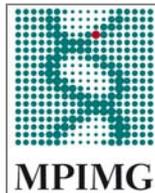


# Neue DNA-Sequenzierungstechnologien im Überblick

Dr. Bernd Timmermann



**Next Generation Sequencing Core Facility**  
Max Planck Institute for Molecular Genetics  
Berlin, Germany

# Max-Planck-Gesellschaft

## Max-Planck-Gesellschaft



80 Institute und Forschungseinrichtungen

21514 Mitarbeiter

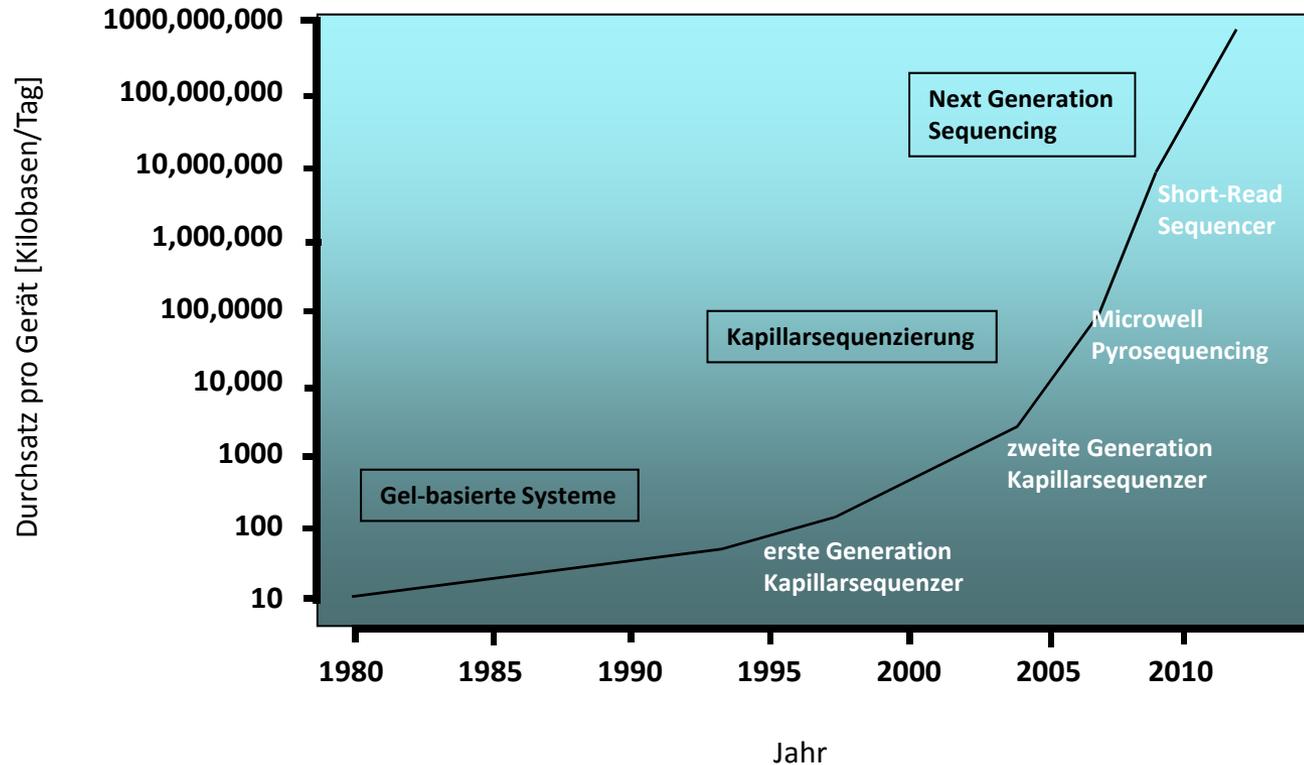
Budget 1,400 Mrd. Euro



**Max-Planck Institut  
für molekulare Genetik**



# Entwicklung des Sequenzierungsdurchsatzes



Modifiziert nach MR Stratton *et al. Nature* 458, 719-724 (2009)



# Entwicklung der Sequenzierungstechnologien

Humanes Genomprojekt

1000 Genomes Project



- *96 Sequenzen parallel*

- *3.2 Milliarden Sequenzen pro Gerätelauf*

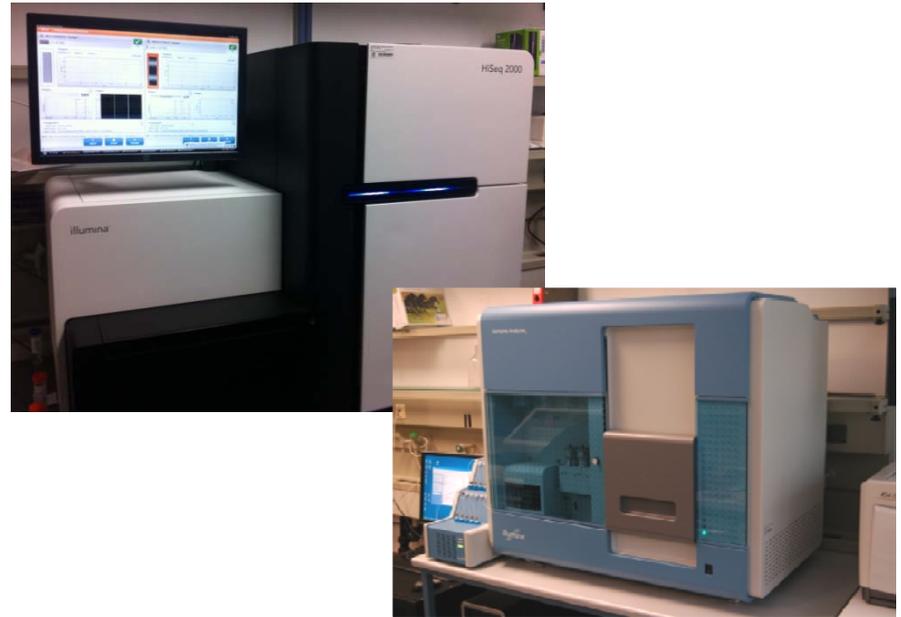


# Roche/454 Genome Sequencer FLX



- *de novo* Sequenzierung
- Metagenome Analysen
- Sequenzierung von PCR Produkten
- Analyse von Transkripten
- Sequenzierung von Zielregionen

