

Project Information

Item	Value
Project Name	R_jondoe_XWGS1
Manifest	R_jondoe_XWGS1_M001
PI Name	Dr John Doe
PI Email	John.Doe@unimelb.edu.au
PI Address	
Received Date	2014-08-10
Report Date	2015-01-19
Number of Samples	10
Lanes Used	15

Overall Summary

The table below summarises the overall performance of each sample. It provides the raw sequencing Yield (Megabases), the percentage of bases that had quality \geq Q30, and compares the requested sequencing coverage to the mean sequencing coverage that we achieved. It is important to note that to determine coverage statistics, we aligned the sequencing reads to the reference genome using a fast genome aligner (iSAAC; see methods). Your values may differ if you use a different short-read aligner.

	SampleID	Yield (Mb)	% Bases \geq Q30	Requested Coverage	Mean Coverage (Illumina)
1	6282	130548	86.05	30	36.30
2	T1157	128200	87.59	30	36.25
3	T21432	254645	87.64	30	70.03
4	T21437	252844	87.06	30	69.85
5	T21454	258057	82.41	30	59.63
6	T21573	265852	87.68	30	70.34
7	T22770	268429	87.35	30	70.10
8	T23485	132053	86.56	30	35.81
9	T23547	134321	87.23	30	37.00
10	T23548	132965	87.09	30	36.37

Samples Information

Samples are sequenced on multiple lanes in order to achieve the requested sequencing coverage and quality. In some cases, even a 30x genome requires more than one sequencing lane. We refer to each lane that a sample is sequenced on as an Indivisible Unit of Sequencing (IUS).

Each IUS is named using the following pattern: FlowCellID_Lane_SampleID_Species_Index_Manifest

Item	Description
------	-------------

FlowCellID	The ID of the flow cell used for sequencing
Lane	The lane on the flow cell used for the sample
SampleID	The sample ID consist of the date of submission and Sample name
Species	The species the sample was taken from (default is human)
Index	The index sequence for multiplexing (not currently used for HiSeq X)
Manifest	Manifest consist of the project name followed by the manifest id

The following table shows the samples for each IUS.

	IUS	FlowCellID	Lane	SampleID	Species	Index	Manifest
1	H078KALXX_1_6282_Human__R_jondoe_XWGS1_M001	H078KALXX	1	6282	Human		R_jondoe_XWGS1_M001
2	H078KALXX_5_T1157_Human__R_jondoe_XWGS1_M001	H078KALXX	5	T1157	Human		R_jondoe_XWGS1_M001
3	H078KALXX_6_T21432_Human__R_jondoe_XWGS1_M001	H078KALXX	6	T21432	Human		R_jondoe_XWGS1_M001
4	H09ATALXX_4_T21432_Human__R_jondoe_XWGS1_M001	H09ATALXX	4	T21432	Human		R_jondoe_XWGS1_M001
5	H078KALXX_7_T21437_Human__R_jondoe_XWGS1_M001	H078KALXX	7	T21437	Human		R_jondoe_XWGS1_M001
6	H09ATALXX_5_T21437_Human__R_jondoe_XWGS1_M001	H09ATALXX	5	T21437	Human		R_jondoe_XWGS1_M001
7	H0ANHALXX_3_T21454_Human__R_jondoe_XWGS1_M001	H0ANHALXX	3	T21454	Human		R_jondoe_XWGS1_M001
8	H0ATNALXX_2_T21454_Human__R_jondoe_XWGS1_M001	H0ATNALXX	2	T21454	Human		R_jondoe_XWGS1_M001
9	H067GALXX_1_T21573_Human__R_jondoe_XWGS1_M001	H067GALXX	1	T21573	Human		R_jondoe_XWGS1_M001
10	H09ATALXX_6_T21573_Human__R_jondoe_XWGS1_M001	H09ATALXX	6	T21573	Human		R_jondoe_XWGS1_M001
11	H067GALXX_2_T22770_Human__R_jondoe_XWGS1_M001	H067GALXX	2	T22770	Human		R_jondoe_XWGS1_M001
12	H0AP7ALXX_8_T22770_Human__R_140911_jondoe_FGS00088_M001	H0AP7ALXX	8	T22770	Human		R_140911_jondoe_FGS00088
13	H067GALXX_3_T23485_Human__R_jondoe_XWGS1_M001	H067GALXX	3	T23485	Human		R_jondoe_XWGS1_M001
14	H067GALXX_4_T23547_Human__R_jondoe_XWGS1_M001	H067GALXX	4	T23547	Human		R_jondoe_XWGS1_M001
15	H067GALXX_5_T23548_Human__R_jondoe_XWGS1_M001	H067GALXX	5	T23548	Human		R_jondoe_XWGS1_M001

