Technote

EmbryoMap SNP Workflow QC Guidance

Overview

This technical note provides guidance ranges for key metrics of the EmbryoMap SNP Workflow. It also outlines the various factors affecting the key metrics and includes steps for troubleshooting.

Introduction

The EmbryoMap SNP workflow simultaneously amplifies DNA from a group of cells (3-10) and prepares sequencing-ready libraries for Illumina platforms. This document will help to assess if the prepared libraries and the sequencing run are within the validated sample metric ranges (Guidance ranges are shown in Table 1) when laboratories are validating their sample collection, transportation, amplification/ library preparation and sequencing processes. Refer to the **EmbryoMap SNP Sample Prep Guide** and the **EmbryoMap eMap Software Guide** for more information on how to perform the test and analyse the results.

Sample input recommendations

All biopsy samples for EmbryoMap processing must be suspended in molecular grade 1× PBS (Phosphate Buffered Saline) and, if used, with a maximum concentration of 0.5% PVP (Polyvinylpyrrolidone) in a total volume of 2.5 µl. Biopsy stability under storage and shipment conditions used (including any alternative sample collection volume, wash or collection buffers) must be fully validated by the user before use and may impact the quality and integrity of the biopsy samples, leading to suboptimal library quality that may compromise your test results.

EmbryoMap amplification

Incorporation of extraneous material containing DNA prior to amplification can lead to misleading results. This is a concern with any DNA amplification technology. Pre-PCR environmental and sampling control measures (i.e., single use of molecular grade reagents and consumables) are required to reduce the risk of sample contamination. Failure to follow Pre-PCR best practices during the reaction setup may lead to contamination of the PCR products, affecting subsequent analysis and/or compromise your test results.

Human DNA contamination

Contamination with human DNA can obscure the CNV status of the tested sample or lead to shifts in sex chromosome calling. It can also interfere with the ability to detect the true ploidy status. The sequencing reads of human DNA contamination will also map to the human genome and the system cannot distinguish it from the DNA of the samples with CNV analysis alone. The EmbryoMap SNP workflow now enables detection of external contamination of samples from an unrelated source at a threshold of ≥30%. External contamination will be reflected in the Allelic Consistency Score. If the level of contamination is high enough so that the ploidy status can no longer be called confidently, the SNP QC will fail.

Identification of contaminants

When contamination is suspected, sequencing can be used to confirm the presence or absence of extraneous DNA. For the investigation and identification of the contaminants, a 2x75 cycle or 2x100 cycle sequencing run can be performed on the libraries in question with the MiSeq Reagent Kit v3. Set up a sequencing run using the Library QC workflow from the Illumina Experiment Manager, as described in the MiSeq System User Guide (Illumina Part # 15027617). Metagenomic apps can be used on Illumina BaseSpace Sequence Hub for the analysis of contaminants. BaseSpace Sequence Hub offers a wide variety of next generation sequencing (NGS) data analysis apps that are developed or optimised by Illumina and a growing ecosystem of third-party app providers.

Non-human DNA contamination

Microorganism contamination can occur during the embryo biopsy process, from non-sterile reagents or equipment, or even airborne DNA. These extraneous materials can produce PCR products that will generate sequenceable libraries. The resulting sequencing reads will not map against the human genome and will be filtered out by the system. When the number of reads generated by the contaminant(s) is elevated, it may in turn decrease the reads of the sample below your validated sample metric ranges (Guidance ranges are described in Table 1). This may compromise your test results due to the reduction in usable reads.

EmbryoMap SNP Amplification Products

EmbryoMap SNP amplification of genomic DNA generates PCR products between approximately 0.2 to 2.5 kb. After Size Selection library fragments are optimised for Illumina sequencing, with two distinct peaks at fragment lengths of 250 bp and 550bp.

This can be seen in Figure 1. Fluorometric quantification of **Size Selected** library duplex DNA shall result in positive amplification yield for an input mass of 60 pg or 5-cell isolates that is clearly separable from the yield of 'no template control' (NTC) samples. NTC or suspected amplification failures shall produce fluorescence that is equivalent to the background of the quantification kit.

