Mutation-specific pathophysiological mechanisms define different neurodevelopmental disorders associated with SATB1 dysfunction

Joery den Hoed[1, 2, 72], Elke de Boer[3, 4, 72], Norine Voisin[5, 72], Alexander J.M. Dingemans[3, 4], Nicolas Guex[5, 6], Laurens Wiel[3, 7, 8], Christoffer Nellaker[9, 10, 11], Shivarajan M. Amudhavalli[12, 13], Siddharth Banka[14, 15], Frederique S. Bena[16], Bruria Ben-Zeev[17], Vincent R. Bonagura[18, 19], Ange-Line Bruel[20, 21], Theresa Brunet[22], Han G. Brunner[3, 4, 23], Hui B. Chew[24], Jacqueline Chrast[5], Loreta Cimbalistienė[25], Hillary Coon[26], The DDD study[27], Emmanuelle C. Délot[28], Florence Démurger[29], Anne-Sophie Denommé-Pichon[20, 21], Christel Depienne[30], Dian Donnai[14, 15], David A. Dyment[31], Orly Elpeleg[32], Laurence Faivre[20, 33, 34], Christian Gilissen[3, 7], Leslie Granger[35], Benjamin Haber[36], Yasuo Hachiya[37], Yasmin Hamzavi Abedi[38, 39], Jennifer Hanebeck[36], Jayne Y. Hehir-Kwa[40], Brooke Horist[41], Toshiyuki Itai[42], Adam Jackson[14], Rosalyn Jewell[43], Kelly L. Jones[44, 45], Shelagh Joss[46], Hirofumi Kashii[37], Mitsuhiro Kato[47], Anja A. Kattentidt-Mouravieva[48], Fernando Kok[49, 50], Urania Kotzaeridou[36], Vidya Krishnamurthy[41], Vaidutis Kučinskas[25], Alma Kuechler[30], Alinoë Lavillaureix[51], Pengfei Liu[52, 53], Linda Manwaring[54], Naomichi Matsumoto[42], Benoît Mazel[33], Kirsty McWalter[55], Vardiella Meiner[32], Mohamad A. Mikati[56], Satoko Miyatake[42], Takeshi Mizuguchi[42], Lip H. Moey[57], Shehla Mohammed[58], Hagar Mor-Shaked[32], Hayley Mountford[59], Ruth Newbury-Ecob[60], Sylvie Odent[51], Laura Orec[36], Matthew Osmond[31], Timothy B. Palculict[55], Michael Parker[61], Andrea K. Petersen[35], Rolph Pfundt[3], Egle Preikšaitienė[25], Kelly Radtke[62], Emmanuelle Ranza[16, 63], Jill A. Rosenfeld[52], Teresa Santiago-Sim[55], Caitlin Schwager[12, 13], Margje Sinnema[23, 64], Lot Snijders Blok[1, 3, 4], Rebecca C. Spillmann[65], Alexander P.A. Stegmann[3, 23], Isabelle Thiffault[12, 66, 67], Linh Tran[56], Adi Vaknin-Dembinsky[32], Juliana H. Vedovato-dos-Santos[49], Samantha A. Vergano[44], Eric Vilain[28], Antonio Vitobello[20, 21], Matias Wagner[22, 68], Androu Waheeb[31, 69], Marcia Willing[54], Britton Zuccarelli[70], Usha Kini[71], Dianne F. Newbury[59], Tjitske Kleefstra[3, 4], Alexandre Reymond[5, 73], Simon E. Fisher[1, 4, 73*], Lisenka E.L.M. Vissers[3, 4, 73]

Affiliations

1. Language and Genetics Department, Max Planck Institute for Psycholinguistics, 6500 AH Nijmegen, the Netherlands.

2. International Max Planck Research School for Language Sciences, Max Planck Institute for Psycholinguistics, 6500 AH Nijmegen, The Netherlands.

3. Department of Human Genetics, Radboudumc, 6500 HB Nijmegen, the Netherlands.

4. Donders Institute for Brain, Cognition and Behaviour, Radboud University, 6500 GL Nijmegen, the Netherlands.

5. Center for Integrative Genomics, University of Lausanne, 1015 Lausanne, Switzerland.

6. Bioinformatics Competence Center, University of Lausanne, 1015 Lausanne, Switzerland.

7. Radboud Institute for Molecular Life Sciences, Radboud University Medical Center, 6500 HB Nijmegen, The Netherlands.

8. Center for Molecular and Biomolecular Informatics of the Radboudumc, 6500 HB Nijmegen, the Netherlands.

9. Nuffield Department of Women's and Reproductive Health, University of Oxford, Women's Centre, John Radcliffe Hospital, Oxford OX3 9DU, UK.

10. Institute of Biomedical Engineering, Department of Engineering Science, University of Oxford, Oxford OX3 7DQ, UK.

11. Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford OX3 7LF, UK.

12. University of Missouri-Kansas City School of Medicine, Kansas City, MO 64108, USA.

13. Department of Pediatrics, Division of Clinical Genetics, Children's Mercy Hospital, Kansas City, MO 64108, USA.

14. Manchester Centre for Genomic Medicine, Division of Evolution and Genomic Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester M13 9PL, UK.

15. Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester University NHS Foundation Trust, Health Innovation Manchester, Manchester M13 9WL, UK.

16. Service of Genetic Medicine, University Hospitals of Geneva, 1205 Geneva, Switzerland.
17. Edmomd and Lilly Safra Pediatric Hospital, Sheba Medical Center and Sackler School of Medicine, Tel Aviv University, Ramat Aviv 69978, Israel.
18. Institute of Molecular Medicine, Feinstein Institutes for Medical Research, NY 11030, USA.
19. Pediatrics and Molecular Medicine, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, NY 11549, USA.
20. UMR-Inserm, Génétique des Anomalies du développement, Université de Bourgogne Franche-Comté, 21070 Dijon, France.
21. Laboratoire de Génétique chromosomique et moléculaire, UF Innovation en diagnostic génomique des maladies rares, Centre Hospitalier Universitaire de Dijon, 21070 Dijon, France.
22. Institute of Human Genetics, Technical University of Munich, 81675 Munich, Germany.
23. Department of Clinical Genetics, Maastricht University Medical Center+, azM, 6202 AZ Maastricht, The Netherlands.
24. Department of Genetics, Kuala Lumpur Hospital, Jalan Pahang, 50586 Kuala Lumpur, Malaysia.
25. Department of Human and Medical Genetics, Institute of Biomedical Sciences, Faculty of Medicine, Vilnius University, LT-08661 Vilnius, Lithuania.
26. Department of Psychiatry, University of Utah School of Medicine, Salt Lake City, UT 84018, USA.
27. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, Cambridge CB10 1SA, UK.
28. Center for Genetic Medicine Research, Children's National Hospital, Children's Research Institute and Department of Genomics and Precision Medicine, George Washington University, Washington, DC 20010, USA.
29. Department of clinical genetics, Vannes hospital, 56017 Vannes, France.
30. Institute of Human Genetics, University Hospital Essen, University of Duisburg-Essen, 45147 Essen, Germany.

31. Children's Hospital of Eastern Ontario Research Institute, Ottawa, Ontario K1H 5B2, Canada.
32. Department of Genetics, Hadassah, Hebrew University Medical Center, Jerusalem 91120, Israel.
33. Centre de Génétique et Centre de Référence Anomalies du Développement et Syndromes Malformatifs de l'Interrégion Est, Centre Hospitalier Universitaire Dijon, 21079 Dijon, France.
34. Fédération Hospitalo-Universitaire Médecine Translationnelle et Anomalies du Développement (TRANSLAD), Centre Hospitalier Universitaire Dijon, 21079 Dijon, France.
35. Department of Rehabilitation and Development, Randall Children's Hospital at Legacy Emanuel Medical Center, Portland, Oregon 97227, USA.
36. Division of Child Neurology and Inherited Metabolic Diseases, Centre for Paediatrics and Adolescent Medicine, University Hospital Heidelberg, 69120 Heidelberg, Germany.
37. Department of Neuropediatrics, Tokyo Metropolitan Neurological Hospital, Fuchu, Tokyo 183-0042, Japan.
38. Division of Allergy and Immunology, Northwell Health, Great Neck, NY 11021, USA.
39. Departments of Medicine and Pediatrics, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY 11549, USA.
40. Princess Máxima Center for Pediatric Oncology, 3584 CS Utrecht, The Netherlands.
41. Pediatrics & Genetics, Alpharetta, GA 30005, USA.
42. Department of Human Genetics, Yokohama City University Graduate School of Medicine, Yokohama, Kanagawa 236-0004, Japan.
43. Yorkshire Regional Genetics Service, Chapel Allerton Hospital, Leeds LS7 4SA, UK.
44. Division of Medical Genetics & Metabolism, Children’s Hospital of the King’s Daughters, Norfolk, VA 23507, USA.
45. Department of Pediatrics, Eastern Virginia Medical School, Norfolk, VA 23507, USA.
46. West of Scotland Centre for Genomic Medicine, Queen Elizabeth University Hospital, Glasgow G51 4TF, UK.

