A. Filtering the Variant calls to identify the Candidates

The full text of the GATK call used for the MSSM sample is shown below. Only loci found in the file SureSelect_All_Exon_G3362_with_names_v2_hg19.nochr.bed were searched for variants. The minimum mapping quality for a read to be included and the minimum base quality were both set at ten (Phred scale). This is a lenient threshold, allowing some lower quality reads to pass through, but this was motivated by the desire to avoid the possibility that blood/cell line disagreements (or mismatches) could be attributed to the use of too-stringent criteria at this stage.

Three initial filters were applied to the variant calls as follows:

1. The number of reads that covered the site was required to be at least 20 for both the blood and the cell line. Note that these counts exclude any reads that did not meet the quality thresholds described above.
2. The Phred-scaled likelihood for each of the calls was required to be at least 30. The genotyper returns a set of three “likelihoods,” one for each of the three possible genotypes at the locus (homozygous reference, heterozygous, homozygous alternative), expressed relative to the best fit. Hence, the best fit will have a score of 0, the other two will be positive, with larger values indicating more confidence in the call.
3. Each call is assigned a label of either “pass” or “low quality” by the genotyper; only those which “pass” are retained.

Any candidate that does not pass all three of these tests is excluded from further consideration.

The final step taken was to apply the GATK Variant Recalibrator to detect low-quality calls. This method can be summarized as follows: Using sites which are known to hold common variants (e.g., they are present in HapMap) a Gaussian mixture model was trained on a suite of quality metrics. The motivation behind this approach is that most of these calls will be correct, and that they will largely be of high quality, hence the quality metrics at these sites are reflective of what one would expect of those that yield correct calls. This model was then applied to the entire set of calls, and a score (the Variant Quality Score Log Odds, or VQSLOD) is assigned to each. The VQSLOD is interpreted as the log odds ratio that a true variant exists versus no variant. The commands used for the MSSM sample are shown in the Supplementary Material.

B. Results of Filtering

Table 1 summarizes the results of the preliminary calls. The second row gives the number of loci for which at least one of the sequences yielded a non-reference call. For

instance, for the MSSM data, there are twelve variant call files (vcfs), two for each of the six subjects. There were 63,222 loci for which at least one of these twelve sequences was called a variant. This creates 6 × 63,222 = 379,332 total blood/cell line pairs for which calls are made. (Calls are made even in the case the individual is homozygous reference.) Of these, 183,973 (48.5%) were such that either the blood or cell line call included a variant. This high percentage can be explained by the fact that many of these sites are common variants. Of these, 16,847 (9.2%) showed disagreement between the blood and cell line calls. The Vanderbilt subjects yielded similar percentages with 38,947 disagreements.

	MSSM	Vanderbilt
Number of Subjects	6	10
Loci at which variant called on at least one subject	63,222	75,232
Blood or cell line called with non-reference base	183,973	327,728
Number for which blood and cell line calls differ	16,847	38,947

Table 1: Basic Properties of the Sample

These 16,847 + 38,947 = 55,794 instances were then passed through the three first-level filters. Any candidate that did not pass all three of these tests was excluded from further consideration. There were 15,099 sites remaining after this step.

	MSSM	Vanderbilt
Number for which blood and cell line calls differed	16,847	38,947
Number for which depth was at least 20 on both blood and cell line	8,918	25,472
Number for which Phred-scaled likelihood of both calls was at least 30	3,883	13,131
Number characterized as a “PASS” by GATK	15,590	36,199
Number that satisfied all three criteria	3,294	11,805

Table 2: Basic filters applied to the calls.

A larger percentage of the Vanderbilt calls survive this filter, in large part because overall depth for the Vanderbilt subjects was higher, and hence there was more confidence in the calls. If the same filters are applied to the cases of blood/cell line agreement, there are 806,233 pairs remaining implying a preliminary 1.8% disagreement rate.

C. GATK Commands

The following is the command used to call the variants.

```
   java -Xmx8g \
       -jar /Users/Shared/GATK/GenomeAnalysisTK.jar \
       -R /Users/Shared/References/ForGATK/b37/human_g1k_v37.fasta \
       -T UnifiedGenotyper \
       -I /Users/Shared/Autism/BloodCell/MSSM/BAMs/004ACL_markdup.bam \
       -I /Users/Shared/Autism/BloodCell/MSSM/BAMs/004AC_markdup.bam \
       -I /Users/Shared/Autism/BloodCell/MSSM/BAMs/007ACL_markdup.bam \
       -I /Users/Shared/Autism/BloodCell/MSSM/BAMs/007AC_markdup.bam \
       -I /Users/Shared/Autism/BloodCell/MSSM/BAMs/009ACL_markdup.bam \
       -I /Users/Shared/Autism/BloodCell/MSSM/BAMs/009A_markdup.bam \
       -I /Users/Shared/Autism/BloodCell/MSSM/BAMs/201ACL_markdup.bam \
       -I /Users/Shared/Autism/BloodCell/MSSM/BAMs/201AC_markdup.bam \
       -I /Users/Shared/Autism/BloodCell/MSSM/BAMs/8CL_markdup.bam \
       -I /Users/Shared/Autism/BloodCell/MSSM/BAMs/8_markdup.bam \
       -I /Users/Shared/Autism/BloodCell/MSSM/BAMs/9CL_markdup.bam \
       -I /Users/Shared/Autism/BloodCell/MSSM/BAMs/9_markdup.bam \
       -L
/Users/Shared/ExonBEDfiles/SureSelect_All_Exon_G3362_with_names_v2_hg19.nochr.b
ed \
       -B:dbsnp,VCF /Users/Shared/dbSNPdatabase/dbsnp_132.b37.vcf \
       -A AlleleBalance \
       -A DepthOfCoverage \
       -A FisherStrand \
       -A LowMQ \
       -A ReadPosRankSumTest \
       -BTIMR UNION \
       -glm SNP \
       -nt 7 \
       -im ALL \
       -gt_mode DISCOVERY \
       -hets 0.001 \
       -mmq 10 \
       -stand_call_conf 20.0 -stand_emit_conf 10.0 -dcov 500 \
       --min_base_quality_score 10 \
       -l INFO \
       -log MSSM_BloodCell.raw.log \
       -o ./MSSM_BloodCell.raw.vcf
```

The following is the command used to construct the Gaussian mixture model.

```
   java -Xmx8g \
       -jar /Users/Shared/GATK/GenomeAnalysisTK.jar \
       -T VariantRecalibrator \
       -R /Users/Shared/References/ForGATK/b37/human_g1k_v37.fasta \
```

```
    -B:input,VCF ./MSSM_BloodCell.raw.vcf \
    -B:dbsnp,VCF,known=true,training=false,truth=false,prior=8.0 \
         /Users/Shared/dbSNPdatabase/dbsnp_132.b37.vcf \
    -B:hapmap,VCF,known=false,training=true,truth=true,prior=15.0 \
         /Users/Shared/GATK/HapMap/hapmap_3.3.b37.vcf \
    -B:omni,VCF,known=false,training=true,truth=false,prior=12.0 \
         /Users/Shared/GATK/Omni/1000G_omni2.5.b37.vcf \
    -an QD \
    -an HaplotypeScore \
    -an MQRankSum \
    -an ReadPosRankSum \
    -an FS \
    -an MQ \
    -nt 8 \
    --maxGaussians 6 \
    -mode SNP \
    -log ./MSSM_BloodCell.Recal.log \
    -recalFile ./MSSM_BloodCell.Recal.recal \
    -tranchesFile ./MSSM_BloodCell.Recal.tranches \
    -rscriptFile ./MSSM_BloodCell.Recal.R
```

The following is the command used to apply the Gaussian mixture model to the observed variants, in order to obtain the VQSLOD scores.

```
java -Xmx8g \
    -jar /Users/Shared/GATK/GenomeAnalysisTK.jar \
    -T ApplyRecalibration \
    -R /Users/Shared/References/ForGATK/b37/human_g1k_v37.fasta \
    -B:input,VCF ./MSSM_BloodCell.raw.vcf \
    -recalFile ./MSSM_BloodCell.Recal.recal \
    -tranchesFile ./MSSM_BloodCell.Recal.tranches \
    -ts_filter_level 99.0 \
    -log ./MSSM_BloodCell.ApplyRecal.log \
    -o ./MSSM_BloodCell.Recal.vcf
```

