# SUPPLEMENTARY NOTES

**The constitutional gain-of-function variant p.Glu1099Lys in *NSD2* is associated with a novel syndrome**

Bernt Popp 1,2,\*,#, Melanie Brugger 3,\*, Sibylle Poschmann 4, Tobias Bartolomaeus 1, Maximilian Radtke 1, Julia Hentschel 1, Nataliya Di Donato 5, Andreas Rump 5, Janina Gburek-Augustat 6, Elisabeth Graf 3, Matias Wagner 3,7,8, Ina Sorge 9, Johannes Lemke 1, Thomas Meitinger 3, Rami Abou Jamra 1, Vincent Strehlow 1,\*, Theresa Brunet 3,7,\*,#

1 Institute of Human Genetics, University of Leipzig Medical Center, Leipzig, Germany

2 Berlin Institute of Health at Charité - Universitätsmedizin Berlin, Center of Functional Genomics, Hessische Straße 4A, 10115 Berlin, Germany

3 Institute of Human Genetics, Klinikum rechts der Isar, School of Medicine, Technical University of Munich, Munich, Germany

4 Division of Neuropediatrics, Clinic for Children and Adolescents Dritter Orden, Munich, Germany

5 Institute for Clinical Genetics, University Hospital, TU Dresden, Dresden, Germany

6 Division of Neuropaediatrics, Hospital for Children and Adolescents, University of Leipzig Medical Center, Leipzig, Germany

7 Institute of Neurogenomics, Helmholtz Zentrum München, Neuherberg, Germany.

8 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

9 Department of Pediatric Radiology, University of Leipzig Hospitals and Clinics, Leipzig, 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

# SUPPLEMENTARY METHODS

## DNA Extraction from blood, buccal swabs and fingernails

## DNA from EDTA blood was extracted using a chemagic™ 360 instrument (PerkinElmer, Waltham, Massachusetts, USA) according to the manufacturer’s recommendations. DNA from buccal swabs (ORACollect DNA OCR-100, DNA Genotek) was extracted using the prepIT L2P reagent kit (DNA Genotek).

## Sanger Sequencing

## Polymerase chain reaction was performed according to standard procedures. DNA fragments were Sanger-sequenced using a SeqStudio Genetic Analyzer (ThermoFisher Scientific) according to the manufacturer’s recommendations.

## Exome Sequencing (ES)

Genomic DNA was extracted using standard methods from peripheral blood samples of Ind\_1 and his parents. Trio exome sequencing and analysis was performed as described previously.1,2 In brief, enrichment for exome sequencing of Ind\_1 and both his parents was done using the BGI Exome capture 59M kit (BGI, Shenzhen, China) and sequenced with 100bp paired end reads on a BGISEQ-500 system (BGI, Shenzhen, China). The resulting sequencing data was processed using the cloud based “varfeed” pipeline (Limbus Medical Technologies GmbH, Rostock, Germany). The “varvis” webtool (Limbus Medical Technologies GmbH, Rostock, Germany) was used to filter the resulting variant files.

ES in Ind\_2 was performed using genomic DNA which was extracted from leukocytes or whole blood using a chemagic 360 Instrument (PerkinElmer). SureSelect Human All Exon 60Mb V6 Kit (Agilent) was used for exome enrichment. Libraries were sequenced on an Illumina NovaSeq6000 system (Illumina, San Diego, California, USA) and reads were aligned to the UCSC human reference assembly (hg19) with BWA v.0.7.5a. On average, more than 98% of targeted regions were covered at least 20x. Single-nucleotide variants (SNVs) and small insertions and deletions were detected using both SAMtools v.0.1.19 and GATK 4.1. Copy number variations (CNVs) were detected using ExomeDepth and Pindel. Variant prioritization was performed based on autosomal recessive (minor allele frequency (MAF) < 0.1%) and autosomal dominant (MAF < 0.01%) inheritance.

Trio ES 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. No further candidate variants were prioritized.

