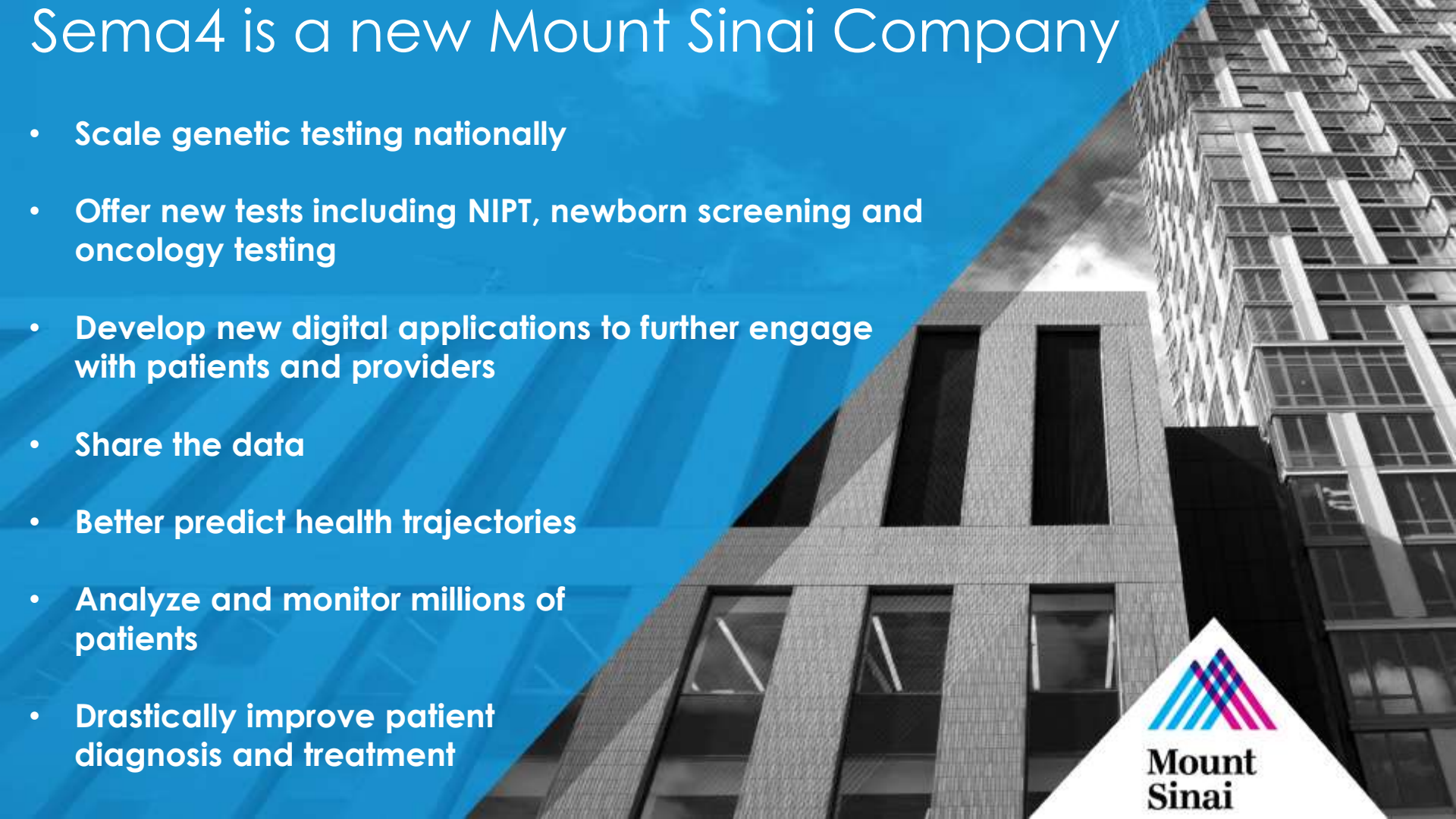
Capturing the entire clinically actionable genome with high-throughput long-read DNA sequencing technologies for comprehensive diagnostic testing

sema4

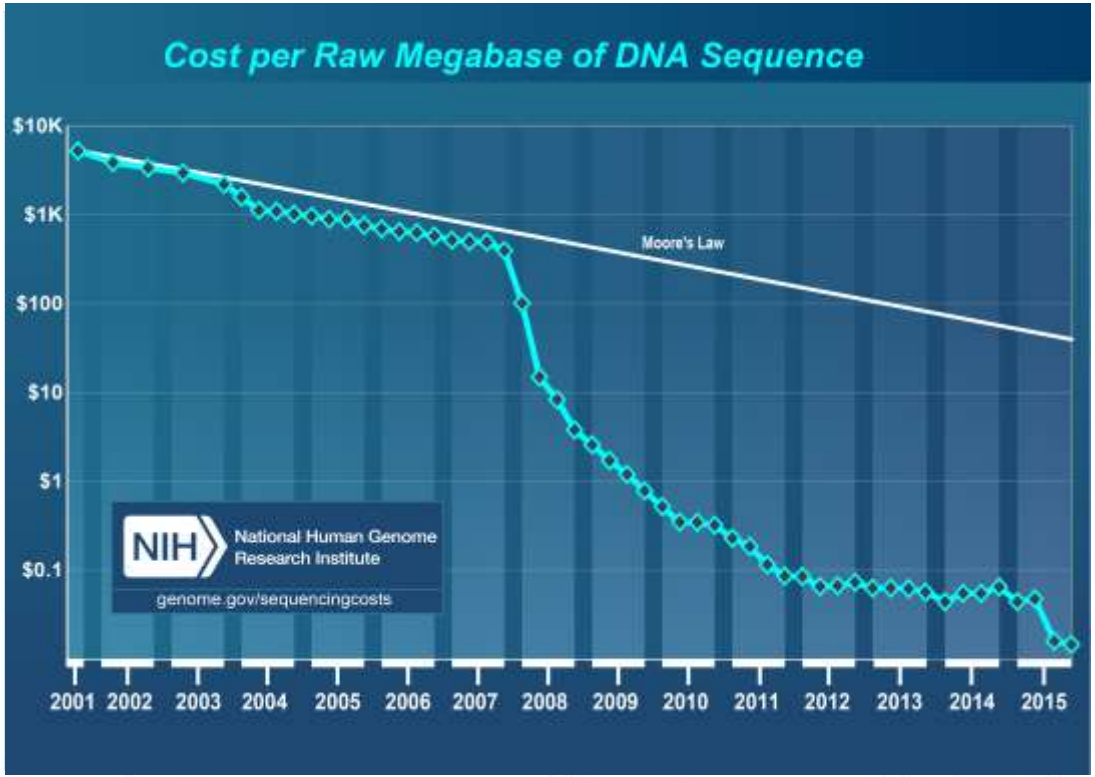
Sema4 is a new Mount Sinai Company

- Scale genetic testing nationally
- Offer new tests including NIPT, newborn screening and oncology testing
- Develop new digital applications to further engage with patients and providers
- Share the data
- Better predict health trajectories
- Analyze and monitor millions of patients
- Drastically improve patient diagnosis and treatment



Advanced technologies have given rise to an explosion in the digital universe

Advances in DNA sequencing technologies have now added to this explosion

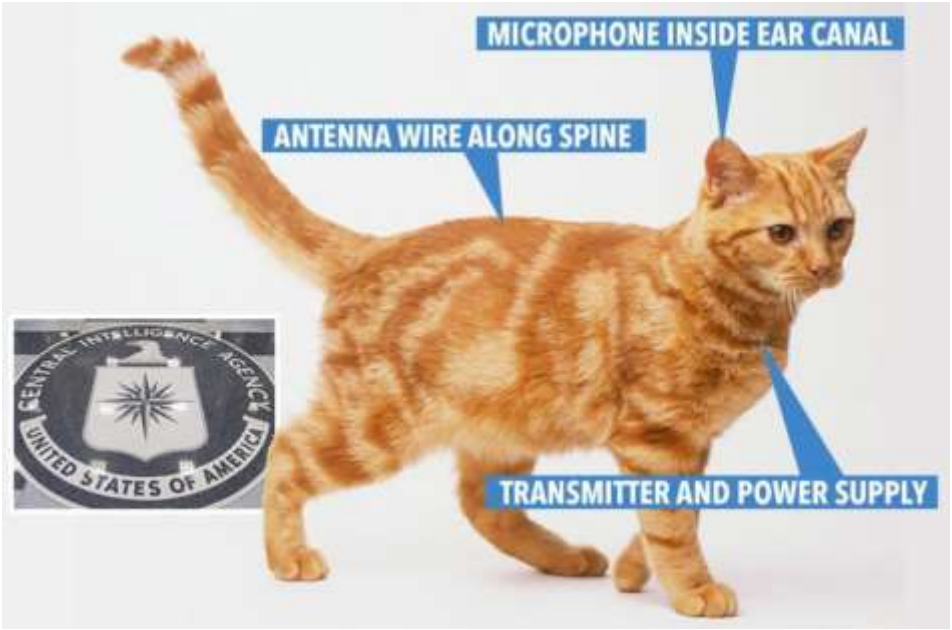


Organisms → Tissues → Single cells → Single cell, real-time, continuous?

FUR YOUR EYES ONLY CIA ‘implanted microphones into CATS’ in a bizarre attempt to spy on Russia

Mobile + Social Networks

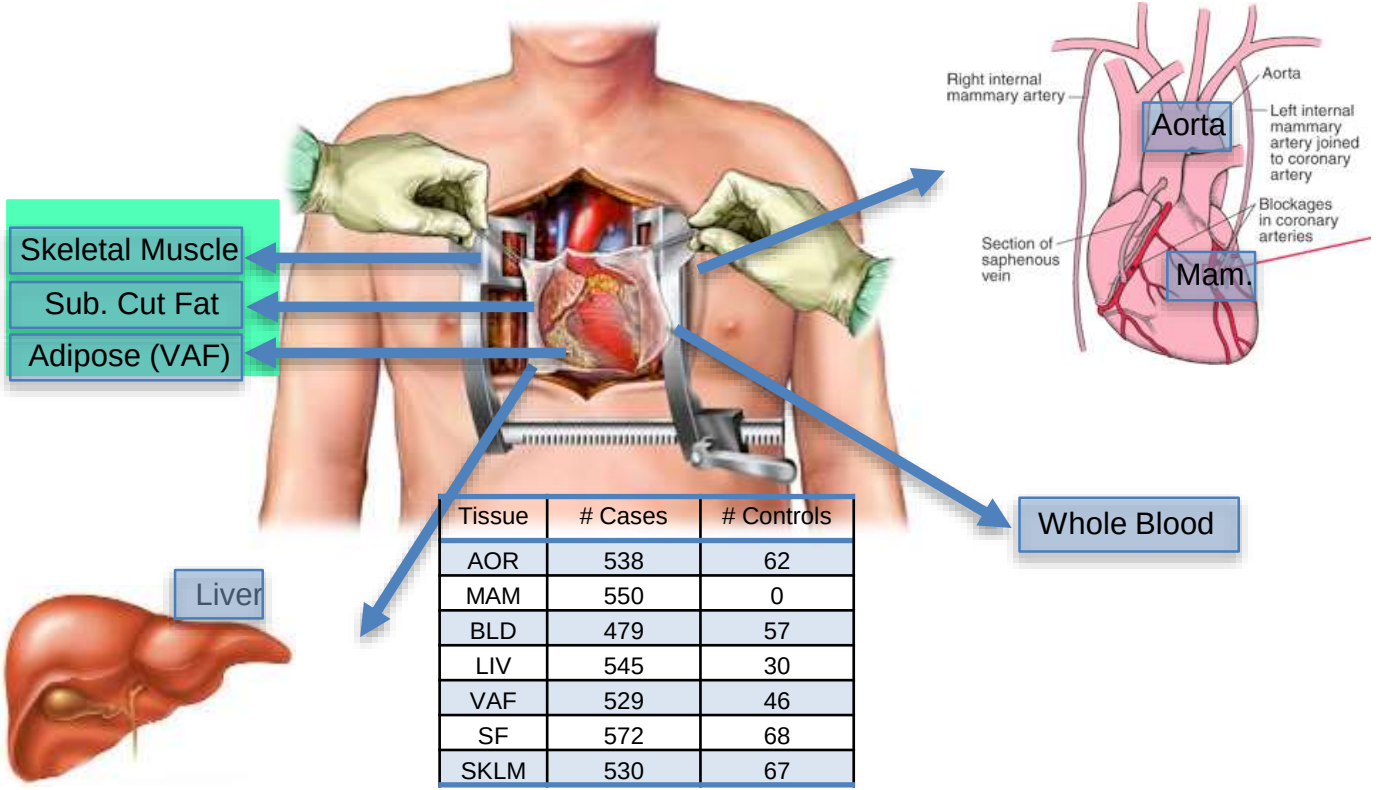
Gesture-based , Interactive Computing



Next-Gen Genomics

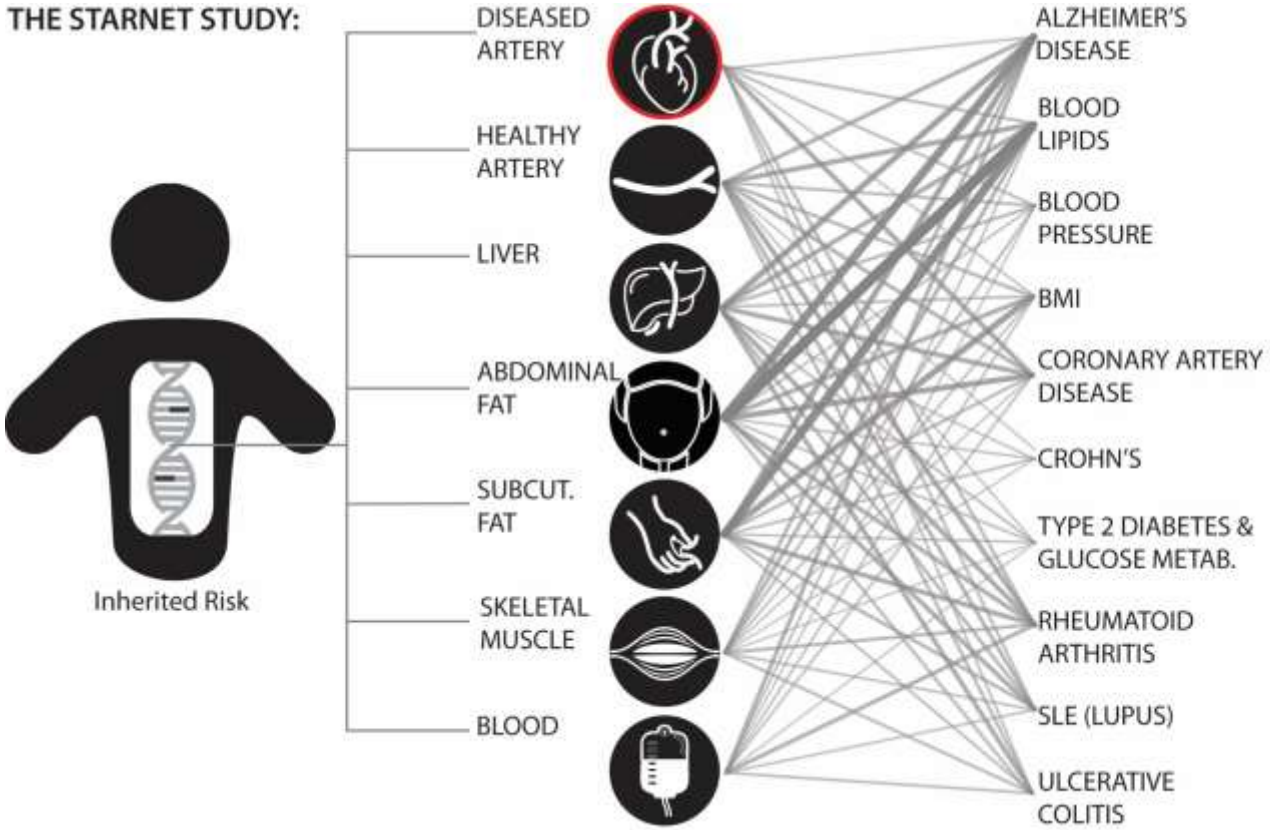
The “Internet of Things”

