This is a repository copy of *First genotype-phenotype study in TBX4 syndrome : gain-of-function mutations causative for lung disease*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/191107/

Version: Accepted Version

Article:
Prapa, M., Lago-Docampo, M., Swietlik, E.M. et al. (37 more authors) (2022) First genotype-phenotype study in TBX4 syndrome : gain-of-function mutations causative for lung disease. American Journal of Respiratory and Critical Care Medicine. ISSN 1073-449X

https://doi.org/10.1164/rccm.202203-0485oc

© 2022 American Thoracic Society. This article is open access and distributed under the terms of the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Reuse
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

First Genotype-Phenotype Study in TBX4 Syndrome: Gain-of-Function Mutations Causative for Lung Disease

Matina Prapa[1,2,#], Mauro Lago-Docampo[3,4,#], Emilia M. Swietlik[1,5,6], David Montani[7], Mélanie Eyries[8], Marc Humbert[7], Carrie C.L. Welch[9], Wendy Chung[10], Rolf M.F. Berger[14], Ham Jan Bogaard[15], Olivier Danhaive[16,17], Pilar Escribano-Subías[18,19], Henning Gall[14], Barbara Girerd[7], Ignacio Hernandez-Gonzalez[20], Simon Holden[21], David Hunt[22], Samara M.A. Jansen[15], Wilhelmina Kerstjens-Frederikse[23], David Kiely[11,24], Pablo Lapunzina[25,26,27], John McDermott[28,29], Shahin Moledina[30], Joanna Pepke-Zaba[6], Gary J. Polwarth[6], Gwen Schotte[15], Jair Tenorio-Castaño[25,26,27], A.A. Roger Thompson[11,24], John Wharton[31], Stephen J. Wort[31], NIHR BioResource for Translational Research – Rare Diseases[32], National Cohort Study of Idiopathic and Heritable PAH[33], PAH Biobank Enrolling Centers' Investigators[33], Karyn Megy[1,5], Rutendo Mapeta[1,5], Carmen M. Treacy[1], Jennifer M Martin[1], Wei Li[1], Andrew J. Swift[11], Paul D. Upton[1], Nicholas W. Morrell[1,5,6,12,*], Stefan Gräf[1,12,13,*], Diana Valverde[3,4,*]

# Equal contribution

[1] Department of Medicine, University of Cambridge, Cambridge Biomedical Campus

[2] St George's University Hospitals NHS Foundation Trust

[3] CINBIO, Universidade de Vigo, 36310 Vigo, Spain

[4] Rare Diseases and Pediatric Medicine, Galicia Sur Health Research Institute (IIS Galicia Sur), SERGAS-UVIGO, 36312 Vigo, Spain

[5] Addenbrooke's Hospital NHS Foundation Trust, Cambridge Biomedical Campus

[6] Royal Papworth Hospital NHS Foundation Trust, Cambridge Biomedical Campus

[7] Université Paris-Sud, Faculté de Médecine, Université Paris-Saclay; AP-HP, Service de Pneumologie, Centre de référence de l'hypertension pulmonaire; INSERM UMR_S 999, Hôpital Bicêtre, Le Kremlin-Bicêtre

[8] Département de génétique, hôpital Pitié-Salpêtrière, Assistance Publique-Hôpitaux de Paris, and UMR_S 1166-ICAN, INSERM, UPMC Sorbonne Universités

[9] Columbia University Medical Center

[10] Department of Pediatrics, Columbia University

[11] Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield

[12] NIHR BioResource for Translational Research, Cambridge Biomedical Campus

[13] Department of Haematology, University of Cambridge, Cambridge Biomedical Campus

[14] Centre for Congenital Heart Diseases, Pediatric Cardiology, Beatrix Children's Hospital, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands.

[15] Department of Pulmonary Medicine, Amsterdam University Medical Centre, Vrije Universiteit Amsterdam, Amsterdam Cardiovascular Sciences, the Netherlands.

[16] Division of Neonatology, St-Luc University Hospital, Catholic University of Louvain, Brussels, Belgium.

[17] Department of Pediatrics, University of California San Francisco, San Francisco, CA, USA.

[18] Unidad Multidisciplinar de Hipertensión Pulmonar, Servicio de Cardiología, Hospital Universitario 12 de Octubre, 28041 Madrid, Spain

[19] CIBERCV, Centro de Investigación Biomédica en Red de Enfermedades Cardiovasculares, ISCIII, 28029 Madrid, Spain

[20] Department of Cardiology, Hospital Universitario Río Hortega, 47012 Valladolid, Spain

[21] Department of Clinical Genetics, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK.

[22] Wessex Clinical Genetics Service, Princess Anne Hospital, Southampton SO16 5YA, UK

[23] Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands.

[24] Sheffield Pulmonary Vascular Disease Unit, Royal Hallamshire Hospital

[25] Institute of Medical and Molecular Genetics (INGEMM)-IdiPAZ, Hospital Universitario La Paz-UAM, 28046 Madrid, Spain

[26] CIBERER, Centro de Investigación Biomédica en Red de Enfermedades Raras, ISCIII, 28029 Madrid, Spain

[27] ITHACA, European Reference Network on Rare Congenital Malformations and Rare Intellectual Disability, 1000 Brussels, Belgium

[28] Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester University NHS Foundation Trust, UK.

[29] Division of Evolution and Genomic Sciences, School of Biological Sciences, University of Manchester, UK.

[30] Great Ormond Street Hospital

[31] National Heart & Lung Institute, Imperial College London

[32] www.ipahcohort.com

[33] www.pahbiobank.org

*** Corresponding authors**

Nicholas W. Morrell (nwmorrel23@cam.ca.uk)

University of Cambridge, Department of Medicine

Box 157, Level 5, Addenbrooke's Hospital, Hills Road,

Cambridge, CB2 0QQ, United Kingdom Tel: (+44) 1223 331666

Stefan Gräf (sg550@cam.ac.uk)

University of Cambridge, Department of Medicine

Addenbrooke's Hospital, Hills Road,

Cambridge, CB2 0QQ, United Kingdom

Diana Valverde (dianaval@uvigo.es)

CINBIO, Universidade de Vigo

Campus Universitario As Lagoas-Marcosende s/n

Vigo 36310, Spain

Tel: (+34) 986 811 953

Author contributions: All authors collected data and provided constructive criticism of the study manuscript. MP, ML-D, EMS, WL, AJS, PDU, NWM, SG and DV designed and interpreted the study. MP, ML-D and EMS analyzed the data.

