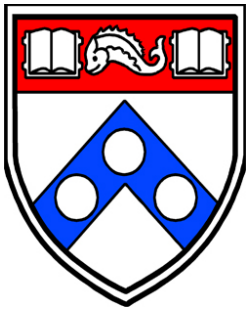
The Children's Hospital of Philadelphia

Discovery Science and Electronic Health Records: Experience from CHOP

Hakon Hakonarson, MD, PhD,
Associate Professor of Pediatrics
Director, Center for Applied Genomics
The Children's Hospital of Philadelphia
University of Pennsylvania, School of Medicine



Genomics and Health Information Technology Systems: Exploring the Issues

NHGRI, April 27-28, 2011, Bethesda MD

Financial & competing interests disclosure

NONE

Center for Applied Genomics

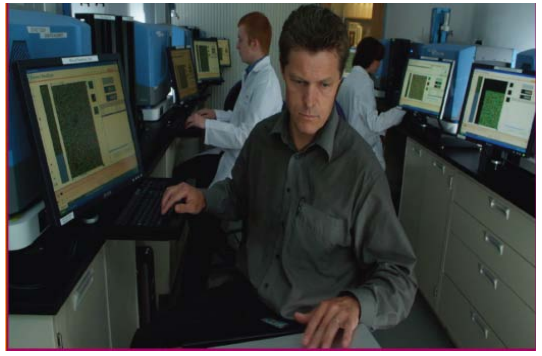
The Children's Hospital of Philadelphia® | RESEARCH INSTITUTE

Impact of gene-centric rare homozygous or compound heterozygous variants in human

- Apart from newborn screening of 30+ conditions, genetic risk assessment is generally only done for individuals who are at increased risk based on family history
- We now have the ability to perform WGS where:
  - Ability to uncover all mono-genic medical conditions and novel mutations
  - All variants that modulate drug response can be identified
  - All variants involved in complex disease traits can be profiled to establish risk
- Where is the medical community in terms of taking this on?

The Center for Applied Genomics at CHOP

- **Founded in June 2006**
- **Staff of 86**
- **Over 30 active disease projects with CHOP/Penn collaborators**
- **TARGET: Genotype 100,000 children**
  - Over 150,000 samples genotyped to date (60k kids)
    - Over 100B genotypes reside in DB
  - IC - participation in future studies >85%
- **Database**
  - Electronic Health Records
  - extensive information on each child
  - 1 million visits per year to CHOP
- **Automation/IT Infrastructure InforSense Analysis Pipeline**



Population Genomics Research
Recruitment of CHOP/PENN HealthCare Network Patients

Autism, Asthma, ADHD, Type 1 Diabetes, IBD, Obesity, Cancer etc.
- all high priority

CAG Repository – complex diseases

Project	Samples
CHOP/CAG	59,787
Children (0-21)	52,764
CAD	38,352
Asthma	12,015
Congenital disorders	9,674
Autism	9,493
Alzheimer's	7,317
IBD	7,021
T1 Diabetes	6,852
Heart Defect	5,743
Cancer	5,240
Cytogenomics	5,117
Neurologic Disorder	4,936
Neuroblastoma	4,901
Lung Cancer	4,589
Developmental Delay	4,189
ADHD	3,675
Obesity	3,089
Endocrine Disorder	2,812
PSP	2,678
Breast Cancer	2,580
Metabolic Disorder	2,171
Cranio Facial Anomaly	2,138
T2 Diabetes	1,978
Addiction/Lung Cancer	1,812
Hematologic Disorder Hereditary	1,789
Cornelia de Lang Syndrome	1,668
Autoimmune	1,474
Leukemia	1,468
Seizures	1,437

Project	Samples
Acute Lung Injury	1,341
Immunodeficiency	1,262
Genetic Anomaly	1,260
Anorexia	1,255
Brain Malformations	1,198
Juvenile Idiopathic Arthritis	1,164
Beryliosis	1,062
Biliary Atresia	828
AA epidemilolgy	811
CerebralPalsy	800
AGS /Clinical/Fetal Malformation	694
Mental Retardation	617
Lupus	572
Schizophrenia	562
Cleft Lip/Palate	498
Liver Transplant	446
Sickle Cell	393
Trisomy 21	385
Growth Failure	379
APoE Cardiac	364
Eosinophilc Enteritis	345
Canine	288
Bone Density	225
Cystic Fibrosis	210
Intestine	101
Polycythemia vera	100
Kidney Stones	73
Hyperinsulinism	47
Brain cancer	40
Total	145,719

- All major pediatric and several adult diseases are represented
- EHR have unlimited potential regarding
  - Birth history (mother/child)
  - Acute/chronic illnesses
  - Medication use and compliance
  - Developmental trajectories
  - AEs/SAEs/DDIs
  - Longitudinal f/u
- We have established over 60 collaborations world-wide for both discovery and replication purposes further stengthening the value of EHRs

CAG Repository – Rare Diseases

EPIC DIAGNOSIS	UNIQUE PATIENTS
ACQUIRED CLUBFOOT,CLUBFOOT,CONGE	24
ACROMEGALY AND GIGANTISM	7
AMYLOIDOSIS	8
ANDROGEN INSENSITIVITY SYNDROME,A	6
ANIRIDIA	8
APERT SYNDROME	2
ARTHROGRYPOSIS	2
ASYMPTOMATIC HEMOPHILIA A CARRIER	2
ATRIAL FIBRILLATION	30
BENIGN ROLANDIC EPILEPSY OF CHILDHC	2
BLEPHAROPHIMOSIS	23
CARPAL TUNNEL SYNDROME	18
CENTRAL PRECOCIOUS PUBERTY	2
CHARGE SYNDROME	11
CHOLESTASIS,CHOLESTASIS OF PARENTE	7
CHOREA NEC,HUNTINGTON'S CHOREA,RI	18
CHRONIC GRANULOMATOUS DISEASE,CG	2
COLOBOMA OF OPTIC DISC,FUNDUS COL	17
CONG ECTODERMAL DYSPLAS,CONGENIT	11
CONGENITAL ECTODERMAL DYSPLASIA,C	6
CONGENITAL FACTOR XI DEFICIENCY	3
CORTICOADRENAL INSUFFICIENCY	14
CRANIOSYNOSTOSIS	6
CRITICAL ILLNESS POLYNEUROPATHY,POI	21
DIGEORGE SYNDROME	2
EMPHYSEMA NEC,EMPHYSEMA (SUBCUT/	15
EXOSTOSES,EXTERNAL AUDITORY CANAL	2
FAMILIAL MEDITERRANEAN FEVER,FMF (	19
FEMALE INFERTILITY ASSOCIATED WITH A	4
FRAGILE X SYNDROME	8
FULMINANT HEPATIC FAILURE,HEPATIC F	2
GAUCHER DISEASE	2
GOUT NOS,GOUT, UNSPECIFIED,GOUTY N	8
GROWTH HORMONE DEFICIENCY	5
HERED SPASTIC PARAPLEGIA,HEREDITAR	9
HEREDITARY FRUCTOSE INTOLERANCE,FR	4
HEREDITARY HEMORRHAGIC TELANGIECT	8
HEREDITARY PERIODIC FEVER SYNDROME	4
HETEROTAXY	3
HIP DYSPLASIA, CONGENITAL,CONGENIT	8

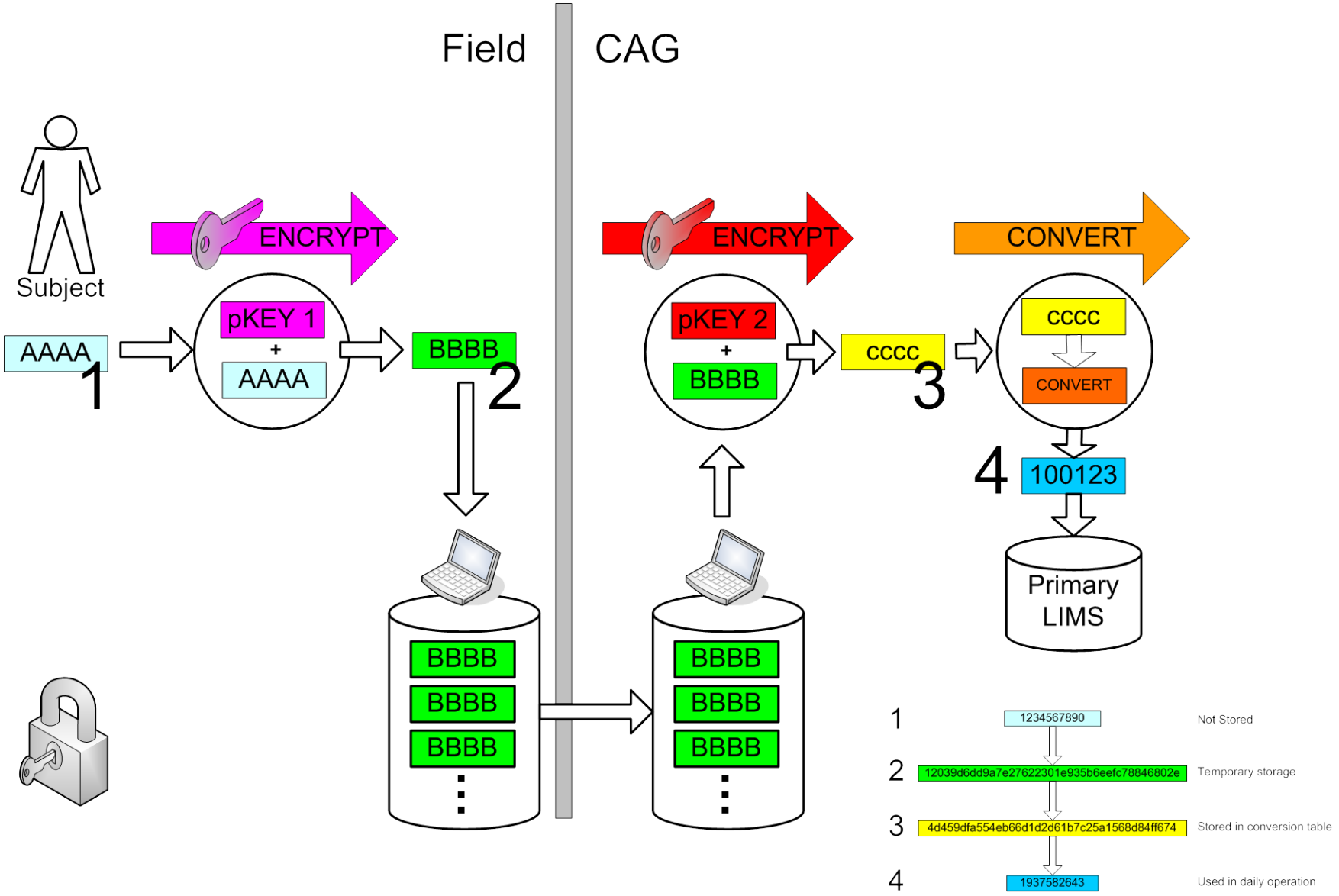
EPIC DIAGNOSIS	UNIQUE PATIENTS
HOMOCYSTINURIA	2
HUS (HEMOLYTIC UREMIC SYNDROME)	2
HYDROPS FETALIS NO ISOIM,HYDROPS FE	20
HYPERALDOSTERONISM,GLUCOCORTICO	2
HYPERANDROGENISM	2
HYPERBILIRUBINEMIA,CONJUGATED HYP	9
HYPERCHYLOMICRONEMIA	2
HYPERPARATHYROIDISM, UNSPECIFIED,S	16
HYPOCALCEMIA AND HYPOMAGNESEMIA	21
ICHTHYOSIS CONGENITA,ICHTHYOSIS	26
IDIO PULM HEMOSIDEROSIS,IDIOPATHIC	5
IDIOPATHIC ANGIOEDEMA,ANGIOEDEMA	11
INDETERMINATE SEX AND PSEUDOHERM	9
INSULIN RESISTANCE	3
JUVENILE MYOCLONIC EPILEPSY,INFANTI	5
KLIPPEL-FEIL SYNDROME	14
LACTASE DEFICIENCY	3
LEUKODYSTROPHY	22
LIPODYSTROPHY	5
LIVER FAILURE, ACUTE,ACUTE LIVER FAILU	6
LOW GRADE MYELODYSPLASTIC SYNDRO	2
MALIGNANT MELANOMA OF SKIN OF LO	7
MALIGNANT MELANOMA OF SKIN OF LO	5
MATERNAL HYPERTHYROIDISM,FAMILY H	4
METHEMOGLOBINEMIA,ACQUIRED METH	6
MICROCEPHALY	10
MICROPHTHALMIA,MICROPHTHALMIA, B	5
MITRAL VALVE PROLAPSE,MVP (MITRAL	2
MUCINOUS CYSTADENOMA OF OVARY,FA	13
MULTIPLE ENDOCRINE NEOPLASIA (MEN	4
MULTIPLE EPIPHYSEAL DYSPLASIA	3
MYASTHENIA GRAVIS WITHOUT (ACUTE)	16
MYASTHENIA GRAVIS WITHOUT (ACUTE)	17
MYASTHENIC SYNDROMES IN DISEASES C	2
MYELODYSPLASTIC SYNDROME, UNSPECI	13
MYOGLOBINURIA	3
NAFL (NONALCOHOLIC FATTY LIVER),FAT	5
NEPHROTIC SYNDROME WITH LESION OF	11
NEUROBLASTOMA,NEUROBLASTOMA OF	9
OPTIC NERVE HYPOPLASIA	22