# Illumina HiSeq 2000/GAIIx

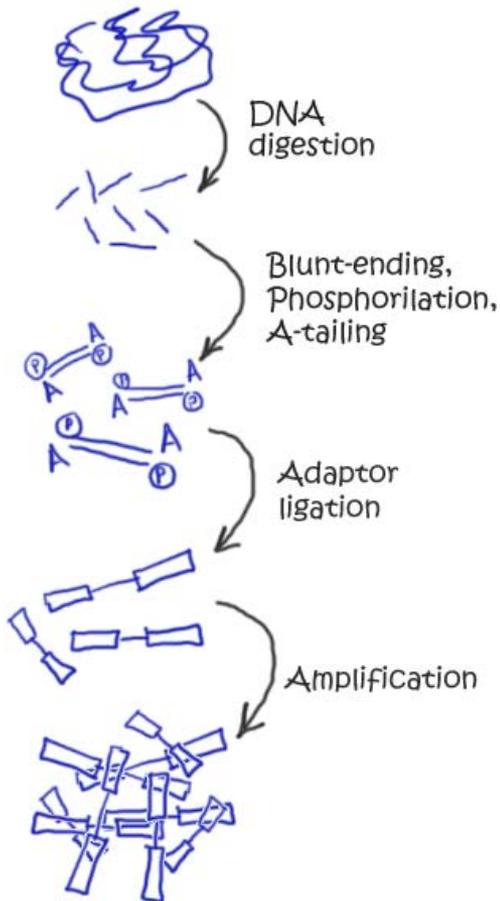


- Identifizierung von Transkriptionsfaktoren
- Methylierungsanalysen
- Identifizierung von *miRNAs*
- Expressionsanalysen
- Sequenzierung von Zielregionen
- Re-Sequenzierung von Genomen



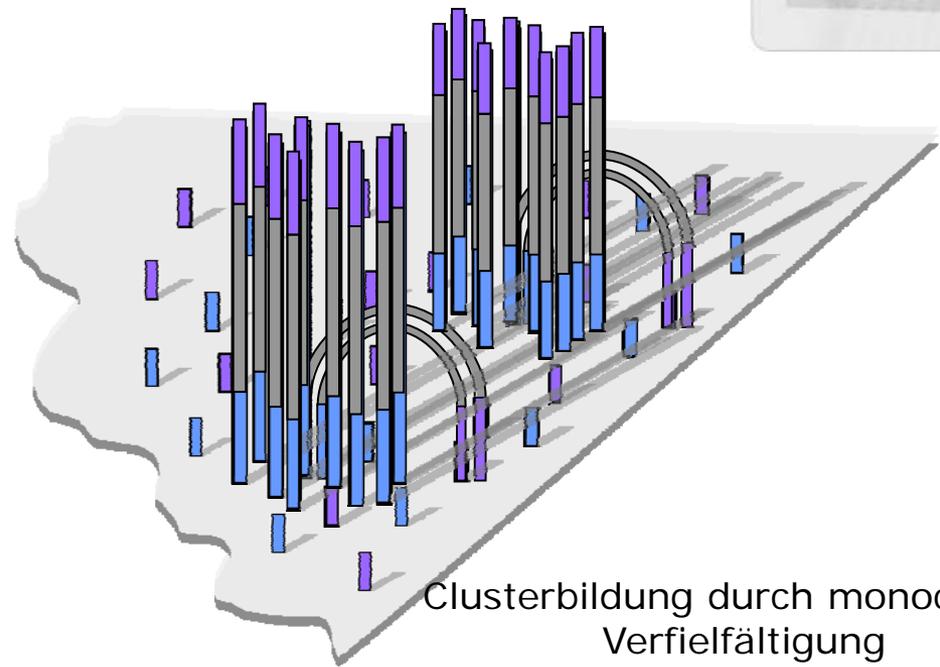
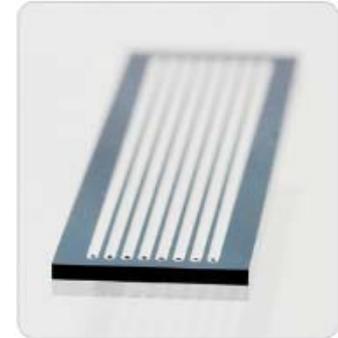
# Grundlagen Illumina Sequenzierung

## Probenvorbereitung



## Generierung von DNA Clustern

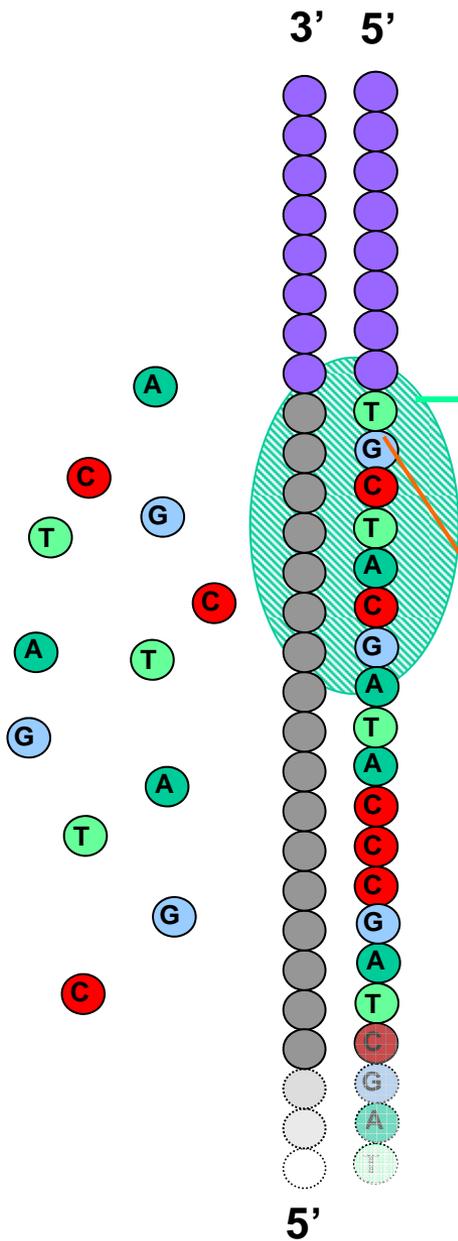
Bindung von Einzelmolekülen auf einer festen Oberfläche



