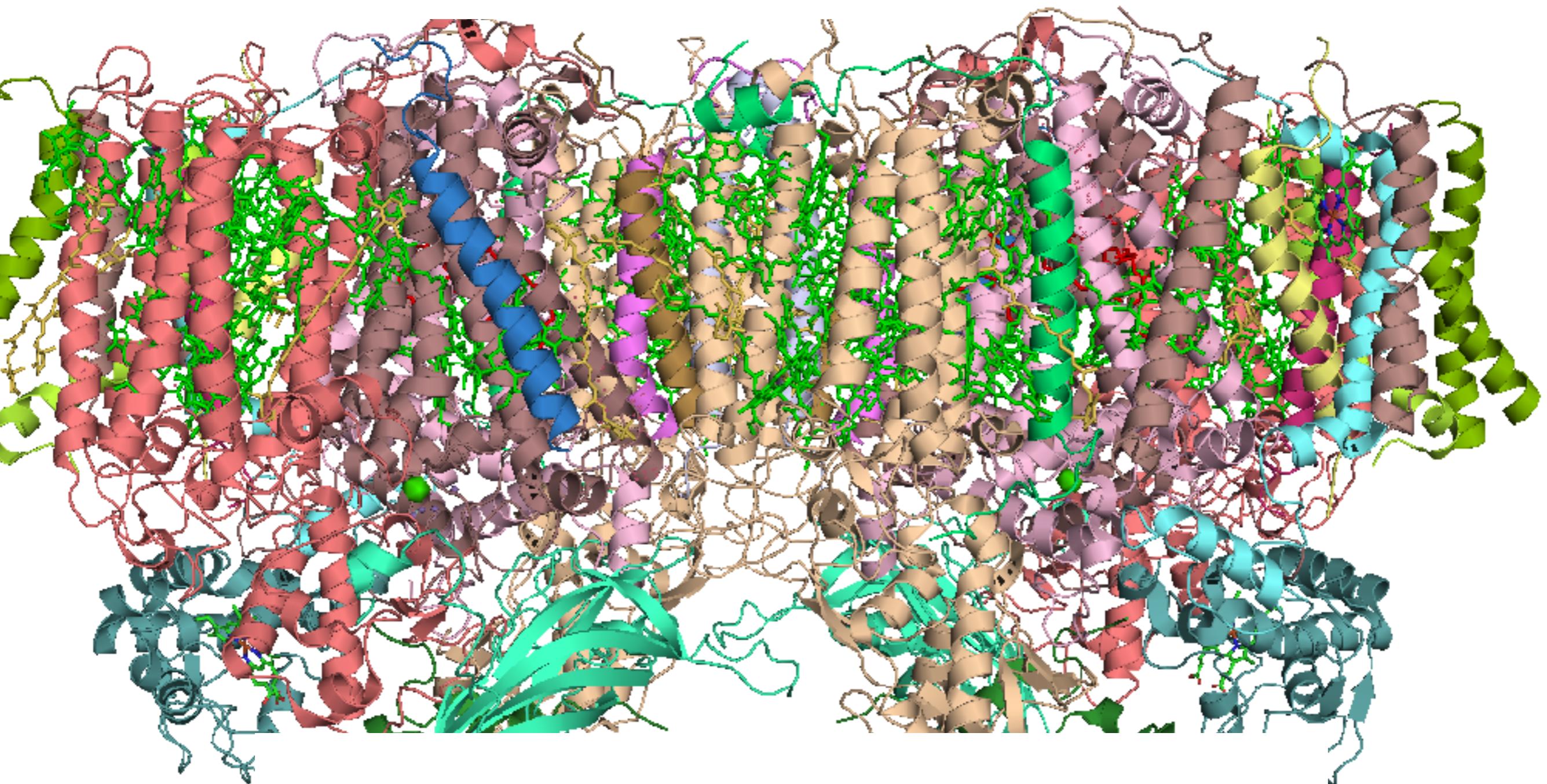




waag
wetlab amsterdam



Bio Informatics

- US VV



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 ▾](#) ① Midterm - Amy Traylor - Bioshield

