



# Canadian Bioinformatics Workshops

[www.bioinformatics.ca](http://www.bioinformatics.ca)

This page is available in the following languages:

Afrikaans Български Català Dansk Deutsch Ελληνικά English English (CA) English (GB) English (US) Esperanto Castellano Castellano (AR) Español (CL) Castellano (CO) Español (Ecuador) Castellano (MX) Castellano (PE) Euskara Suomeksi français français (CA) Galego ລາວ hrvatski Magyar Italiano 日本語 한국어 Macedonian Melayu Nederlands Norsk Sesotho sa Leboa polski Português română slovenščina srpski (latinica) Sotho svenska 中文 華語 (台灣) IsiZulu



## Attribution-Share Alike 2.5 Canada

### You are free:



- to Share — to copy, distribute and transmit the work
- to Remix — to adapt the work



### Under the following conditions:



- Attribution.** You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).
- Share Alike.** If you alter, transform, or build upon this work, you may distribute the resulting work only under the same or similar licence to this one.

- For any reuse or distribution, you must make clear to others the licence terms of this work.
- Any of the above conditions can be waived if you get permission from the copyright holder.
- The author's moral rights are retained in this licence.

[Disclaimer](#)

Your fair dealing and other rights are in no way affected by the above.

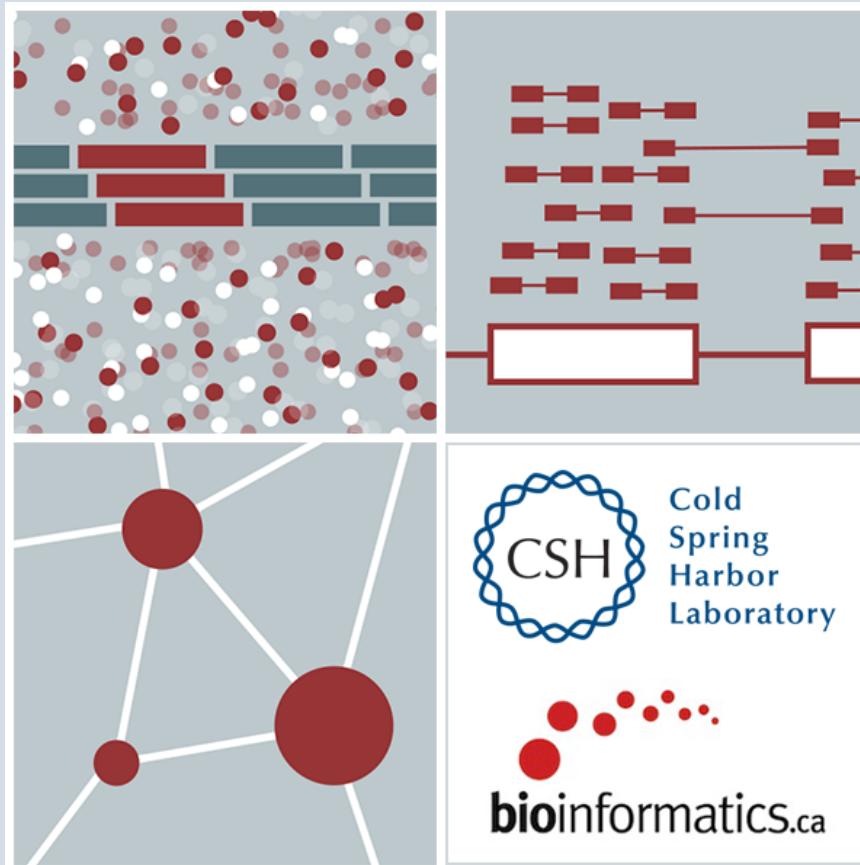
This is a human-readable summary of the Legal Code (the full licence) available in the following languages:

[English](#) [French](#)

