208,979	511,637
266R05	1:50	50,000,000	516,344	13,222	60	74%	33%	37%	16%	100%	3%	82%	92,942	423,402
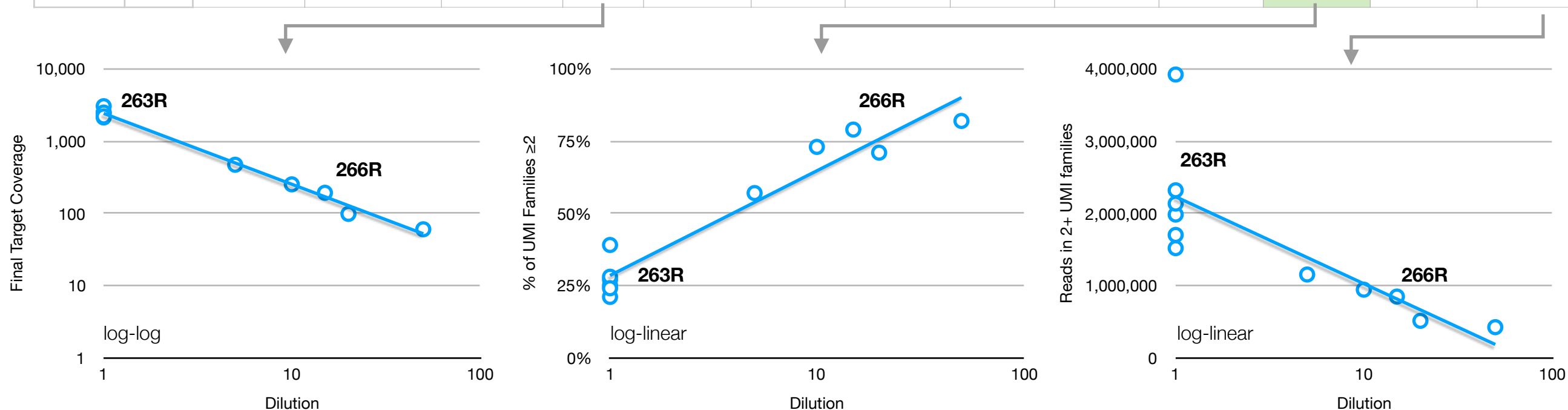
266R

266R

Sample				Pre-PCR			Post-PCR		Pooling prior to hybridization	
Name	size	ng/uL	Fetal fraction	WellId	Input	Theoretical max coverage	Avg. Size [bp]	Conc. [ng/µl]	Mass added to pre-hyb pool (ng)	Moles added to pre-hyb pool
266R01	188	0.194ng/uL	—	A7	**1:5**	400x	361	11.9	125	5.33E-13
266R02	188	0.194ng/uL	—	B7	**1:10**	193x	374	29.9	125	5.14E-13
266R03	188	0.194ng/uL	—	C7	**1:15**	130x	372	33.5	125	5.17E-13
266R04	188	0.194ng/uL	—	D7	**1:20**	96.7x	370	30.6	125	5.2E-13
266R05	188	0.194ng/uL	—	E7	**1:50**	40x	367	31.9	125	5.24E-13

- Limit the amount of cell-free DNA carried forward from adaptor ligation into pre-capture PCR, to:

    ◦ hopefully increase the number of duplicate reads

    ◦ determine which dilution would be most effective in maximizing the fraction of UMI families with a family size of at least 2.

◦ Five remnant cell-free DNAs from the PDNAS assay (due to having few to zero flagged bins and no detected aneuploidies), pooled together.

◦ Pool-level quality control
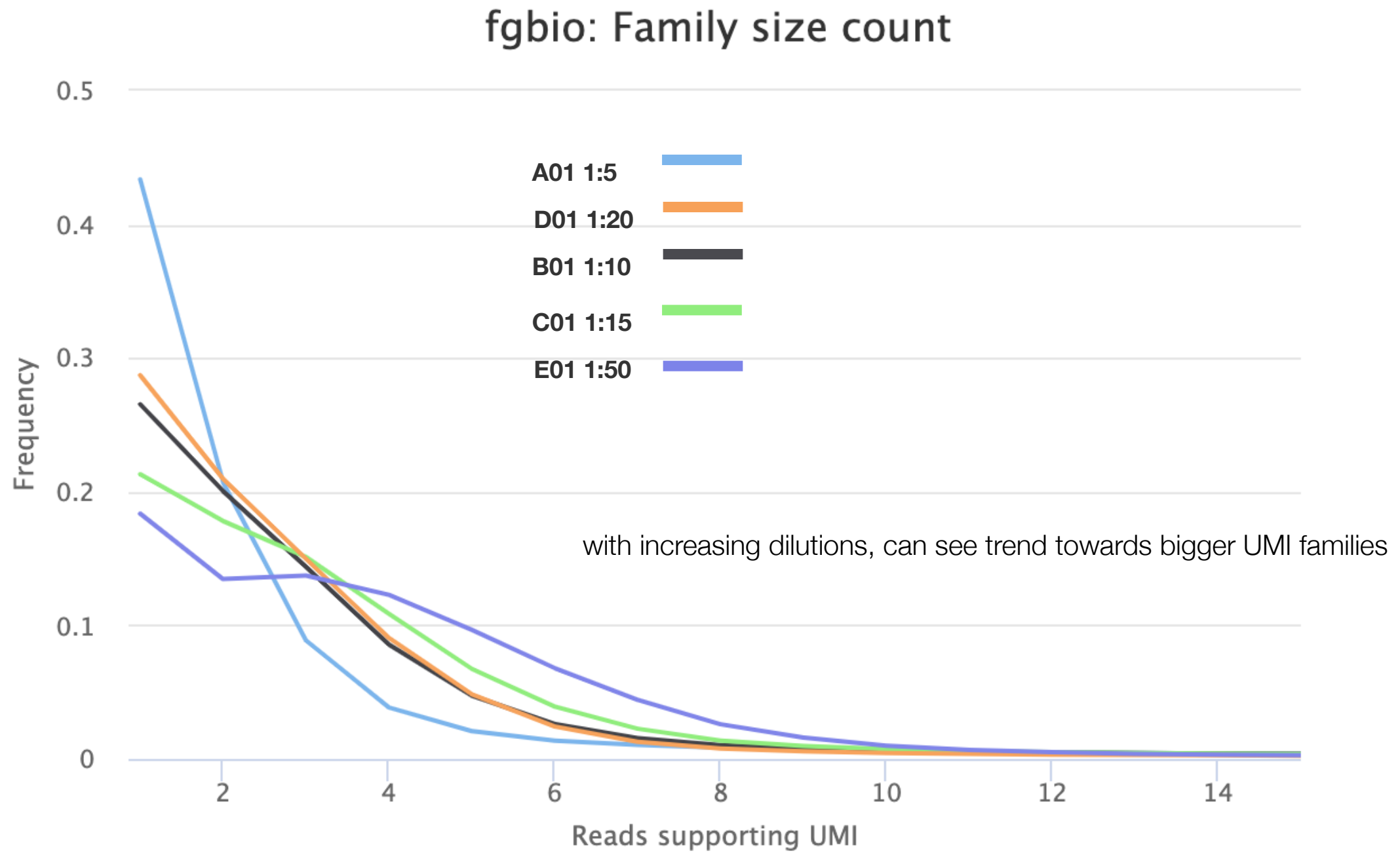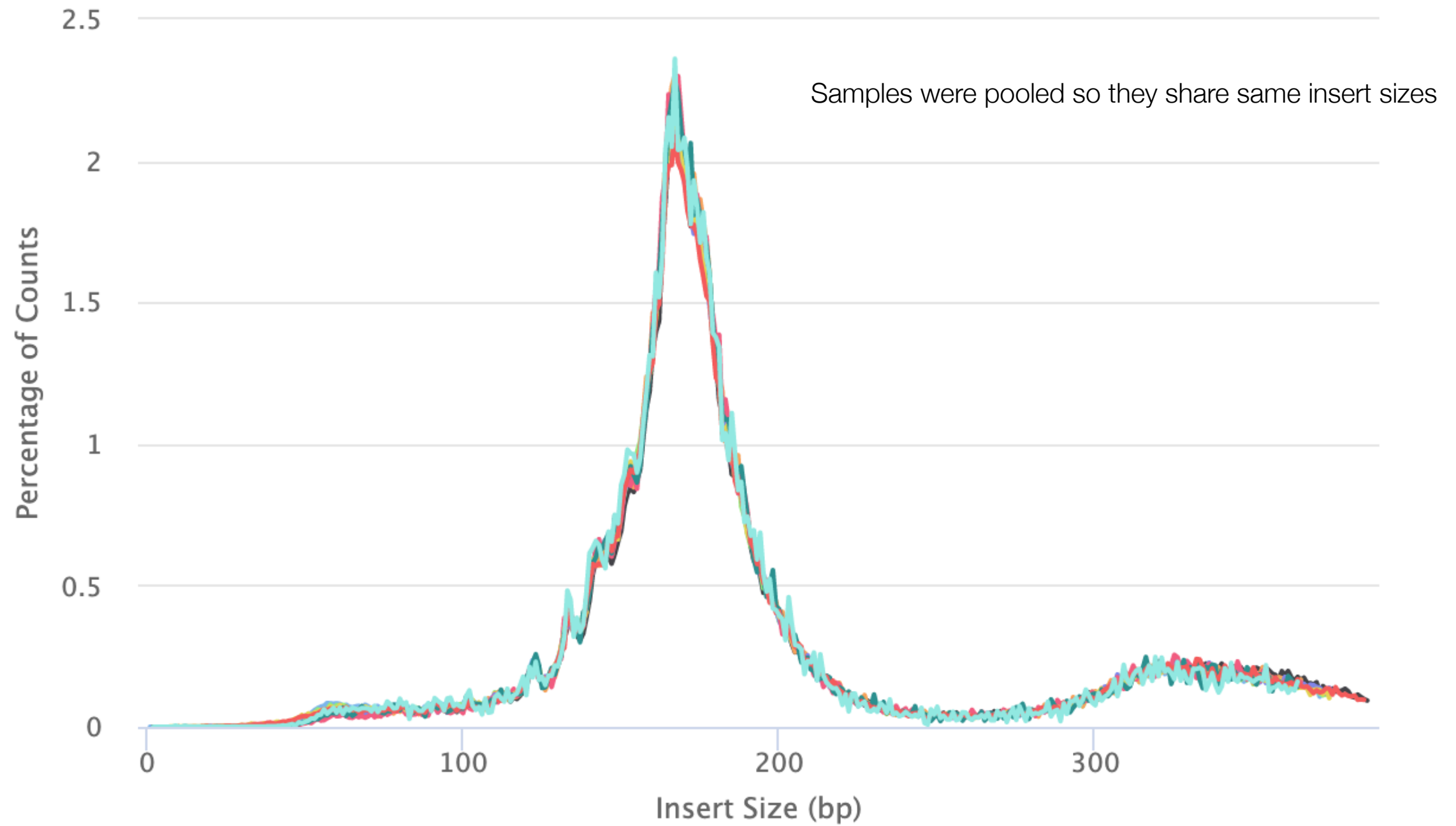   (tape station): 0.194ng/ul = 5.82ng in 30ul

