Clinical Study

Systematic detection of mosaicism by using digital NGS reveals 3 new MEN1 mosaicisms

Arnaud Lagarde¹, Grégory Mougel¹, Lucie Coppin², Magalie Haissaguerre³, Lauriane Le Collen⁴, Amira Mohamed⁶, Marc Klein⁷, Marie-Françoise Odou⁸, Antoine Tabarin³, Hedia Brixi¹⁰, Thomas Cuny¹¹, Brigitte Delemer⁴, Anne Barlier¹£, Pauline Romanet¹£

¹ Aix Marseille Univ, APHM, INSERM, MMG, Laboratory of Molecular Biology Hospital La Conception, Marseille, France; arnaud.lagarde@ap-hm.fr, gregory.mougel@ap-hm.fr, anne.barlier@univ-amu.fr, pauline.romanet@univ-amu.fr
² Univ. Lille, CNRS, Inserm, CHU Lille, UMR9020-U1277 - CANTHER - Cancer - Heterogeneity Plasticity and Resistance to Therapies, F-59000, Lille, France; lucie.verstraete@chru-lille.fr
³ Service d’Endocrinologie, Centre Hospitalier Universitaire, Hôpital du Haut Levêque, Pessac, France; magalie.haissaguerre@chu-bordeaux.fr, antoine.tabarin@chu-bordeaux
⁴ Endocrinology, Diabetology and Nutrition Unit, University Hospital of Reims, Reims, France; lle-collen@chu-reims.fr, bdelemer@chu-reims.fr
⁵ Inserm/CNRS UMR 1283/8199, Pasteur Institute of Lille, EGID, Lille, France
⁶ Laboratory of Molecular Biology, Hospital La Conception, APHM, Marseille; amira.mohamed@ap-hm.fr
⁷ Service endocrinologie, CHU de Nancy, hôpital de Brabois, Vandoeuvre-lès-Nancy, France; m.klein@chu-nancy.fr
⁸ CHU Lille, Service de Biochimie et Biologie moléculaire « Hormonologie, Métabolisme-Nutrition, Oncologie », Lille, France ; mf.paris@chu-lille.fr
⁹ Univ. Lille, Inserm, CHU Lille, U1286 – Infinite – Institute for Translational Research in Inflammation, F-59000 Lille, France
¹⁰ Department of Gastroenterology and Digestive Oncology, Reims University Hospital, Reims, France; hbrixi@chu-reims.fr
¹¹ Aix Marseille Univ, APHM, INSERM, MMG, Department of Endocrinology, Hospital La Conception, Marseille, France
£ These authors contributed equally to this work.

Correspondence:
Abstract:

Purpose: Mosaicism is a feature of several inherited tumor syndromes. Only few cases of mosaicism in Multiple Endocrine Neoplasia type 1 (MEN1) have been described. Next generation sequencing (NGS) offers new possibilities for detecting mosaicism. Here we report the first study to systematically look for MEN1 mosaicism, using blood DNA, in MEN1-suspected patients but without MEN1 pathogenic variants (PV) in a heterozygous state.

Methods: digital targeted NGS, including unique molecular identifiers (UMIs), was performed in routine practice and the analytic performance of this method was verified. Results: Among a cohort of 119 patients harboring from 2 to 5 MEN1 lesions, we identified 3 patients with MEN1 mosaic PVs. The allele frequencies ranged from 2.3 to 9.5%. The detection rate of MEN1 mosaicism in patients bearing at least 3 MEN1 lesions was 17% (3/18). No cases were detected in patients with 2 lesions. Conclusion: We report here 3 new cases with MEN1 mosaicism. This study examined the performance of UMI in the diagnosis of MEN1 mosaicism in routine
practice and our results underline that the frequency of mosaicism is probably underestimated in patients with suspected MEN1.

**Keywords:** Mosaicism; MEN1; NGS; unique molecular identifier; mosaic; deep-sequencing; de novo; digital NGS; thymic tumor; pituitary adenoma; hyperparathyroidism; pancreatic neuroendocrine tumor
INTRODUCTION

Multiple endocrine neoplasia type 1 (MEN1, OMIM 131100) is an autosomal dominant disease due to mutation in the MEN1 gene, characterized by a broad spectrum of clinical manifestations (1). The classic clinical triad includes primary hyperparathyroidism (PHPT), pituitary neuroendocrine tumors (PitNET), and duodeno-pancreatic neuroendocrine tumors (DPNET). Other endocrine tumors including adrenal cortical tumors, and neuroendocrine thymic or bronchopulmonary tumors may also be present. Several non-endocrine manifestations have also been associated with MEN1: facial angiofibromas, facial collagenomas, lipomas, and meningiomas. Twenty-eight to 70% of patients with MEN1 die as a consequence of the disease, particularly due to the pancreatic and carcinoid lesions (2, 3).

