

Machine Learning Supporting Precision Medicine: Real-World Examples from Oncology

Dr.-Ing. Matthieu-P. Schapranow TMF Workshop on Omics in Medical Research Dec 5, 2017

The Challenge Distributed Heterogeneous Data Sources





Hospital information systems

600GB per full genome

Human genome/biological data

15PB+ in databases of leading institutes

Often more than 50GB

	+9.5	144	
	1 1 1 1 1		
- 28	- E.		100
- 88	10 m	- 0	Ш.
8	2012		
- 10	10.	- 223	
- 10			
- 88			
- 88			
. 80	ι.		
- 88	i		п.

Cancer patient records >160k records at NCT



Human proteome

160M data points (2.4GB) per sample >3TB raw proteome data in ProteomicsDB



PubMed database >23M articles



Medical sensor data Scan of a single organ in 1s creates 10GB of raw data



Prescription data

1.5B records from 10,000 doctors and 10M Patients (100 GB)



Routine Data

Medical and treatment costs



Clinical trials Currently more than 30k

recruiting on ClinicalTrials.gov Schapranow, Omics in

Machine Learning Supporting Precision Medicine

Our Motivation Turn Precision Medicine Into Clinical Routine



- Can we enable clinicians to:
 - □ Incorporate <u>all available patient specifics</u>, e.g. *-omics data, lab results,
 - Access latest worldwide medical knowledge, and
 - □ Get questions answered <u>interactively</u>, e.g. during tumor board discussions.



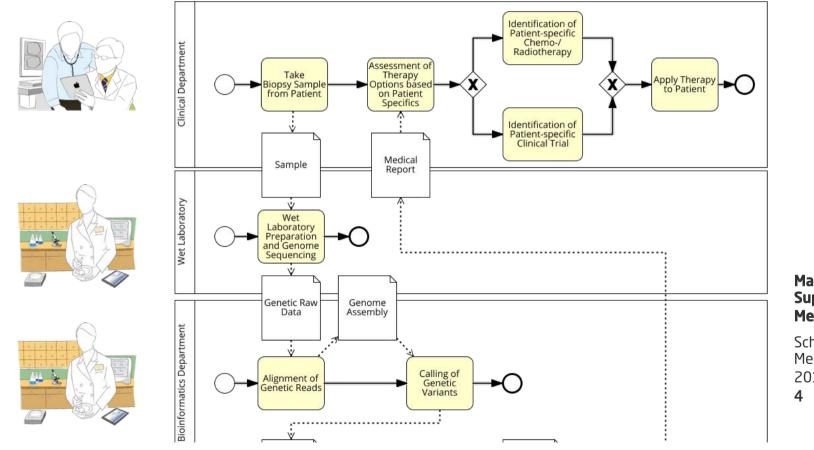




Machine Learning Supporting Precision Medicine

Simplified Clinical Oncology Process (1/2)



