Evaluation of Targeted Sequencing Technology to Screen 17 Genes for Actionable Conditions in Healthy Individuals

Alicia Brandt[1], Edgar A Rivera-Munoz[1], Lonna Mollison[1], Chelsea Gustafson[1], Daniela DeCristo[1], Falecia Metcalf[1], James Evans[1], Gail Henderson[2], Jonathan S. Berg[1]

[1]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC

[2]Department of Social Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC

Introduction

Screening programs, such as newborn screening, in healthcare serve a public health role in that they have the potential for early detection and prevention of diseases prior to clinical manifestation of symptoms

Advancements in next-generation sequencing (NGS) provide opportunities to implement genomic screening and "precision medicine" in the general population

GeneScreen explores the feasibility and ethics of screening an adult population for 11 highly actionable conditions via targeted sequencing of 17 genes, mitigating ELSI concerns raised from genome-scale sequencing in healthy populations

As part of this work, we evaluated and compared targeted sequencing technologies that could provide a cost-effective alternative to genome-scale sequencing (GSS) approaches:

- Roche/Nimblegen Heat-Seq molecular inversion probes (MIPs)
- Integrated DNA Technologies xGen lockdown hybridization capture probes

Category	Condition (gene)	Interventions
Cancer	Familial adenomatous polyposis (*APC*)	Colonoscopy, endoscopy screening, thyroid ultrasound, surgery
	MUTYH-associated polyposis (*MUTYH*)	Colonoscopy, endoscopy
	Lynch syndrome (*MLH1*)	Colonoscopy, endoscopy, endometrial biopsy, possible surgery (prophylactic hysterectomy and salpingo-oophorectomy)
	Lynch syndrome (*MSH2*)	
	Lynch syndrome (*MSH6*)	
	Lynch syndrome (*PMS2*)	
	Familial breast/ovarian cancer (*BRCA1*)	Breast imaging, prophylactic mastectomy and/or salpingo-oophorectomy
	Familial breast/ovarian cancer (*BRCA2*)	
	MEN2A/2B (*RET*)	Prophylactic thyroidectomy, serum metanephrine blood test
Cardiovascular	Long QT syndrome (*KCNQ1*)	Cardiology consultation, ECG, β-blocker medication if ECG is positive; implantable cardioverter-defibrillator if symptomatic
	Long QT syndrome (*KCNH2*)	
	Long QT syndrome (*SCN5A*)	
	Familial hypercholesterolemia (*LDLR*)	Lipid biochemical screening, pharmacotherapy if needed
	Marfan syndrome (*FBN1*)	Echocardiography, ophthalmologic screening
Other	Malignant hyperthermia (*RYR1*)	Avoidance of specific anesthetics
	Hereditary hemochromatosis (*HFE*)	Ferritin biochemical screening, phlebotomy
	α-1 Antitrypsin deficiency (*SERPINA1*)	Avoidance of exposure to smoke

Table 1. Characteristics of 11 screened conditions and 17 candidate genes. The GeneScreen Committee and Community Advisory Board reviewed and weighted that these genes, when mutated, confer high risk of these potentially detectable and preventable disorders.[3, 4]

Methods

We targeted those 17 genes in a subset of 58 participants enrolled in the GeneScreen study, using Heat-Seq MIPs and xGen lockdown probes.

Our metrics assessed the performance on three aspects important for clinical sequencing:

- Gene-level "adequate" coverage
- Variant calling comparability

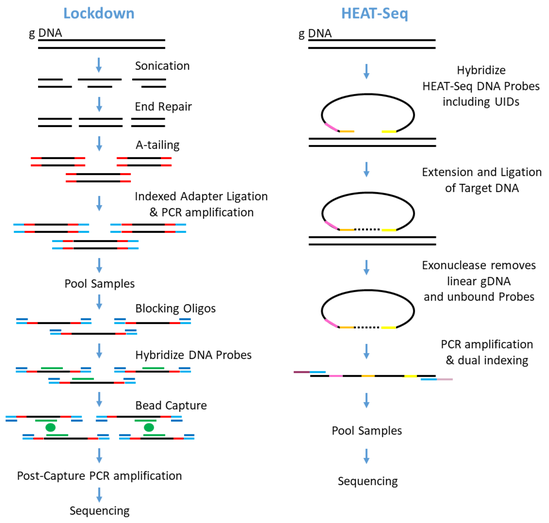

Figure 1. Library preparation workflows for IDT's Lockdown Probes and Roche's HEAT-Seq Probes