266R

Table 1

Sample Name		Total reads		Median Target Coverage		% on bait bases		% on target bases		% Bases >100x		% UMI ≥2
		standard	final	standard	final	standard	final	standard	final	standard	final	
266R01	1:5	50,000,000	2,021,550	13,477	469	72.7%	62.5%	36.5%	30.9%	100.0%	99.4%	57%
266R02	1:10	50,000,000	1,291,048	13,597	251	73.5%	52.8%	36.7%	26.0%	100.0%	97.8%	73%
266R03	1:15	50,000,000	1,072,306	13,481	192	73.5%	48.9%	36.7%	24.2%	100.0%	96.4%	79%
266R04	1:20	50,000,000	720,616	13,537	98	74.3%	37.7%	37.1%	18.4%	100.0%	47.4%	71%
266R05	1:50	50,000,000	516,344	13,222	60	74.4%	32.7%	37.0%	16.0%	100.0%	3.0%	82%

Table 1-1

263R Mean	1:1	50,000,000	8,076,796	12,282	2,448	78%	67%	40%	33%	100%	100%	27%

- ▸ All FASTQs downsampled to have 25m read pairs

- ▸ With increasing dilution:

  - ▸ Fewer UMI families (lower final total reads)

  - ▸ Lower median target coverage, lower % on bait/target bases

  - ▸ Above 1:20, significant loss of % bases with >100x coverage

- ▸ Our bait-to-target efficiency is about 50%

- ▸ 266R01-A01 has lower percentage of UMI families with ≥2 reads, however no clear trend for this value in other dilutions?
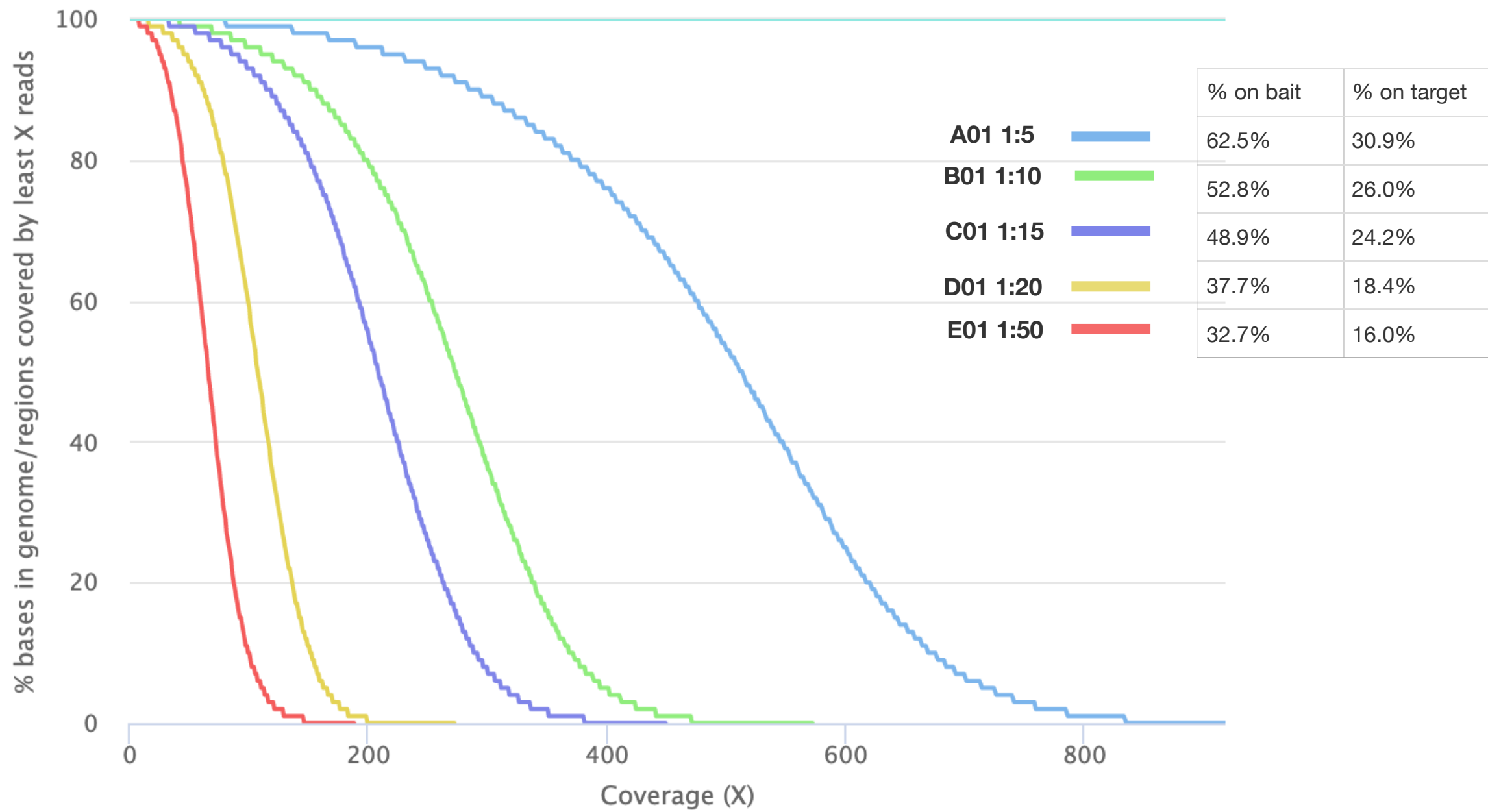
266R UMI family sizes

fgbio: Family size count

with increasing dilutions, can see trend towards bigger UMI families

Created with MultiQC

266R insert size

Samples were pooled so they share same insert sizes

Created with MultiQC

Coverage of final bam

	% on bait	% on target
A01 1:5	62.5%	30.9%
B01 1:10	52.8%	26.0%
C01 1:15	48.9%	24.2%
D01 1:20	37.7%	18.4%
E01 1:50	32.7%	16.0%

