Discovery of genetic defects in unexplained colorectal cancer syndromes

Jansen, A.M.L.

Chapter 4

RNA analysis of cancer predisposing genes in formalin-fixed paraffin-embedded tissue determines aberrant splicing

AML Jansen[1,2], HM van der Klift[3], MAE Roos[1], JDH van Eendenburg[1], CMJ Tops[3], JT Wijnen[3], FJ Hes[3], H Morreau[2], T van Wezel[2]

EJHG, in press

Department of [1]Pathology, [2]Human Genetics and [3]Clinical Genetics, Leiden University Medical Center, Leiden, The Netherlands

Abstract

High-throughput sequencing efforts in molecular tumor diagnostics detect increasing numbers of novel variants, including variants predicted to affect splicing. *In silico* prediction tools can reliably predict the effect of variant disrupting canonical splice sites; however, experimental validation is required to confirm aberrant splicing. Here, we present RNA analysis performed for 13 canonical splice site variants predicted- or known to result in splicing in the cancer predisposition genes *MLH1*, *MSH2, MSH6, APC* and *BRCA1.* Total nucleic acid was successfully isolated for 10 variants from eight formalin-fixed paraffin-embedded (FFPE) tumor tissues and two B-cell lines. Aberrant splicing was confirmed in all six variants known to result in splicing. Of one known variant in the B-cell line, aberrant splicing could only be detected after formalin fixation, which indicated that formalin fixation could possibly inhibit RNA degradation. Aberrant splicing was concluded in three of four predicted splice variants of uncertain significance, supporting their pathogenic effect. With this assay, somatic splice variants can be easily and rapidly analysed, enabling retrospective analysis to support the pathogenicity of variants predicted to result in splicing when only FFPE material is available.

Keywords: Colorectal cancer; splice site; splice prediction; FFPE; RNA

Introduction

Most mammalian genes consist of multiple exons interspersed by long intronic sequences.[1] To create mature mRNA, the introns must be correctly identified and 'spliced out', and the exons joined together.[1] The spliceosome, the splicing machinery responsible for this process, recognizes conserved motifs at or near the intron ends and a branch site within the intron.[1] Splice regulatory elements (SRE) near exon-intron boundaries, such as SR (serine-arginine rich) or hnRNP (heterogeneous nuclear ribonucleoparticle) proteins, are indispensable for correct splice site identification.[2] These elements can enhance or repress splicing and play an important role in alternative splicing.[1, 2]

Of the variants that cause disease, 15-60% are proposed to disrupt splicing.[3] Included are variants in the canonical splice site, which directly alter the canonical splice site efficiency, but also intronic and exonic variants that alter the SREs or result in creation of a new splice site or activation of a cryptic splice site.[3, 4] The latter could result in inclusion of a pseudoexon, an intronic sequence wrongly interpreted as an exon. Exon skipping or inclusion of a pseudoexon often results in a shift of the open reading frame, resulting in a premature stop codon or leading to a non- or less-functional protein. The mechanism of nonsense-mediated decay (NMD) in which mRNAs with a premature termination codon are degraded can remove aberrant mRNAs encoding for truncated proteins, ensuring mRNA quality.[5]

Current assessment whether variants result in aberrant RNA transcripts often consists of *in silico* prediction with bioinformatics prediction tools, sometimes followed by reverse transcriptase PCR (RT-PCR) analysis of RNA extracted from blood[6-8] or functional splicing reporter minigene assays.[4, 9] Splicing microarrays can be used for large-scale identification of splicing differences but are not always implemented in current diagnostics.[3] Although high-quality patient RNA analysis is usually preferred, this RNA is not always available, or the analysis is hampered because of degradation of aberrant transcripts through NMD.[1]

With the current high-throughput sequencing methods applied in molecular tumor diagnostics, many variants are found, most of uncertain significance (VUS); hence, functional tests are required to classify these variants. Of these unclassified variants, a percentage is predicted to affect splicing. Specific kits are available to isolate RNA from formalin-fixed paraffin-embedded (FFPE) tissue blocks, and previous studies show that PCR, RT-PCR and even next-generation sequencing (NGS) are possible on these RNA samples.[10, 11] RNA analysis on RNA isolated from FFPE tissue is currently not standardly performed but would enable analysis of somatic splice site variants.

In the current study, the effect of splice site variants was examined in multiple cancer susceptibility genes, *MLH1, MSH2, MSH6, APC* and *BRCA1*. *MLH1, MSH2* and *MSH6* are part of the mismatch repair (MMR) pathway. Pathogenic heterozygous germ line variants in the MMR genes cause Lynch Syndrome, an autosomal dominant predisposition for colorectal, endometrial and other cancers.[12] Other known causes of MMR-deficiency are somatic *MLH1* promoter hypermethylation and the recently described biallelic somatic inactivation of the MMR genes caused by somatic variants.[13-15] Pathogenic germ line variants in the *APC* gene are known to result in familial adenomatous polyposis, a dominant disorder characterized

by the occurrence of hundreds to even thousands of adenomas throughout the colon.[16, 17] In a small percentage of patients, the tumor phenotype can be explained by mosaic *APC* variants.[18-20] These variants can be easily detected by screening multiple adenomas, because the *APC* variant is present with a higher variant allele frequency in the tumor.[18, 21] Pathogenic variants in the *BRCA1* gene, a key player in the nucleotide excision repair pathway, result in a high susceptibility to breast and ovarian cancers.[22, 23] Because BRCA-mutation status affects treatment strategies (PARP-inhibitors); the ability to detect and functionally assess both germ line and somatic mutations in *BRCA1* and *BRCA2* must increase.[24, 25] With the shift towards increased diagnostic screening of tumor tissue for all three syndromes, more somatic variants are found, which require functional tests to assess their pathogenicity.

