Bioinformatics platform for routine diagnostics of chronic lymphocytic leukemia patients

---

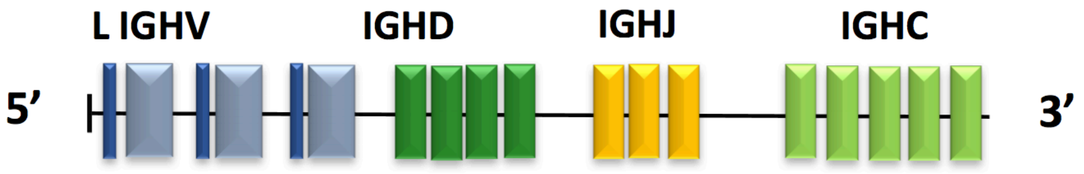
Tomáš Reigl, CEITEC MU Brno, Czech Republic

CEITEC
Central European Institute of Technology
BRNO | CZECH REPUBLIC

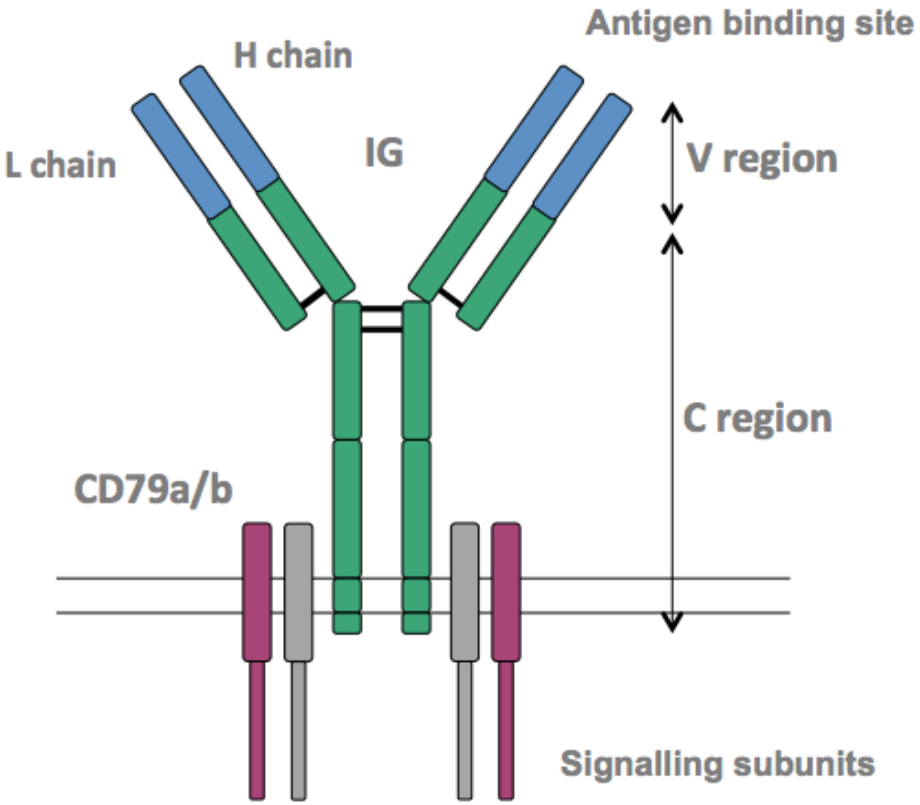
MASARYK
UNIVERSITY

Chronic lymphocytic leukemia (CLL)

- a heterogeneous disease with varying clinical outcome
- a clone of mature B cells resistant to apoptosis
- BcR immunoglobulin sequence analysis



