# Rare *de novo* gain-of-function missense variants in *DOT1L* are associated with developmental delay and congenital anomalies


## Authors

Zelha Nil, Ashish R. Deshwar, Yan Huang, ..., Shinya Yamamoto, Gregory Costain, Hugo J. Bellen

## Correspondence

gregory.costain@sickkids.ca (G.C.), hbellen@bcm.edu (H.J.B.)



**This study identified individuals with *de novo DOT1L* variants associated with developmental delay and congenital anomalies. *DOT1L* encodes a histone methyltransferase and the variants led to increased methylation in flies and human cells. Hence, *DOT1L* now joins the list of histone lysine methyltransferases associated with Mendelian developmental disorders.**



Nil et al., 2023, The American Journal of Human Genetics *110*, 1919–1937
November 2, 2023



https://doi.org/10.1016/j.ajhg.2023.09.009

# Rare *de novo* gain-of-function missense variants in *DOT1L* are associated with developmental delay and congenital anomalies

Zelha Nil,[1,2,26] Ashish R. Deshwar,[3,4,26] Yan Huang,[1,2] Scott Barish,[1] Xi Zhang,[5,6] Sanaa Choufani,[4] Polona Le Quesne Stabej,[7] Ian Hayes,[8] Patrick Yap,[8] Chad Haldeman-Englert,[9] Carolyn Wilson,[9] Trine Prescott,[10] Kristian Tveten,[10] Arve Vøllo,[11] Devon Haynes,[12,13] Patricia G. Wheeler,[12] Jessica Zon,[3] Cheryl Cytrynbaum,[3,4] Rebekah Jobling,[3] Moira Blyth,[14] Siddharth Banka,[15,16] Alexandra Afenjar,[17] Cyril Mignot,[18,19] Florence Robin-Renaldo,[20] Boris Keren,[21] Oguz Kanca,[1,2] Xiao Mao,[6,22] Daniel J. Wegner,[23] Kathleen Sisco,[23] Marwan Shinawi,[23] Undiagnosed Disease Network,[27] Michael F. Wangler,[1,2] Rosanna Weksberg,[3,5,24] Shinya Yamamoto,[1,2,25] Gregory Costain,[3,4,24,*] and Hugo J. Bellen[1,2,25,*]

## Summary

Misregulation of histone lysine methylation is associated with several human cancers and with human developmental disorders. *DOT1L* is an evolutionarily conserved gene encoding a lysine methyltransferase (KMT) that methylates histone 3 lysine-79 (H3K79) and was not previously associated with a Mendelian disease in OMIM. We have identified nine unrelated individuals with seven different *de novo* heterozygous missense variants in *DOT1L* through the Undiagnosed Disease Network (UDN), the SickKids Complex Care genomics project, and GeneMatcher. All probands had some degree of global developmental delay/intellectual disability, and most had one or more major congenital anomalies. To assess the pathogenicity of the *DOT1L* variants, functional studies were performed in *Drosophila* and human cells. The fruit fly *DOT1L* ortholog, *grappa*, is expressed in most cells including neurons in the central nervous system. The identified *DOT1L* variants behave as gain-of-function alleles in flies and lead to increased H3K79 methylation levels in flies and human cells. Our results show that human *DOT1L* and fly *grappa* are required for proper development and that *de novo* heterozygous variants in *DOT1L* are associated with a Mendelian disease.

## Introduction

Dynamic changes in chromatin state through post-translational modifications (PTMs) of histone proteins play a key role in gene expression regulation and genomic stability.[1] One of the most common types of chromatin modifications is the methylation of lysine residues at histone tails, which is known to affect DNA repair, transcription, stress response, and differentiation.[2] Misregulation of histone lysine methylation is associated with several human cancers and other disorders.[3]

Methylation of different types of histones at specific lysine residues is catalyzed by specific histone lysine methyltransferases (KMTs) and there are thirty-four KMT-encoding genes in humans.[4] Of these, dysregulation (typically haploinsufficiency) of sixteen genes encoding KMTs have been previously associated with dominant human developmental disorders (Table S1). Most of these disorders of

[1]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA; [2]Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, TX 77030, USA; [3]Division of Clinical and Metabolic Genetics, The Hospital for Sick Children, Toronto, ON, Canada; [4]Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON, Canada; [5]Department of Neurology, Xiangya Hospital, Central South University, Changsha 410008, China; [6]National Health Commission Key Laboratory for Birth Defect Research and Prevention, Hunan Provincial Maternal and Child Health Care Hospital, Changsha 410005, China; [7]Department of Molecular Medicine and Pathology, Faculty of Medical and Health Sciences, the University of Auckland, Auckland, New Zealand; [8]Genetic Health Service New Zealand- Northern Hub, Auckland District Health Board, Auckland, New Zealand; [9]Mission Fullerton Genetics Center, Asheville, NC 28803, USA; [10]Department of Medical Genetics, Telemark Hospital Trust, 3710 Skien, Norway; [11]Department of Pediatrics, Hospital of Østfold, 1714 Grålum, Norway; [12]Division of Genetics, Arnold Palmer Hospital for Children – Orlando Health, Orlando, FL, USA; [13]Clinical Genetics Service, Guy's Hospital, Guy's and St Thomas' NHS Trust, London, England, UK; [14]North of Scotland Regional Genetics Service, Clinical Genetics Centre, Ashgrove House, Foresterhill, Aberdeen, UK; [15]Division of Evolution, Infection and Genomics, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, M13 9WL Manchester, UK; [16]Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester University NHS Foundation Trust, Health Innovation Manchester, M13 9WL Manchester, UK; [17]Service de génétique, CRMR des malformations et maladies congénitales du cervelet et CRMR déficience intellectuelle, hôpital Trousseau, AP-HP, Paris, France; [18]Sorbonne Université, Département de Génétique, Groupe Hospitalier Pitié-Salpêtrière and Hôpital Trousseau, Paris, France; [19]Centre de Référence Déficiences Intellectuelles de Causes Rares, Paris, France; [20]AP-HP, Sorbonne Université, Service de Neurpoédiatrie, Hôpital Trousseau, Paris, France; [21]AP-HP, Hôpital de la Pitié-Salpêtrière, Département de Génétique, 75013 Paris, France; [22]Clinical Research Center for Placental Medicine in Hunan Province, Hunan Provincial Maternal and Child Health Care Hospital, Changsha 410005, China; [23]Department of Pediatrics, Washington University School of Medicine, St. Louis, MO 63110, USA; [24]Department of Molecular Genetics, University of Toronto, Toronto, ON, Canada; [25]Department of Neuroscience, Baylor College of Medicine, Houston, TX 77030, USA
[26]These authors contributed equally
[27]Further details can be found in the supplemental information
*Correspondence: gregory.costain@sickkids.ca (G.C.), hbellen@bcm.edu (H.J.B.)
https://doi.org/10.1016/j.ajhg.2023.09.009.

epigenetic regulation are also now known to have unique patterns of genome-wide DNA methylation alterations ("DNAm signatures") in peripheral blood (Table S1). A review of the orthologs in available animal models of genes encoding KMTs demonstrates that both disease-associated and non-disease-associated genes encoding KMTs are important in animal development.[5] Therefore, it is likely that additional KMTs will be associated with Mendelian disease over time.

*DOT1 Like histone lysine methyltransferase* (*DOT1L* [MIM: 607375]) is an evolutionarily conserved KMT found in a broad range of eukaryotic species including yeast, fruit fly, zebrafish, mice, and human.[6] Compared to other KMTs, DOT1L has a number of distinct characteristics. It is the only KMT that does not contain the SET catalytic domain; instead, it contains an N-terminal Dot1 catalytic domain.[7,8] Unlike SET-domain containing KMTs, which target lysine residues in the histone tails where most modifications occur, DOT1L methylates a lysine residue (K79) in the core globular domain of histone H3, and to our knowledge, it is the only enzyme known to do so.[9] DOT1L, in contrast to other KMTs, seems to be a distributive enzyme that sequentially adds methyl groups to H3K79.[10,11] It also differs from other KMTs as it methylates H3K79 only in the form of intact chromatin and not as free histones.[12–15]

DOT1L and H3K79 methylation have been implicated in several fundamental cellular processes, including gene expression,[6] cell cycle regulation,[16] DNA damage response,[17] telomeric silencing,[18] proliferation,[19] and differentiation.[20] Distribution of H3K79 methylation throughout the chromatin as well as DOT1L are strongly associated with actively transcribed genes.[21–24] Indeed, DOT1L directly interacts with elongating RNA Polymerase II and gets recruited to actively transcribed genes.[25] Possibly owing to its aforementioned cellular roles, DOT1L is a critical regulator during development, particularly in hematopoiesis of some metazoans. In mice, germline deletion of *Dot1l* leads to embryonic lethality due to defects in the yolk sack and heart and failure of primitive hematopoiesis.[26] Similarly, *dot1l*-depleted zebrafish morphants show impaired growth, defective angiogenesis, and cardiac dilatation.[27] Loss-of-function (LoF) mutants for the fly ortholog of *DOT1L*, called *grappa* (*gpp*), exhibit slow growth and die at larval stages.[28] Furthermore, knockdown of *DOT1L* orthologs in the above listed animal models decreases expression of several Wnt signaling target genes.[29–31] This suggests that loss of *DOT1L* affects Wnt signaling in multiple species but the precise mechanism as to how DOT1L regulates development of specific tissues and organs remains elusive.

Here, we present a cohort of nine individuals with a total of seven different *de novo* missense variants in *DOT1L*. Their phenotypes include central nervous system (CNS) dysfunction and a range of congenital anomalies, consistent with the variability associated with other KMT disorders (Table S1). We provide functional data in flies showing that six of the variants cause an increase in DOT1L enzymatic activity and disrupt development. We also show the remaining variant alters DOT1L activity in human cells. Our data support that *de novo* missense variants in *DOT1L* cause a KMT disorder through a gain-of-function mechanism.

## Subjects and methods

### Subject recruitment and sequencing

The two index families were identified through the Undiagnosed Disease Network (UDN)[32,33] and the SickKids Complex Care genomics project,[34,35] respectively. The five additional participants were identified through GeneMatcher.[36,37] The initial GeneMatcher entry was made on October 17, 2017, and all matches made until 2020 were considered in this study (Table S2). Two additional individuals (probands #5 and #8) were identified through a pre-existing collaboration. Two families (three total individuals) with rare *DOT1L* missense variants suspected to be non-diagnostic also participated in aspects of this study for comparison purposes; they were identified through GeneMatcher (Table S2) and through a prior publication,[5,38] respectively.

Genome- and exome-wide sequencing was performed at each participating center in either a clinical diagnostic laboratory or a research laboratory, using standard methods for sequencing, quality control, and analysis (Table S3). All *DOT1L* variants were high-confidence calls, and maternity and paternity were confirmed by a trio genome- or exome-wide sequencing design and/or other standard methods. This research was conducted with the voluntary, informed consent of any research participants, free of coercion or coercive circumstances, and received Institutional Review Board (IRB) or Research Ethics Committee (REC) approval consistent with the principles of research ethics and the legal requirements of the lead authors' jurisdiction(s) (The Hospital for Sick Children, Canada; Baylor College of Medicine and NIH, USA).

### DNAm profiling and data processing

Genomic DNA was extracted from peripheral blood and bisulfite converted using the EpiTect Bisulfite Kit (EpiTect PLUS Bisulfite Kit, QIAGEN). Sodium-bisulfite-converted DNA was then hybridized to the Illumina Infinium Human Methylation EPIC BeadChip to interrogate more than 850,000 CpG sites in the human genome at The Center for Applied Genomics (TCAG), Hospital for Sick Children Research Institute, Toronto, Ontario, Canada. On each microarray chip, cases and controls were randomly assigned a chip position. The minfi Bioconductor package in R was used to preprocess data including quality control, Illumina normalization, and background subtraction, followed by extraction of beta (β) values.[39] Standard quality control metrics in minfi were used, including median intensity QC plots, density plots, and control probe plots. Probes with detection flaws (n = 691), probes near SNPs with minor allele frequencies above 5% (n = 91,246), cross-reactive probes (n = 39,371), probes with raw beta of 0 or 1 in >0.25% of samples (n = 236), non-CpG probes (n = 2,635), and X and Y chromosome probes (n = 18,653) were removed, resulting in a total of n = 716,401 probes remained for differential methylation analysis.

### DNAm signature exploration

To assess DNAm patterns, we identified differentially methylated sites in whole-blood-derived DNA from individuals with suspected

deleterious variants in *DOT1L*, sex- and age-matched typically developing controls, and individuals with suspected non-deleterious variants in *DOT1L* (Table S7). For all samples, we applied the blood cell-type proportion estimation tool in minfi based on Illumina EPIC array data from FACS-sorted blood cells.[40] Two of the samples were not useable as part of a discovery cohort for technical reasons (age outliers), corresponding to the variants p.Cys45Gly (DNA sample obtained when the individual was an infant) and p.Lys1025Met (DNA sample from adult), and no sample was available for the deceased study participant. The DOT1L discovery cohort ($n = 4$) therefore included 2 females and 2 males with mean age at sample collection of 4.2 ± 2.2 years (range 2–8 years). The 12 sex- and age-matched control subjects included 8 females and 4 males and mean age at sample collection of 4.8 ± 2.3 years (range 1.3–9 years). We identified differentially methylated CpG sites using Limma[41] regression modeling with age and PC1 as covariates. The thresholds for differentially methylated CpG sites were false discovery rate (FDR)-adjusted p value $< 0.3$ and a $|\Delta\beta| > 0.10$. $\Delta\beta$ represents the difference in average DNAm ($\beta$) between groups. Principal component analysis (PCA) and hierarchical clustering were generated using Qlucore Omics Explorer (QOE, www.qlucore.com).

## Fly husbandry and fly stocks

All flies were raised at room temperature, unless otherwise noted, on standard cornmeal and molasses medium in plastic vials.

All fly strains used in this study were created in house or obtained from the Bloomington *Drosophila* Stock Center (BDSC) or Kyoto Stock Center. The $gpp^{TG4}$ allele was created by insertion of an artificial exon containing a *T2AGAL4* cassette into the third coding intron of *gpp* via CRISPR-mediated homologous recombination.[42,43] $gpp^{xxv}$ allele was created via X-ray mutagenesis and characterized previously.[28] For complementation tests, immunostaining, and RNAi assays, see genotypes in Table S4. UAS-Empty is an empty *pGW-HA.attB* plasmid injected into the identical docking site as a control and has been published previously.[44]

## Generation of human *UAS-DOT1L* transgenic lines

The *DOT1L* cDNA clone used in this study corresponds to GenBank: NM_032482.3, which encodes the canonical full-length DOT1L isoform and is defined as reference here. *DOT1L* variants were created via site directed mutagenesis (Q5 Site-Directed Mutagenesis Kit, NEB) on a *pENTR223.1-DOT1L* cDNA construct (Horizon Discovery, clone ID 100062384). Primers are detailed in Table S5. Reference and mutagenized cDNA products were carried into *pGW-HA-attB*- gateway expression vector[45] using LR clonase II (ThermoFisher). Since the *DOT1L* cDNA clone does not contain a translational stop codon, all constructs are tagged with 3X-HA tag and a linker sequence. The constructs were sequence verified and injected into *VK37* (BDSC #24872) docking site by phiC31-mediated transgenesis.[46]

## Imaging of fly wings and eyes and of whole animals

For imaging of wings, fly wings were plucked and placed on a white filter paper. For imaging of eyes, whole flies were placed on a white filter paper in an angle so that the whole eye can be viewed. For imaging of whole larvae and adults, animals were placed on a black filter paper. Images were acquired using a digital camera (MicroFire; Olympus) mounted on a stereomicroscope (MZ16; Leica) with ImagePro Plus 5.0 acquisition software (Media Cybernetics). z stacks were taken by the "extend depth of field" function of the AxioVision software (web resources). Selected range of z stack images were max-projected using Imaris (web resources).

## Real-time qPCR

Total RNA was extracted from larvae and adult flies using TRIzol (Invitrogen) according to manufacturer's protocol. Biological triplicate samples of whole animals were processed per condition. qPCR was performed as previously described.[44] High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems), SYBR Green Fast Master Mix (Thermo Fisher), and a BioRad C1000 Touch Cycler were used. A housekeeping gene, *RpL32*, was included for normalization. qPCR primers are listed in Table S5.

## Immunostaining and confocal microscopy

Immunostaining of larval brains and imaginal discs, and adult brains were done as previously described.[47] Briefly, dissected tissues were fixed in 4% paraformaldehyde (PFA) and blocked in 5% normal goat serum in 0.1% Triton X-100 in 1× PBS. Primary antibodies and dilutions used as follows: anti-Elav (DSHB #7E8A10; 1:250), anti-Repo (DSHB #8D12; 1:50). Donkey-derived secondary antibodies were used at 1:250 (Jackson ImmunoResearch Laboratories). A Leica Sp8x with lightning deconvolution was used for confocal microscopy. Images were taken with either a 20× or 40× Leica objective and processed using Imaris.

## Expression of human *DOT1L* in *Drosophila*

We overexpressed reference and variant *DOT1L* cDNA molecules in flies in *gpp* expression domains by crossing the *UAS-DOT1L* males to virgin female flies from $gpp^{TG4}$ driver stock. The progeny of these crosses was cultured at 25°C. More than 100 flies were assessed for the expected Mendelian ratios. Percent viabilities (o/e ratios) from three independent experiments were plotted as mean ± SEM, and statistical significance was determined by one-way ANOVA for multiple groups.

For wing phenotypic distribution, the surviving overexpression flies were collected. For each genotype, more than 100 wings (>50 flies) were randomly selected and separated into different groups based on the severity of their wing phenotypes.

## Detection of DOT1L and H3K79 methylation levels using western blot

We performed western blot analysis to detect DOT1L and H3K79 methylation status from larva and adult flies. For larval lysates, 10 animals were used. For adult lysates, 5 female and 5 male flies were used. Animals were lysed using a pestle with a cordless motor (VWR) in a buffer (30 μL per animal) containing 150 mM NaCl, 50 mM Tris-Cl (pH 8.0), 1.0% NP-40, 0.1% Triton X-100, 0.4% Na- Deoxycholate, 0.1% SDS in the presence of EDTA-free protease inhibitor (Roche), followed by rotation at 4°C for 20 min. After addition 1× lysis buffer without SDS and Na-Deoxycholate, lysates were sonicated at 20 kHz for 10 cycles of 10 s "on" and 10 s "off" and centrifuged at 1,000 rpm for 2 min. Supernatants were collected and treated with 50U Benzonase (EMD Milipore), 20 μg RNase A (QIAGEN), and 4 mM $MgCl_2$ (final concentration) at 4°C for 30 min. Samples were centrifuged at 5,000 rpm at 4°C for 10 min and the cleared supernatant was collected in fresh 1.5 mL Eppendorf tube. Total of 50 μg protein for each sample was boiled with 2× Laemmli sample buffer. For detection of histones, samples were resolved on a 12% Bis-Tris SDS-polyacrylamide gel (Invitrogen) and immunoblotted with Rabbit

anti-H3K79Me (Ab177185, Abcam, 1:1,000) and Mouse anti-H3 (05-499, EMD-Milipore, 1:1,000) antibodies. For detection of DOT1L, samples were resolved on a 3%–8% Tris-Acetate SDS-polyacrylamide gel (Invitrogen) and immunoblotted with Mouse anti-HA.11 (16B12, Biolegend, 1:1,000) and Rabbit anti-α-Tubulin (11H10, Cell signaling, 1:5,000) antibodies. Secondary antibodies Goat anti-Rabbit IRDye680 (Licor) and Goat anti-Mouse IRDye800 (Licor) were used in 1:5,000 dilution.

### Recombinant plasmid construction for human and mouse cell culture assays

The full-length human reference *DOT1L* coding sequence (GenBank: NM_032482.3) was downloaded from NCBI and inserted into pcDNA5 vector, which was added a Flag exogenous tag (DYDDDK) at N-terminal of DOT1L. Meanwhile, two variants of *DOT1L* were constructed into the same vector pcDNA5 via site directed mutagenesis. The knockdown efficiency of the *Dot1l*-shRNA plasmid was verified through N2A cell line. The shRNA sequence: 5′-CCGGTCGCCAACACGAGTGTTATATCTCGAGATA TAACACTCGTGTTGGCGATTTTTT-3′.

### HEK293 cell culture

HEK293 FLP-In cell lines were bought from Thermo Fisher (Invitrogen, K6010–01), which was cultured with DMEM basic (GIBCO, C11995500BT), 10% fetal bovine serum (FBS, GIBCO, 12483020), and penicillin and streptomycin (50 μg/mL) (Invitrogen) and placed into 37°C and 5% $CO_2$ incubator. We used Lipofectamine 3000 Reagent (Invitrogen) to transfect *DOT1L* reference and variant plasmids into cell lines according to the manual.