Library and Read Specifications

Below are statistics that describe the short-read sequencing library prepared from the sequenced samples. Sequencing read type and read length(s) are also indicated

Read Type: Paired

Read 1 Length: 150

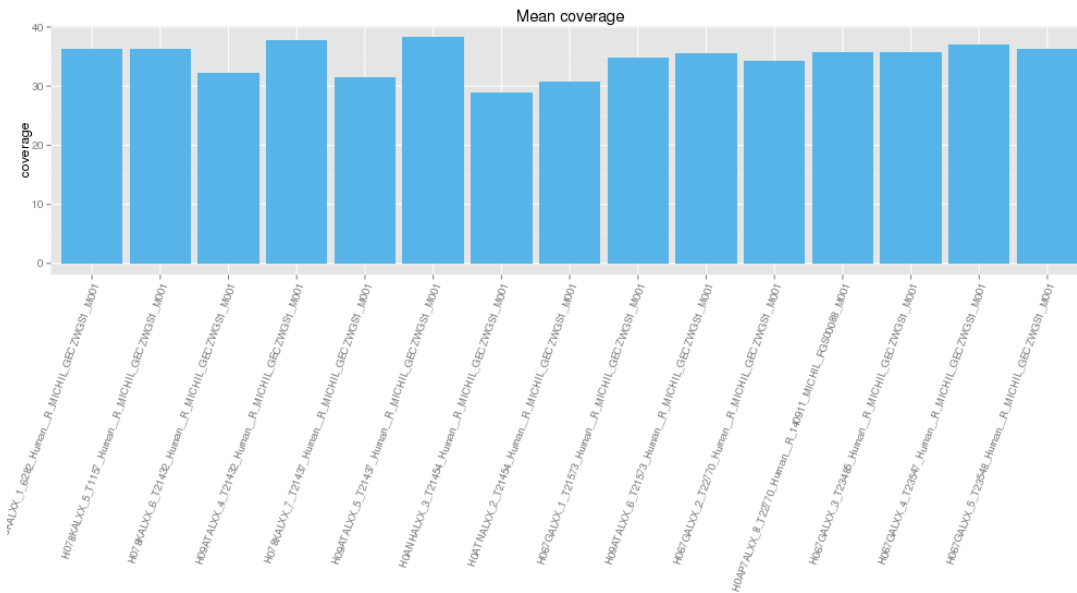
Read 2 Length: 150

Data Quality and Coverage

The following table summarises the performance of each indivisible unit of sequencing (IUS).

	IUS	Genome Territory	Mean Coverage (Illumina)	Yield (MB)	% >= Q30
1	H078KALXX_1_6282_Human__R_jondoe_XWGS1_M001	2864785223	36.30	130548	86.04
2	H078KALXX_5_T1157_Human__R_jondoe_XWGS1_M001	2864785223	36.25	128200	87.59
3	H078KALXX_6_T21432_Human__R_jondoe_XWGS1_M001	2864785223	32.22	120837	88.59
4	H09ATALXX_4_T21432_Human__R_jondoe_XWGS1_M001	2864785223	37.81	133808	86.78
5	H078KALXX_7_T21437_Human__R_jondoe_XWGS1_M001	2864785223	31.50	116627	88.59
6	H09ATALXX_5_T21437_Human__R_jondoe_XWGS1_M001	2864785223	38.35	136217	85.74
7	H0ANHALXX_3_T21454_Human__R_jondoe_XWGS1_M001	2864785223	28.90	119460	82.75