On the molecular side, significant cohorts profiled at an unprecedented depth



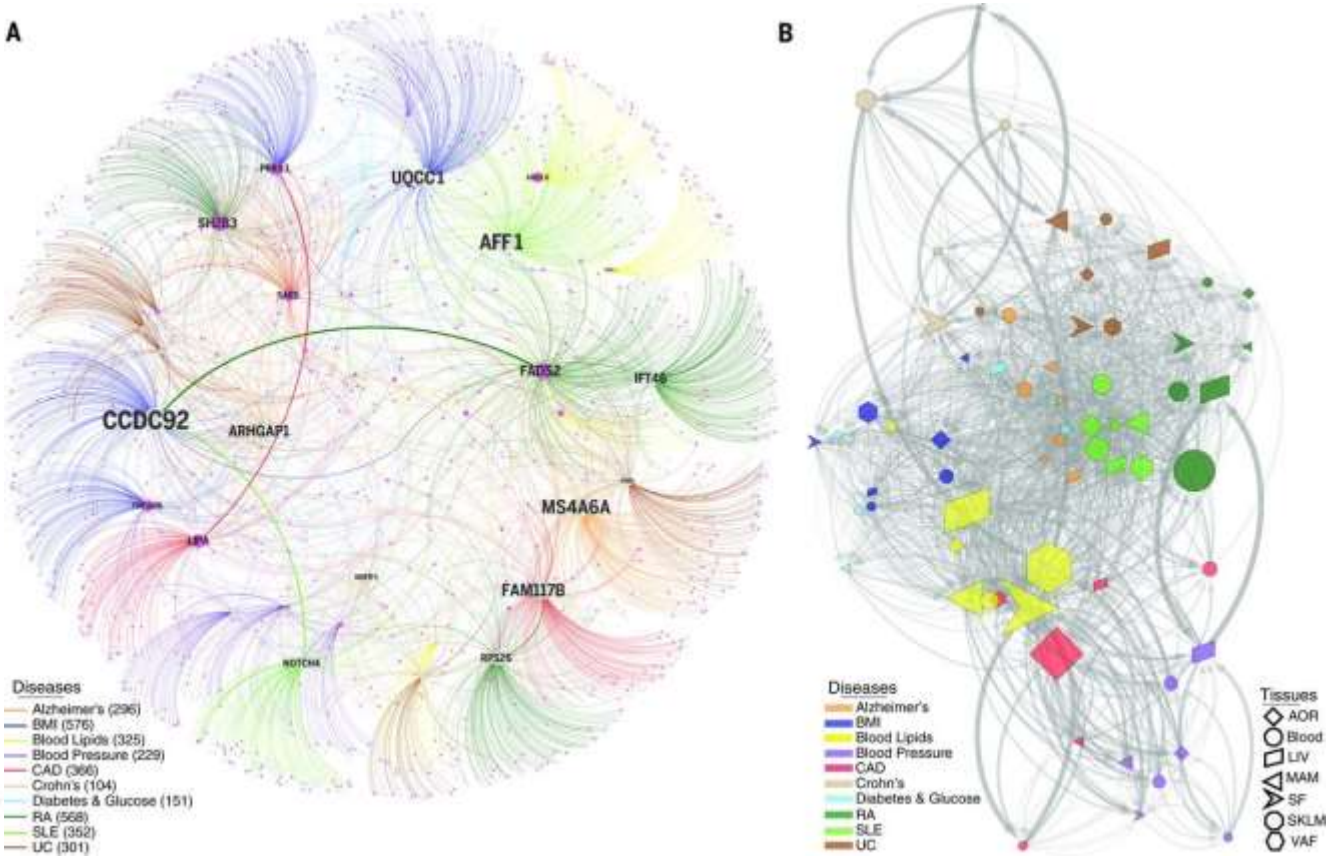
Tissue	# Cases	# Controls
AOR	538	62
MAM	550	0
BLD	479	57
LIV	545	30
VAF	529	46
SF	572	68
SKLM	530	67

Providing for the first time extensive maps of disease

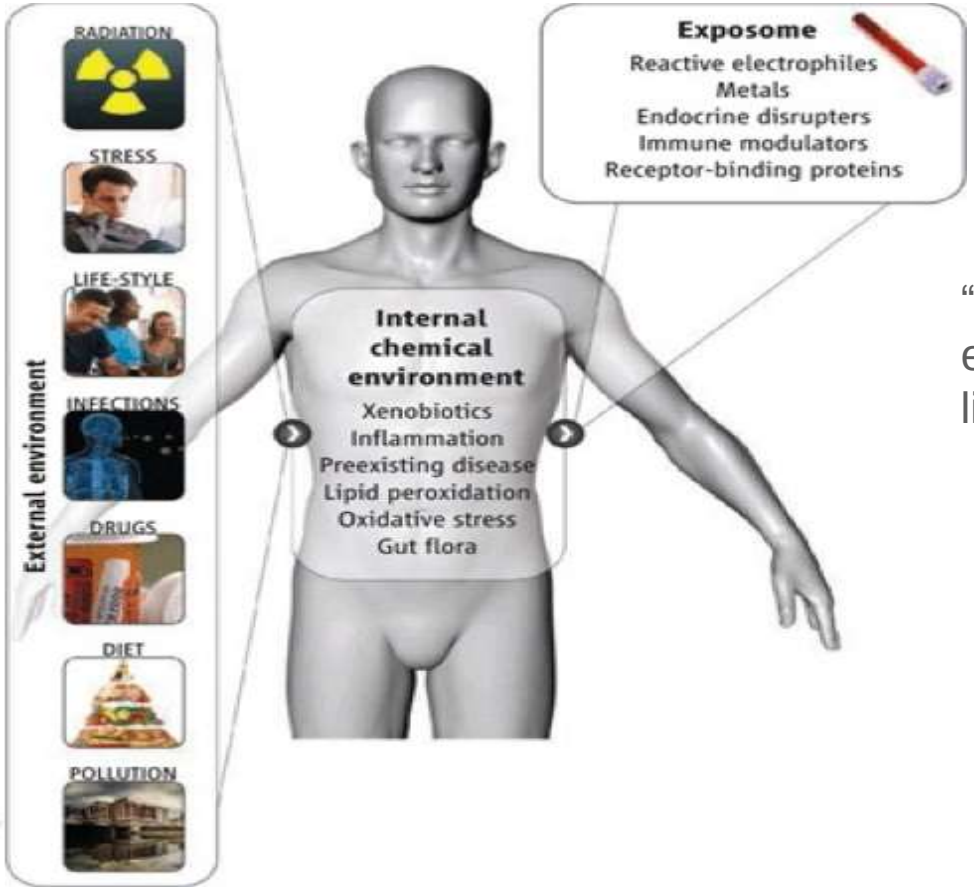


The thickness of the edges between the tissue and the disease reflect how frequently a disease-causing gene is shared between pairs of diseases and tissues.

Which in turn has enabled the construction of predictive network models elucidating the molecular underpinnings of disease and wellness



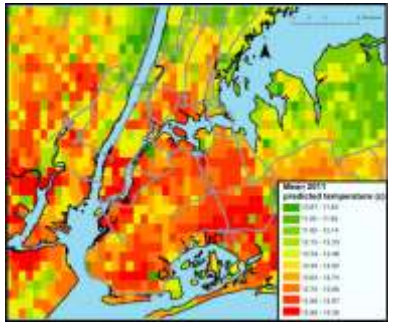
Complementing the molecular dimensions is the “Exposome”, the newest “Omic”



“The totality of environmental exposures throughout a lifetime”

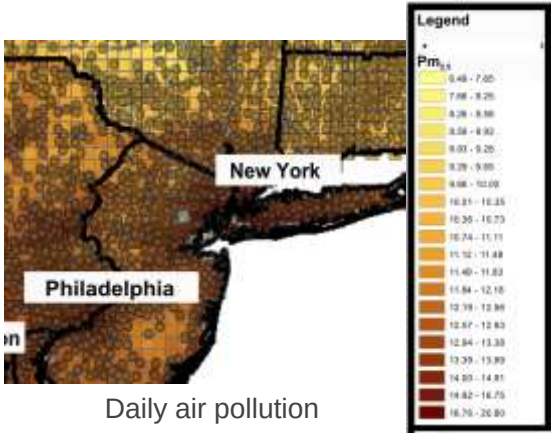
-Analogous to the Genome

App enabled use of Geographical Information Systems for the "*External Exposome*"

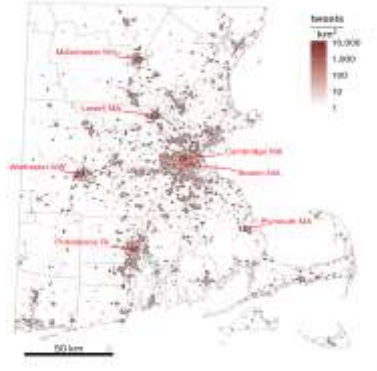


Air Temperature/Climate

Exposure to green, natural areas



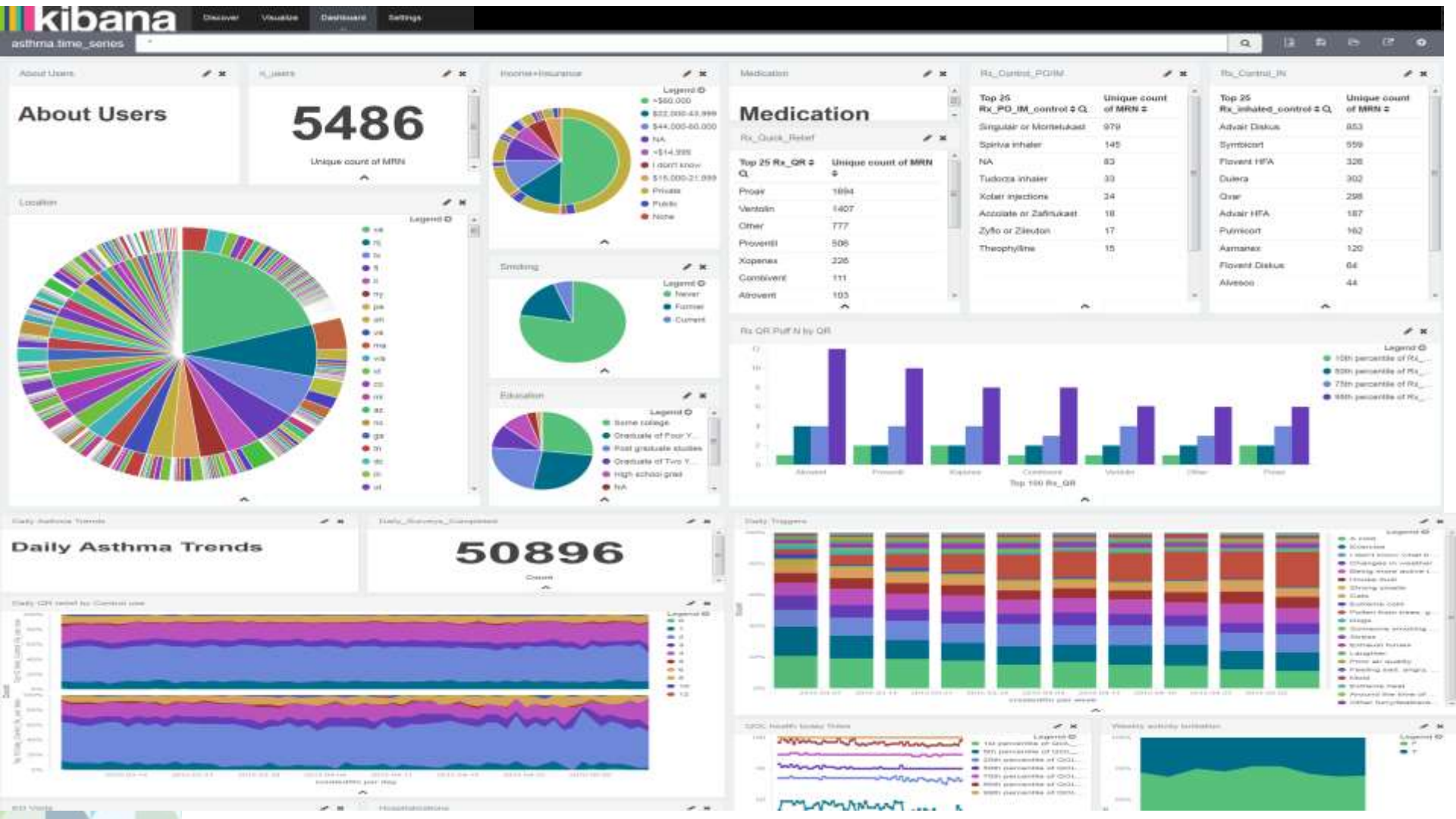
Daily air pollution



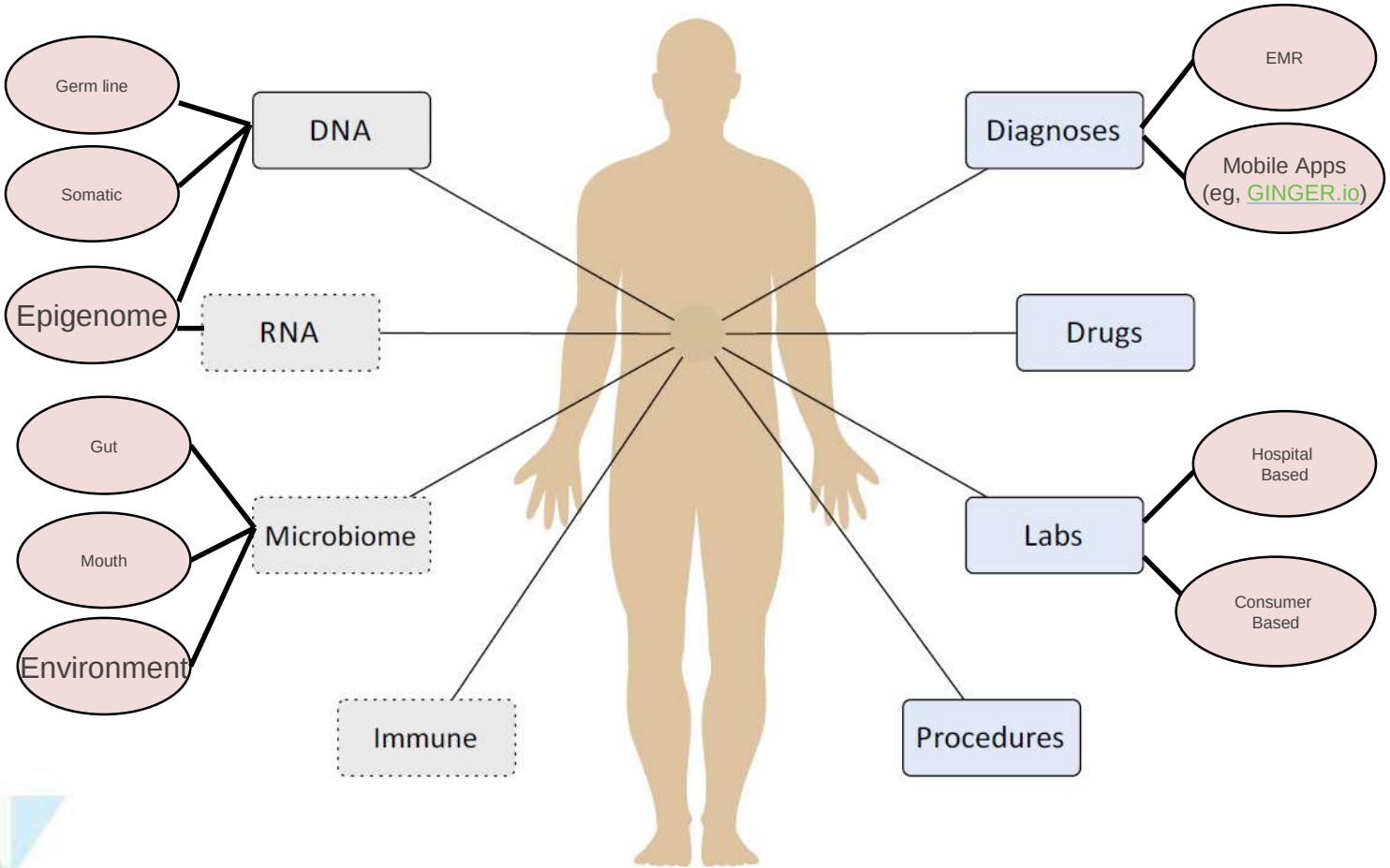
Social Media Content

Access to Healthy Foods

Traffic patterns


kibana
Discover
Visualize
Dashboard
Settings
asthma.time_series
About Users
About Users
5486
Unique count of MRN
Income+Insurance
Legend
>$60,000
$22,000-43,999
$44,000-60,000
NA
<$14,999
I don't know
$15,000-21,999
Private
Public
None
Location
Legend
Smoking
Legend
Never
Former
Current
Education
Legend
Some college
Graduate of Four Y...
Post graduate studies
Graduate of Two Y...
High school grad
NA
Medication
Medication
Rx_Quick_Relief
Top 25 Rx_QR
Unique count of MRN
Proair 1894
Ventolin 1407
Other 777
Proventil 506
Xopenex 226
Combivent 111
Atrovent 103
Rx_Control_PO/IM
Top 25 Rx_PO_IM_control
Unique count of MRN
Singulair or Montelukast 979
Spiriva inhaler 145
NA 83
Tudorza inhaler 33
Xolair injections 24
Accolate or Zafirlukast 18
Zyflo or Zileuton 17
Theophylline 15
Rx_Control_IN
Top 25 Rx_inhaled_control
Unique count of MRN
Advair Diskus 853
Symbicort 559
Flovent HFA 326
Dulera 302
Qvar 298
Advair HFA 187
Pulmicort 162
Aamanex 120
Flovent Diskus 64
Alvesco 44
Rx QR Puff N by QR
Legend
10th percentile of Rx_...
50th percentile of Rx_...
75th percentile of Rx_...
95th percentile of Rx_...
Albuterol
Proventil
Xopenex
Combivent
Ventolin
Other
Proair
Top 100 Rx_QR
Daily Asthma Trends
Daily Asthma Trends
50896
Count
Daily Triggers
Legend
Daily QR relief by Control use
Legend
Count
controller use week
Daily Health Index Trends
Legend
10th percentile of Daily_...
50th percentile of DAILY_...
75th percentile of DAILY_...
90th percentile of DAILY_...
95th percentile of DAILY_...
Weekly activity limitation
Legend
Y
N
ED Visits
Hospitalizations

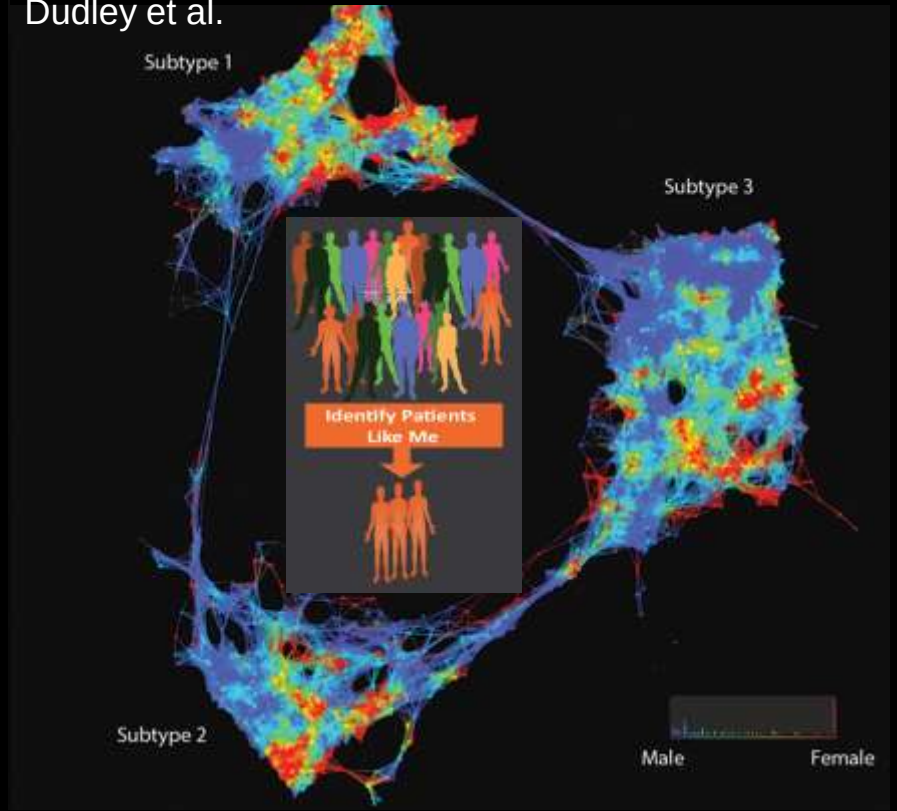
The value of a more completely phenotyped and molecularly profiled population



Using EHR data, diabetic patients organized into 3 distinct groups representing different severity, comorbidities and genetic components

Science Translational Medicine AAAS

Dudley et al.

