Design a super panel for comprehensive genetic testing

Rong Chen, Ph.D.
Assistant Professor
Director of Clinical Genome Sequencing
Dept. of Genetics and Genomic Sciences
Institute for Genomics and Multiscale Biology
Icahn School of Medicine at Mount Sinai

Icahn School of Medicine at
Mount Sinai

Goal

- Enable genetic testing
  - Comprehensive coverage on cancer variants
  - Reasonable coverage on cardiovascular diseases
  - Comprehensive ethnicity report for background risk assessment
- Enable the identification of unexpected hero for resilience project
- Move Mount Sinai screening tests to next-gen
- Build a large number of sample for risk modeling

Comprehensive cancer variants

- 139 cancer genes
- 405 cancer GWAS SNPs
- 908 noncoding regions

Resilience sub-panel

- 209 genes with very high penetrance, leading to early onset of severe Mendelian phenotypes
- 861 mutations/indels
  - 270 outside exon regions (most covered by +/-50 bp)
- Covering 183 Mendelian diseases

Mount Sinai screening panel

- 551 mutations/indels in 133 genes for 126 Mendelian diseases, covering all screening panels from
  - Councyl_mssm: 293 mutations/indels in 97 genes for 94 diseases
  - Ambry,: 212 mutations/indels in 70 genes for 66 diseases
  - Tessarae: 380 mutations/indels in 88 genes for 83 diseases
- 130 out of 133 genes have been covered by the super hero panel

Comprehensive disease risk model

- 2029 SNPs in 537 genes that Rong has built risk models for 166 complex diseases

Final super panel

- **761 distinct genes with +/- 50 bp flanking exons**
  - ▫ 139 cancer genes
  - ▫ 150 Cardiovascular genes
  - ▫ 112 PGX genes
  - ▫ 112 genes from Invitae
  - ▫ 226 resilience genes
  - ▫ 41 Maturity Onset Diabetes of the Young (MODY) genes
- **4259 hot spots**

Over 4000 regions

- 405 cancer GWAS SNPs
- 908 cancer non-coding regions
- 751 ethnicity SNPs
- 270 mssm resilience regions
- 8 additional regions from mssm screening panel
- 2028 mssm disease risk model SNPs

A Multi-faceted Sequencing Technology Suite

Multi-Platform DNA Sequencing

- 1 x Applied Biosystems 3730xl
- 6 x Illumina HiSeq 2500
- 2 x MiSeq
- **2 x Ion Proton (NYC)**
- **8 x Ion Proton (CT)**
- **2 x Ion PGM**
- **1 x Ion Chef (NYC)**
- **8 x Ion Chef (CT)**
- 3 x PacBio RS II

Microarray Genotyping and Expression
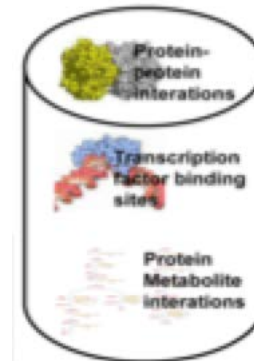
- 1 x Illumina Bead Array platform

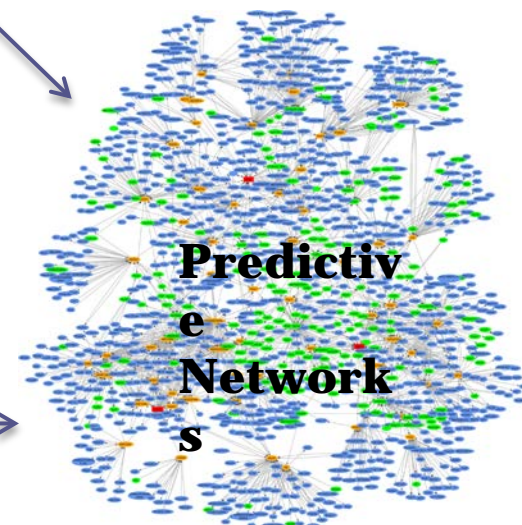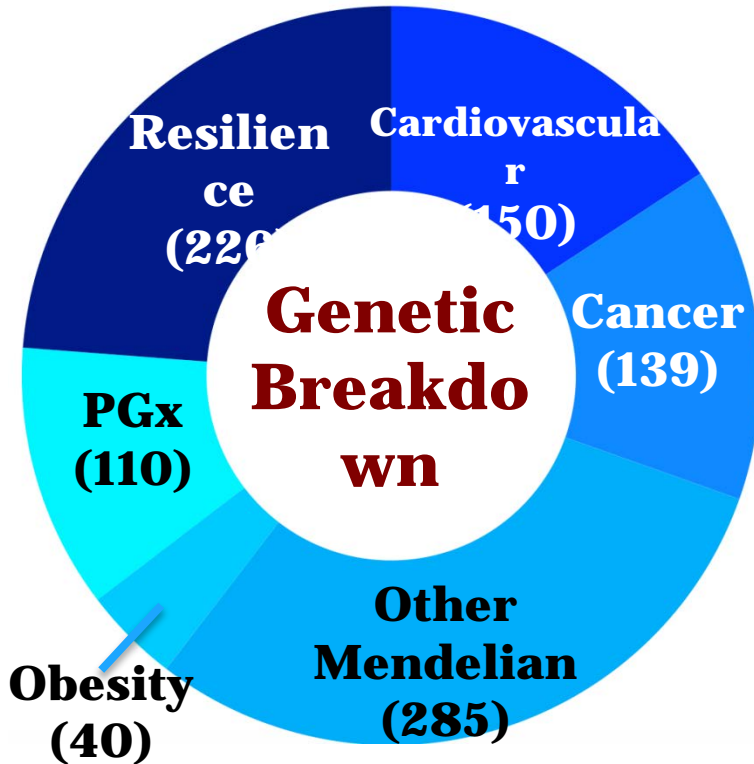## Liquid Handling Automation (CT)