47. Department of Pediatrics, Showa University School of Medicine, Shinagawa-ku, Tokyo 142-8666, Japan.

48. Zuidwester, 3241 LB Middelharnis, The Netherlands.

49. Mendelics Genomic Analysis, Sao Paulo, SP 04013-000, Brazil.

50. University of Sao Paulo, School of Medicine, Sao Paulo, SP 01246-903, Brazil.

51. CHU Rennes, Univ Rennes, CNRS, IGDR, Service de Génétique Clinique, Centre de Référence Maladies Rares CLAD-Ouest, ERN ITHACA, Hôpital Sud, 35033 Rennes, France.

52. Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA.

53. Baylor Genetics, Houston, TX 77021, USA.

54. Department of Pediatrics, Division of Genetics and Genomic Medicine, Washington University School of Medicine, St. Louis, MO 63110-1093, USA.

55. GeneDx, 207 Perry Parkway Gaithersburg, MD 20877, USA.

56. Division of Pediatric Neurology, Duke University Medical Center, Durham, NC 27710, USA.

57. Department of Genetics, Penang General Hospital, Jalan Residensi, 10990 Georgetown, Penang, Malaysia.

58. Clinical Genetics, Guy's Hospital, Great Maze Pond, London SE1 9RT, UK.

59. Department of Biological and Medical Sciences, Headington Campus, Oxford Brookes University, Oxford OX3 0BP, UK.

60. Clinical Genetics, St Michael's Hospital Bristol, University Hospitals Bristol NHS Foundation Trust, Bristol BS2 8EG, UK.

61. Sheffield Clinical Genetics Service, Sheffield Children's Hospital, Sheffield S5 7AU, UK.

62. Clinical Genomics Department, Ambry Genetics, Aliso Viejo, CA 92656, USA.

63. Medigenome, Swiss Institute of Genomic Medicine, 1207 Geneva, Switzerland.

64. Department of Genetics and Cell Biology, Faculty of Health Medicine Life Sciences, Maastricht University Medical Center+, Maastricht University, 6229 ER Maastricht, The Netherlands.

65. Department of Pediatrics, Division of Medical Genetics, Duke University Medical Center, Durham, NC 27713, USA.

66. Center for Pediatric Genomic Medicine, Children's Mercy Hospital, Kansas City, MO 64108, USA.

67. Department of Pathology and Laboratory Medicine, Children's Mercy Hospital, Kansas City, MO 64108, USA.

68. Institute of Neurogenomics, Helmholtz Zentrum München, D-85764 Munich, Germany.

69. Department of Genetics, Children's Hospital of Eastern Ontario, Ottawa, Ontario K1H 8L1, Canada.

70. The University of Kansas School of Medicine Salina Campus, Salina, KS 67401, USA.

71. Oxford Centre for Genomic Medicine, Oxford University Hospitals NHS Foundation Trust, Oxford OX3 7LE, UK.

72. These authors contributed equally to this work.

73. These authors contributed equally to this work.

* To whom correspondence should be addressed:

Prof. Dr. S.E. Fisher

Simon.Fisher@mpi.nl

Abstract

Whereas large-scale statistical analyses can robustly identify disease-gene relationships, they do not accurately capture genotype-phenotype correlations or disease mechanisms. We use multiple lines of independent evidence to show that different variant types in a single gene, *SATB1,* cause clinically overlapping but distinct neurodevelopmental disorders. Clinical evaluation of 42 individuals carrying *SATB1* variants identified overt genotype-phenotype relationships, associated with different pathophysiological mechanisms, established by functional assays. Missense variants in the CUT1 and CUT2 DNA-binding domains result in stronger chromatin binding, increased transcriptional repression and a severe phenotype. In contrast, variants predicted to result in haploinsufficiency are associated with a milder clinical presentation. A similarly mild phenotype is observed for individuals with premature protein truncating variants that escape nonsense-mediated decay, which are transcriptionally active but mislocalized in the cell. Our results suggest that in-depth mutation-specific genotype-phenotype studies are essential to capture full disease complexity and to explain phenotypic variability.

Main text

SATB1 encodes a dimeric/tetrameric transcription factor[1] with crucial roles in development and maturation of T-cells[2-4]. Recently, a potential contribution of *SATB1* to brain development was suggested by statistically significant enrichment of *de novo* variants in two large neurodevelopmental disorder (NDD) cohorts[5; 6], although its functions in the central nervous system are poorly characterized.

Through international collaborations[7-9] conforming to local ethical guidelines and the declaration of Helsinki, we identified 42 individuals with a rare (likely) pathogenic variant in *SATB1* (NM_001131010.4), a gene under constraint against loss-of-function and missense variation (pLoF: o/e=0.15 (0.08-0.29); missense: o/e=0.46 (0.41-0.52); gnomAD v2.1.1)[10]. Twenty-eight of the *SATB1* variants occurred *de novo,* three were inherited from an affected parent, and five resulted from (suspected) parental mosaicism (Figure S1). Reduced penetrance is suggested by two variants inherited from unaffected parents (identified in individual 2 and 12; Table S1A), consistent with recent predictions of incomplete penetrance being more prevalent in novel NDD syndromes[6]. Inheritance status of the final four could not be established (Table S1A). Of note, two individuals also carried a (likely) pathogenic variant affecting other known disease genes, including *NF1* (MIM #162200; individual 27) and *FOXP2 (*MIM #602081; individual 42) which contributed to (individual 27) or explained (individual 42) the observed phenotype (Table S1A).

Thirty individuals carried 15 unique *SATB1* missense variants, including three recurrent variants (Figure 1A), significantly clustering in the highly homologous DNA-binding domains CUT1 and CUT2 (p=1.00e-7; Figure 2A, Figure S2)[11; 12]. Ten individuals carried premature protein truncating variants (PTVs; two nonsense, seven frameshift, one splice site; Table S1A, Table S2), and two individuals had a (partial) gene deletion (Figure S3). For 38 affected individuals and one mosaic parent, clinical information was available. Overall, we observed a broad phenotypic spectrum, characterized by neurodevelopmental delay (35/36, 97%), ID (28/31, 90%), muscle tone abnormalities (abnormal tone 28/37, 76%; hypotonia 28/37, 76%; spasticity 10/36, 28%), epilepsy (22/37, 61%) behavioral problems (24/34, 71%), facial

dysmorphisms (24/36, 67%; Figure 1B-1C, Figure S4A), and dental abnormalities (24/34, 71%) (Figure 1D, Table 1, Figure S4B, Table S1). Individuals with missense variants were globally more severely affected than those with PTVs: 57% of individuals with a missense variant had severe/profound ID whereas this level of ID was not observed for any individuals with PTVs. Furthermore, hypotonia, spasticity and (severe) epilepsy were more common in individuals with missense variants than in those with PTVs (92% versus 42%, 42% versus 0%, 80% versus 18%, respectively) (Figure 1F, Table 1, Table S1A). To objectively quantify these observations, we divided our cohort into two variant-specific clusters (missense versus PTVs) and assessed the two groups using a Partitioning Around Medoids clustering algorithm[13] on 100 features derived from standardized clinical data (Human Phenotype Ontology (HPO); Figure S5A and Suppl. JSON)[14]. Thirty-eight individuals were subjected to this analysis, of which 27 were classified correctly as either belonging to the PTV or missense variant group (p=0.022), confirming the existence of at least two separate clinical entities (Figure 1G, Figure S5B). Moreover, computational averaging of facial photographs[15] revealed differences between the average facial gestalt for individuals with missense variants when compared to individuals with PTVs or deletions (Figure 1B-E, Figure S4, Table S1B).

