Hasso Plattner Institut

IT Systems Engineering | Universität Potsdam

Exercise III

Harry Freitas da Cruz

Data Management for Digital Health

Summer 2017

Exercise III

- Genome Browser
- Medical Knowledge Cockpit
- Use Case Heart Failure (Systems Medicine)

Hasso Plattner Institut
IT Systems Engineering | Universität Potsdam

Exercise III Part 1
Genome Browser and Medical Knowledge Cockpit

Harry Freitas da Cruz, Milena Kraus

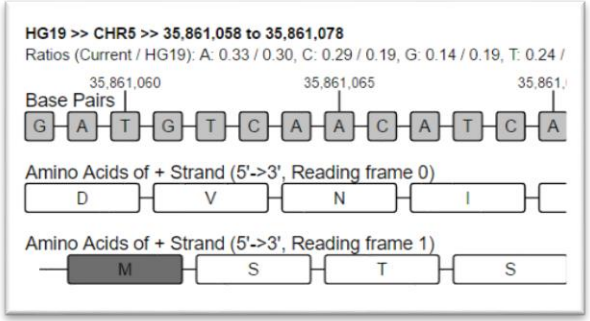
Hasso Plattner Institut

July 18, 2017

Assessment of Genetic Variants and Medical Knowledge Exploration

Hasso Plattner Institut

- **Genome Browser**
  - Interactive exploration of genetic variants
  - Identify SNP/InDels
  - Distinguish between functional and non-functional variants
  - Identify associated diseases harvesting int'l knowledge databases
- **Medical Knowledge Cockpit**
  - Interaction exploration of medical knowledge data sources
  - Patient-centered queries
  - Analysis medical data in real time
  - Links back to primary data sources



Genome Browser
Tasks and variants

Alignment tasks

- User-configured pipelines
- Parallelization of jobs
- Task scheduling

List of variants

- Chromosome positions
- Associated diseases
- Affected genes

» Tasks

Task List

Pages: 1 2 3 4 5 6 ... 23 (30 rows per page)

New alignment task

Results	Cost
1942578	Free
327960	Free
333180	Free
276012	Free
274805	Free
1601151	Free
1507363	Free
1377275	Free

» Tasks » Task #2276

Results for task #2276

Pages: 1 2 3 4 5 6 ... 64753 (30 rows per page)

Statistics: Jobs | Workers

Chr	Pos	Ref	Alt	Qual	Other syndrome disease	Name	Gene
5	35861068	T	C	35.80	Severe combined immune deficiency	interleukin 7 receptor	IL7R
14	81444967	G	A	32.80	Hereditary nonautoimmune hyperthyroidism; subclinical hypothyroidism	thyroid stimulating hormone receptor	TSHR
10	123330541	C	T	16.90	Crouzon, Pfeiffer, and Apert syndromes	fibroblast growth factor receptor 2	FGFR2
2	223151465	C	T	30.80	Waardenburg syndrome; craniofacial-deafness-hand syndrome	paired box gene 3	HUP2
14	81454564	C	T	3.54	Hereditary nonautoimmune hyperthyroidism; subclinical hypothyroidism	thyroid stimulating hormone receptor	TSHR
3	12352688	G	C	6.98	Insulin resistance ; lipodystrophy, familial partial L;diabetes mellitus, insulin-resistantl, with acanthosis nigricans and hypertension	peroxisome proliferative activated receptor, gamma	PPARG
5	176682734	C	T	5.46	Sotos Syndrome	nuclear receptor binding SET domain protein 1	NSD1
3	69883623	G	A	3.54	Waardenburg syndrome type 2, Tietz syndrome	microphthalmia-associated transcription factor	MITF

Exercise III

Data Management for Digital Health, Summer 2017

Code of Life - When Computer Science Meets Genetics Dr.-Ing. Matthieu-P. Schapranow

Course is available

Test Exercise

Instructions:

These are the instructions for the test.

This is an ungraded exercise. 0.0 points

Previous Launch exercise tool Next

ANALYZE GENOMES
by Hasso-Plattner-Institut

200

TASKS | TRIAL DESIGN | MEDICAL KNOWLEDGE COCKPIT

» Tasks

Task List

Pages: 1 2 3 4 5 6 ... 23 (30 rows per page)

New alignment task

ID	Progress	Pipeline	Results	Cost
2276	Completed in 07h06m47s	BWA_s	1942578	Free
2265	Completed in 05h54m31s	BWA_s	327960	Free
2264	Completed in 02h25m56s	BWA_s	333180	Free
2263	Completed in 05h22m28s	BWA_s	276012	Free
2262	Completed in 05h17m58s	BWA_s	274805	Free
2261	Completed in 14h39m52s	BWA_s	1601151	Free
2260	Completed in 14h59m17s	BWA_s	1507363	Free
2259	Completed in 17h34m42s	BWA_s	1377275	Free
2258	Completed in 16h31m59s	BWA_s	1900591	Free
2257	Completed in 08h21m43s	BWA_s	1647173	Free
2256	Completed in 08h12m04s	BWA_s	1698232	Free
2255	Completed in 20h50m50s	BWA_s	187279	Free
2250	Completed in 10h34m54s	BWA_s	2044480	Free