This study phenotypically characterizes *TBX4* monoallelic variants and investigates the effect of missense variation using luciferase reporter constructs with the novel finding of both gain- and loss-of-function effects. It highlights gene-specific refinements for *TBX4* variant interpretation and draws important conclusions on genotype-phenotype correlations, including lower age at diagnosis of lung disease with shorter survival for variants located in the critical T-BOX domain.

Running title: *TBX4* genotype-phenotype associations

Descriptor number: 9.35 Pulmonary Hypertension: Clinical-Diagnosis/Pathogenesis/Outcome

Total number of words: 3336 (no abstract/title/references)

This article has an online data supplement, which is accessible from this issue's table of content online at www.atsjournals.org.

Some of the results of these studies have been previously reported in the form of a preprint (medRxiv, [28 February 2022] https://doi.org/10.1101/2022.02.06.22270467).

This article is open access and distributed under the terms of the Creative Commons Attribution Non-Commercial No Derivatives License 4.0 (http://creativecommons.org/licenses/by-nc-nd/4.0/). For commercial usage and reprints please contact Diane Gern (dgern@thoracic.org).

At a Glance

What is the current scientific knowledge on this subject?

TBX4 is the second most common gene culprit for development of pulmonary arterial hypertension, especially in paediatric-onset cases. A growing body of literature has expanded its phenotypic spectrum, which now includes developmental lung lesions, raging from mild to lethal.

What does this study add to the field?

Our study identifies genetic determinants of TBX4 disease heterogeneity, inclusive of newly described gain-of-function missense variants associated with later onset lung disease. It draws important conclusions on genotype-phenotype associations, informative to TBX4 variant interpretation and genetic counselling of affected families.

Abstract (250 words)

Rationale: Despite the increased recognition of *TBX4*-associated pulmonary arterial hypertension (PAH), genotype-phenotype associations are lacking and may provide important insights.

Methods: We assembled a multi-center cohort of 137 patients harboring monoallelic *TBX4* variants and assessed the pathogenicity of missense variation (n = 42) using a novel luciferase reporter assay containing T-BOX binding motifs. We sought genotype-phenotype correlations and undertook a comparative analysis with PAH patients with *BMPR2* causal variants (n = 162) or no identified variants in PAH-associated genes (n = 741) genotyped via the NIHR BioResource - Rare Diseases (NBR).

Results: Functional assessment of *TBX4* missense variants led to the novel finding of gain-of-function effects associated with older age at diagnosis of lung disease compared to loss-of-function ($p = 0.038$). Variants located in the T-BOX and nuclear localization domains were associated with earlier presentation ($p = 0.005$) and increased incidence of interstitial lung disease ($p = 0.003$). Event-free survival (death or transplantation) was shorter in the T-BOX group ($p = 0.022$) although age had a significant effect in the hazard model ($p = 0.0461$). Carriers of *TBX4* variants were diagnosed at a younger age ($p < 0.001$) and had worse baseline lung function (FEV1, FVC) ($p = 0.009$) compared to the *BMPR2* and no identified causal variant groups.

Conclusions: We demonstrated that *TBX4* syndrome is not strictly the result of haploinsufficiency but can also be caused by gain-of-function. The pleiotropic effects of *TBX4* in lung disease may be in part explained by the differential effect of pathogenic mutations located in critical protein domains.

Keywords: pulmonary arterial hypertension, *TBX4*, interstitial lung disease, lung developmental disease, gain-of-function

Introduction (326 words)

Pulmonary Arterial Hypertension (PAH) is a rare progressive vasculopathy characterized by abnormal cell proliferation in the pulmonary arterioles leading to increased pulmonary artery pressure and ultimately, right ventricular failure (1). It can occur idiopathically or in association with other medical conditions, such as congenital heart disease, or exposure to certain drugs and toxins. Since the landmark discovery of BMPR2 gene in 2000 as the main cause of heritable PAH (2), high-throughput sequencing has implicated more than 20 genes with identified causal variants in up to 25% of individuals diagnosed with idiopathic PAH (IPAH) (3–5). Monoallelic pathogenic variants in the T-BOX transcription factor 4 (*TBX4*) gene are the second commonest heritable cause of PAH, often enriched in pediatric cohorts (6–8). *TBX4* belongs to the T-BOX family of transcription factors playing a critical role in early hindlimb development (9–11) and branching of the lungs (12) regulating the expression of Fibroblast Growth Factor 10 (*FGF10*) together with *TBX5* (13, 14). Transcriptome analysis suggests it continues to be active following organogenesis and plays an important role in the cellular homeostasis of adult lung fibroblasts (15).

TBX4 sequence variants and contiguous gene deletions, as part of the recurrent chromosome 17q23.2 microdeletion, were originally reported in association with small patella syndrome (SPS) [MIM# 147891] (16–18). More recently, evidence is emerging that *TBX4* single nucleotide variants (SNVs)/deletions are not only causative of PAH, but also of a wide spectrum of developmental parenchymal lung disorders (19, 20). Partial or complete loss of a single *TBX4* functional allele is sufficient for the production of the phenotype, termed haploinsufficiency. Although the above disorders are dominantly inherited, both variable expressivity and reduced penetrance have been observed (6).

Herein, we assembled a large cohort of patients harboring *TBX4* sequence variants to establish genotype-phenotype correlations. We developed *in vitro* assays to assess the pathogenicity of non-truncating variants. This led to the novel finding of functionally distinguishable missense variation, causing either gain- (GoF) or loss-of-function (LoF).

Methods (528 words)

Using the term “TBX4”, our PubMed search identified 22 publications limited to human studies dating from 2004 to 2021 (supplementary data, xlsx). Cases with the recurrent 17q23.2 deletion were excluded from this study’s scope. We collected phenotypic information on 137 heterozygous carriers of *TBX4* sequence variants, including 15 novel cases from the National Institute for Health Research BioResource–Rare Diseases (NBR) study (21), the UK National Cohort Study of Idiopathic and Heritable PAH (3, 5), the Spanish registry of PAH (22), the registry of the French PAH network (23, 24), and the DECIPHER database (25). Demographic and phenotypic information at the time of diagnosis was captured from relevant publications and databases (supplementary data, xlsx). Follow-up information was obtained, where available.