#32 opened a day ago by trembl

 ① Midterm - Rowan Roberts - tear gas / pepper spray

#31 opened a day ago by trembl

 ① Midterm - Sophia Montaya - Persephone

#30 opened a day ago by trembl

 ① Midterm - June Lee - Microbe Portraiture

#29 opened a day ago by trembl

1

 ① Midterm - Inger Le Gue - Resistance is fertile

#28 opened a day ago by trembl

 ① Midterm - Candyce Dryburgh - Cultivated Heritage

#27 opened a day ago by trembl



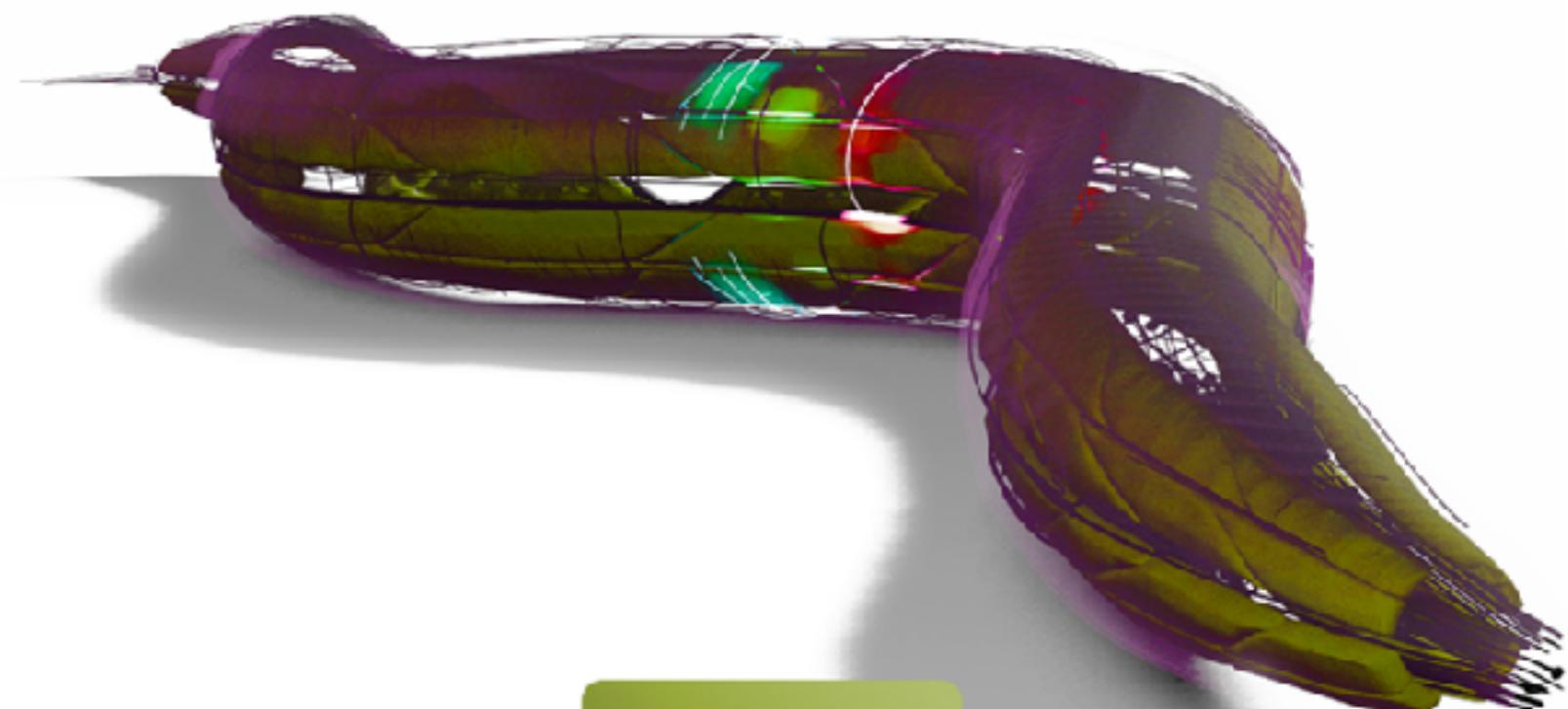
The information / “Omics” age

- “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





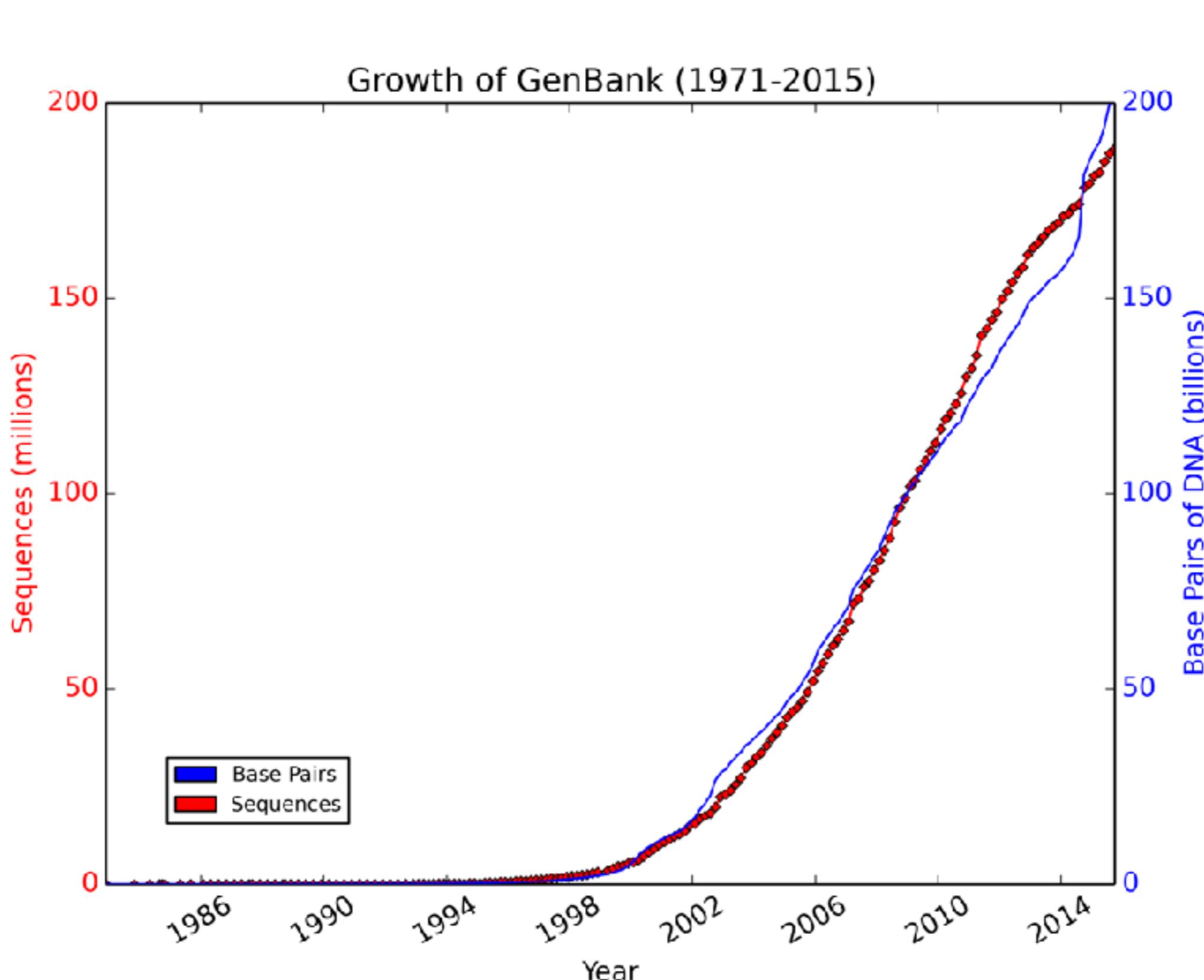
Genomics

- Functional genomics
- Metagenomics
- Personal Genomics
- Epigenomics





DNA database GenBank





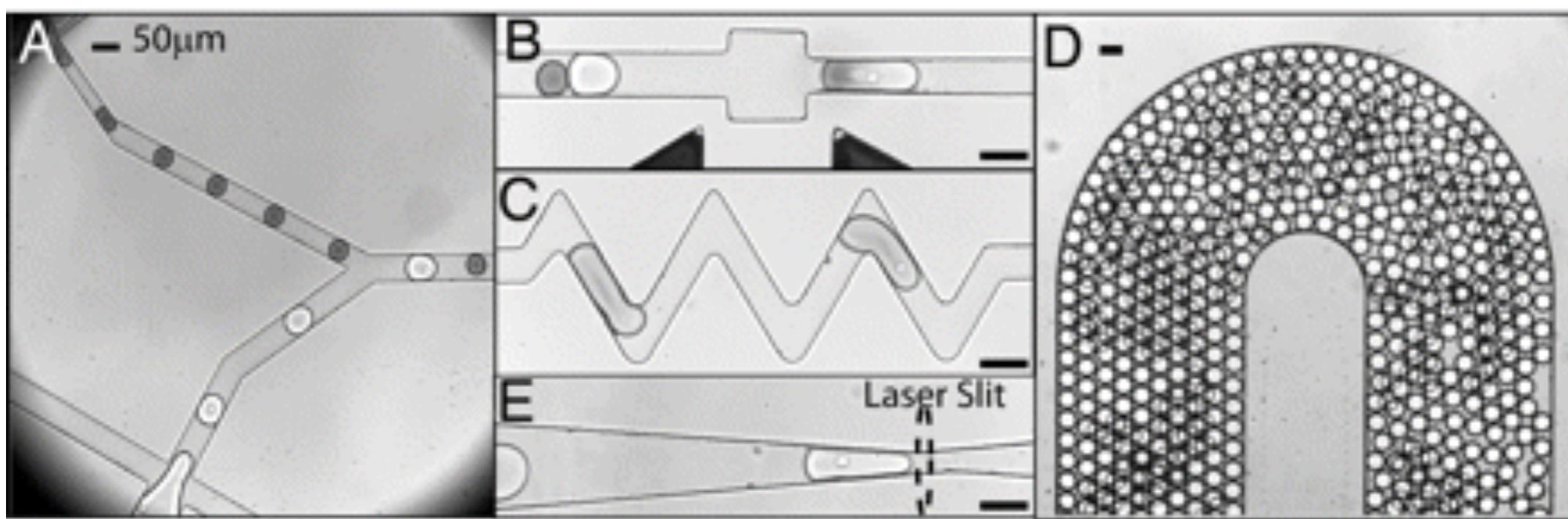
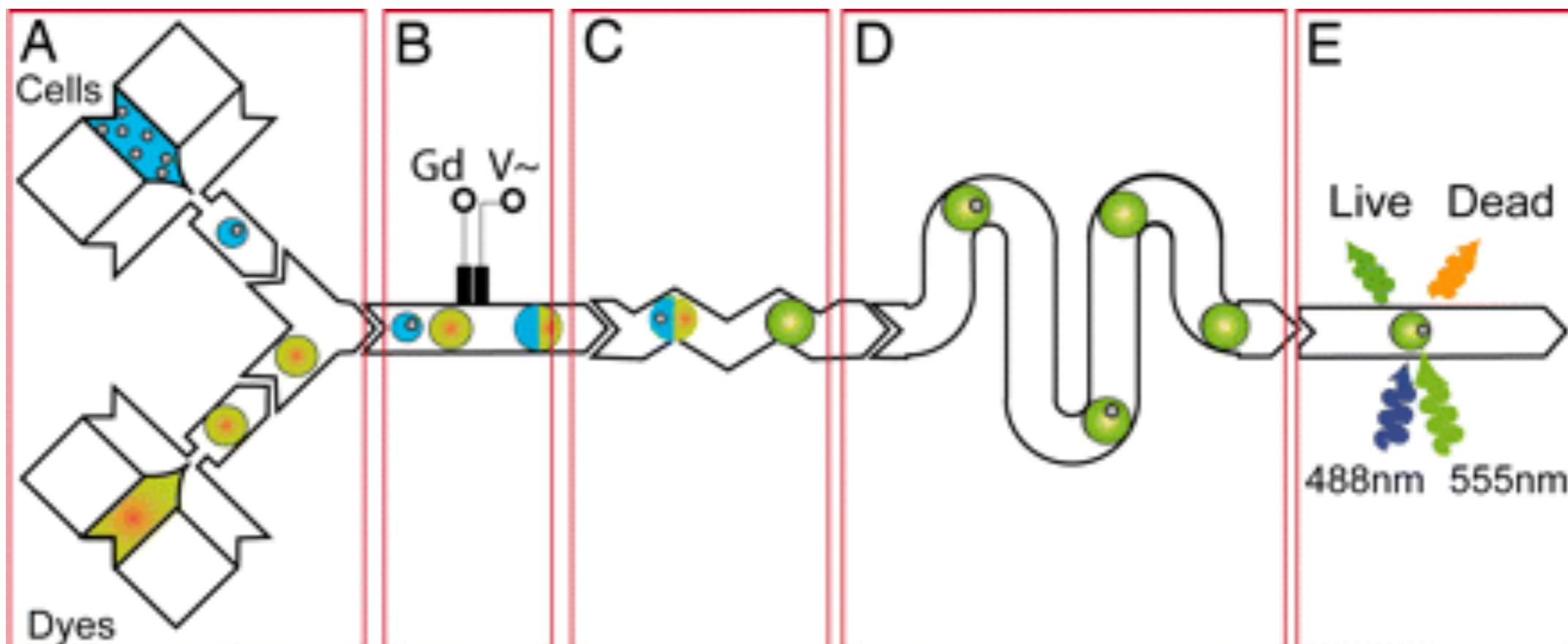
Drivers

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



High Throughput Screening

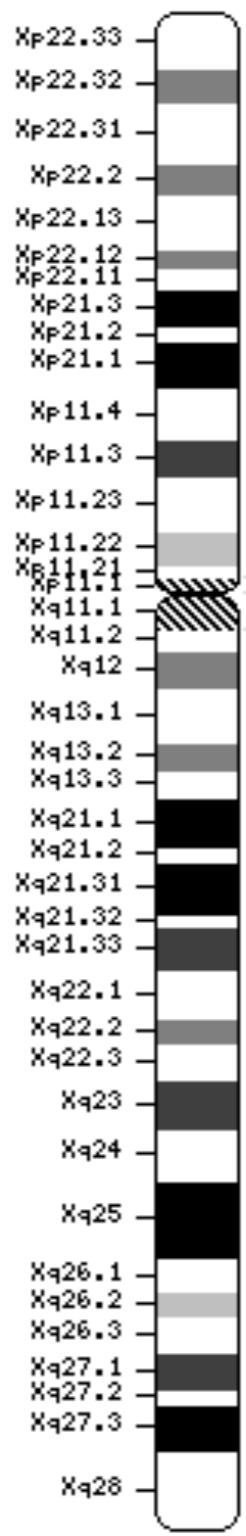
Eric Brouzes et al. PNAS 2009;106:14195-14200





Gene annotations

Ideogram → X





Sequence Alignment

14	SIKLWPPSQTTTRLLLVERMANNLST..PSIFTRK..YGSLSKEEAREN AKQIEEVACSTANQ.....HYEKEPDGDGGSAVQLYAKECSKLILEVLK	101
13	SIKLWPPSESTRIMLVDRMTNNLST..ESIFSRK..YRLLGKQEAHENAKT IEELCFALADE.....HFREEPDGDGSSAVQLYAKETSKMMLEVLK	100
23	VFKLWPPSQGTREAVRQKMAKLSS..ACFESQS..FARIELADAQE HARAIEEVAFGAAQE.....ADSGGDKTGSAGVVMVYAKHASKLMLET LR	109
13	SVKLWPPGQSTRMLVERMTKNFIT..PSFISRK..YGLLSKEEA EEDAKKIEEVAFAAANQ.....HYEKQPDGDGSSAVQIYAKESSRLM LEVLK	100
30	SFSIWPPPTQRTRDAVVRRLVDTLGG..DTILCKR..YGAVPA ADAEPAAARGIEAEAFDAAAA..SGEAAAATASVEEGIKALQLYS KEVSRRLLDFVK	120
44	SLSIWPPSQRTRDAVVRRLVQTLVA..PSILSQR..YGAVPEAE AGRAAAAVEAEAYAAYTES..SSAAAAPASVEDGI EVLQAYSKEVSRRLLELAK	135
56	SFSIWPPPTQRTRDAIIISRLIETLST..TSVLSKR..YGTIPKEE EASEASRRRIEEEAFSGAST.....VASSEKDGL EVLQLQLYSKEISKRMLETVK	141
29	SFAVWPPTRRTRDAVVRRLVA VLSGDTTTALRKRYRYGAVPA ADAERAARAVEAQAFDAASA.....SSSSSS VEDGIETLQLQLYSREVSNRLLAFVR	121
13	SIKLWPPSESTRMLVERMTDNLSS..VSFFSRK..YGLLSKEEA AENAKRIEETAFLAAND.....HEAKEPNLDDSSVV QFYAREASKLMLEALK	100
57	SLRIWPPTQKTRDAVLNRLIETLST..ESILSKR..YGT LKSDDATTVA KLIEEEAYGVASN.....AVSSDDD GIKILELYSKEISKRMLESVK	142
25	NYSIWPPKQRTRDAVKNRLIETLST..PSVLT K..YGTMSADE ASAAAIQIEDEAF SVANA.....SSSTSNDNV TILEVYSKEISKRM IETVK	110
28	SFKIWPPPTQRTR EAVVRRLVETLTS..QSVLSKR..YGV IPEEDAT SAARIIEEA FSVAVS ASA A ASTGG RPE DEW JIE V LHI Y S Q E I X Q R V V E S A K	119
25	SFSIWPPPTQRTRDAVINRLIESLST..PSILSKR..YGTLPQ DE A SET ARL IEEEAF AAAGS.....TASDADD GIEILQV Y S K E I S K R M I D T V K	110
14	SVKMWPPSKSTRMLVERMTKNITT..PSIFSRK..YGLLSVE EA EQDAK RIEDLA FATANK.....HFQNEPDGD GTS AVH VY AK E SS K L M D V K	101
13	SIKLWPPSLPTRKALIERITNNFSS..KTIFTEK..YGS LTKDQ ATENA KRIEDIA FSTANQ.....QFEREPDG DGG SAVQLY AKECS KLILE VLK	100
48	SLSIWPPPTQRTRDAV ITRLIETLSS..PSVLSKR..YGT ISH DEA ESA ARR IEDEAF GVANT.....ATS AED DG LEILQL Y S K E I S R R M L D T V K	133



BLAST: Basic Local Alignment Search Tool

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

BLAST® Basic Local Alignment Search Tool

NCBI BLAST! blastn suite Standard Nucleotide BLAST

Enter Query Sequence

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

Or, upload file no file selected [?](#)

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.):
[Nucleotide collection \(nr/nt\)](#) [?](#) [+](#) [-](#)

Organism Optional Enter organism name or id—completers will be suggested Exclude [?](#)
Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown [?](#)

Exclude Optional Models (XM/XP) Uncultured/environmental sample sequences

Limit to Optional Sequences from type material

Entrez Query Optional [YouTube](#) [Create custom database](#)
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



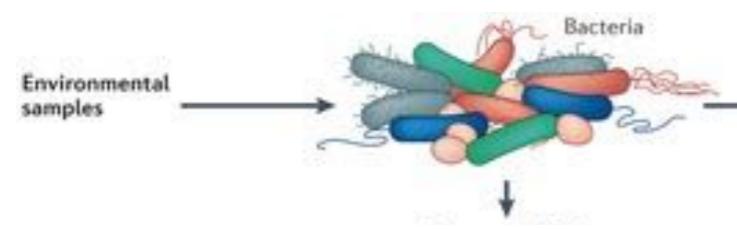
Scoring Matrix BLOSUM

(BLOcks SUbstitution Matrix)

