

ARTICLE

The genomic architecture of *NLRP7* is Alu rich and predisposes to disease-associated large deletions

This article has been amended since online publication. A corrigendum also appears in this issue

Ramesh Reddy^{1,2}, Ngoc MP Nguyen^{1,2,15}, Guillaume Sarrabay^{3,15}, Maryam Rezaei⁴, Mayra CG Rivas⁵, Aysenur Kavasoglu⁶, Hakan Berkil⁷, Alaa Elshafey⁸, Ebtesam Abdalla⁹, Kristin P Nunez⁹, H el ene Dreyfus¹⁰, Merviel Philippe¹¹, Zahra Hadipour¹², Asude Durmaz⁶, Erin E Eaton¹³, Brittany Schubert⁹, Volkan Ulker¹⁴, Fatemeh Hadipour¹², Fatemeh Ahmadpour⁴, Isabelle Touitou³, Majid Fardaei⁴ and Rima Slim^{*,1,2}

NLRP7 is a major gene responsible for recurrent hydatidiform moles. Here, we report 11 novel *NLRP7* protein truncating variants, of which five deletions of more than 1-kb. We analyzed the transcriptional consequences of four variants. We demonstrate that one large homozygous deletion removes *NLRP7* transcription start site and results in the complete absence of its transcripts in a patient in good health besides her reproductive problem. This observation strengthens existing data on the requirement of *NLRP7* only for female reproduction. We show that two other variants affecting the splice acceptor of exon 6 lead to its in-frame skipping while another variant affecting the splice donor site of exon 9 leads to an in-frame insertion of 54 amino acids. Our characterization of the deletion breakpoints demonstrated that most of the breakpoints occurred within Alu repeats and the deletions were most likely mediated by microhomology events. Our data define a hotspot of Alu instability and deletions in intron 5 with six different breakpoints and rearrangements. Analysis of *NLRP7* genomic sequences for repetitive elements demonstrated that Alu repeats represent 48% of its intronic sequences and these repeats seem to have been inserted into the common *NLRP2/7* primate ancestor before its duplication into two genes.

European Journal of Human Genetics (2016) 24, 1445–1452; doi:10.1038/ejhg.2016.9; published online 9 March 2016

INTRODUCTION

Hydatidiform mole (HM) is an aberrant human pregnancy characterized by abnormal embryonic development, hydropic degeneration of chorionic villi and excessive trophoblastic proliferation. Hydatidiform moles affect 1 in 600 pregnancies in western countries,¹ but have higher frequencies in Southern and Asian countries.² These moles are mostly sporadic and not recurrent. Recurrent hydatidiform moles (RHMs) are defined by the occurrence of at least two moles in the same patient and affects 1.5–9.3% of women with a prior HM.^{3–7}

To date, two genes, *NLRP7* and *KHDC3L*, responsible for RHMs have been identified.^{8,9} Studies by several groups and on various populations demonstrated that *NLRP7* is a major gene for RHMs. *NLRP7* pathogenic variants are found in 48–80% of patients with RHMs depending on populations and patients inclusion criteria.^{10–15} To date, 48 distinct pathogenic variants observed in recessive state in *NLRP7* have been reported in a total of 131 patients.¹⁶ *NLRP7* codes for a NOD-like receptor pyrin containing protein 7 that plays roles in inflammatory response,^{17,18} trophoblastic tissue differentiation^{19,20} and proliferation,²⁰ and is part of the oocyte cortical cytoskeleton.²¹

KHDC3L is a second gene responsible for RHMs and is mutated in 14% of patients who are negative for *NLRP7* mutations.^{9,22} So far, four different pathogenic variants in *KHDC3L* have been reported in a total of six patients. *KHDC3L* codes for a KH-domain containing protein, of which another member, *KHDC1B*, has been shown to bind mRNA and is believed to play a role in the regulation of their translation during oocyte maturation.²³

The closest *NLRP* gene to *NLRP7* is *NLRP2*, which is located 25-kb distal to *NLRP7* on human chromosome 19q13.4 in a head-to-head orientation. *NLRP7* does not have rodent or bovine orthologues and is believed to have originated from a common *NLRP2/7* ancestor in primates.²⁴

In this study, we report 11 novel protein-truncating variants in *NLRP7*, of which five are large deletions. We characterized the effects of four variants at the transcriptional level. We demonstrate that six of the large deletion breakpoints occurred in intron 5 and eight of the nine large deletions reported so far in *NLRP7* may have been mediated by Alu repeats and microhomology of 16-bp to 44-bp. Our analysis demonstrates that human *NLRP7* is highly rich in Alu

¹Department of Human Genetics, McGill University Health Centre Research Institute, Montr eal, Canada; ²Department of Obstetrics and Gynaecology, McGill University Health Centre Research Institute, Montr eal, Canada; ³Laboratoire de G en etique des Maladies Rares et Autoinflammatoires, H opital Arnaud de Villeneuve, Montpellier, France; ⁴Genetics Department, Shiraz University of Medical Sciences, Shiraz, Iran; ⁵Hospital Aranda de la Parra, Laboratorio-s otano, Hidalgo, Centro Leon Gto, Mexico; ⁶Ege University Medical Faculty, Department of Medical Genetics, Izmir, Turkey; ⁷Genetik Genetic Laboratory, Istanbul, Turkey; ⁸Kuwait Medical Genetic Center, Kuwait Pediatric Department, Kuwait City, Kuwait; ⁹Division of Maternal-Fetal Medicine, Comprehensive Hemostasis & Thrombosis Center, Duke University Medical Centre, Durham, NC, USA; ¹⁰Institut Ste Catherine, Avignon, France; ¹¹Gynecologic and Obstetrics, Department University Hospital Amiens, Amiens, France; ¹²Medical Genetics Department, Sarem Fertility & Infertility Research Center (SAFIR), Tehran, Iran; ¹³Department of Obstetrics and Gynecology, Division of Maternal Fetal Medicine, The University of North Carolina, Chapel Hill, USA; ¹⁴Department of Obstetrics and Gynaecology, Oncology Unit, Kanuni Sultan Suleyman Training and Research Hospital, Istanbul, Turkey

*Correspondence: Dr R Slim, Department of Human Genetics, McGill University Health Centre Research Institute, Glen site, EM03210 (office) E012379 (lab), 1001 D ecarie Blvd, Montr eal, QC H4A 3J1, Canada. Tel: +1 514 934 1934 ext 44550 (office) or 44519 (lab); Fax: +1 514 933 4149; E-mail: rima.slim@muhc.mcgill.ca

¹⁵These authors contributed equally to this work.