### Mouse primary neural cell culture

The pregnant mice were sacrificed at day 16.5 and the embryos were dissected out in 1× pre-cooled HBSS. The olfactory bulb and vascular membrane of embryo brains was removed and discarded. The remaining part of the brains were cut it into pieces with scissors and digested in papain at 37°C for 30 min shaking gently every 10 min. After addition of 2 volumes of pre-warmed FBS medium, the neurons were separated gently and filtered with a 40 μm cell strainer. The cells were seeded (150,000 per well) into 24-well plates which contained a poly-lysine-coated coverslip at the bottom. After cells adhered to the cover slips, medium was replaced with pre-warmed B27 medium. We used Lipofectamine 3000 Reagent (Invitrogen) to transfect *Dot1l*-shRNA plasmids and control plasmids into the primary neurons according to the manual.

### Western blot from human cells

The total protein was extracted using RIPA cell lysis buffer (Solarbio #R0010) containg 1% Phenylmethylsulfonyl fluoride (PMSF, Solarbio). The BioRad Mini-PROTEANTM Tetra System was used to electrophorese with ExpressPlus 8% precast gel (GenScript). 25 μg of total protein per sample were run and transferred to a nitrocellulose membrane and immunoblotted with Rabbit anti-H3K79Me3 (A2369, Abclonal), anti-Flag and anti-β-Tubulin antibodies. The protein levels were detected by ECL chemiluminescence.

## Results

### Summary of clinical findings

Through the UDN,[32,33] SickKids Complex Care genomics project,[34,35] and GeneMatcher,[36,37] we identified nine unrelated individuals (probands 1–9) in whom exome or genome sequencing highlighted candidate variants in *DOT1L*. All seven distinct variants are missense variants, and in all but one family (with the recurrent p.Glu123Lys variant, where the child was deceased and parents declined further testing), the variant was confirmed to be *de novo*. The comprehensive clinical information of the probands can be found in Tables 1 and S6. Throughout this report, all sequence variants are described based on reference sequence GenBank: NM_032482.3.

All identified probands had some degree of global developmental delay (9/9) and most (5/9) were found to have hypotonia. Whereas motor delay was usually relatively mild, significant speech and language delay was observed in all individuals in which details were provided (seven individuals): two were non-verbal and five were able to speak only individual words. All nine probands were found to have various congenital anomalies although there was considerable heterogeneity across the cohort. Seven had brain structural anomalies, five had cardiac defects (four with septation defects), and five presented with varied urogenital features (of which two had hypospadias). Other common features included hip laxity (3/9), variable ophthalmological features (3/9), and sensorineural hearing loss (3/9). Growth restriction was also a feature seen in the cohort, with five individuals reported to have post-natal microcephaly (<3$^{rd}$ centile for occipital-frontal circumference [OFC]) and four with short stature (<3$^{rd}$ centile for height). While distinctive facial features were described for all probands (9/9), no recognizable pattern was appreciated in this cohort.

In addition to the nine probands described above, we also identified two families with missense variants in *DOT1L* that were not suspected to be diagnostic by the referring providers. One was heterozygous for a missense variant (c.1352A>G [p.Asp451Gly]) inherited from an unaffected father that had an allele frequency inconsistent with a high-penetrance Mendelian developmental disorder (Table S8), and the other was heterozygous for a *de novo* missense variant (c.874C>T [p.Arg292Cys]) reported previously[5,38] where an alternative genetic diagnosis likely explains the entirety of her clinical features. Clinical features for both individuals can be found in Table S7, and they were included in aspects of our functional studies as negative controls (see below).

### Bioinformatic analyses of *DOT1L de novo* variants are inconclusive

We performed *in silico* analyses using MARRVEL (Model organism Aggregated resources for Rare Variant ExpLoration)[48,49] to collect information about the human *DOT1L* and its orthologous genes in genetic model organisms. *DOT1L* is intolerant LoF with a probability LoF intolerant (pLI) score of 1.0 and a LoF variant observed/expected (o/e) ratio of 0.046 (0.02–0.12). Variants in *DOT1L* are predicted to have a high "probability of being autosomal dominant" based on DOMINO[50] (Figure 1A).

**Table 1. Clinical features of individuals with *DOT1L* variants**

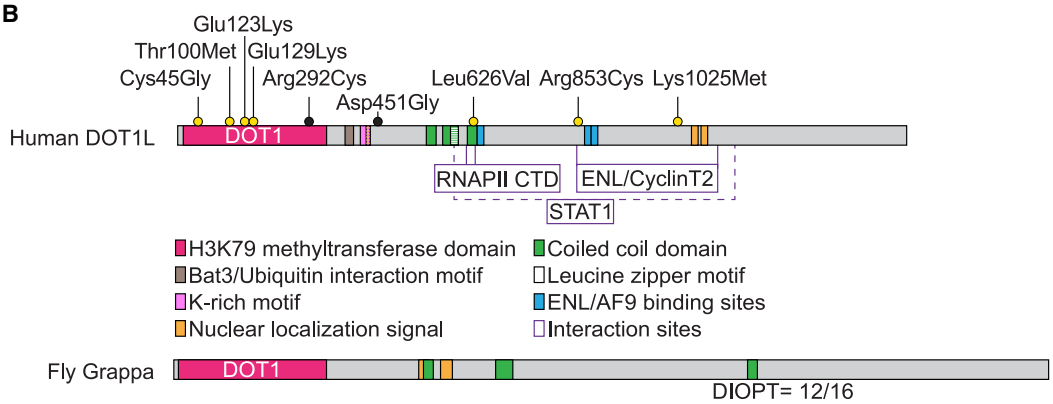
| Proband | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| *DOT1L* variant (NM_032482.3) | c.133T>G (p.Cys45Gly) | c.299C>T (p.Thr100Met) | c.367G>A (p.Glu123Lys) | c.367G>A (p.Glu123Lys) | c.367G>A (p.Glu123Lys) | c.385G>A (p.Glu129Lys) | c.1876C>G (p.Leu626Val) | c.2557C>T (p.Arg853Cys) | c.3074A>T (p.Lys1025Met) |
| Inheritance | *de novo* | *de novo* | unknown | *de novo* | *de novo* | *de novo* | *de novo* | *de novo* | *de novo* |
| Sex | male | female | male | male | female | male | male | male | male |
| Age at last assessment | 5 years | 4 years 4 months (age at death) | 4 months (later died in infancy) | 5 years | 3 years 11 months | 6 years | 7 years | 4 years | 21 years |
| **Medical history** | | | | | | | | | |
| Brain anomalies (MRI/CT) | cortical dysplasia, periventricular heterotopia | no | HIE | bilateral brain atrophy, bilateral frontotemporal extra brain space widening, multiple small cysts in the corpus callosum | ONH/tract truncated corpus callosum | enlarged VRS, tentorial herniation | normal brain MRI | bilateral brain atrophy, focal cortical dysplasia | cerebellar atrophy, megalencephaly |
| Cardiac anomalies | ASD, PLSVC | myocarditis with cardiomyopathy; resolved by 1 year old | large VSD, moderate size ASD/PDA | ASD | no | no | VSD | no | no |
| Hypotonia | yes | yes | yes | no | yes | yes | no | no | no |
| Musculoskeletal anomalies | hip subluxation | hyperflexible hip joints | tall vertebral bodies, anomalous ribs (first 3), non ossified patella | N/D | hip laxity | no | pain after physical exercise, low endurance | N/D | no |
| Urogenital anomalies | hypospadias with chordee | no | hypospadias | cryptorchidism | grade 1 VUR | no | small penis | no | no |
| Hearing loss | sensorineural and conductive; stenosis external auditory canal | lack of reaction to loud stimuli or startle responses at 9 months | no | mild hearing loss in both ears | no | no | sensorineural w/ bilateral SCC dysplasia | no | no |
| **Growth and development** | | | | | | | | | |
| Global developmental delay | severe | yes | severe | severe | severe | severe | yes | mild | yes |
| Language | non-verbal | non-verbal | N/A (died in infancy) | only a few words | non-verbal | only a few words | only a few words | only a few words | only a few words |
| Stature (percentile) | short (2.94%) | short (4.95%) | N/D | N/D | normal (93.3%) | short (0.64%) | short (0.40%) | N/D | normal (39.7%) |
| Microcephaly (percentile) | no (97%) | no (60%) | yes (<3.0%) | N/D | yes (4.0%) | yes (3.0%) | yes (<1.0%) | yes (N/D) | no (>99%) |

Abbreviations are as follows: HIE, hypoxic ischemic encephalopathy; ONH, optic nerve hypoplasia; VRS, Virchow Robin spaces; ASD, atrial septal defect; PLSVC, persistent left supervior vena cava; VSD, ventricular septal defect; PDA, patent ductus arteriosus; VUR, vesicoureteral reflux; SCC, semi-circular canal; N/D, no data.

**A**

| Probability of LoF Intolerance | | | | Missense Constraint | | | DOMINO | |
|---|---|---|---|---|---|---|---|---|
| Expected | Observed | pLI score | o/e score | Expected | Observed | z score | Inheritance prediction | Probability of being AD |
| 65.8 | 3 | 1.0 | 0.05 | 975.5 | 736 | 2.72 | Either dominant or recessive | 0.599 |

**B**




**C** Protein sequence alignment of human DOT1L with fly grappa

```
Human --MGEKLELRLKSPVGAE-PAVYPWPLPVY--DKHHDAAHEIIETIRWVCEEIPDLKLAMENYVLIDYDTKSFESMQRLCDKYNRAIDSIHQLWKGTTQPM-KLNTRPSTGLLRHI [110]
Fly   MATPQVKDLVLRSPAGSSDVISFAWPLQIGHGQDKHDNGIDIIDTIKFVCDELPSMSSAFEETNLHQIDTACYKTMTGLVDRFNKAVDSIVALEKGTSLPAERLNKFAHPSLLRHI [116]
      :  :* *:**.*:.    : *** :   :.:** . :**:**::**:*:*.:. *:*:  * : ** .:::*  * *::*:*:*** * ***: * :**.    .*****
Human LQQVYNHSVTDPEKLNNYEPFSPEVYGETSFDLVAQMIDEIKMTDDDLFVDLGSGVGQVVLQVAAATNCKHHYGVEKADIPAKYAETMDREFRKWMKWYGKKHAEYTLERGDFLSE [226]
Fly   LQLVYNAAVLDPDKLNQYEPFSPEVYGETSYELVQQMLKHVTVSKEDTFIDLGSGVGQVVLQMAGSFPLKTCIGIEKADTPARYAERMDVIFRQYMGWFGKRFCEYKLIKGDFLVD [232]
      ** *** :* **:***:************::** **:.:.:.:.:* *:************:*.:   *   *:**** **:*** **  **::* *:**:..**.* :**** :
Human EWRERIANTSVIFVNNFAFGPEVDHQLKERFANMKEGGRIVSSKPFAPLNFRINSRNLSDIGTIMRVVELSPLKGSVSWTGKPVSYYLHTIDRTILENYFSSLKNPKLREEQEAAR [342]
Fly   EHRENITSSTLVFVNNFAFGPTVDHQLKERFADLRDGARVVSSKSFCPLNFRITDRNLSDIGTIMHVSEIPPLKGSVSWTCKPVSYYLHVIDRTILERYFQRLKTKGGNDHESVGT [348]
      * **.*:.:::::********* **********:::*.*:**** *.******..**********:* *: ********* ********.*******.**. **.   .:.:...


Human LDALHAQTVSQTAASSPQDAYR-SPHSPFYQLPPSVQR-----HSPNPLLVAPTPPALQKLLESFKIQYLQFLAYTKTPQYKASLQELLGQEKEKNAQLLGAAQQLLSHCQAQKEE [542]
Fly   LEALHKQTVLSTSLDTMTKKLPAAPGTVDQQLTALLTENMSHAELDIPTAPQDTPYALQILLDVFRSQYTSMIEHMKSSAYVPQVQKQIAQEQERMARLKNRASQLDKQIKVLIDD [716]
      *:*** *** .*: .:  .    :* :   **   : .     .   *     ** *** **: *: ** .:: : *:  *  .:*: :.**:*: *:* . *.** .: :.  ::
Human IRRLFQQKLDELGVKALTYNDLIQAQKEISAHNQQLREQSEQLEQDNRALRGQSLQLLKARCEELQLDWATLSLEKLLKEKQALKSQISEK--------------QRHCLELQIS [651]
Fly   SVALLKVRMNELGIHVNSPNDLIAQAKEIVGRHKDLQHTASRMRNEVTFYEGEQKLLLNKQLKNLP-EYQKLCGTVNGKVKLEVPPELSETTAQELVLKEIANTLSQRKKLYAQVS [839]
         *:: :::***::. : ****   *** .::::*:. :.:::.::    .*:.  **: : ::*  :: .*.     * *  :  ::**.                **: *  *:*


Human APDYTRLSPAKIVLRRHLSQDHTVPGRPAASELHSRAEHTKENGLPYQSPSV--------PGSMKLSPQDPRPLSPGALQLAGEKSS-EKGLRER-------AYGS----SGELIT [863]
Fly   APSSQVNGKNKAI--------IEVPPPPATAPVSIK----SSPGHHYKDTTLMPAPKQQQQQQMTLS-QLPKCELPGLSTSRKQESPKVANFEDRLKSIITTALNEDQEQRSKAVE [1045]
      **.    .  * :          **  **:: :  :    .. *  *:. ::          .*.** * *:   **      ::*    .:.:*       * ..     .: :


Human SR--SLALAPAGFSYAG--SVAISGALAGSPASLTPGAEPATLDESSSSGSLFATVGSRSSTPQHPLLLAQPRNSLPASPAHQLSSSPRLGGAAQGPLPEASKGDLPSDSGFSDPE [1038]
Fly   HLNANTTISPITPPLPGPGAGATASTAPPPPANLPYGAYGGAVAKTTISGKYQA--------AKEPKYSPVRQAPLPPPPSHMASLYP-----------AGQQTTPADLGYQRRR [1187]
         . :::*      *  : * :.:    **.*  **  .:: ::: **.  *         :.*      :  **  *:*  *  *           *.:   *:* *:.  .
```

**Figure 1. *DOT1L* is loss of function constrained and conserved in flies**
(A) *DOT1L* is variant constrained based upon the control population database, gnomAD.[51] Variants in *DOT1L* are predicted to be inherited "either recessive or dominant" and have a score of 0.599 for "probability of being autosomal dominant" by DOMINO.
(B) Protein domain organization of human DOT1L and its fly (gpp) ortholog. Domains of DOT1L and the region they span specified with amino acid numbers are as follows: catalytic domain DOT1 (16–330),[9,10] Bat3/Ubiquitin interaction motif (361–380),[17] K-rich motif (390–407)[9] with a partially overlapping nuclear localization signal (NLS) (395–417),[52] three coiled-coil domains (529–549, 564–595, 616–636),[53] leucine zipper motif (576–594),[54] CTD binding patch (618–627),[25] three ENL/AF9 binding sites (628–653, 863–878, 877–900),[55–57] two more NLSs (1,088–1,111 and 1,164–1,171),[52] and STAT1 interaction region (580–1,138).[58] Variants suspected to be diagnostic are indicated above the protein as yellow dots and the variants suspected to be non-diagnostic are indicated as black dots. DOT1L's ortholog in fly, gpp, has a DOT1 domain, three coiled-coil domains, and two NLSs. Gpp has a homology score 12/16 (DIOPT, v.8.5).[59]
(C) Protein sequence alignment of human DOT1L (Uniprot: Q8TEK3) and fly gpp (UniProt: Q8INR6). The variants are highlighted with yellow or black boxes. Pink letters correspond to the DOT1 domain and green letters correspond to a coiled-coil domain. Symbols in the protein alignment: fully conserved (*), strongly similar (:), weakly similar (.), absent (–).

Based on gnomAD,[51] *DOT1L* is missense constrained with a *Z* score of 2.72 and a missense variant o/e ratio of 0.754 (0.709–0.802). Of the seven missense variants in our cohort (including one variant that was recurrent in three unrelated families), four variants map to the amino acid residues in the enzymatic domain (DOT1, residues 16–330) of DOT1L (Figure 1B).[9,10] The remaining three variants map to residues that are important for interaction of DOT1L with its interaction partners. The variant p.Leu626Val is found in a coiled-coil domain which overlaps with RNA Polymerase II C-Terminal Domain (RNAPII-CTD) binding patch (residues 618–627).[20,25,53]

The variants p.Arg853Cys and p.Lys1025Met are located in disordered regions of DOT1L. While seeming nonfunctional, these regions have been previously shown to be involved in DOT1L's interaction with Eleven Nineteen Leukemia (ENL), Cyclin T2, and STAT1.[53,55–57]

Variant allele frequencies and outputs from *in silico* analysis tools[60] are found in Table S8. All proband variants were extremely rare (n = 3) or absent (n = 4) in population-scale databases of genomic variation from approximately 430,000 total individuals,[51,61,62] consistent with the variants potentially causing a developmental disorder characterized by variable expressivity. *In silico* predictions were often conflicting, inconclusive, and/or suggestive of a benign impact (e.g., REVEL scores of <0.5 for six of the seven variants),[63] emphasizing the need for additional functional studies. We initially considered the possibility of haploinsufficiency. However, based on the control population data in Database of Genomic Variants (DGV),[64] there are forty control individuals that harbor deletions or copy number loss containing *DOT1L*. There are also two individuals with stop gain variants (c.4064C>A [p.Ser1355*]) in gnomAD[51] control population. Together these data indicate that unlike other KMTs, haploinsufficiency of *DOT1L* is unlikely to be associated with a disease. Therefore, we hypothesize that *DOT1L* variants harbored by the individuals may be gain-of-function mutations that correspond to hypermorphs that are more active than the wild-type protein, or they correspond to antimorphs that act in a dominant-negative fashion or to neomorphs that exhibit novel functions.

### DNA methylation profiling of proband blood samples

About half of the KMT-associated diseases have unique DNA methylation (DNAm) signatures (Table S1). Even though the molecular basis for these DNAm signatures remains to be determined, they can be used as one indicator to classify variants of uncertain significance (VUSs) as pathogenic or benign.[65] To determine whether *DOT1L* variants impact DNAm patterns, we assessed genome-wide DNAm in peripheral blood from seven individuals with *DOT1L* variants (p.Cys45Gly, p.Thr100Met, p.Glu123Lys, p.Glu129Lys, p.Leu626Val, p.Lys1025Met, and the suspected non-diagnostic variant p.Arg292Cys) as well as twelve sex- and age-matched controls. We identified 184 differentially methylated CpG sites that meet thresholds of false discovery rate (FDR) $< 0.3$ and $|\Delta\beta| > 0.10$ (10% difference in DNAm), using linear regression modeling. We visualized DNAm data at profile sites using principal component analysis (PCA) and hierarchal clustering. Two of the variants, p.Cys45Gly and p.Lys1025Met, were excluded from the discovery cohort as they are age outliers. As seen in Figures S1A and S1B, the significant CpGs could be used to segregate four individuals with suspected diagnostic variants (p.Thr100Met, p.Glu123Lys, p.Glu129Lys, and p.Leu626Val) from control subjects. The suspected non-diagnostic *DOT1L* variant, p.Arg292Cys, clustered with control individuals. Despite the small sample size, these results suggest that there is a distinct *DOT1L*-specific methylation signature.

### Loss of *gpp* in *Drosophila* causes developmental delay, loss of H3K79 methylation, and lethality

The ortholog of *DOT1L* in the fruit fly *Drosophila melanogaster* is *grappa* (Flybase: FBgn0264495). Based on the DRSC Integrative Ortholog Prediction Tool (DIOPT, v.8.5),[59] human DOT1L and fly Gpp are highly conserved with a DIOPT score of 12/16. The two proteins exhibit 24% identity and 35% similarity with 28% gaps. There is greater conservation in the DOT1 enzymatic domain (65% identity), to which most variants (5/9, including the suspected non-diagnostic variants) map. Three variants map to disordered regions and the other variant map to a coiled-coil domain (Figure 1B). Three of the nine variants affect fully conserved amino acid residues between DOT1L and Gpp. Four variants occur at highly similar amino acid residues and the last two variants occur at weakly similar amino acid residues between human and fly (Figure 1C).

Previous studies have shown that LoF mutants of *gpp* exhibit slow growth and larval lethality.[28] To further assess the phenotypes associated with loss of *gpp* in flies, we generated a mutant allele, $gpp^{TG4}$, by insertion of a *Splice acceptor-T2A-GAL4-polyA* cassette into a coding intron (i.e., an intron that is flanked by two coding exons) of *gpp* shared by all transcripts via CRISPR-Cas9. The $gpp^{TG4}$ allele is expected to behave as a severe LoF allele since the splice acceptor allows the artificial exon to be integrated into the mRNA and the polyA-tail leads to early termination of transcription.[42,43] Real-time qPCR data confirm that the *gpp* transcript levels are severely decreased (<17%) in $gpp^{TG4}$/*Df(3R)BSC738* when compared to wild-type larvae (Figure 2B). We also obtained a deletion allele, $gpp^{xxv}$, which was generated by X-ray mutagenesis and one of its breakpoints is in the coding region of *gpp*.[28] We also used three publicly available *UAS-RNAi* lines targeting *gpp* (Figure 2A). *RNAi #1* leads to a 50% ± 0.03% reduction and *RNAi #3* leads to a 65% ± 0.02% reduction of expression of *gpp* based on real-time qPCR data, while *RNAi #2* does not affect the levels of *gpp* serving as a negative control in our experiments (Figure 2B).

