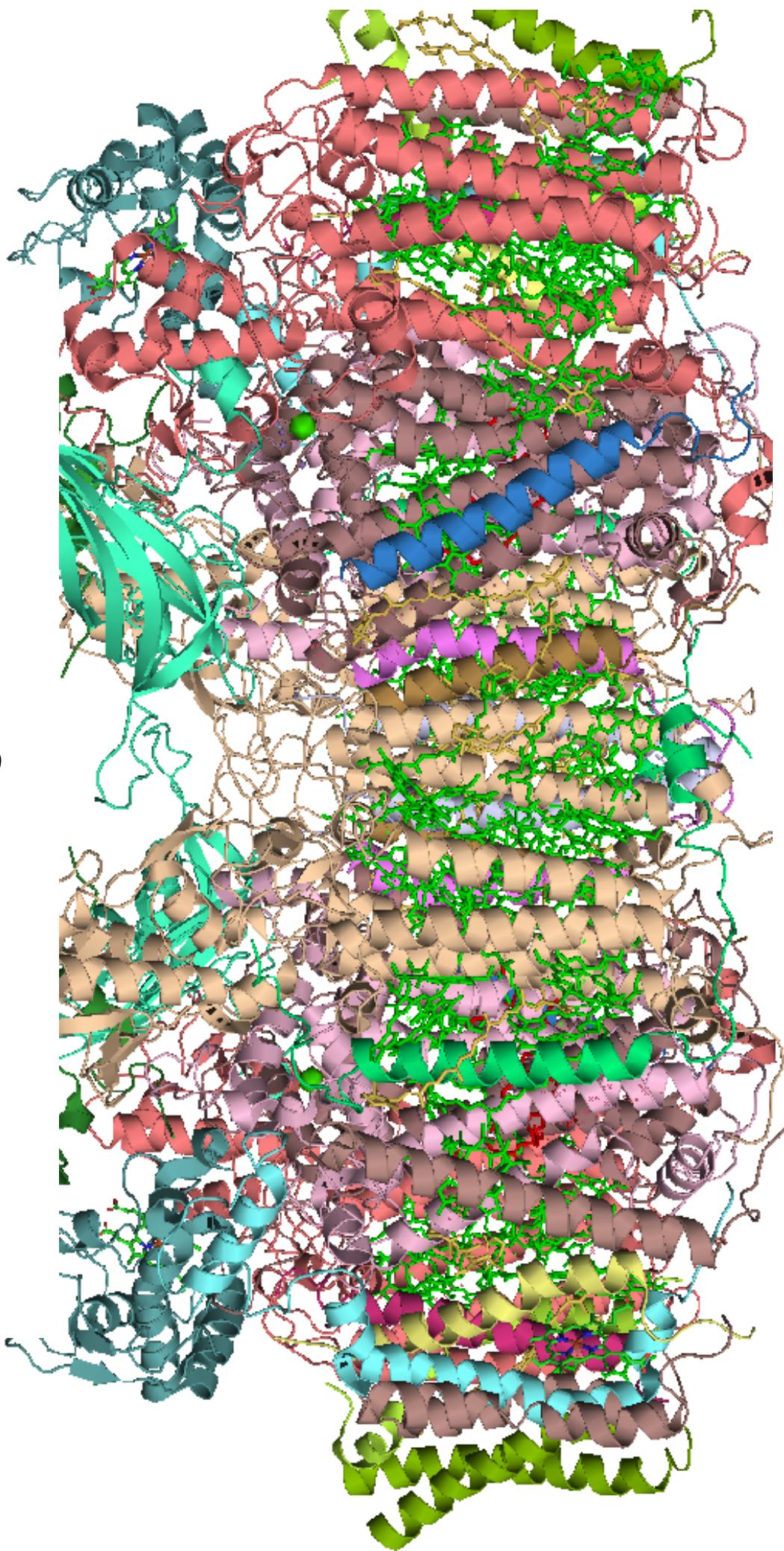




Waag
wetlab
amsterdam



Bio Informatics



Peer to Peer Feedback on the Forum

Label issues and pull requests for new contributors

[Dismiss](#)

Now, GitHub will help potential first-time contributors discover issues labeled with **help wanted** or **good first issue**

Filters ▾ is:issue is:open

Labels 8 Milestones 0

New issue

① 30 Open ✓ 2 Closed

Author ▾ Projects ▾ Labels ▾ Milestones ▾ Assignee ▾ Sort ▾

④ Midterm - Amy Traylor - Bioshield

#32 opened a day ago by trembl

④ Midterm - Rowan Roberts - tear gas / pepper spray

#31 opened a day ago by trembl

④ Midterm - Sophia Montaya - Persephone

#30 opened a day ago by trembl

④ Midterm - June Lee - Microbe Portraiture

#29 opened a day ago by trembl

④ Midterm - Inger Le Gue - Resistance is fertile

#28 opened a day ago by trembl

④ Midterm - Candyce Dryburgh - Cultivated Heritage

#27 opened a day ago by trembl



The information / “Omnics” age

- “Genomics” DNA sequence analysis
- “Transcriptomics” DNA expression analysis
- “Proteomics” Protein (structure) prediction / analysis
- “Interactomics” Protein – Protein, DNA – Protein interaction
- “Metabolomics” Metabolism modeling



What is it used for

- Optimizing yield
- Predicting organisms behaviour
- Medical diagnostics
- Personal medicine
- Drug discovery





