Development of a Targeted Second-Tier Genetic Test for Conditions Examined during Newborn Screening

[1]Lonna Mollison, [1]Chelsea Gustafson, [1]Edgar A Rivera-Munoz, [1]Alicia Brandt, [1]Daniela DeCristo, [1]Falecia Metcalf, [2]Cynthia Powell, [1]Jonathan Berg

[1]Department of Genetics, [2]Department of Pediatrics and Genetics, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina

NCNexus — North Carolina Newborn Exome Sequencing for Universal Screening

UNC SCHOOL OF MEDICINE

Introduction

- Approximately four million children are born annually in the United States, most undergo state mandated newborn screening
- In 2006 the American College of Medical Genetics and Genomics (ACMG) developed a recommended uniform screening panel (RUSP) to minimize variability between states
- Most RUSP conditions are detected by tandem mass spectrometry
- Through the use of genetic sequencing it is possible to detect the underlying genetic cause of RUSP conditions
- Here we propose a step-wise approach to enhance traditional newborn screening and integrate genetic screening into population health using cost-effective, targeted sequencing to examine current RUSP conditions

Core Condition	Associated Gene(s)	ACMG Code
Maple Syrup Urine Disease	*BCKDHA, BCKDHB, DBT*	MSUD
Homocystinuria	*MTRR, CBS, MTHFR, MTR*	HCY
Propionic Acidemia	*PCCA, PCCB*	PROP
Medium-chain Acyl-CoA Dehydrogenase Deficiency	*ACADM*	MCAD
Very Long-chain Acyl-CoA Dehydrogenase Deficiency	*ACADVL*	VLCAD
β-Ketothiolase Deficiency	*ACAT1*	βKT
Argininosuccinic Aciduria	*ASL*	ASA
Citrullinemia, Type I	*ASS1*	CIT
Biotinidase Deficiency	*BTD*	BIOT
Tyroseinemia, Type I	*FAH*	TYR1
Glycogen Storage Disease Type II	*GAA*	GSDII
Classic Galactosemia	*GALT*	GALT
Glutaric Acidemia Type I	*GCDH*	GA1
Long-chanin L-3 Hydroxyl-CoA Dehydrogenase Deficiency	*HADHA*	LCHAD
Trifunctional Protein Deficiency	*HADHB*	TFP
Holocarboxylase Synthase Deficiency	*HLCS*	MCD
3-Hydroxy-3-Methyglutaric Aciduria	*HMGCS1, HMGCS2*	HMG
Isovaleric Acidemia	*IVD*	IVA
3-Methylcrotonyl-CoA Carboxylase Deficiency	*MCCC1, MCCC2*	3-MCC
Methylmalonic Acidemia (Cobalamin disorders)	*MMAA, MMAB*	Cbl A, Cbl B
Methylmalonic Acidemia (Methylmalonyl-CoA mutase)	*MUT*	MUT
Classic Phenylketonuria	*PAH*	PKU
Carnitine Uptake Defect/Carnitine Transport Defect	*SLC22A5*	CUD
Primary Congenital Hypothyroidism	*SLC5A5, THRA, THRB, THSR, DUOXA2, DUOX2, NKX2-5, PAX8*	CH
Secondary Condition	**Associated Gene(s)**	**ACMG Code**
Isobutyrylglycinuria	*ACAD8*	IBG
Short-chain acyl-CoA dehydrogenase deficiency	*ACADS*	SCAD
2-Methylbutyrylglycinuria	*ACADSB*	2MBG
Hypermethioninemia	*ADK, AHCY, MAT1A*	MET
Argininemia	*ARG1*	ARG
3-Methylglutaconic aciduria	*AUH*	3MGA, Type 1
Carnitine palmitoyltransferase type I deficiency	*CPT1A*	CPT IA
Carnitine palmitoyltransferase type II deficiency	*CPT2*	CPT II
Glutaric acidemia type IIA	*ETFA*	GA2
Glutaric acidemia type IIB	*ETFB*	GA2
Glutaric acidemia type IIC	*ETFDH*	GA2
Galactoepimerase deficiency	*GALE*	GALE
Galactokinase deficiency	*GALK1*	GALK
Biopterin defect in cofactor biosynthesis	*GCH1*	BIOPT(BS)
Hypermethioninemia	*GNMT*	MET
Medium/short-chain L-3-hydroxyacl-CoA dehydrogenase deficiency	*HADH*	M/SCHAD
Tyrosinemia, type III	*HPD*	Tyr III
2-Methyl-3-hydroxybutyric aciduria	*HSDI7B10*	2M3HBA
Malonic acidemia	*MLYCD*	MAL
Methylmalonic acidemia with homocystinuria	*MMACHC, MMADHC*	Cbl C, Cbl D
2,4 Dienoyl-CoA reductase deficiency	*NADK2*	DE RED
3-Methylglutaconic aciduria	*OPA3*	3MGA, TYPE III
Benign hyperphenylalaninemia	*PAH*	H-PHE
Biopterin defect in cofactor regeneration	*PCBD1, QDPR*	BIOPT(REG)
Biopterin defect in cofactor biosynthesis	*PTS*	BIOPT(BS)
Citrullinemia, type II	*SLC25A13*	CIT II
Carnitine acylcarnitine translocase deficiency	*SLC25A20*	CACT
Tyrosinemia, type II	*TAT*	Tyr II
3-Methylglutaconic aciduria	*TAZ*	3MGA, Type II