We performed functional analyses assessing consequences of different types of *SATB1* variants for cellular localization, transcriptional activity, overall chromatin binding, and dimerization capacity. Based on protein modeling (Figure 2, Suppl. Notes), we selected five missense variants (observed in 14 individuals) in CUT1 and CUT2 affecting residues that interact with, or are close to, the DNA backbone (mosaic variant c.1220A>G; p.Glu407Gly and *de novo* variants c.1259A>G; p.Gln420Arg, c.1588G>A; p.Glu530Lys, c.1588G>C; p.Glu530Gln, c.1639G>A; p.Glu547Lys), as well as the only homeobox domain variant (c.2044C>G; p.Leu682Val, *de novo*). As controls, we selected three rare missense variants from the UK10K consortium, identified in healthy individuals with a normal IQ: c.1097C>T; p.Ser366Leu (gnomAD allele frequency 6.61e-4), c.1555G>C; p.Val519Leu (8.67e-6) and c.1717G>A; p.Ala573Thr (1.17e-4) (Figure 1A, Table S3)[16]. When overexpressed as YFP-fusion proteins in HEK293T/17 cells, wildtype SATB1 localized to the nucleus in a granular

pattern, with an intensity profile inverse to the DNA-binding dye Hoechst 33342 (Figure 3A-B). In contrast to wildtype and UK10K control missense variants, the p.Glu407Gly, p.Gln420Arg, p.Glu530Lys/p.Glu530Gln and p.Glu547Lys variants displayed a cage-like clustered nuclear pattern, strongly co-localizing with the DNA (Figure 3A-B, Figure S6).

To assess the effects of SATB1 missense variants on transrepressive activity, we used a luciferase reporter system with two previously established downstream targets of SATB1, the *IL2*-promoter and IgH-MAR (matrix associated region)[17-19]. All five functionally assessed CUT1 and CUT2 missense variants demonstrated increased transcriptional repression of the *IL2*-promoter, while the UK10K control variants did not differ from wildtype (Figure 3C). In assays using IgH-MAR, increased repression was seen for both CUT1 variants, and for one of the CUT2 variants (Figure 3C). The latter can be explained by previous reports that the CUT1 domain is essential for binding to MARs, whereas the CUT2 domain is dispensable[20; 21]. Taken together, these data suggest that etiological SATB1 missense variants in CUT1 and CUT2 lead to stronger binding of the transcription factor to its targets.

To study whether SATB1 missense variants affect the dynamics of chromatin binding more globally, we employed fluorescent recovery after photobleaching (FRAP) assays. Consistent with the luciferase reporter assays, all CUT1 and CUT2 missense variants, but not the UK10K control variants, affected protein mobility in the nucleus. The CUT2 variant p.Gln420Arg demonstrated an increased half time, but showed a maximum recovery similar to wildtype, while the other CUT1 and CUT2 variants demonstrated both increased halftimes and reduced maximum recovery. These results suggest stabilization of SATB1 chromatin binding for all tested CUT1 and CUT2 variants (Figure 3D).

In contrast to the CUT1 and CUT2 missense variants, the homeobox variant p.Leu682Val did not show functional differences from wildtype (Figure 3A-D, Figure S6), suggesting that, although it is absent from gnomAD, highly intolerant to variation and evolutionarily conserved (Figure S2, Figure S7A-B), this variant is unlikely to be pathogenic. This conclusion is further supported by the presence of a valine residue at the equivalent position in multiple homologous homeobox domains (Figure S7C). Additionally, the mild

phenotypic features in this individual (individual 42) can be explained by the fact that the individual carries an out-of-frame *de novo* intragenic duplication of *FOXP2*, known to cause NDD through haploinsufficiency[22].

We went on to assess the impact of the CUT1 and CUT2 missense variants (p.Glu407Gly, p.Gln420Arg, p.Glu530Lys, p.Glu547Lys) on protein interaction capacities using bioluminescence resonance energy transfer (BRET). All tested variants retained the ability to interact with wildtype SATB1 (Figure 3E), with the potential to yield dominant-negative dimers/tetramers *in vivo* and to disturb normal activity of the wildtype protein.

The identification of *SATB1* deletions suggests that haploinsufficiency is a second underlying disease mechanism. This is supported by the constraint of *SATB1* against loss-of-function variation, and the identification of PTV carriers that are clinically distinct from individuals with missense variants. PTVs are found throughout the locus and several are predicted to undergo NMD by *in silico* models of NMD efficacy (Table S4)[23]. In contrast to these predictions, we found that one of the PTVs, c.1228C>T; p.Arg410*, escapes NMD (Figure S8A-B). However, the p.Arg410* variant would lack critical functional domains (CUT1, CUT2, homeobox) and indeed showed reduced transcriptional activity in luciferase reporter assays when compared to wildtype protein (Figure S8), consistent with the haploinsufficiency model.

Four unique PTVs that we identified were located within the final exon of SATB1 (Figure 1A) and predicted to escape NMD (Table S4). Following experimental validation of NMD escape (Figure 4A-B), three such variants (c.1877delC; p.Pro626Hisfs*81, c.2080C>T; p.Gln694* and c.2207delA; p.Asn736Ilefs*8) were assessed with the same functional assays that we used for missense variants. When overexpressed as YFP-fusion proteins, the tested variants showed altered subcellular localization, forming nuclear puncta or (nuclear) aggregates, different from patterns observed for missense variants (Figure 4C, Figure S9A-B). In luciferase reporter assays, the p.Pro626Hisfs*81 variant showed increased repression of both the *IL2*-promoter and IgH-MAR, whereas p.Gln694* only showed reduced repression of IgH-MAR (Figure 4D). The p.Asn736Ilefs*8 variant showed repression comparable to that of wildtype protein for both targets (Figure 4D). In further pursuit of pathophysiological

mechanisms, we tested protein stability and SUMOylation, as the previously described p.Lys744 SUMOylation site is missing in all assessed NMD-escaping truncated proteins (Figure 4A)[24]. Our observations suggest the existence of multiple SATB1 SUMOylation sites (Figure S10) and no effect of NMD-escaping variants on SUMOylation of the encoded proteins (Figure S10) nor any changes in protein stability (Figure S9C). Although functional assays with NMD-escaping PTVs hint towards additional disease mechanisms, HPO-based phenotypic analysis or qualitative evaluation could not confirm a third distinct clinical entity (p=0.932; Figure S5, Figure S11, Table S5).

Our study demonstrates that while statistical analyses[5; 6] can provide the first step towards identification of new NDDs, a mutation-specific functional follow-up is required to gain insight into the underlying mechanisms and to understand phenotypic differences within patient cohorts (Table S6). Multiple mechanisms and/or more complex genotype-phenotype correlations are increasingly appreciated in newly described NDDs, such as those associated with *RAC1, POL2RA, KMT2E* and *PPP2CA*[25-28]. Interestingly, although less often explored, such mechanistic complexity might also underlie well-known (clinically recognizable) NDDs. For instance, a CUT1 missense variant in *SATB2*, a paralog of *SATB1* that causes Glass syndrome through haploinsufficiency (MIM #612313)[29], affects protein localization and nuclear mobility in a similar manner to the corresponding *SATB1* missense variants (Figure S12, Figure S13)[30]. Taken together, these observations suggest that mutation-specific mechanisms await discovery both for new and well-established clinical syndromes.

In summary, we demonstrate that at least two different previously uncharacterized NDDs are caused by distinct classes of rare *(de novo)* variation at a single locus. We combined clinical investigation, *in silico* models and cellular assays to characterize the phenotypic consequences and functional impacts of a large patient series uncovering distinct pathophysiological mechanisms of the *SATB1*-associated NDDs. This level of combined analyses is recommended for known and yet undiscovered NDDs to fully understand disease etiology.

Acknowledgements

We are extremely grateful to all families participating in this study. In addition, we wish to thank the members of the Genome Technology Center and Cell culture facility, Department of Human Genetics, Radboud university medical center, Nijmegen, for data processing and cell culture of patient-derived cell lines. This work was financially supported by Aspasia grants of the Dutch Research Council (015.014.036 to TK and 015.014.066 to LV), Netherlands Organization for Health Research and Development (91718310 to TK), the Max Planck Society (JdH, SF), Oxford Brookes University, the Leverhulme Trust and the British Academy (DN), and grants from the Swiss National Science Foundation (31003A_182632 to AR), Lithuanian-Swiss cooperation program to reduce economic and social disparities within the enlarged European Union (AR, VK) and the Jérôme Lejeune Foundation (AR). We wish to acknowledge ALSPAC, the UK10K consortium, the 100,000 Genomes Project, “TRANSLATE NAMSE” and Genomic Answers for Kids program (see Suppl. Acknowledgements). In addition, the collaborations in this study were facilitated by ERN ITHACA, one of the 24 European Reference Networks (ERNs) approved by the ERN Board of Member States, co-funded by European Commission. The aims of this study contribute to the Solve-RD project (EdB, HB, SB, ASDP, LF, CG, AJ, TK, AV, LV) which has received funding from the European Union’s Horizon 2020 research and innovation programme under grant agreement No 779257.

Conflict of interest

KM, TBP, and TSS are employees of GeneDx, Inc. KR is employee of Ambrygen Genetics.

References

1. Wang, Z., Yang, X., Chu, X., Zhang, J., Zhou, H., Shen, Y., and Long, J. (2012). The structural basis for the oligomerization of the N-terminal domain of SATB1. Nucleic Acids Res 40, 4193-4202.

