UAMS

MYELOMA INSTITUTE FOR RESEARCH AND THERAPY

UNIVERSITY OF ARKANSAS FOR MEDICAL SCIENCES

Variant Consensus Reporter:

Increasing the Confidence of Variants in Whole Exome Sequencing via a Consensus Approach

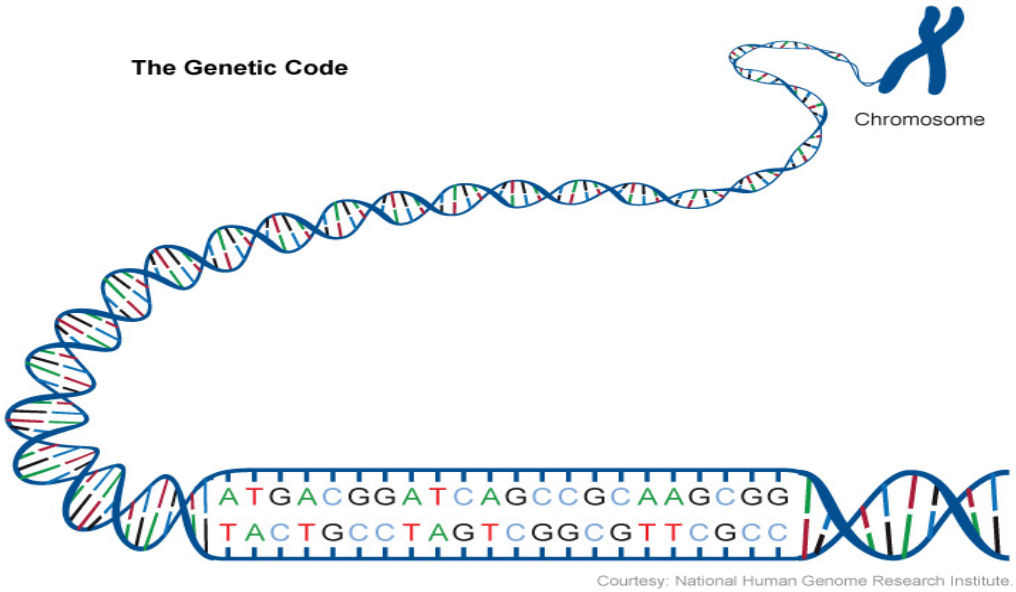
Erich A. Peterson, Shweta S. Chavan,
Michael A. Bauer, Christoph Heuck,
and Donald J. Johann Jr.

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

Molecular Medicine is Based on DNA & its Derivatives



- DNA: 4 base code → instructions → cellular actions
- Genome: entire DNA structure
- Gene: sections of the genome that code for protein

Fruits of the Human Genome Project

- Human Genome Project:

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

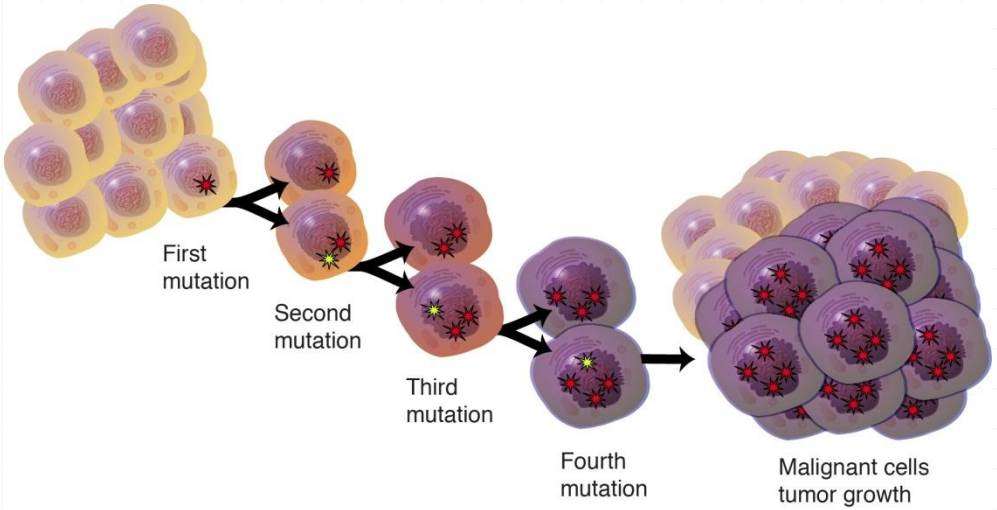
  - Yielding a nearly complete map of the DNA sequence