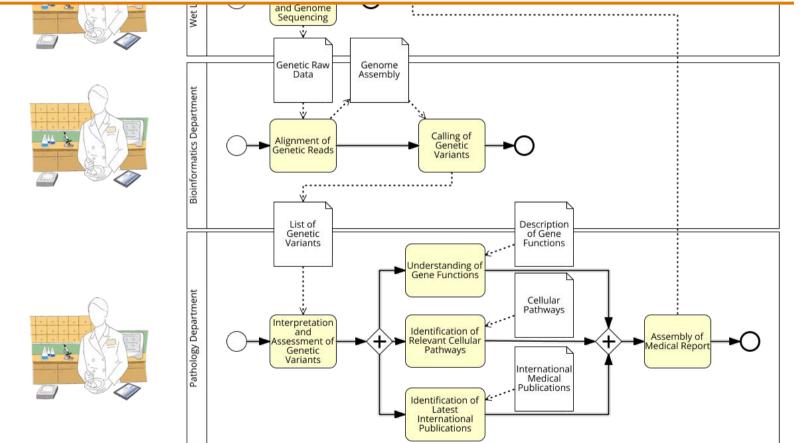


Machine Learning Supporting Precision Medicine



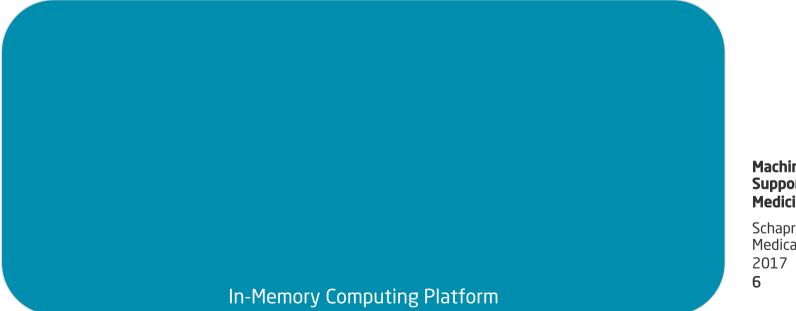
Simplified Clinical Oncology Process (2/2)





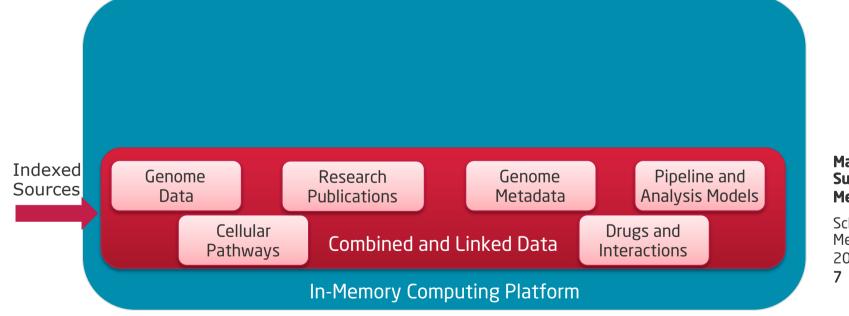
Machine Learning Supporting Precision Medicine





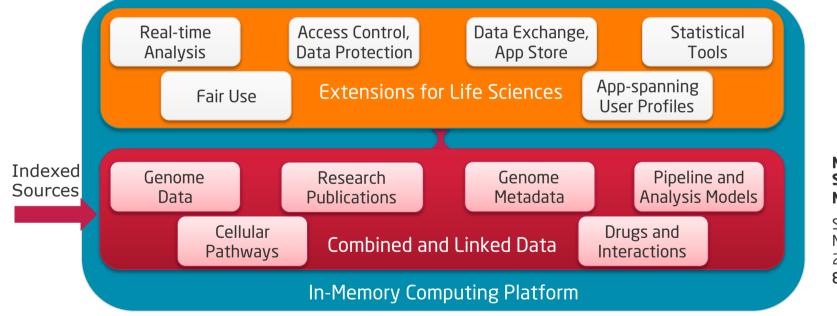
Machine Learning Supporting Precision Medicine