Annotation of variants was harmonized to the *TBX4* canonical transcript NM_018488.3 of the human reference genome assembly GRCh37/hg19 using the Mutalyzer and Ensembl Variant Effect Predictor web services. We assessed variant pathogenicity according to the American College of Medical Genetics and Genomics (ACMG) guidelines using the VarSome Clinical tool followed by manual curation (26). Non-truncating variants (missense, indels) and variants predicted to affect splicing were selected for functional characterization (supplementary data, docx). Truncating variants (frameshift, nonsense) were presumed to cause LoF; p.Tyr127Ter pathogenic variant was used as a positive control in functional experiments.

Among the T-BOX family members, TBX4 has the highest similarity to TBX5 (27). Since no structure for TBX4 has been reported, we analyzed the effect of variants using the crystal structure of the TBX5 complex with DNA (Figure 3) (28). Variants were sub-grouped by protein domains, including the highly-conserved T-BOX domain (codons 71-251) containing the first nuclear localization segment NLS1 (91-103), and the predicted NLS2 (338-351) and transactivating region (351-393) (29, 30).

We sought genotype-phenotype associations in carriers of likely pathogenic/pathogenic *TBX4* variants (excluded functionally assessed variants classed as benign/likely benign or of uncertain significance) and undertook a comparative analysis with genotyped PAH patients recruited to the NBR study, termed in this paper as *BMRP2* (n = 162) and no identified causal variant (n = 741) groups (Table 1). The latter included patients with idiopathic PAH and likely pathogenic/pathogenic *BMRP2* variants or no identified likely pathogenic/pathogenic variants in PAH-associated genes (*BMPR2*, *TBX4*, *EIF2AK4*, *SMAD1/4/9*, *CAV1*, *KCNK3*, *ENG*, *ALK1*, *GDF2*, *AQP1*, *ATP13A3*, *SOX17*) (3). We additionally obtained computed tomography (CT) chest images from a subset of patients from the *BMPR2* (n = 34) and no identified causal variant (n = 143) groups and compared these to 11 *TBX4* cases from the NBR study. Details of the radiological sub-study can be found in the online supplement (docx). We assessed genotype-phenotype differences in the range of phenotypic features expected to differentiate the three selected genotypes. As this analysis was preplanned and the results consistent we did not perform correction for multiple comparisons.

Statistical analysis was performed using R (www.r-project.org). The R package "survival" was used to compare event-free survival between different groups. Survival was estimated by the Kaplan-Meier method from the time of diagnosis to death or

transplantation. To avoid immortal time bias, this was limited to a 10-year interval. Gender and age at diagnosis of lung disease were included as covariates in the semi-parametric Cox-proportional hazard models.

Results (1096 words)

Study population

We identified 137 heterozygous carriers of *TBX4* variants, the majority of which were sporadic cases (n=127, 93%). Out of four identified families, eight related individuals with available detailed phenotypic information were included in our analyses. Twenty-one cases had a primary phenotype of SPS. In the remaining 116 individuals presenting with lung disease, subsequent assessment for SPS was lacking in most, with only 29 cases (25%) reported having associated skeletal features. PAH was the predominant primary lung phenotype (Table 1). Median age at diagnosis of lung disease was 14 years (IQR: 2- 47 years). Fifty-three individuals (45%) presented in adulthood, 41 (36%) in childhood, and 22 (19%) in the perinatal period. In the overall patient cohort, there was an equal female-to-male ratio with an observed female predominance (62%) in the lung disease group.

Spectrum and functional assessment of *TBX4* variants

A total of 108 distinct *TBX4* variants were retrieved from the literature and aforementioned databases (Figure 1). Of these, 43 were missense, 39 frameshift, 15 nonsense, 6 indels including an additional deletion of whole exon 5, and 3 variants predicted to affect splicing (supplementary data, xlsx). A single case with a variant in the *TBX4* promoter region was excluded from our genotype-phenotype analyses (23).

We assessed the pathogenicity of all indels and 42/43 missense variants using a luciferase reporter assay (Figure 2). Variant c.1021G>C was assessed by a minigene

assay instead as predicted to affect the same donor splice site as in c.1021+1G>A, which resulted in a double exon skipping. All indels and 23 missense variants caused LoF with another 11 shown to be benign. Eight missense variants resulted in GoF with a mean Relative Luciferase Units (RLU) of approximately twice the levels of the wild-type construct. We confirmed that GoF was not an artifact by checking TBX4 protein expression levels of several constructs with different outcomes in the luciferase assay by qPCR and western-blot (supplementary eFigure 1). Independent of their functional activity, all assessed variants translocated to the nucleus, replicating the translocation of wild-type and Green Fluorescent Protein tagged TBX4 (supplementary eFigure 2 & 3).

As per ACMG guidelines, our luciferase functional data altered the classification of the majority of respective variants with a total of 33/48 (67%) initially classified as variants of uncertain significance, and an equivalent number (32/48) of likely pathogenic/pathogenic variants following application of the PS3 criterion (in vitro functional studies supportive of a damaging effect) where appropriate (Supplementary eFigure 4). We evaluated the performance of in silico tool predictions for likely pathogenic/pathogenic *TBX4* missense variants, with SIFT generating the highest overall percentage of correct calls (Supplementary eFigure 5). Overall, a total of 4/8 GoF and 22/23 LoF missense variants were classed as likely pathogenic/pathogenic and included in the genotype-phenotype analyses (supplementary data, xlsx).

Finally, we assessed the functional impact of all previously reported *TBX4* variants predicted to affect splicing with recurrent variants c.702+1G>A and c.1021+1G>A inducing exon skipping events (Supplementary eFigure6). Key findings of the structural variant analysis are summarized in Figure 3.