- Today:

Time	Cost
~ few days	< $10K

  - Oncology is seeing some of the biggest impacts
    - Advanced molecular diagnostics – will show an example
    - New classes of therapeutics / drugs

Genetic Origins of Cancer



KinTalk @ UCSF

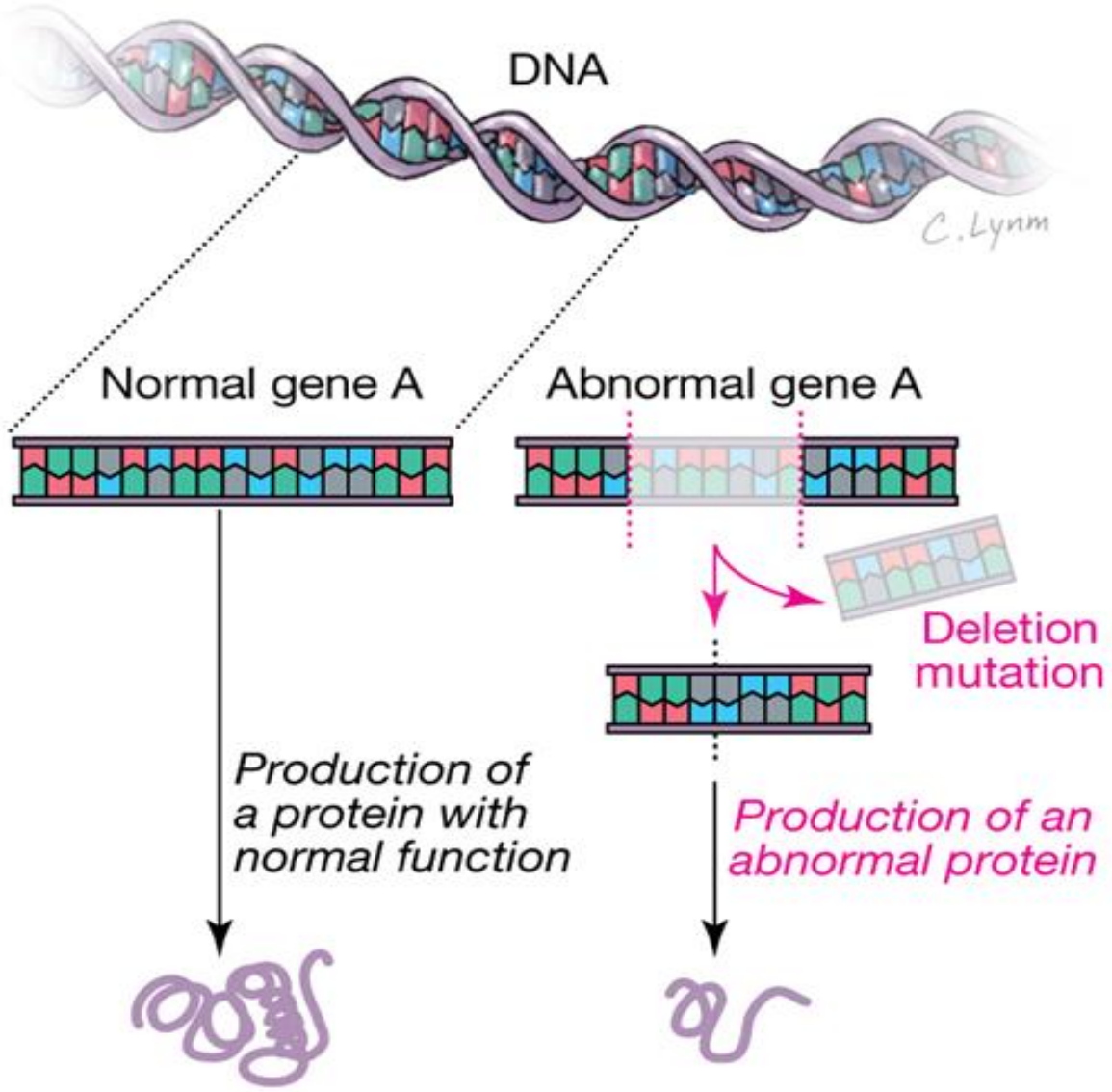
- Cancer is a disease of the genome at the cellular level
- Each cancer has an entire genome in it, only it is mutated
- Personalized or Genomic Medicine – Linking the best drug for a particular clinical condition based on your genetic make-up

