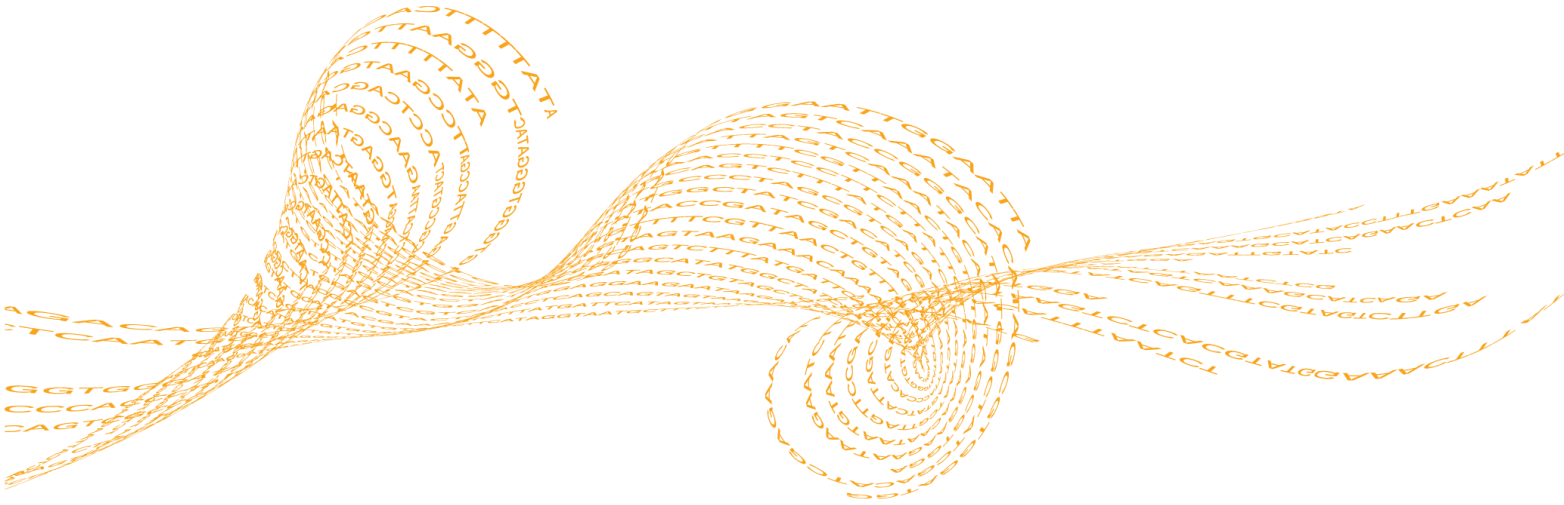
BWA Enrichment v2.1

BaseSpace App Guide

For Research Use Only. Not for use in diagnostic procedures.



ILLUMINA PROPRIETARY
Document # 15050958 v01
January 2016

illumina®

Introduction

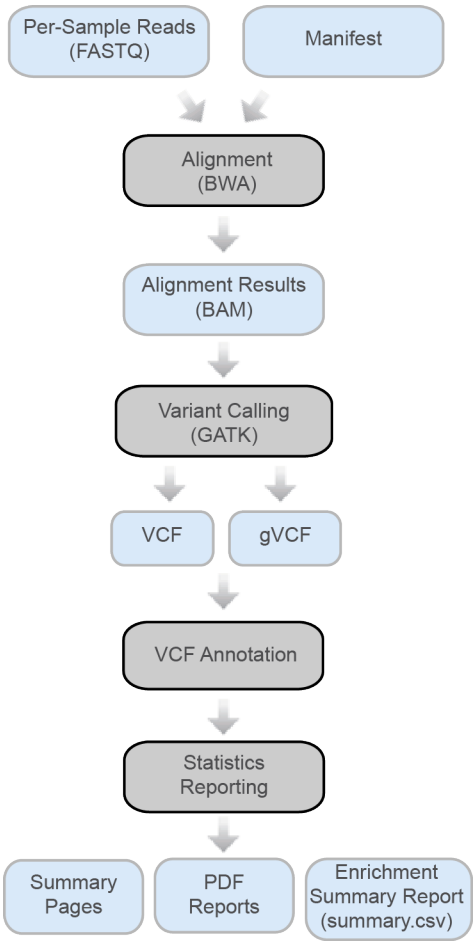
The BaseSpace® BWA Enrichment v2.1 App analyzes DNA that are enriched for particular target sequences using Nextera® Rapid Capture. Burrows-Wheeler Aligner (BWA) aligns the samples to the reference genome and GATK performs variant calls. The app analyzes the variants for the target regions. Statistics reporting accumulates coverage and enrichment-specific statistics for each target as well as overall metrics.

