Table S1. Analysis parameters of the Whole Genome Sequencing Analysis Workflow with BWA, GATK and Manta.

Tool (version)	Parameter	Description	Remarks
Metadata fetch (v0)	/opt/samtools-1.6/samtools view -H input.cram | grep '^@RG' > with_metadata.header.sam	This tool sets metadata on input files using the @RG fields in the SAM/BAM/CRAM file.	Utility tool
md5sum (v0)		This tool invokes md5sum Unix utility to calculate MD5 sums of the input files.	
biobambam2 bamtofastq (v2.0.87)	F=input.pe_1.fastq F2=input.pe_2.fastq O=input.o_1.fastq O2=input.o_2.fastq S=s.fastq	Separate outputs for paired-end FASTQ files (1 and 2) and potential orphaned (1 and 2) and single-end reads.	
	inputformat=cram	Input format is CRAM.	
	reference=Homo_sapiens.GRCh38_15_plus_hs38d1.fa	Reference FASTA corresponding to the CRAM input.	
	filename=input.cram	The location and the filename of the input cram file.	
	outputdir=.	Output directory.	
FastQC (v0.11.5)	--adapters adapters.txt	Input file with adapters.	
SBG Untar fasta (v1.0)	tar -xf GRCh38_primary_assembly_plus_ebv_alt_decoy_hla.fasta.tar		Utility tool
SBG Fasta Indices (v1.0)	samtools faidx GRCh38_primary_assembly_plus_ebv_alt_decoy_hla.fasta picard.jar CreateSequenceDictionary R=GRCh38_primary_assembly_plus_ebv_alt_decoy_hla.fasta O=GRCh38_primary_assembly_plus_ebv_alt_decoy_hla.dict		Utility tool
SBG Prepare Intervals (v1.0)	python sbg_prepare_intervals.py --bed Homo_sapiens_primary_assembly_38_81_intervals.bed --mode 4	Based on the input file containing intervals, this tool creates per interval files and a file with ALT contigs in a single file.	Utility tool
GATK IndexFeatureFile (v4.0.12.0)		Indexing VCF and VCF.GZ input reference files - creating .IDX and .TBI indices, respectively.	Utility tool - only used once, not in the workflow
SBG FlattenLists (v1.0)			Utility tool
BWA-MEM Bundle:	-K 100000000	Reads chunk size to process.	This option is used to

BWA MEM (v0.7.17) Sambamba View (v0.6.7)		Used to achieve deterministic results in multi-threaded mode.	ensure that tool results do not depend on the number of processing threads.
	-Y	Use soft-clipping for supplementary reads.	
	-R '@RG\tID:value\tCN:SC\tLB:value\tPL:ILLUMINA\tPU:value\tSM:input\tDT:value\tDS:UK Biobank Study: UK Biobank Whole Genome sequencing study\tPG:value'	Read group header line.	This parameter is not in Broad Institute Best Practices, but is an optional permitted parameter in CCDG and not expected to affect results.
	-t $NT	Number of threads.	Not expected to affect results (see -K).
	GRCh38_primary_assembly_plus_ebv_alt_decoy_hla.fasta	Reference FASTA file.	Corresponding indices and the .alt file supplied alongside it.
	$FASTQ1 $FASTQ2	Sample inputs.	
	| sambamba view -f -t 32 bam -S	Conversion of native bwa mem SAM output (-S) to BAM (-f bam) using multiple threads (-t).	
Picard MarkDuplicates (v2.18.26)	INPUT=input.bam OUTPUT=input.bwa.bam METRICS_FILE=input.metrics	Sample inputs and outputs.	
	OPTICAL_DUPLICATE_PIXEL_DISTANCE=2500	Maximum offset between two duplicate clusters in order to consider them optical duplicates.	
	ASSUME_SORT_ORDER=queryname	Consider inputs query name sorted.	Output of BWA MEM is query-grouped, which is considered similar enough by Broad Institute's Best Practices.
	VALIDATION_STRINGENCY=SILENT	Validation level of SAM-type inputs. SILENT improves performance when processing BAMs if not all data needs to be decoded.	
Sambamba Sort (v0.5.9)	--nthreads=$NT --memory-limit=$MEM	Number of threads. Memory for sorting.	This step performs coordinate-sorting of BAM inputs.
	--out=input.bwa.sorted.bam input.bwa.bam	Sample inputs and outputs.	
Picard CollectAlignmentSummaryMetrics (v2.18.26)	INPUT=input.bwa.sorted.bam OUTPUT=input.bwa.sorted.summary_metrics.txt	Sample inputs and outputs.	

