



Open

Constitutional mismatch repair deficiency is the diagnosis in 0.41% of pathogenic *NF1/SPRED1* variant negative children suspected of sporadic neurofibromatosis type 1

Juan A. Perez-Valencia, PhD¹, Richard Gallon, PhD², Yunjia Chen, PhD³, Jakob Koch, MSc¹, Markus Keller, PhD¹, Klaus Oberhuber, Dipl. MTA¹, Alicia Gomes, MS, CGC³, Johannes Zschocke, MD, PhD¹, John Burn, MD², Michael S. Jackson, PhD², Mauro Santibanez-Koref, MD², Ludwine Messiaen, PhD³ and Katharina Wimmer, PhD¹

Purpose: Biallelic germline mismatch repair (MMR) gene pathogenic variants (PVs) cause constitutional MMR deficiency (CMMRD), a highly penetrant childhood cancer syndrome phenotypically overlapping with neurofibromatosis type 1 (NF1). CMMRD testing in suspected NF1 children without *NF1/SPRED1* PVs enables inclusion of CMMRD positives into monitoring programs prior to tumor onset. However, testing is associated with potential harms and the prevalence of CMMRD among these children is unknown.

Methods: Using a simple and scalable microsatellite instability (MSI) assay of non-neoplastic leukocyte DNA to detect CMMRD, we retrospectively screened >700 children suspected of sporadic NF1 but lacking *NF1/SPRED1* PVs.

Results: For three of seven MSI-positive patients germline MMR gene PVs confirmed the diagnosis of CMMRD. Founder variants NM_000535.5(PMS2):c.736_741delinsTGTGTGTGAAG, prevalent in Europe and North America, and NM_000179.2(MSH6):

c.10C>G, affecting 1:400 French Canadians, represented two of five PVs. The prevalence of CMMRD was 3/735 (0.41%, 95% confidence interval [CI]: 0.08–1.19%).

Conclusion: Our empirical data provide reliable numbers for genetic counseling and confirm previous prevalence estimations, on which Care for CMMRD consortium guidelines are based. These advocate CMMRD testing of preselected patients rather than offering reflex testing to *all* suspected sporadic NF1 children lacking *NF1/SPRED1* PVs. The possibility of founder effects should be considered alongside these testing guidelines.

Genetics in Medicine (2020) 22:2081–2088; <https://doi.org/10.1038/s41436-020-0925-z>

Keywords: constitutional mismatch repair deficiency; neurofibromatosis type 1; childhood cancer; microsatellite instability; founder variant

INTRODUCTION

Constitutional mismatch repair deficiency (CMMRD; MIM 276300) is a recessive childhood cancer syndrome caused by pathogenic variants (PVs) in both alleles of one of the mismatch repair (MMR) genes (*MLH1*, MIM *120436; *MSH2*, MIM *609309; *MSH6*, MIM *600678; *PMS2*, MIM *600259).¹ CMMRD is associated with an extraordinarily high tumor risk, with 321 malignancies in a cohort of 197 CMMRD patients analyzed in 2017.² Nearly all known CMMRD patients developed a malignancy within the first three decades of life, predominantly in the first decade.² The spectrum of CMMRD-associated malignancies is very broad, including primarily hematological malignancies, tumors of the central nervous system, and carcinomas associated with Lynch syndrome.³ Therefore, rigorous surveillance protocols, including annual to biannual brain magnetic resonance image (MRI) from the time of diagnosis and colonoscopy starting at

age 6–8 years, are recommended.^{4–6} Timely diagnosis of CMMRD also has implications for the patient's family. The risk of recurrence in a sibling is 25%, and the parents are carriers of heterozygous MMR gene PVs that predispose to Lynch syndrome-associated cancers in adulthood.⁶

CMMRD shows phenotypic overlap with neurofibromatosis 1 (NF1), a common neurocutaneous disorder (incidence 1:2000–3000).⁷ Hallmark features of NF1 are café-au-lait macules (CALMs), skinfold freckling, and neurofibromas. These, and additional characteristic features, are included in the National Institutes of Health (NIH) (Consensus Conference) diagnostic criteria for NF1. Presence of at least two of these NIH criteria confirms NF1.⁸ However, young children often do not fulfill these criteria as NF1 is a progressive disorder and approximately half of NF1 patients are sporadic cases. Highly sensitive *NF1* variant analysis protocols^{9,10} enable early diagnosis in these children without waiting for

¹Institute of Human Genetics, Medical University of Innsbruck, Innsbruck, Austria; ²Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK; ³Department of Genetics, University of Alabama at Birmingham, Birmingham, AL, USA. Correspondence: Katharina Wimmer (katharina.wimmer@i-med.ac.at)

These authors contributed equally: Juan A. Perez-Valencia, Richard Gallon

Submitted 6 May 2020; revised 16 July 2020; accepted: 22 July 2020

Published online: 10 August 2020

unequivocal clinical presentation. If no germline *NF1* PV is found, mosaic or segmental *NF1* due to postzygotic mutations¹¹ and Legius syndrome caused by a *SPRED1* PV are the most important differential diagnoses¹² expected to account for at least 10% and 2.4%, respectively.¹³

