OPEN ACCESS

Original research

Transcript capture and ultradeep long-read RNA sequencing (CAPLRseq) to diagnose HNPCC/Lynch syndrome

Vincent Schwenk,[1] Rafaela Magalhaes Leal Silva ,[1] Florentine Scharf ,[1] Katharina Knaust,[1] Martin Wendlandt,[1] Tanja Häusser,[1] Julia M A Pickl,[1,2] Verena Steinke-Lange,[1] Andreas Laner,[1] Monika Morak,[1,2] Elke Holinski-Feder,[1,3] Dieter A Wolf [1,4]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/jmg-2022-108931).

[1]Medizinisch Genetisches Zentrum (MGZ), Munich, Germany
[2]Klinikum der Universität München, Munich, Germany
[3]Medizinische Klinik und Poliklinik IV, Campus Innenstadt, Klinikum der Universität München, Munich, Germany
[4]Department of Medicine II, Technical University Munich, Munich, Germany

Correspondence to
Dr Dieter A Wolf, MGZ – Medizinisch Genetisches Zentrum, Munich, Germany; dieter.wolf@mgz-muenchen.de
Professor Elke Holinski-Feder; elke.holinski-feder@mgz-muenchen.de

VS, RMLS and FS are joint first authors.
EH-F and DAW are joint senior authors.

Received 8 September 2022
Accepted 10 December 2022
Published Online First 2 January 2023

To cite: Schwenk V, Leal Silva RM, Scharf F, *et al*. *J Med Genet* 2023;**60**:747–759.

ABSTRACT

Purpose Whereas most human genes encode multiple mRNA isoforms with distinct function, clinical workflows for assessing this heterogeneity are not readily available. This is a substantial shortcoming, considering that up to 25% of disease-causing gene variants are suspected of disrupting mRNA splicing or mRNA abundance. Long-read sequencing can readily portray mRNA isoform diversity, but its sensitivity is relatively low due to insufficient transcriptome penetration.

Methods We developed and applied capture-based target enrichment from patient RNA samples combined with Oxford Nanopore long-read sequencing for the analysis of 123 hereditary cancer transcripts (capture and ultradeep long-read RNA sequencing (CAPLRseq)).

Results Validating CAPLRseq, we confirmed 17 cases of hereditary non-polyposis colorectal cancer/Lynch syndrome based on the demonstration of splicing defects and loss of allele expression of mismatch repair genes *MLH1*, *PMS2*, *MSH2* and *MSH6*. Using CAPLRseq, we reclassified two variants of uncertain significance in *MSH6* and *PMS2* as either likely pathogenic or benign.

Conclusion Our data show that CAPLRseq is an automatable and adaptable workflow for effective transcriptome-based identification of disease variants in a clinical diagnostic setting.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ A large fraction of disease-causing variants are known to disrupt mRNA structure or expression.
⇒ Long-read RNA sequencing is a powerful tool to assess mRNA structure, but its sensitivity is limited.

WHAT THIS STUDY ADDS

⇒ We developed capture and ultradeep long-read RNA sequencing (CAPLRseq) as an automatable and adaptable workflow for effective transcriptome-based identification of disease variants in a clinical diagnostic setting.
⇒ CAPLRseq can evaluate a wide range of simple and complex DNA variants that affect mRNA structure and expression.
⇒ We validated CAPLRseq for the diagnosis of Lynch syndrome.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ CAPLRseq may be incorporated into the diagnostic workflow to unambiguously classify DNA variants of uncertain significance in hereditary cancer predisposition genes

INTRODUCTION

DNA genetic testing for germline variants in tumour suppressor genes can identify individuals with hereditary cancer predisposition. However, up to 44% of genetic variants identified by DNA sequencing are classified as variants of uncertain significance (VUS).[1] VUS present significant clinical challenges as their uncertain impact can impede optimal preventive and therapeutic care of patients and their family members. Fifteen to twenty-five percent of VUS are predicted to disrupt mRNA splicing,[2] although the functional impact of individual variants is rarely confirmed. This is particularly worrisome as hereditary cancer genes are highly susceptible to splicing defects.[3–5] For example, studying genes encoding mismatch repair (MMR) proteins MLH1, PMS2, MSH2 and MSH6 involved in hereditary non-polyposis colorectal cancer (HNPCC)/Lynch syndrome, a form of hereditary colorectal cancer predisposition, we previously reported that 16% of missense variants and 12% of VUS actually cause splicing defects.[6] These considerations highlight the potential of RNA-based analysis in aiding the classification of VUS. Indeed, a recent study demonstrated successful interpretation of VUS identified at the DNA level by subsequent RNA-based analysis in 88% of cases.[7]

Long-read sequencing is ideally suited for detecting mRNA isoforms that might arise from DNA variants.[8–11] For example, using PCR amplification of cDNA, a study identified 32 alternatively spliced isoforms of *BRCA1* mRNA by Oxford Nanopore Technology (ONT) sequencing, 20 of which were novel.[12] Using a similar approach, a study identified single-nucleotide variants (SNVs) and an aberrantly spliced form of *NF1* mRNA.[13] However, the requirement for PCR amplification limits this approach to distinct loci and is thus unsuitable for highly parallel panel analyses and automation. Although whole transcriptome sequencing on the ONT platform

is possible, sequencing depth is often limiting in confidently assigning new isoforms to mRNAs with low-level expression (see the Results section). For example, many variants causing premature termination codons (PTCs) trigger nonsense-mediated mRNA decay (NMD), causing severe depletion of PTC carrying mRNA species of potential diagnostic value.[14 15] Likewise, the relatively high error rate of ONT sequencing necessitates ultradeep coverage for variant calling and mRNA isoform profiling.

Recent studies established the feasibility of RNA capture by hybridisation to enrich parts of the transcriptome for deep long-read sequencing.[11 16 17] While being a powerful tool for the de novo discovery of coding and non-coding transcripts, this methodology requires custom capture probe design and validation. To perform RNA-based analysis in a routine diagnostic setting, we sought to develop an approach that employs validated probe sets and is amenable to automation. Here, we describe a facile protocol merging Agilent's SureSelectXT Low Input Target Enrichment System with Oxford Nanopore's cDNA-PCR Barcoding Library Preparation Kit for highly efficient capture of transcripts from hereditary cancer predisposition genes. Studying samples from patients with suspected HNPCC/Lynch syndrome, we demonstrate that the technique readily enables the interpretation of variants affecting mRNA structure or expression, including (1) PTCs or promoter methylation causing allelic reduction in mRNA expression, (2) alterations in mRNA splicing resulting in exon skipping or intron retention, and (3) structural variants such as SINE-VNTR-Alu (SVA) insertions and fusion transcripts.

RESULTS

Performance of RNA capture sequencing (capture and ultradeep long-read RNA sequencing (CAPLRseq))

The CAPLRseq method established here is summarised in figure 1. It involves extraction of total cellular RNA from either whole human blood or short-term cultures of peripheral blood mononuclear cells (PBMCs). The RNA is then reversely transcribed by oligo-dT priming and template switching according to the ONT cDNA-PCR sequencing Kit (SQK-PCS109), followed by PCR amplification with or without barcoding. The amplified cDNA is employed as input for the Agilent SureSelect^{XT} capture workflow according to the Low Input Target Enrichment System. Captured cDNAs are PCR-amplified and subjected to ONT Rapid Adapter Ligation and sequencing on a GridION instrument.

Table 1 compares run statistics of cDNA sequencing (cDNA-seq) performed with total RNA or after enrichment of 123 hereditary cancer-related transcripts with the CAPLRseq method developed here. In single runs on the GridION platform using R.9.4.1 flow cells, we obtained 4.8–9.8 million reads, with multiplexed samples typically being at the lower end of this range. A minimum of 80% and as many as 94% of reads were typically aligned to the reference genome. Mean read length varied between 625 and 1111 nucleotides, whereas medians ranged from 625 to 917 nucleotides. We did not observe any systematic differences in read length between total cDNA-seq versus CAPLRseq samples.

Whereas we obtained reasonable sequencing depth of 82× across the exome by total RNA sequencing (RNA-seq), average read depth was relatively shallow for 123 cancer-related transcripts (48; table 1, top row) which appear to be expressed at a comparatively low level. This sequencing depth was deemed insufficient for reliable use in disease diagnostics. The CAPLRseq approach raised average sequencing depth of the cancer gene transcripts in duplexed samples to over 5000×, a >100-fold improvement over total RNA-seq (table 1). High depth was also obtained by CAPLRseq of RNA derived from whole blood (~2000×). This depth was not increased in a substantial way by depleting >80% of haemoglobin (HB)-encoding mRNAs *HBA1*/*HBA2* and *HBB* using the GlobinLock strategy,[18] even though globin mRNAs were sequenced to a depth of ~75 000× in non-depleted samples (table 1). This indicates that, at least in a duplex format, flow cell capacity

