

## **A novel syndrome caused by the constitutional gain-of-function variant p.Glu1099Lys in *NSD2***

Bernt Popp <sup>1,\*,#</sup>, Melanie Brugger <sup>2,\*</sup>, Sibylle Poschmann <sup>3</sup>, Tobias Bartolomaeus <sup>1</sup>, Maximilian Radtke <sup>1</sup>, Julia Hentschel <sup>1</sup>, Nataliya Di Donato <sup>4</sup>, Andreas Rump <sup>4</sup>, Janina Gburek-Augustat <sup>5</sup>, Elisabeth Graf <sup>2</sup>, Matias Wagner <sup>2,6,7</sup>, Johannes Lemke <sup>1</sup>, Thomas Meitinger <sup>2</sup>, Rami Abou Jamra <sup>1</sup>, Vincent Strehlow <sup>1,\*</sup>, Theresa Brunet <sup>2,6,\*,#</sup>

<sup>1</sup> Institute of Human Genetics, University of Leipzig Medical Center, Leipzig, Germany

<sup>2</sup> Institute of Human Genetics, Klinikum rechts der Isar, School of Medicine, Technical University of Munich, Munich, Germany

<sup>3</sup> Division of Neuropediatrics, Clinic for Children and Adolescents Dritter Orden, Munich, Germany

<sup>4</sup> Institute for Clinical Genetics, University Hospital, TU Dresden, Dresden, Germany

<sup>5</sup> Division of Neuropaediatrics, Hospital for Children and Adolescents, University of Leipzig Medical Center, Leipzig, Germany

<sup>6</sup> Institute of Neurogenomics, Helmholtz Zentrum München, Neuherberg, Germany.

<sup>7</sup> Division of Pediatric Neurology, Developmental Medicine and Social Pediatrics, Department of Pediatrics, Dr. von Hauner Children's Hospital, Munich University Hospital (Ludwig Maximilians University), Munich, Germany

\* These authors contributed equally to this work

# Corresponding author

### **CORRESPONDING AUTHOR:**

Dr. med. Bernt Popp  
Institute of Human Genetics  
University of Leipzig Medical Center  
Leipzig  
Germany  
Tel: +49 341 97 23840  
Fax: +49 341 97 23819  
Email: [Bernt.Popp@medizin.uni-leipzig.de](mailto:Bernt.Popp@medizin.uni-leipzig.de)

**KEYWORDS:** NSD2, gain-of-function, Glu1099Lys, neurodevelopmental disorder, Rauch-Steindl syndrome, Wolf-Hirschhorn syndrome

## ABSTRACT

**Purpose:** NSD2 dimethylates histone H3 at lysine 36 (H3K36me2) and is located in the Wolf-Hirschhorn syndrome (WHS) region. Recent descriptions delineated loss-of-function (LoF) variants in *NSD2* with a distinct disorder. The oncogenic missense variant p.Glu1099Lys occurs somatically in leukemia and has a gain-of-function (GoF) effect.

**Methods:** We describe two unrelated individuals carrying c.3295G>A, p.Glu1099Lys as heterozygous *de novo* germline variants identified by exome sequencing of blood DNA and subsequently confirmed in two ectodermal tissues. We use omics data from the Cancer Cell Line Encyclopedia to analyze the GoF effect.

**Results:** Clinically these individuals are characterized by intellectual disability, coarse/square facial gestalt, abnormalities of the hands and organomegaly. We confirmed increased K36me2 methylation in lines with either *NSD2* GoF variants or duplications. Cells with GoF variants showed increased DNA promoter methylation and dysregulated RNA expression, influencing cellular modules involved in white blood cell activation, cell growth and organ development.

**Conclusion:** NSD2 GoF caused by the missense variant p.Glu1099Lys leads to a novel syndromic phenotype distinct from both the previously described LoF phenotypes. Other variants causing NSD2 hyperactivation or overexpression may cause a similar phenotype. This syndrome should be distinguished from the recently named Rauch-Steindl syndrome caused by *NSD2* haploinsufficiency.

## INTRODUCTION

Variants affecting genes encoding proteins involved in epigenetic processes are a prominent cause for neurodevelopmental disorders (NDDs) and have recently been summarized as “Mendelian Disorders of the Epigenetic Machinery”.<sup>1</sup> Clinically, these entities often comprise intellectual disability (ID) and growth abnormalities.

Post-translational modifications (acetylation or methylation) of lysine (K) residues in the N-termini of histones are among the most common mechanisms of epigenetic regulation in the cell. These modifications can determine the accessibility of DNA and therefore its transcriptional activity. They mainly affect the H3 core histone and, depending on the modified amino acid (AA) residues, can serve as a marker for repressive (like H3K27) or active (like H3K36) chromatin states.

The nuclear receptor binding SET domain protein 2 (*NSD2*) catalyzes writing of H3K36 methylation marks.<sup>2,3</sup> The previous name “Wolf-Hirschhorn syndrome candidate 1” (*WHSC1*) indicates that the gene is located in the Wolf-Hirschhorn syndrome (WHS, OMIM #194190) deletion region on the short arm of chromosome 4. WHS has been cytogenetically characterized prior to the chromosomal microarray (CMA) era and subsequently refined as a contiguous gene deletion syndrome with two candidate genes (*NSD2* and *LETM1*) in a 1.6 Megabase (Mb) critical genomic region on 4p16.3 (OMIM #194190). The core clinical features of WHS include besides NDD a typical facial dysmorphism described to resemble a “Greek warrior helmet”, ID and intrauterine/ postnatal growth retardation.<sup>4,5</sup> With the broad availability of exome sequencing, individuals with NDDs carrying heterozygous *NSD2* variants have been reported subsequently.<sup>6–8</sup> A recent article described truncating and missense variants in *NSD2* in 18 individuals from 16 families with a rather mild developmental phenotype that only partially overlaps with WHS (Rauch-Steindl syndrome, RAUST, OMIM #619695).<sup>9</sup> Functional analyses of the five different missense variants showed decreased ability to generate H3K36me<sub>2</sub> in all but one (c.2606G>A, p.(Cys869Tyr)), indicating that loss of methylation activity is the main pathomechanism.<sup>9</sup>

