UAMS

MYELOMA INSTITUTE FOR RESEARCH AND THERAPY

UNIVERSITY OF ARKANSAS FOR MEDICAL SCIENCES

# Revealing the Inherent Heterogeneity of Human Malignancies by Variant Consensus Strategies Coupled with Cancer Clonal Analysis

Erich A. Peterson, Shweta S. Chavan, Michael A. Bauer, Christoph Heuck, and Donald J. Johann Jr.

# Outline

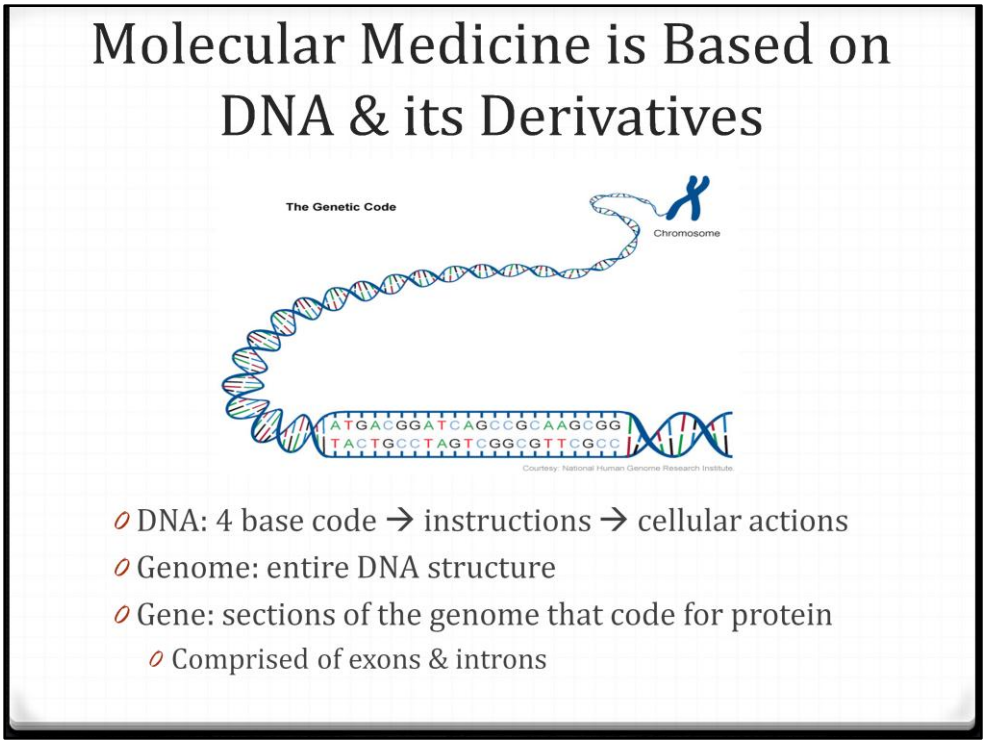
- Background
- Whole Exome Sequencing (WES) & Variant Consensus Analysis
- Tumor Clonality Analysis

# Biomedicine: Unprecedented Views of Our Molecular Make-up & Disease

*There are three things extremely hard, Steel, a Diamond, and to "know one's self"*

- Benjamin Franklin, Poor Richard's Almanac

- While Mr. Franklin was speaking about knowing one's self in the philosophical sense
  - Here we are talking about truly understanding our genetic make-up at the molecular level (the lowest meaningful level)
- Understanding disease in a way that was not previously possible.
- We have learned in the last 10-15 yrs. that cancer is a disease of the genome at the cellular level.
- We've now the technology to see genomic aberrations & derangements at the base level.




