SVT – Tagung vom 29. März 2012, Universität Irchel, Zürich

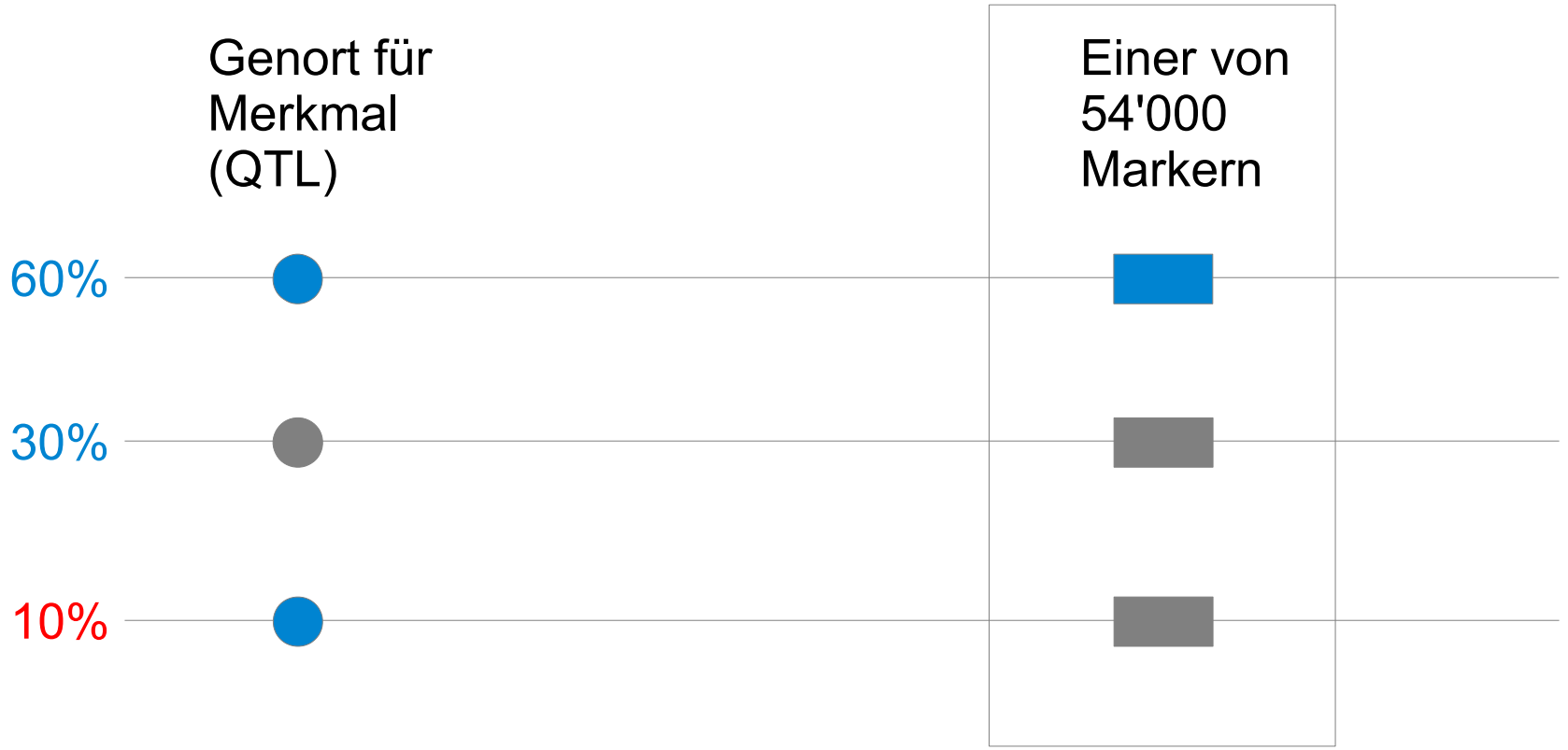
Zukünftige Entwicklungen und Anwendungen der genomischen Selektion (GS)

Ruedi Fries und Hubert Pausch
Lehrstuhl für Tierzucht
Technische Universität München

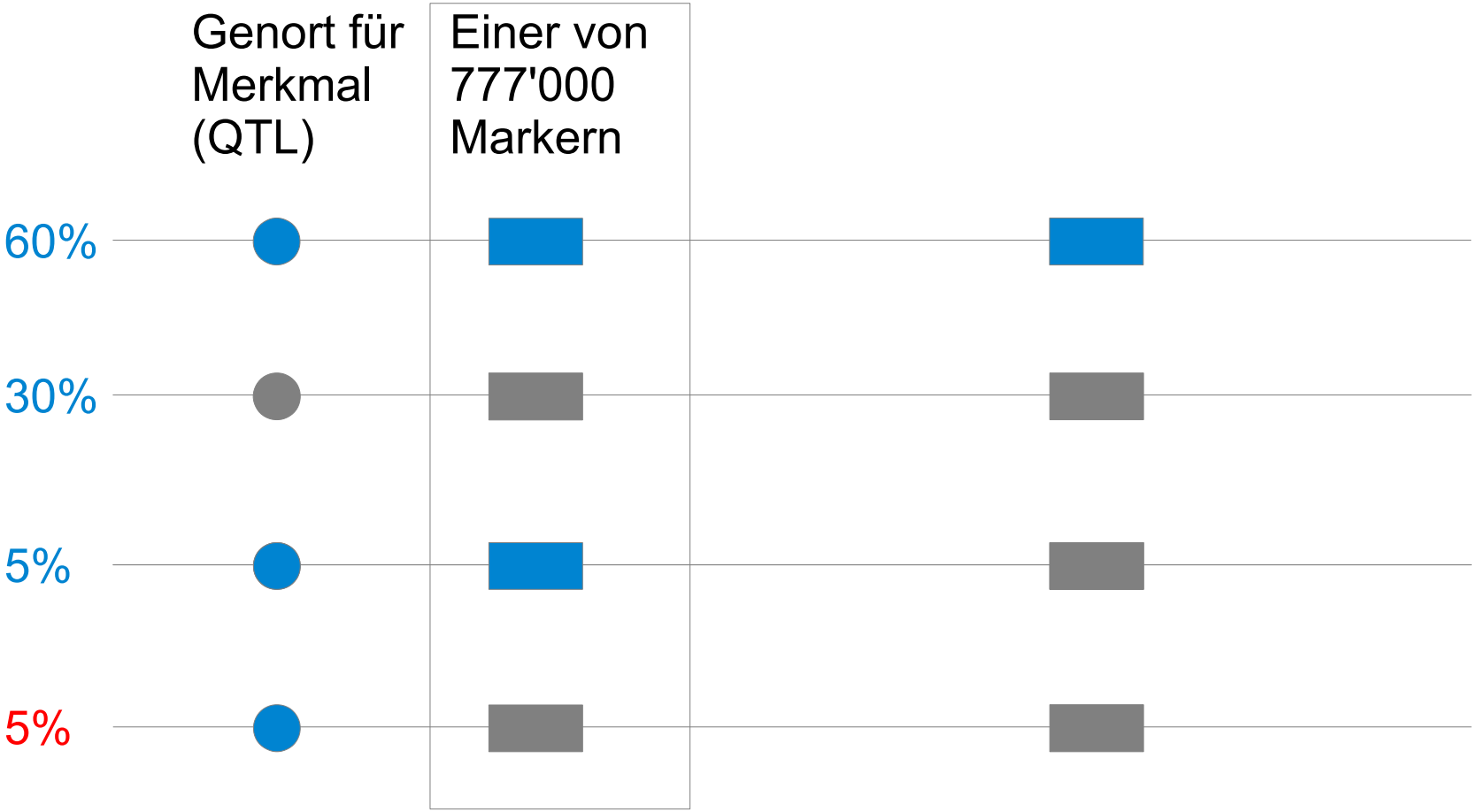
GS 1.0 beim Milchrind

- Die Vorhersage des QTL – Status basiert auf gekoppelten Markern.
  - Das Kopplungsungleichgewicht ist über Generationen und Populationen hinweg nicht stabil.
- Die QTL-Architektur ist nicht bekannt
  - Nicht-additive Effekte werden nicht berücksichtigt.
- Ist teuer und deshalb im Wesentlichen auf männliche Tiere beschränkt.

