CDK13-related disorder: Report of a series of 18 previously unpublished individuals and description of an epigenetic signature

Flavien Rouxel, Raissa Relator, Jennifer Kerkhof, Haley Mcconkey, Michael Levy, Patricia Dias, Mouna Barat-Houari, Nathalie Bednarek, Odile Boute, Nicolas Chatron, et al.

CDK13-related disorder: report of a series of 18 previously unpublished individuals and description of an epigenetic signature

Flavien Rouxel[1*], Raissa Relator[2*], Jennifer Kerkhof[2], Haley McConkey[2], Michael Levy[2], Patricia Dias[3], Mouna Barat-Houari[4], Nathalie Bednarek[5], Odile Boute[6], Nicolas Chatron[7] Florian Cherik[8], Andrée Delahaye-Duriez[9], Martine Doco-Fenzy[10], Laurence Faivre[11,12], Lucas W. Gauthier[7], Delphine Heron[13], Michael S. Hildebrand[14,15], Gaëtan Lesca[7], James Lespinasse[16], Benoit Mazel[11], Leonie A. Menke[17],Angela T. Morgan[18], Lucile Pinson[1],Chloe Quelin[19], Massimiliano Rossi[7,20], Nathalie Ruiz-Pallares[4], Frederic Tran-Mau-Them[12,21], Imke N. Van Kessel[17], Marie Vincent[22], Mathys Weber[11], Marjolaine Willems[1], Gwenael Leguyader[23] ,Bekim Sadikovic[2,24&], David Genevieve[1&].

* These authors contributed equally to this work

& These authors contributed equally to this work

1.Montpellier University, ERN ITHACA, Génétique clinique, Département de Génétique Médicale, Maladies Rares et Médecine Personnalisée, CHU Montpellier, Centre de référence anomalies du développement SOOR, INSERM U1183, Montpellier, France.

2. Verspeeten Clinical Genome Centre; London Health Sciences Centre, London, ON N6A 5W9, Canada.

3. Genetics department, hospital center of Lisbon North, Lisbon, ERN ITHACA, Portugal

4. Département de Génétique Médicale, Maladies Rares et Médecine Personnalisée, Génétique des Maladies Rares et Auto-inflammatoires, CHU Montpellier, Université de Montpellier, France.

5. Genetics department, CHU Reims, Medical school IFR53, EA3801, Reims, France

6. Genetics department, Guy Fontaine Medical Center, CLAD Nord de France, Jeanne de Flandre Hospital, CHRU Lille, Lille, France

7. Genetics department, Lyon University Hospital, and Institut NeuroMyoGène. CNRS UMR 5310 - INSERM U1217, Claude Bernard Lyon 1 University, Bron, France

8. Genetics department, CHU Clermont-Ferrand, Clermont-Ferrand, France

9. Department of Histology Embryology and Cytogenetics, Jean Verdier Hospital; Paris 13 University, Sorbonne Paris Cité, UFR SMBH Bobigny; PROTECT, INSERM, Paris Diderot University, Bondy, France.

10. Genetics department, CHU Reims, Medical school IFR53, EA3801, Reims, France

11. Centre de Référence Anomalies du Développement et Syndromes Malformatifs, FHU TRANSLAD, CHU Dijon, Dijon, France

12. INSERM - Bourgogne Franche-Comté University, UMR 1231 GAD team, Genetics of Developmental disorders, Dijon, France

13. Genetics department, Hospital Pitié-Salpétrière, Paris, France

14. Epilepsy Research Center, Department of Medicine, The University of Melbourne, Austin Health, Heidelberg, Australia

15. Murdoch Children's Research Institute, Melbourne, Australia

16. Chromosomic genetic laboratory, CH Général, Chambéry, France

17. Department of Pediatrics, Emma Children's Hospital, Amsterdam UMC, University of Amsterdam, Amsterdam, Netherlands, ERN ITHACA

18. Department of Audiology and Speech Pathology, The University of Melbourne, Australia

19. Genetics department, CHU de Rennes, Rennes, France

20. Genetics department, Referral centre for developmental abnormalities, Lyon University Hospital, and INSERM U1028, CNRS UMR5292, Lyon Neuroscience Research Centre, GENDEV Team, Claude Bernard Lyon 1 University, Bron, France

21. Functional unit 6254 Innovation in genomic diagnosis of rare diseases, CHU Dijon Bourgogne, Dijon, France

22. Genetics department, CHU de Nantes, Nantes, France

23. Genetics department, CHU de Poitiers, Poitiers, France

24. Department of Pathology and Laboratory Medicine, Western University, London, ON N6A 3K7, Canada.

Corresponding authors:

Pr David GENEVIEVE

Montpellier University, Département de génétique médicale, maladies rares et médecine personnalisée, Centre de référence anomalies du développement et syndromes malformatifs, Plateforme recherche de microremaniements chromosomiques, CHU Arnaud de Villeneuve, Faculté de Médecine Montpellier-Nîmes, ERN ITHACA, France

Tel: 00334 67 33 67 33 Fax: 00334 67 33 60 52 d-genevieve@chu-montpellier.fr

Dr Bekim Sadikovic

Department of Pathology and Laboratory Medicine, Western University,

London, ON N6A 3K7, Canada.

Phone: +1 5196858500 53074

Email: bekim.sadikovic@lhsc.on.ca

Abstract:

Purpose: Rare genetic variants in *CDK13* are responsible for CDK13-related disorder (CDK13-RD), with main clinical features being developmental delay or intellectual disability, facial features, behavioral problems, congenital heart defect and seizures. Here we report 18 novel individuals with CDK13-RD and provide characterization of genome-wide DNA methylation.

Methods: We obtained clinical phenotype and neuropsychological data for 18 and ten individuals respectively and compared this series to the literature. We also compared peripheral blood DNA methylation profiles in individuals, controls and other neurodevelopmental disorders episignatures. Finally, we developed a support vector machine (SVM) based classifier distinguishing CDK13-RD and non-CDK13-RD samples.

Results: We reported health and developmental parameters, clinical data and neuropsychological profile of individuals with CDK13-RD. Genome-wide differential methylation analysis revealed a global hypomethylated profile for CDK13-RD-affected individuals in a highly sensitive and specific model that could aid in reclassifying variants of unknown significance.

Conclusions: We describe the novel features anxiety disorder, cryptorchidism and disrupted sleep in CDK13-RD. We define a CDK13-RD DNA methylation episignature as a diagnostic tool and a defining functional feature of the evolving clinical presentation of this disorder. We also demonstrate overlap of the CDK13 DNA methylation profile in an individual with a functionally and clinically related CCNK-RD.

Introduction

Cyclin-dependent kinase 13(*CDK13*)-related disease was first reported in 2016 with exome sequencing of individuals with congenital heart disease, revealing seven individuals with a *CDK13* pathogenic variant [1]. Since then, pathogenic variants in *CDK13* were found to cause CDK13-related disorder (CDK13-RD), also known as CHDFIDD (for congenital heart defects, dysmorphic facial features, and intellectual developmental disorder; OMIM 617360), a rare genetic disorder characterized by developmental delay or intellectual disability, recognizable facial features, behavioral problems, and feeding difficulties, sometimes secondary to gastroesophageal reflux, congenital heart defects, audition loss, speech and language disorder and seizures [2-7].