	REFERENCE_SEQUENCE=GRCh38_primary_assembly_plus_ebv_alt_decoy_hla.fasta	Reference FASTA file.	
	VALIDATION_STRINGENCY=SILENT	Validation level of SAM-type inputs. SILENT improves performance when processing BAMs if not all data needs to be decoded.	
VerifyBamID (v1.1.3)	--bam input.bwa.sorted.bam --bai input.bwa.sorted.bam.bai --out input.bwa.sorted	Sample inputs and outputs.	
	--vcf Standard-GRCh38_15_plus_hs38d1.vcf.gz	Input VCF file with allele frequencies.	
	--minAF 0.05	Minimum allele frequency of the markers to include.	
	--self	Only compare the ID-matching individuals between the VCF and BAM file.	
	--ignoreRG	Ignore read group level comparison and compare samples only.	
	--precise	Calculate the likelihood in log-scale for high-depth data (recommended when --maxDepth is greater than 20).	
	--maxDepth 500	Maximum read depth.	
	--minQ 20	Minimum base quality to include.	
	--noPhoneHome	Disable sending of usage statistics.	
GATK BaseRecalibrator (v4.0.12.0)	--input input.bwa.sorted.bam --output input.bwa.sorted.recal_data.grp	Sample inputs and outputs.	
	--known-sites Homo_sapiens_assembly38.dbsnp138.vcf --known-sites Mills_and_1000G_gold_standard.indels.hg38.vcf --known-sites Homo_sapiens_assembly38.known_indels.vcf	Known sites resources for recalibration.	
	--reference GRCh38_primary_assembly_plus_ebv_alt_decoy_hla.fasta	Reference FASTA file.	

	--intervals interval.bed	Interval to run on (parallelization option).	Broad Best Practices uses contig-grouping, CCDG allows parallelization in this way and permits autosomes only.
	--use-original-qualities	Use the base quality scores from the OQ tag	Broad Best Practices parameter; optional, permitted in CCDG.
GATK GatherBQSRReports (v4.0.12.0)	--input 1.grp ... --input n.grp --output input.bwa.sorted.recal_data.all.grp	This tool merges scattered BaseRecalibrator tool outputs.	Utility tool
GATK ApplyBQSR (v4.0.12.0)	--input input.bwa.sorted.bam --output input.bwa.sorted_interval.recalibrated.bam	Sample inputs and outputs.	
	--intervals interval.bed	Interval to run on.	
	--reference GRCh38_primary_assembly_plus_ebv_alt_decoy_hla.fasta	Reference FASTA file.	
	--bqsr-recal-file input.bwa.sorted.recal_data.all.grp	Input base recalibration table.	
	--static-quantized-quals 10 --static-quantized-quals 20 --static-quantized-quals 30	BQSR binning scheme.	This option compresses recalibrated base quality scores to significantly reduce output sizes with minimum impact on variant calling (CCDG, Broad Institute Best Practices).
	--add-output-sam-program-record	Adds a PG tag to created BAM files.	Broad Best Practices parameter, optional in CCDG
	--use-original-qualities	Use the base quality scores from the OQ tag.	Broad Best Practices parameter, optional in CCDG
MultiQC (v1.3)	-n input.multiqc_report --pdf	Output naming; PDF report	
Picard GatherBamFiles (v2.18.26)	INPUT=1.bam … INPUT=n.bam OUTPUT = input.gathered.bam	This tool merges BAM files after applying BQSR.	Utility tool. In Broad's Best Practices CREATE_INDEX is True. We do not create index in this tool as it doesn't work in our implementation.
SAMtools View (v1.9)	--output-fmt CRAM -h	Convert to CRAM and output header.	