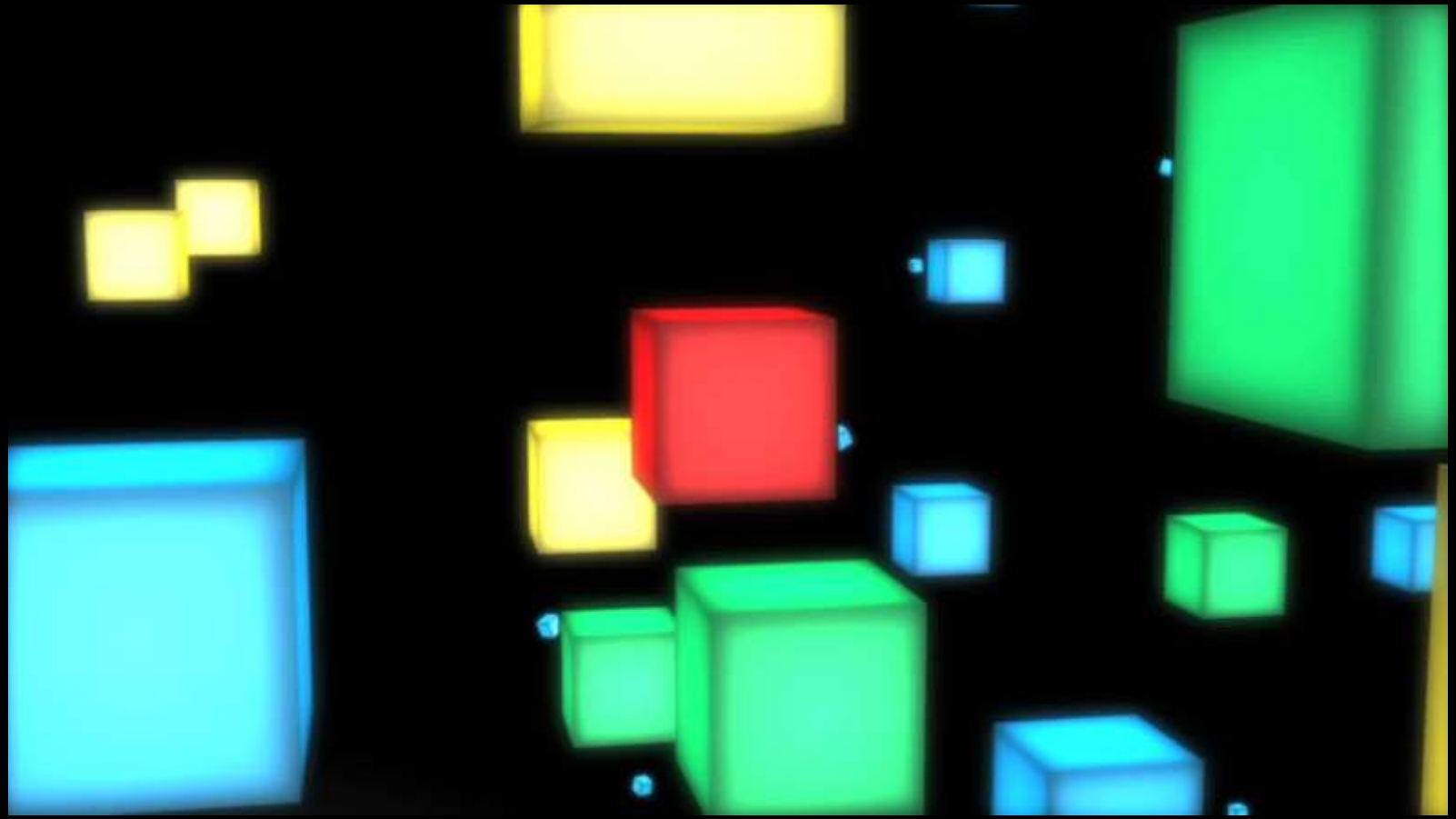


Subtype 1:
- More likely to suffer from blindness and vision defects
- Grouping genetically supported

Subtype 2:
- Greater risk of infections and cancer
- More immune deficient

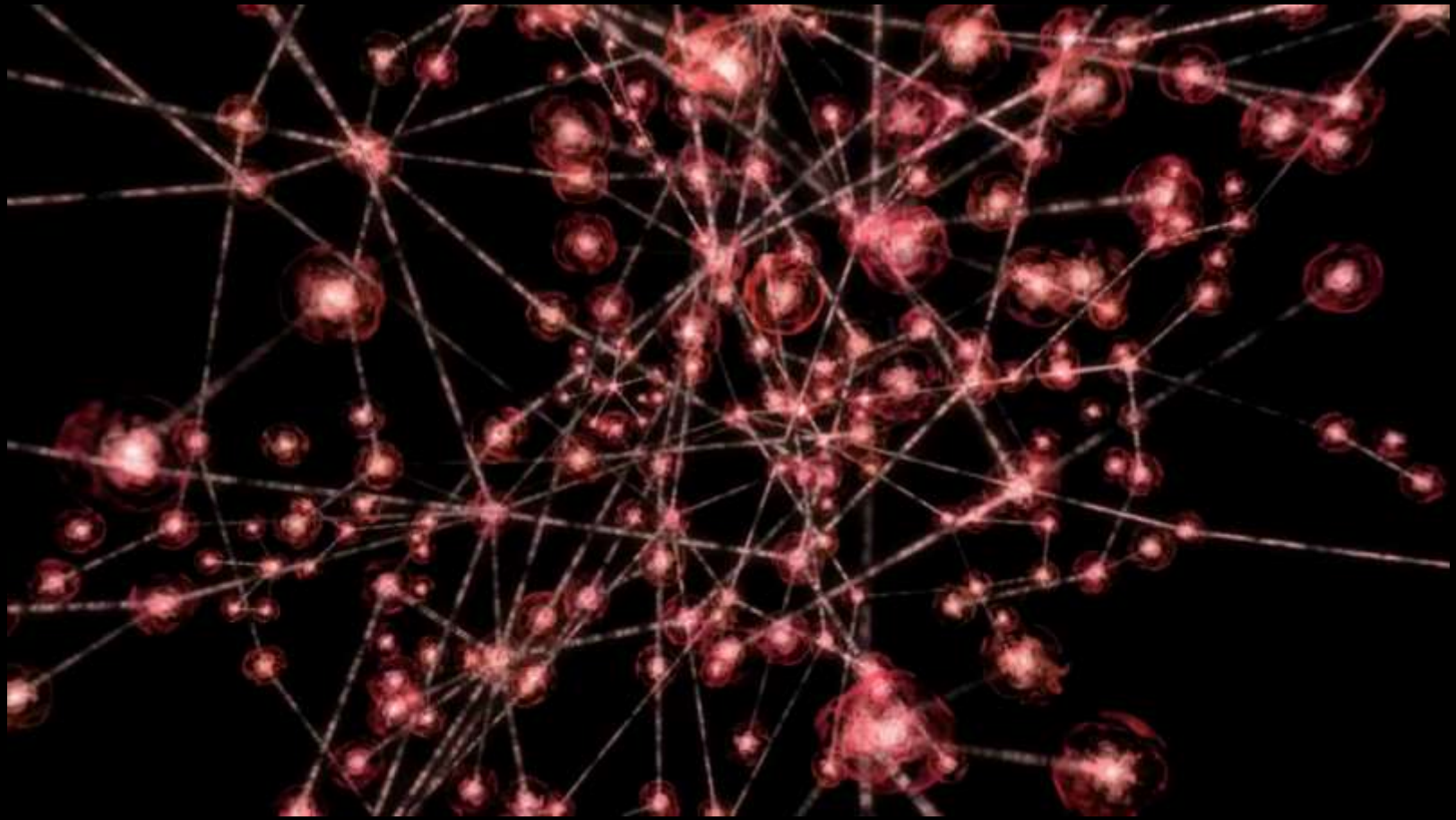
Subtype 3:
- Higher blood pressure, blood clots, more metabolic syndrome like
- Grouping genetically supported

We can go further to integrate these different dimensions of data to build models

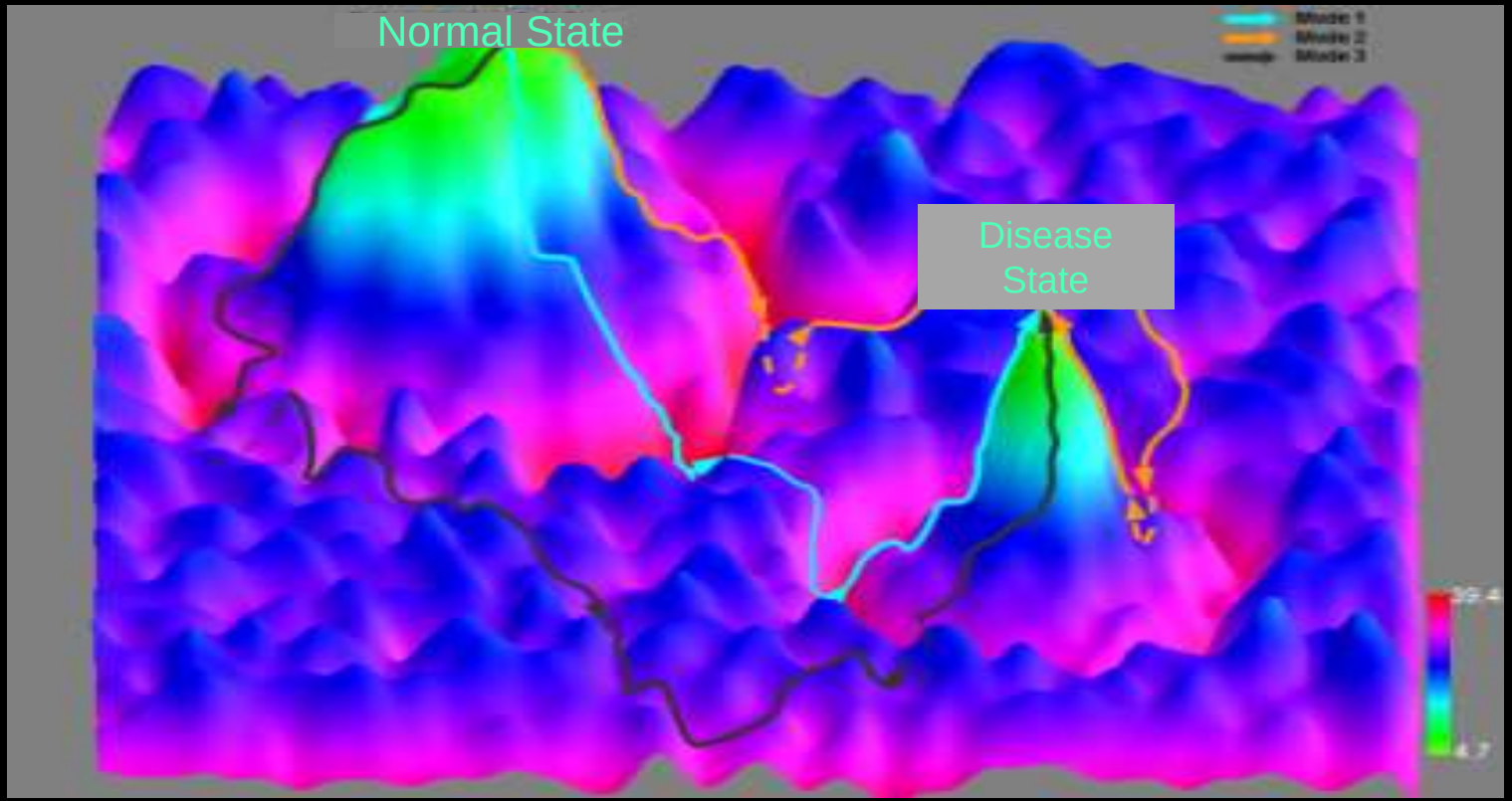


Host Molecular(DNA, RNA, Proteins) Microbiome Molecular Clinical (EMR) Consumer Acquired

With such models, we can carry out biology in silico, perturbing systems “experimentally” to understand how information flows through it



Ultimate Objective: Model individual health course trajectories to enhance clinical decision making



Adapted from Rui Chang et al. PLoS Computational Biology

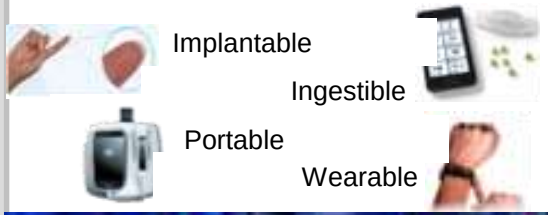
Medical systems of the future…

Largest car company in the world owns no cars (founded 2009)

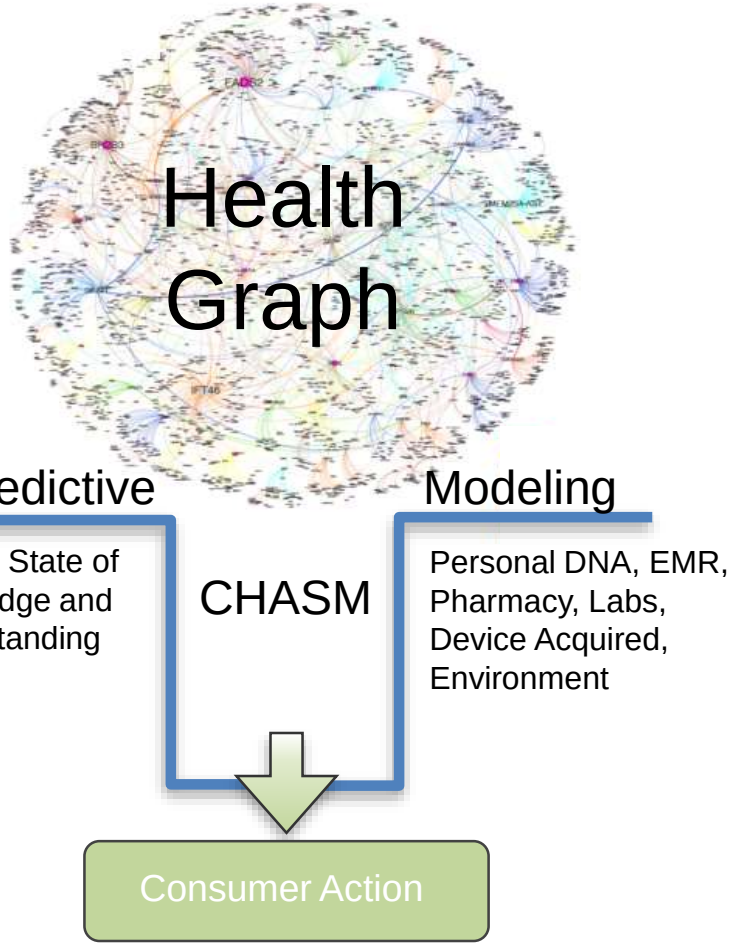
> 200,000 drivers in US
> 1,000,000 drivers world wide

Largest hotel chain in the world owns no hotels (founded in 2008)

Soon the largest medical system in the world will own no hospitals



THE PROBLEM: We do not have the scale of content needed to build these models to realize this vision



Unique capabilities of coupling what people ***need*** to know with what they ***want*** to know for their future

The pregnancy journey is a great example of the opportunity that exists to engage patients along their health trajectory

Our solution: Engage patients along their [illegible] providing comprehensive, medically actiona[illegible]g

KNOWLEDGEBASE
(enabling better tracking

DATA ACQUISITION AND MANAGEMENT
(taking into account more complete information)

