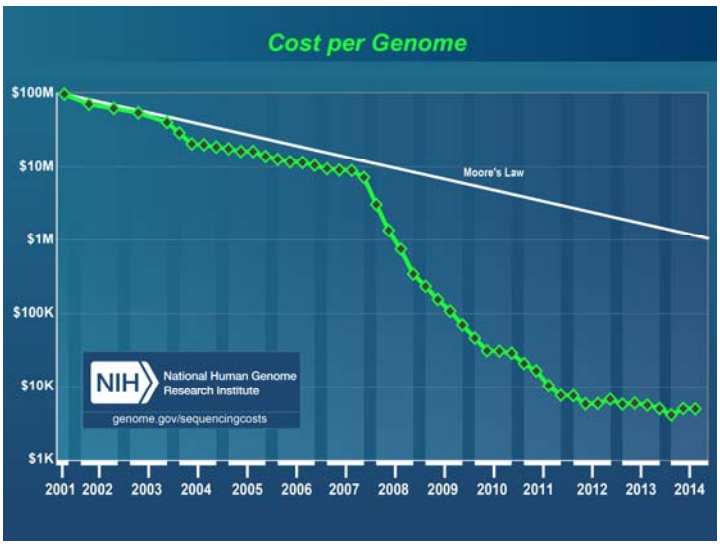
Genome Sequencing for Diagnosing and Predicting Human Disease Risk

Michael Snyder
Stanford University

October 31, 2016

Conflicts: Personalis, Genapsys, SensOmics

The Cost of DNA Sequencing is Dropping



Human Genome Cost <$2K

http://www.genome.gov/

Impact of Genomics on Medicine

- **Understand and Treat Disease**
  - Cancer
  - Mystery diseases
  - Prenatal diagnostics

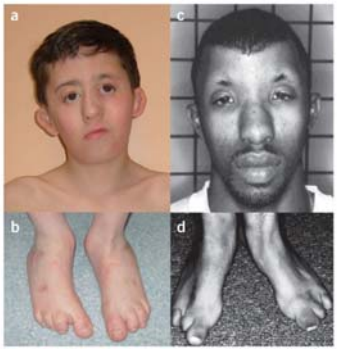
- **Pharmacogenomics**
  - Determining which drug side effects and doses
- **Managing Health Care in Healthy Individuals?**

Topics Covered

1) Solving Mendelian and Undiagnosed Diseases
2) Genome Sequencing for the “Healthy person”
3) Analysis of Noncoding regions

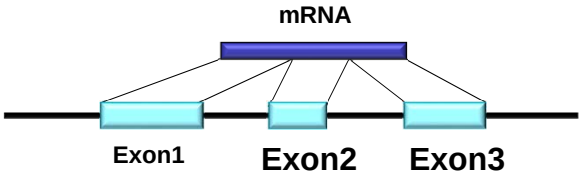
Undiagnosed Mystery Diseases

- 0.4% of live births
- 8% of adults have genetic disorder recognized by adulthood
- 25 M US Citizens
- $5M/individual/lifetime

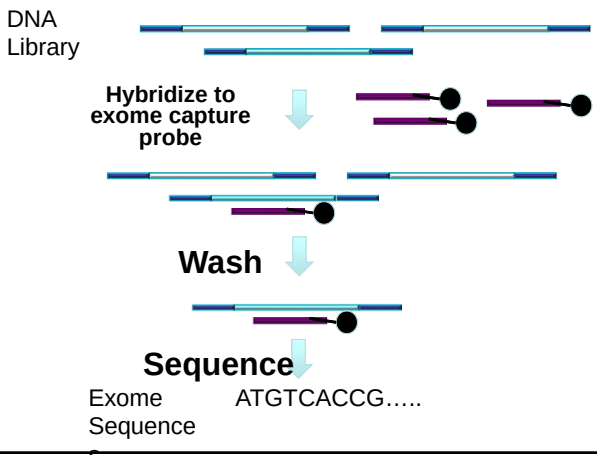


Ng et al., 2010 Nat. Genetics

Exome = 1.5% of Human DNA



Exome Capture & Sequencing



Lessons Learned

1) Overall success rate for identifying causative mutations is ~30%

2) Information not always directly actionable but still valuable.

3) Best success with
   a) Specific phenotypes
   b) Large families

4) Need large database to share information: Recurrence is key.

Example

- 14 yr old Patient
- hypogammaglobulinemia, mild clinical immune phenotype and growth hormone deficiency
- Exome sequencing of parents and child → c.2596A>C (p.S866R) in NFKappaB2 gene
- Similar to that described by Quentien et al.

