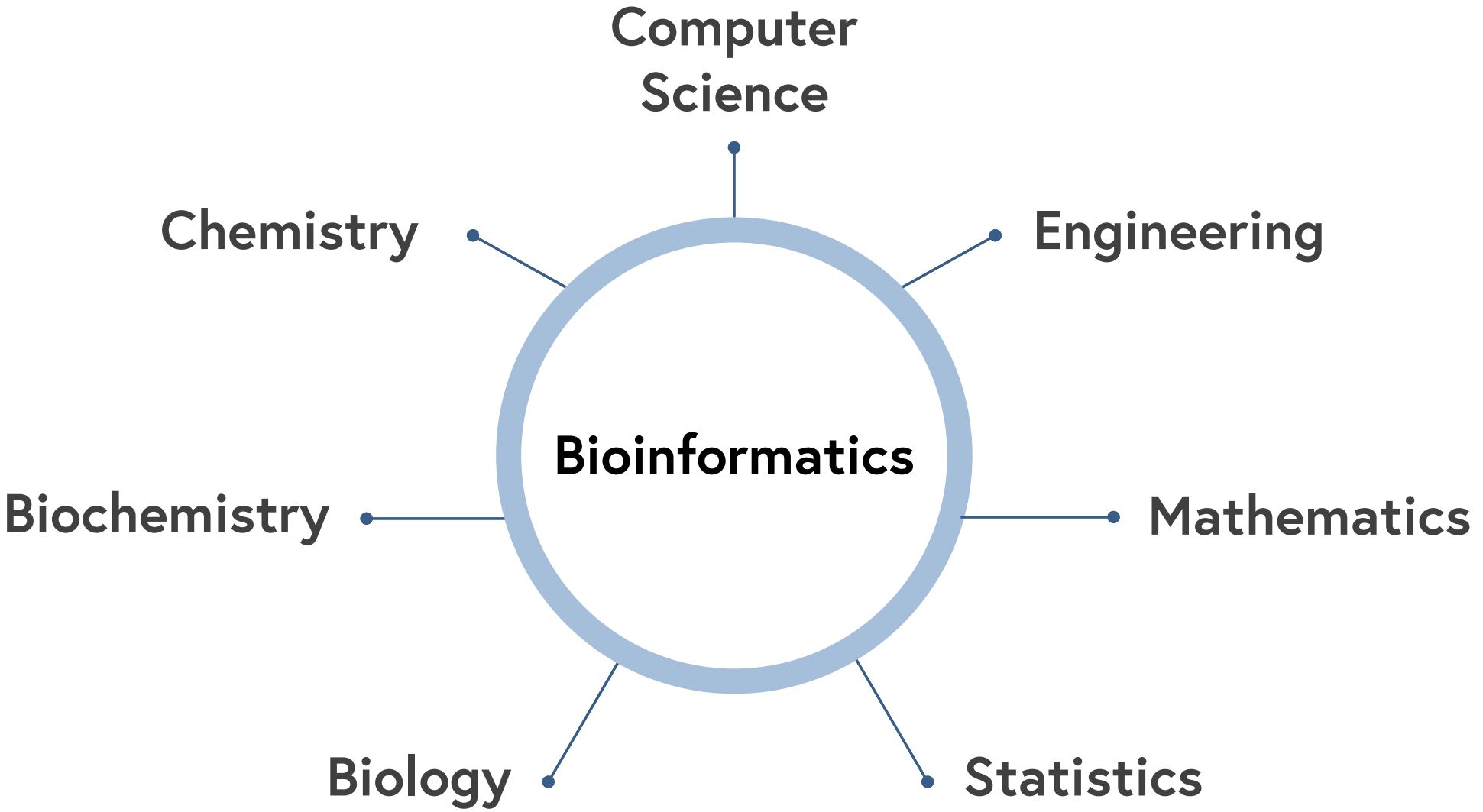
Overview on Bioinformatics

Outline

- Overview of Bioinformatics
- Public Resources for Bioinformatics
  - Different public databases for bioinformatics data
- Bioinformatics Data
  - What are different kinds of Bioinformatics data?
- How To Collect Data
  - How to collect Bioinformatics data from public resources?

Overview of Bioinformatics


Computer Science
Chemistry
Engineering
Bioinformatics
Biochemistry
Mathematics
Biology
Statistics