Molecular

Outcomes

Mount Sinai

NLP
Curatio
ETL

Networks
Knowledg
Simulator

V. Gen.
V. Onc.
V. Res.

Data → Knowledge → Understanding

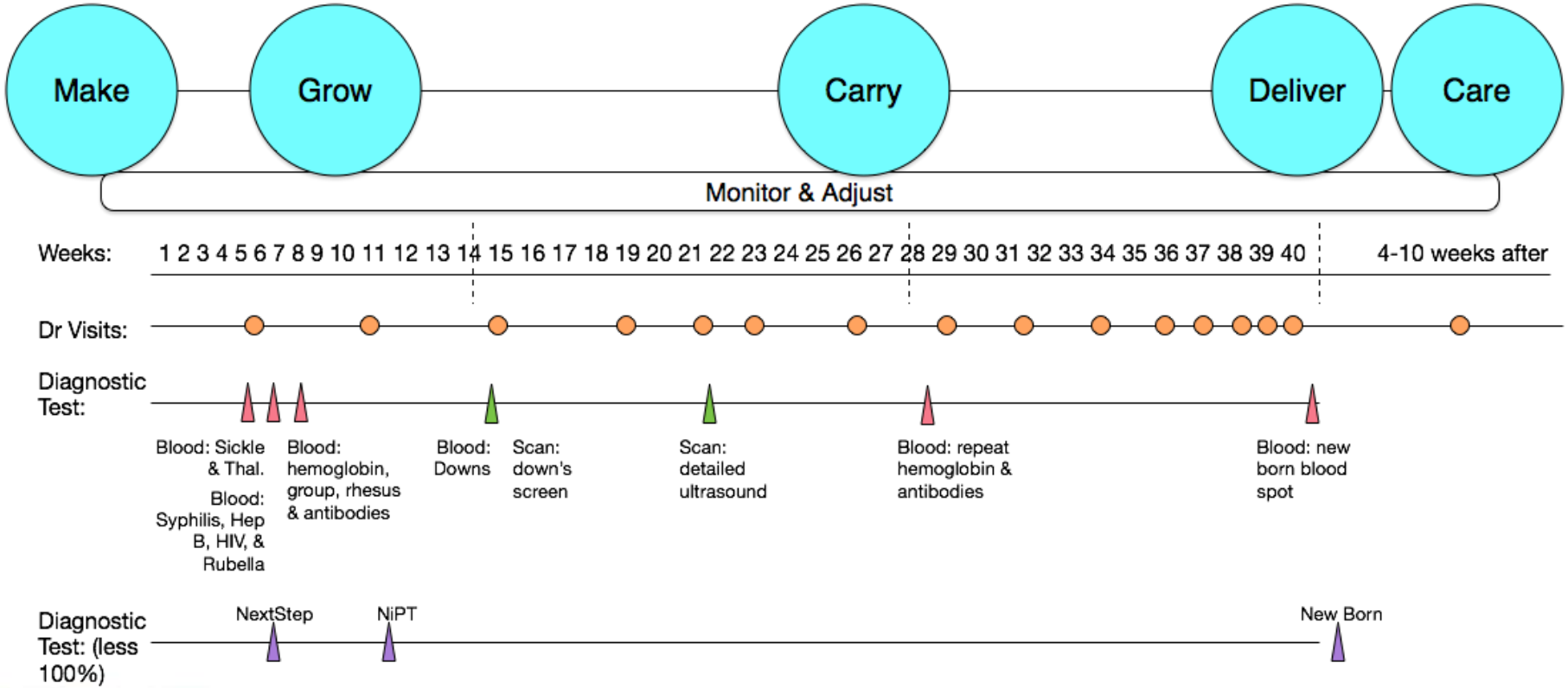
Deep, Dynamic, Adaptive Learning Platform

ADVANCED ANALYTICS
(providing better solutions for the future)

We don't just run a test, we engage patient and physician as partner

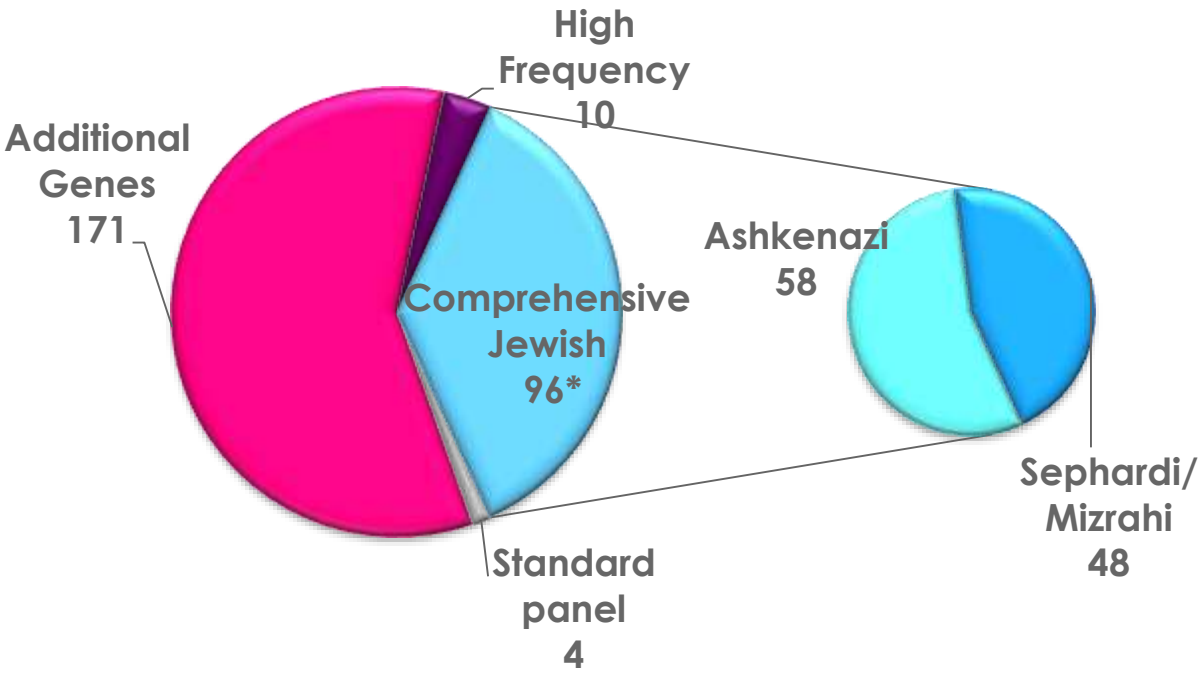
- Aid physicians in maximizing patient outcomes
- Streamline physician workflow
- Partnering to enable learning healthcare systems

Milestone connection between diagnostic testing and information system



Development of an Expanded Pan Ethnic Carrier Screen

281 autosomal recessive and X-linked diseases chosen by literature review, internal research, and physician input



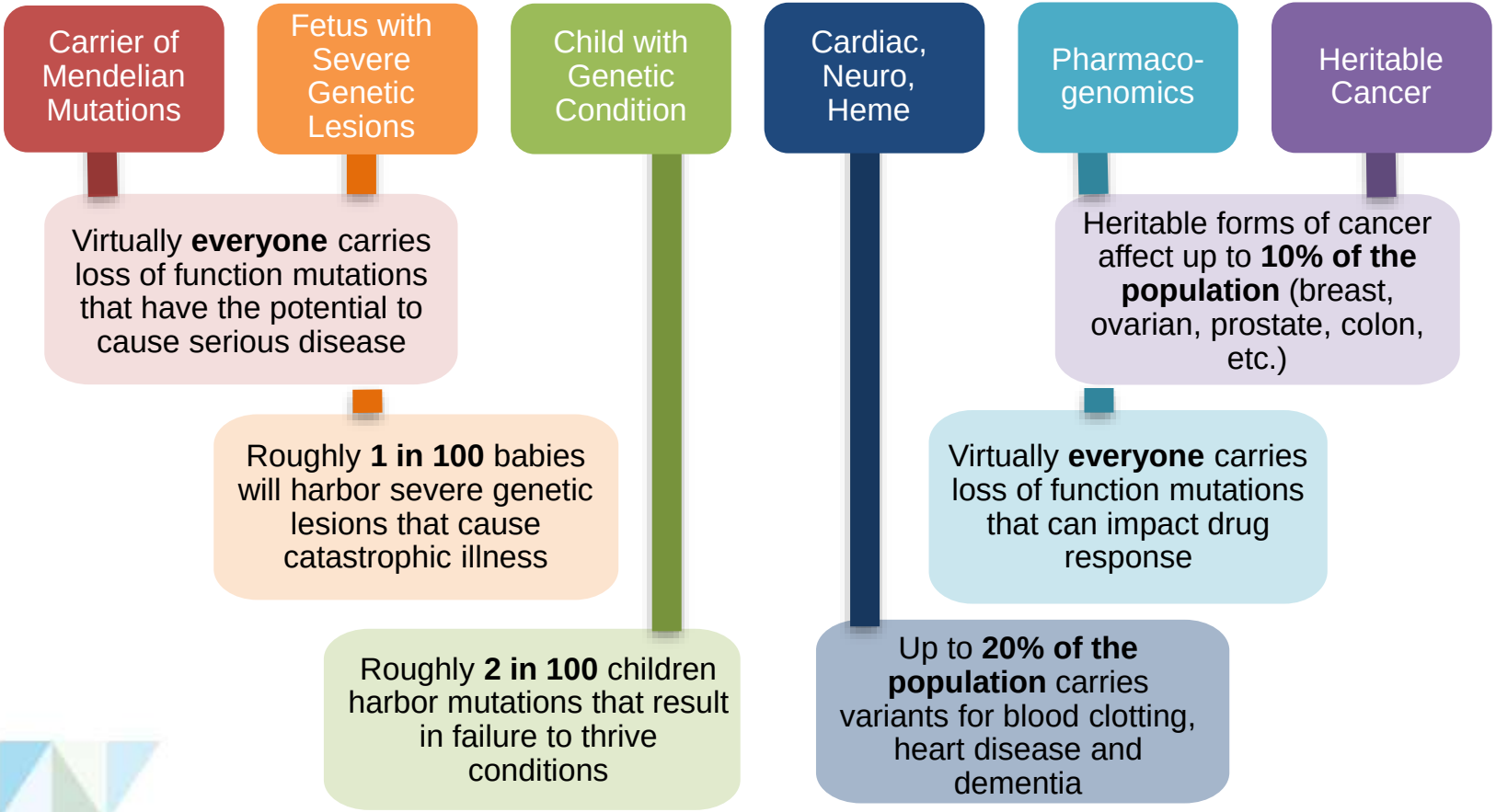
Jewish Migration Paths (Ashkenazi, Sephardi, and Mizrahi)

- 48 disorders with Ashkenazi founder alleles
- 38 disorders with Sephardi/Mizrahi founder alleles
- 10 common to all groups

>800 Sephardi/Mizrahi samples screened from the greater NY metropolitan area

>Exhaustive literature search

Beyond the pregnancy journey we now know enough to make genomic testing relevant for everyone

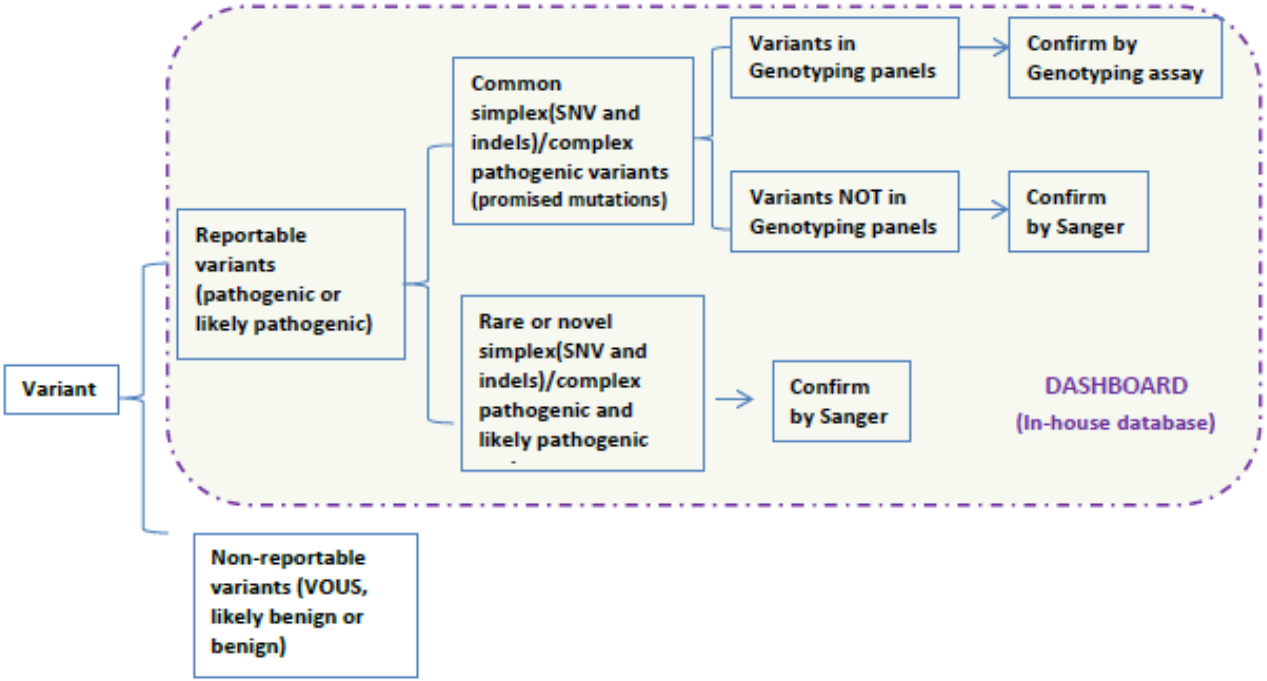


But we need 3rd gen sequencing to extract the most meaning from individual genomes we push it?

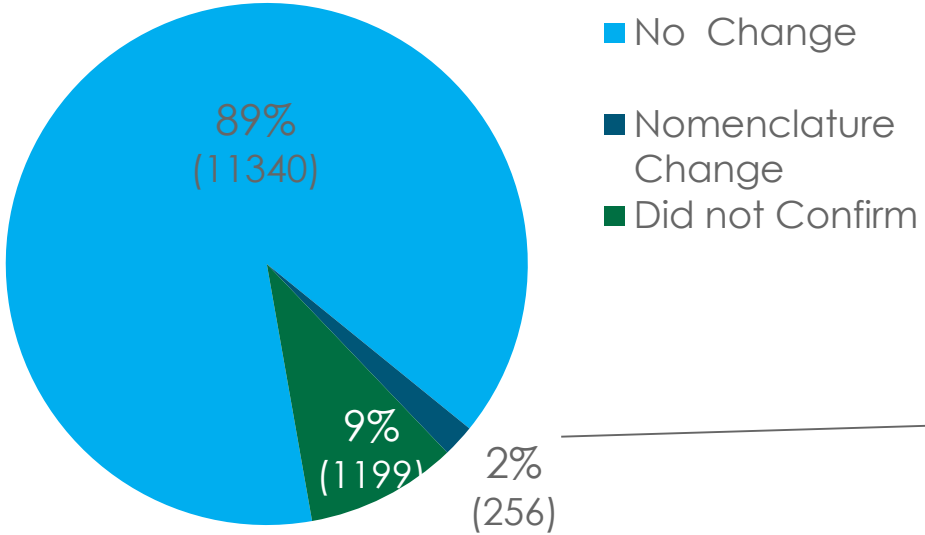
- **Many advantages:**
  - Long read lengths → chromosome-scale assemblies
  - Can use native DNA (no amplification required)
  - *Genome-wide DNA methylation detection (4mC, 6mA) at single-base, strand-specific resolution*

- Has opened up new fields of research in epigenomics and de novo genome assembly

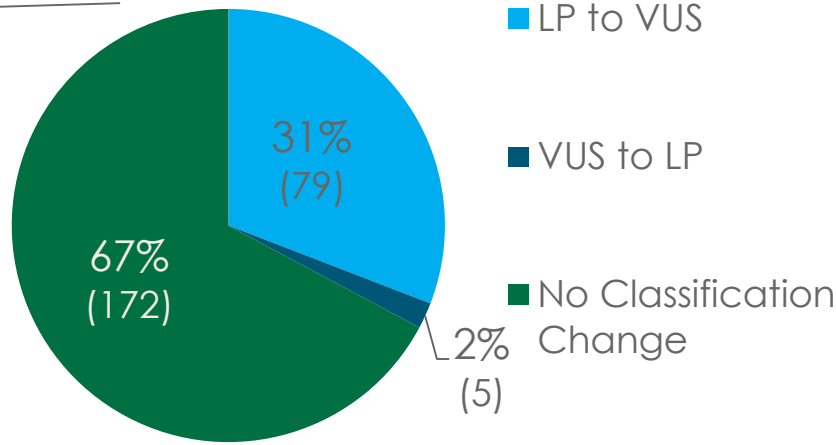
NGS variant calling for single nucleotide variants is highly accurate >99.9%
Complex insertion/deletion calling is not 99.9% accurate (~95%)



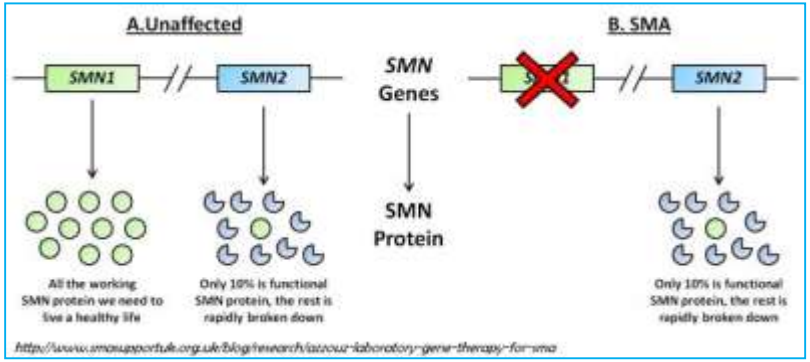
Using PacBio sequencing as an orthogonal technology to clinically validate

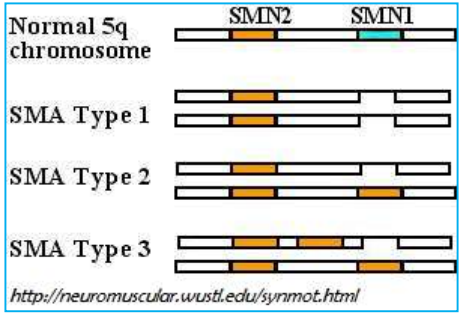


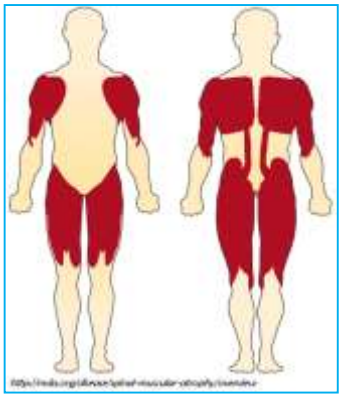
Reclassification of Variants with Nomenclature Change (n=256)