2. Alvarez, J.D., Yasui, D.H., Niida, H., Joh, T., Loh, D.Y., and Kohwi-Shigematsu, T. (2000). The MAR-binding protein SATB1 orchestrates temporal and spatial expression of multiple genes during T-cell development. Genes Dev 14, 521-535.

3. Cai, S., Lee, C.C., and Kohwi-Shigematsu, T. (2006). SATB1 packages densely looped, transcriptionally active chromatin for coordinated expression of cytokine genes. Nat Genet 38, 1278-1288.

4. Kitagawa, Y., Ohkura, N., Kidani, Y., Vandenbon, A., Hirota, K., Kawakami, R., Yasuda, K., Motooka, D., Nakamura, S., Kondo, M., et al. (2017). Guidance of regulatory T cell development by Satb1-dependent super-enhancer establishment. Nat Immunol 18, 173-183.

5. Satterstrom, F.K., Kosmicki, J.A., Wang, J., Breen, M.S., De Rubeis, S., An, J.Y., Peng, M., Collins, R., Grove, J., Klei, L., et al. (2020). Large-Scale Exome Sequencing Study Implicates Both Developmental and Functional Changes in the Neurobiology of Autism. Cell 180, 568-584.e523.

6. Kaplanis, J., Samocha, K.E., Wiel, L., Zhang, Z., Arvai, K.J., Eberhardt, R.Y., Gallone, G., Lelieveld, S.H., Martin, H.C., McRae, J.F., et al. (2020). Evidence for 28 genetic disorders discovered by combining healthcare and research data. Nature.

7. 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.

8. Thompson, R., Johnston, L., Taruscio, D., Monaco, L., Beroud, C., Gut, I.G., Hansson, M.G., t Hoen, P.B., Patrinos, G.P., Dawkins, H., et al. (2014). RD-Connect: an integrated platform connecting databases, registries, biobanks and clinical bioinformatics for rare disease research. J Gen Intern Med 29 Suppl 3, S780-787.

9. Firth, H.V., Richards, S.M., Bevan, A.P., Clayton, S., Corpas, M., Rajan, D., Van Vooren, S., Moreau, Y., Pettett, R.M., and Carter, N.P. (2009). DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. Am J Hum Genet 84, 524-533.

364 10. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins,
365 R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational
366 constraint spectrum quantified from variation in 141,456 humans. Nature 581, 434-
367 443.
368 11. Lelieveld, S.H., Reijnders, M.R., Pfundt, R., Yntema, H.G., Kamsteeg, E.J., de Vries, P.,
369 de Vries, B.B., Willemsen, M.H., Kleefstra, T., Lohner, K., et al. (2016). Meta-analysis
370 of 2,104 trios provides support for 10 new genes for intellectual disability. Nat
371 Neurosci 19, 1194-1196.
372 12. Lelieveld, S.H., Wiel, L., Venselaar, H., Pfundt, R., Vriend, G., Veltman, J.A., Brunner,
373 H.G., Vissers, L., and Gilissen, C. (2017). Spatial Clustering of de Novo Missense
374 Mutations Identifies Candidate Neurodevelopmental Disorder-Associated Genes. Am
375 J Hum Genet 101, 478-484.
376 13. Kaufman L., R.P.J. (1987). Clustering by means of medoids
377 https://wis.kuleuven.be/stat/robust/papers/publications-1987/kaufmanrousseeuw-
378 clusteringbymedoids-l1norm-1987.pdf.
379 14. Köhler, S., Carmody, L., Vasilevsky, N., Jacobsen, J.O.B., Danis, D., Gourdine, J.P.,
380 Gargano, M., Harris, N.L., Matentzoglu, N., McMurry, J.A., et al. (2019). Expansion of
381 the Human Phenotype Ontology (HPO) knowledge base and resources. Nucleic
382 Acids Res 47, D1018-d1027.
383 15. Reijnders, M.R.F., Miller, K.A., Alvi, M., Goos, J.A.C., Lees, M.M., de Burca, A.,
384 Henderson, A., Kraus, A., Mikat, B., de Vries, B.B.A., et al. (2018). De Novo and
385 Inherited Loss-of-Function Variants in TLK2: Clinical and Genotype-Phenotype
386 Evaluation of a Distinct Neurodevelopmental Disorder. Am J Hum Genet 102, 1195-
387 1203.
388 16. Walter, K., Min, J.L., Huang, J., Crooks, L., Memari, Y., McCarthy, S., Perry, J.R., Xu, C.,
389 Futema, M., Lawson, D., et al. (2015). The UK10K project identifies rare variants in
390 health and disease. Nature 526, 82-90.

17. Pavan Kumar, P., Purbey, P.K., Sinha, C.K., Notani, D., Limaye, A., Jayani, R.S., and Galande, S. (2006). Phosphorylation of SATB1, a global gene regulator, acts as a molecular switch regulating its transcriptional activity in vivo. Mol Cell 22, 231-243.
18. Kumar, P.P., Purbey, P.K., Ravi, D.S., Mitra, D., and Galande, S. (2005). Displacement of SATB1-bound histone deacetylase 1 corepressor by the human immunodeficiency virus type 1 transactivator induces expression of interleukin-2 and its receptor in T cells. Mol Cell Biol 25, 1620-1633.
19. Siebenlist, U., Durand, D.B., Bressler, P., Holbrook, N.J., Norris, C.A., Kamoun, M., Kant, J.A., and Crabtree, G.R. (1986). Promoter region of interleukin-2 gene undergoes chromatin structure changes and confers inducibility on chloramphenicol acetyltransferase gene during activation of T cells. Mol Cell Biol 6, 3042-3049.
20. Ghosh, R.P., Shi, Q., Yang, L., Reddick, M.P., Nikitina, T., Zhurkin, V.B., Fordyce, P., Stasevich, T.J., Chang, H.Y., Greenleaf, W.J., et al. (2019). Satb1 integrates DNA binding site geometry and torsional stress to differentially target nucleosome-dense regions. Nat Commun 10, 3221.
21. Dickinson, L.A., Dickinson, C.D., and Kohwi-Shigematsu, T. (1997). An atypical homeodomain in SATB1 promotes specific recognition of the key structural element in a matrix attachment region. J Biol Chem 272, 11463-11470.
22. MacDermot, K.D., Bonora, E., Sykes, N., Coupe, A.M., Lai, C.S., Vernes, S.C., Vargha-Khadem, F., McKenzie, F., Smith, R.L., Monaco, A.P., et al. (2005). Identification of FOXP2 truncation as a novel cause of developmental speech and language deficits. Am J Hum Genet 76, 1074-1080.
23. Lindeboom, R.G.H., Vermeulen, M., Lehner, B., and Supek, F. (2019). The impact of nonsense-mediated mRNA decay on genetic disease, gene editing and cancer immunotherapy. Nat Genet 51, 1645-1651.
24. Tan, J.A., Sun, Y., Song, J., Chen, Y., Krontiris, T.G., and Durrin, L.K. (2008). SUMO conjugation to the matrix attachment region-binding protein, special AT-rich

