Introduction to high-throughput deep-sequencing platforms: 454 (Roche), SOLiD (ABI) en Solexa (Illumina)

Wouter Van Delm
wouter.vandelm@vib.be

VIB MicroArray Facility

Basic Problem

Researcher:	which mutations in a genomic region "REG" cause disease "DIS" ?
Experiment:	?? … … how do we collect data about the DNA sequence?

Inspiration from cell biology: *http://www.youtube.com/watch?v=983lhh20rGY*

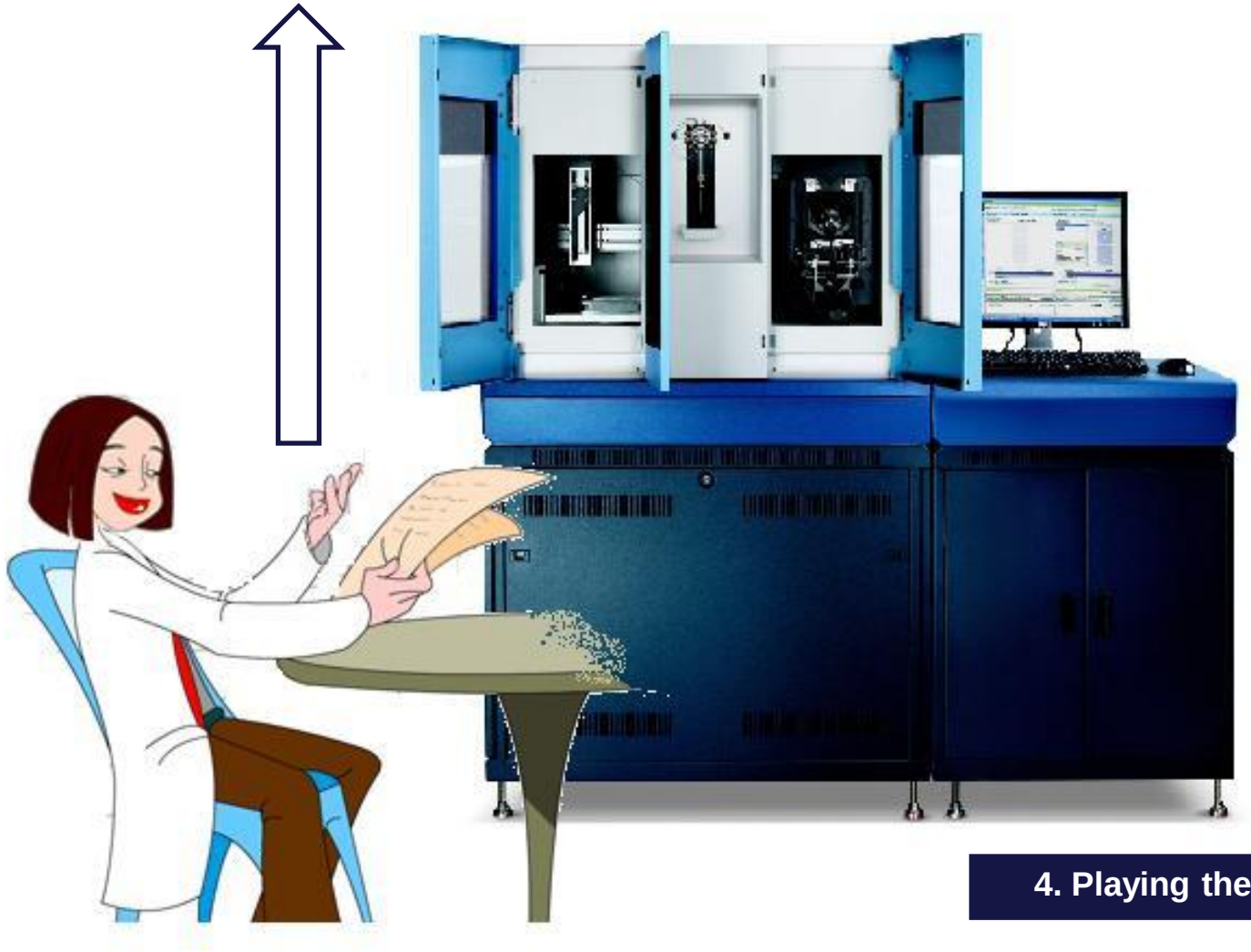
↓

Sequencing-by-synthesis:

Can we also read the genome base-by-base and output not mRNA but a text-file?