Clusterbildung durch monoclonale Verfielfältigung



# Sequenzierung durch Synthese



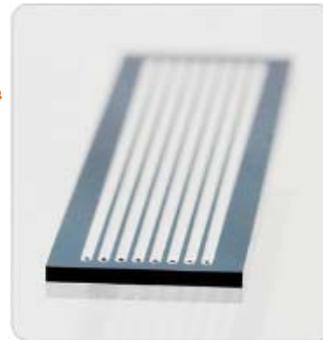
Cycle 1: Zuführung der Nukleotide und Polymerase

Einbau der ersten Base

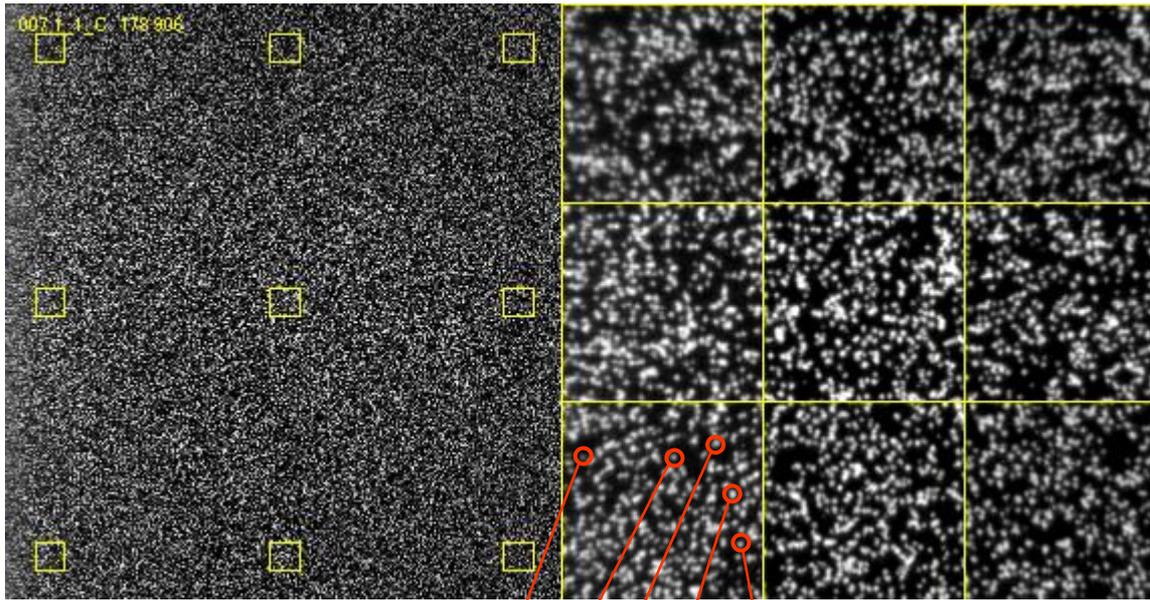
Entfernen der nicht eingebauten Daten

Signaldetektion

Cycle 2-n: Erneute Zufuhr von Nukleotiden und Enzymen



# Umwandlung der Bilddaten in die DNA Sequenz



Sequence Reads

TCGGGAGTCTAATGAGCCCGTAATCCCGTTAGTA  
TGAAGTCGGGAGTCTAATGAGCCCGTAATCCCGTT  
CGAATGAAGTCGGGAGTCTAATGAGCCCGTAATCC  
GAGCGAATGAAGTCGGGAGTCTAATGAGCCCGTAA  
CGAGCGAATGAAGTCGGGAGTCTAATGAGCCCGTA

Referenzsequenz

...CGAGCGAATGAAGTCGGGAGTCGTAATGAGCCCGTAATCCCGTTAGTA...



# Fakten Illumina Sequenzierung (HiSeq 2000)

Ausgangsmaterial: ~ 1-3 µg DNA

Library Preparation: ~ 1 Tag

Cluster Generation: ~ 1 Tag

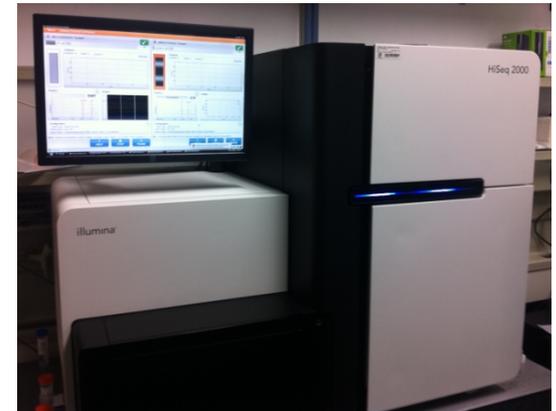
Laufzeit/ Single read ~ 2 Tage (36 b)

Leselänge: Paired End ~ 10 Tage (2 x 100 b)

Datenprozessierung: ~ 1 Tag

Datenmenge: Paired End ~ 500 Gigabasen

Anzahl der Sequenzen: bis zu 3200 Mio



# Fakten 454 Sequenzierung

Ausgangsmaterial: ~ 0.5 µg DNA

Probenvorbereitung: ~ 4 Stunden

Emulsions PCR: ~ 1 Tag

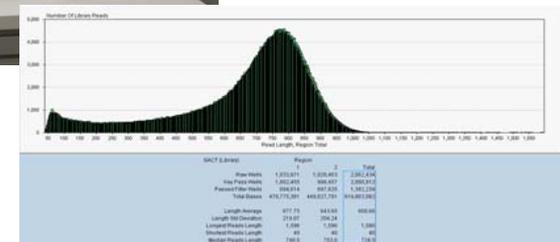
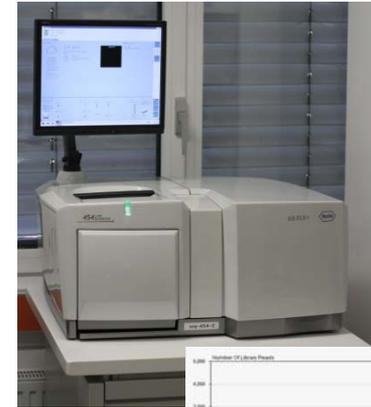
Gerätelaufzeit: 20 Stunden