Quentien MH, et al. "Deficit in anterior pituitary function and variable immune deficiency (DAVID) in children presenting with adrenocorticotropin deficiency and severe infections". J Clin Endocrinol Metab. 2012 Jan;97(1):E121-8.

Solving Mystery Diseases: Child With Variety of Conditions

Developmentally Delayed, Significant Health Issues

Father + **Mother** → **Child**

Candidate Mutations:
TCP10L2, SUPV3L1, PIEZO1, TMIE RPGRIP1L, FANCA, WFS1, NGLY1

Match!

Second Family → **Child**

NGLY1

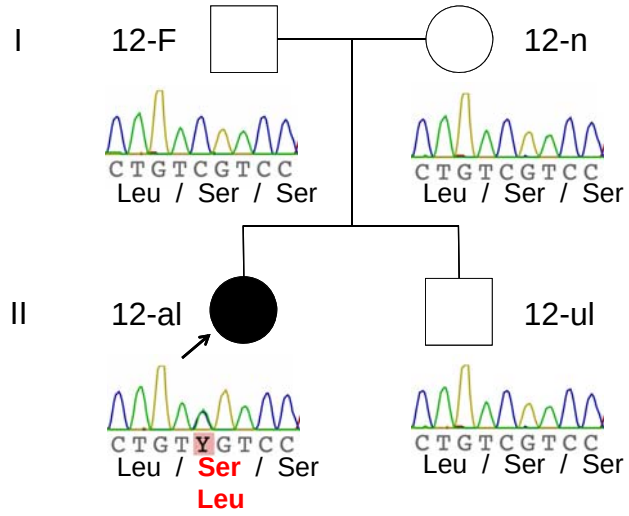
Enns et al 2014

Inflammatory Bowel Disease ~26 yr old

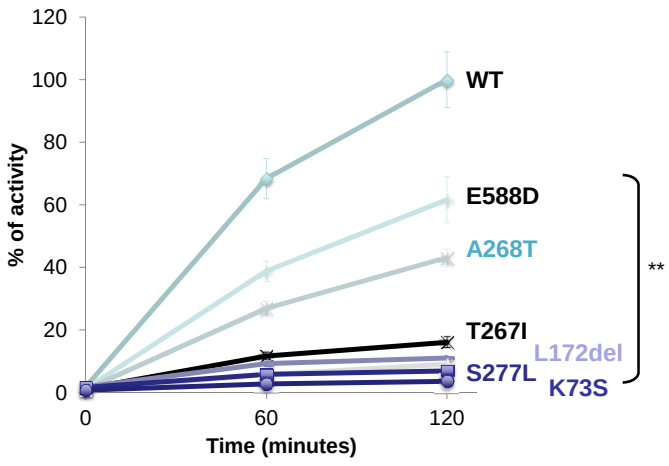
- Unaffected Mother, Father, Brother
- Affected daughter
- Exome sequenced the family

- List of about ~10 candidate genes

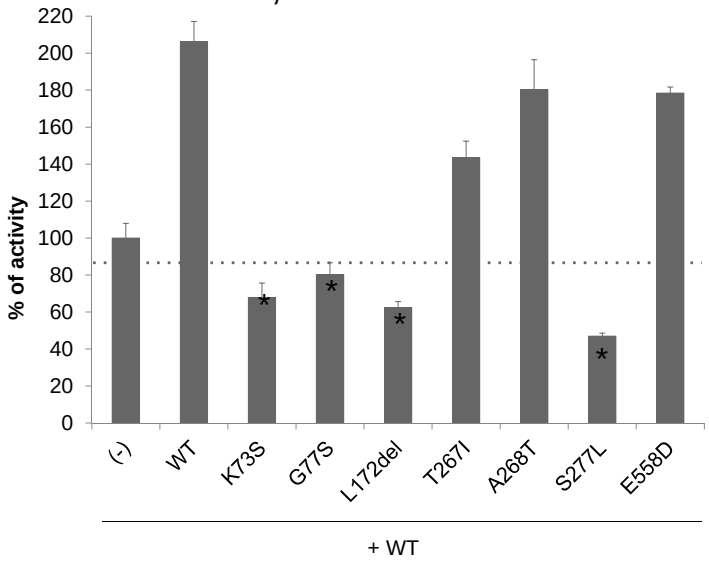
Pedigree and Sanger Traces of Family A (Heat Shock Protein Gene)



Effects of the HSP variants on HSP70/40 mediated-refolding heat denatured luciferase