Compatible Libraries

See the BaseSpace support page for a list of library types that are compatible with the BWA Enrichment v2.1 App.

Workflow Diagram

Figure 1 BWA Enrichment v2.1 App Workflow



Versions

The following components are used in the BWA Enrichment v2.1 App.

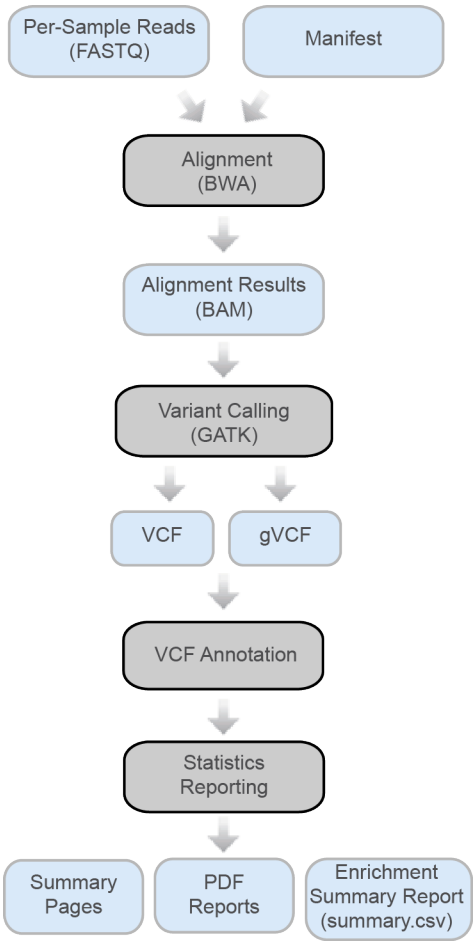
Software	Version
BWA	0.7.7-isis-1.0.0
GATK	1.6-23-gf0210b3
Picard	1.79
IAS (Annotation Service)	v3.0
SAMtools	0.1.19-isis-1.0.1
Tabix	0.2.5 (r1005)

Reference Genomes

- Human, UCSC hg19
  The human reference genome is PAR-Masked, which means that the Y chromosome sequence has the Pseudo Autosomal Regions (PAR) masked (set to N) to avoid mismapping of reads in the duplicate regions of sex chromosomes.

Workflow Diagram

Figure 2 BWA Enrichment v2.1 App Workflow



Set Analysis Parameters

1. Navigate to BaseSpace, and then click the **Apps** tab.
2. Click **BWA Enrichment**.
3. From the drop-down list, select **version 2.1.0**, and then click **Launch** to open the app.
4. In the **Analysis Name** field, enter the analysis name.
   By default, the analysis name includes the app name, followed by the date and time that the analysis session starts.
5. From the **Save Results To** field, select the project that stores the app results.
6. From the **Sample(s) [96 maximum]** field, browse to the sample you want to analyze, and select the checkbox.
   You can select multiple samples, with a maximum of 96 samples in one session.
7. From the **Reference Genome** field, select the reference genome you want to align.
8. From the **Targeted Regions** field, select the targeted region of your enrichment.
9. From the **Custom Targeted Manifest** field, select the custom targeted manifest file for analysis.
   This option is available if you selected **Custom Manifest** from the **Targeted Regions** field drop-down list.
10. From the **Aligner** field, select the aligner. The default is **BWA-MEM**.
    BWA-MEM is optimized for alignment of modern Illumina sequencing data. BWA-backtrack can be used for consistency with legacy data.
11. From the **Base Padding** field, select the padding. The default is **150**.
    Padding defines the amount of sequence immediately upstream and downstream of the targeted regions that is also used in enrichment analysis.
12. From the **Annotation** field, select the gene and transcript annotation reference database. The default is **RefSeq**.
13. Optional, click the **Advance** drop-down list for additional parameter fields.
    a. From the **Trim Adapters** field, select the adapters to trim. Use this setting if adapter trimming has not already been applied as a sample sheet setting.
    b. From the **Flag PCR Duplicates** field, select the checkbox if you want the app to flag PCR duplicates in the BAM files and to not used for variant calling.
    PCR duplicates are defined as two clusters from a paired-end run where both clusters have the exact same alignment positions for each read. Optical duplicates are already filtered out during RTA processing. PCR duplicates are not applicable for single-end samples.
    c. From the **Generate Picard HS Metrics and Per Target Coverage Information** field, select the checkbox to generate Picard HS metrics. See *Picard Metrics* on page 8 for more information.
    d. From the **Custom Probes Manifest** field, select the custom probes manifest file for analysis. This field is required when you select the **Custom Targeted Manifest** and **Generate Picard HS Metrics and per target coverage information** options.
14. Click **Continue**.
    The BWA Enrichment v2.1 App begins analysis of the samples.

