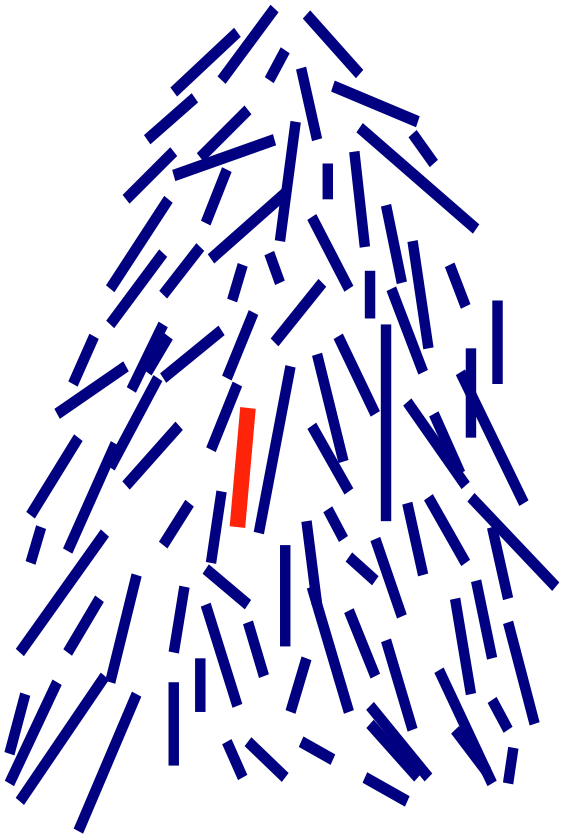
INSTITUTE OF APPLIED BIOSCIENCES
ΙΝΣΤΙΤΟΥΤΟ ΕΦΑΡΜΟΣΜΕΝΩΝ ΒΙΟΕΠΙΣΤΗΜΩΝ
CERTH

Molecular MRD analysis in CLL

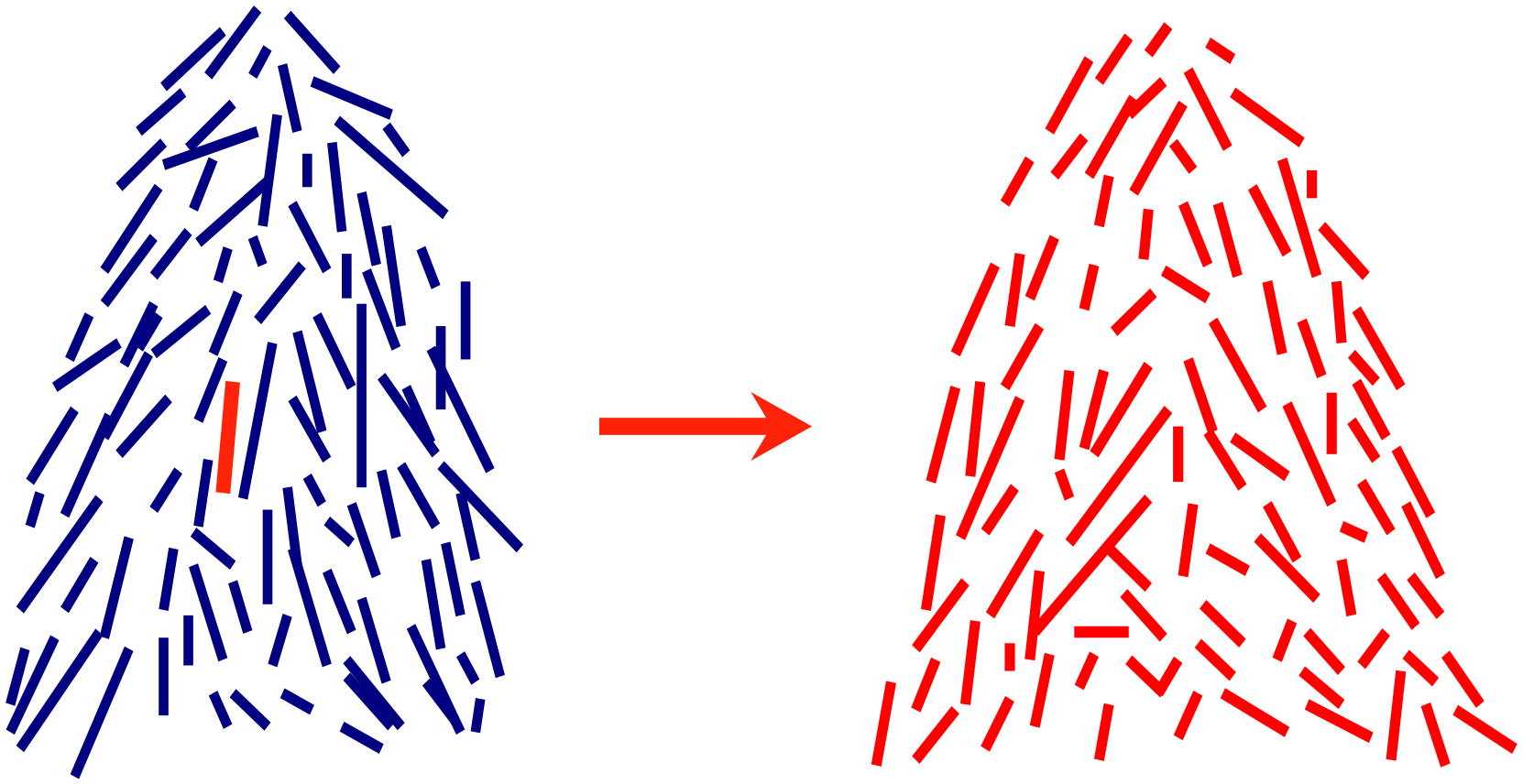
Kostas Stamatopoulos

Institute of Applied Biosciences
CERTH, Thessaloniki, Greece

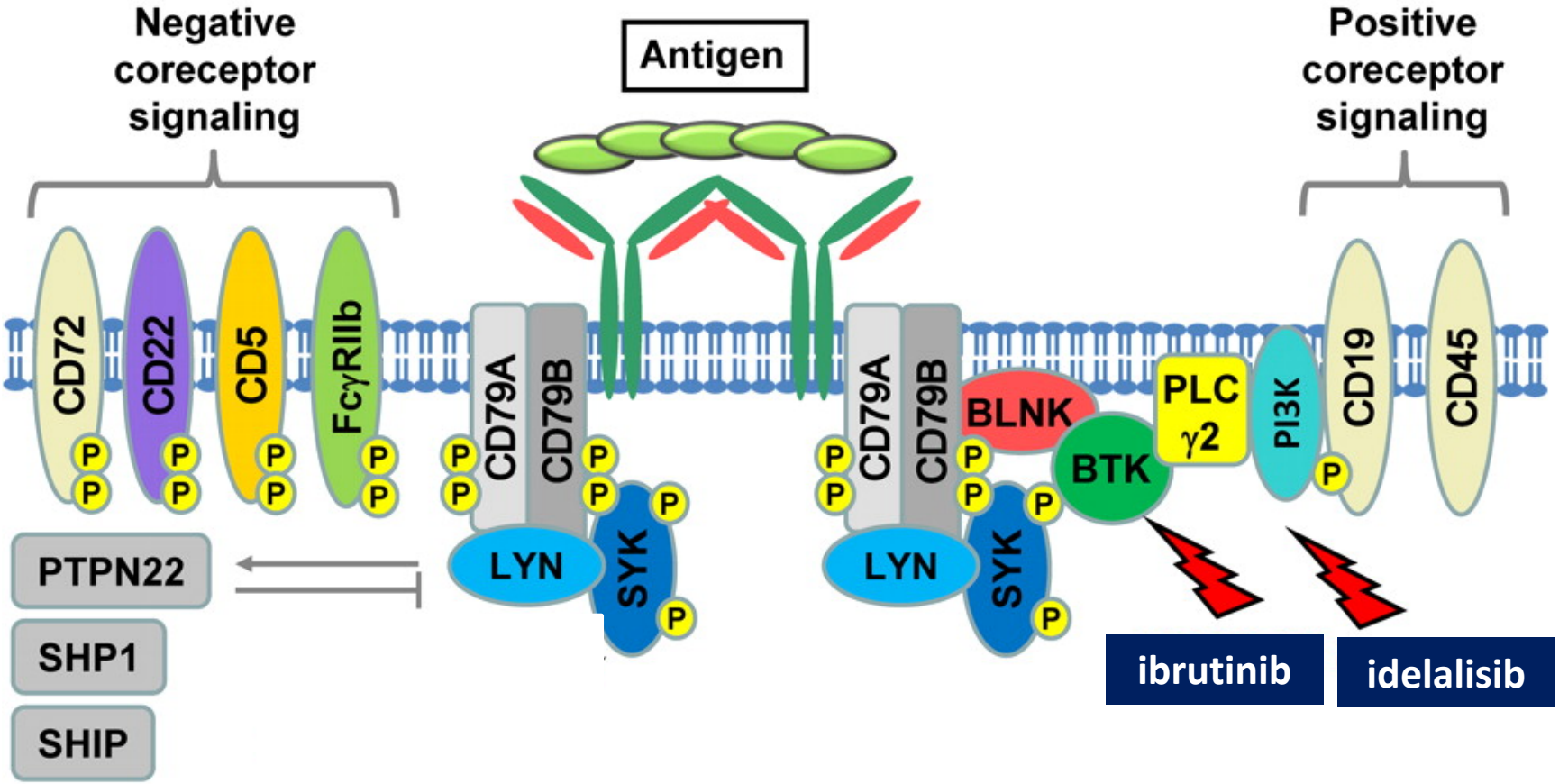
minimal residual disease



MRD by PCR



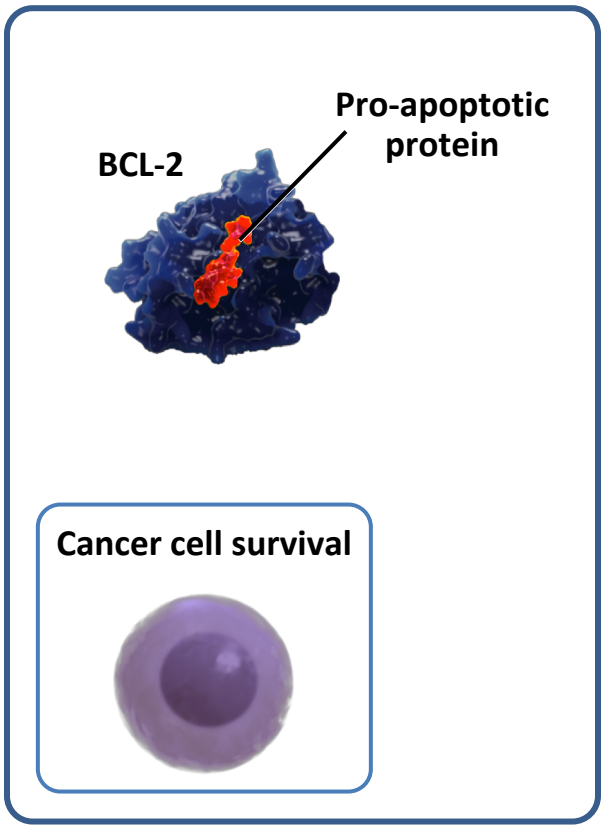
2014-2018
BTK and PI3Kδ inhibition
a rational and feasible way of treating CLL

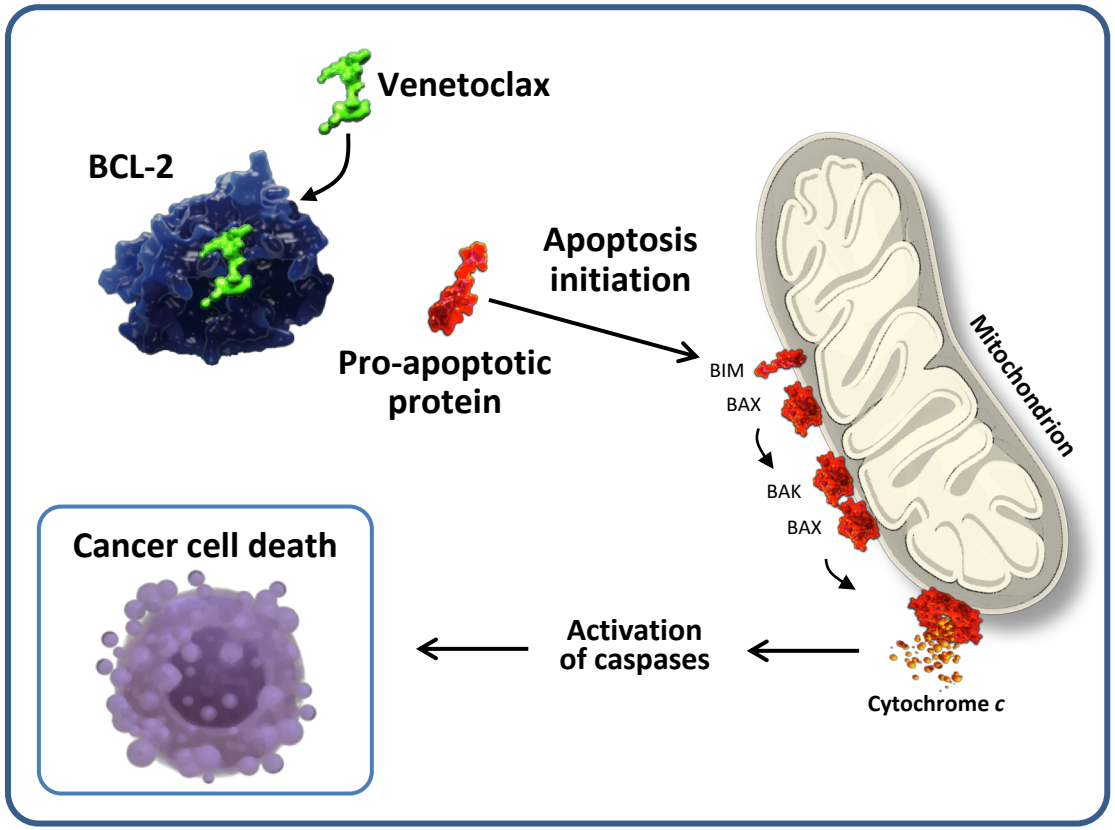


Stevenson F K et al. Blood 2011

2016-2018

restoration of apoptosis through **BCL2 inhibition with Venetoclax**





precision medicine is becoming a **reality** in the clinical management of **CLL**

CLL in 2019

CLL in 2019: on the road to cure?

blood®

Prepublished online March 14, 2018;
doi:10.1182/blood-2017-09-806398

Guidelines for diagnosis, indications for treatment, response assessment and supportive management of chronic lymphocytic leukemia

Michael Hallek, Bruce D. Cheson, Daniel Catovsky, Federico Caligaris-Cappio, Guillermo Dighiero, Hartmut Döhner, Peter Hillmen, Michael Keating, Emili Montserrat, Nicholas Chiorazzi, Stephan Stilgenbauer, Kanti R. Rai, John C. Byrd, Barbara Eichhorst, Susan O'Brien, Tadeusz Robak, John F. Seymour and Thomas J. Kipps

blood®
Prepublished online March 14, 2018;
doi:10.1182/blood-2017-09-806398

Guidelines for diagnosis, indications for treatment, response assessment and supportive management of chronic lymphocytic leukemia

Michael Hallek, Bruce D. Cheson, Daniel Catovsky, Federico Caligaris-Cappio, Guillermo Dighiero, Hartmut Döhner, Peter Hillmen, Michael Keating, Emili Montserrat, Nicholas Chiorazzi, Stephan Stilgenbauer, Kanti R. Rai, John C. Byrd, Barbara Eichhorst, Susan O'Brien, Tadeusz Robak, John F. Seymour and Thomas J. Kipps

the complete eradication of the leukemia is a desired endpoint

how?

blood®
Prepublished online March 14, 2018;
doi:10.1182/blood-2017-09-806398

Guidelines for diagnosis, indications for treatment, response assessment and supportive management of chronic lymphocytic leukemia

Michael Hallek, Bruce D. Cheson, Daniel Catovsky, Federico Caligaris-Cappio, Guillermo Dighiero, Hartmut Döhner, Peter Hillmen, Michael Keating, Emili Montserrat, Nicholas Chiorazzi, Stephan Stilgenbauer, Kanti R. Rai, John C. Byrd, Barbara Eichhorst, Susan O'Brien, Tadeusz Robak, John F. Seymour and Thomas J. Kipps

MRD negativity in CLL **threshold**

Refinement and harmonization of technologies has established that a typical flow cytometry-based assay comprises a core panel of six markers (i.e. CD19, CD20, CD5, CD43, CD79b and CD81).

Patients will be defined as having undetectable MRD (MRD^{neg}) remission if they have blood or marrow with less than one CLL cell per 10,000 leukocytes

blood®
Prepublished online March 14, 2018;
doi:10.1182/blood-2017-09-806398

Guidelines for diagnosis, indications for treatment, response assessment and supportive management of chronic lymphocytic leukemia

Michael Hallek, Bruce D. Cheson, Daniel Catovsky, Federico Caligaris-Cappio, Guillermo Dighiero, Hartmut Döhner, Peter Hillmen, Michael Keating, Emili Montserrat, Nicholas Chiorazzi, Stephan Stilgenbauer, Kanti R. Rai, John C. Byrd, Barbara Eichhorst, Susan O'Brien, Tadeusz Robak, John F. Seymour and Thomas J. Kipps

MRD negativity in CLL
methods

The techniques for assessing MRD have undergone a critical evaluation and have become well standardized.

six-color flow cytometry (MRD flow)
allele-specific oligonucleotide PCR
high-throughput sequencing

are reliably sensitive down to a level of <1 CLL cell in 10,000 leukocytes.

blood®
Prepublished online March 14, 2018;
doi:10.1182/blood-2017-09-806398

Guidelines for diagnosis, indications for treatment, response assessment and supportive management of chronic lymphocytic leukemia

Michael Hallek, Bruce D. Cheson, Daniel Catovsky, Federico Caligaris-Cappio, Guillermo Dighiero, Hartmut Döhner, Peter Hillmen, Michael Keating, Emili Montserrat, Nicholas Chiorazzi, Stephan Stilgenbauer, Kanti R. Rai, John C. Byrd, Barbara Eichhorst, Susan O'Brien, Tadeusz Robak, John F. Seymour and Thomas J. Kipps

MRD negativity in CLL **samples**

The blood generally can be used for making this assessment, as the marrow will have detectable CLL when it is also found in the peripheral blood.

