

1 **Title**

2 *SETBP1* variants outside the degron disrupt DNA-binding

3 and transcription independent of protein abundance to

4 cause a heterogeneous neurodevelopmental disorder

5 **Authors and Affiliations**

6 Maggie MK Wong,^{1,*} Rosalie A Kampen,¹ Ruth O Braden,^{2,3} Gökberk Alagöz,¹ Michael S

7 Hildebrand,^{2,4} Christopher Barnett,⁵ Meghan Barnett,⁵ Alfredo Brusco,^{6,7} Diana Carli,^{8,9} Bert

8 BA de Vries,¹⁰ Alexander JM Dingemans,¹⁰ Frances Elmslie,¹¹ Giovanni B Ferrero,¹² Nadieh

9 A Jansen,¹⁰ Ingrid MBH van de Laar,¹³ Alice Moroni,⁶ David Mowat,^{14,15} Lucinda Murray,¹⁶

10 Francesca Novara,¹⁷ Angela Peron,^{18,19,20} Ingrid E Scheffer,² Fabio Sirchia,²¹ Samantha J

11 Turner,^{2,3} Aglaia Vignoli,²² Arianna Vino,¹ Sacha Weber,²³ Wendy K Chung,^{24,25} Marion

12 Gerard,²³ Vanessa López-González,²⁶ Elizabeth Palmer,^{15,16} Angela T Morgan,^{3,27} Bregje W

13 van Bon,¹⁰ Simon E Fisher^{1,28,*}

14 ¹Language and Genetics Department, Max Planck Institute for Psycholinguistics, 6500AH

15 Nijmegen, the Netherlands.

16 ²Epilepsy Research Centre, Department of Medicine and Paediatrics, University of

17 Melbourne Austin Health Victoria, 3010, Australia.

18 ³Speech and Language, Murdoch Children's Research Institute, Melbourne, Victoria, 3052,

19 Australia.

20 ⁴Neuroscience Group, Murdoch Children's Research Institute, Royal Children's Hospital,

21 Melbourne, Victoria, 3052, Australia.

22 ⁵SA Clinical Genetics Service, Women's and Children's Hospital, North Adelaide 5006,

23 South Australia.

24 ⁶Department of Medical Sciences, University of Torino, 10126 Torino, Italy.

25 ⁷Medical Genetics Unit, Città della Salute e della Scienza, University Hospital, 10126 Turin,
26 Italy.

27 ⁸Department of Public Health and Pediatric Sciences, University of Torino, 10126 Torino,
28 Italy.

29 ⁹Pediatric Onco-Hematology, Stem Cell Transplantation and Cell Therapy Division, Regina
30 Margherita Children's Hospital, Città della Salute e della Scienza di Torino, 10126 Torino,
31 Italy.

32 ¹⁰Department of Human Genetics, Radboud University Medical Center, 6500HB Nijmegen,
33 the Netherlands.

34 ¹¹South West Thames Regional Genetics Service, St George's, University of London,
35 London SW17 0RE, United Kingdom.

36 ¹²Department of Clinical and Biological Sciences, University of Torino, 10124 Torino, Italy.

37 ¹³Department of Clinical Genetics, Erasmus MC, University Medical Center Rotterdam, 3015
38 GD Rotterdam, the Netherlands.

39 ¹⁴Centre for Clinical genetics, Sydney Children's Hospitals Network Randwick, Randwick,
40 NSW Australia.

41 ¹⁵Discipline of Paediatrics and Child Health, Faculty of Clinical Medicine, UNSW, Randwick,
42 NSW, Australia.

43 ¹⁶Genetics of Learning Disability Service, Royal North Shore Hospital, St Leonards, Sydney,
44 NSW 2065, Australia.

45 ¹⁷Microgenomics srl, Pavia, Italy.

46 ¹⁸Human Pathology, ASST Santi Paolo e Carlo, San Paolo Hospital, Milan, Italy.

47 ¹⁹Child Neuropsychiatry Unit - Epilepsy Center, ASST Santi Paolo e Carlo, San Paolo
48 Hospital, Department of Health Sciences, Università degli Studi di Milano, Milan, Italy.

49 ²⁰Division of Medical Genetics, Department of Pediatrics, University of Utah, Salt Lake City,
50 UT, USA.

51 ²¹Department of Molecular Medicine, University of Pavia, 27100, Pavia, Italy.

52 ²²Child Neuropsychiatry Unit, ASST Grande Ospedale Metropolitano Niguarda, Department
53 of Health Sciences, Università degli Studi di Milano, Milan, Italy.

54 ²³Department of Genetics, Reference center for Rare Diseases and Developmental
55 Anomalies, Caen, France.

56 ²⁴Department of Pediatrics, Columbia University Medical Center, New York, NY 10032, USA.

57 ²⁵Department of Medicine, Columbia University Medical Center, New York, NY 10032, USA.

58 ²⁶Sección de Genética Médica, Servicio de Pediatría, Hospital Clínico Universitario Virgen de
59 la Arrixaca, IMIB-Arrixaca, CIBERER-ISCIII, 30120 Murcia, Spain.

60 ²⁷Department of Audiology and Speech Pathology, University of Melbourne, Melbourne,
61 Victoria, 3010, Australia.

62 ²⁸Donders Institute for Brain, Cognition and Behaviour, Radboud University, 6500HE
63 Nijmegen, the Netherlands.

64 *Correspondence:

65 maggie.wong@mpi.nl and simon.fisher@mpi.nl

66 **Abstract**

67 Germline *de novo* *SETBP1* variants cause clinically distinct and heterogeneous
68 neurodevelopmental disorders. Heterozygous missense variants at a hotspot encoding a
69 canonical degron lead to *SETBP1* accumulation and Schinzel-Giedion syndrome (SGS), a
70 rare severe developmental disorder involving multisystem malformations. Heterozygous
71 loss-of-function variants result in *SETBP1* haploinsufficiency disorder which is phenotypically
72 much milder than SGS. Following an initial description of four individuals with atypical SGS
73 carrying heterozygous missense variants adjacent to the degron, a few individual cases of
74 variants outside the degron were reported. Due to the lack of systematic investigation of
75 genotype-phenotype associations of different types of *SETBP1* variants, and limited
76 understanding of the roles of the gene in brain development, the extent of clinical
77 heterogeneity and how this relates to underlying pathophysiological mechanisms remain
78 elusive, imposing challenges for diagnosis and patient care. Here, we present a
79 comprehensive investigation of the largest cohort to-date of individuals carrying *SETBP1*
80 missense variants outside the degron (n=18, including one in-frame deletion). We performed
81 thorough clinical and speech phenotyping with functional follow-up using cellular assays and
82 transcriptomics. Our findings suggest that such variants cause a clinically and functionally
83 variable developmental syndrome, showing only partial overlaps with classical SGS and
84 *SETBP1* haploinsufficiency disorder, and primarily characterised by intellectual disability,
85 epilepsy, speech and motor impairment. We provide evidence of loss-of-function
86 pathophysiological mechanisms impairing ubiquitination, DNA-binding and transcription. In
87 contrast to SGS and *SETBP1* haploinsufficiency, these effects are independent of protein
88 abundance. Overall, our study provides important novel insights into diagnosis, patient care
89 and aetiology of *SETBP1*-related disorders.

90 Introduction

91 Different types of germline variants in the gene encoding SET binding protein 1 (*SETBP1*)
92 (NM_015559.2) lead to distinct and phenotypically heterogeneous developmental
93 syndromes¹. Schinzel–Giedion syndrome (SGS, MIM #269150)² is a rare and severe
94 disorder with multisystem malformations that include recognizable facial characteristics,
95 neurological problems (including severe intellectual disability, intractable epilepsy, cerebral
96 blindness and deafness) and various congenital anomalies (such as heart defects, and
97 genital, kidney and bone abnormalities)^{3–6}. The majority of patients with a molecularly
98 confirmed diagnosis have a shortened lifespan and often do not survive beyond the first
99 decade⁶. In 2010, heterozygous germline *de novo* missense variants in one section of the
100 *SETBP1* gene were identified as a cause of SGS. The variants cluster in a hotspot of 12
101 base pairs coding for four amino acids (residues 868 to 871) in the SKI domain of the
102 *SETBP1* protein^{2,6}. These four amino acid residues are part of a canonical degron sequence
103 recognised by ubiquitin E3 ligases and important for regulating protein degradation, and the
104 variants are thought to cause SGS via gain-of-function mechanisms. Intriguingly, overlapping
105 somatic *SETBP1* variants have been identified recurrently in several forms of myeloid
106 leukaemia^{6,7}. These overlapping somatic variants are more disruptive to the degron, implying
107 a higher functional threshold to cause cancer⁶, and it is thought that germline variants might
108 lead to features similar to cancer cells such as increased proliferation and accumulation in
109 DNA damage but via different cell type-specific pathways⁸.

110 In contrast, *SETBP1*-specific deletions and *de novo* loss-of-function (LoF) point mutations
111 disrupting the gene result in *SETBP1* haploinsufficiency disorder⁹, (MIM #616078), a
112 syndrome that is phenotypically milder than SGS^{10–14}. The heterozygous *SETBP1*-specific
113 deletions lead to a loss of the locus while the LoF variants (nonsense and frameshift
114 variants) are predicted to lead to nonsense-mediated decay of the mutated *SETBP1*
115 transcript, both of which therefore result in reduced dosage of *SETBP1* protein
116 (haploinsufficiency). Unlike classical SGS, individuals carrying such variants do not show

117 major congenital or growth anomalies. Our recent systematic gene-driven studies of a large
118 cohort with confirmed *SETBP1* LoF variants revealed a far broader clinical severity spectrum
119 than previously reported^{15,16}. Despite subtle overlapping facial dysmorphisms, and in
120 contrast to SGS, the affected individuals do not present with a recognizable facial gestalt or
121 specific features of dysmorphisms. The main clinical features include moderate-to-severe
122 speech and language impairments, mild motor developmental delay, wide variability in
123 intellectual functioning, hypotonia, vision impairment, and behavioural problems such as
124 attention/concentration deficits and hyperactivity^{15,16}. Heterozygous pathogenic LoF variants
125 in *SETBP1* have been independently identified by exome/genome sequencing in different
126 cohorts of individuals with childhood apraxia of speech (CAS)^{13,17}. In a comprehensive
127 speech phenotyping study led by Morgan *et al.*¹⁵, poor communication was shown to be a
128 central feature of *SETBP1* haploinsufficiency disorder. These findings further highlight the
129 heterogeneity and complexity of *SETBP1*-related aetiologies.

130 *SETBP1* is expressed in numerous tissues including the brain, and multiple alternative
131 transcripts encoding different isoforms have been found [OMIM 611060]. However, there is
132 little existing knowledge concerning the expression and functions of the *SETBP1* protein as
133 well as its isoforms, with much of our understanding coming from studies of somatic variants
134 in haematopoietic cells or overexpression systems. Accumulation of *SETBP1* has been
135 shown to reduce protease cleavage of its interactor oncoprotein SET, resulting in the
136 formation of a *SETBP1*-SET-PP2A complex that inhibits PP2A phosphatase activity, thus
137 promoting proliferation of leukaemic cells^{6,7,18,19}. *SETBP1* has chromatin remodeller
138 functions, binding to AT-rich genomic regions via two AT-hooks, and has been shown to
139 recruit a HCF1/KMT2A/PHF8 epigenetic complex in HEK cells overexpressing *SETBP1*²⁰.
140 Acting as a transcription factor, *SETBP1* is able to activate or repress expression of genes
141 such as *HOXA9*, *HOXA10*, *RUNX1*, *MYB* and *MECOM* in haematopoietic cells and in HEK
142 cells overexpressing an SGS variant²⁰⁻²³. Nevertheless, the roles of *SETBP1* in the
143 developing brain and the pathophysiological pathways underlying germline pathogenic

144 *SETBP1* variants remain elusive. Thus far, only three recent studies have investigated the
145 molecular consequences of *SETBP1* disruptions in mouse and human neuronal
146 models^{8,20,24}, all of which described impairment in cell proliferation as a shared mechanism.
147 Overexpression of human wild type *SETBP1* and an SGS variant in mouse embryos via *in*
148 *utero* electroporation disrupted neuronal migration and neurogenesis in the neocortex²⁰. In
149 neural progenitor cells derived from patients with SGS, accumulation of *SETBP1* promoted
150 cell proliferation and induced DNA damage via SET stabilization that inhibited P53, and a
151 PARP-1-dependent mechanism⁸. Investigations of neural progenitors in which *SETBP1* had
152 been knocked out identified prolonged proliferation and distorted layer-specific neuronal
153 differentiation with overall decrease in neurogenesis via the WNT/ β -catenin signalling
154 pathway²⁴. However, these studies focused on a few SGS variants or *SETBP1* knockouts,
155 leaving the majority of missense variants uncharacterized.

156 In 2017, Acuna-Hidalgo *et al.* identified four individuals carrying *SETBP1* variants in close
157 proximity to the canonical degron [p.(Glu862Lys), p.(Ser867Arg), p.(Thr873Ile)] who showed
158 a milder developmental phenotype with clinical characteristics that partially overlapped with
159 classical SGS⁶. Additional individuals carrying missense variants in close proximity to the
160 degron have since been reported in the medical literature. Nevertheless, the majority of the
161 variants reported thus far are classified as variants of uncertain significance and their
162 functional impacts have not been thoroughly characterized. Genotype-phenotype
163 associations of germline *SETBP1* variants therefore remain unclear. The highly variable
164 phenotypes and severity seen in these patients have made diagnoses difficult and precluded
165 development of new, personalized therapies, often creating confusion among clinicians and
166 patient families.

167 The current study addresses these important issues using a gene-driven approach. We
168 recruited the largest cohort of individuals, to our knowledge, with a molecular diagnosis of
169 *SETBP1* variants (missense variants and one in-frame deletion) outside the degron region
170 (i.e. not affecting amino acids 868-871) who show clinical features that do not fit with the

171 original SGS diagnostic criteria. We investigated the genotype-phenotype relationships of
172 variants in the *SETBP1* gene by thorough phenotyping of clinical and speech/language
173 features as well as functional follow-up of specific variants in cellular assays. We discovered
174 that *SETBP1* variants outside the degron lead to a neurodevelopmental disorder
175 characterised by a broad spectrum of clinical features, which are much milder than classical
176 SGS despite partial overlaps, but in some cases more severe than *SETBP1*
177 haploinsufficiency disorder. Using functional cellular assays, we showed that effects of
178 *SETBP1* variants involve a combination of mechanisms including deficits in protein
179 degradation, ubiquitination and transcriptional control, independent of elevated protein
180 levels. Overall, our findings point towards pathophysiological mechanisms that act via
181 functional dosage of *SETBP1* protein in a way that is dependent on variant location and
182 specific change in amino-acid residues, rather than merely the amount of protein, explaining
183 the clinical heterogeneity observed in patients.

184 **Material and Methods**

185 **Molecular analyses**

186 Peripheral blood samples were collected in a diagnostic context from the proband, and
187 parents when available. Results originated from diagnostic or research-based genome
188 sequencing, exome sequencing, gene panel testing for intellectual disability or epilepsy or
189 via initial direct Sanger sequencing of the *SETBP1* gene based on clinical phenotypes,
190 which in one case was followed by trio-based exome sequencing (details in Table S1).

191 **Patients and consent**

192 This study analysed the medical data for patients with a variant in *SETBP1*. Only patients
193 with missense variants outside the degron region (not affecting amino acids 868-871) of
194 *SETBP1* were included in this study. The clinical data and patient materials were obtained
195 through international collaborations involving clinicians from various countries (Table S1).
196 Some of these collaborations were established via GeneMatcher²⁵. Figure 2, Table 1, Table
197 S1 and Figure S1 are not included in this preprint and are available from the corresponding
198 author on request. Consent procedures and details of the IRB/oversight body that provided
199 approval or exemption for the research described are described in the ethical statement.