The aim of our study was to investigate the possibility of analysing RNA isolated from FFPE tissue to assess the effect of germ line and somatic variants predicted to affect splicing. We hypothesized that formalin fixation could inhibit RNA degradation, enabling the detection of aberrant RNA in FFPE tissues.

Materials and methods

Selection of variants

In total, 13 variants of the cancer susceptibility genes *MLH1*, *MSH2, MSH6*, *APC* and *BRCA1* were tested for their effect on splicing (Supplementary Table 1). Of all variants, eight were somatic variants found between 2014 and 2017 with next-generation sequencing in a previous study[26] or through molecular tumor diagnostic NGS, with all having a variant allele frequency of at least 12%. Five were germ line splice site variants, all previously demonstrated to result in aberrant RNA (Supplementary Table 1). The *MLH1* c.454_545del, a germ line genomic exon six deletion, was added as a positive control.

RNA isolation and cDNA synthesis

For eleven variants, total nucleic acid was obtained from tissue cores punched from FFPE blocks embedded between 2009 and 2016. Tumor areas were marked on a hematoxylin and eosin stained slide by a pathologist. Tissue cores from the corresponding area on the FFPE block were punched with a 0.6 mm biopsy needle. Total nucleic acid was isolated from the obtained punches and microdissected areas with a Tissue Preparation System with VERSANT Tissue Preparation Reagents (Siemens Healthcare Diagnostics, Tarrytown, NY, USA).[27] For two variants (*MLH1* c.791-1G>C and *MSH2* c.1511-2A>G), no FFPE tissue was available, but EBV-transformed B-cells were cultured. Additionally, RNA from the *MSH2* c.1511-2A>G B-cell line was isolated after incubating the cells with 4% formalin for five hours. RNA from the B-cell lines and three colorectal cancer cell lines (SW480, SW837 and LS180) was isolated using a Nucleospin RNA isolation kit (Macherey-Nagel-06/2015, Rev.17, Düren, Germany) according to the manufacturer's protocol. Colorectal cancer cell lines were used as a positive control for RNA expression.

All cDNA was synthesized using OligoDT's and random primers with a SuperScript VILO cDNA synthesis Kit (Invitrogen, Carlsbad, CA, USA) according to the manufacturer's protocol.

Splice site prediction/variant nomenclature
For (canonical) splice site prediction, Alamut (Interactive Biosoftware, Rouen, France) was used. This software package includes the *in silico* splice site prediction algorithms SpliceSite Finder (SSF), MaxEntScan (MES) (*http://genes.mit.edu/burgelab/maxent/Xmaxentscan_scoreseq.html*), NNSPLICE (*http://www.fruitfly.org/seq_tools/splice.html*) and Human Splicing Finder (HSF, *http://www.umd.be/HSF/*). Variants were annotated according to the Human Genetics Variation Society (HGVS) guidelines. Recommendations of the Human Genome Variation Society (HGVS) were followed to use the term "variant" instead of "mutation" (*http://www.hgvs.org/mutnomen/recs.html*). The following Genbank reference sequences were used: NM_000249.2 for *MLH1,* NM_000251.2 for *MSH2,* NM_000179.2 for *MSH6,* NM_000038.5 for *APC* and NM_007294.3 for *BRCA1.*

Primer design and PCR
For all variants with an (predicted) exon skip, two primer pairs were created to amplify two exon-exon boundaries. Primers were used in three combinations: Forward1/Reverse1, Foward2/Reverse2 and Forward1/Reverse2. For all variants with a partial exon skip or pseudoexon insertion as (predicted) RNA effect, primers were designed to amplify the exon-exon boundary. All primer sequences are listed in Supplementary Table 2. Real-time PCR was used to amplify the exon-exon boundaries and to assess the expression of the affected gene. All PCR reactions were performed on a CFX96 touch Realtime PCR machine (Bio-rad, Hercules, CA, USA) with the following PCR program: 95 °C for 5 min (1 cycle), 95 °C for 15 s, 60 °C for 30 s, and 72 °C for 30 s (38 cycles), followed by a melt curve from 65 °C to 95 °C with a 0.5 °C increment for 5 s with plate read. When no PCR product was detected, PCR was repeated with high cDNA input and 44 instead of 38 cycles. Because of limited cDNA, only F1/R2 was repeated for variants with two primer pairs, when the first PCR failed. All PCR products were analysed on a Qiaxcel capillary electrophoresis system (Qiagen, Hilden, Germany) and sequenced with Sanger Sequencing.

