

Sequencing by Solexa

Outi Monni, PhD

*Biomedicum Genomics
and
Institute of Biomedicine
Genome-Scale Biology Research Program
University of Helsinki*

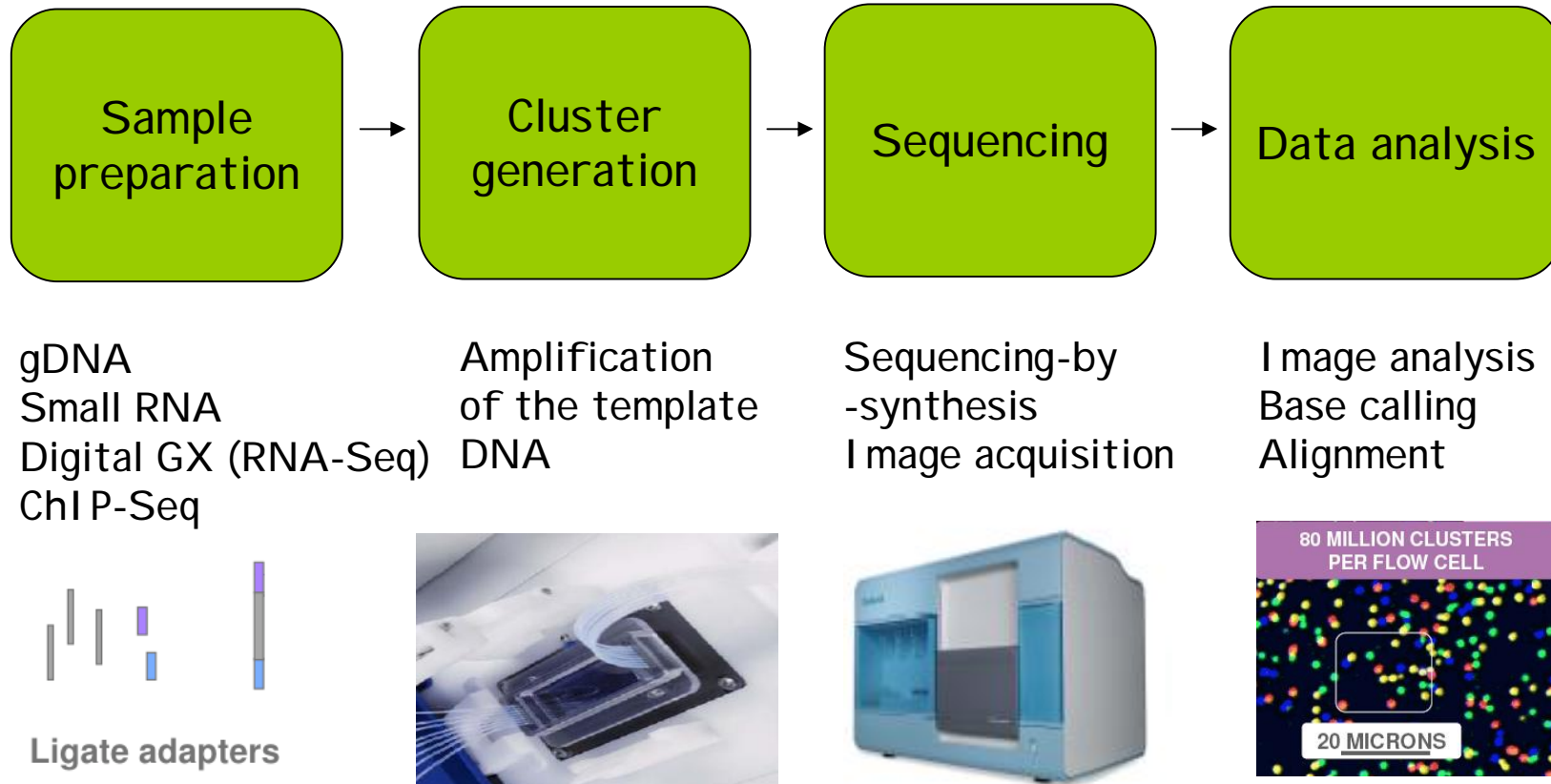
Solexa equipment in Biomedicum

- Two Genome Analyzers (GAII)
 - FIMM: Biomedicum 2U (P-floor)
 - Paired-end module
 - Contact person: Päivi Lahermo
 - Biochip Center (BMGen): Biomedicum 1 (P-floor)
 - Single-read sequencing, cluster station
 - Contact person: Outi Monni
- Services
 - Offered jointly through the common portal

Applications

- Whole genome sequencing and re-sequencing
 - Mutation detection
 - SNP genotyping
 - Mapping of genomic rearrangements
 - Copy number profiling
- Gene expression profiling (RNA-seq)
- Discovery and analysis of small RNA (miRNA-seq)
- Chromatin immunoprecipitation sequencing (ChIP-seq)

