

Next generation sequencing for immunogenetics

Christian Gabriel, M.D.,
Red Cross Transfusion Service
of Upper Austria

**Scientific Session of the BBTS
annual conference in Leeds**
25-09-2014





Purpose of Next Generation Sequencing (NGS) and some cautionary notes

- NGS-systems are designed primarily for biology, metagenomics and plant genomics:
 - High genomic complexity
 - Long ranges of repetitive sequences
 - Multiple overlaying genomes
- NGS is a testing strategy in which:
 - Brute force is applied to generate high volumes reads
 - Massive parallel sequencing is a prerequisite
 - Bioinformatics might distort the data analysed

Not designed primarily for medical applications



AUSTRIAN RED CROSS

UPPER AUSTRIA

Most developed medical areas for next generation sequencing

- Immunogenetics
- Virology:
 - ultradeep sequencing
 - susceptibility
- Immunology:
 - IgH and TCR
- Oncology:
 - Multiple polymorphisms
 - Regions of resistance polymorphisms



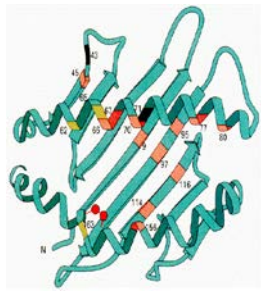
AUSTRIAN RED CROSS

UPPER AUSTRIA

Why?

- Higher sensitivity than Sanger
- Finding small populations, but high relevance for the patient's fate (f.i. HIV resistance)
- Detecting multiple polymorphisms
deletions, inversions at once:
 - Compound mutations
 - In alleles
 - Haplotypes



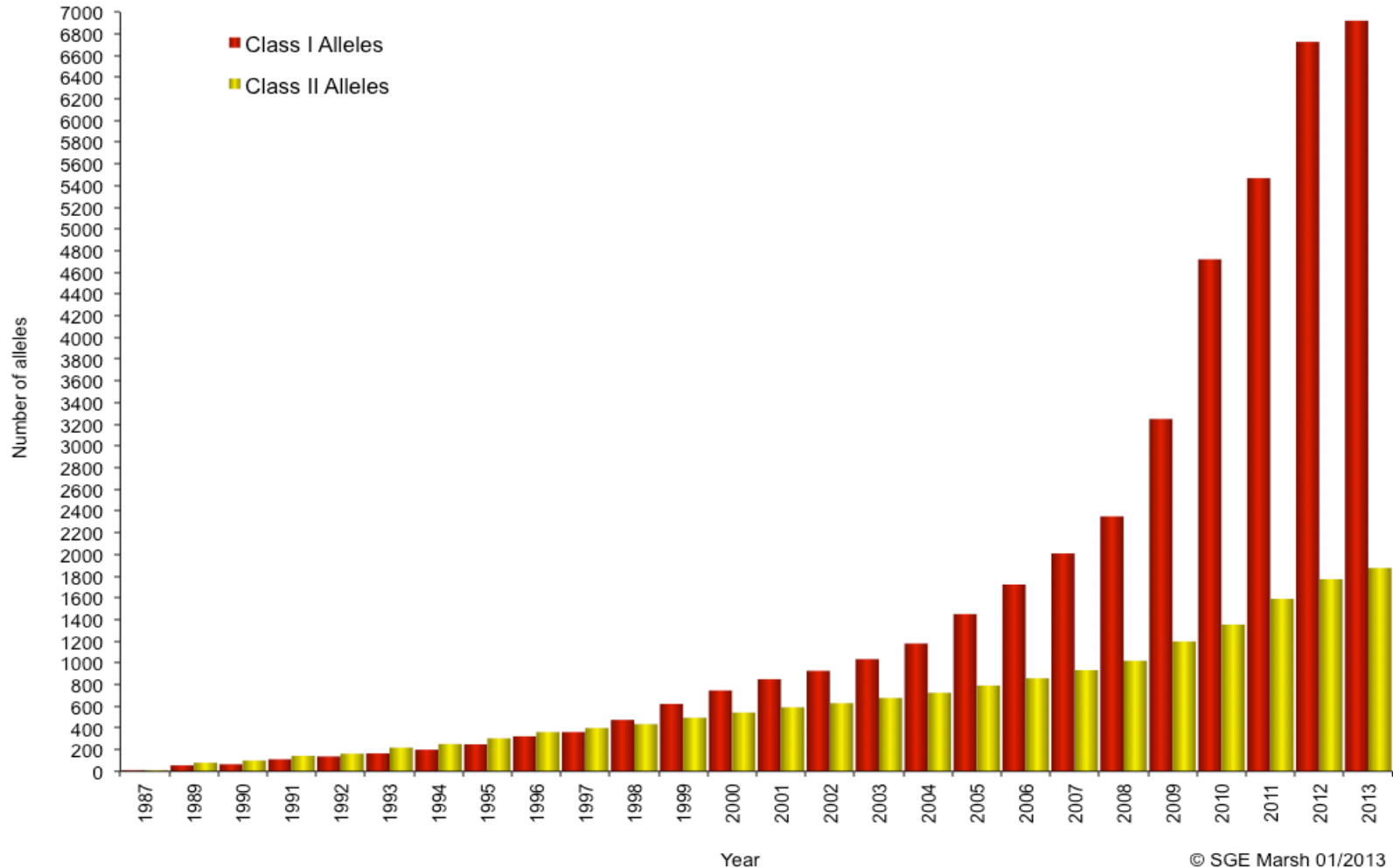


New alleles – every third day



AUSTRIAN RED CROSS

UPPER AUSTRIA

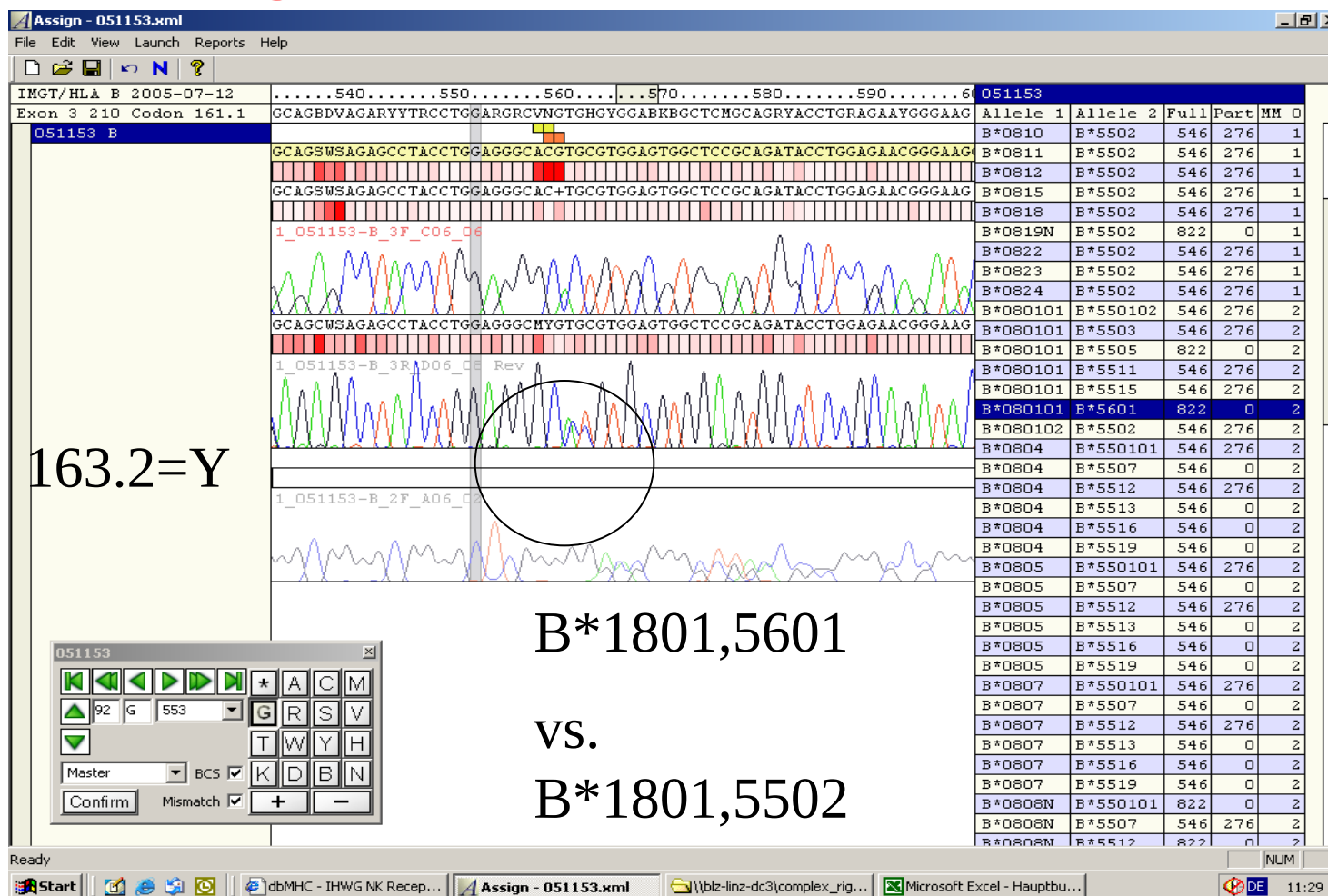




AUSTRIAN RED CROSS

UPPER AUSTRIA

Sequencing (sequencing-based typing, SBT)

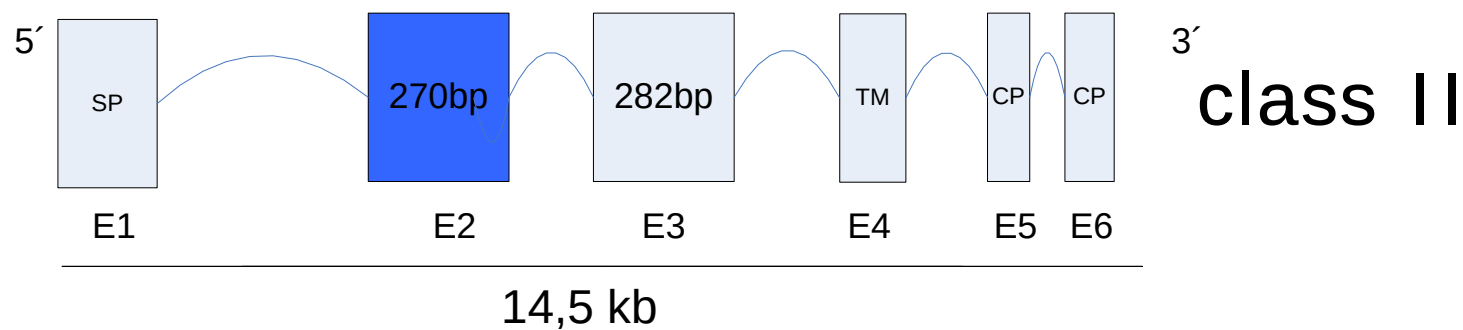
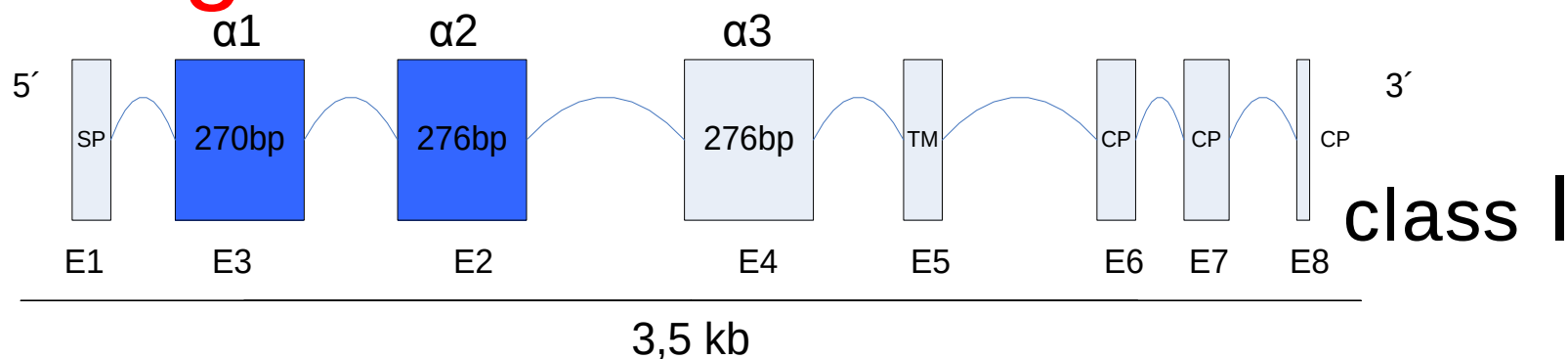


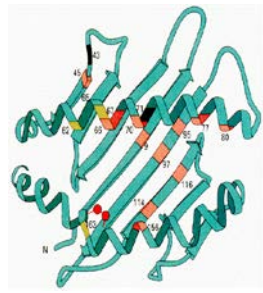


AUSTRIAN RED CROSS

UPPER AUSTRIA

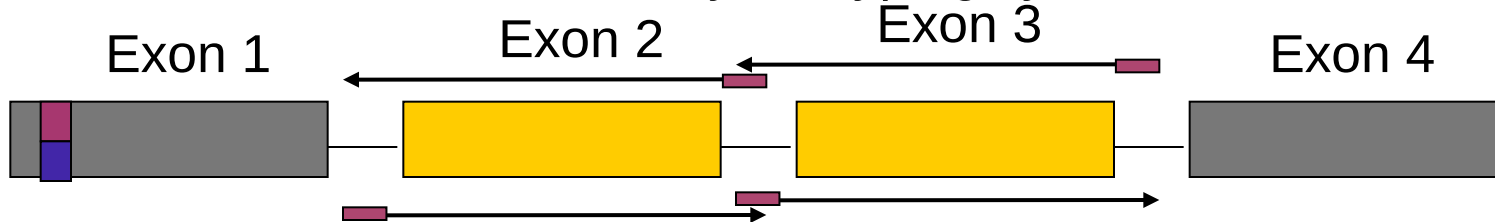
Regions of interest





Allele ambiguity

outlier mutations: allele ambiguity results when polymorphisms that distinguish alleles fall outside of the regions examined by the typing system



