APPENDIX

SUPPLEMENTAL FIGURES AND TABLES

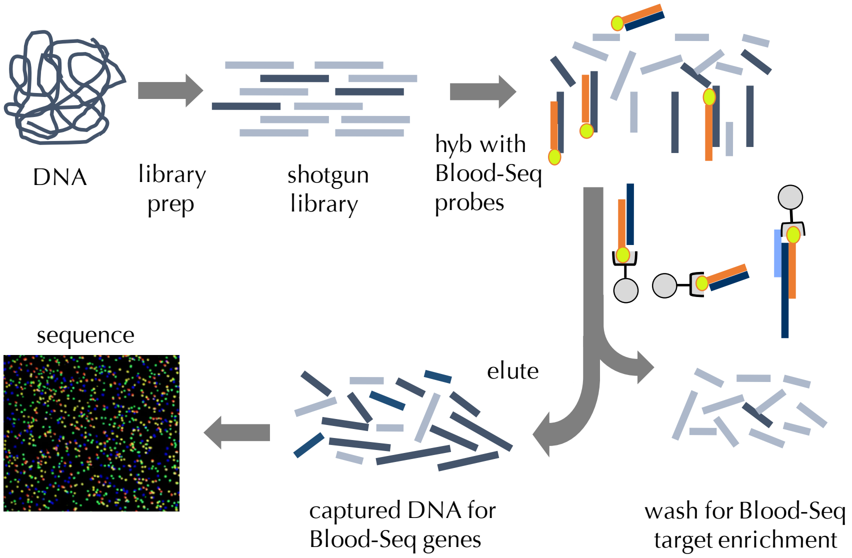
FOR

NEXT-GENERATION *ABO* GENETICS AND GENOMICS

PATRICK KEOLU OZER FOX

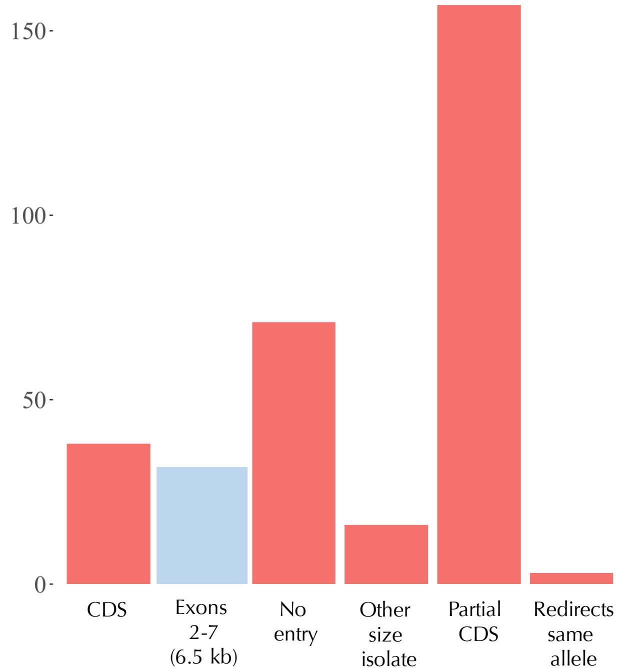
(THIS PAGE LEFT INTENTIONALLY BLANK)

SUPPLEMENTAL FIGURES — CHAPTER 3: BUILDING A SEQUENCE-BASED CALLER FOR *ABO*



Supplemental Figure 3.1: **Overview of Blood-Seq targeted capture.** Nimblegen oligo probes were designed to capture 41 blood-group gene regions (~1.5 MB total). Next we used a Covaris shearing technique to sonicate the DNA into a targeted inset size of ~200 base pair, with a goal of sequencing paired end 100 BP reads. We prepared the library first by performing end repair, then attaching poly A tails, adding sample specific bar codes, and finally PCR to amplify the desired DNA product. We then performed the targeted multiplex capture (24 samples each). We then pulled down the biotinylated probes with the desired DNA attached, performed a bead clean up and were left with the targeted capture product readying for Illumina next-generation sequencing. 96 samples were sequenced per lane with an average of 150X coverage.

a.)

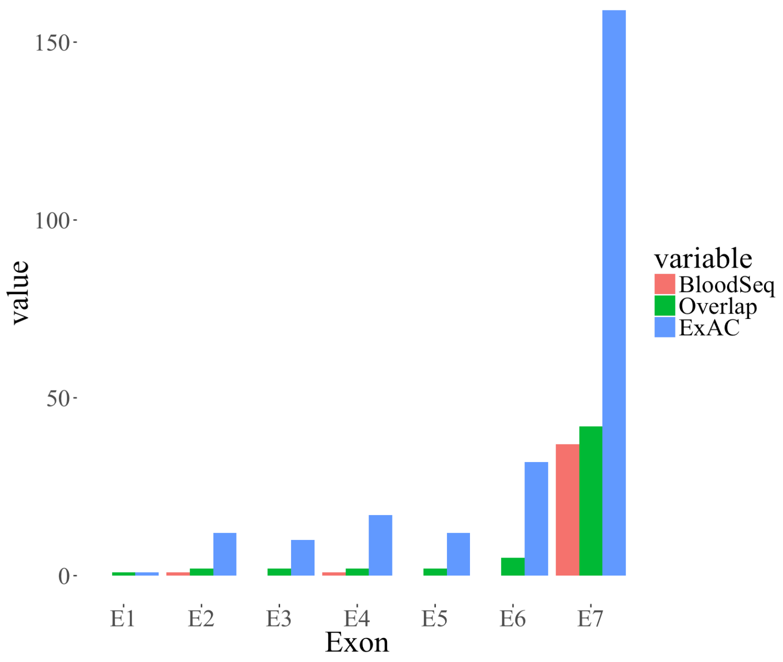


Supplemental Figure 3.2a: **Filtering *ABO* BGMUT allele entries on size.** In order to curate the allelic variation existing in *ABO* on BGMUT with corresponding FASTA we designated 6 categories for *ABO* allelic classification, (1) coding sequence (CDS) n=38, (2) partial CDS n=157, (3) multiple entries classified to the same allele n=3, (4) No submission n=71, (5) exons 2-7 n=31, and (6) other size isolate n=16. None of the entries included in BGMUT included the 13.5 kb stretch of sequence beginning in exon 1 through the end of intron 1. Finally, we converted all qualifying 151 entries with corresponding FASTA sequence available on Genbank to hg19 coordinates for our BGMUT coordinate lookup table database.

b.)