EPIC DIAGNOSIS	UNIQUE PATIENTS
OSTEOGENESIS IMPERFECTA,OSTEOGENE	19
OSTEOMALACIA NOS,OSTEOMALACIA, U	7
OSTEOPETROSIS	4
OSTEOSARCOMA,OSTEOSARCOMA OF HU	12
OTHER CEREBELLAR ATAXIA,CEREBELLAR	11
OTHER HEART BLOCK,CONGENITAL HEAR	30
OTHER LYMPHEDEMA,OTHER NONINFEC	20
OTHER OVARIAN FAILURE,POSTABLATIVI	15
PERIPHERAL ANGIOPATHY IN DISEASES C	6
PHENYLKETONURIA (PKU),PHENYLKETON	13
POLYCYTHEMIA VERA	12
POLYMICROGYRIA	2
PORENCEPHALY,CONGENITAL PORENCEF	2
POST-INFLAMMATORY HYPERPIGMENTA	15
POSTINFLAMMATORY PULMONARY FIBR	13
PRADER-WILLI SYNDROME	26
PRIMARY CARNITINE DEFICIENCY,CARNIT	7
PSEUDOPOLYPOSIS OF COLON,FAP (FAM	18
PULMONARY ALVEOLAR PROTEINOSIS	2
RETINAL DEGENERATION,UNSPECIFIED P	2
RHABDOMYOLYSIS	28
RHABDOMYOSARCOMA OF FOREARM,RH	3
SARCOIDOSIS	19
SCHIZOPHRENIA NEC-UNSPEC,SCHIZOPH	13
SCREENING FOR GALACTOSEMIA,GALACT	24
SENILE DEMENTIA UNCOMP,DEMENTIA I	4
SENSORY RETINAL DYSTROPHY,PIGMENT	10
SLEEP RELATED LEG CRAMPS,FOOT CRAM	11
SYMPTOM TORSION DYSTONIA,GENETIC	27
THROMBOCYTOSIS	5
TRICHOTILLOMANIA	2
UNSPECIFIED KERATOCONUS,KERATOCO	2
UNSPECIFIED MITRAL AND AORTIC VALVI	3
UNSPECIFIED SPINAL MUSCULAR ATROPH	15
URIC ACID NEPHROLITHIASIS,NEPHROLIT	3
VACTERL ASSOCIATION	7
VENTRICULAR FIBRILLATION	24
VERTIGO,PERIPHERAL VERTIGO NOS,VER	27
VON WILLEBRAND DISEASE	3
WILSON DISEASE	2

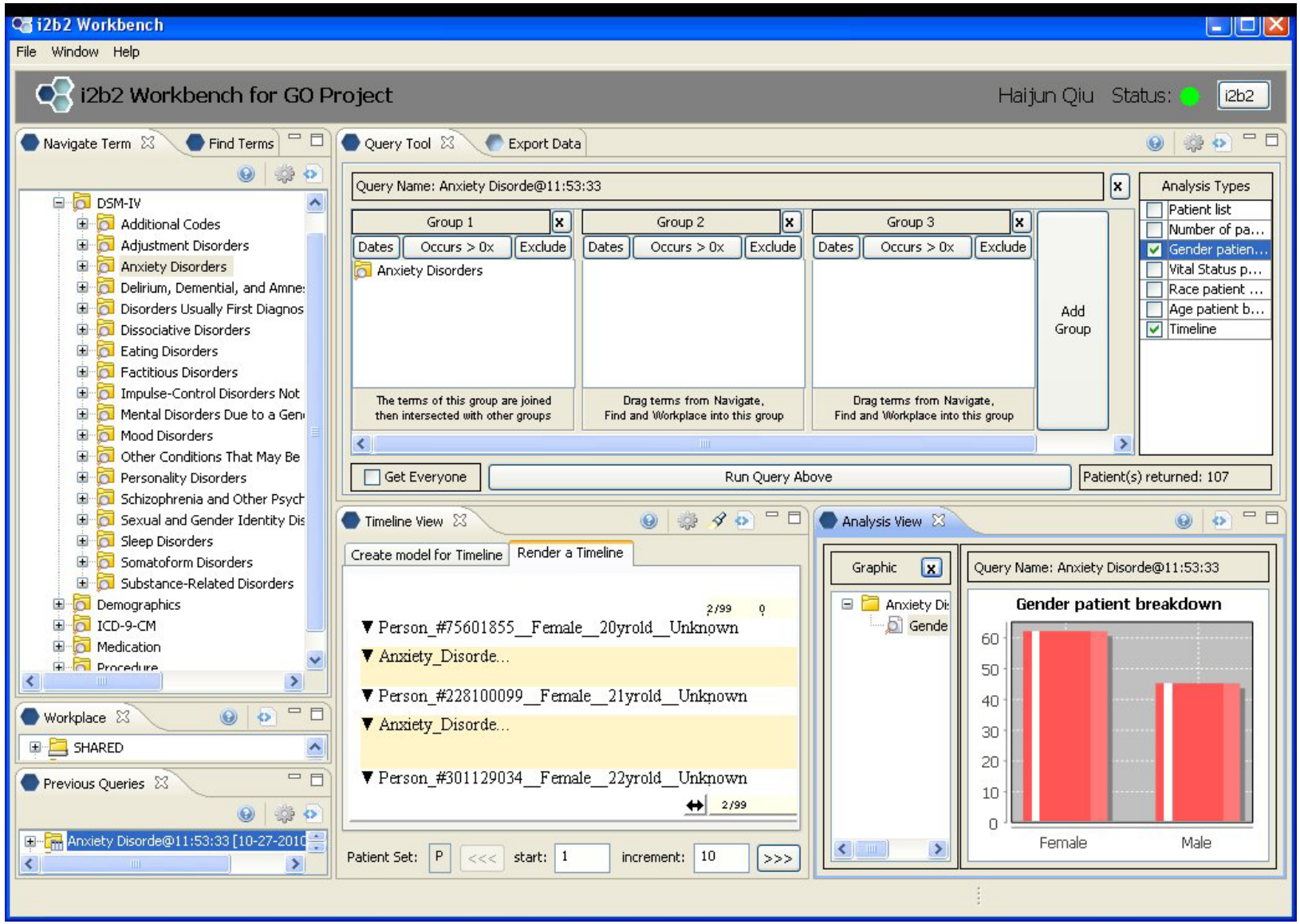
- Numerous rare pediatric diseases – also based on EHR
- Families available
- Mendelian inheritance pattern observed
- Those with known conditions have been tested
- Sequencing is helping resolve