Polymorphic positions



Core heterozygous sequence data

example: HLA-B

B*0702, 4402

B*0702, 4419N

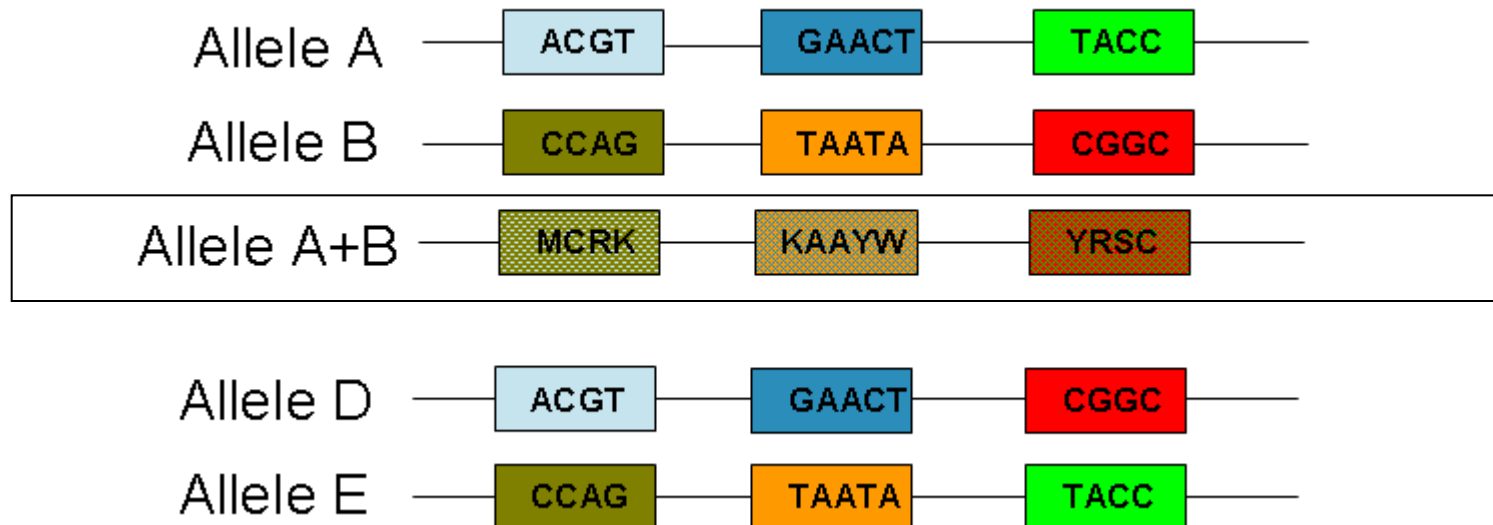
Genotype ambiguities



AUSTRIAN RED CROSS

UPPER AUSTRIA

Results from an inability to establish **phase** between closely linked polymorphisms identified by the typing system



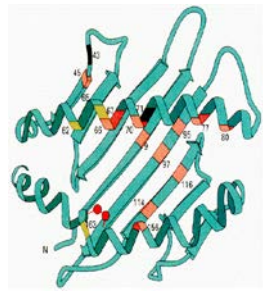
$$A + B = D + E$$

example: HLA-B

B*0702, 4402

B*0720, 4416

B*0724, 4421



cis/trans Problems

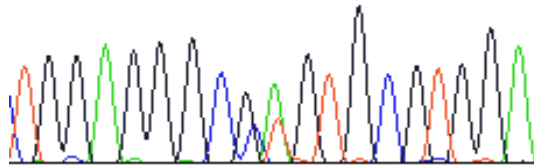


AUSTRIAN RED CROSS

UPPER AUSTRIA



TGGAGGGCSMGTGCGTGGGA



TGGAGGGCSMGTGCGTGGGA

Number of possible
linkages = 2^n

$n=2$; 4 combinations

$n=4$; 16 combinations

S = G und C

M = A und T

-----SM-----

-----GA-----

-----CT-----

-----GT-----

-----CA-----

IUB Code	K	S	W	M	Y	R
Bases	G,T	G,C	A,T	A,C	C,T	A,G



without
ambiguities?

Hey dumbheads!
How can we get faster and
more reliable results?



Strategies

- Amplicon-based:
 - Patients and donors prior to transplantation VHRT
 - For registries – IRT („quick and dirty“)
- Genome-based:
 - Transcript-Sequencing

Strategies



AUSTRIAN RED CROSS

UPPER AUSTRIA

1

Exon-based typing



flexibility: high
capacity: gentle
full allele drop out: rare
DNA quality: not critical
Scability: medium/ high/ ultra high
library prep: complex
automation: yes

Shorter amplicons allow the complete coverage in reverse and forward direction !



Gabriel C et al: **Rapid high-throughput human leukocyte antigen typing by massively parallel pyrosequencing for high-resolution allele identification.** *Hum Immunol* 2009, **70**:960-964.

? Ambiguities
library prep

Bentley G et al: **High-resolution, high-throughput HLA genotyping by next-generation sequencing.** *Tissue Antigens* 2009, **74**:393-403.

Erlich RL et al: **Next-generation sequencing for HLA typing of class I loci.** *BMC Genomics* 2011, **12**:42.

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

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

Grumbt B et al: **Diagnostic applications of next generation sequencing in immunogenetics and molecular oncology.** *Transfus Med Hemother.* 2013 Jun

Trachtenberg EA et al: **Next-generation HLA sequencing using the 454 GS FLX system.** *Methods Mol Biol.* 2013

Holcomb CL et al: **Multi-site study using high-resolution HLA genotyping by next generation sequencing.** *Tissue Antigens* 2011, **77**:206-217.



AUSTRIAN RED CROSS

UPPER AUSTRIA

Amplicons

- HLA -A, -B, -C, -DP, -DQ, -DR
- class I:
 - A: exons 2,3,4
 - B: exons 1,2,3,4
 - C: exons 1,2,3,4,6,7
- class II:
 - DP: exon 2
 - DQ: exons 2,3
 - DR: exons 2,3

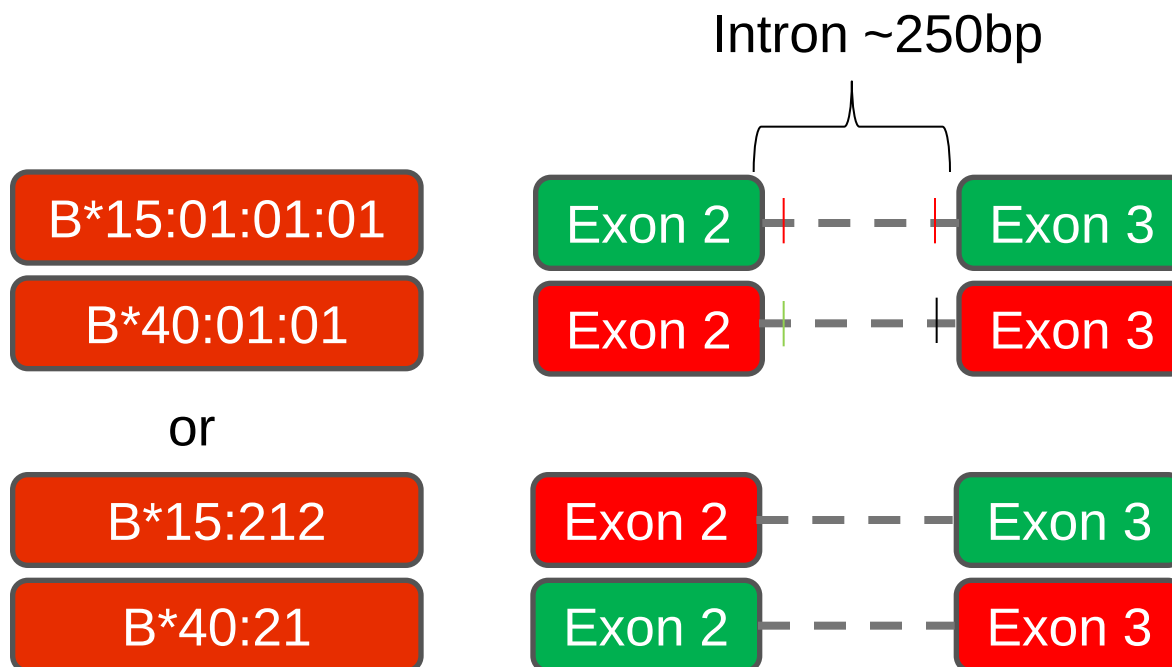


AUSTRIAN RED CROSS

UPPER AUSTRIA

Limitations

Ambiguities between Exons



Strategies



AUSTRIAN RED CROSS

UPPER AUSTRIA

2 cDNA-based typing



flexibility: medium
capacity: ideal
full allele drop out: possible
RNA quality: very critical
Scability: : medium/ high
library prep: reduced complexity
automation: yes

? N- alleles
RNA quality

Lank SM et al: **A novel single cDNA amplicon pyrosequencing method for high-throughput, cost-effective sequence-based HLA class I genotyping.** *Hum Immunol* 2010, **71**:1011-1017.

Lank SM et al: **Ultra-high resolution HLA genotyping and allele discovery by highly multiplexed cDNA amplicon pyrosequencing.** *BMC Genomics* 2012, **13**:378.

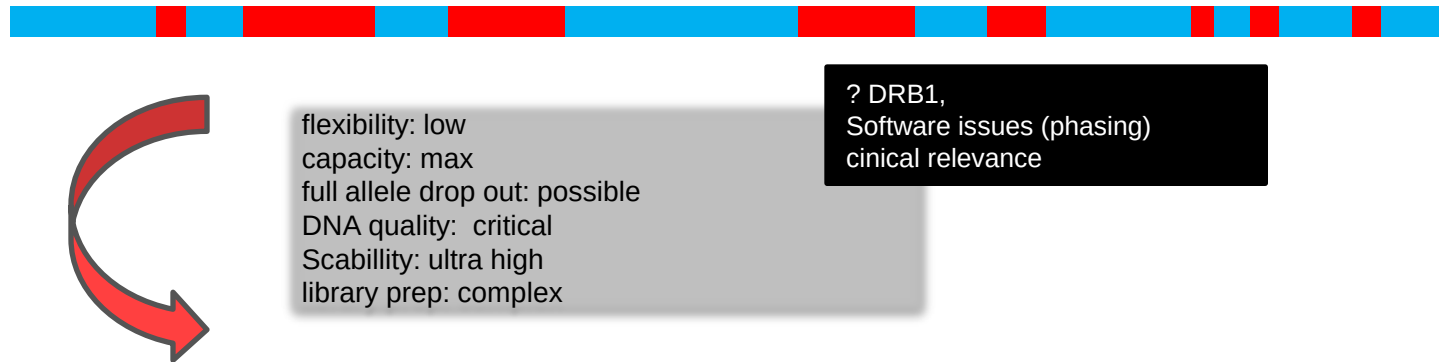
Strategies



AUSTRIAN RED CROSS

UPPER AUSTRIA

3 Full genomic typing



Lind C et al: **Next-generation sequencing: the solution for high-resolution, unambiguous human leukocyte antigen typing.** *Hum Immunol* 2010, **71**:1033-1042.

Wang C et al: **High-throughput, high-fidelity HLA genotyping with deep sequencing.** *Proc Natl Acad Sci U S A* 2012, **109**:8676-8681.

Shiina T et al: **Super high resolution for single molecule-sequence-based typing of classical HLA loci at the 8-digit level using next generation sequencers.** *Tissue Antigens* 2012, **80**:305-316.

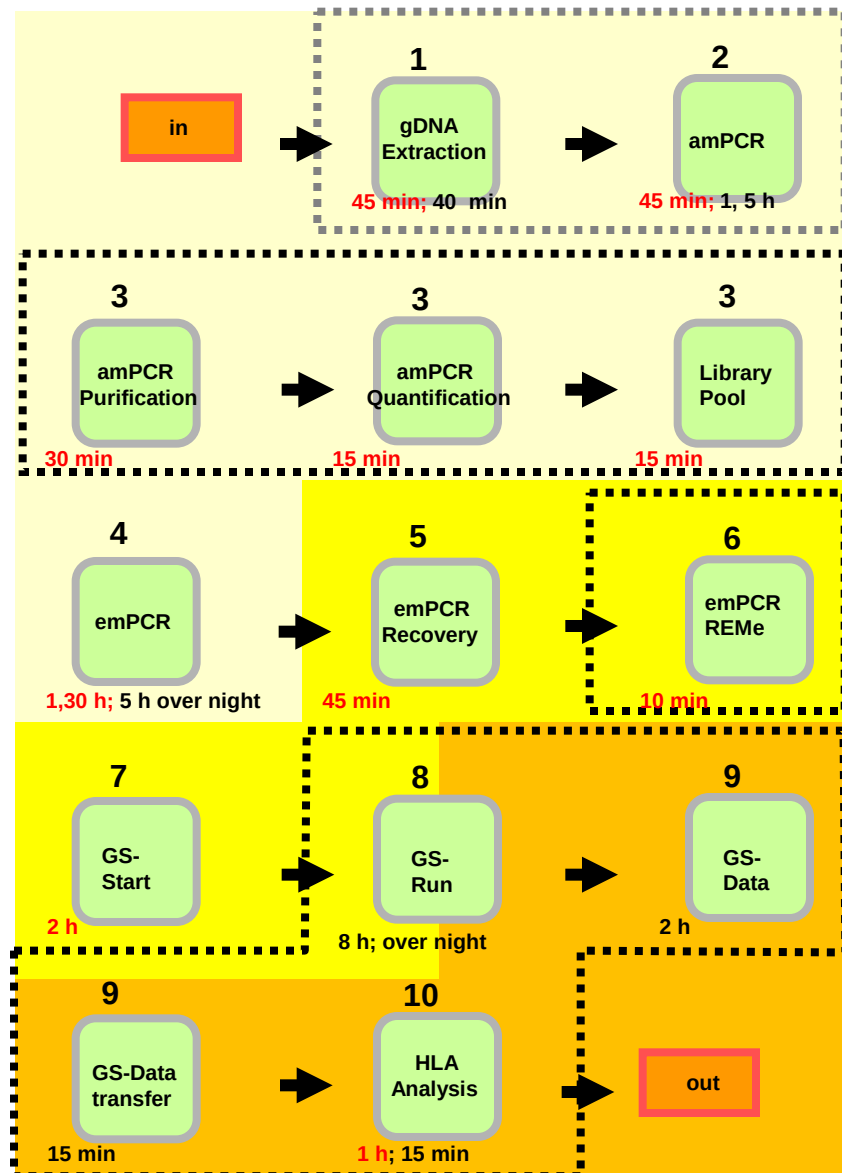
Hosomichi K et al: **Phase-defined complete sequencing of the HLA genes by next-generation sequencing.** *BMC Genomics* 2013, **28**:14:355.