Received 22 September 2015; revised 28 December 2015; accepted 14 January 2016; published online 9 March 2016

Table 1 Recapitulation of the 11 identified novel protein truncating variants and the reproductive outcomes of the patients

Family	ID	Patient	Origin/Ethnicity	Exon/Intron	cDNA	Variants	Predicted protein/RNA data	Reproductive history and gynecological morbidities
MoMx436	1243		Mexican	Exon 6/intron 9	c.[2248C>G]; [2810+2T>G]		p.[Leu750Val]; p.[Arg937_Leu938Ins54]	PHM-GTN, 2 SA, SA(46,XX), 6 SA, PHM
F5108	6152		Egyptian	Exon 4	c.[1374_1375delAG]; [1374_1375delAG]		p.[Arg458Serfs*69]; [Arg458Serfs*69]	3 PHM; several failed ART
F5587	6477		French, Caucasian	Exon 4/6	c.[1908dup]; [2161C>T]		p.[Glu637*]; [Arg721Trp]	SA, 2 HM, CHM-GTN
F4915	5729		French, Caucasian	Exon4/intron 10	c.[939_952dup14]; [2982-2A>G]		p.[Tyr318Cysfs*7]; [?]	BO, 3 CHM
F5669	6576		Persian	Exon 9	c.[2759G>A]; [2759G>A]		p.[Trp920*]; [Trp920*]	2 HM, 3 failed ART, SA
F5007	5868		Egyptian	Exon 9	c.[2777T>G]; [2777T>G]		p.[Leu926*]; [Leu926*]	SA, BO, 4 CHM
MoUs457	5869 (sister)		American, Caucasian	Intron 5	c.[2130-6_2132del]; [=]		r.[2130_2299del169]; [=]	SA, 4 CHM
F5303	6202		Syrian	Intron 5 & 6	c.[2130-266_2300+782del]; [2130-266_2300+782del]		p.?	SA(69,XXX) ⁻ , SA(trisomy 9) ⁻ , SA(45,X), SA(69,XXY, Digynic) 4 HM, ICSI-HM, ICSI-HM
MoTu497	1341	sister	Turkish	Before exon 1 & intron 5	c.[- 39-231_2130-510del]; [-39-231_2130-510del]		r.0	2 HM HM, SA, HM, PHM, SA, HM, PHM
MoTu546	1424		Turkish	Intron 1 & intron 5/exon 8	c.[40+251_2130-681del]; [2571dup]		p.[?]; [Ile858Hisfs*11]	HM, LB, HM, LB
	1426 (sister 1)							2 SA, HM, HM (Dip Bip), SA, HM
	1428 (sister 2)							HM, HM (Dip Bip) HM (Dip Bip XY ^e)
MoI256	P120		Iranian	Intron 1 & intron 5	c.[3998_2130-668del]; [-3998_2130-668del]		p.?	3 HM, HM-GTN, HM
MoMx505	1359		Mexican	Before exon 1 /exon 6	c.[13413_2982-344del]; [2248C>G]		p.[?]; [Leu750Val]	3 HM 3 HM, HM (twins by ultrasound)

Novel variants are in bold font. HM, stands for hydatidiform mole and is used when no tissues were available to re-evaluate the diagnosis and available pathology report does not distinguish between partial and complete HM; PHM, for partial HM; CHM, for complete HM; LB, for live birth; SA, for spontaneous abortion; GTN, for gestational trophoblastic neoplasia; ART, for assisted reproductive technologies; ICSI, for intracytoplasmic sperm injection; BO, for blighted ovum. Reproductive outcomes are listed by chronological order starting from the left; the absence of a number indicates one such reproductive outcome. '*' indicates conceptions after clomide treatment. The positions of different mutations in exons or introns are separated by '/'. The positions of the two breakpoints of the same mutation are indicated by '&'.
^eIndicates the POC genotyped by our group. Variants nomenclature is given according to the following references, NM_001127255.1 (cDNA), NG_008056.1 (genomic), and NP_001120727.1 (protein).

repeats, which represent half of its intronic sequences, that may have accumulated in the primate ancestor of *NLRP2/7* before its duplication into two distinct genes, *NLRP2* and *NLRP7*.

MATERIALS AND METHODS

Patients

Patients included in this study were referred to our laboratories at the Research Institute of the McGill University Health Centre, the 'Laboratoire de génétique des maladies rares et autoinflammatoires', or the Genetics Department of the Shiraz University of Medical Sciences, for *NLRP7* mutation analysis. All the patients included in the mutation analysis were new with the exception of patient 1424 and her two affected sisters, 1426 and 1428, who were previously reported as being heterozygous for a 60-kb deletion encompassing portions of *NLRP7* and *NLRP2*.²⁵ In the transcriptional analysis of the consequences of the variants, two other patients, 1200 and 1074, whose mutations were previously reported, were included.²⁰ The reproductive history of the new patients recapitulated in Table 1. The products of conception (POCs) of the patients listed in Table 1 were not available for analysis with the exception of one molar tissue from patient 1428 that was genotyped by our group and found diploid biparental confirming therefore its previously reported genotype.²⁵ All patients had RHMs with the exception of patient 1291 who had four spontaneous abortions (SAs), of which two were found triploid by karyotype analysis. Her products of conception were not available for us for histopathological review and genotyping, but were analyzed as part of the patient clinical care. One of her triploid products of conception was genotyped by microsatellite DNA markers and reported to be digynic. Her two other SAs were found diploid, one with trisomy 9 and one with monosomy X (Table 1).