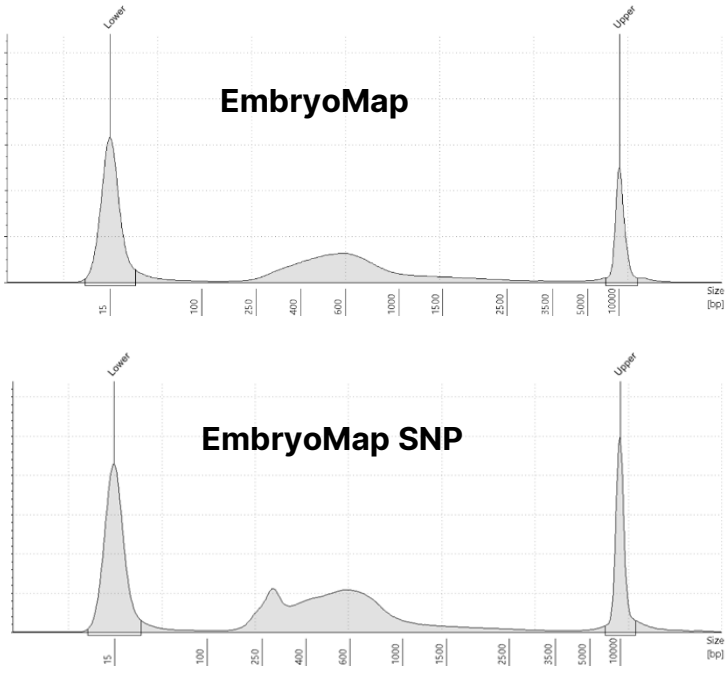

Figure 1: Fragment analysis of EmbryoMap and EmbryoMap SNP library amplification products Post- Size Selection.

NTC or suspected amplification failures may still be sequenced in an EmbryoMap run. During preparation of the 'Pooling Guide', select the **'Default'** pooling volume checkbox to include a nominal 5 µl of these samples into the final library pool. Similarly, if contamination is suspected, the NTC sample should be sequenced and reads associated with the sample's unique index should be interrogated to identify the source of contamination.

Library Handling and Sequencing

The eMap software (v2.0.0 or later) can help to assess the quality of data on two different levels, at the sequencing run (flow cell) level (Figure 2) or at a sample level for both CNVs and SNPs (Figure 3).

Sequencing run-level metrics

The percentage of successful clusters is shown by Clusters Passing Filter (%). A low percentage of clusters PF might indicate the amount of DNA loaded onto the instrument requires optimization for subsequent sequencing runs.

The number of successful reads is shown by Total reads passing filter. Low reads passing filter might indicate low library quality or the presence of non-human DNA contamination.

The workflow and sequencing read lengths will detail the type of run that has been performed. An EmbryoMap SNP run should contain the workflow "EmbryoMap SNP".

Sample-level metrics

The Sample QC tab provided by the eMap software displays metrics describing the quality of the data for each sample as it is processed from the initial base calling to alignment, filtering, and copy-number calling.

Reads that pass filters in the copy-number calling algorithm are indicated by 'Number of reads after filtering' in the eMap software. The eMap software then automatically checks the key CNV metrics (colour shaded) for each sample against the user defined acceptance criteria specified in the CNV QC settings in the eMap account.

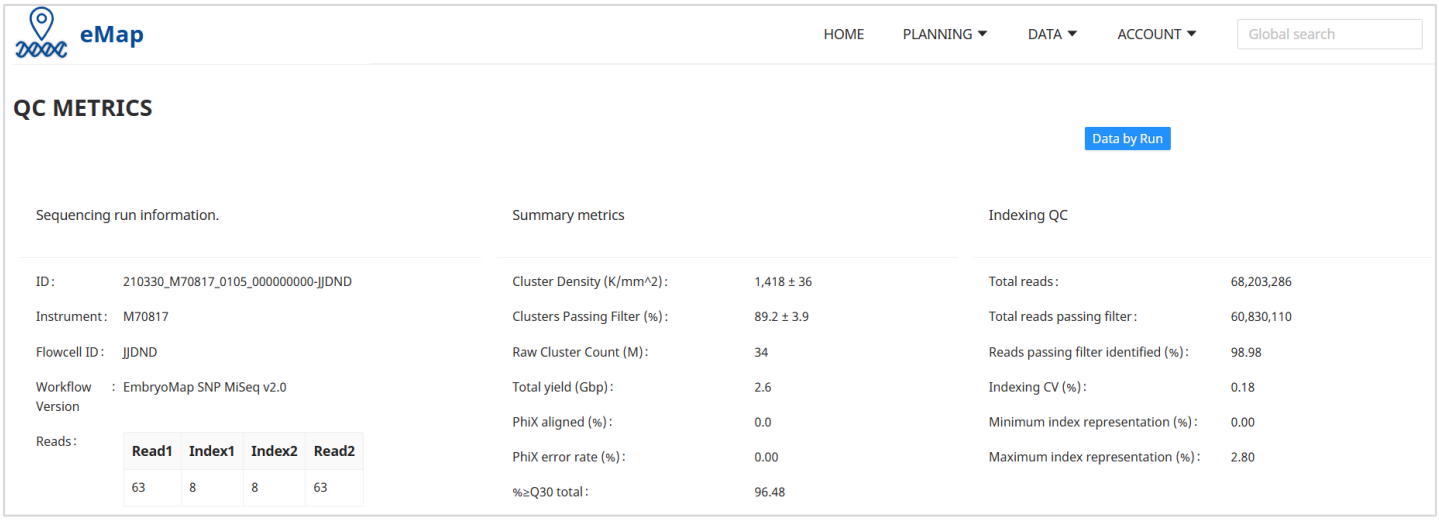

Figure 2: Example of eMap Software sequencing measurements for an EmbryoMap SNP Run.