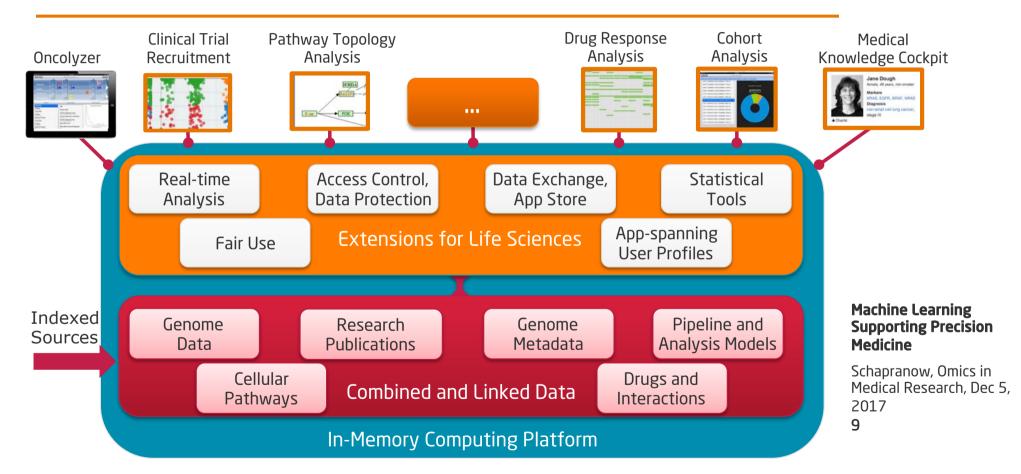


Machine Learning Supporting Precision Medicine





Machine Learning Supporting Precision Medicine



Hasso Plattner

Institut

Reproducibility Data Processing Modeling of Analysis Pipelines

- 1. <u>Design time</u> (researcher, process expert)
 - Definition of parameterized process model
 - $\hfill\square$ Uses graphical editor and jobs from repository
- 2. <u>Configuration time</u> (researcher, lab assistant)
 - Select model and specify parameters, e.g. aln opts
 - Results in model instance stored in repository
- 3. <u>Execution time</u> (researcher)
 - Select model instance
 - □ Specify execution parameters, e.g. input files

ALIGNMENT TRIAL SEARCH	COHORT ANALYSIS PAT	HWAYS SIDE EFFECT	S	
> Administration » Pipelines » Optimize	d			
Name	12270			
Optimized	*to		Alignment (Sepil_count)	
96	+ +			
			see Asignment_opt imized	
lefine a tag for this pipeline. Tags can se used to allow certain pipelines to be	_			
mbedded in other pipelines.	0		1	
lumber of files	0		Variant Calling (24)	
25	\otimes			
pecify the number of files that this lpeline accepts as input.		\bigcirc	San2Vid_s Volimport	\rightarrow \neg
	-		\square \square	
97 Hasso-Plattner-Institut	لما			
Hasso-Plattner-Institut Logged in as matthieu.schap Pofile Admin Log out 992600 ALIGNMENT TRIAL SEARCH	anow	PATHWAYS SI	DE EFFECTS	
Hasso-Plattner-Institut Logged in as matthieu.schapr Poffile Admin Log out 992600 ALIGNMENT TRIAL SEARCH >> Alignment	anow	PATHWAYS SI Pipeline category	DE EFFECTS Pipeline	PROCESS
Hasso-Plattner-Institut	anow 4 COHORT ANALYSIS			PROCESS

