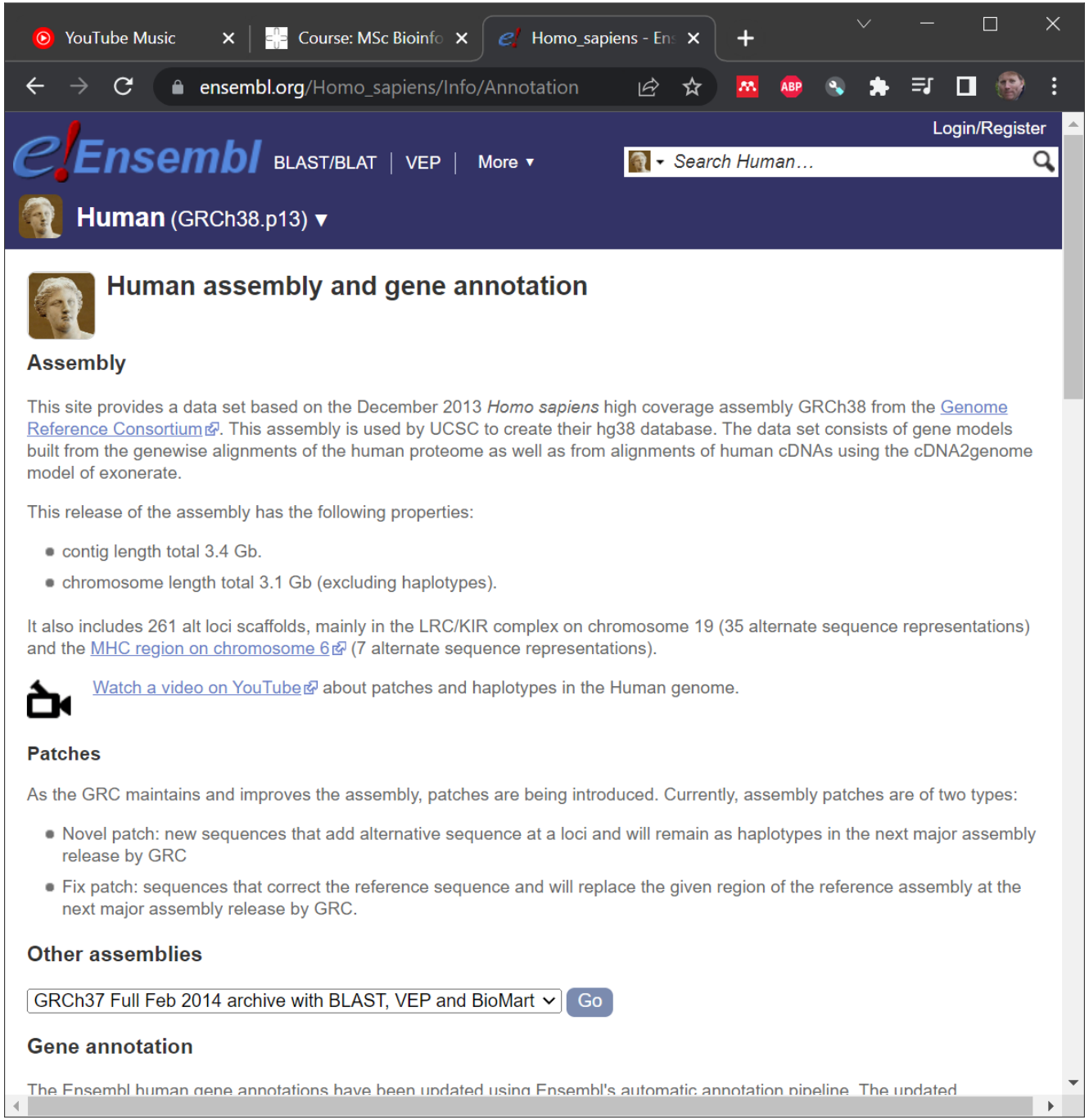
Queen Mary
University of London

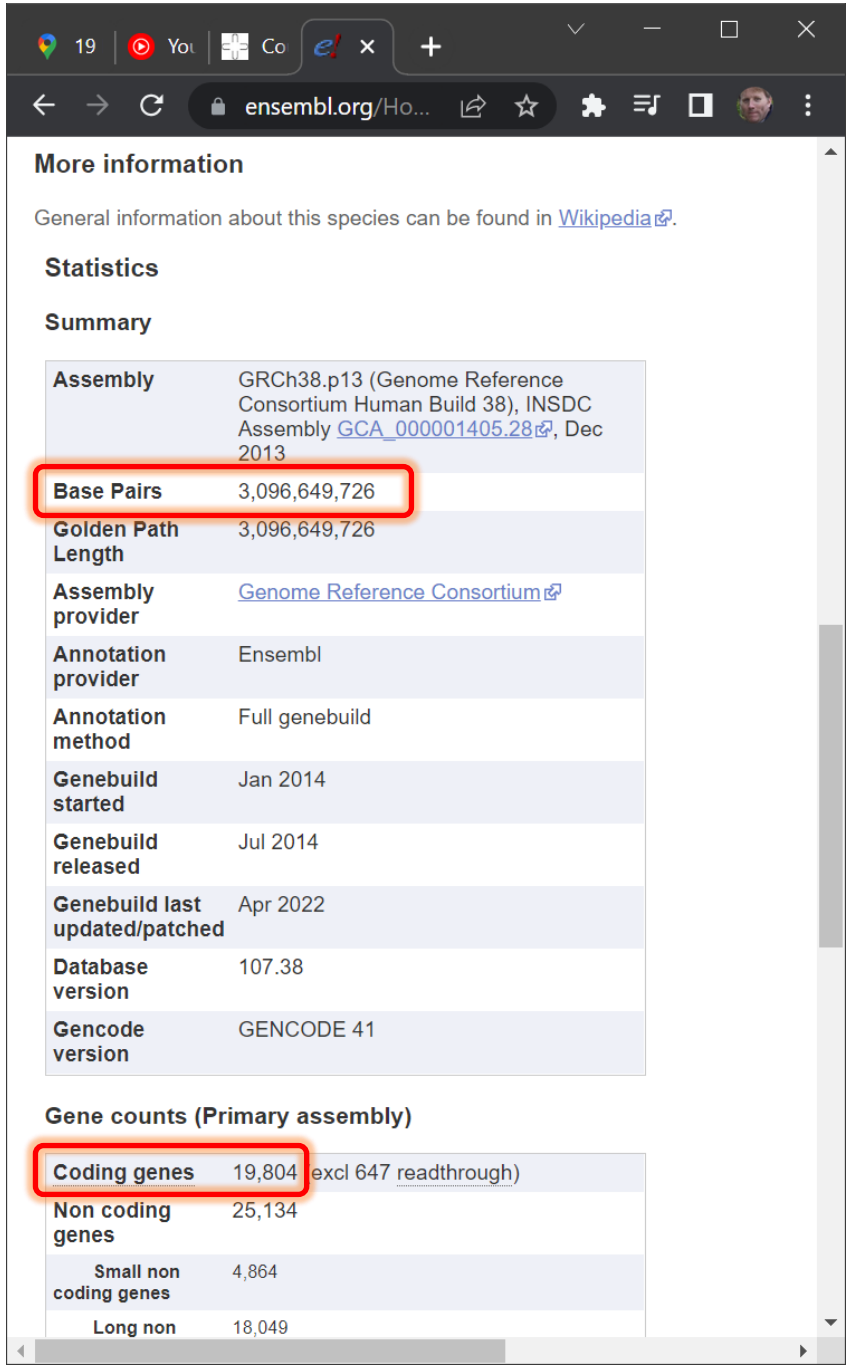
Welcome to MSc Bioinformatics

Prof Conrad Bessant
MSc Bioinformatics Programme Leader
c.bessant@qmul.ac.uk

