The Francis Crick Institute
1 Midland Road
London, NW1 1AT, United Kingdom

To whom it may concern;

Here, we present a short report entitled, **"Evaluation of single nucleotide variants in exon 5 of *BRCA1* using a redesigned Saturation Genome Editing assay"**.

We have chosen to share the data included here directly with expert clinicians as an interim report in case it may support efforts to classify variants in *BRCA1*. We are taking this unusual step because our results contradict previously published findings (Findlay et al. 2018) regarding the functional consequences of certain variants that have been reported in patients.

Critically, this work has not been peer reviewed. It is part of a larger study on variants in *BRCA1* that is ongoing. All data presented were obtained as part of a research study being conducted at the Francis Crick Institute. As such, we make no claims regarding the data's ultimate validity for adjudicating genetic variants identified in patients. "Function scores" provided here are subject to change at a later date and may be altered in conjunction with a broader analysis that is ongoing.

For further information on the Saturation Genome Editing assay, please see Findlay et al, 2018 (PMID: 30209399).

Regards,

Gregory M. Findlay, MD PhD
Group Leader
The Francis Crick Institute, London

Evaluation of single nucleotide variants in exon 5 of *BRCA1* using a redesigned Saturation Genome Editing assay

Phoebe Dace[1] and Gregory M. Findlay[1,2]

[1] The Genome Function Laboratory, The Francis Crick Institute, London, UK
[2] Correspondence to greg.findlay@crick.ac.uk

Introduction

Functional evidence from laboratory assays can be used to guide the clinical interpretation of genetic variants[1]. Previously, we performed a method called Saturation Genome Editing (SGE) to functionally classify single nucleotide variants (SNVs) across 13 coding exons of *BRCA1*[2]. Resulting data for *n* = 3,893 SNVs was calibrated using established ClinVar annotations available at time of publication[3]. The high predictive power with which the SGE data distinguished pathogenic variants from benign variants in *BRCA1* has led to the data being widely used to aid clinical variant interpretation[4–7].

Variants included in the initial SGE assay that remain beyond definitive clinical interpretation include those potentially affecting splicing at the 3' end of exon 5. Specifically, multiple variants at c.301+1G currently have conflicting interpretations in ClinVar. Although these variants alter the canonical exon 5 donor site, uncertainty stems from reports of an in-frame transcript with a 9-bp deletion observed in patients, as noted in ClinVar[8].

Here, we employ an optimized SGE assay designed specifically to assay variants at the 3' end of exon 5. Similarly to previous work, SGE function scores successfully distinguish known pathogenic and benign variants across the complete exon. However, in contrast to previous findings, our new results indicate that variants which disrupt canonical splicing at the exon 5 donor site do not lead to loss of *BRCA1* function in HAP1 cells. This indicates that the previous SGE data for variants disrupting splicing near the 3'-end of exon 5 should not be used as evidence for pathogenicity.

Methods

SGE was performed in the human HAP1 cell line, similarly to before[2]. Briefly, a library of variants was introduced to a large pool of HAP1 cells via CRISPR editing to include all possible SNVs at each position across exon 5 of *BRCA1* (NM_007294.3) and extending into adjacent introns. Each SNV's effect on cell growth was quantified by performing next-generation sequencing (NGS) of edited cell populations sampled on day 5 and day 14 post-transfection. The ratio of each variant's abundance in sequencing data on day 14 normalized to day 5 was calculated to determine a "function score", which was averaged and normalized across *N* = 2 replicates. Variants with significant function scores were determined by fitting a normal distribution to synonymous SNVs and performing a correction for multiple hypothesis testing. Variants with

significantly low function scores ($q < 0.01$) were deemed loss-of-function (LoF), whereas those with function scores not significantly different to 0 ($q > 0.05$) were deemed neutral.

Results

The SGE method relies upon CRISPR-mediated genome editing to introduce variants into human cells via homology-directed DNA repair. During editing, at least one variant is required to efficiently block re-cutting of edited DNA. Without a “blocking variant”, successfully edited alleles can be depleted over time due to ongoing CRISPR/Cas9 cleavage and mutation, an effect that precludes acquisition of accurate SGE data.