# Informatics for RNA-seq Analysis

Malachi Griffith and Obi Griffith

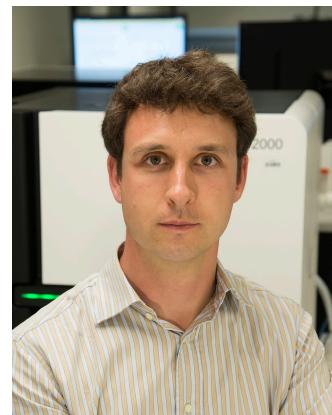
May 28-30, 2018



# Introductions to MGI/WUSTL instructors



**Malachi Griffith**  
Assistant Professor of Medicine  
Assistant Professor of Genetics  
Assistant Director, MGI



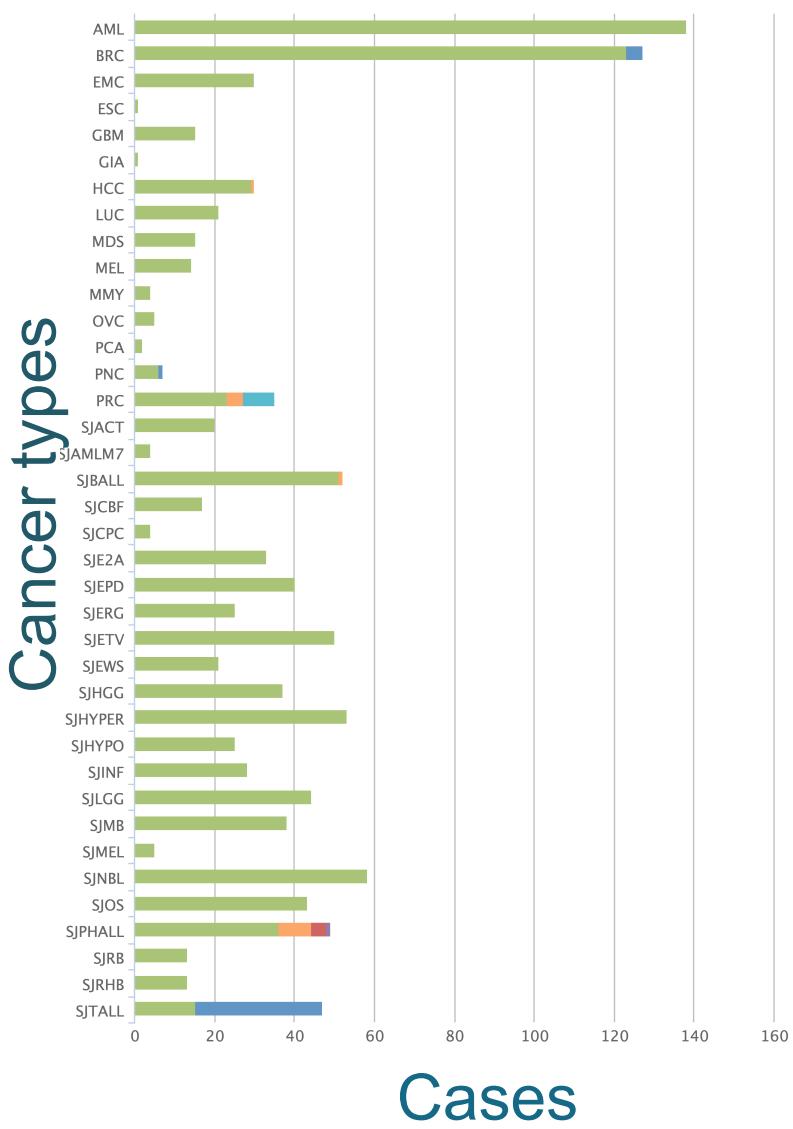
**Obi Griffith, PhD**  
Assistant Professor of Medicine  