## Clinical data collection and comparison

The clinical terms in our Excel (Microsoft, Redmond, Washington, United States) based questionnaire were based on the phenotypes observed in Ind\_1 and Ind\_2 and a review of previous descriptions of 26 individuals with RAUST due to (likely) pathogenic and presumed LoF variants from six publications in which case specific phenotype data was reported 3–8 which were standardized using the Human Phenotype Ontology (HPO) 9. Clinical descriptions of patients with single nucleotide variants or indels in *NSD2* associated with RAUST were extracted from the reports or supplements of the primary literature source, individually reviewed and updated based on the published textual clinical descriptions or images. Physical growth measurements were compared using World Health Organization (WHO) child growth standards using the pedz calculator (https://www.pedz.de/).

## Variant spectrum and 3D structure analysis

The *NSD2* germline variants were harmonized to a common reference with VariantValidator 10 (NM\_001042424.2 transcript, hg19 reference).

The distribution of NDD associated *NSD2* variants in the linear protein representation was compared to variants reported in the COSMIC cancer 11 database (file “V95\_38\_MUTANT.csv” downloaded on 2021-01-30) and in the public gnomAD 12 population database. Annotation, plotting and analysis of and protein regions constrained for missense variation was performed as described previously 13,14 with updated versions of all annotations.

For analysis of the missense variant distribution in the tertiary protein structure, we used the published Protein data bank (PDB) format structure 7CRO 15 of the SET and AWS domain of NSD2 together with the AlphaFold 16 homology model of human NSD2 downloaded from the EMBL website (https://alphafold.ebi.ac.uk/) with a pipeline based on the software PyMOL version 2.5.0 (Open-Source) as described before 13.

## Analysis of omics data from cancer cell lines

We queried the Cancer Cell Line Encyclopedia (CCLE) 17,18 data using the “depmap” (Cancer **Dep**endency **Map**; https://depmap.org/portal/) 19 and “ExperimentHub” packages in R version 4.1.0 from within RStudio version 1.4.1717. We used the metadata dataset "EH7294" together with the mutation calls ("EH7293") and copy number calls ("EH7291") to identify cell lines with NSD2 GoF/LoF or copy-number variants (CNVs) affecting the *NSD2* region. Cell lines with the known NSD2 GoF variants E1099K and T1150A were filtered using these annotation terms and classified into the GoF category while all nonsense or frameshifting variants were classified into the LoF category. Lines with a variant in *NSD2* not falling into these categories were classified as “other” and subsequently excluded from downstream analyses. The copy number data set was divided into five classes based on the log2 values ("duplication": log2 >= log2(3/2+1) ~ 1.32; "duplication purity 50%": log2 >= log2(5/4+1) ~ 1.17; "deletion purity 50%": log2 >= log2(3/4+1) ~ 0.81; "deletion": log2 >= log2(1/2+1) ~ 0.59; “neutral”). Cell lines with possible copy number changes at lower purity estimates were excluded from analysis.

To investigate chromatin modifications, we downloaded the “CCLE\_GlobalChromatinProfiling\_20181130” dataset directly from the depmap 19 portal. The Z-score values of the chromatin marks at K27 and K36 were visualized as boxplots and group wise p-values were calculated using the “Wilcox rank sum test” as implemented in R.

To investigate differential methylation and differential expression, we next uploaded the lists of nine cell lines classified as GoF and 1192 lines classified as having no *NSD2* mutation and being copy number neutral (Supplementary File S02) to the depmap portal in the “Data Explorer” tool using the “Custom Analyses” tab. We used the “Two class comparison” option to calculate the difference between two groups (GoF vs. neutral, excluding all cell lines with CNVs or any small *NSD2* variants) regarding both the methylation (“Methylation (1kb upsteam TSS)”) and the expression (“Expression 21Q4 Public”) datasets. Differentially methylated/ expressed gene lists were downloaded from depmap portal. Effect sizes, grouped by strength, were plotted and -log10 scaled p-values were visualized as scatter plots using ggplot2 in R.

We filtered the list of differentially expressed genes to select 887 genes with a very high effect size (> 1.2 or < -1.2). This list was then uploaded to the humanbase 20 web tool (https://hb.flatironinstitute.org/) to predict functionally enriched gene modules with the “global” network setting.