- Please excuse me if this is too elementary, but I just want everyone on the same page
- Contained within the chromosomes (proper) is the entirety of the genome (or our genetic code)
    - DNA is coded using a 4 base code (ATCG)
        - It is the DNA bases which code instructions
        - Instructions confer cellular action
    - 1% of the genome codes for proteins (we call these sections exons)

# Fruits of the Human Genome Project

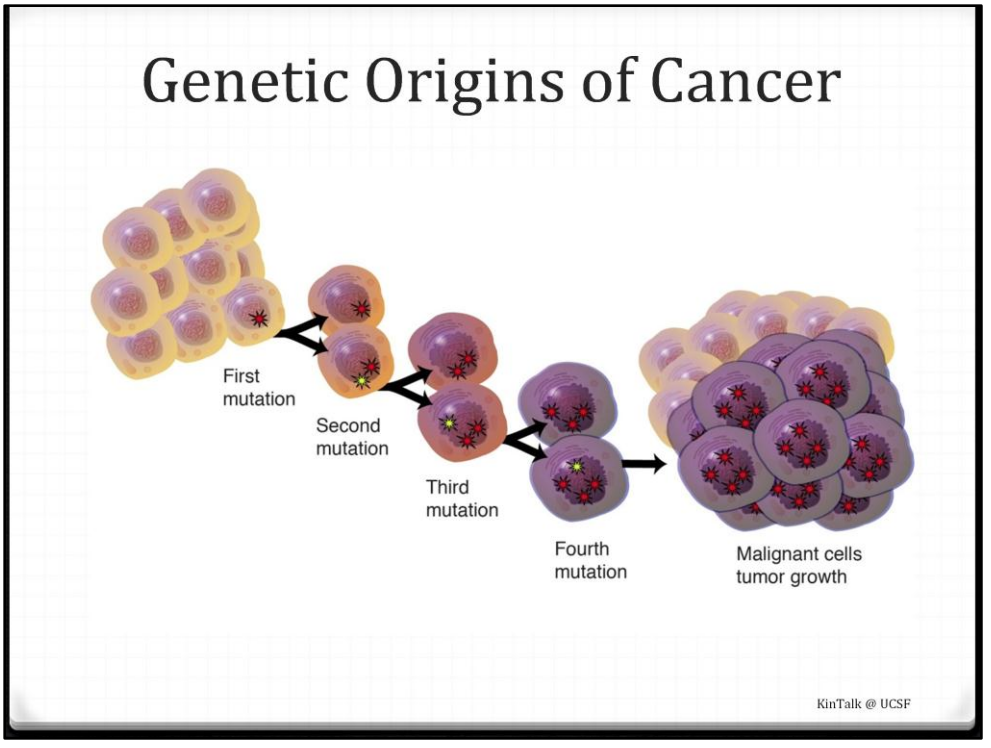
- Human Genome Project :
  - Near complete map of human genome

| Time | Cost |
|---|---|
| 1988 – 2003 (~15 yrs.) | $2.7 billion |

- Today:

| Time | Cost |
|---|---|
| 1 day | $1,000 |

- HGP yielded a nearly complete map of the human DNA sequence
- Illumina just announced a few days ago, WGS can be done in 1 day at a cost of $1K
- Because of the cheapness, it will soon enter clinical practice
- Oncology is seeing some of the biggest impacts:
  - In the form of new classes of therapeutics / drugs
  - Current treatments are far too toxic and we are finding tumors too late

