

The Elizabeth H.  
and James S. McDonnell III

**McDONNELL  
GENOME INSTITUTE**  
at Washington University



**Washington**  
University in St. Louis  

---

SCHOOL OF MEDICINE

# PMBIO Module 0

## Introductions

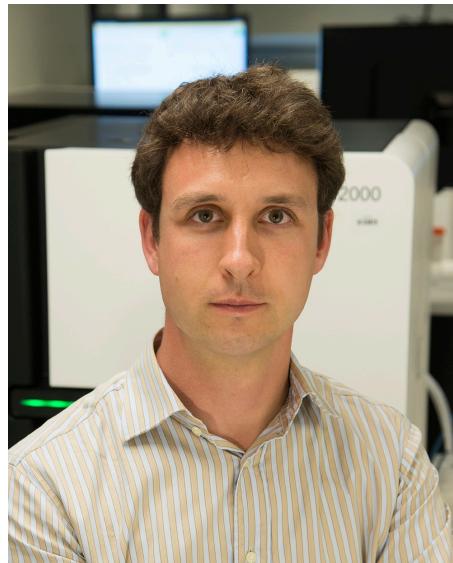
Malachi Griffith, Obi Griffith, Zachary Skidmore  
Introduction to bioinformatics for DNA and RNA sequence  
analysis (IBDR01)

29 October - 2 November, 2018  
Glasgow

# Introductions to course instructors



**Malachi Griffith, PhD**  
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



**Zachary Skidmore, MSc**  
Staff Scientist, MGI

McDonnell Genome Institute, Washington University School of Medicine

# Other major contributors

Erica Barnell



Sharon Freshour



Jeff Szymanski



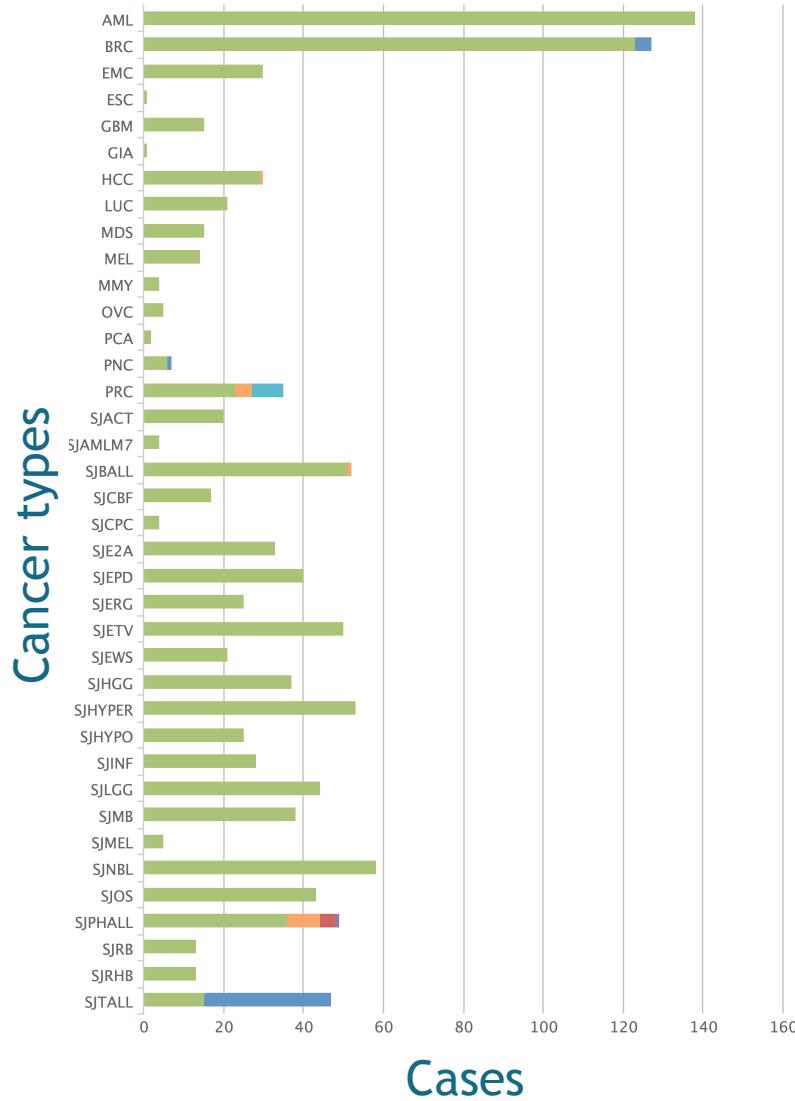
Huiming Xia



Additional guidance and expertise from members of the Griffith Lab, the McDonnell Genome Institute, and elsewhere:

Chris Miller, Jason Walker, Alex Wagner, Thomas Mooney, Susanna Kiwala, Jasreet Hundal, Yang-Yang Feng, Adam Coffman, ...