Spinal Muscular Atrophy







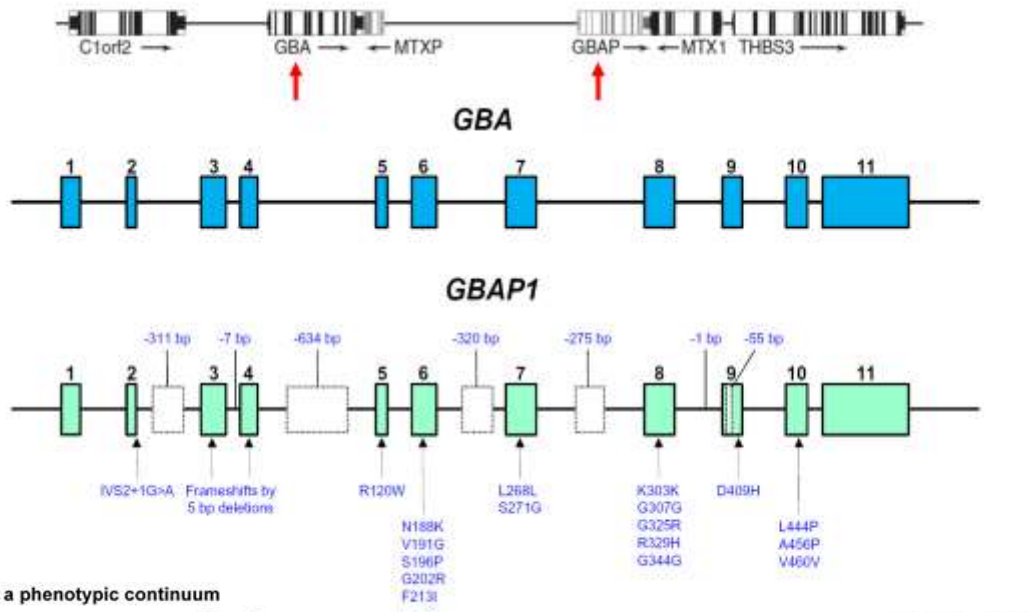
Type	Usual age of onset	Characteristics
I: Infantile	0–6 months	• manifests in the first months of life • "floppy baby syndrome" • rapid motor neuron death (especially of the respiratory system) • pneumonia-induced respiratory failure is the most frequent COD • do not generally live past two years of age

II: Intermediate	6–18 months	• never able to stand and walk (but are able to maintain a sitting position)
		• progress varies greatly:
		  • some patients gradually grow weaker over time
		  • others show little worsening of the disease manifestations
		• body muscles are weakened, and the respiratory system is a major concern
		• life expectancy is somewhat reduced but most SMA II patients live well into adulthood

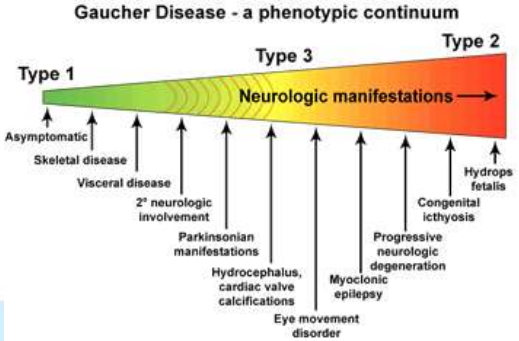
III: Juvenile	>18 months	• able to walk without support at some time (although many later lose this ability) • respiratory involvement is less noticeable • life expectancy is normal or near normal
IV: Adult-onset	Adulthood	• gradual weakening of muscles • frequently requiring the patient to use a wheelchair for mobility • other complications are rare • life expectancy is unaffected

adapted from: http://en.wikipedia.org/wiki/Spinal_muscular_atrophy

Gaucher's Disease: Assessing GBA using SMRT Sequencing



Jeong et al., 2011



A spectrum of presentations with as many as 300 mutations have been observed in GBA and the diversity is what makes clinical presentation so tough to differentiate and drives the need for genetic screening

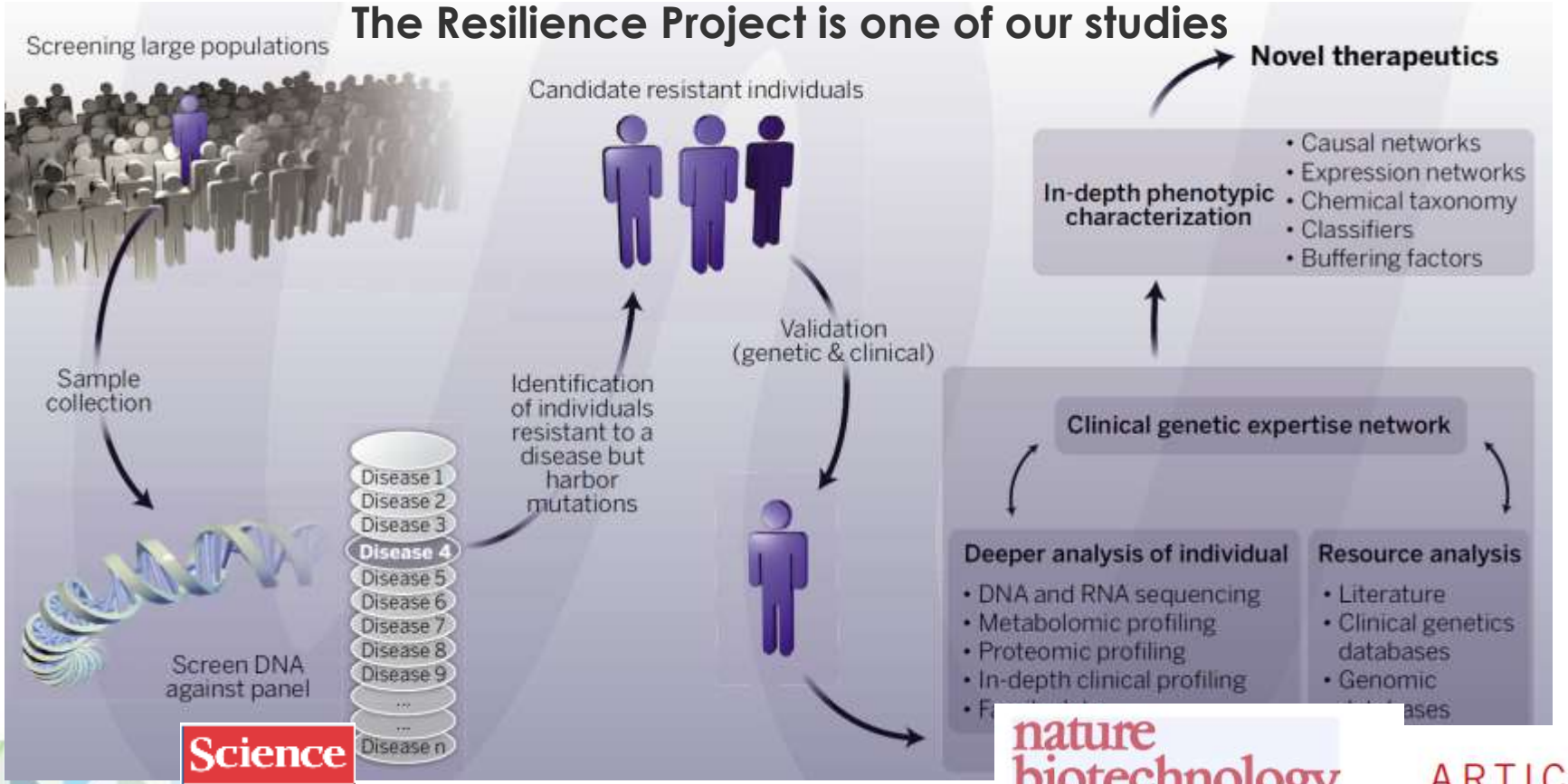
Gaucher's Disease: Assessing GBA using SMRT Sequencing

Sample ID	L29Afs*18 c.84dupG	c.115+1G>A	R159Q, c.476G>A	N409S, c.1226A>G	V433L, c.1297G>T	L422PfsX4, c.1265_1319del	D448H, c.1342G>C	P454R, c.1361C>G	L483P, c.1448T>C	R535H, c.1604G>A
GM00878									HET	
NA20270		HET							HET	
NA08752									HOM	
NA01607				HET	HET					
NA00852	HET			HET						
NA00877									HOM	
NA01260								HET	HET	
NA04394									HET	
NA10870				HOM						
NA10874			HET	HET						
NA20270		HET							HET	
NA20273									HOM	
Internal S1				HOM						
Internal S2				HET						
Internal S3									HET	
Internal S4	HET									
Internal S5							HET			
Internal S6				HET						HET
Internal S7									HET	
Internal S8						HET	HET		HET	
Internal S9 (-)										
Internal S10 (-)										

-12 control cell lines and 8 internal clinical samples confirm 100% of pathogenic variants
-NA20270 repeated twice to show both het calls and general reproducibility
-Two internal normal controls show negative results, as expected

Collab w/ CG Lab / Yao Yang

In fact given wide expressivity of GD, many cases go undiagnosed given inadequate testing

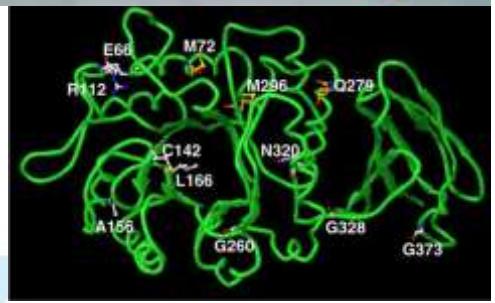


Science
AAAS

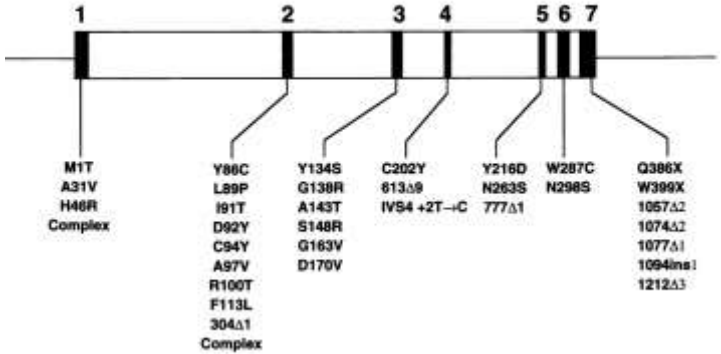
nature biotechnology
ARTICLES

Fabry's Disease Spectrum Due to Varied Mutations in GLA

- Manifests as range of systemic symptoms, often misdiagnosed in childhood/early adults
- Deficiency of alpha galactocidase (GLA) results in insufficient lipid metabolism; accumulation of lipids causes widespread organ damage
- Spectrum of symptoms and severities due to multiple pathogenic mutations across GLA locus and penetrance (X-linked)

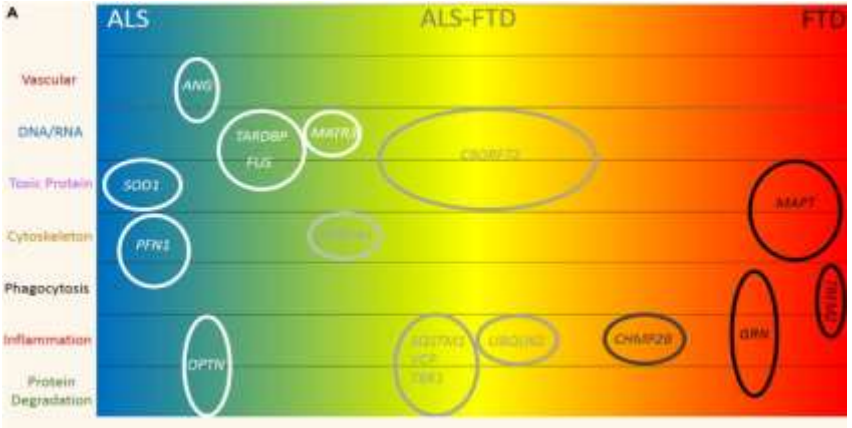


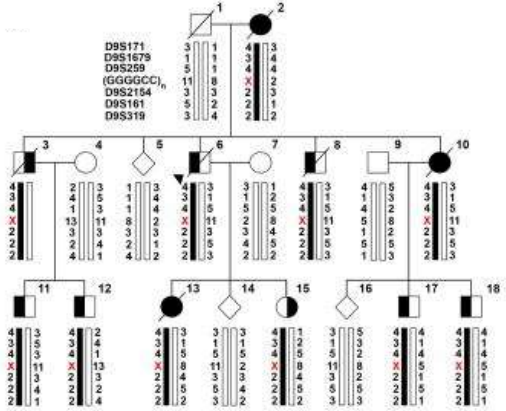
GLA locus: 7 exons; ~10kb
chrX: 101,397,803-101,407,925



GLA variants: Eng, CM. et al. *Mol. Med.* 1997
GLA structural model: Matsuzawa, F. et al. *Hum Genet.* 2005

Resolving Repeat Expansions: Clinical & Pathologic Overlap of ALS and FTD





Heterogeneous clinical presentation in C9orf72 patients

Central hypothesis: **there exists a relationship between the C9orf72 repeat expansion length, sequence, and/or heterogeneity and the clinical manifestation of motor neuron disease across the ALS-FTD spectrum.**

Genetic modifiers primarily contribute to ALS-FTD phenotypic heterogeneity (e.g. *TMEM106B, Ataxin-2*).

Dejesus-Hernandez et al. Neuron 2011
Radford et al. Front. Cell Neurosci 2015.

Loci w/ Known Disease Association That Require Advances

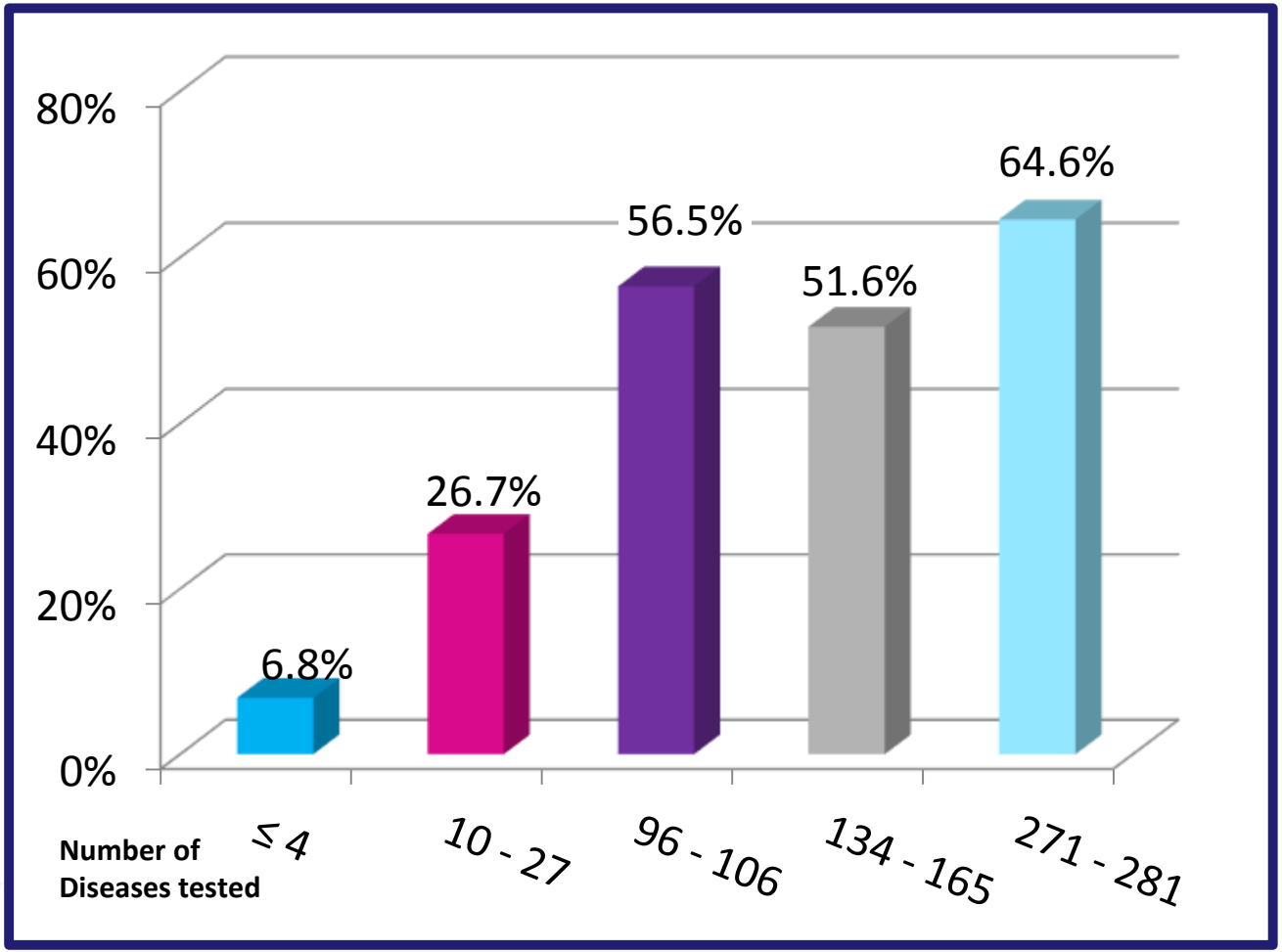
Disease Association	Gene(s)	Incl. Intronic Regions	Normal Repeat #	Pathogenic Repeat #
PKU	PAH	121525	NA	NA
Non-PKU HPA	PAH	121525*	NA	NA
PKU	GTPCH1	61144	NA	NA
PKU	PTPRS	182308	NA	NA
PKU	DHPR (QDPR)	51973	NA	NA
PKU	PCD (PCBD1)	6504	NA	NA
CF	CFTR	202881	NA	NA
CF	TGFBR1	50174	NA	NA
CF	CEACAM6	16784	NA	NA
CF	CEACAM3	15222	NA	NA
3-MCC	MCCC1	100857	NA	NA
3-MCC	MCCC2	71416	NA	NA
MCAD	ACADM	63224	NA	NA
VLCAD	ACADVL	8148	NA	NA
SCAD	ACADS	14273	NA	NA
C3	PCCA	441417	NA	NA
C3	PCCB	87590	NA	NA
C3	MUT	32831	NA	NA
C3	MMAA	41772	NA	NA
C3	MMAB	18424	NA	NA
C3	MMACHC	11014	NA	NA
C3	MMADHC	18182	NA	NA
C3	LMBRD1	121309	NA	NA
Krabbe	GALC	155845	NA	NA
Krabbe	PSAP	35071	NA	NA
Pompe	GAA	18323	NA	NA
Gaucher	GBA	10414	NA	NA
Fabry	GLA	10122	NA	NA
NPA/B	SMPD1	4573	NA	NA
DRPLA	ATN1	17858	7 to 25	49 to 88
SBMA	AR	185996	11 to 24	40 to 62
SCA1	ATXN1	462379	6 to 39	39 to 83
SCA2	ATXN2	147462	15 to 29	34 to 59