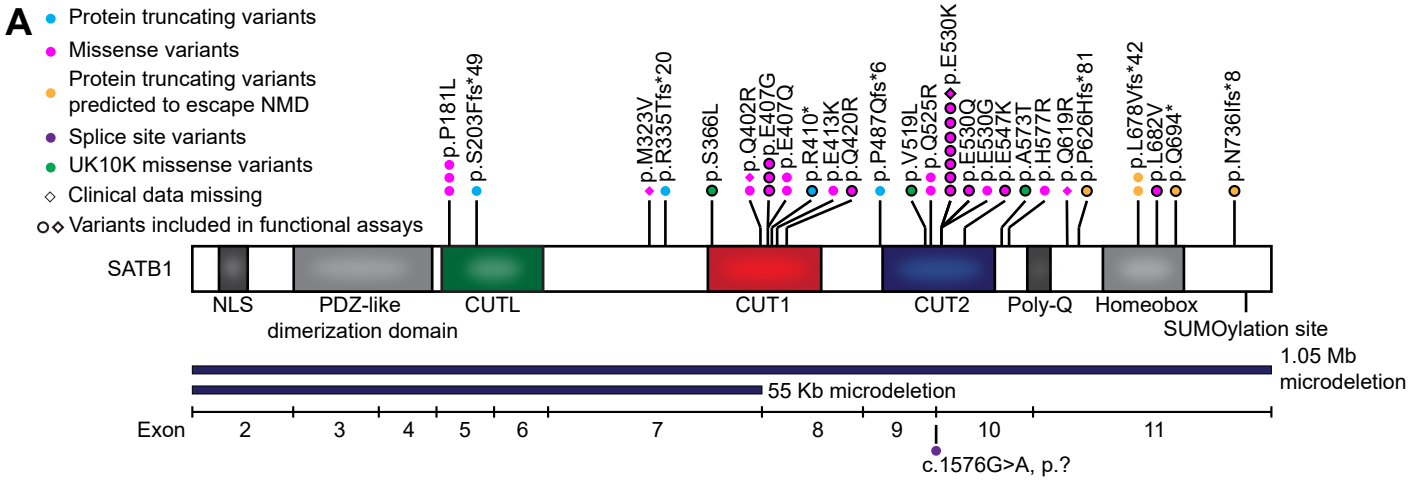
418 sequence-binding protein-1 (SATB1), targets SATB1 to promyelocytic nuclear bodies
419 where it undergoes caspase cleavage. J Biol Chem 283, 18124-18134.
420 25. Haijes, H.A., Koster, M.J.E., Rehmann, H., Li, D., Hakonarson, H., Cappuccio, G.,
421 Hancarova, M., Lehalle, D., Reardon, W., Schaefer, G.B., et al. (2019). De Novo
422 Heterozygous POLR2A Variants Cause a Neurodevelopmental Syndrome with
423 Profound Infantile-Onset Hypotonia. Am J Hum Genet 105, 283-301.
424 26. O'Donnell-Luria, A.H., Pais, L.S., Faundes, V., Wood, J.C., Sveden, A., Luria, V., Abou
425 Jamra, R., Accogli, A., Amburgey, K., Anderlid, B.M., et al. (2019). Heterozygous
426 Variants in KMT2E Cause a Spectrum of Neurodevelopmental Disorders and
427 Epilepsy. Am J Hum Genet 104, 1210-1222.
428 27. Reynhout, S., Jansen, S., Haesen, D., van Belle, S., de Munnik, S.A., Bongers, E.,
429 Schieving, J.H., Marcelis, C., Amiel, J., Rio, M., et al. (2019). De Novo Mutations
430 Affecting the Catalytic Calpha Subunit of PP2A, PPP2CA, Cause Syndromic
431 Intellectual Disability Resembling Other PP2A-Related Neurodevelopmental
432 Disorders. Am J Hum Genet 104, 139-156.
433 28. Reijnders, M.R.F., Ansor, N.M., Kousi, M., Yue, W.W., Tan, P.L., Clarkson, K., Clayton-
434 Smith, J., Corning, K., Jones, J.R., Lam, W.W.K., et al. (2017). RAC1 Missense
435 Mutations in Developmental Disorders with Diverse Phenotypes. Am J Hum Genet
436 101, 466-477.
437 29. Zarate, Y.A., Bosanko, K.A., Caffrey, A.R., Bernstein, J.A., Martin, D.M., Williams, M.S.,
438 Berry-Kravis, E.M., Mark, P.R., Manning, M.A., Bhambhani, V., et al. (2019). Mutation
439 update for the SATB2 gene. Hum Mutat 40, 1013-1029.
440 30. Lee, J.S., Yoo, Y., Lim, B.C., Kim, K.J., Choi, M., and Chae, J.H. (2016). SATB2-
441 associated syndrome presenting with Rett-like phenotypes. Clin Genet 89, 728-732.

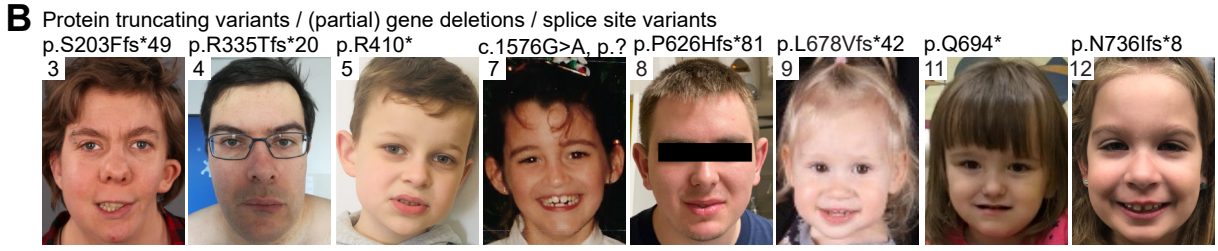
442

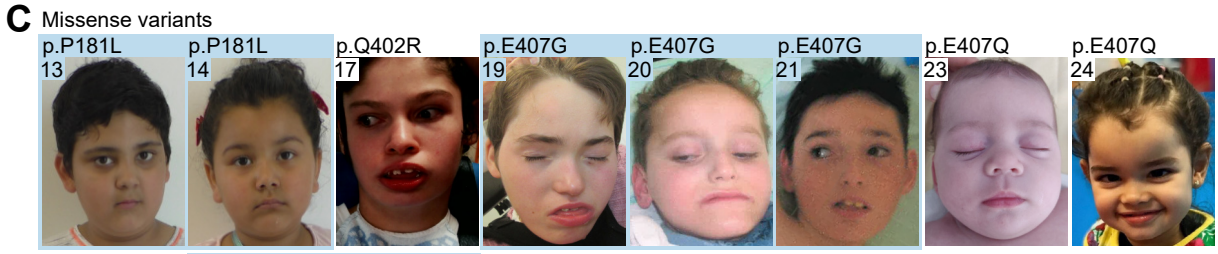
Figure legends

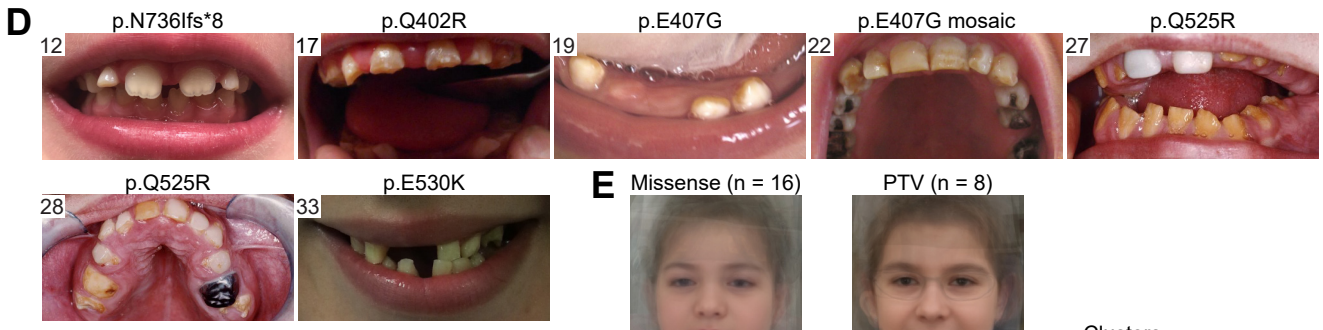
Figure 1. Clinical evaluation of *SATB1* variants in neurodevelopmental disorders. A) Schematic representation of the SATB1 protein (NM_001131010.4/NP_001124482.1), including functional domains, with truncating variants labeled in cyan, truncating variants predicted to escape NMD in orange, splice site variants in purple, missense variants in magenta, and UK10K rare control missense variants in green. Deletions are shown in dark blue below the protein schematic, above a diagram showing the exon boundaries. We obtained clinical data for all individuals depicted by a circle. **B-C**) Facial photographs of individuals with (partial) gene deletions and truncations (**B**), and of individuals with missense variants (**C**). All depicted individuals show facial dysmorphisms and although overlapping features are seen, no consistent facial phenotype can be observed for the group as a whole. Overlapping facial dysmorphisms include facial asymmetry, high forehead, prominent ears, straight and/or full eyebrows, puffy eyelids, downslant of palpebral fissures, low nasal bridge, full nasal tip and full nasal alae, full lips with absent cupid's bow, prominent cupid's bow or thin upper lip vermillion (Table S1B). Individuals with missense variants are more alike than individuals in the truncating cohorts, and we observed recognizable overlap between several individuals in the missense cohort (individual 17, 27, 31, 37, the siblings 19, 20 and 21, and to a lesser extent individual 24 and 35). A recognizable facial overlap between individuals with (partial) gene deletions and truncations could not be observed. Related individuals are marked with a blue box. **D**) Photographs of teeth abnormalities observed in individuals with *SATB1* variants. Dental abnormalities are seen for all variant types and include widely spaced teeth, dental fragility, missing teeth, disorganized teeth implant, and enamel discoloration (Table S1B). **E**) Computational average of facial photographs of 16 individuals with a missense variant (left) and 8 individuals with PTVs or (partial) gene deletions (right). **F**) Mosaic plot presenting a selection of clinical features. **G**) The Partitioning Around Medoids analysis of clustered HPO-standardized clinical data from 38 individuals with truncating (triangle) and missense variants (circle) shows a significant distinction between the clusters of individuals with missense variants (blue) and individuals with PTVs (red). Applying Bonferroni correction, a *p*-value

smaller than 0.025 was considered significant. For analyses displayed in (F) and (G), individuals with absence of any clinical data and/or low level mosaicism for the *SATB1* variant were omitted (for details, see Suppl. Materials and Methods).