CDKs are proteins that regulate the cell cycle and transcription in higher eukaryotes. CDK13 (OMIM 603309) forms a complex with cyclin K, coded by *CCNK*, to phosphorylate certain serine residues of the RNA polymerase II C-terminal domain. This phosphorylation is tightly linked to the phases of RNA polymerase II-mediated transcription [8]. This allows CDK13 to control the processing activity of RNA polymerase II and regulate growth signaling pathways. CDK13 has also been shown to share regulation of DNA damage-associated pathways with CDK12 [9]. It may have other roles in regulating gene transcription [10], but more studies are needed to assess its exact functions.

Recently, some diseases affecting DNA methylation were found to have specific methylation signatures, called episignatures [11;13]. Episignature analysis has recently been implemented with the diagnostic clinical genomic DNA methylation test EpiSign, in individuals with rare disorders, thus providing strong evidence of clinical utility including the ability to provide conclusive diagnostic findings in most individuals tested [14]. Abnormal transcription may have indirect consequences on DNA methylation. Therefore, individuals with CDK13-RD may have different DNA methylation than other individuals.

This study aimed to better delineate the clinical presentation of CDK13-RD by performing a literature review to describe the clinical phenotype of 44 previously published individuals and by describing an additional 18 previously unpublished individuals with *CDK13* pathogenic variants, and to investigate evidence of an epigenetic signature for CDK13-RD.

Materials and methods

Clinical data

Ascertainment

Individuals from this series with CDK13-RD due to rare *CDK13* variants were mainly identified by clinical exome sequencing after being referred to genetic centers for an indication of neurodevelopmental delay.

Previously reported individuals were retrieved by using the terms “CDK13” OR “CDK13-related disorder” OR “CHDFIDD” in articles indexed in PubMed between inception and February 28, 2021. The full texts of articles were reviewed. The criteria for including published individuals in the present study were 1) availability of clinical details, 2) unambiguous description of the *CDK13* variant and 3) not duplicated in any previous report.

Data for 18 previously unreported individuals with heterozygous *CDK13* pathogenic variants were examined and compared to a series of 44 individuals from the literature.

We also assembled front and lateral view pictures for 13 CDK13-RD individuals and profile pictures for 11 CDK13-RD individuals. We created a facial mask combining all of our front pictures by using Face2Gene (https://www.face2gene.com/) (Figure 1).

We also identified an individual with a *CCNK* variant, which we decided to include in the methylation study because of phenotypic similarities with CDK13-RD individuals and because CDK13 and cyclin K are a part of the same functional protein complex [1].

Data and sample collection (Supplementary table S1)

Data were collected by different means; by using clinical networks (Groupe DI France, AnDDI-Rares (http://anddi-rares.org/), ERN ITHACA (https://ern-ithaca.eu/)), and by directly contacting clinicians from all centers in France and Australia via a larger *CDK13*-RD reverse-phenotyping project. Clinical phenotypes were evaluated by analysing health, medical and neuropsychological data systematically collected by using a standardized form sent to collaborators. Written consent was obtained from all individuals or their legal guardians.

This research protocol was validated by the institutional review board of the University hospital of Montpellier (29/05/2020, IRB-MTP_2020_05_202000466); ClinicalTrials.gov identifier: NCT04382573; the Western University Research Ethics Board (REB 106302); and the Australian Human Research Ethics Committee (HREC 37353). Procedures were in accordance with local ethical standards.

For each individual, we obtained a DNA sample from previous genetic studies, a minimum sample of 0.5µg (1.76e-8oz) of DNA from peripheral blood, with information on the individual's sex and age at the time of sample collection.

Statistical analyses

For calculating of frequencies of features, we excluded individuals for whom that feature was coded as "unknown" (which includes when the presence or absence of a particular feature was not clearly documented or a feature may not be applicable owing to the individual's sex or age) in the clinical form (Supplementary table S2). Absolute and relative frequencies (expressed as n [%], n being the number of individuals with the feature and percentage was calculated as n/number of individuals with information on the clinical feature) were used for describing categorical variables, and median (m), interquartile range (IQR) and range for

describing continuous variables. Medians and IQRs were calculated by using Microsoft Excel functions.

Composite gestalt of individuals with a pathogenic variant in *CDK13*

We used the Face2Gene online tool to create a computed gestalt for individuals with CDK13-RD based on facial pictures of this series of individuals.

DNA methylation analysis

Individuals in the study

The initial series of individuals included in the genome-wide DNA methylation analysis included eight males and nine females, with age ranging from 3.5 to 16 years at the time of sample collection, and who had a clinical diagnosis of CHDFIDD and carried *CDK13* missense variants classified as pathogenic or likely pathogenic according to the American College of Medical Genetics /Association for Molecular Pathology (ACMG/AMP) guidelines and . One individual (individual C2) with a pathogenic variant in CDK13 was included only in the clinical description, but not in the methylation analysis because no usable sample could be obtained. Four additional individuals (two males and two females; age 1.5-7 years old) also diagnosed with CHDFIDD were also included in a secondary analysis after the differential methylation analysis of the initial series of 17 individuals but not in the clinical description: two individuals (IDs B8 and B9) had large duplications in *CDK13*, one individual (ID A5) carried a missense variant of unknown significance (VUS) and one (ID B5) had a missense variant in *CCNK* (OMIM 603544), which is also a VUS. DNA samples were collected as described above, and all individuals were given pseudonym substitutes for record de-identification. All information about the type of data collected for each individual can be found in Supplementary table S1.

Methylation data preprocessing

DNA was extracted from peripheral blood previously sampled from all individuals in the series and subjected to bisulfite conversion by using the Illumina Infinium MethylationEPIC BeadChip kit in accordance with the manufacturer's protocol. The resulting data files (Illumina IDAT files) containing the methylated and unmethylated signal intensities as detailed in previous works [11,14-15] were subjected to methylation analysis. Quality control of methylation data involved using the minfi R package [16]. Standard preprocessing with background correction and reference control normalization for Illumina microarrays was implemented, and beta values were extracted. Density plots and concordance between recorded and predicted sex and age were evaluated. Thereafter, probes with at least one of the following conditions were removed: detection p-value > 0.1, located in the X or Y chromosome, targeting a CpG site that overlaps a known single nucleotide polymorphism, or cross-reactive, which resulted in 776314 probes. In addition,we excluded existing batches in our database that were previously identified to cause batch effects as well as samples with > 5% failed probes. A total of 260 samples were removed after this filtering step.

Sample selection

The 17 study samples with ACMG/AMP category 1 and 2 variants were used for methylation signature discovery, and the other four samples were used for assessing use of the signature for diagnostics and VUS reclassification. Matched unaffected control samples were selected from the London Health Sciences Centre EpiSign Knowledge Database (EKD) [11,14] according to age, sex and array type by using the MatchIt package in R [17]. This step resulted in a total of 68 samples for training data, 17 cases and 51 controls, with case-control ratio 1:3. The existence of batch effects and outliers in the original cohort data and the matched case-control training data were inspected by using principal component analysis (PCA) as shown in Supplementary figure S1, but no outliers and no batch structure were detected.