Cancer is a Disease of the Genome

- DNA & Spelling
  - In the world of DNA spelling counts!
  - Even a single base substitution can have deleterious effects on protein function

- Genomics
  - What are the causes of disease?
  - How to better threat and intervene?
  - Molecular medicine based on your genetic code

Example of a DNA Deletion Mutation



JAMA 2013

Types of Variants in DNA

- Variant
  - Differences from the "normal" / reference genome
  - If present, is it meaningful? (possible cause of disease)

Variant Type	Example
SNP (Single-Nucleotide Polymorphism)	A → C
Insertion	T → GTC
Deletion	GTC → T
MNP (Multiple-Nucleotide Polymorphism)	CCTG → TGCT

Tumor DNA targeted re-sequencing assay "Foundation One" commercial assay

THERAPEUTIC IMPLICATIONS

Genomic Alterations Detected	FDA Approved Therapies (in patient's tumor type)	FDA Approved Therapies (in another tumor type)	Potential Clinical Trials
KRAS Q61H	None	Trametinib	Yes, see clinical trials section
MTOR V2006L	None	Temsirolimus Everolimus	Yes, see clinical trials section
TSC1 A944T	None	Everolimus Temsirolimus	Yes, see clinical trials section
TET2 Q888*	None	None	Yes, see clinical trials section
LRP1B loss	None	None	None

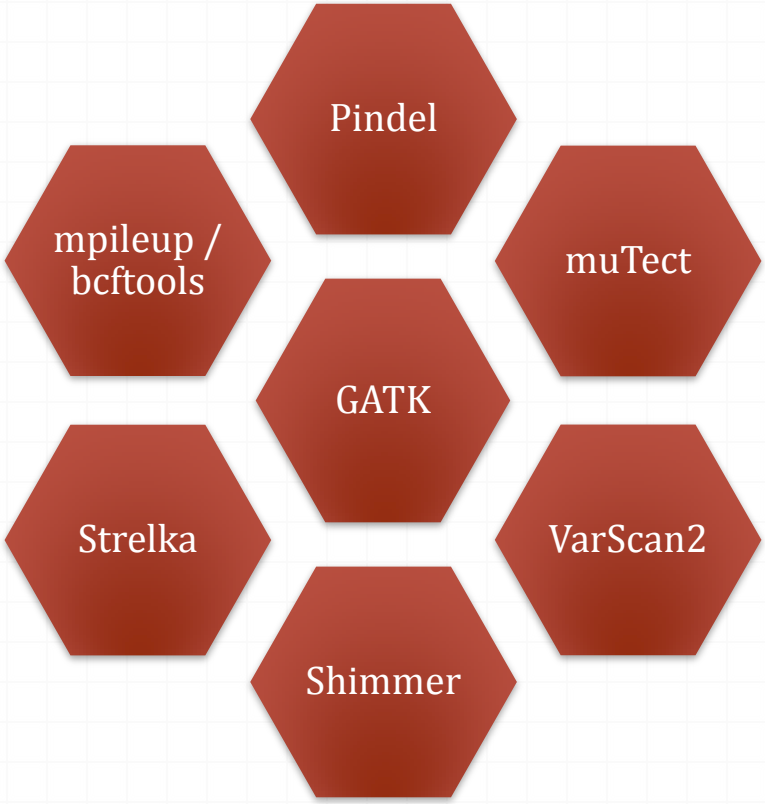
Sample source: FFPE tumor block, 10 slices ~5µm thick, macro-dissected, no germ line material
Instrumentation: Illumina HiSeq 2500
Time: 2 wks
Cost: $5800
Business model: Information company

Whole Exome Sequencing (WES) Methods

- DNA extracted from Multiple Myeloma (MM) cell lines
  - RPMI-8226 and U266
- Whole Exome Sequencing (WES) was performed:
  - Illumina HiSeq 2500
- Alignment of short reads:
  - BWA / STAMPY
- Variant calling performed:
  - GATK, VarScan2, and mpileup
- VCF files used as input to Variant Consensus Reporter