When analysis is complete, the app updates the status of the session and sends an email to notify you.

Analysis Methods

The BWA Enrichment v2.1 App uses these methods to analyze the sequencing data.

BWA

The BWA Enrichment workflow uses the Burrows-Wheeler Aligner (BWA, which adjusts parameters based on read lengths and error rates, and then estimates the insert size distribution).

For more information, see github.com/lh3/bwa.

After BWA alignment, GATK performs variant calling.

GATK

The Genome Analysis Toolkit (GATK) is the standard variant caller after BWA alignment.

Developed by the Broad Institute, the Genome Analysis Toolkit (GATK) first calls raw variants for each sample read. Then GATK analyzes the variants against known variants, and applies a calibration procedure to compute a false discovery rate for each variant. Variants are flagged as homozygous (1/1) or heterozygous (0/1) in the VCF file sample column.

The GATK best practices were guidelines for the app; they are described here: www.broadinstitute.org/gatk/guide/topic?name=best-practices.

For more information about GATK, see www.broadinstitute.org/gatk.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 43(5): 491-8.

Picard Metrics

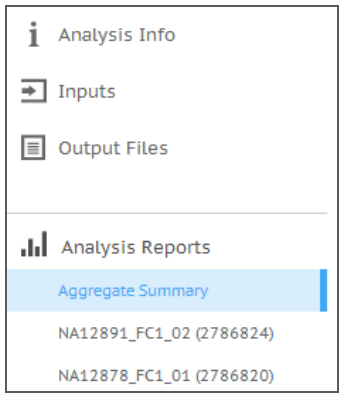
Picard is a suite of tools in Java that work with next-generation sequencing data in BAM format. BWA Enrichment uses the CalculateHsMetrics tool in Picard to compute a set of Hybrid Selection specific metrics from an aligned SAM or BAM file. If a reference sequence is provided, AT/GC dropout metrics are calculated. GC and mean coverage information for every target can also be computed.

For more information, see http://broadinstitute.github.io/picard/.

Analysis Output

To view the results, click the **Projects** tab, then the project name, and then the analysis.

Figure 3 BWA Enrichment v2.1 Output Navigation Bar



After analysis is complete, access the output through the left navigation bar.

- **Analysis Info**—Information about the analysis session, including log files.
- **Inputs**—Overview of input settings.
- **Output Files**—Output files for the sample.
- **Aggregate Summary**—Analysis metrics for the aggregate results.
- **Sample Analysis Reports**—Analysis reports for each sample.

Analysis Info

The Analysis Info page displays the analysis settings and execution details.

Row Heading	Definition
Name	Name of the analysis session.
Application	App that generated this analysis.
Date Started	Date and time the analysis session started.
Date Completed	Date and time the analysis session completed.
Duration	Duration of the analysis.
Session Type	Number of nodes used.
Status	Status of the analysis session. The status shows either Running or Complete.

Log Files

File Name	Description
CompletedJobInfo.xml	Contains information about the completed analysis session.

File Name	Description
EnrichmentStatistics.xml	Contains statistics about the completed analysis session.
Logging.zip	Contains all detailed log files for each step of the workflow.
output-{timestamp}.log	Shows the raw console output from the app.

Output Files

The Output Files page provides access to the output files for each sample analysis.