O40	**100**	99.91	99.88	99.86	99.89	99.88	99.32	99.34	99.55	99.18	99.18	99.18	99.2	99.2	99.22	99.2	99.2	99.22	99.46	99.26	99.26	99.26	99.28	99.29	99.29	99.31	99.31	99.31	99.31	99.31
O06	99.91	**100**	99.91	99.92	99.95	99.94	99.26	99.28	99.58	99.18	99.18	99.18	99.2	99.2	99.22	99.2	99.2	99.22	99.49	99.2	99.29	99.29	99.31	99.32	99.32	99.34	99.34	99.34	99.25	99.25
O73	99.88	99.91	**100**	99.92	99.95	99.94	99.2	99.22	99.55	99.15	99.15	99.15	99.17	99.17	99.18	99.17	99.17	99.18	99.43	99.14	99.23	99.23	99.25	99.26	99.26	99.28	99.28	99.28	99.18	99.18
O63	99.86	99.92	99.92	**100**	99.97	99.95	99.22	99.23	99.54	99.14	99.14	99.14	99.15	99.15	99.17	99.15	99.15	99.17	99.42	99.12	99.22	99.22	99.23	99.25	99.25	99.26	99.26	99.26	99.17	99.17
O65	99.89	99.95	99.95	99.97	**100**	99.98	99.25	99.26	99.57	99.17	99.17	99.17	99.18	99.18	99.2	99.18	99.18	99.2	99.45	99.15	99.25	99.25	99.26	99.28	99.28	99.29	99.29	99.29	99.2	99.2
O01 - allele	99.88	99.94	99.94	99.95	99.98	**100**	99.23	99.25	99.55	99.15	99.15	99.15	99.17	99.17	99.18	99.17	99.17	99.18	99.43	99.14	99.23	99.23	99.25	99.26	99.26	99.28	99.28	99.28	99.18	99.18
Aw08	99.32	99.26	99.2	99.22	99.25	99.23	**100**	99.98	99.49	99.68	99.68	99.68	99.69	99.69	99.71	99.69	99.69	99.71	99.49	99.54	99.54	99.54	99.55	99.57	99.57	99.58	99.58	99.58	99.58	99.58
O03	99.34	99.28	99.22	99.23	99.26	99.25	99.98	**100**	99.51	99.69	99.69	99.69	99.71	99.71	99.72	99.71	99.71	99.72	99.51	99.55	99.55	99.55	99.57	99.58	99.58	99.6	99.6	99.6	99.6	99.6
Ax08	99.55	99.58	99.55	99.54	99.57	99.55	99.49	99.51	**100**	99.6	99.6	99.6	99.62	99.62	99.63	99.62	99.62	99.63	99.57	99.52	99.68	99.68	99.69	99.71	99.71	99.72	99.72	99.72	99.57	99.57
B116	99.18	99.18	99.15	99.14	99.17	99.15	99.68	99.69	99.6	**100**	99.97	99.97	99.98	99.98	99.94	99.95	99.95	99.97	99.54	99.49	99.65	99.65	99.69	99.71	99.68	99.69	99.69	99.69	99.54	99.54
Bw07	99.18	99.18	99.15	99.14	99.17	99.15	99.68	99.69	99.6	99.97	**100**	99.97	99.98	99.98	99.94	99.95	99.95	99.97	99.54	99.49	99.65	99.65	99.69	99.71	99.68	99.69	99.69	99.69	99.54	99.54
Bw08	99.18	99.18	99.15	99.14	99.17	99.15	99.68	99.69	99.6	99.97	99.97	**100**	99.98	99.98	99.94	99.95	99.95	99.97	99.54	99.49	99.65	99.65	99.69	99.71	99.68	99.69	99.69	99.69	99.54	99.54
B101	99.2	99.2	99.17	99.15	99.18	99.17	99.69	99.71	99.62	99.98	99.98	99.98	**100**	**100**	99.95	99.97	99.97	99.98	99.55	99.51	99.66	99.66	99.71	99.72	99.69	99.71	99.71	99.71	99.55	99.55
B114	99.2	99.2	99.17	99.15	99.18	99.17	99.69	99.71	99.62	99.98	99.98	99.98	**100**	**100**	99.95	99.97	99.97	99.98	99.55	99.51	99.66	99.66	99.71	99.72	99.69	99.71	99.71	99.71	99.55	99.55
B(A)03	99.22	99.22	99.18	99.17	99.2	99.18	99.71	99.72	99.63	99.94	99.94	99.94	99.95	99.95	**100**	99.98	99.95	99.97	99.57	99.52	99.68	99.68	99.72	99.74	99.71	99.72	99.72	99.72	99.57	99.57
B113	99.2	99.2	99.17	99.15	99.18	99.17	99.69	99.71	99.62	99.95	99.95	99.95	99.97	99.97	99.98	**100**	99.97	99.98	99.55	99.51	99.66	99.66	99.71	99.72	99.69	99.71	99.71	99.71	99.55	99.55
B - Weak Isolate	99.2	99.2	99.17	99.15	99.18	99.17	99.69	99.71	99.62	99.95	99.95	99.95	99.97	99.97	99.95	99.97	**100**	99.98	99.55	99.51	99.66	99.66	99.71	99.72	99.69	99.71	99.71	99.71	99.55	99.55
B - Allele Isolate	99.22	99.22	99.18	99.17	99.2	99.18	99.71	99.72	99.63	99.97	99.97	99.97	99.98	99.98	99.97	99.98	99.98	**100**	99.57	99.52	99.68	99.68	99.72	99.74	99.71	99.72	99.72	99.72	99.57	99.57
A109	99.46	99.49	99.43	99.42	99.45	99.43	99.49	99.51	99.57	99.54	99.54	99.54	99.55	99.55	99.57	99.55	99.55	99.57	**100**	99.71	99.8	99.8	99.82	99.83	99.83	99.85	99.85	99.85	99.75	99.75
A215	99.26	99.2	99.14	99.12	99.15	99.14	99.54	99.55	99.52	99.49	99.49	99.49	99.51	99.51	99.52	99.51	99.51	99.52	99.71	**100**	99.82	99.78	99.77	99.78	99.82	99.8	99.8	99.8	99.89	99.89
A214	99.26	99.29	99.23	99.22	99.25	99.23	99.54	99.55	99.68	99.65	99.65	99.65	99.66	99.66	99.68	99.66	99.66	99.68	99.8	99.82	**100**	99.94	99.92	99.94	99.97	99.95	99.95	99.95	99.83	99.83
Aw07	99.26	99.29	99.23	99.22	99.25	99.23	99.54	99.55	99.68	99.65	99.65	99.65	99.66	99.66	99.68	99.66	99.66	99.68	99.8	99.78	99.94	**100**	99.92	99.94	99.97	99.95	99.95	99.95	99.8	99.8
A108	99.28	99.31	99.25	99.23	99.26	99.25	99.55	99.57	99.69	99.69	99.69	99.69	99.71	99.71	99.72	99.71	99.71	99.72	99.82	99.77	99.92	99.92	**100**	99.95	99.95	99.97	99.97	99.97	99.82	99.82
B115	99.29	99.32	99.26	99.25	99.28	99.26	99.57	99.58	99.71	99.71	99.71	99.71	99.72	99.72	99.74	99.72	99.72	99.74	99.83	99.78	99.94	99.94	99.95	**100**	99.97	99.98	99.98	99.98	99.83	99.83
O21	99.29	99.32	99.26	99.25	99.28	99.26	99.57	99.58	99.71	99.68	99.68	99.68	99.69	99.69	99.71	99.69	99.69	99.71	99.83	99.82	99.97	99.97	99.95	99.97	**100**	99.98	99.98	99.98	99.83	99.83
A101	99.31	99.34	99.28	99.26	99.29	99.28	99.58	99.6	99.72	99.69	99.69	99.69	99.71	99.71	99.72	99.71	99.71	99.72	99.85	99.8	99.95	99.95	99.97	99.98	99.98	**100**	**100**	**100**	99.85	99.85
Ael01	99.31	99.34	99.28	99.26	99.29	99.28	99.58	99.6	99.72	99.69	99.69	99.69	99.71	99.71	99.72	99.71	99.71	99.72	99.85	99.8	99.95	99.95	99.97	99.98	99.98	**100**	**100**	**100**	99.85	99.85
Ael03	99.31	99.34	99.28	99.26	99.29	99.28	99.58	99.6	99.72	99.69	99.69	99.69	99.71	99.71	99.72	99.71	99.71	99.72	99.85	99.8	99.95	99.95	99.97	99.98	99.98	**100**	**100**	**100**	99.85	99.85
O26	99.31	99.25	99.18	99.17	99.2	99.18	99.58	99.6	99.57	99.54	99.54	99.54	99.55	99.55	99.57	99.55	99.55	99.57	99.75	99.89	99.83	99.8	99.82	99.83	99.83	99.85	99.85	99.85	**100**	99.97
O66	99.31	99.25	99.18	99.17	99.2	99.18	99.58	99.6	99.57	99.54	99.54	99.54	99.55	99.55	99.57	99.55	99.55	99.57	99.75	99.89	99.83	99.8	99.82	99.83	99.83	99.85	99.85	99.85	99.97	**100**

Supplemental Figure 3.2b: **Percent identity matrix for BGMUT.** After deriving *ABO* haplotypes composed of criteria qualifying data (i.e., the beginning of exon 2-exon 7 or ~6.5 kb) from 248 *ABO* alleles with corresponding BGMUT alleles we analyzed sequence identity (uniqueness) using Multiple Sequence Comparison by Log- Expectation (MUSCLE). Of the 33 entries that included exon 2 – the end of exon 7, (i.e., the most complete *ABO* FASTAs available on BGMUT, size= ~6.5 kp), 28 had unique *ABO* sequence qualifying them as independent alleles. Five of the ~6.5 kb BGMUT entries had the exact same sequence listed under different accession numbers leaving us with 28 high quality (~6.5 kb) allele entries with unique sequence. Those 28 high quality BGMUT reference FASTAs were then used as reference haplotypes in our matrix based calling algorithm (i.e., ABO-Seq).



Supplemental figure 3.3: **Comparing *ABO* SNVs in the BloodSeq to the ExAC database by exon.**

SUPPLEMENTAL TABLES — CHAPTER 3: BUILDING A SEQUENCE-BASED CALLER FOR *ABO*

Supplemental Table 3.1: **Blood-Seq annotated *ABO* SNVs for 1,140 individuals (2,280 *ABO* chromosomes) using SeattleSeq.**

DBSNP	Position	Reference	Sample	Function GVS	rsID	amino acid	protein	poly-Phen	GERP	ExAC
dbSNP_129	136131056	CG	C	frameshift	56392308	none	NA	unknown	-5.32	delG=5059/ref=40137
dbSNP_129	136131064	G	G/A	missense	56390333	ARG,TRP	352/355	1	1.03	C=2/A=41/G=54483
dbSNP_129	136131109	T	T/C	missense	56231718	ARG,GLY	337/355	0.47	-0.002	C=29/T=109011
dbSNP_138	136131118	C	C/T	missense	369695939	ALA,THR	334/355	0.013	-8.76	T=2/C=113272
none	136131154	C	C/T	missense	0	GLU,LYS	322/355	1	4.38	T=5/C=119169
none	136131166	C	C/A	missense	0	VAL,LEU	318/355	0.024	3.48	A=1/C=119981
dbSNP_117	136131188	C	C/T	synonymous	8176749	none	310/355	unknown	1.36	T=14613/C=106423
none	136131236	G	G/A	synonymous	0	none	294/355	unknown	-8.76	A=2/G=121076
dbSNP_129	136131247	C	C/T	missense	56409303	ASP,ASN	291/355	1	4.38	T=1/C=120997
dbSNP_117	136131289	C	T/C	missense	8176748	VAL,MET	277/355	1	4.38	A=1/T=31193/C=87126
dbSNP_117	136131315	C	C/G	missense	8176747	GLY,ALA	268/355	0.006	-8.39	G=14032/C=93378
dbSNP_127	136131316	C	C/T	missense	41302905	GLY,ARG	268/355	0.99	3.28	T=1726/C=105746
dbSNP_117	136131322	G	G/T	missense	8176746	LEU,MET	266/355	0.045	-8.76	A=1/T=13903/G=91164
dbSNP_117	136131347	G	A/G	synonymous	8176745	none	257/355	unknown	-1.59	A=25722/G=61360
dbSNP_117	136131350	G	G/T	synonymous	8176744	none	256/355	unknown	-3.34	T=2927/G=70809
dbSNP_138	136131392	C	C/T	synonymous	376910781	none	242/355	unknown	-2.95	A=1/T=1/C=52470
dbSNP_129	136131397	G	G/A	missense	55951833	ARG,TRP	241/355	1	3.44	A=6/G=51626
dbSNP_137	136131407	G	G/A	synonymous	201439325	none	237/355	unknown	-0.929	A=48/G=50292
dbSNP_117	136131415	C	C/T	missense	8176743	GLY,SER	235/355	0.858	2.51	T=8378/C=44320
dbSNP_129	136131429	C	C/T	missense	56116432	GLY,ASP	230/355	1	4.39	T=223/C=5058