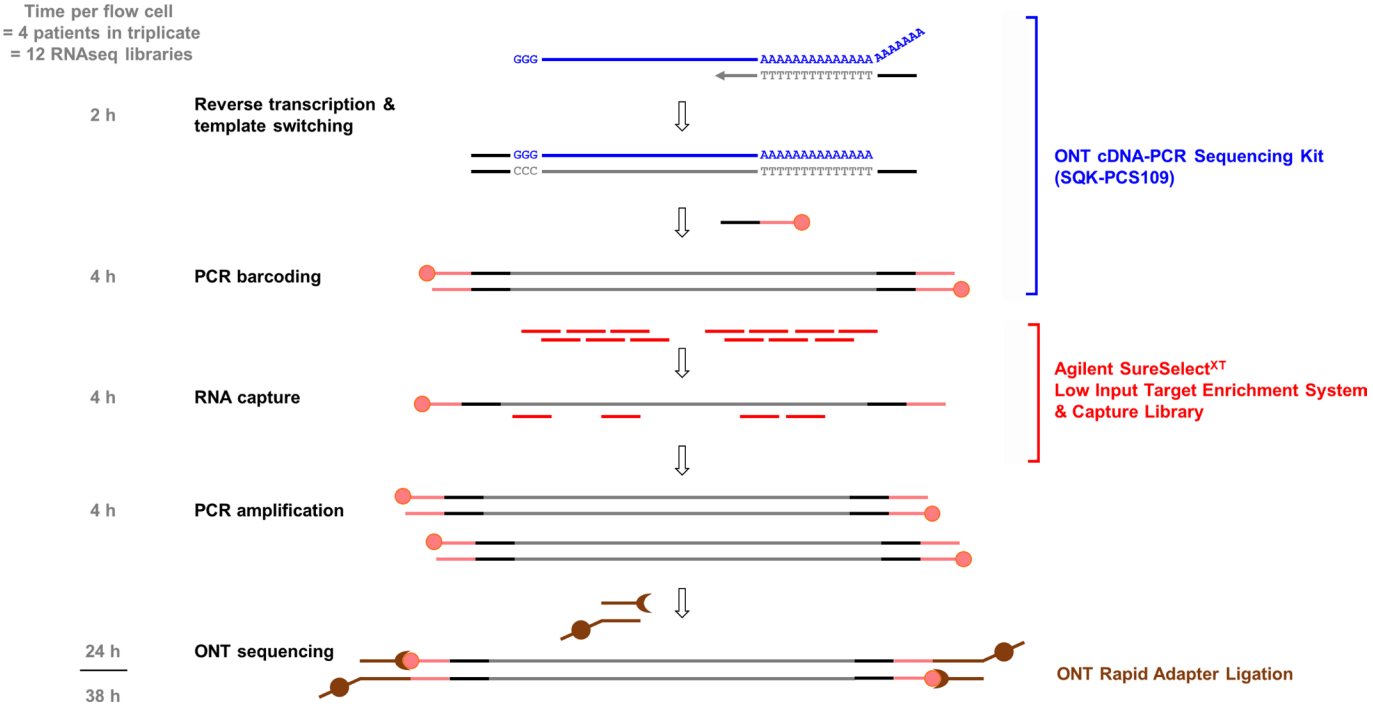


Figure 1 Outline of capture and ultradeep long-read RNA sequencing. The different experimental steps and approximate times are shown for a representative mRNA (see text for details). ONT, Oxford Nanopore Technology.

Table 1 Capture and ultradeep long-read RNA sequencing run statistics

			Reads		Read lengths (bp)		Sequencing depth (x)		
Approach	RNA source	RIN	Number (Million)	Aligned (%)	Mean	Median	Cancer genes*	Exome	Globin
Total cDNA-seq	Human PBMCs	9.8	9.03	92.92	918	732	48	82	43
Total cDNA-seq (2-plex)	Human PBMCs	10.0	3.52	92.99	1077	872	39	50	2
		9.9	4.14	88.89	998	809	30	47	1
		Total per flow cell	7.66		Average per flow cell	841	35	48	2
Capture-seq (2-plex)	Human PBMCs	10.0	4.21	86.15	943	797	4815	58	0
		9.9	5.65	82.85	847	740	5151	61	1
		Total per flow cell	9.86		Average per flow cell	769	4983	59	0
Capture-seq (2-plex)	Whole Blood	10.0	4.54	87.76	756	702	2516	34	14 897†
		10.0	3.22	94.63	826	750	1750	33	74 215‡
		Total per flow cell	7.76		Average per flow cell	726	2133	34	44 556
Capture-seq (6-plex)	Human PBMC	10.0	0.75	92.4	963	819	652	11	0
		10.0	0.69	89.75	953	788	460	9	0
		10.0	0.72	91.01	1015	852	837	11	0
		9.9	1.38	83.73	711	639	548	11	0
		9.9	0.41	92.65	1111	917	409	7	0
		9.9	0.86	85.51	732	659	447	8	0
		Total per flow cell	4.81		Average per flow cell	779	559	9	0
Capture-seq (12-plex)	Human PBMC	10.0	0.35	89.35	903	755	214	4	0
		10.0	0.39	91.8	894	760	252	4	0
		10.0	0.65	81.51	680	625	196	4	1
		10.0	0.44	90.31	876	746	234	5	0
		10.0	0.45	87.91	847	734	272	4	0
		10.0	0.33	92.89	988	833	273	4	0
		10.0	0.61	82.01	695	630	208	4	1
		10.0	0.53	82.85	758	684	256	4	1

Continued

Table 1 Continued

Approach	RNA source	RIN	Reads		Read lengths (bp)		Sequencing depth (x)		
			Number (Million)	Aligned (%)	Mean	Median	Cancer genes*	Exome	Globin
		10.0	0.44	81.73	731	674	187	3	1
		10.0	0.55	86.32	749	684	223	4	1
		10.0	0.62	85.12	742	673	225	5	2
		10.0	0.44	90.09	854	739	244	4	1
		Total per flow cell	5.80		Average per flow cell	711	232	4	1

*A set of 123 cancer genes listed in Supplementary table 2.
†With globin mRNA depletion.
‡Without globin mRNA depletion.
cDNA-seq, cDNA sequencing; PBMC, peripheral blood mononuclear cell.

is not limiting for CAPLRseq of cancer-related mRNAs from whole blood RNA samples even without depletion of globin mRNA.

Although average sequencing depth of the 123 hereditary cancer gene transcripts in the 12-plex format was 187×–273× (table 1), relative quantification based on transcripts per million reads (TPM) revealed that the abundance of individual mRNAs of the cancer panel varied in a representative sample across four orders of magnitude (online supplemental figure S1A). Thus, 75% of the total reads mapped to the 15 most abundant mRNAs, while the remaining 108 transcripts were represented in 25% of all reads (online supplemental figure S1B). The median TPM value of the 123 cancer gene transcripts was 1590. Thirteen transcripts on our cancer panel (*ALK*, *CASR*, *CDKN2B*, *CFTR*, *CTRC*, *GREM1*, *HOXB13*, *KIT*, *MITF*, *PDGFRA*, *PHOX2B*, *PTCH2*, *SPRED1* and *WT1*) could not be reliably detected in the 12-plex format due to low abundance in PBMCs. The wide range of expression levels illustrates the importance of enriching transcripts in order to assess low abundance cancer gene transcripts.

Despite these limitations, quantification of mRNA expression by CAPLRseq was remarkably reproducible with correlation coefficients between technical replicates of the same RNA sample typically above 0.9 (online supplemental figure S1C). Correlations between samples from distinct patients was also high, varying between 0.62 and >0.9. Poor correlation ($r<0.5$) was obtained only for RNA samples of low integrity (online supplemental figure S1C).

Figure 1 also summarises the timeline of CAPLRseq. Times refer to the duration of the RNA-seq procedure for one Gridion flow cell. Considering the time needed for tissue culture and RNA isolation, the entire duration per sample from blood draw to result is ca. 1 week. Since 12 libraries fit on one flow cell (12-plex), four patients per flow cell can be analysed in triplicate. On the fully loaded Gridion instrument with five flow cells, the maximum number of patients that can be analysed in parallel is 20. If required, the throughput of 20 patients per week can be increased by stacking the entire process, such that several 24 hours runs on the Gridion instrument are scheduled per week.

Validation of CAPLRseq for the diagnosis of patients with suspected HNPCC/Lynch syndrome

Using a panel of 25 samples, we had previously demonstrated that PCR-based analysis of mRNAs encoding each of the four MMR proteins involved in HNPCC/Lynch syndrome considerably increases diagnostic yield.[6] Although relative expression values (TPM) of the four mRNAs differed more than 10-fold (*PMS2*=34 312, *MSH6*=17 458, *MLH1*=7215 and *MSH2*=3031), all were expressed well beyond the median TPM value of the 123-cancer gene panel (1590) and thus readily analysable.

We initially used eight of these PCR verified samples harbouring a wide array of disease variants to test whether our newly developed CAPLRseq method would recapitulate the PCR-based variant classification. As shown in table 2, all eight variants were confirmed by CAPLRseq. This included splicing changes in *MSH2* and *MLH1* caused by intronic SNVs as well as nucleotide and whole exon duplications in *MSH6* and *PMS2*, respectively. An SNV creating a PTC in *MSH2* mRNA was confirmed to cause allelic imbalance in mRNA expression due to NMD as the variant transcript was rescued in PBMC cultures treated with puromycin (figure 2A).

To assess the effect of an intronic variant affecting a splice donor region of *MSH2* mRNA, we determined percent spliced-in (PSI) values which represent the fraction of reads containing a certain exon relative to all reads spanning that exonic region.[19] The analysis confirmed that the splice donor variant caused in-frame skipping of exon 5 with a PSI of 41.93%±4.16%. (figure 2B). Exon 5 skipping was similar in puromycin-treated PBMCs (45.21%±0.79%), suggesting that the aberrantly spliced isoform is not degraded by NMD as was expected for a variant that does not lead to a frame shift. Control samples show almost complete inclusion of exon 5 in *MSH2* transcripts with a PSI of 90.84±3.62 (figure 2B).