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

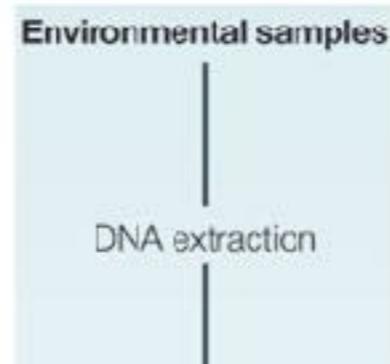
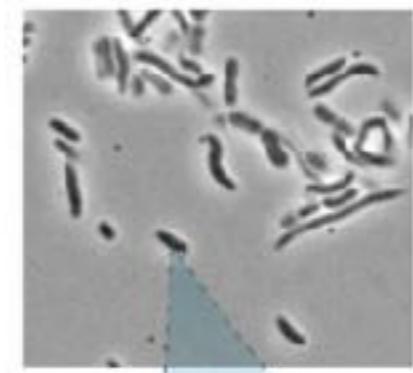


Environmental DNA analysis



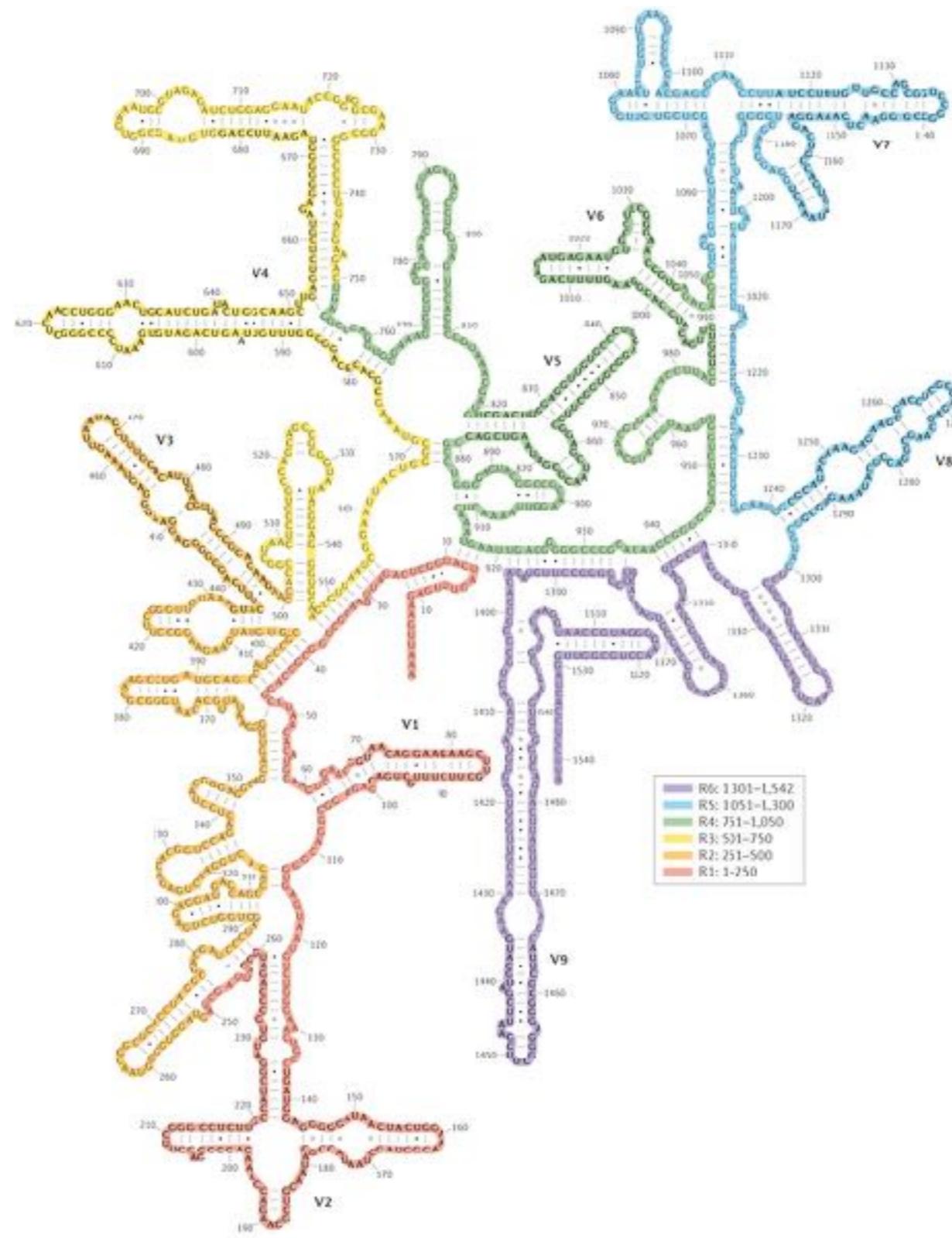


16S RNA





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

[sign in](#) [register kit](#)

welcome health ancestry how it works research buy help

**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

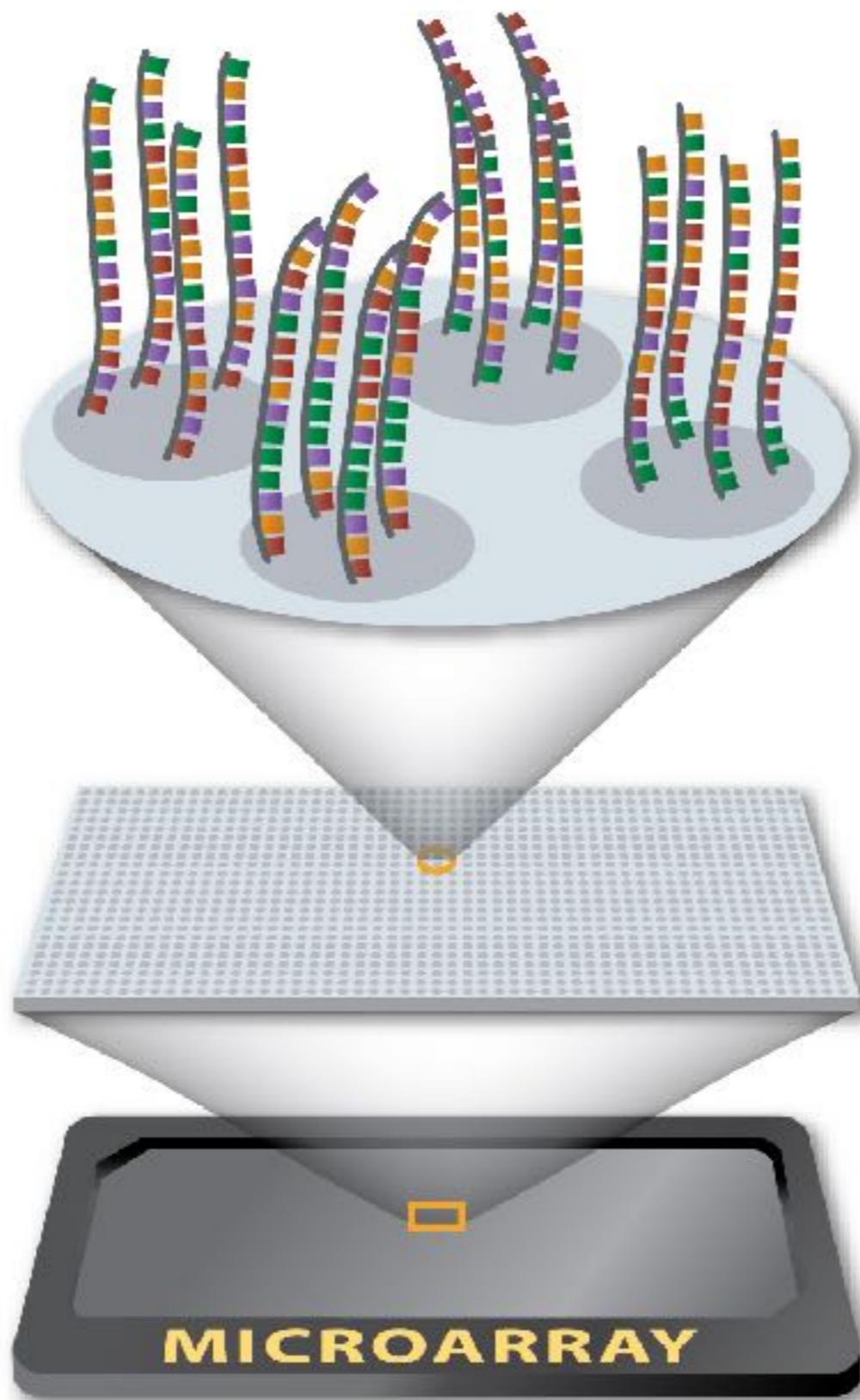
[order now](#) **€169**
shipping included

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 metabolises caffeine quickly or if you're likely lactose intolerant. We believe the more you know about



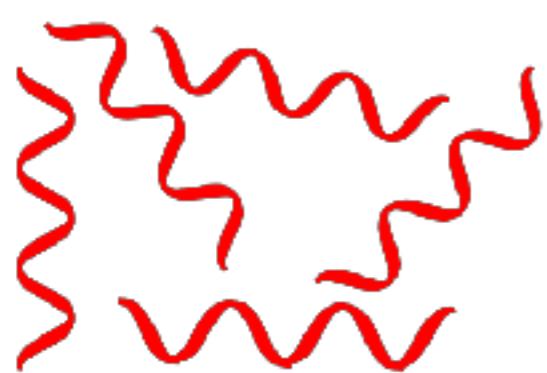
DNA Microarray





DNA Microarray Expression Analysis

DNA extraction and digestion



Test = Tumoral DNA
labeled with **Cy5**



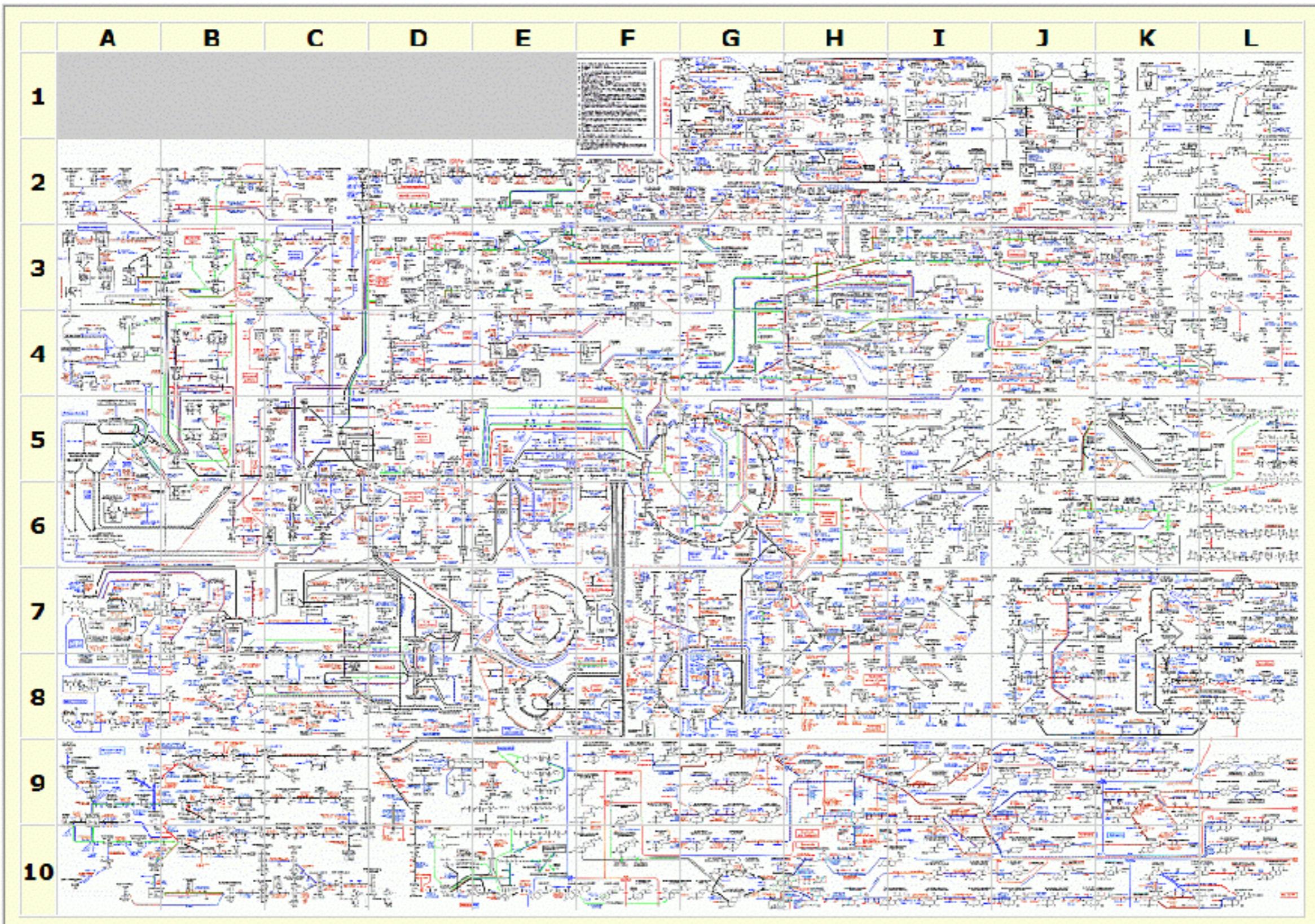
Reference = Normal DNA
labeled with **Cy3**