Dominant negative effects of G77S, L172del, and S277L on HSP

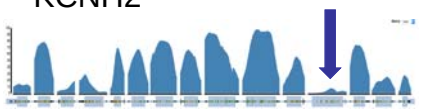


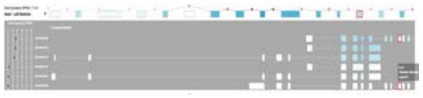
Coverage of *KCNH2*: Standard Exome vs. ACE Clinical Exome

Rachel Wood and Euan Ashley

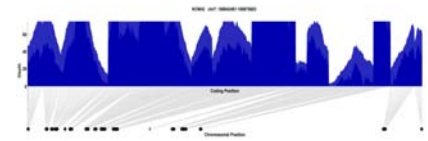
19 pathogenic or likely pathogenic variants in Clinvar

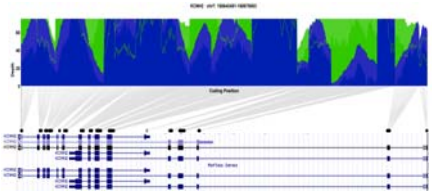
KCNH2





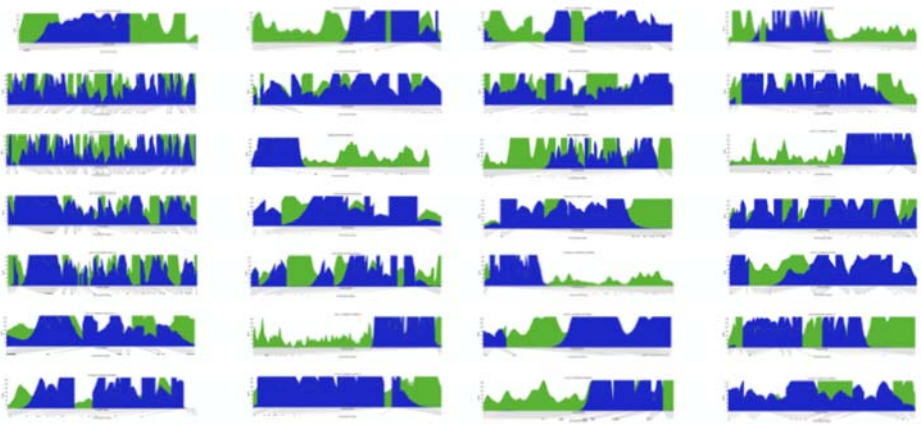
exac.broadinstitute.org





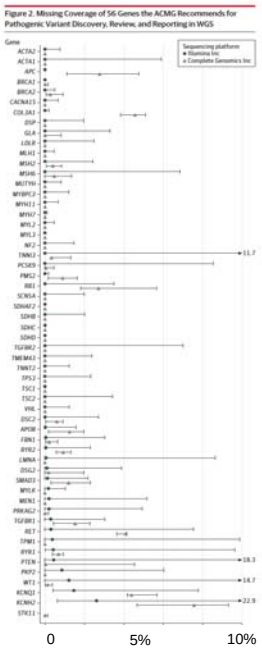
Augmented exome

Many Medically Relevant Genes Have Poor Coverage



Michael Clark PhD

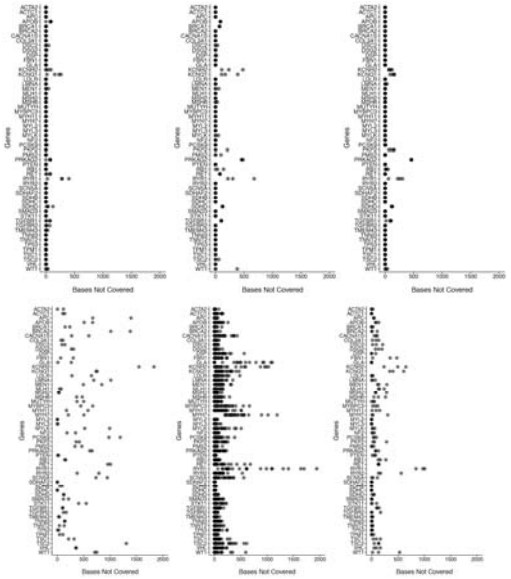
How well does 30x “clinical” whole genome sequencing perform?