	--reference GRCh38_primary_assembly_plus_ebv_alt_decoy_hla.fasta	Reference FASTA file.	
	--threads 7	Number of threads.	
	-o input.gathered.sorted.cram input.gathered.sorted.bam	Sample inputs and outputs.	
SAMtools Index (v1.9)	-c input.cram	This tool indexes input BAM or CRAM files.	
Manta (v1.4.0)	--bam input.gathered.sorted.bam	Sample input.	
	--referenceFasta GRCh38_primary_assembly_plus_ebv_alt_decoy_hla.fasta	Reference FASTA file.	
GATK HaplotypeCaller (v4.0.12.0)	--input input.bwa.sorted_chr15.recalibrated.bam --output input.bwa.sorted_interval.g.vcf	Sample inputs and outputs.	The optional Broad Best Practices parameter -contamination was not explicitly included in the command line as its recommended default value matches the tool default (0).
	--reference GRCh38_primary_assembly_plus_ebv_alt_decoy_hla.fasta	Reference FASTA file.	
	--intervals interval.bed	Interval to run on.	
	--emit-ref-confidence GVCF	Create GVCF output	
GATK GenotypeGVCFs (v4.0.12.0)	--variant input.bwa.sorted_interval.g.vcf --output input.bwa.sorted_interval.vcf	Sample inputs and outputs.	
	--reference GRCh38_primary_assembly_plus_ebv_alt_decoy_hla.fasta	Reference FASTA file.	
	--dbsnp Homo_sapiens_assembly38.dbsnp138.vcf	dbSNP file (only used for populating output ID and DB INFO fields)	
	--intervals interval.bed	Interval to run on	
	--only-output-calls-starting-in-intervals	Only calls starting in the current interval (--intervals) will be output.	This option prevents duplication of calls falling across consecutive intervals.
	-G StandardAnnotation	Annotations added to the output variant calls.	
	--use-new-qual-calculator	Use the new AF model.	This option is the default tool behaviour as of 4.1.0.0 (explicitly included in the Broad

			Institute Best Practices for earlier GATK versions).
GATK MergeVCFs (v4.0.12.0)	--INPUT 1.vcf … --INPUT 2.vcf --OUTPUT input.bwa.sorted.vcf	Sample inputs and outputs.	Utility tool
Picard CollectVariantCallingMetrics (v2.18.26)	INPUT=input.bwa.sorted.vcf OUTPUT=input.bwa.sorted.summary_metrics.txt	Sample inputs and outputs.	
	DBSNP=Homo_sapiens_assembly38.dbsnp138.vcf	dbSNP input file.	
chrM-rename-chrMT (v0)	sed "s/^chrM/chrMT/g" input.bwa.sorted.vcf > input.bwa.sorted.MTed.vcf	Renaming chrM variant calls to match entries in SnpEff database file.	Utility tool
SnpEff (v4.3k)	-csvStats	Create CSV summary file.	
	GRCh38.86	Use database for GRCh38.86.	
Tabix BGZIP (v0.2.6)	-c -f input.vcf > input.vcf.gz	BGZIP compress the input file while writing to stdout.	Utility tool
Tabix INDEX (v0.2.6)	-f -p vcf input.vcf.gz	Create a TBI index for the input VCF.GZ file.	Utility tool
QC per sample aggregate (v0)		Scripts for collecting QC metrics.	Utility tool
Merge QC aggregate (v0)			Utility tool
Picard CollectWgsMetrics (v2.18.26)	INPUT=input.bwa.sorted.bam OUTPUT=input.bwa.sorted.summary_metrics.txt	Sample inputs and outputs.	
	REFERENCE_SEQUENCE=GRCh38_primary_assembly_plus_ebv_alt_decoy_hla.fasta	Reference sequence.	
	MINIMUM_MAPPING_QUALITY=0 MINIMUM_BASE_QUALITY=0	Overriding default filters per agreement on 30 May 2019.	

Table S2. Genotype concordance workflow tools and parameters.