GS 1.0 – 54'000 Marker



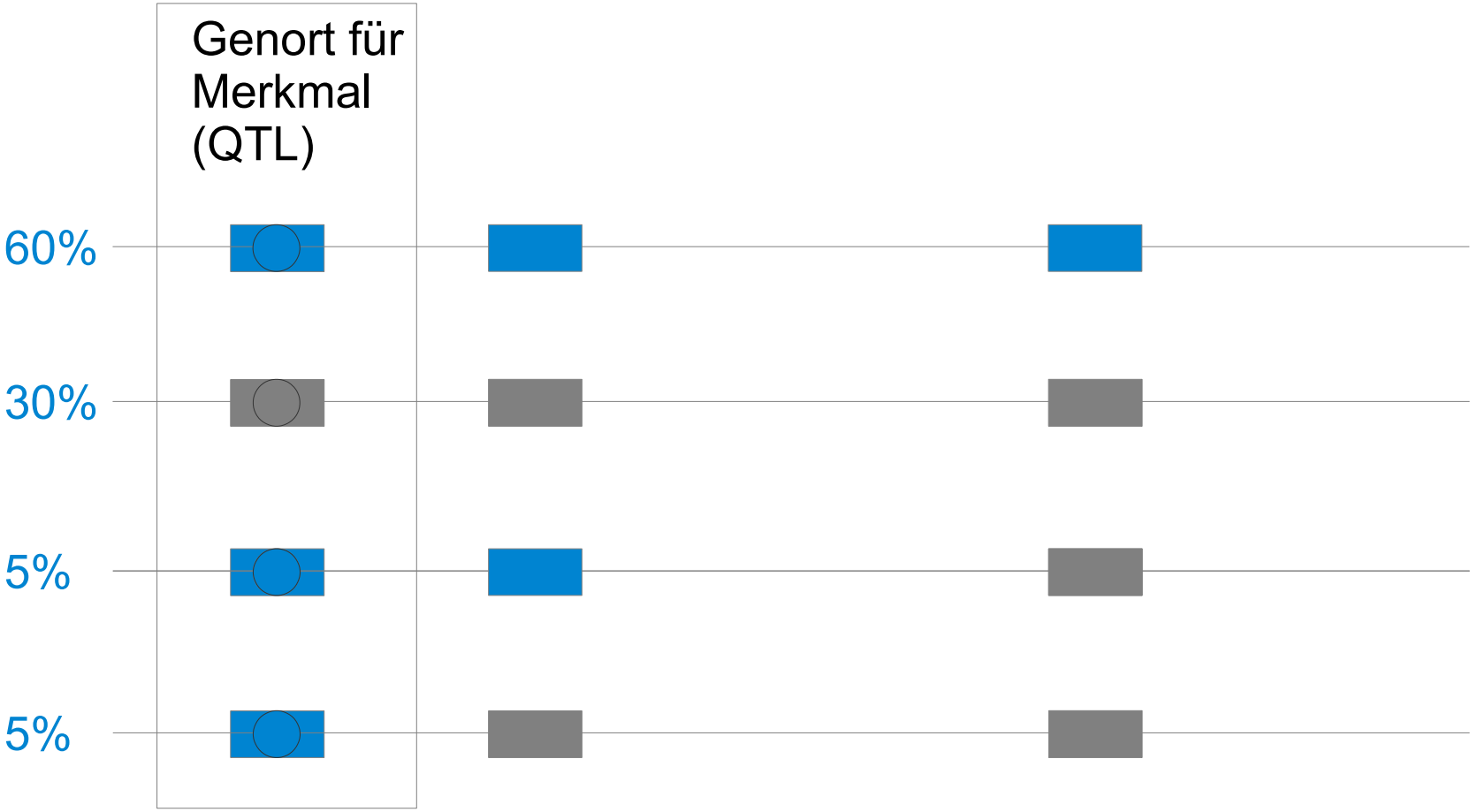
GS 1.5 – 777'000 Marker



GS 2.0 beim Milchrind

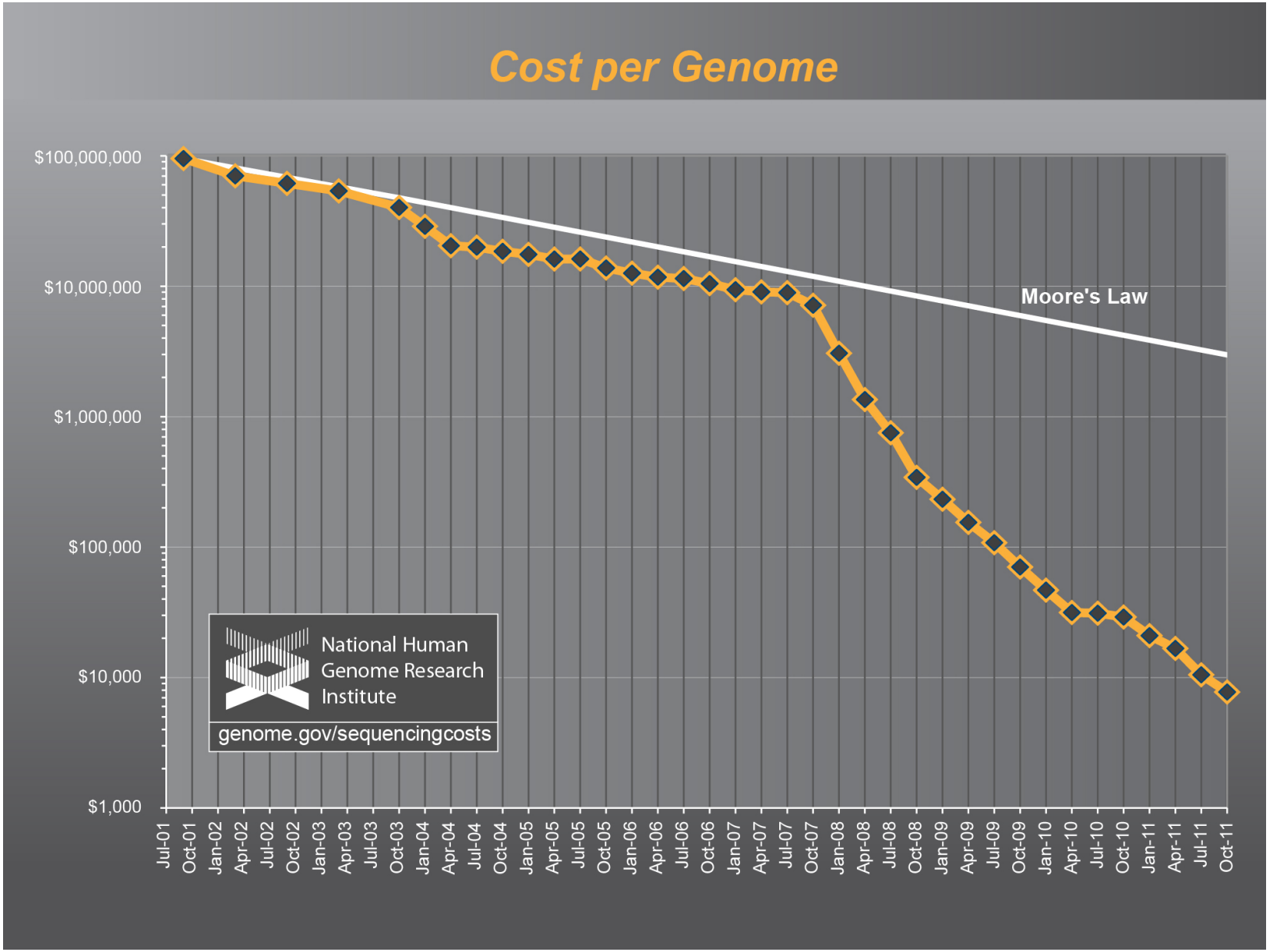
- Direkte Erfassung der QTL durch die Sequenzierung individueller Genome
- Berücksichtigung der QTL-Architektur
- “Flächendeckende”, kostengünstige genomische Evaluierung
- Vorhersage des genomischen Produktionswertes von Tieren

GS 2.0 – Sequenz



Sequenzierung individueller Genome

„More than Moore“



Hochdurchsatz-Sequenzierung: „State-of-the-art” Frühling 2012

	ABI	Roche	ABI*	Illumina
	ABI 3730	GSFLX	SOLiD 4	HiSeq 2000
Basen / Lauf	70 kb	400 Mb	300 Gb	600 Gb
Leselänge	750 bp	400 bp	75 bp	100 bp

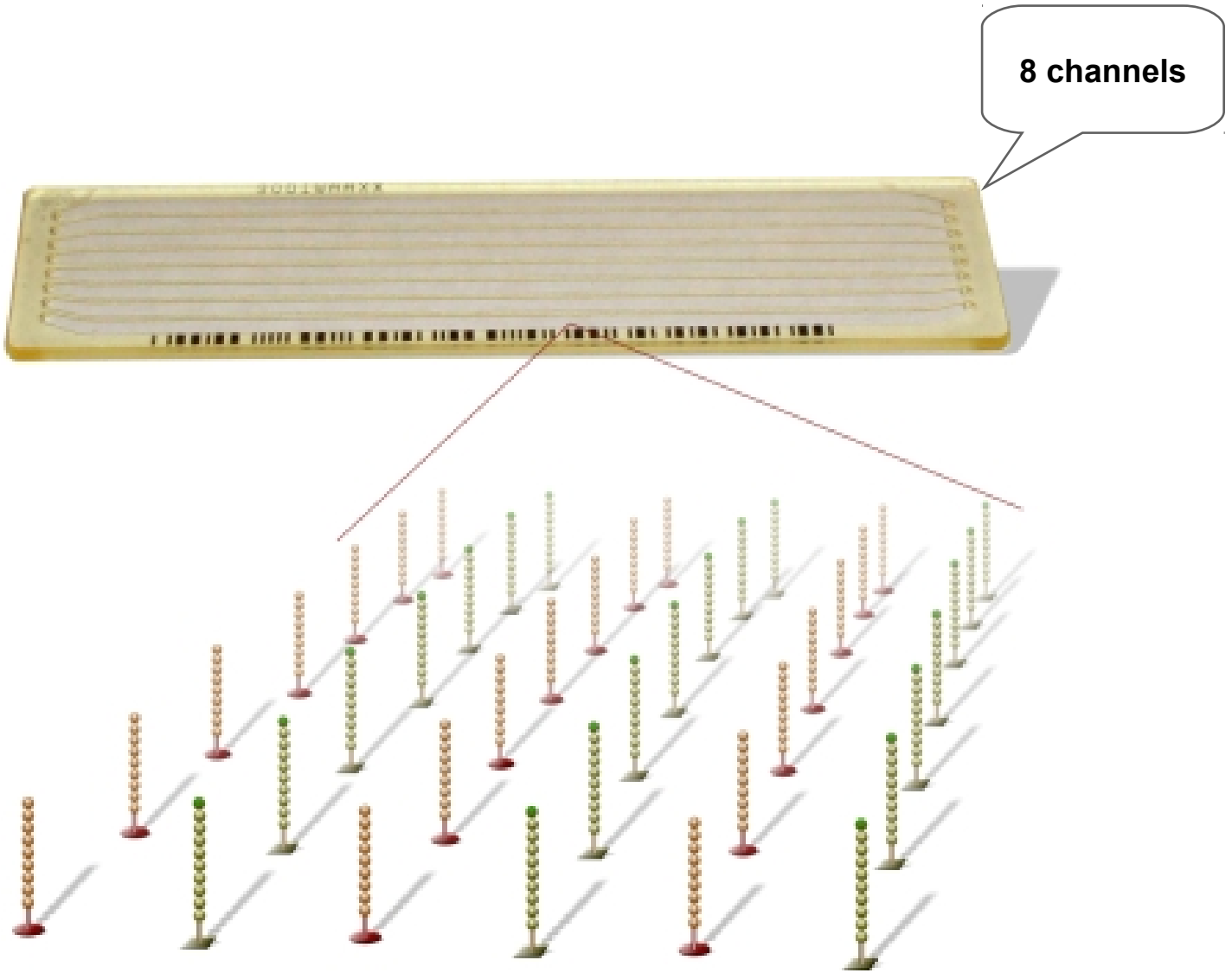
*Life Technologies

Herbst 2012?

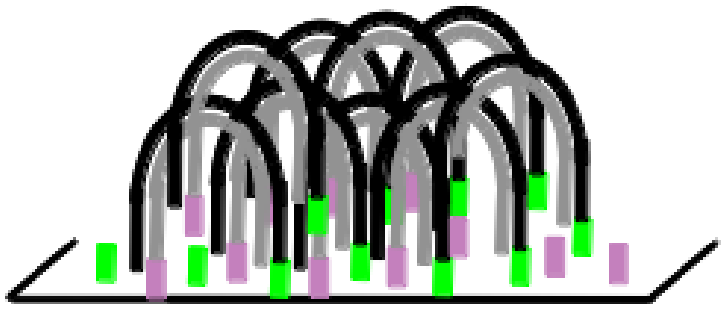
„Sequenzierautomat“ der 3. Generation (Oxford Nanopore)

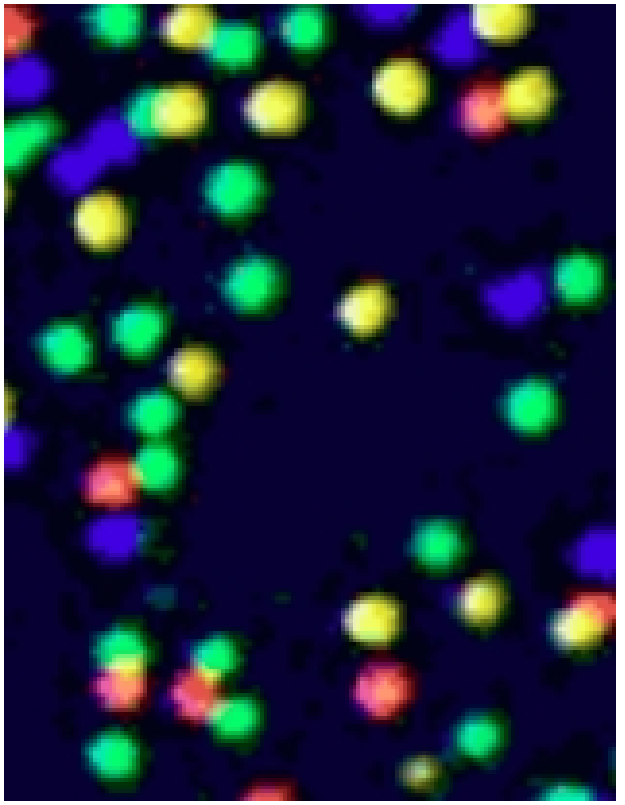
“BGI's sequencing power”

HiSeq 2000: Flow cell



2 x n* pictures / cluster





*n = Sequenzierlänge

HiSeq2000 output: Two *fastq* files per lane (20 – 50 Gigabytes per file)