The two *gpp* mutant alleles over a molecularly defined deletion line [*DF(3R)BSC193*] covering *gpp* are recessive lethal. Notably, lethality is observed starting from first instar larval stage and animals do not survive beyond the third instar larval stage (Figure 2C). Indeed, a comparison of age-matched mutants, $gpp^{TG4}$/*Df* or $gpp^{TG4}/gpp^{xxv}$, with wild-type third-instar larvae shows severe developmental delay in mutant animals (Figure 2E), consistent with previous findings.[28] Mutant larvae also show a severe decrease in total H3K79 methylation levels compared to wild-type animals (Figure 2F). The lethality, larval developmental phenotype and the decrease in H3K79 methylation levels are rescued by a genomic rescue (GR) construct, $GR^{gpp}$, which carries an independent copy of the entire *gpp* locus [*P[acman]CH321-05N03*)].[66] Consistently,

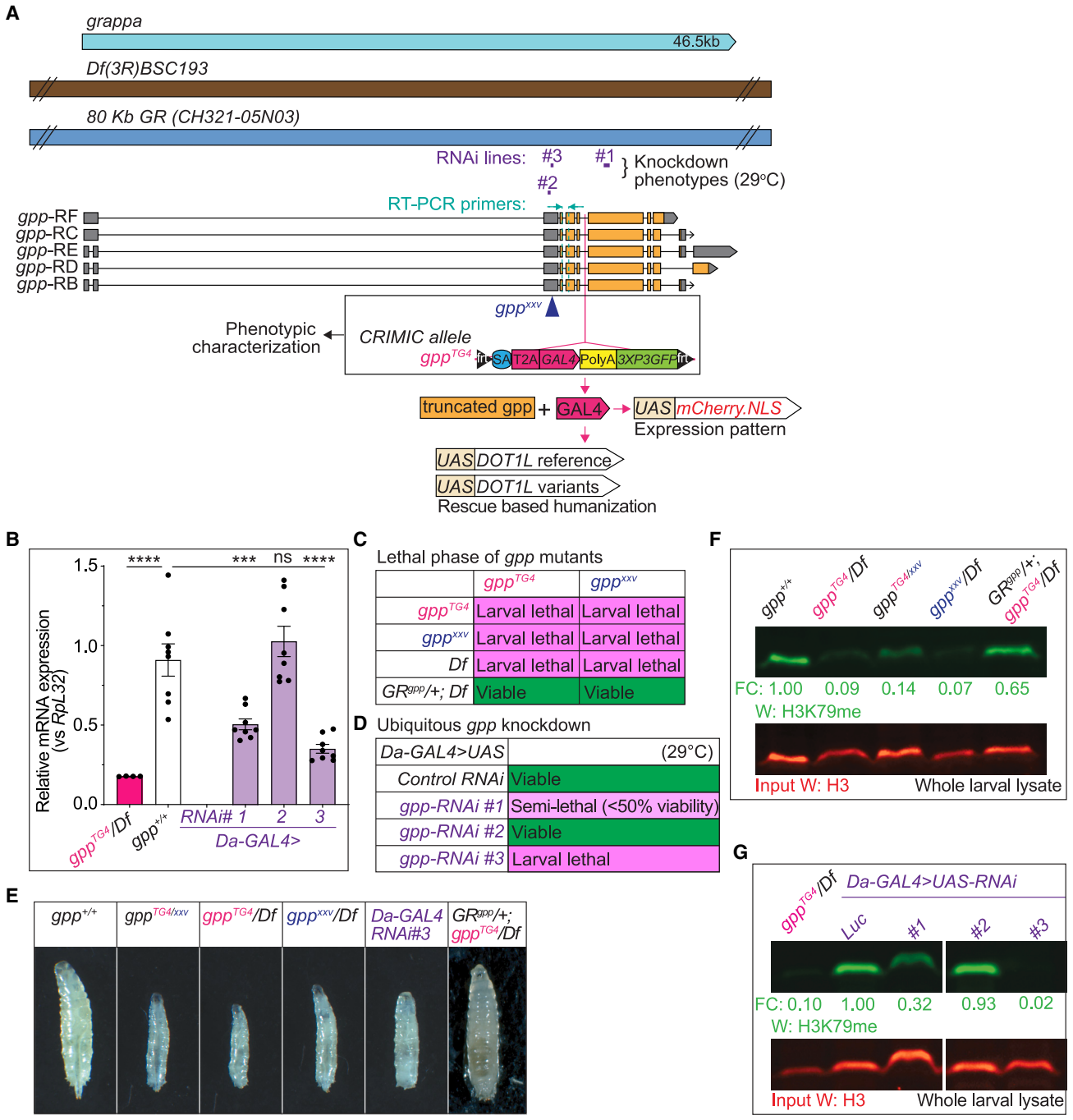



**Figure 2. Fly *grappa* is essential for development and loss of *gpp* leads to developmental delay and a reduction in H3K79 methylation**

(A) Gene structure of fly *gpp*, fly reagents, and assays used in this study. The gene region is shown in light blue. A deficiency and genomic rescue (GR) construct are shown in brown and dark blue, respectively. Three independent *RNAi* lines (#1, *P{TRiP.JF01284}attP2*; #2, *P{TRiP.GL01325}attP2*; and #3, *P{TRiP.HMJ02129}attP40*) tested in this study and their target regions are shown in purple. Different isoforms of *gpp* (RB, RC, RD, RE, and RF) are shown in the middle panel and exons are indicated in orange, introns as black lines, and untranslated regions (UTR) in gray. *gpp* mutants used in phenotypic characterization studies include $gpp^{xxv}$ (X-ray mutagenesis)[28] and $gpp^{TG4}$ (CRIMIC allele).

(B) Relative *gpp* mRNA expression levels are lower than 20% in $gpp^{TG4}/Df$ mutant larvae when compared to controls ($gpp^{+/+}$) based on real-time qPCR using primers shown in (A). Real-time qPCR analysis of *gpp* mRNA levels in ubiquitous *gpp* knockdown larvae shows that *Da-GAL4 > gpp-RNAi* reduces *gpp* expression to different levels. *Da-GAL4 > gpp-RNAi #1* decreases *gpp* mRNA levels by ~50% and *RNAi #3* by 70%. *RNAi #2* does not have any effect on *gpp* levels. Normalized *gpp* levels from three independent experiments were plotted as mean ± SEM, and statistical significance was determined by one-way ANOVA for multiple groups (***$p < 0.001$, ****$p < 0.0001$).

(C) Animals homozygous for *gpp* mutant alleles are larval lethal. None of the alleles or a large deficiency allele which lacks *gpp*, *Df(3R)BSC193*, can complement each other. One copy of a genomic rescue construct inserted in second chromosome (VK37), $GR^{gpp}/+$, rescues lethality of *trans*-heterozygous allelic combinations.

*(legend continued on next page)*

ubiquitous expression of *RNAi #1* using *da-GAL4* causes a decrease in viability (<50%) (Figure 2D), and ubiquitous expression of *RNAi #3* causes lethality at larval stages with developmental delay (Figures 2D and 2E). Further, ubiquitous knockdown of *gpp* with different *RNAi* lines also leads to reduction in H3K79 methylation levels (Figure 2G). Altogether, these data provide compelling evidence that *gpp* is essential for fly development and is responsible for H3K79 methylation in flies.

### *gpp* is broadly expressed in flies and primarily expressed in neurons in CNS

Next, we determined the expression pattern of *gpp* in flies. We crossed $gpp^{TG4}$ animals to *UAS-mCherry.NLS* (nuclear localized mCherry fluorescent protein) animals to label the nuclei of the cells that express *gpp*. We find that *gpp* is expressed broadly in most tissues in both larvae and adult flies (Figure 3A). In the third-instar larvae, *gpp* is expressed in all imaginal discs including wing, eye, leg, and haltere discs, which gives rise to respective adult structures (Figures 3B and 3C). To determine the cell types in the CNS that express *gpp*, we counterstained the larval and adult brains with the nuclear pan-neuronal marker Elav[67] or nuclear pan-glial marker Repo.[68] In both larval and adult brain, many but not all Elav-positive cells express *gpp* based on whole-mount z stacked confocal microscopy images (Figure 3D, left). In agreement, single slice images also show that *gpp* is expressed in most Elav-positive cells, indicating that *gpp* is expressed in a large subset of neurons (Figure 3D, right). Whole-mount z stacked images and 3-slice stack images of Repo staining in both larval and adult brain show a small group of Repo-positive cells that express *gpp* (Figure 3E). This suggests that *gpp* is expressed in a small subset of glial cells in addition to its broad neuronal expression.

To assess whether knockdown of *gpp* in the nervous system can cause an obvious defect, we performed partial knockdown of *gpp* in neurons using a neuronal (*Elav-GAL4*) or a glial (*Repo-GAL4*) driver. We found that neuronal knockdown of *gpp* using *RNAi #3* causes lethality while glial knockdown does not cause any observable phenotype. Moreover, knockdown of *gpp* in the developing eye (*ey-GAL4*) causes a rough eye phenotype (Figure S2A), and knockdown in developing tissues (*en-GAL4*) leads to lethality and morphological phenotypes of the wing (Figure S2B). These data suggest that *gpp* have important roles in nervous system function as well as in imaginal disc development.

### Functional assays in flies indicate that *DOT1L* variants are gain-of-function alleles

We next attempted to rescue the lethality observed in $gpp^{TG4}$ flies by expressing human *DOT1L* in flies. We took advantage of *GAL4* expression in $gpp^{TG4}$ flies. In these flies, even though the transcription is terminated early, the produced transcript is translated into a short protein and the presence of the viral T2A sequence allows ribosomal skipping and re-initiation of translation to produce a functional GAL4. Therefore, GAL4 is produced under the control of *gpp*'s regulatory elements in the same spatial and temporal pattern as *gpp*.[42,43] To use "UASxGAL4" system, we also created transgenic lines expressing C-terminal 3XHA-tagged human reference or variant *DOT1L* under the control of the UAS promoter (Figure 1A). In addition to modeling the variants seen in our cohort, we also modeled the p.Arg292Cys variant that was suspected to be nonpathogenic as a negative control. Unfortunately, expression of reference or variant *DOT1L* in $gpp^{TG4}/Df$ flies failed to rescue the lethality at all temperatures (18°C, 25°C, 29°C) tested (Figure S3).

Intriguingly, heterozygous mutant flies, $gpp^{TG4/+}$, expressing reference or variant *DOT1L* cDNA or fly *gpp* ($gpp^{GS13895}$, a fly line containing UAS regulatory sequence at 5′ UTR of fly *gpp*) have reduced viability compared to control flies ($gpp^{TG4/+}$>*UAS-Empty*) as shown by lower-than-expected Mendelian ratios of eclosion (percent o/e ratios ≤ 60%), indicating that overexpression of human *DOT1L* in the pattern of the fly *gpp* causes toxicity. Moreover, the variants p.Glu123Lys and p.Glu129Lys shows a severe reduction, and the variant p.Leu626Val shows a moderate reduction in viability compared to flies expressing reference *DOT1L* (Figure 4A). A real-time qPCR analysis for mRNA levels of *DOT1L* in $gpp^{TG4/+}$ flies expressing variant *DOT1L* cDNAs show no significant difference compared to flies expressing reference *DOT1L* (Figure 4B). Expression of reference or variant human cDNAs also cause morphological wing defects with high penetrance (>95%)

(D) Ubiquitous knockdown of *gpp* results in larval lethality. While *Da-GAL4 > RNAi #3* flies are larval lethal, *-RNAi #1* flies are semi-lethal, as shown by lower-than-expected genotypic ratios of survival into adulthood. *RNAi #2* flies are completely viable. *Da-GAL4 > UAS-gpp-RNAi* flies were compared to *Da-GAL4 > control-RNAi* (*control-RNAi = UAS-luciferase-RNAi*). All the crosses were performed at 29°C.
(E) Complete loss or ubiquitous knockdown (<50%) of *gpp* leads to severe developmental delay. Images of age-matched larvae at third instar larval stage is shown. One copy of a genomic rescue construct inserted in second chromosome (VK37), $GR^{gpp}/+$, can rescue the developmental delay phenotype.
(F) *Gpp* mutant animals have a drastic decrease in H3K79 methylation levels. Protein lysate from 10 larvae was prepared for each sample. H3K79 methylation levels were normalized with loading control, H3, and fold change (FC) for each sample were calculated by comparing normalized H3K79 methylation levels to wild-type larvae ($gpp^{+/+}$). One copy of the genomic rescue construct, $GR^{gpp}/+$, can increase the H3K79 methylation levels to ∼65%.
(G) Ubiquitous knockdown (<50%) of *gpp* causes a severe decrease in H3K79 methylation levels. *Da-GAL4 > gpp-RNAi #1* decreases H3K79 methylation levels to ∼30% and *RNAi #3* to ∼2%. *RNAi #2* does not have any effect on methylation levels. Protein lysate from 10 larvae was prepared for each sample. H3K79 methylation levels were normalized with loading control, H3, and FC for each sample were calculated by comparing normalized H3K79 methylation levels to *control-RNAi* (*control-RNAi = UAS-luciferase-RNAi*). All the crosses were performed at 29°C. A sample which is excluded from this study is cropped out from the gel image.

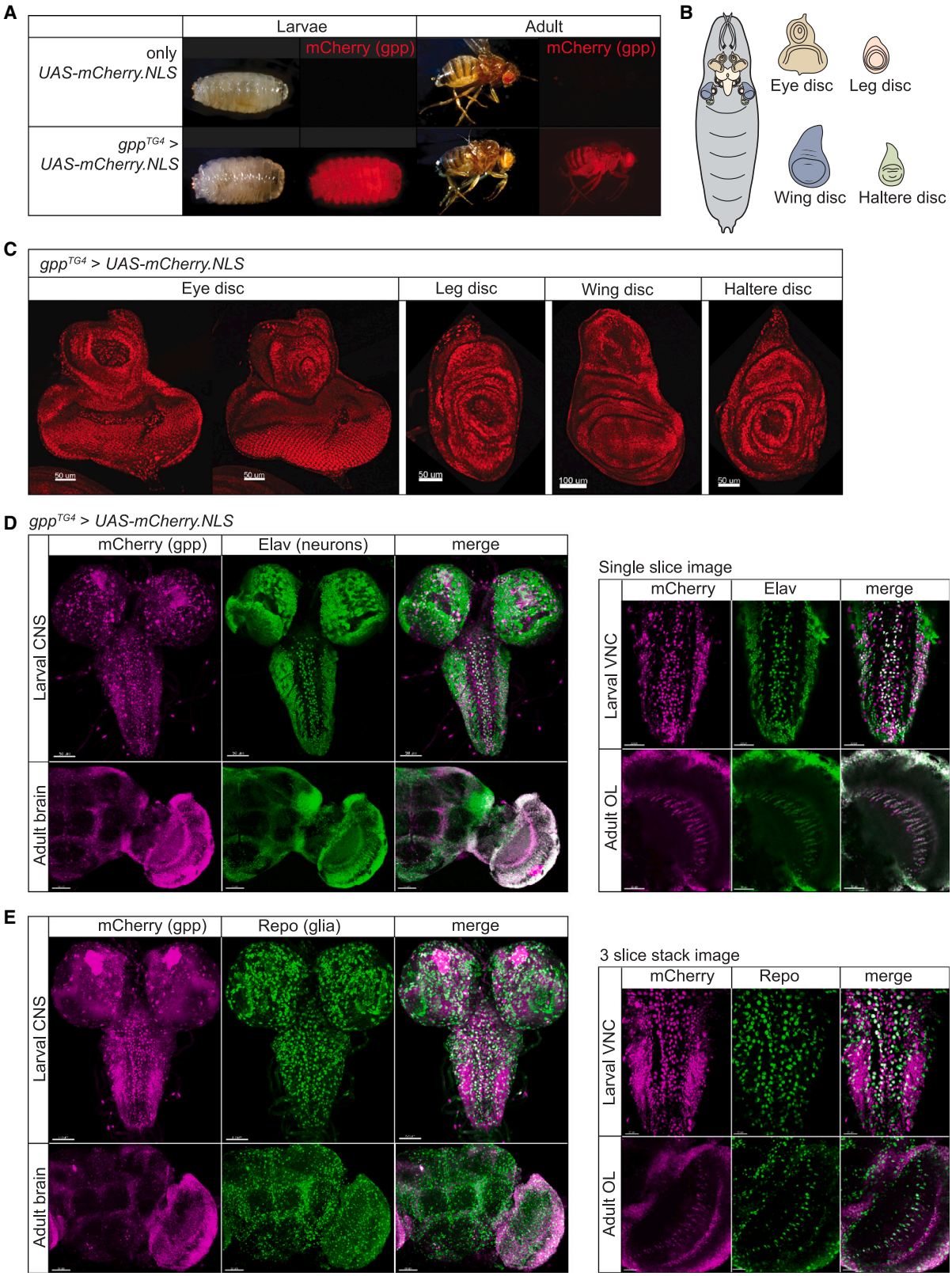



**Figure 3. Fly *grappa* is broadly expressed in both larvae and adult flies**

The *gpp*$^{TG4}$ allele was used to drive expression of *UAS-fluorescent reporter transgene* (*UAS-mCherry.NLS*).

(A) *gpp* is expressed broadly and highly throughout the body in both larvae and adult flies. Nuclear mCherry was present and whole larvae or adult was imaged.

(B) Schematic of different tissues of larvae.

*(legend continued on next page)*

(Figure S4). We classified the wing phenotypes of these flies into three different groups based on their severity. Mild phenotypes include loss of cross-veins and extra vein branching; moderate phenotype includes blistering in addition to mild phenotypes; and strong phenotype includes necrosis in addition to moderate phenotypes (Figure 4C). The variants p.Glu123Lys and p.Glu129Lys cause a more severe wing phenotype as none of the animals have wild-type wings and strong wing phenotypes in these flies are associated with a significantly higher penetrance compared to reference (Figures S4B and 4D). In addition, the variants p.Thr100Met and p.Leu626Val cause significantly more strong phenotypes compared to reference (Figure 4D). Furthermore, expression of human reference or variant *DOT1L* in developing tissues (*en-GAL4*) also causes morphological wing defects with full penetrance, and variants p.Glu123Lys and p.Glu129Lys again lead to more severe phenotypes (Figure S5).

To assess the effect of the variants on enzymatic function of DOT1L, we next measured H3K79 methylation levels in heterozygous mutant flies, $gpp^{TG4/+}$, expressing reference or variant *DOT1L* cDNA. We prepared total protein lysate from 5- to 7-day-old flies to extract histones and histone-bound DOT1L. Previously it was shown that the overall level of methylation at H3K79 residues rather than the level of one specific methylation state, i.e., mono-, di-, or trimethylation, determines the downstream effect of this particular modification.[11] Therefore, we detected the methylation status of our samples using an antibody that recognizes all methylation states at H3K79.[69] We found the flies expressing human reference *DOT1L* have a 4.41-fold increase in H3K79 methylation compared to control flies (*UAS-empty*) suggesting that human DOT1L methylates H3K79 in *Drosophila* (Figure $4E_i$). Flies expressing all the *DOT1L* variants, except p.Arg292Cys, show significantly higher H3K79 methylation levels when compared to the reference *DOT1L*-expressing flies (Figures $4E_i$ and $4F_i$). The variants p.Glu123Lys and p.Glu129Lys cause the highest increase in H3K79 methylation levels possibly due to the higher DOT1L levels in these flies (Figures $4E_{ii}$ and $4F_{ii}$). DOT1L levels in the other variants do not show a statistically significant increase based on one-way ANOVA multiple group comparison analysis. When we ubiquitously expressed reference or variant *DOT1L* using the *Actin-GAL4* driver, we found that the flies expressing human reference *DOT1L* have a 1.7-fold increase in H3K79 methylation compared to control flies (*UAS-empty*). We also observed similarly higher H3K79 methylation levels with variants p.Thr100Met, p.Glu123Lys, and p.Glu129Lys when compared to the reference cDNA. DOT1L levels in the variant-expressing flies are not significantly different from the reference-expressing flies (Figures S6A and S6B). Furthermore, there is also no significant change in the viability of flies expressing reference or variant *DOT1L* ubiquitously when compared to control flies (*UAS-Empty*) (Figure S6C). To understand the reason underlying the discrepancy in methylation levels and viability in flies expressing *DOT1L* cDNAs ubiquitously and cDNAs expressed in the endogenous pattern of *gpp*, we compared the levels of DOT1L reference in these flies. As shown in Figure S6D, the DOT1L level is ~2-fold higher in $gpp^{TG4/+}$ flies when compared to ubiquitously expressing flies (Figure S6D). Consistent with these observations, ubiquitous expression of fly *gpp* is less toxic than its expression in the endogenous pattern (Figure S6E).

In summary, our data suggest that all variants tested in flies except p.Arg292Cys (the suspected non-diagnostic variant) alter DOT1L activity or levels and cause higher H3K79 methylation. Moreover, the severity of the phenotypes expressing variant DOT1L correlate with the level of H3K79 methylation.