Assistant Professor of Genetics  
Assistant Director, MGI

**McDonnell Genome Institute, Washington University School of Medicine**

RNA sequencing and analysis

**bioinformatics.ca**

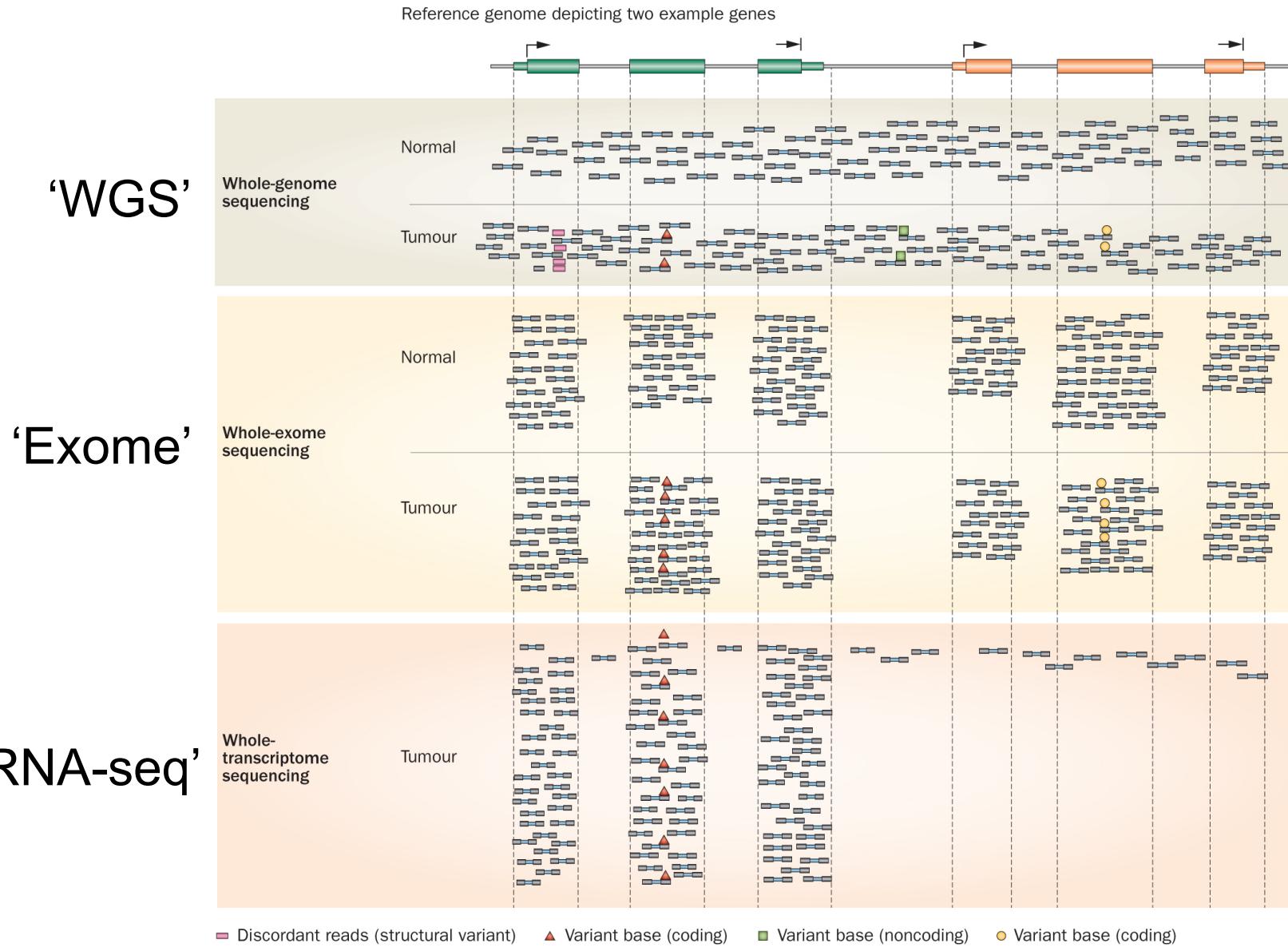
The McDonnell Genome Institute has pursued the field of genomics since inception: >1000 whole genomes, >5000 exomes, >1000 transcriptomes for dozens of tumor types