Genotype-phenotype associations

Patients with lung disease and variants within the DNA binding T-BOX domain presented at a younger age (median [IQR]: 7.5 [1 - 18.5] years) compared to carriers of variants outside this domain (18 [3 - 51.5] years, p = 0.028). This remained true for sequence variants located in either the T-BOX domain containing NLS1 or NLS2 at the C-terminus (p = 0.005, Figure 4). Individuals with LoF variants (missense, indels) were also diagnosed at a younger age compared to GoF variant carriers (14 [5 - 31] years vs. 57 [42.5 - 59.5], p = 0.038, Figure 4).

Baseline clinical features and hemodynamic parameters did not differ significantly between protein domain groups with the exception of interstitial lung disease which was more frequently reported in carriers of *TBX4* variants located in the T-BOX domain alone (75% vs. 21%, p = 0.003) or in combination with NLS2 variants (p = 0.001). Similarly, no significant differences were observed between LoF and GoF variation and this was consistent when protein-truncating variants were added to the LoF group.

The observed primary phenotypes varied significantly between protein domains with a greater frequency of developmental lung disorders (including acinar dysplasia and congenital alveolar dysplasia) in the T-BOX group (15.4% vs. 3.8%, overall p = 0.046); when combined with variants in the NLS2 domain, this remained statistically significant (overall p = 0.003). Lung histology was available in 17 previously published *TBX4* cases (see supplementary docx and xlsx for details). A greater frequency of a confirmed secondary phenotype of SPS was observed in carriers of variants located outside the T-BOX and NLS2 domains (29.7% vs. 11.4%, p = 0.038) and was more prevalent in protein-truncating versus missense variation (27.8% vs. 7.1%, p = 0.044). Variant exon location did not appear to have any phenotypic impact (data not shown).

As previously reported (31), patients with *TBX4*-associated lung disease presented at a younger age (median [IQR]: 14 [2 - 48] years) compared to patients in the *BMPR2* (39 [31 – 51] years) and no identified causal variant (51 [38 - 66] years, $p < 0.001$) groups. They performed better at the six-minute walk test with no significant differences in functional class at presentation (Table 1). They also had worse baseline lung function (FEV1, FVC) despite presentation at a younger age (Table 1). The frequency of airway / acinar abnormalities was significantly greater in the *TBX4* (87.5%) versus the *BMPR2* (32.4%) and no identified causal variant groups (33.6%, $p = 0.009$) (Table 2).

Impact of genotype on survival

Clinical outcomes were available for 89/115 TBX4 cases presenting with lung disease with a median follow-up of 8 years (IQR: 3 - 10 years). Event-free survival was shorter in the T-BOX domain variant group ($p = 0.022$ for log-rank test) although age at diagnosis also had a significant effect ($p = 0.0461$ for Cox-proportional hazard model, Figure 5). There were no observed differences in outcomes between GoF and LoF variants or missense versus protein-truncating variants. Overall, event-free survival was longer in the *TBX4* group compared to *BMPR2* and no identified causal variants groups ($p = 0.0025$). Pairwise comparisons showed no significant differences between *TBX4* and *BMPR2* variant carriers ($p = 0.69$). Compared to individuals with no causal sequence variants, both patients with *TBX4* ($p = 0.035$) and *BMPR2* variants ($p = 0.016$) had higher event-free survival rates. However, genotype did not have a significant effect on survival following correction for age and sex; both male gender ($p = 0.0002$) and older age at the time of diagnosis ($p < 2e\text{-}16$) were associated with shorter survival in the Cox-proportional hazard model.

Discussion (1498 words)

Distinct mutational mechanisms disrupt *TBX4* function

A wide range of *TBX4* mutations can result in human disease. Intragenic sequence variants occur throughout the gene and are mostly private. The functional impact of *TBX4* variants was not previously investigated and a haploinsufficient effect was assumed. Our assessment of *TBX4* missense variants reported to date led to the discovery of two distinct functional classes of variants, GoF or LoF. Although this constitutes a novel finding, GoF variants have been reported in other members of the T-BOX gene family in association with a similar phenotypic spectrum caused by LoF. This includes carriers of missense variants in the *TBX1* gene with velocardiofacial syndrome, a single *TBX5* variant in a family with Holt-Oram syndrome, and a *TBX20* variant in a family with atrial septal defects and valvular disease (32–34). All of the above were located in the highly-conserved T-BOX domain with the exception of a single *TBX1* variant in exon 8 of 9 (c.928G>A). Structural analyses were suggestive of increased protein stability; our respective analysis was indicative of the detrimental effects of LoF variants with no simple explanation for GoF variation (Figure 3).

Out of 31 pathogenic missense *TBX4* variants reported in our study, 8 resulted in GoF including the recurrent variant c.432G>T (22). These were located across the gene with 3 in the T-BOX region and 2 in the transactivation domain (Figure 1). Two variants (c.743G>T and c.1592A>G) were originally reported in 2004 in 2 Dutch families with classical SPS and constituted one of the first reports implicating *TBX4* in this phenotype (16). The remaining ones were described in association with adult-onset lung disease diagnosed as late as 81 years with the exception of a neonate presenting with interstitial lung disease (Table 3). Notably, 6/8 GoF variants were present in control populations (gnomAD database) including variant c.104C>T with the highest

overall population frequency in the cumulative *TBX4* variant dataset. As a result, many of these were classified as likely benign/uncertain clinical significance and cannot be considered as responsible for the underlying phenotype at present (Table 3).

It remains to be seen whether these GoF variants are purely hypermorphic, increasing the protein's function, or neomorphic, causing ectopic expression or acquisition of a new function. This may in turn influence their exerted phenotypic effects. From a mechanistic perspective, our knowledge of the consequences of *TBX4* overexpression during embryogenesis and/or following organogenesis is lacking. Transcription can be affected not only by decreased but also by excessive amounts of transcription factors (35). Comparison of likely pathogenic/pathogenic gain- and loss-of-function missense variation showed a later onset lung disease in the former group (Figure 4), which may suggest a milder influence on the phenotype with additional genetic modifiers likely at play (20). A number of functionally assessed *TBX4* missense variants appeared to be hypomorphic, resulting in reduced levels of transcriptional activity but not complete LoF (Figure 2, eFigure 7); the vast majority of null variants were located in the T-BOX domain whereas hypomorphic variants spread across the gene. However, the above differences in transcriptional activity alone were not sufficient to explain the diverse phenotypic spectrum of *TBX4* disease (no significant genotype-phenotype correlations).