Ozaki Y. et al: **HLA-DRB1, -DRB3, -DRB4 and -DRB5 genotyping at a super-high resolution level by long range PCR and high-throughput sequencing.** *Tissue Antigens* 2014



AUSTRIAN RED CROSS

UPPER AUSTRIA



Lab 1 and 2
DAY 1

MagNa 96
STARlet

Auto Block 1

Lab 3

STAR

Auto Block 2

Lab 4 and 3
DAY 2

STARlet
REMe

Auto Block 3

Lab 5 and 6
DAY 3

Bioinformatics

Auto Block 4

Lab 6 and Office
DAY 3

~ 25-26 h

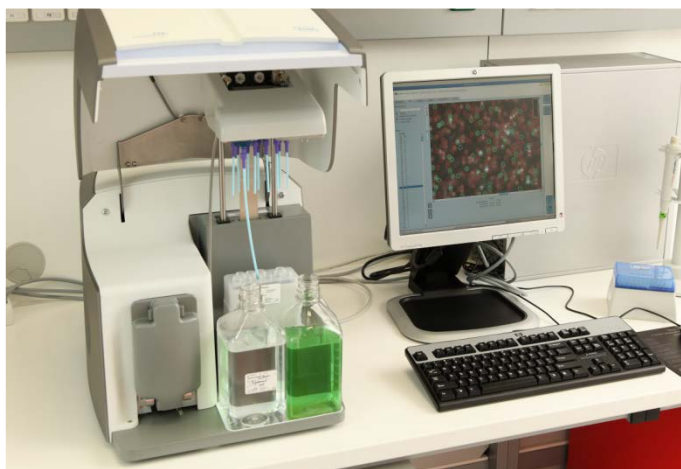


AUSTRIAN RED CROSS

UPPER AUSTRIA



**Hamilton STAR
Hamilton STARlet
454 REM e**



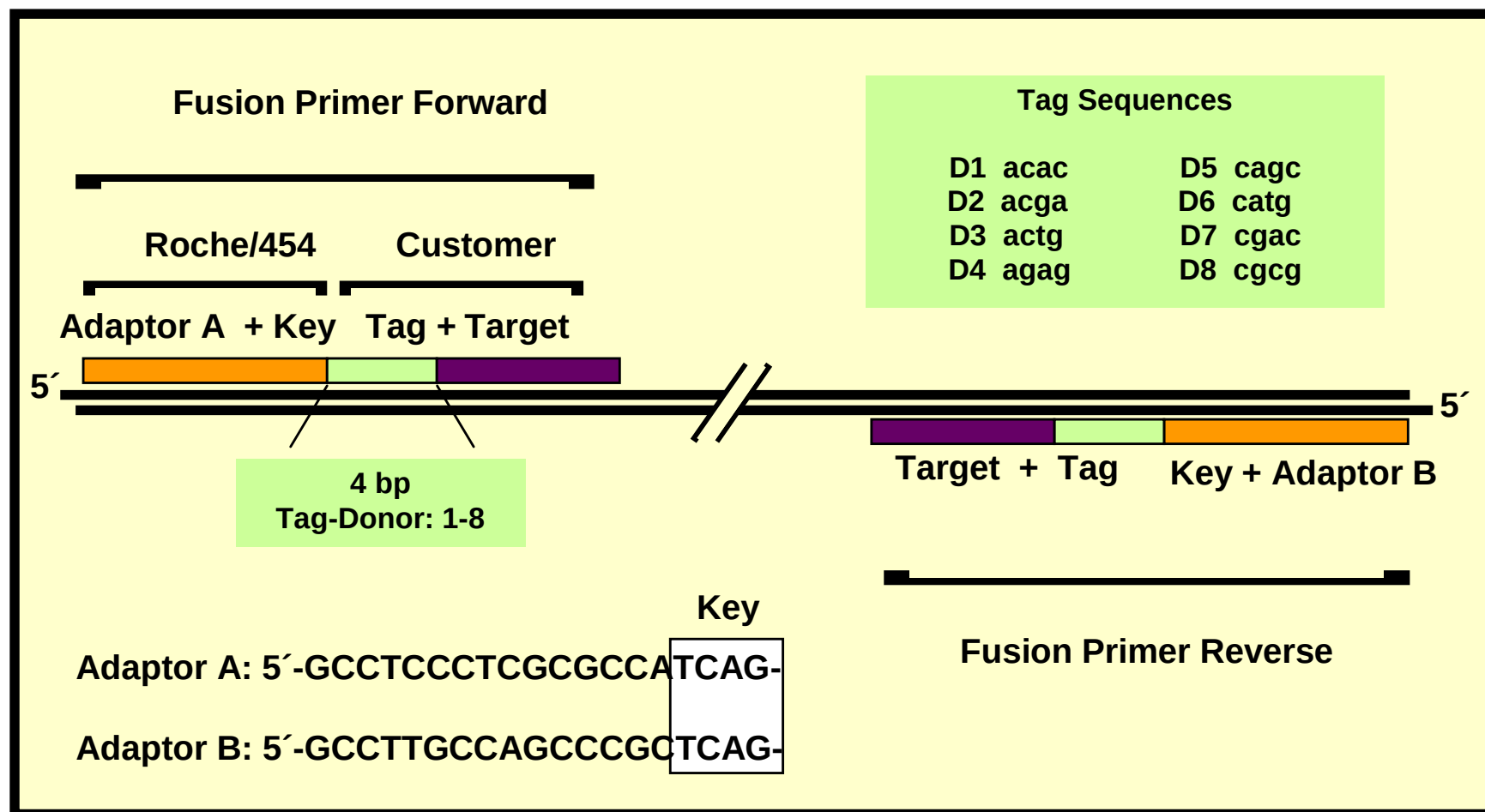
**GS Junior
GS FLX**



AUSTRIAN RED CROSS

UPPER AUSTRIA

Composition of primers with individual identification tags





AUSTRIAN RED CROSS

UPPER AUSTRIA

DEBUG HLA-B Ultra High Resolution

2.3 pipeline.xml reg2.xml reg3.xml **reg4.xml** x

Exon 1-2
>1.36-184 HLAB.E
>1.25-146 HLAB.E
<1.29-281
<1.14-256

Exon 3
<1.3-280
<1.2-226
>1.17-108
>1.12-119

Exon 4-5
>1.18-126 HLAB.E
>1.16-180 HLAB.E
<1.9-323 HLAB.
<1.24-228 HLAB

fB 2.25.1
Base 2914 (2914)
C A 2 1 R
gA
DPA1
DQA1
DQB1
DQB2
H
Sample 01* gB
gC
Sample 03
Sample 04
Sample 05
Sample 06
Sample 08
Sample 10
Sample 11
Sample 13

..2891.....2901... 2911.....2921.....
TCCTGHGGGCTCTGACCAGGTCTGTTTTRGTTCTACTCCA
Sample 01 gB
Start: 285 (285) Exon 1-2 1
Stop: 2350 (2350) Exon 4-5 497

Allele 1	Allele 2	MM 0	MM 2	MM 3	MM 4	Difference
B*0706	B*550201	0	0	0	0	Exon 3
B*0780	B*5537	0	0	1	0	
B*070201	B*550201	1	NC	NC	NC	
B*070501	B*550201	1	NC	NC	NC	
B*070501	B*550202	1	NC	NC	NC	
B*070501	B*550203	1	NC	NC	NC	
B*070501	B*550204	1	NC	NC	NC	
B*070501	B*550205	1	NC	NC	NC	
B*070501	B*5507	1	NC	NC	NC	
B*070501	B*5513	1	NC	NC	NC	
B*070501	B*5516	1	NC	NC	NC	
B*070501	B*5519	1	NC	NC	NC	
B*070501	B*5526	1	NC	NC	NC	
B*070501	B*5530	1	NC	NC	NC	
B*070501	B*5537	1	NC	NC	NC	
B*070501	B*5539	1	NC	NC	NC	
B*070502	B*550201	1	NC	NC	NC	
B*070503	B*550201	1	NC	NC	NC	
B*0706	B*550101	1	NC	NC	NC	
B*0706	B*550202	1	NC	NC	NC	
B*0706	B*550203	1	NC	NC	NC	
B*0706	B*550204	1	NC	NC	NC	
B*0706	B*550205	1	NC	NC	NC	

Sample 01

2914 972.1 0

A C G T * + -

No Offset Master ☒ BCS ☐ Edits ☐ MM ☒ Var



Sequencing errors

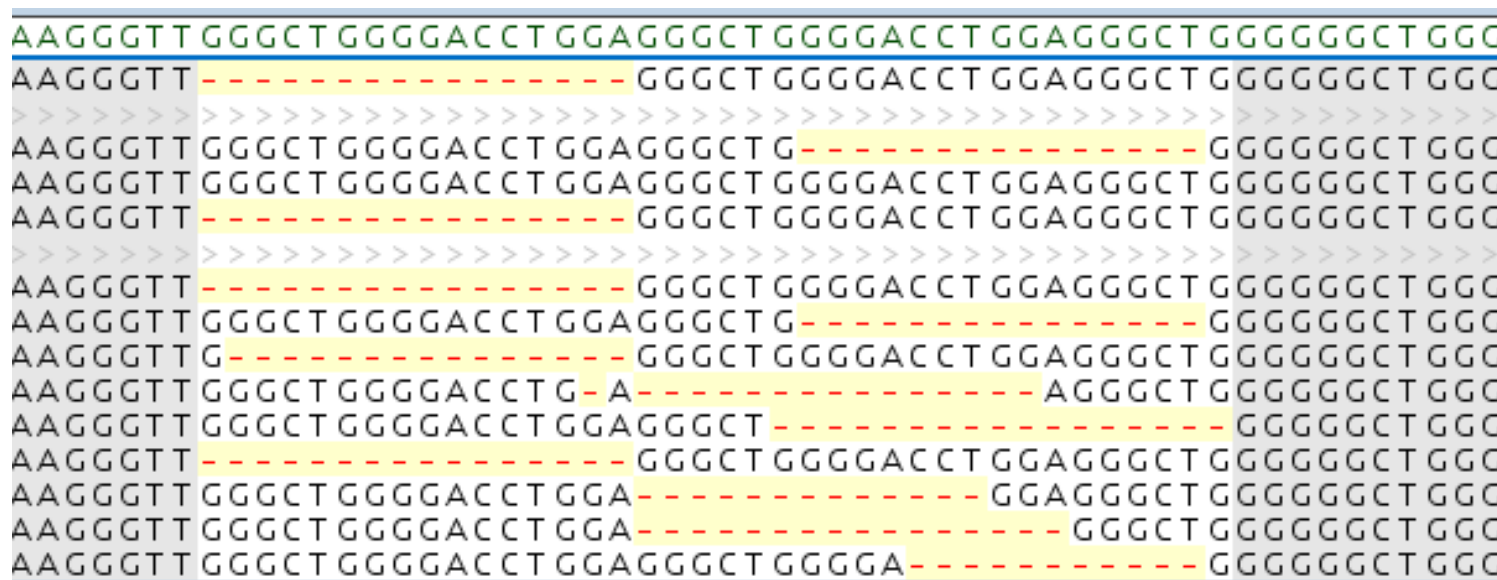
- not just homopolymers, also subsequent regions
- „difficult“ sequence pattern





Alignment problems

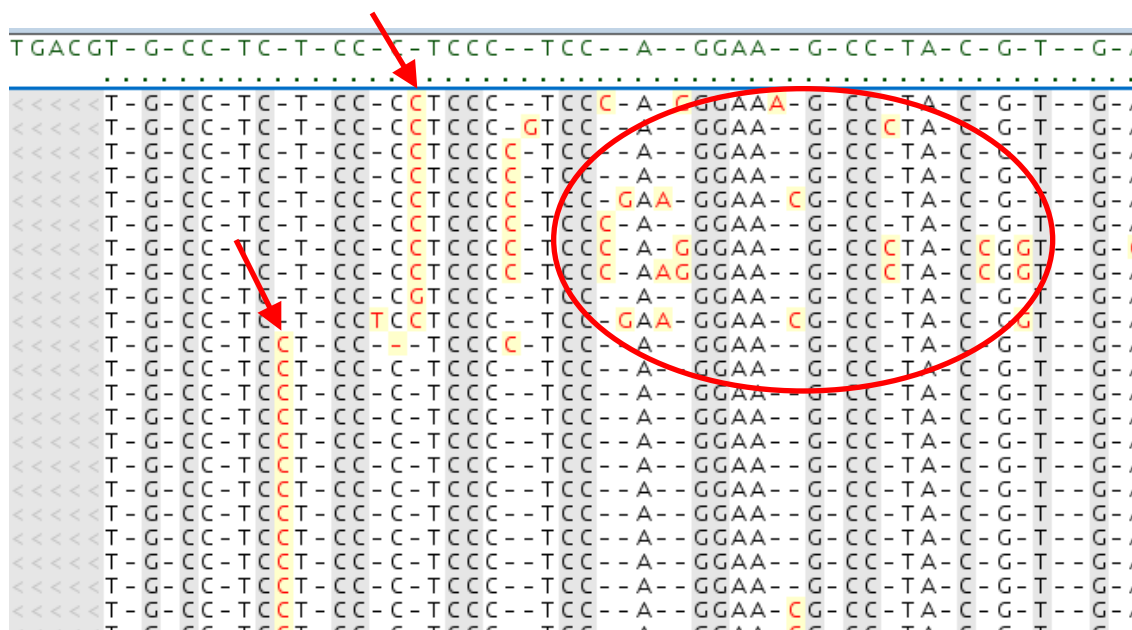
- alignments over multiple reads not consistent