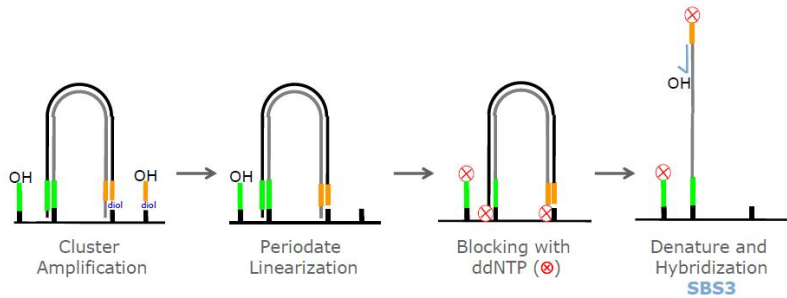
Solexa workflow



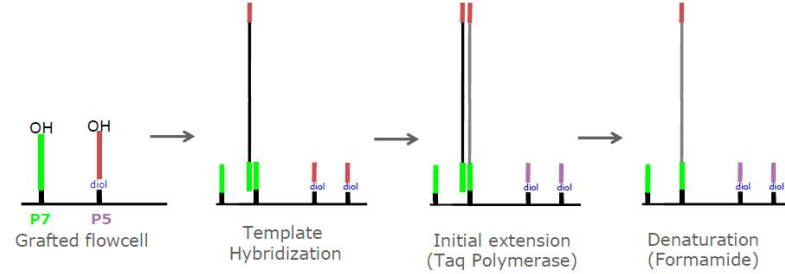
Clusteration steps

1. Hybridization of template DNA
2. Amplification of template DNA
3. Linearization
4. Blocking
5. Denaturation and hybridization of sequencing primer

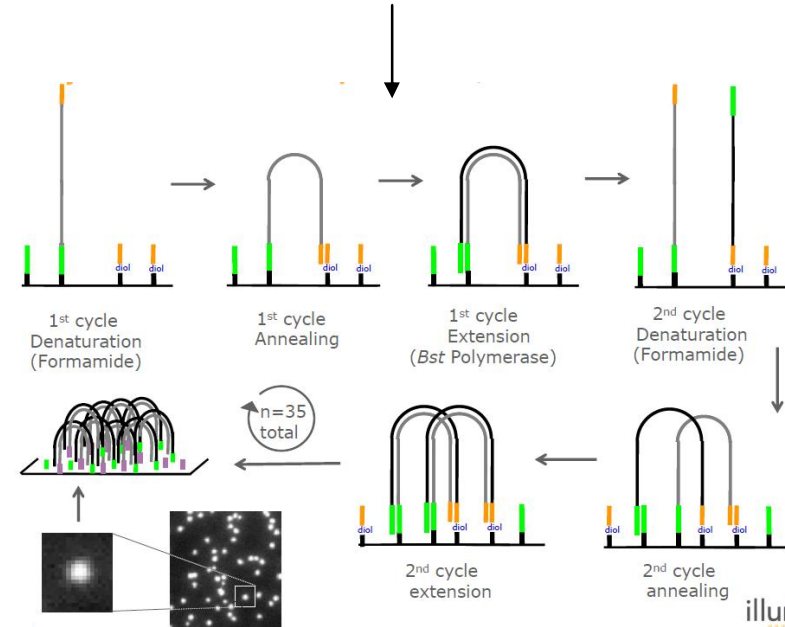
3.



1.

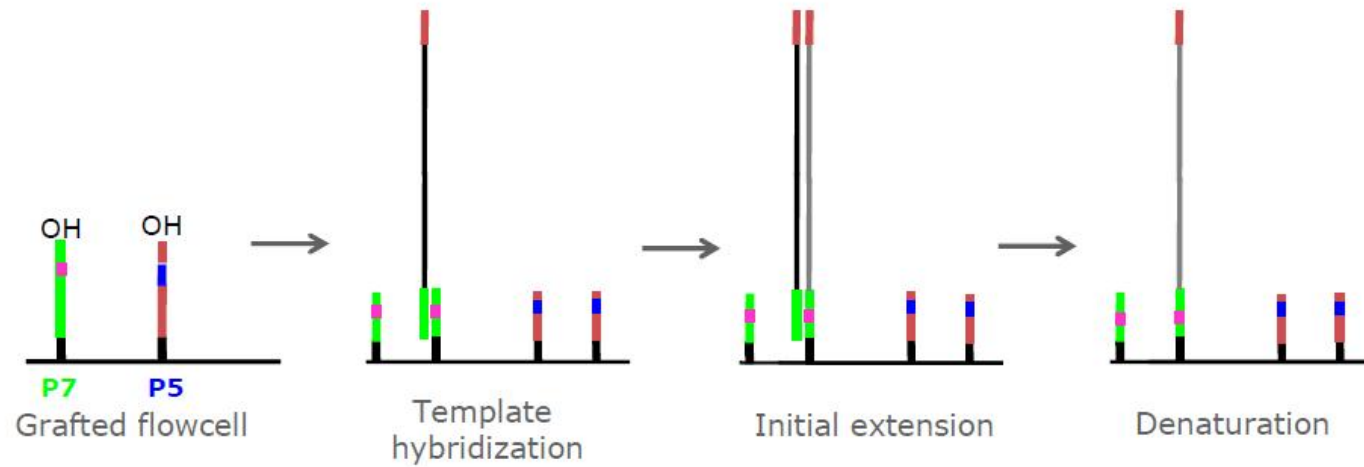


2.