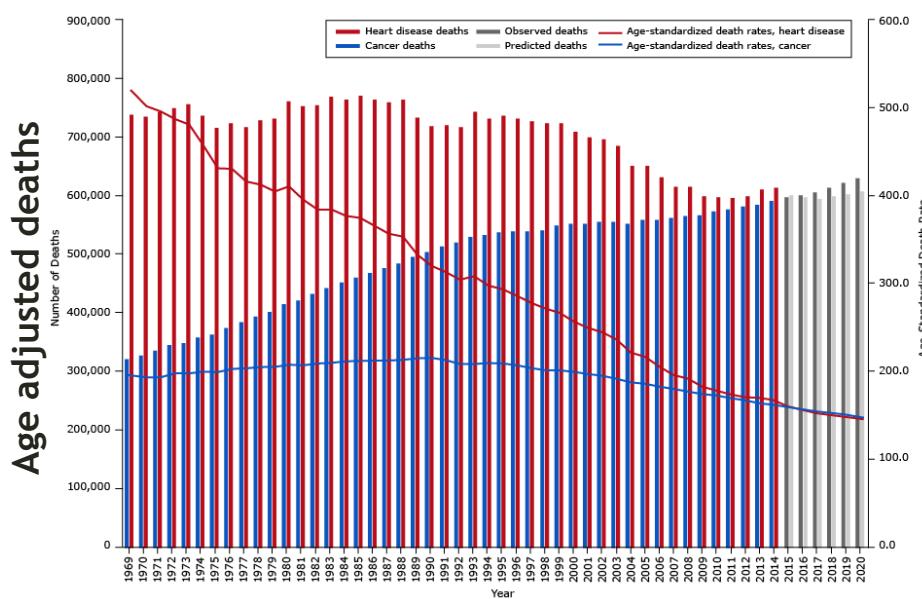
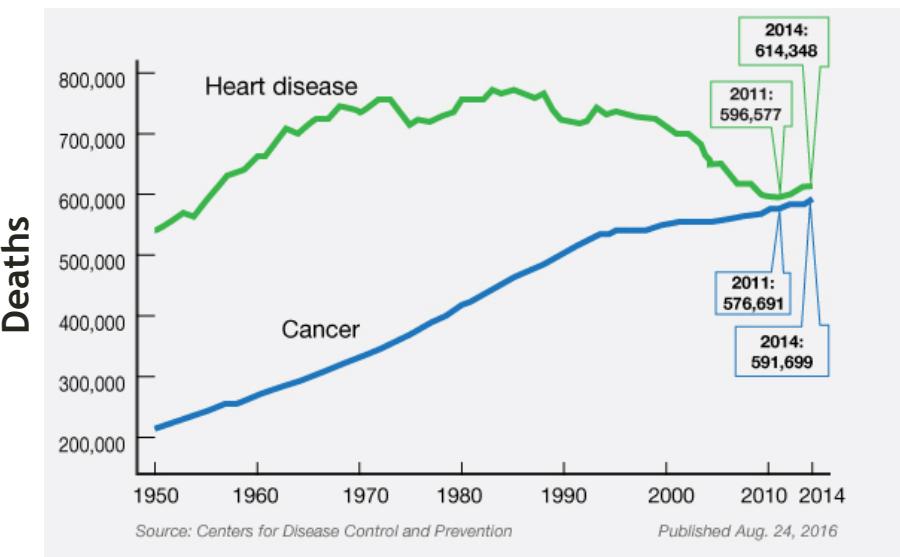
The McDonnell Genome Institute has pursued the field of genomics since inception:  
>>1000 whole genomes, >5000 exomes, >1000 transcriptomes for dozens of tumor types



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

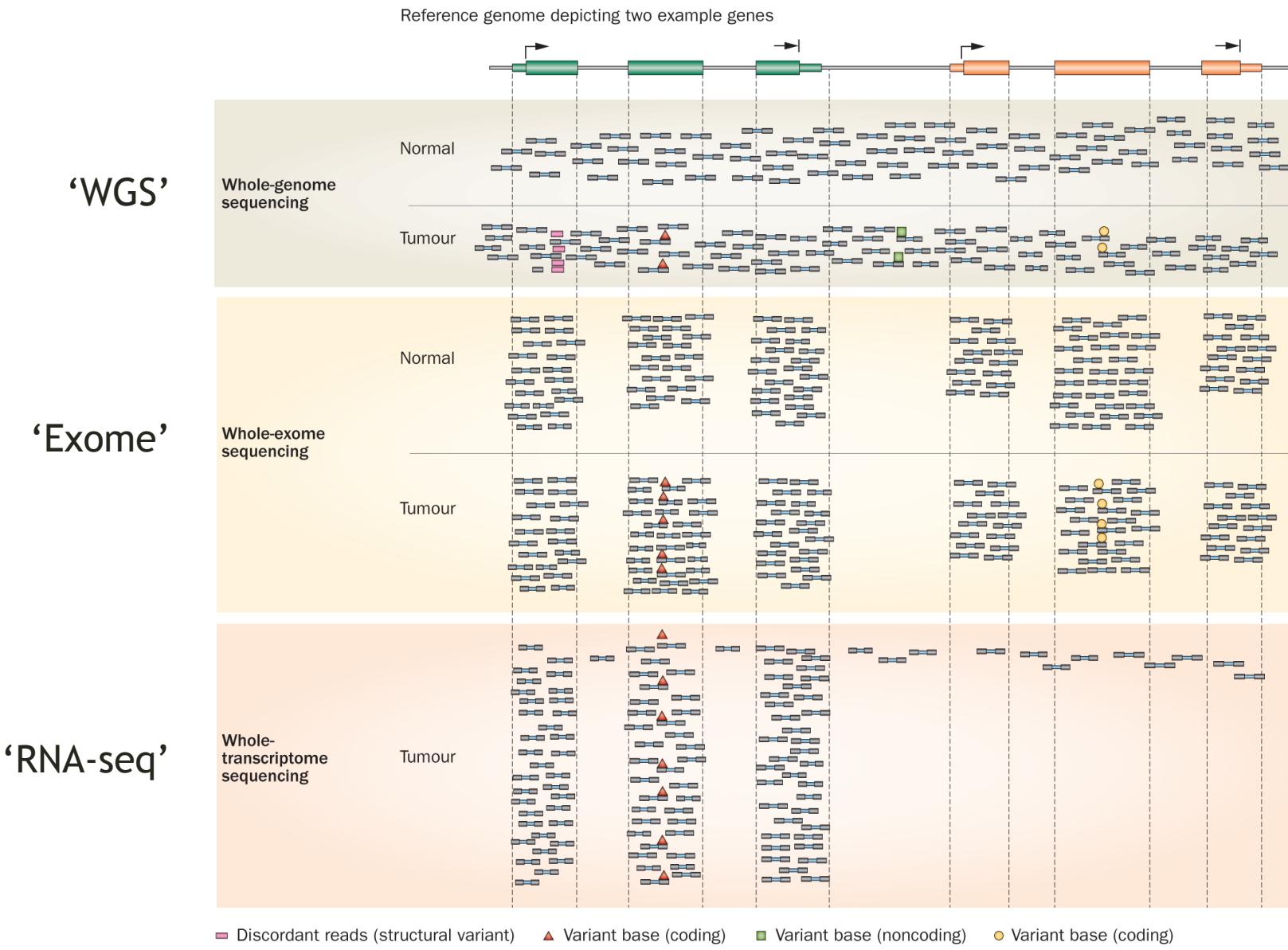
- Human Genome Project
- First to sequence and analyze a tumor whole genome sequence (Ley et al, 2008)
- Major contributor to TCGA, PCGP, etc.
- Most comprehensively sequenced single patient tumor ever published (Griffith et al, 2015)
- Early proof-of-principle for cancer precision medicine (Griffith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