The other previous name of *NSD2*, “multiple myeloma SET domain containing protein” (MMSET), highlights the gene's involvement in hematologic malignancies, which was initially identified by characterization of the somatic t(4;14) translocation present in 15-25% of multiple myeloma cases.<sup>10</sup> This rearrangement fuses *NSD2* to the immunoglobulin heavy-chain (IGH) and thus causes its overexpression. Sequencing studies later identified a recurring somatic hotspot mutation in *NSD2* (E1099K; c.3295G>A, p.Glu1099Lys) in cases of pediatric acute lymphoblastic leukemia.<sup>11,12</sup> Functional studies showed that the IGH-*NSD2* fusion causes expansion of H3K36me2 methylation marks and differential gene expression with oncogenic potential.<sup>13</sup> Similarly the p.Glu1099Lys oncogenic missense variant within *NSD2* enhanced global H3K36me2 in lymphoid cells and reduced H3K36 trimethylation.<sup>11</sup> Recent investigations using cryo-electron microscopy solved the structure of *NSD2* bound to the nucleosome and indicated that the oncogenic p.Glu1099Lys *NSD2* missense variant destabilizes the autoinhibitory loop of the protein to enhance substrate recognition and increase the nucleosomal H3K36 methyltransferase activity of *NSD2*.<sup>14,15</sup>

It is known<sup>16</sup> that genes involved in clonal hematopoiesis, hematologic malignancy and other types of cancer are also overrepresented in curated NDD databases. On a variant level, there is a significant overlap between somatic driver mutations in (hematologic) malignancies and (*de novo*) germline variants causing NDD. Both *de novo* missense variants identified in large NDD cohorts and somatic driver variants in cancer<sup>17</sup> are enriched in constrained genomic regions. These observations seem plausible, because many of the overlapping genes are central to cellular processes and similar mutational mechanisms affect both somatic and germline cells.

Prominent examples for cancer associated genes with two distinct associated NDD entities based on different molecular alterations caused by the underlying variants include epigenetic regulators like *DNMT3A* (OMIM #618724, “Heyn-Sproul-Jackson syndrome”; OMIM #615879, “Tatton-Brown-Rahman syndrome”<sup>18</sup>) and *SETD2* (OMIM #616831, “Luscan-Lumish syndrome”; a clinically distinguishable disorder associated with GoF missense variants at AA position Arg1740 recently described by Rabin and colleagues<sup>19</sup>). Likely

further similar entities caused by distinct functional effects of disease-causing variants in the same gene are to be discovered.

Here, we describe the identification of the GoF variant p.Glu1099Lys in *NSD2* as a constitutional, germline heterozygous *de novo* variant in two individuals with a clinically recognizable syndromic NDD entity.

## **MATERIALS AND METHODS**

### Ethics approval and consent to participate

The Ethical Committee of the Medical Faculty, Leipzig University, approved genetic testing in a research setting for individual Ind\_1 (224/16-ek and 402/16-ek). For individual Ind\_2, the study was approved by the local ethics committee of the Technical University of Munich (#5360/12S). Written consent of the parents of both individuals to publish genetic, clinical data and images was received and archived by the authors.

### Genetic analyses and comparison of phenotypes and genotypes

Variants in the index cases were identified by trio exome sequencing (ES) as described previously<sup>20,21</sup>. We used an Excel based questionnaire for retrospective phenotyping of the individuals described here and for review of published cases. Details regarding ES, including primer sequences for Sanger sequencing using DNA from three different tissues, and the clinical sheet are provided in Supplementary notes.

### Variant spectrum and 3D structure analysis

The *NSD2* germline variants were harmonized to a common reference with VariantValidator<sup>22</sup> (NM\_001042424.2 transcript, hg19 reference). The distribution of NDD associated *NDS2* variants in the linear protein representation and in the tertiary protein structure (PDB 7CRO<sup>14</sup>) was compared to variants from databases as described previously<sup>23</sup> (details in Supplementary notes).

### Analysis of omics data from cancer cell lines

We analyzed data from the Cancer Cell Line Encyclopedia (CCLE)<sup>12</sup> using “depmap<sup>24</sup> the ExperimentHub R-package regarding mutation and copy number calls, chromatin modifications, differential methylation and differential expression. Humanbase<sup>25</sup> (<https://hb.flatironinstitute.org/>) was used to predict functionally enriched gene modules. Compare File S2 and Supplementary notes for details.

## RESULTS

### Genetic analyses and review of *NSD2* variants

Trio exome sequencing detected a heterozygous *de novo* missense variant in *NSD2* (NM\_001042424.3: c.3295G>A, p.Glu1099Lys; chr4[hg19]:g.1962801G>A) in both individuals described here. Constitutional status of the c.3295G>A, p.Glu1099Lys *NSD2* variant was confirmed through Sanger sequencing of DNA from buccal swabs and fingernails in Ind\_1 or from buccal swabs in Ind\_2 (Fig. S01).

The *NSD2* variant c.3295G>A, p.Glu1099Lys, had not been previously described as disease-causing in the literature when mutated in the germline and was listed once (1/251,470 alleles) in gnomAD in an individual assigned to the age group of 75-80 years. However, when visualizing the variant listed in gnomAD, it was present in only 32/89 reads (variant allele fraction of 36%) indicating somatic mosaicism (compare Fig. S02). As the phenotype of the individuals did not match with LoF descriptions (RAUST or WHS) the variant was classified as uncertain significance (VUS) according to the ACMG (American College of Medical Genetics and Genomics) recommendations<sup>26</sup> applying the criteria (PS2\_moderate, PP3, PM2\_supporting). Based on the significant clinical overlap of the two individuals, we upgraded the ACMG criterion PS2\_moderate to PS2\_strong and added PM1\_strong (multiple functional analyses) and PP2 (missense Z-score=3.9), allowing classification as “pathogenic”.

Previously described LoF variants affecting *NSD2* are dispersed throughout the protein (Fig. 2A). The distribution in samples from the COSMIC database is similar and shows peaks for the recurrent GoF variants (Fig. 2A). Missense variants with a presumed LoF effect are located in functional domains, but show no clustering in the tertiary protein structure while the two missense variants with a proven GoF effect are relatively close to each other and the lysine at AA position 36 of H3 (Fig. 2B).