Classical genetic testing on blood DNA is positive in 90% to 95% of familial cases, and in 30-45% of sporadic cases who present with the classical triad (4, 5, 6). Mosaicism can explain some of these unresolved cases. Mosaicism has been described in several inherited tumor syndromes and corresponds to the spontaneous acquisition of a genetic variant during cell division during post-zygotic embryonic development (7, 8, 9, 10, 11). Mosaicism thus results in a fetus comprised of a variable proportion of mutated cells, depending on how early and in which cell lines the variant occurs. Mosaic variants may be undetectable in blood samples using classical sequencing methods. Only few cases of MEN1 mosaicism have been reported to date (12, 13, 14, 15, 16). Indeed, identification of MEN1 mosaicism remains challenging in routine practice of diagnosis laboratories and consequently is not systematically performed.

Next generation sequencing (NGS) offers new possibilities for detecting mosaic variants (7, 8, 14, 16, 17, 18). Here we set up targeted NGS using unique molecular identifiers (UMI, the UMI principle described in supplemental data) to systematically search for MEN1 mosaicism using
blood DNA in unresolved MEN1 index cases showing at least two lesions, and determined the performance of such analysis in a cohort of 119 patients.

MATERIALS AND METHODS

Next generation sequencing workflow

Genomic DNA was extracted from peripheral blood samples using QIAsymphony DSP DNA Midi Kit (Qiagen, Germany) according to the manufacturer’s protocols. Between 20 to 40 ng of genomic DNA was used to produce the library and perform target enrichment using the QIAseq Targeted DNA Custom Panel kit (Qiagen). The custom library included 62 kb of coding exons and 20 bp flanking regions of MEN1 (NM_130799) and 27 other genes involved in endocrine diseases (see supplemental data). This library used unique molecular identifiers (UMIs). UMIs are unique oligonucleotide sequences which are added to DNA prior to any amplification and differentially label each molecule in the native DNA fragment. UMIs are usually used for improving molecular detection of rare events in somatic DNA (19). Indeed, UMIs allow for a computational correction of amplification bias and sequencing errors by identifying PCR duplicates (see principle in supplemental data).

Quantitation and qualification of libraries was done using the Qubit Fluorometer (Thermo Fisher Scientific Inc., USA) and the TapeStation instrument (D100 ScreenTape, Qiagen) to enable equimolar pooling of barcoded samples. Twenty-four samples were sequenced during the same paired-end run (V2 2x150bp) on a MiSeqDx (Illumina, USA).

Alignment and variant calling were performed on the CLC Genomics WorkBench 20.0.4 (Qiagen) standalone analysis system against the Human genome reference GRCH37. The elements of optimization and validation of the variant calling for detection of mosaicism are
described in supplemental data. All variants that had an allele frequency (AF) greater than 0.5% were considered.

Validation and analytical performance of mosaicism detection using library preparation with UMIs (see also supplemental data)

Firstly, we analyzed the DNA from a patient who presented with a known mosaic MEN1 pathogenic variant, quantified at an AF of 5.1% using ultra-deep NGS (patient A, Supplemental Table 1) (14).

Secondly we created artificial mosaic variants (AMVs) to simulate mosaic MEN1 variants at different frequencies. For this, DNA from two patients (B and C) carrying MEN1 pathogenic variants in a heterozygous state were mixed at 20%, 10%, 4% and 2% with DNA from a wild-type sample with a known genotype (patient D). The samples were then sent to the molecular laboratory at the La Conception hospital for genetic testing. Patient B carried a heterozygous MEN1 pathogenic variant c.1546dupC, p.(Arg516Profs*15) which is located in a repeat region (homopolymer track with n=7), and patient C carried a heterozygous MEN1 pathogenic variant: c.1252G>A, p.(Asp418Asn)(4). The expected mutated AFs were respectively 10%, 5%, 2% and 1% in diluted samples. Variant calling using UMI groups was compared to variant calling using the same workflow but discarding the UMI group creating tool. The error rate of the method was taken as the number of false positives detected in a patient divided by the number of sequenced nucleotides and expressed as number of false positives / kb.

Cohort of unresolved MEN1 cases
This study was performed on patients referred between March 2018 and March 2021 for MEN1 genetic testing to the molecular laboratory of Marseille La Conception Hospital. Patients presented with at least two MEN1-related tumors but without MEN1 PV in heterozygous state. Written informed consent for genetic analysis was obtained from all patients during one-on-one genetic counselling. NGS data produced during routine practice processes were retrospectively realigned and reanalyzed using the pipeline optimized for mosaic detection. All variants with an allele frequency superior to 0.5% were analyzed. The study was approved by the ethics committee of Aix-Marseille University (approval number: 2018-13-12-004).

