Deciphering the consequence of deep intronic variants: 
a progeroid syndrome caused by a $TAPT1$ mutation
is revealed by combined RNA/SI-NET sequencing

Nasrin Sadat Nabavizadeh$^{1,2,3}$, Annkatrin Bressin$^4$, Poh Hui Chia$^1$, Ricardo Moreno Traspas$^1$, Nathalie Escande-Beillard$^{1,3}$, Carine Bonnard$^1$, Zohreh Hojati$^2$, Scott Drutman$^5$, Susanne Freier$^4$, Mohammad El-Khateebe$^6$, Rajaa Fathallah$^6$, Jean-Laurent Casanova$^{5,7,8,9,10}$, Wesam Soror$^6$, Alaa Arafat$^6$, Mohammad Shboul$^{11}$, Andreas Mayer$^{4*}$, Bruno Reversade$^{1,3,12,13*}$

$^1$ Laboratory of Human Genetics & Therapeutics, Genome Institute of Singapore, A*STAR, Singapore.
$^2$ Division of Genetics, Department of Cell and Molecular Biology & Microbiology, Faculty of Biological Science and Technology, University of Isfahan, Isfahan, Iran.
$^3$ Medical Genetics Department, Koç University School of Medicine, 34010 Istanbul, Turkey.
$^4$ Max Planck Institute for Molecular Genetics, Berlin, Germany
$^5$ St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, Rockefeller University, New York, NY
$^6$ National Center for Diabetes, Endocrinology and Genetics, Amman
$^7$ Laboratory of Human Genetics of Infectious Diseases, Necker Branch, INSERM U1163, Necker Hospital for Sick Children, Paris, France.
$^8$ University of Paris, Imagine Institute, Paris, France.
$^9$ Howard Hughes Medical Institute, New York, NY, USA.
$^{10}$ Pediatric Hematology and Immunology Unit, Necker Hospital for Sick Children, AP-HP, Paris, France.
$^{11}$ Department of Medical Laboratory Sciences, Jordan University of Science and Technology, Irbid 22110, Jordan
$^{12}$ Institute of Molecular and Cell Biology, A*STAR, 61 Biopolis Drive, Proteos, Singapore 138673.
$^{13}$ National University of Singapore, Department of Paediatrics, 1E Kent Ridge Road, Singapore 119228.

* Correspondence should be sent to:
mayer@molgen.mpg.de and bruno@reversade.com

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
ABSTRACT

Exome sequencing has introduced a paradigm shift for the identification of germline variations responsible for Mendelian diseases. However, non-coding regions, which make up 98% of the genome, cannot be captured. The lack of functional annotation for intronic and intergenic variants makes RNA-seq a powerful companion diagnostic. Here, we illustrate this point by identifying six patients with a recessive Osteogenesis Imperfecta (OI) and neonatal progeria syndrome. By integrating homozygosity mapping and RNA-seq, we delineated a deep intronic TAPT1 mutation (c.1237-52G>A) that segregated with the disease. Using patients’ fibroblasts, we document that TAPT1’s nascent transcription was not affected, indicating instead that this variant leads to an alteration of pre-mRNA processing. Predicted to serve as an alternative splicing branchpoint, this mutation causes TAPT1 exon 12 skipping, creating a protein-null allele. Additionally, our study reveals dysregulation of pathways involved in collagen and extracellular matrix biology in disease-relevant cells. Overall, our work highlights the power of transcriptomic approaches in deciphering the repercussion of non-coding variants as well as in illuminating the molecular mechanisms and underlying pathways of human diseases.
INTRODUCTION

Whole exome sequencing (WES) targets less than 2% of our genome whereas the majority of non-coding sequences are still understudied. These sequences are crucial for gene regulation, are to a large extent transcribed and form a significant portion of our genome which are also susceptible to harbor variants responsible for human diseases\textsuperscript{1–4}. Indeed, from the more than 4000 Mendelian phenotypes reported to date, approximately 50% still lack the identification of the underlying genetic cause\textsuperscript{5}. This speaks to the necessity to further explore non-coding sequences. Whole-genome sequencing (WGS) provides a more comprehensive method to cover the full genome, however, a key challenge to its implementation is the prioritization of the vast amount of non-coding variants identified. This barrier to interpretation is in part driven by the lack of annotated information in intronic and intergenic regions which together comprise up to 98% of our genome. RNA-sequencing (RNA-seq) has proven to be a powerful complementary approach that can help overcome these challenges by revealing the functional impact of the genetic variants at the transcriptome level. The use of RNA-seq in conjunction with WGS permits cross-referencing of endogenous RNA levels and splicing events to help prioritize disease-causing mutations at the DNA level\textsuperscript{6–8}.

Here we report the study of six affected children from two consanguineous families that presented with a congenital syndrome consisting of osteogenesis imperfecta (OI), severe developmental delay and neonatal progeria. By combining homozygosity mapping, RNA-seq and targeted Sanger sequencing, we identified an intronic homozygous variant (c.1237-52 G>A) in TAPT1 (MIM612758) which entirely segregated with the disease. Our downstream characterization assays using patient-derived fibroblasts showed how this private non-coding mutation induces the complete skipping of exon 12 leading to a TAPT1 protein-null allele.
TAPT1 codes for a predicted transmembrane protein which has been related to ER/Golgi pathways, human Cytomegalovirus (HCMV) infection and primary ciliogenesis \(^9\)–\(^{15}\). Our functional studies using patient derived TAPT1-knockout cells could not detect patent anomalies in the pathways previously linked to TAPT1 indicating that its precise molecular function has yet to be ascertained. Notwithstanding, our RNA-seq and NET-seq (native elongating transcript sequencing) analyses revealed a role for TAPT1 in collagen and ECM biology, which is consistent with clinical presentation of our patients. Overall, our study highlights the capacity of applying robust transcriptomic approaches to prioritize disease-causing genes and understand the underlying pathogenesis of Mendelian disease.