FEEDBACK

Results for task #2243

Pages: 1 2 3 4 5 6 ... 71067 (30 rows per page)

Statistics: Jobs | Workers

Chr	Pos	Ref	Alt	Qual	Other syndrome disease	Name	Gene
5	35871190	G	A	82.30	Severe combined immune deficiency	interleukin 7 receptor	IL7R
5	35861068	T	C	7.80	Severe combined immune deficiency	interleukin 7 receptor	IL7R
22	36684354	T	C	6.98	Deafness, autosomal dominant 17, Epstein syndrome, Fechtner syndrome, May-Hegglin anomaly, Sebastian syndrome	myosin, heavy polypeptide 9, non-muscle	MYH9
3	12393125	C	G	7.80	Insulin resistance ; lipodystrophy, familial partial L;diabetes mellitus, insulin-resistantl, with acanthosis nigricans and hypertension	peroxisome proliferative activated receptor, gamma	PPARG
10	8116241	GAAAAAAAAAAA	GAAAAAAAAAAAA	4.58	HDR syndrome (HYPOPARATHYROIDISM, SENSORINEURAL DEAFNESS, AND RENAL DISEASE)	GATA binding protein 3	GATA3
3	12429385	atttt	aTtttt	11.50	Insulin resistance ; lipodystrophy, familial partial L;diabetes mellitus, insulin-resistantl, with acanthosis nigricans and hypertension	peroxisome proliferative activated receptor, gamma	PPARG
4	55567888	A	C	7.80	Piebald trait	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog	KIT
10	123249057	T	C	4.13	Crouzon, Pfeiffer, and Apert syndromes	fibroblast growth factor receptor 2	FGFR2
10	43585621	T	G	21	Hirschsprung disease	ret proto-oncogene	RET
3	69874515	A	G	6.20	Waardenburg syndrome type 2, Tietz syndrome	microphthalmia-associated transcription factor	MITF
14	81457257	A	C	3.54	Hereditary nonautoimmune hyperthyroidism; subclinical hypothyroidism	thyroid stimulating hormone receptor	TSHR
3	69845553	A	T	58	Waardenburg syndrome type 2, Tietz syndrome	microphthalmia-associated transcription factor	MITF
7	140454973	G	T	14.20	Cardio-facio-cutaneous syndrome	v-raf murine sarcoma viral oncogene homolog B1	BRAF

ANALYZE GENOMES by Hasso-Plattner-Institut

200

TASKS | TRIAL DESIGN | MEDICAL KNOWLEDGE COCKPIT

» Tasks » Task #2243 » Chromosome 5 » Position 35861068

HG19 >> CHR5 >> 35,861,058 to 35,861,078

Ratios (Current / HG19): A: 0.33 / 0.30, C: 0.29 / 0.19, G: 0.14 / 0.19, T: 0.24 / 0.30, N: 0.00 / 0.02

35,861,060 35,861,065 35,861,070 35,861,075

Base Pairs: G A T G T C A A C A T C A C C A A T C T G

Amino Acids of + Strand (5'->3', Reading frame 0): D V N I T N L

Amino Acids of + Strand (5'->3', Reading frame 1): M S T S P I

Amino Acids of + Strand (5'->3', Reading frame 2): C Q H H Q S

Amino Acids of - Strand (3'->5', Reading frame 2): T L M V L R

Amino Acids of - Strand (3'->5', Reading frame 1): H * C * W D

Amino Acids of - Strand (3'->5', Reading frame 0): I D V D G I Q

35,861,060 35,861,065 35,861,070 35,861,075

Opposite: C T A C A G T T G T A G T G G T T A G A C

35,861,060 35,861,065 35,861,070 35,861,075

2243: G A T G T C A A C A C C A C C A A T C T G X

Amino Acids of + Strand (5'->3', Reading frame 0): D V N T T N L

Amino Acids of + Strand (5'->3', Reading frame 1)