Confirmation of mosaicism

Putative mosaic variants responsive for MEN1 were confirmed by searching for the variant either in the MEN1 lesions when these were available or in a second blood sample from the patient. NGS or Sanger sequencing was performed on DNA from formalin-fixed paraffin embedded (FFPE) MEN1 lesions. FFPE-tissue DNA was extracted from samples using a QIAamp DNA FFPE tissue kit (Qiagen, Courtaboeuf, France). For Sanger sequencing, DNA was amplified using PCR targeting the identified MEN1 variation (primers available upon request) using the AmpliTaq Gold 360 Master Mix (ThermoFisher Scientific, Waltham, MA, USA). After ExoSap-IT purification (ThermoFisher Scientific), PCR products were sequenced using the Sanger method on an AB3500XL-DX genetic analyzer (ThermoFisher Scientific). NGS sequencing used the same method as described above.

Statistical methods
Data were compiled using R (R Core Team, 2020 (20)). Statistical analyses were performed using Prism software v9.0 (GraphPad Software, La Jolla, CA, USA). All results are expressed as median (range min – max values). The correspondence between the theoretical and observed values was evaluated by calculating the R\(^2\) of the Pearson correlation coefficient.

**RESULTS**

*Validation and analytical performance of MEN1 mosaicism detection using library preparation with UMIs* (see also supplementary data)

Firstly, we confirmed the presence of the known mosaic MEN1 variant in patient A, with an AF of 5.9% (vs. 5.1% by ultra-deep NGS (7, 14) Supplementary table 1).

Then, a total of 10 DNA samples were sequenced: 2 undiluted samples from patients B and C, plus 4 dilutions for each sample (Supplementary table 2). The sensitivity of the mosaic detection process was 100% with all MEN1 artificial mosaic variants (AMVs) being detected. The observed AF of the MEN1 AMVs correlated well with the expected AFs (correlation coefficient R\(^2\)=0.95, y=0.8255*x+0.3102, Figure 1A). Using the UMI specific bioinformatic process, false positive MEN1 variants were detected in only 3 out of the 10 samples (Figure 1B) and only at low AF. One was a single nucleotide variation at an AF of 0.7% (MEN1: c.787C>G, p.(Leu263Val)), two were the same deletion of one nucleotide in a homopolymer region at an AF of 1% (MEN1: c.1546delC, p.(Arg516Glyfs*43)), supported in both cases by only one UMI group.

Finally, at AF values greater than 0.5%, the error rate for MEN1 was only 0.27 per kb (Figure 1C), and at AF values greater than 1%, the error rate was zero, that is, no false positive MEN1 variants were found (Figure 1C). In contrast, without considering the UMIs in the
bioinformatic process, the median number of false positives in MEN1 was 37.5 per sample (min 20 – max 53) (Figure 1B). At an AF superior to 0.5% the error rate for MEN1 was 17 per kb (Figure 1C). These data validated the mosaic detection process using UMI for MEN1 (see also supplemental data for the rest of the 62 kb panel).

Cohort of unresolved MEN1 cases

One hundred and nineteen patients (40 males and 79 females) were included in the study. Their average age at the time of genetic testing was 54 years (range 17 to 86 years). Patients presented with at least 2 MEN1-associated lesions (Table 1). PHPT and PitNET were the most frequent lesions (80% and 67% of patients respectively). Sixty-five patients (54%) presented with both PHPT and PitNET.

The median coverage of depth for MEN1 sequencing was 584X (48 -2753X). In view of the results of the validation study, we restricted the analysis to variants supported by two UMI groups to avoid false positives. Among the 119 patients, 14 presented with variants at an AF greater than 0.5%. Eight patients harbored the MEN1 variant c.787C>G, previously identified as a false positive in our validation study, at an AF between 0.6% and 1.22% (mean 0.9%). Two patients harbored the c.655-6dupC, p.(?) MEN1 variant (AF 0.8% and 0.62%), one patient the c.655-5delC, p.(?) variant at an AF of 1.1%. These two variants were localized in a homopolymer region and were each supported by only one UMI-group in two other patients. These variants were present in the general population in a heterozygous state (gnomAD v2.1 last access 07/05/2022) and using SpliceAI were not predicted to alter splicing, they were thus not selected as putative mosaic variants. The last 3 variants were not found in other patients even supported by only one UMI group (Table 2). These 3 variants have been reported in the
literature, considered as pathogenic MEN1 variants and consequently were selected as putative mosaic variants (Table 2 patients #1, #2 and #3).