8	H0ATNALXX_2_T21454_Human__R_jondoe_XWGS1_M001	2864785223	30.73	138597	82.12
9	H067GALXX_1_T21573_Human__R_jondoe_XWGS1_M001	2864785223	34.85	131719	84.88
10	H09ATALXX_6_T21573_Human__R_jondoe_XWGS1_M001	2864785223	35.50	134133	90.43
11	H067GALXX_2_T22770_Human__R_jondoe_XWGS1_M001	2864785223	34.34	131672	85.06
12	HOAP7ALXX_8_T22770_Human__R_140911_jondoe_FGS00088_M001	2864785223	35.77	136757	89.56
13	H067GALXX_3_T23485_Human__R_jondoe_XWGS1_M001	2864785223	35.81	132053	86.56
14	H067GALXX_4_T23547_Human__R_jondoe_XWGS1_M001	2864785223	37.00	134321	87.23
15	H067GALXX_5_T23548_Human__R_jondoe_XWGS1_M001	2864785223	36.37	132965	87.09



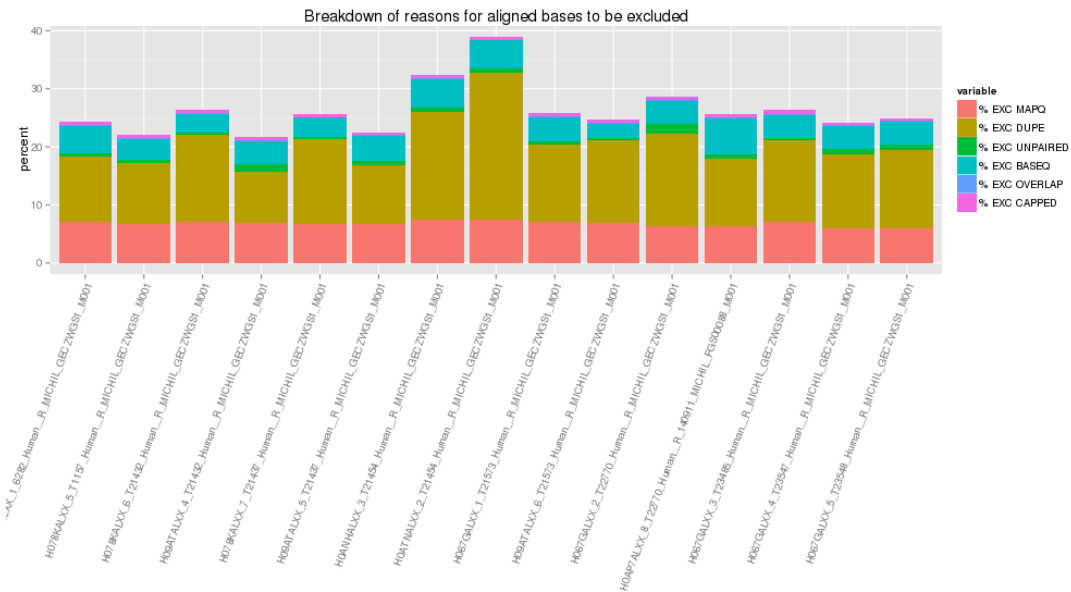
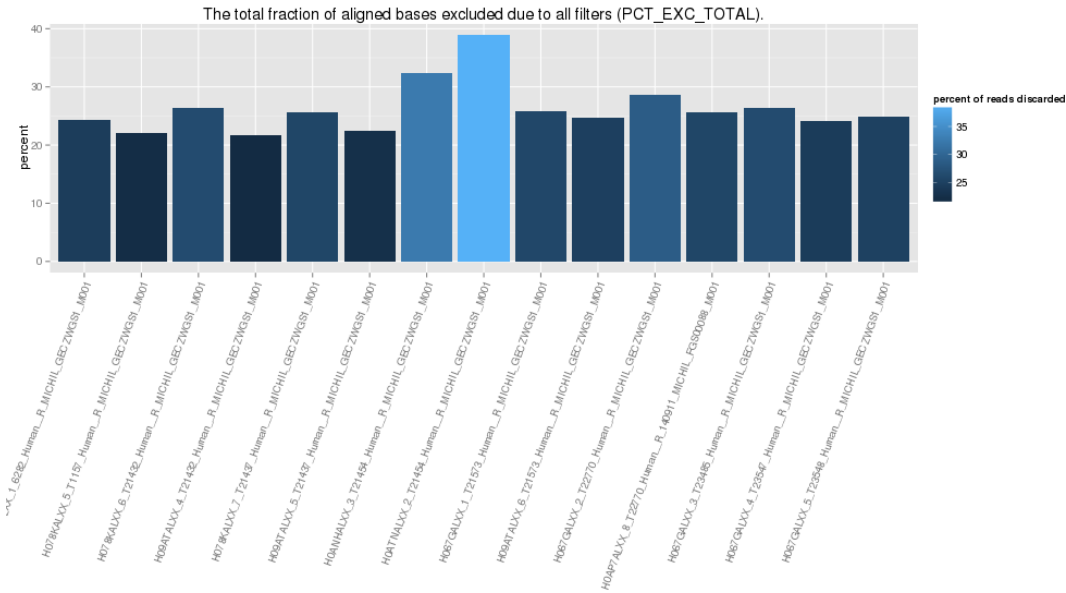
Poor Quality Reads