# SUPPLEMENTARY RESULTS

## Case of Individual 1

The boy Ind\_1 (main Fig. 1A) was the first child of healthy, non-consanguineous parents. The family history was uneventful for genetic disorders, intellectual disability, or seizures. Pregnancy was complicated by polyhydramnion and a prominent nuchal fold in first trimester ultrasound (US) screening. A horseshoe kidney was suspected in follow-up US examinations. Subsequent amniocentesis followed by conventional karyotyping and panel sequencing for Noonan syndrome were unremarkable. Due to premature rupture of the amniotic sac, pathological cardiotocography (CTG) and green discolored amniotic fluid, vacuum delivery was performed at 37+5 weeks gestational age. At birth his weight was 2,880 g (-0.85 SD), his length was 48.0 cm (-1.22 SD), his occipital frontal circumference (OFC) 34.0 cm (-0.54 SD). Apgar scores were 8/9/9 and the umbilical cord pH was 7.09. Due to dyspnoea and decreased oxygen saturation he required intensive care with continuous positive airway pressure (CPAP) ventilation. His respiratory situation improved only slowly with continuous need for respiratory support. He required long-term total parenteral nutrition followed by duodenal tube feeding. Because of early cholestasis with conjugated hyperbilirubinemia, liver biopsy was performed, and histologic analysis of the liver tissue showed hepatocellular and canalicular cholestasis but normal intrahepatic bile ducts with no evidence of storage disease, viral inclusions, granulomata, or steatosis. On his 21st day of life he developed sepsis and had to be intubated again for mechanical ventilation. Overall, he required intensive care until his 3rd month of life.

Postnatally, multiple organ anomalies were identified, including persistent ductus arteriosus with aneurysm, splenomegaly with splenic cyst, cryptorchidism, bilateral dysplastic and enlarged kidneys with nephrocalcinosis, duplex kidney on the right side, bilateral vesicoureteral reflux and hypertrophic pyloric stenosis which required pyloromyotomy at age of 2 months. He had preretinal hemorrhage of the left eye and astigmatism/ myopia were corrected using glasses. Hearing was normal. He had no seizures and electroencephalogram (EEG) was normal. Laboratory tests for etiological clarification of the suspected syndromic disease including metabolic testing were unremarkable.

He showed delayed developmental milestones with coordinated crawling at the age of 1.5 years, free sitting at the age of 2 years, first walking steps while holding at the age of 2.5 years, free walking at the age of 3.5 years and first words at the age of about 3 years. Formal IQ testing at the age of 6 years and 6 months using the Snijders-Oomen non-verbale test (SON-R 2½-7) confirmed mild to moderate intellectual disability with an overall IQ of 50 and estimated developmental age of about 2.5 years.

The boy was last reviewed at age 8 years 3 months. His height was 128.0 cm (-0.62 SD), his weight was 33.4 kg (1.01 SD; BMI 20,4), and his head circumference was 57.6 cm (3.47 SD). He presented with mild to moderate intellectual disability, could speak approximately five words and mainly communicated via gestures. Medical interventions regarding the malformations of the internal organs have not been necessary since the first year of life. In particular, the laboratory parameters of renal and hepatic function were within the normal range. He had coarse facial features with square facial shape and bitemporal narrowing, rather sparse and thin lateral eyebrows, low anterior hairline, multiple eyelid creases, periorbital fullness, narrow palpebral fissures, hypertelorism and convergent strabismus, a wide nasal bridge and a bulbous tip of the nose, a long philtrum, a wide mouth, a thin upper lip with exaggerated cupid's bow, retrognathia and a short neck (Fig. 1A-D). His chest was broad with widely spaced nipples, and the posterior thorax was broad with widely spaced scapulae and a flat vertebral column. The umbilicus is prominent after surgery for umbilical hernia. He showed a fixed elbow flexion and limited supination of the forearms. The hands had relatively broad palms, with brachydactyly and small nails and bridged palmar crease on the right side and a single palmar crease on the left side (Fig. 1E-H). Similarly, the feet were short and broad, with bilateral pes valgus and hypoplastic nails (Fig. 1I-J). He had genu valgum and a broad-based insecure gait (Fig. 1K). Examination of the external genitalia showed a micropenis.