Proteins

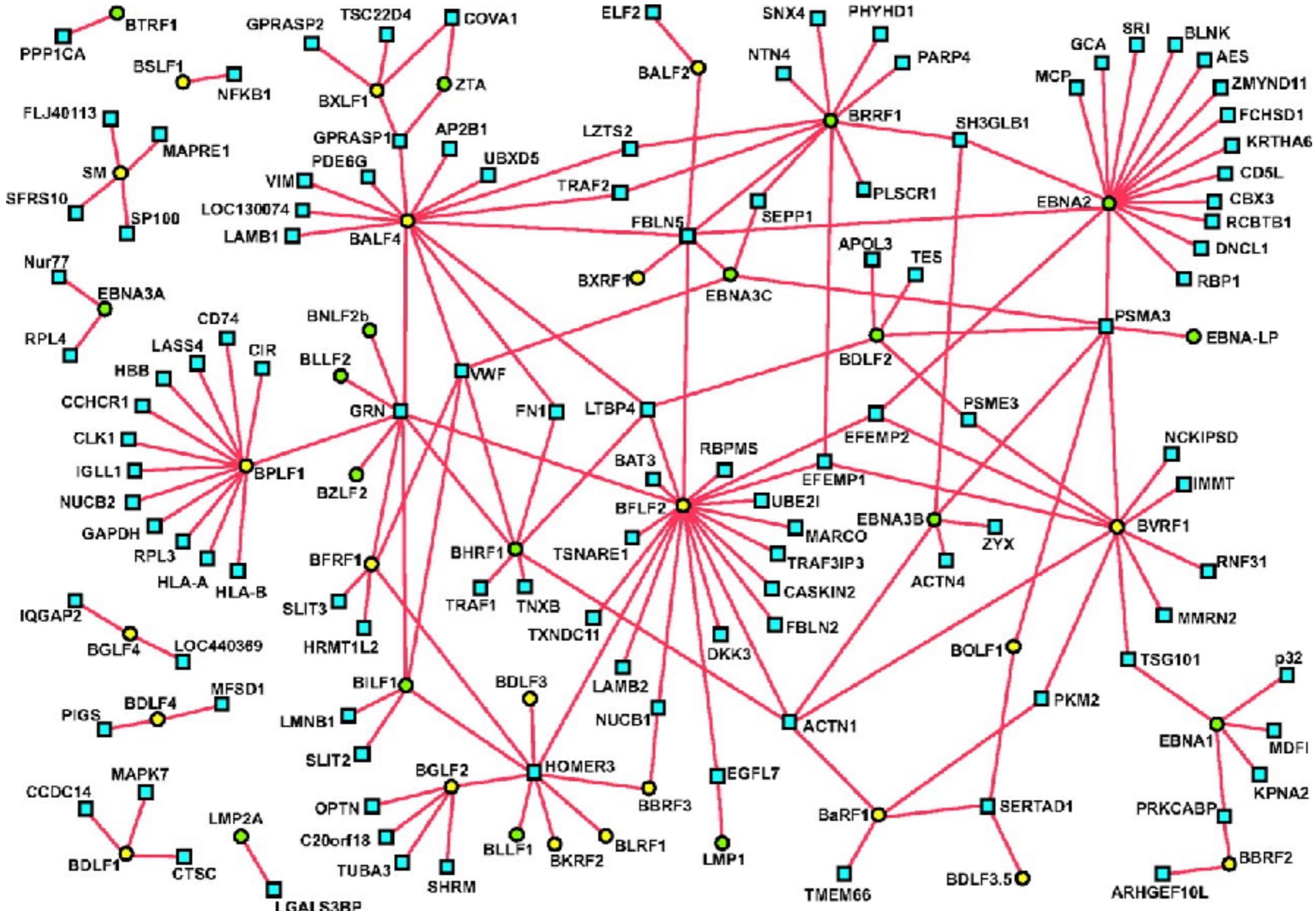


Biochemical Pathways of the Cell





Protein interaction mapping: MS





What is this?

Simon Eugstar - CC-BY-SA 3.0



Debstart - CC-BY-SA 3.0





NMR Machines

MartinSaunders

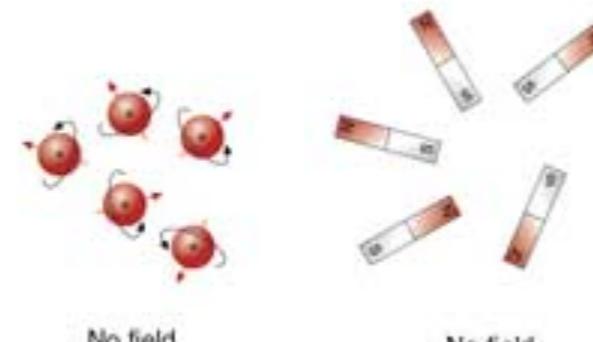
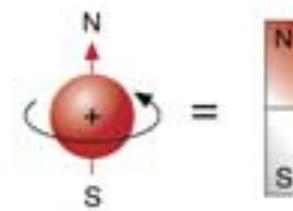
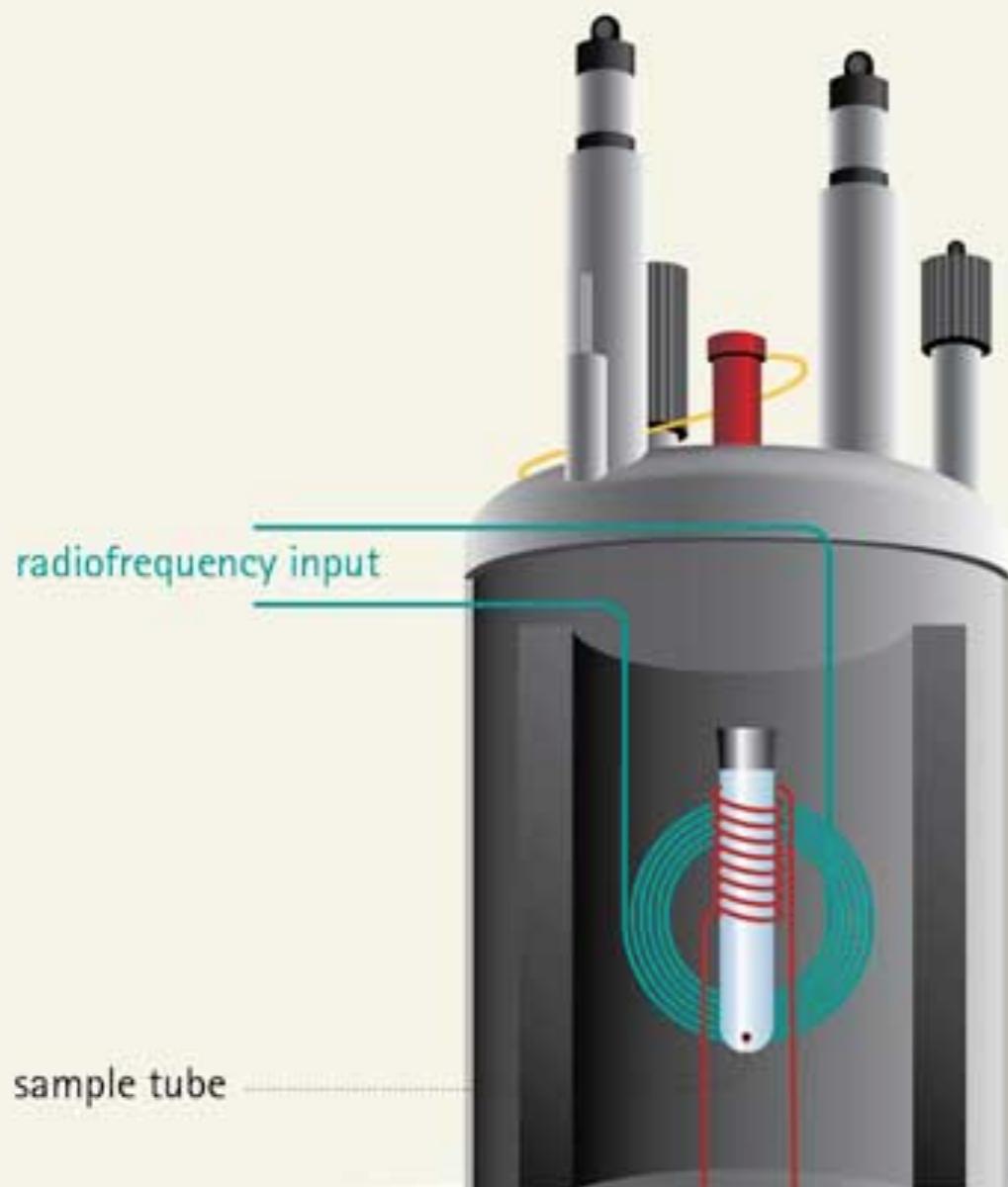


Public Domain

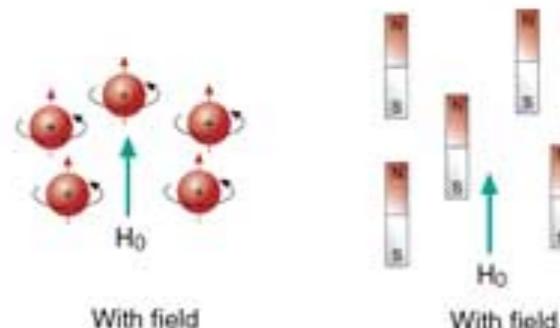




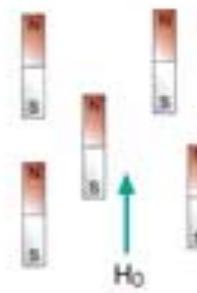
NMR principles



No field No field



With field

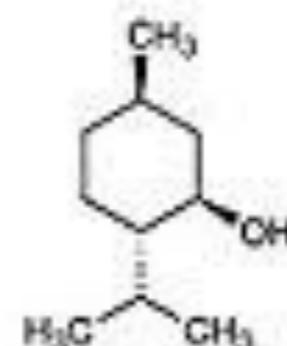
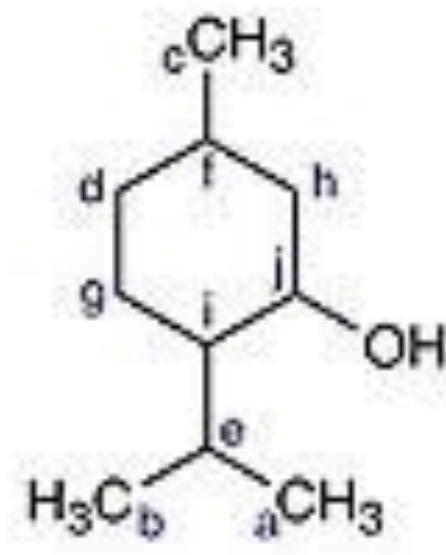