RESULTS

A severe recessive progeroid syndrome with osteogenesis imperfecta

We investigated six severely-affected children from two consanguineous families (Fig. 1A, 1B) manifesting growth retardation, short stature, multiple bone deformities, lipodystrophy and neonatal progeria. The patients had various craniofacial abnormalities including prominent forehead, plagiocephaly, depressed nasal bridge, nasal septum deviation, low set ears, ear deformities, micrognathia, and occult cleft palate (Fig. 1C - 1E, S1). The patients also suffered from microphthalmia, cataract, and bilateral esotropia. They had translucent, wrinkled skin with patent acrogeria and sparse hair with premature depigmentation (Fig. 1C - 1E, S1). They also displayed pectus excavatum and brachydactyly of both hands and feet (Fig. 1C - 1E, S1). X-rays of patient V.1 from family 1 showed extensive deformity of the bones, bone dysplasia with bowing, and evidence of previous multiple fractures (Fig. 1F). The proband had spared joints, a flattened epiphysis of the humeral bone, irregular growth of arm bones resulting in small deformed radius.
bone, and a bowed ulnar bone. She also presented a deformed clavicular bone with displacement of both clavicolosternal and acromioclavicular joints, deformed shoulders, irregular development of the scapula, bilateral shallow acetabulum, abnormal contour of bilateral femoral head, and absent femoral diaphysis. X-rays also revealed severe calcification defects involving premature atherosclerotic vascular calcification, periarticular soft tissue calcification, and irregular calcification of carpal bones (Fig. 1F). Brain abnormalities were also reported in patient V.1 (F1) with cranial MRI showing defects in the white matter of the frontal and occipital lobes with pachygyria, possibly representing some form of leukodystrophy. The proband V-1 (F1) and V-5 (F1) died of severe respiratory infection and inflammation at the age of 10 and 4.5 years, respectively. The history of a similar disease was remarkable in this extended kindred. Two affected girls (IV-7 (F1) and V-13 (F1)) born to the mother’s aunts who showed similar clinical manifestation and died of severe respiratory distress at the age of 5 years. Another case (V-12 (F1)) of 2 years of age is alive and manifests similar clinical features.

**A deep intronic TAPT1 variant segregates with the disease**

Although the two families were reported to be unrelated, both originated from Redacted. Assuming a founder mutation, we carried out homozygosity mapping for a total of 15 individuals including 3 affected patients (V.1 and V.5 (F1), IV.1 (F2)), 3 pairs of parents, and 6 unaffected siblings from F1 (IV.4, IV.5, IV.6, V.2, V.3 and V.4). Homozygosity mapping confirmed distant relatedness between the 2 families with a minimal shared locus on chromosome 4 (4p16.1- p15.31) (hg19). The length of this Identical-by-Descent (IBD) locus was 8.4 Mb spanning a total of 39 candidate genes (Fig. 2A, S2A). We first performed whole-exome sequencing (WES) for V.1 (F1) and IV.1 (F2), but no compelling recessive mutations were found. To expand our search, we turned to an unbiased RNA-seq approach using primary cutaneous
fibroblasts from 2 affected individuals (V.1 and V.5 (F1)), and 2 unrelated wildtype (WT1 and WT2) controls. Out of the 39 candidate genes in the IBD region, our differential expression analysis data disclosed that TAPT1 was the only significantly dysregulated transcript in the patient primary dermal fibroblasts (Log2 fold change = -2.5) (Fig. 2B, S2B, S2D). Moreover, the alternative splicing analysis identified 63 genes with splicing defects in patient samples, TAPT1 being the top significant transcript with an exon 12 skipping event (Fig. 2C, 2D, S2C). Interestingly, homozygous mutations in TAPT1 were previously reported as the genetic cause of complex osteochondrodysplasia (MIM616897)\(^14\). Although less severe, this disease bears strong clinical overlap with the syndrome reported in this study. The exon 12-skipping event prompted us to search for the presence of possible TAPT1 intronic mutations. Targeted Sanger sequencing for this exon and its neighboring nucleotides (~300 bps) disclosed a deep intronic single nucleotide polymorphism (c.1237-52 G>A) within intron 11 that entirely segregated with the disease in all available family members (Fig. 2D, 3A). Together, these findings indicate that the c.1237-52 G>A mutation within intron 11 of TAPT1 most likely caused the disease for the 6 affected children.

**Exon 12 skipping targets TAPT1 mutant transcript for NMD, resulting in a protein-null allele**

How does the deep intronic mutation in TAPT1 lead to disease? To gain insights into the underlying disease-causing molecular mechanism, we applied a combined computational and experimental approach. We found that the private homozygous c.1237-52 G>A transition was predicted to serve as an alternative splicing branchpoint (Fig. 3B), thereby resulting in the exclusion of TAPT1 exon 12 (Fig. 2D, 3C). As the complete loss of exon 12 creates a premature stop codon, we used orthogonal qPCR validation tests to investigate whether the mutant TAPT1 transcript is targeted for nonsense-mediated decay (NMD) (Fig. 3C). Our data confirmed the
statistically significant reduction of endogenous TAPT1 mRNA levels in three of the patients compared to WT individuals (Fig 3D). We next examined the effect of the identified deep intronic TAPT1 mutation on the expression of its encoded protein. We employed two different commercial antibodies to detect endogenous TAPT1 in protein extracts from primary fibroblast cultures from two distinct patients and two WT individuals. Western blotting with both antibodies showed a complete loss of endogenous TAPT1 protein in all patient cells carrying the c.1237-52 G>A mutation in homozygosity (Fig. 3E). Notably, the fibroblasts from the mother IV.3 (F1) showed intermediate TAPT1 protein levels, which were consistent with her heterozygous genotype (Fig. 3E). These findings indicate that the novel mutant variant reported in this study behaves as a protein-null allele by creating an aberrant mis-spliced TAPT1 transcript, which undergoes degradation before being translated.

TAPT1 is situated head-to-head with its sequence-related antisense gene TAPT1-AS1 (Fig. 3A), which encodes a long non-coding RNA. Such upstream antisense transcripts can play a critical role in the regulation of gene expression\(^{16-19}\), in particular towards their associated protein-coding genes\(^{20,21}\). Here, because of the manifest physical proximity of TAPT1 and TAPT1-AS1, it is likely that both genes are expressed in a coordinated manner through shared regulatory elements as previously described for the majority of long non-coding RNA:mRNA gene pairs\(^{22}\). As such we examined whether the downregulation of TAPT1 may also alter the expression levels of TAPT1-AS1. qPCR data showed no overt changes in TAPT1-AS1 levels in patients' fibroblasts relative to control cells (Fig. 3D), thereby indicating that TAPT1 downregulation does not affect the expression of its neighbor antisense transcript. To further investigate the possible regulatory function of TAPT1-AS1 on its target gene, we knocked down the endogenous transcript in WT and TAPT1 mutant fibroblasts by transient transfection of two different TAPT1-AS1 GapmeRs. As
evidenced in our qPCR results, although both GapmeRs achieved the near complete depletion of \textit{TAPT1-ASI} transcript in both control and patient cells, no significant alterations were detected in \textit{TAPT1} mRNA levels (Fig. S3A). In addition, TAPT1 protein expression was also found unaffected in a \textit{TAPT1-ASI} knocked down context (Fig. S3B). These results argue against the potential regulatory role of \textit{TAPT1-ASI} on \textit{TAPT1} expression, and hence exclude the possibility that this antisense transcript could be having an impact on the pathogenesis of the disease found in our patients.