Episignature discovery

Probe selection for defining the methylation profile for CDK13-RD involved using data for matched cases and controls. Differentially methylated probes (DMPs) between the two groups were determined by initially fitting a linear regression model with use of the limma R package [18]. We used the methylation M-values converted from beta values (log2(beta)/(1-beta)) as predictors, case/control labels as the response and estimated blood cell proportions as confounding variables. Moderated t-statistics and their corresponding p-values for the fitted model were computed with an empirical Bayes method and controlled for 5% false discovery rate with the Benjamini-Hochberg procedure. The top 700 statistically significant DMPs with at least 5% methylation difference between cases and controls were considered. Individual variable importance was calculated with the filterVarImp function from the caret R package [25], which implements receiver operating characteristic curve analysis on each probe, and the 140 most differentiating probes were identified.

Unsupervised and supervised analysis

Separation of groups using the selected probes from previous analysis was examined with unsupervised analysis by hierarchical clustering and multidimensional scaling (MDS). Furthermore, the discrimination power of the chosen probes specific to CHDFIDD was assessed by building a prediction model with a linear support vector machine (SVM) in the e1071 R package [19]. The supervised classifier was trained by using all 17 study samples included in the signature probe selection and select samples from the EKD database: 75% of all controls and 75% of the individuals in the 38 neurodevelopmental disorder (NDD) cohort. Consequently, the respective remaining 25% and the supplemental four subjects in the study cohort were used as test data. Prediction outputs of the classifier are probability scores, called methylation variant pathogenicity (MVP) scores [10], ranging from 0 to 1, that represent the degree of confidence in the predicted class. Finally, differentially methylated regions (DMRs)

were identified as regions with at least 5 CpGs within 1 kbp with at least 10% methylation difference between classes, with Fisher's multiple comparison $p < 0.01$, by using the DMRcate package [20].

Results

Clinical results

Clinical delineation of CHDFIDD

Table 1 summarizes the data for this series and the literature

Frequent morphological findings across both novel and previously reported cases included anomalies of ear morphology, a wide nasal base, hypertelorism and epicanthic folds.

Novel participants

Across this series, growth parameters were in the low normal range, with mean height -0.58 SD, mean weight -0.05 SD and mean head circumference -1 SD (Figure 2A). In terms of general development, 17/18 individuals showed global developmental delay or intellectual disability. Motor skills were markedly delayed: mean age at walking of 26 months. Similarly, first words were further delayed: median, 35 months (Figure 2B).

Beyond the development of first words, communication remained delayed, with five individuals (aged 8 to 17) using sentences, but six (aged 6 to 14 years) not (typical age of sentence use is 3-4 years). One child had a diagnosis of childhood apraxia of speech, as previously reported [7]. Similarly, beyond early walking milestones, motor impairments remained: 15/18 individuals presented with at least one motor feature, most frequently axial hypotonia. Other motor features included hyperreflexia, tremor, dysmetria and gait pattern. Most individuals (14/18) had at least one ophthalmological issue, including strabismus, astigmatism and myopia. Psychiatric comorbidity was also frequent, with one feature seen in 12/18 individuals: anxiety disorder, attention deficit hyperactivity disorder, aggression and autism spectrum disorder: 11/18 individuals had at least one other neurological feature, including disrupted sleep, dysphagia (feeding difficulties) and sensorineural hearing loss.

Another significant and frequently encountered medical issue was recurrent infections, present in 6/18 individuals. Also, 6/8 males had cryptorchidism.

All clinical data are detailed in Supplementary table S2.

Previously reported cases

We gathered clinical data for 44 individuals across 6 publications (see Supplementary Table S3) [1-6]. If data were left as a blank space in a publication, it was not counted in the statistics.

When adding data for our series to that for previously published individuals with *CDK13* pathogenic variations, growth parameters remained in the low normal range: mean height -0.88 SD, mean weight -0.22 SD and mean head circumference -0.69 SD(Figure 2C).

Neuropsychological studies

We gathered data from ten neuropsychological studies. Overall, cognitive level was very variable in individuals with a *CDK13* pathogenic variant.

Intellectual quotient (IQ) data were available for ten participants; full-scale IQ for six participants ranged from 59-71 to 84. Four individuals had heterogeneous indexes, so a full-scale IQ could not be accurately calculated according to the recommendations of the manufacturer.

For individuals with a pathogenic variant in *CDK13*, the verbal comprehension index seemed the most preserved field of competence.

Neuropsychological data are fully described in Supplementary table S4.

Epigenetic signature

A hypomethylated signature profile distinguishes CDK13-RD from unaffected controls

We included 17 individuals clinically diagnosed with CHDFIDD and carrying pathogenic and likely pathogenic missense variants for identifying a methylation profile specific to CDK13-RD. Detailed clinical and molecular descriptions of these individuals are given in Supplementary Table S2. Genome-wide DNA methylation analysis of these 17 individuals versus 51 matched unaffected and unrelated controls revealed a DNA methylation episignature including 140 DMPs that most effectively segregated cases from controls (Supplementary Table S5). As compared with controls, for cases, all probes were severely hypomethylated, exhibiting approximately 25% methylation (Figure 3A). Hierarchical clustering and MDS analysis of the selected probes confirmed the robustness of the methylation profile in differentiating the groups (Figure 3B, C), which reveals a clear separation of the two classes. Leave-one-out cross-validation on the case samples also supported this claim: *CDK13* samples were always correctly grouped together, and their profiles were visibly distinct from that of controls (Supplementary Figure S2). We also identified differentially methylated regions (regions including multiple contiguous probes with significant methylation difference) (Supplementary Table S6). Consistently, the case samples exhibited hypomethylated levels for all 107 DMRs detected (Supplementary Figure S3).

CHDFIDD case with CCNK variant exhibits similar episignature

The methylation pattern specifically identified for CDK13-RD was used to test the additional 2 CHDFIDD samples with large *CDK13* duplications, one sample with an unclassified *CDK13* missense, and one sample with an unclassified *CCNK* missense variant. A similar analysis of these additional individuals using the selected 140 DMPs revealed that individuals carrying large duplications in the *CDK13* gene did not exhibit the same signature as the training cohort. Hierarchical clustering and MDS analysis revealed that these two samples grouped with the unaffected controls. The individual carrying a *CDK13* missense VUS also

demonstrated negative results, which suggests evidence against the pathogenicity of this variant. Of note, the individual with the *CCNK* missense VUS clustered with the *CDK13* cohort and had an identical DNA methylation profile, as evident in the heatmap and the MDS plots (Figures 4A, B).