Disease Association	Gene(s)	Incl. Intronic Regions	Normal Repeat #	Pathogenic Repeat #
SCA3	ATXN3	48069	13 to 36	55 to 84
SCA6	CACNA1A	417548	4 to 18	21 to 30
SCA7	ATXN7	138905	4 to 35	34 to 300
SCA17	TBP	18567	25 to 44	45 to 66
DM1	DMPK	12835	5 to 37	>50 to 2000
DM2	CNBP	14438	<27	75 to 11,000
EPM1	CSTB	3933	2 to 3	30 to 75
FXS	FMR1	39176	6 to 52	55 to 2000
FRAXE MR	AFF2 or FMR3	500054	6 to 25	>200
FRA12A MR	DIP2B	243682	6 to 23	unknown
FRDA	FXN	64919	7 to 22	66 to 900
SCA10	ATXN10	173509	10 to 29	280 to 4500
SCA8	ATXN8 or ATXN8	24333	6 to 37	107 to 250
HDL-2	JPH3	96321	<50	>50
SCA12	PPP2R2B	496411	<66	>66
HD	HTT	169268	10 to 35	>35
SMA	SMN1	19654	1 to 5	0
SMA	SMN2	19654*	0 to 5	NA
FXTAS	FMR1	39176*	6 to 52	55 to 2000
HLA	HLA-B	3316	NA	NA
Celiac disease	HLA-DQA1	11190*	NA	NA
Celiac disease	HLA-DQB1	11190	NA	NA
Narcolepsy	HLA-DQB1	8916	NA	NA
Ankylosing spond	HLA-B	3316*	NA	NA
6-Mercaptopurine	TPMT	26763	NA	NA
Irinotecan	UGT1A1	13051	NA	NA
Warfarin	CYP2C9	50732	NA	NA
Warfarin	VKORC1	5138	NA	NA
Tricyclic Antidepre	CYP2D6	4407	NA	NA
Abacavir	HLA-B	3316*	NA	NA
Carbamamazepin	HLA-B	3316*	NA	NA
Ximelagatran	HLA-DRB1	34102	NA	NA
Ximelagatran	HLA-DQA1	22353	NA	NA

-We need to gain a better understanding of pseudogenes, repeat expansions, and polymorphic loci for disease diagnostics

-Every individual has ~1000 SVs that are > 2500bp in length, requiring long read sequencing

Breakdown of positive rates of all panels

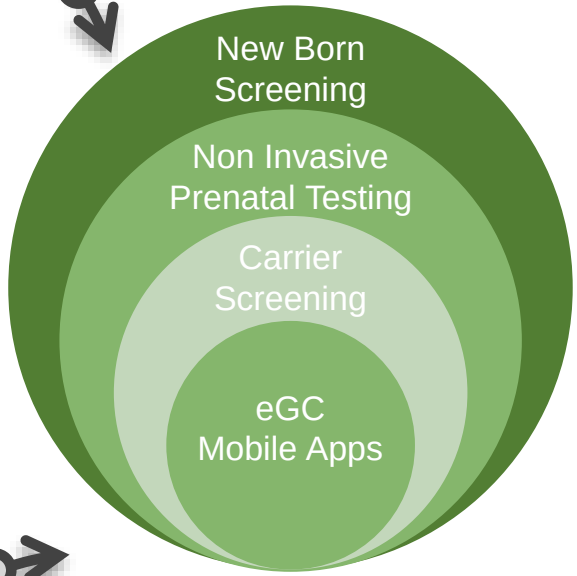


Direct to Consumer

Helix Carrier Screening

NBS from Dx pathways

Search for Resilience



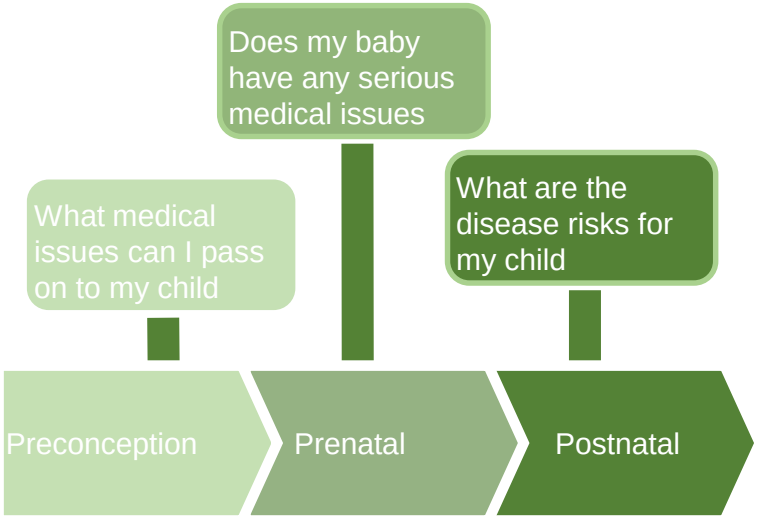
DT Medical Systems

NBS Dx Pathway for Pharma

Current Carrier and NiPT DTP

Clinical Genetic Testing

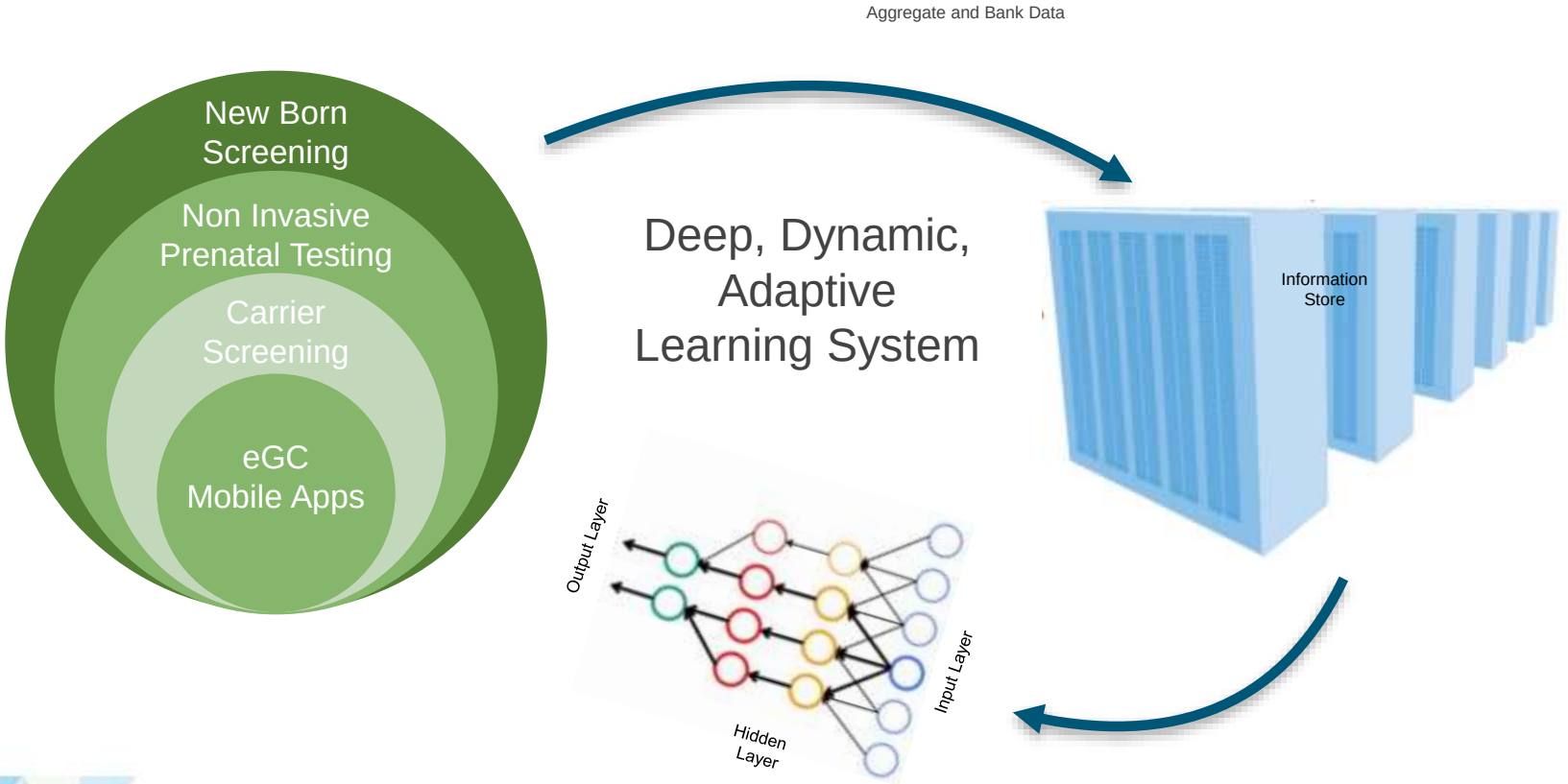
General surveillance and managing wellness



Web & Mobile Apps

Managing pregnancy and postnatal journey

Creating a feedback loop to continually refine test interpretations and expand test utility



Product Roadmap

TODAY

RH Dx Test

- NextStep
- NIPT
- ETL
- Clinical Reporting
- EMR Integration

Q4 2017

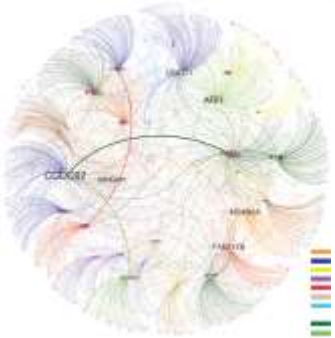
eGC/Onc

- Multiple onc panels
- Heritable cancers
- Clinical Reporting
- Electronic patient engagement (eGC)

2018

V. Res/V. Ge

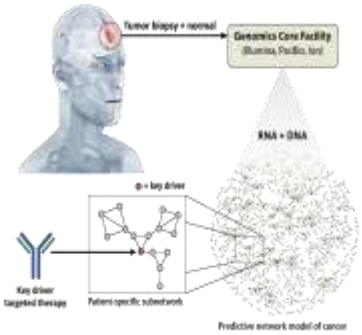
- Information Store
- Health Graph for pharma
- eGC with decision support (virtual geneticist)
- Consumer: Apps and DNA test



2019

Virtual Onc

- Treatment decision support
- Treatment outcomes analytics
- Molecular mechanisms
- Adverse events
- Cohort analytics



sema4
powered by Mount Sinai

Development of SMRT Seq Targeted Assays with Dx potential

Cancer

- Whole gene: p53 GynOnc
- Whole gene: BRCA 1&2
- Whole gene: XPO1, CDK12, EGFR
- FLT3: ITD detection in AML
- **Validation of PCT patient-specific loci**
- Fusion detection w/IsoSeq
- **Inverted PCR to detect damage**

Immunology

- HLA Class I
- HLA Class II
- Immune Repertoire Profiling
  - **FL TCR**, BCR, scFv (R&D)
- Other SV loci (R&D)

Inherited Disease

- IBD SV Loci (R&D)
- FMR1: FRGX TNR
- c9orf72: ALS G4C2
- mtDNA
- **GLA: Fabry's Disease**
- Large Rearrangements
- Rare Diseases

PGx

- **CYP2D6**
- HLA Class I
- HLA Class II

Pathogen Surveillance

- Full genomes and m6A for microbes
- Target plasmids
- Viral Sequencing (Influenza, HIV, HCV)
- HERV activation in HIV patients
- Metagenomics

Population PGx Studies: CYP2D6 Metabolism

- *CYP2D6* metabolizes ~20-25% of all medications.
  - Antidepressants, antipsychotics, antiarrhythmics, opiates, antiemetics, □-adrenoceptor blockers, tamoxifen, etc.
- >100 alleles identified, including CNVs
- *CYP2D6* SMRT sequencing benefits
  - Novel allele characterization; duplication allele-specific sequencing, genotype phasing

Coverage across CYP2D6

CYP2D6 on Sequel (Oct 2016)

372,143 Mapped Polymerase Reads
12,510 kb Polymerase Read N50

Can **multiplex 384 samples** per Sequel 1M chip (100X/sample)

Collaboration with Stuart Scott and Yao

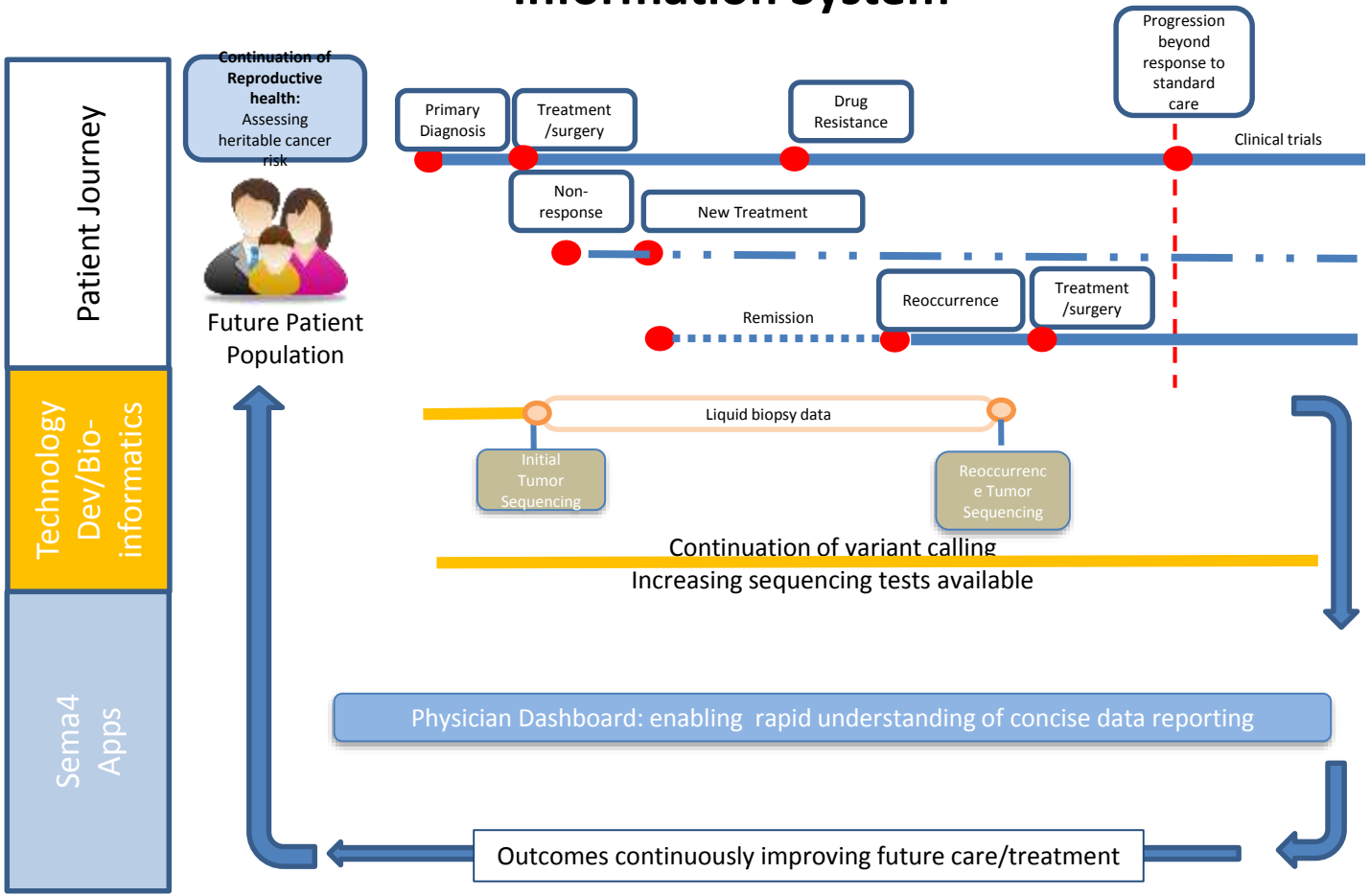
Population PGx Studies: CYP2D6 Utility

	CYP2D6 Genotype		TaqMan Copy Number		SMRT
Samples	Reported [a]	Luminex v3	Intron 2	Exon 9	Genotype
NA17289	*2/*4 (?)	*2/*4	2	2	*2M/*4
NA17084	*1/*10 (?)	*1/*10	3	2	*1A/*36-*10B
NA17252	*4/*5 (?)	*4/*5	1	1	*4/*5
NA17244	*2A/*4,DUP (?)	*2/*4,DUP	4	4	*2Mx2/*4x2
NA17287	*1/*1(*36/?)	*1/*1	2	1	*1A/*83
NA09301	DUP (?)	*1/*2,DUP	3	3	*1A/*2x2
NA17218	*2/*2(*35)	*2/*35	2	2	*2M/*35
NA17213	*1/*2(*35)	*1/*35	2	2	*1A/*35
NA17256	*2(*35)/*2(*35)	*35/*35	2	2	*35/*35
NA17243	*2(*35)/*4	*4/*35	2	2	*4/*35 + 2D7
NA17261	*2(*35)/*4	*4/*35	2	2	*4/*35
NA17119	*1/*2 (?)	*1/*2	2	2	*1A/*2M
CAUC073	-	*8/*10(*29?)	2	2	*8/*10B

Qiao et al, *Human Mutation*. Nov. 2015

***CYP2D6* reference genotype data:** Pratt V, et al. *J Mol Diagn*. 2010

Milestone Connection Between Oncology and Information System

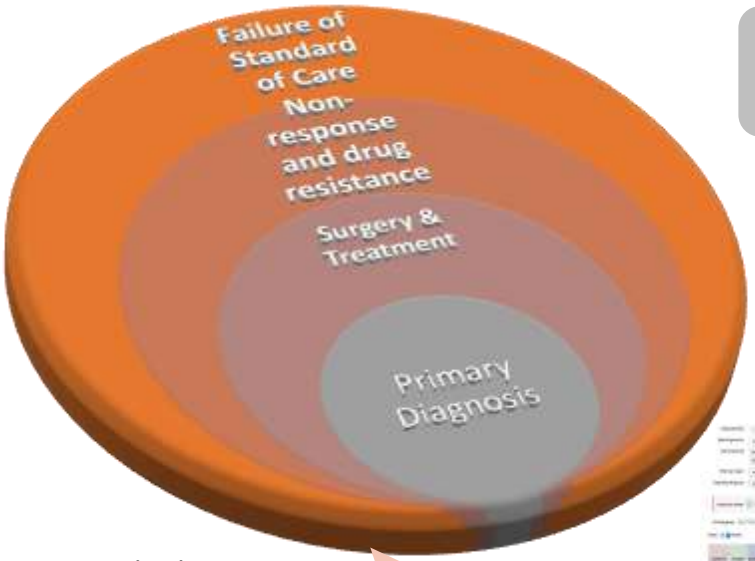


Direct to Consumer

Personal odyssey cases

Referrals from practices

Patient advocacy groups



DT Medical Systems

Pharma clinical trials

Current testing business

Molecular Genetic Testing

Disease surveillance and managing active disease

How has my tumor change? What drug now?