Personal Data

EMR

Public Databases

Predictive Networks

Integrated Sequencing Data

The 25K Amplicon Panel Design Includes a Diverse Array of Loci

Genetic Breakdown

- Cardiovascular (150)
- Cancer (139)
- Other Mendelian (285)
- Obesity (40)
- PGx (110)
- Resilience (220)

Panel Design Criteria	Outcome
Number of Distinct Genomic Loci	12,905
Panel Size (unique genomic bases)	4,365,777 nt
% desired "Genes" covered at all	98.0%
% desired "HotSpots" covered at all	91.0%
% desired bases covered ("Genes")	97.9%
% desired bases covered ("Hotspots")	86.0%

**Gene = Input to AmpliSeq as gene exons w/ 50bp (5') and 25bp (3') flanks
HotSpot = 4495 targeted SNVs with flanks**

Panel design:

- \> 700 disease-associated genes (all coding exons covered)
- \> 4,000 disease-associated genomic HotSpots (including intronic and intergenic loci)
- Total of 4.4 MB – covered by approximately 24,644 amplicons in only 2 pools (products range from 125bp to 275bp)
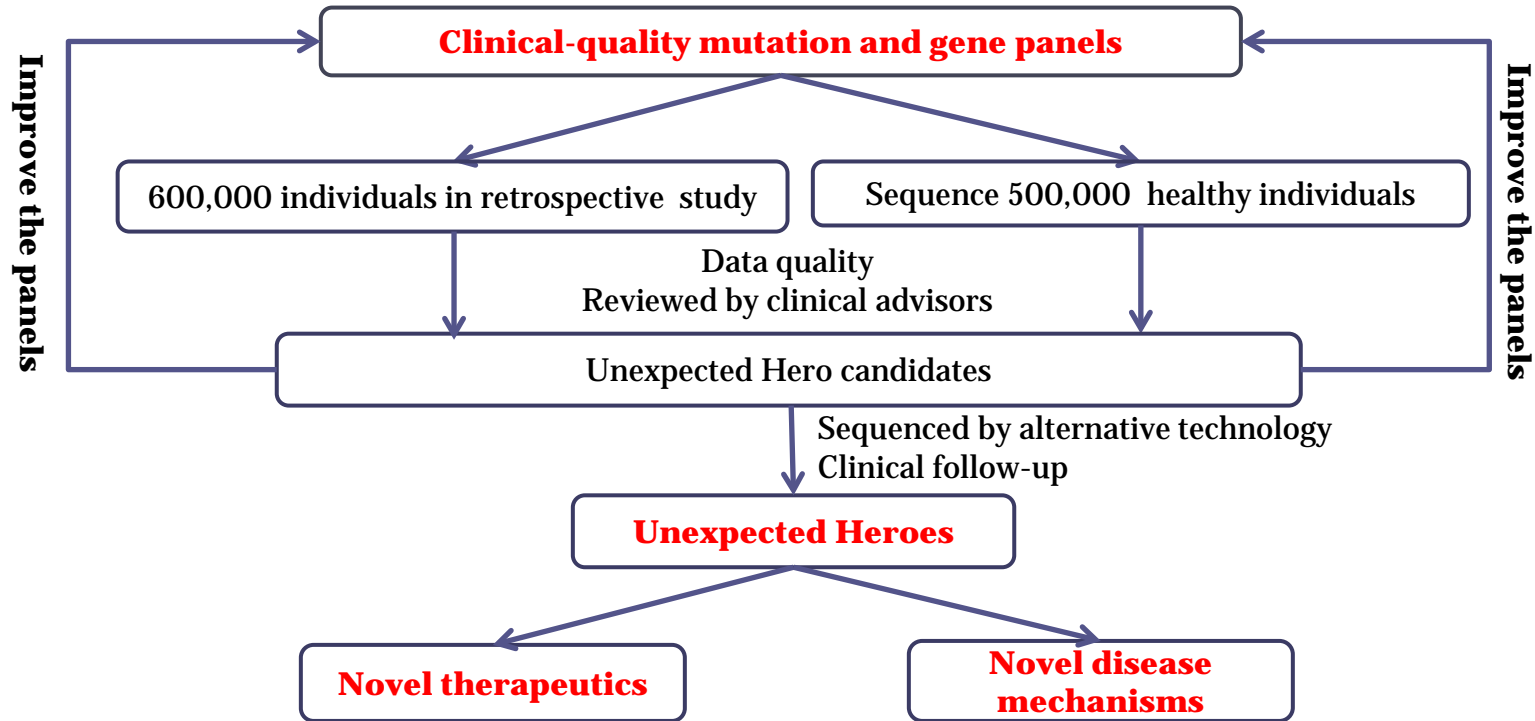
The Resilience Project: A Retrospective View