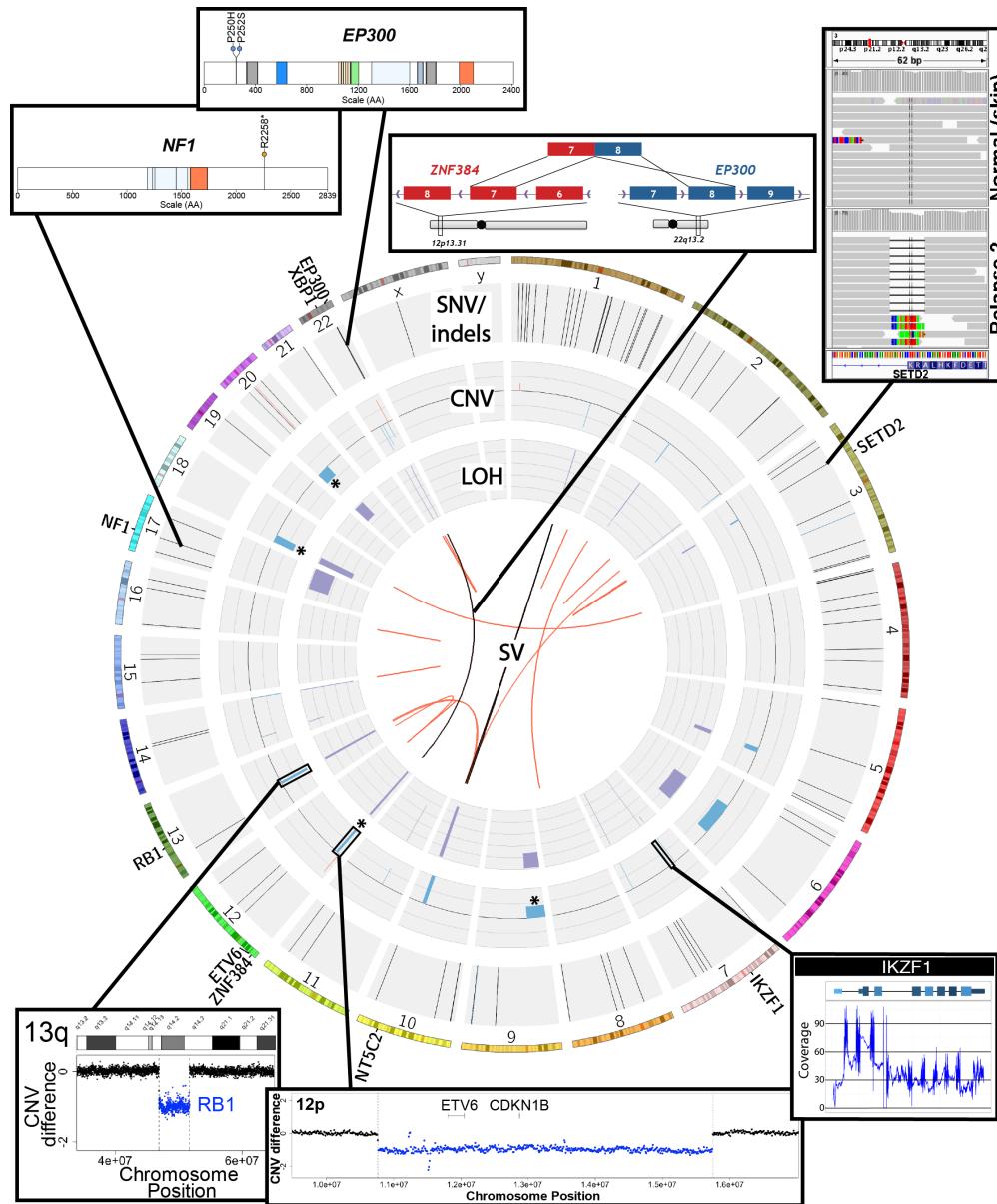
- Human Genome Project
- First to sequence and analyze a tumor whole genome sequence (Ley et al, 2008)
- Major contributor to TCGA, PCGP, etc.
- Most comprehensively sequenced single patient tumor ever published (Griffith et al, 2015)
- Early proof-of-principle for cancer precision medicine (Griffith et al, 2016)
- Analysis and tools for first personalized cancer vaccine design in humans (Carreno et al, 2015)
- Many other widely used tools