Next, we sequenced a series of nine samples with known DNA variants in MMR genes graded as pathogenic based on American College of Medical Genetics and Genomics (ACMG) criteria for which we had no prior RNA data available. CAPLRseq confirmed all nine variants as either class 4 or 5 pathogenic

Table 2 Summary of mismatch repair gene variants analysed by capture and ultradeep long-read RNA sequencing

Patient	Gene	Variant type	ACMG class	Variant (Human Genome Variation Society, HGVS)	Consequence for mRNA structure/ expression	Multiplexing	Sequencing depth –P/+P*	Variant frequency –P/+P	RT-PCR variant frequency –P/+P
Disease variants previously analysed by PCR and Sanger sequencing† and confirmed by RNA capture-seq									
1	MSH2	PTC	5	NM_000251.3(MSH2):c.1147C>T (p.Arg383Ter)	Premature termination of translation and NMD of mRNA	12-plex	754X/502X	17%/32%	10%/30%
4	MSH2	Intronic SNV	5	NM_000251.3(MSH2):c.942+3A>T	In-frame skipping of exon 5	12-plex	3358X/2694X	50%/55%	50%/50%
6	MSH2	Intronic SNV	5	NM_000251.3(MSH2):c.942+3A>T	In-frame skipping of exon 5	12-plex	4805X/3725X	50%/50%	50%/50%
7	MSH2	Intronic SNV	4	NM_000251.3(MSH2):c.2635–1G>T	Splice acceptor defect resulting in exon 16 starts four nt downstream	12-plex	3033X/2741X	39%/40%	45%/45%
11	MLH1	Intronic SNV	5	NM_000249.4(MLH1):c.1558+1G>A	Partial intron 13 retention (141 nt)	6-plex	1376X/808X	24%/29%	Not applicable/20%
12	MLH1	Intronic SNV	5	NM_000249.4(MLH1):c.1559–1G>C	Skipping of exon 14	12-plex	4223X/2188X	24%/29%	Not applicable/20%
13	MSH6	PTC	5	NM_000179.3(MSH6):c.1421_1422dup (p.Gln475fs)	2 nt duplication, premature termination of translation and NMD	12-plex	5398X/3489X	25%/30%	10%/30%
15	PMS2	Duplication	4‡	NM_000535.7:c.2174+1_2175-1dup p.(?)	In-frame duplication of exon 12	12-plex	4491X/3632X	20%/20%	45%/50%
Previously classified disease variants confirmed by RNA capture-seq									
2	MSH2	PTC	5	NM_000251.3(MSH2):c.2096C>G (p.Ser699Ter)	Premature termination of translation and NMD of mRNA	12-plex	3040X/1266X	7%/23%	Not tested
3	MSH2	Deletion	5	NM_000251.3:c.1077_1276del (p.Leu360LysfsTer16)	Deletion of exon 7 resulting in frame shift, NMD	12-plex	4042X/10 277X	8%/15%	0%/15%
5	MSH2	Intronic	5	NM_000251.3(MSH2):c.2459–12A>G	New splice acceptor site resulting in intron retention (11 nt)	12-plex	6287X/3502X	12%/17%	Not tested
8	MLH1	Methylation	5	Germline promoter methylation MLH1 in MLPA	Monoallelic loss of MLH1 expression	12-plex	683X/522X	Not applicable	15%/50%
10	MLH1	Missense	5	NM_000249.4(MLH1):c.2103G>C (p.Gln701His)	In-frame skipping of exon 18	12-plex	2309X/2571X	21%/22%	50%/50%

Continued

Table 2 Continued

Patient	Gene	Variant type	ACMG class	Variant (Human Genome Variation Society, HGVS)	Consequence for mRNA structure/ expression	Multiplexing	Sequencing depth –P/+P*	Variant frequency –P/+P	RT-PCR variant frequency –P/+P
9	MLH1	Fusion transcript	5	NC_000003.12:g.(37036415-36762824inv;insTGTTAC)	Fusion transcript of MLH1 (exon 1) and DCLK3 (exons 4 and 5)	6-plex	1924X/2122X	Not applicable	Not tested
14	MSH6	Duplication	4	NM_000179.3(MSH6):c.356dup (p.Ile120HisfsTer16)	1 nt duplication, premature termination of translation and NMD	12-plex	868X/1162X	20%/40%	Not tested
16	PMS2	SVA insertion into intron 7	4	c.804–60_804-59insJN866832.1 r.803_804ins{804–73_804–60;JN866863.1:g.1_57} p.(?)	Insertion of 71 nt of SVA sequence upstream of exon 8	12-plex	13 036X/10 275X	27%/55%	20%/35%
19	PMS2	Duplication	4	NM_000535.7(PMS2):c.1076dup (p.Leu359fs)	1 nt duplication, premature termination of translation and NMD	12-plex	380X/279X	8%/69%	Not tested
Variants of unknown significance reclassified by RNA capture-seq									
18	MSH6	Intronic SNV	3–>2	NM_000179.3(MSH6):c.628–4G>C	Intact splicing of exons 3 and 4	12-plex	260X/95X		Not tested
17	PMS2	Intronic SNV	3–>4	NM_000535.7(PMS2):c.163+5G>C (p.Ser8Argfs*5)	Out-of-frame skipping of exon 2	12-plex	1140X/1062X	59%/57%	Not tested

*–P/+P: without puromycin/with puromycin.
†Morak *et al*.[6]
‡Classification based on criteria of the INSiGHT Variant Interpretation Committee (publication pending).
MLPA, multiplexed ligation-dependent probe analysis; n.a., not available; NMD, nonsense-mediated mRNA decay; PTC, premature termination codon; SVA, SINE-VNTR-Alu.

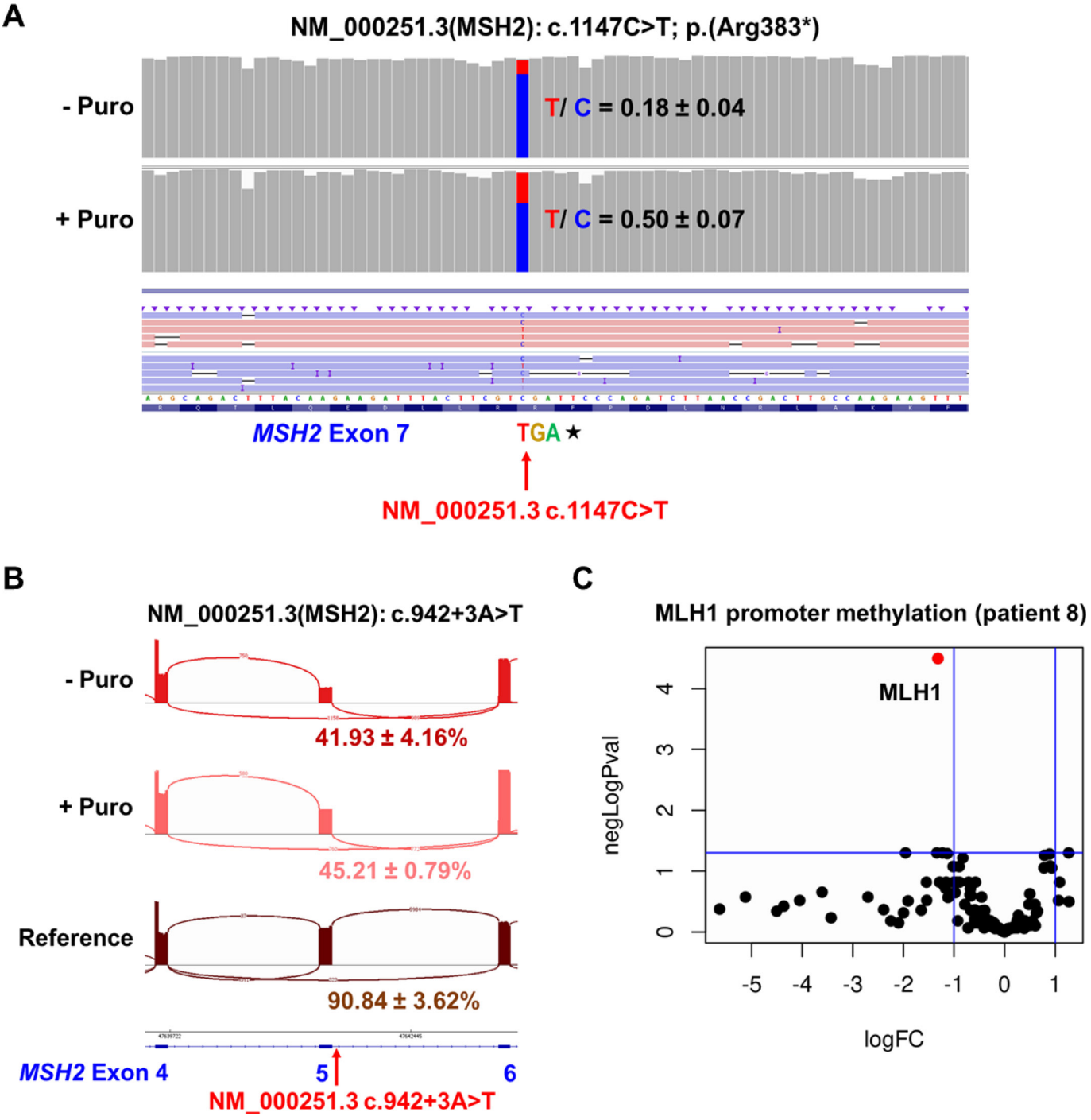


Figure 2 Effect of single-nucleotide variants on *MSH2* mRNA expression and splicing as determined by CAPLRseq. (A) RNA isolated from PBMCs of a patient carrying the indicated genomic variant was sequenced with the CAPLRseq protocol. Prior to RNA isolation, one of two parallel cultures received puromycin to inhibit nonsense-mediated mRNA decay (see the Methods section). Allele-specific expression was assessed based on the ratio of the T variant to the C allele (representative example shown, averages of three technical replicates with SD). (B) CAPLRseq analysis of a patient carrying the indicated splice site variant in *MSH2*. Percent skipping of exon 5 is indicated. Numbers are averages of two technical replicates (three for the reference sample) with SD. (C) *MLH1* expression levels in a sample from patient 8, which showed an MLH1 promoter methylation by multiplexed ligation-dependent probe analysis. *MLH1* mRNA levels were quantified as described in the Methods section relative to 37 reference RNA samples of subjects without a known variant in *MLH1*. Log2-fold change of *MLH1* mRNA (red dot) is shown in a volcano plot (n=3). CAPLRseq, capture and ultradeep long-read RNA sequencing; PBMC, peripheral blood mononuclear cell; SD, standard deviation.

variants (table 2). For example, a germline promoter methylation identified by multiplexed ligation-dependent probe analysis was apparent as a ~2-fold downregulation of MLH1 mRNA in PBMCs (figure 2C).