Modifiers of *TBX4* disease spectrum

Our study captures the variable expressivity of *TBX4* pathogenic variation, with additionally observed phenotypic differences in recurring variants (supplementary data, xlsx). Aside from hypertensive pulmonary vascular disease, distal lung development can be disrupted to a variable degree raising the issue of *TBX4* disease classification under World Health Organization Group 3, pulmonary hypertension

associated with hypoxia and lung disease, as opposed to Group 1, idiopathic/heritable PAH. Although the underlying mechanisms are not yet fully elucidated, the impact of genotype was discernible with a greater frequency of developmental lung disorders and interstitial lung disease in individuals harboring pathogenic variants located in the T-BOX domain and nuclear localization segments, NLS1 and NLS2. In contrast, SPS was more frequently observed as a secondary phenotype when variants occurred outside the above *TBX4* regions.

Despite the younger age at presentation in TBX4 compared to BMPR2 variant carriers, baseline lung function was worse with evidence of possible disrupted alveolarization on CT imaging suggestive of underlying developmental lung lesions. It can be postulated that this partly accounts for the variable age of onset of *TBX4* associated pulmonary hypertension with less pulmonary reserve and increased susceptibility to external insults acting as environmental modifiers of penetrance. There is insufficient data at present to suggest a specific PAH treatment approach, distinct from other aetiologies, for TBX4 disease. However, at the point of diagnosis, it is important to evaluate the potential impact of any parenchymal abnormalities versus pure vascular disease to adapt treatment accordingly.

Gene-specific variant classification and prognostic value

Inconsistent variant interpretation can not only lead to misdiagnosis of individual patients but also have significant consequences for at-risk relatives through inappropriate predictive testing. Gene-specific knowledge overcomes some of these pitfalls, especially when semi-automated impact analysis tools are used for variant classification. *In silico* predictions did not reliably reflect the true effects of *TBX4* missense variants on gene activity (supplementary eFigure 4). Structural analysis was also not capable of discerning *TBX4* GoF variants. In light of these limitations, an

integrated pipeline incorporating molecular testing and functional assessment of novel *TBX4* missense variants by standardized assays would be of high diagnostic value.

Our results indicate that both protein-truncating and missense variants contribute to *TBX4* disease with the majority of the latter (26/42) classified as likely pathogenic/pathogenic following functional assessment. In light of this, we removed the ACMG BP1 criterion (missense variant in a gene for which primarily truncating variants are known to cause disease) from *TBX4* variant annotation. In addition, the PM1 criterion (variant located in critical/well-established functional domain) could be expanded to include not only variants located in the T-BOX domain but also the predicted nuclear localization segments (NLS1 and NLS2) as our study was suggestive of an equivalent effect on produced phenotypes, including a higher rate of developmental (early-onset) lung disorders.

Estimating the true penetrance of *TBX4* disease remains a high priority and impacts on variant interpretation as well as genetic counseling of respective families. A characteristic example exhibiting inter- and intra-familial variability is that of splice variant c.1021+1G>A reported by 3 independent studies (supplementary data, xlsx); index cases had adult-onset lung disease with positive family history, including first-degree relatives with severe PAH resulting in death in infancy/childhood or absent patella only (22, 36, 37). Leaky splicing variants (reducing but not completely abolishing the production of normal transcripts) can result in reduced penetrance although this phenomenon would still not explain the variable phenotypes observed in association with other types of recurrent variants (supplementary data, xlsx). We applied the ACMG BS1 criterion (allele frequency is greater than expected for the disorder) using a maximum tolerated population frequency of 5.00×10^{-8} arising from a generous estimate of 50% for lung disease penetrance (38); out of 17/42 missense

TBX4 variants present in gnomAD, several GoF and LoF variants remained of uncertain clinical significance (online supplement, docx and xlsx). Where there is enough evidence to support (likely) pathogenicity for missense variation in critical protein domains, important prognostic information can be inferred based on our genotype-phenotype analysis.

Limitations

We were limited by the multi-center nature of this study including retrospective data collection with variable follow-up duration. This methodology introduced selection bias as the vast majority of cases were previously published, a proportion of which consisted of severe cases with lung histopathology from post-mortem examination or explants at the time of lung transplantation. Published cases may not be truly representative of the natural history of TBX4 disease at large, which is yet to be elucidated via recall-by-genotype studies. We were unable to account for missing phenotypic information (presence/absence of SPS features in individuals with a primary lung phenotype), although estimation of *TBX4* disease-penetrance was not the focus of this work. The main limitation of our functional assay is that we used overexpression plasmids, forcing the expression of variants whose effect could be lowered by post-transcriptional regulation. Our reporter system was designed using the standard T-BOX motif and *FGF10* promoter sequences. A recent report of *TBX4* chromatin immunoprecipitation in human fetal lung fibroblasts identified several other potential genome-wide target sites whose effects we have not tested (39). Therefore, a variant shown to have a modest or benign effect by our luciferase assay may still have significant damaging effects on interactions with other crucial binding sites in biologically relevant tissues.

Conclusion

In summary, we combined functional and phenotypic characterization of all reported *TBX4* sequence variants to date to determine the hallmarks of *TBX4*-mediated lung disease. We used *in vitro* analyses to assess the pathogenicity of missense, indel, and splice variants, resulting in either loss- or gain-of-function effects with phenotypic and prognostic implications when also taking into account variant location in functional domains. Our knowledge of *TBX4* genotype-phenotype associations can only be furthered by active collaborations between molecular scientists and clinicians, requiring both an in-depth understanding of the biological aspects of the disease and a systematic approach to phenotyping.

References

1. Simonneau G, Montani D, Celermajer D, Denton C, Gatzoulis M, Krowka M, Williams P, Souza R. Hemodynamic definitions and updated clinical classification of pulmonary hypertension. *Eur Respir J* 2018;1–20.doi:10.1183/13993003.01913-2018.