### *DOT1L* p.Glu123Lys is a gain-of-function variant in vertebrate cells

*DOT1L* is a conserved H3K79 methyltransferase in both invertebrates and vertebrates. Previous mouse studies indicate that germline *Dot1L* deletion causes embryonic lethality and defects in heart and yolk sac.[26] Conditional knockout of *Dot1l* during embryonic development impairs cortical development and decreases neural progenitor cells.[70] To assess the effect of *Dot1l* loss at the cellular level in nervous system, we performed *RNAi* assays in mouse primary neuronal cells. Knockdown of *Dot1l* in primary neuronal cells causes a significant reduction in axon length (Figures 5A and 5B). It also decreases the length of all neurites significantly (Figure 5C).

To determine how *DOT1L* variants affect histone methylation, we transfected HEK293T cells with cDNA from a reference and two variants. The p.Glu123Lys variant, which causes the most severe fly phenotypes, leads to an almost 4-fold increase in H3K79 methylation level compared to reference DOT1L, similar to what was observed in flies. The p.Arg853Cys variant also increases H3K79 methylation level but not to a statistically significant level (Figures 5D and 5E).

Together, these data confirm the conserved roles of *DOT1L* in neuronal development and show that the DOT1L variants are gain-of-function alleles.

(C) *gpp* is expressed in most tissues in larvae. Nuclear mCherry was present and tissue were dissected from L3 larvae including eye, leg, wing, and haltere discs.
(D and E) *gpp* is broadly expressed in the nervous system. Nuclear mCherry was driven by $gpp^{TG4/+}$ and tissues were dissected from L3 larvae (CNS, includes central brain and VNC) or adults (brain). Shown is half of the adult brain. Tissue was counterstained with markers for neurons (Elav) (D) or glia (Repo) (E). Z stacked images of *gpp* expression pattern compared to neurons or glia are shown in the left panels in (D) and (E), respectively. Single slice images (right panels) were used to better visualize cellular co-localization of Nuclear mCherry signal with neurons (D) or glia (E). Note that *gpp* is expressed in a large subset of neurons and in a small subset of glia. VNC, ventral nerve cord; OL, optic lobe.

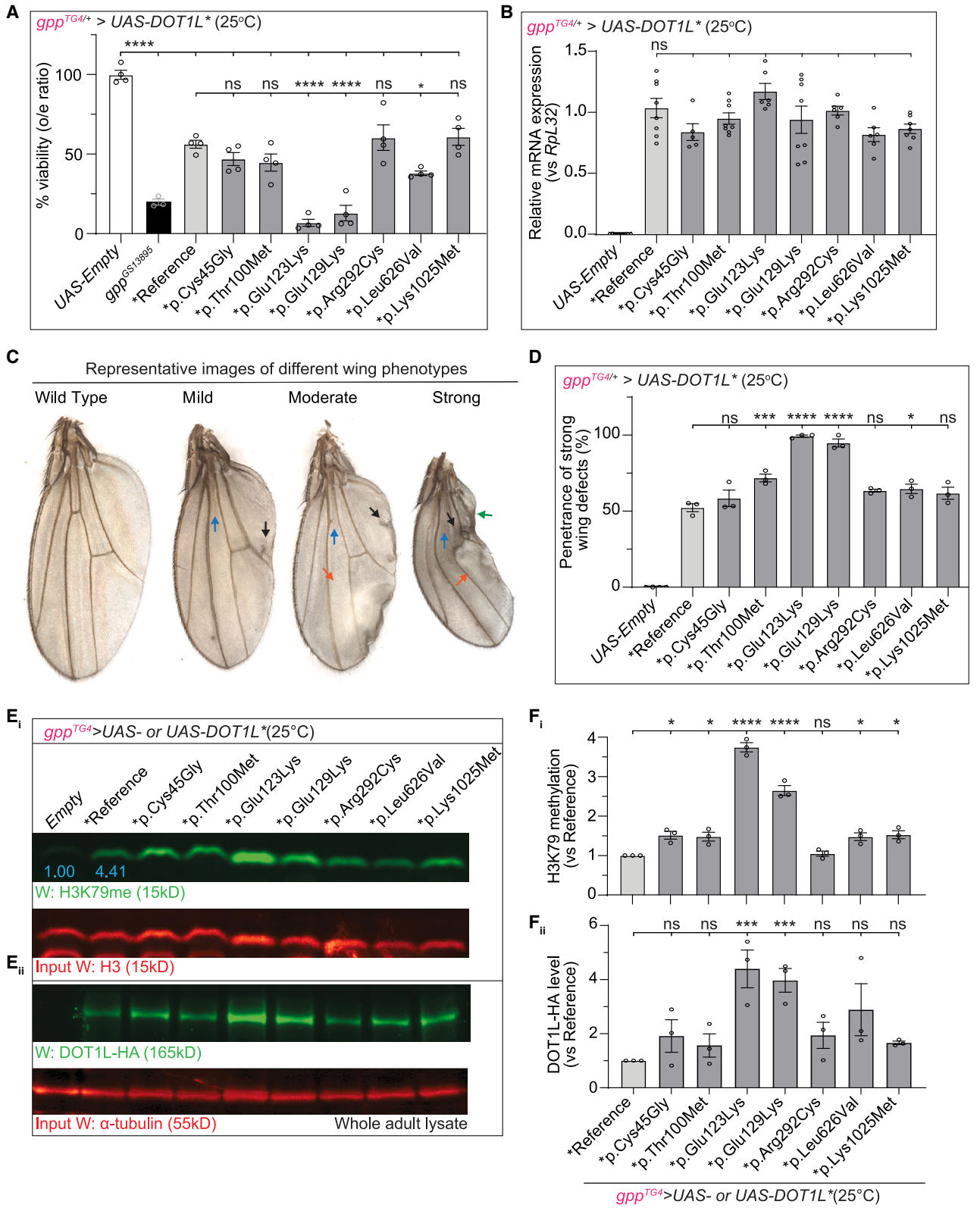



**Figure 4. Expression human *DOT1L* disrupts wing morphology and different variants display different amounts of toxicity and H3K79 methylation**

(A) Heterozygous mutant flies, $gpp^{TG4/+}$, expressing reference or variant *DOT1L* cDNA have reduced viability compared to control flies ($gpp^{TG4/+}$ > *UAS-Empty*) as shown by lower-than-expected genotypic ratios of survival into adulthood. The variants p.Glu123Lys and p.Glu129Lys show significant reduction in viability compared to flies expressing reference *DOT1L*. All the crosses were performed at 25°C. Percent viabilities (o/e ratios) from three independent experiments were plotted as mean ± SEM, and statistical significance was determined by one-way ANOVA for multiple groups (****$p < 0.0001$).

*(legend continued on next page)*

## Discussion

We identified nine individuals with rare heterozygous missense variants in *DOT1L*, encoding a KMT that has not been associated with a Mendelian genetic disorder in humans in OMIM previously. The core clinical features in this cohort includes a range of congenital anomalies and CNS dysfunction (global developmental delay, feeding difficulties, hypotonia, and/or intellectual disability), similar to other KMT disorders (Table S1). We confirm that the fly ortholog *gpp* is an essential gene and that it is responsible for H3K79 methylation in flies. We also show that it is required for proper development of eyes, wings, and neurons. Expression of the human reference *DOT1L* increases H3K79 methylation in flies, showing that the human protein can recognize and act on H3K79 in flies. Interestingly, expression of some *DOT1L* variants decrease the viability of flies when compared to animals expressing the reference *DOT1L*. All variants, except p.Arg853Cys, increases H3K79 methylation levels compared to reference, consistent with a gain-of-function mechanism. Moreover, the variants associated with the most severe phenotypes in flies also show evidence of a distinct blood-derived DNAm profile in humans. In summary, the clinical, molecular, and biochemical data presented here indicate that select heterozygous missense variants in human *DOT1L* cause an autosomal-dominant KMT disorder through a gain-of-function mechanism.

The number of Mendelian disorders resulting from dysregulation of epigenetic machinery, including KMTs, have greatly expanded in the past decade.[71,72] So far, around 50% of all genes encoding KMTs have been associated with a Mendelian disorder and all these disorders have an autosomal-dominant inheritance pattern. Heterozygous deletions or protein-truncating variants underlie most of these KMT-associated disorders indicating haploinsufficiency as the predominant mechanism (Table S1). Even though missense variants in several genes encoding KMTs have also been reported for individuals with other symptoms including autism, ID, bipolar disorder, and congenital anomalies,[5] their causal relationship has not been extensively investigated through functional studies. Here, our functional analysis in flies and human cells show that missense variants in a KMT, *DOT1L*, correspond to gain-of-function alleles.

DOT1L is the only KMT known to lack SET domain but contains a DOT1 enzymatic domain.[7,8] Besides its structure, it is also mechanistically different from SET domain containing KMTs, which has direct consequences for its regulation and function. Compared to the processive mechanism of all SET domain KMTs, DOT1L acts in a distributive manner such that it can only sequentially add methyl groups to each H3K79 residue. For each methylation cycle, it needs to dissociate from and reassociate with the same lysine residue to achieve the higher methylation states (i.e., di- or tri-).[10,11] In contrast to the other methylated lysines, H3K79 methylation states (i.e., mono-, di-, or tri-) are similarly distributed over a gene and have redundant roles. Indeed, the overall level of methylation at H3K79, rather than the level of one state, determines the outcome of this histone modification.[11] In addition to our knowledge, so far no protein has been identified that can specifically bind to methylated H3K79 *in vivo*, and there is no known demethylase that can remove H3K79 methylation. Hence, the cells may not be able to easily compensate for the downstream effects of excess methylation due to elevated DOT1L enzymatic activity or elevated levels. Consistent with this hypothesis, we observed H3K79 hypermethylation when *DOT1L* is expressed in flies that endogenously express *gpp*. These flies also exhibit reduced viability and wing deformities showing a dominant effect due to H3K79 hypermethylation. Finally, the observed wing phenotypes are consistent with the previous observation of DOT1L-dependent regulation of Wnt target gene expression.[29–31]

(B) Relative *DOT1L* mRNA expression levels in $gpp^{TG4/+}$ flies expressing reference or variant *DOT1L*. Normalized *DOT1L* levels from three independent experiments were plotted as mean ± SEM, and statistical significance was determined by one-way ANOVA for multiple groups.
(C) Representative images of different wing phenotypes observed in in survivors of heterozygous mutant flies, $gpp^{TG4/+}$, expressing each *DOT1L* cDNA. Blue arrow, loss of cross-vein; black arrow, extra vein branching; orange arrow, blistered areas; green arrow, necrotic areas.
(D) Penetrance of strong wing phenotypes in survivors of heterozygous mutant flies, $gpp^{TG4/+}$, expressing each *DOT1L* cDNA. Percent penetrance for strong phenotype from three independent experiments were plotted as mean ± SEM, and statistical significance was determined by one-way ANOVA for multiple groups (*$p < 0.05$, ***$p < 0.001$, ****$p < 0.0001$).
(E) All the *DOT1L* variants, except p.Arg292Cys, behave as gain-of-function (GoF) mutations. Survivors of heterozygous mutant flies, $gpp^{TG4/+}$, expressing reference or variant *DOT1L* cDNA show increased H3K79 methylation compared to control flies ($gpp^{TG4/+}$ > *UAS-Empty*). Flies expressing variant *DOT1L* cDNAs, corresponding to p.Cys45Gly, p.Thr100Met p.Glu123Lys, p.Glu129Lys, p.Leu626Val, and p.Lys1025Met, show higher H3K79 methylation levels when compared to reference ($E_i$). Flies expressing variant *DOT1L* cDNAs p.Glu123Lys and p.Glu129Lys show higher DOT1L levels when compared to reference ($E_{ii}$). Protein lysate from 10 adult flies were prepared for each sample. H3K79 methylation levels were normalized with loading control, H3, and fold change for each sample were calculated by comparing normalized H3K79 methylation levels to reference *DOT1L*-expressing flies. DOT1L levels were normalized with loading control, α-tubulin, and fold change for each sample were calculated by comparing normalized DOT1L levels to reference *DOT1L*-expressing flies. All the crosses were performed at 25°C. Blue numbers indicate the fold change of H3K79 methylation level in reference when compared to control (*UAS-Empty*).
($F_i$) Normalized H3K79 methylation band intensities for each group from three independent experiments were plotted as mean ± SEM, and statistical significance was determined by one-way ANOVA for multiple groups (*$p < 0.05$, ****$p < 0.0001$).
($F_{ii}$) Normalized DOT1L band intensities for each group from three independent experiments were plotted as mean ± SEM, and statistical significance was determined by one-way ANOVA for multiple groups (***$p < 0.001$).

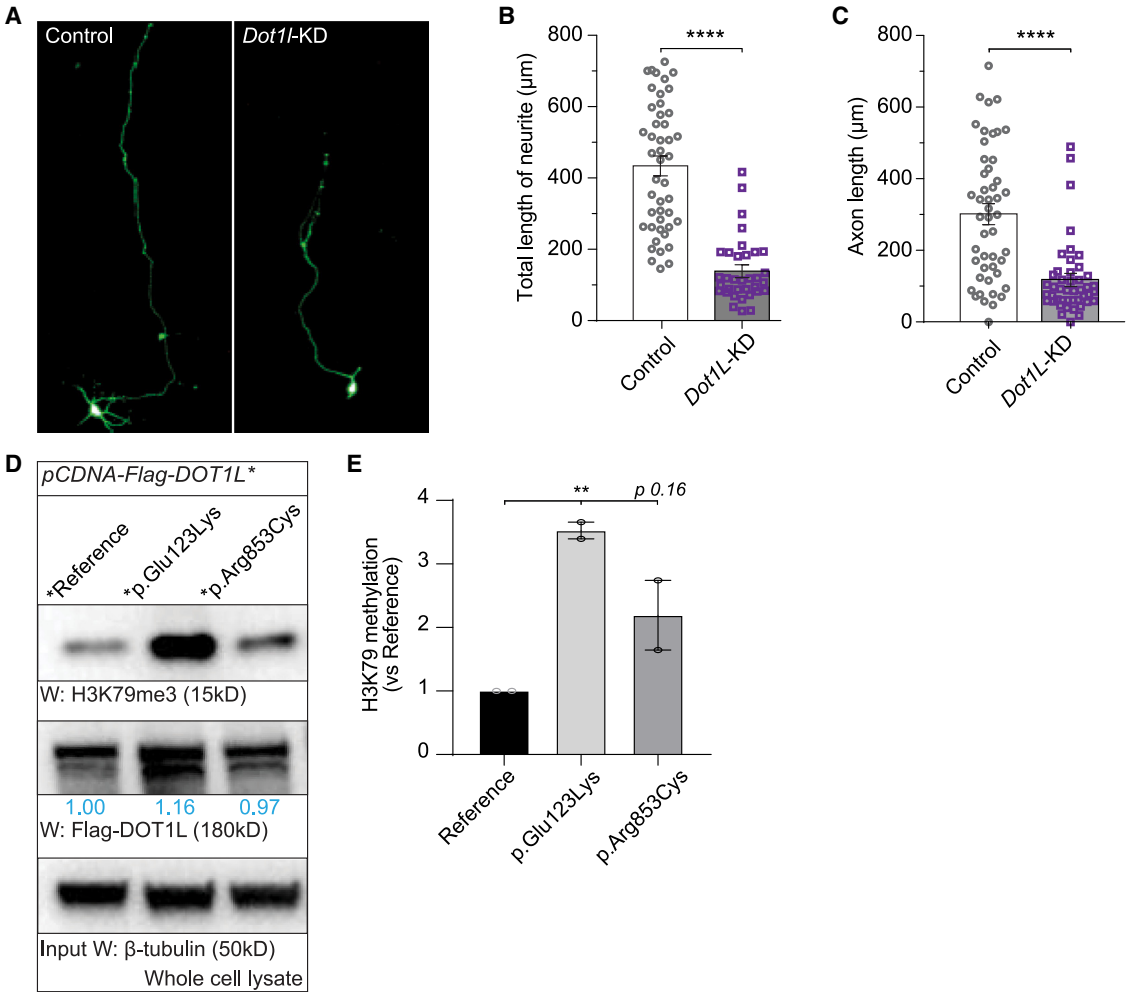



**Figure 5. Knockdown of *Dot1l* decreases axonal length and human *DOT1L* variants increase H3K79 methylation levels**

(A) Representative images of different primary neuronal cells. Left panel shows mouse primary neurons transfected with control plasmid and right panel shows mouse primary neurons transfected with *Dot1l*-shRNA plasmid.

(B and C) Total neurite length (B) and total axonal length (C) was measured using ImageJ (n = 50 per condition). Results were plotted as mean ± SEM, and statistical significance was determined by unpaired t test (****$p < 0.0001$).

(D) *DOT1L* variants behave as gain-of-function (GoF) mutations. HEK293T cells were transfected with N-terminal flag tagged reference or variant (p.Glu123Lys and p.Arg853Cys) *DOT1L* cDNAs. Whole cell lysate was prepared for each sample. H3K79 methylation levels were normalized with loading control, β-tubulin, and fold change for each sample were calculated by comparing normalized H3K79 methylation levels to reference *DOT1L*-expressing cells. Blue numbers indicate the fold change of DOT1L levels in cells expressing variant compared to cells expressing reference *DOT1L*.

(E) Normalized H3K79 methylation band intensities for each group from two independent experiments were plotted as mean ± SEM, and statistical significance was determined by one-way ANOVA for multiple groups (**$p < 0.01$).

Since *DOT1L* reference and variant transgenes are inserted in the same genomic site in *Drosophila* and heterozygous mutant flies express similar levels of reference or variant *DOT1L* mRNAs, we can perform comparative functional analyses of the variants with the reference cDNA. We find that the severity of the phenotypes of the flies expressing *DOT1L* variants correlates with the levels of DOT1L and H3K79 methylation. The p.Glu123Lys and p.Glu129Lys variants show the highest H3K79 methylation as well as DOT1L levels. This causes a very severe reduction in viability and a more severe wing phenotype as no animals have wild-type wings. The p.Glu123Lys variant causes a drastic increase in H3K79 methylation in human cells as well. These two variants are located in the predicted S-adenosylmethionine (SAM, methyl group donor) binding pocket in the DOT1 domain.[73] They might either increase the DOT1L stability or levels or alter SAM binding and lead to enhanced DOT1L activity observed in these flies. The variant p.Leu626Val causes a moderately higher reduction in viability and higher penetrance of strong wing phenotypes compared to reference despite a relatively mild increase in H3K79 methylation. This variant is in a coiled-coil domain that is important for direct interaction of DOT1L with elongating RNAPII on actively transcribed genes.[25] Therefore, it might have a direct effect on expression of the target genes leading to a milder but a different phenotype than the reference. Although the other variants (p.Cys45Gly, p.Thr100Met,

and p.Lys1025Met) have similar viability and they (except p.Thr100Met) cause similar wing phenotypes as the reference, they show a mild increase in H3K79 methylation. Even though p.Cys45Gly and p.Thr100Met map to the DOT1 domain, which includes regions that are important not only for biochemical activity but also for SAM and histone binding of DOT1L, it is unknown what specific role they may play in DOT1L activity. Any alterations in these bindings, in addition to the modest increase in DOT1L levels in flies expressing these two variants, may lead to the observed increase in H3K79 methylation. The variant p.Lys1025Met is located in a disordered region which is known to be involved in DOT1L's interaction with its partners.[55–57] Alterations in the interaction of DOT1L with its partners could lead to higher H3K79 methylation seen in p.Lys1025Met flies. The nondiagnostic variant p.Arg292Cys shows a similar phenotypic profile as p.Cys45Gly and p.Lys1025Met except for H3K79 methylation levels. Further biochemical and structural studies will be needed to understand the specific impact of these variants on DOT1L activity and the interaction with its substrate or interaction partners. Taken together, our data suggest that the *DOT1L* variants reported here result in a gain of function, are associated with increased activity of DOT1L, and are associated with developmental phenotypes. Understanding the precise molecular mechanisms as to how these variants lead to increased activity *in vivo* may allow for the use of specific DOT1L inhibitors as a potential therapeutic strategy.

Our study emphasizes the value of functional studies and non-human model systems for rare variant interpretation. Limitations of *in silico* prediction tools for missense variants are well established,[74] and these limitations are exacerbated in situations where the mechanism is not LoF.[75] Classification of novel missense variation in *DOT1L* in the clinical laboratory setting may remain challenging in the short term. Although the number of useable human DNA samples was too low in this study to confidently assign and validate a DNAm signature, this remains a promising area of research that would support clinical diagnostics. We also emphasize that it remains possible that other variants in *DOT1L* are associated with distinct Mendelian disorders through pathogenic mechanisms beyond gain of function. Furthermore, it should be highlighted that there is some degree of variability with respect to clinical features across the cohort. For example, five individuals have microcephaly, including one with brain atrophy shown by MRI, whereas two others (with p.Cys45Gly and p.Lys1025Met) had large head circumferences, including one with megalencephaly. Further study will be required in order to better understand the specific impact of each missense variant on DOT1L function and the specific phenotypes associated with each.[76]

In summary, we find that *de novo* pathogenic *DOT1L* variants are associated with a Mendelian disorder in humans. We provide functional analysis in flies that supports a gain-of-function model for *DOT1L* variants. Further studies of the underlying mechanism will be necessary to provide a better understanding of the pathological mechanisms and may provide therapeutic strategies.