A synonymous blocking variant, c.297G>A, was installed in exon 5 of *BRCA1* during previous SGE experiments, as described[2]. Effectively, all SNVs assayed in exon 5 before were tested in the context of this synonymous variant. c.297G>A was chosen to block re-cutting based on a lack of sequence conservation at the position and by virtue of being more than 3 bp away from the exon’s end. However, whether c.297G>A may impact splicing of alternative transcripts has not been assessed. Indeed, the proximity of c.297G>A to the reported alternative splice site in exon 5 raises the possibility that it may interfere with splicing of transcript isoforms that arise due to canonical splice site disruption. Therefore, we created a new SGE library without any synonymous blocking variants within 70 bp of the 3’ end of the exon 5.

We next performed SGE with this library to measure function scores for $n = 338$ SNVs. Similarly to before, function scores distinguished ClinVar “pathogenic” and ”likely pathogenic” variants from “benign” and “likely benign” variants with 100% accuracy (**Fig. 1A**), validating the assay’s technical quality. However, in contrast to published data, SNVs disrupting the canonical donor site did not score as LoF (**Fig. 1B**). Whereas the median nonsense function score was -1.89, scores for SNVs from c.301+1 to c.301+10 ranged from -0.26 to 0.70, with a median score of 0.09.

In light of these results, we have defined a set of discordant variants that previously scored as LoF by SGE but are deemed neutral in the present study (**Table 1**). In total, this includes 25 SNVs**,** the majority of which are located in close proximity to the 3’ end of the exon.

Discussion

We applied knowledge of an alternative transcript isoform identified in patients to design and implement an improved SGE assay for SNVs in exon 5 of *BRCA1*. Our results once more validate the high accuracy of SGE data for informing pathogenicity. However, in comparison to previous findings, many SNVs at or near the exon 5 splice donor site did not lead to LoF, an effect attributable to improved library design allowing splice-altering SNVs to be introduced without other variants in close proximity.

This result indicates that disruption of the canonical donor site does not lead to *BRCA1* LoF in HAP1 cells. Indeed, the finding supports the hypothesis that disruption of the canonical exon 5

donor site leads to increased production of the in-frame isoform previously described, which in turn rescues BRCA1 function. This would indicate that the last three amino acids encoded by exon 5 are dispensable for normal BRCA1 function, although this was not directly tested.

Considering the data reported here were generated using the same cell line and assay as before (HAP1 essentiality), this time without simultaneous introduction of c.297G>A, we suggest that the previous data for variants from c.301+1 to c.301+10 no longer be used as evidence supporting pathogenicity. All SNVs previously deemed LoF that scored discordantly here are provided in **Table 1**, independent of whether they've been previously reported in ClinVar. Further work drawing on orthogonal lines of evidence will prove valuable for definitively classifying these variants in the future.

References

1. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med.* **17**, 405–424 (2015).
2. Findlay, G. M. *et al.* Accurate classification of BRCA1 variants with saturation genome editing. *Nature* **562**, 217–222 (2018).
3. Landrum, M. J. *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* **44**, D862–8 (2016).
4. Kim, H.-K. *et al.* Impact of proactive high-throughput functional assay data on BRCA1 variant interpretation in 3684 patients with breast or ovarian cancer. *J. Hum. Genet.* **65**, 209–220 (2020).
5. Fayer, S. *et al.* Closing the gap: Systematic integration of multiplexed functional data resolves variants of uncertain significance in BRCA1, TP53, and PTEN. *Am. J. Hum. Genet.* **108**, 2248–2258 (2021).
6. Lyra, P. C. M., Jr *et al.* Integration of functional assay data results provides strong evidence for classification of hundreds of BRCA1 variants of uncertain significance. *Genet. Med.* **23**, 306–315 (2021).
7. Schiabor Barrett, K. M. *et al.* Clinical validation of genomic functional screen data: Analysis of observed BRCA1 variants in an unselected population cohort. *HGG Adv* **3**, 100086 (2022).
8. Thomassen, M. *et al.* Characterization of BRCA1 and BRCA2 splicing variants: a collaborative report by ENIGMA consortium members. *Breast Cancer Res. Treat.* **132**, 1009–1023 (2012).