										5
dbSNP_129	136131431	G	G/A	synonymous	56215404	none	229/355	unknown	-3.81	A=1/G=50913
dbSNP_117	136131437	C	T/C	synonymous	8176742	none	227/355	unknown	-8.65	T=13460/C=43502
dbSNP_117	136131461	G	G/A	synonymous	8176741	none	219/355	unknown	-4.2	A=9491/G=50041
dbSNP_129	136131469	G	G/A	missense	56408700	ARG,CYS	217/355	0.037	-9.38	A=86/G=58342
dbSNP_117	136131472	A	T/A	missense	8176740	PHE,ILE	216/355	1	3.55	T=16226/A=47766
dbSNP_117	136131523	G	G/A	missense	8176739	ARG,CYS	199/355	0.868	2.8	A=1405/G=66839
dbSNP_129	136131539	A	A/G	synonymous	55764262	none	193/355	unknown	3.79	G=74/A=77166
dbSNP_129	136131576	C	C/T	stop-gained	55727303	TRP,stop	181/355	unknown	4.56	T=3278/C=98528
dbSNP_129	136131579	C	C/T	missense	56258249	ARG,HIS	180/355	0.171	3.66	T=1/C=102671
dbSNP_116	136131592	G	G/C	missense	7853989	ARG,GLY	176/355	0.011	-1.09	A=1/C=15935/G=94822
none	136131595	C	C/T	missense	0	VAL,MET	175/355	1	3.77	T=51/C=112005
none	136131604	C	C/G	missense	0	VAL,LEU	172/355	0.308	0.934	unknown
dbSNP_129	136131616	G	G/C	missense	56043861	ARG,GLY	168/355	1	-4.42	A=1/C=2/G=115095
dbSNP_129	136131621	GT	G	frameshift	56284703	none	NA	unknown	-8.32	delT=410/ref=115308
dbSNP_137	136131629	C	C/T	synonymous	201567722	none	163/355	unknown	1.62	G=1/A=2/T=20/C=116177
dbSNP_137	136131635	G	G/A	synonymous	200932155	none	161/355	unknown	-3.15	A=139/G=116133
dbSNP_117	136131636	C	C/T	missense	8176738	ARG,HIS	161/355	0	-9.38	A=1/T=84/C=116231
none	136131647	G	G/A	synonymous	0	none	157/355	unknown	-8.34	A=2/G=116552
dbSNP_86	136131651	G	G/A	missense	1053878	PRO,LEU	156/355	0.95	3.77	A=10390/G=106232
dbSNP_137	136131740	G	G/A	synonymous	201105186	none	126/355	unknown	2.29	A=23/G=115525
dbSNP_117	136132852	G	G/A	synonymous	8176721	none	106/355	unknown	-3.12	A=1601/G=120723
none	136132864	G	G/A	synonymous	0	none	102/355	unknown	0.112	A=1/G=122311
dbSNP_117	136132873	T	C/T	synonymous	8176720	none	99/355	unknown	-8.39	C=48708/T=73496
none	136132906	G	G/A	synonymous	0	none	88/355	unknown	-4.38	T=7/G=103137

dbSNP_117	136132908	T	TC	frameshift	8176719	none	NA	unknown	4.2	insC=45730/ref=75728
none	136133490	G	G/T	missense	0	PRO,GLN	79/355	0.992	2.04	C=1/A=2/G=121973
dbSNP_83	136133506	A	A/G	missense	512770	SER,PRO	74/355	unknown	2.04	G=92871/A=29225
dbSNP_83	136135237	A	A/G	coding-unknown	549443	none	NA	unknown	-2.61	G=90060/A=31798
dbSNP_83	136135238	T	T/C	missense	549446	HIS,ARG	63/355	unknown	-2.61	unknown
dbSNP_83	136136770	A	A/C	missense	688976	PHE,VAL	36/355	unknown	-3.95	C=66731/A=21321
dbSNP_117	136136773	C	C/T	missense	8176696	GLY,ARG	35/355	0.806	2.04	A=2/T=1262/C=58966
dbSNP_129	136137547	C	C/A	missense	55876802	ARG,LEU	18/355	0.001	-2.13	T=1/A=1785/C=109688
none	136137553	G	G/C	missense	0	ALA,GLY	16/355	0.005	-2.13	unknown
dbSNP_129	136137554	C	C/T	missense	55917063	ALA,THR	16/355	0.062	-4.6	G=2/T=215/C=108557
dbSNP_138	136150600	G	G/A	synonymous	367824410	none	2/355	unknown	-1.44	A=11/G=4317

Supplemental Table 3.2: **BGMUT*ABO* reference alleles annotated SNVs for (151 ABO reference alleles total) using SeattleSNPs.**

DBSNP	position	reference Base	sampleGenotype	GVSfunction	rsID	proteinPosition	polyPhen	consScoreGERP	genomesExAC
dbSNP_129	136131056	CG	C	frameshift	56392308	NA	unknown	-5.32	delG=5059/ref=40137
dbSNP_129	136131063	C	T	missense	56202119	352/355	0.999	3.2	unknown
dbSNP_129	136131064	G	A/C	coding-unknown	56390333	352/355	unknown	1.03	C=2/A=41/G=54483
dbSNP_129	136131082	T	C	missense	56016583	346/355	0.869	3.19	unknown
dbSNP_129	136131109	T	C	missense	56231718	337/355	0.47	-0.002	C=29/T=109011
dbSNP_129	136131122	C	T	stop-gained	55788852	332/355	unknown	2.48	T=1/C=114933
none	136131142	C	A	missense	0	326/355	0.78	4.38	unknown
dbSNP_129	136131153	T	C	missense	55927860	322/355	1	4.38	unknown
none	136131166	C	T	missense	0	318/355	0.996	3.48	A=1/C=119981
none	136131178	T	C	missense	0	314/355	0.999	4.38	unknown
none	136131180	T	G	missense	0	313/355	0.999	3.21	unknown
dbSNP_117	136131188	C	T	synonymous	8176749	310/355	unknown	1.36	T=14613/C=106423
none	136131195	T	C	missense	0	308/355	0.209	-0.639	unknown

none	136131213	T	C	missense	0	302/355	1	1.84	unknown
none	136131218	C	G	missense	0	300/355	0.923	4.38	unknown
dbSNP_129	136131225	G	A	missense	55783488	298/355	0.999	3.46	unknown
dbSNP_138	136131229	C	T	missense	370138477	297/355	1	4.38	unknown
dbSNP_129	136131245	G	C	missense	56355240	291/355	1	3.47	A=1/G=121025
dbSNP_129	136131247	C	T	missense	56409303	291/355	1	4.38	T=1/C=120997
dbSNP_129	136131255	A	C	missense	55805279	288/355	0.995	4.38	unknown
dbSNP_117	136131289	C	T	missense	8176748	277/355	1	4.38	A=1/T=31193/C=87126
none	136131310	A	T	missense	0	270/355	1	4.2	unknown
dbSNP_117	136131315	C	G	missense	8176747	268/355	0.006	-8.39	G=14032/C=93378
dbSNP_127	136131316	C	T	missense	41302905	268/355	0.99	3.28	T=1726/C=105746
dbSNP_117	136131322	G	T	missense	8176746	266/355	0.045	-8.76	A=1/T=13903/G=91164
dbSNP_129	136131340	C	T	missense	56323007	260/355	0.976	-1.85	T=1/C=90659
dbSNP_129	136131342	T	C	missense	56396277	259/355	0.134	3.2	unknown
dbSNP_117	136131347	G	A	synonymous	8176745	257/355	unknown	-1.59	A=25722/G=61360
dbSNP_117	136131350	G	T	synonymous	8176744	256/355	unknown	-3.34	T=2927/G=70809
dbSNP_129	136131351	A	G	missense	55751228	256/355	1	4.38	unknown
none	136131375	C	G	missense	0	248/355	1	1.42	T=2/G=7/C=57099
none	136131376	G	A	missense	0	248/355	0.997	3.46	unknown
none	136131396	C	T	missense	0	241/355	1	2.51	unknown
dbSNP_129	136131397	G	A	missense	55951833	241/355	1	3.44	A=6/G=51626
dbSNP_117	136131415	C	T	missense	8176743	235/355	0.858	2.51	T=8378/C=44320
dbSNP_129	136131423	A	G	missense	55739900	232/355	1	4.39	unknown
dbSNP_129	136131429	C	T	missense	56116432	230/355	1	4.39	T=223/C=50585
none	136131430	C	G	missense	0	230/355	1	4.39	unknown
dbSNP_117	136131437	C	T	synonymous	8176742	227/355	unknown	-8.65	T=13460/C=43502
dbSNP_129	136131449	C	A	missense	56159791	223/355	1	3.48	unknown
dbSNP_117	136131461	G	A	synonymous	8176741	219/355	unknown	-4.2	A=9491/G=50041
dbSNP_129	136131469	G	A	missense	56408700	217/355	0.037	-9.38	A=86/G=58342
dbSNP_117	136131472	A	T	missense	8176740	216/355	1	3.55	T=16226/A=47766
dbSNP_129	136131477	A	C	missense	55827808	214/355	1	4.69	unknown
dbSNP_138	136131484	C	T	missense	376840879	212/355	1	4.69	T=1/C=61973
dbSNP_138	136131511	C	T/G	coding-unknown	374698850	203/355	unknown	1.72	T=5/C=64639
dbSNP_129	136131526	G	A	missense	56223957	198/355	0.998	3.76	unknown
none	136131530	G	C	missense	0	196/355	0.996	-9.38	unknown
dbSNP_129	136131539	A	G	synonymous	55764262	193/355	unknown	3.79	G=74/A=77166
none	136131543	A	G	missense	0	192/355	0.993	4.69	unknown
none	136131553	T	C	missense	0	189/355	0.999	4.69	C=1/T=83405