CALMs are also present in 62–97% of CMMRD patients, and approximately 20% of CMMRD patients show more than one NIH *NF1* feature.² Therefore, *NF1* features are included as one diagnostic criterion in the Care for CMMRD (C4CMMRD) scoring system for the clinical suspect diagnosis of CMMRD in pediatric cancer patients.³ Prior to malignancy, CMMRD may be indistinguishable from *NF1*, as exemplified by several cases who received an incorrect initial diagnosis of *NF1*.^{14–17} Hence, CMMRD is a differential diagnosis in suspected sporadic *NF1* children for whom no *NF1* or *SPRED1* PV is identified. Testing for CMMRD in these children would allow for cancer surveillance before their first malignancy, predictive genetic testing and surveillance in relatives (who are at risk for both CMMRD and Lynch syndrome), and may impact family planning. However, testing and counseling for CMMRD could cause significant, and potentially unnecessary, harms, including anxiety in the patient and family, difficult diagnostic and management decisions where a variant of unknown significance (VUS) is identified, and the diagnosis of Lynch syndrome in a minor (for details see Suerink *et al*.¹⁸). Therefore, several interdisciplinary teams and consortia have started discussing if and when to counsel and test for CMMRD in *NF1/SPRED1* PV negatives.^{18,19}

The C4CMMRD consortium estimated the prevalence of CMMRD to be 0.39% (1/258) among malignancy-free children who are suspected to have sporadic *NF1* but lack *NF1/SPRED1* PVs. Based on this estimate, and balancing the benefits of a diagnosis against the potential harms, C4CMMRD has proposed that counseling and testing should be restricted to cases where CMMRD is most likely.¹⁸ However, this prevalence estimation is based on a number of assumptions. To perform a reliable benefit–risk evaluation of the C4CMMRD guidelines and to inform genetic counseling, it is thus important to establish a robust empirical data basis. Due to the expected low prevalence, it will be necessary to analyze a large retrospective cohort of genetically and clinically well-characterized patients. Equally, a reliable and scalable CMMRD screening pipeline is needed, since direct mutational testing of the four MMR genes in every patient has two major limitations. First, *PMS2*, which is mutated in the majority of CMMRD patients, is a recognized “dead zone” for genomic sequencing due to its multiple pseudogenes²⁰ and requires specialized, non-automatable variant analysis protocols.^{21,22} Second, identification of a VUS will preclude a definitive diagnosis unless additional assays can prove its pathogenicity or a complementary assay confirms the diagnosis. In particular, microsatellite instability (MSI) analysis can be used to detect low-frequency microsatellite length variants in the patient’s constitutional DNA, which is a distinctive hallmark of CMMRD.^{23–26}

We have previously developed a scalable and low cost MSI assay that detects microsatellite length variants in DNA extracted from non-neoplastic peripheral blood leukocytes (PBLs) of CMMRD patients.²⁵ It analyzes 24 mononucleotide repeat microsatellites that are sensitive to loss of function of any MMR gene, which is not the case for alternative assays that analyze dinucleotide repeats insensitive to MSH6 deficiency.^{23,25} The molecular inversion probe (MIP)-based protocol allows scalable amplicon sequencing, and the custom analysis pipeline, which utilizes molecular barcodes to reduce polymerase chain reaction (PCR) and sequencing errors, allows accurate and automated detection of low-frequency microsatellite variants from read depths <5000×. As such, 96 samples can be analyzed on a single MiSeq v2 flow cell, requiring far fewer reads than other sequencing-based methods.²⁶ The assay generates an easy to interpret MSI score: a higher MSI score means a higher probability that the sample has increased MSI relative to a control set. We showed that the assay identifies CMMRD with 98% sensitivity and 100% specificity using a classification threshold of score >2.00 (>99% probability of increased MSI). The single false negative result was attributed to the patient’s chemotherapy-induced aplasia when this sample was collected, as additional samples collected after recovery from aplasia were correctly classified.²⁵ As aplasia will not affect suspected *NF1* children without malignancy, the chance of false negatives when screening these children will be negligible.

Here, we used an updated version of this MSI assay to screen >700, genetically and clinically well-characterized patients fulfilling the testing prerequisites specified by C4CMMRD guidelines,¹⁸ to provide a reliable empirical estimate of CMMRD prevalence as a differential diagnosis among malignancy-free children who are suspected of sporadic *NF1* but lack *NF1/SPRED1* PVs.

MATERIALS AND METHODS

Study ethics

This study was approved by the Medical University of Innsbruck Review Board, EK number 1012/2018, and by the University of Alabama at Birmingham Institutional Review Board, protocol X040517014.