Datenprozessierung: ~ 10 Stunden

Datenmenge: 700 - 1000 Megabasen

Anzahl der Sequenzen: 1.000.000 - 1.600.000

Leselänge: 700 - 800 Basen

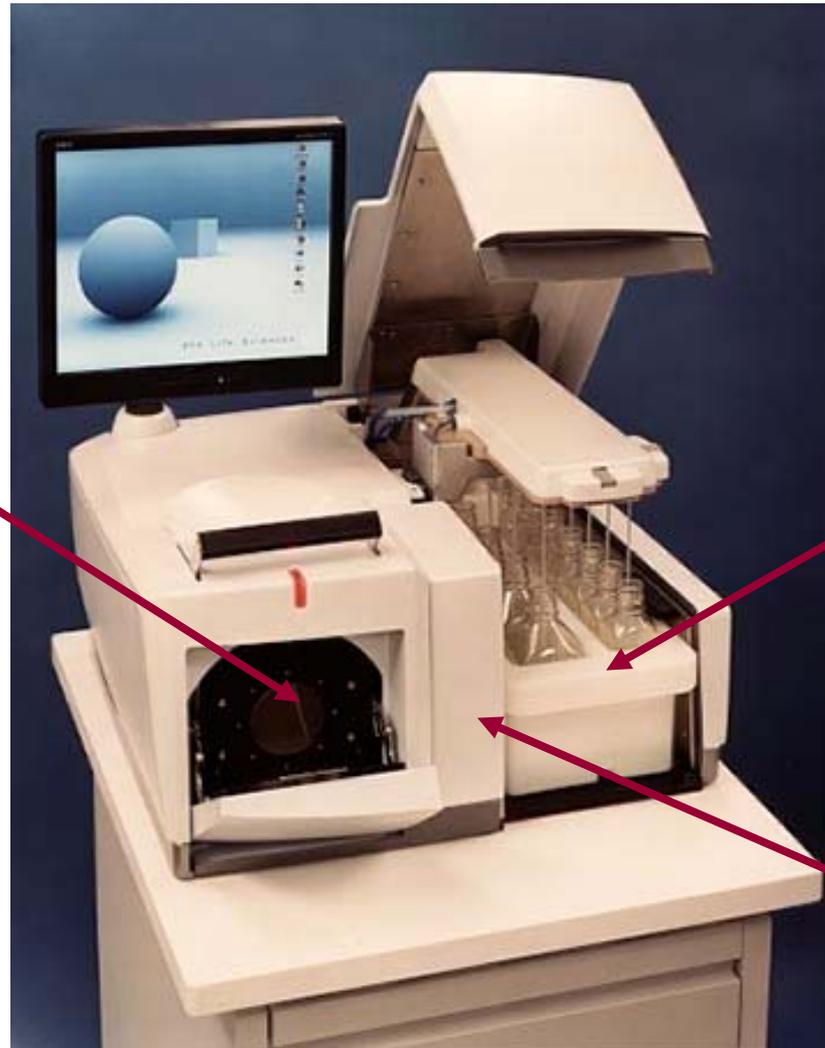


# 454 Sequencing Instrument

2. Load PicoTiter plate into instrument



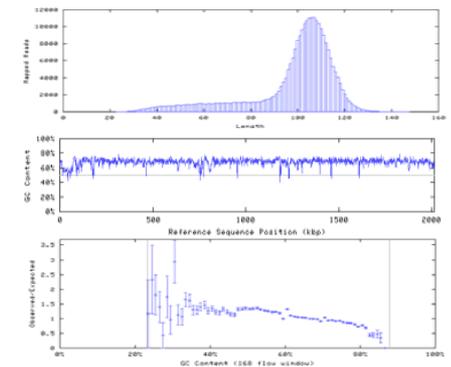
1. Genome is loaded into a PicoTiter™ plate



3. Load Reagents in a single rack



4. Sequencing



# 1000 Genomes

A Deep Catalog of Human Genetic Variation

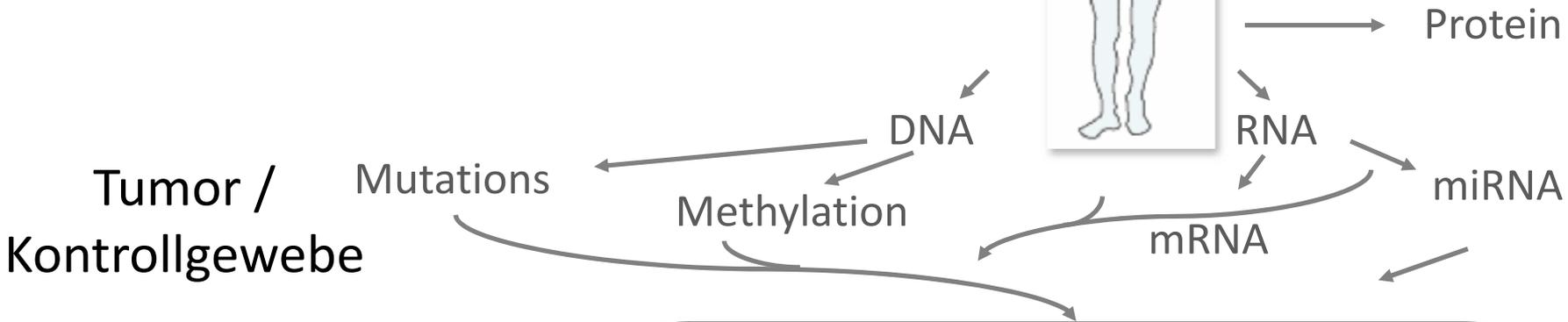
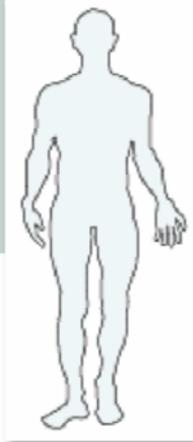
## Ziele

- Erstellung einer öffentlichen Datenbank mit allen SNVs und CNVs mit einer Allelfrequenz von  $>1\%$  in verschiedenen Populationen
- Entwicklung und Evaluierung von Methoden:
  - Generierung von Daten durch verschiedene Next-Generation Sequencing Plattformen
  - Austausch und Kombination von Daten und analytischen Methoden
  - Identifizierung der genetischen Variation aus Nextgen Daten

*(Nature 2010, Science 2010 and Nature 2011)*

[www.1000genomes.org](http://www.1000genomes.org)