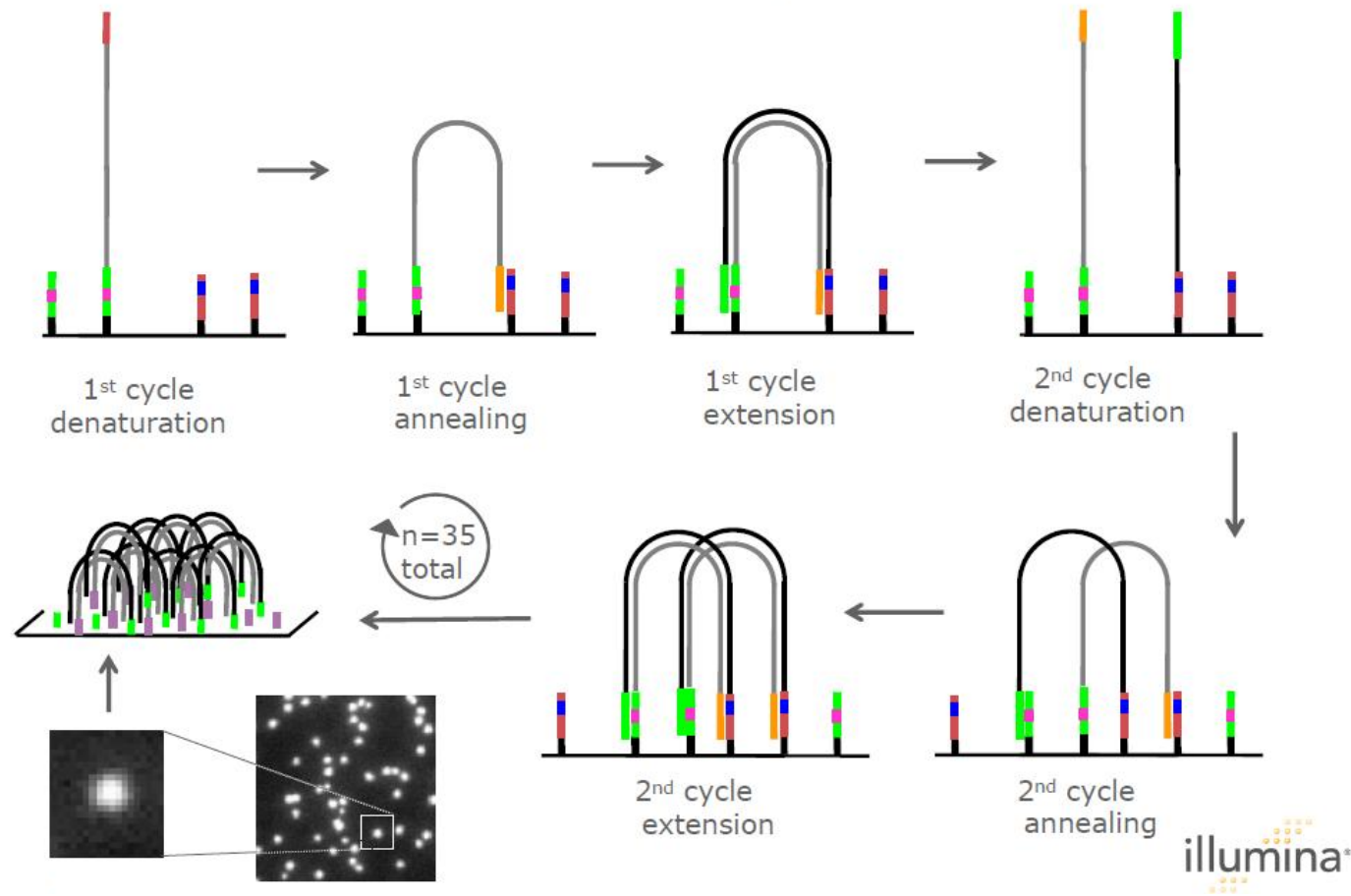


illumina

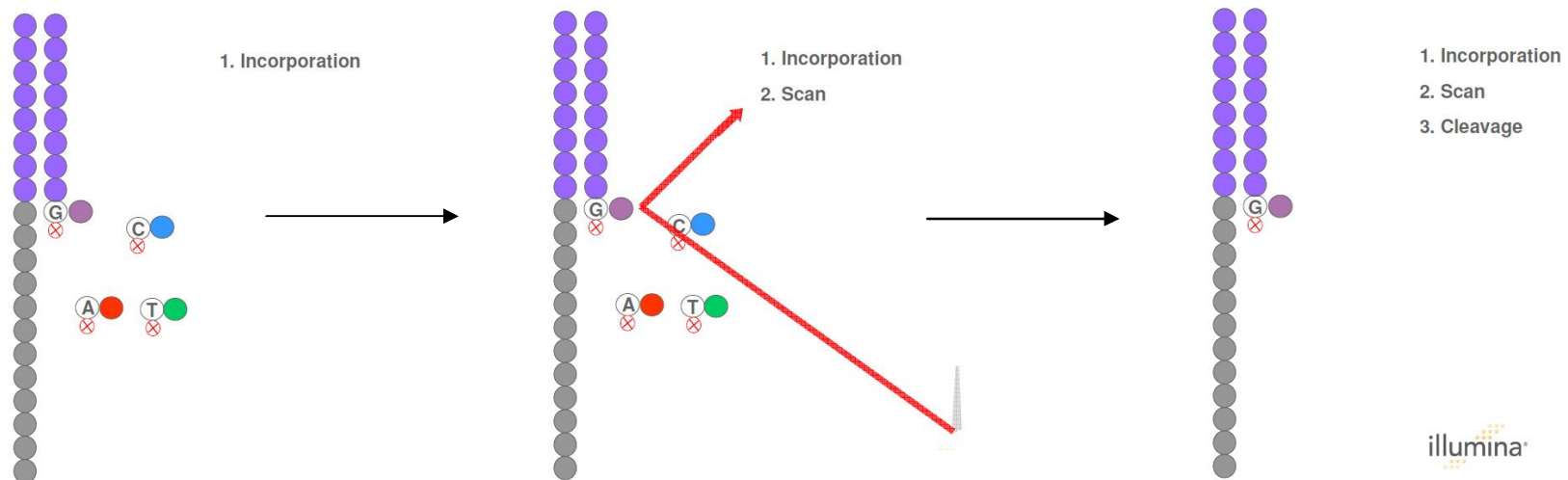
Paired-end Cluster