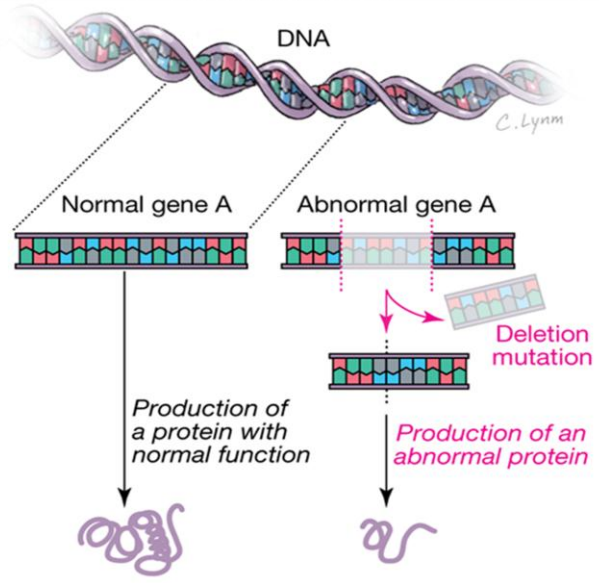



- This is just a graphic
- Cancer is a disease of the genome at the cellular level
- Each cancer has an entire genome in it, only it is mutated
    - That is, within a someone with cancer there exist two genomes; one from the tumor (somatic), the other he/she was born with (germline)
- Next-generation sequencing and its associated analyses can differentiate between somatic and germline mutations

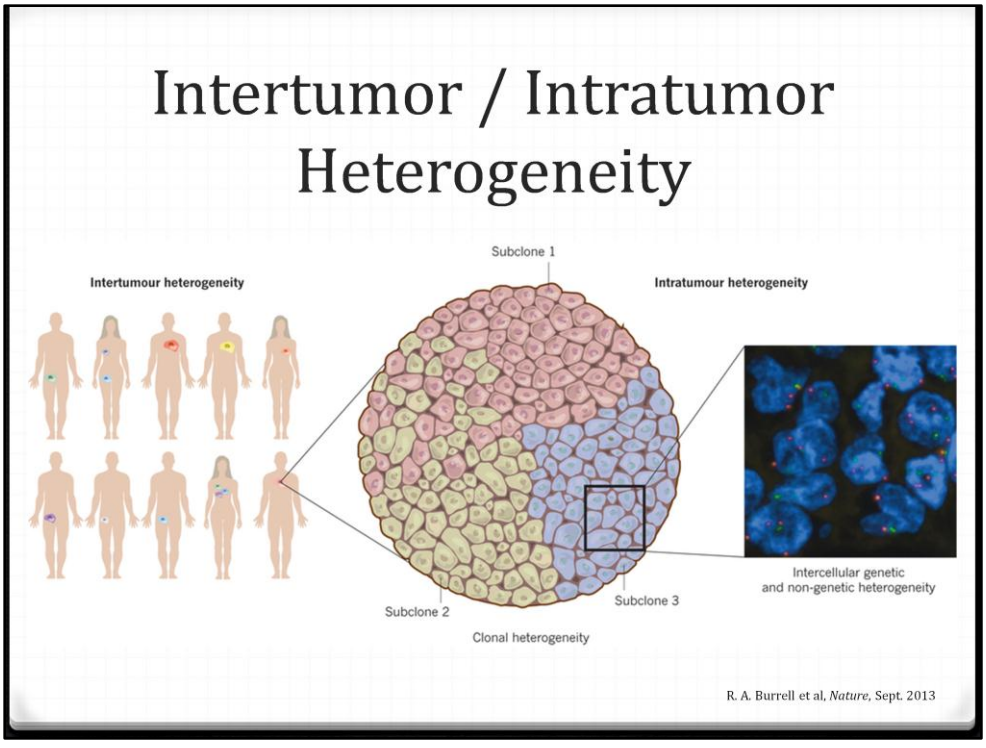
# Variability in the Human Genome

- For everyone here has a DNA sequence that is 99% identical
    - If we were all dogs, we would all be of the same breed
- However, it is the variability that defines the human population
- Underlies different prognoses & outcomes in disease states
- A primary reason why categorical therapeutic approaches do not always work (lack specificity)