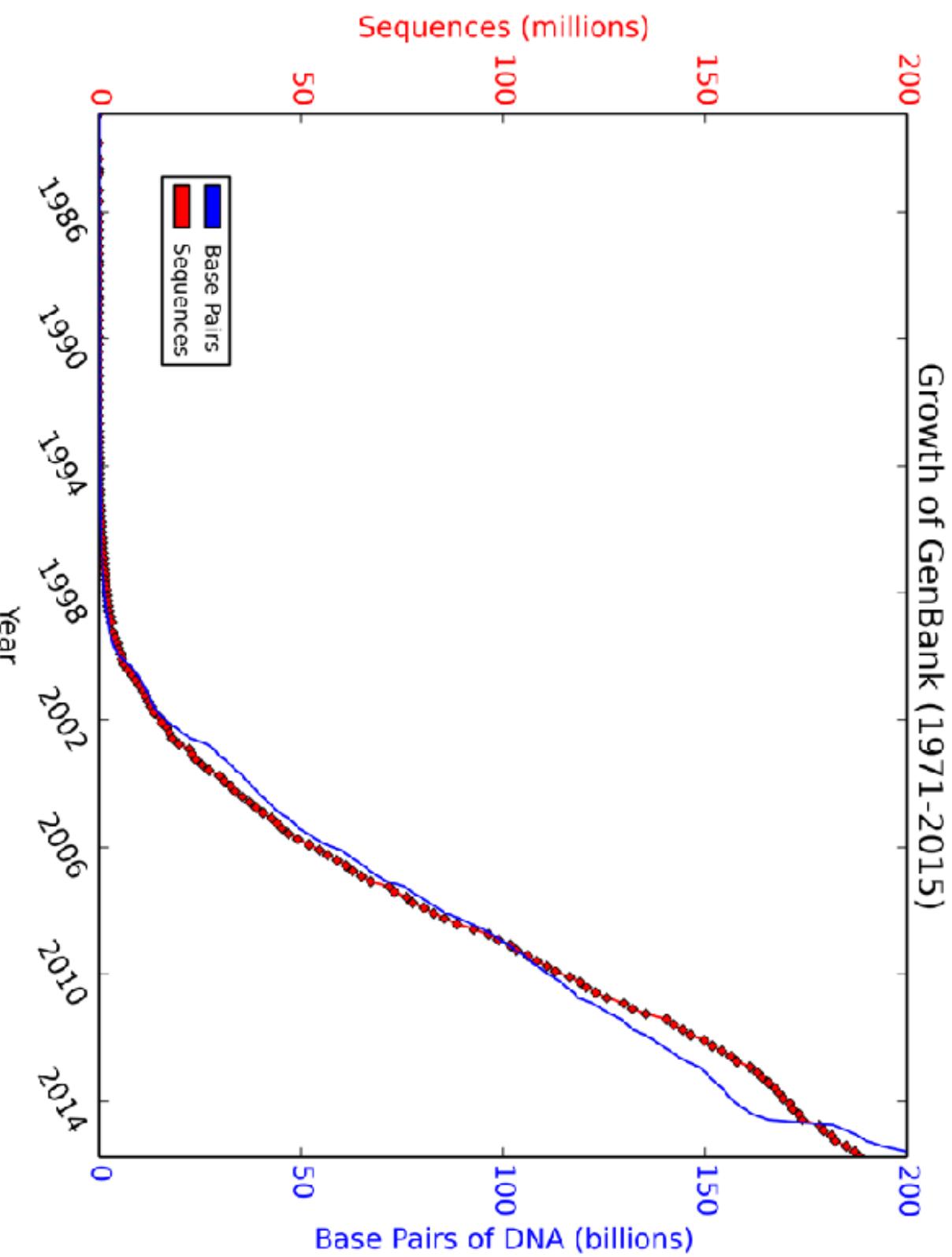
Genomics

- Functional genomics
- Metagenomics
- Personal Genomics
- Epigenomics





DNA database GenBank





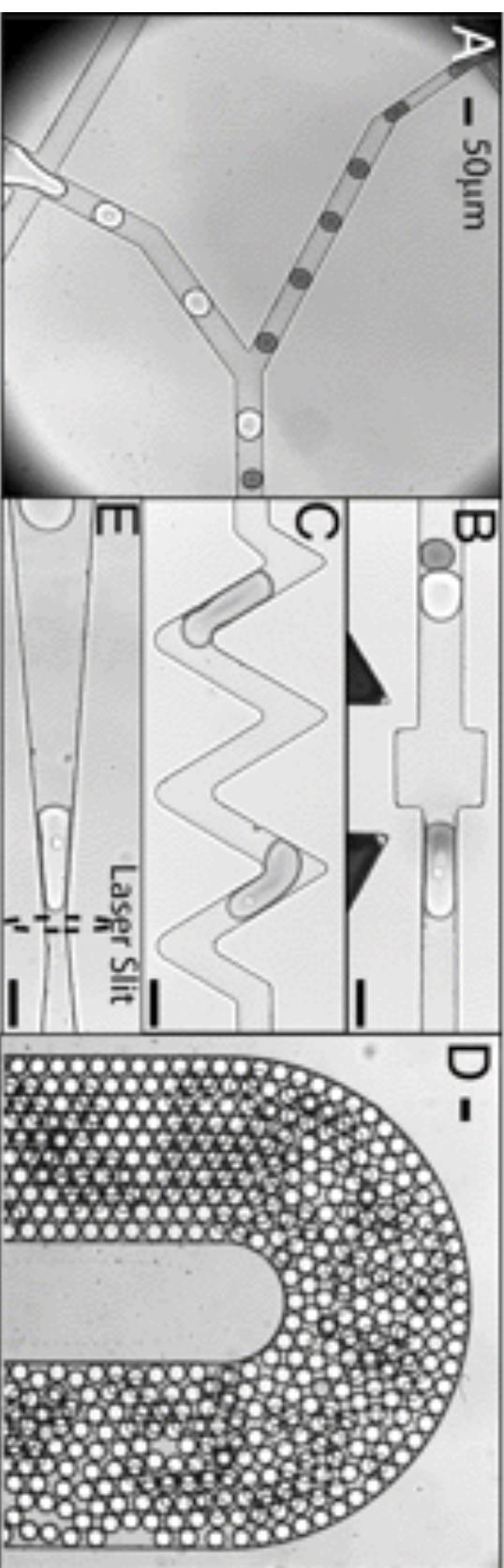
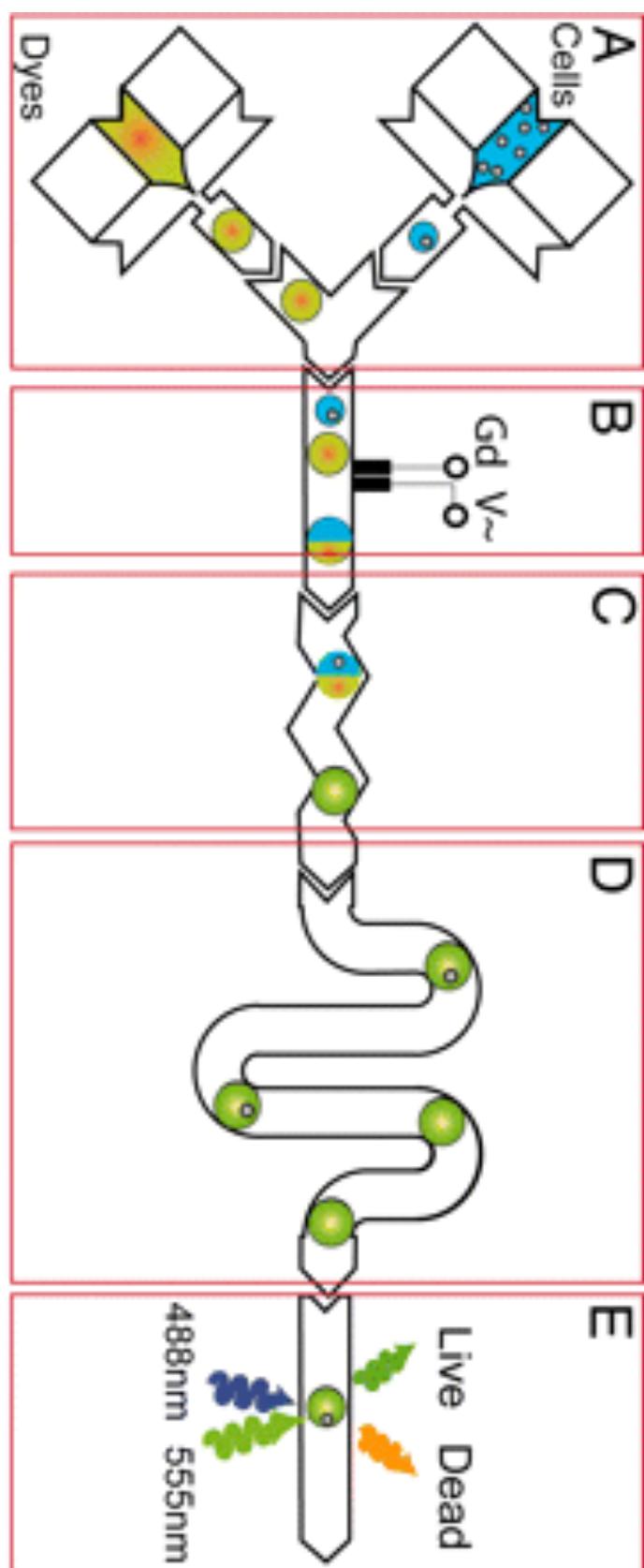
Drivers

- “High Throughput Research”
- Robotics
- Databases
- Visualisation
- Public tools
- Open data



High Throughput Screening

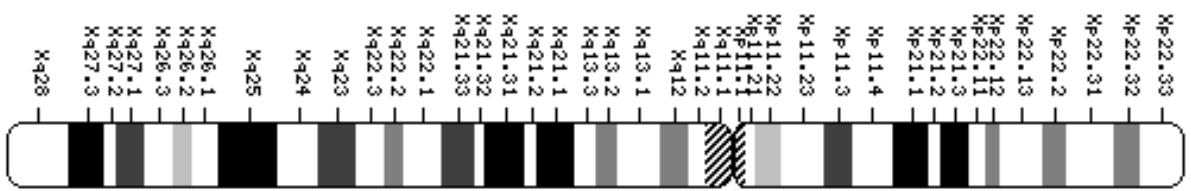
Eric Brouzes et al. PNAS 2009;106:14195-14200





Gene annotations

Ideogram





Sequence Alignment

14	SIKLWPPSQTTRLLVERMANNLST	PSIFTRK	YGSLSKEEARENAKQIEEVACSTANQ	RHEKEPDGGGSAYQLYAKECSKLILEVLK	101
13	SIKLWPPSESTRIMLVDRMTNNLST	ESIFSRK	YRLLGKQEAHENAKTIEELCFALADE	HFRREEPDGGGSAYQLYAKEETSKMLLEVLK	10C
23	VFKLWPPSQGTREARQKMALKLSS	ACFESQS	FARIELADAQEHRARAEIEEVAFGAQE	ADSGGDGTGSAVMVYAKHASKLMLETLR	10C
13	SVKLWPPGOSTRMLVERMNTKNFIT	PSFISRK	YGLLSKEEEDAKKIEEVAFAAHQ	HYEKQPDGGSSAVQIYAKESSRMLLEVLK	10C
30	SFSIWIWPPTQRTRDAVWRRLVDTLGG	DTILCKR	YGAVPAADEPAARGIEAEFDAAA	SGEAAATASVEEGIKALQDLYSKERVRLLDFVK	12C
44	SLSIWIWPPSQRTRDAVWRRLVDTLVA	PSILSQR	YGAVPAAEAGRAAAAVEAEAYAVTES	SSAAAPASVEDGIEVLQYSKEYSRRLLDFVK	13C
56	SFSIWIWPPTQRTRDAITSLRIETLST	TSVLSKR	YGTIPKEEASEASRRIEEAFSGAST	VASSEKDGLEVLQDLYSKERISKRMLETVK	141
29	SFAWWPPTRRTRDAVWRRLVAVLSGDTTALKRYRYGAVPAADEAERAARAYEADQAFDAASA		SSSSSSVEDGIETLQDLYSREVSNRLLA	YFVR	121
13	SIKLWPPSESTRMLVERMNTDNLSS	VSFFSRK	YGLLSKEEAAENAKRIEETFLAAND	HEAKEPNLDDSSVVQFYAREASKLMLEALK	10C
57	SRLIWPPPTQKTRDAVLNRLIETLST	ESILSKR	YGTLSKSDATTYAKLIEEAYGVASH	AVSSDDDGIKILELYSKERISKRMLESVK	14C
25	NYSIWIWPPKQRTDAVKNRLIETLST	PSVLTKR	YGTMSADEASAAIAQIEDEAFSYNA	SSSTSNDNVNTILEVYYSKEEISKRMIETYVK	11C
28	SFKIWIWPPTQRTREAVVRLVETLTS	QSYLSKR	YGVIPPEEDATSAAARIIEEEAFSYASY	ASAASTGRPEDIIEWIYLHYSQEIXQRVVESAK	11C
25	SFSIWIWPPTQRTRDAVINRLIESLST	PSILSKR	YGTLPQDEASETARLIEEEAFAAAGS	TASDADDGIEILQDVYSKEEISKRMIETYVK	11C
14	SVKMWPPPSKSTRMLVERMTKNITT	PSIFSRK	YGLLSVVEAEQQDAKRIEDLAFATANK	HQNEDGGGTSAYHYAKESSSKLMLDVTK	101
13	SIKLWPPPSLPTRKALIERITNNFSS	KTIFTEK	YGSLTQDKQATEAKRIEDIAFSTANG	QFEREPDGGSAYQLYAKECSKLILEVLK	10C
48	SLSIWIWPPTQRTRDAVITRLIETLSS	PSVLSKR	YGTISHDEAESAAARRIEDEAFGWANT	ATSAEDDGLLEILQDLYSKERISKRMLETVK	13C



BLAST: Basic Local Alignment Search Tool

<http://blast.ncbi.nlm.nih.gov/Blast.cgi>

Basic Local Alignment Search Tool

NCBI BLAST! blastn

Home Recent Results Saved Strategies Help

Standard Nucleotide BLAST

BLASTN programs search nucleotide databases using a nucleotide query. more...

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [?](#)

Align two or more sequences [?](#)

Or, upload file [Choose File](#) no file selected [\(i\)](#)

Job Title

Enter a descriptive title for your BLAST search [?](#)

Align two or more sequences [?](#)

Choose Search Set

Database

Human genomic + transcript Mouse genomic + transcript Others (nr etc.): [Human](#) [Create custom database](#)

Organism

Optional
Enter organism name or ID—compliant.org will be suggested Exclude [?](#)

Exclude

Optional
 Models (XML/XP) Uncultured/environmental sample sequences

Limit to

Optional
 Sequences from type material [You Tube](#) [Create custom database](#)

Optional
Enter an Entrez query to limit search [?](#)

Program Selection

Optimize for

Highly similar sequences (megablast)
 More dissimilar sequences (discontiguous megablast)
 Somewhat similar sequences (blastn)

Choose a BLAST algorithm [?](#)

BLAST

Search database Nucleotide collection (nr/nt) using Megablast (Optimize for highly similar sequences)
 Show results in a new window



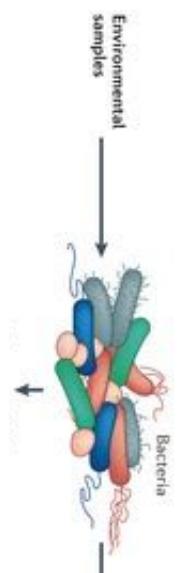
Scoring Matrix BLOSUM

(BLOCKS Substitution Matrix)

Ala	4
Arg	-1
Asn	-2
Asp	-2
Cys	0
Gln	-1
Glu	-1
Gly	0
His	-2
Ile	-1
Leu	-1
Lys	-1
Met	-1
Phe	-2
Pro	-1
Ser	1
Thr	0
Trp	-3
Tyr	-2
Val	0
Ala	4
Arg	5
Asn	6
Asp	6
Cys	9
Gln	5
Glu	5
Gly	6
His	8
Ile	4
Leu	2
Lys	4
Met	5
Phe	4
Pro	6
Ser	7
Thr	5
Trp	11
Tyr	7
Val	4