The AF values of the 3 putative mosaic variants ranged from 2.3 to 9.7%, with coverage of depth between 312 to 921X. The number of UMI groups supporting mosaicism ranged between 3 and 23. The three mosaic MEN1 variants were confirmed by a second method; the patient #1 variant was detected using NGS in somatic DNA from a parathyroid adenoma (AF: 78%) and from a thymic tumor (AF: 46%); the patient #2 variant was detected in somatic DNA from a thymic tumor using NGS (AF: 46%) and from a duodenal neuroendocrine tumor by Sanger sequencing; and the patient #3 variant was confirmed by Sanger sequencing using peripheral blood DNA and by NGS on a second blood sample.

Overall, no mosaic variant was detected in the 101 patients with 2 MEN1-lesions (0/101), whereas 2 mosaic variants were detected in 2 out of the 16 patients with 3 lesions (2/16, 12.5%; patient #1 and #3), and in 1 of the 2 patients having at least 4 lesions (1/2; patient #2).

DISCUSSION

The diagnosis of mosaicism remains an unmet medical need in routine practice of genetic laboratories (7, 8, 9, 10, 11). NGS offers the possibility of detecting mosaic variants using DNA from blood (7, 8, 14, 16, 17). The challenge is to lower the threshold of variant detection without including sequencing artifacts and also to distinguish variants from artifacts (17, 21, 22). UMIs are unique oligonucleotide sequences which are added to DNA prior to any amplification and these differentially label each molecule in the native DNA fragment. UMIs were initially developed to eliminate PCR duplicates in DNA or RNA seq in order to count the
absolute number of molecules (23). In somatic or germline DNA context, UMI can also be used to improve variant calling, by eliminating some artifacts coming from the amplification steps or from sequencing, thus improving the selection of true variants (24). This method is particularly appropriate for detecting variants at low AF. In specific regions, such as homopolymer tracks, UMI increases the sensitivity and specificity of variant detection. Thus, in our validation study we were able to detect a pathogenic mosaic variant in a homopolymer region at an AF as low as 1% (Supplementary table 2 and supplementary data), but found two false positive variants in the same region at an AF of 1%, supported by only one UMI group. In our retrospective study, in order to increase the specificity of analysis we considered variants supported by at least 2 UMIs groups. When compared to bioinformatics methods without using UMIs, the number of false positives was reduced by 98.4% for MEN1 and by 97.6% on the whole 62kb panel, at an AF greater than 0.5% (supplementary data). Our method, using UMIs, enables a high level of specificity and consequently rapid decisions made regarding further molecular explorations to confirm the MEN1 mosaicism, and thus improved management for both the patient and their family. In our study, we detected MEN1 mosaicism at a very low AF (2.3%), a threshold that has not been reported previously for this pathology. Due to the technical difficulties involved in setting up systematic mosaicism detection in routine practice of laboratories, the frequency of mosaicism is probably underestimated in most genetic diseases. We report here the first blood-based systematic analysis of MEN1 mosaicism in a cohort of clinically suspected MEN1 patients without heterozygous pathogenic variant. In our cohort of 119 patients with at least two-MEN1 related lesions, we found mosaicisms in 2.5% of patients, and notably only in those with 3 or more lesions (3/18, 17%), showing that searching for MEN1 mosaicism is required in such a context.
In neurofibromatosis type 2 (NF2), in which mosaicism is more frequent (25, 26), it has been proposed that two independent tumor samples are analyzed in each patient with clinically diagnosed NF2, based on the Manchester criteria (27). MEN1 is a lifelong disease in which the lesions may be widely spaced and earlier biopsy specimens are not always available. Moreover, some lesions may have benefited from first-line medical treatment or are not always systematically surgically removed. In our cohort, the rate of mosaicism was significantly lower than that reported in NF2, although 91 patients presented with clinically-diagnosed MEN1(28), based on the presence of at least two MEN1 primary lesions. No MEN1 mosaicism was identified in patients bearing only 2 MEN1-related lesions. This is in agreement with the low rate of heterozygous MEN1 mutations in such a population, ranging from 2 to 19% depending on the type of MEN1 lesions involved (5, 6). In the Dutch cohort, three quarters of MEN1 genetically-negative patients who had 2 MEN1-related lesions had a PTHP and a PitNET and they had developed their lesions at a later age than MEN1 genetically-positive patients (29). In our series of patients with 2 MEN1-related lesions, 54% had a PHPT and PitNET and the other patients had a mild phenotype, with one of the tumors occurring at over 50 years of age. In these conditions, a complementary strategy involving systematic analysis of multiple tissues seems to be difficult to justify, both medically and economically, in all patients with two MEN1-related lesions. Nevertheless, somatic analysis of MEN1-related lesions should not be forgotten in case of strong clinical suspicion and negative blood-based genetic testing, in particular in patients with multiple lesions including DPNET and thymic neuroendocrine tumor. Recently, a patient with a macroprolactinoma at 24 years of age, harboring a MEN1 mosaic variant at an allele frequency of 11%, has been described (30). This case shows that we should not limit the detection of mosaicism to unresolved cases with very
specific phenotypes. Our work highlights the need to optimize the sequencing process in diagnostic laboratories to allow detection of mosaicism on blood samples in routine practice.

