



Characterization of a novel transcript of the EHMT1 gene reveals important diagnostic implications for Kleefstra Syndrome

Willy M Nillesen, Helger G Yntema, Marco Moscarda, Nienke E Verbeek, Louise C. Wilson, Frances Cowan, Marga Schepens, Annick Raas-Rothschild, Orly Orly Yehouda -Weinstein, Marcella Zollino, et al.

► To cite this version:

Willy M Nillesen, Helger G Yntema, Marco Moscarda, Nienke E Verbeek, Louise C. Wilson, et al.. Characterization of a novel transcript of the EHMT1 gene reveals important diagnostic implications for Kleefstra Syndrome. Human Mutation, 2011, 32 (7), pp.853. 10.1002/humu.21523 . hal-00655181

HAL Id: hal-00655181

<https://hal.science/hal-00655181>

Submitted on 27 Dec 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Characterization of a novel transcript of the EHMT1 gene reveals important diagnostic implications for Kleefstra Syndrome

Journal:	<i>Human Mutation</i>
Manuscript ID:	humu-2010-0547.R2
Wiley - Manuscript type:	Research Article
Date Submitted by the Author:	07-Apr-2011
Complete List of Authors:	<p>Nillesen, Willy; Radboud University Nijmegen Medical Centre, 849 Department of Human Genetics Yntema, Helger; Radboud University Nijmegen Medical Centre, 849 Department of Human Genetics Moscarda, Marco; Catholic University, Rome, Italy, Institute of Medical Genetics Verbeek, Nienke; UMC Utrecht, Utrecht, The Netherlands, 3Department of Medical Genetics Wilson, Louise; Great Ormond Street Hospital for Children NHS Trust and Institute of Child Health, University College London, UK, Clinical Genetics Unit Cowan, Frances; Imperial College Healthcare Trust, London, UK, Perinatal Neurology Department of Paediatrics Schepens, Marga; Radboud University Nijmegen Medical Centre, 849 Department of Human Genetics Raas-Rothschild, Annick; Hadassah Hebrew University Medical Center Orly Yehouda -Weinstein, Orly; Hadassah Hebrew University Medical Center Zollino, Marcella; Catholic University, Rome, Italy, Institute of Medical Genetics Vijzelaar, R; 7MRC Holland bv, Amsterdam, the Netherlands Neri, Giovanni; Catholic University, Rome, Italy, Institute of Medical Genetics Nelen, Marcel; Radboud University Nijmegen Medical Centre, 849 Department of Human Genetics; UMC Utrecht, Utrecht, The Netherlands, 3Department of Medical Genetics van Bokhoven, Hans; Radboud University Nijmegen Medical Centre, 849 Department of Human Genetics Giltay, Jacques; University Medical Center, Utrecht, The Netherlands, Department of Biomedical Genetics Kleefstra, Tjitske; Radboud University Nijmegen Medical Centre, 849 Department of Human Genetics</p>

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Key Words:	Kleefstra syndrome , Chromosome 9q34 microdeletion, 9qSTDS, EHMT1

SCHOLARONE™
Manuscripts

For Peer Review

Characterization of a novel transcript of the *EHMT1* gene reveals important diagnostic implications for Kleefstra Syndrome

Willy M Nillesen^{1†}, Helger G Yntema^{1†}, Marco Moscarda², Nienke E. Verbeek³,
Louise C Wilson⁴, Frances Cowan⁵, Marga Schepens¹, Annick Raas-Rothschild⁶,
Orly Gafni-Weinstein⁶, Marcella Zollino², Raymon Vijzelaar⁷, Giovanni Neri², Marcel
Nelen^{1,3}, Hans van Bokhoven¹, Jacques Giltay³, Tjitske Kleefstra¹

¹ Department of Human Genetics, RUNMC, Nijmegen, The Netherlands

² Institute of Medical Genetics, Catholic University, Rome, Italy

³ Department of Medical Genetics, UMC Utrecht, Utrecht, The Netherlands

⁴ Clinical Genetics Unit, Great Ormond Street Hospital for Children NHS Trust and Institute of Child Health, University College London, UK

⁵ Perinatal Neurology, Department of Paediatrics, Imperial College Healthcare Trust, London, UK

⁶ Department of Human Genetics and Metabolic Diseases, Hadassah Hebrew University Medical Center, Jerusalem, Israel

⁷ MRC Holland bv, Amsterdam, the Netherlands

† These authors contributed equally to this manuscript

Corresponding author:

Tjitske Kleefstra MD, PhD

849 Department of Human Genetics

PO Box 9101

6500 HB Nijmegen

The Netherlands

0031-(0)243613946

Key words: Kleefstra syndrome, Chromosome 9q, Subtelomere deletion, 9qSTDS,
EHMT1

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

ABSTRACT

The core phenotype of Kleefstra syndrome (KS) is characterized by intellectual disability, childhood hypotonia and a characteristic facial appearance. This can be caused by either submicroscopic 9q34 deletions or loss of function mutations of the *EHMT1* gene. Remarkably, in three patients with a clinical suspicion of KS, molecular cytogenetic analysis revealed an interstitial 9q34 microdeletion proximal to the coding region of the *EHMT1* gene based on the NM_024757.3 transcript. Since we found a mono-allelic *EHMT1* transcript suggestive for haploinsufficiency of *EHMT1* in two of these patients tested, we hypothesized that a deletion of regulatory elements or so far unknown coding sequences in the 5' region of the *EHMT1* gene, might result in a phenotype compatible with KS. We further characterized the molecular content of deletions proximal to the transcript NM_024757.3 and confirmed presence of a novel predicted open reading frame comprising 27 coding exons (NM_024757.4). Further analysis showed that all three deletions included the presumed novel first exon of the *EHMT1* gene and subsequent testing of 75 individuals without previously detectable *EHMT1* aberrations, showed one additional case with a deletion comprising only this 5' part of the gene. These results have important implications for the genetic screening of KS and for studies of the functional significance of *EHMT1*.

