I am applying deep learning to harmonise high throughput sequencing data from different modalities. From this data, it will be possible to identify global patterns of gene regulation.

Size of data is in the order of $10^{11}$ samples per dataset across $10^2$ different studies.