However, there are therapies that preferentially clear the blood but not the marrow (such as monoclonal antibodies).

Therefore, it may be important to confirm that the marrow aspirate also is MRD-neg when the blood is found to be MRD-neg.

Minimal Residual Disease in CLL recommendations

Aspect	Recommendation
Sensitivity threshold (Lower limit of quantification)	< 1 cell cell/10,000 leucocytes (10^{-4})
Methods of detection	Multicolor- Flow Cytometry ASO-qPCR NGS
Target tissue	**Screening in Peripheral blood** **Confirmation in Bone Marrow**

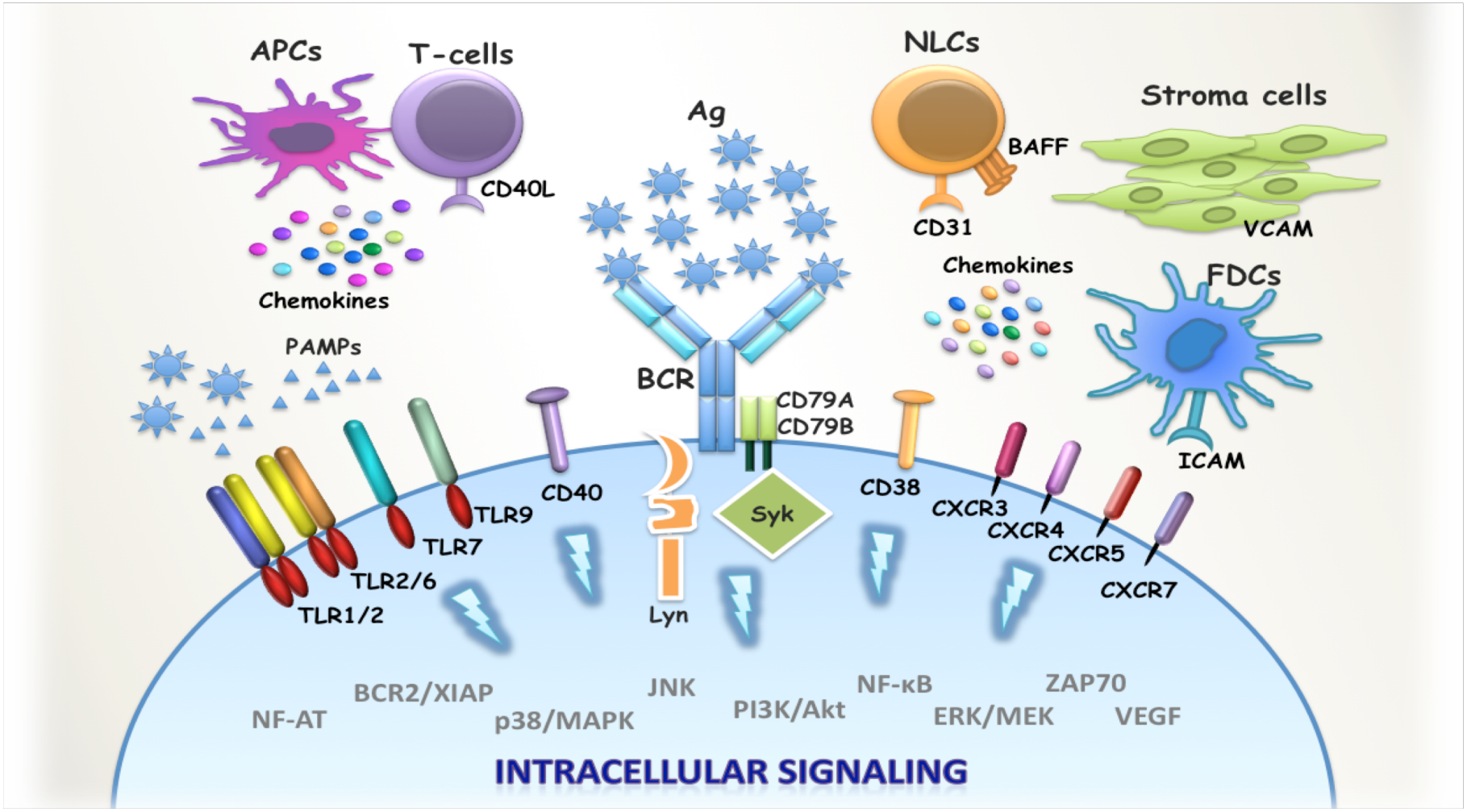
ERIC european research initiative on CLL

Ghia & Rawstron, Leukemia 2018

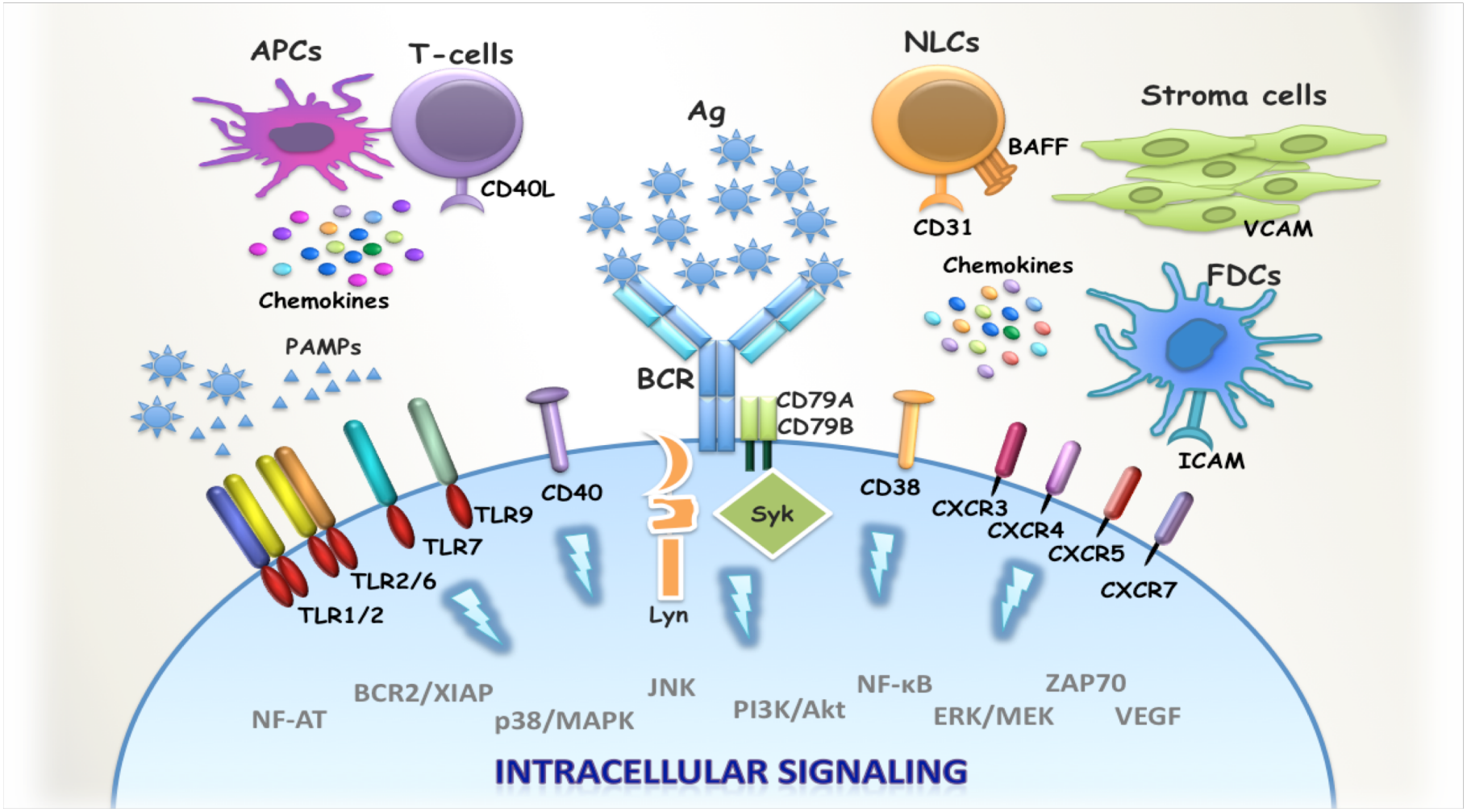
molecular MRD analysis in CLL

CLL is a clone of mature B cells

mature B cells interact with the microenvironment

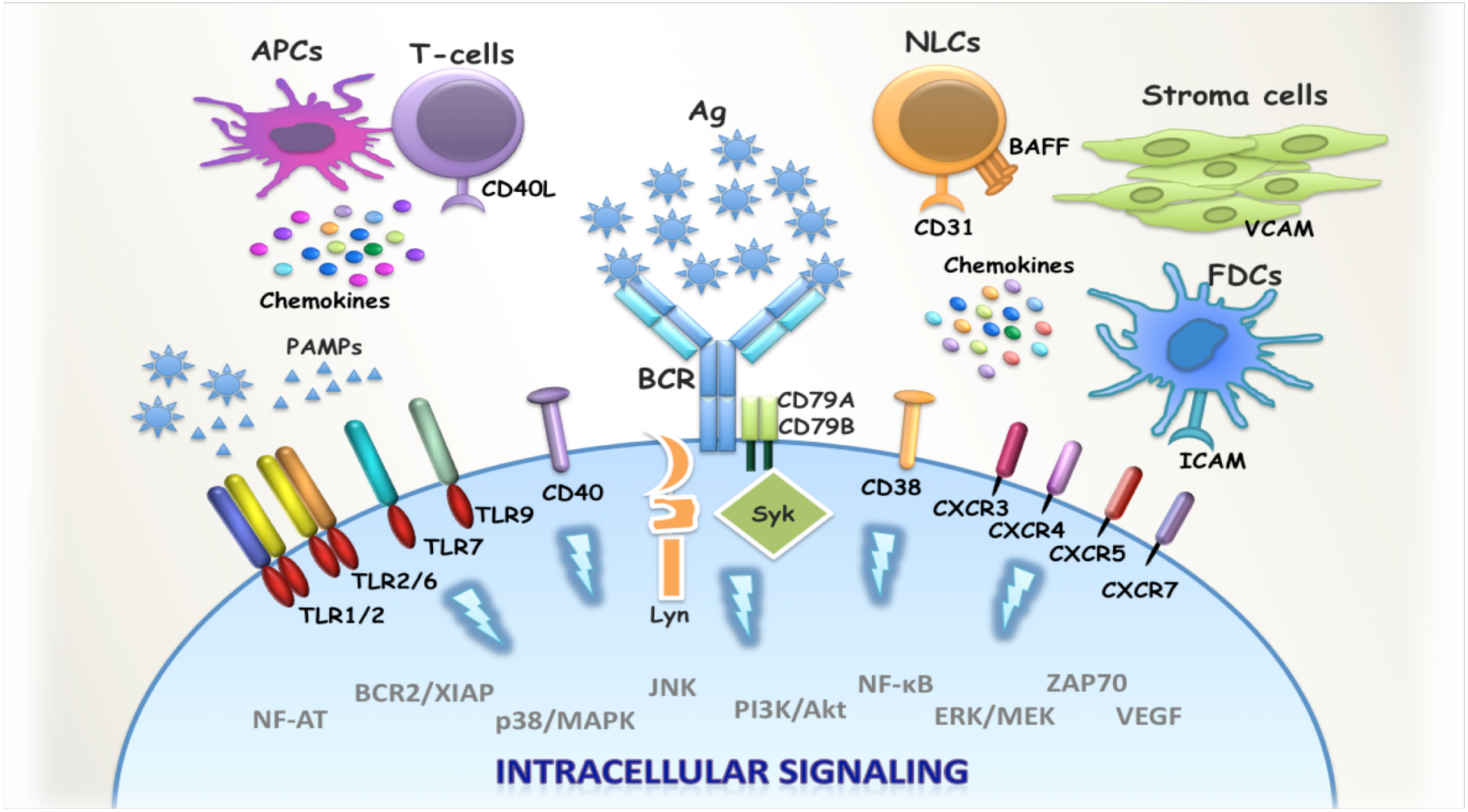


CLL cells are doing the same!