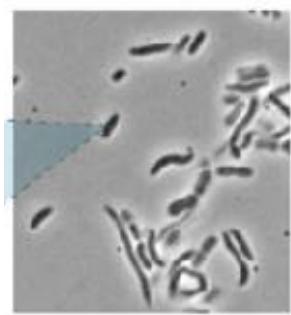


Environmental DNA analysis





16S RNA

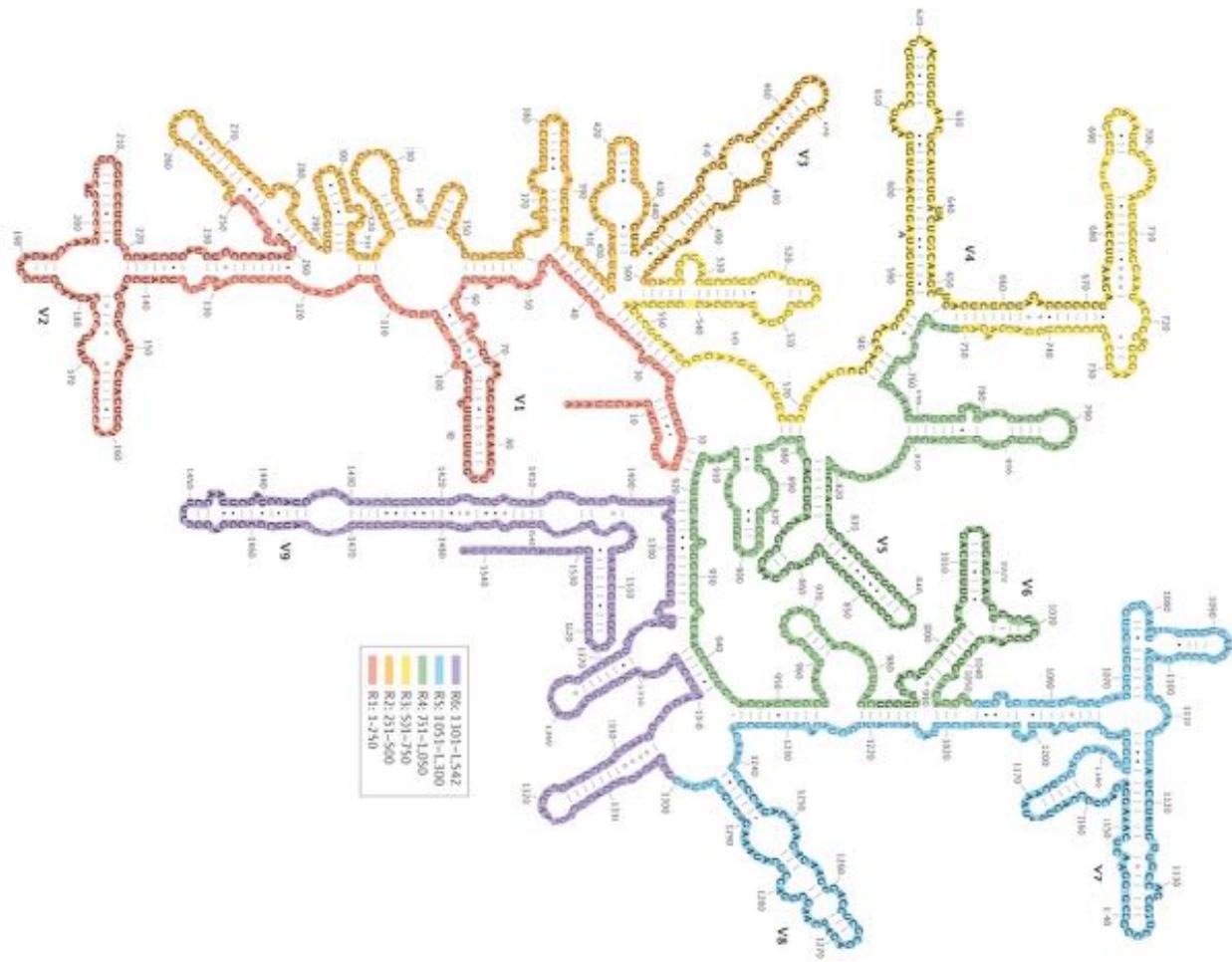


Environmental samples

—
DNA extraction



16S RNA molecule





Challenges

- Cross linking data / Data mining
- Relate Genomics to Transcriptomics, Proteomics
- Relate to structure
- Relate to disease



Bio Informatics for the public



23andMe

welcome

health ancestry

how it works research buy help

sign in register kit 0

welcome
to you®

DNA Collection Kit



Get to know you.
Health and ancestry
start here.

- View reports on over 100 health conditions and traits
- Find out about your inherited risk factors and how you might respond to certain medications
- Discover your lineage and find DNA relatives

order now

€169

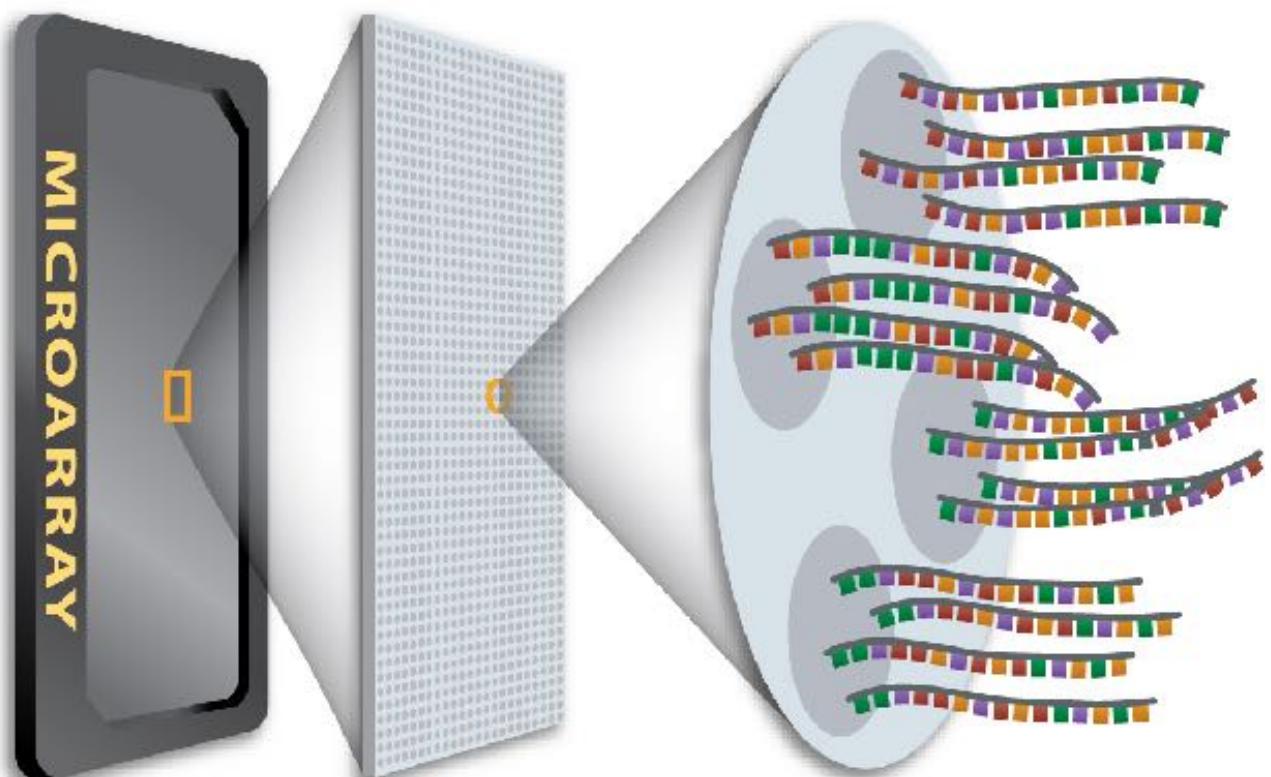
shipping included

What your DNA says about you.

Find out how your genetics relate to things like abnormal blood clotting, cystic fibrosis or response to certain medications. You can also see if your body metabolizes caffeine quickly or if you're likely lactose intolerant. We believe the more you know about



