



Genomic data visualization and interpretation

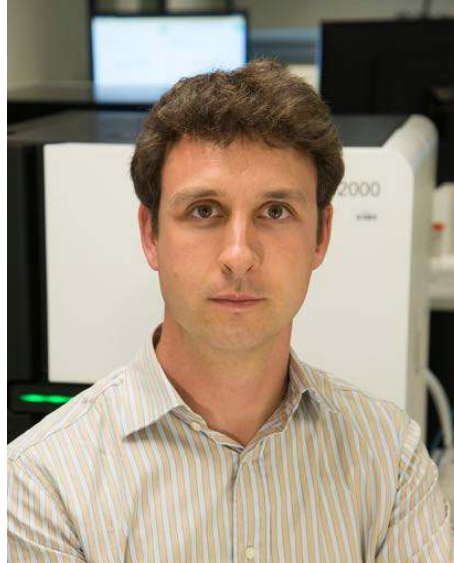
Malachi Griffith, Obi Griffith, Zachary Skidmore
Evomics, Workshop on Genomics
Jan 11, 2019
Český Krumlov, Czech Republic

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



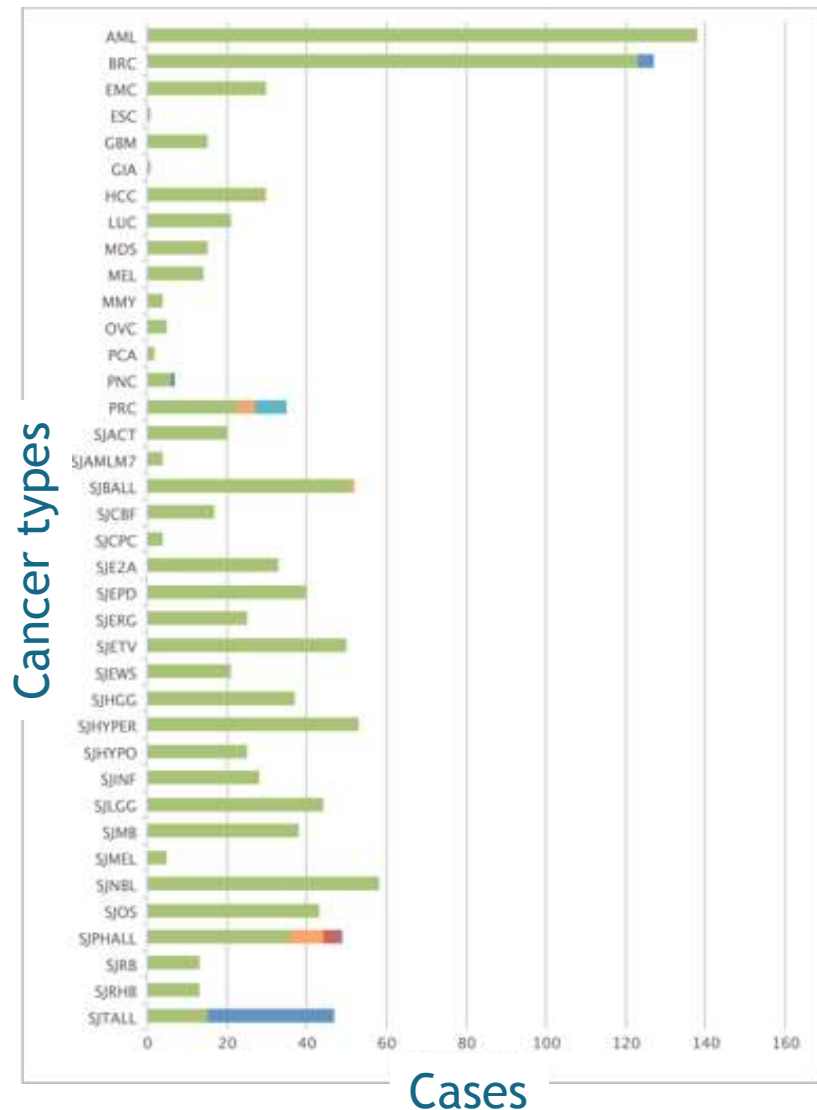
Zachary Skidmore, MSc

Staff Scientist, MGI

GenVisR creator

McDonnell Genome Institute, Washington University School of Medicine

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

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

Overview of lab's research (griffithlab.org)

- **Cancer genome analysis**
 - Breast cancer, Liver cancer, Lung cancer, Head and neck cancer, etc.
 - Variant Interpretation
 - Immunogenomics
- **Precision medicine for cancer**
 - [Genomics Tumor Board](#)
 - [Case Reports](#)
 - [Clinical Trials](#)
 - [Personalized Cancer Vaccines](#)
- **Education projects**
 - [RNA-seq analysis and cloud computing](#) ([CBW](#), Toronto)
 - [Advanced Sequencing Technologies and Applications](#) ([CSHL](#), New York)
 - [Genomic Data Visualization/Interpretation](#) ([Physalia Courses](#), Berlin)
 - [Precision Medicine Bioinformatics](#) ([PR Informatics](#), Glasgow)
 - [High-Throughput Biology: From Sequence to Networks](#) ([CSHL](#) / [CBW](#), New York)
 - [Workshop on Genomics](#) ([Evomics](#), Český Krumlov)
- **Tool development**

Where tools/resources do not exist we build them



www.dgidb.org

Search genes for known and potentially druggable interactions



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

Personalize vaccine design



www.docm.info

Filter against highly curated set of mutations known to cause cancer



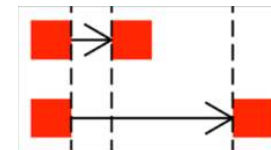
<https://github.com/griffithlab/GenVisR>

Create genomic visualizations



www.civicdb.org

Identify highly curated summaries of clinical interpretations for variants in cancer

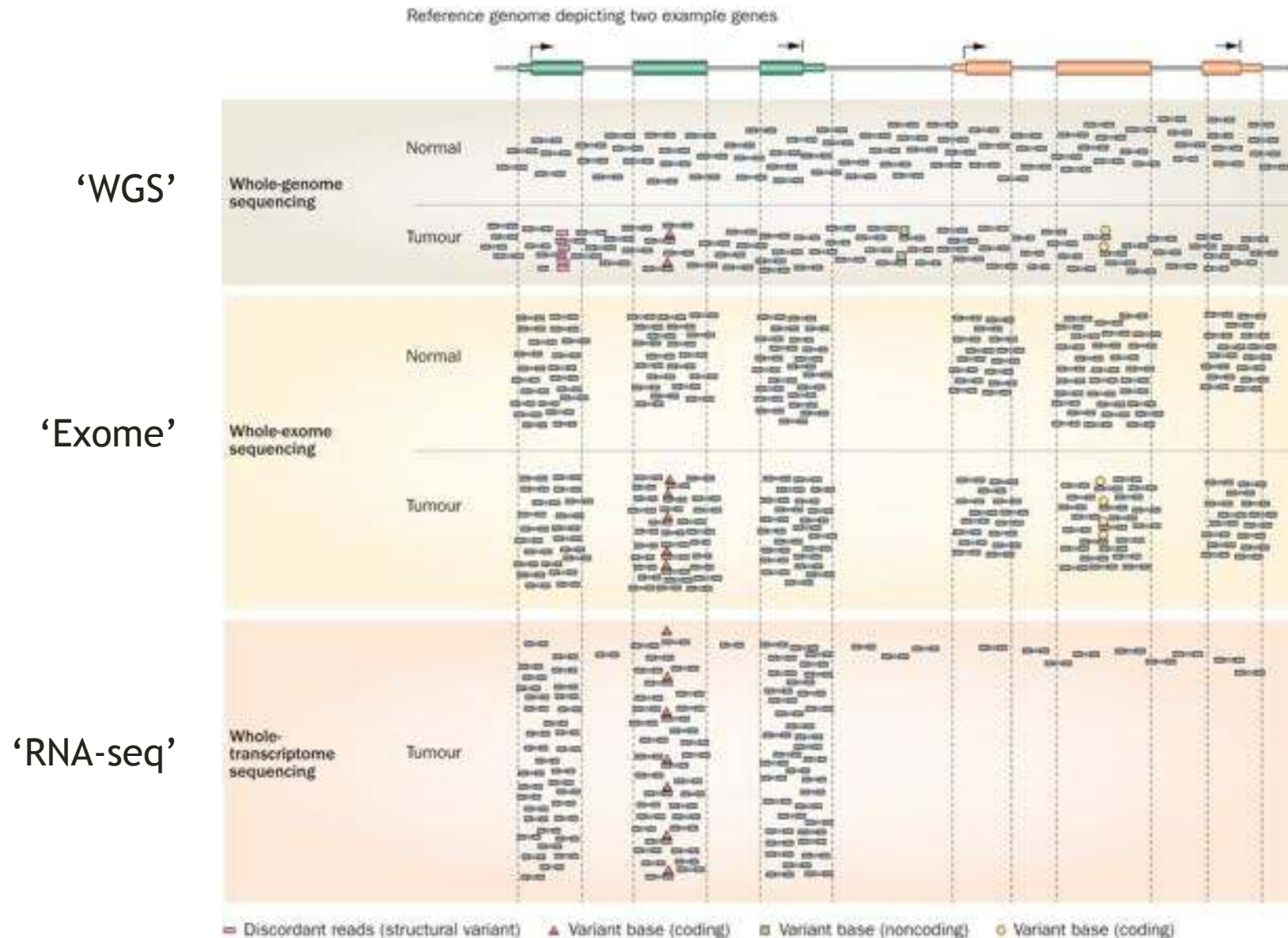


regtools

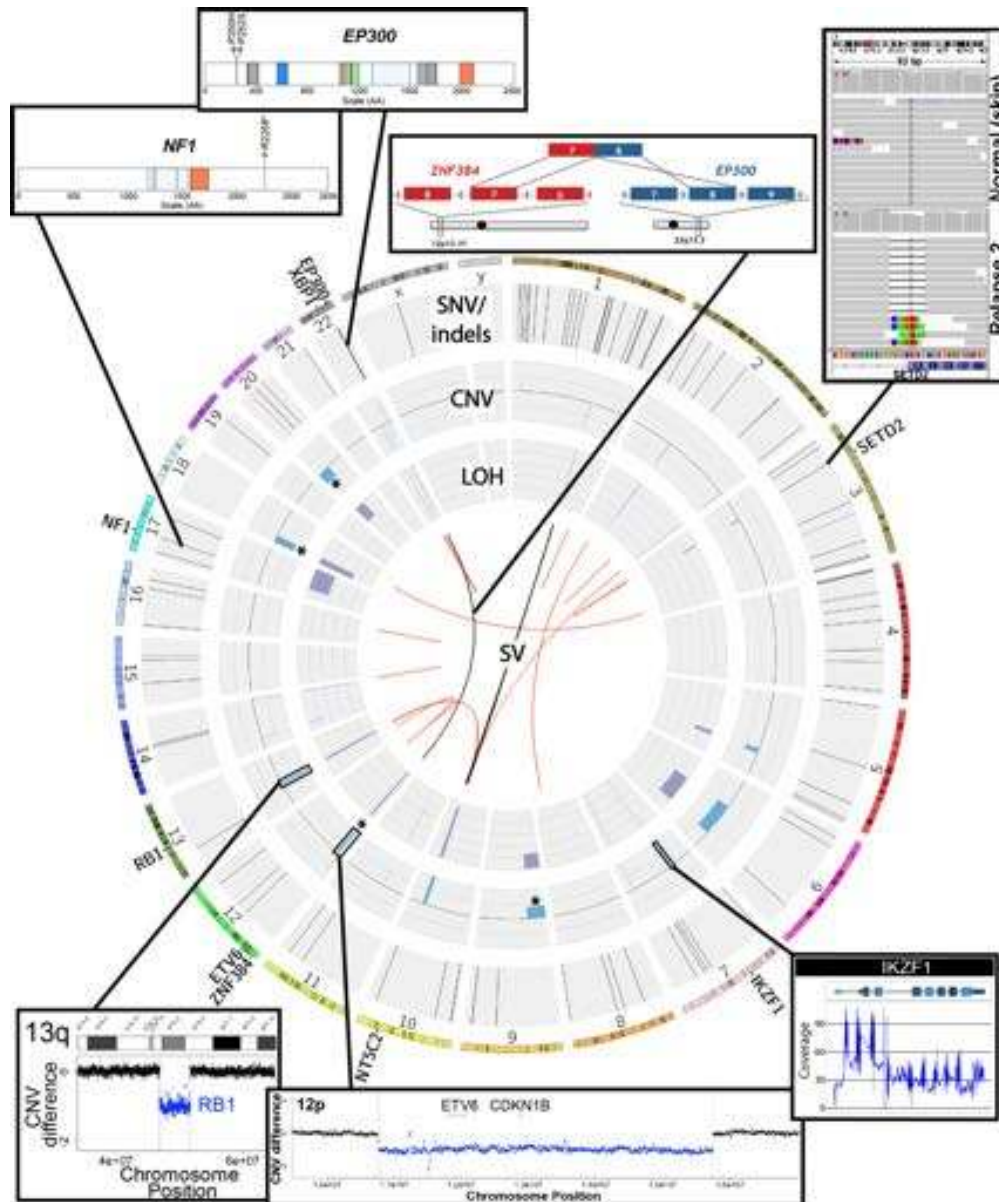
<https://github.com/griffithlab/regtools>

Identify regulatory variants

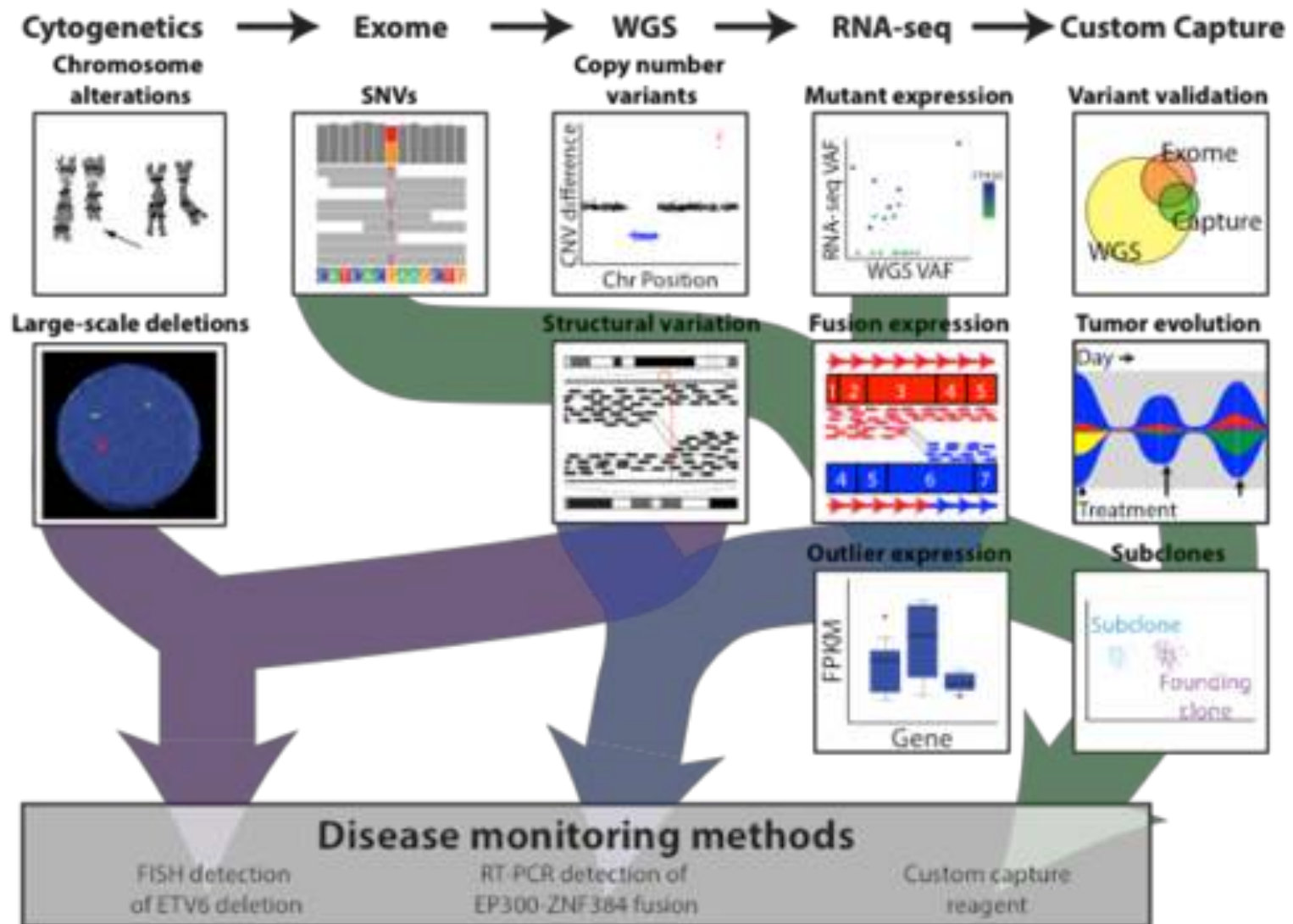
Whole genome, exome, transcriptome and other ‘omic’ sequencing allows us to detect and confirm many different variant types



