PIPADOR REVISION

MRCOG Part 1 Revision Notes

Bioinformatics

MOLECULAR BIOLOGY • LECTURE 100

VISUAL EXAMINATION REVISION NOTES

2026 Edition

MRCOG Part 1 Revision Notes

Lecture 100: Bioinformatics

Molecular biology • Visual revision edition — Lecture 100

COLOUR CODE USED THROUGHOUT THESE NOTES

ALIGN / SEQUENCE	CLINICAL / PATHOGENIC	CALLING / OMICS	ANNOTATE / CANCER	POPULATION / FILTER

LEARNING OBJECTIVES — tick off as you go

- ☐ Follow the pipeline from sequencing reads to a clinical report.
- ☐ Know reference genomes, coordinates and core tools (BLAST, alignment).
- ☐ Name the key sequence, clinical, population and cancer databases.
- ☐ Interpret a variant with ACMG/ACGS evidence and in silico predictors.
- ☐ Describe UK genomic infrastructure and the ethics of genomic data.

1 The Pipeline

IN ONE LINE: *Bioinformatics turns billions of sequence reads into one sentence on a report: align → call → annotate → filter → classify.*

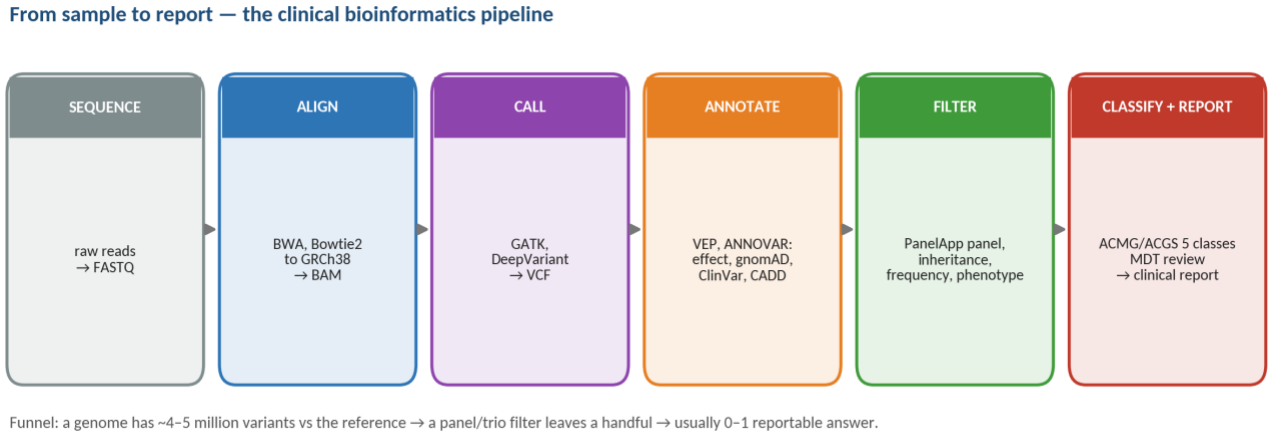


Figure 1. Clinical bioinformatics pipeline.

File / tool	What it is
FASTQ	Raw reads + quality scores
BAM	Reads aligned to the reference (BWA, Bowtie2)
VCF	List of variant calls (GATK; DeepVariant uses deep learning; MuTect/Strelka for somatic)
Annotation	VEP / ANNOVAR add consequence, frequency, ClinVar, predictors
Predictors	SIFT, PolyPhen-2, CADD, REVEL (missense), AlphaMissense (AI, 2023), SpliceAI — supporting evidence only