Deleted: This new *EHMT1* transcript therefore.

INTRODUCTION

Kleefstra syndrome (MIM# 610253) also known as 9q Subtelomeric Deletion Syndrome; 9qSTDS, is characterized by intellectual disability, childhood hypotonia, a characteristic facial appearance and in some patients, by congenital heart and renal defects, (brachy)microcephaly, epilepsy, obesity and behavioural problems as well (Stewart et al., 2004; Yatsenko et al., 2005; Stewart and Kleefstra, 2007; Kleefstra et al., 2005, 2006, 2009; Verhoeven et al., 2010). The syndrome can be caused by either submicroscopic deletions of the distal long arm of chromosome 9q or intragenic loss of function mutations in the *EHMT1* (*Euchromatic Histone Methyltransferase 1*) gene (MIM# 607001), both leading to haploinsufficiency of the *EHMT1* gene. Since the identification of this gene, we have provided regular diagnostic services for mutational analysis comprising direct sequencing of the coding region and screening for deletions by Multiplex Ligation-dependent Probe Amplification (MLPA) of the transcript NM_024757.3 (<http://www.ncbi.nlm.nih.gov/>).

So far, we have published information about 27 cases with *EHMT1* deletions or mutations (Kleefstra et al., 2006, 2009).

Remarkably, in three patients with a clinical suspicion of Kleefstra syndrome (KS), molecular cytogenetic analysis revealed a submicroscopic interstitial 9q deletion proximal to the coding region of the *EHMT1* gene. We therefore hypothesized that regulatory elements or so far unknown coding sequences in the 5' region of the *EHMT1* gene, might be affected by the deletion resulting in a phenotype compatible with KS. In this study we aimed to further characterize the molecular content of deletions proximal to the transcript NM_024757.3 causing KS and test a large group of patients suspected of having KS with this hitherto negative findings.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

PATIENTS AND METHODS

Patients

All patients were referred to our centre for further genetic analysis because of the clinical suspicion of Kleefstra syndrome. Informed consent was obtained for further publication of data and photographs.

Patient 1 (Figure 1A,B) was born by vaginal delivery at 40 weeks gestation after a pregnancy complicated by first trimester bleedings, treated with progesterone, and contractions at 17 weeks, treated with isoxsuprine. Her birth weight was 3020 g (25-50th centile), birth length 47 cm (10-25th centile) and head circumference 32 cm (10-25th centile). Her APGAR scores were 8/9 at 1/5 minutes. During the first day after birth, she was bradycardic and generally hypotonic. She suffered from gastro-esophageal reflux until 8 months of age. Reduced head circumference centile was noted at 5 months. Sleep, feeding and bowel transit and an EEG and echocardiography were normal.

At 16 months her weight was 10 kg (~25th centile), height 80 cm (~50th centile), and head circumference 43 cm (<2nd centile). She had mild synophrys, upslanting palpebral fissures, hypoplastic nasal saddle, normal ears with folded helix (also in the mother), a narrow palate and persistent fetal pads on her fingers. She had normal female external genitalia and a normal anus. She remained generally hypotonic and had joint hyperlaxity.

At three years of age she did not use complete words, but communicated with gestures (i.e. bringing the parent's hand to the item she wanted). She could walk but

not run or climb stairs independently. She had sporadic stereotypic movements. She was longsighted (+7.5 dioptries).

Patient 2 (Figure 1 C,D) was born at term after an uneventful pregnancy. Delivery was induced because pregnancy was overdue. Her birth weight was 2460 g (< 2nd centile) , length 49 cm (25th centile) and head circumference 32 cm (10-25th centile), and her APGAR scores were 5/8 at 1/5 min. There were neonatal feeding problems and hypotonia.

Her developmental milestones were delayed. She could sit at one year, stand at two years and walk without support at three years of age. She was first referred to us at 15 months for investigation of her psychomotor retardation, hypotonia and joint hypermobility. At five years moderate to severe psychomotor retardation and autism were diagnosed. At this time her height was on the 50th centile, her length on the 10th centile and head circumference just <3rd centile but at latest examination at 12 years it was at -1/-2 SD. There were facial dysmorphisms such as pronounced eyebrows, small mouth, irregularly placed teeth, full lips, anteverted nostrils, protruding ears with a thick ear helix in addition to generalized hypotonia and joint hypermobility which her mother also had.

Auditory brainstem evoked responses (ABER), routine metabolic screening of plasma and urine, cardiac examination including cardiac echography, brain MRI and repeated chromosome analysis and FISH 7q11.21-23 (Williams syndrome) did not reveal any abnormalities.

Patient 3 (Figure 1E) was born at 40 weeks gestation with a birth weight of 2878 g (10th centile). Polyhydramnios had been diagnosed at week 30. During the first

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

months after birth she had feeding difficulties and needed nasogastric tube feeding.

Her parents were healthy

At 23 months the patient presented with hypotonia and psychomotor delay, left hydronephrosis, and a patent foramen ovale. She had an abnormal hearing test (ABER was positive only at 60db). Her brain MRI showed agenesis of the corpus callosum. She sat at 24 months. At 34 months her height was 91 cm (10th centile), weight was 13.5 kg (12th centile), head circumference 45 cm (<3rd centile), palm length 6.2 cm (25th centile), third finger 4 cm (3rd centile). She was brachycephalic and had a low frontal hairline, synophrys, hypertrichosis, adenoid hypertrophy, a bifid uvula and a protruding tongue. At 3 years 9 months her height was 91.8 cm (3rd centile), weight 13.5 kg (25th centile) and head circumference 45.5 cm (-3SD). At this age she had no behavioral problems, could walk with help but did not say any words.

Routine chromosomal analysis showed a normal female karyotype. Molecular analysis of the SNRP gene region allowed ruling out of uniparental disomy and methylation abnormality compatible with Angelman syndrome and transferrin immunoelectrophoresis was negative for carbohydrate deficient glycoprotein syndromes.