Commercial Screening Panels

Curated Mutation Databases

GENE Reviews | Genetics Home Reference *Your Guide to Understanding Genetic Conditions*

- Disease prevalence > 1:250,000
- Extremely high penetrance
- Severity phenotype
- Age of onset < 18

Clinical-quality mutation and gene panels

600,000 individuals in retrospective study

Sequence 500,000 healthy individuals

Data quality
Reviewed by clinical advisors

Unexpected Hero candidates

Improve the panels

Improve the panels

Sequenced by alternative technology
Clinical follow-up

Unexpected Heroes

Novel therapeutics

Novel disease mechanisms

From retrospective analyses, 13 potential unexpected heroes can be identified from 600,000 candidates screened.

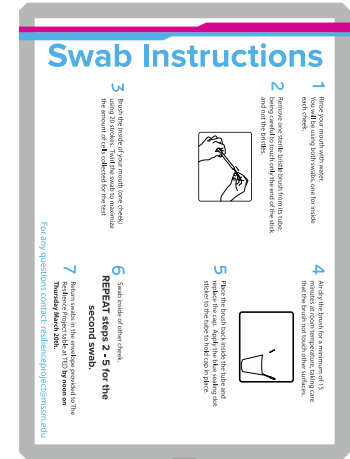# The Resilience Project Requires High-Throughput End-to-End
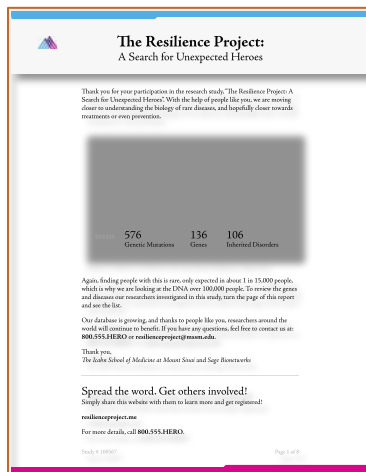
Online recruitment
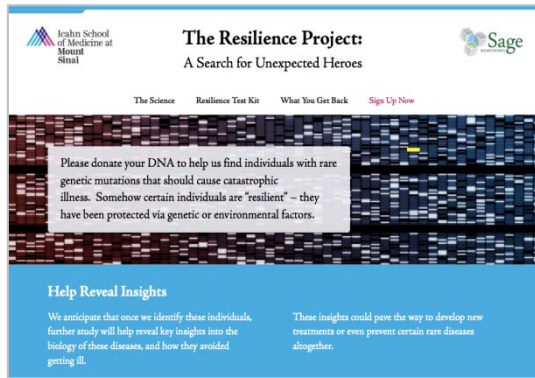
e-Consent