# Cancer is an exemplar disease for precision/personalized medicine

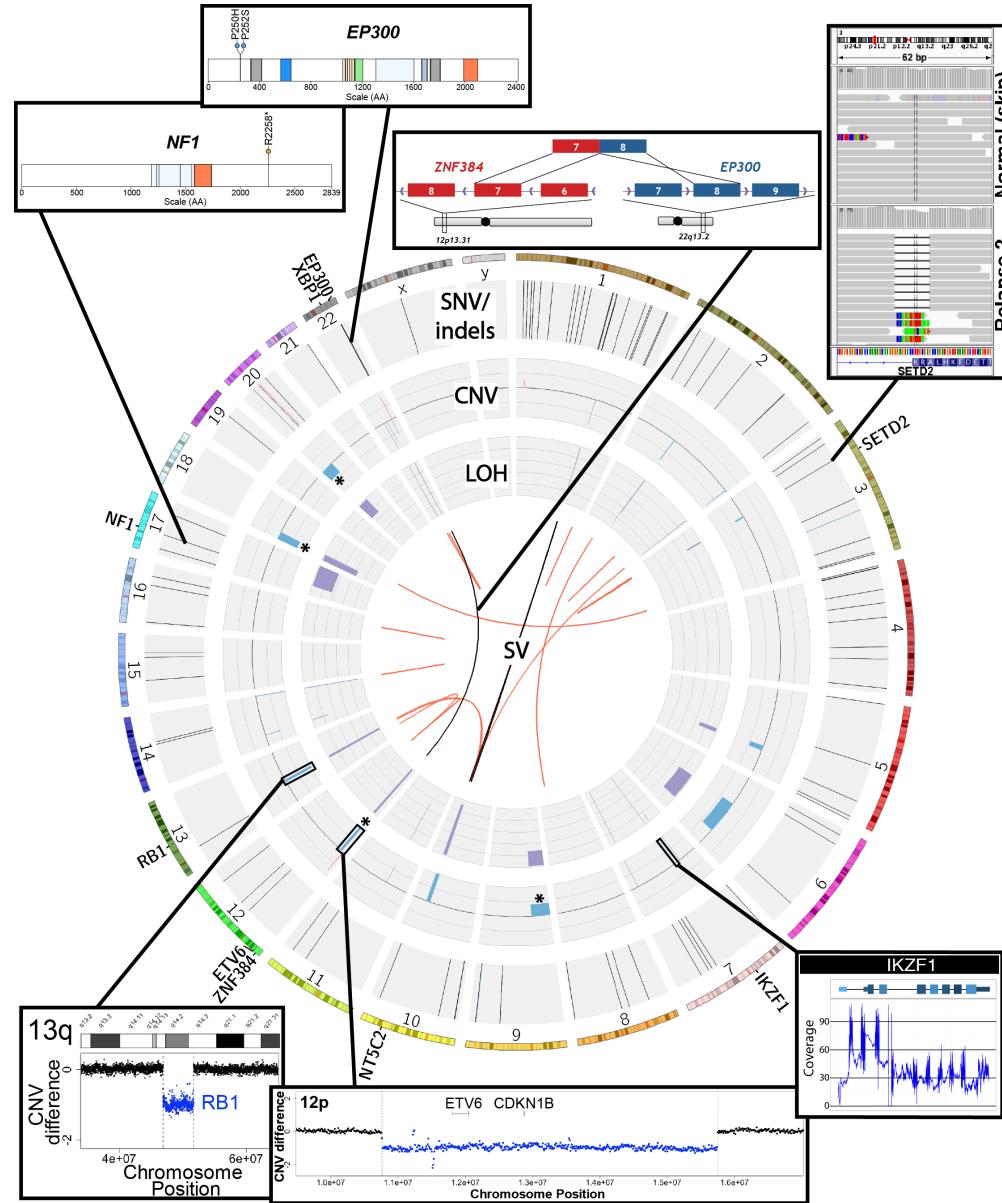


- High prevalence
- Treatment options are already incredibly complex
- A disease of genome\* aberrations
  - Both inherited and acquired
- In some cancer types, existing therapies are effective but “brute force” with harsh side effects
  - A few cancer types that started at ~100% mortality now have >90% cure rate
  - In some cancer types, existing therapies remain largely ineffective