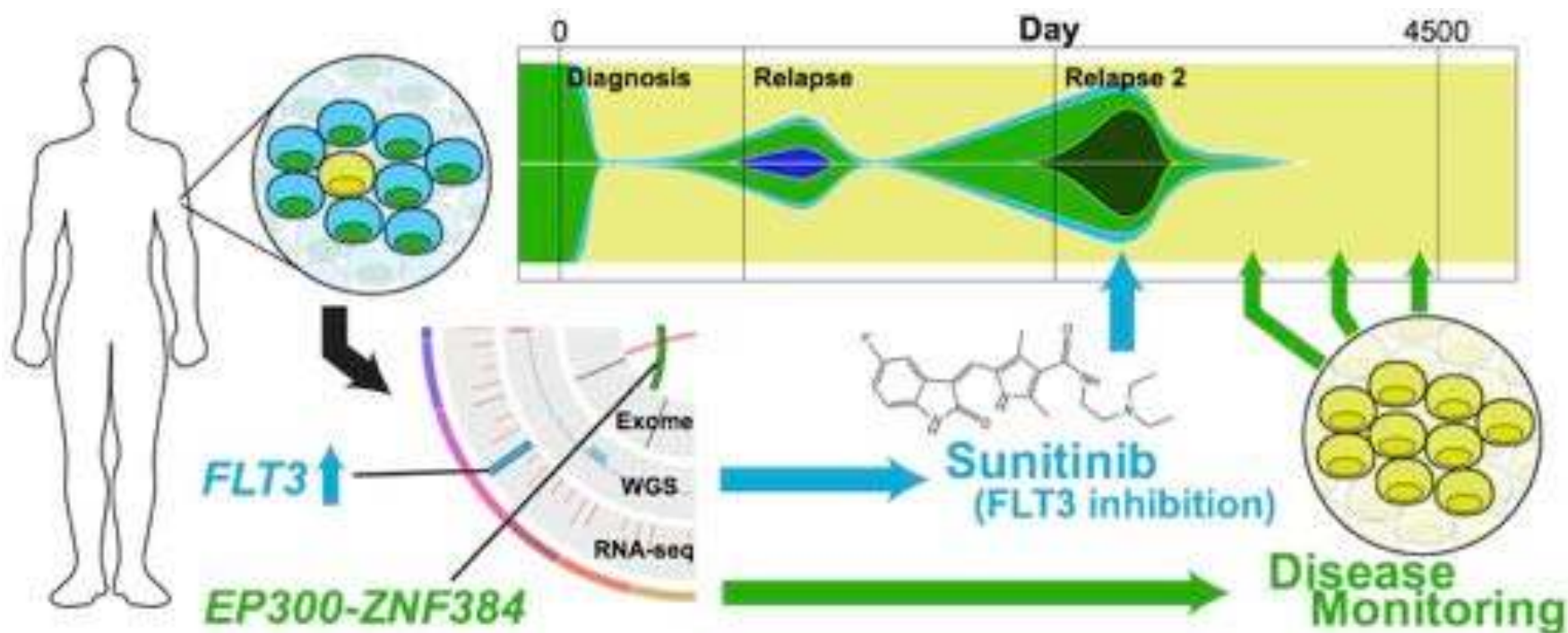
SNVs, Indels, CNVs, SVs, fusions, LOH, expression changes, methylation changes, and more



Comprehensive and integrative analysis methods are needed



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](#)

Strategies to bring genomics information to bear for as many cancer patients as possible

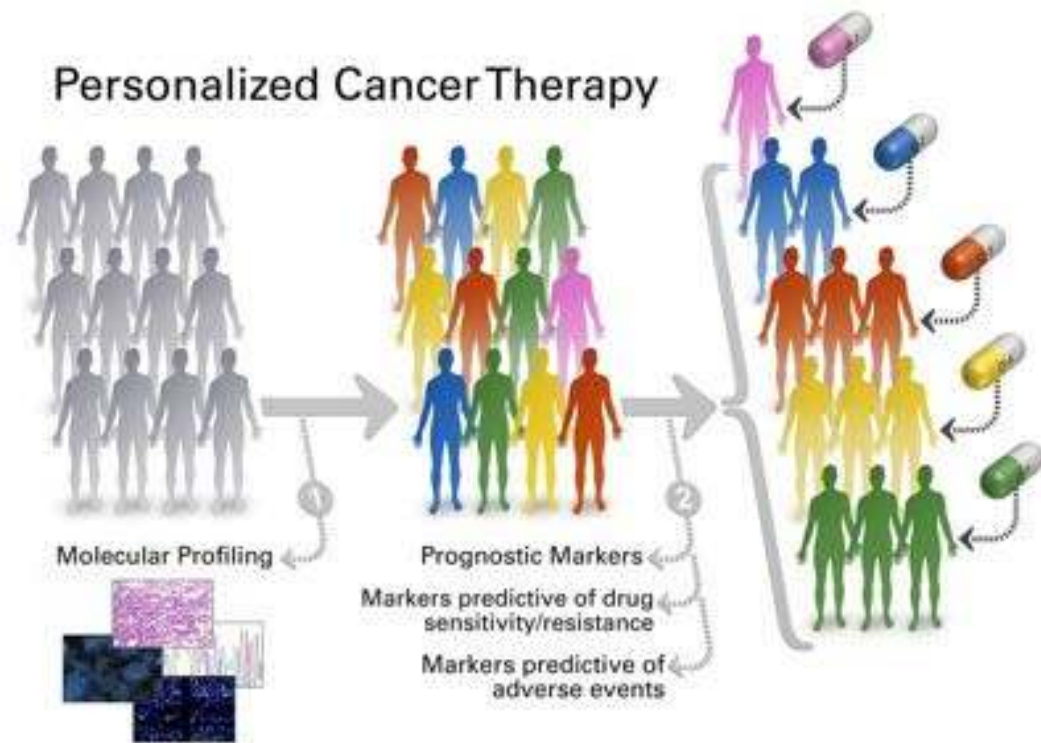
1. Precision medicine targeting of driver mutations

2. Leveraging passenger variants

- a. Tracking minimal residual disease
- b. Identifying neoepitopes
 - Predicting response to immunotherapy
 - Developing personalized vaccines

Precision medicine targeting of driver mutations

Personalized Cancer Therapy



1980s: Development of Targeted Therapies

2000: Human Genome Sequencing Project

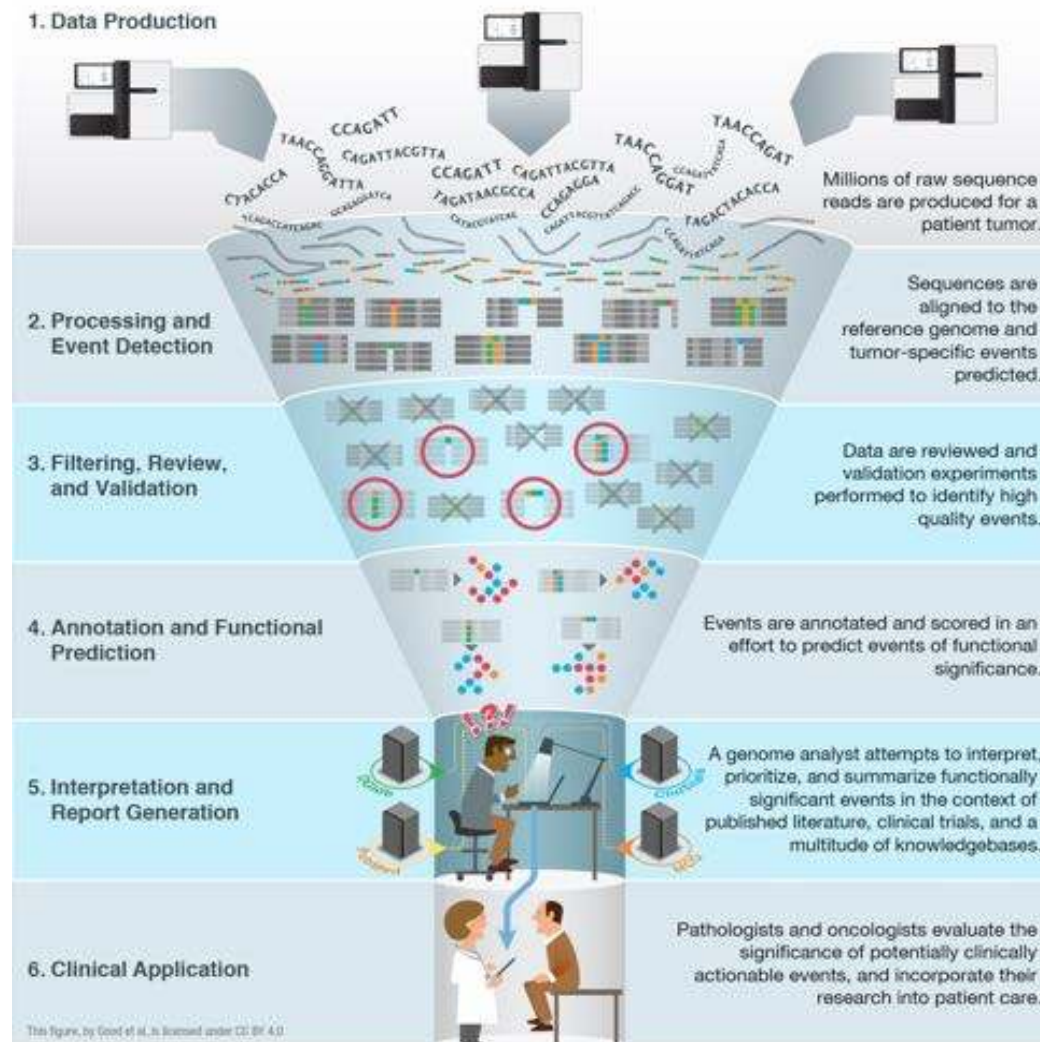
2001: FDA Approval for Imatinib in BCR-ABL1 CML

2014: FDA Approval for BRCA Testing

2017: FDA Approval of Pembrolizumab for MSI tumors

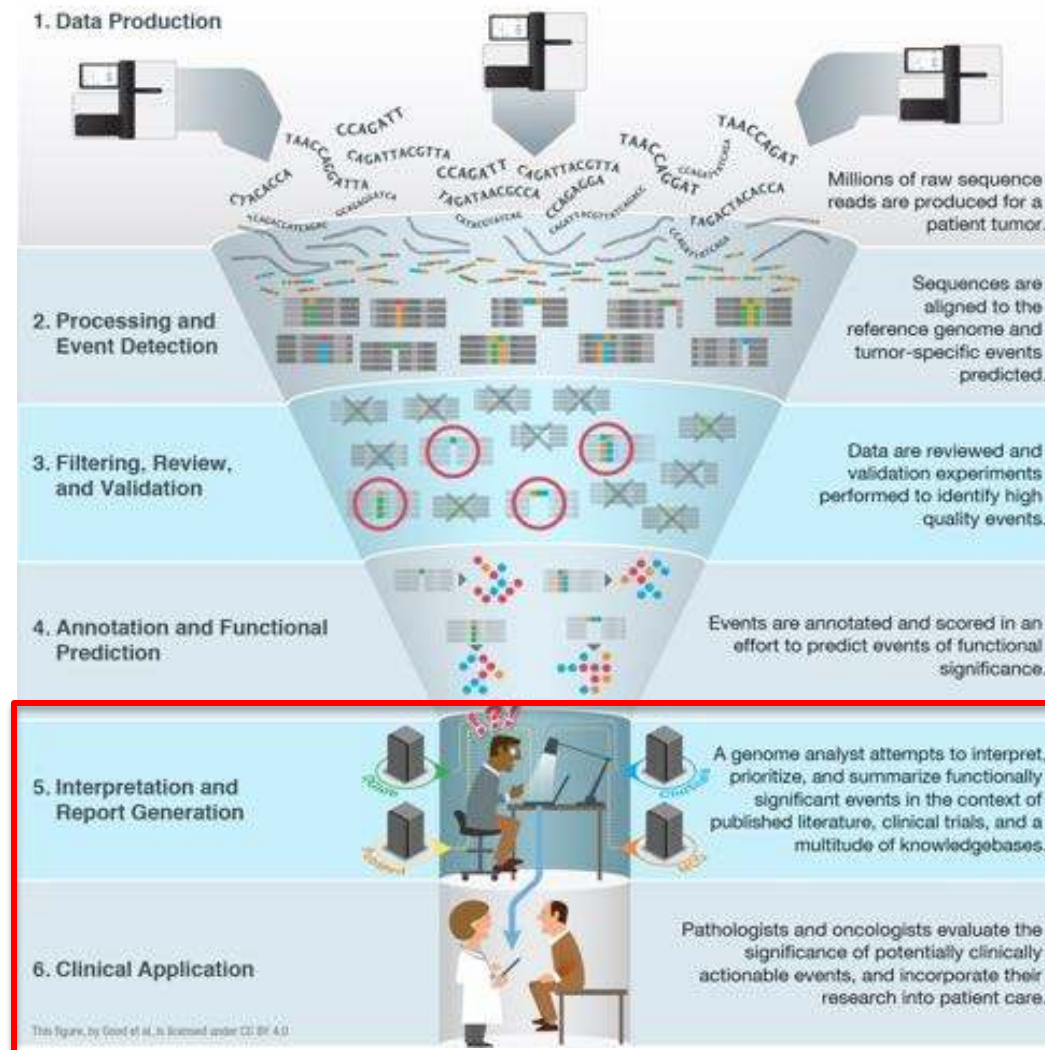
BRAF -> V600E -> Melanoma -> Predictive -> Vemurafenib
ERBB2 -> Amplification -> Breast -> Predictive -> Trastuzumab
EGFR -> L858R -> Lung -> Predictive -> Erlotinib
ALK -> Fusions -> Lung -> Predictive -> Crizotinib
EWSR1-FLI1 -> Fusions -> Ewing Sarcoma -> Diagnostic
DNAJB1-PRKACA -> Fusions -> fHCC -> Diagnostic
VHL -> Loss of function mutations -> Kidney -> Predisposing
... an increasingly long tail of rare but clinically relevant variants

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



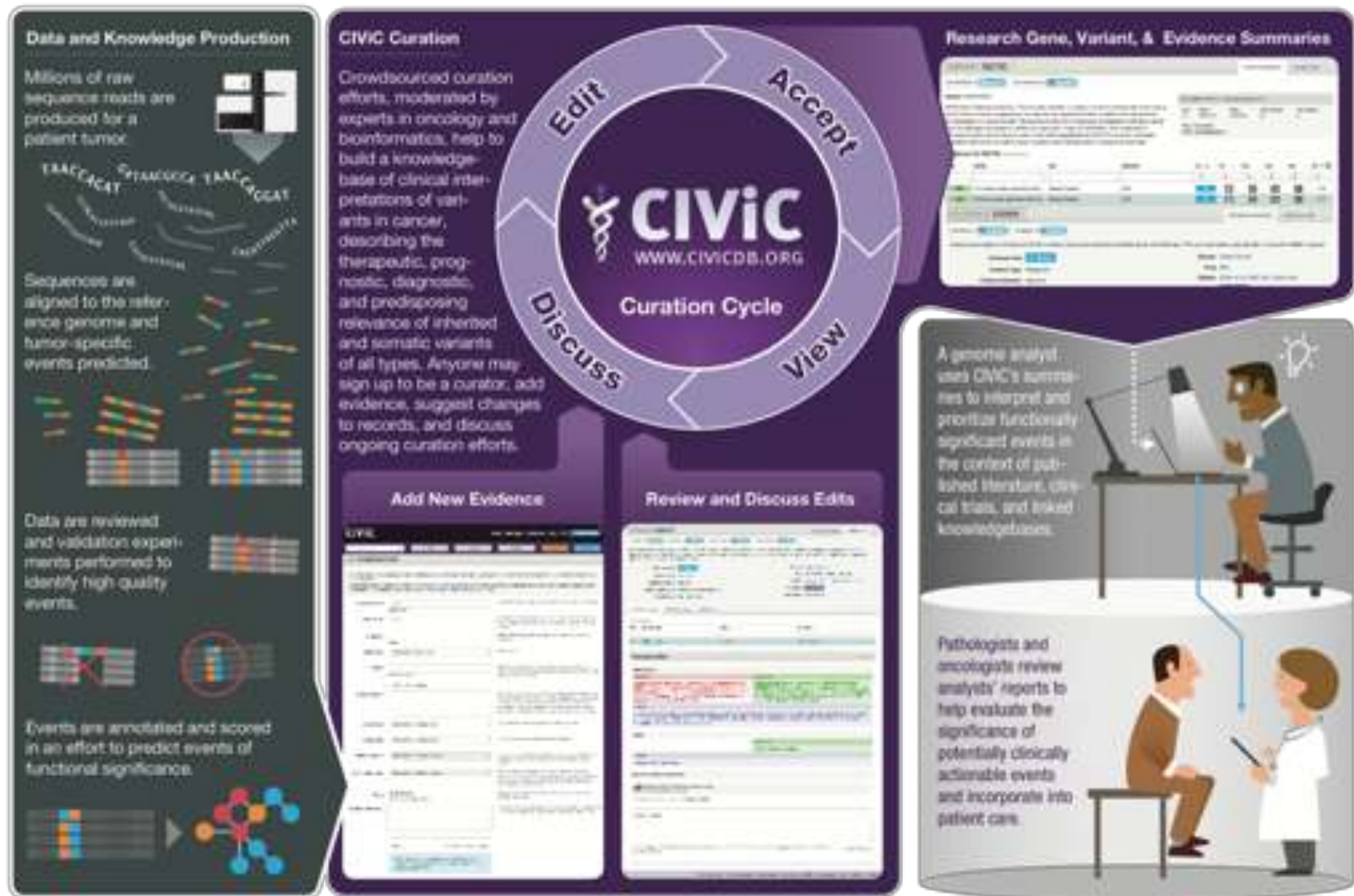
Good BM, Ainscough BJ, McMichael JF, Su AI, Griffith OL. 2014. Genome Biology. 15(8):438.