Percentage of gene not callable

JAMA 2014 311:1035-44 2014

Augmented exomes outperform 30x whole genome sequencing



ACMG 56 most "actionable" genes
x axis = bases not covered at clinical grade

Augmented exomes

Left to right
- Company A
- Company B
- Company C

Left to right
- Off the shelf exome
- Hiseq genome
- X10 genome

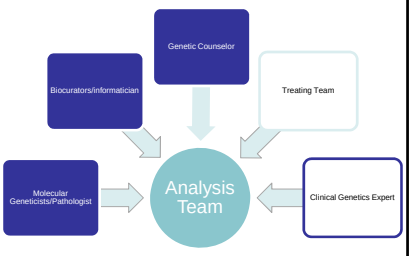
PSB 2015
precision.fda.gov

precisionFDA

Stanford Medicine Clinical Genomics Service

- Run by Jason Merker (Pathology) and Euan Ashley (Medicine)
- Use genome sequencing to determine the cause of disease in patients with suspected genetic disease
- Focused on several major disease areas:
  - Childhood mystery diseases
  - Familial cancer
  - Familial heart disease

Workflow for Clinical Exome Sequencing
1 – Test ordering & review

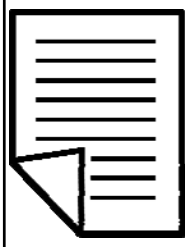


Test request by Stanford health care provider

Review by Genetic Test Optimization Service

Establish questions being posed by patient and treating team

Workflow for Clinical Exome Sequencing
2 – Counseling, consent & specimen collection



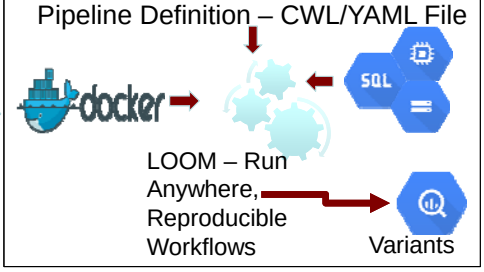
Insurance authorization

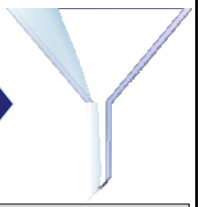
Clinical counseling & consent
- **Molecular diagnostic yield**
- **Incidental/Secondary findings**
- **Biobanking and data sharing**

Specimen collection
- **Genome sequencing**
- **Specimen ID/confirmatory studies**
- **Biobanking***

Workflow for Clinical Exome Sequencing
3 – Sequencing and data analysis

Pipeline Definition – CWL/YAML File



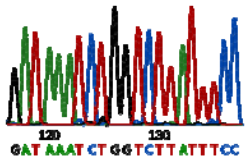


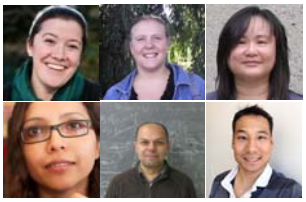
Exome sequencing (Illumina chemistry)

Data analysis (MedGAP)
- **Alignment**
- **SNV/indel calling**
- **QM program**
- **ID checks to confirm specimen identity**

Annotation, filtration & prioritization
- **Phenotype**
- **Variant impact**
- **Inheritance**
- **Pop freq**
- **ClinVar**

Workflow for Clinical Exome Sequencing
4 – Verification and report generation

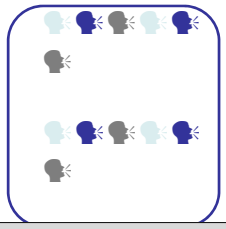




Genomics Review Groups
- **Genomics Service**
- **Treating Team**
- **Clinical genetics expert**

Variant verification by orthogonal method

Curation meeting & draft variant classification
- **Genetic counselors**
- **Biocurators**
- **Bioinformaticians**
- **Molecular pathologist/geneticists**



