Genomic Medicine in Pediatric Patients – Obstacles and Future Directions

Center for Applied Genomics, Children's Hospital of Philadelphia

Hakon Hakonarson, MD, PhD

Bethesda, January 22, 2014

Three Areas

- Review of Current Pediatric Projects
  - Phenotyping
  - Consent
  - Sequencing
- New approaches to analyzing existing data
- Prospective directions – a custom-based informative chip

1. Current Pediatric Projects

1a. Phenotypes - Pediatric-Led Algorithms

Phenotype	Primary	Secondary	Status
Asthma	CAG	Marshfield CCHMC	Completed by all Centers GWAS ongoing
Atopic Dermatitis	CAG	Marshfield	Ready for eMERGE-wide dissemination
Obesity	CCHMC/ Boston	CAG	Validated
Autism	CCHMC/ Boston	CAG	Undergoing validation
ADHD	CAG	?	In development
GERD	CAG	?	In development
Lipids	CAG	?	In development
Others?	CCHMC?	?	?

Asthma

Center	Cases	Controls	C:C Ratio
CCHMC	20	1,582	79.1 : 1
CHOP	4,598	9,470	2.1 : 1
Geisinger	204	1,098	5.4 : 1
Group Health	131	949	7.2 : 1
Marshfield	255	869	3.4:1
Mayo	205	3,117	15.2 : 1
Mount Sinai	743	1,062	1.4 : 1
Northwestern	234	1,943	8.3 : 1
Vanderbilt	326	1,336	4.1 : 1
TOTAL	**6,716**	**21,426**	**3.2 : 1**

Adult-Led Algorithms

	Cases	Controls
C.Diff		
CCHMC	15	0
CHOP	165	178
All Sites	1,919	10,437
VTE		
CCHMC	--	--
CHOP	140	469
All Sites	4,460	23,153
Diverticulosis		
CCHMC	--	--
CHOP	--	--
All Sites	6,060	4,049
Zoster		
CCHMC	--	--
CHOP	--	--
All Sites	2,446	24,396
Extreme Obesity		
CCHMC	--	--
CHOP	2	42
All Sites	1,293	7,239

	Cases	Controls
AAA		
CCHMC	--	--
CHOP	--	--
All Sites	1,103	16,643
Occular Hypertension		
CCHMC	--	--
CHOP	--	--
All Sites	771	7,477
Glaucoma		
CCHMC	--	--
CHOP	--	--
All Sites	1,124	4,568
Ace-I Cough		
CCHMC	--	--
CHOP	--	--
All Sites	1,792	8,476
TOTAL		
CCHMC	15	0
CHOP	307	689
All Sites	20,968	106,438

Adult-Led Algorithms

	Cases	Controls
C.Diff		
CCHMC	15	0
CHOP	165	178
All Sites	1,919	10,437
VTE		
CCHMC	--	--
CHOP	140	469
All Sites	4,460	23,153
Diverticulosis		
CCHMC	--	--
CHOP	--	--
All Sites	6,060	4,049
Zoster		
CCHMC	--	--
CHOP	--	--
All Sites	2,446	24,396
Extreme Obesity		
CCHMC	--	--
CHOP	2	42
All Sites	1,293	7,239

	Cases	Controls
AAA		
CCHMC	--	--
CHOP	--	--
All Sites	1,103	16,643
Occular Hypertension		
CCHMC	--	--
CHOP	--	--
All Sites	771	7,477
Glaucoma		
CCHMC	--	--
CHOP	--	--
All Sites	1,124	4,568
Ace-I Cough		
CCHMC	--	--
CHOP	--	--
All Sites	1,792	8,476
TOTAL		
CCHMC	15	0
CHOP	307	689
All Sites	20,968	106,438

Major Obstacle

- Incongruity between pediatric and adult data sets:
  - Adult algorithms that exclude or have low frequency in pediatric patients
  - Pediatric algorithms that exclude or have low frequency in adult populations