MG - 1000+ tumor whole genomes, many more exomes, X10 & NovaSeq will accelerate this!

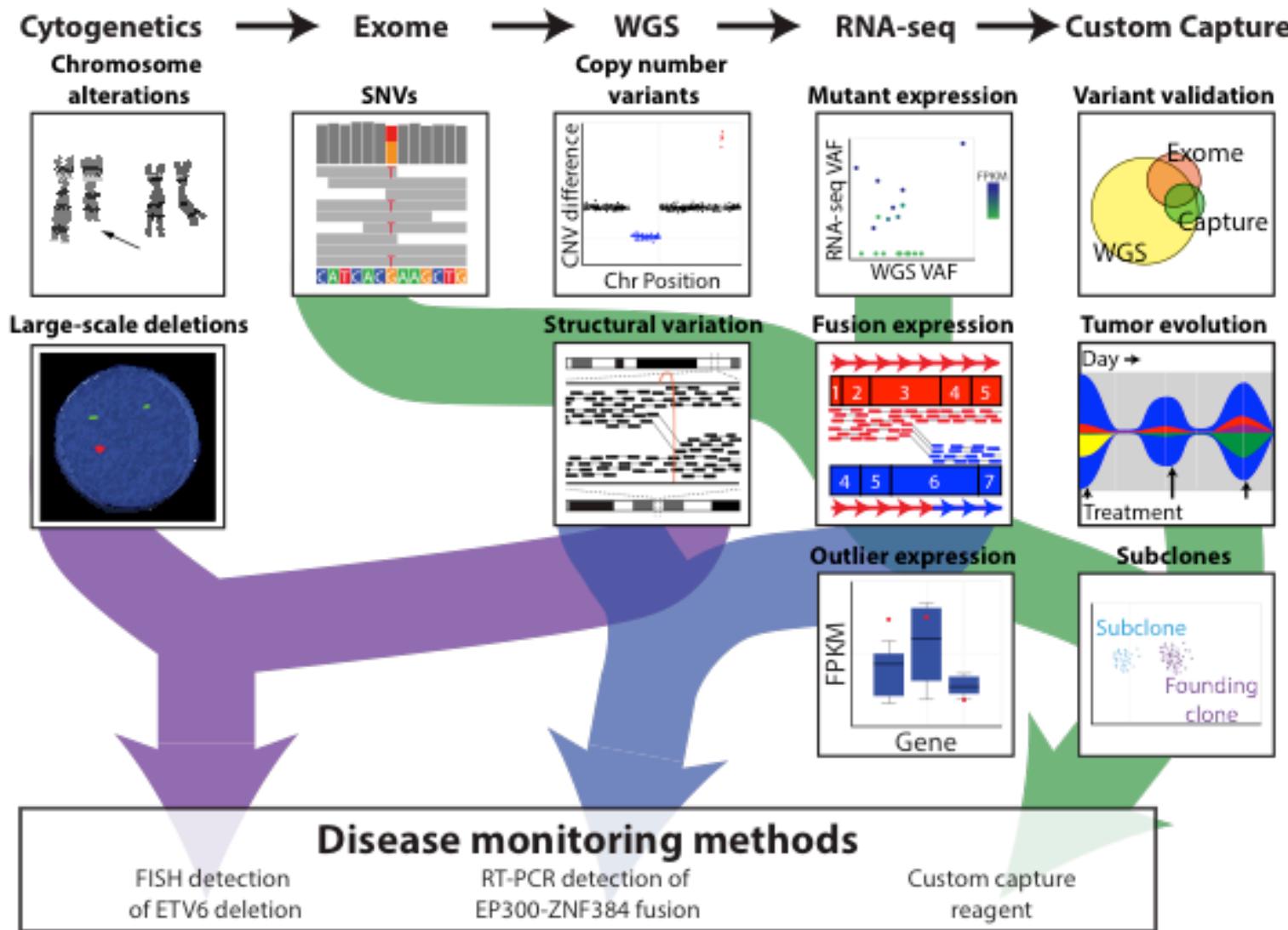
# Whole genome, exome, transcriptome and other ‘omic’ sequencing allows us to detect and confirm many different variant types