Genomics Review Board
- **Additional genetics expertise**
- **Bioethics/legal**

Workflow for Clinical Exome Sequencing
5 – Reporting & post-test counseling

Generate final report
- **Coverage metrics of relevant genes**

Patient meets with genetic counselor and relevant members of treatment team

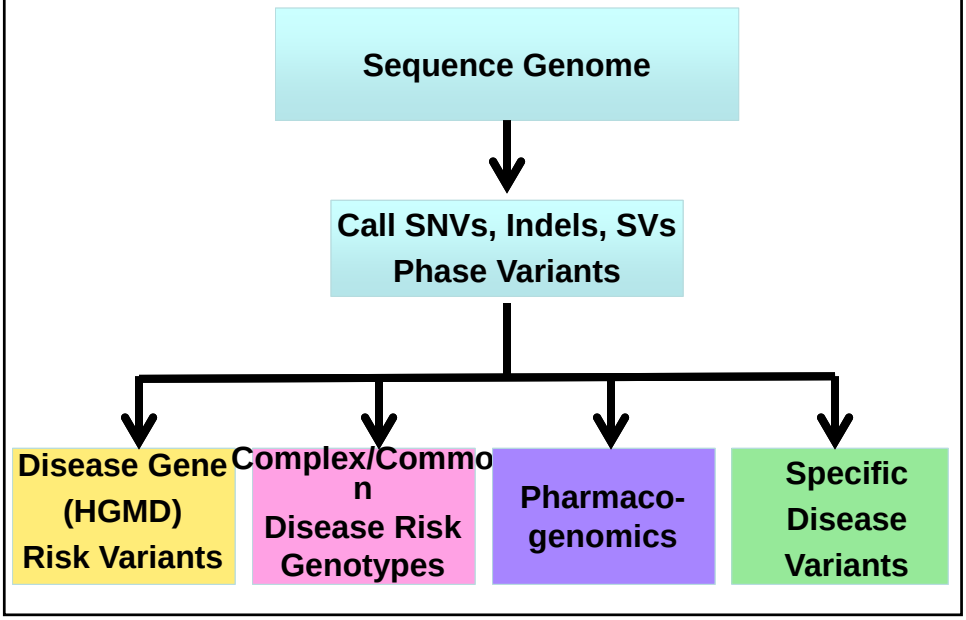
Reanalysis upon request

Impact of Genomics on Medicine

- **Understand and Treat Disease**
  - Cancer
  - Mystery diseases
  - Prenatal diagnostics

- **Pharmacogenomics**
  - Determining which drug side effects and doses
- **Managing Health Care in Healthy Individuals?**

Genome Interpretation Pipeline



Genome Analysis of 12 Healthy People

Dewey, Grove, Pan, Ashley, Quertermous et al JAMA 2014

Ethnicity:
7 Asians
5 Europeans

Sequence genomes with Illumina (all 12;
Mean depth: 50X (38-62)

9 also sequenced with Complete Genomics

Inherited Disease Risk and Carrier Status

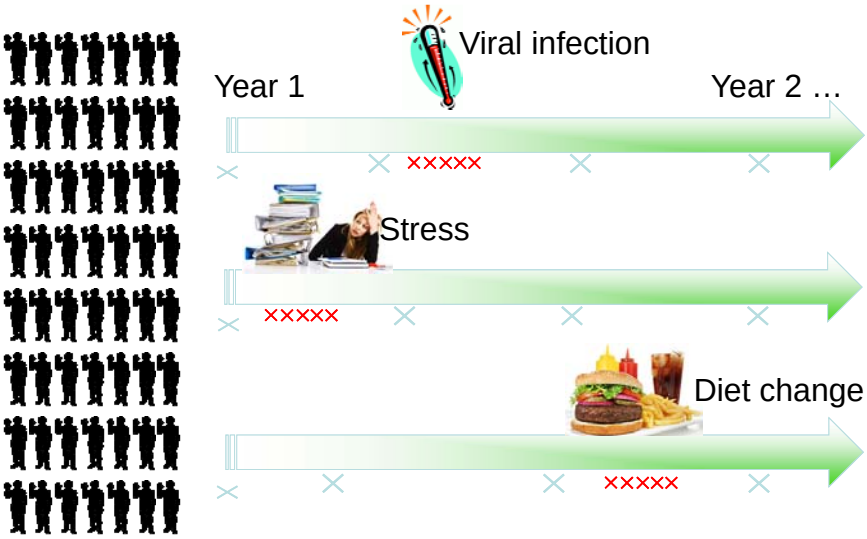
	# Variants Per Subject Median (Range)
Candidate Variants Manually Curated	108 (90-127)
- Previously reported or potential pathogenic variants in ACMG genes	3 (1-7)
Reportable variants associated with disease risk (HGMD)	5 (2-6)
- Reported disease-associated variants	0 (0-2)
- Rare expected pathogenic variants	0 (0-1)
- Genetic variants of unknown significance	3 (1-6)
Reportable variants associated with carrier status	13 (8-18)
- Reported disease-associated variants	2 (0-4)
- Rare expected pathogenic variants	2 (1-4)
- Genetic variants of unknown significance	9 (4-12)

Study of 12 Healthy People

Dewey, Grove, Pan, Ashley, Quertermous et al

- 3 followup diagnostic tests (range 0-10)
  - Cost ~$400-$1400 per individual (median $663-$773
- 54 minutes per variants
- One individual had a BRCA1 nonsense mutation—no known family history

Longitudinal Profilng of 100 individuals over periods of health, stress and disease