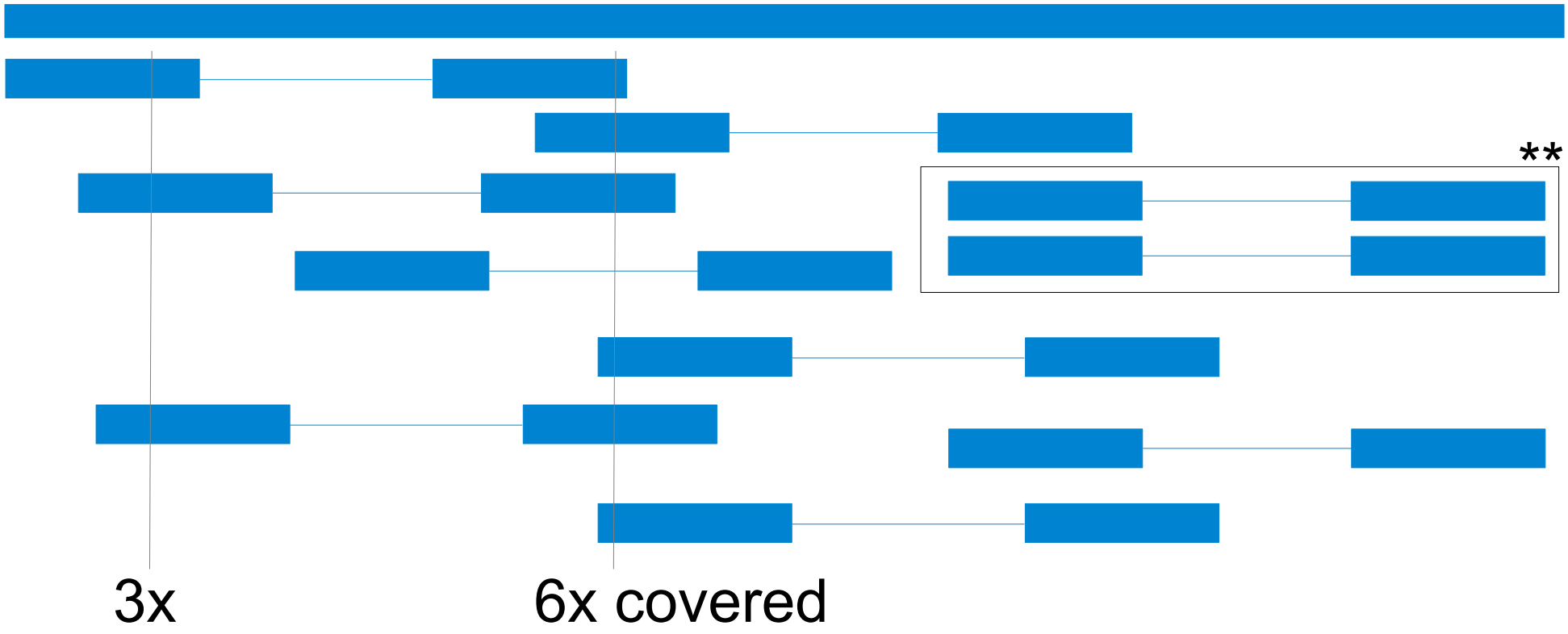
```
@HWI-ST506_0058:2:1101:6926:174667#0/1
CAGCTTCAGGTGTCCACCCGAGACCCTCTCCCCACCCTCTCCTGTGGGCTAAAGG
AACTACACTATCTGAAATCCTATCTCTGGAGTTTCTTGTCTGGCT
+
HHHHHHHHHHHHHHHCGGGGHGHHFHHHEHHHHCGHHHHFHHFHFHHHDHHFG
HHFHDHHHFFFHBHEHHFHEFHFHHHFHEHHEDCFGGEE1?DD<EA=

@HWI-ST506_0058:2:1104:17618:61439#0/1
GTTCGCGTGTGCAGCTTGCCTACCAGGTGGTGGGCAGCCACTGACAGGCCCTGT
GCTGACTCTGTCCTGCAGCCAGTGCCTCCCCCAGACAAGAGCCAGG
+
FFFFEFFEDEAB@@@EDEEEFFFEDFFFFFCECEEFFFFDBEEEECB>=EC5C7B<==
>=EE?D>ADDCDF@>F=413)807*44B=CCCBB37;09;76
```

Phred quality: 0 - 93, ASCII 33 - 126 (python: *ord* und *chr* functions)

Alignment with *bwa* → *bam**

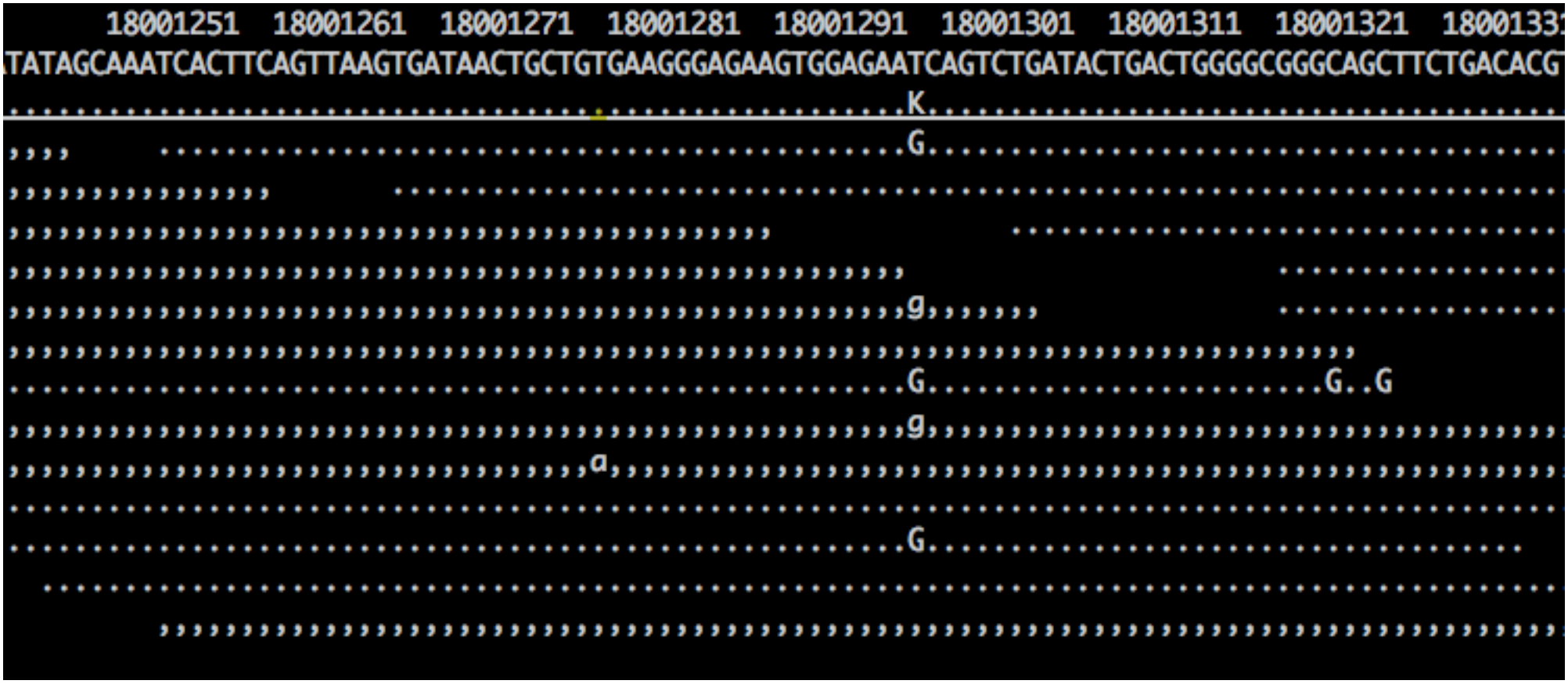
Reference sequence: UMD3.1, 2.66 billion bases



* Binary version of *sam* (sequence aligment / map) - format

** PCR duplicate (arising during library construction)

Visualization of *bam* file

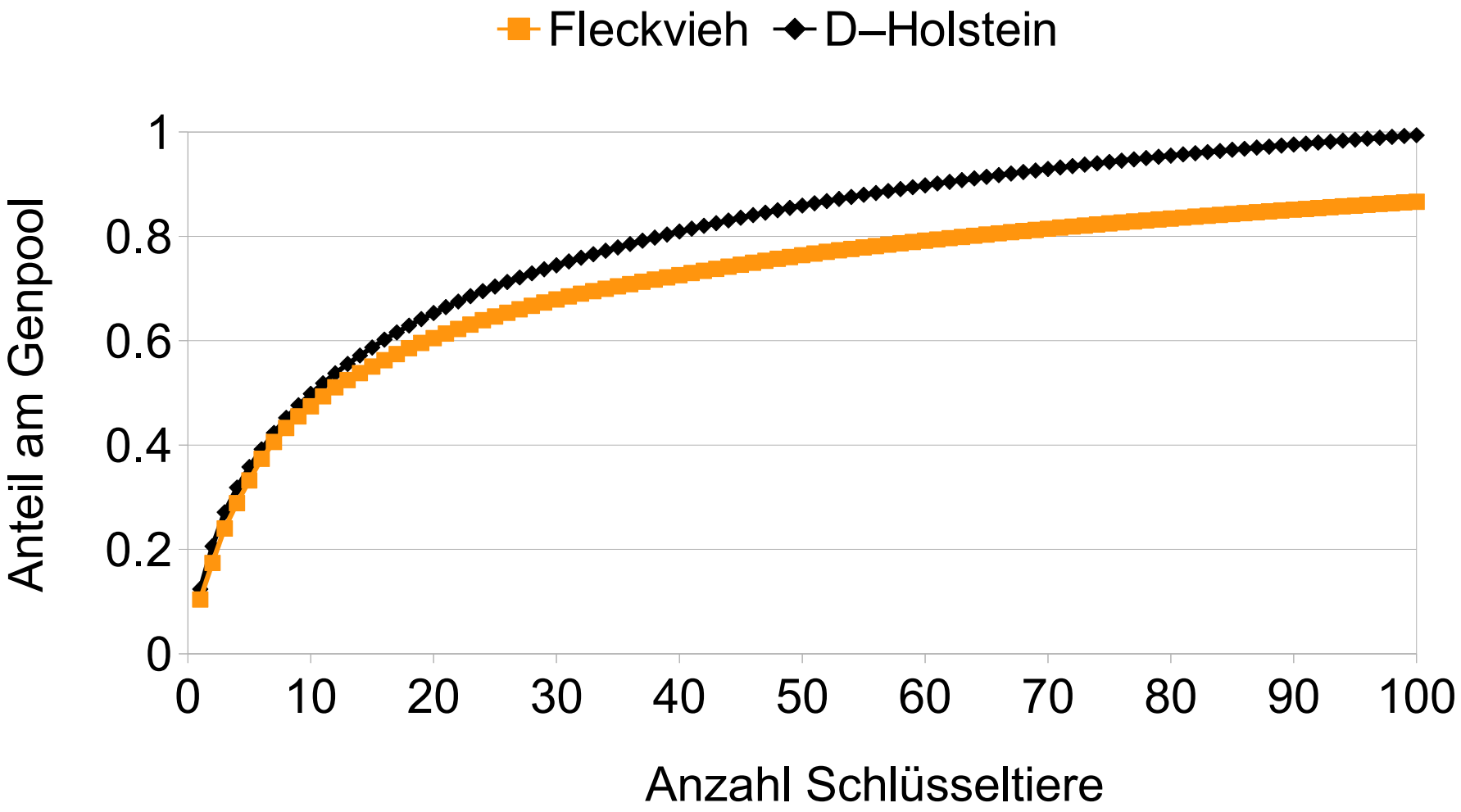


Sequenzierung von Schlüsseltieren

- Die Sequenzierung von Tausenden von Individuen einer Populationen ist (noch) nicht möglich.
- Sequenzierung von Schlüssel-Tieren!
- Populationsweite Imputation der Sequenz