Tool (version)	**Nested tool (if it exists)**	**Parameter**	**Description**
VCF concordance - Sample Preprocessing (SBG Internal Tools v0, VT Toolkit v0.5572+, Tabix v1.2.1)	rename_sample_column_ukb_vcf (v0)		Utility tool. This tool does renaming of sample column in the WGS VCF files to match UKB array data identifiers, with matching done via UKB-supplied bridging files.
	VT preprocessing	vt sort input.vcf.gz |	It consists of preprocessing with the vt toolkit

	(SBG Internal Tools v0, VT Toolkit v0.5572+, Tabix v1.2.1)	vt decompose -s - | vt normalize -r ref.fa - | vt uniq | bgzip -c > out.vcf.gz; tabix out.vcf.gz	(sorting, normalization, decomposition and dropping potential duplicate variant entries introduced by normalization and decomposition), followed by compressing the output with bgzip and indexing with tabix.
VCF concordance - bcftools stats to TGZ (SBG Internal Tools v0, Bcftools v1.9)	Bcftools Stats - Simplified Edition (v1.9)	--collapse snps	Any SNP records are compatible, regardless of whether the ALT alleles match or not. For duplicate positions, only the first SNP record will be considered and appear on output.
		--apply-filters .,PASS	Include only sites which have no filters set.
		-s sample	Include additional sample level statistics in the output.
		--verbose	Produce verbose per-site and per-sample output
		--fasta-ref GRCh38_primary_assembly_plus_ebv_alt_decoy_hla.fasta	Reference sequence.
		Input_1.vcf.gz input_2.vcf.gz	Input files to be compared.
		> output.stats	The location and the filename of the output stats file.
	compress-stats (v0)		Utility tool. Multiple bcftools stats outputs per chromosome were gathered together in one TGZ archive.
parse_bcftools_stats (v0)			Utility tool. It outputs NRD values for SNPs and indels for all samples, chromosome-level NRD values for SNPs and indels and genotype concordance tables for all samples, for SNPs and indels.
tgz-repack-per-sample (v0)			Utility tool. It gathers individual sample results into one archive.

Table S3. Cram Check Phase 1 workflow

Tool (version)	**Parameter**	**Description**
Samtools View (1.9)	samtools view -u --reference Homo_sapiens.GRCh38_15_plus_hs38d1.fa input.cram | samtools view -c	This command line is used to check the format of the input CRAM. Success codes of piped processes are evaluated via a bash wrapper around the command given.

Table S4. Analysis parameters for the WGS phase 2 workflow.

Tool (version)	Parameter	Description	Remarks
Metadata fetch (v0)	/opt/samtools-1.9/samtools view -H input.cram | grep '^@RG' > with_metadata.header.sam	This tool sets metadata on input files using the @RG fields in the SAM/BAM/CRAM file.	Utility tool
md5sum (v0)	md5sum input.file > input.file.md5	This tool invokes md5sum Unix utility to calculate MD5 sums of the input files.	
SBG Pair FASTQs by Metadata (v1)			Utility tool.
biobambam2 bamseqchksum (v2.0.144)	reference=Homo_sapiens.GRCh38_15_plus_hs38d1.fa/GRCh38_primary_assembly_plus_ebv_alt_decoy_hla.fasta	Reference FASTA corresponding to the CRAM input. For CRAM files produced by the workflow GRCh38_primary_assembly_plus_ebv_alt_decoy_hla.fasta reference file is used.	
	inputformat=cram	Input file format.	
	< input_cram.cram > output.bamseqchksum.txt	Input and output filenames.	
biobambam2 bamtofastq (v2.0.144)	outputperreadgroupsuffixF=input.pe_1.fastq outputperreadgroupsuffixF2=input.pe_2.fastq outputperreadgroupsuffixO=input.o_1.fastq outputperreadgroupsuffixO2=input.o_2.fastq outputperreadgroupsuffixS=input.s.fastq	Separate outputs (read group level) for paired-end FASTQ files (1 and 2) and potential orphaned (1 and 2) and single-end reads.	
	inputformat=cram	Input format is CRAM.	
	reference=Homo_sapiens.GRCh38_15_plus_hs38d1.fa	Reference FASTA corresponding to the CRAM input.	
	filename=input.cram	The location and the filename of the input cram file.	
	outputdir=.	Output directory.	
	outputperreadgroup=1	Output FASTQ files per read	Read-group

