# Illumina Services
# Data Report

ASWX-20200626a

| Customer | | Fasteris | |
|---|---|---|---|
| Account | Jérémie Vidal-Dupiol | Project | ASWX-20200626a |
| | | Date / ref. | DR_2020-08-03_ASWX-1-6 |
| | | Author | Manuele Castelnuovo |

**Summary**

| | |
|---|---|
| Sequenced libraries | **ASWX-1-6** |
| Number of libraries | **48** |
| Sample type | **Unknown** |
| Run ID | **200729_A00902_A** |
| Number of cycle | **2x100+8** |
| Instrument | **NovaSeq** |
| Yield(GB) | **272.7** |
| Quality control | **Within specifications** |

Author's comments:

No comment

## 1 Introduction

DNA or RNA samples are processed using illumina technology to prepare libraries, which are then sequenced on Next-Generation DNA sequencing (NGS) instruments.

The present report describes the pipeline used and provides quality and yield information about the data produced.

Upon request, our bioinformatics team is happy to help you to extract the biologically useful information from your data.

**NB: Although the illumina NGS instruments deliver high-quality data, you must obtain independent validation of the data.**

Illumina Propel Certified™ 2017 MiSeq® NextSeq® HiSeq® 2500 HiSeq® 4000

## 2 Library preparation

| Sample name | Sample type | Library ID | Library type | Index * |
|---|---|---|---|---|
| E50-3-400 | Unknown | ASWX-1-1 | Ready-to-Run | ATGTCA |
| I49-3-375 | Unknown | ASWX-2-1 | Ready-to-Run | TAGCTT |
| D48-5-64 | Unknown | ASWX-3-1 | Ready-to-Run | CGATGT |
| J41-4-416 | Unknown | ASWX-4-1 | Ready-to-Run | ATGTCA |
| E45-8-727 | Unknown | ASWX-5-1 | Ready-to-Run | TAGCTT |
| J56-3-208 | Unknown | ASWX-6-1 | Ready-to-Run | CGATGT |
| E54-5-508 | Unknown | ASWX-1-2 | Ready-to-Run | CCGTCC |
| E55-2-45 | Unknown | ASWX-2-2 | Ready-to-Run | GGCTAC |
| I42-1-253 | Unknown | ASWX-3-2 | Ready-to-Run | TGACCA |
| I55-4-459 | Unknown | ASWX-4-2 | Ready-to-Run | CCGTCC |
| J23-7-684 | Unknown | ASWX-5-2 | Ready-to-Run | GGCTAC |
| J19-3-113 | Unknown | ASWX-6-2 | Ready-to-Run | TGACCA |
| C49-2-57 | Unknown | ASWX-1-3 | Ready-to-Run | GTCCGC |
| D27-1-250 | Unknown | ASWX-2-3 | Ready-to-Run | GTGGCC |
| D53-2-10 | Unknown | ASWX-3-3 | Ready-to-Run | GCCAAT |
| C2-4-183 | Unknown | ASWX-4-3 | Ready-to-Run | GTCCGC |
| D25-1-242 | Unknown | ASWX-5-3 | Ready-to-Run | GTGGCC |
| B25-2-73 | Unknown | ASWX-6-3 | Ready-to-Run | GCCAAT |
| C52-7-118 | Unknown | ASWX-1-4 | Ready-to-Run | GTGAAA |
| E5-1-255 | Unknown | ASWX-2-4 | Ready-to-Run | GTTTCG |
| B51-8-16 | Unknown | ASWX-3-4 | Ready-to-Run | ACAGTG |
| B26-4-13 | Unknown | ASWX-4-4 | Ready-to-Run | GTGAAA |
| E19-8-84 | Unknown | ASWX-5-4 | Ready-to-Run | GTTTCG |
| I39-4-429 | Unknown | ASWX-6-4 | Ready-to-Run | ACAGTG |
| B60-3-59 | Unknown | ASWX-1-5 | Ready-to-Run | ATCACG |
| D30-4-172 | Unknown | ASWX-2-5 | Ready-to-Run | CGTACG |
| E51-8-728 | Unknown | ASWX-3-5 | Ready-to-Run | CAGATC |
| B4-1-44 | Unknown | ASWX-4-5 | Ready-to-Run | ATCACG |
| B39-5-161 | Unknown | ASWX-5-5 | Ready-to-Run | CGTACG |
| I35-7-678 | Unknown | ASWX-6-5 | Ready-to-Run | CAGATC |
| B52-5-144 | Unknown | ASWX-1-6 | Ready-to-Run | TTAGGC |
| J55-1-265 | Unknown | ASWX-2-6 | Ready-to-Run | GAGTGG |
| E28-8-132 | Unknown | ASWX-3-6 | Ready-to-Run | CTTGTA |
| E23-1-221 | Unknown | ASWX-4-6 | Ready-to-Run | TTAGGC |
| C18-1-55 | Unknown | ASWX-5-6 | Ready-to-Run | GAGTGG |
| E57-4-124 | Unknown | ASWX-6-6 | Ready-to-Run | CTTGTA |
| E49-3-61 | Unknown | ASWX-1-7 | Ready-to-Run | ACTTGA |
| J16-2-339 | Unknown | ASWX-2-7 | Ready-to-Run | ACTGAT |
| C19-8-699 | Unknown | ASWX-3-7 | Ready-to-Run | AGTCAA |