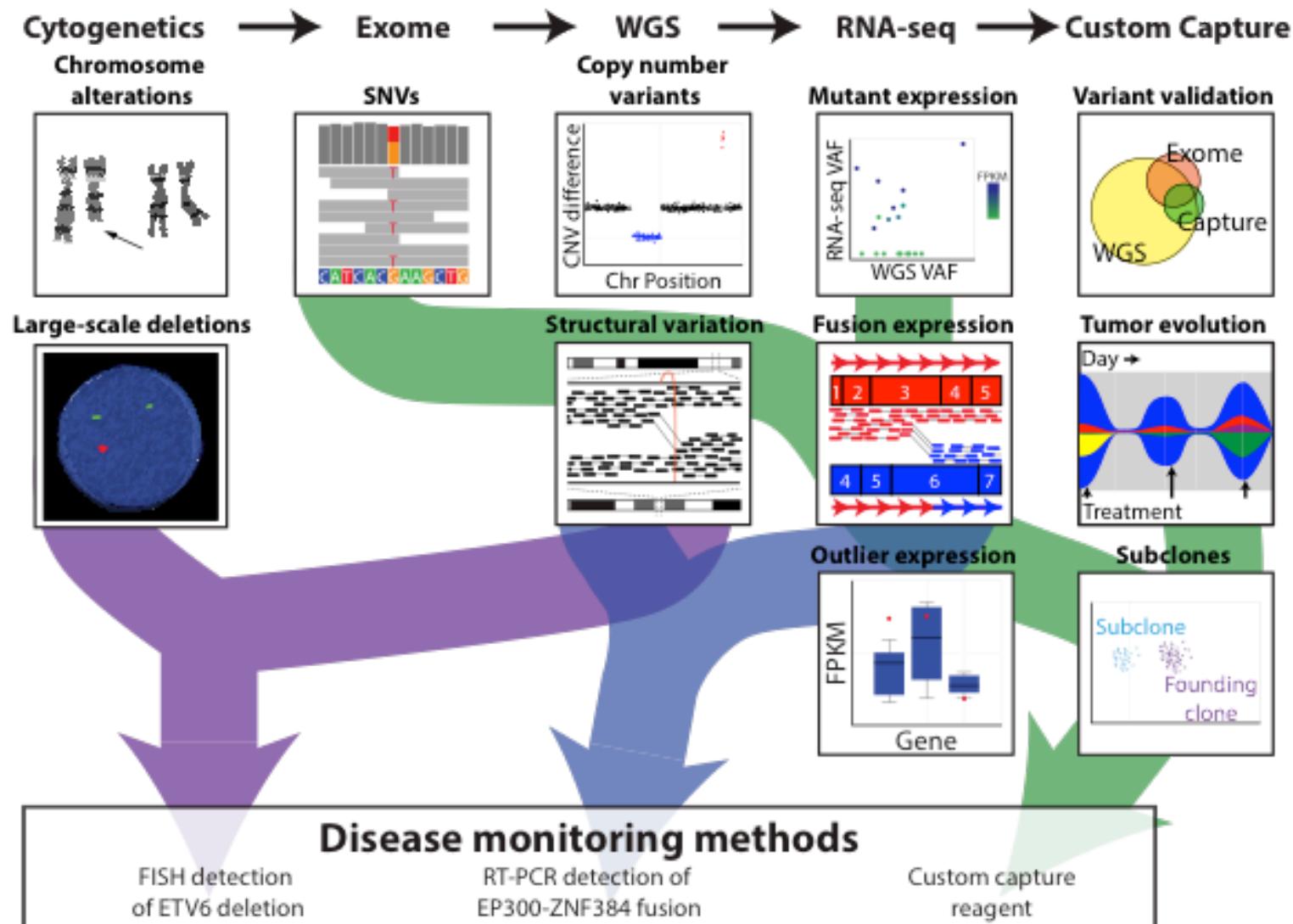
# Whole genome, exome, transcriptome and other ‘omic’ sequencing allows us to detect and confirm many different variant types



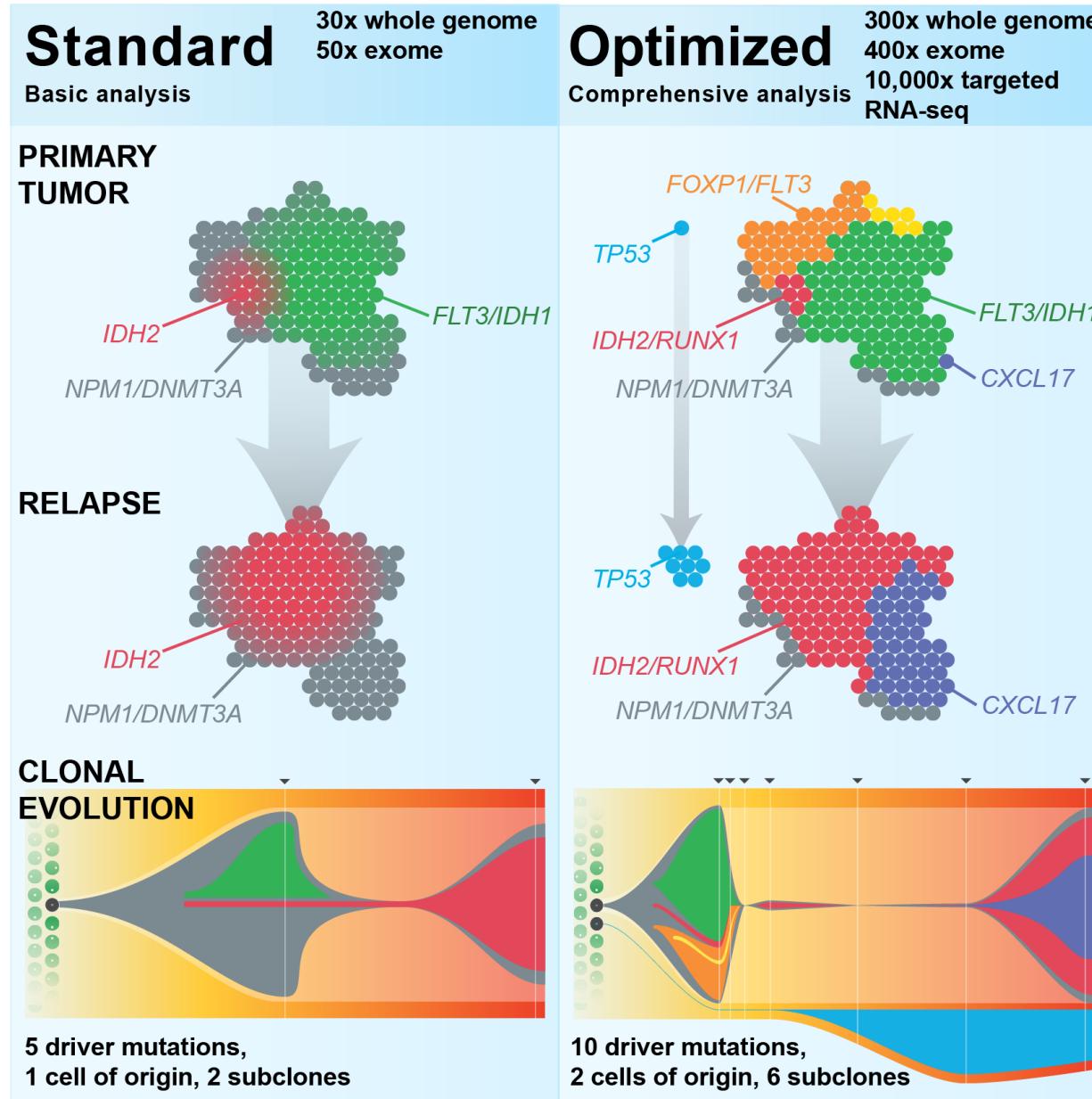
# SNVs, Indels, CNVs, SVs, fusions, LOH, expression changes, methylation changes, and more