### Comparing NSD2 allelic disorders

Both individuals described here showed a remarkable phenotypic overlap with complicated pregnancies and postnatal periods, prenatally detectable organomegaly and postnatally evident dysmorphic features characterized by coarse and rather square facial features (Fig. 1B-D, K) and large hands with tapering fingers (please compare the Supplementary notes for detailed clinical case reports of Ind\_1 and Ind\_2). Noticing this phenotypic overlap we wanted to attempt a genotype-phenotype correlation in comparison with published LoF cases.

The literature review identified 26 individuals with (likely) pathogenic and presumed LoF variants from six publications in which case specific phenotype data was reported.<sup>6-9,27,28</sup> They carried 23 different variants, mostly truncating (n=18, 78.3%) and few missense (n=5, 21.7%). As the descriptions were reported in different publications we first normalized and reviewed the clinical descriptions to compare these with the two cases identified in our study carrying the GoF variant c.3295G>A, p.Glu1099Lys. Despite the sample size of two for the GoF group and the LoF group being inhomogeneous, we identified several facial dysmorphologies significantly varying. These include the coarse facial features with a rather square face in the GoF cases compared to the rather triangular face of individuals with LoF variants. The LoF cases had a broad forehead with high anterior hairline, while the low anterior hairline reported on both GoF cases indicated a smaller forehead area. Facial features exclusively reported in both the GoF cases included anteverted nares and an exaggerated cupid's bow, while a long philtrum was described on both GoF and 1/15 LoF cases. While 6/20 facial features reached nominal significance, due to the small sample size no features remained significant after correction for multiple testing (Table 1).

When comparing other phenotype categories, a short neck was present in both GoF cases but in 0/12 LoF cases with data. Tapered fingers were described in both GoF cases but in only 1/18 LoF cases. Remarkably, four other hand phenotypes were described in both GoF cases, but were not phenotyped in the LoF cases (Table 2; compare Fig. 1E-H and Q-R).



Regarding organic abnormalities only abnormal heart morphology reached nominal significance, but the patent ductus arteriosus in *Ind\_1* is a relatively nonspecific and frequent finding which indicates that the associated phenotype needs to be refined in further descriptions. Abnormalities of the genitourinary system were described in both GoF cases but also in 7/23 (30.4%) of LoF cases. It should be noted that abnormal liver morphology was present in both GoF cases but data on this phenotype was only reported for one LoF; assuming that the other LoF cases were healthy in this regard, this feature would also be nominally significant. Intellectual disability or developmental delay was reported as moderate in both GoF cases, while most LoF cases had mild intellectual disability (73.7%; 14/19). Behavioral anomalies (47.4%; 9/19) and autistic behavior (30.4%; 7/23) were reported in the LoF cases but absent in the two GoF cases. Again, 4/23 phenotypic features analyzed reached nominal significance but did not withstand correction for multiple testing (Table 2).

#### Altered H3K36 modification and DNA methylation changes in *NSD2* GoF cell lines

Our CCLE<sup>12</sup> analysis of global chromatin modification showed that cell lines with GoF variants in *NSD2* had a significantly higher mono- (me1) and dimethylation (me2) level at H3 lysine 36 (K36) with mirroring significant reduction of unmethylated (me0) K36 (Fig. 3A). Cell lines with a genomic duplication affecting the *NSD2* gene region, similarly showed increased ratios of K36me1 and K36me2 with reduction of K36me0, with a lesser effect (Fig. 3B). Differential DNA methylation analysis leveraging the CCLE omics dataset identified a large fraction of promoter regions being differentially methylated. Most loci had a small depmap effect size between the GoF and *NSD2* variant neutral groups and most loci were hypermethylated (n=5,165) while only few showed DNA hypomethylation (n=117). When comparing large depmap effect sizes (defined as > 0.5 or < -0.5) only 12 loci were hypermethylated compared to 28 hypomethylated (Fig. 3C). Of the promoter loci showing a high hypermethylation effect, the *TTC12* ("Ciliary dyskinesia, primary, 45", OMIM #618801) and the *NR2F2* ("46,XX sex reversal 5", OMIM #618901; "Congenital heart defects, multiple

types, 4", OMIM #615779) are associated with human disease in the OMIM database. The *NR2F2*-disorders are associated with congenital heart defects and abnormalities of genitalia. The genes *KPNA7* and *CDC42BPB* have the two most highly hypermethylated promoters when considering NDD-associations curated in the SysID<sup>29</sup> database (compare File S2 sheet "GoFvsNeutral\_Methylation").

#### Dysregulated pathways in NSD2 GoF cell lines

Differential RNA expression analysis using the sets of NSD2 GoF and neutral cell lines resulted in a large set of dysregulated genes (n=2,284 with q-value < 0.05; Fig. 3D). Considering only genes with a very large depmap effect size (n=887; defined as > 1.2 or < -1.2), 311 showed overexpression and 576 showed a reduced expression. The OMIM disease genes *TTC12* and *NR2F2* identified to have a hypermethylated promoter region were also in the downregulated gene set with very high depmap effect size. Similarly, *CDC42BPB*, identified in the SysID overlap analysis and recently associated with NDD<sup>30</sup>, was in the group of genes with very high depmap effect size. Less than expected (n=54; p-value ~ 0.035; two sided Binomial test) NDD genes from SysID (status November 18th 2021 with 1534 primary genes; 1523 present in the depmap expression dataset) overlapped with the very high effect expression geneset (compare File S2 sheet "GoFvsNeutral\_Expression").

Finally, we sought to functionally cluster the 887 genes with a very high effect on expression to explore dysregulated cellular mechanisms. The humanbase analysis identified eight functionally enriched modules with four of them (M1-4) containing most of the genes (Fig. 3E). The M1 module comprised 179 genes and was significantly enriched for blood cell activation terms (e.g. top two terms: "lymphocyte activation", GO:0046649; "leukocyte activation", GO:0045321). The M2 module contained 66 genes enriched for tyrosine kinase pathway (e.g. term number two "regulation of peptidyl-tyrosine phosphorylation", GO:0050730) and kidney development (e.g. term number four "positive regulation of metanephros development", GO:0072216). Module M3 contained 186 genes with an

enrichment cell adhesion (top term “cell-substrate adhesion”, GO:0031589) and growth terms (e.g. term four “hippo signaling”, GO:0035329). Module M4 contained 194 genes enriched for cellular growth (top term “integrin-mediated signaling pathway”, GO:0007229) and embryonic development (e.g. term four “gastrulation”, GO:0007369). Overall, the dysregulated modules are in line with the known function of NSD2 GoF in leukemia and the organ overgrowth phenotype in the two individuals described here.