Identifizierung von Schlüsseltieren

- Schlüsseltiere = Tiere, die möglichst viel zum aktuellen Genpool beitragen.
- Anzahl hängt von der effektiven Populationsgröße ab.
  - N_e Holstein ca. 80
  - N_e Fleckvieh ca. 140


Fleckvieh
D–Holstein
Anteil am Genpool
1
0.8
0.6
0.4
0.2
0
0
10
20
30
40
50
60
70
80
90
100
Anzahl Schlüsseltiere

Sequenzierung von 43 Schlüssel-Tieren der Fleckviehrasse

- Decken 68% des Genpools ab
- Durchschnittlich 7.4-fache Sequenzabdeckung mit Illumina Instrumenten
- 17.3 Millionen variable Positionen

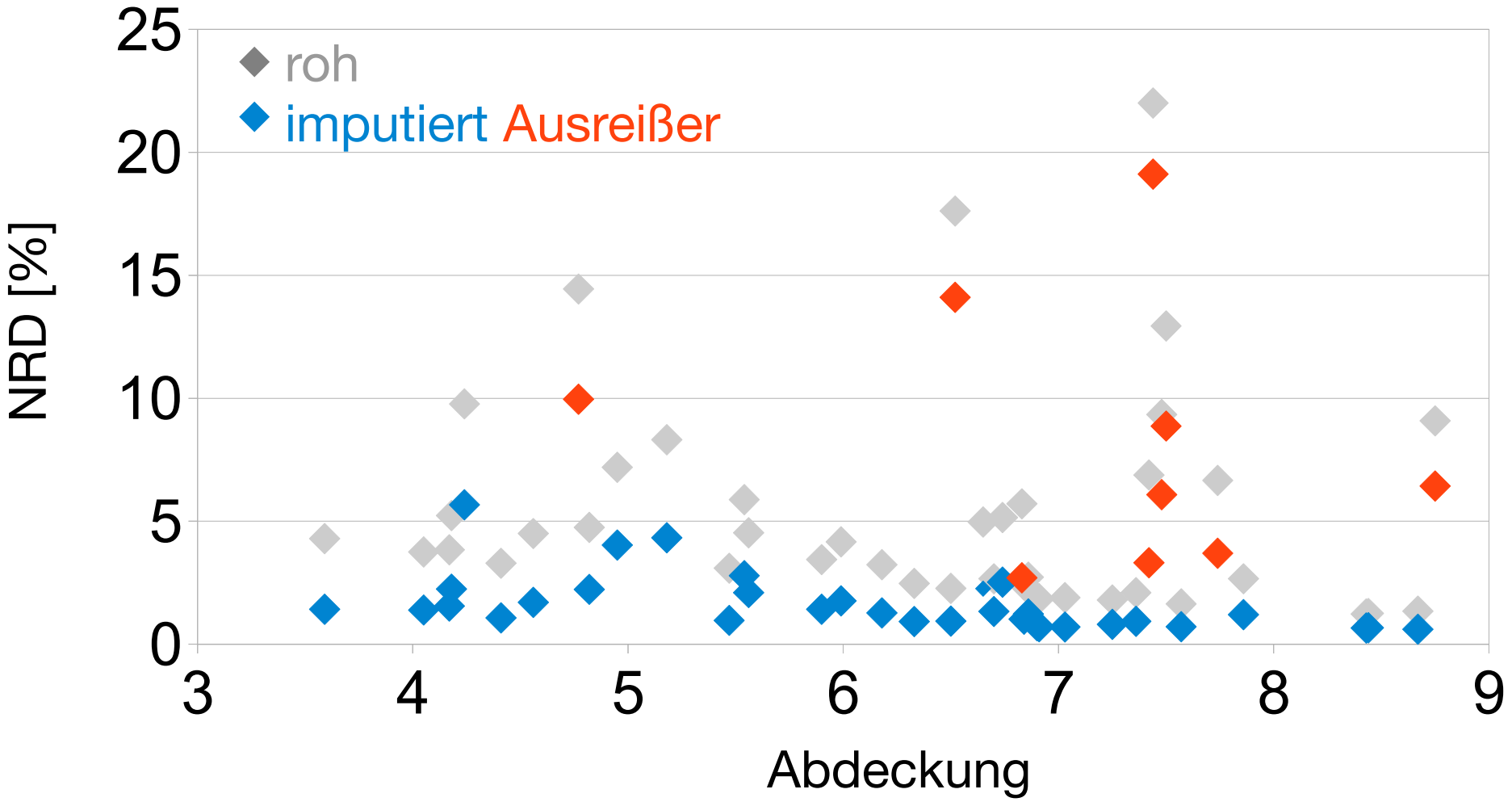
Abweichung von HD (777K) -Genotypen
(Non-Reference-Discrepancy-Rate, NRD)

HD-Genotypen

Seq. Genotypen	AA	AB	BB
AA	1	2	3
AB	4	5	6
BB	7	8	9
- -	10	11	12

$$NRD = \frac{\text{rot}}{\text{blau} + \text{rot}}$$

Effekt der *Beagle-Imputation* auf NRD



„Reference homozygosity discrepancy rate“ (RHD)

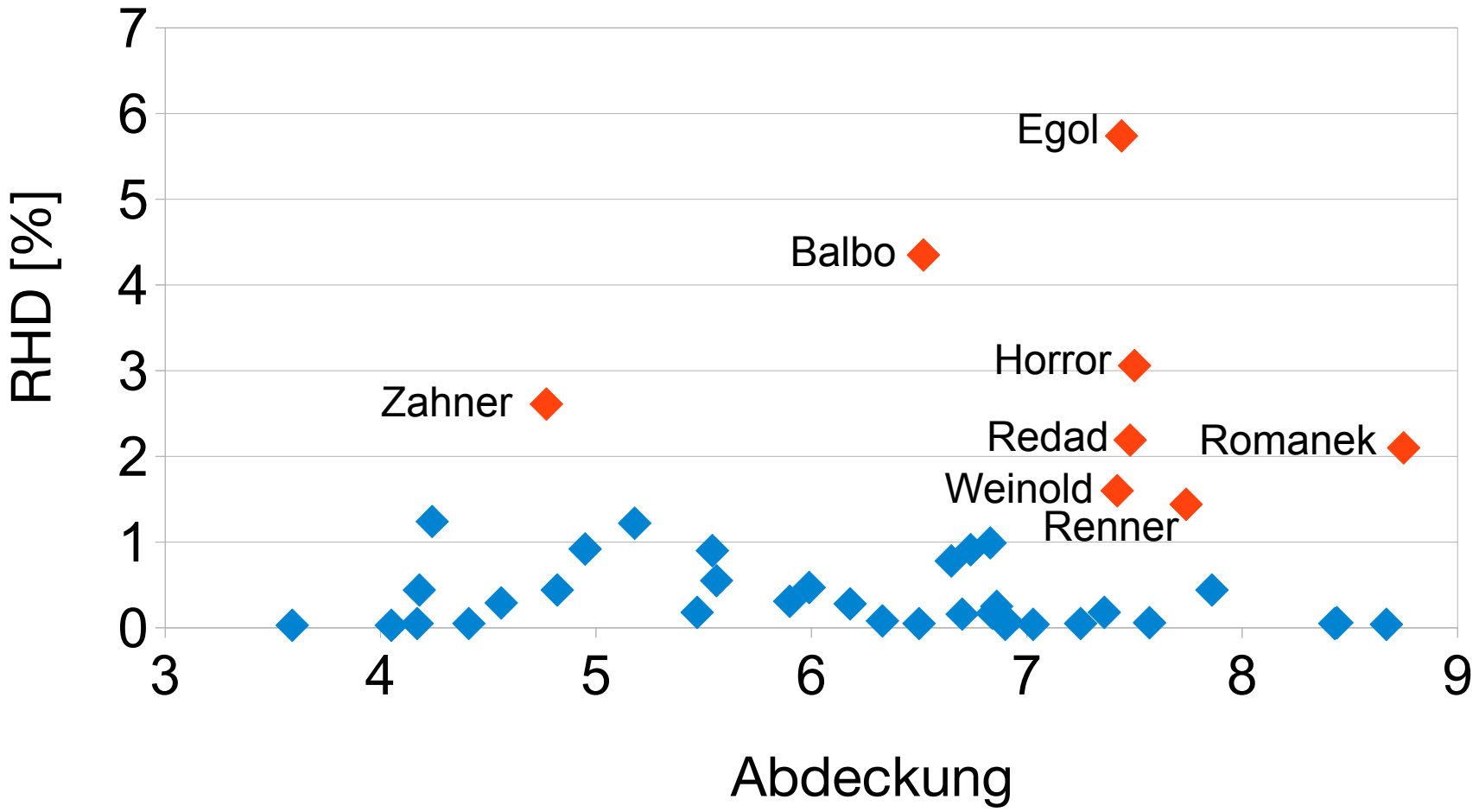
HD Genotypen

Seq. Genotypen

	AA	AB	BB
AA	1	2	3
AB	4	5	6
BB	7	8	9
- -	10	11	12

$$\text{RHD} = \frac{\text{red}}{\text{blue} + \text{red}}$$

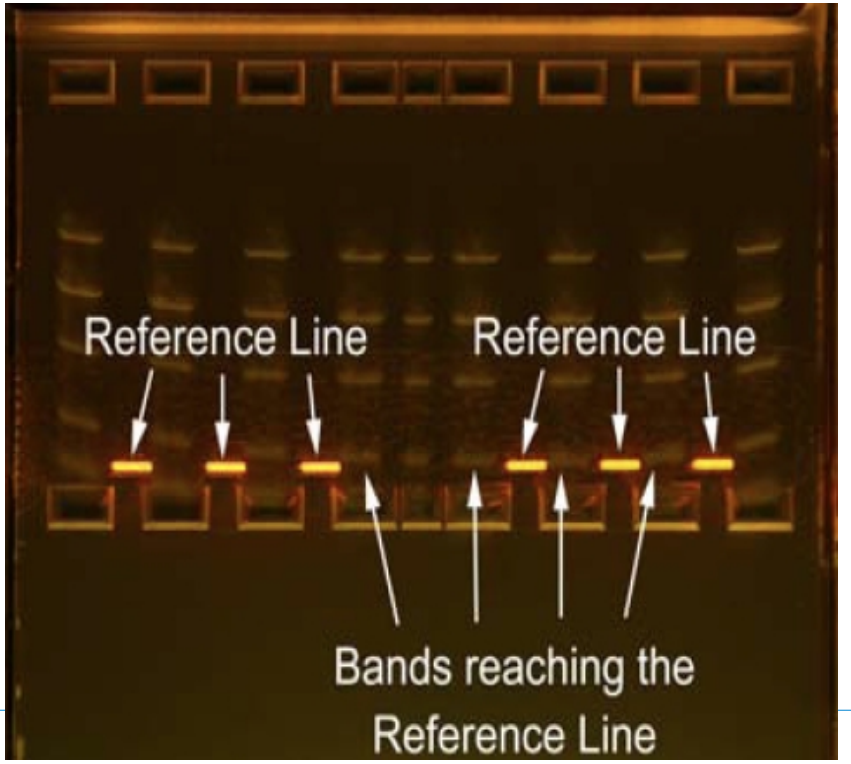
RHD: Kontamination



Contamination during fragment-size-selection step of library preparation

- Open electrophoresis system
- Samples can diffuse, leading to cross-well contamination

E-Gel® SizeSelect Agarose Gels (Invitrogen)



Populationsweite Sequenz-Imputation

Imputation 1. Schritt: Erstelle Haplotypen-Bibliothek bei den sequenzierten Schlüsseltieren

A	G	A
A	G	G
C	C	T
G	G	C
T	A	T
A	A	T
T	C	T
T	T	C
A	A	C
T	C	T

...