If a mosaic variant is also present in germinal tissue, the variant can be transmitted to the offspring. Thus, undiagnosed mosaicism in someone planning children leads to the lack of adjusted genetic counselling and to a possible loss of chance for having children, particularly in the case of inherited cancer syndromes. Conversely, molecular diagnosis of mosaicism could have a strong benefit for patient care and also for genetic counseling. Family genetic investigation revealed no mutation in the 3 children of patient #2. The children of patient #1 refused genetic testing, and patient #3 has no children to date.

Of the 2 patients who had at least 4 MEN1 lesions, we failed to identify a MEN1 mosaic pathogenic variant in blood of one patient. This patient presented with a somatotroph adenoma at 46 years, a thymic tumor at 56 years, a PHPT and a left adrenal tumor at 57 years of age. CDKN1B analysis for this patient was also negative. Further analysis, such as whole genome sequencing, is required to identify variants in other regions, possibly as a deep intronic variant or a variant located in the promoter.

Regarding our entire panel of 62kb, during our validation study the median false positive per sample was 25.5 versus 1,073 per sample when UMIs were not taken into account, showing the interest of using UMIs for the genetic diagnosis of other syndromes, such as Neurofibromatosis type 1, von Hippel Lindau syndrome or tuberous sclerosis (see supplementary material).

CONCLUSIONS

We present here the first study to systematically search for MEN1 mosaicism using blood DNA in unresolved MEN1 cases in which we identified mosaicism in 17% of patients bearing 3
or more MEN1-related lesions. This study suggests that the frequency of MEN1 mosaicism may be underestimated and underlines the need to develop tools to detect it in routine practice, with UMI being potentially an accurate and powerful tool.

Acknowledgments: We thank all the patients and their physicians: Dr Albarel, Dr Alesandre, Dr Andrieu, Dr Barriquand, Dr Zatta, Dr Chevalier, Dr Corsini, Dr D’Anella, Dr Gall, Dr Grouthier, Dr Grulet, Dr Jelloul, Dr Julier, Dr Lagrave-Vergnet, Dr Lazoussnie, Dr Laveille, Dr Mallet, Dr Moreau, Dr Moutton, Dr Nassouri, Pr Niccoli, Dr Philippon, Dr Plas, Dr Priou Guesdon, Dr Puerto, Dr Radaoui, Dr Raingeard, Dr Morange, Dr Rigalleau, Dr Rochette, Dr Solmon, Dr Scheyer, Pr Taillard, Dr Tchknavorian, Dr Vermalle, Dr Vital. The authors thank Ian Darby from Editingbiomed for English revisions.

Author Informations: The authors declare no conflict of interest. Contributions:


Funding: All phases of this study were supported by grants from the Institut National de lutte contre le Cancer (INCa), the MarMaRa Institute, and the French Ministry of Health. The project resulting in this publication received funding from the Excellence Initiative of Aix Marseille
University -A*MideX- a French "Investissement d’Avenir" - Institute MarMaRa AMX-19-IET-007.

**Ethics declaration:** Informed consent was obtained from all subjects involved in the study. The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Ethics Committee of Aix Marseille Univ (ref 2018-13-12-004, date of approval: XII/14/2018).

**References**


22 53–59. (doi:10.1038/s41436-019-0598-7)


Diagnosis of mosaic mutations in the MEN1 gene by next generation sequencing.


Figure 1: Analytic performance of the method.

AF: allele frequency UMI: unique molecular identifier A. Linear regression between expected and observed AF of the artificial mosaic variants in MEN1. B. Comparison of the false positive variant count in MEN1 with and without taking into account UMIs. C. Error rate for variant calling in MEN1.
Figure 1: Analytic performance of the method.
AF: allele frequency UMI: unique molecular identifier A. Linear regression between expected and observed AF of the artificial mosaic variants in MEN1. B. Comparison of the false positive variant count in MEN1 with and without taking into account UMIs. C. Error rate for variant calling in MEN1.
Table 1. Clinical characteristics of the cohort of 119 unresolved MEN1 cases.