Results

MMR variants
Five MMR splice variants and one *MLH1* genomic exon deletion in RNA isolated from FFPE tissue were analysed for their effect on splicing (Table 1). Total nucleic acid was isolated from FFPE blocks and converted to cDNA using OligoDT's and random primers. The quality of cDNA was evaluated by detecting the expression of housekeeping genes (HKG) *HNRNPM* and/or *CPSF6.* From five FFPE tissue blocks, HKG expression was detected, and in three of the five, cDNA from the affected MMR gene could be amplified and analysed (*MLH1* c.454_545del, *MLH1* c.2104 G>C and *MSH6* c.3801+1_3801+5del). The amplified products of the three MMR cDNAs from FFPE tissues were measured with Qiaxcel (Figure 1A) and sequenced (Figure 1B). Size determination of the cDNAs carrying the *MLH1* c.454_545del and *MSH6* c.3801+1_3801+5del variants showed only a product size smaller than that of the WT control, whereas *MLH1* c.2104-1G>C only showed a product comparable in size with that of the WT product. Sequencing showed an aberrant product in two of the three FFPE samples, the *MLH1* genomic exon 6 deletion and a skip of exon 8 in the *MSH6* c.3801+1_3801+5del sample (Figure 1B), whereas for *MLH1* c.2104-1G>C, sequencing was normal, as was that for the WT control.

Additionally, RNA isolated from the two EBV-transformed B-cell lines carrying an *MSH2* c.1511-2A>G and an *MLH1* c.791-1G>C splice site variant was tested. Size determination of the cDNA from the two cell lines showed a product of approximately 350 bp (comparable with the WT control) and a product of 250 bp for the *MLH1* c.791-1G>C sample and a product comparable in size with that of the WT for the *MSH2* c.1511-2A>G sample. Sequencing detected an aberrant product only in the *MLH1* c.791-1G>C sample, which showed a skip of exon 10. To mimic FFPE conditions, 4% formalin was added to fixate EBV-transformed cells carrying the *MSH2* variant. After fixation, RNA was isolated, and cDNA was synthesized following the same protocol as that for the non-formalin fixed cells. cDNA from the formalin-fixed cells was tested and showed a size comparable with that of the WT (Figure 1A), and an aberrant product was detected with Sanger Sequencing (Figure 1B).

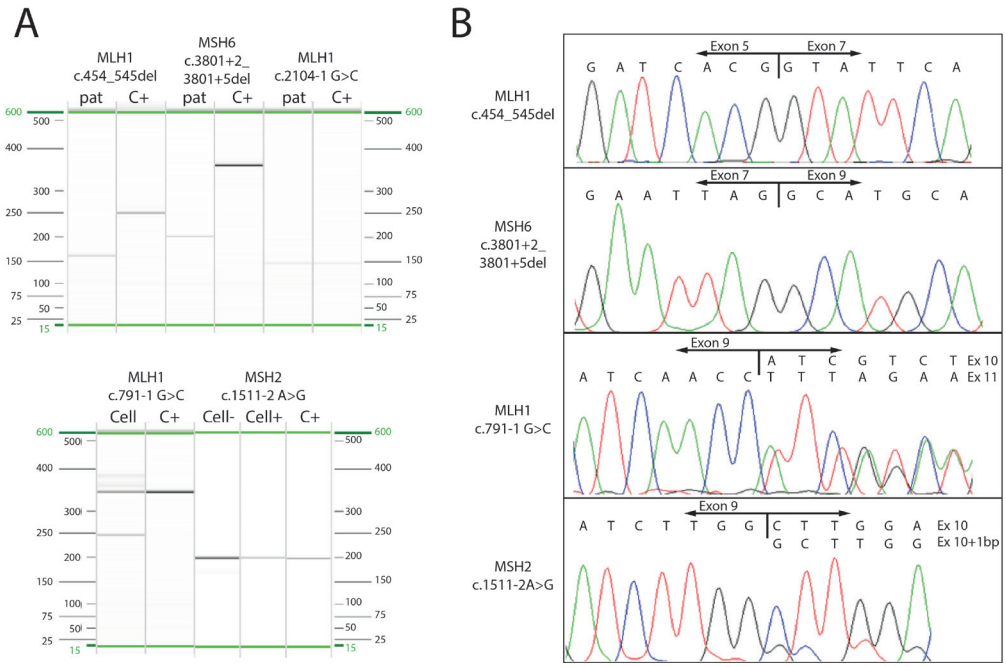

Figure 1A and B: Size determination and sequencing results MMR

Qiaxcel results showing the size of the MMR variants [A] of patient material (pat), cell lines (Cell) and control RNA isolated from colorectal cell lines (C+). The MSH2 cell line was analysed with (Cell+) and without (Cell-) formalin fixation of the cells. [B] Sanger sequencing results of variants showing aberrant products.