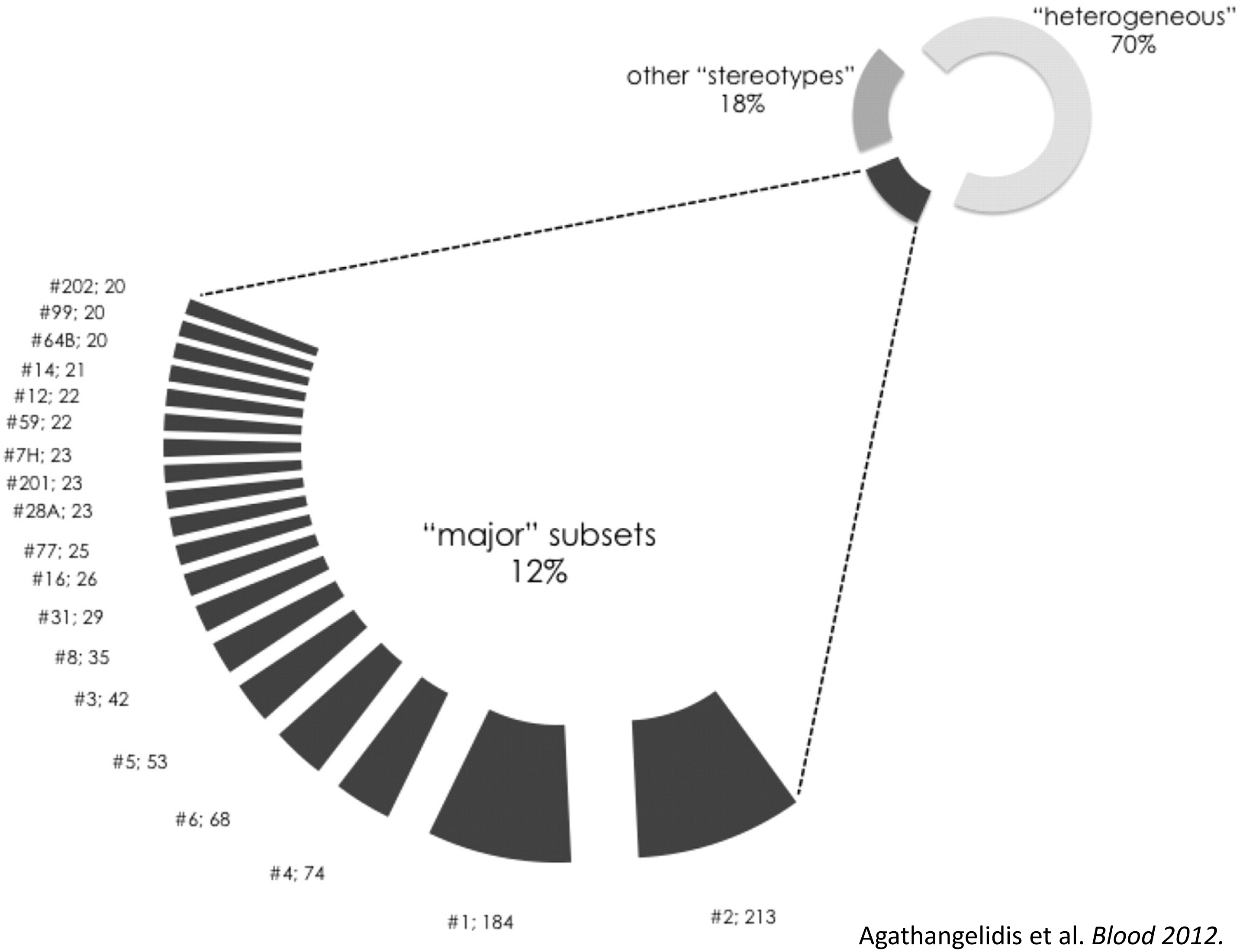
39-46 IGHV x 23 IGHD x 6 IGHJ



B cell receptor (BcR)

Chronic lymphocytic leukemia (CLL) - subsets

- one in three cases with stereotyped B cell receptors (BcR) in studied cohorts
- shared biological and clinical markers with implications for personalised care
- management and annotation of BcR sequence data



Agathangelidis et al. *Blood 2012.*

Antigen Receptors Research Tool / ARResT

- a bioinformatics platform for deep understanding of antigen receptor sequences
- a cascade of algorithms and databases
- Bioinformatics Analysis Team / BAT + many colleagues and collaborators
  - Central European Institute of Technology - Masaryk University Brno
  - Universty Hospital Brno
  - EuroClonality-NGS Consortium on Next-Generation Sequencing for IG / TR immunogenetic analysis

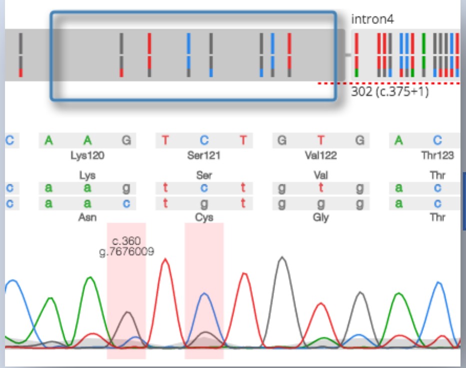
A prototype of the ARResT-based bioinformatics platform for the University Hospital Brno

IG sequences
(.fasta/.fastq)

alignment data
(.abi)

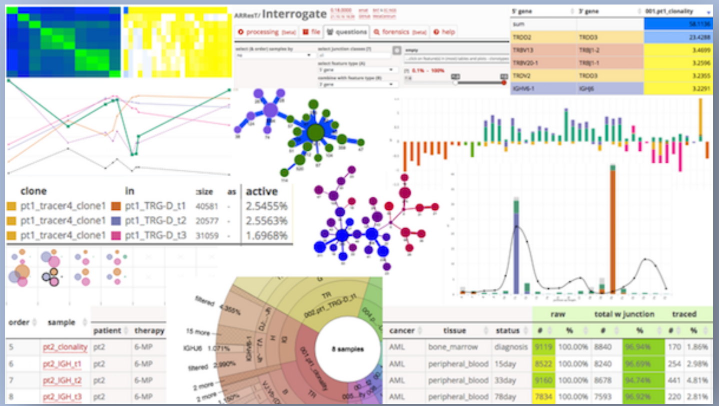
metadata
(.xls/.xlsx)