Options

- Adult/pediatric sites pursue entirely separate paths for phenotyping
- Revise list of candidate phenotypes to increase overlap
- Proceed as-is (i.e. on a case-by-case basis)
- Utilize divergent primary/validation strategy:
  - Adult sites for primary analysis, pediatric for validation

    and

  - Pediatric sites for primary analysis, adult sites for validation

1b. Consent (Kyle Brothers)
Practical Guidance on Informed Consent:

- Consent from one parent is adequate
- Children should be asked to provide assent starting with ages 7 through 10.
- Older adolescents, perhaps those older than 14 years of age, in a “co-consent” process.
- Sharing de-identified data is appropriate for pediatric biobanks
- Identified pediatric data should generally not be retained beyond age of majority without consent
- It is acceptable for a pediatric biorepository to return results but should take individual preferences into account
- Result should only be returned when both the adolescent and her parents agree

1c. Sequencing Recap: Pediatric Centers

	BCH	CCHMC	CHOP	GHC	Geisinger	Marsh	Mayo	MSSM	NU	VU
Site Information	X	X	X		X	X	X	X	X	X
Recruitment Process Descriptive Meta-Data	X	X	X		In progress		X	In progress	X	X
Recruitment Statistics		X	X				X		X	X
PGRNSeq Sequencing Descriptive Meta-Data	n/a	n/a	X			n/a	X		n/a	n/a
PGRNSeq Quantitative Measures	n/a	n/a	X			n/a	X		n/a	n/a
Validation Descriptive Meta-Data		In progress	X				X		X	X
Validation Quantitative Measures			X				X		In progress	In progress
EHRIntegration and CDS Descriptive Meta-Data			X				X		X	In progress

2. New approaches to analyzing existing data

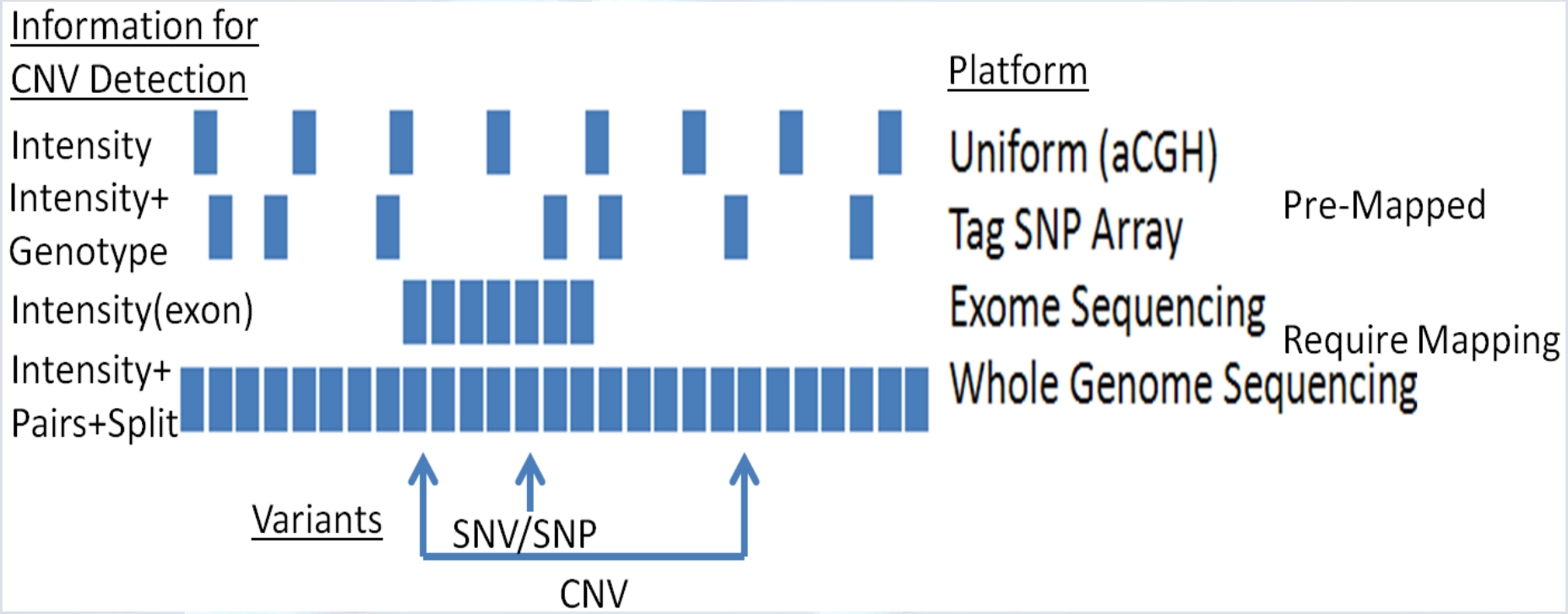
Proposed New Approaches to Existing Data