A recent cranial MRI of Ind\_1 at age 9 years and 2 months showed hyperostosis of the calvaria and a deformation of both eyeballs (compare Fig. S5). Also, a recent immunological blood analysis showed a reduced B-cell count with unremarkable further differentiation, while the T-cell differentiation showed age appropriate naive T-cells but the CD8+ memory cells were reduced.

## Case of Individual 2

The boy Ind\_2 was the fourth child of non-consanguineous parents, who had three older and one younger healthy children. Their family history was unremarkable for developmental disorders. Prenatal US revealed hepatomegaly, large echo-rich kidneys, right ventricular hypertrophy, unilateral clubfoot, macroglossia, single umbilical cord artery and polyhydramnios. Amniocentesis followed by prenatal karyotyping showed an unremarkable male karyotype (46,XY). High-throughput sequencing based panel testing for macrosomia and testing for Beckwith-Wiedemann Syndrome were negative. He was born preterm (33+4 gestational weeks, birth weight 2,740 g (1.9 SD), length 45.5 cm (0.47 SD), OFC 33.0 cm (1.13 SD)) via cesarean section due to missing fetal movements and pathological CTG. Apgar scores were 4/7/9, umbilical cord pH was 6.89. Acidosis resolved during the first 24 hours after birth. Respiratory insufficiency required intubation on the first day of life. Intracerebral hemorrhage of the right nucleus caudatus, grade 2 intraventricular hemorrhage, small cortical hemorrhages, and signs of hypoxia of the occipital cortex were identified by MRI on day 9. Multiple congenital malformations were confirmed after birth including a congenital heart defect (hypoplastic aortic arch, ventricular septal defect), hepatomegaly, nephromegaly with small renal cysts, unilateral clubfoot (left foot), inguinal hernia (right side) and cryptorchidism. X-ray of the pelvis showed enlarged proximal femur metaphysis and prominent incisura ischiadica indicating skeletal dysplasia (Fig. 1L). A ventricular septal defect leading to increased right ventricular load was corrected via patch at the age of 5 weeks. Surgical repair of the hypoplastic aortic arch was performed. The club foot was treated by casting. Laboratory investigations revealed mild hyperammonemia (up to 150 µmol/l). Liver MRI (Fig. 1M-N) identified a patent ductus venosus, which was treated by endovascular coiling. Repetitive episodes of respiratory distress treated by CPAP and oxygen therapy required hospitalization for the first 4.5 months of his life. Recurrent pyelonephritis with fever due to grade 2 vesicoureteral reflux were diagnosed and antibiotic prophylaxis was given. Nephrocalcinosis was identified by ultrasound but resolved spontaneously. Reduced counts of CD4+ T-cells and B-cells as well as increased numbers of NK-cells were identified; immunoglobulin levels (IgG and IgM) were borderline reduced. Severe recurrent infections were not present. Early on, deficits in head control and lifting the head and upper body while in prone prop position as well as muscular hypotonia were observed. Additionally, feeding difficulties and failure to thrive were prominent features in the first months of life. After the gastric tube was removed at the age of 6 months, he was able to eat soft foods and the BMI gradually normalized.

At the last investigation at age 3 years and 0 months the individual had short stature and microcephaly (measurements: weight 12.0 kg (-1.46 SD; BMI 17,4), height 83.0 cm (-3.43 SD), OFC 46,5 cm (-3.32 SD)). Growth hormone stimulation test was abnormal. Hearing was noted to be impaired (failed BERA at the age of 6 months), most likely due to Eustachian tube dysfunction. Ophthalmologic assessment revealed myopia and astigmatism but was otherwise normal. Additionally, the individual presented with an eczema of unknown etiology of upper arms and legs as well as the face which did not resolve to treatment with steroids. Neurologic assessment revealed global developmental delay and a mild muscular hypotonia, but no signs of a movement disorder. He showed no clinical signs of seizures and EEG was normal. Motor development was markedly delayed with turning at age of 12 months, crawling at the age of 24 months, sitting at the age of 18 months and pulling to stand at the age of 37 months. Cognitive and speech development was also delayed; at 37 months the subject babbled repetitive syllables but was not able to speak any words. Evaluation for dysmorphic facial features identified multiple distinct features, including thick hair, synophrys, long palpebral fissures with infraorbital crease, coarse facial features, depressed nasal bridge, anteverted nares, long philtrum, thin upper lip, microdontia, macroglossia, deep-set, and posteriorly rotated ears. Additionally, coarse, and large hands with short fingers and a single transverse palmar crease as well as hypoplastic nails were present, while the feet were short and broad with hypoplastic nails. He had unilateral maldescensus testis.