Reading frames

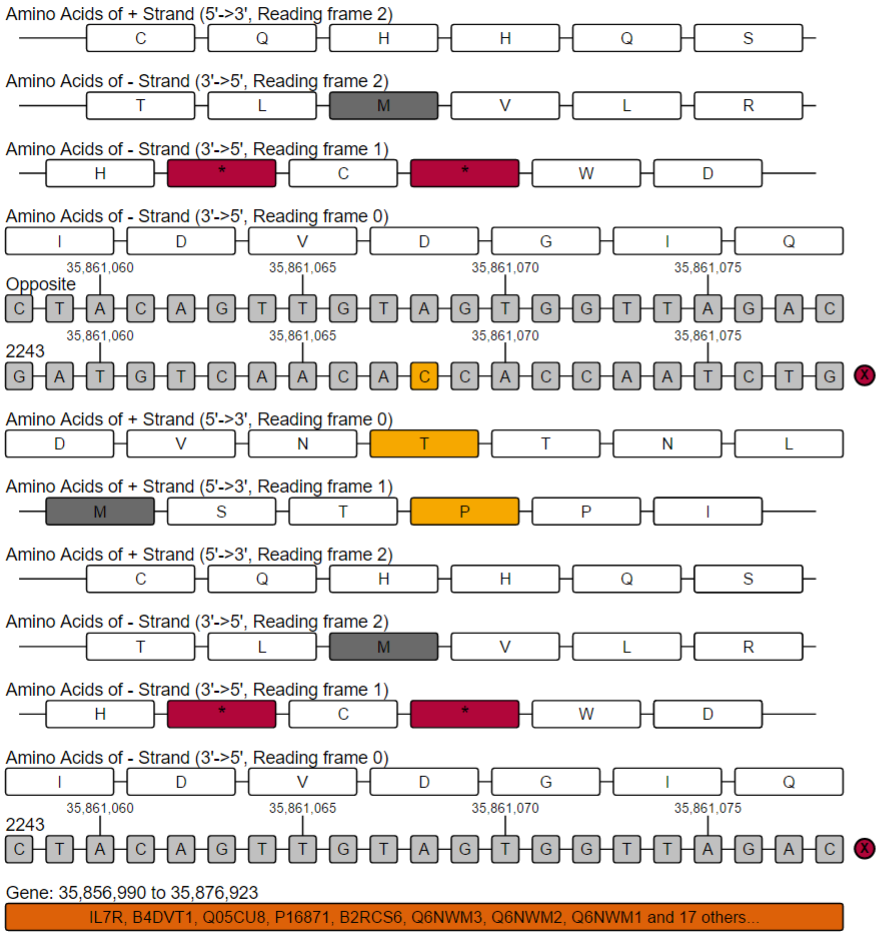
Add a Cell Line

Please Choose a Cell Line

26

ADD

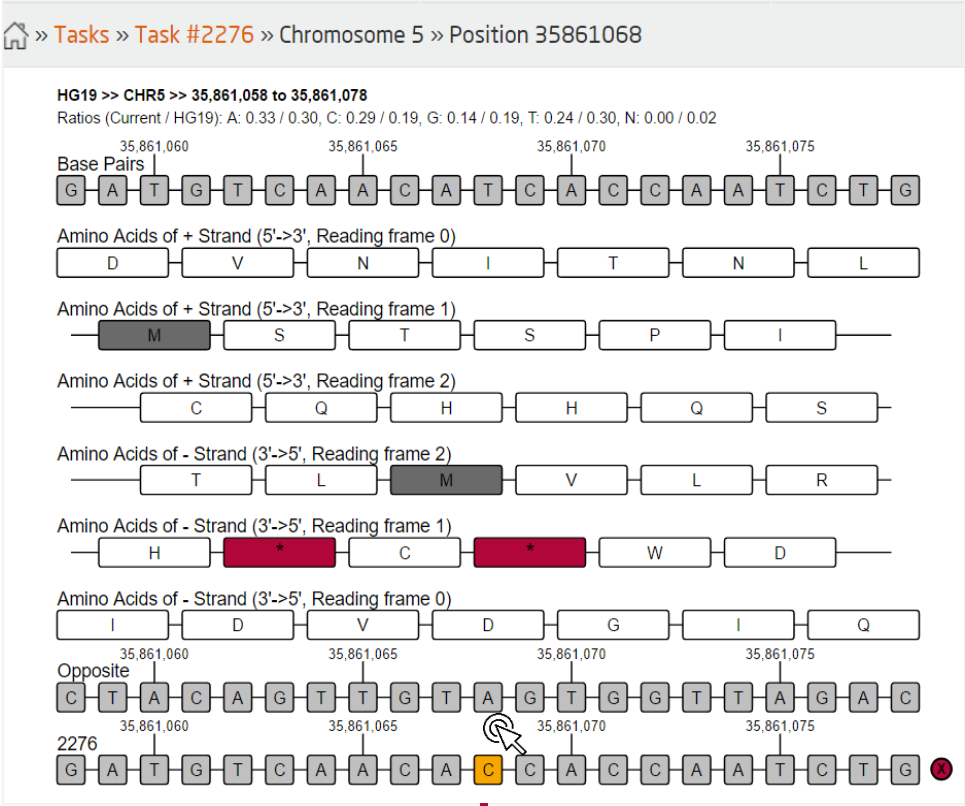
FEEDBACK



Mutations

FEEDBACK

Genome Browser
Assessment of variants



MUTATION DETAILS FOR CHR5 AT 35,861,068

DGV

› Pubmed ID: 15286789
› Short description: Twenty samples were obtained from normal individuals whose DNA are stored at the Coriell Cell Repository (Camden, New Jersey) with the following ethnic representation: 10 European/Caucasian, 4 Native American, 2 Chinese, 2 Indo-Pakistani, and 2 Sub-Saharan African. Twelve samples were from fixed cell pellets of cytogenetically normal individuals from the Brigham and Women's Cytogenetics Laboratory and 7 samples from cytogenetically normal individuals were obtained from the Hospital for Sick Children.