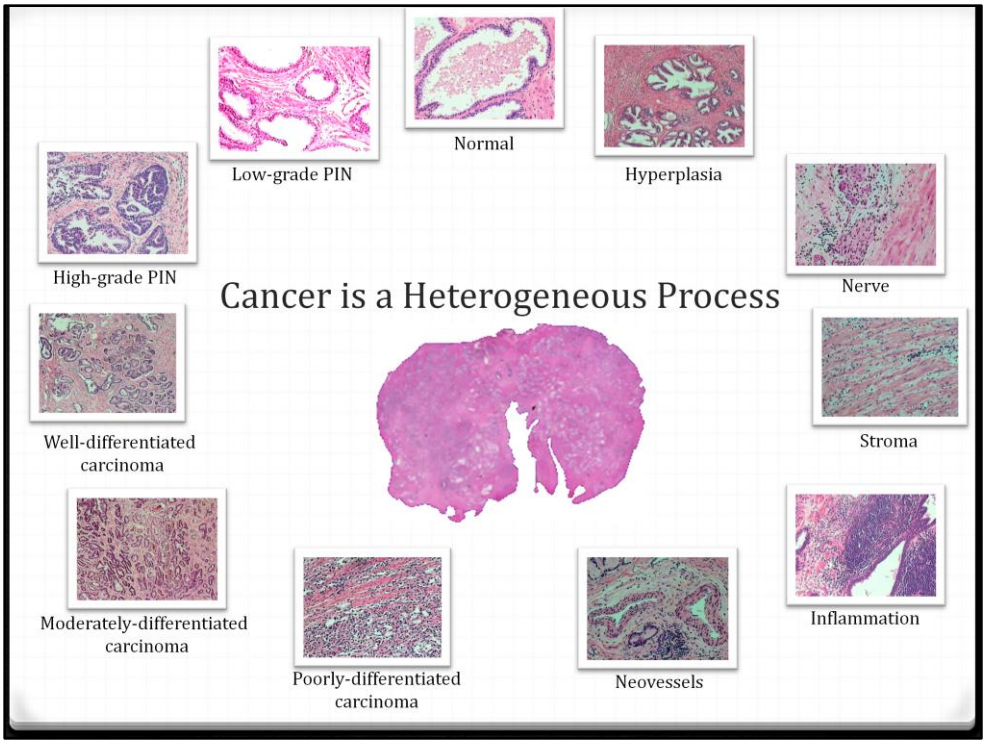
# Example of a DNA Deletion Mutation



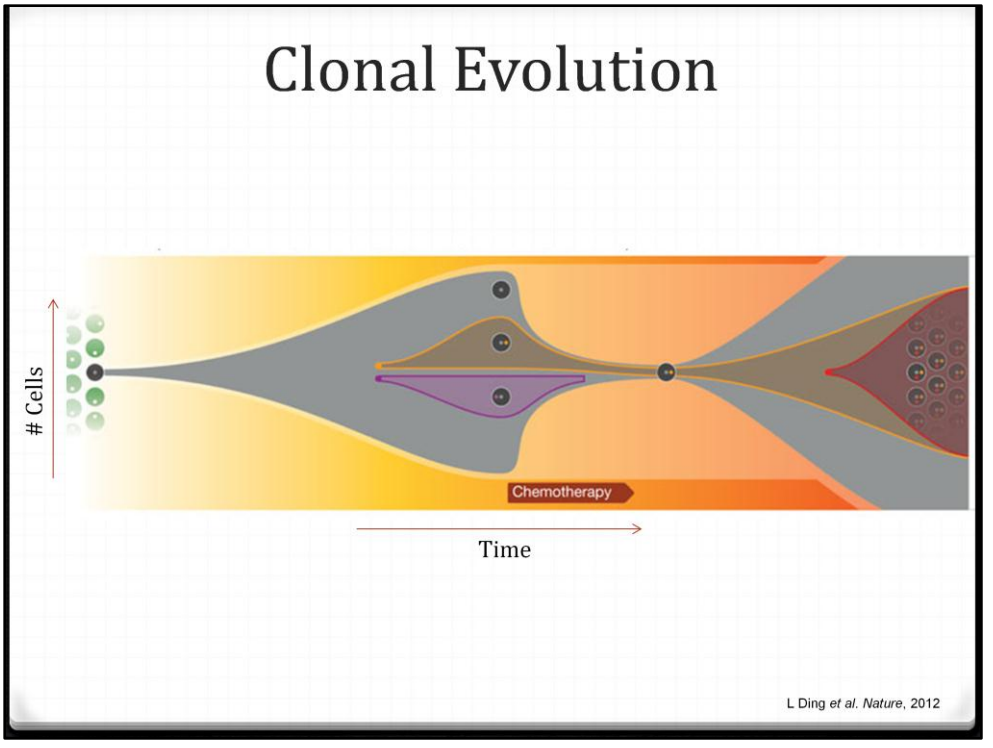

JAMA 2013




- This graphic illustrates the heterogeneity that exists not only between tumor, but within a particular person's tumor
- We know that not every tumor cell within a particular tumor are identical
- Tumors are a mosaic.




- Here we are looking at the cancer at a tissue and morphological level
- In the center, we see a tissue cross-section of a human prostate gland that is cancerous
- Around the periphery are the tissue subtypes within the cancerous prostate gland
    - “PIN” is not normal but not cancerous
    - “Differentiated carcinoma” show different levels of aggressiveness of the cancer

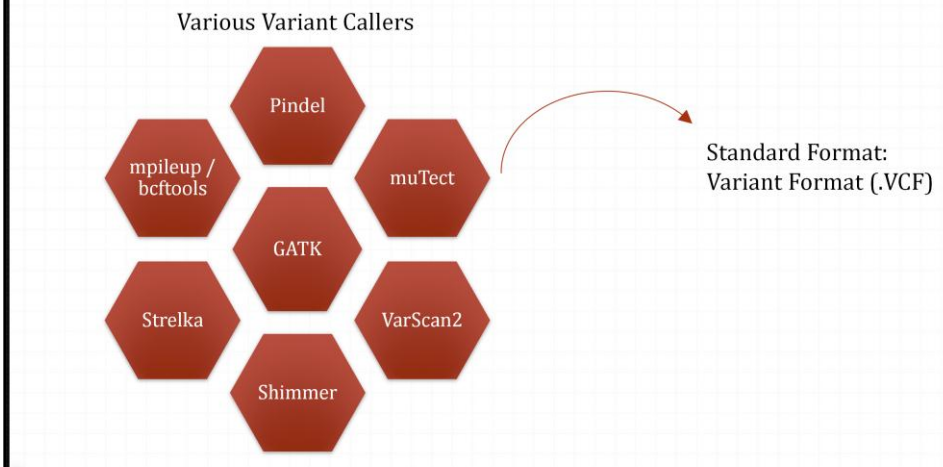



- Here we are looking at the clonal / molecular level
    - At the clonal evolution of cancerous cells