- BAM Files
- VCF Files
- gVCF Files
- Enrichment_summary.csv
- Manifest Output Files

BAM File Format

A BAM file (*.bam) is the compressed binary version of a SAM file that is used to represent aligned sequences up to 128 Mb. SAM and BAM formats are described in detail at https://samtools.github.io/hts-specs/SAMv1.pdf.

BAM files use the file naming format of SampleName_S#.bam, where # is the sample number determined by the order that samples are listed for the run.

BAM files contain a header section and an alignments section:

- **Header**—Contains information about the entire file, such as sample name, sample length, and alignment method. Alignments in the alignments section are associated with specific information in the header section.
- **Alignments**—Contains read name, read sequence, read quality, alignment information, and custom tags. The read name includes the chromosome, start coordinate, alignment quality, and the match descriptor string.

The alignments section includes the following information for each or read pair:

- **RG:** Read group, which indicates the number of reads for a specific sample.
- **BC:** Barcode tag, which indicates the demultiplexed sample ID associated with the read.
- **SM:** Single-end alignment quality.
- **AS:** Paired-end alignment quality.
- **NM:** Edit distance tag, which records the Levenshtein distance between the read and the reference.
- **XN:** Amplicon name tag, which records the amplicon tile ID associated with the read.

BAM index files (*.bam.bai) provide an index of the corresponding BAM file.

VCF File Format

VCF is a widely used file format developed by the genomics scientific community that contains information about variants found at specific positions in a reference genome.

VCF files use the file naming format SampleName_S#.vcf, where # is the sample number determined by the order that samples are listed for the run.

VCF File Header—Includes the VCF file format version and the variant caller version. The header lists the annotations used in the remainder of the file. If MARS is listed, the Illumina internal annotation algorithm annotated the VCF file. The VCF header includes the reference genome file and BAM file. The last line in the header contains the column headings for the data lines.

VCF File Data Lines—Each data line contains information about a single variant.

VCF File Headings

Heading	Description
CHROM	The chromosome of the reference genome. Chromosomes appear in the same order as the reference FASTA file.
POS	The single-base position of the variant in the reference chromosome. For SNPs, this position is the reference base with the variant; for indels or deletions, this position is the reference base immediately before the variant.
ID	The rs number for the SNP obtained from dbSNP.txt, if applicable. If there are multiple rs numbers at this location, the list is semicolon delimited. If no dbSNP entry exists at this position, a missing value marker ('.') is used.
REF	The reference genotype. For example, a deletion of a single T is represented as reference TT and alternate T. An A to T single nucleotide variant is represented as reference A and alternate T.
ALT	The alleles that differ from the reference read. For example, an insertion of a single T is represented as reference A and alternate AT. An A to T single nucleotide variant is represented as reference A and alternate T.
QUAL	A Phred-scaled quality score assigned by the variant caller. Higher scores indicate higher confidence in the variant and lower probability of errors. For a quality score of Q, the estimated probability of an error is $10^{-(Q/10)}$. For example, the set of Q30 calls has a 0.1% error rate. Many variant callers assign quality scores based on their statistical models, which are high in relation to the error rate observed.

VCF File Annotations

Heading	Description
FILTER	If all filters are passed, **PASS** is written in the filter column. • **LowDP**—Applied to sites with depth of coverage below a cutoff. • **LowGQ**—The genotyping quality (GQ) is below a cutoff. • **LowQual**—The variant quality (QUAL) is below a cutoff. • **LowVariantFreq**—The variant frequency is less than the given threshold. • **R8**—For an indel, the number of adjacent repeats (1-base or 2-base) in the reference is greater than 8. • **SB**—The strand bias is more than the given threshold. Used with the Somatic Variant Caller and GATK.

