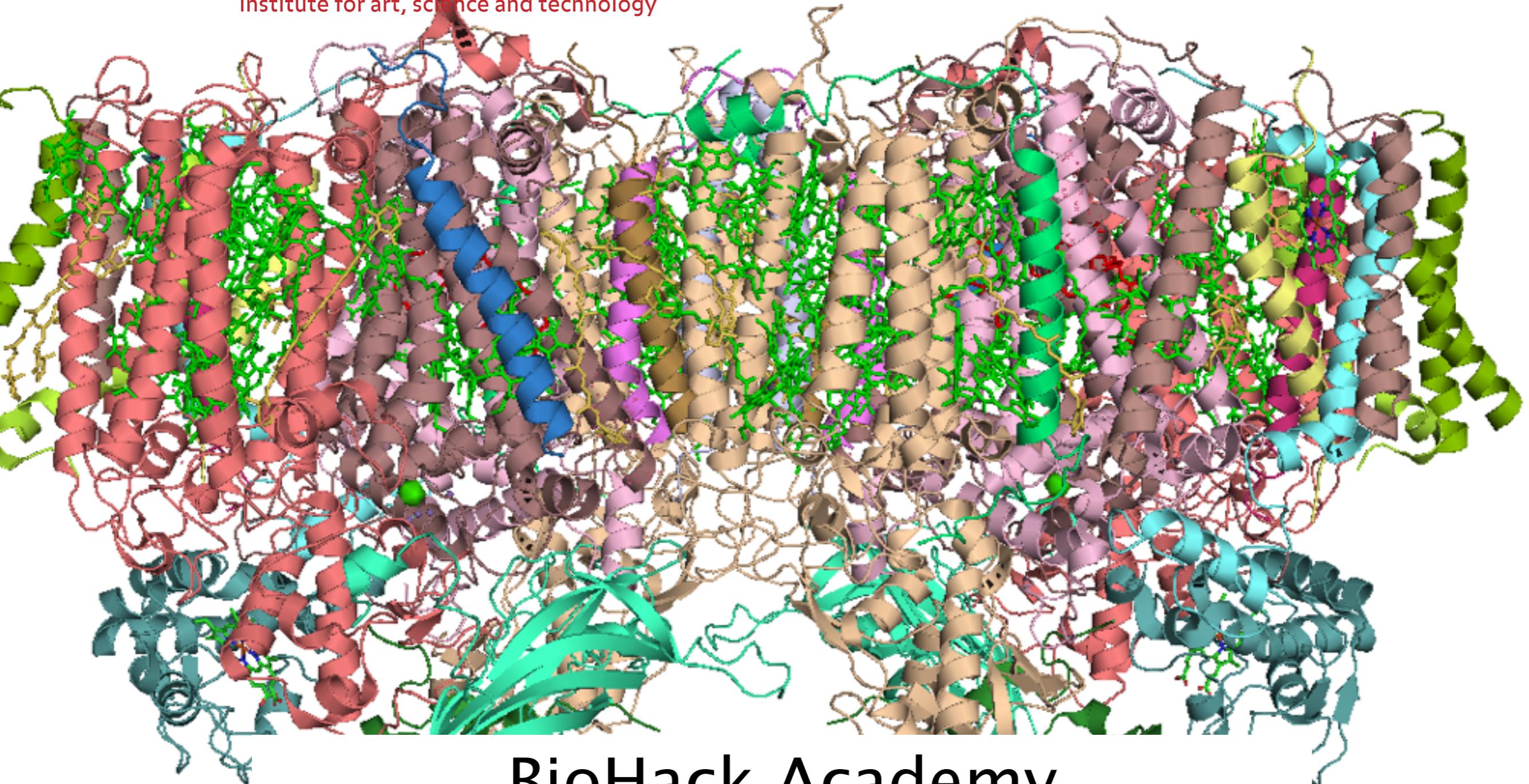




waag society

institute for art, science and technology



**BioHack Academy
Bio Informatics**

- US US

Neveu,Curtis CC-BY-SA 3.0



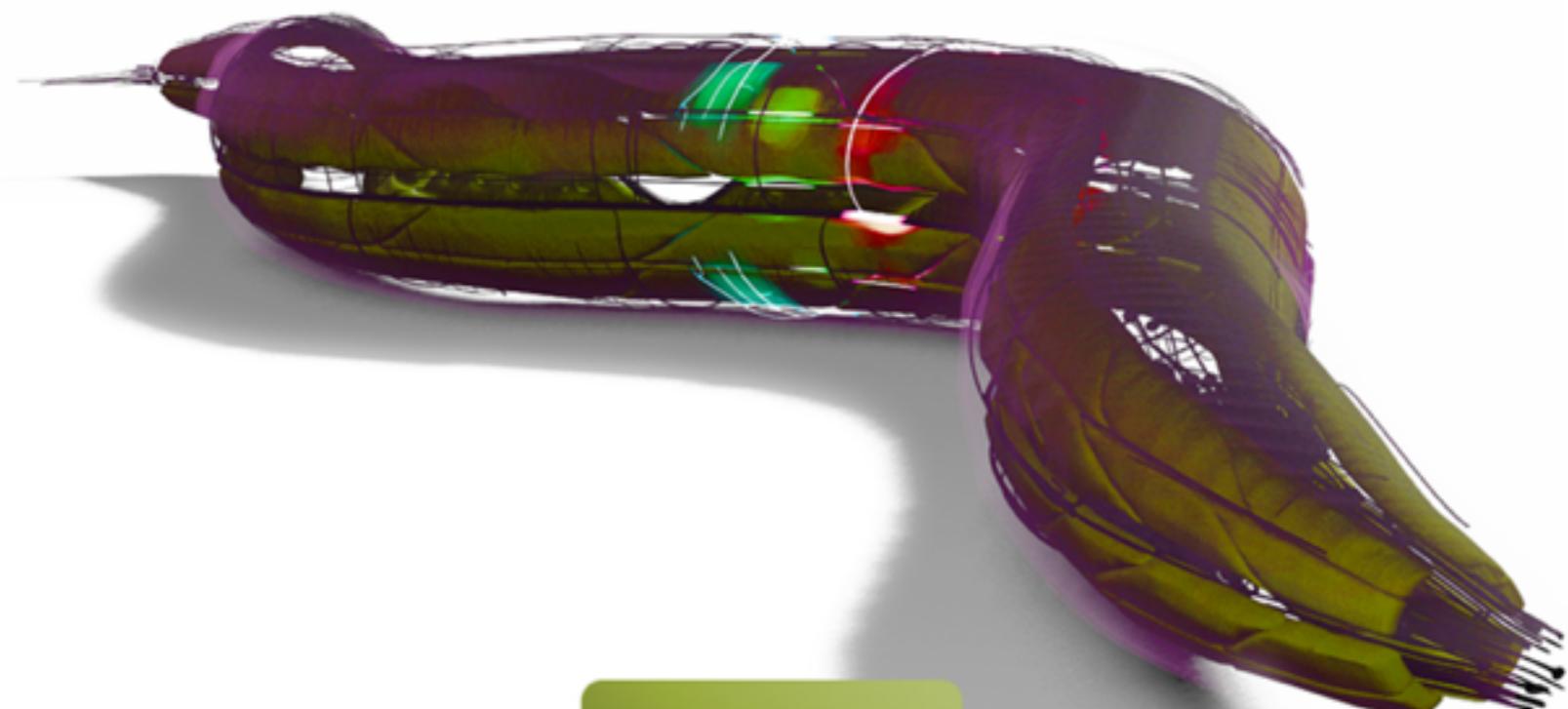
The “Omics”