Mutation analysis and characterization of the deletions

NLRP7 mutation analysis was performed as previously described by PCR amplification of the 11 exons of *NLRP7* from genomic DNA followed by Sanger sequencing of the PCR products in the two directions.⁸ Various other strategies were performed to identify the exact variants in patients with suspected deletions. These strategies included the design of additional primers for various genomic regions, genotyping of *NLRP7* fragments containing known single nucleotide polymorphisms (SNPs) in all family members and establishment of the haplotypes, regular and long range PCR, and genomic DNA dosage by quantitative real-time PCR. The primer pairs that allowed the amplification of the junction fragment for each of the described large deletions are provided in Supplementary Table S1. Variants are described with reference to the following transcript, NM_001127255.1, protein NP_001120727.1 and genomic sequence NG_008056.1. Variant numbering starts with the initiation codon. All variants detected have been deposited in the Leiden Open Variation Database (<http://databases.lovd.nl/shared/genes/NLRP7>) (patient IDs 00054739-00054741, 00054760-00054768 and 00054773-00054779) and INFEVERS (<http://fmf.igh.cnrs.fr/ISSAID/infevers/>). Exons are numbered as previously described in Messaed *et al.*²⁶

Quantitative real time PCR

Quantitative real time PCR to define the deletions in patients 1359 and 1424 was performed using Quantifast SYBR-green PCR kit (Qiagen, Toronto, ON, Canada) on 25-ng of genomic DNA. Each sample was tested in duplicates using the Bio-Rad MiniOpticon Real Time PCR system (Bio-Rad Laboratories, Mississauga, ON, Canada) and analyzed using the Opticon Monitor software (Bio-Rad Laboratories). The comparative C_T method ($\Delta\Delta C_T$ method) was used for relative quantification, and data was normalized against an endogenous control sequence, exon 11 of *NLRP7* in patient 1359 and exon 8 of *NLRP7* in patient 1424, both of which have two normal copies documented by their heterozygous status for SNPs within the amplicons that amplify these exons. Melting curve analysis was also done to verify PCR specificity.

Reverse transcription PCR

Reverse transcription followed by PCR (RT-PCR) was performed on RNA extracted from Epstein-Barr virus (EBV) transformed cells of patients 1074,

1200, 1243, 1291 and 1341 using Trizol (Invitrogen, Carlsbad, CA, USA) and verified by electrophoresis. cDNA was synthesized using a reverse transcription kit (Life Technologies, Thermo Scientific, Carlsbad, CA, USA). PCR after reverse transcription was performed using primers located in exons 5 and 7 of *NLRP7* for patients 1291 and 1200, primers from exons 9 and 11 for patients 1074 and 1243 and various primers located in exons 5 to 11 for patient 1341. The amplified fragments were separated on 1.5% agarose gel, purified using QIAquick PCR purification kit (Qiagen, Toronto, ON, Canada) and sequenced using Sanger sequencing. To document the absence of *NLRP7* transcripts in patient 1341, primers from exons 5 and 7 of *ZNF28* gene were used in combination with the various *NLRP7* primers in the same reaction as technical positive control for the quality and quantity of RNA and the success of the RT-PCR reactions. The various primers used in RT-PCR are provided in Supplementary Table S2.

RESULTS

Identification of 11 novel protein truncating variants including five large deletions

NLRP7 mutation analysis was performed by PCR amplification of all exons from genomic DNA of 12 unrelated patients and direct sequencing of the PCR products. The identified variants are listed in Table 1 and some of them are shown in (Supplementary Figure S1). These variants include two previously reported missense variants, a previously reported splice variant, a new splice variant, two novel single nucleotide substitutions leading to stop codons, two small duplications (one previously reported and one novel), two small deletions, and five large deletions. All patients were either homozygous for one, or compound heterozygous for two, potentially pathogenic variants with the exception of one patient, 1291, who had a small deletion in a heterozygous state and no other variants that are believed to be pathogenic. The variant found in this patient deletes six nucleotides from intron 5 and the first three nucleotides of exon 6, c.2130-6_2132del.

In patient 6202, regular PCR primers designed to amplify exon 6 did not yield any DNA amplification and suggested the presence of a homozygous deletion. Other primers were designed and allowed the amplification of a fragment overlapping the junction (Supplementary Table S1). The sequencing of this fragment revealed the presence of a 1219-bp deletion, c.2130-266_2300+782del, in a homozygous state.

Similarly, in patients P120 and her affected sister (P140), no amplifications were obtained with primers designed to amplify exons 1–5. A series of primers, mapping around the suspected deleted region were used in PCR to amplify the junction fragment that was sequenced (Supplementary Table S1) and the analysis of its sequence defined the deletion at the nucleotide level as c.-3998_2130-668del.

In patient 1341, no PCR fragments were amplified with the primers for exons 2–6. We designed new primers and performed primer walking using regular and long range PCR amplification on the genomic DNA of the patient. This resulted in the amplification of an ~1-kb fragment only in the patient, but not in controls (Supplementary Table 1 and Supplementary Figure S2). Sequencing this fragment and the analysis of its sequence revealed a 5041-bp deletion extending from introns 1 to 5 and defined the deletion at the nucleotide level to c.-39-231_2130-510del. This abnormal 1-kb fragment overlapping the deletion was also amplified from DNA of the two parents of the patient. We note that the size of the normal fragment that does not contain the deletion is ~6-kb and was not amplified on DNA from the parents of the patient or the controls under our experimental conditions (Supplementary Figure S2).

Patients 1424, 1426 and 1428 are sisters and were previously reported as carriers of a 60-kb deletion eliminating substantial portions of *NLRP2* and *NLRP7* in a heterozygous state and with

