# Tyrone Chen

PhD Student
Computational Biology
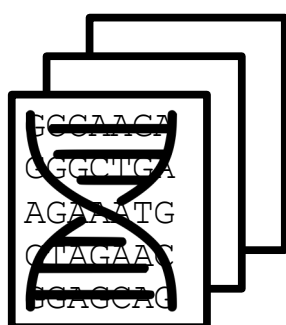School of Biological Sciences, Monash University, Australia

## Using deep learning to unravel hidden signals in big biological data
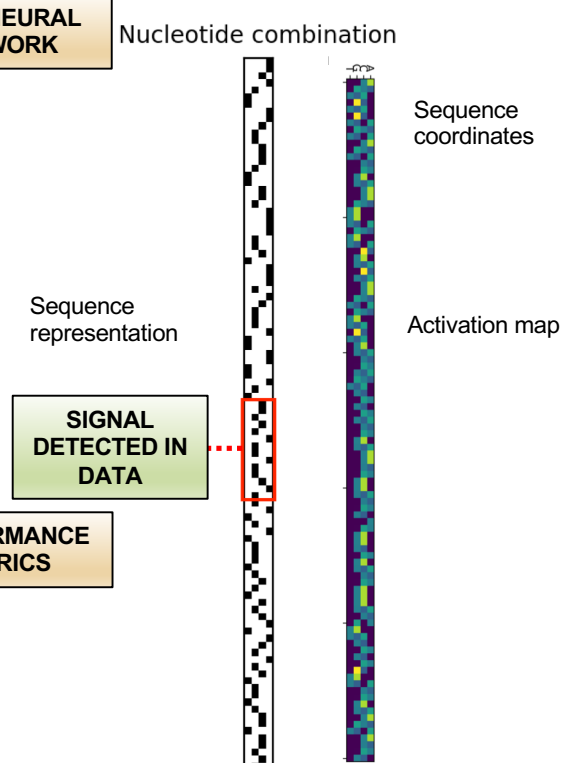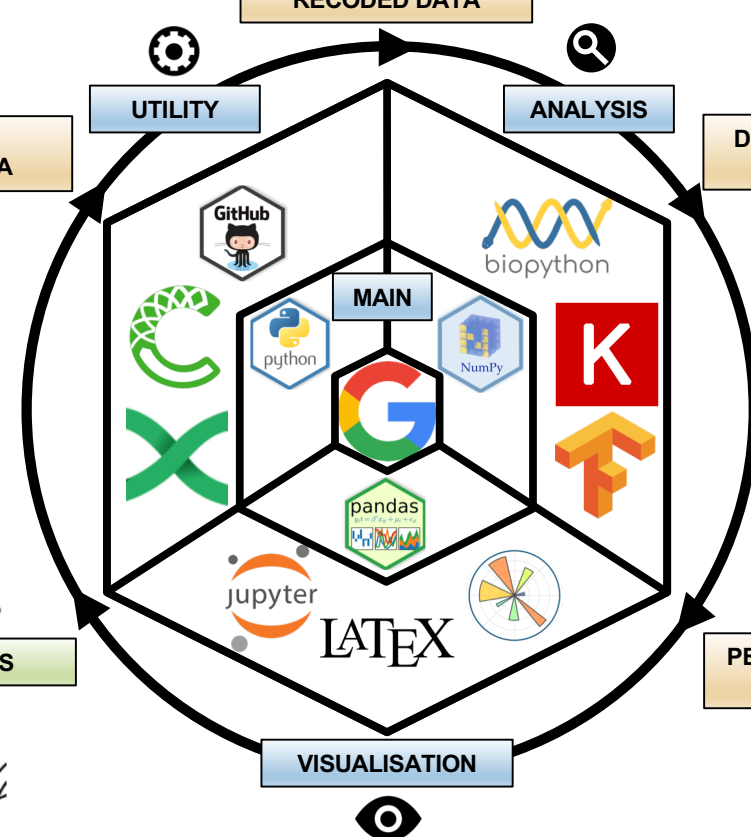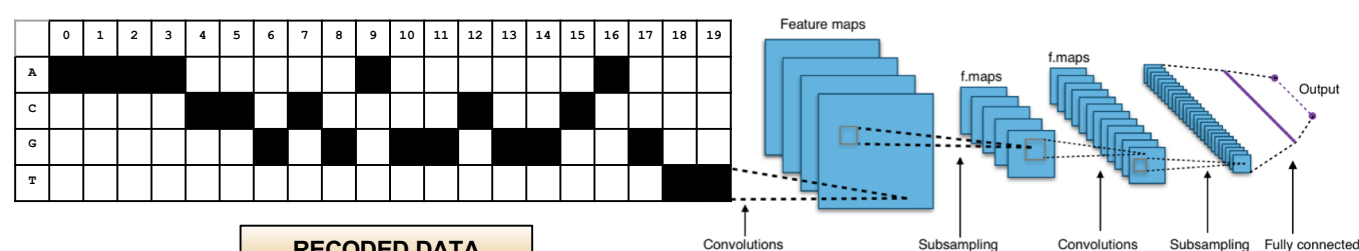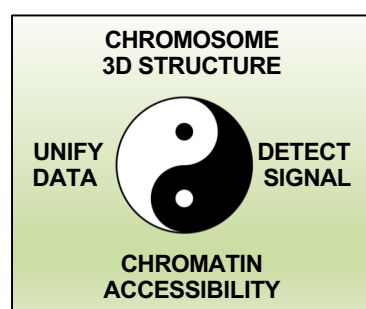
**MY RESEARCH:**