\textbf{TAPT1 is enriched in the ER/Golgi and is dispensable for HCMV gH infection}

\textit{TAPT1} (MIM#612758) codes for a protein termed Transmembrane Anterior Posterior Transformation 1, with 5 membrane-spanning helices (Fig. 4A). Its cellular localization has been reported to be either in the endoplasmic reticulum (ER) or at the centrosome \cite{11,12,15}. In order to gain further insights in its cellular localization, we performed immunofluorescence (IF) staining with both commercial TAPT1 antibodies in WT and mutant primary dermal fibroblasts. Each antibody yielded different staining patterns which were identical between control and TAPT1 knockout cells (Fig. S4A, S4B). This clearly indicated that these antibodies are not suitable for IF purposes. Since we validated the use of the same antibodies for western blotting, we opted to conduct subcellular fractionation on patient and WT fibroblasts as an alternative strategy to examine its subcellular localization. Endogenous TAPT1 was enriched in the Mito/ER/Golgi fractions and, to a lesser extent, in the nuclear fractions (Fig. 4B). This data is consistent with the previous report of EMP65, the homologous TAPT1 protein in yeast \cite{11,12}, and pfcarl, the homologous TAPT1 protein in plasmodium \cite{9}, localizing in the ER/Golgi apparatus. Additional evidence of localization of TAPT1, with its partner protein SUCO, in the secretory pathway was also obtained in various human cell lines \cite{23}.
We carried out a series of functional tests by comparing WT and patient fibroblasts in order to gain a better understanding of TAPT1’s cellular function. Several studies in yeast have reported that EMP65 is critically involved in the Unfolded Protein Response (UPR) and ER-Associated Degradation (ERAD) pathways. However, we could not document significant alterations in the expression levels of a panel of ER stress-associated markers in TAPT1 knockout cells by qPCR. IF staining did not reveal obvious defects in ER- (using an anti-CANX antibody), GOLGI- (using an anti-GLG1 antibody) or mitochondrial-morphology (using anti-TOM20 antibody) in patient TAPT1 knockout cells compared to WT cells.

Two early publications had reported that TAPT1 encodes for a receptor of the human cytomegalovirus (HCMV) gH. We revisited this claim by testing whether human TAPT1-null patient cells were resistant to HCMV strain RC256 infection. HCMV strain RC256 is a recombinant virus carrying the Escherichia coli lacZ gene as a marker under the control of the β gene promoter. Our β-gal reporter assay showed no discernable differences between WT and mutant patient cells, hence indicating that TAPT1 is not essential for HCMV infection and suggesting the likely presence of other cellular receptors which permit HCMV cellular entry in the absence of TAPT1.

**Extracellular matrix and collagen-related pathways are dysregulated in TAPT1-null cells**

To gain insights into the cellular role of TAPT1 and the disease mechanism, we investigated which genes and pathways were changed in TAPT1-null patient cells. The combined analysis of RNA-seq and SI-NET-seq data was highly informative in this context. RNA-seq quantifies the steady-state RNA levels whereas SI-NET-seq provides a quantitative measure of the Pol II occupancy with single-nucleotide precision genome-wide. SI-NET-seq is an improved
variant of the NET-seq approach\textsuperscript{26} that relies on spike-ins and therefore allows quantitative comparisons between conditions\textsuperscript{25}. While the RNA-seq data identified \textit{TAPT1} as the only dysregulated gene on the candidate Chr. 4 locus, it also provided an unbiased list of 172 significantly (p<0.05) dysregulated genes in the patients’ fibroblasts at the mRNA level (Fig. 5A) (Table S1). Of these significantly altered genes, a similar fraction was up- and downregulated (Fig. 5A) (Table S1). Beyond \textit{TAPT1} which is the fifth most significantly downregulated gene in mutant cells, dysregulation of several other target genes was also validated by qPCR, such as \textit{RARRES2}, \textit{ZIC1}, and \textit{ZIC4} (Fig. 5B). Moreover, SI-NET-seq results revealed a total of 317 genes with aberrant Pol II occupancy and dysregulated nascent RNAs in the patient cells (Fig. 5C, S5A, S5B) (Table S2). For the majority of these genes (70\%) the density of transcriptionally engaged Pol II was significantly increased in patient cells (Fig. 5C) (Table S2). The Pol II occupancy at \textit{TAPT1} was not changed in patient cells indicating that Pol II transcription of \textit{TAPT1} was not impacted by the mutation (Fig. 5C). Importantly, the integrated analysis of RNA-seq and SI-NET-seq data provided a comprehensive view on the molecular pathways that were affected by the \textit{TAPT1} mutation. A pathway analysis of genes with either a significantly altered mRNA level or Pol II occupancy consistently revealed that extracellular matrix (ECM) organization and collagen-related pathways were highly enriched (Fig. 5D), suggesting a role of \textit{TAPT1} in ECM and collagen dynamics.

**DISCUSSION**

**Phenotypic spectrum of \textit{TAPT1} insufficiency**

Here, we report the successful identification of a genetic variant causing a recessive Mendelian syndrome in six affected children presenting with severe bone defects, developmental
delay and premature aging. As we did not detect any recessive mutations by WES, we followed an alternative analytical pipeline which involved homozygosity mapping, RNA-sequencing and targeted Sanger sequencing. We eventually identified a deep intronic mutation (c.1237-52 G>A) in the TAPT1 gene that entirely segregated with the disease. It is known that pathogenic deep intronic mutations can induce splicing abnormalities, which in most cases lead to mRNA NMD due to the introduction of premature termination codons (PTCs) 27–30. Our prediction analysis suggested that the novel c.1237-52 G>A mutation is likely behaving as an alternative splicing branchpoint which triggers aberrant exon 12 skipping in the TAPT1 pre-mRNA. This splicing aberration disrupts the reading frame and introduces a premature stop codon which targets mutant TAPT1 transcript for NMD, as confirmed by qPCR in two distinct patient’s fibroblast cell lines. As expected and evidenced by western blotting data, the significant drop in TAPT1 mRNA levels prevents the translation of a truncated protein. Moreover, SI-NET-seq showed that nascent transcription at the TAPT1 gene was not affected by the mutation, confirming that the disease arises from a post-transcriptional dysregulation.

