High-Throughput Allele-Specific Expression Analysis Can Detect Allelic Imbalance in Clinical Patient Samples as a Proxy for Quantitative Measures of Gene Expression

Hoda Mirsafian[1], Heather Zimmermann[1], Laura Hudish[1], Sami Belhadj[1], Donovan Cheng[2], Elizabeth Chao[1], Rachid Karam[1], Matthew Schultz[2]

[1]Ambry Genetics, Aliso Viejo, CA, [2]REALM IDX Aliso Viejo, CA

hmirsafian@ambrygen.com

Ambry Genetics®

BACKGROUND

- RNA analysis is a powerful tool for variant classification.
- Most RNA analysis performed on clinical RNA-Seq data involves the detection of qualitative differences resulting from aberrant splicing.
- Identifying quantitative differences in allele-specific transcript production is more difficult, historically relying on quantitative real-time PCR (qRT-PCR).
- We evaluated whether allele-specific expression (ASE) analysis of RNA-seq data could be used to identify allelic differences in expression.

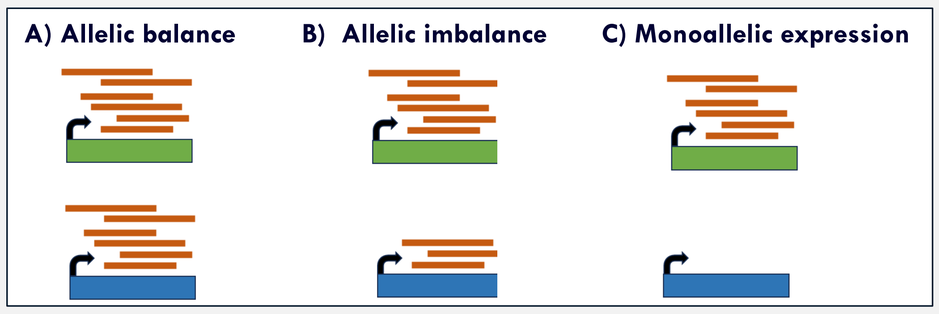


FIGURE 1. Schemes illustrating different patterns of allelic expression A) Allelic balance: both alleles are expressed equally. B) Allelic imbalance: one allele has significantly higher expression than the other. C) Monoallelic expression: one allele is expressed in the presence of a heterozygous genome.

METHODS

FIGURE 2. A flowchart of the allele-specific expression analysis pipeline. The RNA-Seq and mapping was carried out using STAR (*Dobin A, et al. 2013*) per individual. The VCF file containing SNPs was generated from DNA-Seq data from the same individual. phASER package (*Castel SE, et al. 2016*) employed RNA-Seq reads to phase heterozygous variants relative to one another within a given gene and produced a gene-level haplotypic expression matrix for each individual. The magnitude of the allelic imbalance was quantified by allelic fold change (aFC). The statistical significance of the imbalance was evaluated using binomial-based statistics to account for the count-based nature of the data.

RESULTS

- As proof of concept, ASE analysis was performed on samples from two patients with the *MLH1* c.-27C>A pathogenic variant. This variant is associated with reduced *MLH1* promoter activity and is expected to result in ASE leading to Lynch syndrome-related cancers.
- Significant ASE was detected for c.-27C>A as well as downstream heterozygous polymorphisms in both patients (c.655A>G in patient 1; c.85G>T and c.655A>G in patient 2) (Figure 3A). Controls displayed allelic balance (Figure 4A).
- Analysis was extended to patients with novel likely pathogenic variants impacting the canonical donor splice sites of the noncoding first exons of *TP53* (c.-29+1G>C) and BRCA2 (c.-40+2T>C).
  - Significant ASE was observed for heterozygous polymorphisms in *TP53* (c.215C>G) and *BRCA2* (c.-52A>G, c.-26G>A, c.1114A>G, c.3396A>G, c.7242A>G) in patients (Figure 3B,C) but not in controls (Figure 4B, C).