Patients and samples

Informed by a case number analysis (Supplementary Information S1), we aimed to screen at least 666 children aged 1–17 years who fulfilled the prerequisites for CMMRD counseling and testing as defined by C4CMMRD:¹⁸ (1) absence of diagnostic *NF1* sign(s) in both parents, (2) absence of an identifiable *NF1* or *SPRED1* germline PV after comprehensive and highly sensitive variant analysis, and (3) presence of multiple (>5) CALMs, or at least two hyperpigmented skin patches reminiscent of CALMs in addition to one other diagnostic NIH *NF1* feature. Patients were selected from two academic diagnostic centers: The Medical Genomics Laboratory, University of Alabama–Birmingham (UAB), and the Institute of Human Genetics, Medical University of

Innsbruck (MUI). Both centers perform comprehensive *NF1* and *SPRED1* variant analysis, which includes *NF1* transcript analysis using direct complementary DNA (cDNA) sequencing to detect (splice) PVs that may escape detection by genomic DNA (gDNA) analysis.¹⁰ This ensured exclusion of *NF1* germline PVs by highly sensitive methods and availability of RNA for *PMS2* variant analysis by cDNA sequencing, one of the most sensitive and reliable methods to analyze this gene.²¹ Following a standardized phenotypic checklist (<http://www.genetics.uab.edu/medgenomics>), all patients in the UAB cohort had at least one diagnostic *NF1* feature, including a minimum of two hyperpigmented skin patches reminiscent of CALMs. In the MUI cohort, any child referred as a suspected sporadic *NF1* patient, regardless of whether a phenotypic checklist was completed, was included except if parental consanguinity or presence of *NF1* signs in a parent were reported. Therefore, the patients selected at UAB and MUI were analyzed separately.

In total, samples from 672 and 80 selected patients, sent for genetic analysis from November 2012 to April 2016 and from June 2009 to December 2018, were available from UAB and MUI, respectively. The UAB patients were divided by phenotype and age, and their samples were assigned a subgroup identifier before being anonymized so that basic clinical data would be available for identified CMMRD patients while maintaining patient anonymity (Table 1). The MUI patients were not divided by phenotypic subgroup and age prior to anonymization, due to the small sample size and lack of data for some patients (Table 1). Two samples were collected for each patient: gDNA extracted from PBLs, and RNA extracted from a short-term lymphocyte culture treated with the translation inhibitor puromycin to prevent nonsense mediated decay of transcripts prior to cell harvest.

MSI assay: marker amplification and sequencing

PBL gDNA samples were quantified using the QuBit dsDNA Broad Range Assay Kit (Invitrogen, ThermoFisher Scientific, Q32850, Waltham, MA USA). Twenty-four mononucleotide repeat microsatellite markers, and a single-nucleotide polymorphism (SNP) within 30 bp of each, were amplified from 200 ng of gDNA using the single molecule molecular inversion probe (smMIP)-based protocol of Gallon et al.,²⁵ adapted from Hiatt et al.²⁷ Amplicons were purified using Agencourt AMPure XP Beads (Beckman Coulter, A63881, Brea, CA, USA) following the manufacturer’s protocols. Purified amplicons were quantified using Qubit dsDNA High Sensitivity Assay Kit (Invitrogen, ThermoFisher Scientific, Q33231), and diluted to 4 nM using 10 mM Tris buffer at pH 8.5. Amplicons at 4 nM from up to 96 samples, with each batch including one CMMRD-positive and two CMMRD-negative control samples, were pooled to create each amplicon library. These were sequenced to a target read depth >4000× using the MiSeq platform (Illumina, San Diego, CA, USA) and custom sequencing primers,²⁷ following Illumina protocols.

Table 1 Demographics of the patient cohorts.

Distribution of <i>NF1</i> features: Phenotype:	UAB cohort			MUI cohort			Total
	Segmental			All phenotypes			
	1–7 years	8–16 years	>5 CALM +/– freckling only	1–7 years	8–16 years	0–17 years	
Number of patients	44	55	177	43	42	80	752
Median age in years (range)	5.0 (1.4–8.0)	12.1 (8.1–16.2)	11.0 (8.1–16.7)	4.6 (1.1–7.9)	12.3 (8.5–16.9)	4.7 (0.2–17.5)	6.2 (0.2–17.5)
Males:females	25:18 ^a	32:23	105:72	28:15	19:23	44:36	450:301 ^a

The 672 patients of UAB cohort were classified into three phenotypic subgroups that were each further subdivided into two age groups of 1–7 years and 8–16 years, respectively. Since a segmental distribution of neurofibromatosis type 1 (*NF1*) signs has frequently been reported for constitutional mismatch repair deficiency (CMMRD) patients, one phenotypic group consisted of patients with one or more *NF1* features restricted to one body part. Patients with a generalized distribution (i.e., found in multiple body areas) of *NF1* features were separated into two groups: one with patients with >5 CALMs, with or without freckling, and no other National Institutes of Health (NIH) feature; and one with patients with a generalized distribution of >5 CALMs, with or without freckling, with at least one other NIH feature. The relatively small number of 80 patients in the MUI cohort was not subdivided.

^aOne patient without recorded sex. CALM café-au-lait macule, MUI Medical University of Innsbruck, UAB University of Alabama–Birmingham.

MSI assay: sample scoring

Fastq files were aligned to the human reference genome build hg19 using BWA v0.7.17.²⁸ Aligned reads were analyzed as described by Gallon *et al.*²⁵ using R v3.5.3 and the *metap* R package. The *EnvStats* R package was used here instead of the *ExtDist* R package due to incompatibility with R v3.5.3. The method produces an MSI score for each sample: a higher score represents a higher probability that the sample has increased MSI relative to a reference set of controls.²⁵ There were several amendments to the scoring method: (1) the reference set contained 58 control samples sequenced at the Medical University of Innsbruck in addition to data from the original 40 control samples sequenced at Newcastle University, (2) markers suspected of being germline heterozygous were excluded from scoring by a different threshold, and (3) markers with low counts were excluded from sample scoring (Supplementary Information S2).