Overview

1. Speaking the Language	2. Mastering the Lab	3. Understanding Data Files

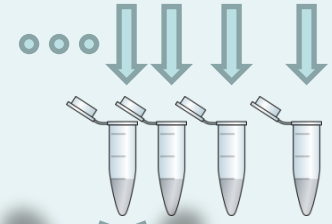


4. Playing the Game!

Speaking the language (1/2)

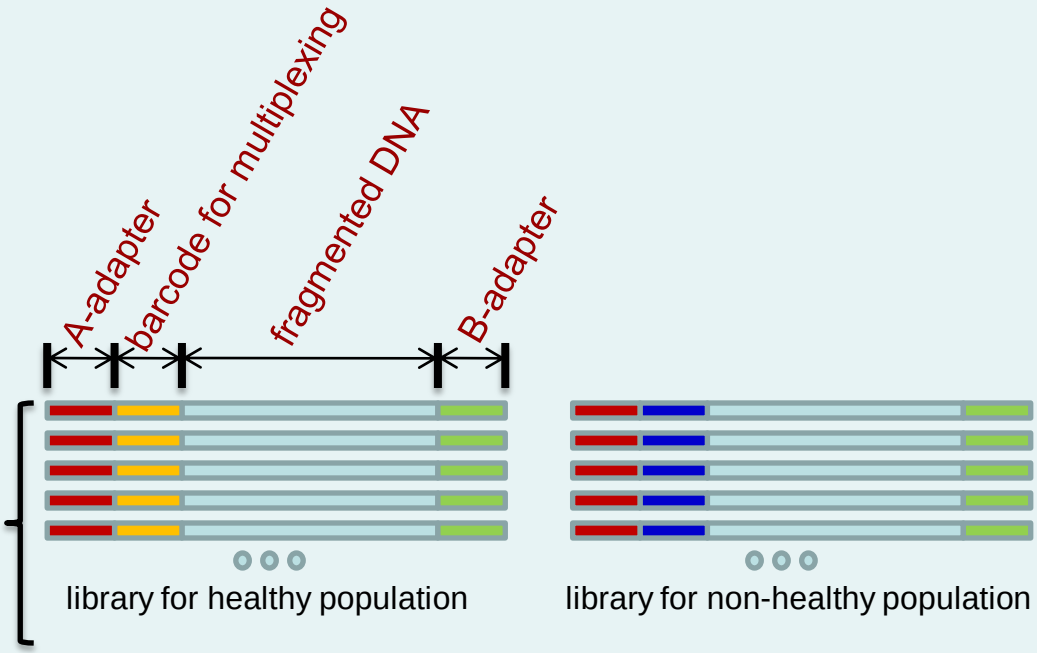
Sample preparation

biological replicates



technical replicates

Library preparation

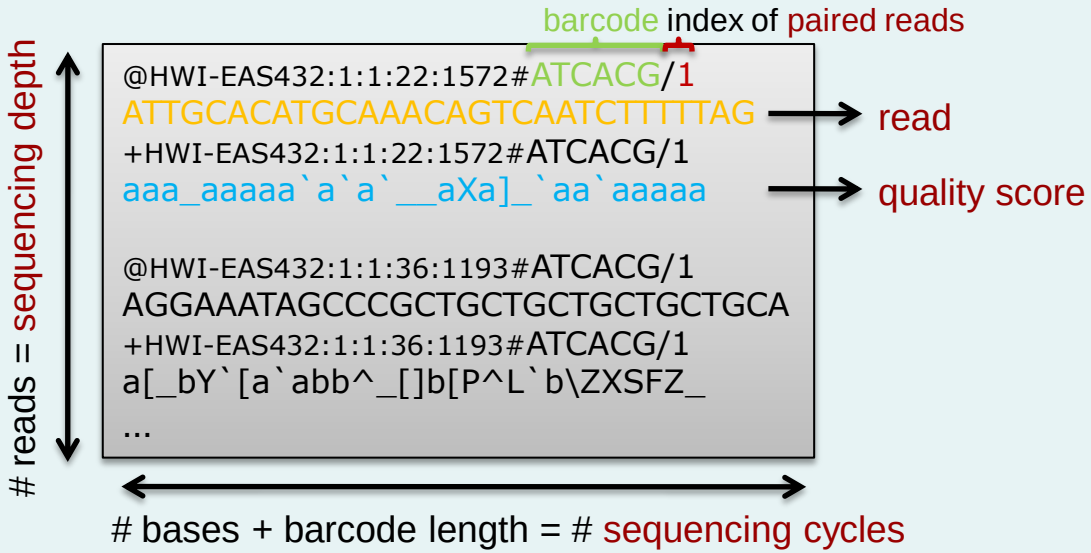


Speaking the Language (2/2)