		group.	level FASTQ files are created.
FastQC (v0.11.7)	--adapters adapters.txt	Input file with adapters.	
	--threads $NT	Number of threads	Matches the number of input read-group level FASTQ files.
BWA-MEM: BWA MEM (v0.7.17) Sambamba View (v0.6.7)	-K 100000000	Reads chunk size to process. Used to achieve deterministic results in multi-threaded mode.	This option is used to ensure that tool results do not depend on the number of processing threads.
	-Y	Use soft-clipping for supplementary reads.	
	-R '@RG\tID:value\tCN:SC\tLB:value\tPL:ILLUMINA\tPU:value\tSM:input\tDT:value\tDS:UK Biobank Study: UK Biobank Whole Genome sequencing study\tPG:value'	Read group header line.	This parameter is not in Broad Institute Best Practices, but is an optional permitted parameter in CCDG and not expected to affect results.
	-t $NT	Number of threads.	Not expected to affect results (see -K).
	GRCh38_primary_assembly_plus_ebv_alt_decoy_hla.fasta	Reference FASTA file.	Corresponding indices and the .alt file supplied alongside it.
	$FASTQ1 $FASTQ2	Sample inputs.	
	| sambamba view -f bam -t $NT -S /dev/stdin -o RG_input.bam	Conversion of native bwa mem SAM output (-S) to BAM (-f bam) using multiple threads (-t).	Adding mate tags with samblaster (-a --addMateTags) was omitted in agreement with WSI.
Picard MarkDuplicates (v2.18.26)	INPUT=RG1_input.bam INPUT=RG2_input.bam OUTPUT=input.bwa.bam	Sample inputs and outputs.	

	METRICS_FILE=input.metrics		
	OPTICAL_DUPLICATE_PIXEL_DISTANCE=2500	Maximum offset between two duplicate clusters in order to consider them optical duplicates.	
	ASSUME_SORT_ORDER=queryname	Consider inputs query name sorted.	Output of BWA MEM is query-grouped, which is considered similar enough by Broad Institute's Best Practices.
	VALIDATION_STRINGENCY=SILENT	Validation level of SAM-type inputs. SILENT improves performance when processing BAMs if not all data needs to be decoded.	
Sambamba Sort (v0.5.9)	--nthreads=$NT --memory-limit=$MEM	Number of threads. Memory for sorting.	This step performs coordinate-sorting of BAM inputs.
	--out=input.bwa.sorted.bam input.bwa.bam	Sample inputs and outputs.	
VerifyBamID (v1.1.3)	--bam input.bwa.sorted.bam --bai input.bwa.sorted.bam.bai --out input.bwa.sorted	Sample inputs and outputs.	
	--vcf Standard-GRCh38_15_plus_hs38d1.vcf.gz	Input VCF file with allele frequencies.	
	--minAF 0.05	Minimum allele frequency of the markers to include.	
	--self	Only compare the ID-matching individuals between the VCF and BAM file.	
	--ignoreRG	Ignore read group level comparison and compare samples only.	
	--precise	Calculate the likelihood in log-scale for high-depth data (recommended when --maxDepth is greater than 20).	

	--maxDepth 500	Maximum read depth.	
	--minQ 20	Minimum base quality to include.	
	--noPhoneHome	Disable sending of usage statistics.	
SAMtools View (v1.9)	--output-fmt CRAM -h	Convert to CRAM and output header.	
	--reference GRCh38_primary_assembly_plus_ebv_alt_decoy_hla.fasta	Reference FASTA file.	
	--threads $NT	Number of threads.	
	-o input.gathered.sorted.cram input.gathered.sorted.bam	Sample inputs and outputs.	
SAMtools Index (v1.9)	-c	Create a CSI index.	
Samtools Stats (1.9-48-g2d4907c with htslib version 1.9-gbcf9bff)	--threads $NT	Number of threads.	
	--cov-threshold 14	Only bases with coverage above this value will be included in the target percentage computation.	Value is set to 14 in agreement with WSI.
	--ref-seq GRCh38_primary_assembly_plus_ebv_alt_decoy_hla.fasta	Reference sequence.	
	--remove-overlaps	Remove overlaps of paired-end reads from coverage and base count computations.	
	--filtering-flag 0xF04	Filtering flag.	SUPPLEMENTARY, SECONDARY, QCFAIL, DUP, UNMAPPED
	--target-regions Homo_sapiens.GRCh38_15_plus_hs38d1_AUTOS.fa.interval_list	Perform the calculation of statistics using these regions only.	
	input.bwa.sorted.bam	Input BAM file.	
	> input.bwa.sorted_F0xF04.autosomes.stats	The location and the filename of the output stats file.	
Samtools idxstats (1.9)	input.bwa.sorted.bam >	This tool outputs read	