APC variants

Three *APC* variants were analysed for their effect on splicing (Table 1). RNA was successfully isolated from all three FFPE blocks, shown by detection of HKG expression. From all three samples, cDNA from *APC* could be amplified (Figure 1C) and sequenced (Figure 1D). Compared with the control without the variant, the difference in size of *APC* c.1548G>A was almost 125 bp. Sequencing showed an aberrant product for all three variants: a skip of the last 5 nucleotides of *APC* exon 7 in the *APC* c.834+2 T>A sample; a skip of the first nucleotide of exon 15 in the *APC* c.1959-1G>A sample; and an exon 11 skip in the *APC* c.1548G>A sample (Figure 1D).

Table 1: Results splicing assay

Gene	Variant	Source RNA	Variant frequency	(predicted) RNA effect	Result
			A. variants known to result in splicing		
MLH1	c.454_545del*	FFPE, CRC45	100%	Skip exon 6	Aberrant splicing as reported‡
MLH1	c.791-1G>C*	Cell line	50%	Skip exon 10	Aberrant splicing as reported
MLH1	c.793C>T*	FFPE, CRC63	NA	Skip exon 10 (major), WT (minor)	Unable to isolate nucleic acid
MLH1	c.1731G>A	FFPE, CRC51	35%	Skip exon 15	No MMR RNA expression detected
MSH2	c.1511-2A>G*	Cell line	80%	Last nt of intron 9 included	Aberrant splicing as reported§
MSH6	c.3801+1_ 3801+5del*	FFPE, Pol67	50%	Skip exon 8	Aberrant splicing as reported‡
APC	c.1548G>A	FFPE, Pol68	28%	Skip exon 11	Aberrant splicing as reported‡
BRCA1	c.213-12A>G	FFPE, EC74	93%	Insertion sequence corresponding to c.213-12_c.213-1	Aberrant splicing as reported‡
			B. variants predicted to result in splicing		
MLH1	c.885-2A>G	FFPE, CRC45	57%	Skip 5nt of exon 11†	No MMR RNA expression detected
MLH1	c.2104-1G>C	FFPE, EC55	36%	Skip 2nt of exon 19†	Only WT product
APC	c.834+2T>A	FFPE, Pol60	12%	Skip 5nt of exon 7†	Aberrant splicing as predicted
APC	c.1959-1G>A	FFPE, Pol67	40%	Skip 1nt of exon 15†	Aberrant splicing as predicted
BRCA1	c.212+3A>T	FFPE, OC46	54%	Skip 22nt of exon 5†	Aberrant splicing as predicted‡

*Source RNA shows cell line or formalin-fixed paraffin embedded (FFPE) tissue block, with type of tumor and age of onset. CRC = colorectal cancer, EC = endometrial cancer, Pol = polyps, WT = Wildtype, nt = nucleotides *Germline variants, variant allele frequency based on Sanger Sequencing of tumor material (estimated peak height. NA; not applicable, variant frequency in FFPE-tissue could not be assessed due to low quality †predicted RNA effect is based on nearest cryptic splice site, which could possibly act as a new 3' or 5' splice site. ‡no WT product detected §aberrant product only detected with formalin fixation of the cells.*

BRCA1 variants

Two *BRCA1* splice site variants were analysed for their effect on splicing (Table 1). RNA was successfully isolated from both FFPE blocks (shown by HKG detection), and *BRCA1* cDNA was amplified and analysed. PCR was performed with the same primers for both variants. Size determination indicated a smaller (*BRCA1* c.212+3A>T) and a slightly larger (*BRCA1* c.213-12A>G) product compared with that of the wild-type control (C+, Figure 1E). Sequencing showed a skip of the last 22 nucleotides of exon 5 (*BRCA1* c.212+3A>T) and an inclusion of the last 11 nucleotides of intron 5 corresponding to the *BRCA1* c.213-11_c.213-1 sequence (*BRCA1* c.213-12A>G, Figure 1F).