Some bases are filtered out from the alignments. This section shows a detailed breakdown of the criteria to exclude data and the percentage of reads excluded in each category.

- PCT_EXC_BASEQ: The fraction of aligned bases that were filtered out because they were of low base quality (default is < 20).
- PCT_EXC_CAPPED: The fraction of aligned bases that were filtered out because they would have raised coverage above the capped value (default cap = 250x).
- PCT_EXC_DUPE: The fraction of aligned bases that were filtered out because they were in reads marked as duplicates.
- PCT_EXC_MAPQ: The fraction of aligned bases that were filtered out because they were in reads with low mapping quality (default is < 20).
- PCT_EXC_OVERLAP: The fraction of aligned bases that were filtered out because they were the second observation from an insert with overlapping reads.

	IUS	% EXC TOTAL	% EXC MAPQ	% EXC DUPE	% EXC UNPAIRED	% EXC BASEQ	% EXC OVERLAP	% EXC CAPPED
1	H078KALXX_1_6282_Human__R_jondoe_XWGS1_M001	24.35	7.130	11.23	0.5979	4.532	0.3326	0.5296
2	H078KALXX_5_T1157_Human__R_jondoe_XWGS1_M001	22.03	6.653	10.58	0.4601	3.719	0.1667	0.4543
3	H078KALXX_6_T21432_Human__R_jondoe_XWGS1_M001	26.32	7.061	14.98	0.3664	3.172	0.1987	0.5396
4	H09ATALXX_4_T21432_Human__R_jondoe_XWGS1_M001	21.68	6.912	8.82	1.2707	3.889	0.2200	0.5711
5	H078KALXX_7_T21437_Human__R_jondoe_XWGS1_M001	25.58	6.686	14.70	0.3816	3.206	0.1571	0.4417
6	H09ATALXX_5_T21437_Human__R_jondoe_XWGS1_M001	22.49	6.694	10.19	0.6668	4.331	0.1583	0.4479
7	H09ATALXX_3_T21454_Human__R_jondoe_XWGS1_M001	32.42	7.522	18.42	0.8275	4.887	0.2144	0.5449