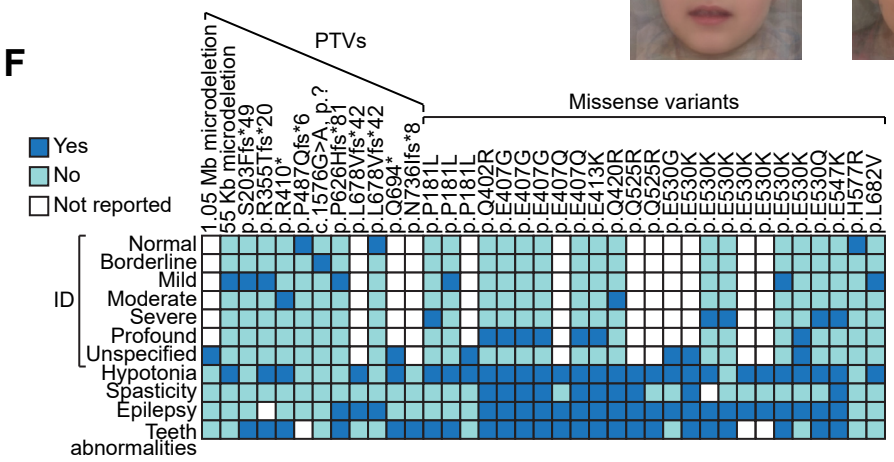

A
Protein truncating variants
Missense variants
Protein truncating variants predicted to escape NMD
Splice site variants
UK10K missense variants
Clinical data missing
Variants included in functional assays
p.P181L
p.S203Ffs*49
p.M323V
p.R335Tfs*20
p.S366L
p.Q402R
p.E407G
p.E407Q
p.R410*
p.E413K
p.Q420R
p.P487Qfs*6
p.V519L
p.Q525R
p.E530K
p.E530Q
p.E530G
p.E547K
p.A573T
p.H577R
p.Q619R
p.P626Hfs*81
p.L678Vfs*42
p.L682V
p.Q694*
p.N736Ifs*8
SATB1
NLS
PDZ-like dimerization domain
CUTL
CUT1
CUT2
Poly-Q
Homeobox
SUMOylation site
1.05 Mb microdeletion
55 Kb microdeletion
Exon 2 3 4 5 6 7 8 9 10 11
c.1576G>A, p.?


B
Protein truncating variants / (partial) gene deletions / splice site variants
p.S203Ffs*49
p.R335Tfs*20
p.R410*
c.1576G>A, p.?
p.P626Hfs*81
p.L678Vfs*42
p.Q694*
p.N736Ifs*8
3
4
5
7
8
9
11
12


C
Missense variants
p.P181L
p.P181L
p.Q402R
p.E407G
p.E407G
p.E407G
p.E407Q
p.E407Q
13
14
17
19
20
21
23
24
p.Q420R
p.Q525R
p.Q525R
p.E530K
p.E530K
p.E530K
p.E530K
p.E530K
26
27
28
31
33
34
35
37


D
p.N736Ifs*8
p.Q402R
p.E407G
p.E407G mosaic
p.Q525R
12
17
19
22
27
p.Q525R
p.E530K
28
33

E
Missense (n = 16)
PTV (n = 8)


F
PTVs
Missense variants
Yes
No
Not reported
1.05 Mb microdeletion
55 Kb microdeletion
p.S203Ffs*49
p.R335Tfs*20
p.R410*
p.P487Qfs*6
c.1576G>A, p.?
p.P626Hfs*81
p.L678Vfs*42
p.L678Vfs*42
p.Q694*
p.N736Ifs*8
p.P181L
p.P181L
p.Q402R
p.E407G
p.E407G
p.E407Q
p.E407Q
p.E413K
p.Q420R
p.Q525R
p.E530G
p.E530K
p.E530K
p.E530K
p.E530K
p.E530K
p.E530Q
p.E547K
p.H577R
p.L682V
ID
Normal
Borderline
Mild
Moderate
Severe
Profound
Unspecified
Hypotonia
Spasticity
Epilepsy
Teeth abnormalities

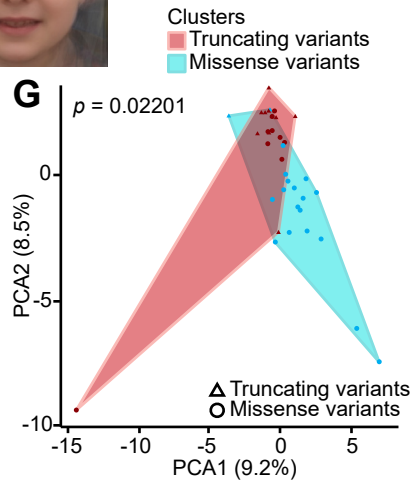

G
Clusters
Truncating variants
Missense variants
p = 0.02201
PCA2 (8.5%)
PCA1 (9.2%)
0
-5
-10
-15
-10
-5
0
5
Truncating variants
Missense variants

Figure 2. 3D protein modeling of SATB1 missense variants in DNA-binding domains. A) Schematic representation of the aligned CUT1 and CUT2 DNA-binding domains. CUT1 and CUT2 domains have a high sequence identity (40%) and similarity (78%). Note that the recurrent p.Q402R, p.E407G/p.E407Q and p.Q525R, p.E530G/p.E530K/p.E530Q variants affect equivalent positions within the respective CUT1 and CUT2 domains, while p.Q420R in CUT1 and p.E547K in CUT2 affect cognate regions. **B**) 3D-model of the SATB1 CUT1 domain (left; PDB 2O4A) and CUT2 domain (right; based on PDB 2CSF) in interaction with DNA (yellow). Mutated residues are highlighted in red for CUT1 and cyan for CUT2, along the ribbon visualization of the corresponding domains in burgundy and dark blue, respectively. **C**) 3D-homology model of the SATB1 homeobox domain (based on PDB 1WI3 and 2D5V) in interaction with DNA (yellow). The mutated residue is shown in light gray along the ribbon visualization of the corresponding domain in dark gray. **B-C**) For more detailed descriptions of the different missense variants in our cohort, see Suppl. Notes.

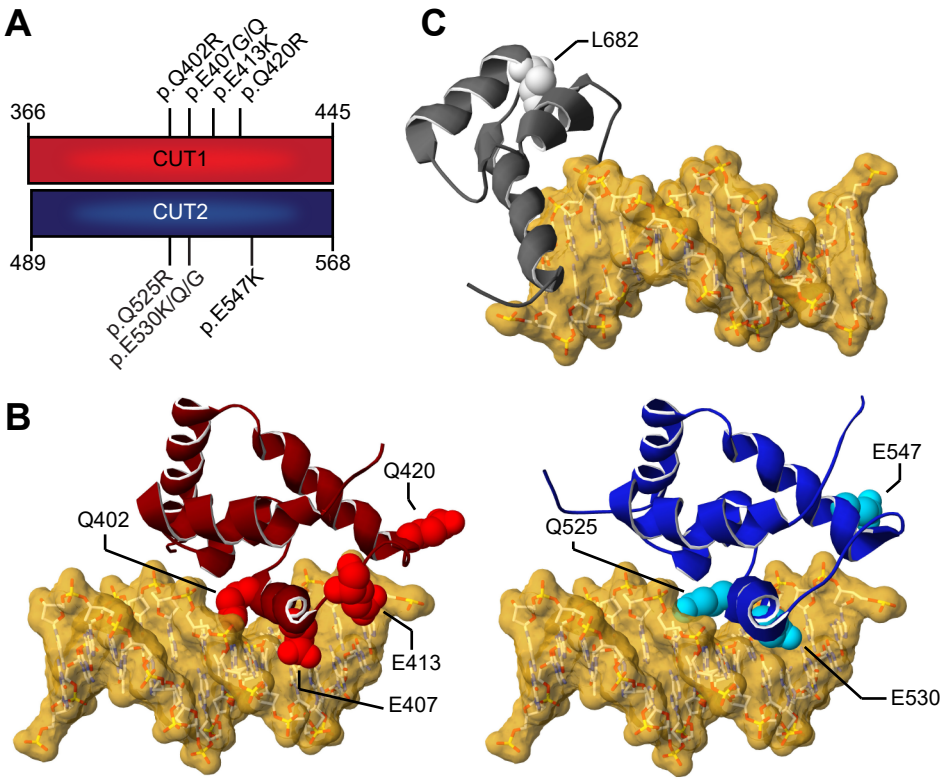

A
p.Q402R
p.E407G/Q
p.E413K
p.Q420R
366
445
CUT1
CUT2
489
568
p.Q525R
p.E530K/Q/G
p.E547K
C
L682
B
Q420
Q402
E413
E407
E547
Q525
E530