Furthermore, an intronic variant in *MSH2* caused an 11-nucleotide extension of exon 15 predicted to cause a frame shift and a stop codon three residues downstream (figure 3A). The frequency of the extension was increased from a PSI of

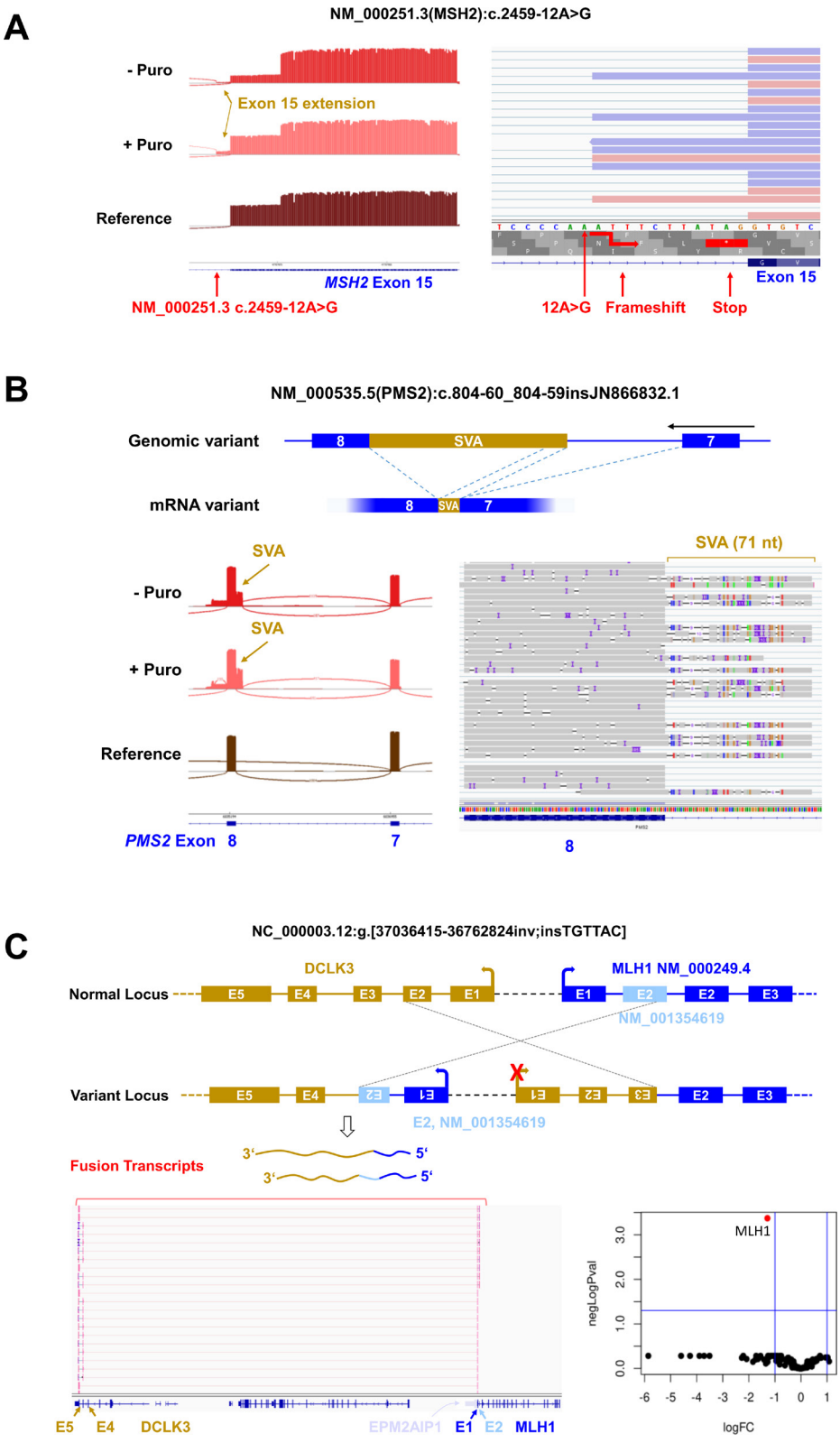


Figure 3 Effect of single-nucleotide and structural variants and on mRNA structure and expression as determined by CAPLRseq. (A) CAPLRseq analysis of a patient carrying the indicated intronic variant in *MSH2*. The extension to exon 15 (left panel) and the resulting frame shift (right panel) are shown. Numbers indicate the percentage of the variant mRNA detected with and without puromycin (averages of three technical replicates with SD). (B) CAPLRseq analysis of a patient carrying the indicated SINE-VNTR-Alu insertion in *PMS2* intron 7 schematically shown in the top panel. The 71 nucleotide extensions to exon 8 and the mismatches to the reference sequence are shown at the bottom right. Numbers indicate the percentage of the variant mRNA detected with and without puromycin (averages of three technical replicates with SD). (C) CAPLRseq analysis of a patient carrying the indicated structural variant involving the *MLH1* locus and the predicted *MLH1–DCLK3* fusion transcripts arising from it. The bottom left panel shows the fusion transcripts rendered in the Integrative Genomics Viewer. The bottom right plot shows *MLH1* expression levels relative to 37 reference RNA samples of subjects without a known variant in *MLH1*. Log2-fold change of *MLH1* mRNA (red dot) is shown in a volcano plot (n=2). CAPLRseq, capture and ultradeep long-read RNA sequencing.

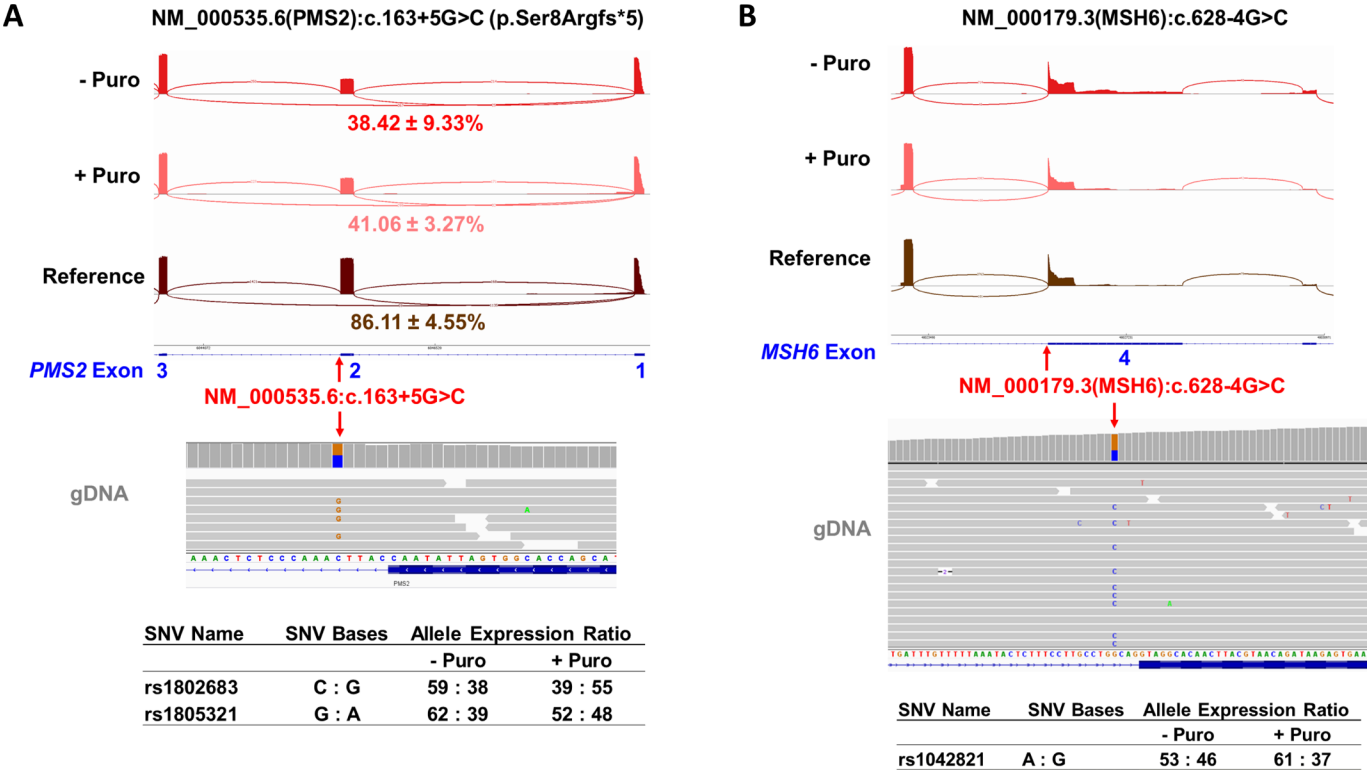
20.19±6.75 to a PSI of 26.49±3.62 in puromycin-treated PBMCs.

We also documented the deleterious effect of insertion of an SVA retrotransposon on the *PMS2* mRNA. This SVA insertion was previously shown by PCR amplification and Sanger sequencing to add 71 nucleotides to exon 8 of *PMS2* mRNA, thus resulting in a premature stop codon.[20] In the CAPLRseq data, the SVA insertion was readily detectable in the Integrated Genomics Viewer as a shoulder in the coverage track formed by a stretch of 71 nucleotides mapping to the reference sequence with multiple mismatches (figure 3B). The sequence of the insertion corresponded to the relevant region of the SVA retrotransposon (data not shown). The frequency of the variant allele was ~2-fold increased in PBMCs treated with puromycin (37.91%±5.57% without puromycin vs 54.73%±1.40% with puromycin), suggesting that it is degraded by NMD as expected for a frameshift variant.

Lastly, we confirmed fusion transcripts between *MLH1* and *DCLK3*, a gene located ~230 kb upstream of *MLH1*.[21] The fusion arises from a genomic inversion with breakpoints in *MLH1* and *DCLK3* resulting in the fusion of *MLH1* exon 1 (and parts of an alternative exon 2; see further) with exons 4 and 5 of *DCLK3* (figure 3C and data not shown). In the rearranged allele, transcription is likely driven by the inverted *MLH1* promoter. Some of the fusion transcripts also contained exon 2 of a unique isoform of *MLH1* mRNA (NM_001354619) corresponding to the genomic region in which the breakpoint is located (figure 3C). We also detected heterogeneity at the 3′ end of the *MLH1–DCLK3* fusion transcript with some isoforms missing exon 4 of *DCLK3* presumably due to alternative splicing (figure 3C). Puromycin led to a ~2.2-fold upregulation of the fusion transcripts (16.52%±0.28% without puromycin vs 36.81%±5.42% with puromycin) indicating that they are subjected to NMD. Consistent with this interpretation, we detected a ~2-fold downregulation of *MLH1* mRNA (figure 3C), which is known to cause severe deficiency in MMR activity.[22] We did not detect the predicted reverse *DCLK3–MLH1* fusion transcript, indicating that the promoter of *DCLK3* is either disrupted through the rearrangement or otherwise silenced.