| Sample name | Sample type | Library ID | Library type | Index * |
|---|---|---|---|---|
| B38-1-180 | Unknown | ASWX-4-7 | Ready-to-Run | ACTTGA |
| D21-8-70 | Unknown | ASWX-5-7 | Ready-to-Run | ACTGAT |
| D7-2-153 | Unknown | ASWX-6-7 | Ready-to-Run | AGTCAA |
| J30-7-650 | Unknown | ASWX-1-8 | Ready-to-Run | GATCAG |
| C60-8-702 | Unknown | ASWX-2-8 | Ready-to-Run | ATTCCT |
| D46-5-30 | Unknown | ASWX-3-8 | Ready-to-Run | AGTTCC |
| J37-2-312 | Unknown | ASWX-4-8 | Ready-to-Run | GATCAG |
| D41-8-71 | Unknown | ASWX-5-8 | Ready-to-Run | ATTCCT |
| D39-1-9 | Unknown | ASWX-6-8 | Ready-to-Run | AGTTCC |

*Table 1: Description of the samples*

** index sequences used for the basecalling*

## 3 Sequencing

### 3.1 Instrument

| | |
|---|---|
| Serial number | **A00902** |
| Manufacturer | Illumina |
| Version | NovaSeq 6000 |
| Slot used | A |
| Basecalling pipeline | - NovaSeq Control Software 1.6.0<br>- RTA v3.4.4<br>- bcl2fastq2.20 v2.20.0.422 |

### 3.2 Run

| | |
|---|---|
| Run ID | **200729_A00902_A** |
| Mode | Xp |
| Number of cycles | 2x100+8 |
| Number of lanes | 2 |
| Flow cell ID | HMNGCDRXX |
| Flow cell version | S1 |
| Kit version | NV2864786-RGSBS |

### 3.3 Fasteris specifications[1]

| | |
|---|---|
| Service | **Full flow-cell** |
| PF clusters (minimal yield) | **1'300 Mreads** |
| Q30 | 80 % |

[1] Fasteris specifications for the run described in points above

## 4 Basecalling summary

### 4.1 Parameters

| Lane | Number of mismatch in the index selection[1] | Phix spiked[2] | Other multiplexed libraries[3] |
|---|---|---|---|
| 1 | 0 | Yes | No |
| 2 | 0 | Yes | No |