Table 1. Primary and secondary ACMG RUSP conditions and associated genes.

Methods

- We used molecular inversion probes (HEAT-Seq, Roche-NimbleGen) to examine 72 genes associated with RUSP primary and secondary conditions as a possible second-tier genetic screen
- We performed HEAT-Seq library preparation on eight samples that had previously undergone whole-exome sequencing (WES) (SureSelectXT, Agilent) and compared the exon coverage, base-level coverage and variant detection between the two methods

Acknowledgements & Sources

We would like to acknowledge funding from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) and The National Human Genome Research Institute (NHGRI) for RFA-HD-13-010.

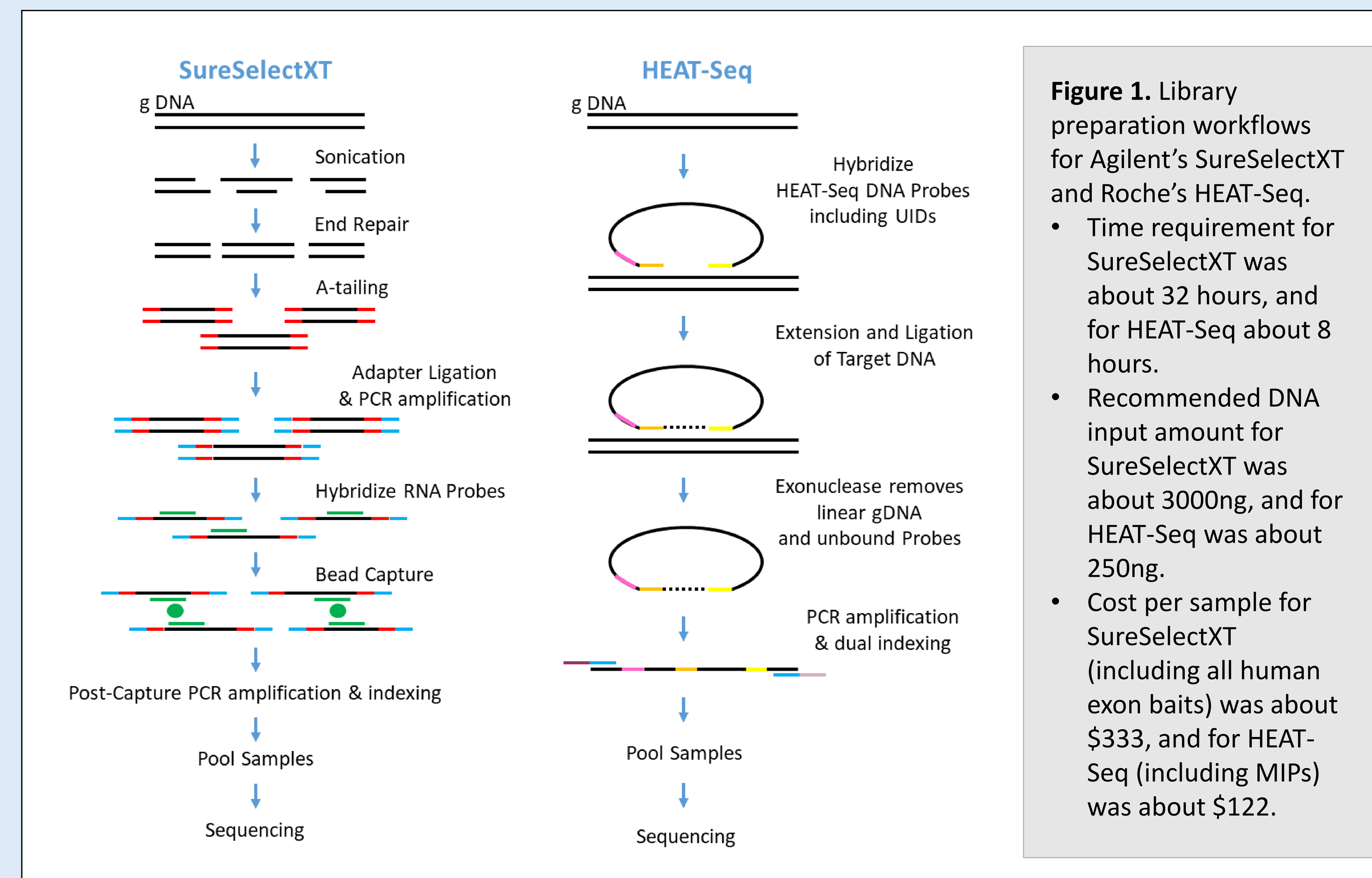

Figure 1. Library preparation workflows for Agilent's SureSelectXT and Roche's HEAT-Seq.
- Time requirement for SureSelectXT was about 32 hours, and for HEAT-Seq about 8 hours.
- Recommended DNA input amount for SureSelectXT was about 3000ng, and for HEAT-Seq was about 250ng.
- Cost per sample for SureSelectXT (including all human exon baits) was about $333, and for HEAT-Seq (including MIPs) was about $122.