D. Detailed Validation Results

Indiv.	Chr.	Position	Blood Call	Validation	Cell Call	Validation	Comment
3	X	90691093	AAC[C/C]GCC	TRUE	AAC[C/T]GCC	TRUE	MOSAIC
6	X	123538932	CCT[C/C]TGC	TRUE	CCT[C/T]TGC	TRUE	MOSAIC
7	4	123533877	TTC[A/A]CTT	TRUE	TTC[A/G]CTT	TRUE	MOSAIC
11	2	228882979	ACC[G/G]TTG	TRUE	ACC[G/A]TTG	TRUE	MOSAIC
11	8	42256366	AAA[T/T]CTC	TRUE	AAA[T/A]CTC	TRUE	MOSAIC
12	6	131188570	GCT[C/C]GAA	TRUE	GCT[C/T]GAA	TRUE	MOSAIC
12	20	13846164	CTG[T/T]ACA	TRUE	CTG[T/G]ACA	TRUE	MOSAIC

16	1	167403251	AAC[A/A]CTC	TRUE	AAC[A/G]CTC	TRUE	MOSAIC
16	12	109684116	CTG[C/C]GGG	TRUE	CTG[C/T]GGG	TRUE	MOSAIC
16	18	58038892	TGC[C/C]GGG	TRUE	TGC[C/T]GGG	TRUE	MOSAIC

Table 3: Group 0 Validation Results

Indiv.	Chr.	Position	Blood Call	Validation	Cell Call	Validation	Comment
1	4	69180008	GAC[G/G]AAG	TRUE	GAC[G/A]AAG	TRUE	MOSAIC
4	1	10706375	AGG[C/C]AAG	TRUE	AGG[C/A]AAG	TRUE	MOSAIC
4	6	32629194	AAC[G/G]CCA		AAC[G/A]CCA		Impossible primer design.
4	6	32629199	CAC[T/T]CAG		CAC[T/C]CAG		Impossible primer design.
4	10	98078170	CTT[C/C]AAG	TRUE	CTT[C/T]AAG	FALSE	All “C” reads on forward strand.
4	11	47644317	CAT[C/C]CAC	TRUE	CAT[C/A]CAC	TRUE	MOSAIC
4	11	56143156	AGA[C/C]CAT	TRUE	AGA[C/T]CAT	FALSE	Possibly an alignment problem? Note next candidate is adjacent.
4	11	56143158	ACC[A/A]TCA	TRUE	ACC[A/G]TCA	FALSE	See previous.
4	12	57662804	CTG[C/C]TGG	TRUE	CTG[C/G]TGG	FALSE	All “G” reads on the forward strand.
4	12	77424135	TAA[C/T]GAA	FALSE	TAA[C/C]GAA	TRUE	Almost all reads on the reverse strand.
4	22	50553671	GAC[G/G]GAA	TRUE	GAC[G/T]GAA	FALSE	All “T” reads on the reverse strand.
4	X	55249176	GAA[G/G]CAG	TRUE	GAA[G/A]CAG	TRUE	MOSAIC
5	11	67378012	CCC[G/T]CCT	FALSE	CCC[G/G]CCT	TRUE	All “T” reads on reverse strand.
5	16	89246670	GGA[T/T]GAG	TRUE	GGA[T/C]GAG	FALSE	All but one “C” read is highly skewed.
6	X	99661793	TCG[G/G]CCA	TRUE	TCG[G/A]CCA	TRUE	MOSAIC
6	X	118767373	GGA[C/C]GAA	TRUE	GGA[C/T]GAA	TRUE	MOSAIC
7	12	54803111	GTG[C/C]GCC	TRUE	GTG[C/T]GCC	TRUE	MOSAIC
7	20	210435	ACT[C/C]CAA	TRUE	ACT[C/T]CAA	TRUE	MOSAIC
8	5	137801559	GAG[G/G]AGA	TRUE	GAG[G/C]AGA	TRUE	MOSAIC
8	9	119106793	CTC[C/C]TTC	TRUE	CTC[C/T]TTC	TRUE	MOSAIC
9	2	198299652	TTA[G/G]GCT	TRUE	TTA[G/C]GCT	TRUE	MOSAIC, but noisy background
9	11	134241679	AAT[C/C]GTG	TRUE	AAT[C/T]GTG	TRUE	MOSAIC
11	2	116599816	ATC[T/T]GAG	TRUE	ATC[T/G]GAG	TRUE	MOSAIC
11	7	72397315	ACA[G/G]CTA	TRUE	ACA[G/A]CTA	TRUE	MOSAIC (very small additional peak)
11	19	56300185	TTA[A/A]GTA	TRUE	TTA[A/G]GTA	TRUE	MOSAIC
12	11	27389932	TTT[C/C]GGG	TRUE	TTT[C/T]GGG	TRUE	MOSAIC
15	11	116797981	AGC[T/T]TTT	TRUE	AGC[T/C]TTT	TRUE	MOSAIC
16	11	18313224	GTC[G/G]TTC	TRUE	GTC[G/T]TTC	TRUE	MOSAIC

Table 4: Group 1 Validation Results

Indiv.	Chr.	Position	Blood Call	Validation	Cell Call	Validation	Comment
1	1	65313240	GGG[T/T]CTA	TRUE	GGG[T/G]CTA	FALSE	All “G” reads on the forward strand, and these reads are skewed.
1	3	122354873	ACA[C/C]GAT	TRUE	ACA[C/T]GAT	TRUE	MOSAIC
1	4	46264063	GTC[C/C]ATG	TRUE	GTC[C/A]ATG	FALSE	Large total depth, and very few “A” reads on cell line.
1	9	135982045	CTT[C/C]CCA	TRUE	CTT[C/T]CCA	FALSE	All “T” reads on the forward strand.
1	11	62378536	CCG[C/A]CAC	FALSE	CCG[C/C]CAC	TRUE	Blood “A” reads are skewed, and few.
1	12	132623810	GAG[C/A]CCC	FALSE	GAG[C/C]CCC	TRUE	Blood “A” reads are skiewed, and few.
1	17	78931437	GGT[G/G]GTG	TRUE	GGT[G/T]GTG	FALSE	The forward “T” reads are skewed to the left, the reverse “T” reads are skewed to the right.
1	19	15278133	TGG[T/T]GCC	TRUE	TGG[T/G]GCC	FALSE	All “G” reads are on the forward strand.
1	19	51958638	AGG[T/T]GTG	TRUE	AGG[T/G]GTG	FALSE	All “G” reads are on the forward strand.
1	20	3928845	GGG[A/A]AAA	TRUE	GGG[A/G]AAA	FALSE	All but one “G” read is on the forward strand, and are highly skewed.
1	20	61459302	CCC[G/G]GAG	TRUE	CCC[G/T]GAG	FALSE	The forward “T” reads are skewed to the left, the reverse “T” reads are skewed to the right.
1	Y	14958873	TCC[C/C]CAA	TRUE	TCC[C/T]CAA	TRUE	MOSAIC
2	7	44613258	CGA[G/G]CCA	TRUE	CGA[G/T]CCA	FALSE	All “T” reads are skewed, and there are few.
2	9	8484293	TAT[G/G]ATT	TRUE	TAT[G/T]ATT	FALSE	Low proportion of “T” reads on cell line, and almost all are skewed.
2	9	101498842	TGG[C/A]ACT	FALSE	TGG[C/C]ACT	TRUE	The forward “A” reads are skewed to the left, the reverse “A” reads are skewed to the right.
2	9	101498847	CTG[C/A]TCT	FALSE	CTG[C/C]TCT	TRUE	The forward “A” reads are skewed to the left, the reverse “A” reads are skewed to the right. Also, note proximity to previous.
2	18	63511013	AAC[C/C]GGT	TRUE	AAC[C/T]GGT	TRUE	MOSAIC
2	19	54725995	GCT[A/A]TAG	FALSE	GCT[G/A]TAG	TRUE	This appears to be a mosaic on both blood and cell line. Large imbalance in “A” versus “G” calls, but consistent between blood and cell line.
2	21	35169715	TGG[A/A]GCA	TRUE	TGG[A/G]GCA	FALSE	All “G” reads are on the forward strand.
2	X	10107414	AAT[C/C]AAT	TRUE	AAT[C/A]AAT	FALSE	All “A” reads are on the forward strand.