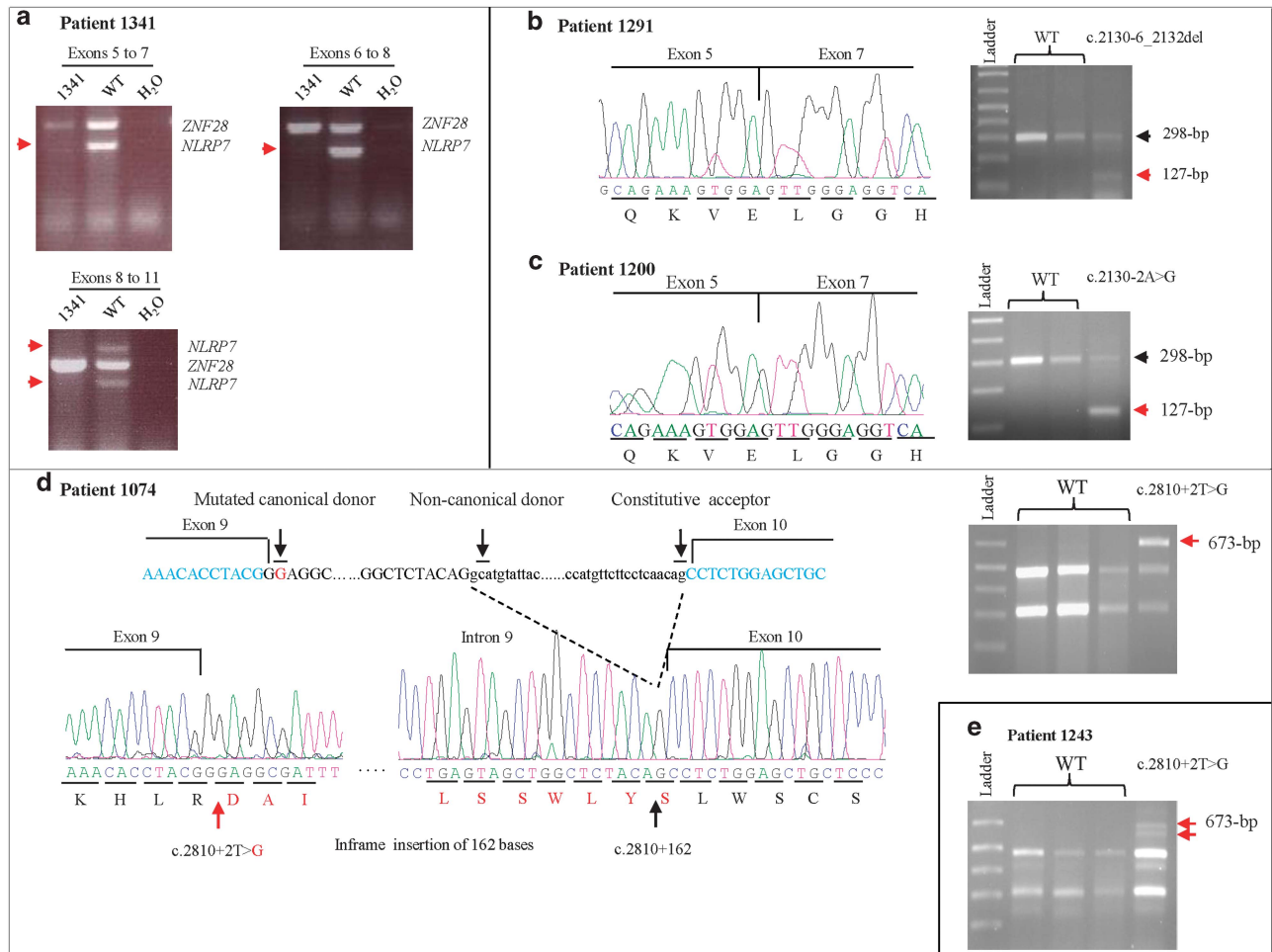


Figure 1 Transcriptional characterization of four *NLRP7* variants using RT-PCR on RNA from EBV-transformed patient cells. Red arrows denote missing or abnormal fragments observed only in the patients. (a) RT-PCR on RNA from patient 1341 showing complete loss of *NLRP7* transcripts using various combinations of primers located in exons 5–11. Primers from exons 5 and 7 of *ZNF28* gene were used in combination with those of *NLRP7* as positive controls for the RT-PCR. (b) RT-PCR on RNA from patient 1291 with a 9-bp deletion removing the invariant splice acceptor sites of exon 6 showing the deletion of exon 6 from the cDNA. (c) RT-PCR on RNA from patient 1200 with an invariant splice acceptor site variant showing the loss of exon 6. (d) RT-PCR on RNA from patients 1074 and 1243 (e) with the same variant affecting the invariant splice donor of exon 9. This variant leads to one common abnormal fragment of 673-bp in the two patients resulting from the insertion of 162-bp of intron 9 between exons 9 and 10. In patient 1243, we observed other faint fragments due to minor aberrant splicing, of them a fragment smaller than 673-bp is shown in this figure and is indicated by a red arrow.

no other *NLRP7* pathogenic variants.²⁵ *NLRP7* sequencing on DNA from patient 1424 revealed a previously reported mutation, c.2571dup, p.(Ile858Hisfs*11) in a heterozygous state. To define accurately the deletion, we sequenced *NLRP7* amplicons containing common SNPs in all family members, genotyped additional SNPs in the family, established the haplotypes and used quantitative real-time PCR to define the 5' breakpoint of the deletion close to *NLRP7* exon 1 (Supplementary Figure S2) and the 3' breakpoint close to *NLRP7* intron 5. Using primers flanking the deletion, we amplified a fragment of ~1.6-kb in the patient but not in controls (Supplementary Table S1). Sanger sequencing defined the deletion at the nucleotide level to c.-40+251_2130-681del (Table 1). The two protein truncating variants are most likely pathogenic and segregated from the two parents to the three affected sisters.

Patient 1359 was found homozygous for a previously reported missense mutation, p.(Leu750Val) that was present only in her father in a heterozygous state, but not in her mother. This raised the

possibility that the patient is hemizygous for the fragment containing p.(Leu750Val). We then performed a high density SNP microarray using Cytoscan HD (Affymetrix, Santa Carla, CA, USA), which suggested that the patient is heterozygous for a deletion spanning from introns 5 to 10 (Supplementary Figure S2). Sequencing of *NLRP7* amplicons containing common SNPs in all family members, as well as genotyping additional SNPs from the region, haplotype establishment, and genomic DNA dosage analysis using quantitative real-time PCR using various genomic primers defined the 5' breakpoint to ~13-kb upstream of exon 1 and the 3' breakpoint to intron 10. We then amplified the junction fragment in the patient, sequenced it and defined the deletion at the nucleotide level to c.-13413_2982-344del (Table 1). Because this deletion was not detected by PCR amplification of the 11 *NLRP7* exons from genomic DNA, we checked by PCR for its presence in three other unrelated patients of the same ethnic group to that of patient 1359 (Mexican origin), and in which homozygosity for a mutation in *NLRP7* was detected but the parents

