



The Elizabeth H.  
and James S. McDonnell III

**McDONNELL  
GENOME INSTITUTE**  
at Washington University

# Introduction to Cancer Genomics and Precision Medicine

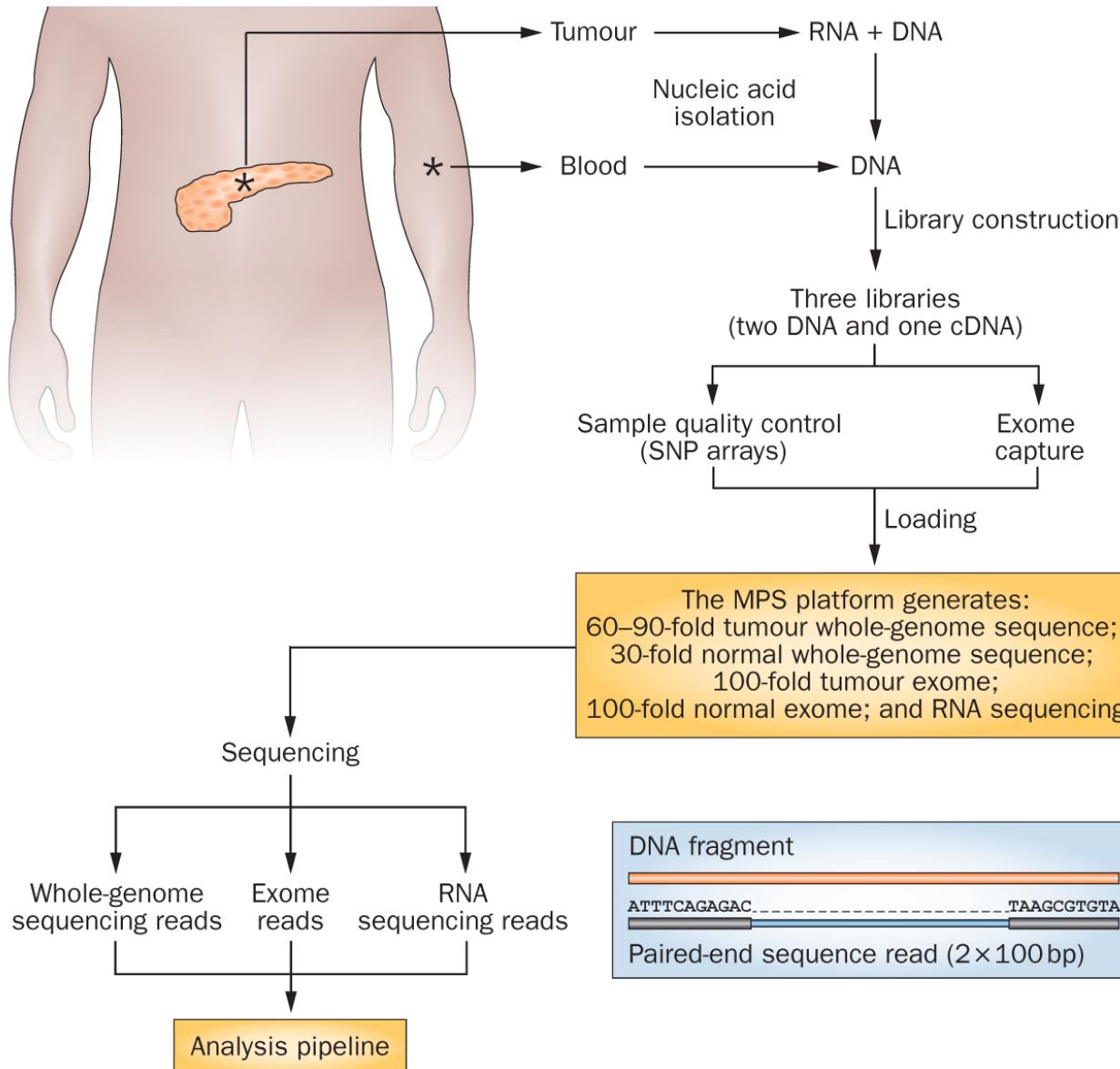
Obi L. Griffith, PhD  
Assistant Professor of Medicine  
Assistant Director, McDonnell Genome Institute

HLA Course  
22 November 2016

# Outline

- Introduction to Cancer Genomics
  - Sequencing approaches
  - Alteration types
  - Progress to date
- Introduction to Precision Medicine
  - Feasibility studies
    - Gene panels
    - Whole exome sequencing
    - Comprehensive approaches
- Key challenges
  - Visualization
  - Interpretation

# Cancer genomics research has exploded with the rapid advances in DNA sequencing technologies



# How does it work? Short read alignments are the fundamental currency of cancer genome analysis



- Alignment is about fitting individual pieces (reads) into the correct part of the puzzle
- The human genome project gave us the picture on the box cover (the reference genome)
- Imperfections in how the pieces fit can indicate damage to a copy of the picture

Reference: AGCCTGAGACCGTAAAAAA**A**GTCAAG

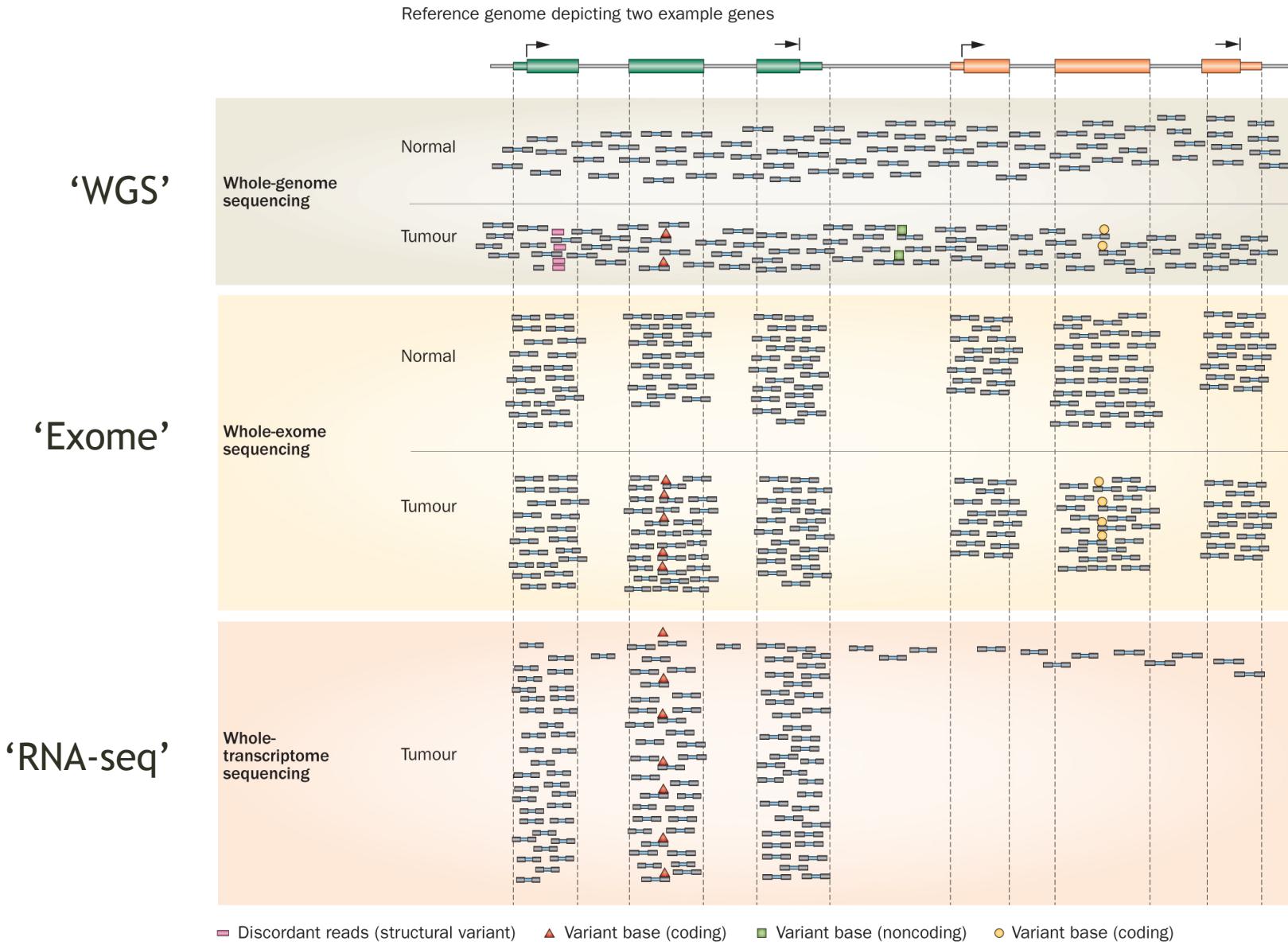
||||||||||||||||||||||

A read sequence: GAGACCGTAAAAAA**C**GTC

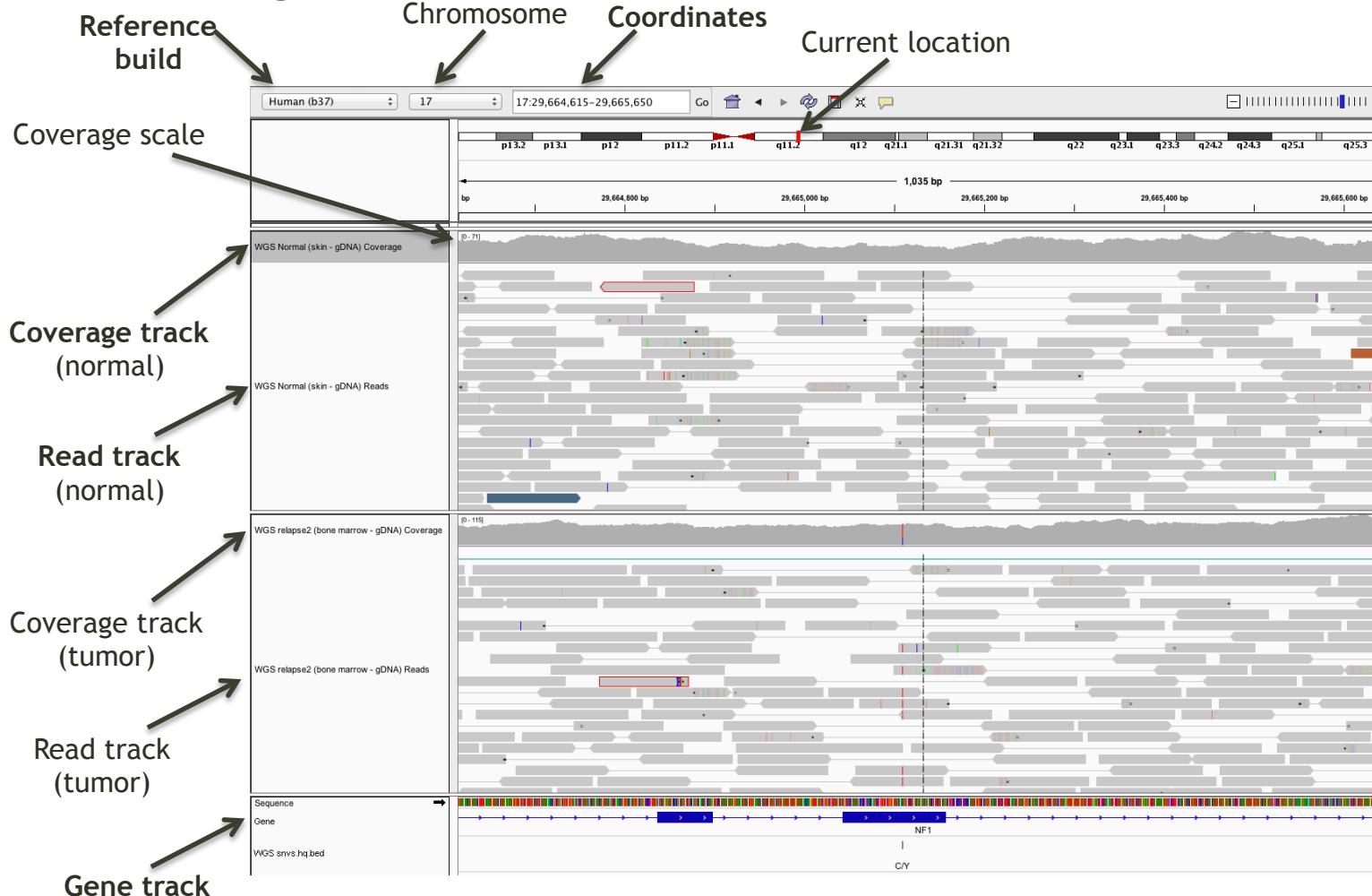


A variant

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



# The Integrative Genomics Viewer showing individual read alignments



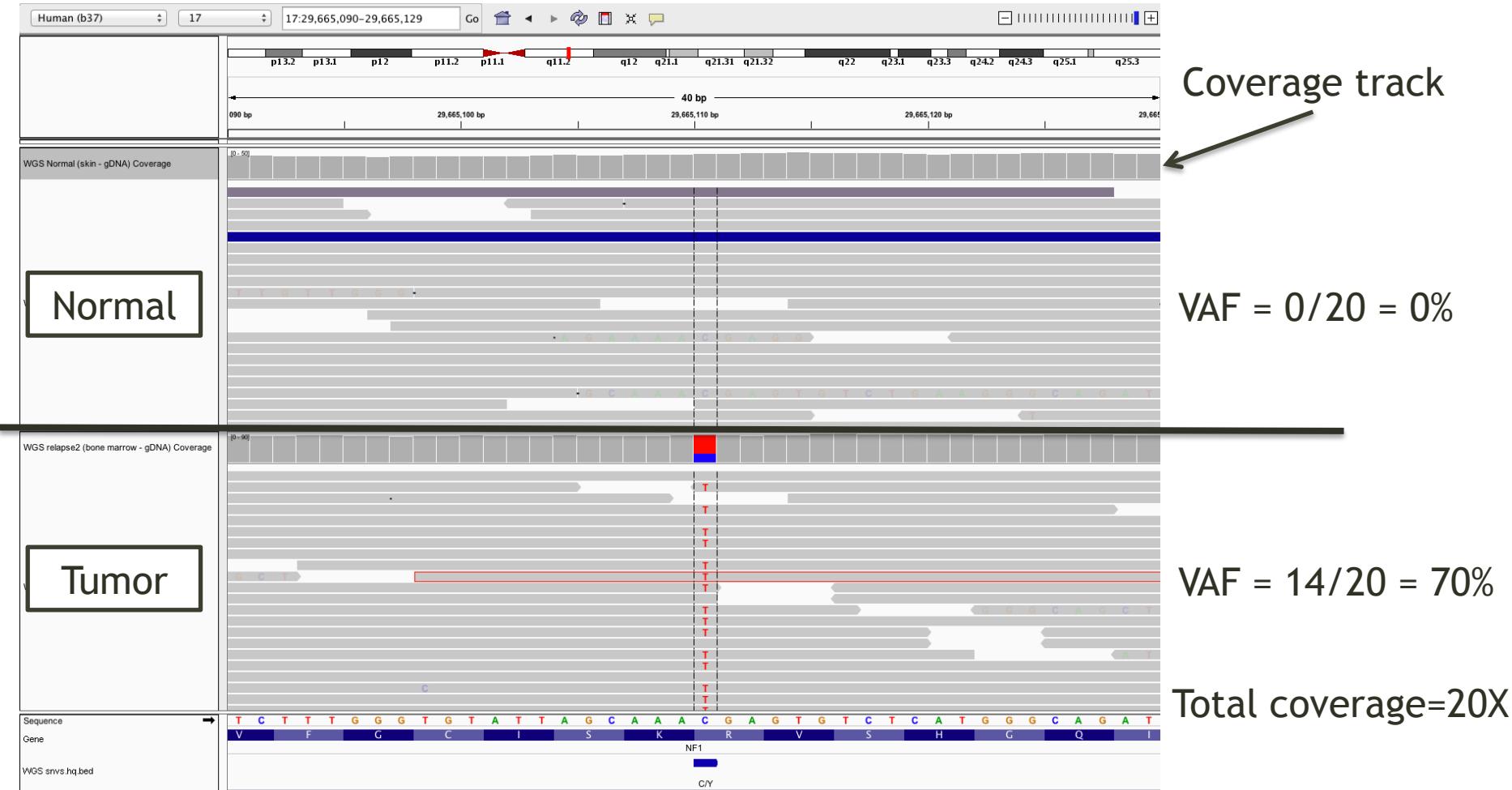
- Each grey bar represents a single paired Illumina read
- Grey indicates sequence match to the reference sequence
- Color indicates variation from the reference sequence

# SNVs and related terminology

- SNV - Single Nucleotide Variant. Single base substitution relative to the reference genome or normal genome
  - Missense, nonsense, etc.
- Tier 1-4
  - Every base is divided into 4 mutually exclusive categories
  - ‘Tier1’ - coding regions of protein and RNA genes
  - ‘Tier2’ - regulatory regions
  - ‘Tier3’ - everything not in one of the other tiers
  - ‘Tier4’ - repetitive regions
- VAF - variant allele frequency (next slide)
- BAM files and BAM read counts
  - BAM file - Binary SAM (sequence alignment/map) file. Stores the alignments of reads to the reference genome for a single sample (e.g. tumor)
  - BAM read counts - summary of bases supported by a BAM file at a single position
    - Chr17:29,665,110 -> A=0, C=13, G=0, T=14

# Variant allele fraction (VAF) and coverage

VAF = Variant reads / Total reads



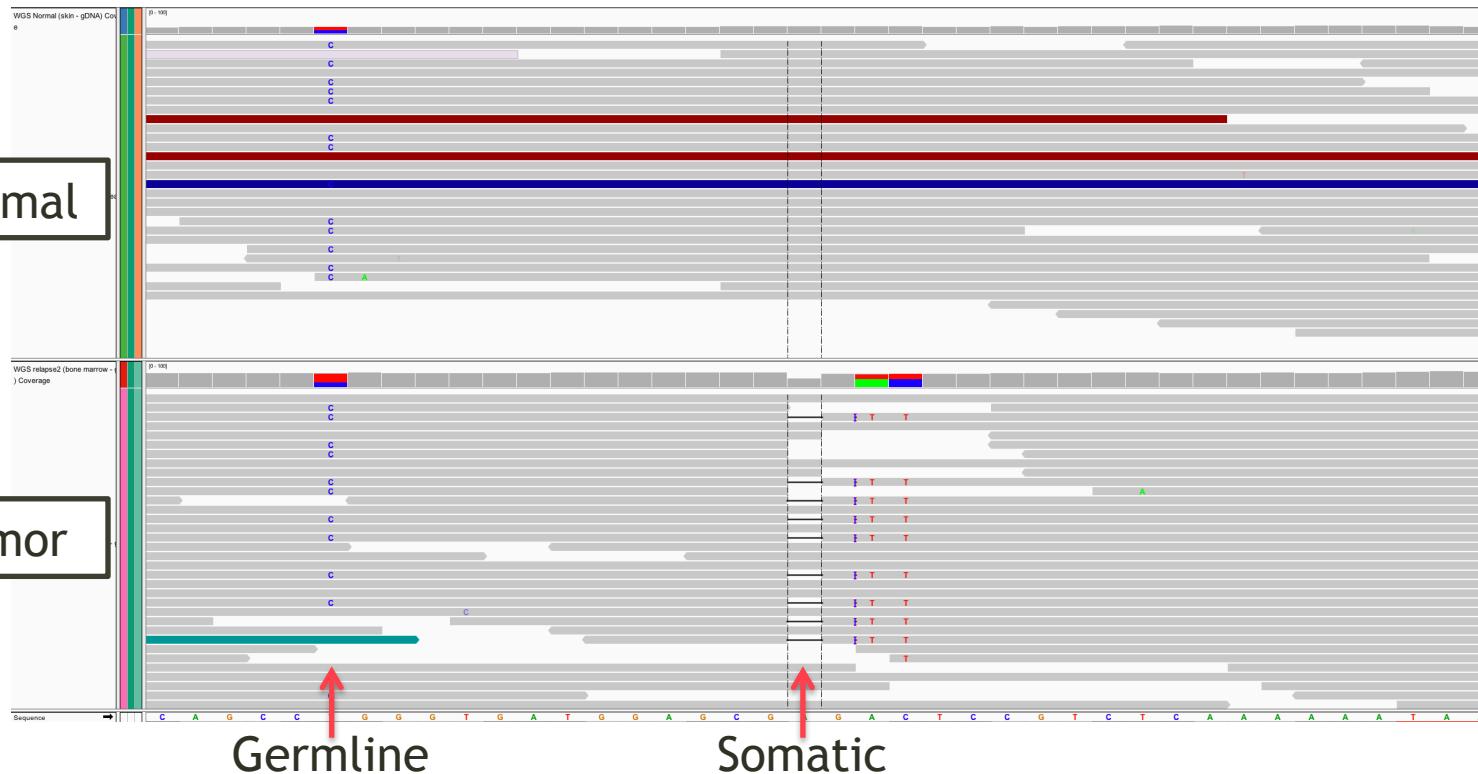
Recall that we have two copies of our genome ( $2n$ ). Mutations are typically in one copy. Therefore, a heterozygous variant (one copy) is expected to have  $VAF = 50\%$ . Often not true due to sample purity, tumor heterogeneity, sampling error, alignment issues, copy number variation, etc.