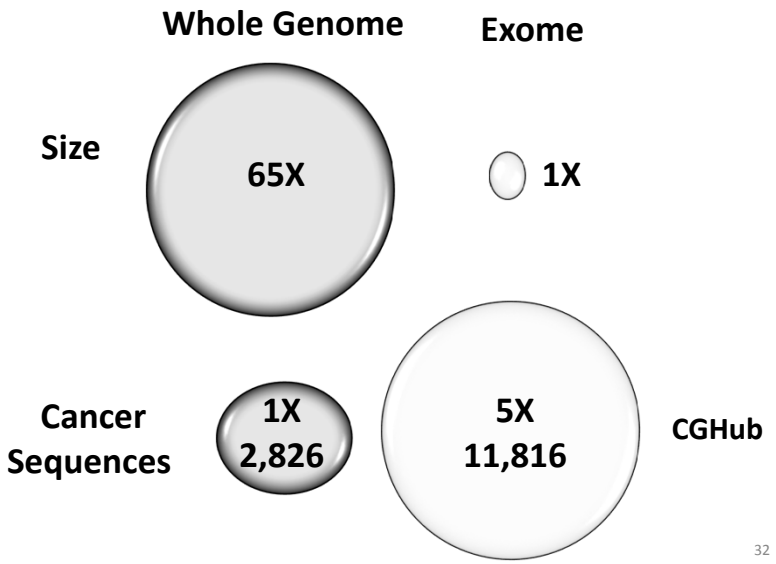


Genome Sequencing – First ~50 People

- Three have important mutations to know about
  - SHBD (2X): high freq. of paraganglioma
  - PROC (2X): Affects coagulation
  - RBM20: cardiomyopathy
  - One MODY mutation
- All have carrier mutations and pharmacogenetic variants

NonCoding Variants

Most Effort on Cancer is Performed on Exomes



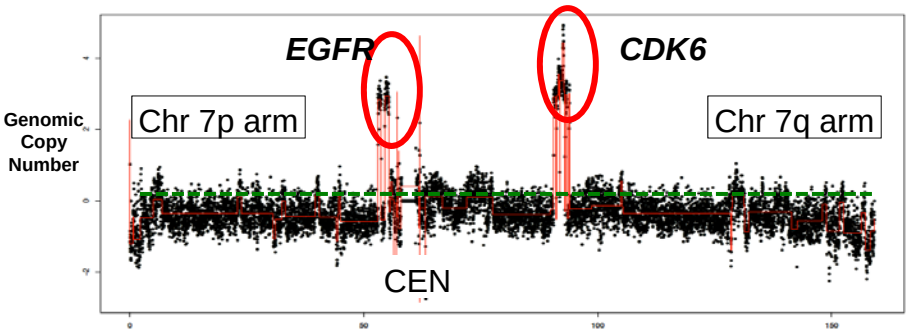
32

Cancer is Often Driven by Altered Gene Expression

Examples:

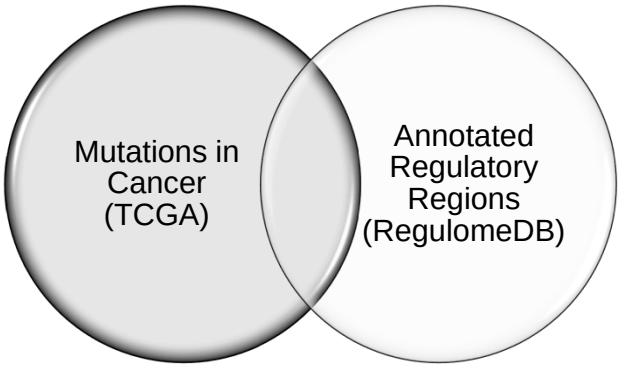
- Amplifications and Deletions
- Promoter fusions to Oncogenes

Colon Cancer Patient With Hanlee Ji



33

Are there functionally important mutations in regulatory regions in cancer?

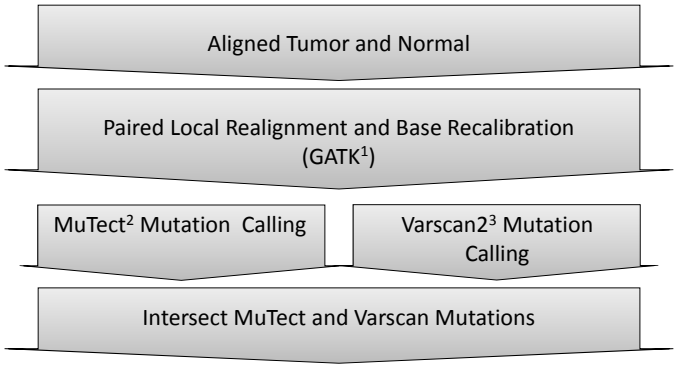


RegulomeDB:
ENCODE
RoadMap
+ Many Other

Boyle et al. Genome Res. 2012.