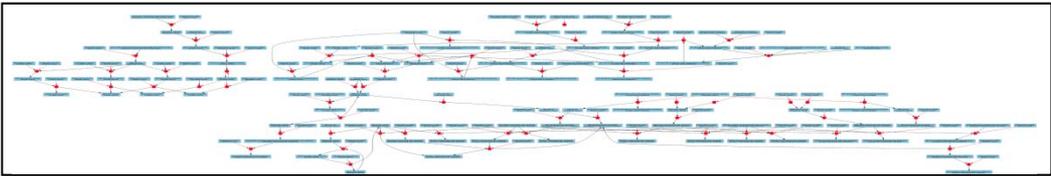




Sequenzierung



Bioinformatik

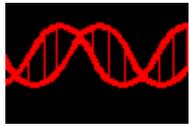


[www.oncotrack.eu](http://www.oncotrack.eu)



# Anreicherung von Zielregionen

## DNA Preparation

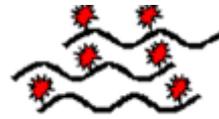


genomic  
DNA

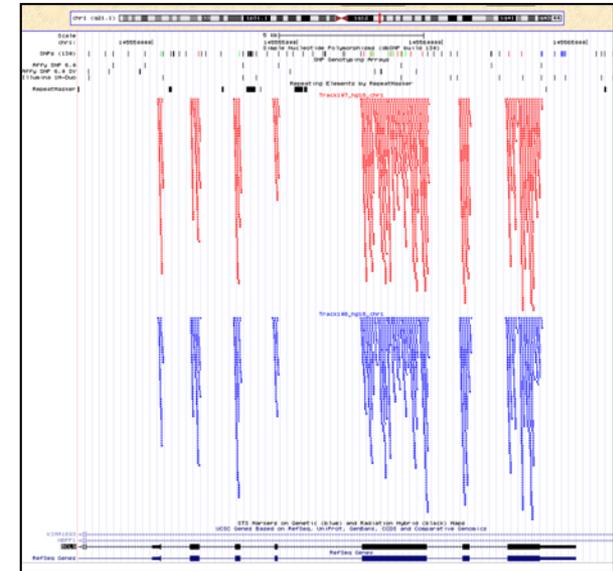


## Anreicherung

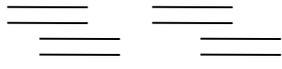
Hybridization



## Sequenzierung



Fragments (200-500bp)



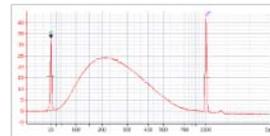
Ligation of adapters



Selection with  
streptavidin beads



Amplification and  
Quantification



# Sequenzierung in der Diagnostik

## IRON Study

### Interlaboratory **Robustness** of **NGS**

#### 'Ring trial like' Setup

<b>Goal</b>	<b>Lab to lab comparison</b> of performance of prototype primer set and sequencing protocol in ultra deep sequencing workflow using the 454 GS FLX System
<b>Purpose</b>	<b>Technical study</b> to compare ultra deep 454 sequencing within 10 different centers. Evaluate the interlaboratory performance, precision and reliability of an early-stage research use of the 454 ultra-deep amplicon sequencing assay system
<b>Timeframe</b>	July to Mid August 2010
<b>Investigators</b>	<b>10 'experienced' 454 users</b>
<b>Samples</b>	<b>18 blinded reference samples of hematological malignancies</b> (DNA) selected and send to all sites by MLL
<b>Primers</b>	31 primer sets labeled wit 3 MIDs for amplicon generation of TET2, CBL and KRAS (MTPs with total dried primers)
<b>Software</b>	<b>Toolbox</b> analysis provided by Prof. M. Dugas, Münster