Exercise III

Data Management for Digital Health, Summer 2017

ANALYZE GENOMES
by Hasso-Plattner-Institut

200

TASKS | TRIAL DESIGN | MEDICAL KNOWLEDGE COCKPIT

» Tasks

Task List

Pages: 1 2 3 4 5 6 ... 23 (30 rows per page)

New alignment task

ID	Progress	Pipeline	Results	Cost
2276	Completed in 07h06m47s	BWA_s	1942578	Free
2265	Completed in 05h54m31s	BWA_s	327960	Free
2264	Completed in 02h25m56s	BWA_s	333180	Free
2263	Completed in 05h22m28s	BWA_s	276012	Free
2262	Completed in 05h17m58s	BWA_s	274805	Free
2261	Completed in 14h39m52s	BWA_s	1601151	Free
2260	Completed in 14h59m17s	BWA_s	1507363	Free
2259	Completed in 17h34m42s	BWA_s	1377275	Free
2258	Completed in 16h31m59s	BWA_s	1900591	Free
2257	Completed in 08h21m43s	BWA_s	1647173	Free
2256	Completed in 08h12m04s	BWA_s	1698232	Free
2255	Completed in 20h50m50s	BWA_s	187279	Free
2250	Completed in 10h34m54s	BWA_s	2044480	Free

FEEDBACK

Medical Knowledge Cockpit

Enter search terms | Search

How about A2M, AADAC or PIK3CA?

Search terms

Jane Dough

female, 51 years, non-smoker

Markers

KRAS EGFR BRAF NRAS

Diagnosis

non-small cell lung cancer, stage IV

Charité

FEEDBACK

Medical Knowledge Cockpit

Enter search terms Search

How about A2M, AADAC or PIK3CA?

A2M × AADAC ×

Jane Dough

female, 51 years, non-smoker

Markers

KRAS EGFR BRAF NRAS

Diagnosis

non-small cell lung cancer, stage IV

Charité

Cancer Gene Census

Definition

AADAC

Microsomal arylacetamide deacetylase competes against the activity of cytosolic arylamine N-acetyltransferase,which catalyzes one of the initial biotransformation pathways for arylamine and heterocyclic amine carcinogens(provided by RefSeq, Jul 2008).

A2M

Alpha-2-macroglobulin is a protease inhibitor and

PubMed Links

AADAC

- Human liver arylacetamide deacetylase. Molecular cloning of a novel esterase involved in the metabolic activation of arylamine carcinogens with high sequence similarity to hormone-sensitive lipase.

Clinical Trials

Internal

External

- Long-Term Extension Study in Subjects With Multiple Sclerosis Who Have Completed Study 205MS301 (NCT01064401) to Evaluate the Safety and Efficacy of DAC HYP (EXTEND)

Exercise III

Data Management for Digital Health, Summer 2017

FEEDBACK

Medical Knowledge Cockpit
Context-sensitive definitions for search terms

Enter search terms

How about A2M, AADAC or PIK3CA?

BRAF × lung × cancer ×

Jane Dough

female, 51 years, non-smoker

Markers

KRAS EGFR BRAF NRAS

Diagnosis

non-small cell lung cancer, stage IV

Charité

Definition

BRAF

This gene encodes a protein belonging to the raf/mil family of serine/threonine protein kinases.

AADAC

Microsomal arylacetamide deacetylase competes against the activity of cytosolic arylamine N-acetyltransferase,which catalyzes one of the initial biotransformation pathways for arylamine and heterocyclic amine carcinogens(provided by RefSeq, Jul 2008).

Cancer Gene Census

BRAF

Somatic tumors

melanoma, colorectal, papillary thyroid, borderline ov, Non small-cell lung cancer (NSCLC), cholangiocarcinoma, pilocytic astrocytoma

Exercise III

Data Management for Digital Health, Summer 2017

Medical Knowledge Cockpit
Latest research papers and clinical trials

Hasso Plattner Institut

PubMed Links

BRAF

- Complete coding sequence of a human B-raf cDNA and detection of B-raf protein kinase with isozyme specific antibodies
- B-raf and a B-raf pseudogene are located on 7q in man
- Mutations of the BRAF gene in human cancer
- The ins and outs of Raf kinases
- Autoregulation of the Raf-1 serine/threonine kinase

Clinical Trials

Internal

External

- CEP-32496 in Patients With Advanced Solid Tumors in Phase 1 and Advanced Melanoma and Metastatic Colorectal Cancer in Phase 2
- GSK2141795 and Dabrafenib in Treating Patients With Stage IIIC-IV Cancer
- Selumetinib and Akt Inhibitor MK2206 in Treating Patients With Stage III or Stage IV Melanoma Who Failed Prior Therapy With Vemurafenib or Dabrafenib
- Dabrafenib/Trametinib/Navitoclax In Braf Mutant Melanoma
- Vemurafenib, Cetuximab, and Irinotecan in Advanced Solid Cancers

Collapsible list

PubMed

ClinicalTrials.gov

Exercise III

Data Management for Digital Health, Summer 2017

Cancer Gene Ce

Definition

AADAC

Microsomal aryl
against the activ
acetyltransferas
biotransformatio
heterocyclic ami
2008).