none	136131559	G	A	missense	0	187/355	1	1.46	unknown
none	136131560	C	A	missense	0	186/355	0.809	3.66	unknown
none	136131564	G	A	missense	0	185/355	0.92	3.79	unknown
none	136131568	C	T	missense	0	184/355	0.997	-2.66	T=5/C=95009
dbSNP_129	136131570	T	C	missense	56089890	183/355	0.995	4.56	unknown
none	136131577	A	G	missense	0	181/355	0.928	4.56	unknown
dbSNP_129	136131579	C	T	missense	56258249	180/355	0.171	3.66	T=1/C=102671
dbSNP_129	136131580	G	A	missense	56164697	180/355	0.922	1.41	unknown
dbSNP_129	136131589	C	T	missense	55687199	177/355	0.048	-6.69	T=96/C=107234
dbSNP_129	136131591	C	T	missense	56039827	176/355	0.024	-1.05	T=143/C=108359
dbSNP_116	136131592	G	C	missense	7853989	176/355	0.011	-1.09	A=1/C=15935/G=94822
none	136131595	C	T	missense	0	175/355	1	3.77	T=51/C=112005
none	136131615	C	A	missense	0	168/355	1	2.83	unknown
dbSNP_129	136131620	G	A	synonymous	55964869	166/355	unknown	-8.4	C=1/A=2/G=115543
dbSNP_86	136131651	G	A	missense	1053878	156/355	0.95	3.77	A=10390/G=106232
none	136131684	T	C	missense	0	145/355	1	4.69	unknown
none	136131692	C	G	missense	0	142/355	1	3.77	unknown
none	136131693	A	G	missense	0	142/355	1	3.55	unknown
dbSNP_129	136131711	G	A	missense	55658842	136/355	1	4.95	A=6/G=116318
none	136132800	T	C	missense	0	124/355	0.154	4.33	unknown
dbSNP_129	136132819	C	T	synonymous	55748923	117/355	-0.779	ABO	unknown
none	136132820	C	G	missense	0	117/355	1	4.33	G=1/T=1/C=122320
none	136132823	A	G	missense	0	116/355	0.611	3.16	G=1/A=122321
dbSNP_131	136132848	G	A	stop-gained	77805226	108/355	un-known	3.42	unknown
dbSNP_117	136132852	G	A	synonymous	8176721	106/355	unknown	-3.12	A=1601/G=120723
dbSNP_117	136132873	T	C	synonymous	8176720	99/355	unknown	-8.39	C=48708/T=73496
dbSNP_117	136132908	T	TC	frameshift	8176719	NA	unknown	4.2	insC=45730/ref=75728
none	136133490	G	A	missense	0	79/355	0.425	2.04	C=1/A=2/G=121973
dbSNP_83	136133506	A	G	missense	512770	74/355	unknown	2.04	G=92871/A=29225
dbSNP_129	136135236	C	T	missense	56335272	64/355	0.137	1.3	unknown
dbSNP_83	136135237	A	G	coding-unknown	549443	63/355	unknown	-2.61	G=90060/A=31798
dbSNP_83	136135238	T	C	missense	549446	63/355	unknown	-2.61	unknown
dbSNP_83	136136770	A	C	missense	688976	36/355	unknown	-3.95	C=66731/A=21321
dbSNP_129	136137554	C	T	missense	55917063	16/355	0.062	-4.6	G=2/T=215/C=108557
none	136150605	T	C	missense	0	1/355	0.059	0.72	unknown

Supplemental Table 3.3: **Overlap in identification of *ABO* in variants in blood-Seq and BGMUT using VCFtools compare command**.

CHROM	POS1	POS2	IN_FILE	REF1	REF2	ALT1	ALT2
9	136131056	136131056	B	AC	CG	A	C
9	136131058	.	1	GG .	G	.	
9	136131063	.	1	C	.	T	.
9	136131064	136131064	B	G	G	A,C A	
9	136131082	.	1	T	.	C	.
9	136131109	136131109	B	T	T	C	C
9	.	136131118	2	.	C	.	T
9	136131122	.	1	C	.	T	.
9	136131142	.	1	C	.	A	.
9	136131153	.	1	T	.	C	.
9	.	136131154	2	.	C	.	T
9	136131166	136131166	B	C	C	T	A
9	136131178	.	1	T	.	C	.
9	136131180	.	1	T	.	G	.
9	136131188	136131188	B	C	C	T	T
9	136131195	.	1	T	.	C	.
9	136131213	.	1	T	.	C	.
9	136131218	.	1	C	.	G	.
9	136131225	.	1	G	.	A	.
9	136131229	.	1	C	.	T	.
9	.	136131236	2	.	G	.	A
9	136131245	.	1	G	.	C	.
9	136131247	136131247	B	C	C	T	T
9	136131255	.	1	A	.	C	.
9	136131289	136131289	B	C	C	T	T
9	136131310	.	1	A	.	T	.
9	136131315	136131315	B	C	C	G	G
9	136131316	136131316	B	C	C	T	T
9	136131319	.	1	CC .	CCC,C .		
9	136131322	136131322	B	G	G	T	T

9	136131340	.	1	C	.	T	.
9	136131342	.	1	T	.	C	.
9	136131347	136131347	B	G	G	A	A
9	136131350	136131350	B	G	G	T	T
9	136131351	.	1	A	.	G	.
9	136131375	.	1	C	.	G	.
9	136131376	.	1	G	.	A	.
9	.	136131392	2	.	C	.	T
9	136131396	.	1	C	.	T	.
9	136131397	136131397	B	G	G	A	A
9	.	136131407	2	.	G	.	A
9	136131415	136131415	B	C	C	T	T
9	136131423	.	1	A	.	G	.
9	136131429	136131429	B	C	C	T	T
9	136131430	.	1	C	.	G	.
9	.	136131431	2	.	G	.	A
9	136131437	136131437	B	C	C	T	T
9	136131449	.	1	C	.	A	.
9	136131461	136131461	B	G	G	A	A
9	136131469	136131469	B	G	G	A	A
9	136131472	136131472	B	A	A	T	T
9	136131477	.	1	A	.	C	.
9	136131484	.	1	C	.	T	.
9	136131511	.	1	C	.	T,G	.
9	.	136131523	2	.	G	.	A
9	136131526	.	1	G	.	A	.
9	136131530	.	1	G	.	C	.
9	136131539	136131539	B	A	A	G	G
9	136131543	.	1	A	.	G	.
9	136131553	.	1	T	.	C	.
9	136131559	.	1	G	.	A	.
9	136131560	.	1	C	.	A	.
9	136131564	.	1	G	.	A	.
9	136131568	.	1	C	.	T	.
9	136131570	.	1	T	.	C	.
9	.	136131576	2	.	C	.	T

9	136131577	.	1	A	.	G	.
9	136131579	136131579	B	C	C	T	T
9	136131580	.	1	G	.	A	.
9	136131589	.	1	C	.	T	.
9	136131591	.	1	C	.	T	.
9	136131592	136131592	B	G	G	C	C
9	136131595	136131595	B	C	C	T	T
9	.	136131604	2	.	C	.	G
9	136131615	.	1	C	.	A	.
9	.	136131616	2	.	G	.	C
9	136131620	.	1	G	.	A	.
9	.	136131621	2	.	GT .	G	
9	.	136131629	2	.	C	.	T
9	.	136131635	2	.	G	.	A
9	.	136131636	2	.	C	.	T
9	.	136131647	2	.	G	.	A
9	136131651	136131651	B	G	G	A	A
9	136131684	.	1	T	.	C	.
9	136131692	.	1	C	.	G	.
9	136131693	.	1	A	.	G	.
9	136131711	.	1	G	.	A	.
9	.	136131740	2	.	G	.	A
9	136132800	.	1	T	.	C	.
9	136132819	.	1	C	.	T	.
9	136132820	.	1	C	.	G	.
9	136132823	.	1	A	.	G	.
9	136132848	.	1	G	.	A	.
9	136132852	136132852	B	G	G	A	A
9	.	136132864	2	.	G	.	A
9	136132873	136132873	B	T	T	C	C
9	.	136132906	2	.	G	.	A
9	136132907	.	1	G	.	GC .	
9	.	136132908	2	.	T	.	TC
9	136133486	.	1	CCAC .	C	.	
9	136133490	136133490	B	G	G	A	T
9	136133506	136133506	B	A	A	G	G

9	136135236	.	1	C	.	T	.
9	136135237	136135237	B	A	A	G	G
9	136135238	136135238	B	T	T	C	C
9	136136770	136136770	B	A	A	C	C
9	.	136136773	2	.	C	.	T
9	136137512	.	1	C	.	CC	.
9	.	136137547	2	.	C	.	A
9	.	136137553	2	.	G	.	C
9	136137554	136137554	B	C	C	T	T
9	.	136150600	2	.	G	.	A
9	136150605	.	1	T	.	C	.

32 sites common to both files.

58 sites only in BGMUT file.

23 sites only in Blood-Seq file.

Supplemental Table 3.4: **Blood-Seq phased (PHASE 2.2.1) *ABO* haplotypes for 1,140 individuals (2,280 *ABO* chromosomes): 62 unique haplotypes total from 2,280 *ABO* chromosomes.**