34

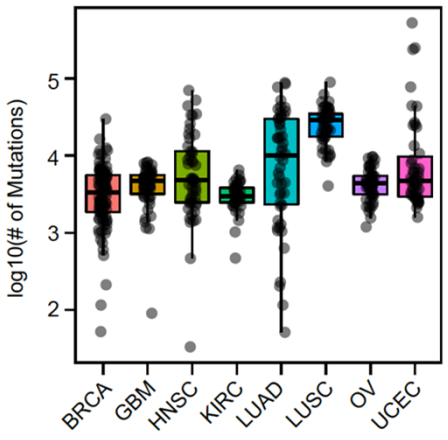
Identify Cancer Mutations from WGS Data: 438 Genomes



1. McKenna et al. Genome Res. 2010.
2. Cibulskis K et al. Nat Biotechnol. 2013.
3. Koboldt, D. et al. Genome Res. 2012.

35

Identified Mutations in 438 Individuals from 8 Cancer Types



BRCA: Breast Cancer
GBM: Glioblastoma Multiforme
HNSC: Head and Neck Squamous Cell Carcinoma
KIRC: Kidney Clear Cell Renal Carcinoma
LUAD: Lung Adenomcarcinoma
LUSC: Lung Squamous Cell Carcinoma
OV: Ovarian Cancer
UCEC: Uterine Corpus Endometrial Carcinoma

36

Are Mutations Enriched in Regulatory Regions?

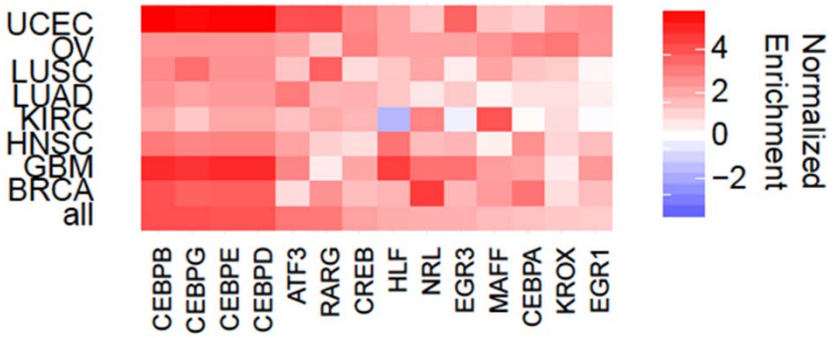
Fraction of Mutations in Regulatory Regions

vs

Fraction of Simulated Mutations in Regulatory Regions

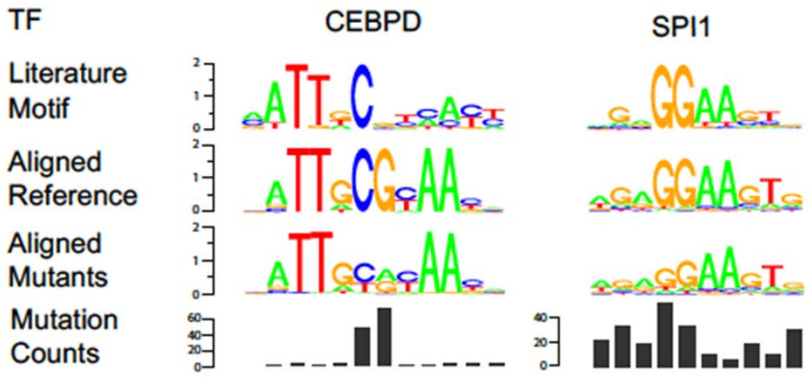
37

General Characterization: Sites Bound by Specific Transcription Factors are Enriched for Mutations