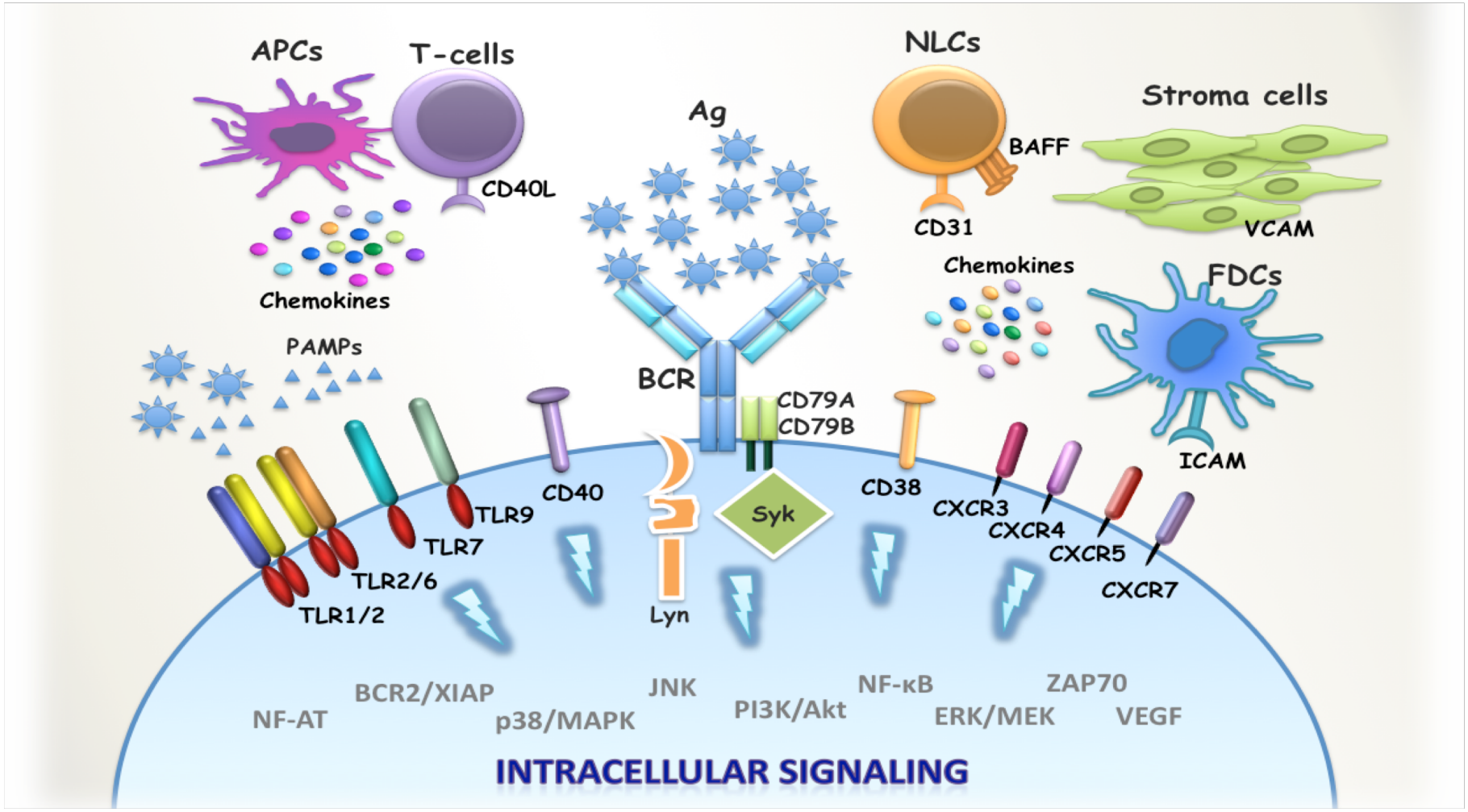
microenvironmental interactions

signal perception



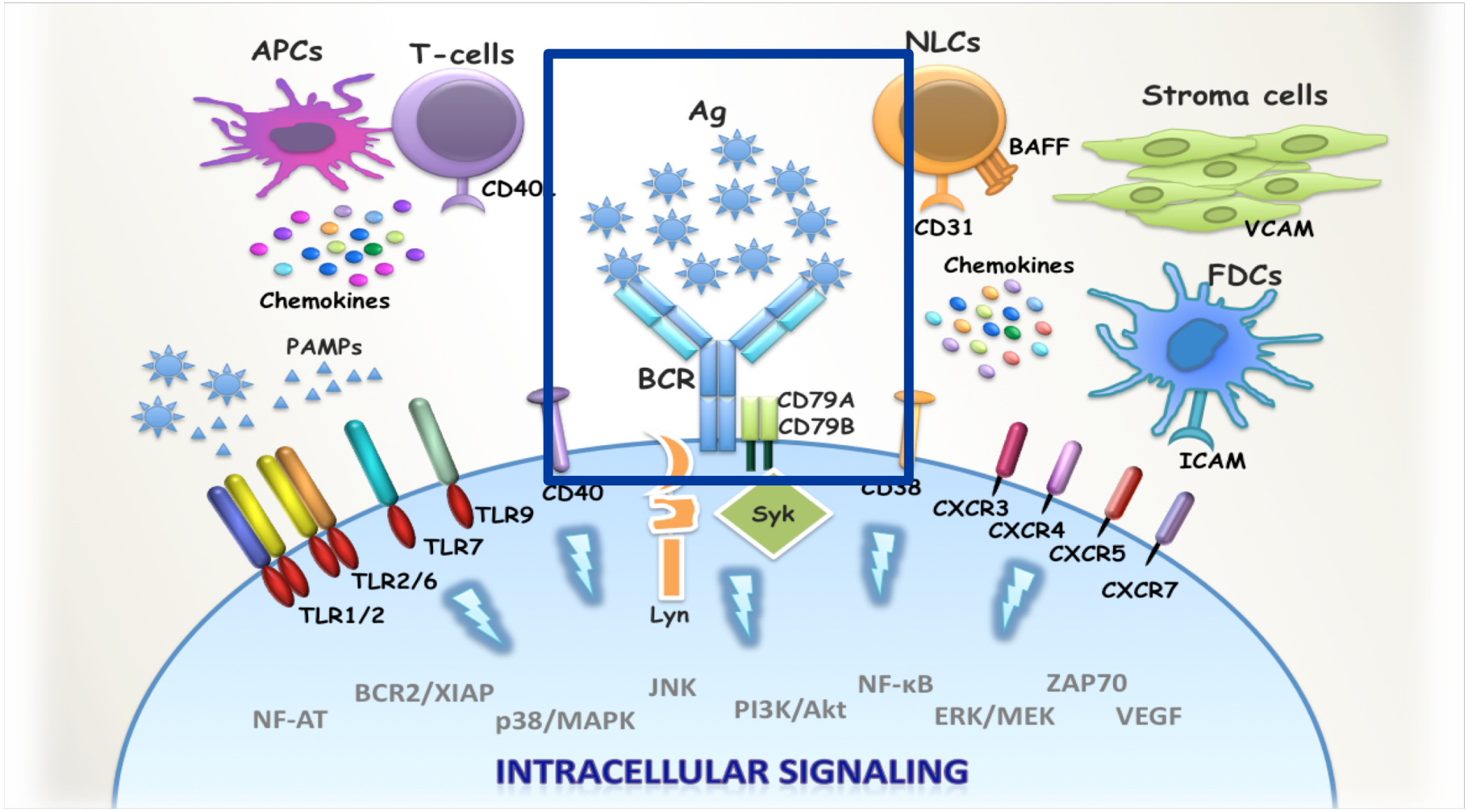
microenvironmental interactions

receptors



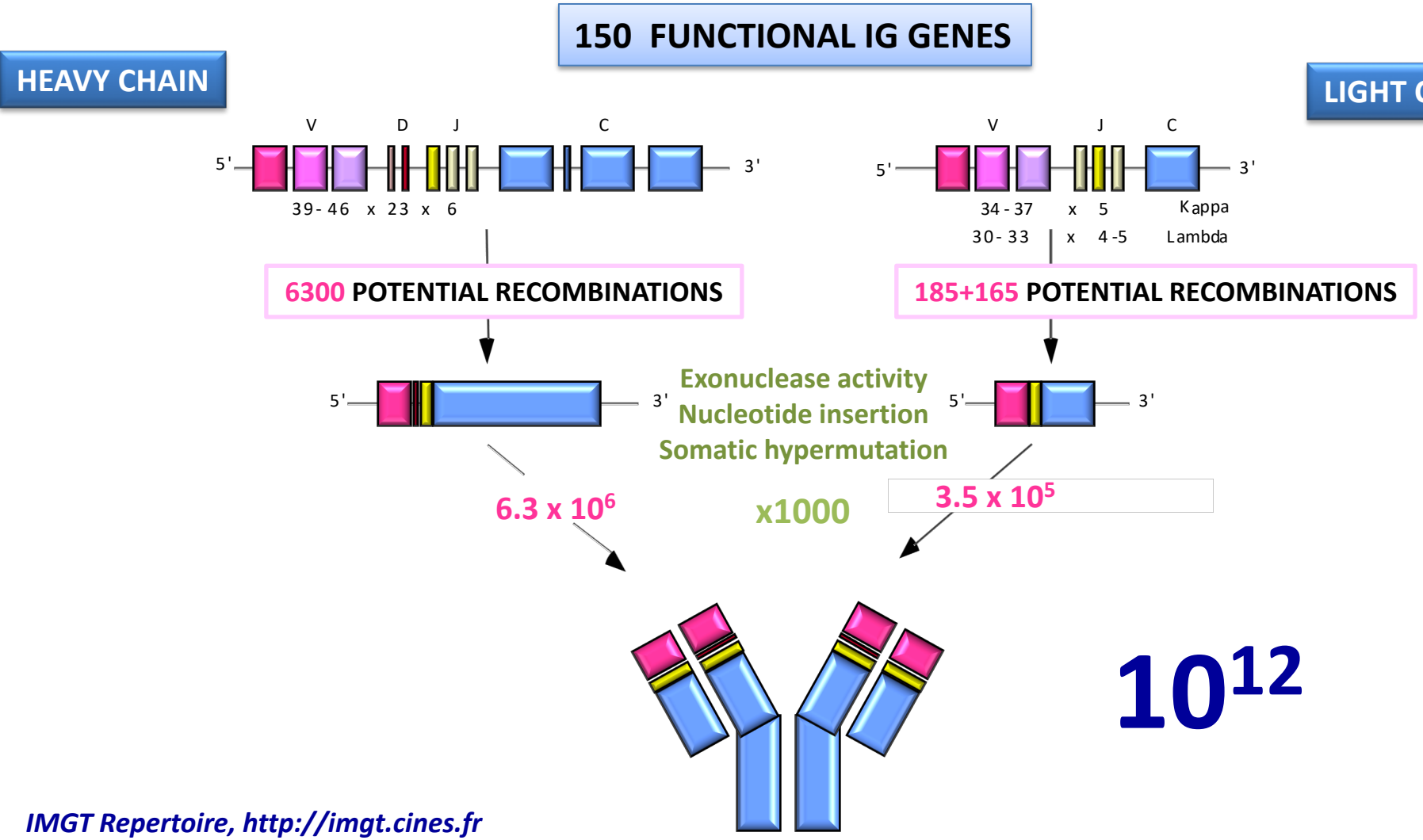
B cell receptor immunoglobulin

a unique molecular signature for every B cell clone



Logistics of IG synthesis

the LEGO approach



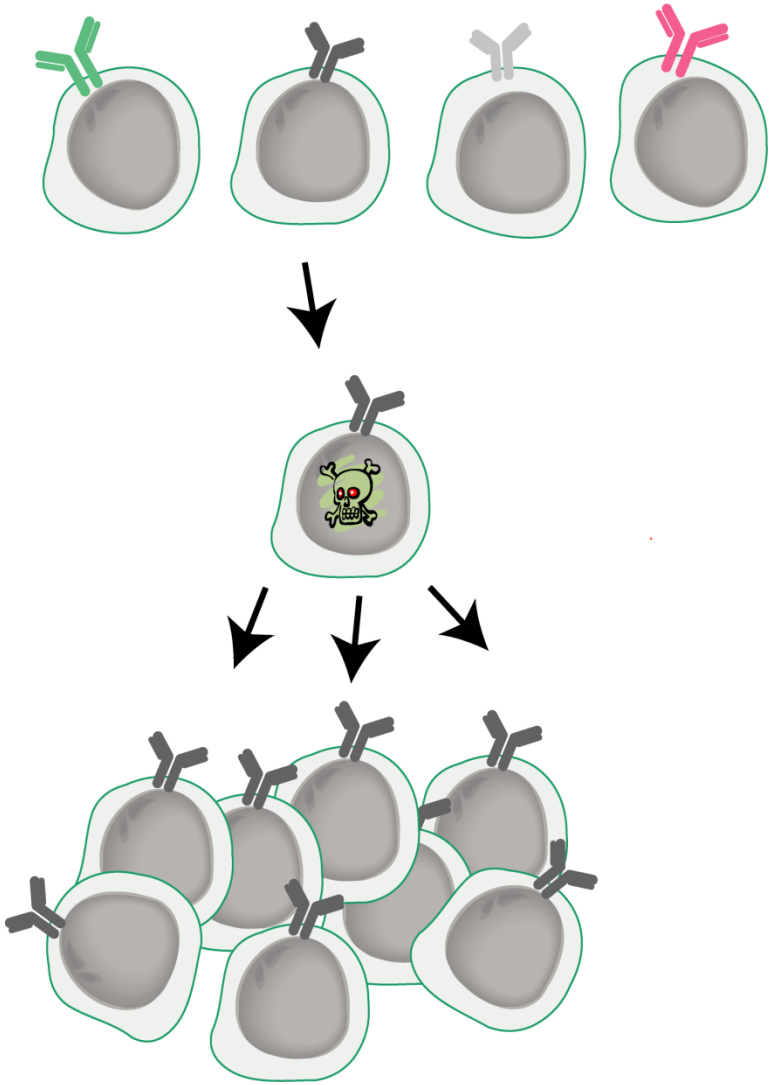
IMGT Repertoire, http://imgt.cines.fr

immunology and mathematics

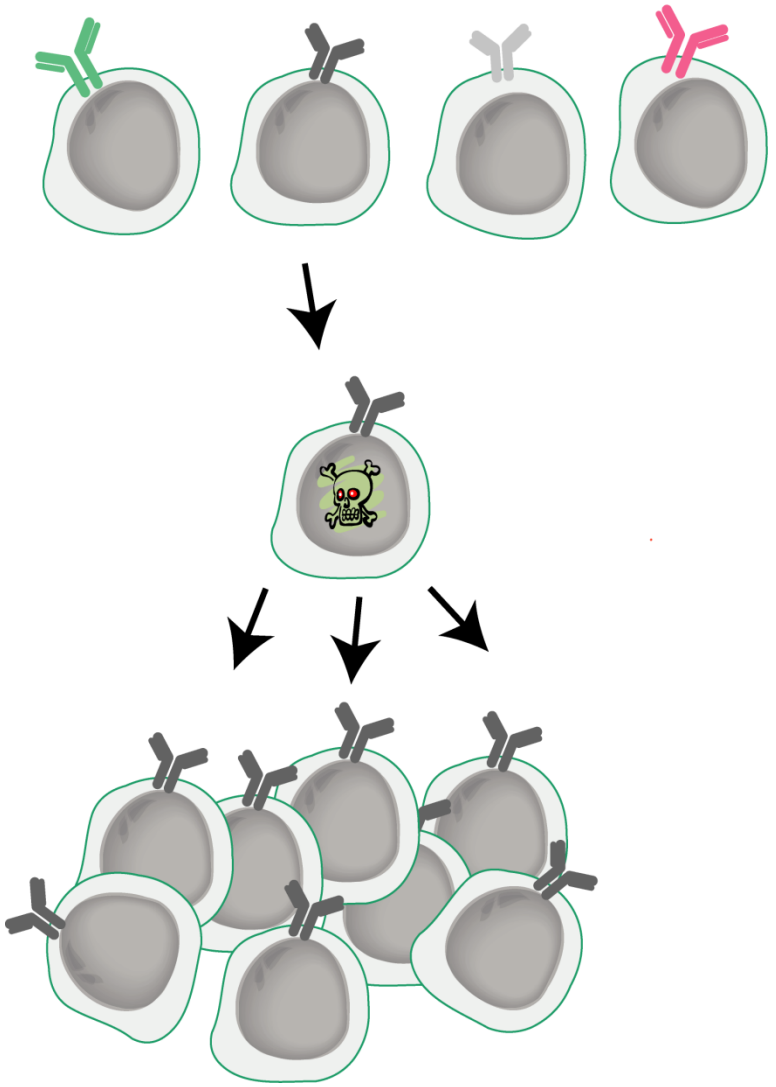
probability that ***two different B cell*** clones carry identical B cell receptors

$$1:10^{-12}$$

CLL is a clone of mature B cells



IG genes
a unique molecular identity for CLL



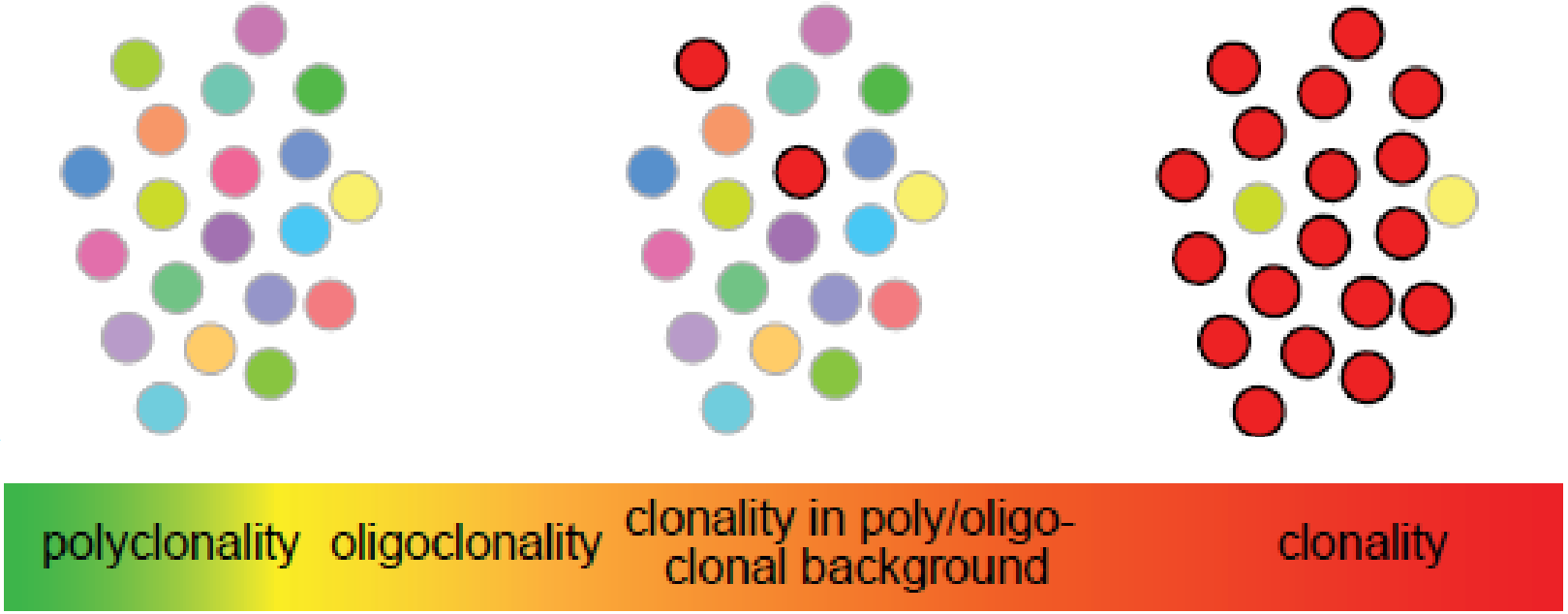
CLL = mature B cells

rearranged

IG gene loci

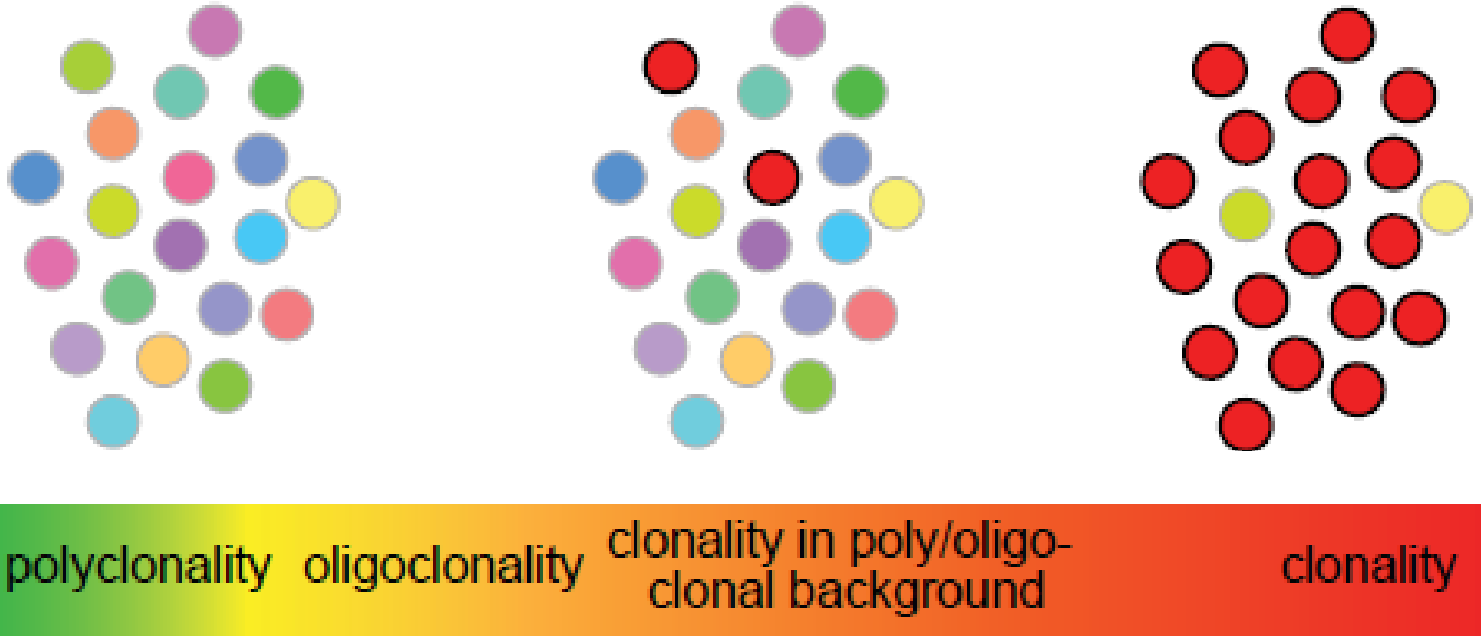
Immunogenetics
the study of IG | TR genes