- “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





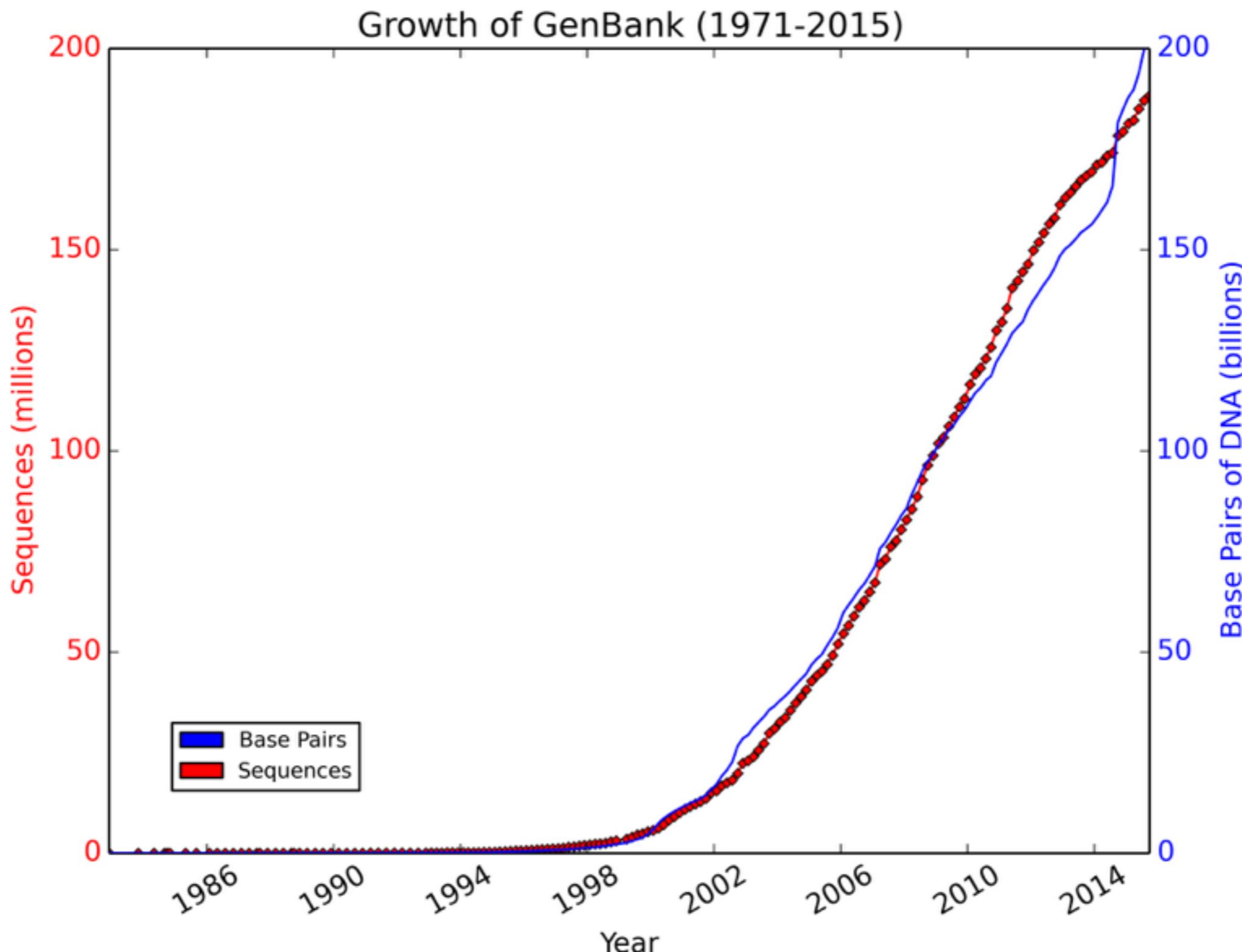
Genomics

- Functional genomics
- Metagenomics
- Personal Genomics
- Epigenomics





DNA database GenBank





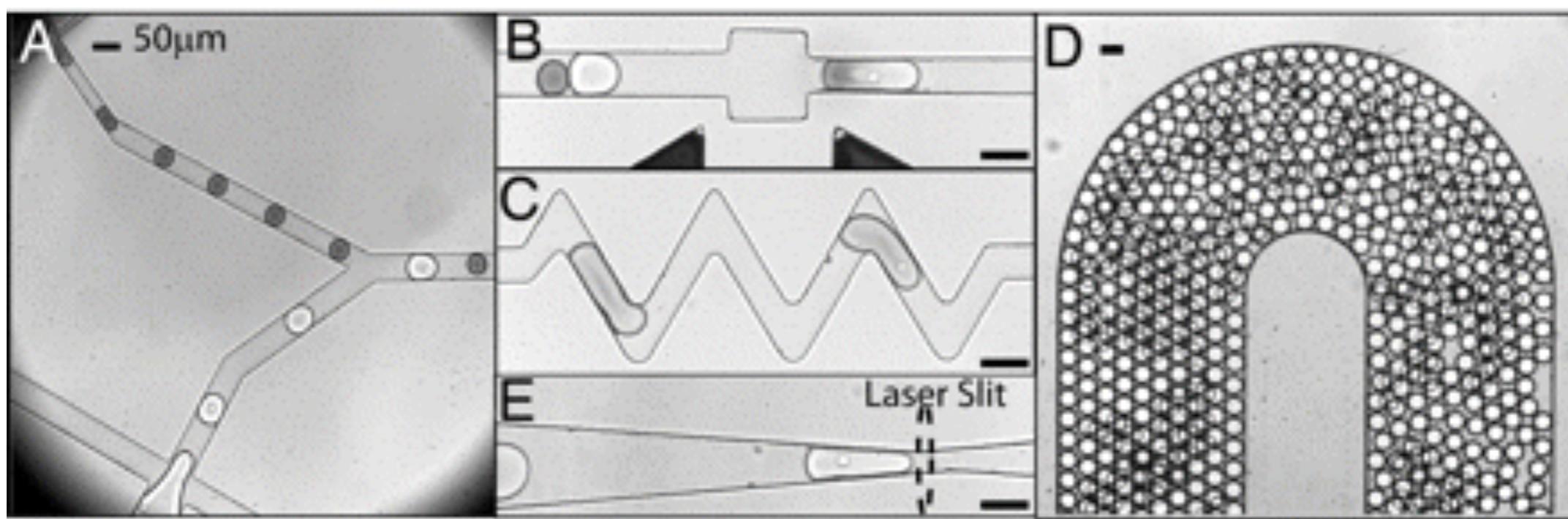
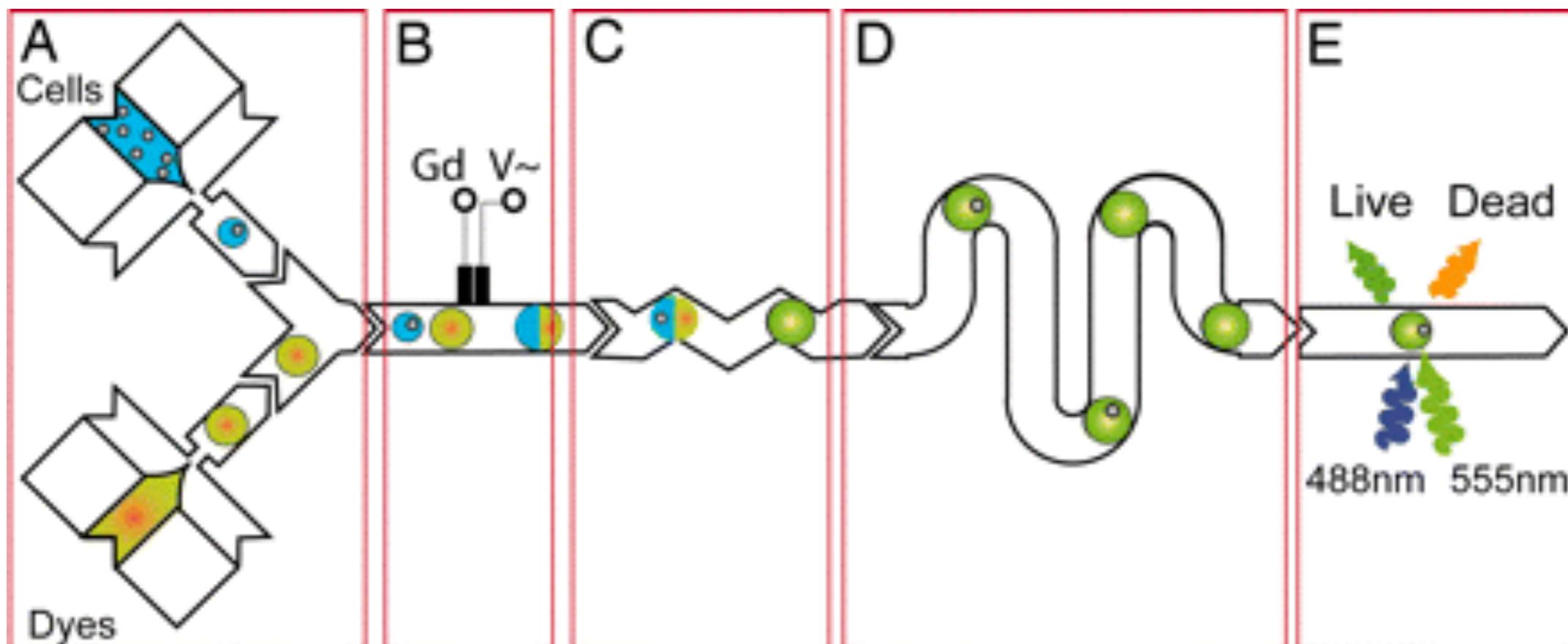
Drivers