Bioinformatics Branches

Sequence analysis

- DNA sequencing
- Genome annotation
- Pan genomics

Structural bioinformatics

- Molecular docking
- Virtual screening

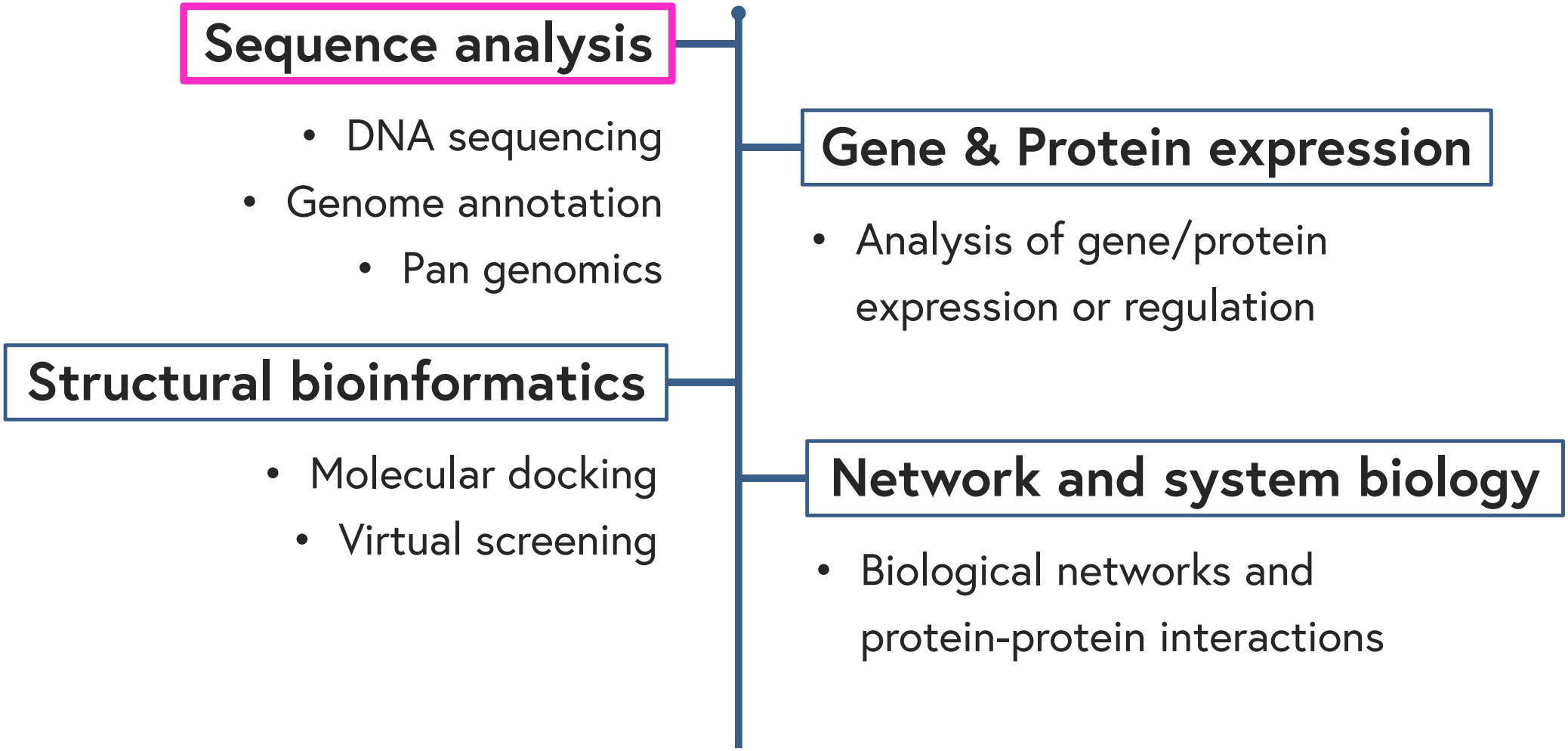
Gene & Protein expression

- Analysis of gene/protein expression or regulation

Network and system biology

- Biological networks and protein-protein interactions

Bioinformatics Branches



Bioinformatics Branches

Sequence analysis

- DNA sequencing
- Genome annotation
- Pan genomics

Structural bioinformatics

- Molecular docking
- Virtual screening

Gene & Protein expression

- Analysis of gene/protein expression or regulation

Network and system biology

- Biological networks and protein-protein interactions

Bioinformatics Branches

Sequence analysis

- DNA sequencing
- Genome annotation
- Pan genomics

Structural bioinformatics

- Molecular docking
- Virtual screening

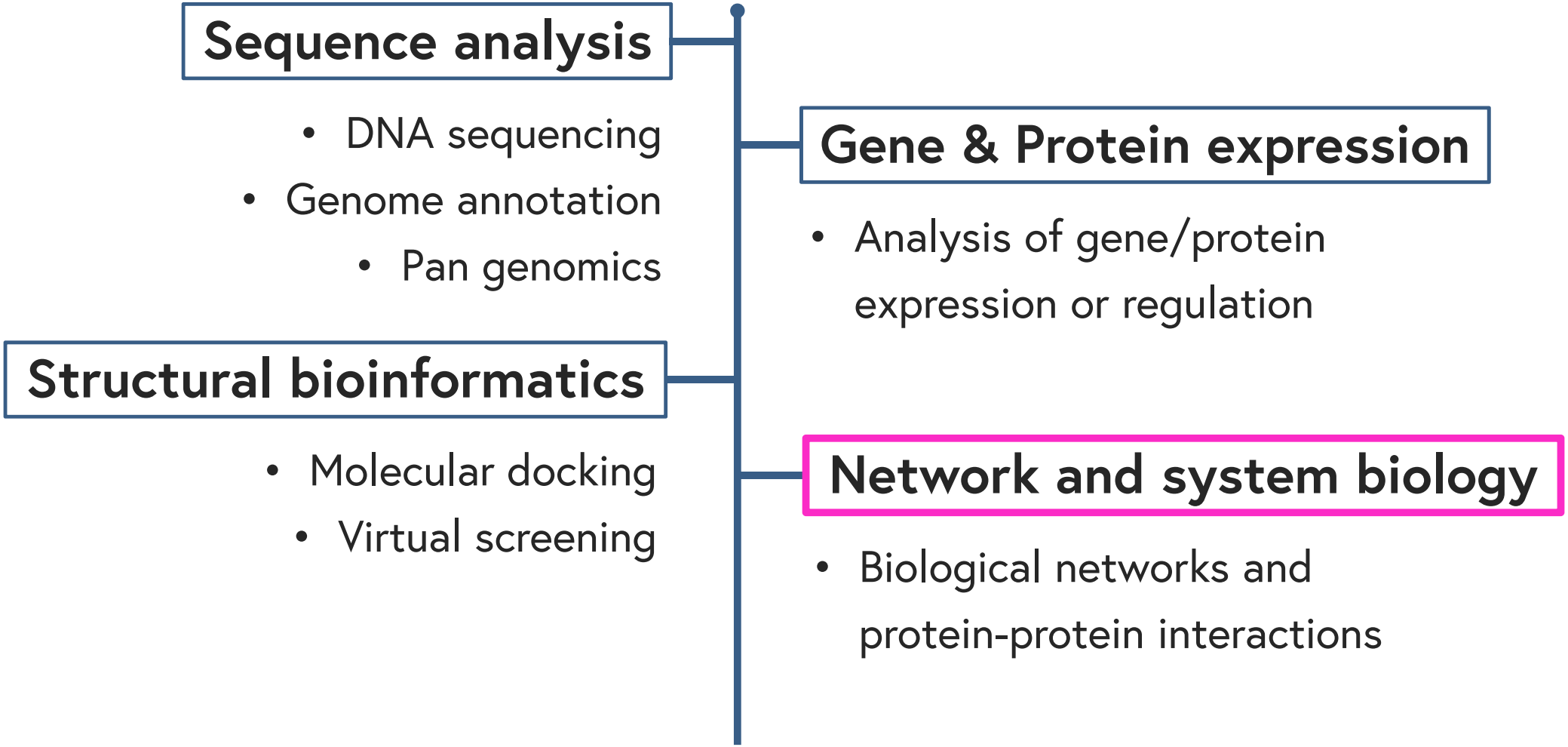
Gene & Protein expression

- Analysis of gene/protein expression or regulation

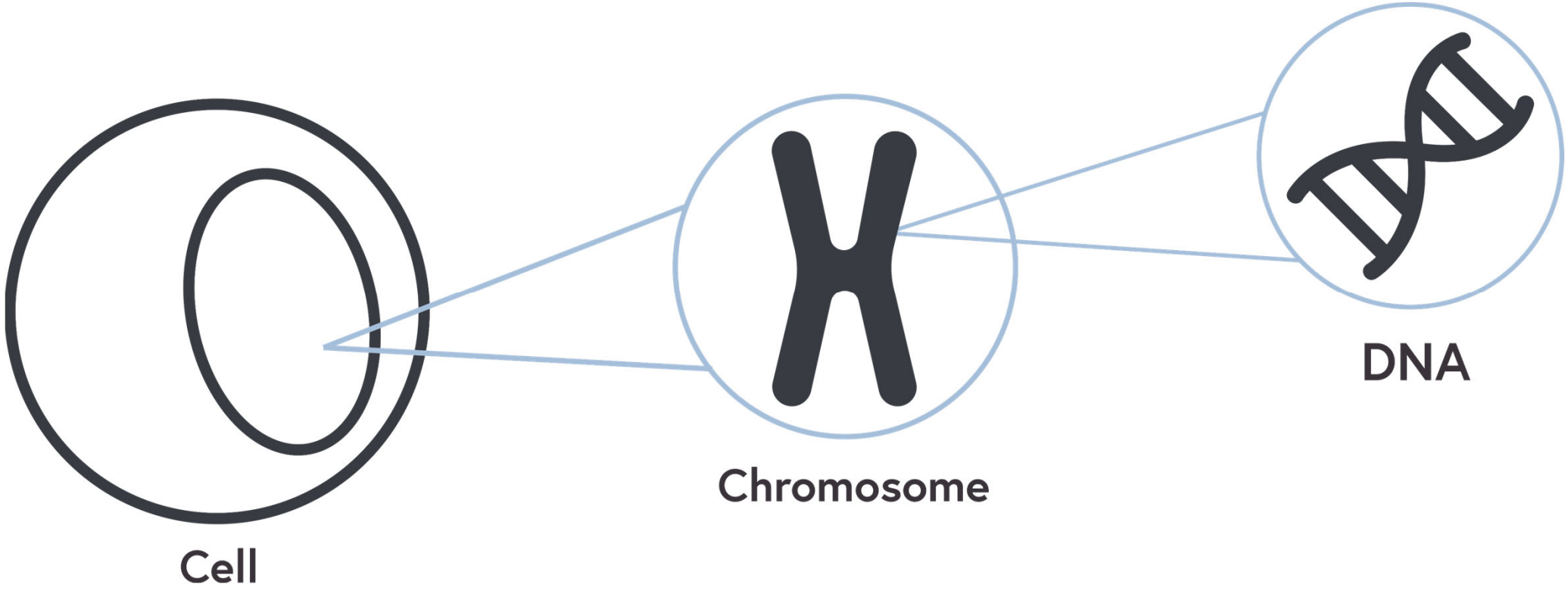
Network and system biology

- Biological networks and protein-protein interactions

Bioinformatics Branches



