NGS-based HLA and Bloodgroup Typing -

One-Stop Shop (?)

Prof. Andre Franke
Institute of Clinical Molecular Biology | Kiel University

Swisstransfusion, 25.8.16 in Bern

ikmb
Institut für Klinische Molekularbiologie

Focus on two methods

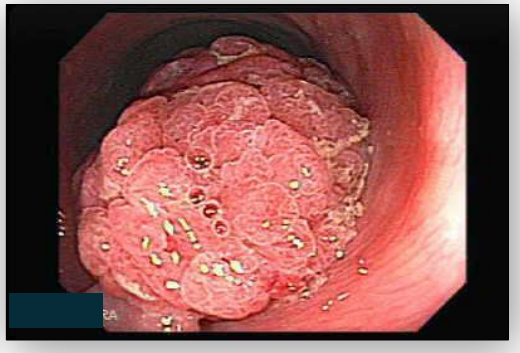
1) Imputation of genotypes
2) In-solution capture of DNA using RNA baits

Inflammation

An important mechanism

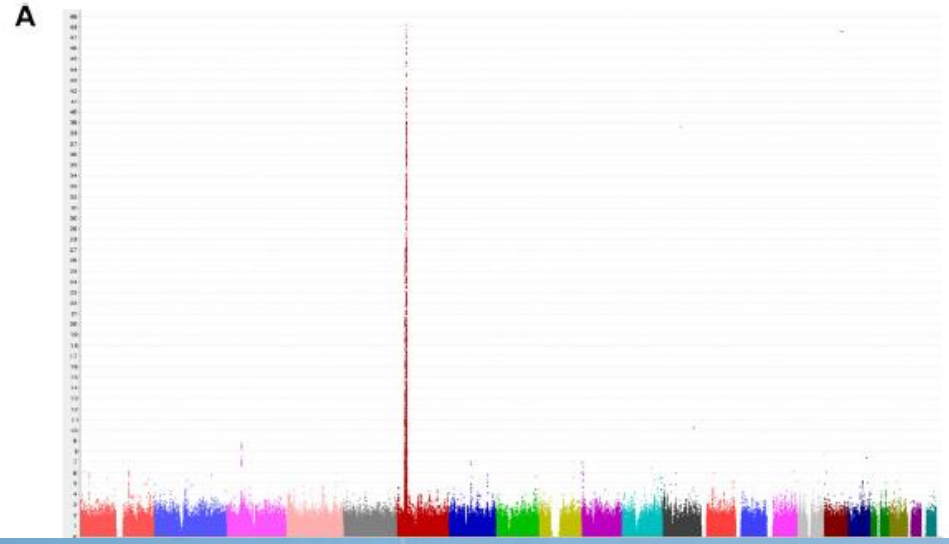
Exzellenzcluster
Entzündungsforschung







Importance of HLA association



PSC (Ulcerative colitis)

disease

iabetes

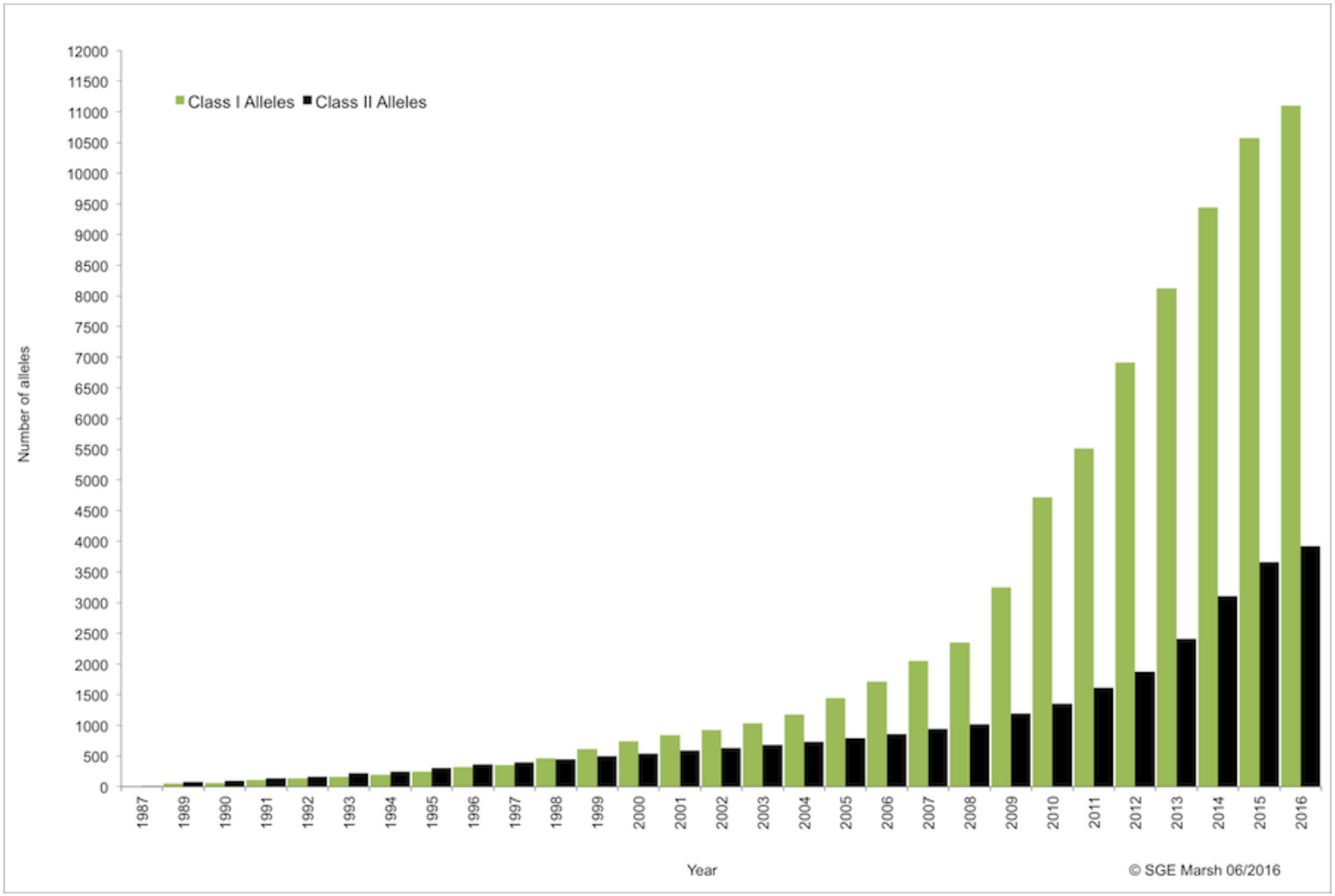
sclerosis

Number of different alleles

- Number of alleles annotated for the most important HLA loci, IMGT/HLA database v3.25 , July 2016