Heading	Description
INFO	Possible entries in the INFO column include: • **AC**—Allele count in genotypes for each ALT allele, in the same order as listed. • **AF**—Allele Frequency for each ALT allele, in the same order as listed. • **AN**—The total number of alleles in called genotypes. • **CD**—A flag indicating that the SNP occurs within the coding region of at least 1 RefGene entry. • **DP**—The depth (number of base calls aligned to a position and used in variant calling). • **Exon**—A comma-separated list of exon regions read from RefGene. • **FC**—Functional Consequence. • **GI**—A comma-separated list of gene IDs read from RefGene. • **QD**—Variant Confidence/Quality by Depth. • **TI**—A comma-separated list of transcript IDs read from RefGene.
FORMAT	The format column lists fields separated by colons. For example, GT:GQ. The list of fields provided depends on the variant caller used. Available fields include: • **AD**—Entry of the form X,Y, where X is the number of reference calls, and Y is the number of alternate calls. • **DP**—Approximate read depth; reads with MQ=255 or with bad mates are filtered. • **GQ**—Genotype quality. • **GQX**—Genotype quality. GQX is the minimum of the GQ value and the QUAL column. In general, these values are similar; taking the minimum makes GQX the more conservative measure of genotype quality. • **GT**—Genotype. 0 corresponds to the reference base, 1 corresponds to the first entry in the ALT column, and so on. The forward slash (/) indicates that no phasing information is available. • **NL**—Noise level; an estimate of base calling noise at this position. • **PL**—Normalized, Phred-scaled likelihoods for genotypes. • **SB**—Strand bias at this position. Larger negative values indicate less bias; values near 0 indicate more bias. Used with the Somatic Variant Caller and GATK. • **VF**—Variant frequency; the percentage of reads supporting the alternate allele.
SAMPLE	The sample column gives the values specified in the FORMAT column.

Genome VCF Files

Genome VCF (gVCF) files are VCF v4.1 files that follow a set of conventions for representing all sites within the genome in a reasonably compact format. The gVCF files include all sites within the region of interest in a single file for each sample.

The gVCF file shows no-calls at positions with low coverage, or where a low-frequency variant (< 3%) occurs often enough (> 1%) that the position cannot be called to the reference. A genotype (GT) tag of **./.** indicates a no-call.

For more information, see sites.google.com/site/gvcftools/home/about-gvcf.

Enrichment Summary Report

The BWA Enrichment v2.1 app produces an enrichment summary report and the aggregate results in a comma-separated values (CSV) format: *.summary.csv. These files are located in the results folder for each sample and the aggregate results.

NOTE
PCR duplicate reads are not removed from statistics. Results are not directly comparable to Picard HsMetrics.

Statistic	Definition
Sample ID	IDs of samples reported on in the file.
Sample Name	Names of samples reported on in the file.
Run Folder	Run folders for samples reported on in the file.
Reference Genome	Reference genome selected.
Target Manifest	The target manifest file used for analysis. This file specifies the targeted regions for the aligner and variant caller.
Total Length of Targeted Reference	The total length of the sequenced bases in the target region.
Padding Size	The length of sequence immediately upstream and downstream of the enrichment targets that is included for a padded target.
Total PF Reads	The number of reads passing filter for the sample.
Total Aligned Reads	The total number of reads passing filter present in the data set that aligned to the reference genome.
Percent Aligned Reads	
Percent Duplicate Paired Reads	Percentage of paired reads that have duplicates.
Targeted Aligned Reads	Number of reads that aligned to the target.
Padded Target Aligned Reads	Number of reads that aligned to the padded target.
Read Enrichment	100*(Target aligned reads/Total aligned reads).
Padded Read Enrichment	100*(Padded target aligned reads/Total aligned reads).
Total PF Bases	The number of bases passing filter for the sample.
Percent Q30	The percentage of bases with a quality score of 30 or higher.
Percent Q30 Aligned	Percent of bases with a quality score of 30 or higher that aligned to the reference genome.

Statistic	Definition
Total Aligned bases	The total number of bases present in the data set that aligned to the reference genome.
Percent Aligned bases	Percent aligned bases in the target region.
Targeted Aligned bases	Total aligned bases in the target region.
Padded Target Aligned bases	Total aligned bases in the padded target region.
Base Enrichment	100*(Total Aligned Bases in Targeted Regions/Total Aligned Bases).
Padded Base Enrichment	100*(Total Aligned Bases in Padded Targeted Regions/Total Aligned Bases).
Mean Region Coverage Depth	The total number of aligned bases to the targeted region divided by the targeted region size.
Uniformity of Coverage (Pct > 0.2*mean):	The percentage of targeted base positions in which the read depth is greater than 0.2 times the mean region target coverage depth.
Target Coverage at 1X	Percentage targets with coverage greater than 1X.
Target Coverage at 10X	Percentage targets with coverage greater than 10X.
Target Coverage at 20X	Percentage targets with coverage greater than 20X.
Target Coverage at 50X	Percentage targets with coverage greater than 50X.
Fragment Length Median	Median length of the sequenced fragment. The fragment length is calculated based on the locations at which a read pair aligns to the reference. The read mapping information is parsed from the BAM files.
Fragment Length Min	Minimum length of the sequenced fragment.
Fragment Length Max	Maximum length of the sequenced fragment.
Fragment Length SD	Standard deviation of the sequenced fragment length.
SNVs, Indels, Insertions, Deletions	Total number of variants present in the data set that pass the quality filters.
SNVs (All), Indels (All), Insertions (All), Deletions (All)	Total number of predicted variants in the data set.