Alignment problems

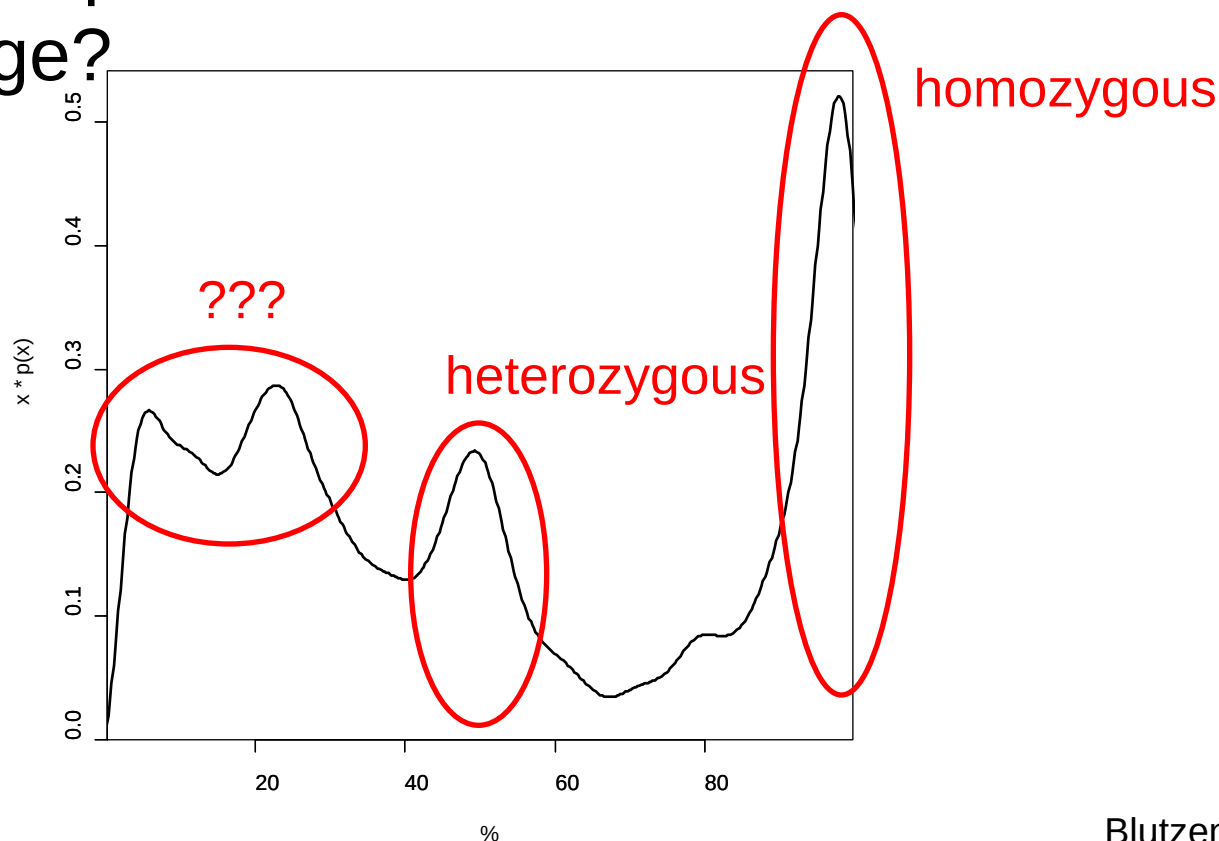
- existing variants are not reported
- many single sequencing errors





Variants

- how to interpret variants with low percentage?



HLA Study Participants *from Europe and U.S.*

Dr. **Wassmuth**, DKMS, Dresden

Dr. **Klein**, SEQIT Kaiserslautern, Germany

Dr. **Tyan**, Stanford University
(GSFLX at Stanford SGTC)

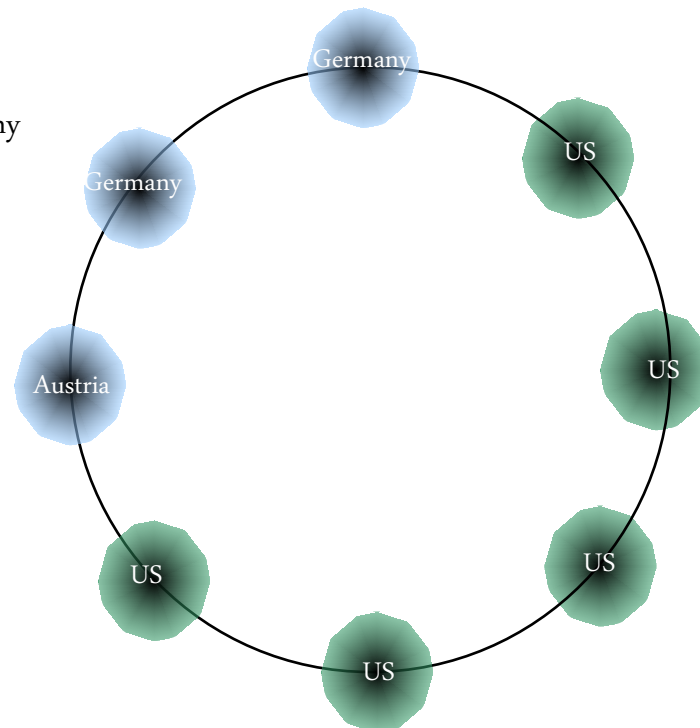
Dr. **Trachtenberg**, Childrens Hospital Oakland

Dr. **Monos**, Childrens Hospital Philadelphia (454
run and GS FLX rental)

Dr. **Erlich**, RMS, Pleasanton

Dr. **Simen**, 454, Branford

Dr. **Gabriel**, Blood Bank, Linz



The following slides of this study have been kindly provided by Ch Holcomb, Roche



AUSTRIAN RED CROSS

UPPER AUSTRIA

Genotype Assignment

Assignment for 95% of the 2240 genotypes examined
Causes for Failure to Assign Genotype

Cause	# of Genotypes Omitted	# of Sites Affected
New allele	24 (2%)	8
Plate seal failure in gPCR	16	1
Amplicon not added to pool	7	4
forward sequencing primer not annealed	4	1
Bias against B-2 reverse reads	6	1

3%

4/8 sites had no genotypes omitted due to procedural/technical issues

Summary of Agreement and Concordance

Locus	% Agreement	% Concordance with Known Variants		% Req'd Man. Ed.
		Genotype	Allele	
HLA-A	89	91	94	13
HLA-B	93	96	98	15
HLA-C	94	94	97	4
DPB1	99	100	100	3
DQA1	100	100	100	5
DQB1	99	100	100	4
DRB1	97	98	99	10
DRB3,4,5	97	98	99	13
Overall	96	97	98	8

Agreement = Identical ambiguity string obtained

Concordance = Reported genotype/allele in a limited ambiguity string matched submitted

Conclusions



AUSTRIAN RED CROSS

UPPER AUSTRIA

- 454 Life Sciences GS FLX System with Conexio ATF software can deliver high-throughput, high resolution sequencing: 20 samples can be typed in a single run for HLA-A, HLA-B, HLA-C, DPB1, DQA1, DQB1, DRB1 and DRB3, 4, 5.
- The genotyping system gives robust and reproducible results
 - 1280 genotypes considered
 - Assignment possible for 95%
 - 97% overall concordance with known genotypes
 - 98% overall concordance with known alleles
 - New alleles were correctly identified by the software as “No matches” with the database, and location of sequence variation is displayed

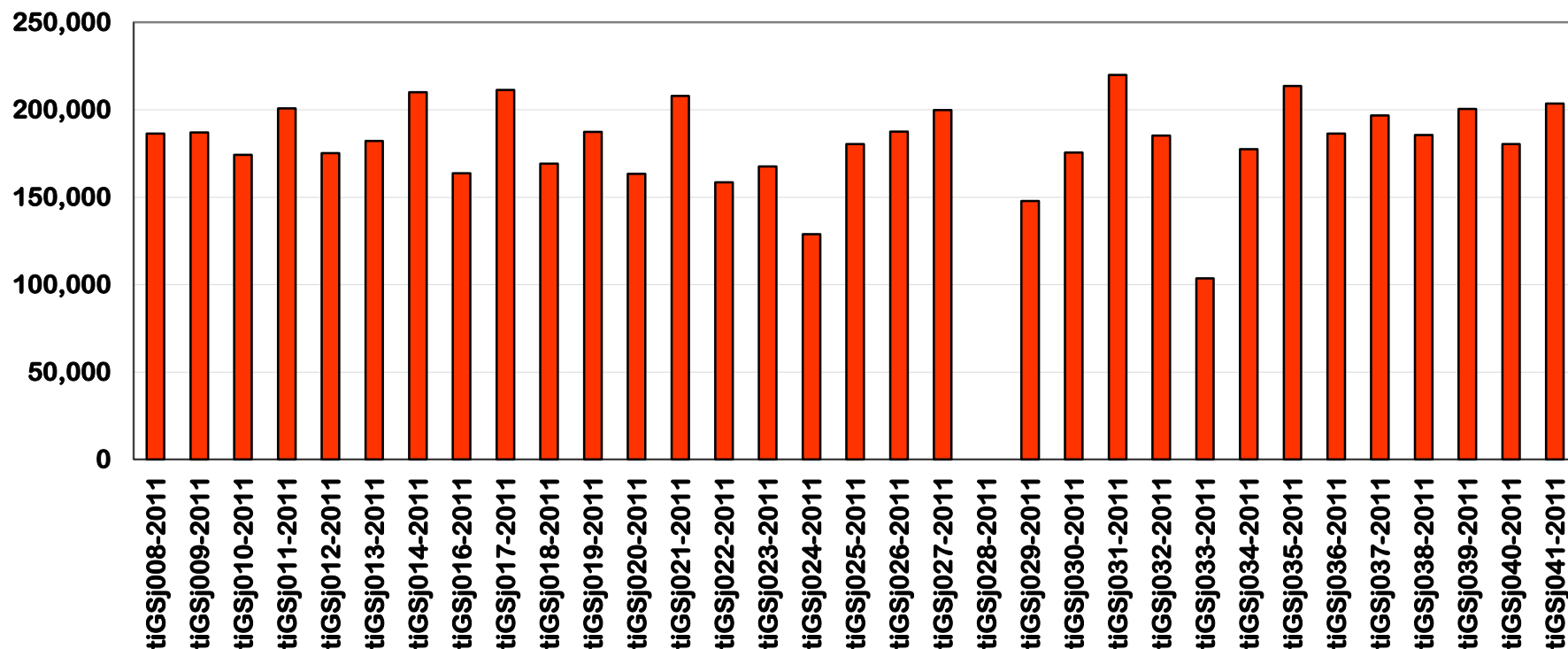


GS Junior Performance in our hands

AUSTRIAN RED CROSS

UPPER AUSTRIA

Raw Wells



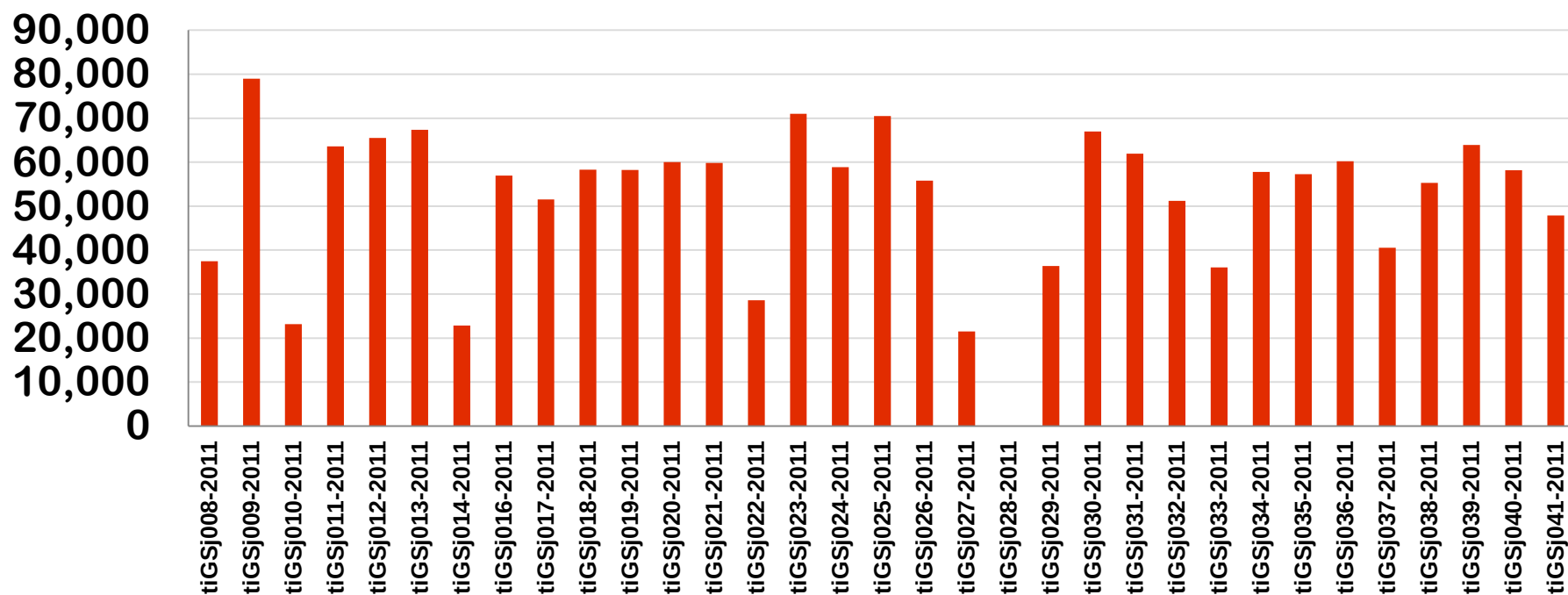


AUSTRIAN RED CROSS

UPPER AUSTRIA

GS Junior Performance in our hands

HQ Reads





AUSTRIAN RED CROSS

UPPER AUSTRIA

GS Junior Sequencing Performance - Validation Study.