HEAT-Seq Probe Performance

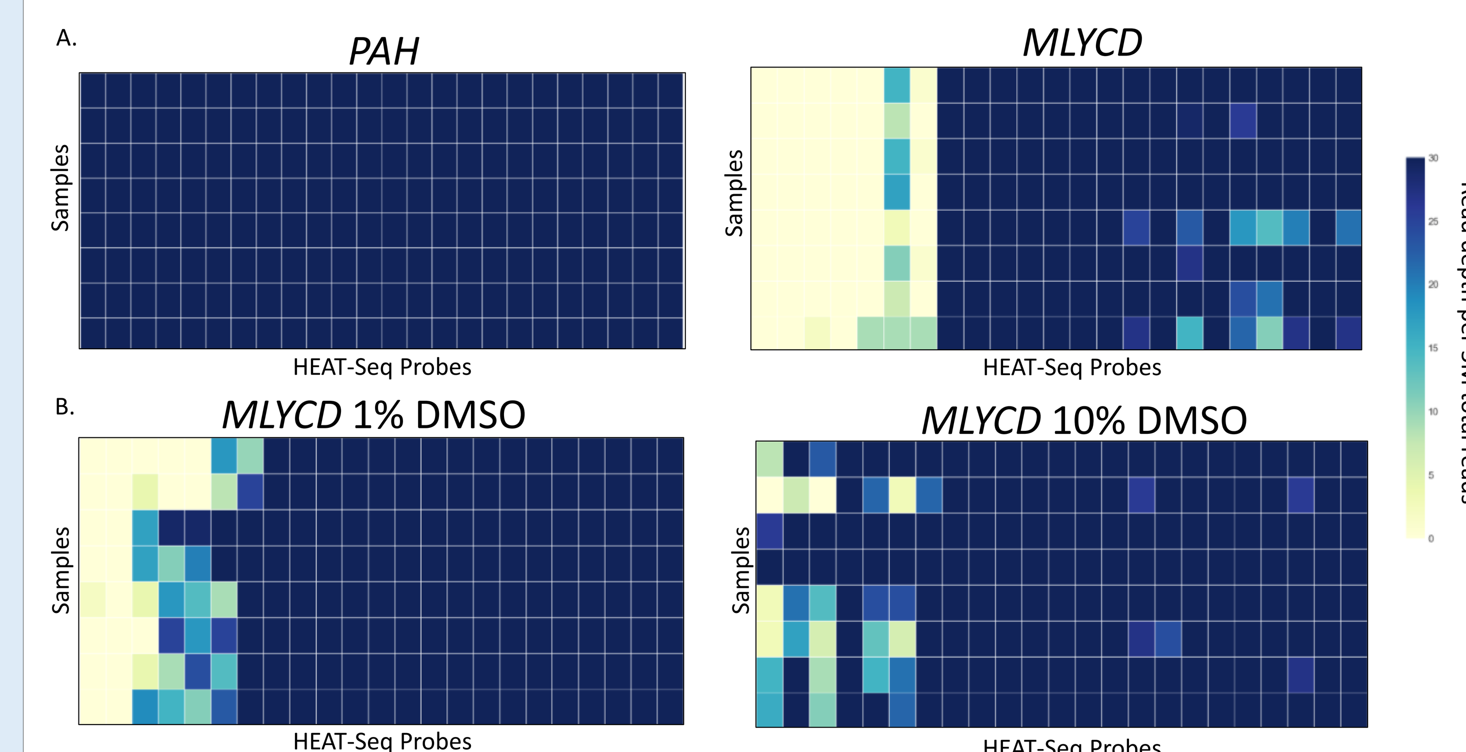

Figure 2. (A) Representative heat maps depicting the number of reads (normalized to 5M reads per sample) for every probe in each of eight samples. All probes yielded greater than or equal to 30 fold coverage in every sample for *PAH*, whereas *MLYCD* had 7 probes that performed poorly in every sample. **(B)** Many of the poor performing probes targeted G-C rich exons. The HEAT-Seq probes were rerun with 0.1%, 1.0%, and 10% DMSO treatment. The DMSO treatment improved coverage in genes with G-C rich exons, as seen with MLYCD.

	Condition	Exons Covered >30X: All	All but 1 or 2	None or > 2 not at 30X
Standard Protocol	Primary	13/40	19/40	10/40
	Secondary	13/32	12/32	4/32
DMSO Treatment	Primary	19/40	15/40	8/40
	Secondary	19/32	8/32	2/32

Table 2. Of the 72 genes targeted by the HEAT-Seq probes, 26 genes had 100% of the protein-coding targeted exons at 30 fold coverage on average across all samples when normalized to 5M sequenced reads. The addition of DMSO improved the number of genes with all protein-coding exons at 30 fold coverage to 38 genes.