A little biology …

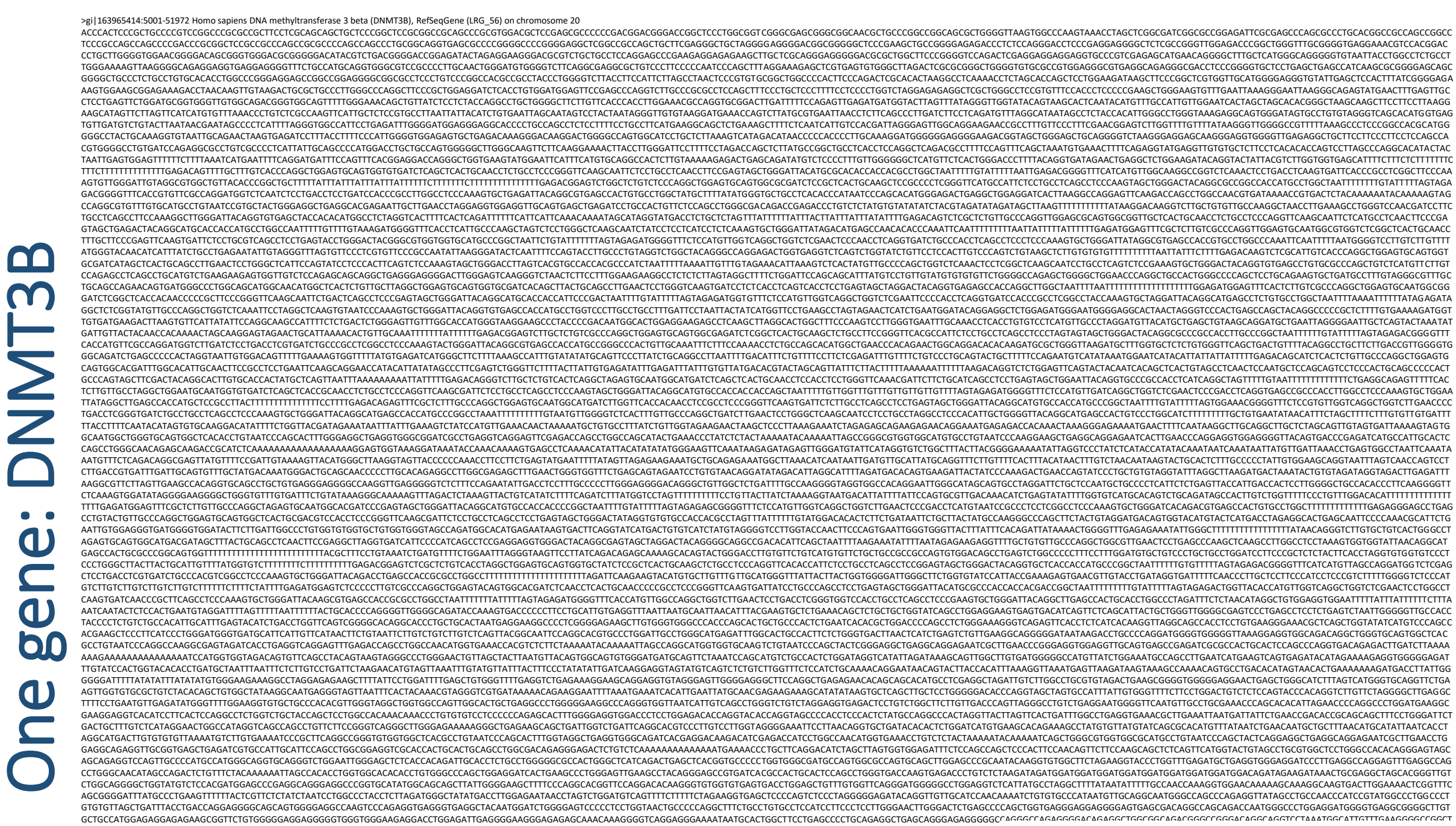
The human genome





One gene: DNMT3B

>gi|163965414:5001-51972 Homo sapiens DNA methyltransferase 3 beta (DNMT3B), RefSeqGene (LRG_56) on chromosome 20