To investigate the use of the identified episignature as a diagnostic tool, a linear SVM model was developed and trained using data for 17 individuals in the series of unaffected controls and individuals known to be affected by different neurodevelopmental and rare disorders (NDD/RD) included in the EKD. LOOCV was implemented on the case samples to validate the reproducibility of the episignature and evaluate the model. All cases had MVP scores > 0.5, whereas non-CDK13 samples all had very low MVP scores. These scores are summarized in Supplementary Figure S2, which reveals a highly sensitive and specific model. In addition, the whole training data (cases, matched controls and NDD/RD samples) underwent Monte Carlo cross-validation, a repeated random subsampling cross-validation, for ten times to estimate overall classifier performance. The cross-validation results showed an average of 100% accuracy, which further supports the robustness of our episignature-based predictive model. Performance of the model was additionally examined with samples A5, B5, B8 and B9, and EKD cohorts. All control and NDD/RD samples had low probability scores, which implies that the classifier can predict with high specificity (Figure 4C). Results for the four CHDFIDD test samples also agreed with the outcomes of the previous unsupervised analysis. Calculated MVP scores were low for samples B8 and B9 with large duplications, as well as sample A5 with the *CDK13* missense VUS. However, the individual with the *CCNK* missense VUS had a positive MVP score.

Discussion

CDK13-RD is a rare monogenic disorder, with 44 individuals reported in the literature thusfar. Here we expand the clinical phenotype by describing a series of 18 previously

unreported individuals, providing novel clinical signs to aid future detection of the condition in others, and demonstrate evidence of a unique DNA methylation episignature for CDK13-RD.

Clinical description

Most clinical features had a similar distribution in this series and in the literature.

Growth parameters seemed to be mostly in the low normal range, and psychomotor development was delayed for all individuals. However, few individuals seemed to have developmental regression. Special attention must be provided for individuals with a *CDK13* pathogenic variant, regarding psychomotor development, with the help of a physiotherapist, a psychomotrician and speech therapist if needed.

Neurodevelopmental disorders such as attention deficit hyperactivity disorder and autism spectrum disorder were present in 11/33 and 14/46 individuals when adding data from this series and the literature. Other psychiatric comorbidities notably included anxiety disorder, which is a new finding from this study and was observed in half of individuals (9/18).

Other neurological findings included axial hypotonia for 13/18 individuals of our series, sensorineural hearing loss and epilepsy. A new clinical finding in this study was prevalence of disrupted sleep in 8/18 individuals of our series. Additional studies are needed to identify which care will be efficient.

We noted epilepsy in only one individual in this series, which is less than previously observed. However, the P-value was not significant based on Fisher's exact test (p=0.13)

For neuropsychological studies, verbal comprehension seemed the most preserved field of competence, which is strange considering that individuals with CDK13-RD often have a delay in expressive language. This could be a strength, but will require more studies to confirm.

Regarding ophthalmological features, strabismus was frequent when adding our series and the literature (25/61), and we also found 6/18 individuals with astigmatism in this series, which leads us to believe that a thorough ophthalmological examination with regular evaluations seem important.

Congenital heart malformations were present in 27/59 individuals when adding this series and the literature. According to these data, we recommend a cardiac sonography for every individual diagnosed with CDK13-RD.

Constipation was observed in 10/18 individuals and gastroesophageal reflux disease was observed in 4/18 individuals with CDK13-RD in our series.

A new finding is the presence of cryptorchidism in 6/8 males with CDK13-RD in our series. We did not find cryptorchidism in previously published individuals. Cryptorchidism could be searched for in males with CDK13-RD.

Facial features included, from most to least frequent, anomalies of ear morphology, wide nasal root, epicanthic folds, curly hair, upslanting palpebral fissures and short columella. Other less frequent clinical signs included deep set eyes, arched eyebrows and blepharophimosis. Deep-set eyes were significantly more frequent in this series than in previous studies, perhaps because investigators were not asked specifically to look for deep-set eyes in previous studies.

We also found hemangiomas of different localizations in 11/49 individuals with CDK13-RD when adding results from our series and the literature, so a thorough dermatologic evaluation could be of interest for the clinical diagnosis of CHDFIDD. In fact, when seeing an individual with clinical signs evoking CDK13-RD, the finding of a hemangioma is another clue to this diagnosis.

Immunological complications, mostly recurrent otitis media, were described in 11/27 individuals when adding results from our series and the literature. According to a previous publication, [21] CDK13 seems to play on immunity, with *CDK13* described as a virus-regulating factor regulating viral mRNA splicing. Other studies could help better delineate the immune phenotype in CDK13-RD.

We also describe an individual with a *CCNK* pathogenic variant whose DNA methylation profile matched those of individuals with *CDK13* pathogenic variants. The identification of a shared episignature between *CDK13* and *CCNK* is probably related to these two genes coding for two partner proteins that were shown to interact by forming a complex [8,22] that regulates phosphorylation of Ser2 in the C-terminal domain of RNA polymerase II and expression of a small subset of human genes. The consequences of the transcription modifications are similar between CDK13 and CCNK, which would lead to a similar abnormal DNA hyper- or hypomethylation of probes. Individuals with a *CCNK* pathogenic variant also shared several clinical features with CDK13 individuals, namely facial features including hypertelorism, thin eyebrows, abnormal ears, broad nasal bridge and tip, thin upper vermilion and narrow jaw, although only four individuals with such variants have been reported so far [23]. Additional studies are needed to conclude on a common epigenetic signature between *CCNK* and *CDK13* as well as overlapping clinical features

Epigenetic signature

Currently, the diagnosis of CDK13-RD is established when a heterozygous pathogenic variant in *CDK13* is identified by molecular testing. However, one of the main difficulties for confirming the diagnosis is the classification of a VUS. We previously demonstrated the clinical utility of the EpiSign classifier, which is based on this technology, to effectively reclassify variants of unknown clinical significance and help resolve the clinical diagnosis in a growing number of Mendelian disorders [11;14-15] Determining the clinical relevance of a

CDK13 VUS identified by NGS sequencing can be challenging and time-consuming, which indicates the need for reliable biomarkers to support the diagnosis and validate the significance of variants.

The methylation profile for CDK13-RD that we have defined here consists of a relatively small number of hypomethylated CpG sites (<150), and is variant type-specific, characterizing pathogenic missense variants, among others. However, it has been found to be a robust and effective discriminator of individuals with CDK13-RD versus individuals with other NDDs, as well as healthy controls. Aside from a case-control prediction with high specificity, using the signature revealed possible sharing of epigenetic markers between individuals with *CDK13* and *CCNK* variants who, notably, also present similar phenotypic traits. The identified 140 DMPs overlap with 116 unique genes (Supplementary Table S5). Functional annotation analysis using DAVID [24] showed that this set is enriched (adjusted $p < 0.01$) with genes whose molecular functions are related to (poly(A)) RNA-binding and protein binding and have cellular component ontology terms including nucleoplasm and nucleus. Furthermore, UniProt key terms indicate a significant subset of the list as phosphoproteins, and proteins that are altered by attachment to at least one acetyl or methyl group post-translation. In addition to the episignature biomarker, 107 differentially hypomethylated regions were identified, overlapping 145 unique genes (Supplementary Table S6). Similar to the DMP genes, this set is enriched with phosphoproteins and proteins with post-translational modifications affected by acetylation or methylation ($p < 0.01$), as well as ubiquitin-like modifier proteins ($p = 0.03$) and isopeptide bonds ($p = 0.047$). Molecular function and biological processes Gene Ontology enriched terms include RNA binding ($p = 0.024$) and protein biosynthesis ($p = 0.033$). These results indicate possible underlying pathophysiological processes associated with CDK13-RD, warranting further investigations.