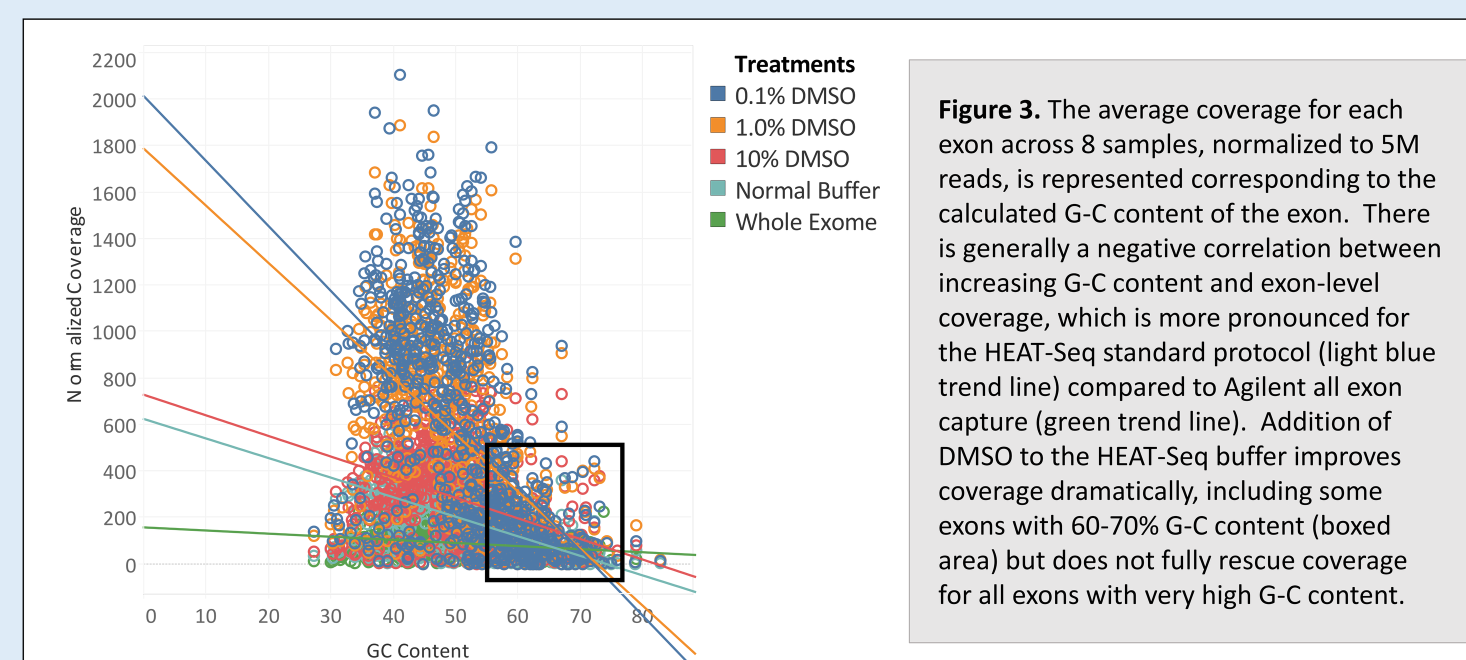

Figure 3. The average coverage for each exon across 8 samples, normalized to 5M reads, is represented corresponding to the calculated G-C content of the exon. There is generally a negative correlation between increasing G-C content and exon-level coverage, which is more pronounced for the HEAT-Seq standard protocol (light blue trend line) compared to Agilent all exon capture (green trend line). Addition of DMSO to the HEAT-Seq buffer improves coverage dramatically, including some exons with 60-70% G-C content (boxed area) but does not fully rescue coverage for all exons with very high G-C content.

