



# Alignment and splice junction identification

## RNA-Seq Hands-on Practical

Never Stand Still

Zhiliang Chen  
Systems Biology Initiative  
University of New South Wales

# Introduction

- The goal of this hands-on session is to perform some basic tasks in the downstream analysis of RNA-seq data. We will start from RNA-seq data aligned to the zebrafish genome using Tophat.
  - Prepare the environment
  - Tophat alignment
  - Viewing the alignment using IGV

# Preparing the alignment

- Sample
  - Zebra fish
  - 2 conditions : 2cells and 6h
- Type of sequencing
  - 76bp pair-end
  - Illumina
- File name and formats
  - Fastq file
  - 2cells 1.fastq and 2cells 2.fastq
  - 6h 1.fastq and 6h 2.fastq

# Alignment

- Reference genome
  - Danio\_rerio.Zv9.66
  - Index has already been created
- Alignment program and parameters
  - The 2cells data is pre-aligned
  - Alignment needs to be done for 6h dataset

# Alignment Visualisation in IGV

- The Integrative Genomics Viewer (IGV) is able to provide a visualisation of read alignments given a reference sequence and a BAM file.
  - Visualise the alignment
  - Look at the splice junctions
- Interpreting the alignment in IGV
  - <http://www.broadinstitute.org/igv/AlignmentData>