# Single nucleotide variants (SNVs) and insertions/deletions (indels) appear as short alignment discrepancies from reference genome

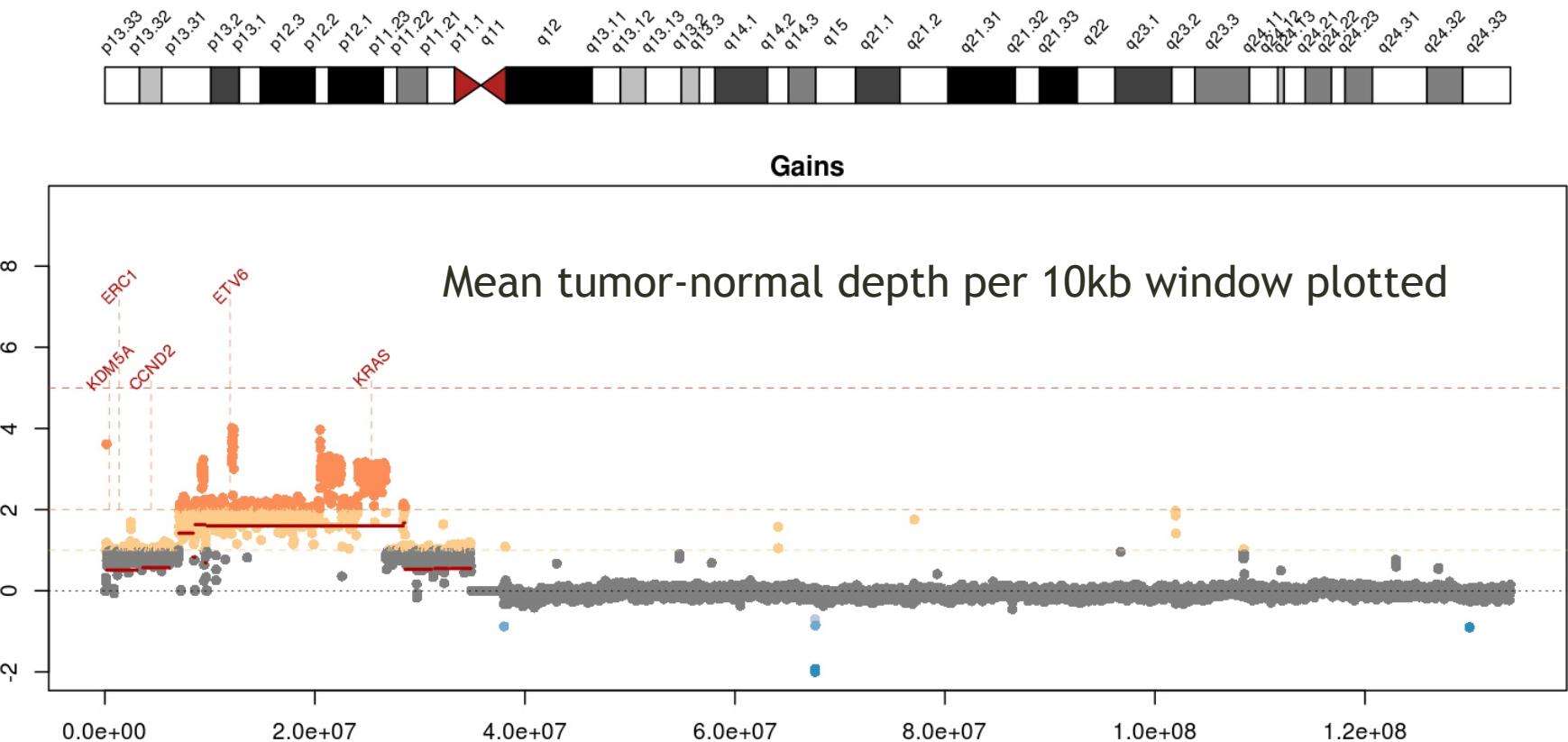


# Both somatic and germline mutations are important

- Germline mutations
  - Present in egg or sperm
    - All cells of affected offspring
  - Heritable
  - Cause of familial cancers
- Somatic mutations
  - Occur in non-germline tissues
    - Only tumor cells (breast, lung, blood, etc)
  - Non-heritable
  - Cause of sporadic cancers



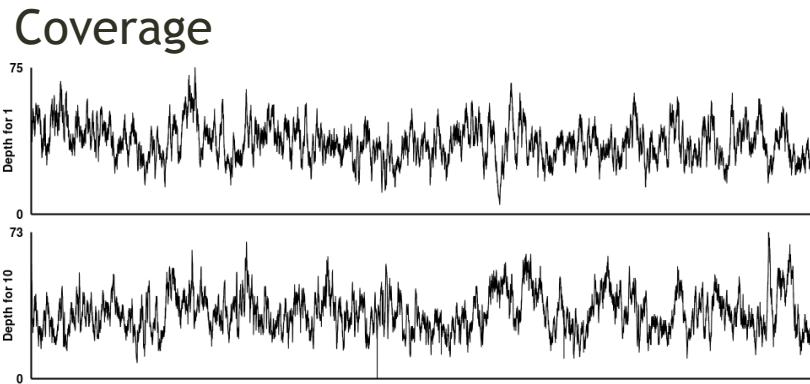
# Copy number variants (CNVs) appear as deviations from in alignment “depth” or “coverage”



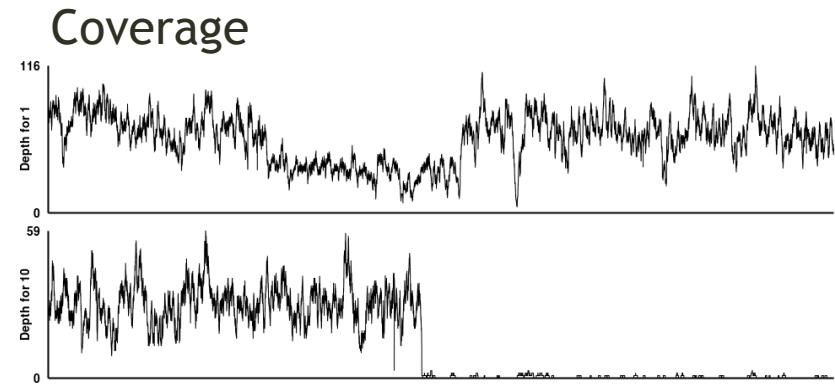
KRAS amplification in a metastatic breast cancer

# Structural variants (SVs) can be identified using a combination of coverage and discordant read alignments

Normal



Tumor



Chr10

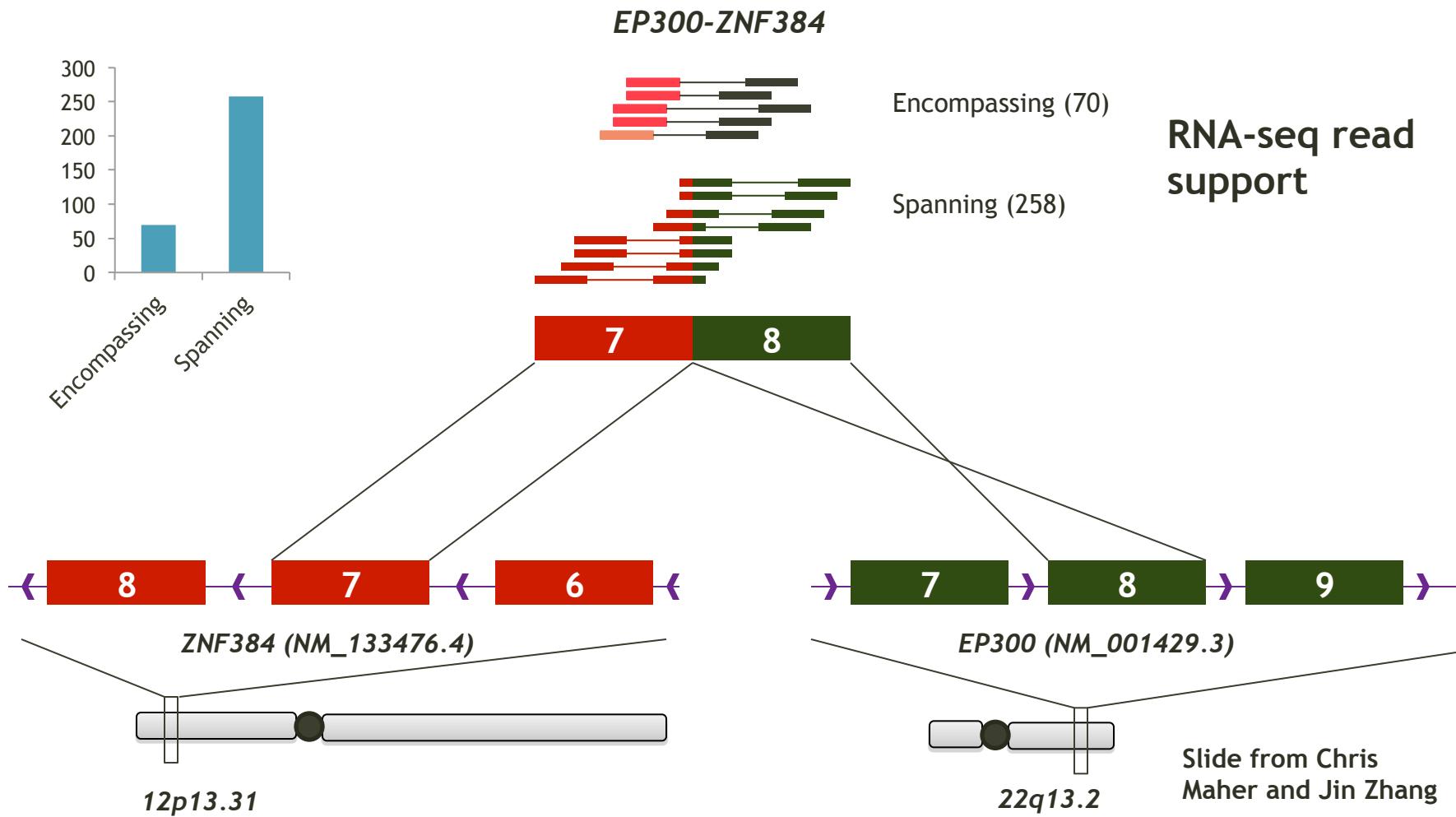
10:104273357 10:104314357 10:104273357 10:104314357

Discordant read support

A Chr1-Chr10 (TBX19-SUFU) unbalanced translocation identified in an adult acute lymphocytic leukemia.

Discordant read support

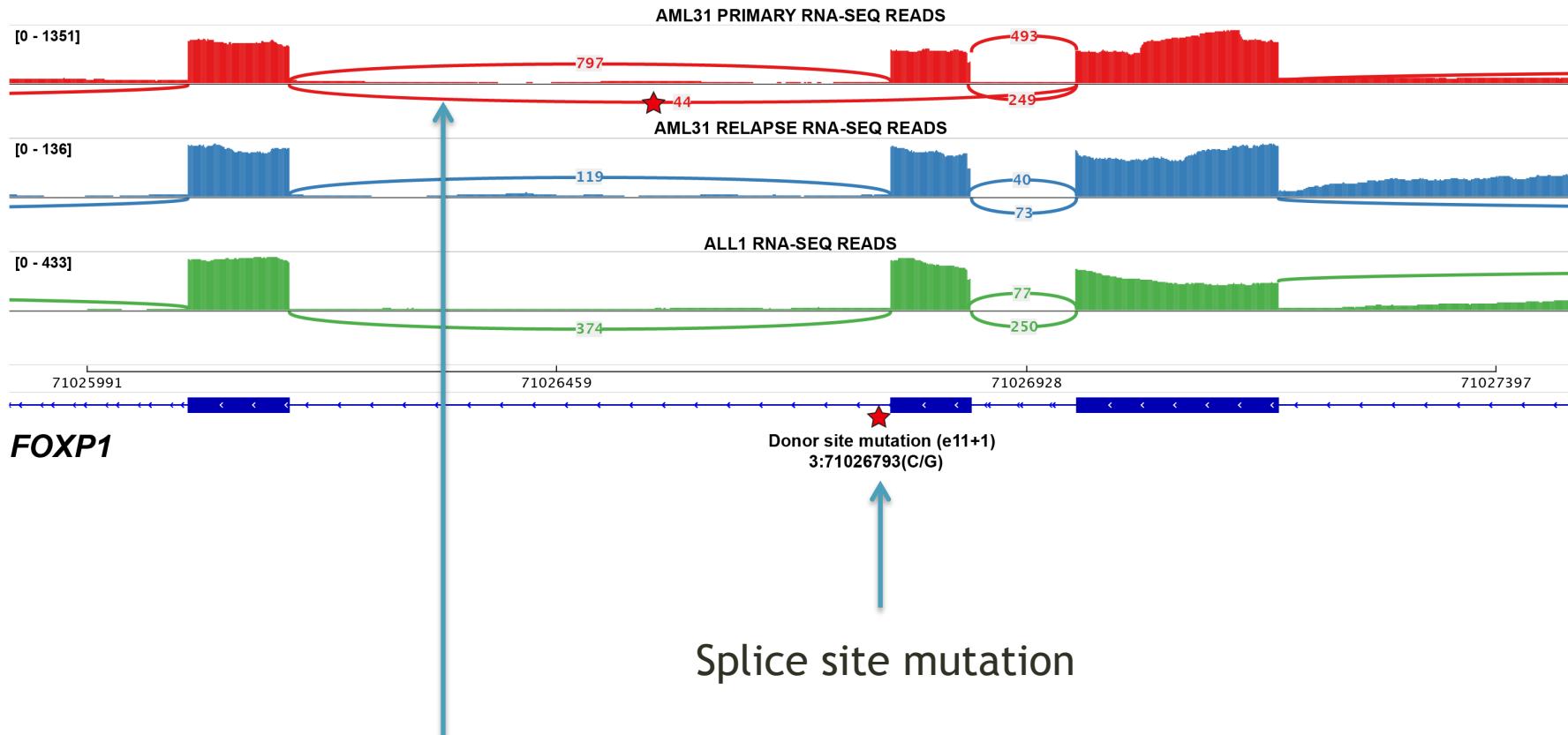
# Expressed gene fusions can be identified by discordant read alignments spanning known exons from RNA-seq data



Exons 1-8 of EP300 fused to exons 7-10 of ZNF384 in head-to-tail fashion.

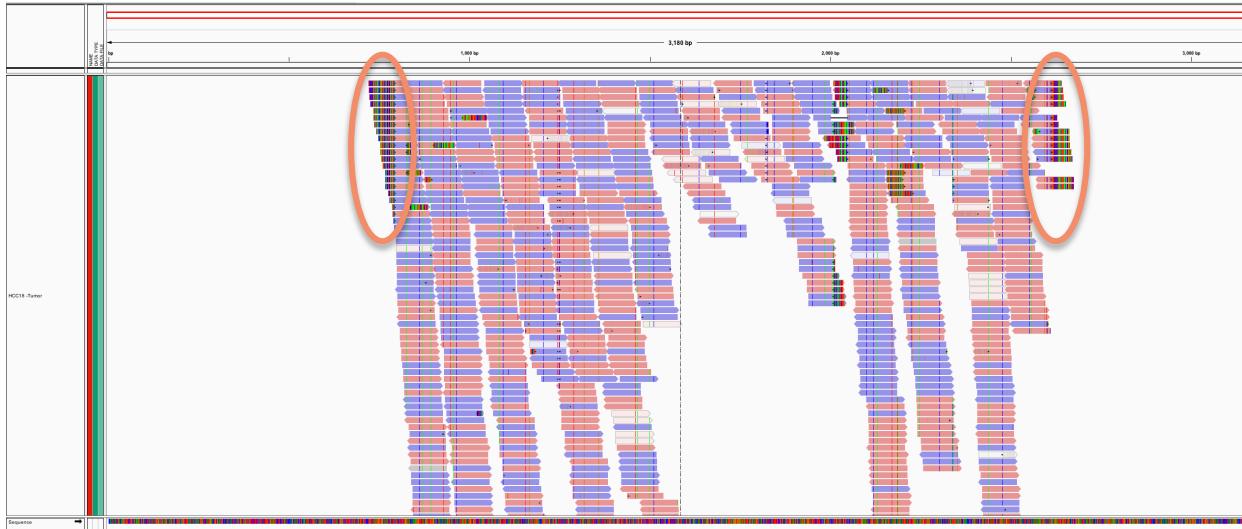
# RNA-seq can also reveal the splicing consequence of somatic mutations detected in WGS

## A. Sub-clonal somatic splicing event in *FOXP1* observed in the primary tumor but cleared in relapse



Read alignments support the skipping of affected exon

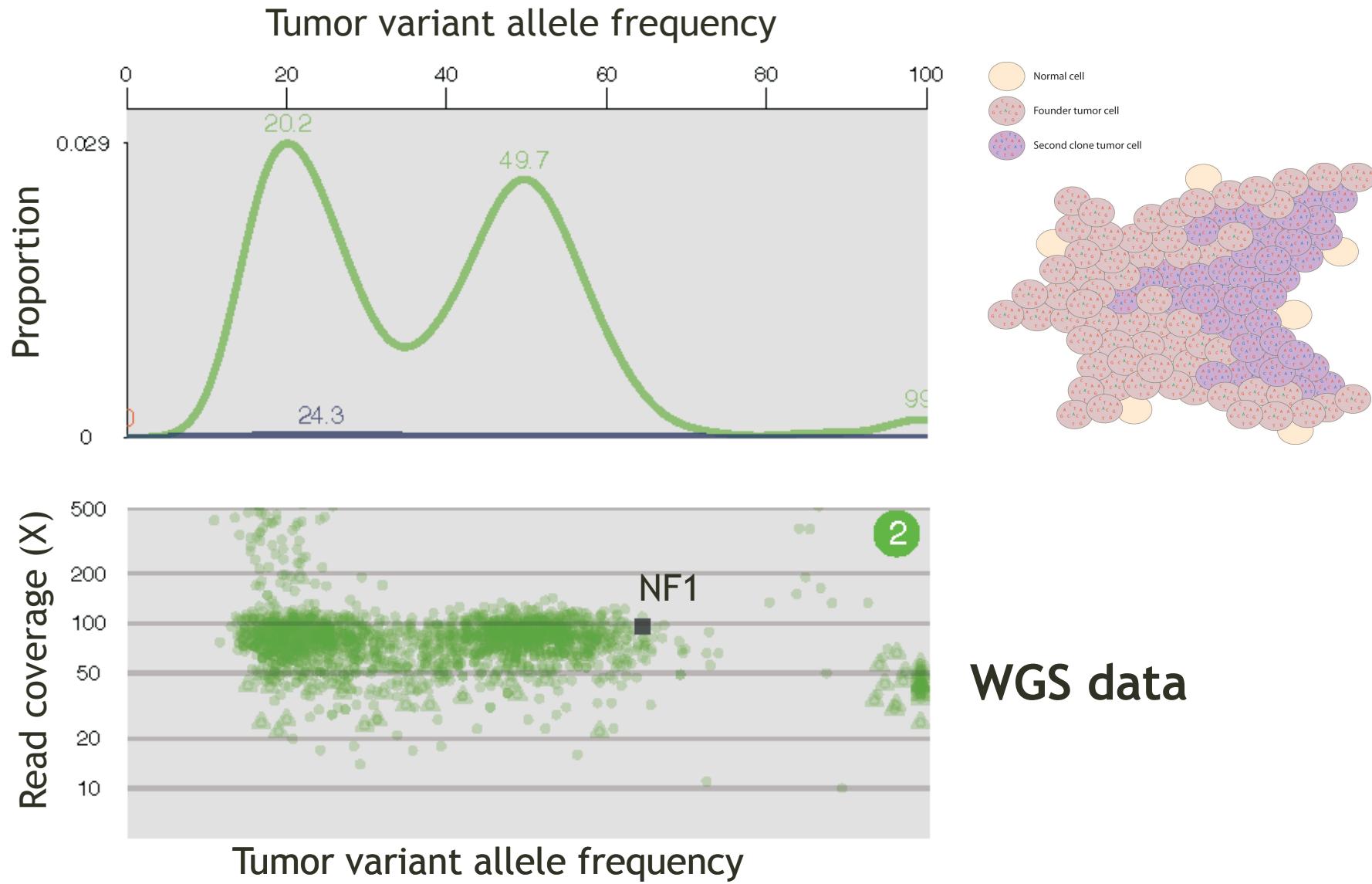
# Viral expression and integrations can be identified by competitive alignment with human and viral genome databases