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

Our CCLE 17 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. S4A). Cell lines with a genomic duplication affecting the *NSD2* gene region, similarly, showed increased ratios of K36me1 and K36me2 with a reduction of K36me0, with a lesser effect (Fig. S4B).

Differential DNA methylation analysis leveraging the CCLE omics dataset identified a large fraction of promoter regions as being differentially methylated. Most loci had a small effect size (as calculated and reported in the depmap tool) between the GoF and *NSD2* variant neutral groups, and most loci were hypermethylated (n = 5,165) while only a few showed DNA hypomethylation (n = 117). When comparing large effect sizes (defined as > 0.5 or < -0.5) only 12 loci were hypermethylated compared to 28 hypomethylated (Fig. S4C). 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 the genitalia. The genes *KPNA7* and *CDC42BPB* have the two most highly hypermethylated promoters when considering NDD-associations curated in the SysID 21 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. S4D). Considering only genes with a very large effect size (n = 887; defined as > 1.2 or < -1.2), 311 showed overexpression and 576 showed 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 effect size. Similarly, *CDC42BPB*, identified in the SysID overlap analysis and recently associated with NDD 22, was in the group of genes with very high 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 portal based 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. S4E). 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.

# SUPPLEMENTARY FIGURES

## 

## **Figure S1** | Sanger sequencing of the missense variant c.3295G>A, p.Glu1099Lys in *NSD2* DNA from DNA from additional tissues.

The missense variant in *NSD2* found in exome sequencing was confirmed in peripheral blood, buccal swab, and fingernails in individual 1 **(A)** and in peripheral blood and buccal swab in individual 2 **(B)**, thus indicating germline origin.