Generation: Initial Extension



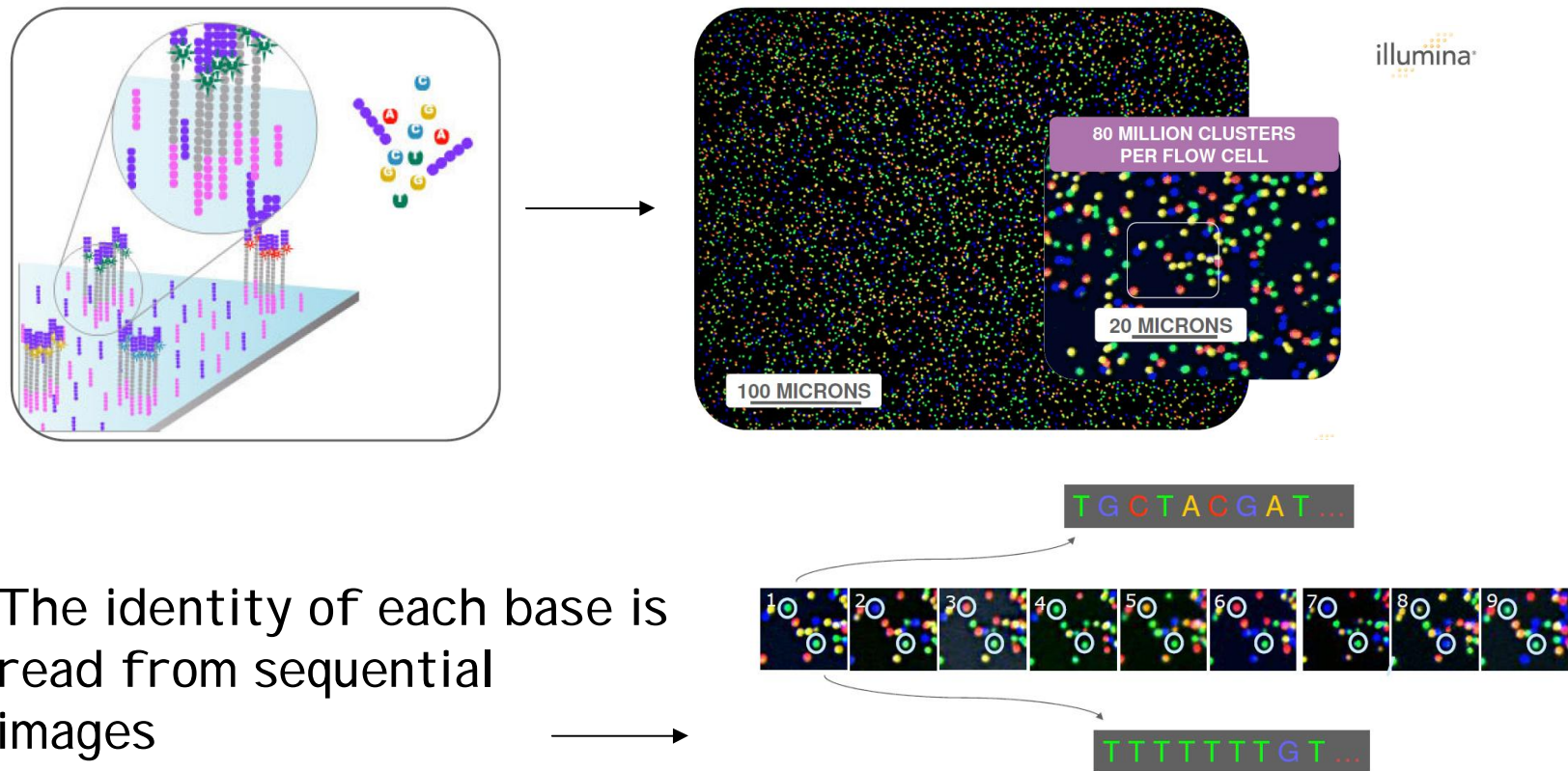
Paired-end Cluster Generation: Amplification