Quality control (QC) criteria were used to ensure MSI scores were reliable. These included (1) mean *smSequence* count across all markers ≥ 200 , (2) < 3 marker exclusions due to *smSequence* count < 100 , and (3) inspection of SNP allele distributions to check for evidence of reaction contamination (Supplementary Information S2). The SNPs sequenced with each microsatellite were used to confirm the identity of any sample repeats as there is only a 3.6×10^{-10} probability that any two individuals share the same genotype.²⁹

Germline MSI (gMSI) assay

gMSI analysis was performed according to the protocols described by Ingham *et al.*,²³ as adapted by Gallon *et al.*²⁵

Variant analysis of the MMR genes

The MMR and *EPCAM* genes were analyzed using the TruSight™ Cancer V2 Sequencing Panel (Illumina), following enrichment by Nextera® Rapid Capture from 50 ng of gDNA, according to manufacturer's protocols. Sequencing was performed on a MiSeq (Illumina) reaching an average read depth of 500×. Sequences were aligned to the human reference genome GRCh37 (hg19) and analyzed using SeqNext software (JSI, Ettlenheim, Germany). Copy-number variations (CNVs) were also assessed using the *panelcnMOPS* software package.³⁰ All variants with a frequency of more than 10% were evaluated. Direct *PMS2* and *MSH6* cDNA sequencing was performed according to previously developed protocols.²¹

Variants were evaluated according to the consensus recommendations of the American College of Medical Genetics and Genomics (ACMG)³¹ and the Mismatch Repair Gene Variant Classification Criteria v1.9 (InSiGHTgroup; https://www.insight-group.org/content/uploads/2017/05/2013-08_InSiGHT_VIC_v1.9.pdf). They are described in accordance with the Human Genome Variation Society guidelines (<http://www.hgvs.org/mutnomen>) using the reference sequences NM_000179.2 for *MSH6* and NM_000535.5 as well as NG_008466.1 for *PMS2*. The start codon A is used as position c.1.

Statistical analysis

Prevalence 95% confidence intervals (CIs) were calculated with the binomial Clopper–Pearson “exact” method using *Epitools* (<https://epitools.ausvet.com.au/ciproportion>).

RESULTS

MSI analysis of *NF1/SPRED1* PV negative patients detected seven potential CMMRD cases

PBL gDNA samples were available from 752 patients (UAB $n = 672$; MUI $n = 80$) who had clinical features indicative of sporadic NF1 but a negative genetic diagnosis, and no tumor incidence to suggest a differential diagnosis of CMMRD. The workflow and summary of the CMMRD screening are presented in Fig. 1. Detailed results for each patient can be found in Supplementary Table S1. Twelve patients from the UAB cohort were excluded from the study due to insufficient gDNA for screening. The remaining 740 patients were screened for CMMRD using the updated MSI assay and QC criteria (see “Materials and Methods”). Samples with reliable scores > 2.00 were classified as MSI-positive. Of the 740 patients included, 25 had an unreliable MSI score from the initial assay. Two of these had insufficient gDNA available to repeat MSI analysis, and a further three had an unreliable MSI score again in the repeat assay; these five patients were also excluded from the study. Therefore, 735 patients had a reliable MSI score to interpret.

Samples were amplified and sequenced in batches. One batch, including results for 75 patients, had an exceptional increase in MSI score: 27 samples had a score > 2.00 and the batch median score was 2.12 points higher than the median score across all other batches ($p < 10^{-15}$). There was no difference in the quality of sequencing between this batch and others. Unfortunately, there was insufficient gDNA remaining to repeat MSI analysis of these patients. To avoid excluding these patients and potentially alter the phenotype distribution of the anonymized cohort, the classification threshold for this batch was increased by 2.12 (the increase in median score) to MSI score > 4.12 (Supplementary Table S1).

In total, seven patients were classified as MSI-positive from either a single result or, where possible, from a result validated by a repeat assay to reduce false positives. This included UAB117 (scores = 30.08 and 29.0), UAB620 (scores = 4.15 and 4.16), UAB332 (score = 34.90), UAB146 (score = 8.47), UAB350 (score = 5.77), UAB399 (score = 4.27), and UAB410 (score = 4.99). Notably, the last five of these patients were analyzed exclusively on the batch classified by an alternative threshold of MSI score > 4.12 . These seven patients were taken forward for additional analyses to confirm a diagnosis of CMMRD, and the 728 patients with MSI negative PBL gDNA were classified as CMMRD-negative.

