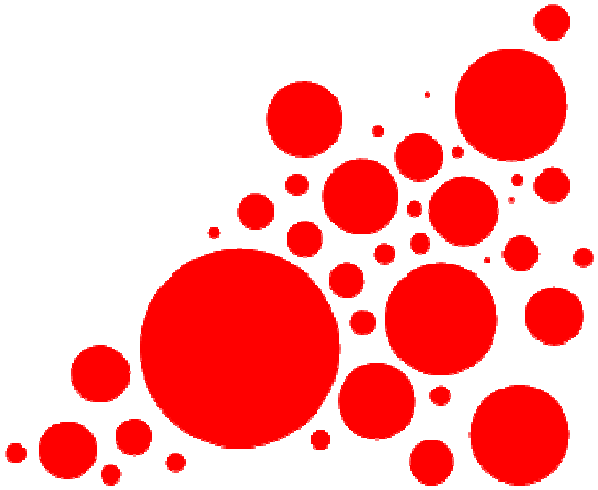
How AI can support diagnostic approaches in haematology

Torsten Haferlach
MLL Munich Leukemia Laboratory

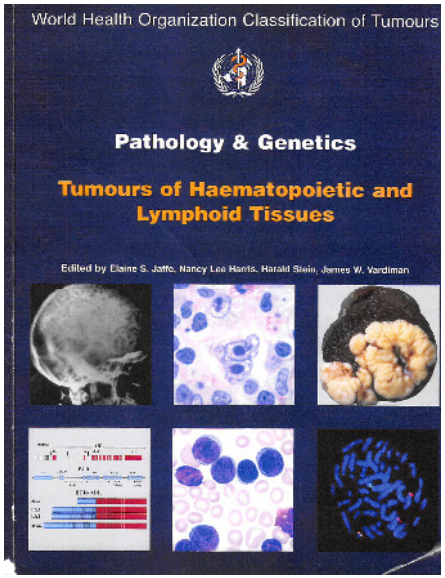


Disclosures

- Dr. Haferlach is part owner of MLL Munich Leukemia Laboratory

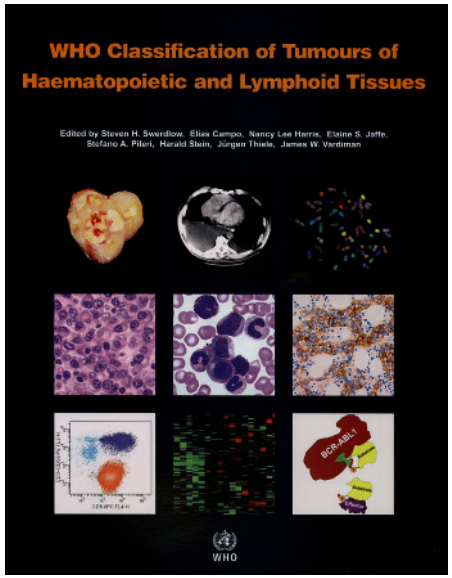
WHO Classification 2001 to 2017

Genetically defined entities: **5**
AML (n=4), MDS (n=1)



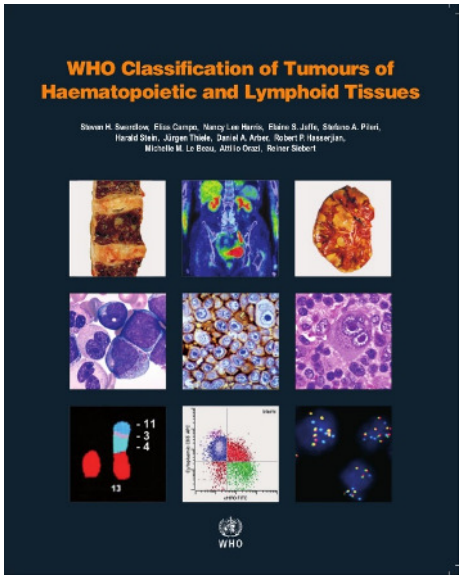
2001

Genetically defined entities: **24**
AML (n=10), MDS (n=2),
MPN (n=3), ALL (n=9)



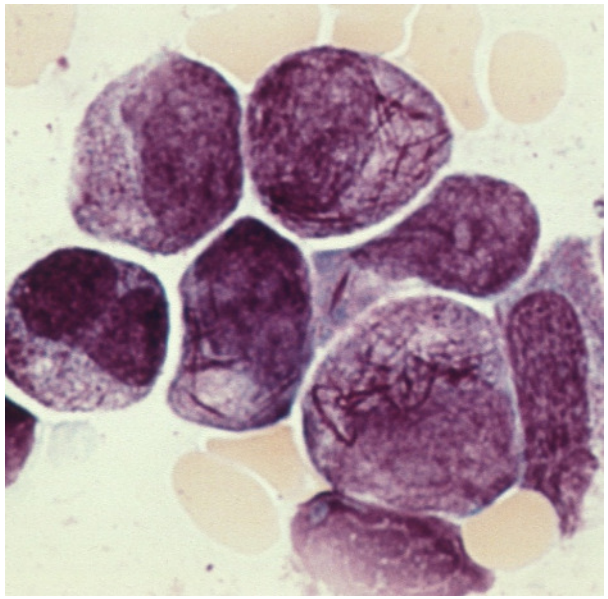
2008

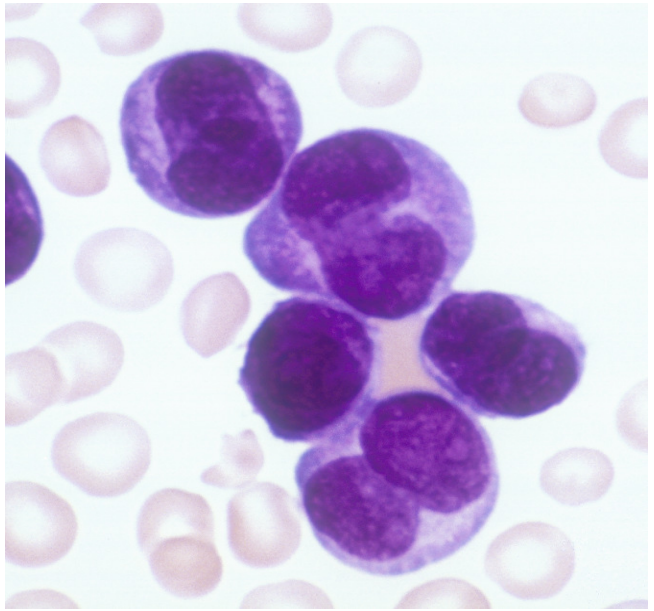
Genetically defined entities: **31**
AML (n=12), MDS (n=2),
MPN (n=6), ALL (n=11)



2017

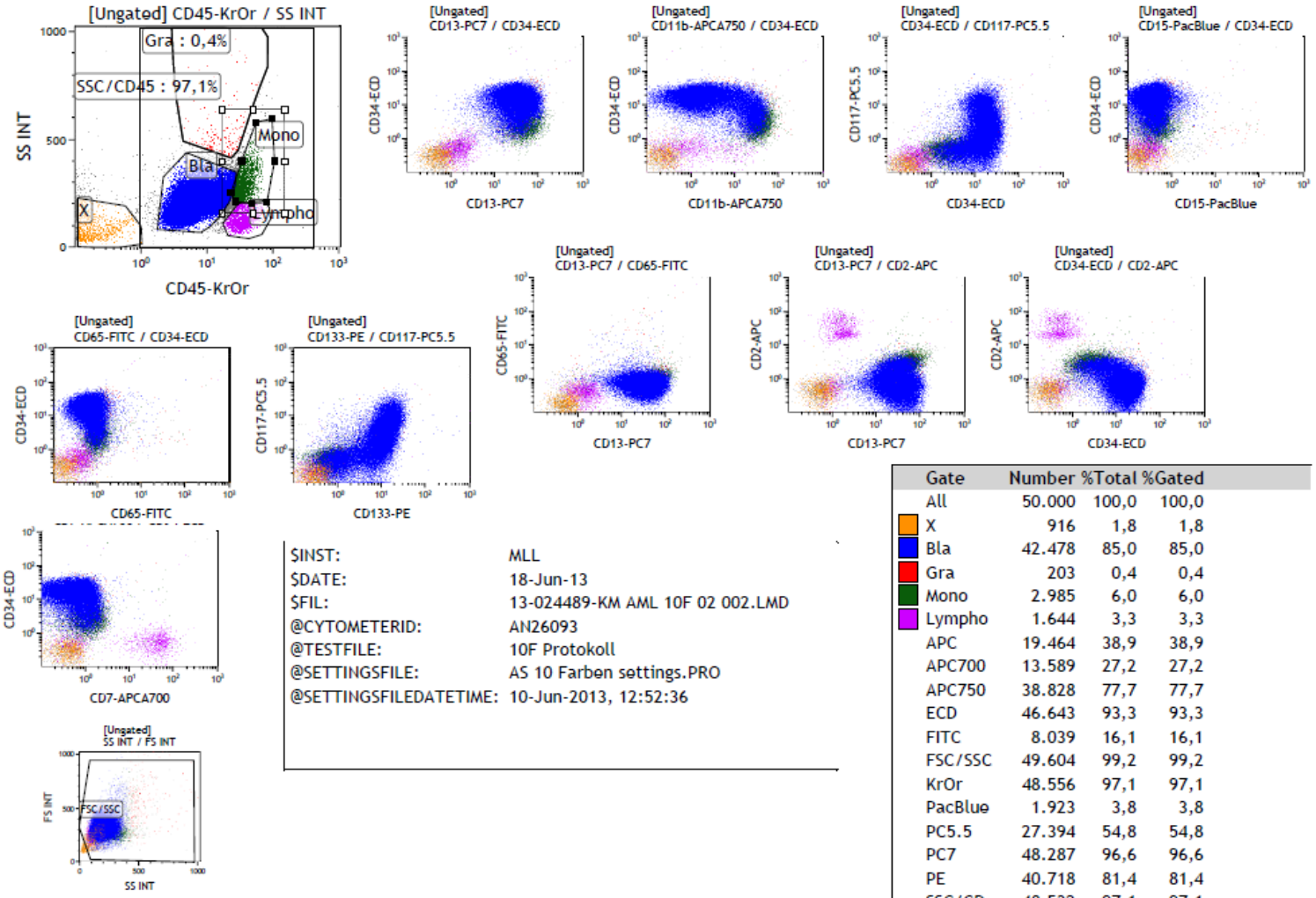
Cytomorphology





MGG

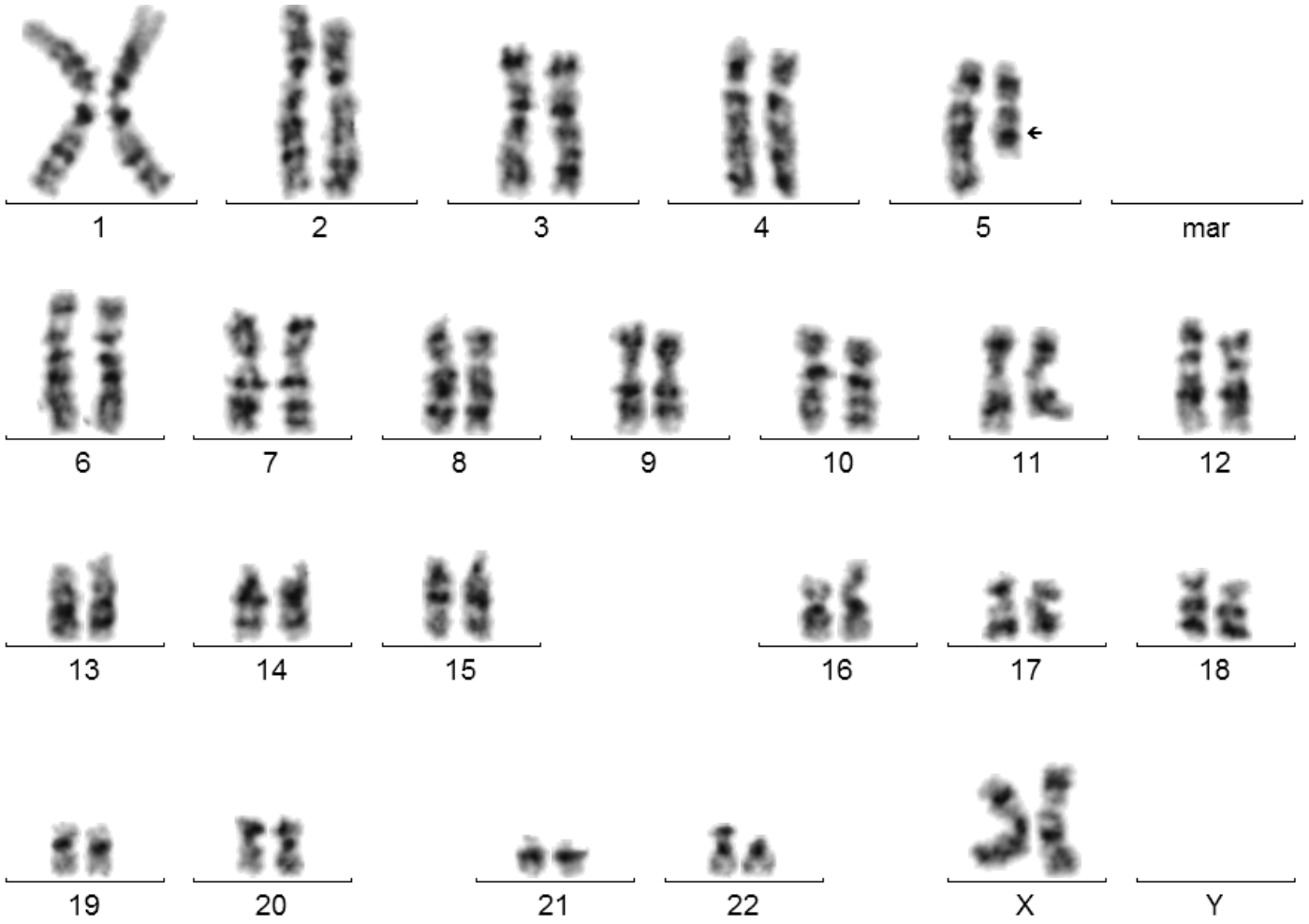
Immunophenotyping: AML (10-color-staining)