Cell & Genetics



DNA Sequence

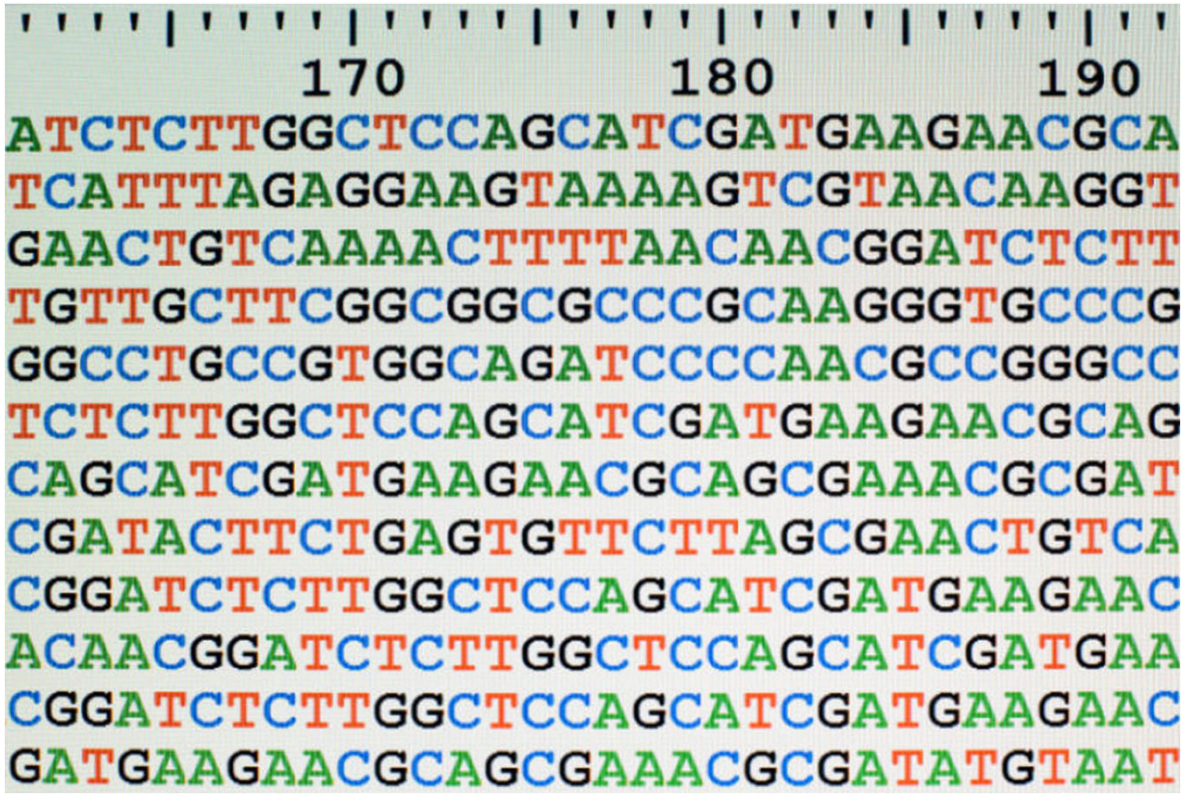
Nucleobases

↓

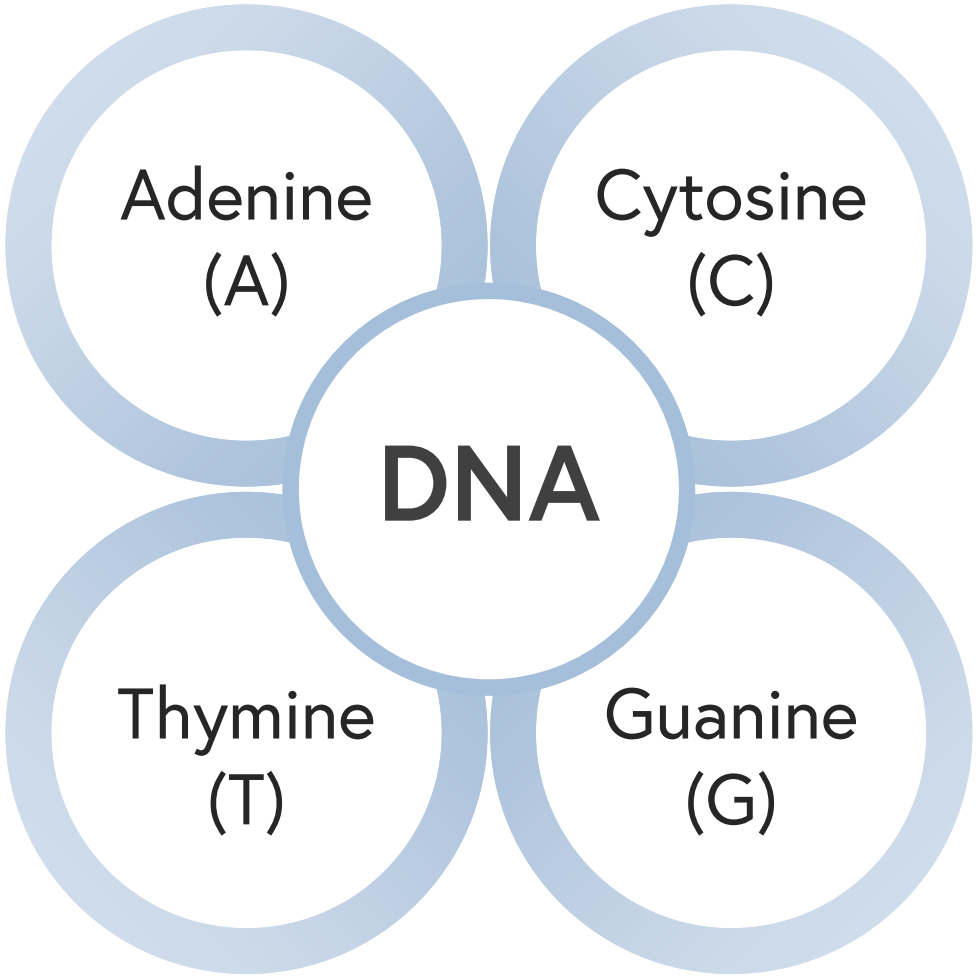
Nucleic acids

↓

DNA



DNA Sequence

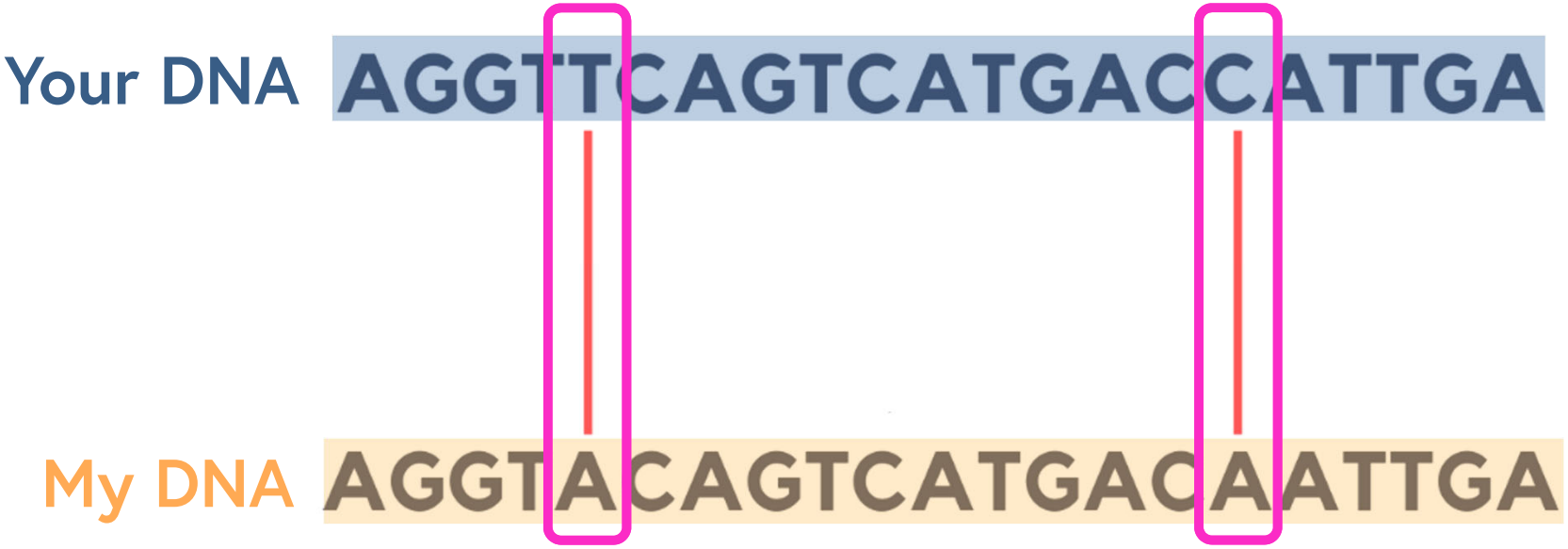


Genome

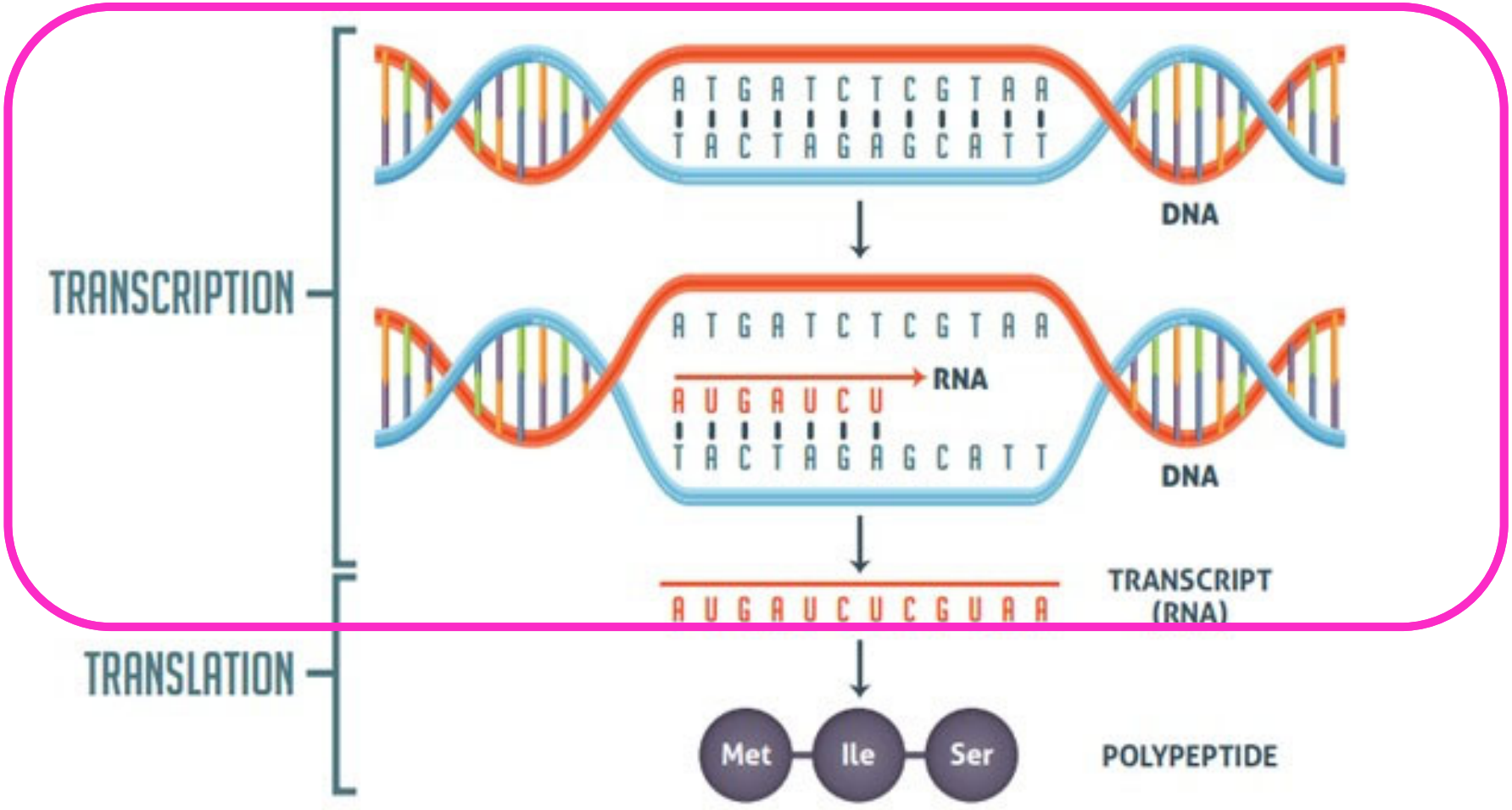
The collective sequence

DNA Sequence

- Different people has different DNA sequences

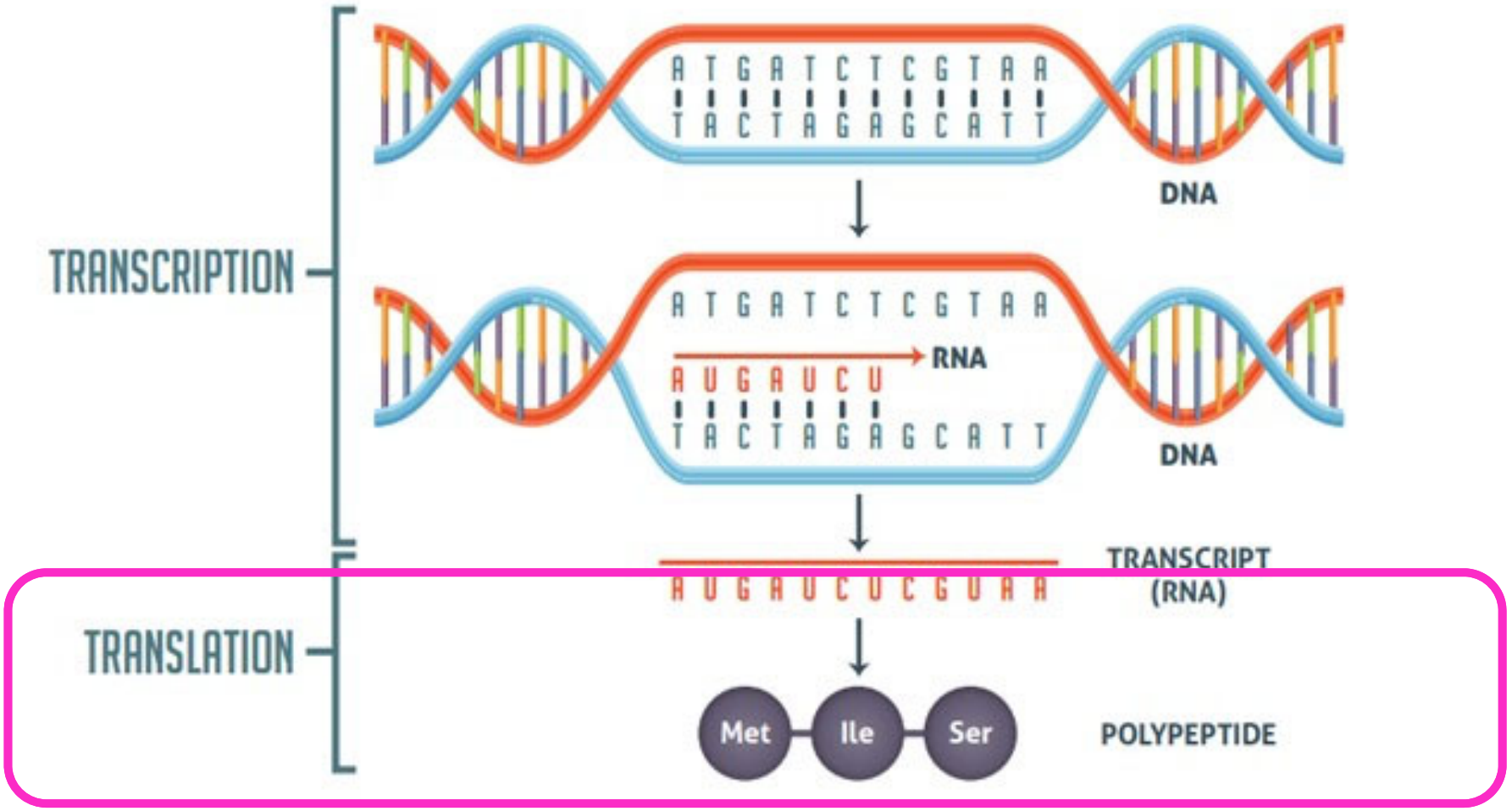


From DNA to Protein



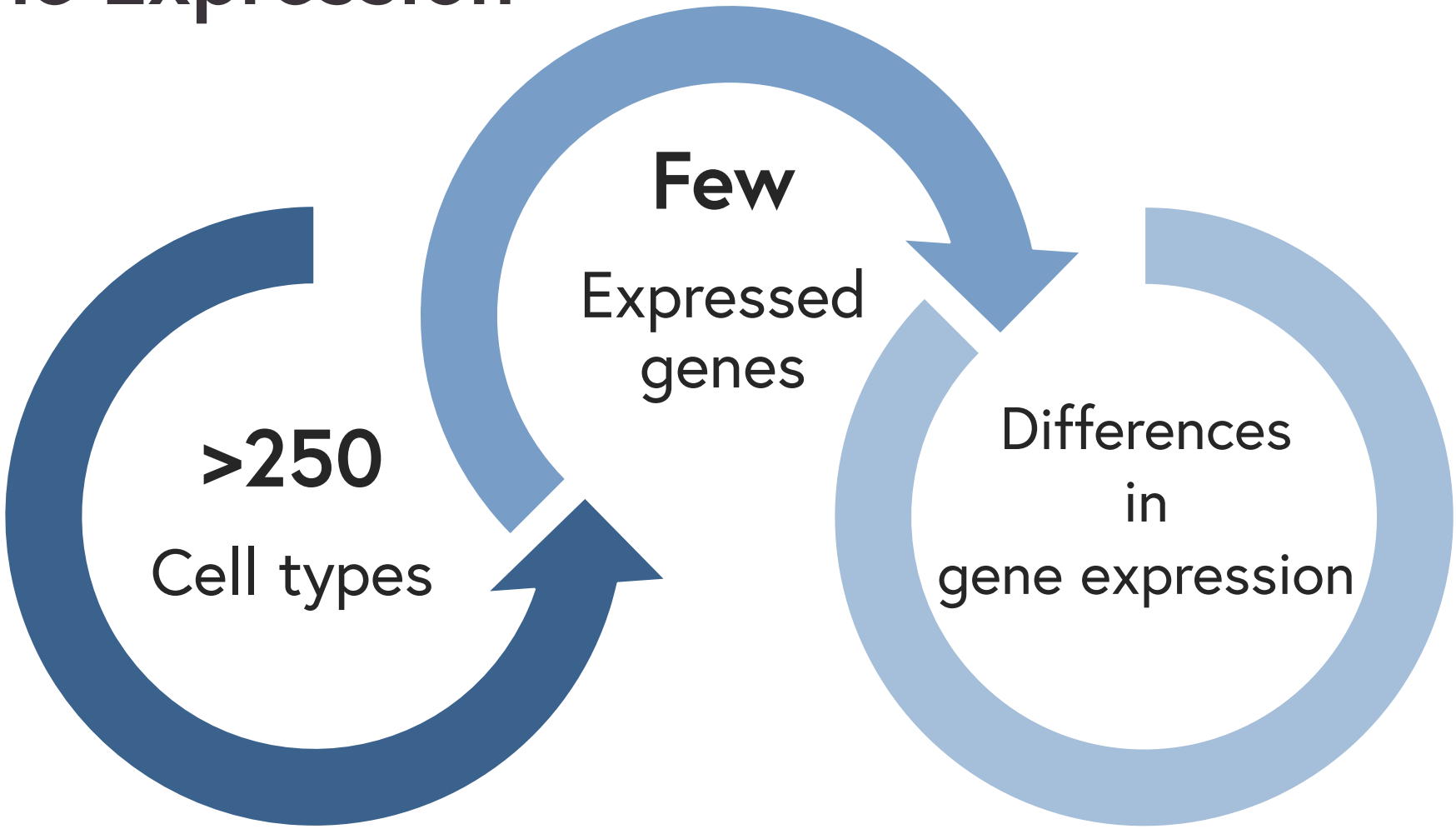
https://www.technologynetworks.com/genomics/articles/transcription-vs-translation-worksheet-323080

From DNA to Protein

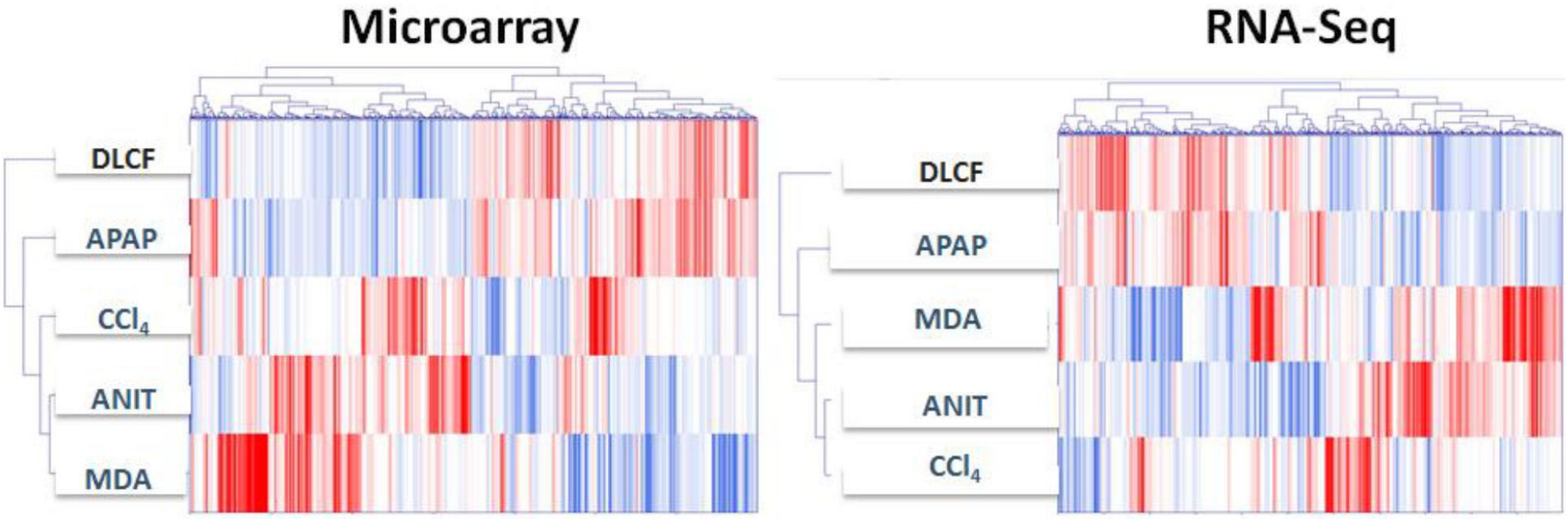


https://www.technologynetworks.com/genomics/articles/transcription-vs-translation-worksheet-323080

Gene Expression



Different Platforms for Gene Expression



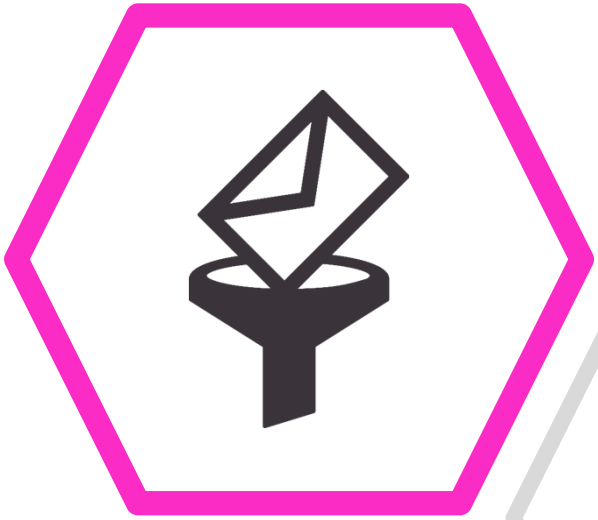
Rao MS, Van Vleet TR, Ciurlionis R, Buck WR, Mittelstadt SW, Blomme EAG and Liguori MJ (2019) Comparison of RNA-Seq and Microarray Gene Expression Platforms for the Toxicogenomic Evaluation of Liver From Short-Term Rat Toxicity Studies. Front. Genet. 9:636. doi: 10.3389/fgene.2018.00636

Why AI in Bioinformatics?