*Table 2: Basecalling parameters*
*[1] According to Fasteris specifications, 1 mismatch is only authorized when all the indexes differ by at least 3 bases.*
*[2] Indicates if a PhiX reference is spiked (i.e. a low concentration of a PhiX library is added before sequencing and selectively retrieved through its related index) in your lane to estimate the error rate for your sequences.*
*[3] Indicates whether libraries from other projects are present in the lane*

The base calling pipeline proceeds to the demultiplexing prior to the generation of fast-q sequence files, *i.e.* by separating the libraries according to their indexes. When different libraries are multiplexed in the same lane, a very low proportion of cross-talk may happen (reads sorted to wrong index)

### 4.2 Results

The sequences are sorted according to their index code.

| Lane | Expected Read nb[1] (PF[2]) | Library ID | Yield (Mb) | %PF | Cluster (PF) | Q30[3] | Mean qual. (PF) |
|---|---|---|---|---|---|---|---|
| 1 | 27’000’000 | ASWX-1-1 | 3'933 | 100 | 19'668'789 | 92.01 | 35.65 |
| 1 | 27’000’000 | ASWX-1-2 | 2'995 | 100 | 14'978'061 | 92.61 | 35.76 |
| 1 | 27’000’000 | ASWX-1-3 | 3'466 | 100 | 17'331'783 | 92.37 | 35.72 |
| 1 | 27’000’000 | ASWX-1-4 | 4'195 | 100 | 20'976'014 | 92.26 | 35.69 |
| 1 | 27’000’000 | ASWX-1-5 | 4'081 | 100 | 20'409'245 | 92.68 | 35.77 |
| 1 | 27’000’000 | ASWX-1-6 | 3'138 | 100 | 15'694'776 | 91.98 | 35.65 |
| 1 | 27’000’000 | ASWX-1-7 | 4'076 | 100 | 20'381'930 | 92.57 | 35.75 |
| 1 | 27’000’000 | ASWX-1-8 | 3'518 | 100 | 17'591'749 | 92.83 | 35.80 |
| 1 | 27’000’000 | ASWX-2-1 | 8'107 | 100 | 40'538'415 | 92.73 | 35.79 |
| 1 | 27’000’000 | ASWX-2-2 | 6'611 | 100 | 33'057'387 | 91.82 | 35.62 |