Amino acid											268		266						235											176																								
PHASE	136131056	136131064	136131109	136131118	136131154	136131166	136131188	136131236	136131247	136131289	136131315	136131316	136131322	136131347	136131350	136131392	136131397	136131407	136131415	136131429	136131431	136131437	136131461	136131469	136131472	136131523	136131539	136131576	136131579	136131592	136131595	136131604	136131616	136131621	136131629	136131635	136131636	136131647	136131651	136131740	36132852	136132864	136132873	136132908	136133490	136133506	136135237	136135238	136136770	136136773	136137547	136137553	136137554	Number of haplotypes
1	I	G	T	C	C	C	C	G	C	T	C	C	G	A	G	C	G	G	C	C	G	T	G	G	T	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	C	D	G	A	A	T	A	C	C	G	C	510
2	I	G	T	C	C	C	C	G	C	T	C	C	G	A	G	C	G	G	C	C	G	T	G	G	T	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	C	D	G	A	A	T	A	T	C	G	C	15
3	I	G	T	C	C	C	C	G	C	T	C	C	G	A	G	C	G	G	C	C	G	T	G	G	T	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	C	D	G	A	A	T	A	T	C	G	T	1
4	I	G	T	C	C	C	C	G	C	T	C	C	G	A	G	C	G	G	C	C	G	T	G	G	T	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	C	D	G	G	A	T	A	C	C	G	C	27
5	I	G	T	C	C	C	C	G	C	T	C	C	G	A	G	C	G	G	C	C	G	T	G	G	T	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	C	D	G	G	G	C	C	C	C	G	C	1
6	I	G	T	C	C	C	C	G	C	T	C	C	G	A	G	C	G	G	C	C	G	T	G	G	T	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	C	D	T	A	A	T	A	C	C	G	C	2
7	I	G	T	C	C	C	C	G	C	T	C	C	G	A	G	C	G	G	C	C	G	T	G	G	T	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	T	D	G	G	G	C	C	C	C	G	C	1
8	I	G	T	C	C	C	C	G	C	T	C	C	G	A	G	C	G	G	C	C	G	T	G	G	T	G	A	C	C	G	C	C	G	I	C	G	C	G	G	A	G	G	C	D	G	A	A	T	A	C	C	G	C	1
9	I	G	T	C	C	C	C	G	C	T	C	C	G	A	G	C	G	G	C	C	G	T	G	G	T	G	A	C	C	G	C	G	G	I	C	G	C	G	G	G	G	G	C	D	G	A	A	T	A	T	C	G	C	1
10	I	G	T	C	C	C	C	G	C	T	C	C	G	A	G	C	G	G	C	C	G	T	G	G	T	G	A	T	C	G	C	C	G	I	C	G	C	G	G	G	G	G	C	D	G	A	A	T	A	C	C	G	C	46
11	I	G	T	C	C	C	C	G	C	T	C	C	G	A	G	C	G	G	C	C	G	T	G	G	T	A	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	C	D	G	G	A	T	A	C	C	G	C	25
12	I	G	T	C	C	C	C	G	C	T	C	C	G	A	G	C	G	G	C	C	G	C	G	G	T	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	C	D	G	A	A	T	A	C	C	G	C	2
13	I	G	T	C	C	C	C	G	C	T	C	C	G	A	G	C	G	G	C	C	G	C	G	G	T	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	T	D	G	G	G	C	C	C	C	G	C	1
14	I	G	T	C	C	C	C	G	C	T	C	C	G	A	G	C	G	G	T	C	G	C	A	G	A	G	A	C	C	C	C	C	G	I	C	G	C	G	G	G	G	G	C	I	G	G	G	C	C	C	C	G	C	1
15	I	G	T	C	C	C	C	G	C	T	C	C	G	A	G	C	A	G	C	C	G	T	G	G	T	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	C	D	G	A	A	T	A	C	C	G	C	2
16	I	G	T	C	C	C	C	G	C	T	C	C	G	A	G	T	G	G	C	C	G	T	G	G	T	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	C	D	G	G	A	T	A	C	C	G	C	1
17	I	G	T	C	C	C	C	G	C	T	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	T	D	G	G	G	C	C	C	C	G	C	1
18	I	G	T	C	C	C	C	G	C	C	C	C	G	A	G	C	G	G	C	C	G	T	G	G	T	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	C	D	G	G	A	T	A	C	C	G	C	1
19	I	G	T	C	C	C	C	G	C	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	C	D	G	A	A	T	A	C	C	G	C	5
20	I	G	T	C	C	C	C	G	C	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	C	D	G	G	G	T	A	C	C	G	C	1
21	I	G	T	C	C	C	C	G	C	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	T	D	G	A	A	T	A	C	C	G	C	2
22	I	G	T	C	C	C	C	G	C	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	T	D	G	G	A	T	A	C	C	G	C	1
23	I	G	T	C	C	C	C	G	C	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	T	D	G	G	G	T	C	C	C	G	C	2
24	I	G	T	C	C	C	C	G	C	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	T	D	G	G	G	C	A	C	C	G	C	2
25	I	G	T	C	C	C	C	G	C	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	T	D	G	G	G	C	C	C	C	G	C	727
26	I	G	T	C	C	C	C	G	C	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	T	I	G	G	A	T	A	C	C	G	C	1
27	I	G	T	C	C	C	C	G	C	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	T	I	G	G	G	T	A	C	C	G	C	1
28	I	G	T	C	C	C	C	G	C	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	T	I	G	G	G	C	C	C	C	G	C	190
29	I	G	T	C	C	C	C	G	C	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	C	G	C	C	G	I	C	G	C	G	A	G	G	G	T	I	G	G	A	T	A	C	C	G	C	1
30	I	G	T	C	C	C	C	G	C	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	C	G	C	C	G	I	C	G	C	G	A	G	G	G	T	I	G	G	G	C	C	C	C	G	C	240
31	I	G	T	C	C	C	C	G	C	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	C	G	C	C	G	I	C	G	C	G	A	G	A	G	T	D	G	G	G	C	C	C	C	G	C	3
32	I	G	T	C	C	C	C	G	C	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	C	G	C	C	G	I	C	G	C	A	G	G	G	G	T	D	G	G	G	C	C	C	C	G	C	1
33	I	G	T	C	C	C	C	G	C	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	C	G	C	C	G	I	C	G	T	G	G	G	G	G	T	I	G	G	G	C	C	C	C	G	C	1
34	I	G	T	C	C	C	C	G	C	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	C	G	C	C	G	I	C	A	C	G	G	G	G	G	T	D	G	G	G	C	C	C	C	G	C	1
35	I	G	T	C	C	C	C	G	C	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	C	G	C	C	G	I	T	G	C	G	G	G	G	G	T	D	G	G	G	C	C	C	C	G	C	2
36	I	G	T	C	C	C	C	G	C	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	C	G	C	C	G	D	C	G	C	G	G	G	G	G	T	D	G	G	G	C	C	C	C	G	C	7
37	I	G	T	C	C	C	C	G	C	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	C	G	C	C	C	I	C	G	C	G	G	G	G	G	T	I	G	G	G	C	C	C	C	G	C	1
38	I	G	T	C	C	C	C	G	C	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	T	G	C	C	G	I	C	G	C	G	G	G	G	G	T	D	G	G	G	C	C	C	C	G	C	1
39	I	G	T	C	C	C	C	G	C	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	G	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	T	D	G	G	G	C	C	C	C	G	C	31
40	I	G	T	C	C	C	C	G	C	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	G	C	C	G	C	C	G	I	C	G	C	G	G	G	G	A	T	D	G	G	G	C	C	C	C	G	C	1
41	I	G	T	C	C	C	C	G	C	C	C	C	G	G	G	C	G	G	C	C	A	C	G	G	A	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	T	D	G	G	G	C	C	C	C	G	C	1
42	I	G	T	C	C	C	C	G	C	C	C	C	G	G	G	C	G	A	C	C	G	C	G	G	A	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	T	I	G	G	G	C	C	C	C	G	C	2
43	I	G	T	C	C	C	C	G	C	C	C	C	G	G	T	C	G	G	C	C	G	C	G	G	A	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	T	D	G	G	G	C	C	C	C	G	C	30
44	I	G	T	C	C	C	C	G	C	C	C	T	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	C	C	C	C	G	I	C	G	C	G	G	G	G	G	C	I	G	A	G	C	C	C	A	G	C	10
45	I	G	T	C	C	C	C	G	C	C	C	T	G	G	G	C	G	G	C	T	G	C	G	G	A	G	A	C	C	C	C	C	G	I	C	G	C	G	G	G	G	G	C	I	G	A	G	C	C	C	A	G	C	2
46	I	G	T	C	C	C	C	G	C	C	C	T	G	G	G	C	G	G	C	T	G	C	G	A	A	G	A	C	C	C	C	C	G	I	C	G	C	G	G	G	G	G	C	I	G	A	G	C	C	C	A	G	C	2
47	I	G	T	C	C	C	C	G	C	C	G	C	T	G	G	C	G	G	T	C	G	C	A	G	A	G	A	C	C	C	C	C	G	I	C	G	C	G	G	G	G	G	C	I	G	G	G	C	C	C	C	G	C	1
48	I	G	T	C	C	C	C	G	T	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	T	I	G	G	G	C	C	C	C	G	C	1
49	I	G	T	C	C	C	C	A	C	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	C	G	C	C	G	I	C	G	C	G	A	G	G	G	T	I	G	G	G	C	C	C	C	G	C	1
50	I	G	T	C	C	C	T	G	C	C	G	C	T	G	G	C	G	G	T	C	G	C	A	G	A	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	C	I	G	G	G	C	C	C	C	G	C	2
51	I	G	T	C	C	C	T	G	C	C	G	C	T	G	G	C	G	G	T	C	G	C	A	G	A	G	A	C	C	C	C	C	G	I	C	G	C	G	G	G	G	G	C	I	G	G	G	C	C	C	C	G	C	311
52	I	G	T	C	C	C	T	G	C	C	G	C	T	G	G	C	G	G	T	C	G	C	A	G	A	G	A	C	C	C	C	C	G	I	C	G	C	G	G	G	G	G	C	I	G	G	G	C	C	C	C	C	C	1
53	I	G	T	C	C	C	T	G	C	C	G	C	T	G	G	C	G	G	T	C	G	C	A	G	A	G	A	C	C	C	C	C	G	I	C	G	C	G	G	G	G	G	T	D	G	G	G	C	C	C	C	G	C	1
54	I	G	T	C	C	C	T	G	C	C	G	C	T	G	G	C	G	G	T	C	G	C	A	G	A	G	A	C	C	C	C	C	G	I	C	G	C	G	G	G	G	G	T	I	G	G	G	C	C	C	C	G	C	1
55	I	G	T	C	C	C	T	G	C	C	G	C	T	G	G	C	G	G	T	C	G	C	A	G	A	G	A	C	C	C	T	C	G	I	C	G	C	G	G	G	G	G	C	I	G	G	G	C	C	C	C	G	C	2
56	I	G	T	C	C	A	C	G	C	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	T	D	G	G	G	C	C	C	C	G	C	1
57	I	G	T	C	T	C	C	G	C	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	T	D	G	G	G	C	C	C	C	G	C	2
58	I	G	T	T	C	C	C	G	C	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	C	G	C	C	G	I	C	G	C	G	A	G	G	G	T	I	G	G	G	C	C	C	C	G	C	1
59	I	G	C	C	C	C	C	G	C	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	C	G	C	C	G	I	C	G	C	G	A	G	G	G	T	I	G	G	G	C	C	C	C	G	C	1
60	I	A	T	C	C	C	C	G	C	T	C	C	G	A	G	C	G	G	C	C	G	C	G	G	T	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	C	D	G	A	A	T	A	C	C	G	C	2
61	D	G	T	C	C	C	C	G	C	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	C	G	C	C	G	I	C	G	C	G	G	G	G	G	T	I	G	G	G	C	C	C	C	G	C	2
62	D	G	T	C	C	C	C	G	C	C	C	C	G	G	G	C	G	G	C	C	G	C	G	G	A	G	A	C	C	G	C	C	G	I	C	G	C	G	A	G	G	G	T	I	G	G	G	C	C	C	C	G	C	42

Supplemental Table 3.4.1: **Blood-Seq phased *ABO* haplotypes for 1,140 individuals (2,280 *ABO* chromosomes).** Color code legend. ABO active site loci are highlighted in yellow with corresponding amino acids substitutions.