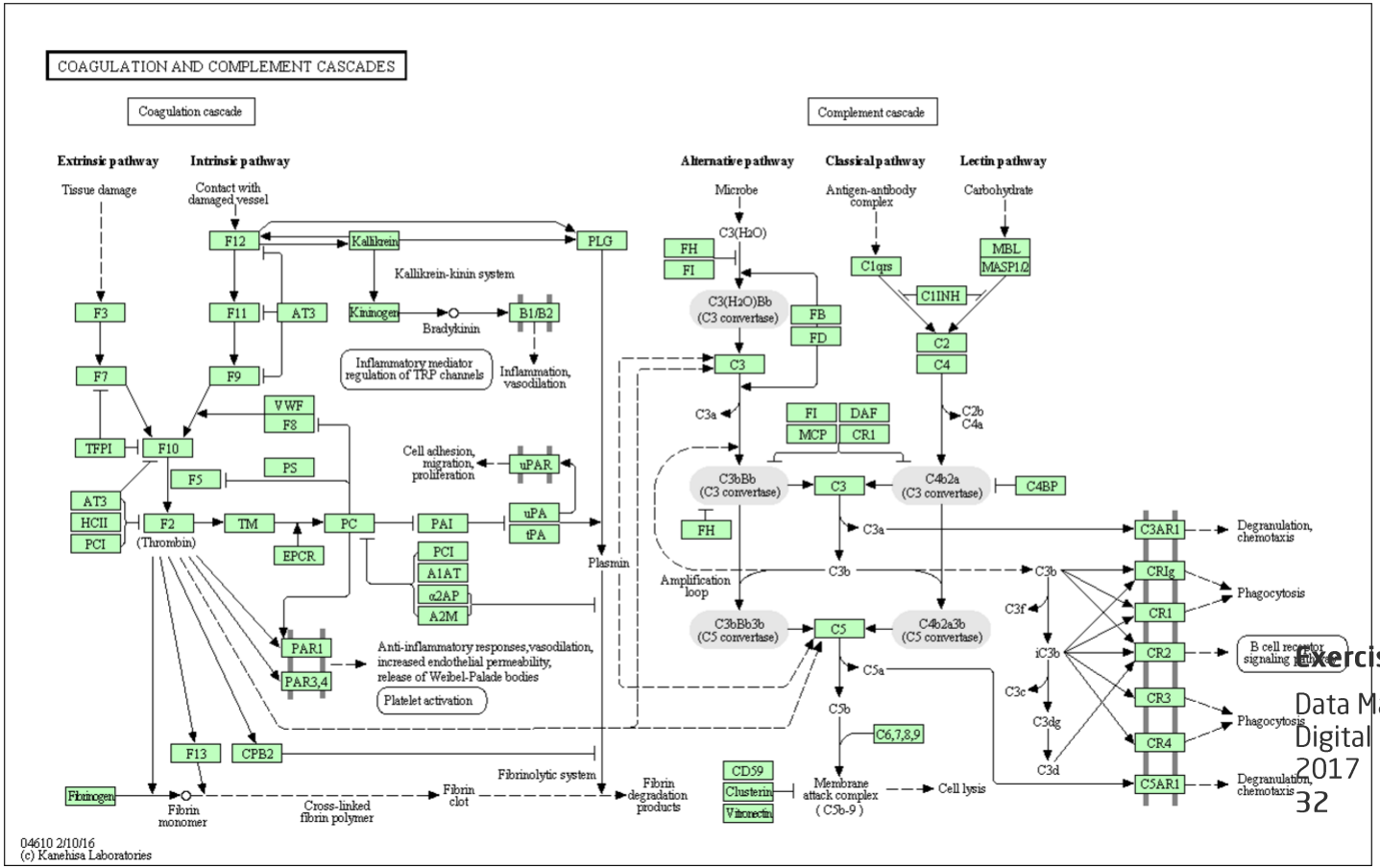
A2M

Alpha-2-macrogl
cytokine transpo

Pathways

Complement an
Homo sapiens (h

Complement and coagulation cascades



Exercise III

Data Management for Digital Health, Summer 2017

FEEDBACK

Hasso Plattner Institut
IT Systems Engineering | Universität Potsdam

Exercise III Part 2
Systems Medicine Approach for Heart Failure

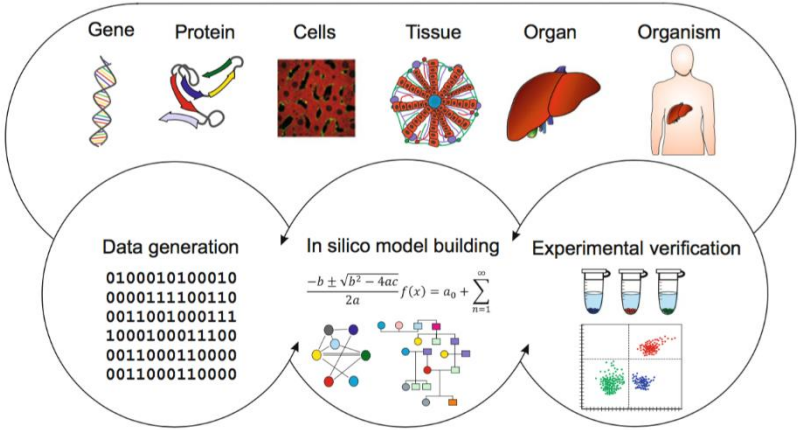
Harry Freitas da Cruz, Milena Kraus

Hasso Plattner Institut