3	1	2441539	TGA[G/G]GGA	TRUE	TGA[G/A]GGA	FALSE	
3	1	27995021	GTC[C/A]GAG	FALSE	GTC[C/C]GAG	TRUE	
3	15	75970085	GCA[G/T]CCT	FALSE	GCA[G/G]CCT	TRUE	
3	17	33459447	CTT[C/C]GGG	TRUE	CTT[C/T]GGG	FALSE	
3	19	55148043	GTC[T/T]GGG	TRUE	GTC[T/C]GGG	FALSE	
3	21	33077817	TTA[C/C]TTA	TRUE	TTA[C/T]TTA	FALSE	
4	1	16748468	CTC[T/T]GGA	TRUE	CTC[T/G]GGA	FALSE	
4	1	35870652	TGG[A/A]GCC	TRUE	TGG[A/G]GCC	FALSE	
4	1	46499441	TCC[A/A]GCT	TRUE	TCC[A/C]GCT	FALSE	
4	1	57398975	CTT[C/C]CAC	TRUE	CTT[C/T]CAC	FALSE	
4	1	94461686	TCC[A/A]GCA	TRUE	TCC[A/C]GCA	FALSE	
4	2	11780500	CAG[T/T]TCC	TRUE	CAG[T/G]TCC	FALSE	
4	2	242204025	TGG[A/A]CCA	TRUE	TGG[A/C]CCA	FALSE	
4	5	43173577	GTT[C/C]TTA	TRUE	GTT[C/T]TTA	FALSE	
4	6	11735807	CTG[A/A]TGT	TRUE	CTG[A/G]TGT	FALSE	
4	6	32632795	CTC[C/C]GTC	TRUE	CTC[C/T]GTC	FALSE	
4	6	32632850	GCG[G/G]GGA	TRUE	GCG[G/A]GGA	FALSE	
4	6	32634318	TCG[C/C]CAG		TCG[C/A]CAG		Failure
4	7	6461451	CGT[T/T]GCA	TRUE	CGT[T/G]GCA	FALSE	
4	7	44151241	CTG[T/T]GGG	TRUE	CTG[T/G]GGG	FALSE	
4	8	2020398	GCC[C/C]TGA	TRUE	GCC[C/A]TGA	FALSE	
4	9	100315518	CTG[T/T]TCC	TRUE	CTG[T/G]TCC	FALSE	
4	10	94822655	TGG[A/A]CGA	TRUE	TGG[A/G]CGA	FALSE	
4	10	97397146	TCC[A/A]CTG	TRUE	TCC[A/C]CTG	FALSE	
4	10	103772671	CAG[T/T]AGC	TRUE	CAG[T/C]AGC	TRUE	MOSAIC
4	11	68131385	AGG[T/T]GCT	TRUE	AGG[T/G]GCT	FALSE	
4	11	116706900	TCC[A/A]CCT	TRUE	TCC[A/C]CCT	FALSE	
4	13	21063460	GGG[A/A]GAC	TRUE	GGG[A/G]GAC	FALSE	
4	14	23564357	CTT[C/C]GGG	TRUE	CTT[C/T]GGG	FALSE	
4	14	23867949	CAG[T/T]TGC	TRUE	CAG[T/G]TGC	FALSE	
4	14	54416937	TGG[T/T]CAG	TRUE	TGG[T/C]CAG	FALSE	
4	14	57085352	CCG[T/T]CCT	TRUE	CCG[T/G]CCT	FALSE	
4	16	2498902	CTC[A/A]CGG	TRUE	CTC[A/C]CGG	FALSE	
4	16	4445265	CCC[A/A]CCC	TRUE	CCC[A/C]CCC	FALSE	

4	17	7707597	AGG[A/A]GCT	TRUE	AGG[A/G]GCT	FALSE	
4	17	21319369	ACC[G/G]AGG	TRUE	ACC[G/A]AGG	FALSE	
4	19	313490	CCC[A/A]CCT	TRUE	CCC[A/C]CCT	FALSE	
4	19	44252108	CCC[T/T]CCT	TRUE	CCC[T/C]CCT	FALSE	
4	19	50322605	GTG[A/A]ATG	TRUE	GTG[A/C]ATG	FALSE	
4	20	47592567	CTC[A/A]CGG	TRUE	CTC[A/C]CGG	FALSE	
4	21	46640865	AGC[T/T]GGG	TRUE	AGC[T/G]GGG	FALSE	
5	3	179527471	AAA[G/A]AAG	TRUE	AAA[G/G]AAG	FALSE	
5	5	140079996	TCC[A/C]AGC	FALSE	TCC[A/A]AGC	TRUE	
5	5	148712424	CAA[G/G]CCC	TRUE	CAA[G/T]CCC	FALSE	
5	6	7405175	TAT[C/C]ATT	TRUE	TAT[C/A]ATT	FALSE	
5	7	66648157	GAA[C/T]GTG	TRUE	GAA[T/T]GTG	FALSE	
5	11	117789342	GCC[T/C]GGG	TRUE	GCC[T/T]GGG	FALSE	
5	11	117789345	TGG[G/C]CTG	TRUE	TGG[G/G]CTG	FALSE	
5	15	43712957	GGA[G/A]AAG	FALSE	GGA[G/G]AAG	TRUE	
5	20	62658494	CGG[G/A]AAG	FALSE	CGG[G/G]AAG	TRUE	
6	6	32632598	CGG[A/T]ACG		CGG[A/A]ACG		Failure
6	16	135354	GAG[G/T]ACA	FALSE	GAG[G/G]ACA	TRUE	
6	22	24583727	GTG[G/C]CAG	FALSE	GTG[G/G]CAG	TRUE	
7	9	4663267	CAC[C/C]ATC	TRUE	CAC[C/A]ATC	TRUE	MOSAIC
7	9	72006567	CTC[G/G]CGG	TRUE	CTC[G/A]CGG	TRUE	MOSAIC
7	9	87425462	TTT[G/G]TTT	TRUE	TTT[G/A]TTT	TRUE	MOSAIC
7	10	15138591	CGA[C/C]TTT	TRUE	CGA[C/T]TTT	TRUE	MOSAIC
7	22	42912029	CTT[G/G]GGA	TRUE	CTT[G/T]GGA	FALSE	
8	1	54360125	AGC[G/G]ACT	TRUE	AGC[G/A]ACT	TRUE	MOSAIC
8	11	63672427	GTG[A/C]CCC	TRUE	GTG[A/A]CCC	FALSE	
8	20	10036141	TTT[C/T]TTT	TRUE	TTT[C/C]TTT	FALSE	
9	2	130912789	CAA[A/A]CAA	TRUE	CAA[A/C]CAA	FALSE	
9	5	840785	GAG[C/A]CAC	TRUE	GAG[C/C]CAC	FALSE	
9	6	75823294	CAT[G/G]TGC	TRUE	CAT[G/A]TGC	TRUE	MOSAIC
9	11	2416725	AGG[C/C]TGT	TRUE	AGG[C/A]TGT	FALSE	
9	15	48459567	ATC[C/C]TTA	TRUE	ATC[C/T]TTA	TRUE	MOSAIC
10	1	68942568	CTT[C/T]ATT	TRUE	CTT[C/C]ATT	FALSE	
10	5	60241142	GAA[A/A]GAG	TRUE	GAA[G/A]GAG	FALSE	