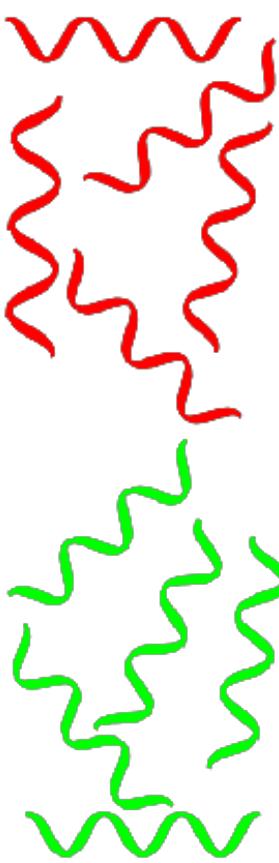
DNA Microarray





DNA Microarray Expression Analysis

DNA extraction and digestion



Test = Tumoral DNA
labeled with **Cy5**

Reference = Normal DNA
labeled with **Cy3**

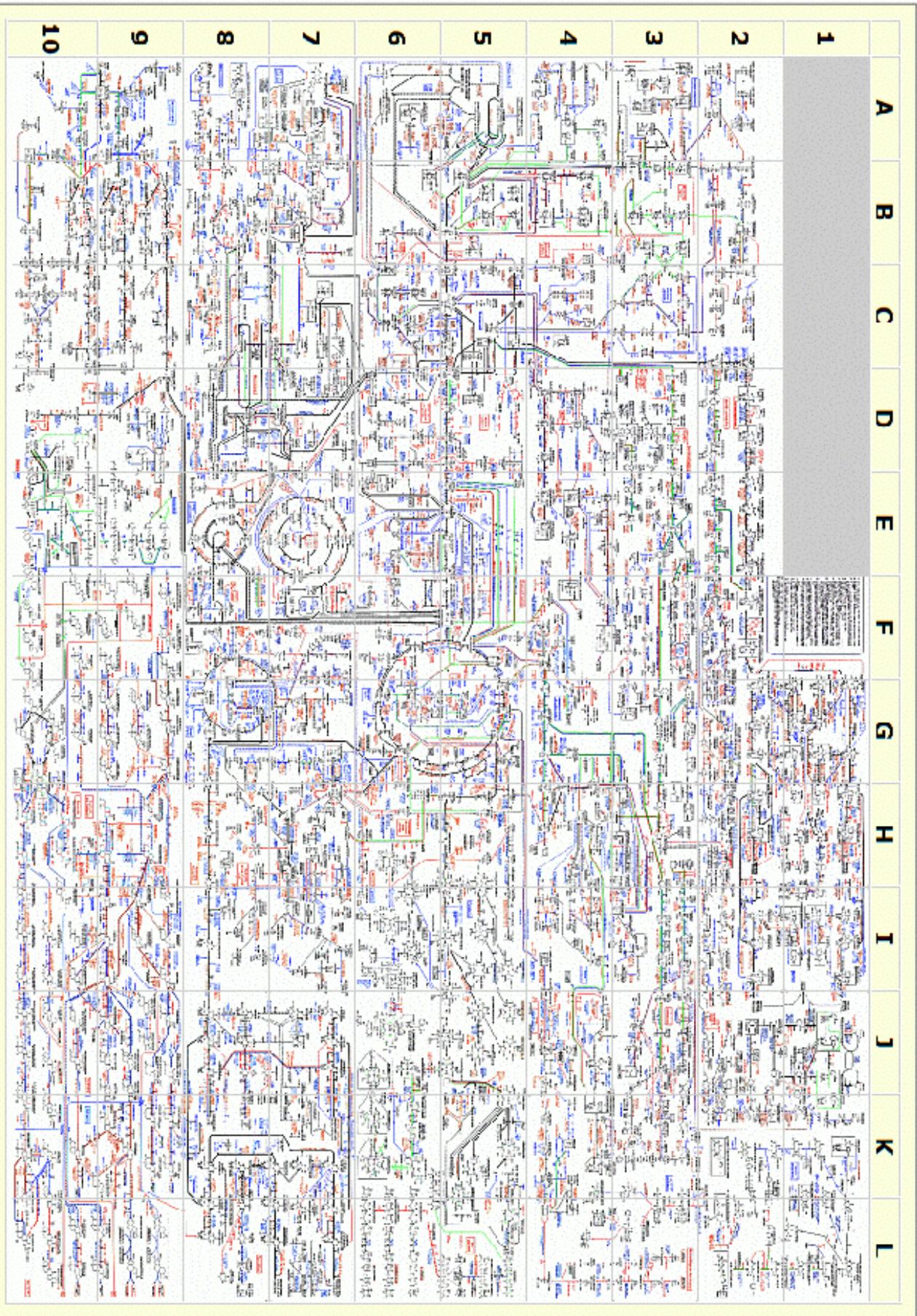


waag
wetlab amsterdam

Proteins



Biochemical Pathways of the Cell





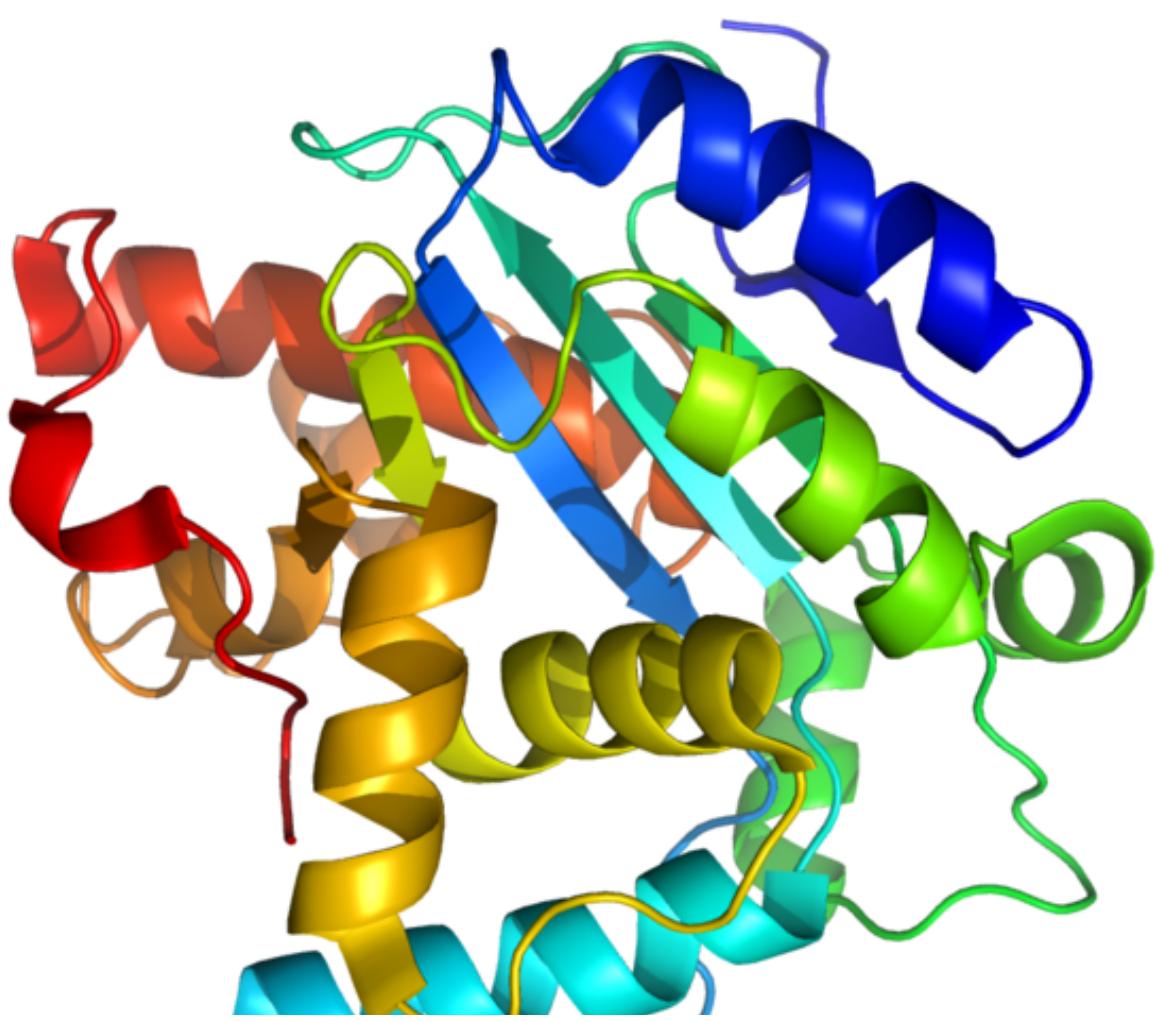
Protein interaction mapping: MS



What is this?