Putative sero	haplotypes (n)	chrm (n)
A hap	12	440
A weak hap	2	44
B hap	5	316
Cis-AB	1	2
O hap	38	1,462
Non-del O hap	3	15
Possible recombinant (LOF)	1	1
Total	62	2280

Supplemental Table 3.5: **ABO-Seq haplotype/subtype calls for training set 2.** Calls for 467 individuals included in the Blood-Seq dataset with corresponding serology. Samples highlighted in yellow are discorant.

BloodSeq ID	H 1	H 2	Subtype 1	Subtype 2	ABOSeq Call	Serology
102692	30	30	A	A	A	A
102701	25	30	O	A	A	A
102702	25	30	O	A	A	A
102706	25	28	O	A	A	A
102751	28	44	A	Non-del O	A	A
102833	25	28	O	A	A	A
102844	28	36	O	A	A	A
104878	25	42	O	A	A	A
104891	28	43	O	A	A	A
104915	25	30	O	A	A	A
104920	27	28	A	A	A	A
104929	11	30	O	A	A	A
104932	1	30	O	A	A	A
104934	1	30	O	A	A	A
104937	30	30	A	A	A	A

104939	28	31	O	A	A	A
104945	25	30	O	A	A	A
104951	25	28	O	A	A	A
104952	30	30	A	A	A	A
104953	1	30	O	A	A	A
104954	25	28	O	A	A	A
104961	25	30	O	A	A	A
104963	1	28	O	A	A	A
104968	1	30	O	A	A	A
104970	4	28	O	A	A	A
104973	25	30	O	A	A	A
104974	25	30	O	A	A	A
104977	25	28	O	A	A	A
104978	1	30	O	A	A	A
104980	11	28	O	A	A	A
104981	1	28	O	A	A	A
104982	25	30	O	A	A	A
104983	25	30	O	A	A	A
104985	25	28	O	A	A	A
104987	25	30	O	A	A	A
104993	28	43	O	A	A	A
104999	25	28	O	A	A	A
105002	28	28	A	A	A	A
105007	1	30	O	A	A	A
105013	4	30	O	A	A	A
105027	28	28	A	A	A	A
105028	1	28	O	A	A	A
105032	30	39	O	A	A	A
105034	11	30	O	A	A	A
105037	1	28	O	A	A	A
105038	25	28	O	A	A	A
105039	1	30	O	A	A	A
105040	1	28	O	A	A	A
105053	21	30	O	A	A	A

105056	1	30	O	A	A	A
105060	1	28	O	A	A	A
105064	19	28	O	A	A	A
106933	25	28	O	A	A	A
106941	15	30	O	A	A	A
106943	28	28	A	A	A	A
106946	1	30	O	A	A	A
106955	1	28	O	A	A	A
106956	1	30	O	A	A	A
106959	25	28	O	A	A	A
106962	1	30	O	A	A	A
106963	1	28	O	A	A	A
106965	25	28	O	A	A	A
106966	1	30	O	A	A	A
106967	2	30	O	A	A	A
106969	1	28	O	A	A	A
106970	28	30	A	A	A	A
106977	30	39	O	A	A	A
106981	1	30	O	A	A	A
106983	25	30	O	A	A	A
106984	30	30	A	A	A	A
106985	25	28	O	A	A	A
106986	28	28	A	A	A	A
106992	1	30	O	A	A	A
106993	1	28	O	A	A	A
106996	1	28	O	A	A	A
106998	28	30	A	A	A	A
107001	1	30	O	A	A	A
107003	1	26	O	A	A	A
107004	25	30	O	A	A	A
107013	28	57	O	A	A	A
107017	25	28	O	A	A	A
107023	1	30	O	A	A	A
107029	25	28	O	A	A	A

107032	25	30	O	A	A	A
107033	28	36	O	A	A	A
107038	30	35	O	A	A	A
107039	25	30	O	A	A	A
107040	25	30	O	A	A	A
107042	9	30	O	A	A	A
107045	1	30	O	A	A	A
107048	30	39	O	A	A	A
107049	30	30	A	A	A	A
107056	1	59	O	A	A	A
107058	25	30	O	A	A	A
107060	1	30	O	A	A	A
107063	25	30	O	A	A	A
107070	30	30	A	A	A	A
107071	1	30	O	A	A	A
107075	4	30	O	A	A	A
107078	25	30	O	A	A	A
107080	4	28	O	A	A	A
107081	28	30	A	A	A	A
107089	30	43	O	A	A	A
107270	4	30	O	A	A	A
107272	25	30	O	A	A	A
107292	25	28	O	A	A	A
107297	30	30	A	A	A	A
107301	4	30	O	A	A	A
107303	25	28	O	A	A	A
107304	1	28	O	A	A	A
107305	10	28	O	A	A	A
107317	28	46	A	Non-del O	A	A
107323	28	30	A	A	A	A
107330	1	28	O	A	A	A
107331	25	28	O	A	A	A
107332	28	30	A	A	A	A
107336	1	30	O	A	A	A

107340	25	30	O	A	A	A
107346	25	28	O	A	A	A
107347	25	28	O	A	A	A
107348	1	30	O	A	A	A
107356	25	28	O	A	A	A
107357	7	30	O	A	A	A
107359	1	28	O	A	A	A
107361	28	44	A	Non-del O	A	A
107363	30	30	A	A	A	A
107364	1	28	O	A	A	A
107366	25	30	O	A	A	A
107372	25	28	O	A	A	A
107373	30	30	A	A	A	A
107378	1	30	O	A	A	A
107379	25	28	O	A	A	A
107382	1	28	O	A	A	A
107383	25	30	O	A	A	A
107396	25	30	O	A	A	A
108989	25	28	O	A	A	A
109113	28	44	A	Non-del O	A	A
112755	1	28	O	A	A	A
113314	30	39	O	A	A	A
104898	1	62	O	A weak	A weak	A
104946	25	62	O	A weak	A weak	A
104969	25	62	O	A weak	A weak	A
105011	25	62	O	A weak	A weak	A
105022	4	62	O	A weak	A weak	A
105045	25	62	O	A weak	A weak	A
106957	4	62	O	A weak	A weak	A
106991	6	62	O	A weak	A weak	A
107307	8	62	O	A weak	A weak	A
107351	25	62	O	A weak	A weak	A
108964	25	62	O	A weak	A weak	A
107377	25	51	O	B	B	A

102718	25	28	O	A	A	A
107327	1	28	O	A	A	A
113320	25	25	O	O	O	A
102675	30	51	A	B	AB	AB
104925	30	51	A	B	AB	AB
104955	30	54	A	B	AB	AB
104958	28	51	A	B	AB	AB
104997	30	51	A	B	AB	AB
105003	30	51	A	B	AB	AB
105004	30	50	A	Cis-AB*	AB	AB
105006	51	62	B	A weak	AB	AB
105024	28	51	A	B	AB	AB
105026	28	51	A	B	AB	AB
105029	28	51	A	B	AB	AB
105035	30	51	A	B	AB	AB
105042	30	51	A	B	AB	AB
105059	30	51	A	B	AB	AB
106971	51	62	B	A weak	AB	AB
106999	30	51	A	B	AB	AB
107002	28	51	A	B	AB	AB
107012	30	51	A	B	AB	AB
107031	28	51	A	B	AB	AB
107037	30	51	A	B	AB	AB
107055	49	51	A	B	AB	AB
107083	28	51	A	B	AB	AB
107214	28	51	A	B	AB	AB
107321	28	51	A	B	AB	AB
108962	28	51	A	B	AB	AB
109110	30	51	A	B	AB	AB
102696	1	51	O	B	B	B
102737	1	51	O	B	B	B
102743	25	51	O	B	B	B
104904	1	51	O	B	B	B
104908	25	51	O	B	B	B

104909	51	51	B	B	B	B
104912	1	51	O	B	B	B
104914	1	51	O	B	B	B
104922	25	51	O	B	B	B
104933	44	51	Non-del	B	B	B
104936	25	51	O	B	B	B
104938	1	51	O	B	B	B
104944	51	52	B	B	B	B
104947	10	51	O	B	B	B
104950	25	51	O	B	B	B
104959	1	51	O	B	B	B
104966	25	51	O	B	B	B
104967	25	51	O	B	B	B
104971	25	51	O	B	B	B
104994	51	51	B	B	B	B
104996	25	51	O	B	B	B
105009	25	51	O	B	B	B
105012	25	51	O	B	B	B
105019	25	51	O	B	B	B
105023	25	51	O	B	B	B
105031	38	51	O	B	B	B
105033	25	51	O	B	B	B
105036	25	51	O	B	B	B
105041	1	51	O	B	B	B
105044	1	51	O	B	B	B
105050	25	51	O	B	B	B
105062	4	51	O	B	B	B
106935	25	51	O	B	B	B
106936	51	51	B	B	B	B
106937	1	51	O	B	B	B
106938	1	51	O	B	B	B
106939	25	51	O	B	B	B
106940	51	51	B	B	B	B
106944	1	51	O	B	B	B

106953	1	51	O	B	B	B
106960	1	51	O	B	B	B
106975	25	51	O	B	B	B
106976	25	51	O	B	B	B
106987	25	51	O	B	B	B
106988	25	51	O	B	B	B
106997	25	51	O	B	B	B
107009	25	51	O	B	B	B
107014	25	51	O	B	B	B
107019	25	51	O	B	B	B
107022	25	51	O	B	B	B
107024	39	51	O	B	B	B
107027	25	51	O	B	B	B
107028	25	51	O	B	B	B
107043	1	51	O	B	B	B
107047	25	51	O	B	B	B
107053	1	51	O	B	B	B
107059	1	51	O	B	B	B
107061	1	51	O	B	B	B
107073	25	51	O	B	B	B
107074	1	51	O	B	B	B
107079	25	51	O	B	B	B
107082	51	51	B	B	B	B
107100	1	51	O	B	B	B
107114	25	51	O	B	B	B
107118	25	51	O	B	B	B
107219	1	51	O	B	B	B
107267	1	51	O	B	B	B
107269	1	51	O	B	B	B
107291	25	51	O	B	B	B
107293	1	51	O	B	B	B
107296	43	51	O	B	B	B
107298	25	51	O	B	B	B
107299	25	51	O	B	B	B