Germline variant analysis confirmed CMMRD in three cases identified by MSI analysis

We used gMSI as an initial validation of the MSI results and to stratify patients for germline variant analysis. gMSI uses dinucleotide repeat markers and is therefore insensitive to

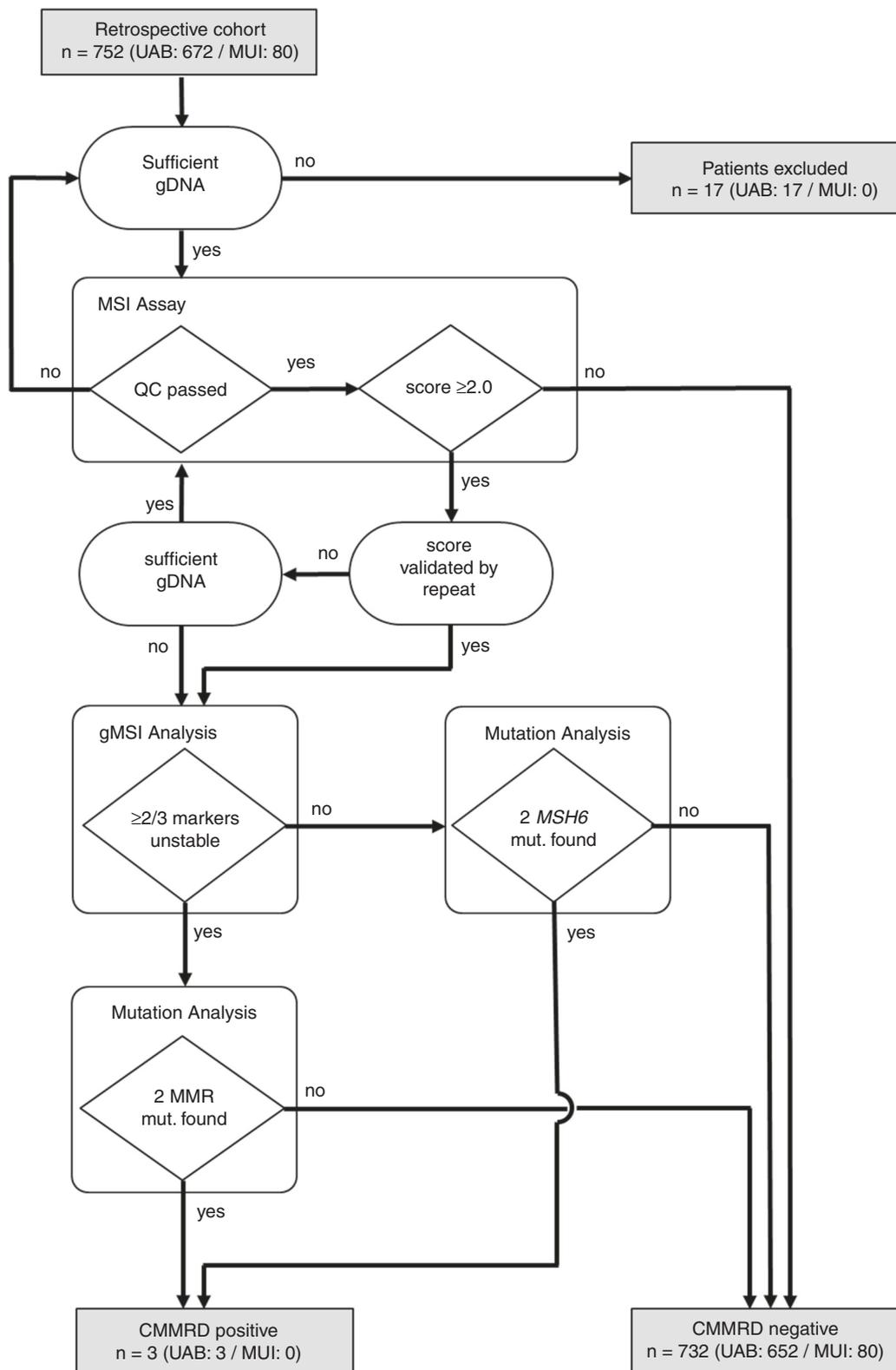


Fig. 1 Workflow of the constitutional mismatch repair deficiency (CMMRD) screening pipeline and summary of results. The number of patients and their outcomes are presented in gray boxes, assays used in the screening pipeline are presented in white boxes, criteria by which the assay results were assessed are presented in rhombi, and additional assessments are presented in ovals. *gDNA* genomic DNA, *gMSI* germline MSI, *MSI* microsatellite instability, *MUI* Medical University of Innsbruck, *QC* quality control, *UAB* University of Alabama–Birmingham.

MSI caused by *MSH6* deficiency, but has shown 100% sensitivity for MSI caused by *MLH1*, *MSH2*, or *PMS2* deficiency.^{23,25} Therefore, gMSI-negative patients had only *MSH6* analyzed. UAB117 and UAB332 were both gMSI-positive, whereas UAB146, UAB350, UAB399, UAB410, and UAB620 were gMSI-negative.

Variant analysis of all four MMR genes was performed in the gDNA sample from UAB117 by gene panel sequencing. Direct cDNA sequencing of *PMS2*,³² the most frequently mutated MMR gene in patients with CMMRD,^{1,3} was performed for UAB332 as only RNA was available following exhaustion of the gDNA during MSI analysis. For UAB620, gDNA was available for *MSH6* analysis by gene panel sequencing. For UAB146, UAB350, UAB399, and UAB410, direct cDNA sequencing of *MSH6* was performed as, like UAB332, only RNA was available.