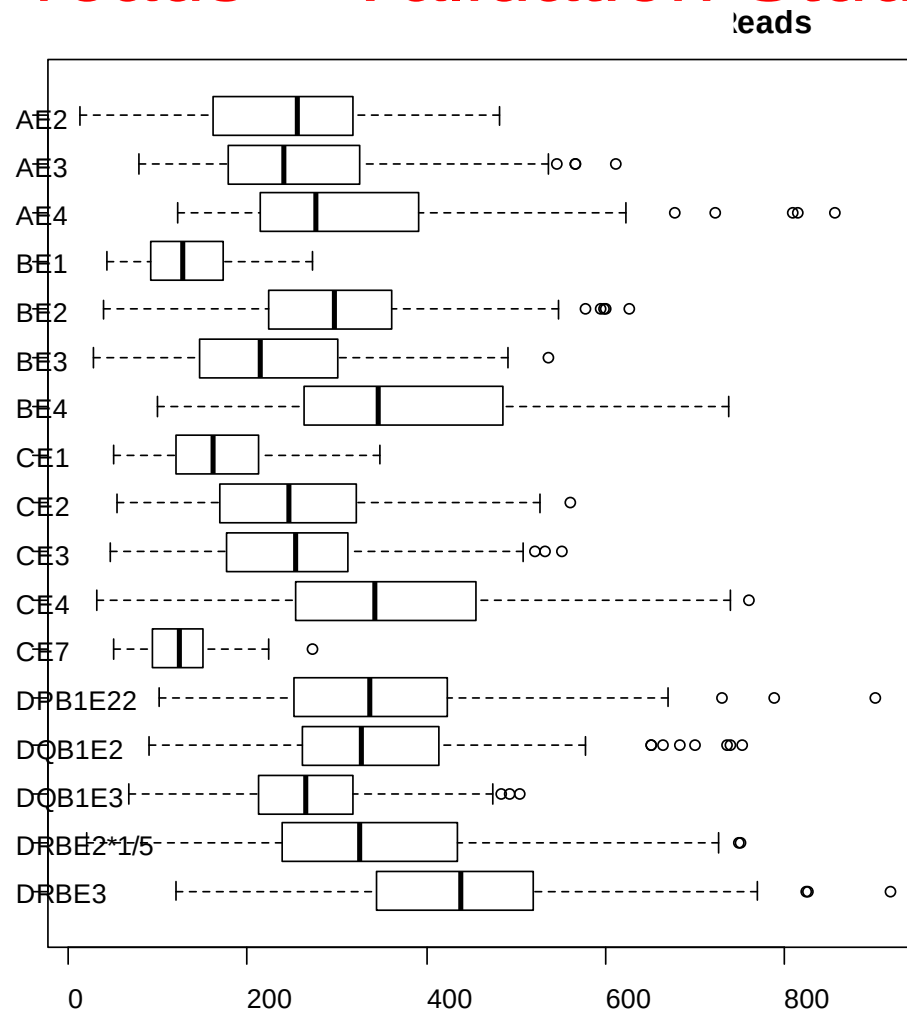
Parameter	Avg.	S.D.	CV (%)
Raw Wells [wells]	182565	32665	18
HQ Reads per Run [reads]	66078	16821	25
Median Read Length [bp]	425.33	24.55	6
QC (400bp, 98%) [%]	77.14	4.03	5



AUSTRIAN RED CROSS

UPPER AUSTRIA

Raw 454 reads – Validation Study.



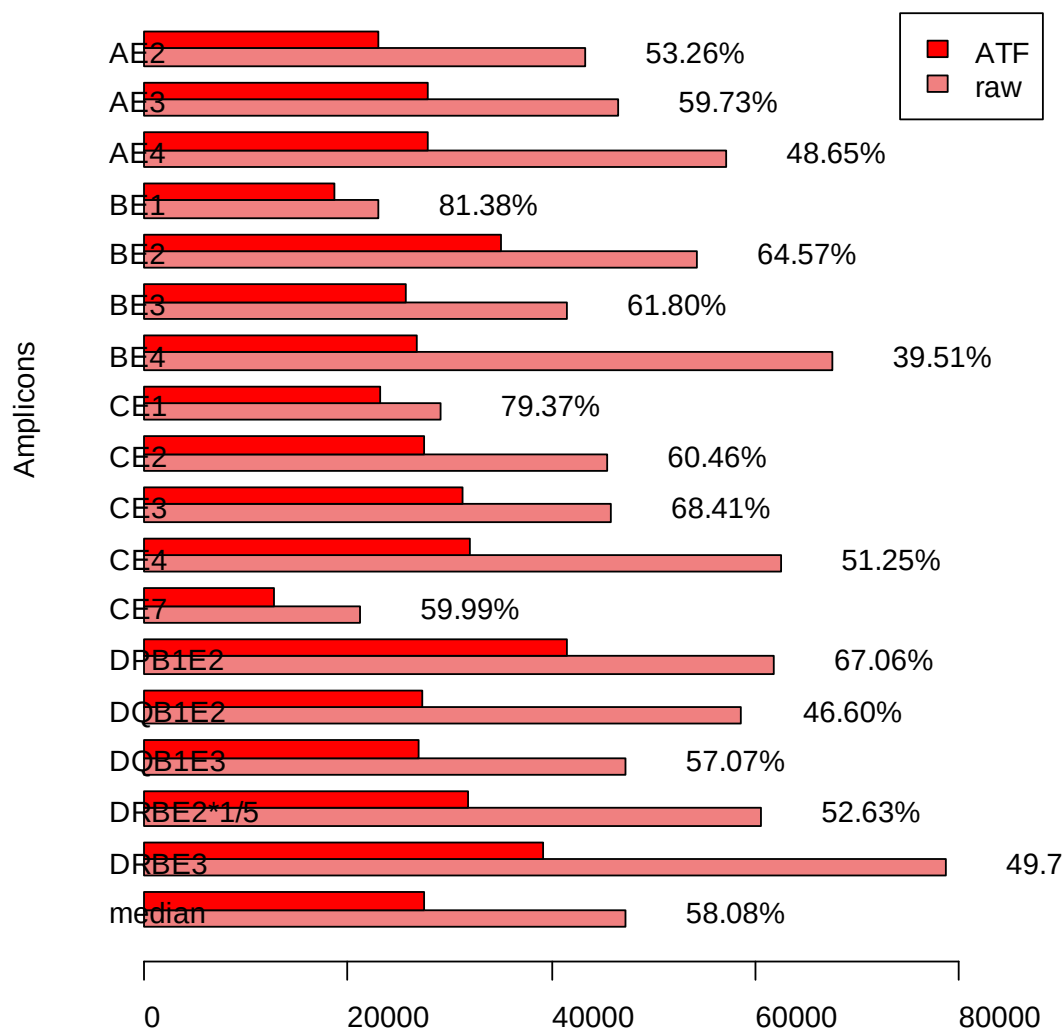


AUSTRIAN RED CROSS

UPPER AUSTRIA

Raw 454 reads vs used ATF reads

Validation Study



Reasons:

Recognized
Sequencing
Errors

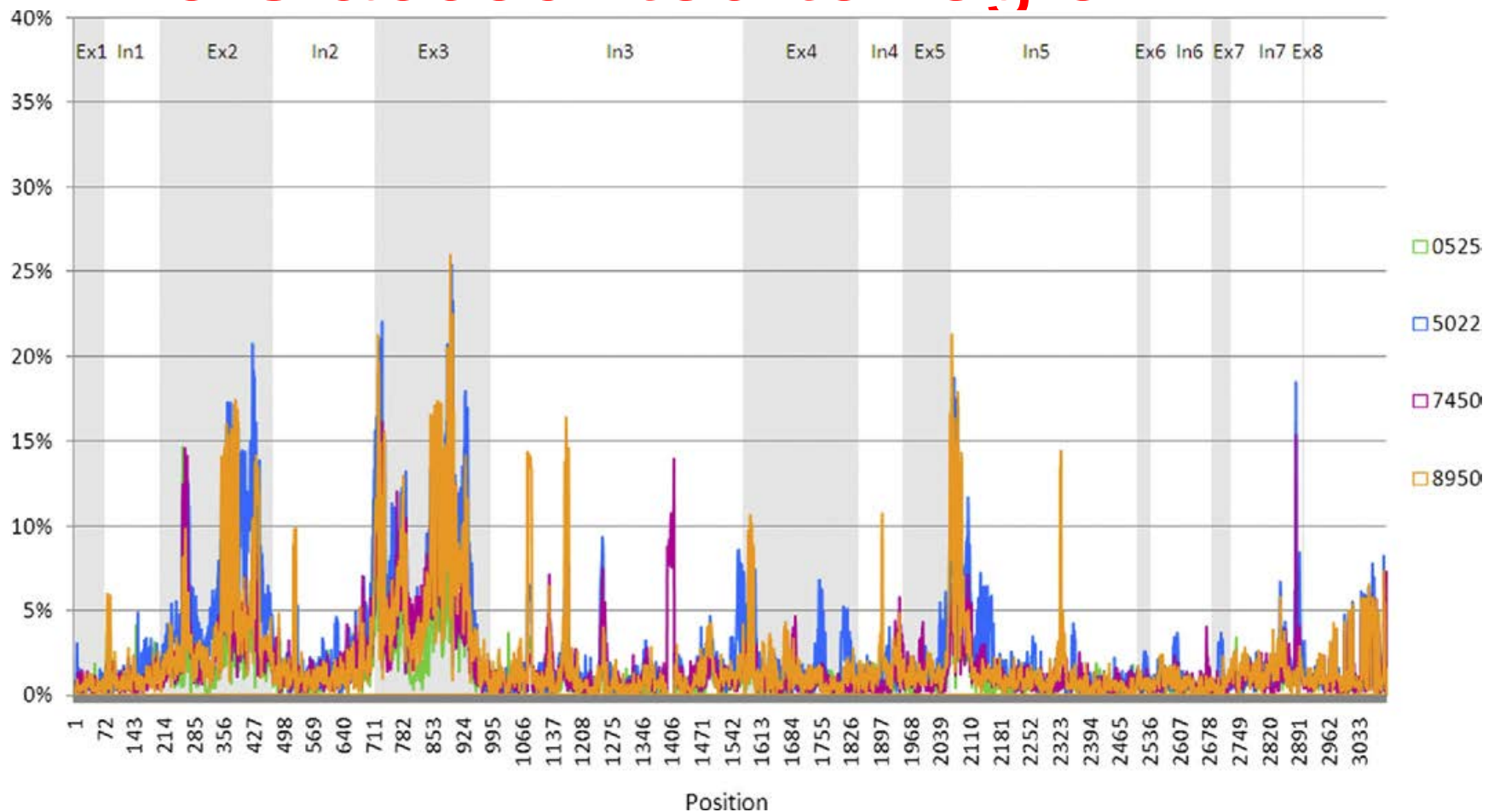
(Filter settings,
Coverage)