2. Thomson JR, Machado RD, Pauciulo MW, Morgan NV, Humbert M, Elliott GC, Ward K, Yacoub M, Mikhail G, Rogers P, Newman J, Wheeler L, Higenbottam T, Gibbs JS, Egan J, Crozier A, Peacock A, Allcock R, Corris P, Loyd JE, Trembath RC, Nichols WC. Sporadic primary pulmonary hypertension is associated with germline mutations of the gene encoding BMPR-II, a receptor member of the TGF-beta family. *J Med Genet* 2000;37:741–745.

3. Gräf S, Haimel M, Bleda M, Hadinnapola C, Southgate L, Li W, Hodgson J, Liu B, Salmon RM, Southwood M, Machado RD, Martin JM, Treacy CM, Yates K, Daugherty LC, Shamardina O, Whitehorn D, Holden S, Aldred M, Bogaard HJ, Church C, Coghlan G, Condliffe R, Corris PA, Danesino C, Eyries M, Gall H, Ghio S, Ghofrani H-AA, *et al.* Identification of rare sequence variation underlying heritable pulmonary arterial hypertension. *Nat Commun* 2018;9:1416.

4. Swietlik EM, Greene D, Zhu N, Megy K, Cogliano M, Rajaram S, Pandya D, Tilly T, Lutz KA, Welch CCL, Pauciulo MW, Southgate L, Martin JM, Treacy CM, Penkett CJ, Stephens JC, Bogaard HJ, Church C, Coghlan G, Coleman AW, Condliffe R, Eichstaedt CA, Eyries M, Gall H, Ghio S, Girerd B, Grünig E, Holden S, Howard L, *et al.* Bayesian Inference Associates Rare KDR Variants With Specific Phenotypes in Pulmonary Arterial Hypertension. *Circ Genomic Precis Med* 2020;14:e003155.

5. Zhu N, Swietlik EM, Welch CL, Pauciulo MW, Hagen JJ, Zhou X, Guo Y, Karten J, Pandya D, Tilly T, Lutz KA, Martin JM, Treacy CM, Rosenzweig EB, Krishnan U, Coleman AW, Gonzaga-Juaregui C, Lawrie A, Trembath RC, Wilkins MR, Morrell NW, Shen Y, Gräf S, Nichols WC, Chung WK. Rare variant analysis of 4241 pulmonary arterial hypertension cases from an international consortium implicates FBLN2, PDGFD, and rare de novo variants in PAH. *Genome Med* 2021;13:.

6. Swietlik EM, Prapa M, Martin JM, Pandya D, Auckland K, Morrell NW, Gräf S. 'There and Back Again'-Forward Genetics and Reverse Phenotyping in Pulmonary Arterial Hypertension. *Genes* 2020;11:.

7. Welch CL, Chung WK. Genetics and Genomics of Pediatric Pulmonary Arterial Hypertension. *Genes* 2020;11:1213.

8. Southgate L, Machado RD, Gräf S, Morrell NW. Molecular genetic framework underlying pulmonary arterial hypertension. *Nat Rev Cardiol* 2019;doi:10.1038/s41569-019-0242-x.

9. Naiche LA, Harrelson Z, Kelly RG, Papaioannou VE. T-box genes in vertebrate development. *Annu Rev Genet* 2005;39:219–239.

10. Chapman DL, Garvey N, Hancock S, Alexiou M, Agulnik SI, Gibson-Brown JJ, Cebra-Thomas J, Bollag RJ, Silver LM, Papaioannou VE. Expression of the T-box family genes, Tbx1-Tbx5, during early mouse development. *Dev Dyn Off Publ Am Assoc Anat* 1996;206:379–390.

11. Gibson-Brown JJ, I Agulnik S null, Silver LM, Papaioannou VE. Expression of T-box genes Tbx2-Tbx5 during chick organogenesis. *Mech Dev* 1998;74:165–169.

12. Naiche LA, Arora R, Kania A, Lewandoski M, Papaioannou VE. Identity and fate of Tbx4-expressing cells reveal developmental cell fate decisions in the

allantois, limb, and external genitalia. *Dev Dyn Off Publ Am Assoc Anat* 2011;240:2290–2300.

13. Goss AM, Tian Y, Tsukiyama T, Cohen ED, Zhou D, Lu MM, Yamaguchi TP, Morrisey EE. Wnt2/2b and beta-catenin signaling are necessary and sufficient to specify lung progenitors in the foregut. *Dev Cell* 2009;17:290–298.

14. Arora R, Metzger RJ, Papaioannou VE. Multiple roles and interactions of Tbx4 and Tbx5 in development of the respiratory system. *PLoS Genet* 2012;8:e1002866.

15. Horie M, Miyashita N, Mikami Y, Noguchi S, Yamauchi Y, Suzukawa M, Fukami T, Ohta K, Asano Y, Sato S, Yamaguchi Y, Ohshima M, Suzuki HI, Saito A, Nagase T. TBX4 is involved in the super-enhancer-driven transcriptional programs underlying features specific to lung fibroblasts. *Am J Physiol-Lung Cell Mol Physiol* 2018;314:L177–L191.

16. Bongers EMHF, Duijf PHG, van Beersum SEM, Schoots J, van Kampen A, Burckhardt A, Hamel BCJ, Lošan F, Hoefsloot LH, Yntema HG, Knoers NVAM, van Bokhoven H. Mutations in the Human TBX4 Gene Cause Small Patella Syndrome. *Am J Hum Genet* 2004;74:1239–1248.

17. Ballif BC, Theisen A, Rosenfeld JA, Traylor RN, Gastier-Foster J, Thrush DL, Astbury C, Bartholomew D, McBride KL, Pyatt RE, Shane K, Smith WE, Banks V, Gallentine WB, Brock P, Rudd MK, Adam MP, Keene JA, Phillips JA, Pfotenhauer JP, Gowans GC, Stankiewicz P, Bejjani BA, Shaffer LG. Identification of a recurrent microdeletion at 17q23.1q23.2 flanked by segmental duplications associated with heart defects and limb abnormalities. *Am J Hum Genet* 2010;86:454–461.

18. Levin ML, Shaffer LG, Lewis RAp6 null, Gresik MV, Lupski JR. Unique de novo interstitial deletion of chromosome 17, del(17) (q23.2q24.3) in a female newborn with multiple congenital anomalies. *Am J Med Genet* 1995;55:30–32.