	input.idxstats.txt	statistics based on the index file.	
Determine-xy (v0)	python determineXY.py input.idxstats.txt	A python script implementing two coverage-based sample sex estimation methods.	
Picard CollectSequencingArtifactMetrics (v2.18.26)	OUTPUT=input.bwa.sorted	Output file name prefix.	
	R=GRCh38_primary_assembly_plus_ebv_alt_decoy_hla.fasta	Reference sequence.	
	INPUT=input.bwa.sorted.bam	Input BAM file.	
QC report (v0)		This tool gathers the outputs of other QC tools in the workflow and creates a JSON report and ZIP archive.	Utility tool

Table S5. Analysis parameters of the WGS phase 3 workflow.

Tool (version)	**Parameter**	**Description**	**Remarks**
SBG Prepare Intervals (v1.0)	python sbg_prepare_intervals.py --bed Homo_sapiens_primary_assembly_38_81_intervals.bed --mode 4	Based on the input file containing intervals, this tool creates per interval files and a file with ALT contigs in a single file.	Utility tool
SBG FASTA Indices (v1.0)	samtools faidx GRCh38_primary_assembly_plus_ebv_alt_decoy_hla.fasta && java -Xmx2048M -jar /opt/picard.jar CreateSequenceDictionary R=GRCh38_primary_assembly_plus_ebv_alt_decoy_hla.fasta O=GRCh38_primary_assembly_plus_ebv_alt_decoy_hla.dict		Utility tool.
md5sum (v0)		This tool invokes md5sum Unix utility to calculate MD5 sums of the input files.	
GATK Base	--input	Sample inputs and	

Recalibrator (4.0.12.0)	input.bwa.sorted.bam --output input.bwa.sorted.recal_data.grp	outputs.	
	--known-sites Homo_sapiens_assembly38.dbsnp138.vcf --known-sites Mills_and_1000G_gold_standard.indels.hg38.vcf --known-sites Homo_sapiens_assembly38.known_indels.vcf	Known sites resources for recalibration.	
	--reference GRCh38_primary_assembly_plus_ebv_alt_decoy_hla.fasta	Reference FASTA file.	
	--intervals interval.bed	Interval to run on (parallelization option).	Broad Best Practices uses contig-grouping, CCDG allows parallelization in this way and permits autosomes only.
	--use-original-qualities	Use the base quality scores from the OQ tag	Broad Best Practices parameter; optional, permitted in CCDG.
GATK GatherBQSRReports (4.0.12.0)	--input 1.grp ... --input n.grp --output input.bwa.sorted.recal_data.all.grp	This tool merges scattered BaseRecalibrator tool outputs.	Utility tool
GATK ApplyBQSR (4.0.12.0)	--input input.bwa.sorted.bam --output input.bwa.sorted_interval.recalibrated.bam	Sample inputs and outputs.	
	--intervals interval.bed	Interval to run on.	
	--reference GRCh38_primary_assembly_plus_ebv_alt_decoy_hla.fasta	Reference FASTA file.	
	--bqsr-recal-file input.bwa.sorted.recal_data.all.grp	Input base recalibration table.	
	--static-quantized-quals 10	BQSR binning scheme.	This option compresses recalibrated base quality