HEAT-Seq and SureSelectXT Comparison

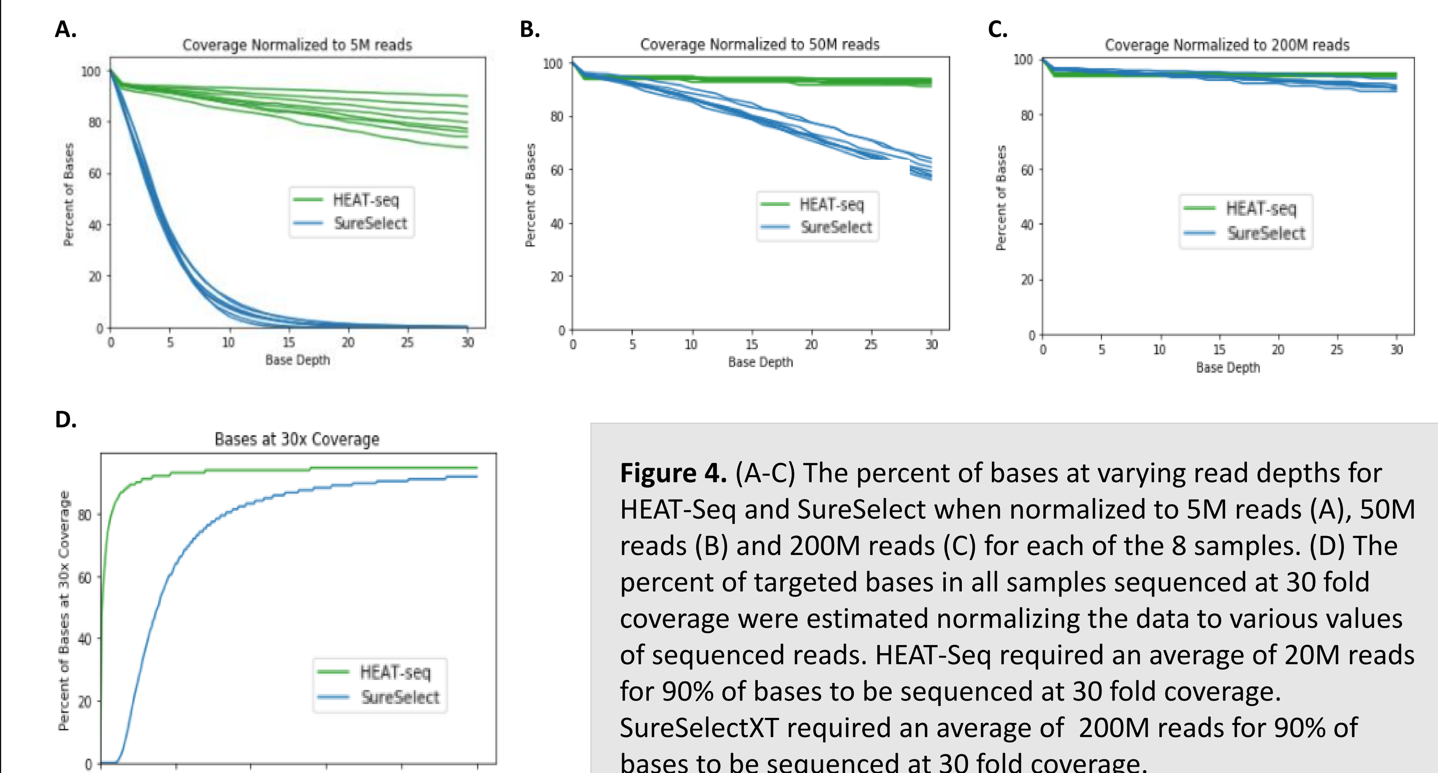

Figure 4. (A-C) The percent of bases at varying read depths for HEAT-Seq and SureSelect when normalized to 5M reads (A), 50M reads (B) and 200M reads (C) for each of the 8 samples. (D) The percent of targeted bases in all samples sequenced at 30 fold coverage were estimated normalizing the data to various values of sequenced reads. HEAT-Seq required an average of 20M reads for 90% of bases to be sequenced at 30 fold coverage. SureSelectXT required an average of 200M reads for 90% of bases to be sequenced at 30 fold coverage.