Application of CAPLRseq for the diagnosis of patients with suspected HNPCC/Lynch syndrome

Finally, we sequenced two samples in which variants of unknown significance were identified in MMR genes. The first proband had a positive family history for colorectal cancer with one parent and a grandparent affected. An SNV in intron 2 of the *PMS2* gene was predicted to alter a conserved splice donor site, but in the absence of any evidence at the mRNA level, the variant was graded a VUS. CAPLRseq revealed a skipping of exon 2 in *PMS2* transcripts with a PSI of 38.42%±9.33% (figure 4A). The skipping is predicted to result in a frame shift, although we found no evidence that the variant transcript is degraded by NMD (PSI of exon 2=41.06%±3.27% in puromycin-treated PBMCs, figure 4A). Consistent with this conclusion, we did not see a major allelic imbalance in *PMS2* mRNA expression based on transcribed SNVs (figure 4A). Regardless, the frame shift predicts a defective PMS2 protein, thus presumably causing a loss-of-function of *PMS2*. These results provided evidence to reclassify the variant as class 4 according to ACMG criteria (PS3_VRS, PM2_SUP).



SNV Name	SNV Bases	Allele Expression Ratio	
		- Puro	+ Puro
rs1802683	C : G	59 : 38	39 : 55
rs1805321	G : A	62 : 39	52 : 48

SNV Name	SNV Bases	Allele Expression Ratio	
		- Puro	+ Puro
rs1042821	A : G	53 : 46	61 : 37

Figure 4 Variants of uncertain significance reclassified by CAPLRseq analysis (A) CAPLRseq analysis of a patient carrying the indicated intronic variant in *PMS2*. Percent skipping of exon 2 is indicated. Numbers are averages of three technical replicates with SD. Allelic expression was evaluated based on the ratio of reads for the indicated SNVs. (B) CAPLRseq analysis of a patient carrying the indicated intronic variant in *MSH6*. Splicing of intron 3 was assessed by visualisation in the Integrative Genomics Viewer. Images show representative data of three technical replicates. Allelic expression was evaluated based on the ratio of reads for the indicated SNVs. CAPLRseq, capture and ultradeep long-read RNA sequencing; SNV, single-nucleotide variant; gDNA, genomic DNA.

The second proband had ovarian cancer (mother) and bladder cancer (brother) in the family, and a VUS was found at the end of intron 3 of the *MSH6* gene, potentially affecting a splice acceptor site. CAPLRseq did not identify a change in splicing of exon 4 nor a change in the balance of allele expression (figure 4B), and the variant was hence reclassified as likely benign (ACMG class 2, BS3_SUP, BP4).

DISCUSSION

The CAPLRseq method we developed, validated and deployed for the diagnosis of HNPCC/Lynch syndrome provides a powerful diagnostic workflow. Its main strengths comprise its versatility in identifying the consequences of a diverse set of genomic variants for the structural integrity and expression of cognate mRNAs. The spectrum of variants includes coding and non-coding (intronic) SNVs as well as structural variants such as insertion of retrotransposons and large genomic rearrangements resulting in the formation of fusion transcripts. Whereas the use of individual splice junctions can also be assessed by short-read RNA-seq, only long-read sequencing can mirror the full diversity of the transcriptome at a single molecule level. As such, the de novo detection of altered transcripts arising from structural variants such as the *MLH1-DCLK3* inversion is not readily accomplished by short-read sequencing. Likewise, genes such as *PMS2*, which has several highly homologous pseudogenes, cannot be fully assessed by short-read RNA-seq.[4] Finally, the enrichment approach provides deep penetration of the hereditary cancer transcriptome, which spans an abundance range of four orders of magnitude, thus enabling the application of the approach in clinical genetic diagnosis.

Our standard input material for CAPLRseq was total RNA isolated from short-term cultures of patient PBMCs. Apart from straightforward sampling, PBMC cultures have several advantages that, in our view, outweigh their main disadvantage of increased handling time:

- Unlike RNA isolated from whole blood (eg, PAXgene Blood RNA System), PBMC RNA does not contain the vast amounts of *HBA* and *HBB* mRNA that can limit transcriptome penetration. Although CAPLRseq is compatible with whole blood RNA samples (table 1) because HB RNA is partially depleted during the enrichment process, HB sequencing depth remains high even after enrichment despite the inclusion of an additional HB mRNA depletion step (GlobinLock, table 1). It is thus likely that HB will negatively impact capacity to sequence cancer-related transcripts at higher levels of multiplexing, such as the 12-plex format we typically use on an R.9.4.1 flow cell. Availability of patient PBMC RNA samples free of HB mRNA also provides the option of subsequent follow-up by standard total RNA-seq, if applicable.
- A second major advantage is the ability to supplement the PMBC cultures with puromycin to inhibit NMD. The necessity to inhibit NMD-mediated degradation of aberrant transcripts for the reliable RNA-based diagnosis of HNPCC/Lynch syndrome has been repeatedly demonstrated.[6 15 23 24] Confirming this, we are showing here that NMD inhibition improves the detection of a splicing variant in *MSH2* (NM_000251.3 c.2459–12A>G, figure 3A). It is likely that NMD-mediated degradation of aberrantly spliced transcripts is equally pervasive for many other hereditary cancer genes included in our panel as was, for example, previously demonstrated for *BRCA1*.[14]
- A secondary benefit of PBMC cultures is that cells can be frozen and banked for later reuse as a source of additional RNA, DNA and protein samples for confirmation studies without the need of resampling the patient, which is often inconvenient or impossible.
- Lastly, although we have not systematically addressed this, it is conceivable that the surprisingly modest interindividual variability in the expression of hereditary cancer genes we observed within our diverse cohort of patients subjected to different environmental influences is, at least in part, due to the standardised conditions under which the patient-derived PBMCs were cultured for several days. The resulting stabilisation of expression signatures by passaging cells through uniform culture conditions might cancel out extraneous influences thus facilitating the identification of genetically driven changes of potential diagnostic value.

In summary, CAPLRseq is a highly efficient diagnostic method that readily integrates into existing workflows of modern clinical genetics laboratories. The method is automatable and cost effective with an approximate material cost of ~€400 per patient in our setting.

METHODS

Patient samples

Nineteen patients meeting at least the revised Bethesda criteria[25] were retrospectively enrolled in this study. All patients underwent genetic counselling and genetic diagnostic testing by DNA and RNA-seq with consent according to German laws. ACMG guidelines[26] were used to categorise variants as class 5 (pathogenic), class 4 (likely pathogenic), class 3 (variant of uncertain significance), class 2 (likely not pathogenic) or class 1 (not pathogenic).

RNA isolation from clinical samples

Short-term PBMC cultures were established from 3 mL whole blood anticoagulated with heparin. The blood sample was diluted with 3 mL sterile saline solution and PBMCs were isolated by centrifugation in 10 mL Leucosep tubes (Greiner). The layer containing PBMCs was removed, washed and transferred into two 15 mL conical tubes containing 5 mL PB-MAX Karyotyping Media (Thermo Fisher Scientific, catalogue number 12557021). Cells were incubated at 37°C for 72–96 hours. Five hours prior to harvesting, one of the two cultures received 50 µg/mL of the protein synthesis inhibitor puromycin to prevent NMD. Cells were collected by centrifugation, and RNA was isolated with the Qiagen RNA Blood Mini-Kit (Qiagen) or the NEB Monarch Total RNA Miniprep Kit (T2010S) according to the manufacturer's instructions. RNA yield was determined by spectrophotometry, and RNA integrity was assessed using the Agilent Fragment Analyzer.

For extraction of total RNA from whole blood, 2.5 mL blood obtained by standard venipuncture was collected into PAXgene Blood RNA Tubes (Qiagen). Blood samples were kept at room temperature for at least 2 hours prior to storage at 4°C for up to 3 days or at −20°C for long-term storage. RNA was purified on spin columns according the instructions provided in the PAXgene Blood RNA Kit. RNA yield was determined by spectrophotometry, and RNA integrity was assessed using the Agilent Fragment Analyzer. RNA samples with an RNA integrity number (RIN) >9.5 were used as input for reverse transcription.

Reverse transcription and PCR amplification

Fifty nanograms total RNA were used for cDNA synthesis following the ONT SQK-PCB109 kit protocol with the adjustments described as follows. After cDNA synthesis, four parallel PCRs per sample were performed with the respective barcoding primer pairs (online supplemental table 1) using LongAmp Taq Master Mix (New England Biolabs, catalogue number M0287) with the following cycling conditions: denaturation at 95°C for 30 s, 14 cycles of 95°C for 15 s, 62°C for 15 s, 65°C for 8 min and 20 s, 65°C for 6 min, hold at 4°C. After digestion with 20 units of exonuclease 1 (New England Biolabs, catalogue number M0293) at 37°C for 15 min, samples were heated to 80°C for 15 min, and the four PCR reactions were pooled in a 1.5 mL Eppendorf LoBind tube, followed by purification with 0.8× equivalents of resuspended AMPure XP beads (Beckman Coulter). Samples were incubated on an Intelli-Mixer ELMI RM-2M (mode F-F7, 15 RPM) for 5 min at room temperature, followed by magnetic capture and two washes with 200 µL of freshly prepared 70% ethanol. After brief drying, the pelleted beads were resuspended in 12 µL of elution buffer (EB) and incubated on an Intelli-Mixer ELMI RM-2M (mode F-90, 15 RPM) for 10 min at room temperature. Beads were retained on a magnetic rack and 12 µL of eluate was transferred to a new LoBind tube. Quantification of the amplified DNA was done with Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, catalogue numbser Q32851) using 1 µL of amplified library. Typical yields ranged between 30 ng/µL and 100 ng/µL.

cDNA capture

Purified library (between 350 ng and 850 ng) in a volume of 11 µL was used for cDNA capture with the Agilent SureSelectXT Low Input Enrichment System. Prehybridisation blocking was done by adding 5 µL of SureSelect XT HS and XT Low Input Blocker Mix followed by incubation in a thermal cycler with the following settings: heated lid on, 95°C for 5 min, 65°C for 10 min, and pause at 65°C for 1 min, during which the hybridisation mix was added. The capture library hybridisation mix was prepared at room temperature and contained 2 µL 25% RNase Block solution, 1 µL Cancer Panel Capture Library (<3 Mb; see online supplemental table 2 for individual genes), 6 µL SureSelect Fast Hybridization Buffer and 3 µL nuclease-free water. The amount of capture library was optimised by titration from 0.1 µL to 2 µL (data not shown). The mix was added to the samples while in the thermal cycler at 65°C and mixed by slowly pipetting up and down 8–10 times. Brief vortexing was followed by brief centrifugation, and the thermal cycler programme continued for 60 cycles each for 1 min at 65°C and 3 s at 37°C, after which samples were held at 65°C for a maximum of 10 min before probe binding to streptavidin beads.