A proportion of the reads that are filtered from the total reads will align to specific SNP targets in the genome, these are indicated by 'Valid SNP Reads'. The eMap software interrogates the Valid SNP Reads to assess the number of Het SNPs with the minimum read depth against a pre-defined Vitrolife threshold. The minimum SNP depth may be modified by the user. The SNP algorithm then automatically checks the key SNP metrics (colour shaded) for each sample against the pre-defined Vitrolife acceptance criteria.

Idiogram | Sample QC | Experimental details

SAMPLE ID	SNP-Sample01
CNV QC STATUS	PASS
AVERAGE ALIGNMENT SCORE	56.7
NUMBER OF TOTAL READS	1,134,200
NUMBER OF MAPPED READS	1,099,320
NUMBER OF READS AFTER FILTERING	790,654
OVERALL NOISE	0.17
SD ROBUST	0.19
SDI	0.71
SNP QC STATUS	PASS
NUMBER OF VALID SNP READS	275,128
NUMBER OF HETEROZYGOUS SNP CALLS	332
PLOIDY CONFIDENCE	0.99
ALLELIC CONSISTENCY	0.99

Figure 3: eMap sample-level metrics. Coloured shading indicates where metrics have been compared to the thresholds specified.

CNV QC Status

The CNV QC status is set to PASS or FAIL by default according to the sample-level QC criteria specified in the eMap account Settings. The QC Status can be manually changed by the user during the sample sign off procedure.

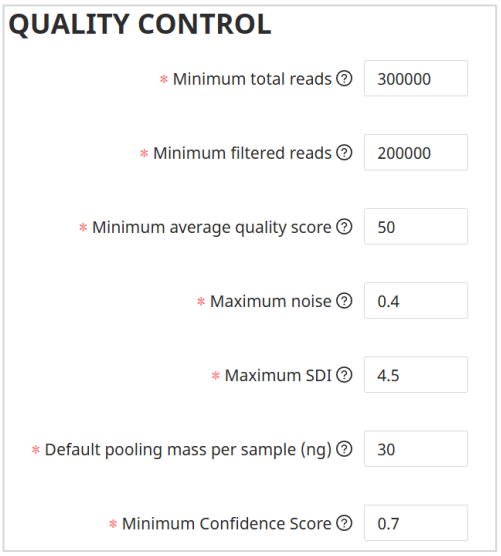

Figure 4: eMap Software CNV quality control criteria may be customised by the user in Account Settings. Quality criteria are used for automatic CNV QC flagging and status.

SNP QC Status

The SNP QC status is set to PASS or FAIL automatically according to the sample-level QC criteria specified by Vitrolife (Valid SNP reads, Het SNP Calls, Allelic Consistency, Ploidy Confidence). These thresholds may not be modified by the user.

However, the user may increase the minimum coverage (Min SNP depth) required for SNPs to be included in the analysis which may influence the metrics above. This can be changed via the eMap Account Settings to a value in the range 10-100. The QC Status **cannot** be manually changed by the user during the sample sign off procedure.

Average Alignment Score

The alignment software calculates this score and considers the base quality score (Q-score), read length, and the number of alignments for every read. The Q-score is a measure of confidence in the base that was called at a given position. The Q-score is calculated as mean (mean [Phred33-based base quality] per read) for all reads per sample.

Reads (Total, Mapped, Filtered)

The run-level metrics described above can be used to check whether the libraries have been pooled in relatively equal proportions for sequencing. The typical total number of reads per sample range from 1.5% to 4.0% of reads passing filter for 48-plex sequencing runs. The guidance range of the total reads and reads after filtering is described in Table 1. "Filtered Reads" reflects the reads available for the CNV analysis and excludes QC-failed reads, multi-mappers and duplicated reads (which include SNP reads).
Low values of Mapped (aligned) reads and Reads after filtering might indicate the presence of non-human reads in the DNA source, or an elevated proportion of PCR artefacts caused by suboptimal amplification and library preparations. It is expected that out of ~50 million paired-end reads per sequencing run, several hundred reads may randomly align against random sequences.

Overall Noise (DLR-Derivative Log Ratio)

The overall sample noise measures the spread of the difference in copy number values between all bins of a chromosome. High DLR values might indicate low quality sample DNA or problems during the amplification or library preparation steps.

SDI Noise

The Summed Deviation to nearest Integer metric describes the step-wise noise across all chromosomes in a sample. CNV charts with high SDI values are more challenging to interpret for chromosome-level copy-number.