HLA Class I					
Gene	*A*	*B*	*C*		
Alleles	3,492	4,358	3,111		
Proteins	2,480	3,221	2,196		
HLA Class II					
Gene	*DRB*	*DQA1*	*DQB1*	*DPA1*	*DPB1*
Alleles	2,135	73	940	43	671
Proteins	1,569	33	647	21	552

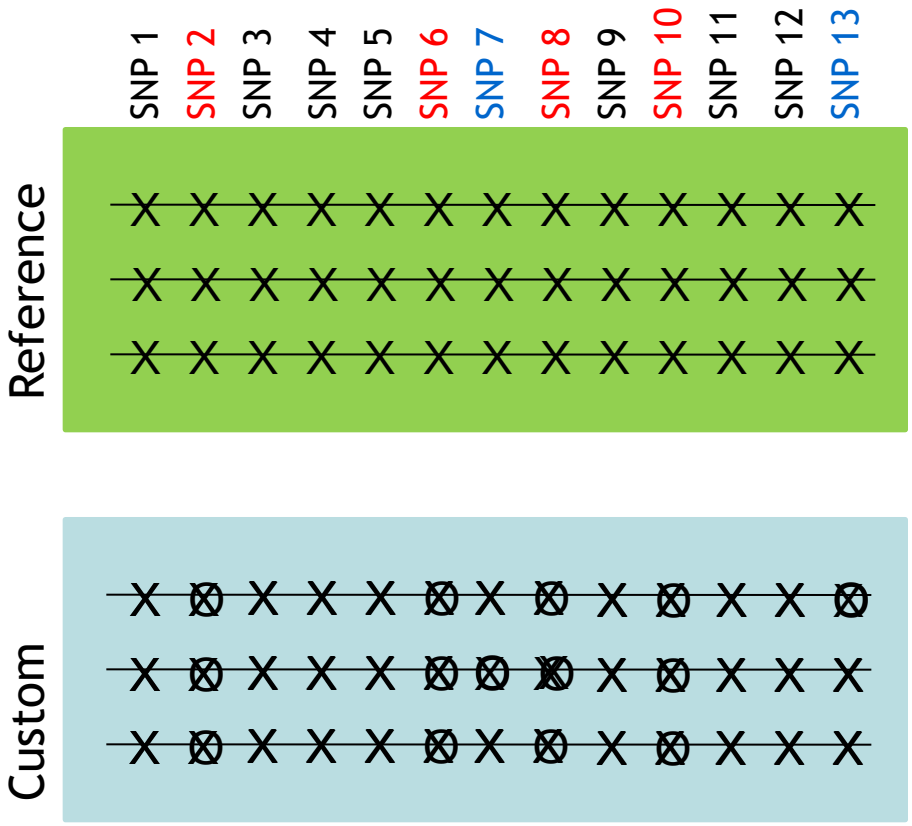
Growing number of alleles



„The HLA is a monster“*

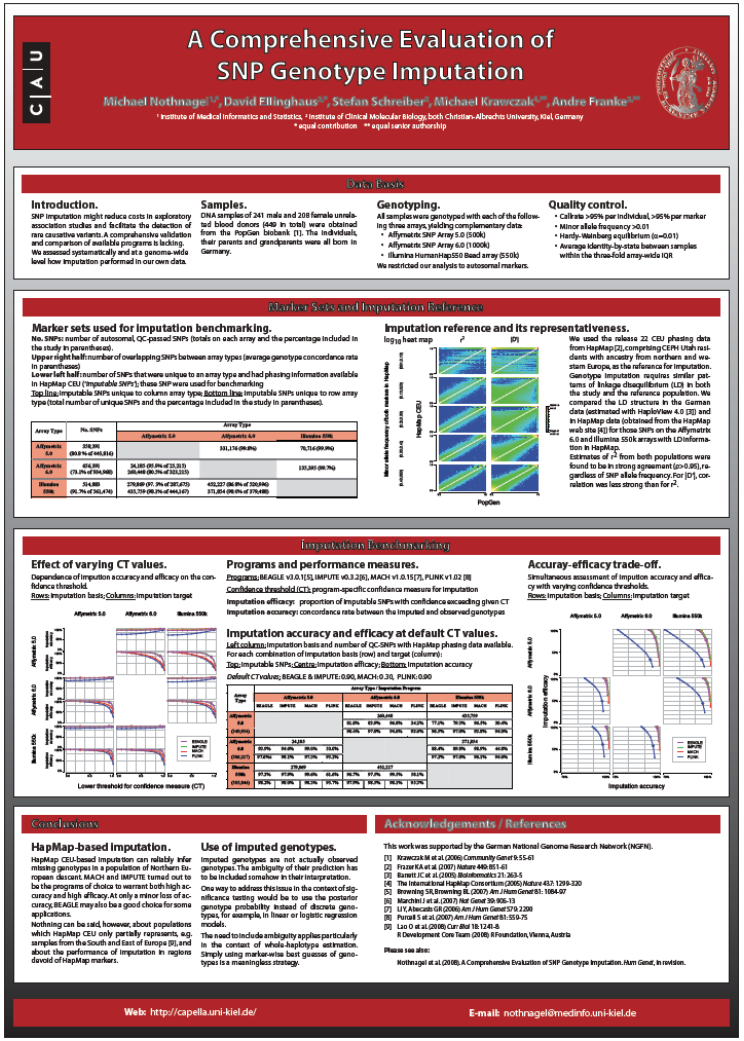
*Prof. Tom H. Karlsen 2008, Oslo, Norway