Sequencing

- Amplification
- Sequencing chemistry
- Image processing
- Base calling

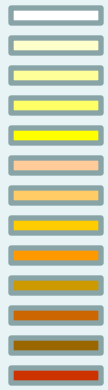
Fastq-output



De novo assembly

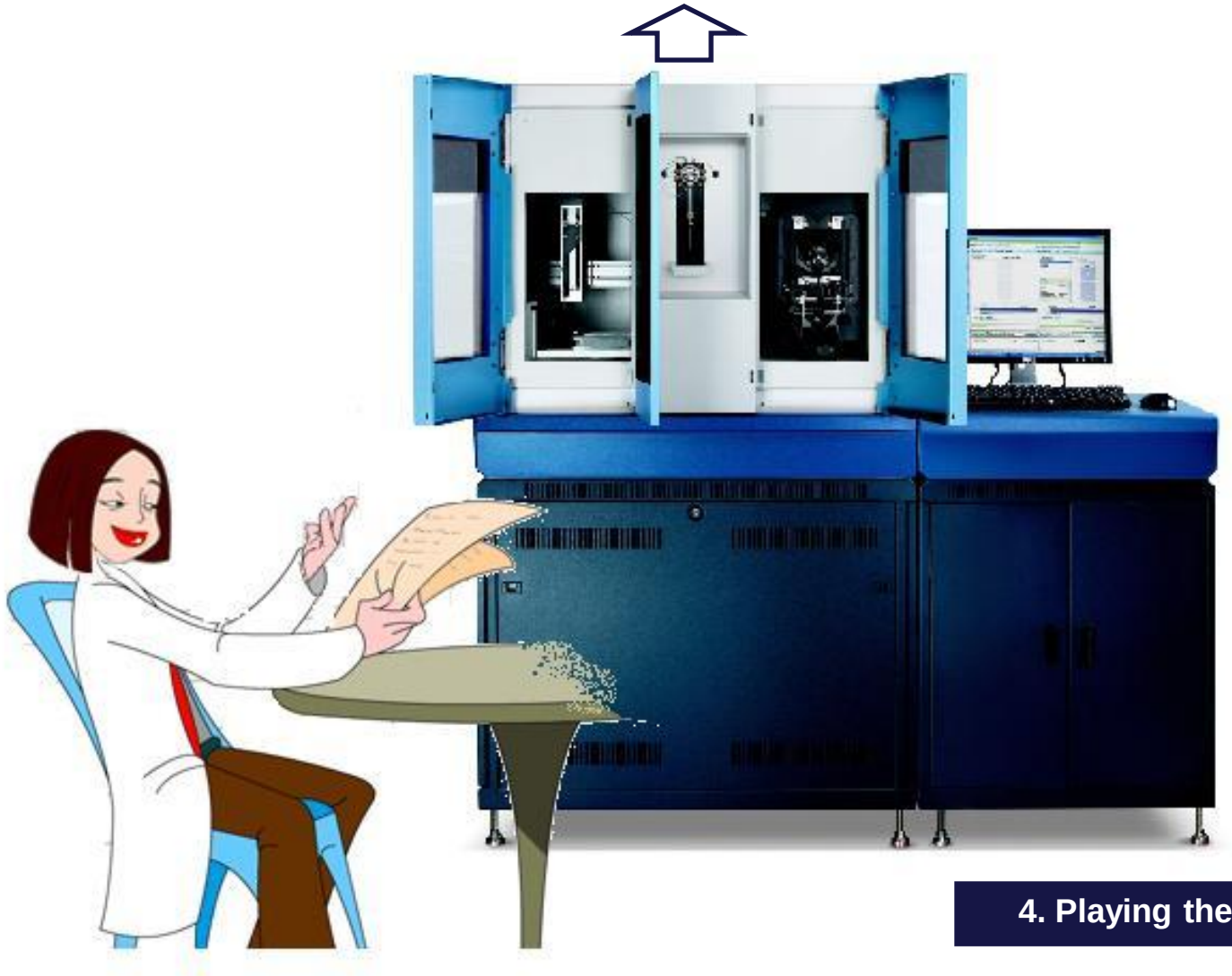
contigs

Alignment



Overview

1. Speaking the Language
2. Mastering the Lab
3. Understanding Data Files



4. Playing the Game!

Mastering the Lab

	GS FLX Titanium	GA IIx Solexa	SOLiD 3 Plus
Sample requirements	>5 µg of dsDNA in a 10 µL volume	0.1-1µg of dsDNA ; 10µg for mate pairs	10ng-5µg of dsDNA ; 5µg-20µg for mate pairs
Multiplexing	16 regions x 12 barcodes	8 lanes x 12 barcodes	2 cells x 8 samples/cell x 16 barcodes
Paired ends separation	3kb-20kb (half of the read length for each end)	200bp-10kb	600bp-10kb
Amplification	Emulsion PCR	Bridge amplification	Emulsion PCR
Sequencing chemistry	Pyrosequencing	Polymerase based sequencing-by-synthesis	Sequencing by ligation
Read length	*400 bp average - 500 bp modal (variable)*	2 x 35, 50, 75 and 100bp (fixed)	35 and 50 (fixed)
Reads per run	*> 1 M*	225-250M	>1000M
Mbp per run	*400 Mbp – 600 Mbp*	25 Gbp	60Gbp
Accuracy	*99% accuracy at the 400th base and higher for preceding bases*	99,9% accuracy and higher for more than 70% of the bases;	99,9% accuracy and higher for more than 80% of the bases