Figure 3. SATB1 missense variants stabilize DNA binding and show increased transcriptional repression. A) Direct fluorescence super-resolution imaging of nuclei of HEK293T/17 cells expressing YFP-SATB1 fusion proteins. Scale bar = 5 µm. **B**) Intensity profiles of YFP-tagged SATB1 and variants, and the DNA binding dye Hoechst 33342. The graphs represent the fluorescence intensity values of the position of the red lines drawn in the micrographs on the top (SATB1 proteins in green, Hoechst 33342 in white, scale bar = 5 µm). For each condition a representative image and corresponding intensity profile plot is shown. **C**) Luciferase reporter assays using reporter constructs containing the *IL2*-promoter region and the IgH matrix associated region (MAR) binding site. UK10K control variants are shaded in green, CUT1 domain variants in red, CUT2 domain variants in blue and the homeobox variant in gray. Values are expressed relative to the control (pYFP; black) and represent the mean ± S.E.M. (*n* = 4, *p*-values compared to wildtype SATB1 (WT; white), one-way ANOVA and *post-hoc* Bonferroni test). **D**) FRAP experiments to assess the dynamics of SATB1 chromatin binding in live cells. Left, mean recovery curves ± 95% C.I. recorded in HEK293T/17 cells expressing YFP-SATB1 fusion proteins. Right, violin plots with median of the halftime (central panel) and maximum recovery values (right panel) based on single-term exponential curve fitting of individual recordings (*n* = 60 nuclei from three independent experiments, *p*-values compared to WT SATB1, one-way ANOVA and *post-hoc* Bonferroni test). Color code as in C. **E**) BRET assays for SATB1 dimerization in live cells. Left, mean BRET saturation curves ± 95% C.I. fitted using a non-linear regression equation assuming a single binding site (y = BRETmax * x / (BRET50 / x); GraphPad). The corrected BRET ratio is plotted against the ratio of fluorescence/luminescence (AU) to correct for expression level differences between conditions. Right, corrected BRET ratio values at mean BRET50 level of WT SATB1, based on curve fitting of individual experiments (*n* = 4, one-way ANOVA and *post-hoc* Bonferroni test, no significant differences). Color code as in C. **A-E**) When compared to WT YFP-SATB1 or UK10K variants, most variants identified in affected individuals show a nuclear cage-like localization (**A**), stronger co-localization with the DNA-binding dye Hoechst 33342 (**B**),

increased transcriptional repression (**C**), reduced protein mobility (**D**) and unchanged capacity of interaction with WT SATB1 (**E**).

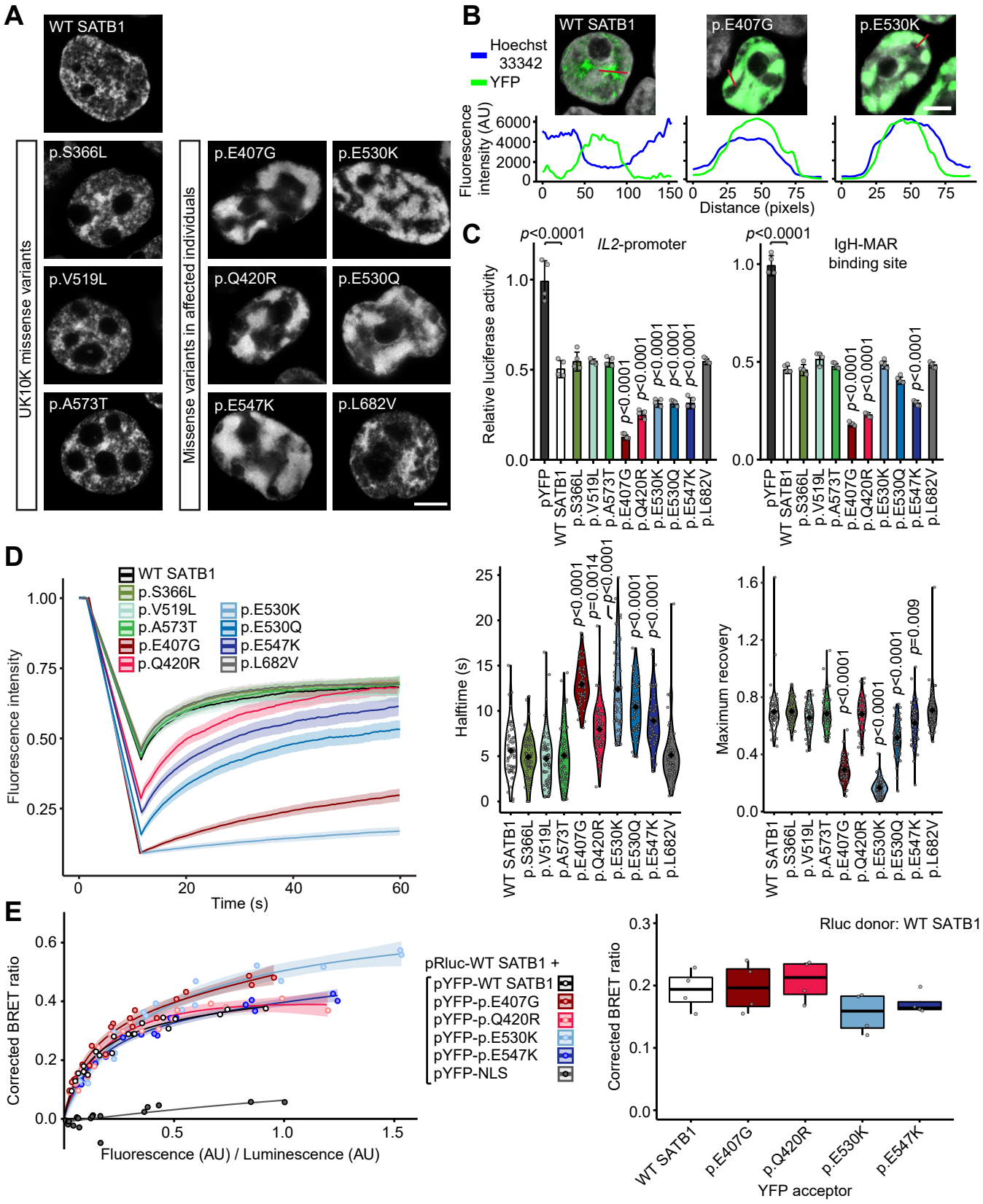
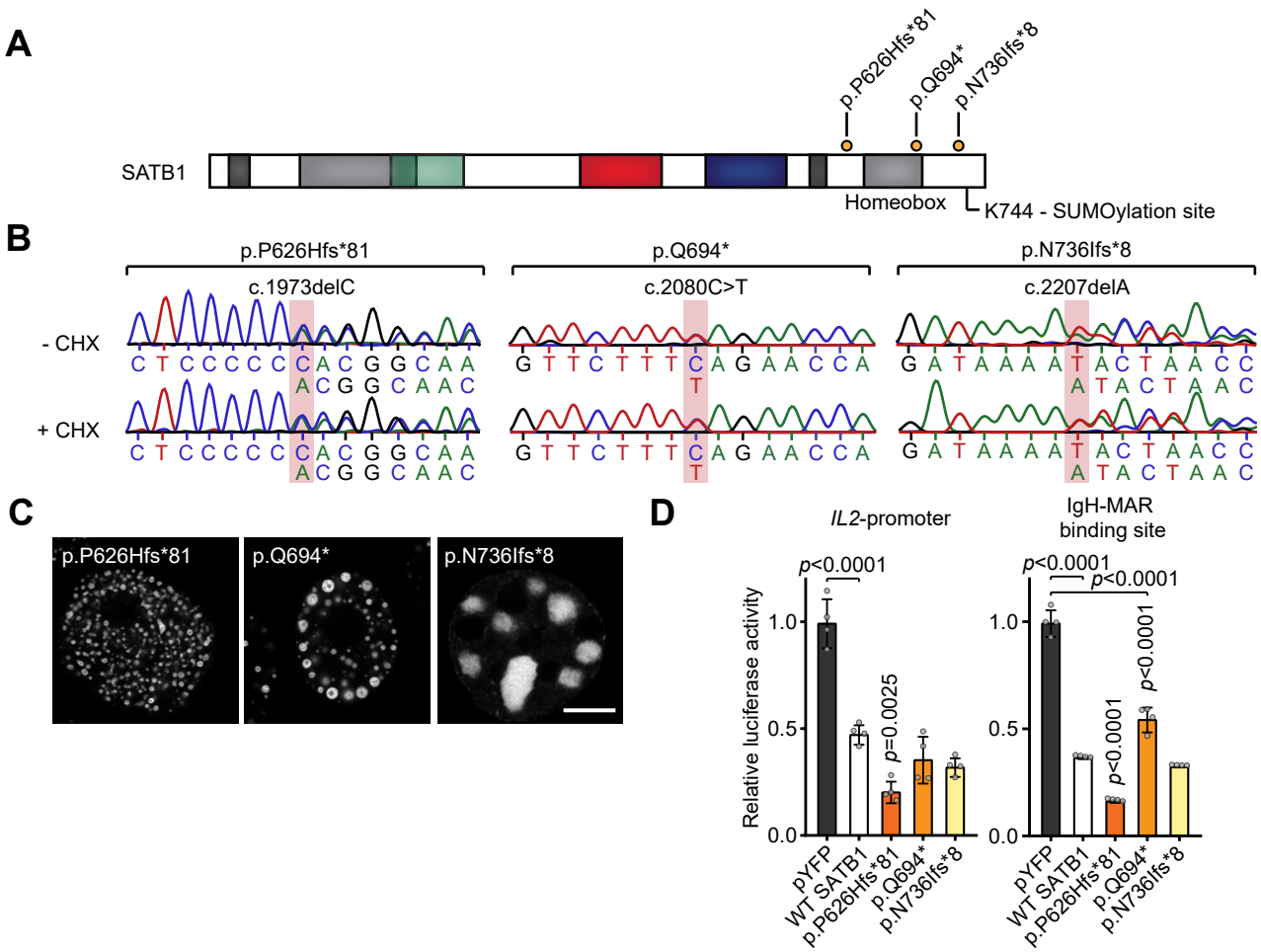