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.

We created CIViC to address this need - an open knowledgebase and curation system for clinical interpretation of variants in cancer

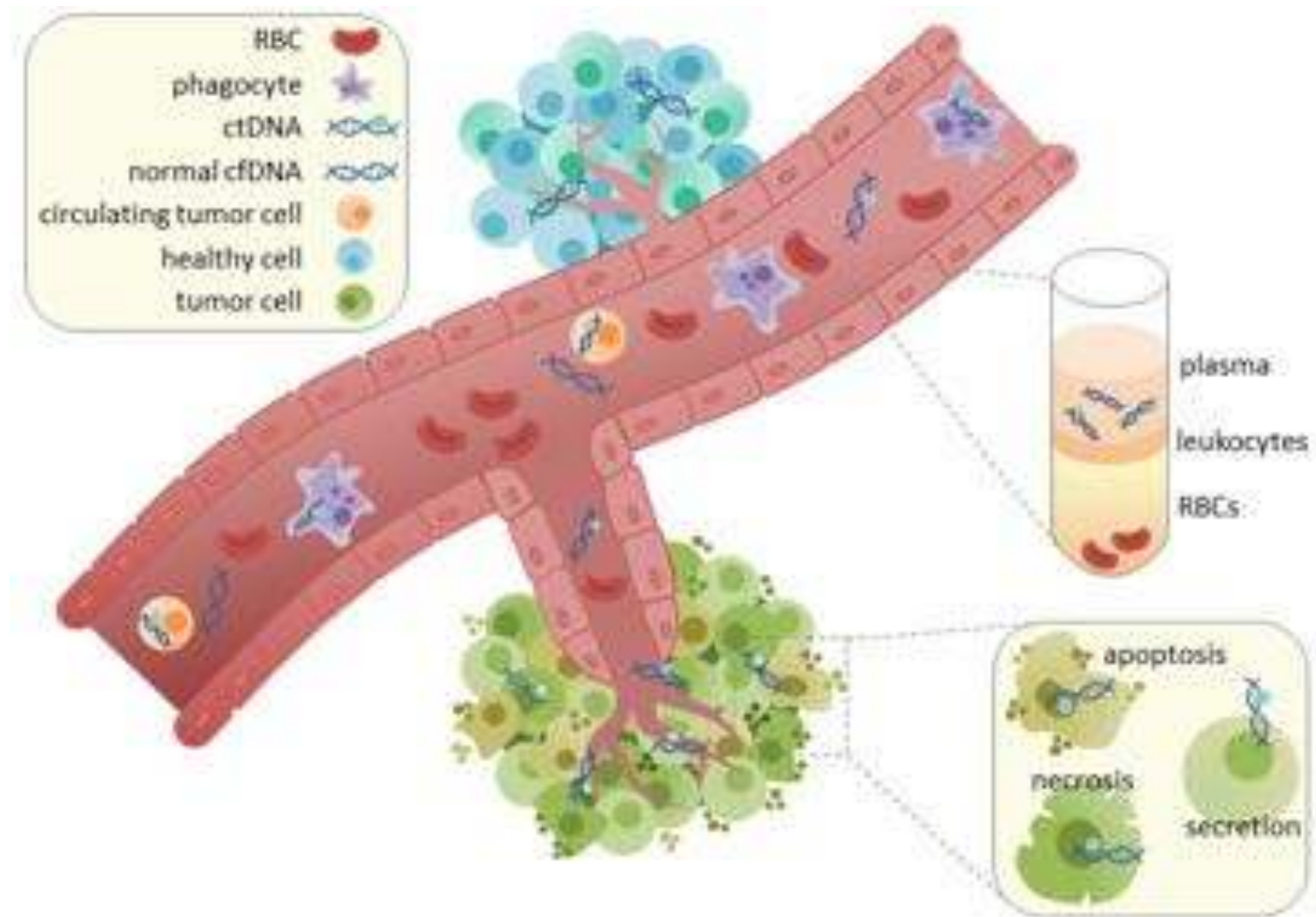


www.civicdb.org

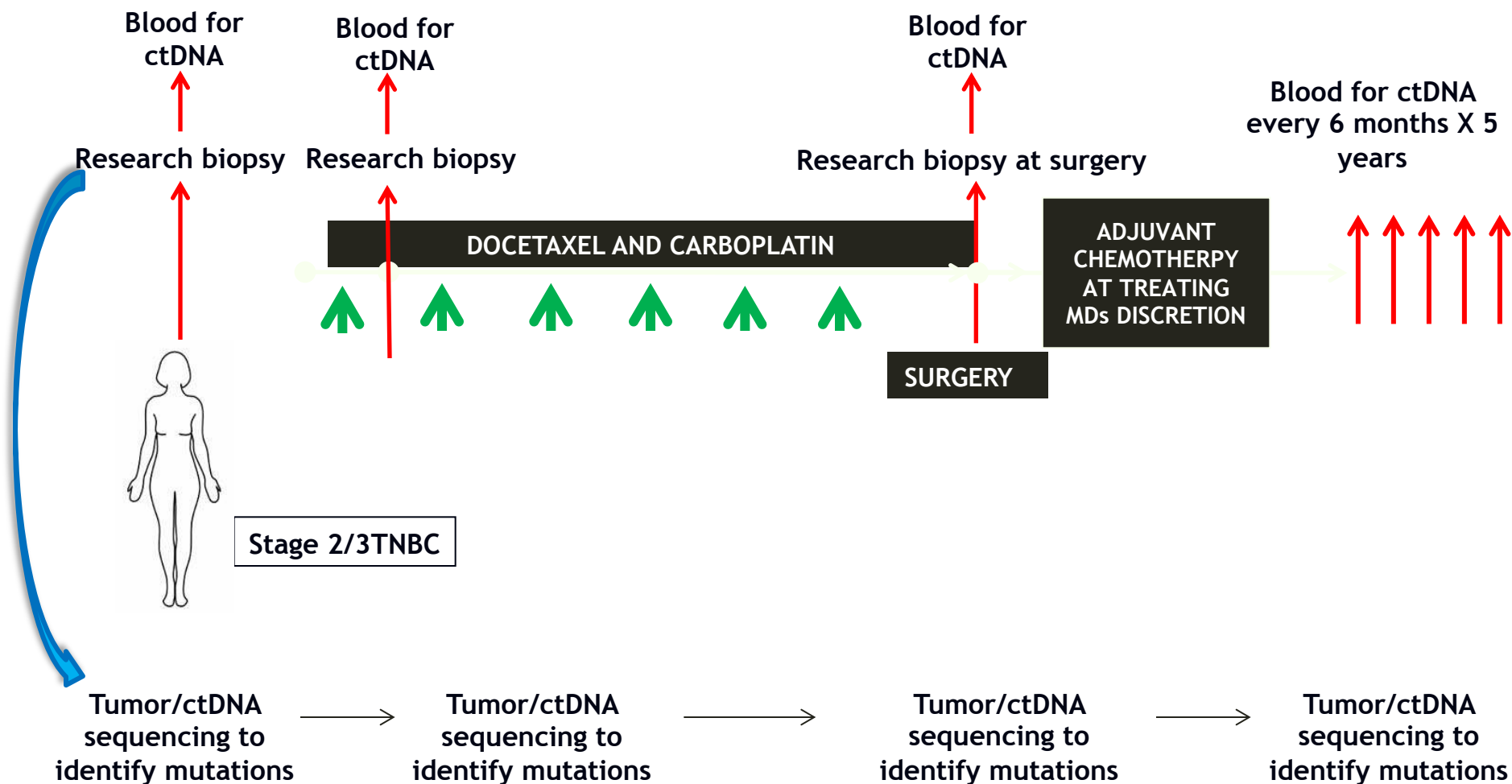
Strategies to bring genomics information to bear for as many cancer patients as possible

1. Precision medicine targeting of driver mutations
2. Leveraging passenger variants
 - a. Tracking minimal residual disease
 - b. Identifying neoepitopes
 - Predicting response to immunotherapy
 - Developing personalized vaccines

Circulating tumor DNA (ctDNA) could allow generalized tracking in any cancer type



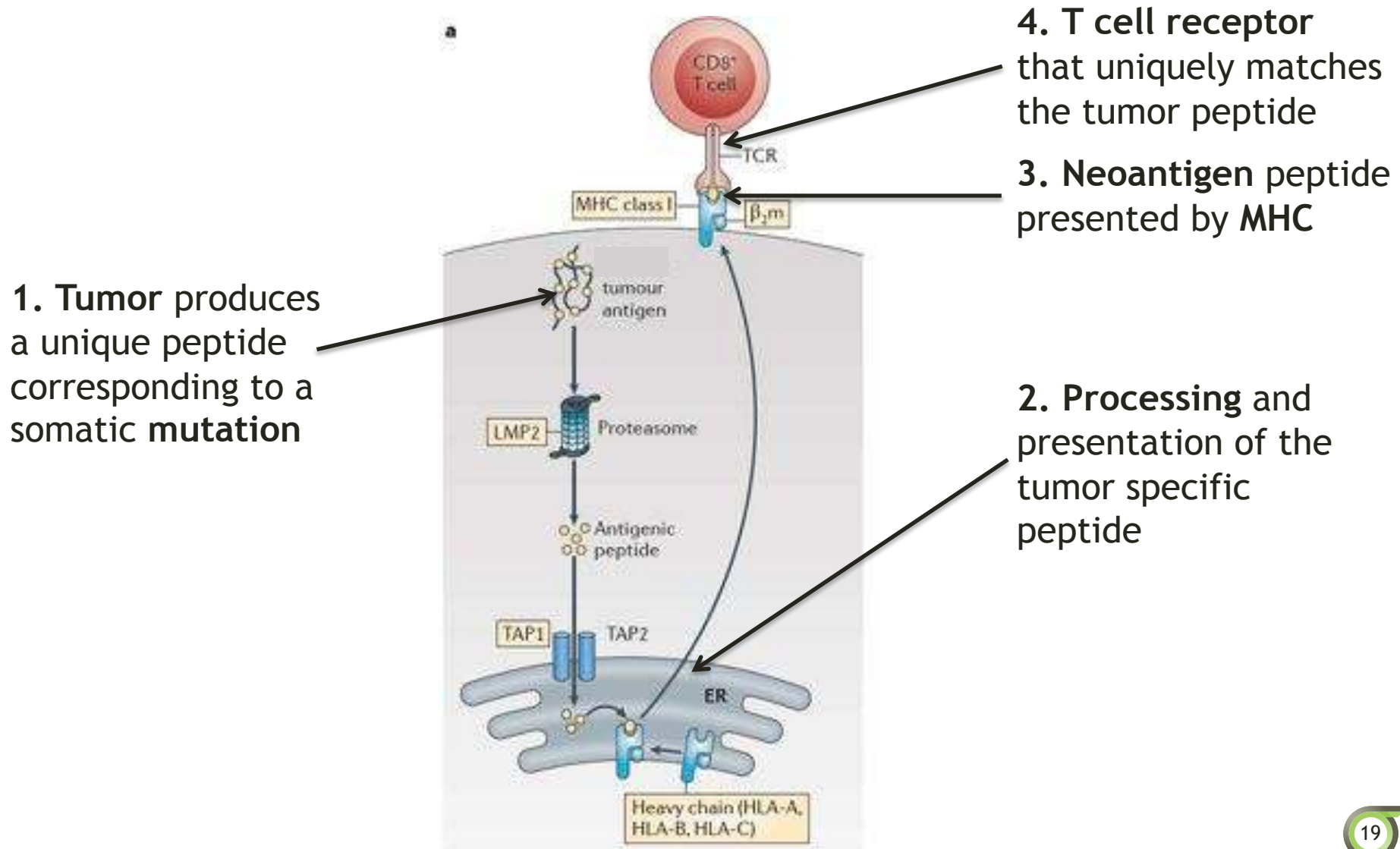
ctDNA tracking in triple negative breast cancer



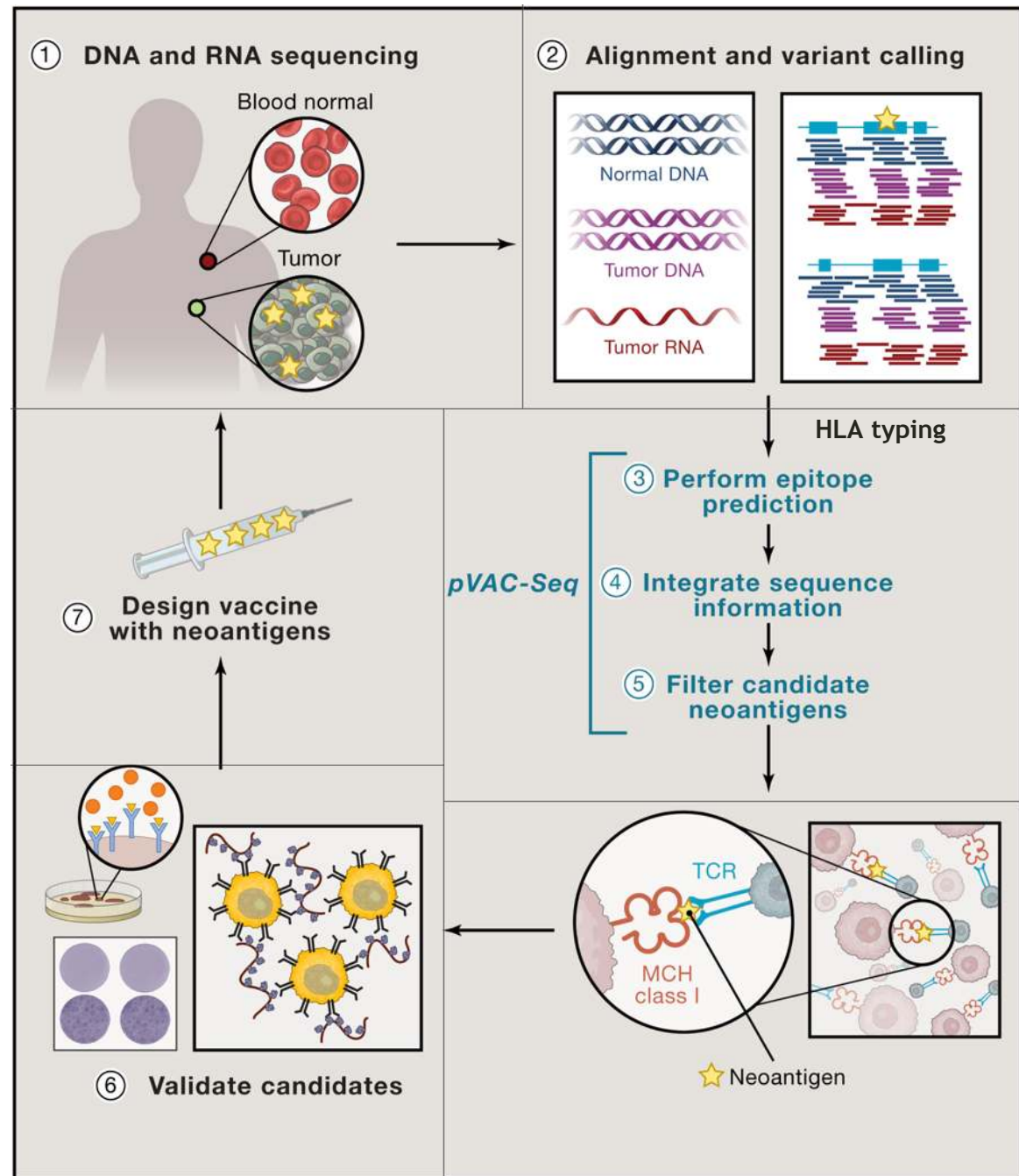
Strategies to bring genomics information to bear for as many cancer patients as possible

1. Precision medicine targeting of driver mutations
2. Leveraging passenger variants
 - a. Tracking minimal residual disease
 - b. Identifying neoepitopes
 - Predicting response to immunotherapy
 - Developing personalized vaccines

