limma, *edgeR* and *Glimma* ENABLE DIFFERENTIAL ANALYSES OF GENOMIC DATA

Essential Open Source Software for Science
Kick-off Meeting

25th Feb, 2020

Chan
Zuckerberg
Initiative

Walter+
Eliza Hall
Institute of Medical Research
DISCOVERIES
FOR HUMANITY

OUR PROJECTS

- Address the most fundamental genomic problem:

  Detecting differences in the abundance of genes or other genomic features between experimental conditions, cell types or disease states

	Wild-type		**Mutant**	
	Mouse1	**Mouse2**	**Mouse1**	**Mouse2**
Gene 1	45	60	30	39
Gene 2	0	4	3	7
Gene 3	1010	800	3099	3450
…	…	…	…	…
Total	**20m**	**16m**	**15m**	**17m**

- We are Bioconductor R packages
  - *limma* – a linear model approach for both continuous and count data
  - *edgeR* – a count-based method using NB generalized linear model
  - *Glimma* – creates interactive plots for exploration of results
- The usage can be extended to various fields

GOALS AND EXPECTATIONS FOR THE COMING YEAR

- Improve ease of use and inter-operability
  - Revise packages to accept Bioconductor's *SummarizedExperiment* containers
  - Reduce computation speed (program optimization)
- New data challenges
  - Deal with zero inflation in datasets arising from new technologies
  - Handle large-scale datasets
- Documentation, training and outreach
  - Update user's guides
  - Write teaching material describing proper use of design and contrast matrices
  - Continue devoting time and effort to the Bioconductor user community

ABOUT THE KICK-OFF MEETING

- What we hope to learn
  - Ideas to improve usability, documentation and implementation of our software
  - Insight into the directionality of established, world-class software

- Our team and what we have to share
  - Mathematical/statistical skills and strategies
  - Intensive experience in the medical research field
  - Efficient programming in R

Gordon Smyth

Charity Law

Yunshun Chen

Goknur Giner

FIAT LUX
Walter+Eliza Hall
Institute of Medical Research