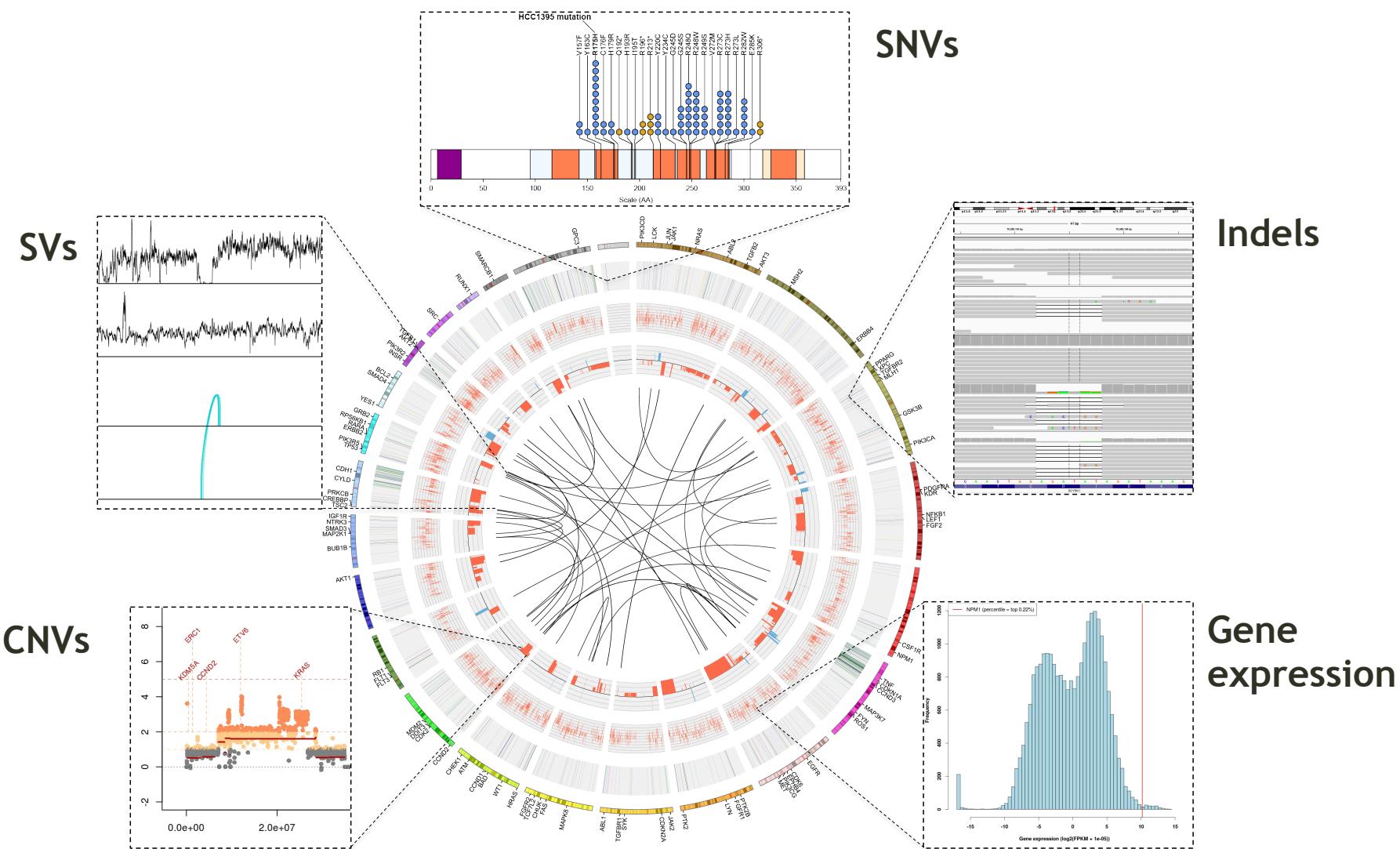
- Many reads align directly to viral genome (Hepatitis B)
- Soft-clipped reads in human reference help identify integration site

Example from HCC case thought to be virus negative

# Tumor clonal architecture can be inferred from distributions of variant allele frequencies

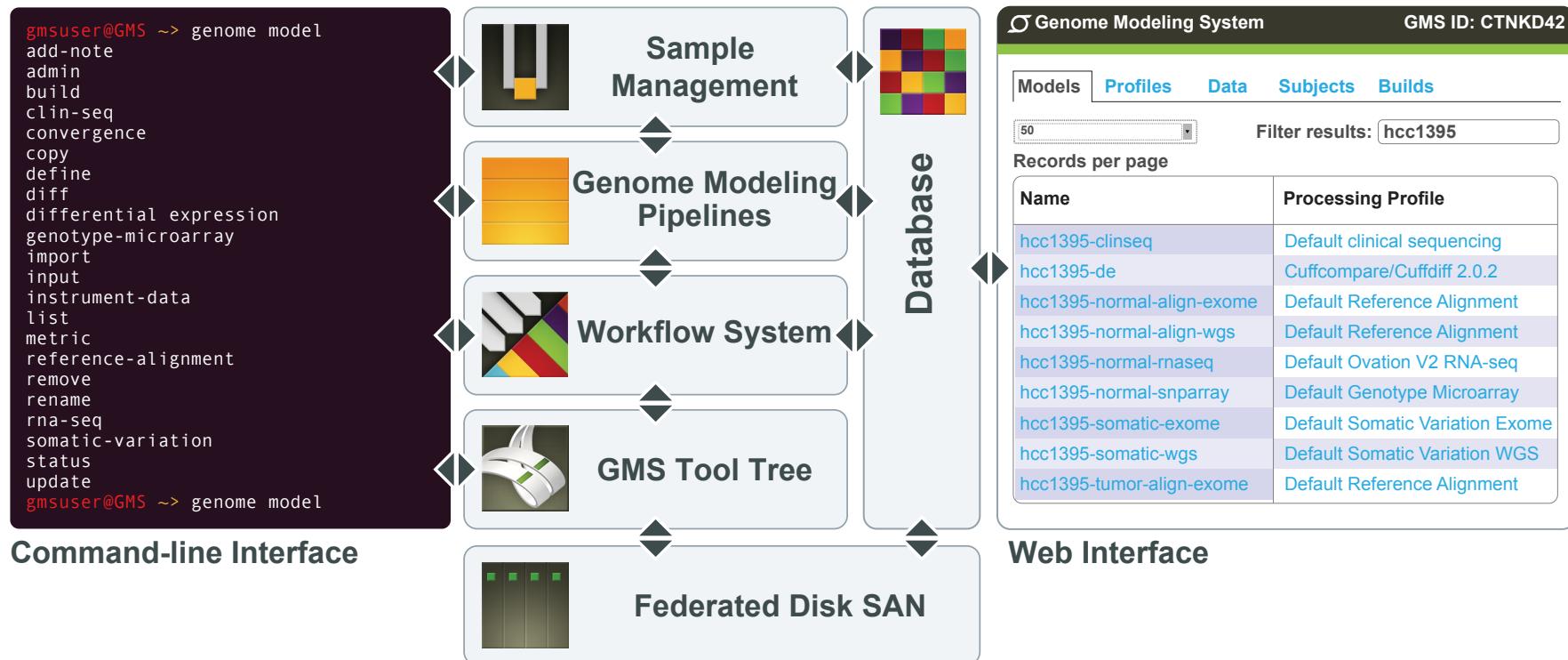


Tumor genome analysis will typically reveal dozens to thousands of alterations of multiple types



# Comprehensive and integrated sequence analysis methods are non-trivial - Genome Modeling System

An Analysis Information Management System (AIMS) that manages and tracks all analysis

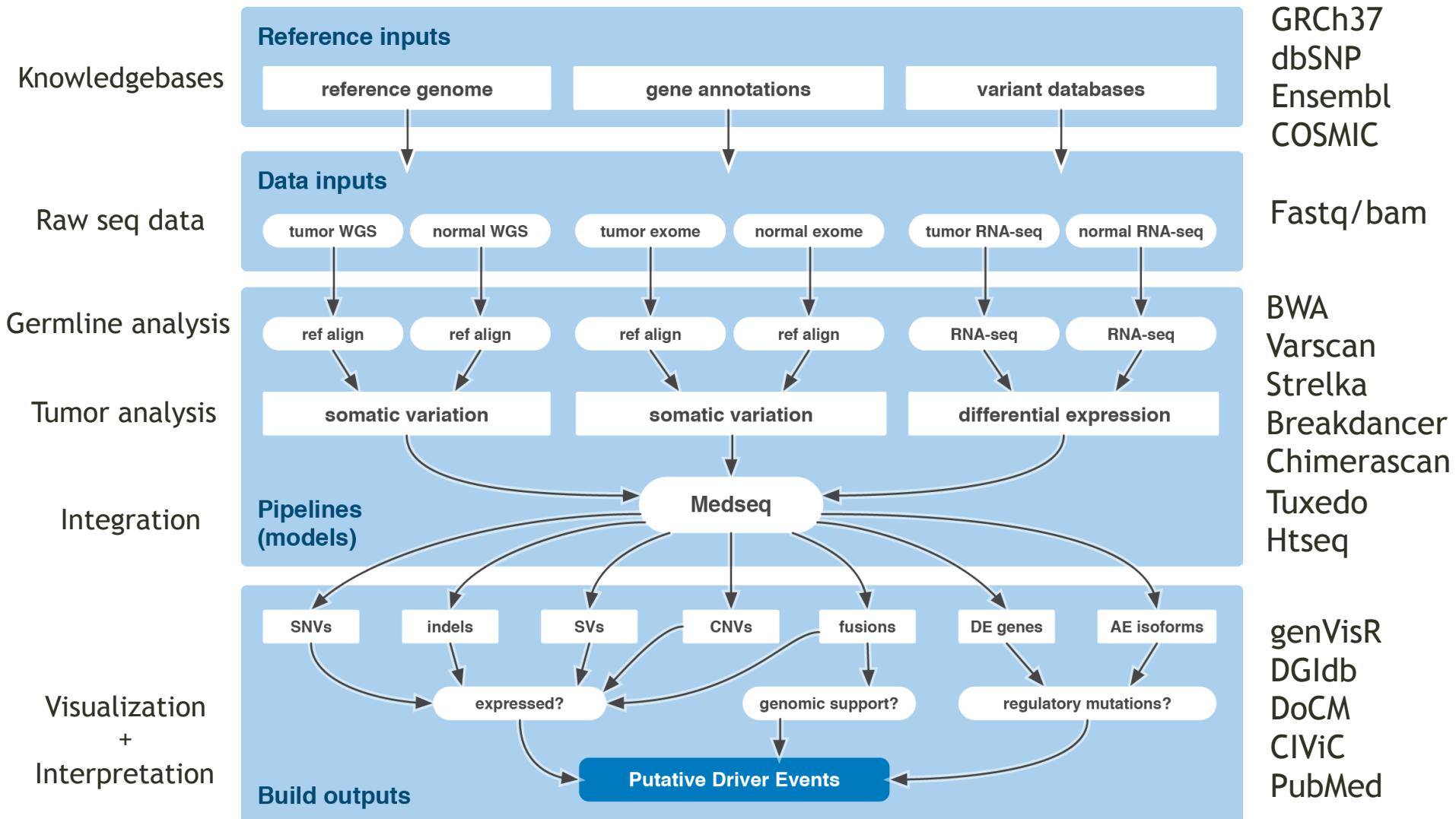


- Key Features:
  - Automates most routine analysis
  - Manages cluster and disk usage
  - Provides history of all analysis and exact parameters
  - Allows short-cutting!

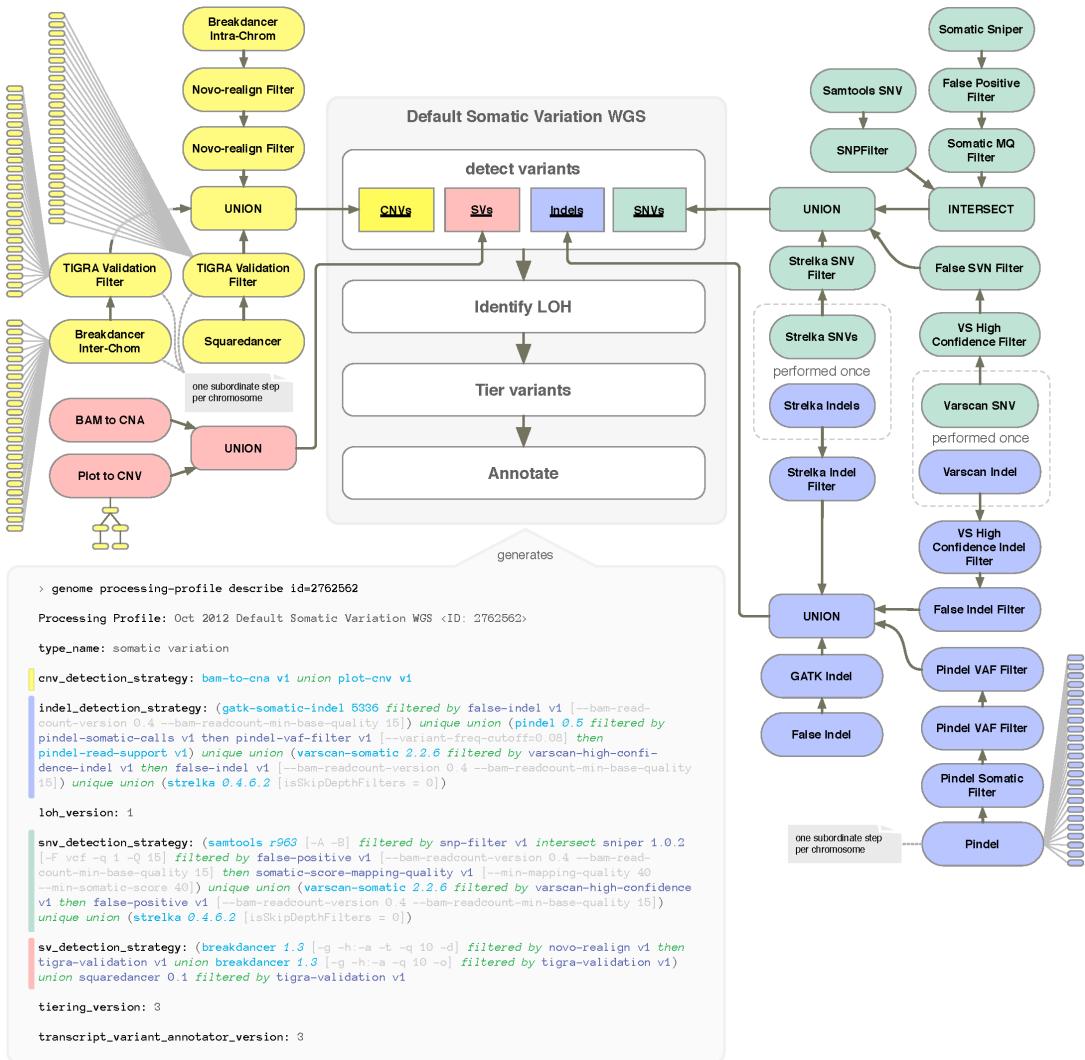
<https://github.com/genome/gms>

Griffith et al, 2015, PLoS Comp Biol.

# A single tumor analysis requires many pipelines involving dozens of different software/algorithms



# Each component on the previous slide is itself made of many, many components



For example, the somatic variation pipeline shown here is a complex workflow involving multiple different variant callers for different variant types, intersections, unions, filters, annotations, and visualization steps

# First cancer whole genomes sequenced in recent years

nature

Vol 456 | 6 November 2008 | doi:10.1038/nature07485

## ARTICLES

### DNA sequencing of a cytogenetically normal a

Vol 461 | 8 October 2009 | doi:10.1038/nature08489

Timothy J. Ley<sup>1,2,3,4\*</sup>, El Brian H. Dunford-Shore<sup>1</sup>, Dan C. Koboldt<sup>3</sup>, Craig Tracie Miner<sup>3</sup>, Lucinda Nathan Sander<sup>3</sup>, Xiaozhong Rhonda E. Ries<sup>1</sup>, Jacqueline Ivanovich<sup>4,7</sup>, S Daniel C. Link<sup>1,4</sup>, Timo

### Mutational evolution profiled at single

Sohrab P. Shah<sup>1,2,\*</sup>, Ryan D. Moore<sup>1</sup>, Allen Delaney<sup>3</sup>, Karen Gelmon<sup>4</sup>, Mark Sun<sup>1</sup>, Gillian Leung<sup>1</sup>, Richard Gulisa Turashvili<sup>1</sup>, Richard Varley<sup>1</sup>, David Huntsman<sup>2,5</sup>, Martin Hirsch<sup>6</sup>

2008 - AML

nature

## LETTERS

2009 - Breast

Vol 463 | 14 January 2010 | doi:10.1038/nature08629

## ARTICLES

2010 - SCLC

### A small-cell lung cancer genome with complex

Vol 463 | 14 January 2010 | doi:10.1038/nature08629

Erin D. Pleasance<sup>1</sup>, Paul J. Meltzer<sup>1</sup>, Meng-Lay Lin<sup>1</sup>, David J. Jones<sup>1</sup>, Helen R. Davies<sup>1</sup>, Goi Mingming Jia<sup>1</sup>, Catherine Jonathon Mangion<sup>2</sup>, Clarence C. Lee<sup>3</sup>, John M. Stratton<sup>1</sup>

### A comprehensive catalogue of somatic mutations

## ARTICLE

ARTICLES 2010 - MEL

2011 - PRC

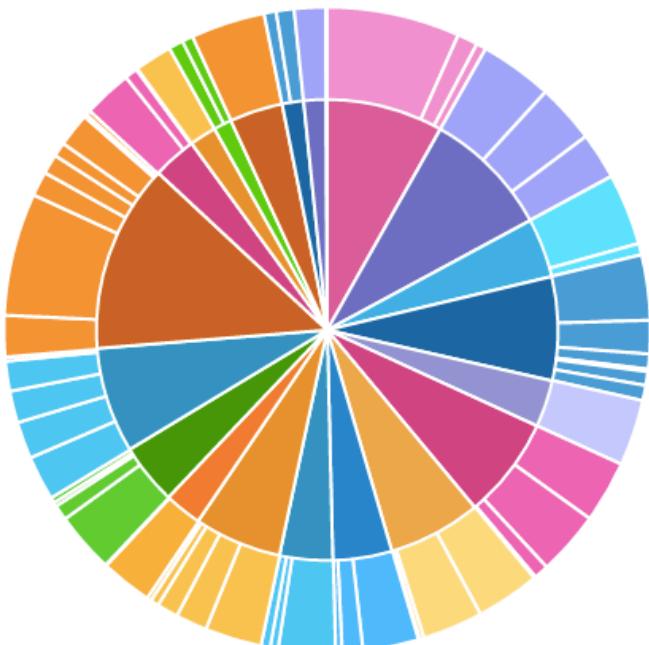
### The genomic complexity of primary human prostate cancer