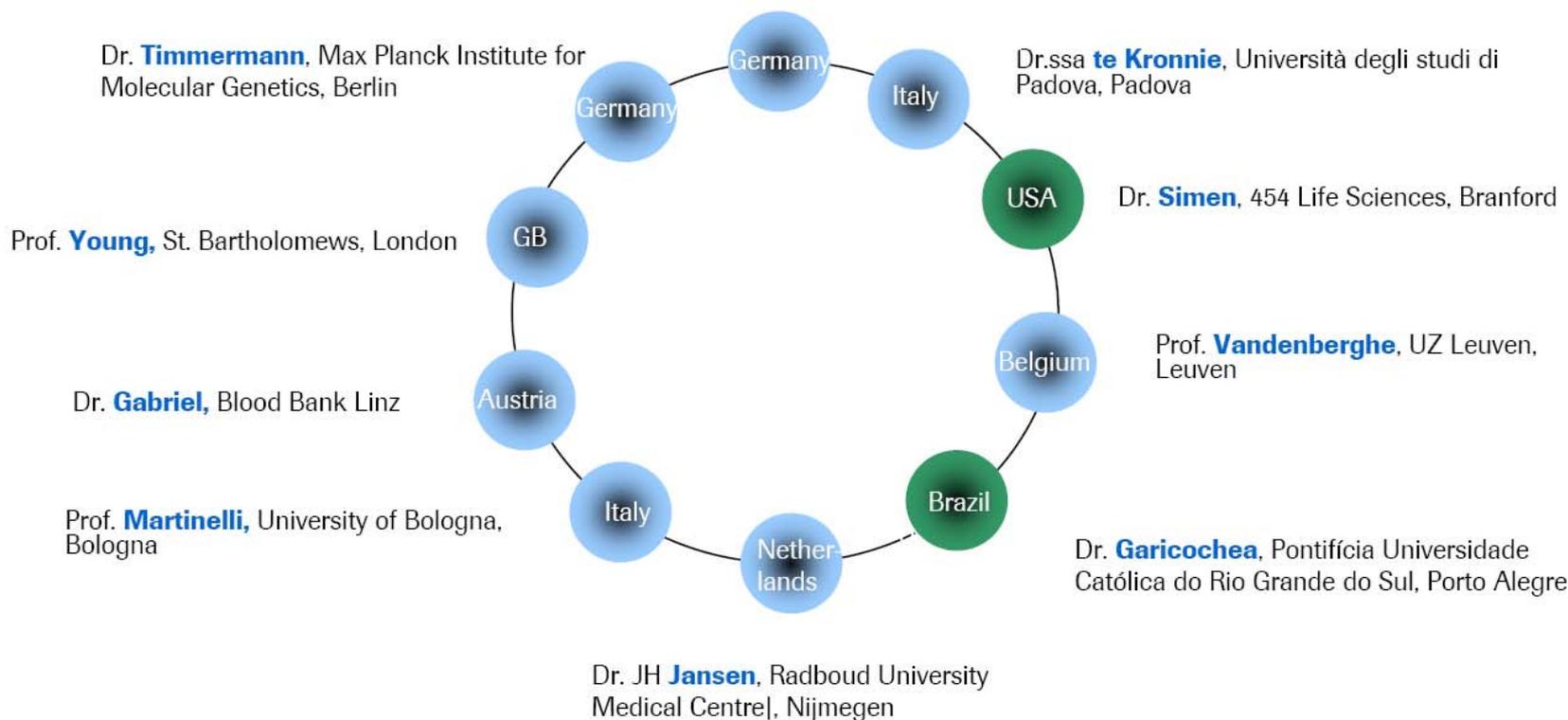


# Sequenzierung in der Diagnostik

## IRON Study

### Interlaboratory **R**obustness of **N**GS

Prof. **Haferlach**, Münchner Leukämie Labor, München



# Sequenzierung in der Diagnostik

## IRON Studie Ergebnisse

- Insgesamt wurden 92 Varianten identifiziert
- Im Vergleich zu den Vergleichsdaten aus der Sanger Sequenzierung, konnten mit der verwendeten 454 Technologien alle bekannten Varianten von allen Teilnehmern der Ringstudie sicher bestimmt werden
- In dieser Multicenterstudie konnte gezeigt werden, daß Amplicon-basiertes NGS in der Diagnostik technisch machbar ist, eine hohe Concordance zwischen den verschiedenen Laboren erreichbar ist und so auch für die Charakterisierung von hämatologischen Malignomen geeignet ist.

Kohlmann *et al.* (2011), *Leukemia*



# Bestimmung von Varianten

sun-awt-X11-XFramePeer

Project: 1F5map Parameters incomplete

Overview	Project	Parameters	Result files	HCDiffs	Alignment results	Flowgrams												
Reference Accession ▲ Number	Start Position in Ref	End Position in Ref	Reference Bases	Variation Bases	Total Percent	Total Depth	Reference Amino Acids	Variation Amino Acids	Coding Frame	Region Name	Known SNP Info	Percent Forward With Variation	Percent Reverse With Variation	Num Forward With Variation	Num Reverse With Variation	Total Num Forward Reads	Total Num Reverse Reads	
chr1	873488	873488	A	G	75.0	4			-	NOC2L	rs4970378	100.0	66.7	1	2	1	3	
chr1	877423	877423	A	C	100.0	5			-	NOC2L	rs3748595	100.0	100.0	3	2	3	2	
chr1	972325	972325	T	C	100.0	3			+	AGRN	rs3128100	100.0	100.0	1	2	1	2	
chr1	1168334	1168334	G	A	100.0	3	I	I	-2	FAM132A		100.0	100.0	1	2	1	2	
chr1	1221450	1221450	G	A	50.0	6			-	ACAP3		50.0	50.0	1	2	2	4	
chr1	1244704	1244704	C	G	100.0	4	G	G	-3	CPSF3L	rs10907179	100.0	100.0	2	2	2	2	
chr1	1836625	1836625	G	A	100.0	3			+	CALML6	rs2247560	100.0	100.0	2	1	2	1	
chr1	1837969	1837969	G	C	100.0	3			+	CALML6	rs2803296	100.0	100.0	2	1	2	1	
chr1	1839389	1839389	A	G	100.0	5	M	T	-2	TMEM52	rs28640257,r...	100.0	100.0	3	2	3	2	
chr1	1950786	1950786	T	C	100.0	3			+	GABRD	rs28581504	100.0	100.0	1	2	1	2	
chr1	2224763	2224763	C	T	100.0	3			+	SKI	rs2256178	100.0	100.0	1	2	1	2	
chr1	2227108	2227108	A	G	100.0	3			+	SKI	rs11580302	100.0	100.0	1	2	1	2	
chr1	2442429	2442429	T	C	83.3	6	K	K	-1	PANK4	rs57877814,r...	100.0	66.7	3	2	3	3	