## 

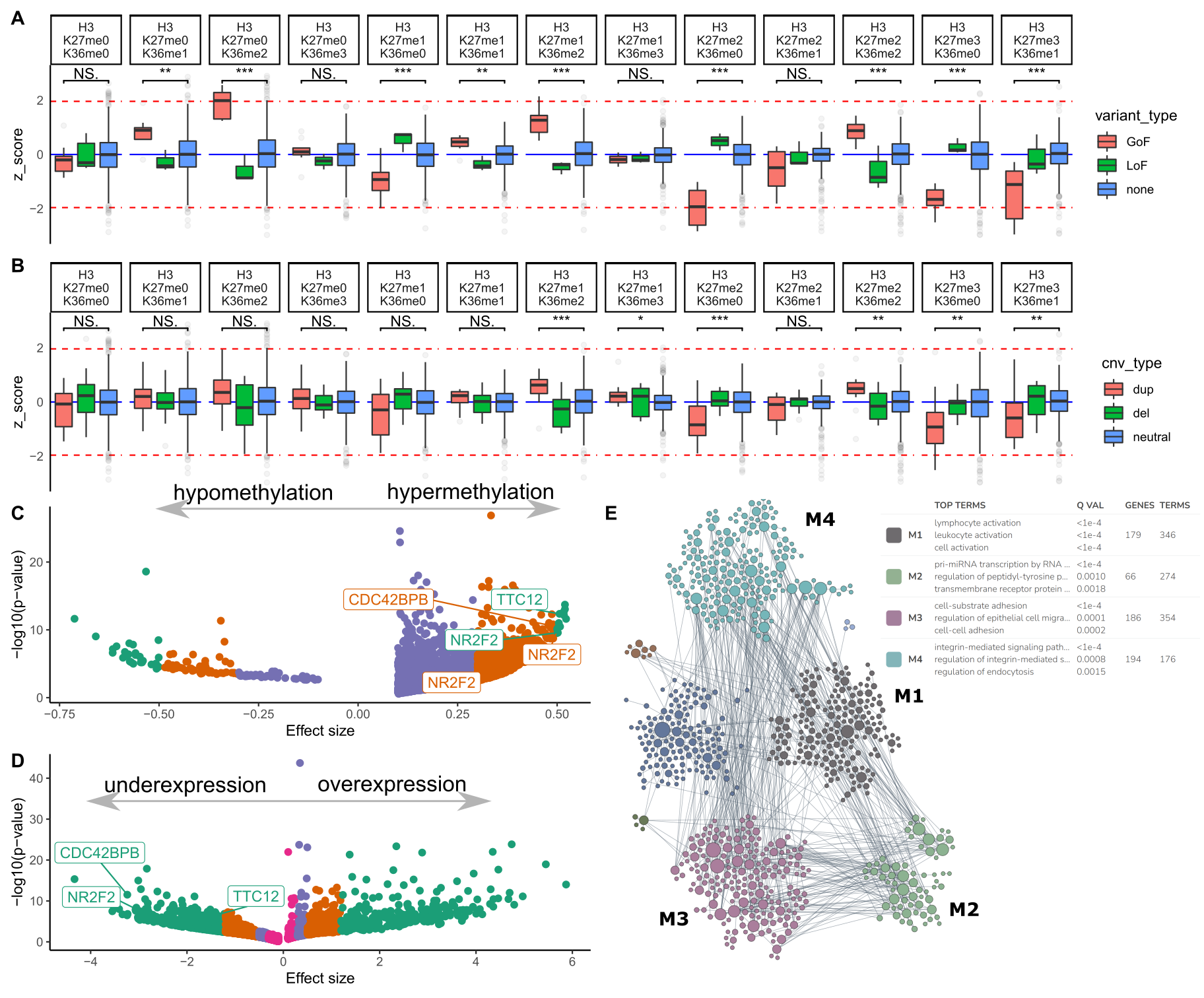
## **Figure S2** | Somatic NSD2 variant c.3295G>A, p.Glu1099Lys in an elderly person

**(A)** Screenshot of the “Age Distribution” from gnomAD browser (https://gnomad.broadinstitute.org/variant/4-1962801-G-A?dataset=gnomad\_r2\_1) taken on 2021-01-30 showing that one of 4,136 individuals in the 75-80 years age range carries the c.3295G>A, p.Glu1099Lys *NSD2* variant. **(B)** Screenshot of the interactive IGV.js visualization of the read data of this individual indicates somatic status of the variant because of the skewed variant allele fraction with 32/96 (36%). Somatic driver mutations deriving from a unique hematopoietic cell lineage are regularly encountered in genomic databases of healthy populations such as gnomAD and pose a pitfall when misinterpreted as germline variant.23 It is known 24 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 25 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. Diagnostic labs should be aware of this known pitfall and either use whitelists of somatic variants or allow a low allele count in *de novo* filters. 26

## 

## **Figure S3**. Variant distribution on NSD2

**(A)** Upper panel with schematic of the linear NSD2 protein with domains and variant distribution (missense variants in violet, truncating variants in red). Variants above the protein scheme have been described in germline 3–6, while the ones below are somatic from the COSMIC 11 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 27. Domains based on UniProt O96028. Lower panel with the generalized additive model of the CADD PHRED v1.6 values (blue line) for all possible NSD2 missense variants as indicator of conserved protein domains indicates highest constraint in the SET domain (red line represent the recommended CADD score cutoff). **(B)** Pseudo model (based on 7CRO 15) 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 violet and from the literature review as green spheres. The AA of the second described GoF missense (Thr1150) variant is also colored in violet. 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 S4** | 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 promotor 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).