200 **Speech and language phenotyping**

201 History of speech and language development was recorded. Speech was analysed for the
202 presence of articulation, phonological disorder, stuttering, CAS^{17,26} and dysarthria^{27,28}.
203 Clinical diagnostic reports from local treating speech pathologists were used to validate the
204 diagnoses with 100% agreement. The Intelligibility in Context Scale²⁹ examined how often
205 the participant's speech was understood, with a 5-point scale of responses ranging from
206 never to always. Children were assessed for verbal, written and social language with the
207 standardised Vineland Adaptive Behaviour Scales-Parent/Caregiver III³⁰. Adult participants
208 were assessed for receptive language with the Peabody Picture Vocabulary Test-4³¹ and

209 clinical speech pathology (ST, RB, AM) and neurology (IS) assessment determined
210 expressive language, social skills and written language ability. Table 2 is not included in this
211 preprint and is available from the corresponding authors on request.

212 **Spatial clustering analysis of missense variants**

213 All independently observed *SETBP1* missense variants outside the canonical degtron were
214 included in the spatial clustering analysis^{32,33}. The geometric mean was computed over the
215 locations of observed missense variants in the canonical transcript of *SETBP1* (9,899 bp;
216 NM_015559.2) and subsequently compared to 100,000,000 permutations by randomly
217 redistributing the variant locations over the total size of the *SETBP1* coding region and
218 calculating the resulting geometric mean from each of these permutations. The corrected *p*-
219 value was computed by checking how often the observed geometric mean distance was
220 smaller than the permuted geometric mean distance and considered significant if <0.05.

221 **Cell culture and transfection**

222 HEK293T/17 cells (CRL-11268, ATCC) were cultured in Dulbecco's modified Eagle's
223 medium (DMEM) (Gibco) supplemented with 10% foetal bovine serum (FBS) (Gibco) and
224 100 U/ml Penicillin-Streptomycin (Thermo-Fisher) at 37°C and 5% CO₂. For
225 immunofluorescence analysis, cells were seeded onto coverslips coated with 100 µg/ml
226 poly-L-lysine (Merck-Millipore). Fibroblast cell lines were established from skin biopsies of
227 patients with *SETBP1* variants and controls at the Cell Culture Facility, Department of
228 Human Genetics at Radboud university medical center. Human dermal fibroblasts were
229 cultured in DMEM supplemented with 20% FBS, 100 U/ml Penicillin-Streptomycin and 1%
230 sodium pyruvate (Thermo-Fisher) at 37°C and 5% CO₂. Transfections were performed using
231 transfection reagent GeneJuice (Merck-Millipore) following the manufacturer's instructions or
232 polyethylenimine (PEI) (Sigma) in a 3:1 ratio with the total mass of DNA transfected.

233 **DNA Constructs and site-directed mutagenesis**

234 The full-length SETBP1 construct fused to a C-terminal Myc-DDK tag under a human CMV
235 promoter (pCMV-Entry-SETBP1) was purchased from Origene (RC229443). Constructs
236 carrying *SETBP1* variants (pCMV-Entry-SETBP1) were generated using QuikChange
237 Lightning Site-Directed Mutagenesis Kit (Agilent) following the manufacturer's protocol. To
238 generate constructs carrying *SETBP1* fused to an YFP-tag (pYFP-SETBP1), *SETBP1*
239 cDNAs were first subcloned using EcoRI/XhoI restriction sites into a modified pEGFP-C2
240 vector (Clontech) where the N-terminal EGFP-tag was then replaced with an YFP-tag using
241 BshTI/Bsp1407I restriction sites. To generate AT-rich reporter vectors used in luciferase
242 reporter assay, 100 μ M sense and antisense single-stranded oligonucleotides carrying 5'-
243 AAAATAA-3' or 5'-AAAATAT-3' repeats were first annealed using 2x annealing buffer
244 [20mM Tris (Thermo-Fisher), 2mM EDTA (Sigma), 100mM NaCl (Sigma), pH 8.0]. Annealed
245 oligonucleotides were phosphorylated using 1x T4 DNA ligase buffer and 30 units of T4 PNK
246 (New England Biolabs), then inserted into a pGL4-luc2 vector using KpnI/HindIII restriction
247 sites. A firefly luciferase reporter construct containing FOXP2 promoters, TSS1 and TSS2,
248 respectively were a gift from the Vernes group³⁴. For mammalian one-hybrid (M1H) assays,
249 a pGL4-luc2-GAL4UAS-adenovirus promotor reporter construct and a pGL4.23 construct
250 with an adenovirus major late promotor (pBIND2) were generated as previously described³⁵.
251 *SETBP1* cDNAs were subcloned from pCMV-Entry-SETBP1 into the empty pBIND2 using
252 Sall/NotI restriction sites generating pBIND2-SETBP1 expression vectors. A GFP-SET
253 construct was generated as previously described⁶. All constructs were verified by Sanger
254 sequencing. Primer sequences are listed in Table S11.

255 **Immunofluorescence microscopy**

256 HEK293T/17 cells grown on poly-L-lysine-coated coverslips were transiently transfected with
257 500ng of a pCMV-SETBP1 or pYFP-SETBP1 construct. Cells were fixated 48 hours after
258 transfection using 4% paraformaldehyde solution (Electron Microscopy Supplies Ltd) for 20
259 minutes at room temperature (RT). Human dermal fibroblasts were grown on coverslips
260 coated with 1 μ g/ml fibronectin (Corning) and fixated as above. Fixated cells were

261 permeabilized with 0.4% Triton-X-100/PBS for 20 minutes at RT. Following incubation in a
262 blocking solution for 1 hour at RT, cells were incubated in primary antibodies overnight at
263 4°C and then with conjugated secondary anti-IgG antibodies for one hour at RT. Hoechst
264 33342 (Invitrogen) was used for nuclear staining, before mounting with VECTASHIELD®
265 Antifade Mounting Medium (Vectorlab). Fluorescence images were obtained using an
266 LSM880 AxioObserved confocal microscope (Zeiss). For images of single nuclei, the
267 Airyscan unit (Zeiss) was used with a 4.0 zoom factor. Images were analyzed with an
268 ImageJ “Analyze particle” plugin. A list of antibodies used can be found in Table S12.

269 **Co-immunoprecipitation**

270 HEK293T/17 cells grown in a 10-cm culture dish were transiently transfected with constructs
271 expressing FLAG-SETBP1 only or together with GFP-SET under a human CMV promoter.
272 An empty pCR2.1-TOPO (Invitrogen) was used as a filler to top up to 12 µg DNA in total per
273 transfection. Cells were lysed in 1ml of Pierce® IP Lysis Buffer (Thermo-Fisher)
274 supplemented with protease inhibitors (Roche) and 1% PMSF (Thermo-Fisher) 48 hours
275 post-transfection. All following steps were performed at 4°C. Cell lysates were incubated for
276 10 minutes under rotation, centrifuged at 13200 rpm for 20 minutes. The supernatant was
277 then quantified with a Pierce BCA protein assay kit (Thermo-Fisher). 5% protein lysate was
278 collected as input and denatured in Laemmli buffer (Bio-Rad). 400 µg protein was
279 immunoprecipitated with 30 µl FLAG-agarose beads (Thermo-Fisher) or 25 µl GFP-Trap
280 (Chromotek) on a rotating wheel overnight. Beads were then washed three times with PBS
281 and eluted with Laemmli buffer for immunoblotting analysis.

282 **Immunoblotting and band intensity quantification**

283 HEK293T/17 cells grown in a 6-well culture plate were transiently transfected with 3 µg DNA.
284 Cells were lysed in 1 ml of Pierce® RIPA Buffer (Thermo-Fisher) supplemented with
285 protease inhibitors (Roche) and 1% PMSF (Thermo-Fisher) 48 hours post-transfection. Total
286 protein was quantified using a BCA assay. Proteins were resolved on 4 –15% Tris-Glycine

287 gels and transferred to PDVF membranes (Bio-Rad). After blotting, membranes were
288 incubated overnight at 4°C with the appropriate primary antibodies, followed by HRP-
289 conjugated secondary antibodies (Table S12). Proteins were visualized using the Novex
290 ECL Chemiluminescent Substrate Reagent kit (Invitrogen) or SuperSignal West Femto
291 Maximum Sensitivity Substrate (Thermo-Fisher) and the ChemiDoc XRS+ System (Bio-
292 Rad). Band intensity was quantified using ImageJ. Background-subtracted band intensity
293 was divided by the background-subtracted band intensity of β -actin for normalization. For
294 quantification of (co-) immunoprecipitation experiments, the background-subtracted band
295 intensity of the immunoprecipitated fraction was normalised with respect to the input fraction
296 for each condition.

297 **Fluorescence-based quantification of protein stability and degradation**

298 HEK293/T17 cells were transfected in triplicate in clear-bottomed black 96-well plates with
299 YFP-tagged SETBP1 variants. After 24 hours, cycloheximide (Sigma) at a final concentration
300 of 50 μ g/ml was added, MG132 (R&D Systems) at 5 μ g/ml and Bafilomycin A1 (R&D
301 Systems) at 100 nM. Cells were incubated at 37°C with 5% CO₂ in the Infinite M200Pro
302 microplate reader (Tecan), and the fluorescence intensity of YFP (Ex: 505 nm, Em: 545 nm)
303 was measured over 24 hours at 3-hour intervals.

304 **Luciferase reporter assay and mammalian-one-hybrid (M1H) assay**

305 HEK293/T17 cells were seeded in clear-bottomed white 96-well plates (Greiner Bio-One)
306 and transfected in triplicates using GeneJuice (Merck-Millipore). Cells were co-transfected
307 with 350 ng of firefly luciferase reporter construct containing six repeats of AT-rich
308 consensus sequence (5'-AAAATAA-3' or 5'-AAAATAT-3') or *FOXP2* promoters
309 (TSS1/TSS2)³⁴, 6.5ng of pGL4.74 Renilla luciferase normalization control, and 700 ng of
310 pYFP construct with or without SETBP1. For M1H assay, cells were co-transfected with
311 1,433 ng (50 nM) of expression construct containing only GAL4 or GAL4 including SETBP1,
312 416 ng of reporter construct with or without GAL4-binding site, and 165 ng of Renilla

313 luciferase normalization control. After 48 hours, firefly luciferase and Renilla luciferase
314 activities were measured using a Dual-Luciferase Reporter Assay system (Promega)
315 according to manufacturer's instructions at an Infinite M Plex Microplate reader (Tecan).

316 **Cell proliferation assay**

317 Fibroblasts were seeded in triplicate in clear-bottomed black 96-well plates (Greiner Bio-
318 One). Cell proliferation was measured using a CyQUANT™ Direct Cell Proliferation Assay
319 (Thermo-Fisher) according to manufacturer's instructions at an Infinite M Plex Microplate
320 reader. Fluorescence intensity were measured daily for four days. Background fluorescence
321 was subtracted and values were normalized to day 0. Growth curves were plotted as
322 fluorescence versus time.

323 **RNA sequencing (RNA-seq) and data analysis**

324 Three technical replicates of each fibroblast cell line were included in the RNA-seq
325 experiment. Total RNA was extracted from one to two million cells. 1 µg of total RNA
326 extracted (Qiagen) was used to generate RNA libraries using NEBNext® Ultra™ RNA
327 Library Prep Kit for Illumina® (New England Biolabs) following the manufacturer's protocol
328 and index codes were added to attribute sequences to each sample. The libraries were
329 sequenced on a NovaSeq6000 platform using 150bp pair-end reads. Image processing and
330 basecall was performed using the Illumina Real Time Analysis Software. Fastq files were
331 aligned to the human genome (GRCh38/hg39, Ensembl) using HISAT2³⁶ software together
332 with the corresponding splice junctions Ensembl GTF annotation. At least 89% of the reads
333 for each sample were uniquely mapped to the human genome entailing at least 36.4 million
334 unique reads in the sample with the lowest sequencing depth. Principal Component Analysis
335 (PCA) was performed using the SeqMonk tool (version 1.47.1). Differential gene expression
336 was performed using DESeq2³⁷ and intensity differences [$p > 0.05$, false discovery rate
337 (FDR)] were calculated via the SeqMonk tool (version 1.47.1) for female and male samples
338 (patients versus controls) separately. Gene set enrichment analysis of the differentially

339 expressed genes (DEGs) that were present in both female and male samples was
340 performed using GSEA software (version 4.1.0)³⁸ using gene sets for biological pathways
341 (c5.go.bp.v7.2.symbols.gmt) with 1000 permutations and multiple testing correction ($p > 0.05$,
342 FDR). Gene ontology analysis was performed using g:GOST within the g:Profiler web
343 server³⁹. A list of expressed genes in our fibroblast samples was used as the custom
344 background for gene ontology analysis with multiple testing correction ($p < 0.05$, Benjamini-
345 Hochberg FDR). Degrees of overlap between the differentially expressed genes in our
346 fibroblast samples and autism (v.0.22)- or intellectual disability (ID, v.3.2)-associated genes
347 in the PanelApp database were calculated using a Fisher's exact test ($p > 0.05$). A list of
348 expressed genes in our fibroblast samples was used as the custom background. Datasets
349 generated and analysed in the current study are not included in this preprint, and are
350 available from the corresponding authors on request.

351 **Quantitative real-time PCR**

352 Total RNA was extracted with Qiazol or an RNeasy Plus mini kit (Qiagen) following
353 manufacturer's protocols. 1 μ g of total RNA was used to synthesis cDNA using
354 SuperScriptIII Reverse Transcriptase (Thermo-Fisher). Real-time PCR was performed using
355 iQ SYBR Green Supermix (Bio-Rad) at a CFX384 Touch Real-Time PCR Detection System
356 (Bio-Rad).

357 **Statistical analysis of cell-based functional assays**

358 Statistical analyses for cell-based functional assays were done using a one- or two-way
359 ANOVA followed by a Bonferroni, Tukey, or Dunnett *post-hoc* test, with GraphPad Prism
360 (version 8.0) Software.

361 Results

362 **SETBP1 variants outside the canonical degron of the SETBP1 protein cluster in the** 363 **SKI domain**

364 In our study, we identified 18 unrelated individuals carrying rare heterozygous variants with
365 uncertain functional impact in *SETBP1* (NM_015559.2), a gene under constraint against
366 loss-of-function and missense variation [pLoF: o/e = 0.02 (0.01 – 0.11); missense: o/e = 0.9
367 (0.84 – 0.95); gnomAD v.2.1.1⁴⁰] (Figure 1A; Tables 1, S1 and S3). Variants were identified
368 via exome or genome sequencing. In one case, the variant was first identified with direct
369 Sanger sequencing of the *SETBP1* gene based on clinical observations in the patient
370 followed by trio-based exome sequencing. None of the 18 individuals met the diagnostic
371 criteria of Schinzel-Giedion Syndrome (SGS)⁴¹. Fifteen individuals carried a *de novo*
372 *SETBP1* variant, while the other two had inherited a variant from an affected parent. One
373 individual harboured a missense variant that was not inherited from the mother; the father
374 was unavailable for testing. Among the 15 individuals with a *de novo* variant, 14 carried a
375 missense variant and one had an in-frame deletion [c.2885_2887del(CCA) p.(Thr962del)].
376 Within our cohort, there were multiple cases of recurrent identical *de novo* variants including
377 c.2572G>A p.(Glu858Lys) found in four children and c.2584G>A p.(Glu862Lys) in two
378 individuals, revealing an independent mutational hotspot located in close proximity to the
379 canonical degron region. None of the *SETBP1* variants included in our study were present in
380 the gnomAD v.2.1.1 database. Two individuals also carried variants affecting other known
381 disease genes. In proband 3, who has a *de novo* c.2561C>T p.(Ser854Phe) *SETBP1*
382 variant, an *EHMT1* variant was identified. In proband 1, who has a c.1332C>G
383 p.(Ser444Arg) *SETBP1* variant, an inherited 5q22.31 dup and *WWOX* variants were
384 identified. However, these variants in probands 1 and 3 are unlikely to be pathogenic based
385 on the patients' clinical features, the inheritance model and the lack of functional impact
386 observed in assays performed in cellular models (unpublished data with the *EHMT1* and
387 *WWOX* variants).