- Copy Number Variants – Tools and Opportunities
- Imputing Drug-Gene Interactions from GWAS data
- CNV Analysis and Sequencing Data
- High-sensitivity GWAS and Functional/biological annotation using publicly accessible resources
  - Gene-Based Association Testing (GBAT)
  - Tissue Specific Gene Set Enrichment Analysis
  - Immuno-Cell Types Gene Set Enrichment Analysis
  - Pathways, Protein Interaction, and Text-Based Enrichment Analysis: Dapple, *Webgestalt, String, David, IPA, Other*

2a. Copy Number Variants and Existing Data

- CNVs are the primary mode by which an individual acquires a mutation, and occur at a rate of approximately 1.7×10^{-6} per locus as opposed to 1.8×10^{-8} for sequence variation (Lupski et al., 2007)
- eMERGE includes >56,000 genotypes linked with electronic medical records (EMRs).
- Largely untapped resource

Range of approaches are available for detecting CNVs



Opportunity

- Considerable CNV expertise in eMERGE:
  - CHOP Developed PennCNV
  - CHOP Developed PareseCNV
  - PennCNV-Seq currently in development
- Revisit existing phenotypes
- Catalog Pathogenicity

Pathogenicity - Obstacle:

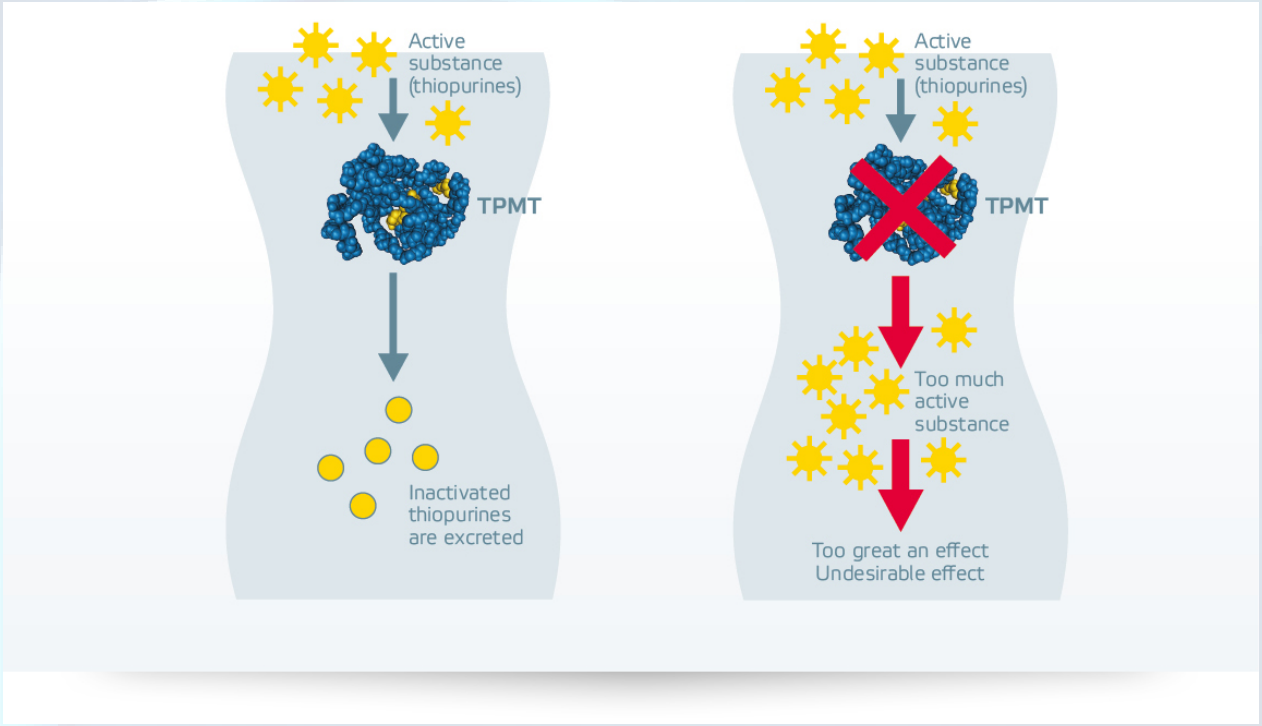
- The Database of Genomic Variation (DGV) currently lists over 100,000 published, unique, CNVs across the genome.
- However, underlining studies are inconsistent in terms of platforms, QC, methodology, etc.
- Duclos et al. (2011) “Urgent need to validate the frequencies and boundaries of the CNVs recorded in the DGV.

Pathogenecity - eMERGE to the Rescue?

- CNV expertise
- Genotypes centralized
- Critically, records are EMR-linked
  - Provide proper control data (i.e. not just absence of one particular phenotype)