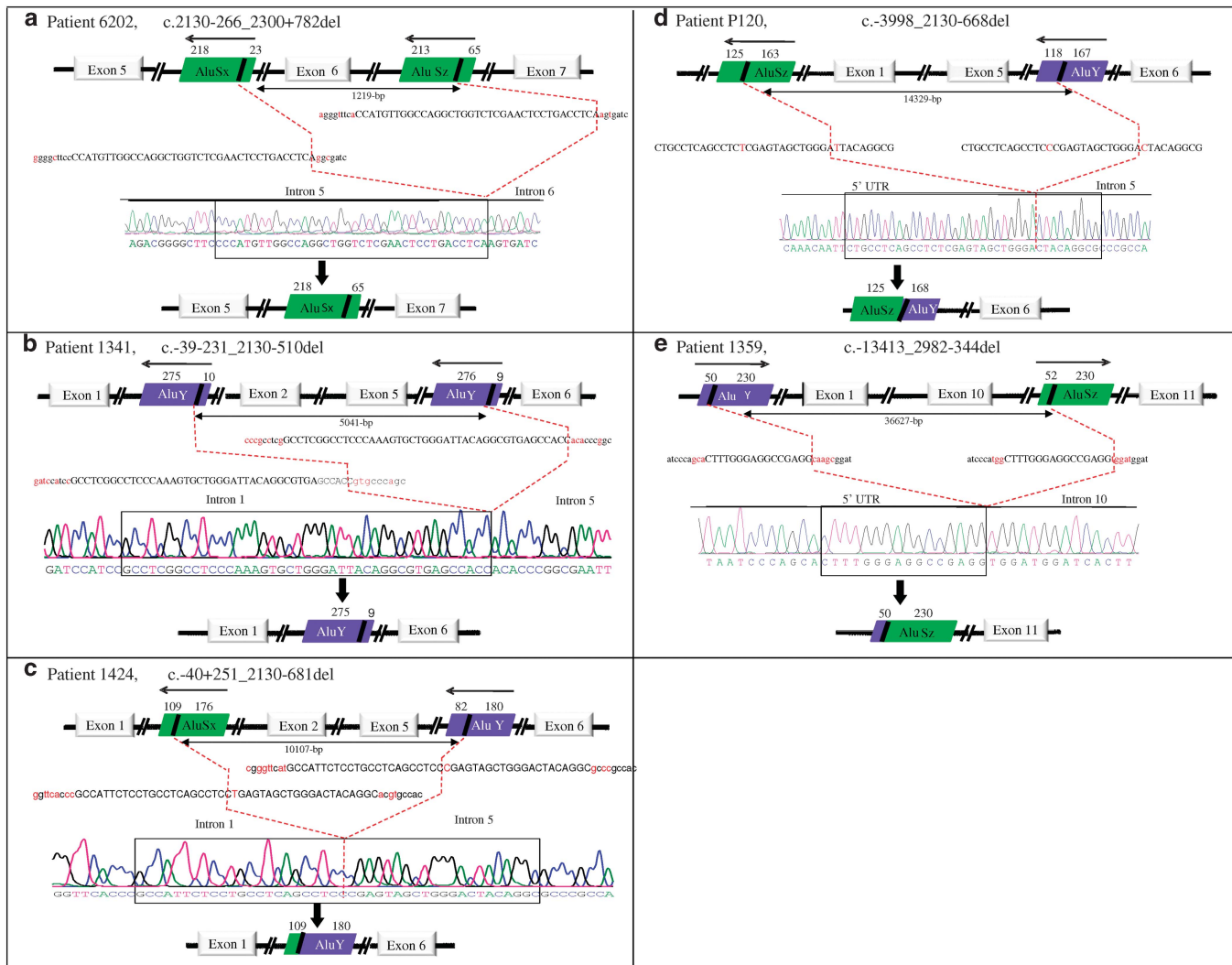


Figure 2 Characterization of five large *NLRP7* deletions identified in this study and schematic representation of the Alu elements at the breakpoints. The microhomology sequences surrounding the deletions are shown in capital letters and the unique sequences on each side of the microhomology sequences are shown in small letters. Red letters indicate differences between the two microhomology elements or the flanking sequences. Dashed red lines delimit the deletions. The orientation of the Alu repeats is indicated by arrows above each repeat.

were not available for DNA testing. This analysis did not detect the junction fragment in any of the three unrelated patients.

Analysis of the effects of four variants on *NLRP7* transcription

The 5041-bp deletion identified in patient 1341, c.-39-231_2130-510del, is the first reported homozygous deletion in *NLRP7* that removes the transcription start site (+1). To analyze its effect on the gene transcription, we performed reverse transcription followed by PCR on RNA extracted from EBV-transformed lymphoblastoid cell lines from the patient with three combinations of primers that amplify three cDNA fragments from exons 5 to 11 (Figure 1). However, no cDNA fragments were amplified with the three combinations of primers demonstrating the complete absence of *NLRP7* transcripts in the cell line from this patient.

Patient 1291 has a novel heterozygous 9-bp deletion that deletes the splice acceptor of exon 6, in which we previously reported an invariant splice site mutation, c.2130-2A>G, in a heterozygous state in patient 1200.²⁰ To determine the effects of these two variants on the splicing of exon 6, we performed RT-PCR on RNA from EBV-transformed cells from the two patients, 1291 and 1200. This revealed the presence

of a normal (298-bp) and an abnormal shorter cDNA fragments (127-bp) in the two patients (Figure 1). Purification of the short fragments from the agarose gel and their sequencing revealed their identical sequences and their lack of exon 6, presumably due to the splicing variants (Figure 1).

We also assessed the effects of a previously reported splice variant that affects the invariant splice donor site of exon 9, c.2810+2T>G, in two patients, 1074 and 1243. We found that this variant leads to different splicing in the two patients. In patient 1074, this variant led to an abnormal cDNA fragment of 673-bp that contains the first 162-bp of intron 9 and is expected to result in an in-frame insertion of 54 amino acids in the protein. In the second patient, 1243, the same variant yielded, in addition to the 673-bp fragment, to another smaller cDNA fragment of ~650-bp (Figure 1) and to some other faint fragments larger than 673-bp (data not shown) that could not be cloned and were not observed in more than 10 control subjects.

Eight Alu mediated large deletions and rearrangements in *NLRP7*

For all the five novel large deletions described in this study, the breakpoints were within Alu elements and were found flanked