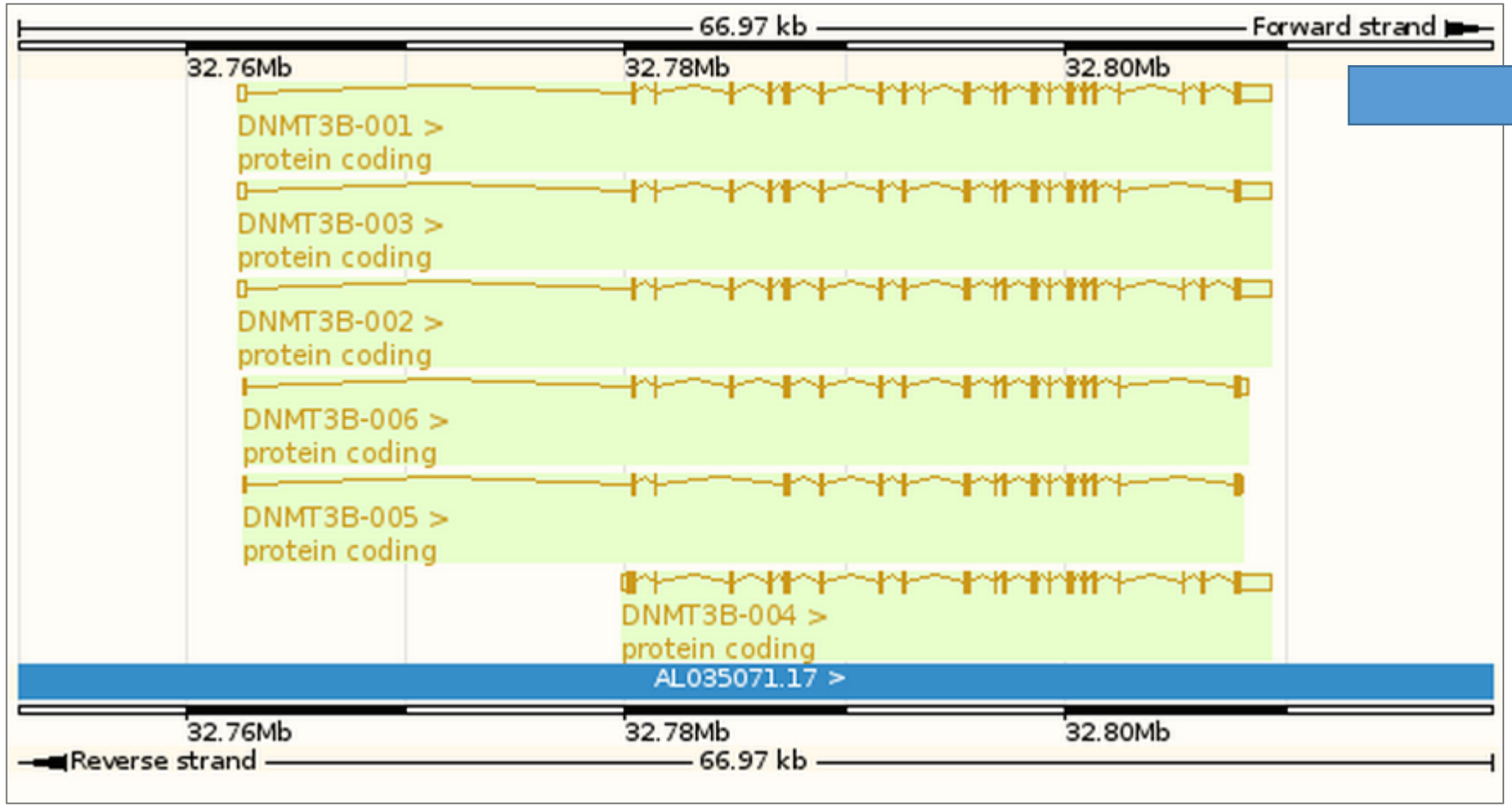
Continues for a couple more pages …

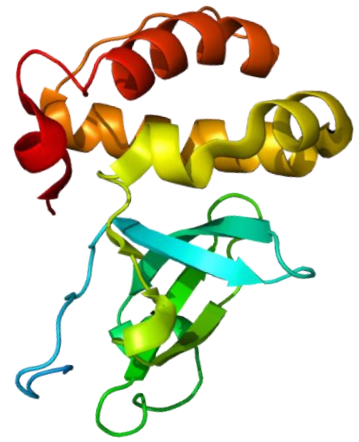
DNMT3B: Transcript

ATGGAACCAAGTCCTGAGCCTCCAAGCTTGGAAAGCATGAAGGGAGACACCAGGCATCTCAATGGAGAGG
AGGACGCCGGCGGGAGGGAAGACTCGATCCTCGTCAACGGGGCCTGCAGCGACCAGTCCTCCGACTCGCC
CCCAATCCTGGAGGCTATCCGCACCCCGGAGATCAGAGGCCGAAGATCAAGCTCGCGACTCTCCAAGAGG
GAGGTGTCCAGTCTGCTAAGCTACACACAGGACTTGACAGGCGATGGCGACGGGGAAGATGGGGATGGCT
CTGACACCCCAGTCATGCCAAAGCTCTTCCGGGAAACCAGGACTCGTTCAGAAAGCCCAGCTGTCCGAAC
TCGAAATAACAACAGTGTCTCCAGCCGGGAGAGGCACAGGCCTTCCCCACGTTCCACCCGAGGCCGGCAG
GGCCGCAACCATGTGGACGAGTCCCCCGTGGAGTTCCCGGCTACCAGGTCCCTGAGACGGCGGGCAACAG
CATCGGCAGGAACGCCATGGCCGTCCCCTCCCAGCTCTTACCTTACCATCGACCTCACAGACGACACAGA
GGACACACATGGGACGCCCCAGAGCAGCAGTACCCCCTACGCCCGCCTAGCCCAGGACAGCCAGCAGGGG
GGCATGGAGTCCCCGCAGGTGGAGGCAGACAGTGGAGATGGAGACAGTTCAGAGTATCAGGATGGGAAGG
AGTTTGGAATAGGGGACCTCGTGTGGGGAAAGATCAAGGGCTTCTCCTGGTGGCCCGCCATGGTGGTGTC
TTGGAAGGCCACCTCCAAGCGACAGGCTATGTCTGGCATGCGGTGGGTCCAGTGGTTTGGCGATGGCAAG
TTCTCCGAGGTCTCTGCAGACAAACTGGTGGCACTGGGGCTGTTCAGCCAGCACTTTAATTTGGCCACCT
TCAATAAGCTCGTCTCCTATCGAAAAGCCATGTACCATGCTCTGGAGAAAGCTAGGGTGCGAGCTGGCAA
GACCTTCCCCAGCAGCCCTGGAGACTCATTGGAGGACCAGCTGAAGCCCATGTTGGAGTGGGCCCACGGG
GGCTTCAAGCCCACTGGGATCGAGGGCCTCAAACCCAACAACACGCAACCAGAGAACAAGACTCGAAGAC
GCACAGCTGACGACTCAGCCACCTCTGACTACTGCCCCGCACCCAAGCGCCTCAAGACAAATTGCTATAA
CAACGGCAAAGACCGAGGGGATGAAGATCAGAGCCGAGAACAAATGGCTTCAGATGTTGCCAACAACAAG
AGCAGCCTGGAAGATGGCTGTTTGTCTTGTGGCAGGAAAAACCCCGTGTCCTTCCACCCTCTCTTTGAGG
GGGGGCTCTGTCAGACATGCCGGGATCGCTTCCTTGAGCTGTTTTACATGTATGATGACGATGGCTATCA
GTCTTACTGCACTGTGTGCTGCGAGGGCCGAGAGCTGCTGCTTTGCAGCAACACGAGCTGCTGCCGGTGT
TTCTGTGTGGAGTGCCTGGAGGTGCTGGTGGGCACAGGCACAGCGGCCGAGGCCAAGCTTCAGGAGCCCT
GGAGCTGTTACATGTGTCTCCCGCAGCGCTGTCATGGCGTCCTGCGGCGCCGGAAGGACTGGAACGTGCG
CCTGCAGGCCTTCTTCACCAGTGACACGGGGCTTGAATATGAAGCCCCCAAGCTGTACCCTGCCATTCCC
GCAGCCCGAAGGCGGCCCATTCGAGTCCTGTCATTGTTTGATGGCATCGCGACAGGCTACCTAGTCCTCA
AAGAGTTGGGCATAAAGGTAGGAAAGTACGTCGCTTCTGAAGTGTGTGAGGAGTCCATTGCTGTTGGAAC
CGTGAAGCACGAGGGGAATATCAAATACGTGAACGACGTGAGGAACATCACAAAGAAAAATATTGAAGAA
TGGGGCCCATTTGACTTGGTGATTGGCGGAAGCCCATGCAACGATCTCTCAAATGTGAATCCAGCCAGGA
AAGGCCTGTATGAGGGTACAGGCCGGCTCTTCTTCGAATTTTACCACCTGCTGAATTACTCACGCCCCAA
GGAGGGTGATGACCGGCCGTTCTTCTGGATGTTTGAGAATGTTGTAGCCATGAAGGTTGGCGACAAGAGG
GACATCTCACGGTTCCTGGAGTGTAATCCAGTGATGATTGATGCCATCAAAGTTTCTGCTGCTCACAGGG
CCCGATACTTCTGGGGCAACCTACCCGGGATGAACAGGCCCGTGATAGCATCAAAGAATGATAAACTCGA
GCTGCAGGACTGCTTGGAATACAATAGGATAGCCAAGTTAAAGAAAGTACAGACAATAACCACCAAGTCG
AACTCGATCAAACAGGGGAAAAACCAACTTTTCCCTGTTGTCATGAATGGCAAAGAAGATGTTTTGTGGT
GCACTGAGCTCGAAAGGATCTTTGGCTTTCCTGTGCACTACACAGACGTGTCCAACATGGGCCGTGGTGC
CCGCCAGAAGCTGCTGGGAAGGTCCTGGAGCGTGCCTGTCATCCGACACCTCTTCGCCCCTCTGAAGGAC
TACTTTGCATGTGAATAG