In conclusion, we define a CDK13-RD DNA methylation episignature as a diagnostic tool and a defining functional feature of the evolving clinical presentation of this disorder. We also demonstrate overlap of a CDK13 DNA methylation profile in an individual with a functionally and clinically related CCNK-related disorder. We also add new features of anxiety disorder, disrupted sleep and cryptorchidism in individuals of our series and establish hemangioma as an important diagnostic clue for CDK13-RD based on our series and the literature.

The ongoing research in episignature syndromes continues to reveal novel episignatures related to various genetic conditions [11]. The expanding complexity of the reference databases and increased sample numbers are revealing additional episignature disorders as well as an increasing complexity, including protein network-specific, gene-specific, sub gene and domain-specific episignatures. Of note, limitations including the use of peripheral blood as well as targeted nature of data analysis (EPIC arrays focus on subset 850K CpG probes) may limit the ability of this approach to distinguish some disorders and not others. Addressing these questions and assessing the potential for episignature detection across the span of Mendelian conditions will require systematic expansion of databases, refinement of analytical tools and bioinformatic approaches as well as more comprehensive genomic methylation analyses.

An ongoing research trial (EpiSign-CAN), which is a national level health systems impact study, is designed to answer whether clinical episignature analysis is better suited as a primary screen for the expanding list of related disorders in the developmental delay/neurodevelopmental disorder patient population, or as a reflex post-extensive genomic analysis. The study is currently recruiting 2000 patients to the two arms and should provide empirical data addressing this question. Preliminary findings in the initial cohort of >200 patients that are being tested clinically by EpiSign analysis showed that 57 (27.6%) were

positive for a diagnostic episignature including 48/136 (35.3%) in the targeted cohort and 8/71 (11.3%) in the screening cohort [14]. Whether this technology proves optimal for use as a primary screen or a reflex test, or in combination with other genomic assessments including exome/genome sequencing (because it is comparatively low cost with potential for additive diagnostic yield), the primary objective is to facilitate the identification of the underpinning genetic defects. However, our group has now observed multiple instances (data not shown) of patients in whom the only molecular finding supporting the clinical diagnosis is the diagnostic episignature, which leads to an intriguing question of the utility of an episignature as a molecular diagnosis in and of itself.

References

1. Sifrim A, Hitz M-P, Wilsdon A et al. Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. Nat Genet. 2016;48(9):1060-5.
2. Bostwick BL, McLean S, Posey JE, Streff HE, Gripp KW, Blesson A, et al. Phenotypic and molecular characterisation of CDK13-related congenital heart defects, dysmorphic facial features and intellectual developmental disorders. Genome Med. 14 2017;9(1):73.
3. Van den Akker WMR, Brummelman I, Martis LM et al. De novo variants in CDK13 associated with syndromic ID/DD: Molecular and clinical delineation of 15 individuals and a further review. Clin Genet. 2018;93(5):1000-7.
4. Uehara T, Takenouchi T, Kosaki R et al. Redefining the phenotypic spectrum of de novo heterozygous CDK13 variants: Three patients without cardiac defects. Eur J Med Genet. mai 2018;61(5):243-7.
5. Hamilton MJ, Caswell RC, Canham N, Cole T, Firth HV, Foulds N, et al. Heterozygous mutations affecting the protein kinase domain of CDK13 cause a syndromic form of developmental delay and intellectual disability. J Med Genet. 2018;55(1):28-38.
6. Yakubov R, Ayman A, Kremer AK, van den Akker M. One-month-old girl presenting with pseudohypoaldosteronism leading to the diagnosis of CDK13-related disorder: a case report and review of the literature. J Med Case Rep. 29 déc 2019;13(1):386.
7. Hildebrand MS, Jackson VE, Scerri TS, Van Reyk O, Coleman M, Braden RO, et al. Severe childhood speech disorder: Gene discovery highlights transcriptional dysregulation. Neurology. 19 mai 2020;94(20):e2148-67.

8. Greifenberg AK, Hönig D, Pilarova K et al. Structural and Functional Analysis of the Cdk13/Cyclin K Complex. Cell Rep. 12 janv 2016;14(2):320-31.
9. Fan Z, Devlin JR, Hogg SJ, Doyle MA et al. CDK13 cooperates with CDK12 to control global RNA polymerase II processivity. Science Advances. 1 mai 2020;6(18):eaaz5041.
10. Greenleaf AL. Human CDK12 and CDK13, multi-tasking CTD kinases for the new millenium. Transcription. 2019;10(2):91-110.
11. Aref-Eshghi E, Kerkhof J, Pedro VP, et al. Evaluation of DNA Methylation Episignatures for Diagnosis and Phenotype Correlations in 42 Mendelian Neurodevelopmental Disorders. *Am J Hum Genet*. 2020;106(3):356-370. doi:10.1016/j.ajhg.2020.01.019.
12. Aref-Eshghi E, Bend EG, Colaiacovo S, Caudle M, Chakrabarti R, Napier M, et al. Diagnostic Utility of Genome-wide DNA Methylation Testing in Genetically Unsolved Individuals with Suspected Hereditary Conditions. Am J Hum Genet. 4 avr 2019;104(4):685-700.
13. Aref-Eshghi E, Rodenhiser DI, Schenkel LC, Lin H, Skinner C, Ainsworth P, et al. Genomic DNA Methylation Signatures Enable Concurrent Diagnosis and Clinical Genetic Variant Classification in Neurodevelopmental Syndromes. Am J Hum Genet. 4 janv 2018;102(1):156-74.
14. Sadikovic B, Levi MA, Kerkhof J, et al. Clinical epigenomics: genome-wide DNA methylation analysis for the diagnosis of Mendelian disorders. *Genet Med*. Published online February 5, 2021. doi:10.1038/s41436-020-01096-4.
15. Sadikovic B, Levy MA, Aref-Eshghi E. Functional annotation of genomic variation: DNA methylation episignatures in neurodevelopmental Mendelian disorders. *Hum Mol Genet*. 2020;29(R1):R27-R32. doi:10.1093/hmg/ddaa144.

16. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30(10):1363-1369. doi:10.1093/bioinformatics/btu049.

17. Ho DE, Imai K, King G, et al. MatchIt: Nonparametric Preprocessing for Parametric Causal Inference. *Journal of Statistical Software*, 42(8), 1-28.

18. Ritchie ME, Phipson B, Wu D, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res*. 2015;43(7), e47. doi:10.1093/nar/gkv007.

19. Meyer D, Dimitriadou E, Hornik K, et al. e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien. R package ver 1.7-6. https://CRAN.R-project.org/package=e1071.