IG repertoire, from diverse to restricted



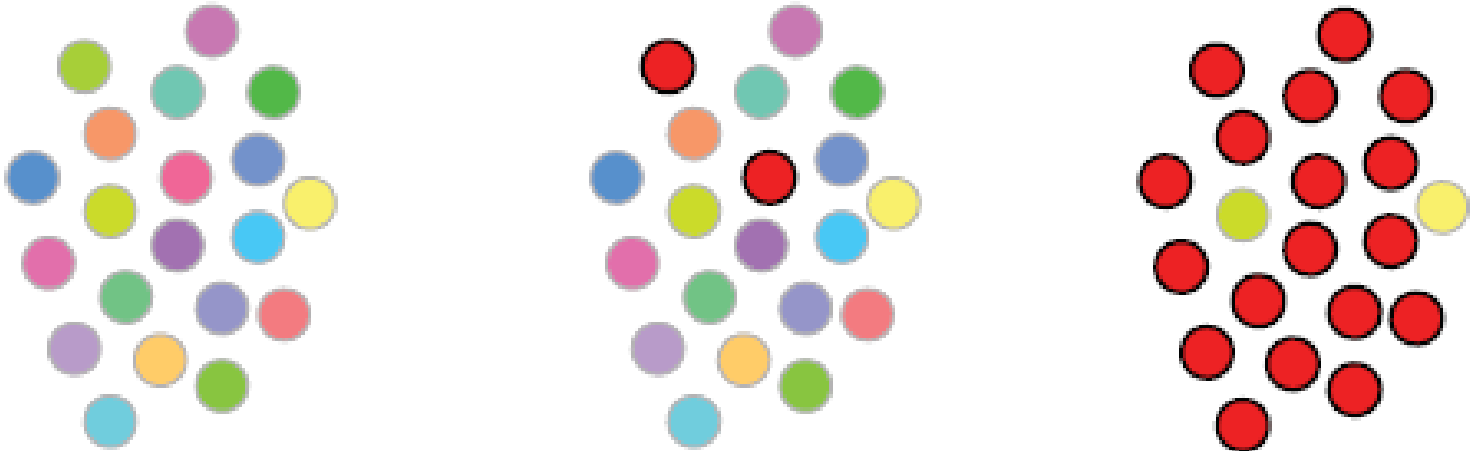
Langerak, J Immunol 2017

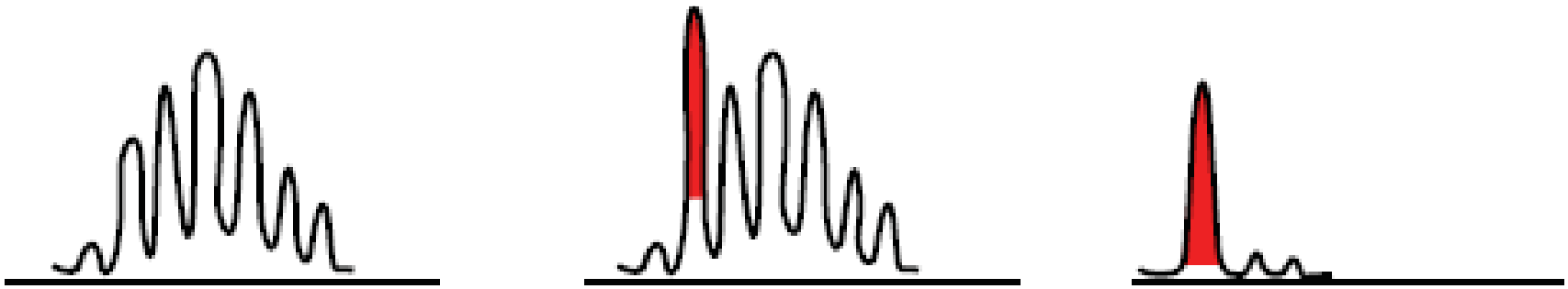
Immunogenetics



Langerak, J Immunol 2017

Immunogenetics





repertoire analysis

MRD monitoring

clonality assessment

Immunogenetics - challenges

General

- huge number of V / D / J genes to cover (multiplexing)

Clonality

- distinguishing clonal from polyclonal (via GeneScan spectratyping)

MRD

- high analytical sensitivity required (patient-specific ASO qPCR)

Repertoire

- unbiased; coverage of entire V gene (Sanger sequencing)

Standardization of optimized <u>low-throughput</u> assays in European networks

EuroClonality

Leukemia 2003;17:2257
Leukemia 2007;21:201
Leukemia 2012;26:2159

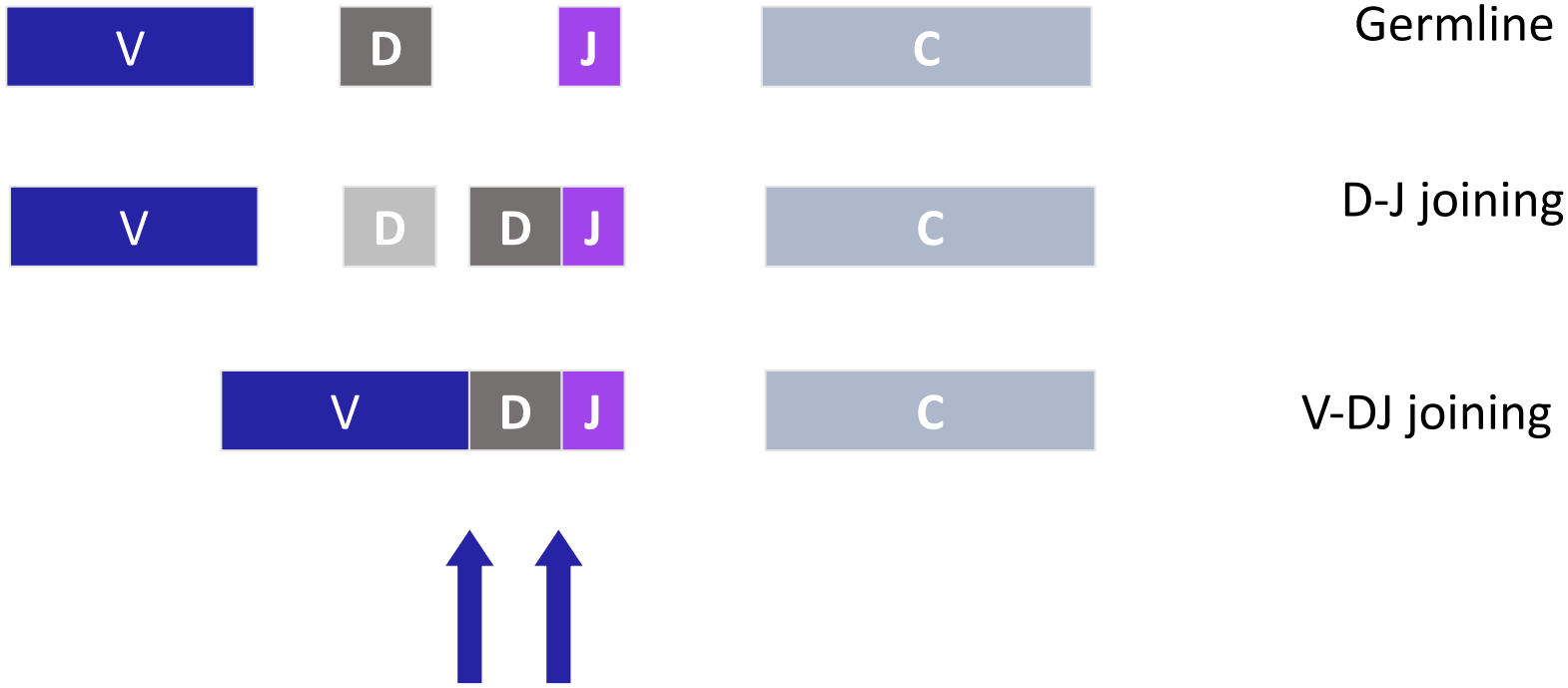
EuroMRD

Leukemia 2007;21:604
Leukemia 2008;22:771

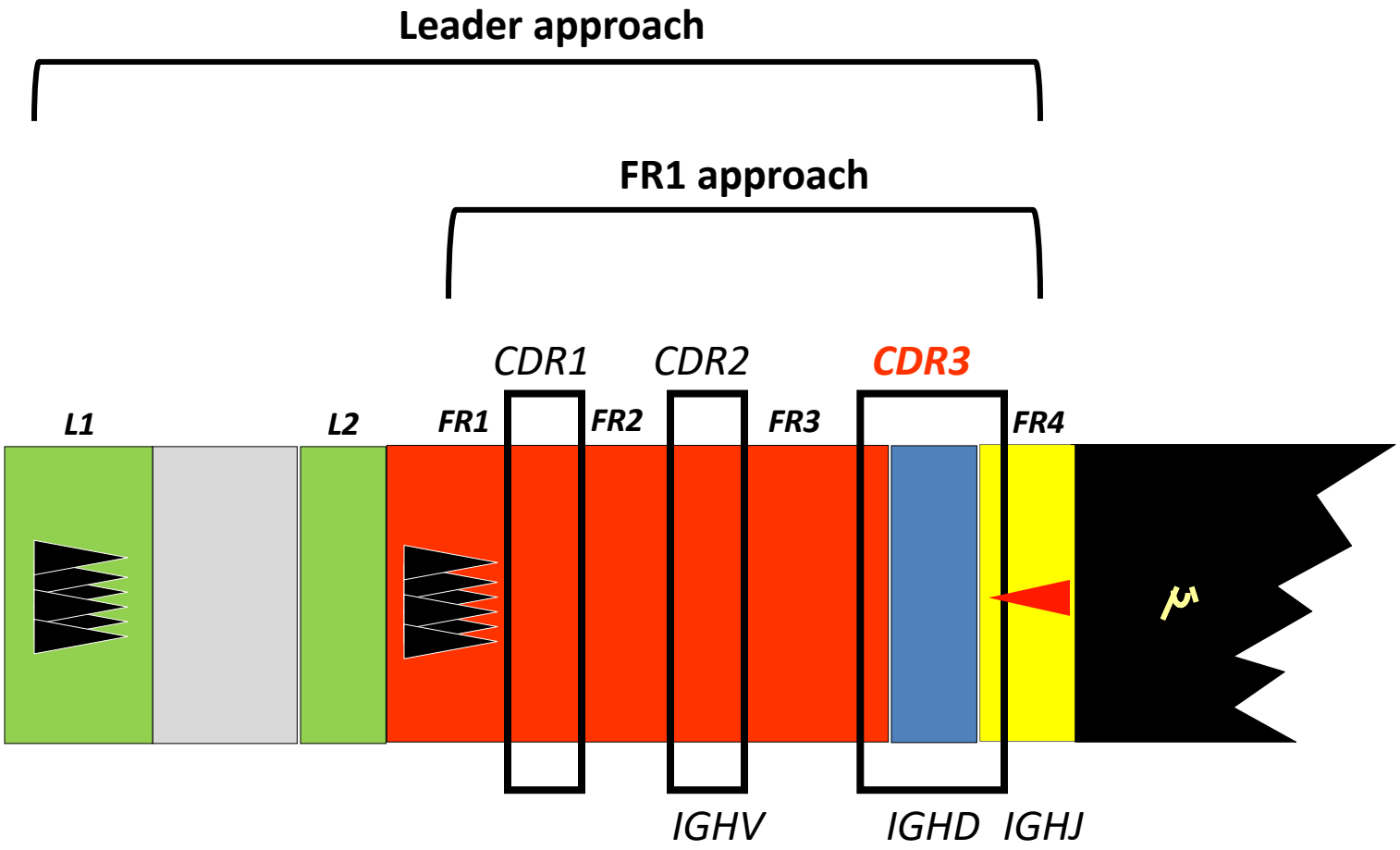
ERIC
european research initiative on CLL

Leukemia 2007;21:1
Leukemia 2011;25:979
Leukemia 2017;31:

IGHV-IGHD-IGHJ gene rearrangement

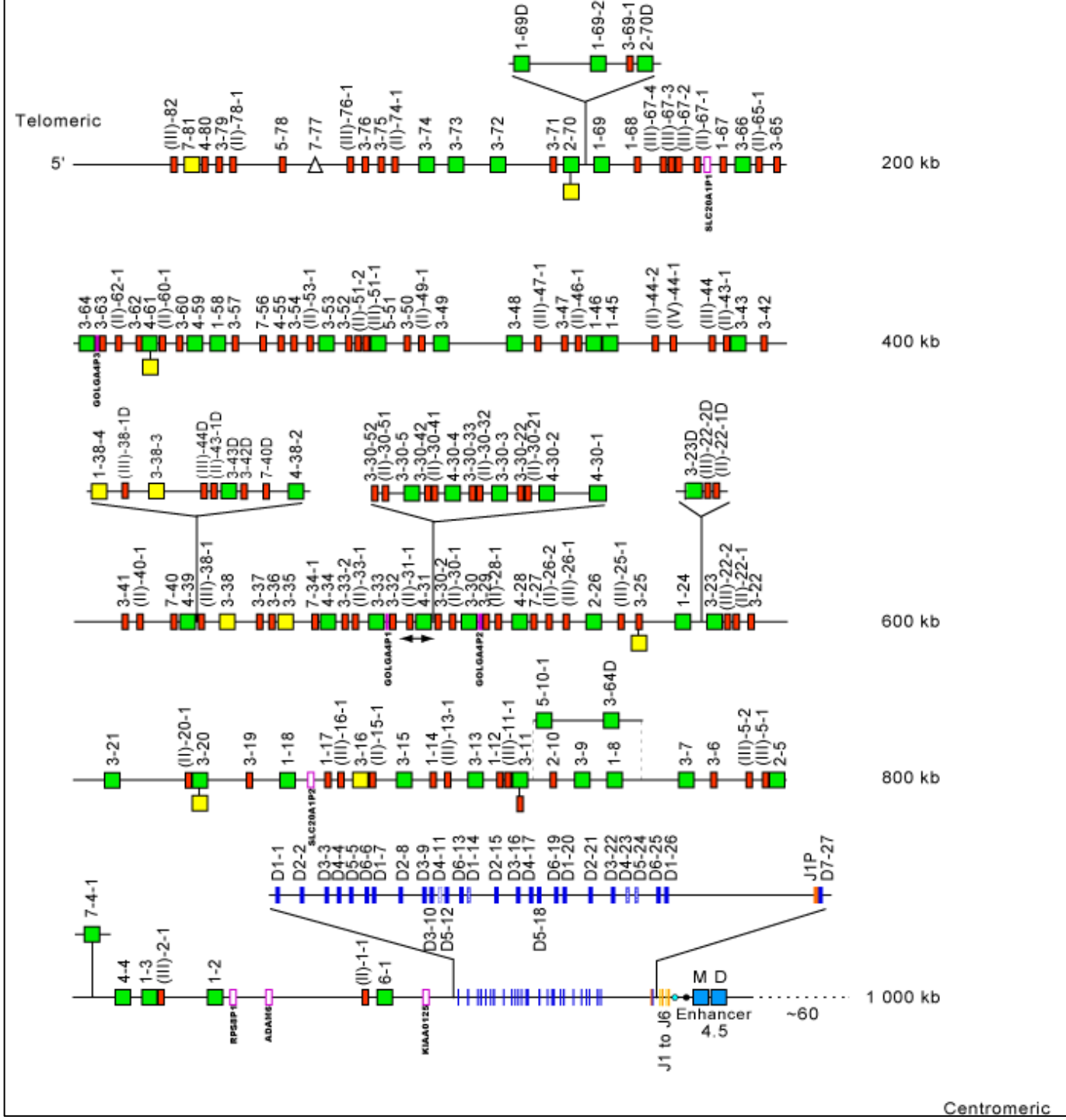


Analytical strategies: find the clonal signature



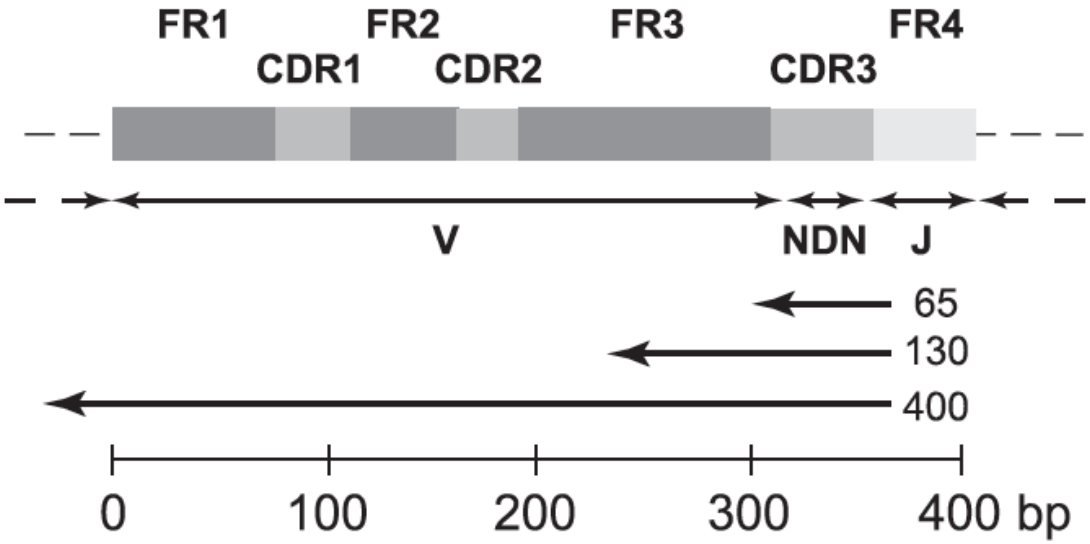
IGH locus

7 IGHV subgroups



Analytical strategies: shooting in the dark

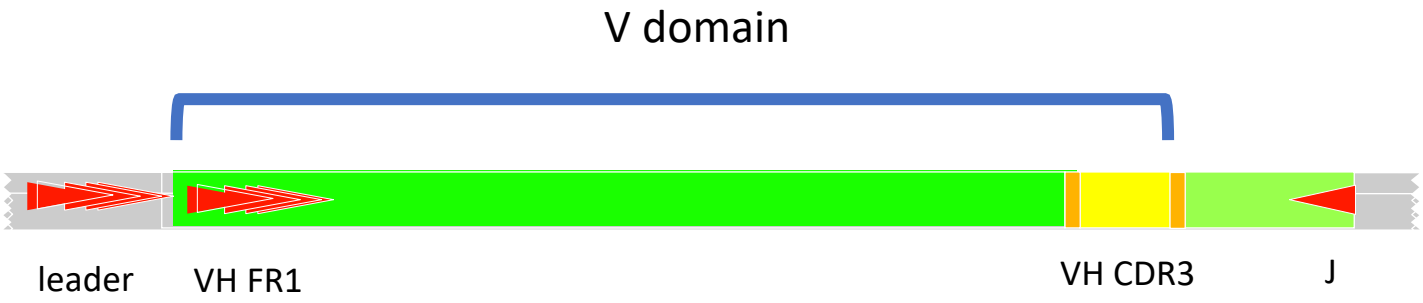
when you do not know your target in advance