Added complexity: Alternate splicing

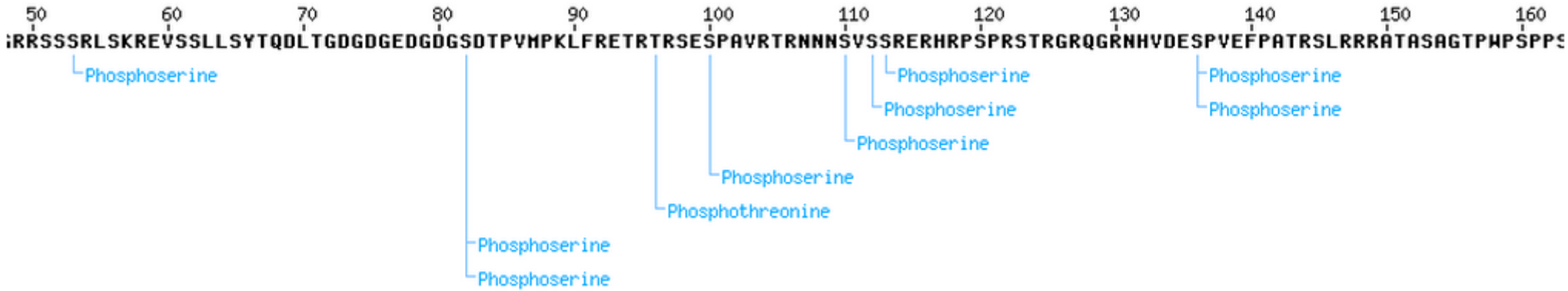
Example: The DNMT3B gene produces six distinct proteins.





More complexity: Proteins modified by PTM

DNMT3B post translational modifications (PTMs):

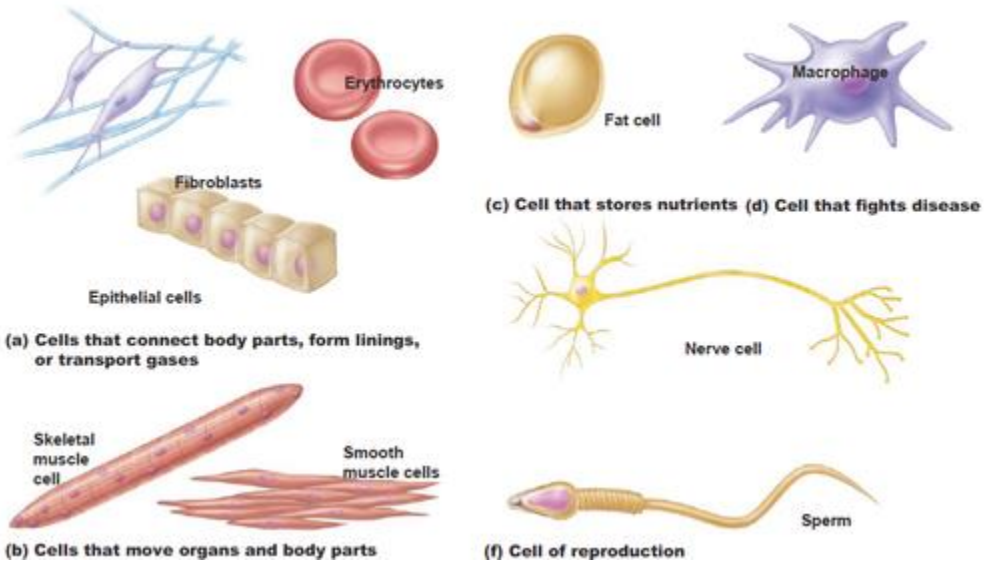


dbptm.mbc.nctu.edu.tw

Even more complexity: Abundance variation

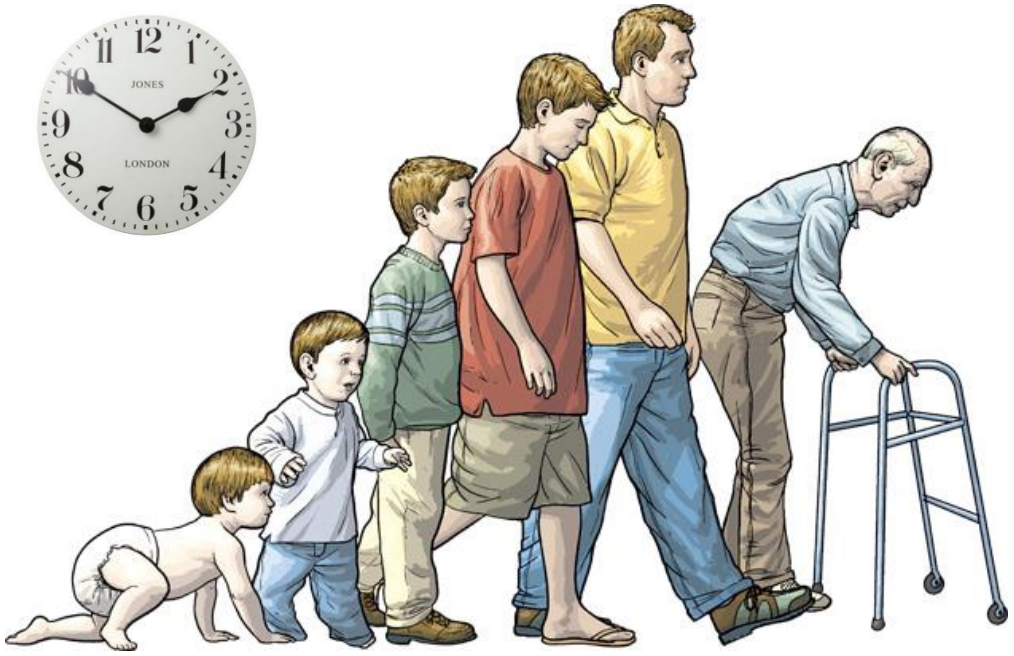
- Expression of genes (and hence proteins) varies according to:

Cell Type



anatomyandphysiologyi.com

Time



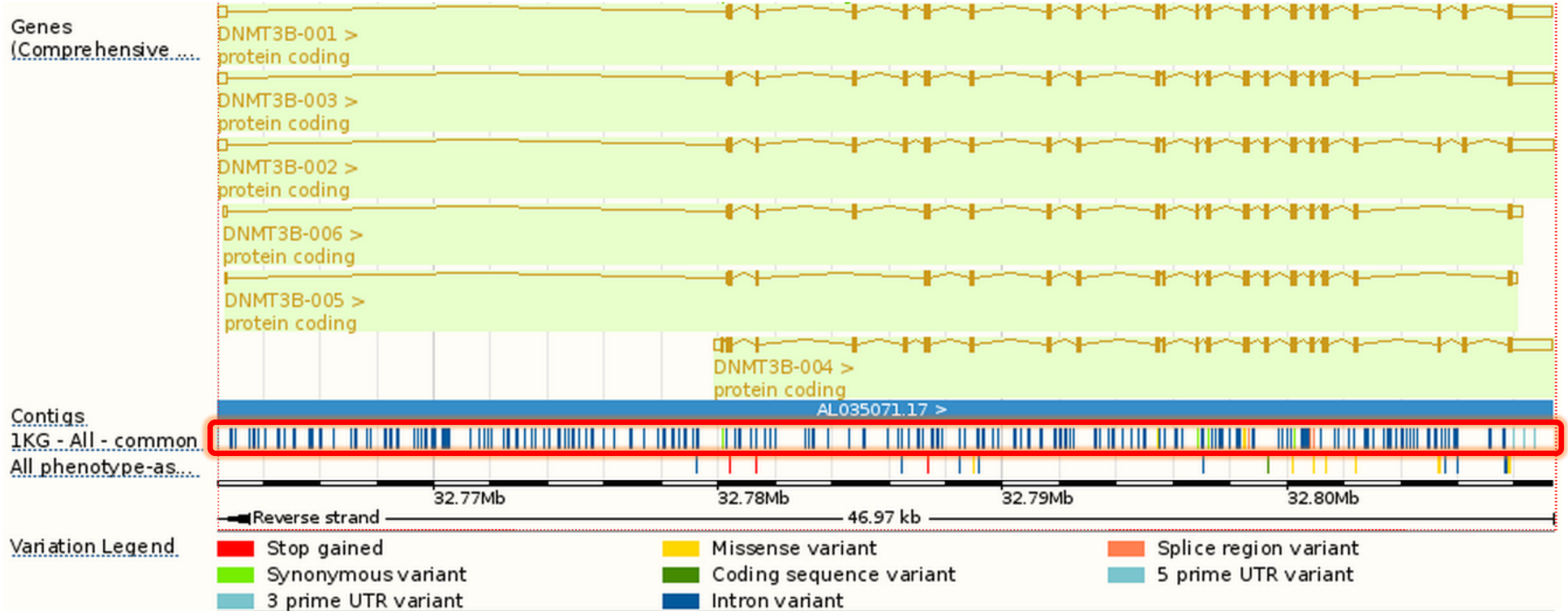
www.nicaraguaenvivo.com

High throughput analytical techniques

Sequencers to quantify gene expression (RNA-seq).

Mass spectrometers to quantify proteins and metaboloites.

And there's more complexity: Genome variation



Just in humans …

billions of different states and individual variations

1M+ distinct proteins

90,000+ gene variants

20,000+ genes

Bioinformatics in 2024

High throughput laboratory methods are producing a tsunami of biological data.

Not only DNA sequence data. Also RNA transcripts, data on proteins, metabolites, etc. etc.

The bottleneck in making new discoveries in the biosciences has moved from the lab to the computer.

Good bioinformaticians are in high demand.