$INST:	MLL
$DATE:	18-Jun-13
$FIL:	13-024489-KM AML 10F 02 002.LMD
@CYTOMETERID:	AN26093
@TESTFILE:	10F Protokoll
@SETTINGSFILE:	AS 10 Farben settings.PRO
@SETTINGSFILEDATETIME:	10-Jun-2013, 12:52:36

Gate	Number	%Total	%Gated
All	50.000	100,0	100,0
X	916	1,8	1,8
Bla	42.478	85,0	85,0
Gra	203	0,4	0,4
Mono	2.985	6,0	6,0
Lympho	1.644	3,3	3,3
APC	19.464	38,9	38,9
APC700	13.589	27,2	27,2
APC750	38.828	77,7	77,7
ECD	46.643	93,3	93,3
FITC	8.039	16,1	16,1
FSC/SSC	49.604	99,2	99,2
KrOr	48.556	97,1	97,1
PacBlue	1.923	3,8	3,8
PC5.5	27.394	54,8	54,8
PC7	48.287	96,6	96,6
PE	40.718	81,4	81,4
SSC/CD...	48.532	97,1	97,1

Karyotype: 46,XX,del(5)(q15q32)



MLL Myeloid panel

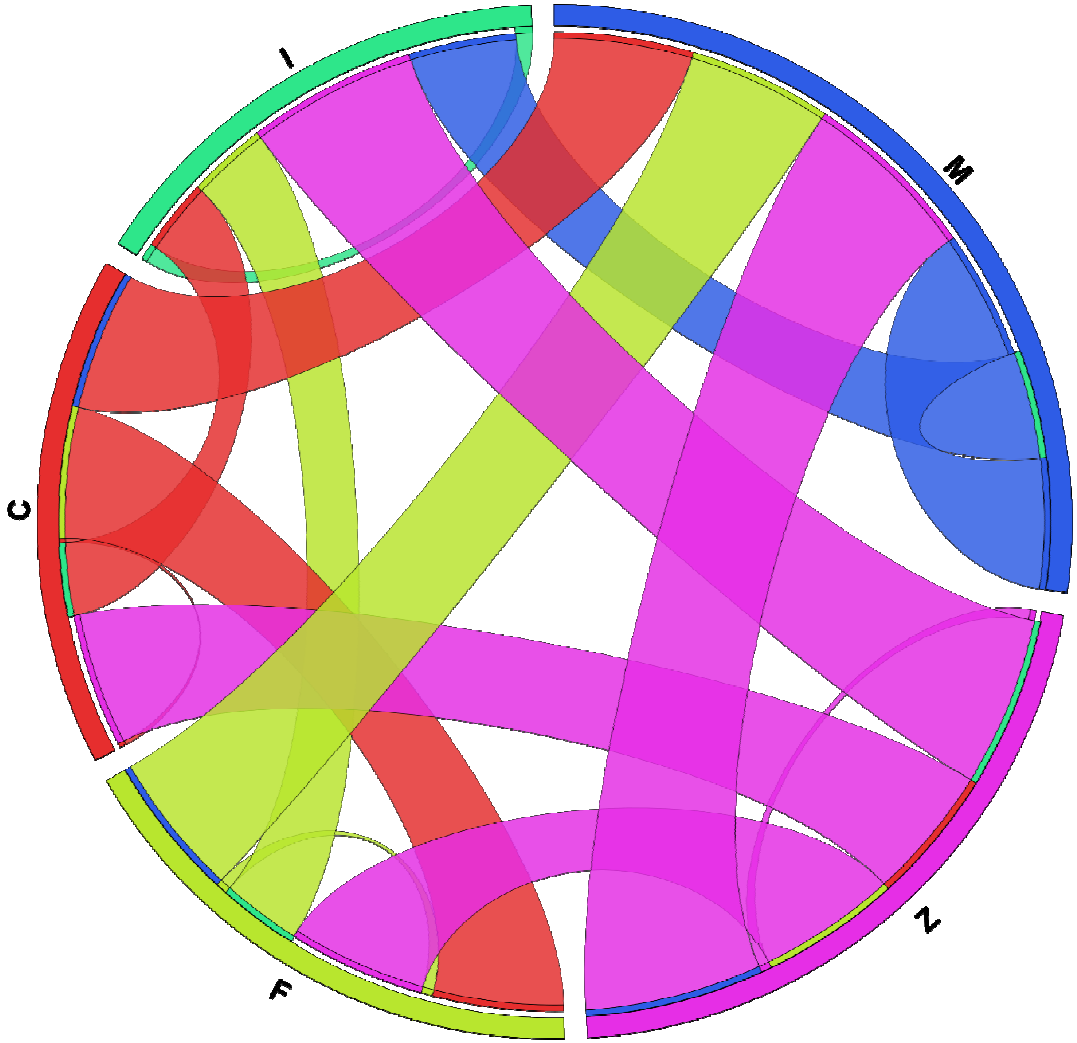
- *APC*
- *ASXL1*
- *ASXL2*
- *ATM*
- *ATRX*
- *BCOR*
- *BCORL1*
- *BRAF*
- *BRCC3*
- *CALR*
- *CBL*
- *CDH23*
- *CDKN2A*
- *CEBPA*
- *CREBBP*
- *CSF3R*
- *CSNK1A1*
- *CTCF*
- *CUX1*
- *DDX41*
- *DDX54*
- *DHX29*
- *DNMT3A*
- *EP300*
- *ETNK1*
- *ETV6*
- *EZH2*
- *FANCL*
- *FBXW7*
- *FLT3*
- *GATA1*
- *GATA2*
- *GNAS*
- *GNB1*
- *IDH1*
- *IDH2*
- *JAK2*
- *KDM5A*
- *KDM6A*
- *KIT*
- *KMT2D*
- *KRAS*
- *MPL*
- *MYC*
- *NF1*
- *NOTCH1*
- *NPM1*
- *NRAS*
- *PHF6*
- *PIGA*
- *PPM1D*
- *PRPF8*
- *PTPN11*
- *RAD21*
- *RB1*
- *RUNX1*
- *SETBP1*
- *SF1*
- *SF3A1*
- *SF3B1*
- *SH2B3*
- *SMC1A*
- *SMC3*
- *SRSF2*
- *STAG2*
- *SUZ12*
- *TET2*
- *TP53*
- *U2AF1*
- *U2AF2*
- *WT1*
- *ZBTB7A*
- *ZRSR2*

73 genes
+ 23 loci for Pat ID

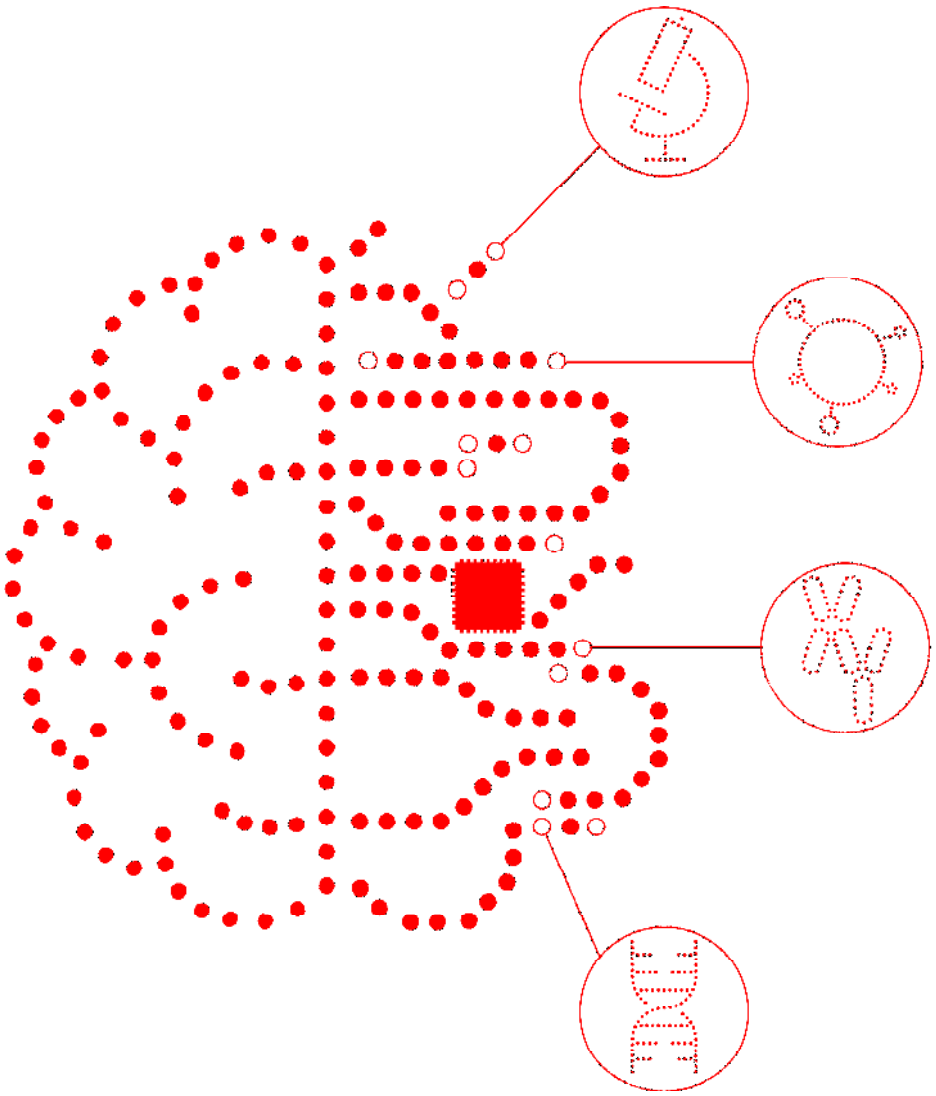
MLL data

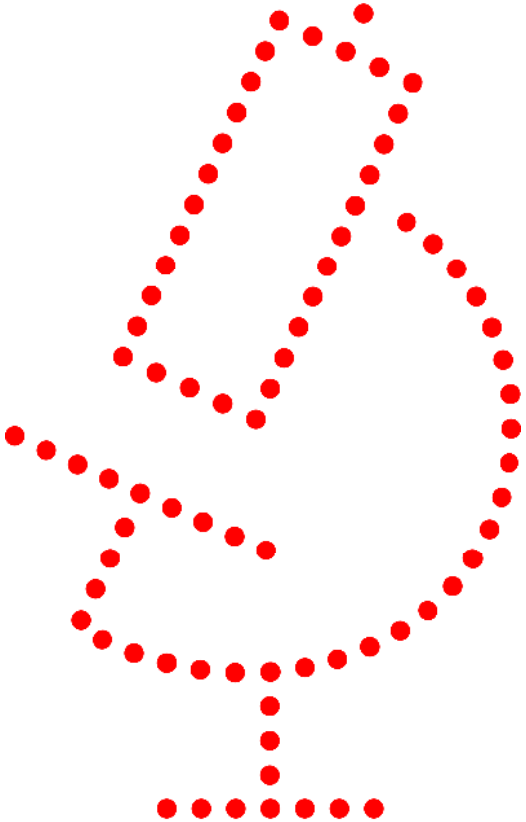
Composition of goldstandards in hematology

Z = Cytomorphology
C = Cytogenetics
F = FISH
M = Moleculargenetics
I = Immunophenotyping



MLL data





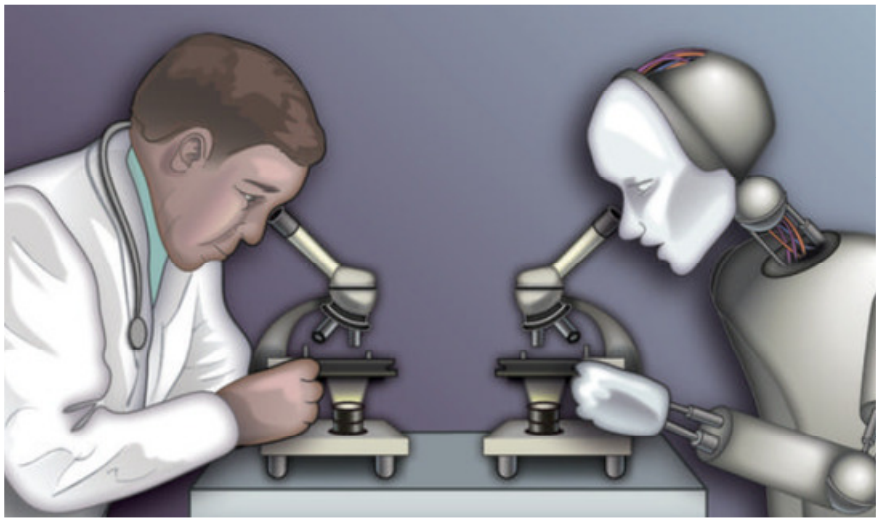
AI-based peripheral blood cell differentiation