# SNVs, Indels, CNVs, SVs, fusions, LOH, expression changes, methylation changes, and more

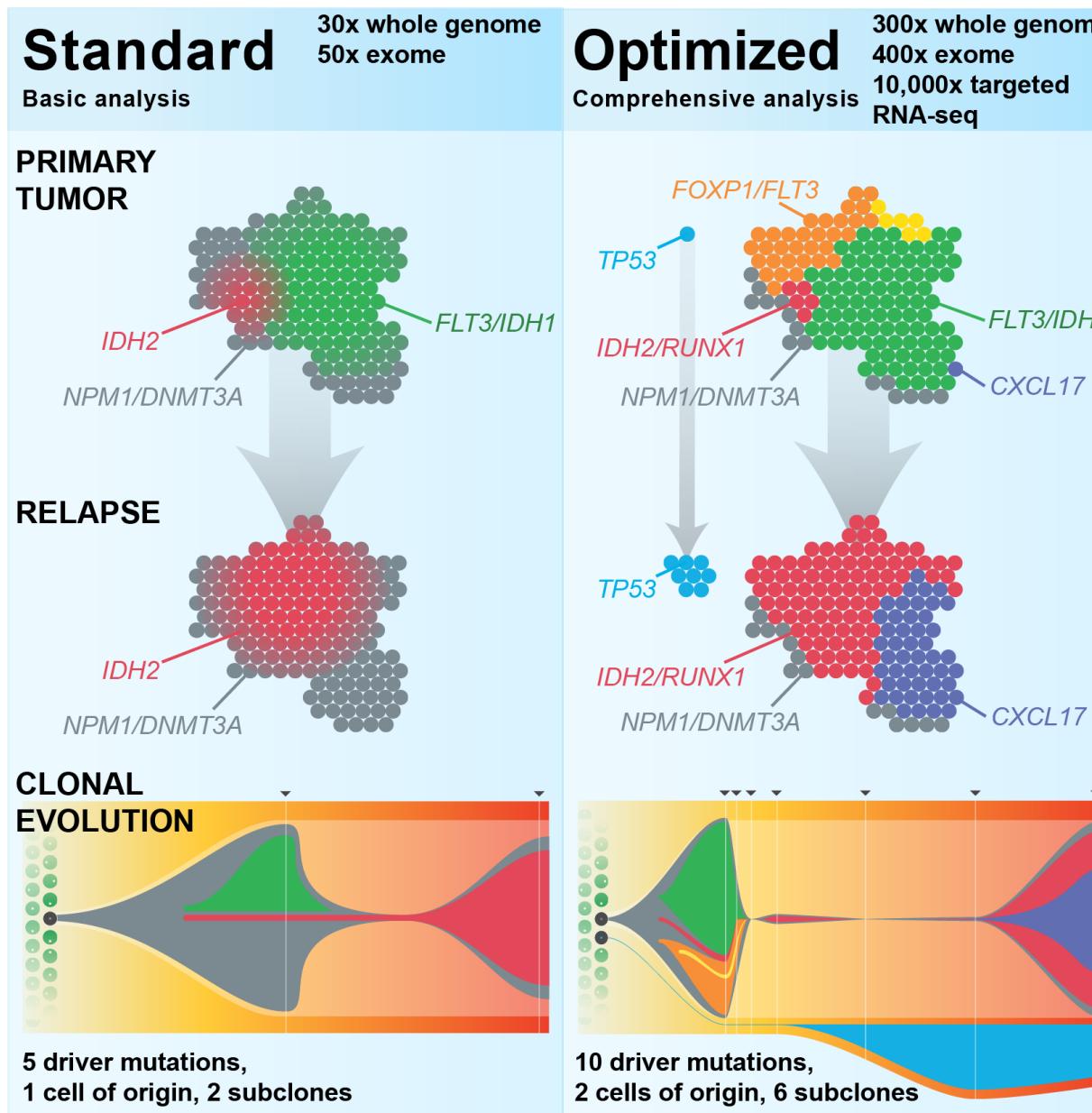


# Comprehensive and integrative analysis methods are needed



Kilannin Krysiak

# Each study often requires considerable customization



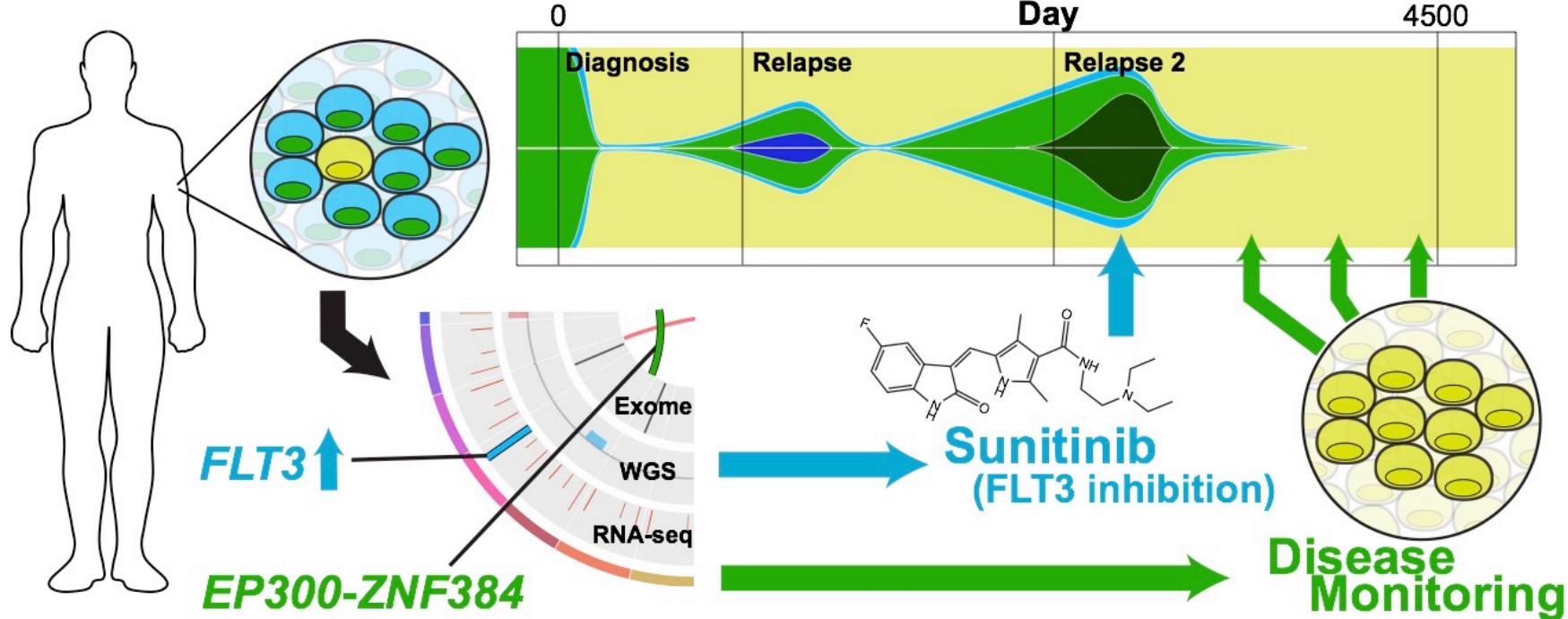
Kilannin Krysiak

RNA sequencing and analysis

Griffith et al. 2015

bioinformatics.ca

# Personalized medicine requires personalized strategies



ELSEVIER



CrossMark

Experimental Hematology 2016;44:603–613

**Experimental  
Hematology**

Comprehensive genomic analysis reveals *FLT3* activation and a therapeutic strategy for a patient with relapsed adult B-lymphoblastic leukemia