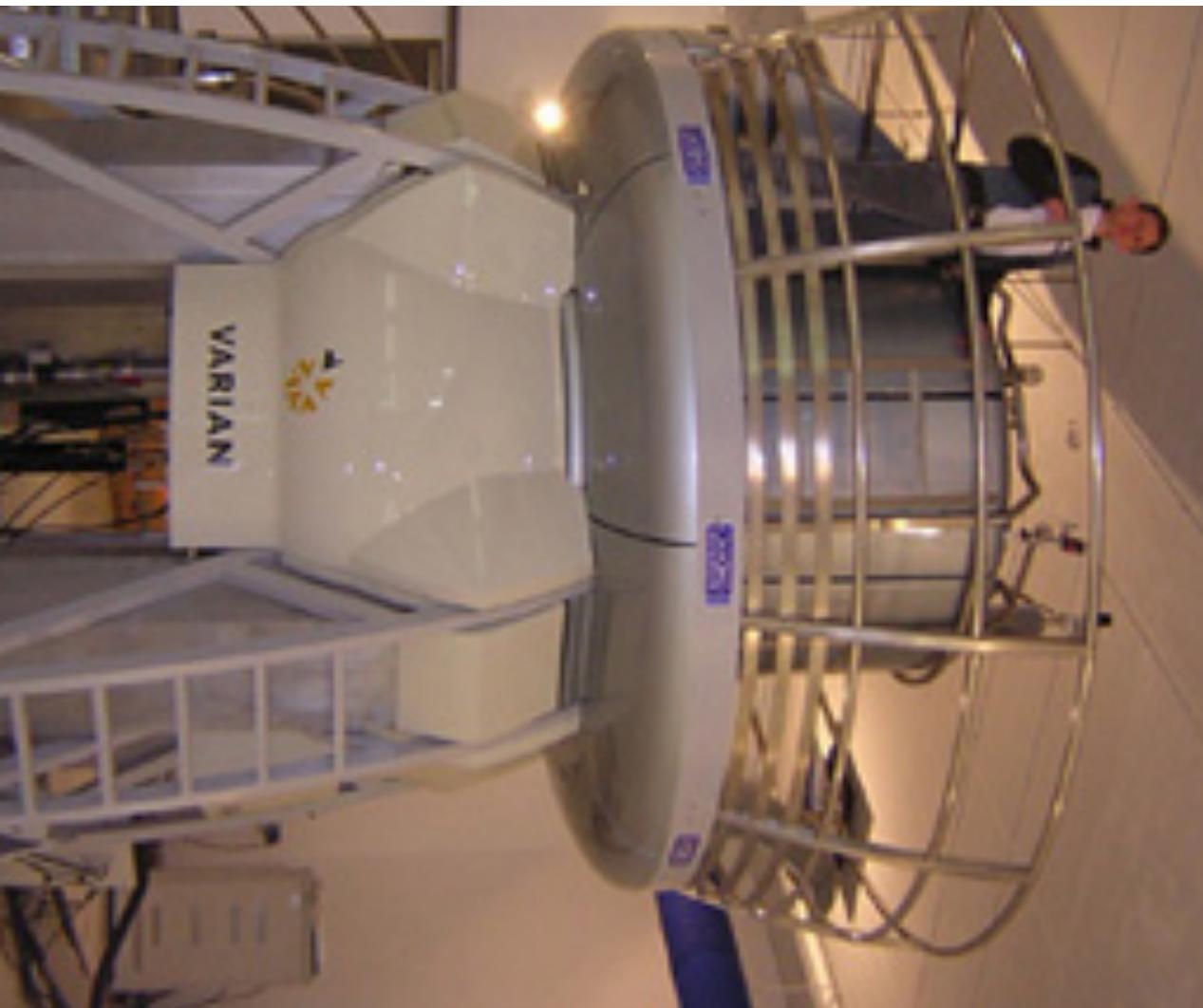
Simon Eugstar - CC-BY-SA 3.0



Debstart - CC-BY-SA 3.0

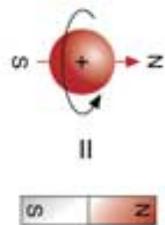
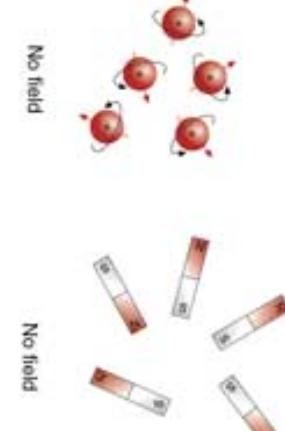
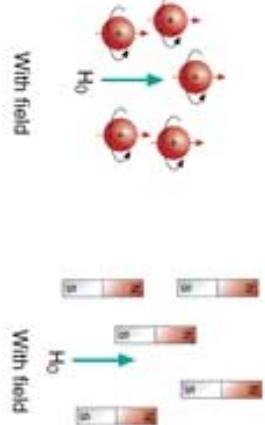
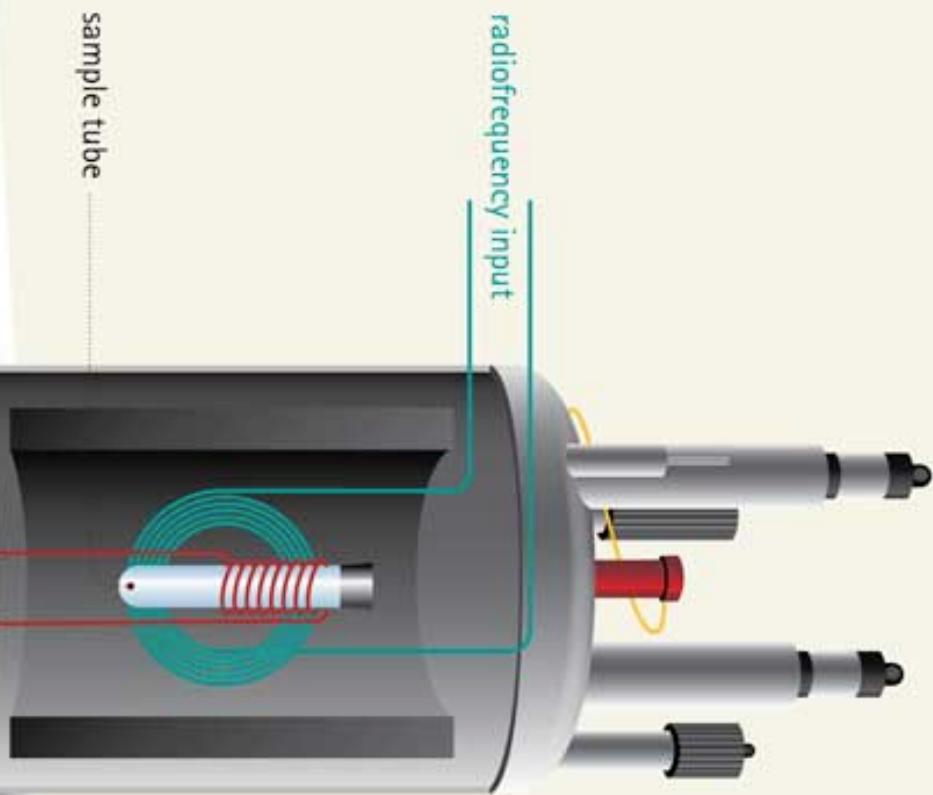


NMR Machines





NMR principles

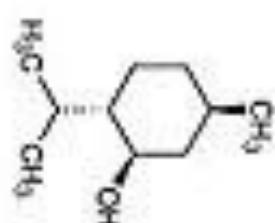
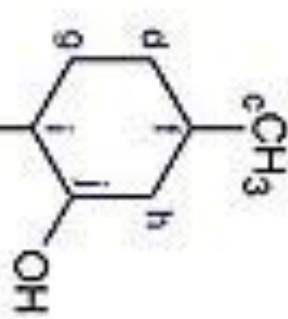




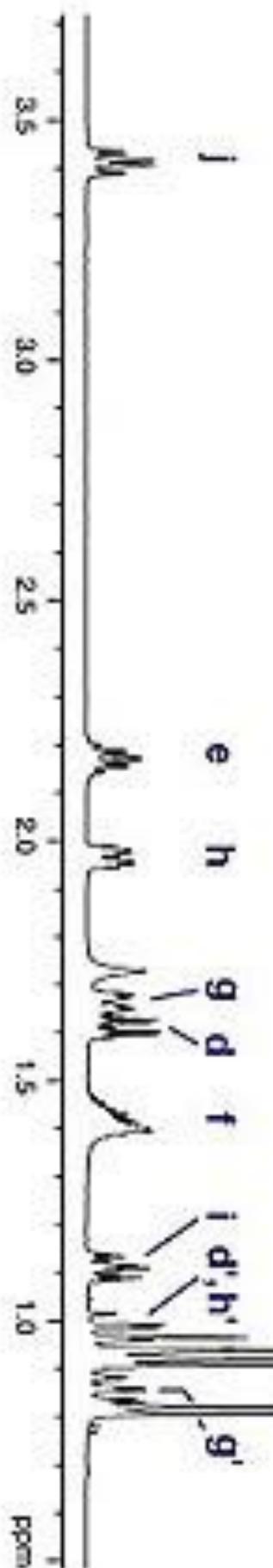
NMR Spectrum

1D PROTON SPECTRUM

b
c
a



(\pm)-Menthol

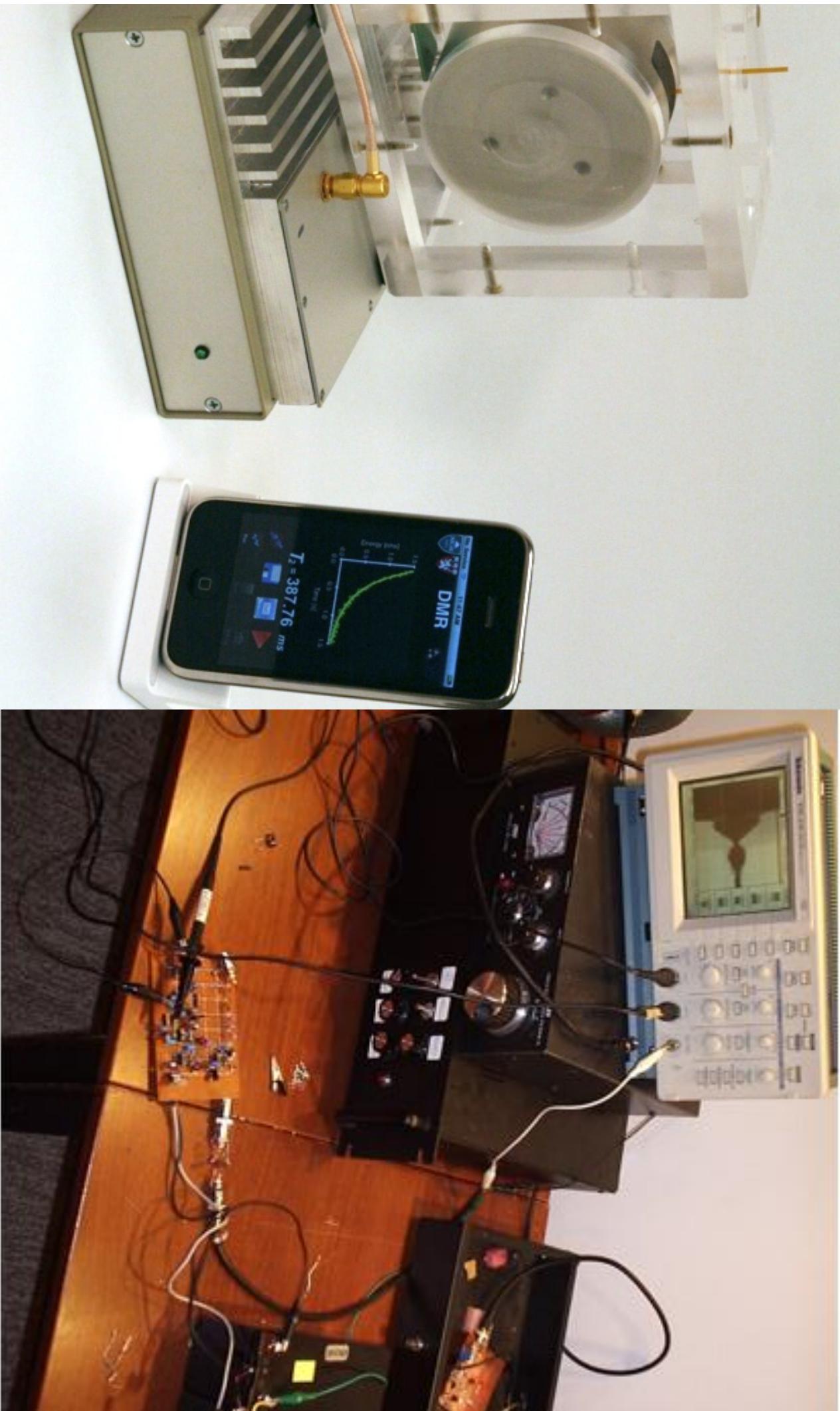




DIY NMR?