Patient 4 (Figure 1 F,G) was the second child of healthy unrelated parents. The pregnancy was unremarkable until 22 weeks when routine ultrasound scans showed borderline cerebral ventriculomegaly, choroid plexus cysts and echogenic bowel. Amniocentesis and screening for congenital infections were normal. Rescanning at 27 weeks gestation showed greater dilatation of the ventricles and a fetal MRI at 32 weeks gestation showed bilateral marked ventriculomegaly. She was delivered by elective caesarean section at 39 weeks gestation weighing 2094 g (9-25th centile)

with a head circumference of 35cm (50-75th centile) and did not require resuscitation – her Apgar scores were 7/8 (after 1/5 minutes). She had global hypotonia, feeding difficulties secondary to a disorganised suck, gastro-esophageal reflux which was treated medically, and laryngomalacia. She had some dysmorphic facial features as described below. Her development was delayed. She walked at around 18 months of age. At 3 years her cognitive abilities were assessed at around a 13 month level. At 4 years she had around 100 words and used some 2-3 word phrases although her articulation was poor. She was sociable and good natured but had poor concentration. She has never had seizures. She has had recurrent middle ear infections and had mild-moderate conductive hearing loss and glue ear for which she has had grommets. She was long-sighted with a strabismus for which she wears glasses. She had 3 urinary tract infections but a renal ultrasound and DMSA scan did not show any significant abnormality.

When last formally examined at 3 years 3 months her height was 91.8cm (9-25th centile), weight 14.0 kg (25-50th centile) and head circumference 50.9 cm (50-75th centile). She had up-slanting palpebral fissures, mild ptosis, a flattened mid face and nose with a rounded down-turned nasal tip and short columella, a tented mouth with down turned corners, a smallish chin and a protruding tongue. She had fine hair with a frontal upsweep and slightly sparse eyebrows. Her ears appeared normal. She had marked dimpling over both her sacroiliac joints and behind her shoulders. She had increased joint laxity and her fifth toes were curled under. An MRI brain at 7 weeks of age confirmed bilateral lateral ventricular dilation, right greater than left, unusually shaped posterior horns, a relatively small cisterna magna, a thinned corpus callosum and a slightly thickened cortex in the medial temporal region.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Investigations have included a postnatal karyotype, telomere FISH, genome wide array-CGH analysis at approximately 1Mb resolution, Prader-Willi methylation testing, full blood count, urea, electrolytes, lactate, creatinine kinase, thyroid function, very long chain fatty acids, transferrin isoelectric focussing, plasma amino-acids, urine organic acids and glycosaminoglycans, CSF lactate, pyruvate, glucose and neurotransmitters, all of which were normal. There was a slightly low CSF/blood glucose ratio of 46% that was not investigated further in the absence of seizures. An EEG did not show any significant abnormalities and echocardiography was normal.

***EHMT1* mutation negative patient cohort**

An additional cohort of 75 patients, who were referred to the DNA diagnostic service of the Department of Human Genetics in Nijmegen, the Netherlands with a clinical suspicion of Kleefstra syndrome, were incorporated in this study. No mutations had been identified by sequencing and MLPA analysis of the coding region of *EHMT1* based on the NM_024757.3 transcript. Furthermore, all patients have been screened for clinically relevant Copy Number Variations as part of their diagnostic routine by using high-density single nucleotide polymorphism (SNP) microarray platforms. MLPA analysis was performed using a home-designed MLPA kit containing synthetic probes in all coding *EHMT1* exons, except for exon 21 (Kleefstra et al., 2006, 2009).

Genomic and Molecular analysis

Genome wide array studies

Several array platforms were used to identify chromosomal abnormalities.

In patient 1, Array-CGH was performed using Agilent oligonucleotide-array (Human Genome CGH microarray 4x44K, Agilent Technologies Santa Clara, CA), with an average resolution of 75 kb, following the manufacturer's instructions. The array was analyzed using GenePix 4000B scanner (Axon, Union City, CA) and Feature Extraction V.9.5.1 software. A graphical overview of the results was obtained using CGH Analytics V.3.5.14 software.

In patient 2, the 105K custom *Agilent* oligo array (Oxford design, AMADID 019015) containing over 99,000 unique *in situ* synthesized 60-mer oligonucleotides, with a resolution of 21.7 kb, was applied. Hybridizations were performed according to the manufacturer's protocols (*Agilent* Technologies, Santa Clara CA). Images of the arrays were acquired using a G2505B microarray scanner (*Agilent* Technologies, Santa Clara CA). Data-analysis was done using the CGH-analytics software (*Agilent* Technologies, Santa Clara CA).

In patient 3, the Affymetrix Cytogenetics Whole-Genome 2.7M Array (Affymetrix, Santa Clara, CA) was used to screen for Copy Number Variations (CNVs). Genome-wide SNP-array hybridization was performed following the manufacturer's instructions. Genome-wide copy number analysis was performed using the Affymetrix Chromosome Analysis Suite software (ChAS) and a pooled reference DNA sample data set.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

For data interpretation, CNVs in the patient were compared to those previously identified in more than 840 control individuals and those reported in the Database of Genomic Variants (<http://projects.tcag.ca>).

Expression analysis

Expression studies were performed in patients 1 and 3 in order to investigate if the deletion had an effect on the expression level of *EHMT1*. Lymphoblastoid cell lines were established by Epstein–Barr virus transformation from peripheral blood lymphocytes of the patient and healthy controls. Lymphoblasts were grown in RPMI-1640 medium supplemented with 10% fetal bovine serum and penicillin/streptomycin at 37°C with 5% CO₂. Cell medium was changed every 48h.