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.

**MY DIGITAL TOOLKIT:**



**RECODED DATA**

**UTILITY**

**ANALYSIS**

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

**INPUT BIG DATA**

**DEEP NEURAL NETWORK**

Nucleotide combination

Sequence coordinates

**CHROMOSOME 3D STRUCTURE**

UNIFY DATA — DETECT SIGNAL

CHROMATIN ACCESSIBILITY

**MAIN**

Sequence representation

Activation map

**SIGNAL DETECTED IN DATA**

**RESULTS**

**PERFORMANCE METRICS**

**VISUALISATION**

**VISUALISATION OF SIGNAL DETECTED IN DATA**

A biological signal is detected in the data. This signal is tiny relative to the search space and regulates many critical biological processes.
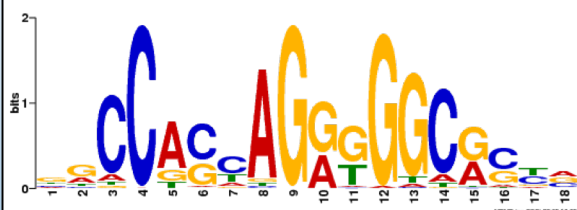
**HOW I USE MY FAVOURITE TOOL TO ENHANCE MY RESEARCH:**

**Ease of use:** A combination of TensorFlow and the higher level Keras API automates and greatly simplifies the process of developing and testing neural networks.

**Scalability:** Methods written in TensorFlow and Keras are scalable and can be run on a laptop for interactive testing, or expand to use a high performance compute cluster with minimal changes to code.

**Flexibility:** TensorFlow and Keras allow users to customise and modify their own models as needed.

**Enhancement:** Together, these factors allow me to use time effectively and focus on obtaining results from data.

**NEW TOOLS I'M INTERESTED IN:**

**Analysis: PyTorch** is functionally similar to TensorFlow, automating and simplifying the process of developing deep neural networks. It is intuitive and easy to use.

**Visualisation: ggplot2** is a powerful data visualisation library with both python and R implementations. It may be a useful alternative to matplotlib and seaborn.

**Utility: snakemake** is a common workflow language similar to nextflow, automating the process of running and validating long software pipelines created from many independent programs. Compared to nextflow, it is output focused, which may be useful when constructing different types of software pipelines.

## MONASH DATA FLUENCY

✉ tyrone.chen@monash.edu        iD 0000-0002-9207-0385

🐦 tyrone__chen        tyrone__chen

 tyronechen        in tyrone-chen