2 References, Coordinates and Tools

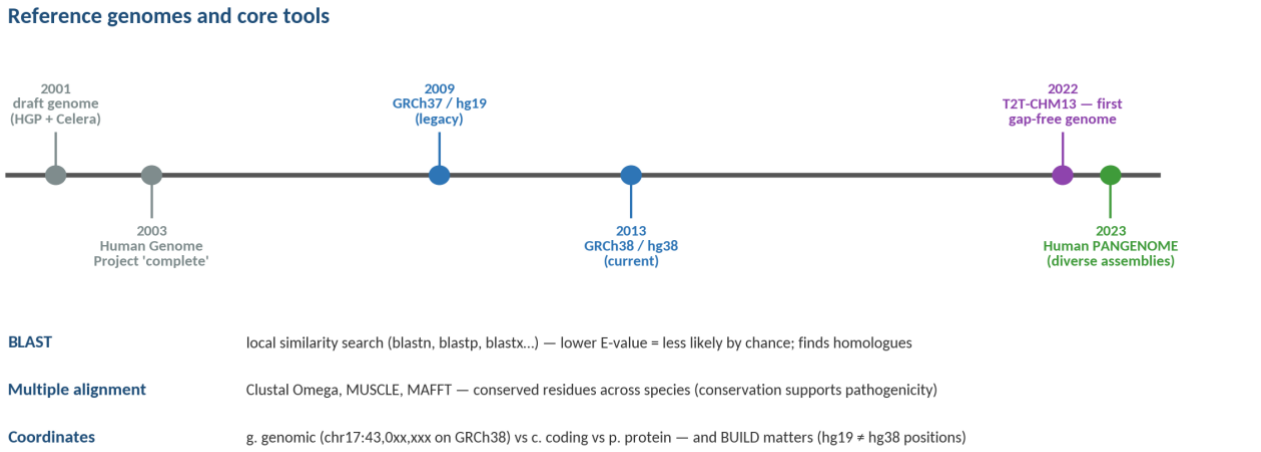


Figure 2. Reference genomes and tools.

3 Databases

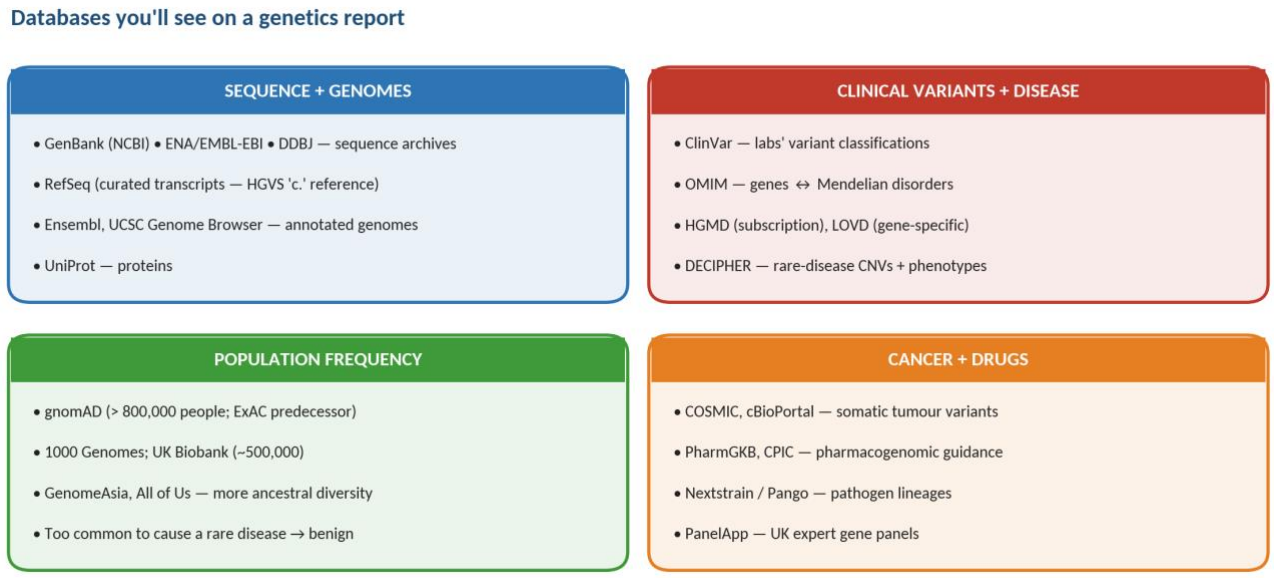


Figure 3. Key databases.

MNEMONIC / MEMORY HOOK

'gnomAD = how COMMON; ClinVar = what OTHERS said; OMIM = which DISEASE.'