10,082 patient samples (Jan 2021 – Jul 2021)

∑ = 988,130 cells differentiated

Highly skilled technicians (median 5y in lab)

Model for 21 predefined classes

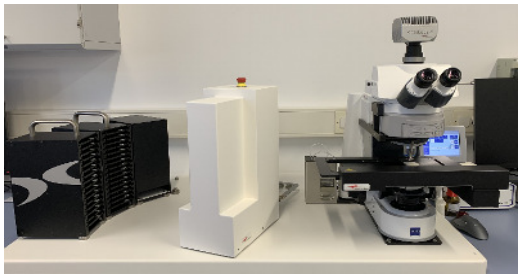


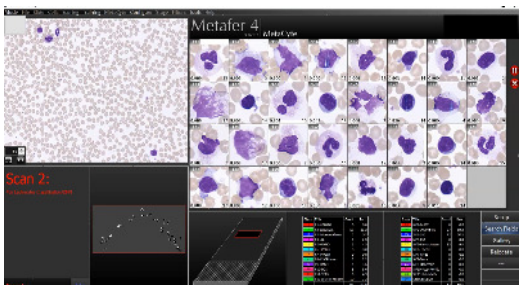
52%	Segmented neutrophils	53%
2.25%	[illegible]	3.36%
0.72%		0.72%
7.5%		6.64%
31.7%	Lymphocytes	24%
0.97 %	Pathogenic blasts	1.65%

Concordance of 95% for pathogenic cases

∑ = 4,937,389 cells differentiated

Capture time for 500 cells: 4:37 min



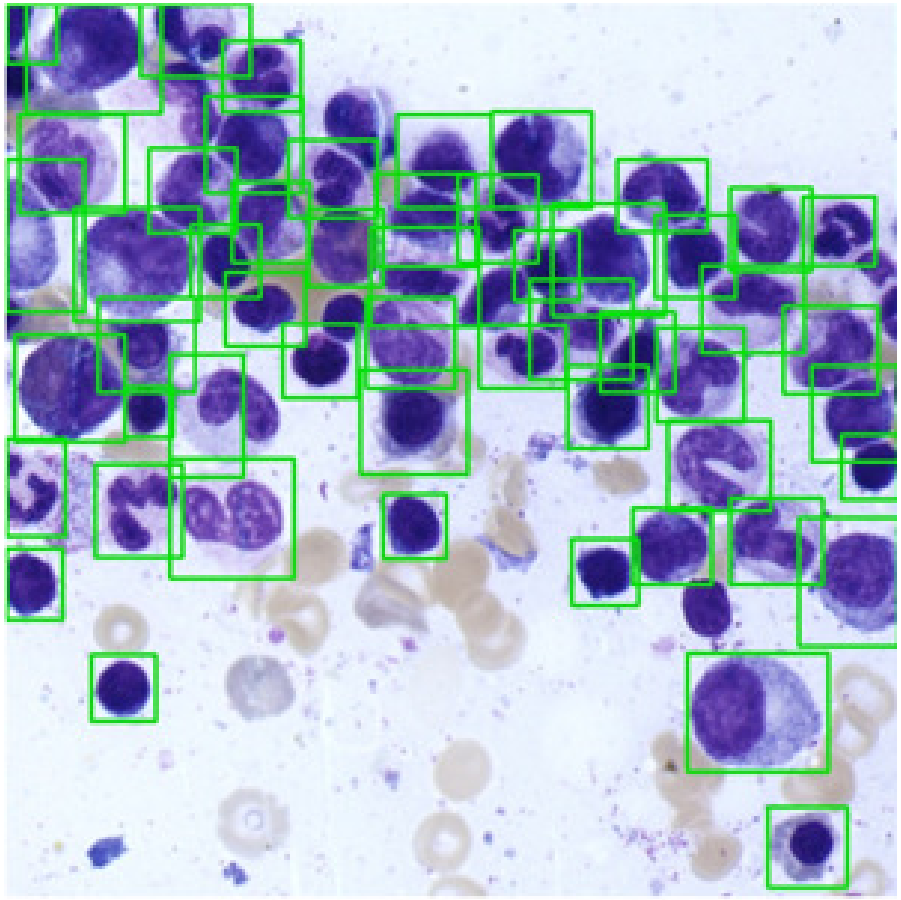


MLL data

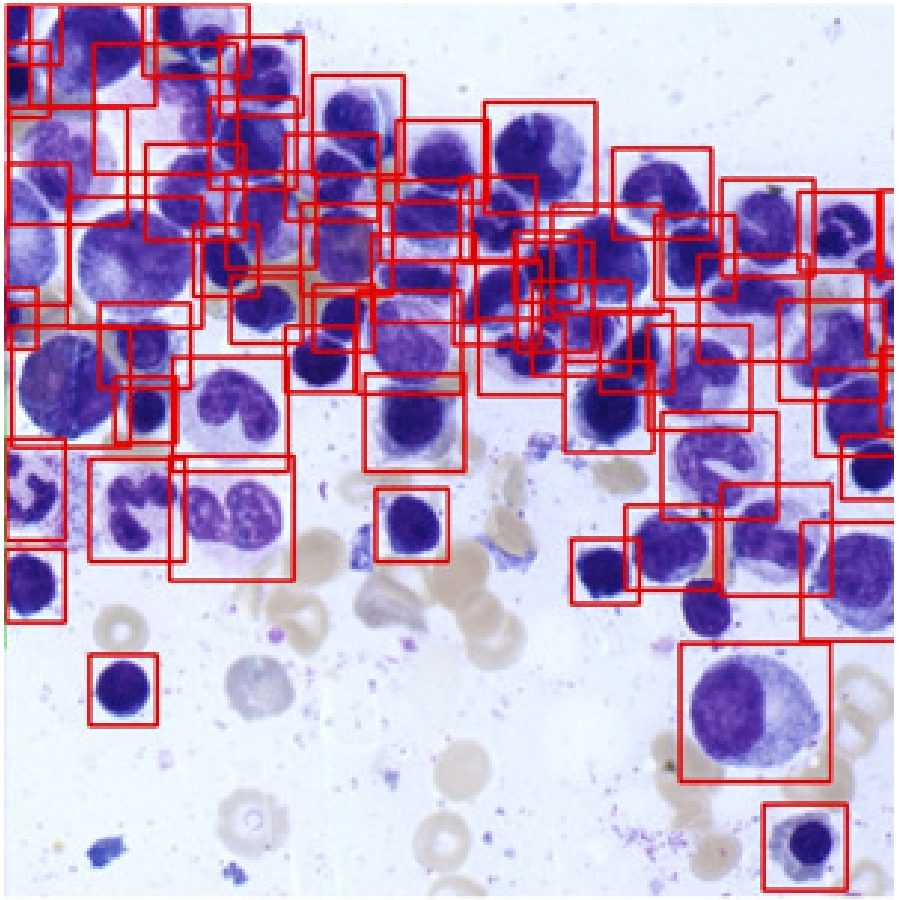
NIH U.S. National Library of Medicine
ClinicalTrials.gov