CAG Encryption Process

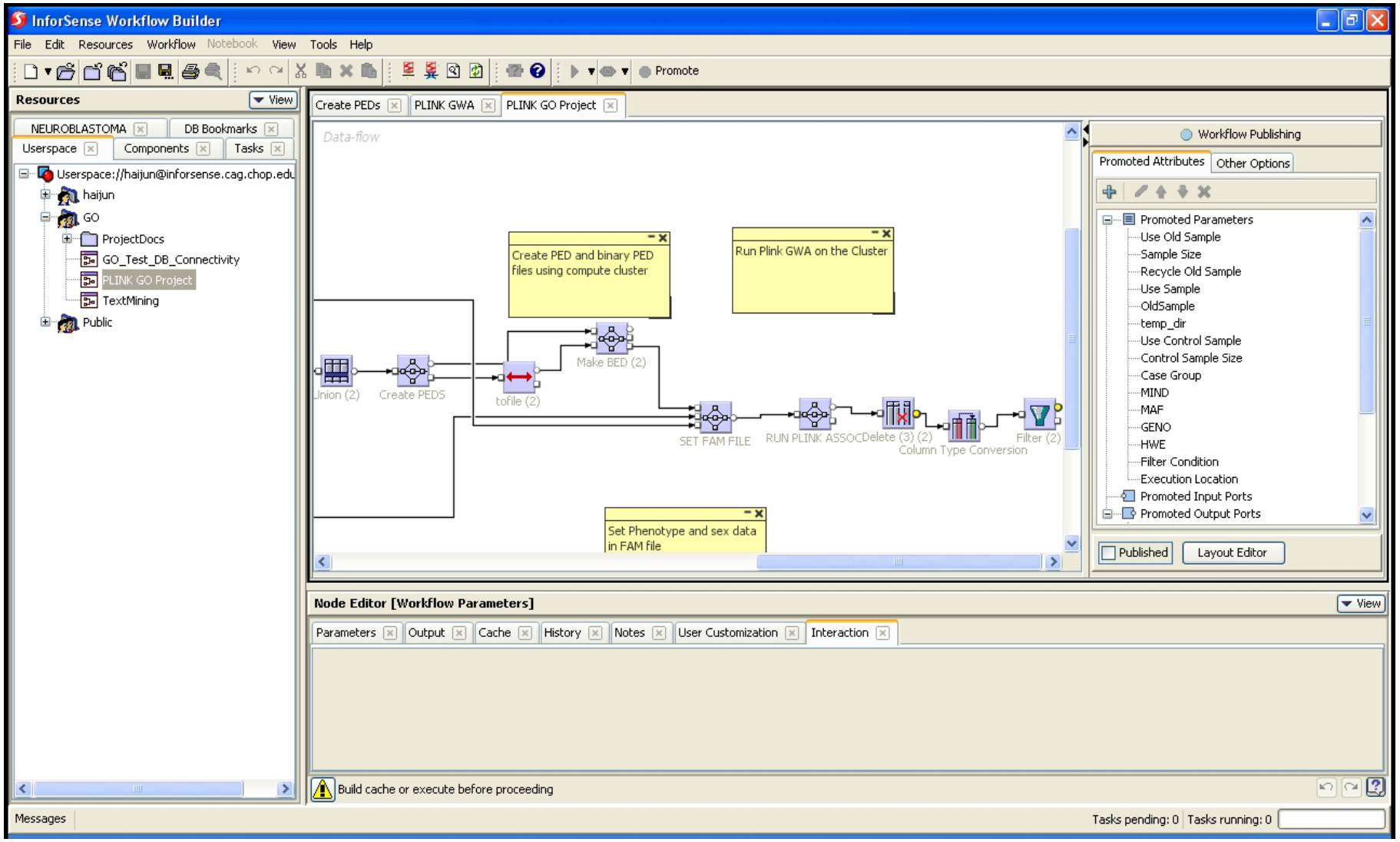
Stringent measures for privacy protection



I2b2 Phenotype Browser



InforSense – analytics workflows



Phenotyping Project: GO Grant-RC2 (NIMH)

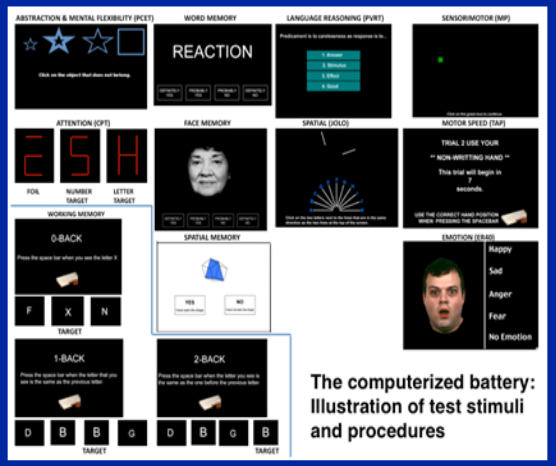
Neurodevelopmental Genomics: Trajectories of Complex Phenotype

- Large-scale phenotyping program at CHOP and Penn (Hakonarson/Gur) **driven by information from EHRs**
- 10,000 genotyped children from CHOP and family members
- Detailed neurocognitive phenotyping (3+ hour battery of testing)
- MRI – both structural and functional
- Methylation profiling
- Questionnaires and assessment – **validation of EHRs**
- Participation in future studies

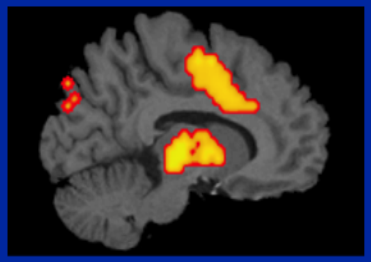
Clinical Assessment

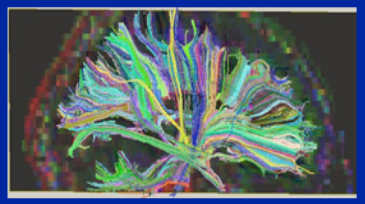
Computerized Neurocognitive Battery (CNB)