A
WT SATB1
UK10K missense variants
p.S366L
p.V519L
p.A573T
Missense variants in affected individuals
p.E407G
p.E530K
p.Q420R
p.E530Q
p.E547K
p.L682V
B
WT SATB1
p.E407G
p.E530K
Hoechst 33342
YFP
Fluorescence intensity (AU)
Distance (pixels)
C
p<0.0001
IL2-promoter
IgH-MAR binding site
Relative luciferase activity
pYFP
WT SATB1
p.S366L
p.V519L
p.A573T
p.E407G
p.Q420R
p.E530K
p.E530Q
p.E547K
p.L682V
D
Fluorescence intensity
Time (s)
Halftime (s)
Maximum recovery
E
Corrected BRET ratio
Fluorescence (AU) / Luminescence (AU)
pRluc-WT SATB1 +
pYFP-WT SATB1
pYFP-p.E407G
pYFP-p.Q420R
pYFP-p.E530K
pYFP-p.E547K
pYFP-NLS
Rluc donor: WT SATB1
YFP acceptor

Figure 4. *SATB1* frameshift variants in the last exon escape NMD. A) Schematic overview of the SATB1 protein, with truncating variants predicted to escape NMD that are included in functional assays labeled in orange. A potential SUMOylation site at position p.K744 is highlighted. **B**) Sanger sequencing traces of patient-derived EBV-immortalized lymphoblastoid cell lines treated with or without cycloheximide (CHX) to test for NMD. The mutated nucleotides are shaded in red. Transcripts from both alleles are present in both conditions showing that these variants escape NMD. **C**) Direct fluorescence super-resolution imaging of nuclei of HEK293T/17 cells expressing SATB1 truncating variants fused with a YFP-tag. Scale bar = 5 µm. Compared to WT YFP-SATB1, NMD-escaping variants show altered localization forming nuclear puncta or aggregates. **D**) Luciferase reporter assays using reporter constructs containing the *IL2*-promoter and the IgH matrix associated region (MAR) binding site. Values are expressed relative to the control (pYFP; black) and represent the mean ± S.E.M. (n = 4, p-values compared to WT SATB1 (white), one-way ANOVA and *post-hoc* Bonferroni test). All NMD-escaping variants are transcriptionally active and show repression of the *IL2*-promoter and IgH-MAR binding site.


A
p.P626Hfs*81
p.Q694*
p.N736Ifs*8
SATB1
Homeobox
K744 - SUMOylation site
B
p.P626Hfs*81
p.Q694*
p.N736Ifs*8
c.1973delC
c.2080C>T
c.2207delA
- CHX
+ CHX
C
p.P626Hfs*81
p.Q694*
p.N736Ifs*8
D
IL2-promoter
IgH-MAR binding site
Relative luciferase activity
p<0.0001
p=0.0025
p<0.0001
p<0.0001
p<0.0001
p<0.0001
1.0
0.5
0.0
pYFP
WT SATB1
p.P626Hfs*81
p.Q694*
p.N736Ifs*8

595 **Table 1. Summary of clinical characteristics associated with (*de novo*) *SATB1* variants**

	All individuals		Individuals with PTVs and (partial) gene deletions		Individuals with missense variants	
	%	Present / total assessed	%	Present / total assessed	%	Present / total assessed
Neurologic						
Intellectual disability	90	28/31	80	8/10	95	20/21
Normal	10	3/31	20	2/10	5	1/21
Borderline	0	0/31	0	0/10	0	0/21
Mild	26	8/31	60	6/10	10	2/21
Moderate	10	3/31	10	1/10	10	2/21
Severe	19	6/31	0	0/10	29	6/21
Profound	19	6/31	0	0/10	29	6/21
Unspecified	16	5/31	10	1/10	19	4/21
Developmental delay	97	35/36	100	12/12	96	23/24
Motor delay	92	34/37	92	11/12	92	23/25
Speech delay	89	32/36	83	10/12	92	22/24
Dysarthria	30	6/20	9	1/11	56	5/9
Epilepsy	61	22/36	18	2/11	80	20/25
EEG abnormalities	79	19/24	29	2/7	100	17/17
Hypotonia	76	28/37	42	5/12	92	23/25
Spasticity	28	10/36	0	0/12	42	10/24
Ataxia	22	6/27	17	2/12	27	4/15
Behavioral disturbances	71	24/34	58	7/12	77	17/22
Sleep disturbances	41	12/29	27	3/11	50	9/18
Abnormal brain imaging	55	17/31	43	3/7	58	14/24
Regression	17	6/35	8	1/12	22	5/23
Growth						
Abnormalities during pregnancy	24	8/33	27	3/11	23	5/22
Abnormalities during delivery	32	10/31	55	6/11	20	4/20
Abnormal term of delivery	6	2/31	10	1/10	5	1/21
Preterm (<37 weeks)	6	2/31	10	1/10	5	1/21
Postterm (>42 weeks)	0	0/31	0	0/10	0	0/21
Abnormal weight at birth	16	5/32	22	2/9	13	3/23
Small for gestational age (<p10)	9	3/32	11	1/9	9	2/23
Large for gestational age (>p90)	6	2/32	11	1/9	4	1/23
Abnormal head circumference at birth	7	1/14	17	1/6	0	0/8
Microcephaly (<p3)	0	0/14	0	0/6	0	0/8
Macrocephaly (>p97)	7	1/14	17	1/6	0	0/8
Abnormal height	21	6/29	9	1/11	28	5/18
Short stature (<p3)	14	4/29	0	0/11	22	4/18
Tall stature (>p97)	7	2/29	9	1/11	6	1/18
Abnormal head circumference	26	7/31	11	1/9	32	6/22
Microcephaly (<p3)	26	7/31	11	1/9	32	6/22
Macrocephaly (>p97)	0	0/31	0	0/9	0	0/22
Abnormal weight	48	13/27	11	1/9	67	12/18
Underweight (<p3)	22	6/27	11	1/9	28	5/18
Overweight (>p97)	26	7/27	0	0/9	39	7/18
Other phenotypic features						
Facial dysmorphisms	67	24/36	64	7/11	68	17/25
Dental/oral abnormalities	71	24/34	55	6/11	78	18/23
Drooling/dysphagia	38	12/32	25	3/12	45	9/20
Hearing abnormalities	7	2/30	18	2/11	0	0/19
Vision abnormalities	55	17/31	73	8/11	45	9/20
Cardiac abnormalities	19	6/32	27	3/11	14	3/21
Skeleton/limb abnormalities	38	13/34	18	2/11	48	11/23
Hypermobility of joints	30	8/27	30	3/10	29	5/17
Gastrointestinal abnormalities	53	17/32	27	3/11	67	14/21
Urogenital abnormalities	17	5/30	0	0/11	26	5/19
Endocrine/metabolic abnormalities	30	9/30	0	0/11	47	9/19
Immunological abnormalities	32	8/25	25	2/8	35	6/17
Skin/hair/nail abnormalities	24	8/34	9	1/11	30	7/23
Neoplasms in medical history	0	0/34	0	0/11	0	0/23