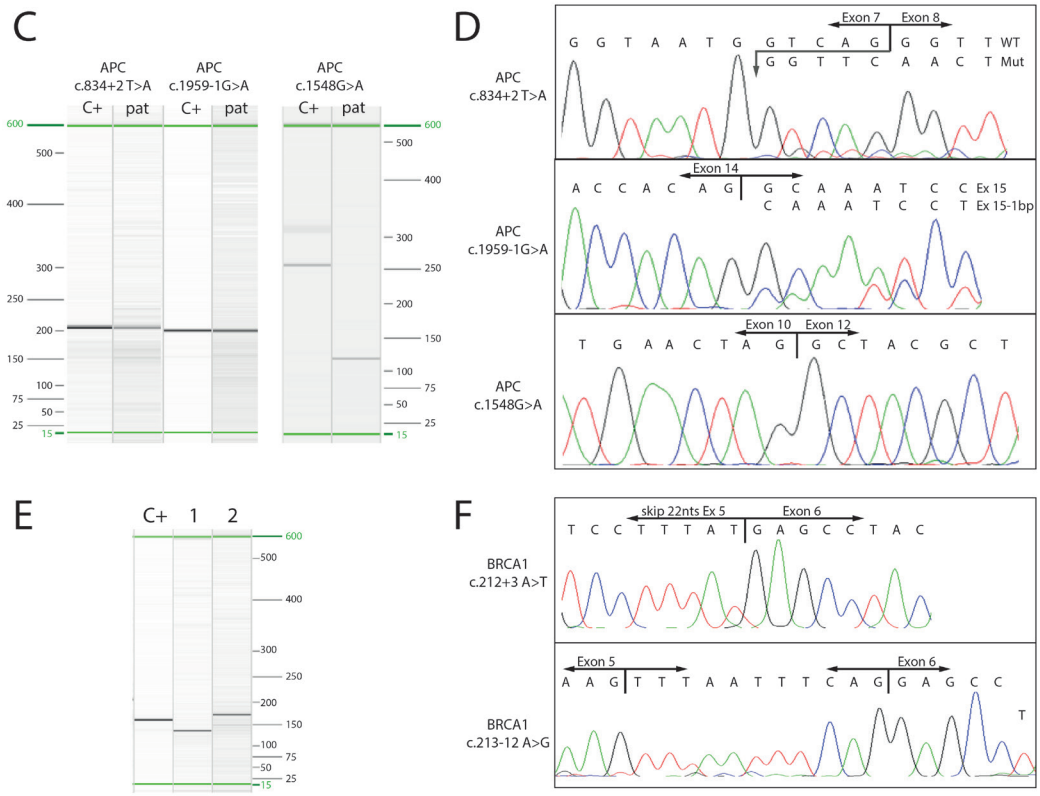

Figure 1C-F: Size determination and sequencing results *APC* and *BRCA1*.

Qiaxcel results showing the size of the APC variants [C] and BRCA1 variants [E] of patient material (pat) and control RNA isolated from colorectal cell lines (C+). The BRCA1 variants were analysed with the same primers, pat1 shows the BRCA1 c.212+3A>T and the pat2 shows the BRCA1 c.213-12A>G. [D] and [F] Sanger sequencing results of variants showing aberrant products.

Discussion

We performed RNA analysis for six splice site variants known to result in aberrant splicing and five variants predicted to result in aberrant splicing using RNA isolated from FFPE tissues. For the six variants shown previously to result in aberrant splicing, the reported splice effect was confirmed for four, and for the other two variants, RNA analysis was not possible because either RNA isolation from FFPE tissue failed (no expression of HKG) or the affected gene (i.e., *MLH1*) did not show expression in the presence of positive HKG expression. In all four confirmed splice effects, no WT product was identified. For two variants (*MLH1* c.454_545del and *BRCA1* c.213-12 A>G), this result was expected because of the high variant allele frequencies (100% and 93%, respectively). The *APC* c.1548G>A and *MSH6* c.3801+1_3801+5del had variant allele frequencies of 28% and 50%, respectively, but only aberrant product was detected, which could be due to preferential amplification of the smaller (aberrant) product or possible FFPE-induced RNA degradation. For the *APC* variant, the F2/R2 primers that amplified the boundary of exon 11-exon 12 (with a part of exon 11) produced a product, which showed that exon 11 was part of the cDNA.

For two variants, *MLH1* c.791-1G>C and *MSH2* c.1511-2A>G, RNA was isolated from EBV-transformed cell lines because FFPE tissue was not available. Both previously resulted in an (partial) exon skip,[4, 28-30] which changed the reading frame and led to a premature stop codon; however, the aberrant splicing was only confirmed for one of the two variants (*MLH1* c.791-1G>C), whereas only WT transcript was detected for the *MSH2* c.1511-2A>G. The EBV-transformed cells were cultured without NMD inhibitors, and we hypothesized that the aberrant MSH2 RNA was possibly degraded through nonsense-mediated decay (NMD). NMD of the *MSH2* c.1511-2A>G RNA and no NMD of the *MLH1* c.791-1G>C RNA is consistent with previous studies in which NMD-inhibitors were omitted.[31, 32] Notably, the aberrant RNA was detected after formalin fixation of the EBV cells carrying the *MSH2* c.1511-2A>G, which is consistent with our hypothesis that aberrant RNA in FFPE tissue can be detected because formalin fixation prevents the degradation of RNA.

For the five variants predicted to affect the canonical splice site, three resulted in aberrant splicing, showing the predicted splice effect. For the other two, in one (*MLH1* c.2104-1G>C), only the WT transcript was detected, and in one, no expression of the affected gene (*MLH1*) was detected in the presence of normal HKG expression. The *APC* c.834+2T>A and *APC* c.1959-1G>A showed WT and aberrant product, which can be explained by the variant allele frequencies (12% and 40%, respectively). The *BRCA1* c.212+3A>T had a variant allele frequency of 54%, although only aberrant product was detected, which was possibly due to preferential amplification of the smaller (aberrant) product or because no RNA expression of the WT allele occurred. However, FFPE-related RNA degradation could not be excluded. With the detection of aberrant RNA in three of four variants predicted to affect splicing, this assay confirmed that *in silico* splice prediction tools are reliable in their predictions, particularly for variants disrupting the canonical splice site, although experimental analysis is required.[4, 7, 33] The three predicted splice variants shown to result in aberrant splicing in this study are currently classified as variants of uncertain significance, but the RNA analysis in this study supports their pathogenic effect.