Michael F. Berger<sup>1,1\*</sup>, Michael S. Lawrence<sup>1,\*</sup>, Francesca Demichelis<sup>2,3\*</sup>, Yotam Drier<sup>4\*</sup>, Kristian Cibulskis<sup>1</sup>, Andrey Y. Sivachenko<sup>1</sup>, Andrea Sboner<sup>1,6</sup>, Raquel Esgueva<sup>2</sup>, Dorothée Pfleiderer<sup>2</sup>, Carrie Sogne<sup>1</sup>, Robert Onofrio<sup>1</sup>, Scott L. Carter<sup>1</sup>, Kyung Park<sup>2</sup>, Lukas Habegger<sup>6</sup>, Lauren Ambrogio<sup>3</sup>, Timothy Fennell<sup>1</sup>, Melissa Parkin<sup>1</sup>, Gordon Saksena<sup>2</sup>, Douglas Voet<sup>1</sup>, Alex H. Ramos<sup>1,7</sup>, Trevor J. Pugh<sup>1,7,8</sup>, Jane Wilkinson<sup>1</sup>, Sheila Fisher<sup>1</sup>, Wendy Winckler<sup>1</sup>, Scott Mahan<sup>1</sup>, Kristin Ardlie<sup>1</sup>, Jennifer Baldwin<sup>1</sup>, Jonathan W. Simons<sup>9</sup>, Naoki Kitabayashi<sup>2</sup>, Theresa Y. MacDonald<sup>2</sup>, Philip W. Kantoff<sup>7,8</sup>, Lynda Chin<sup>1,7,8,10</sup>, Stacey B. Gabriel<sup>1</sup>, Mark B. Gerstein<sup>1,6,11</sup>, Todd R. Golub<sup>1,12,13,14</sup>, Matthew Meyerson<sup>1,7,8,14</sup>, Ashutosh Tewari<sup>15</sup>, Eric S. Lander<sup>1,7,16</sup>, Gad Getz<sup>1</sup>, Mark A. Rubin<sup>2</sup> & Levi A. Garraway<sup>1,7,8,14</sup>

TCGA and ICGC have sequenced thousands of exomes (and some genomes) and surveyed the landscape of mutated genes for 65 cancer types

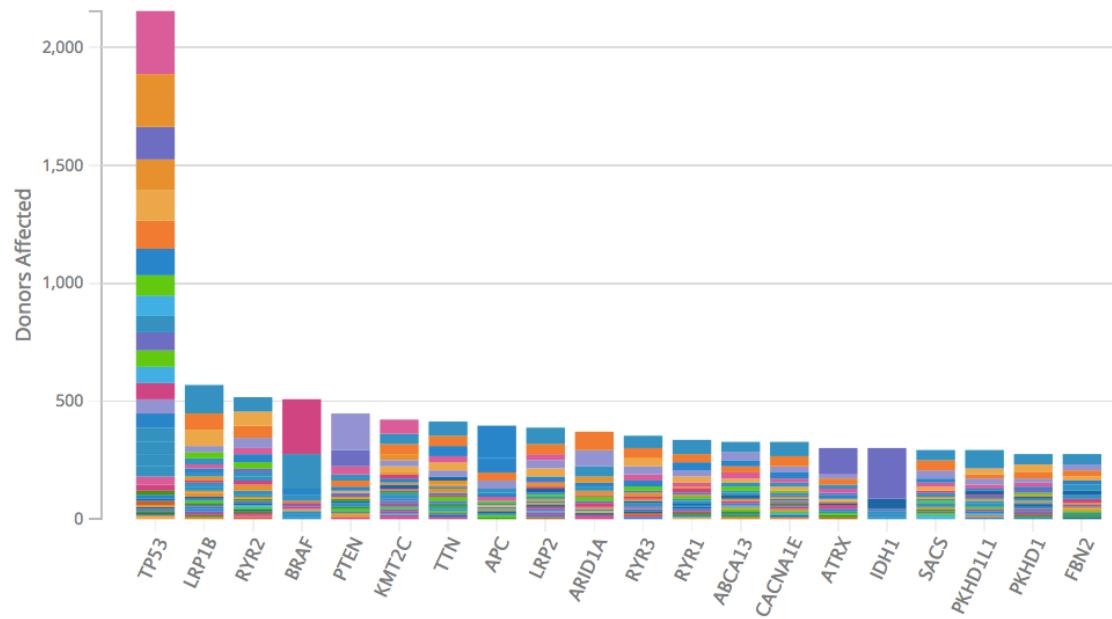
### Donor Distribution

16,318 Donors across 65 Projects



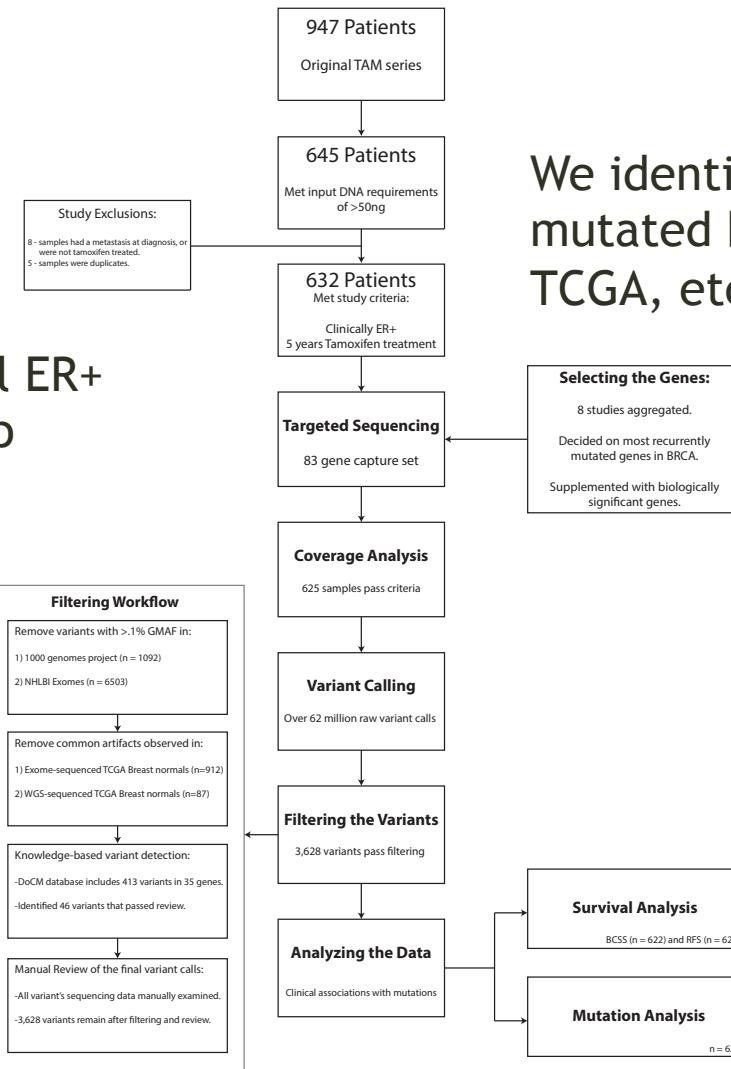
Top 20 Mutated Genes with High Functional Impact SSMs

8,038 Unique SSM-Tested Donors



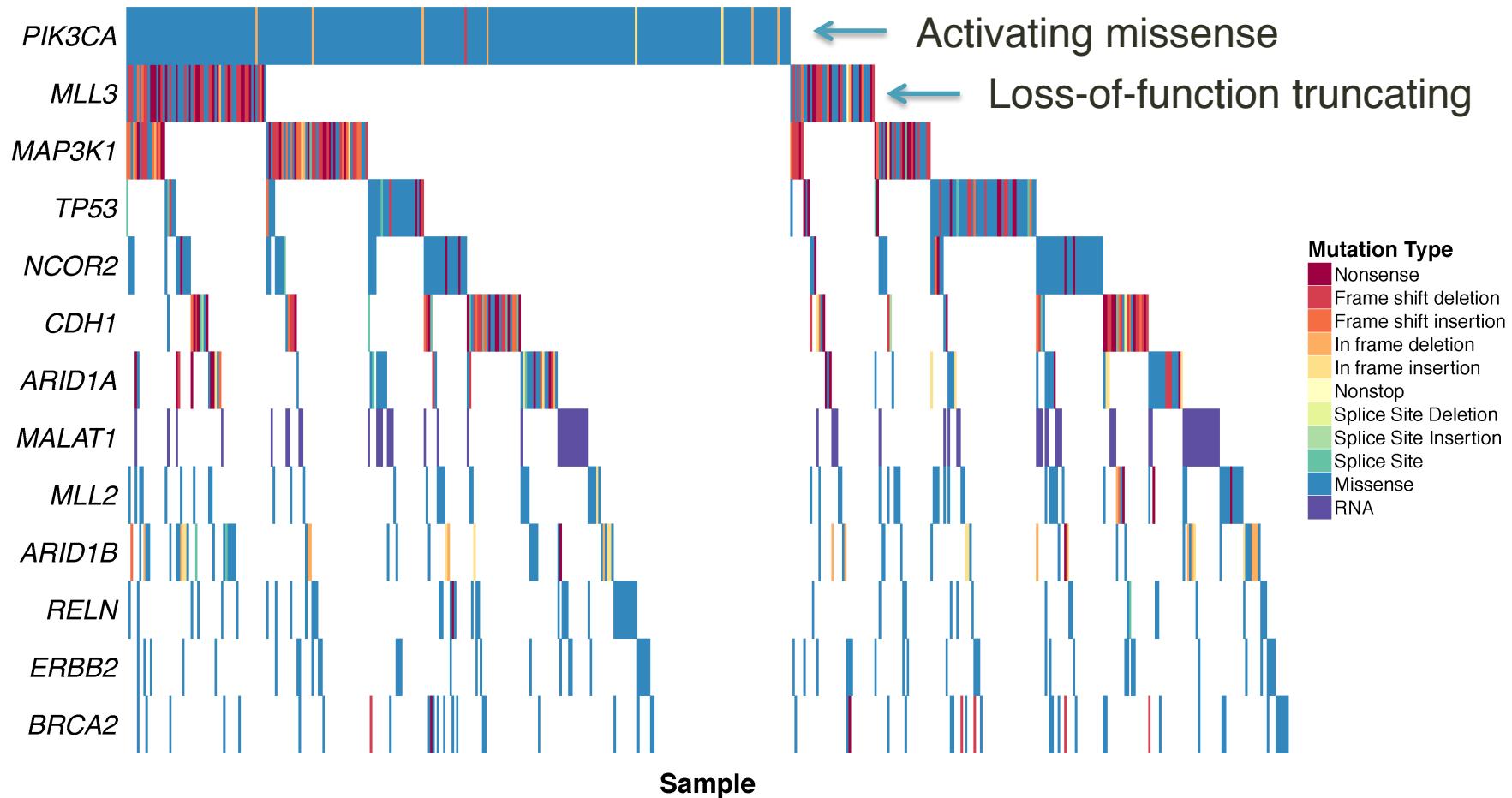
# Important to remember that these surveys are preliminary/incomplete

Sequenced 625 additional ER+ breast cancers with deep targeted sequencing



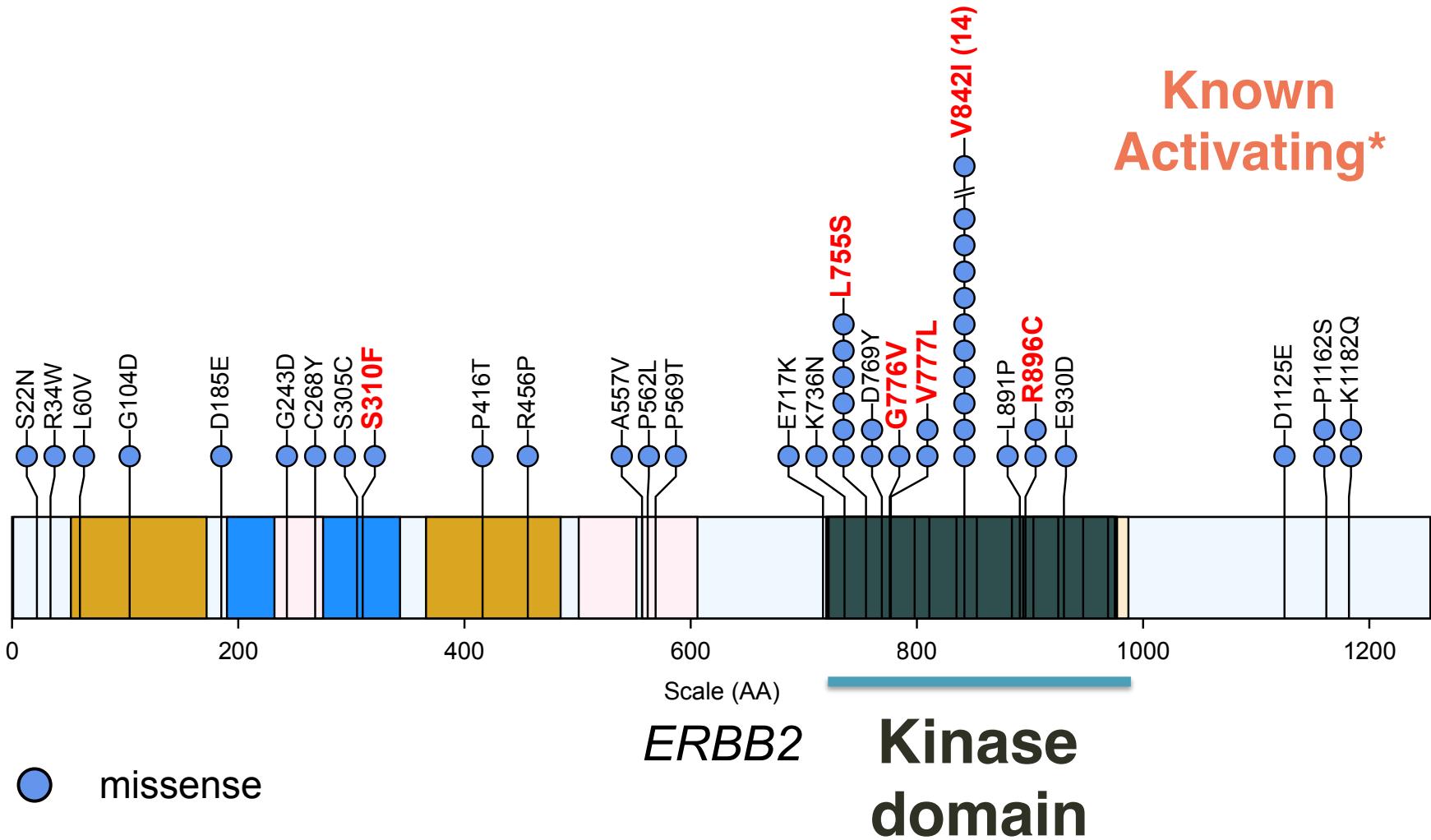
We identified 83 recurrently mutated breast cancer genes from TCGA, etc

# Mutation rates generally recapitulate TCGA exome-sequencing results



5 KRAS G12/13; 1 BRAF V600E; 28 AKT1 E17K

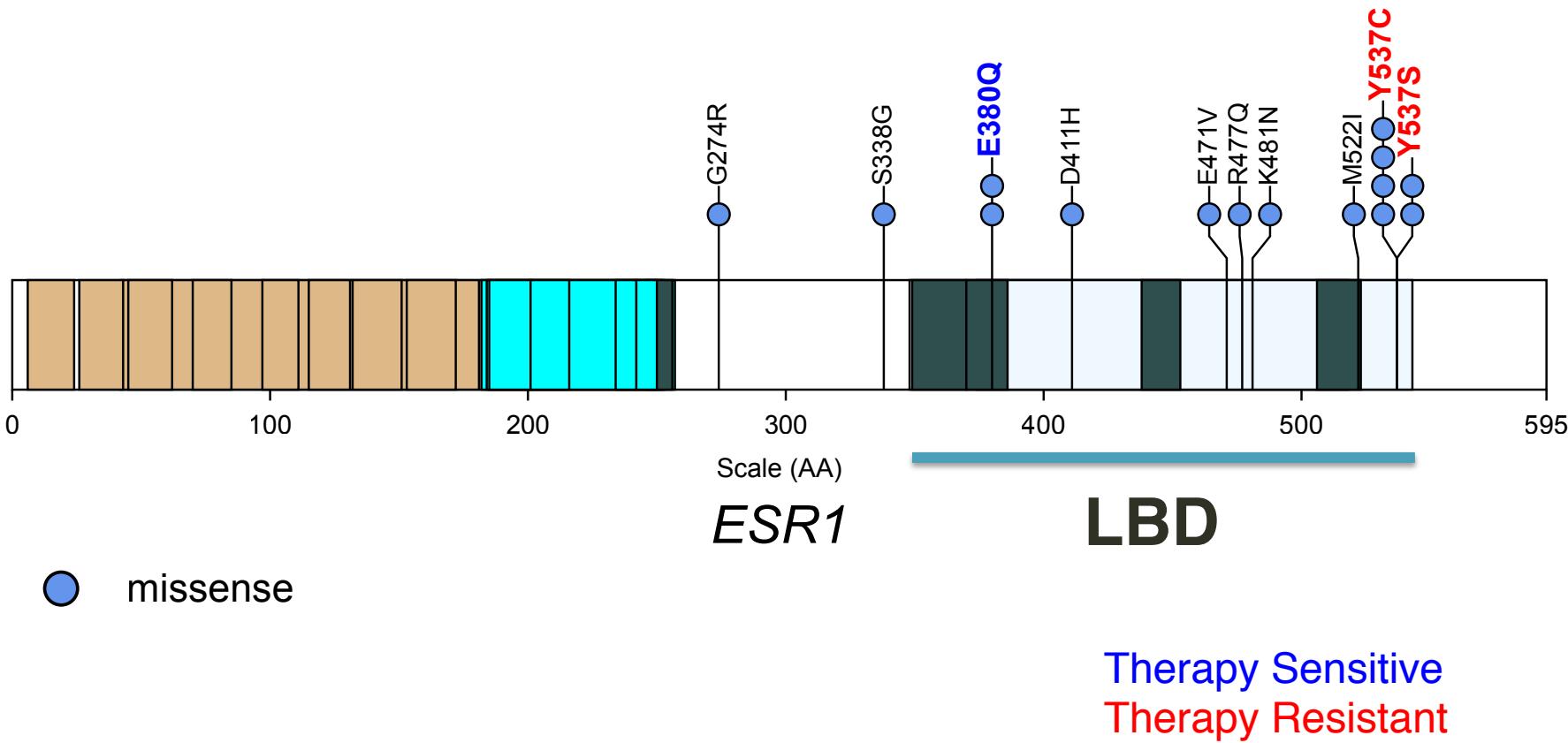
# Potential ERBB2 kinase domain mutations in up to 5.0% samples



\*Bose. et al, 2012, Cancer Discovery.