Created with MultiQC

280R

280R

Sample				Pre-PCR			Post-PCR		Pooling prior to hybridization	
Name	size	ng/uL	Fetal fraction	WellId	Input mass (ng)	Theoretical max coverage	Avg. Size [bp]	Conc. [ng/µl]	Mass added to pre-hyb pool (ng)	Moles added to pre-hyb pool
210P03	182	0.763	0.14 (XX)	A2	**5ng**	1667x	420	1.48	125	4.58E-13
210P03	182	0.763	0.14 (XX)	D2	**1ng**	333x	369	29.5	125	5.21E-13
211P12	198	0.725	0.056 (XY)	B2	**5ng**	1667x	377	26.7	125	5.10E-13
211P12	198	0.725	0.056 (XY)	E2	**1ng**	333x	335	22.1	125	5.74E-13
222P09	194	0.441	0.076 (XY)	C2	**5ng**	1667x	366	28.5	125	5.25E-13
222P09	194	0.441	0.076 (XY)	F2	**1ng**	333x	388	37.3	125	4.96E-13

- Repeat the best conditions from NGS266R (1/5 ligation into PCR)

- See whether approximately 1/5 of the input mass with the normal 20ul into PCR would produce similar results.

- Three remnant cell-free DNAs from the PDNAS assay were selected because they had enough leftover for two library preps (5ng, and 1ng). These were also normal samples with few flagged bins.
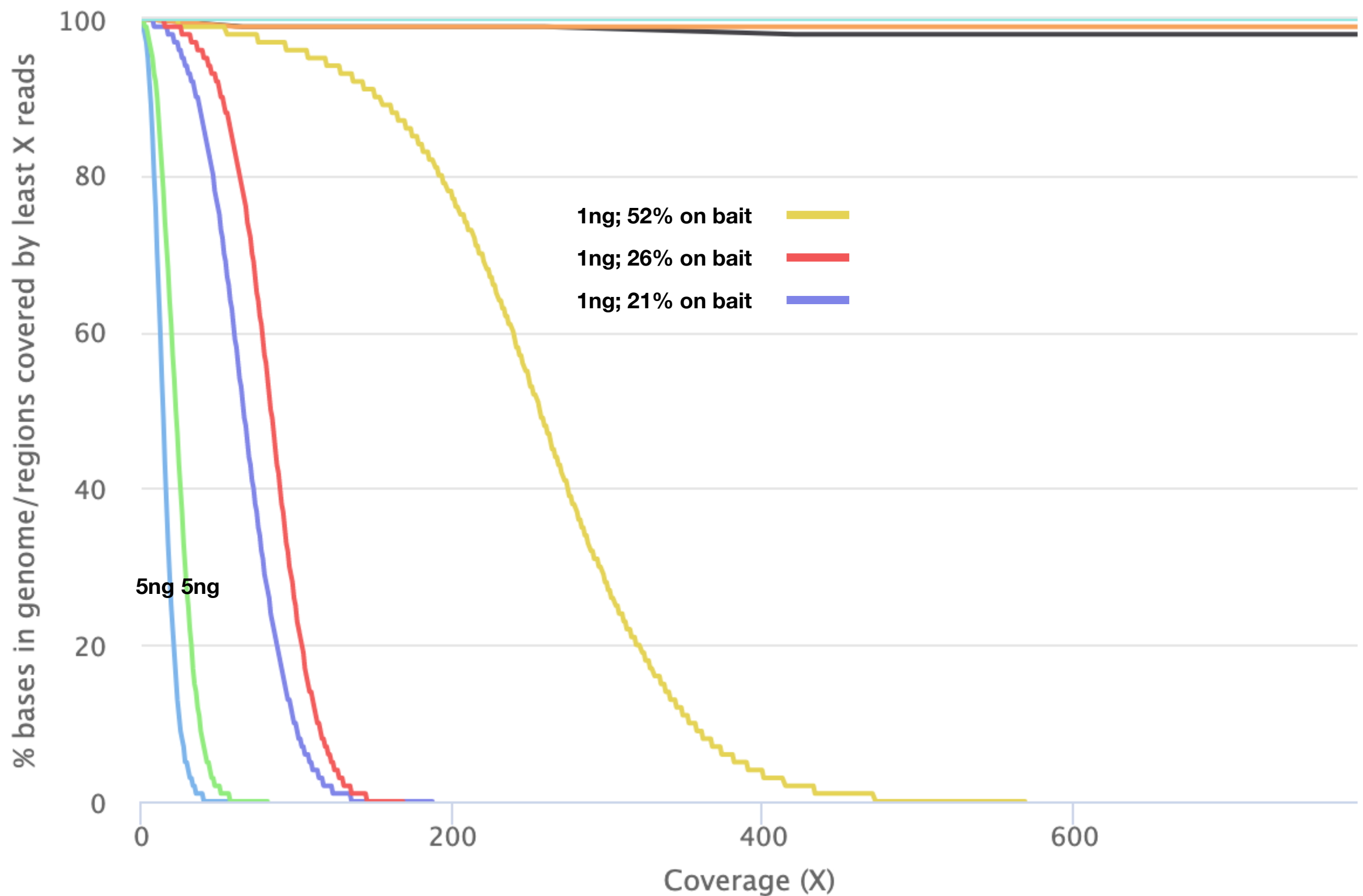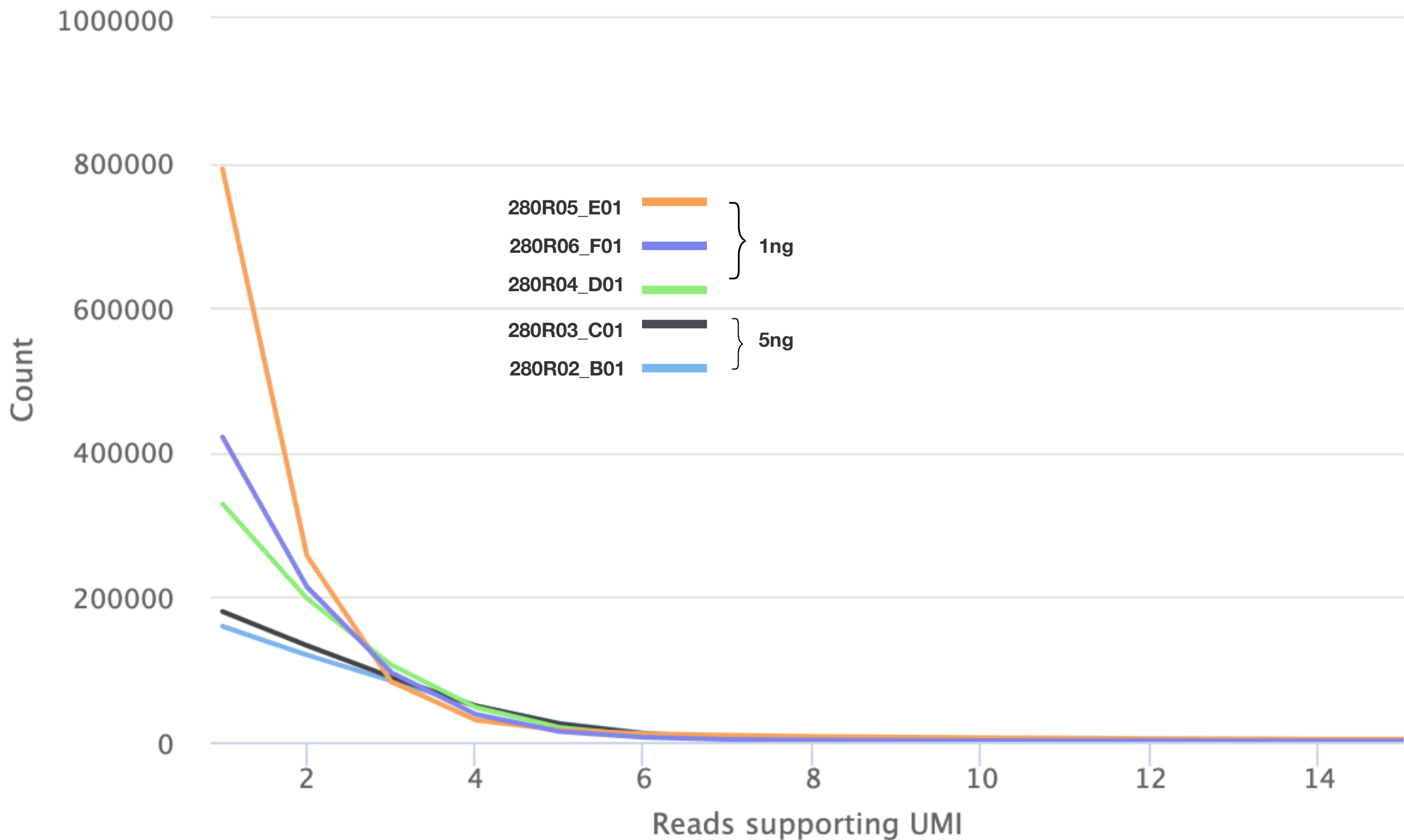
280R