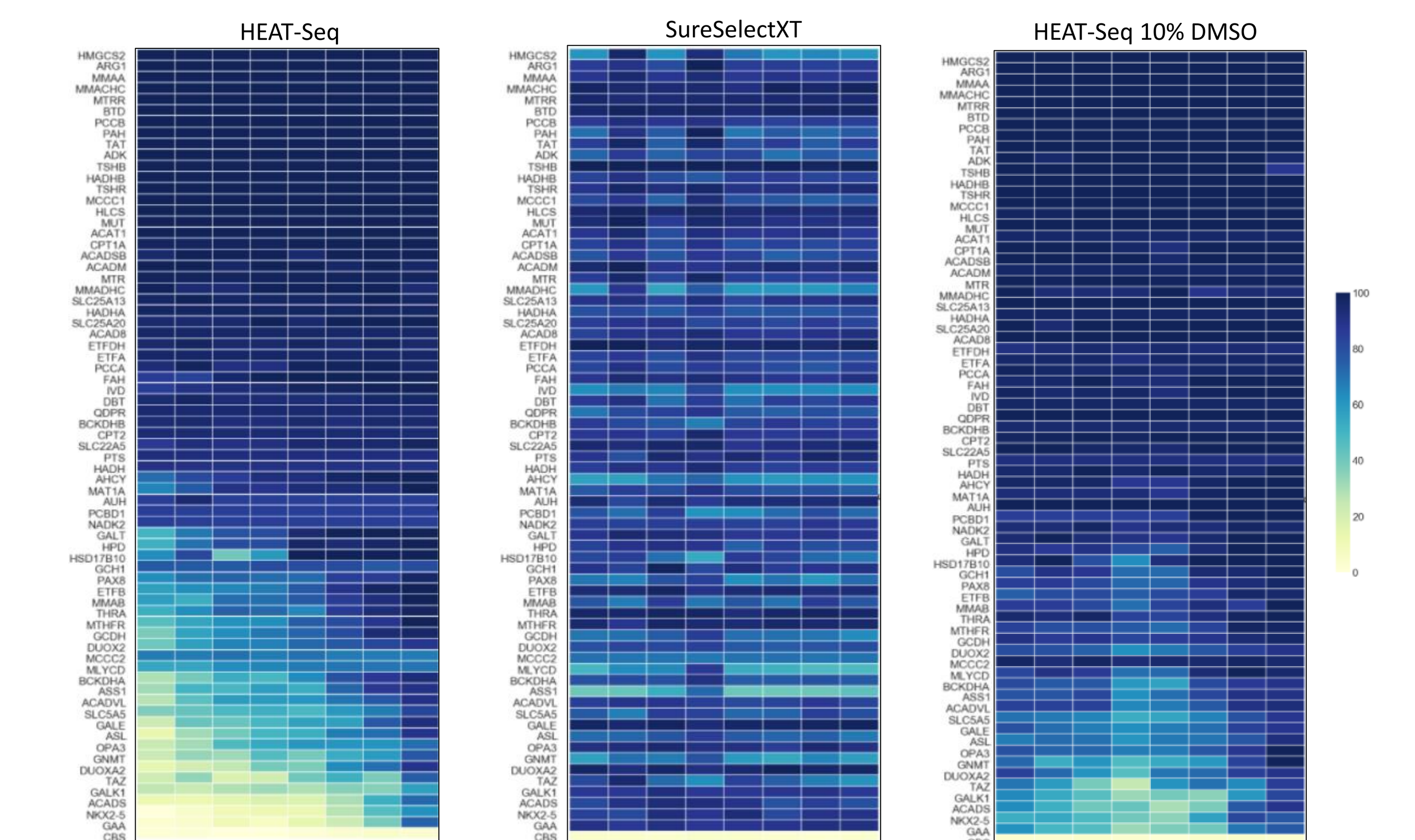

Figure 5. Heatmaps illustrating the percentage of nucleotides in each coding region covered at 30X or greater. HEAT-Seq samples (left) were normalized to 5M reads, which correlated to 80% of nucleotides covered at 30X or greater. The SureSelect XT data (center) was normalized to 100M reads to provide comparable coverage with 80% of nucleotides at 30X or greater. A separate assay was conducted using HEAT-Seq with 10% DMSO treatment (right), showing improvement of coverage for many genes, which presumably corresponds to better performance in GC-rich regions.

HEAT-Seq and SureSelectXT Variant Detection

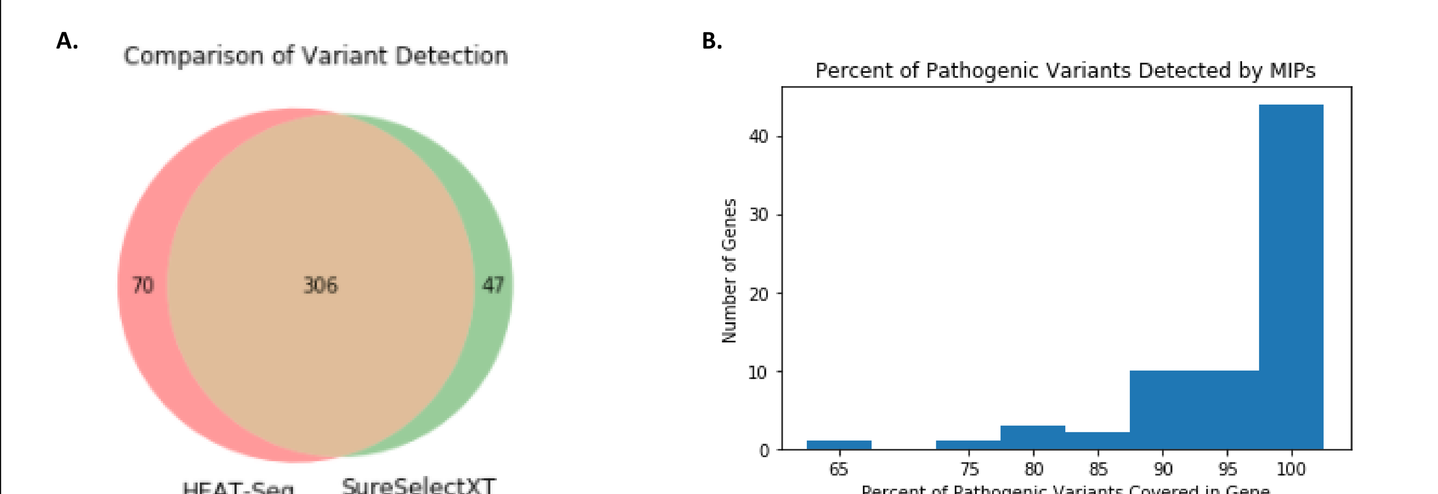

Figure 6. Comparison of variants detected by HEAT-Seq and SureSelectXT when confined to protein-coding exons targeted by HEAT-Seq probes. **(A)** A total of 423 variants were called across the participants using both HEAT-Seq and SureSelectXT technologies. 72% (306) of these variants were called using both technologies. 17% (70) of the total variants were only detected by Heat-Seq and 11% (47) of variants were only detected using SureSelectXT. **(B)** Theoretical depiction of the ability of Heat-Seq probes to detect pathogenic variants in RUSP genes. HEAT-Seq probe design should detect at least 75% or greater of the pathogenic variants in 69 RUSP genes.

Conclusions & Future Implications

We anticipate that this approach could be translated as an economical secondary genetic screen for current newborn screening, and serve as a proof of concept for adding other medically actionable conditions to the current recommended list for newborn screening.