chr1	2479633	2479633	C	T	100.0	3			-	TNFRSF14	rs8725	100.0	100.0	2	1	2	1	
chr1	2483213	2483213	G	A	100.0	3			-	TNFRSF14	rs2234161	100.0	100.0	2	1	2	1	
chr1	2513072	2513072	G	C	100.0	6			-	MMEL1	rs876938	100.0	100.0	5	1	5	1	
chr1	2525618	2525618	A	G	100.0	3			-	MMEL1	rs4648652	100.0	100.0	1	2	1	2	
chr1	2532513	2532513	A	G	100.0	5			-	MMEL1	rs10910111	100.0	100.0	3	2	3	2	
chr1	3335622	3335622	A	G	100.0	3			+	PRDM16		100.0	100.0	2	1	2	1	
chr1	3532046	3532046	G	A	100.0	3			+	TPRG1L		100.0	100.0	2	1	2	1	
chr1	3635704	3635704	G	A	100.0	3			i	TP73	rs2096224	100.0	100.0	1	2	1	2	
chr1	3635997	3635997	C	T	100.0	3			+	TP73	rs11589885	100.0	100.0	1	2	1	2	
chr1	3637279	3637279	G	A	80.0	5			+	TP73	rs41301973	75.0	100.0	3	1	4	1	
chr1	3639263	3639263	G	A	71.4	7	A	A	+1	TP73	rs61735051	66.7	75.0	2	3	3	4	
chr1	3639422	3639422	C	A	88.9	9	A	A	+1	TP73	rs9662633	100.0	80.0	4	4	4	5	
chr1	3732969	3732969	A	G	100.0	4			-	KIAA0562	rs2275834	100.0	100.0	1	3	1	3	
chr1	3732992	3732992	T	C	100.0	3			-	KIAA0562	rs4648344	100.0	100.0	1	2	1	2	
chr1	3735583	3735583	-	ACAA	42.9	7			-	KIAA0562		50.0	33.3	2	1	4	3	
chr1	3758579	3758579	T	C	100.0	3			-	KIAA0562	rs1539647	100.0	100.0	1	2	1	2	
chr1	3758589	3758589	T	C	100.0	3			-	KIAA0562	rs1539648	100.0	100.0	1	2	1	2	
chr1	3765314	3765314	C	G	100.0	3			+	DFFB	rs4376673	100.0	100.0	2	1	2	1	
chr1	4118916	4118916	A	G	100.0	3			-		rs3890798	100.0	100.0	2	1	2	1	
chr1	4734466	4734466	T	C	100.0	6			+	AJAP1	rs389959	100.0	100.0	3	3	3	3	
chr1	6081149	6081149	A	G	60.0	5	S	S	+1	KCNAB2	rs2229002	50.0	66.7	1	2	2	3	
chr1	6093011	6093012	CG	-	100.0	3			-	CHD5		100.0	100.0	2	1	2	1	
chr1	6126853	6126853	A	G	100.0	4			-	CHD5	rs9435102	100.0	100.0	2	2	2	2	
chr1	6516119	6516119	G	A	100.0	3			-	NOL9	rs11122083	100.0	100.0	1	2	1	2	
chr1	6532438	6532438	A	G	100.0	3			-	NOL9	rs10864622	100.0	100.0	1	2	1	2	
chr1	6559758	6559758	G	A	100.0	3			+	TAS1R1	rs12040340	100.0	100.0	1	2	1	2	
chr1	6612727	6612727	A	G	100.0	4			+	THAP3	rs742394	100.0	100.0	3	1	3	1	
chr1	6635701	6635701	C	T	100.0	4			-	DNAJC11	rs2235564	100.0	100.0	2	2	2	2	
chr1	7660386	7660386	A	G	100.0	4			+	CAMTA1	rs57934614,r...	100.0	100.0	2	2	2	2	
chr1	7719228	7719228	A	G	100.0	4			+	CAMTA1	rs2071985	100.0	100.0	3	1	3	1	
chr1	7785486	7785486	T	C	100.0	3			+	PER3	rs228641	100.0	100.0	2	1	2	1	
chr1	7792635	7792635	T	C	100.0	4	S	S	+1	PER3	rs228669	100.0	100.0	2	2	2	2	