Specifics	no complex optics or lasers		data is recorded in ABI color space

SOLiD (ABI)

Introduction video	*http://www.appliedbiosystems.com/solid4*

	SOLiD 3 Plus
Sample requirements	10ng-5µg of dsDNA ; 5µg-20µg for mate pairs
Multiplexing	2 cells x 8 samples/cell x 16 barcodes
Paired ends separation	600bp-10kb
Amplification	Emulsion PCR
Sequencing chemistry	Sequencing by ligation
Read length	35 and 50 (fixed)
Reads per run	>1000M
Mbp per run	60Gbp
Accuracy	99,9% accuracy and higher for more than 80% of the bases
Specifics	data is recorded in ABI color space

Solexa (Illumina)

Introduction video	*http://www.illumina.com/technology/sequencing_technology.ilmn*

	GA IIx Solexa
Sample requirements	0.1-1µg of dsDNA ; 10µg for mate pairs
Multiplexing	8 lanes x 12 barcodes
Paired ends separation	200bp-10kb
Amplification	Bridge amplification
Sequencing chemistry	Polymerase based sequencing-by-synthesis
Read length	2 x 35, 50, 75 and 100bp (fixed)
Reads per run	225-250M
Mbp per run	25 Gbp
Accuracy	99,9% accuracy and higher for more than 70% of the bases;
Specifics	