10	7	128483043	TGC[A/C]CCT	TRUE	TGC[A/A]CCT	FALSE	
10	10	3174669	CTC[A/A]CCG		CTC[A/C]CCG		Failure
10	19	46375892	GCC[A/A]GAA	TRUE	GCC[A/G]GAA	FALSE	
11	1	12921127	ATT[A/G]GAA	TRUE	ATT[A/A]GAA	FALSE	All G calls are in a single direction
11	2	139318349	AAC[A/A]TTT	TRUE	AAC[A/T]TTT	FALSE	All T calls are in a single direction
11	3	48464209	AAG[C/C]GAT	TRUE	AAG[C/T]GAT	TRUE	MOSAIC
12	3	100091564	GCG[G/T]TAC	FALSE	GCG[G/G]TAC	TRUE	All T calls are in a single direction
12	4	155506864	CAC[G/G]GGA	TRUE	CAC[G/A]GGA	TRUE	MOSAIC
12	10	117825111	GCC[C/C]GAC	TRUE	GCC[C/T]GAC	TRUE	All T calls are in a single direction
12	19	33467325	TTC[A/A]TTG	TRUE	TTC[A/G]TTG	TRUE	MOSAIC
13	17	21318770	CGC[G/A]GCG	FALSE	CGC[G/G]GCG	TRUE	All A calls are in a single direction
13	17	21318773	GGC[G/A]CAG	FALSE	GGC[G/G]CAG	TRUE	All A calls are in a single direction
14	7	26251843	GTT[C/T]TTT	FALSE	GTT[C/C]TTT	TRUE	Very slight bump for T on blood
14	15	20588693	CAA[C/C]ATC	FALSE	CAA[C/T]ATC	TRUE	Appears to be mosaic on both
15	14	96180232	AAG[C/C]TCC	TRUE	AAG[C/T]TCC	FALSE	All T calls are in a single direction
16	3	48505302	GGC[A/A]TCT	TRUE	GGC[A/G]TCT	FALSE	No clear explanation for mistake
16	6	151334926	TGG[T/T]GTT	TRUE	TGG[T/G]GTT	FALSE	All G calls are in a single direction
16	11	19251484	TGG[T/T]TTG	TRUE	TGG[T/G]TTG	TRUE	MOSAIC
16	14	81223208	ACA[A/A]TAA	TRUE	ACA[A/C]TAA	FALSE	All C calls are in a single direction

Table 5: Group 2 Validation Results

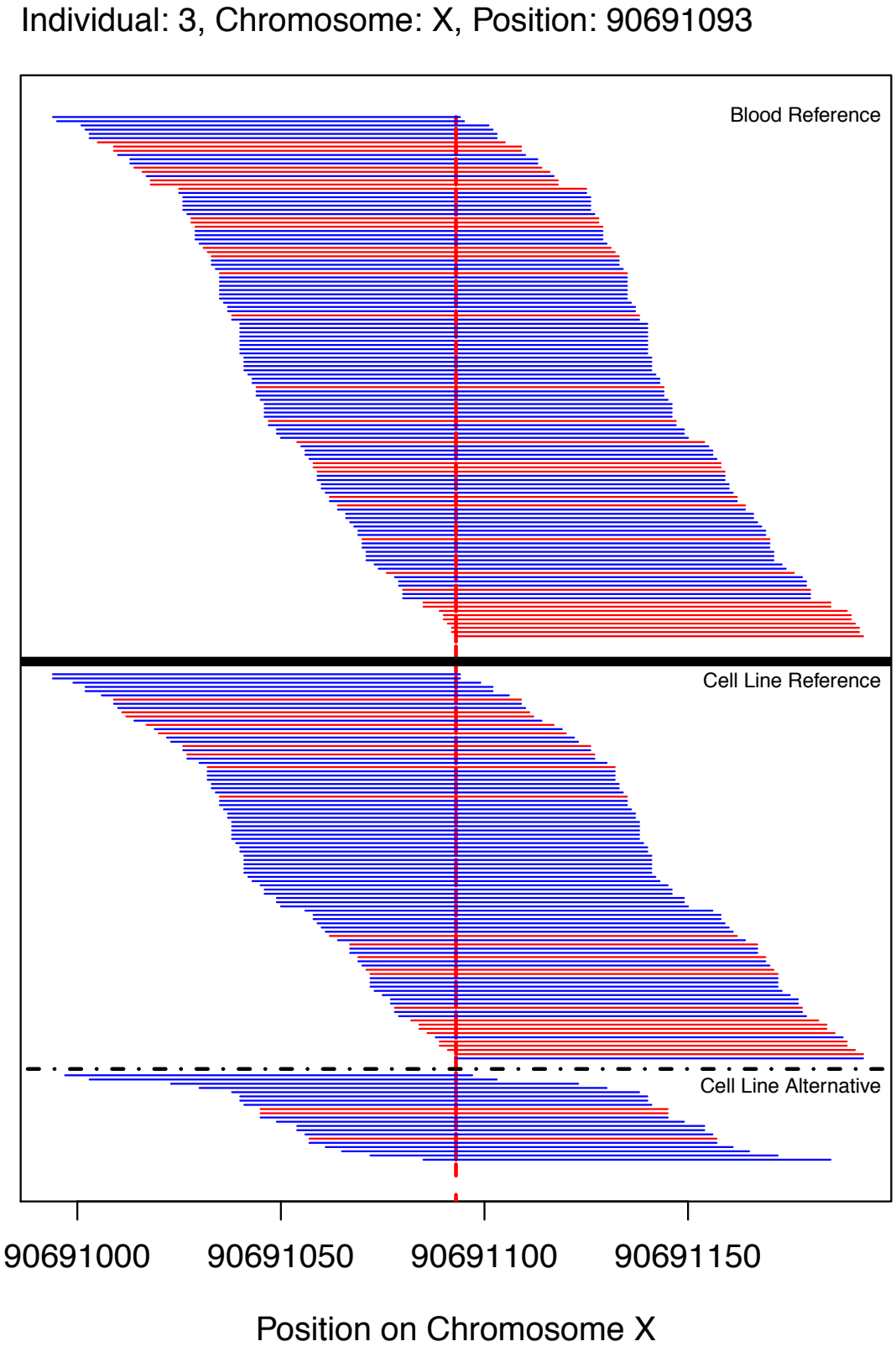
Indiv.	Chr.	Position	Blood Call	Validation	Cell Call	Validation	Comment
1	7	44099733	TGC[C/C]CTG	TRUE	TGC[C/A]CTG	FALSE	Strong Candidate
1	20	1895794	CGT[G/G]TTG		CGT[G/A]TTG		Strong Candidate; Failure
1	20	1895797	GTT[G/G]GTT		GTT[G/A]GTT		Strong Candidate; Failure
2	1	185137463	TAT[T/T]TAT		TAT[T/A]TAT		Strong Canddiate; Failure
2	3	16264288	AGA[G/G]CCT	TRUE	AGA[G/T]CCT	FALSE	Strong Candidate
2	7	912157	CAC[G/G]TAT	TRUE	CAC[G/T]TAT	FALSE	Strong Candidate
2	8	73480055	TCC[G/G]GAG	TRUE	TCC[G/A]GAG	FALSE	Strong Candidate
3	1	240371031	AGC[A/A]GGA		AGC[A/G]GGA		Strong Candidate; Failure
4	5	115230799	ACA[T/T]TAG	TRUE	ACA[T/A]TAG	FALSE	Strong Candidate
4	5	115230800	CAT[T/T]AGA	TRUE	CAT[T/C]AGA	FALSE	Strong Candidate
4	6	32712996	AGT[A/A]CAC	TRUE	AGT[A/T]CAC	FALSE	Strong Candidate
4	X	154721357	AAA[C/C]AGA		AAA[G/C]AGA		Strong Candidate; Failure
5	6	148861617	CTC[G/G]TTG	TRUE	CTC[G/A]TTG	TRUE	Strong Candidate; MOSAIC

6	19	51919894	CAC[A/A]CCC	TRUE	CAC[A/G]CCC	FALSE	Strong Candidate
7	19	54942009	CAG[C/C]GAG		CAG[C/A]GAG		Strong Candidate; Failure
8	12	51740409	CCA[T/T]AAA		CCA[T/G]AAA		Strong Candidate; Failure
8	12	51740410	CAT[A/A]AAG		CAT[A/G]AAG		Strong Candidate; Failure
9	6	32609368	CTT[A/A]AGT		CTT[A/G]AGT		Strong Candidate; Failure
9	8	28651299	CAA[A/A]AAA	TRUE	CAA[A/C]AAA	FALSE	Strong Candidate
11	2	27840357	TGG[A/G]AAA	FALSE	TGG[A/A]AAA	TRUE	
11	3	10114944	GAA[A/C]CAA	FALSE	GAA[A/A]CAA	TRUE	
11	5	180335598	AGG[T/T]GTT	TRUE	AGG[T/G]GTT	FALSE	
12	4	72319317	CAT[T/T]TAT	TRUE	CAT[T/G]TAT	TRUE	Strong Candidate; MOSAIC
12	16	88926388	GCC[C/C]TAC	TRUE	GCC[C/T]TAC	FALSE	
16	1	15808767	TCC[G/G]GGA		TCC[G/A]GGA		Strong Candidate; Failure
16	1	202743892	ATC[T/T]CAT	TRUE	ATC[T/C]CAT	FALSE	
16	5	120022221	GAT[T/T]CCT	TRUE	GAT[T/C]CCT	TRUE	Strong Candidate; MOSAIC
16	7	99361466	GTA[T/T]CAT	TRUE	GTA[C/T]CAT	FALSE	
16	8	7718187	AAT[T/G]AGA	TRUE	AAT[G/G]AGA	FALSE	
16	8	87229948	CGA[A/A]GAT		CGA[T/A]GAT		Strong Candidate; Failure
16	9	135153668	CAA[T/T]AGT	TRUE	CAA[C/T]AGT	FALSE	
16	19	52249211	AAC[G/G]CCA	TRUE	AAC[T/G]CCA	FALSE	Strong Candidate
16	21	15516948	TTT[T/T]TGG	TRUE	TTT[C/T]TGG	FALSE	