Imputation 2. Schritt: Populationsweite Genotypisierung (54K, 777K)

A	G	A
A	G	**G**
C	C	T
G	G	C
T	A	T
A	A	T
T	C	T
T	T	**C**
A	A	C
T	C	T

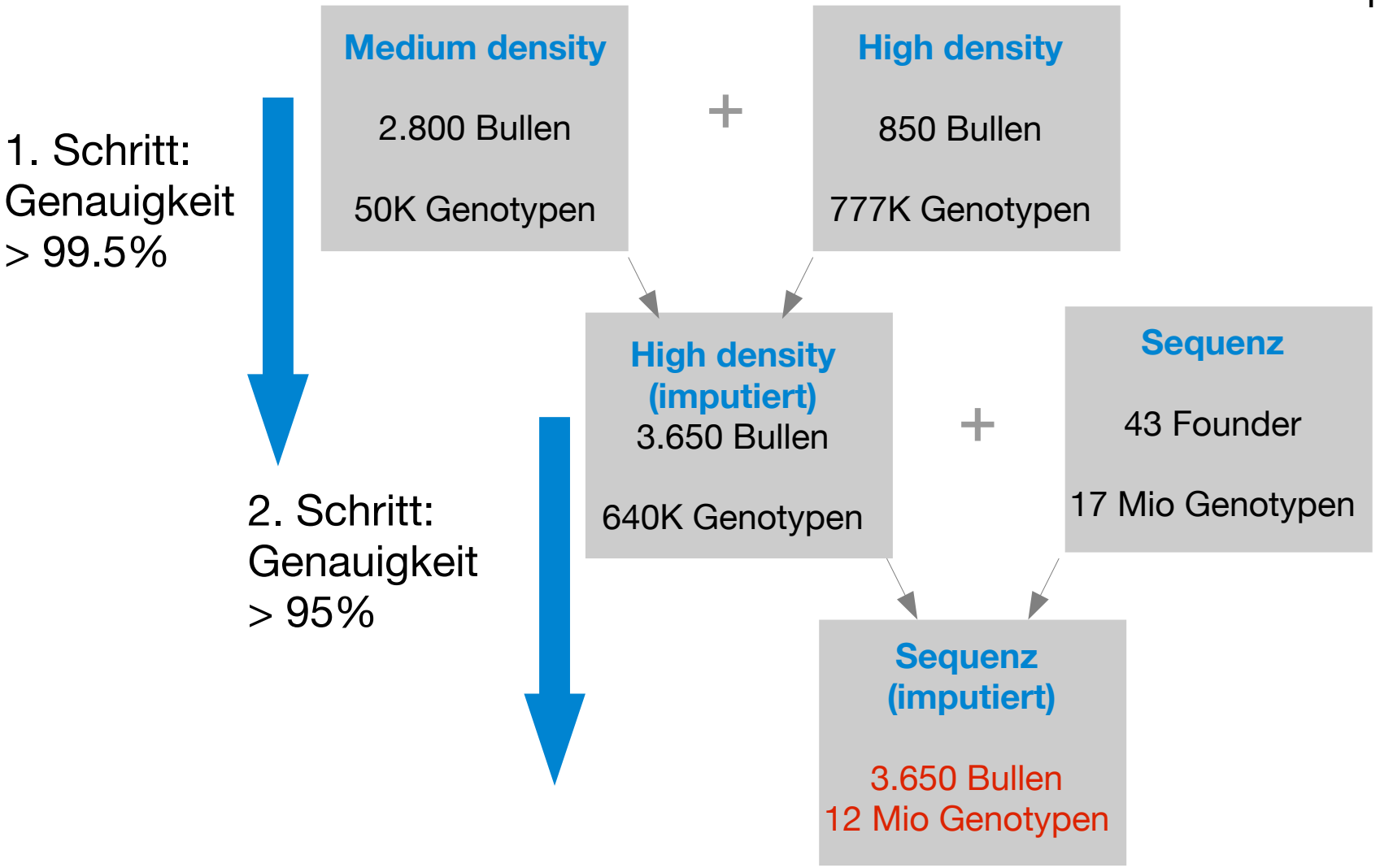
...

Imputation 3. Schritt: Leite Sequenz ab

A	G	A		AA
A	G	G		AG
C	C	T		CT
G	G	C		GC
T	A	T		TT
A	A	T	→	AT
T	C	T		TT
T	T	C		TC
A	A	C		AC
T	C	T		TT

Sequenzimputation

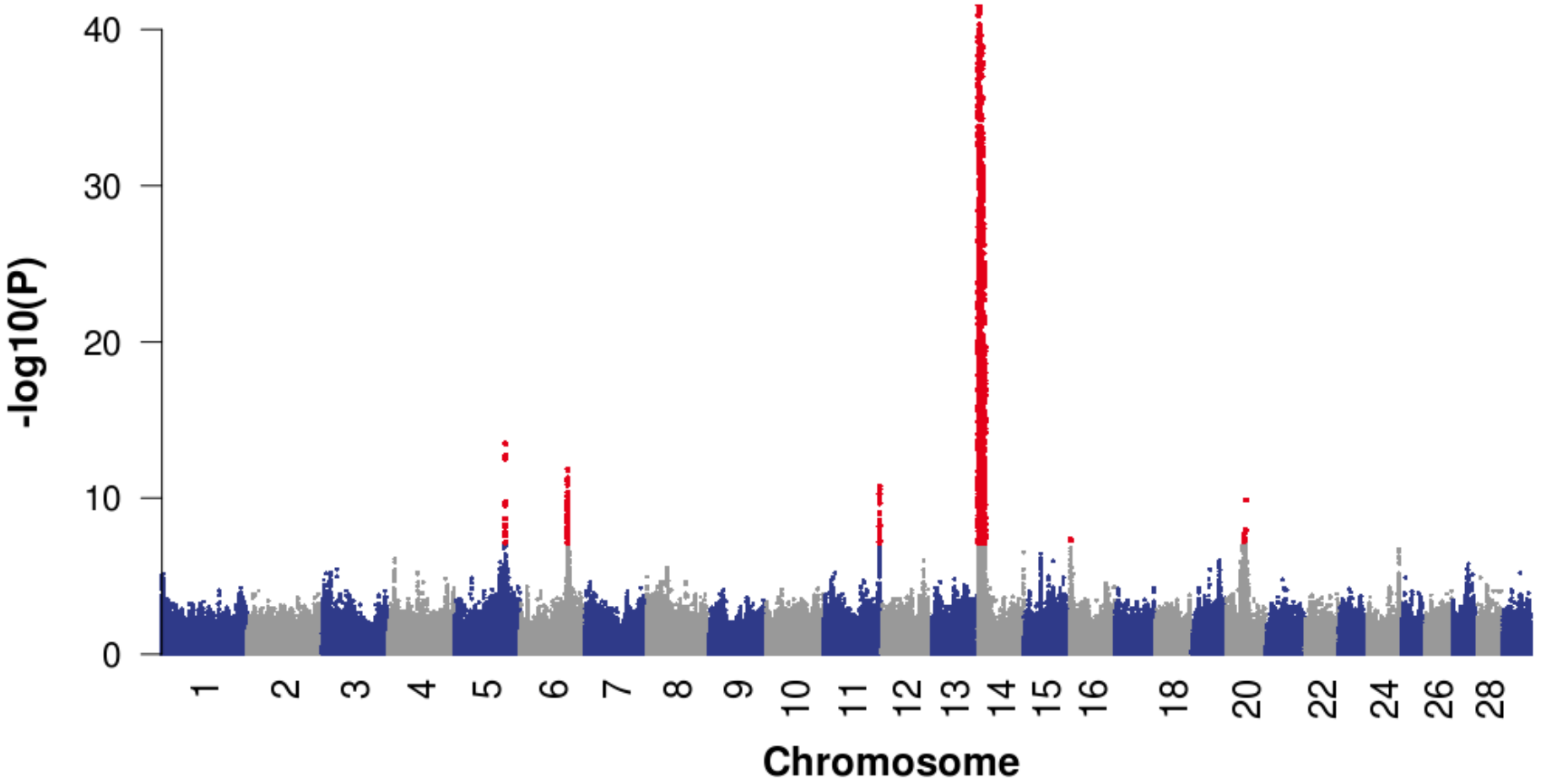
Intel E5620
48 GB RAM
16 threads



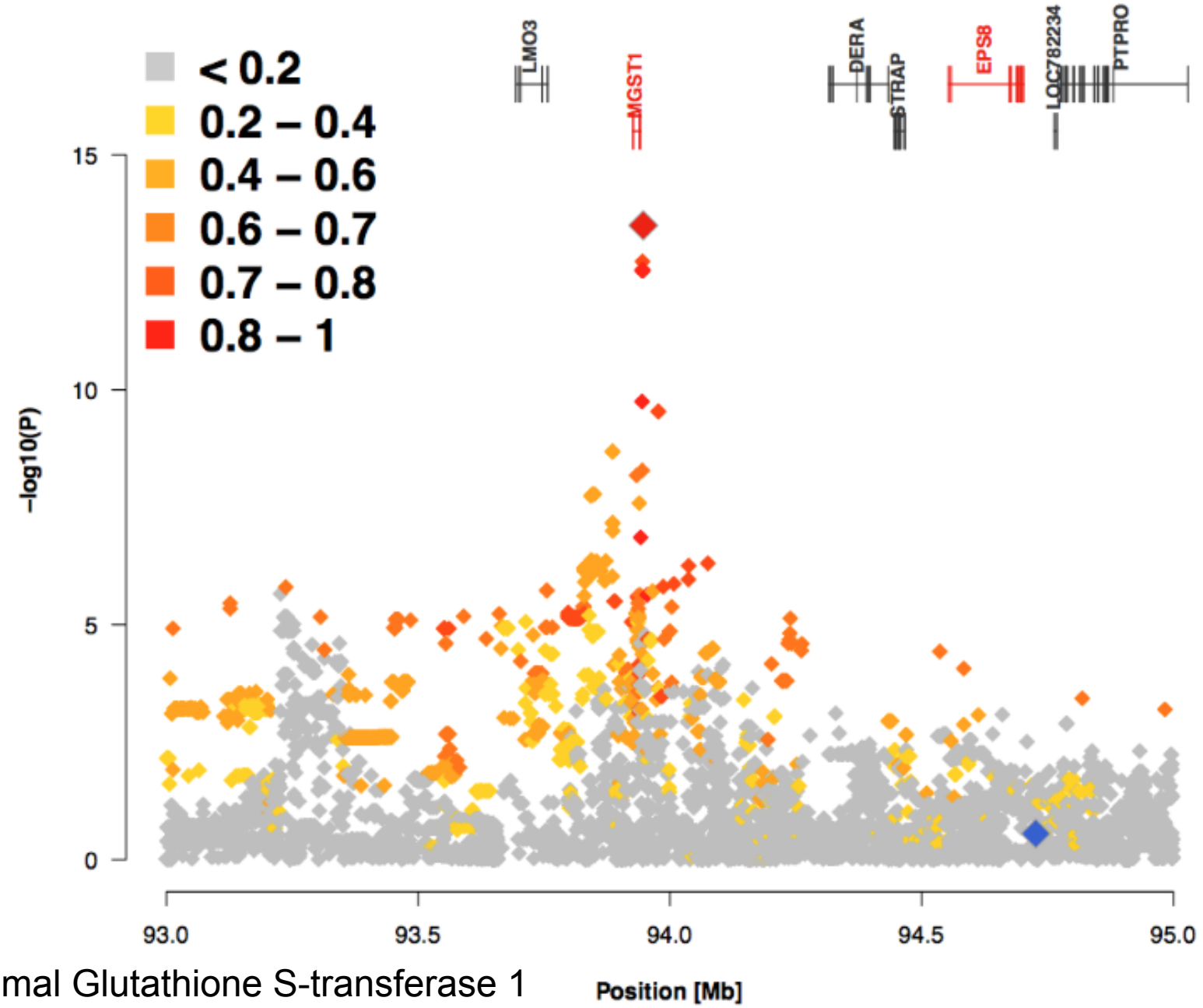
Auf dem Weg zur GS 2.0 (1)