July 18, 2017

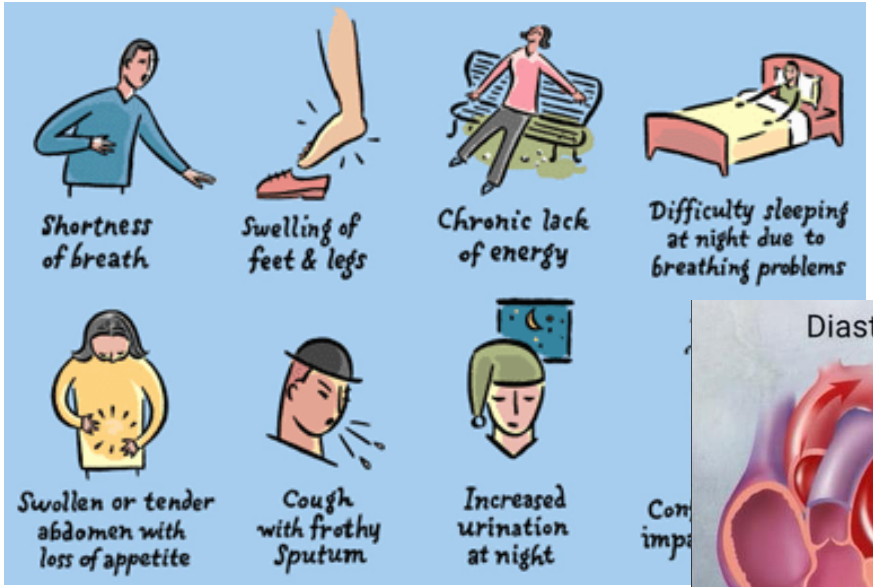
Systems Medicine Approach for Heart Failure

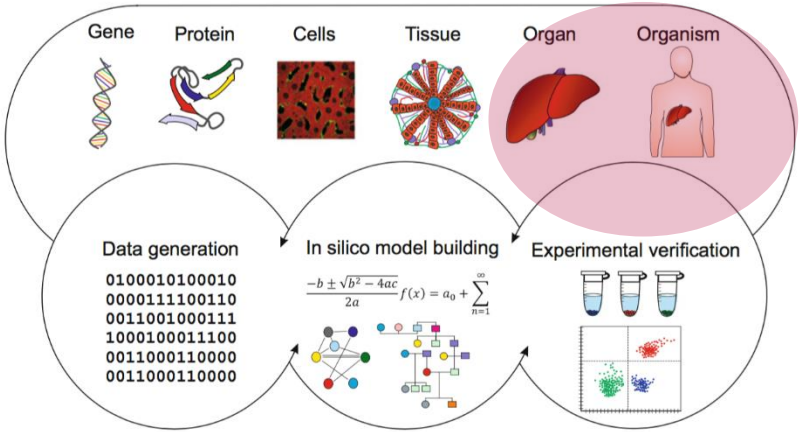
- Systems medicine concept
- Biomedical and clinical facts regarding heart failure
- RNA sequencing and differential gene expression analysis
- Clustering of mixed-type data



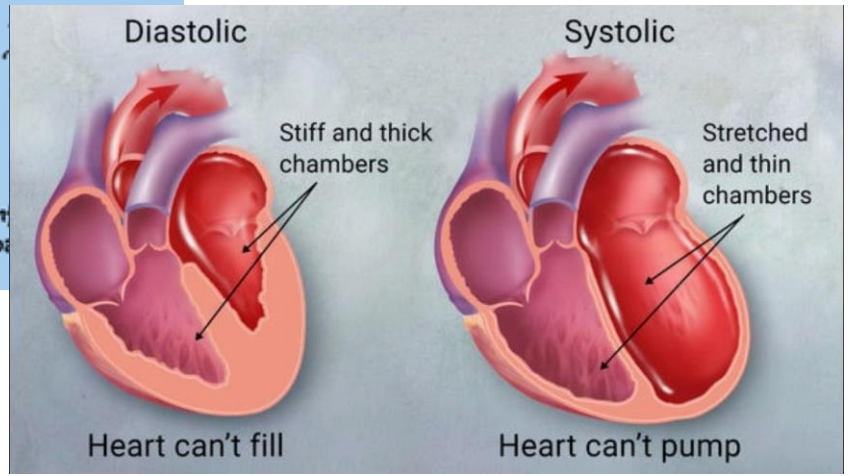
Schmitz, U. and Wolkenhauer, O. "Systems medicine methods and protocols: methods in molecular biology." (2016).