388 ***SETBP1* variants outside the canonical degron cause a broad spectrum of clinical**
389 **features**

390 An overview of the main clinical features per individual is shown in Table 1. More detailed
391 data for each patient are shown in Table S1. Individuals (n=5) with a variant in close
392 proximity to the degron within the SKI domain (affecting amino acids 862-874, excluding the
393 degron 868-871) showed severe or profound intellectual disability and severe motor
394 impairment with inability to walk without support. Four of these individuals were unable to
395 speak. Three individuals showed spasticity. Focal and tonic-clonic seizures were noted in
396 two of these individuals. Two individuals showed renal abnormalities: one had mild kidney
397 dilatation, another medullary cystic kidneys. Shared facial features were present in at least
398 three out of the five individuals, including prominent ears, shallow orbits, midface retraction
399 and microcephaly. Overall, the phenotypes of these five children did not fulfil the original
400 Lehman et al. criteria for SGS⁴¹. However, based on the severity of the phenotypes and
401 facial features, they appeared more similar to (but less severe than) SGS compared to
402 *SETBP1* haploinsufficiency disorder.

403 Individuals (n=7) with variants (amino acids 854-858) located slightly further away from the
404 degron but still within the SKI domain showed mild or moderate intellectual disability. Two
405 out of four individuals with a c.2572G>A p.(Glu858Lys) variant were minimally verbal (see
406 the speech and language section; the relevant data on this from the other two individuals
407 were unavailable). All seven individuals showed motor delay, but the degree was much
408 milder compared to the aforementioned cases. All those aged four years and older were able
409 to walk without support, although one individual walked only limited distances. One individual
410 of 3.5 years did not walk yet. Absence seizures were noted in two of these cases. One
411 individual had asplenia. One individual had a non-progressive heart tumour of unknown
412 origin. These individuals did not show recognizable or similar facial features, nor did they
413 show overlapping facial features with either classical SGS or *SETBP1* haploinsufficiency
414 disorder (data not shown).

415 The individuals with inherited variants located furthest away from the degron and the SKI
416 domain [c.1332C>G p.(Ser444Arg) and c.1970T>C p.(Val657Ala)] showed mild intellectual
417 disability or a low non-verbal IQ. In both cases, parents carrying the variant were similarly
418 affected. For the remainder of the study, we therefore did not distinguish these inherited
419 variants from the *de novo* variants outside the degron. All variants outside the degron were
420 considered in functional assays as one group versus those within the degron (causing
421 classical SGS).

422 Variants located outside the SKI domain (n=4) were associated with a more variable clinical
423 phenotype. One individual with an in-frame deletion removing a threonine residue
424 c.2885_2887del(CCA) p.(Thr962del) showed a severe phenotype with severe speech delay,
425 inability to walk and tonic-clonic seizures. This individual had bilateral ptosis and had surgery
426 to the right upper eyelid and left strabismus surgery. This proband's facial features appeared
427 similar to the individual with the c.2984C>T p.(Leu957Pro) variant. They both showed a
428 round face, blepharophimosis, hypertelorism and a short nose with a bulbous tip, features
429 also often noted in individuals with *SETBP1* haploinsufficiency disorder. The latter individual
430 had a less severe neurodevelopmental phenotype. This individual started to walk at 22
431 months and was able to use sign language at the age of three years.

432 **Individuals carrying *SETBP1* variants outside the canonical degron show generally**
433 **low speech/language ability**

434 Speech/language data were available from seven individuals in this cohort, four of whom
435 carry a variant within the SKI domain close to the degron (proband 3, 6, 8 and 14) while
436 three individuals harbour a variant located far from the degron and outside the SKI domain
437 (proband 1 and 18, and affected mother of proband 1). In this cohort, speech development
438 during infancy was characterised by limited babbling. A history of early feeding difficulties
439 was also present for two children (proband 3 and 18). Language ability was generally low
440 across the group (n=7) for all subdomains including expressive, receptive, written and social

441 language (Table 2 and Figure 3). The youngest children (< 9 years of age; probands 3, 6, 8
442 and 18) present with a severe speech and language impairment. They are minimally verbal,
443 defined as the presence of less than 50 spoken words (Table 2), and augment verbal
444 communication with sign language, gestural communication and digital devices. The speech
445 motor system is impaired across all individuals in the group, with Childhood Apraxia of
446 Speech (CAS) the most common diagnosis (n=5), followed by mild dysarthria (proband 1
447 and affected mother) (Table 2). CAS features included hesitancy, groping, inconsistency of
448 production, increased errors with increasing word length, simplified word and syllable
449 structures relative to age, and vowel and prosodic errors. Dysarthria was typically
450 characterised by slower rate of speech, imprecision of consonants, altered nasal resonance
451 and monotonous speech. The adult participants (proband 1 and affected mother) had a
452 history of poor speech development but are now able to hold appropriate conversations and
453 speak in full sentences with speech that is usually to always intelligible. All individuals in this
454 cohort who performed speech/language assessment are attending (probands 3, 6, 8 and
455 18), or had attended (probands 1 and 14, and affected mother of proband 1), speech
456 therapy.

457 ***SETBP1* variants outside the canonical degron are predicted to be damaging and**
458 **cluster in the SKI domain**

459 We used an array of computational tools to predict the functional effects of the observed
460 *SETBP1* variants. Among the 14 variants observed in 18 individuals, eight were located in
461 the SKI domain while six were outside any known functional domain (Figure 1A). Using a
462 spatial clustering analysis³³, we showed that these variants outside the degron significantly
463 clustered in exon 4 of the canonical *SETBP1* transcript (NM_015559.2) corresponding to the
464 SKI domain (corrected *p*-value = 9.99e-09, Bonferroni correction). All of the mutated amino
465 acid sites were highly conserved across species with the exception of the threonine residue
466 at position 962, which was conserved only in mammals (Figure 1B). All observed variants

467 were predicted to be (likely) pathogenic by PolyPhen-2 and/or SIFT, and showed CADD-
468 PHRED scores above 21 (Table S2).

469 We went on to use the MetaDome web tool (v.1.0.1) to visualise all *SETBP1* missense
470 variants in the tolerance landscape of the gene (Figure S2A and B). Variants in the SKI
471 domain are located in the regions of high intolerance while the remaining variants, including
472 those adjacent to the HCF1-binding site, are located in the less tolerant regions (Figure S2A
473 and B). Of note, *SETBP1* does not have a particularly high Z-score (1.1) for missense
474 variants in the gnomAD database (v.2.1.1), indicating that the complete coding region of this
475 gene is not extremely intolerant to missense variation. This observation is consistent with the
476 results from the MetaDome analysis, which show that only a few regions of *SETBP1* have a
477 high intolerance for missense variants, including the part of the SKI domain in which the
478 majority of the observed variants are located (Figure S2B). We observed that 33% of the
479 amino acid residues that were mutated by a germline *SETBP1* missense variant were also
480 mutated in somatic cells, particularly in haematopoietic and lymphoid cells, according to the
481 COSMIC database (Catalogue of Somatic Mutations in Cancer v.94), including the recurrent
482 missense variant p.(Glu858Lys). While missense variants within the degron (affecting amino
483 acids 868-871) showed different frequencies in germline and somatic cells, consistent with
484 the previously reported higher functional threshold in somatic cells⁶, all observed missense
485 variants outside the degron showed similar frequencies in germline and somatic cells (Figure
486 S3 and Table S3).

487 ***SETBP1* missense variants outside the degron cause variable disruption of protein**
488 **degradation in contrast to accumulating stable SGS variants**

489 We went on to study the functional consequences of a representative selection of variants
490 across *SETBP1* on protein localization, protein stability and degradation, cell proliferation
491 and transcriptional activity using HEK293T/17 cells transiently transfected with *SETBP1*
492 expression constructs, as well as in patient fibroblasts. Based on the location and their

493 distance to the canonical degron (Figure 1A), we included in our assays two missense
494 variants located furthest from the SKI domain [p.(Ser444Arg) and p.(Val657Ala)], three
495 missense variants close to the degron [p.(Glu858Lys), p.(Glu862Lys)], and p.(Leu957Pro)],
496 and one *de novo* in-frame deletion [p.(Thr962del)]. In addition, we included in our assays
497 four classical SGS variants located within the canonical degron [p.(Asp868Asn),
498 p.(Ser869Asn), p.(Gly870Ser), and p.(Ile871Thr)] for functional comparisons to those outside
499 the degron.

500 We first assessed the abundance of FLAG-tagged SETBP1 in transfected HEK293T/17
501 cells, comparing cells with variant expression constructs to those with a wild type construct.
502 All variants showed higher SETBP1 protein levels than the wild type (Figures 4A, 4B and
503 S4A). Next, we hypothesized that missense variants outside the degron might alter SETBP1
504 stability. We therefore treated HEK239T/17 cells expressing YFP-tagged SETBP1 with
505 cycloheximide to inhibit translation and measured relative fluorescence intensity over 24
506 hours. We found that all classical SGS variants (within the degron) showed increased protein
507 stability whereas all variants outside the degron displayed similar stability to wild type
508 (Figures 4C and S6). To evaluate the impact of variants on proteasome-mediated
509 degradation, we treated HEK293T/17 cells expressing YFP-tagged SETBP1 with
510 proteasome inhibitor MG132. Surprisingly, the classical SGS variants did not show impaired
511 proteasome degradation, except for p.(Gly870Ser) (Figures 4C and S7), unlike previously
512 reported⁶. Moreover, three variants outside the degron [p.(Ser444Arg), p.(Glu858Lys), and
513 p.(Leu957Pro)] demonstrated disrupted proteasome degradation to various extents (Figure
514 4C). To assess whether the degradation of *SETBP1* variants might be compensated by
515 other protein degradation pathways, such as mTOR-dependent autophagy, we used the
516 autophagy inhibitor BafilomycinA1 to treat HEK293T/17 cells expressing YFP-tagged
517 SETBP1. While the majority of variants outside the degron [p.(Ser444Arg), p.(Val657Ala),
518 p.(Glu858Lys), and p.(Leu957Pro)] differed significantly in degradation via autophagy, most
519 of the SGS variants were similar to wild type (Figures 4C and S8). It is noteworthy that the

520 direction and extent of degradation of the variant proteins was variable and appeared to
521 depend on the distance of the variant from the degron region. Consistent with the
522 overexpression system, patient fibroblasts [p.(Glu858Lys), p.Leu957Pro, and p.(Thr962del)]
523 also showed significantly more abundant SETBP1 protein compared to healthy controls,
524 although mRNA expression appeared to be more variable (Figure 4D). Interestingly, protein
525 stability and degradation of the in-frame deletion p.(Thr962del) was not affected (Figures 4C
526 and S6A), suggesting that it might operate via a different pathophysiological mechanism, in
527 spite of increased abundance.

528 *In silico* modelling of germline variants within the canonical degron that cause classical SGS
529 has suggested effects on the interaction between the degron of β -catenin, which has similar
530 sequence to the β TrCP1 binding site in the SETBP1 degron, and ubiquitin E3 ligase
531 β TrCP1⁶. To investigate whether the observed differences in protein degradation were due
532 to alterations in SETBP1 ubiquitination, we performed immunoprecipitation of FLAG-
533 SETBP1 and assessed its ubiquitin level. Even though impaired proteasome degradation
534 and autophagy were observed in two classical SGS variants and several variants outside the
535 degron (Figure 4C), ubiquitination was not significantly reduced in the majority of the tested
536 variants [p.(Gly870Ser), p.(Glu858Lys), p.(Leu957Pro), and p.(Thr962del)] (Figures 4E and
537 4F). Intriguingly, variants furthest from the degron showed significantly lower SETBP1
538 ubiquitination compared to wild type (Figures 4E and 4F), consistent with the degradation
539 assay results (Figure 4C). Although based on *in silico* modelling of interaction with ubiquitin
540 E3 ligase β TrCP1⁶, the classical SGS variant p.(Asp868Asn) would be expected to show the
541 strongest disruption in SETBP1 degradation, followed by p.(Gly870Ser), we did not see such
542 a pattern in our proteasome and autophagy inhibition assays, nor in the level of
543 ubiquitination. Taken together, these results suggest that accumulations of SETBP1 protein
544 observed for a subset of variants are caused by variable disruptions in SETBP1 protein
545 degradation via the proteasome and autophagy pathways. Other mechanisms are likely to
546 contribute to higher SETBP1 protein levels in addition to pathways involving ubiquitination.

547 **SETBP1 variants outside the degron, but not SGS variants, reduce binding to AT-rich**
548 **DNA sequences and transcriptional activation**

549 SETBP1 can bind to genomic DNA via its AT-hooks and function as a regulator of
550 transcription²⁰. We went on to assess the effects of SETBP1 variants on the capacity of the
551 protein to bind to AT-rich DNA sequences, using a luciferase reporter system. We generated
552 two luciferase reporters inserted respectively with six repeats of the previously reported
553 consensus AT-rich DNA binding sequences of SETBP1 (5'-AAAATAA-3' or 5'-AAAATAT-
554 3')²⁰. The majority of the variants tested could still bind to AT-rich DNA sequences (Figure
555 5A). Of note, both variants [p.(Leu957Pro) and p.(Thr962del)] located close to the HCF
556 binding domain (amino acids 991-994) showed significantly reduced AT-rich sequence
557 binding capacity (Figure 5A).

558 We next used a mammalian-one-hybrid (M1H) assay to further delineate whether SETBP1
559 can induce transcriptional activity in the proximity of promoter regions without direct DNA
560 binding. Wild type SETBP1 fused with GAL4 showed significantly increased luciferase
561 activity compared to empty controls and a reporter construct without a GAL4-binding site
562 (Figure 5B). These results confirmed the capacity of the protein to activate transcription in
563 the vicinity of a promoter region without direct binding to DNA, consistent with its role as a
564 chromatin remodeller. The majority of the variants could activate transcription (Figure 5B).
565 Interestingly, two SGS variants and two variants close to the degron [p.(Glu862Lys) and
566 p.(Leu957Pro)] showed significantly higher transcriptional activity compared to wild type
567 (Figure 5B). In contrast, the two variants furthest from the degron [p.(Ser444Arg) and
568 p.(Thr962del)] failed to activate transcription, appearing to be loss-of-function (Figure 5B).

569 Moreover, our experiments identified *FOXP2*, rare genetic disruptions of which lead to
570 CAS⁴²⁻⁴⁴, as a novel direct transcriptional target of SETBP1 (Figure 5C). Most of the variants
571 that we tested led to reduced *FOXP2* transcription activation (Figure 5C). Notably,
572 p.(Thr962del), which lacks only one threonine residue in the encoded SETBP1 protein,

573 resulted in complete loss of function in all of our luciferase reporter assays, highlighting the
574 importance of this residue for SETBP1 transcriptional activity. SETBP1 has been shown to
575 be a largely nuclear protein and so its potential mislocalization could lead to disruption of its
576 function. However, we found that all *SETBP1* variants localized to the nucleus as puncta
577 similar to wild type when assessed in transiently transfected HEK293T/17 cells and
578 endogenously in patient fibroblasts (Figure S5). Taken together, these data suggest that
579 pathogenic *SETBP1* variants outside the degron reduce AT-rich DNA binding capacity and
580 transcriptional activity of SETBP1 within the nucleus while preserving gross intracellular
581 localization.