Table 1

Sample Name			Total reads		Median Target Coverage		% on bait bases		% on target bases		% Bases >100x	
			standard	final	standard	final	standard	final	standard	final	standard	final
280R01_A01	210P03	5ng	50,000,000	1,236,226	7	1	67.4%	1.7%	34.0%	0.9%	44.8%	0.0%
280R04_D01	210P03	1ng	50,000,000	593,438	12,643	**60**	70.7%	**21.7%**	35.0%	**10.6%**	99.9%	**2.5%**
280R02_B01	211P12	5ng	50,000,000	816,094	12,333	12	71.4%	6.5%	35.1%	3.1%	99.3%	0.0%
280R05_E01	211P12	1ng	50,000,000	547,794	12,789	219	74.6%	**52.0%**	34.9%	23.9%	99.9%	97.4%
280R03_C01	222P09	5ng	50,000,000	772,752	12,469	20	71.6%	9.6%	35.5%	4.7%	99.5%	0.0%
280R06_F01	222P09	1ng	50,000,000	343,036	13,085	76	70.6%	**25.6%**	35.7%	12.6%	100.0%	10.0%

‣ All FASTQs downsampled to have 25m read pairs

‣ A01 failed — disregard; note that it had normal *mean* coverage; only median target coverage is low.

‣ Metrics on standard bam do not predict final bam metrics.

‣ Comparing 5ng to 1ng:

  ‣ 1ng has MORE on-bait and on-target bases and GREATER median target coverage.

  ‣ BUT 1ng has FEWER reads in final/collapsed bam!

  ‣ **did 5ng and 1ng get switched?**

‣ **These results look like dilutions of 1:20 and 1:50, but don't follow the 263/266 trends**
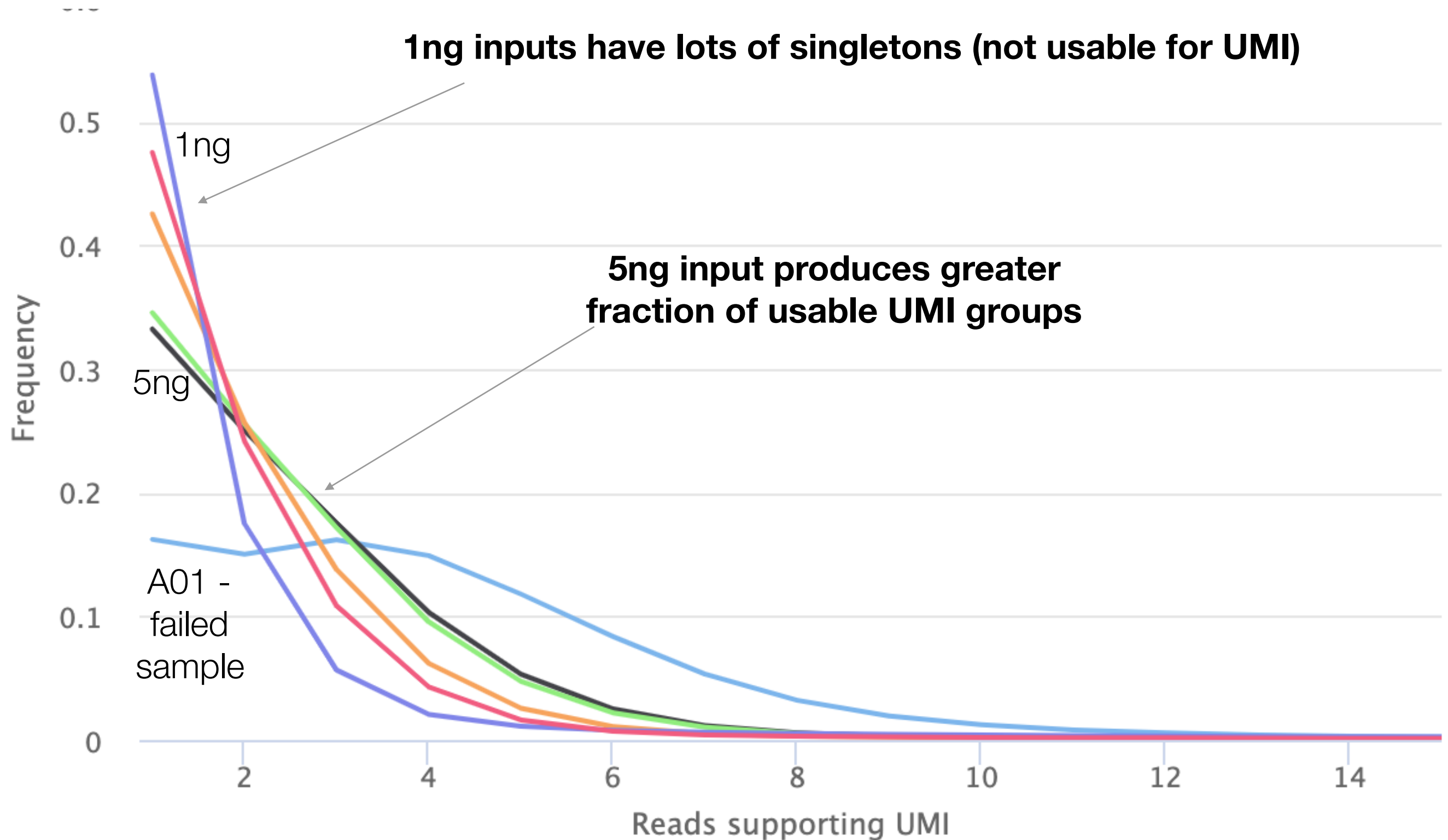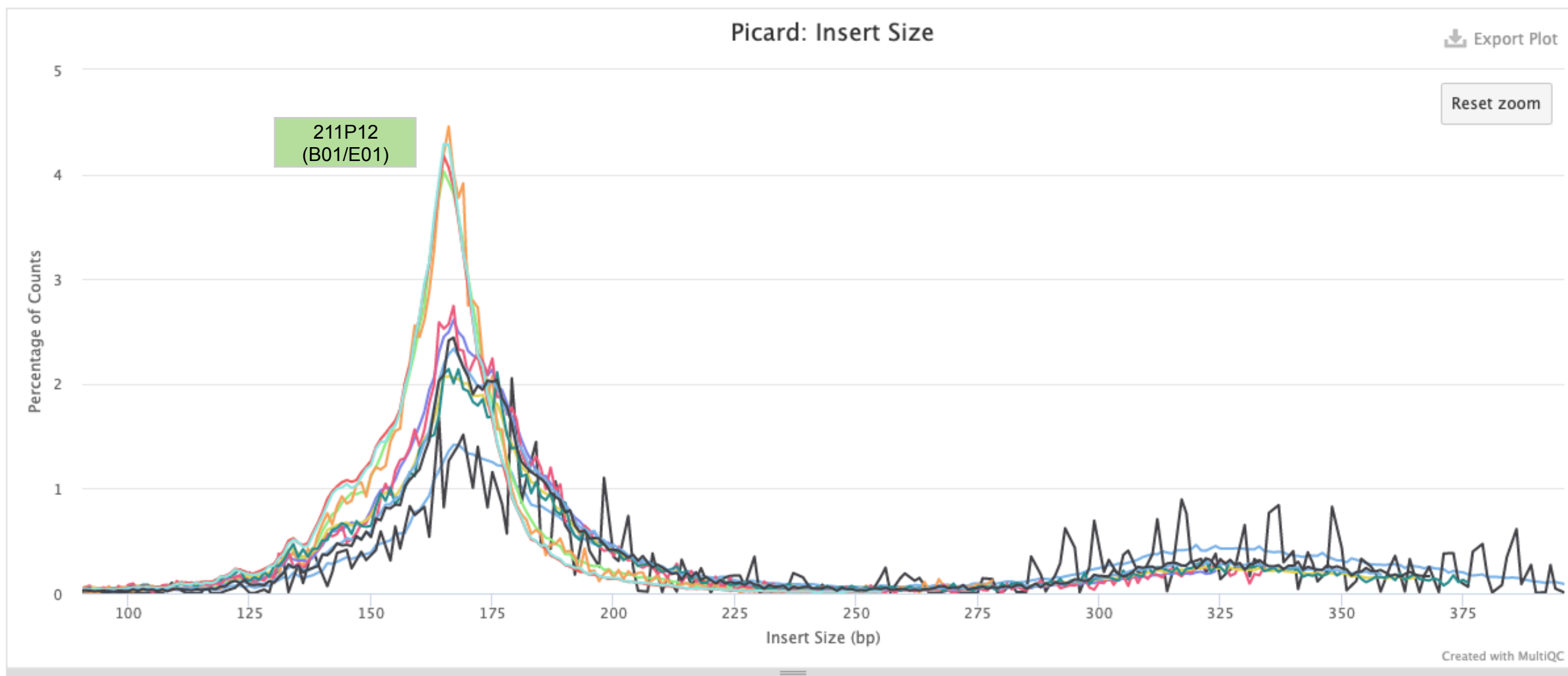
Coverage of final bam