App Example: Medical Knowledge Cockpit for Patients and Clinicians



the squarts		# DOP toral change
CONTRACT OF TAXABLE PARTY OF TAXABLE PARTY.		mile Carean
Contraction in contraction of the local division of the local divi		and and a second second
	(here)	round .
Read and the Addation (1995), 49	- Tell - 1	tend had using planam.
		and dear one many a
	AND AND DOUBLE TO A DOUBLE T	2573
tanuar bana banaus (?	8 Arrupe hours 2	
stat	Found ADD documents in 1273 indiana anna	
Name official States of Antony 7 and Antony Results Results parameter of Antony States and Antony	train. The point therapy in him bank Call Long Cannar (2014) 2 Januar	
Electronic balance in the second seco	2012 Temprined Thermative for Non-analitiCall Long Cannon Transition Longitude	
Manada and Andrew State	1000 Recented to Descentional Parage and Targeted Mintechel Testage Investories and the Investories Respirations	
and an experimental point for the point of the second statements and the second statements and the second statements are set of the second statements are	2010 Rear Infrastructure In the Provision Balance of Long Earson Proof Tagle Trans. Name Log and Analysis Inter-	
Anni in orbit, security more; Eastan Apostone : Facilitat anglorector Apostone : Apostone : Apostonector	2010 Epidemic Second Participation Institutions in the Designed of Second (14) Lang Cancer Participation for Second Participation (14)	
Environmental learning (1/13 pr) 1 Denominational gain, (1)(1)	100 EVFE Relations, Other Reference International Related To Security in 2009 Interfaces, and Reference Long California	Teating for \$12795. Terrigeneek Transporter to Non-Board Card
Moderan Ages A. D. Sto Turner Agest garmony . 2022.2	Partit Derival	
	BUT Bale of decoupting to files that Cell Long Conner Textment (ant-branes table frequest, state-bags Rear Sev. Salar Sev.	
Selected ()	2 2010 Research progress is the sam of contributors of platious based channels rays and aptioned growth anyon vary factor they limite.	factor receptor to come interface.
RALA	BPE EXPECTATION AND A CONTRACT AND A	
This proc. a forefact on on open follows from the representation on gene letting - microsof a protein that is a neutrinor of the simal UT follow apportances.		
100	Pathetins /	
60/8		-
The prime excessed by the point 4.2 households are promotion flight 4.4 metrics for prime increase reportering	 Mass Massission and expension of the function of any processing and a second processing and procesing and processing and processing and processing and processin	
	· Charles and a statistics of these hadrest- by price the study character hadronic trans-	
1717a	Plan mining and parameters and become the second and the seco	
Tellines 2	 Contrary Mail and MAM metalant is caulty for a colonization 	
	F	
Dania	194	
	· TTP advanta is increased by material plantations and advantation in the advantation	
trad equating pathway	 Excepts to application growth factor includes factory 	
on designed (unter	• Externa from Factor constraints for providing the second • External from Factor constitutions appropriate of the Pactor Constitution Factor Income.	
ton-staff out tung carrier	 A Approximation size on the assessed point large research 	
Net agent (salar)		
and the second s		
	Desit fue ()	
	 Measured Trace 1 Marc of Across and and Trace Acad control (47) which the Charles of A Second Secon	
	 ED-10. Caregorite/Pacificati (Htt or Nithout Benargaria) and to "Statement in Stage V or Records from Study Co in Nites and Discontinger, of Long Adamse processor or Study offers Result Records. 	create cases
	a bit state ma counter o had county (and	
	the second se	

 Search for affected genes in distributed and heterogeneous data sources

- Immediate exploration of relevant information, such as
 - Gene descriptions,
 - Molecular impact and related pathways,
 - Scientific publications, and
 - □ Suitable clinical trials.
- No manual searching for hours or days: In-memory technology translates searching into interactive finding!

Machine Learning Supporting Precision Medicine

Schapranow, Omics in Medical Research, Dec 5, 2017 11





Automatic clinical trial matching build on text analysis features

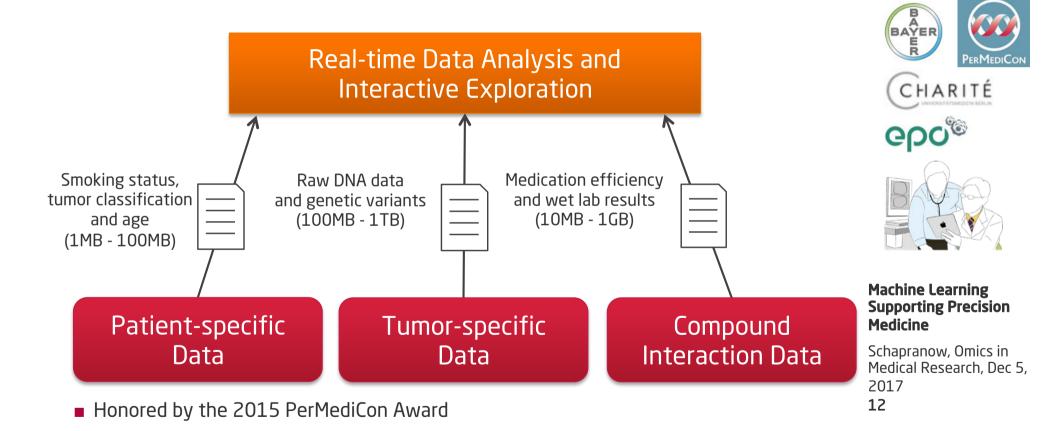
and un-structured data

sources

Unified access to structured

App Example: Identification of Optimal Chemotherapy





	ଟ – ≐ ଅ≵ଇ
HPI Hasso-Plattner-Institut GENOMES	Design IT. Create Knowledge.
Configuration Results	
	First step: Choose which drugs'values should be predicted
5-FU	First step: Choose which drugs values should be predicted
 Methotrexat Carboplatin Docetaxel 	or choose
Cetuximab	platin to c
	carbowin

Second step: Choose which data should be factored into the prediction. You can choose a preset (click on the row) or create an own set.