19. German K, Deutsch GH, Freed AS, Dipple KM, Chabra S, Bennett JT. Identification of a deletion containing TBX4 in a neonate with acinar dysplasia by rapid exome sequencing. *Am J Med Genet A* 2019;179:842–845.

20. Karolak JA, Vincent M, Deutsch G, Gambin T, Cogné B, Pichon O, Vetrini F, Mefford HC, Dines JN, Golden-Grant K, Dipple K, Freed AS, Leppig KA, Dishop M, Mowat D, Bennetts B, Gifford AJ, Weber MA, Lee AF, Boerkoel CF, Bartell TM, Ward-Melver C, Besnard T, Petit F, Bache I, Tümer Z, Denis-Musquer M, Joubert M, Martinovic J, *et al.* Complex Compound Inheritance of Lethal Lung Developmental Disorders Due to Disruption of the TBX-FGF Pathway. *Am J Hum Genet* 2019;104:213–228.

21. Turro E, Astle WJ, Megy K, Gräf S, Greene D, Shamardina O, Allen HL, Sanchis-Juan A, Frontini M, Thys C, Stephens J, Mapeta R, Burren OS, Downes K, Haimel M, Tuna S, Deevi SVV, Aitman TJ, Bennett DL, Calleja P, Carss K, Caulfield MJ, Chinnery PF, Dixon PH, Gale DP, James R, Koziell A, Laffan MA, Levine AP, *et al.* Whole-genome sequencing of patients with rare diseases in a national health system. *Nature* 2020;583:96–102.

22. Hernandez-Gonzalez I, Tenorio J, Palomino-Doza J, Meñaca AM, Ruiz RM, Lago-Docampo M, Gomez MV, Roman JG, Valls ABE, Perez-Olivares C, Valverde D, Carbonell JG, Rodríguez-Monte EGL, del Cerro MJ, Lapunzina P, Escribano-Subias P. Clinical heterogeneity of Pulmonary Arterial Hypertension associated with variants in TBX4. In: West J, editor. *PLoS ONE* 2020;15:e0232216.

23. Thoré P, Girerd B, Jaïs X, Savale L, Ghigna M-R, Eyries M, Levy M, Ovaert C, Servettaz A, Guillaumot A, Dauphin C, Chabanne C, Boiffard E, Cottin V, Perros F, Simonneau G, Sitbon O, Soubrier F, Bonnet D, Remy-Jardin M, Chaouat A, Humbert M, Montani D. Phenotype and outcome of pulmonary arterial hypertension patients carrying a TBX4 mutation. *Eur Respir J* 2020;55:1902340.

24. Galambos C, Mullen MP, Shieh JT, Schwerk N, Kielt MJ, Ullmann N, Boldrini R, Stucin-Gantar I, Haass C, Bansal M, Agrawal PB, Johnson J, Peca D, Surace C, Cutrera R, Pauciulo MW, Nichols WC, Griese M, Ivy D, Abman SH, Austin ED, Danhaive O. Phenotype characterisation of TBX4 mutation and deletion carriers with neonatal and paediatric pulmonary hypertension. *Eur Respir J* 2019;54:.

25. Bragin E, Chatzimichali EA, Wright CF, Hurles ME, Firth HV, Bevan AP, Swaminathan GJ. DECIPHER: database for the interpretation of phenotype-linked plausibly pathogenic sequence and copy-number variation. *Nucleic Acids Res* 2014;42:D993–D1000.

26. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015;17:405–424.

27. Papaioannou VE. The T-box gene family: emerging roles in development, stem cells and cancer. *Dev Camb Engl* 2014;141:3819–3833.

28. Stirnimann CU, Ptchelkine D, Grimm C, Müller CW. Structural basis of TBX5-DNA recognition: the T-box domain in its DNA-bound and -unbound form. *J Mol Biol* 2010;400:71–81.

29. Kulisz A, Simon H-G. An evolutionarily conserved nuclear export signal facilitates cytoplasmic localization of the Tbx5 transcription factor. *Mol Cell Biol* 2008;28:1553–1564.

30. Murakami M, Nakagawa M, Olson EN, Nakagawa O. A WW domain protein TAZ is a critical coactivator for TBX5, a transcription factor implicated in Holt-Oram syndrome. *Proc Natl Acad Sci U S A* 2005;102:18034–18039.

31. Zhu N, Gonzaga-Jauregui C, Welch CL, Ma L, Qi H, King AK, Krishnan U, Rosenzweig EB, Ivy DD, Austin ED, Hamid R, Nichols WC, Pauciulo MW, Lutz KA, Sawle A, Reid JG, Overton JD, Baras A, Dewey F, Shen Y, Chung WK. Exome Sequencing in Children With Pulmonary Arterial Hypertension Demonstrates Differences Compared With Adults. *Circ Genomic Precis Med* 2018;11:e001887.

32. Zweier C, Sticht H, Aydin-Yaylagül I, Campbell CE, Rauch A. Human TBX1 missense mutations cause gain of function resulting in the same phenotype as 22q11.2 deletions. *Am J Hum Genet* 2007;80:510–517.

33. Postma AV, Van De Meerakker JBA, Mathijssen IB, Barnett P, Christoffels VM, Ilgun A, Lam J, Wilde AAM, Deprez RHL, Moorman AFM. A gain-of-function TBX5 mutation is associated with atypical Holt-Oram syndrome and paroxysmal atrial fibrillation. *Circ Res* 2008;102:1433–1442.

34. Posch MG, Gramlich M, Sunde M, Schmitt KR, Lee SHY, Richter S, Kersten A, Perrot A, Panek AN, Al Khatib IH, Nemer G, Mégarbané A, Dietz R, Stiller B, Berger F, Harvey RP, Ozcelik C. A gain-of-function TBX20 mutation causes congenital atrial septal defects, patent foramen ovale and cardiac valve defects. *J Med Genet* 2010;47:230–235.

35. Veitia RA, Birchler JA. Dominance and gene dosage balance in health and disease: why levels matter! *J Pathol* 2010;220:174–185.