- X-axis shows time, Y-axis shows the expansion of mutated cells
- Over time we see that a so-called “founding clone”, that initiated the cancer, gains mutations over time
- At presentation, a sample of the patient’s tumor will appear as a cross-section
    - Exhibiting different abundances of various subclones
- During treatment a “selection event” occurs, some mutations are killed, others survive
- Some mutations might be a result of treatment (chemo) and ultimately proliferate and kill the patient

# Whole Exome Sequencing (WES) Methods

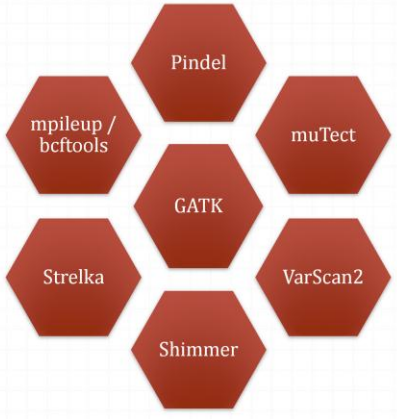
- Germline & Tumor DNA extracted from a Multiple Myeloma patient
- Whole Exome Sequencing (WES) was performed:
  - Illumina HiSeq 2500
- Paired End Alignment of Reads:
  - BWA / STAMPY
- Base recalibration with GATK
- Variant calling performed:
  - GATK, VarScan2, and Strelka
- VCF files used as input to Variant Consensus Reporter