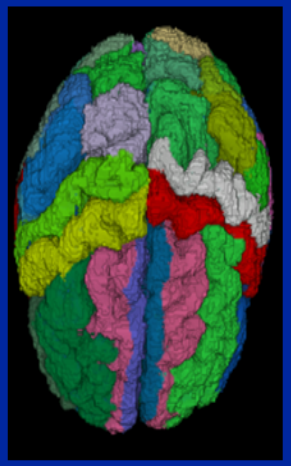
Neuroimaging sMRI, DTI, fMRI



The computerized battery: Illustration of test stimuli and procedures







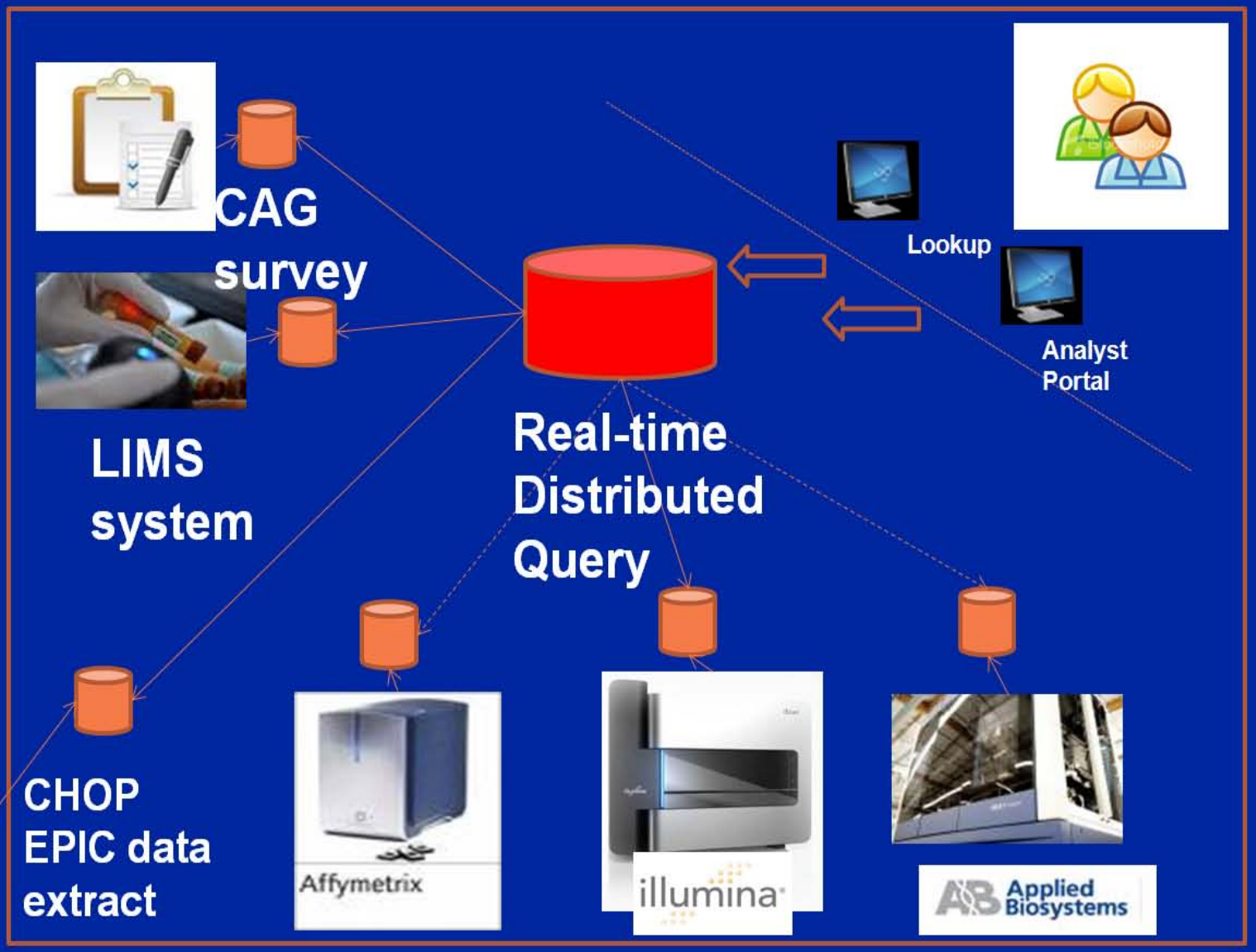
PHENOTYPING

CAG

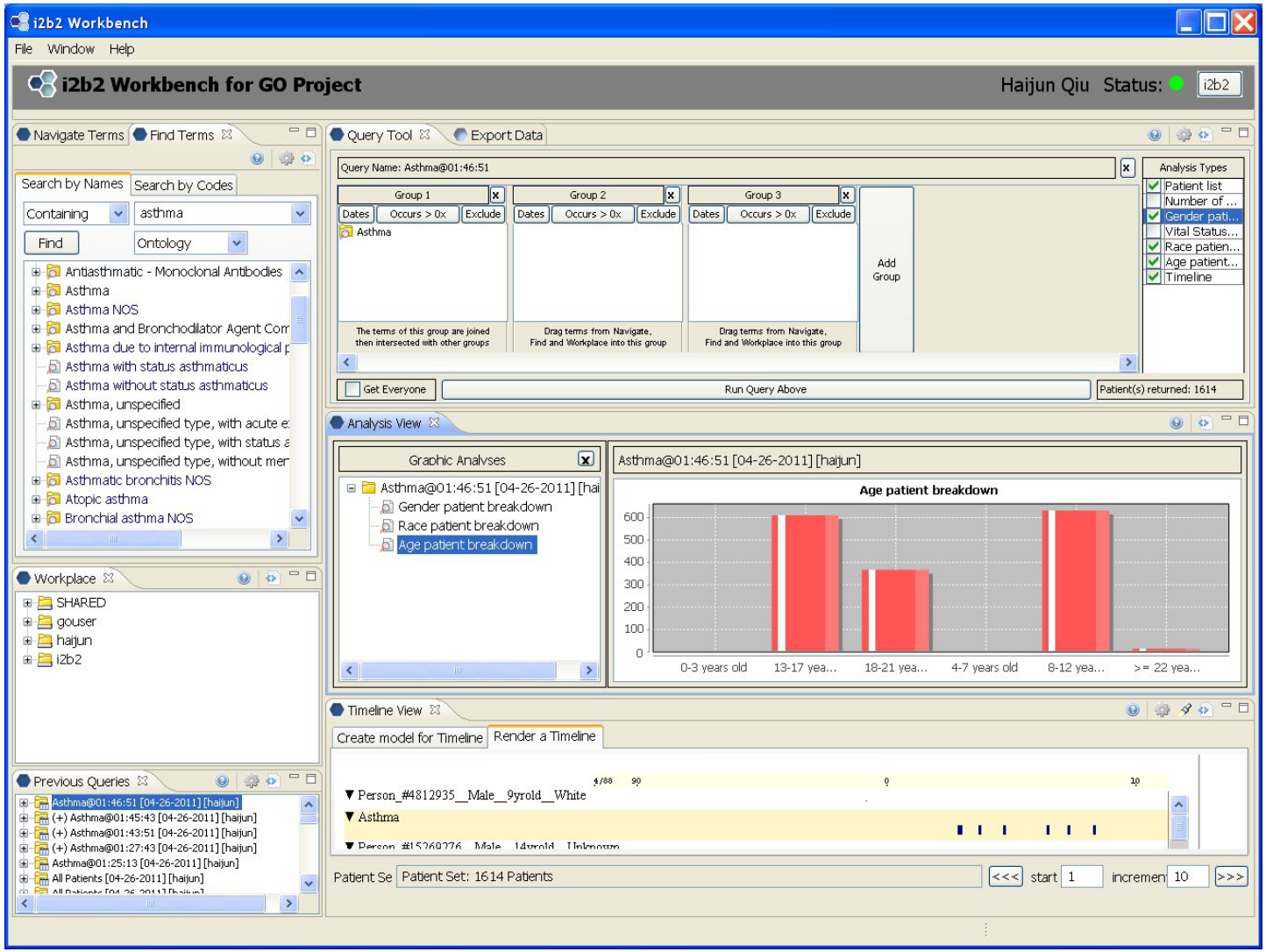
EMR

Center for Applied Genomics

CAG real-time biobank query system

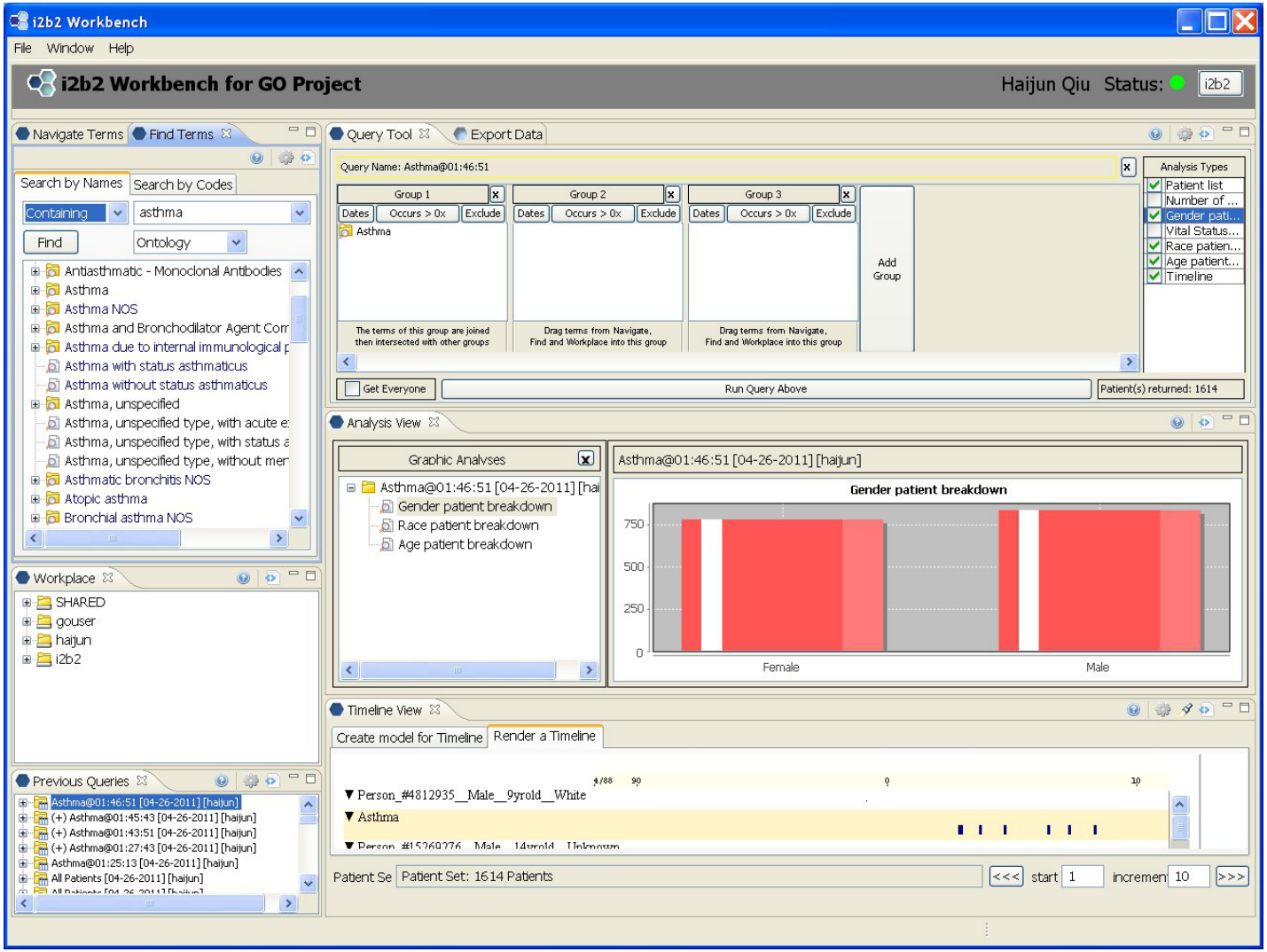


Asthma – age breakdown



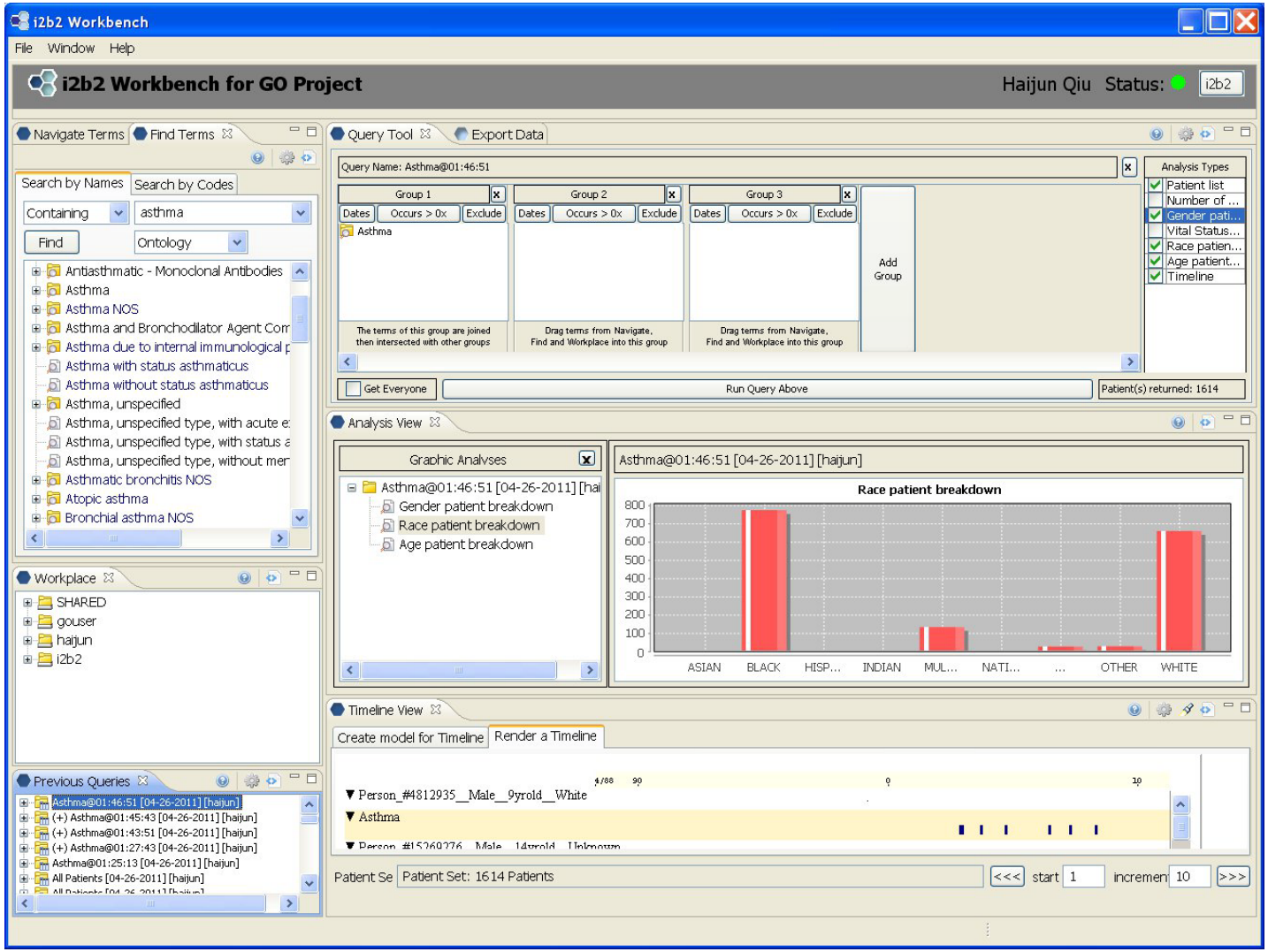
n=1614

Asthma – gender breakdown



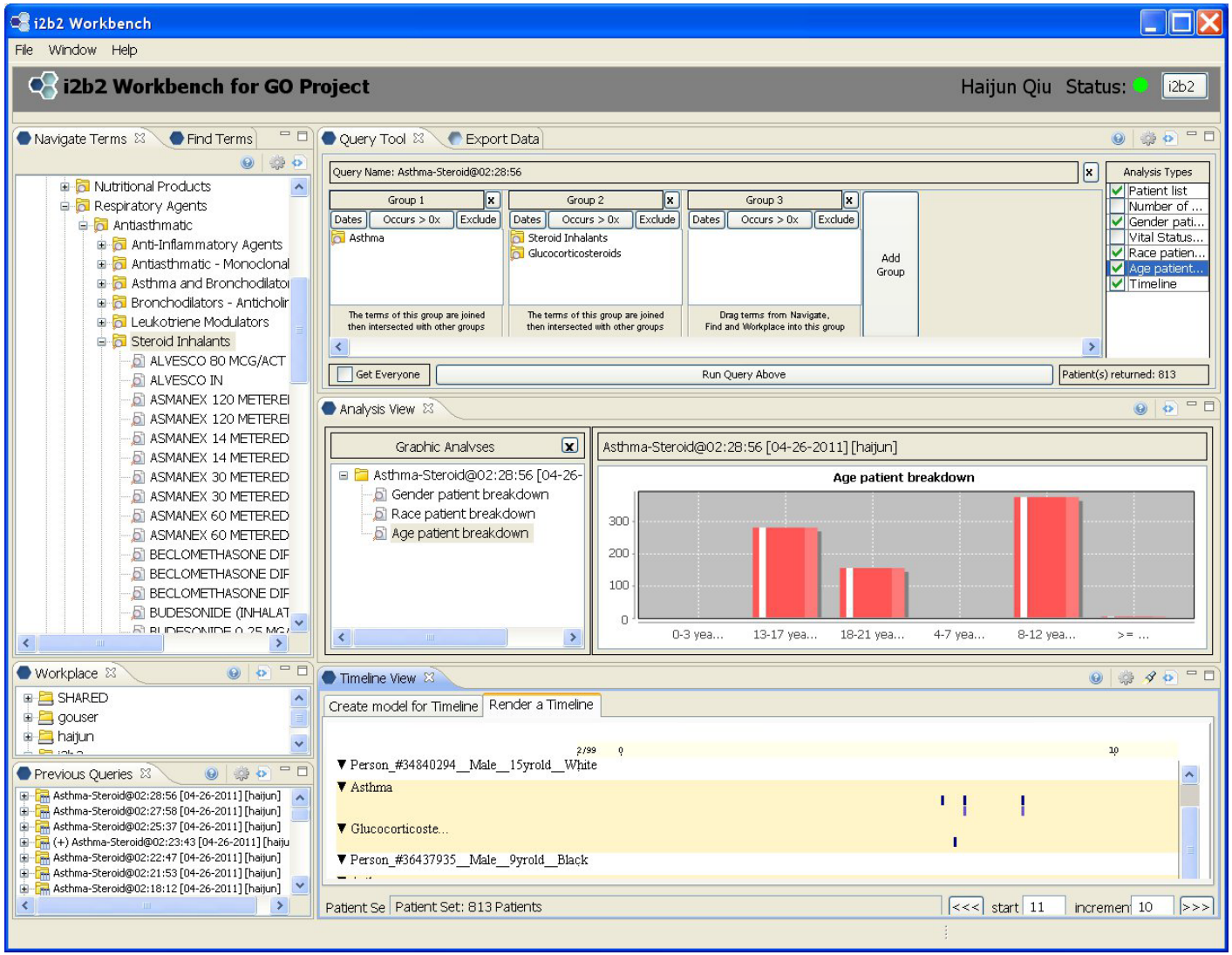
n=1614

Asthma – race breakdown



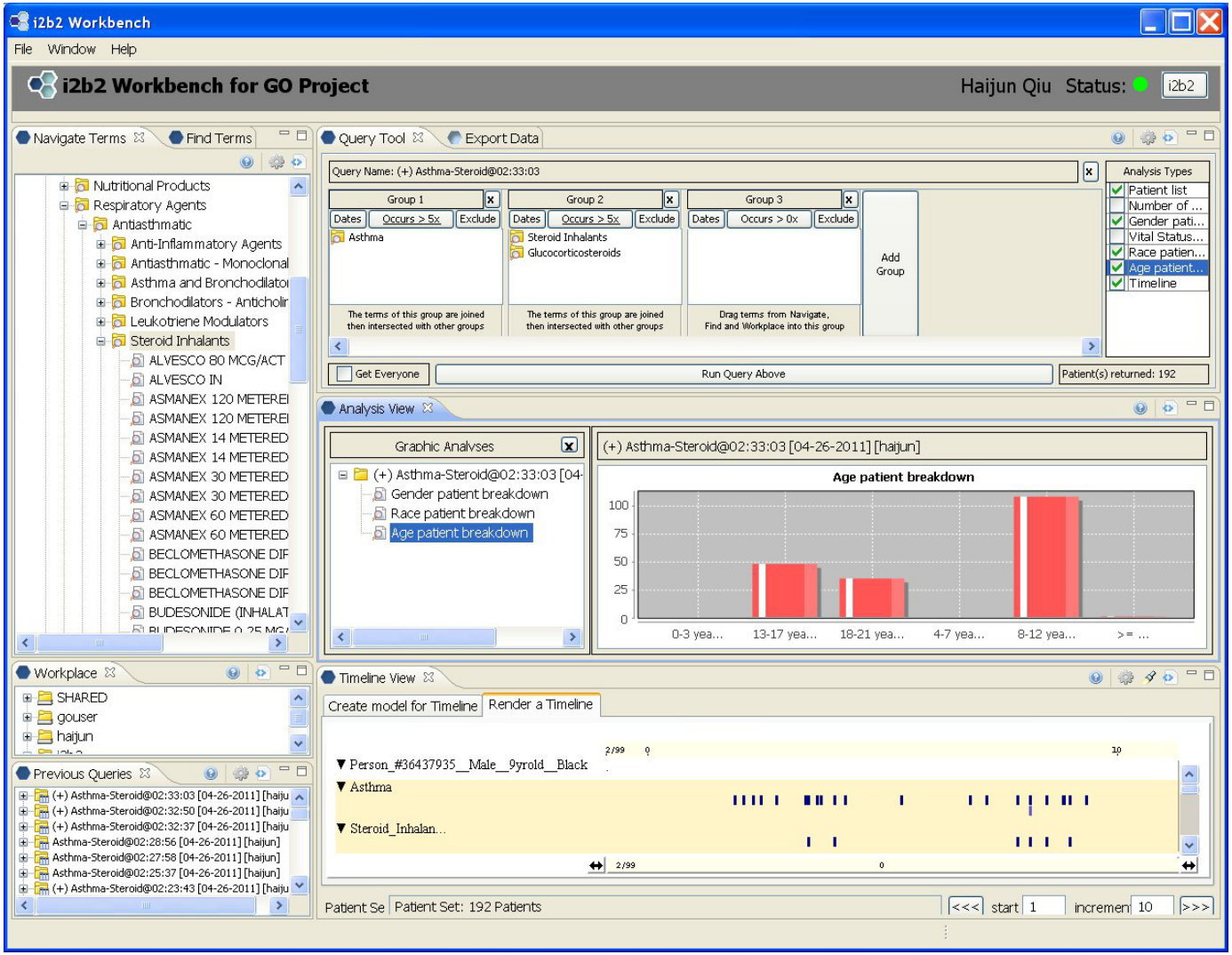
n=1614

Asthma – on preventive corticosteroids



n=813

