Detection of Novel Alu Insertions by Next-generation Sequencing of Hereditary Cancer Genes

Michael Kennemer[1], Nila Patil[1], Michelle Zeman[1], Martin Powers[1], and Karen Ouyang[1]
[1]Invitae, San Francisco, CA

Introduction

Alu elements are the most prevalent mobile repetitive DNA sequences in humans, accounting for approximately 11% of the human genome[1]. While the majority of Alu elements represent variation in the human genome and are not deleterious, Alu-mediated pathogenic rearrangements and insertions have been demonstrated in a number of diseases through disruption of either a coding region or splice site[2]. It has been estimated that there is one new Alu insertion per 20 human births[3], which could lead to approximately one in every 1,000 cases of human genetic disease[4]. Here we identify novel Alu insertions in coding regions of ATM, BRCA2, and BRIP1 where the ATM event is found in three unrelated individuals. These findings suggest that the incidence of this type of variation in not exceedingly rare and illustrates the importance of detection of Alu insertions in hereditary cancer testing.

Methods

Invitae offers a hereditary cancer syndromes panel as a clinical diagnostic next-generation sequencing (NGS) test using a custom designed hybridization-based targeted enrichment bait set and 150bp paired-end sequence reads. NGS coverage depth varies, with a minimum of 50X depth and an average of 350X. While analysis tools for detecting readthrough and copy-number variants are well established, variants that are outside this spectrum are difficult to detect. Additional tools and analysis are generally required to detect and characterize complex variants such as Alu element insertions and genomic rearrangements. As a starting point, rudimentary tools for identifying low confidence reference regions that require manual review can be incorporated into the analysis workflow. Our analysis pipeline, for example, will identify split reads signals where a high proportion of NGS reads are soft-clipped. Many split reads signals are artifactual and can be annotated and filtered accordingly. The remaining split reads signals, however, may be indicative of real structural variation and are manually reviewed. High quality soft-clipped bases can be consolidated into a consensus sequence and aligned to the reference genome or to a restricted set of sequence contigs, including representative Alu sequences. Currently, these ad hoc steps are done outside of our analysis pipeline but can be incorporated into analysis through either a Structural Variant (SV) caller or a local de novo (LDN) approach.

Results

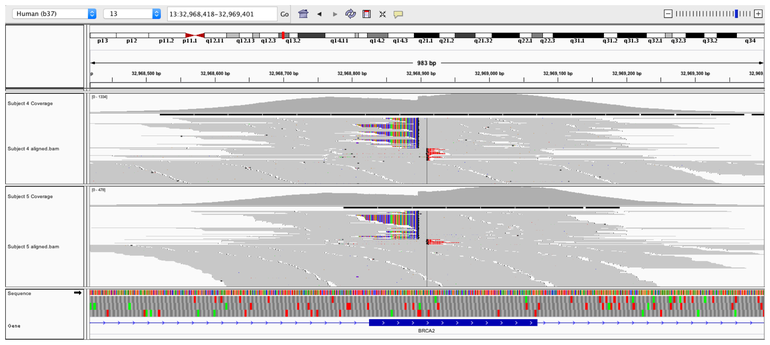
In addition to many known intronic Alu insertion events, including the known Portuguese founder mutation[5] of an Alu insertion in exon 3 of BRCA2, we have identified three novel Alu insertion events that interrupt the coding region of 6 clinical samples tested on the hereditary cancer syndromes panel (Table 1).

Table 1: Novel Alu insertion events detected in exons

	Gene	Variant Description
Subject 1	ATM	Exon 50, NM_000051.3:c.7387_7388insAlu
Subject 2	ATM	Exon 50, NM_000051.3:c.7387_7388insAlu
Subject 3	ATM	Exon 50, NM_000051.3:c.7387_7388insAlu
Subject 4	BRCA2	Exon 25, NM_000059.3:c.9342_9343insAluY
Subject 5 (sibling of 4)	BRCA2	Exon 25, NM_000059.3:c.9342_9343insAluY
Subject 6	BRIP1	Exon 5, NM_032043.2:c.505_506insAlu