GLASS



chromatograms

ARResT/Interrogate



5' gene	3' gene	001.pt1_clonality
sum		58.1136
TRDD2	TRDD3	23.4288
TRBV13	TRBJ1-2	3.4699
TRBV20-1	TRBJ1-1	3.2596
TRDV2	TRDD3	3.2355
IGHV6-1	IGHJ6	3.2291

clone	in	size	as	active
pt1_tracer4_clone1	pt1_TRG-D_t1	40581	-	2.5455%
pt1_tracer4_clone1	pt1_TRG-D_t2	20577	-	2.5563%
pt1_tracer4_clone1	pt1_TRG-D_t3	31059	-	1.6968%

order	sample	patient	therapy
5	pt2_clonality	pt2	6-MP
6	pt2_IGH_t1	pt2	6-MP
7	pt2_IGH_t2	pt2	6-MP
8	pt2_IGH_t3	pt2	6-MP

			raw		total w junction		traced	
cancer	tissue	status	#	%	#	%	#	%
AML	bone_marrow	diagnosis	9119	100.00%	8840	96.94%	170	1.86%
AML	peripheral_blood	15day	8522	100.00%	8240	96.69%	254	2.98%
AML	peripheral_blood	33day	9160	100.00%	8678	94.74%	441	4.81%
AML	peripheral_blood	78day	7834	100.00%	7593	96.92%	220	2.81%

storage, visualisations, reports

IMGT/V-QUEST

ImMunoGeneTics Information system®

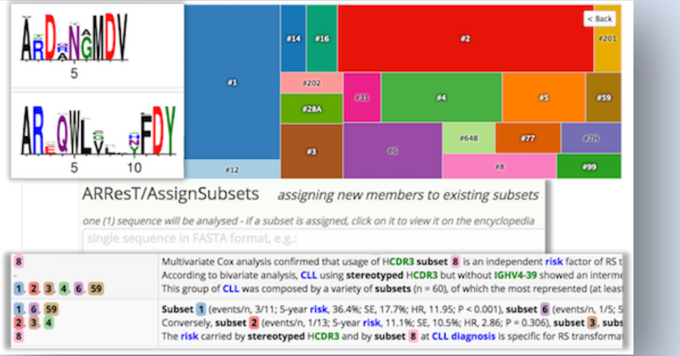
sequence annotation

ARResT/AssignSubsets

SeqCure	subset	confidence
OK	CLL#1	extreme
OK	CLL#2	high
warning	unassigned	extreme
fatal	skipped/unhealthy	

subset assignment

ARResT/Subsets



ARResT/AssignSubsets *assigning new members to existing subsets*

one (1) sequence will be analysed - if a subset is assigned, click on it to view it on the encyclopedia

single sequence in FASTA format, e.g.:

8 - 1, 2, 3, 4, 6, 59	Multivariate Cox analysis confirmed that usage of HCDR3 subset 8 is an independent risk factor of RS t According to bivariate analysis, CLL using stereotyped HCDR3 but without IGHV4-39 showed an interme This group of CLL was composed by a variety of subsets (n = 60), of which the most represented (at least
1, 6, 59 2, 3, 4 8	Subset 1 (events/n, 3/11; 5-year risk, 36.4%; SE, 17.7%; HR, 11.95; P < 0.001), subset 6 (events/n, 1/5; 5 Conversely, subset 2 (events/n, 1/13; 5-year risk, 11.1%; SE, 10.5%; HR, 2.86; P = 0.306), subset 3, subs The risk carried by stereotyped HCDR3 and by subset 8 at CLL diagnosis is specific for RS transformat

subsets context and information

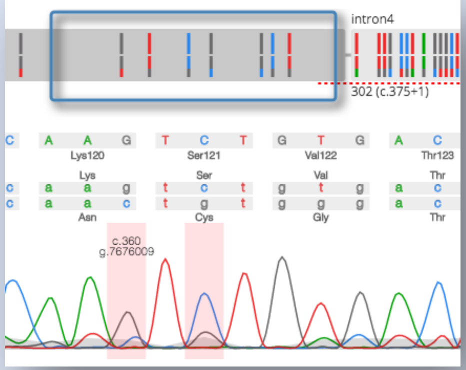
A prototype of the ARResT-based bioinformatics platform for the University Hospital Brno

IG sequences
(.fasta/.fastq)

alignment data
(.abi)