## Data and code availability

There are restrictions to the availability of the genome-wide (exome and genome) sequencing data because the sequencing was performed as a clinical diagnostic test and/or because of country-specific factors. Fly reagents generated in this study will be deposited to Bloomington Stock Center and Drosophila Genomics Research Center. Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contacts Gregory Costain (gregory.costain@sickkids.ca) and Hugo J. Bellen (hbellen@bcm.edu).

## Supplemental information

Supplemental information can be found online at https://doi.org/10.1016/j.ajhg.2023.09.009.


## Acknowledgments

We thank all participating individuals and their family members for supporting this study. We thank Hongling Pan for injections to create transgenic flies and Burak Tepe and Lindsey Goodman for suggestions. We thank the Bloomington Drosophila Stock Center (BDSC) for fly stocks and the Developmental Studies Hybridoma Bank for antibodies. This work was supported by a grant from the NIH Commonfund to the Model Organisms Screening Center of the UDN through U54 NS093793 (NINDS), the Office of Research Infrastructure Programs of the NIH (awards R24 OD022005 and R24 OD031447), the Huffington Foundation, the Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital to H.J.B., and the Baylor College of Medicine IDDRC P50HD103555 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development for use of the Microscopy Core facilities. S. Banka acknowledges the support by the NIHR Manchester Biomedical Research Centre (NIHR203308). This work was supported by a Canadian Institutes of Health Research (CIHR) grant to R.W. (PJT-178315).


## Declaration of interests

The authors declare no competing interests.


Received: May 25, 2023
Accepted: September 18, 2023
Published: October 11, 2023


## Web resources

Alamut, https://www.sophiagenetics.com/platform/alamut-visual-plus/
AxioVision, https://www.micro-shop.zeiss.com/en/us/system/software-axiovision+software-products/1007/
CADD, https://cadd.gs.washington.edu/
deCAF, https://decaf.decode.com/
DIOPT, https://www.flyrnai.org/cgi-bin/DRSC_orthologs.pl

Ensembl Variant Effect Predictor, http://useast.ensembl.org/Tools/VEP
Franklin, https://franklin.genoox.com/clinical-db/home
GeneDx ClinVar submission page, https://www.ncbi.nlm.nih.gov/clinvar/submitters/26957/
GeneMatcher, https://genematcher.org/statistics/
gnomAD, https://gnomad.broadinstitute.org/
Imaris, https://imaris.oxinst.com/
MARRVEL, http://www.marrvel.org/
OMIM, https://www.omim.org/
TOPMed, https://bravo.sph.umich.edu/freeze8/hg38/
Uniprot, https://www.uniprot.org/

## References


1. Kouzarides, T. (2007). Chromatin Modifications and Their Function. Cell *128*, 693–705. https://doi.org/10.1016/j.cell.2007.02.005.
2. Greer, E.L., and Shi, Y. (2012). Histone methylation: A dynamic mark in health, disease and inheritance. Nat. Rev. Genet. *13*, 343–357. https://doi.org/10.1038/nrg3173.
3. Husmann, D., and Gozani, O. (2019). Histone lysine methyltransferases in biology and disease. Nat. Struct. Mol. Biol. *26*, 880–889. https://doi.org/10.1038/s41594-019-0298-7.
4. Allis, C.D., Berger, S.L., Cote, J., Dent, S., Jenuwien, T., Kouzarides, T., Pillus, L., Reinberg, D., Shi, Y., Shiekhattar, R., et al. (2007). New Nomenclature for Chromatin-Modifying Enzymes. Cell *131*, 633–636. https://doi.org/10.1016/j.cell.2007.10.039.
5. Faundes, V., Newman, W.G., Bernardini, L., Canham, N., Clayton-Smith, J., Dallapiccola, B., Davies, S.J., Demos, M.K., Goldman, A., Gill, H., et al. (2018). Histone Lysine Methylases and Demethylases in the Landscape of Human Developmental Disorders. Am. J. Hum. Genet. *102*, 175–187. https://doi.org/10.1016/j.ajhg.2017.11.013.
6. Wood, K., Tellier, M., and Murphy, S. (2018). DOT1L and H3K79 methylation in transcription and genomic stability. Biomolecules *8*, 11. https://doi.org/10.3390/biom8010011.
7. Van Leeuwen, F., Gafken, P.R., and Gottschling, D.E. (2002). Dot1p modulates silencing in yeast by methylation of the nucleosome core. Cell *109*, 745–756. https://doi.org/10.1016/S0092-8674(02)00759-6.
8. Singer, M.S., Kahana, A., Wolf, A.J., Meisinger, L.L., Peterson, S.E., Goggin, C., Mahowald, M., and Gottschling, D.E. (1998). Identification of high-copy disruptors of telomeric silencing in Saccharomyces cerevisiae. Genetics *150*, 613–632. https://doi.org/10.1093/genetics/150.2.613.
9. Min, J., Feng, Q., Li, Z., Zhang, Y., and Xu, R.M. (2003). Structure of the catalytic domain of human Dot1L, a non-SET domain nucleosomal histone methyltransferase. Cell *112*, 711–723. https://doi.org/10.1016/S0092-8674(03)00114-4.
10. Vlaming, H., and van Leeuwen, F. (2016). The upstreams and downstreams of H3K79 methylation by DOT1L. Chromosoma *125*, 593–605. https://doi.org/10.1007/s00412-015-0570-5.
11. Frederiks, F., Tzouros, M., Oudgenoeg, G., Van Welsem, T., Fornerod, M., Krijgsveld, J., and Van Leeuwen, F. (2008). Nonprocessive methylation by Dot1 leads to functional redundancy of histone H3K79 methylation states. Nat. Struct. Mol. Biol. *15*, 550–557. https://doi.org/10.1038/nsmb.1432.
12. Ng, H.H., Xu, R.M., Zhang, Y., and Struhl, K. (2002). Ubiquitination of histone H2B by Rad6 is required for efficient Dot1-mediated methylation of histone H3 lysine 79. J. Biol. Chem. *277*, 34655–34657. https://doi.org/10.1074/jbc.C200433200.
13. McGinty, R.K., Kim, J., Chatterjee, C., Roeder, R.G., and Muir, T.W. (2008). Chemically ubiquitylated histone H2B stimulates hDot1L-mediated intranucleosomal methylation. Nature *453*, 812–816. https://doi.org/10.1038/nature06906.
14. Briggs, S.D., Xiao, T., Sun, Z.W., Caldwell, J.A., Shabanowitz, J., Hunt, D.F., Allis, C.D., and Strahl, B.D. (2002). Trans-histone regulatory pathway in chromatin. Nature *418*, 498. https://doi.org/10.1038/nature00970.
15. Worden, E.J., Hoffmann, N.A., Hicks, C.W., and Wolberger, C. (2019). Mechanism of Cross-talk between H2B Ubiquitination and H3 Methylation by Dot1L. Cell *176*, 1490–1501.e12. https://doi.org/10.1016/j.cell.2019.02.002.
16. Barry, E.R., Krueger, W., Jakuba, C.M., Veilleux, E., Ambrosi, D.J., Nelson, C.E., and Rasmussen, T.P. (2009). ES cell cycle progression and differentiation require the action of the histone methyltransferase Dot1L. Stem Cell. *27*, 1538–1547. https://doi.org/10.1002/stem.86.
17. Wakeman, T.P., Wang, Q., Feng, J., and Wang, X.-F. (2012). Bat3 facilitates H3K79 dimethylation by DOT1L and promotes DNA damage-induced 53BP1 foci at G1/G2 cell-cycle phases. EMBO J. *31*, 2169–2181. https://doi.org/10.1038/emboj.2012.50.
18. Ng, H.H., Feng, Q., Wang, H., Erdjument-Bromage, H., Tempst, P., Zhang, Y., and Struhl, K. (2002). Lysine methylation within the globular domain of histone H3 by Dot1 is important for telomeric silencing and Sir protein association. Genes Dev. *16*, 1518–1527. https://doi.org/10.1101/gad.1001502.
19. Kim, W., Kim, R., Park, G., Park, J.W., and Kim, J.E. (2012). Deficiency of H3K79 histone methyltransferase Dot1-like protein (DOT1L) inhibits cell proliferation. J. Biol. Chem. *287*, 5588–5599. https://doi.org/10.1074/jbc.M111.328138.
20. Kim, W., Choi, M., and Kim, J.E. (2014). The histone methyltransferase Dot1/DOT1L as a critical regulator of the cell cycle. Cell Cycle *13*, 726–738. https://doi.org/10.4161/cc.28104.
21. Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q., and Zhao, K. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nat. Genet. *40*, 897–903. https://doi.org/10.1038/ng.154.
22. Steger, D.J., Lefterova, M.I., Ying, L., Stonestrom, A.J., Schupp, M., Zhuo, D., Vakoc, A.L., Kim, J.-E., Chen, J., Lazar, M.A., et al. (2008). DOT1L/KMT4 Recruitment and H3K79 Methylation Are Ubiquitously Coupled with Gene Transcription in Mammalian Cells. Mol. Cell Biol. *28*, 2825–2839. https://doi.org/10.1128/mcb.02076-07.
23. Vakoc, C.R., Sachdeva, M.M., Wang, H., and Blobel, G.A. (2006). Profile of Histone Lysine Methylation across Transcribed Mammalian Chromatin. Mol. Cell Biol. *26*, 9185–9195. https://doi.org/10.1128/mcb.01529-06.
24. Schübeler, D., MacAlpine, D.M., Scalzo, D., Wirbelauer, C., Kooperberg, C., Van Leeuwen, F., Gottschling, D.E., O'Neill, L.P., Turner, B.M., Delrow, J., et al. (2004). The histone modification pattern of active genes revealed through genome-wide chromatin analysis of a higher eukaryote. Genes Dev. *18*, 1263–1271. https://doi.org/10.1101/gad.1198204.
25. Kim, S.K., Jung, I., Lee, H., Kang, K., Kim, M., Jeong, K., Kwon, C.S., Han, Y.M., Kim, Y.S., Kim, D., and Lee, D. (2012). Human histone H3K79 methyltransferase DOT1L methyltransferase

binds actively transcribing RNA polymerase II to regulate gene expression. J. Biol. Chem. *287*, 39698–39709. https://doi.org/10.1074/jbc.M112.384057.

26. Jones, B., Su, H., Bhat, A., Lei, H., Bajko, J., Hevi, S., Baltus, G.A., Kadam, S., Zhai, H., Valdez, R., et al. (2008). The histone H3K79 methyltransferase Dot1L is essential for mammalian development and heterochromatin structure. PLoS Genet. *4*, e1000190. https://doi.org/10.1371/journal.pgen.1000190.
27. Morello, G., Porazzi, P., Moro, E., Argenton, F., Basso, G., Felix, C.A., and Germano, G. (2012). Zebrafish Ortholog of Human DOT1L Regulates Primitive and Transient Definitive Hematopoiesis and Controls hoxa9 and meis1 Expression. Blood *120*, 849. https://doi.org/10.1182/blood.v120.21.849.849.
28. Shanower, G.A., Muller, M., Blanton, J.L., Honti, V., Gyurkovics, H., and Schedl, P. (2005). Characterization of the grappa gene, the Drosophila Histone H3 lysine 79 methyltransferase. Genetics *169*, 173–184. https://doi.org/10.1534/genetics.104.033191.
29. Castaño Betancourt, M.C., Cailotto, F., Kerkhof, H.J., Cornelis, F.M.F., Doherty, S.A., Hart, D.J., Hofman, A., Luyten, F.P., Maciewicz, R.A., Mangino, M., et al. (2012). Genome-wide association and functional studies identify the DOT1L gene to be involved in cartilage thickness and hip osteoarthritis. Proc. Natl. Acad. Sci. USA *109*, 8218–8223. https://doi.org/10.1073/pnas.1119899109.
30. Mahmoudi, T., Boj, S.F., Hatzis, P., Li, V.S.W., Taouatas, N., Vries, R.G.J., Teunissen, H., Begthel, H., Korving, J., Mohammed, S., et al. (2010). The leukemia-associated Mllt10/Af10-Dot1l are Tcf4/β-catenin coactivators essential for intestinal homeostasis. PLoS Biol. *8*, e1000539. https://doi.org/10.1371/journal.pbio.1000539.
31. Mohan, M., Herz, H.M., Takahashi, Y.H., Lin, C., Lai, K.C., Zhang, Y., Washburn, M.P., Florens, L., and Shilatifard, A. (2010). Linking H3K79 trimethylation to Wnt signaling through a novel Dot1-containing complex (DotCom). Genes Dev. *24*, 574–589. https://doi.org/10.1101/gad.1898410.
32. Splinter, K., Adams, D.R., Bacino, C.A., Bellen, H.J., Bernstein, J.A., Cheatle-Jarvela, A.M., Eng, C.M., Esteves, C., Gahl, W.A., Hamid, R., et al. (2018). Effect of Genetic Diagnosis on Patients with Previously Undiagnosed Disease. N. Engl. J. Med. *379*, 2131–2139. https://doi.org/10.1056/nejmoa1714458.
33. Ramoni, R.B., Mulvihill, J.J., Adams, D.R., Allard, P., Ashley, E.A., Bernstein, J.A., Gahl, W.A., Hamid, R., Loscalzo, J., McCray, A.T., et al. (2017). The Undiagnosed Diseases Network: Accelerating Discovery about Health and Disease. Am. J. Hum. Genet. *100*, 185–192. https://doi.org/10.1016/J.AJHG.2017.01.006.
34. Costain, G., Walker, S., Marano, M., Veenma, D., Snell, M., Curtis, M., Luca, S., Buera, J., Arje, D., Reuter, M.S., et al. (2020). Genome Sequencing as a Diagnostic Test in Children with Unexplained Medical Complexity. JAMA Netw. Open *3*, e2018109. https://doi.org/10.1001/jamanetworkopen.2020.18109.
35. Deshwar, A.R., Yuki, K.E., Hou, H., Liang, Y., Khan, T., Celik, A., Ramani, A., Mendoza-Londono, R., Marshall, C.R., Brudno, M., et al. (2023). Trio RNA sequencing in a cohort of medically complex children. Am. J. Hum. Genet. *110*, 895–900. https://doi.org/10.1016/j.ajhg.2023.03.006.
36. Sobreira, N., Schiettecatte, F., Valle, D., and Hamosh, A. (2015). GeneMatcher: A Matching Tool for Connecting Investigators with an Interest in the Same Gene. Hum. Mutat. *36*, 928–930. https://doi.org/10.1002/HUMU.22844.
37. Hamosh, A., Wohler, E., Martin, R., Griffith, S., Rodrigues, E.d.S., Antonescu, C., Doheny, K.F., Valle, D., and Sobreira, N. (2022). The impact of GeneMatcher on international data sharing and collaboration. Hum. Mutat. *43*, 668–673. https://doi.org/10.1002/HUMU.24350.
38. Deciphering Developmental Disorders Study, Clayton, S., Fitzgerald, T.W., Kaplanis, J., Prigmore, E., Rajan, D., Sifrim, A., Aitken, S., Akawi, N., Alvi, M., et al. (2017). Prevalence and architecture of de novo mutations in developmental disorders. Nature *542*, 433–438. https://doi.org/10.1038/nature21062.
39. Aryee, M.J., Jaffe, A.E., Corrada-Bravo, H., Ladd-Acosta, C., Feinberg, A.P., Hansen, K.D., and Irizarry, R.A. (2014). Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics *30*, 1363–1369. https://doi.org/10.1093/bioinformatics/btu049.
40. Salas, L.A., Koestler, D.C., Butler, R.A., Hansen, H.M., Wiencke, J.K., Kelsey, K.T., and Christensen, B.C. (2018). An optimized library for reference-based deconvolution of whole-blood biospecimens assayed using the Illumina HumanMethylationEPIC BeadArray. Genome Biol. *19*, 64. https://doi.org/10.1186/s13059-018-1448-7.
41. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47. https://doi.org/10.1093/nar/gkv007.
42. Kanca, O., Zirin, J., Garcia-Marques, J., Knight, S.M., Yang-Zhou, D., Amador, G., Chung, H., Zuo, Z., Ma, L., He, Y., et al. (2019). An efficient CRISPR-based strategy to insert small and large fragments of DNA using short homology arms. Elife *8*, e51539. https://doi.org/10.7554/eLife.51539.
43. Lee, P.-T., Zirin, J., Kanca, O., Lin, W.-W., Schulze, K.L., Li-Kroeger, D., Tao, R., Devereaux, C., Hu, Y., Chung, V., et al. (2018). A gene-specific T2A-GAL4 library for Drosophila. Elife *7*, e35574. https://doi.org/10.7554/eLife.35574.
44. Goodman, L.D., Cope, H., Nil, Z., Ravenscroft, T.A., Charng, W.L., Lu, S., Tien, A.C., Pfundt, R., Koolen, D.A., Haaxma, C.A., et al. (2021). TNPO2 variants associate with human developmental delays, neurologic deficits, and dysmorphic features and alter TNPO2 activity in Drosophila. Am. J. Hum. Genet. *108*, 1669–1691. https://doi.org/10.1016/j.ajhg.2021.06.019.
45. Bischof, J., Björklund, M., Furger, E., Schertel, C., Taipale, J., and Basler, K. (2013). A versatile platform for creating a comprehensive UAS-ORFeome library in Drosophila. Development *140*, 2434–2442. https://doi.org/10.1242/DEV.088757.
46. Venken, K.J.T., He, Y., Hoskins, R.A., and Bellen, H.J. (2006). P[acman]: A BAC transgenic platform for targeted insertion of large DNA fragments in D. melanogaster. Science *314*, 1747–1751. https://doi.org/10.1126/science.1134426.
47. Wang, L., Lin, G., Zuo, Z., Li, Y., Byeon, S.K., Pandey, A., and Bellen, H.J. (2022). Neuronal activity induces glucosylceramide that is secreted via exosomes for lysosomal degradation in glia. Sci. Adv. *8*, 3326. https://doi.org/10.1126/sciadv.abn3326.
48. Wang, J., Al-Ouran, R., Hu, Y., Kim, S.Y., Wan, Y.W., Wangler, M.F., Yamamoto, S., Chao, H.T., Comjean, A., Mohr, S.E., et al. (2017). MARRVEL: Integration of Human and Model Organism Genetic Resources to Facilitate Functional Annotation of the Human Genome. Am. J. Hum. Genet. *100*, 843–853. https://doi.org/10.1016/j.ajhg.2017.04.010.