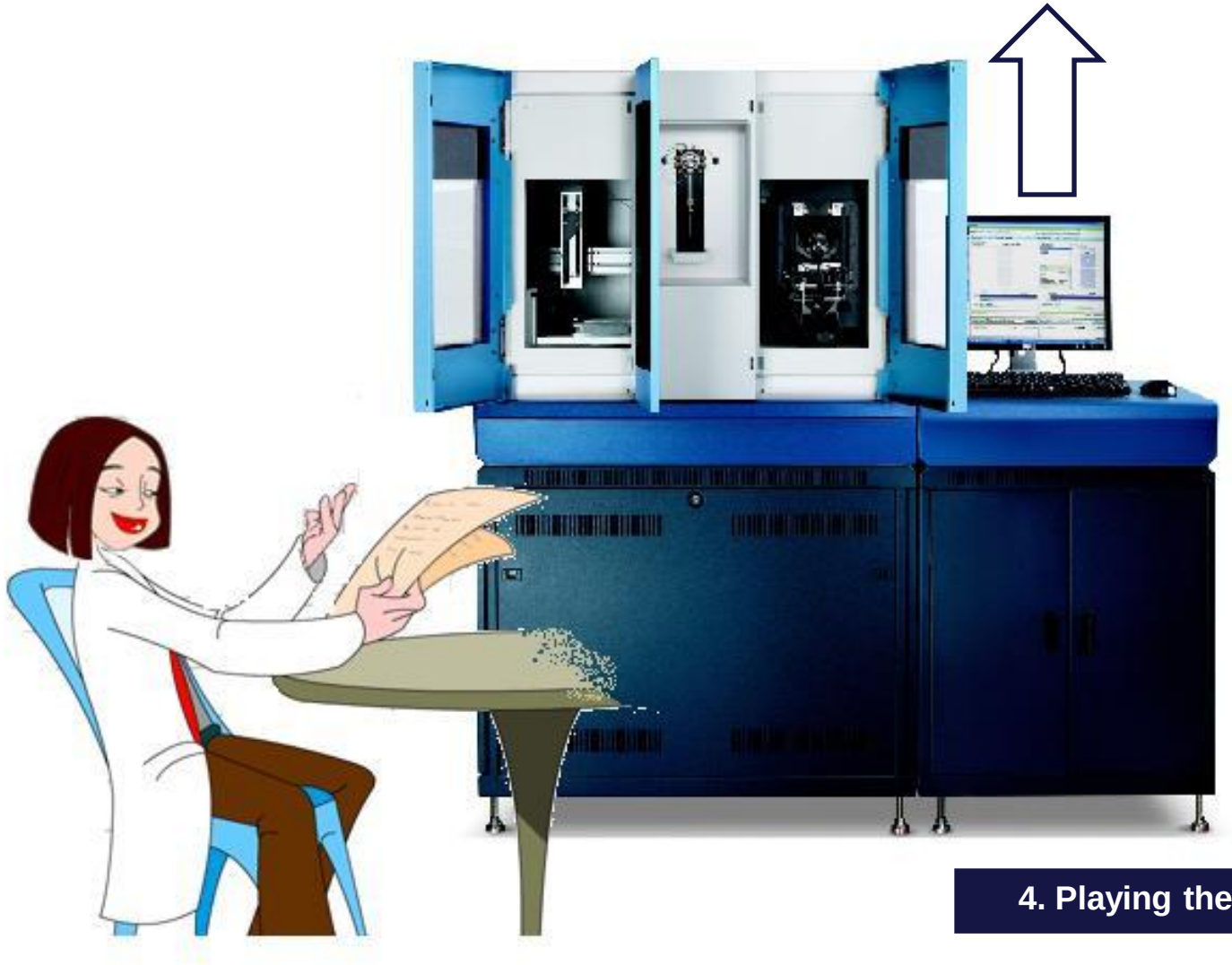
454 platform (Roche)

Introduction video	*http://www.youtube.com/watch?v=bFNjxKHP8Jc*

	GS FLX Titanium
Sample requirements	>5 µg of dsDNA in a 10 µL volume
Multiplexing	16 regions x 12 barcodes
Paired ends separation	3kb-20kb (half of the read length for each end)
Amplification	Emulsion PCR
Sequencing chemistry	Pyrosequencing
Read length	*400 bp average - 500 bp modal (variable)*
Reads per run	*> 1 M*
Mbp per run	*400 Mbp – 600 Mbp*
Accuracy	*99% accuracy at the 400th base and higher for preceding bases*
Specifics	no complex optics or lasers

Overview

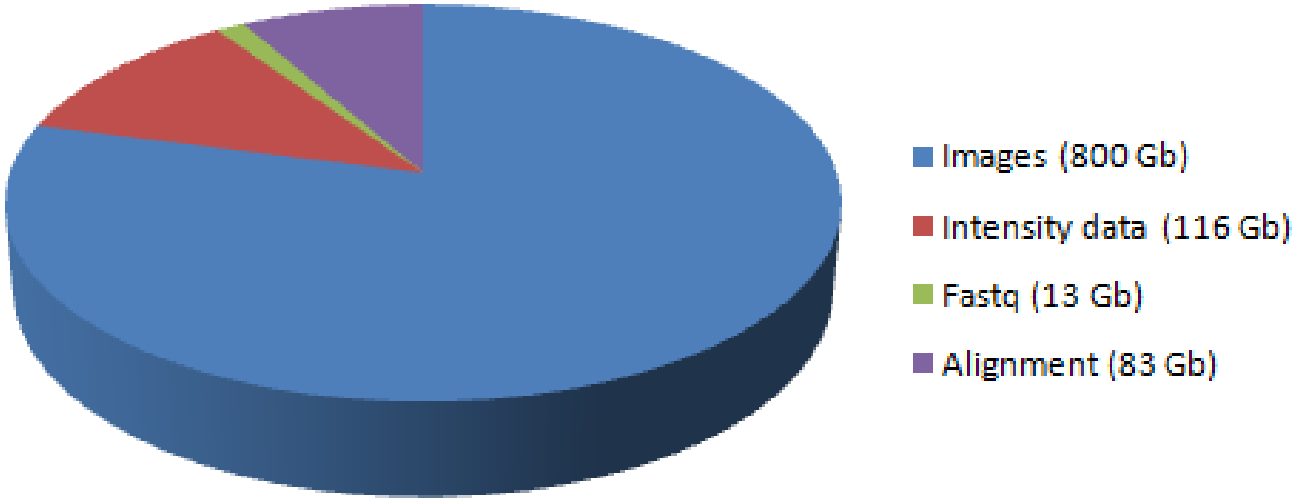
1. Speaking the Language	2. Mastering the Lab	3. Understanding Data Files