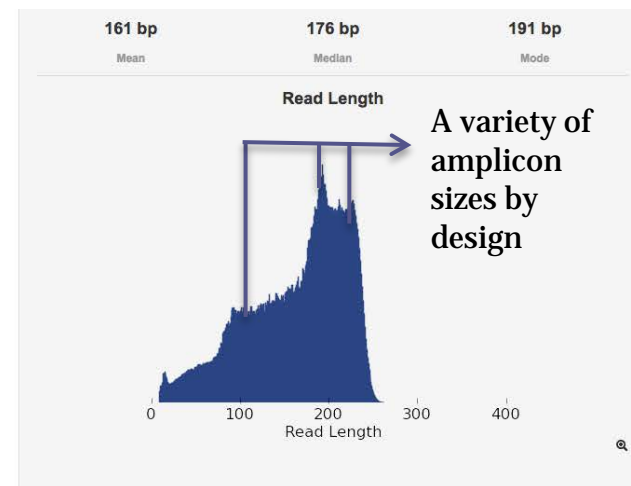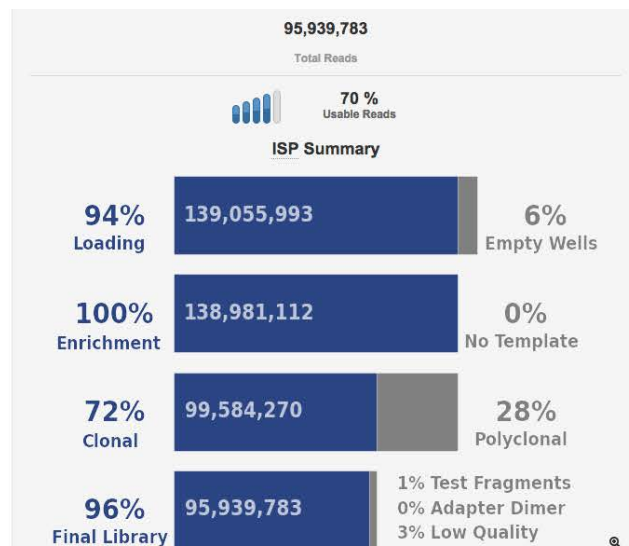**At-home sample collection**

e-Results

Bioinformatic analysis

Sequence analysis

On-Target and Uniformity Remains High, Despite Panel Diversity

A variety of amplicon sizes by design
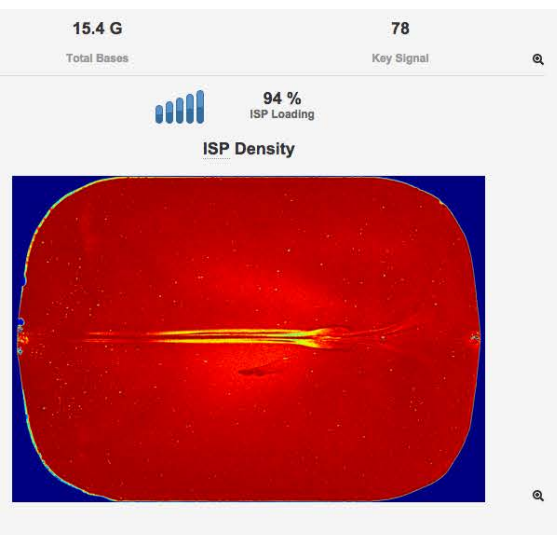
Amplicon Read Coverage

Number of amplicons	24,698
Percent assigned amplicon reads	96.19%
Average reads per amplicon	331.0
Uniformity of amplicon coverage	92.64%
Amplicons with at least 1 read	99.71%
Amplicons with at least 20 reads	97.38%
Amplicons with at least 100 reads	87.32%
Amplicons with at least 500 reads	19.48%
Amplicons with no strand bias	93.74%
Amplicons reading end-to-end	17.54%

Target Base Coverage

Bases in target regions	4,365,777
Percent base reads on target	96.28%
Average base coverage depth	251.3
Uniformity of base coverage	90.20%
Target base coverage at 1x	98.95%
Target base coverage at 20x	95.67%
Target base coverage at 100x	76.58%
Target base coverage at 500x	11.33%
Target bases with no strand bias	72.45%