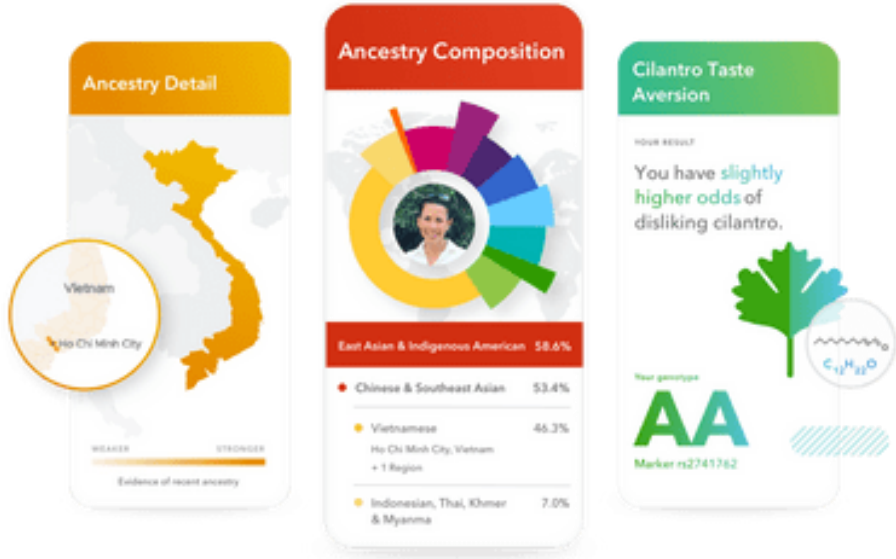
A modern sequencing lab

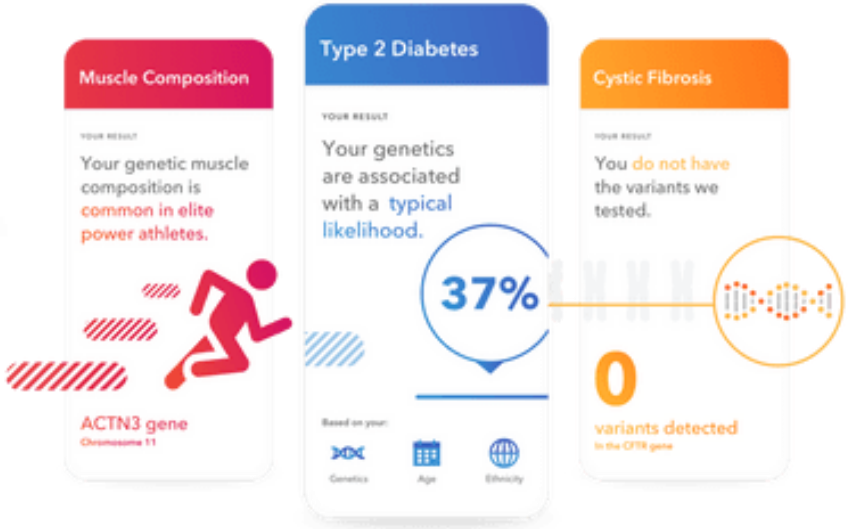
Nanopore Sequencing

- ✓ Portable
- ✓ Low cost
- ✓ Whole genome
- ✓ Long read lengths

Personal genotyping

£159

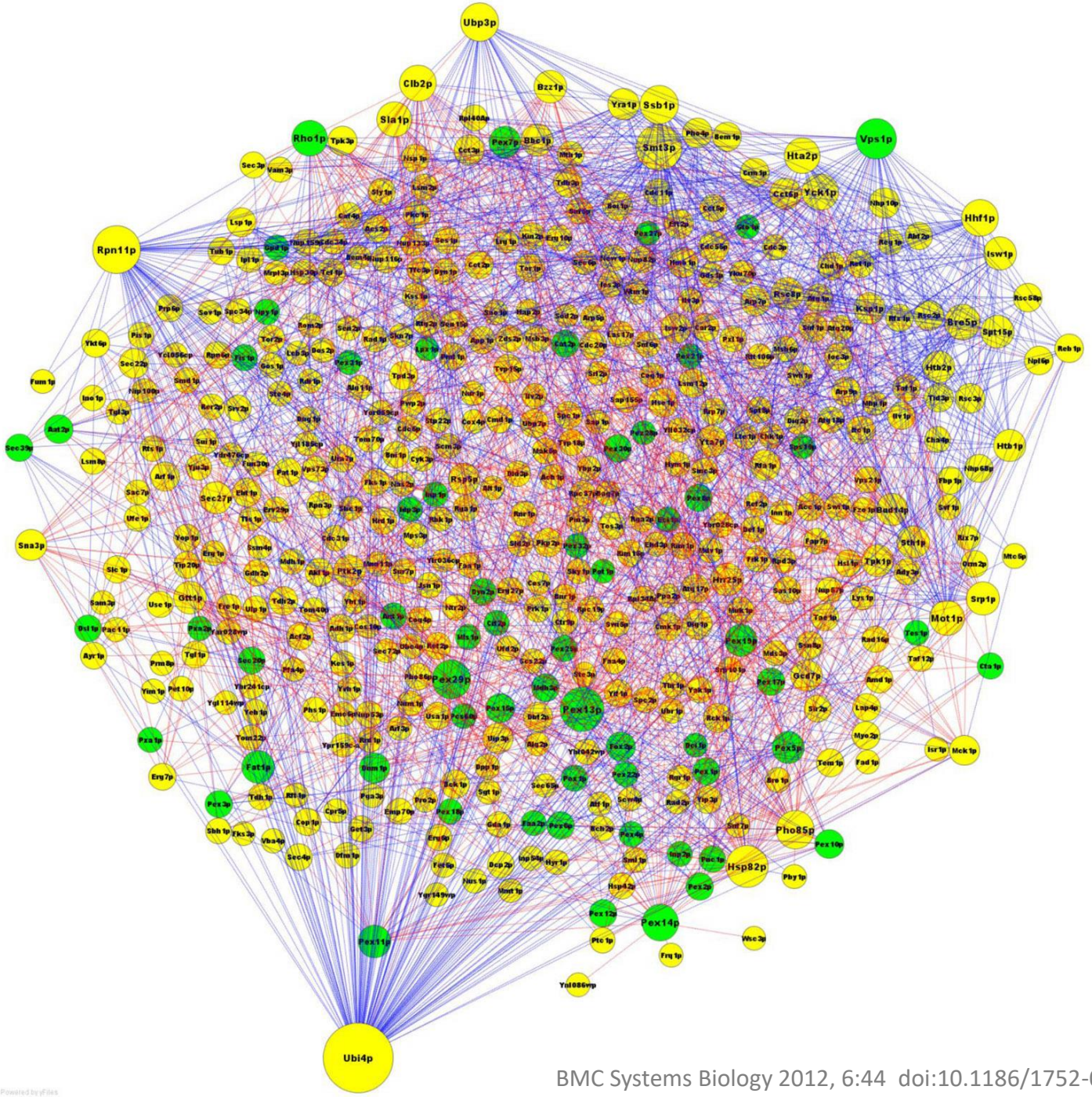




Systems biology

Aims to understand what is actually happening at a molecular level.

Potentially millions of interactions between proteins, metabolites and other molecules.



BMC Systems Biology 2012, 6:44 doi:10.1186/1752-0509-6-44

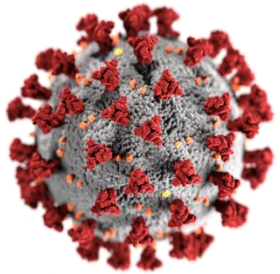
Even more complexity …

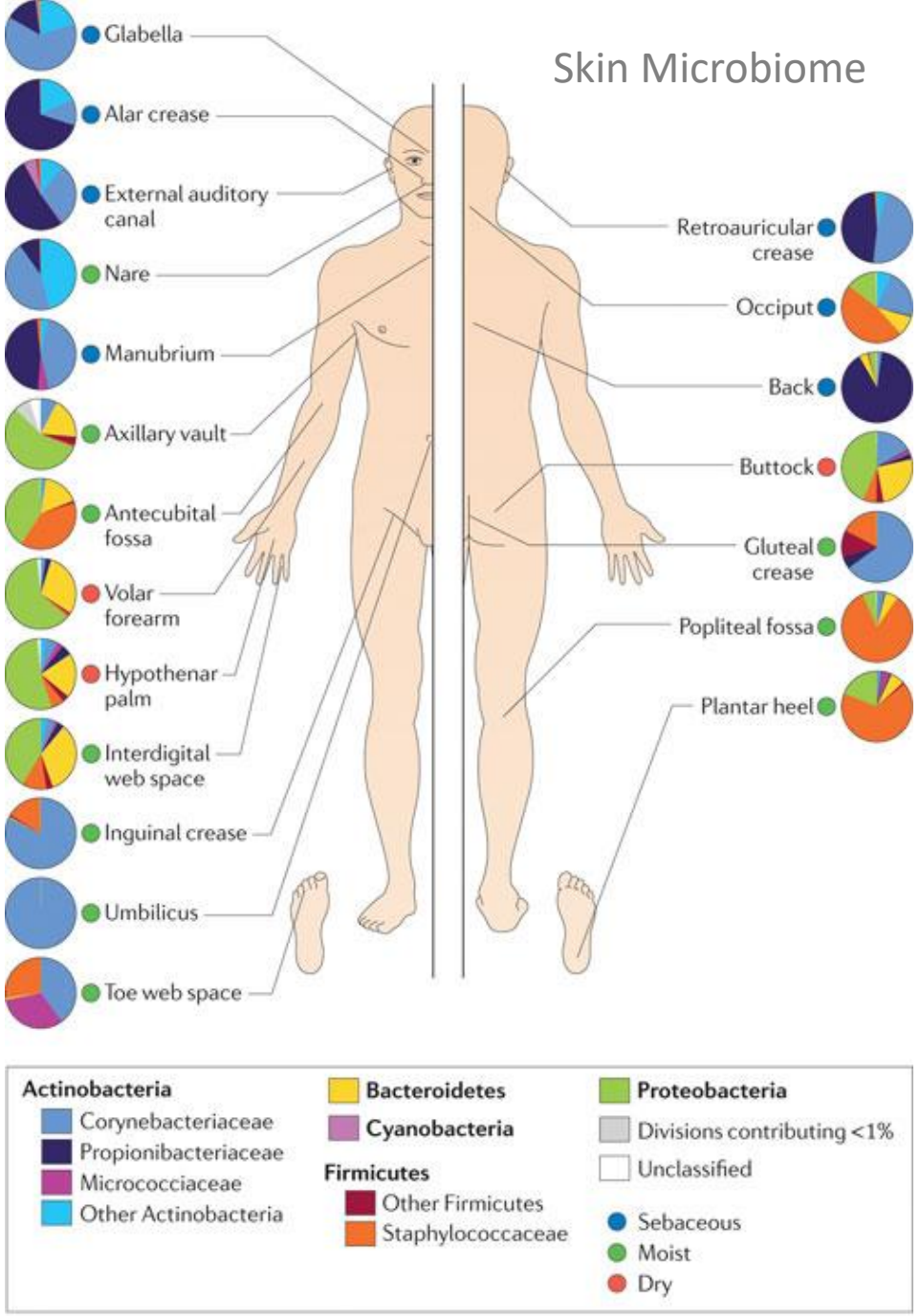
We are more microbe than human:

- 100s of other genomes to consider.