Genotype Imputation

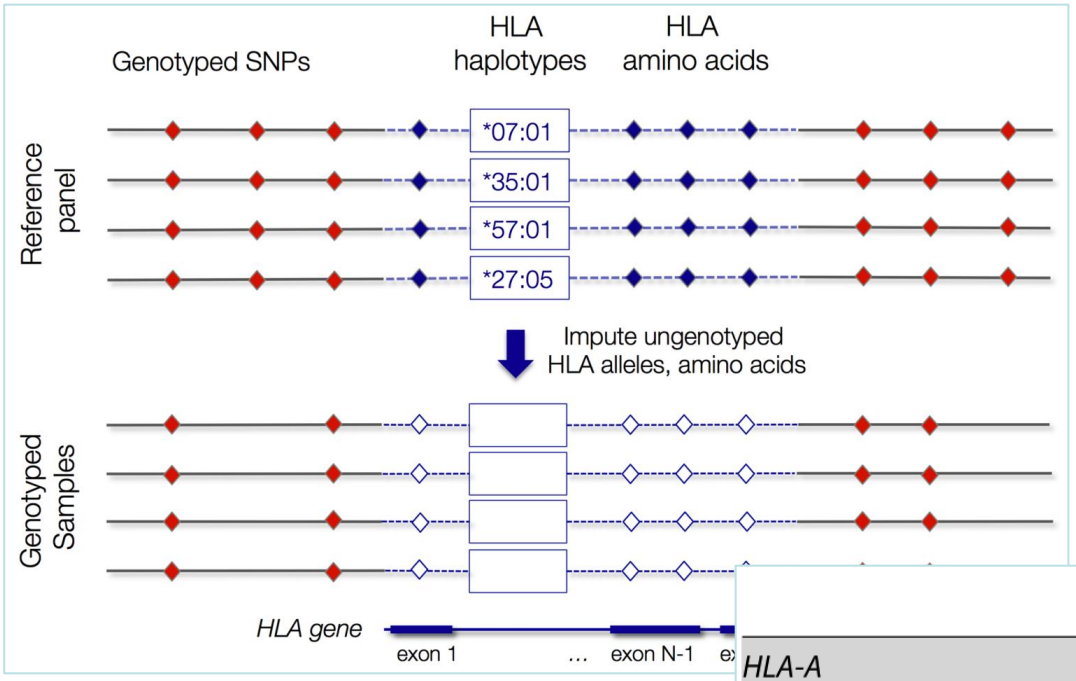


O Missing genotypes
X Available genotype data

Human Mutation, 2008



SNP2HLA



	CEU/CEPH	YRI	CHB+JPT
HLA-A	99.1%	69.9%	98.1%
HLA-B	96.8%	90.5%	65.6%
HLA-C	99.1%	98.4%	68.8%
HLA-DQA1	98.5%	64.9%	96.3%
HLA-DQB1	99.1%	96.1%	96.5%
HLA-DRB1	96.9%	20.3%	92.3%
All loci	98.3%	72.9%	86.4%