Sequencing-by-synthesis by Genome Analyzer II



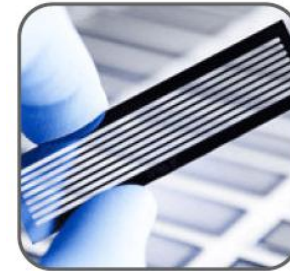
Sequencing-by-synthesis by Genome Analyzer II



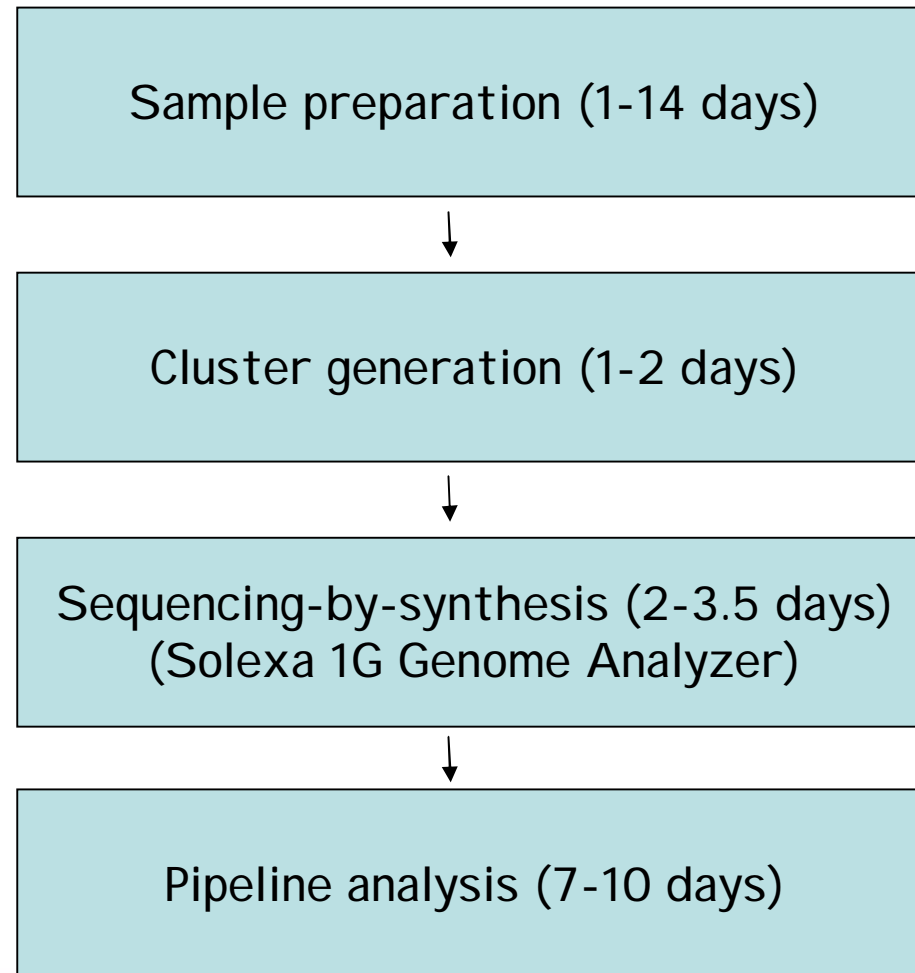
The identity of each base is read from sequential images