FFPE blocks were collected from the archives and were embedded between 2000 and 2016, with most (n=9) embedded after 2009. Notably, from the two samples embedded before 2009, no RNA was successfully isolated from the FFPE block, indicating that isolating RNA from FFPE tissue might not be possible in older blocks. From eight of nine blocks embedded after 2009, RNA was successfully isolated, showing WT and/or aberrant RNA, and combined with the formalin fixation results from the cell lines, showed that formalin fixation possibly inhibited RNA degradation. With analysis of RNA from FFPE tissues, the splicing effect of somatic variants that are only present in the tumor can be analysed. Furthermore, available material can be retrospectively analysed, without having to request RNA from leukocyte DNA, which is not always available.

However, a few limitations of this study deserve discussion. First, RNA from FFPE tissue blocks is often degraded to fragments less than ~300 bases in length.[34] To analyse this RNA, small amplicons must be designed instead of large amplicons containing multiple exons, as is preferred in leukocyte RNA testing. The design of the primers is very specific for the variant and is based on splicing prediction software. Although these algorithms are described as accurate for variants in the canonical splice site,[4, 7, 8] aberrant products that are not predicted and fall outside the amplification window of the assay will not be detected. Second, when an aberrant product is not detected, poor RNA quality, a wrongly predicted effect or no splice effect of the variant can be implicated, but no expression or RNA degradation of the mutant product by NMD can also occur. Therefore, the assay can confirm aberrant splicing, but from negative assay results, one cannot accurately conclude that no aberrant splicing occurs or that lack of aberrant product is due to other factors. Negative results should be confirmed with other methods, such as a minigene splicing assay.[4, 9] Further evidence of FFPE-based RNA analysis should be obtained from a larger study with more samples and variants. Targeted RNA-seq of RNA isolated from FFPE tissue might enable high-throughput analysis of somatic splice site variants.

RNA analysis on total nucleic acid isolated from FFPE tissue blocks is a valuable tool for the fast and easy detection of aberrant splicing, offering additional support for the pathogenicity of a (predicted) splice variant. With this assay, we correctly showed the splice effect of six known splice site variants and showed the splice effect of three variants predicted to affect splicing. This assay can be used to analyse somatic variants found in FFPE tumor tissue, with formalin fixation possibly inhibiting RNA degradation, and can be easily implemented in current molecular tumor diagnostics to help classify the high number of variants of uncertain significance currently found with high-throughput sequencing.

Supplementary information accompanies this paper on European Journal of Human Genetics' website (*http://www.nature.com/ejhg*).

Acknowledgements
The authors would like to acknowledge Marina Ventayol Garcia for technical assistance.