NCT04466059

AI-based bone marrow object detection

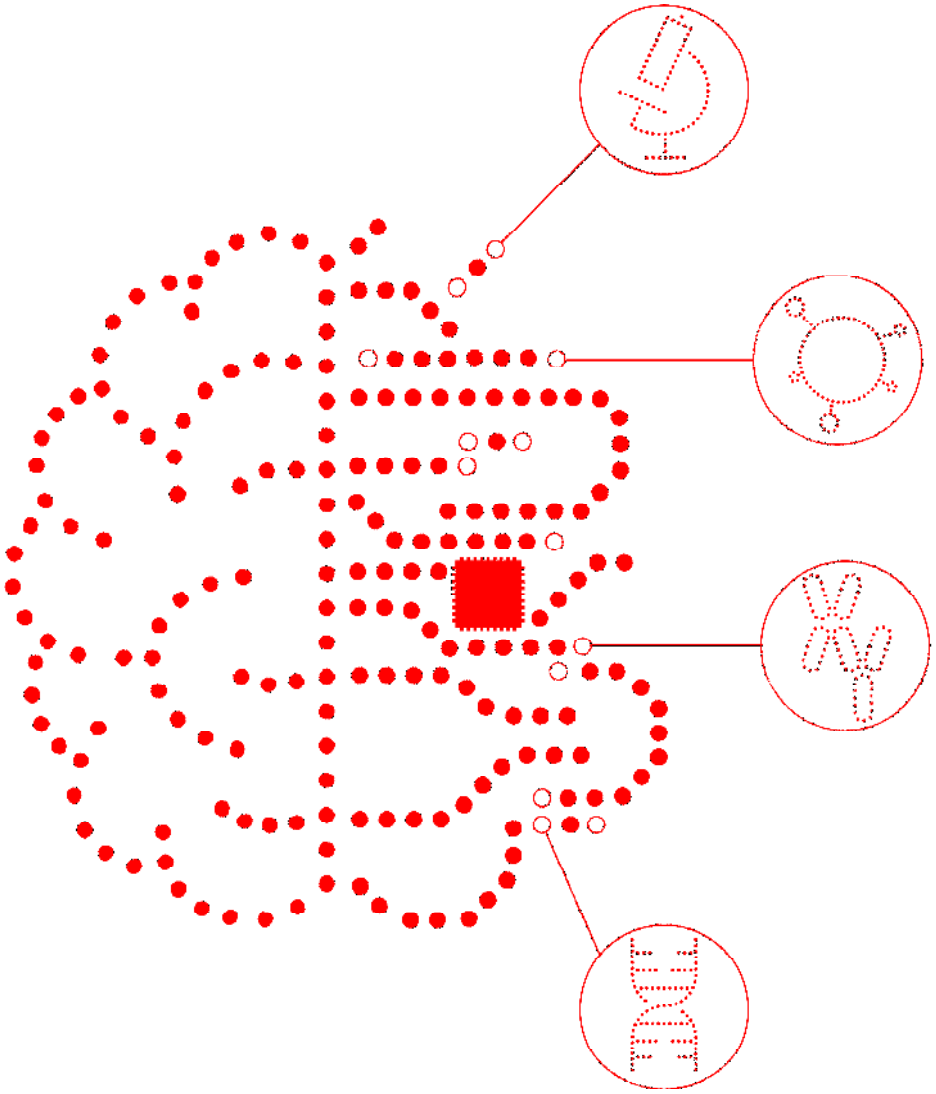


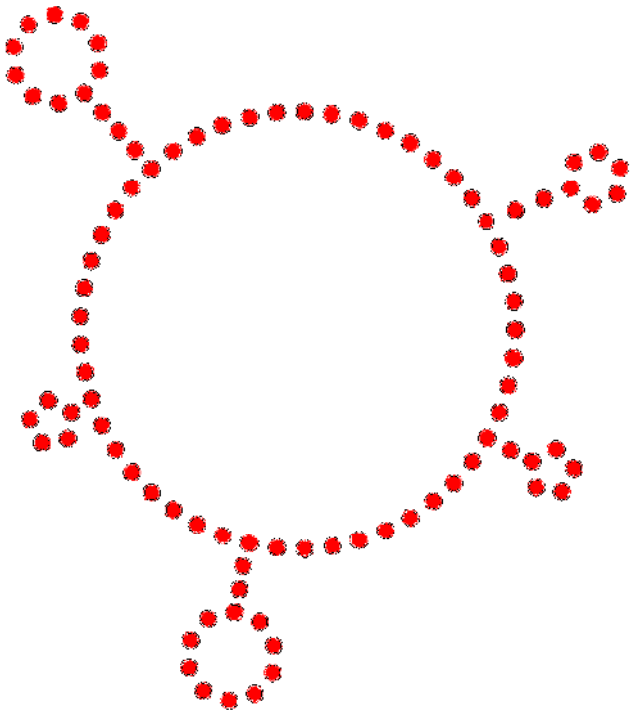
Manual object definition



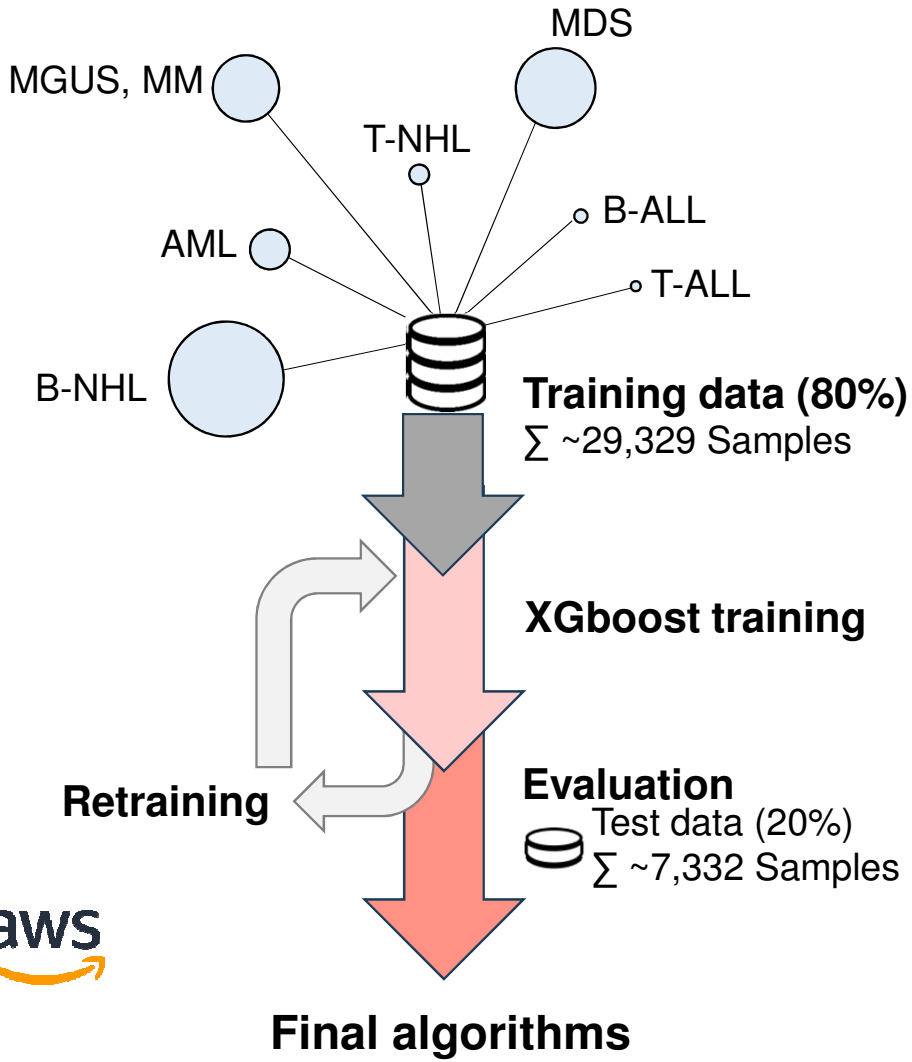
AI based object detection

MLL data

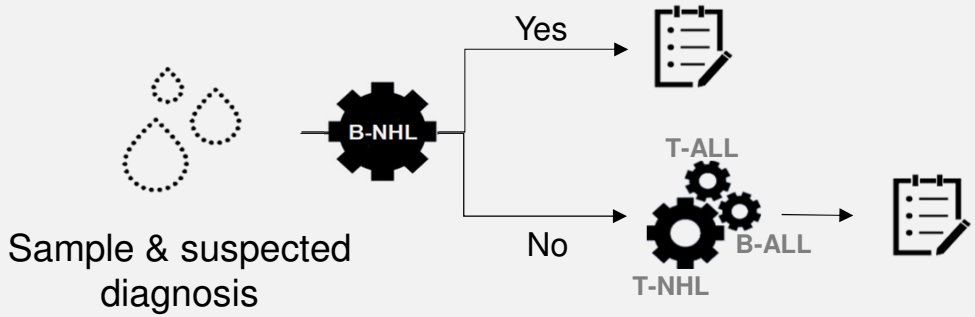




AI-based diagnostic using flow cytometric data



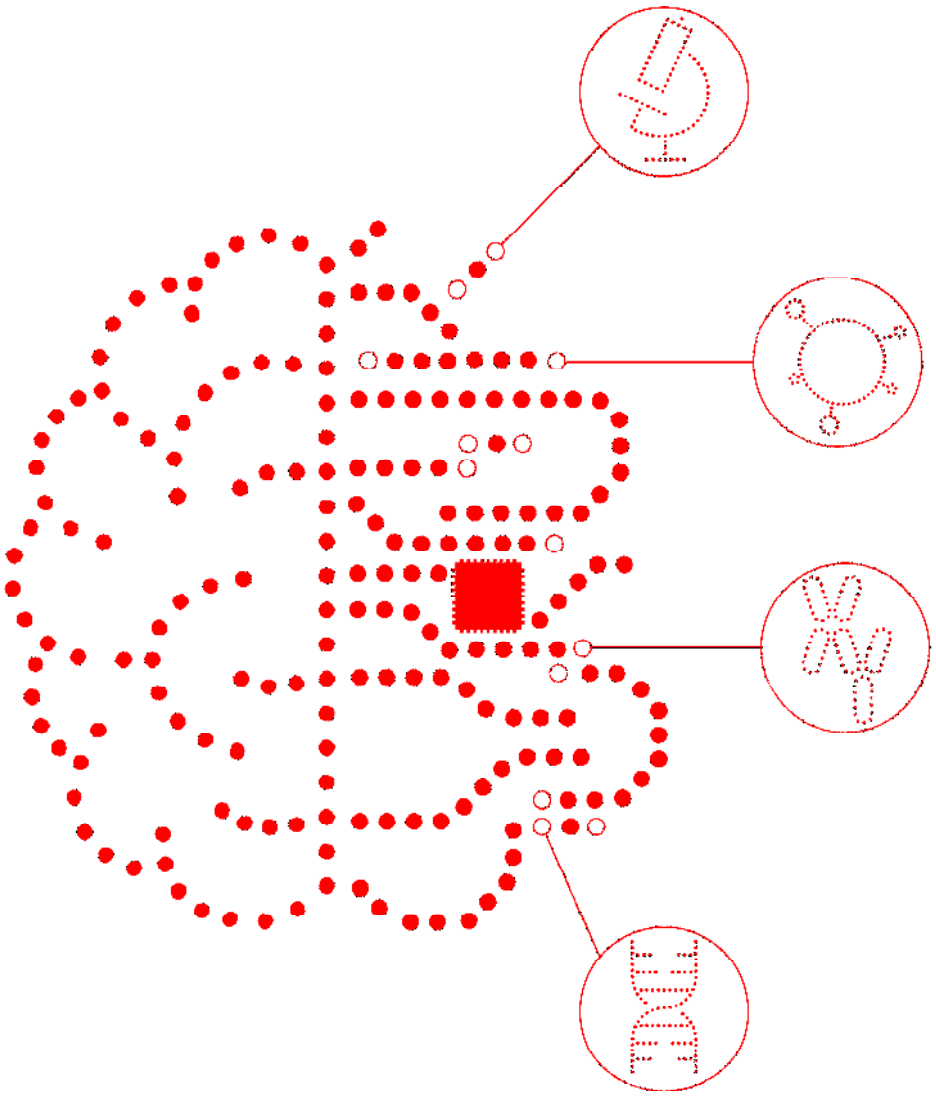
Intended diagnostic workflow



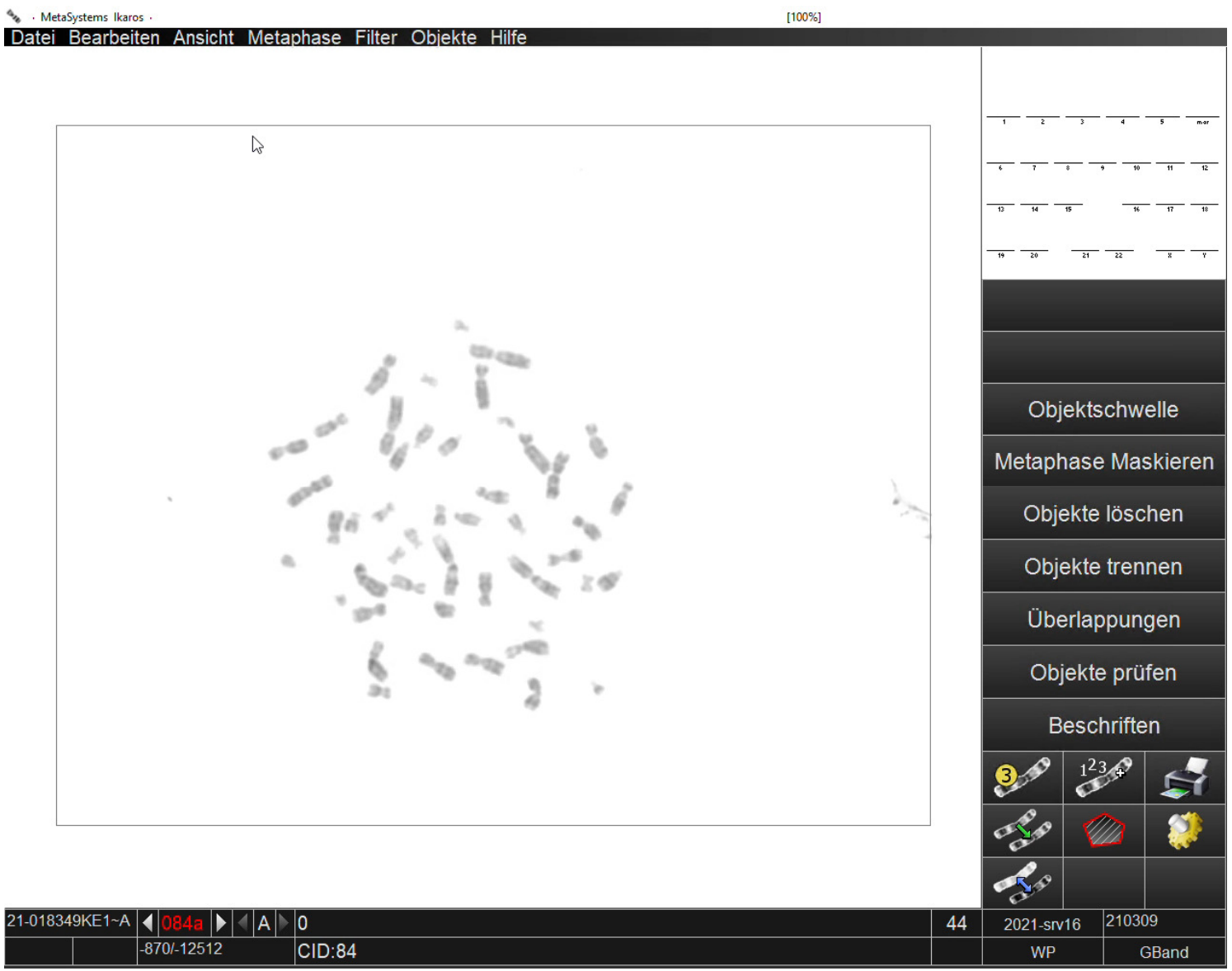
- Optimization to detect small populations & perform MRD
- To be used for all flow machines/workflows