fgbio: Family size count

280R05_E01
280R06_F01   } 1ng
280R04_D01

280R03_C01   } 5ng
280R02_B01

Count

Reads supporting UMI

Created with MultiQC

UMI Histogram - percentages

1ng inputs have lots of singletons (not usable for UMI)

1ng

5ng input produces greater fraction of usable UMI groups

5ng

A01 - failed sample

Frequency

Reads supporting UMI

Created with MultiQC

280R Insert Size

Insert Size

Plot shows the number of reads at a given insert size. Reads with different orientations are summed.

Y-Limits: ● on

Counts | Percentages

Picard: Insert Size

211P12 (B01/E01)

Percentage of Counts

Insert Size (bp)

Created with MultiQC

Export Plot

Reset zoom

211P12 (B01/E01) has slightly smaller insert size than other two samples.
(small bimodal peak seen on post-hyb TapeStation as well)
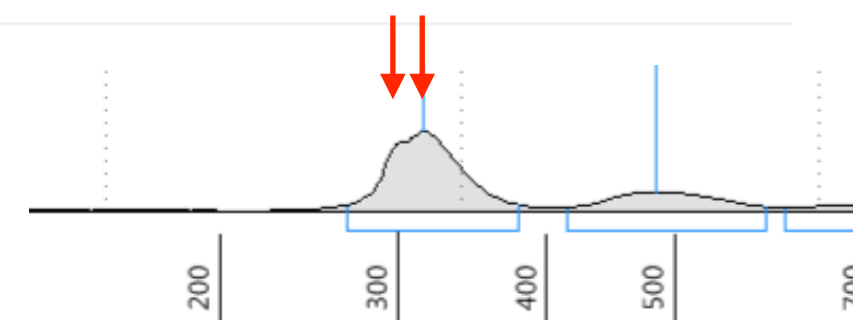
Comparing 266 vs 280

▸ 266R01 (1:5) should be comparable to 5ng condition of 280R? Very different numbers. What happened? **did 5ng and 1ng get switched??**

Table 1-1

Sample Name		Total reads		Median Target Coverage		% on bait bases		% on target bases		% Bases >100x	
		standard	final	standard	final	standard	final	standard	final	standard	final
266R01	**1:5**	50,000,000	2,021,550	13,477	469	72.7%	62.5%	36.5%	30.9%	100.0%	99.4%
266R02	**1:10**	50,000,000	1,291,048	13,597	251	73.5%	52.8%	36.7%	26.0%	100.0%	97.8%
266R03	**1:15**	50,000,000	1,072,306	13,481	192	73.5%	48.9%	36.7%	24.2%	100.0%	96.4%
266R04	**1:20**	50,000,000	720,616	13,537	98	74.3%	37.7%	37.1%	18.4%	100.0%	47.4%
266R05	**1:50**	50,000,000	516,344	13,222	60	74.4%	32.7%	37.0%	16.0%	100.0%	3.0%

Table 1

Sample Name			Total reads		Median Target Coverage		% on bait bases		% on target bases		% Bases >100x	
			standard	final	standard	final	standard	final	standard	final	standard	final
280R02_B01	211P12	5ng	50,000,000	816,094	12,333	12	71.4%	6.5%	35.1%	3.1%	99.3%	0.0%
280R03_C01	222P09	5ng	50,000,000	772,752	12,469	20	71.6%	9.6%	35.5%	4.7%	99.5%	0.0%

Table 1-2

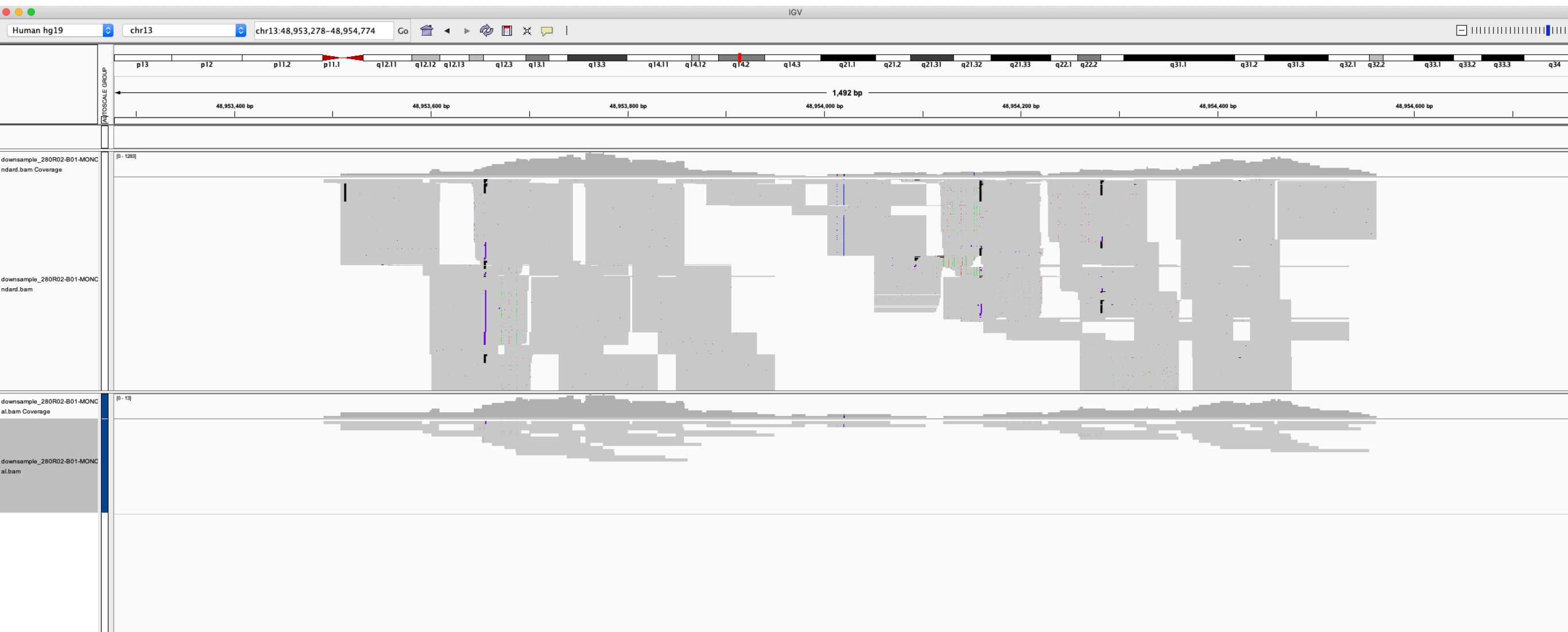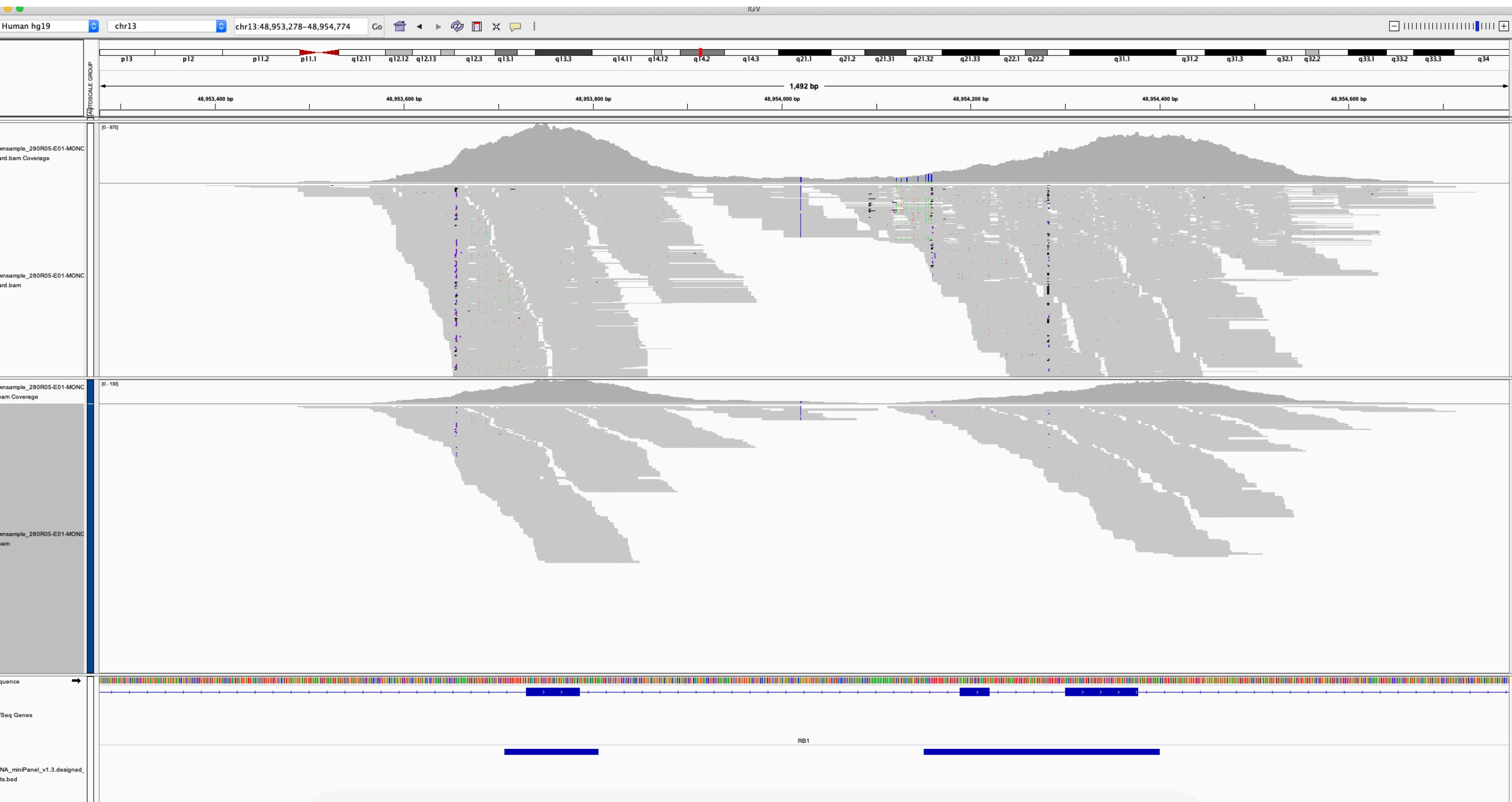|												
280R05_E01	211P12	1ng	50,000,000	547,794	12,789	**219**	74.6%	**52.0%**	34.9%	**23.9%**	99.9%	**97.4%**
280R06_F01	222P09	1ng	50,000,000	343,036	13,085	**76**	70.6%	**25.6%**	35.7%	**12.6%**	100.0%	**10.0%**