References

1. Cartegni L, Chew SL, Krainer AR. Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet 2002;3:285-98.
2. Brillen AL, Schoneweis K, Walotka L, et al. Succession of splicing regulatory elements determines cryptic 5'ss functionality. Nucleic Acids Res 2016.
3. Wang GS, Cooper TA. Splicing in disease: disruption of the splicing code and the decoding machinery. Nat Rev Genet 2007;8:749-61.
4. van der Klift HM, Jansen AM, van der Steenstraten N, et al. Splicing analysis for exonic and intronic mismatch repair gene variants associated with Lynch syndrome confirms high concordance between minigene assays and patient RNA analyses. Mol Genet Genomic Med 2015;3:327-45.
5. Hentze MW, Kulozik AE. A perfect message: RNA surveillance and nonsense-mediated decay. Cell 1999;96:307-10.
6. Whiley PJ, de la Hoya M, Thomassen M, et al. Comparison of mRNA splicing assay protocols across multiple laboratories: recommendations for best practice in standardized clinical testing. Clin Chem 2014;60:341-52.
7. Vreeswijk MP, Kraan JN, van der Klift HM, et al. Intronic variants in BRCA1 and BRCA2 that affect RNA splicing can be reliably selected by splice-site prediction programs. Hum Mutat 2009;30:107-14.
8. Spurdle AB, Couch FJ, Hogervorst FB, et al. Prediction and assessment of splicing alterations: implications for clinical testing. Hum Mutat 2008;29:1304-13.
9. Cooper TA. Use of minigene systems to dissect alternative splicing elements. Methods 2005;37:331-40.
10. Hedegaard J, Thorsen K, Lund MK, et al. Next-generation sequencing of RNA and DNA isolated from paired fresh-frozen and formalin-fixed paraffin-embedded samples of human cancer and normal tissue. PLoS One 2014;9:e98187.
11. Gouveia GR, Ferreira SC, Ferreira JE, et al. Comparison of two methods of RNA extraction from formalin-fixed paraffin-embedded tissue specimens. Biomed Res Int 2014;2014:151724.
12. Lynch HT, de la Chapelle A. Hereditary colorectal cancer. N Engl J Med 2003;348:919-32.
13. Mensenkamp AR, Vogelaar IP, van Zelst-Stams WA, et al. Somatic mutations in MLH1 and MSH2 are a frequent cause of mismatch-repair deficiency in Lynch syndrome-like tumors. Gastroenterology 2014;146:643-646 e8.
14. Geurts-Giele WR, Leenen CH, Dubbink HJ, et al. Somatic aberrations of mismatch repair genes as a cause of microsatellite-unstable cancers. J Pathol 2014;234:548-59.
15. Kuismanen SA, Holmberg MT, Salovaara R, et al. Genetic and Epigenetic Modification of MLH1 Accounts for a Major Share of Microsatellite-Unstable Colorectal Cancers. The American Journal of Pathology 2000;156:1773-1779.
16. Powell SM, Petersen GM, Krush AJ, et al. Molecular diagnosis of familial adenomatous polyposis. N Engl J Med 1993;329:1982-7.
17. Groden J, Thliveris A, Samowitz W, et al. Identification and characterization of the familial adenomatous polyposis coli gene. Cell 1991;66:589-600.
18. Spier I, Drichel D, Kerick M, et al. Low-level APC mutational mosaicism is the underlying cause in a substantial fraction of unexplained colorectal adenomatous polyposis cases. J Med Genet 2016;53:172-9.
19. Yamaguchi K, Komura M, Yamaguchi R, et al. Detection of APC mosaicism by next-generation sequencing in an FAP patient. J Hum Genet 2015;60:227-31.
20. Out AA, van Minderhout IJ, van der Stoep N, et al. High-resolution melting (HRM) re-analysis of a polyposis patients cohort reveals previously undetected heterozygous and mosaic APC gene mutations. Fam Cancer 2015;14:247-57.
21. Jansen AM, Crobach S, Geurts-Giele WR, et al. Distinct Patterns of Somatic Mosaicism in the APC Gene in Neoplasms From Patients With Unexplained Adenomatous Polyposis. Gastroenterology 2017;152:546-549 e3.
22. Hall JM, Lee MK, Newman B, et al. Linkage of early-onset familial breast cancer to chromosome 17q21. Science 1990;250:1684-9.
23. Rahman N, Stratton MR. The genetics of breast cancer susceptibility. Annu Rev Genet 1998;32:95-121.
24. Mafficini A, Simbolo M, Parisi A, et al. BRCA somatic and germline mutation detection in paraffin embedded ovarian cancers by next-generation sequencing. Oncotarget 2016;7:1076-83.
25. Koczkowska M, Zuk M, Gorczynski A, et al. Detection of somatic BRCA1/2 mutations in ovarian cancer - next-generation sequencing analysis of 100 cases. Cancer Med 2016;5:1640-6.
26. Jansen AM, van Wezel T, van den Akker BE, et al. Combined mismatch repair and POLE/POLD1 defects explain unresolved suspected Lynch syndrome cancers. Eur J Hum Genet 2016;24:1089-92.
27. van Eijk R, Stevens L, Morreau H, et al. Assessment of a fully automated high-throughput DNA extraction method from formalin-fixed, paraffin-embedded tissue for KRAS, and BRAF somatic mutation analysis. Exp

Mol Pathol 2013;94:121-5.
28. Wijnen J, Khan PM, Vasen H, et al. Majority of hMLH1 mutations responsible for hereditary nonpolyposis colorectal cancer cluster at the exonic region 15-16. American Journal of Human Genetics 1996;58:300-307.
29. Wijnen J, Khan PM, Vasen H, et al. Hereditary nonpolyposis colorectal cancer families not complying with the Amsterdam criteria show extremely low frequency of mismatch-repair-gene mutations. Am J Hum Genet 1997;61:329-35.
30. Casey G, Lindor NM, Papadopoulos N, et al. Conversion analysis for mutation detection in MLH1 and MSH2 in patients with colorectal cancer. Jama 2005;293:799-809.
31. Bianchi F, Rosati S, Belvederesi L, et al. MSH2 splice site mutation and endometrial cancer. Int J Gynecol Cancer 2006;16:1419-23.
32. Thompson BA, Goldgar DE, Paterson C, et al. A multifactorial likelihood model for MMR gene variant classification incorporating probabilities based on sequence bioinformatics and tumor characteristics: a report from the Colon Cancer Family Registry. Hum Mutat 2013;34:200-9.
33. Colombo M, De Vecchi G, Caleca L, et al. Comparative in vitro and in silico analyses of variants in splicing regions of BRCA1 and BRCA2 genes and characterization of novel pathogenic mutations. PLoS One 2013;8:e57173.
34. Cronin M, Pho M, Dutta D, et al. Measurement of gene expression in archival paraffin-embedded tissues: development and performance of a 92-gene reverse transcriptase-polymerase chain reaction assay. Am J Pathol 2004;164:35-42.