Dynabeads MyOne Streptavidin T1 magnetic beads were prepared according to the manufacturer (Thermo Fisher Scientific). Briefly, 200 µL of SureSelect Binding Buffer was mixed with 50 µL of the resuspended beads. Beads were pelleted in a magnetic rack and the supernatant was discarded. A total of three washes were done, and beads were resuspended in 200 µL of SureSelect Binding Buffer. After hybridisation, samples were transferred and mixed with the washed streptavidin beads. A 30 min incubation in the Intelli-Mixer ELMI RM-2M was done at room temperature and low speed (F-F30, RPM20). Beads were gently mixed by flicking the tube every 5 min to prevent beads from settling. During this incubation, PCR tubes containing 200 µL of SureSelect Wash Buffer 2 were prewarmed at 70°C in a thermal cycler. Exact temperature control is essential at this step to maintain capture specificity. In order to remove non-hybridised DNA, beads were collected in a magnetic rack and the supernatant was discarded. Beads were fully resuspended in 200 µL of SureSelect Wash Buffer 1 by pipetting up and down 15–20 times. Using the magnet, the supernatant was removed and beads were fully resuspended in 200 µL of 70°C prewarmed Wash Buffer 2 by pipetting up and down 15–20 times. Samples were vortexed at high speed for 8 s and spun briefly, taking care that beads did not pellet but remained in suspension. Beads were incubated for 5 min at 70°C in the thermal cycler. Beads were pelleted for 1 min in the magnetic rack, and the supernatant was discarded. These steps were repeated for a total of 6 washes. After removal of the wash buffer, beads were resuspended in 25 µL of nuclease-free water.

A second PCR was done using 14 µL of the enriched library, the same ONT Barcode Primers used in the first PCR (see reverse transcription and PCR amplification) and 2× LongAmp Taq Master Mix. Cycling conditions were 95°C for 30 s, 20 cycles (95°C for 15 s, 62°C for 15 s, 65°C for 8 min and 20 s), 65°C for 6 min, hold at 4°C. After digestion with 20 units of NEB Exonuclease 1 at 37°C for 15 min and heat inactivation at 80°C for 15 min, amplification products were purified with 0.8× equivalents of resuspended AMPure XP beads as described previously, followed by elution in 12 µL of EB. Quantification of the enriched DNA was done with Qubit dsDNA HS Assay Kit, using 1 µL of sample. Typical yields ranged between 50 ng/µL and 100 ng/µL.

HB depletion

Depletion of HBA and HBB mRNA was done with the GlobinLock procedure.[18] For this, the ONT SQK-PCB109 protocol was modified by annealing locked nucleic acid (LNA) containing oligonucleotides complementary to the 3′-untranslated regions (UTRs) of *HBA* (oligo LNA-A) and *HBB* (oligo LNA-B) mRNAs (online supplemental table 1) prior to first strand cDNA synthesis. Reactions of 10 µL contained 50 ng of total RNA, 3 µM each of LNA-A and LNA-B, 125 mM KCl, 1 mM dNTPs, and 1× RT buffer. Samples were heated to 95°C for 30 s and incubated at 60°C for 5 min. After this, 1 µL VNP primer was added (0.2 µM final concentration), followed by incubation at 60°C for another 5 min, after which the sample was placed on ice. Reverse transcription was started by adding 2 µL 5× RT buffer, 1 µL RNAseOUT, 2 µL SSP, 3 µL H_2O and 1 µL Maxima RT and incubation as per ONT SQK-PCB109 protocol.

Adapter ligation and ONT sequencing

Amplified library of 100 fmol in a volume of 11 µL was used for sequencing. Assuming a mean length of 1.5 kb, 100 fmol corresponds to ~100 ng. If multiplexed libraries were sequenced on one flow cell, barcoded libraries were mixed at equal ratio to obtain a total of 100 fmol in 11 µL. ONT Rapid Adapter of 1 µL was added and incubated for 5 min at room temperature. The prepared library was kept on ice until loading onto the sequencing flow cell.

Priming and sample loading onto R.9.4.1. flow cells were done according to standard protocols described by ONT.

Sequencing was performed on the GridION platform using the ONT Flo-Min 106D flow cell setting and the standard MinKNOW protocol script (NC_48 hours_sequencing_FLO-MIN106_SQK-PBC109) with accurate base calling and demultiplexing turned on.

Data analysis

Mapping for the analysis of aberrant splicing and transcript fusion detection was performed using minimap2 V.2.17-r941. Fusion transcripts were detected using JAFFA V.2.2. Percentages of aberrant transcripts due to splicing defects were determined either by manual inspection of mapped reads in Integrative Genomics Viewer (IGV) and deduction of affected read counts or by calculation of PSI scores using PSI-Sigma V.1.9r[19] as applicable. Differential gene expression (DGE) analysis was performed using ONT's pipeline for DGE and differential transcript usage analysis of long reads (https://github.com/nanoporetech/pipeline-transcriptome-de). Accordingly, mapping was performed using minimap2 V.2.18, expression was quantified using salmon V.1.5.0, and DGE analysis was performed using edgeR V.3.34.0. Log2-fold changes and adjusted p values were calculated relative to the cohort of all samples, and data were plotted in volcano plots.

Acknowledgements We thank all members of the MGZ Medical Genetics Center for supporting this work.

Contributors Conceptualisation: EH-F and DAW. Data curation: FS, VS, JMAP, AL and MM. Investigation: VS, RMLS, FS, TH, MW and DAW. Formal analysis: VS, RMLS, FS, KK, AL and MM. Visualisation: VS, FS and DAW. Writing: DAW and FS. Guarantor: DAW.

Funding Initial phases of this project were funded by German Cancer Aid (project no. 111222) and the Wilhelm Sander-Stiftung (#2012.081.1).

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval This study involves human participants and was approved by the ethics committee of the Ludwig-Maximilians University (project number 20-0839). The participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request.

ORCID iDs
Rafaela Magalhaes Leal Silva http://orcid.org/0000-0002-0062-1865
Florentine Scharf http://orcid.org/0000-0002-3848-312X
Dieter A Wolf http://orcid.org/0000-0002-3761-1070