4. Playing the Game!

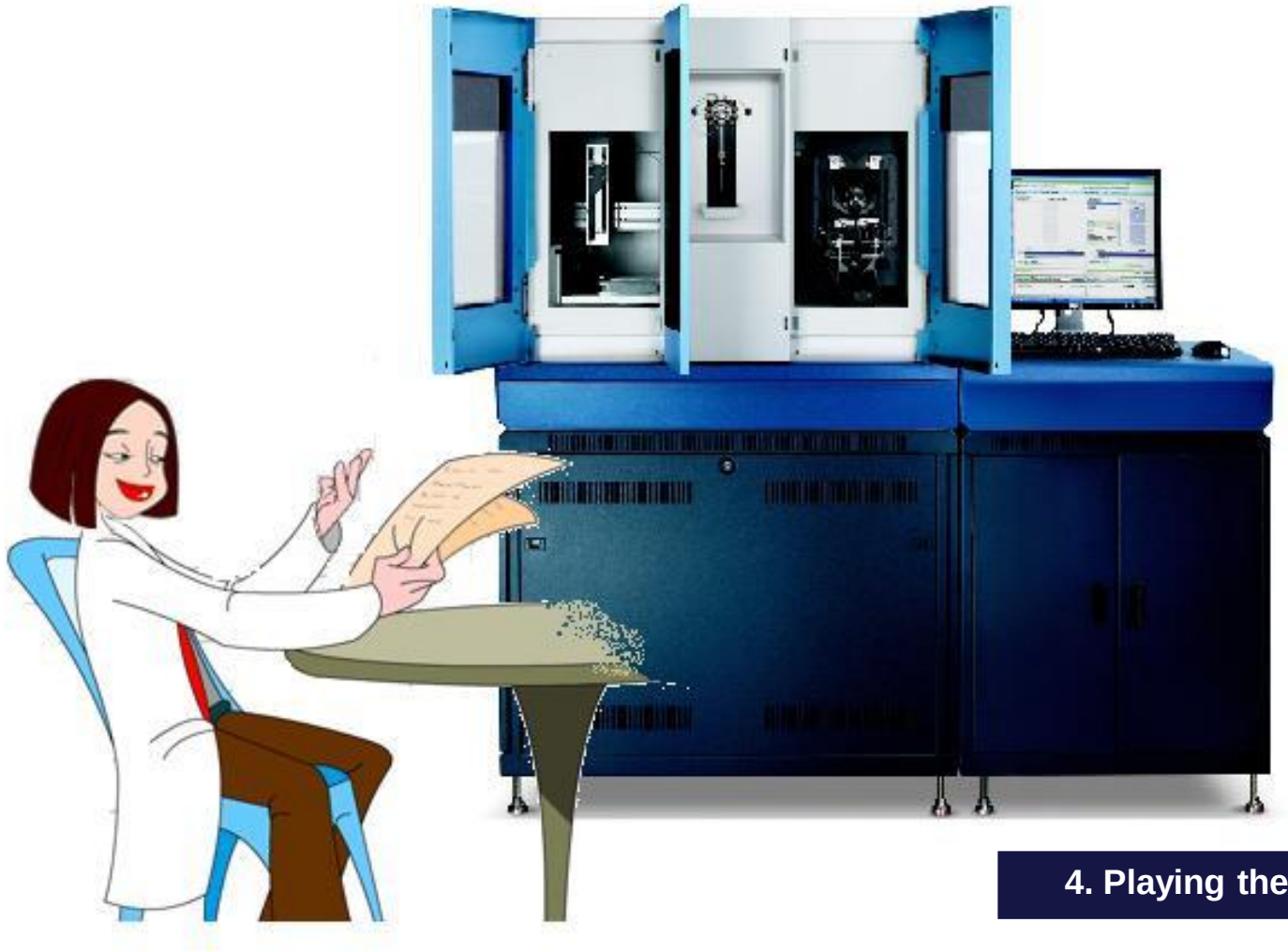
Understanding Data Files

Platform	**Solexa – GA II / GAP v1.3**
Single reads	1000 Gb for 8 lanes
Paired reads & 76 cycles	4000 Gb for 8 lanes