AUSTRIAN RED CROSS

UPPER AUSTRIA

Errors accounted to region

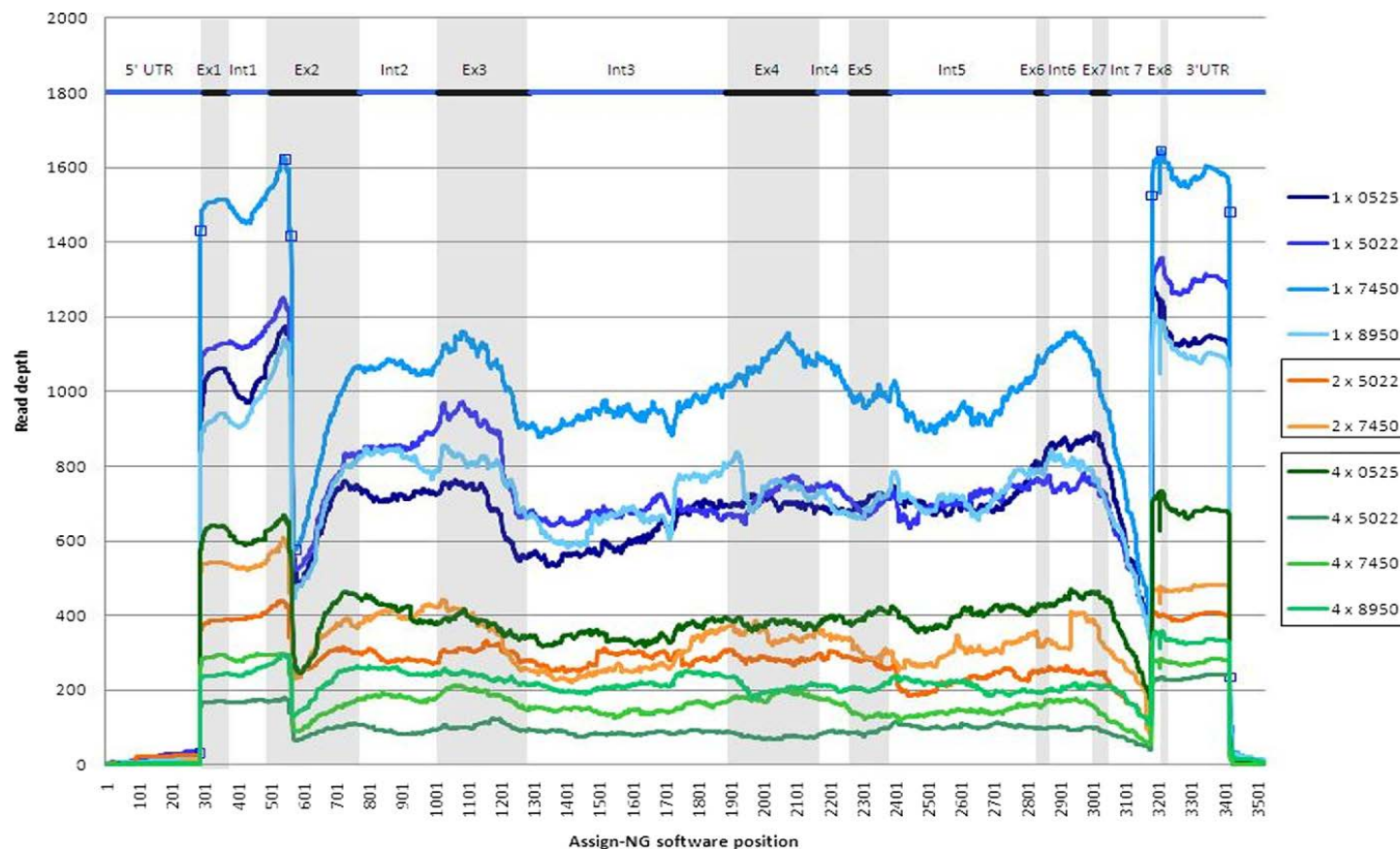




AUSTRIAN RED CROSS

UPPER AUSTRIA

Depth of base calls varies at each position





AUSTRIAN RED CROSS

UPPER AUSTRIA

Summary Typing Results – Validation Study

■

Locus	Total	successful	Insufficient Read counts	wrong	No Sanger	successfull Typings [%]	Concordance Rate [%]
A	173	167	5	0	1	97,09%	100,00%
B	173	172	1	0	0	99,42%	100,00%
C	172	166	4	0	2	97,65%	100,00%
DRB1	172	169	2	0	1	98,83%	100,00%
DRB3	115	91	20	0	4	81,98%	100,00%
DRB4	77	73	1	0	3	98,65%	100,00%
DRB5	53	52	0	0	1	100,00%	100,00%
DQB1	162	158	0	0	4	100,00%	100,00%
DPB1	30	29	1	0	0	96,67%	100,00%
all	1127	1077	34	0	16	96,94%	100,00%



AUSTRIAN RED CROSS

UPPER AUSTRIA

High throughput genotyping

- no full exploitation of sequencing capacity with 26 patients
- according to the manufacturer's promotion up to 500 patients possible - in theory
- in HLA typing sequencing of 48 patients (à 14 amplicons) per run is performed
- simple amendment of further amplicons for blood group genotyping (ABO, etc.)



AUSTRIAN RED CROSS

UPPER AUSTRIA

NGS HLA useability

- Variation in the selection of primers and exons – there is no standard
- NGS is useful but only under certain conditions:
 - High throughput
 - Automation
 - Full exploitation of bioinformatics



AUSTRIAN RED CROSS

UPPER AUSTRIA

Conclusions

- Primers from Sanger SBT to NGS are not transferable. Some need new design
- Clonal sequencing by NGS enables
 - resolution of cis/trans linkage of SNPs
 - clear identification of duplications
 - *de novo* alterations can be identified easily
- NGS suitable for immunogenetics in very high resolution mode

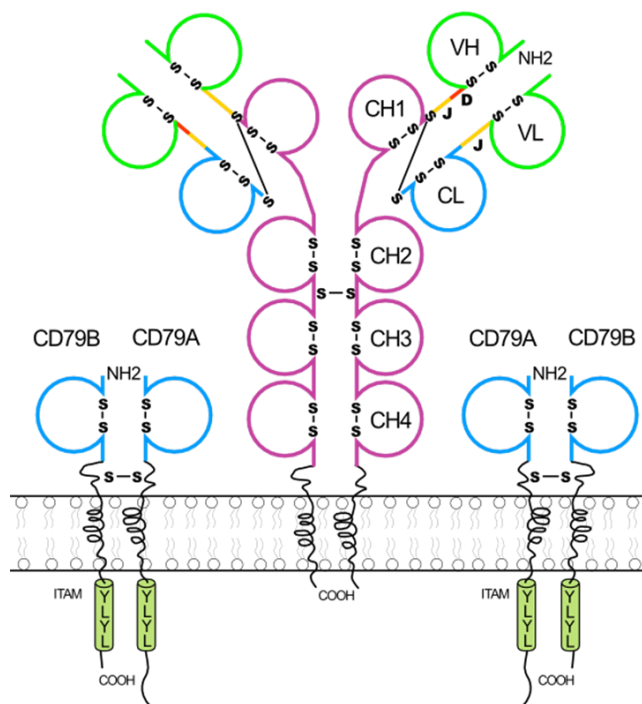


AUSTRIAN RED CROSS

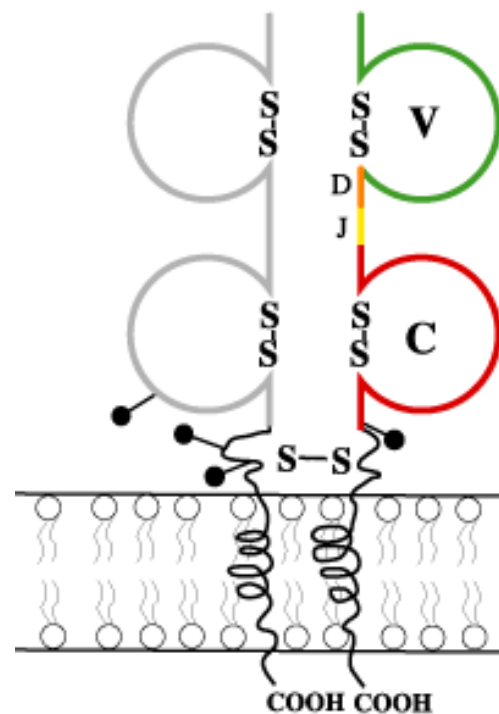
UPPER AUSTRIA

SEQUENCING IGH AND TCR

Structure



BCR



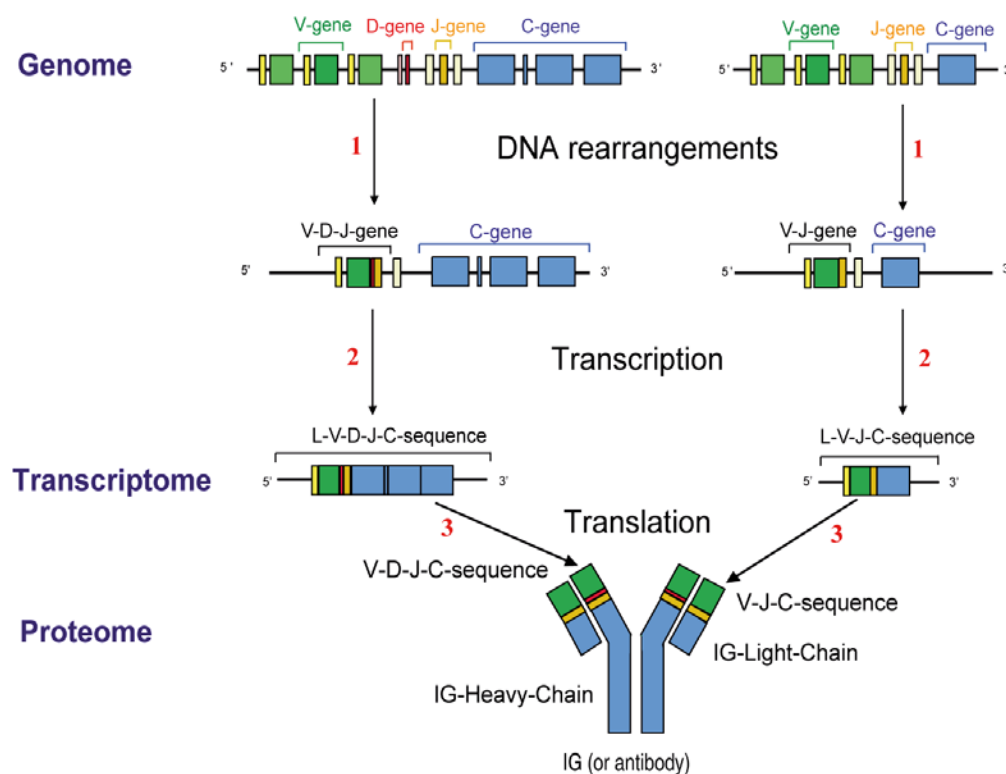
TCR



Receptor-recombination

ca. 10^{12} variants of receptors

One specific receptor per cell

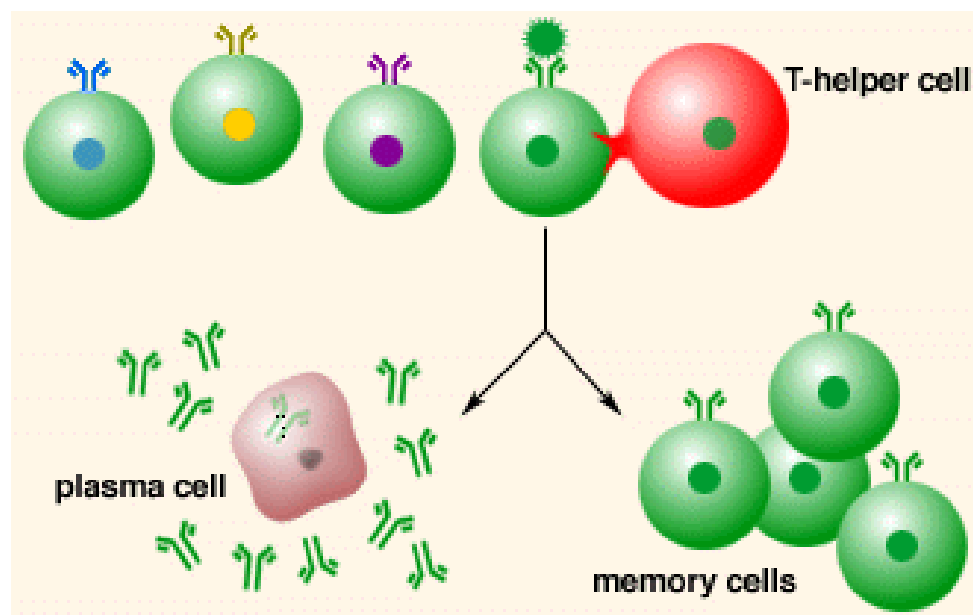




AUSTRIAN RED CROSS

UPPER AUSTRIA

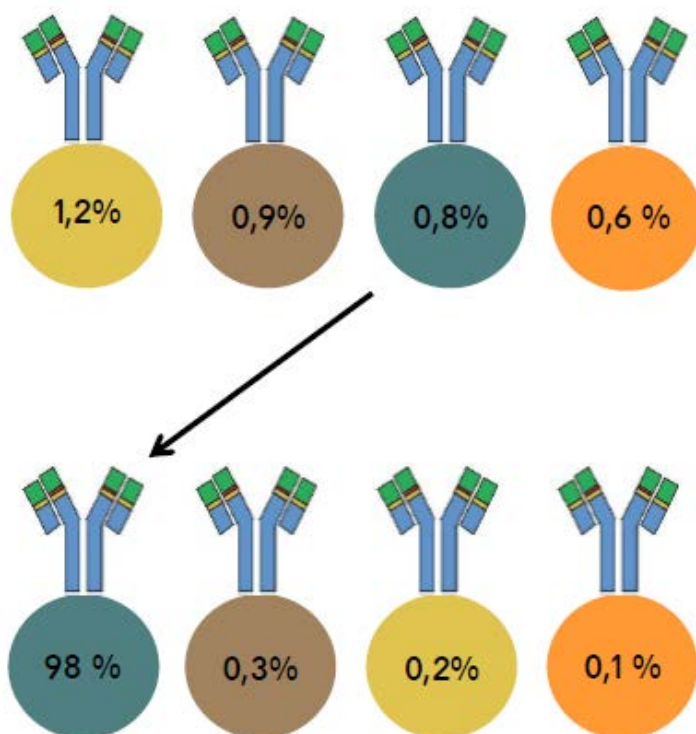
Clonal selection and proliferation



Cells with high affinity to epitope proliferate



Clonal Repertoire



Clonal development of the BCR/TCR repertoire by directed or undirected proliferation