With field

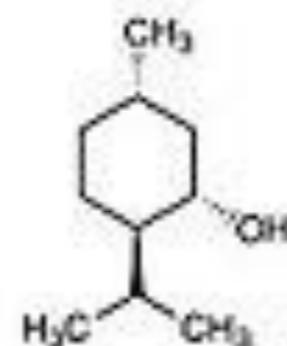


NMR Spectrum

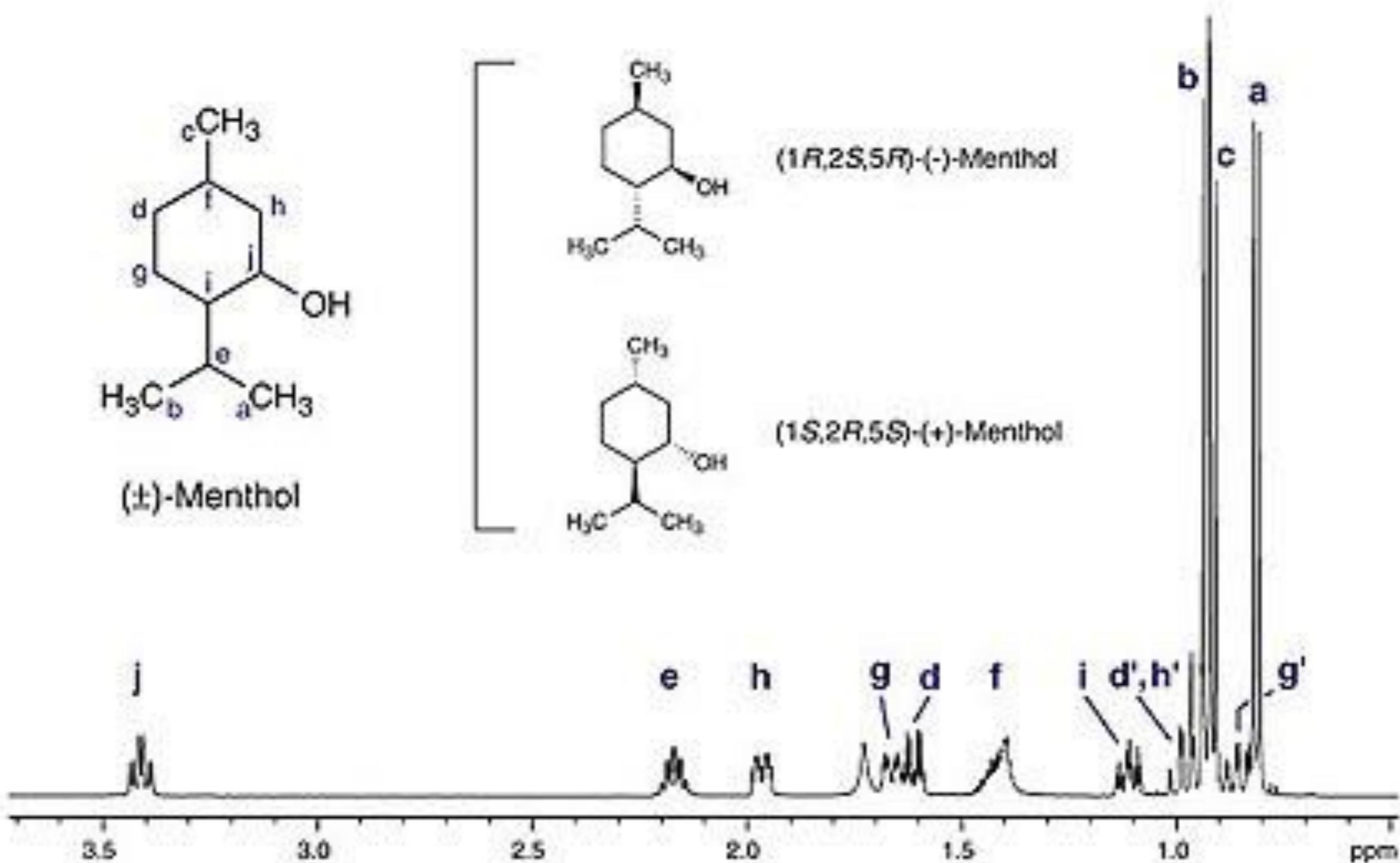
1D PROTON SPECTRUM



(1*R*,2*S*,5*R*)-(-)-Menthol



(1*S*,2*R*,5*S*)-(+)-Menthol

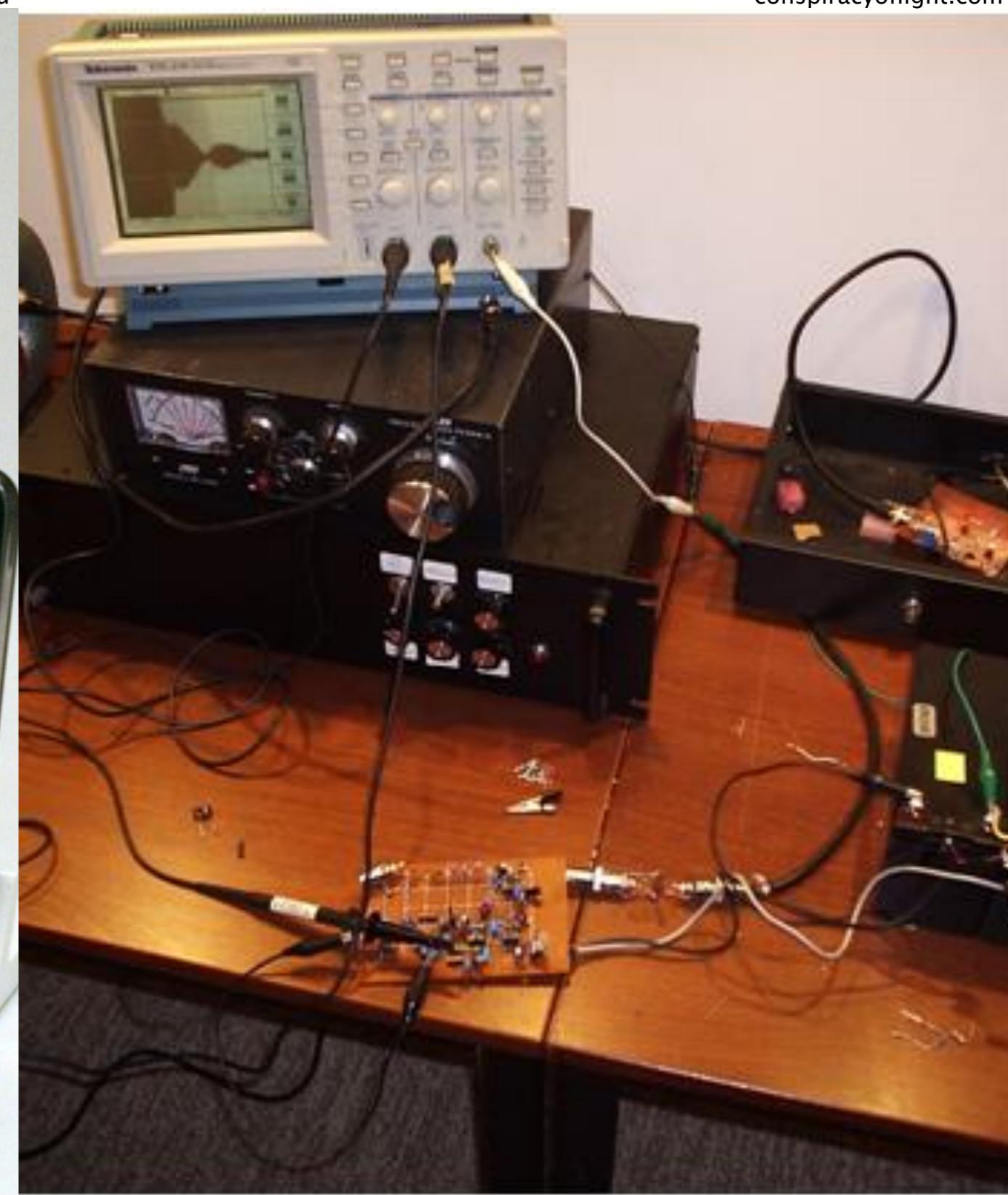
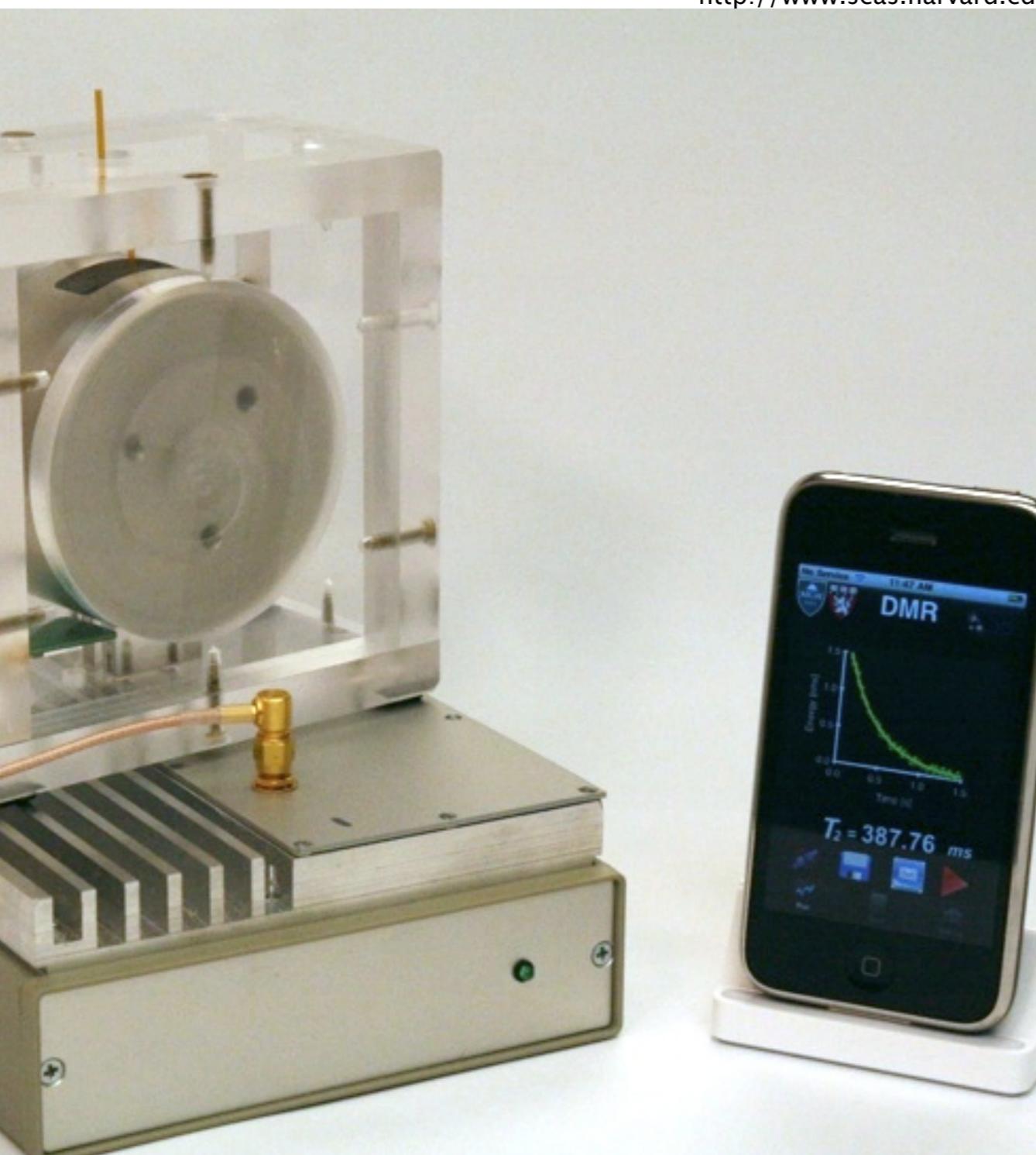




DIY NMR?