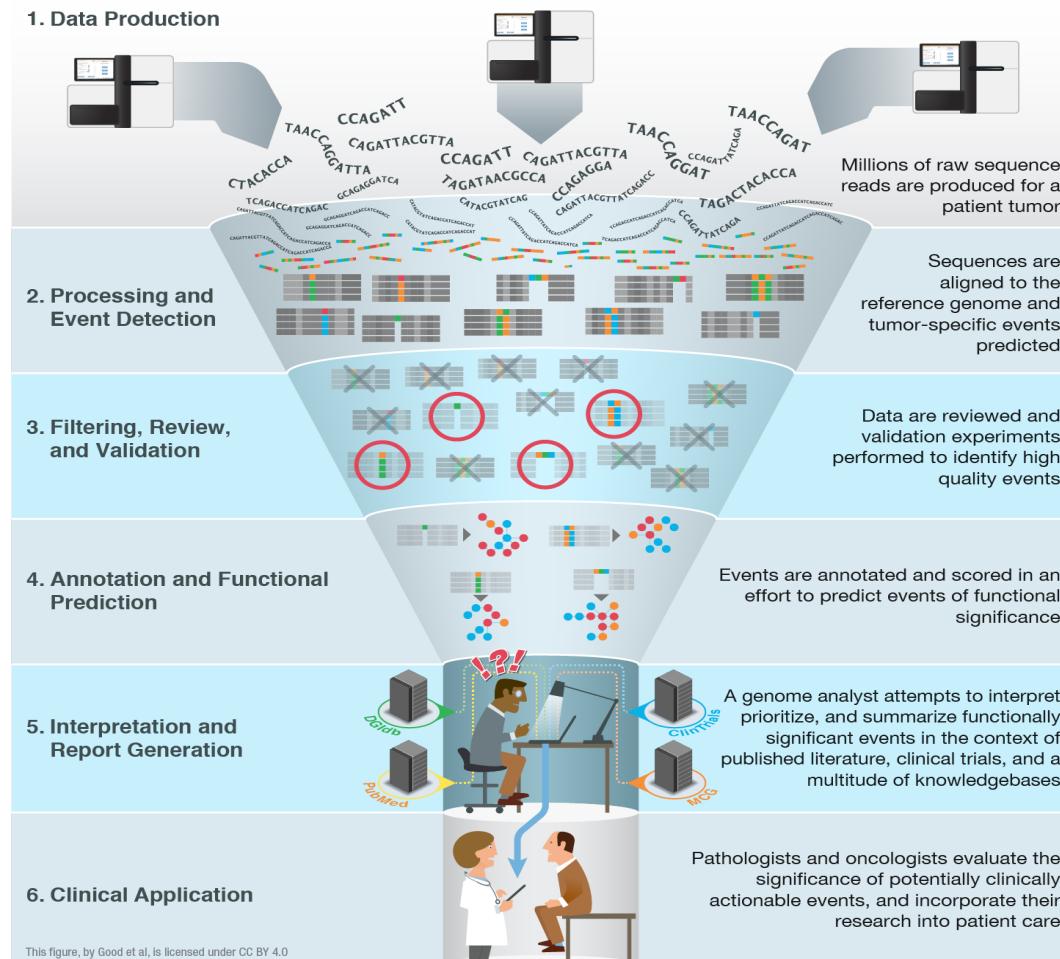
Kilannin Krysiak

[Griffith et al. 2016](#)

RNA sequencing and analysis

**bioinformatics.ca**

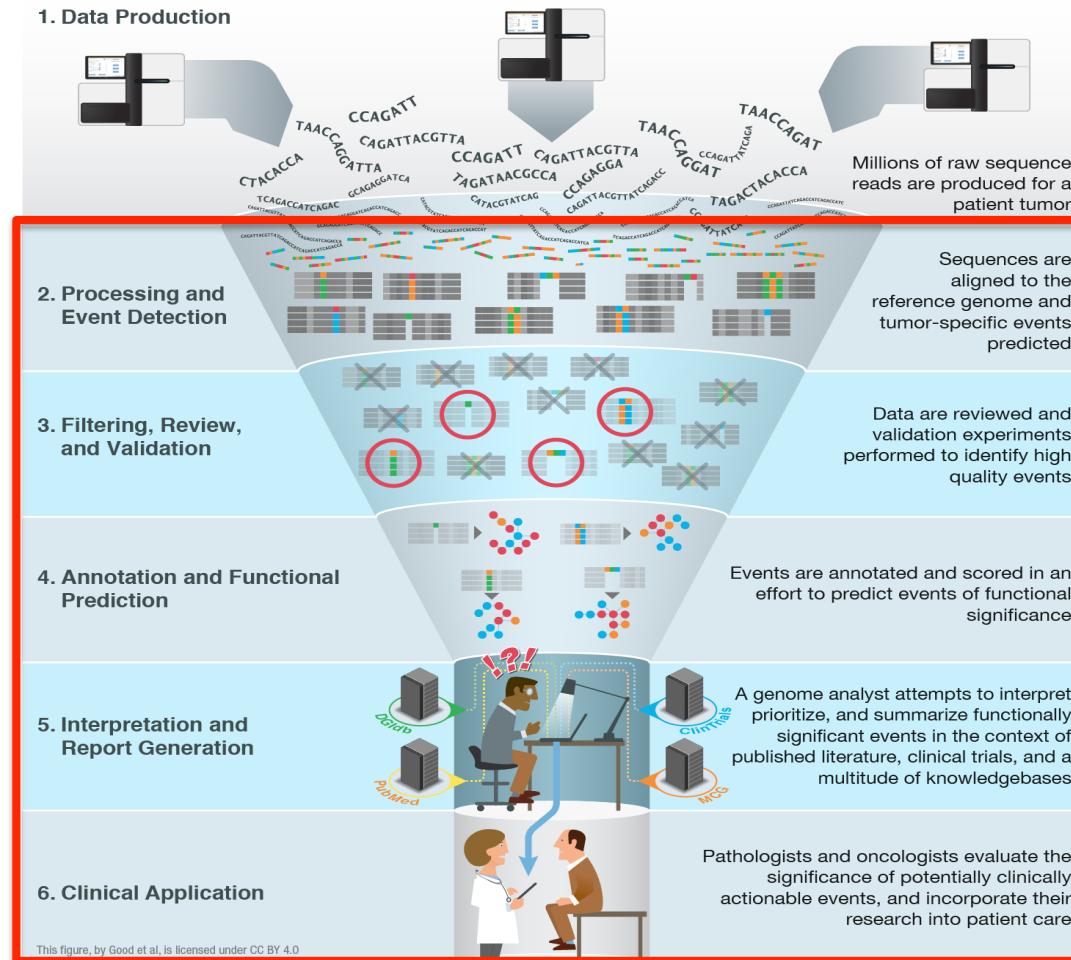
# High-throughput sequencing has been largely automated



Good BM, Ainscough BJ, McMichael JF, Su AI, Griffith OL. 2014.  
Genome Biology. 15(8):438.

Joshua McMichael

The rest of this workshop will focus on the methods and tools needed to take raw sequence data to interpretation and application – still the bottleneck!



Good BM, Ainscough BJ, McMichael JF, Su AI, Griffith OL. 2014.  
Genome Biology. 15(8):438.