## DISCUSSION

A genetic disorder can be described most unambiguously by the affected gene, the proposed inheritance pattern, the phenotypic description of associated disease and finally the precise effect of pathogenic variants on protein function. This differentiation is important to guide standardized genetic diagnostics to enable well-characterized rare disease cohorts. *NSD2* was previously recognized as a candidate gene in the WHS region and truncating/missense variants with a loss-of-function (LoF) effect in *NSD2* had been associated with a milder NDD entity with distinct facial appearance (Rauch-Steindl syndrome; RAUST). Thus, *NSD2* haploinsufficiency explains only part of the WHS phenotype.

Next to these entities, the *NSD2* missense variant p.Glu1099Lys has been extensively studied, because of its somatic recurrence in leukemia caused by a gain-of-function (GoF) effect. Here, we report two individuals with a syndromic NDD presentation harboring the germline GoF variant c.3295G>A, p.Glu1099Lys in *NSD2*. Comparing the facial gestalt of individuals with WHS is complicated as further genes in the deletion are expected to contribute to some of the most distinguishing features like the wide nose bridge devolving into the forehead ("Greek warrior helmet"), but there are some overlaps with individuals carrying *NSD2* LoF variants such as high anterior hairline, broad forehead, arched eyebrows, large palpebral fissures, short philtrum. Notably, when comparing the described individuals with *NSD2* LoF individuals, some facial features seem inverted. While GoF carriers have a square facial appearance with low anterior hairline and small forehead, LoF carriers have a triangular face with high anterior hairline and large appearing forehead. Similarly the nasal bridge is wide and depressed on the two individuals with the c.3295G>A, p.Glu1099Lys variant while a thin and elevated nasal bridge was described as specific in the LoF cohort. With the organomegaly and the abnormalities of the skeleton and hands/ feet, the appearance of the *NSD2* GoF individuals seems rather syndromic compared to the relatively unspecific presentation in the LOF individuals.

These phenotypic dissimilarities are expected, given the opposite cellular effects of GoF and LoF variants. Utilizing publically available omics data from the Cancer Cell Line

Encyclopedia (CCLE)<sup>12</sup> we confirm<sup>15,31</sup> dysregulated methylation of histone H3 at the lysine residue 36, especially increased dimethylation, (H3K36me2) as primary molecular effect of the c.3295G>A, p.Glu1099Lys variant. Looking at differential promoter methylation, we identified extensive DNA hypermethylation in GoF cell lines (Fig. 3). The DNA-(cytosine-5)-methyltransferase enzymes (DNMTs) recognize H3K36me2 with their PWWP (AAs 'Pro-Trp-Trp-Pro') domain and bind to these regions of open chromatin.<sup>32</sup> This effect of NSD2 on DNA methylation can be used in future studies to generate specific epi-signatures from peripheral blood DNA, which could aid to characterize variants.

As DNA methylation is involved in gene expression, we next investigated RNA data from the CCLE, which identified >800 genes differentially expressed with very large effect size between GoF and LoF cell lines. Interestingly, less than expected genes currently associated with NDDs in the SysID database overlapped with these dysregulated geneset. This observation however is in line with the relatively mild developmental phenotype observed in Ind\_1 (brain hemorrhage and hypoxia complicated by the course of Ind\_2), considering the genome wide effect of the *NSD2* GoF variant c.3295G>A, p.Glu1099Lys. Using the humanbase tool<sup>33</sup> to cluster the dysregulated geneset, we identified eight main functional modules with the largest and most significantly enriched four being involved in white blood cell activation, cell growth and organ development (Fig. 3E). These enriched functions overlap well with known cellular functions of NSD2 and with the phenotype observed in carriers of the c.3295G>A, p.Glu1099Lys GoF variant. Other “Mendelian Disorders of the Epigenetic Machinery” entities like Sotos syndrome (OMIM #606681; *NSD1* gene), Tatton-Brown-Rahman syndrome (OMIM #615879; *DNMT3A* gene), and Luscan-Lumish syndrome (OMIM #616831; *SETD2* gene) are also associated with overgrowth symptoms like macrocephaly and tall stature.

While the c.3295G>A, p.Glu1099Lys variant in Ind\_1 was identified by clinical ES about six years prior to this report, its highly suspected pathogenicity remained unclear until a second case bearing the same variant was identified. This anecdotal diagnostic odyssey for Ind\_1 together with no other germline descriptions of this variant, despite the relatively large *NSD2*

cohort<sup>9</sup> collected through matchmaking, argues for a very rare disorder. The description of this rare entity was enabled through the practice of openly sharing all diagnostic classifications through ClinVar. Another possible explanation of this association being described after the LoF variants, is that the c.3295G>A, p.Glu1099Lys *NSD2* variant is a known somatic variant in leukemia, which is present in one elderly individual from gnomAD, and could thus be excluded in some very strict filtering pipelines<sup>34</sup>.

Next to the p.Glu1099Lys missense substitution, other *NSD2* missense variants causing similar GoF effects have either been described in blood malignancies (T1150A; c.3448A>G, p.Thr1150Ala) or investigated in vitro<sup>14,15</sup>. Also structural variants like the recurrent somatic t(4;14) translocation, can result in overexpression and increased activity of *NSD2*. This is intriguingly confirmed by our observation that genomic duplications in cell lines cause a similar hypermethylation profile to GoF variants (Fig. 3B), which indicates that *NSD2* could be triplosensitive and contribute to the not well defined 4p16.3 microduplications.