  - Increase confidence that what is catalogued as benign is indeed benign

2b. Imputing Drug-Gene Interactions from GWAS Data

- TPMT: enzyme involved in the metabolism of purine analogs
- Used as chemotherapeutic and immunosuppressant agents
- Due to the potential cytotoxicity and narrow therapeutic index, the FDA recommends TPMT testing prior to treatment



Imputing Drug-Gene Interactions, Methods

- N = 87,979 (CHOP) genotyped with
  - Infinium II HumanHap550 (550; N=45,893)
  - Human610-Quad version 1 (Quad; N=42,086)
- Imputation with IMPUTE2
- Four most common defective alleles imputed
  - *2 (rs1800462)
  - *3A (rs1800460 and rs1142345)
  - *3B (rs1800460)
  - *3C (rs1142345)

Imputing Drug-Gene Interactions, Results

	Caucasian (N=63,997)		AA (N=16,519)		Hispanic (N=5,764)		Asian (N=1,698)		Total (N=87,978)	
Allele	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
***1**	122,787	95.93	31,225	94.51	11,020	95.59	3,302	97.23	168,333	95.67
***3A**	4,305	3.36	303	0.92	334	2.90	19	0.56	4,961	2.82
***3B**	86	0.07	1	0.00	12	0.10	0	0.00	99	0.06
***3C**	817	0.64	1509	4.57	162	1.41	75	2.21	2,563	1.46
Genotype	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
***1/*1**	58,981	92.16	14,761	89.36	5,275	91.52	1,606	94.58	80,623	91.64
***1/*3A**	4,119	6.44	286	1.73	322	5.59	19	1.12	4,746	5.39
***1/*3B**	10	0.02	1	0.01	1	0.02	0	0.00	12	0.01
***1/*3C**	697	1.09	1,416	8.57	147	2.55	71	4.18	2,331	2.65
***3A/*3A**	81	0.13	1	0.01	5	0.09	0	0.00	87	0.10
***3A/*3B**	1	0.00	0	0.00	0	0.00	0	0.00	1	0.00
***3A/*3C**	23	0.04	15	0.09	2	0.03	0	0.00	40	0.05
***3B/*3C**	75	0.12	0	0.00	11	0.19	0	0.00	86	0.10
***3C/*3C**	11	0.02	39	0.24	1	0.02	2	0.12	53	0.06