The Good

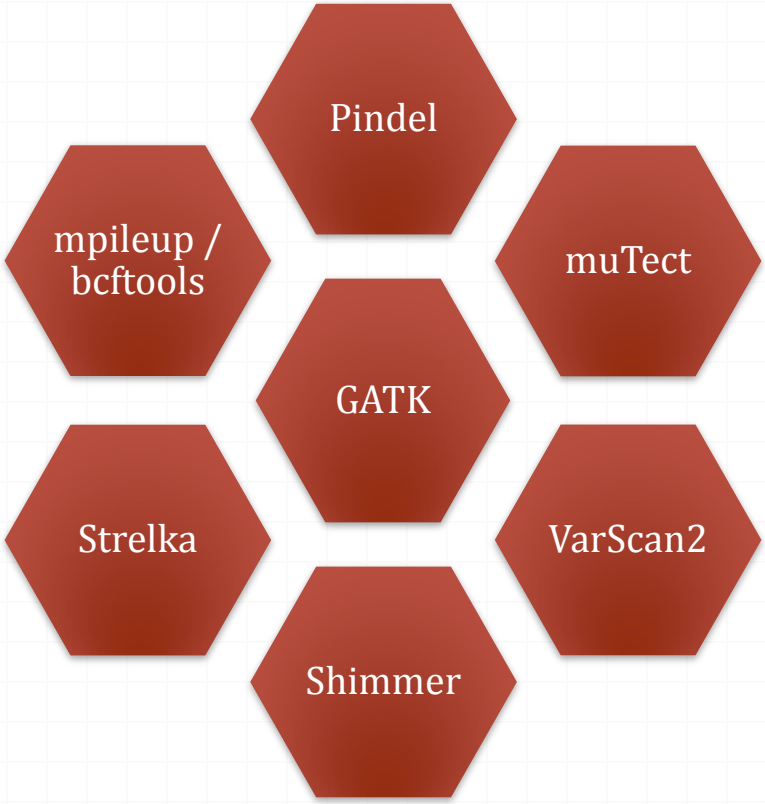
Various Variant Callers



Standard Format:
Variant Format (.VCF)

The Bad

Various Variant Callers



Different Computation Algorithms

Vis-à-vis

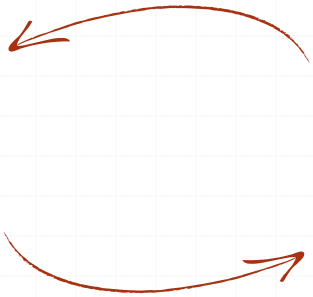
Different Results

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)