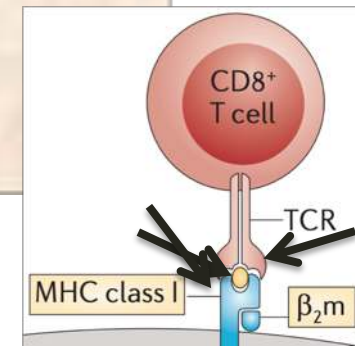
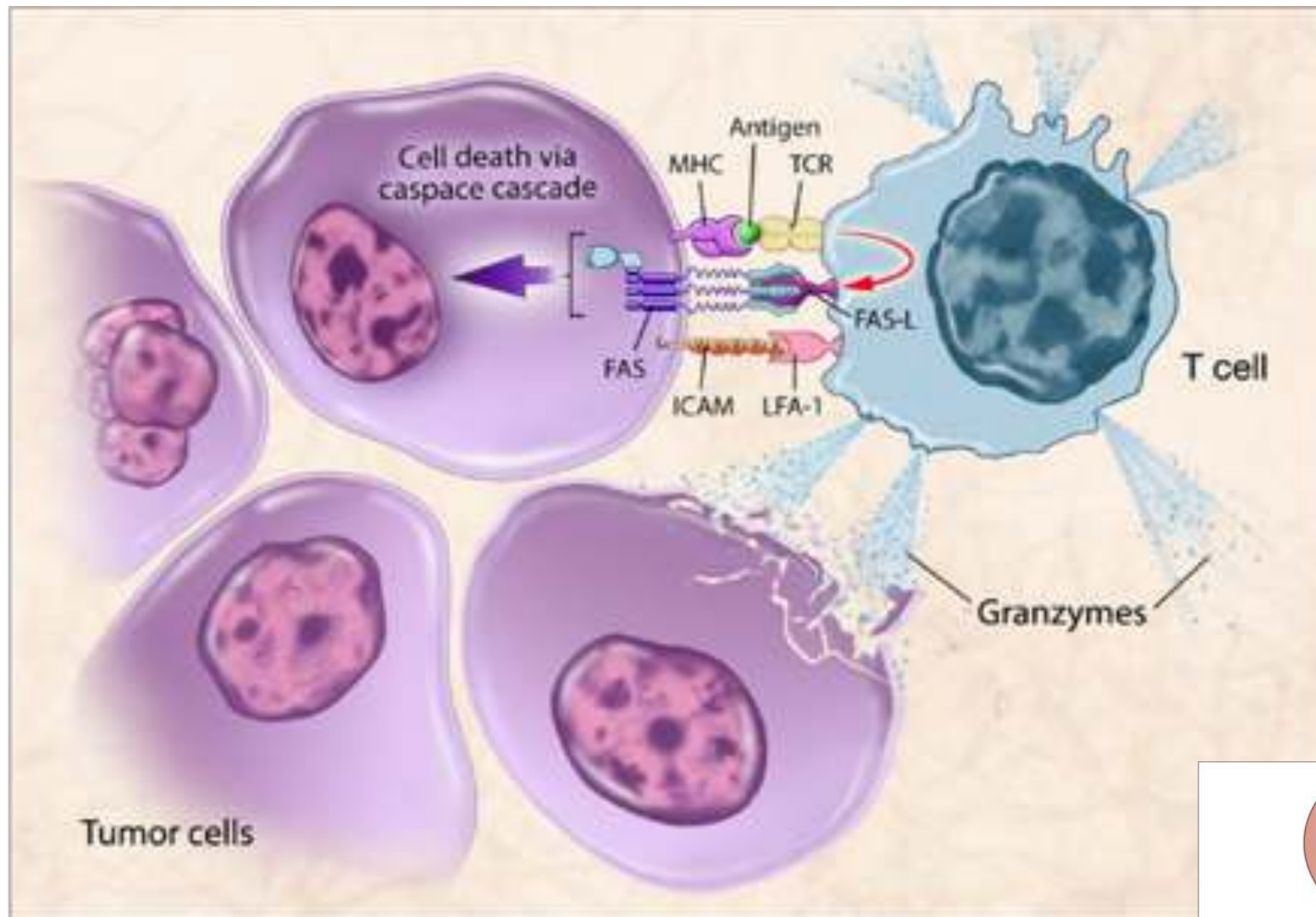
Invoking an adaptive immune response against the tumor (focus on CD8+ T cells)



Neoepitope characterization workflow



T cell mediated cell death

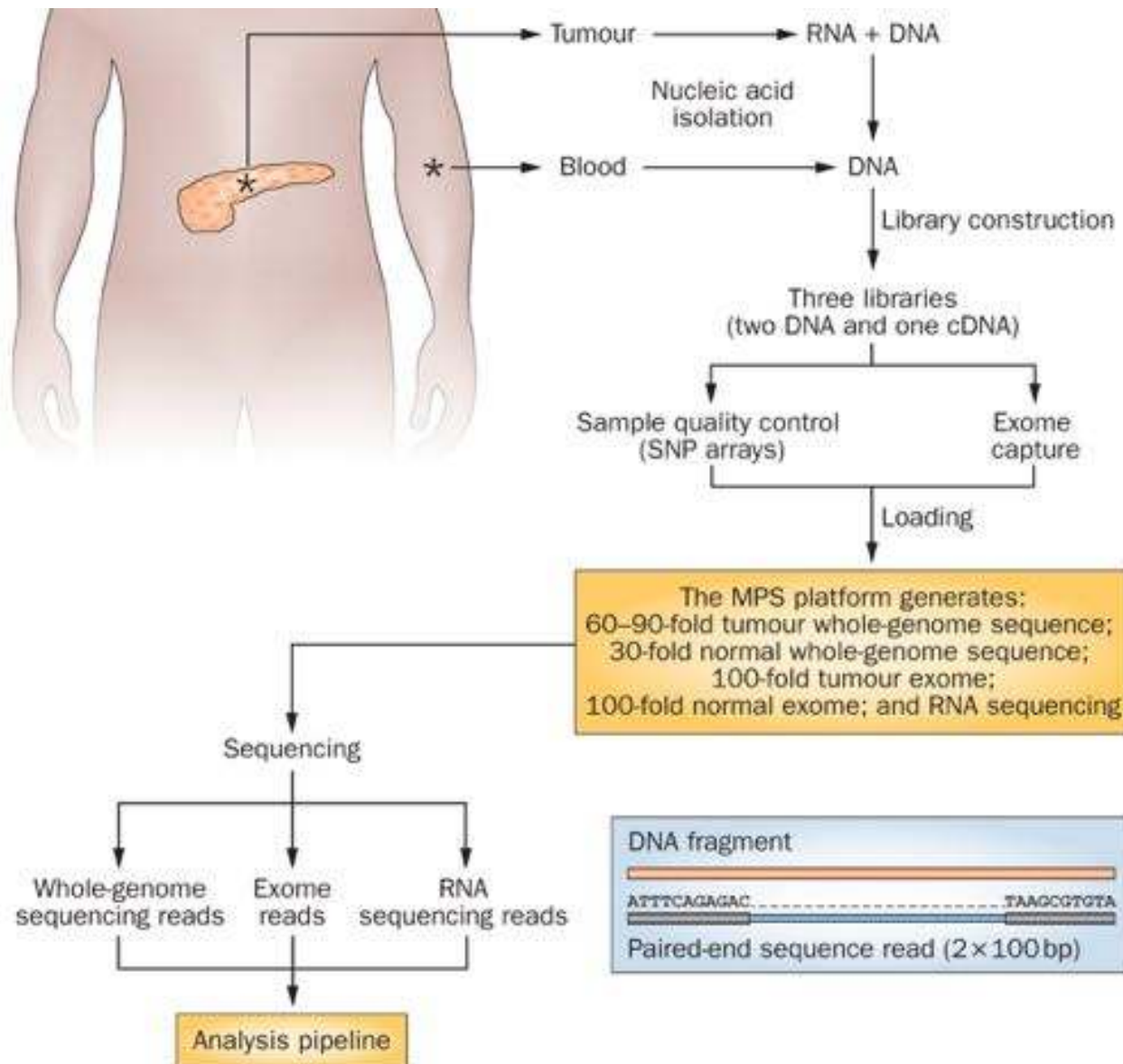


Personalized cancer vaccine trials

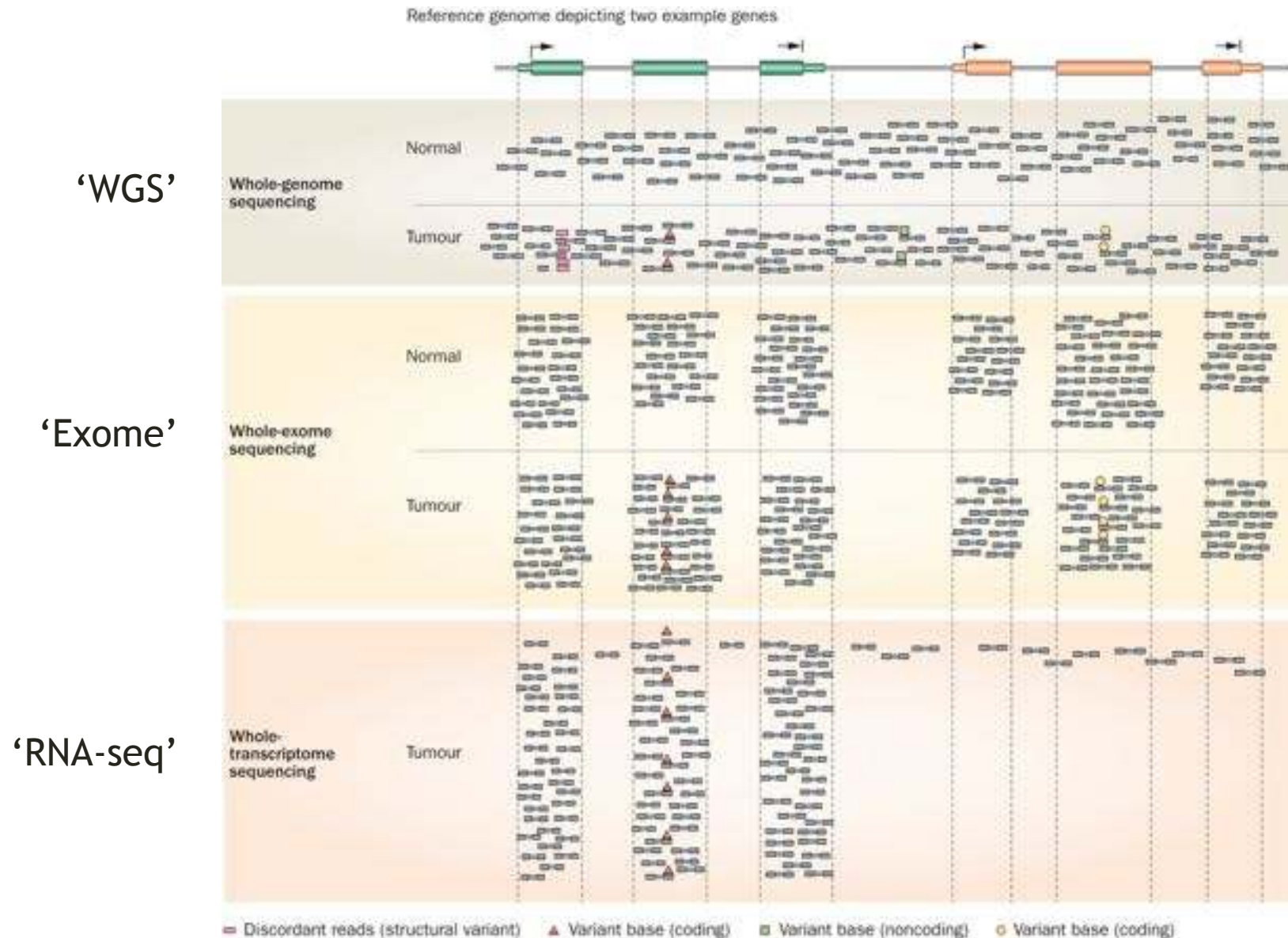
- Kidney Cancer (n = 15 patients)
 - PolyImmune {Durvalumab (MEDI4736) and Tremelimumab} & Vaccine Orchestrated Treatment for Patients With Advanced/Metastatic Renal Cell Carcinoma ([NCT03598816](#)). Collaboration with MedImmune.
- Lung Cancer (n = 15 patients)
 - A Personal Cancer Vaccine (NEO-PV-01) With Pembrolizumab and Chemotherapy for Patients With Lung Cancer ([NCT03380871](#)). Collaboration with Neon Therapeutics and Merck.
- Glioblastoma (n = 30 patients)
 - Neoantigen-based Personalized Vaccine Combined With Immune Checkpoint Blockade Therapy in Patients With Newly Diagnosed, Unmethylated Glioblastoma ([NCT03422094](#))
- Breast Cancer (n = 54 patients)
 - Neoantigen DNA Vaccine Alone vs. Neoantigen DNA Vaccine Plus Durvalumab in Triple Negative Breast Cancer Patients Following Standard of Care Therapy ([NCT03199040](#)). Collaboration with MedImmune.
 - Safety and Immunogenicity of a Personalized Polyepitope DNA Vaccine Strategy in Breast Cancer Patients With Persistent Triple-Negative Disease Following Neoadjuvant Chemotherapy ([NCT02348320](#))
- Follicular Lymphoma (n = 20 patients)
 - Personalized Tumor Vaccine Strategy and PD-1 Blockade in Patients With Follicular Lymphoma ([NCT03121677](#)). Collaboration with Bristol-Myers Squibb.
- Prostate Cancer (n = 20 patients)
 - Neoantigen DNA Vaccine in Combination With Nivolumab/Ipilimumab and PROSTVAC in Metastatic Hormone-Sensitive Prostate Cancer ([NCT03532217](#)). Collaboration with Bristol-Myers Squibb.
- Pancreatic Cancer (n = 15 patients)
 - Neoantigen DNA Vaccine in Pancreatic Cancer Patients Following Surgical Resection and Adjuvant Chemotherapy ([NCT03122106](#)).
- Melanoma (n = 12 patients)
 - Dendritic Cell Vaccination in Patients With Advanced Melanoma ([NCT03092453](#)). Collaboration with UPenn/Parker ICI.

Genomic data visualization and interpretation

Genomics research has exploded with the rapid advances in DNA sequencing technologies



Whole genome, exome and transcriptome sequencing allows us to detect and confirm many different variant types

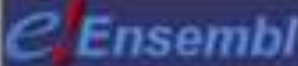


Why do we create visualizations of genomic data?

- Data exploration and interpretation of results
 - QC analysis
 - Understanding whether/how an experiment worked
 - Discovery
- Communication
 - Slides for presentations
 - e.g. Keynote, Powerpoint, etc.
 - Figures for publications
 - e.g. PDFs, PNGs, etc.
 - Illustrator, Gimp, Inkscape, etc.
 - R and R Studio

Genome browsers

Genome browsers - Ensembl

 [BLAST/BLAT](#) | [BioMart](#) | [Tools](#) | [Downloads](#) | [Help & Documentation](#) | [Blog](#) | [Mirrors](#)

Search:  

e.g. BRCA2 or rat 5:62797383-63627669 or rs899 or coronary heart disease

Browse a Genome

Ensembl is a genome browser for vertebrate genomes that supports research in comparative genomics, evolution, sequence variation and transcriptional regulation. Ensembl annotates genes, computes multiple alignments, predicts regulatory function and collects disease data. Ensembl tools include BLAST, BLAT, BioMart and the Variant Effect Predictor (VEP) for all supported species.

Favourite genomes

 **Human**
GRCh38.p13

 **Mouse**
GRCm38.p6

 **Zebrafish**
GRCz10

[Edit favourites](#)

All genomes



[View full list of all Ensembl species](#)

Find a Data Display

Not sure how to find the data visualisation you need? With our new [Find a Data Display](#) page, you can choose a gene, region or variant and then browse a selection of relevant visualisations

[Try it now!](#)

