Novel Processed Pseudogenes Identified in a Hereditary Pan-Cancer NGS Panel

Debora Mancini-DiNardo[1]; Kenneth Pond[1]; Jeff Terry[1]; Shujuan Pan[1]

1. Myriad Genetics, Inc., Salt Lake City, UT, USA

Background

- Pseudogenes, whether non-processed (with introns) or processed (without introns), pose great challenges to short-read next generation sequencing (NGS) platforms due to the high level of homology they share with their native genes.[a]
- Although filtration of pseudogene-specific reads can reduce the interference of known pseudogene sequences, insufficient filtering can produce false positive signals in copy number and/or sequencing variants. Thorough investigation is needed to distinguish pseudogene interference from true pathogenic variations in the native genes.
- Here we present the identification and characterization of novel processed pseudogenes found in multiple genes initially identified via a hereditary pan-cancer NGS Panel.

Methods

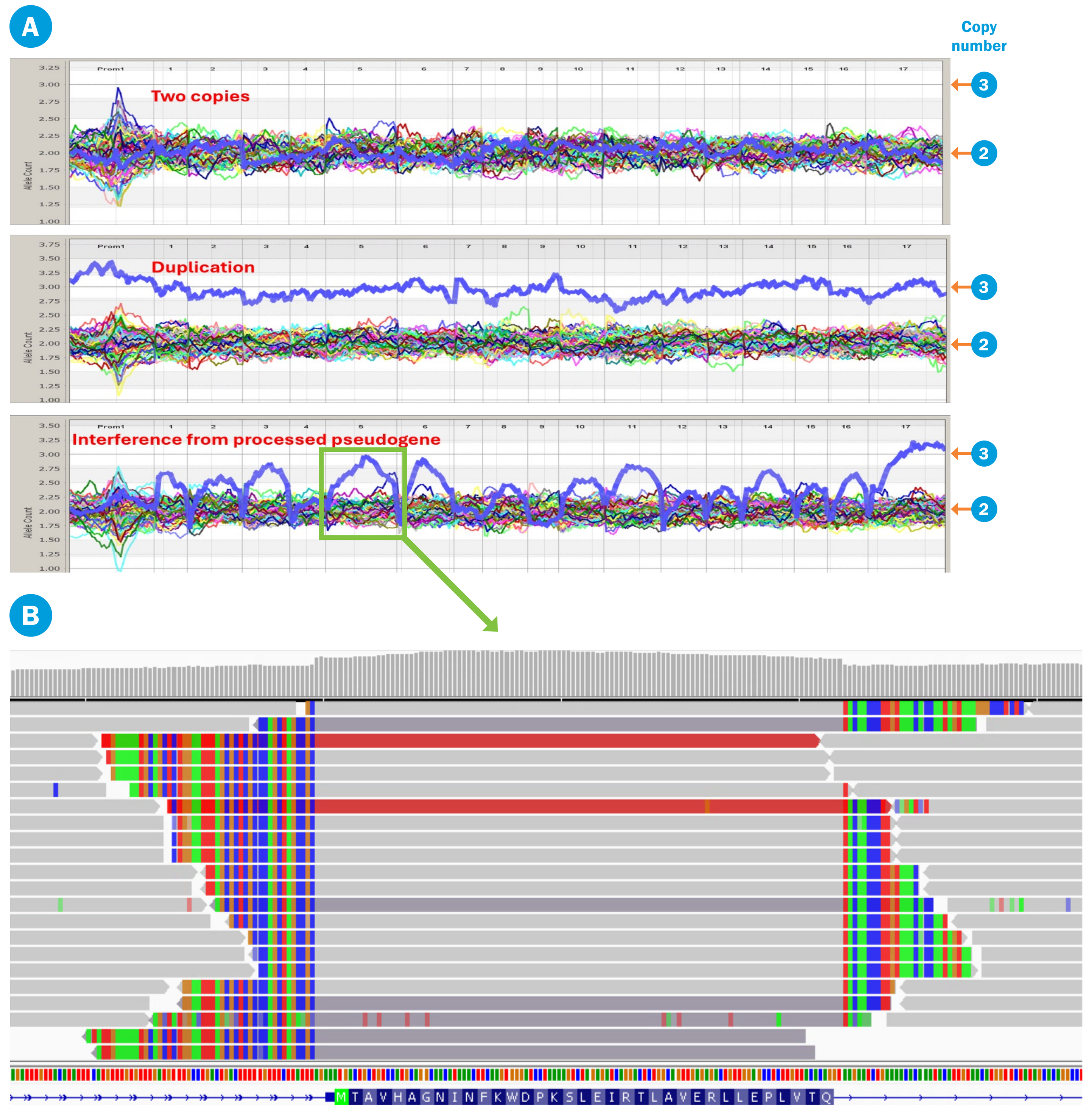
- Genes included in the hereditary pan-cancer NGS panel from December 2018 to April 2026 were evaluated for sequencing and copy number changes.
- In cases where the presence of a processed pseudogene was suspected, PCR primers were designed to flank the potential pseudogene insertion points, and the regions in question were subsequently amplified by long-range PCR and sequenced by Sanger.
- The number of genes in the panel ranged from 25 (2018) to 63 (current). Some of the affected genes were only included in the panel during certain periods (e.g., *RPS20* was included for analysis in the hereditary cancer panel from 2018 to 2022, and from end of 2025 to present; *CTNNA1* and *VHL* were included from 2022 to present; *MAX* was included from the end of 2025 to present).