Jia *et al.*, PLoS One 2013

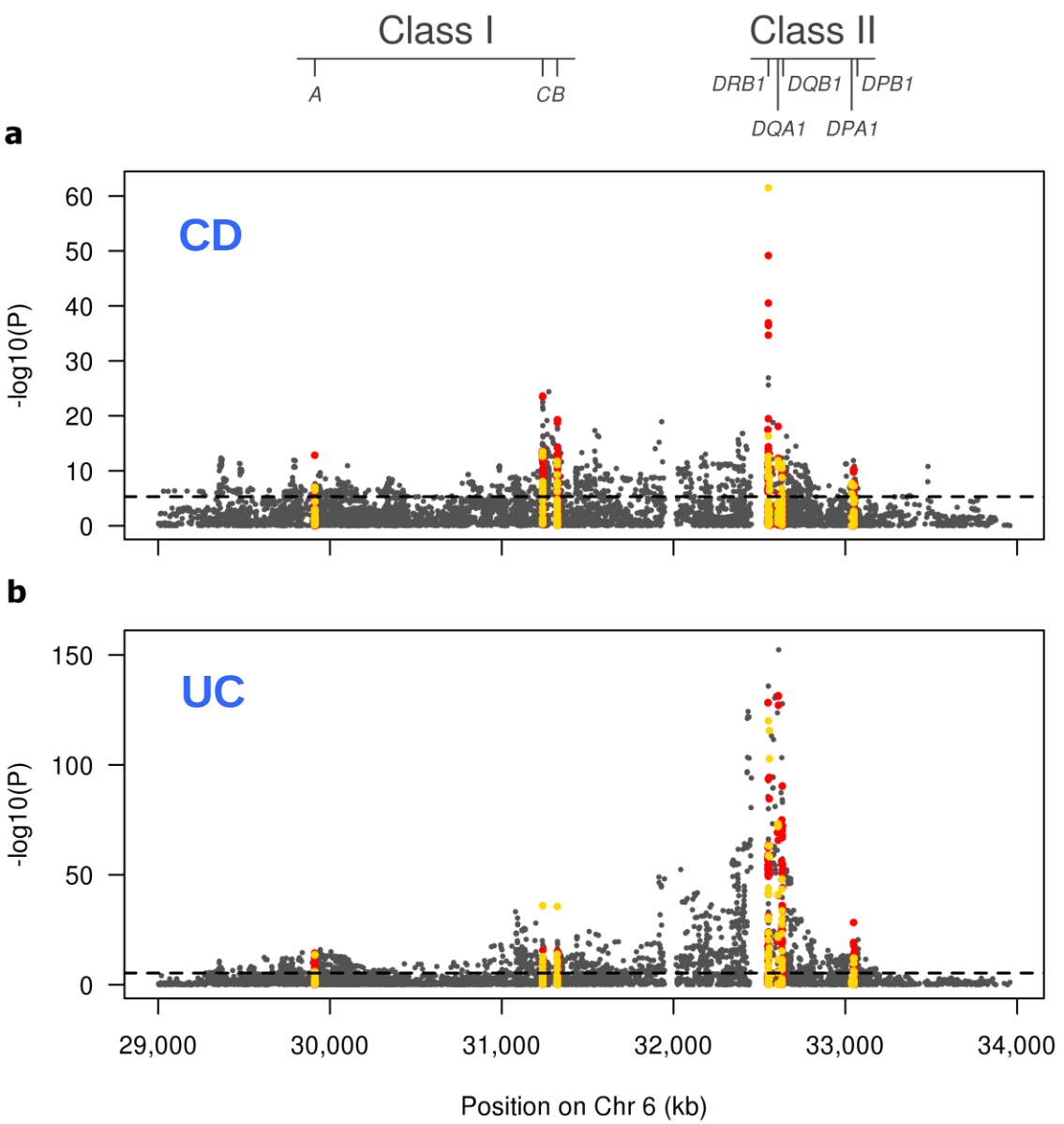
HLA fine mapping

34,241 controls

18,405 CD

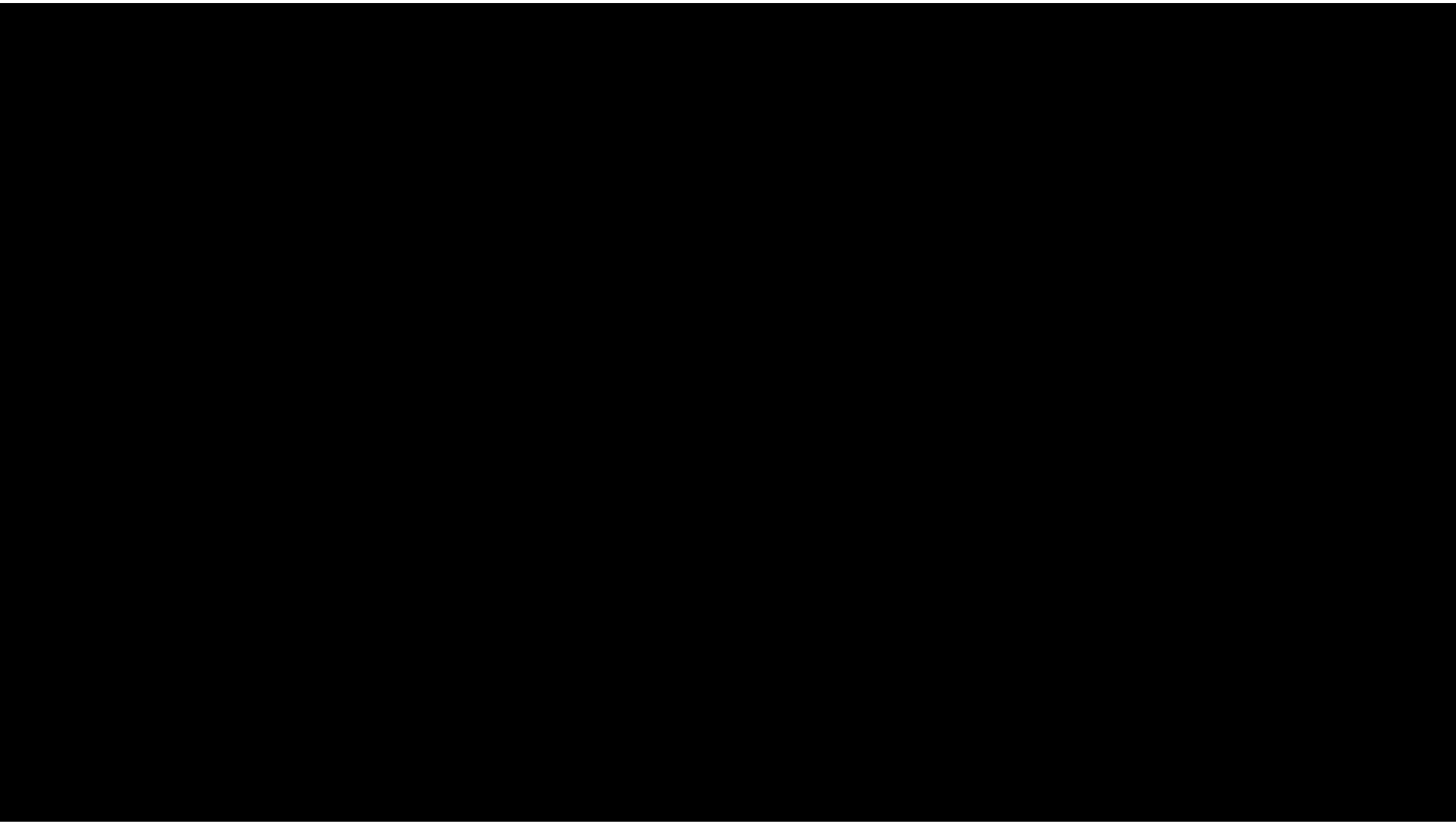
14,308 UC

- 8939 SNPs
- 741 AA
- 90/138 class. genes

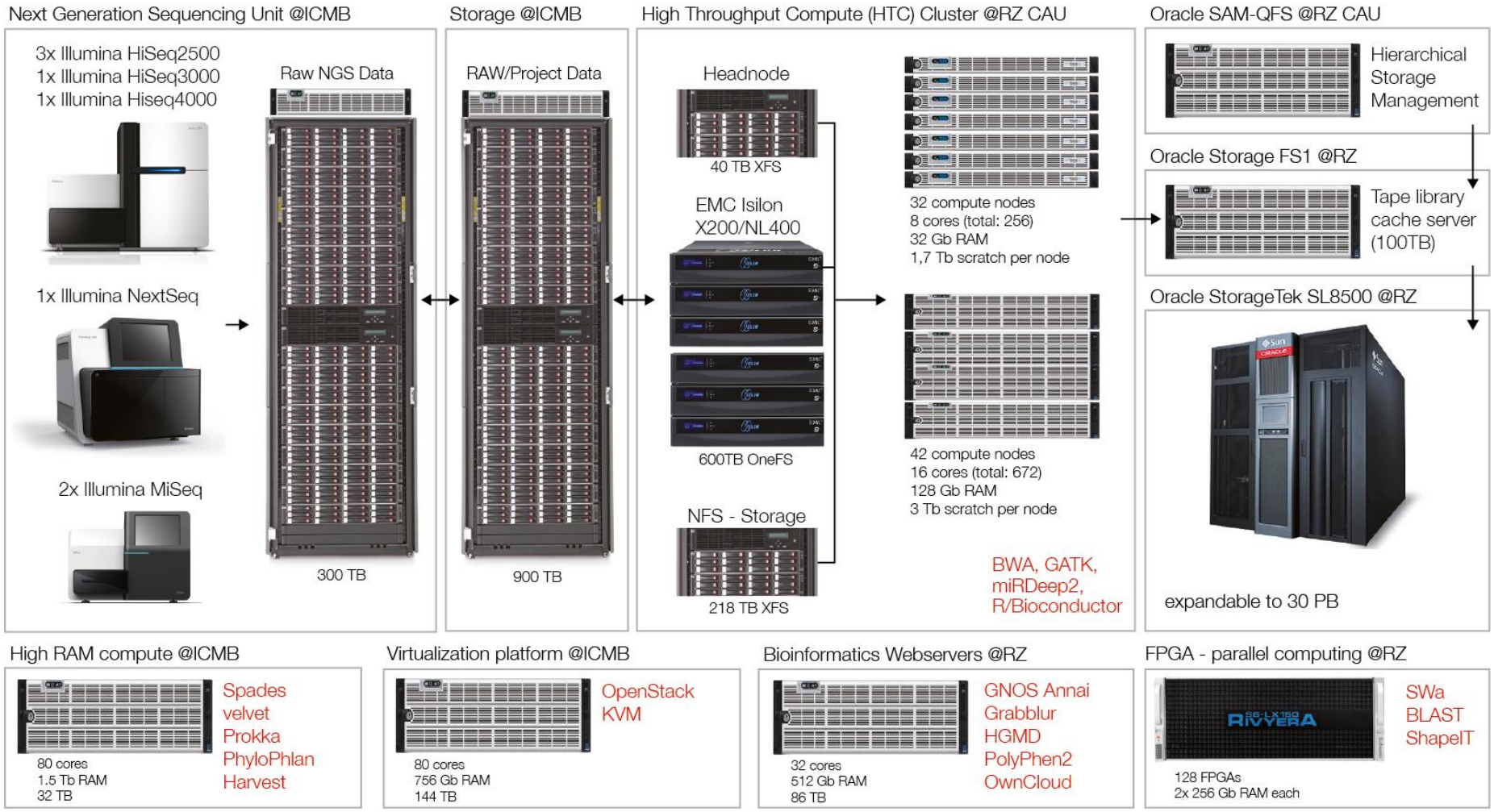


Franke & Rioux, *Nature Genetics 2015*

Classical analysis



IKMB Sequencing Center (since 2004)