582 **Patient fibroblasts carrying *SETBP1* variants outside the degron show increased** 583 **proliferation**

584 Increased cell proliferation has been reported in EBV-transformed lymphoblastoid cell lines
585 (LCLs) derived from patients carrying germline classical SGS variants⁶ and in leukaemic
586 cells with somatic *SETBP1* variants that drive development of myeloid malignancies^{6,45}. We
587 therefore investigated the proliferation of fibroblasts derived from three individuals carrying a
588 germline *SETBP1* variant outside the degron [p.(Glu858Lys), p.(Leu957Pro), and
589 p.(Thr962del)] compared to those from sex-matched healthy controls. For two of the three
590 variants, we observed that fibroblasts displayed significantly faster proliferation (Figures 6A
591 and S10A) and shorter doubling time than healthy controls in a time course experiment
592 (Figure 6B). Interaction between SETBP1 and SET has been shown to stabilise SET,
593 protecting SET from cleavage by protease, subsequently inhibiting PP2A activity and
594 therefore promoting proliferation in HEK and leukaemia cells^{7,18,45}. To determine whether
595 SETBP1 interaction with SET is affected by variants, we performed co-immunoprecipitation
596 assays in HEK293T/17 cells co-transfected with GFP-SET and FLAG-SETBP1 variants. We
597 observed more abundant GFP-SET expression with increasing SETBP1 levels when co-
598 expressed with FLAG-SETBP1 variants [p.(Leu957Pro) and p.(Thr962del)] compared to wild
599 type (Figures 6C, 6D and S10B), consistent with the previously reported upregulation of SET

600 with increasing SETBP1 levels¹⁸. We showed that mutated versions of FLAG-SETBP1,
601 including those that led to faster fibroblast proliferation, retained interaction with GFP-SET
602 similar to wild type (Figures 6C and 6D). Unexpectedly, we saw no differences from wild type
603 for cell proliferation (Figure 6A) and interaction with SET (Figures 6C and 6D) in cells
604 carrying a recurrent missense variant [p.(Glu858Lys)] which has also been reported in
605 leukaemia cells in atypical chronic myeloid leukaemia (aCML) patients⁷. Moreover, we did
606 not find any differences in *SET* expression between patient and control fibroblasts (Figure
607 S10C), further suggesting that the aetiology involving identical variants in germline and
608 somatic cells is likely to be cell-type specific. Overall, a subset of germline *SETBP1* variants
609 outside the degron promote cell proliferation via a mechanism that is not driven by
610 alterations in SET/SETBP1 interaction.

611 **Patient fibroblasts carrying germline *de novo* SETBP1 variants outside the degron**
612 **show distinct transcriptomic profiles from healthy controls**

613 To assess whether the observed *SETBP1* variants lead to a distinct gene expression
614 signature compared to wild type, we performed RNA-seq on fibroblasts derived from three
615 individuals carrying germline *SETBP1* variants outside the degron [p.(Glu858Lys) (female),
616 p.(Leu957Pro) (male), and p.(Thr962del) (female)] as well as four healthy controls (two
617 females and two males). Principal component analysis (PCA) revealed that transcriptomic
618 profiles of patient fibroblasts did not cluster with those from healthy individuals (Figure 7A).
619 We performed differential gene expression analysis comparing patient cells in females and
620 males separately to controls. This analysis identified 4,110 differentially expressed genes
621 (DEGs) in females (48.1% upregulated) and 2,479 DEGs in males (47.2% upregulated)
622 ($p < 0.05$, FDR) (Figure 7B, Tables S4 and S5). To further narrow down DEGs of interest, we
623 used an intensity difference test, which assesses the statistical magnitude of effect, to find
624 genes that showed unusually highly changing values⁴⁶ in our samples. There were 218
625 (female) and 264 (male) DEGs that showed significant intensity differences ($p < 0.05$, FDR)
626 compared to healthy controls (Figure 7B, Tables S6 and S7). 32 DEGs ($p < 0.05$, FDR)

627 showed significant intensity differences ($p < 0.05$, FDR) that were present in both female and
628 male patient fibroblasts (Figures 7C and 7D, Table S8). Relative quantification of a subset of
629 target genes (*BDKRB2*, *BRINP1*, *CHRM2*, *KRT19*, *RUNX3*, *SFRP2* and *LHX9*) by means of
630 RT-qPCR confirmed the differential expression detected by RNA-seq (Figure 7E). We next
631 performed gene ontology and gene set enrichment analyses (GSEA)^{38,47} of the consistent
632 DEGs to delineate the most relevant biological pathways and cellular components.
633 Functional annotation demonstrated over-representation ($p < 0.05$, Benjamini-Hochberg FDR)
634 of ontologies related to tissue development, cell proliferation and differentiation, cell surface
635 receptor signalling, and response to stimuli (Figure 7F and Table S9). GSEA showed
636 significant enrichment in patient samples of four gene sets including the G protein-coupled
637 receptor signalling pathway, chemical homeostasis, ion transport and secretion (FDR < 25%;
638 Figure S11E and Table S10). Cell proliferation was also found as one of the dysregulated
639 pathways (Figures 7F and Table S9), consistent with the faster cell proliferation observed in
640 patient fibroblasts (Figures 6A and 6B). Of note, enrichment of signalling pathways and
641 cellular components comprising plasma membrane and vesicles was also detected in results
642 of different GO analyses (Figures 7F and S11E). Six of the 32 differentially expressed genes
643 (*ABCC9*, *ADGRG6*, *COMP*, *ELN*, *HLA-A*, and *MITF*) were present (significant overlap, p -
644 value = 0.036, Fisher's exact test) in the PanelApp database (intellectual disability v3.2 and
645 autism v.0.22) (Figure S11A) showing their association with neurodevelopmental disorders
646 including autism and intellectual disability. Nine of the genes (*RUNX3*, *SFRP2*, *CA12*, *SIM2*,
647 *CPE*, *SCIN*, *FAM118A*, *ADGRG6*, and *PMEPA1*) were also dysregulated in HEK cells
648 overexpressing an SGS variant [p.(Gly870Ser)] compared to empty controls, in a prior
649 study²⁰, suggesting a partial overlap with classical SGS pathology (Figure S11B).
650 Furthermore, we cross-referenced these 32 DEGS with 70 putative SETBP1 target genes
651 that were previously reported²⁰ and identified *RUNX3*, a regulator of chromatin dynamics⁴⁸,
652 as a potential direct transcriptional target that was dysregulated in patient fibroblasts carrying
653 *SETBP1* variants outside the degron (Figure S11C). Several direct transcriptional targets of
654 SETBP1 (*MECOM*, *RUNX1*, *HOXA9*, *HOXA10* and *MYB*) have shown differential expression

655 in leukaemia cells from aCML patients²⁰⁻²³ and in HEK cells overexpressing the
656 p.(Gly870Ser) variant²⁰. However, we did not observe differential expression of these genes
657 in our RNA-seq data (Tables S4-8), further suggesting that the aetiological pathways are
658 likely to be cell-type specific. Taken together, *SETBP1* variants outside the degron are
659 associated with distinct transcriptomic profiles from healthy individuals and linked to
660 ontologies related to tissue development, cellular signalling, transmembrane transport and
661 membrane structure.

662 **Discussion**

663 Our study demonstrates that variant-specific functional follow-up is crucial to understand
664 biological underpinnings of overlapping phenotypes and heterogeneity within cohorts. Work
665 from our group and others have previously delineated the phenotypic heterogeneity within
666 cohorts carrying germline *SETBP1* loss-of-function variants causing *SETBP1*
667 haploinsufficiency disorder, which have much milder clinical correlates than classical
668 SGS^{6,15,16}. Here, we present, to our knowledge, the largest cohort of individuals who have
669 *SETBP1* variants (missense and in-frame deletion) outside the canonical degron region, with
670 clinical features that do not fit with the original diagnostic criteria of classical SGS. We used
671 a genotype-driven approach combining deep clinical and speech phenotyping with functional
672 follow-up of specific variants to investigate the phenotype-genotype associations and
673 delineate the underlying pathophysiological mechanisms. Our results show that individuals
674 carrying variants outside the degron display a broad spectrum of clinical features of variable
675 severity that only partially overlap with either SGS or *SETBP1* haploinsufficiency disorder.
676 This wide clinical spectrum could be explained by the heterogeneous pathophysiological
677 mechanisms resulting from these genetic variations, as shown in the cell-based
678 experiments. Using an array of functional assays, we showed that variants outside the
679 degron disrupt *SETBP1* protein functions via aetiological mechanisms including impairments
680 in ubiquitination, DNA binding capacity, transcription and cell proliferation, which are
681 independent of *SETBP1* protein levels. Interestingly, we found that *SETBP1* variants led to

682 reduced transcription of its direct target *FOXP2*, of which rare variants are a known cause of
683 monogenic speech disorder characterised by CAS⁴²⁻⁴⁴.

684 **Variable severity of broad clinical features depends on proximity of variants to the**
685 **degron**

686 The variable severity of the broad spectrum of clinical features observed in our cohort could
687 largely be categorised into three groups based on the distance of the variant from the
688 canonical degron. We found that patients with variants in close proximity to the degron within
689 the SKI domain (affecting amino acids 862-874, excluding the degron 868-871) showed the
690 most severe phenotype with severe intellectual disability, inability to speak and are often
691 unable to walk. These individuals are much more severely affected than those with *SETBP1*
692 haploinsufficiency disorder¹⁶ and yet do not fulfil the original diagnostic criteria of classical
693 SGS as proposed by Lehman *et al.*⁴¹, which were proposed at a time when the causative
694 gene was not identified and may have led to an ascertainment bias in the first studies after
695 identification of the gene^{2,6}. In 2020, Sullivan *et al.*⁴⁹ described an individual with a *SETBP1*
696 variant affecting amino acid 871 within the hotspot region, who displayed an atypical milder
697 phenotype which was not concordant with the original diagnostic criteria of SGS. She had a
698 moderate intellectual disability, no congenital anomalies, and showed less apparent
699 dysmorphisms than patients with classical SGS including mild midface retraction,
700 hypertelorism, short upturned nose and prominent low-set ears. This more mildly affected
701 individual than those with classical SGS may point to a broader SGS-associated phenotype
702 than originally defined, with a different (non-lethal) life expectancy, possibly involving
703 interactions with other variants in the genomic background. In addition, this wider spectrum
704 of SGS may also be applicable to cases with aetiological mutations in the 862-874 region
705 who displayed similar facial features including mild midface retraction, microcephaly and
706 prominent low-set ears.

707 The group of *SETBP1* variants affecting residues at positions 854-858 (located further from
708 the degron) showed a phenotype including mild to moderate intellectual disability. Motor
709 development was delayed but they all achieved the milestone to walk without support.
710 Growth was normal and we did not observe a consistent pattern of dysmorphisms, but this
711 may be partly because we had only a limited number of patient photographs, and these
712 came from different age ranges. Future studies including a larger cohort whose facial
713 features are examined at standardised age ranges longitudinally (with photographs taken
714 from different angles) will be beneficial to profiling any dysmorphism and thus improving
715 facial feature-based clinical diagnosis. The p.(Glu858Lys) variant was noted in four
716 individuals. Leonardi *et al.*⁵⁰ also described an individual with the same variant, showing
717 severe intellectual disability and epilepsy. In that study, the variant was reported to be
718 inherited from an asymptomatic mother whose variant was *de novo*. Unfortunately, no other
719 tissues were tested in the mother to exclude the possibility of mosaicism.

720 Individuals with variants located outside the SKI domain (n=4) showed a more variable
721 clinical phenotype ranging from mild to severe intellectual disability. Two of these individuals
722 [carrying p.(Leu957Pro) and p.(Thr962del) variants] showed a very similar facial phenotype
723 including a round face, blepharophimosis, hypertelorism and a short nose with a bulbous tip,
724 features also often noted in individuals with *SETBP1* haploinsufficiency disorder. The two
725 individuals with variants located furthest from the SKI domain [p.(Ser444Arg) and
726 p.(Val657Ala)] had inherited them from affected mothers who also have low IQ
727 (grandparents and other family members were not available for testing). We therefore also
728 included these variants in the functional cellular assays together with the *de novo* variants
729 outside the degron in our study. Indeed, these two variants showed significant functional
730 impacts on the *SETBP1* protein as shown in our cellular assays. In light of this, we
731 recommend that clinicians should examine the clinical phenotypes of family members
732 carrying the same variants as well as performing cellular assays to test pathogenicity when
733 assessing pathogenicity of inherited variants of uncertain significance based on ACMG

734 (American College of Medical Genetics and Genomics) guidelines⁵¹. Clinicians could also
735 consider performing functional prediction using algorithms where available, such as those for
736 known epilepsy ion channel genes⁵².

737 The variable phenotypic severity and partial clinical overlaps with SGS or *SETBP1*
738 haploinsufficiency disorder suggest the potential existence of a third clinical entity involving
739 disruptions of this gene. Alternatively, different types of *SETBP1* variants could lead to a
740 continuum of clinical features of SETBP1-related disorders with SGS and *SETBP1*
741 haploinsufficiency disorder positioned at two extremes of severity. Investigations of future
742 larger cohorts including individuals with different variants in combination with genome-wide
743 analysis, for example, DNA methylation analysis (episignatures)^{53,54}, will provide clarification
744 over the clinical definition and aetiology, as well as improving clinical diagnosis. Our
745 participants had long-term chronic difficulties across a number of developmental domains.
746 Individuals with *SETBP1* variants outside the degron region should receive careful
747 assessment across core domains of speech and language, attention, motor and sleep as
748 early as possible following diagnosis, leading to better targeted and earlier intervention to
749 optimise children's health and developmental outcomes.

750 ***FOXP2* is a direct transcriptional target of SETBP1**

751 Individuals with *SETBP1* haploinsufficiency disorder showed significant impairments in both
752 receptive and expressive language, suggesting SETBP1 as a strong candidate for speech
753 and language disorders¹⁵. Of note, we have here demonstrated that individuals carrying
754 *SETBP1* variants outside the degron, where speech and language data were available,
755 showed generally low language ability for all subdomains including expressive, receptive,
756 written and social language. Intriguingly, our cell-based experiments identified *FOXP2* as a
757 novel direct transcriptional target of SETBP1 (Figure 5C). All functionally tested *SETBP1*
758 variants outside the degron displayed significantly decreased *FOXP2* transcription in
759 luciferase reporter assays. *FOXP2* is one of the few genes identified to-date for which

760 disruptions yield disproportionate effects on speech and language, yet there is little known
761 about its upstream regulators. Our results provide molecular evidence linking two strong
762 candidate genes implicated in childhood apraxia of speech, which might help provide insight
763 into the severe speech and language impairments observed in our, and other, cohorts. We
764 did not detect *FOXP2* as a differentially expressed gene in our transcriptomic analysis of
765 patient fibroblasts or in those performed on patient leukaemia cells and cell lines expressing
766 *SETBP1* constructs in previous studies, further pointing towards potential cell type-specific
767 aetiological pathways.

768 **A broad range of functions are disrupted by variants outside the degron via**
769 **pathophysiological mechanisms independent of SETBP1 protein dosage**

770 As seen in the summary of our functional analyses (Figure 8), a broad spectrum of cellular
771 functions was affected. In these cellular assays, variants outside the degron led to molecular
772 consequences distinct from classical SGS variants. Classical SGS variants showed
773 increased protein stability and higher SETBP1 protein levels as a result of impaired
774 degradation by proteasome and autophagy. Moreover, while the binding capacity of SGS
775 variants to AT-rich DNA sequences was not affected, a subset of SGS variants showed
776 higher transcriptional activation, suggesting gain of function. These findings are consistent
777 with prior studies showing global upregulation of SETBP1 binding to genomic regions and
778 increased chromatin accessibility when overexpressing an SGS variant [p.(Gly870Ser)] in
779 HEK cells²⁰.