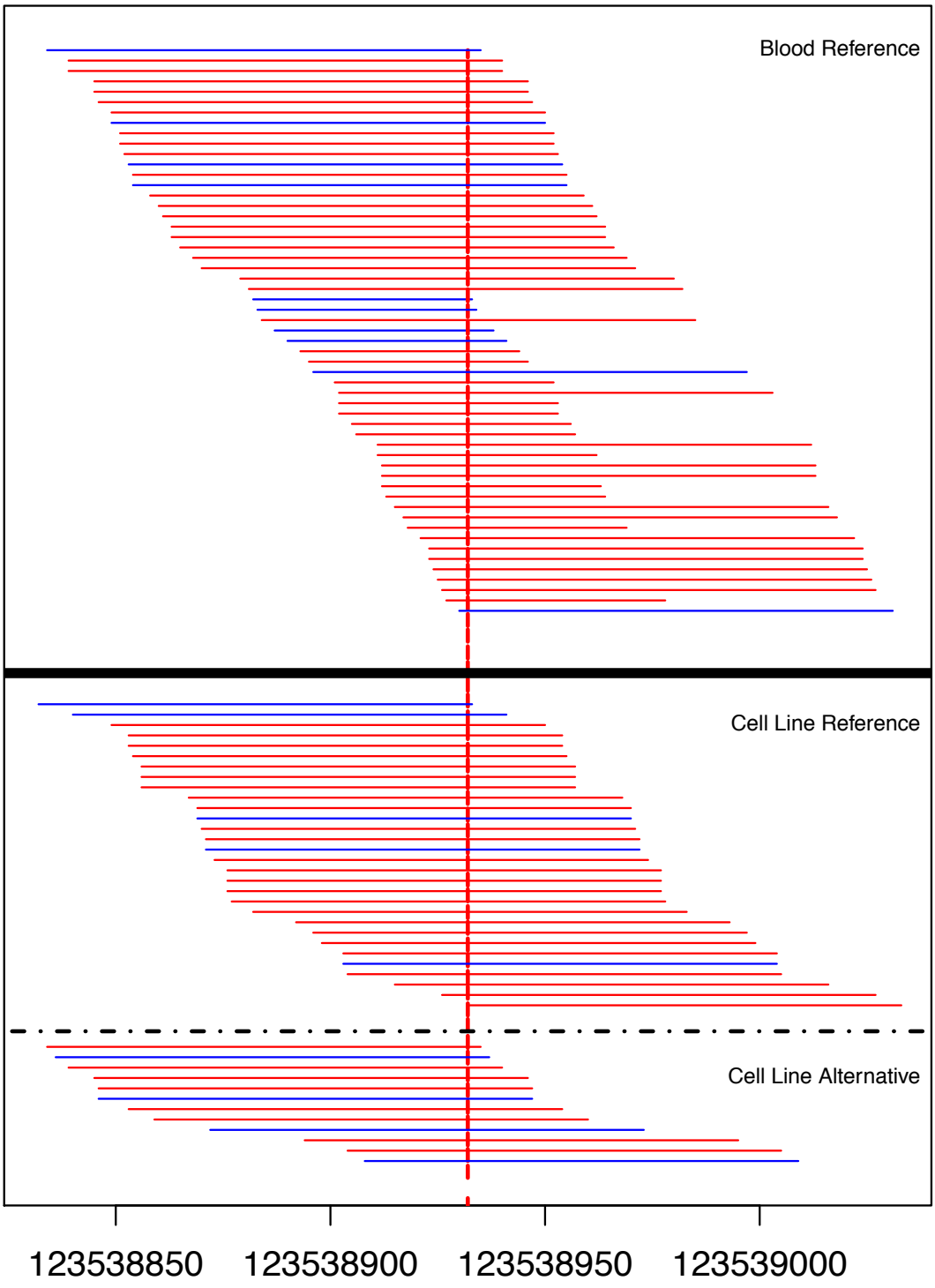
Table 6: Group 3 Validation Results

E. Pileup Diagrams of Group 0 Candidates

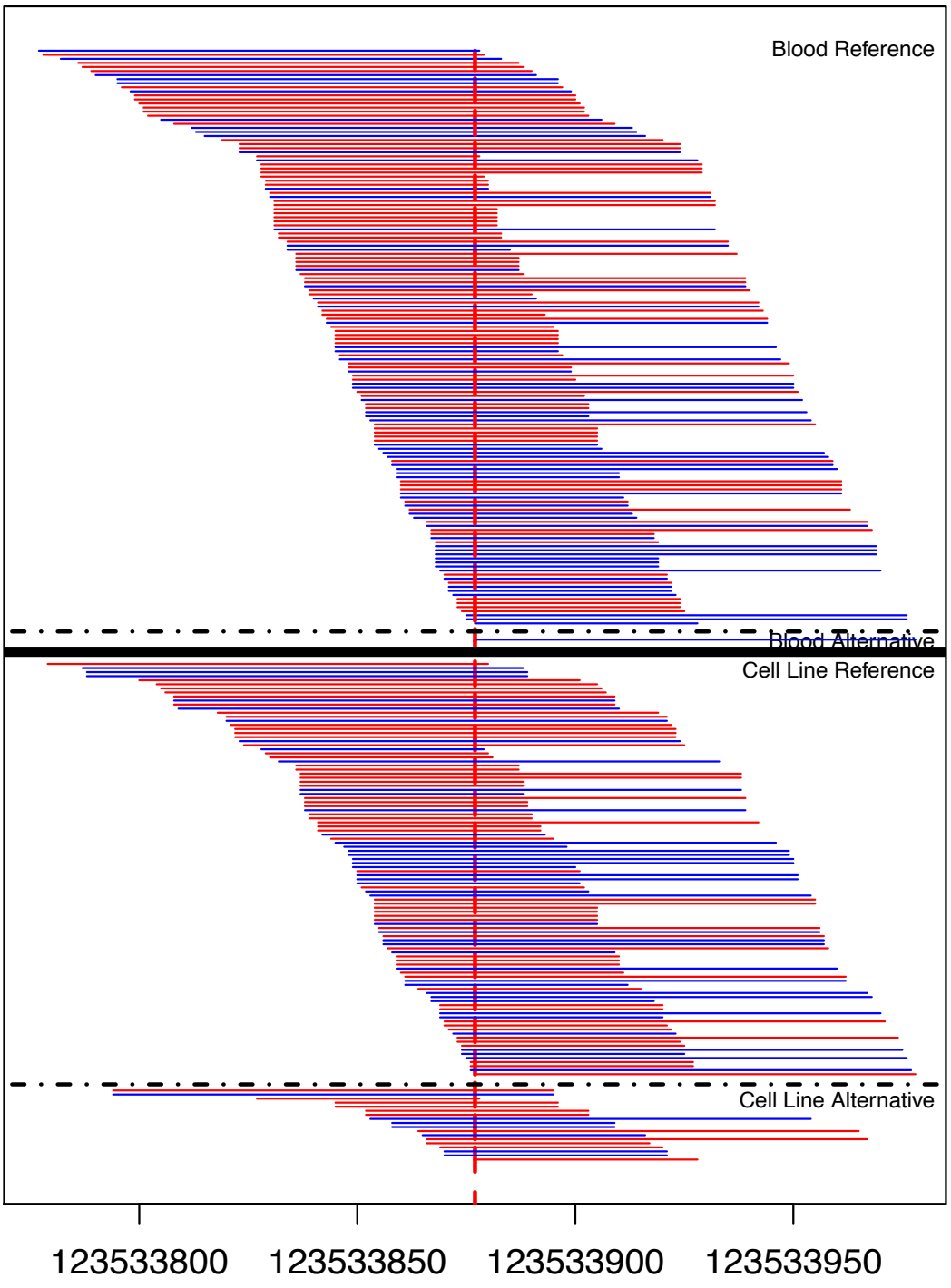
The plots on the following pages illustrate the sequencing results for the ten Group 0 Candidates. To read these plots: The range of bases spanned by a read is depicted as a horizontal line. The color of that line indicates the direction of the strand to which the read mapped: red indicates a read on the forward strand, blue is on the reverse. The vertical line passes through the target locus.