In summary of these experiments, we used publicly available omics data from the Cancer Cell Line Encyclopedia (CCLE) 17 to confirm 28,29 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. 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.30 This effect of NSD2 on DNA methylation can be used in future studies to generate specific epi-signatures from peripheral blood DNA of primary patient material, 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, fewer than expected genes currently associated with NDDs in the SysID database overlapped with these dysregulated genesets. This observation, however is in line with the relatively mild developmental phenotype observed in Ind\_1 (postnatal 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 31 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. S4E). These enriched functions overlap well with known cellular functions of NSD2, previous expression and pathway analyzes 32 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 associated with overgrowth symptoms like macrocephaly and tall stature. Overall, these omic analyses presented here lack so far as only blood cancer cell lines were used. Future functional studies should investigate primary patient cells for methylation and expression, but also, for example, other cell lineages (brain, kidney, liver) through induced pluripotent cell lines and organoids to elucidate organ specific effects.

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 15,29. 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. S4B), which indicates that NSD2 could be triplosensitive (also confirmed by a recent publication 33 describing the “cross-disorder dosage sensitivity map of the human genome” with pHaplo=0.99528727797659 and pTriplo=0.986112792649751 for “WHSC1” as accessed from the supplementary file 34 downloaded from Zenodo) and contribute to the not well defined 4p16.3 microduplications. Descriptions of submicroscopic duplications encompassing *NSD2* are rare, and the phenotypes are often complicated by the size of the aberration or additional cytogenetic aberrations (e.g. Bi et al. 35).

# 

# 

# Figure S5 | cranial MRI of Ind\_1 at age 9 years and 2 months

# (A) T1 3D transversal shows marked hyperostosis of the calvaria (arrow). (B) Transversal MRI T2 weighting shows the deformation of both eyeballs (arrows).

# SUPPLEMENTARY TABLES

## **Table S1** | Phenotypic abnormalities of the face

| Phenotype | HPO | GoF cases | LoF cases |
| --- | --- | --- | --- |
| Abnormal facial shape | HP:0001999 | 100% (2/2) | 100% (16/16) |
| Coarse facial features | HP:0000280 | 100% (2/2) | 12% (2/16) |
| Triangular face | HP:0000325 | 0% (0/2) | 81% (13/16) |
| Square face | HP:0000321 | 100% (2/2) | 0% (0/16) |
| Thick hair | HP:0100874 | 100% (2/2) | 0% (0/1) |
| High anterior hairline | HP:0009890 | 0% (0/2) | 80% (12/15) |
| Low anterior hairline | HP:0000294 | 100% (2/2) | 13% (2/15) |
| 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) |
| Anteverted nares | HP:0000463 | 100% (2/2) | 0% (0/15) |
| Wide nasal bridge | HP:0000431 | 100% (2/2) | 39% (7/18) |
| Long philtrum | HP:0000343 | 100% (2/2) | 7% (1/15) |
| Short philtrum | HP:0000322 | 0% (0/2) | 40% (6/15) |
| Thin upper lip | HP:0000219 | 100% (2/2) | 35% (6/17) |
| Exaggerated cupid's bow | HP:0002263 | 100% (2/2) | 0% (0/14) |
| Posteriorly rotated ears | HP:0000358 | 50% (1/2) | 80% (4/5) |

## 

## 

## **Table S2** | Comparison of other phenotypes between of GoF and LoF cases

| group | Phenotype | HPO | GoF cases | LoF cases |
| --- | --- | --- | --- | --- |
| 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) |
| Tapered fingers | HP:0001182 | 100% (2/2) | 6% (1/18) |
| Short neck | HP:0000470 | 100% (2/2) | 0% (0/12) |
| Brachydactyly | HP:0001156 | 100% (2/2) | NaN% (0/0) |
| Hypoplastic fingernails and toenails | HP:0001804; HP:0001800 | 100% (2/2) | NaN% (0/0) |
| Large hands | HP:0001176 | 100% (2/2) | NaN% (0/0) |
| Single transverse palmar crease | HP:0000954 | 100% (2/2) | NaN% (0/0) |
| 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) |
| moderate | HP:0011343; HP:0002342 | 100% (2/2) | 0% (0/19) |
| mild | HP:0011342; HP:0001256 | 0% (0/2) | 74% (14/19) |
| 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) |
| Abnormal heart morphology | HP:0001627 | 100% (2/2) | 14% (3/21) |
| 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) |

# SUPPLEMENTARY DATA FILES

These data files used for analyses and tables/ figures are available for download from Zenodo:

## **File S2** | Clinical information of individuals described here and reviewed individuals together with all variant data.36

## **File S3** | Cell line analyses data.36

## **File S4** | humanbase functional modules data.36

# REFERENCES

1. 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).