8	H0ATNALXX_2_T21454_Human__R_jondoe_XWGS1_M001	38.94	7.485	25.25	0.8335	4.710	0.1725	0.4834
9	H067GALXX_1_T21573_Human__R_jondoe_XWGS1_M001	25.78	7.058	13.39	0.5606	4.008	0.2908	0.4791
10	H09ATALXX_6_T21573_Human__R_jondoe_XWGS1_M001	24.68	6.873	14.34	0.2815	2.429	0.2464	0.5033
11	H067GALXX_2_T22770_Human__R_jondoe_XWGS1_M001	28.60	6.387	15.89	1.5992	4.085	0.2063	0.4341
12	H0AP7ALXX_8_T22770_Human__R_140911_jondoe_FGS00088_M001	25.57	6.367	11.51	0.8415	6.242	0.1750	0.4398
13	H067GALXX_3_T23485_Human__R_jondoe_XWGS1_M001	26.31	7.070	13.99	0.4142	4.034	0.1781	0.6178
14	H067GALXX_4_T23547_Human__R_jondoe_XWGS1_M001	24.18	5.960	12.80	0.9535	3.840	0.2033	0.4250
15	H067GALXX_5_T23548_Human__R_jondoe_XWGS1_M001	24.92	5.930	13.61	0.9149	3.873	0.1807	0.4096



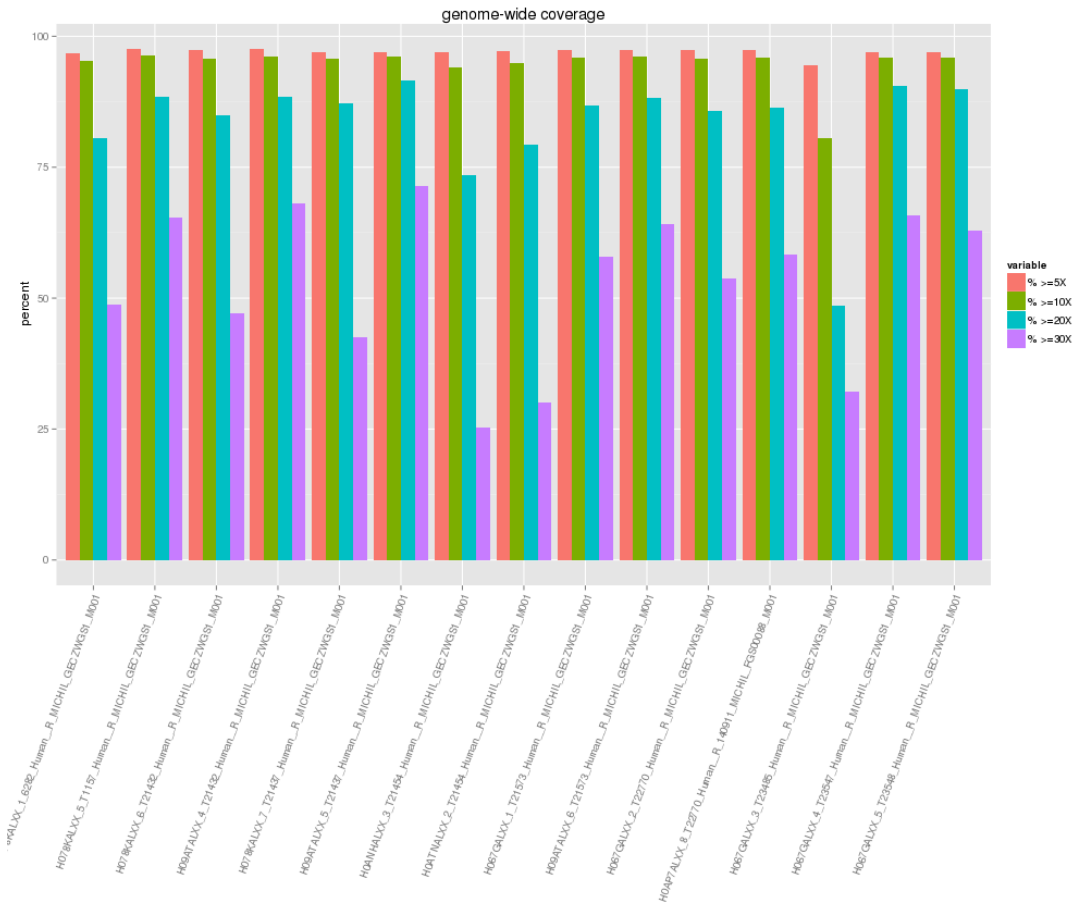
Coverage, expressed as % of target covered to certain depth

The figure and table below summarise the proportion of the genome that is sequenced above a certain average (e.g. >20X).