Individual: 3, Chromosome: X, Position: 90691093
Blood Reference
Cell Line Reference
Cell Line Alternative
90691000
90691050
90691100
90691150
Position on Chromosome X

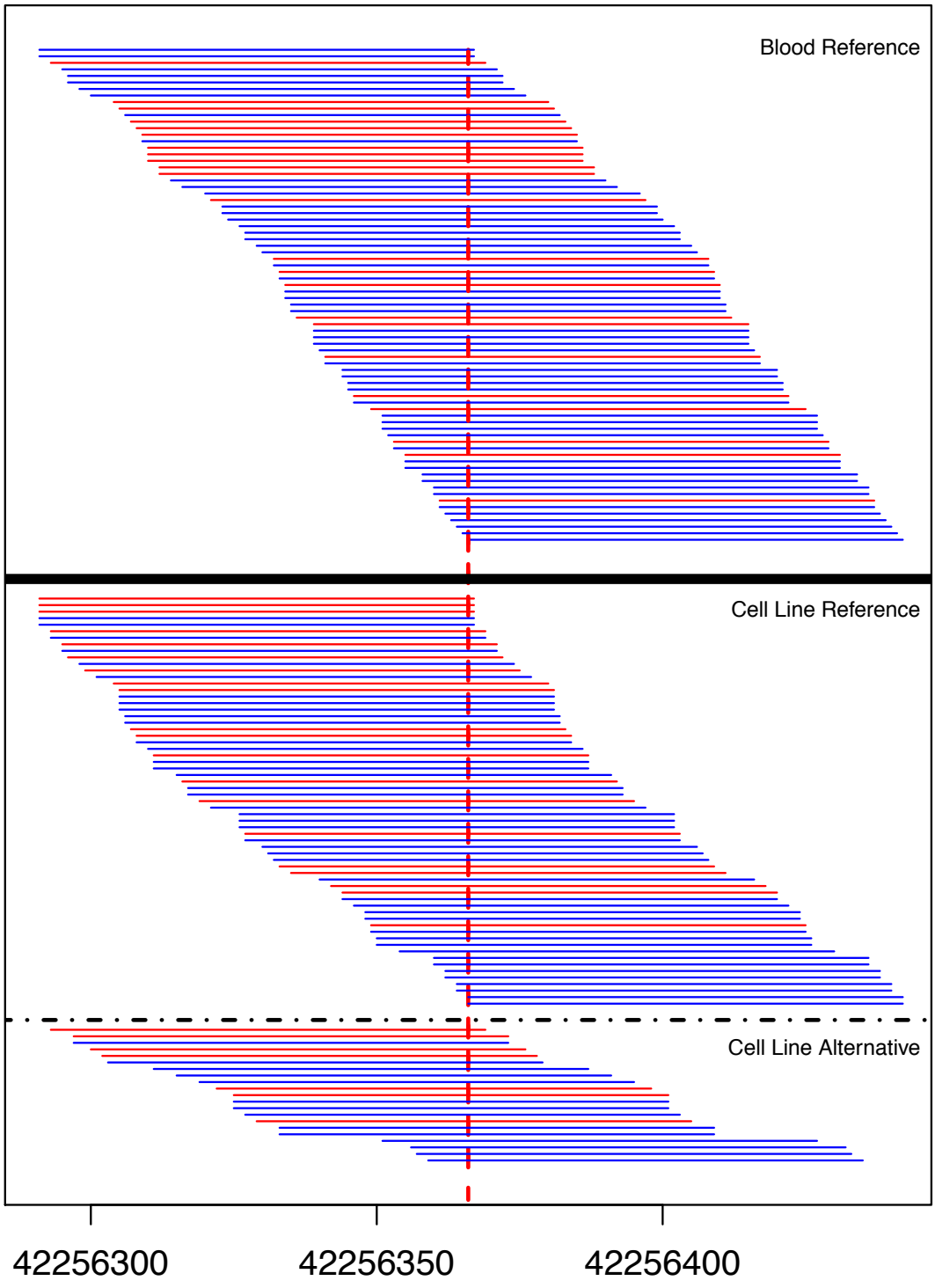
Individual: 6, Chromosome: X, Position: 123538932



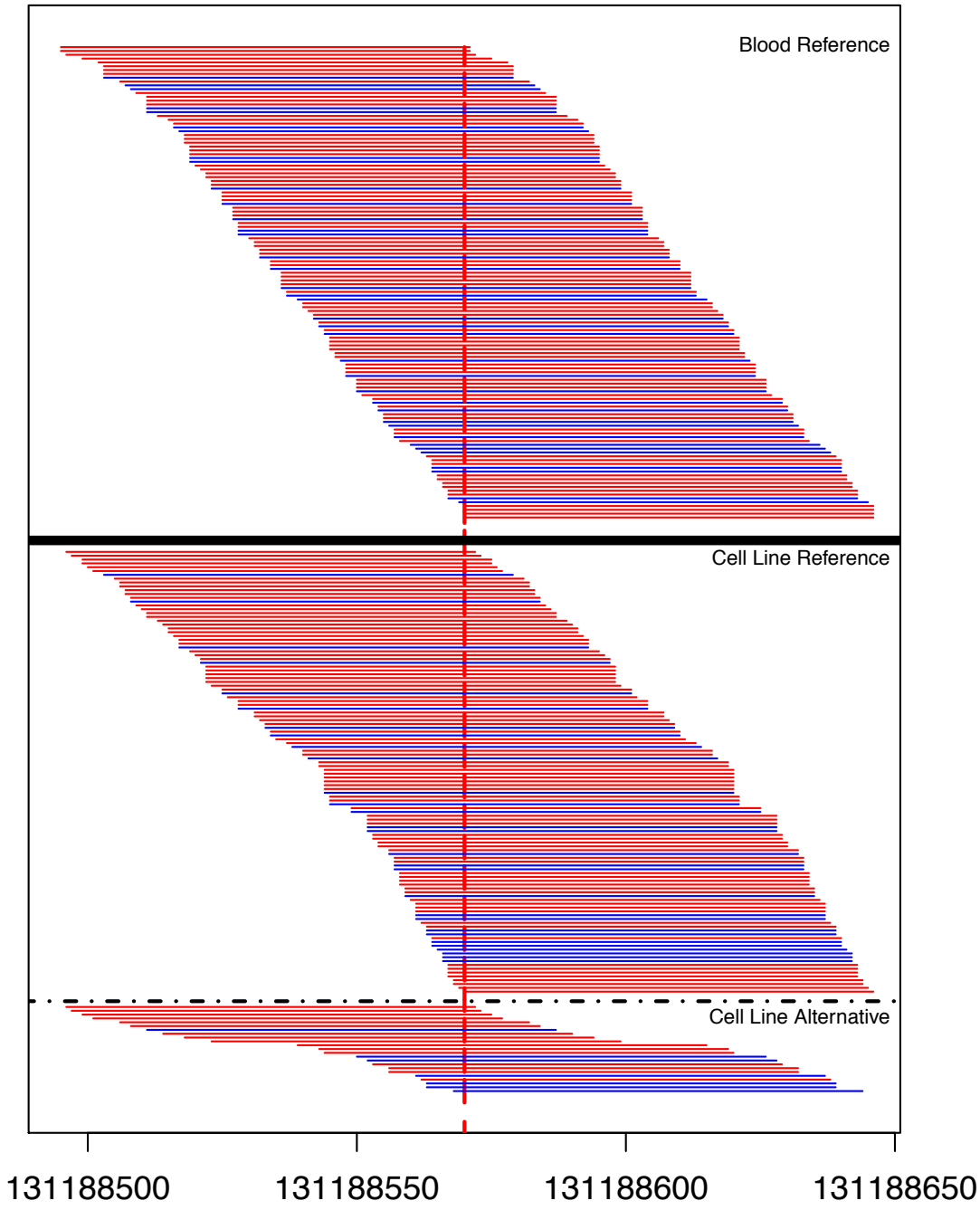
Individual: 7, Chromosome: 4, Position: 123533877