a) Zhang Z, Harrison PM, Liu Y, Gerstein M. Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. Genome Res. 2003 Dec;13(12):2541-58. doi: 10.1101/gr.1429003. PMID: 14656962; PMCID: PMC403796.

Results

- 1,735,680 individuals were tested with NGS between December 2018 and April 2026, though not all individuals were tested for the same set of genes.
- Based on specific attributes in the NGS data, processed pseudogenes were detected in 69 individuals in 10 genes, including *RPS20, CTNNA1, MAX, CDH1, PMS2, SDHB, PTEN, FH, VHL,* and *EPCAM.*

Figure 1. Characteristics of processed pseudogenes on NGS. Each panel shows a batch run for *CTNNA1* with the sample of interest highlighted in blue. **(A)** Presentation of the copy number change caused by pseudogene interference compared to two copies or duplication (three copies) of the entire gene. **(B)** Breakpoints are visible at most intron-exon junctions in cases demonstrating processed pseudogene interference.


A
Copy number
Two copies
Duplication
Interference from processed pseudogene
3
2
B

- Among these, *RPS20* and *CTNNA1* were the genes in which processed pseudogenes were most commonly detected.
- These 69 observations represent at least 31 unique processed pseudogenes.
- 23 of these 31 unique processed pseudogenes were resolved as benign because the native genes were determined to be unaffected.

Figure 2. Localization of the processed pseudogene in *CTNNA1.* **(A)** The breakpoint sequence upstream of exon 1 provides information on the potential location of the processed pseudogene on Chr12; **(B)** Forward and reverse primers (Chr12_F and Chr12_R) flanking the insertion point on Chr12 were designed; **(C)** Long-range PCR using the primer combination followed by sequencing identifies the processed *CTNNA1* pseudogene on Chr12.