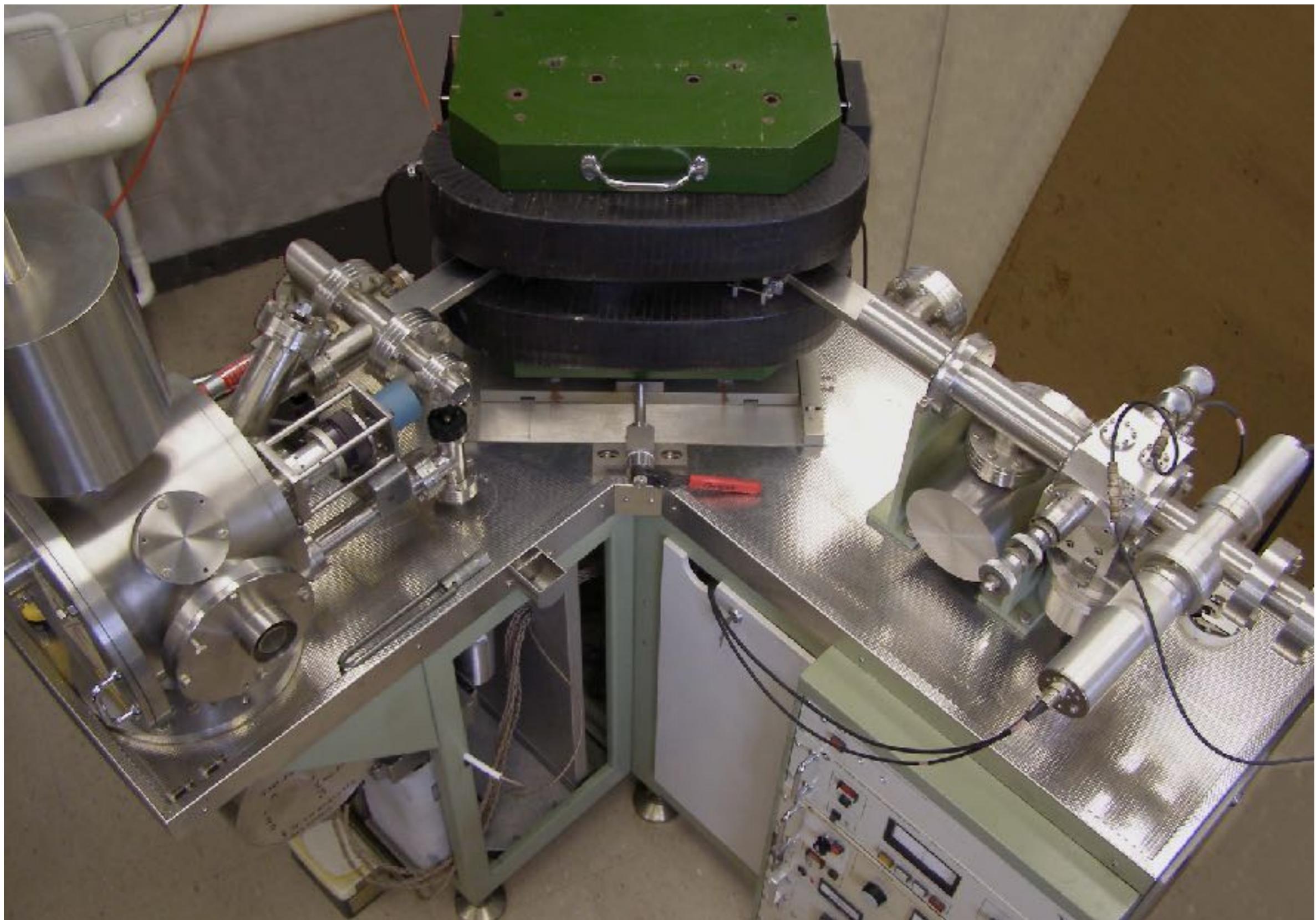
<http://www.seas.harvard.edu>

conspiracyoflight.com



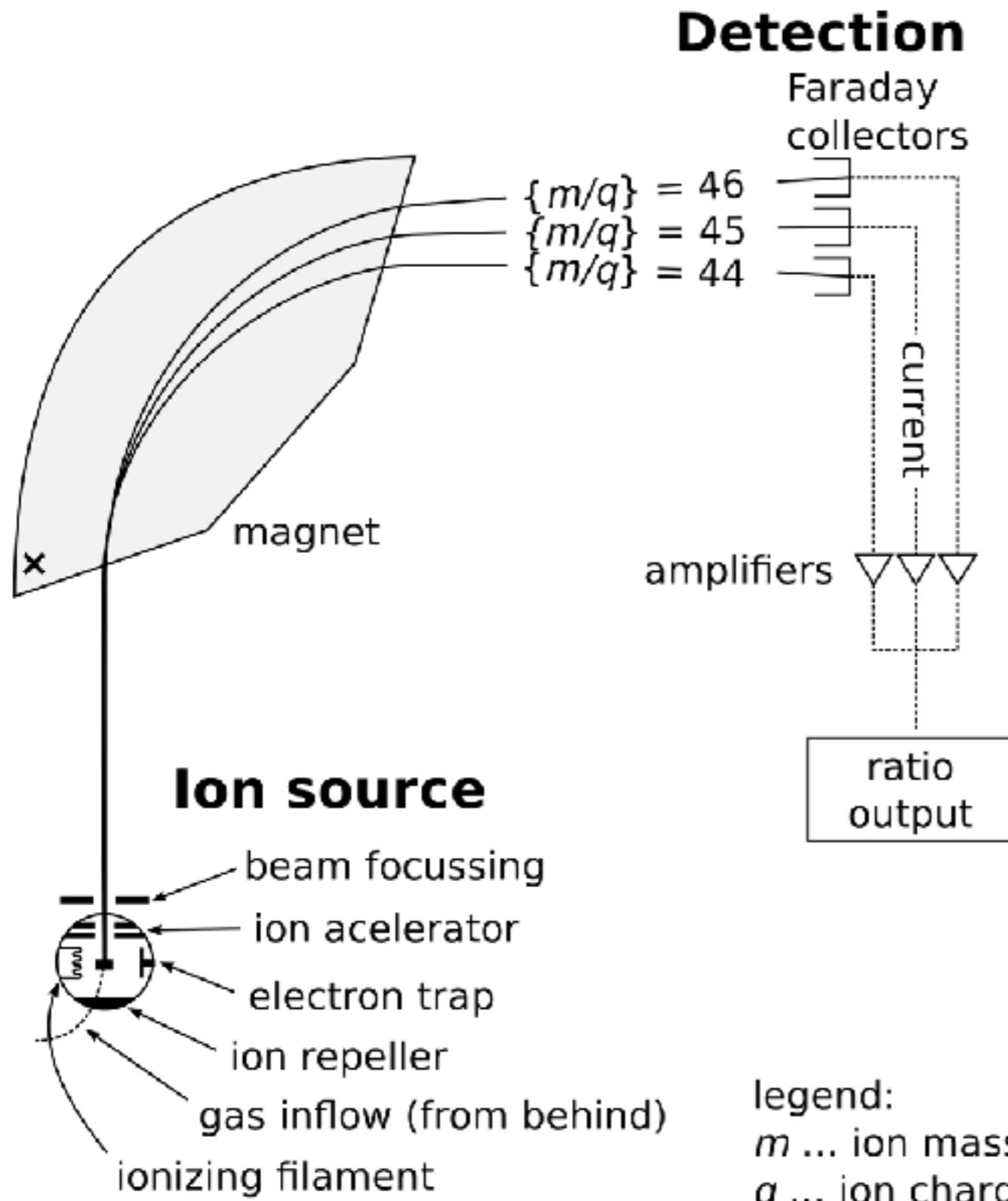


Mass Spectrometer





Simple Mass Spectrometry



High mass = high m/q

Big charge = low m/q

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



Mass Spectrometry



cells or tissue

TOF

MALDI

Tandem



Procedure



Sampling the colony

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

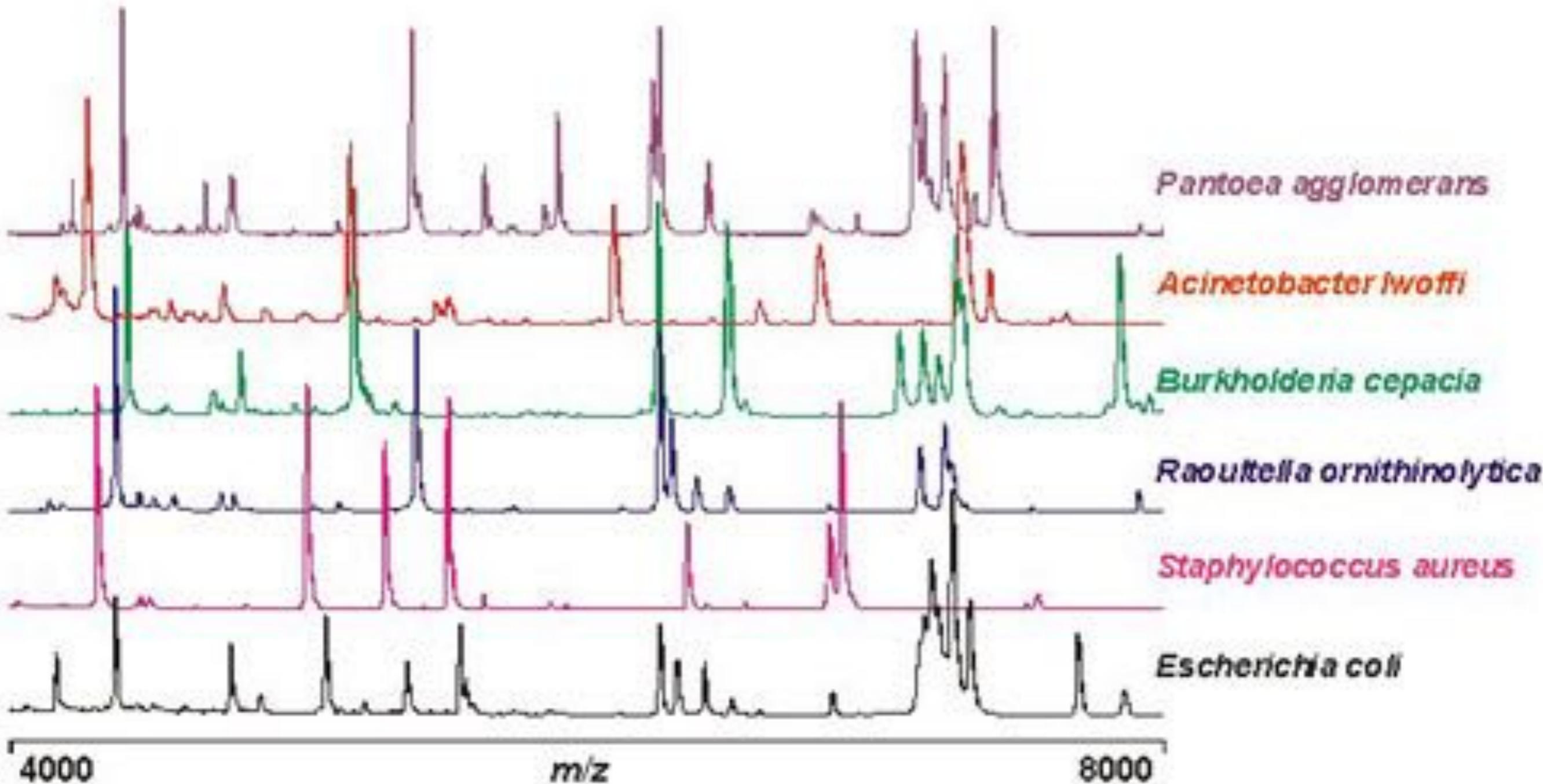
TOOLS: Petri dish, swab, 70% formic acid, absolute ethanol