In UAB117, a heterozygous deletion of the *PMS2* exon 10 (NG_008466.1[NM_000535.6]:c.[988+1_989-1]_[1144+1_1145-1]del), expected to cause loss of 52 amino acids p. (Glu330_Glu381del), and a heterozygous 1-basepair duplication in *PMS2* (NM_000535.6:c.1831dupA; p.Ile611Asnfs*2) were found. In sample UAB332, a complex frameshift PV (NM_000535.6:c.736_741delinsTGTGTGTGAAG; p.Pro246-Cysfs*3), as well as a C>T transition in a CpG site causing a premature stop codon (NM_000535.6:c.2404C>T; p.Arg802*), were identified in *PMS2*. In both UAB117 and UAB332, the MSI assay and gMSI results confirm the diagnosis. Hence, although not formally proven, it can be inferred that the *PMS2* PVs are located in *trans* and cause CMMRD in these patients. In UAB620, a homozygous nonsense variant in *MSH6* (NM_000179.2:c.10C>T; p.Gln4*) was identified, also confirming a diagnosis of CMMRD. Hemizyosity of this PV was excluded by CNV analysis. For UAB146, UAB350, UAB399, and UAB410 no *MSH6* PVs were identified and regular biallelic expression of this gene was evidenced in three of these patients (UAB146, UAB399, and UAB410) by expression of a heterozygous polymorphism (rs1042821). These four false positive results of the MSI assay were all patients analyzed exclusively in the one batch that had a general increase in assay score across all samples. All four patients were hence classified as CMMRD-negative.

CMMRD is the diagnosis in 0.41% of *NF1/SPRED1* PV negative patients

A diagnosis of CMMRD was confirmed in three of the 735 *NF1/SPRED1* PV negative patients screened in this study. This excludes 17 patients for whom the gDNA sample was insufficient for screening, or for whom the MSI analysis only produced unreliable results. Therefore, the prevalence of CMMRD in these *NF1/SPRED1* PV negative patients is 3/735 (0.41%, 95% CI: 0.08–1.19%). All three CMMRD patients were derived from the UAB cohort. Two (UAB332 and UAB620) were from the subgroup of patients aged 1–7 years with a generalized distribution of >5 CALMs, with or without freckling, and no other NIH NF1 feature. UAB117 was from

the subgroup of patients aged 8–16 years with a generalized distribution of at least two CALMs, with or without freckling, and at least one other NIH NF1 feature.

DISCUSSION

Recently, CMMRD has been recognized as an important differential diagnosis in malignancy-free children suspected of sporadic NF1 but lacking *NF1/SPRED1* PVs.^{18,19,33} The here empirically determined prevalence of 0.41% (3/735) is most likely the best approximation to the true prevalence that is currently achievable as UAB holds, to our knowledge, the largest collection of such systematically characterized patients globally. The empirically defined prevalence is almost identical to the previously calculated estimate of 0.39%. This suggests the assumptions used in these calculations and the C4CMMRD consensus guidelines for counseling and testing for CMMRD in these patients (for details see Suerink *et al.*¹⁸) are based on good approximations to true numbers. Specifically, given this low CMMRD prevalence, C4CMMRD proposes that only children with a higher probability of having CMMRD are counseled and tested so that the benefits outweigh the potential harms.¹⁸ The full anonymization of patient samples is a limitation of this study as the clinical phenotypes and family histories of the identified CMMRD patients cannot be assessed against these selection criteria. The three CMMRD patients identified were more or less proportionally distributed over the six phenotypic and age subgroups, but a greater number of such patients would be needed to draw conclusions on phenotype or age associations. Nonetheless, the empirical prevalence provided here can be used in a prospective benefit–risk assessment study of the C4CMMRD guidelines, and gives reliable numbers for counseling and informed decision making.

In this study we used MSI analysis of non-neoplastic tissues to screen a cohort of several hundred patients for CMMRD. In total, seven patient samples had an MSI score greater than the classification threshold. Of these, three were confirmed to have CMMRD by germline variant analysis, and four had no MMR gene PVs detected, equating to a positive predictive value of 42.9%. However, the MSI results for these four false positives were exclusively generated in one batch that had a score distribution distinct from all other batches. Further exploration of the exceptional score distribution of this batch is beyond the scope of this study. Batch effects may have implications for the assay's clinical deployment as a functional assay to support genetic testing. However, batch effects can be expected of any high throughput method and can be mitigated by inclusion of control samples in each batch (to help their identification) and repeat testing (precluded in the false positives in this study due to sample exhaustion). These are facilitated by the assay's low cost, at approximately \$15–55 per sample for reagents, and scalability, being fully automatable and able to analyze 7–174 samples on a single MiSeq flow cell (Supplementary Information S3).^{25,29} In further support of its clinical utility, we have previously validated the MSI assay's sensitivity and specificity using a blinded sample

cohort,²⁵ and, here, we have demonstrated that it can be deployed in a separate laboratory from the laboratory in which it was developed (including perfect concordance of sample classification between laboratories), shown that control sample scoring is equivalent between laboratories, updated it with automatically flagged QCs based on sequencing metrics, and expanded the reference set of controls to make it more robust (Supplementary Information S2). Therefore, while the MSI assay has not been formally accredited for clinical diagnostics, its performance in this study suggests it may not only be useful in research, but also in clinical settings as a prescreening assay on selected individuals or to resolve diagnostic uncertainties.^{18,25,29} Having shown the feasibility of large-scale MSI analysis to screen for CMMRD, the prevalence of CMMRD in other patient groups could be defined using this screening pipeline. This is particularly important given the broad phenotypic spectrum of CMMRD, including the wide variety of associated cancer types, and the clinical implications of its differential diagnosis with respect to clinical surveillance, therapy, and genetic counseling.⁶