Total RNA was extracted with the single-step acid phenol method using Trizol (Invitrogen). Two independent RNA extractions were performed from lymphoblastoid cell lines established from both patients. RNA was reverse-transcribed by MoMLV-reverse transcriptase (Invitrogen). PCR products containing a coding SNP were designed. The cDNA of Patient 1 was amplified by primers surrounding the SNP g.139,731,389G>A (UCSC Genome Browser database, March 2006 release (<http://genome.ucsc.edu>; hg18). The cDNA of patient 3 was amplified by primers surrounding the SNP rs1129768. Primer sequences and PCR conditions are available upon request.

***In silico* analysis and expression studies of a novel *EHMT1* transcript**

A search for novel coding exons in the *EHMT1* gene was performed at the UCSC Genome Browser database, February 2009 release (<http://genome.ucsc.edu>; hg19),

and the consensus CDS project (CCDS) database at NCBI (<http://www.ncbi.nlm.nih.gov/CCDS/CcdsBrowse.cgi>).

MLPA and Sequencing

Multiplex ligation dependent probe amplification was performed by applying MLPA kit P340-A1, which was developed by MRC-Holland (www.mrc-holland.com) comprising 31 probes in *EHMT1*, covering almost the entire coding region, based on transcript NM_024757.4. Since the high C/G content in exon 1 makes it impossible to design an MLPA probe, a probe in intron 1 is included. In this MLPA kit also two probes in the proximally located gene *C9orf37* (exons 1 and 2) and one probe in the distally located gene *CACNA1B* (exon 3) are present. Tests were carried out according to the manufacturer's protocols (Version 2007, MRC-Holland, The Netherlands). Analysis was done using the MLPA module of Genemapper software, version 4.0 (Applied Biosystems).

Direct sequencing of *EHMT1* exons 1 and 2, based on transcript NM_024757.4, was performed by standard procedures.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

RESULTS

Genome wide array studies

The genome wide array studies in patients 1-3 are based on different array platforms. The probe locations mentioned below are based on the UCSC Genome Browser database, build NCBI 36, UCSC hg18, Mar 2006 (<http://genome.ucsc.edu>).

In patient 1, the first deleted probe (A_14_P122901) is located on chromosome 9 position 139,407,449, the last deleted probe (A_14_P100411) on position 139,633,014. The first preserved probes (A_14_P125110 (chr9:139,405,999) proximally, A_14_P105992 (chr9:139,680,613 distally)). The deletion included the following genes: partially *EXD3*, *NOXA1*, *ENTPD8*, *NELF*, *PNPLA7*, *MRPL41*, *WDR85*, *ZMYND19*, *ARRDC1*, and *C9orf37*.

In patient 2, the first deleted probe is A_14_P125110, the last one A_14_P105992 (chr9:139,405,999-139,680,613). The first preserved probes are A_16_P18802737 and A_16_P02195868 (chr9: 139,390,002-139,697,348), indicating that the deletion contains 11 known genes (partially *EXD3*, *NOXA1*, *ENTPD8*, *NELF*, *PNPLA7*, *MRPL41*, *WDR85*, *ZMYND19*, *ARRDC1* and *C9orf37*).

In patient 3, probe C-1CKEK (chr9: 139,609,502) was found to be deleted by SNP array analysis, while the proximal probe C-1CK64 showed a normal ratio (139,273,509). Fine mapping of the proximal deletion breakpoint by MLPA analysis with MRC Holland Kit P286 (www.mrcholland.com) showed that the proximal deletion breakpoint is located between the proximal probe C-1CK64 and the deleted probe in

NELF exon 12 (chr9: 139,464,509). The distal deletion breakpoint is located between the probes S-2JMMP (chr9:139,649,316) and S-2IGFE (chr9: 139,659,097). The deletion in patient 3 contains (at least) the genes *NELF*, *PNPLA7*, *MRPL41*, *WDR85*, *ZMYN19*, *ARRDC1* and *C9orf37*.

Expression analysis

Expression analysis of *EHMT1* performed in cases 1 and 3, showed mono-allelic expression of the *EHMT1* gene. On the DNA level, heterozygosity for a coding SNP is shown (figures 2A and 2C). On the cDNA level only one of the alleles is expressed (figures 2B and 2D). This indicated haploinsufficiency for *EHMT1*, though this gene did not seem to be deleted on the basis of array results.

In silico analysis and RT-PCR

According to the UCSC Genome Browser database, February 2009 release (<http://genome.ucsc.edu>), and the consensus CDS project (CCDS) database at NCBI (<http://www.ncbi.nlm.nih.gov/CCDS/CcidsBrowse.cgi>), the coding region of the *EHMT1* gene has been updated to extend the N-terminus with an extra 5' exon. The previously defined *EHMT1* gene transcript (NM_024757.3) contained 26 exons, the translation start site being located in exon 2. Diagnostic testing so far has been directed towards the 25 coding exons of the *EHMT1* gene, however the novel open reading frame comprises 27 coding exons (NM_024757.4). The translation start site is located in a "novel" exon 1, 97.6 kb proximal to the 'old' ATG start codon and surrounded by a more than 75% CG rich region (Figure 3). The previously determined non-coding first exon is now a coding exon (exon 2) in the new transcript.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

To confirm that the novel exons are expressed we performed RT-PCR analysis with primers specific for the novel *EHMT1* exons and obtained a PCR product corresponding in size and sequence to NM_024757.4 (data not shown). The *EHMT1* protein therefore, might appear to be 31 amino acids larger comprising in total 1299 amino acids. This presumed new *EHMT1* N-terminus does not contain any predicted domains, repeats, motifs or other features. Both the old (NM_024757.3) and new transcript (NM_024757.4) are indicated in figure 3.

MLPA

The MLPA probes located in exons 1 and 2 of *C9orf37* and in intron 1 of *EHMT1*, clearly confirmed the presence of a heterozygous deletion in patients 1-3 whereas the distally located probe in novel exon 2 and the other distal exons of *EHMT1* showed normal results (Figure 4).