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

Introduction of the target into the apparatus

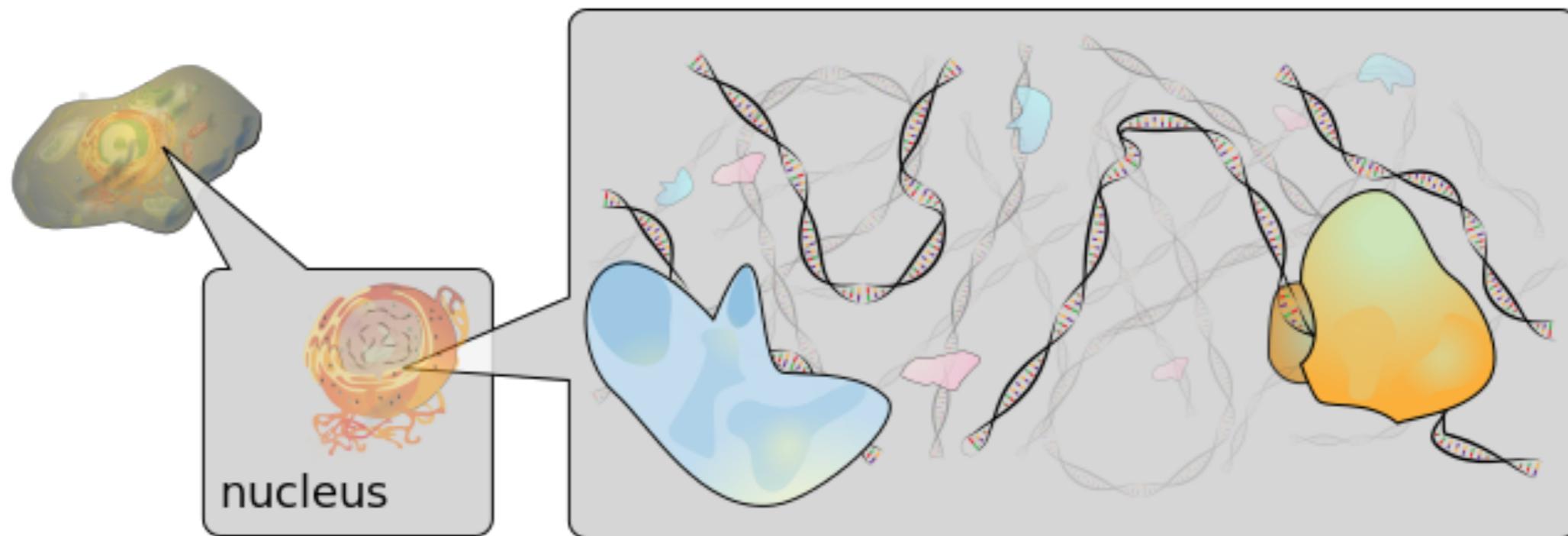


Bacterial profiles

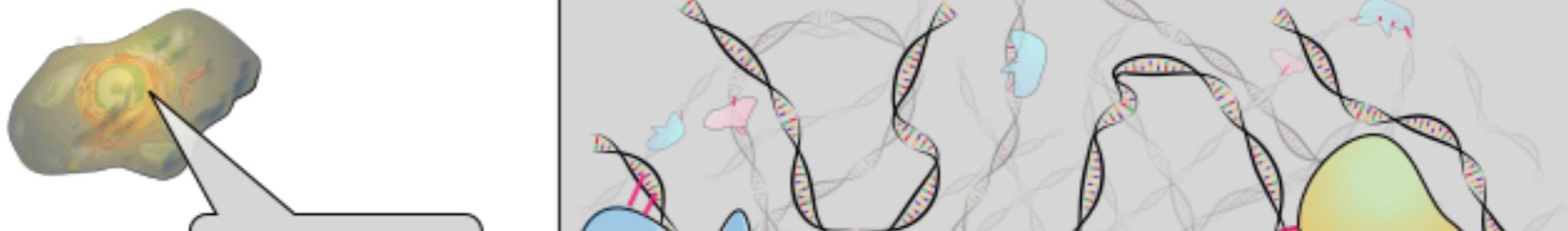




Chromatin ImmunoPrecipitation ChIP

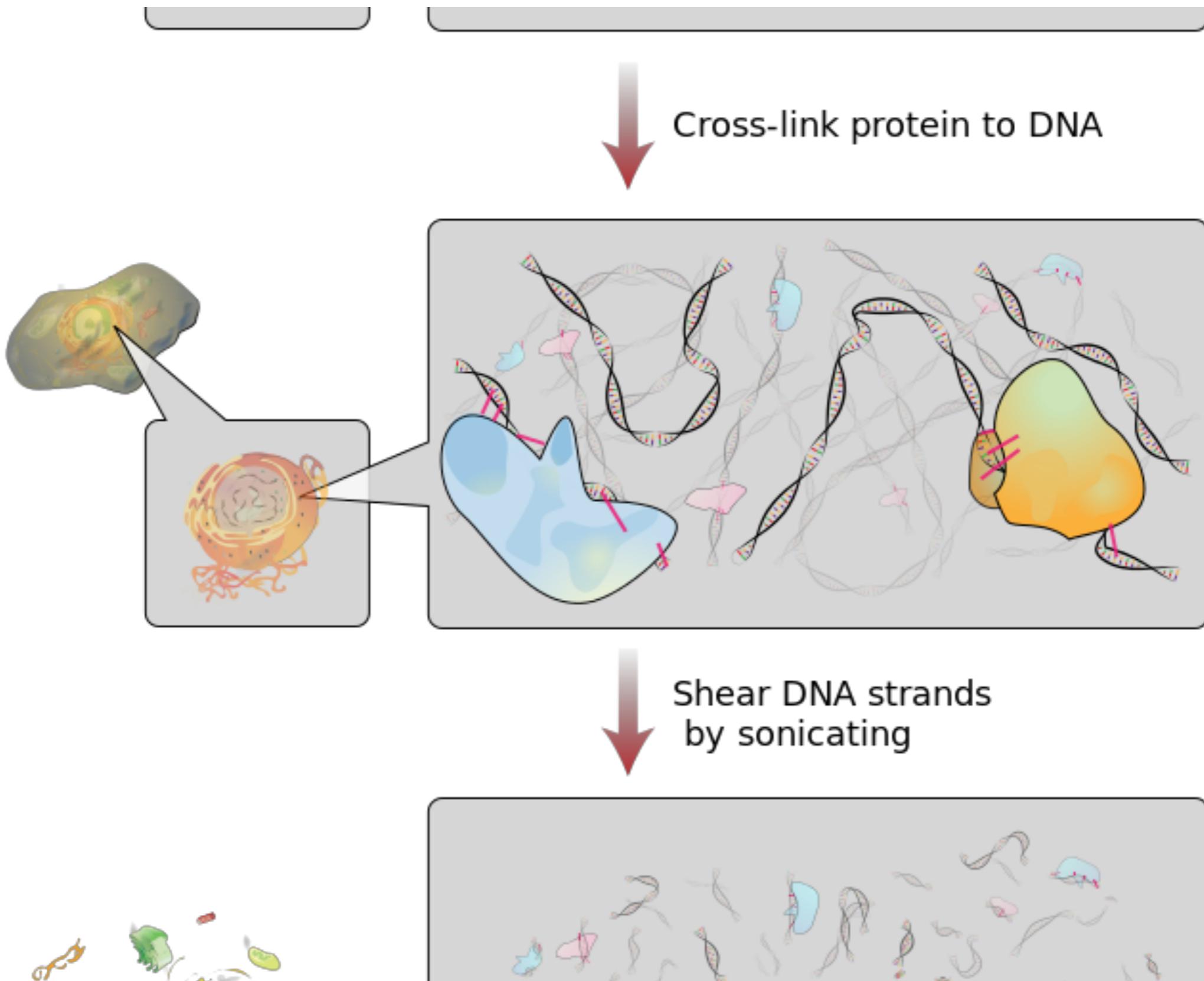


Cross-link protein to DNA



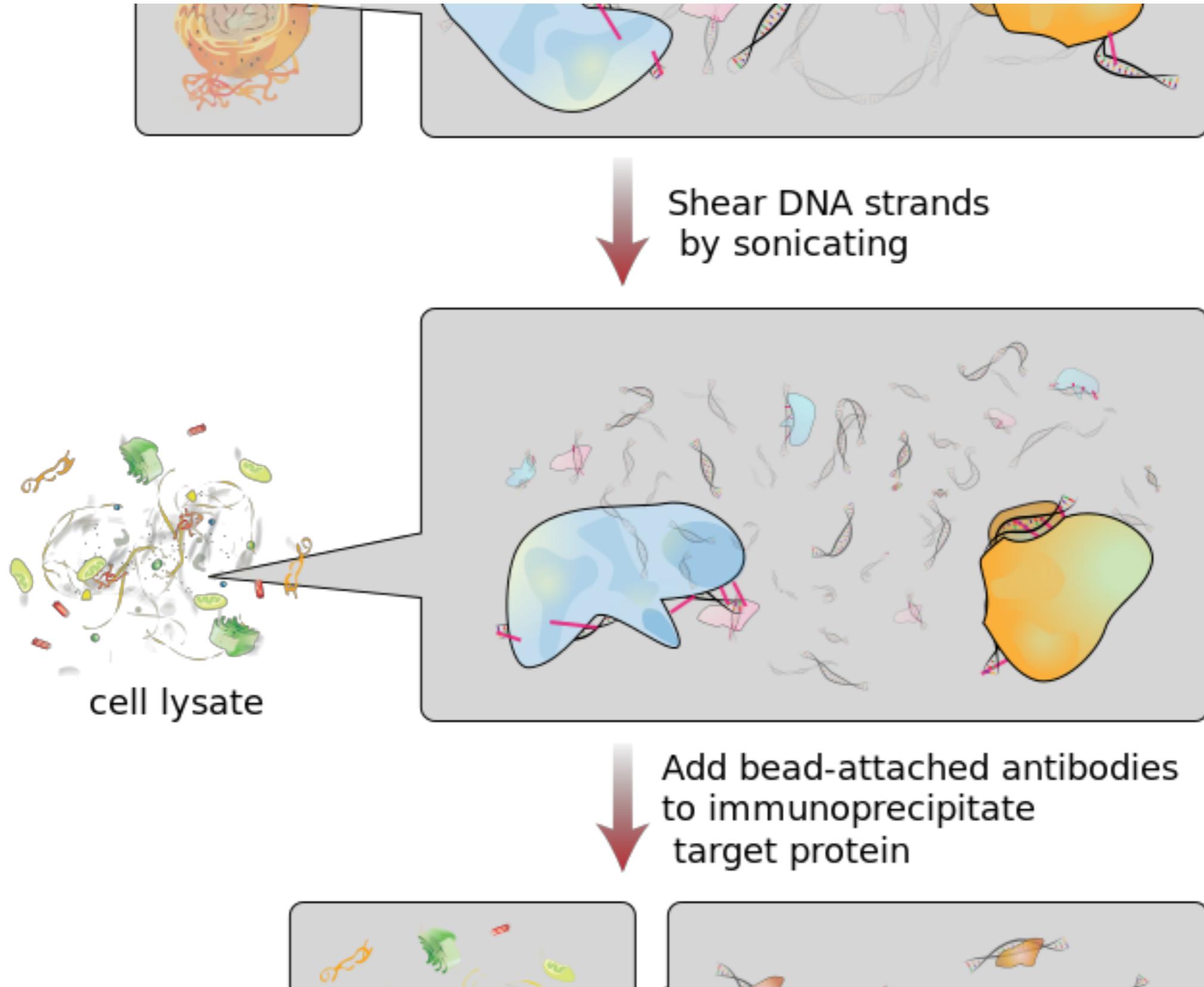


Chromatin ImmunoPrecipitation ChIP



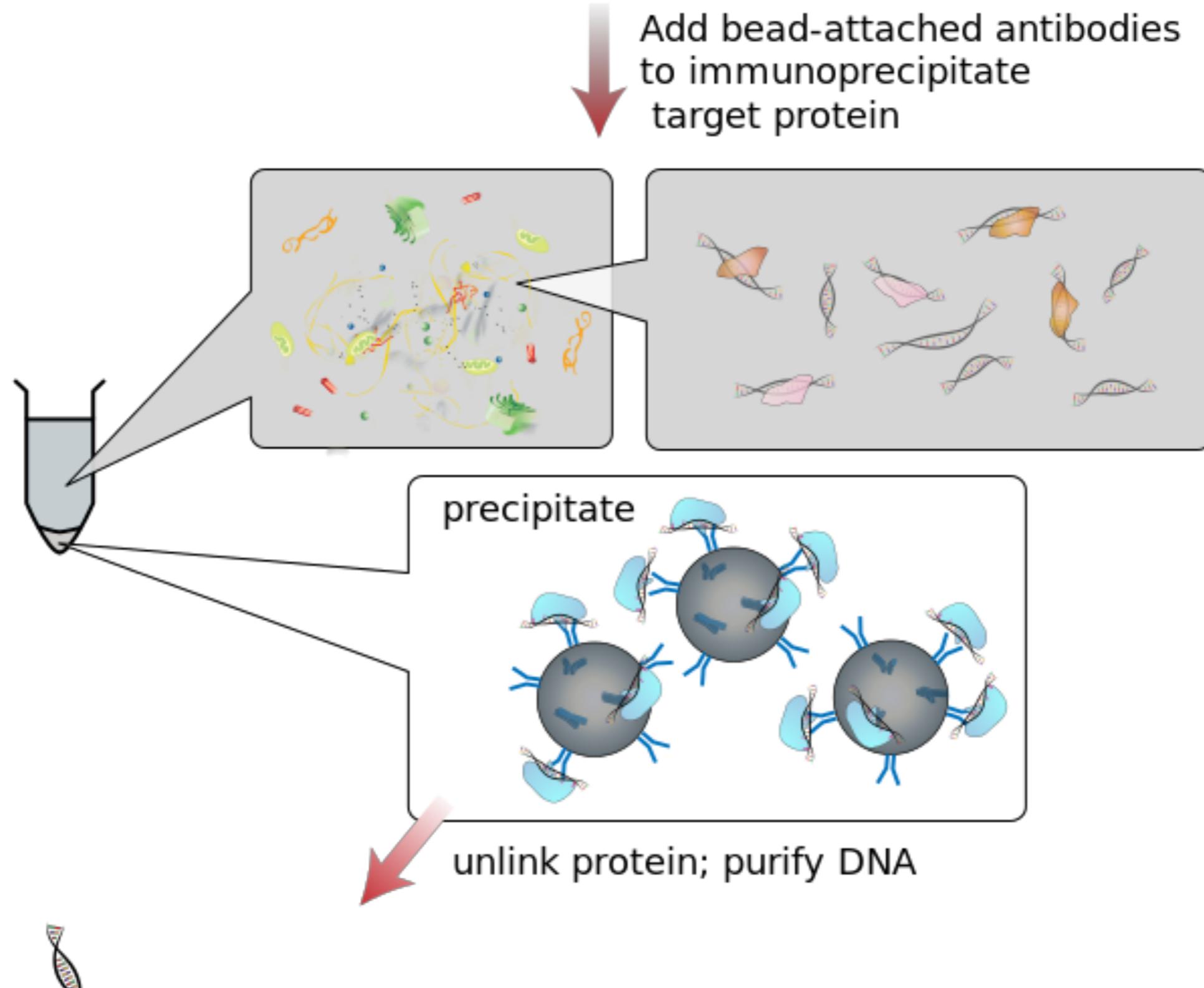


Chromatin ImmunoPrecipitation ChIP



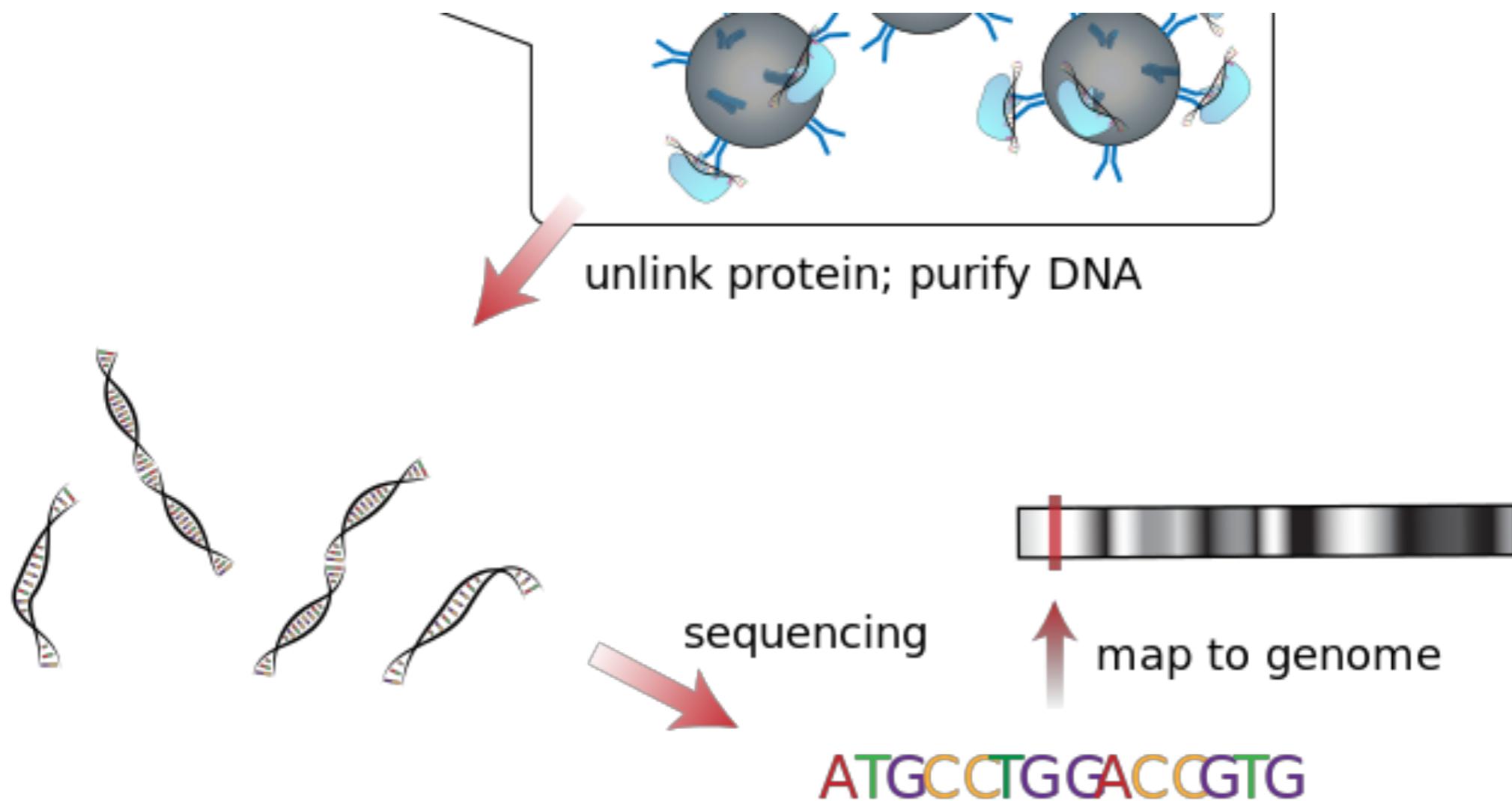