Many other important species to study:

- Organisms of environmental and ecological importance.
- Pathogens.
- Food.
- Industrial organisms.





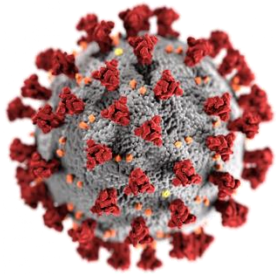
Even more complexity …

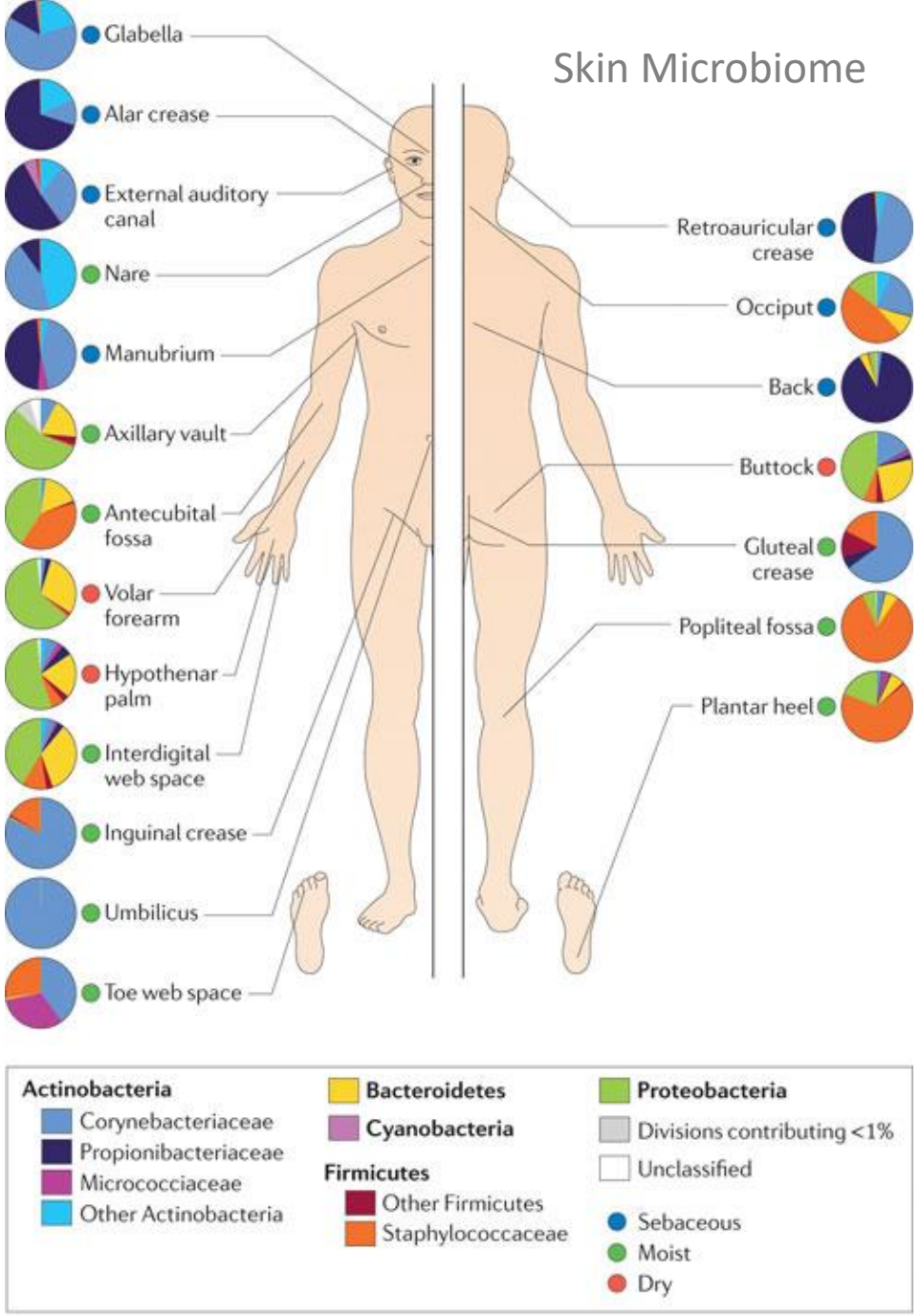
We are more microbe than human:

- 100s of other genomes to consider.

Many other important species to study:

- Organisms of environmental and ecological importance.
- Pathogens.
- Food.
- Industrial organisms.





We need computation … a lot of computation

Computational biology

2000 -

2010 -

Practicalities:

- Data storage and management.
- High performance computing.
- Visualisation.

Research:

- Application of existing algorithms to new data, to answer biological questions.
- Development of new algorithms.

2020 -

Essential computational biology skills

Technical
Molecular biology
Analytical science
Data analysis
Coding (programming)

Transferable
Scientific method
Teamwork
Communication
Attention to detail

Your MSc

Taught Modules
W1-3: Unix and analysis of large genomic datasets
W4-6: Coding for bioscientists
W7-9: Statistics for bioinformatics
W10-12: AI and data science in biology

Individual Research Project (5½ months)

A focussed piece of computer-based research within a research group at QMUL (or elsewhere)

23 Sept 2024

Oct Nov Dec Jan Feb Mar Apr May Jun Jul Aug


Yannick Wurm


Fabrizio Smeraldi


Matteo Fumagalli


Conrad Bessant

Group Project (six weeks)

Work in a team of 4-5 people to produce a piece of software to solve a biological problem

Queen Mary University of London

Examples of this previous MSc research projects

1. Tracking ongoing genome evolution using single cell sequencing
2. Can heritable DNA methylations increase disease resistance?
3. Inference of genomic signatures of natural selection using deep learning
4. Using machine learning for thyroid disease prediction and multimorbidity
5. Mapping epigenomic data to repetitive elements
6. Integrative analysis of miRNA and mRNA expression profiles in arthritis
7. Exploring how genome size impacts plant extinction risk
8. Multi-modal analysis of single cell data using machine learning methods
9. Systematic identification of biochemical networks in cancer cells








“We cannot teach people anything; we can only help them discover it within themselves”

Galileo Galilei
(1564-1642)

Less of this …

More of this …

King Size Homer (Episode 3F05), Simpsons Season 7, 1995.

A word about cheating

Plagiarism, collusion, outsourcing,
excessive use of chatGPT.

Don’t do it!