Number of Heterozygous SNP Calls

The number of heterozygous (het) SNP calls details the number of genome wide SNPs which have a minimum coverage (or depth) of 10, these are the SNPs which are used downstream for ploidy and allelic consistency assessment. The minimum number of het SNP calls is a fixed threshold pre-defined by Vitrolife and cannot be adjusted by the user.

Ploidy Confidence

After the assessment of the number of Het SNPs available for the SNP analysis, the algorithm will process the SNP data and will output the most likely ploidy status based on a predictive model. The Ploidy confidence is a score that is outputted based on the confidence of the algorithm in the given ploidy call. A score close to 1 indicates very high confidence, while a score below 0.7 would be considered a QC Fail as the quality of the SNP data is not good enough to reliably predict the ploidy status. A ploidy confidence close to but above 0.7 indicates a low confidence call.

Allelic Consistency

The specific alleles at SNP loci are analysed for unusual patterns. An increase in the number of het SNPs or the presence of a third allele are indicators that a sample is low-quality or there is a foreign DNA source contaminating the sample. This will be reflected in the Allelic Consistency which is measured on a scale of 0 to 1, with 1 being a high-quality sample with no foreign DNA detected. If the allelic consistency falls below 0.7, this is considered a SNP QC Fail, as the SNP data is showing signs of noise/contamination and the sample ploidy state cannot be predicted.

Sequencing Issues

You can perform a more detailed analysis of your data using Illumina's Sequence Analysis Viewer (SAV) to troubleshoot the most common sequencing problems.
If your run does not meet your expected run metrics, you should monitor the following sequencing metrics:

- For hardware and software performance of the instrument, a PhiX validation run needs to be performed in accordance with the MiSeq System guide (part# 15039740) and MiSeq System User Guide (Part # 15027617).
- See the Illumina bulletins '*What is the PhiX Control v3 Library and what is its function in Illumina Next Generation Sequencing?*' and '*How to set up the sample sheet for a PhiX validation run on the MiSeq system using Illumina Experiment Manager*' for additional information.
- The EmbryoMap recipe is a custom recipe designed for paired-end, dual-index read sequencing of libraries prepared with the EmbryoMap Sample Prep kit only. This recipe is not suitable to assess the MiSeq instrument hardware and software performance.
- Note that the EmbryoMap Q30 metrics vary based on the library pool quality, cluster density, and the percentage of clusters passing filter.
- To assess the sample level performance, refer to your validated sample metric ranges (See Table 1 for guidance ranges).

Cluster Density

When assessing the quality of your EmbryoMap run, first check the cluster density against your validated sample metric ranges (Guidance ranges are described in Table 1). If optimisation is required, please refer to the Cluster Optimization guide (Illumina Part #1000000071511).
When performing the Template Line Wash, it is important that the correct concentration and volume (1 mL) of NaOCl is used. If the concentration or volume used is too high, it can impact subsequent runs by impeding cluster generation or even complete failure to generate clusters.

Flow Check Error

If the MiSeq fails to measure the flow rate while starting a run, inspect the gasket of the flow cell for any damage that may have occurred during cleaning. Please see the Illumina bulletin *'What should I do if my MiSeq fails to measure flow rate when starting a run?'* for additional information.
Clean the flow cell as described in the MiSeq System User Guide. Not adhering to the flow cell best cleaning practices might raise or distort the gasket, causing flow check errors.

Table 1: EmbryoMap SNP MiSeq flow cell and sample guidance QC metrics for CNVs and SNPs

Metrics	Parameters	Guidance Range	Sources of Values
Amplification Test	Concentration of dsDNA ng/µl (5-cells)	3-15	Final library quantification after Size Selection step
Flow cell level metrics	Density (K/mm^2)	1,100 – 1,600	MiSeq instrument screen Display, eMap Run QC page, Illumina SAV software
	Total number of reads	50,000,000	
	Total number of reads (PF)	38,000,000	
	% of clusters passing filter	≥ 75 %	
CNV metrics	Total number of reads per sample	≥ 300,000	CNV analysis in eMap Software (User Defined)
	Number of reads after filtering	≥ 200,000	
	Average quality score	≥ 50	
	Overall noise (DLR)	≤ 0.4	
	SDI noise	≤ 4.5	
Chromosome-level metric	Region confidence (value in the Regions tab)	≥ 0.70	
SNP Metrics	Number of valid SNP reads per sample	≥ 150,000	SNP Analysis in eMap Software (Vitrolife Defined)
	Number of heterozygous SNP calls	≥ 100	
	Ploidy confidence	≥ 0.7	
	Allelic consistency	≥ 0.7	

Note: The above metrics are guidance ranges. Individual laboratories must validate their own sample metric ranges in accordance with their quality management system to ensure that they meet their performance and specification requirements.