4 Variant Interpretation

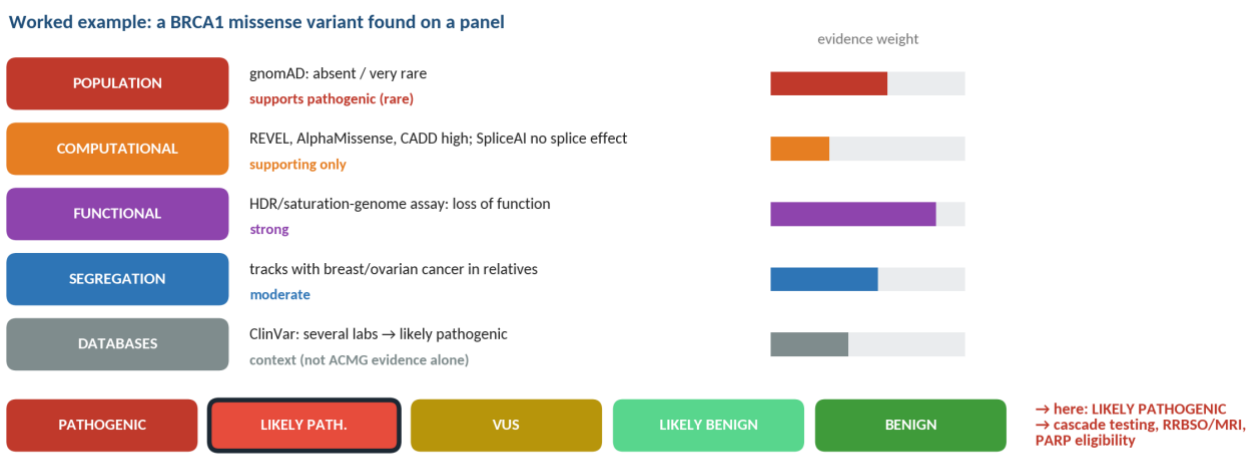


Figure 4. Interpreting a variant.

5 UK Infrastructure

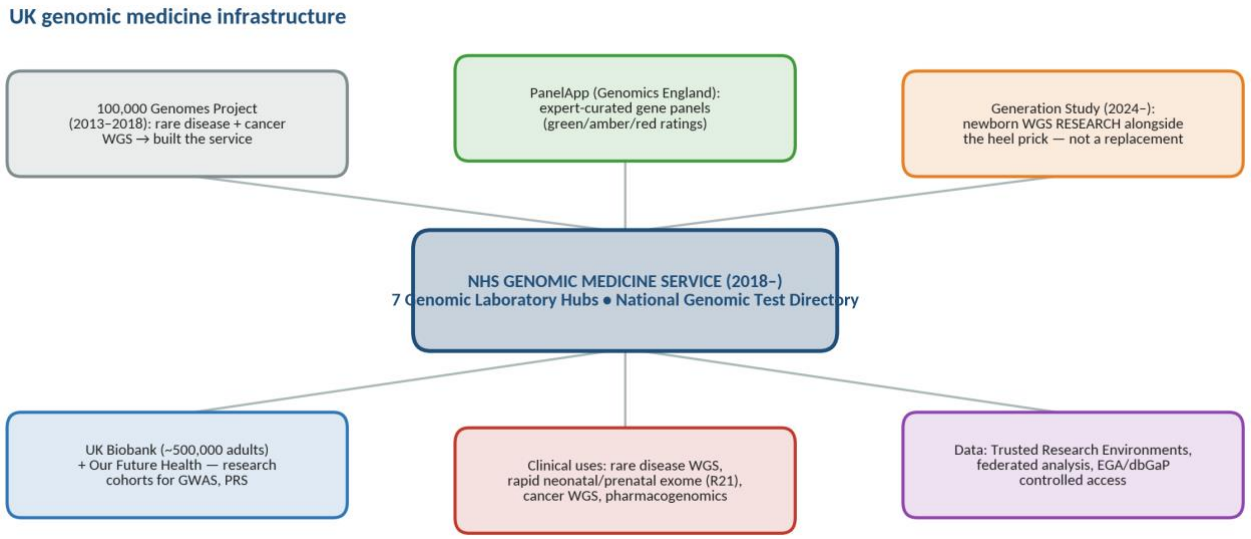


Figure 5. UK genomic medicine.