Due to the rarity of this novel *NSD2* GoF entity it is hard to predict the clinical outcome and give recommendations regarding treatment and surveillance. Both individuals with the GoF variant required intensive care postnatally. The course of Ind\_2 was complicated by brain hypoxia, which hampers assessing his development. Ind\_1 on the other hand shows a mild to moderate intellectual disability, which seems surprising given the strong histone methylation changes induced by the GoF variant. Similarly, while both individuals had organ abnormalities and organomegaly (kidneys, liver), laboratory findings were within normal ranges indicating no signs of beginning organ failure. Due to the strong association of the c.3295G>A, p.Glu1099Lys variant with leukemia, the most obvious concern in the two cases remains a possibly increased blood cancer risk, despite unremarkable repeated blood counts. An increased cancer risk has been described in other “Mendelian Disorders of the Epigenetic Machinery” caused by certain variants in *NSD1*, *DNMT3A*, *SETD2* and others.<sup>35–37</sup> A murine conditional knock-in model of the p.Glu1099Lys variant in the b-cell lineage did not lead to spontaneous development of leukemia in immunocompetent mice during the expected life span.<sup>31</sup> Enrichment in relapses of acute lymphoblastic leukemia and functional

studies showing induction of glucocorticoid resistance indicate that the p.Glu1099Lys NSD2 GoF variant might cause clonal advantage and not act as a primary driver event.<sup>38</sup> Based on our current knowledge, we would still recommend at least yearly blood counts, coupled with the regular developmental assessment and abdominal organ ultrasound. Additionally, efforts are made to develop small molecule inhibitors targeting NSD2 as a novel drug in the context of hematologic malignancies<sup>39,40</sup> posing a ground to investigate therapeutic strategies in GoF germline carriers.

Here, we describe a syndromic NDD entity associated with increased enzymatic activity of the *NSD2* gene product causing disturbed histone methylation and dysregulation of gene networks correlating with the observed phenotypes of organomegaly. This relatively specific phenotype, coupled with recognizable coarse facial features and abnormalities of the hands, may allow the clinically supported diagnosis of this “Mendelian Disorders of the Epigenetic Machinery” in individuals with other *NSD2* variants identified by ES. Due to the clinical distinguishability from the previously described syndromes associated with genomic loss of the *NSD2* gene region (WHS) or LoF variants affecting *NSD2* and an antagonistic molecular consequence, it is necessary to differentiate this *NSD2*-related disorder and choose a disease name different from Rauch-Steindl syndrome.

## **DATA AVAILABILITY**

All data generated or analyzed during this study can be found either in the online version of this article at the publisher's website or has been uploaded to Zenodo (File S2,S3,S4: DOI: 10.5281/ZENODO.6206868).

## **ACKNOWLEDGEMENTS**

We thank all involved families for participating in this study.

## **AUTHORS' CONTRIBUTIONS**

B.P., M.B., T.Br., R.A.J. and V.S. conceived the initial study concept. B.P., M.B., M.W., T.Ba., R.A.J., T.Br. and V.S. analyzed genetic data. V.S., M.B. and T.Br. coordinated collection of clinical and genetic data. M.B., T.Br. and B.P. reviewed literature data and standardized the clinical HPO terms. M.B., S.P., M.R., N.D.D., J.G.A., E.G., M.W., T.Ba., J.H., T.M., R.A.J., T.Br. and V.S. provided clinical and genetic data and performed clinical assessments. B.P. performed structural protein analysis. B.P. analyzed all data and created all main figures. B.P., T.Br. and M.B. created the Supplementary files. B.P., M.B., V.S. and T.Br. wrote and edited the manuscript. All authors reviewed, commented and agreed on the final draft manuscript.

## **FUNDING**

B.P. is supported by the Deutsche Forschungsgemeinschaft (DFG) through grant PO2366/2-1.



## **SUPPLEMENTARY**

**File S1 | Supplementary notes with supplementary figures and tables.**

## REFERENCES

1. Fahrner, J. A. & Bjornsson, H. T. Mendelian disorders of the epigenetic machinery: postnatal malleability and therapeutic prospects. *Hum. Mol. Genet.* **28**, R254–R264 (2019).
2. Martinez-Garcia, E. *et al.* The MMSET histone methyl transferase switches global histone methylation and alters gene expression in t(4;14) multiple myeloma cells. *Blood* **117**, 211–220 (2011).
3. Morishita, M., Mevius, D. & di Luccio, E. In vitro histone lysine methylation by NSD1, NSD2/MMSET/WHSC1 and NSD3/WHSC1L. *BMC Struct. Biol.* **14**, 25 (2014).
4. Stec, I. *et al.* WHSC1, a 90 kb SET domain-containing gene, expressed in early development and homologous to a Drosophila dysmorphism gene maps in the Wolf-Hirschhorn syndrome critical region and is fused to IgH in t(4;14) multiple myeloma. *Hum. Mol. Genet.* **7**, 1071–1082 (1998).
5. Zollino, M. *et al.* On the nosology and pathogenesis of Wolf-Hirschhorn syndrome: Genotype-phenotype correlation analysis of 80 patients and literature review. *Am. J. Med. Genet. C Semin. Med. Genet.* **148C**, 257–269 (2008).
6. Barrie, E. S. *et al.* De novo loss-of-function variants in NSD2 (WHSC1) associate with a subset of Wolf-Hirschhorn syndrome. *Cold Spring Harb. Mol. Case Stud.* **5**, a004044 (2019).
7. Lozier, E. R. *et al.* De novo nonsense mutation in WHSC1 (NSD2) in patient with intellectual disability and dysmorphic features. *J. Hum. Genet.* **63**, 919–922 (2018).
8. Boczek, N. J. *et al.* Developmental delay and failure to thrive associated with a loss-of-function variant in WHSC1 (NSD2). *Am. J. Med. Genet. A.* **176**, 2798–2802 (2018).
9. Zaroni, P. *et al.* Loss-of-function and missense variants in NSD2 cause decreased methylation activity and are associated with a distinct developmental phenotype. *Genet. Med.* **23**, 1474–1483 (2021).
10. Chesi, M. *et al.* The t(4;14) Translocation in Myeloma Dysregulates Both FGFR3 and a Novel Gene, MMSET, Resulting in IgH/MMSET Hybrid Transcripts. *Blood* **92**, 3025–