Variant Effect Predictor



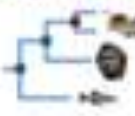
Gene expression in different tissues



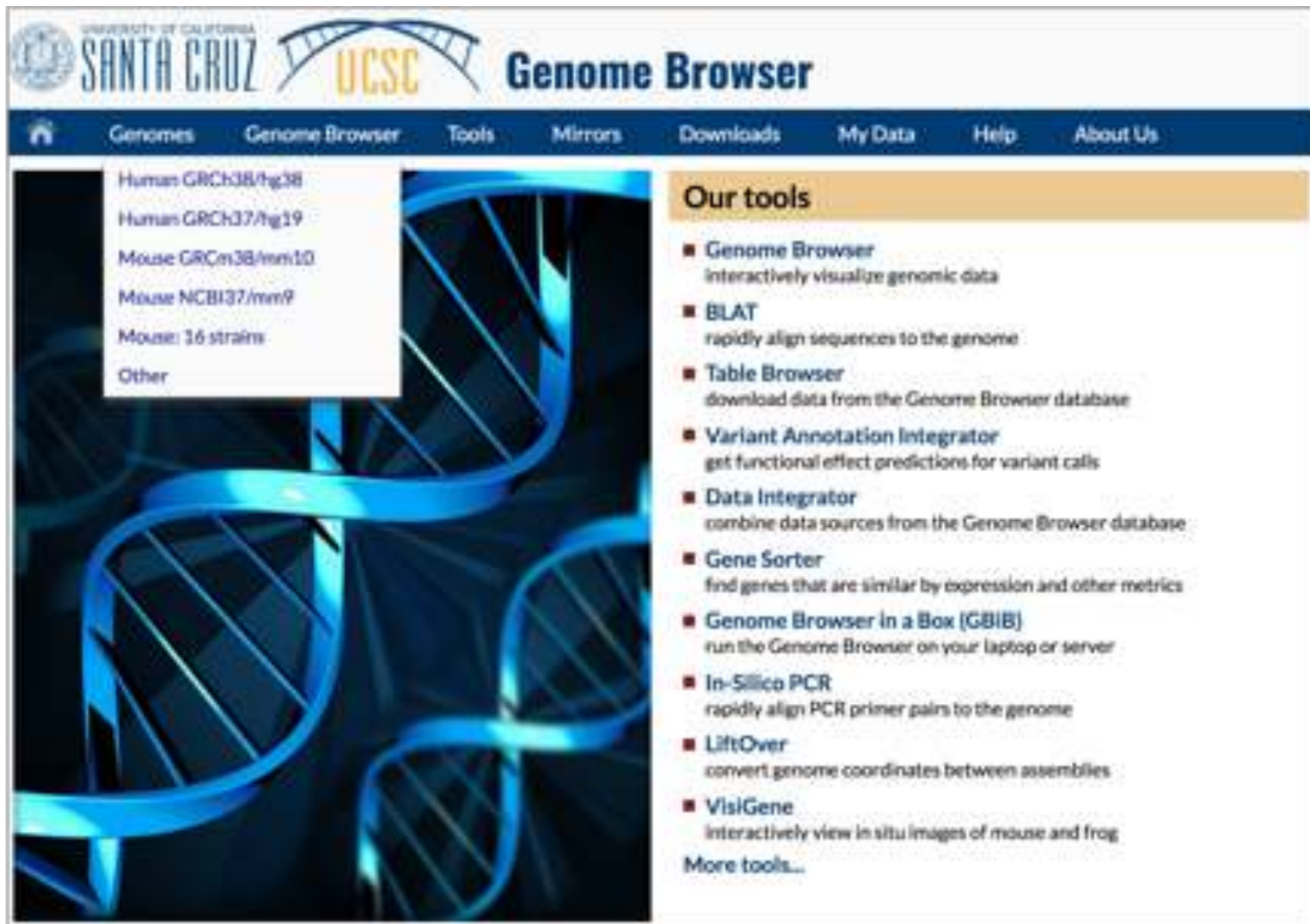
Retrieve gene sequence



Compare genes across species



Genome browsers - UCSC

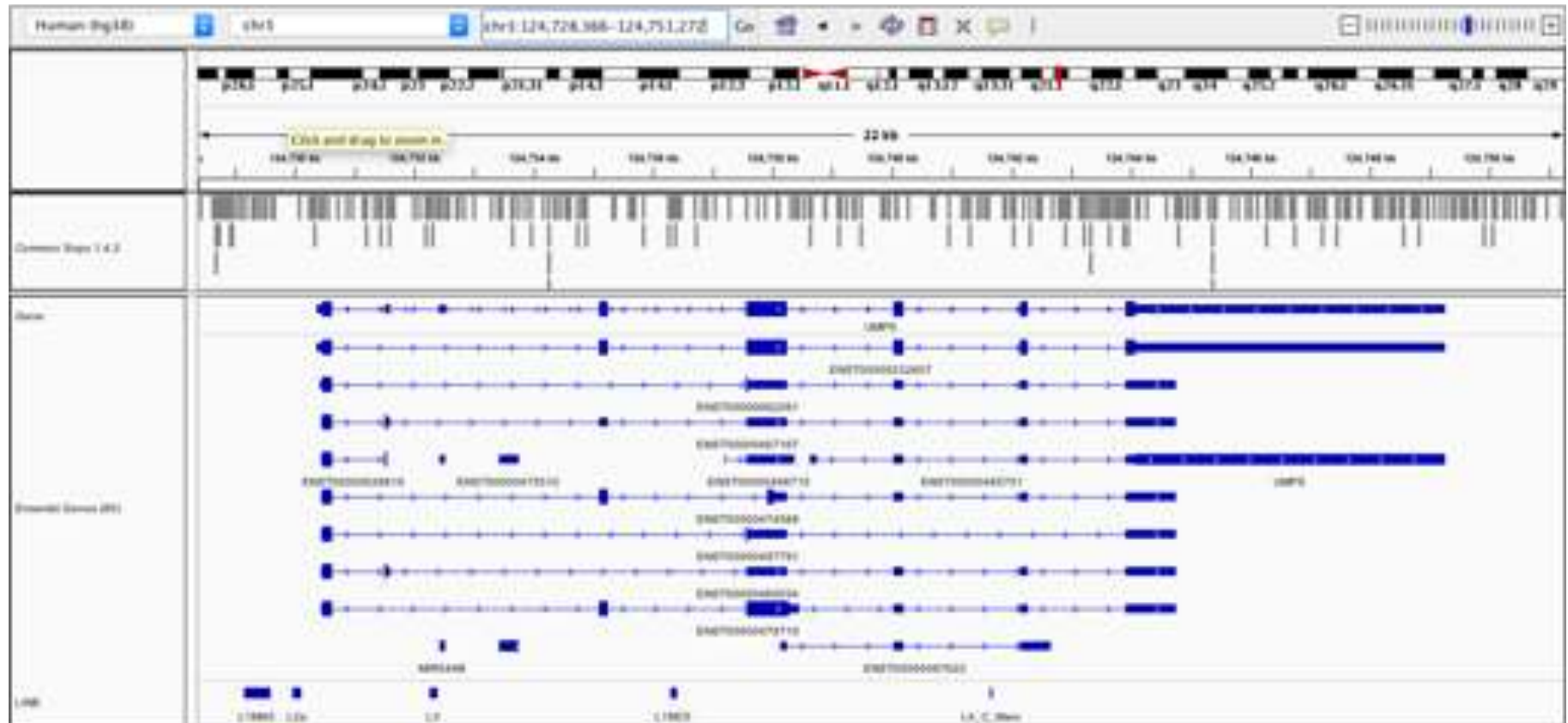


The screenshot shows the UCSC Genome Browser homepage. At the top, there is a header with the University of California Santa Cruz logo, the UCSC logo, and the text "Genome Browser". Below the header is a navigation bar with links: Genomes, Genome Browser, Tools, Mirrors, Downloads, My Data, Help, and About Us. On the left side, there is a dropdown menu for selecting a genome assembly, with options: Human GRCh38/hg38, Human GRCh37/hg19, Mouse GRCh38/mm10, Mouse NCBI37/mm9, Mouse: 16 strains, and Other. The background of the page features a large, stylized blue DNA double helix. On the right side, there is a section titled "Our tools" with a list of tools and their descriptions:

- **Genome Browser**
interactively visualize genomic data
- **BLAT**
rapidly align sequences to the genome
- **Table Browser**
download data from the Genome Browser database
- **Variant Annotation Integrator**
get functional effect predictions for variant calls
- **Data Integrator**
combine data sources from the Genome Browser database
- **Gene Sorter**
find genes that are similar by expression and other metrics
- **Genome Browser in a Box (GBIB)**
run the Genome Browser on your laptop or server
- **In-Silico PCR**
rapidly align PCR primer pairs to the genome
- **LiftOver**
convert genome coordinates between assemblies
- **VisiGene**
interactively view in situ images of mouse and frog

More tools...

Genome browsers - IGV

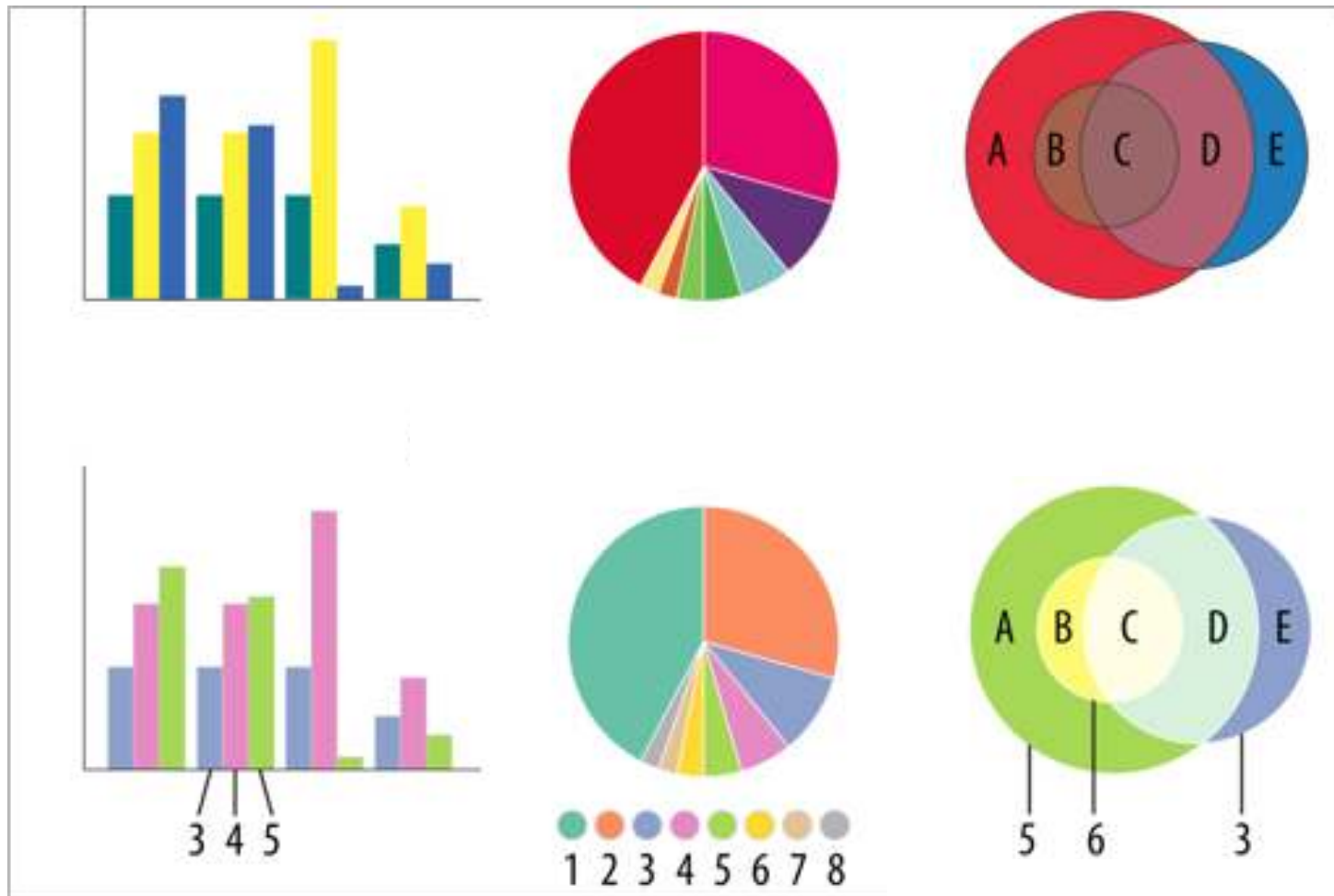


Fundamentals of data visualization

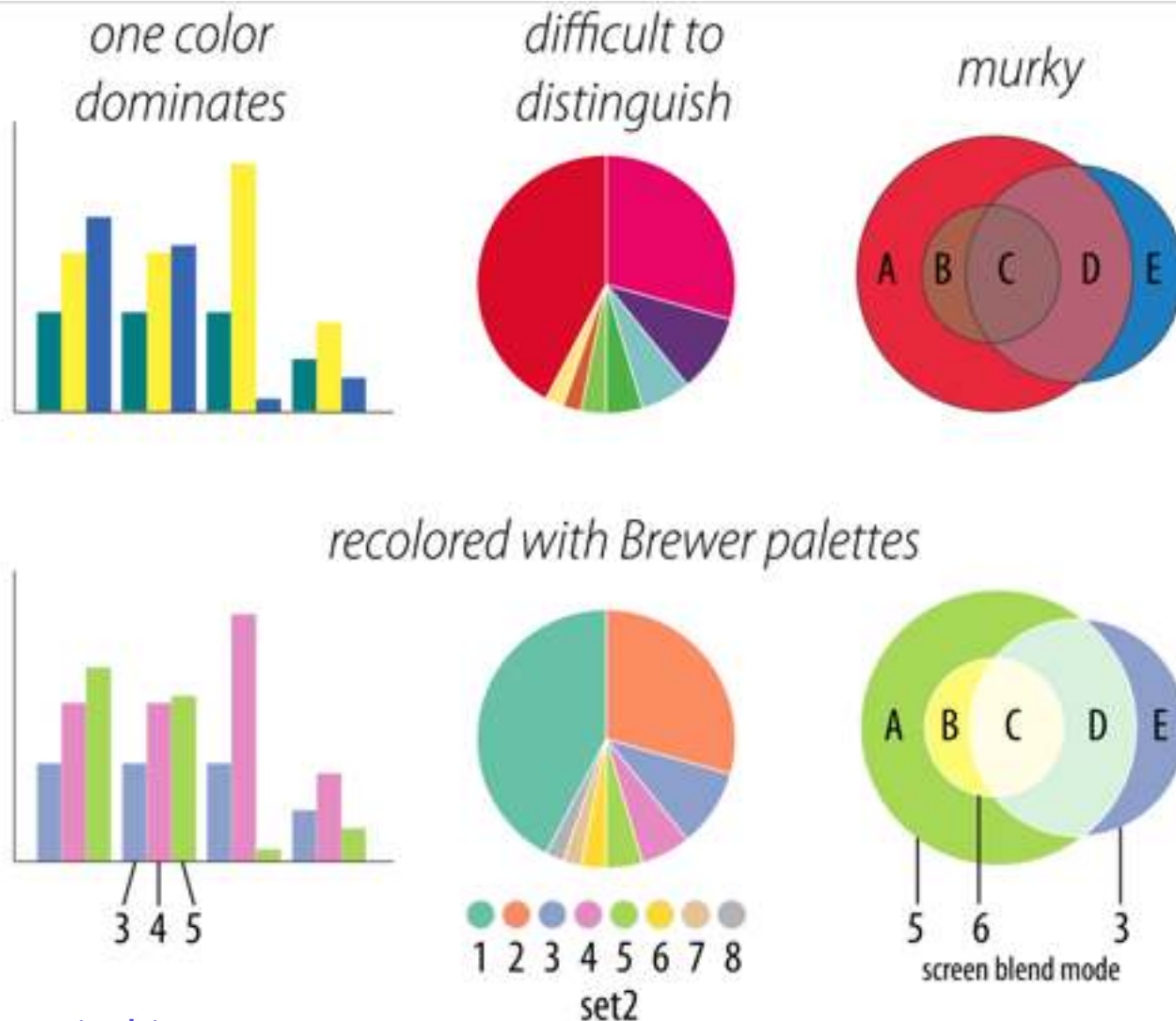
Fundamentals of data visualization

- Where to learn more about the art and science of visualization:
 - Collection of 40 Nature Methods articles on data visualization:
[“Points of View” Articles](#)
 - [Visual design principles lecture](#)

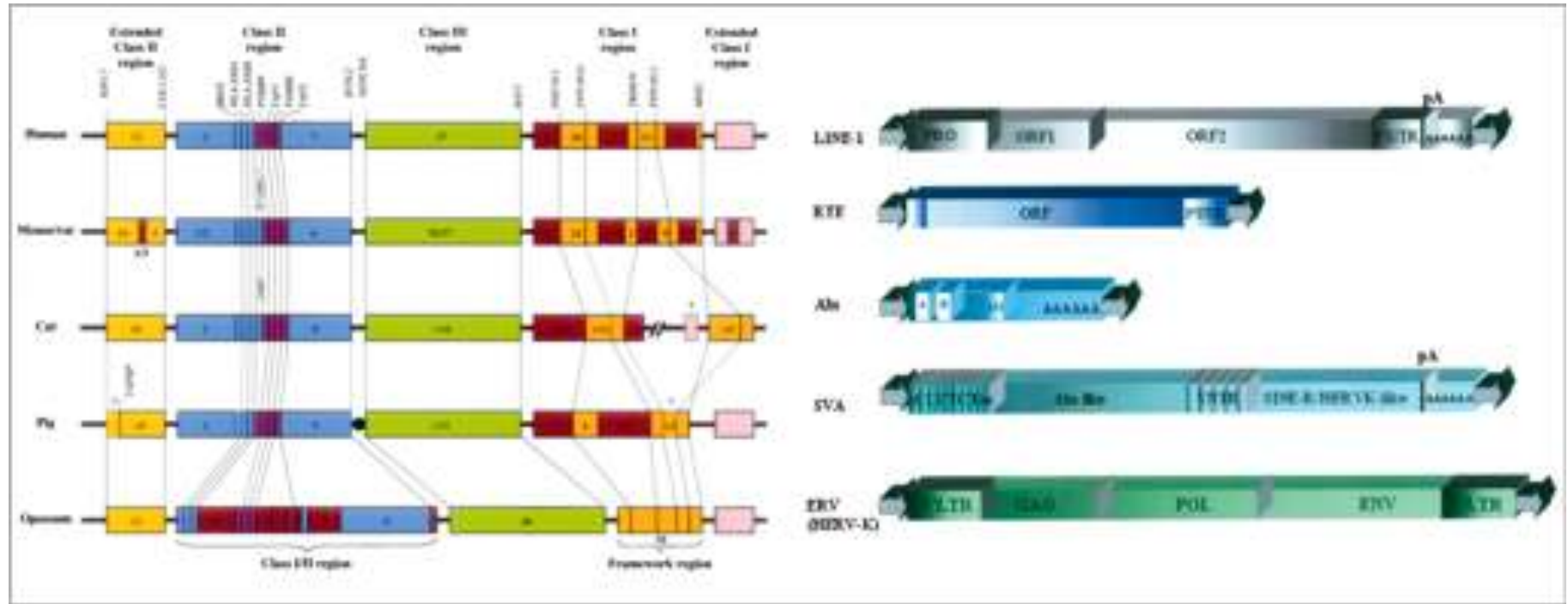
Which series is more effective (top or bottom)?