- Identifizierung von “very important QTL” (VIQ) mit genomweiten Assoziationsstudien
  - ca. 5 pro Merkmal und Population
- Identifizierung von quasi-kausalen Varianten im Bereich der VIQ
- Untersuchung nicht-additiver Effekte der VIQ bei Kühen

FV: Genomweite Assoziationsstudie 3600 Bullen, 12 Mio SNPs, Fett% 305 Tage

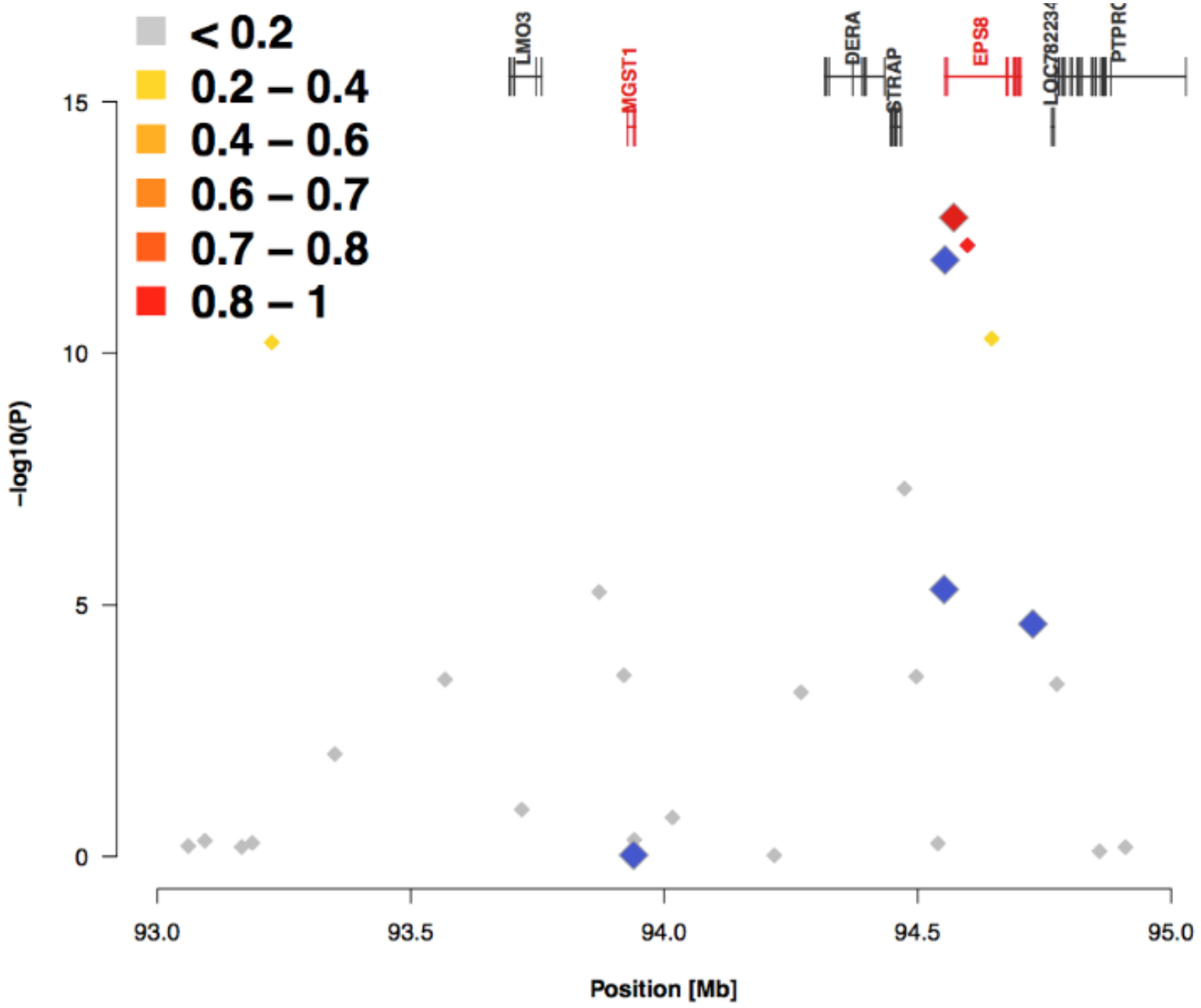


FV: Fett-% QTL auf Chromosom 5 – 12 Mio SNPs

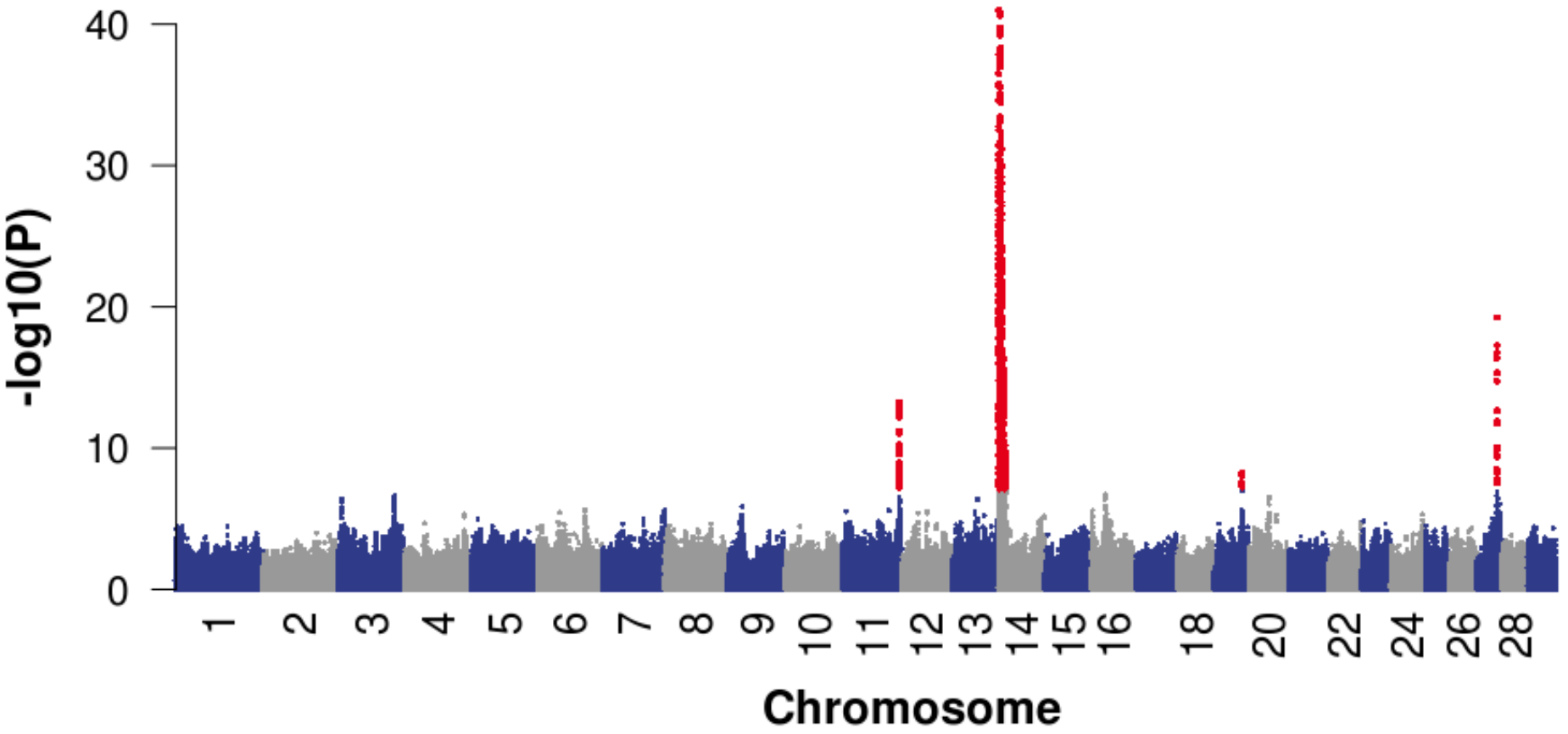


Microsomal Glutathione S-transferase 1

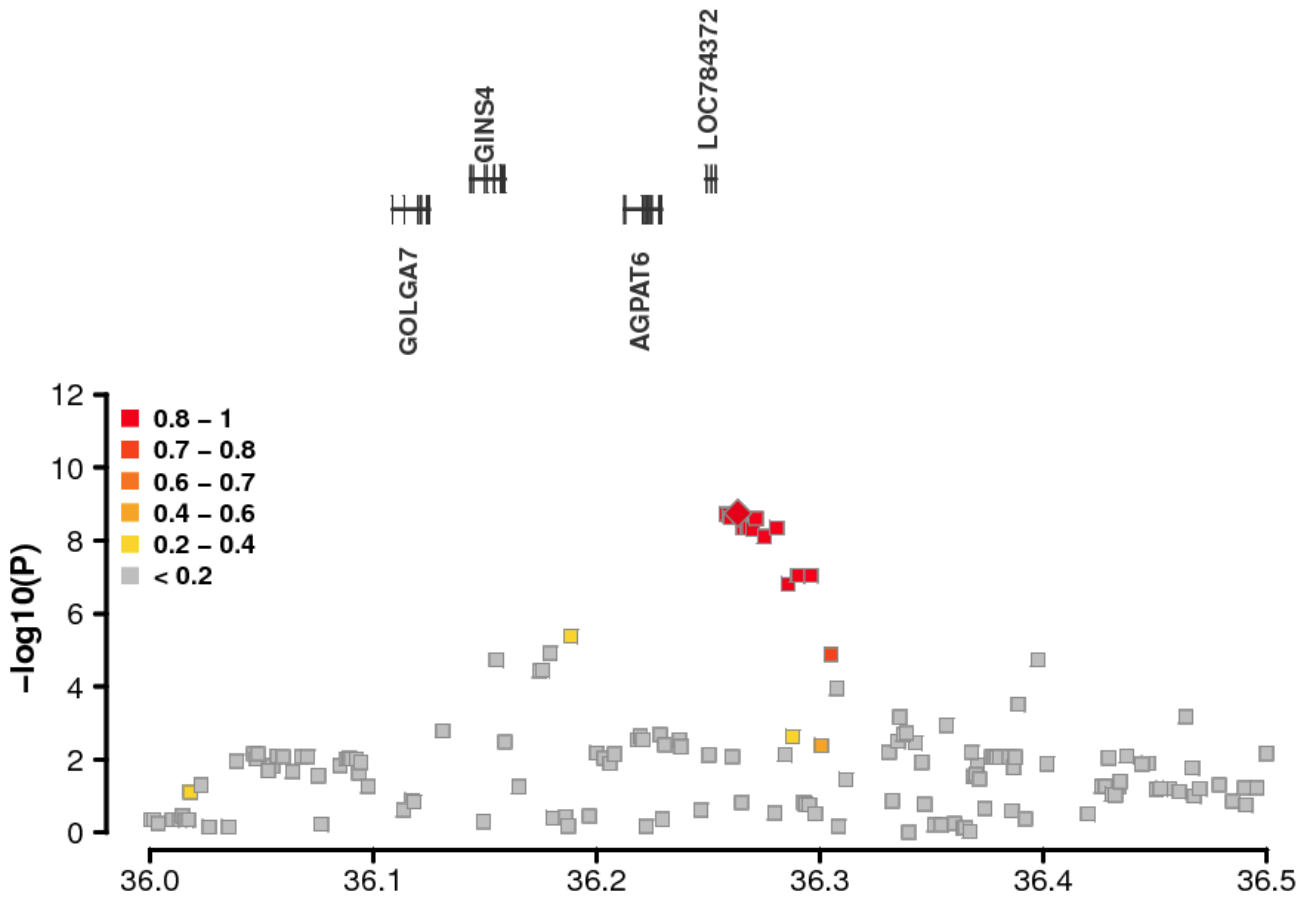
HF: Fett-% QTL auf Chromosom 5 – 54K SNPs