REFERENCES

1 LaDuca H, Stuenkel AJ, Dolinsky JS, Keiles S, Tandy S, Pesaran T, Chen E, Gau C-L, Palmaer E, Shoaepour K, Shah D, Speare V, Gandomi S, Chao E. Utilization of multigene panels in hereditary cancer predisposition testing: analysis of more than 2,000 patients. *Genet Med* 2014;16:830–7.
2 Nielsen FC. Hansen T van O, Sørensen Cs. hereditary breast and ovarian cancer: new genes in confined pathways. *Nat Rev Cancer* 2016;16:599–612.
3 Rhine CL, Cygan KJ, Soemedi R, Maguire S, Murray MF, Monaghan SF, Fairbrother WG. Hereditary cancer genes are highly susceptible to splicing mutations. *PLoS Genet* 2018;14:e1007231.
4 Landrith T, Li B, Cass AA, Conner BR, LaDuca H, McKenna DB, Maxwell KN, Domchek S, Morman NA, Heinlen C, Wham D, Koptiuch C, Vagher J, Rivera R, Bunnell A, Patel G, Geurts JL, Depas MM, Gaonkar S, Pirzadeh-Miller S, Krukenberg R, Seidel M, Pilarski R, Farmer M, Pyrtel K, Milliron K, Lee J, Hoodfar E, Nathan D, Ganzak AC, Wu S, Vuong H, Xu D, Arulmoli A, Parra M, Hoang L, Molparia B, Fennessy M, Fox S, Charpentier S, Burdette J, Pesaran T, Profato J, Smith B, Haynes G, Dalton E, Crandall JR-R, Baxter R, Lu H-M, Tippin-Davis B, Elliott A, Chao E, Karam R. Splicing profile by capture RNA-seq identifies pathogenic germline variants in tumor suppressor genes. *npj Precision Oncology* 2020;4:1–9.
5 Thompson BA, Walters R, Parsons MT, et al.. Contribution of mRNA Splicing to Mismatch Repair Gene Sequence Variant Interpretation. Frontiers in Genetics 11:798, 2020[cited 2022 Aug 30];11. Available from:. Available: https://www.frontiersin.org/articles/10.3389/fgene.2020.00798
6 Morak M, Schaefer K, Steinke-Lange V, Koehler U, Keinath S, Massdorf T, Mauracher B, Rahner N, Bailey J, Kling C, Haeusser T, Laner A, Holinski-Feder E. Full-Length transcript amplification and sequencing as universal method to test mRNA integrity and biallelic expression in mismatch repair genes. *Eur J Hum Genet* 2019;27:1808–20.
7 Karam R, Conner B, LaDuca H, McGoldrick K, Krempely K, Richardson ME, Zimmermann H, Gutierrez S, Reineke P, Hoang L, Allen K, Yussuf A, Farber-Katz S, Rana HQ, Culver S, Lee J, Nashed S, Toppmeyer D, Collins D, Haynes G, Pesaran T, Dolinsky JS, Tippin Davis B, Elliott A, Chao E. Assessment of diagnostic outcomes of RNA genetic testing for hereditary cancer. *JAMA Netw Open* 2019;2.
8 Byrne A, Beaudin AE, Olsen HE, Jain M, Cole C, Palmer T, DuBois RM, Forsberg EC, Akeson M, Vollmers C. Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells. *Nat Commun* 2017;8:1–11.
9 Byrne A, Supple MA, Volden R, et al.. Depletion of Hemoglobin Transcripts and Long-Read Sequencing Improves the Transcriptome Annotation of the Polar Bear (Ursus maritimus). Front Genet [Internet]., 2019;10. Available from:. Available: https://www.frontiersin.org/articles/10.3389/fgene.2019.00643/full [Accessed cited 2019 Sep 18].
10 Volden R, Palmer T, Byrne A, Cole C, Schmitz RJ, Green RE, Vollmers C. Improving nanopore read accuracy with the R2C2 method enables the sequencing of highly multiplexed full-length single-cell cDNA. *Proc Natl Acad Sci U S A* 2018;115:9726–31.
11 Hardwick SA, Bassett SD, Kaczorowski D, et al.. Targeted, High-Resolution RNA Sequencing of Non-coding Genomic Regions Associated With Neuropsychiatric Functions. Front Genet [Internet]., 201910. Available from:. Available: https://www.frontiersin.org/articles/10.3389/fgene.2019.00309/full [Accessed cited 2020 Mar 13].
12 de Jong LC, Cree S, Lattimore V, Wiggins GAR, Spurdle AB, Miller A, Kennedy MA, Walker LC, kConFab Investigators. Nanopore sequencing of full-length BRCA1 mRNA transcripts reveals co-occurrence of known exon skipping events. *Breast Cancer Res* 2017;19.
13 Suzuki A, Suzuki M, Mizushima-Sugano J, Frith MC, Makalowski W, Kohno T, Sugano S, Tsuchihara K, Suzuki Y. Sequencing and phasing cancer mutations in lung cancers using a long-read portable sequencer. *DNA Res* 2017;24:585–96.
14 Perrin-Vidoz L, Sinilnikova OM, Stoppa-Lyonnet D, Lenoir GM, Mazoyer S. The nonsense-mediated mRNA decay pathway triggers degradation of most BRCA1 mRNAs bearing premature termination codons. *Hum Mol Genet* 2002;11:2805–14.
15 Sumitsuji I, Sugano K, Matsui T, Fukayama N, Yamaguchi K, Akasu T, Fujita S, Moriya Y, Yokoyama R, Nomura S, Yoshida T, Kodama T, Ogawa M. Frequent genomic disorganisation of MLH1 in hereditary non-polyposis colorectal cancer (HNPCC) screened by RT-PCR on puromycin treated samples. *J Med Genet* 2003;40:30e–30.
16 Lagarde J, Uszczynska-Ratajczak B, Carbonell S, Pérez-Lluch S, Abad A, Davis C, Gingeras TR, Frankish A, Harrow J, Guigo R, Johnson R. High-Throughput annotation of full-length long noncoding RNAs with capture long-read sequencing. *Nat Genet* 2017;49:1731–40.
17 Deveson IW, Brunck ME, Blackburn J, Tseng E, Hon T, Clark TA, Clark MB, Crawford J, Dinger ME, Nielsen LK, Mattick JS, Mercer TR. Universal alternative splicing of noncoding exons. *Cell Syst* 2018;6:245–55.
18 Krjutškov K, Koel M, Roost AM, Katayama S, Einarsdottir E, Jouhilahti E-M, Söderhäll C, Jaakma Ülle, Plaas M, Vesterlund L, Lohi H, Salumets A, Kere J. Globin mRNA reduction for whole-blood transcriptome sequencing. *Sci Rep* 2016;6:31584.
19 Lin K-T, Krainer AR. PSI-Sigma: a comprehensive splicing-detection method for short-read and long-read RNA-seq analysis. *Bioinformatics* 2019;35:5048–54.
20 van der Klift HM, Tops CM, Hes FJ, Devilee P, Wijnen JT. Insertion of an SVA element, a nonautonomous retrotransposon, in *PMS2* intron 7 as a novel cause of Lynch syndrome. *Hum Mutat* 2012;33:1051–5.
21 Arnold AM, Morak M, Benet-Pagès A, Laner A, Frishman D, Holinski-Feder E. Targeted deep-intronic sequencing in a cohort of unexplained cases of suspected Lynch syndrome. *Eur J Hum Genet* 2020;28:597–608.
22 Kansikas M, Kasela M, Kantelinen J, Nyström M. Assessing how reduced expression levels of the mismatch repair genes MLH1, MSH2, and MSH6 affect repair efficiency. *Hum Mutat* 2014;35:1123–7.

23 Andreutti-Zaugg C, Scott RJ, Iggo R. Inhibition of nonsense-mediated messenger RNA decay in clinical samples facilitates detection of human MSH2 mutations with an in vivo fusion protein assay and conventional techniques. *Cancer Res* 1997;57:3288–93.
24 Vreeswijk MPG, van der Klift HM. Analysis and Interpretation of RNA Splicing Alterations in Genes Involved in Genetic Disorders. In: Aartsma-Rus A, ed. *Exon Skipping: Methods and Protocols*. Totowa, NJ: Humana Press, 2012: 49–63(Methods in Molecular Biology). Available from:. (cited 2022 Apr 14).
25 Umar A, Boland CR, Terdiman JP, Syngal S, Chapelle Adl, Ruschoff J, Fishel R, Lindor NM, Burgart LJ, Hamelin R, Hamilton SR, Hiatt RA, Jass J, Lindblom A, Lynch HT, Peltomaki P, Ramsey SD, Rodriguez-Bigas MA, Vasen HFA, Hawk ET, Barrett JC, Freedman AN, Srivastava S. Revised Bethesda guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *JNCI Journal of the National Cancer Institute* 2004;96:261–8.
26 Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, ACMG Laboratory Quality Assurance Committee. 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* 2015;17:405–24.

Supplemental Information

Transcript Capture and Ultra-Deep Long-Read RNA Sequencing (CAPLRseq) to Diagnose HNPCC/Lynch Syndrome

Vincent Schwenk [1,4], Rafaela Magalhaes Leal Silva [1,4], Florentine Scharf [1,4], Katharina Knaust [1], Martin Wendlandt [1], Tanja Häusser [1], Julia M. A. Pickl [1,2], Verena Steinke-Lange [1,2], Andreas Laner [1], Monika Morak [2], Elke Holinski-Feder [1,2,5], Dieter A. Wolf [1,3,5]

Schwenk V, *et al. J Med Genet* 2023;0:1–13. doi: 10.1136/jmg-2022-108931

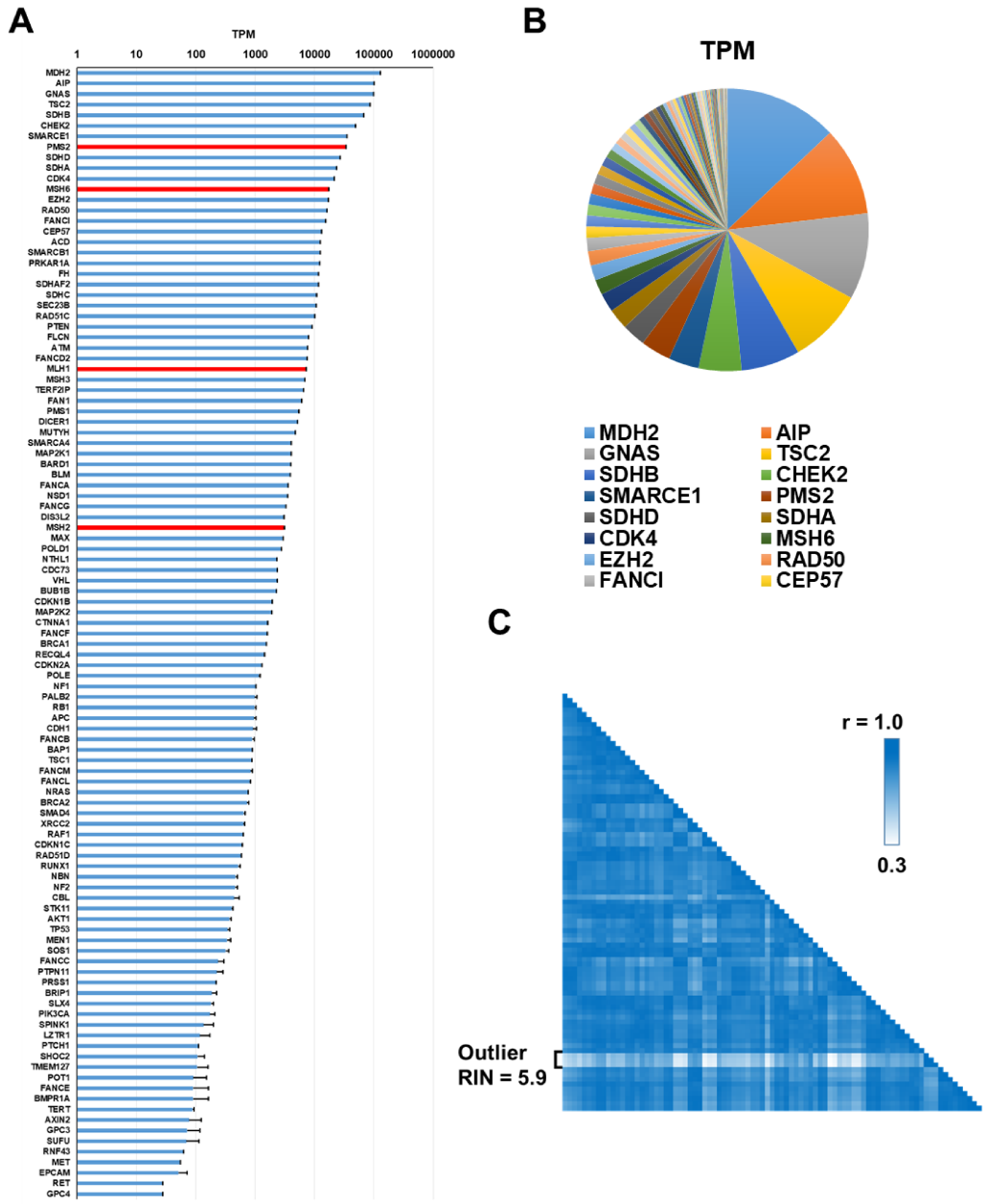


Figure S1 Quantification of transcripts from 123 hereditary cancer genes by CAPLRseq

A. Relative abundance of the indicated transcripts (in transcripts per million reads) plotted on a log scale. Error bars represent relative standard deviations of technical triplicates sequenced in a 12-plex format on a single R9.4 flow cell. Transcripts encoding the mismatch repair proteins MSH2, PMS2, MLH1, and MSH6 are highlighted in red.