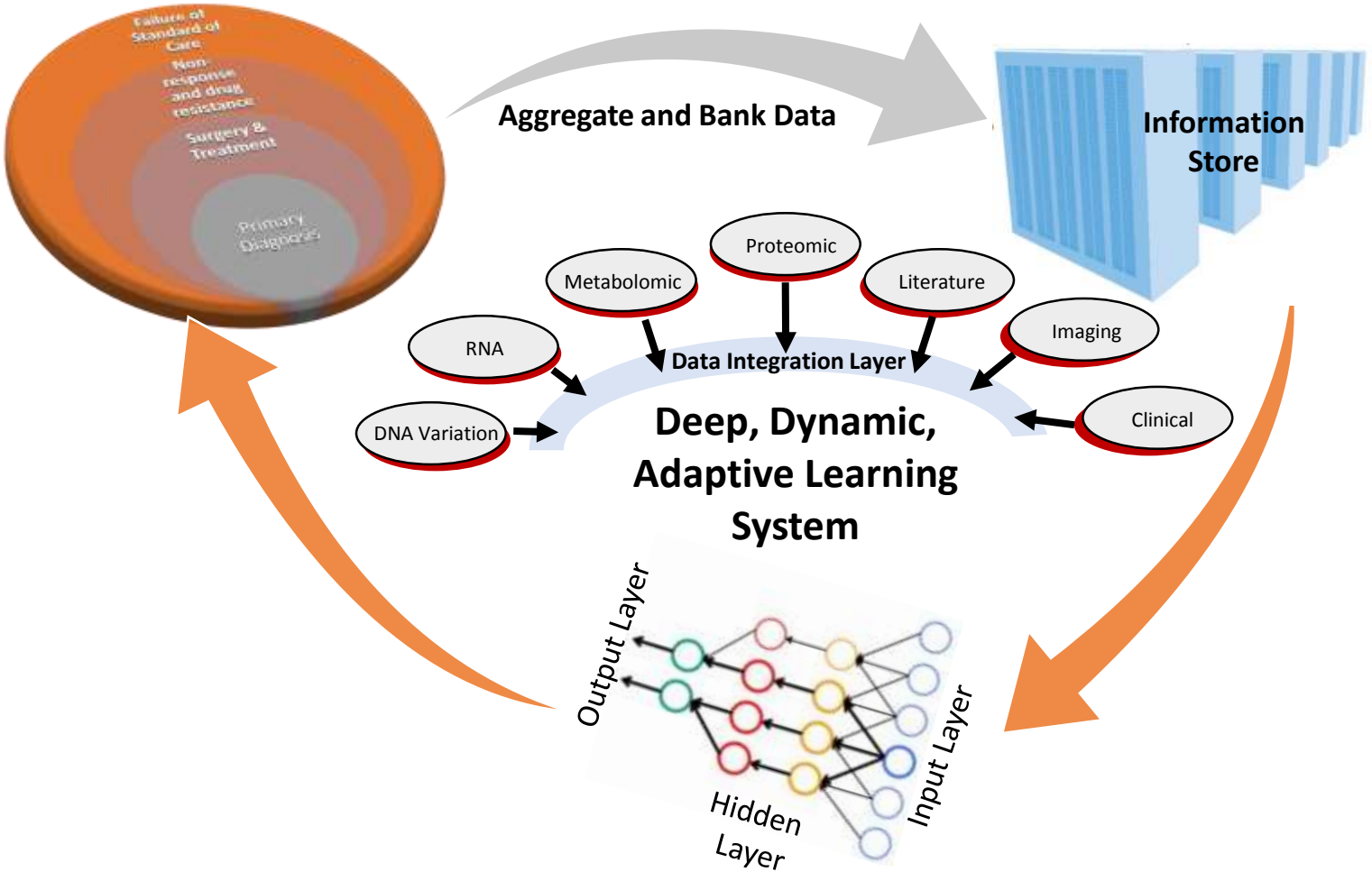
What treatment is best for me

What clinical trials or non-standard Tx?

Web & Mobile Apps

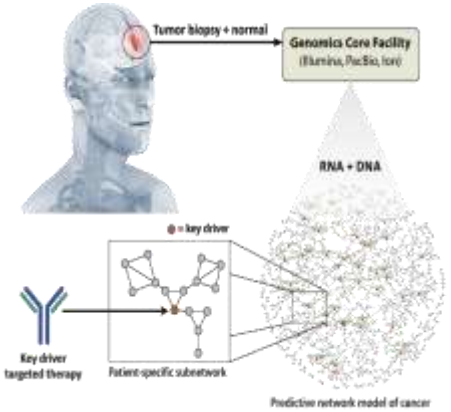
Helping patient navigate journey to cure

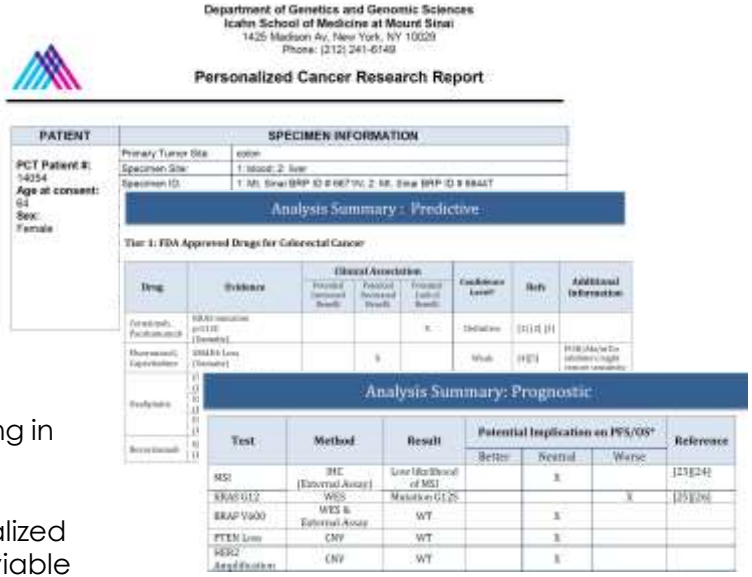
Creating the virtual oncologist of the future



Current Clinical Grade Molecular Oncology Services (Comprehensive Tumor/Germline Characterization)

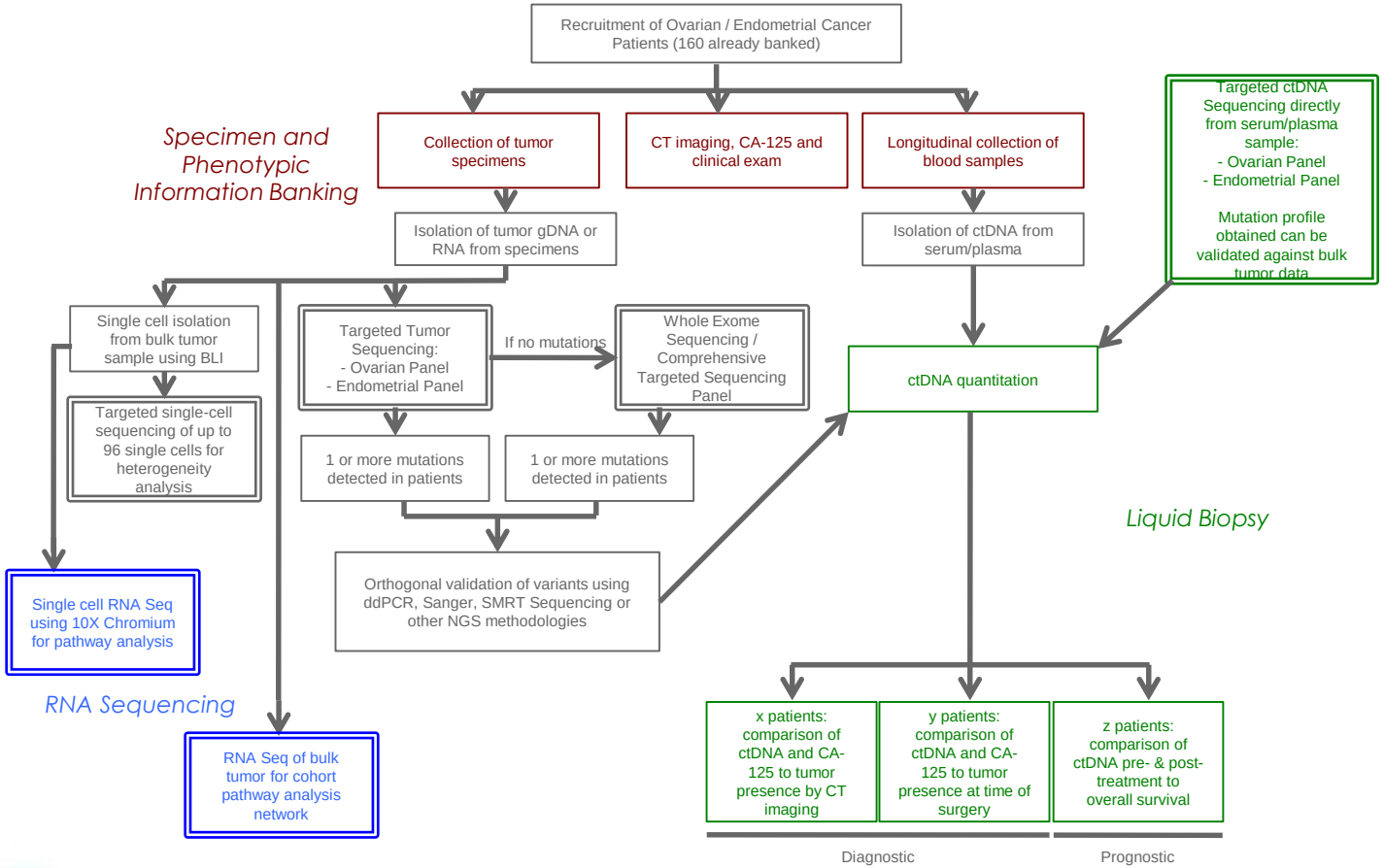
- Approved by New York state for multiple types of cancer
- Expanding panels in collaboration with Life Tech (Oncomine)
- Model-based interpretations that include pan-omic data





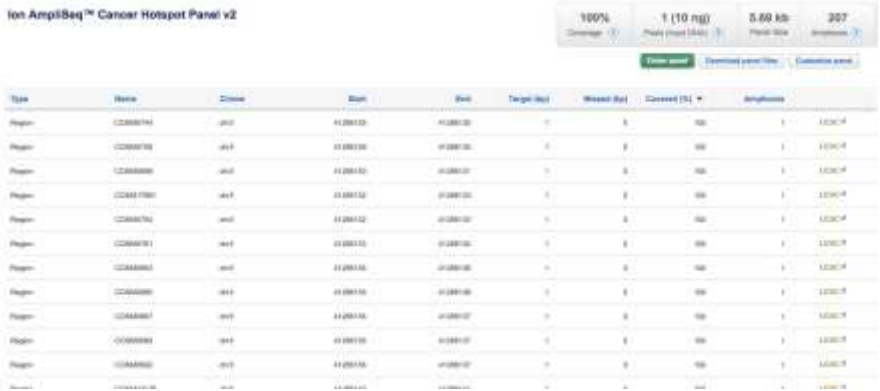
- Employing the model driven decision making in these cancer panels
- Some progress made in scaling our personalized cancer therapy to become commercially viable

Cancer Genomics Requires Integrated Technologies



Collab w/ Martignetti, Dottino, Reva

Cancer Hotspot Panel v2 (207 Amplicons) Validation



ABL1	EZH2	JAK3	PTEN
AKT1	FBXW7	IDH2	PTPN11
ALK	FGFR1	KDR	RB1
APC	FGFR2	KIT	RET
ATM	FGFR3	KRAS	SMAD4
BRAF	FLT3	MET	SMARCB1
CDH1	GNA11	MLH1	SMO
CDKN2A	GNAS	MPL	SRC
CSF1R	GNAQ	NOTCH1	STK11
CTNNB1	HNF1A	NPM1	TP53
EGFR	HRAS	NRAS	VHL
ERBB2	IDH1	PDGFRA	
ERBB4	JAK2	PIK3CA	

Variants are annotated back to loci in variant .VCF files

Library type: AmpliSeq

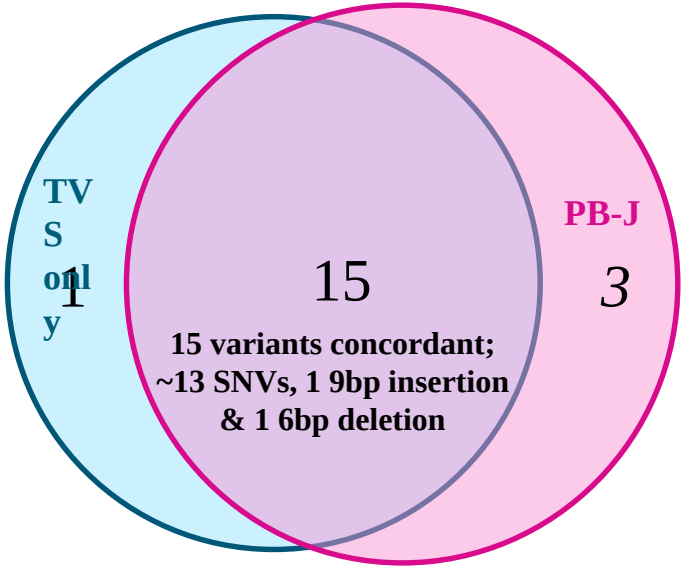
Targeted regions: 4475346_CHP2_designed_20120806

Hotspot regions: 4475346_CHP2_hotspots_20120927

Configuration: Somatic - Low Stringency

Barcode Name	Sample Name	Variants	Hotspot Variants	Download Links
IonXpress_001	none	13	2	[VCF.GZ] [VCF.GZ.TBI]
IonXpress_002	none	14	2	[VCF.GZ] [VCF.GZ.TBI]
IonXpress_003	none	14	1	[VCF.GZ] [VCF.GZ.TBI]
IonXpress_004	none	12	1	[VCF.GZ] [VCF.GZ.TBI]
IonXpress_005	none	20	4	[VCF.GZ] [VCF.GZ.TBI]
IonXpress_006	none	13	2	[VCF.GZ] [VCF.GZ.TBI]

SMRT-Sequencing Validation using Juliet vs Torrent Data



-3 Additional variants are in downstream regions of the amplicon that would have likely fallen victim to edge effects with the short read data. (i.e. – the HotSpot amplicon wasn't designed to detect them properly)

-1 TVS-only variant is completely absent though a separate minor allele is detected by Juliet and is known OMIM variant. Negative call needs validation.