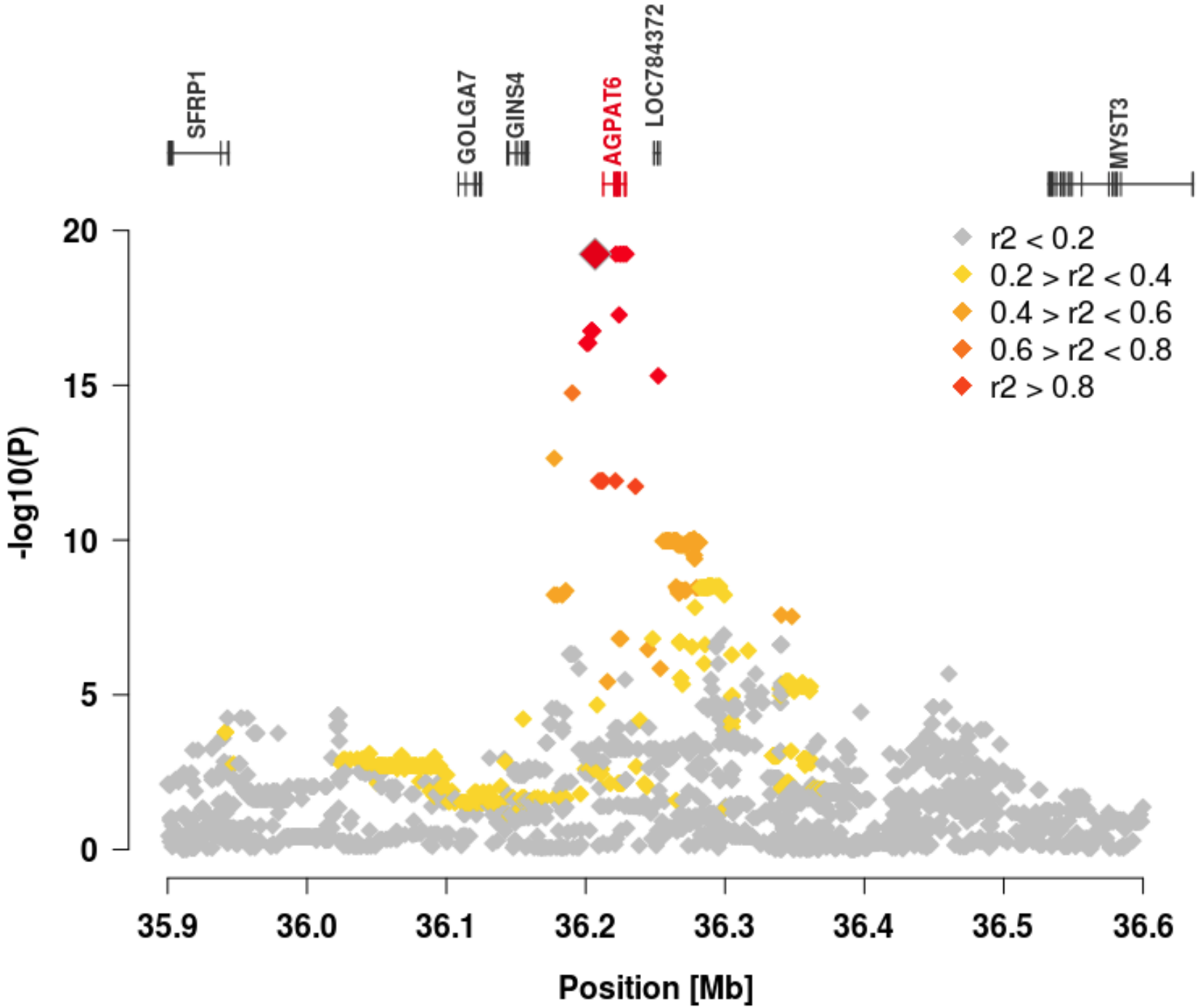
FV: Genomweite Assoziationsstudie 3600 Bullen, 12 Mio SNPs, Fett% Tage 8-12

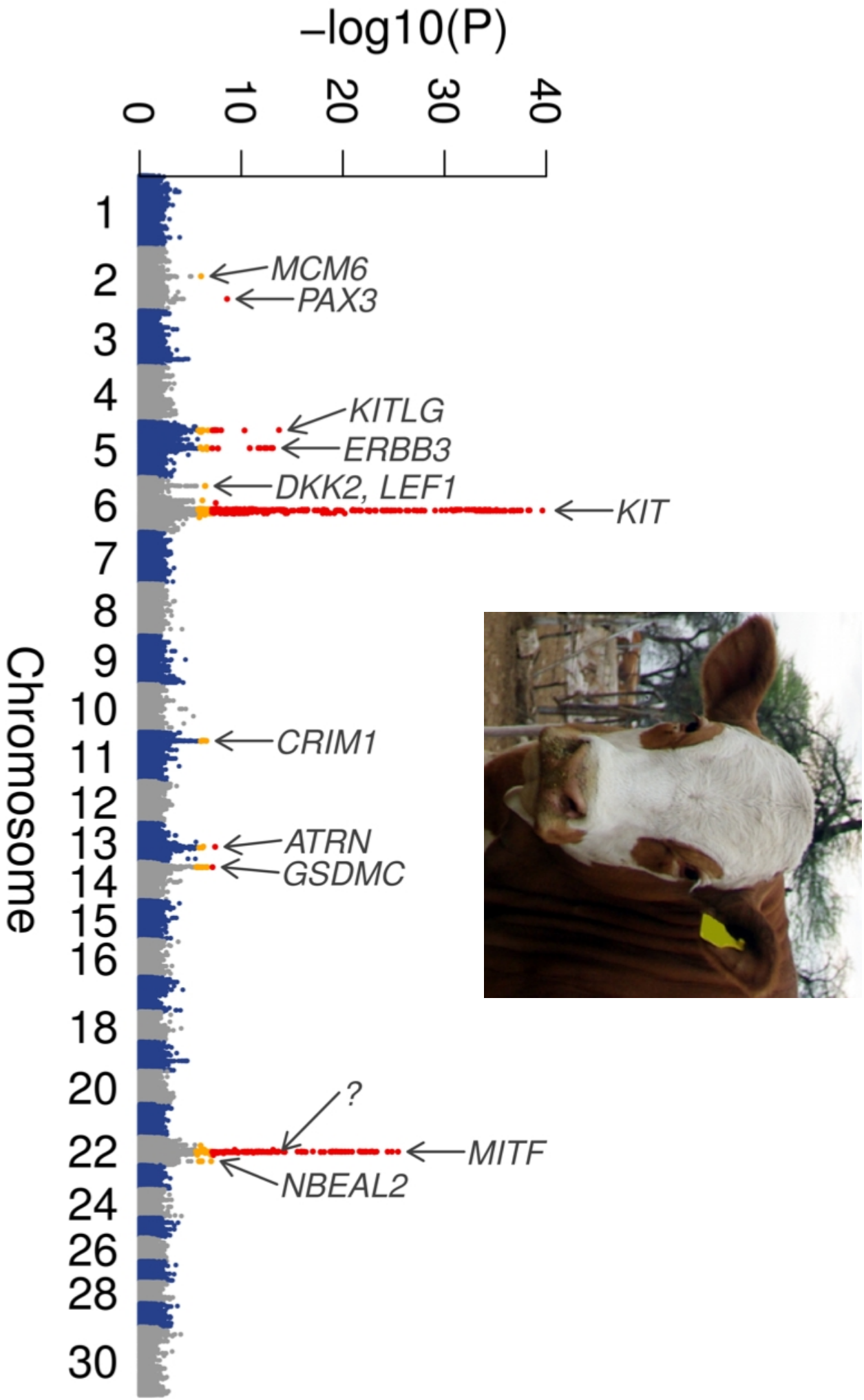


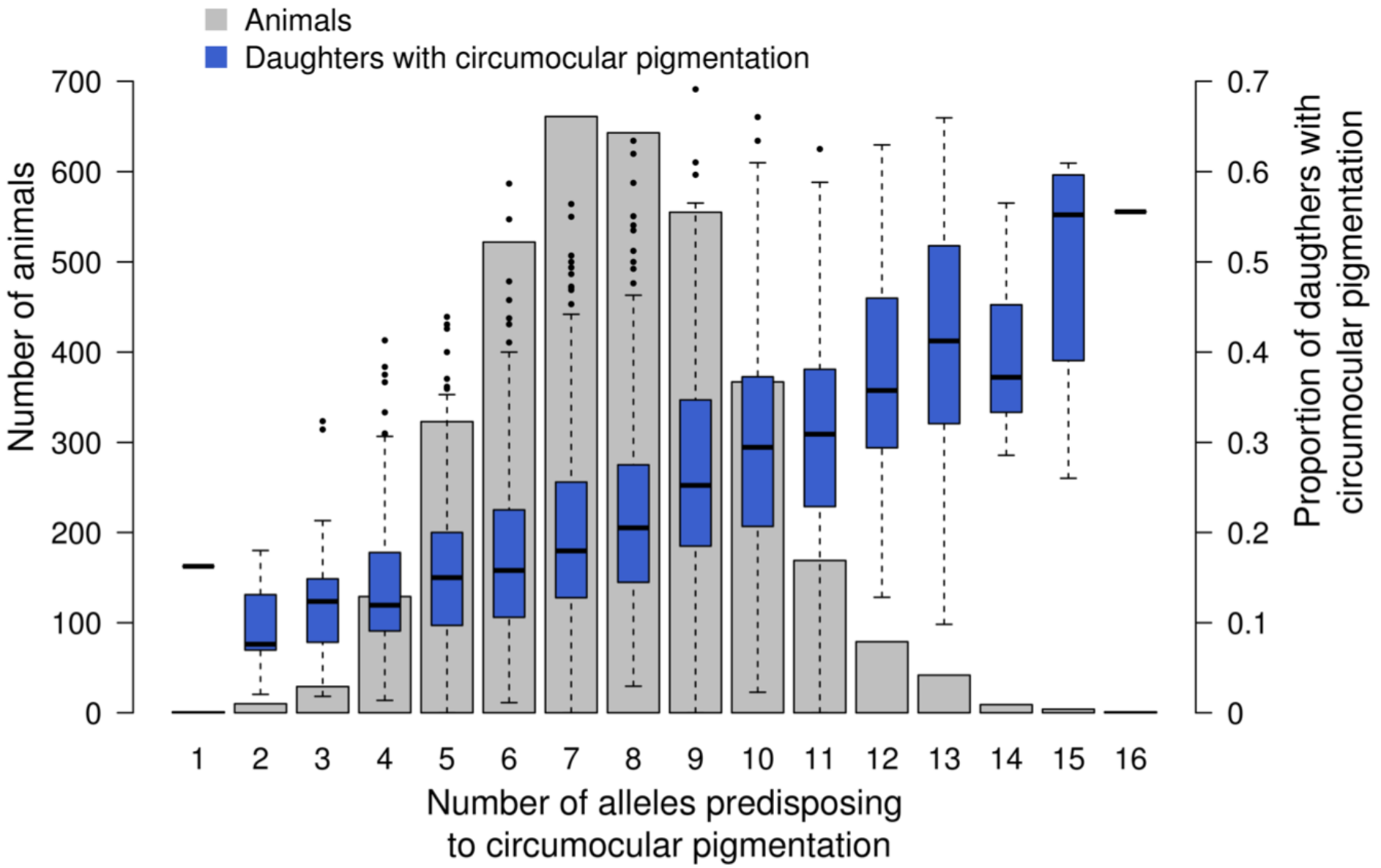
FV: Fett-% QTL auf Chromosom 27 – 777K SNPs