| Lane | Expected Read nb[1] (PF[2]) | Library ID | Yield (Mb) | %PF | Cluster (PF) | Q30[3] | Mean qual. (PF) |
|---|---|---|---|---|---|---|---|
| 1 | 27'000'000 | ASWX-2-3 | 7'681 | 100 | 38'407'547 | 92.81 | 35.80 |
| 1 | 27'000'000 | ASWX-2-4 | 5'643 | 100 | 28'216'362 | 92.19 | 35.69 |
| 1 | 27'000'000 | ASWX-2-5 | 5'498 | 100 | 27'490'908 | 92.71 | 35.79 |
| 1 | 27'000'000 | ASWX-2-6 | 8'410 | 100 | 42'053'593 | 92.34 | 35.72 |
| 1 | 27'000'000 | ASWX-2-7 | 7'667 | 100 | 38'339'069 | 92.93 | 35.83 |
| 1 | 27'000'000 | ASWX-2-8 | 10'131 | 100 | 50'657'803 | 92.70 | 35.78 |
| 1 | 27'000'000 | ASWX-3-1 | 7'008 | 100 | 35'043'421 | 92.93 | 35.83 |
| 1 | 27'000'000 | ASWX-3-2 | 6'566 | 100 | 32'831'653 | 92.61 | 35.77 |
| 1 | 27'000'000 | ASWX-3-3 | 6'542 | 100 | 32'711'089 | 92.89 | 35.82 |
| 1 | 27'000'000 | ASWX-3-4 | 5'770 | 100 | 28'854'258 | 92.98 | 35.84 |
| 1 | 27'000'000 | ASWX-3-5 | 298 | 100 | 1'494'006 | 93.03 | 35.83 |
| 1 | 27'000'000 | ASWX-3-6 | 7'545 | 100 | 37'726'900 | 92.87 | 35.82 |
| 1 | 27'000'000 | ASWX-3-7 | 6'554 | 100 | 32'773'525 | 92.47 | 35.74 |
| 1 | 27'000'000 | ASWX-3-8 | 7'067 | 100 | 35'335'954 | 92.43 | 35.73 |
| 2 | 27'000'000 | ASWX-4-1 | 4'385 | 100 | 21'926'082 | 92.27 | 35.70 |
| 2 | 27'000'000 | ASWX-4-2 | 5'174 | 100 | 25'874'171 | 92.83 | 35.81 |
| 2 | 27'000'000 | ASWX-4-3 | 4'094 | 100 | 20'473'431 | 92.89 | 35.82 |
| 2 | 27'000'000 | ASWX-4-4 | 4'397 | 100 | 21'989'172 | 92.89 | 35.82 |
| 2 | 27'000'000 | ASWX-4-5 | 6'392 | 100 | 31'964'011 | 93.07 | 35.85 |
| 2 | 27'000'000 | ASWX-4-6 | 4'937 | 100 | 24'686'909 | 92.67 | 35.78 |
| 2 | 27'000'000 | ASWX-4-7 | 5'432 | 100 | 27'162'864 | 93.26 | 35.88 |
| 2 | 27'000'000 | ASWX-4-8 | 4'506 | 100 | 22'532'070 | 93.28 | 35.89 |
| 2 | 27'000'000 | ASWX-5-1 | 2'464 | 100 | 12'320'952 | 93.44 | 35.92 |
| 2 | 27'000'000 | ASWX-5-2 | 2'804 | 100 | 14'023'852 | 92.79 | 35.80 |
| 2 | 27'000'000 | ASWX-5-3 | 2'384 | 100 | 11'922'179 | 93.23 | 35.88 |
| 2 | 27'000'000 | ASWX-5-4 | 3'118 | 100 | 15'592'448 | 93.07 | 35.85 |
| 2 | 27'000'000 | ASWX-5-5 | 2'448 | 100 | 12'243'831 | 93.27 | 35.89 |
| 2 | 27'000'000 | ASWX-5-6 | 3'144 | 100 | 15'722'253 | 93.48 | 35.93 |
| 2 | 27'000'000 | ASWX-5-7 | 3'201 | 100 | 16'005'815 | 93.50 | 35.93 |
| 2 | 27'000'000 | ASWX-5-8 | 3'353 | 100 | 16'765'072 | 93.46 | 35.92 |
| 2 | 27'000'000 | ASWX-6-1 | 11'548 | 100 | 57'741'689 | 92.56 | 35.75 |
| 2 | 27'000'000 | ASWX-6-2 | 8'347 | 100 | 41'736'263 | 92.53 | 35.75 |
| 2 | 27'000'000 | ASWX-6-3 | 10'743 | 100 | 53'719'031 | 92.69 | 35.78 |
| 2 | 27'000'000 | ASWX-6-4 | 7'846 | 100 | 39'234'844 | 92.79 | 35.80 |
| 2 | 27'000'000 | ASWX-6-5 | 9'020 | 100 | 45'102'531 | 92.94 | 35.82 |
| 2 | 27'000'000 | ASWX-6-6 | 9'905 | 100 | 49'525'469 | 92.71 | 35.78 |
| 2 | 27'000'000 | ASWX-6-7 | 8'676 | 100 | 43'381'514 | 92.36 | 35.72 |
| 2 | 27'000'000 | ASWX-6-8 | 7'922 | 100 | 39'614'101 | 92.54 | 35.75 |

*[1] Total number for all lanes combined*

*[2] PF stands for 'passed filter' i.e. clusters that fulfill the default Illumina quality criteria*
*[3] % of bases (PF) with a quality score greater or equal to 30*

### 4.3 Quality control

**Spiked-PhiX:**
Fasteris developed an “in-lane” control spike in each lane of the flow-cell. These spiked control reads are mapped on the PhiX reference genome. (Details can be found in the document “Fasteris_data-quality-and-specifications.pdf” ).