# Comprehensive and integrative analysis methods are needed



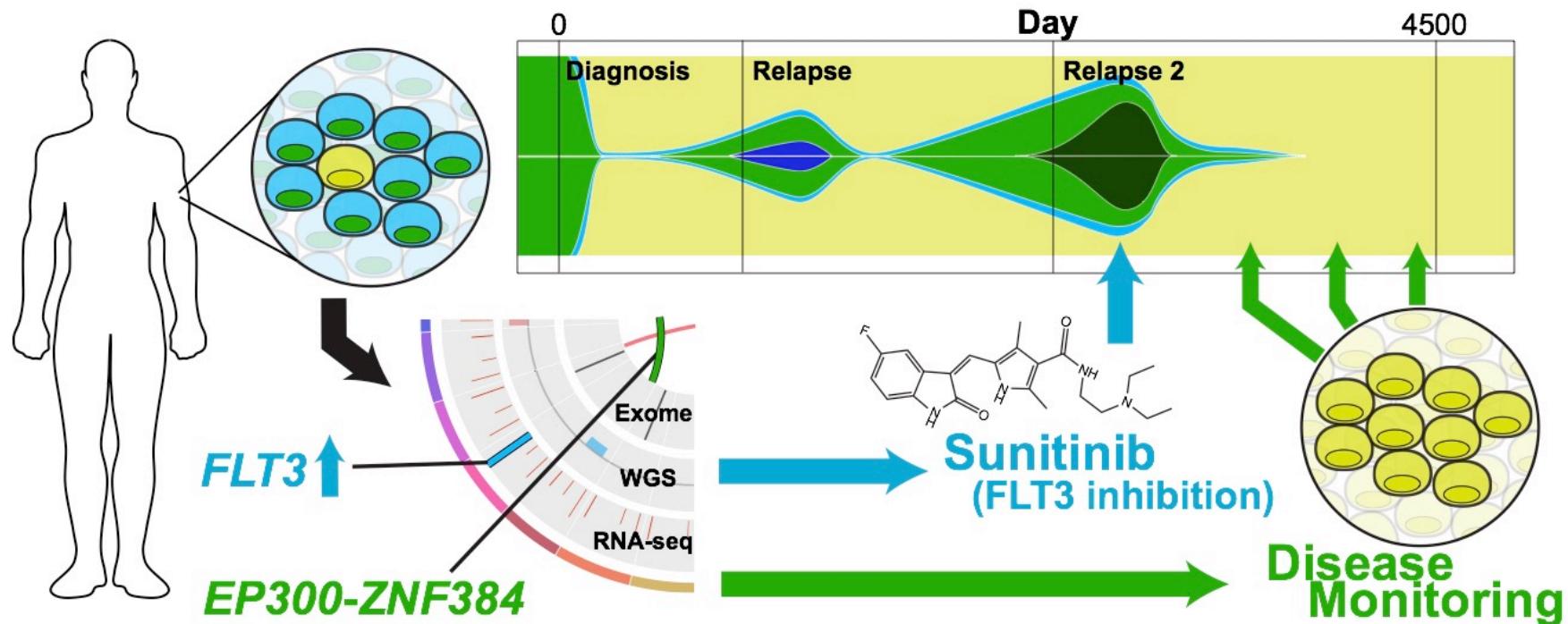
# Each study often requires considerable customization