B. Pie chart of the data in A. showing that ~75% of all reads represent the 16 most abundant mRNAs.

C. Reproducibility of cancer gene transcript quantification by RNA capture -seq. Heat map of Pearson correlation coefficients of technical triplicates of RNA samples from 29 distinct patients.

Table S1 Barcoding primers and GlobinLock LNAs

	Name	Sequence
Top	BC01	ATCGCCTACCGTGACAAGAAAGTTGTCGGTGTCTTTGTGACTTGCCTGTCGCTCTATCTTC
	BC02	ATCGCCTACCGTGACTCGATTCCGTTTGTAGTCGTCTGTACTTGCCTGTCGCTCTATCTTC
	BC03	ATCGCCTACCGTGACGAGTCTTGTGTCCCAGTTACCAGGACTTGCCTGTCGCTCTATCTTC
	BC04	ATCGCCTACCGTGACTTCGGATTCTATCGTGTTTCCCTAACTTGCCTGTCGCTCTATCTTC
	BC05	ATCGCCTACCGTGACCTTGTCCAGGGTTTGTGTAACCTTACTTGCCTGTCGCTCTATCTTC
	BC06	ATCGCCTACCGTGACTTCTCGCAAAGGCAGAAAGTAGTCACTTGCCTGTCGCTCTATCTTC
	BC07	ATCGCCTACCGTGACGTGTTACCGTGGGAATGAATCCTTACTTGCCTGTCGCTCTATCTTC
	BC08	ATCGCCTACCGTGACTTCAGGGAACAAACCAAGTTACGTACTTGCCTGTCGCTCTATCTTC
	BC09	ATCGCCTACCGTGACAACTAGGCACAGCGAGTCTTGGTTACTTGCCTGTCGCTCTATCTTC
	BC10	ATCGCCTACCGTGACAAGCGTTGAAACCTTTGTCCTCTCACTTGCCTGTCGCTCTATCTTC
	BC11	ATCGCCTACCGTGACGTTTCATCTATCGGAGGGAATGGAACTTGCCTGTCGCTCTATCTTC
	RB12A	ATCGCCTACCGTGACGTTGAGTTACAAAGCACCGATCAGACTTGCCTGTCGCTCTATCTTC
Bottom	BC01	ATCGCCTACCGTGACAAGAAAGTTGTCGGTGTCTTTGTGTTTCTGTTGGTGCTGATATTGC
	BC02	ATCGCCTACCGTGACTCGATTCCGTTTGTAGTCGTCTGTTTTCTGTTGGTGCTGATATTGC
	BC03	ATCGCCTACCGTGACGAGTCTTGTGTCCCAGTTACCAGGTTTCTGTTGGTGCTGATATTGC
	BC04	ATCGCCTACCGTGACTTCGGATTCTATCGTGTTTCCCTATTTCTGTTGGTGCTGATATTGC
	BC05	ATCGCCTACCGTGACCTTGTCCAGGGTTTGTGTAACCTTTTTCTGTTGGTGCTGATATTGC
	BC06	ATCGCCTACCGTGACTTCTCGCAAAGGCAGAAAGTAGTCTTTCTGTTGGTGCTGATATTGC
	BC07	ATCGCCTACCGTGACGTGTTACCGTGGGAATGAATCCTTTTTCTGTTGGTGCTGATATTGC
	BC08	ATCGCCTACCGTGACTTCAGGGAACAAACCAAGTTACGTTTTCTGTTGGTGCTGATATTGC
	BC09	ATCGCCTACCGTGACAACTAGGCACAGCGAGTCTTGGTTTTTCTGTTGGTGCTGATATTGC
	BC10	ATCGCCTACCGTGACAAGCGTTGAAACCTTTGTCCTCTCTTTCTGTTGGTGCTGATATTGC
	BC11	ATCGCCTACCGTGACGTTTCATCTATCGGAGGGAATGGATTTCTGTTGGTGCTGATATTGC
	BC12A	ATCGCCTACCGTGACGTTGAGTTACAAAGCACCGATCAGTTTCTGTTGGTGCTGATATTGC
GlobinLock LNAs	LNA-HBA	TTTTTG+CYGCCC+ACTCAG+ACTTTA+TTC [a]
	LNA-HBB	TTTTTTTTTTG+CAATGA+AAATAA+ATGTTT+TTTATTAGG

[a] "+" denotes a LNA nucleotide, Y denotes a pyrimidine nucleotide

Supplementary Table 2: Cancer Genes

#	Systematic Name	Transcript	OMIM
1	ACD	NM_001082486.1	609377
2	AIP	NM_003977.3	605555
3	AKT1	NM_001014432.1	164730
4	ALK	NM_004304.4	105590
5	APC	NM_000038.4	611731
6	ATM	NM_000051.3	607585
7	AXIN2	NM_004655.3	604025
8	BAP1	NM_004656.3	603089
9	BARD1	NM_000465.3	601593
10	BLM	NM_000057.3	604610
11	BMPR1A	NM_004329.2	601299
12	BRCA1	NM_007294.3	113705
13	BRCA2	NM_000059.3	600185
14	BRIP1	NM_032043.2	605882
15	BUB1B	NM_001211.5	602860
16	CASR	NM_000388.3	601199
17	CBL	NM_005188.3	165360
18	CDC73	NM_024529.4	607393
19	CDH1	NM_004360.4	192090
20	CDK4	NM_000075.3	123829
21	CDKN1B	NM_004064.4	600778
22	CDKN1C	NM_000076.2	600856
23	CDKN2A	NM_000077.4	600160
24	CDKN2B	NM_004936.3	600431
25	CEP57	NM_014679.4	607951
26	CFTR	NM_000492.3	602421
27	CHEK2	NM_007194.3	604373
28	CTNNA1	NM_001290307.2	116805
29	CTRC	NM_007272.2	601405
30	DICER1	NM_177438.2	606241
31	DIS3L2	NM_152383.4	614184
32	EPAS1	NM_001430.4	603349
33	EPCAM	NM_002354.2	185535
34	EZH2	NM_004456.4	601573
35	FAN1	NM_014967.4	613534
36	FANCA	NM_000135.2	607139
37	FANCB	NM_001018113.2	300515
38	FANCC	NM_000136.2	613899
39	FANCD2	NM_001018115.2	613984
40	FANCE	NM_021922.2	613976
41	FANCF	NM_022725.3	613897
42	FANCG	NM_004629.1	602956
43	FANCI	NM_001113378.1	611360
44	FANCL	NM_001114636.1	608111
45	FANCM	NM_020937.3	609644
46	FH	NM_000143.3	136850
47	FLCN	NM_144997.5	607273
48	GNAS	NM_000516.5	139320
49	GPC3	NM_004484.3	300037
50	GPC4	NM_001448.2	300168
51	GREM1	NM_013372.6	603054
52	HOXB13	NM_006361.5	604607
53	KIT	NM_000222.2	164920
54	LZTR1	NM_006767.3	600574
55	MAP2K1	NM_002755.3	176872
56	MAP2K2	NM_030662.3	601263
57	MAX	NM_002382.4	154950
58	MDH2	NM_005918.3	154100
59	MEN1	NM_000244.3	613733
60	MET	NM_001127500.2	164860
61	MITF	NM_000248.3	156845
62	MLH1	NM_000249.3	120436
63	MRE11	NM_005591.3	600814
64	MSH2	NM_000251.2	609309
65	MSH3	NM_002439.4	600887
66	MSH6	NM_000179.2	600678
67	MUTYH	NM_001128425.1	604933
68	NBN	NM_002485.4	602667
69	NF1	NM_000267.3	613113
70	NF2	NM_000268.3	607379
71	NRAS	NM_002524.4	164790
72	NSD1	NM_022455.4	606681
73	NTHL1	NM_002528.6	602656
74	PALB2	NM_024675.3	610355
75	PDGFRA	NM_006206.4	173490
76	PHOX2B	NM_003924.3	603851
77	PIK3CA	NM_006218.3	171834
78	PMS1	NM_000534.4	600258
79	PMS2	NM_000535.6	600259
80	POLD1	NM_001256849.1	174761
81	POLE	NM_006231.3	174762
82	POT1	NM_015450.2	606478
83	PRKAR1A	NM_002734.4	188830
84	PRSS1	NM_002769.4	276000
85	PTCH1	NM_000264.3	601309
86	PTCH2	NM_003738.4	603673
87	PTEN	NM_000314.6	601728
88	PTPN11	NM_002834.3	176876
89	RAD50	NM_005732.3	604040
90	RAD51C	NM_058216.2	602774
91	RAD51D	NM_002878.3	602954
92	RAF1	NM_002880.3	164760
93	RB1	NM_000321.2	614041
94	RECQL4	NM_004260.3	603780
95	RET	NM_020975.4	164761
96	RNF43	NM_017763.5	612482
97	RUNX1	NM_001754.4	151385
98	SDHA	NM_004168.3	600857
99	SDHAF2	NM_017841.2	613019
100	SDHB	NM_003000.2	185470
101	SDHC	NM_003001.3	602413
102	SDHD	NM_003002.3	602690
103	SEC23B	NM_006363.4	610512
104	SHOC2	NM_007373.3	602775
105	SLX4	NM_032444.2	613278
106	SMAD4	NM_005359.5	600993
107	SMARCA4	NM_001128849.1	603254
108	SMARCB1	NM_003073.4	601607
109	SMARCE1	NM_003079.4	603111
110	SOS1	NM_005633.3	182530
111	SPINK1	NM_003122.4	167790
112	SPRED1	NM_152594.2	609291
113	STK11	NM_000455.4	602216
114	SUFU	NM_016169.3	607035
115	TERF2IP	NM_018975.3	605061
116	TERT	NM_198253.2	187270
117	TMEM127	NM_017849.3	613403
118	TP53	NM_000546.5	191170
119	TSC1	NM_000368.4	605284
120	TSC2	NM_000548.4	191092
121	VHL	NM_000551.3	608537
122	WT1	NM_024426.4	607102
123	XRCC2	NM_005431.1	600375