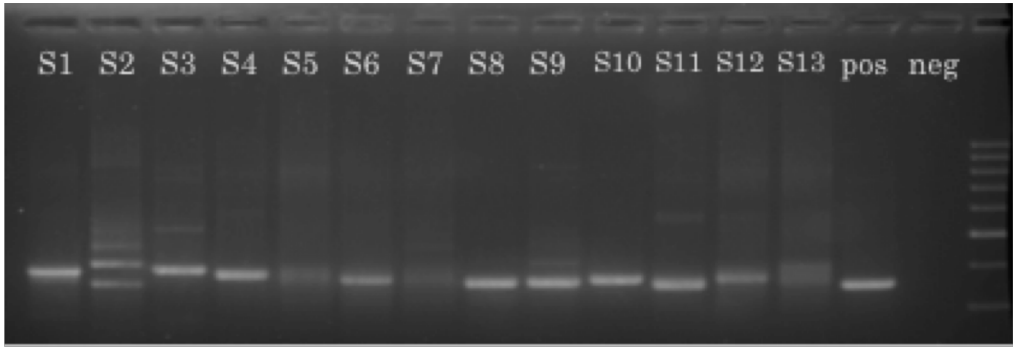
PCR analysis of IGH gene rearrangements

BIOMED-2 Concerted Action BMH4-CT98-3936: PCR-based clonality studies

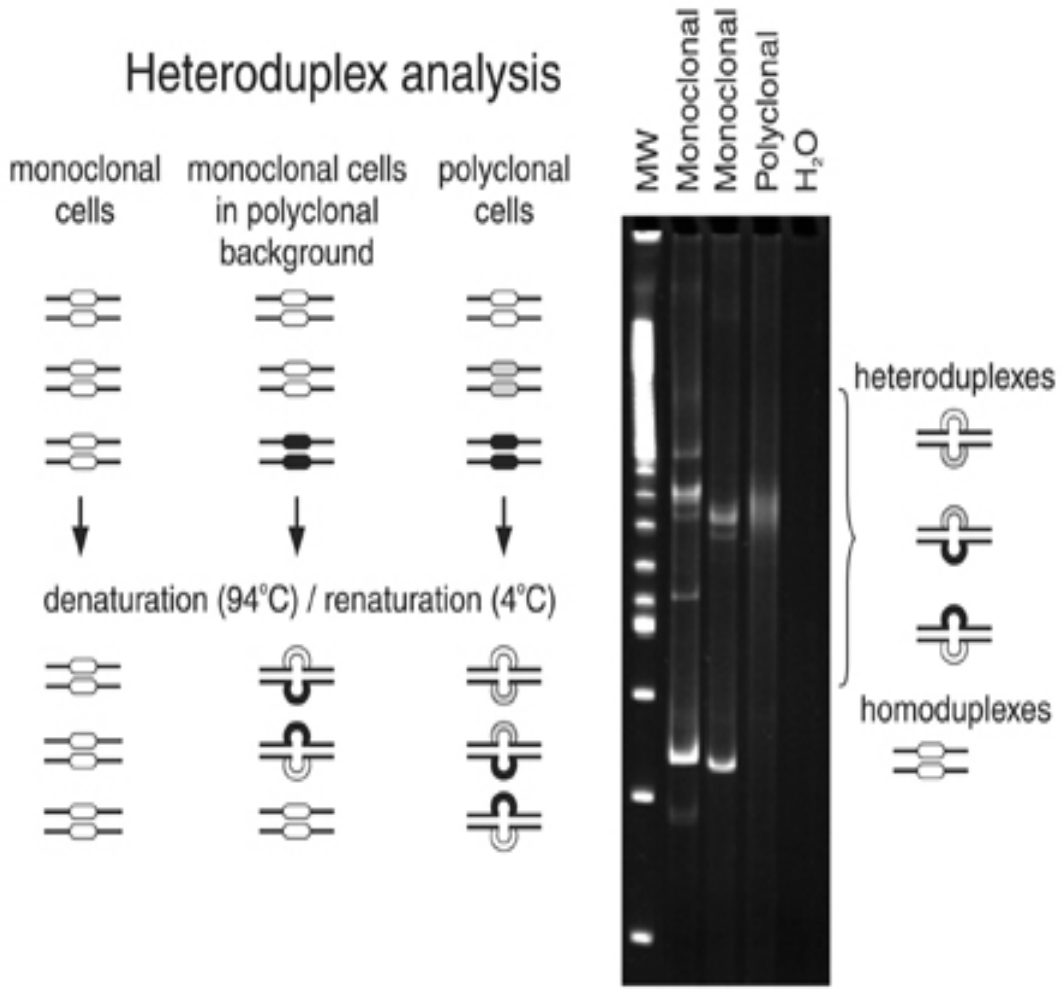
PCR amplification of IGHV-IGHD-IGHJ gene rearrangements

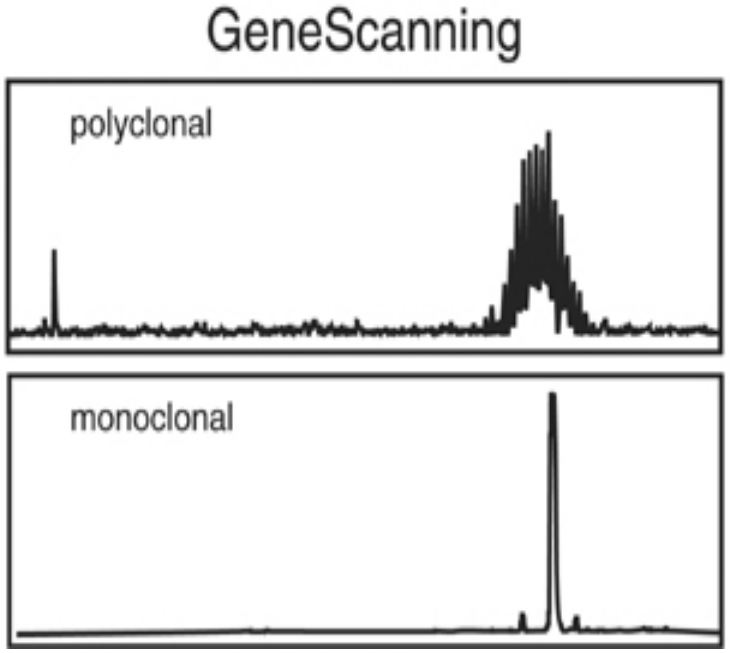


7 IGHV subgroups – a multiplex approach is required

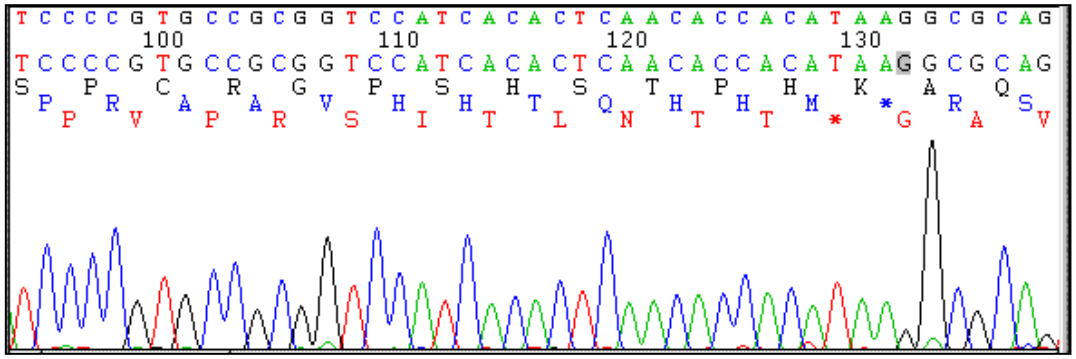


Assessment of clonality



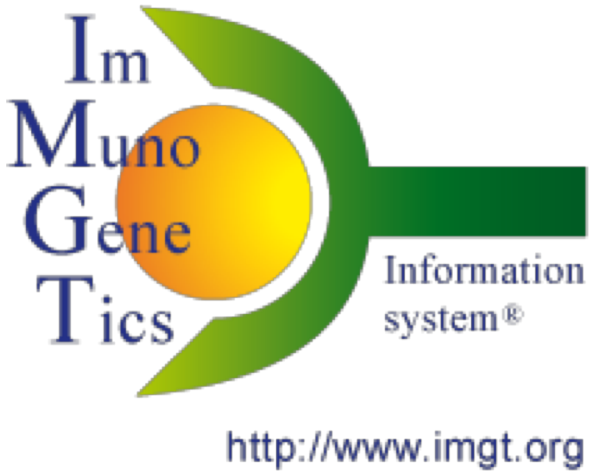


Sequence analysis and interpretation

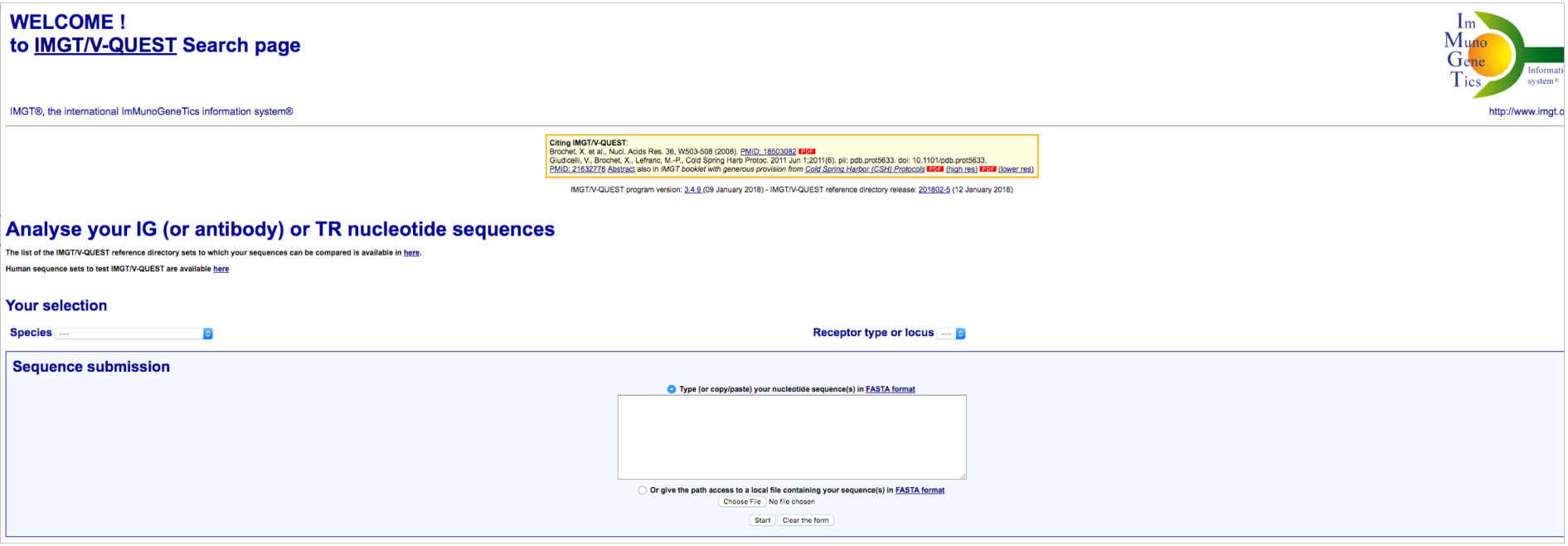


bidirectional sequencing

>P1276
CAGGTCCAGCTGGTGCAATCTGGGGCTGAGGTGAAGAAGCCTGGG
TCCTCGGTGAAGGTCTCCTGCAAGGCTTCTGGAGGCACCTTCAGCA
GCTATACTATCAGCTGGGTGCGACAGGCCCCTGGACAAGGGCTTGA
GTGGATGGGAAGGATCATCCCTATCCTTGGTATAGCAAACTACGCAC
AGAAGTTCCAGGGCAGAGTCACGATTACCGCGGACAAATCCACGA
GCACAGCCTACATGGAGCTGAGCAGCCTGAGATCTGAGGACACGG
CCGTGTATTACTGTGCGAGAGGTTACGATTTTTGGAGTGGTTACCGA
TACTGGGGCCAGGGAACCCTGGTCACCGTCTCCT



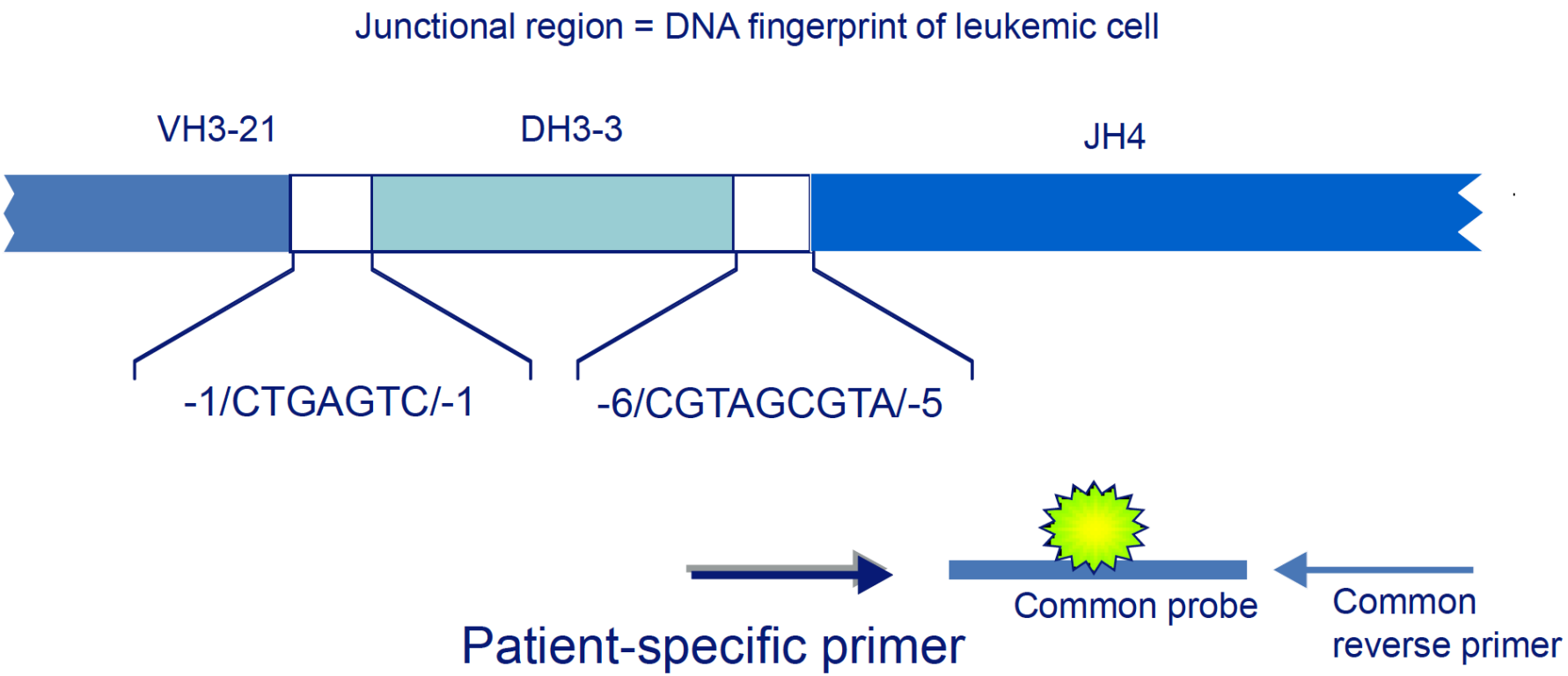
Sequence analysis and interpretation



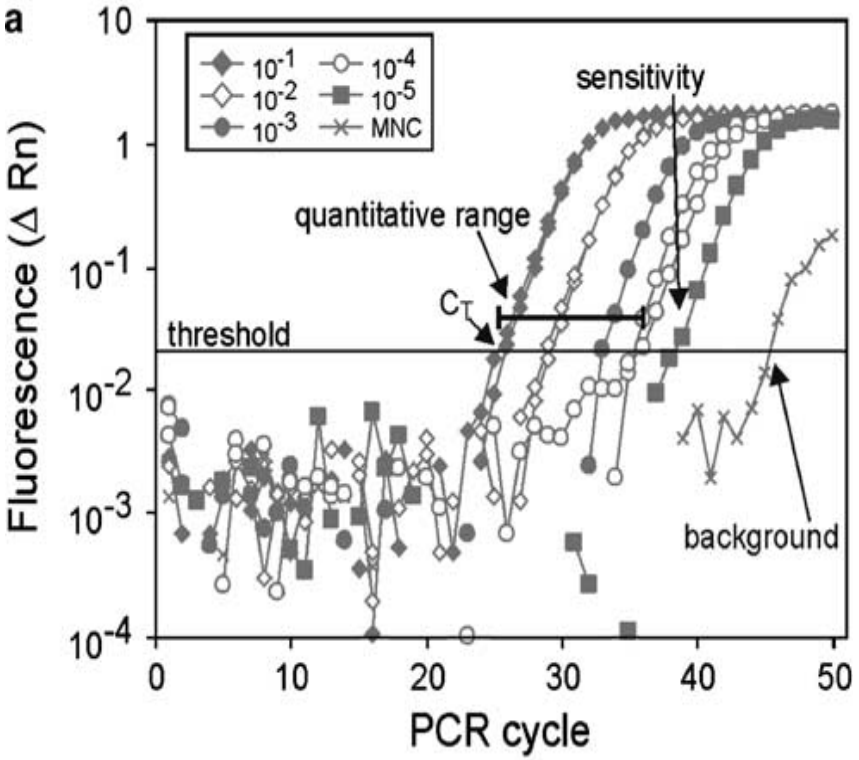
Result summary:	Productive IGH rearranged sequence: (no stop codon and in-frame junction)		
V-GENE and allele	Homsap IGHV4-34*01 F	score = 1411	identity = 99.65% (284/285 nt)
J-GENE and allele	Homsap IGHJ6*02 F	score = 291	identity = 98.33% (59/60 nt)
D-GENE and allele by IMGT/JunctionAnalysis	Homsap IGHD3-3*01 F	D-REGION is in reading frame 1	
FR-IMGT lengths, CDR-IMGT lengths and AA JUNCTION	[25.17.38.10]	[8.7.22]	CARGLPLLEWLLGPYYYYYGMDVW