Genetic defects in TAPT1 were firstly reported by S. Symoens et al. (2016) in two consanguineous families with a complex and lethal osteochondroplasia syndrome (MIM616897) 15. Furthermore, N. Patel et al. (2017) reported a homozygous truncating mutation (c.846+2insT) in TAPT1 segregating with pediatric cataract, although these patients did not show any evidence of skeletal defects 31. In addition to the shared clinical features with these previously reported TAPT1-deficient patients including bone abnormalities and cataract, our affected children also suffered from neonatal progeria, characterized by wrinkled and thin skin, premature depigmentation and lipodystrophy. This vast phenotypic variation may be driven by the severity of the alleles identified. Our six new patients carry a complete loss-of-function mutation resulting
in a protein-null allele, whereas both prior studies showed partial loss-of-function mutations including missense and in-frame exon 6 and exon 10 skipping\textsuperscript{15,31}. Another possibility is that, as is the case for \textit{LMNA} \textsuperscript{32}, a wide range of \textit{TAPT1} diseases exist depending on which domain of the protein is mutated, thus accounting for the observed phenotypic heterogeneity.

\textbf{TAPT1-deficiency resembles a collagenopathy}

To date, RNA-seq stands out as the gold-standard technique to identify affected signaling pathways underlying a certain disease. To identify cellular processes that are affected upon TAPT1 mutation, we performed an integrated pathway enrichment analysis combining RNA-seq and SI-NET-seq results. RNA-seq and SI-NET-seq uncovered genes with a significant change in transcript levels and nascent transcription in patient cells, respectively. Despite the different types of data, we observed a strong overlap in dysregulated pathways between both datasets. Collagen- and ECM-related pathways standout as most significant hits from this combined analysis indicating a dysregulation of these processes in patient cells. This interesting finding is consistent with our patients’ phenotype, which manifests with severe bone defects and skin abnormalities. Collagens are the most abundant proteins made by the human body and serve to provide structural support, tensile strength while mediating cell adhesion, and migration\textsuperscript{33,34}. The bone tissue and the skin dermis account for 80\% of the total collagen content of the body\textsuperscript{35}. Importantly, the majority of genetic alterations causing bone defects affect collagen themselves e.g. \textit{COL1A1} (MIM114000, MIM619115, MIM130060, MIM166200, MIM166210, MIM259420, MIM166220, MIM166710), \textit{COL1A2} (MIM619120, MIM617821, MIM225320, MIM166210, MIM259420, MIM166220, MIM166710), or enzymes dedicated to their processing and secretion such as \textit{P3H1} (MIM610915), \textit{CRTAP} (MIM610682) and \textit{TANGO1} (MIM619269)\textsuperscript{36-38}. The clinical
manifestations and transcriptomics results shown here support the hypothesis that TAPT1-deficiency belongs to the heterogeneous group of collagenopathies.

Previous computational and experimental interactome analyses proposed that TAPT1 physically interacts with two additional ER-resident proteins: SUCO (SUN Domain Containing Ossification Factor) $^{23,39}$ and P4HB, also known as PDI1 (protein disulfide isomerase 1) $^{39}$. Notably, mutations in SUCO and P4HB have been linked to skeletal dysplasia $^{40-45}$ in humans, which aligns with the TAPT1 loss-of-function clinical presentation. $Tapt1$ and $Suco$ mutant mice successfully phenocopy their corresponding human disease as they also present with severe skeletal defects $^{46,47}$. These two proteins form a highly conserved complex which is present in all eukaryotic cells from yeast $^{11,48}$ to humans $^{23}$. EMP65 and SLP1, the yeast homologues for TAPT1 and SUCO respectively, have been shown to be involved in ER quality-control machinery including UPR $^{11}$ and ERAD pathways $^{12}$. Surprisingly, our functional analyses did not reveal major abnormalities in the ER morphology and expression levels of ER stress markers in TAPT1-null cells. Previous research actually reported unaltered protein levels of ER chaperones, including BIP/GRP78, Calnexin, and GRP94, in $Suco$-null mouse osteoblasts $^{47}$, which is consistent with our results assuming a common functional pathway for TAPT1 and SUCO. P4HB, the other proposed interacting partner for TAPT1, serves as a prototypic thiol isomerase that is involved in the hydroxylation of proline residues in collagen fibers $^{49-51}$. Therefore, our data adds to previous studies supporting the hypothesis of TAPT1, SUCO, and P4HB may form a functional complex residing in the ER/GOLGI and playing a key role in collagen post-translational processing with a particular relevance to skeletal development in vertebrate species. In accordance with this idea, delayed collagen folding and secretion was documented in $TAPT1$ mutant fibroblasts $^{15}$.

**What could be TAPT1’s universal function in eukaryotic cells?**
Although loss-of-function mutations in TAPT1, SUCO and P4HB in humans all result in osteochondroplasia-like phenotypes, the homologues of these genes in lower organisms lead to unrelated phenotypes when absent. POD1, the TAPT1 homologue in Arabidopsis, was shown to be involved in pollen tube formation. F26F2.7, the Caenorhabditis elegans homologue of TAPT1 is a critical gene for embryonic viability with undetermined function yet. In Plasmodium falciparum, the causative pathogen for malaria, mutations in TAPT1’s homologue, pfcarl, confer resistance to various structurally unrelated antimalarial compounds which appear to target the ER/Golgi function of the parasite. In the unicellular fungus Saccharomyces cerevisiae, TAPT1 which is known as Emp65, is required for the stability of soluble proteins that are targeted to the secretory pathway. Notably, none of these species of plants, invertebrates or fungus possess genes coding for collagens, arguing that TAPT1’s role in all eukaryotic cells must be unrelated to collagen biology per se, but instead fulfill a more essential cellular role.
MATERIALS AND METHODS

Sample Collection and Clinical Assessment

The affected children were firstly diagnosed with severe osteogenesis imperfecta. In total, 15 saliva samples were collected from members of the two families including parents, affected and unaffected siblings. Genomic DNA from saliva samples was isolated using the Origene DNA Collection Kit (OG-500, DNAGenotek). Skin biopsies from three affected (V.I (F1), V.5 (F1) and IV.1 (F2)) and one unaffected (IV.3 (F1)) family members were also collected. Informed consent was obtained from all individuals in accordance with local ethical review board requirements in and Singapore (A*STAR IRB reference code #2019-087, Singapore). Additionally, parents gave their consent to participate in the present study and to have the results of this research work published.

Genotyping and Homozygosity Mapping

SNP genotyping was performed on the genomic DNA from 15 affected and unaffected individuals from both families using Illumina HumanCoreExome-12v1 Bead-Chips. Identity-by-descent (IBD) mapping detected common homozygous regions in the 3 affected individuals using Wolfram Mathematica data-analysis software. IBD homozygous blocks were identified as regions >2 cM. Candidate homozygous regions were refined by excluding the shared homozygous regions with unaffected individuals. Finally, a single identical and homozygous region was revealed on Chr. 4 (4p16.1- p15.31) (hg19).