49. Wang, J., Liu, Z., Bellen, H.J., and Yamamoto, S. (2019). Navigating MARRVEL, a Web-Based Tool that Integrates Human Genomics and Model Organism Genetics Information. J. Vis. Exp.e59542 https://doi.org/10.3791/59542.
50. Quinodoz, M., Royer-Bertrand, B., Cisarova, K., Di Gioia, S.A., Superti-Furga, A., and Rivolta, C. (2017). DOMINO: Using Machine Learning to Predict Genes Associated with Dominant Disorders. Am. J. Hum. Genet. *101*, 623–629. https://doi.org/10.1016/j.ajhg.2017.09.001.
51. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141,456 humans. Nat *581*, 434–443. https://doi.org/10.1038/s41586-020-2308-7.
52. Reisenauer, M.R., Wang, S.W., Xia, Y., and Zhang, W. (2010). Dot1a contains three nuclear localization signals and regulates the epithelial Na+ channel (ENaC) at multiple levels. Am. J. Physiol. Renal Physiol. *299*, 63–76. https://doi.org/10.1152/ajprenal.00105.2010.
53. Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G.A., Sonnhammer, E.L.L., Tosatto, S.C.E., Paladin, L., Raj, S., Richardson, L.J., et al. (2021). Pfam: The protein families database in 2021. Nucleic Acids Res. *49*, D412–D419. https://doi.org/10.1093/NAR/GKAA913.
54. Zhang, W., Hayashizaki, Y., and Kone, B.C. (2004). Structure and regulation of the mDot1 gene, a mouse histone H3 methyltransferase. Biochem. J. *377*, 641–651. https://doi.org/10.1042/bj20030839.
55. Mueller, D., García-Cuéllar, M.P., Bach, C., Buhl, S., Maethner, E., and Slany, R.K. (2009). Misguided Transcriptional Elongation Causes Mixed Lineage Leukemia. PLoS Biol. *7*, 1000249. https://doi.org/10.1371/journal.pbio.1000249.
56. Mueller, D., Bach, C., Zeisig, D., Garcia-Cuellar, M.P., Monroe, S., Sreekumar, A., Zhou, R., Nesvizhskii, A., Chinnaiyan, A., Hess, J.L., and Slany, R.K. (2007). A role for the MLL fusion partner ENL in transcriptional elongation and chromatin modification. Blood *110*, 4445–4454. https://doi.org/10.1182/blood-2007-05-090514.
57. Kuntimaddi, A., Achille, N.J., Thorpe, J., Lokken, A.A., Singh, R., Hemenway, C.S., Adli, M., Zeleznik-Le, N.J., and Bushweller, J.H. (2015). Degree of Recruitment of DOT1L to MLL-AF9 Defines Level of H3K79 Di- and Trimethylation on Target Genes and Transformation Potential. Cell Rep. *11*, 808–820. https://doi.org/10.1016/j.celrep.2015.04.004.
58. Shah, S., and Henriksen, M.A. (2011). A novel disrupter of telomere silencing 1-like (DOT1L) interaction is required for signal transducer and activator of transcription 1 (STAT1)-activated gene expression. J. Biol. Chem. *286*, 41195–41204. https://doi.org/10.1074/jbc.M111.284190.
59. Hu, Y., Flockhart, I., Vinayagam, A., Bergwitz, C., Berger, B., Perrimon, N., and Mohr, S.E. (2011). An integrative approach to ortholog prediction for disease-focused and other functional studies. BMC Bioinf. *12*, 357. https://doi.org/10.1186/1471-2105-12-357.
60. McLaren, W., Gil, L., Hunt, S.E., Riat, H.S., Ritchie, G.R.S., Thormann, A., Flicek, P., and Cunningham, F. (2016). The Ensembl Variant Effect Predictor. Genome Biol. *17*, 122–214. https://doi.org/10.1186/s13059-016-0974-4.
61. Halldorsson, B.V., Eggertsson, H.P., Moore, K.H.S., Hauswedell, H., Eiriksson, O., Ulfarsson, M.O., Palsson, G., Hardarson, M.T., Oddsson, A., Jensson, B.O., et al. (2022). The sequences of 150,119 genomes in the UK Biobank. Nature *607*, 732–740. https://doi.org/10.1038/s41586-022-04965-x.
62. Taliun, D., Harris, D.N., Kessler, M.D., Carlson, J., Szpiech, Z.A., Torres, R., Taliun, S.A.G., Corvelo, A., Gogarten, S.M., Kang, H.M., et al. (2021). Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. Nature *590*, 290–299. https://doi.org/10.1038/s41586-021-03205-y.
63. Ioannidis, N.M., Rothstein, J.H., Pejaver, V., Middha, S., McDonnell, S.K., Baheti, S., Musolf, A., Li, Q., Holzinger, E., Karyadi, D., et al. (2016). REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. Am. J. Hum. Genet. *99*, 877–885. https://doi.org/10.1016/j.ajhg.2016.08.016.
64. MacDonald, J.R., Ziman, R., Yuen, R.K.C., Feuk, L., and Scherer, S.W. (2014). The Database of Genomic Variants: A curated collection of structural variation in the human genome. Nucleic Acids Res. *42*, D986–D992. https://doi.org/10.1093/nar/gkt958.
65. Chater-Diehl, E., Goodman, S.J., Cytrynbaum, C., Turinsky, A.L., Choufani, S., and Weksberg, R. (2021). Anatomy of DNA methylation signatures: Emerging insights and applications. Am. J. Hum. Genet. *108*, 1359–1366. https://doi.org/10.1016/j.ajhg.2021.06.015.
66. Venken, K.J.T., Carlson, J.W., Schulze, K.L., Pan, H., He, Y., Spokony, R., Wan, K.H., Koriabine, M., de Jong, P.J., White, K.P., et al. (2009). Versatile P[acman] BAC libraries for transgenesis studies in Drosophila melanogaster. Nat. Methods *6*, 431–434. https://doi.org/10.1038/nmeth.1331.
67. Soller, M., and White, K. (2004). Curr. Biol. *14*, R53. https://doi.org/10.1016/j.cub.2003.12.041.
68. Xiong, W.C., Okano, H., Patel, N.H., Blendy, J.A., and Montell, C. (1994). repo encodes a glial-specific homeo domain protein required in the Drosophila nervous system. Genes Dev. *8*, 981–994. https://doi.org/10.1101/gad.8.8.981.
69. Park, P.H., Yamamoto, T.M., Li, H., Alcivar, A.L., Xia, B., Wang, Y., Bernhardy, A.J., Turner, K.M., Kossenkov, A.V., Watson, Z.L., et al. (2020). Amplification of the mutation-carrying BRCA2 allele promotes RAD51 loading and PARP inhibitor resistance in the absence of reversion mutations. Mol. Cancer Ther. *19*, 602–613. https://doi.org/10.1158/1535-7163.MCT-17-0256.
70. Franz, H., Villarreal, A., Heidrich, S., Videm, P., Kilpert, F., Mestres, I., Calegari, F., Backofen, R., Manke, T., and Vogel, T. (2019). DOT1L promotes progenitor proliferation and primes neuronal layer identity in the developing cerebral cortex. Nucleic Acids Res. *47*, 168–183. https://doi.org/10.1093/nar/gky953.
71. Fahrner, J.A., and Bjornsson, H.T. (2019). Mendelian disorders of the epigenetic machinery: postnatal malleability and therapeutic prospects. Hum. Mol. Genet. *28*, R254–R264. https://doi.org/10.1093/hmg/ddz174.
72. Luo, X., Schoch, K., Jangam, S.V., Bhavana, V.H., Graves, H.K., Kansagra, S., Jasien, J.M., Stong, N., Keren, B., Mignot, C., et al. (2021). Rare deleterious de novo missense variants in Rnf2/Ring2 are associated with a neurodevelopmental disorder with unique clinical features. Hum. Mol. Genet. *30*, 1283–1292. https://doi.org/10.1093/hmg/ddab110.
73. Feng, Q., Wang, H., Ng, H.H., Erdjument-Bromage, H., Tempst, P., Struhl, K., and Zhang, Y. (2002). Methylation of H3-Lysine 79 Is Mediated by a New Family of HMTases

without a SET Domain. Curr. Biol. *12*, 1052–1058. https://doi.org/10.1016/S0960-9822(02)00901-6.

74. Costain, G., and Andrade, D.M. (2023). Third-generation computational approaches for genetic variant interpretation. Brain *146*, 411–412. https://doi.org/10.1093/brain/awad011.
75. Gerasimavicius, L., Livesey, B.J., and Marsh, J.A. (2022). Loss-of-function, gain-of-function and dominant-negative mutations have profoundly different effects on protein structure. Nat. Commun. *13*, 3895–3915. https://doi.org/10.1038/s41467-022-31686-6.
76. Hou, S.M., Hsia, C.W., Tsai, C.L., Hsia, C.H., Jayakumar, T., Velusamy, M., and Sheu, J.R. (2020). Modulation of human platelet activation and in vivo vascular thrombosis by columbianadin: Regulation by integrin αiIbβ3 inside-out but not outside-in signals. J. Biomed. Sci. *27*, 60. https://doi.org/10.1186/s12929-020-0619-5.

**The American Journal of Human Genetics, Volume *110***

## Supplemental information

# Rare *de novo* gain-of-function missense variants in *DOT1L* are associated with developmental delay and congenital anomalies

**Zelha Nil, Ashish R. Deshwar, Yan Huang, Scott Barish, Xi Zhang, Sanaa Choufani, Polona Le Quesne Stabej, Ian Hayes, Patrick Yap, Chad Haldeman-Englert, Carolyn Wilson, Trine Prescott, Kristian Tveten, Arve Vøllo, Devon Haynes, Patricia G. Wheeler, Jessica Zon, Cheryl Cytrynbaum, Rebekah Jobling, Moira Blyth, Siddharth Banka, Alexandra Afenjar, Cyril Mignot, Florence Robin-Renaldo, Boris Keren, Oguz Kanca, Xiao Mao, Daniel J. Wegner, Kathleen Sisco, Marwan Shinawi, Undiagnosed Disease Network, Michael F. Wangler, Rosanna Weksberg, Shinya Yamamoto, Gregory Costain, and Hugo J. Bellen**

## SUPPLEMENTAL INFORMATION

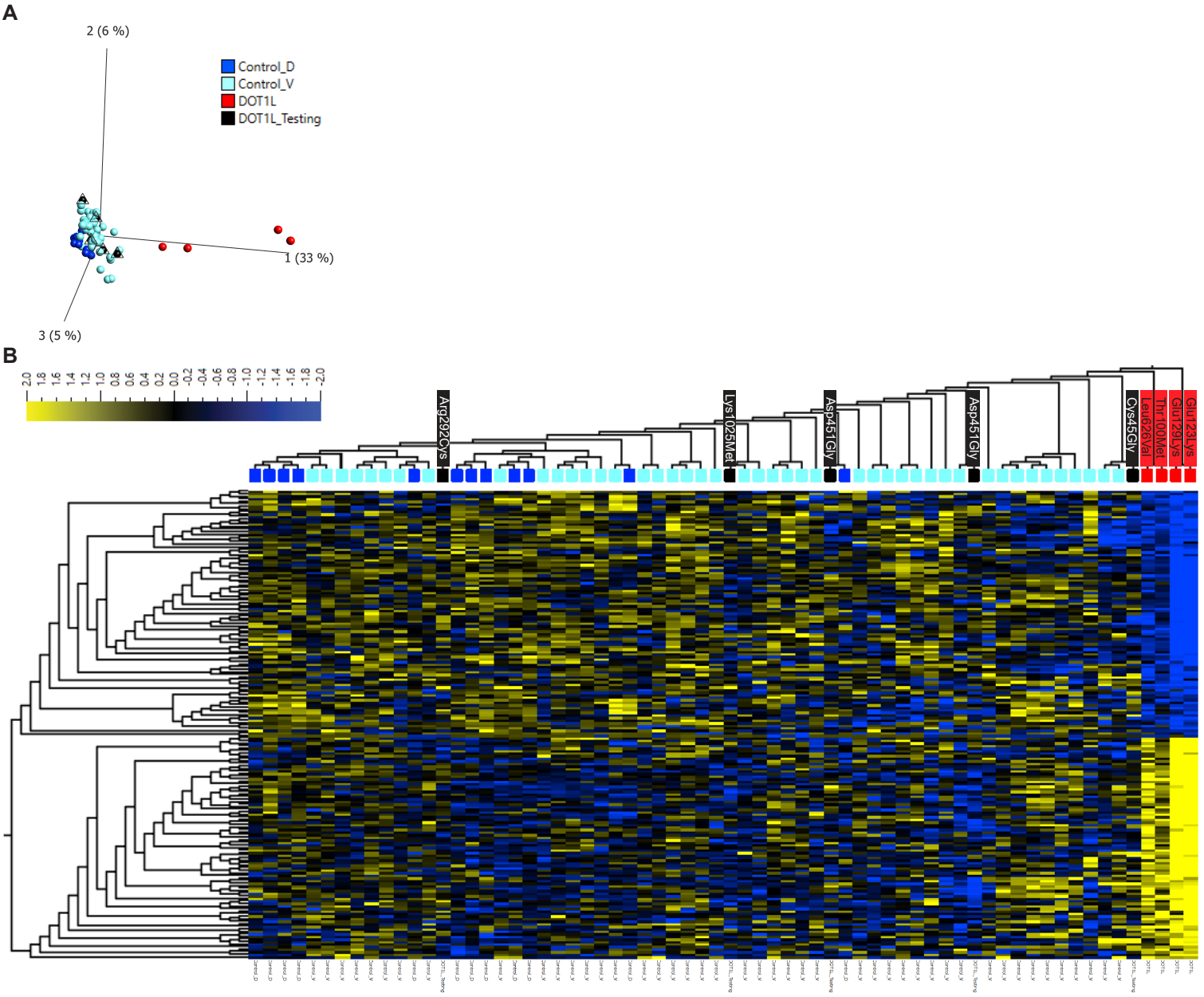



**Figure S1. DNA methylation profile of proband blood samples**

(A) Principal component analysis (PCA) and (B) heatmap showing clustering of the useable DOT1L samples (red) and matched control samples (blue) using DNAm values at the 185 CpG sites associated with a distinct DOT1L specific DNAm profile. The heatmap color gradient indicates the normalized DNAm value ranging from −2.0 (blue) to 2.0 (yellow). Euclidean distance metric is used in the heatmap clustering dendrograms. Please see Table S9 for significantly enriched CpG sites.

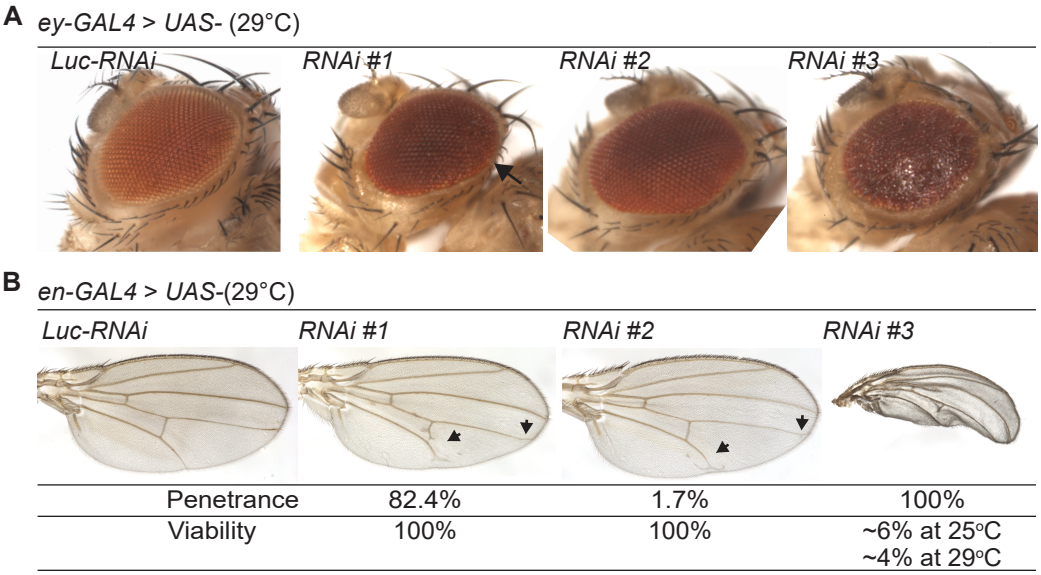



**Figure S2. Tissue specific knockdown of gpp causes eye and wing phenotypes**

(A) Eye specific knockdown of *gpp* causes a rough eye phenotype. *Ey-GAL4 > gpp-RNAi #1* causes a mild phenotype in a region of the eye indicated with an arrow. *Ey-GAL4 > RNAi #3* causes a severe phenotype in the whole eye. *Ey-GAL4 > RNAi #2* does not cause any phenotype. All the crosses were performed at 29ºC. (ey: eyeless)

(B) Knockdown of *gpp* in developing tissues causes wing phenotypes and lethality. *en-GAL4 > gpp-RNAi #1* causes cross vein branching with a penetrance of 82.4%. *en-GAL4 > RNAi #3* causes lethality and the survivors have a severe phenotype in the whole wing with a full penetrance. *en-GAL4 > RNAi #2* causes vein branching phenotype with a very low penetrance. All the crosses were performed at 29ºC. (en: engrailed)

| *UAS-DOT1L*/+* | $gpp^{TG4}$/*Df* (18, 25 & 29°C) |
|---|---|
| *Reference | Lethal |
| *p.Cys45Gly | Lethal |
| *p.Thr100Met | Lethal |
| *p.Glu123Lys | Lethal |
| *p.Glu129Lys | Lethal |
| *p.Arg292Cys | Lethal |
| *p.Leu626Val | Lethal |
| *p.Lys1025Met | Lethal |
| $GR^{gpp}$ | Viable |

**Figure S3. Human reference or variant *DOT1L* fails to rescue lethality**

Expression of reference or variant *DOT1L* in $gpp^{TG4}$/*Df* flies failed to rescue the lethality at all temperatures (18°C, 25°C and 29°C) tested while reintroduction of GR construct fully rescued lethality.

A

| $gpp^{TG4/+}$ > *UAS-DOT1L** (25°C) | | |
|---|---|---|
| | Phenotype | Penetrance |
| *Reference | wing defects | 96.6% |
| *p.Cys45Gly | wing defects | 97.7% |
| *p.Thr100Met | wing defects | 97.3% |
| *p.Glu123Lys | wing defects | 100% |
| *p.Glu129Lys | wing defects | 100% |
| *p.Arg292Cys | wing defects | 98.3% |
| *p.Leu626Val | wing defects | 98.5% |
| *p.Lys1025Met | wing defects | 97.8% |

B

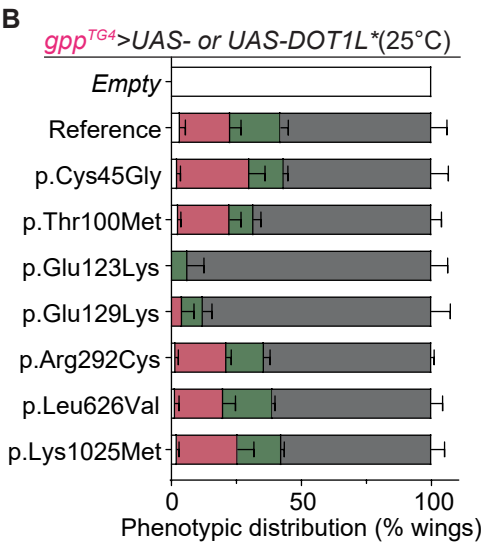
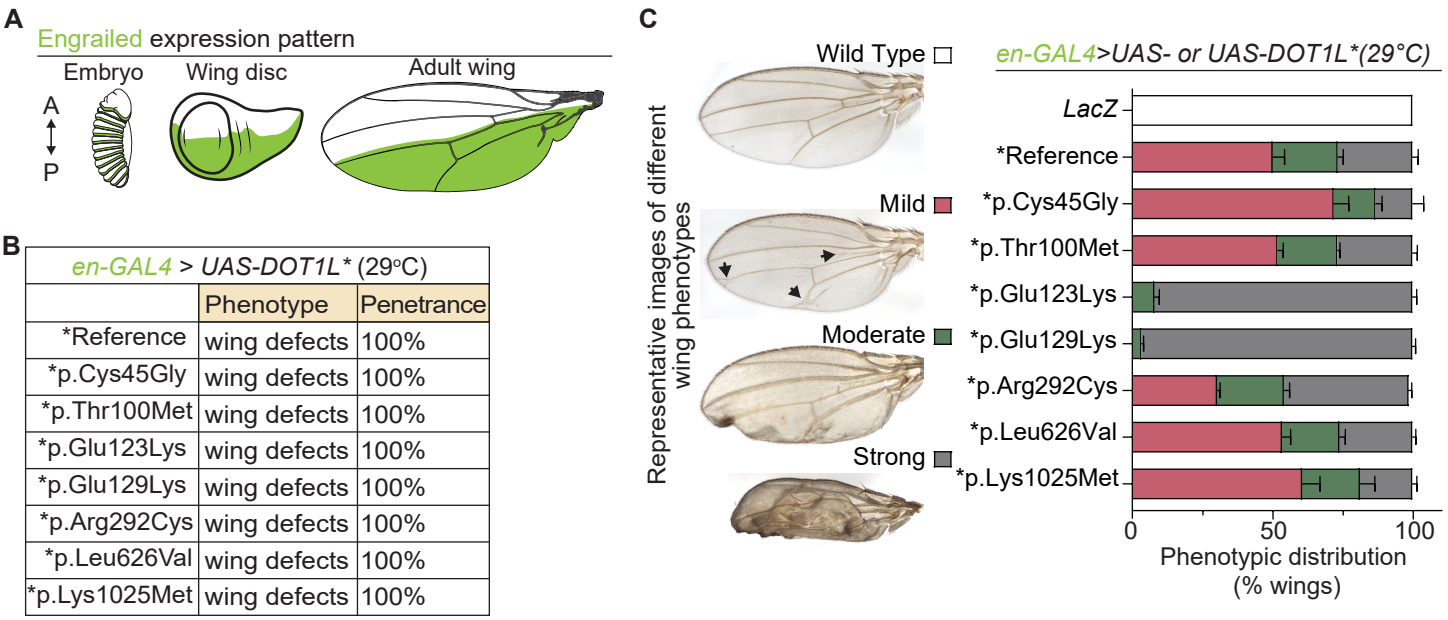



**Figure S4. Expression of reference or variant DOT1L in gpp expressing cells causes wing deformities**

(A) Survivors of heterozygous mutant flies, $gpp^{TG4/+}$, expressing reference or variant *DOT1L* cDNA show morphological wing defects such as whole wing blistering, necrosis, loss of cross-veins and extra vein branching with penetrance levels >95%.

(B) The distribution of different wing phenotypes in survivors of heterozygous mutant flies, $gpp^{TG4/+}$, expressing each *DOT1L* cDNA.




B

| *en-GAL4* > *UAS-DOT1L** (29°C) | | |
|---|---|---|
| | Phenotype | Penetrance |
| *Reference | wing defects | 100% |
| *p.Cys45Gly | wing defects | 100% |
| *p.Thr100Met | wing defects | 100% |
| *p.Glu123Lys | wing defects | 100% |
| *p.Glu129Lys | wing defects | 100% |
| *p.Arg292Cys | wing defects | 100% |
| *p.Leu626Val | wing defects | 100% |
| *p.Lys1025Met | wing defects | 100% |

**Figure S5. Expression of reference or variant DOT1L in developing tissues causes wing deformities**

(A) Schematic of engrailed expression from embryos to larval wing disc and adult wings.