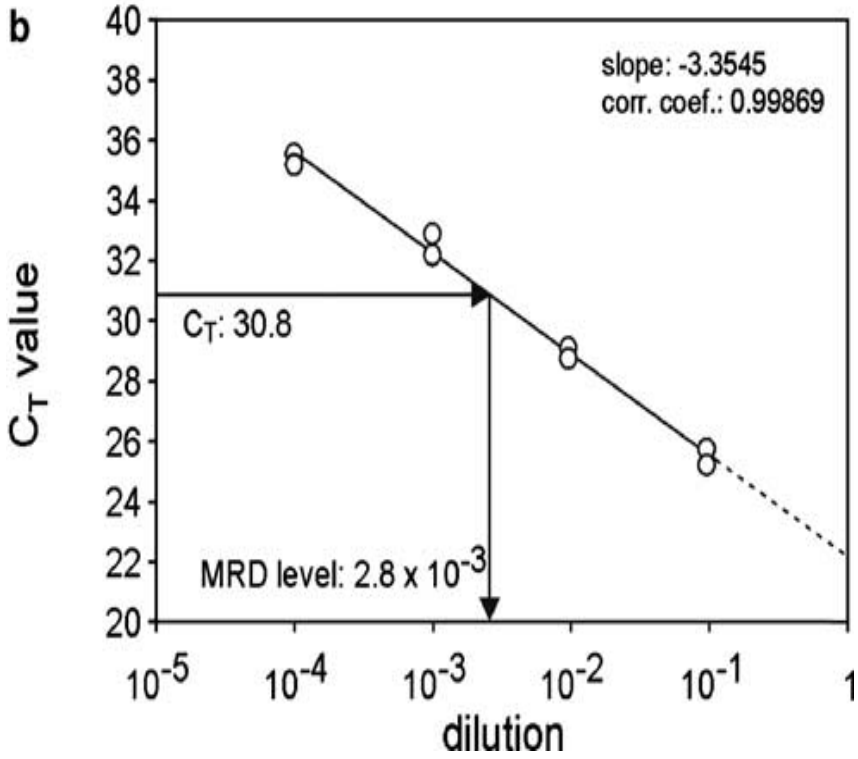
Allele-specific RQ-PCR

Allele-specific RQ-PCR



Allele-specific RQ-PCR





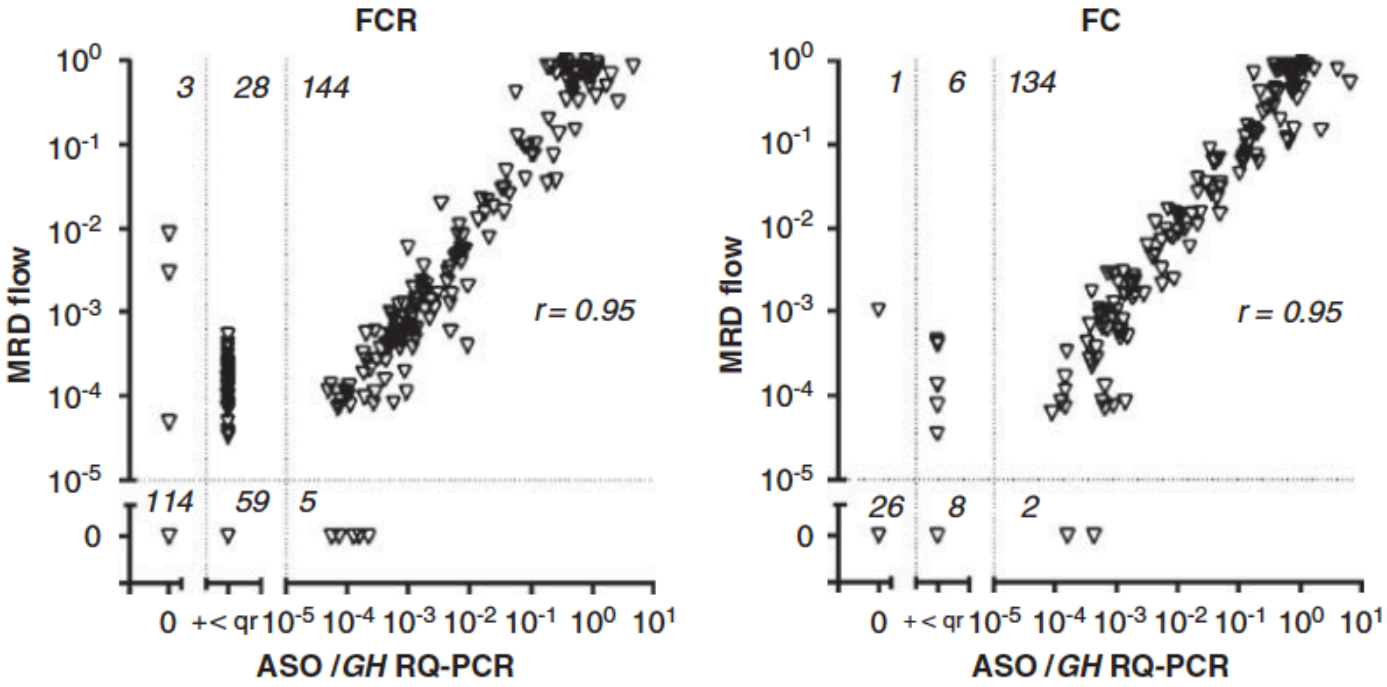
Leukemia (2009) 23, 2007–2017

www.nature.com/leu

ORIGINAL ARTICLE

Standardized MRD flow and ASO *IGH* RQ-PCR for MRD quantification in CLL patients after rituximab-containing immunochemotherapy: a comparative analysis

S Böttcher[1], S Stilgenbauer[2], R Busch[3], M Brüggemann[1], T Raff[1], C Pott[1], K Fischer[4], G Fingerle-Rowson[4], H Döhner[2], M Hallek[4], M Kneba[1] and M Ritgen[1] on behalf of the German CLL study group



Leukemia (2009) 23, 2007–2017

www.nature.com/leu

ORIGINAL ARTICLE

Standardized MRD flow and ASO *IGH* RQ-PCR for MRD quantification in CLL patients after rituximab-containing immunochemotherapy: a comparative analysis

S Böttcher[1], S Stilgenbauer[2], R Busch[3], M Brüggemann[1], T Raff[1], C Pott[1], K Fischer[4], G Fingerle-Rowson[4], H Döhner[2], M Hallek[4], M Kneba[1] and M Ritgen[1] on behalf of the German CLL study group

Quantitative MRD levels assessed by both techniques were closely correlated irrespective of therapy

Leukemia (2009) 23, 2007–2017

www.nature.com/leu

ORIGINAL ARTICLE

Standardized MRD flow and ASO *IGH* RQ-PCR for MRD quantification in CLL patients after rituximab-containing immunochemotherapy: a comparative analysis

S Böttcher[1], S Stilgenbauer[2], R Busch[3], M Brüggemann[1], T Raff[1], C Pott[1], K Fischer[4], G Fingerle-Rowson[4], H Döhner[2], M Hallek[4], M Kneba[1] and M Ritgen[1] on behalf of the German CLL study group

Discordant samples were typically negative by MRD flow and simultaneously positive close to the detection limit of the PCR assays, indicating a higher sensitivity of PCR for very low MRD levels.

Leukemia (2009) 23, 2007–2017

www.nature.com/leu

ORIGINAL ARTICLE

Standardized MRD flow and ASO *IGH* RQ-PCR for MRD quantification in CLL patients after rituximab-containing immunochemotherapy: a comparative analysis

S Böttcher[1], S Stilgenbauer[2], R Busch[3], M Brüggemann[1], T Raff[1], C Pott[1], K Fischer[4], G Fingerle-Rowson[4], H Döhner[2], M Hallek[4], M Kneba[1] and M Ritgen[1] on behalf of the German CLL study group

93.8% of all samples were concordantly classified by both methods using a threshold of 10^4 to determine MRD positivity.

Allele-specific RQ-PCR phases and steps

tedious and time consuming

MRD analysis of follow-up samples

- RQ-PCR analysis
- RQ-PCR data interpretation
- Calculation of MRD levels

Next generation sequencing

proof of principle studies by NGS

Clonality

- clonal relationship

- intra-clonal diversity

MRD

- target identification;

- sensitive monitoring

- subclonal heterogeneity

-clonal evolution

Repertoire (normal, clonal)

- depth and coverage

Measurement and Clinical Monitoring of Human Lymphocyte Clonality by Massively Parallel V-D-J Pyrosequencing

Scott D. Boyd,[1] Eleanor L. Marshall,[2] Jason D. Merker,[1,3] Jay M. Maniar,[2] Lyndon N. Zhang,[4] Bita Sahaf,[2] Carol D. Jones,[1] Birgitte B. Simen,[5] Bozena Hanczaruk,[5] Khoa D. Nguyen,[6] Kari C. Nadeau,[6] Michael Egholm,[5] David B. Miklos,[7] James L. Zehnder,[1,7] Andrew Z. Fire[1,2]*

(Published 23 December 2009; Volume 1 Issue 12 12ra23)

PNAS

High-throughput VDJ sequencing for quantification of minimal residual disease in chronic lymphocytic leukemia and immune reconstitution assessment

Aaron C. Logan[a], Hong Gao[b], Chunlin Wang[b], Bita Sahaf[a], Carol D. Jones[c], Eleanor L. Marshall[c], Ismael Buño[d], Randall Armstrong[e], Andrew Z. Fire[c], Kenneth I. Weinberg[f], Michael Mindrinos[b], James L. Zehnder[g], Scott D. Boyd[c], Wenzhong Xiao[b,h], Ronald W. Davis[b,1], and David B. Miklos[a,1]

Divisions of [a]Blood and Marrow Transplantation and [g]Hematology, Department of Medicine, [c]Department of Pathology, and [f]Division of Pediatric Stem Cell Transplantation, Department of Pediatrics, Stanford University School of Medicine, Stanford, CA 94305; [b]Stanford Genome Technology Center, Stanford, CA 94304; [d]Department of Hematology, Hospital General Universitario Gregorio Marañón, 28007 Madrid, Spain; [h]Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114; and [e]Stanford Cellular Therapeutics and Transplantation Laboratory, Stanford Hospital and Clinics, Stanford, CA 94305

PNAS

High throughput sequencing reveals a complex pattern of dynamic interrelationships among human T cell subsets

Chunlin Wang[a], Catherine M. Sanders[b], Qunying Yang[b], Harry W. Schroeder, Jr.[c], Elijah Wang[b], Farbod Babrzadeh[a], Baback Gharizadeh[a], Richard M. Myers[b], James R. Hudson, Jr.[b], Ronald W. Davis[a,1], and Jian Han[b,1]

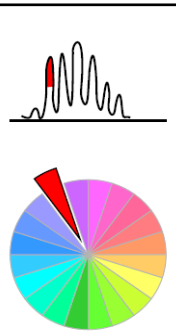
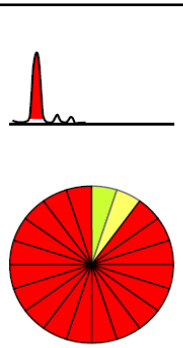
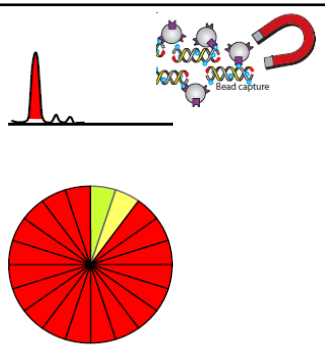
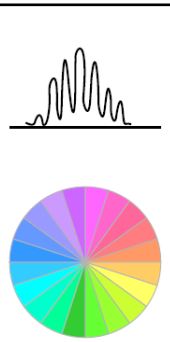
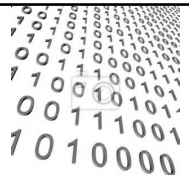
[a]Stanford Genome Technology Center, Palo Alto, CA 94304; [b]HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806; and [c]Departments of Medicine and Microbiology, University of Alabama at Birmingham, Birmingham, AL 35294

Contributed by Ronald W. Davis, December 8, 2009 (sent for review October 9, 2009)

NGS

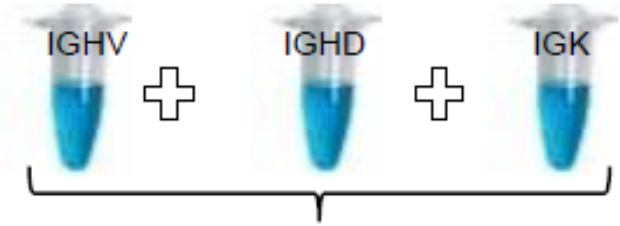
EuroClonality

EuroClonality-NGS core projects

Steering committee (coordinator: Langerak)				
Primer design	IGH V-J Pott Garcia Sanz	IGHD-J Davi Stamatopoulos	IGK Groenen Langerak	TRB Brüggemann Hummel TRG Cazzaniga TRD Macintyre
Project	1. MRD	2. Clonality (amplicon-based)	3. Clonality + transloc. (capture-based)	4. Repertoire (amplicon-based)
				
Leaders	Brüggem/Pott	Groenen, Hummel	Gonzalez	Stamatopoulos, Davi
	5. Bioinformatics Darzentas			

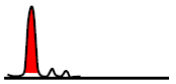
NGS-based clonality assessment

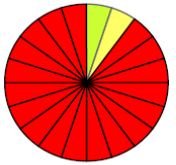
NGS EuroClonality



Multiplex PCR:
40 ng input DNA (Qubit)

Adaptor ligation and Library preparation
- Ion Plus Fragment Library Kit
- Ion Xpress Barcode adaptors
- Agencourt AMPure XP Beads





Sequencing by the local sequence facility
- Ion PGM template OT2 200
- Ion Chef (or Ion OneTouch 2 System)
  (4.5-6 M reads per 318 Chip)