20. Peters TJ, Buckley MJ, Statham AL, et al. De novo identification of differentially methylated regions in the human genome. *Epigenetics Chromatin*. 2015;8(1):6. doi:10.1186/1756-8935-8-6.

21. Berro R, Pedati C, Kehn-Hall K et al. CDK13, a new potential human immunodeficiency virus type 1 inhibitory factor regulating viral mRNA splicing. *J virol.* 2008; 82(14):7155-66. doi: 10.1128/JVI.02543-07.

22. Blazek D, Kohoutek J, Bartholomeeusen K, Johansen E, Hulinkova P, Luo Z, et al. The Cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. Genes Dev. 15 oct 2011;25(20):2158-72.

23. Fan Y, Yin W, Hu B, Kline AD, Zhang VW, Liang D, et al. De Novo Mutations of CCNK Cause a Syndromic Neurodevelopmental Disorder with Distinctive Facial Dysmorphism. Am J Hum Genet. 6 sept 2018;103(3):448-55

24. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res. 2009;37(1):1-13.*

25. Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, *28*(5), 1–26. https://doi.org/10.18637/jss.v028.i05

Data availability:

The authors confirm that the data supporting the findings of this study are available within the article and its supplementary materials.

Acknowledgments:

Funding for this study was provided, in part, by the London Health Sciences Molecular Diagnostics Development Fund and Genome Canada Genomic Applications Partnership Program Grant (Beyond Genomics: Assessing the Improvement in Diagnosis of Rare Diseases using Clinical Epigenomics in Canada, EpiSign-CAN) awarded to B.S.

We deeply thank the families for their participation in this work.
We thank the ERN Ithaca network (https://ern-ithaca.eu) and the AnDDI-Rares network (http://anddi-rares.org) for the diffusion of the partnership proposal

Author information:

Conceptualization: D.G.,B.S.; Data curation: R.R.,F.R.,N.R-P, M.B-H., N.C.; Formal analysis: R.R., F.R.; Funding acquisition: B.S.; Investigation: D.G., M.W., L.P., M.W.,

F.C., J.K., H.M., P.D.;N.B., O.B., A.D-D, M.D-F,L.F.,L.G.,D.H.,J.L.,B.M.,L.A.M.,I.V.,A.M.,C.Q.,M.R.,L.G; Methodology: D.G.,B.S.; Project administration: D.G.; Supervision: D.G.,B.S.; Writing – original draft: F.R.,R.R.,D.G.,B.S.; Writing – review and editing: all authors

Ethics declaration: University hospital of Montpellier (29/05/2020; IRB-MTP_2020_05_202000466); ClinicalTrial.gov identifier: NCT04382573. Informed consent was obtained from all participants as required by the IRB. Consent was received for the use of photos. All institutions involved in human participant research received local IRB approval.

Figure and table legends

- **Table 1: Clinical findings and their distribution for 62 new and already published individuals with CDK13-related disorder (CDK13-RD).** [a]The number of responders are detailed for every feature, and their frequencies/distributions were calculated according to that number. [b]Sacrum anomalies include sacrococcygial dimples and teratoma. [c]MRI brain findings include anomalies of corpus callosum, ptosis of cerebellar tonsils, thickening of occipital horns and temporal lobe asymmetry. All findings are described in Supplementary table 2.
- **Figure 1: Pictures from individuals with CDK13-RD and composite gestalt for CDK13-RD. (A)** Frontal pictures from 13 individuals with CDK13-RD. **(B)** Side pictures from 11 individuals with CDK13-RD. **(C)** Composite gestalt for CDK13-RD created by using Face2Gene
- **Figure 2: Box plot representing different growth parameters and psychomotor development. (A)** Growth parameters for individuals from this series. **(B)** Growth parameters for all individuals with a CDK13 pathogenic or likely pathogenic variant. **(C)**

Psychomotor development for individuals from this series. Data are median (horizontal line); interquartile range: box edges; whiskers: range. HC, head circumference

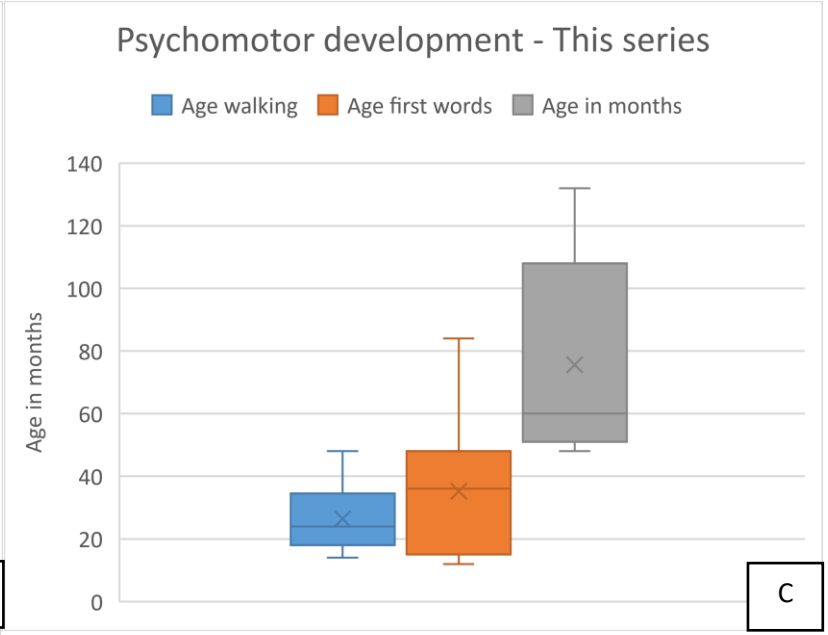
- **Figure 3 CDK13-RD episignature identification. (A)** Volcano plot showing the mean methylation difference between cases and controls, and their corresponding -log(p) values. Red data points correspond to the probes selected for defining the methylation signature. **(B)** Multidimensional scaling (MDS) plot of the training data using the selected probes projected in 2 dimensions. Blue data points represent healthy controls and red data points individuals with CDK13-RD, with strong independent grouping. **(C)** Heatmap and hierarchical clustering of the training data. Rows of the heatmap correspond to the signature probes and columns to the samples. Methylation beta values are color scaled to show intensity values, from 0 (blue) to 1 (red). A clear and distinct profile separates the cases from the controls. This is further supported by the hierarchical clustering dendrogram generated by using Ward's method resulting in the 2 disjoint sets at the top level.
- **Figure 4 Signature-specific analysis shows promising functionality for CHDFIDD diagnosis with high specificity.** Unsupervised analysis using **(A)** hierarchical clustering and **(B)** MDS was performed to classify other variant types and categories (testing set) with respect to the selected samples used for signature discovery (training set). **(C)** A support vector machine (SVM) model was developed and used to predict probability to classify CDK13-RD produced full specificity. The classifier was built using the cohort samples with non-variant of unknown significance (VUS) missense variants and 75% of control and neurodevelopmental disorder/related disorder (NDD/RD) samples as training data (blue), and the rest of the cohort samples and 25% of the control and NDD/RD samples as testing data (grey).