appendix

263R02
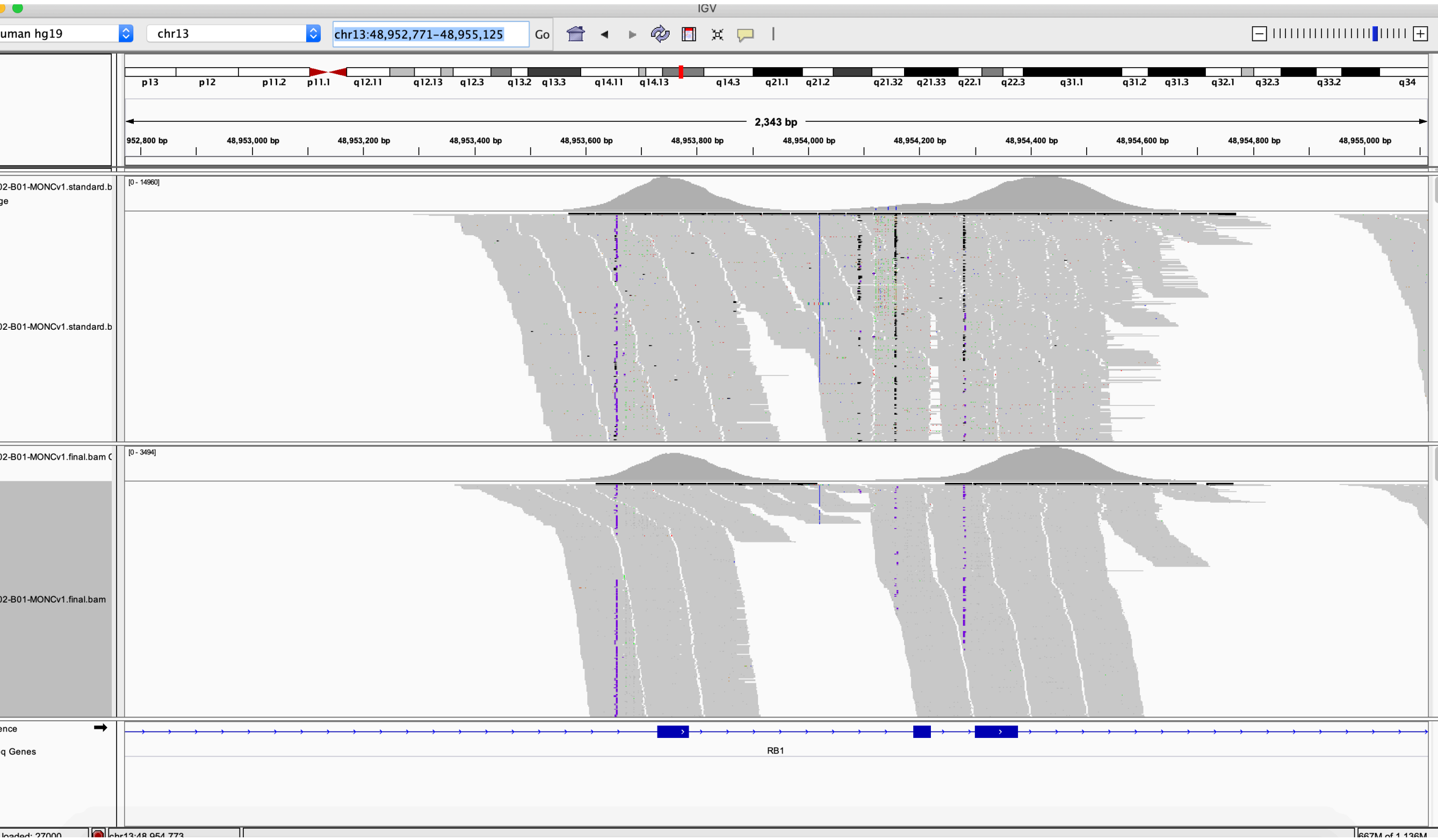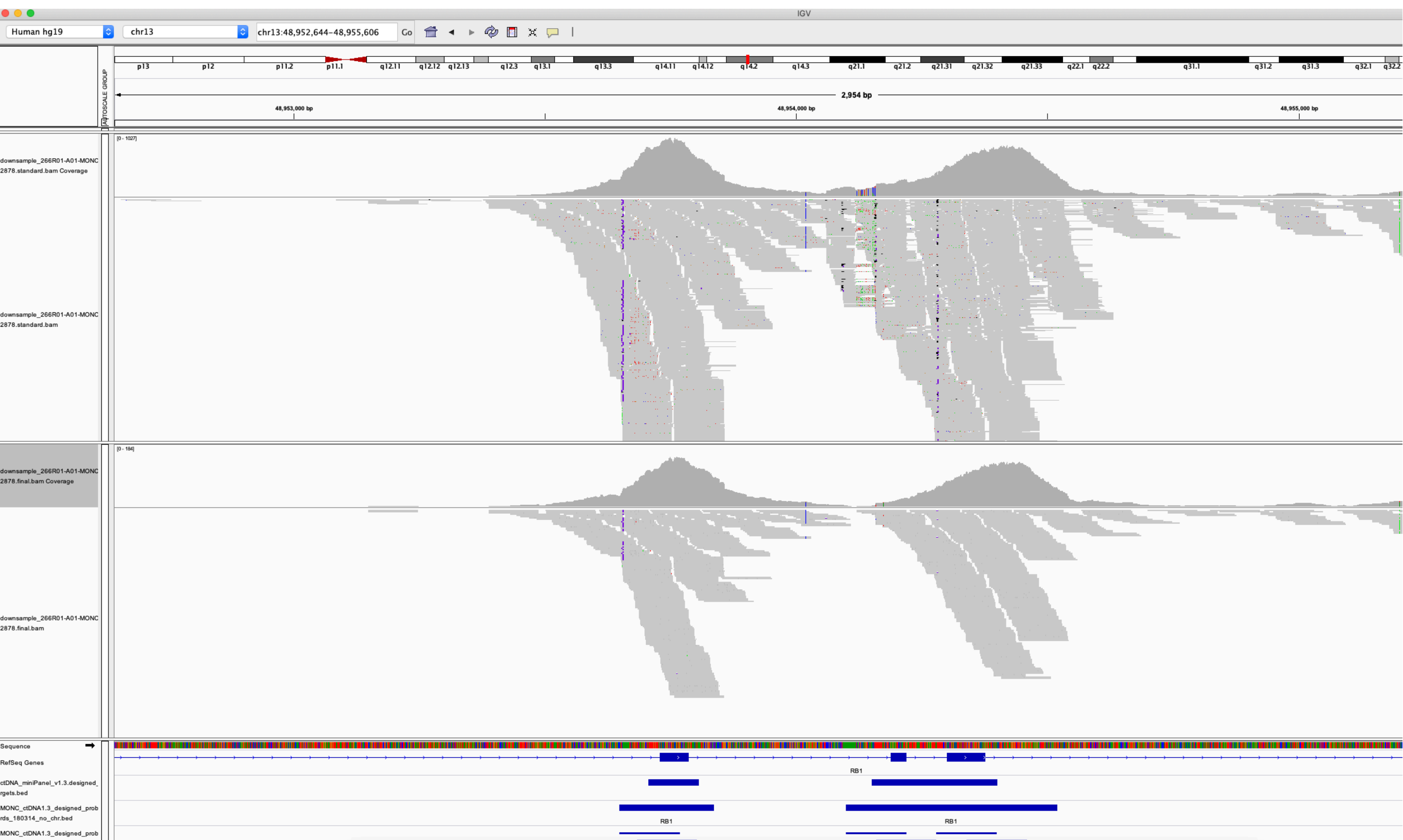