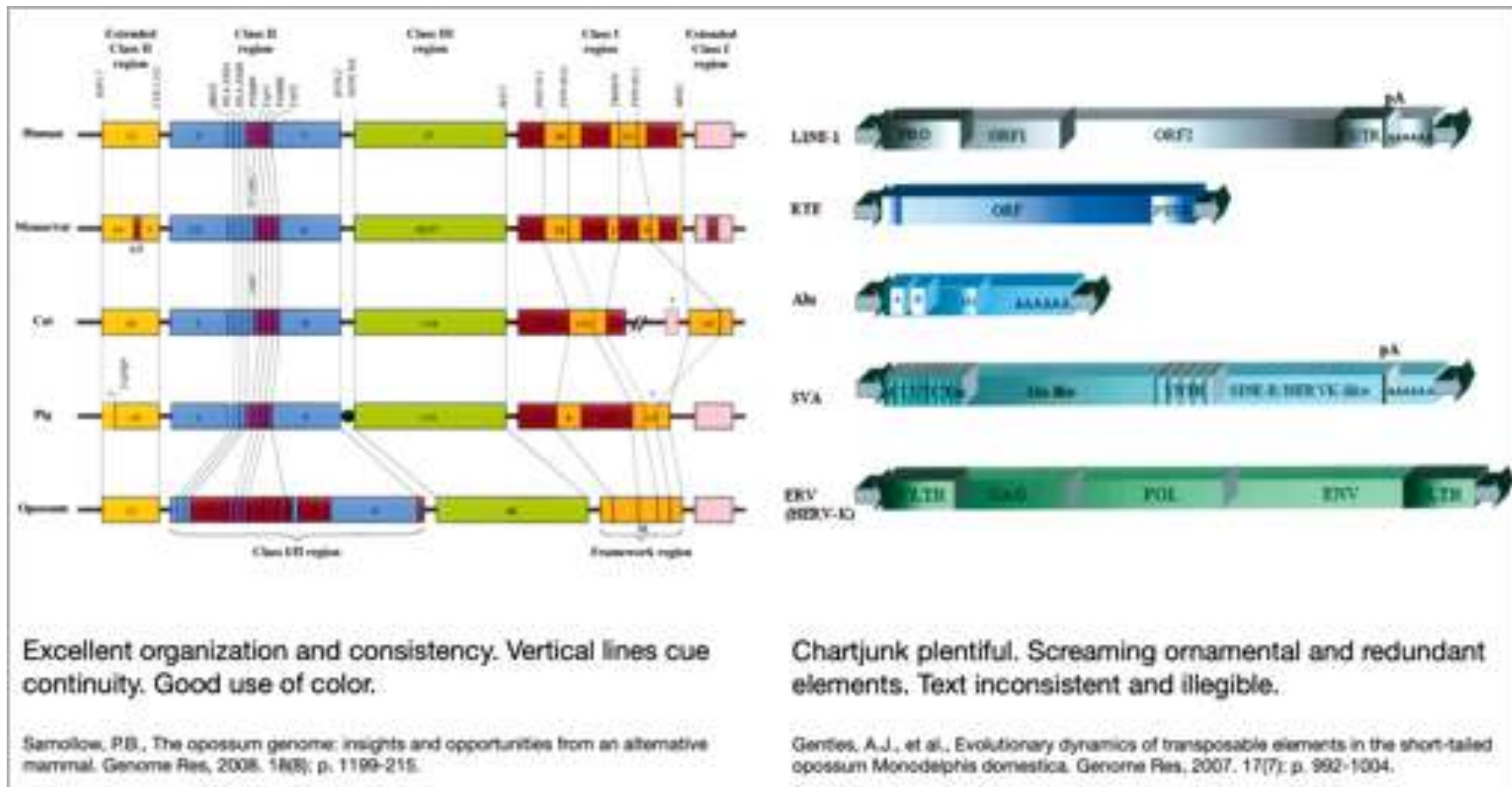
Which series is more effective (top or bottom)?



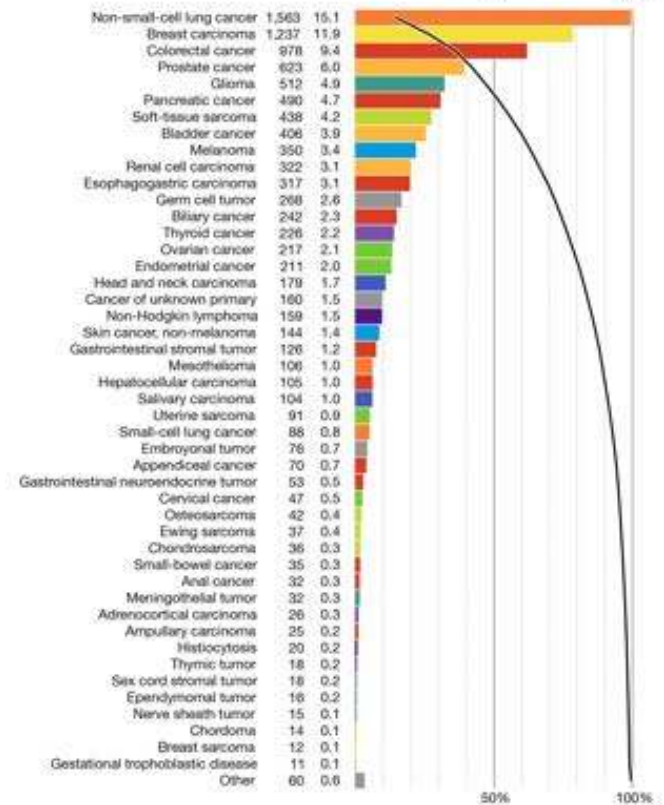
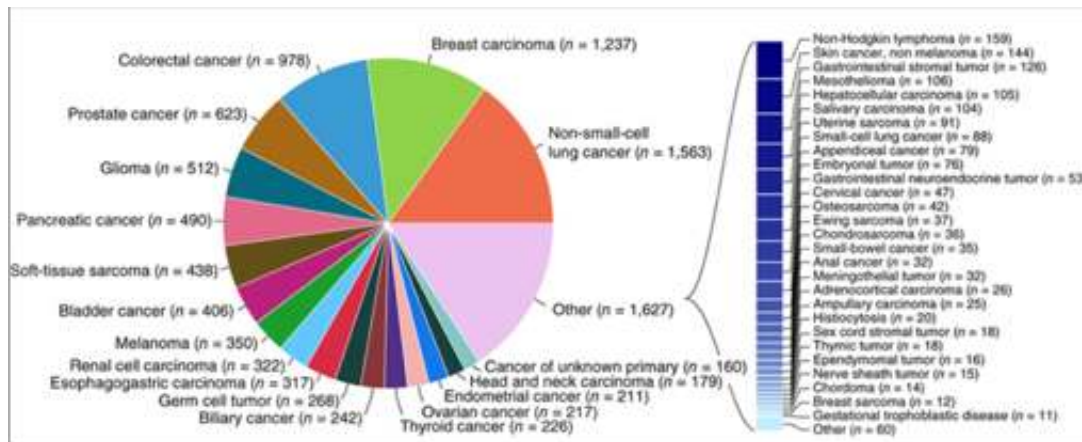
Which is more effective (left or right?)



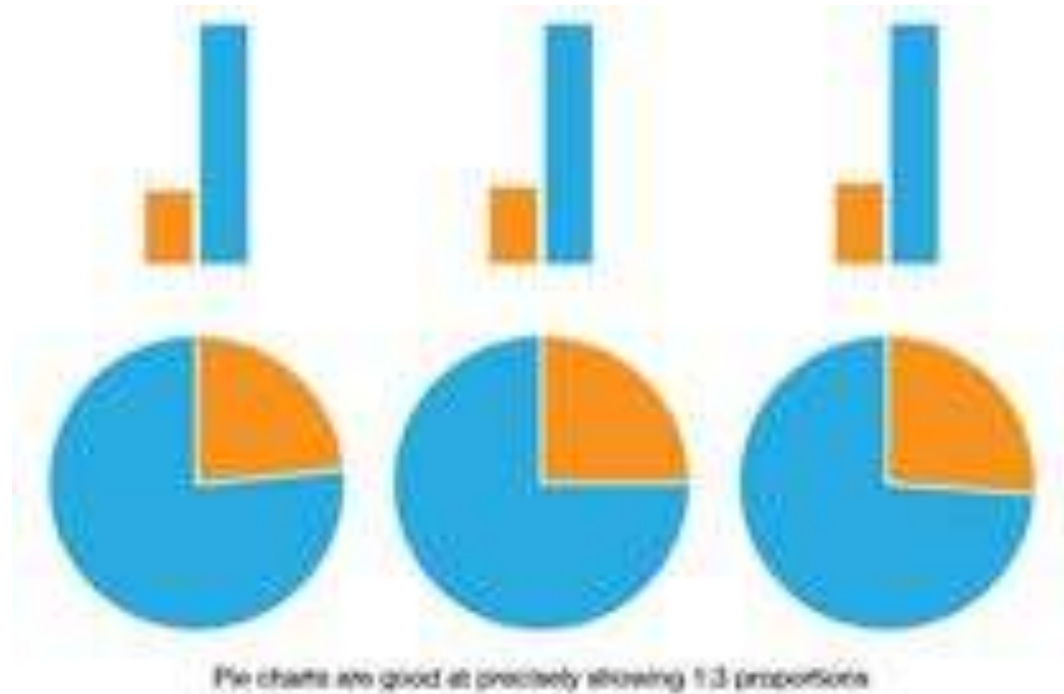
Which is more effective (left or right?)



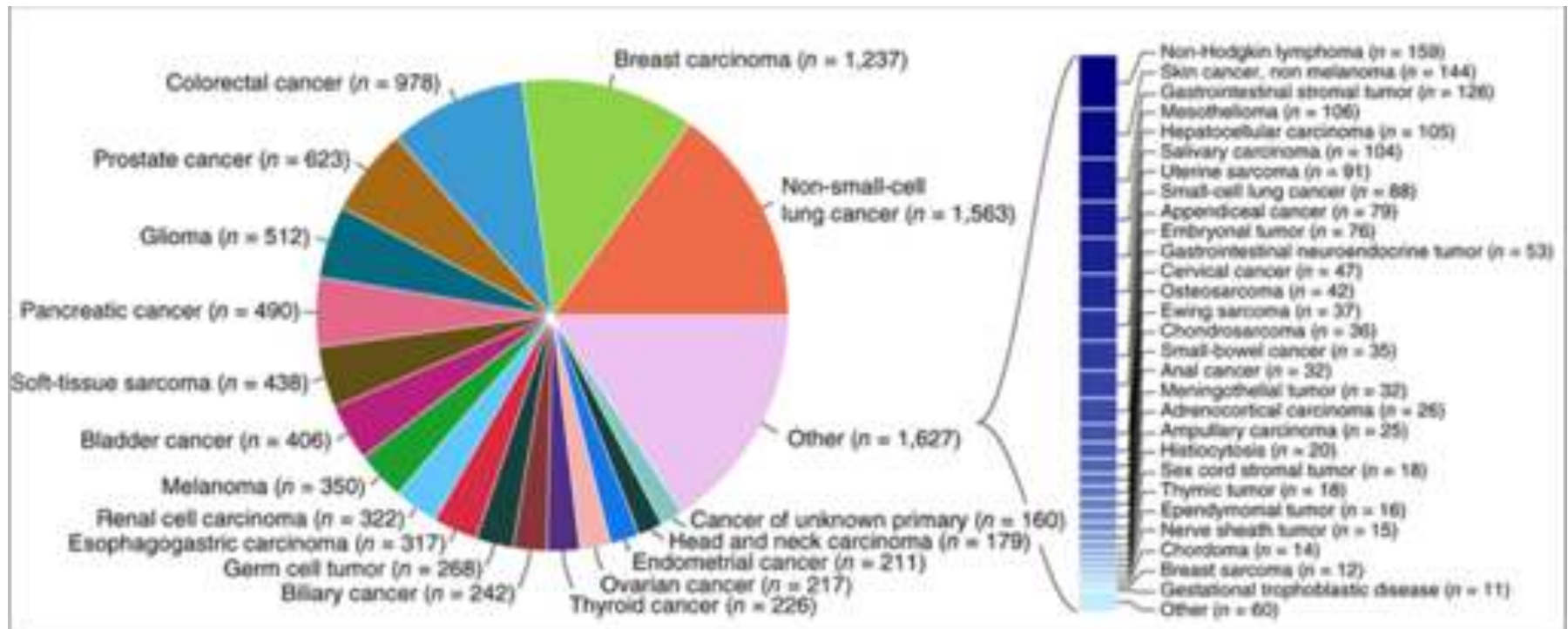
Which is more effective (left or right?)



When to use a pie chart



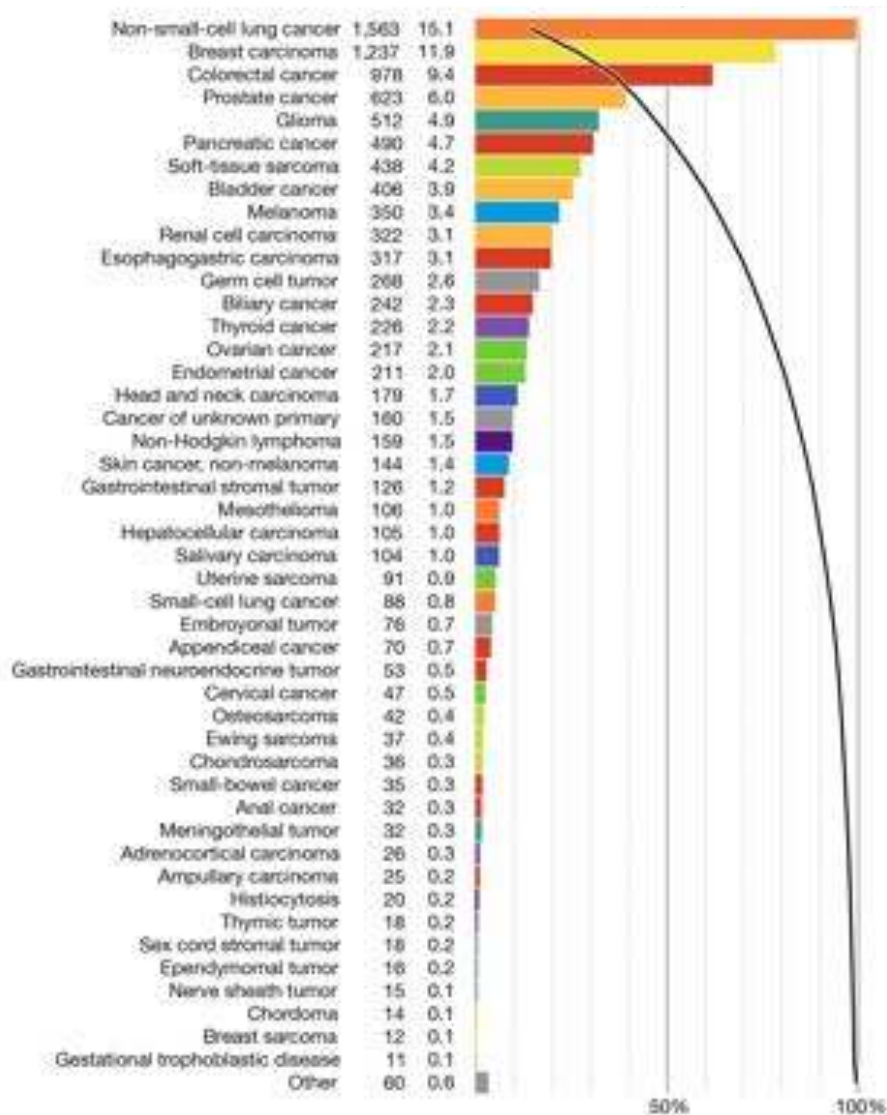
When to not use a pie chart



Ahmet Zehir et al. (2017) Nature Medicine doi:10.1038/nm.4333

- Hard to judge proportions
- Poor use of color
- Hard to read labels
- Over ½ of the categories had to be broken out of the pie chart

Same data with a redesigned approach



Selected articles on fundamentals of data viz

- [Visualizing samples with box plots](#)
- [Circos plots](#)
- [When to use \(and not use\) pie charts](#)
- Resources for choosing colors
 - <http://colorbrewer2.org/>
 - <http://mkweb.bcgsc.ca/color/>
 - [Understanding and using Color Palettes](#)
 - [Color palettes for color blindness](#)
 - [Names for >9000 colors](#)
 - Including 40 beer colors
- Credit to Martin Krzywinski for his extensive work in this area and many of the above resources

Best practices in visualization

Best practices from this workshop

- *Always* label axes
- Consider readability of font size
- Avoid vertical or angled text if possible
- Avoid unnecessary use of color, point shapes, etc.
- Chose colors wisely
- If individual data points are being plotted and have started to really pile up on top of each other consider using a density function
- Always be transparent about what data manipulation is taking place (e.g. log scale, filtering of outliers, etc.)