Specifications of Solexa sequencing

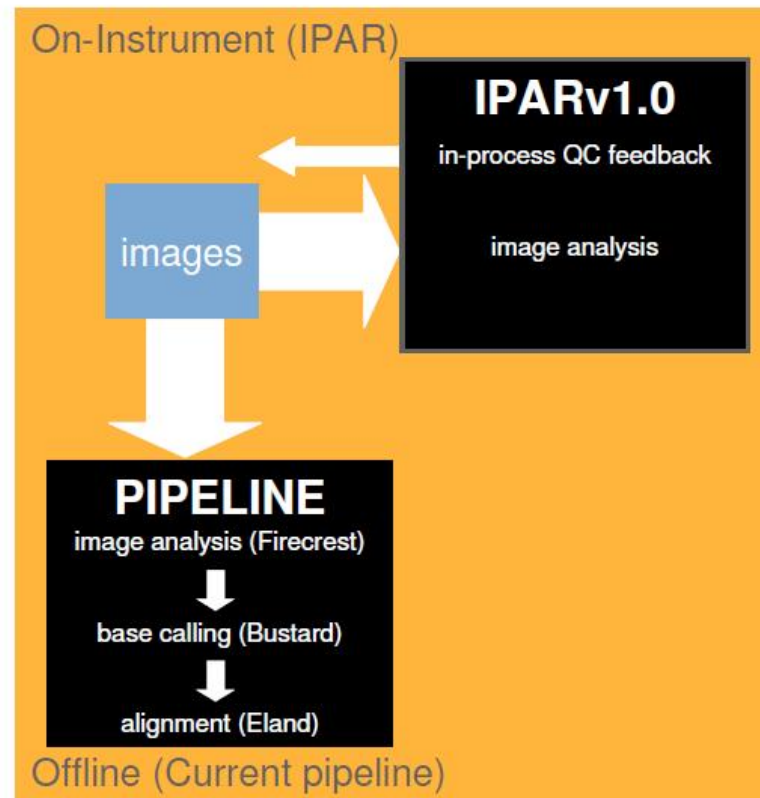
- 8 lanes per flow cell = 8 samples
- Each lane consists of 100 tiles
- supported read length 35 base pairs
- 85.000 - 130.000 clusters per tile
- 68 - 106 million (106.000.000) clusters per flow cell
- >187 megabases (187.500.000) per lane
- >1.5 gigabases (1.500.000.000) per flow cell (single-read)
- >3.0 gigabases per flow cell (paired-end read)
- raw accuracy 98.5%



Throughput (max 7 samples/run)