PHPT: primary hyperparathyroidism; PitNET: Pituitary neuroendocrine tumor; DPNET: duodeno-pancreatic neuroendocrine tumor; ADRE: adrenal tumor; THYM: thymic neuroendocrine tumor; LUNG: bronchial neuroendocrine tumors. *percentage among patients with 2 lesions

<table>
<thead>
<tr>
<th></th>
<th>PHPT</th>
<th>PitNET</th>
<th>DPNET</th>
<th>ADRE</th>
<th>THYM</th>
<th>LUNG</th>
<th>≥4 lesions</th>
<th>3 lesions</th>
<th>2 lesions</th>
<th>PHPT+ PitNET</th>
<th>PHPT+ DPNET</th>
<th>PitNET+ DPNET</th>
</tr>
</thead>
<tbody>
<tr>
<td>number of patients</td>
<td>95</td>
<td>80</td>
<td>29</td>
<td>28</td>
<td>10</td>
<td>4</td>
<td>2</td>
<td>16</td>
<td>101</td>
<td>54</td>
<td>17</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>(80%)</td>
<td>(67%)</td>
<td>(24%)</td>
<td>(24%)</td>
<td>(8%)</td>
<td>(3%)</td>
<td>(2%)</td>
<td>(13%)</td>
<td>(85%)</td>
<td>(54%*)</td>
<td>(17%*)</td>
<td>(3%*)</td>
</tr>
<tr>
<td>Mean age at diagnosis (years ; min-max)</td>
<td>57</td>
<td>53</td>
<td>55</td>
<td>58</td>
<td>52</td>
<td>56</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>(30-86)</td>
<td>(18-80)</td>
<td>(34-86)</td>
<td>(19-86)</td>
<td>(35-72)</td>
<td>(43-84)</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 2. Clinical and genetic characteristics of patients with MEN1 mosaicism in the cohort of unresolved MEN1 cases (patients #1, #2, #3).

PHPT: primary hyperparathyroidism, PitNET: Pituitary neuroendocrine tumor, DPNET: duodenopancreatic neuroendocrine tumor, THYM: neuroendocrine thymic tumor

<table>
<thead>
<tr>
<th></th>
<th>Patient #1</th>
<th>Patient #2</th>
<th>Patient #3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>Male</td>
<td>Male</td>
<td>Male</td>
</tr>
<tr>
<td>Age at molecular diagnosis (years)</td>
<td>54</td>
<td>60</td>
<td>31</td>
</tr>
<tr>
<td>Lesions (years)</td>
<td>PHPT (43)</td>
<td>PHPT (56)</td>
<td>PHPT (24)</td>
</tr>
<tr>
<td></td>
<td>THYM (43)</td>
<td>DPNET (56)</td>
<td>DPNET (24)</td>
</tr>
<tr>
<td></td>
<td>DPNET (53)</td>
<td>PitNET (56)</td>
<td>PitNET (27)</td>
</tr>
<tr>
<td>MEN1 mosaic pathogenic variant</td>
<td>exon 3 c.496=/C&gt;T p.(Gln166=/*)</td>
<td>intron 4 c.784-9=/G&gt;A p.(?)</td>
<td>exon 2 c.252=/dup p.(Ile85=/Tyrfs*32)</td>
</tr>
<tr>
<td>Allelic frequency</td>
<td>9.7%</td>
<td>2.3%</td>
<td>9.3%</td>
</tr>
<tr>
<td>Number of reads supporting the mutation (singleton/UMI group)</td>
<td>80 (57/23)</td>
<td>21 (18/3)</td>
<td>29 (23/6)</td>
</tr>
<tr>
<td>Proportion (singleton/UMIs)</td>
<td>0.71</td>
<td>0.86</td>
<td>0.79</td>
</tr>
<tr>
<td>QUAL</td>
<td>200</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td>Coverage of depth of the mutation</td>
<td>828X</td>
<td>921X</td>
<td>312X</td>
</tr>
<tr>
<td></td>
<td>Patient A</td>
<td></td>
<td></td>
</tr>
<tr>
<td>--------------------------------</td>
<td>----------------------------</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>Male</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age at molecular diagnosis (years)</td>
<td>41</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lesions (years)</td>
<td>PHPT (28)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>THYM (35)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PitNET (37)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Exon</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEN1 mosaic pathogenic variant</td>
<td>c.375_376=/delAT</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>p.(Ile125=/Metfs*54)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allele frequency</td>
<td>5.9%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of reads supporting the mutation (singleton/UMI group)</td>
<td>49 (31/18)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proportion (singleton/UMIs)</td>
<td>0.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Coverage of depth of the mutation</td>
<td>834X</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Supplementary Table 1.** Clinical and genetic characteristics of the patient with known mosaic MEN1 variant explored for method validation (patient A).