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



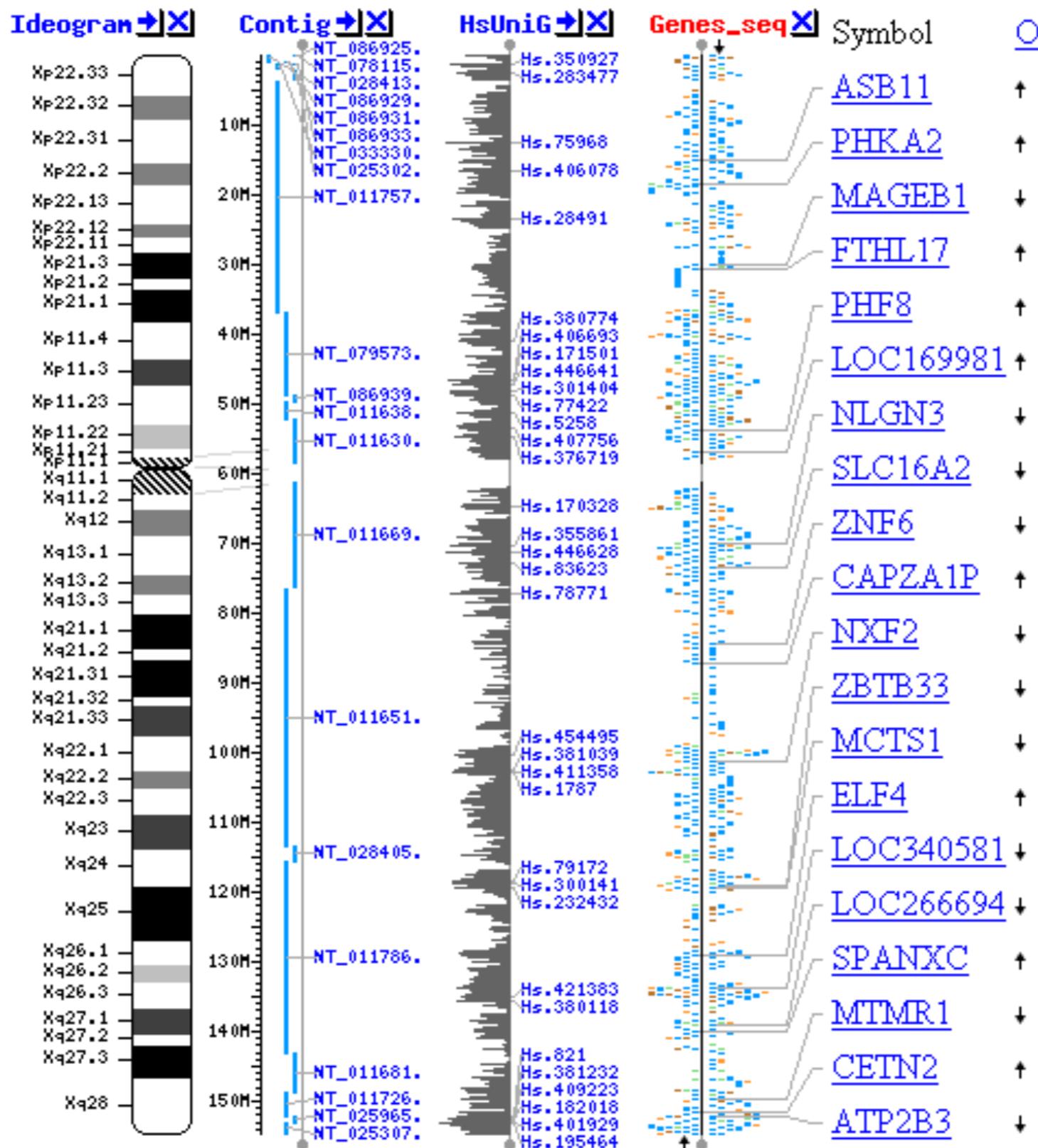
High Throughput Screening

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





Gene annotations





BLAST: Basic Local Alignment Search Tool

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

BLAST® Basic Local Alignment Search Tool

Home Recent Results Saved Strategies Help

NCBI/ BLAST/ blastn suite Standard Nucleotide BLAST

blastn blastp blastx tblastn tblastx

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s) [more...](#) Query subrange [?](#)

From
To

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—completions 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



Sequence Alignment

14	SIKLWPPSQTTRLLLVERMANNLST..PSIFTRK..YGSLSKEEAREN AKQIEEVACSTANQ.....HYEKEPDGDGGSAVQLYAKECSKLILEVLK	101
13	SIKLWPPSESTRIMLVDRMTNNLST..ESIFSRK..YRLLGKQEAHEN AKTIEELCFALADE.....HFREEPDGDGSSAVQLYAKETSKMMLEV LK	100
23	VFKLWPPSQGTREAVRQKMAKLSS..ACFESQS..FARIELADAQE HARAIEEVAFGAAQE.....ADSGGDKTGS AVVMVYAKHASKLMLET LR	109
13	SVKLWPPGQSTRMLVERMTKNFIT..PSFISRK..YGLLSKEEAEED AKKIEEVAFAAANQ.....HYEKQPDGDGSSAVQIYAKESSRLMLEV LK	100
30	SFSIWPPPTQRTRDAVVRRLVDTLGG..DTILCKR..YGAVPAADA EPAARGIEAEAFDAAAA..SGEAAAATASVEEGIKALQLYSKEVS RRLLDFVK	120
44	SLSIWPPSQRTRDAVVRRLVQTLVA..PSILSQR..YGAVPEAEAG RAAAA AVEAEAYAAYTES.SSAAAAPASVEDGI EVLQAYSKEVS RRLLELAK	135
56	SFSIWPPPTQRTRDAIIISRLIETLST..TSVLSKR..YGTIPKEE ASEASRRRIEEEAFSGAST.....VASSEKDGL EVLQLYSKEISKRMLET VK	141
29	SFAVWPPTRRTRDAVVRRLVAVL SGDTTTALRKRYRYGAVPAADAERA ARAVEAQAFDAASA...SSSSSSVEDGI ETLQLYSREVSNRLLAF VR	121
13	SIKLWPPSESTRMLVERMTDNLSS..VSFFSRK..YGLLSKEEAA ENAKRIEETAFLAAND.....HEAKEPNLDDSSVVQFYAREASK LMLEALK	100
57	SLRIWPPTQKTRDAVLNRLIETLST..ESILSKR..YGT LKSDDATT VAKLIEEEAYGVASN.....AVSSDDDGI KILELYSKEISKRMLES VK	142
25	NYSIWPPKQRTRDAVKNRLIETLST..PSVLT KR..YGTMSADE ASAAAIQIEDEAF SVANA.....SSSTSNDNV TILEVY SKEISKRM IET VK	110
28	SFKIWPPPTQRTR EAVV RRLVETL TS..QSVLSKR..YGV IPEEDAT SAARI IEEEAF SVASV. ASA A AST GG RPE DE WIE EV LHI Y SQ E IX Q RV V ES AK	119
25	SFSIWPPPTQRTRDAVINRLIESLST..PSILSKR..YGTLPQD EASE TARL IEEEAF AAAGS.....TASDADD GIEILQV Y SKE ISK RM ID TV K	110
14	SVKMWPPSKSTRMLVERMTKNITT..PSIFSRK..YGLLSVE EA EQDAK RIED LA FAT ANK.....HFQNEPDGD GT SAV HV Y AKE ESS KL ML DV IK	101
13	SIKLWPPSLPTRKALIERITNNFSS..KTIFTEK..YGS LTKD QAT ENA KRIED IAF STANQ.....QFEREPDG DG GS AV QLY AKE CS KL L I E VL K	100
48	SLSIWPPPTQRTRDAVITRLIETLSS..PSVLSKR..YGT ISH DEA ESA ARR IEDEAF GV ANT.....ATS AED DG LEIL QLY SKE IS RR ML DT VK	133



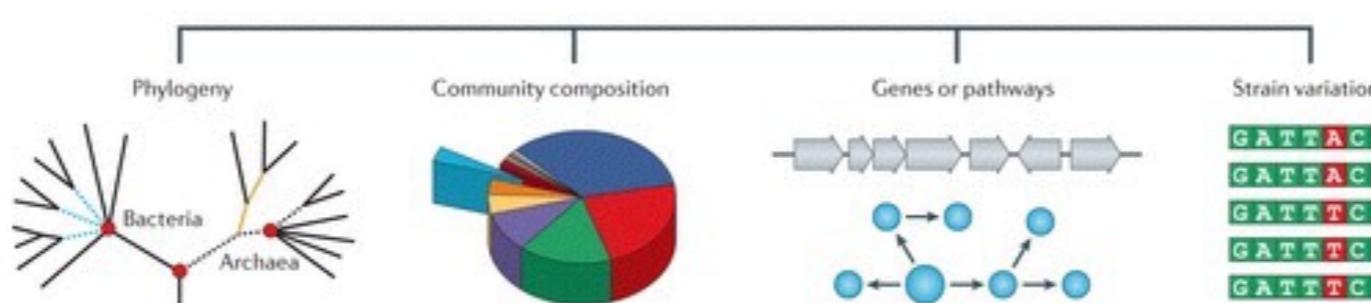
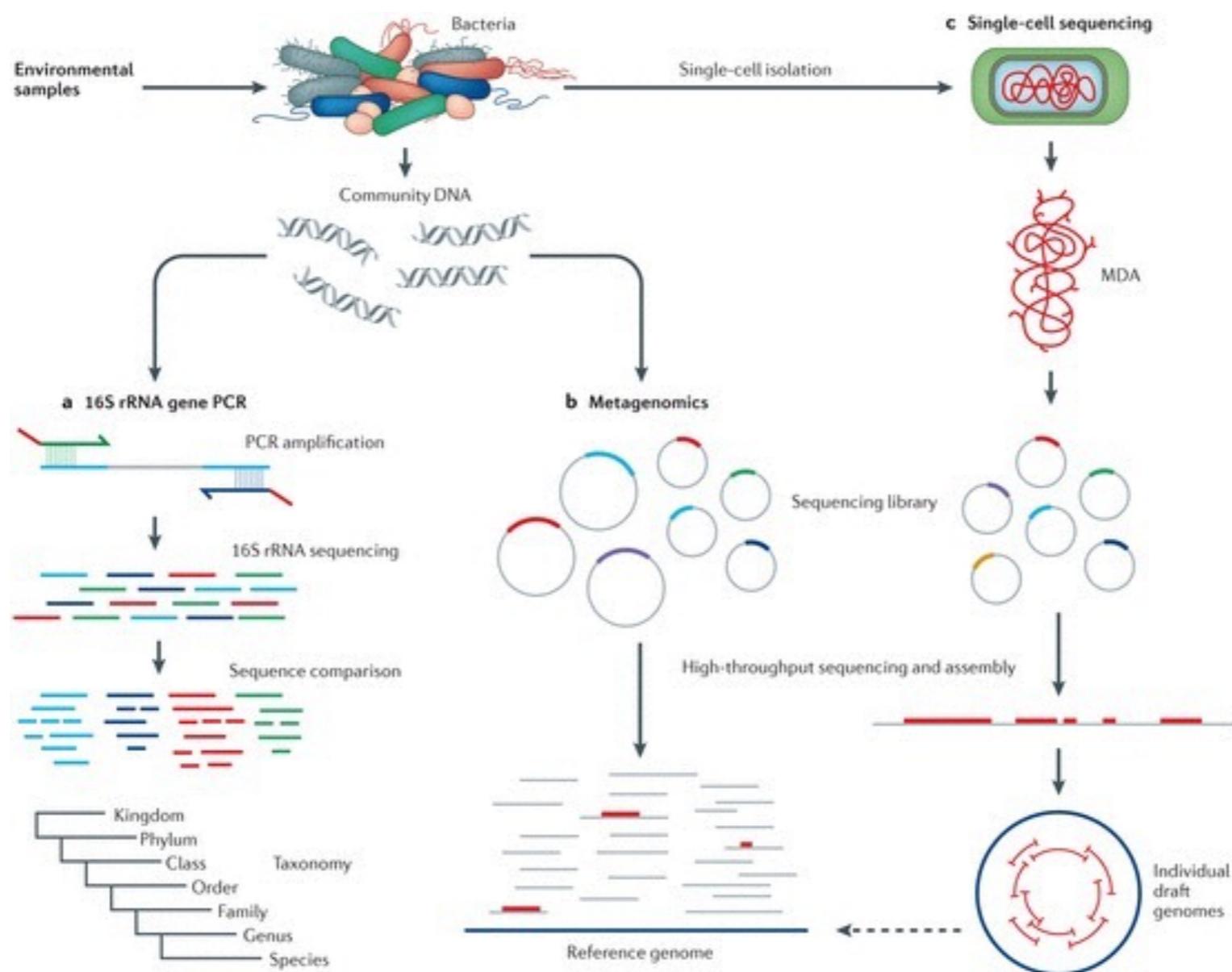
Scoring Matrix BLOSUM