780 In contrast, *SETBP1* variants outside the degron demonstrated an array of functional
781 disruptions that only partially overlapped with those observed for SGS variants. Although the
782 majority of variants outside the degron resulted in more abundant SETBP1, there were
783 variable extents of disruption in degradation of this protein via the proteasome and
784 autophagy machinery, suggesting that impaired protein degradation might not be a core
785 pathogenic mechanism. The p.(Glu862Lys) variant that is in close proximity to the degron

786 led to increased protein levels and transcriptional activation, similar to SGS variants, but
787 neither protein stability nor degradation were affected. This points to a functional landscape
788 partially overlapping with SGS variants but distinct from the other variants located further
789 from the degron. The variant further away from the degron but still in the SKI domain
790 [p.(Glu858Lys)] led to more abundant SETBP1 protein, reduced degradation via proteasome
791 and autophagy and lower *FOXP2* transcription. However, even though p.(Glu858Lys)
792 showed functional results overlapping with those of SGS variants, the individual carrying the
793 variant still presented milder clinical features compared to classical SGS (amino acids 868-
794 871) and to those with variants nearest to the degron (amino acids 862-867 and 873-874).

795 Interestingly, functional impairments of variants further away from SKI domain could be
796 largely divided into two groups, matching with the variable clinical features observed in the
797 individuals that carried them. In our cell-based assays, two variants [p.(Leu957Pro) and
798 p.(Thr962del)] led to a loss of DNA-binding and transcriptional activity despite the increased
799 protein levels detected in transfected cells and patient fibroblasts, consistent with the
800 observed clinical features being closer to individuals with *SETBP1* haploinsufficiency
801 disorder. These two variants are located close to the HCF binding site (amino acids 991–
802 994), which is an important component of the COMPASS complex. Piazza *et al.* reported
803 that mutant SETBP1 still retained the ability to bind to HCF, PHF6/8 and KMT2A in HEK
804 cells overexpressing an SGS variant [p.(Gly870Ser)]²⁰. This raises a hypothesis that the loss
805 of transcriptional activity for variants such as p.(Leu957Pro) and p.(Thr962del) could result
806 from disruption of interaction with HCF, a core protein of the SET1/KMT2A COMPASS-like
807 complex responsible for H3K4 mono- and di-methylation, and chromatin accessibility^{20,55}. In
808 particular, the threonine residue at the 962 position is likely crucial for SETBP1
809 transcriptional activity, potentially via HCF1-related pathways given its proximity to the HCF-
810 binding site and the complete loss of transcriptional activation in cells expressing
811 p.(Thr962del). Intriguingly, variants that were furthest from the degron region [p.(Ser444Arg)
812 and p.(Val657Ala)] showed mostly impaired SETBP1 degradation by both proteasome and

813 autophagy due to disrupted ubiquitination. It is possible that these variants are physically
814 closer to the degron in the three-dimensional conformation of the protein. However, we could
815 not predict how *SETBP1* variants might affect its protein folding or interactions with
816 interactors due to limited knowledge about the overall structure of the protein.

817 SETBP1 has been identified as a protein with intrinsically disordered regions (IDRs), which
818 are regions that are prone to mutations and are often found in proteins associated with
819 cancers⁵⁶. Indeed, the majority of *SETBP1* missense variants identified to-date, both somatic
820 and germline, are located within an IDR (amino acids 858-880)⁵⁶. Many proteins containing
821 IDRs do not have a single well-defined conformation, and the conformation changes
822 depending on interactions with cofactors. Most proteins with IDRs that are elements of the
823 transcriptional machinery are multifunctional, involved in processes such as regulating
824 degradation via a linear motif, binding to genomic DNA, activating/repressing transcription,
825 and/or modulating histone modifications⁵⁶. This wide range of functions could explain the
826 broad spectrum of impairments found in our cell-based assays and make it more difficult to
827 predict how variants affect SETBP1 function. Future studies that profile and delineate the
828 structural impact of *SETBP1* variants and how they affect interactions with cofactors will aid
829 the understanding of their impacts on protein functions and thus aetiology.

830 **Cell-type specific aetiology of *SETBP1* variants**

831 We identified a subset of differentially expressed genes in our RNA-seq data that also
832 showed differential expression in prior analyses of p.(Gly870Ser) expressing cells²⁰,
833 suggesting partly shared aetiological pathways with SGS, complementing the observed
834 partial overlap in clinical manifestation. Nevertheless, we did not observe overlaps with
835 genes dysregulated in aCML patient cells, nor did we observe any changes in proliferation in
836 fibroblasts derived from individuals carrying recurrent variant p.(Glu858Lys) that has also
837 been reported in individuals with leukaemia⁷. This is consistent with the findings of a recent
838 study where disruption of SET/PP2A pathway was not detected in SGS neural progenitors

839 and cortical organoids⁸, suggesting that the aetiological pathways underlying germline and
840 somatic *SETBP1* variants are likely cell type-/tissue-specific. Future studies investigating the
841 spatial and temporal expression of *SETBP1* and the functional impact of variants during
842 brain development are much needed to understand the pathways that go awry. This could
843 potentially be achieved by using brain organoids that carry *SETBP1* variants. A recent study
844 investigating SGS variants in young cortical organoids and neural progenitors suggested
845 p53-related pathways as a novel mechanism involved in this disorder⁸. *SETBP1* knockout
846 neural progenitors showed protracted proliferation and distorted layer-specific neuronal
847 differentiation with overall decrease in neurogenesis via WNT/ β -catenin signaling pathway in
848 a recent BioRxiv preprint²⁴. These investigations focused on relatively young neural
849 progenitors as well as either an SGS variant or gene deletions. Future work comparing the
850 functional impacts of different types of variants in neuronal models is essential to delineate
851 the aetiology, complexity and pleiotropy of *SETBP1*-related disorders.

852 ***SETBP1* is a central upstream element of biological networks disturbed in**
853 **neurodevelopmental disorders**

854 Our RNA-seq analysis of patient fibroblasts carrying variants outside the degron revealed a
855 number of DEGs including transcription factors that are important for neurodevelopment and
856 associated with intellectual disability and/or autism (Figure S11D). Previous chromatin
857 immunoprecipitation and mass spectrometry analyses in HEK cells overexpressing *SETBP1*
858 have shown its role as a potential epigenetic hub²⁰. Moreover, *SETBP1* directly regulates
859 transcription of *FOXP2*, which has been shown to interact with an array of transcription
860 factors implicated in neurodevelopmental disorders characterised by speech and language
861 deficits^{57,58}. Given the broad expression of *SETBP1* in neural progenitors and neurons, and
862 its roles as a chromatin remodeller and transcription factor, our findings further suggest that
863 this protein is a central element of biological networks regulating brain development, and
864 could be important in pathophysiology of neurodevelopmental disorders caused by genetic
865 disruptions of its interactors or downstream targets.

866 In conclusion, in this study, we investigated the genotype-phenotype associations of
867 germline *SETBP1* variants by thoroughly profiling the clinical and speech-language
868 phenotypes. We have established an array of functional assays for testing pathogenicity of
869 the identified variants, which can be applied to variants of uncertain significance in the
870 future. Despite the partial overlap of results in clinical and functional analyses with SGS, our
871 data suggest that *SETBP1* variants outside the degron cause a clinically and functionally
872 variable form of neurodevelopmental disorder that is milder than SGS and in certain aspects
873 more severe than *SETBP1* haploinsufficiency disorder. In particular, we have shown
874 impairments in ubiquitination, DNA binding capacity, transcription and cell proliferation
875 depending on the proximity of variants relative to the degron in the SKI domain. Overall, our
876 findings reveal that variants outside the degron act via a range of loss-of-function
877 pathophysiological mechanisms independent of protein abundance, unlike SGS and
878 *SETBP1* haploinsufficiency disorder, providing valuable new insights into diagnosis and
879 aetiology of SETBP1-related disorders.

880 **Supplemental Data**

881 Supplemental Data include 11 figures and 12 tables.

882 **Declaration of Interests**

883 The authors declare no other competing interests.

884 **Acknowledgments**

885 We are grateful to all individuals and families for their contribution. We would like to thank
886 the members of the Cell Culture Facility, Department of Human Genetics, Radboud
887 university medical center, Nijmegen, for cell culture of proband-derived cell lines. We
888 especially thank Joery den Hoed, Else Eising (Language and Genetics Department, MPI for
889 Psycholinguistics, Nijmegen), Alexander Hoischen, Lot Snijders-Blok (Department of Human
890 Genetics, Radboud university medical center, Nijmegen) and Rocio Acuna-Hidalgo (Nostos

891 Genomics, Berlin) for scientific discussion. This work was supported by the Max Planck
892 Society (M.M.K.W., R.A.K., G.A., A.V., and S.E.F.). W.K.C was funded by SFARI and the
893 JPB Foundation. M.S.H., I.E.S., and A.T.M. were funded by a National Health and Medical
894 Research Council (NHMRC) Centre of Research Excellence Grant (1116976), an Australian
895 Research Council (ARC) Discovery Project (DP120100285), an NHMRC Project Grant
896 (1127144), and the March of Dimes Grant Scheme. M.S.H. was funded by an NHMRC
897 Career Development Fellowship (1063799). I.E.S. was funded by an NHMRC Investigator
898 Grant (1172897), an NHMRC Practitioner Fellowship (1006110), and NHMRC Development
899 Grant (1153614). A.T.M. was funded by an NHMRC Investigator Grant (1195955), an
900 NHMRC Practitioner Fellowship (1105008), and an NHMRC Development Grant (1153614).
901 E.P. was supported by the National Health and Medical Research Council (GNT11149630),
902 Australia and the research WGS was supported by NHMRC.

903 **Web Resources**

904 CADD, <https://cadd.gs.washington.edu>

905 gnomAD, <https://gnomad.broadinstitute.org>

906 MetaDome, <https://stuart.radboudumc.nl/metadome>

907 OMIM, <https://www.omim.org>

908 UniProt, <https://www.uniprot.org>

909 Spatial Clustering, <https://github.com/laurensvdwiels/SpatialClustering>

910 g:GOST (part of g:Profiler), <https://biit.cs.ut.ee/gprofiler/gost>

911 The Gene Ontology Resource, <http://geneontology.org/docs/go-enrichment-analysis/>

912 COSMIC: <https://cancer.sanger.ac.uk/cosmic>

913 Seqmonk: <https://www.bioinformatics.babraham.ac.uk/projects/seqmonk/>

914 **Data and Code Availability**

915 Code used in the spatial clustering analysis is available at:

916 <https://github.com/laurensvdwiel/SpatialClustering>³³. Codes of RNA-seq data analysis are

917 available on request.

918 **Figure Legends**

919 **Figure 1: *SETBP1* variants outside the degron cluster in the SKI domain and are**
920 **highly conserved across species.** (A) Schematic representation of the SETBP1 protein
921 (UniProt: Q9Y6X0) indicating the locations of variants included in this study. These comprise
922 eleven novel germline variants (circles) including ten missense variants (blue) and one in-
923 frame deletion (*de novo*, green) outside the canonical degron. Three missense variants
924 outside (blue) and nine within (orange) the degron that were previously reported (diamond)
925 are also annotated in the schematic. § represent variants for which speech phenotyping data
926 were available. # represent variants included in functional assays. Five exons (black bars)
927 encode isoform A of the protein (NP_056374.2, 1596 amino acids). The SETBP1 protein
928 sequence contains three AT hook domains (Ath; orange; amino acids 584–596, 1016–1028,
929 1451–1463), a SKI homologous region (SKI; purple; amino acids 706–917), an HCF1-
930 binding motif (HCF; black; amino acids 991–994), a SET-binding domain (SET; green; amino
931 acids 1292–1488), three bipartite NLS motifs (brown; amino acids 462-477, 1370-1384,
932 1383-1399), six PEST sequences (grey; amino acids 1-13, 269-280, 548-561, 678-689, 806-
933 830, 1502-1526) and a repeat domain (Rpt; black; amino acids 1520–1543)^{7,19,20}. An
934 overview with variant details per subject is provided in Table 1. (B) Sequence alignment of
935 the region containing part of the SETBP1 amino acid sequence in human (Uniprot:
936 Q9Y6X0), chimpanzee (H2QEG8), mouse (Q9Z180), chicken (A0A1D5PT15), African
937 clawed frog (F6TBV9) and zebrafish (B0R147). The canonical degron is highlighted in bold.

938 Residues in which germline missense variants located within (orange), outside the degnon
939 (blue) and deleted by an in-frame deletion (green) are highlighted.

940 **Figure 2: Clinical evaluation of individuals with *SETBP1* variants.** Not available in this
941 preprint.

942 **Figure 3: Performance on Vineland-3 subtests.** Lines denote median scores; X denotes
943 mean scores; ABC, Adaptive Behaviour Composite. Standard scores between 85 and 115
944 are considered within the average range, with a mean of 100 and a standard deviation of 15.

945 **Figure 4: Variable impairment of *SETBP1* degradation via proteasome and autophagy
946 pathways is associated with partial reduction of ubiquitination.** (A) Immunoblot of whole
947 cell lysates of HEK293T/17 cells expressing FLAG-tagged *SETBP1* variants probed with
948 anti-*SETBP1* and anti-FLAG antibodies. β -actin was used as a loading control. Low and high
949 exposures of immunoblot probed with anti-FLAG antibody are shown. Representative blots
950 of three independent experiments are shown. (B) Quantification of protein levels of FLAG-
951 tagged *SETBP1* variants normalised to β -actin. Values are expressed relative to wild type
952 (WT) and represent the mean \pm SEM of three independent experiments ($*p < 0.05$, $**p < 0.01$,
953 $***p < 0.001$, using one-way ANOVA and a *post-hoc* Dunnett's test). (C) Relative
954 fluorescence intensity of YFP-tagged *SETBP1* variants overexpressed in HEK293T/17 cells
955 treated with translation inhibitor cycloheximide (CHX; 50 μ g/ml; top), proteasomal
956 degradation inhibitor MG132 (5 μ g/ml; middle), or autophagy inhibitor Bafilomycin A1 (BafA1;
957 100nM; bottom). An equal volume of DMSO was used as a vehicle control. Fluorescence
958 intensity was measured for 24 hours with three-hour intervals and normalised to an mCherry
959 transfection control. Values are expressed relative to t= 0 hour and represent the mean \pm
960 SEM of three independent experiments, each preformed in triplicate ($*p < 0.05$, $**p < 0.01$,
961 $***p < 0.001$, $****p < 0.0001$; two-way ANOVA and a *post-hoc* Dunnett's test). (D) Immunoblot
962 of whole cell lysates of control and patient human dermal fibroblasts (HDF) probed with anti-
963 *SETBP1* antibody. β -actin was used as a loading control (top). Normalised *SETBP1*

964 transcript expression (bottom) in control and patient HDFs. Bars represent mean \pm SEM of
965 three independent experiments (n.s., $*p < 0.05$, versus WT; one-way ANOVA and a *post-hoc*
966 Dunnett's test). (E) Immunoprecipitation of FLAG-tagged SETBP1 wild type and variants
967 using FLAG-conjugated magnetic agarose and blotted with an anti-FLAG, anti-SETBP1 or
968 anti-ubiquitin antibody. β -actin was used as a loading control. (F) Quantification of SETBP1
969 (top), ubiquitin (middle) in FLAG-IP fractions for inherited (outside degron, grey) and SGS
970 variants (orange, left), and *de novo* variants outside the degron (right). Ratio of
971 ubiquitin/SETBP1 in the FLAG-IP fractions was plotted (bottom). Bars represent mean \pm SEM
972 of three independent experiments ($*p < 0.05$, versus WT; one-way ANOVA and a *post-hoc*
973 Dunnett's test).