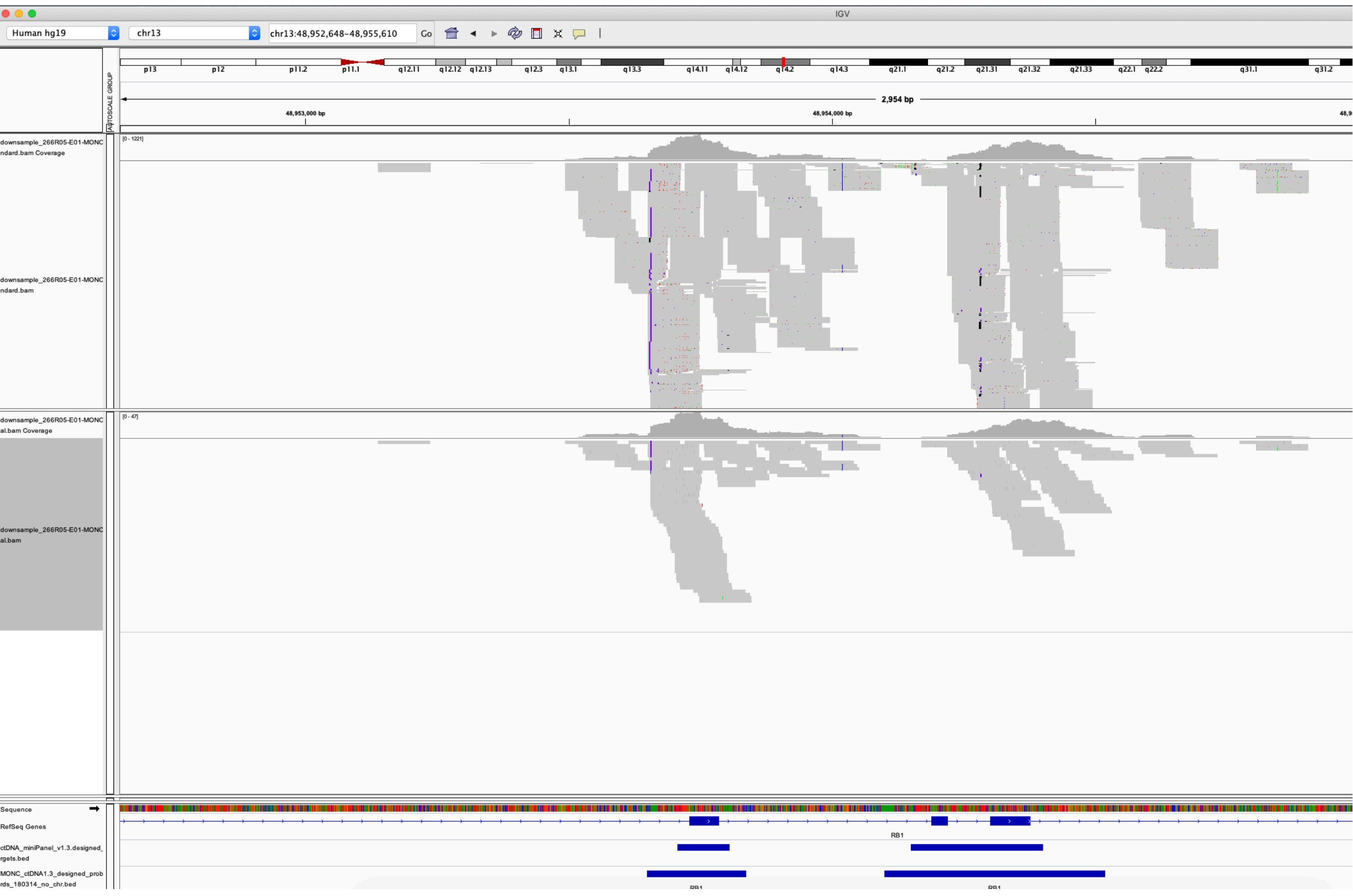
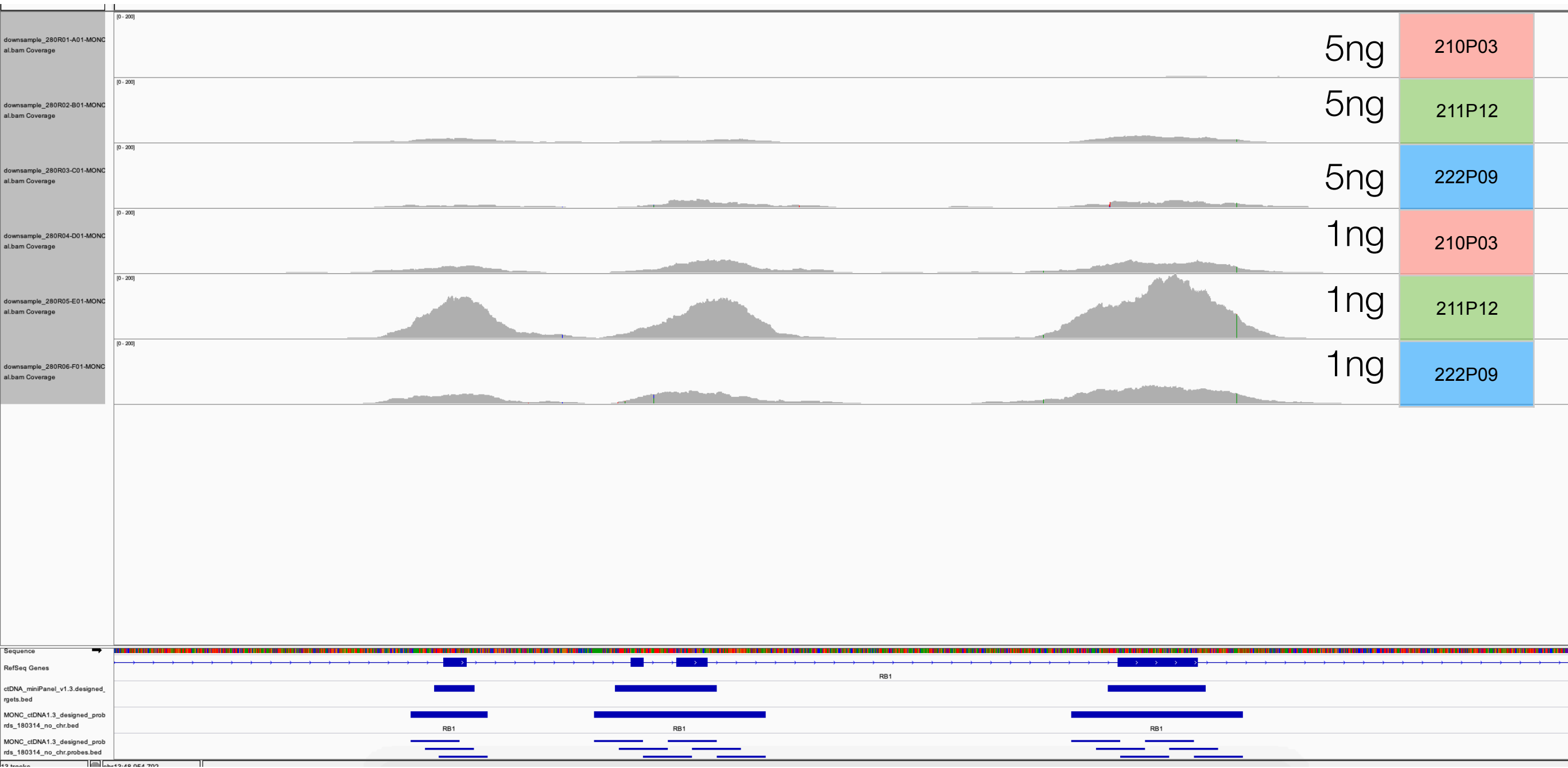# 280R02 (5ng) standard vs final