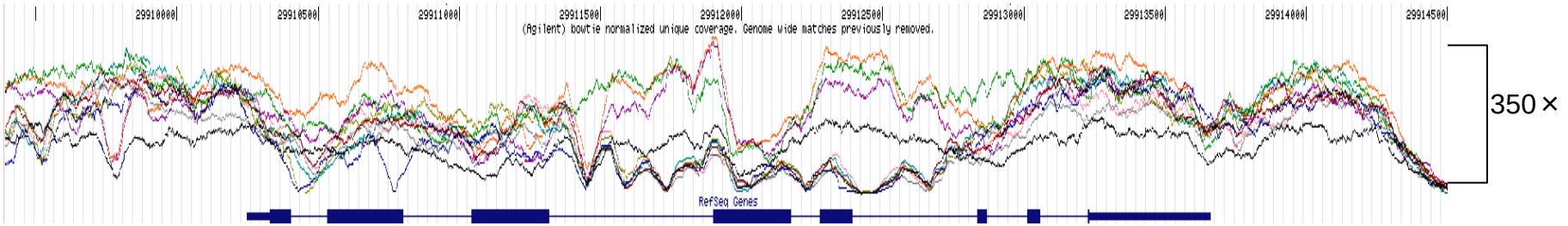


2012 study with reference-based design

- Alignment of the enriched NGS runs to pgf haplotype
  - Amb132, a heterozygous sample (2 replicates)
  - Lev337, a heterozygous sample
  - 7 haploid cell line samples of known haplotype

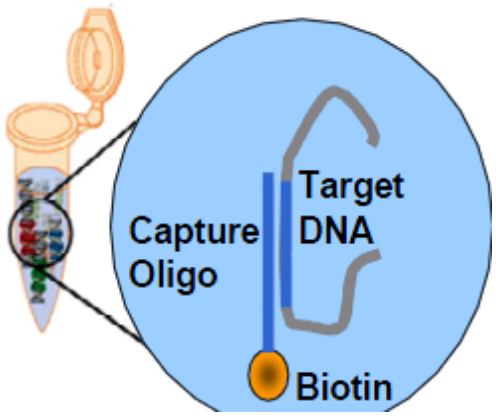
Horton et al. Immunogenetics (2008) 60: 1-18

Lev337
Amb132
Amb132
qbl
cox
mou_mann
ssto
pgf
apd
dbb

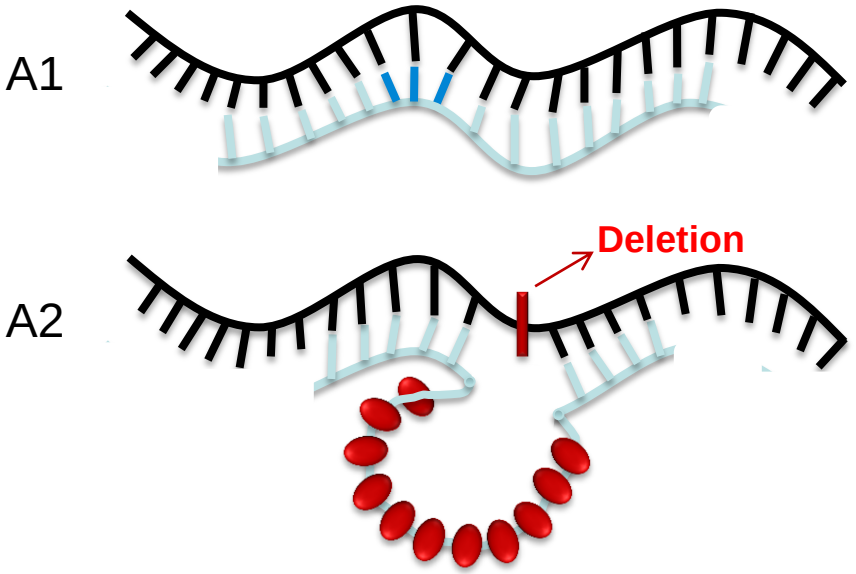


Setup of benchmark

- 357 samples (enriched for diversity)
- 48 samples/lane on HiSeq 2500
- SureSelect Automated LibraryPrep and Capture System



In-solution



Diversity of test samples

	Sample #	A	B	C	DPA	DPB	DQA	DQB	DRB
NGS Kiel	299	74	111	53	73	20	17	5	42
Classical Typing Sanger	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.	ref.
* **Fluidigm**	192	N.A	N.A	N.A	N.A	N.A	N.A	N.A	N.A
** **Roche 454 Amplicon Seq.**	173	29	63	26	-	22	-	19	19
*** **PNAS Amplicon Seq.**	79	41	88	35					41

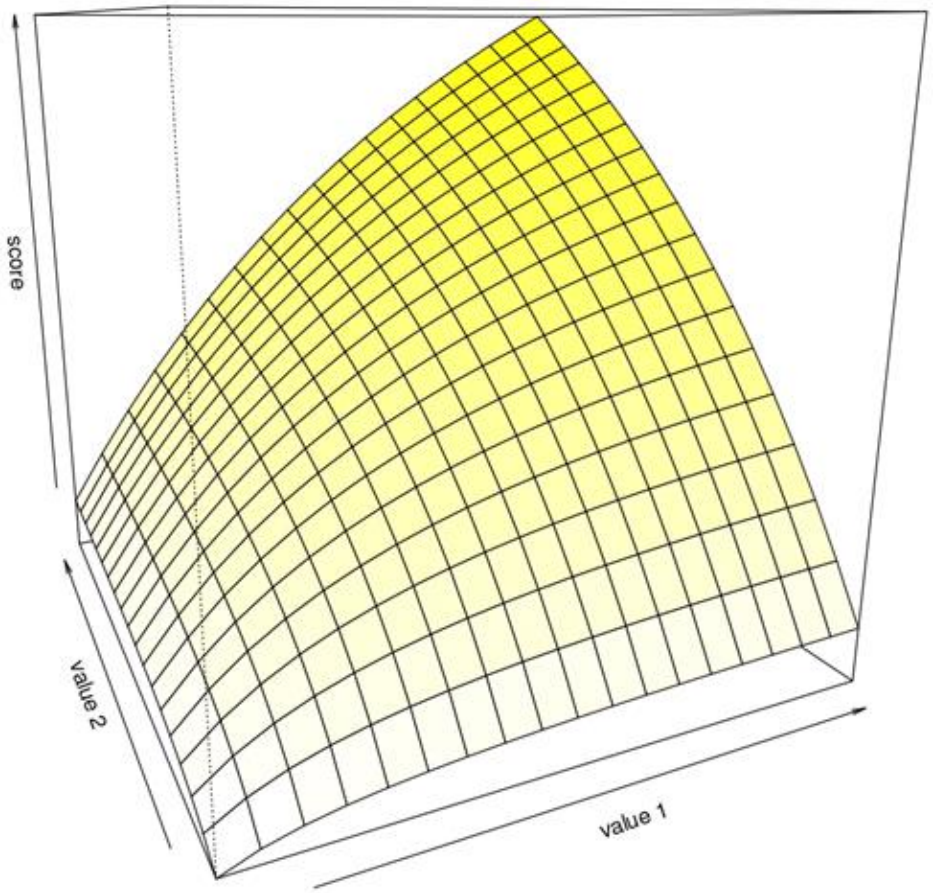
* Moonsamy PV *et al.* 2013. High throughput HLA genotyping using 454 sequencing and the Fluidigm Access Array™ System for simplified amplicon library preparation. Tissue Antigens. 81(3):141-9.