Asthma - severe



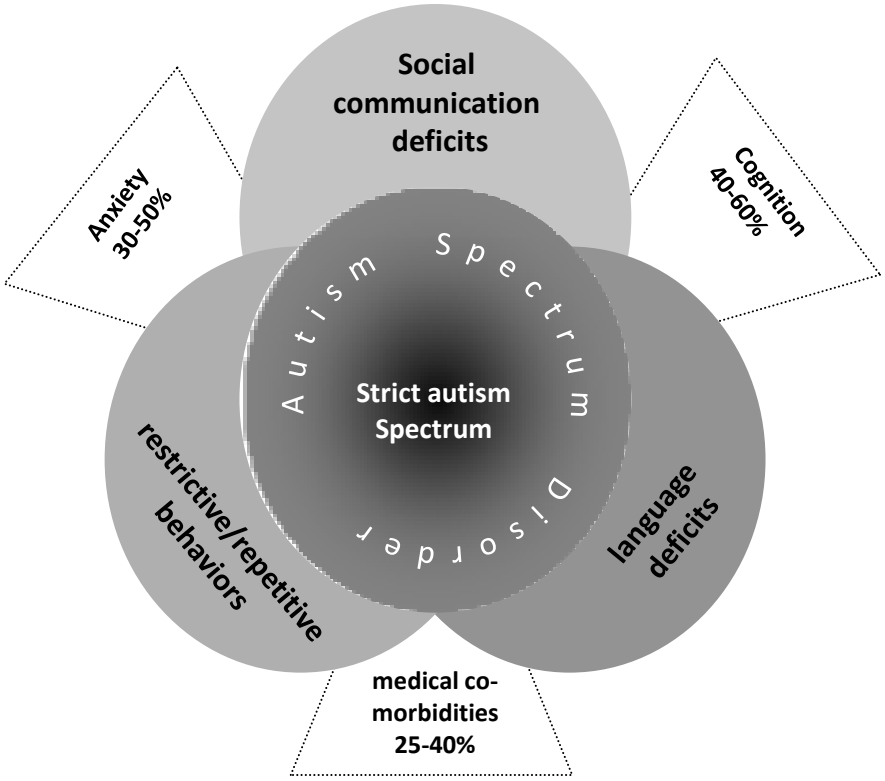
n=192

Deliverables from EHRs

- Numerous discoveries been made in the areas of pediatric disease, using EHR: Autism, ADHD, Anorexia, Asthma, IBD, T1D, Obesity, Childhood Cancer etc
- Characterization of a large (>10,000) healthy control cohort – been critical for our cytogenomics program where we deliver CNV results to clinicians and patients
- Characterization of various sub-cohorts we are targeting for clinical development – reposition of old drugs
  - ADHD program is the first program to be launched by CAG with autism program to follow
- Consented to participate in future studies