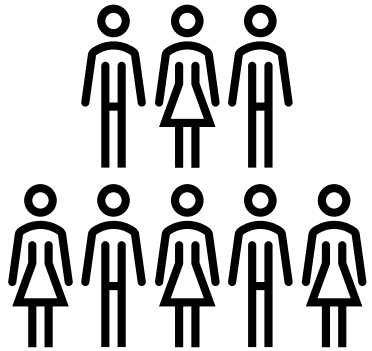
**Ask lectures, supervisor
or student support for help instead.**

Supporting your learning

Programme Leader

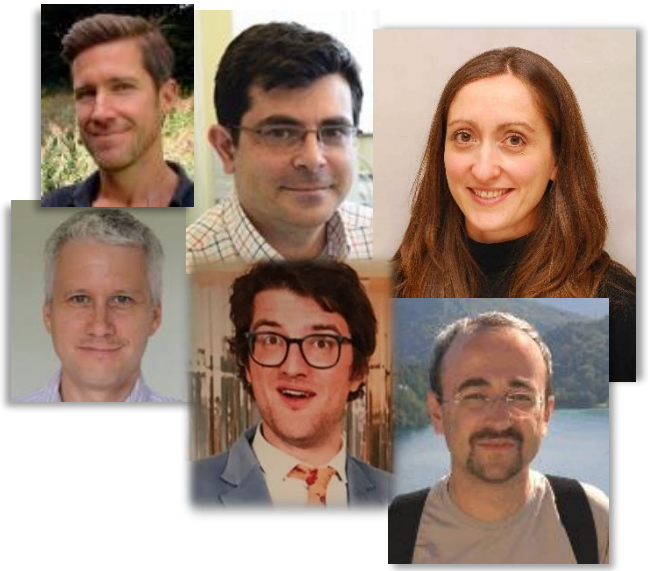
Conrad Bessant
c.bessant@qmul.ac.uk

Admin



SBBS Admin Team
sbbs-office@qmul.ac.uk

So Many Others

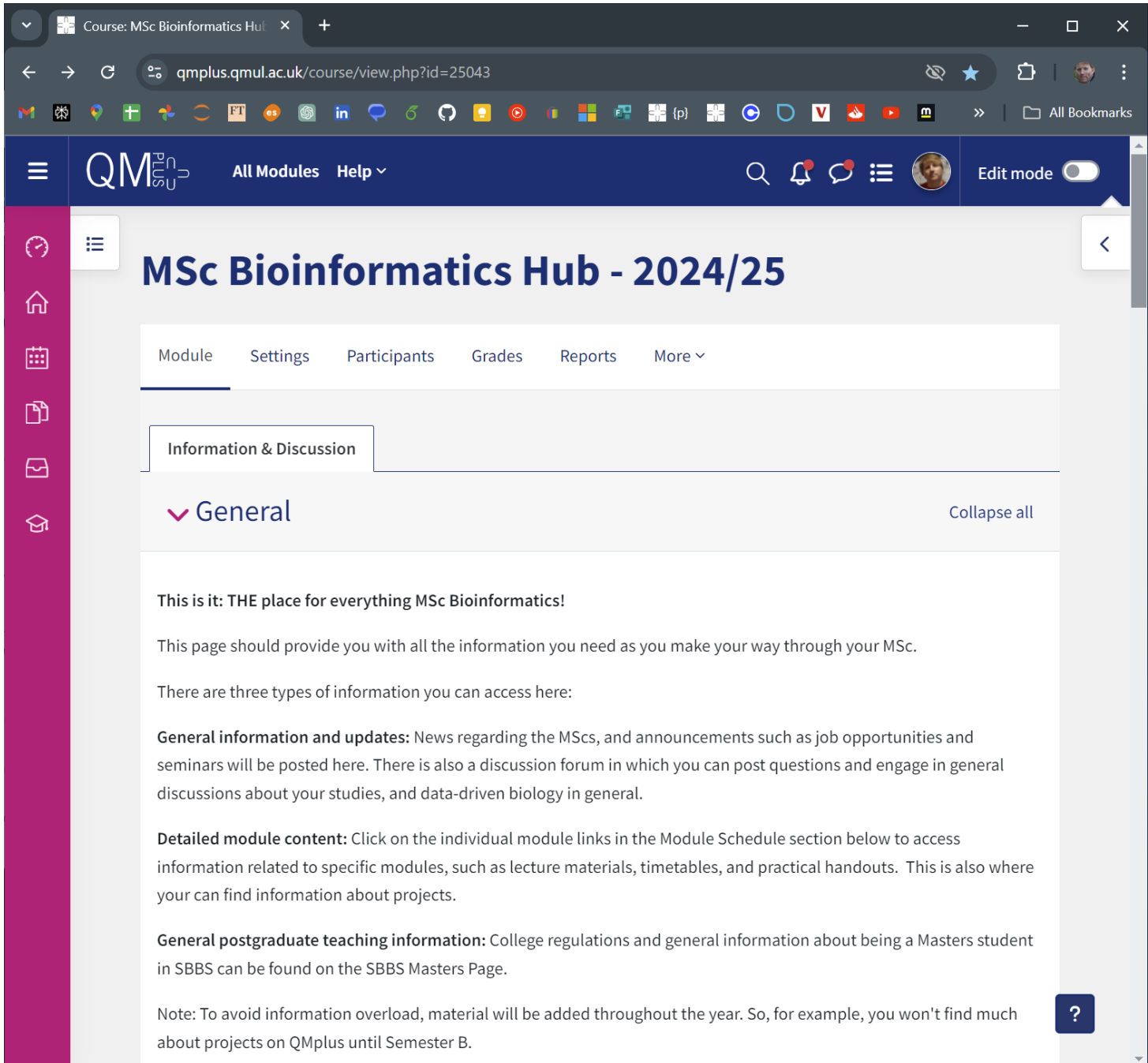


Module Leads
+ Lecturers
+ Demonstrators
+ Project Supervisors

MSc Bioinformatics Hub

bit.ly/bix24

(you need to be logged into QMplus to access)



Get ready for next week …

Unix and Analysis of Large Genomic Datasets (BIO726P)

Starts on Monday

Head over to the BIO726P QMplus page **from Friday** to get up to speed – there are things you need to do!

First live session in on Monday: see Qmplus for details.


Room QB209
Queens Building
Tower Hamlets
History Library...
The Octagon
At Queen Mary...
University
of London
G. E. Fogg Building
.E. Fogg Building
Queens' Building
QMUL School of
Engineering and...
A11
Harford St

A word about attendance

We expect everyone to be present in person at the hands-on practical sessions in Semester A.

We will not be streaming these sessions online.

Why?

linkedin.com

Is it worth it?

Justyna Gredecka
Full Stack Software Engineer at Cambridge Cancer Genomics

Modupeh Betts
Bioinformatics fellow at Medical Research Council(UK) The Gambia

Rachael Turner
Data Engineer at Sainsbury's

Numaan Iqbal
Data Analyst at Knight Frank

Nazrath Nawaz
Senior Bioinformatician at Kinomica

yasmine benbrahim
Inside Sales Account Manager (Nordics) at illumina
Connected 2 years ago

Karol Antoniuk
Bioinformatics Associate at Exact Sciences

Alejandra Carriero
PhD Student at University of Sussex

Josephine Mensah-Kane
Data Scientist at Mologic

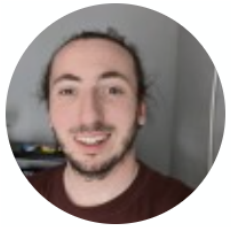
Dionysios Grigoriadis
Bioinformatician at European Bioinformatics Institute

Suresh Hewapathirana
Bioinformatician/ Software Engineer
at European Bioinformatics Institute

Hajar Saihi
PhD at the Blizard Institute (Centre for Immunobiology)

Alumni Panel Q&A

2022



Liam Johnson
Bioinformatician - Clinical Genomics
The Royal Marsden NHS Foundation Trust

2022

Janeesh Kaur Bansal
PhD Student at Queen Mary University of London
London

2022

Shiloh Alleyne
Financial Model Developer at Gallagher Re
Ascot