(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





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](#) [cart 0](#)

 **23andMe**

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

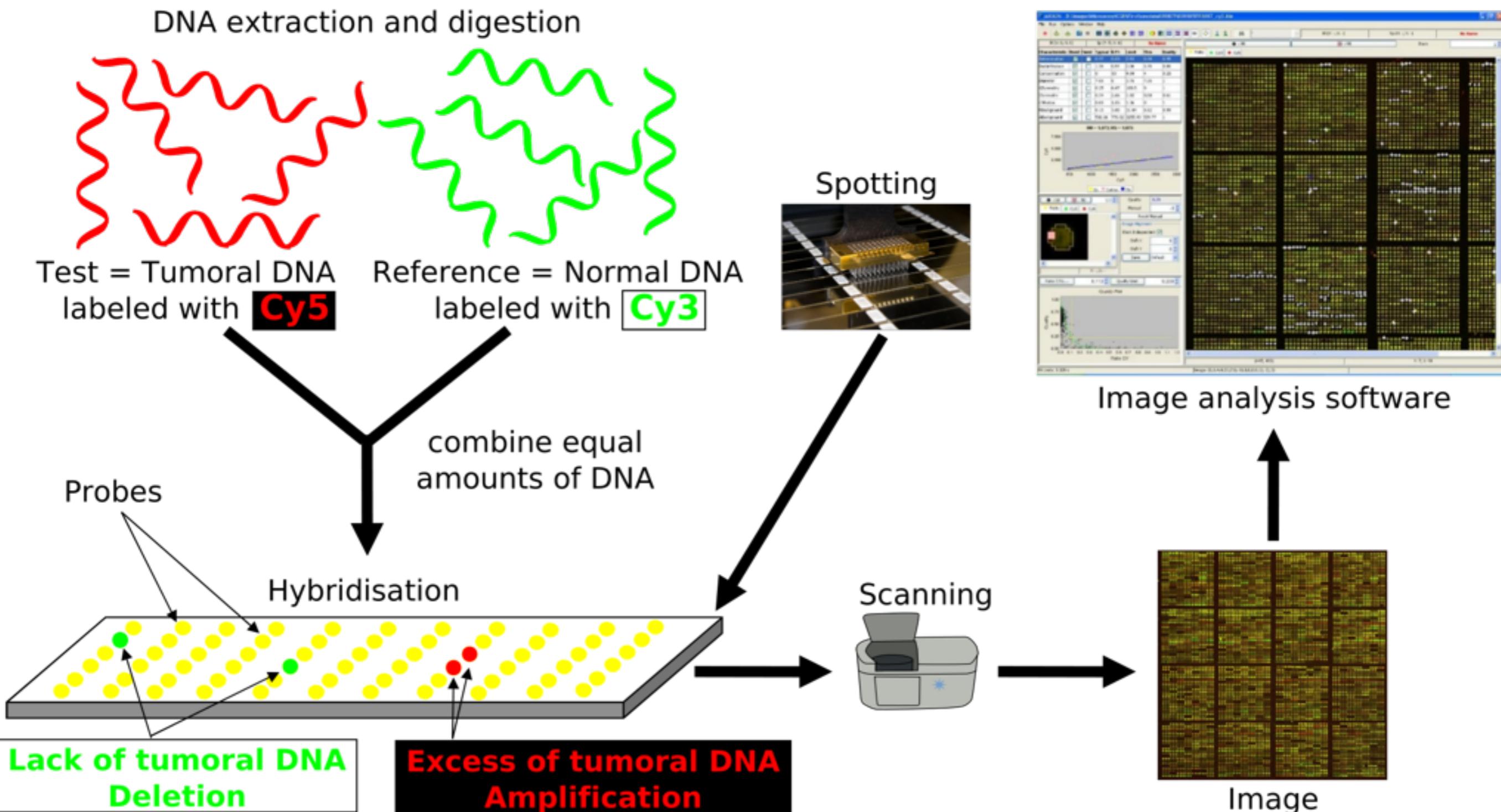
[order now](#) **€169**
shipping included