	--static-quantized-quals 20 --static-quantized-quals 30		scores to significantly reduce output sizes with minimum impact on variant calling (CCDG, Broad Institute Best Practices).
	--add-output-sam-program-record	Adds a PG tag to created BAM files.	Broad Best Practices parameter, optional in CCDG
	--use-original-qualities	Use the base quality scores from the OQ tag.	Broad Best Practices parameter, optional in CCDG
GATK GenotypeGVCFs (4.0.12.0)	--variant input.bwa.sorted_interval.g.vcf --output input.bwa.sorted_interval.vcf	Sample inputs and outputs.	
	--reference GRCh38_primary_assembly_plus_ebv_alt_decoy_hla.fasta	Reference FASTA file.	
	--dbsnp Homo_sapiens_assembly38.dbsnp138.vcf	dbSNP file (only used for populating output ID and DB INFO fields)	
	--intervals interval.bed	Interval to run on	
	--only-output-calls-starting-in-intervals	Only calls starting in the current interval (--intervals) will be output.	This option prevents duplication of calls falling across consecutive intervals.
	-G StandardAnnotation	Annotations added to the output variant calls.	
	--use-new-qual-calculator	Use the new AF model.	This option is the default tool behaviour as of 4.1.0.0 (explicitly included in the Broad Institute Best Practices for earlier GATK versions).
GATK MergeVcfs (4.0.12.0)	--INPUT 1.vcf … --INPUT 2.vcf --OUTPUT input.bwa.sorted.vcf	Sample inputs and outputs.	Utility tool
GATK IndexFeatureFile (4.0.12.0)	Indexing VCF and VCF.GZ input reference files - creating .IDX and .TBI indices, respectively.	Utility tool	

GATK HaplotypeCaller (4.0.12.0)	--input input.bwa.sorted_chr15.recalibrated.bam --output input.bwa.sorted_interval.g.vcf	Sample inputs and outputs.	The optional Broad Best Practices parameter -contamination was not explicitly included in the command line as its recommended default value matches the tool default (0).
	--reference GRCh38_primary_assembly_plus_ebv_alt_decoy_hla.fasta	Reference FASTA file.	
	--intervals interval.bed	Interval to run on.	
	--emit-ref-confidence GVCF	Create GVCF output	
Tabix BGZIP (v0-2-6)	-c -f input.vcf > input.vcf.gz	BGZIP compress the input file while writing to stdout.	Utility tool
Tabix Index (v0-2-6)	-f -p vcf input.vcf.gz	Create a TBI index for the input VCF.GZ file.	Utility tool
read_haps	--fa GRCh38_primary_assembly_plus_ebv_alt_decoy_hla.fasta	Reference sequence.	Default values were used for all tool parameters.
	input.bwa.sorted.bam	Input BAM file.	
	high_quality_markers_deCODE_2015.txt	Default reference file distributed alongside the tool. If a different set of markers should be used, these should be provided to Seven Bridges.	
	input.bwa.sorted.vcf.gz	Input VCF file.	
VT preprocessing (SBG Internal Tools v0, VT Toolkit v0.5572+, Tabix v1.2.1)	vt sort input.vcf.gz | vt decompose -s - | vt normalize -r ref.fa - | vt uniq | bgzip -c > out.vcf.gz; tabix out.vcf.gz	It consists of preprocessing with the vt toolkit (sorting, normalization, decomposition and dropping potential duplicate variant entries introduced by normalization and decomposition), followed by compressing the output with bgzip and indexing with tabix.	This tool is only used to preprocess variant calls before NRD checks with BCFtools stats.

Bcftools Stats (v1.9)	--collapse snps	Any SNP records are compatible, regardless of whether the ALT alleles match or not. For duplicate positions, only the first SNP record will be considered and appear on output.	
	--apply-filters .,PASS	Include only sites which have no filters set.	
	-s sample	Include additional sample level statistics in the output.	
	--verbose	Produce verbose per-site and per-sample output	
	--fasta-ref GRCh38_primary_assembly_plus_ebv_alt_decoy_hla.fasta	Reference sequence.	
	Input_1.vcf.gz input_2.vcf.gz	Input files to be compared.	
	> output.stats	The location and the filename of the output stats file.	
WGS QC Collate (v0)			Utility script