- PCT_5X: The fraction of bases that attained at least 5X sequence coverage in post-filtering bases.
- PCT_10X: The fraction of bases that attained at least 10X sequence coverage in post-filtering bases.
- PCT_20X: The fraction of bases that attained at least 20X sequence coverage in post-filtering bases.

PCT_30X: The fraction of bases that attained at least 30X sequence coverage in post-filtering bases.

	IUS	% >=5X	% >=10X	% >=20X	% >=30X
1	H078KALXX_1_6282_Human__R_jondoe_XWGS1_M001	96.85	95.27	80.63	48.74
2	H078KALXX_5_T1157_Human__R_jondoe_XWGS1_M001	97.57	96.32	88.54	65.45
3	H078KALXX_6_T21432_Human__R_jondoe_XWGS1_M001	97.42	95.73	84.82	47.01
4	H09ATALXX_4_T21432_Human__R_jondoe_XWGS1_M001	97.51	96.22	88.54	68.16
5	H078KALXX_7_T21437_Human__R_jondoe_XWGS1_M001	96.97	95.75	87.18	42.57
6	H09ATALXX_5_T21437_Human__R_jondoe_XWGS1_M001	97.07	96.07	91.53	71.50
7	H0ANHALXX_3_T21454_Human__R_jondoe_XWGS1_M001	97.04	93.98	73.55	25.30
8	H0ATNALXX_2_T21454_Human__R_jondoe_XWGS1_M001	97.26	94.90	79.30	29.97
9	H067GALXX_1_T21573_Human__R_jondoe_XWGS1_M001	97.35	95.88	86.85	57.99
10	H09ATALXX_6_T21573_Human__R_jondoe_XWGS1_M001	97.47	96.18	88.15	64.16
11	H067GALXX_2_T22770_Human__R_jondoe_XWGS1_M001	97.43	95.82	85.77	53.79
12	H0AP7ALXX_8_T22770_Human__R_140911_jondoe_FGS00088_M001	97.47	95.92	86.38	58.40
13	H067GALXX_3_T23485_Human__R_jondoe_XWGS1_M001	94.44	80.52	48.54	32.14
14	H067GALXX_4_T23547_Human__R_jondoe_XWGS1_M001	96.99	95.94	90.59	65.80
15	H067GALXX_5_T23548_Human__R_jondoe_XWGS1_M001	97.00	95.91	89.98	62.97



Methods

Libraries were created from DNA using [TruSeq Nano DNA HT Sample Prep Kit \(V1\)](#) protocol as per manufacturer's instructions. One sample was loaded per flow cell lane.

The flow cells were loaded onto an Illumina HiSeq X sequencer and 2x150bp paired-end sequencing was performed. The raw data

from the sequencers was converted to FastQ file format using Illumina's Bcl2FastQ v 2.15.0.4.

To verify data quality, FastQC was run on the FastQ files. Furthermore, the sequences are aligned to b37 human reference genome (human.g1k.v37) using Illumina's ISAAC aligner v01.14.07.17 ([Raczy et al. 2013](#); [link to whitepaper](#)) to generate BAM files. Additional quality metrics were calculated using [Picard WgsMetrics v1.119](#).

To calculate the mean coverage the following equation was applied as described by the [manufacturer](#).

$$((\text{MEAN_COVERAGE}) \times (1 - \text{PCT_EXC_DUPE} - \text{PCT_EXC_OVERLAP})) / (1 - \text{PCT_EXC_TOTAL})$$

Further Information

For more details please visit the [KCCG Public Wiki](#).

Garvan Institute of Medical Research

384 Victoria Street, Darlinghurst, NSW 2010

T: + 61 (0)2 9355 5846 | E: kccgseq@garvan.org.au

<http://www.garvan.org.au/research/clinical-genomics>

<https://cgg.garvan.org.au/confluence/display/KP/KCCG+Public+Wiki>