Artificial Intelligence

Artificial Intelligence

Mail Filter



Icons made by Freepik from Flaticon

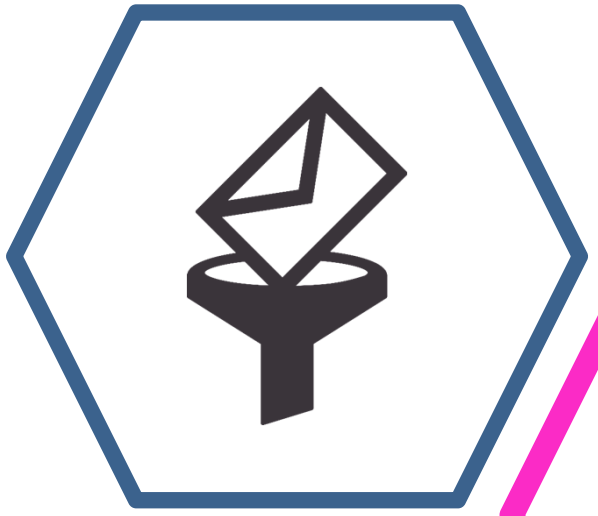
Icons made by Freepik from Flaticon

Icons made by Smashicons from Flaticon

Icons made by wanicon from Flaticon

Artificial Intelligence

Mail Filter



Icons made by Freepik from Flaticon

Auto Drive Car

Icons made by Freepik from Flaticon

Icons made by Smashicons from Flaticon

Icons made by wanicon from Flaticon

Artificial Intelligence

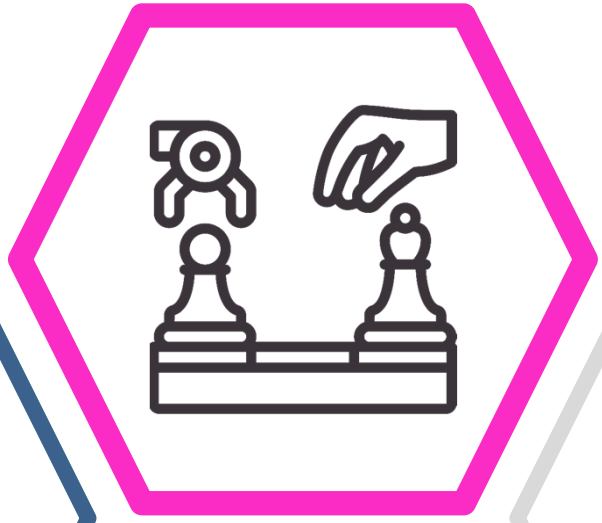
Mail Filter

Icons made by Freepik from Flaticon

Auto Drive Car

Icons made by Freepik from Flaticon

Chess



Icons made by Smashicons from Flaticon

Icons made by wanicon from Flaticon

Artificial Intelligence

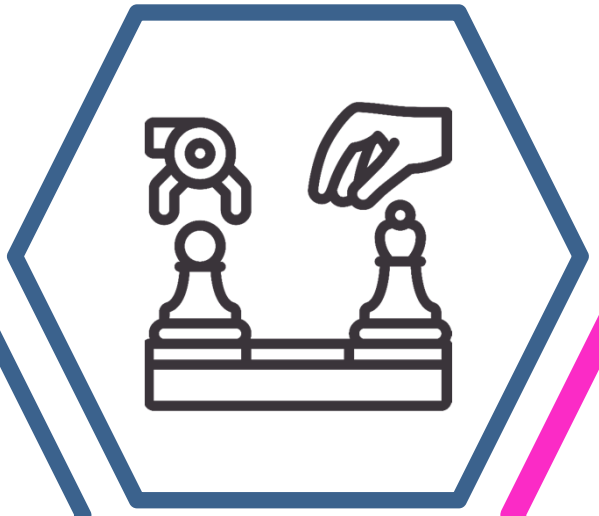
Mail Filter

Icons made by Freepik from Flaticon

Auto Drive Car

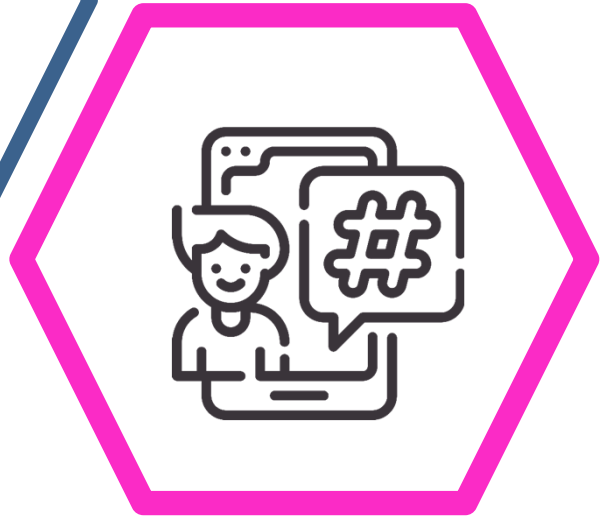
Icons made by Freepik from Flaticon

Chess



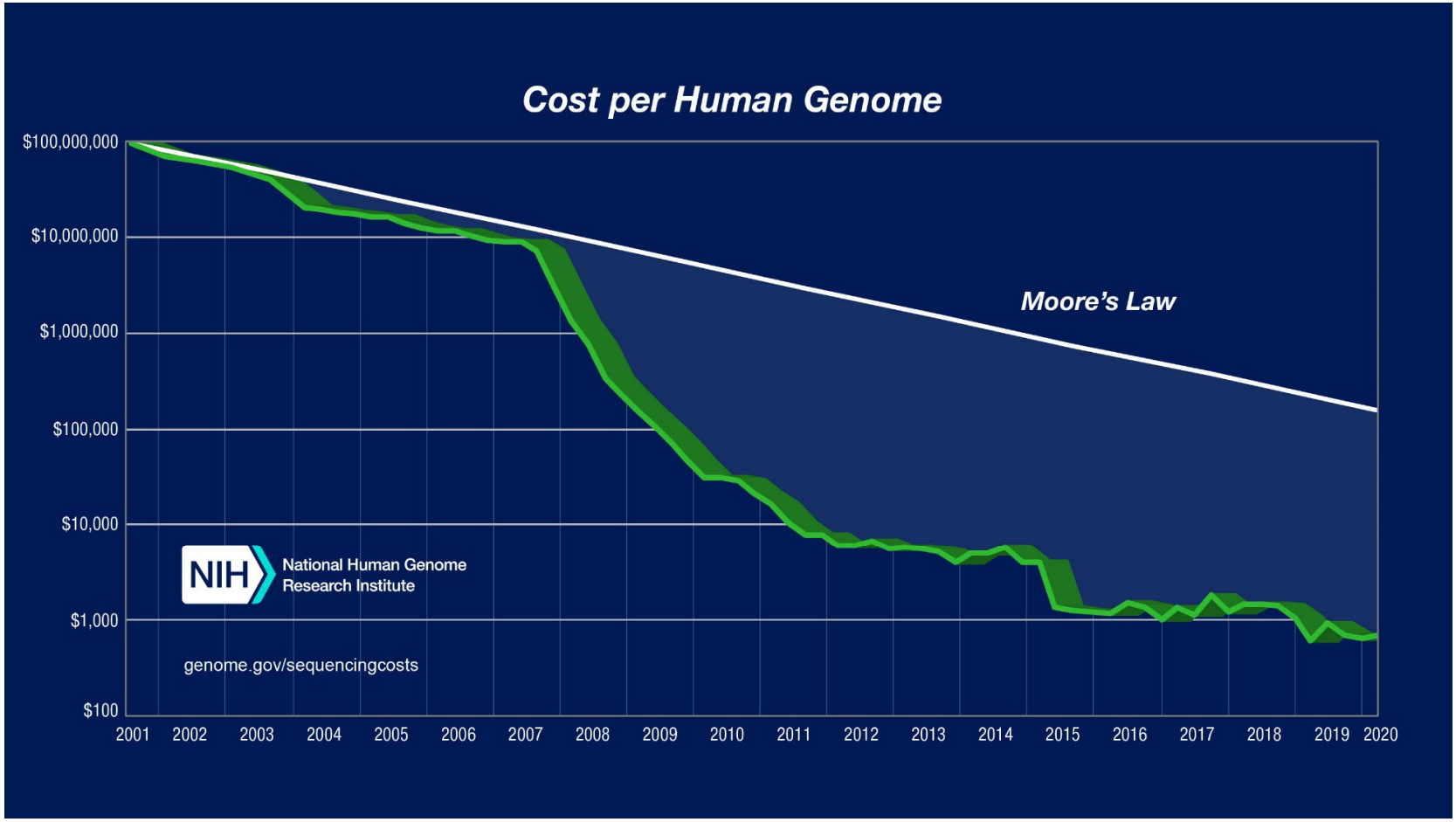
Icons made by Smashicons from Flaticon

Auto tag



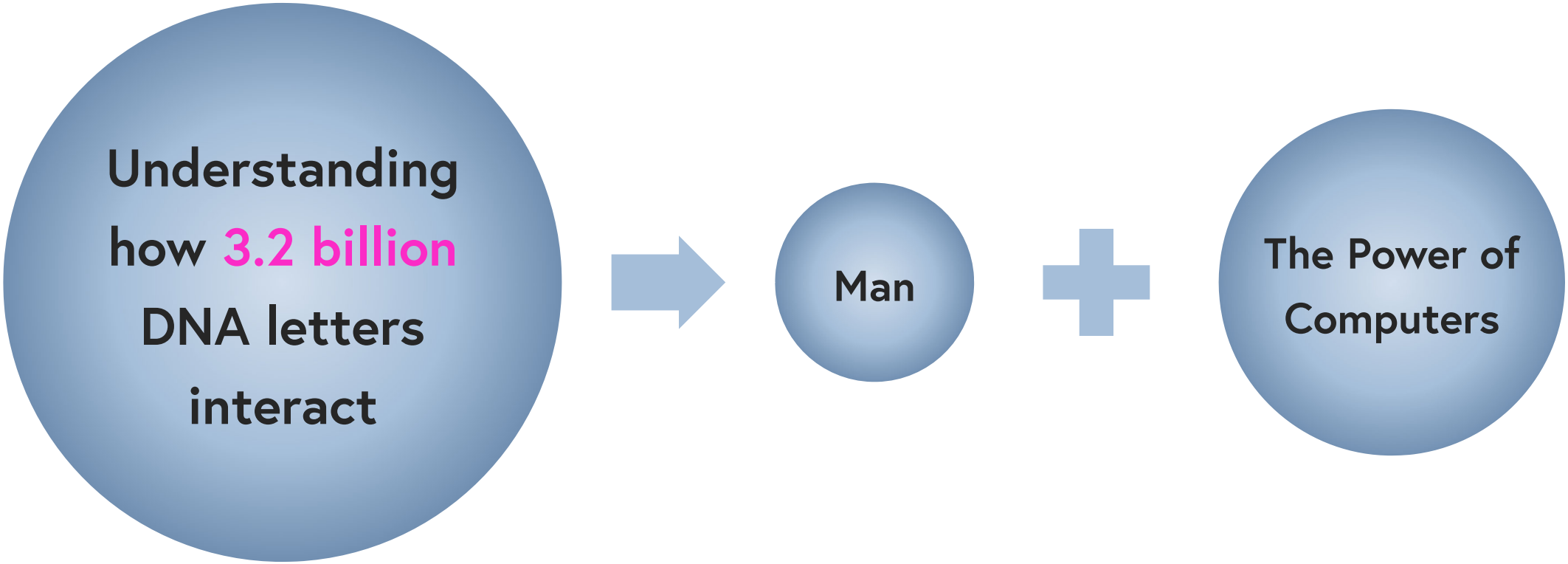
Icons made by wanicon from Flaticon

Why AI in Bioinformatics?