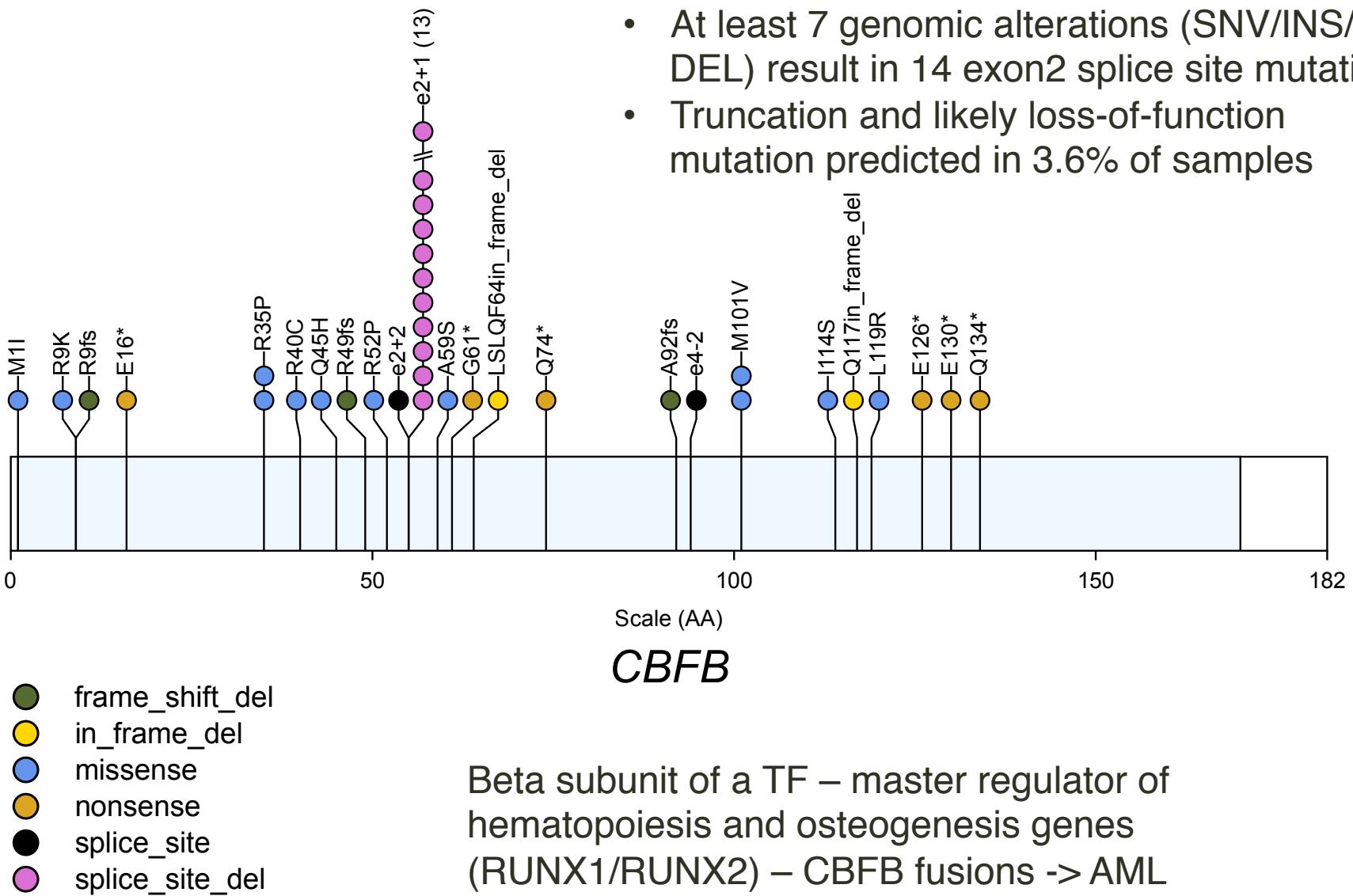
Note: many mutations (especially V842I) at low variant allele frequency

# ESR1 ligand binding domain mutated in up to 2.1% samples



# Novel splice site mutation hotspot in CBFB

- At least 7 genomic alterations (SNV/INS/DEL) result in 14 exon2 splice site mutations
- Truncation and likely loss-of-function mutation predicted in 3.6% of samples



# What is Precision (Personalized) medicine?

- Precision medicine (aka ‘Personalized medicine’ or ‘Individualized medicine’)
- The NIH Precision Medicine Initiative defines Precision medicine as - *an emerging approach for disease treatment and prevention that takes into account individual variability in genes, environment, and lifestyle for each person*
  - This definition makes regular (imprecision) medicine sound pretty lame
    - Doctors already consider genetics (family history), environment (smoking status, etc) and lifestyle (e.g., obesity) into account for practically every medical decision
- For our purposes let’s define precision medicine as the *patient-specific tailoring of medical decisions (diagnoses, prognoses, and treatments) through advanced genetic, molecular or cellular analyses*
  - A very specific subset of this definition would be the **use of genome-wide (sequencing) technologies to identify genetic features/events associated with some form of clinical actionability**

# Low-throughput (single-gene) molecular assays have already transformed treatment of some cancer types

- Many individually proven biomarkers or biomarker panels
  - ER/PR/HER2 in breast cancer
    - ER+: Tamoxifen (adjuvant chemo depends on risk or relapse, see below)
    - HER2+: amplification -> Trastuzumab
    - ER-/PR-/HER2-: (neoadjuvant)/adjuvant chemo + RT
- Expression signatures
  - Oncotype DX - A microarray derived, 21-gene signature, that predicts prognosis in ER+ breast cancer and can be used to assign risk of relapse allowing some patients to avoid adjuvant chemo)
- Many other examples

# Van Allen et al present an approach for high quality clinical WES from FFPE and algorithm for interpretation

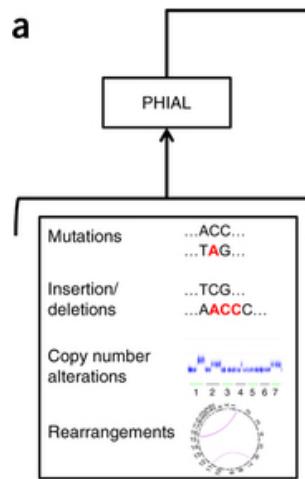
TECHNICAL REPORTS

nature  
medicine

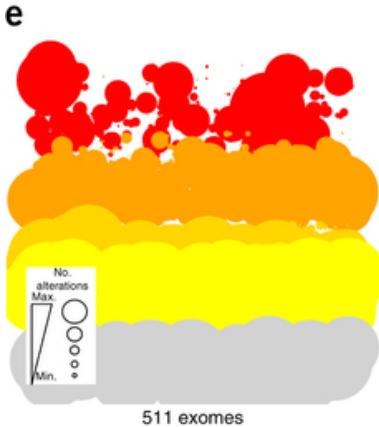
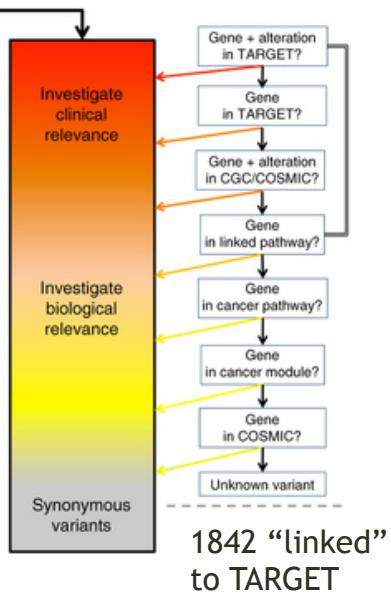
## Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine

Eliezer M Van Allen<sup>1,2,8</sup>, Nikhil Wagle<sup>1,2,8</sup>, Petar Stojanov<sup>1,2</sup>, Danielle L Perrin<sup>2</sup>, Kristian Cibulskis<sup>2</sup>, Sara Marlow<sup>1,2</sup>, Judit Jane-Valbuena<sup>1,2</sup>, Dennis C Friedrich<sup>2</sup>, Gregory Kryukov<sup>2</sup>, Scott L Carter<sup>2</sup>, Aaron McKenna<sup>2,3</sup>, Andrey Sivachenko<sup>2</sup>, Mara Rosenberg<sup>2</sup>, Adam Kiezun<sup>2</sup>, Douglas Voet<sup>2</sup>, Michael Lawrence<sup>2</sup>, Lee T Lichtenstein<sup>2</sup>, Jeff G Gentry<sup>2</sup>, Franklin W Huang<sup>1,2</sup>, Jennifer Fostel<sup>2</sup>, Deborah Farlow<sup>2</sup>, David Barbie<sup>1</sup>, Leena Gandhi<sup>1</sup>, Eric S Lander<sup>2</sup>, Stacy W Gray<sup>1</sup>, Steven Joffe<sup>1,4</sup>, Pasi Janne<sup>1</sup>, Judy Garber<sup>1</sup>, Laura MacConaill<sup>1,5</sup>, Neal Lindeman<sup>1,5</sup>, Barrett Rollins<sup>1</sup>, Philip Kantoff<sup>1</sup>, Sheila A Fisher<sup>2</sup>, Stacey Gabriel<sup>2,9</sup>, Gad Getz<sup>2,6,7,9</sup> & Levi A Garraway<sup>1,2,9</sup>

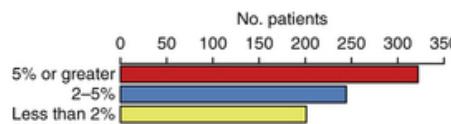
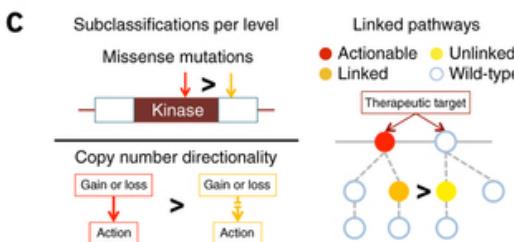
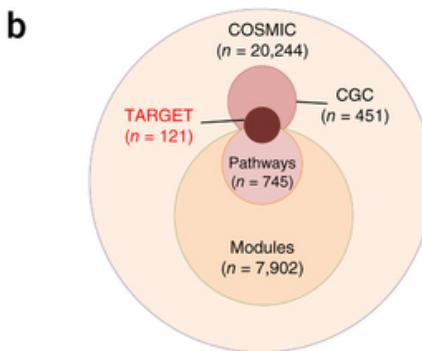
# Developed a “database” of tumor alterations relevant for genomics-driven therapy (TARGET) and precision heuristics for interpreting the alteration landscape (PHIAL)



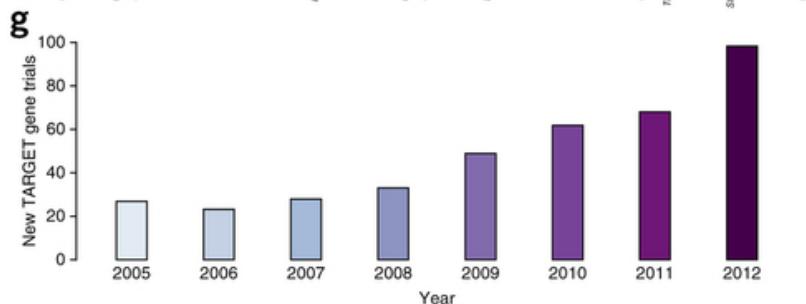
258,226 somatic coding mutations



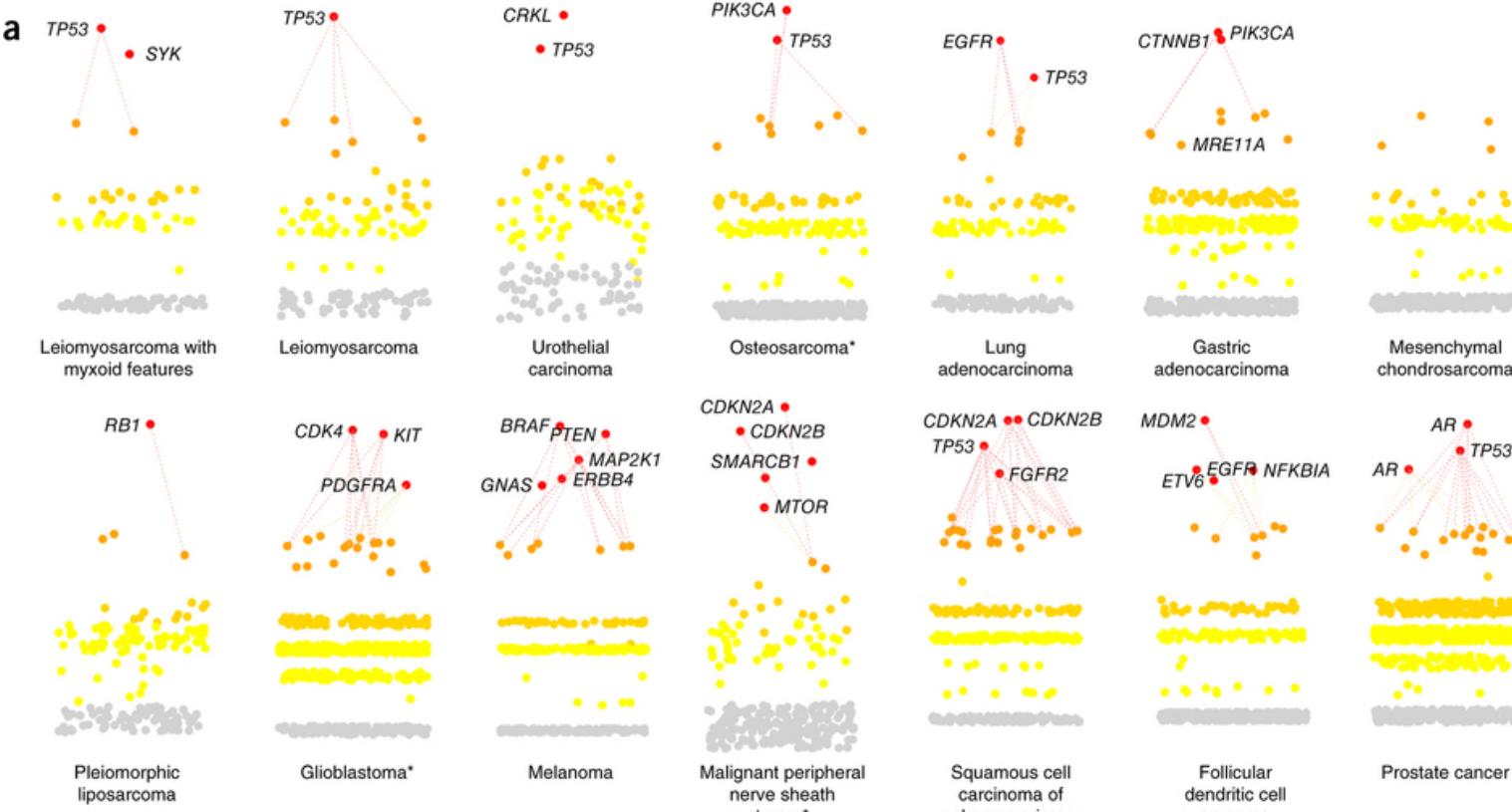
Applied method to 511 published exomes from lung, breast, prostate, head/neck SCC, melanoma and DLBCL



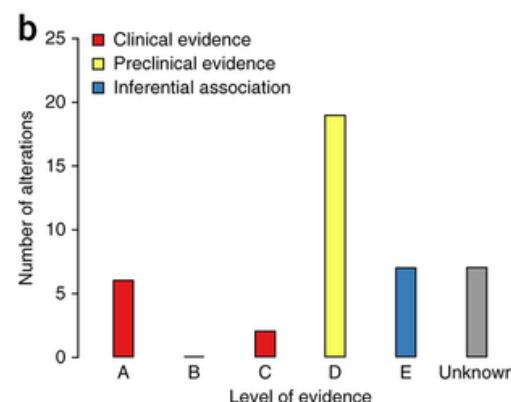
80% of patients have mutation in at least one TARGET-linked gene



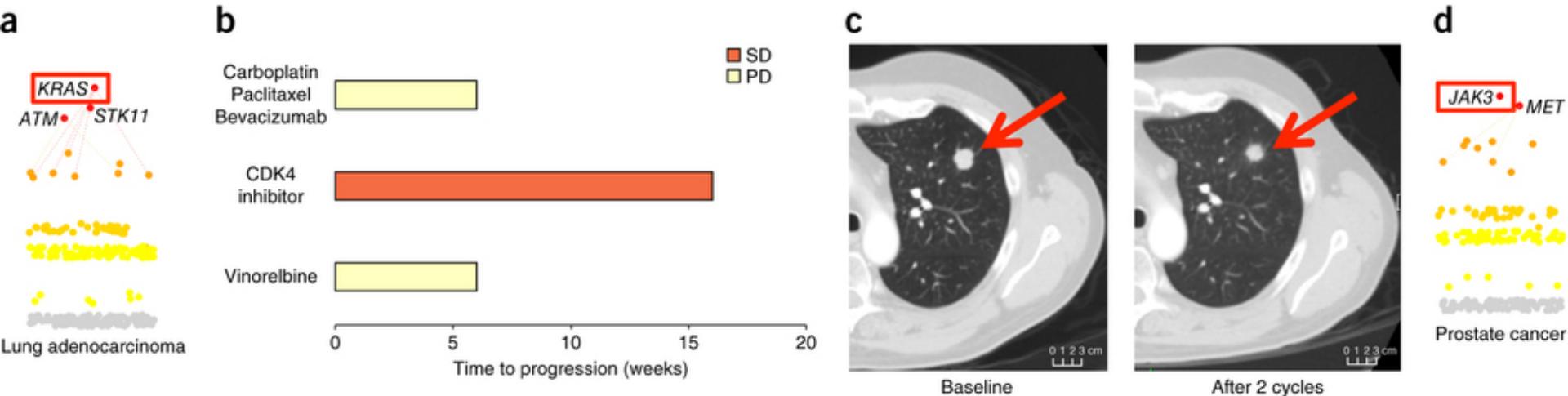
# Performed a prospective pilot in 16 patients with advanced cancers



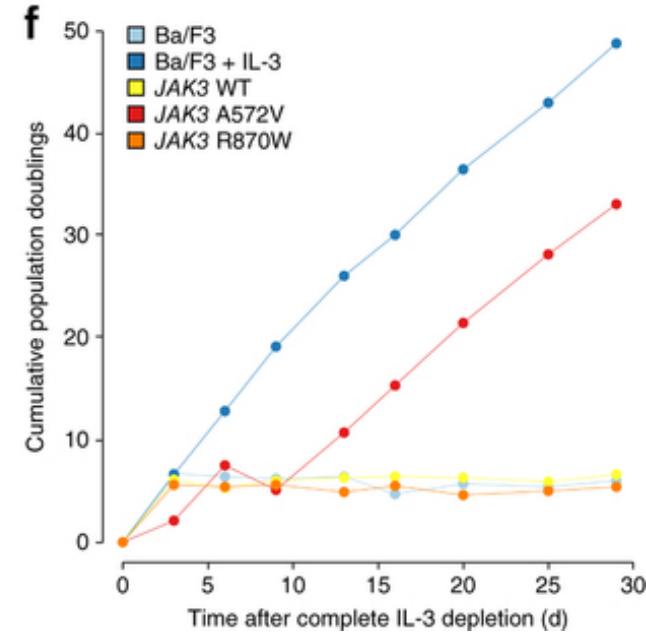
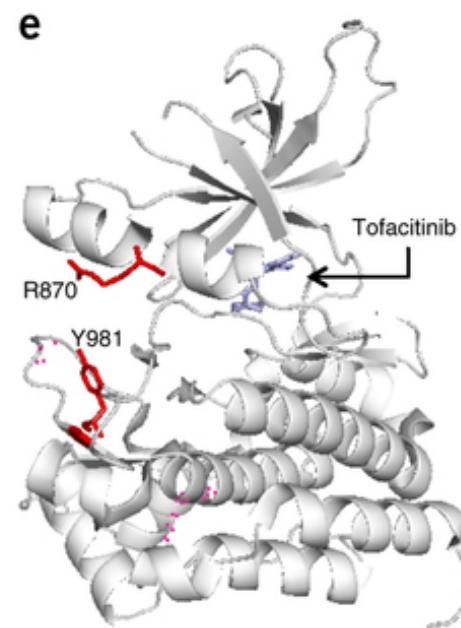
- 29 unique TARGET genes found for 16 patients (14 shown) (median 2, range 0-5)
- manually reviewed and assigned to evidence level
- 41 Clinically relevant alterations in 15/16 patients: EGFR L858R, PIK3CA alterations
  - 46.3% based on preclinical evidence



# One patient's clinical treatment was directly affected



- Metastatic lung cancer
- Targeted testing: wild-type for EGFR, KRAS, ALK
  - STK11 frameshift
- WES and PHIAL confirmed STK11 and identified a KRAS A146V and ATM mutations
- KRAS A146V is a known activating mutation
- Preclinical data implicated synthetic lethal relationship between KRAS and CDK4
- Patient put on CDK4i phase I trial
  - Achieved stable disease (RECIST), 7.9% tumor reduction - this was the patients "best and only response to any cancer-directed therapy"



Anecdotal cases have begun to demonstrate the benefit of genome-wide approaches for patients

The New York Times

In Treatment for Leukemia, Glimpses of the Future



 **Second Chance:** Lukas Wartman, a leukemia doctor and researcher, developed the disease himself. As he faced death, his colleagues sequenced his cancer genome. The result was a totally unexpected treatment.