Autism Spectrum Disorders

A heterogeneous 'spectrum' disorder involving deficits in 3 domains of function



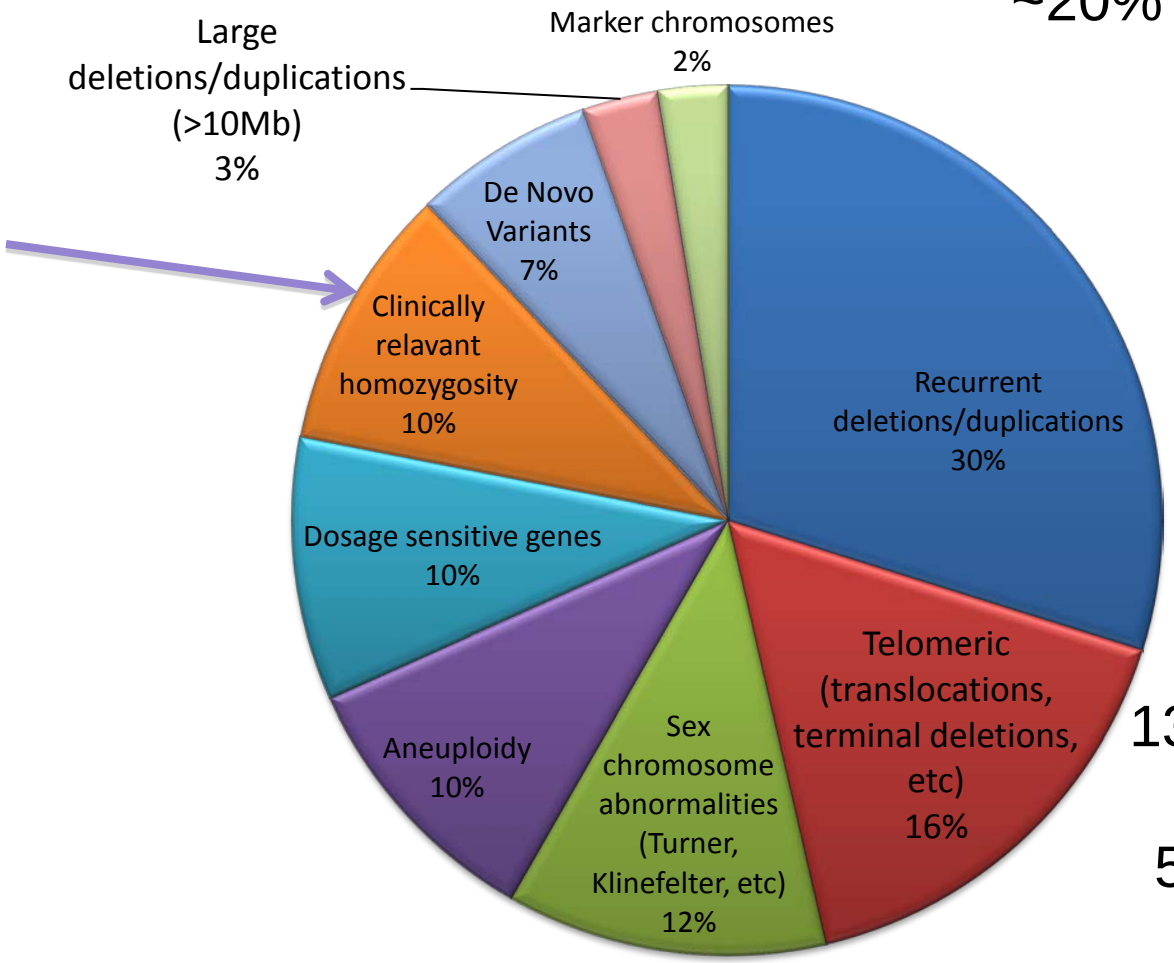
- 0.7-0.8 % prevalence
- ~15-20% of sibs have an ASD
- ~15 % of all cases have genomic finding (rare single-gene disorders, chr. rearrangements)
- Several CNVs identified as risk factors
- Few common GWAs hits

Bringing Genomic Discoveries to the Clinic

- We recently launched a study focused on recruiting 2000 children with autism, genotyping them on a high-density SNP array and analyzing and delivering the results back to the patient/family – targeted through EHR information
- These samples are processed under our CLIA/CAP-based workflow
- We report only on CNVs that have been established as playing a role in autism, such as 22q, 16p, 15q, SHANK3..
- A genetic counselor provides feedback to participants
- Future contact is established with updates if new content is identified and invitation to participate in other studies

Analysis of CNVs for Clinical Referrals

Over 6500 referrals processed
~20% abnormal



13% cases mosaic
and
5% at low level

Dilemmas for the Clinical Interpretation

Variable Expressivity

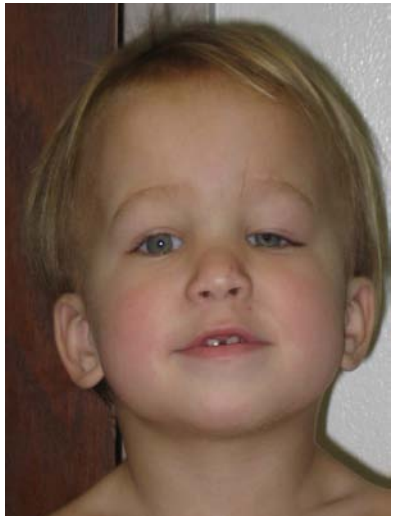
- Does inheritance from unaffected parent make it benign?
- How to counsel normal carriers of variably expressed syndromes

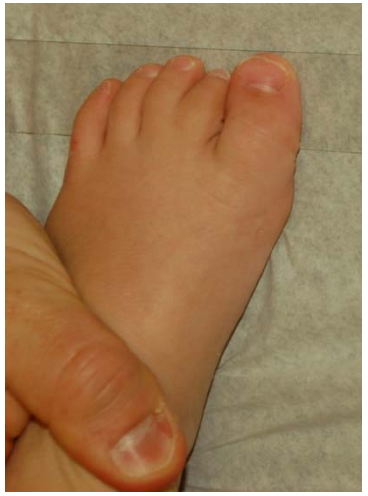
Non-recurrent CNVs

Incidental finding

Case 1

- Developmental delay
- Left eye ptosis
- Third nerve palsy
- Syndactyly
- Hip anomaly

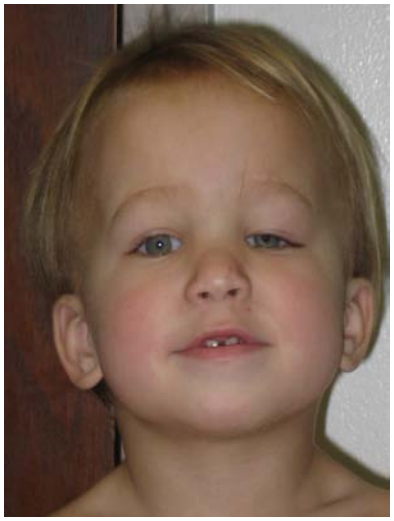


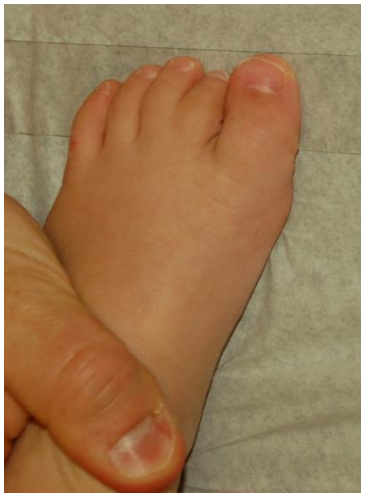


1q21 Deletion (1.6Mb, >15 genes)

Case 1

- Developmental delay
- Left eye ptosis
- Third nerve palsy
- Syndactyly
- Hip anomaly





INHERITED from father:

-Normal intelligence (PhD, Biochemist)
-Marathon runner
-Mildly dysmorphic

Pre-symptomatic findings

Case 2

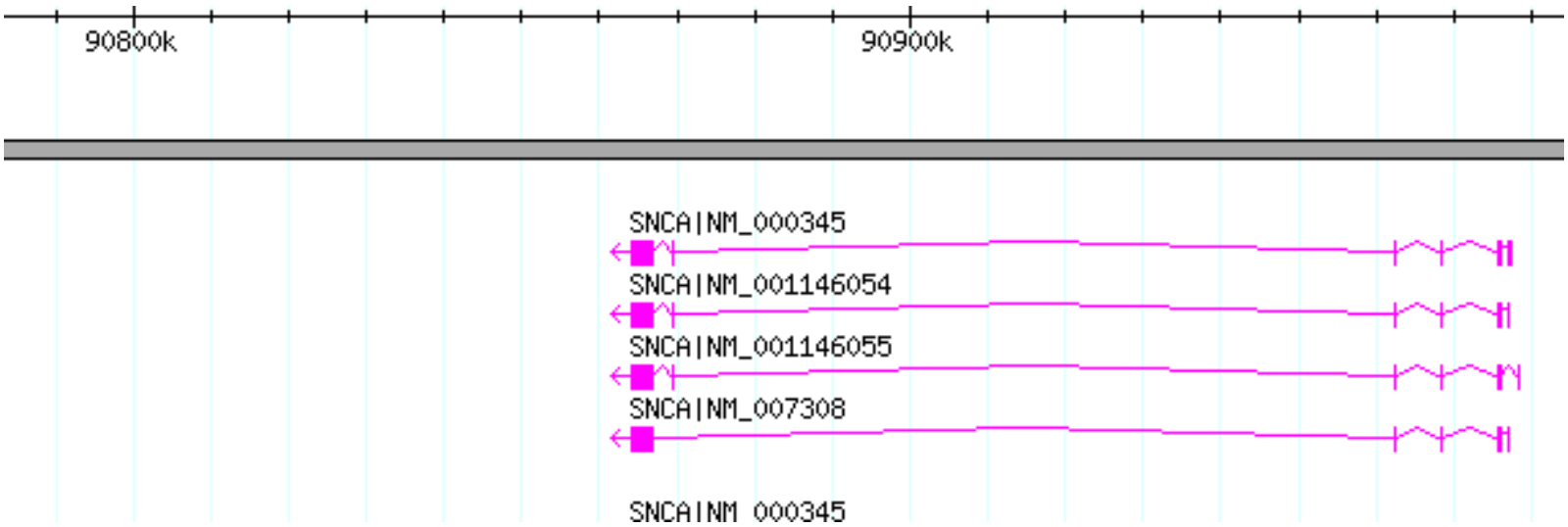
Developmental Delay

Small genomic region missing chromosome 17p

Includes a cancer predisposition gene (p53)

Eye Anomaly

400kb 6q duplication



Alpha-synuclein (SNCA) duplication
associated with autosomal dominant Parkinson disease

Incidental Findings

- Unexpected and unanticipated findings
  - Pre-symptomatic
    - Useful: cancer predisposition
    - Useful: carrier state for recessive mutations
    - Useful?: pre-senile dementia, Parkinson's
  - Parentage discrepancy
  - Reveals degree of inbreeding (incest?)