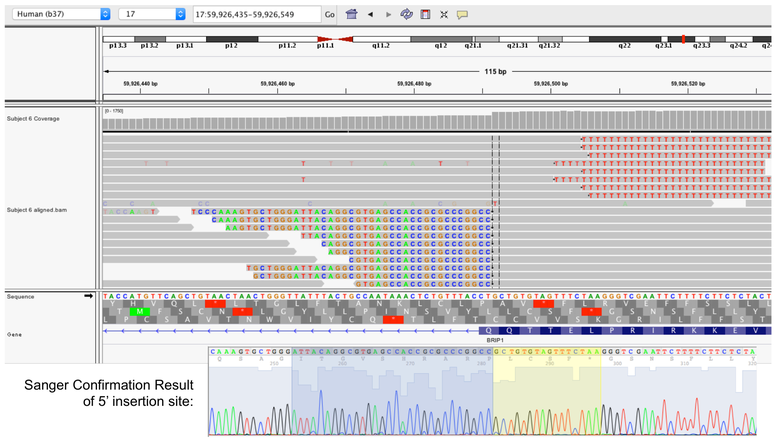
Each detected Alu insertion event is supported by soft-clipping in 20-40% of reads and identified by two corresponding split reads signals at both sides of the insertion site (Figures 1 & 2).

Fig. 1: IGV screenshot of BRCA2 exon 25 Alu insertions



The Alu insertions were all confirmed by Sanger sequencing at the insertion site (Figure 2). While gel shift assays were also performed, PCR bias towards the shorter product often resulted in non-interpretable gel shift assay results.

Fig. 2: BRIP1 exon 5 Alu insertion with Sanger confirmation.



While soft-clipped segments of reads supporting Alu were observed to cover 50-90bp of the Alu insertion, local de novo assembly performed using Kragle[6], an Invitae-developed LDN assembler, was able to extend the consensus insertion sequence roughly 250bp into each Alu. This enables the confident identification of the Alu subfamily, as all 6 of these Alu insertions most closely resemble AluY.

Additional studies are needed to determine sensitivity and specificity.

Clinical Significance

Each of these variants is expected to result in a frame-shift followed by a premature stop codon, likely resulting in loss of the mRNA through nonsense-mediated decay or production of a non-functional truncated protein.

ATM, Exon 50, NM_000051.3:c.7387_7388insAlu

- The ATM gene is associated with an increased risk for autosomal dominant breast cancer and possibly other cancers (PMID: 15928302). Females who are carriers of one pathogenic ATM finding may have a 2- to 5-fold increased risk for breast cancer (PMID: 3574400, 15928302, 16832357).
- ATM is also associated with autosomal recessive ataxia-telangiectasia (A-T) (MedGen UID: 439).

BRCA2, Exon 25, NM_000059.3:c.9342_9343insAlu

- The BRCA2 gene is associated with autosomal dominant hereditary breast and ovarian cancer syndrome (MedGen UID: 151793). The lifetime risk for female breast cancer in individuals with a pathogenic BRCA2 sequence variant is 40-85%. The lifetime risk for ovarian, fallopian tube, or peritoneal cancer is 16-27% (PMID: 9145676, 9497246). The risk for male breast cancer is 7-8% (PMID: 20587410). There are also increased risks for melanoma, prostate cancer (20%), and pancreatic cancer (2-3%) (PMID: 10433620).
- BRCA2 is also associated with autosomal recessive Fanconi anemia (MedGen UID: 325420).

BRIP1, Exon 5, NM_032043.2:c.505_506insAlu

- The BRIP1 gene is associated with an increased risk for autosomal dominant breast and ovarian cancer (PMID: 17033622, 21964575). Pathogenic variants in BRIP1 have been associated with an approximate two-fold increased risk for breast cancer and an eight-fold risk for ovarian cancer (PMID: 21964575).
- BRIP1 is also associated with autosomal recessive Fanconi anemia (MedGen UID: 323015).

Conclusions

While we have demonstrated the ability to detect Alu insertions in NGS targeted enrichment data, better tools are needed to do so reproducibly and further studies are required to establish sensitivity of these new methods.

Furthermore, multiple observations of the ATM exon 50 insertion implies that Alu insertion detection needs to be an assay design consideration for clinical diagnostic testing of heredity cancer syndromes. PCR-based enrichment assays may miss Alu insertions completely while hybridization-based assays may lose sensitivity due to reference bias, especially in regions with minimal redundancy in probe design.

References

1. Lander et al., Nature. 2001 Feb 15;409(6822):860-921
2. Xing et al., Genome Res. 2009 Sep;19(9):1516-26
3. Belancio et al., Genome Res. 2008 Mar;18(3):343-58
4. Deininger et al., Mol Genet Metab. 1999 Jul;67(3):183-93
5. Machado et al., JCO May 20, 2007 vol. 25 no. 15 2027-2034
6. Konvicka et al., ASHG, 2014