- Time requirement for Lockdown was about 8 - 20 hours[§], and for HEAT-Seq about 8 hours
- Recommended DNA input amount for Lockdown was about 10ng[§], and for HEAT-Seq was about 250ng
- Cost per sample for Lockdown (including probes) was about $556[§], and for HEAT-Seq (including MIPs) was about $122

[§]Used Kapa Hyper DNA kit, but differs with other library prep kits

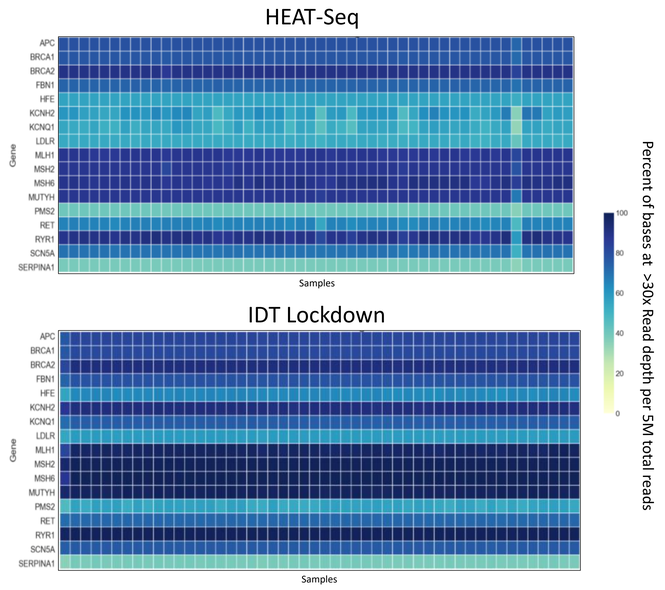

Figure 2. Heat maps illustrating the percent of nucleotides in Locus Reference Genomic (LRG) sequence regions in each gene in each sample covered at 30X or greater, normalized to 5M reads per sample.

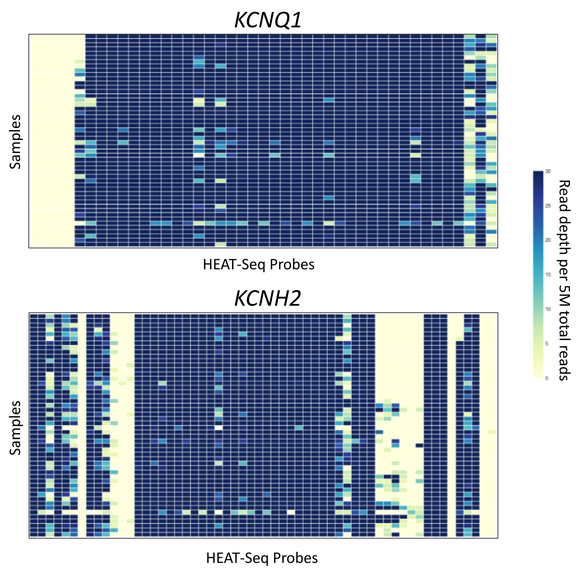

Figure 3. Representative heat maps depicting the number of reads (normalized to 5M reads per sample) for every HEAT-Seq probe in each sample for in *KCNQ1* and *KCNH2*. Many of the poor performing probes in *KCNQ1* and *KCNH2* targeted G-C rich exons.