Statistic	Definition
SNVs, Indels, Insertions, Deletions (Percent Found in dbSNP)	100*(Number of variants in dbSNP/Number of variants).
SNV Ts/Tv ratio	The number of Transition SNVs that pass the quality filters divided by the number of Transversion SNVs that pass the quality filters. Transitions are interchanges of purines (A, G) or of pyrimidines (C, T). Transversions are interchanges of purine and pyrimidine bases (for example, A to T).
SNVs, Indels, Insertions, Deletions Het/Hom ratio	Number of heterozygous variants/Number of homozygous variants.
SNVs, Insertions, Deletions in Genes	The number of variants that fall into a gene.
SNVs, Insertions, Deletions in Exons	The number of variants that fall into an exon.
SNVs, Insertions, Deletions in Coding Regions	The number of variants that fall into a coding region.
SNVs, Insertions, Deletions in Mature miRNA	The number of variants that fall into a mature microRNA.
SNVs, Insertions, Deletions in UTR Region	The number of variants that fall into an untranslated region (UTR).
SNVs, Insertions, Deletions in Splice Site Region	The number of variants that fall into a splice site region.
Stop Gained SNVs, Insertions, Deletions	The number of variants that cause an additional stop codon.
Stop Lost SNVs, Insertions, Deletions	The number of variants that cause the loss of a stop codon.
Frameshift Insertions, Deletions	The number of variants that cause a frameshift.
Non-synonymous SNVs, Insertions, Deletions	The number of variants that cause an amino acid change in a coding region.
Synonymous SNVs	The number of variants that are within a coding region, but do not cause an amino acid change.

Manifest Output Files

The BWA Enrichment v2.1 app produces BED and TXT manifest output files that specify the regions that were used in the analysis. If there were any duplicates or overlapping regions, those files contain the corrected version. The BED file can be used in VariantStudio or IGV to highlight the targeted regions.

The two output files are in the aggregate output directory for multi-sample inputs and in the single sample directory for single-sample inputs.

Sample Summary Page

The BWA Enrichment v2.1 app provides an overview of statistics per sample in the Analysis Reports sample pages. To download the statistics, click **PDF Summary Report**.

- Enrichment Summary
- Small Variants Summary
- Coverage Summary
- Fragment Summary
- Duplicate Summary

NOTE
PCR duplicate reads are not removed from statistics. Results are not directly comparable to Picard HsMetrics.

Enrichment Summary

Table 1 Enrichment Summary

Statistic	Definition
Total Length of Targeted Reference	The total length of the sequenced bases in the target region.
Padding Size	The length of sequence immediately upstream and downstream of the enrichment targets that is included for a padded target.

Table 2 Read Level Enrichment

Statistic	Definition
Total Aligned Reads	The total number of reads passing filter present in the data set that aligned to the reference genome.
Percent Aligned Reads	The percentage of reads passing filter that aligned to the reference genome.
Target Aligned Reads	Number of reads that aligned to the target.
Read Enrichment	100*(Target aligned reads/Total aligned reads).
Padded Target Aligned Reads	Number of reads that aligned to the padded target.
Padded Read Enrichment	100*(Padded target aligned reads/Total aligned reads).

Table 3 Base Level Enrichment

Statistic	Definition
Total Aligned Bases	The total number of bases present in the data set that aligned to the reference genome.
Target Aligned Bases	Total aligned bases in the target region.
Bases Enrichment	100*(Total Aligned Bases in Targeted Regions/Total Aligned Bases).
Padded Target Aligned Bases	Total aligned bases in the padded target region.
Padded Base Enrichment	100*(Total Aligned Bases in Padded Targeted Regions/Total Aligned Bases).