Best practices from the experts

Ten Simple Rules for Better Figures (Rougier et al. 2014):

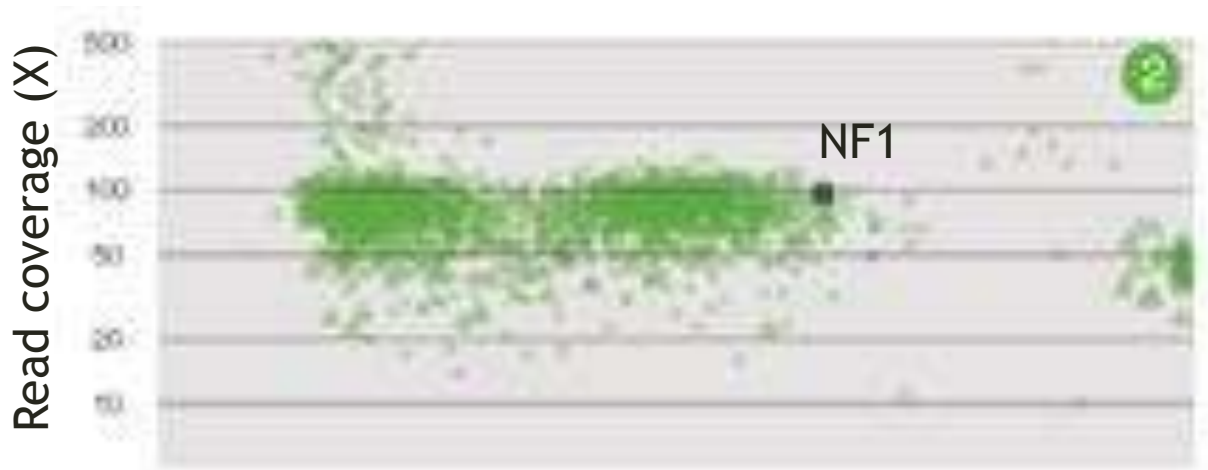
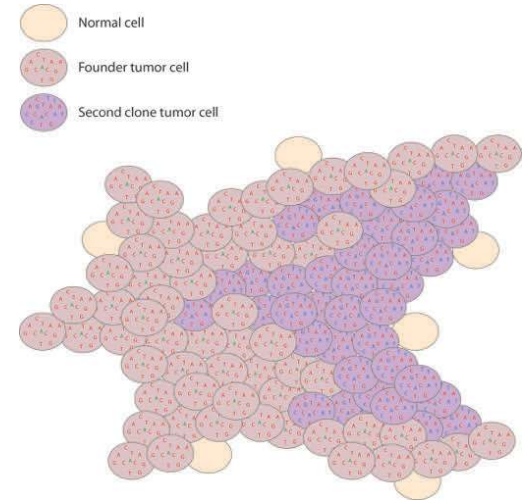
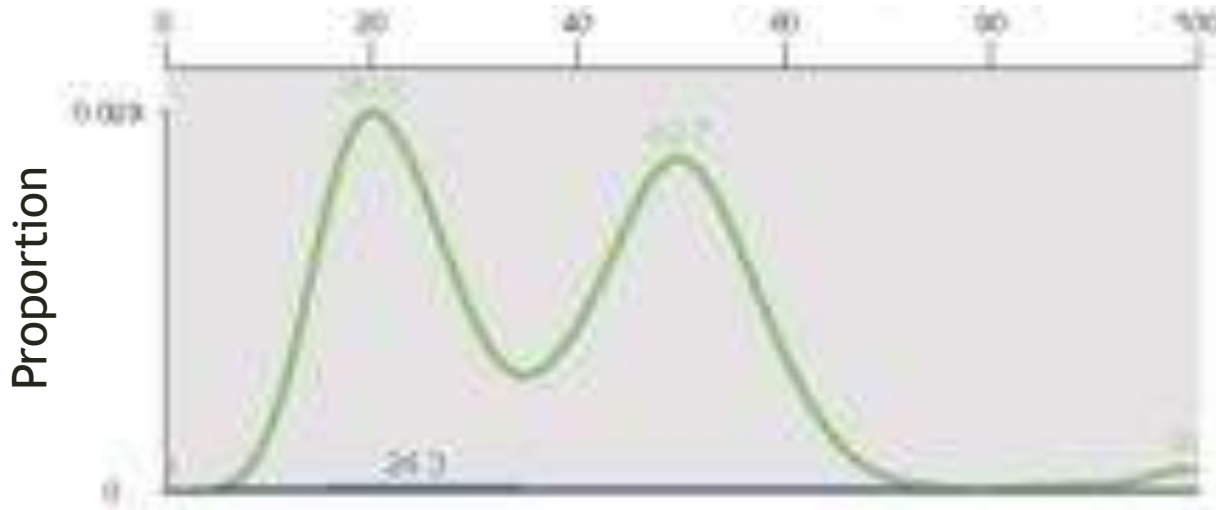
Scientific visualizations should act as a “a graphical interface between people and data”. Try to follow these rules.

1. Know Your Audience
2. Identify Your Message
3. Adapt the Figure to the Support Medium
4. Captions Are Not Optional
5. Do Not Trust the Defaults
6. Use Color Effectively
7. Do Not Mislead the Reader
8. Avoid “Chartjunk”
9. Message Trumps Beauty
10. Get the Right Tool

Example visualizations using R

Histogram and scatterplot to define cell populations

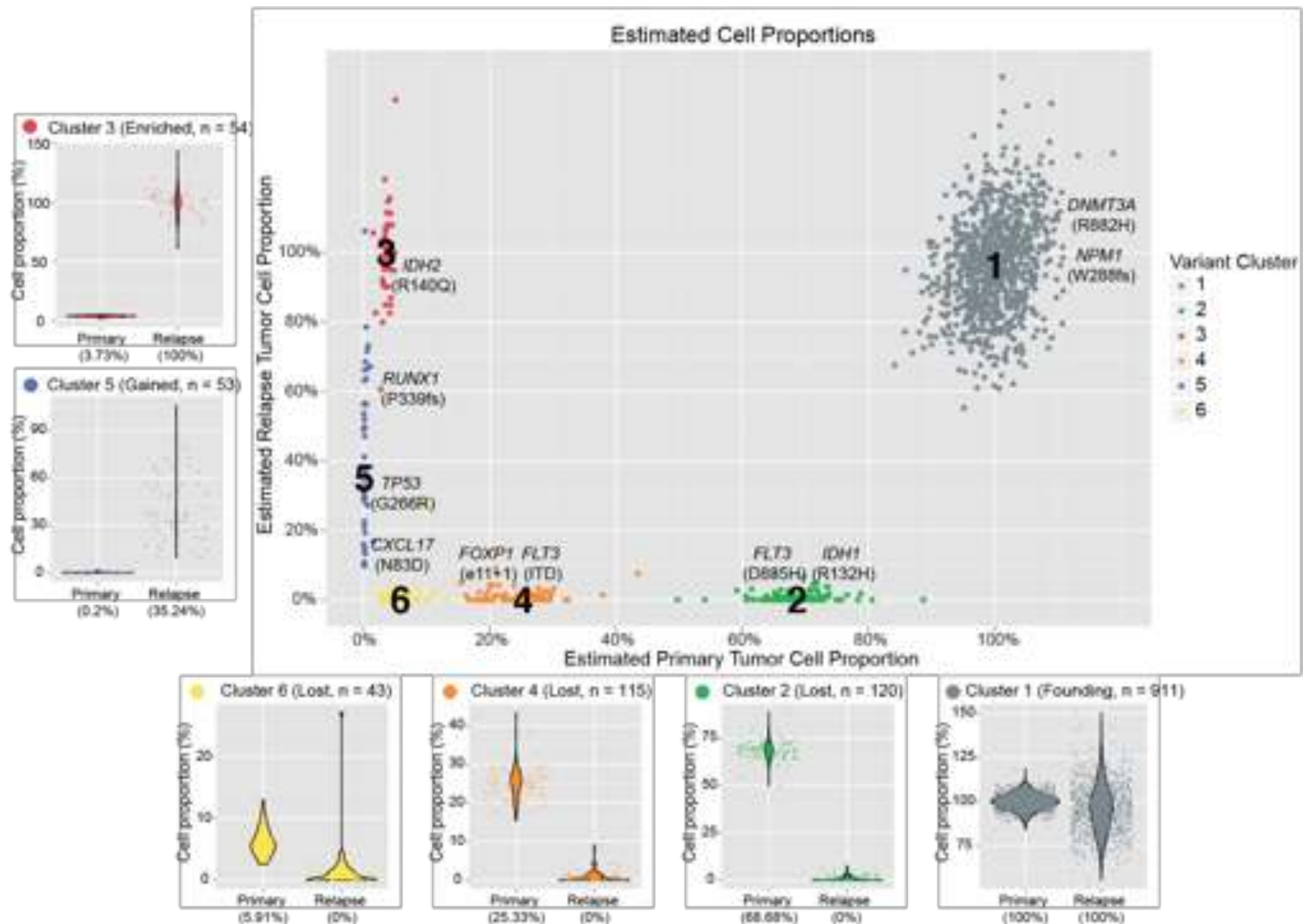
Tumor variant allele frequency



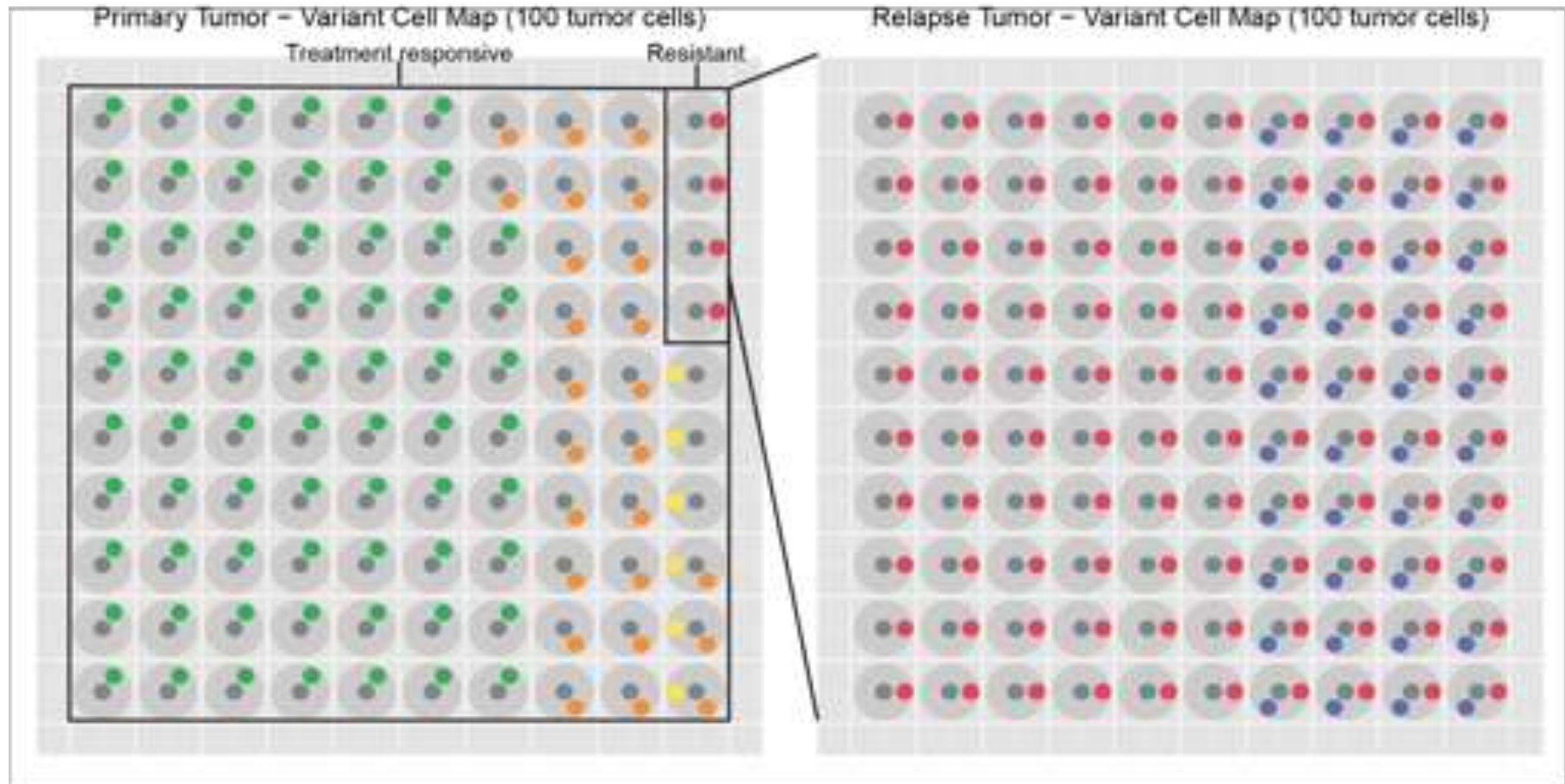
WGS data

Tumor variant allele frequency

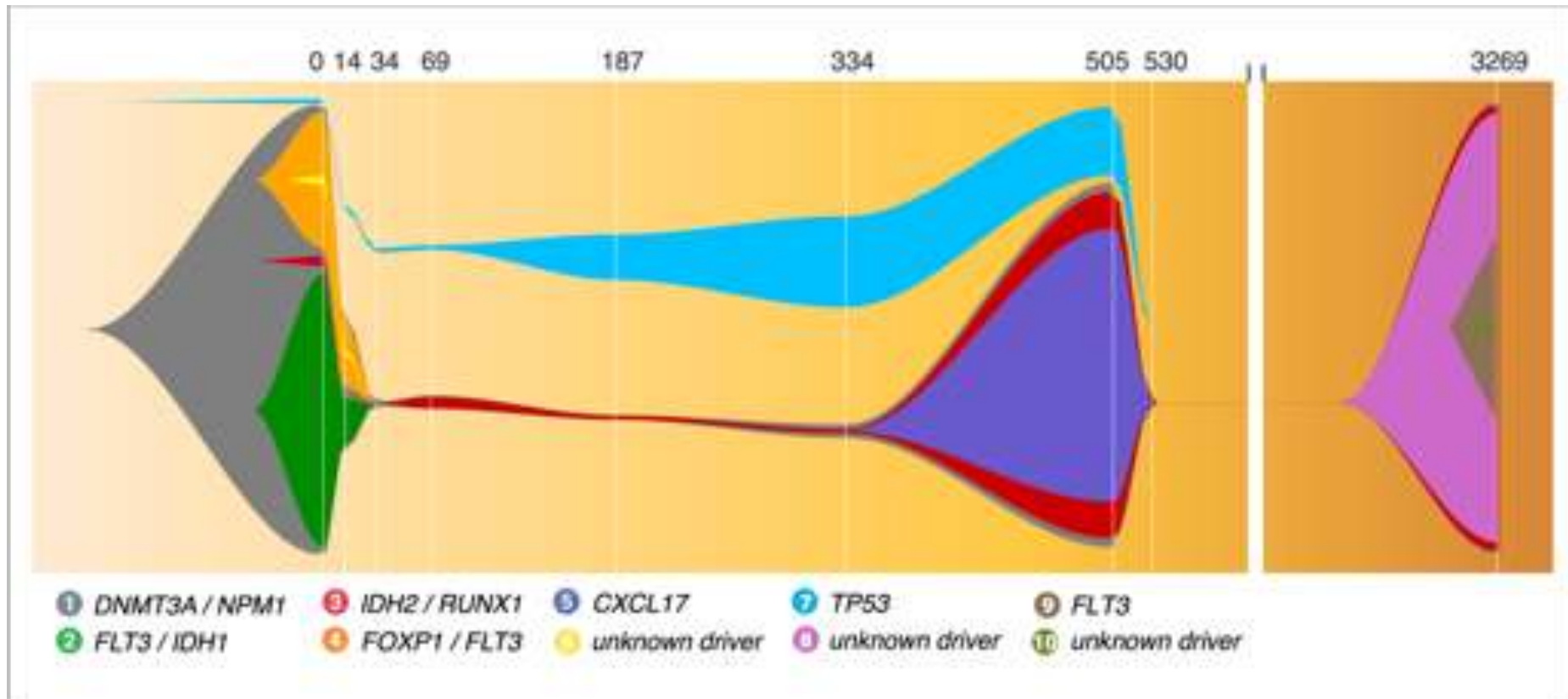
Clustering to define to distinct cell populations