Griffith et al. 2015

Kilannin Krysiak

# Personalized medicine requires personalized strategies



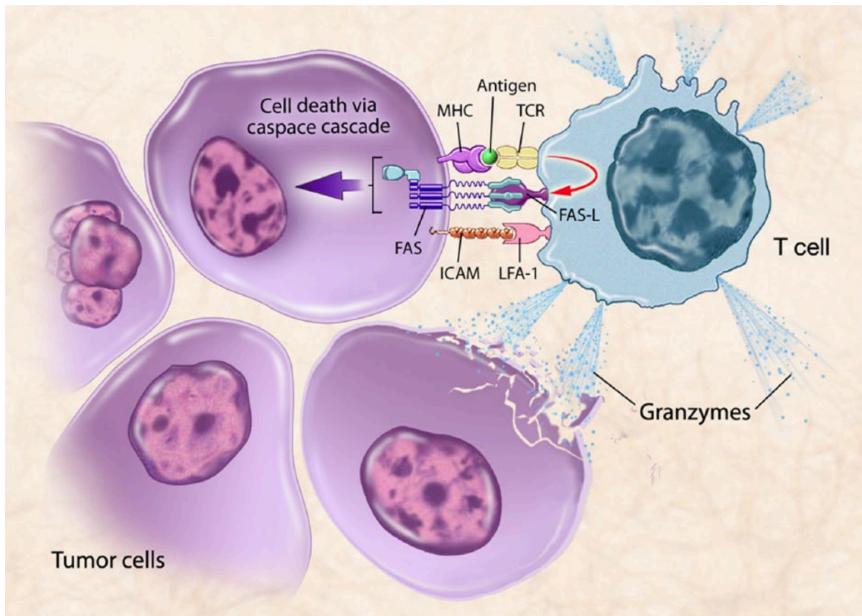
Experimental Hematology 2016;44:603–613

**Experimental  
Hematology**

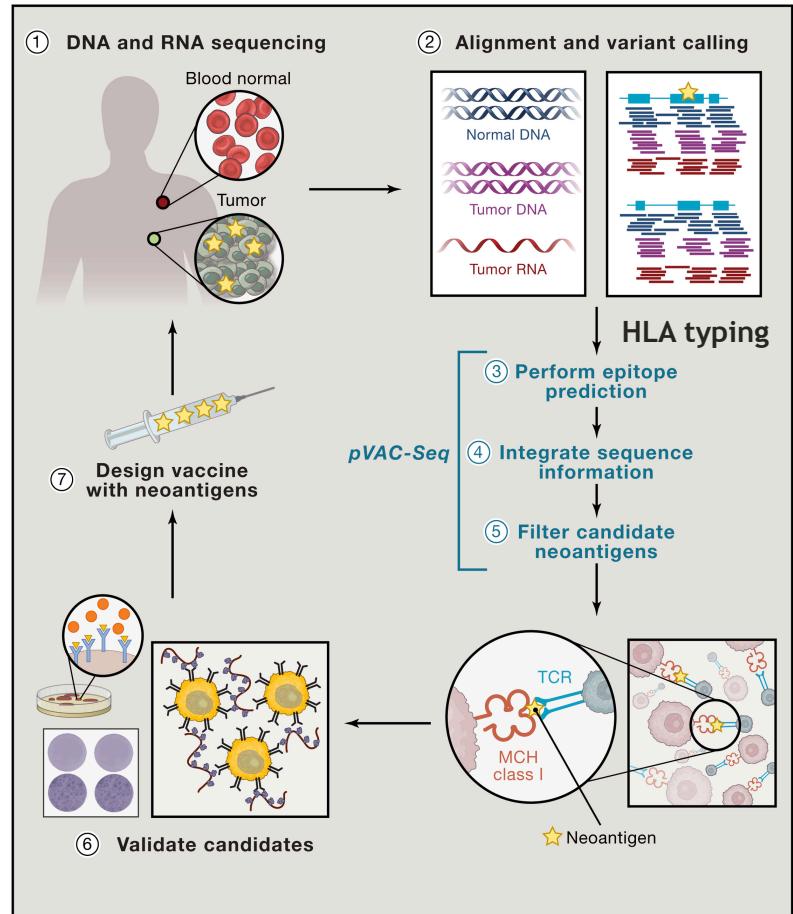
Comprehensive genomic analysis reveals *FLT3* activation and a therapeutic strategy for a patient with relapsed adult B-lymphoblastic leukemia

[Griffith et al. 2016](#)

# Immune therapy opens exciting new avenues for personalized therapy



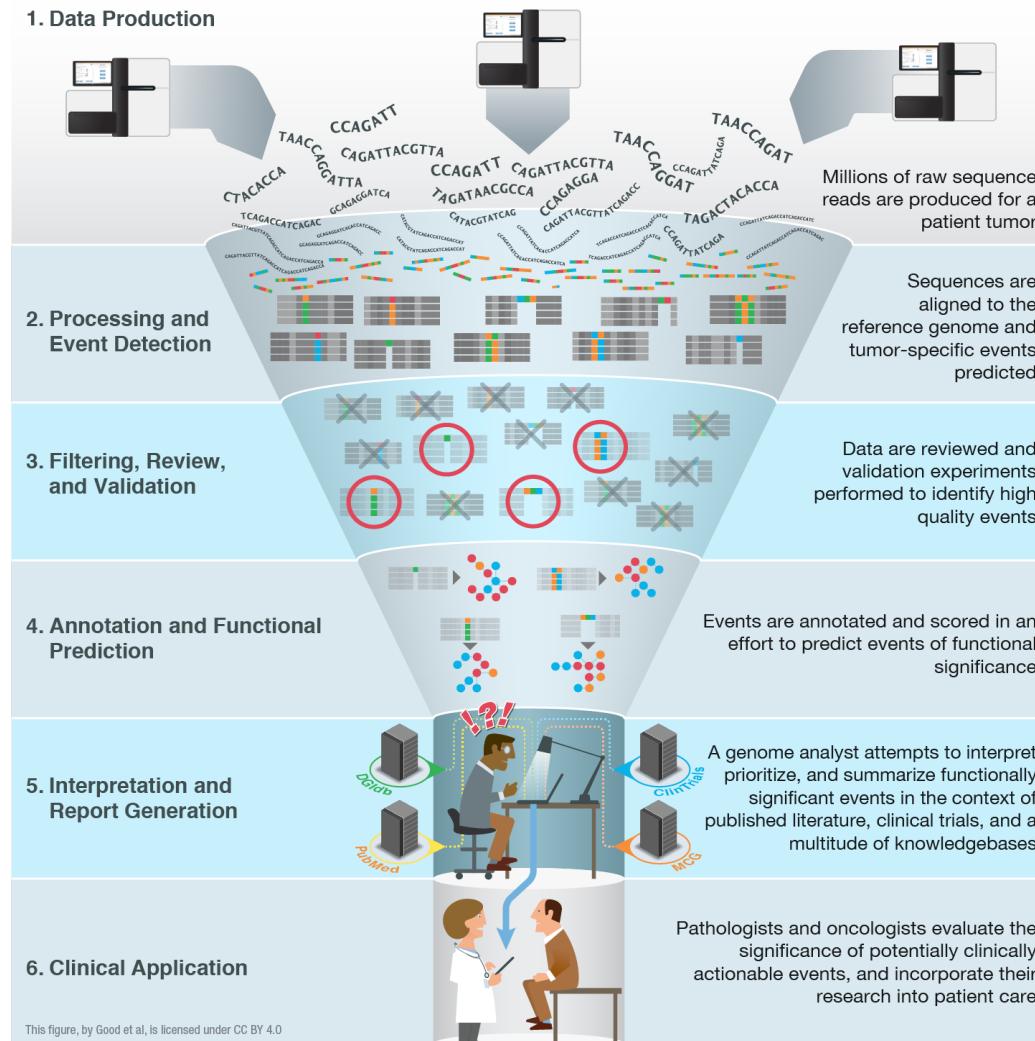
[Singh and Gulley, 2014](#)