PB Barcode	Chr	Primer	HG19 POS	HotSpot Variant	HotSpot AF	PB coverage	Juliet Variant	Juliet AF	Notes
1	chr7	BRAF_2	140753355	c.1780 G>A	30.09%	1175X	G>A	15.5%	concordant
2	chr7	BRAF_2	140735354	c.1781 A>G	4.80%	960X	A>G	3.4%	concordant
3	chr3	CTNNB1_1	41224633	c.121 A>G	24.32%	872X	T>C	26.5%	concordant; compliment
4	chr3	CTNNB1_1	41224633	c.121 A>G	48%	716X	G>T	3.4%	discordant for HotSpot; **novel variant** downstream at 3.4% is known OMIM variant
5	chr7	EGFR_6	55181312	c.2302 G>T	40.64%	1083X	C>A	39.7%	concordant; compliment. **Novel variant** 10bp downstream; C>T at 39.7%
6	chr7	EGFR_6	55181370	G>A	44.05%	798X	G>A	56.5%	concordant
7	chr7	EGFR_6	55181329	c.2319_2320ins9	32.27%	980X	T>C	40.7%	~50% of sequence shows 9bp insertion; **novel T>C variant** seen just upstream of insertion
8	chr4	FBXW7_2	152329731	c.1177 C>T	18.32%	1124X	G>A	16.5%	concordant; compliment
9	chr20	GNAS_1	58909365	c.601 C>T	9.84%	1014X	C>T	10.2%	concordant
10	chr4	KIT_4	54727495	T>C	90.69%	1022X	T>C	94.0%	concordant
11	chr4	KIT_4	54727495	T>C	92.89%	875X	T>C	94.3%	concordant
69	chr4	KIT_4	54727491	c.1723_1728del6	83.18%	N/A	N/A	N/A	noted 6bp indel identified via alignment; Juliet being optimized for indels now
12	chr12	KRAS_1	25245350	c.34_35 GG>TT	32.93%	895X	GG>TT	21.3%	concordant
13	chr12	KRAS_1	25245321	c.64 C>A	7.11%	1012X	C>A	7.1%	concordant
14	chr12	PTPN11_2	112489081	c.1505 C>T	20.40%	783X	C>T	19.0%	concordant
15	chr12	PTPN11_2	112489081	c.1508 G>T	3.62%	698X	G>T	3.0%	concordant

What d
patient

c

Molecular Analysis Summary : Tumor Classification

Analysis Summary : Predictive

colon cancer

Clinical Trial Connection

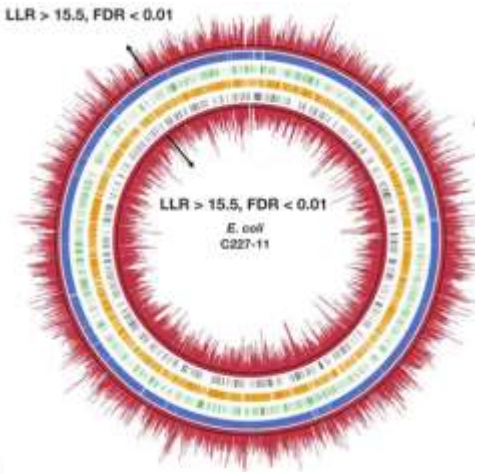
colon cancer

Protocol	Phase	Title	Target	Contact
NCT01750918	Phase I/II	BRAF/MEK/EGFR Inhibitor Combination Study in Colorectal Cancer	BRAF, MEK, EGFR	US GSK Clinical Trials Call Center 877-379-3718 GSKClinicalSupportHD@gsk.com
NCT01347866	Phase I	Clinical Study Of PI3K/mTOR Inhibitors In Combination With An Oral MEK Inhibitor Or Irinotecan In Patients With Advanced Cancer	MEK, PI3K/mTOR	Pfizer CT. gov Call Center 1-800-718-1021
NCT01927341	Phase Ib/II	Phase Ib/II Study of Efficacy and Safety of MEK162 and Panitumumab, in Adult mCRC Patients With Mutant or Wild-type RAS Tumors	MEK	Novartis Pharmaceuticals 1-888-669-6682
NCT02079740	Phase I/II	Trametinib and Navitoclax in Treating Patients With Advanced or Metastatic Solid Tumors	MEK, BCL2	Principal Investigator: Ryan Corcoran Dana-Farber Cancer Institute 617-726-8599 rbcorcoran@partners.org
NCT01351103	Phase I	A Study of Oral LGK974 in Patients With Malignancies Dependent on Wnt Ligands	PORCN (Wnt Signaling pathway)	Novartis Pharmaceuticals 1-888-669-6682

Our Continuing Developments

(we are only now just scratching the surface!!)

Genome-wide DNA methylation surveys



Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing

The methylomes of six bacteria

Methylome Diversification through Changes in DNA Methyltransferase Sequence Specificity

Detecting DNA Modifications from SMRT Sequencing Data by Modeling Sequence Context Dependence of Polym

The complex methylome of the human gastric pat

Methyltransferases acquired by lactococcal 936 rest

ModM DNA methyltransferase methylome analysis reveals a potential role for *Moraxella catarrhalis* ph

Identification of Restriction-Modification Systems of *Bifidobacterium animalis* subsp. *lactis* CNCM I-2494 by SMRT Sequencing and Associated Methylome Analysis

Comprehensive Methylome Characterization of *Mycoplasma genitalium* and *Mycoplasma pneumoniae* at Single-B

Global methylation state at base-pair resolution of the *Caulobacter* genome throughout the cell cycle

Jennifer B. Kozdon[a,b,1], Michael D. Melfi[a,b,1], Khai Luong[c,1], Tyson A. Clark[c], Matthew Boitano[c], Susana Wang[c], Bo Zhou[d], Diego Gonzalez[d], Justine Collier[d], Stephen W. Turner[c], Jonas Korlach[c], Lucy Shapiro[a,2], and Harley H. McAdams[a,2]

CTGCAG

GATC

ACCACC

GGATC

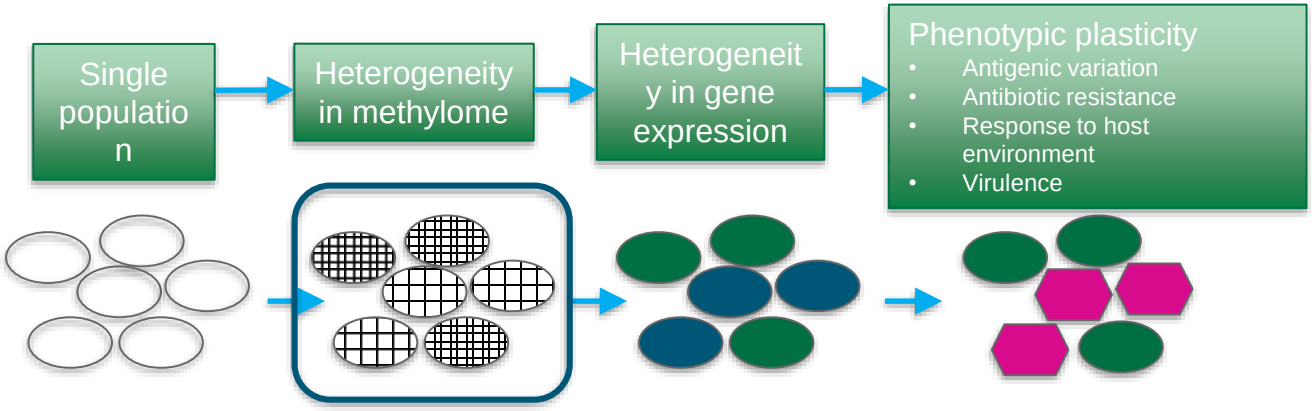
ATGCAT

TCCAGG

GATCC

Why single-molecule detection?

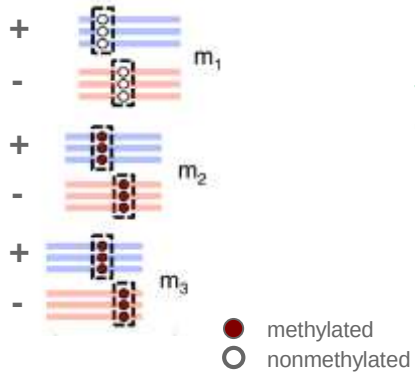
- Heterogeneity in methylome can lead to **phenotypic plasticity**
  - Change phenotype in response to environment
  - For bacteria: subpopulations of cells with distinct phenotypes (phase variation)
  - Fitness advantage in certain environmental conditions



- **Need new methods** to directly observe epigenetic heterogeneity in bacteria

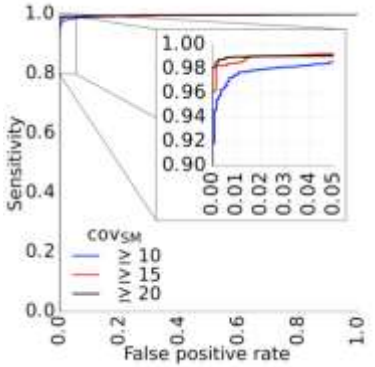
Proposed: two novel **single-molecule methods**

#1: Single molecule, single nucleotide **(SM_{SN})**

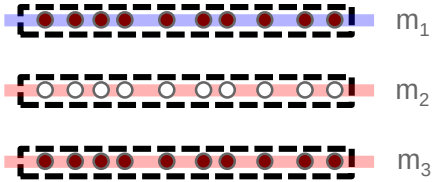


Short libraries (<2kb)

Single site detection

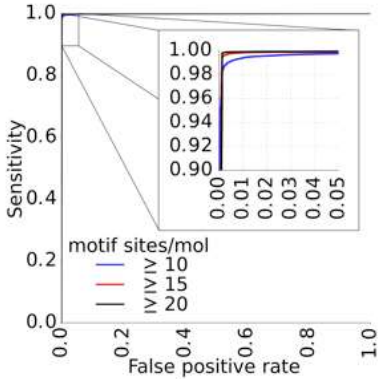


#2: Single molecule, pooled **(SM_P)**

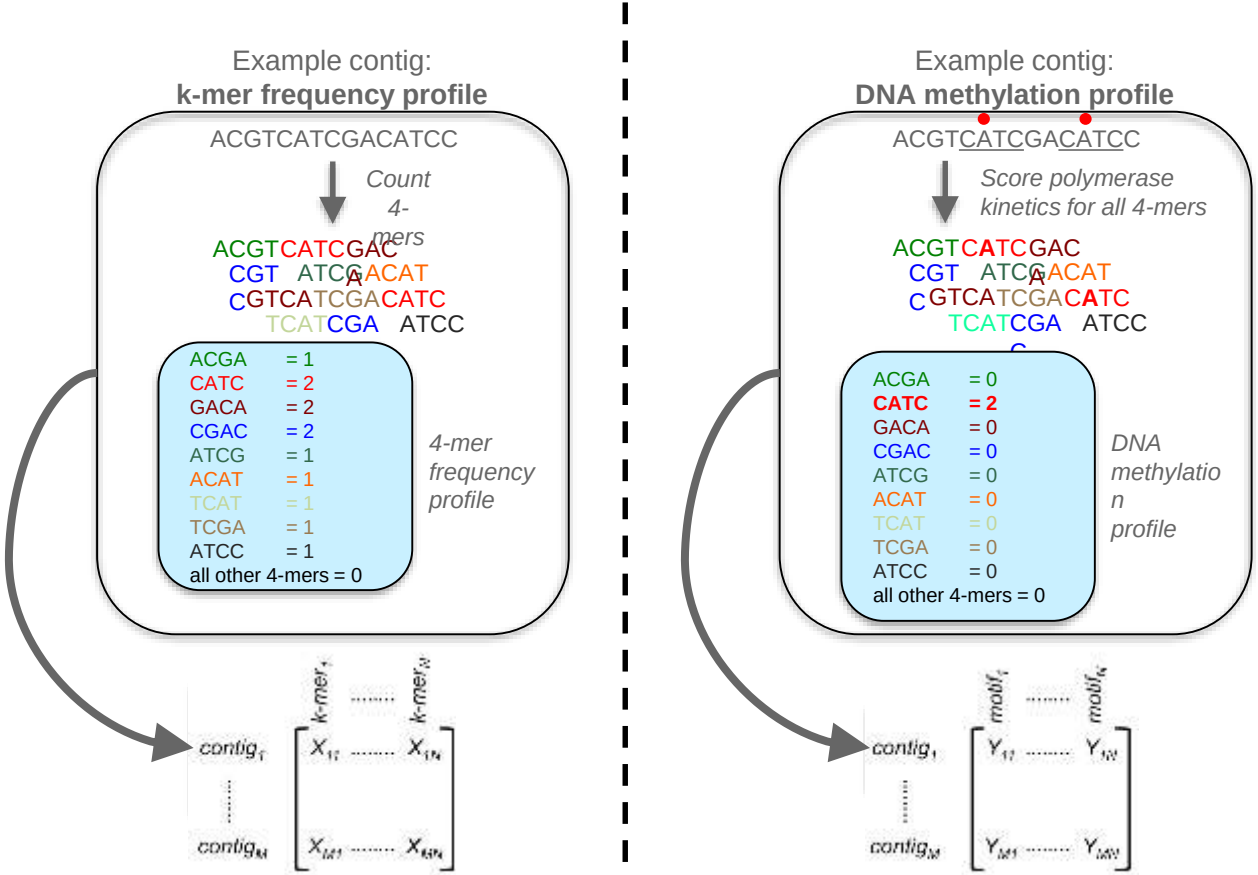


Long libraries (>10kb)

Epigenetic phasing



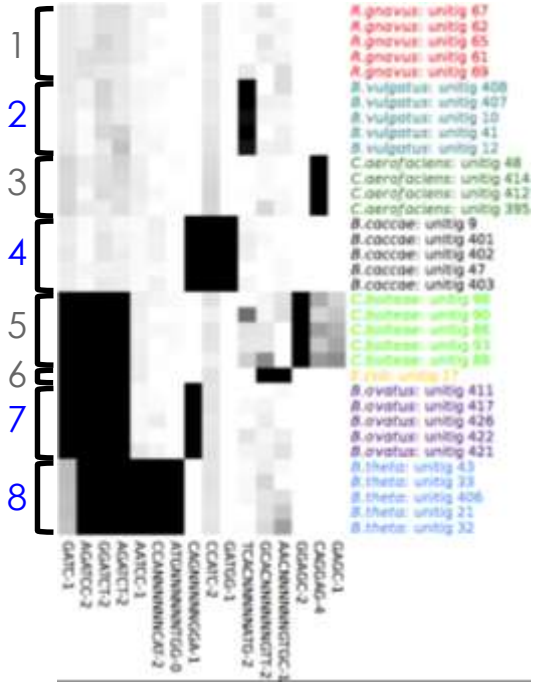
Binning SMRT read-assembled contigs



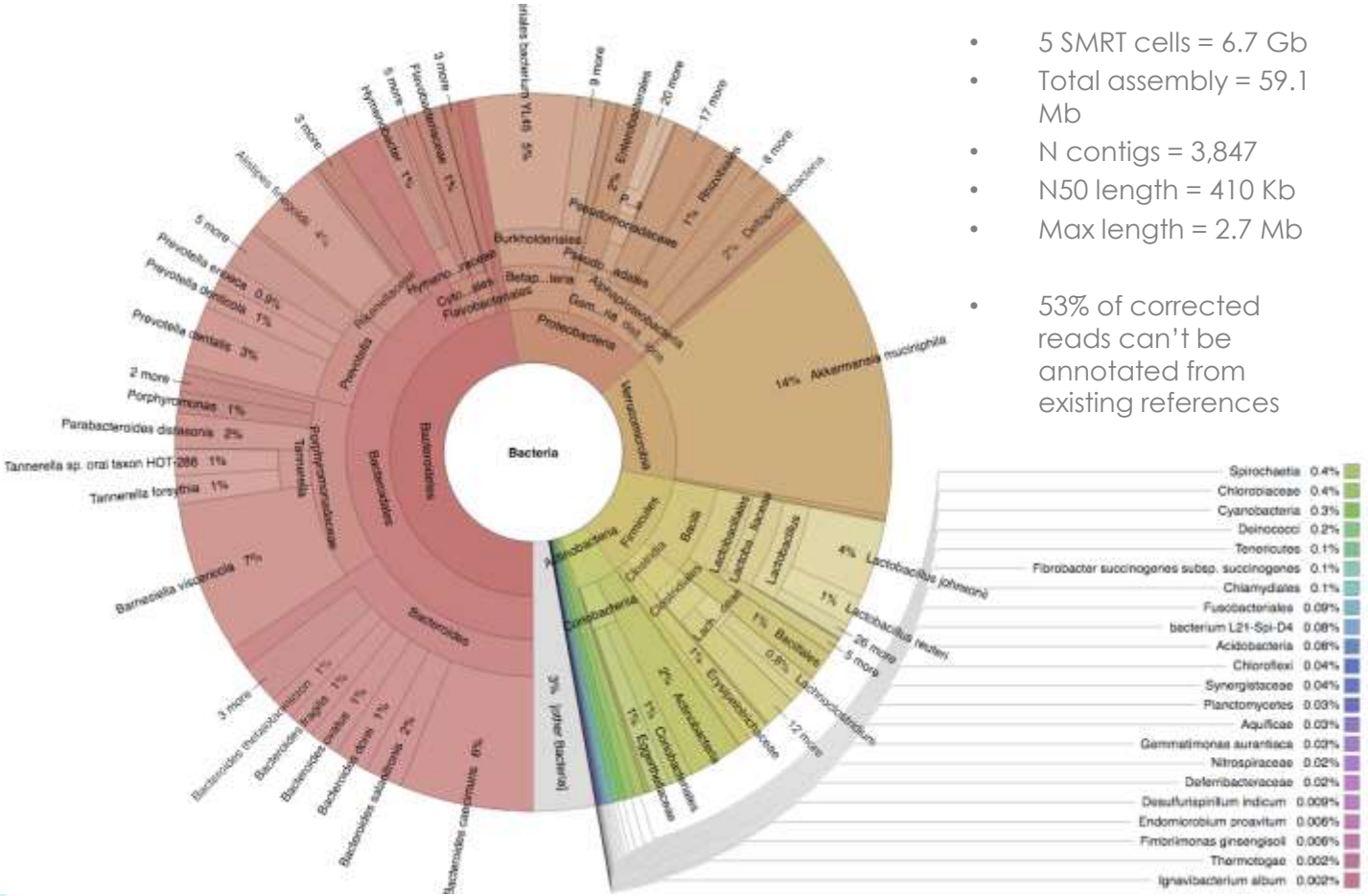
Contigs assembled from SMRT reads

Synthetic mixture of reads from **eight bacterial species**

- Methylation scores from 16 motifs organize contigs by species
- Four species from genus *Bacteroides*
  - Similar sequences
  - Distinct methylation



SMRT sequencing of mouse fecal pellets



- 5 SMRT cells = 6.7 Gb
- Total assembly = 59.1 Mb
- N contigs = 3,847
- N50 length = 410 Kb
- Max length = 2.7 Mb

- 53% of corrected reads can't be annotated from existing references

Realizing an information driven approach to reinventing medicine