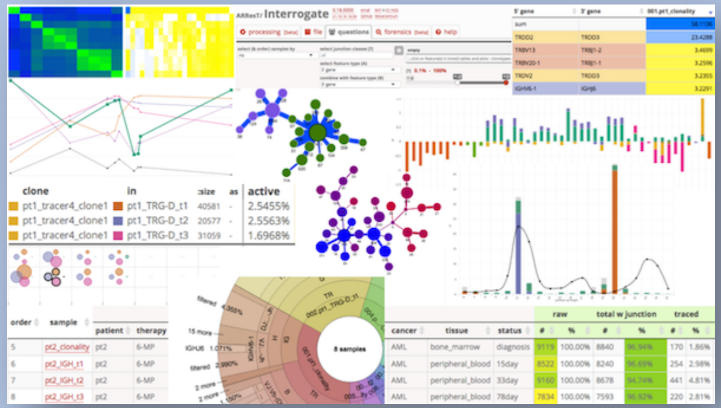
metadata
(.xls/.xlsx)

GLASS



chromatograms

ARResT/Interrogate



storage, visualisations, reports

IMGT/V-QUEST

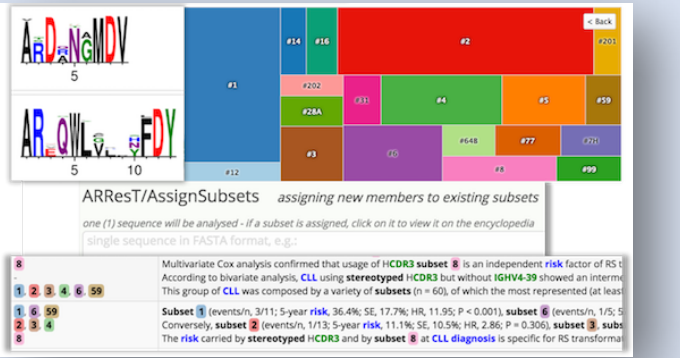
sequence annotation

ARResT/AssignSubsets

SeqCure	subset	confidence
OK	CLL#1	extreme
OK	CLL#2	high
warning	unassigned	extreme
fatal	skipped/unhealthy	

subset assignment

ARResT/Subsets



ARResT/AssignSubsets *assigning new members to existing subsets*

one (1) sequence will be analysed - if a subset is assigned, click on it to view it on the encyclopedia

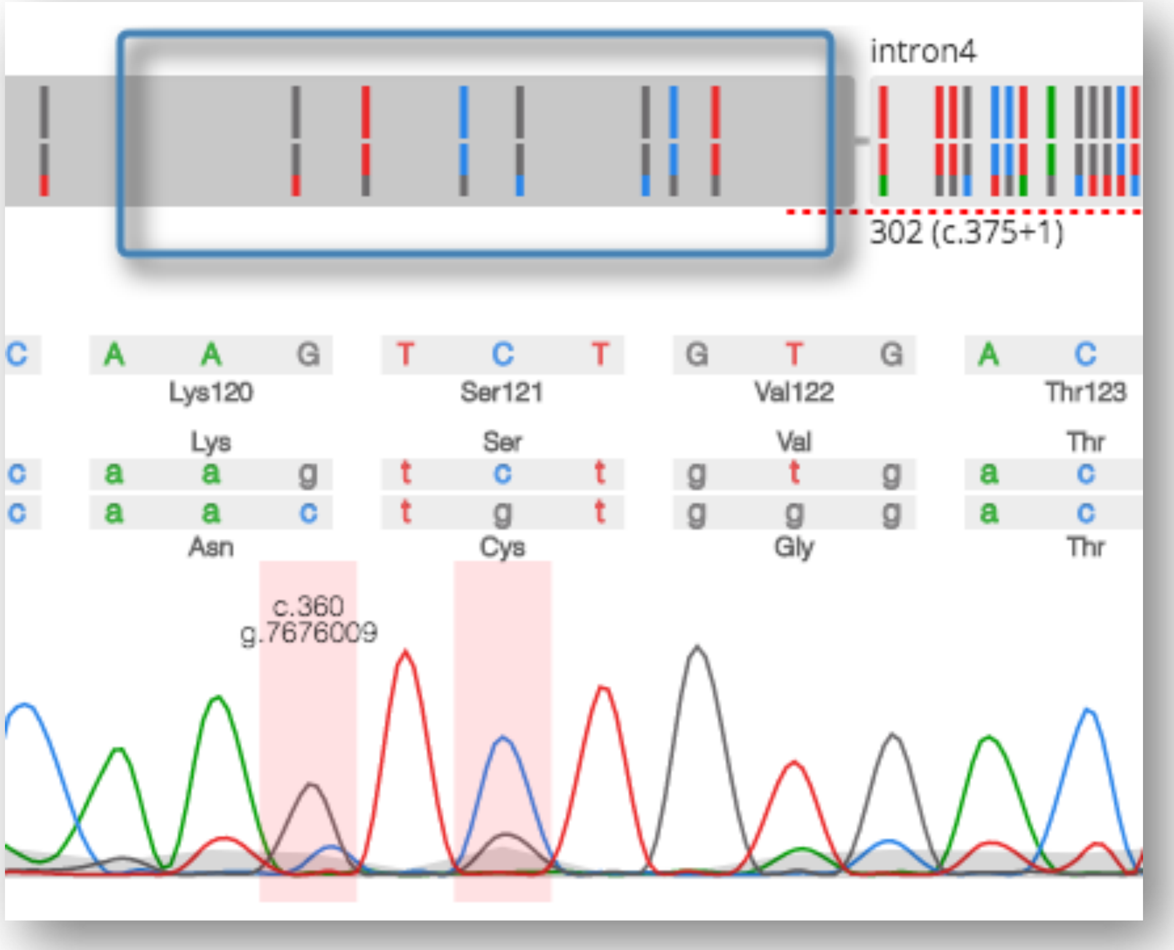
single sequence in FASTA format, e.g.:

8 - 1, 2, 3, 4, 6, 59	Multivariate Cox analysis confirmed that usage of HCDR3 subset 8 is an independent risk factor of RS t According to bivariate analysis, CLL using stereotyped HCDR3 but without IGHV4-39 showed an interm This group of CLL was composed by a variety of subsets (n = 60), of which the most represented (at least
1, 6, 59 2, 3, 4 8	Subset 1 (events/n, 3/11; 5-year risk, 36.4%; SE, 17.7%; HR, 11.95; P < 0.001), subset 6 (events/n, 1/5; 5 Conversely, subset 2 (events/n, 1/13; 5-year risk, 11.1%; SE, 10.5%; HR, 2.86; P = 0.306), subset 3, subs The risk carried by stereotyped HCDR3 and by subset 8 at CLL diagnosis is specific for RS transforma

subsets context and information

GLASS

- initial inspection of chromatograms
- sequence consensus from both forward and reverse strands
- automatic/manual error correction



http://apps.bat.infspire.org/genomepd/glass/

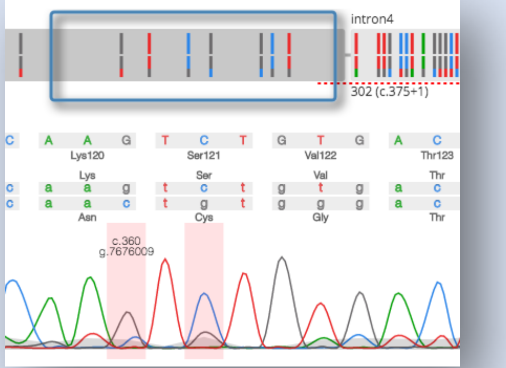
A prototype of the ARResT-based bioinformatics platform for the University Hospital Brno

IG sequences (.fasta/.fastq)

alignment data (.abi)

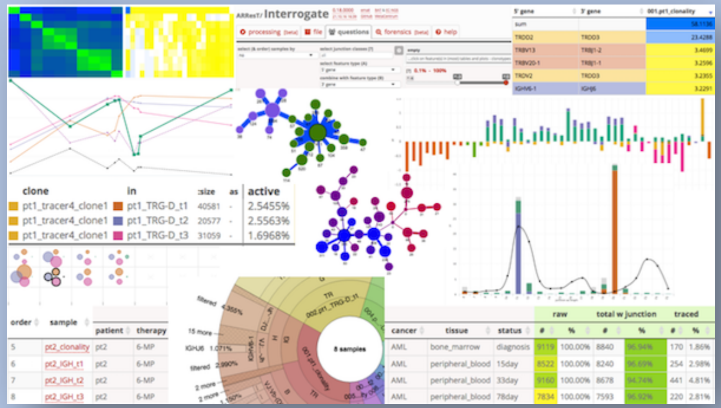
metadata (.xls/.xlsx)