Whole Exome Sequencing (WES)
The Ion TargetSeq™ Exome and Custom Enrichment Kit (Life Technologies) was used for exome capture from 1 µg of genomic DNA from individuals V.I (F1) and IV.1 (F2). The Ion OneTouch System (Life Technologies) was used for exome library preparation. Sequencing was performed using the Ion Proton Instrument (Life Technologies) with one Ion PI chip (Life Technologies). The variants were annotated with their associated gene and location. No candidate variant was found using various filtering parameters.

**RNA-Sequencing**

RNA from primary dermal fibroblasts from 2 patients (V.I (F1) and V.5 (F1)), and 2 unrelated wildtypes (WT1 and WT2) was extracted using the RNeasy Mini Kit (Qiagen). After measuring RNA quantity and integrity using the Agilent Bioanalyzer 2100 (Agilent Technologies), libraries were sequenced on a Illumina HiSeq/Novaseq sequencer. Reads were aligned to the GRCh38.p12 human reference genome using STAR v2.5.3a with default parameters in paired-end mode.

For differential gene expression analysis, we quantified the transcript abundance of the annotated genes from GENCODE v28 using HTSeq v0.11.4 in ‘union’ mode. Significant changes between the conditions were tested using DESeq2 v1.25.4. We defined genes as significantly dysregulated when they had an FDR adjusted p-value of <0.05. For alternative splicing analysis, we focused on alternative exon inclusion and exclusion events between wildtype and patient samples. After read mapping, we identified all exons from GENCODE v28 annotation showing an ‘exon inclusion level’ difference of at least 10% using rMATs v3.1.0. The ‘exon inclusion level’ of an exon describes the fraction of reads accounting for the inclusion of the exon. We defined alternative exon usage as an event between conditions with a significant (FDR <0.05)
difference in the ‘exon inclusion level’. Splicing events that were supported by less than 5 reads were excluded.

**SI-NET-sequencing**

For spike-in NET-seq (SI-NET-seq) 15 x 10^6 primary dermal fibroblasts were mixed with 3 x 10^6 murine NIH 3T3 cells. Murine NIH 3T3 cells served as spike-in controls. All subsequent steps of the SI-NET-seq experiments were performed as recently described with the following modification. For reverse transcription of nascent RNAs the SuperScript IV Reverse Transcriptase (ThermoFisher) was used.

Processing of SI-NET-seq data was performed as previously described with some modifications. Briefly, adaptor sequences and unique molecular identifiers (UMIs) were trimmed by cutadapt v2.4 and a custom python script, which preserves information of UMI sequences for the corresponding reads. The obtained reads were aligned to a joined reference genome from human GRCh38.p12 and mouse GRCm38.p6 using the STAR v2.5.3a aligner. For uniquely mapped reads, the position corresponding to the 3’-end of the nascent RNA fragment was recorded. We excluded reads that originated from reverse transcription mispriming and from PCR duplication using the UMI sequence information as described previously. Additionally, sequenced splicing intermediates were excluded. We masked regions that were transcribed by Pol I and III, as well as loci of short chromatin-associated RNAs, which were extracted from annotations in GENCODE v28/v29 (mouse: M18 and M22), RefSeq v109, miRBase v22.1 and the UCSC’s RepeatMasker. In the final step of data processing, we split the spiked-in mouse observations from sample observations.

We statistically tested the significance of changes in the Pol II occupancy. First, we quantified the Pol II occupancy at actively transcribed genes using SI-NET-seq data. Active genes had a
calculated TPM value of at least one using RSEM v1.3.1\textsuperscript{66} quantifications from wildtype RNA-seq data. Second, we tested for significant changes in the Pol II occupancy using DEseq2 v1.25.4\textsuperscript{59}. For data normalization we calculated the ‘Relative Log Expression’ on Pol II occupancy measurements from spiked-in mouse cells. Quantification of Pol II occupancy in mouse was calculated as for sample observations. To define actively transcribed genes, we used RNA-seq data available for NIH3T3 mouse cells (ENCODER: ENCSR000CLW)\textsuperscript{67}. Changes in the Pol II occupancy at genes with an FDR adjusted p-value of 0.05 or smaller were considered significant.

**Segregation Analysis**

The position coordinates and sequence of the candidate gene were obtained from the UCSC database. The region of the candidate mutation was amplified by PCR from genomic DNA from all 15 individuals using specific primers. Direct Sanger sequencing was performed on the PCR products using the BigDye Terminator Cycle Sequencing Kit (Applied Biosystems). Primer sequences are shown in Table S3.

**Cell Culture**

Primary dermal fibroblasts of affected and unaffected individuals were derived from skin biopsies following standard procedures\textsuperscript{68}. All human cell lines were cultured at 5% CO\textsubscript{2} and 37°C in high glucose DMEM (HyClone) supplemented with 10% fetal bovine serum (FBS) (HyClone), 1X penicillin/streptomycin (Thermo Fisher Scientific) and 2 mM L-glutamine (Biological Industries), and tested negative for mycoplasma using the MycoAlert\textsuperscript{TM} Mycoplasma Detection Kit (Lonza). Murine NIH 3T3 cells (ATCC: CRL-1658) were grown in DMEM containing 10% FBS (Bovine Calf Serum, iron-fortified, Sigma) and 5% penicillin-streptomycin.
Quantitative PCR

Total RNA was isolated from primary dermal fibroblasts using the RNeasy Mini Kit (Qiagen). RNA (1 µg) was reverse transcribed using the Iscript™ cDNA Synthesis Kit (Bio-Rad) according to the manufacturer’s instructions. Transcript levels were assessed using the Power SYBR™ Green PCR Master Mix (Applied Biosystems) and specific primers (Table S3) on the ABI Prism 7900HT Fast qPCR System (Applied Biosystems). qPCR assays involved three biological replicates per condition and three technical replicates per sample (N = 3, n = 3). GAPDH was used as the housekeeping gene to normalize gene expression.