# The Good


Various Variant Callers
Pindel
mpileup / bcftools
muTect
GATK
Strelka
VarScan2
Shimmer
Standard Format:
Variant Format (.VCF)

# The Bad

Various Variant Callers




Different Computation Algorithms

Vis-à-vis

Different Results

# One Solution: Consensus Analysis

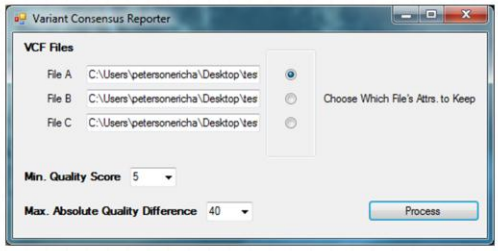
- Why: Common Approach used in Machine Learning
- Variant Consensus Reporter (VCR)
  - Input: Various .vcf files from variant callers
  - Output: Union and/or intersection of records in .vcfs
    - Output could be fed to functional annotators (i.e., ANNOVAR, snpEff)

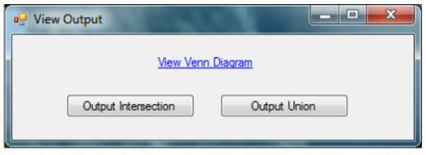
# Essential Contents of a VCF File

- Chromosome
- Position within chromosome
- Reference base(s)
- Alternate base(s)
- Quality score