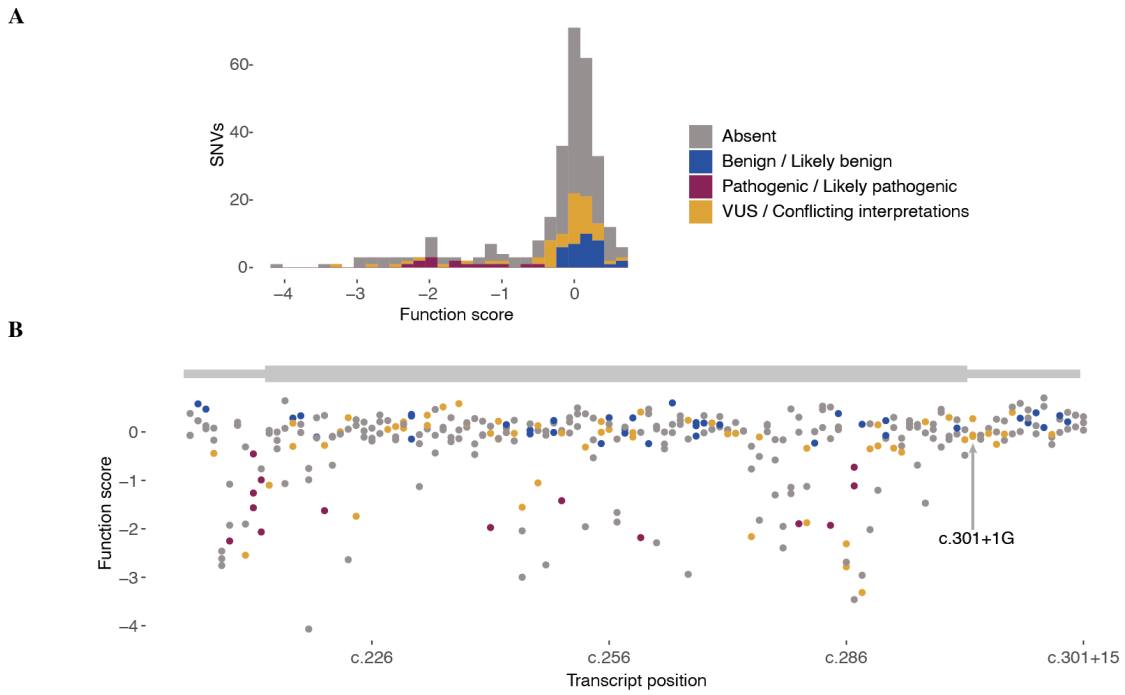

Figure 1. A redesigned SGE assay *BRCA1* exon 5 reveals a lack of LoF variants near the donor splice site. **A**, Function scores were derived for *n* = 338 SNVs using an optimized version of a previously published protocol[2]. Function scores successfully distinguish ClinVar "Pathogenic" and "Likely pathogenic" variants from those deemed "Benign" or "Likely benign". **B**, Function scores for SNVs are plotted by transcript position and colored by ClinVar status. The arrow marks the position of three SNVs with conflicting interpretations in ClinVar at the canonical donor site (c.301+1G). None of these three variants scored as LoF, nor did others in intron 5.