Small Variants Summary

Table 4 Small Variants Summary

Statistic	Definition
Total Passing	The total number of variants present in the data set that passed the variant quality filters.
Percent Found in dbSNP	100*(Number of variants in dbSNP/Number of variants).
Het/Hom Ratio	Number of heterozygous variants/Number of homozygous variants.
Ts/Tv Ratio	Transition rate of SNVs that pass the quality filters divided by transversion rate of SNVs that pass the quality filters. Transitions are interchanges of purines (A, G) or of pyrimidines (C, T). Transversions are interchanges between purine and pyrimidine bases (for example, A to T).

Table 5 Variants by Sequence Context

Statistic	Definition
Number in Genes	The number of variants that fall into a gene.
Number in Exons	The number of variants that fall into an exon.
Number in Coding Regions	The number of variants that fall into a coding region.
Number in UTR Region	The number of variants that fall into an untranslated region (UTR).
Number in Mature microRNA	The number of variants that fall into a mature microRNA.
Number in Splice Site Regions	The number of variants that fall into a splice site region.

Table 6 Variants by Consequence

Statistic	Definition
Frameshifts	The number of variants that cause a frameshift.
Non-synonymous	The number of variants that cause an amino acid change in a coding region.
Synonymous	The number of variants that are within a coding region, but do not cause an amino acid change.
Stop Gained	The number of variants that cause an additional stop codon.
Stop Lost	The number of variants that cause the loss of a stop codon.

Coverage Summary

Table 7 Coverage Summary

Statistic	Definition
Mean Coverage	The total number of aligned bases to the targeted region divided by the targeted region size.
Uniformity of Coverage (Pct > 0.2*mean):	The percentage of targeted base positions in which the read depth is greater than 0.2 times the mean region target coverage depth.
Target Coverage at 1X	Percentage targets with coverage greater than 1X.
Target Coverage at 10X	Percentage targets with coverage greater than 10X.
Target Coverage at 20X	Percentage targets with coverage greater than 20X.
Target Coverage at 50X	Percentage targets with coverage greater than 50X.

Fragment Length Summary

Table 8 Fragment Length Summary

Statistic	Definition
Fragment Length Median	Median length of the sequenced fragment. The fragment length is calculated based on the locations at which a read pair aligns to the reference. The read mapping information is parsed from the BAM files.
Minimum	Minimum length of the sequenced fragment.
Maximum	Maximum length of the sequenced fragment.
Standard Deviation	Standard deviation of the sequenced fragment length.

Duplicate Summary

Table 9 Duplicate Summary

Statistic	Definition
Percent Duplicate Paired Reads	Percentage of paired reads that have duplicates.

Enrichment Sequencing Report

The BWA Enrichment v2.1 App provides an aggregate summary in PDF for all samples combined on the Summary page. To download the statistics, click **PDF Summary Report**.

- Enrichment Summary
- Manifest Information
- SNV Summary
- Indel Summary
- Coverage Summary
- Fragment Summary
- Duplicate Summary

NOTE
PCR duplicate reads are not removed from statistics. Results are not directly comparable to Picard HsMetrics.

Enrichment Summary

Table 10 Enrichment Summary

Statistic	Definition
Total Length of Targeted Reference	The total length of the sequenced bases in the target region.
Padding Size	The length of sequence immediately upstream and downstream of the enrichment targets that is included for a padded target.

Table 11 Read Level Enrichment

Statistic	Definition
Total Aligned Reads	The total number of reads passing filter present in the data set that aligned to the reference genome.
Percent Aligned Reads	The percentage of reads passing filter that aligned to the reference genome.
Target Aligned Reads	Number of reads that aligned to the target.
Read Enrichment	100*(Target aligned reads/Total aligned reads).

Statistic	Definition
Padded Target Aligned Reads	Number of reads that aligned to the padded target.
Padded Read Enrichment	100*(Padded target aligned reads/Total aligned reads).