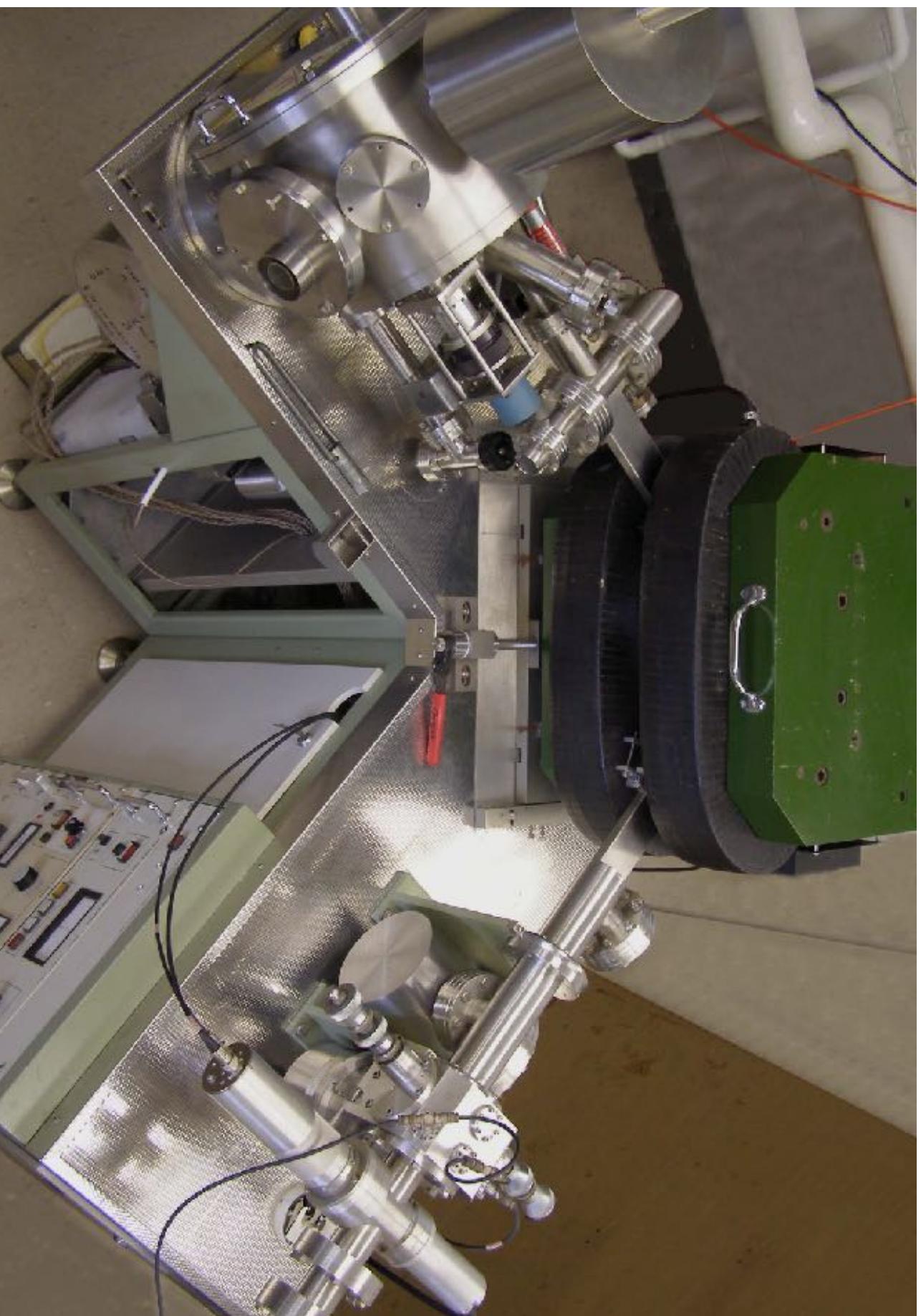
<http://www.seas.harvard.edu>

conspiracyofflight.com





Mass Spectrometer





Simple Mass Spectrometry

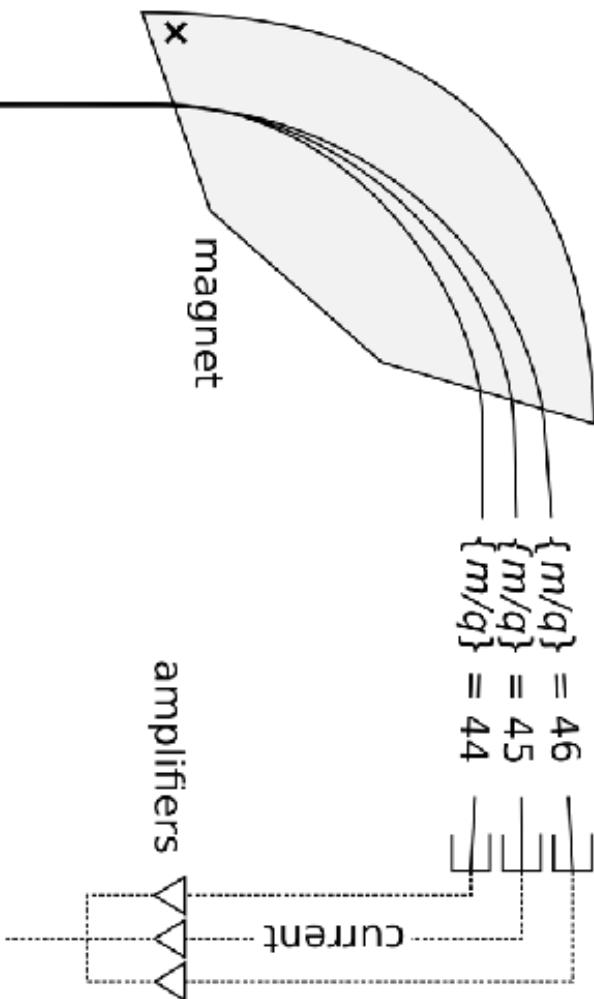
Detection

Faraday
collectors

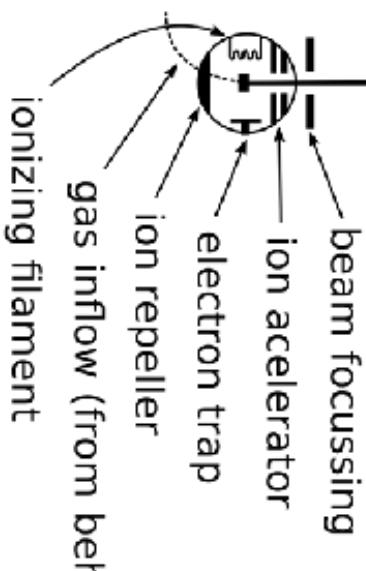
$\{m/q\} = 46$
 $\{m/q\} = 45$
 $\{m/q\} = 44$

High mass = high m/q

Big charge = low m/q



Ion source



legend:
 m ... ion mass
 q ... ion charge



Mass Spectrometry



cells or tissue

MALDI

Tandem

TOF



Procedure



—

Sampling the colony

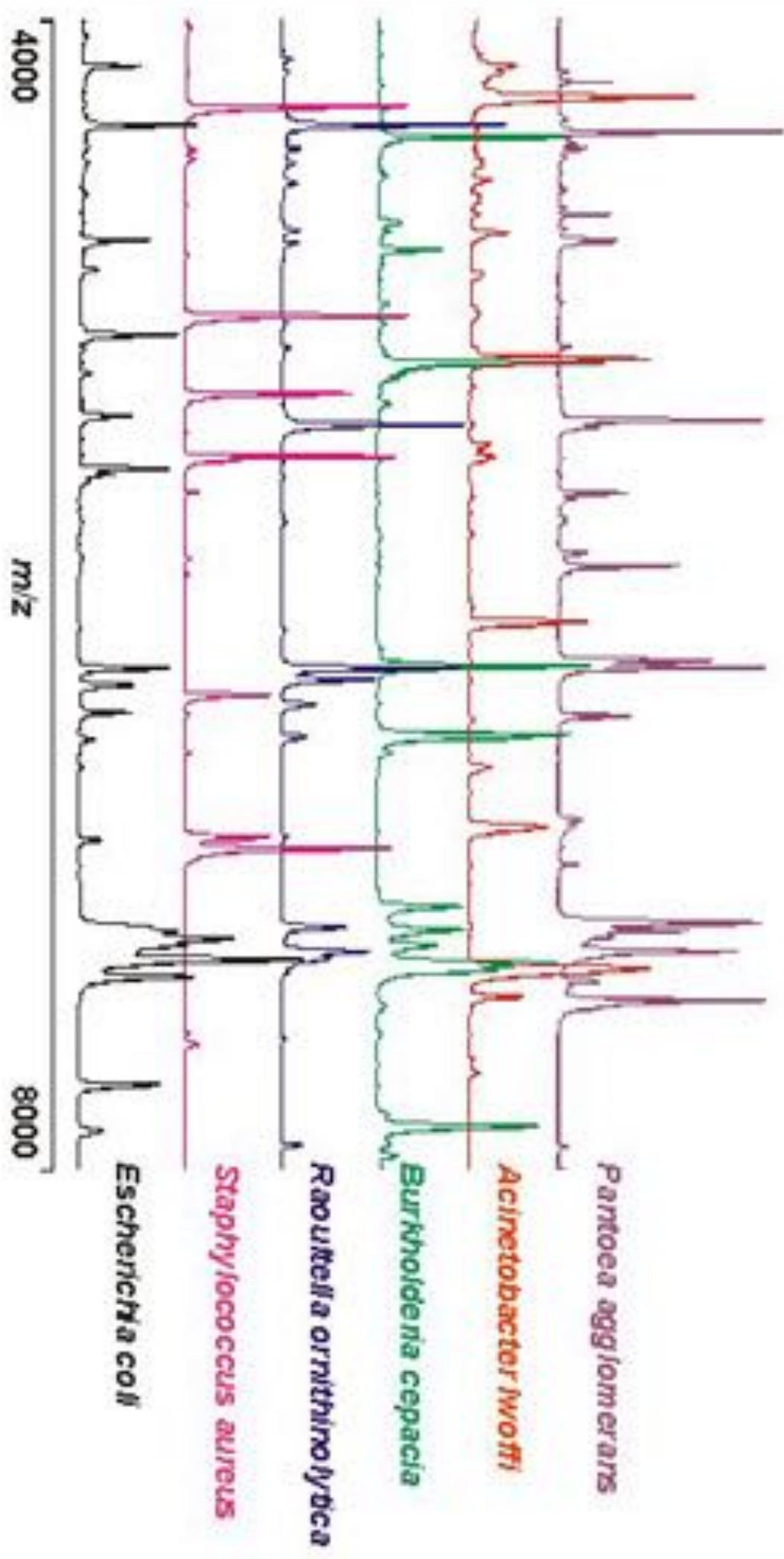
± Extraction of intracellular protein in 70% formic acid and absolute ethanol