Table 1. Variants functionally reclassified using new SGE data for BRCA1 exon 5. All variants assayed in the new SGE experiment that were previously measured to be LoF but are now scored as functional (i.e. neutral) are listed. ClinVar data was obtained on 1 May 2023. Old function scores were obtained from Findlay et al. 2018.

pos hg38	ref	alt	cHGVS	pHGVS	consequence	clinvar	function.score.old	function.score.new	p.value	q.value	old.func.class	new.func.class
43104956	C	G	c.213G>C	p.R71S	NON_SYNONYMOUS	Absent	-2.485	-0.029	0.217	0.482	LOF	FUNC
43104954	C	T	c.215G>A	p.S72N	NON_SYNONYMOUS	Absent	-1.375	0.073	0.406	0.679	LOF	FUNC
43104938	C	A	c.231G>T	p.T77=	SYNONYMOUS	Likely benign	-1.907	-0.148	0.079	0.250	LOF	FUNC
43104926	T	A	c.243A>T	p.Q81H	NON_SYNONYMOUS	Absent	-1.394	0.083	0.425	0.693	LOF	FUNC
43104925	G	A	c.244C>T	p.L82F	NON_SYNONYMOUS	Absent	-1.799	-0.186	0.053	0.193	LOF	FUNC
43104919	C	T	c.250G>A	p.E84K	NON_SYNONYMOUS	Uncertain significance	-1.415	-0.022	0.228	0.494	LOF	FUNC
43104918	T	A	c.251A>T	p.E84V	NON_SYNONYMOUS	Absent	-1.578	-0.022	0.228	0.494	LOF	FUNC
43104915	T	C	c.254A>G	p.E85G	NON_SYNONYMOUS	Absent	-1.456	-0.164	0.067	0.232	LOF	FUNC
43104908	C	T	c.261G>A	p.L87=	SYNONYMOUS	Likely benign	-2.272	-0.243	0.028	0.106	LOF	FUNC
43104908	C	A	c.261G>T	p.L87F	NON_SYNONYMOUS	Absent	-2.191	0.375	0.914	0.987	LOF	FUNC
43104867	C	T	c.301+1G>A	NA	CANONICAL_SPLICE	Conflicting interpretations of pathogenicity	-2.043	-0.063	0.168	0.417	LOF	FUNC
43104867	C	G	c.301+1G>C	NA	CANONICAL_SPLICE	Conflicting interpretations of pathogenicity	-3.592	0.272	0.792	0.950	LOF	FUNC
43104867	C	A	c.301+1G>T	NA	CANONICAL_SPLICE	Conflicting interpretations of pathogenicity	-2.079	-0.095	0.128	0.349	LOF	FUNC
43104866	A	C	c.301+2T>G	NA	CANONICAL_SPLICE	Absent	-3.785	-0.151	0.076	0.247	LOF	FUNC
43104866	A	G	c.301+2T>C	NA	CANONICAL_SPLICE	Uncertain significance	-2.222	-0.068	0.161	0.406	LOF	FUNC
43104866	A	T	c.301+2T>A	NA	CANONICAL_SPLICE	Absent	-2.424	-0.069	0.160	0.406	LOF	FUNC
43104865	T	A	c.301+3A>T	NA	SPLICE_SITE	Absent	-2.006	0.072	0.403	0.678	LOF	FUNC
43104865	T	C	c.301+3A>G	NA	SPLICE_SITE	Conflicting interpretations of pathogenicity	-2.165	-0.032	0.212	0.481	LOF	FUNC
43104864	T	A	c.301+4A>T	NA	SPLICE_SITE	Absent	-1.693	-0.002	0.262	0.520	LOF	FUNC
43104864	T	C	c.301+4A>G	NA	SPLICE_SITE	Uncertain significance	-2.487	-0.255	0.024	0.095	LOF	FUNC
43104864	T	G	c.301+4A>C	NA	SPLICE_SITE	Absent	-3.058	0.262	0.778	0.935	LOF	FUNC
43104863	C	G	c.301+5G>C	NA	SPLICE_SITE	Absent	-2.604	-0.173	0.061	0.216	LOF	FUNC
43104863	C	T	c.301+5G>A	NA	SPLICE_SITE	Uncertain significance	-4.460	-0.045	0.193	0.461	LOF	FUNC
43104862	A	G	c.301+6T>C	NA	SPLICE_SITE	Uncertain significance	-1.949	0.404	0.935	0.994	LOF	FUNC
43104858	C	G	c.301+10G>C	NA	INTRONIC	Absent	-1.336	0.526	0.985	0.999	LOF	FUNC