PHPT: primary hyperparathyroidism; PitNET: Pituitary neuroendocrine tumor; DPNET: duodenopancreatic neuroendocrine tumor; THYM: neuroendocrine thymic tumor
Supplementary Table 2. Validation study for the detection of MEN1 mosaicism

DNA from patient B & C were diluted with DNA of known genotype (not harboring MEN1 pathogenic variant) in order to create artificial mosaic variants at different allele frequencies.

Sample fraction is the percentage of DNA from patient B & C in the diluted samples. UMI: Unique molecular identifier
Supplementary material: validation and analytical performance of mosaicism detection using library preparation with UMIs

1/ PRINCIPLE OF LIBRARIES PREPARATION AND VARIANTS DETECTION USING UMIS.

A

1) DNA fragmentation, end repair and A-tailing
   Genomic DNA
   5’ A
   3) Target enrichment using single primer extension
   Universal Primer with adapter
   Forward Primer
   5’ A
   2) UMI & 5’ sample index ligation
   5’ A
   3) Target enrichment using single primer extension
   Universal Primer with adapter & 5’ sample index
   Target Primer
   5’ A
   4) Amplification

B

Classic library, reference

Library using UMIs

Supplementary Figure 1. Principle of libraries preparation and variants detection using UMIs (adapted from Qiagen documentation).

A: UMI-Library preparation. UMI: Unique Molecular Identifier. 1) Enzymatic fragmentation of DNA, end-repair and A-addition, 2) Ligation of an adapter containing a UMI, the 5’ sample
index, and a universal primer sequence, 3) Target enrichment using single primer extension with specific reverse primers and a universal forward primer, 4) Universal amplification and addition of the 3’ sample index.

B: Principle of variant detection using UMIs, for a theoretical example with coverage of depth at 12X. By using a classic library, the true variant (dark grey) is detected with an allele frequency of 25% (4/12 reads), and the false-positive variant (light grey) is detected at 17% (2/12 reads). By using the UMI-library, after removal of PCR duplicates by melting identical UMI reads within one UMI group, the final coverage of depth is 5X, and only the true heterozygous variant is detected, with an allele frequency of 20%. The false positive variant was excluded (variant is called if more than 75% of the reads within a UMI group carry them).

2/ MATERIAL AND METHODS

Next generation sequencing workflow

*MEN1* analysis was performed by NGS using a custom library including 62 kb of coding exons and 20 bp flanking regions of the following 27 genes involved in endocrine diseases: *AIP* (NM_003977), *AP2S1* (NM_004069), *CASR* (NM_000388), *CDC73* (NM_024529), *CDKN1B* (NM_004064), *EPAS1* (NM_001430), *FH* (NM_000143), *FLCN* (NM_144997), *GCM2* (NM_004752), *GNA11* (NM_002067), *GNAS* (NM_000516), *GPR101* (NM_054021), *MAX* (NM_002382), *MDH2* (NM_005918), *MEN1* (NM_130799), *MET* (NM_001127500), *NF1* (NM_001042492), *RET* (NM_020975), *SDHAF2* (NM_017841), *SDHB* (NM_003000), *SDHC* (NM_003001), *SDHD* (NM_003002), *SLC25A11* (NM_003562), *TMEM127* (NM_017849), *TSC1* (NM_000368), *TSC2* (NM_000548), *VHL* (NM_000551). The NGS workflow was the same than that used in the study.
Control sample and preparation of DNA samples harboring artificial mosaic variants

The performance of the detection of mosaicism was also evaluated on the entire 62 kb panel. Mosaic variants (AF < 30%) in genes other than MEN1 were analyzed on the DNA from the patients B and C diluted with wild-type DNA from patient D. The variant calling using UMI groups was compared to variant calling using the same workflow but discarding the UMI group creating tool. The error rate of the method was taken as the number of false positives detected in a patient divided by the number of sequenced nucleotides and expressed as number of false positives / kb.

Bioinformatics pipeline using UMI.

Raw sequencing data comes out of the MiseqDx sequencer as fastq files after demultiplexing. Alignment and variant calling were performed on the CLC Genomics WorkBench 20.0.4 (Qiagen) standalone analysis system. The Workflow Identify Qiaseq DNA variant panel analysis ready-to-use tool contains a step for exploiting the UMIs. As they are added to native DNA fragments before any amplification, UMIs allow for a computational correction of amplification bias and sequencing errors by identifying PCR duplicates (Figure 1B). Groups of reads with the same UMI are termed UMI groups. Reads with a unique UMI, not found on another read, are termed singletons. The principle of variant detection with UMIs is simple: all reads generated from the same original DNA fragment bear the same UMI and must carry the same variants. A variant is a true positive if it is present in all reads within a UMI group, whereas a variant is a sequencing artifact (termed false positive in this study) if it is not present in all reads within a UMI group. In practice, variants are considered as true positives and are called for variant calling if they belong to a singleton or if they are present in at least 75% of reads within a UMI group (Figure 1B).
Adjustment of the pipeline settings for detection of mosaicism.