Accuracy: 82-94%

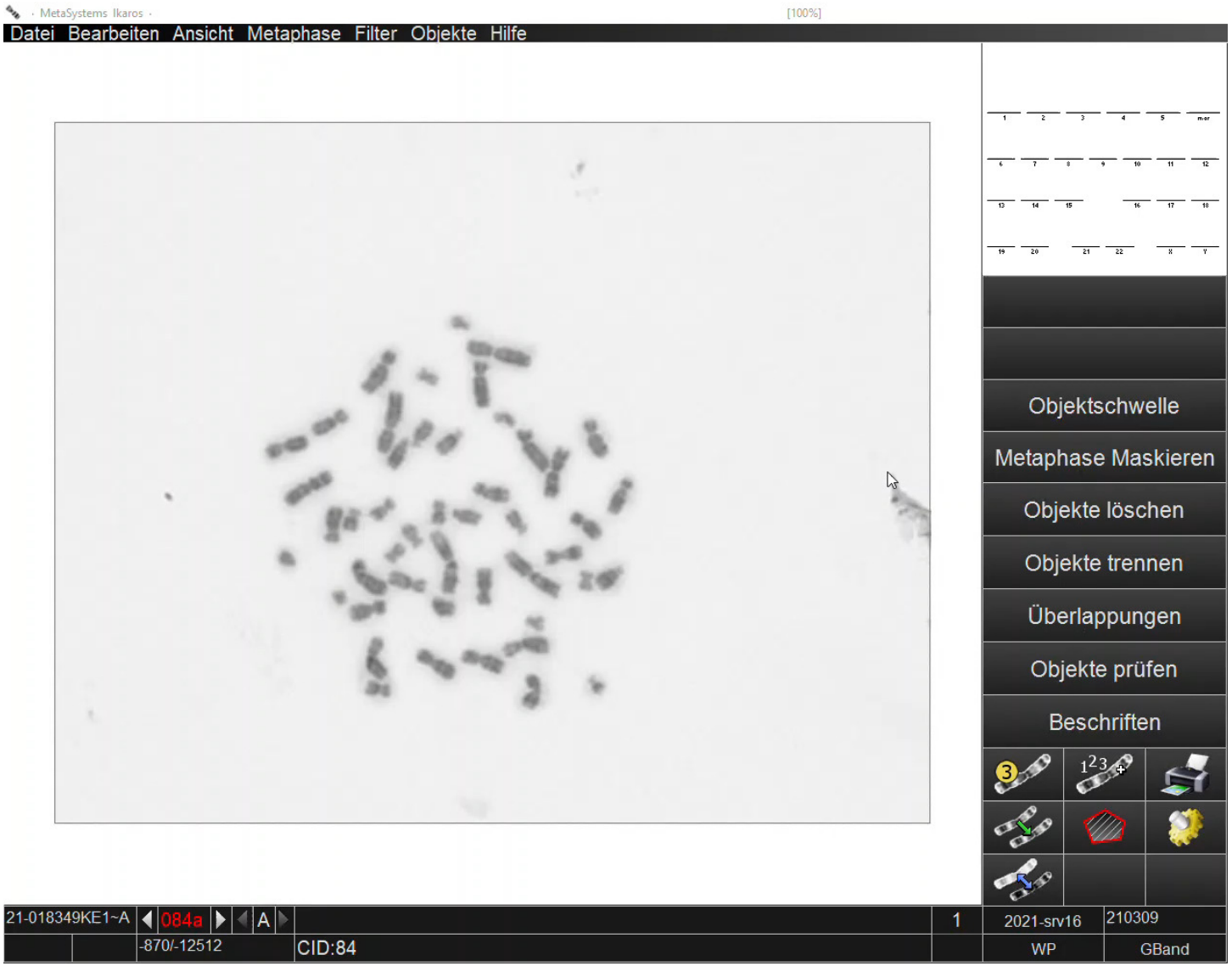
MLL data



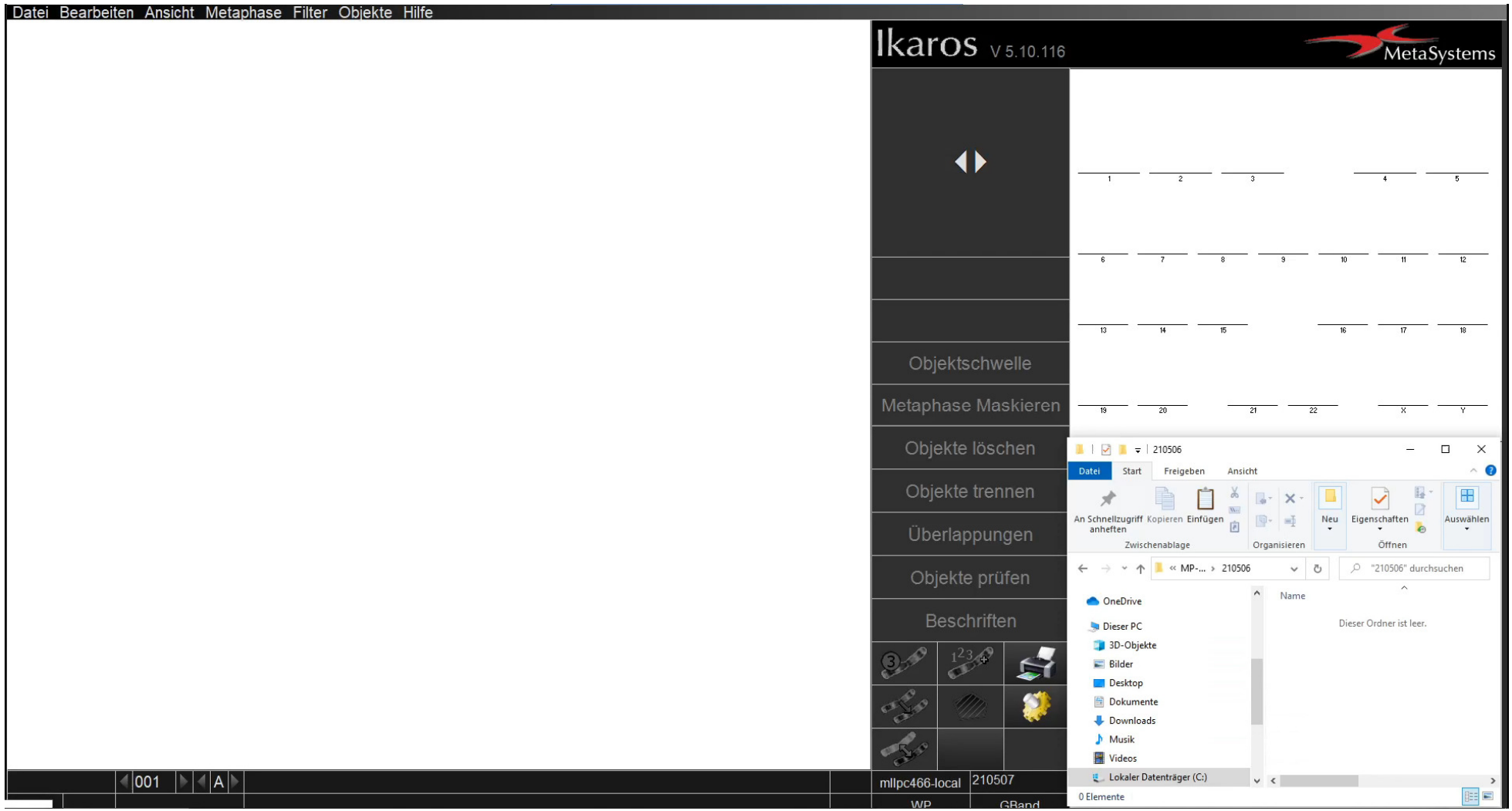
Manual classification

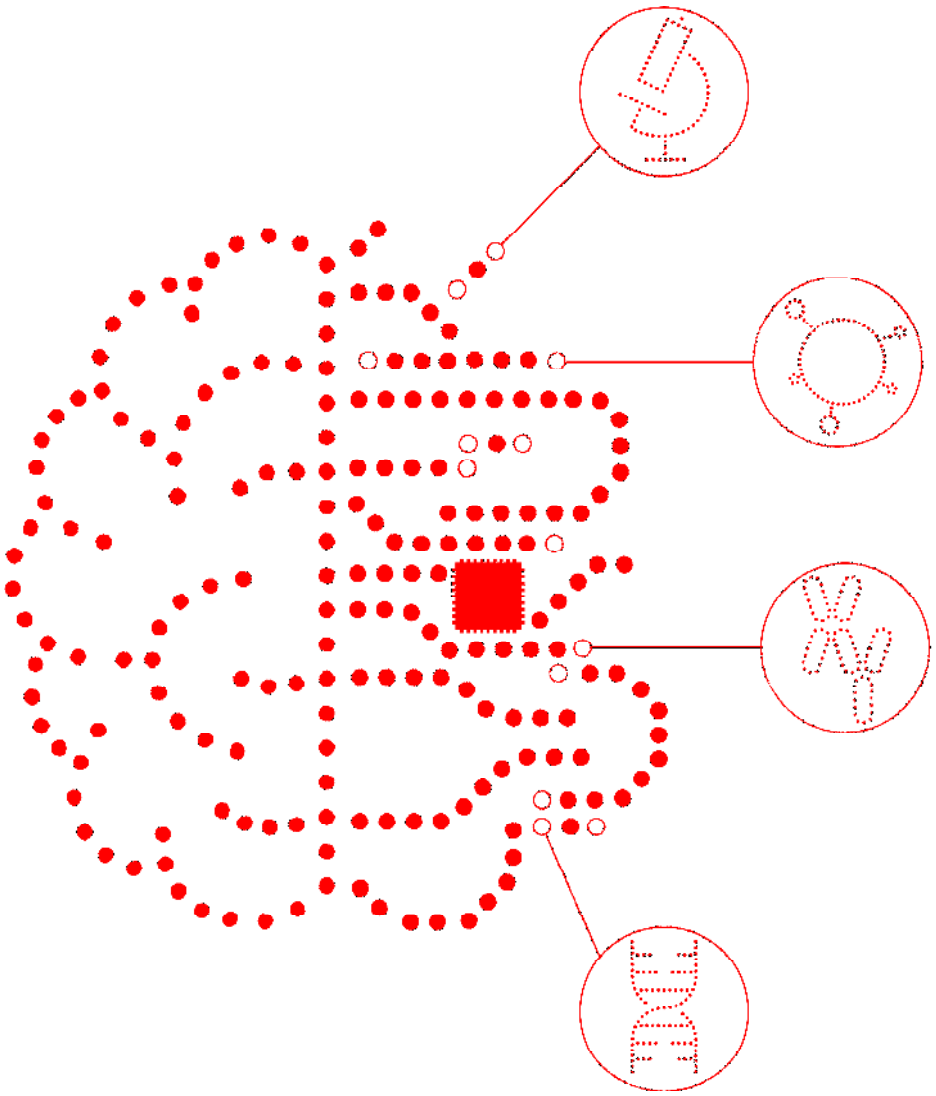


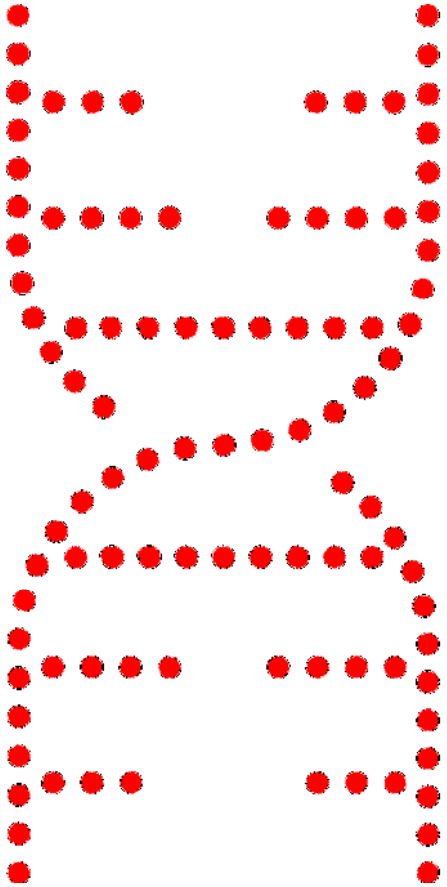
AI-based classification



AI-based classification – optimized version







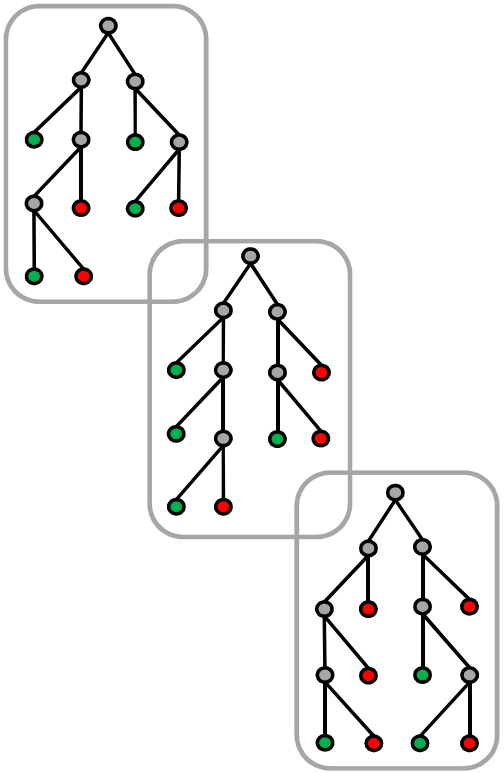
MLL Predictor: A hematological ensemble predictor

***TP53*: c.743G>A**

A variant with pathogenic *in silico* predictions

Predictor	Score
PROVEAN	-3.92
VEST3	0.872
M-CAP	0.735
SIFT	0.005
Polyphen-2	1.000
FATHMM	-7.28
FATHMM-MKL	0.980
Mutation Assessor	2.935
LRT	0.000
Mutation Taster	1.0

Random forest model



- Model trained on ~500 manually curated variants
- Variants where observed at least 10x in the MLL data set
- Variants where unambiguously classified as either “somatic mutation” or “benign polymorphism”

MLL Predictor	0.997

MLL data

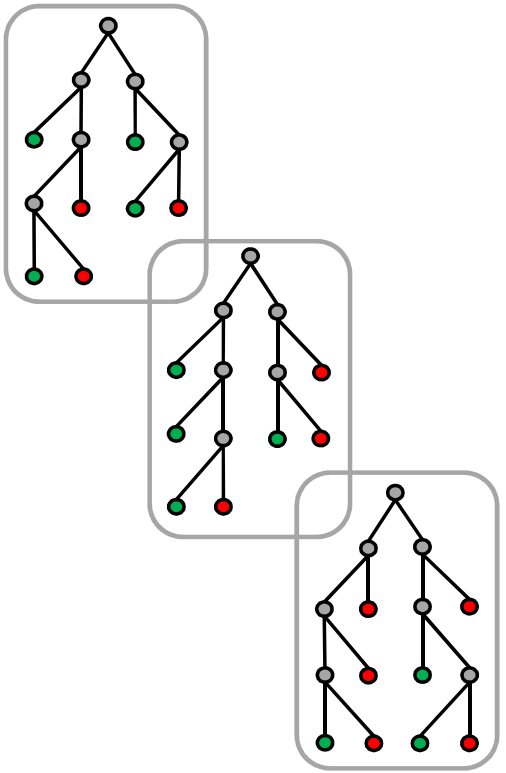
MLL Predictor: A hematological ensemble predictor

***TP53*: c.1027G>C**

A variant with conflicting *in silico* predictions

Predictor	Score
PROVEAN	-0.26
VEST3	0.175
M-CAP	0.047
SIFT	0.045
Polyphen-2	0.238
FATHMM	-3.23
FATHMM-MKL	0.815
Mutation Assessor	2.05
LRT	0.023
Mutation Taster	1.0

Random forest model



MLL Predictor	0.066

MLL data

Data Interpretation: NGS

= Variant annotation & interpretation

COSMIC

COSMIC ID	DNA	Protein	SNP	Somatic status	FATHMM-MKL	Count	Samples
COSM53042	c.2644C>T	p.R882C	No	Reported in another cancer sample as somatic Confirmed somatic variant	PATHOGENIC	442	more...
COSM87001	c.2644C>A	p.R882S	No	Reported in another cancer sample as somatic Confirmed somatic variant	PATHOGENIC	44	more...