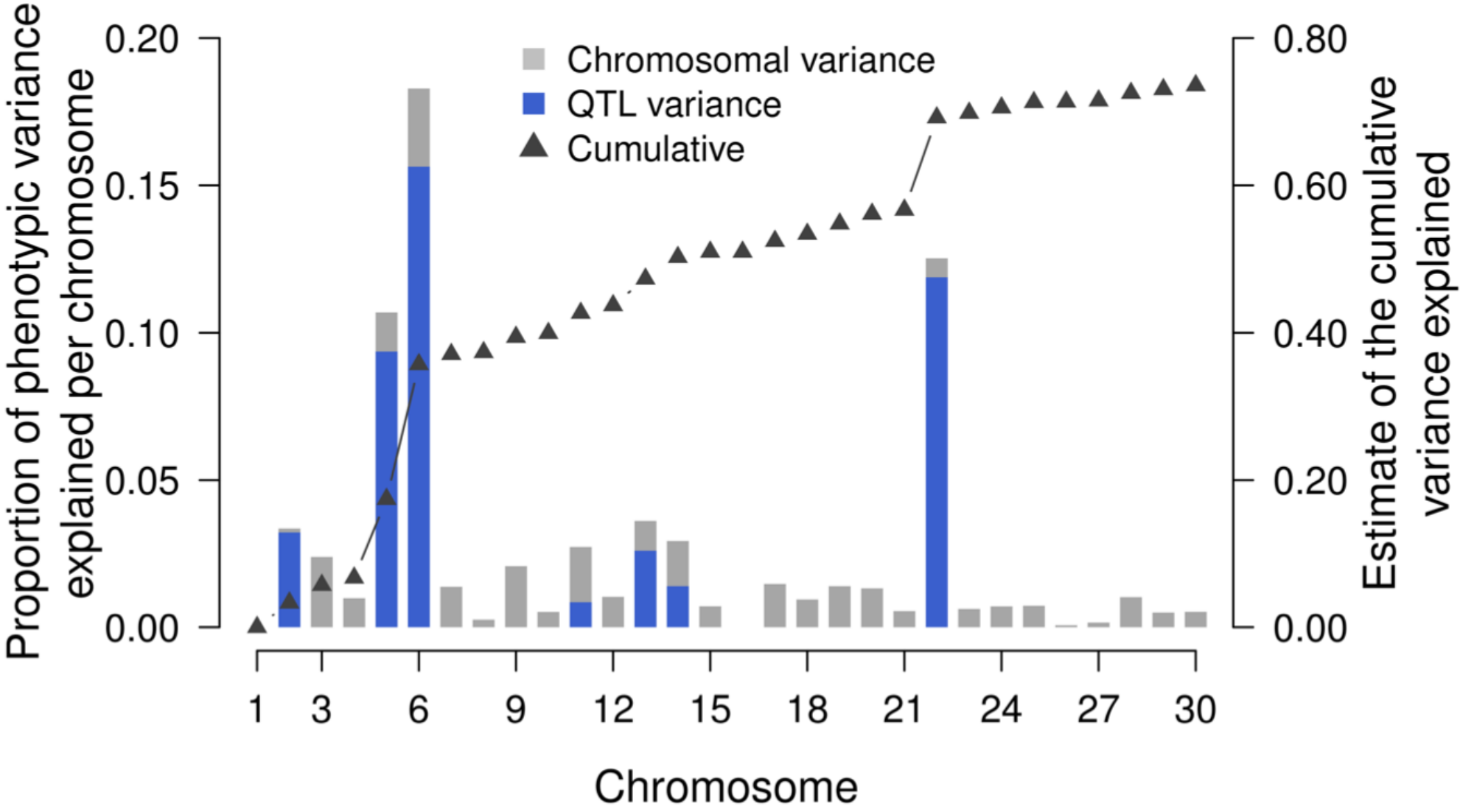


FV: Fett-% QTL auf Chromosom 27 – 12 Mio SNPs




−log10(P)
0
10
20
30
40
Chromosome
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
18
20
22
24
26
28
30
MCM6
PAX3
KITLG
ERBB3
DKK2, LEF1
KIT
CRIM1
ATRN
GSDMC
?
MITF
NBEAL2

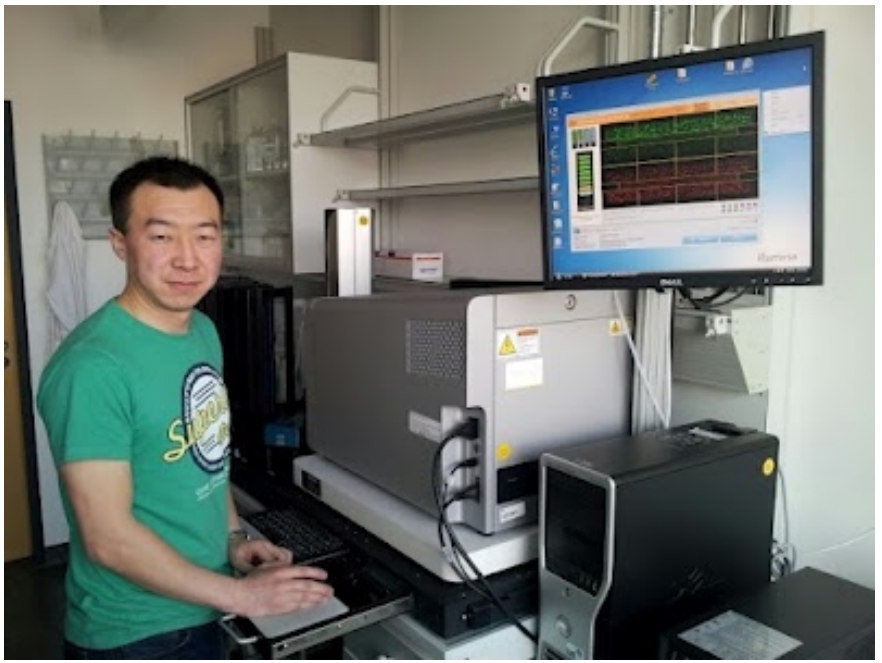

Animals
Daughters with circumocular pigmentation
Number of animals
0
100
200
300
400
500
600
700
Proportion of daugthers with circumocular pigmentation
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
Number of alleles predisposing to circumocular pigmentation

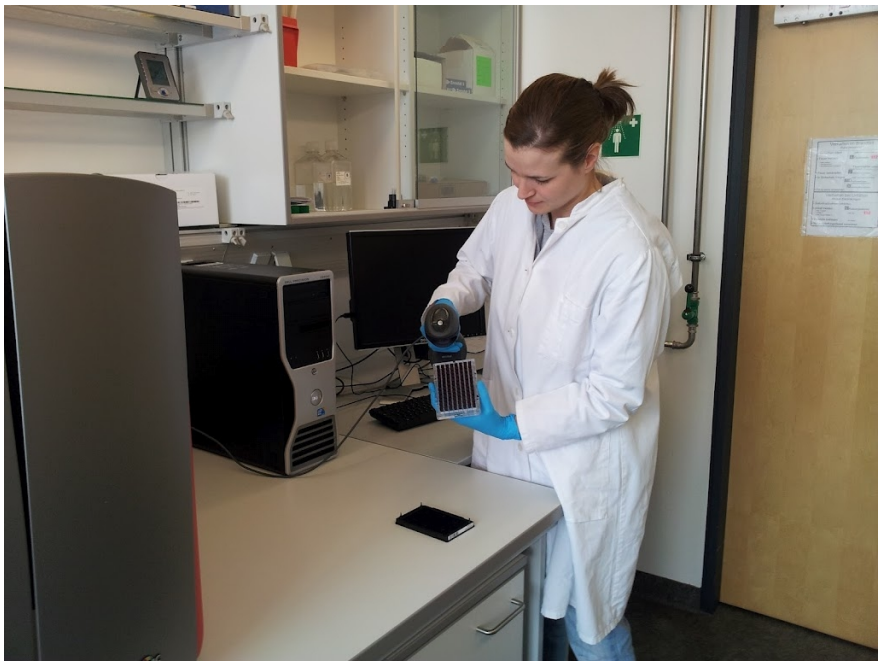

Chromosomal variance
QTL variance
Cumulative
Proportion of phenotypic variance explained per chromosome
Estimate of the cumulative variance explained
0.00
0.05
0.10
0.15
0.20
0.00
0.20
0.40
0.60
0.80
1
3
6
9
12
15
18
21
24
27
30
Chromosome

Auf dem Weg zur GS 2.0 (2)

- Vorbereitung der flächendeckenden HD-Genotypisierung (> 600 SNPs für 30 € pro Tier)
- Entwicklung der Schätzung des genomischen Produktionswertes auf der Basis
  - der imputierten Genomsequenz,
  - der VIQ-Information und
  - der allgemeinen (“ = polygenen”) genomischen Formel

30 € pro Tier für HD-Genotypisierung?





illumina

777K

Affymetrix

650K

Hubert Pausch und ich danken ...

- dem BMBF für die Unterstützung im Rahmen von GenoTrack und Synbreed,
- Sebastian Eck, Anna Benet-Pagès, Elisabeth Graf, Thomas Wieland, Tim Strom und Thomas Meitinger vom Lehrstuhl für Humangenetik der TUM / Institut für Humangenetik des Helmholtz-Zentrums München für die Sequenzierung,
- Sandra Jansen, Bernhard Aigner, Xiaolong Wang, Tini Wurmser und Michal Wysocki vom Lehrstuhl für Tierzucht für die Organisation der DNA-Logistik, Genotypisierung und Datenanalyse.

Steve Jobs' quote on product design applied to genomic evaluation (1)

- *"When you start looking at a problem and it seems really simply, you don't really understand the complexity of the problem."*
  - Genomic evaluation and improvement of animals
- *"Then you get into the problem, and you see it's really complicated, and you come up with all these convoluted solutions."*
  - Bayes A – Z, ... methods for genomic evaluation

Steve Jobs' quote on product design applied to genomic evaluation (2)

- *"That's sort of the middle, and that's where most people stop"*
  - Genomic selection 1.0: keep genes in the "black box"
- *"But let's keep going and find the key, the underlying principle of the problem – and come up with an elegant, really beautiful solution that works."*
  - Genomic selection 2.0: gene-based evaluation and selection