974 **Figure 5: Genotype-specific reduction of binding capacity to AT-rich DNA sequences**
975 **and transcriptional activation for *SETBP1* variants outside the degron.** (A) Results of
976 luciferase assays with constructs containing WT and *SETBP1* variants, and the reporter
977 constructs with the consensus SETBP1 binding sequences. Values are expressed relative to
978 the control condition which used a pCMV-YFP construct without SETBP1. (B) Results of the
979 M1H assay for SETBP1 transcriptional regulatory activity with WT and *SETBP1* variants
980 fused with an N-terminal GAL4 in combination with a reporter construct with or without the
981 GAL4-binding site. Values are expressed relative to the control condition which used a
982 pBIND2-GAL4 construct without SETBP1. (C) Results of luciferase assays with constructs
983 containing WT and *SETBP1* variants, and reporter constructs with *FOXP2* promoters: TSS1
984 (left) and TSS2 (right). Values are expressed relative to the control condition which used a
985 pCMV-YFP construct without SETBP1. All graphs for luciferase assays show the mean \pm
986 SEM of three independent experiments, each performed in triplicate ($*p < 0.05$, $**p < 0.01$,
987 $***p < 0.001$, $****p < 0.0001$ versus WT; one-way ANOVA and a *post-hoc* Dunnett's test). All
988 graphs for the mammalian-one-hybrid assay show the mean \pm SEM of three independent
989 experiments, each performed in triplicate ($*p < 0.05$, $***p < 0.001$, $****p < 0.0001$ versus reporter

990 without GAL4-binding site; # $p < 0.05$, ### $p < 0.001$, #### $p < 0.0001$ versus WT; two-way ANOVA
991 and a *post-hoc* Tukey's test).

992 **Figure 6: Increased proliferation of fibroblasts derived from patients carrying *SETBP1***
993 **variants outside the canonical degron.** (A) CyQuant cell proliferation assay of fibroblasts
994 from healthy individuals (controls) and patients carrying a *SETBP1* variant outside the
995 degron. Nuclei of fibroblasts were stained with a GFP fluorescence dye. Fluorescence
996 activity was measured daily for four days. Values are expressed relative to 1 day after
997 seeding (day 1) and represent the mean \pm SEM of three independent experiments, each
998 performed in triplicate (* $p < 0.05$, *** $p < 0.001$ versus controls; two-way ANOVA and a *post-hoc*
999 Tukey's test). (B) Cell doubling time of fibroblasts. Values represent the mean \pm SEM of
1000 three independent experiments (* $p < 0.05$ versus healthy controls; one-way ANOVA and a
1001 *post-hoc* Dunnett's test). (C) Co-immunoprecipitation (Co-IP) was performed in whole cell
1002 lysates co-expressing FLAG-*SETBP1* and GFP-SET. Wild type FLAG-*SETBP1* and variants
1003 were co-immunoprecipitated using FLAG-conjugated magnetic agarose. Immunoblots were
1004 probed with an anti-FLAG, anti-*SETBP1*, anti-GFP or anti-SET antibody. β -actin was used
1005 as the loading control in the input fraction. (D) Quantification of *SETBP1* (top) and SET
1006 (middle) levels in the input fraction. Quantification of SET co-immunoprecipitated with
1007 *SETBP1* in the IP fraction was plotted (bottom). Values are expressed as the mean \pm SEM
1008 of four independent experiments (** $p < 0.01$, *** $p < 0.001$ versus WT; one-way ANOVA and a
1009 *post-hoc* Dunnett's test).

1010 **Figure 7: Fibroblasts carrying *SETBP1* variants showed distinct transcriptomic**
1011 **profiles from healthy controls.** (A) Fibroblasts derived from patients harbouring a *SETBP1*
1012 variant (missense or in-frame deletion) outside the degron did not cluster with those from
1013 healthy individuals (controls). 3D principal component analysis (PCA) plot of variance
1014 distribution of four control fibroblast lines (blue; two females and two males), three patient
1015 lines (red; two females and one male); three technical replicates were included for each line.
1016 Principal component (PC) 1, PC2 and PC3 showed 25.9%, 17.8% and 13.5% of total

1017 variance respectively. (B) MA-plots showing the gene expression data of control vs. patient
1018 fibroblast lines (left: females; right: males) as a function of log ratios and mean average gene
1019 counts. (C) Venn diagram showing the overlap of genes that demonstrated significant
1020 differential expression ($p < 0.05$, multiple testing correction with FDR) with significant intensity
1021 differences (adjusted $p < 0.05$, multiple testing correction with FDR) in female control vs.
1022 female patient fibroblasts (left, blue) and male control vs. male patient fibroblasts (right, red).
1023 (D) Gene expression heatmap of the 32 DEGS with significant intensity differences in patient
1024 fibroblasts. (E) Validation of RNA-seq DEGs using RT-qPCR. Bars represent the mean
1025 \pm SEM of three independent experiments ($*p < 0.05$, $***p < 0.001$, $****p < 0.0001$ versus female
1026 or male controls; one-way ANOVA and a *post-hoc* Dunnett's test). (F, G) Dysregulated GO
1027 biological process and cellular components revealed by over-representation analysis of the
1028 32 DEGs in patient fibroblasts using (F) over-representation analysis with g:Profiler ($p < 0.05$,
1029 multiple testing correction with FDR).

1030 **Figure 8: Summary of functional assays: variants outside the degron disrupt SETBP1**
1031 **functions by dysregulating transcription, reducing DNA binding capacity and**
1032 **promoting cell proliferation independent of SETBP1 level.** A heat map summarising the
1033 functional characterisation of *SETBP1* variants within (classical SGS) and outside the
1034 degron. Classical SGS variants showed increased protein stability and higher SETBP1
1035 protein levels. While binding capacity of SGS variants to AT-rich DNA sequences was not
1036 affected, a subset showed higher transcriptional activation, suggesting a gain of function. In
1037 contrast, *SETBP1* variants outside the degron demonstrated a broad spectrum of functional
1038 disruptions which could be largely categorised into two groups. Although the majority of
1039 these variants resulted in more abundant SETBP1 protein, there were variable degrees of
1040 disruption in SETBP1 degradation via the proteasome and autophagy machinery. Variants
1041 furthest from the degron (grey) showed mostly reduced SETBP1 degradation by both
1042 proteasome and autophagy due to disrupted ubiquitination, while only a subset of those in
1043 the proximity of the degron demonstrated impaired degradation and mildly reduced

1044 ubiquitination. Two variants outside the degron led to lower affinity to AT-rich DNA
1045 sequences, suggesting loss of function. Transcriptional activation was affected in the
1046 majority of variants outside the degron to various degrees. Increased proliferation was seen
1047 for patient fibroblasts carrying two variants outside the degron, suggesting a partially
1048 overlapping mechanism with classical SGS. Patient fibroblasts carrying variants outside the
1049 degron showed distinct transcriptomic profiles from healthy controls, implicating biological
1050 pathways involved in system/tissue development, cell proliferation and differentiation, cell
1051 surface receptor signalling, and membrane composition.

1052 **Table 1: Summary of clinical phenotypes in individuals with *SETBP1* variants outside the degen.**

1053 Not available in this preprint.

1054 **Table 2: Summary of speech phenotypes in individuals with *SETBP1* missense variants outside the degron.**

1055 Not available in this preprint.

1056 **Permissions**

1057 No permission was needed.

1058 **References**

1059 1. Cogliati, F., Forzano, F., and Russo, S. (2021). Editorial: Overlapping Phenotypes and
1060 Genetic Heterogeneity of Rare Neurodevelopmental Disorders. *Front. Neurol.* *12*, 1031.

1061 2. Hoischen, A., van Bon, B.W.M., Gilissen, C., Arts, P., van Lier, B., Steehouwer, M., de
1062 Vries, P., de Reuver, R., Wieskamp, N., Mortier, G., et al. (2010). De novo mutations of
1063 SETBP1 cause Schinzel-Giedion syndrome. *Nat. Genet.* *42*, 483–485.

1064 3. Schinzel, A., and Giedion, A. (1978). A syndrome of severe midface retraction, multiple
1065 skull anomalies, clubfeet, and cardiac and renal malformations in sibs. *Am. J. Med. Genet.* *1*,
1066 361–375.

1067 4. Minn, D., Christmann, D., De Saint-Martin, A., Alembik, Y., Eliot, M., Mack, G., Fischbach,
1068 M., Flament, J., Veillon, F., and Dollfus, H. (2002). Further clinical and sensorial delineation
1069 of Schinzel-Giedion syndrome: Report of two cases. *Am. J. Med. Genet.* *109*, 211–217.

1070 5. AL-Mudaffer, M., Oley, C., Price, S., Hayes, I., Stewart, A., Hall, C.M., and Reardon, W.
1071 (2008). Clinical and radiological findings in Schinzel–Giedion syndrome. *Eur. J. Pediatr.* *167*,
1072 1399–1407.

1073 6. Acuna-Hidalgo, R., Deriziotis, P., Steehouwer, M., Gilissen, C., Graham, S.A., van Dam,
1074 S., Hoover-Fong, J., Telegrafi, A.B., Destree, A., Smigiel, R., et al. (2017). Overlapping
1075 SETBP1 gain-of-function mutations in Schinzel-Giedion syndrome and hematologic
1076 malignancies. *PLoS Genet.* *13*, e1006683.

- 1077 7. Piazza, R., Valletta, S., Winkelmann, N., Redaelli, S., Spinelli, R., Pirola, A., Antolini, L.,
1078 Mologni, L., Donadoni, C., Papaemmanuil, E., et al. (2013). Recurrent SETBP1 mutations in
1079 atypical chronic myeloid leukemia. *Nat. Genet.* *45*, 18–24.
- 1080 8. Banfi, F., Rubio, A., Zaghi, M., Massimino, L., Fagnocchi, G., Bellini, E., Luoni, M.,
1081 Cancellieri, C., Bagliani, A., Di Resta, C., et al. (2021). SETBP1 accumulation induces P53
1082 inhibition and genotoxic stress in neural progenitors underlying neurodegeneration in
1083 Schinzel-Giedion syndrome. *Nat. Commun.* *12*, 4050.
- 1084 9. Morgan, A., Srivastava, S., Duis, J., and van Bon, B. (2021). SETBP1 Haploinsufficiency
1085 Disorder. In *GeneReviews®*, M.P. Adam, H.H. Ardinger, R.A. Pagon, S.E. Wallace, L.J.
1086 Bean, G. Mirzaa, and A. Amemiya, eds. (Seattle (WA): University of Washington, Seattle), p.
- 1087 10. Filges, I., Shimojima, K., Okamoto, N., Röthlisberger, B., Weber, P., Huber, A.R.,
1088 Nishizawa, T., Datta, A.N., Miny, P., and Yamamoto, T. (2011). Reduced expression by
1089 SETBP1 haploinsufficiency causes developmental and expressive language delay indicating
1090 a phenotype distinct from Schinzel–Giedion syndrome. *J. Med. Genet.* *48*, 117–122.
- 1091 11. Marseglia, G., Scordo, M.R., Pescucci, C., Nannetti, G., Biagini, E., Scandurra, V.,
1092 Gerundino, F., Magi, A., Benelli, M., and Torricelli, F. (2012). 372 kb microdeletion in
1093 18q12.3 causing SETBP1 haploinsufficiency associated with mild mental retardation and
1094 expressive speech impairment. *Eur. J. Med. Genet.* *55*, 216–221.
- 1095 12. Coe, B.P., Witherspoon, K., Rosenfeld, J.A., van Bon, B.W.M., Vulto-van Silfhout, A.T.,
1096 Bosco, P., Friend, K.L., Baker, C., Buono, S., Vissers, L.E.L.M., et al. (2014). Refining
1097 analyses of copy number variation identifies specific genes associated with developmental
1098 delay. *Nat. Genet.* *46*, 1063–1071.
- 1099 13. Eising, E., Carrion-Castillo, A., Vino, A., Strand, E.A., Jakielski, K.J., Scerri, T.S.,
1100 Hildebrand, M.S., Webster, R., Ma, A., Mazoyer, B., et al. (2019). A set of regulatory genes

- 1101 co-expressed in embryonic human brain is implicated in disrupted speech development. *Mol.*
1102 *Psychiatry* 24, 1065–1078.
- 1103 14. Hamdan, F.F., Srour, M., Capo-Chichi, J.-M., Daoud, H., Nassif, C., Patry, L.,
1104 Massicotte, C., Ambalavanan, A., Spiegelman, D., Diallo, O., et al. (2014). De novo
1105 mutations in moderate or severe intellectual disability. *PLoS Genet.* 10, e1004772.
- 1106 15. Morgan, A., Braden, R., Wong, M.M.K., Colin, E., Amor, D., Liégeois, F., Srivastava, S.,
1107 Vogel, A., Bizaoui, V., Ranguin, K., et al. (2021). Speech and language deficits are central to
1108 SETBP1 haploinsufficiency disorder. *Eur. J. Hum. Genet.* 29, 1216–1225.
- 1109 16. Jansen, N.A., Braden, R.O., Srivastava, S., Otness, E.F., Lesca, G., Rossi, M., Nizon,
1110 M., Bernier, R.A., Quelin, C., van Haeringen, A., et al. (2021). Clinical delineation of SETBP1
1111 haploinsufficiency disorder. *Eur. J. Hum. Genet.* 29, 1198–1205.
- 1112 17. Hildebrand, M.S., Jackson, V.E., Scerri, T.S., Van Reyk, O., Coleman, M., Braden, R.O.,
1113 Turner, S., Rigbye, K.A., Boys, A., Barton, S., et al. (2020). Severe childhood speech
1114 disorder: Gene discovery highlights transcriptional dysregulation. *Neurology* 94, e2148–
1115 e2167.
- 1116 18. Cristóbal, I., Blanco, F.J., Garcia-Orti, L., Marcotegui, N., Vicente, C., Rifon, J., Novo,
1117 F.J., Bandres, E., Calasanz, M.J., Bernabeu, C., et al. (2010). SETBP1 overexpression is a
1118 novel leukemogenic mechanism that predicts adverse outcome in elderly patients with acute
1119 myeloid leukemia. *Blood* 115, 615–625.
- 1120 19. Minakuchi, M., Kakazu, N., Gorrin-Rivas, M.J., Abe, T., Copeland, T.D., Ueda, K., and
1121 Adachi, Y. (2001). Identification and characterization of SEB, a novel protein that binds to
1122 the acute undifferentiated leukemia-associated protein SET. *Eur. J. Biochem.* 268, 1340–
1123 1351.

- 1124 20. Piazza, R., Magistrini, V., Redaelli, S., Mauri, M., Massimino, L., Sessa, A., Peronaci,
1125 M., Lalowski, M., Soliymani, R., Mezzatesta, C., et al. (2018). SETBP1 induces transcription
1126 of a network of development genes by acting as an epigenetic hub. *Nat. Commun.* 9, 2192.
- 1127 21. Vishwakarma, B.A., Nguyen, N., Makishima, H., Hosono, N., Gudmundsson, K.O., Negi,
1128 V., Oakley, K., Han, Y., Przychodzen, B., Maciejewski, J.P., et al. (2016). Runx1 repression
1129 by histone deacetylation is critical for Setbp1-induced mouse myeloid leukemia
1130 development. *Leukemia* 30, 200–208.
- 1131 22. Nguyen, N., Vishwakarma, B.A., Oakley, K., Han, Y., Przychodzen, B., Maciejewski,
1132 J.P., and Du, Y. (2016). Myb expression is critical for myeloid leukemia development
1133 induced by Setbp1 activation. *Oncotarget* 7, 86300–86312.
- 1134 23. Oakley, K., Han, Y., Vishwakarma, B.A., Chu, S., Bhatia, R., Gudmundsson, K.O., Keller,
1135 J., Chen, X., Vasko, V., Jenkins, N.A., et al. (2012). Setbp1 promotes the self-renewal of
1136 murine myeloid progenitors via activation of Hoxa9 and Hoxa10. *Blood* 119, 6099–6108.
- 1137 24. Cardo, L.F., and Li, M. (2021). WNT/ β -catenin dependant alteration of cortical
1138 neurogenesis in a human stem cell model of SETBP1 disorder. *bioRxiv* 2021.10.12.464034.
- 1139 25. Sobreira, N., Schiettecatte, F., Valle, D., and Hamosh, A. (2015). GeneMatcher: a
1140 matching tool for connecting investigators with an interest in the same gene. *Hum. Mutat.* 36,
1141 928–930.
- 1142 26. American-Speech-Language-Hearing-Association [ASHA] (2007). *Childhood Apraxia of*
1143 *Speech*.
- 1144 27. Duffy, J.R. (2013). *Motor Speech Disorders - E-Book: Substrates, Differential Diagnosis,*
1145 *and Management (Elsevier Health Sciences)*.

