Exploring a Transcriptome-Forward Approach for Genetic Evaluation in Thoracic Aortopathy

Katherine Stanley[1], Benjamin Landis[2]

[1]Indiana University School of Medicine; [2]Indiana University School of Medicine, Department of Pediatrics

INTRODUCTION: Thoracic Aortic Aneurysm

- Thoracic aortic aneurysm (TAA): life-threatening aortopathy
  - Predisposes to aortic dissection and rupture
- Associated Disorders:
  - Marfan (MFS), Loeys-Dietz, and Vascular Ehlers-Danlos Syndromes
  - Genetic cause often not identified
  - Poor prediction of severity
- Transcriptomic analysis:
  - Diagnostic utility in other diseases
  - Not yet developed for TAA

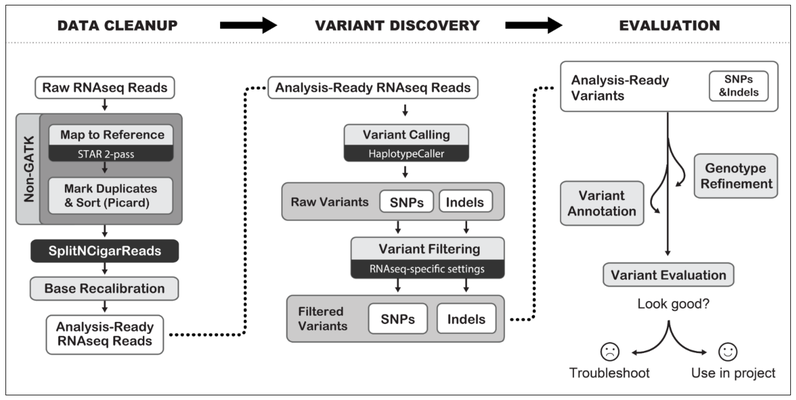
OBJECTIVE:

Implement the Genome Analysis Toolkit as a tool for high-throughput detection of novel variants in RNA-sequencing data from patients with TAA.

METHODS: Genome Analysis Toolkit (GATK)[1]

- RNA-Sequencing (RNA-Seq)
  - Proximal Aorta Samples: 63 TAA patients & 15 Controls
  - Smooth Muscle Cell (SMC) Lines in Culture
  - RNA-seq: 150 base pair paired-end reads at early passage
- Variant Calling and Annotation
  - GATK RNA-Seq Pipeline Best Practices
    - Exclusion of Mark Duplicates
  - Variant Annotation: ANNOVAR[2]
- Variant Selection and Evaluation
  - Selection of Singletons: heterozygous variants unique to one sample
  - Variant Evaluation: Filtration Parameters & Franklin by Genoox[3]

Figure 1: GATK RNA-Seq Pipeline[1]



ACKNOWLEDGEMENTS

- This project was funded, in part, with support from the Short-Term Training Program in Biomedical Sciences Grant funded, in part by T35 HL 110854 from the National Institutes of Health. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.
- Use of the Quartz Research System was supported in part by Lilly Endowment, Inc., through its support for the Indiana University Pervasive Technology Institute.

REFERENCES

1. Auwera GA, O'Connor BD. Genomics in the cloud: using Docker, GATK, and WDL in Terra. 2020.
2. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic acids research. 2010 Sep 1;38(16):e164.
3. https://franklin.genoox.com - Franklin by Genoox

RESULTS: Validation of GATK Pipeline

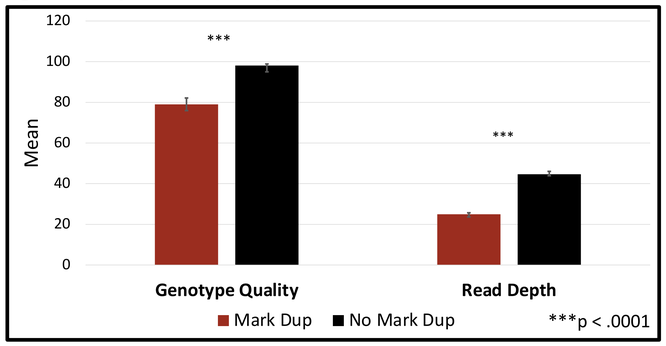
- Inspection with Integrated Genomics View (IGV)
- Confirmation of Known Variants

Table 1: Confirmation of Known Causal Variants

Gene	Patient	Variant	Functional Impact	Ref. Read Depth	Alt. Read Depth	Syndrome
FBN1	0802	chr15:48465820 G>A	Stopgain	287	37	Marfan
TGFBR1	0424	chr9:99137985 TCTC>T	Inframe Deletion	25	41	Loeys-Dietz