** Martin Danzer *et al.* 2013. Rapid, scalable and highly automated HLA genotyping using next-generation sequencing: a transition from research to diagnostics. BMC Genomics 2013, 14:221

*** Chunlin Wang *et al.* 2012. High-throughput, high-fidelity HLA genotyping with deep sequencing. PNAS 109(22):8676-81

HLA allele calling

P: parameters characterizing the mapping
w: weighting of the different parameters
P = { asm,req,msl,mppr,auc }
w= { 0.5, 1, 0.1, 0.1, 1 }



$$H = \frac{\left(\sum_{k=1}^{5} \frac{1}{W_k}\right) * \prod_{k=1}^{5} P_k}{\left(\sum_{k=1}^{5} \frac{1}{W_k} P_k\right)}$$

$$H = \frac{(2+1+10+10+1) * asm * req * msl * mppr * auc}{2 * asm + 1 * req + 10 * msl + 10 * mppr + 1 * auc}$$

HLA calling success rate

Table 1 Benchmark statistics for allele calls that could be achieved in a fully automated manner at three-field resolution (note that some of the reference samples only had two-field resolution, Supplementary Table S2)

Locus	Genotypes available	Different alleles	Allele call rate		Mean alternatives/sample
			Exact	Including alternatives	
A	341	114	**0.98**	**0.99**	2.21
B	344	122	**0.99**	**1.00**	1.03
C	285	53	**0.98**	**0.99**	1.58
DRB1	307	80	**0.98**	**0.98**	1.08
DQA1	152	20	**1.00**	**1.00**	0.06
DQB1	264	18	**0.99**	**0.99**	0.06
DPA1	87	5	**0.98**	**0.98**	0.00
DPB1	209	40	**0.97**	**0.97**	0.57

Columns - *Locus*: Lists the typed classical gene locus. *Genotypes available*: Shows the number of samples for which reference genotypes were available. Only alleles for which a reference call was available were used for calculating the call rate. *Different alleles*: Lists the number of different alleles within the reference data set for each locus. The automated calling generates two probable allele calls per diploid sample. *Allele calling exact*: Shows the validated allele call rate that was achieved by the automated calling. *Allele calling including alternatives*: Shows the call rate when including possible alternative alleles (ambiguities). *Mean alternatives/sample*: Shows the average number of alternative allele calls (ambiguities) per sample.

Wittig *et al.*, Nucleic Acids Research 2015

Evolution of HLAssign

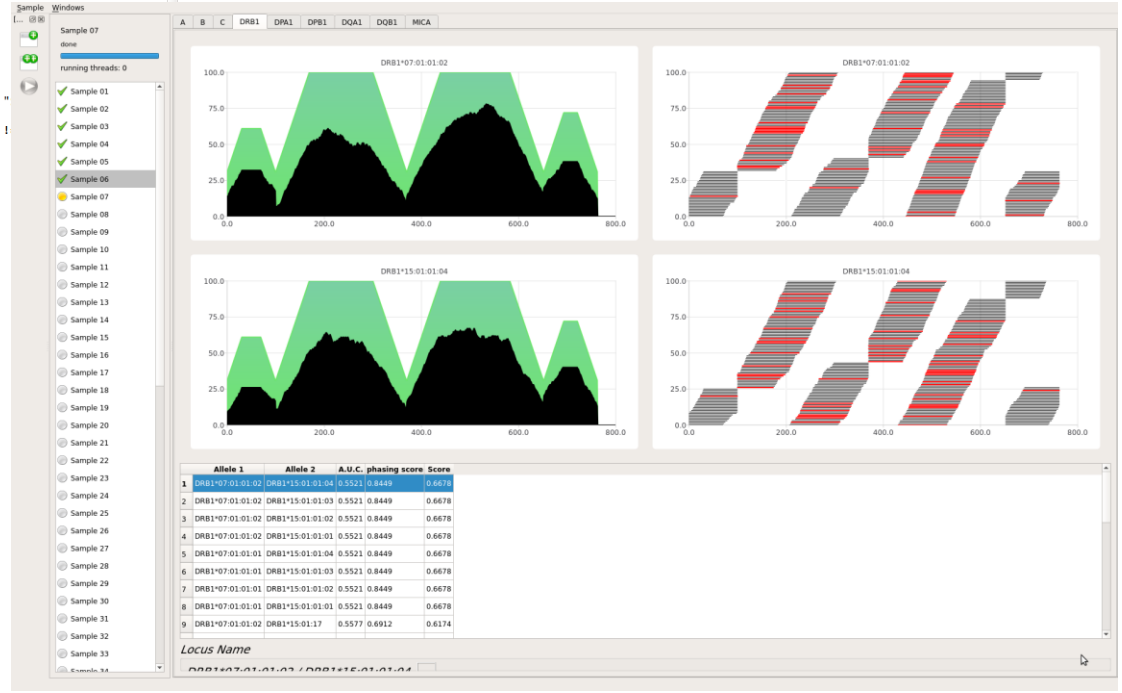
```
mwittig@MW-X1C: ~/ramDisk
mwittig@MW-X1C:~/ramDisk$ cat dummy.txt+
for i in *.fastq.gz
do
zcat $i | awk '{if( (NR+3)%4==0 ) {split($0, a, " "); print">"a[1]"/"NR} if( (NR+2)%4==0 ) {print(substr($0,1,100))
}}' > ${TMPDIR}/reads_to_map.fasta
done

# create required directories
mkdir -p ${THERESULTDIR}/pics
cd ${TMPDIR}
mkdir ${TMPDIR}/pics
mkdir ${TMPDIR}/scratch

# run the analysis
nextcallhla --cDNA ${THEALLELEDIR}/THELOCUS_nuc.txt --gDNA ${THEALLELEDIR}/THELOCUS_gen.txt --scratch ${TMPDIR}/scr
atch --out ${TMPDIR}/result.txt --outPic ${TMPDIR}/pics --bias-alleles /foo/bar/incomplete_alleles.txt --amb-table
/foo/bar/ambiguity_table.txt --reads ${TMPDIR}/reads_to_map.fasta --read-length 100

cp ${TMPDIR}/pics/* ${THERESULTDIR}/pics/
cp ${TMPDIR}/result.txt ${THERESULTDIR}/THELOCUS_calls_unsorted.txt

# extract calls and store in a file called the_calls.txt
for j in A B C DRB1 DPA DPB DQA DQB
do
   THECALL=`grep -A 1 "[Result]" ${THERESULTDIR}/${j}_calls_unsorted.txt | tail -1`
   echo $i $j $THECALL | awk 'BEGIN {IFS=" "}; END {if( NF <= 7 ){print $0} else {print $1" "
}}' >> the_calls.txt
   for k in `sort -k 8,8n ${THERESULTDIR}/${j}_calls_unsorted.txt | awk '{if( NF == 8 && $8 !
d -5 | cut -f 1 | sed -re 's/\*/-/g'`
   do
```