Sequence analysis by using ARResT/Interrogate

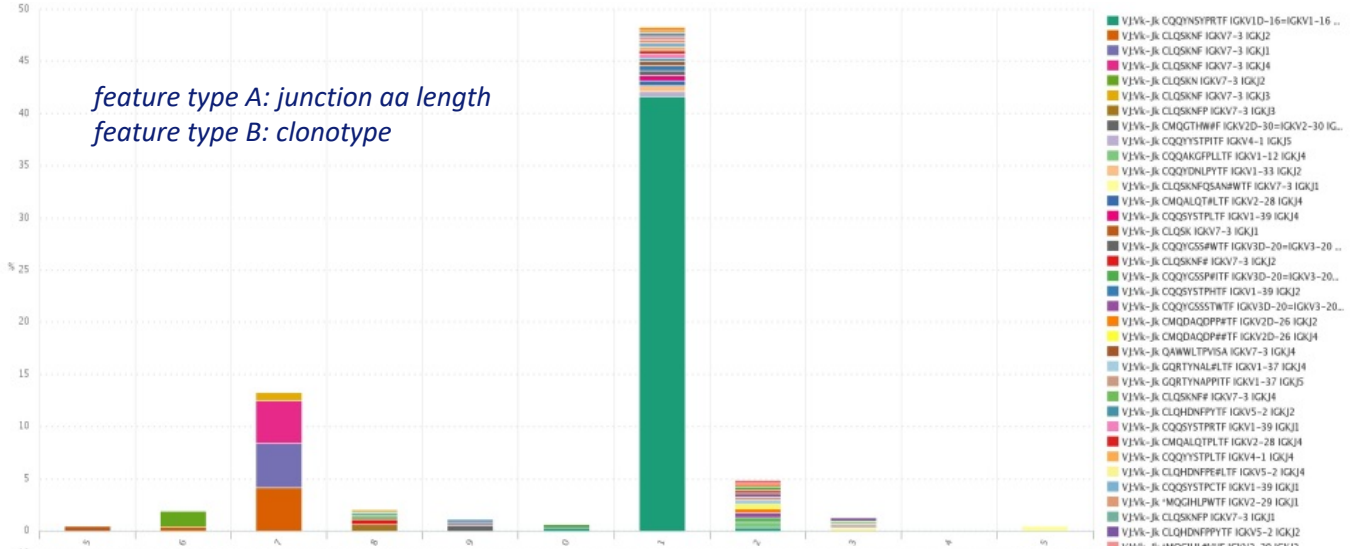
Courtesy: P. Groenen

Scheijen, under revision

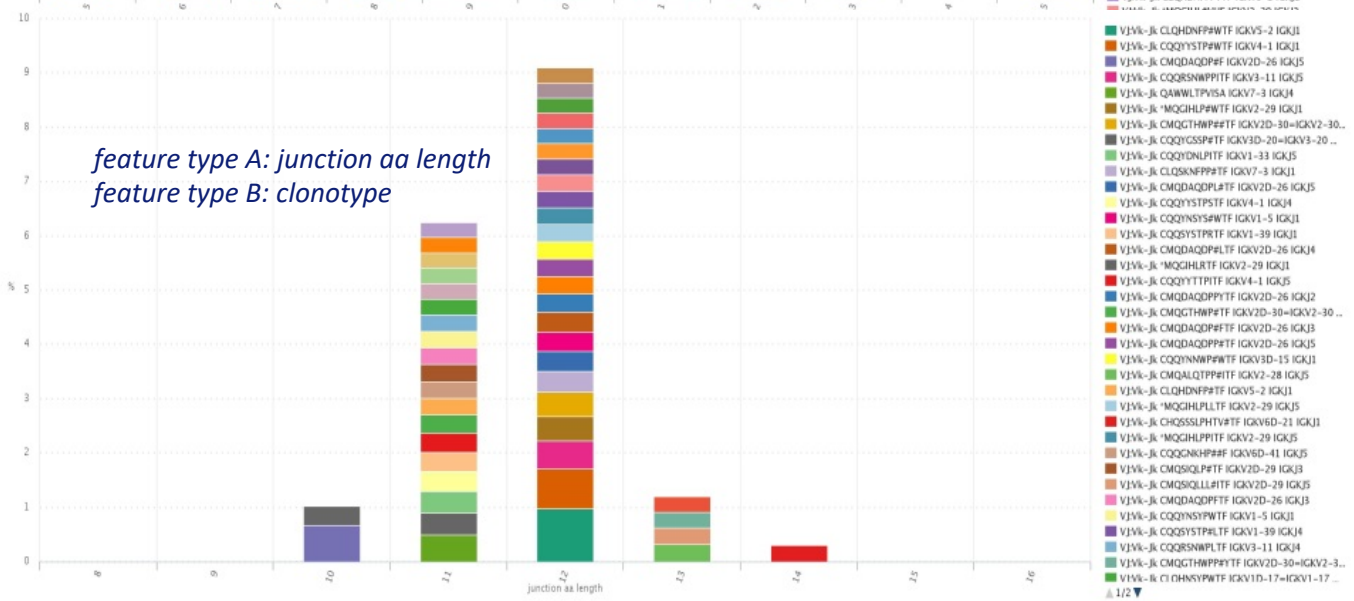
NGS-based clonality assessment

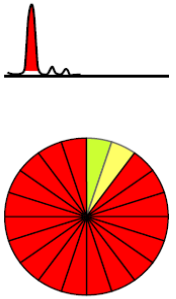
NGS EuroClonality

Sample: CLL



Sample normal



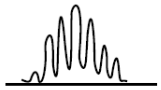


Software: ARResT

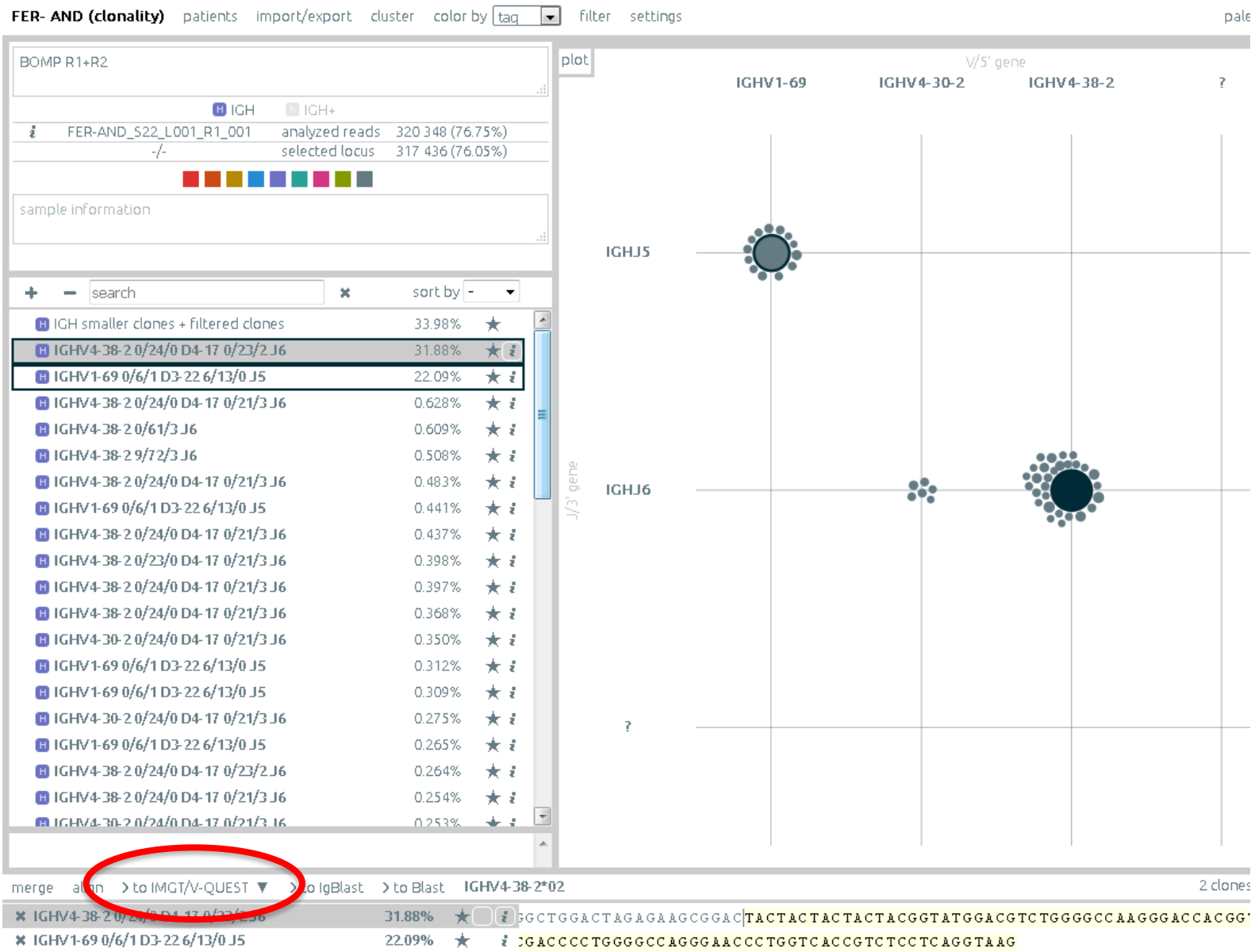
Courtesy: P. Groenen

NGS-based clonality assessment

NGS EuroClonality

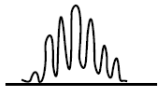


Software: Vidjil



NGS-based clonality assessment

NGS EuroClonality



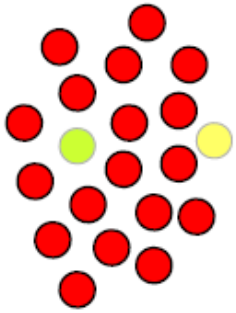
Software: ARResT

Courtesy: F. Davi

MRD: NGS-based IG/TR workflow

NGS EuroClonality

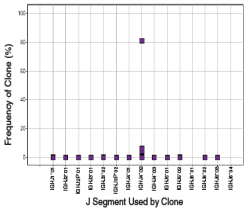
Screening of diagnostic sample



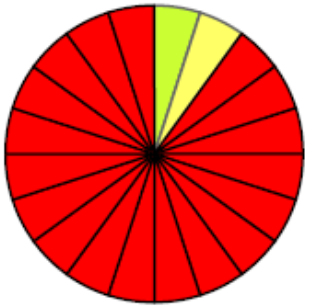
IG/TR Multiplex PCR

Preparation of amplicons for NGS

NGS



Bioinformatic identification of index-sequence



Brüggemann, Leukemia 2019

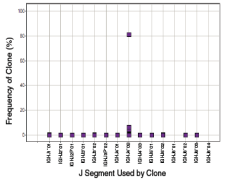
MRD-Analysis

Courtesy: M. Brueggemann

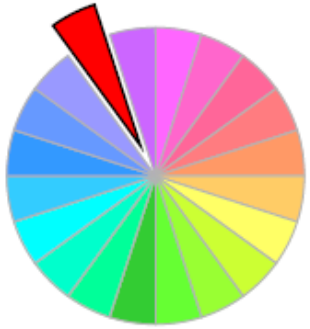
IG/TR Multiplex PCR

Preparation of amplicons for NGS

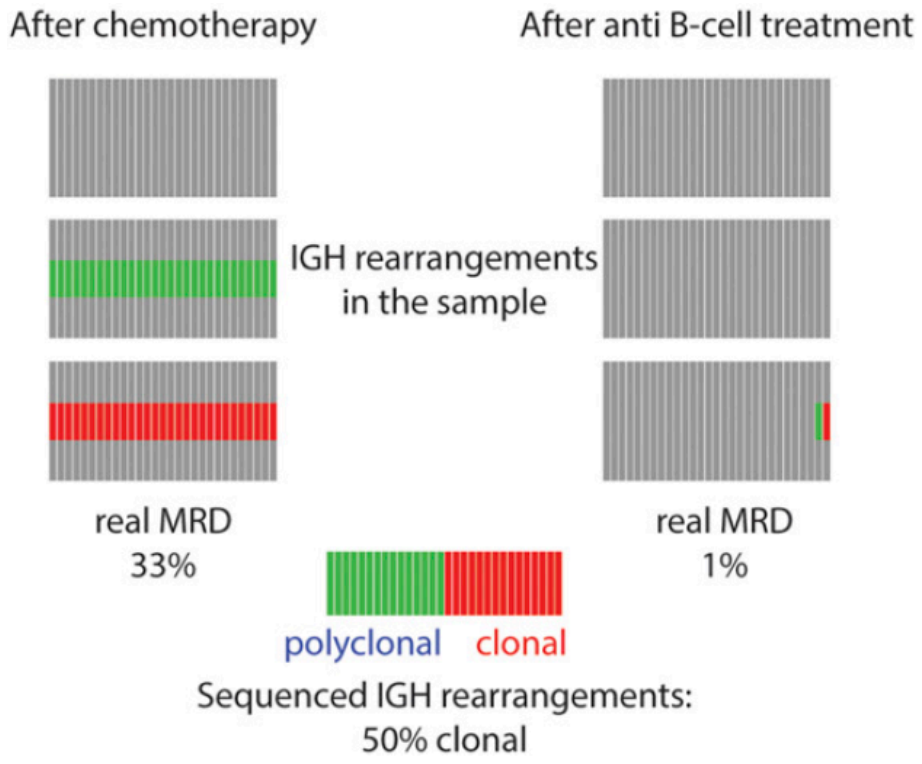
NGS



Bioinformatic search for index-sequence



Correct MRD quantification is dependent on the background level of polyclonal B lymphocytes

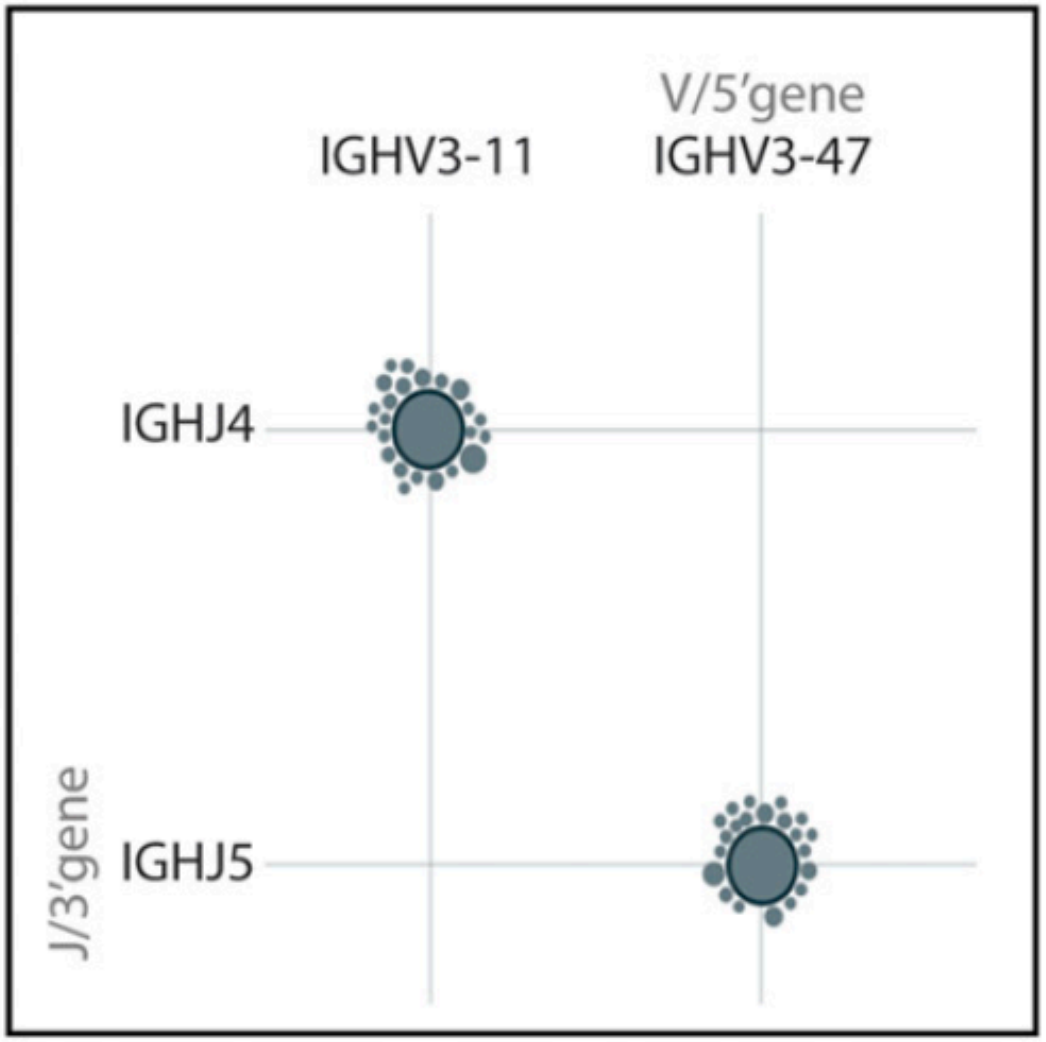


Cell type	% IGH reads	% MRD of B cells	% MRD of cells	
			post Chemo	post CD20
Leukemic B	50%	50%	33%	1%
Normal B	50%	50%	33%	1%
Other	n.a.	n.a.	33%	98%

MRD levels may greatly differ following chemotherapy versus B cell depletion therapy, yet might give rise to the same relative frequency of index sequence

use internal references for accurately calculating MRD levels

Satellite clonotypes



fact or artefact?

Work in progress

NGS EuroClonality

Clonality testing	Multiplexing and complementarity of IG/TR targets
	Reappraisal / redefinition of the term clonality
	Defining the value of clonal size and numerical cut-off value
	Data analysis pipeline incl. visualization

Langerak, J Immunol 2017

Work in progress

NGS EuroClonality

MRD monitoring	Multiplexing and complementarity of IG/TR targets
	Amount and type of starting material
	Use of internal controls (spike-ins) for quantitation
	Definition of limits of detection (quantifiable, sensitivity)
	Correct for disproportional PCR amplification of rearrangements
	Data analysis pipeline incl. visualization

Langerak, J Immunol 2017

Work in progress

NGS EuroClonality

Repertoire analysis	Multiplexing and complete coverage of genes
	Equal amplification (thus, representation) of genes
	Sequence information of entire V gene (IG loci) → long reads
	Error correction prior to accurately defining mutations, polymorphisms
	Data analysis pipeline incl. visualization

Langerak, J Immunol 2017

causes for optimism

MRD by IG NGS

blood®

Prepublished online December 28, 2017;
doi:10.1182/blood-2017-09-806521

Measurable residual disease detection by high throughput sequencing improves risk stratification for pediatric B-ALL

Brent Wood, David Wu, Beryl Crossley, Yunfeng Dai, David Williamson, Charles Gawad, Michael J. Borowitz, Meenakshi Devidas, Kelly W. Maloney, Eric Larsen, Naomi Winick, Elizabeth Raetz, William L. Carroll, Stephen P. Hunger, Mignon Loh, Harlan Robins and Ilan Kirsch

blood®
Prepublished online December 28, 2017;
doi:10.1182/blood-2017-09-806521

Measurable residual disease detection by high throughput sequencing improves risk stratification for pediatric B-ALL

Brent Wood, David Wu, Beryl Crossley, Yunfeng Dai, David Williamson, Charles Gawad, Michael J. Borowitz, Meenakshi Devidas, Kelly W. Maloney, Eric Larsen, Naomi Winick, Elizabeth Raetz, William L. Carroll, Stephen P. Hunger, Mignon Loh, Harlan Robins and Ilan Kirsch

The higher analytic sensitivity and lower false negative rate of NGS improves upon flow cytometry for measurable residual disease detection in pediatric B ALL

NGS identifies a novel subset of patients at end of induction who are essentially cured using current chemotherapy

causes for optimism in CLL

european research initiative on CLL

OPEN

Leukemia (2016) **30,** 929–936

www.nature.com/leu

ORIGINAL ARTICLE

A complementary role of multiparameter flow cytometry and high-throughput sequencing for minimal residual disease detection in chronic lymphocytic leukemia: an European Research Initiative on CLL study

AC Rawstron[1], C Fazi[2], A Agathangelidis[2], N Villamor[3], R Letestu[4], J Nomdedeu[5], C Palacio[6], O Stehlikova[7], K-A Kreuzer[8], S Liptrot[9], D O'Brien[9], RM de Tute[1], I Marinov[10], M Hauwel[11], M Spacek[12], J Dobber[13], AP Kater[13], P Gambell[14], A Soosapilla[15], G Lozanski[16], G Brachtl[17,18], K Lin[19], J Boysen[20], C Hanson[20], JL Jorgensen[21], M Stetler-Stevenson[22], C Yuan[22], HE Broome[23], L Rassenti[23], F Craig[24], J Delgado[3], C Moreno[5], F Bosch[6], A Egle[17], M Doubek[7], S Pospisilova[7], S Mulligan[25], D Westerman[14], CM Sanders[26], R Emerson[26], HS Robins[26], I Kirsch[26], T Shanafelt[20], A Pettitt[19], TJ Kipps[23], WG Wierda[21], F Cymbalista[4], M Hallek[8], P Hillmen[27], E Montserrat[3], and P Ghia[2,28] on behalf of ERIC (European Research Initiative on CLL)

Sample types

51 CLL, 6 HC-MBL
available IGHV-IGHD-IGHJ gene sequence from Sanger analysis

Factors taken into consideration for case selection
representation of 'difficult' cases

CDR3 features | e.g. very short CDR3s
cases with multiple rearrangements (2-3)
cases with single unproductive rearrangement
different timepoints

sensitivity
serial dilutions, starting from 1:10 up to 10^{-6} from 3 CLL patient samples

Methodology

Material: 400 ng or 6-7 µg DNA from PB or BM samples from patients with CLL before and after treatment, respectively.