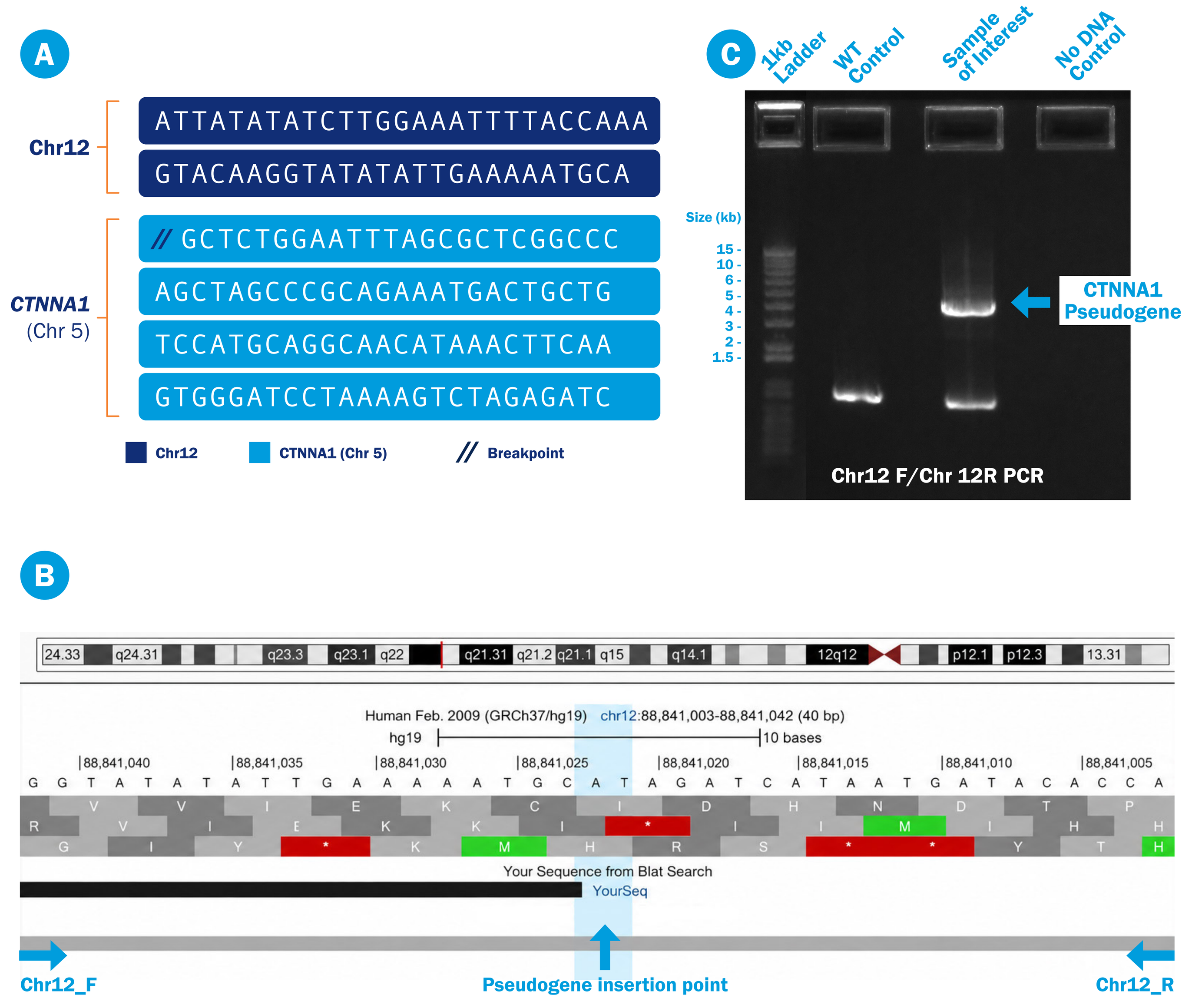

A
Chr12
ATTATATATCTTGGAAATTTTACCAAA
GTACAAGGTATATATTGAAAAATGCA
CTNNA1
(Chr 5)
// GCTCTGGAATTTAGCGCTCGGCCC
AGCTAGCCCGCAGAAATGACTGCTG
TCCATGCAGGCAACATAAACTTCAA
GTGGGATCCTAAAAGTCTAGAGATC
Chr12
CTNNA1 (Chr 5)
Breakpoint
C
1kb Ladder
WT Control
Sample of Interest
No DNA Control
Size (kb)
CTNNA1 Pseudogene
Chr12 F/Chr 12R PCR
B
Human Feb. 2009 (GRCh37/hg19) chr12:88,841,003-88,841,042 (40 bp)
10 bases
Your Sequence from Blat Search
YourSeq
Chr12_F
Pseudogene insertion point
Chr12_R

Table 1. List of unique processed pseudogenes identified in this study.