280R05 (1ng) standard vs final
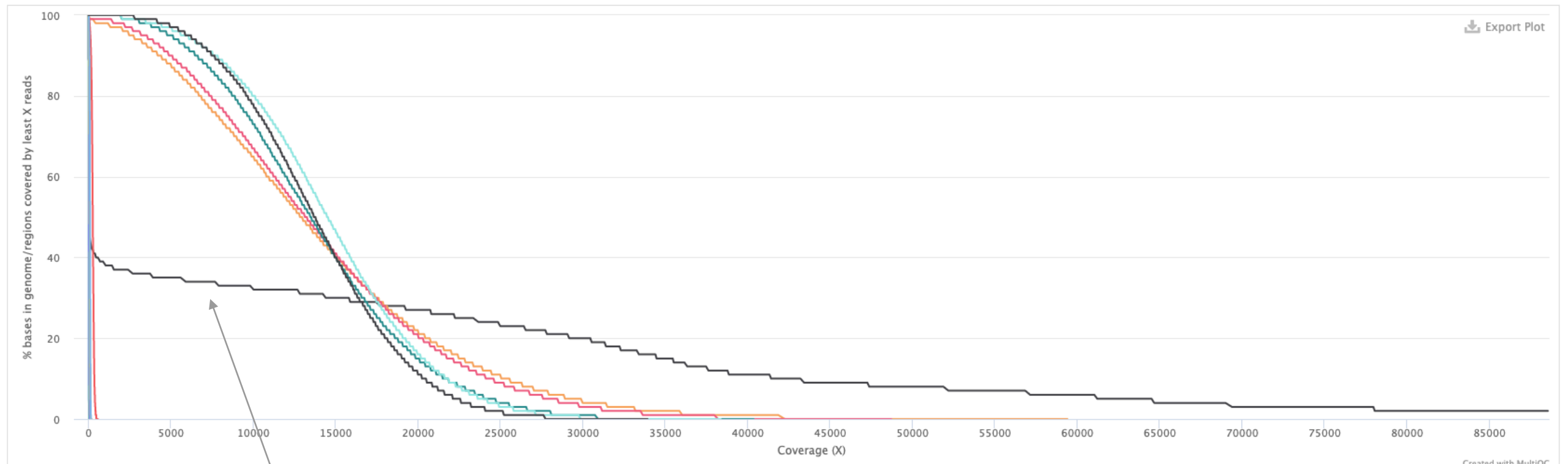
280R coverage snapshot (scale 0-200)

280R coverage (autoscale)
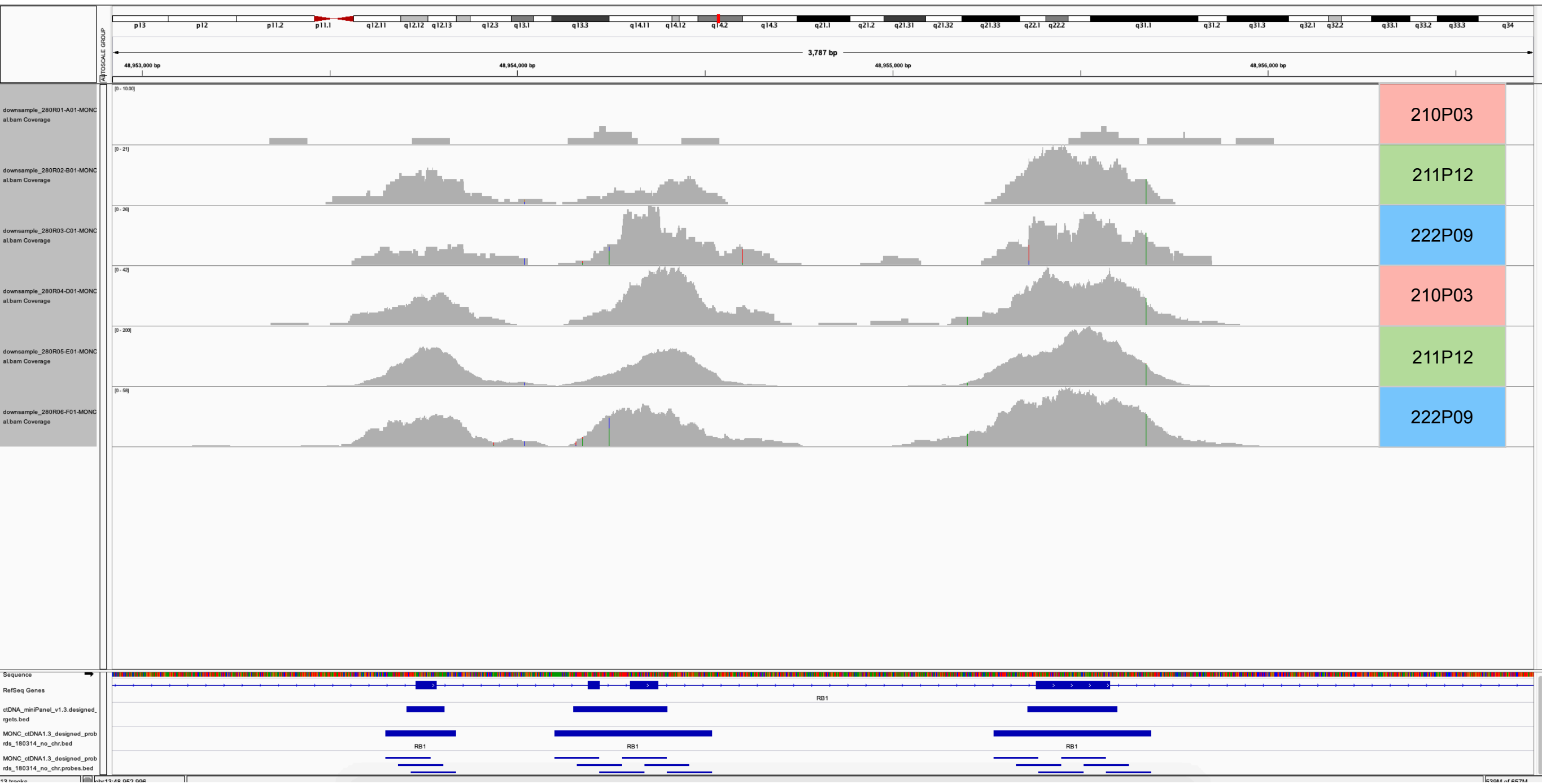
Coverage distribution

Distribution of the number of locations in the reference genome with a given depth of coverage

Indicates ~60% of targets have no coverage, remaining 40% have all reads

280R01-A01 failed but note that the MEAN coverage looks great, the <u>median coverage</u> is bad;

→ represents very strong jackpot effect (max target coverage =150,000x vs 20-50k average)