- 3034 (1998).
11. Oyer, J. A. *et al.* Point mutation E1099K in MMSET/NSD2 enhances its methyltransferase activity and leads to altered global chromatin methylation in lymphoid malignancies. *Leukemia* **28**, 198–201 (2014).
  12. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).
  13. Kuo, A. J. *et al.* NSD2 Links Dimethylation of Histone H3 at Lysine 36 to Oncogenic Programming. *Mol. Cell* **44**, 609–620 (2011).
  14. Li, W. *et al.* Molecular basis of nucleosomal H3K36 methylation by NSD methyltransferases. *Nature* **590**, 498–503 (2021).
  15. Sato, K. *et al.* Structural basis of the regulation of the normal and oncogenic methylation of nucleosomal histone H3 Lys36 by NSD2. *Nat. Commun.* **12**, 6605 (2021).
  16. Qi, H., Dong, C., Chung, W. K., Wang, K. & Shen, Y. Deep Genetic Connection Between Cancer and Developmental Disorders: HUMAN MUTATION. *Hum. Mutat.* **37**, 1042–1050 (2016).
  17. Vitsios, D. *et al.* Cancer-driving mutations are enriched in genic regions intolerant to germline variation. <http://biorxiv.org/lookup/doi/10.1101/2022.01.07.475416> (2022) doi:10.1101/2022.01.07.475416.
  18. Shen, W. *et al.* The spectrum of DNMT3A variants in Tatton-Brown-Rahman syndrome overlaps with that in hematologic malignancies. *Am. J. Med. Genet. A.* **173**, 3022–3028 (2017).
  19. Rabin, R. *et al.* Genotype–phenotype correlation at codon 1740 of SETD2. *Am. J. Med. Genet. A.* **182**, 2037–2048 (2020).
  20. Zacher, P. *et al.* The genetic landscape of intellectual disability and epilepsy in adults and the elderly: a systematic genetic work-up of 150 individuals. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **23**, 1492–1497 (2021).
  21. Brunet, T. *et al.* De novo variants in neurodevelopmental disorders-experiences from a tertiary care center. *Clin. Genet.* **100**, 14–28 (2021).

22. Freeman, P. J., Hart, R. K., Gretton, L. J., Brookes, A. J. & Dalgleish, R.  
VariantValidator: Accurate validation, mapping, and formatting of sequence variation descriptions. *Hum. Mutat.* **39**, 61–68 (2018).
23. Popp, B. *et al.* Targeted sequencing of FH-deficient uterine leiomyomas reveals biallelic inactivating somatic fumarase variants and allows characterization of missense variants. *Mod. Pathol.* **33**, 2341–2353 (2020).
24. Laurent Gatto, Theo Killian. *depmap*. (Bioconductor).  
doi:10.18129/B9.BIOC.DEPMAP.
25. Greene, C. S. *et al.* Understanding multicellular function and disease with human tissue-specific networks. *Nat. Genet.* **47**, 569–576 (2015).
26. Richards, S. *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **17**, 405–424 (2015).
27. Derar, N. *et al.* De novo truncating variants in WHSC1 recapitulate the Wolf-Hirschhorn (4p16.3 microdeletion) syndrome phenotype. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **21**, 185–188 (2019).
28. Jiang, Y. *et al.* De novo truncating variant in NSD2 gene leading to atypical Wolf-Hirschhorn syndrome phenotype. *BMC Med. Genet.* **20**, 134 (2019).
29. Kochinke, K. *et al.* Systematic Phenomics Analysis Deconvolutes Genes Mutated in Intellectual Disability into Biologically Coherent Modules. *Am. J. Hum. Genet.* **98**, 149–164 (2016).
30. Chilton, I. *et al.* De novo heterozygous missense and loss of function variants in *CDC42BPB* are associated with a neurodevelopmental phenotype. *Am. J. Med. Genet. A.* **182**, 962–973 (2020).
31. Li, J. *et al.* NSD2-E1099K Mutation Leads to Glucocorticoid-Resistant B Cell Lymphocytic Leukemia in Mice. *Blood* **136**, 3–4 (2020).
32. Weinberg, D. N. *et al.* The histone mark H3K36me2 recruits DNMT3A and shapes

- the intergenic DNA methylation landscape. *Nature* **573**, 281–286 (2019).
33. Krishnan, A. *et al.* Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. *Nat. Neurosci.* **19**, 1454–1462 (2016).
  34. Brunet, T. *et al.* Clonal Hematopoiesis as a pitfall in germline variant interpretation in the context of Mendelian disorders. *Hum. Mol. Genet.* ddac034 (2022)  
doi:10.1093/hmg/ddac034.
  35. Pappas, J. & Rabin, R. SETD2 Neurodevelopmental Disorders. in *GeneReviews®* (eds. Adam, M. P. *et al.*) (University of Washington, Seattle, 1993).
  36. Hollink, I. H. I. M. *et al.* Acute myeloid leukaemia in a case with Tatton-Brown-Rahman syndrome: the peculiar *DNMT3A* R882 mutation. *J. Med. Genet.* **54**, 805–808 (2017).
  37. Kosaki, R., Terashima, H., Kubota, M. & Kosaki, K. Acute myeloid leukemia-associated *DNMT3A* p.Arg882His mutation in a patient with Tatton-Brown-Rahman overgrowth syndrome as a constitutional mutation. *Am. J. Med. Genet. A.* **173**, 250–253 (2017).
  38. Pierro, J. *et al.* The NSD2 p.E1099K Mutation Is Enriched at Relapse and Confers Drug Resistance in a Cell Context-Dependent Manner in Pediatric Acute Lymphoblastic Leukemia. *Mol. Cancer Res. MCR* **18**, 1153–1165 (2020).
  39. Shen, Y. *et al.* Identification of LEM-14 inhibitor of the oncoprotein NSD2. *Biochem. Biophys. Res. Commun.* **508**, 102–108 (2019).
  40. Coussens, N. P. *et al.* High-throughput screening with nucleosome substrate identifies small-molecule inhibitors of the human histone lysine methyltransferase NSD2. *J. Biol. Chem.* **293**, 13750–13765 (2018).

## FIGURES

### Figure 1. Facial dysmorphology and clinical images

(A) and (B) frontal facial images of Ind\_1. (C) and (D) lateral facial images of Ind\_1. facial images of Ind\_1. (E) and (F) dorsal and (G) and (H) palmar sides of the right hand of Ind\_1. (I) and (J) dorsal side of the feet of Ind\_1. Paired images at different ages. (K) Full frontal body view of Ind\_1. (L) Hip X-ray of Ind\_2 showing almost horizontal acetabulum, widened proximal femoral metaphysis and prominent incisura ischiadica indicative of skeletal dysplasia. (M) and (N) abdominal MRI images of Ind\_2 showing hepatomegaly and nephromegaly. (O) abdominal ultrasound of Ind\_2 displaying kidney hypertrophy, diminished corticomedullary differentiation, irregular parenchyma, multiple small renal cysts. and (O) hepatomegaly with rounded caudal margin in Ind\_2. (Q) and (R) Dorsal and palmar sides of the right hand of Ind\_2 showing short and tapered fingers as well as hypoplastic fingernails. (S) dorsal side of the feet of Ind\_2 illustrating hypoplastic toenails.