**Q30:**
The illumina pipelines estimates the reads quality according to the percentage of bases having a base quality value greater or equal to 30 (Q30), ie less than 1 error in 1000 bases. (c.f. : “Fasteris_data-quality-and-specifications.pdf” ).

| Lane | Q30 Average % | Within specifications |
|---|---|---|
| 1 | 92.80 | Yes |
| 2 | 93.33 | Yes |

*N.B.: We do not perform quality filters on sequence files that are within our quality specifications, besides the default 'failed chastity' filter done by the pipeline itself. We have observed that additional filters can alter the representation/variability of the sequences (e.g. specific removal of sequences with secondary structures).*

The base-calling pipeline can sometimes call bases as blanks ('N'). Blanks correspond to unattributed bases. If the blank rate is lower than 0.05%, no further information is provided.

## 5 Nomenclature

| | |
|---|---|
| Library: | The DNA or RNA samples are processed into short fragments of 20-1000 bp, depending on the protocols (e.g. genomic shotgun, transcriptome, ChIP-SEQ, small RNA, etc..) and cloned *in vitro* between the 3' and 5' adapters. Most libraries are amplified by PCR to generate enough DNA for precise measurement of its concentration, a key factor for maximizing the yield (too low concentration will mean not enough DNA clusters and too high will results in too many overlapping DNA clusters that will be eliminated at quality filter step during base-calling) |
| Insert: | Sample fragment that has been incorporated between two adapters during library preparation after fragmentation and size selection. |
| Adapter: | 3' and 5' sequences added during library preparation (used for PCR amplification, DNA cluster generation on the flow cell and sequencing) |
| Index: | A 6-to-8 bases DNA sequence tag found in adapter sequence to uniquely identify each library. The index is read separately from the inserts and its sequence is used at demultiplexing step. |
| DNA Cluster: | A DNA Colony generated on the flow-cell from a single DNA molecule of the library. It is perceived as a single sequencing unit by the base-calling pipeline, even if 2 DNA clusters are overlapping. In such case, the double sequence produced is eliminated at filtering step. |
| Read: | A sequence obtained after base calling. Its length is determined by the number of sequencing cycles. All the reads have the same length.<br>Runs done using only forward sequencing primer, will generate one **single-read** per DNA cluster. Runs done using the forward and the reverse sequencing primers will produce two **paired-reads** (one pair) per DNA cluster. |
| PF Clusters: | Pass filter clusters. The illumina pipeline uses the chastity filter *(c.f. : "FAS_MKT_X_NGS-QCandSpecifications.pdf" )* to remove sequences produced from clusters with low signal to noise ratio (e.g. overlapping DNA clusters). |
| PhiX Spike: | A Fasteris-developped quality control to measure the real error rate in the lane. About 0.5% of a PhiX library is added in each lane. |
| Demultiplexing: | Sorting the reads according to the indexes of each library |

## 6 File format

| **Index selected sequences**<br>e.g. 200729_A00902_A_L001_ASWX-1-1_R1.fastq.gz<br>200729_A00902_A_L001_ASWX-1-1_R2.fastq.gz |
|---|
| *File extension: fastq.gz* |
| *Format: Illumina fast-q format. A text file listing the sequences and their quality. Each group of 4 lines describe one sequence:*<br>*Line 1: is the sequence name (a unique identifier of the sequence)*<br>*Line 2: the base sequence itself*<br>*Line 3- orientation of the sequence (always + to indicate forward strand)*<br>*Line 4- quality value for each base, encoded as a Phred score* |

The sequence files are available to download on our secured server.

*N.B. The sequence files are compressed as .gz archives. The archives can be uncompressed on linux OS using a gzip -d command. We cannot guarantee, due to their large size, that they can be uncompressed on Windows or MacOS systems.*