Western Blot

Total cellular protein extracts from primary dermal fibroblasts were obtained using RIPA buffer supplemented with 1X Protease Inhibitor Cocktail (Roche). Nuclear, Mito/ER/Golgi, and Cytoplasmic fractions were prepared using the Cell Fractionation Kit Standard (Abcam, ab109719) following the manufacturer’s instructions. Protein concentrations were measured using the Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific). Samples were reduced in Laemmli loading buffer containing dithiothreitol, and denatured at 95°C for 5 minutes. Equal amounts of protein were loaded on precast 10% Tris/Glycine/SDS polyacrylamide gradient gels (Bio-Rad), followed by transferring on PVDF membranes (Bio-Rad) using the Trans-Blot® Turbo™ Transfer System (Bio-Rad). Membranes were blocked in 5% milk in TBST for 1 hour at room temperature, and subsequently probed with the following primary antibodies diluted in 5% milk in TBST overnight at 4°C: mouse anti-GAPDH (1:1000; Santa-Cruz, sc-47724), rabbit anti-TAPT1 (1:1000; Sigma, HPA042567, reacted with TAPT1 sequence covering exon 13-14), rabbit anti-TAPT1 (1:1000; Sigma, HPA048658, reacted with TAPT1 sequence covering exon 6-8),
mouse anti-BIP (1:1000; BD Biosciences, 610978), rabbit anti-TGN-46 (1:1000; Abcam, ab50595), rabbit anti-AK2 (1:1000; Proteintech, 11014-1-AP) and mouse anti-Lamin A/C (1:1000; EMD Millipore, MAB3211). After washes in TBST, secondary anti-mouse/HRP or anti-rabbit/HRP antibodies were used at 1:4000 dilution in 5% milk in TBST for 1 hour at room temperature. The signal was revealed with the SuperSignal™ West Chemiluminescent Substrate System (Thermo Fisher Scientific, #34080/34076/34096) and developed using CL-Xposure™ Films (Thermo Fisher Scientific) in a Carestream Kodak developer.

**Immunofluorescence Analysis**

Primary dermal fibroblasts were cultured on 8-well glass chamber slides (Millicell EZ SLIDES) and fixed for 15 minutes in 4% paraformaldehyde in PBS at room temperature. The cells were permeabilized with 0.3% Triton-X100 in PBS for 15 minutes, and blocked in 1% BSA in PBS for 1 hour at room temperature. Samples were then incubated with the following primary antibodies diluted in 1% BSA in PBS overnight at 4°C: rabbit anti-TAPT1 (1:1000; Sigma, HPA042567), rabbit anti-TAPT1 (1:1000; Sigma, HPA048658), rabbit anti-TOM20 (1:1000; Proteintech, 11802-1-AP), rabbit anti-Calnexin (1:2000; Abcam, ab22595) and mouse anti-GLG1 (1:500; Abcam, ab103439). For visualization, 1:500 secondary antibodies conjugated to Alexa Fluor 568 or Alexa Fluor 488 (Invitrogen, Molecular Probes) were incubated for 1 hour at room temperature in the dark. 1 μg/ml DAPI (Life Technologies) was used for DNA staining, and cells were mounted using ProLong™ Diamond Antifade Mountant (Invitrogen). Images were captured using a FV1000 Olympus inverted confocal microscope equipped with a Leica camera.

**GapmeR Transfection**
Primary dermal fibroblasts were seeded on 6-well plates at a density of 100,000 cells per well. The following day, cells were transfected using the Lipofectamine RNAiMAX Reagent (Invitrogen) with two GapmeRs specific for TAPTI-AS1 and a non-targeted GapmeR as a negative control at a 40 nM concentration. The GapmeRs were purchased from Qiagen (Germany), and their sequences are given in Table S3. 72 hours post-transfection, RNA and protein were harvested for downstream experiments.

**CMV β-Gal Assay**

20,000-40,000 cells were plated per well in 96-well plates. MRC5 (human lung fibroblast cells) was also used as a positive control. The next day after seeding, the virus (CMV strain RC256 ATCC VR-2356) was added in DMEM supplemented with 10% FBS at MOI=0.1 and absorbed for 1 hour at room temperature. Then the virus was aspirated off and the plates were carefully washed twice with PBS. 80 µl of DMEM supplemented with 10% FBS were added back to each well and the plate was returned to the incubator. β-gal activity was read at different time points (Days 0, 1, 2, 3), involving three replicates per time point and per cell line. For that purpose, 20 µl of 5X lysis buffer (500 mM K-phos pH 7.8, 1% Triton X-100) were first added to each well. After pipetting up and down, samples were incubated for 15 minutes at 37°C. Then, 10 µl of the lysates were transferred to a plate with Galacto-Star, which was subsequently covered with foil and incubated at room temperature for 20 minutes. The signal was finally measured in a luminescence plate reader.
ACKNOWLEDGEMENTS

We are profoundly grateful to all patient family members for their participation in this study. We would like to thank all members of Reversade laboratory for their kind help and support. We would also like to thank Dr. Shokouh Karimi for her help in careful clinical evaluation.

AUTHOR CONTRIBUTIONS


CONFLICT OF INTEREST STATEMENT

None of the authors have any financial interest related to this work and therefore declare no conflict of interest.

FUNDING

B.R. is a fellow of the Branco Weiss Foundation and EMBO Young Investigator. This work was also supported by a Strategic Positioning Fund on Genetic Orphan Diseases (GODAFIT) and an Use-Inspired Basic Research (UIBR) grant from Agency for Science, Technology and Research (A*STAR) in Singapore to B.R. This work was also funded by the Max Planck Society (to A.M.) and the Deutsche Forschungsgemeinschaft (DFG, grant 418415292 to A. M.).
References


48. Friederichs JM, Gardner JM, Smoyer CJ, et al. Genetic Analysis of Mps3 SUN Domain Mutants in *Saccharomyces cerevisiae* Reveals an Interaction with the


FIGURE LEGENDS

Figure 1. Patients from two distantly related families present with a recessively inherited syndrome characterized by osteogenesis imperfecta and neonatal progeria. (A, B) Pedigrees of two distantly related consanguineous families from Israel, showing an autosomal recessive mode of inheritance of the disease. Black symbols and crossed symbols represent affected and deceased individuals respectively. (C-E) Pictures of investigated patients showing severe bone deformities and fractures, neonatal progeria, wrinkled skin, prominent forehead and pectus excavatum. (F) Radiographs of affected V.1 (F1) showing several deficits in the bones including deformity, dysplasia, spared joints and evidence of previous fractures. Severe calcification defects can also be noticed, involving premature atherosclerotic vascular calcification, periarticular soft tissue calcification and irregular calcification of carpal bones. The pedigrees and patients’ images have been removed per medRxiv policy on identifying information.