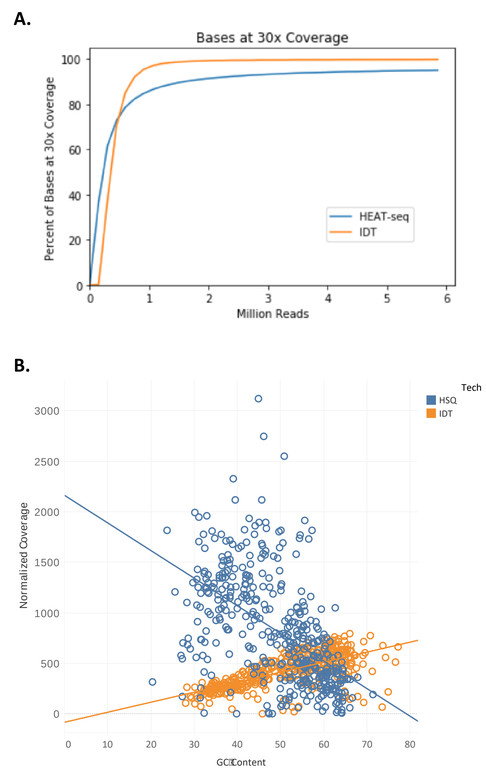

Figure 4. (A) The percent of targeted bases in coding regions in 50 samples sequenced at 30x coverage were estimated normalizing the data to various values of sequenced reads. HEAT-Seq required an average of 0.75M reads for 80% of bases to be sequenced at 30x coverage. Lockdown required an average of 0.6M reads for 80% of bases to be sequenced at 30x coverage. (B) The average coverage for each exon across 50 samples, normalized to 1M reads, is represented corresponding to the calculated G-C content of the exon. HEAT-Seq has a negative correlation between increasing G-C content and exon-level coverage, while Lockdown has a positive correlation.

Variant detection

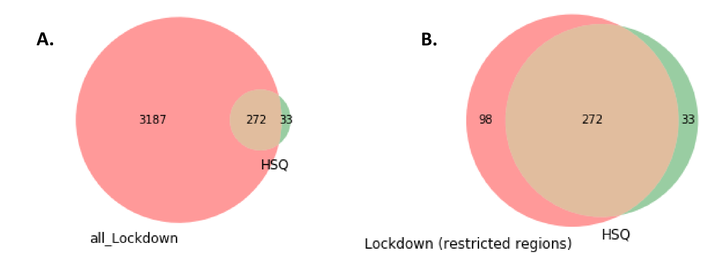

Figure 5. Variant detection comparison of (A) unique variants detected by all Lockdown vs HEAT-Seq probes, and (B) unique variants detected only in LRG regions by HEAT-Seq probes. 259/272 variants (95%) have been reported to ClinVar, excluding possible technological sequencing artifacts.

Conclusions & Future Implications

Both targeted probe technologies have their strengths and weaknesses:

- Roche HEAT-Seq has G-C rich limitations, but cost-effective
- IDT Lockdown has extra variant noise, but effective for G-C rich regions and smaller panels

With further optimization, targeted genomic sequencing could be a feasible and ethical option of screening the general population as it not only promises lower cost than GSS but would avoid generating large numbers of variants in genes with unknown or non-clinical significance.

Acknowledgments & References

We would like to acknowledge the GeneScreen Collaborative Group and funding from the National Human Genome Research Institute of the National Institutes of Health under grants P50HG004488 and U01HG006487.

[3]Anya E.R. Prince JD, MPP; Jonathan S. Berg MD, PhD; James P. Evans MD, PhD; Daniel E. Jonas MD, MPH; and Gail Henderson PhD. Genomic Screening of the General Adult Population: Key Concepts for Assessing Net Benefit with Systematic Evidence Reviews. *Genetics in Medicine* **17**, 441–443 (2015). doi:10.1038/gim.2014.129

[4]Michael C. Adams MS, James P. Evans MD, PhD; Gail E. Henderson PhD & Jonathan S. Berg MD, PhD. *Genetics in Medicine* **18**, 593–599 (2016). doi:10.1038/gim.2015.136