Gene	Affected regions	Breakpoint location in the native gene	Potential insertion point	Report result'
RPS20	Exon 1	chr8:56,987,064-56,987,065	Undetermined	Benign, native gene unaffected
RPS20	Exon 1	chr8:56,987,064-56,987,065	chr4:276,885-276,886	Benign, native gene unaffected
RPS20	Promoter (3') - exon 1	chr8:56,987,064-56,987,065	chr15:40,792,169-40,792,170	Benign, native gene unaffected
RPS20	Promoter (3') - exon 4	chr8:56,987,065-56,987,066	chr16:90,037,668-90,037,669	Benign, native gene unaffected
RPS20	Exon 1	chr8:56,987,065-56,987,066	chr14:74,286,653-74,286,654	Benign, native gene unaffected
RPS20	Exon 1	chr8:56,987,065-56,987,066	chr9:24,513,248-24,513,249	Benign, native gene unaffected
RPS20	Exon 1	chr8:56,987,065-56,987,066	chr19:20,352,620-20,352,621	Benign, native gene unaffected
RPS20	Promoter (3') - exon 1	chr8:56,987,065-56,987,066	chr5:71,474,487-71,474,488	Benign, native gene unaffected
RPS20	Promoter (3') - exon 1	chr8:56,987,065-56,987,066	chr18:37,393,160-37,393,161	Inconclusive
RPS20	Promoter (3') - exons 1, 4	chr8:56,987,065-56,987,066	chr2:147,480,675-147,480,676	Inconclusive
RPS20	Promoter (3') - exon 1	chr8:56,987,065-56,987,066	chr4:24,512,600-24,512,601	Benign, native gene unaffected
RPS20	Promoter (3') - exons 1, 4	chr8:56,987,066-56,987,067	chr14:93,661,981-93,661,982	Benign, native gene unaffected
RPS20	Promoter (3') - exons 1, 4	chr8:56,987,067-56,987,068	chr19:54,210,158-54,210,159	Benign, native gene unaffected
RPS20	Promoter (3')-2(5')	chr8:56,987,067-56,987,068 & chr8:56,986,938-56,986,939	chr14:65,506,550; chr14:23,105,850	Benign, native gene unaffected
CTNNA1	Exons 1-17	chr5:138,089,135-138,089,136	chr9:75,969,227-75,969,228	Benign, native gene unaffected
CTNNA1	Exons 1-17	chr5:138,089,165-138,089,166	chr12:88,841,022-88,841,023	Benign, native gene unaffected
CTNNA1	Exons 1-17	chr5:138,089,168-138,089,169	chr13:61,266,427-61,266,428	Benign, native gene unaffected
CTNNA1	Exons 1-17	chr5:138,089,169-138,089,170	chrX:30,947,493-30,947,494.	Inconclusive
CTNNA1	Exons 1-17	chr5:138,089,169-138,089,170	chr9:77,731,611-77,731,612	Benign, native gene unaffected
CTNNA1	Exons 5-17	chr5:138,160,215-138,160,216	chrX:143,135,865-143,135,866	Benign, native gene unaffected
CTNNA1	Exons 16-17	chr5:138,268,266-138,268,267	chr6:126,848,498-126,848,499	Benign, native gene unaffected
MAX	Exons 1-5	chr14:65,569,235-65,569,236	chrUn_gl000220:125,431-125,432	Inconclusive
MAX	Exons 1-5	chr14:65,569,196-65,569,197	chr12:112,701,369-112,701,370	Benign, native gene unaffected
CDH1	Exons 12(3') - 16	chr16:68,856,089-68,856,090	chr6:154,718,338-154,718,339	Benign, native gene unaffected
PMS2	Exons 11(3') - 15	chr7:6,026,626-6,026,627	chr3:46,241,849-46,241,850	Benign, native gene unaffected
SDHB	Exons 1-8	chr1:17,380,527-17,380,528	chrX:121,521,923-121,521,924	Benign, native gene unaffected
SDHB	Exons 1-8	chr1:17,380,534-17,380,535	chr2:175,448,961-175,448,962	Benign, native gene unaffected
PTEN	Promoter (3'), exons 5, 7, 9	chr10:89,623,386-89,623,387	chr10:42,400,509-42,400,510	Inconclusive
FH	Promoter (3') - exon 10	chr1:241,683,055-241,683,056	chr16:22,486,440-22,486,441 or chr16:21,475,143-21,475,144	Inconclusive
VHL	Exons 1-3	chr3:10,183,480-10,183,481	chr12:54,848,491-54,848,492	Benign, native gene unaffected
EPCAM	Exon 9	N/A		Inconclusive

Conclusions

- Novel processed pseudogenes can produce false positive copy number or sequencing changes on NGS.
- Thorough investigation is needed to determine the location of the processed pseudogene in order to differentiate the interference from these pseudogenes from true pathogenic variations in the native genes.

Disclosures: All authors were employees of Myriad Genetics, Inc. at the time of this study and received salaries and stock as compensation.

Debora Mancini-DiNardo (dmancini@myriad.com)