Figure 2. Homozygosity mapping followed by RNA-seq uncovers a deep intronic recessive mutation in TAPT1. (A) Schematic representation of the shared IBD region between both families, located on Chromosome 4 (4p16.1 - p15.31) with a size of ~12 cM. Whereas WES analysis did not reveal any mutations in the coding sequences located in the IBD region, RNA-seq analysis helped us to identify the disease causative gene from this locus. (B) Volcano plot showing differentially expressed genes between WT (WT1 and WT2) and patient (V.1 (F1), V.5 (F1)) primary dermal fibroblasts. The vertical axis (y-axis) shows the –log10 P-value, whereas the horizontal axis (x-axis) displays the log2 fold change value. The red dots represent the upregulated transcripts; the blue dots represent the downregulated transcripts. A total of 172 genes were found significantly dysregulated. TAPT1, a gene located in the IBD region, appeared among the most significantly downregulated genes in the patients. (C) Plot showing the alternative
splicing analysis results from WT (WT1 and WT2) and patient (V.1 (F1), V.5 (F1)) primary dermal fibroblasts. The vertical axis (y-axis) shows the –log10 FDR (False Discovery Rate), whereas the horizontal axis (x-axis) represents the exon inclusion level (value ranging from -1 to 1). The red dots represent transcripts with exon inclusion events; the blue dots represent transcripts affected by exon skipping. A total of 63 aberrantly spliced genes were found in the patient cells, being TAPT1 the most significant exon skipping event. (D) (Left) Schematic representation showing the complete loss of exon 12 from TAPT1 transcript in patient cells, as defined by our splicing analysis data. (Right) Chromatogram showing the novel intronic mutation (c.1237-52 G>A) we found entirely segregating with the disease in all available family members. For display purposes, results from the targeted Sanger sequencing in WT, IV.3 (F1) and V.5 (F1) individuals are shown. The mutation is present in heterozygosis in IV.3 (F1) (unaffected mother) and in homozygosis in V.5 (F1) (affected patient).

Figure 3. TAPT1 mutant transcript lacks exon 12 and undergoes NMD to create a protein-null allele. (A) Schematic representation of TAPT1 and TAPT1-ASI, indicating the causative intronic mutation (c.1237-52 G>A). The transcription start sites and the direction of transcription are indicated by arrows. Scale bar represents 2 kb. (B) Diagram showing the branchpoint scores for the target c.1237-52 position and flanking nucleotides in TAPT1 intron 11 in both WT (+/+) and patient cells (-/-), as obtained from the RNABPS ⁶⁹, LaBranchoR ⁷⁰ and BPP ⁷¹ softwares. High branchpoint scores were predicted for the G>A transition in the patient cells using the RNABPS and LaBranchoR methods. The x-axis represents the nucleotide distance to the 3´ splice site (3´SS). (C) Schematic representation showing that the complete loss of exon 12 in TAPT1 results in a premature stop codon, which targets the transcript for nonsense-mediated mRNA decay. (D) qPCR results using specific primers for TAPT1 and TAPT1-ASI in 3 WT (WT1, WT2 and V.2 (F1)) and
3 affected (V.1 (F1), V.5 (F1), IV.1 (F2)) individuals. TAPT1 mRNA is significantly reduced in all patients compared to WTs, whereas TAPT1-AS1 transcript levels are unaffected. Fold change relative to V.2 (F1) is plotted as mean ± SD. Asterisks indicate conventional statistical significance (Student t-test; n.s. p-value > 0.05, **** p-value < 0.0001). (E) Western blot analysis of endogenous TAPT1 protein (~60 kDa) using whole protein extracts from primary dermal fibroblasts from WT (WT1 and WT2), heterozygous (IV.3 (F1)) and homozygous (V.1 (F1), V.5 (F1) and IV.1 (F2)) individuals and two different commercial antibodies (top: Sigma, HPA042567; bottom: Sigma, HPA048658). Results show a complete absence of TAPT1 protein in patient samples. GAPDH was used as a loading control.

**Figure 4. TAPT1 cellular localization and functional data.** (A) TAPT1 predicted topology: a membrane-spanning protein consisting of 5 transmembrane helices (Uniprot database). (B) Western blot analyses for TAPT1 (~60 kDa) using cytosolic, Mito/ER/Golgi and nuclear protein extracts from primary dermal fibroblasts of two WTs (WT1 and WT2) and two patients (V.5 (F1) and IV.1 (F2)). TAPT1 protein is highly enriched in the Mito/ER/Golgi fraction, and to a lower extent in the nuclear fraction. GAPDH, TGN46 and BiP served as a cytosolic, Golgi network and ER markers, respectively. Adenylate Kinase (AK2) was used as a mitochondrial marker. Laminin A/C was used as a nuclear marker. (C) Immunofluorescence staining of mitochondria using anti-TOM20 (green), ER using anti-CANX (red) and Golgi using anti-GLG1 (red) in primary dermal fibroblasts from WT1 and V.5 (F1). Similar staining patterns are observed with the three antibodies in both cell lines. Scale bar represents 10 µm. (D) qPCR analysis of a panel of canonical ER stress markers shows no significant differences in 3 patients (-/-) (V.I (F1), V.5 (F1) and IV.1 (F2)) primary dermal fibroblasts compared to WTs (+/+)(WT1, WT2 and WT3) cells. Fold change relative to WT is plotted as mean ± SD. Statistical significance was tested by Student t-test (n.s. p-
value > 0.05). (E) CMV cell infection assay on 2 patient (V.1 (F1) and V.5 (F1)) and 3 WT (WT1, WT4 and WT5) primary dermal fibroblast cell lines, using β-galactosidase activity as a readout. MRC5 cell line was used as a positive control. All of the cells were infected by the HCMV strain RC256 at a MOI=0.1. Data are shown as mean ± SD. Statistical significance was tested by Student t-test (n.s. p-value > 0.05).

Figure 5. Integrated analysis of SI-NET-seq and RNA-seq data revealed enrichment of collagen and ECM-related pathways in TAPT1-null cells. (A) Volcano plot showing differentially expressed genes as determined by RNA-seq in patient primary dermal fibroblasts (V.1 (F1), V.5 (F1)) compared to WT (WT1 and WT2) cells. The y-axis shows the –log10 P-value, whereas the x-axis displays the log2 fold change value. The red dots represent 75 significantly upregulated genes, and the blue dots represent 97 significantly downregulated genes. (B) qPCR validation test for 3 top dysregulated genes (RARRES2, ZIC1, and ZIC4) detected by RNA-seq. The analysis was performed on primary dermal fibroblasts from 2 WTs (WT1 and WT2) and 2 patients (V.5 (F1) and IV.1 (F2)). Fold change relative to WT1 is plotted as mean ± SD. Asterisks indicate statistical significance (Student t-test; ** p-value < 0.01, **** p-value < 0.0001). (C) Volcano plot showing genes with an altered occupancy of transcriptionally engaged Pol II in patient (V.1 (F1), V.5 (F1)) compared to WT (WT1 and WT2) primary fibroblast cells. The y-axis shows the –log10 P-value, whereas the x-axis indicates the log2 fold change value for the Pol II occupancy. The Pol II density is increased in 222 genes (red dots), and decreased in 95 genes (blue dots). The yellow dot represents TAPT1. (D) Bubble plot showing enrichment of collagen and extracellular matrix (ECM) pathways from the integrated reactome pathway analysis of the SI-NET-seq (light blue circles) and RNA-seq (dark blue circles) data. Enriched pathways are
indicated on the y-axis, and the corresponding p-values are shown on the x-axis. The size of the circles represents the number of altered genes from each pathway.