- 1146 28. Morgan, A.T., Masterton, R., Pigdon, L., Connelly, A., and Liégeois, F.J. (2013).
1147 Functional magnetic resonance imaging of chronic dysarthric speech after childhood brain
1148 injury: reliance on a left-hemisphere compensatory network. *Brain J. Neurol.* 136, 646–657.
- 1149 29. McLeod, S., Crowe, K., and Shahaieian, A. (2015). Intelligibility in Context Scale:
1150 Normative and Validation Data for English-Speaking Preschoolers. *Lang. Speech Hear.*
1151 *Serv. Sch.* 46, 266–276.
- 1152 30. Sparrow, S., Cicchetti, D., and Saulnier, C. (2016). Vineland Adaptive Behaviour Scales
1153 (Bloomington: Pearson).
- 1154 31. Dunn, L.M., and Dunn, D.M. (2007). Peabody Picture Vocabulary Test | Fourth Edition.
- 1155 32. Lelieveld, S.H., Reijnders, M.R.F., Pfundt, R., Yntema, H.G., Kamsteeg, E.-J., de Vries,
1156 P., de Vries, B.B.A., Willemsen, M.H., Kleefstra, T., Löhner, K., et al. (2016). Meta-analysis
1157 of 2,104 trios provides support for 10 new genes for intellectual disability. *Nat. Neurosci.* 19,
1158 1194–1196.
- 1159 33. Lelieveld, S.H., Wiel, L., Venselaar, H., Pfundt, R., Vriend, G., Veltman, J.A., Brunner,
1160 H.G., Vissers, L.E.L.M., and Gilissen, C. (2017). Spatial Clustering of de Novo Missense
1161 Mutations Identifies Candidate Neurodevelopmental Disorder-Associated Genes. *Am. J.*
1162 *Hum. Genet.* 101, 478–484.
- 1163 34. Becker, M., Devanna, P., Fisher, S.E., and Vernes, S.C. (2018). Mapping of Human
1164 FOXP2 Enhancers Reveals Complex Regulation. *Front. Mol. Neurosci.* 11, 47.
- 1165 35. Dias, C., Estruch, S.B., Graham, S.A., McRae, J., Sawiak, S.J., Hurst, J.A., Joss, S.K.,
1166 Holder, S.E., Morton, J.E.V., Turner, C., et al. (2016). BCL11A Haploinsufficiency Causes an
1167 Intellectual Disability Syndrome and Dysregulates Transcription. *Am. J. Hum. Genet.* 99,
1168 253–274.

- 1169 36. Kim, D., Paggi, J.M., Park, C., Bennett, C., and Salzberg, S.L. (2019). Graph-based
1170 genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* 37,
1171 907–915.
- 1172 37. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and
1173 dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 550.
- 1174 38. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A.,
1175 Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment
1176 analysis: a knowledge-based approach for interpreting genome-wide expression profiles.
1177 *Proc. Natl. Acad. Sci. U. S. A.* 102, 15545–15550.
- 1178 39. Raudvere, U., Kolberg, L., Kuzmin, I., Arak, T., Adler, P., Peterson, H., and Vilo, J.
1179 (2019). g:Profiler: a web server for functional enrichment analysis and conversions of gene
1180 lists (2019 update). *Nucleic Acids Res.* 47, W191–W198.
- 1181 40. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins,
1182 R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint
1183 spectrum quantified from variation in 141,456 humans. *Nature* 581, 434–443.
- 1184 41. Lehman, A.M., McFadden, D., Pugash, D., Sangha, K., Gibson, W.T., and Patel, M.S.
1185 (2008). Schinzel-Giedion syndrome: report of splenopancreatic fusion and proposed
1186 diagnostic criteria. *Am. J. Med. Genet. A.* 146A, 1299–1306.
- 1187 42. Lai, C.S.L., Fisher, S.E., Hurst, J.A., Vargha-Khadem, F., and Monaco, A.P. (2001). A
1188 forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 413,
1189 519–523.
- 1190 43. Reuter, M.S., Riess, A., Moog, U., Briggs, T.A., Chandler, K.E., Rauch, A., Stampfer, M.,
1191 Steindl, K., Gläser, D., Joset, P., et al. (2017). FOXP2 variants in 14 individuals with

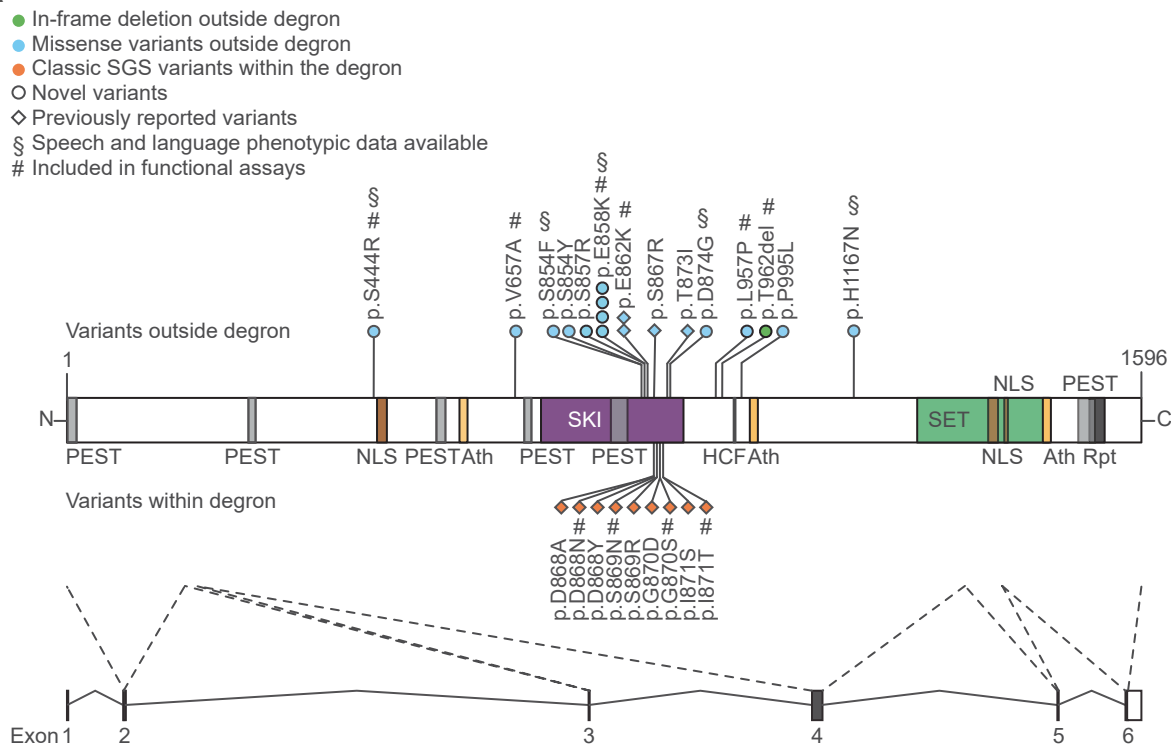
- 1192 developmental speech and language disorders broaden the mutational and clinical
1193 spectrum. *J. Med. Genet.* *54*, 64–72.
- 1194 44. Morgan, A., Fisher, S.E., Scheffer, I., and Hildebrand, M. (2016). FOXP2-Related
1195 Speech and Language Disorders. In *GeneReviews®*, M.P. Adam, H.H. Ardinger, R.A.
1196 Pagon, S.E. Wallace, L.J. Bean, G. Mirzaa, and A. Amemiya, eds. (Seattle (WA): University
1197 of Washington, Seattle), p.
- 1198 45. Inoue, D., Kitaura, J., Matsui, H., Hou, H.-A., Chou, W.-C., Nagamachi, A., Kawabata,
1199 K.C., Togami, K., Nagase, R., Horikawa, S., et al. (2015). SETBP1 mutations drive leukemic
1200 transformation in ASXL1-mutated MDS. *Leukemia* *29*, 847–857.
- 1201 46. Robinson, D.G., Wang, J.Y., and Storey, J.D. (2015). A nested parallel experiment
1202 demonstrates differences in intensity-dependence between RNA-seq and microarrays.
1203 *Nucleic Acids Res.* *43*, e131.
- 1204 47. Mootha, V.K., Lindgren, C.M., Eriksson, K.-F., Subramanian, A., Sihag, S., Lehar, J.,
1205 Puigserver, P., Carlsson, E., Ridderstråle, M., Laurila, E., et al. (2003). PGC-1alpha-
1206 responsive genes involved in oxidative phosphorylation are coordinately downregulated in
1207 human diabetes. *Nat. Genet.* *34*, 267–273.
- 1208 48. Lee, J.-W., Kim, D.-M., Jang, J.-W., Park, T.-G., Song, S.-H., Lee, Y.-S., Chi, X.-Z., Park,
1209 I.Y., Hyun, J.-W., Ito, Y., et al. (2019). RUNX3 regulates cell cycle-dependent chromatin
1210 dynamics by functioning as a pioneer factor of the restriction-point. *Nat. Commun.* *10*, 1897.
- 1211 49. Sullivan, J.A., Stong, N., Baugh, E.H., McDonald, M.T., Takeuchi, A., and Shashi, V.
1212 (2020). A pathogenic variant in the SETBP1 hotspot results in a forme-fruste Schinzel–
1213 Giedion syndrome. *Am. J. Med. Genet. A.* *182*, 1947–1951.
- 1214 50. Leonardi, E., Bettella, E., Pelizza, M.F., Aspromonte, M.C., Polli, R., Boniver, C., Sartori,
1215 S., Milani, D., and Murgia, A. (2020). Identification of SETBP1 Mutations by Gene Panel

- 1216 Sequencing in Individuals With Intellectual Disability or With “Developmental and Epileptic
1217 Encephalopathy.” *Front. Neurol.* *11*, 593446.
- 1218 51. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W.,
1219 Hegde, M., Lyon, E., Spector, E., et al. (2015). Standards and guidelines for the
1220 interpretation of sequence variants: a joint consensus recommendation of the American
1221 College of Medical Genetics and Genomics and the Association for Molecular Pathology.
1222 *Genet. Med.* *17*, 405–423.
- 1223 52. Heyne, H.O., Baez-Nieto, D., Iqbal, S., Palmer, D.S., Brunklaus, A., May, P.,
1224 Collaborative, E., Johannesen, K.M., Lauxmann, S., Lemke, J.R., et al. (2020). Predicting
1225 functional effects of missense variants in voltage-gated sodium and calcium channels. *Sci.*
1226 *Transl. Med.* *12*, eaay 8648.
- 1227 53. Haghshenas, S., Bhai, P., Aref-Eshghi, E., and Sadikovic, B. (2020). Diagnostic Utility of
1228 Genome-Wide DNA Methylation Analysis in Mendelian Neurodevelopmental Disorders. *Int.*
1229 *J. Mol. Sci.* *21*, 9303.
- 1230 54. Aref-Eshghi, E., Kerkhof, J., Pedro, V.P., Barat-Houari, M., Ruiz-Pallares, N., Andrau, J.-
1231 C., Lacombe, D., Van-Gils, J., Fergelot, P., Dubourg, C., et al. (2020). Evaluation of DNA
1232 Methylation Episignatures for Diagnosis and Phenotype Correlations in 42 Mendelian
1233 Neurodevelopmental Disorders. *Am. J. Hum. Genet.* *106*, 356–370.
- 1234 55. Rao, R.C., and Dou, Y. (2015). Hijacked in cancer: the KMT2 (MLL) family of
1235 methyltransferases. *Nat. Rev. Cancer* *15*, 334–346.
- 1236 56. Mészáros, B., Hajdu-Soltész, B., Zeke, A., and Dosztányi, Z. (2020). How mutations of
1237 intrinsically disordered protein regions can drive cancer. *bioRxiv* 2020.04.29.069245.
- 1238 57. den Hoed, J., and Fisher, S.E. (2020). Genetic pathways involved in human speech
1239 disorders. *Curr. Opin. Genet. Dev.* *65*, 103–111.

- 1240 58. den Hoed, J., Devaraju, K., and Fisher, S.E. (2021). Molecular networks of the FOXP2
1241 transcription factor in the brain. EMBO Rep. 22, e52803.
- 1242

Figure 1: SETBP1 variants outside the degnon cluster in the SKI domain and are highly conserved across species.

A



B

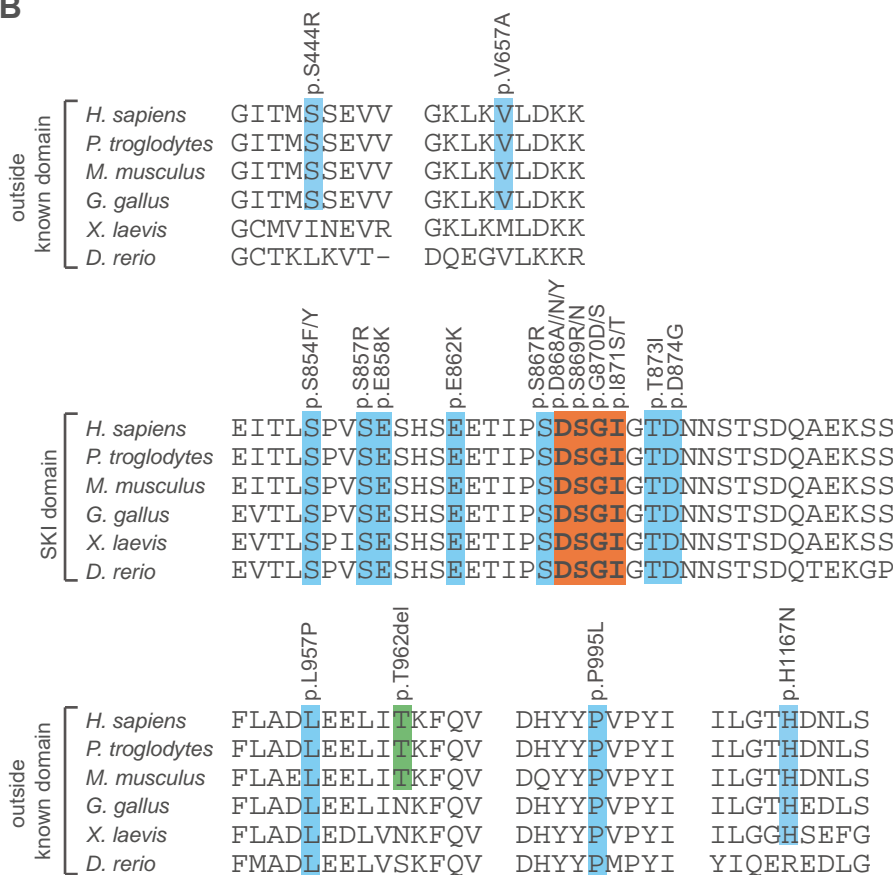


Figure 2: Clinical evaluation of individuals with *SETBP1* variants

Facial photographs not available in this preprint

Figure 3: Performance on Vineland-3 subtests

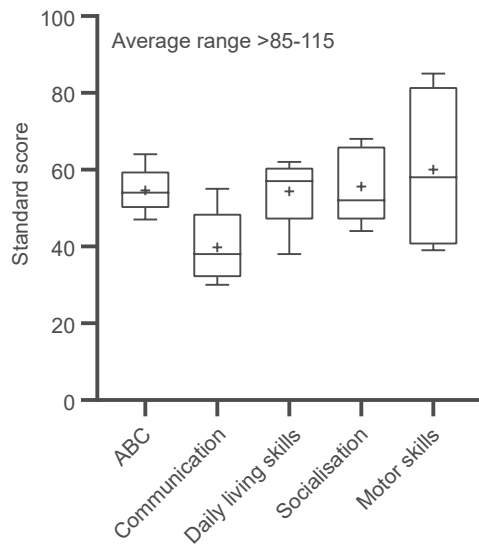


Figure 4: Variable impairment of SETBP1 ubiquitination via the proteasome and autophagy pathways is associated with partial reduction of ubiquitination.