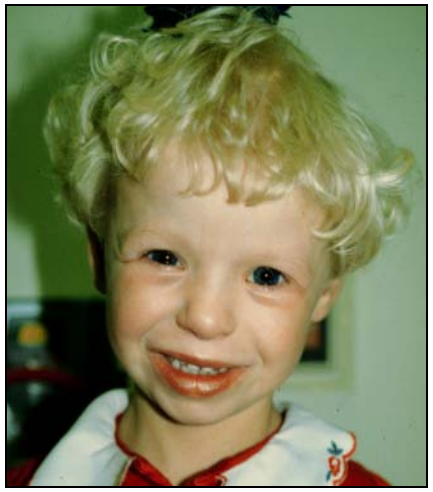
Examples of Recurrent Genomic Disorders

Smith Magenis

17p11.2 del

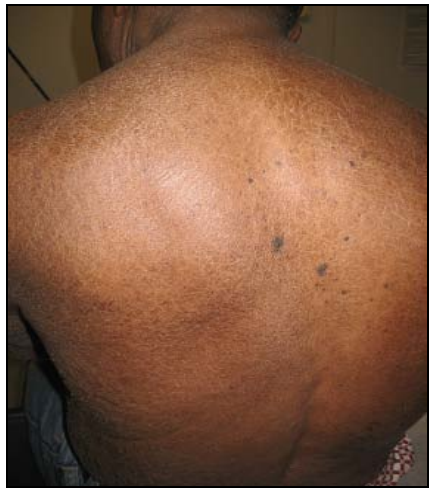
DiGeorge/VCSF

22q11.2 del



William syndrome

7q11.23 del



X linked ichthyosis

Xp22.31 del

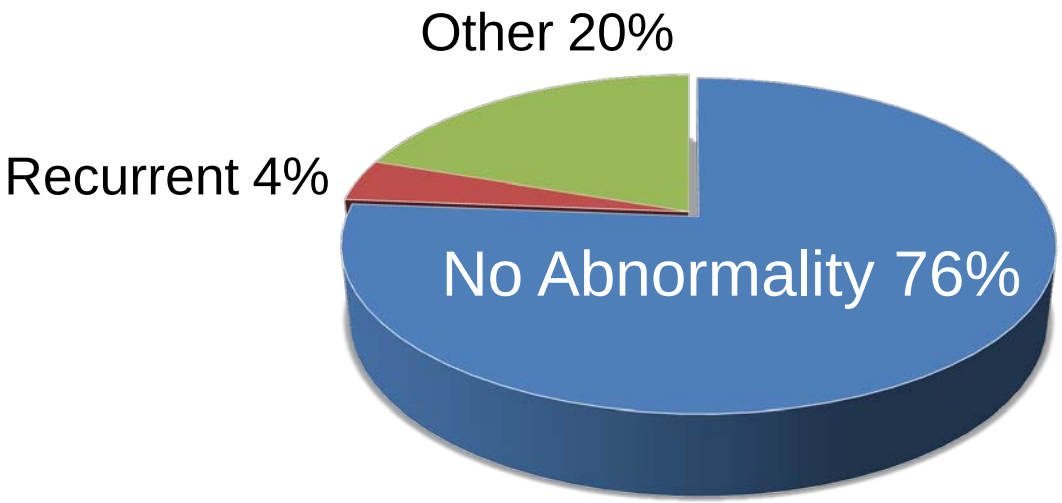
21 New Recurrent Disorders….. and Counting

Locus	Type	Size
1q21.1	Deletion / Duplication	1.35Mb
2q11.2	Deletion / Duplication	1.40Mb
2q13	Deletion / Duplication	1.70Mb
3q29	Deletion / Duplication	1.50Mb
5q35.2q35.3	Duplication	2.10Mb
7q11.23	Duplication	
7q36.1	Deletion	2.35Mb
8p23.1	Duplication	5.48Mb
10q22q23	Deletion	7.90Mb
15q11.2	Deletion	0.50Mb
15q13.3	Deletion / Duplication	1.50Mb
15q24	Deletion	1.80Mb
16p11.2	Deletion / Duplication	0.58Mb
16p11.2p12.2	Deletion	6.00Mb
16p12.1	Deletion	0.52Mb
16p13.11	Deletion / Duplication	1.00Mb
17p11.2	Duplication	3.60Mb
17q12	Deletion / Duplication	1.50Mb
17q23.2	Deletion	2.14Mb
17q21.31	Deletion	0.70Mb
Xp22.31	Duplication	1.5Mb

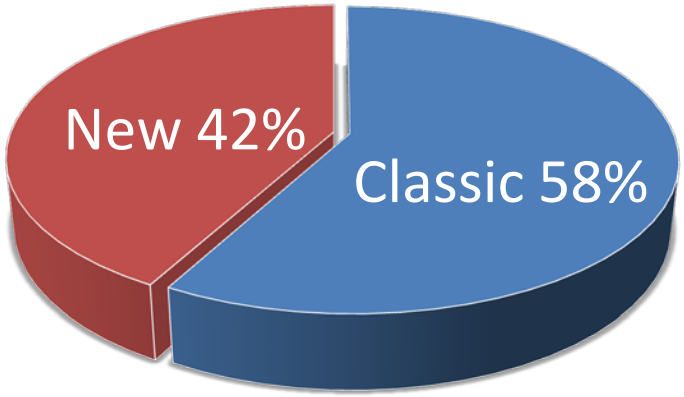
- Intellectual disability
- Autism
- Multiple congenital anomalies
- Schizophrenia

Identification of new recurrent deletions/duplications

Recurrent Genomic Disorders in the CHOP Cohort



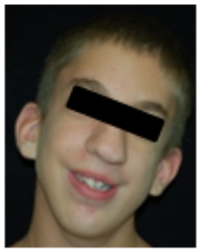
Results

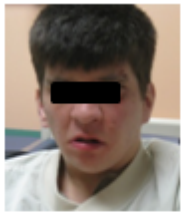


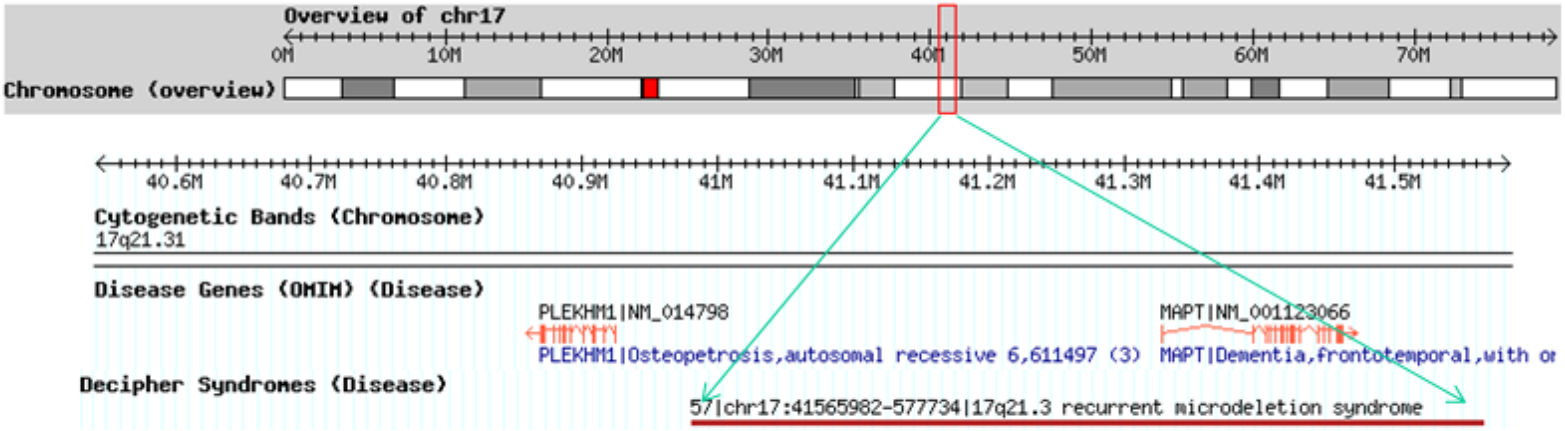
Recurrent Rearrangements

Classifying Newly Characterized Syndromes

17q21.31 deletion







- 5 Patients: all affected
- *De-novo*
- Dysmorphia + multiple anomalies including agenesis of corpus callosum, developmental delay, hypotonia and +/- cleft lip/palate

Summary-
New Recurrent Genomic Disorders

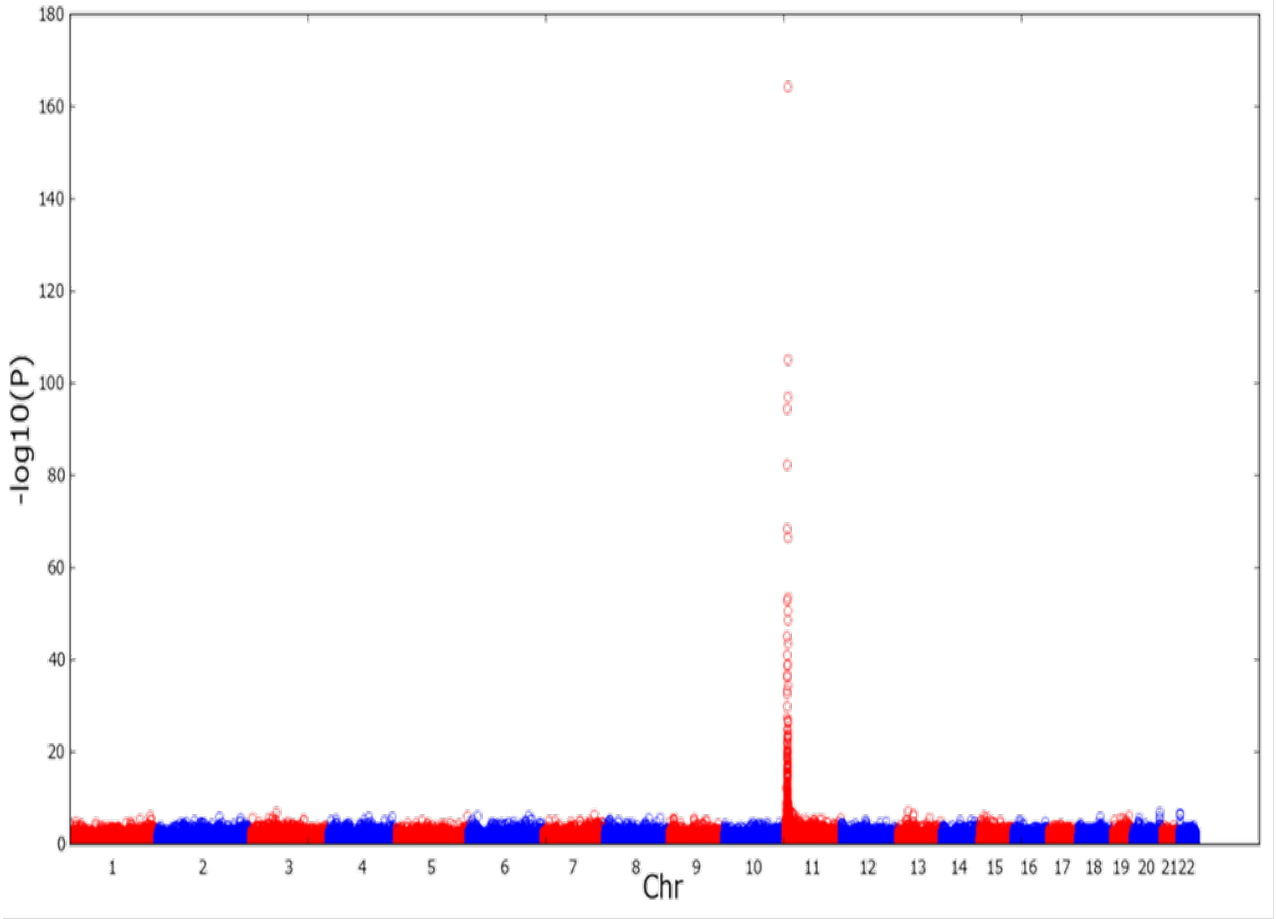
Five groups:

- Category 1: De-novo, consistent phenotype
- Category 2: Inherited, consistent phenotype
- Category 3: De-novo, variable features
- Category 4: Inconsistent inheritance and features
- Category 5: Likely benign

We are currently advancing this program to begin report on sequencing results

Rare Disease Example:

Sickle Cell Disease Dx from EMRs

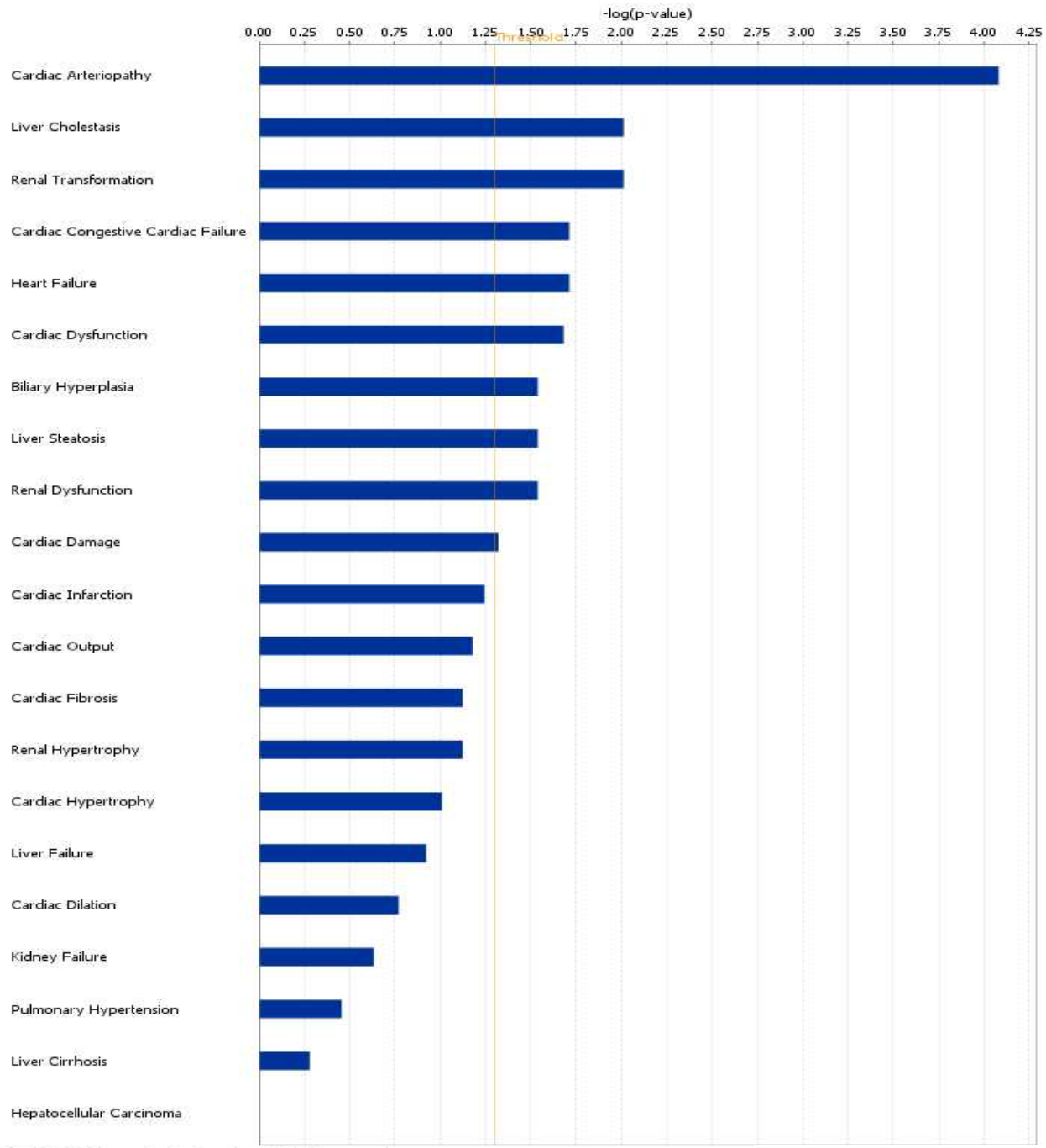


Linkage disequilibrium (r2) between SNPs at the 11p15.5 in SCA: Plotted are −log10(P-value) of allelic chi-squared tests

Sickle Cell Disease

		SCA (n=78)		Control (n= 763)		SCA	Control	
ICD-9	Phenotype	Yes	No	Yes	No	%	%	p
79.99	Viral infection	21	57	327	436	27	43	0.007606
389.9	Hearing loss	11	67	102	661	14	13	0.861611
429.3	Cardiomegaly	14	64	8	755	18	1	1.97E-10
462	Acute pharyngitis	32	46	190	573	41	25	0.002953
477.9	Allergic rhinitis	28	50	205	558	36	27	0.110147
493.9	Mild persistent asthma	37	41	241	522	47	32	0.007556
517.3	ACS	50	28	5	758	64	1	2.75E-54
564	Constipation	42	36	162	601	54	21	2.97E-09
692.9	Dermatitis NOS	22	56	369	394	28	48	0.000763
724.2	Lumbago	20	58	4	759	26	1	1.55E-18
724.5	Backache	36	42	17	746	46	2	1.16E-28
729.5	Pain in soft tissue	43	35	54	709	55	7	3.34E-24
733.4	Aseptic necrosis of bone	10	68	0	763	13	0	2.72E-11
780.57	Sleep apnea	20	58	28	735	26	4	4.47E-10
782.4	Jaundice	30	48	13	750	38	2	4.69E-24
784	Headach	39	39	72	691	50	9	4.24E-17
785	Tachycardia	34	44	32	731	44	4	3.59E-21
786.2	Cough	32	46	210	553	41	28	0.017476
787.03	Vomiting	32	46	146	617	41	19	2.98E-05
789	Abdorminal pain	37	41	91	672	47	12	7.04E-13
V12.54	Transient ischemic attack	9	69	1	762	12	0	3.04E-09
V12.69	Diseae of respiratory system	41	37	22	741	53	3	5.06E-32
V58.62	Long-term use of antibiotics	36	42	7	756	46	1	1.96E-34

Sickle Cell Disease



- Gene function analysis by Ingenuity (looking for modifying factors)
- Total 209 genes impacted by CNVs in SCD were used for function analysis

.

- Genes related to cardiac arteriopathy are highly enriched and more significantly linked to the SCA patients than other toxicity related genes (p-value = 8.29E-05).
- This result suggests that genetic variations increasing cardiac arteriopathy may play an important role in manifestation of SCA phenotypes

Sickle Cell Disease

Symbol	Entrez Gene Name	Location
ABCB11	ATP-binding cassette, sub-family B (MDR/TAP), member 11	Extracellular Space
ADK	adenosine kinase	Nucleus
ANKRD26P1	ankyrin repeat domain 26 pseudogene 1	unknown
BIN1	bridging integrator 1	Nucleus
CADPS	Ca++-dependent secretion activator	Plasma Membrane
CD36	CD36 molecule (thrombospondin receptor)	Plasma Membrane
CNTN4	contactin 4	Plasma Membrane
CPNE4	copine IV	Cytoplasm
FAM19A4	family with sequence similarity 19 (chemokine (C-C motif)-like), member A4	Extracellular Space
IMMP2L	IMP2 inner mitochondrial membrane peptidase-like (S. cerevisiae)	Cytoplasm
KIAA1370	KIAA1370	unknown
LINGO2	leucine rich repeat and Ig domain containing 2	unknown
MYO3B	myosin IIIB	unknown
NPR1	natriuretic peptide receptor A/guanylate cyclase A (atrionatriuretic peptide receptor A)	Plasma Membrane
NRXN1	neurexin 1	Plasma Membrane
PBX3	pre-B-cell leukemia homeobox 3	Nucleus
PHACTR3	phosphatase and actin regulator 3	Nucleus
PKD1L2	polycystic kidney disease 1-like 2	unknown
PLXDC2	plexin domain containing 2	Extracellular Space
PPP1R9A	protein phosphatase 1, regulatory (inhibitor) subunit 9A	Cytoplasm
PRKCE	protein kinase C, epsilon	Cytoplasm
SCARB1	scavenger receptor class B, member 1	Plasma Membrane
SDK1	sidekick homolog 1, cell adhesion molecule (chicken)	Plasma Membrane
SNTG1	syntrophin, gamma 1	Nucleus
SPOCK3	sparc/osteonectin, cwcv and kazal-like domains proteoglycan (testican) 3	Extracellular Space
WWOX	WW domain containing oxidoreductase	Cytoplasm
ZNF295	zinc finger protein 295	Nucleus
CADPS2	Ca++-dependent secretion activator 2	Plasma Membrane
SPAG11A	sperm associated antigen 11A	Extracellular Space
SPAG11B	sperm associated antigen 11B	Extracellular Space
AK2	adenylate kinase 2	Cytoplasm

31 genes with 'harmful' function in cardiac arteriopathy

Thank you