To screen for additional deletions covering only this extended 5' part of the *EHMT1* gene, we screened 75 patients who had been previously referred for diagnostic analysis of *EHMT1* but with normal results based on sequencing and MLPA analysis of transcript NM_024757.3. One case (patient 4) showed a deletion of *C9orf37* exon 1 (chr9: 139,632,979) and the *EHMT1* intron 1 probe (chr9:139,633,958) in MRC Holland kit P340-A1. The proximally located probe *C9orf37* exon 2 (chr9; 139,632,295) and the distally located probe in *EHMT1* exon 2 (chr9:139,725,213) showed a normal dosage.

Deletion mapping

The results of the deletion mapping, including array and MLPA probes, of patients 1-4 are schematically shown in figure 5. All aberrations comprised *de novo* interstitial deletions ranging from 1.0 kb to 386 kb. For probe alignments: build NCBI 36, UCSC hg18, Mar 2006 build was used.

The deletion in patient 1 was spanning at least 226 kb, based on the first deleted array probe (A_14_P122901) and last deleted MLPA probe in *EHMT1* intron 1. The maximal size of the deletion is 275 kb, based on the first and last preserved array probes (A_14_P125110 and A_14_P105992, respectively).

In patient 2, the smallest deleted region is about 275 kb in size, and the maximal size of the deletion is 307 kb, based on the array results (deleted A_14_P125110- A_14_P105992; preserved A_1618802737- A_16_P02195868). The results of the MLPA did not add to the deletion size.

The deletion in patient 3 ranges minimally from the MLPA probe in *NELF* exon 12 to array probe S-2JMMP, and maximally from array probes C-1CK64 to S-2IGFE. The deletion is therefore between 185 kb and 386 kb in size.

Patient 4 has one of the smallest interstitial deletions described in Kleefstra syndrome so far: between 1.0 and 93 kb in size (based on the deleted MLPA probes *C9orf37* exon 1 and *EHMT1* intron 1 and the preserved probes *C9orf37* exon 2 and *EHMT1* exon 2.1). Since the results in this patient proves that a deletion of only the novel *EHMT1* exon 1 can be causative of Kleefstra syndrome, we wanted to investigate whether point mutation in the novel part of the transcript are present in Kleefstra syndrome. We therefore PCR amplified and sequenced exons 1 and 2 in the remaining 74 individuals with Kleefstra syndrome, but this revealed no mutations.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

DISCUSSION

Here we describe the identification and important clinical relevance of an elongated *EHMT1* transcript. This novel transcript might contribute 31 additional amino acids to the 5' coding region of the previous transcript which would result in a novel EHMT1 protein of 1299 amino acids. So far, it is not known what biological properties this additional part possesses. No recognisable protein domain sequences could be traced within this transcript. However, there is large homology with transcripts in other species, as mouse and rat (Protein blast search at <http://blast.ncbi.nlm.nih.gov>).

The finding of 3 cases with Kleefstra syndrome harbouring interstitial 9q microdeletions encompassing only the novel first exon in the *EHMT1* sequence (encoding 7 amino acids) in addition to several proximally located genes, prompted us to investigate this additional stretch in 75 cases with the clinical suspicion of Kleefstra syndrome but without any *EHMT1* abnormalities found after molecular diagnostics of the previous transcript. Interestingly, in one patient (patient 4), we observed the presence of a small deletion of only the novel first *EHMT1* exon and the first exon of the proximal flanking *C9orf37* gene. Strikingly, in all 4 patients described here the distal breakpoint is located in intron 1 of the *EHMT1* gene. This breakpoint is located between the MLPA probes in intron 1 and exon 2 (chr9:139,633,920-139,725,201; hg18). This 91 kb region harbours several repetitive sequences, such as LINE and SINE. This locus could be considered as locus 4, in addition to the 3 loci described by Yatsenko et al (2009). We hypothesize that this region is the most common distal breakpoint locus involved in KS, but detailed analysis of the breakpoint sequences in all patients should be performed in order to prove this statement.

Though the size of deletions differed in all 4 patients, no significant differences in the degree of intellectual disability or severity of medical problems was seen. This is in agreement with previous observations that haploinsufficiency of the *EHMT1* gene is responsible for the core phenotype of KS (Kleefstra et al., 2006, 2009; Yatsenko et al., 2009). The transcript of *C9orf37* with a so far unknown function, is located antisense to *EHMT1* (head-to-head gene orientation). It is suspected that the *EHMT1* and *C9orf37* genes are activated by the same (bidirectional) promoter. It is not certain whether a deletion of the *C9orf37* transcript has any molecular or clinical consequences. We and others published many cases with 9q deletions encompassing this gene and other proximal flanking genes and they do not have a phenotype significant different from those having deletions located more distally (Kleefstra et al., 2006, 2009; Yatsenko et al., 2009). However, a deletion of only *C9orf37*, has not been identified today but might have an effect on the expression of *EHMT1*. In addition, sequencing of the coding region of *C9orf37* in 30 patients with suggestive clinical Kleefstra syndrome phenotype but negative for *EHMT1* mutations did not reveal any mutations (unpublished observation).

The finding of novel or modified transcripts by *in silico* analyses having important clinical genetic consequences has been recognised before. A significant example is the identification of a previously unidentified *MECP2* open reading frame encoding a new protein isoform relevant to Rett syndrome (Kriaucionis and Bird, 2004; Mnatzakanian et al., 2004). These and our studies emphasise the need to adjust routine diagnostics whenever such modifications of known genes arise. The update of transcripts in the databases however, requires the renaming of the identified

1 mutations accordingly, based on the recommendations for the description of
2 sequence variants (Human Genome Variation Society (HGVS), Den Dunnen and
3 Antonarakis, 2000). A significant departure from these recommendations is to
4 designate the "A" of the ATG startcodon as the number 1 position. Since the new
5 *EHMT1* transcript contains 31 novel amino acids, the previously described mutations
6 should be adjusted by adding 93 nucleotides to the cDNA position and 31 amino
7 acids to the protein position of the mutation.