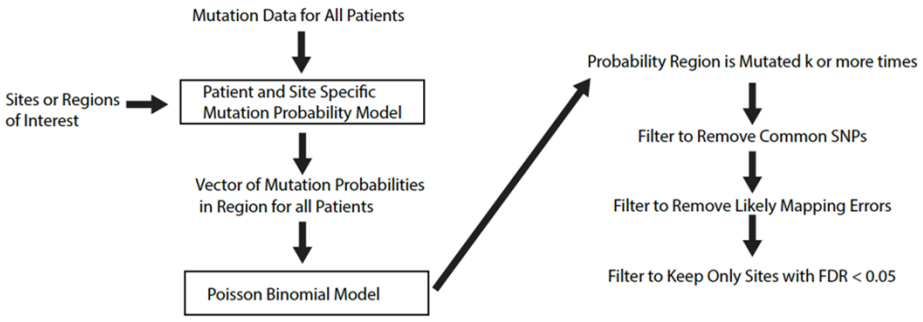
38

Occasionally Specific Residues within a Motif are Selectively Mutated



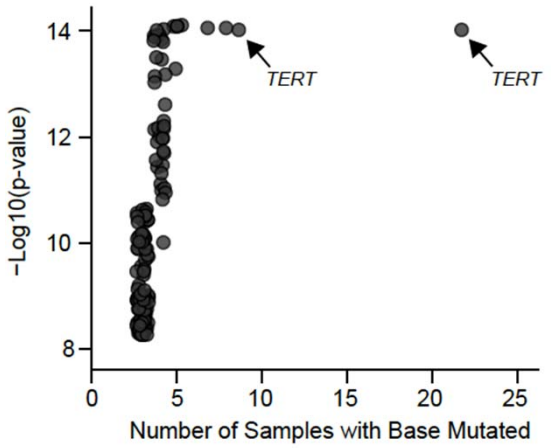
39

Approach to Identify Recurrently Mutated Sites



40

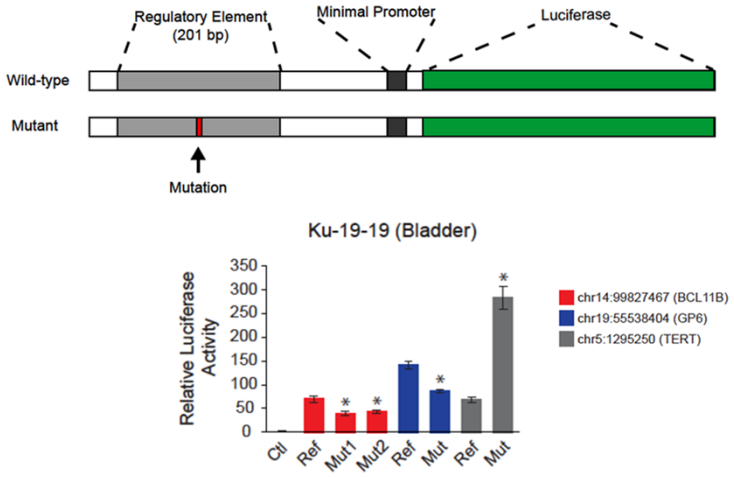
Significant Repeatedly Mutated Loci (~123 Regulatory Sites, Near ~200 genes)



Cancer Associated Genes in Vicinity of Regulatory Mutations
GNAS, INPP4B, MAP2K2, BCL11B, NEDD4L, ANKRD11, TRPM2, P2RY8

41

Mutations Alter Enhancer Function in Validation Assays



42

Conclusion

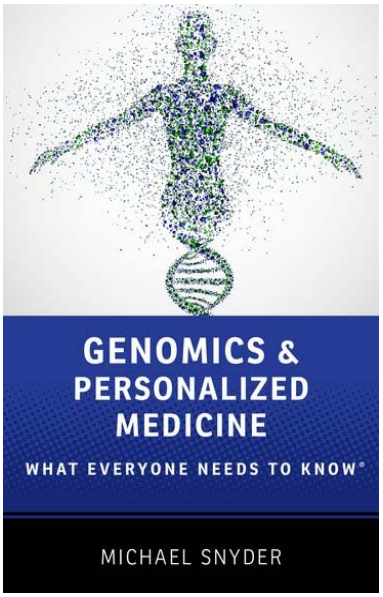
1) Can identify disease associated variants using a variety of approaches
2) Augmented exomes give better coverage
3) Genome sequencing provides useful information healthy people
4) Noncoding variants will be valuable to integrate into Genome interpretation

43

Acknowledgments

- **Mendelian Disease**
  - **NFKappaB** Shannon Rego, Dave Lewis
  - **NGly1:** Greg Enns, Jon Bernstein, Michael Clark, Rui Chen, Richard Gibbs, Huda + Many More
  - **IBD**: Shinichi Takashi, Sara Ellis
  - **Clinical Service:** Jason Merker, Euan Ashley + many others
- **Healthy People**
  - Euan Ashley, Rachel Wood, Rick Dewey, Tom Quertermous, Shannon Rego, Orit, The iPOP Team
- **Cancer**
  - Collin Melton, Jason Reuter, Damek Spacek, Alan Boyle (RegulomeDB)

44



Genomics and Personalized Medicine

What Everyone Needs to Know®

Michael Snyder

Available from Amazon

Stanford | Genetics and Genomics Certificate

New professional certificate - 100% online

Taught by faculty and industry experts

Certificate is comprised of 2 core courses in genetics and genomics and 4 elective courses

For more Information:

http://geneticscertificate.stanford.edu