(B) *en-GAL4 > UAS-DOT1L* reference or variant expressing flies show morphological wing defects such as whole wing blistering, necrosis, loss of cross-veins and extra vein branching.(en: engrailed)

(C) Representative images of different wing phenotypes (left panel) and the distribution of different wing phenotypes in *en-GAL4 > UAS-DOT1L* reference or variant expressing flies (right panel).

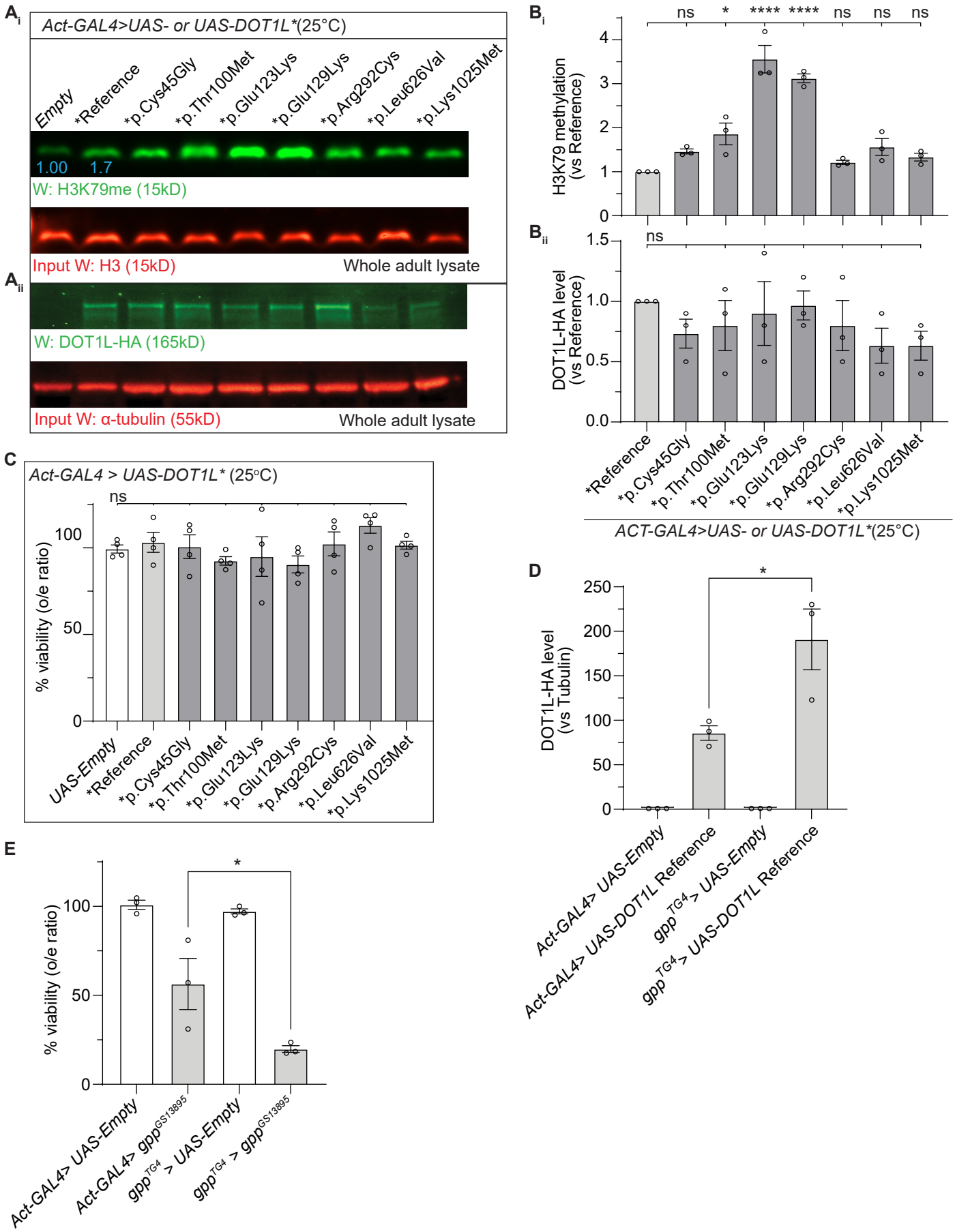

A
Act-GAL4>UAS- or UAS-DOT1L*(25°C)
Empty
*Reference
*p.Cys45Gly
*p.Thr100Met
*p.Glu123Lys
*p.Glu129Lys
*p.Arg292Cys
*p.Leu626Val
*p.Lys1025Met
1.00
1.7
W: H3K79me (15kD)
Input W: H3 (15kD)
Whole adult lysate
A
W: DOT1L-HA (165kD)
Input W: α-tubulin (55kD)
Whole adult lysate
B
H3K79 methylation (vs Reference)
ns
*
****
****
ns
ns
ns
B
DOT1L-HA level (vs Reference)
ns
*Reference
*p.Cys45Gly
*p.Thr100Met
*p.Glu123Lys
*p.Glu129Lys
*p.Arg292Cys
*p.Leu626Val
*p.Lys1025Met
ACT-GAL4>UAS- or UAS-DOT1L*(25°C)
C
Act-GAL4 > UAS-DOT1L* (25°C)
ns
% viability (o/e ratio)
UAS-Empty
*Reference
*p.Cys45Gly
*p.Thr100Met
*p.Glu123Lys
*p.Glu129Lys
*p.Arg292Cys
*p.Leu626Val
*p.Lys1025Met
D
*
DOT1L-HA level (vs Tubulin)
Act-GAL4> UAS-Empty
Act-GAL4> UAS-DOT1L Reference
gpp TG4 > UAS-Empty
gpp TG4 > UAS-DOT1L Reference
E
*
% viability (o/e ratio)
Act-GAL4> UAS-Empty
Act-GAL4> gppGS13895
gppTG4 > UAS-Empty
gppTG4 > gppGS13895

**Figure S6. H3K79 methylation pattern of flies ubiquitously expressing human DOT1L and variants**

($A_i$) H3K79 methylation levels and ($A_{ii}$) DOT1L levels in flies ubiquitously expressing reference or variant *DOT1L* cDNAs. Flies expressing reference *DOT1L* show increased H3K79 methylation compared to control flies (*Act-GAL4* > *UAS-Empty*). Protein lysate from 10 adult flies were prepared for each sample. H3K79 methylation levels were normalized with loading control, H3, and fold change for each sample were calculated by comparing normalized H3K79 methylation levels to reference *DOT1L* expressing flies. DOT1L levels were normalized with loading control, α-tubulin, and fold change for each sample were calculated by comparing normalized DOT1L levels to reference *DOT1L* expressing flies. All the crosses were performed at 25$^{o}$C. Blue numbers indicate the fold change of H3K79 methylation level in reference when compared to control (*UAS-Empty*).

($B_i$) Normalized H3K79 methylation band intensities for each group from three independent experiments were plotted as mean ± SEM, and statistical significance was determined by one-way ANOVA for multiple groups (* $p<0.05$, **** $p < 0.0001$). ($B_{ii}$) Normalized DOT1L band intensities for each group from three independent experiments were plotted as mean ± SEM, and statistical significance was determined by one-way ANOVA for multiple groups.

(C) Percent viability of flies ubiquitously (*Act-GAL4*) expressing reference or variant *DOT1L* cDNAs. All the crosses were performed at 25$^{o}$C. Percent viabilities (o/e ratios) from three independent experiments were plotted as mean ± SEM, and statistical significance was determined by one-way ANOVA for multiple groups.

(D) Expression of DOT1L cDNA in gpp expression domains using gppTG4 results in higher protein levels than ubiquitous expression of DOT1L using Act-GAL4 driver. Normalized DOT1L band intensities for control (UAS-empty) and UAS-DOT1L-Reference from three independent experiments (Figure 4E and Figure S6A) were plotted as mean ± SEM, and statistical significance was determined by one-way ANOVA for multiple groups.

(E) Percent viability of flies expressing fly gpp (gppGS13895) ubiquitously (Act-GAL4) or in gpp expression domains (gppTG4). All the crosses were performed at 25oC. Percent viabilities (o/e ratios) from three independent experiments were plotted as mean ± SEM, and statistical significance was determined by one-way ANOVA for multiple groups (* $p<0.05$).

**Table S2. Details on match outcomes from Genematcher for *DOT1L* from October 2017 to June 2019**

| Event Type | Date of Match | Country of Clinician/Individual | Outcome | Variant | Sample Provided for DNAm Profiling |
|---|---|---|---|---|---|
| Original individual | N/A | Canada | Included in study | c.133T>G, p.Cys45Gly | Yes* |
| Previously published | N/A | England | Included in study | c.874C>T, p.Arg292Cys | Yes |
| Match | October 17th 2017 | Netherlands | No response | - | - |
| Match | October 17th 2017 | Netherlands | No response | - | - |
| Match | October 17th, 2017 | France | No response | - | - |
| Match | October 17th, 2017 | USA | Included in study | c.299C>T, p.Thr100Met | Yes |
| Match | December 5th, 2017 | USA | No response after initial contact | [frameshift variant] | - |
| Match | December 19th, 2017 | USA | Included in study | c.1352A>G, p.Asp451Gly | Yes (for both family members) |
| Match | March 19th, 2018 | USA | Included in study | c.367G>A, p.Glu123Lys | Yes |
| Match | May 7th, 2018 | France | Included in study | c.385G>A, p.Glu129Lys | Yes |
| Match | May 7th, 2018 | France | Included in study | c.3074A>T, p.Lys1025Met | Yes* |
| Match | July 18th, 2018 | New Zealand | Included in study | c.367G>A, p.Glu123Lys | No (not available; individual deceased) |
| Match | June 11th 2019 | Norway | Included in study | c.1876C>G, p.Leu626Val | Yes |

*not utilized in analysis due to lack of age matched controls

**Table S3. Sequencing methods for individuals with *DOT1L* variants**

| Proband | 1 | 2 | 3 NZ | 4 Ch | 5 OR | 6 | 7 | 8 | 9 | 10 | 11 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| ***DOT1L* variant (NM-032482.3)** | c.133T>G p.Cys45Gly | c.299C>T p.Thr100Met | c.367G>A p.Glu123Lys | c.367G>A p.Glu123Lys | c.367G>A p.Glu123Lys | c.385G>A p.Glu129Lys | c.1876C>G p.Leu626Val | c.2557C>T p.Arg853Cys | c.3074A>T p.Lys1025Met | c.874C>T p.Arg292Cys | c.1352A>G p.Asp451Gly |
| **Sequencing approach** | Trio genome sequencing | Trio genome sequencing | Singleton exome sequencing | Whole exome sequencing | Trio exome sequencing | Trio exome sequencing | Trio exome sequencing | Whole exome sequencing | Trio exome sequencing | Trio exome sequencing | Trio exome sequencing |
| **Capture reagent** | - | N/A | SOPHiA Genetics Clinical Exome Solution v1 | N/A | GeneDx Proprietary System | Roche SeqCap EZ MedExome | Nextera Rapid Capture Exome Kit (Illumina) | N/A | Roche SeqCap EZ MedExome | SureSelect RNA baits (Agilent) | N/A |
| **Sequencer** | HiSeq X platform (Illumina Inc) | HiSeq X platform (Illumina Inc) | Illumina Inc | N/A | Illumina Inc | Illumina NextSeq 500 | NextSeq 500 (Illumina) | N/A | Illumina NextSeq 500 | Illumina HiSeq | N/A |
| **Location of Sequencing** | The Center for Applied Genomics, Toronto, Canada | HudsonAlpha Clinical Services Lab, Huntsville, Alabama | SOPHiA GENETICS | N/A | GeneDx | Genetic laboratory, Pitié-Salpêtrière Hospital, Paris, France | Telemark Hospital Trust, Skien, Norway | N/A | Genetic laboratory, Pitié-Salpêtrière Hospital, Paris, France | Genetics services of the UK National Health Service and the Republic of Ireland | N/A |
| **Publication of Sequencing Methods** | PMID: 32960281 | - | - | - | - | PMID: 31580924 | PMID: 26534809 | - | PMID: 31580924 | PMID: 28135719 | - |

**Table S4. Publicly available fly lines used in this study**

| Fly Line | Genotype | BDSC # |
|---|---|---|
| *gpp$^{XXV}$* | *In(3R)gppXXV, gppXXV/TM3, P{ActGFP}JMR2, Ser1* | 42231 |
| *Deficiency* | *w[1118]; Df(3R)BSC193/TM6B, Tb[+]* | 9620 |
| *Genomic rescue* | *w[1118]; Dp(3;2)GV-CH321-05H03, PBac{y[+mDint2] w[+mC]=GV-CH321-05H03}VK00037/CyO* | 90095 |
| *gpp RNAi #1* | *y1 v1; P{TRiP.JF01284}attP2* | 31327 |
| *gpp RNAi #2* | *y1 v1; P{TRiP.JF01283}attP2/TM3, Ser1* | 31481 |
| *gpp RNAi #3* | *y[1] sc[*] v[1] sev[21]; P{y[+t7.7] v[+t1.8]=TRiP.GL01325}attP2* | 41893 |
| *gpp RNAi #4* | *y[1] v[1]; P{y[+t7.7] v[+t1.8]=TRiP.HMJ02129}attP40* | 42556 |
| *Luciferase RNAi* | *y[1] v[1];; P{y[+t7.7] v[+t1.8]=TRiP.JF01355}attP2* | 31603 |
| *ey-GAL4 (on II)* | *w[*]; P{w[+m*] =GAL4-ey.H}3-8* | 5534 |
| *da-GAL4 (on III)* | *w[*]; P{w[+mW.hs]=GAL4-da.G32}UH1, Sb[1]/TM6B, Tb[1]* | 55851 |
| *en-GAL4 (on II)* | *w[1118]; P{w[+mW.hs]=en2.4-GAL4}e16E, P{w[+mC]=UAS-RFP.W}2/CyO* | 30557 |
| *act-GAL4 (on II)* | *y[1] w[*]; P{w[+mC]=Act5C-GAL4}25FO1/CyO, y[+]* | 4414 |
| *repo-GAL4 (on III)* | *w[1118]; P{w[+m*]=GAL4}repo/TM3, Sb[1]* | 7415 |
| *UAS-mCherry. NLS (on II)* | *w[*]; P{w[+mC]=UAS-mCherry.NLS}2; MKRS/TM6B, Tb[1]* | 38425 |
| *UAS-Empty (on II)* | *w[*]; P{w[+mC]=UAS}/CyO* | |
| *gpp$^{GS13895}$* | *y[1] w[67c23]; P{w[+mC]=GSV6}GS13895/TM3, Sb[1] Ser[1]* | Kyoto 205608 |

**Table S5. Mutagenesis and qPCR primers used in this study**

| Name | Assay | Forward primer (5'-3') | Reverse primer (5'-3') |
|---|---|---|---|
| p.Cys45Gly | Mutagenesis | CCGATGGGTCgGTGAAGAAATC | ATGGTCTCGATGATTTCATG |
| p.Thr100Met | Mutagenesis | AAGCTGAACAtGCGGCCGTCC | CATGGGCTGCGTGGTGCC |
| p.Glu123Lys | Mutagenesis | GACCGACCCCaAGAAGCTCAA | ACCGAGTGGTTGTAGACC |
| p.Glu129Lys | Mutagenesis | CAACAACTACaAGCCCTTCTCCCC | AGCTTCTCGGGGTCGGTC |
| p.Arg292Cys | Mutagenesis | CACCATCATGtGCGTGGTGGAGC | CCGATGTCACTCAAGTTTC |
| p.Leu626Val | Mutagenesis | GAAGCAGGCCgTGAAGAGCCA | TCCTTCAACAGCTTCTCCAG |
| p.Lys1025Met | Mutagenesis | GAGGCCAGCAtGGGAGACCTGCCC | GGGCAACGGGCCCTGGGC |
| qPrimer1 | qPCR | ATCGTTGCATTGGAAAAAGG | GTACGGCGGCATTGTAAACT |
| qPrimer2 | qPCR | CCATTGGAAGGTCTAGCAGC | CTGCTGCACGTCGTTGAC |
| RpL32 | qPCR | TAAGCTGTCGCACAAATGGCG | AACGCGGTTCTGCATGAGCA |

**Table S7. Clinical features of individuals with suspected non-diagnostic variants in *DOT1L***

| Individual | 10 | 11 |
|---|---|---|
| ***DOT1L* variant (NM_032482.3)** | c.874C>T<br>p.Arg292Cys | c.1352A>G<br>p.Asp451Gly |
| **Inheritance** | *de novo* | Inherited from unaffected father |
| **Sex** | Female | Female |
| **Age at last assessment** | 11 years | 11 months |
| **Medical History** | | |
| **Brain anomalies (MRI/ CT)** | N/A | N/A |
| **Cardiac Anomalies** | No | No |
| **Hypotonia** | No | No |
| **Musculoskeletal anomalies** | No | Yes<br>Torticollis<br>flexible hips, leg length discrepancy, right sided weakness with facial asymmetry |
| **Urogenital anomalies** | No | No |
| **Hearing loss** | No | Yes |
| **Ophthalmological anomalies** | No | Yes<br>Amblyopia, anisometropia |
| **Growth and Development** | | |
| **Global developmental delay** | Yes | - |
| **Intellectual Disability** | Yes | - |
| **Language** | Non-verbal | Speech delay, articulation concern |
| **Height percentile** | 27.8% | 3.3% |
| **Microcephaly (percentile)** | No<br>($10^{th}$) | No<br>($63^{rd}$) |

Abbreviations are as follows: N/A: Not available

**Table S8. Summary of variant details, allele frequencies and *in-silico* predictions for all *DOT1L* variants reported in our study**

| **Variant Details** | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Type** | Missense | Missense | Missense | Missense | Missense | Missense | Missense | Missense | Missense |
| **g. coordinates [GRCh37 (Chr19)]** | g.2185861T>G | g.2191045C>T | g.2191113G>A | g.2191131G>A | g.2207590C>T | g.2211098A>G | g.2214548C>G | g.2217783C>T | g.2222242A>T |
| **c. variant (NM_032482.3)** | c.133T>G | c.299C>T | c.367G>A | c.385G>A | c.874C>T | c.1352A>G | c.1876C>G | c.2557C>T | c.3074A>T |
| **p. variant (NP_115871.1)** | p.Cys45Gly | p.Thr100Met | p.Glu123Lys | p.Glu129Lys | p.Arg292Cys | p.Asp451Gly | p.Leu626Val | p.Arg853Cys | p.Lys1025Met |
| **Variant Allele Frequencies (searched December 2022)** | | | | | | | | | |
| **gnomAD v2.1.1** | 0 | 0 | 0 | 0 | 0 | 31 | 0 | 0 | 1 |
| **gnomAD v3.1.2** | 0 | 0 | 0 | 0 | 0 | 19 | 0 | 0 | 0 |
| **TOPMed Bravo** | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 2 | 0 |
| **deCAF** | 0 | 2 | 0 | 0 | 0 | 97 | 0 | 0 | 0 |
| ***In Silico* Predictions & Conservation Metrics** | | | | | | | | | |
| **Aggregated Prediction (Franklin)** | Deleterious (0.71) | Benign (0.12) | Uncertain (0.55) | Uncertain (0.55) | Uncertain (0.36) | Benign (0.09) | Uncertain (0.29) | Uncertain (0.41) | Benign (0.07) |
| **Revel** | Deleterious (Supporting) (0.67) | Benign (Moderate) (0.12) | Uncertain (0.47) | Uncertain (0.47) | Benign (Supporting) (0.24) | Benign (Moderate) (0.09) | Benign (Moderate) (0.17) | Uncertain (0.31) | Benign (Moderate) (0.07) |
| **Varity** | Deleterious (0.96) | Deleterious (low) (0.43) | Deleterious (0.84) | Deleterious (0.72) | Deleterious (0.64) | Benign (0.08) | Benign (low) (0.28) | Deleterious (low) (0.44) | Benign (low) (0.25) |
| **MutationAssessor** | Low deleterious probability (1.83) | Medium deleterious probability (1.94) | Medium deleterious probability (1.98) | Medium deleterious probability (1.99) | Medium deleterious probability (2.09) | Medium deleterious probability (2.05) | Medium deleterious probability (1.96) | Medium deleterious probability (2.33) | Medium deleterious probability (1.98) |
| **FATHMM** | Uncertain (1.93) | Uncertain (2.06) | Uncertain (2.02) | Uncertain (2.05) | Uncertain (1.95) | Uncertain (1.81) | Uncertain (1.5) | Uncertain (1.39) | Uncertain (1.68) |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **MetaLR** | Benign (low) (0.15) | Benign (0.13) | Benign (0.11) | Benign (0.11) | Benign (0.05) | Benign (low) (0.25) | Benign (low) (0.17) | Benign (low) (0.17) | Benign (0.15) |
| **BayesDel** | Uncertain (0.09) | Benign (Supporting) (-0.35) | Uncertain (-0.11) | Uncertain (-0.14) | Uncertain (-0.06) | Benign (Moderate) (-0.43) | Benign (Supporting) (-0.35) | Uncertain (-0.17) | Benign (Moderate) (-0.43) |
| **CADD** | Phred: 27.0 Raw score: 3.912343 | Phred: 25.4 Raw score: 3.620297 | Phred: 29.2 Raw score: 4.170345 | Phred: 26.0 Raw score: 3.745109 | Phred: 24.7 Raw score: 3.406433 | Phred: 25.6 Raw score: 3.652334 | Phred: 23.8 Raw score: 3.080506 | Phred: 24.8 Raw score: 3.453433 | Phred: 23.0 Raw score: 2.738641 |
| **Metadome** | Tolerance score (dn/ds): 0.07 (highly intolerant) | Tolerance score (dn/ds): 0.4 (intolerant) | Tolerance score (dn/ds): 0.06 (highly intolerant) | Tolerance score (dn/ds): 0.06 (highly intolerant) | Tolerance score (dn/ds): 0.12 (highly intolerant) | Tolerance score (dn/ds): 0.62 (slightly intolerant) | Tolerance score (dn/ds): 0.34 (intolerant) | Tolerance score (dn/ds): 0.58 (slightly intolerant) | Tolerance score (dn/ds): 0.62 (slightly intolerant) |
| **Grantham distance** | Large physicochemical difference (159) | Moderate physicochemical difference (81) | Small physicochemical difference (56) | Small physicochemical difference (56) | Large physicochemical difference (180) | Moderate physicochemical difference (94) | Small physicochemical difference (32) | Large physicochemical difference (180) | Moderate physicochemical difference (95) |
| **GERP** | Uncertain (4.68) | Uncertain (4.75) | Uncertain (4.75) | Uncertain (4.75) | Uncertain (3.63) | Uncertain (3.62) | Uncertain (3.96) | Uncertain (4.15) | Uncertain (3.49) |
| **Amino acid residue conservation (13 species)** | Highly conserved (11/13) | Moderately conserved (8/13) | Highly conserved (11/13) | Highly conserved (11/13) | Moderately conserved (10/13) | Moderately conserved (9/13) | Moderately conserved (10/13) | Moderately conserved (9/13) | Moderately conserved (8/13) |
| **PhyloP** | Moderately conserved nucleotide (7.08) | Moderately conserved nucleotide (5.64) | Highly conserved nucleotide (9.17) | Highly conserved nucleotide (9.17) | Weakly conserved nucleotide (2.94) | Moderately conserved nucleotide (6.78) | Weakly conserved nucleotide (2.83) | Weakly conserved nucleotide (2.29) | Weakly conserved nucleotide (1.45) |