As expected, the MMR PVs were found in *PMS2* and *MSH6*, which are mutated in >50% and approximately 20% of CMMRD cases, respectively,³ and all PVs have previously been identified in CMMRD and/or Lynch syndrome patients. Significantly, two of the five MMR PVs identified are known founder variants. The French Canadian *MSH6* founder variant c.10C>T (p.Gln4*),³⁴ homozygous in UAB620, is a C>T transition outside a CpG dinucleotide with a heterozygote frequency of 1 in 400 French Canadian newborns in the province of Quebec. One homozygous carrier, who developed colorectal cancer aged 10 years, has already been reported.³⁴ It is likely—albeit not proven by haplotype analysis—that both parents of UAB620 are carriers of this founder variant without necessarily being closely related. Consanguinity assessment was not performed to protect the anonymity of UAB620. The *PMS2* founder variant c.736_741delinsTGTGTGTGAAG of Scandinavian origin³⁵ was compound heterozygous with c.2404C>T (p.Arg802*) in UAB332. It is estimated that >10,000 individuals in the United States are heterozygous for c.736_741delinsTGTGTGTGAAG,³⁵ and it is the most frequent cause of Lynch syndrome in the Icelandic population, where 1 in 427 individuals carries it.³⁶ It has previously been found in two Dutch CMMRD patients, both compound heterozygotes,^{37,38} and in a third CMMRD patient who is homozygous for this variant.³⁹ PV c.2404C>T (p.Arg802*) has also been described as a founder variant in individuals from Pakistani origin living in England.⁴⁰ However, this C to T transition is likely a recurrent PV as it is found in a CpG dinucleotide and has also been observed in a CMMRD patient with a different ethnic origin (unpublished data). Identification of founder variants in two of three patients suggests that the prevalence of CMMRD in *NFI/SPRED1* negatives may be substantially higher in populations with (and potentially lower in populations without) founder effects. For instance, the calculated incidence of CMMRD due to the *MSH6* founder

variant is 1:640,000 in newborns to French Canadians from Quebec, 1.6 times higher than the estimated CMMRD incidence of 1:1,000,000 to unrelated parents. In Iceland, where c.736_741delinsTGTGTGTGAAG together with a second *PMS2* founder variant, c.2T>A, have a combined incidence of 1:307, the expected CMMRD incidence due to founder variants is 1:380,000, 2.6 times higher than current estimates. This should be considered in addition to C4CMMRD guidelines when counseling *NFI/SPRED1* PV negative children from populations with prevalent founder variants.

SUPPLEMENTARY INFORMATION

The online version of this article (<https://doi.org/10.1038/s41436-020-0925-z>) contains supplementary material, which is available to authorized users.

ACKNOWLEDGEMENTS

This work was funded by Austrian Science Funds (FWF; grant number KLI 734-B26).

DISCLOSURE

R.G., M.S.-K., M.S.J., and J.B. are named as inventors on patents held by Newcastle University, UK, covering the markers used in the MSI assay used in this study (patent ID: PCT/GB2017/052488, published 1 March 2018; and PCT application number: PCT/GB2019/052148, unpublished, filing date 31 July 2019). The other authors declare no conflicts of interest.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Wimmer K, Etzler J. Constitutional mismatch repair-deficiency syndrome: have we so far seen only the tip of an iceberg? *Hum Genet.* 2008;124:105–122.
2. Wimmer K, Rosenbaum T, Messiaen L. Connections between constitutional mismatch repair deficiency syndrome and neurofibromatosis type 1. *Clin Genet.* 2017;91:507–519.
3. Wimmer K, et al. Diagnostic criteria for constitutional mismatch repair deficiency syndrome: suggestions of the European consortium 'Care for CMMRD' (C4CMMRD). *J Med Genet.* 2014;51:355–365.
4. Durno C, et al. Recommendations on surveillance and management of biallelic mismatch repair deficiency (BMMRD) syndrome. *J Pediatr Gastroenterol Nutr.* 2017;64:836–843.
5. Tabori U, et al. Clinical management and tumor surveillance recommendations of inherited mismatch repair deficiency in childhood. *Clin Cancer Res.* 2017;23:e32–e37.
6. Vasen HFA, et al. Guidelines for surveillance of individuals with constitutional mismatch repair-deficiency proposed by the European Consortium 'Care for CMMRD' (C4CMMRD). *J Med Genet.* 2014;51:283–293.
7. Uusitalo E, et al. Incidence and mortality of neurofibromatosis: a total population study in Finland. *J Invest Dermatol.* 2015;135:904–906.
8. National Institutes of Health Consensus Development Conference statement: neurofibromatosis. Bethesda, Md., USA, July 13–15, 1987. *Neurofibromatosis.* 1988;1:172–178.
9. Messiaen LM, et al. Exhaustive mutation analysis of the *NF1* gene allows identification of 95% of mutations and reveals a high frequency of unusual splicing defects. *Hum Mutat.* 2000;15:541–555.
10. Messiaen LM, Wimmer K. *NF1* mutational spectrum. *Monogr Hum Genet.* 2008;16:63–77.
11. García-Romero MT, Parkin P, Lara-Corrales I. Mosaic neurofibromatosis type 1: a systematic review. *Pediatr Dermatol.* 2016;33:9–17.