AUSTRIAN RED CROSS

UPPER AUSTRIA

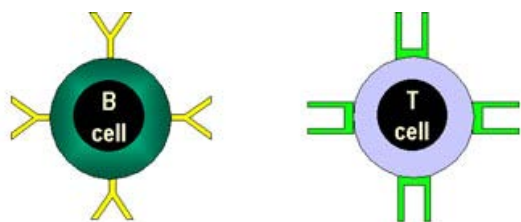
METHODS



AUSTRIAN RED CROSS

UPPER AUSTRIA

Workflow



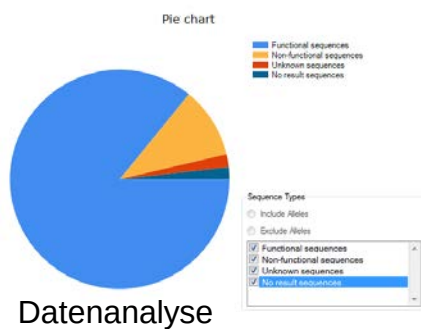
B- und T-Zell Isolation



DNA-
Extraktion



NGS Sample
Preparation



NGS

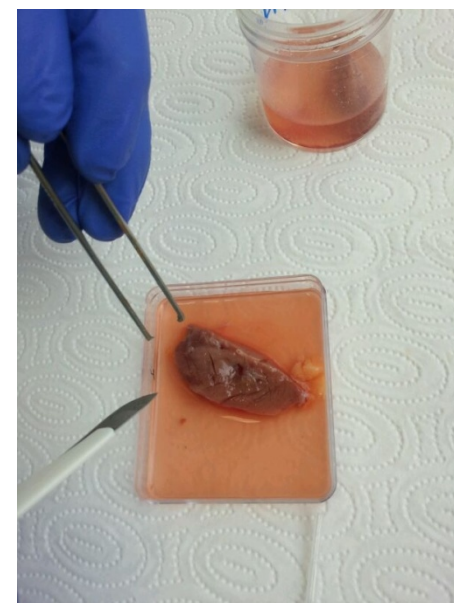


AUSTRIAN RED CROSS

UPPER AUSTRIA

Material

- Blood
- Urine
- Kidney tissue



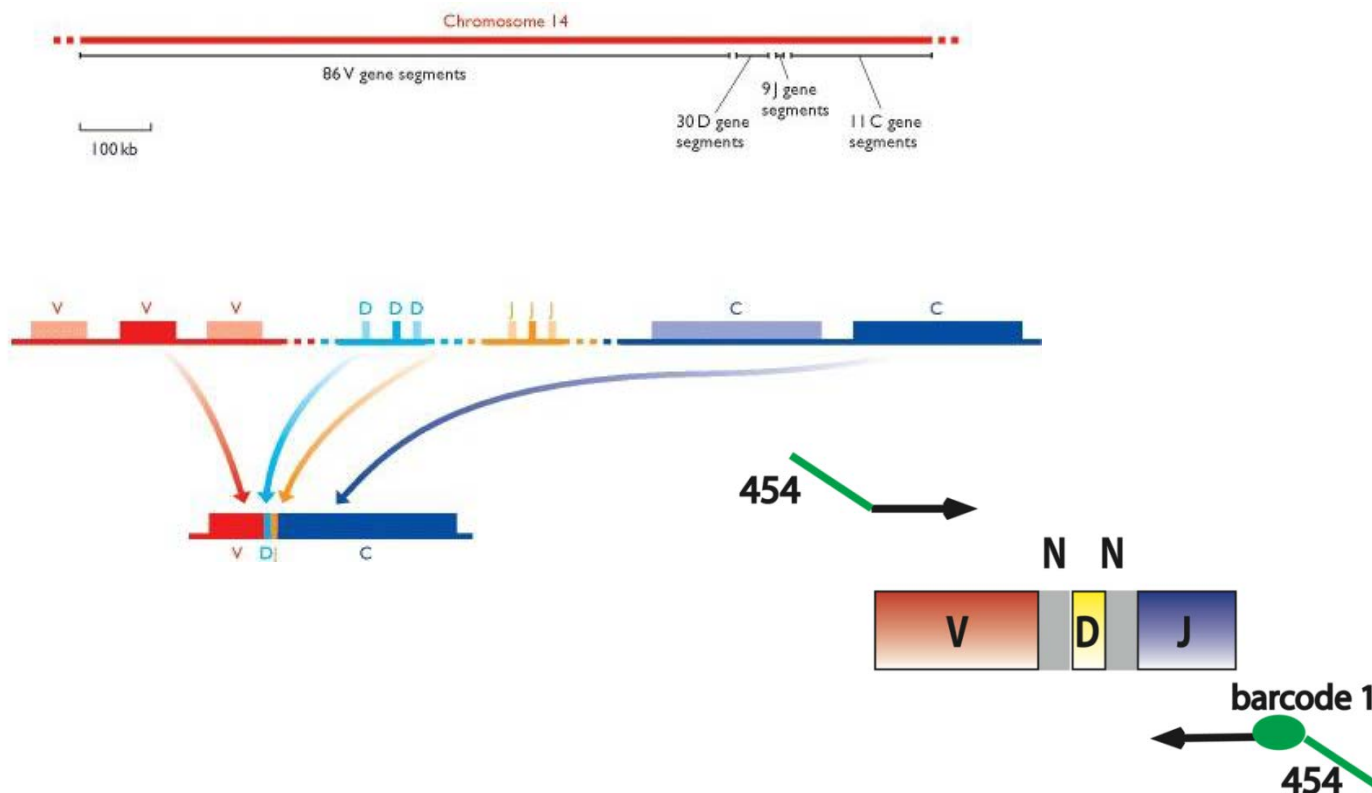
Johannes Weinberger



AUSTRIAN RED CROSS

UPPER AUSTRIA

Amplicons of the recombinant IGH and TCRB Loci



Johannes Weinberger

NGS Sequencing



AUSTRIAN RED CROSS

UPPER AUSTRIA

GSj



100,000 reads
Länge: 400bp

GS FLX(+)

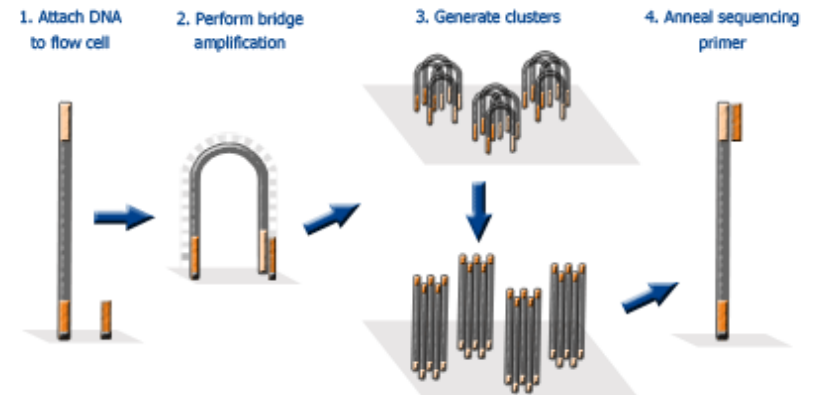
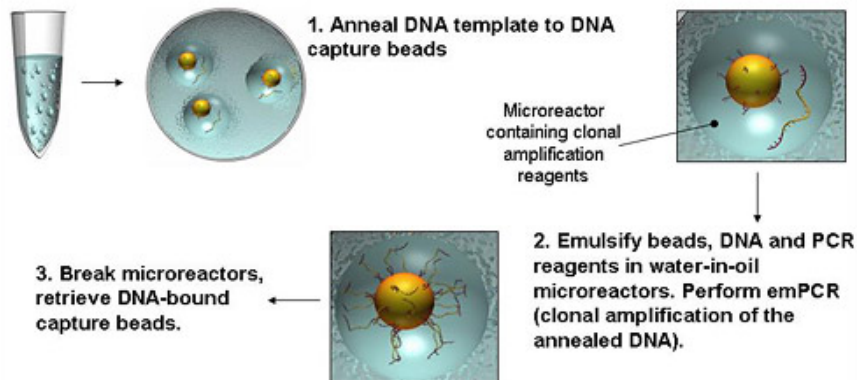


700,000 reads
Länge: 400-700bp

MiSeq



15,000,000 reads
Länge: 250bp



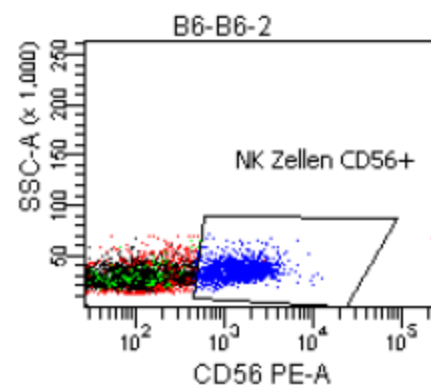
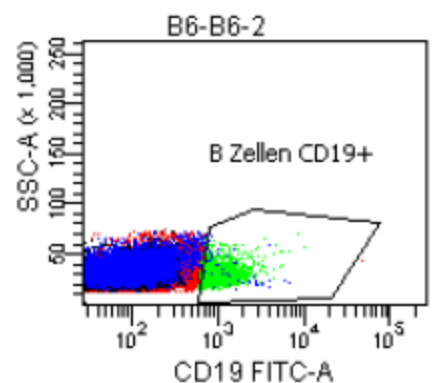
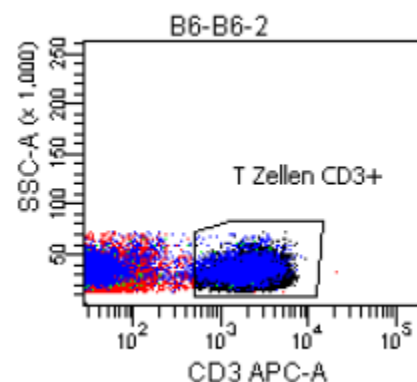
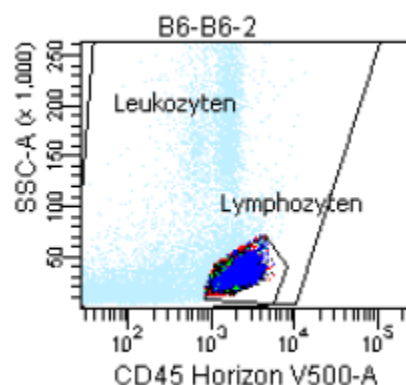
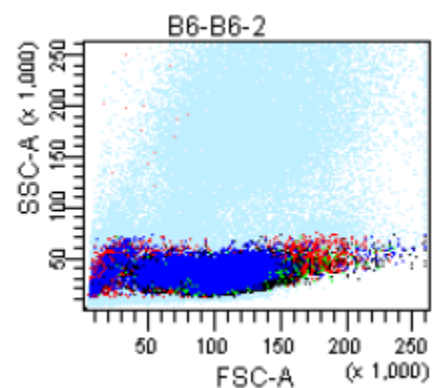
Johannes Weinberger



AUSTRIAN RED CROSS

UPPER AUSTRIA

FACS Blood



Tube: B6-2

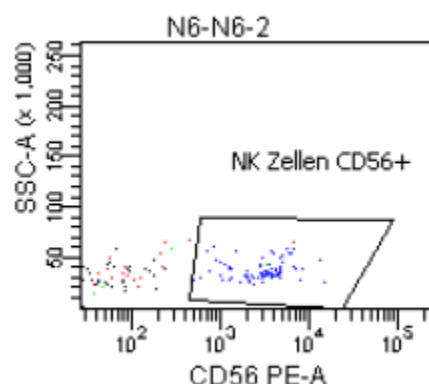
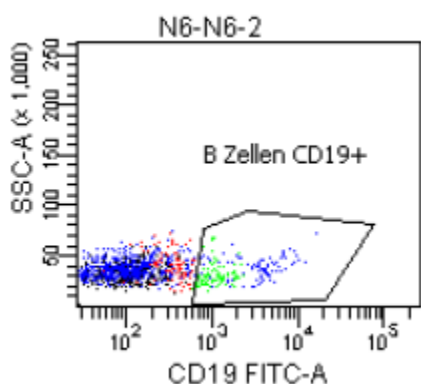
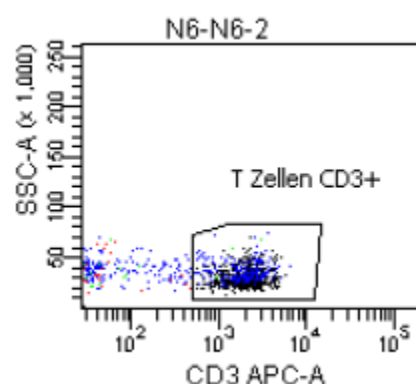
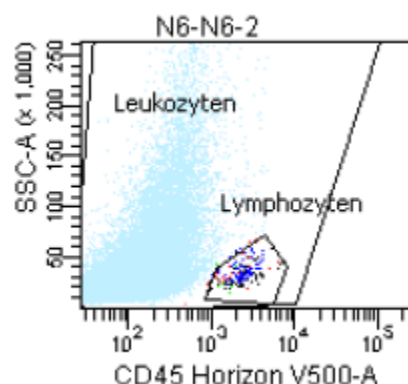
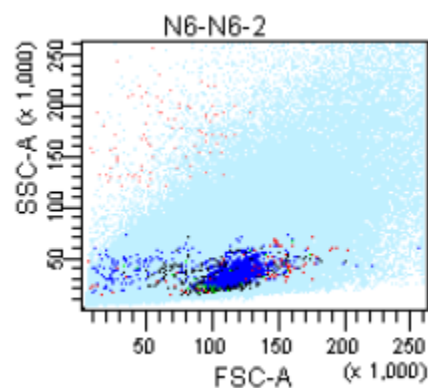
Population	#Events	%Parent	%Total
All Events	300,000	####	100.0
Lymphozyten	61,790	20.6	20.6
T Zellen CD3+	35,282	57.1	11.8
B Zellen CD19+	1,900	3.1	0.6
NK Zellen CD56+	16,699	27.0	5.6
Leukozyten	299,880	100.0	100.0



AUSTRIAN RED CROSS

UPPER AUSTRIA

FACS tissue



Tube: N6-2

Population	#Events	%Parent	%Total
All Events	300,000	####	100.0
Lymphozyten	1,846	0.6	0.6
T Zellen CD3+	835	45.2	0.3
B Zellen CD19+	188	10.2	0.1
NK Zellen CD56+	949	51.4	0.3
Leukozyten	299,828	99.9	99.9

Johannes Weinberger



AUSTRIAN RED CROSS

UPPER AUSTRIA

CDR3 distribution of healthy volunteers

Sequence	Count	Percentage
ACWSSDTLNSTGP	325	1,55405728
ASS*GEQSAPA	275	1,31497155
ASSETVDYGYT	182	0,87027208
ATEGYGYT	120	0,57380577
ASSFLRGGYGYT	109	0,5212069
ASNQNTGAPYEQY	101	0,48295319
ASSHRDRGRSLKL	85	0,40644575
ASSYGRQAISPS	78	0,37297375
ASSPGLTVRSYEQY	78	0,37297375
ASRQSYGYT	77	0,36819203
TSS	74	0,35384689