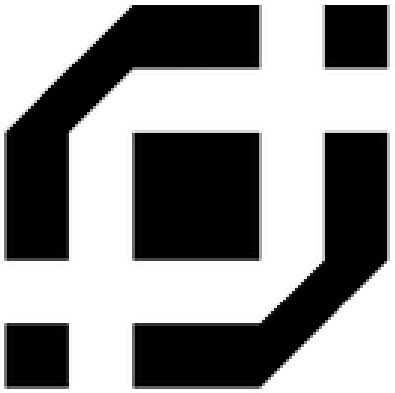


https://www.genome.gov/sequencingcosts/

Why AI in Bioinformatics?

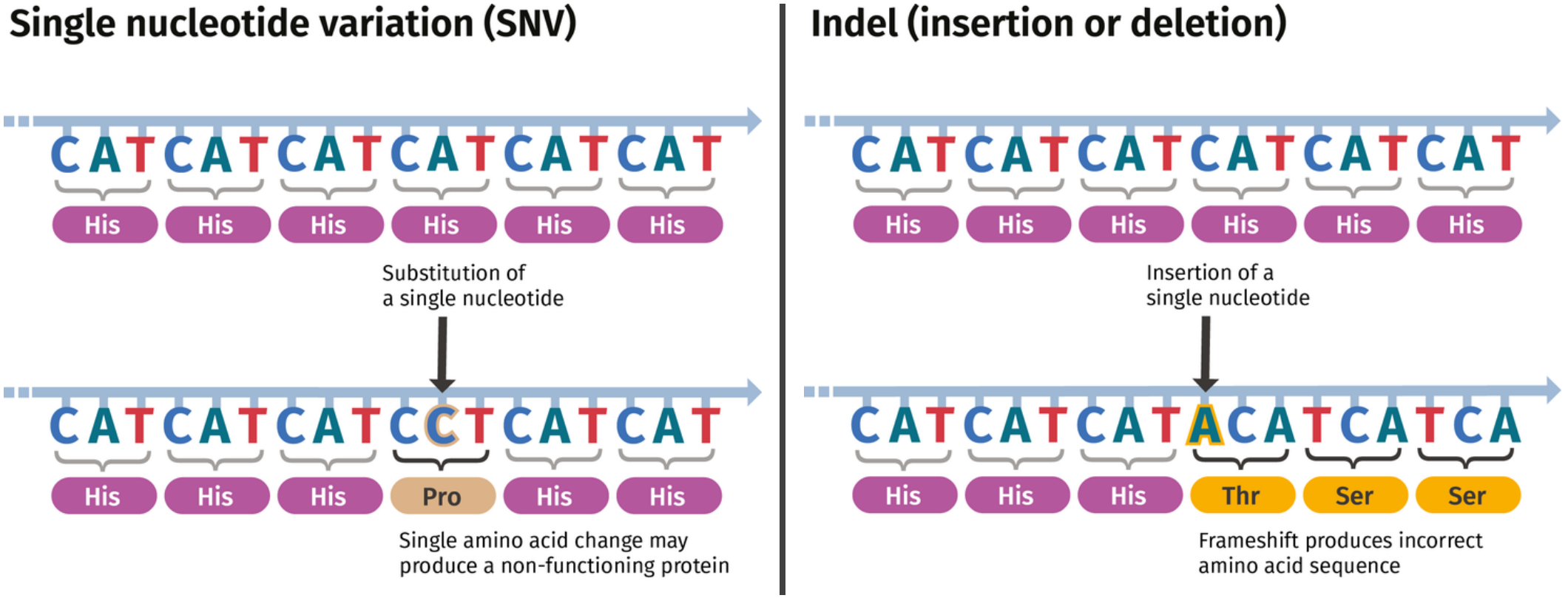


AI in Bioinformatics Companies



DESKTOP
GENETICS

AI in Variant Calling



https://www.garvan.org.au/research/kinghorn-centre-for-clinical-genomics/learn-about-genomics/for-gp/genetics-refresher-1/types-of-variants

AI in Variant Calling

Published: 24 September 2018

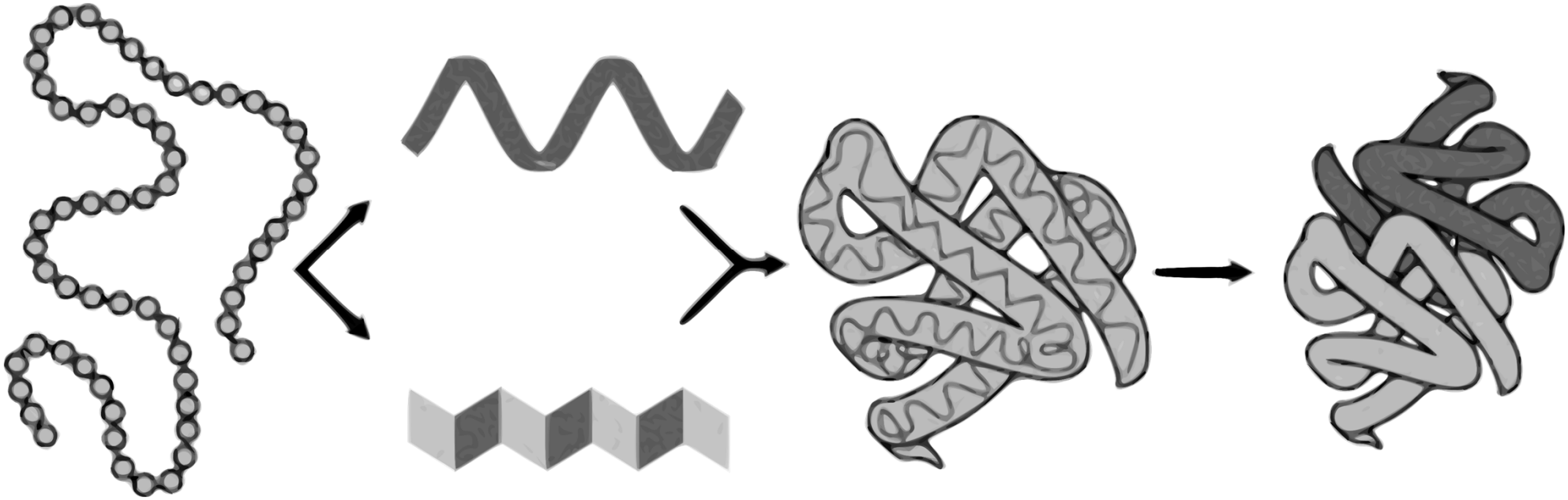
A universal SNP and small-indel variant caller using deep neural networks

Ryan Poplin, Pi-Chuan Chang, David Alexander, Scott Schwartz, Thomas Colthurst, Alexander Ku, Dan Newburger, Jojo Dijamco, Nam Nguyen, Pegah T Afshar, Sam S Gross, Lizzie Dorfman, Cory Y McLean & Mark A DePristo

Nature Biotechnology **36**, 983–987(2018) | Cite this article

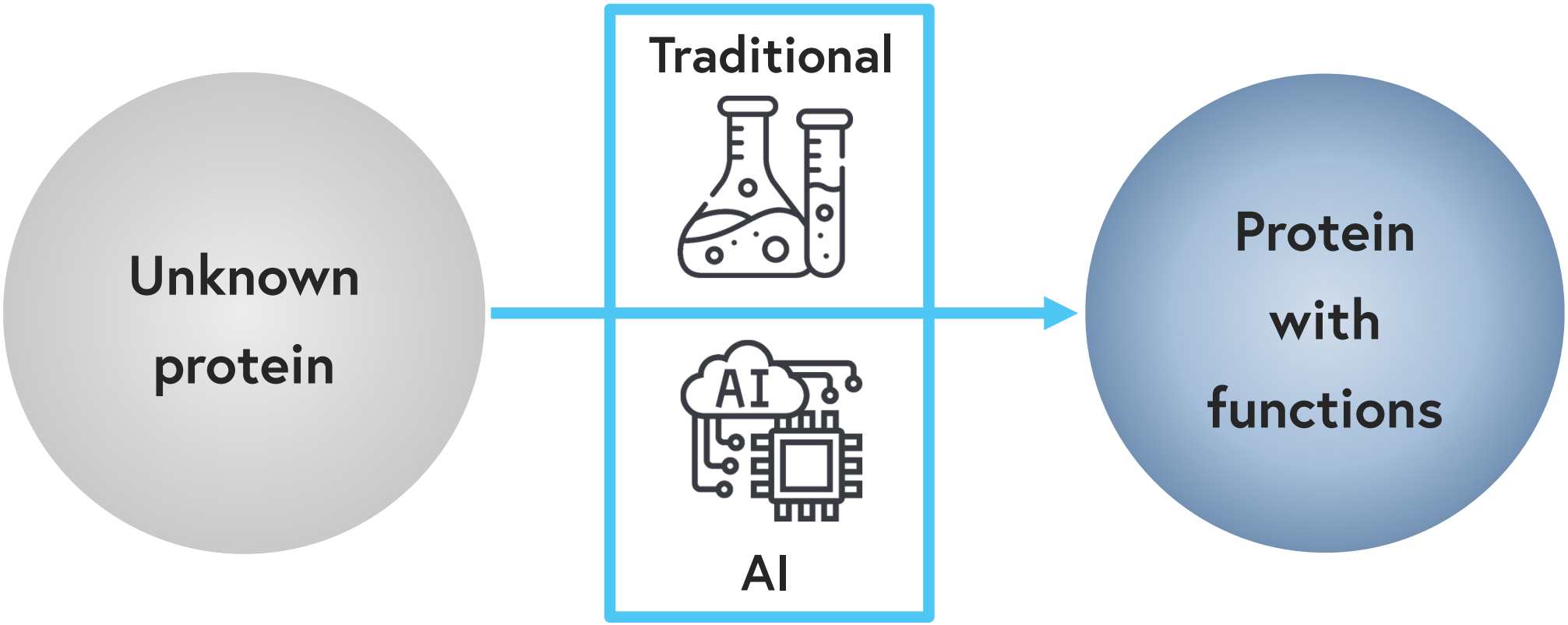
Poplin, R., Chang, PC., Alexander, D. et al. A universal SNP and small-indel variant caller using deep neural networks. Nat Biotechnol 36, 983–987 (2018). https://doi.org/10.1038/nbt.4235

AI in Protein Function Prediction



By NHGRI - Courtesy: National Human Genome Research Institute, Public Domain, https://commons.wikimedia.org/w/index.php?curid=1318488

AI in Protein Function Prediction



Flask free icon by Freepik from flaticon

Artificial Intelligence free icon by Eucalyp from flaticon

AI in Protein Function Prediction

Sequence analysis

DeepGOPlus: improved protein function prediction from sequence

Maxat Kulmanov and Robert Hoehndorf *

Computational Bioscience Research Center, Computer, Electrical and Mathematical Sciences & Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955, Saudi Arabia

*To whom correspondence should be addressed.

Associate Editor: Lenore Cowen

Received on April 25, 2019; revised on July 1, 2019; editorial decision on July 20, 2019; accepted on July 24, 2019

Abstract

Motivation: Protein function prediction is one of the major tasks of bioinformatics that can help in wide range of biological problems such as understanding disease mechanisms or finding drug targets. Many methods are available for predicting protein functions from sequence based features, protein–protein interaction networks, protein structure or literature. However, other than sequence, most of the features are difficult to obtain or not available for many proteins thereby limiting their scope. Furthermore, the performance of sequence-based function prediction methods is often lower than methods that incorporate multiple features and predicting protein functions may require a lot of time.

Predicting protein function from sequence and structure

David Lee, Oliver Redfern and Christine Orengo

Abstract | While the number of sequenced genomes continues to grow, experimentally verified functional annotation of whole genomes remains patchy. Structural genomics projects are yielding many protein structures that have unknown function. Nevertheless, subsequent experimental investigation is costly and time-consuming, which makes computational methods for predicting protein function very attractive. There is an increasing number of noteworthy methods for predicting protein function from sequence and structural data alone, many of which are readily available to cell biologists who are aware of the strengths and pitfalls of each available technique.

There are now >600 completely sequenced genomes of cellular organisms[1], contributing to more than five million unique protein sequences in the publicly accessible databases[2,3]. Experimental determination of the functions of all these proteins would be a hugely time-consuming and costly task and, in most instances, has not been carried out. Currently, approximately 20%, 7%, 10% and 1% of annotated proteins in the *Homo sapiens*, *Mus musculus*, *Drosophila melanogaster* and *Caenorhabditis elegans* genomes, respectively, have been experimentally characterized (traceable author source (TAS) annotations in Gene Ontology (GO))[4]. However, as the volume of data has increased, so too have the number and sophistication possible, but the scope of this review is to focus on what can be achieved by exploiting sequence and structural data using computational means alone.