107300	39	51	O	B	B	B
107311	25	51	O	B	B	B
107313	43	51	O	B	B	B
107318	11	51	O	B	B	B
107320	25	51	O	B	B	B
107325	1	51	O	B	B	B
107328	1	51	O	B	B	B
107329	25	51	O	B	B	B
107335	1	51	O	B	B	B
107337	1	51	O	B	B	B
107349	47	51	B	B	B	B
107354	1	51	O	B	B	B
107355	25	51	O	B	B	B
107362	1	51	O	B	B	B
107367	1	51	O	B	B	B
107370	1	51	O	B	B	B
107376	1	51	O	B	B	B
107380	25	51	O	B	B	B
107385	25	51	O	B	B	B
107387	1	51	O	B	B	B
107388	25	51	O	B	B	B
109038	25	51	O	B	B	B
109041	1	51	O	B	B	B
109107	51	51	B	B	B	B
113322	25	51	O	B	B	B
107374	25	51	O	B	B	B
107054	1	30	O	A	A	O
107326	10	39	O	O	O	O
109001	25	48	O	A	A	O
113321	11	62	O	A weak	A weak	O
102669	10	25	O	O	O	O
102683	4	11	O	O	O	O
102685	1	25	O	O	O	O
102708	1	25	O	O	O	O

102719	25	25	O	O	O	O
102744	25	25	O	O	O	O
102747	4	25	O	O	O	O
102774	1	1	O	O	O	O
102785	23	25	O	O	O	O
102791	1	1	O	O	O	O
102817	25	25	O	O	O	O
104901	25	25	O	O	O	O
104910	1	25	O	O	O	O
104911	25	25	O	O	O	O
104913	1	25	O	O	O	O
104916	1	43	O	O	O	O
104917	1	10	O	O	O	O
104918	1	25	O	O	O	O
104919	25	25	O	O	O	O
104921	25	25	O	O	O	O
104923	1	25	O	O	O	O
104924	25	39	O	O	O	O
104926	1	25	O	O	O	O
104927	10	43	O	O	O	O
104928	11	11	O	O	O	O
104930	1	25	O	O	O	O
104931	10	43	O	O	O	O
104935	1	25	O	O	O	O
104940	25	25	O	O	O	O
104941	1	25	O	O	O	O
104942	1	39	O	O	O	O
104943	25	25	O	O	O	O
104948	1	25	O	O	O	O
104949	25	25	O	O	O	O
104956	1	25	O	O	O	O
104957	1	25	O	O	O	O
104960	1	1	O	O	O	O
104962	1	39	O	O	O	O

104964	1	1	O	O	O	O
104965	10	10	O	O	O	O
104972	25	25	O	O	O	O
104975	1	25	O	O	O	O
104976	1	25	O	O	O	O
104979	2	31	O	O	O	O
104984	25	25	O	O	O	O
104986	25	25	O	O	O	O
104988	1	60	O	O	O	O
104989	10	25	O	O	O	O
104990	25	39	O	O	O	O
104991	1	1	O	O	O	O
104992	10	25	O	O	O	O
104995	1	25	O	O	O	O
104998	2	25	O	O	O	O
105000	2	25	O	O	O	O
105001	1	25	O	O	O	O
105005	1	25	O	O	O	O
105008	25	43	O	O	O	O
105010	1	1	O	O	O	O
105014	1	25	O	O	O	O
105015	1	25	O	O	O	O
105016	1	25	O	O	O	O
105018	10	10	O	O	O	O
105020	25	25	O	O	O	O
105021	1	25	O	O	O	O
105025	1	10	O	O	O	O
105030	1	25	O	O	O	O
105043	17	25	O	O	O	O
105046	25	25	O	O	O	O
105047	1	4	O	O	O	O
105048	1	25	O	O	O	O
105049	1	25	O	O	O	O
105051	1	25	O	O	O	O

105052	1	25	O	O	O	O
105055	25	25	O	O	O	O
105057	1	25	O	O	O	O
105058	1	10	O	O	O	O
105061	19	25	O	O	O	O
105063	1	36	O	O	O	O
106932	4	25	O	O	O	O
106934	1	43	O	O	O	O
106942	1	11	O	O	O	O
106945	1	25	O	O	O	O
106947	1	25	O	O	O	O
106948	25	39	O	O	O	O
106949	25	35	O	O	O	O
106950	1	57	O	O	O	O
106951	25	25	O	O	O	O
106952	10	36	O	O	O	O
106954	1	25	O	O	O	O
106958	1	10	O	O	O	O
106961	1	25	O	O	O	O
106964	1	25	O	O	O	O
106968	1	25	O	O	O	O
106972	1	25	O	O	O	O
106974	10	25	O	O	O	O
106978	1	10	O	O	O	O
106979	25	25	O	O	O	O
106980	1	1	O	O	O	O
106982	11	25	O	O	O	O
106989	1	1	O	O	O	O
106990	25	25	O	O	O	O
106994	25	25	O	O	O	O
106995	25	25	O	O	O	O
107000	1	11	O	O	O	O
107005	25	25	O	O	O	O
107006	1	1	O	O	O	O

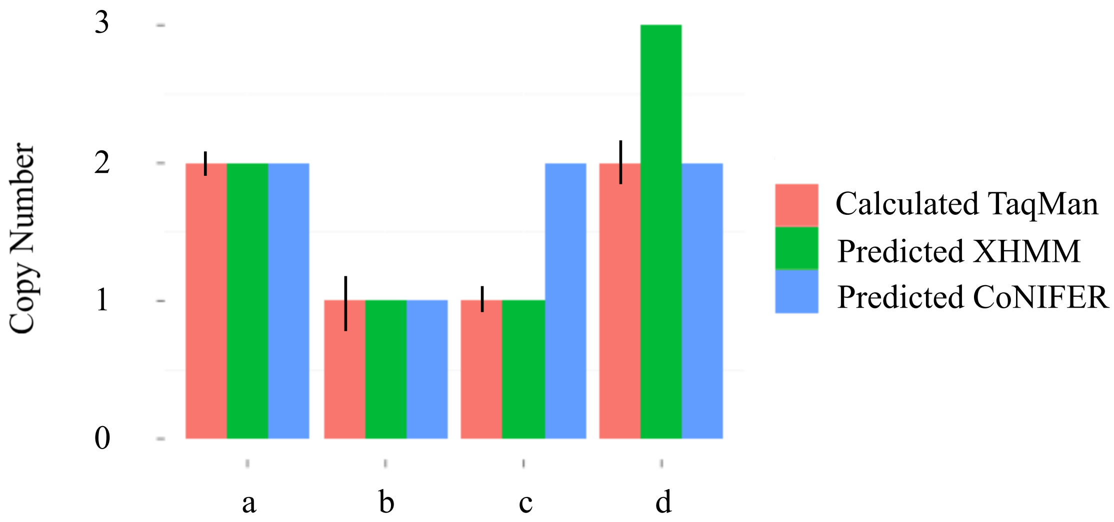
107007	1	25	O	O	O	O
107008	1	1	O	O	O	O
107010	1	25	O	O	O	O
107011	25	25	O	O	O	O
107015	1	25	O	O	O	O
107016	1	25	O	O	O	O
107018	1	25	O	O	O	O
107020	25	46	O	Non-del	O	O
107021	1	25	O	O	O	O
107025	2	25	O	O	O	O
107026	1	25	O	O	O	O
107030	25	25	O	O	O	O
107034	1	25	O	O	O	O
107035	25	43	O	O	O	O
107036	25	25	O	O	O	O
107041	2	43	O	O	O	O
107044	25	39	O	O	O	O
107046	1	25	O	O	O	O
107050	1	25	O	O	O	O
107052	25	25	O	O	O	O
107062	1	1	O	O	O	O
107064	5	43	O	O	O	O
107065	11	25	O	O	O	O
107066	1	44	O	Non-del O	O	O
107067	1	25	O	O	O	O
107068	25	25	O	O	O	O
107069	25	25	O	O	O	O
107077	25	25	O	O	O	O
107084	10	39	O	O	O	O
107101	4	11	O	O	O	O
107109	2	25	O	O	O	O
107112	25	25	O	O	O	O
107231	1	25	O	O	O	O
107266	1	25	O	O	O	O

107273	1	1	O	O	O	O
107289	25	25	O	O	O	O
107290	4	25	O	O	O	O
107294	25	25	O	O	O	O
107295	25	25	O	O	O	O
107302	1	25	O	O	O	O
107306	1	25	O	O	O	O
107308	1	25	O	O	O	O
107309	1	25	O	O	O	O
107310	1	1	O	O	O	O
107312	1	11	O	O	O	O
107314	25	39	O	O	O	O
107315	25	25	O	O	O	O
107316	1	1	O	O	O	O
107319	1	1	O	O	O	O
107322	1	1	O	O	O	O
107324	1	1	O	O	O	O
107333	1	1	O	O	O	O
107334	43	44	O	Non-del O	O	O
107339	1	25	O	O	O	O
107341	1	1	O	O	O	O
107343	25	25	O	O	O	O
107344	10	25	O	O	O	O
107345	1	25	O	O	O	O
107350	25	25	O	O	O	O
107352	1	25	O	O	O	O
107353	1	1	O	O	O	O
107358	1	10	O	O	O	O
107360	25	25	O	O	O	O
107365	1	25	O	O	O	O
107368	25	39	O	O	O	O
107369	1	25	O	O	O	O
107371	1	1	O	O	O	O
107375	25	25	O	O	O	O