2. Neuser, S. *et al.* Prenatal phenotype of PNKP-related primary microcephaly associated with variants affecting both the FHA and phosphatase domain. *Eur. J. Hum. Genet.* **30**, 101–110 (2022).

3. 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).

4. 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).

5. 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).

6. Zanoni, 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).

7. 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).

8. Jiang, Y. *et al.* De novo truncating variant in NSD2gene leading to atypical Wolf-Hirschhorn syndrome phenotype. *BMC Med. Genet.* **20**, 134 (2019).

9. Köhler, S. *et al.* The Human Phenotype Ontology in 2021. *Nucleic Acids Res.* **49**, D1207–D1217 (2021).

10. 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).

11. Tate, J. G. *et al.* COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* **47**, D941–D947 (2019).

12. Karczewski, K. J. *et al. The mutational constraint spectrum quantified from variation in 141,456 humans*. http://biorxiv.org/lookup/doi/10.1101/531210 (2019) doi:10.1101/531210.

13. Hebebrand, M. *et al.* The mutational and phenotypic spectrum of TUBA1A-associated tubulinopathy. *Orphanet J. Rare Dis.* **14**, 38 (2019).

14. 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).

15. Li, W. *et al.* Molecular basis of nucleosomal H3K36 methylation by NSD methyltransferases. *Nature* **590**, 498–503 (2021).

16. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).

17. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603–607 (2012).

18. Ghandi, M. *et al.* Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569**, 503–508 (2019).

19. Laurent Gatto, Theo Killian. depmap. doi:10.18129/B9.BIOC.DEPMAP.

20. Greene, C. S. *et al.* Understanding multicellular function and disease with human tissue-specific networks. *Nat. Genet.* **47**, 569–576 (2015).

21. 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).

22. 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).

23. Chao, E. C. *et al.* Incidental detection of acquired variants in germline genetic and genomic testing: a points to consider statement of the American College of Medical Genetics and Genomics (ACMG). *Genet. Med. Off. J. Am. Coll. Med. Genet.* **23**, 1179–1184 (2021).

24. 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).

25. 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.

26. 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.

27. Rentzsch, P., Witten, D., Cooper, G. M., Shendure, J. & Kircher, M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res.* **47**, D886–D894 (2019).

28. Li, J. *et al.* NSD2-E1099K Mutation Leads to Glucocorticoid-Resistant B Cell Lymphocytic Leukemia in Mice. *Blood* **136**, 3–4 (2020).

29. 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).

30. Weinberg, D. N. *et al.* The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature* **573**, 281–286 (2019).

31. Krishnan, A. *et al.* Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. *Nat. Neurosci.* **19**, 1454–1462 (2016).

32. Swaroop, A. *et al.* An activating mutation of the NSD2 histone methyltransferase drives oncogenic reprogramming in acute lymphocytic leukemia. *Oncogene* **38**, 671–686 (2019).

33. Collins, R. L. *et al.* A cross-disorder dosage sensitivity map of the human genome. *Cell* **185**, 3041-3055.e25 (2022).

34. Collins, Ryan *et al.* A cross-disorder dosage sensitivity map of the human genome. (2022) doi:10.5281/ZENODO.6347673.

35. Bi, W., Cheung, S.-W., Breman, A. M. & Bacino, C. A. 4p16.3 microdeletions and microduplications detected by chromosomal microarray analysis: New insights into mechanisms and critical regions. *Am. J. Med. Genet. A.* **170**, 2540–2550 (2016).

36. Popp, Bernt. Data files for manuscript ‘A novel syndrome caused by the constitutional gain-of-function variant p.Glu1099Lys in NSD2’. (2022) doi:10.5281/ZENODO.6206868.