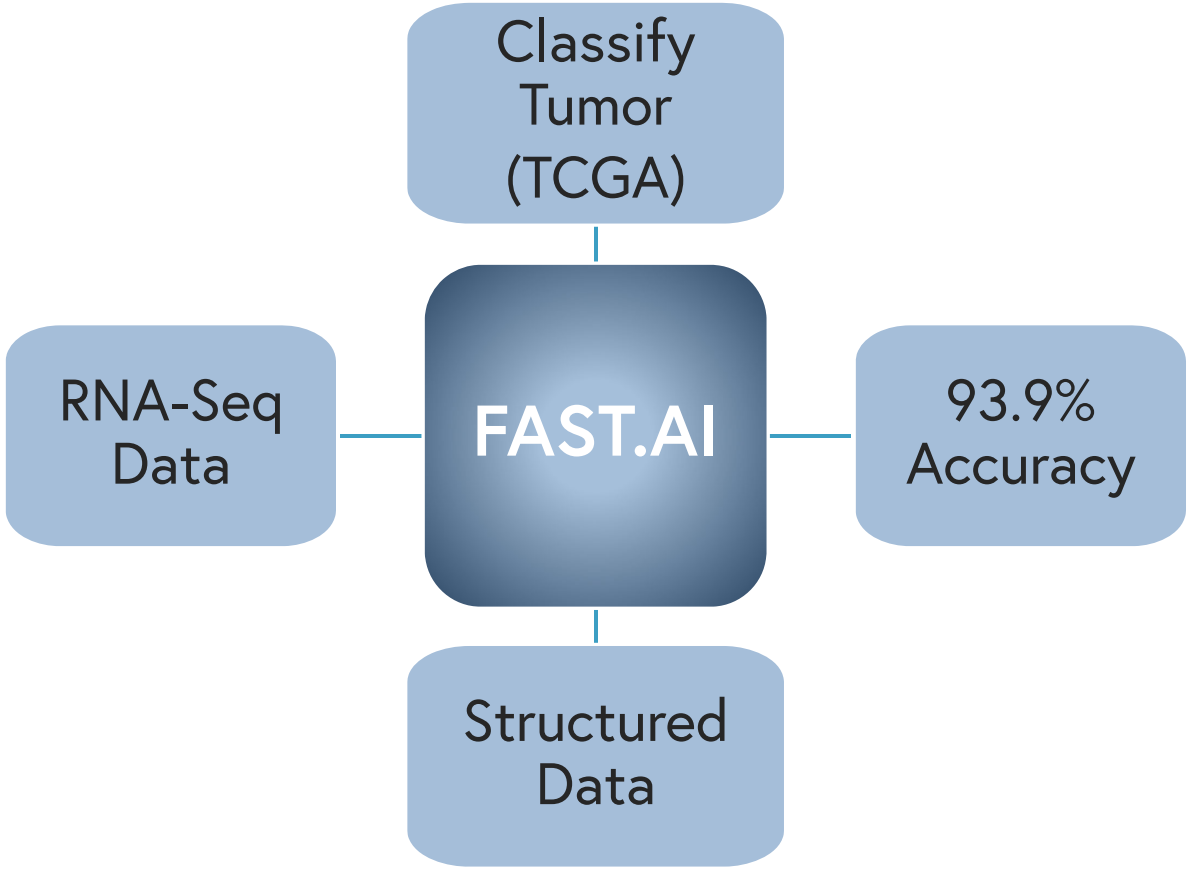
The most common and generally more accessible approach to function prediction is 'inheritance through homology' — that is, the knowledge that proteins with similar sequences frequently carry out similar functions. However, with the recent increase in the number of complete genome sequences, the possibility of establishing orthology has also increased. As discussed later in this review, this greatly improves the reliability of function transfer, although the coverage provided by identifiable orthologues tends to be small compared with that achieved

Maxat Kulmanov, Robert Hoehndorf, DeepGOPlus: improved protein function prediction from sequence, Bioinformatics, Volume 36, Issue 2, 15 January 2020, Pages 422–429

Lee, D., Redfern, O. & Orengo, C. Predicting protein function from sequence and structure. Nat Rev Mol Cell Biol 8, 995–1005 (2007). https://doi.org/10.1038/nrm2281

Tumor Classification Using Gene Expression Profiles

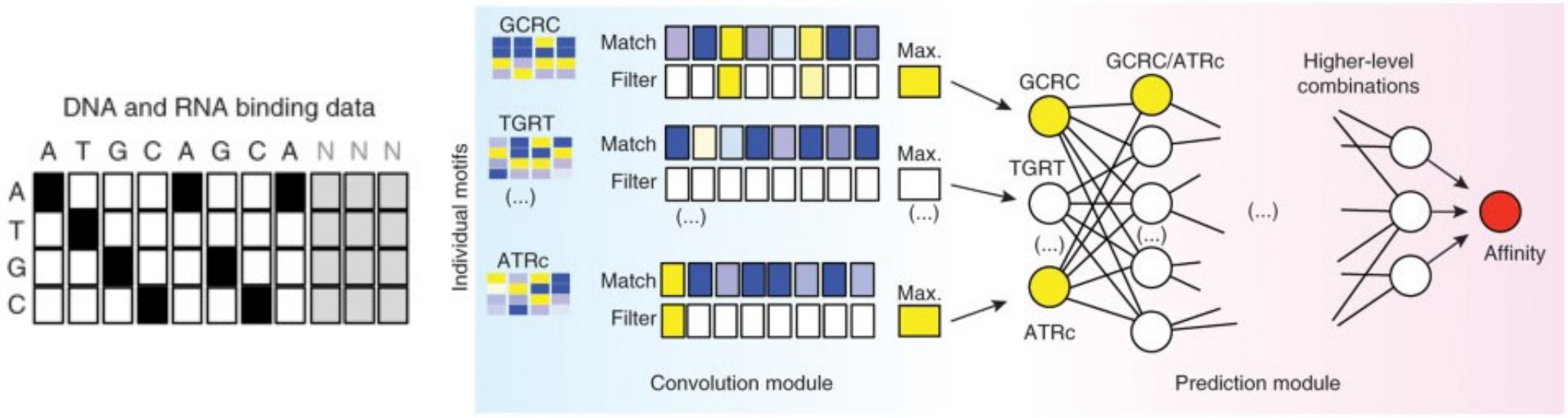
> **Tumor Classification using Gene Expression Data — poking at a problem using Fast.AI again**
>
> Alena Harley



https://medium.com/@alenaharley/tumor-classification-using-gene-expression-data-poking-at-a-problem-using-fast-ai-again-8633c2256c85

More Examples

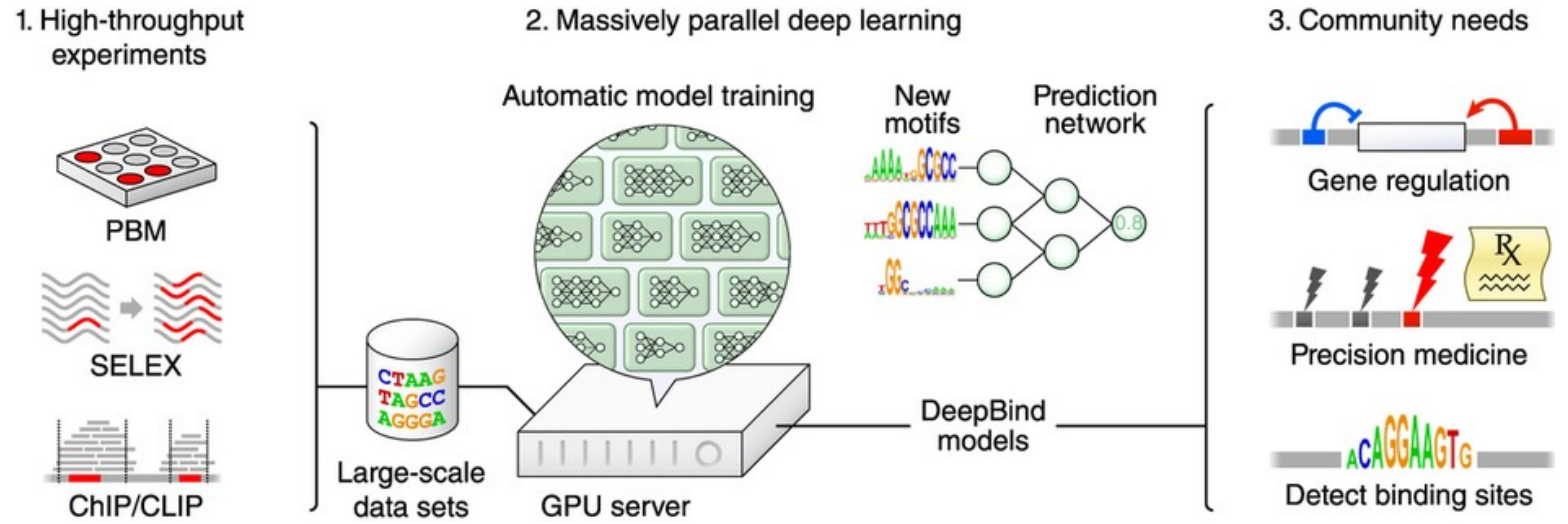
> **Deep learning for regulatory genomics**
>
> Yongjin Park & Manolis Kellis



Park, Y., Kellis, M. Deep learning for regulatory genomics. Nat Biotechnol 33, 825–826 (2015). https://doi.org/10.1038/nbt.3313

More Examples

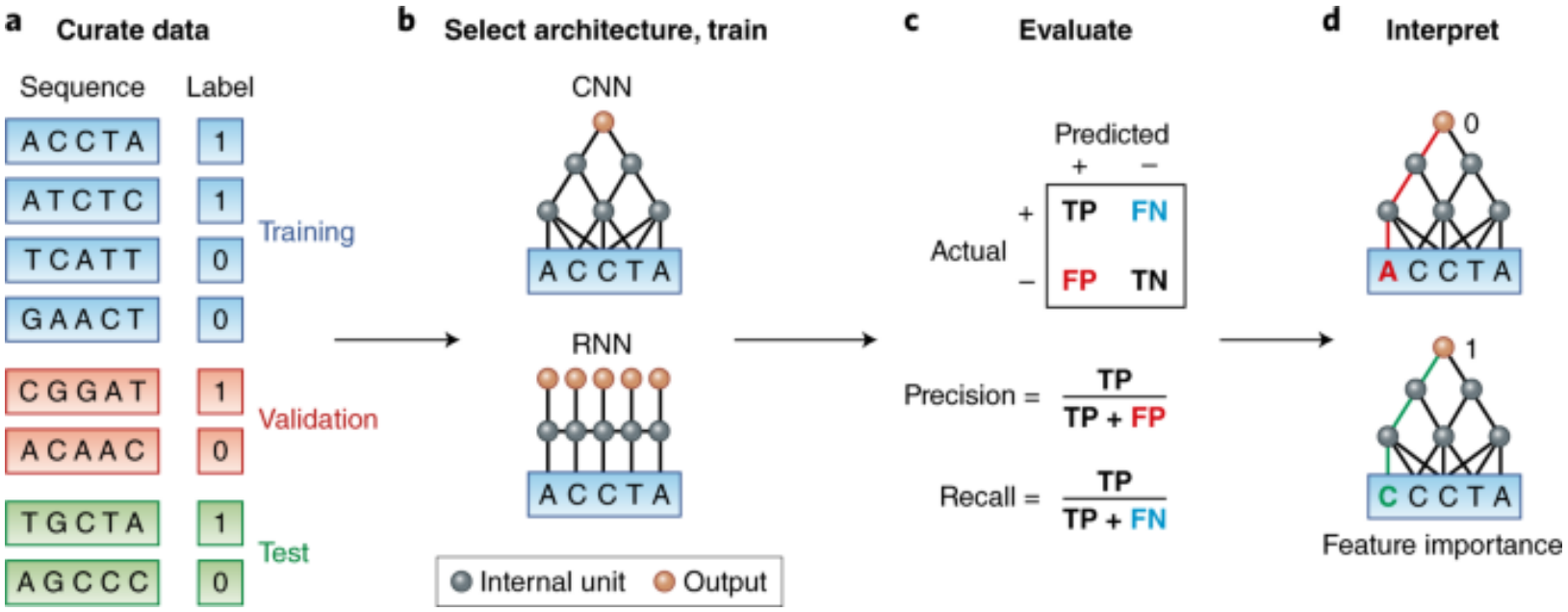
> **Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning**
>
> Babak Alipanahi, Andrew Delong, Matthew T Weirauch & Brendan J Frey



Alipanahi, B., Delong, A., Weirauch, M. et al. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. Nat Biotechnol 33, 831–838 (2015). https://doi.org/10.1038/nbt.3300

More Examples

> **A primer on deep learning in genomics**
>
> James Zou, Mikael Huss, Abubakar Abid, Pejman Mohammadi, Ali Torkamani & Amalio Telenti



Zou, J., Huss, M., Abid, A. et al. A primer on deep learning in genomics. Nat Genet 51, 12–18 (2019). https://doi.org/10.1038/s41588-018-0295-5