DNA pos. may differ, as different transcripts are used. Query based on chromosomal coordinates.

DNMT3A
c.2644C>T

ClinVar

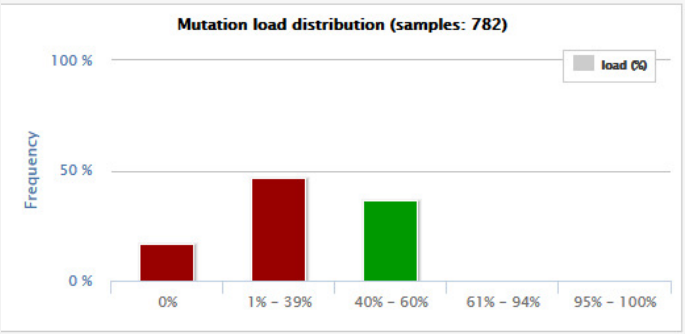
ID	HGVS	Type	Clinical Significance	Origin	ReviewStatus	Number Submitters	Last Evaluated	dbSNP	Cytogenetics	Guidelines	PhenotypeIDs
362761	c.2644C>T (p.Arg882Cys)	single nucleotide variant	Pathogenic/Likely pathogenic	somatic	no assertion criteria provided	1	May 31, 2016	rs377577594	2p23.3		MedGen MedGen MedGen OMIM OMIM Orpha Orpha SNOMED CT
362762	c.2644C>G (p.Arg882Gly)	single nucleotide variant	Pathogenic	somatic	no assertion criteria provided	1	Oct 02, 2014	rs377577594	2p23.3		MedGen OMIM Orpha SNOMED CT
362763	c.2644C>A (p.Arg882Ser)	single nucleotide variant	Pathogenic	somatic	no assertion criteria provided	1	Oct 02, 2014	rs377577594	2p23.3		MedGen OMIM Orpha SNOMED CT

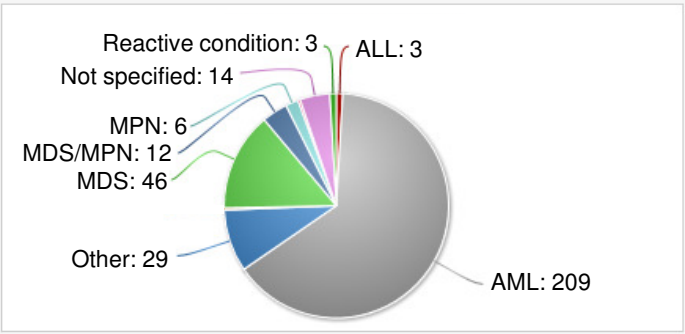
DNA pos. may differ, as different transcripts are used. Query based on chromosomal coordinates.

dbNSFP

Location (hg19)	ref	alt	AAref	AAalt	MLL Predictor	Ensemble Predictions	Individual Predictions	Alt. Allele Freqs
chr2:25457243	G	A	R	C	Pathogenic (1.000)	REVEL CADD DANN Eigen Eigen-PC more...	Mutation Taster PROVEAN VEST3 M-CAP SIFT more...	GnomAD: 0.0126% ESP_EA: 0.0465%

Mouse-over dotted-underlined key words for addtional information.





Variant interpretation

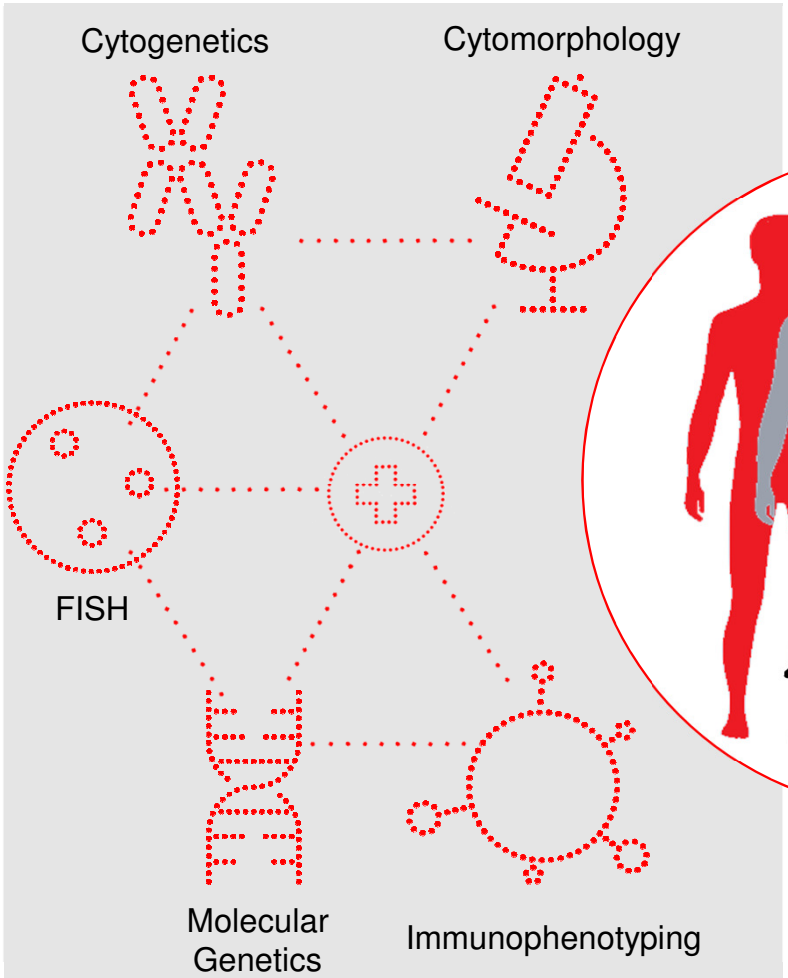
Variant annotation

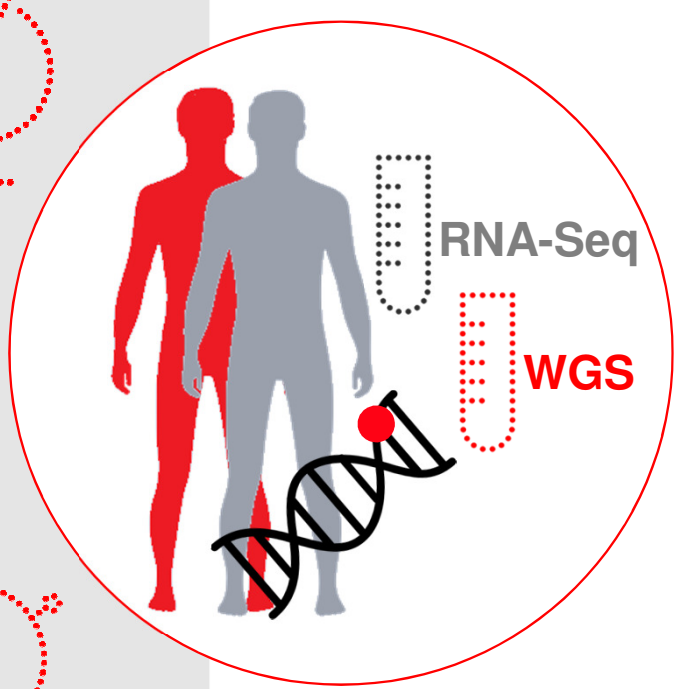
Variant interpretation

- DB (COSMIC, ClinVar, etc.)
- In-house database
- MLL predictor

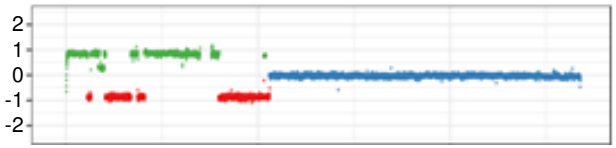
Report-ready variants

WGS and WTS as diagnostic tools

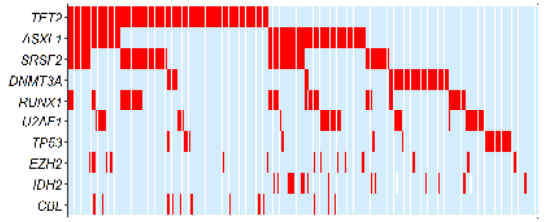




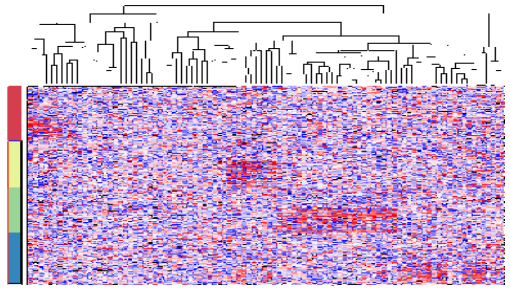
Copy number variants



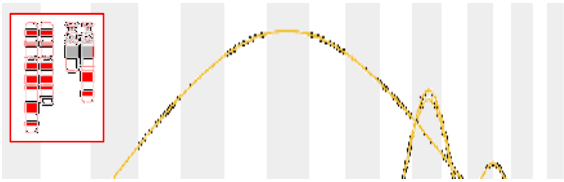
Mutation profiles



Gene expression



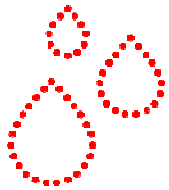
Structural variants
Fusion transcripts



Standard diagnostics

Diagnostics of the future (?)

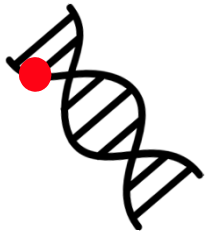
WTS
(1.5d)



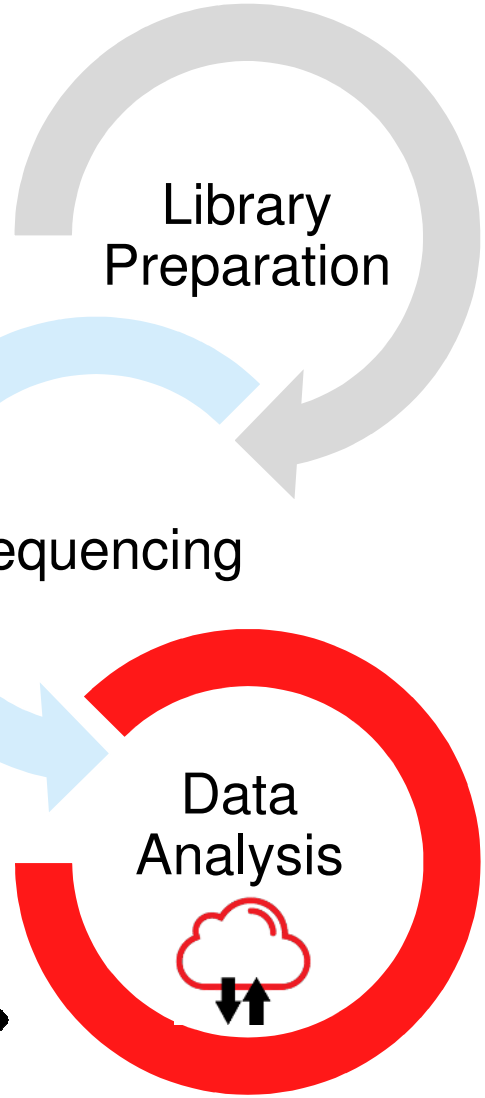
DNA
(~6hr)

WTS (~24hr)
50 mio reads
2x101 cycles

WGS (~44hr)
90x coverage
2x151 cycles

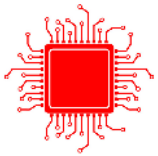


turn around time
5-7 days



Processing steps
- Fragmentation
- End repair
- Adapter Ligation
- Amplification (RNA)

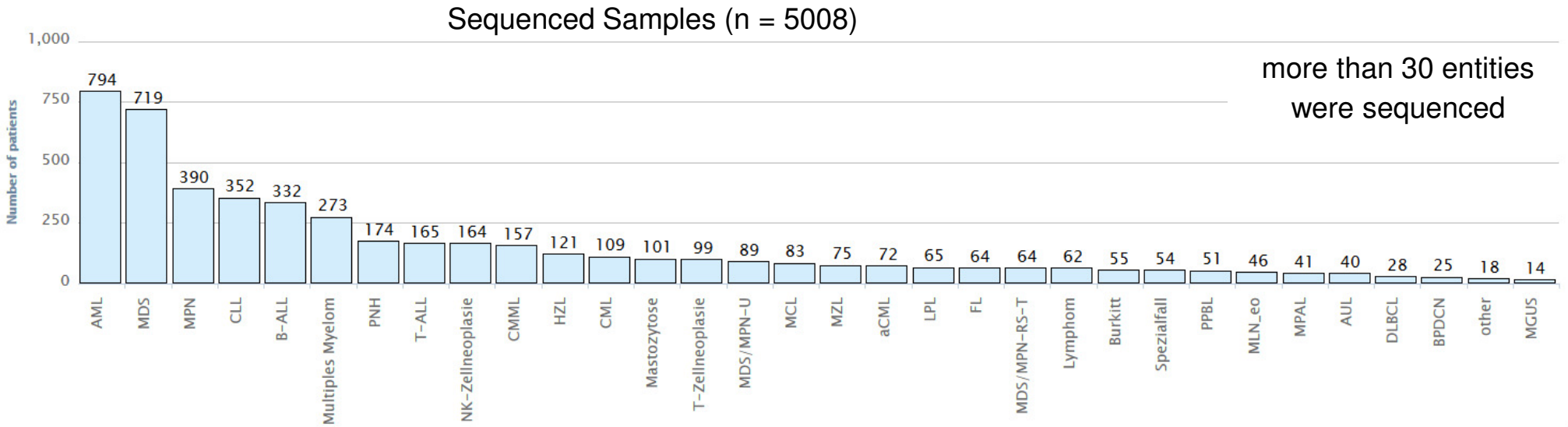
Data preprocessing (~7hr)
- FASTQ generation
- Alignment
- Variant calling (SV, SNV)