Biomedical and Clinical Facts
Heart Failure



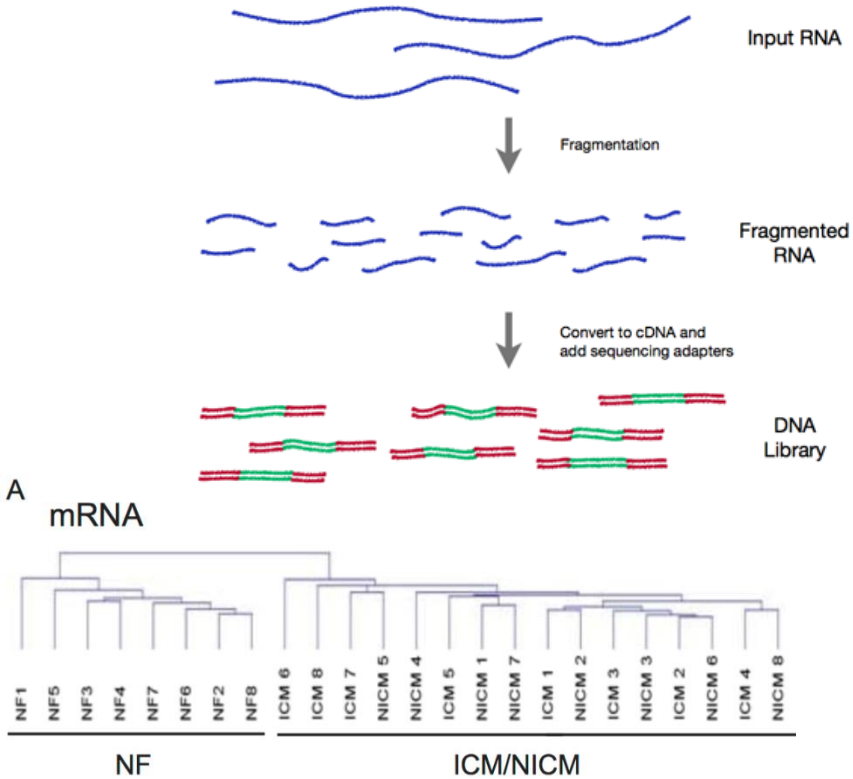


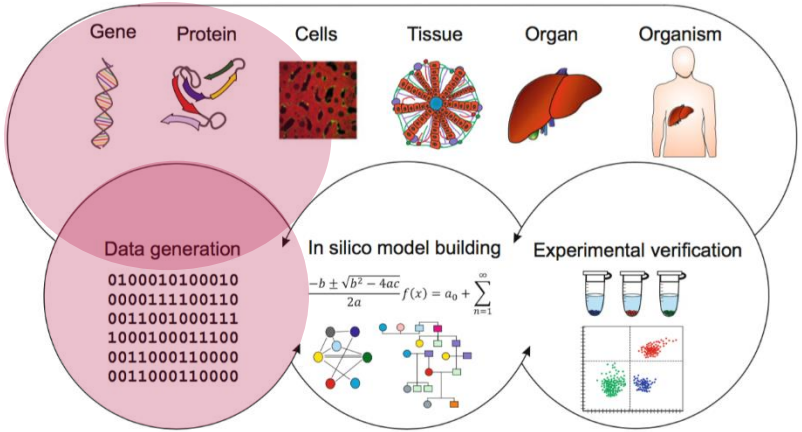
Schmitz, U. and Wolkenhauer, O. "Systems medicine methods and protocols: methods in molecular biology." (2016).



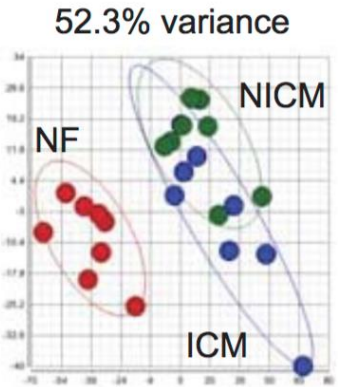
https://www.youtube.com/watch?v=B93TsbJXnMc

RNA Sequencing and Differential Gene Expression Analysis

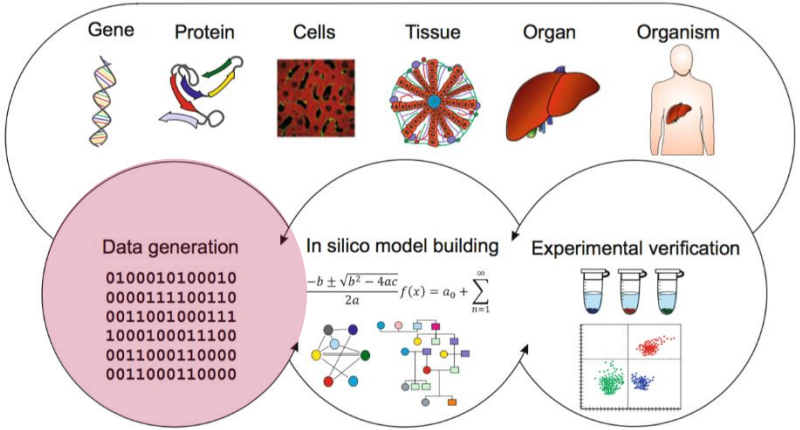




Schmitz, U. and Wolkenhauer, O. "Systems medicine methods and protocols: methods in molecular biology." (2016).



Mixed-type Data



Schmitz, U. and Wolkenhauer, O. "Systems medicine methods and protocols: methods in molecular biology." (2016).