Public Resources for Bioinformatics

Different public databases for bioinformatics data

1000 Genomes Project

- The largest public catalogue of human variation and genotype data

1000 Genomes Release	Variants	Individuals	Populations
Phase 3	84.4 million	2,504	26
Phase 1	37.9 million	1,092	14
Pilot	14.8 million	179	4

https://www.internationalgenome.org

Ensembl Genomes

http://ensemblgenomes.org

Gene Expression Omnibus

https://www.ncbi.nlm.nih.gov/geo/

Gene Ontology Resource

http://geneontology.org

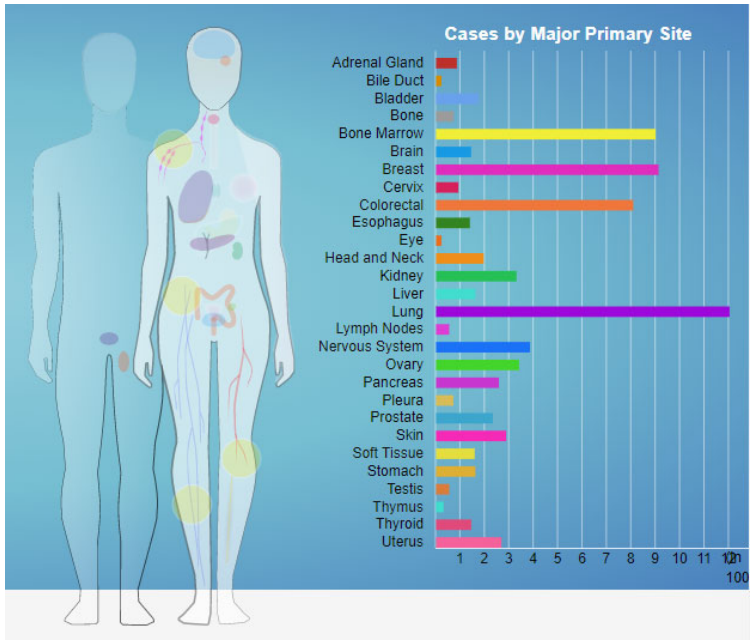
National Center for Biotechnology Information

https://www.ncbi.nlm.nih.gov

UniProt

https://www.uniprot.org

Genomic Data Commons Data Portal

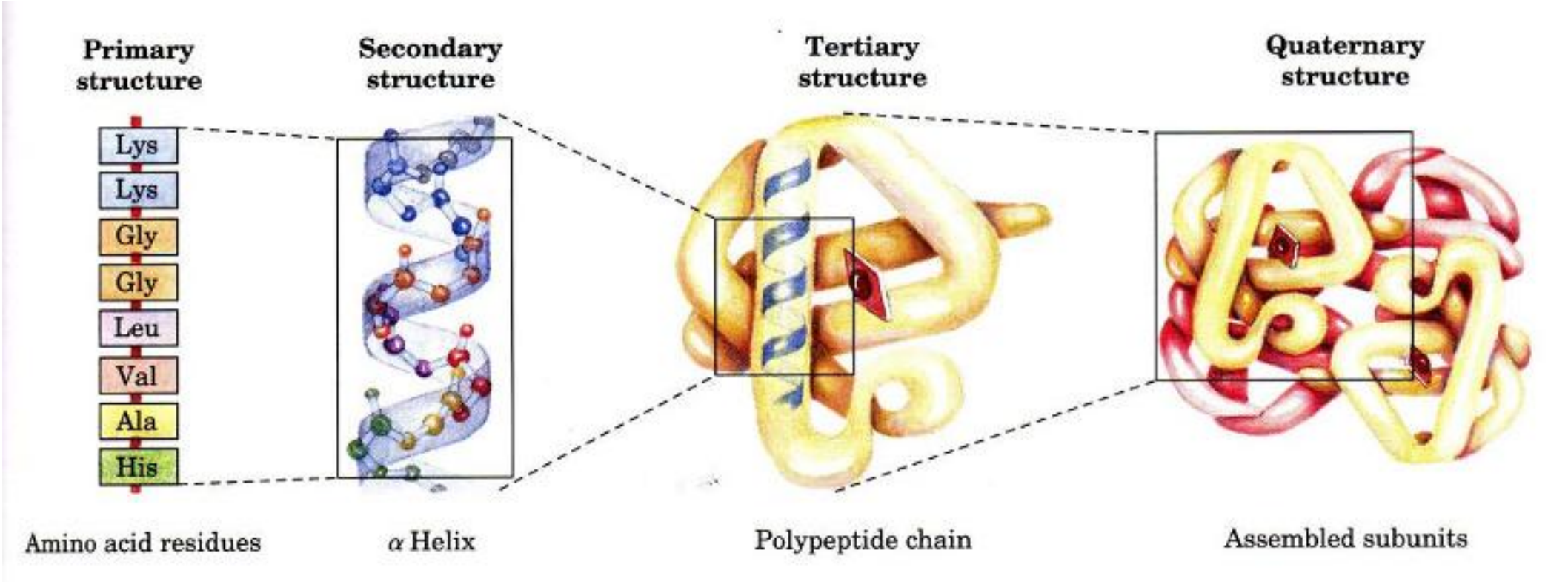


https://portal.gdc.cancer.gov/

Bioinformatics Data

What are different kinds of Bioinformatics data?

Protein Structure



https://comis.med.uvm.edu/VIC/coursefiles/MD540/MD540-Protein_Organization_10400_574581210/Protein-org/Protein_Organization_print.html

3D Structure of Protein

- 6WTT

wwPDB: 6WTT [Internet]. Doi.org. 2020. Available from: http://doi.org/10.2210/pdb6WTT/pdb

Protein Sequence

```
        10         20         30         40         50         60
MANLGCWMLV LFVATWSDLG LCKKRPKPGG WNTGGSRYPG QGSPGGNRYP PQGGGGWGQP

        70         80         90        100        110        120
HGGGWGQPHG GGWGQPHGGG WGQPHGGGWG QGGGTHSQWN KPSKPKTNMK HMAGAAAAGA

       130        140        150        160        170        180
VVGGLGGYML GSAMSRPIIH FGSDYEDRYY RENMHRYPNQ VYYRPMDEYS NQNNFVHDCV

       190        200        210        220        230        240
NITIKQHTVT TTTKGENFTE TDVKMMERVV EQMCITQYER ESQAYYQRGS SMVLFSSPPV

       250
ILLISFLIFL IVG
```

FASTA Format for Sequencing

Hosseini, M.; Pratas, D.; Pinho, A.J. A Survey on Data Compression Methods for Biological Sequences. Information 2016, 7, 56.

FASTA Format for Sequencing

- Protein sequence:

Title

Sequence

```
>sp|Q9NQ39|RS10L_HUMAN Putative 40S ribosomal protein S10-like
MLMPKKNRIAIHELLFKEGVMVAKKDVHMPKHPELADKNVPNLHVMKAMQSLKSRGCVKE
QFAWRHFYWYLTNEGSQYLRDYLHLPPEIVPATLHLPPEIVPATLHRSRPETGRPRPKGL
EGKRPARLTRREADRDTYRRCSVPPGADKKAEAGAGSATEFQFRGRCGRGRGQPPQ
```

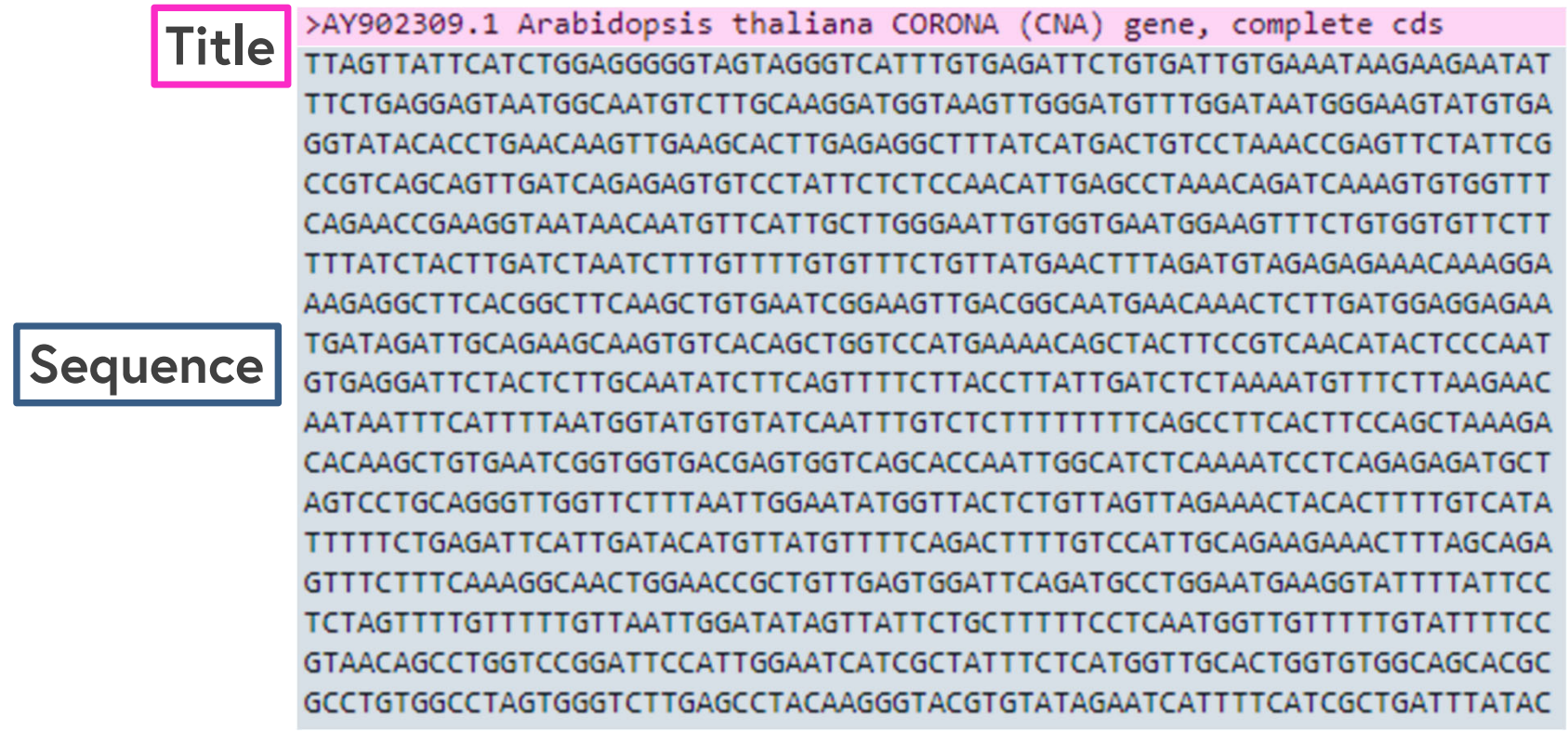
Alanine	A	Ala
Cysteine	C	Cys
Aspartic acid	D	Asp
Glutamic acid	E	Glu
Phenylalanine	F	Phe
Glycine	G	Gly
Histidine	H	His

Isoleucine	I	Ile
Lysine	K	Lys
Leucine	L	Leu
Methionine	M	Met
Asparagine	N	Asn
Pyrrolysine	O	Pyl
Proline	P	Pro

Glutamine	Q	Gln
Arginine	R	Arg
Serine	S	Ser
Threonine	T	Thr
Selenocysteine	U	Sec
Valine	V	Val
Tryptophan	W	Trp
Tyrosine	Y	Tyr

FASTA Format for Sequencing

- DNA sequence:



Gene Expression Level

GENE	GSM767976	GSM767977	GSM767978	GSM767979	GSM767980	GSM76798	GSM76798	GSM76798
MIR4640	11.24	11.47	10.95	11.54	11.3	11.3	11.79	11.95
RFC2	8.55	8.56	8.93	9.52	8.59	8.89	8.7	9.46
HSPA6	4.9	5.89	6.12	6.89	3.39	5.92	5.77	6.27
PAX8	2.83	2.83	2.43	2.75	2.79	2.84	2.18	2.84
GUCA1A	2.08	2.08	2.07	2.07	2.09	2.09	2.07	2.08
MIR5193	7.45	9.8	6.9	6.91	6.69	7.85	8.2	9.1
THRA	7.73	4.62	8.11	7.71	7.91	4.63	8.51	5.74
PTPN21	3.07	2.31	4.06	2.67	2.67	2.61	2.42	2.56
CCL5	2.91	6.6	2.62	2.51	2.68	2.74	2.45	2.77
CYP2E1	4.74	6.58	4.8	5.9	2.61	4.65	4.63	8.95
EPHB3	2.15	2.16	2.96	3.22	2.16	2.16	2.22	3.41
ESRRA	9.65	10.59	9.31	10.16	9.99	10.94	9.76	10.49
CYP2A6	2	2.01	2	2	2.01	2.01	2	2.01
SCARB1	6.36	6.44	5.47	5.62	5.74	5.69	4.87	5.56
TTLL12	10.11	10.89	11.3	11.63	10.89	11.52	10.71	11.33