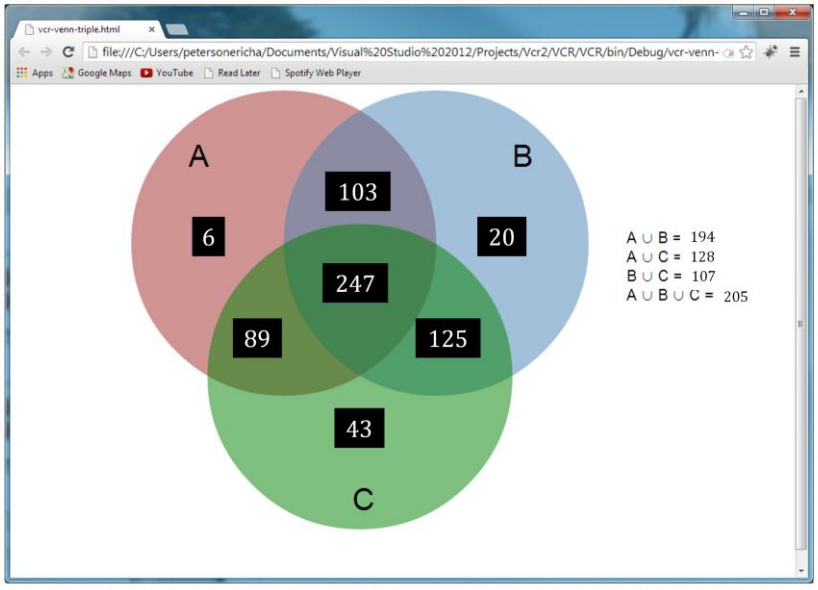
- Chr 1
- 101550
- C
- AT
- 40

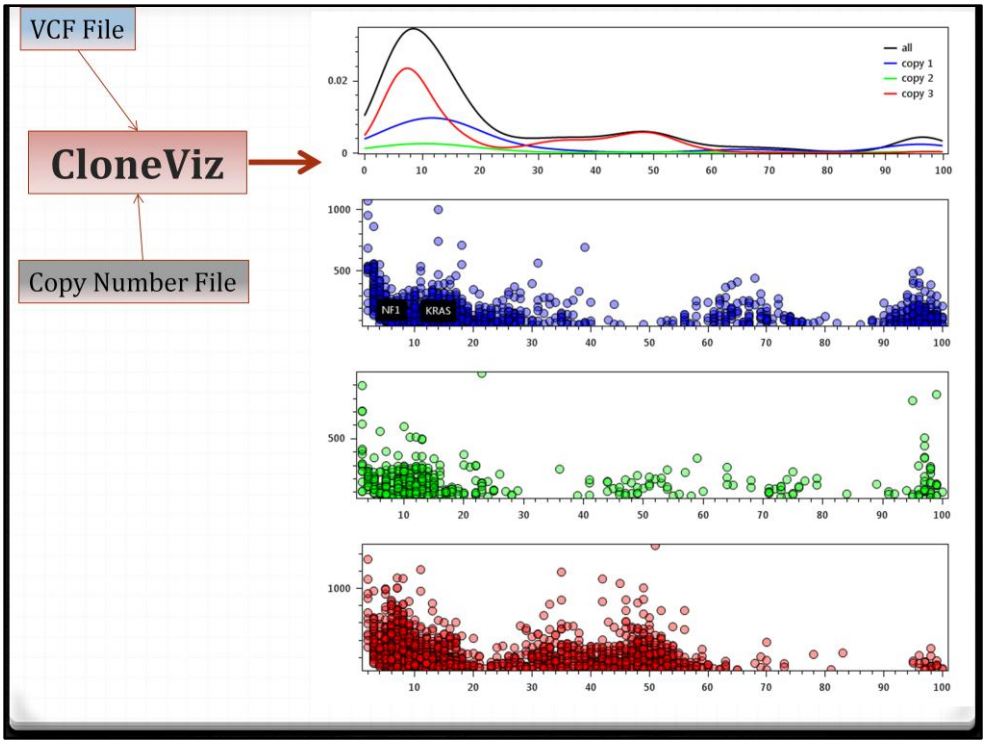
# Interactive User Interface



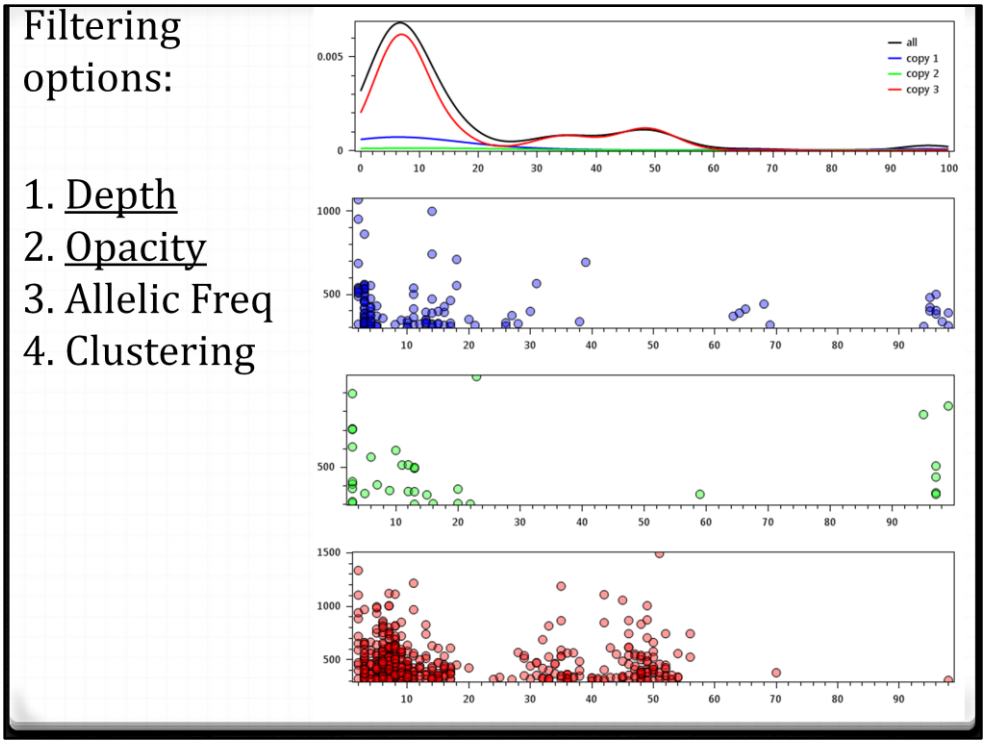



# Automatic Graphical Results

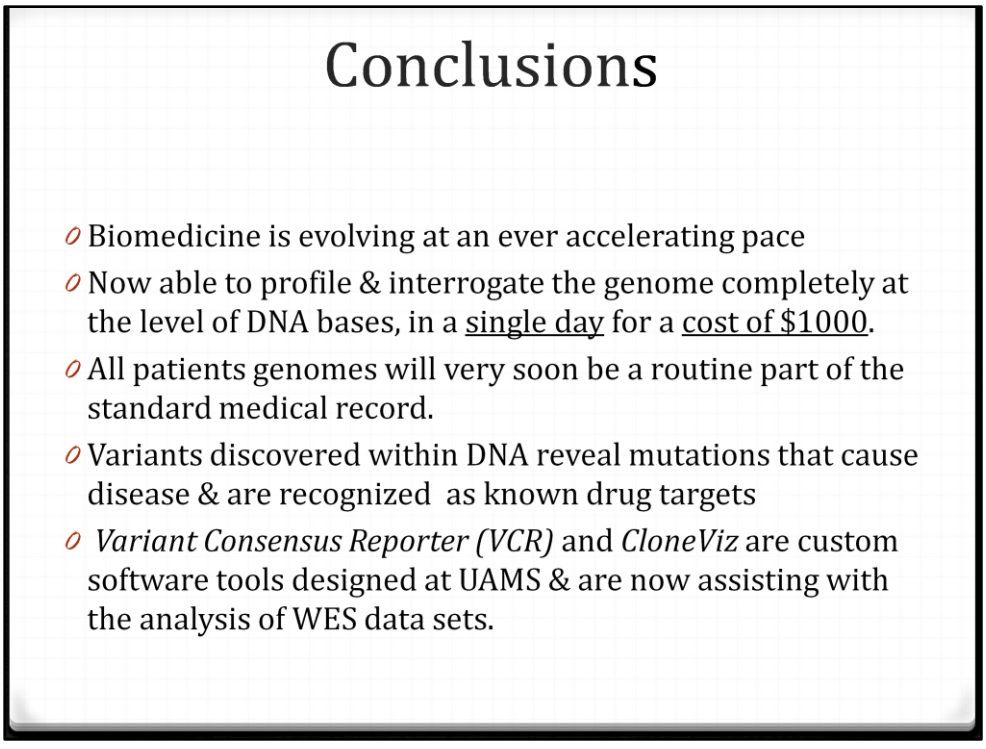





- X-axis in all plots is the variant allele frequency
    - It is basically the relative abundance of the variant within the sample
- Y-axis is depth in the copy number plots and density in the top plot
- Each dot is a variant / mutation found within the sample
- We can label variants in found in certain genes, by providing a custom gene list we are interested in annotating




- Filtered on minimum and maximum depth
- Everything is re-calculated and displayed interactively
- Future work:
    - Filtering on allelic freq.
    - Clustering of subclones




- Biomedicine is evolving at an ever accelerating pace
    - Human Genome Project established the foundation of modern molecular medicine.
    - Subsequent fruits of this public-private partnership continue to advance technology that is now changing the practice of medicine, especially cancer.
- Now able to profile & interrogate the genome completely at the level of DNA bases, in a single day for a cost of $1000.
- All patients genomes will very soon be a routine part of the standard medical record.
- Variants discovered within DNA reveal mutations that cause disease & are recognized as known drug targets
- *Variant Consensus Reporter (VCR)* and *CloneViz* are custom software tools designed at UAMS & are now assisting with the analysis of WES data sets.

# Acknowledgments

Collaborators

Shweta S. Chavan

Michael A. Bauer

Christoph Heuck

Donald J. Johann Jr.

MIRT Director

Bart Barlogie