Available Presets

	Drug	Drug	Non-functional	Functional Changes via	Functional Changes via						
Preset Name	T/C	Recist	Changes	RS	Genes	Age	Gender	т	Ν	M	Grading
T/C Basic	-		~								
Recist Basic		1	1								
T/C Functional RS	-			1							
Recist Functional RS		1		1							
T/C Basic Complete	~		-			~	~	~	~	~	-
Recist Basic Complete		1	1			-	-	-	~	~	-
T/C Functional RS Complete	~			1		~	~	~	~	~	-
Recist Functional RS		-		1		1	-	~	-	-	1

zegenomes.com/dra/clustering/svm_tumor ANALY	ZE D		
HPI Hasso-Plattner-Institut GENOM	ES		Design IT. Create Knowledge
Predict Drug Responses			
OR			
Classify Drug Responses			
	lass borders, which distinguishes effect		
You can choose to use a	predefined set, or define your own cla	asses and save them for later use.	
Available Classi	fication Presets Choose of	na hy diaking an a tabla raw	
		The by clicking on a table row	
Preset Name	Definition	(h) 20 (h-1)	
Recist Basic	0.7 (good) 1.2 (sta		
T/C Basic		stable) 100.0 (bad)	
test2	1.0 (good) 10.0 (o)	k) 50.0 (bad)	
Custom Set Give	a class a name and a class border	r	
	T/C or recist values to be classified, since they		
The lowest class border will go	o from 0 to the defined border. Subsequent	t classes set the end of the defined class (so the class ran	ges from the predecessor class border to the defined border for
	ier (= the last class) will include all values, whic and 100 (bad) - good ranges from 0-25, stable	ch do not fit into any other class, so the last border value is e from >25 to 75, bad from >75 upwards.	s actually not that important.
Add Class:	Name	Border	+
	Current Classes:		
Save Class:	To save this set, enter a name		

🖏 Drug Response Analysis 🛛 🗙 💽		🖆 - 8 📩
+ -> C 🗋 we.analyzegenomes.com/dra/clusteri	ig/svm_tumor	Q 🚖 🖪 🗏
Hasso-Platt		Design IT. Create Knowledge.
Configuration	Results	
	A wun far Carbonlatin	

Result SVM run for Carboplatin

	Drug Class Probabilities Recist			
Tumor	good	stable	bad	Actual Class
10927	0.067	0.205	0.728	bad

Result SVM run for Cetuximab

	Drug Class Probabilities Recist			
Tumor	good	stable	Actual Class	
10927	0.469	0.200	0.331	good

cetuximab might be more beneficial for the current case

transkript

Brauchen wir eine Debatte über künstliche Intelligenz?





30.11.2017

Prof. Dr. Roland Benedikter, Center for Advanced Studies, Eurac Research Bozen (Foto: Univ. Breslau); Dr.-Ing. Matthieu Schapranow, Program Manager E-Health & Life Sciences, HPI, Potsdam

Während der Politikanalytiker Roland Benedikter angesichts großen Transformationspotentials eine breite Debatte über künstliche Intelligenz fordert, wollen Forscher sie endlich nutzen dürfen. Machine Learning Supporting Precision Medicine

Stay in Contact!



Dr.-Ing. Matthieu-P. Schapranow Program Manager E-Health & Life Sciences

> Hasso Plattner Institute August-Bebel-Str. 88 14482 Potsdam, Germany

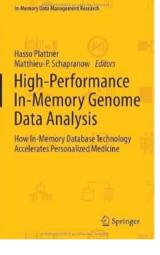
we.analyzegenomes.com



Federal Ministry for Economic Affairs and Energy

on the basis of a decision by the German Bundestag





Machine Learning Supporting Precision Medicine