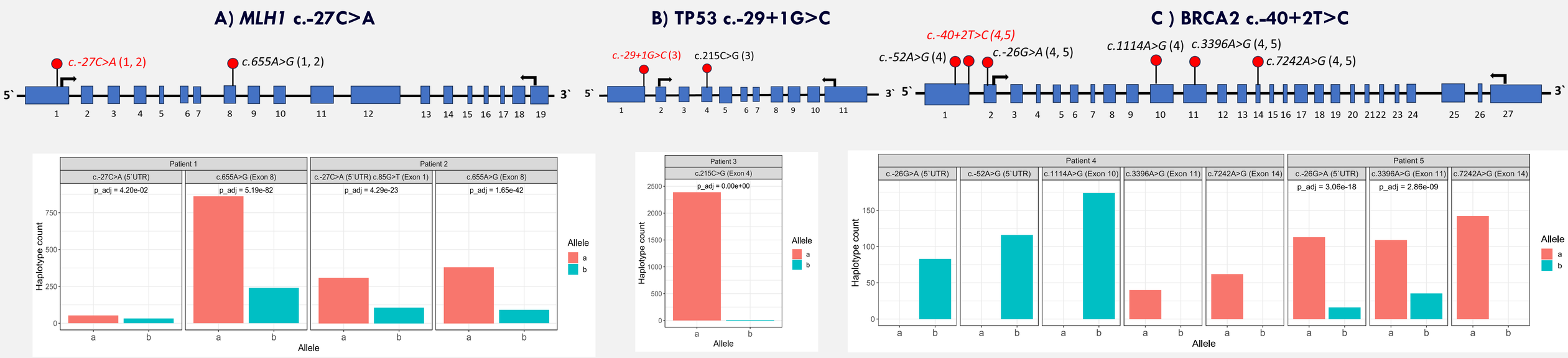


FIGURE 3. Schematic representations of the A) *MLH1* gene, B) *TP53* gene, and C) *BRCA2* gene. The genomic positions of the causative variants (red font) and the polymorphisms with significant ASE (black font) are indicated by the red circles. Numbers in the brackets indicate patient IDs. The boxplots below represent the haplotype count for each of these variant alleles across patients. P-values were calculated using the Binomial test.

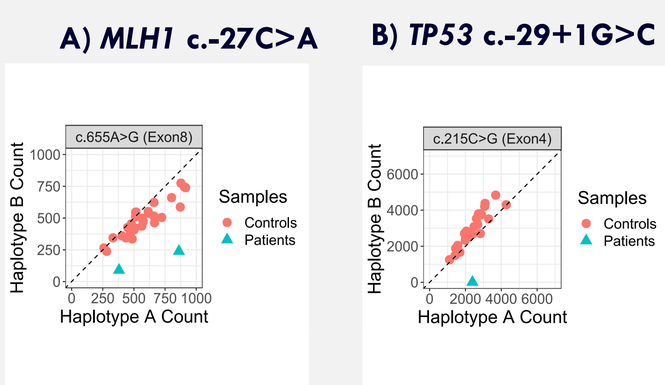
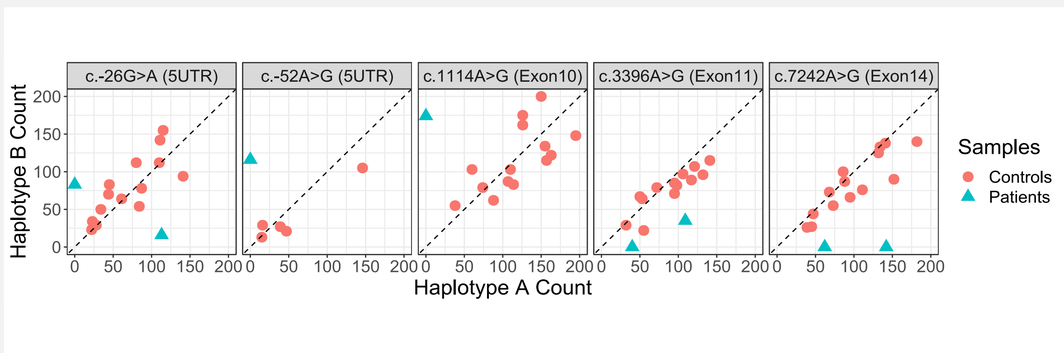


FIGURE 4. Scatter plots representing the haplotype count for each variant allele across patients and controls. The red circles represent the patients, while the blue triangles represent the control samples tested for each of those variants.



TAKE HOME POINTS

- Although preliminary, these data indicate that ASE analysis using phASER is capable of detecting RNA allelic imbalance associated with variants identified through clinical DNA samples.
- These results raise the tantalizing possibility that high-throughput ASE analysis could be used to identify patients with allele-specific expression of unknown origin.
- Future directions of this study include:
  - Validating the method for additional variants and variant types.
  - Assessing potential candidate variants in a variety of genes using our in-house database of >300000 paired DNA-seq and RNA-seq samples.
  - Exploring the range of allelic expression ratios in larger control populations across other SNPS and additional genes.

SCAN ME