Gene Expression Level

GENE	GSM767976	GSM767977	GSM767978	GSM767979	GSM767980	GSM76798	GSM76798	GSM76798
MIR4640	11.24	11.47	10.95	11.54	11.3	11.3	11.79	11.95
RFC2	8.55	8.56	8.93	9.52	8.59	8.89	8.7	9.46
HSPA6	4.9	5.89	6.12	6.89	3.39	5.92	5.77	6.27
PAX8	2.83	2.83	2.43	2.75	2.79	2.84	2.18	2.84
GUCA1A	2.08	2.08	2.07	2.07	2.09	2.09	2.07	2.08
MIR5193	7.45	9.8	6.9	6.91	6.69	7.85	8.2	9.1
THRA	7.73	4.62	8.11	7.71	7.91	4.63	8.51	5.74
PTPN21	3.07	2.31	4.06	2.67	2.67	2.61	2.42	2.56
CCL5	2.91	6.6	2.62	2.51	2.68	2.74	2.45	2.77
CYP2E1	4.74	6.58	4.8	5.9	2.61	4.65	4.63	8.95
EPHB3	2.15	2.16	2.96	3.22	2.16	2.16	2.22	3.41
ESRRA	9.65	10.59	9.31	10.16	9.99	10.94	9.76	10.49
CYP2A6	2	2.01	2	2	2.01	2.01	2	2.01
SCARB1	6.36	6.44	5.47	5.62	5.74	5.69	4.87	5.56
TTLL12	10.11	10.89	11.3	11.63	10.89	11.52	10.71	11.33

Gene Expression Level

GENE	GSM767976	GSM767977	GSM767978	GSM767979	GSM767980	GSM76798	GSM76798	GSM76798
MIR4640	11.24	11.47	10.95	11.54	11.3	11.3	11.79	11.95
RFC2	8.55	8.56	8.93	9.52	8.59	8.89	8.7	9.46
HSPA6	4.9	5.89	6.12	6.89	3.39	5.92	5.77	6.27
PAX8	2.83	2.83	2.43	2.75	2.79	2.84	2.18	2.84
GUCA1A	2.08	2.08	2.07	2.07	2.09	2.09	2.07	2.08
MIR5193	7.45	9.8	6.9	6.91	6.69	7.85	8.2	9.1
THRA	7.73	4.62	8.11	7.71	7.91	4.63	8.51	5.74
PTPN21	3.07	2.31	4.06	2.67	2.67	2.61	2.42	2.56
CCL5	2.91	6.6	2.62	2.51	2.68	2.74	2.45	2.77
CYP2E1	4.74	6.58	4.8	5.9	2.61	4.65	4.63	8.95
EPHB3	2.15	2.16	2.96	3.22	2.16	2.16	2.22	3.41
ESRRA	9.65	10.59	9.31	10.16	9.99	10.94	9.76	10.49
CYP2A6	2	2.01	2	2	2.01	2.01	2	2.01
SCARB1	6.36	6.44	5.47	5.62	5.74	5.69	4.87	5.56
TTLL12	10.11	10.89	11.3	11.63	10.89	11.52	10.71	11.33

How to Collect Data?

How to collect Bioinformatics data from public resources?

Public Resources

NCBI	https://www.ncbi.nlm.nih.gov/
UniProt	https://www.uniprot.org/
GENEONTOLOGY Unifying Biology	http://geneontology.org/
WORLDWIDE PDB PROTEIN DATA BANK	http://www.wwpdb.org/

Data Collection from NCBI

- From NCBI query

Data Collection from GEO

Summary ▾ 20 per page ▾ Sort by Subgroup effect ▾ Send to: ▾

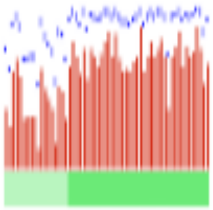
Search results

Items: 1 to 20 of 1050133

<< First < Prev Page 1 of 52507 Next > Last >>

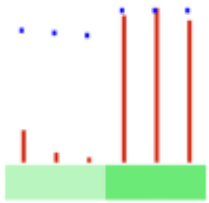
☐ 1. KRT7 - Non-small **lung cancer** subtypes: adenocarcinoma and squamous cell carcinoma

Annotation: KRT7, keratin 7
Organism: Homo sapiens
Reporter: GPL570, 209016_s_at (ID_REF), GDS3627, 3855 (Gene ID), BC002700
DataSet type: Expression profiling by array, transformed count, 58 samples
ID: 62794131
GEO DataSets Gene Profile neighbors Chromosome neighbors Homologene neighbors



☐ 2. ROBO1 - Cisplatin-resistant non-small cell **lung cancer** cell line

Annotation: ROBO1, roundabout guidance receptor 1
Organism: Homo sapiens
Reporter: GPL6244, 8088919 (ID_REF), GDS5247, NM_001145845, NM_002941, NM_133631, AF040990, BC112336, BC115020, BC115022, BC157861, BC171855, chr3:78646390-79639061 (SPOT ID)
DataSet type: Expression profiling by array, transformed count, 6 samples
ID: 119035613
GEO DataSets Gene Profile neighbors Chromosome neighbors Homologene neighbors

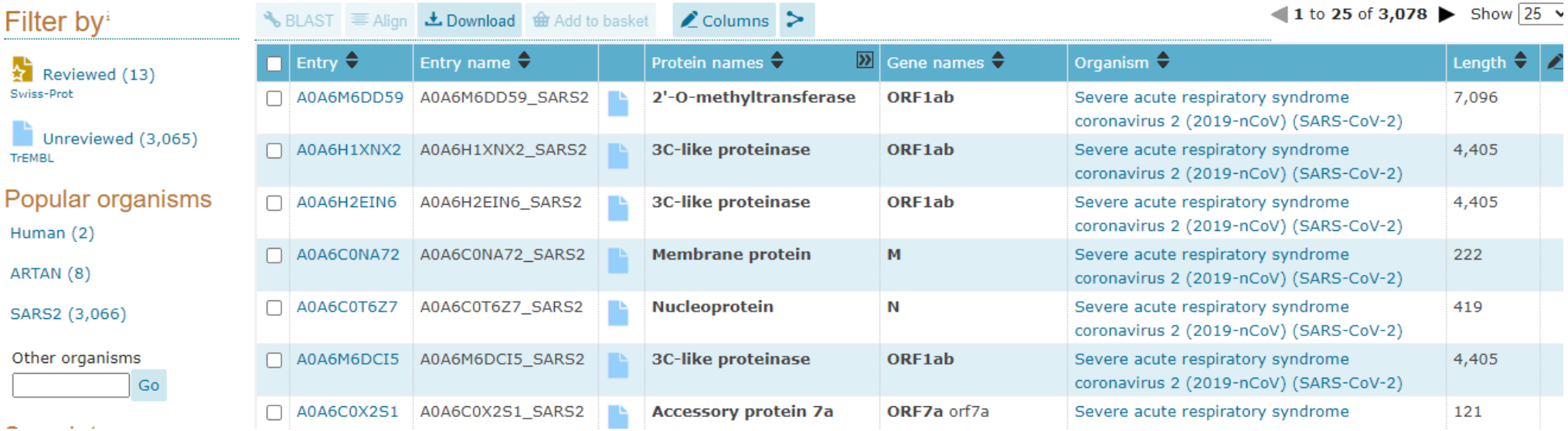


Data Collection from GDC

Data Category	Files (n=56)
Sequencing Reads	3
Transcriptome Profiling	3
Simple Nucleotide Variation	16
Copy Number Variation	7
DNA Methylation	2
Clinical	8
Biospecimen	17

Experimental Strategy	Files (n=56)
Diagnostic Slide	1
Tissue Slide	4
WXS	18
RNA-Seq	4
Genotyping Array	7
Methylation Array	2

Data Collection from UniProt



Filter by

Reviewed (13)
Swiss-Prot

Unreviewed (3,065)
TrEMBL

Popular organisms

Human (2)

ARTAN (8)

SARS2 (3,066)

Other organisms
Go

BLAST | Align | Download | Add to basket | Columns

1 to 25 of 3,078 Show 25

Entry	Entry name	Protein names	Gene names	Organism	Length
A0A6M6DD59	A0A6M6DD59_SARS2	**2'-O-methyltransferase**	**ORF1ab**	Severe acute respiratory syndrome coronavirus 2 (2019-nCoV) (SARS-CoV-2)	7,096
A0A6H1XNX2	A0A6H1XNX2_SARS2	**3C-like proteinase**	**ORF1ab**	Severe acute respiratory syndrome coronavirus 2 (2019-nCoV) (SARS-CoV-2)	4,405
A0A6H2EIN6	A0A6H2EIN6_SARS2	**3C-like proteinase**	**ORF1ab**	Severe acute respiratory syndrome coronavirus 2 (2019-nCoV) (SARS-CoV-2)	4,405
A0A6C0NA72	A0A6C0NA72_SARS2	**Membrane protein**	**M**	Severe acute respiratory syndrome coronavirus 2 (2019-nCoV) (SARS-CoV-2)	222
A0A6C0T6Z7	A0A6C0T6Z7_SARS2	**Nucleoprotein**	**N**	Severe acute respiratory syndrome coronavirus 2 (2019-nCoV) (SARS-CoV-2)	419
A0A6M6DCI5	A0A6M6DCI5_SARS2	**3C-like proteinase**	**ORF1ab**	Severe acute respiratory syndrome coronavirus 2 (2019-nCoV) (SARS-CoV-2)	4,405
A0A6C0X2S1	A0A6C0X2S1_SARS2	**Accessory protein 7a**	**ORF7a** orf7a	Severe acute respiratory syndrome	121

Thank You for Your Attention

References

- Campbell M. Transcription vs Translation Worksheet [Internet]. Technology Networks. 2019. Available from: https://www.technologynetworks.com/genomics/articles/transcription-vs-translation-worksheet-323080
- Rao M, Van Vleet T, Ciurlionis R, Buck W, Mittelstadt S, Blomme E et al. Comparison of RNA-Seq and Microarray Gene Expression Platforms for the Toxicogenomic Evaluation of Liver From Short-Term Rat Toxicity Studies. Frontiers in Genetics. 2019;9.

References

- The Cost of Sequencing a Human Genome [Internet]. Genome.gov. Available from: https://www.genome.gov/about-genomics/fact-sheets/Sequencing-Human-Genome-cost
- Types of variants | Garvan Institute of Medical Research [Internet]. Garvan Institute of Medical Research. Available from: https://www.garvan.org.au/research/kinghorn-centre-for-clinical-genomics/learn-about-genomics/for-gp/genetics-refresher-1/types-of-variants

References

- Poplin R, Chang P, Alexander D, Schwartz S, Colthurst T, Ku A et al. A universal SNP and small-indel variant caller using deep neural networks. Nature Biotechnology. 2018;36(10):983-987.
- Kulmanov M, Hoehndorf R. DeepGOPlus: improved protein function prediction from sequence. Bioinformatics. 2019;.

References

- Lee D, Redfern O, Orengo C. Predicting protein function from sequence and structure. Nature Reviews Molecular Cell Biology. 2007;8(12):995-1005.
- Harley A. Tumor Classification using Gene Expression Data—poking at a problem using Fast.AI again [Internet]. Medium. 2018. Available from: https://medium.com/@alenaharley/tumor-classification-using-gene-expression-data-poking-at-a-problem-using-fast-ai-again-8633c2256c85

References

- Park Y, Kellis M. Deep learning for regulatory genomics. Nature Biotechnology. 2015;33(8):825-826.
- Alipanahi B, Delong A, Weirauch M, Frey B. Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning. Nature Biotechnology. 2015;33(8):831-838.

References

- Zou J, Huss M, Abid A, Mohammadi P, Torkamani A, Telenti A. A primer on deep learning in genomics. Nature Genetics. 2018;51(1):12-18.
- 1000 Genomes | A Deep Catalog of Human Genetic Variation [Internet]. Internationalgenome.org. Available from: https://www.internationalgenome.org/

Image Sources

- Chromosome String free icon is made by Freepik from Flaticon. Available from: https://www.flaticon.com/free-icon/chromosome-string_3254256
- DNA Chromosome free icon is made by Freepik from Flaticon. Available from: https://www.flaticon.com/free-icon/dna-chromosome_17883

Image Sources

- Mail Filter free icon is made by Freepik from Flaticon. Available from: https://www.flaticon.com/free-icon/mail-filter_32016
- Smart Car free icon is made by Freepik from Flaticon. Available from: https://www.flaticon.com/free-icon/smart-car_2228504

Image Sources

- Chess free icon is made by Smashicons from Flaticon. Available from: https://www.flaticon.com/free-icon/chess_910070
- Hashtag free icon is made by wanicon from Flaticon. Available from: https://www.flaticon.com/free-icon/hashtag_2835618

Image Sources

- National Human Genome Research Institute. The structure hierarchy of Proteins [Internet]. Available from: https://commons.wikimedia.org/wiki/File:Protein-structure.png#/media/File:Protein-structure.png
- Flask free icon is made by Freepik from Flaticon. Available from: https://www.flaticon.com/free-icon/flask_3655543

Image Sources

- Artificial Intelligence free icon is made by Eucalyp from Flaticon. Available from: https://www.flaticon.com/free-icon/artificial-intelligence_2752802

Image Sources

- Levels of Protein Organization [Internet]. Comis.med.uvm.edu. 2014. Available from: https://comis.med.uvm.edu/VIC/coursefiles/MD540/MD540-Protein_Organization_10400_574581210/Protein-org/Protein_Organization_print.html

Image Sources

- wwPDB: 6WTT [Internet]. Doi.org. 2020. Available from: http://doi.org/10.2210/pdb6WTT/pdb
- Hosseini M, Pratas D, Pinho A. A Survey on Data Compression Methods for Biological Sequences. Information. 2016;7(4):56.