### Figure 2. Variant distribution on NSD2

(A) Upper panel with schematic of the linear NSD2 protein with domains and variant distribution (missense variants in blue, truncating variants in red). Variants above the protein scheme have been described in germline, while the ones below are somatic from the COSMIC database (note the GoF variant occurring 50 times). The position of the p.Glu1099Lys GoF variant reported here is indicated by the label and arrow. The size of the lollipop circle is proportional to the reported frequency in the literature or database. The segment height is proportional to the CADD-score. Domains based on UniProt O96028. Lower panel with the generalized additive model of the CADD PHRED v1.6 values for all possible NSD2 missense variants as indicator of conserved protein domains indicates highest constraint in the SET domain. (B) Pseudo model of the nucleosome with DNA double helix (light pink), histone octamer (gray) and NSD2 (cyan) bound to H3 (green). The less structured regions of NSD2 (amino acid (AA) positions 1-760 and 1200-1365) are faded.

The position of AA affected by missense variants in NSD2 from our study (Glu1099) are presented as blue and from the literature review as green spheres. The AA of the second described GoF missense (Thr1150) variant is also colored in blue. The H3 lysine 36 (K36) is colored in orange. Right panels show zooms of the K36 region (upper right side) and a view from below the model rotated by 90° (lower right side). Overall, the likely LoF missense variants are further away from K36 and more dispersed throughout the NSD2 protein.

### **Figure 3. NSD2 and cell line omics analyses**

**(A)** Boxplots comparing the z-scores of the global chromatin profiling dataset (CCLE) between cell lines with GoF (red), LoF (green) and no (blue) *NSD2* variant for modification combinations at histone H3 at K27 and K36. Results show a strong hypermethylation effect with increased mono- and dimethylation at K36 and associated reduction of unmethylated (me0) K36 for cell lines with GoF. Cell lines with a LoF variant showed a trend towards an opposite effect on histone modification but did not reach statistical significance (comparison bars not shown), which could be explained by the weaker effect of copy number changes (compare duplications) or other genetic effects in the control cell lines (e.g promoter or non-coding variants) causing similar changes in histone modification. **(B)** Boxplots as in (A) but now comparing cell lines with genomic duplications (“dup”, red), deletions (“del”, green) or copy neutral (blue) state at the *NSD2* gene locus. The results for duplications mirror the observations for GoF in (A) but with a weaker effect. (two sided Wilcoxon signed-rank test; NS, not significant; “\*\*\*\*”, 0.001; “\*\*\*”, 0.01, “\*\*”, 0.05). **(C)** Scatterplot of the differential DNA methylation analysis between cell lines with GoF and no *NSD2* variant using the reduced representation bisulfite sequencing (RRBS) dataset of methylation one kilobase upstream of the transcription start site (TSS1kb). Only genetic loci with an effect size > 0.1 or < -0.1 are displayed and effect sizes are color coded. Overall, most differentially methylated DNA loci are hypermethylated (5,165 vs. 117 with hypomethylation). When comparing large effect sizes (> 0.5 or < -0.5) only few loci are hypermethylated (12 vs. 28 hypomethylated). **(D)** Scatterplot of the differential RNA expression using the same cell line lists as in (C) shows

many genes dysregulated with a very large effect size (887) with 311 showing overexpression and 576 with reduced expression. Dysregulated gene loci with disease association are highlighted by a text box (note that *NR2F2* has different start sites and thus multiple promoter regions marked). Analyses in C and D have been performed using the depmap portal “Data Explorer” tool. **(E)** Network map of functionally enriched modules in the 887 genes with very high effect sizes from (D). The four largest modules (M1: 179 genes, M2: 66 genes, M3: 186 genes, M4: 194 genes) are highlighted with their top associated terms in the accompanying box. The modules are enriched for white blood cell activation (M1), cell adhesion and growth (M2, M3, M4) and organ development (M2, M3).



## TABLES

**Table 1.** Phenotypic abnormalities of the face

| Phenotype                      | HPO               | GoF cases         | LoF cases         | p-value (FDR-corr.)  |
|--------------------------------|-------------------|-------------------|-------------------|----------------------|
| Abnormal facial shape          | HP:0001999        | 100% (2/2)        | 100% (16/16)      |                      |
| <b>Coarse facial features</b>  | <b>HP:0000280</b> | <b>100% (2/2)</b> | <b>12% (2/16)</b> | <b>0.039 (0.189)</b> |
| Triangular face                | HP:0000325        | 0% (0/2)          | 81% (13/16)       |                      |
| <b>Square face</b>             | <b>HP:0000321</b> | <b>100% (2/2)</b> | <b>0% (0/16)</b>  | <b>0.007 (0.090)</b> |
| Thick hair                     | HP:0100874        | 100% (2/2)        | 0% (0/1)          |                      |
| High anterior hairline         | HP:0009890        | 0% (0/2)          | 80% (12/15)       |                      |
| <b>Low anterior hairline</b>   | <b>HP:0000294</b> | <b>100% (2/2)</b> | <b>13% (2/15)</b> | <b>0.044 (0.190)</b> |
| Broad forehead                 | HP:0000337        | 0% (0/2)          | 81% (13/16)       |                      |
| Laterally sparse eyebrows      | HP:0005338        | 50% (1/2)         | 53% (8/15)        |                      |
| Large palpebral fissures       | HP:0001090        | 50% (1/2)         | 67% (10/15)       |                      |
| Hypertelorism                  | HP:0000316        | 50% (1/2)         | 37% (7/19)        |                      |
| Periorbital hyperpigmentation  | HP:0001106        | 0% (0/2)          | 86% (12/14)       |                      |
| Infra-orbital crease           | HP:0100876        | 100% (2/2)        | 33% (4/12)        |                      |
| <b>Anteverted nares</b>        | <b>HP:0000463</b> | <b>100% (2/2)</b> | <b>0% (0/15)</b>  | <b>0.007 (0.090)</b> |
| Wide nasal bridge              | HP:0000431        | 100% (2/2)        | 39% (7/18)        |                      |
| <b>Long philtrum</b>           | <b>HP:0000343</b> | <b>100% (2/2)</b> | <b>7% (1/15)</b>  | <b>0.022 (0.136)</b> |
| Short philtrum                 | HP:0000322        | 0% (0/2)          | 40% (6/15)        |                      |
| Thin upper lip                 | HP:0000219        | 100% (2/2)        | 35% (6/17)        |                      |
| <b>Exaggerated cupid's bow</b> | <b>HP:0002263</b> | <b>100% (2/2)</b> | <b>0% (0/14)</b>  | <b>0.008 (0.090)</b> |
| Posteriorly rotated ears       | HP:0000358        | 50% (1/2)         | 80% (4/5)         |                      |