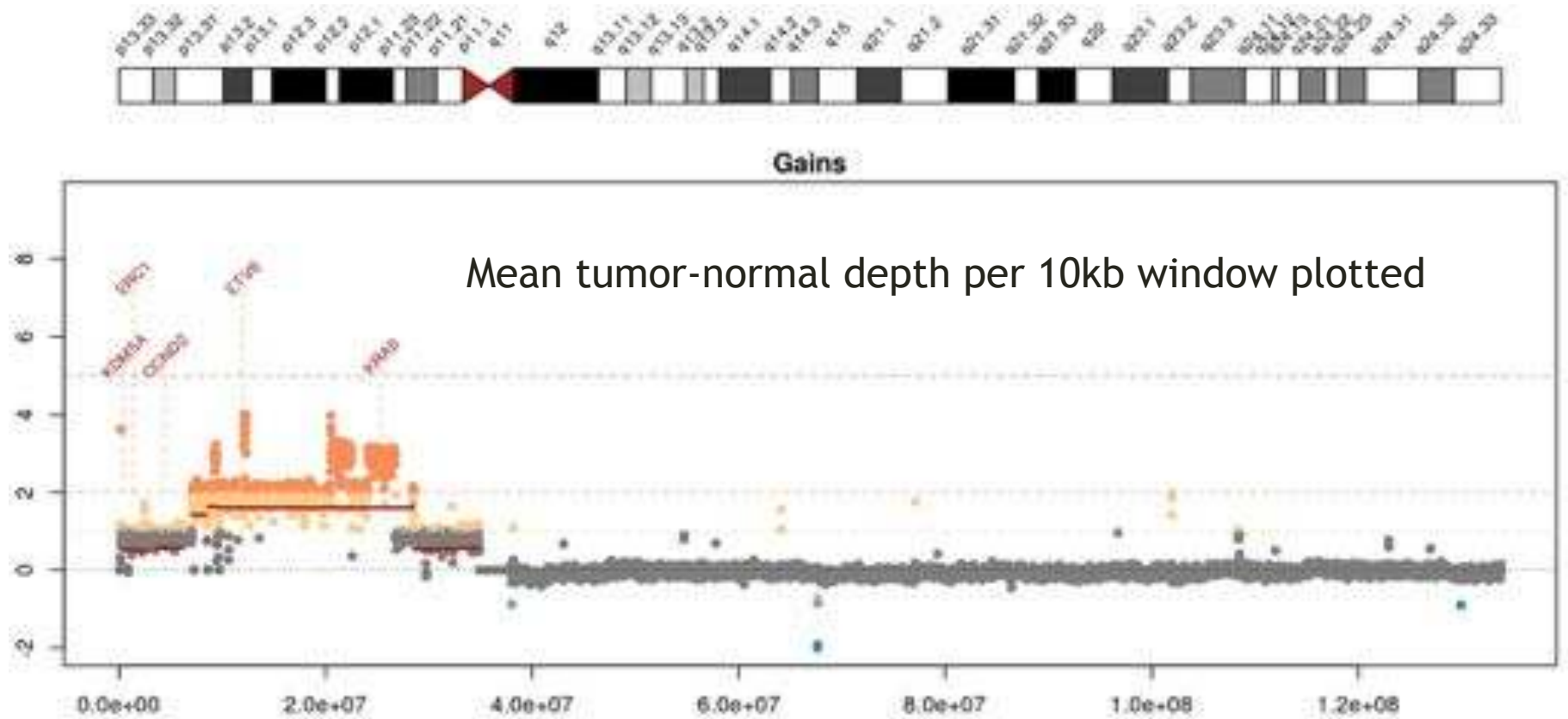
Using a “cell map” to represent the variant clusters in 100 hypothetical cells of a tissue



A 'fish' plot is used to represent sub-clones lost and gained over time

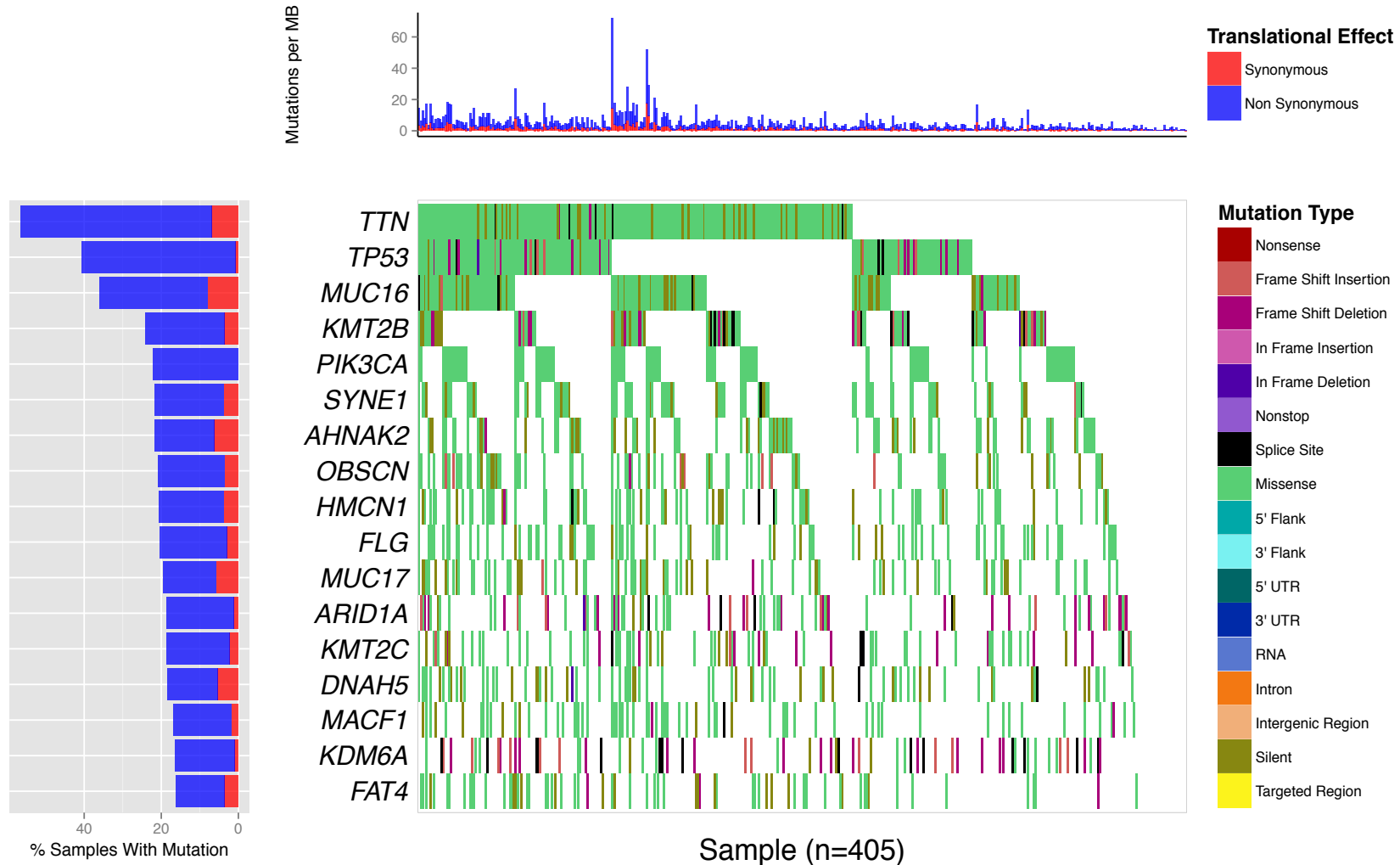


Visualizing copy number variation

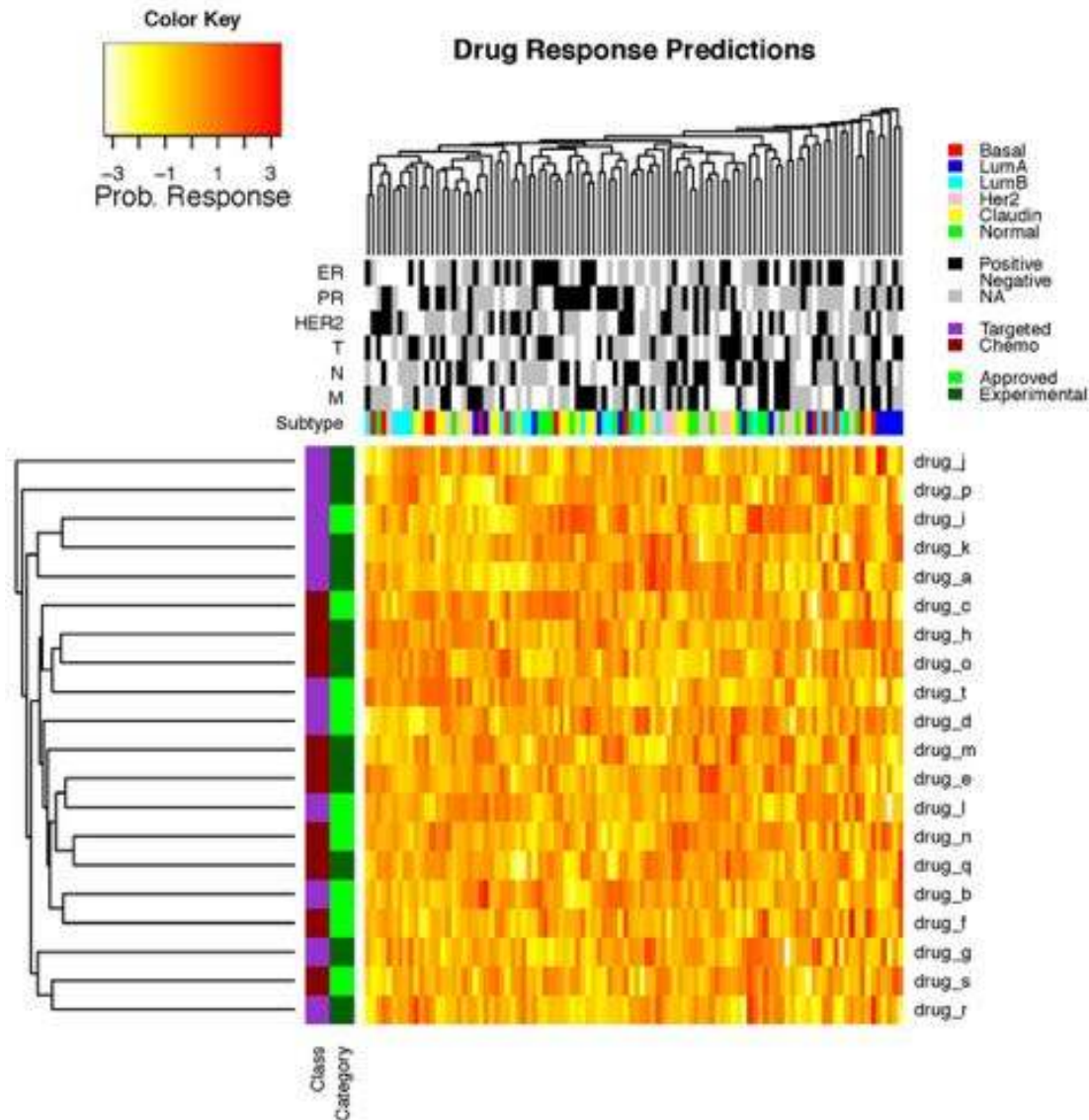


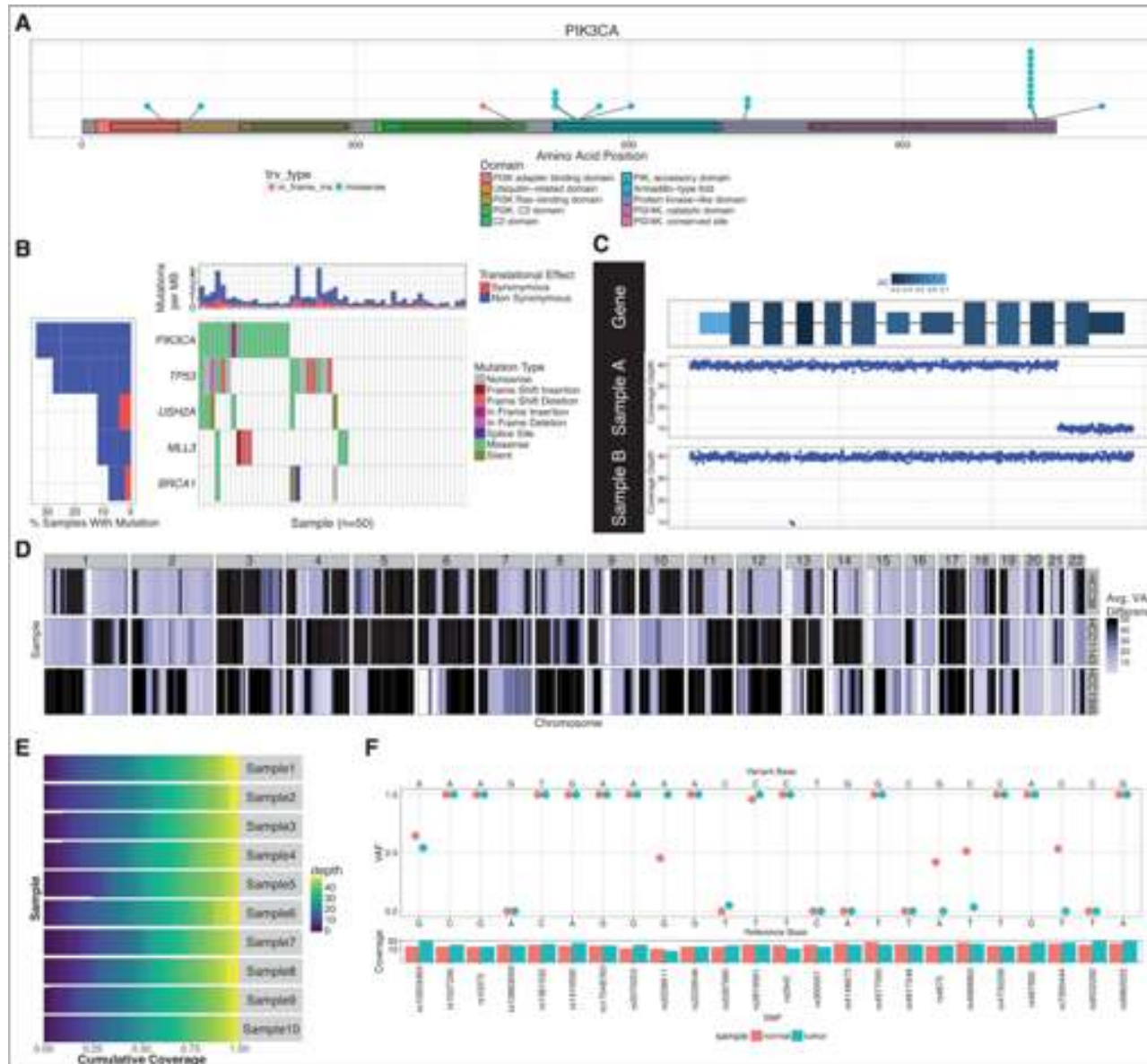
KRAS amplification in a metastatic breast cancer

A 'waterfall' plot is one way to visualize the pattern of variant recurrence in a cohort of samples



Heatmaps are a common way to simultaneously visualize multiple features of a dataset

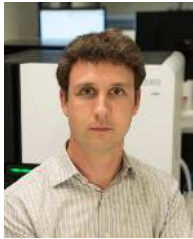




Acknowledgements: Griffith Lab group members



Malachi
Griffith



Obi
Griffith

DoCM
CIViC



Benjamin
Ainscough

CIViC
OSCC



Erica
Barnell

OSCC



Katie
Campbell

CIViC



Kaitlin
Clark

CIViC/DGIdb



Adam
Coffman

DGIdb/Regulome



Kelsy
Cotto

CIViC/Regulo
me



Arpad
Danos

Regtools/pVA
C-seq



Yang-Yang
Feng

Lymphoma



Felicia
Gomez

pVAC-seq



Jasreet
Hundal

pVAC-
seq/CIViC



Susanna
Kiwala

CIViC/Lymphom
a



Kilanin
Krysiak

CIViC



Lynzey
Kujan

HCC



Jason
Kunisaki

CIViC



Josh
McMichael

CIViC
Lymphoma



Cody
Ramirez

genVisR
HCC/SCLC



Zachary
Skidmore

CIViC/Komen



Nick
Spies

HCC/SCLC



Lee
Trani

DGIdb
SCLC



Alex
Wagner

Director
Informatics



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 Blieri, Thomas D. Biewald, Andrew P. Bohannon, Kirk A. Brege, Rose L. Brockhouse, Anthony M. Brummett, Mark M. Burnett, Christopher R. Cobanski, Theresa A. Caesar, Matthew B. Callaway, Francine R. Camacho, Edgar A. Campbell, Darren Casimere, Wan-Ching Chan, Lei Chen, Mari Jo Clark, Adam C. Colfman, Lisa L. Cook, Matthew Cordes, Laura P. Courtney, Amy K. D'Albora, John S. Daibora, Indranisel 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, Efreem 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. Futhy, Xin Gao, Barbara Gillam, Chakravarthy Girdla, Jennifer Godfrey, Jenna Christine Goeckner, Laura Jane Gottschalk, Susan D. Grasso, Malachi Griffith, Obi Griffith, Jeanne M. Grigsby, David S. Guder-muth, 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, Wendell 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. Lakanoni, David E. Larson, Sai Ham 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. MacMil-lan, Vincent J. Magrini, Lenon O. Maguire, Christopher A. Maher, Jennifer Maher, Rony J. 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. Oberkfell, David J. O'Brien, Kerri R. Ochoa, Gretchen A. O'Donnell, Michelle D. O'Laughlin, Philip J. Ozersky, Stephanie J. Par-rish, Josh Peck, Kymberlie A. Pepin, Stephanie K. Pleasant, Craig S. Pohl, Eric Ponce, Allison Ann Regier, Amy D. Reilly, 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'Dasha T'vorle Scott, Harry E. Senaldi, Cher-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, Evanne Trevaskis, Richard Trippier, Rusudan B. Turabelidze, Rahul Tyagi, Joelle M. Veizer, Tammi L. Vick-ery, 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

Susan Dutcher
Tim Ley

Bob Fulton
Lucinda Fulton
Ira Hall
Dave Larson
Chris Maher
Lukas Wartman

Matt Callaway
Chris Miller
Jason Walker

Introduction to GenViz course site

www.genviz.org