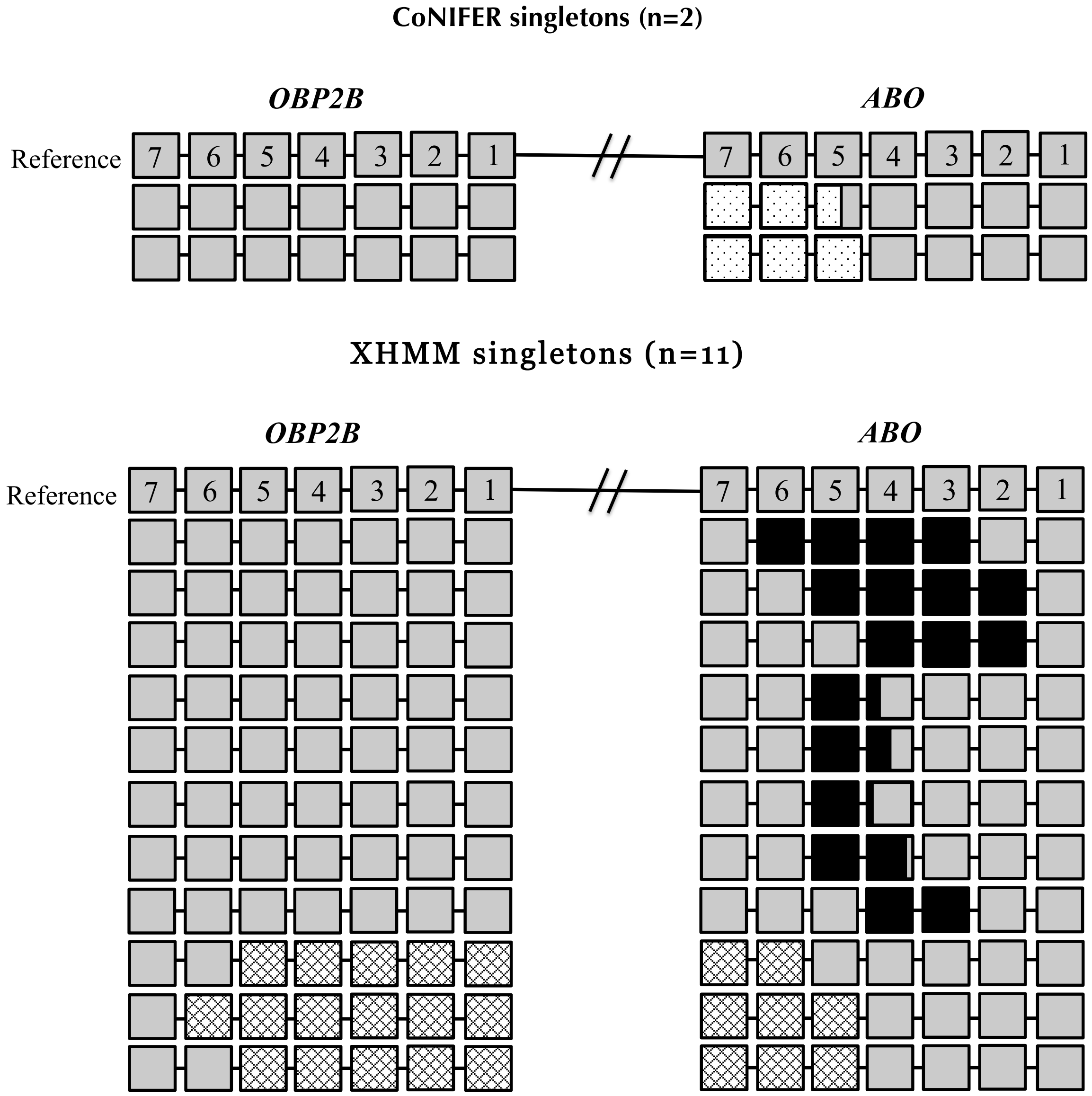
107381	25	25	O	O	O	O
107384	1	25	O	O	O	O
107386	1	11	O	O	O	O
108823	25	25	O	O	O	O
108933	25	25	O	O	O	O
109020	1	25	O	O	O	O
109048	25	25	O	O	O	O
112758	1	25	O	O	O	O
112759	1	4	O	O	O	O
113323	12	60	O	O	O	O

(THIS PAGE LEFT INTENTIONALLY BLANK)

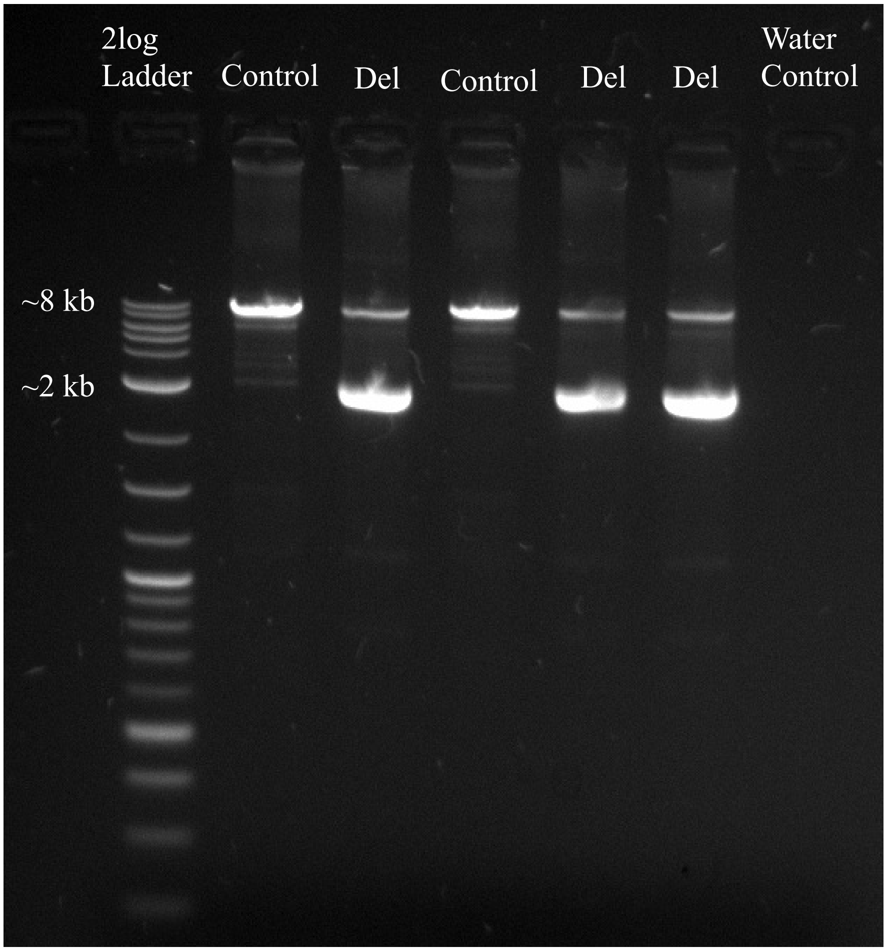
SUPPLEMENTAL FIGURES — CHAPTER 4: ANALYSIS OF EXOME-SEQUENCING DATASETS REVEALS STRUCTURAL VARIATION IN THE CODING REGION OF *ABO* IN INDIVIDUALS OF AFRICAN ANCESTRY



Supplemental Figure 4.1: **rtPCR validation of structural variation in *ABO* exon 7.** Vertical black lines indicate rtPCR error. (a) A control predicted to have two copies of the *ABO* gene by XHMM and CoNIFER. (b) An individual predicted by both XHMM and CoNIFER to have an *ABO* [E5-7] deletion confirmed by TaqMan. (c) An individual predicted by XHMM to have a deletion of only exons 5 and 6. However, manual curation in IGV indicated exon 7 was likely deleted, which was confirmed by TaqMan**;** (d) An individual predicted by XHMM to have a putative duplication of *ABO* [E5-7]. However, TaqMan detected two copies of *ABO*, indicating a false positive duplication call. These validation data highlight the importance of a combined call-set to robustly predict SVs in the *ABO* gene. In total we validated 16 of the 32 subjects predicted to have an *ABO* [E5-7] deletion.



Supplemental Figure 4.2: **Putative SVs discovered in *ABO* in 6,432 exomes using XHMM and CoNIFER (singletons).** *ABO* is a negative strand gene and is shown in negative strand orientation to reflect its spatial relationship with neighboring *OBP2B*. Putative SV discovered in *ABO* and neighboring downstream gene *OBP2B* using two read depth methods are shown. Each gray box represents intact diploid versions of 1-7 exons in the *ABO* and *OBP2B* gene. Dotted boxes represent *ABO* singleton deletions identified using CoNIFER (two events). Black boxes represent *ABO* duplications identified using XHMM (eight events). Cross-hatched boxes represent exons predicted to be deleted using XHMM (three events).



Supplemental Figure 4.3: **Detection of a ~5,800 bp *ABO* deletion in AA samples.** Using a long-range Takara PCR strategy spanning the predicted *ABO* deletion (products shown on a 1% agarose gel), we demonstrate the presence of the predicted wild type amplified PCR product (~8,000 bp) in our control samples (columns 2 and 4) and the presence of both the ~8,000bp product and a <3,000 bp product in samples with a predicted partial *ABO* gene deletion (column 3, 5, and 6). This is consistent with heterozygosity for a >4,800 bp deletion and validates the presence of *ABO* SV. Primer set sequence : Forward: 5′ TTCAGTGGCTCCCATGTCAC 3′ and Reverse: 5′ GGCTAGGCAGCTACTTGTCA 3′

SUPPLEMENTAL TABLE — CHAPTER 4: ANALYSIS OF EXOME-SEQUENCING DATASETS REVEALS STRUCTURAL VARIATION IN THE CODING REGION OF *ABO* IN INDIVIDUALS OF AFRICAN ANCESTRY

Supplemental Table 4.1: **Summary of *ABO* SV discovery using multiple read-depth based algorithms.** Discovery of *ABO* SV using CoNIFER and XHMM read depth algorithms (n= number of chromosomes/events; 2+ indicates a SV found in two or more individuals). For each SV, the number of samples with validated SVs / number of samples tested for the SV is shown. Each dash (-) indicates SV for which no validation was attempted.

	ABO Duplications		*ABO* Deletions	
SV Prevalence/Method	**Number**	**Validated**	**Number**	**Validated**
Singleton (XHMM)	8	-	3	-
Singleton (CoNIFER)	0	-	2	-
2+ Individuals (XHMM)	1 (n=2)	0/1	4 (n=32)	16/16
2+ Individuals (CoNIFER)	0	-	1 (n=5)	5/5
Intersection (XHMM + CoNIFER)	0	-	1 (n=5)	5/5