Spectrum and interpretation are obtained in less than 2 minutes per deposit

Introduction of the target into the apparatus

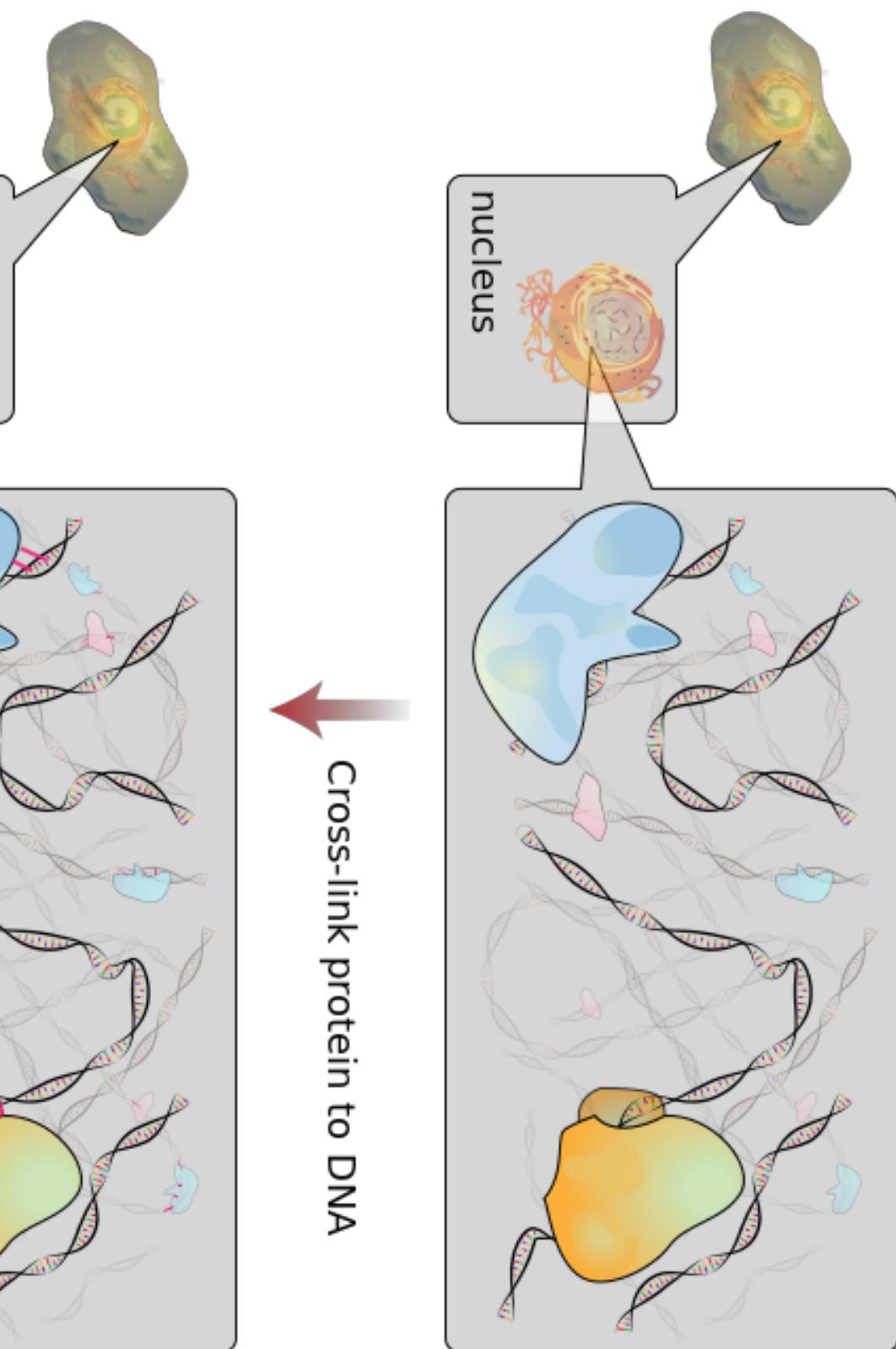


Bacterial profiles



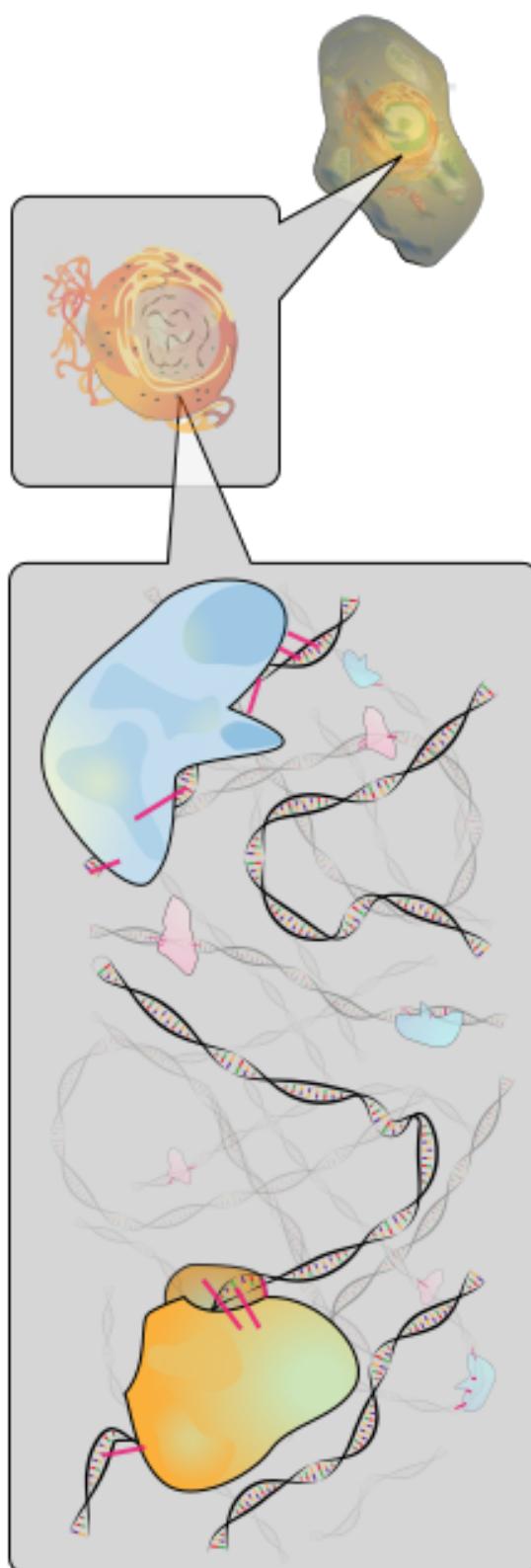


Chromatin Immunoprecipitation ChIP



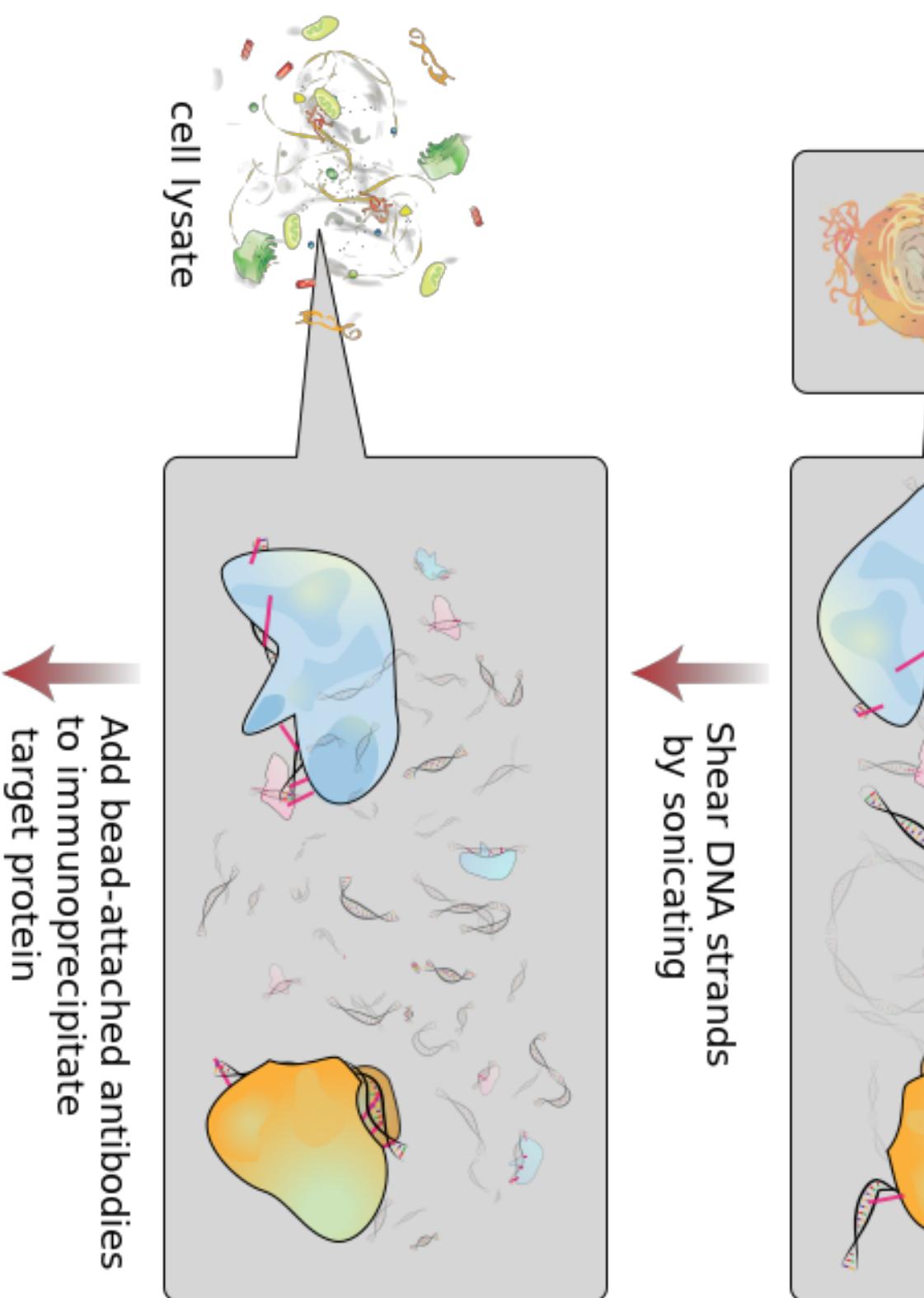


Chromatin Immunoprecipitation ChIP



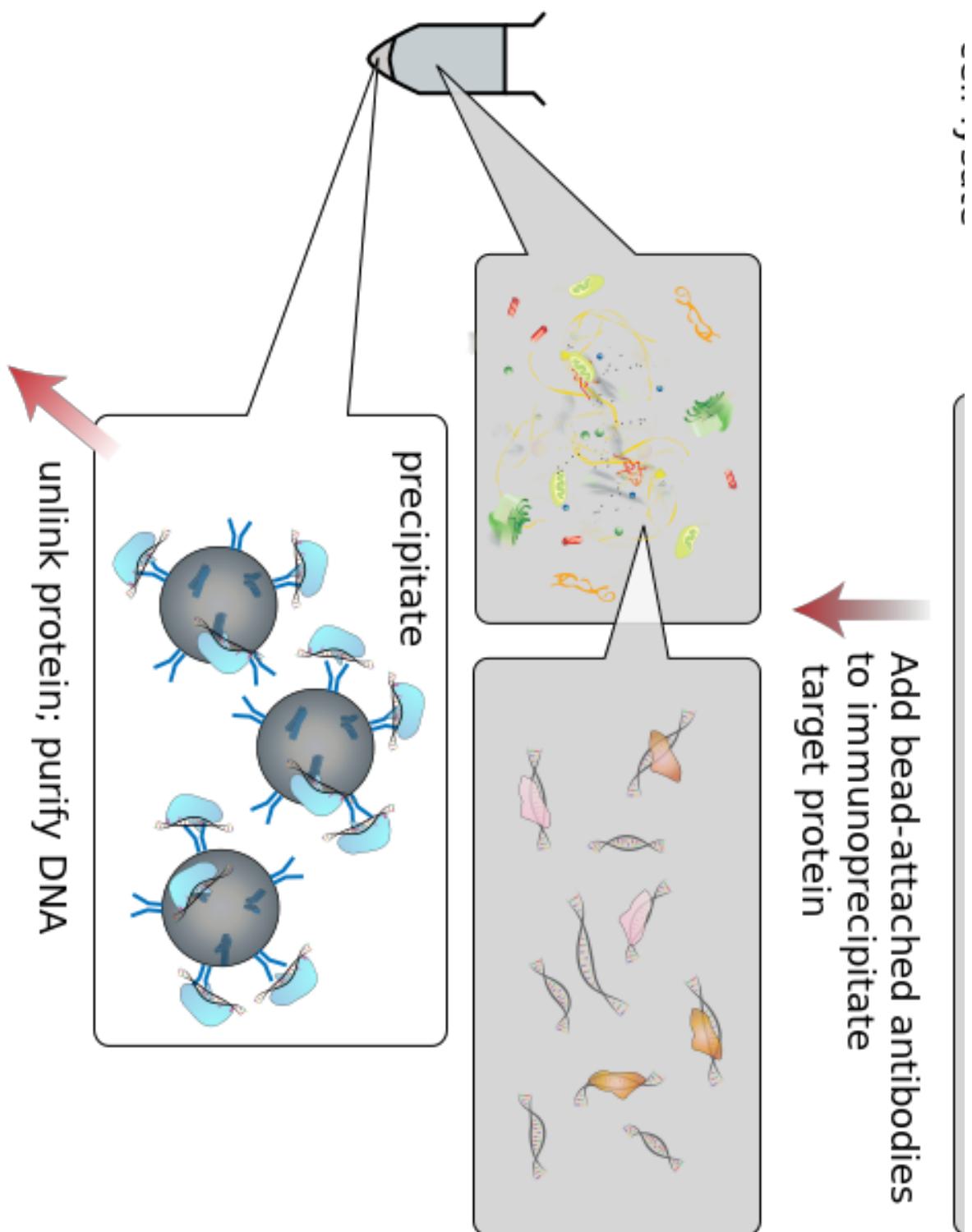


Chromatin Immunoprecipitation ChIP



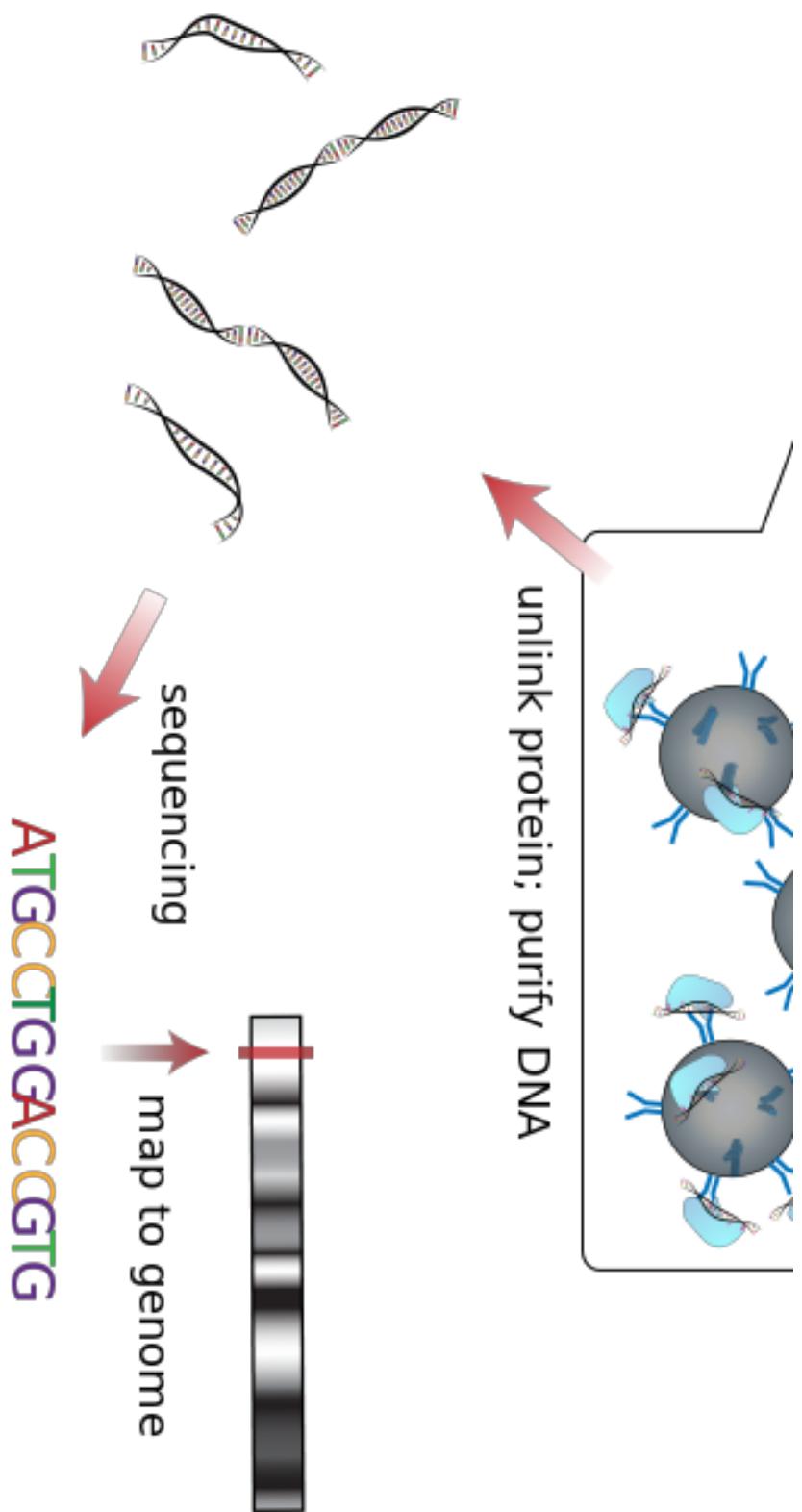


Chromatin Immunoprecipitation ChIP





Chromatin Immunoprecipitation ChIP





From analysis to synthesis

- Genomics: DNA sequence data
- Transcriptomics: Gene expression data
- Proteomics: Protein composition data
- Can we reverse this process and design our own bioproducts?



Bridging protein and dna data

20n

Open-source Platform

Bioreachables Service

Blog

Contact

20n is open sourcing its platform for synthetic biology

Over the last 4 years we have developed a better way to bioengineer organisms. We are now open sourcing our entire software stack, 20n/act. Find it at <https://github.com/20n/act>.

The stack will enumerate all bio-accessible chemicals, called *reachables* ([20n/act/reachables](#)). For each of those chemicals, it will design DNA blueprints. These DNA blueprints can bioengineering organisms with un-natural function. E.g., build organisms to make chemicals that were previously only sourced through petrochemistry.