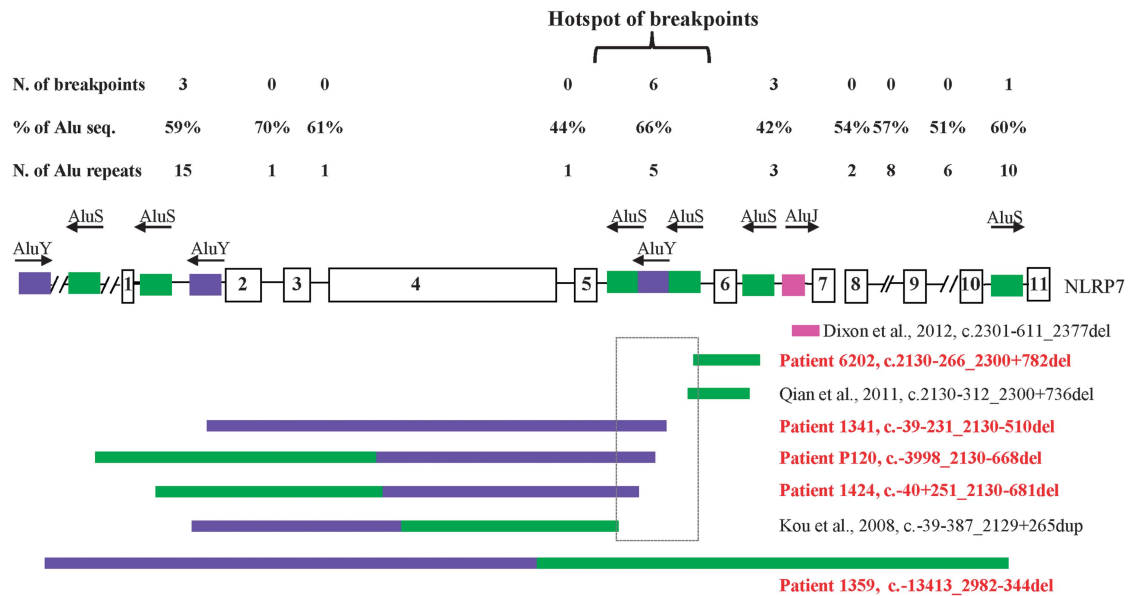


Figure 3 Recapitulation of all *NLRP7* large deletions and distribution of Alu elements in its intronic sequences. To date, eight Alu-repeat mediated deletions and rearrangements in *NLRP7* have been described. Colored boxes refer to specific Alu subfamilies, each with a specific color. Novel and previously reported mutations are in red and black, respectively. Arrows on the top of the colored boxes indicate the orientation of the Alu elements. Six of the 16 deletions and rearrangement breakpoints occurred in intron 5 and define a hotspot of Alu instability, deletions, and rearrangements shown in a rectangle. N, stands for number; seq. stands for sequences. % of Alu seq, indicates the percentage of the length of Alu sequences in each intron divided by the total length of the sequence of each intron.

by microhomology over 16-bp to 44-bp on both sides of the breakpoints with one of the microhomology element being completely or partially removed by the deletion (Figure 2). In addition, the same microhomology element was present in more than one region within the genomic structure of *NLRP7*. For instance, the microhomology element of 37-bp flanking the deletion in patient 6202 is also involved in a previously reported variant, c.2130-312_2300+736del¹⁴ and the microhomology element present at the 5' and 3' breakpoints of the deletion in patient 1424 is also present at the breakpoints of a previously reported deletion by Kou *et al.*¹⁰ We therefore analyzed the whole genomic structure of *NLRP7*, a total of 25,991-bp (starting 1-kb upstream of exon 1 to 1-kb downstream of exon 11) (hg19 at <https://genome.ucsc.edu/>) for repetitive elements using CENSOR,²⁷ a publicly available software that screens query sequences against a reference collection of repeats (<http://www.girinst.org/censor/>). This analysis led to the identification of Alu repetitive elements surrounding all the five deletions reported in this study (Figure 2) and the three previously reported ones^{10,14} (Figure 3). We also found that *NLRP7* is highly rich in Alu repeat elements, which account for ~48% of all its intronic sequences with introns 2, 5 and 10 being the richest with Alu content of 70%, 66% and 60%, respectively (Figure 3). In addition, for four of the novel deletions, their 3' breakpoints were within a 1-kb interval in intron 5, which includes also the breakpoints of two previously reported deletions and rearrangements^{10,14,28} and define a hotspot of Alu instability, deletions and rearrangements (Figure 3).

DISCUSSION

Here we describe eleven novel protein truncating variants in *NLRP7* in unrelated patients, three from familial and eight from simplex cases. All patients had RHMs and were found to have two defective alleles with the exception of one patient, 1291, who had four SAs and a single heterozygous 9-bp deletion. The presence of a single *NLRP7* protein truncating variant in this patient with four SAs supports previous

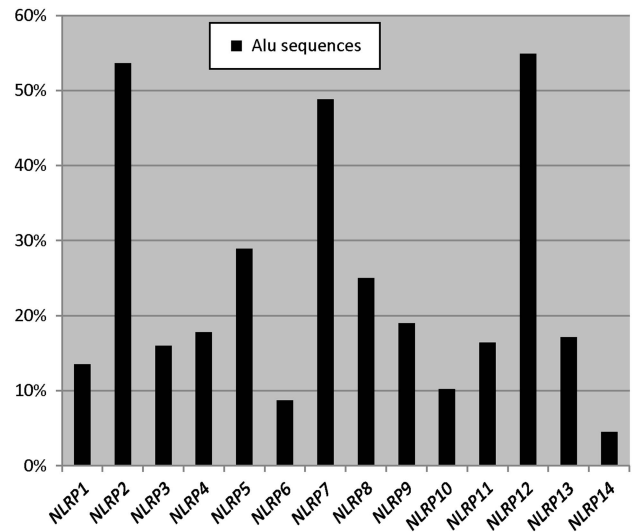


Figure 4 Alu distribution in human *NLRP* genes. For all genes, the genomic structures were downloaded from the UCSC Genome browser (hg19) and each included 1-kb upstream of the first exon and 1-kb downstream of the last exon. The presence of Alu repeats was investigated using CENSOR (<http://www.girinst.org/censor/>). The percentage of Alu sequences in each gene represents the total length of its Alu sequences over the length of its genomic structure. We note that Alu elements were mostly found in introns, which is in agreement with known data about most Alu sequences.

observations by our group²⁶ and others^{29,30} about the involvement of *NLRP7* in the genetic susceptibility for recurrent SAs in a minority of cases.

We also characterized the consequences of four different variants on *NLRP7* transcription. We demonstrated the absence of *NLRP7* transcripts in EBV-transformed cells from patient 1341 with a homozygous large deletion that removes 5041-bp including the transcription start site. This patient is therefore the first with

a demonstrated complete absence of *NLRP7* transcripts. The fact that she is in good health besides her reproductive problem strengthens current knowledge about the main requirement of *NLRP7* for female reproduction. In two other patients, 1291 and 1200, we demonstrated that their different splice variants, c.2130-6_2132del and c.2130-2A>G, lead to the same splice consequence and skipping of exon 6 from the mature RNA. We note that the skipping of exon 6 does not disrupt the open reading frame but is expected to delete 56 amino acids located between the second and the third leucine-rich-repeats and part of the third leucine-rich-repeats. In two other patients, 1074 (reported previously⁸) and 1243, we found that the mutation of the acceptor splice site of exon 9, c.2810+2T>G, activates a new splice site that does not have a canonical GT invariant donor and leads mainly to an in-frame insertion of 162-bp between exons 9 and 10. Other minor aberrant splice isoforms were also observed only in patient 1243 and could be due to the intrinsic degenerate nature of the activated non-canonical splice site. Our data on the consequences of this splice variant in patient 1074 are in agreement with those observed by Kou *et al.*¹⁰ on the same splice variant.