Wittig *et al.*, Nucleic Acids Research 2015

Graphical User Interface (GUI)

	A		B		Equality	Spec. Map	AUC	Length	Score
1	C*03:03:01	C	C*07:02:01:01	C	0.94	512/482	59.71	1096/1096	▲ 91.92
2	C*03:136	ND	C*07:02:01:01	C	0.96	183/176	57.04	546/1096	28.88
3	C*03:39	ND	C*07:02:01:01	C	▲ 0.99	172/174	56.40	546/1096	28.14
4	C*03:39	ND	C*03:136	ND	0.92	102/111	54.34	546/546	12.12
5	C*03:03:01	C	C*03:39	ND	0.92	62/57	58.81	1096/546	9.68
6	C*03:03:01	C	C*03:136	ND	0.92	49/53	59.45	1096/546	8.49
7	C*07:02:01:01	C	C*07:02:01:01	C	0.00	0/0	57.90	1096/1096	0.00
8	C*03:136	ND	C*03:136	ND	0.00	0/0	55.30	546/546	0.00
9	C*03:39	ND	C*03:39	ND	0.00	0/0	53.37	546/546	0.00
10	C*03:03:01	C	C*03:03:01	C	0.00	0/0	▲ 61.52	1096/1096	0.00

Koch *et al.*, in preparation

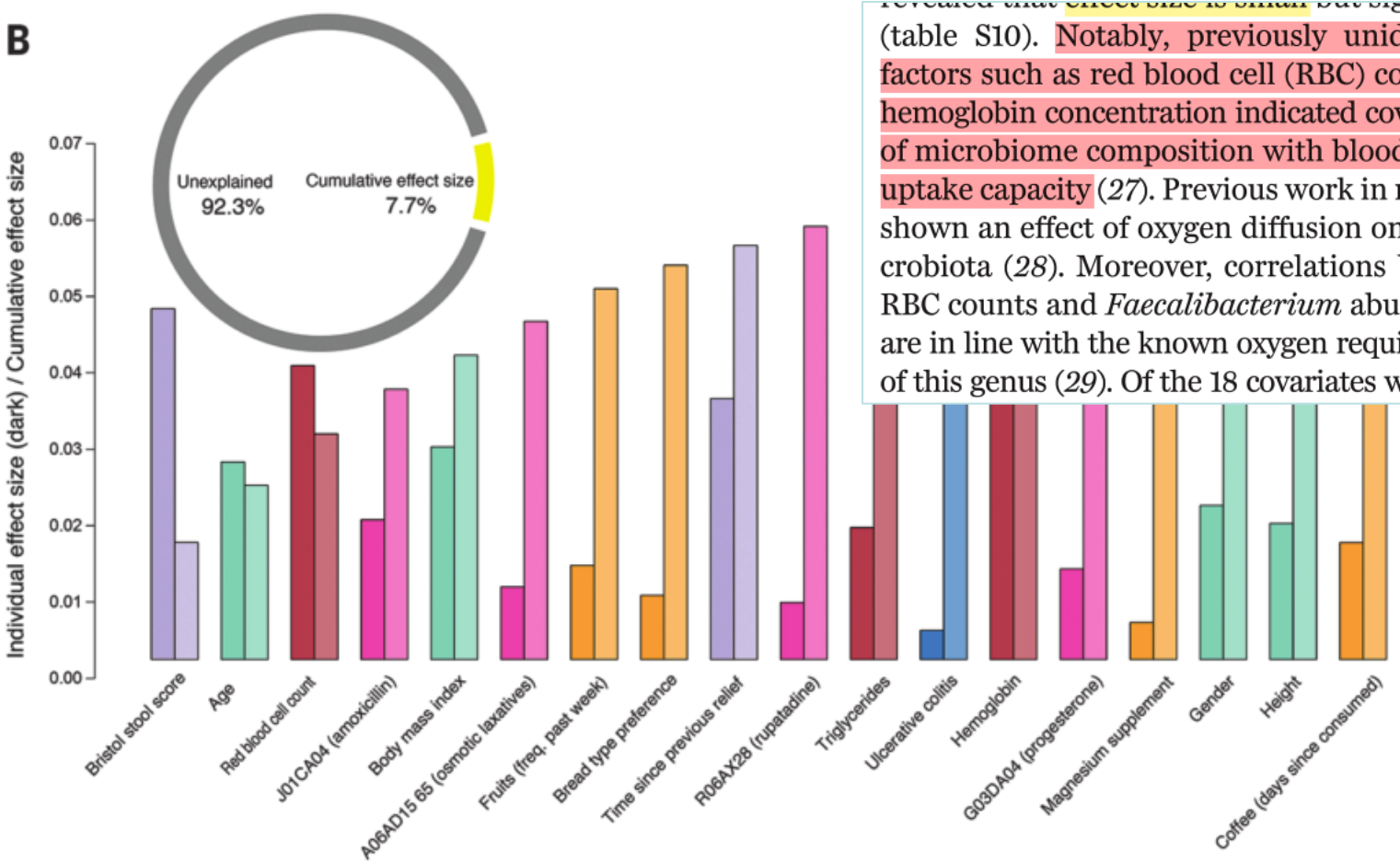
Prices

- Luminex – 100 € per test, needs Luminex-machine, throughput?
- Sanger-based technologies - €100-500 per test
- Service-only competition: DKMS (50 € per test), Univ. Munich (50 € per test); Histogenetics LLC
- http://www.illumina.com/clinical/hla-sequencing.html
  List price 300 € (long-range PCR)
- http://www.gendx.com
  List price 120 €/sample, up to 384 samples on MiSeq
  Hi Throughput: HiSeq 2500 oder NextSeq, 2 x 384 samples
- http://www.omixon.com/holotype-hla/
  Software from Omixon, Kit by CHOP, 192/240 samples per MiSeq run (7/5 loci)
  Hi Throughput price (5000 samples): ca. 55 €/sample regardless whether 5 or 7 loci