Overview

1. Speaking the Language	2. Mastering the Lab	3. Understanding Data Files



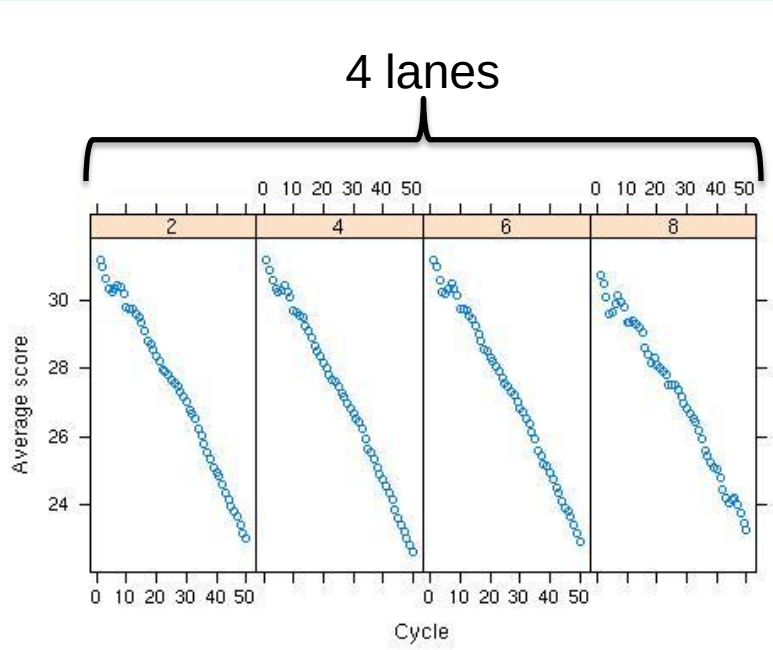
4. Playing the Game!

Playing the Game

Introduction video	*http://www.illumina.com/technology/sequencing_technology.ilmn*