Info 1F5map: Opened Mapper project / amd/schnurziepegal/0/scratch/454/1F5map

Exit

New

Open

Start

Stop

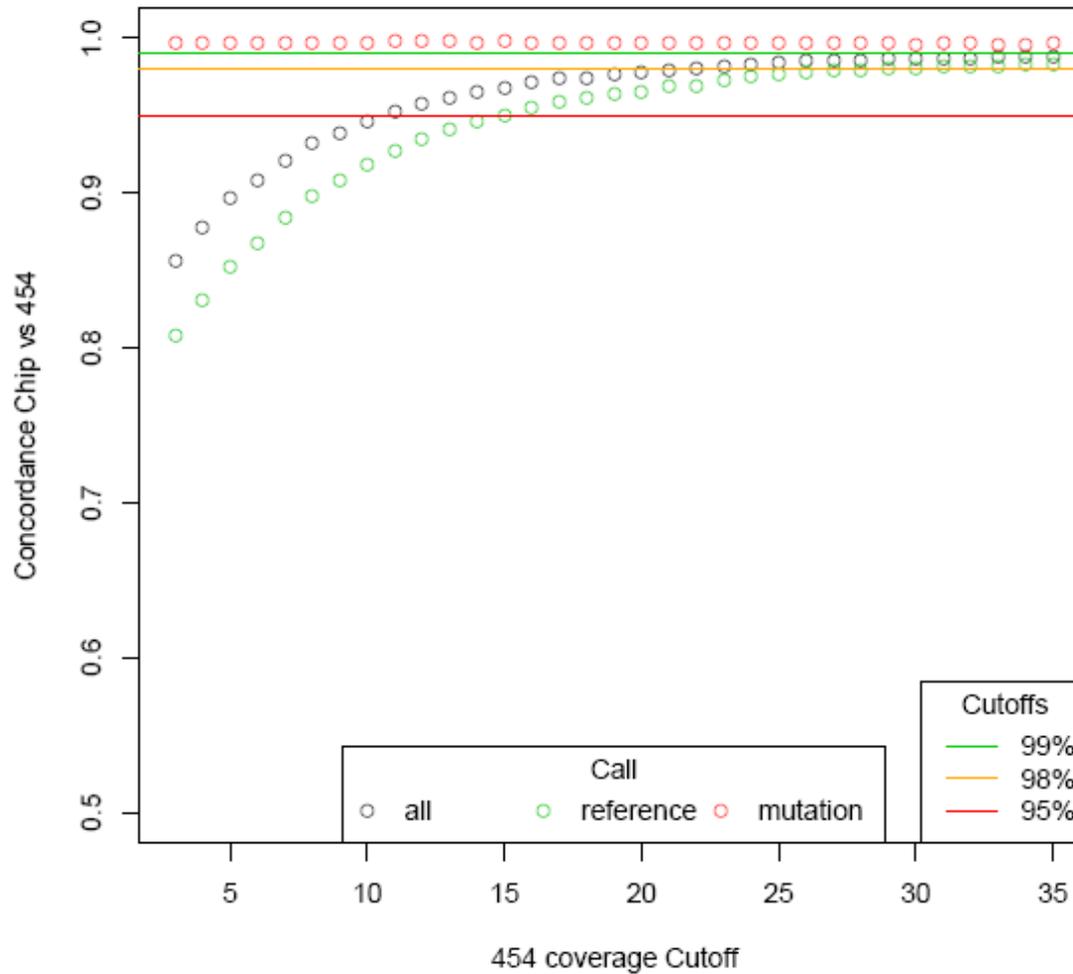
About

Help

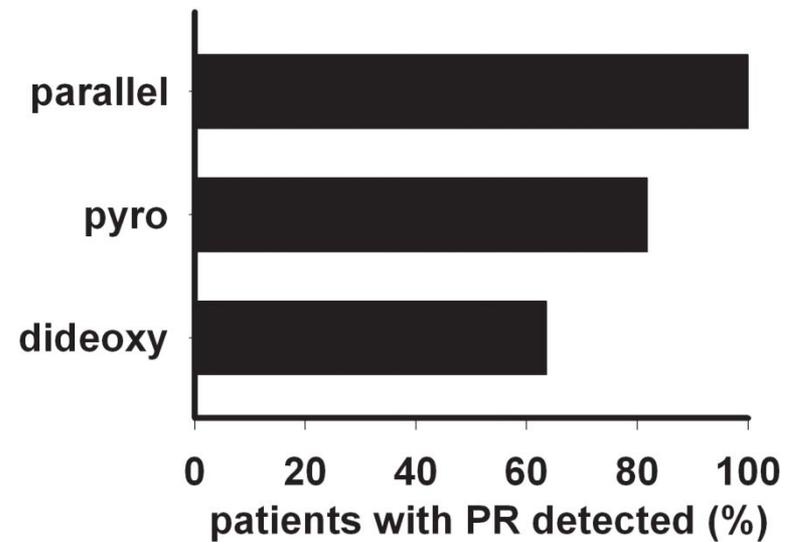
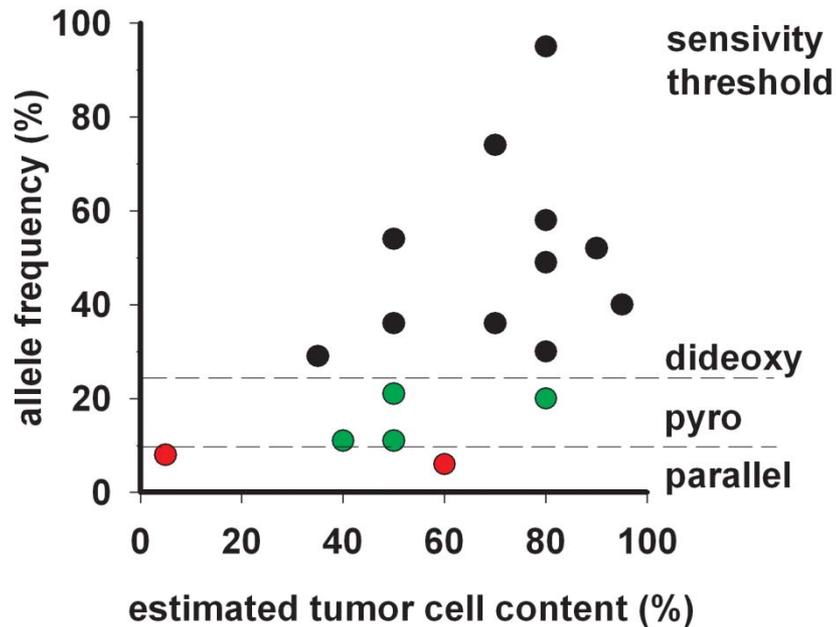


# Qualität der Mutationsdetektion

Concordance mit Affymetrix Array "genome-wide human SNP array 6.0"



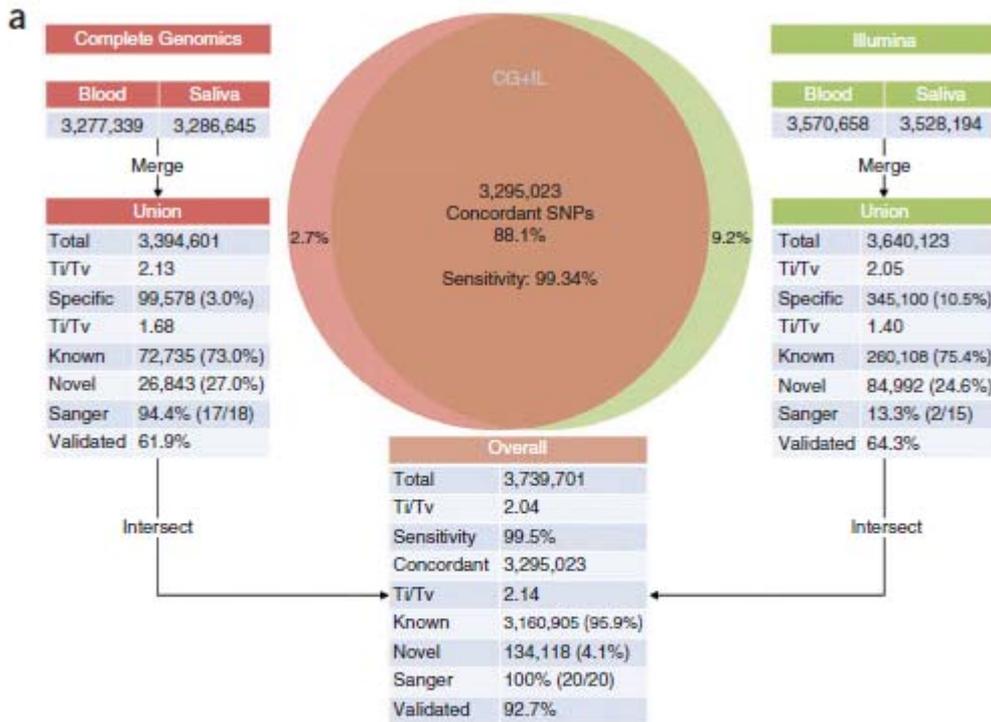
# Sensitivität der Detektion von Mutationen



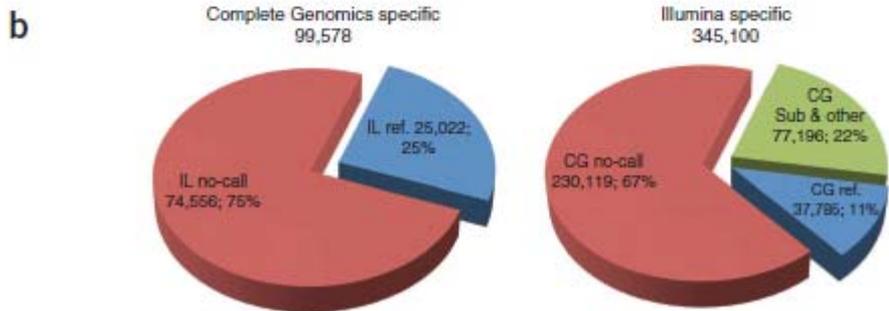
Querings *et al.* (2011), *PlosOne*



# Vergleichende Sequenzierung des menschlichen Genoms



- über 3,7 Millionen Unterschiede zum Referenzgenom
- Abweichungen in Abhängigkeit von der verwendeten Technologie



Lam *et al.* (2012), *Nat Biotechnol*



A large, multi-story building with a modern architectural style, featuring a prominent glass facade and a series of balconies with dark railings. The building is situated on a grassy area with a paved walkway in the foreground. A blue sign is visible in the lower-left corner of the image.

MAX-PLANCK-INSTITUT  
FÜR MOLEKULARE GENETIK  
IHNESTRASSE 63-73

WERKSTÄTTEN  
NACHTPFORTE

**Vielen Dank für Ihr  
Interesse!**