**Table 2.** Comparison of other phenotypes between of GoF and LoF cases

| Group                               | Phenotype   | HPO                           | GoF cases         | LoF cases         | p-value (FDR-corr.)  |
|-------------------------------------|---|-------------------------------|-------------------|-------------------|----------------------|
| Phenotypical abnormalities of body  | Abnormality of the foot                             | HP:0001760                    | 100% (2/2)        | 25% (5/20)        |                      |
|                                     | Abnormality of the hand                             | HP:0001155                    | 100% (2/2)        | 25% (5/20)        |                      |
|                                     | Abnormality of skeletal morphology                  | HP:0011842                    | 100% (2/2)        | 24% (5/21)        |                      |
|                                     | <b>Tapered fingers</b>                              | <b>HP:0001182</b>             | <b>100% (2/2)</b> | <b>6% (1/18)</b>  | <b>0.016 (0.113)</b> |
|                                     | <b>Short neck</b>                                   | <b>HP:0000470</b>             | <b>100% (2/2)</b> | <b>0% (0/12)</b>  | <b>0.011 (0.095)</b> |
|                                     | <b>Brachydactyly</b>                                | <b>HP:0001156</b>             | <b>100% (2/2)</b> | <b>NaN% (0/0)</b> |                      |
|                                     | <b>Hypoplastic fingernails and toenails</b>         | <b>HP:0001804; HP:0001800</b> | <b>100% (2/2)</b> | <b>NaN% (0/0)</b> |                      |
|                                     | <b>Large hands</b>                                  | <b>HP:0001176</b>             | <b>100% (2/2)</b> | <b>NaN% (0/0)</b> |                      |
|                                     | <b>Single transverse palmar crease</b>              | <b>HP:0000954</b>             | <b>100% (2/2)</b> | <b>NaN% (0/0)</b> |                      |
|                                     | Short stature                                       | HP:0004322                    | 50% (1/2)         | 38% (9/24)        |                      |
| Intellectual and social development | Global developmental delay; Intellectual disability | HP:0001263; HP:0001249        | 100% (2/2)        | 87% (20/23)       |                      |
|                                     | <b>moderate</b>                                     | <b>HP:0011343; HP:0002342</b> | <b>100% (2/2)</b> | <b>0% (0/19)</b>  | <b>0.005 (0.090)</b> |
|                                     | mild  | HP:0011342; HP:0001256        | 0% (0/2)          | 74% (14/19)       |                      |

| Group               | Phenotype  | HPO                    | GoF cases         | LoF cases         | p-value (FDR-corr.)  |
|---------------------|--|------------------------|-------------------|-------------------|----------------------|
|                     | Behavioral abnormalities                             | HP:0000708             | 0% (0/2)          | 47% (9/19)        |                      |
|                     | Autistic behavior                                    | HP:0000729             | 0% (0/2)          | 30% (7/23)        |                      |
| Neurological system | Muscular hypotonia                                   | HP:0001252             | 100% (2/2)        | 72% (18/25)       |                      |
| Organs              | Abnormal liver morphology                            | HP:0410042             | 100% (2/2)        | 100% (1/1)        |                      |
|                     | Abnormality of the eye                               | HP:0000478             | 100% (2/2)        | 40% (8/20)        |                      |
|                     | Abnormality of the genitourinary system              | HP:0000119             | 100% (2/2)        | 30% (7/23)        |                      |
|                     | Abnormality of the respiratory system                | HP:0002086             | 100% (2/2)        | 21% (4/19)        |                      |
|                     | <b>Abnormal heart morphology</b>                     | <b>HP:0001627</b>      | <b>100% (2/2)</b> | <b>14% (3/21)</b> | <b>0.040 (0.189)</b> |
|                     | Functional abnormality of the gastrointestinal tract | HP:0011024             | 50% (1/2)         | 44% (11/25)       |                      |
| Miscellaneous       | Feeding difficulties/failure to thrive               | HP:0011968; HP:0001508 | 100% (2/2)        | 100% (17/17)      |                      |

|  |  |  |  |
|--|--|--|--|
| <b>A</b><br><br>removed to<br>comply with<br>medrxiv | <b>B</b><br><br>removed to<br>comply with<br>medrxiv | <b>K</b><br><br>removed to<br>comply with<br>medrxiv | <b>O</b><br><br>removed to<br>comply with<br>medrxiv |
| <b>C</b><br><br>removed to<br>comply with<br>medrxiv | <b>D</b><br><br>removed to<br>comply with<br>medrxiv |  | <b>P</b><br><br>removed to<br>comply with<br>medrxiv |
| <b>E</b><br><br>removed to<br>comply with<br>medrxiv | <b>F</b><br><br>removed to<br>comply with<br>medrxiv | <b>L</b><br><br>removed to<br>comply with<br>medrxiv | <b>Q</b><br><br>removed to<br>comply with<br>medrxiv |
| <b>G</b><br><br>removed to<br>comply with<br>medrxiv | <b>H</b><br><br>removed to<br>comply with<br>medrxiv | <b>M</b><br><br>removed to<br>comply with<br>medrxiv | <b>R</b><br><br>removed to<br>comply with<br>medrxiv |
| <b>I</b><br><br>removed to<br>comply with<br>medrxiv | <b>J</b><br><br>removed to<br>comply with<br>medrxiv | <b>N</b><br><br>removed to<br>comply with<br>medrxiv | <b>S</b><br><br>removed to<br>comply with<br>medrxiv |