Chromatin ImmunoPrecipitation ChIP





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.

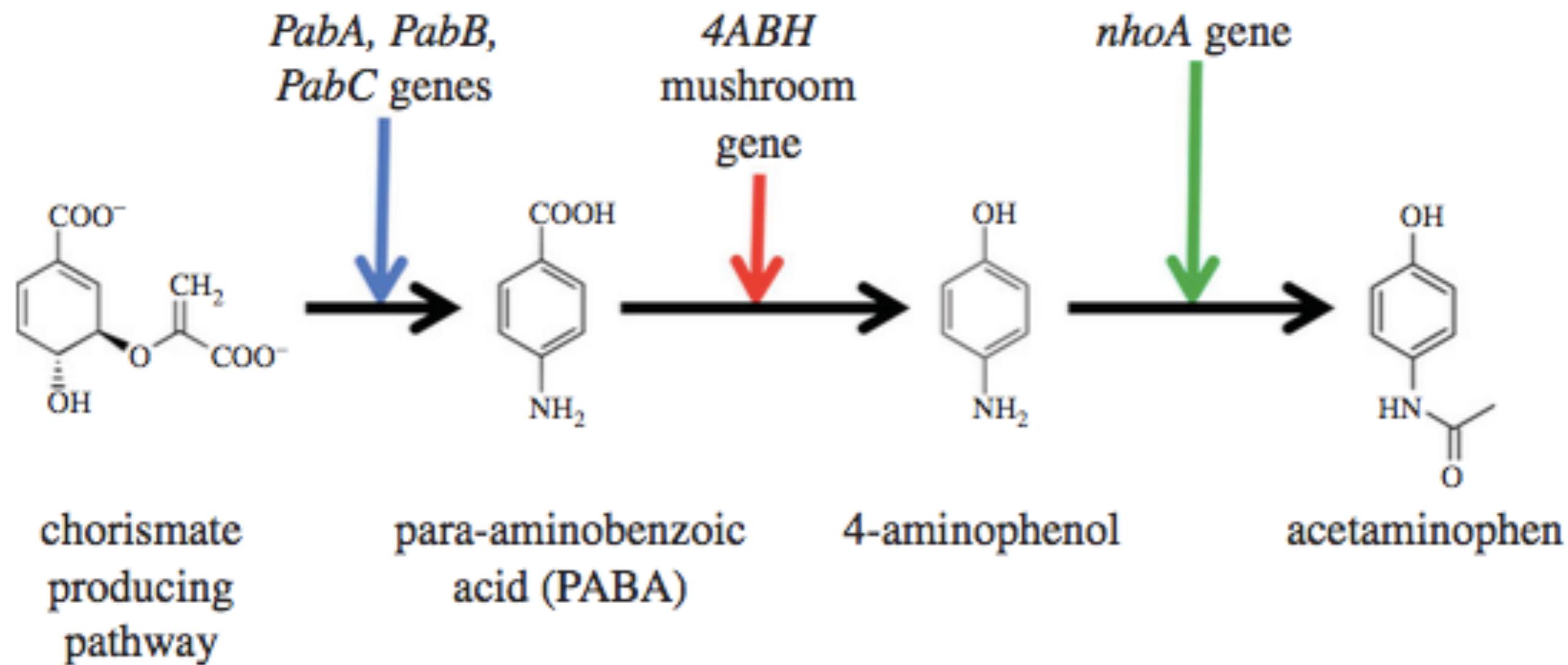
To do that, the stack contains many modules built from scratch in-house. Some of them: mine raw biochemical data, integrate heterogenous sources, learn rules of biochemistry, automatically clean bad data, mine patents, mine plain text, bioinformatic identification of enzymes with desired function.

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?



**some
rights
reserved**



Visualisation PyMol