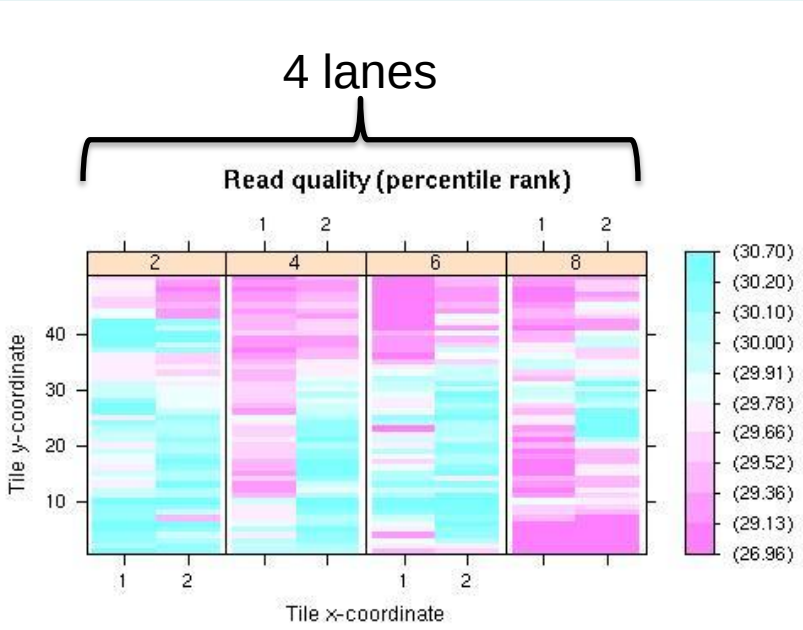
Quality Control Report

Base quality per cycle



50 cycles

Base quality per tile



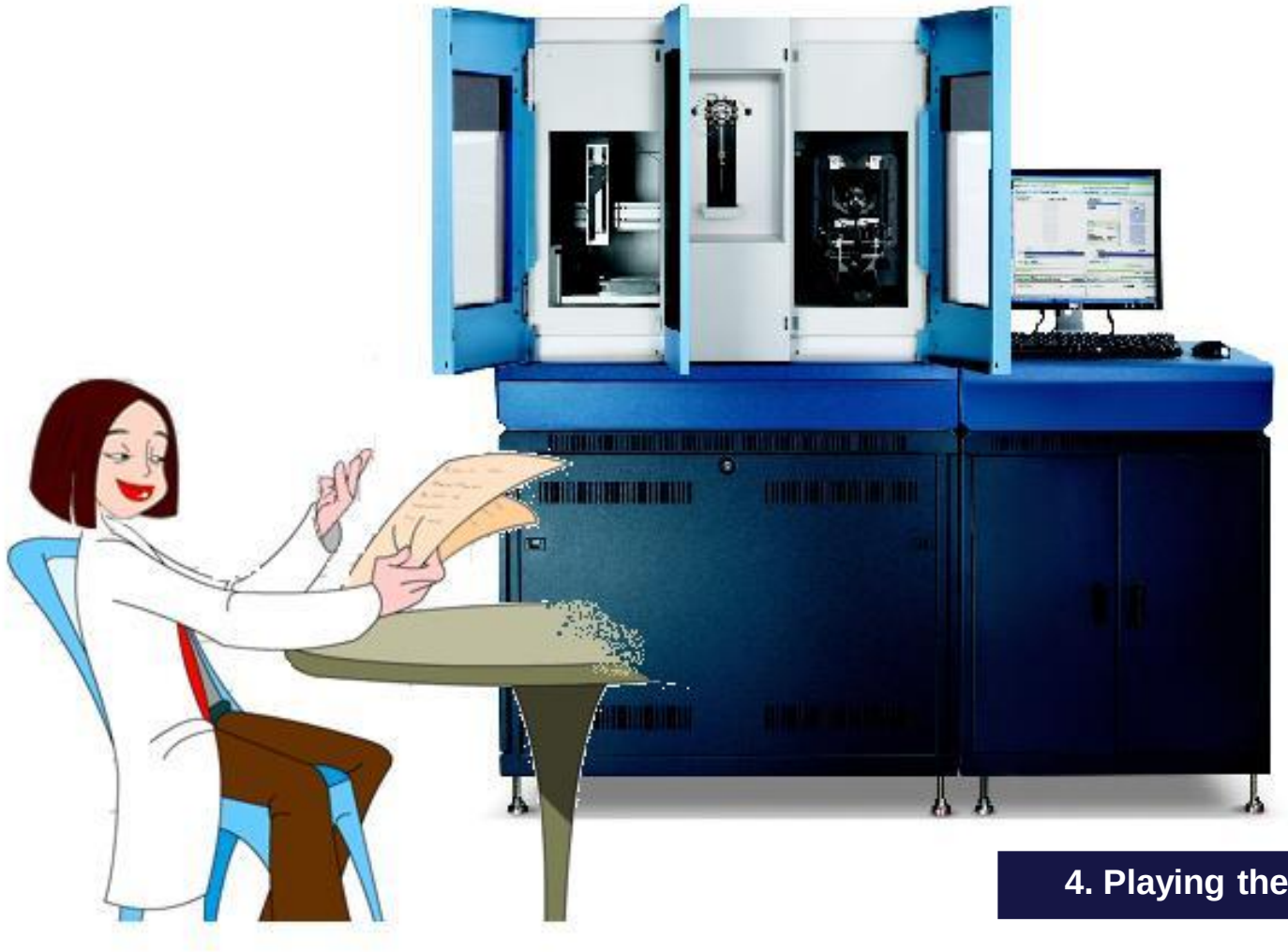
CCD camera takes pictures of 1 tile at a time

Playing the Game

	GS FLX Titanium	GA IIx Solexa	SOLiD 3 Plus
Time per run	10 hours	2 days (2x35) - 9.5 days (2x100)	8.5 days (2x35) – 13 days (2x50)
Cost per run	9,000 EURO	10,000 EURO (1x35) - 35,000 EURO (2x75)	45,000 EURO (1x50)
Cost per Mbp	< 18 EURO	< 1.8 EURO	< 1.8 EURO

Overview

1. Speaking the Language	2. Mastering the Lab	3. Understanding Data Files



4. Playing the Game!

Thanks!

VIB Micro	Array Facility
Torik Ayoubi	Ruth Maes
Joke Allemeersch	Kizi Coeck
Rudy Van Eijsden	Lieze Wolput