Amplification and sequencing
ClonoSEQ™ platform | Adaptive Biotechnologies, Seattle, WA

Primer mix | IGHV and IGHJ genes

Sequencing starts at 3’ end of the IGHJ gene and extends 87 bases upstream

Data analysis: HighV-Quest | IMGT

Metadata analysis and interpretation: proprietary software

NGS vs Sanger

european research initiative on CLL

NGS

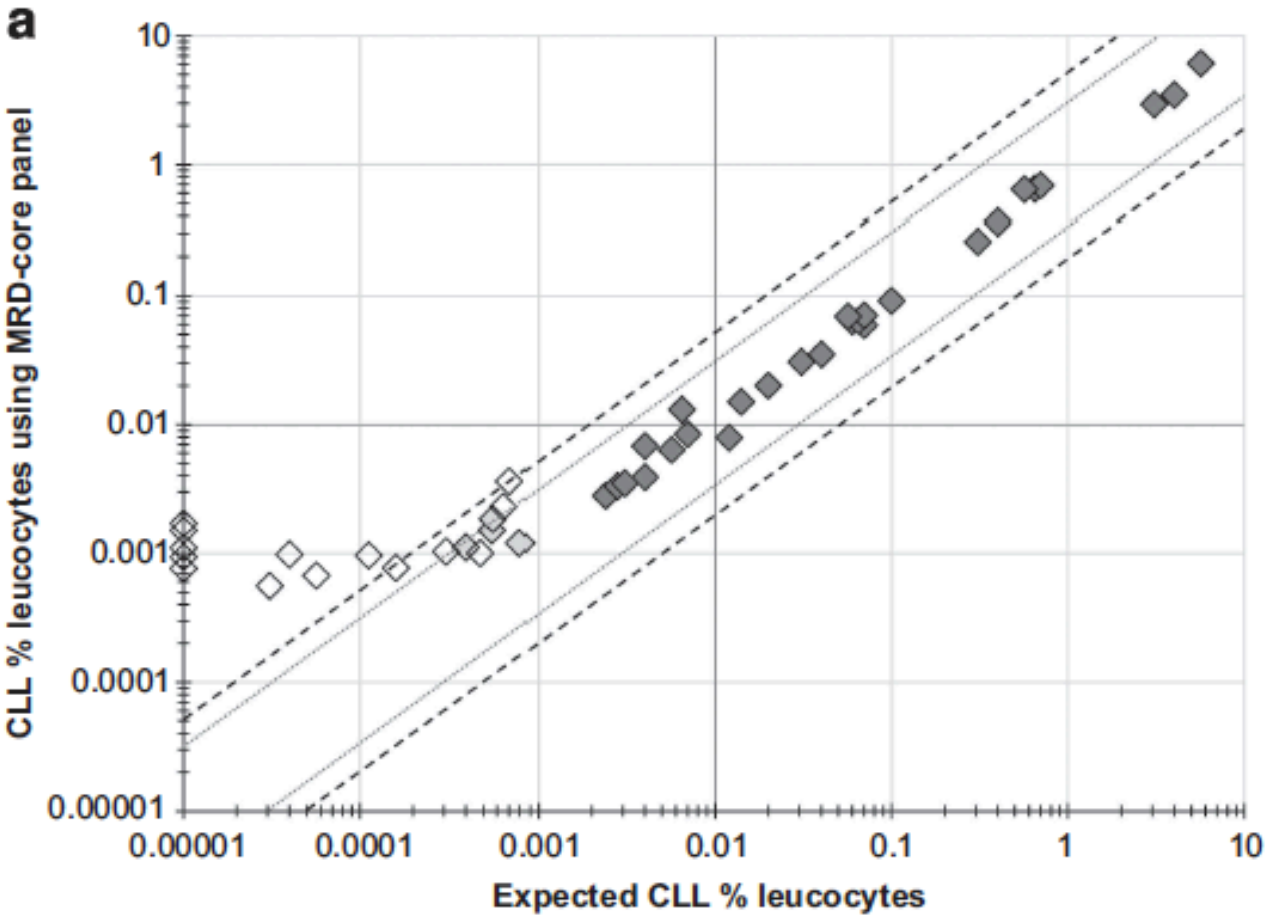
detection of a dominant clonotype in 57/57 samples (100%)
relative frequency of the dominant clonotype: 29-100% (avg, 89%)
detection of a dominant clonotype in samples with multiple rearrangements
concordant cases in samples with serial pre-treatment samples

Sanger

Detection of a dominant clonotype in 51/57 samples (91%)

Six-color flow cytometry

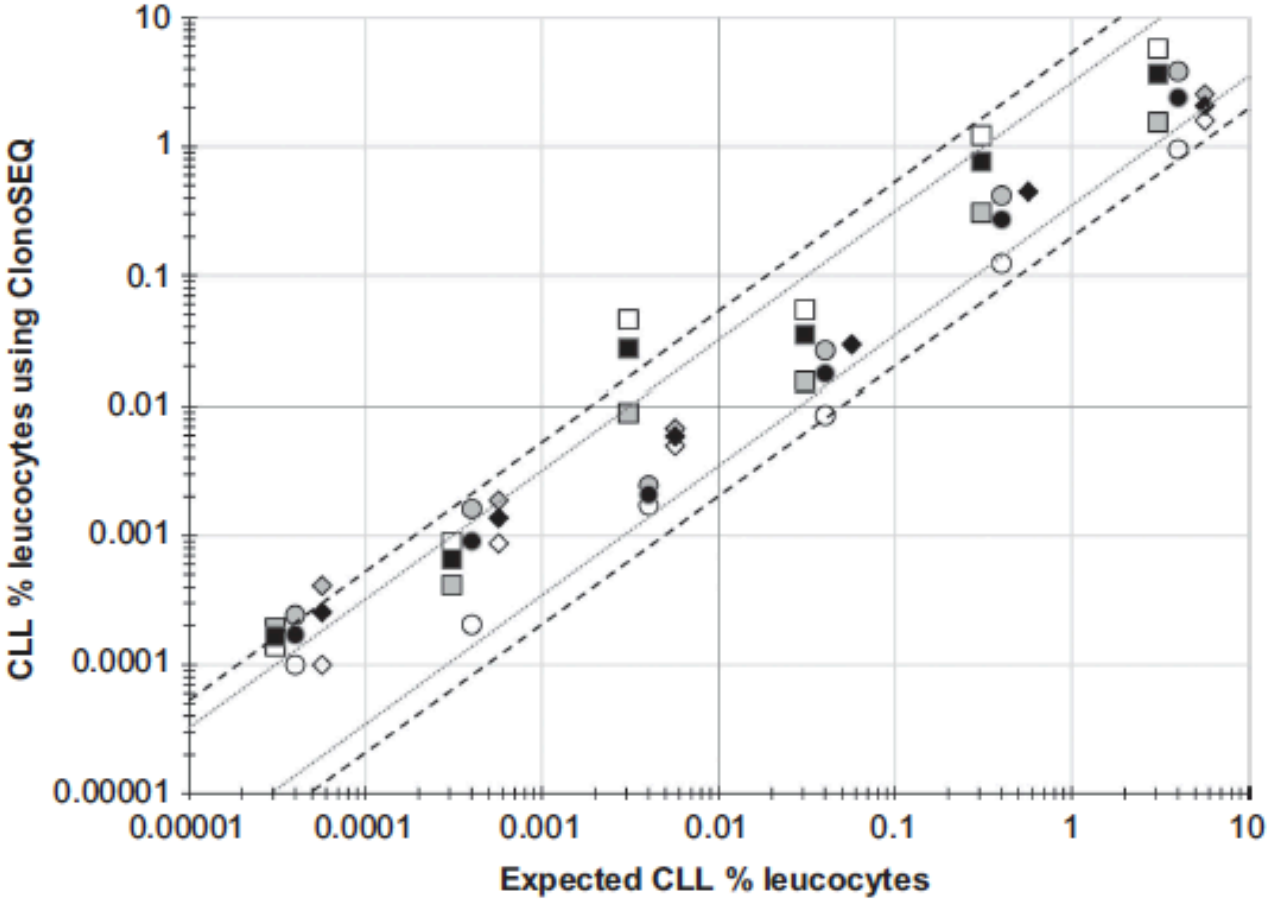
reliable quantitation of CLL cells to the level of 0.0010% (10^{-5})



Rawstron et al. Leukemia 2016

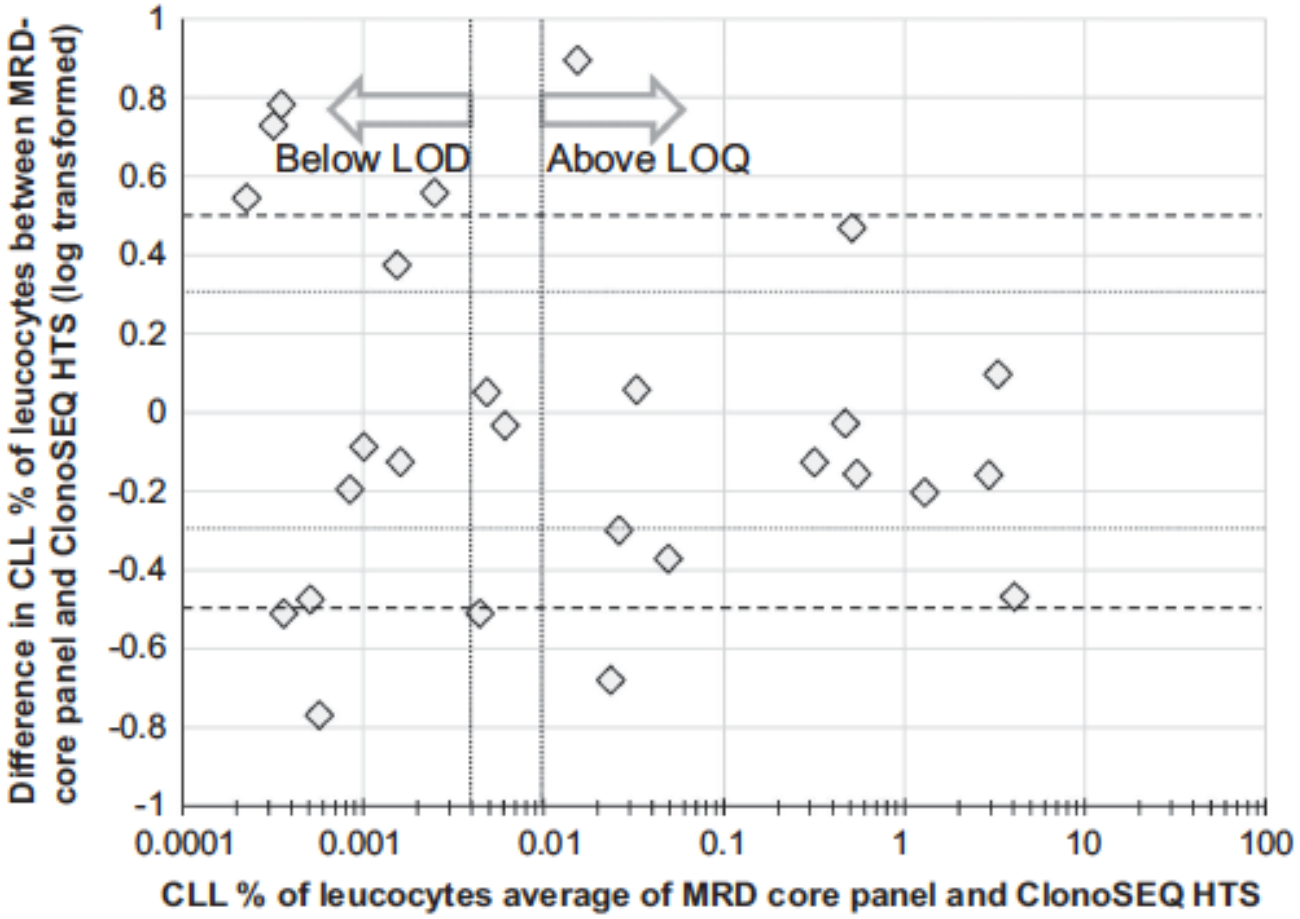
NGS

good linearity to 1 CLL cells in one million leukocytes



Rawstron et al. Leukemia 2016

NGS vs 6-color flow cytometry



NGS: MRD detected in 7/31 samples (22%) at the level of 0.0001% (10^{-6})

90% concordance at the level of 0.01% (10^{-4})

BioStatistics in immunogenetics

the bioinformatics approach

Biostatistics within the Bioinformatics approaches

High-throughput NGS immunogenetic data

- Established bioinformatics workflows (H-Vquest, ARResT/Interrogate/Vidjil, etc)
- Established knowledge databases (IMGT)

However, the final condensed information is still too big to safely assess

- There is a need for concise biostatistical and bioinformatics tools that will:
  - Evaluate significance
  - Visualize results in a meaningful way

Current approaches at CERTH

Represent the aminoacid information of CDR3 sequences in a formal way, easier to process at the mathematical level

Identify correlations between aminoacid clonotypes and the underlying nucleotide sequences the lead to the same aa seq

Construct networks of relevancy between clusters of CDR3 sequences and/or clonotypes

In-depth investigation of the consistency of groupings

Multiple CDR3 sequences: a numerical representation

Any statistical / computational approach heavily relies on numeric values

- Associate letters to numbers
- Define lexical operations (indels, mismatches) as the numerical metric

Multiple sequences can be transformed into single matrices

- Matrix length equal to the length of CDR3
- Matrix size depending on the level of granularity
  - Identity: size 20 (different aminoacids)
  - Similarity: size varies based on grouping (e.g. 11 IMGT groups)

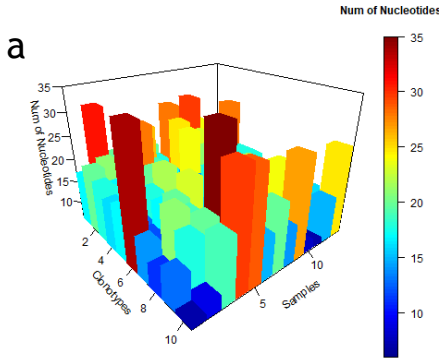
Numerical representations can be used directly in any numerical operation

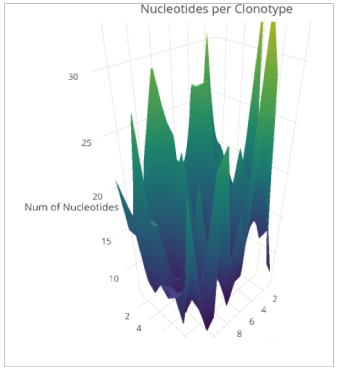
- Statistical correlation
- Significance

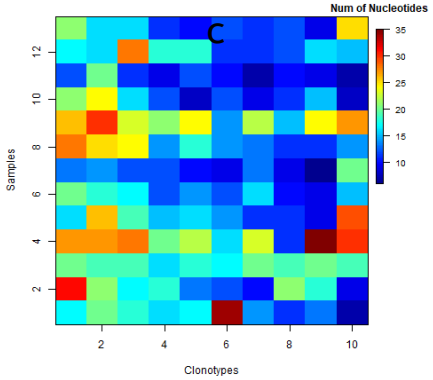
Top 10 clonotypes / nt seqs across multiple samples

Three different representations of the same information:

- 3D “building block” (x: Samples, y: Clonotypes, z: # of different nucleotide sequences of the same clonotype)
- 3D curve (x: Samples, y: Clonotypes, z: # of different nucleotide sequences of the same clonotype)
- 2D heatmap (x: Clonotypes, y: Samples, color: # of nt seqs)







Constructing the relationship network of CDR3 clusters

A numerical matrix-based representation of groups of CDR3 sequences allows for the use of additional metrics.

Distance between groups (i.e. CDR3 clusters) as a function of:

- Statistical metric of identity (i.e. mean, median, IQR3, etc)
- Statistical metric of similarity (i.e. mean, median, etc)
- Tree-based distance (such as average path length, etc)

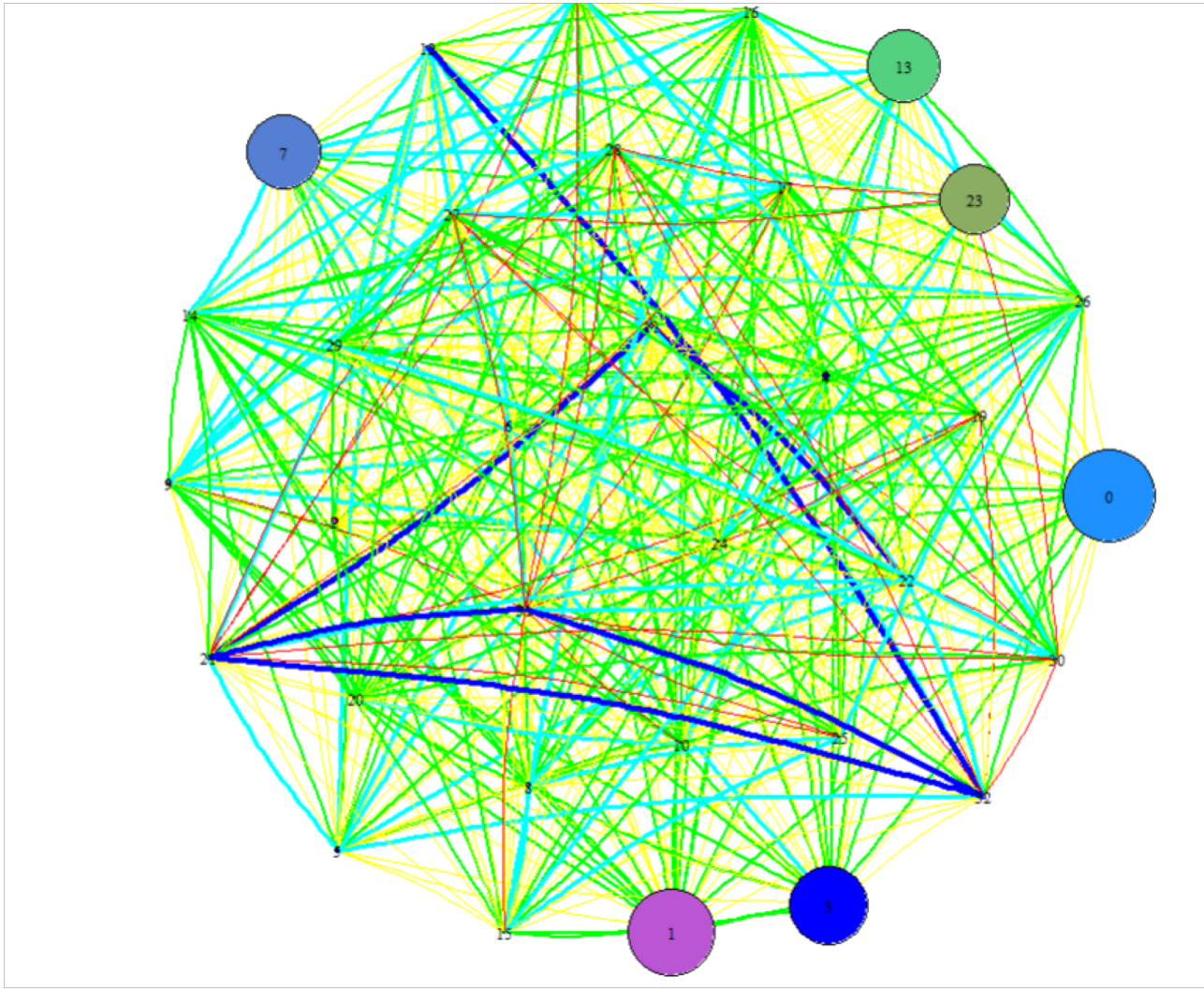
Combining multiple metrics as a single function, allows for both tweaking of the “weight” of each metric, as well as reproducibility of any given setup.

Network of closely similar CDR3 seqs

Using the metrics of:

- Identity
- Similarity
- Distance

Establish network of similar CDR3s at various levels of detail



INAB | CERTH, Thessaloniki
Anastasia Hadzidimitriou
Andreas Agathangelidis
Fotis Psomopoulos
Theodoros Moysiadis
Eva Minga
Katerina Gemenetzi
Chrysi Galigalidou
Elisavet Vlachonikola
Maria Kotouza
Laura Zaragoza

San Raffaele, Milan
Paolo Ghia

Pitié Salpêtrière, Paris
Fred Davi

Erasmus MC, Rotterdam
Anton Langerak

IMGT, Montpellier
Marie-Paule Lefranc
Veronique Giudicelli
Sofia Kossida

ERIC
european research initiative on CLL

NGS
EuroClonality

Im
Muno
Gene
Tics
Information system®