To date and including this study, 59 distinct pathogenic variants in recessive state have been reported in *NLRP7*, of which four previously reported large deletions^{10,14,28} and five novel ones. Analysis of the sequences flanking the nine deletions showed that seven have their two breakpoints within Alu elements, one has only one breakpoint within an Alu element,²⁸ and the remaining does not have any breakpoint in Alu elements, but in exonic sequences (in exon 4).¹⁴ Careful analysis of the deletions revealed that seven of them have microhomology of 16 to 44 nucleotides adjacent to their 5' and 3' breakpoints. In the eight large deletions, the involved microhomology repeats, as well as the Alu elements to which they belong, have the same orientations. Among 15 breakpoints located within Alu elements, 11 occurred in the left arms of the Alu elements, two occurred in the right arms and two in the middle of the Alu elements. In summary, of the 15 breakpoints within Alu elements, one occurred in intron 10, two several kilobases upstream of exon 1, three in intron 1, three in intron 6 and six in intron 5, of which three in the same AluY element. These six different deletions in unrelated subjects define a hotspot of Alu instability and deletions in intron 5 (Figure 3).

Because all the intronic breakpoints of the large *NLRP7* deletions occurred within Alu repeats, we screened *NLRP7* genomic sequence for Alu repeat elements. This analysis revealed that ~48% of the total *NLRP7* genomic structure (25,997-bp) is occupied by Alu sequences, which represent one Alu insertion every 450-bp, and is much higher than the known Alu density in the human genome (one insertion per 3,000-bp).^{31–33} Alu elements are well-known to be primates specific³⁴ and so is *NLRP7*, which is believed to have duplicated from a *NLRP2/7* ancestor in primates.²⁴ To better understand *NLRP7* evolution, we looked at the Alu content of human *NLRP2* and found it, similar to *NLRP7*, highly rich in Alu elements. Indeed, *NLRP2* and *NLRP7* along with *NLRP12* are the richest *NLRPs* in Alu elements (Figure 4). In addition, the human *NLRP2* and *NLRP7* display the same percentages of various Alu subfamilies similar to their orthologues in chimpanzees, gorillas, orangutans, rhesus, baboon and marmoset (data not shown) suggesting that Alu insertion in primates *NLRP2* and *NLRP7* occurred in their common *NLRP2/7* ancestor before its duplication into two genes. Because of the known role of Alu elements in segmental duplications on many chromosomes,^{25,28} it is possible that Alu insertion and expansion in the primate *NLRP2/7* may have mediated, or helped in, its duplication in two distinct genes.

Alu elements are the most abundant and successful short interspersed elements found in the human genome. Alu elements are

estimated to contribute to 0.075% of human mutations³⁵ and form a major part of extensive genomic structural variation.^{33,36,37} There are multiple factors that predispose Alu elements to recombination and include, the number of Alu elements in a chromosome or a region, the close proximity of the Alu elements to each other, the high GC content of their sequences, and the high sequence similarity between the different Alu subfamilies (70–100%).³³ Including the current study, the total number of distinct *NLRP7* mutations is 59 of which eight (13.5%) are mediated by Alu recombination. Therefore, Alu mediated mutations seem to be 180-times more frequent in *NLRP7* than the average of all known human mutations. This suggests that the genomic architecture of *NLRP7* gene may be prone to the occurrence of Alu-mediated deletions and rearrangements that may not be detected by Sanger sequencing or next generation exome sequencing. Our data call for more attention in reporting DNA testing in patients with apparently homozygous *NLRP7* mutations, mainly when the parents are not available for testing. In such patients, compound heterozygous mutations for one detectable mutation by Sanger sequencing and a large deletion may appear as homozygous for the detectable mutation and consequently lead to an inaccurate molecular diagnosis and reproductive genetic counseling.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We thank the patients and their family for their cooperation. The authors wish to acknowledge the use of the Sequencing platform of the McGill University and Génome Québec Innovation Centre. NMPN was supported by a Max E. Binz Fellowship from the McGill University Faculty of Medicine. The study was supported by the Canadian Institute of Health Research grants, MOP-86546 and PPP-122897, to RS.

- 1 Savage P, Williams J, Wong SL *et al*: The demographics of molar pregnancies in England and Wales from 2000–2009. *J Reprod Med* 2010; **55**: 341–345.
- 2 Grimes DA: Epidemiology of gestational trophoblastic disease. *Am J Obstet Gynecol* 1984; **150**: 309–318.
- 3 Eagles N, Sebire NJ, Short D, Savage PM, Seckl MJ, Fisher RA: Risk of recurrent molar pregnancies following complete and partial hydatidiform moles. *Hum Reprod* 2015; **30**: 2055–2063.
- 4 Sebire NJ, Fisher RA, Foskett M, Rees H, Seckl MJ, Newlands ES: Risk of recurrent hydatidiform mole and subsequent pregnancy outcome following complete or partial hydatidiform molar pregnancy. *BJOG* 2003; **110**: 22–26.
- 5 Horn LC, Kowalzik J, Bilek K, Richter CE, Eibenkel J: Clinicopathologic characteristics and subsequent pregnancy outcome in 139 complete hydatidiform moles. *Eur J Obstet Gynecol Reprod Biol* 2006; **128**: 10–14.
- 6 Kronfol NM, Iliya FA, Hajj SN: Recurrent hydatidiform mole: a report of five cases with review of the literature. *J Med Liban* 1969; **22**: 507–520.
- 7 Vargas R, Barroilhet LM, Esselen K *et al*: Subsequent pregnancy outcomes after complete and partial molar pregnancy, recurrent molar pregnancy, and gestational trophoblastic neoplasia: An update from the New England Trophoblastic Disease Center. *J Reprod Med* 2014; **59**: 188–194.
- 8 Murdoch S, Djuric U, Mazhar B *et al*: Mutations in *NALP7* cause recurrent hydatidiform moles and reproductive wastage in humans. *Nat Genet* 2006; **38**: 300–302.
- 9 Parry DA, Logan CV, Hayward BE *et al*: Mutations causing familial biparental hydatidiform mole implicate c6orf221 as a possible regulator of genomic imprinting in the human oocyte. *Am J Hum Genet* 2011; **89**: 451–458.
- 10 Kou YC, Shao L, Peng HH *et al*: A recurrent intragenic genomic duplication, other novel mutations in *NLRP7* and imprinting defects in recurrent biparental hydatidiform moles. *Mol Hum Reprod* 2008; **14**: 33–40.
- 11 Wang CM, Dixon PH, Decordova S *et al*: Identification of 13 novel *NLRP7* mutations in 20 families with recurrent hydatidiform mole; missense mutations cluster in the leucine-rich region. *J Med Genet* 2009; **46**: 569–575.
- 12 Hayward BE, De Vos M, Talati N *et al*: Genetic and epigenetic analysis of recurrent hydatidiform mole. *Hum Mutat* 2009; **30**: E629–E639.
- 13 Slim R, Bagga R, Chebaro W, Srivivasan R, Agarwal N: A strong founder effect for two *NLRP7* mutations in the Indian population: an intriguing observation. *Clin Genet* 2009; **76**: 292–295.