6 Applications and Ethics

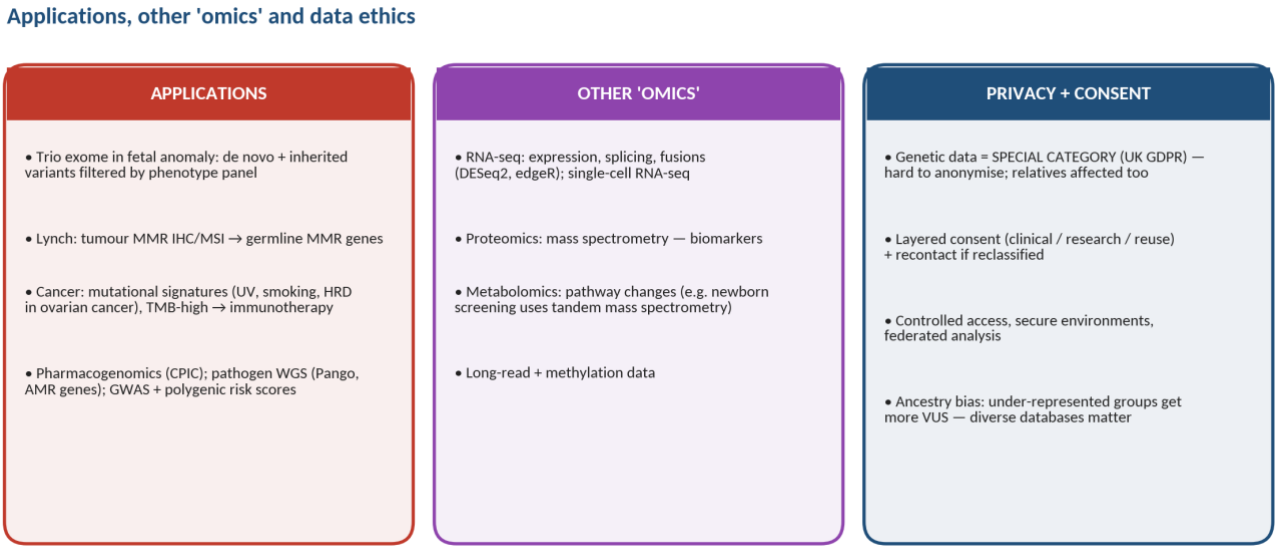


Figure 6. Applications, omics and data ethics.

★ **PEARL**

GRCh38 = current reference (GRCh37 legacy; T2T-CHM13 first complete). Check the build before comparing positions.

A variant common in gnomAD can't cause a rare severe disease. In silico predictions are only supporting evidence. Never act on a VUS.

7 MRCOG Exam Pearls — Last-Minute Page

Numbers to know

GRCh38 Current reference	**2022** T2T complete genome	**> 800,000** People in gnomAD	**~500,000** UK Biobank
7 NHS Genomic Lab Hubs	**2018** GMS launch / 100k done	**5** Variant classes	**~4–5 million** Variants per genome

HIGH-YIELD CHECKLIST

- ☐ FASTQ → BAM (BWA) → VCF (GATK) → VEP → PanelApp filter → ACMG/ACGS.
- ☐ GRCh38 current; GRCh37 legacy; T2T-CHM13 2022; pangenome.
- ☐ g. / c. / p. coordinates; build matters.
- ☐ BLAST (E-value); Clustal/MUSCLE/MAFFT for conservation.
- ☐ GenBank/ENA/DDBJ, RefSeq, Ensembl, UCSC, UniProt.
- ☐ ClinVar, OMIM, HGMD, LOVD, DECIPHER; gnomAD, UK Biobank.
- ☐ Predictors: SIFT, PolyPhen-2, CADD, REVEL, AlphaMissense, SpliceAI.
- ☐ NHS GMS (7 GLHs, Test Directory), PanelApp, 100k Genomes, Generation Study.
- ☐ PharmGKB/CPIC; COSMIC; TMB; signatures; Pango.
- ☐ Genetic data = special category; layered consent; ancestry bias.

Self-assessment — cover the right-hand column

Question	Answer
Current human reference build?	GRCh38 (hg38)
First gap-free human genome?	T2T-CHM13 (2022)
Database of population allele frequencies?	gnomAD
Database of submitted clinical classifications?	ClinVar
Catalogue of genes and Mendelian disorders?	OMIM
What does a low BLAST E-value mean?	Match unlikely to be due to chance
File format for variant calls?	VCF
UK resource defining NHS gene panels?	PanelApp (Genomics England)
Weight of in silico predictions in ACMG?	Supporting only
Legal status of genetic data in UK GDPR?	Special category data

8 References

- Richards S et al. ACMG/AMP standards for interpretation of sequence variants. Genet Med 2015; ACGS Best Practice Guidelines.
- Nurk S et al. The complete sequence of a human genome (T2T). Science 2022.
- NHS England. National Genomic Test Directory; Genomics England PanelApp.