Jasreet Hundal & Katie Campbell

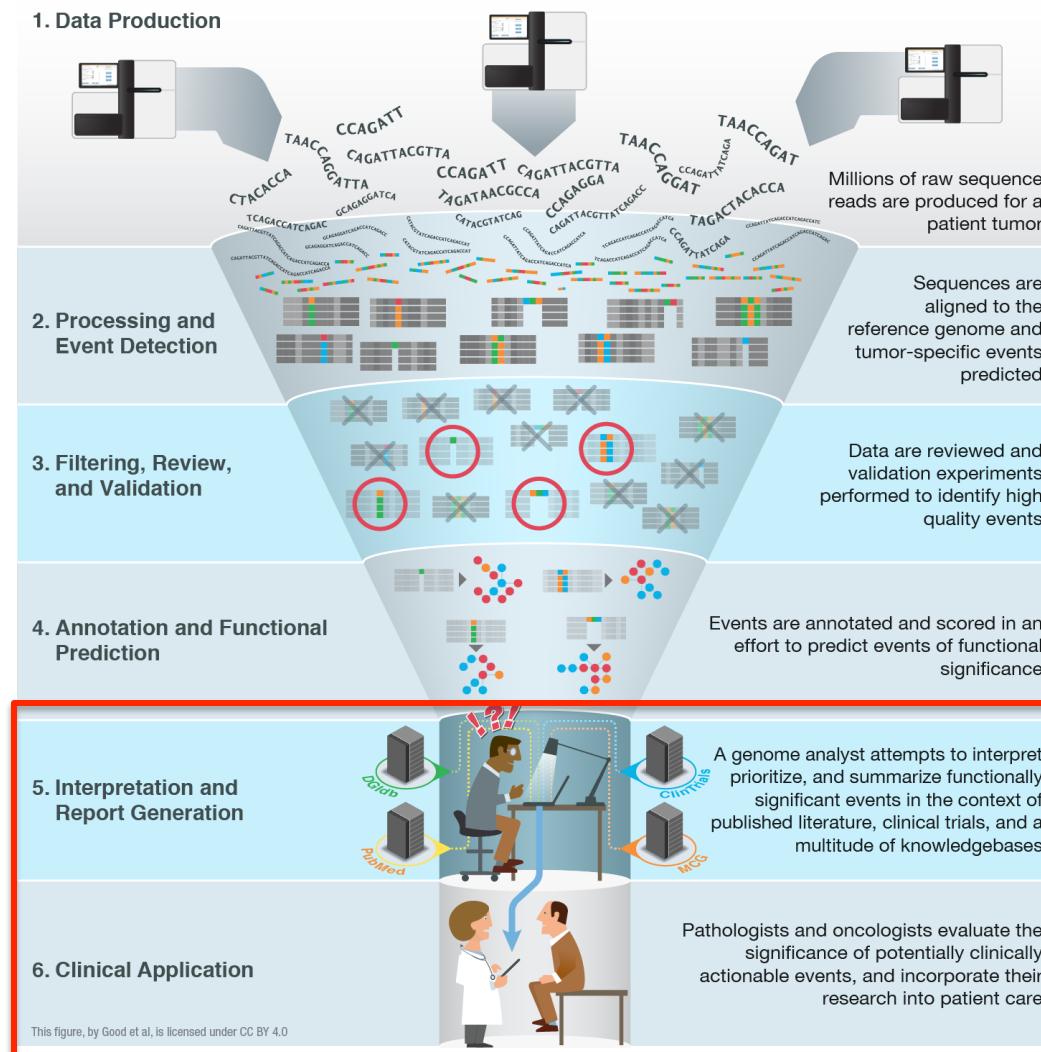
Shirley X. Liu, Elaine R. Mardis.  
Applications of Immunogenomics  
to Cancer. Cell Press. 2017.

# High-throughput sequencing has been largely automated allowing rapid identification of somatic and germline variants



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

# Interpretation and visualization of genomic alterations remains the bottleneck for realizing precision medicine



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

# 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
  - Etc...
- 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



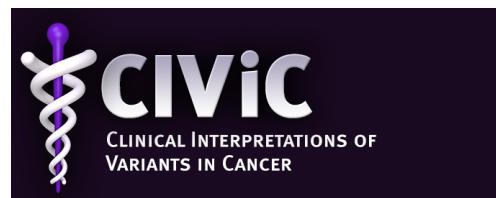
[www.docm.info](http://www.docm.info)

Filter against highly  
curated set of mutations  
known to cause cancer



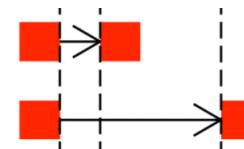
<https://github.com/griffithlab/GenVisR>

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 logo, a search bar, and navigation links for 'Pull requests', 'Issues', and 'Gist'. Below the header, there's a profile picture of two men and the text 'The Griffith Lab' followed by 'Academic Lab of Obi and Malachi Griffith' and a link to 'http://www.griffithlab.org'. A navigation bar below the profile includes 'Repositories' (which is selected), 'People 25', 'Teams 21', 'Projects 0', and 'Settings'. There are also filters for 'Search repositories...', 'Type: All', '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 team 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