Imputing Drug-Gene Interactions, Summary

	Caucasian (N=63,997)		AA (N=16,519)		Hispanic (N=5,764)		Asian (N=1,698)		Total (N=87,978)	
Phenotype	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
Normal	58,980	92.16	14,761	89.36	5,275	91.52	1,606	94.58	80,623	91.64
Intermediate	4,826	7.54	1,703	10.31	470	8.15	90	5.30	7,089	8.06
Low	191	0.30	55	0.33	19	0.33	2	0.12	267	0.30

Imputing Drug-Gene Interactions, Summary – Genotyping, Sanger Seq.

- Genotyping (n=585)
  - Samples validated on Illumina Infinium Immunochip and Omni-Quad (v1), which captured both rs1800460 and rs1142345
  - Concordance between imputed haplotypes and those determined by genotyping with alternative platforms was 99.8%.
- Sanger (n=60)
  - 85% concordance with imputed haplotypes

Validation, Summary

Genotyping and Sanger Sequencing (N=645)	Wild Type	Heterozygotes	Homozygotes
Wild Type	554	1	1
Heterozygote	0	57	5
Homozygote	0	0	25
Total	554	58	33
Concordance	100%	98.28%	75.76%

Imputing Drug-Gene Interactions, Conclusions

- Accuracy of imputation was sufficiently high to allow discrimination of patients carrying one or two defective alleles from those with a wild type genotype
- 1 in 10 individuals tested from our biobank were found to carry at least one high-risk TPMT allele
- Identification of such carriers is especially important in the pediatric population, as thiopurines are commonly prescribed drugs in children
- Large number of other drug-gene pairs can be used to similarly guide sample selection.

3. Prospective Directions – A Custom-Based Informative Chip

Custom-Chips

Facilitate Cost-Effective …

- Gene discovery of a comprehensive repertoire of genes through genomic
- Identification of mutations
- Genotype/Phenotype correlation
- Long-term clinical follow up of patients to determine how genetics influences the clinical outcome.

CAG-Led Custom-Chips

- Cardiochip (2008)
  - $53.14/samples
  - ~53k SNPs
  - n=220,000
  - >150 publications generated to date
- Transplant v1 (2013)
  - <$65/sample
  - 782k SNPs
- PCGC (2010)
  - $80/sample; 785k SNPs

Solid Organ Transplant Chip in Use

- Affy chip, solid organ transplant (excellent data quality):
  - 30k well validated LOF variants
  - ~100k exome generated nsSNPs – putatively pathogenic by various scores.
  - 450k tagging SNPs in the 1-50% MAF range (imputable to >15 million SNPs from the 1000G and ESP)
  - 20k CNV probes capturing all relevant CNVs ever reported with any potential pathogenicity
  - All GWAS loci of genome wide significance (8k)
  - eSNPs (all relevant eSNPs ever reported in the literature) (20k)
  - >780k total variants – cost/sample (>$65)

Future eMERGE Chip

- Affy chip with 780k variants at $65/sample total cost
- Content:
  - All validated LOF variants known (60k)
  - All exome generated nsSNPs – putatively pathogenic by various scores (>100k)
  - 450k tagging SNPs in the rare variant range not currently addressed by existing GWAS (<5% MAF range (imputable to multi-million SNPs from the 1000G and ESP)
  - 20k CNV probes capturing all relevant CNVs ever reported with any potential pathogenicity
  - All GWAS loci of genome wide significance (8-10k)
  - All relevant eSNPs ever reported in the literature) (20k)
  - All PGx variants known