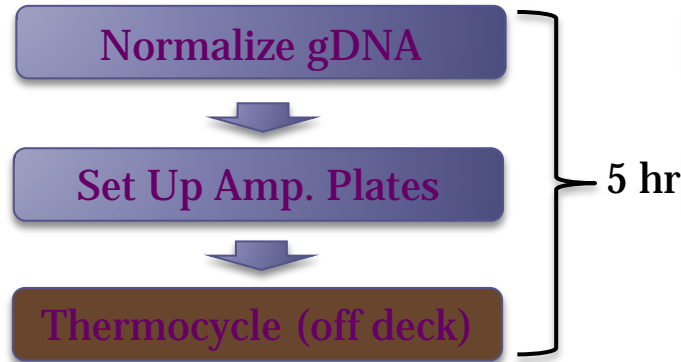**96.2% On-Target 90.2% Uniformity**
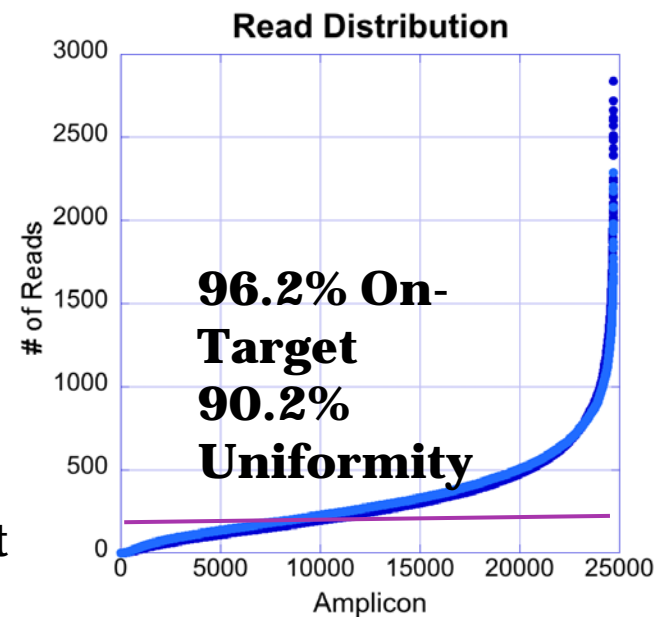
Using a 12-plex sample configuration, we target 200-250X coverage

High-Throughput Production Capacity Using AmpliSeq Panel

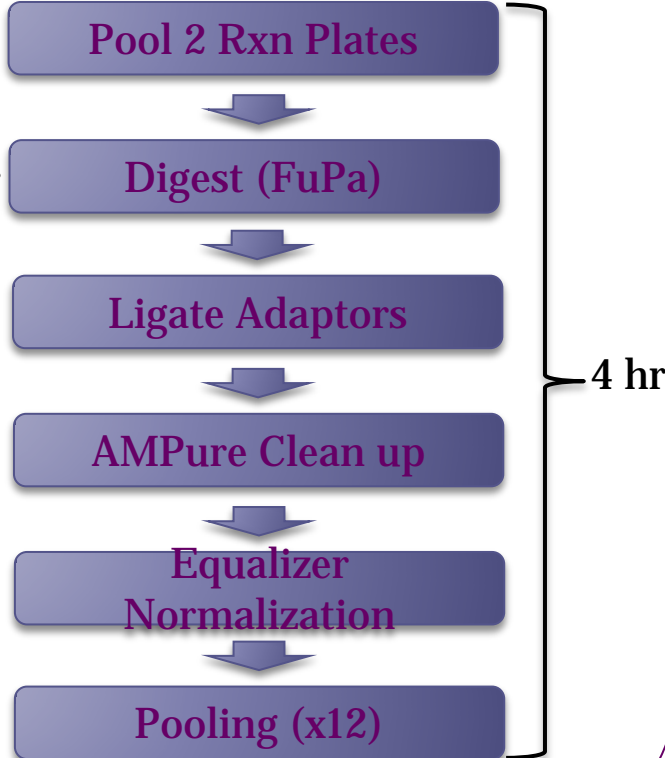## Automated AmpliSeq Production with Hamilton STAR (x2 Units)

Pre-amp

- Normalize gDNA
- Set Up Amp. Plates
- Thermocycle (off deck)

} 5 hr

-Set up for 192 panel preps/day, with 6 FTEs, but can easily easily expand to 384 or greater

Post-amp

- Pool 2 Rxn Plates
- Digest (FuPa)
- Ligate Adaptors
- AMPure Clean up
- Equalizer Normalization
- Pooling (x12)

} 4 hr

Sequencing Pipeline

Chef (x8)

Load 2 Pools @ 12 Preps

~ 14 hours

Proton (x8)

Load AM Chips

~ 3.5 hours

Load PM Chips

~ 3.5 hours

AM Run Data Available

16, 12-plex runs / day with 2 FTEs

Dedicated Run Processing, Storage, and Analysis

- 10 node cluster, expandable to 32 nodes.
- ~340 TB hard drive storage space, expandable to 1 PB.
- Currently supporting 16 Proton runs/day (160 – 224 Gb total filtered bases), can support 2X
- Ion Reporter for variant calling and annotation.

Acknowledgements