Joshua McMichael

# The Griffith lab is focused on developing methods to address this bottleneck for precision medicine in cancer

- **Cancer genome analysis**
  - AML
    - [Optimizing cancer genome analysis](#)
  - Breast cancer
    - [Immunotherapy](#) (cancer vaccines)
    - STAT1-/ mouse model
    - Targeted sequencing of 625 ER+ with long-term follow-up
  - Liver cancer
  - Small cell lung cancer
  - OSCC...
- **Precision Medicine for Cancer**
  - [Genomics Tumor Board](#)
    - [Case Reports](#) and [Clinical Trials](#)
- **Education projects**
  - [RNA-seq analysis and cloud computing](#)
  - [CSHL](#) and [CBW](#)
  - [Genomic Data Visualization/Interpretation](#)
- **Tool development**

# Where tools/resources do not exist we build them



[www.dgidb.org](http://www.dgidb.org)

Search genes for known and potentially druggable interactions



<https://github.com/griffithlab/pVAC-Seq>

Personalize vaccine design



[www.docm.info](http://www.docm.info)

Filter against highly curated set of mutations known to cause cancer



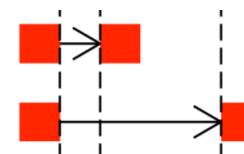
<https://github.com/griffithlab/GenVisR>

Create genomic visualizations



[www.civicdb.org](http://www.civicdb.org)

Identify highly curated summaries of clinical interpretations for variants in cancer



**regtools**

<https://github.com/griffithlab/regtools>

Identify regulatory variants

# Encourage best practices for software development

The screenshot shows the GitHub organization page for 'The Griffith Lab'. At the top, there's a header with a GitHub icon, 'This organization', 'Search', 'Pull requests', 'Issues', 'Gist', and a notification bell. Below the header is a profile picture of two men and the text 'The Griffith Lab' and 'Academic Lab of Obi and Malachi Griffith'. A link to their website is provided. The main navigation bar includes 'Repositories', 'People 25', 'Teams 21', 'Projects 0', and 'Settings'. Below the navigation is a search bar, filters for 'Type: All' and 'Language: All', and a 'Customize pinned repositories' section with a 'New' button. The main content area lists three repositories: 'dgi-db' (Rails frontend to The Genome Institute's drug gene interaction database, Ruby, 23 stars, 14 forks, updated 2 days ago), 'pVAC-Seq' (A cancer immunotherapy pipeline, Python, 37 stars, 33 forks, updated 2 days ago), and 'GenVisR' (Genome data visualizations, R, 95 stars, 37 forks, updated 2 days ago). To the right, there's a 'Top languages' section showing R, Python, Ruby, Perl, and HTML, and a 'People' section showing 25 members with their profile pictures.

- All code deposited in Github
  - 35 repos for different software/projects
  - Virtually all public and open-source
  - Test-driven, code review, pull requests, etc

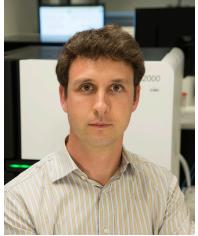
<https://github.com/griffithlab>

<https://github.com/genome>

# WUSTL – MGI: Group members



**Malachi  
Griffith**



**Obi  
Griffith**



**Benjamin  
Ainscough**



**Erica  
Barnell**



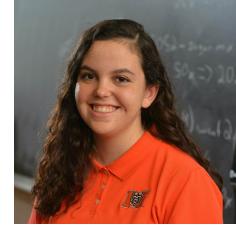
**Katie  
Campbell**



**Kaitlin  
Clark**



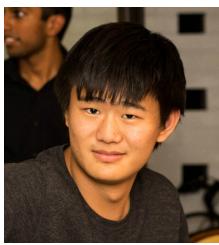
**Adam  
Coffman**



**Kelsy  
Cotto**



**Arpad  
Danos**



**Yang-Yang  
Feng**



**Felicia  
Gomez**



**Jasreet  
Hundal**



**Susann  
a Kiwala**



**Kilanin  
Krysiak**



**Lynzey  
Kujan**



**Jason  
Kunisaki**



**Josh  
McMichael**



**Cody  
Ramirez**



**Zachary  
Skidmore**



**Nick  
Spies**



**Lee  
Trani**



**Alex  
Wagner**



**Jason  
Walker**

# Informatics background poll

- Student poll
  - Are you doing genomics research?
  - Do you consider yourself a bioinformatician?
  - What kind of NGS data are you working with?
    - WGS?
    - Exome?
    - RNAseq?
    - Epigenome?
    - Single cell?
    - Other?
  - What organism do you work with?
    - Does it have a reference genome?
  - Did you bring data?

# Student poll continued

Not counting the pre-requisites and materials for this course:

- Are you familiar with linux/command line?
  - Intermediate?
  - Expert?
- Do you sometimes write code?
  - What language?
- Are you familiar with R?
  - Intermediate?
  - Expert?
- Are you familiar with ggplot?
  - Intermediate?
  - Expert?
- Do you use git/github?