GLASS



chromatograms

ARResT/Interrogate



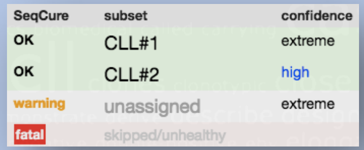
storage, visualisations, reports

IMGT/V-QUEST

IMGT ImMunoGeneTics Information system®

sequence annotation

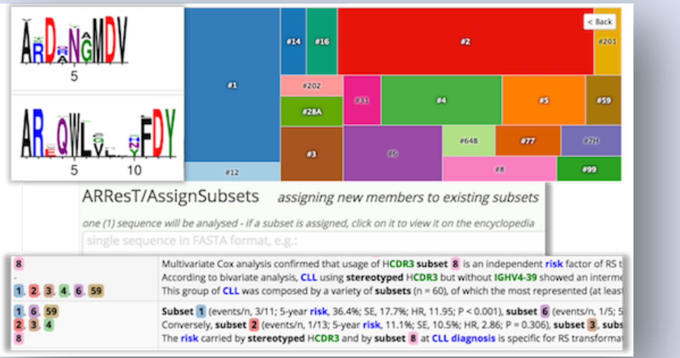
ARResT/AssignSubsets



SeqCure	subset	confidence
OK	CLL#1	extreme
OK	CLL#2	high
warning	unassigned	extreme
fatal	skipped/unhealthy	

subset assignment

ARResT/Subsets



ARResT/AssignSubsets *assigning new members to existing subsets*

one (1) sequence will be analysed - if a subset is assigned, click on it to view it on the encyclopedia

single sequence in FASTA format, e.g.:

8 - 1, 2, 3, 4, 6, 59	Multivariate Cox analysis confirmed that usage of HCDR3 subset 8 is an independent risk factor of RS t According to bivariate analysis, CLL using stereotyped HCDR3 but without IGHV4-39 showed an interme This group of CLL was composed by a variety of subsets (n = 60), of which the most represented (at least
1, 6, 59 2, 3, 4 8	Subset 1 (events/n, 3/11; 5-year risk, 36.4%; SE, 17.7%; HR, 11.95; P < 0.001), subset 6 (events/n, 1/5; 5 Conversely, subset 2 (events/n, 1/13; 5-year risk, 11.1%; SE, 10.5%; HR, 2.86; P = 0.306), subset 3, subs The risk carried by stereotyped HCDR3 and by subset 8 at CLL diagnosis is specific for RS transforma

subsets context and information

ARResT/Interrogate

- annotation of sequences with IMGT/V-QUEST
- assignment of sequences to major CLL subsets with ARResT/AssignSubsets
- access to all stored and new data in more detail
- generating informative and consistent reports

	clone		in	size	as	active
🟧	pt1_tracer4_clone1	🟧	pt1_TRG-D_t1	40581	-	2.5455%
🟧	pt1_tracer4_clone1	🟦	pt1_TRG-D_t2	20577	-	2.5563%
🟧	pt1_tracer4_clone1	🟥	pt1_TRG-D_t3	31059	-	1.6968%

order	sample	patient	therapy
5	pt2_clonality	pt2	6-MP
6	pt2_IGH_t1	pt2	6-MP
7	pt2_IGH_t2	pt2	6-MP
8	pt2_IGH_t3	pt2	6-MP

Odes.:FN Brno
IHOK, Centrum molekulární biologie a genové terapie
Černopolní 9
62500 Brno - Černá Pole

IHOK FN Brno
Oddělení: IHOK-DTC-hematol.amb.3569
Jihlavská 20
62500 Brno

Pro jednotku: IHOK-DTC-hematol.amb.3562

Pacient:
Vyžádal: 72-100-377 MUDr. Brychtová
Diagnóza:
Komentář:
Vyšetření:

Odebrán: 2017-11-07
Přijat: 2017-11-08
Odeslán: 2017-11-08
Číslo: 11549/LYM*1
Kdo hradí: VZP

Homologie IgVH 95
V-D-J přestavba .... klonální expanze

Komentář k homologii:
ZÁVĚR: Pacient/ka má mutovaný IgVH gen.
Procenta vyjadřují nukleotidovou homologii V segmentu k nejbližší zárodečné linii.

1. var. část genu těžkého řetězce Ig**V**H:
2. rozlišovací část genu Ig**D**H:
3. spojovací část genu Ig**J**H:
Délka HCDR3: 20 aminokyselin.

SOP č.SA/CMBGT/L0003

Uvolnil: Mgr. Andrea Marečková
Zhotovil: Mgr. Hana Skuhrová Francová

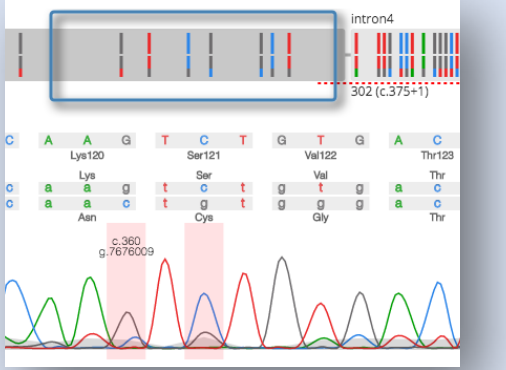
A prototype of the ARResT-based bioinformatics platform for the University Hospital Brno

IG sequences (.fasta/.fastq)

alignment data (.abi)