**Figure S1. Clinical pictures of the affected V.12 (F1) individual.** The patient presented with multiple abnormalities including bone and joint deformities, pectus excavatum, plagioccephaly microphthalmia and bilateral hypotropia. Moreover, she had apparent dysmorphic facial features such as a depressed nasal bridge and low set of ears. The patient's images have been removed per medRxiv policy on identifying information.

**Figure S2. Overlap analysis from homozygosity mapping and RNA-seq data revealed TAPT1 as the only candidate gene.** (A) List of the 39 candidate genes located in the mapped Chr. 4 IBD locus. (B) Lists of the top 10 significantly downregulated (left, blue) and upregulated (right, red) genes obtained from our RNA-seq differential expression analysis. (C) Expression changes (x-axis, log2FC) for genes with at least one alternative splicing event (skipped exon (SE), retained exon (RE), mutually exclusive exon (MXE), alternative 3’ or 5’ splice site (A3SS and A5SS) and retained intron (RI)). (D) (Top) Venn diagram displaying overlapping genes between the Chr. 4 IBD candidate locus, and the top 10 upregulated and downregulated genes from our RNA-seq data analysis. (Bottom) Venn diagram showing the overlapping genes between the differentially expressed set and the alternative spliced set from our RNA-seq data analysis. TAPT1 appears as the only overlapping gene in both diagrams.

**Figure S3. TAPT1-AS1 shows no observable regulatory activity on TAPT1 gene expression.** Knockdown of TAPT1-AS1 transcript using two different GapmeRs (1 and 2) in WT (WT1) and patient (IV.1 (F2)) primary dermal fibroblasts. A non-targeted (NT) GapmeR was used as control. (A) qPCR analysis of TAPT1-AS1 (top) and TAPT1 (bottom) transcript levels in the GapmeR-
transfected cells. Results show the successful knockdown of *TAPT1-AS1* by both GapmeRs 1 and 2 compared to the control NT GapmeR. However, *TAPT1* mRNA levels are unaltered in both WT and patient cells. Fold change relative to WT1-Control NT GapmeR is plotted as mean ± SD. Asterisks indicate conventional statistical significance (Student t-test; n.s. p-value > 0.05, ** p-value < 0.01, *** p-value < 0.001). (B) Western blotting of protein extracts from the GapmeR-transfected cells, probing for TAPT1 (Sigma, HPA042567 antibody). Data shows that TAPT1 protein levels are unaffected by the knockdown of *TAPT1-AS1*. GAPDH was used as loading control.

Figure S4. TAPT1 commercial antibodies are unsuitable for immunofluorescence experiments. Immunofluorescence staining using two different TAPT1 commercial antibodies (A: Sigma, HPA042567; B: Sigma, HPA048658) in WT1 and IV.1 (F2) primary dermal fibroblasts. Similar fluorescent signal was detected in WT and TAPT1-null cells in both cases. Scale bar represents 10 µm.

Figure S5. NET-seq analysis data. (A) Lists of the top 10 genes with significantly decreased (left, blue) or increased (right, red) RNA Pol II occupancy from our NET-seq analysis. (B) High Pearson’s correlation coefficients (r ≥ 0.96) between replicates of Pol II gene occupancy indicate the reproducibility of SI-NET-seq measurements.
Figure 2
Figure 3
Figure 4
Figure 5
**Figure S3**

(A) Relative TAP1:AS1 expression (Fold change) for WT1 (+/+), IV1 (F2) (-/-).

(B) Western blot analysis of TAP1 and GAPDH expression in different genotypes:
- Control NT (WT1 (+/+), IV1 (F2) (+/+))
- TAP1-AS1 GapmeR
- TAP1-AS1 GapmeR1
- TAP1-AS1 GapmeR2

*Note: kDa values are 50 and 37.*
Figure S4

A

TAPT1 (Sigma, HPA042567) / DAPI

WT1 (+/+)

IV.1 (F2) (-/-)

B

TAPT1 (Sigma, HPA048658) / DAPI

WT1 (+/+)

IV.1 (F2) (-/-)
### Table A

#### List of top genes with decreased RNA POL II occupancy

<table>
<thead>
<tr>
<th>Gene name</th>
<th>-log10 (padj)</th>
<th>Log2 fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td>ENOPH1</td>
<td>41.9</td>
<td>-4.96</td>
</tr>
<tr>
<td>COMP</td>
<td>27.55</td>
<td>-4.29</td>
</tr>
<tr>
<td>CRLF1</td>
<td>16.86</td>
<td>-5.11</td>
</tr>
<tr>
<td>PTPRE</td>
<td>13.98</td>
<td>-4</td>
</tr>
<tr>
<td>ANGPTL4</td>
<td>12.1</td>
<td>-3.15</td>
</tr>
<tr>
<td>COL15A1</td>
<td>12.04</td>
<td>-2.36</td>
</tr>
<tr>
<td>CLEC2A</td>
<td>11.94</td>
<td>-5.46</td>
</tr>
<tr>
<td>MMP1</td>
<td>11.55</td>
<td>-4.92</td>
</tr>
<tr>
<td>PTGS1</td>
<td>11.15</td>
<td>-3.38</td>
</tr>
<tr>
<td>SHISAL1</td>
<td>11.12</td>
<td>-2.72</td>
</tr>
</tbody>
</table>

#### List of top genes with increased RNA POL II occupancy

<table>
<thead>
<tr>
<th>Gene name</th>
<th>-log10 (padj)</th>
<th>Log2 fold change</th>
</tr>
</thead>
<tbody>
<tr>
<td>INHBA</td>
<td>12.69</td>
<td>2.74</td>
</tr>
<tr>
<td>CEMIP</td>
<td>12.4</td>
<td>2.64</td>
</tr>
<tr>
<td>NR2F1-AS1</td>
<td>12.04</td>
<td>2.91</td>
</tr>
<tr>
<td>ADGRD1</td>
<td>10.51</td>
<td>2.91</td>
</tr>
<tr>
<td>NTNG1</td>
<td>10.25</td>
<td>3.18</td>
</tr>
<tr>
<td>LNX1</td>
<td>9.86</td>
<td>2.1</td>
</tr>
<tr>
<td>PTGS2</td>
<td>9.3</td>
<td>2.69</td>
</tr>
<tr>
<td>SSC5D</td>
<td>9.04</td>
<td>2.53</td>
</tr>
<tr>
<td>KCND3</td>
<td>8.93</td>
<td>3.03</td>
</tr>
<tr>
<td>TRPC4</td>
<td>8.44</td>
<td>3.29</td>
</tr>
</tbody>
</table>

### Figure S5

[Image of the correlation matrix with significance stars]