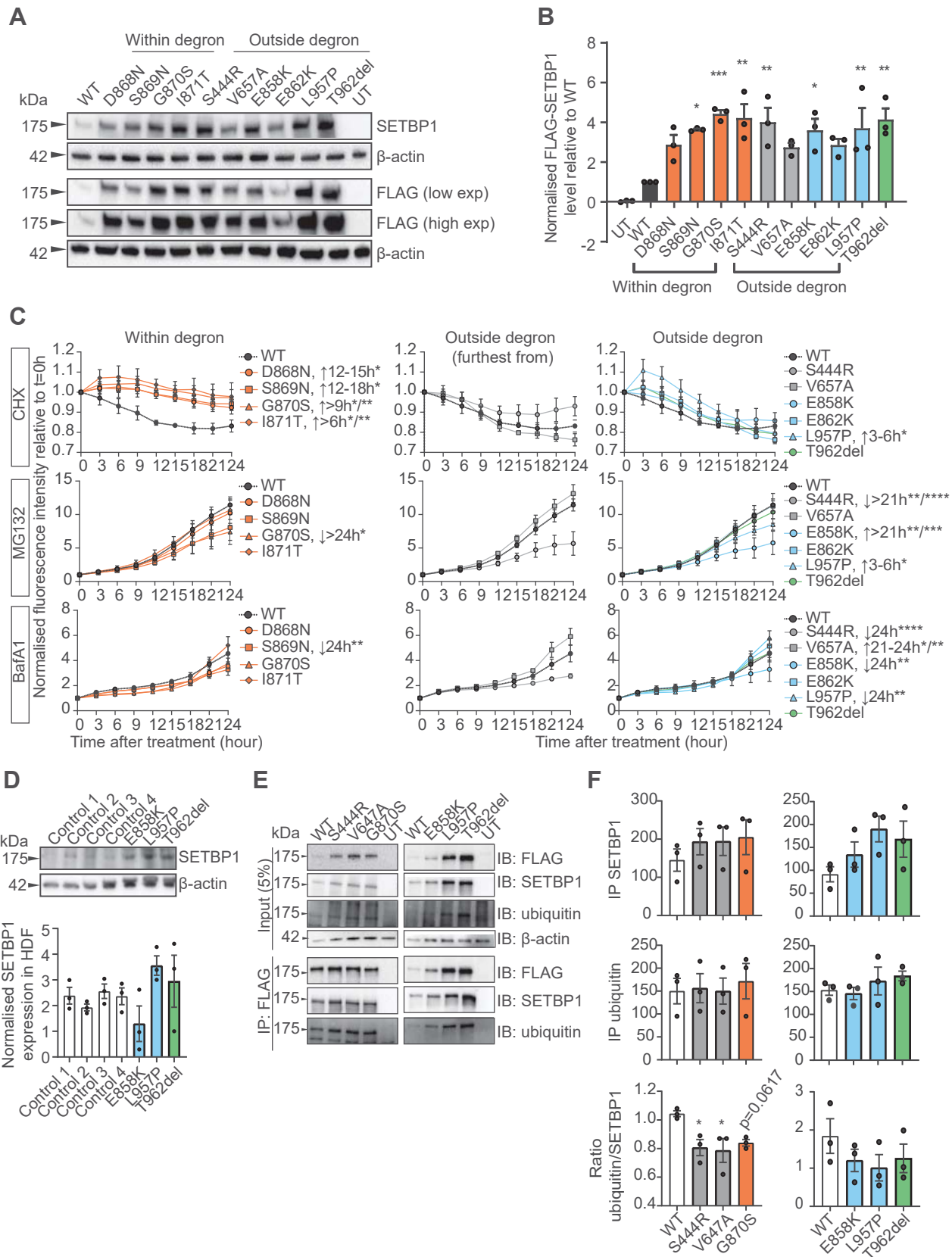


Figure 5: Genotype specific reduction of binding capacity to AT-rich DNA sequences and transcriptional activation in *SETBP1* variants outside the degron.

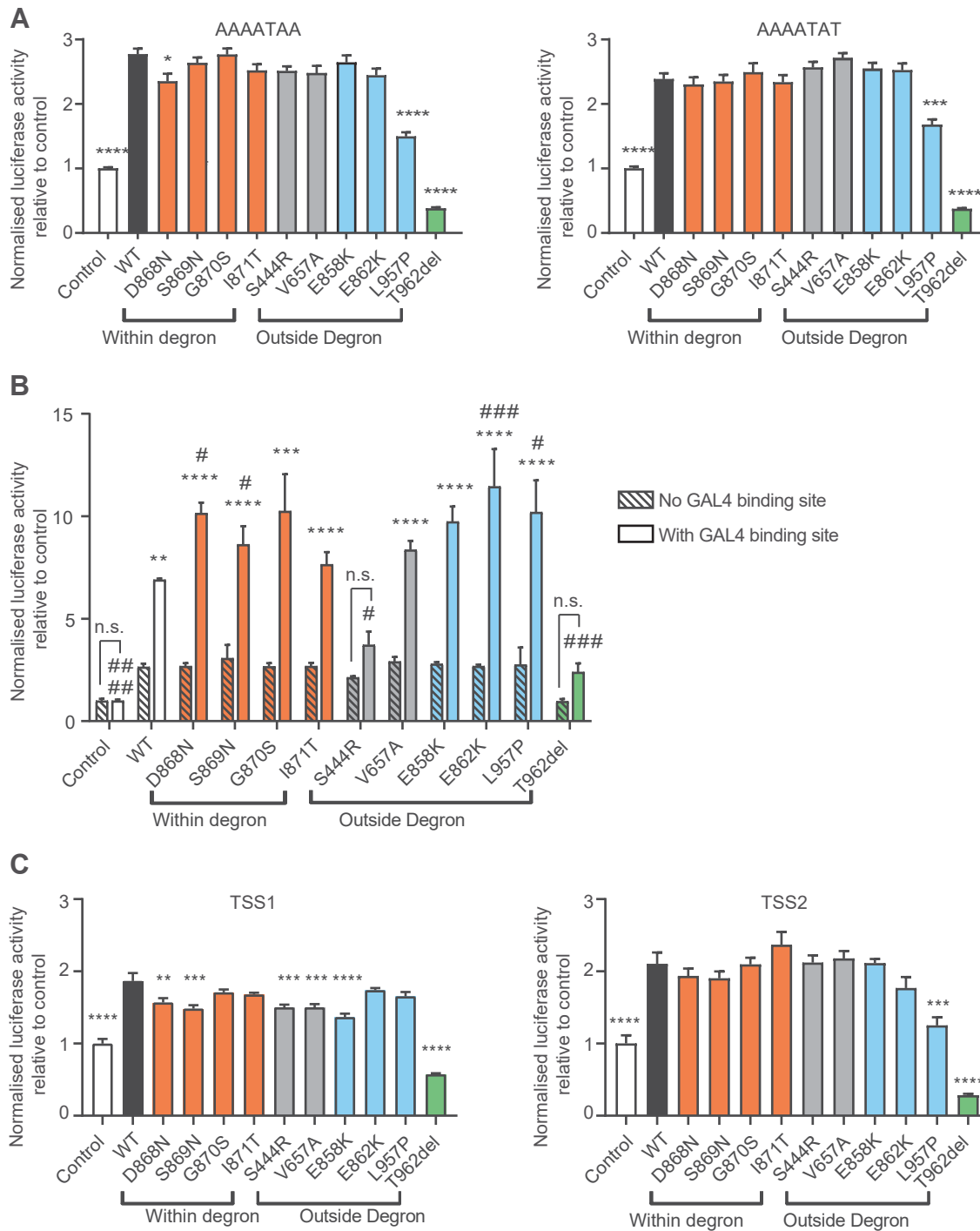


Figure 6: Increased proliferation of fibroblasts derived from patients carrying *SETBP1* variants outside the canonical degron.

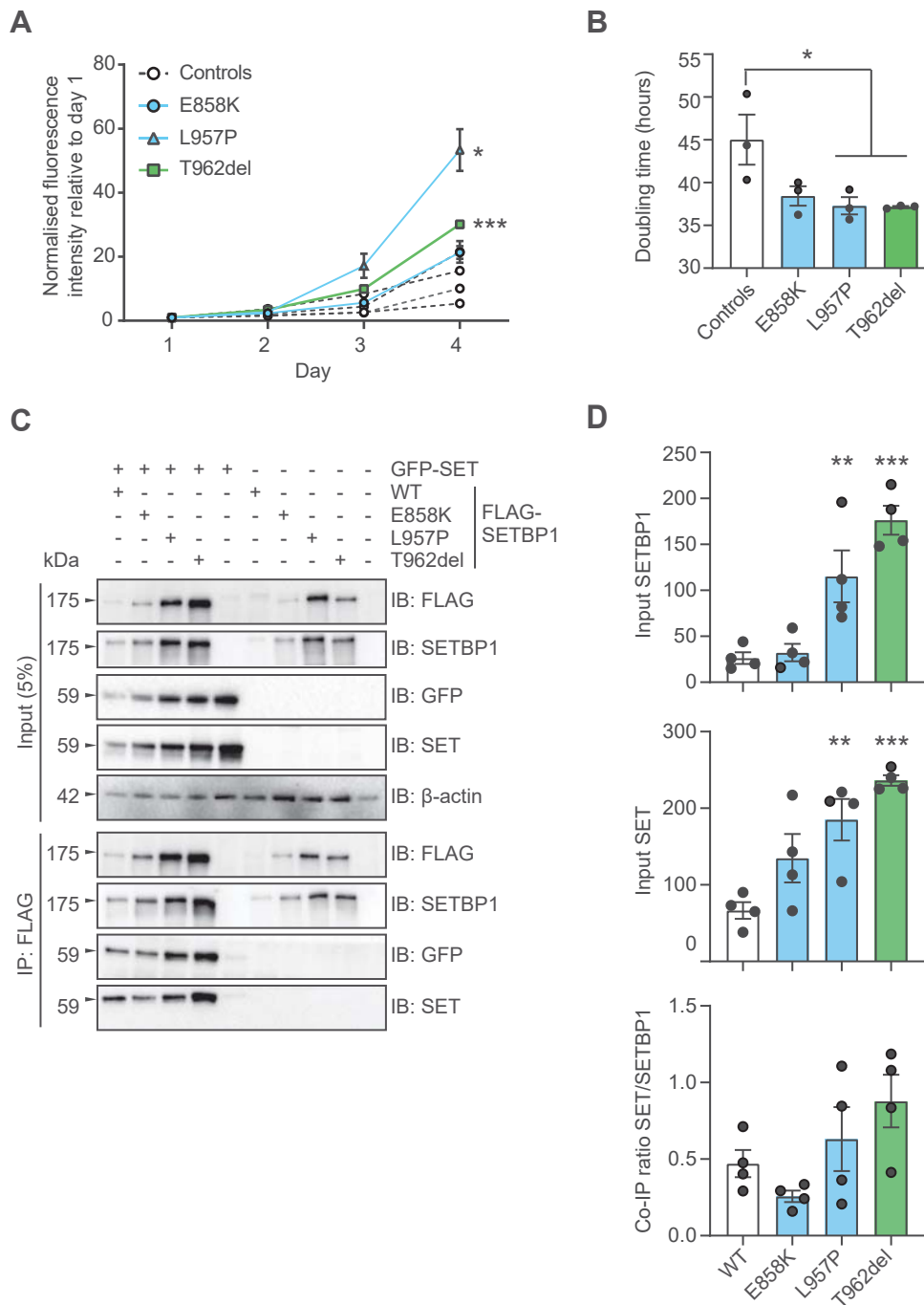


Figure 7: Fibroblasts carrying *SETBP1* variants showed distinct transcriptionomic profiles from healthy controls.

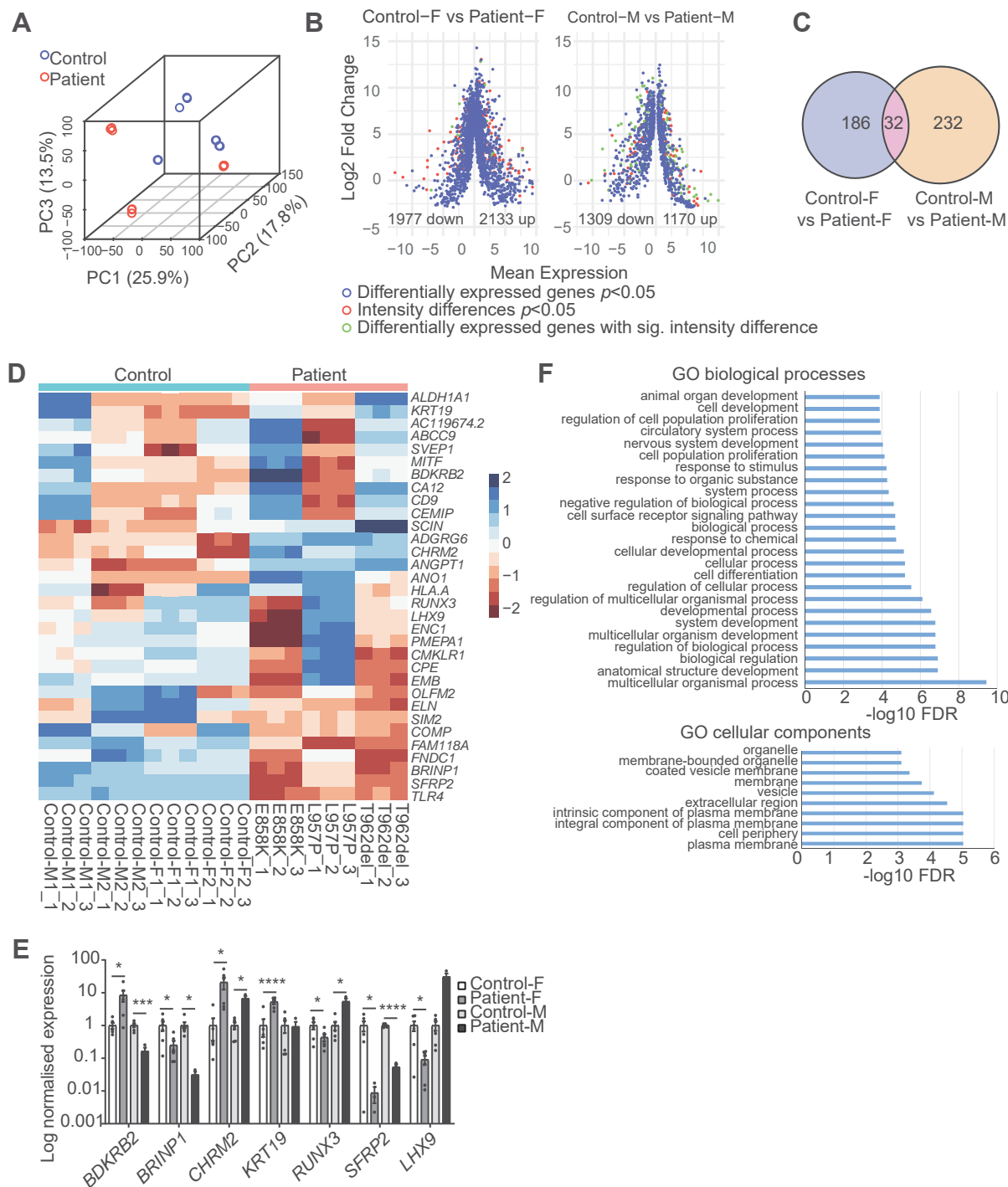


Figure 8: Summary of functional assays: variants outside the degron disrupt SETBP1 functions by dysregulating transcription, reducing DNA binding capacity and promoting cell proliferation independent of SETBP1 level.