By GINA KOLATA

Published: July 7, 2012

# Dr. Lukas Wartman wrote about his experience in the new journal Molecular Case Studies

 COLD SPRING HARBOR  
Molecular Case Studies

HOME | ABOUT | ARCHIVE | SUBMIT | AUTHOR INFO | ADVERTISE | ALERTS / RSS | CONTACT | HELP



Search for Keyword:  Go  
[Advanced Search](#)

## A case of me: clinical cancer sequencing and the future of precision medicine

Lukas D. Wartman

Author Affiliations

Corresponding author: [lwartman@dom.wustl.edu](mailto:lwartman@dom.wustl.edu)

**Abstract**

In this invited Perspective, I detail how my own experience as a patient with acute lymphoblastic leukemia (ALL) exemplifies several key concepts central to the implementation of cancer sequencing and precision medicine into clinical practice.

The field of precision medicine continues to rapidly grow and evolve as a result of the tremendous advances in sequencing technology, the ongoing improvement in genomic data analysis, and the decreasing cost associated with both sequencing data generation and analysis. The overarching principle that defines precision medicine is for health-care providers to optimize the treatment of patients, regardless of their disease, based on the results of comprehensive testing that can reliably and robustly assay for the relevant markers of individual variability that are pertinent for successfully eradicating a given disease, while minimizing the toxicity of the therapy to the patient. Certainly, clinical genomics, irrespective of the sequencing platform, is not synonymous with precision medicine as many other factors weigh into applying individualized therapy to a particular patient, such as environmental exposures and the patient's own treatment preferences. Nonetheless, the potential for clinical sequencing results to guide treatment choices is clear. Moreover, in the future, the results of sequencing studies may provide the basis for targeted efforts for disease prevention in patients found to

« Previous | Next Article »  
[Table of Contents](#)

**This Article**

doi:  
10.1101/mcs.a000349  
*Cold Spring Harb Mol Case Stud* 1: a000349  
© 2015 Wartman; Published by Cold Spring Harbor Laboratory Press

**Abstract**  
» Full Text  
Full Text (PDF)

**Article Category**  
PERSPECTIVE

**Services**  
+ Google Scholar  
+ PubMed/NCBI

**Figures**

  
View larger version:  
» In this page » In a new window  
» Download as PowerPoint Slide

Dr. Lukas Wartman photographed next to an Illumina HiSeq X Ten sequencer at The Genome Institute at Washington University School of Medicine. Photo by Ish Perk

**Current Issue**  
October 2015, 1 (1)



**Featured Articles**

- Elaine Mardis and Ralph DeBerardinis: from "N of 1" to "N of more"
- Cancer genomics in the clinic: a patient's perspective
- Metabolomics as an emerging tool in precision medicine
- Exome sequencing in neurological condition guides successful treatment
- De novo mutations in *POGZ* associated with severe intellectual disability

**Early Release Articles**

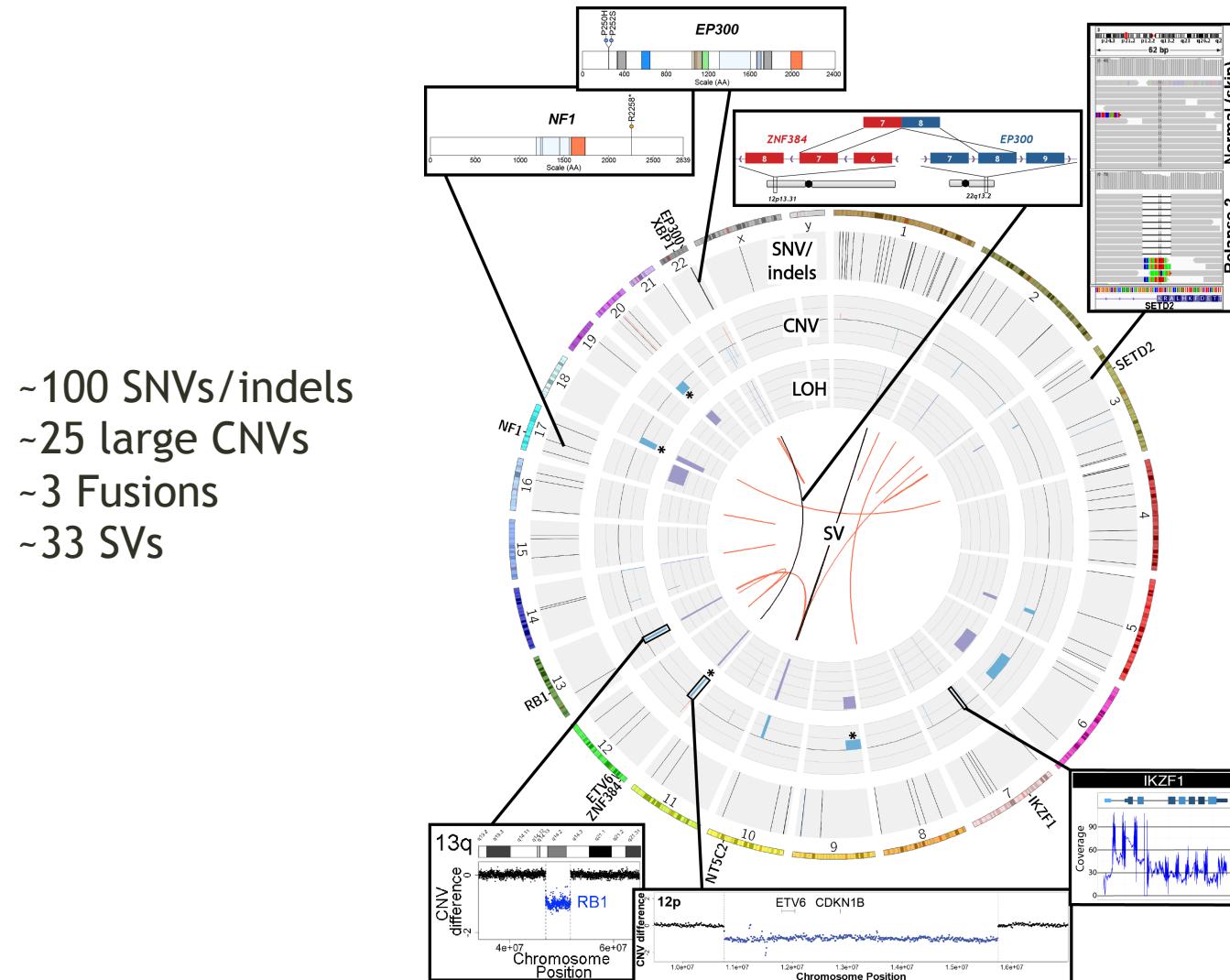
# Dr. Wartman's clinical history with ALL

- Diagnosed with ALL in 2003 while finishing med school
  - Treated with a multidrug chemotherapy regimen
    - Cure rate ~50% for adult ALL
    - Achieved morphologic remission
- Relapsed in 2008 during oncology fellowship
  - Treated with aggressive chemotherapy
    - Again, achieved morphologic remission
    - Received bone marrow transplant from younger brother
    - Overall survival of adults with ALL after first relapse is ~5%-10%
- Relapsed again in 2011
  - Treated with chemo again
    - Remission NOT achieved
    - Adult ALL in second relapse considered incurable, and unvaryingly fatal
    - Enrolled in study to sequence (WGS + RNAseq) 2<sup>nd</sup> relapse

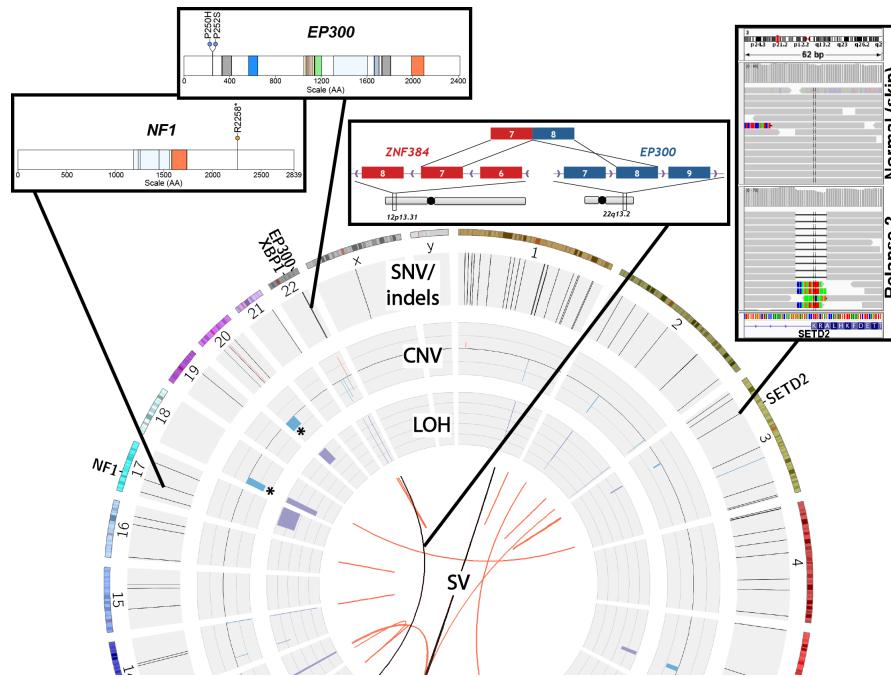
# Dr. Wartman did not expect to benefit

- “I elected to participate because I believed that sequencing my leukemia genome, the first adult ALL sample to be sequenced, could be the first step for researchers to understand why adults with ALL fare worse than children”
- “In my mind, the sequencing of my leukemia genome was “discovery” genomics. I did not expect the results to be clinically important for me.”
- “Yet, after the chemotherapy failed to achieve a remission, I quickly realized that identifying a novel target from the sequencing results was likely my only chance at survival.”

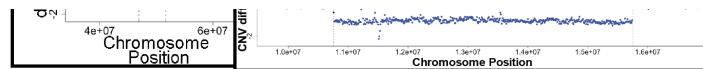
A team of genomic analysts led by Dr. Malachi Griffith identified a large number of somatic mutations but none was “targetable”



A team of genomic analysts led by Dr. Malachi Griffith identified a large number of somatic mutations but none was “targetable”

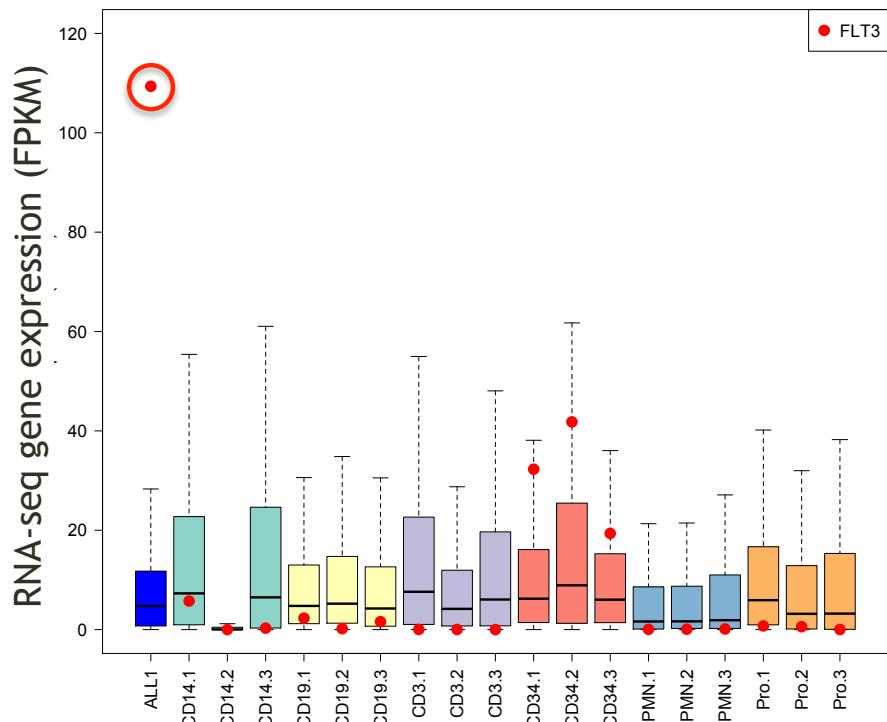


“...my oncologist called me about a week later to tell me that Malachi had identified a clinically actionable target from the RNA-seq data...”



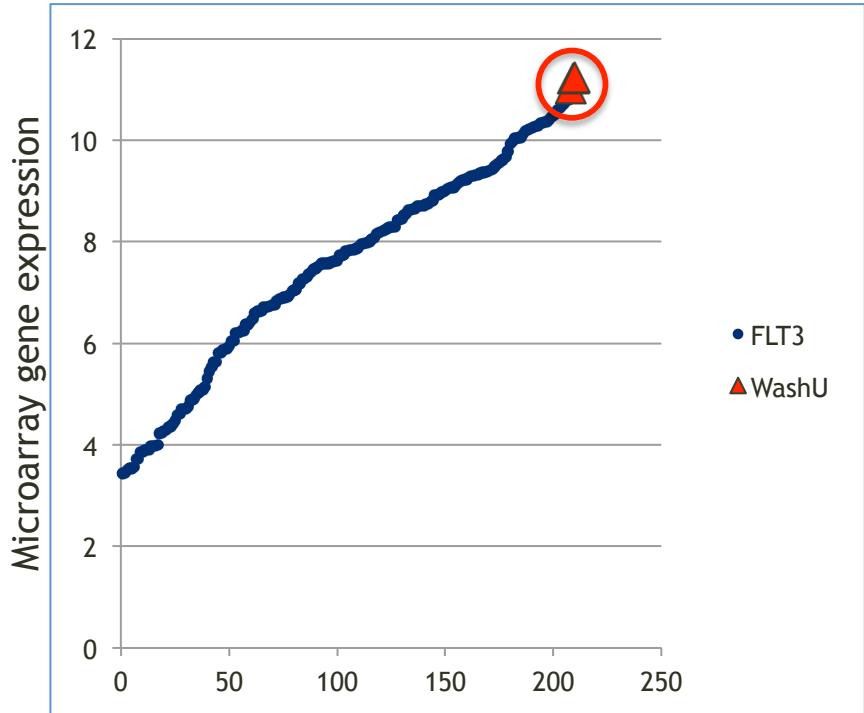
# FLT3 extreme over-expression identified from RNA-seq

RNA-seq expression data



ALL1 and 18 healthy donor samples

Microarray expression data



ALL1 (triplicates) and >200 B-ALL's

- FLT3 is a receptor tyrosine kinase important for hematopoietic cell survival, development, and proliferation.
- Previously work demonstrated that some ALL samples did have high FLT3 expression and that ALL lymphoblasts in culture were sensitive to FLT3 inhibitors
- No clinical trial results of ALL patients had been reported using these drugs

**Unclear whether his ALL would respond to FLT3 inhibitor  
- no other reasonable options available at the time**

- Started treatment with sunitinib (Sutent, Pfizer), FDA-approved for other types of cancer, and known to be a potent wild-type FLT3 inhibitor
- 2 weeks later, in a complete third remission and went on to have a second stem cell transplant from an unrelated matched donor
- Remains in complete remission more than 3 years later

**Unclear whether his ALL would respond to FLT3 inhibitor  
- no other reasonable options available at the time**

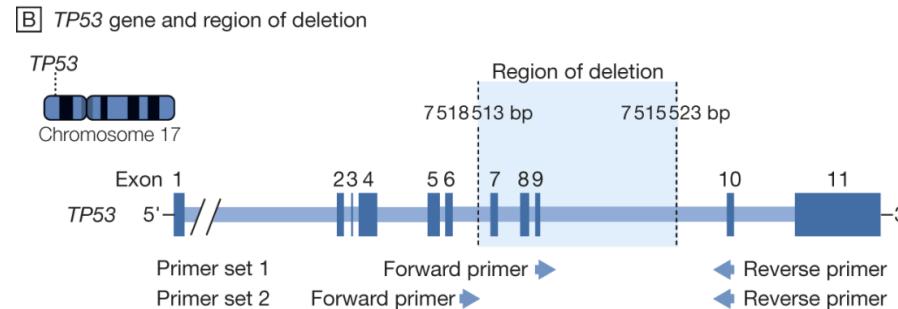
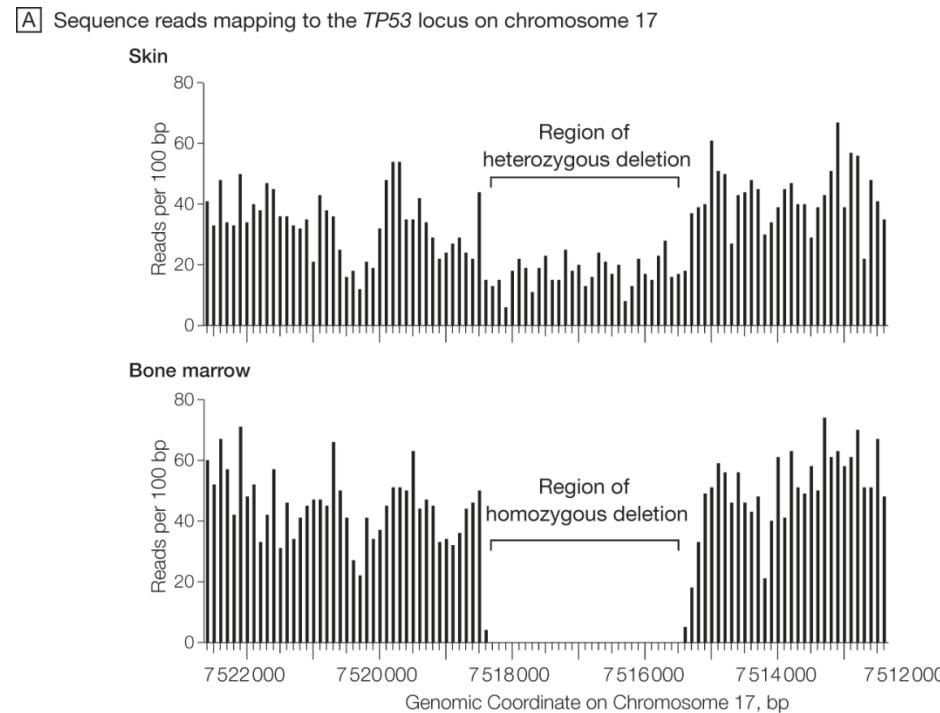
- Started treatment with sunitinib (Sutent, Pfizer), FDA-approved for other types of cancer, and known to be a potent wild-type FLT3 inhibitor
- 2 weeks later, in a complete third remission and went on to have a second stem cell transplant from an unrelated matched donor
- Remains in complete remission more than 3 years later

**“...my case is not worth sharing because of my good luck, but rather because it illustrates how cancer genomics and precision medicine can be clinically relevant.”**

# Link et al used WGS to identify a novel 3kb TP53 deletion that would be missed by targeted approaches

From: Identification of a Novel TP53 Cancer Susceptibility Mutation Through Whole-Genome Sequencing of a Patient With Therapy-Related AML