Samples<<Attributes

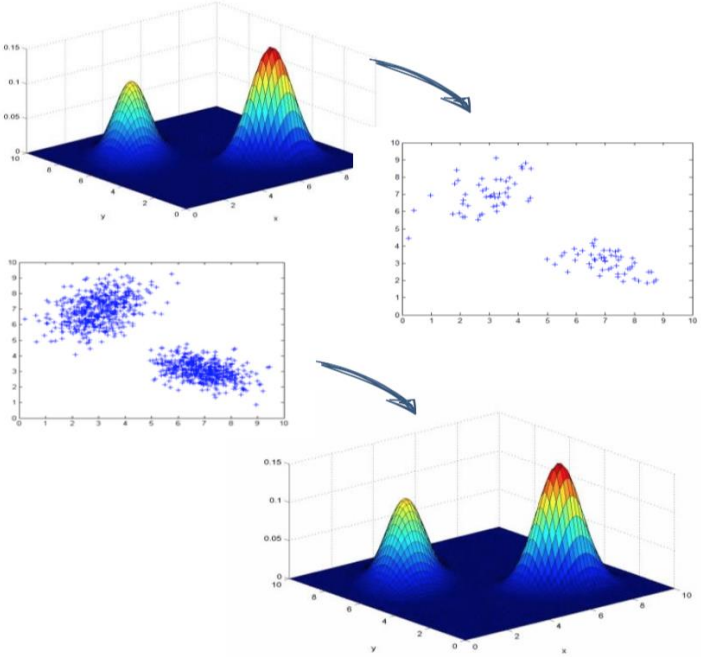
	80338964	9298042	948588	9533780	9724625	age	drug	gender	grading	m	n	primary	smoker	stage
1	0	0	0	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA
2	0	0	0	1	0	NA	NA	NA	NA	NA	NA	NA	NA	NA
3	0	1	0	0	0	NA	NA	NA	NA	NA	NA	NA	NA	NA
4	0	0	0	0	0	65	4.0	0	2	0	2	1	1	4
5	1	0	0	0	0	63	2.7	0	2	0	0	1	0	2
6	0	0	0	0	0	59	1.9	1	2	0	2	1	1	5
7	0	0	0	0	0	46	1.3	0	3	0	2	1	1	4
8	0	0	0	0	0	68	1.1	1	2	0	0	1	0	4
9	0	0	1	0	0	71	6.4	0	2	0	2	1	1	4
10	0	0	0	0	0	56	8.3	0	2	0	2	1	1	5
11	0	0	1	0	0	71	3.5	1	2	0	1	0	0	3
12	0	0	0	0	0	67	1.4	0	2	0	2	1	1	4
13	0	0	0	0	0	59	NA	1	2	0	2	1	1	5
14	0	0	0	0	0	82	0.8	0	0	0	2	1	1	4
15	0	0	0	0	0	71	6.0	0	2	0	2	1	1	4
16	0	0	0	0	0	56	NA	0	2	0	2	1	1	5
17	0	0	0	0	0	75	NA	0	2	0	0	1	1	2
18	0	0	0	0	0	57	2.6	0	2	0	0	1	1	1
19	0	0	0	0	0	55	NA	0	3	0	2	1	0	4
20	0	0	1	0	0	65	NA	0	2	0	0	1	1	2
21	0	0	0	0	0	54	1.8	0	2	0	0	1	1	2
22	0	0	0	0	0	60	7.0	0	3	0	1	1	1	3
23	0	0	0	0	0	50	NA	0	2	0	2	1	1	4

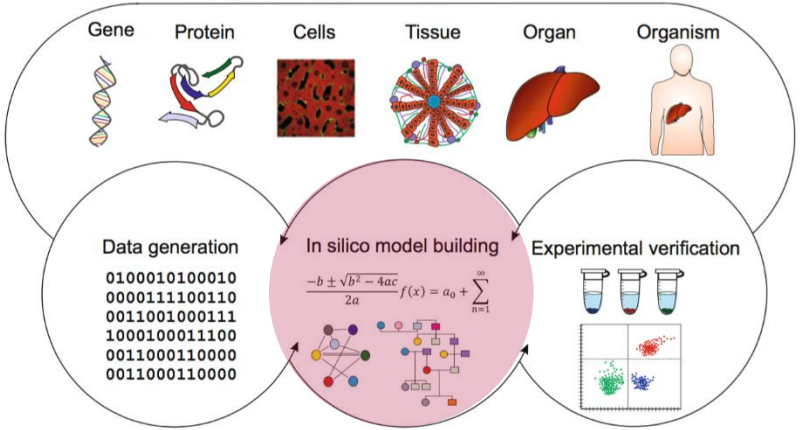
Incomplete

Nominal + Numerical

Sparse

Clustering of Mixed-type Data





Schmitz, U. and Wolkenhauer, O. "Systems medicine methods and protocols: methods in molecular biology." (2016).

Original index	-2	-1	0	1	2	= i
Converted to rank	1	2	3	4	5	= r R = max (r) = 5
Normalized rank	0	$1/4$	$1/2$	$3/4$	1	$x = \frac{r-1}{R-1}$