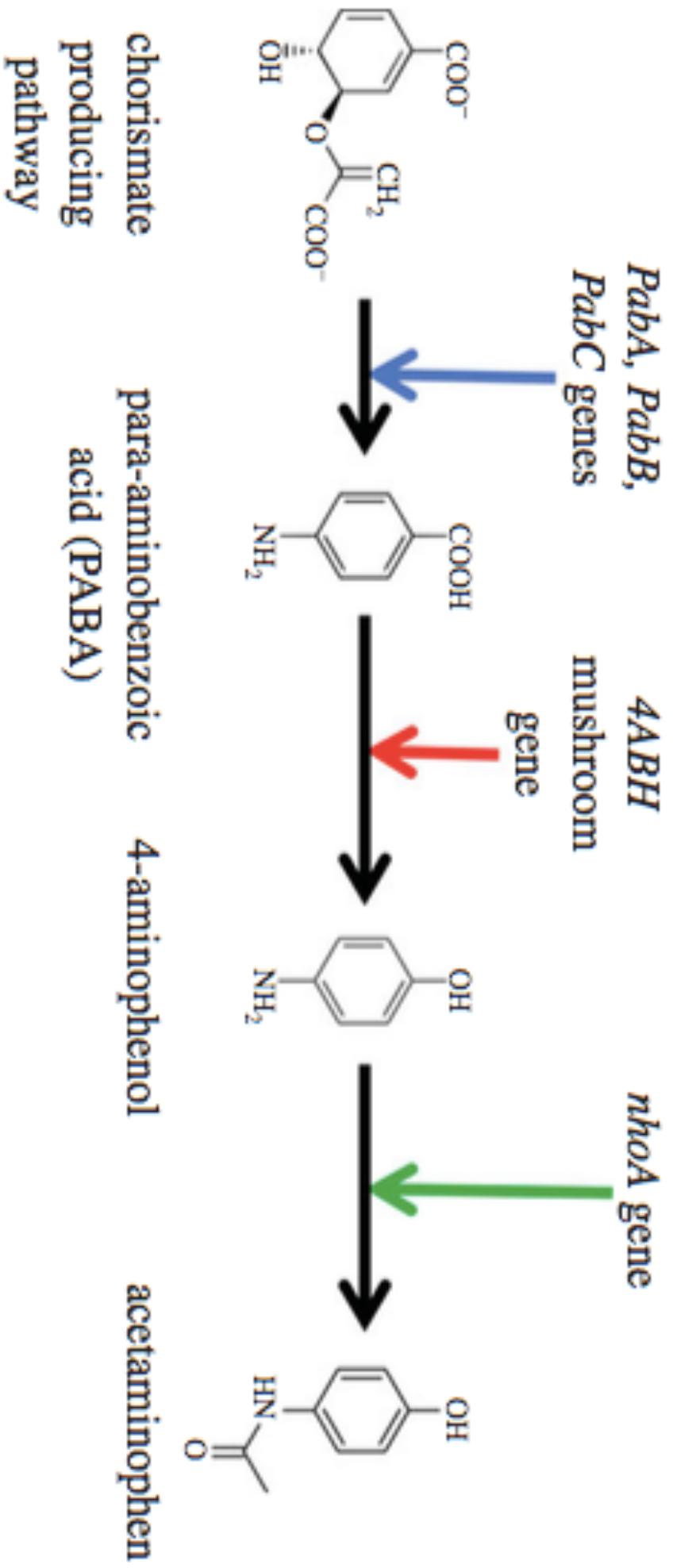
To do that, the stack contains many modules built from scratch in-house. Some of them: mine raw biochemical data, integrate heterogenous sources, learn rules of biochemistry, automatically clean bad data, mine patents, mine plain text, bioinformatic identification of enzymes with desired function.

Once the suggested DNA is used to create new engineered cells those cells can be analyzed with LCMS for function. Our [deep learning-based untargeted metabolomics](#) stack processes the raw data and enumerates all side-effects of the changed genomic structure of the cell. Some would be expected, as the organism making the desired chemical, and some unexpected metabolic changes are highlighted.

We are also releasing a [economic cost model for bioproduction](#). This economic cost model maps the desired the market price of the biological product to the "science needed" to get there. The "science needed" is measured in fermentation metrics, yield,



Predicted pathway





Bioinformatics Issues

- Ethical
 - Who owns bio data?
 - Who decides what to use data for?
 - Is de-personalized bio information possible?
- Imagine:
 - You are one of the only persons immune to Zika virus. Are you entitled to royalties on the vaccine derived from your blood?



reserved
some
rights



Visualisation PyMol