**Consortia**

**Undiagnosed Diseases Network**

Maria T. Acosta, Margaret Adam, David R. Adams, Justin Alvey, Laura Amendola, Ashley Andrews, Euan A. Ashley, Mahshid S. Azamian, Carlos A. Bacino, Guney Bademci, Ashok Balasubramanyam, Dustin Baldridge, Jim Bale, Michael Bamshad, Deborah Barbouth, Pinar Bayrak-Toydemir, Anita Beck, Alan H. Beggs, Edward Behrens, Gill Bejerano, Hugo J. Bellen, Jimmy Bennet, Beverly Berg-Rood, Jonathan A. Bernstein, Gerard T. Berry, Anna Bican, Stephanie Bivona, Elizabeth Blue, John Bohnsack, Devon Bonner, Lorenzo Botto, Brenna Boyd, Lauren C. Briere, Elly Brokamp, Gabrielle Brown, Elizabeth A. Burke, Lindsay C. Burrage, Manish J. Butte, Peter Byers, William E. Byrd, John Carey, Olveen Carrasquillo, Thomas Cassini, Ta Chen Peter Chang, Sirisak Chanprasert, Hsiao-Tuan Chao, Gary D. Clark, Terra R. Coakley, Laurel A. Cobban, Joy D. Cogan, Matthew Coggins, F. Sessions Cole, Heather A. Colley, Cynthia M. Cooper, Heidi Cope, William J. Craigen, Andrew B. Crouse, Michael Cunningham, Precilla D'Souza, Hongzheng Dai, Surendra Dasari, Joie Davis, Jyoti G. Dayal, Matthew Deardorff, Esteban C. Dell'Angelica, Katrina Dipple, Daniel Doherty, Naghmeh Dorrani, Argenia L. Doss, Emilie D. Douine, Laura Duncan, Dawn Earl, David J. Eckstein, Lisa T. Emrick, Christine M. Eng, Cecilia Esteves, Marni Falk, Liliana Fernandez, Elizabeth L. Fieg, Paul G. Fisher, Brent L. Fogel, Irman Forghani, William A. Gahl, Ian Glass, Bernadette Gochuico, Rena A. Godfrey, Katie Golden-Grant, Madison P. Goldrich, Alana Grajewski, Irma Gutierrez, Don Hadley, Sihoun Hahn, Rizwan Hamid, Kelly Hassey, Nichole Hayes, Frances High, Anne Hing, Fuki M. Hisama, Ingrid A. Holm, Jason Hom, Martha Horike-Pyne, Alden Huang, Yong Huang, Wendy Introne, Rosario Isasi, Kosuke Izumi, Fariha Jamal, Gail P. Jarvik, Jeffrey Jarvik, Suman Jayadev, Orpa Jean-Marie, Vaidehi Jobanputra, Lefkothea Karaviti, Jennifer Kennedy, Shamika Ketkar, Dana Kiley, Gonench Kilich, Shilpa N. Kobren, Isaac S. Kohane, Jennefer N. Kohler, Deborah Krakow, Donna M. Krasnewich, Elijah Kravets, Susan Korrick, Mary Koziura, Seema R. Lalani, Byron Lam, Christina Lam, Grace L. LaMoure, Brendan C. Lanpher, Ian R. Lanza, Kimberly LeBlanc, Brendan H. Lee, Roy Levitt, Richard A. Lewis, Pengfei Liu, Xue Zhong Liu, Nicola Longo, Sandra K. Loo, Joseph Loscalzo, Richard L. Maas, Ellen F. Macnamara, Calum

A. MacRae, Valerie V. Maduro, Rachel Mahoney, Bryan C. Mak, May Christine V. Malicdan, Laura A. Mamounas, Teri A. Manolio, Rong Mao, Kenneth Maravilla, Ronit Marom, Gabor Marth, Beth A. Martin, Martin G. Martin, Julian A. Martínez-Agosto, Shruti Marwaha, Jacob McCauley, Allyn McConkie-Rosell, Alexa T. McCray, Elisabeth McGee, Heather Mefford, J. Lawrence Merritt, Matthew Might, Ghayda Mirzaa, Eva Morava, Paolo M. Moretti, Mariko Nakano-Okuno, Stan F. Nelson, John H. Newman, Sarah K. Nicholas, Deborah Nickerson, Shirley Nieves-Rodriguez, Donna Novacic, Devin Oglesbee, James P. Orengo, Laura Pace, Stephen Pak, J. Carl Pallais, Christina GS. Palmer, Jeanette C. Papp, Neil H. Parker, John A. Phillips III, Jennifer E. Posey, Lorraine Potocki, Barbara N. Pusey, Aaron Quinlan, Wendy Raskind, Archana N. Raja, Deepak A. Rao, Anna Raper, Genecee Renteria, Chloe M. Reuter, Lynette Rives, Amy K. Robertson, Lance H. Rodan, Jill A. Rosenfeld, Natalie Rosenwasser, Francis Rossignol, Maura Ruzhnikov, Ralph Sacco, Jacinda B. Sampson, Mario Saporta, Judy Schaechter, Timothy Schedl, Kelly Schoch, C. Ron Scott, Daryl A. Scott, Vandana Shashi, Jimann Shin, Edwin K. Silverman, Janet S. Sinsheimer, Kathy Sisco, Edward C. Smith, Kevin S. Smith, Emily Solem, Lilianna Solnica- Krezel, Ben Solomon, Rebecca C. Spillmann, Joan M. Stoler, Jennifer A. Sullivan, Kathleen Sullivan, Angela Sun, Shirley Sutton, David A. Sweetser, Virginia Sybert, Holly K. Tabor, Amelia L. M. Tan, Queenie K.-G. Tan, Mustafa Tekin, Fred Telischi, Willa Thorson, Cynthia J. Tifft, Camilo Toro, Alyssa A. Tran, Brianna M. Tucker, Tiina K. Urv, Adeline Vanderver, Matt Velinder, Dave Viskochil, Tiphanie P. Vogel, Colleen E. Wahl, Melissa Walker, Stephanie Wallace, Nicole M. Walley, Jennifer Wambach, Jijun Wan, Lee-kai Wang, Michael F. Wangler, Patricia A. Ward, Daniel Wegner, Monika Weisz-Hubshman, Mark Wener, Tara Wenger, Katherine Wesseling Perry, Monte Westerfield, Matthew T. Wheeler, Jordan Whitlock, Lynne A. Wolfe, Kim Worley, Changrui Xiao, Shinya Yamamoto, John Yang, Diane B. Zastrow, Zhe Zhang, Chunli Zhao, Stephan Zuchner

## SUPPLEMENTARY REFERENCES


1. Liu, H., Liu, D.T., Lan, S., Yang, Y., Huang, J., Huang, J., and Fang, L. (2021). ASH1L mutation caused seizures and intellectual disability in twin sisters. J. Clin. Neurosci. 91, 69–74. 10.1016/J.JOCN.2021.06.038.
2. Kleefstra, T., Van Zelst-Stams, W.A., Nillesen, W.M., Cormier-Daire, V., Houge, G., Foulds, N., Van Dooren, M., Willemsen, M.H., Pfundt, R., Turner, A., et al. (2009). Further clinical and molecular delineation of the 9q subtelomeric deletion syndrome supports a major contribution of EHMT1 haploinsufficiency to the core phenotype. J. Med. Genet. 46, 598–606. 10.1136/jmg.2008.062950.
3. Goodman, S.J., Cytrynbaum, C., Chung, B.H.-Y., Chater-Diehl, E., Aziz, C., Turinsky, A.L., Kellam, B., Keller, M., Ko, J.M., Caluseriu, O., et al. (2020). EHMT1 pathogenic variants and 9q34.3 microdeletions share altered DNA methylation patterns in patients with Kleefstra syndrome. J. Transl. Genet. Genomics 4, 144–158. 10.20517/jtgg.2020.23.
4. Gibson, W.T., Hood, R.L., Zhan, S.H., Bulman, D.E., Fejes, A.P., Moore, R., Mungall, A.J., Eydoux, P., Babul-Hirji, R., An, J., et al. (2012). Mutations in EZH2 cause weaver syndrome. Am. J. Hum. Genet. 90, 110–118. 10.1016/j.ajhg.2011.11.018.
5. Choufani, S., Gibson, W.T., Turinsky, A.L., Chung, B.H.Y., Wang, T., Garg, K., Vitriolo, A., Cohen, A.S.A., Cyrus, S., Goodman, S., et al. (2020). DNA Methylation Signature for EZH2 Functionally Classifies Sequence Variants in Three PRC2 Complex Genes. Am. J. Hum. Genet. 106, 596–610. 10.1016/j.ajhg.2020.03.008.
6. Jones, W.D., Dafou, D., McEntagart, M., Woollard, W.J., Elmslie, F. V., Holder-Espinasse, M., Irving, M., Saggar, A.K., Smithson, S., Trembath, R.C., et al. (2012). De novo mutations in MLL cause Wiedemann-Steiner syndrome. Am. J. Hum. Genet. 91, 358–364. 10.1016/j.ajhg.2012.06.008.
7. Foroutan, A., Haghshenas, S., Bhai, P., Levy, M.A., Kerkhof, J., McConkey, H., Niceta, M., Ciolfi, A., Pedace, L., Miele, E., et al. (2022). Clinical Utility of a Unique Genome-Wide DNA Methylation Signature for KMT2A-Related Syndrome. Int. J. Mol. Sci. 23, 1815. 10.3390/ijms23031815.
8. Zech, M., Boesch, S., Maier, E.M., Borggraefe, I., Vill, K., Laccone, F., Pilshofer, V., Ceballos-Baumann, A., Alhaddad, B., Berutti, R., et al. (2016). Haploinsufficiency of KMT2B, Encoding the Lysine-Specific Histone Methyltransferase 2B, Results in Early-Onset Generalized Dystonia. Am. J. Hum. Genet. 99, 1377–1387. 10.1016/j.ajhg.2016.10.010.

9. Mirza-Schreiber, N., Zech, M., Wilson, R., Brunet, T., Wagner, M., Jech, R., Boesch, S., Škorvánek, M., Necpál, J., Weise, D., et al. (2022). Blood DNA methylation provides an accurate biomarker of KMT2B-related dystonia and predicts onset. Brain 145, 644–654. 10.1093/BRAIN/AWAB360.
10. Lee, S., Ochoa, E., Barwick, K., Cif, L., Rodger, F., Docquier, F., Pérez-Dueñas, B., Clark, G., Martin, E., Banka, S., et al. (2022). Comparison of methylation episignatures in KMT2B-and KMT2D-related human disorders. Epigenomics 14, 537–547. 10.2217/epi-2021-0521.
11. Cif, L., Demailly, D., Lin, J.P., Barwick, K.E., Sa, M., Abela, L., Malhotra, S., Chong, W.K., Steel, D., Sanchis-Juan, A., et al. (2020). KMT2B-related disorders: Expansion of the phenotypic spectrum and long-term efficacy of deep brain stimulation. Brain 143, 3242–3261. 10.1093/brain/awaa304.
12. Koemans, T.S., Kleefstra, T., Chubak, M.C., Stone, M.H., Reijnders, M.R.F., de Munnik, S., Willemsen, M.H., Fenckova, M., Stumpel, C.T.R.M., Bok, L.A., et al. (2017). Functional convergence of histone methyltransferases EHMT1 and KMT2C involved in intellectual disability and autism spectrum disorder. PLoS Genet. 13, e1006864. 10.1371/journal.pgen.1006864.
13. Cuvertino, S., Hartill, V., Colyer, A., Garner, T., Nair, N., Al-Gazali, L., Canham, N., Faundes, V., Flinter, F., Hertecant, J., et al. (2020). A restricted spectrum of missense KMT2D variants cause a multiple malformations disorder distinct from Kabuki syndrome. Genet. Med. 22, 980. 10.1038/s41436-020-0784-7.
14. Baldridge, D., Spillmann, R.C., Wegner, D.J., Wambach, J.A., White, F. V., Sisco, K., Toler, T.L., Dickson, P.I., Sessions Cole, F, et al. (2020). Phenotypic expansion of KMT2D-related disorder: Beyond Kabuki syndrome. Am J Med Genet A. 182(5):1053-1065. 10.1002/ajmg.a.61518.
15. Butcher, D.T., Cytrynbaum, C., Turinsky, A.L., Siu, M.T., Inbar-Feigenberg, M., Mendoza-Londono, R., Chitayat, D., Walker, S., Machado, J., Caluseriu, O., et al. (2017). CHARGE and Kabuki Syndromes: Gene-Specific DNA Methylation Signatures Identify Epigenetic Mechanisms Linking These Clinically Overlapping Conditions. Am. J. Hum. Genet. 100, 773–788. 10.1016/j.ajhg.2017.04.004.
16. Ng, S.B., Bigham, A.W., Buckingham, K.J., Hannibal, M.C., McMillin, M.J., Gildersleeve, H.I., Beck, A.E., Tabor, H.K., Cooper, G.M., Mefford, H.C., et al. (2010). Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. Nat. Genet. 42, 790–793. 10.1038/ng.646.

17. O'Donnell-Luria, A.H., Pais, L.S., Faundes, V., Wood, J.C., Sveden, A., Luria, V., Abou Jamra, R., Accogli, A., Amburgey, K., Anderlid, B.M., et al. (2019). Heterozygous Variants in KMT2E Cause a Spectrum of Neurodevelopmental Disorders and Epilepsy. Am. J. Hum. Genet. 104, 1210–1222. 10.1016/j.ajhg.2019.03.021.
18. Niihori, T., Ouchi-Uchiyama, M., Sasahara, Y., Kaneko, T., Hashii, Y., Irie, M., Sato, A., Saito-Nanjo, Y., Funayama, R., Nagashima, T., et al. (2015). Mutations in MECOM, Encoding Oncoprotein EVI1, Cause Radioulnar Synostosis with Amegakaryocytic Thrombocytopenia. Am. J. Hum. Genet. 97, 848–854. 10.1016/j.ajhg.2015.10.010.
19. Kurotaki, N., Imaizumi, K., Harada, N., Masuno, M., Kondoh, T., Nagai, T., Ohashi, H., Naritomi, K., Tsukahara, M., Makita, Y., et al. (2002). Haploinsufficiency of NSD1 causes Sotos syndrome. Nat. Genet. 30, 365–366. 10.1038/ng863.
20. Choufani, S., Cytrynbaum, C., Chung, B.H.Y., Turinsky, A.L., Grafodatskaya, D., Chen, Y.A., Cohen, A.S.A., Dupuis, L., Butcher, D.T., Siu, M.T., et al. (2015). NSD1 mutations generate a genome-wide DNA methylation signature. Nat. Commun. 2015 61 6, 1–7. 10.1038/ncomms10207.
21. Lozier, E.R., Konovalov, F.A., Kanivets, I. V., Pyankov, D. V., Koshkin, P.A., Baleva, L.S., Sipyagina, A.E., Yakusheva, E.N., Kuchina, A.E., and Korostelev, S.A. (2018). De novo nonsense mutation in WHSC1 (NSD2) in patient with intellectual disability and dysmorphic features. J. Hum. Genet. 63, 919–922. 10.1038/s10038-018-0464-5.
22. Kawai, T., Kinoshita, S., Takayama, Y., Onishi, E., Kamura, H., Kojima, K., Kikuchi, H., Terao, M., Sugawara, T., Migita, O., et al. (2023). DNA methylation signature in NSD2 loss-of-function variants appeared similar to that in Wolf-Hirschhorn syndrome. bioRxiv, 2023.01.06.522834. 10.1101/2023.01.06.522834.
23. Li, N., Subrahmanyan, L., Smith, E., Yu, X., Zaidi, S., Choi, M., Mane, S., Nelson-Williams, C., Bahjati, M., Kazemi, M., et al. (2016). Mutations in the Histone Modifier PRDM6 Are Associated with Isolated Nonsyndromic Patent Ductus Arteriosus. Am. J. Hum. Genet. 98, 1082–1091. 10.1016/j.ajhg.2016.03.022.
24. Arndt, A.K., Schafer, S., Drenckhahn, J.D., Sabeh, M.K., Plovie, E.R., Caliebe, A., Klopocki, E., Musso, G., Werdich, A.A., Kalwa, H., et al. (2013). Fine mapping of the 1p36 deletion syndrome identifies mutation of PRDM16 as a cause of cardiomyopathy. Am. J. Hum. Genet. 93, 67–77. 10.1016/j.ajhg.2013.05.015.
25. Yu, X., Yang, L., Li, J., Li, W., Li, D., Wang, R., Wu, K., Chen, W., Zhang, Y., Qiu, Z., et al. (2019). De Novo and Inherited SETD1A Variants in Early-onset Epilepsy. Neurosci. Bull. 35, 1045–1057. 10.1007/s12264-019-00400-w.

26. Kummeling, J., Stremmelaar, D.E., Raun, N., Reijnders, M.R.F., Willemsen, M.H., Ruiterkamp-Versteeg, M., Schepens, M., Man, C.C.O., Gilissen, C., Cho, M.T., et al. (2021). Characterization of SETD1A haploinsufficiency in humans and Drosophila defines a novel neurodevelopmental syndrome. Mol. Psychiatry 26, 2013–2024. 10.1038/s41380-020-0725-5.
27. Hiraide, T., Nakashima, M., Yamoto, K., Fukuda, T., Kato, M., Ikeda, H., Sugie, Y., Aoto, K., Kaname, T., Nakabayashi, K., et al. (2018). De novo variants in SETD1B are associated with intellectual disability, epilepsy and autism. Hum. Genet. 137, 95–104. 10.1007/s00439-017-1863-y.
28. Krzyzewska, I.M., Maas, S.M., Henneman, P., Lip, K. V.D., Venema, A., Baranano, K., Chassevent, A., Aref-Eshghi, E., Van Essen, A.J., Fukuda, T., et al. (2019). A genome-wide DNA methylation signature for SETD1B-related syndrome. Clin. Epigenetics 11, 1–15. 10.1186/S13148-019-0749-3/FIGURES/6.
29. Rabin, R., Radmanesh, A., Glass, I.A., Dobyns, W.B., Aldinger, K.A., Shieh, J.T., Romoser, S., Bombei, H., Dowsett, L., Trapane, P., et al. (2020). Genotype-phenotype correlation at codon 1740 of SETD2. Am. J. Med. Genet. A 182, 2037–2048. 10.1002/AJMG.A.61724.
30. Luscan, A., Laurendeau, I., Malan, V., Francannet, C., Odent, S., Giuliano, F., Lacombe, D., Touraine, R., Vidaud, M., Pasmant, E., et al. (2014). Mutations in SETD2 cause a novel overgrowth condition. J. Med. Genet. 51, 512–517. 10.1136/jmedgenet-2014-102402.
31. Lumish, H.S., Wynn, J., Devinsky, O., and Chung, W.K. (2015). Brief Report: SETD2 Mutation in a Child with Autism, Intellectual Disabilities and Epilepsy. J. Autism Dev. Disord. 45, 3764–3770. 10.1007/s10803-015-2484-8.