- Genotyping for common Single Nucleotide Polymorphism (SNP)
  - Taqman Assay at known SNP: COQ8B chr19:40705151 T>C
  - Validation: GATK-Taqman Assay Genotype Matching
    - Duplicates Marked: 49 out of 51 patients
    - Not Marked: 51 out of 51 patients

Figure 2: Impact of Mark Duplicates Tool at Known SNP



RESULTS: Candidate Causal Variants

Table 3: Candidate Causal Variants**

Gene	Patient	Variant	Franklin Interpretation	Functional Impact	Ref. Read Depth	Alt. Read Depth	Phenotype
FBN1	0826	chr15:48495215 C>G	Pathogenic	Missense	314	309	TAV, MFS
FBN1	0997	chr15:48513657 A>G	Likely Pathogenic	Missense	616	665	TAV, MFS
FBN1	0460	chr15:48520779 C>T	Likely Pathogenic	Missense	581	735	BAV
TGFB3	0790	chr14:75980722 C>A	Likely Pathogenic	Stopgain	23	6	TAV
HEY2	0853	chr6:125759461 GCT>G	Likely Pathogenic	Frameshift Deletion	33	10	TAV
TGFBR1*	0980	chr9:99149218 G>C	Likely Pathogenic	Missense	84	10	BAV
GATA4**†	0758	chr8:11748985 G>A	Likely Pathogenic	Stopgain	15	6	BAV
GATA4†	0771	chr8:11749048 G>A	Likely Pathogenic	Missense	5	4	BAV
ROBO4†	0741	chr11:124886790 GA>G	Likely Pathogenic	Frameshift Deletion	4	2	BAV
ROBO4†	0632	chr11:124891306 CT>C	Likely Pathogenic	Frameshift Deletion	6	2	BAV
ADAMTS19†	0704	chr5:129528547 G>GA	Likely Pathogenic	Frameshift Insertion	8	6	BAV

*Tagged for filtration; **Additional candidate causal variants not included;
† BAV gene, dominant inheritance; BAV: Bicuspid Aortic Valve; TAV: Tricuspid Aortic Valve

RESULTS: Variant Calling in Aortopathy Genes

Table 2: Variant Counts in Aortopathy Genes

	All Variant Sites	Coding Variant Sites	SNV Singleton	Indel Singleton	Candidate Causal Variants	VUS Variants
Total	4,823,462	236,495	139,050	15,349		
31 Genes	10,399	511	330	33	53	227
FBN1	779	20	15	0	6	1
COL3A1	311	8	3	0	0	0
SMAD3	695	4	1	0	0	0
TGFB2	236	3	2	0	0	1
TGFBR1	145	8	5	2	3	2
TGFBR2	241	5	2	0	0	0
ACTA2	241	1	1	0	0	1
MYLK	375	48	29	1	4	18
LOX	151	5	3	0	0	1
MYH11	4	0	0	0	0	0
PRKG1	1,923	17	15	0	0	9
ELN	831	6	4	0	0	2
FLNA	280	10	1	1	1	0
FBN2	200	49	29	3	6	19
SMAD4	294	3	2	1	1	0
NOTCH1**	352	159	123	14	18	94
SKI	373	17	15	0	0	11
BGN	268	3	0	0	0	0
FOXE3	0	0	0	0	0	0
HCN4	12	3	0	0	0	0
MAT2A	42	1	0	0	0	0
MFAP5	368	1	1	0	0	0
SMAD2	530	1	0	0	0	0
TGFB3	156	10	8	1	2	6
ARIH1	601	8	3	0	0	2
HEY2	105	9	5	1	1	5
SMAD6†	378	34	25	6	5	23
GATA4†	274	17	11	0	2	7
GATA5†	81	18	11	0	0	10
ROBO4†	67	27	12	2	3	11
ADAMTS19†	86	16	4	1	1	4

†BAV (Bicuspid Aortic Valve) associated gene, dominant inheritance
**TAA and BAV associated gene

CONCLUSIONS

- GATK RNA-seq pipeline expands our prior IGV analysis
  - Supports diagnostic utility of transcriptomic analysis in aortic SMCs
  - Detection of Novel Variants in Aortopathy-related Genes
- Limitations
  - Confirmation with Whole Genome Sequencing
  - Optimization of filtration parameters in pipeline
- Future Directions
  - Further pipeline optimization and experimentation
  - Exploration of UTR and intronic variants
  - Aberrant splicing analysis
  - Differential expression analysis
  - SNP Array: BAV and TAA vs. BAV alone

SCIENTIFIC IMPACT

Identification of novel aortopathy-associated variants could improve future disease detection and prediction of severity.