18 Consequently, the identification of deletions unique for the previously unknown 5' part
19 of the *EHMT1* transcript has implications for the genetic screening of patients
20 suspected of having KS and for studies of the functional significance of *EHMT1*.
21 Since the deletions identified in this paper are not detected by array CGH, MLPA
22 analysis is highly recommended, and we would suggest this is done in those cases
23 previously not found to have any *EHMT1* abnormalities. MLPA analysis covering the
24 *C9orf37* gene and the novel 5' part of *EHMT1* is now commercially available through
25 MRC-Holland (kit P340, [www. MRC-Holland.com](http://www.MRC-Holland.com)). Routine DNA diagnostic testing
26 for new patients with clinical suspicion of Kleefstra syndrome should involve
27 sequencing and, more importantly, MLPA analysis of the entire coding region of the
28 *EHMT1* transcript NM_024757.4.

43 **ACKNOWLEDGEMENTS**

45 The authors wish to thank the participating patients and parents. They also thank
46 Martina Ruiterkamp-Versteeg and Gaby van de Ven-Schobers for expert technical
47 assistance, and Nicole de Leeuw for performing additional array analysis in patient 3.

This work was supported by grants from GENCODYS, a EU FP7 large-scale integrating project grant (Grant agreement no. 241995) (to T.K).

For Peer Review

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

REFERENCES

Den Dunnen JT and Antonarakis SE. 2000. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. Hum Mut 15:7-12.

Kleefstra T, Smidt M, Banning MJG, Oudakker AR, Van Esch H, de Brouwer APM, Nillesen W, Sistermans EA, Hamel BCJ, de Bruijn D, Fryns JP, Yntema HG, de Vries BBA, van Bokhoven H. 2005. Disruption of the gene Euchromatin Histone Methyl Transferase1 (Eu-HMTase1) is associated with the 9q34 subtelomeric deletion syndrome. J Med Genet 42:299-306.

Kleefstra T, Brunner HG, Amiel J, Oudakker AR, Nillesen WM, Magee A, Genevieve D, Cormier-Daire V, van Esch H, Fryns JP, Hamel BCJ, Sistermans EA, de Vries BBA, van Bokhoven H. 2006. Loss-of-function mutations in euchromatin histone methyl transferase 1 (EHMT1) cause the 9q34 subtelomeric deletion syndrome. Am J Hum Genet 79:370-77.

Kleefstra T, van Zelst-Stams W, Nillesen WM, Cormier-Daire V, Houge G, Foulds N, van Dooren M, Willemsen MH, Pfundt R, Turner A, Wilson M, McGaughran J, Rauch A, Zenker M, Adam MP, Innes M, Davies C, González-Meneses López A , Cassalone R, Weber A, Brueton LA, Delicado Navarro A, Palomares Bralo M, Venselaar H, Stegmann SPA, Yntema HG, van Bokhoven H, Brunner HG. 2009. Further clinical and molecular delineation of the 9q Subtelomeric Deletion Syndrome supports a major contribution of *EHMT1* haploinsufficiency to the core phenotype. J Med Genet 46:598-606

- Kriaucionis S, Bird A. 2004. The major form of MeCP2 has a novel N-terminus generated by alternative splicing. *Nucl Acids Res* 32:1818-23.
- Mnatzakanian GN, Lohi H, Munteanu I, Alfred SE, Yamada T, MacLeod PJM, Jones JR, Scherer SW, Schanen NC, Friez MJ, Vincent JB, Minassian BA. 2004. A previously unidentified MECP2 open reading frame defines a new protein isoform relevant to Rett syndrome. *Nat Genet* 36:339-41.
- Stewart DR, Huang A, Faravelli F, Anderlid BM, Medne L, Cipraro K, Kaur M, Rossi E, Tenconi R, Nordenskjold M, Gripp KW, Nicholson L, Meschino WS, Capua E, Quarrell OWJ, Flint J, Irons M, Giampietro PF, Schowalter DB, Zaleski CA, Malacarne M, Zackai EH, Spinner NB, Krantz ID. 2004. Subtelomeric deletions of chromosome 9q: a novel microdeletion syndrome. *Am J Med Genet A* 128:340-51.
- Stewart D, Kleefstra T. 2007. The chromosome 9q subtelomere deletion syndrome. *Am J Med Genet C Semin Med Genet* 145:383-92.
- Verhoeven WMA, Kleefstra T, Egger JIM. 2010. Behavioral Phenotype in the 9q Subtelomeric Deletion Syndrome: A report about two adult patients. *Am J of Med Genet, B Neuropsychiatr Genet* 153:536-41.
- Yatsenko SA, Cheung SW, Scott DA, Nowaczyk MJ, Tarnopolsky M, Naidu S, Bibat G, Patel A, Leroy JG, Scaglia F, Stankiewicz P, Lupski J. 2005.. Deletion 9q34.3 syndrome: genotype-phenotype correlations and an extended deletion in a patient with features of Opitz C trigonocephaly. *J Med Genet* 42:328-35.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Yatsenko SA, Brundage EK, Roney EK, Cheung SW, Chinault AC, Lupski JR. 2009.
Molecular mechanisms for subtelomeric rearrangements associated with the 9q34.3
microdeletion syndrome. Hum Mol Genet 18:1924-36.

Field Code Changed

For Peer Review

FIGURE LEGENDS

Figure 1: photographs of patient 1 (**A,B**) at the age of 3 years, patient 2 (**C,D**) at the age of 12 years, patient 3 (**E**) at the age of 3 years, and patient 4 (**F** at the age of 3 years; **G** at the age of 3 months). A variety of facial dysmorphisms characteristic for Kleefstra syndrome are seen comprising midface hypoplasia, synophrys, upslanting palpebral fissures, eversion of lower lip, protruding tongue and downturned corners of the mouth.