A.

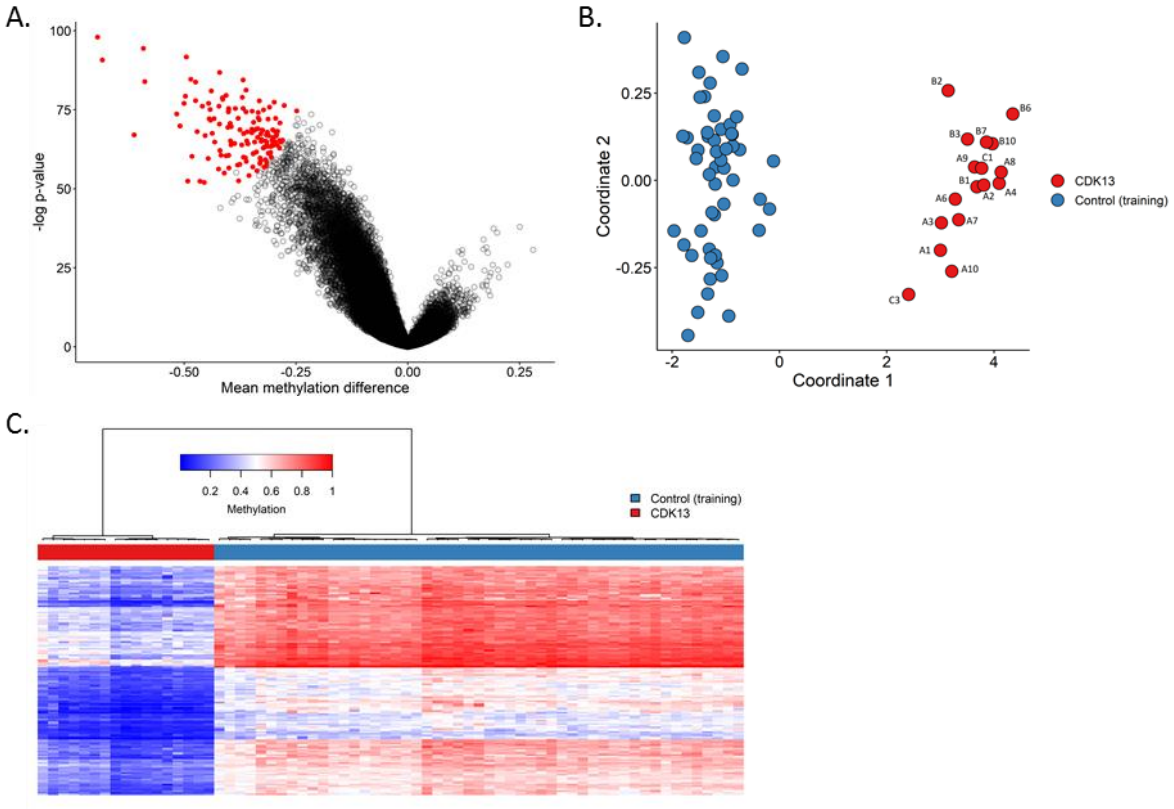
B.

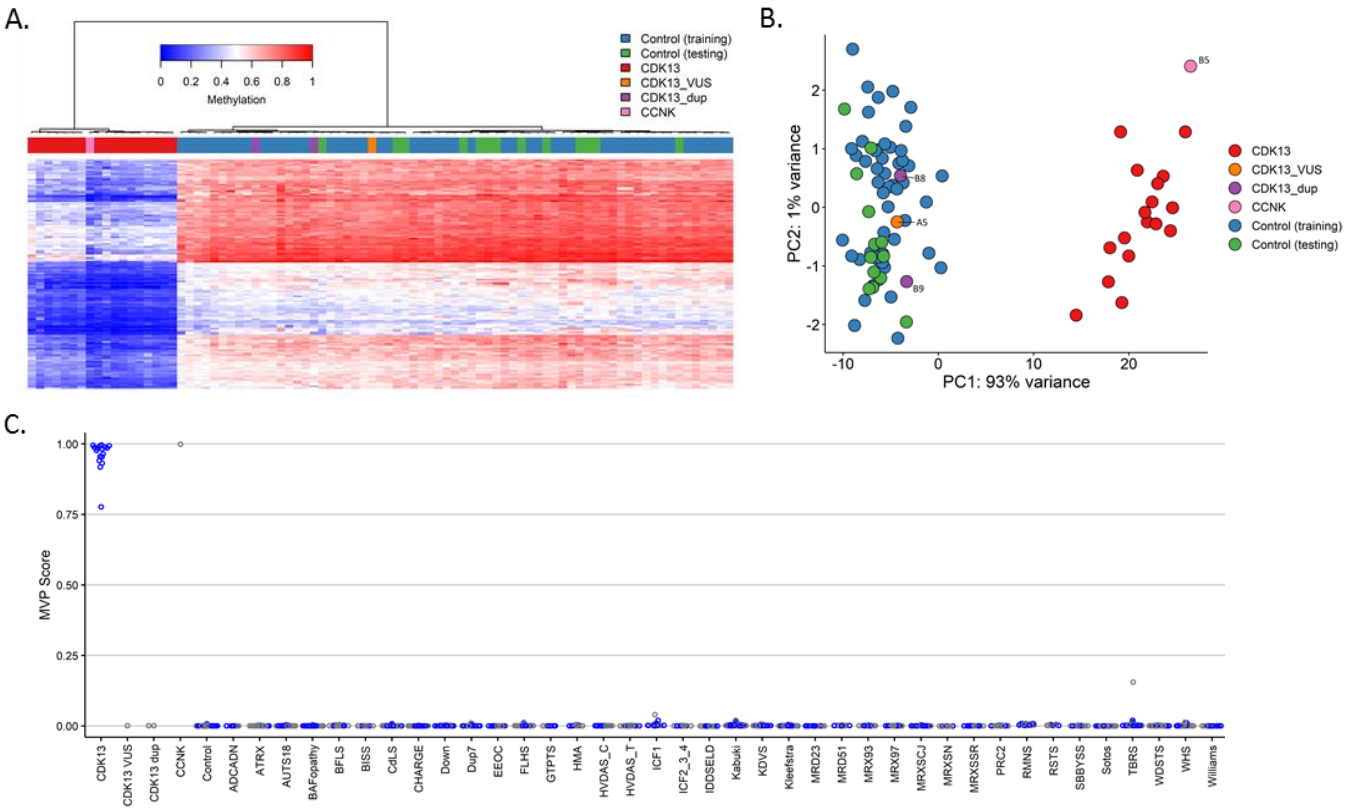
C.