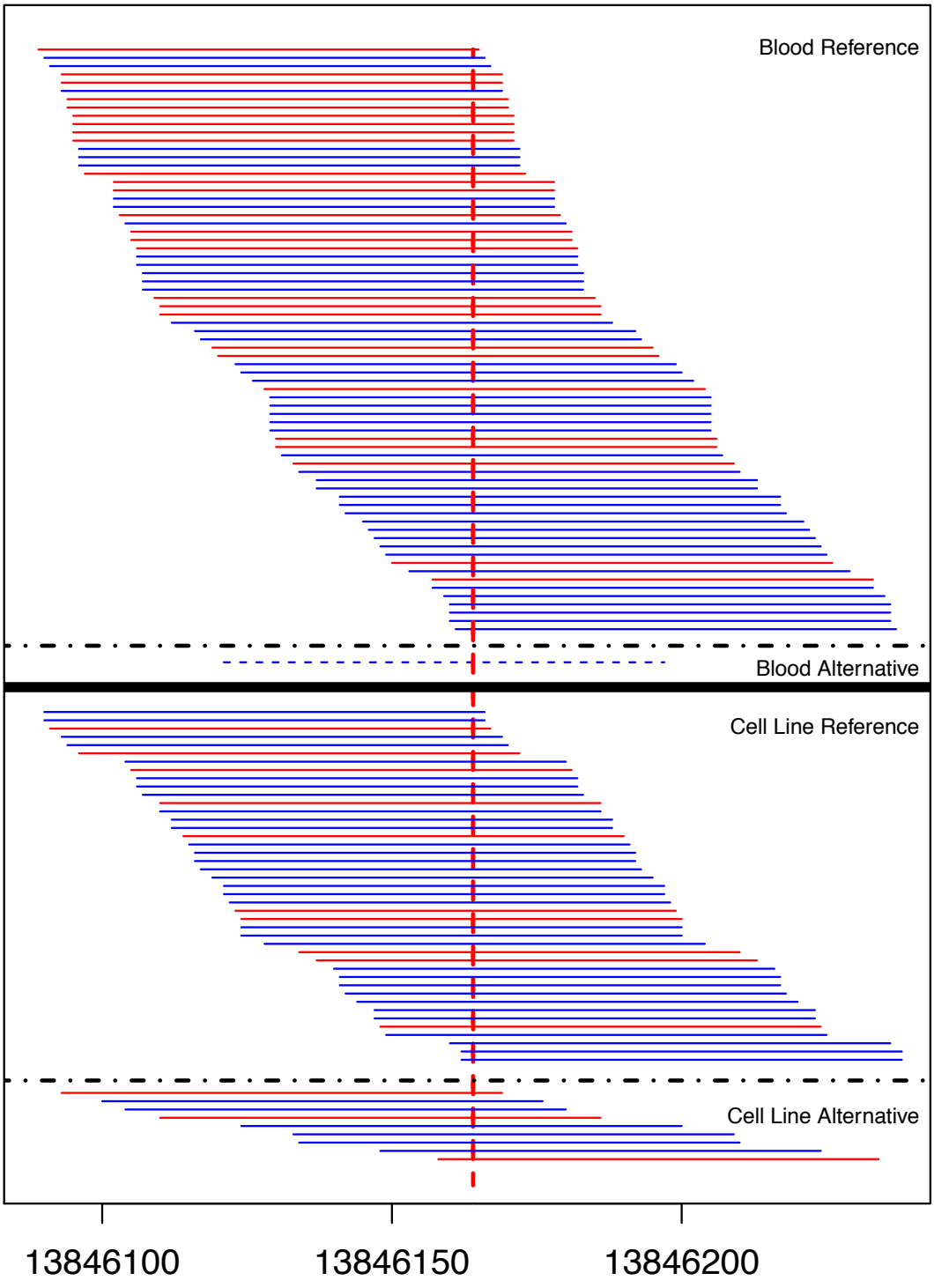
Individual: 11, Chromosome: 8, Position: 42256366



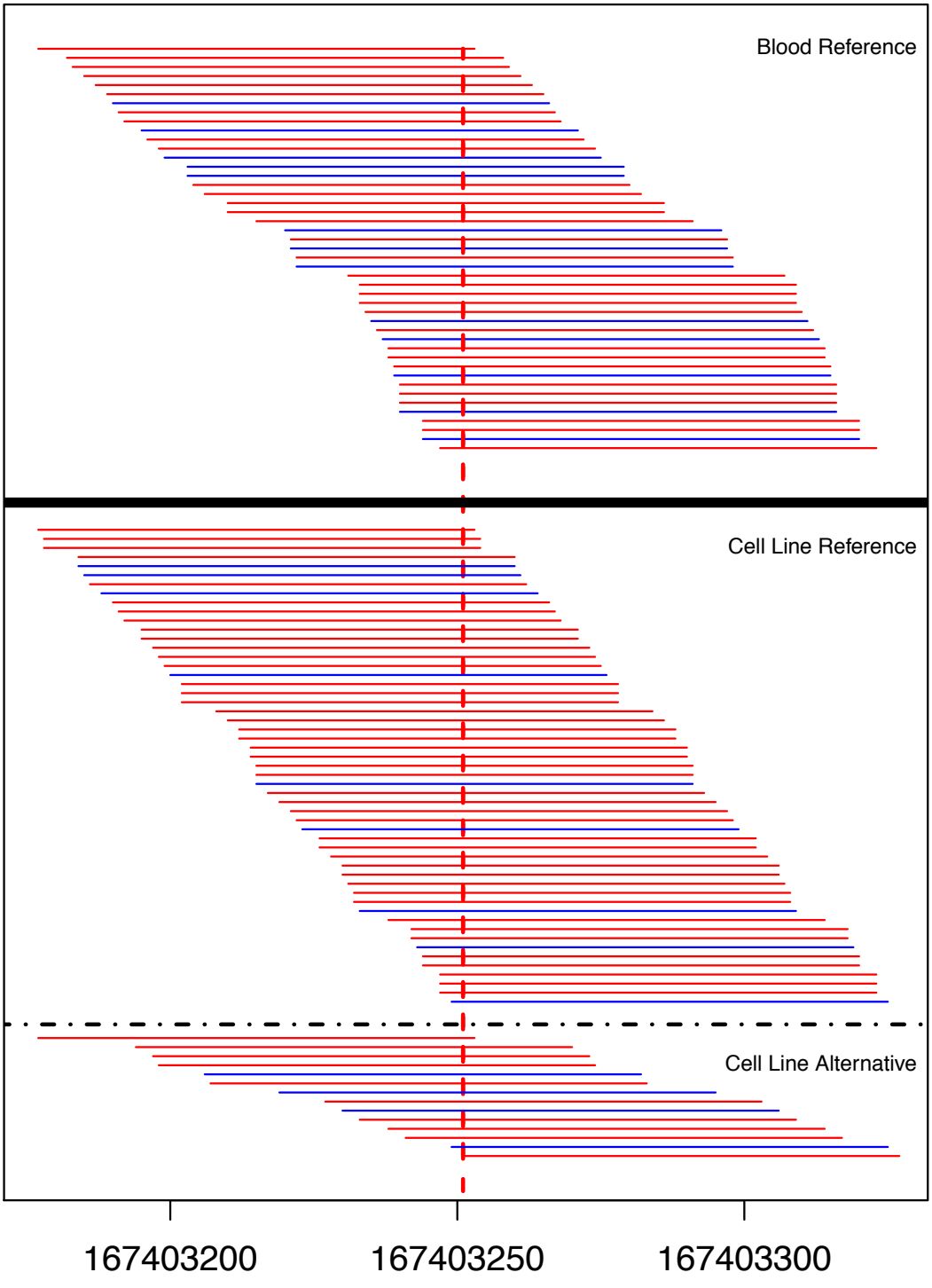
Individual: 12, Chromosome: 6, Position: 131188570