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 Expression Analysis





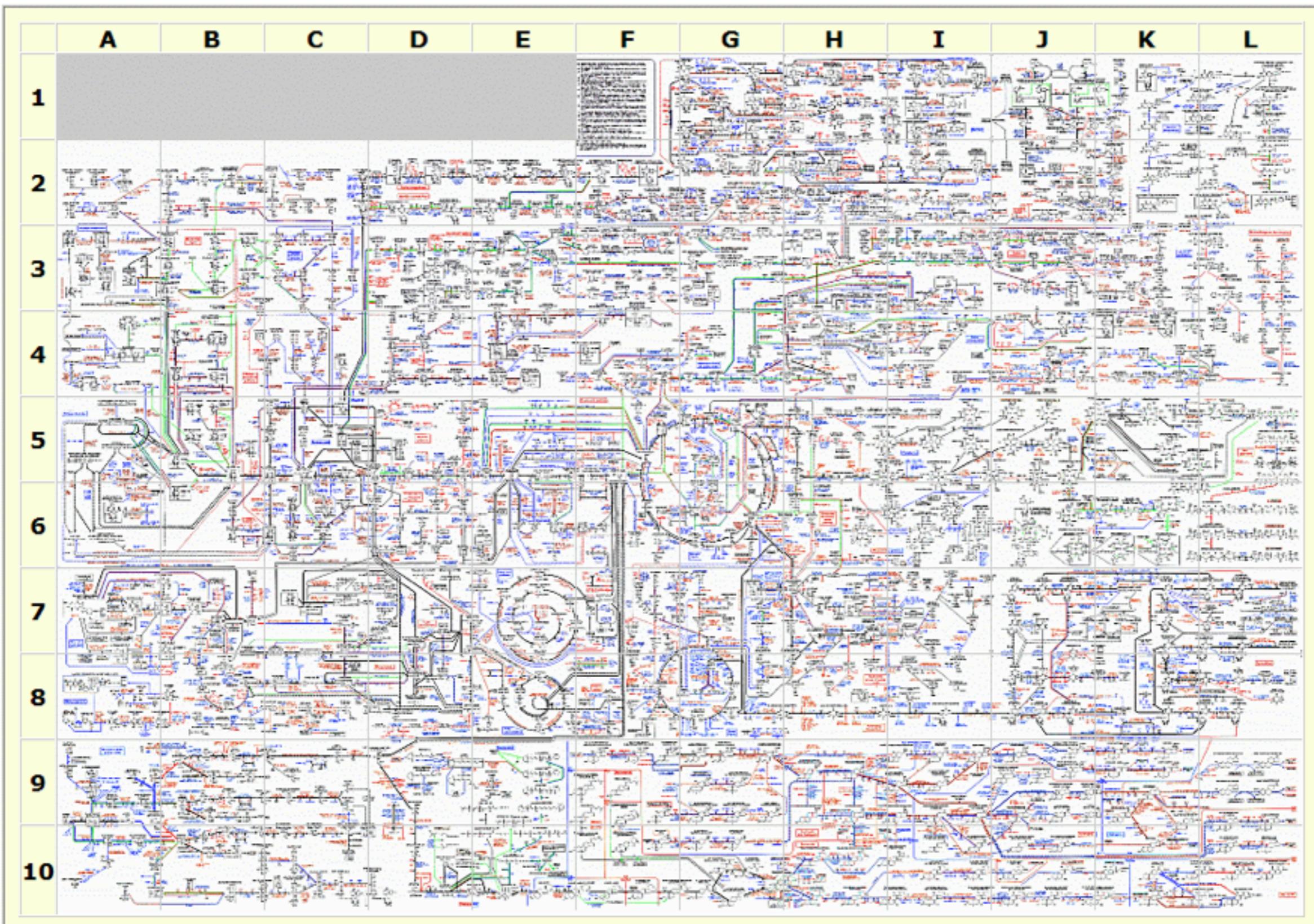
waag society

institute for art, science and technology

Proteins

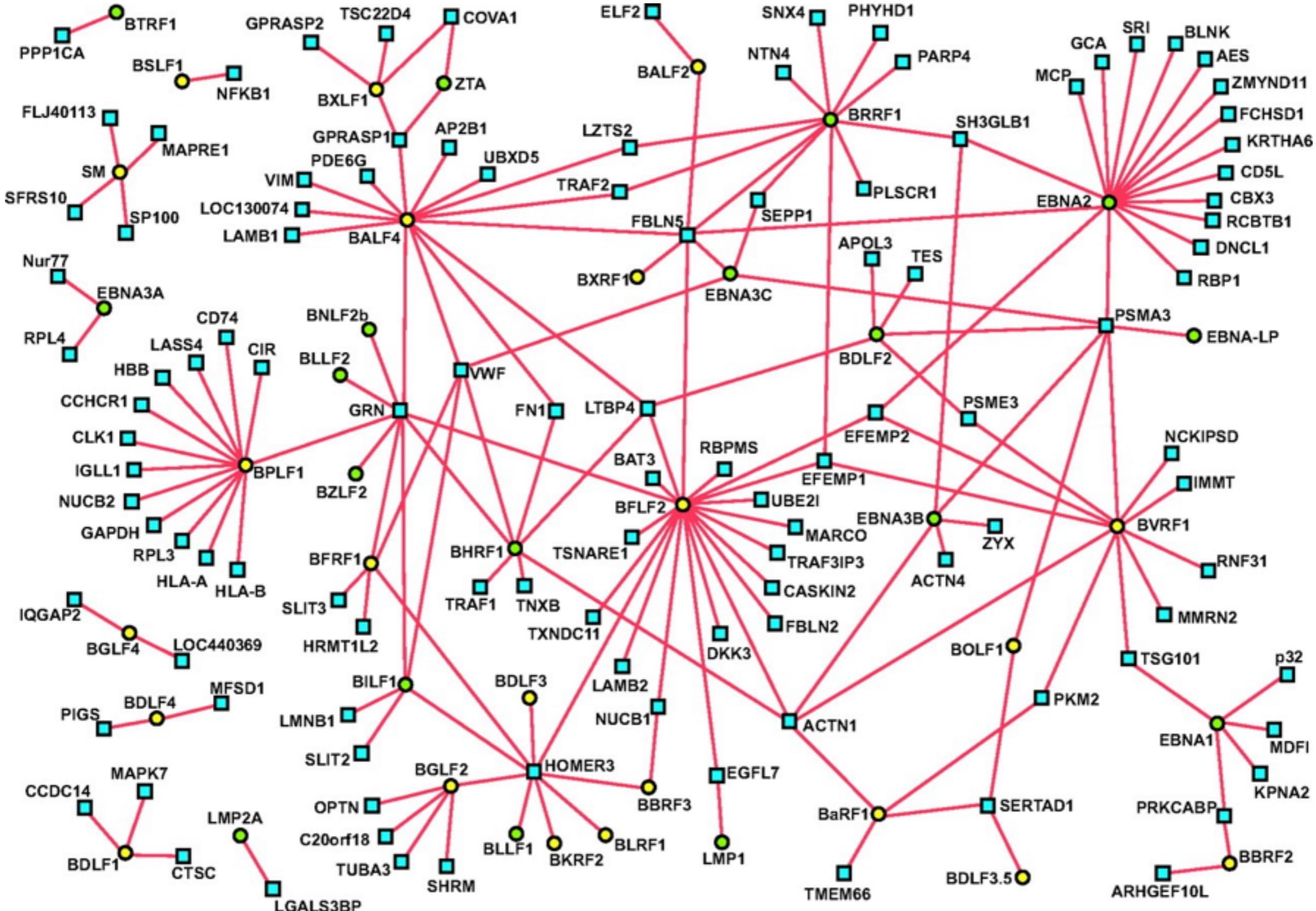


Biochemical Pathways of the Cell





Protein interaction mapping: MS





What is this?

Simon Eugstar - CC-BY-SA 3.0



Debstart - CC-BY-SA 3.0





NMR Machines

MartinSaunders

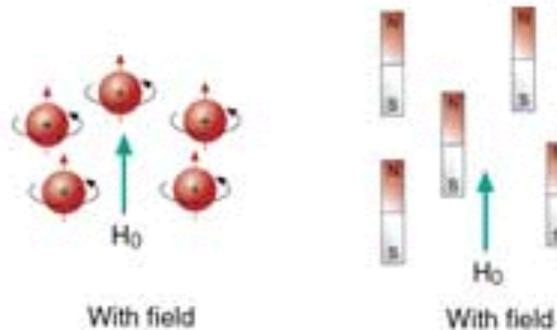
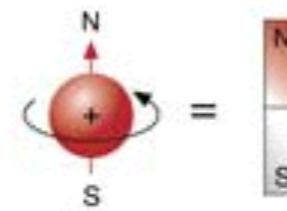
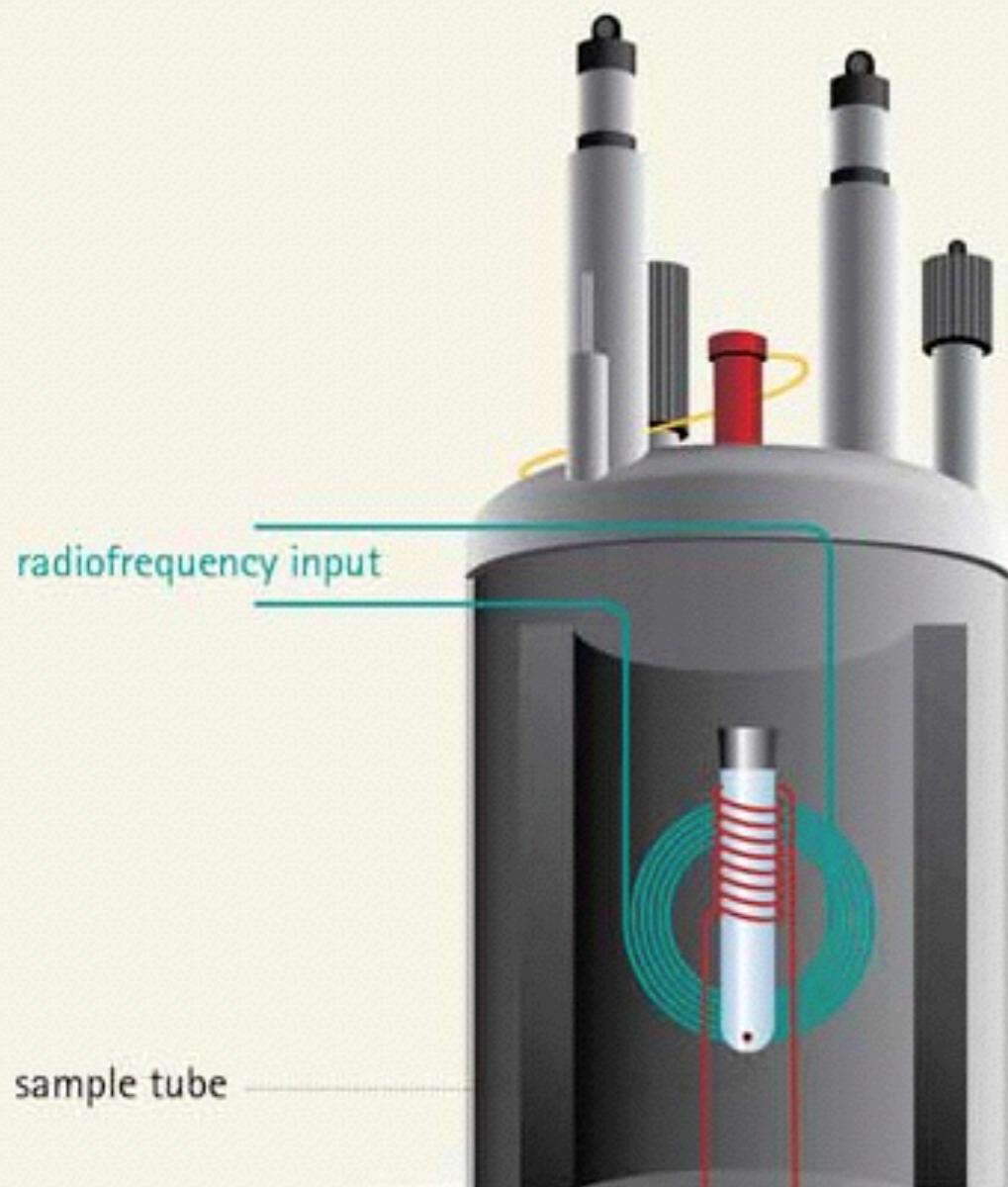