In our clinical practice, the pipeline used as front line is set for constitutional targeted analysis and allows the detection of variants whose allele frequencies (AF) are either 30-70% (heterozygous variants) or superior to 90% (homozygous variants) (See table below).

For detection of mosaicism, quality filters were adjusted to access variants with a low AF up to a 0.5% threshold (Supplemental table 1). In particular, the specific quality score of the software was discarded and the detection of at least one UMI group per variant was imposed to validate the variant calling.

“Variant frequency” represents the number of reads supporting a variant divided by the total number of reads covering its position. “Count” represents the number of reads supporting the allele. “Probability” is calculated by the software by applying a multinomial statistical model where the reported probability is the sum of probabilities for all possible configurations. “Average quality” score is calculated by adding the Q scores of the nucleotides supporting the variant and dividing this sum by the number of nucleotides supporting the variant. “Quality” is calculated as \(-10\log_{10}(1-p)\), \(p\) being the probability that a particular variant exists in the sample. Quality is capped at 200 for \(p=1\). The proportion of singleton / UMI group represents the ratio of singletons to UMI groups. UMI: unique molecular identifier.

<table>
<thead>
<tr>
<th></th>
<th>≥ 5%</th>
<th>≥ 0.5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variant frequency</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Count</td>
<td>≥ 2</td>
<td>≥ 5</td>
</tr>
<tr>
<td>Probability</td>
<td>not filtered</td>
<td>≥ 0.8</td>
</tr>
<tr>
<td>average quality</td>
<td>≥ 30</td>
<td>≥ 30</td>
</tr>
<tr>
<td>Quality</td>
<td>200</td>
<td>not filtered</td>
</tr>
<tr>
<td>proportion singleton / UMI group</td>
<td>not filtered</td>
<td>≤ 0.9</td>
</tr>
</tbody>
</table>

3/ RESULTS

Validation and analytical performance of UMI-libraries for the whole gene panel (62kb)

Using the 10 DNA samples, bearing variant at known AF (2 undiluted samples from patients B and C, plus 4 dilutions for each sample), the analytic performance of the mosaic process was also analyzed on the 62kb whole gene panel base on several heterozygous single
nucleotide variants (SNV), present in the other genes than MEN1, 18 in the DNA from patients B and 13 in the patient C. These 31 (18+ 13) heterozygous SNVs were used as AMVs. All were seen in the 3 diluted samples showing the expected AF values of 10, 5 and 2%, reflecting a sensitivity of 100%. The observed AF values were close to those expected (correlation coefficient $R^2=0.91$, $y=0.9132*x+0.3777$, Supplemental Figure 2A) In the diluted samples corresponding to the expected AF of 1%, 24/31 (77.4%) AMVs were detected. This detection rate was slightly higher in the sample from patient C (11/13, 84.6%) compared to patient B (13/18, 72.2%). At the frequency of 2%, all were detected (Supplemental figure 2A, below). This difference was not explained by a better coverage of depth in patient C (Supplemental table 2 ).

Apart from AMVs, other variations with an AF greater than 0.5% were observed in the diluted samples in the 62kb panel. These were considered as false positives since they were not present in the original sample. The median number of these false positive variants per sample was 25.5 (min 22 – max 44) using UMIs (Supplemental figure 2B). The error rate was calculated at 0.56 per kb at AF greater than 0.5% (Supplemental figure 2C). The median AF of these false positive variants was 1.1% (0.5-8.7%). All false positives found at an AF greater than 2% concerned repeated regions or regions with high homologies leading to alignment errors. In the same conditions but without taking into account UMIs, the median false positive was 1,073 per sample (vs. 25.5 with UMIs) giving an error rate of 17 per kb (vs. 0.56 per kb with UMIs).
### Supplemental figure 2. Analytic performance of the method for the 62 kb panel.

AF: allele frequency  
UMI: unique molecular identifier

A. Linear regression between expected and observed AF of the artificial mosaic variants in all genes of the 62kb panel.

B. Comparison of the false positive variant count with and without taking into account UMIs through the 62 kb panel.

C. Error rate for variant calling in the 62 kb panel.

### 4/ REFERENCES