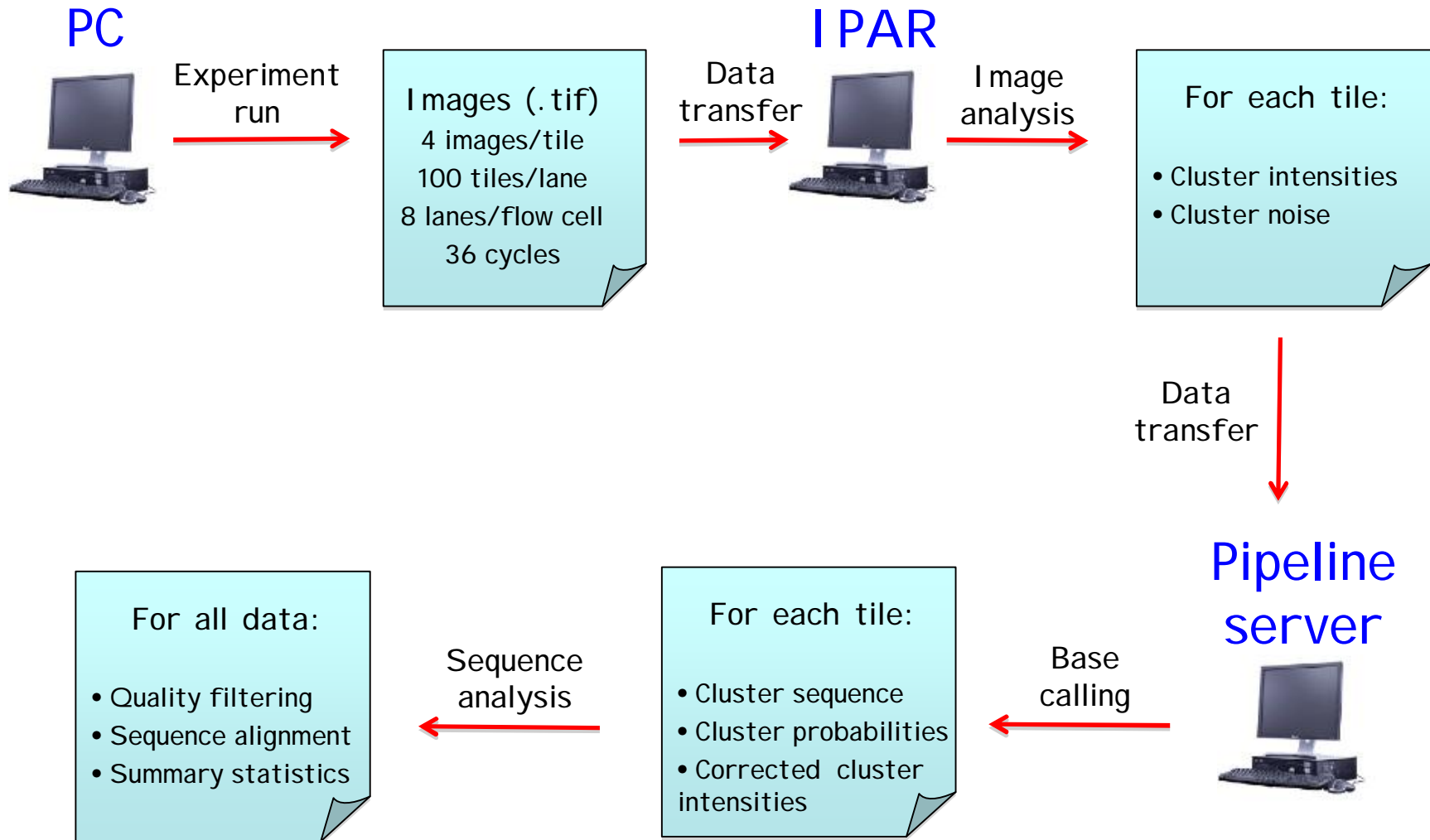


IPAR - Integrated primary analysis and reporting

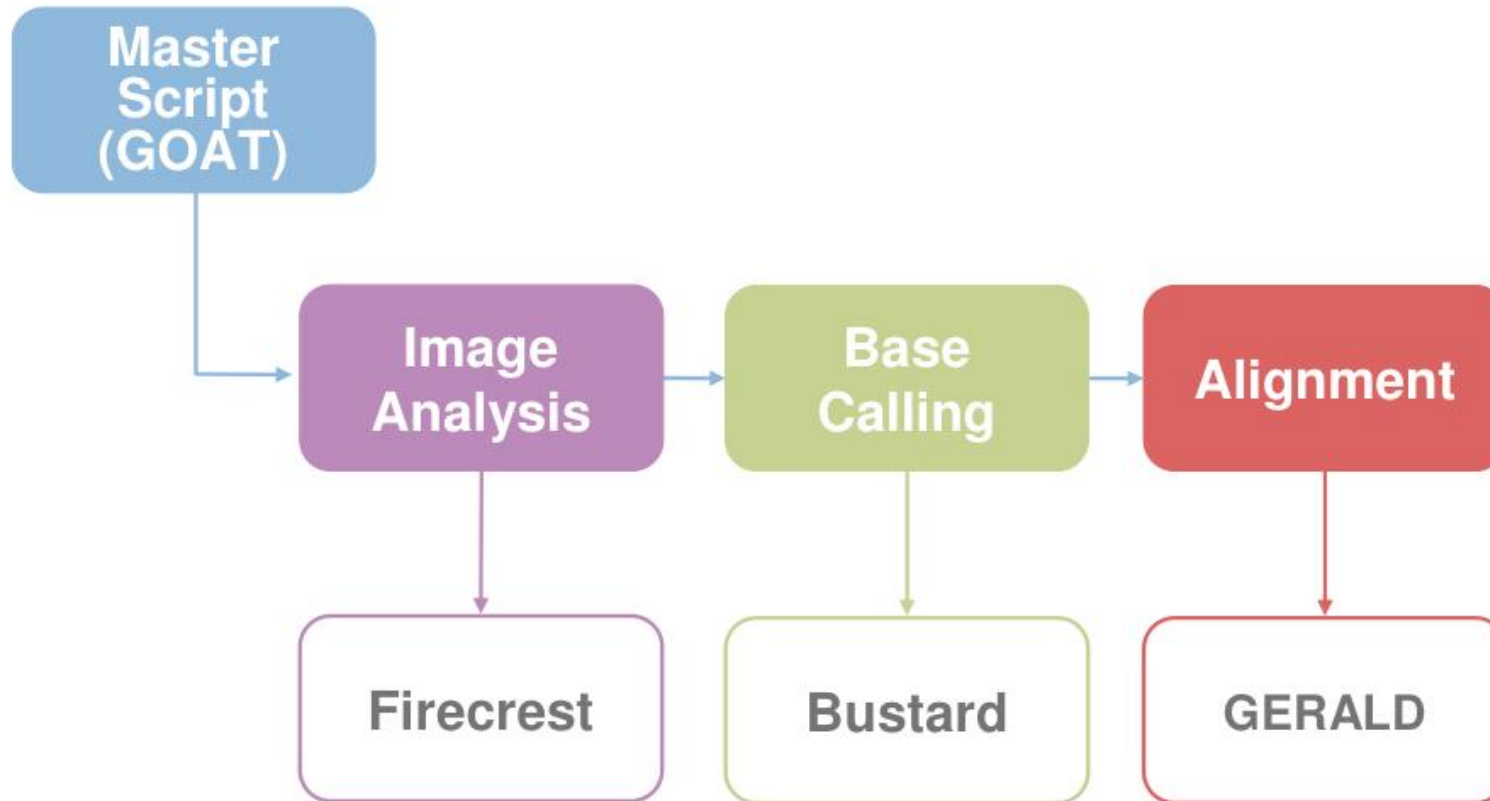


Data Analysis Workflow

Instrument



Genome Analyzer Pipeline Software



Analysis Run Statistics

Lane results summary

Lane Info		Tile Mean +/- SD for Lane							
Lane	Lane Yield (kbases)	Clusters (raw)	Clusters (PF)	1st Cycle Int (PF)	% intensity after 20 cycles (PF)	% PF Clusters	% Align (PF)	Alignment Score (PF)	% Error Rate (PF)
1	195110	105773 +/- 9005	65036 +/- 3289	367 +/- 45	88.55 +/- 3.95	61.81 +/- 4.52	94.92 +/- 1.18	66.91 +/- 5.02	0.48 +/- 0.09