Public Domain





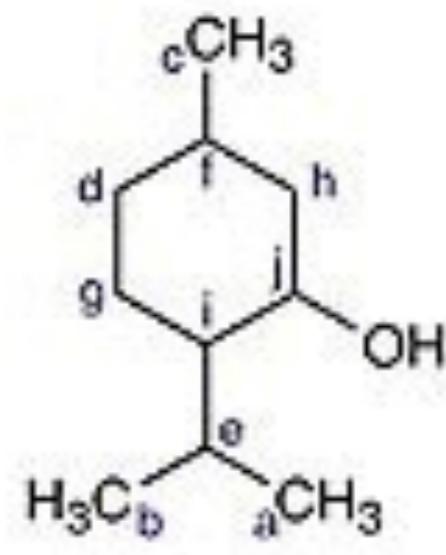
NMR principles



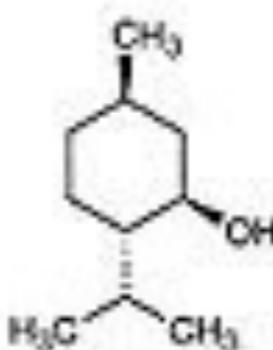


NMR Spectrum

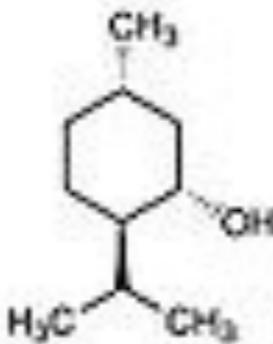
1D PROTON SPECTRUM



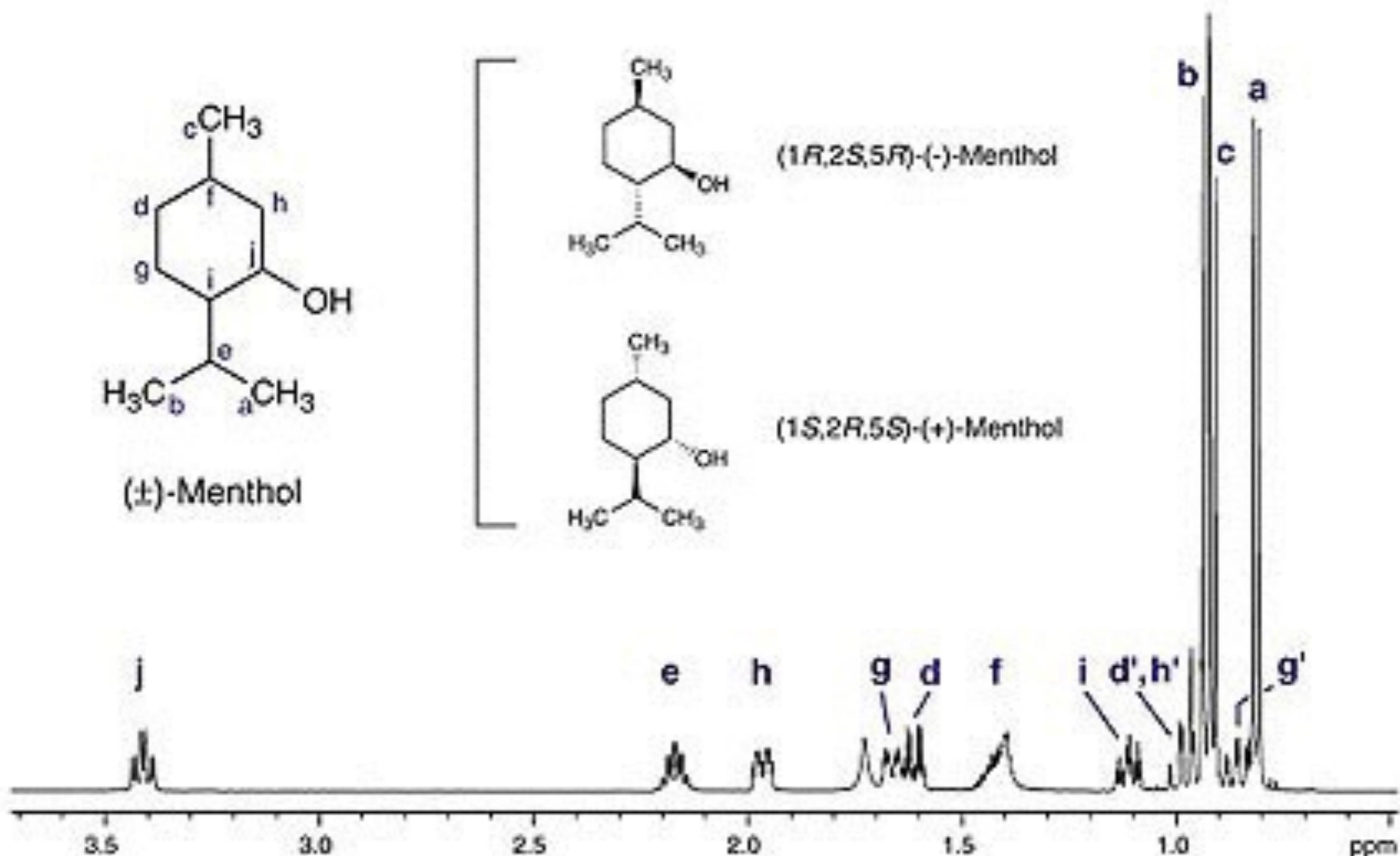
(\pm)-Menthol



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



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

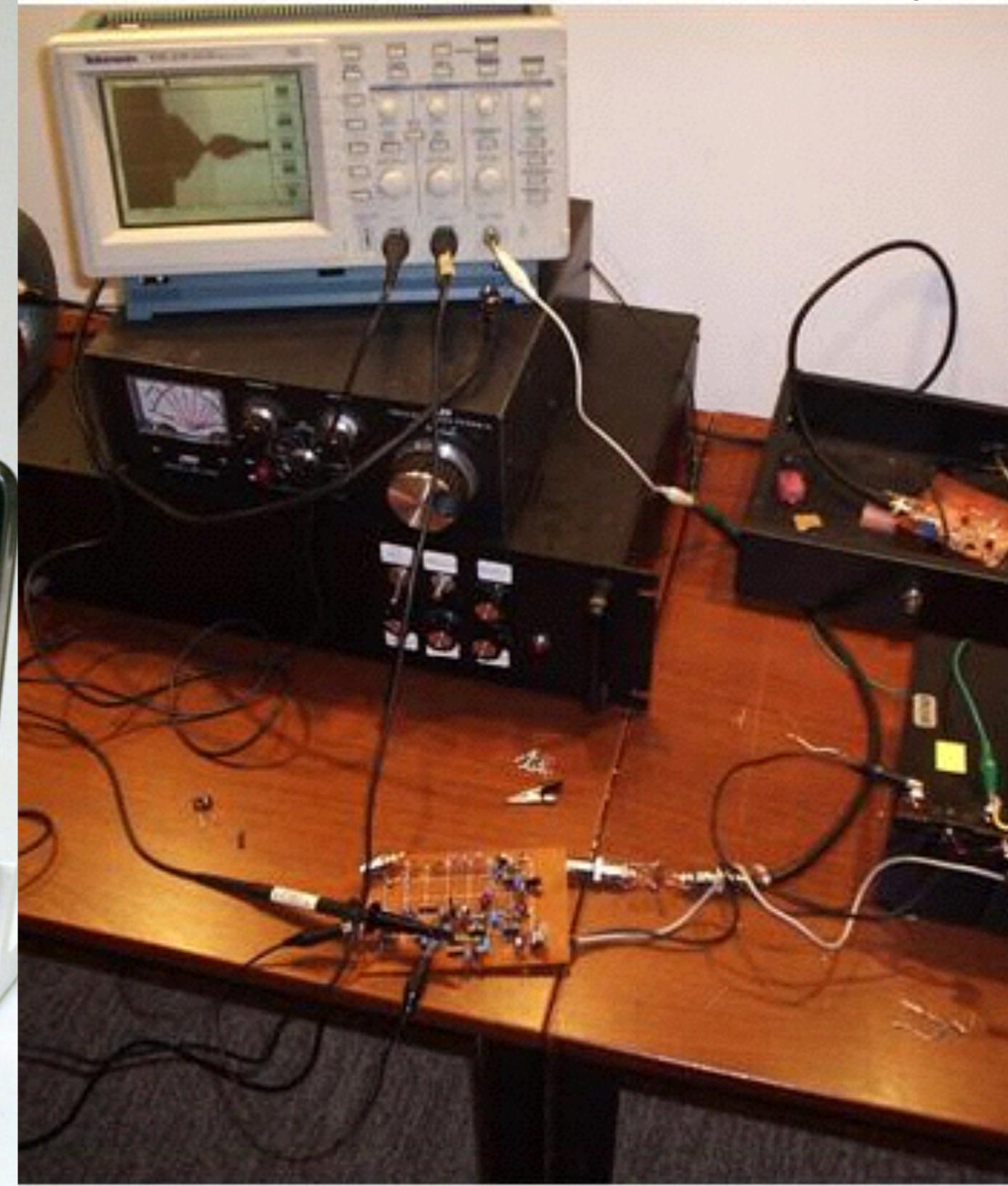
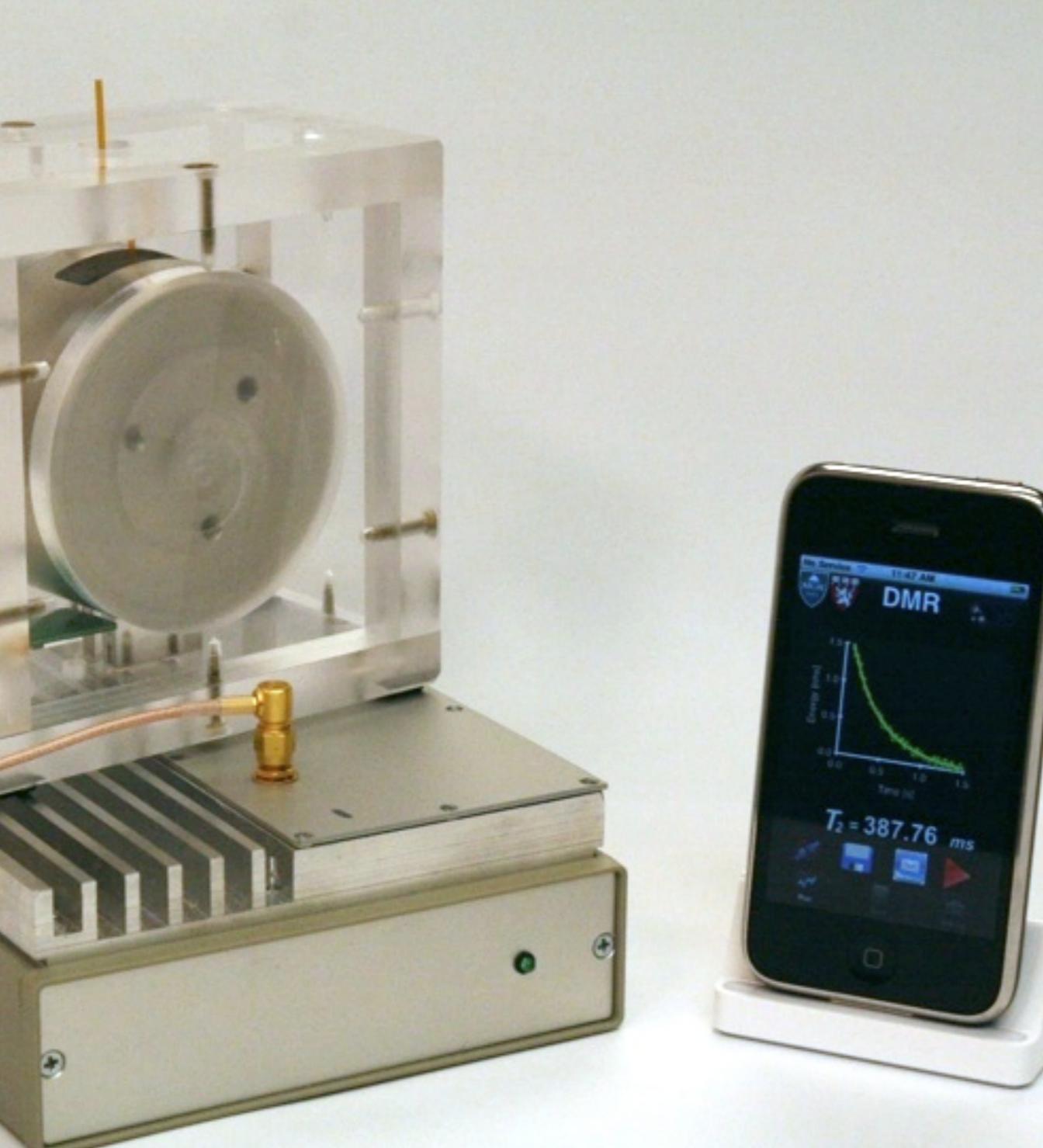




DIY NMR?