TRB

Sequence	Count	Percentage
ARVFRY	311	0,140360695
ARGIDY	261	0,117794667
AR	213	0,09613128
ARCYGMDV	203	0,091618074
ARGGNSDY	196	0,088458831
ARWHDNIPADY	175	0,078981099
ARDVGTLEWYYFDY	164	0,074016572
ASTYYGMDV	155	0,069954687
AASSG	155	0,069954687
ARGYYGMDV	147	0,066344123
ARALVDY	140	0,063184879

IGH



Patient 1

- Nephrectomy due to Hydronephrosis
- Kidneybiopsy: chronic interstitial nephritis with tubulitis and atrophy of tubules, as a response to chronic pyelonephritis
- Tissue sample: 2x3cm



CDR3 9N TCRB vs 9B TCRB

9N

Sequence	Count	Percentage
ASSPYTTGRKLF	12299	18,6249716
ASSKEYRGAGGYT	9588	14,519573
ASSPDRGGNQPQH	6352	9,61914136
ASSYSPTMNTAEF	3527	5,34110699
ASSYGGDGYT	1904	2,88331945
ASSATLEGAGYGYT	835	1,26448096
ATLNYGYT	599	0,90709472
ASSSTA*GHNQPQH	595	0,90103733
ASSILLRTLLRAV	486	0,73597335

9B

Sequence	Count	Percentage
ASSKEYRGAGGYT	6170	4,79562254
ASSYSPTMNTAEF	5997	4,66115857
ASSPYTTGRKLF	5753	4,47150996
ASSYGGDGYT	5510	4,2826386
ASSATLEGAGYGYT	3311	2,5734694
ASSPDRGGNQPQH	847	0,65832938
ASSPRDS*TLKL	670	0,52075642
ASS*GC*QGMAT	622	0,4834485
ASSEGVGTFSAPLH	552	0,42904111



CDR3 9N IGH vs 9B IGH

9N

Sequence	Count	Percentage
ASVMGPLLWFGKSQHRYYFDY	22536	35,4256072
MKRC*RSSD*SSYGP*L	382	0,60048731
AREVGRRPPEAWEVWT	337	0,52974927
ASVMGPLLWFGKSQHRYYFDS	257	0,40399277
ASVMGPLLWFGKSQHTYYFDY	253	0,39770494
ATSLI	174	0,2735204
AKRVGAYSPFEY	169	0,26566061
TKTFSRLVKYFCRD*L	161	0,25308496
F*LAAGAY*FDC	136	0,21378606

9B

Sequence	Count	Percentage
ASVMGPLLWFGKSQHRYYFDY	2164	6,69948299
AKRVGAYSPFEY	1587	4,91316058
IAVVTAFLDVSDI	1168	3,61598712
FYYDSGCLAGST	1018	3,15160521
AKRVGASSPFEY	227	0,70276462
GRDSVETGGLT	138	0,42723136
ARV*TTMRVVGRFGGLT	108	0,33435497
ASVMGPLLWFGESQRRYYFDF	102	0,3157797
ARDWGIQLWLLGGMDV	101	0,31268382



AUSTRIAN RED CROSS

UPPER AUSTRIA

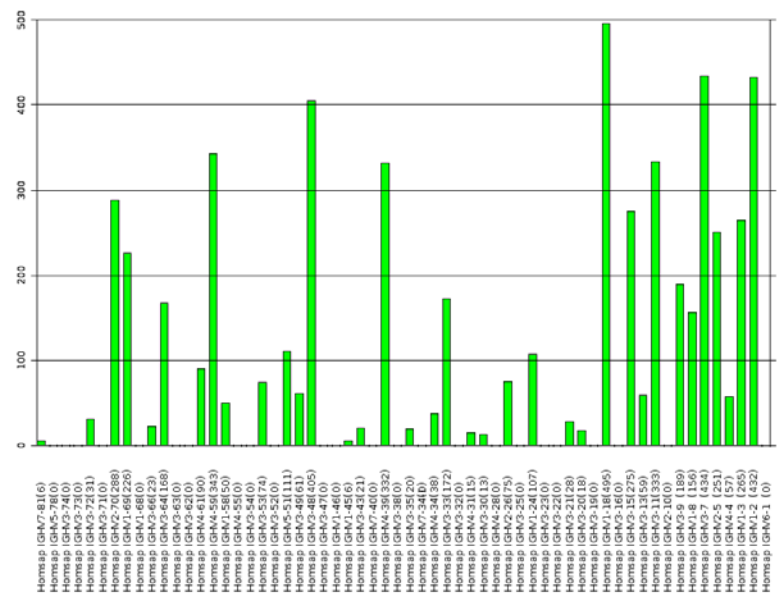
CLL patient

Sequence	Count	Percentage
ARAGINWGPYFDC	17701	95,5158645
ARAGINWAHTLT	79	0,42628966
AKGLTTTGTPDY	65	0,35074466
ARAGINWGPYFD	48	0,25901144
VKGLTTTGTPDY	41	0,22123894
ARAGINWGPYLDC	24	0,12950572
ARAGINWGSYFDC	20	0,10792143
ARAGINWSPYFDC	17	0,09173322
ARVGINWGPYFDC	16	0,08633715

IGH

Sequence	Count	Percentage
ASSRGRGYEQY	2124	3,54035404
ASSIIGNQPQH	1496	2,49358269
ASVQVTMAT	942	1,57015702
ASSRGQSYGYT	584	0,97343068
ASSRGQNYGYT	578	0,96342968
ASREDGRLRAV	472	0,78674534
ASSRGQGYEQY	444	0,74007401
ASSRGLSYEQY	438	0,73007301
ASSRGGSYEQY	381	0,63506351

TRB



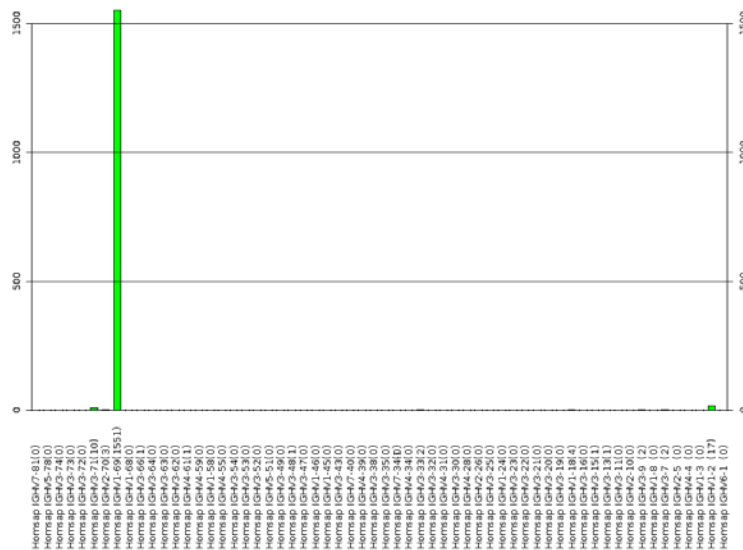


AUSTRIAN RED CROSS

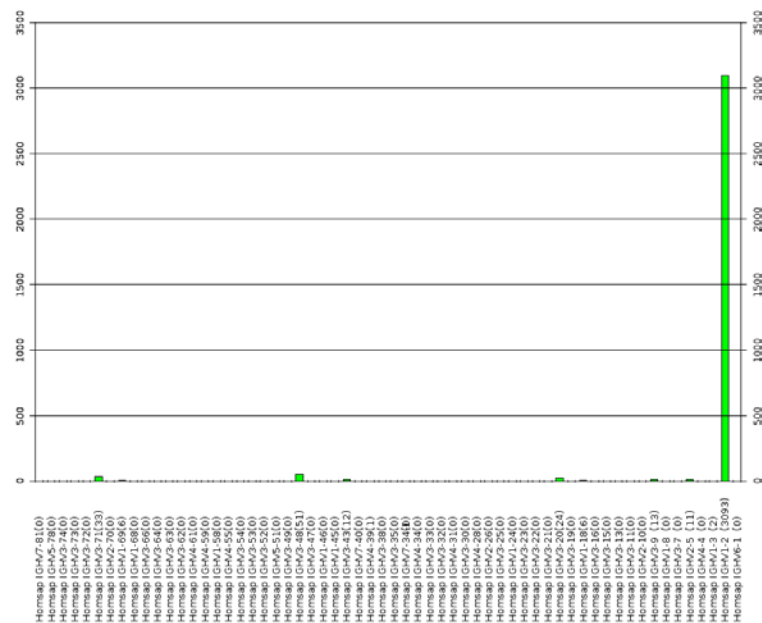
UPPER AUSTRIA

V-Elements Lymphoma

Patient 1



Patient 2

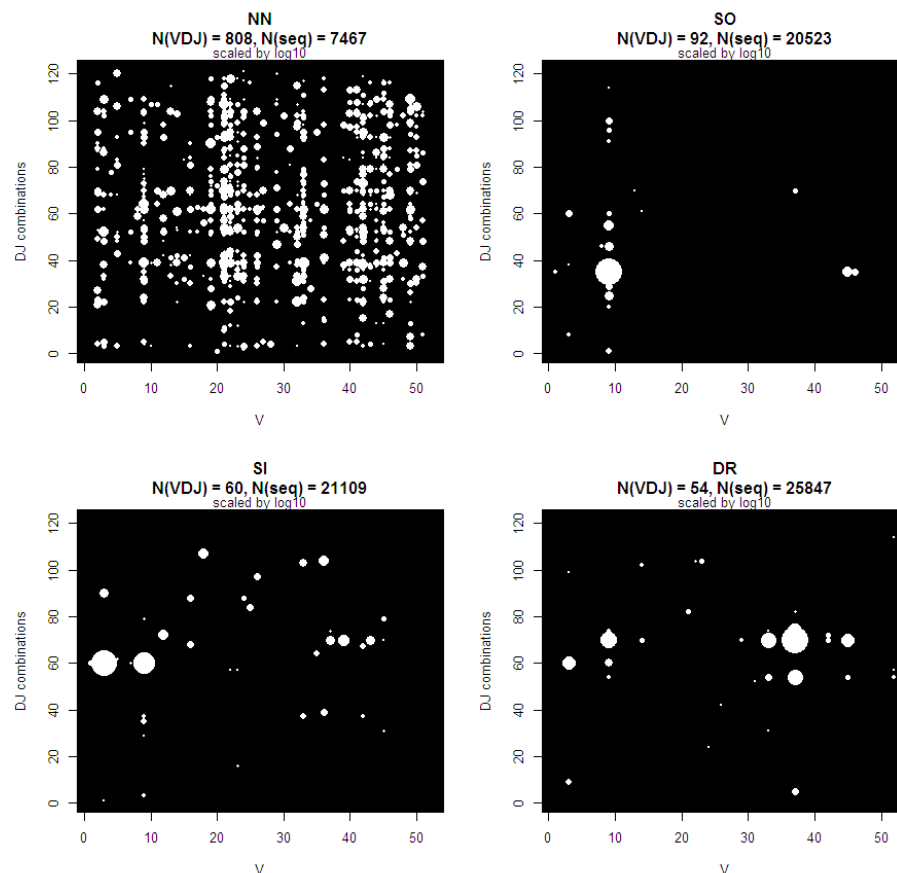




AUSTRIAN RED CROSS

UPPER AUSTRIA

V(D)J recombination: Spectra IgH

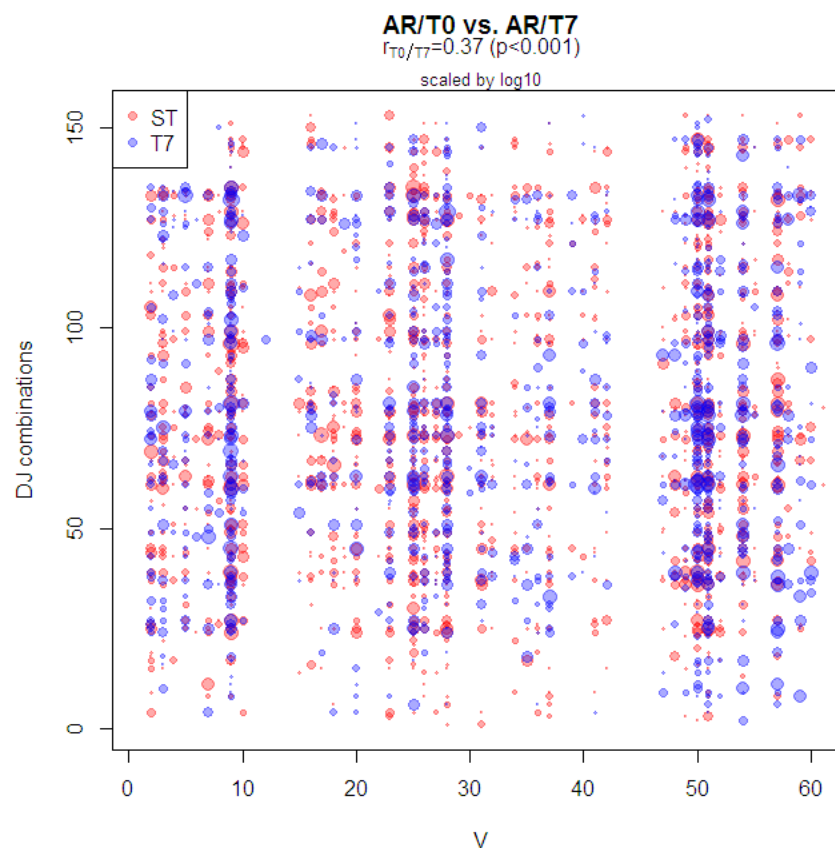




AUSTRIAN RED CROSS

UPPER AUSTRIA

2 V(D)J spectra same patient at different times



Working with NGS is
sometimes slippery



AUSTRIAN RED CROSS

UPPER AUSTRIA



You need a really clear
vision what you want



And sometimes serenity



AUSTRIAN RED CROSS



3.OG Team Genomics

Johannes Pröll, Christa Hackl, Katja Hofer, Steffi Stabentheiner, Martin Danzer, Norbert Niklas