Table 12 Base Level Enrichment

Statistic	Definition
Total Aligned Bases	The total number of bases present in the data set that aligned to the reference genome.
Target Aligned Bases	Total aligned bases in the target region.
Bases Enrichment	100*(Total Aligned Bases in Targeted Regions/Total Aligned Bases).
Padded Target Aligned Bases	Total aligned bases in the padded target region.
Padded Base Enrichment	100*(Total Aligned Bases in Padded Targeted Regions/Total Aligned Bases).

Manifest Information

Table 13 Manifest Information

Statistic	Definition
Total length	The total length of the sequenced bases in the target region.
Padding Size	The length of sequence immediately upstream and downstream of the enrichment targets that is included for a padded target.

SNV Summary

Table 14 SNV Summary

Statistic	Definition
SNVs	Total number of Single Nucleotide Variants (SNVs) present in the data set passing the quality filters.
SNVs (Percent Found in dbSNP)	100*(Number of SNVs in dbSNP/Number of SNVs). The SNVs that were found in the dbSNP are annotated accordingly.
SNV Ts/Tv Ratio	The number of Transition SNVs that pass the quality filters divided by the number of Transversion SNVs that pass the quality filters. Transitions are interchanges of purines (A, G) or of pyrimidines (C, T). Transversions are interchanges of purine and pyrimidine bases (for example, A to T).
SNV Het/Hom Ratio	Number of heterozygous SNVs/Number of homozygous SNVs.

Indel Summary

Table 15 Fragment Length Summary

Statistic	Definition
Indels	Total number of indels present in the data set passing the quality filters.
Indels (Percent Found in dbSNP)	100*(Number of Indels in dbSNP/Number of Indels).
Indel Het/Hom Ratio	Number of heterozygous indels/Number of homozygous indels.

Coverage Summary

Table 16 Coverage Summary

Statistic	Definition
Mean Coverage	The total number of aligned bases to the targeted region divided by the targeted region size.
Uniformity of Coverage (Pct > 0.2*mean):	The percentage of targeted base positions in which the read depth is greater than 0.2 times the mean region target coverage depth.
Target Coverage at 1X	Percentage targets with coverage greater than 1X.
Target Coverage at 10X	Percentage targets with coverage greater than 10X.
Target Coverage at 20X	Percentage targets with coverage greater than 20X.
Target Coverage at 50X	Percentage targets with coverage greater than 50X.

Fragment Length Summary

Table 17 Fragment Length Summary

Statistic	Definition
Fragment Length Median	Median length of the sequenced fragment. The fragment length is calculated based on the locations at which a read pair aligns to the reference. The read mapping information is parsed from the BAM files.
Minimum	Minimum length of the sequenced fragment.
Maximum	Maximum length of the sequenced fragment.
Standard Deviation	Standard deviation of the sequenced fragment length.

Duplicate Summary

Table 18 Duplicate Summary

Statistic	Definition
Percent Duplicate Paired Reads	Percentage of paired reads that have duplicates.

Revision History

Document	Date	Description of Change
Document # 15050958 v01	January 2016	Reorganized topics, updated writing style.

Notes

Technical Assistance

For technical assistance, contact Illumina Technical Support.

Table 19 Illumina General Contact Information

Website	www.illumina.com
Email	techsupport@illumina.com

Table 20 Illumina Customer Support Telephone Numbers

Region	Contact Number	Region	Contact Number
North America	1.800.809.4566	Japan	0800.111.5011
Australia	1.800.775.688	Netherlands	0800.0223859
Austria	0800.296575	New Zealand	0800.451.650
Belgium	0800.81102	Norway	800.16836
China	400.635.9898	Singapore	1.800.579.2745
Denmark	80882346	Spain	900.812168
Finland	0800.918363	Sweden	020790181
France	0800.911850	Switzerland	0800.563118
Germany	0800.180.8994	Taiwan	00806651752
Hong Kong	800960230	United Kingdom	0800.917.0041
Ireland	1.800.812949	Other countries	+44.1799.534000
Italy	800.874909		

Safety data sheets (SDSs)—Available on the Illumina website at support.illumina.com/sds.html.

Product documentation—Available for download in PDF from the Illumina website. Go to support.illumina.com, select a product, then select **Documentation & Literature**.

Illumina
5200 Illumina Way
San Diego, California 92122 U.S.A.
+1.800.809.ILMN (4566)
+1.858.202.4566 (outside North America)
techsupport@illumina.com
www.illumina.com