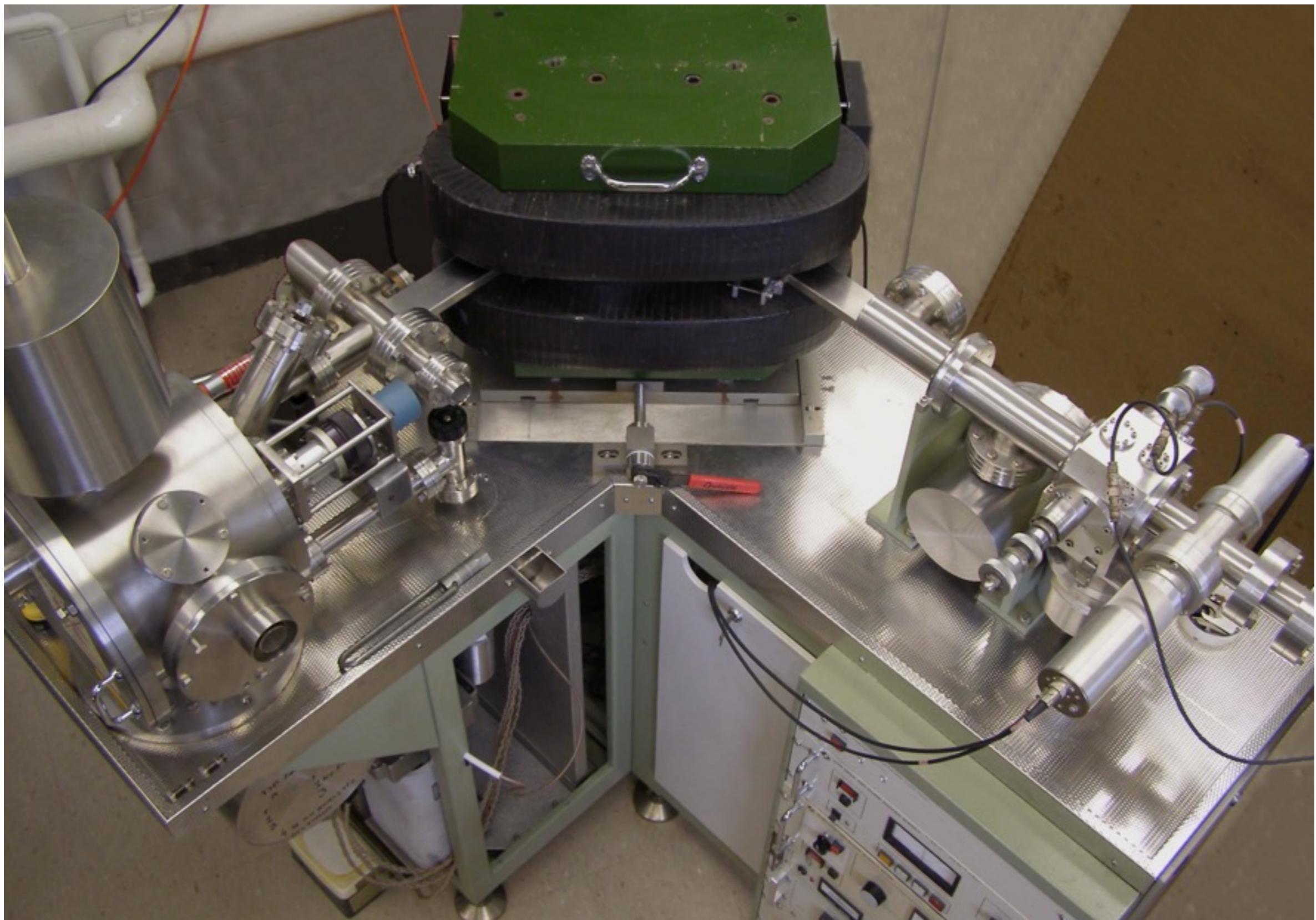
<http://www.seas.harvard.edu>

conspiracyoflight.com



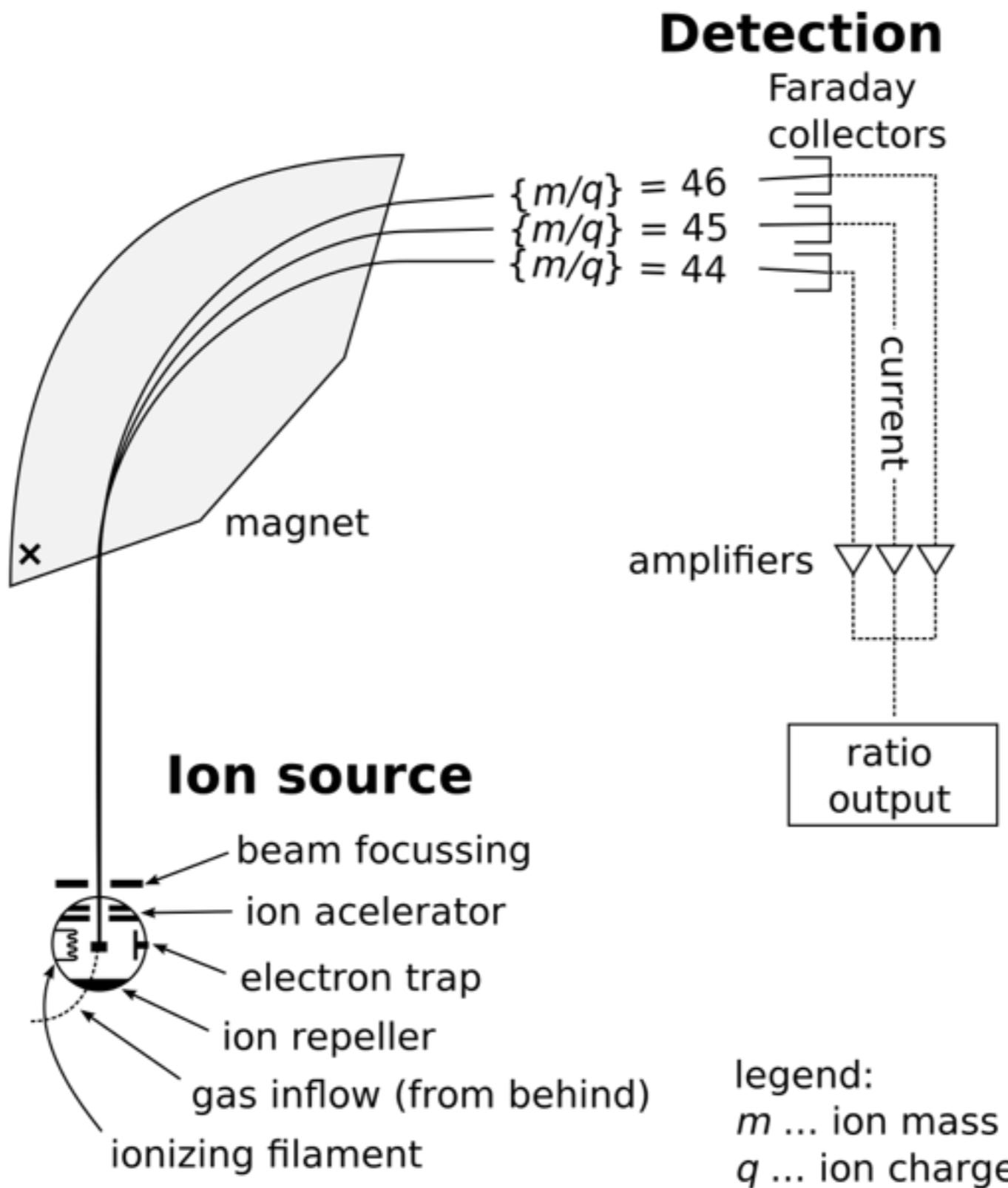


Mass Spectrometer



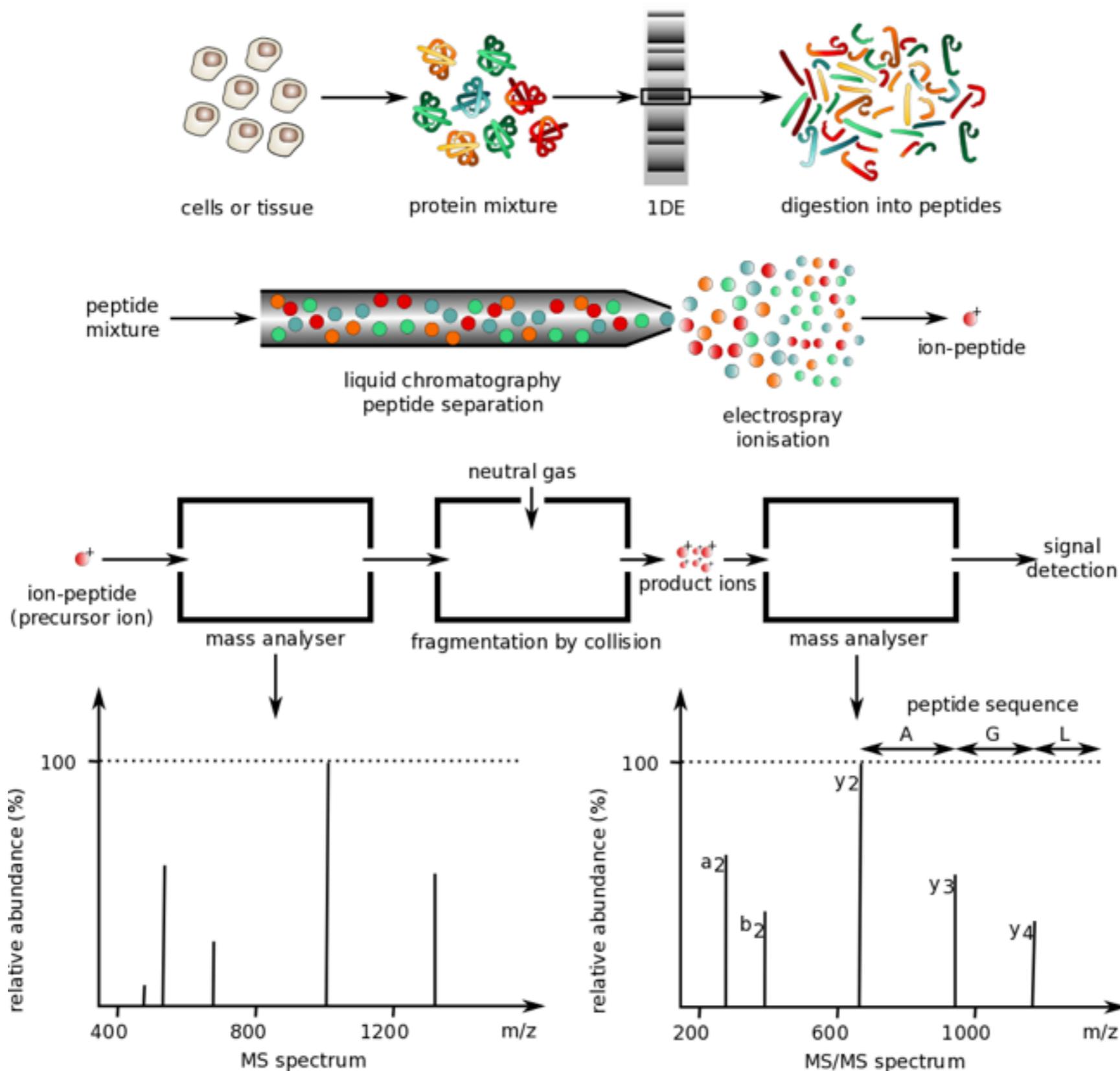


Simple Mass Spectrometry