Expanded lane summary

Lane Info		Phasing Info		Raw Data (tile mean)		Filtered Data (tile mean)						
Lane	Clusters (tile mean) (raw)	% Phasing	% Prephasing	% Error Rate (raw)	Equip Perfect Clusters (raw)	% retained	Cycle 2-4 Av Int (PF)	Cycle 2-10 Av % Loss (PF)	Cycle 10-20 Av % Loss (PF)	% Align (PF)	% Error Rate (PF)	Equip Perfect Clusters (PF)
1	105773	1.0700	0.6000	0.94	70942	61.81	334 +/- 44	0.77 +/- 0.49	-0.25 +/- 0.29	94.92	0.48	60324

Analysis Run Statistics - Examples 1

Lane results summary

Lane Info		Tile Mean +/- SD for Lane							
Lane	Lane Yield (kbases)	Clusters (raw)	Clusters (PF)	1st Cycle Int (PF)	% intensity after 20 cycles (PF)	% PF Clusters	% Align (PF)	Alignment Score (PF)	% Error Rate (PF)
1	195110	105773 +/- 9005	65036 +/- 3289	367 +/- 45	88.55 +/- 3.95	61.81 +/- 4.52	94.92 +/- 1.18	66.91 +/- 5.02	0.48 +/- 0.09

Expanded lane summary

Lane Info		Phasing Info		Raw Data (tile mean)		Filtered Data (tile mean)						
Lane	Clusters (tile mean) (raw)	% Phasing	% Prephasing	% Error Rate (raw)	Equiv Perfect Clusters (raw)	% retained	Cycle 2-4 Av Int (PF)	Cycle 2-10 Av % Loss (PF)	Cycle 10-20 Av % Loss (PF)	% Align (PF)	% Error Rate (PF)	Equiv Perfect Clusters (PF)
1	105773	1.0700	0.6000	0.94	70942	61.81	334 +/- 44	0.77 +/- 0.49	-0.25 +/- 0.29	94.92	0.48	60324

Analysis Run Statistics - Examples 2

Lane results summary

Lane Info		Tile Mean +/- SD for Lane							
Lane	Lane Yield (kbases)	Clusters (raw)	Clusters (PF)	1st Cycle Int (PF)	% intensity after 20 cycles (PF)	% PF Clusters	% Align (PF)	Alignment Score (PF)	% Error Rate (PF)
1	195110	105773 +/- 9005	65036 +/- 3289	367 +/- 45	88.55 +/- 3.95	61.81 +/- 4.52	94.92 +/- 1.18	66.91 +/- 5.02	0.48 +/- 0.09

Expanded lane summary

Lane Info		Phasing Info		Raw Data (tile mean)		Filtered Data (tile mean)						
Lane	Clusters (tile mean) (raw)	% Phasing	% Prephasing	% Error Rate (raw)	Equip Perfect Clusters (raw)	% retained	Cycle 2-4 Av Int (PF)	Cycle 2-10 Av Loss (PF)	Cycle 10-20 Av Loss (PF)	% Align (PF)	% Error Rate (PF)	Equip Perfect Clusters (PF)
1	105773	1.0700	0.6000	0.94	70942	61.81	334 +/- 44	0.77 +/- 0.49	-0.25 +/- 0.29	94.92	0.48	60324

Analysis Run Text Files - Overview

Intermediate Files

Image Analysis

- *_int.txt
- *_nse.txt

Base Calling

- *_prb.txt
- *_qhg.txt
- *_seq
- *_sig2.txt

Alignment

- *_eland.txt
- *_align.txt
- *_prealign.txt
- *_score.txt
- *_prescore.txt
- *_qraw.txt
- *_qval.txt

Final Output Files

Alignment

- *_realign.txt (ELAND)
- *_rescore.txt (ELAND)
- *_sequence.txt
- *_qcal.txt
- *_qcalreport.txt
- *_export.txt (ELAND Extended)
- *_anomaly.txt (ELAND Pairs)

Datafiles take 150 Gb
of space

BMGen Solexa services

- To get more information on the available services, go to www.helsinki.fi/biomedicumgenomics
- More information:
 - Outi Monni (outi.monni@helsinki.fi)
 - Max Gentile (massimiliano.gentile@helsinki.fi)
 - BMGen staff (bmgen-support@helsinki.fi)



Hanna
Pesonen



Taru
Perttula