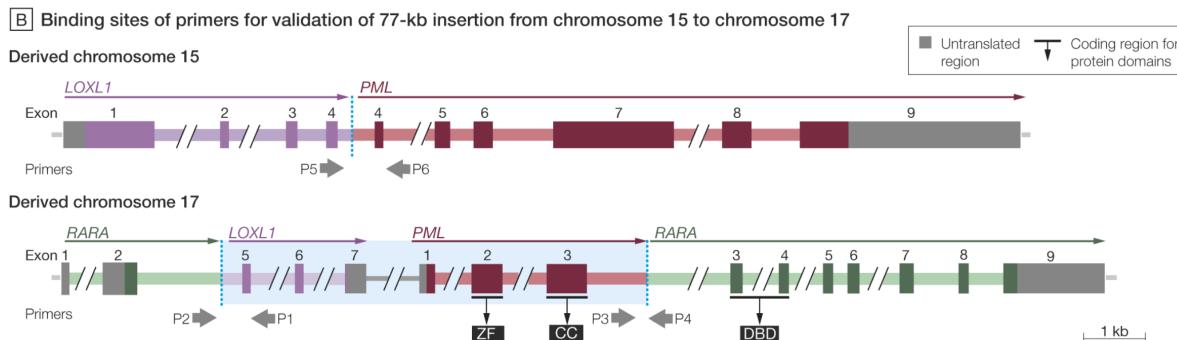
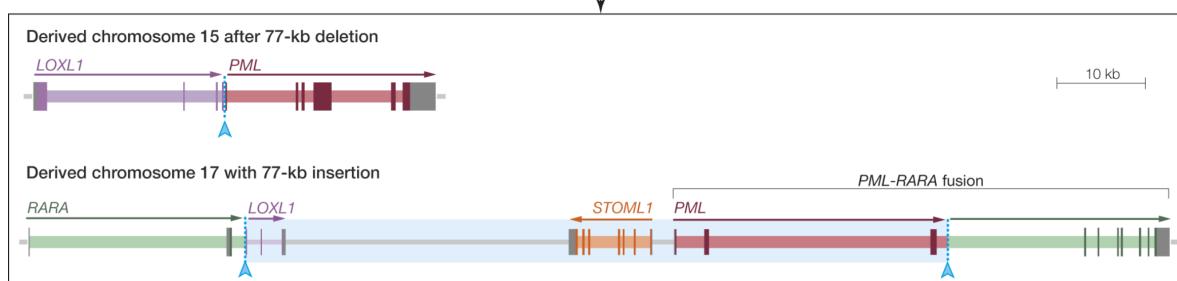
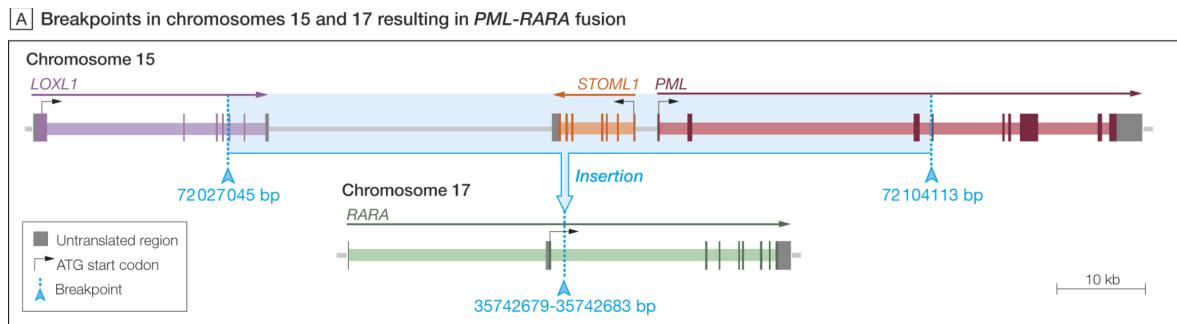
Link et al. JAMA. 2011;305(15):1568-1576. doi:10.1001/jama.2011.473



# Welch et al used WGS to identify a novel 77kb inter-chromosomal insertion resulting in a cytogenetically cryptic PML-RARA fusion

From: Use of Whole-Genome Sequencing to Diagnose a Cryptic Fusion Oncogene

JAMA. 2011;305(15):1577-1584. doi:10.1001/jama.2011.497



# I track (blog) on BioStars about new examples and journals as they are published

LATEST 4 OPEN 3 RNA-SEQ 1 CHIP-SEQ SNP ASSEMBLY TUTORIALS TOOLS JOBS

Obi Griffith ♦ 10k | Logout

 **BioStars**  
BIOINFORMATICS EXPLAINED

Community Messages Votes My Posts My Tags Following

## Forum: Publications for individualized medicine in cancer by whole genome, exome or transcriptome sequencing

 What papers are you aware of that attempt the following: (1) Whole genome, exome and/or transcriptome (RNA) sequencing of (2) live patient tumor samples in an attempt to (3) guide clinical decision making for cancer? The omic events could provide diagnostic, prognostic or treatment response predictions. This approach 41 is widely referred to as personalized medicine, individualized medicine, precision medicine, or precision oncology. There are many reviews describing this idea and many examples that make use of small targeted panels (one to dozens of molecular events). I'm looking for proof-of-principle papers, describing the paradigm where researchers (or tumor genome boards) attempt to use omic (genome-wide) NGS data to alter or inform clinical care. I will also consider larger gene panels that attempt to comprehensively sequence "relevant" cancer genes (e.g., MiSeq/Ion/Proton/NextSeq500 scale approaches). Foundation Medicine provides a list using their targeted panel. Relevant studies could be N-of-1 case reports or overviews describing experiences with small to large cohorts.

 Here is a prototypical example in which an oral adenocarcinoma was sequenced by whole genome and transcriptome sequencing and analysis done to suggest a particular target/pathway for therapy that might not otherwise be considered in this disease type at the time. This is the earliest example that I am aware of but I would appreciate if anyone can point me to earlier examples.  
<http://genomebiology.com/content/11/8/R82>

UPDATE: I am collecting and organizing the responses here and also adding updates as I find them. They are listed chronologically and broken into prospective and retrospective categories (sometimes this distinction is fuzzy).

Prospective or pseudo-prospective studies:

- [Jones et al.](#) Evolution of an adenocarcinoma in response to selection by targeted kinase inhibitors. *Genome Biology*. 11:R82 (9 Aug 2010).
- [Link et al.](#) Identification of a Novel TP53 Cancer Susceptibility Mutation Through Whole-Genome Sequencing of a Patient With Therapy-Related AML. *JAMA*. 305(15):1568-1576 (20 Apr 2011).
- [Welch et al.](#) Use of Whole-Genome Sequencing to Diagnose a Cryptic Fusion Oncogene. *JAMA*. 305(15):1577-1584 (20 Apr 2011).
- [Roychowdhury et al.](#) Personalized Oncology Through Integrative High-Throughput Sequencing: A Pilot Study. *Sci Transl Med*. 3(111):111ra121 (30 Nov 2011).
- [Borad et al.](#) Integrated genomic characterization reveals novel, therapeutically relevant drug targets in FGFR and EGFR pathways in sporadic intrahepatic cholangiocarcinoma. *PLoS Genet*. 10(2):e1004135. (13 Feb 2014).
- [Van Allen et al.](#) Whole-exome sequencing and clinical interpretation of formalin-fixed, paraffin-embedded tumor samples to guide precision cancer medicine. *Nat Med*. 20(6):682-8 (18 May 2014).
- [Wagle et al.](#) Response and acquired resistance to everolimus in anaplastic thyroid cancer. *N Engl J Med*. 371(15):1426-33. (9 Oct 2014).
- [Juric et al.](#) Convergent loss of PTEN leads to clinical resistance to a PI(3)K inhibitor. *Nature*. 518(7538):240-4. (17 Nov 2014).
- [Sekulic et al.](#) Personalized treatment of Sézary syndrome by targeting a novel CTLA4:CD28 fusion. *Mol Genet Genomic Med*. 3(2):130-6. (27 Nov 2014).

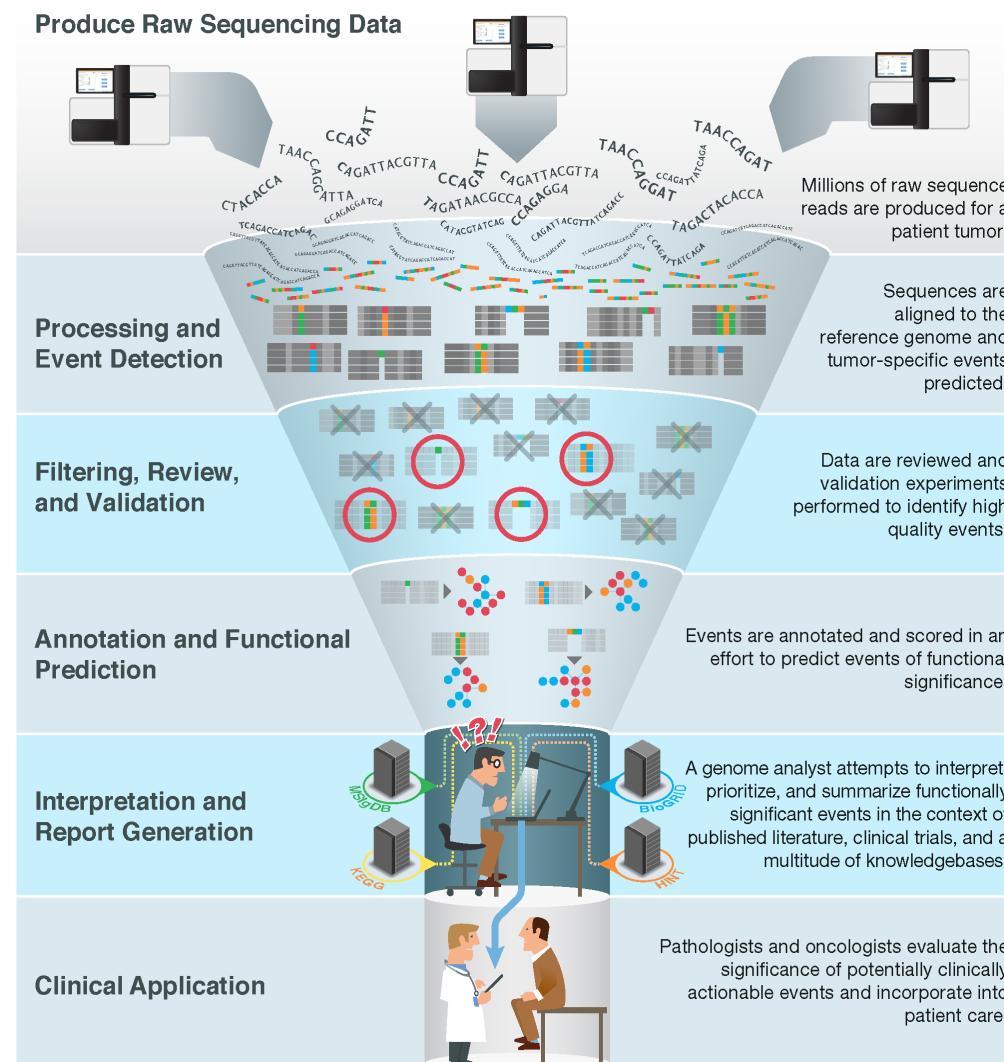
Retrospective studies:

- [Iyer et al.](#) Genome sequencing identifies a basis for everolimus sensitivity. *Science*. 338(6104):221 (23 Aug 2012).

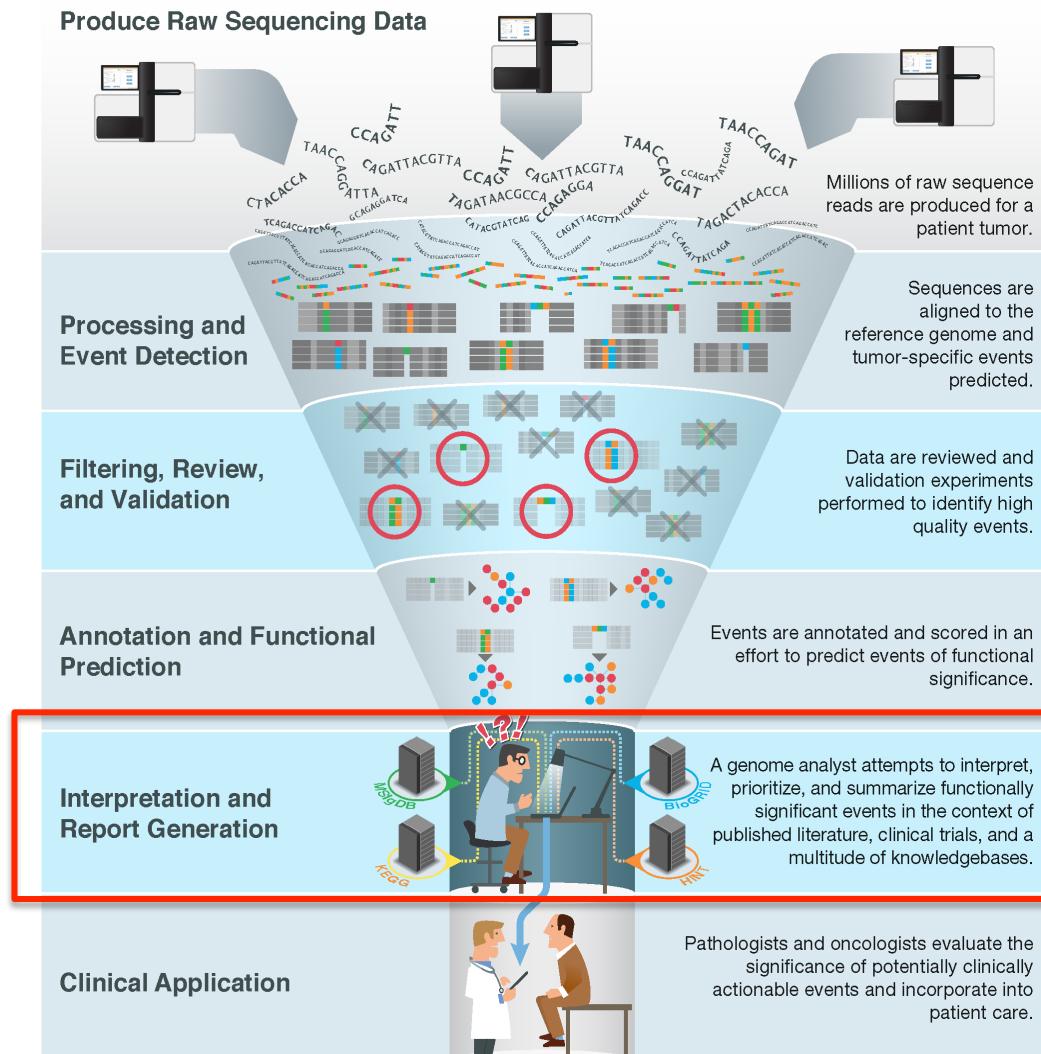
  
14 months ago by  
Obi Griffith ♦ 10K  
Washington University, St Louis, USA

Also, see inaugural Molecular Case Studies journal issue

# How can we extend this success to others - most aspects of sequencing and analysis have been standardized



# Visualization and Interpretation of genomic alterations in the context of the clinical relevance remains the major bottleneck

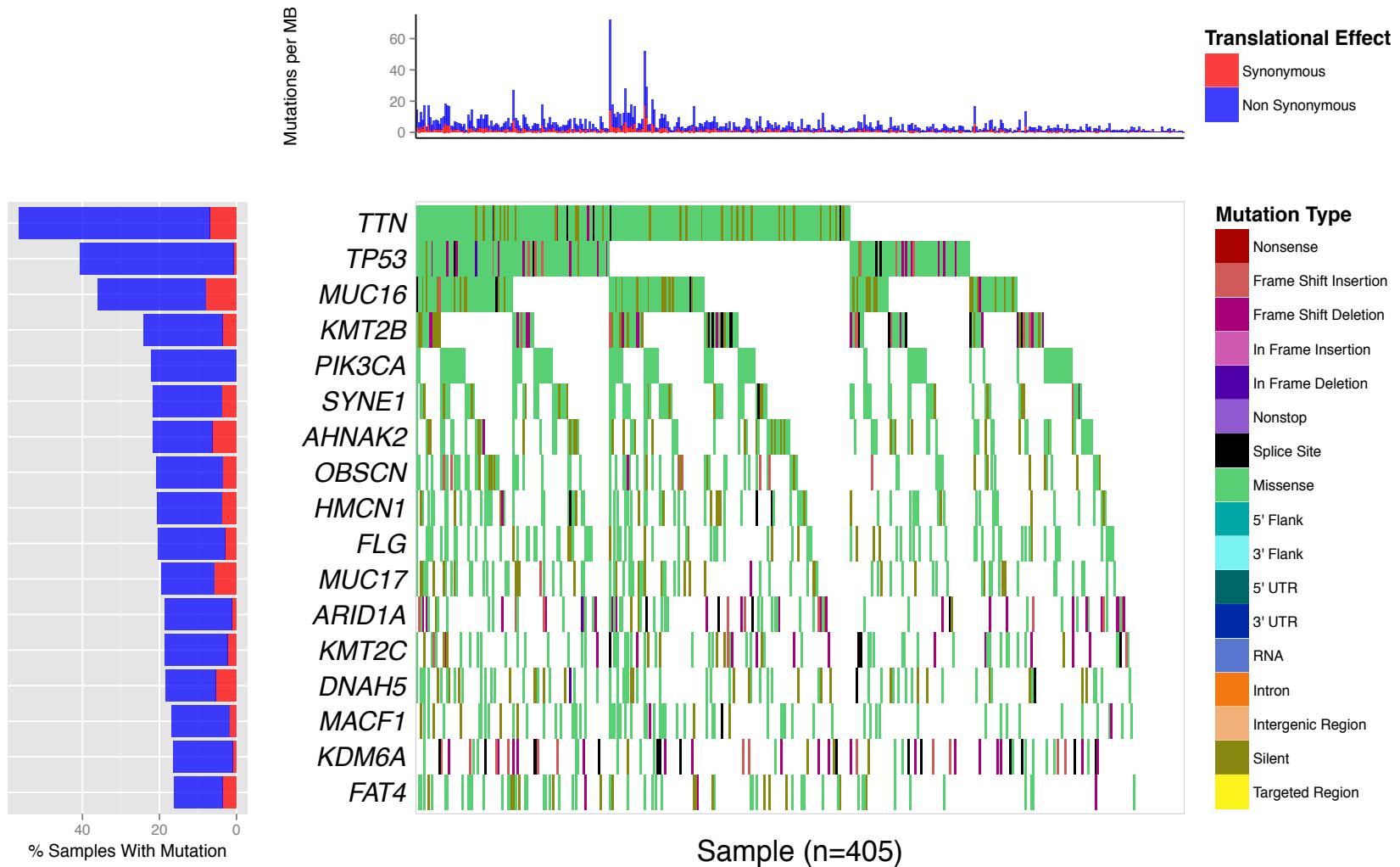


# First problem is visualization: genVisR - genome Visualizations in R

- genVisR attempts to provide highly customizable publication quality graphics for (cancer) genome analysis
  - Builds on existing popular packages, especially ggplots
- Primary Plots
  - Mutation Heatmap aka Waterfall Plot
  - Lollipop
  - Coverage Plot
  - Copy Number Cohort Plot
  - Copy Number Plot
  - Transition/Transversion Plot
- Sub Plots/functions
  - Gene Plot
  - Chromosome Plot
  - Label Plot
  - To be released via Bioconductor in near future
    - Will be accompanied by detailed tutorials at [www.biostars.org](http://www.biostars.org)

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

# genVisR allows non-experts to create commonly used genomics visualizations (e.g., mutation landscape)



# A couple of lines of simple R for sophisticated plots

Usage:

```
> maf_file <- read.delim('TCGA_MAF_FILE.maf')  
  
> mutation_heatmap(maf_file, grid=F, label_x=F, gene_label_size=18,  
file_type='MAF', title='BRCA', coverage_space=63564965, recurrence_cutoff=65)
```

Parameters:

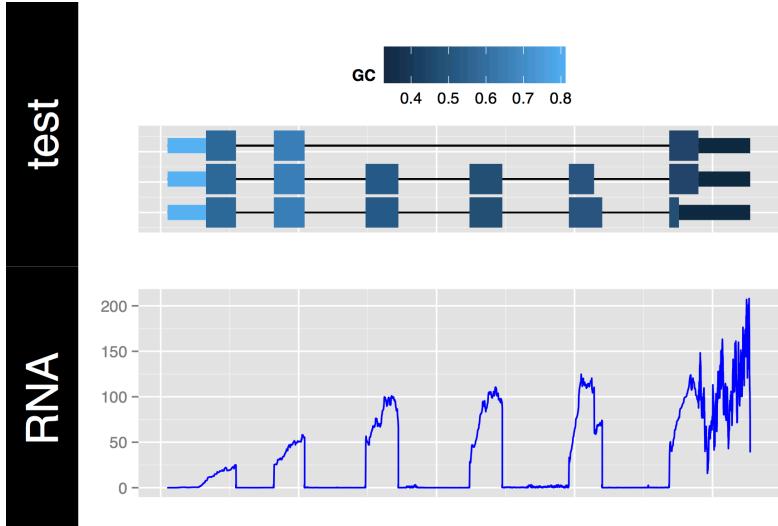
- Coverage\_space = space in bp from which a mutation could be expected to occur to be used in mutations per mb calculation
- Recurrence\_cutoff = integer specifying a requirement that a gene should have at least x number of mutations among the cohort to be plotted
- Grid = boolean specifying whether a grid separating samples/genes should be overlaid
- Label\_x = boolean specifying whether to plot sample names

Notes:

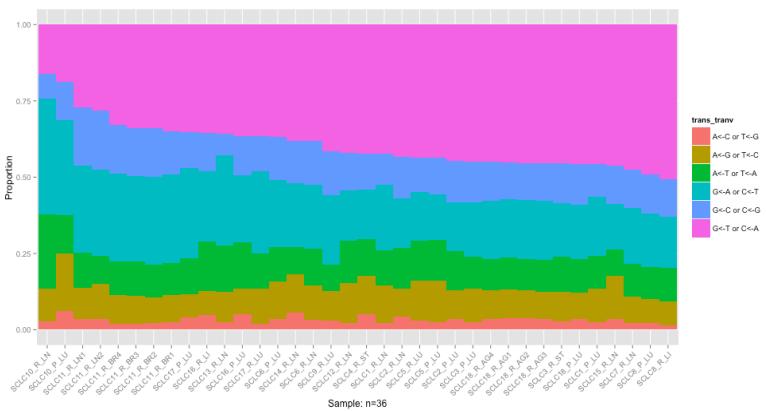
- To Plot samples with no mutations add those samples to the input file under the sample column
- Mutations per MB calculation occurs for the entire MAF file input not a subset resultant of a recurrence\_cutoff

# More genVisR examples

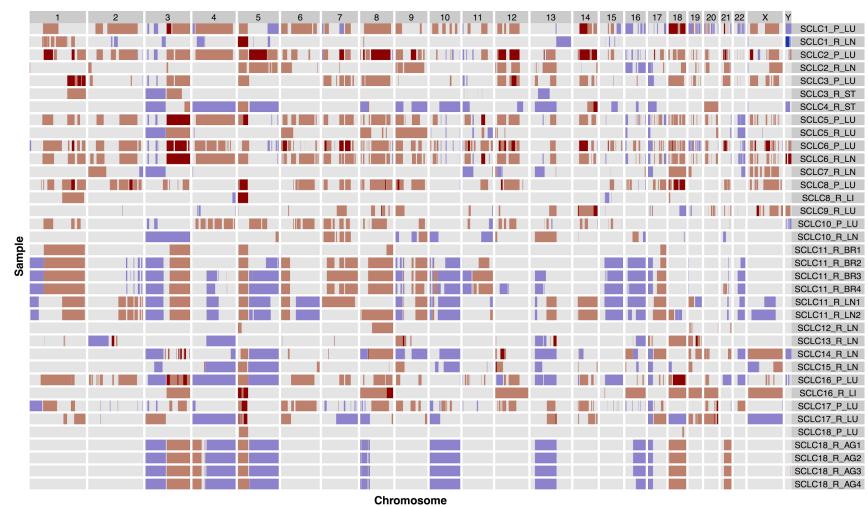
Sequence coverage



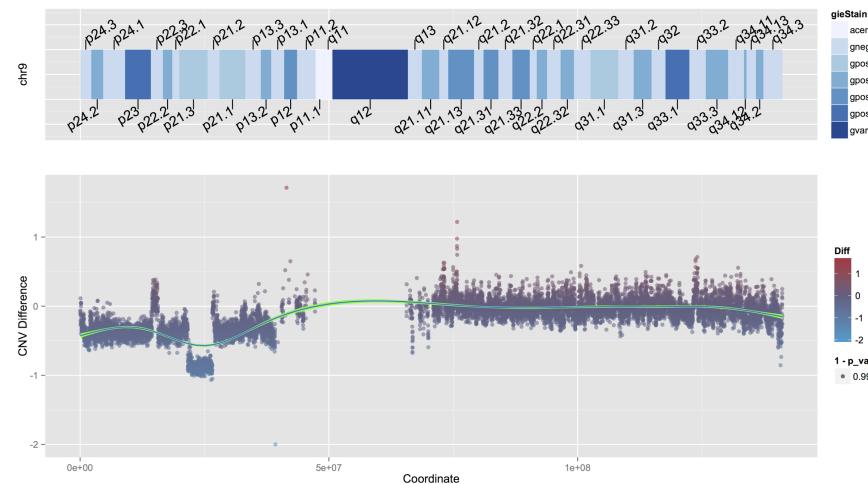
Mutation spectrum



CNV landscape



Genome/chr level CNV views



Many more...

# Next, resources are needed to support interpretation of genes, mutations and evidence for clinical actions



Alex Wagner



Ben Ainscough



Nick Spies, Kilannin Krysiak

- [www.dgidb.org](http://www.dgidb.org)
- Search genes for known and potentially druggable interactions

- [www.docm.info](http://www.docm.info)
- Highly curated set of mutations known to cause cancer

- [www.civicdb.org](http://www.civicdb.org)
- Highly curated summaries of clinical and functional characterization of mutations implicated in cancer

# DGIdb - Drug-gene interactions or druggable categories can be browsed or searched via web interface or API

The screenshot shows the DGIdb web interface. At the top, there's a navigation bar with links for 'Search Interactions', 'Search Categories', 'Browse Categories', and 'Help'. Below the navigation is a section titled 'Search Interactions' with the sub-instruction 'search for drug-gene interactons by gene name'. A 'Show Tour' link is present. On the left, there's a 'Genes' input field containing a list of genes: FLT1, FLT2, FLT3, STK1, MM1, LOC100508755, and FAKE1. To the right of this is a 'Search Interactions Tour' modal window with the text: 'Start by entering a list of gene names (one per line) that you would like to search.' and a 'Next >' button. Below the genes input are buttons for 'Replace Genes with Demo List' and 'Clear All Genes'. Further down, there are filters for 'Gene Category' (39 of 39), 'Interaction Type' (34 of 34), 'Source Database' (5 of 5), and 'Anti-Neoplastic Drugs Only' (checkbox). Under 'Select Output Format', there are radio buttons for 'HTML' (selected) and 'TSV'. At the bottom is a green 'Find Drug Interactions' button.

This screenshot shows the 'Interaction Search Results' page. The title is 'Interaction Search Results drug interactions for your genes'. Below it are tabs for 'Interaction Results' (selected), 'Search Results Summary', 'Search Term Summary', 'By Gene', and 'By Source'. The main area is titled 'Primary Results' with the sub-instruction 'Search terms matching exactly one gene that has one or more drug interactions.' It includes a dropdown for '10 records per page' and a 'Filter results:' input field. A table lists drug-gene interactions for FLT1 and FLT3. Below the table are wavy lines representing chemical structures. Navigation buttons at the bottom include 'Previous', '1', '2', '3', '4', '5', and 'Next →'. The table columns are 'Search Term', 'Gene', 'Drug', 'Interaction Type', and 'Source'.

This screenshot shows the 'Ambiguous Results' section of the search results. The title is 'Ambiguous Results' with the sub-instruction 'Search terms matching multiple genes, where some of those genes have drug interactions.'. It includes a dropdown for '10 records per page' and a 'Filter results:' input field. A table lists drug-gene interactions for STK1 and FLT3. Below the table are wavy lines representing chemical structures. Navigation buttons at the bottom include 'Previous', '1', '2', '3', and 'Next →'. The table columns are 'Search Term', 'Gene', 'Drug', 'Interaction Type', and 'Source'.

This screenshot shows two additional sections. The first is 'Ambiguously Matched Genes With No Interactions' which lists 'PFDN5' and 'PLXNB2' with their respective search terms. The second is 'Search Terms With No Matches' which lists 'LOC100508755' and 'FAKE1'.

Interface allows complex filtering based

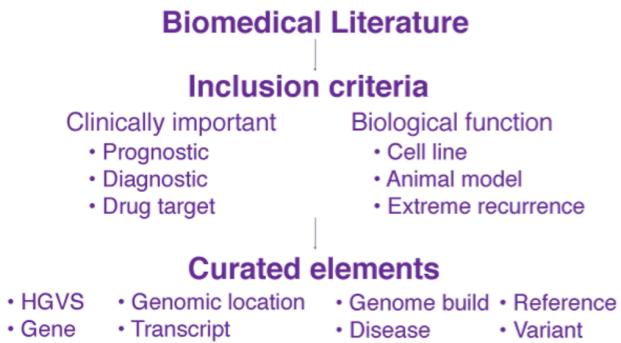
Alex Wagner

Griffith et al. Nature Methods. 2013.  
Wagner et al. Nucleic Acids Research. 2015.

# DoCM - Curated mutations can be browsed by web/API and used for knowledge-driven variant calling

## 1 Publication Curation

Publications are evaluated against the inclusion criteria and curated elements are added to the database



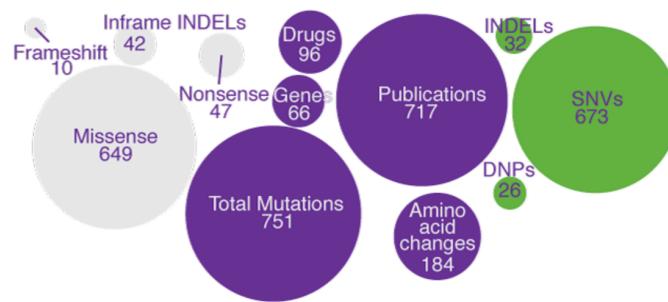
## 3 Web Application

Open source (MIT), openly licensed (CC) web application with API

A screenshot of the DoCM web application. The interface includes a navigation bar with links for About, News, Contact, Sources, API Documentation, and FAQ. On the left, there are filters for Chromosomes, Genes, Diseases, Mutation Types, Amino Acids, and Position. The main area is titled 'Variants' and shows a table of curated mutations. The table columns are HGVS, CHR, Gene, Amino Acid, and Mutation Type. The first few rows show mutations for KRAS genes at various positions, all categorized as missense.

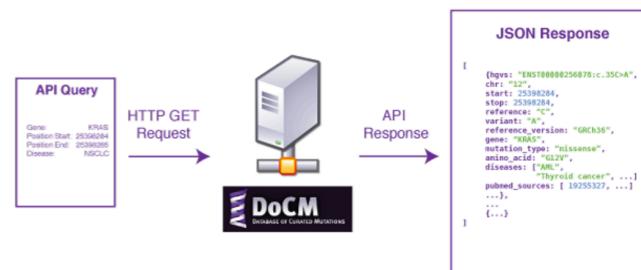
## 2 DoCM Contents

DoCM contains SNVs and INDELs across many genes and cancer subtypes with easy identification of the journal article that outlines the mutation's relevance.



## 4 API

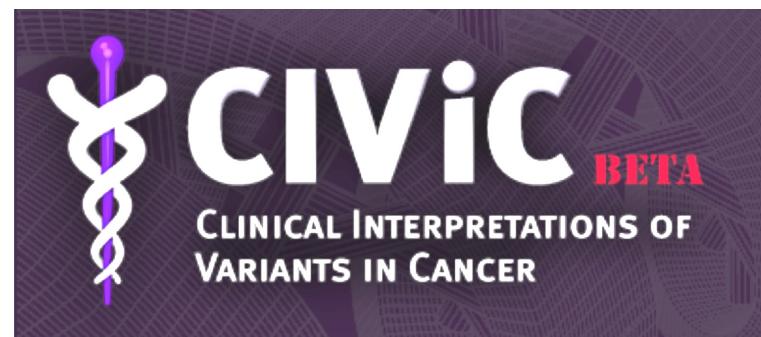
API queries can be made via HTTP requests and returns JSON response



# CIViC - a knowledgebase and forum to create interpretations of variants in cancer

# 55 Cancers 65 Genes 140 Variants

- CIViC currently holds ~500 evidence statements
    - From 397 published sources
    - Growing rapidly with large bulk imports underway to seed the system from other synergistic efforts



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

# Providing current and comprehensive interpretations of clinical significance of variants is a non-trivial task

GENOMIC ALTERATIONS	
GENE	INTERPRETATION
ALTERATION	
● PIK3CA H1047R	Mutations in PIK3CA have been reported in 26% to 33% of breast cancer cases (COSMIC, Jun 2012 and Kalinsky et al., 2009; 19671852). Activating mutations in PIK3CA, such as the one seen here, may predict sensitivity to inhibitors of PI3 kinase or its downstream signaling pathway (the PI3K/Akt/mTOR pathway) (Huang et al., 2007; 18079394). The mTOR inhibitors temsirolimus and everolimus have been tested in several clinical trials in breast cancer, and have been approved by the FDA for use in other tumor types. Inhibitors of PI3K and Akt are currently in clinical trials in breast cancer, alone or in combination with other therapies. PIK3CA mutations may play a role in resistance to hormonal therapy in ER+ breast cancers (Miller et al., 2011; 22114931). Activating mutations in PIK3CA may also confer resistance to anti-Her2 therapies (Chakrabarty et al., 2010; 20581867, Kataoka et al., 2010; 19633047, Wang et al., 2011; 21676217); combined inhibition of Her2 and the PI3K pathway may be required in tumors with ERBB2 amplification and PIK3CA mutation, though this remains an area of active investigation.
● CCND1 amplification	CCND1 amplification has been reported in approximately 10-15% of invasive breast cancers, more frequently in BRCA-negative cancers (Elsheikh et al., 2008; 17653856, Bane et al., 2011; 21327470). There are no approved therapies that directly target the protein product of CCND1 (Cyclin D1); however, CCND1 amplification may predict sensitivity to inhibitors of Cdk4 and Cdk6, which are currently under investigation in clinical trials. Overexpression of Cyclin D1 has also been associated with resistance to endocrine therapy in breast cancer (reviewed in Lange et al., 2011; 21613412; Musgrove and Sutherland, 2009; 19701242, Butt et al., 2005; 16113099).
● CDH1 E167*	CDH1 mutations are present in approximately 17% of breast cancers, and more often in luminal type cancers (COSMIC, Jun 2012, Hollestelle et al., 2010; 19593635). Loss of the E-cadherin protein, which is encoded by the CDH1 gene, has been associated with poor prognosis in triple negative breast cancer (Kashiwagi et al., 2010; 20551954, Tang et al., 2011; 21519872). Presently, there are no targeted therapies to address loss of CDH1/E-cadherin.

- Interpretations are typically produced by paid curators with no provenance and no mechanism for feedback
- I will argue that this effort would be better conducted in the public domain

# CIViC data model (clinical actionability evidence)

## CIViC - Clinical Interpretations of Variants in Cancer

### Clinical Actionability Evidence Item

#### Evidence Statement

Description of evidence from published medical literature detailing the association of or lack of association of a variant with diagnostic, prognostic or predictive value in relation to a specific disease

#### Clinical Significance

« Sensitivity, Resistance or Non-Response; Positive, Negative; Better Outcome, Poor Outcome »

**Rating**  
★☆☆☆☆

**Evidence Direction**  
« Supports, Does Not Support »

**Evidence Level**  
« Validated, Clinical, Preclinical, Case Study, Inferential »

**Variant Origin**  
« Germline, Somatic »

**Variant**  
« Variant ID »

**Evidence Type**  
« Predictive, Prognostic, Diagnostic »

**Gene**  
« Entrez ID »

**Source**  
« Pubmed ID »

**Disease**  
« Disease Ontology ID »

**Drug**  
« PubChem ID »

Entrez

PubMed

Disease Ontology

PubChem

#### Public Databases

<https://civic.genome.wustl.edu/#/help>

# Users create, revise and comment on evidence statements and synthesized into short (clinician friendly) human readable interpretations

NPM1 Variants:

EXON 12 MUTATIONS

W288FS

NPM1 Variant Groups:

NPM1 Exon 12

W288FS

## Variant EXON 12 MUTATIONS [Edit Variant](#)

[Variant Summary](#) [Variant Talk](#)

**Variant Summary:**

NPM1 exon 12 mutations are frequently identified in patients with cytogenetically normal acute myeloid leukemia (AML) and often co-occur with FLT3-ITD. FLT3 status should also be evaluated as co-occurrence with FLT3-ITD may impact prognosis. Exon 12 mutations have been identified as a predictor of good prognostic outcomes in the absence of FLT3-ITD. Due to their high frequency, NPM1 mutations have been retrospectively analyzed in the context of a number of therapies including variable results following ATRA treatment as well as improved response to high-dose daunorubicin or valproic acid. Additionally, multiple groups have shown increased surface expression of CD33 associated with NPM1 mutation, suggesting these patients may respond to anti-CD33 therapy. Cytoplasmic sequestration of NPM1 (NPM1c) is associated with a good response to induction therapy.

Supporting Evidence	Disease	Drug	Level ▲	Rating ▼
AML with mutated NPM1 is a provisional entity in WHO classification ...	Acute Myeloid L...	N/A	A	★★★★★
Complete remission rates were higher and event-free survival was lon...	Acute Myeloid L...	N/A	B	★★★★★
NPM1 mutations were not associated with the M2 FAB subtype of ac...	Acute Myeloid L...	N/A	B	★★★★★☆
NPM1 mutations were associated with M4, M5a and M5b FAB subtyp...	Acute Myeloid L...	N/A	B	★★★★☆☆
NPM1 mutation without FLT3-ITD was associated with reduced relap...	Acute Myeloid L...	N/A	B	★★★★☆☆
In young AML patients (<60 years old), DNMT3A mutation status was ...	Acute Myeloid L...	N/A	B	★★★★☆☆
NPM1 mutation was associated with higher complete remission rates...	Acute Myeloid L...	N/A	B	★★★★☆☆
NPM1 mutations were associated with increased complete remission...	Acute Myeloid L...	N/A	B	★★★★☆☆

1 - 8 of 36 items

1 / 5

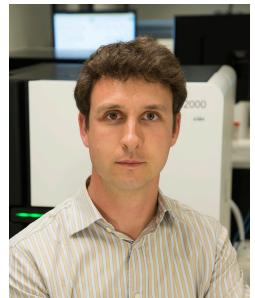
**Evidence ID: EID308 [Edit Evidence](#)** [Evidence Summary](#) [Evidence Talk](#)

AML with mutated NPM1 is a provisional entity in WHO classification of acute myeloid leukemia (AML). This mutation should be tested for in clinical trials and is recommended for testing in patients with cytogenetically normal AML.

# Every aspect of CIViC is freely available, open source, and open access

- The content of CIViC is released with minimal restrictions under a Creative Commons Attribute license (CC BY 4.0)
- No fees or exclusive access will be introduced
- The source code of CIViC is released with minimal restrictions under the MIT license
  - <https://github.com/genome/civic-server>
  - <https://github.com/genome/civic-client>
- Public instance versus running a local instance ...
  - [www.civicdb.org](http://www.civicdb.org)
  - <http://127.0.0.1:3001>

# Group members



Obi  
Griffith



Malachi  
Griffith

DoCM  
CIViC curator



Benjamin  
Ainscough

Breast cancer  
Regulome



Robert  
Lesurf

CIViC developer  
(front end)



Josh  
McMichael

OSCC  
CRC



Katie  
Campbell

genVisR  
HCC/SCLC



Zachary  
Skidmore

CIViC/DGIdb developer  
(back end)



Adam  
Coffman

CIViC lead  
curator



Jasreet  
Hundal



Kilanin  
Krysiak

DGIdb  
SCLC



Alex  
Wagner

Recent additions: Felicia Gomez (postdoc), Jason Kunisaki (thesis student), Cody Ramirez (rotation student), Matt Matlock (rotation student)

# 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. 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, 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. MacMil-lan, 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. Oberkfell, 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'Desha T'verie Scott, Harry E. Senaldi, Cherilyn 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, Eyanne Trevaskis, Richard Trippeer, Rusudan B. Turabelidze, Rahul Tyagi, Joelle M. Veizer, Tammi L. Vick-ery, Jason E. Waligorski, Jason R. Walker, Latricia E. Wallace, John W. Wallis, Qi Wang, Wesley C. Warren, Brian H. Watts, James V. Weible, Matthew R. Weil, George M. Weinstock, Michael C. Wendt, Richard K. Wilson, Roxanne M. Wilson, Aye M. Wollam, Dee Wu, Kristine M. Wylie, Todd N. Wylie, Kai Ye, Xu Zhang, Yanjiao Zhou

**Richard K. Wilson**  
**Elaine R. Mardis**  
**Timothy J. Ley**

**Li Ding**  
**Bob Fulton**  
**Lucinda Fulton**  
**Ira Hall**  
**Dave Larson**  
**Chris Maher**  
**Brad Ozenberger**  
**Lukas Wartman**

**Matt Callaway**  
**Jim Eldred**  
**Dan Koboldt**  
**Chris Miller**  
**Jason Walker**