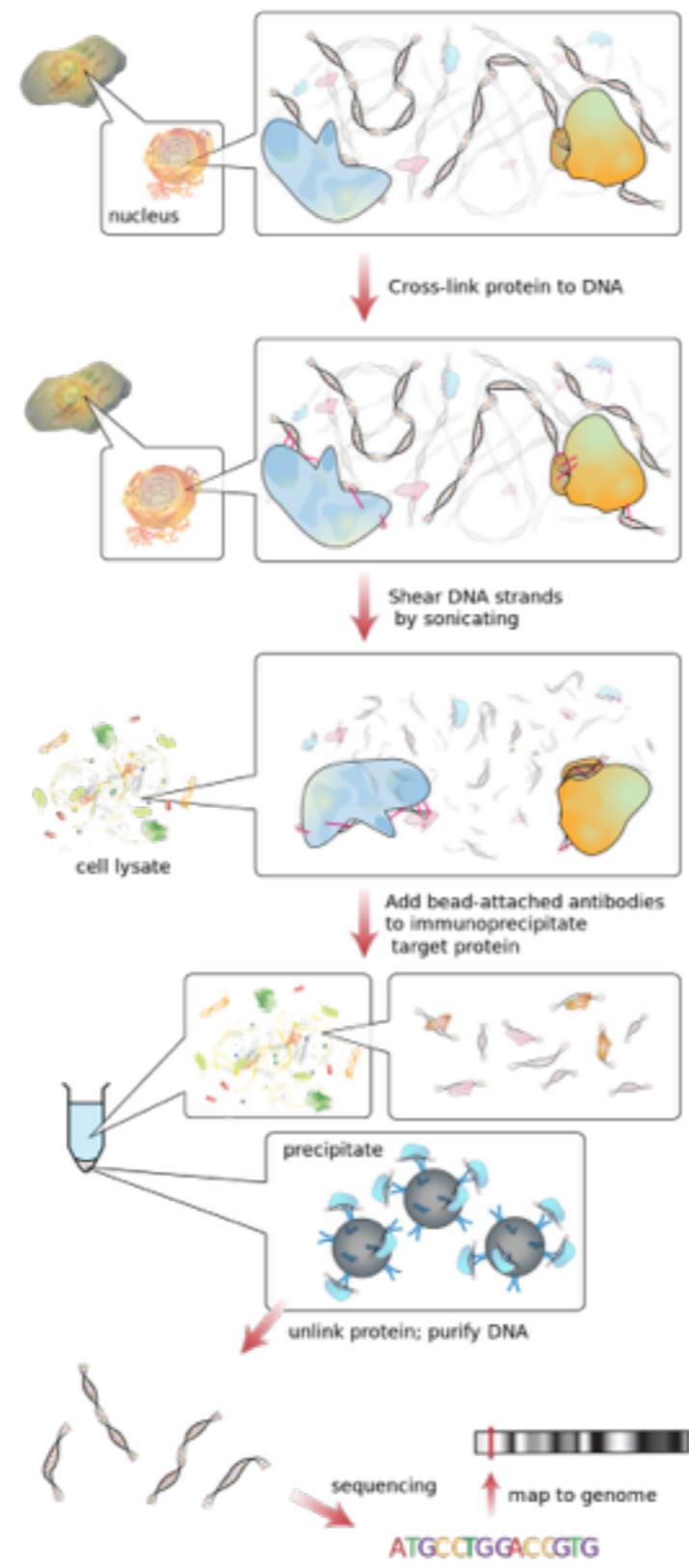


Mass Spectrometry



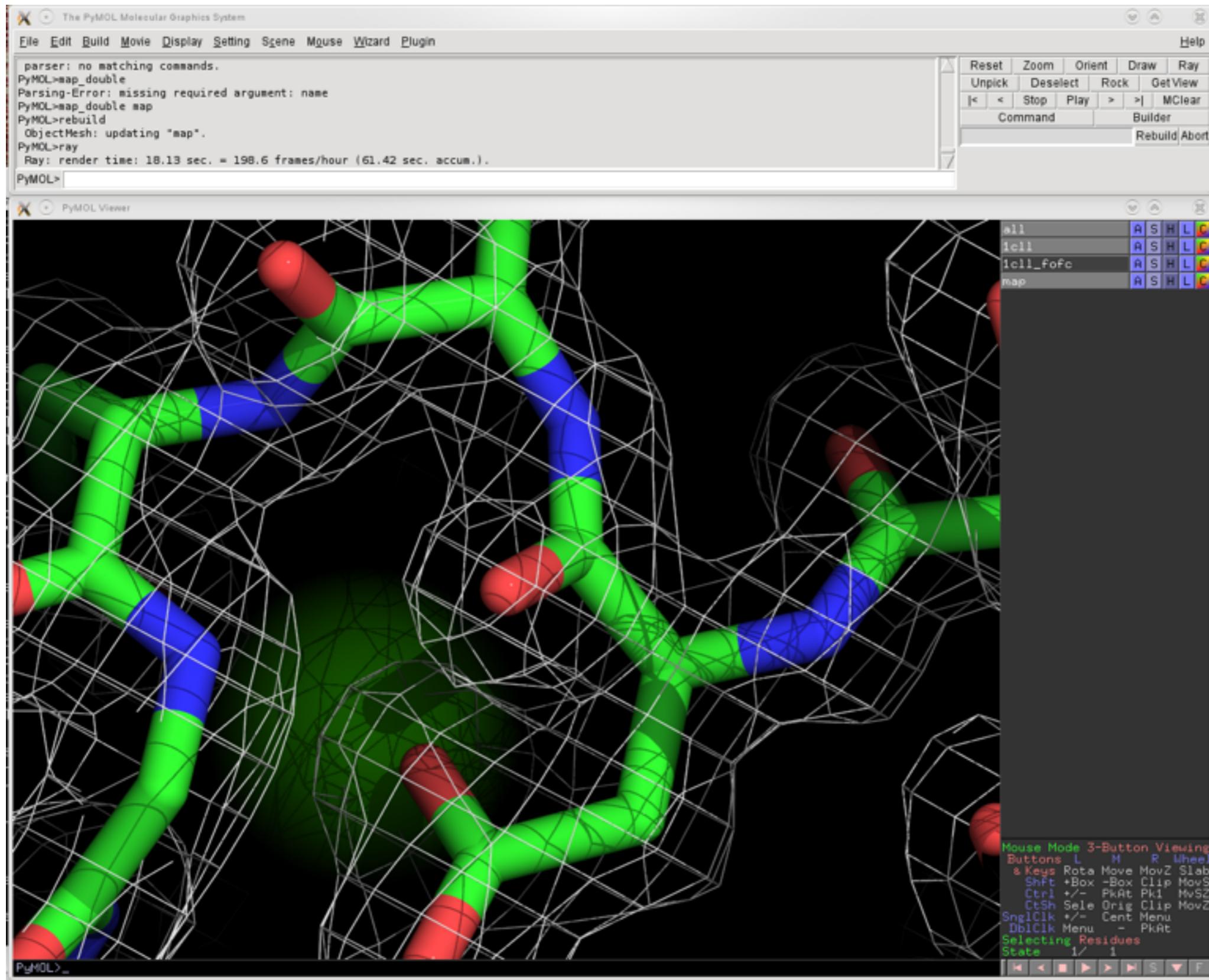


Chromatin ImmunoPrecipitation ChIP





Visualisation PyMol





Issues

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



**some
rights
reserved**