Figure 2: RT-PCR analysis of *EHMT1* performed in lymphoblastoid cell lines of patients 1 and 3, showing mono-allelic expression of the *EHMT1* transcript in RNA. Sequencing results in patient 1 shows heterozygosity for variant g.139,731,389G>A in DNA (**A**), whereas the RNA only shows the A allele (**B**). Sequencing results in patient 3 show heterozygosity for rs1129768 in DNA (**C**) and only one allele in RNA (**D**).

Figure 3: Part of the 9q subtelomeric region indicating schematically the different *EHMT1* transcripts; The translation start site of the novel *EHMT1* transcript (NM_024757.4) is located 97.6 kb proximal to the 'old' ATG start codon of transcript (NM_024757.3).

Figure 4: MLPA analysis with MRC-Holland kit P340-A1 in patients 1-4 (filled symbols) and in 3 control individuals (open symbols). The deletion in patients 1-3 comprises *C9orf37* and the first exon of *EHMT1*. The proximal part of the deletion (detected by array) is not tested in this MLPA analysis. Patient 4 shows the smallest

1
2 deletion, comprising *C9orf37* exon 1, and *EHMT1* exon 1. The probes in *C9orf37*
3
4 exon 2 and *EHMT1* exon 2 show a normal dosage (proximal probe: exon 2.1; distal
5
6 probe: exon 2.2). [The probes in exon 1 and exon 2.1 of *EHMT1* are actually located](#)
7
8 [in intron 1 at positions +625 and -15, respectively.](#) The control probes are located on
9
10 other chromosomes (15q26 and 20q13, respectively).
11
12
13

14 Figure 5:
15
16 The 9q subtelomeric region, with a schematical overview of the deletions identified in
17
18 patients 1-4 (figure not to scale). For all patients, the first preserved and deleted
19
20 probes are indicated (**A**). Agilent probes are shown in bold, and MLPA probes in
21
22 italics. The remainder of the probes are from the Affymetrix platform. The position of
23
24 the 5' and 3' ends of the genes, as well as the location of the probes (the latter in
25
26 bold) on build NCBI 36, UCSC hg18, Mar 2006 is depicted in figure **B**. The 5' parts of
27
28 the reference genes in the region are also depicted in build NCBI 37, UCSC hg19,
29
30 Feb 2009 (**C**). The orientation and location of the genes (**D**), and the minimal size
31
32 (filled grey) and maximum size (shaded grey) and location of the deletion in patients
33
34 1-4 (**E**) is schematically drawn (not to scale).
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Figure 1

photographs of patient 1 (A,B) at the age of 3 years, patient 2 (C,D) at the age of 12 years, patient 3 (E) at the age of 3 years, and patient 4 (F at the age of 3 years; G at the age of 3 months). A variety of facial dysmorphisms characteristic for Kleefstra syndrome are seen comprising midface hypoplasia, synophrys, upslanting palpebral fissures, eversion of lower lip, protruding tongue and downturned corners of the mouth.

184x127mm (400 x 400 DPI)

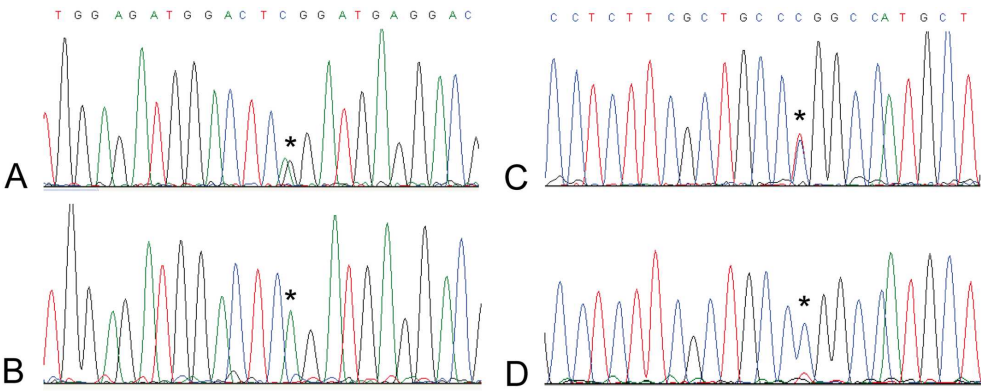


Figure 2: RT-PCR analysis of EHMT1 performed in lymphoblastoid cell lines of patients 1 and 3, showing mono-allelic expression of the EHMT1 transcript in RNA. Sequencing results in patient 1 shows heterozygosity for variant g.139,731,389G>A in DNA (A) , whereas the RNA only shows the A allele (B). Sequencing results in patient 3 show heterozygosity for rs1129768 in DNA (C) and only one allele in RNA (D).
183x72mm (600 x 600 DPI)

Figure 3

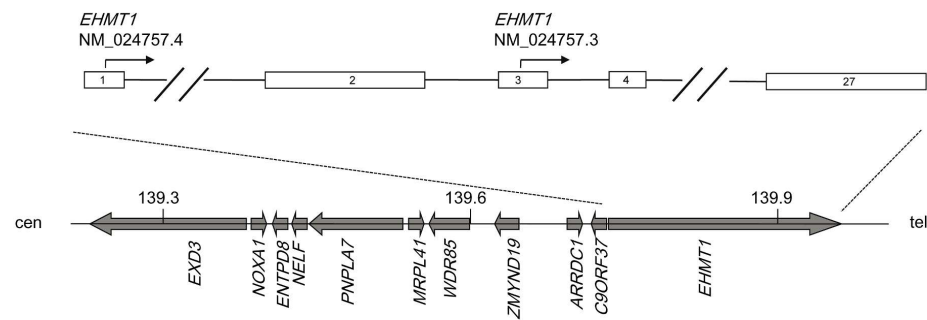


Figure 3: Part of the 9q subtelomeric region indicating schematically the different EHTM1 transcripts; The translation start site of the novel EHTM1 transcript (NM_024757.4) is located 97.6 kb proximal to the 'old' ATG start codon of transcript (NM_024757.3).
251x109mm (600 x 600 DPI)