Measurements - This series
Height
Weight
Standard deviations
Standard deviations
4
3
2
1
0
-1
-2
-3
-4
A


Psychomotor development - This series
Age walking
Age first words
Age in months
Age in months
140
120
100
80
60
40
20
0
C


Measurements - All individuals
Height
Weight
Standard deviation
Standard deviation
4
3
2
1
0
-1
-2
-3
-4
B


A.
100
75
50
25
0
-log p-value
-0.50
-0.25
0.00
0.25
Mean methylation difference
B.
0.25
0.00
-0.25
Coordinate 2
-2
0
2
4
Coordinate 1
B2
B6
B3
B7
B10
A9
C1
A8
B1
A6
A2
A4
A3
A7
A1
A10
C3
CDK13
Control (training)
C.
0.2
0.4
0.6
0.8
1
Methylation
Control (training)
CDK13


A.
0 0.2 0.4 0.6 0.8 1
Methylation
Control (training)
Control (testing)
CDK13
CDK13_VUS
CDK13_dup
CCNK
B.
PC2: 1% variance
PC1: 93% variance
B5
B8
A5
B9
CDK13
CDK13_VUS
CDK13_dup
CCNK
Control (training)
Control (testing)
C.
MVP Score
1.00
0.75
0.50
0.25
0.00
CDK13
CDK13 VUS
CDK13 dup
CCNK
Control
ADCADN
ATRX
AUTS18
BAFopathy
BFLS
BISS
CdLS
CHARGE
Down
Dup7
EEOC
FLHS
GTPTS
HMA
HVDAS_C
HVDAS_T
ICF1
ICF2_3_4
IDDSELD
Kabuki
KDVS
Kleefstra
MRD23
MRD51
MRX93
MRX97
MRXSCJ
MRXSN
MRXSSR
PRC2
RMNS
RSTS
SBBYSS
Sotos
TBRS
WDSTS
WHS
Williams

Clinical finding[a]	This series	Literature	Global
Sex (males/females, n[%]) (n= 18 + 44)	8(44,4)/10(55,6)	15(34)/29(66)	23(37,1)/39(62,9)
Age at last examination (years, m[IQR]) (n= 18 + 44)	12(9,25;14)	8,3(4,6;12,8)	9,4(6,1;13)
Pregnancy complications n[%] (n= 17)	5(29,4)		
Failure to thrive n[%] (n=18)	8(44,4)		
Developmental regression n[%] (n=18)	2(11,1)		
Ability to read n[%] (n=17)	6(35,3)		
Global delay/intellectual disability n[%] (n=17 + 43)	16(94,1)	42(97,7)	58(96,7)
Autism spectrum disorder n[%] (n= 18 + 28)	3(16,7)	11(39,3)	14(30,4)
Attention deficit hyperactivity disorder n[%] (n= 18 + 15)	7(38,9)	4(26,7)	11(33,3)
Anxiety disorder n[%] (n=18)	9(50)		
Aggression n[%] (n=18)	7(38,9)		
Axial hypotonia n[%] (n= 18 + 34)	13(72,2)	24(70,6)	37(69,2)
Appendicular hypertonia n[%] (n=17)	3(17,6)		
Hyperreflexia n[%] (n=17)	3(17,6)		
Tremor n[%] (n=17)	2(11,8)		
Dysmetria n[%] (n=16)	3(18,8)		
Abnormal gait pattern n[%] (n=9)	3(33,3)		
Epilepsy n[%] (n=18 + 29)	1(5,6)	7(24,1)	8(17)
Disrupted sleep n[%] (n=18)	8(44,4)		
Dysphagia n[%] (n=17)	2(11,8)		
Sensorineural hearing loss n[%] (n= 17 + 28)	4(23,5)	4(14,3)	8(17,8)
Central nervous system anomaly n[%] (n= 18 + 18)	5(27,8)	8(44,4)	13(36,1)
Spinal chord anomaly	2(11,1)		
Curly hair n[%] (n= 17 + 16)	12(70,6)	6(37,5)	18(54,5)
Deep set eyes n[%] (n= 17 + 16)	9(52,9)	1(6,3)	10(30,3)
Hypertelorism n[%] (n= 17 + 42)	13(76,5)	24(57,1)	37(62,7)
Epicanthic fold n[%] (n= 17 + 24)	10(58,8)	14(58,3)	24(58,5)
Arched eyebrows n[%] (n= 17 + 24)	4(23,5)	10(41,7)	14(34,1)
Wide nasal base n[%] (n= 18 + 42)	14(77,8)	26(61,9)	40(66,7)
Short columella n[%] (n= 17 + 24)	12(70,6)	9(37,5)	21(51,2)
Thin upper lip n[%] (n= 18 + 39)	12(66,7)	15(38,5)	27(47,4)
Upslanting palpebral fissures n[%] (n= 16 + 19)	8(50)	11(57,9)	19(54,3)
Blepharophimosis n[%] (n= 17 + 31)	6(35,3)	10(32,3)	16(33,3)
Anomaly of ear morphology n[%] (n= 17 + 26)	14(82,4)	17(65,4)	31(72,1)
Posteriorly rotated ears n[%] (n= 17 + 34)	10(58,8)	14(41,1)	24(47)
5th finger clinodactyly n[%] (n= 17 + 36)	9(52,9)	16(44,4)	25(47,2)
Hemangioma n[%] (n= 17 + 32)	3(17,6)	8(25)	11(22,4)
Myopia n[%] (n= 18 + 31)	4(22,2)	2(6,5)	6(10,2)
Strabismus n[%] (n= 18 + 43)	10(55,6)	15(34,9)	25(41)
Astigmatism n[%] (n= 18)	6(33,3)		
High arched palate n[%] (n= 17 + 18)	4(23,5)	2(11,1)	6(17,1)
Atrial septal defect n[%] (n= 18 + 40)	7(38,9)	11(27,5)	18(31)
Ventricular septal defect n[%] (n= 18 + 40)	1(5,6)	4(10)	5(8,7)
Pulmonary artery stenosis n[%] (n= 18 + 31)	5(27,8)	3(9,7)	8(13,6)
At least one cardiac malformation n[%] (n= 18 + 41)	12(66,7)	15(36,6)	27(45,8)
GERD n[%] (n= 16 + 14)	4(25)	5(35,7)	9(30)
Constipation n[%] (n= 18 + 16)	10(55,6)	7(43,8)	17(50)
Cryptorchidism n[%] (n= 8)	6(75)		
Contractures n[%] (n= 18)	4(22,2)		
Sacrum anomalies[b] n[%] (n= 18 + 23)	4(22,2)	6(26,1)	10(24,4)
Recurrent infections n[%] (n= 18 + 9)	6(33,3)	5(55,6)	11(40,7)
MRI brain findings[c] n[%] (n= 16 + 18)	8(50)	8(44,4)	16(47,1)

Table 1 : Clinical findings and their distribution across 18 new cases and 44 previously published cases of CDK13-related disorder

[a]The number of responders are detailed for every feature, and their frequencies/distributions were calculated according to that number, for example for autism spectrum disorder, n= 18 + 28 means data was available for 18 individuals from our series and 28 individuals from literature, 3(16.7) means 3 individuals in our series presented with ASD, which represents 16.7% of individuals in our series.

[b]Sacrum anomalies include sacrococcygial dimples and teratoma

[c]MRI brain findings include anomalies of corpus callosum, ptosis of cerebellar tonsils, thickening of occipital horns and temporal lobe asymmetry. All findings are described in Supplementary table 2