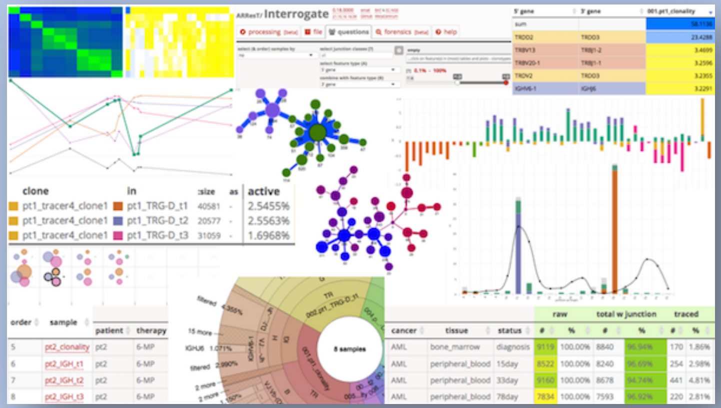
metadata (.xls/.xlsx)

GLASS



chromatograms

ARResT/Interrogate

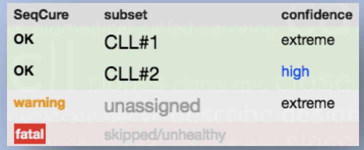


storage, visualisations, reports

IMGT/V-QUEST

sequence annotation

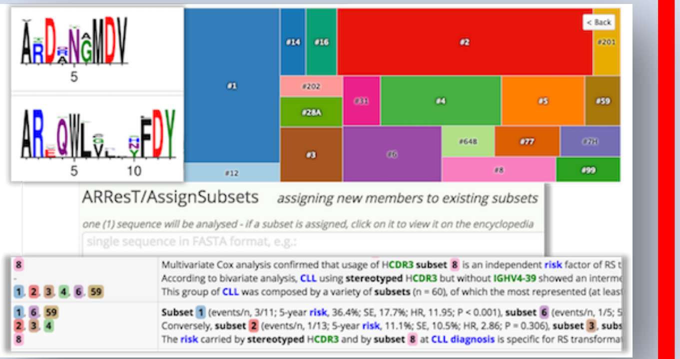
ARResT/AssignSubsets



SeqCure	subset	confidence
OK	CLL#1	extreme
OK	CLL#2	high
warning	unassigned	extreme
fatal	skipped/unhealthy	

subset assignment

ARResT/Subsets



ARResT/AssignSubsets *assigning new members to existing subsets*

one (1) sequence will be analysed - if a subset is assigned, click on it to view it on the encyclopedia

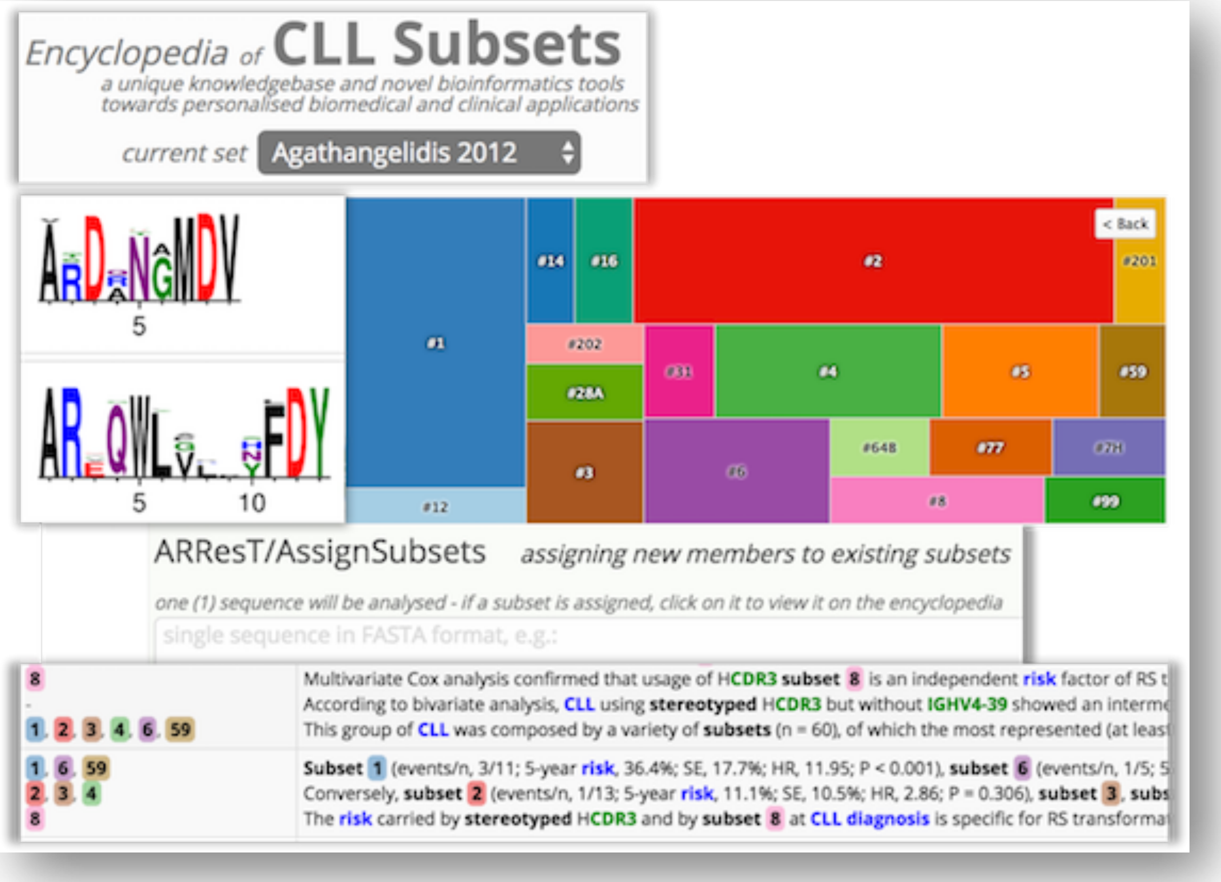
single sequence in FASTA format, e.g.:

8 - 1, 2, 3, 4, 6, 59	Multivariate Cox analysis confirmed that usage of HCDR3 subset 8 is an independent risk factor of RS t According to bivariate analysis, CLL using stereotyped HCDR3 but without IGHV4-39 showed an interme This group of CLL was composed by a variety of subsets (n = 60), of which the most represented (at least
1, 6, 59 2, 3, 4 8	Subset 1 (events/n, 3/11; 5-year risk, 36.4%; SE, 17.7%; HR, 11.95; P < 0.001), subset 6 (events/n, 1/5; 5 Conversely, subset 2 (events/n, 1/13; 5-year risk, 11.1%; SE, 10.5%; HR, 2.86; P = 0.306), subset 3, subs The risk carried by stereotyped HCDR3 and by subset 8 at CLL diagnosis is specific for RS transformat

subsets context and information

ARResT/Subsets

- results in context of other stereotyped CLL cases
- access to publicly available relevant literature, including published biological and clinical data



Encyclopedia of CLL Subsets

a unique knowledgebase and novel bioinformatics tools towards personalised biomedical and clinical applications

current set Agathangelidis 2012

ARResT/AssignSubsets *assigning new members to existing subsets*

one (1) sequence will be analysed - if a subset is assigned, click on it to view it on the encyclopedia

single sequence in FASTA format, e.g.:

8 - 1, 2, 3, 4, 6, 59	Multivariate Cox analysis confirmed that usage of HCDR3 **subset** 8 is an independent risk factor of RS t According to bivariate analysis, CLL using **stereotyped** HCDR3 but without IGHV4-39 showed an interme This group of CLL was composed by a variety of **subsets** (n = 60), of which the most represented (at leas
1, 6, 59 2, 3, 4 8	**Subset** 1 (events/n, 3/11; 5-year risk, 36.4%; SE, 17.7%; HR, 11.95; P < 0.001), **subset** 6 (events/n, 1/5; 5 Conversely, **subset** 2 (events/n, 1/13; 5-year risk, 11.1%; SE, 10.5%; HR, 2.86; P = 0.306), **subset** 3, **subs** The risk carried by **stereotyped** HCDR3 and by **subset** 8 at CLL diagnosis is specific for RS transforma

http://apps.bat.infspire.org/arrest/subsets/

Summary

- a prototype of the ARResT-based bioinformatics platform for a cohort of >2000 of CLL cases in the University Hospital Brno for daily use in the clinic

- annotation, storage, visualisations and comparison of CLL patient data

- direct access to publicly available relevant literature (>250) about any of the major CLL subsets including published biological and clinical data

- personalised biomedical and clinical applications

Thank you for your attention.

bat.infspire.org

tomas.reigl@gmail.com

Tomas Reigl[1], Kamila Stranska[1,2,3], Karol Pal[1,3], Vojtech Bystry[1], Adam Krejci[1,4], Sarka Pospisilova[1,2], Nikos Darzentas[1,5], Karla Plevova[1,2,3]

[1] *CEITEC, Masaryk University Brno, Czechia*
[2] *Department of Internal Medicine – Hematology and Oncology, University Hospital Brno*
[3] *Medical Faculty, Masaryk University, Czechia*
[4] *RECAMO, Masaryk Memorial Cancer Institute, Brno, Czechia*
[5] *University Hospital Schleswig-Holstein, Kiel, Germany*

Acknowledgements: Supported by MZCR-AZV 16-34272A, MZCR-RVO 65269705, and MUNI/A/0968/2017.