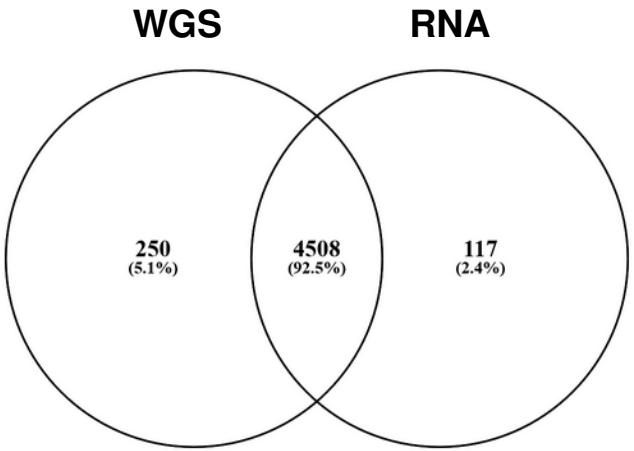
Data analysis
- Variant interpretation
- Gene expression
- CNV analysis
- SV analysis

Bridger

MLL 5k genomes
Project status



more than 30 entities were sequenced



WGS median coverage: 103x

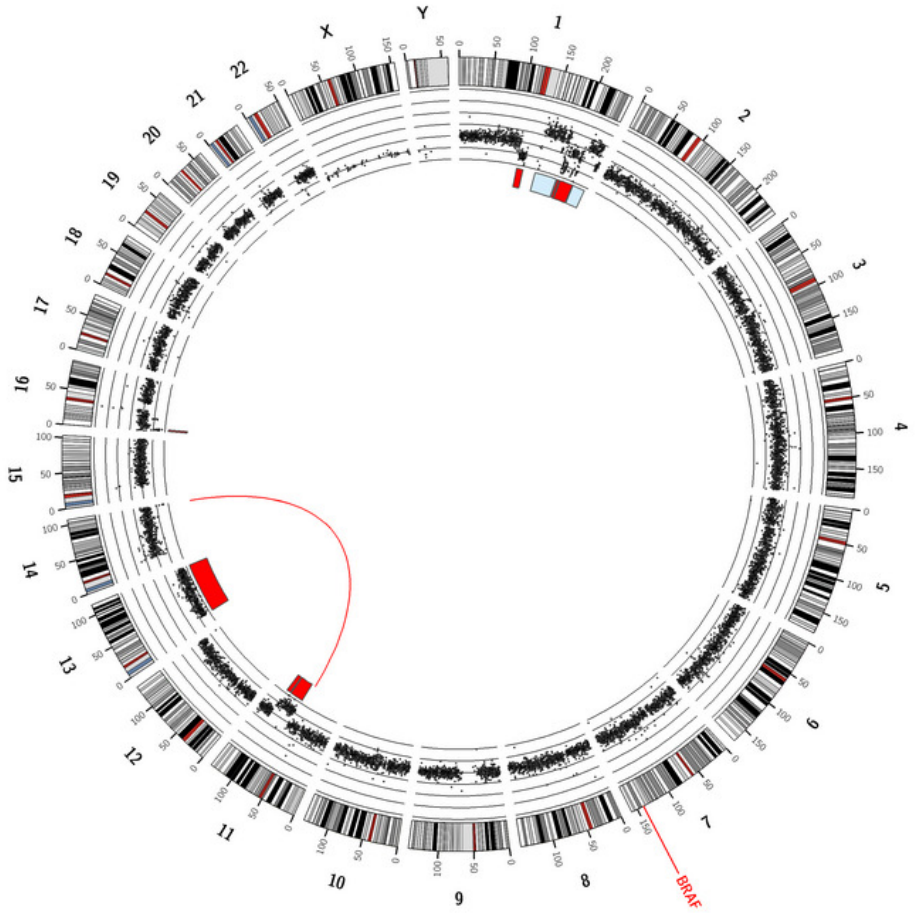
RNA median read count: 68 Mio

Additional 409 cases are in process or sequenced for other projects

Σ **5,417 genomes**

Integrated application for WGS data

Bridger | Select patient | Structural Variants | Copy Number Variants | Small Nucleotide Variants | Exon Coverage | RNA-seq | Final NGS report