Supplementary Table 2: Primers

Variant	Forward primers		Reverse Primers	
MLH1 c.454_545del	F1 F2	GATGGAAAACTGAAAGCCCCT TTTTACAACATAGCCACGAG	R1 R2	GCCAACAACTTCCAAAATTTTCC GGTTGAGGCATTGGGTAGTG
MLH1 c.791-1G>C/ c.793C>T	F1 F2	TGGATGTGAGGATAAAACCC AGAAACAGTGTATGCAGCCT	R1 R2	AGGTACAGGAATGGGTGTGT CTGGAGGAATTGGAGCCCA
MLH1 c.885-2A>G	F1	AGAAACAGTGTATGCAGCCT	R1	CTGGAGGAATTGGAGCCCA
MLH1 c.1731G>A	F1	ATACCTTCTCAACACCACCA	R1	CATCTCAGCCTTCTTCTTCA
MLH1 c.2104-1G>C	F1	AAAGCCTCAGTAAAGAATGC	R1	TGTGTTCCACAGTCCACTTC
MSH2 c.1511-2A>G	F1	CCTTGTAAAACCTTCATTTGATCC	R1	CGAAGGACTTTTTCTTCCTTACA
MSH6 c.3801+1_ 3801+5del	F1 F2	GTGAAACTGCCAGCATACTCA CATTATTTTCAACTCACTACCAT	R1 R2	CATATGTCCTAGGCGCACAG TGGGAGATTAGCAAGCCTTG
APC c.834+2T>A	F1	GGTCATCTCAGAACAAGCAT	R1	CAGATGACTTGTCAGCCTTC
APC c.1548G>A	F1 F2	TCCTGCTGTGTGTGTTCTAA TTATTGCAAGTGGACTGTGA	R1 R2	CAAGTTTGTCAAAGCCATTC TAGTTGGGCCACAAGTGC
APC c.1959-1G>A	F1	AGCCAGACAAACACTTTAGC	R1	GATTCCACAAAGTTCCACAT
BRCA1 c.212+3A>T c.213-12A>G	F1	ATGCTGAAACTTCTCAACCA	R1	AACCTGTGTCAAGCTGAAAA

Supplementary Table 1: Splice variants tested

Gene	Variant	Protein	RNA effect	(Predicted) Effect on ss and/or ESE motifs	% decrease of can. ss strength	RNA effect previously described in:
			A. variants known to result in splicing			
MLH1	c.454_545del	p.(E153Ffs*8)	Skip exon 6	NA	NA	NA
MLH1	c.791-1G>C	p.(H264Lfs*2)	Skip exon 10	Loss of canonical 3' ss	(100, 100, 100, 100)	Wijnen et al, Am. J. Hum. Genet. 1996 & van der Klift et al, Mol Genet Genomic Med, 2015
MLH1	c.793C>T	p.(R265C)	Skip exon 10 (major), WT (minor)	Increase of 1 ESE	NA	van der Klift et al, Mol Genet Genomic Med, 2015
MLH1	c.1731G>A	p.(S556Rfs*14)	Skip exon 15	Decrease in canonical 5' ss, Loss of 1 ESE	(14, 23, 10, 11)	Tournier et al, Hum Mutat, 2008 & Auclair et al, Hum Mutat, 2006
MSH2	c.1511-2A>G	p.(D506Gfs*7)	Last nt of intron 9 included	Loss of canonical 3' ss	(100, 100, 100, 100)	Wijnen et al, Am. J. Hum. Genet. 1997 & Casey et al, JAMA, 2005
MSH6	c.3801+1_3801+5del	p.(R1217Mfs*6)	Skip exon 8	Loss of canonical 5' ss	(100, 100, 100, 24)	LOVD database, ID: MSH6_00908
APC	c.1548G>A	p.(G471Yfs*19)	Skip exon 11	Decrease in canonical 5' ss, loss SR940 site	(14,30, 20, 12)	Kaufmann et al, J Mol Diagn, 2009
BRCA1	c.213-12A>G	p.(?)	Insertion seq. corr. to c.213-12_c.213-1	Loss of canonical 3' ss	(100, 75, 100, 0)	Hoffman et al, Am J Med Genet, 1998
			B. variants predicted to result in splicing			
MLH1	c.885-2A>G	p.(?)	Skip 5nt of exon 11*	Loss of canonical 3' ss	(100, 100, 100, 100)	-
MLH1	c.2104-1G>C	p.(?)	Skip 2nt of exon 19*	Loss of canonical 3' ss	(100, 100, 100, 100)	-
APC	c.834+2T>A	p.(?)	Skip 5nt of exon 7*	Loss of canonical 5' ss	(100, 100, 100, 100)	-
APC	c.1959-1G>A	p.(?)	Skip 1nt of exon 15*	Loss of canonical 3' ss	(100, 100, 100, 100)	-
BRCA1	c.212+3A>T	p.(?)	Skip 22nt of exon 5*	Decrease in canonical 5' ss	(100, 41, 100, 18)	

*ss = splice site, Nt = nucleotides, NA = not applicable *predicted RNA effect is based on nearest cryptic splice site, which could possibly act as a new 3' or 5' splice site. Splice prediction programs SFF (Splice Site Finder), MES (MaxEntScan), NNS (NNSPLICE) and HSF (Human Splicing Finder) were used for splice effect prediction.*