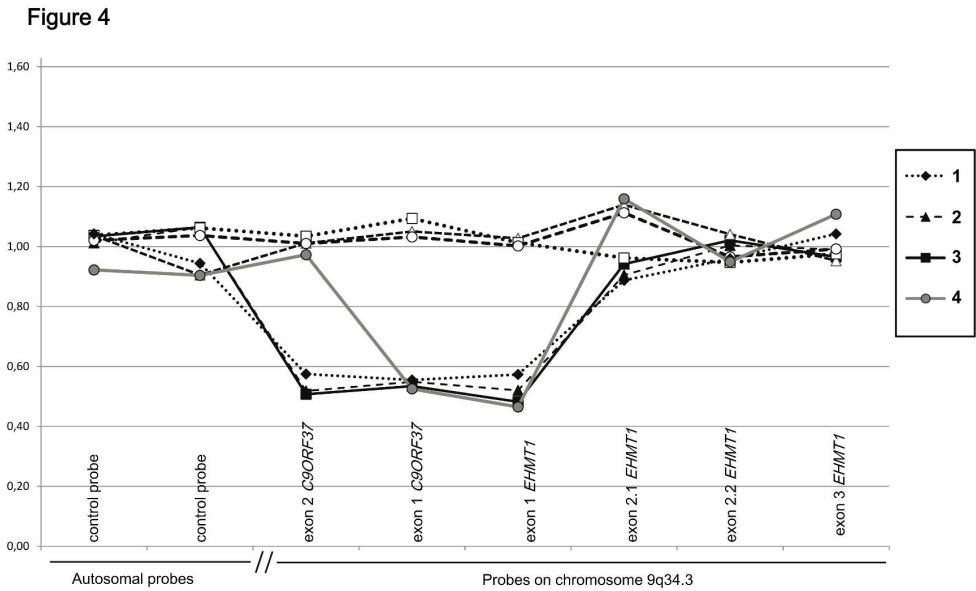
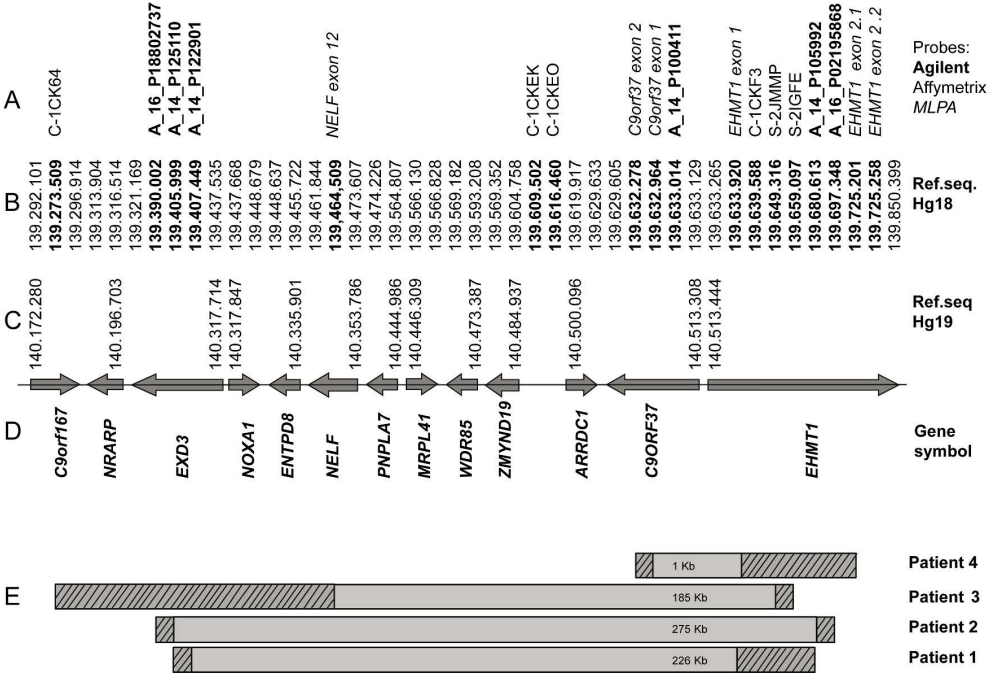


Figure 4: MLPA analysis with MRC-Holland kit P340-A1 in patients 1-4 (filled symbols) and in 3 control individuals (open symbols). The deletion in patients 1-3 comprises C9orf37 and the first exon of EHMT1. The proximal part of the deletion (detected by array) is not tested in this MLPA analysis. Patient 4 shows the smallest deletion, comprising C9orf37 exon 1, and EHMT1 exon 1. The probes in C9orf37 exon 2 and EHMT1 exon 2 show a normal dosage (proximal probe: exon 2.1; distal probe: exon 2.2). The probes in exon 1 and exon 2.1 of EHMT1 are actually located in intron 1 at positions +625 and -15, respectively. The control probes are located on other chromosomes (15q26 and 20q13, respectively).
149x90mm (600 x 600 DPI)



The 9q subtelomeric region, with a schematical overview of the deletions identified in patients 1-4 (figure not to scale). For all patients, the first preserved and deleted probes are indicated (A). Agilent probes are shown in bold, and MLPA probes in italics. The remainder of the probes are from the Affymetrix platform. The position of the 5' and 3' ends of the genes, as well as the location of the probes (the latter in bold) on build NCBI 36, UCSC hg18, Mar 2006 is depicted in figure B. The 5' parts of the reference genes in the region are also depicted in build NCBI 37, UCSC hg19, Feb 2009 (C). The orientation and location of the genes (D), and the minimal size (filled grey) and maximum size (shaded grey) and location of the deletion in patients 1-4 (E) is schematically drawn (not to scale).

149x103mm (600 x 600 DPI)