- 14 Qian J, Cheng Q, Murdoch S *et al*: The genetics of recurrent hydatidiform moles in China: correlations between *NLRP7* mutations, molar genotypes and reproductive outcomes. *Mol Hum Reprod* 2011; **17**: 612–619.
- 15 Estrada H, Buentello B, Zenteno JC, Fiszman R, Aguinaga M: The p.L750V mutation in the *NLRP7* gene is frequent in Mexican patients with recurrent molar pregnancies and is not associated with recurrent pregnancy loss. *Prenat Diagn* 2013; **33**: 205–208.
- 16 Akoury E, Gupta N, Bagga R *et al*: Live births in women with recurrent hydatidiform mole and two *NLRP7* mutations. *Reprod Biomed Online* 2015; **31**: 120–124.
- 17 Khare S, Dorfleutner A, Bryan NB *et al*: An *NLRP7*-containing inflammasome mediates recognition of microbial lipopeptides in human macrophages. *Immunity* 2012; **36**: 464–476.
- 18 Messaed C, Akoury E, Djuric U *et al*: *NLRP7*, a nucleotide oligomerization domain-like receptor protein, is required for normal cytokine secretion and co-localizes with Golgi and the microtubule-organizing center. *J Biol Chem* 2011; **286**: 43313–43323.
- 19 Mahadevan S, Wen S, Wan YW *et al*: *NLRP7* affects trophoblast lineage differentiation, binds to overexpressed YY1 and alters CpG methylation. *Hum Mol Genet* 2014; **23**: 706–716.
- 20 Nguyen NM, Zhang L, Reddy R *et al*: Comprehensive genotype-phenotype correlations between *NLRP7* mutations and the balance between embryonic tissue differentiation and trophoblastic proliferation. *J Med Genet* 2014; **51**: 623–634.
- 21 Akoury E, Zhang L, Ao A, Slim R: *NLRP7* and *KHDC3L*, the two maternal-effect proteins responsible for recurrent hydatidiform moles, co-localize to the oocyte cytoskeleton. *Hum Reprod* 2015; **30**: 159–169.
- 22 Reddy R, Akoury E, Phuong Nguyen NM *et al*: Report of four new patients with protein-truncating mutations in *C6orf221/KHDC3L* and colocalization with *NLRP7*. *Eur J Hum Genet* 2013; **21**: 957–964.
- 23 Cai C, Tamai K, Molyneaux K: *KHDC1B* is a novel CPEB binding partner specifically expressed in mouse oocytes and early embryos. *Mol Biol Cell* 2010; **21**: 3137–3148.
- 24 Tian X, Pascal G, Monget P: Evolution and functional divergence of *NLRP* genes in mammalian reproductive systems. *BMC Evol Biol* 2009; **9**: 202.
- 25 Ulker V, Gurkan H, Tozkir H *et al*: Novel *NLRP7* mutations in familial recurrent hydatidiform mole: are *NLRP7* mutations a risk for recurrent reproductive wastage? *Eur J Obstet Gynecol Reprod Biol* 2013; **170**: 188–192.
- 26 Messaed C, Chebaro W, Di Roberto RB *et al*: *NLRP7* in the spectrum of reproductive wastage: rare non-synonymous variants confer genetic susceptibility to recurrent reproductive wastage. *J Med Genet* 2011; **48**: 540–548.
- 27 Kohany O, Gentles AJ, Hankus L, Jurka J: Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics* 2006; **7**: 474.
- 28 Dixon PH, Trongwongsa P, Abu-Hayyah S *et al*: Mutations in *NLRP7* are associated with diploid biparental hydatidiform moles, but not androgenetic complete moles. *J Med Genet* 2012; **49**: 206–211.
- 29 Slim R, Ao A, Surti U *et al*: Recurrent triploid and dispermic conceptions in patients with *NLRP7* mutations. *Placenta* 2011; **32**: 409–412.
- 30 Huang JY, Su M, Lin SH, Kuo PL: A genetic association study of *NLRP2* and *NLRP7* genes in idiopathic recurrent miscarriage. *Hum Reprod* 2013; **28**: 1127–1134.
- 31 Jurka J, Kohany O, Pavlicek A, Kapitonov VV, Jurka MV: Duplication, coclustering, and selection of human Alu retrotransposons. *Proc Natl Acad Sci USA* 2004; **101**: 1268–1272.
- 32 Batzer MA, Deininger PL: Alu repeats and human genomic diversity. *Nat Rev Genet* 2002; **3**: 370–379.
- 33 Sen SK, Han K, Wang J *et al*: Human genomic deletions mediated by recombination between Alu elements. *Am J Hum Genet* 2006; **79**: 41–53.
- 34 Konkel MK, Walker JA, Batzer MA: LINEs and SINEs of primate evolution. *Evol Anthropol* 2010; **19**: 236–249.
- 35 Deininger PL, Batzer MA: Alu repeats and human disease. *Mol Genet Metab* 1999; **67**: 183–193.
- 36 Flynn EK, Kamat A, Lach FP *et al*: Comprehensive analysis of pathogenic deletion variants in Fanconi anemia genes. *Hum Mutat* 2014; **35**: 1342–1353.
- 37 Hedges DJ, Deininger PL: Inviting instability: Transposable elements, double-strand breaks, and the maintenance of genome integrity. *Mutat Res* 2007; **616**: 46–59.

Supplementary Information accompanies this paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)