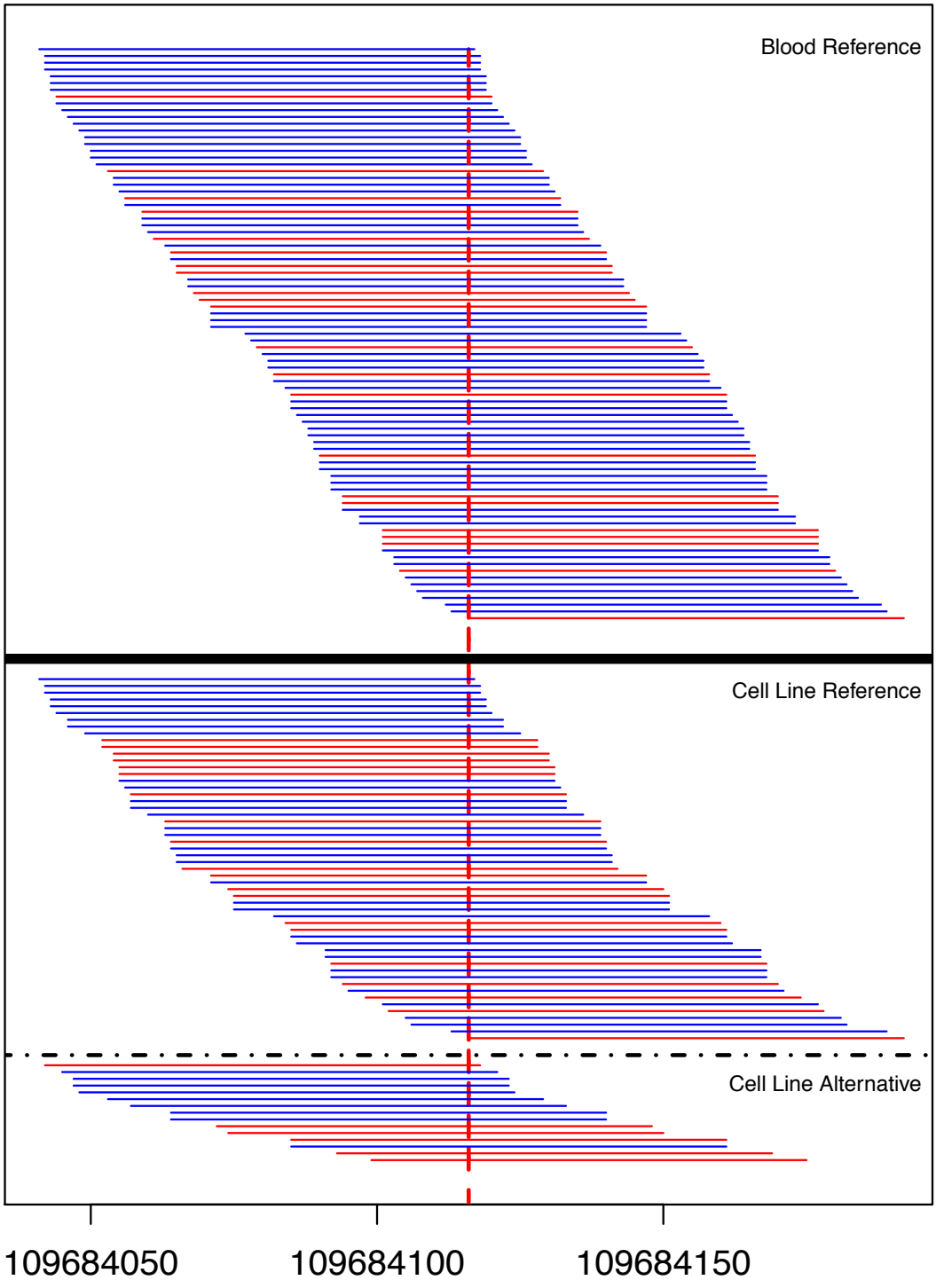
Individual: 12, Chromosome: 20, Position: 13846164



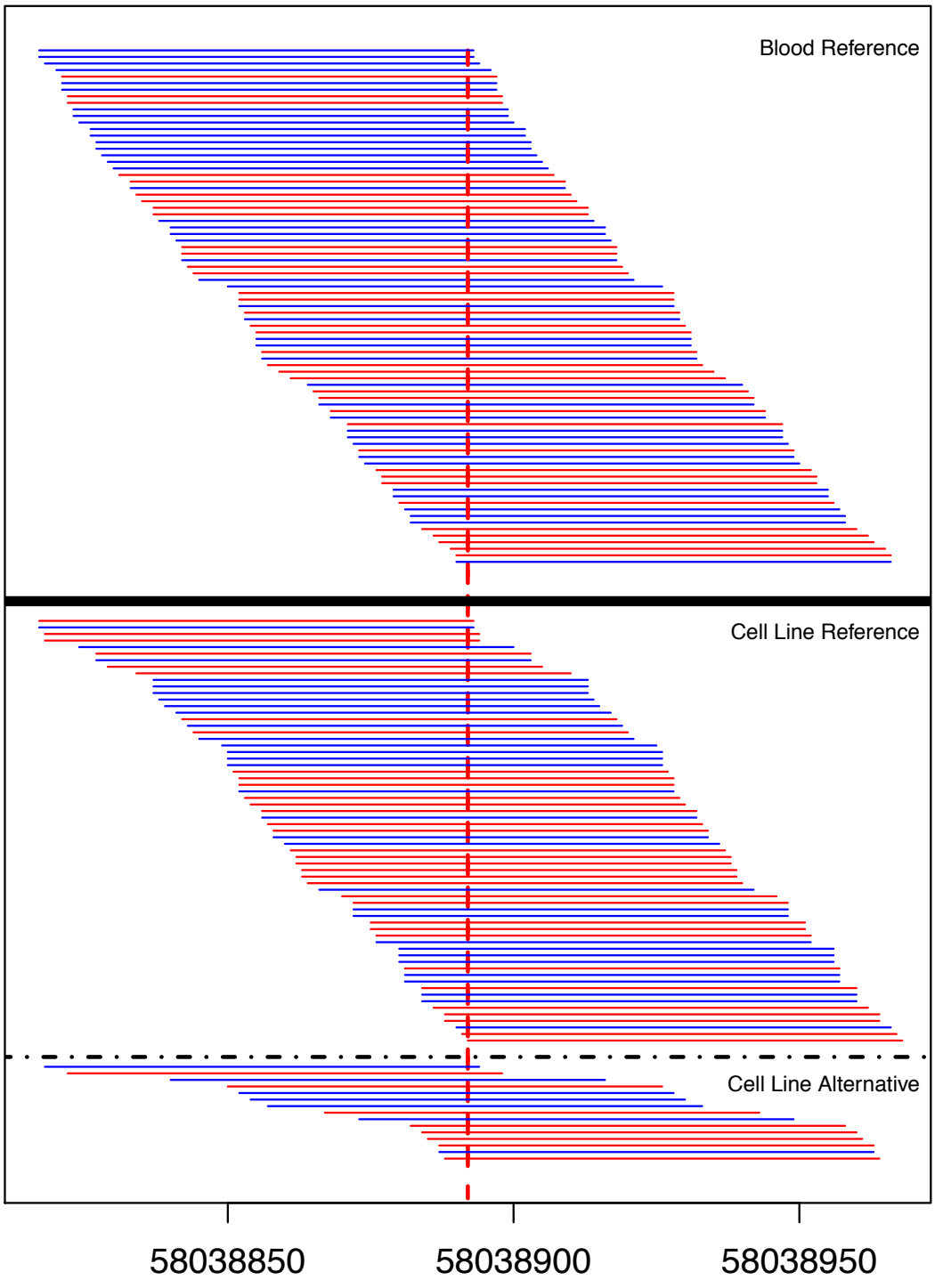
Individual: 16, Chromosome: 1, Position: 167403251



Individual: 16, Chromosome: 12, Position: 109684116



Individual: 16, Chromosome: 18, Position: 58038892



Individual: 11, Chromosome: 2, Position: 228882979