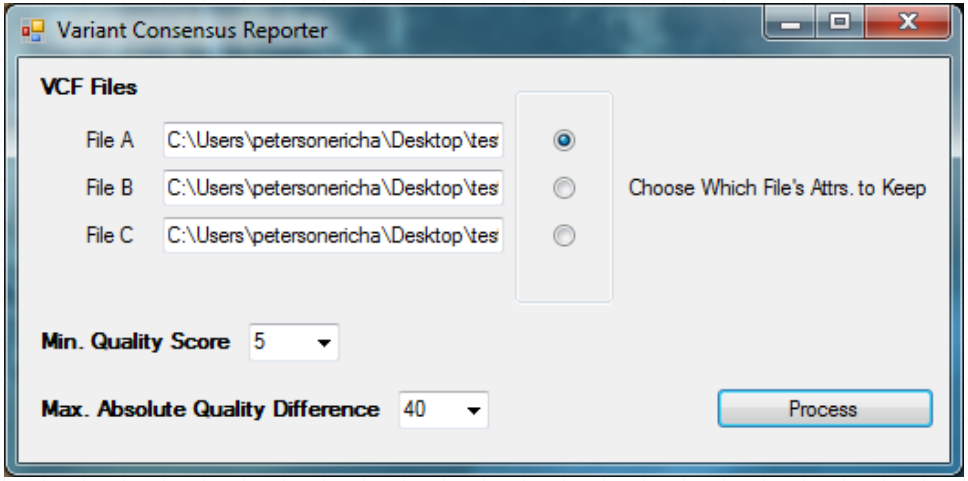
Essential Contents of a VCF File

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

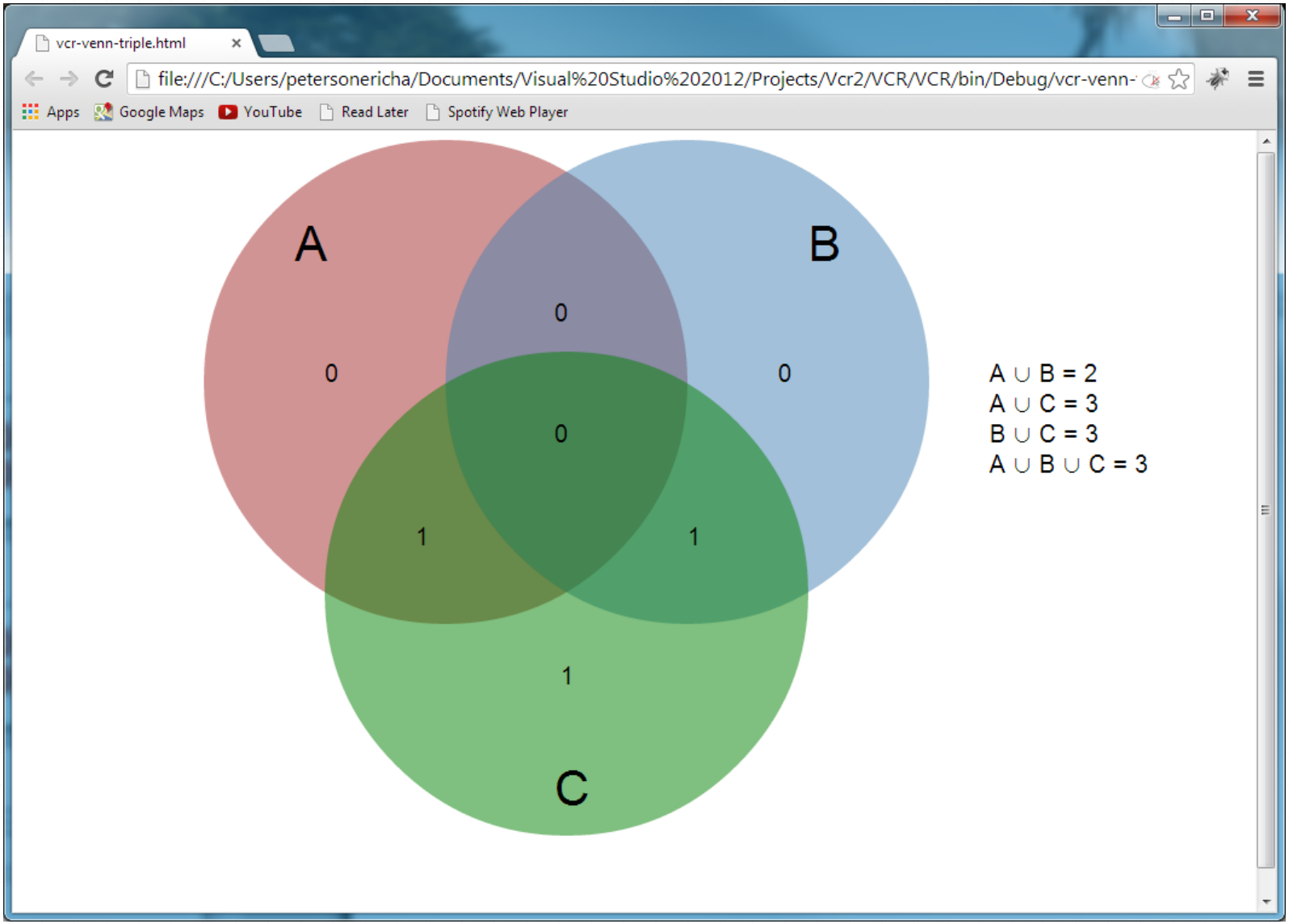


- Chr 1
- 101550
- C
- AT
- 40

Interactive User Interface



Automatic Graphical Results



Conclusion

- Biomedicine is evolving at an ever accelerating pace
  - Human Genome Project was a major precursor
- Now able to interrogate the genome at the molecular level
- Variants discovered within DNA helps elucidate mutations which might cause disease
- Variant Consensus Reporter (VCR) is a custom software tool designed at UAMS
  - Can help to gain more confidence in the variant calls of various variant callers

Acknowledgments

Collaborators

Shweta S. Chavan

Michael A. Bauer

Christoph Heuck

Donald J. Johann Jr.

MIRT Director

Bart Barlogie