Each of these tests have their own problems apart from the price. Problems that can lead to fatal errors. (30%-50% of stem cell transplant recipients die from graft-versus-host disease!)

What about blood group antigens?

Blood groups vs. Microbiome?



(table S10). Notably, previously unidentified factors such as red blood cell (RBC) count and hemoglobin concentration indicated covariation of microbiome composition with blood oxygen uptake capacity (*27*). Previous work in mice has shown an effect of oxygen diffusion on the microbiota (*28*). Moreover, correlations between RBC counts and *Faecalibacterium* abundances are in line with the known oxygen requirements of this genus (*29*). Of the 18 covariates with non-

Falony *et al.*, Science 2016

Poor reference databases...

Table 1. *MNS* alleles with single nucleotide polymorphisms that generate blood group antigens.
A. *GYPA*: Reference allele *MNS*01* encodes M, En^a, ENEH, ENEP, ENAV, ENDA, ENEV.
Note: In most cases, the nucleotide changes also can occur on an N allele; these nucleotide changes are not given.

Phenotype	Allele name	Nucleotide change	Intron/ Exon	Amino acid change	Comments
MNS:1 or M+	*GYPA*01* *GYPA*M*	59C; 71G; 72T	2	Ser20, Gly24	
MNS:2 N+	*GYPA*02* *GYPA*N*	59C>T; 71G>A; 72T>G	2	Ser20Leu, Gly24Glu	
MNS:1,-2,8† M^c+	*GYPA*08* *GYP.Mc*	71G>A,72T>G	2	Gly24Glu	
MNS:7,9,–40 Vw+	*GYPA*09* *GYPA*Vw*	140C>T	3	Thr47Met	
MNS:-1,-2,11 M^g+	*GYPA*11* *GYPA*Mg*	68C>A	2	Thr23Asn	
MNS:12 Vr+	*GYPA*12* *GYPA*Vr*	197C>A	3	Ser66Tyr	
MNS:14 Mt(a+)	*GYPA*14* *GYPA*Mta*	230C>T	3	Thr77Ile	
MNS:16 Ri(a+);	*GYPA*16* *GYPA*Ria*	226G>A	3	Glu76Lys	
MNS:18 Ny(a+)	*GYPA*18*	138T>A	3	Gln46Glu	
MNS:7,19,–40 Hut+	*GYPA*19* *GYPA*Hut*	140C>A	3	Thr47Lys	

from ISBTweb.org

Imputation benchmark

Coa/Cob	5.4%	Jka/Jkb	48.3%	RhCc	31.8%
Cw	3.5%	KEL Kk	3.8%	RhD	18.2%
Dia/Dib	0.1%	KEL Kpa/Kpb	1.0%	RhEe	13.9%
Doa/Dob	37.5%	KEL Jsa/Jsb	0.0%	Sc1/Sc2	0.0%
Fya/Fyb	44.1%	Lua/Lub	3.2%	Vel	0.1%
HPA1a/HPA1b	16.5%	LWa/LWb	1.0%	Wra/Wrb	0.0%
HPA5a/HPA5b	8.3%	MNS MN	45.4%	Yta/Ytb	4.6%
Ina/Inb	0.0%	MNS Ss	31.9%		

NGS experiment

Chromosome	Gene Name	Blood Group
9	*ABO*	ABO blood group
7	*AQP1*	Colton blood group
17	*SLC4A1*	Diego blood group
12	*ART4*	Dombrock blood group
1	*CD55*	Duffy blood group
1	*ACKR1*	Duffy blood group
19	*FUT1*	galactoside 2-alpha-L-fucosyltransferase, H blood group
2	*GYPC*	Gerbich blood group
9	*AQP3*	Gill blood group
X	*GATA1*	globin transcription factor 1
3	*B3GALNT1*	globoside blood group
6	*GCNT2*	I blood group
11	*CD44*	Indian blood group
15	*SEMA7A*	John Milton Hagen blood group
4	*ABCG2*	Junior blood group
7	*KEL*	Kell blood group
18	*SLC14A1*	Kidd blood group
1	*CR1*	Knops blood group
19	*KLF1*	Kruppel-like factor 1
19	*ICAM4*	Landsteiner-Wiener blood group
2	*ABCB6*	Langereis blood group
19	*BCAM*	Lutheran blood group
4	*GYPA*	MNS blood group
4	*GYPB*	MNS blood group
4	*GYPE*	MNS blood group
19	*BSG*	Ok blood group
22	*A4GALT*	Pk antigen of blood histogroup P
11	*CD151*	Raph blood group
1	*RHCE*	Rh blood group, CcEe antigens
1	*RHCE*	Rh blood group, CcEe antigens
1	*RHCE*	Rh blood group, CcEe antigens
1	*RHD*	Rh blood group, D antigen
1	*RHD*	Rh blood group, D antigen
6	*RHAG*	Rh-associated glycoprotein, ammonium transporter Rh type A
1	*ERMAP*	Scianna blood group
X	*XK*	X-linked Kx blood group
X	*XG*	Xg blood group
7	*ACHE*	Yt blood group

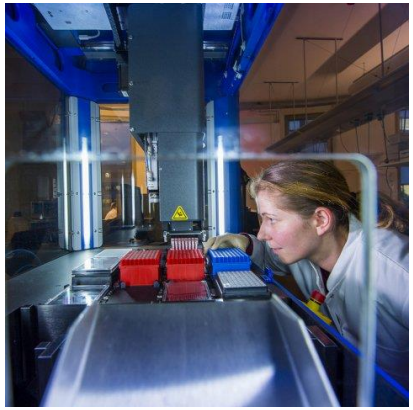
BFX

Christoph Gassner

Michael Wittig

Acknowledgements

BLUTSPENDE ZÜRICH



a.franke@mucosa.de