README.md `cancer_polymorphisms`

This `cancer_polymorphisms` repository describes mapping of Illumina sequencing reads from a glioblastoma sample to the standard human reference sequence genome GRCh38.

The analysis described below can be reproduced by issuing the command:

```
snakemake -np mapped_reads/glioblastoma.bam
```

1. Downloading reference

The GRCh38 reference genome was obtained from the Genome Reference Consortium FTP site (https://ftp.ncbi.nlm.nih.gov/genomes/all/GCA/000/001/405/GCA_000001405.15_GRCh38/), and all files placed in the `reference` subdirectory.

2. Obtaining sequencing reads

Sequencing reads from a human glioblastoma tissue sample were provided by Dr A B Normal. The surgery, diagnosis, treatment and other activities were approved by the institutional research ethics board (Ethical Review Committee of the Medical Faculty of the University of Strathclyde, #144/08-strath; registration numbers: IGOR0021320, IBR00711750) in accordance with the Helsinki Declaration. Tissue preparation and sequencing are described in

- Normal AB, et al. RNA sequencing of glioblastoma tissue slices. Glioblastoma Res. 2021 Feb;2(4):180-193. doi:10.1022/2111-5363.23353

The sequence reads were provided as a FASTQ format file `glioblastoma.fastq`, which were placed in the `data` subdirectory.

3. Mapping sequencing reads

Sequencing reads were mapped using the `bwa` tool (v0.7.17) and default settings, and output was converted to `.bam` format using `samtools` (v1.15.1) as follows:

```
bwa mem reference/GCA_000001405.15_GRCh38_genomic.fna
data/glioblastoma.fastq | samtools view -Sb - >
results/glioblastoma_map_GRCh38.bam
```