**Week 7 – Clinical Genomics**

*Reading Material*

Here are links to papers, projects and tool related to sequencing of human samples in a clinical context and interpreting the resulting data to attempt to identify disease-related variants.

**WGS500 Project**

The WGS500 project (Whole Genome Sequencing of 500 individuals) is an exemplar of an initiative that used genome sequencing to explore a number of different diseases and conditions. Undertaken at the Wellcome Centre for Human Genetics (WHG) about 10 years ago, WGS500 was the first large-scale human sequencing project. It was the precursor to the 100,000 genomes project, the UK’s national effort to sequence at very large scale to investigate rare disease and cancer.

WGS500 was led by:

[Prof Jenny Taylor](https://www.well.ox.ac.uk/people/jenny-taylor): Co-theme Leader of the Genomic Medicine Theme of the [Oxford Biomedical Research Centre (BRC)](http://oxfordbrc.nihr.ac.uk/)

[Prof Gil McVean](https://en.wikipedia.org/wiki/Gilean_McVean), then the Head of the Bioinformatics and Statistical Genetics Core at WHG, former director of the [Big Data Institute](https://www.bdi.ox.ac.uk/) and co-founder of [Genomics plc](https://www.genomicsplc.com/)

[Prof Sir Peter Donnelly](https://en.wikipedia.org/wiki/Peter_Donnelly), former director of WHG, co-founder and CEO of [Genomics plc](https://www.genomicsplc.com/)

The WGS500 'umbrella' paper summarises findings from the project as a whole:

[*Factors influencing success of clinical genome sequencing across a broad spectrum of disorders*](https://pubmed.ncbi.nlm.nih.gov/25985138/)

and this is an example paper showing how the samples were used in specific investigations:

[*Whole-genome sequencing of bladder cancers reveals somatic CDKN1A mutations and clinicopathological associations with mutation burden*](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4010643/)

**Subsequent Large-scale Sequencing Projects**

Human sequencing projects at increasingly larger scales were, and continue to be, conducted in efforts to make progress in applying genomic information to routine healthcare.

**1000 Genomes Project**

<https://www.internationalgenome.org/1000-genomes-summary>

**The 100 000 Genomes Project: bringing whole genome sequencing to the NHS**

<https://www.bmj.com/content/361/bmj.k1687>

<https://www.genomicsengland.co.uk/initiatives/100000-genomes-project>

**Our Future Health (5 million genomes)**

<https://ourfuturehealth.org.uk/>

**Tools**

There is a proliferation of tools for annotating VCF files to provide information such as the predicted pathogenicity of a given variant. They have moved beyond simply querying databases of pre-built values and can now generate context-aware annotations on the fly:

ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data

https://academic.oup.com/nar/article/38/16/e164/1749458

The Ensembl Variant Effect Predictor

https://genomebiology.biomedcentral.com/articles/10.1186/s13059-016-0974-4?report=reader

MutationTaster evaluates disease-causing potential of sequence alterations

https://www.nature.com/articles/nmeth0810-575

There are also tools to assist in viewing and filtering lists of mutations based on these and other annotations:

BrowseVCF: a web-based application and workflow to quickly prioritize disease-causative variants in VCF files

https://academic.oup.com/bib/article/18/5/774/2562769

VarSifter: Visualizing and analyzing exome-scale sequence variation data on a desktop computer

https://academic.oup.com/bioinformatics/article/28/4/599/213473