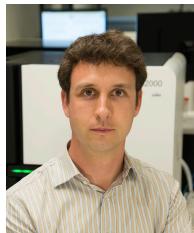
<https://github.com/griffithlab>

<https://github.com/genome>

# Acknowledgements: Griffith Lab group members



Malachi  
Griffith



Obi  
Griffith



Benjamin  
Ainscough



Erica  
Barnell



Katie  
Campbell



Kaitlin  
Clark



Adam  
Coffman



Kelsy  
Cotto

CIViC/  
Regulome

Regtools/  
pVAC-seq

Lymphoma

pVAC-seq

pVAC-  
seq/CIViC

CIViC/  
Lymphoma

CIViC

HCC



Arpad  
Danos



Yang-Yang  
Feng



Felicia  
Gomez



Jasreet  
Hundal



Susanna  
Kiwala



Kilanin  
Krysiak



Lynzey  
Kujan



Jason  
Kunisaki

CIViC

CIViC  
Lymphoma

genVisR  
HCC/SCLC

CIViC/Komen

HCC/SCLC

DGIdb  
SCLC

Director  
Informatics



Josh  
McMichael



Cody  
Ramirez



Zachary  
Skidmore



Nick  
Spies



Lee  
Trani



Alex  
Wagner



Jason  
Walker

# Acknowledgements: McDonnell Genome Institute

## McDonnell Genome Institute @ Washington University School of Medicine

We are The Genome Institute. Scott Abbott, Travis E. Abbott, Derek C. Albracht, Mark A. Ames, Lucinda L. Antonacci-Fulton, Elizabeth L. Appelbaum, Kelly R. Becker, Michael C. Becker, Edward A. Belter Jr., Paul M. Bender, Angela K. Bentlage, Veena Bhonagiri, Tamberlyn Bieri, Thomas D. Biewald, Andrew P. Bohannon, Kirk A. Brege, Rose L. Brockhouse, Anthony M. Brummett, Mark M. Burnett, Christopher R. Cabanski, Theresa A. Caesar, Matthew B. Callaway, Francine R. Camacho, Edgar A. Campbell, Darren Casimere, Wan-Ching Chan, Lei Chen, Mari Jo Clark, Adam C. Coffman, Lisa L. Cook, Matthew Cordes, Laura P. Courtney, Amy K. D'Albora, John S. Dalbora, Indraniel Das, Sara E. Decarlo, Nathan D. Dees, Kimberley D. Delehaunty, Brandon J. Delgado, Anthony M. DeLuca, Tracie L DeLuca, Ryan T. Demeter, Ricardo A. DeMoya, Brian R. Derickson, Li Ding, Kara R. Dix, Gregory P. Dooling, Feiyu Du, James M. Eldred, Efrem Exum, Jason M. Falcone, Candace N. Farmer, Ian T. Ferguson, Grace M. Freeman, Catrina C. Fronick, William L. Fronick, Bradley T. Fulton, Robert S. Fulton, Kristi N. Futhey, Xin Gao, Barbara Gillam, Chakravarthy Girda, Jennifer Godfrey, Jenna Christine Goeckner, Laura Jane Gottschalk, Susan D. Grasso, Malachi Griffith, Obi Griffith, Jeanne M. Grigsby, David S. Gudermann, Priscilla Hale, Terri M. Hall, Kevin V. Haub, Amy E. Hawkins, Todd G. Hepler, Brandi N. Herter, Jennifer S. Hodges, Kimberly Hughes, Jasreet Hundal, Stephanie M. Jackson, Rodney L. Jones, Wendall B. Jones, Krishna L. Kanchi, Cyriac Kandoth, Kimberly Diane Keen, Kyung H. Kim, Michael J. Kiwala, Daniel Koboldt, Sara Kohlberg, Colin L. Kremitzki, Milinn R. Kremitzki, Sheila R. Lakanen, David E. Larson, Sai Harn Lek, Shawn M. Leonard, Shin F. Leong, Andrew D. Levy, Shannon M. Lewis, Timothy J. Ley, Ling Lin, Tina A. Lindsay, Charles Lu, Kaelyn A. Lundry, Amy L. Ly, Sandra K. MacMillan, Vincent J. Magrini, Lenon G. Maguire, Christopher A Maher, Jennifer Maher, Rondy Janjak Malik, Elaine R. Mardis, Christopher M. Markovic, John C. Martin, Judith M. Mc Cart, Suzanne E. McGaugh, Sean D McGrath, Michael D. McLellan II, Joshua F. McMichael, Samantha N. McNulty, Brian Meininger, Karyn N. Meltz Steinberg, Kathie A. Mihindukulasuriya, Lenka Mikalova, Christopher A. Miller, Walter Miller, Patrick J. Minx, Makedonka Mitreva, Deborah S. Moeller, Michael J. Montague, Thomas B. Mooney, Andrew Z. Morrison, David L. Morton, Jennifer B. Murphy, Pamela A. Nangle, William E. Nash, Maze Bi Ndukum Ndonwi, Joanne O. Nelson, Nham Nhan, Beifang Niu, Nathaniel G. Nutter, Benjamin J. Oberfell, David J O'Brien, Kerri R. Ochoa, Gretchen A. O'Donnell, Michelle D. O'Laughlin, Philip Jeffrey Ozersky, Stephanie J. Parish, Josh Peck, Kymberlie A. Pepin, Stephanie K. Pleasant, Craig S. Pohl, Eric Ponce, Allison Ann Regier, Amy D. Reily, Gaia A. Remerowski, Natalia Rivera, Susan M. Rock, Ronald E. Rodriguez-Santiago, Irina R. Ronko, Bruce A. Rosa, Kelsi M. Rotter, Ryan J. Rupp, Gabriel Eugene Sanderson, Amy P. Sansone, Kyriena L. Schatzkamer, Debra E. Scheer, Heather K. Schmidt, Michael T. Schmidt, Zachary L. Schools, William E. Schroeder, N'Desa T'vorie Scott, Harry E. Senaldi, Cheri-lynn R. Shadding, Nicholas T. Sheehan, Dong Shen, Susanna Siebert, Jessica M. Silva-Fisher, Julie T. Smith, Scott M. Smith, Erica J. Sodergren Weinstock, Steven S. Spargur, John G. Spieth, Gary P. Stiehr, Dennis B. Striegel, Cynthia L. Strong, Dawn K. Sutter, Kenneth B. Swanson, Yat T. Tang, Andrea R. Taylor, Thynn K. Thane, Brenda A Theising, Audra S Thomas-Monti, Brianne M. Thomeczek, Valerie L. Thompson, Chad M. Tomlinson, Lee M. Trani, Nikki L. Trapp, John L. Trevaskis III, Evana Trevaskis, Richard Tripper, Rusudan B. Turabelidze, Rahul Tyagi, Joelle M. Veizer, Tammi L. Vickery, Jason E. Waligorski, Jason R. Walker, Patricia E. Wallace, John W. Wallis, Qi Wang, Wesley C. Warren, Brian H. Watts, James V. Weible, Matthew R. Weil, George M. Weinstock, Michael C. Wendl, Richard K. Wilson, Roxanne M. Wilson, Aye M. Wollam, Dee Wu, Kristine M. Wylie, Todd N. Wylie, Kai Ye, Xu Zhang, Yanjiao Zhou

# Introduction to bioinformatics for DNA and RNA sequence analysis

- Course goals - by the end of the course we hope you will:
  - Have a solid understanding of basic computational biology techniques for interpreting data
  - Become comfortable with linux command line and other core bioinformatics skills
  - Learn the basics of cloud computing
  - Run and understand a complex genome analysis workflow (alignment, variant calling, assembly, etc., etc.)
  - Be familiar with a variety of web resources (e.g., genome browsers and portals) to aid in the interpretation of omic data
  - Be able to use R graphical programs for data visualization (e.g., ggplot2, GenVisR) and interactive visualization (e.g., shiny)
  - Be able to perform, visualize, and interpret expression, differential expression and pathway analysis
  - Be able to assess a variant's biological and clinical significance (e.g., VEP, ClinVar, and CIViC)
- This is a brand new workshop for us - feedback welcome!

# Student introductions

---

- Please introduce yourself: Name, affiliation and area of research
- 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?
    - Other?
  - What organism do you work with?
  - Did you bring data?

# Student poll continued

---

- 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?