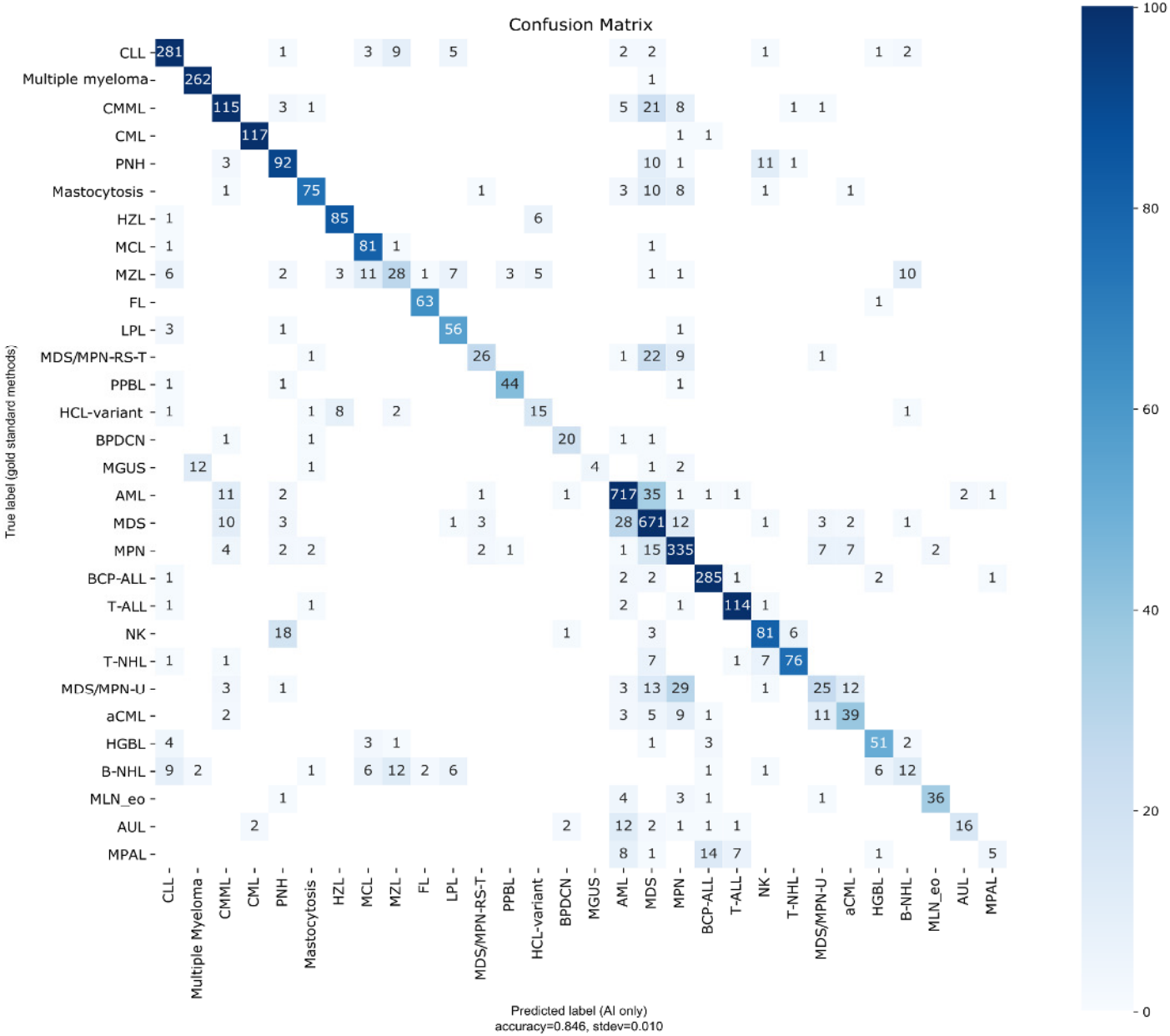
Copy Number Variants

- Gain
- Loss

Structural Variants

- Translocation

Report

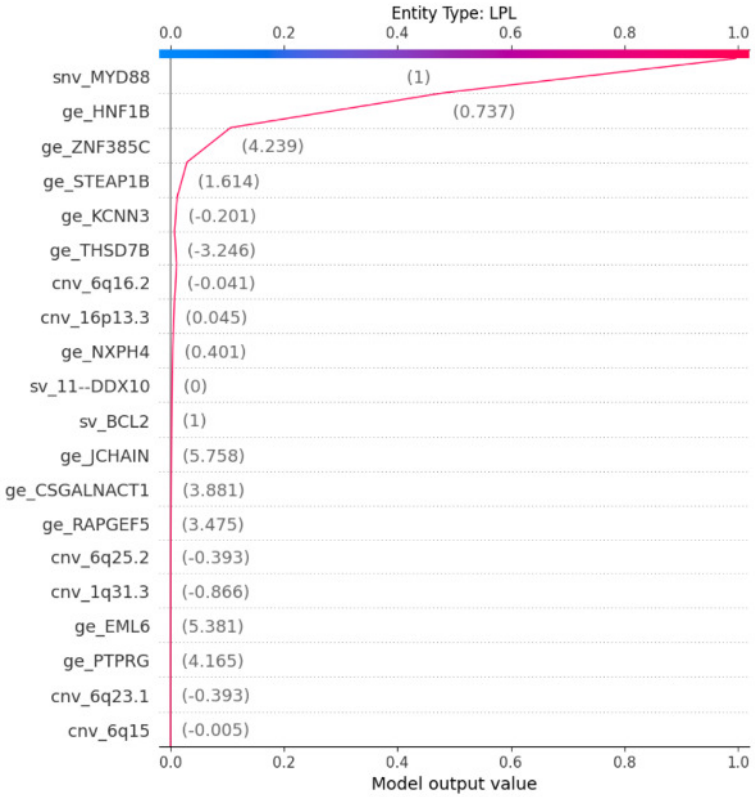


MLL 5k genomes

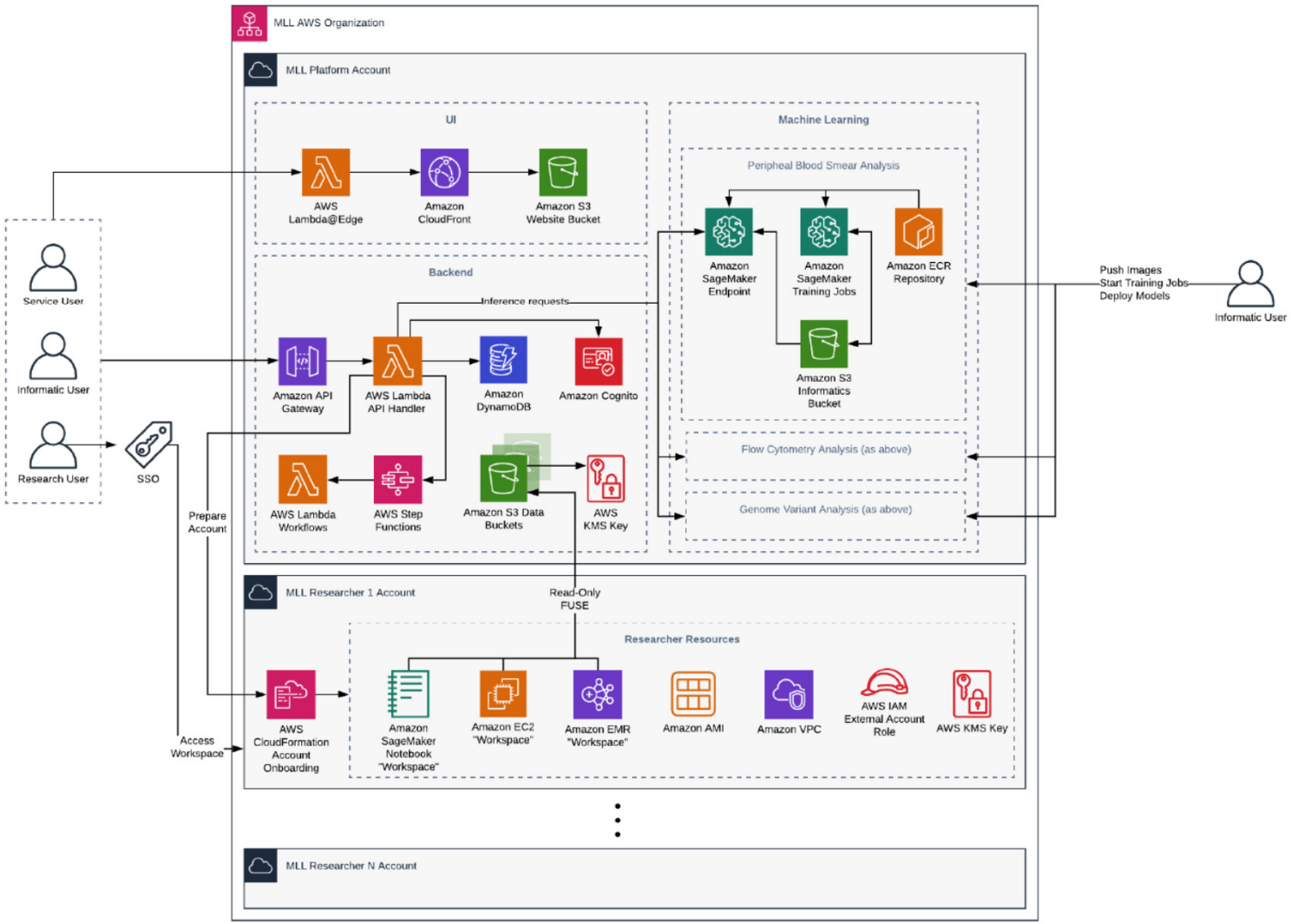
Confusion matrix for prediction

MLL data

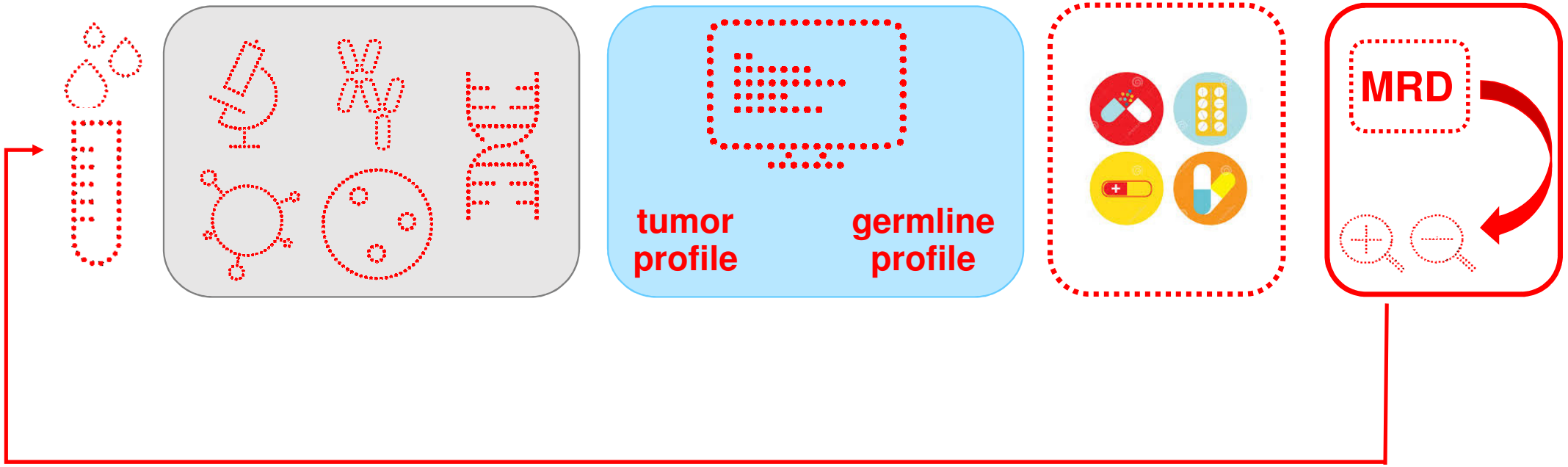
Sample decision plot



Cloud-based infrastructure for application of AI in hematology

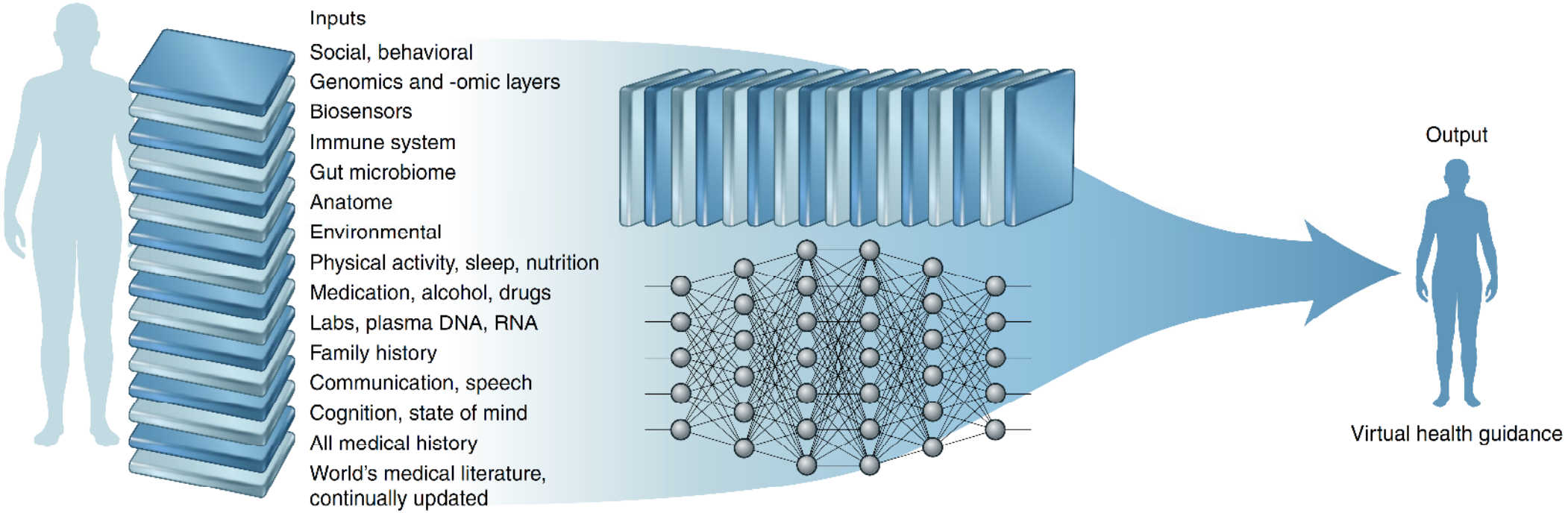


Personalized Medicine driven by Genetics



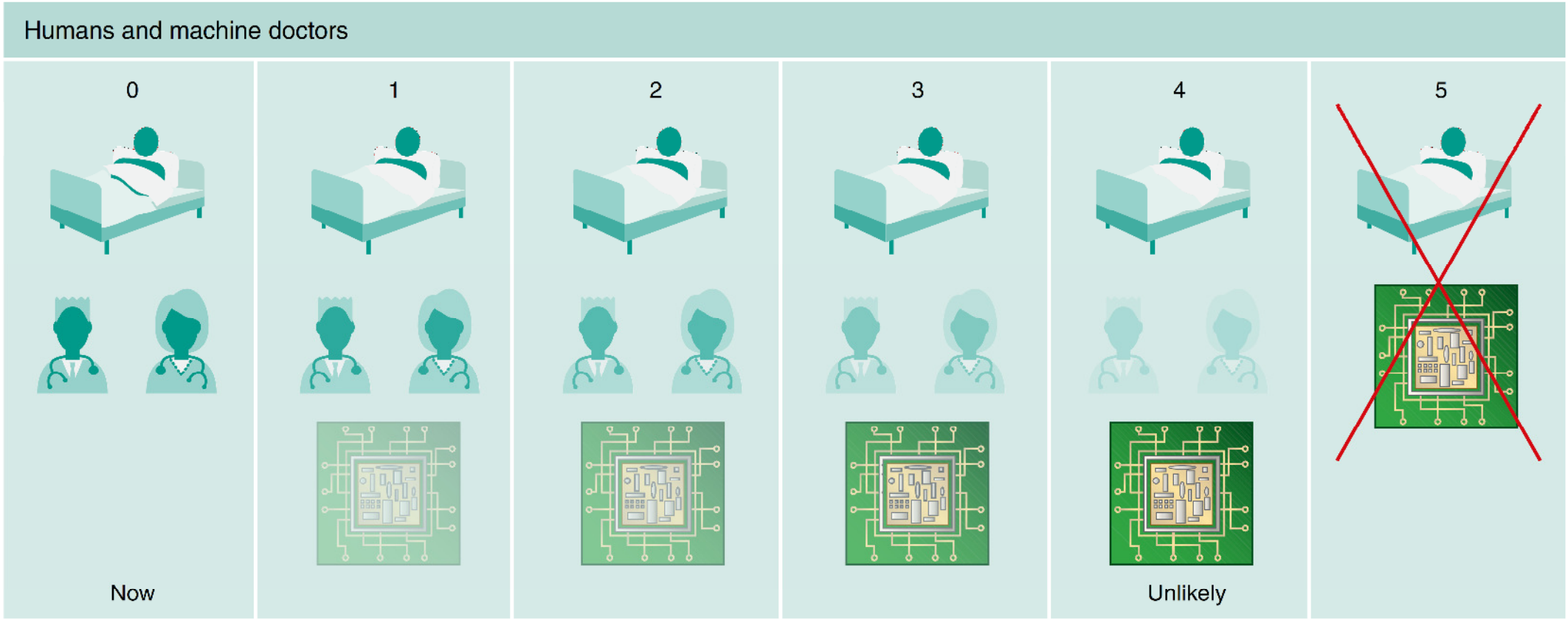
The virtual medical coach

Model for individualized guidance



E. Topol, Nature Medicine, 25, 44-56, 2019

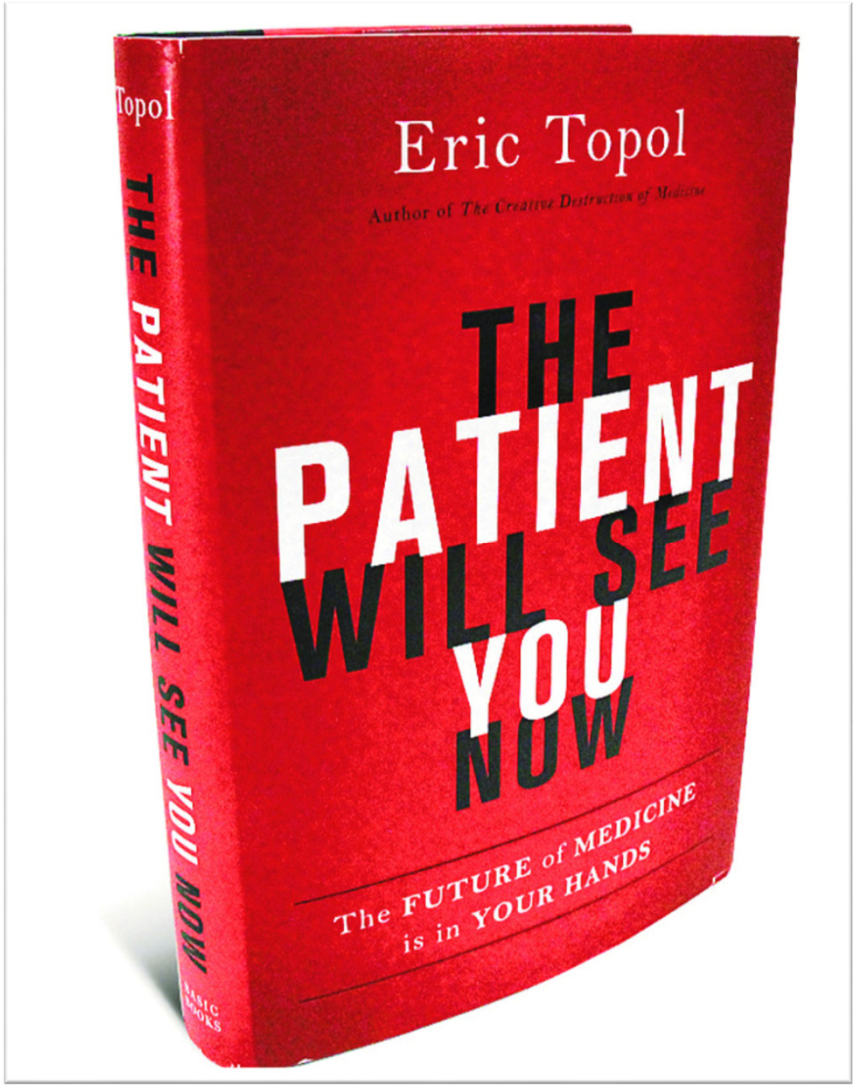
The analogy between self-driving cars and medicine



E. Topol, Nature Medicine, 25, 44-56, 2019

Deep Medicine

"AI will not replace physicians. However, physicians who use AI will replace those who don't."



Claudia Haferlach

Wolfgang Kern

Manja Meggendorfer

Wencke Walter

Niroshan Nadarajah

Stephan Hutter

See behind – Go beyond