12. Brems H, et al. Germline loss-of-function mutations in SPRED1 cause a neurofibromatosis 1-like phenotype. *Nat Genet.* 2007;39:1120–1126.
13. Messiaen L, et al. Clinical and mutational spectrum of neurofibromatosis type 1-like syndrome. *JAMA.* 2009;302:2111–2118.
14. Ricciardone MD, et al. Human MLH1 deficiency predisposes to hematological malignancy and neurofibromatosis type 1. *Cancer Res.* 1999;59:290–293.
15. Wang Q, et al. Neurofibromatosis and early onset of cancers in hMLH1-deficient children. *Cancer Res.* 1999;59:294–297.
16. Suerink M, et al. Constitutional mismatch repair deficiency in a healthy child: on the spot diagnosis? *Clin Genet.* 2018;93:134–137.
17. Urganci N, Genc DB, Kose G, Onal Z, Vidin OO. Colorectal cancer due to constitutional mismatch repair deficiency mimicking neurofibromatosis I. *Pediatrics.* 2015;136:1047–1050.
18. Suerink M, et al. Constitutional mismatch repair deficiency as a differential diagnosis of neurofibromatosis type 1: consensus guidelines for testing a child without malignancy. *J Med Genet.* 2019;56:53–62.
19. Evans DGR, et al. Cancer and central nervous system tumor surveillance in pediatric neurofibromatosis 1. *Clin Cancer Res.* 2017;23:e46–e53.
20. Mandelker D, et al. Navigating highly homologous genes in a molecular diagnostic setting: a resource for clinical next-generation sequencing. *Genet Med.* 2016;18:1282–1289.
21. Etzler J, et al. RNA-based mutation analysis identifies an unusual MSH6 splicing defect and circumvents PMS2 pseudogene interference. *Hum Mutat.* 2008;29:299–305.
22. Vaughn CP, et al. Clinical analysis of PMS2: mutation detection and avoidance of pseudogenes. *Hum Mutat.* 2010;31:588–593.
23. Ingham D, et al. Simple detection of germline microsatellite instability for diagnosis of constitutional mismatch repair cancer syndrome. *Hum Mutat.* 2013;34:847–852.
24. Bodo S, et al. Diagnosis of constitutional mismatch repair-deficiency syndrome based on microsatellite instability and lymphocyte tolerance to methylating agents. *Gastroenterology.* 2015;149:1017. e3.
25. Gallon R, et al. A sensitive and scalable microsatellite instability assay to diagnose constitutional mismatch repair deficiency by sequencing of peripheral blood leukocytes. *Hum Mutat.* 2019;40:649–655.
26. González-Acosta M, et al. High-sensitivity microsatellite instability assessment for the detection of mismatch repair defects in normal tissue of biallelic germline mismatch repair mutation carriers. *J Med Genet.* 2020;57:269–273.
27. Hiatt JB, Pritchard CC, Salipante SJ, O’Roak BJ, Shendure J. Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. *Genome Res.* 2013;23:843–854.
28. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2010;26:589–595.
29. Gallon R, et al. Sequencing-based microsatellite instability testing using as few as six markers for high-throughput clinical diagnostics. *Hum Mutat.* 2020;41:332–341.
30. Povysil G, et al. panelcn.MOPS: Copy-number detection in targeted NGS panel data for clinical diagnostics. *Hum Mutat.* 2017;38:889–897.
31. Richards S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015;17:405–424.
32. Wimmer K, Wernstedt A. PMS2 gene mutational analysis: direct cDNA sequencing to circumvent pseudogene interference. *Methods Mol Biol.* 2014;1167:289–302.
33. Giugliano T, et al. Clinical and genetic findings in children with neurofibromatosis type 1, legius syndrome, and other related neurocutaneous disorders. *Genes (Basel).* 2019;10:580.
34. Castellsagué E, et al. Characterization of a novel founder MSH6 mutation causing Lynch syndrome in the French Canadian population. *Clin Genet.* 2015;87:536–542.
35. Clendenning M, et al. A frame-shift mutation of PMS2 is a widespread cause of Lynch syndrome. *J Med Genet.* 2008;45:340–345.
36. Haraldsdottir S, et al. Comprehensive population-wide detection of Lynch syndrome in Iceland. *J Clin Oncol.* 2016;34:1542–1542.
37. Herkert JC, et al. Paediatric intestinal cancer and polyposis due to biallelic PMS2 mutations: case series, review and follow-up guidelines. *Eur J Cancer.* 2011;47:965–982.
38. Leenen C, et al. Pitfalls in molecular analysis for mismatch repair deficiency in a family with biallelic pms2 germline mutations. *Clin Genet.* 2011;80:558–565.
39. Monika, M et al. CMMRD diagnostics in three cases: verification of PMS2 variants in homozygous status and investigation of their pathogenicity by cDNA analyses. Abstracts of the 4th meeting of the European Hereditary Tumour Group, Barcelona, Spain, October 17–19th, 2019. *Fam Cancer.* 2020;9:77–121.
40. De Vos M, et al. PMS2 mutations in childhood cancer. *J Natl Cancer Inst.* 2006;98:358–361.



Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, and provide a link to the Creative Commons license. You do not have permission under this license to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2020