36. Shrivastava S, Kruisselbrink TM, Mohananey A, Thomas BC, Kushwaha SS, Pereira NL. Rare TBX4 Variant Causing Pulmonary Arterial Hypertension With Small Patella Syndrome in an Adult Man. *JACC Case Rep* 2021;3:1447–1452.

37. Zhu N, Pauciulo MW, Welch CL, Lutz KA, Coleman AW, Gonzaga-Jauregui C, Wang J, Grimes JM, Martin LJ, He H, PAH Biobank Enrolling Centers' Investigators, Shen Y, Chung WK, Nichols WC. Novel risk genes and mechanisms implicated by exome sequencing of 2572 individuals with pulmonary arterial hypertension. *Genome Med* 2019;11:69.

38. Whiffin N, Minikel E, Walsh R, O'Donnell-Luria AH, Karczewski K, Ing AY, Barton PJR, Funke B, Cook SA, MacArthur D, Ware JS. Using high-resolution variant frequencies to empower clinical genome interpretation. *Genet Med Off J Am Coll Med Genet* 2017;19:1151–1158.

39. Karolak JA, Gambin T, Szafranski P, Stankiewicz P. Potential interactions between the TBX4-FGF10 and SHH-FOXF1 signaling during human lung development revealed using ChIP-seq. *Respir Res* 2021;22:26.

40. Kerstjens-Frederikse WS, Bongers EMHF, Roofthooft MTR, Leter EM, Douwes MJ, Dijk AV, Vonk-Noordegraaf A, Dijk-Bos KK, Hoefsloot LH, Hoendermis ES, Gille JJP, Sikkema-Raddatz B, Hofstra RMW, Berger RMF. TBX4 mutations (small patella syndrome) are associated with childhood-onset pulmonary arterial hypertension. *J Med Genet* 2013;50:500–506.

Figure legends

Figure 1. Lolliplot depicting *TBX4* mutation spectrum. Recurrently mutated positions are represented by a proportionally sized lollipop. Critical protein domains are highlighted, inclusive of the DNA binding T-BOX, nuclear localization segments (NLS1 and NLS2), and transactivation domains. Variants are grouped by primary associated phenotype and color-coded taking into account the functional assessment of missense and inframe insertion/deletions (indels). Abbreviations: GoF, gain-of-function; LoF, loss-of-function.

Figure 2. Functionally assessed *TBX4* variants by luciferase assay inducing gain- or loss-of-function. A) Schematic representation of the *in vitro* dual-luciferase reporter assay. We co-transfected three different plasmids: Firefly luciferase with x3 T-BOX motifs as the promoter, Renilla luciferase, and TBX4 overexpression plasmid (wild type/mutated). B) Variants inducing gain- or loss-of-function grouped by primary phenotype; lung disease (perinatal-, childhood-, and adult-onset) or small patella syndrome (SPS). The y-axis represents the Relative Luciferase Units; the dashed line marks the level of the wild type. Data is shown as box plots representing median ± quartiles. Dots represent biological replicates with corresponding batches in different shapes.

Figure 3. Structural analysis of *TBX4* sequence variants. The crystal structure of TBX5 bound to DNA, pdb code 2X6V, was used for structural analysis. They share 52.6% sequence along with the full-length proteins and 93.9% in the T-BOX domain. A) Sequence alignment of TBX4 and TBX5 in the T-BOX region (highlighted in grey) containing the DNA-binding motif as well as the nuclear localization segment 1 (NLS1, in cyan) and the nuclear export segment (NES, in yellow)(29, 30). *TBX4* missense variants are indicated in bold/blue, with indels highlighted in magenta. Residues visible

in the TBX5 structure are shown in light blue letters. B) Mutations plotted on the TBX5 crystal structure as spheres. Cyan, yellow, and magenta spheres correspond to the NLS1, NES regions, and indels as indicated in A. When annotating loss-of-function variants on the TBX4 sequence, they are highly enriched in the T-BOX, particularly the NLS1 and NES. C) Some mutations of the non-interface residues, such as TBX4 p.Glu86 and p.Tyr127 (corresponding residues p.Glu73 and p.Tyr114 in TBX5, respectively), make essential interactions to stabilize the secondary structural elements required for T-BOX binding to DNA. Clustal Omega was used for sequence alignment. Figures were generated using PyMOL Molecular Graphics System.

Figure 4. Age at diagnosis of lung disease by *TBX4* genotype; (A) variants in the T-BOX and second nuclear localization segment (NLS2) versus other domains, (B) likely pathogenic/pathogenic gain-of-function versus loss-of-function missense variants.

Figure 5. Time to death or lung transplantation (years) by TBX4 protein domain genotype. Event-free survival was shorter in the T-BOX domain variant group although age had a significant effect in the hazard model.

Tables

Table 1. Demographic and clinical characteristics at diagnosis of the patient population included in the genotype-phenotype dataset. Heterozygous carriers of *TBX4* variants shown to be benign by our functional studies were excluded.

	BMPR2 **(n=162)**	**No causal variant identified (n=741)**	***TBX4*** **(n=98)**	**p value**	**Available (total n)**
Primary diagnosis					1001
1.1 Idiopathic PAH *including drug- and toxin-induced	107 (66.0%)	741 (100%)	58 (59.2%)		
1.2 Heritable PAH	53 (32.7%)	-	17 (17.3%)		
1.4.1 PAH associated with connective tissue disease	1 (0.62%)	-	2 (2.04%)		
1.4.4 PAH associated with congenital heart disease	1 (0.62%)	-	11 (11.2%)		
1.6 PAH with overt features of venous / capillaries (PVOD/PCH) involvement	-	-	1 (1.02%)		
3.5 Developmental lung disorders	-	-	9 (9.18%)		
Sex: female	107 (66.0%)	530 (71.5%)	62 (63.9%)	0.156	1000
Smoking history: past/current *adults only	53 (39.3%)	293 (53.3%)	10 (47.6%)	0.039	706
Exposure to drug or toxins: yes	6 (3.70%)	45 (6.07%)	7 (19.4%)	0.007	939
Ethnicity					966
European	137 (84.6%)	630 (85.0%)	53 (84.1%)		
Finnish-European	-	1 (0.13%)	-		

African	2 (1.23%)	20 (2.70%)	5 (7.94%)		
East-Asian	2 (1.23%)	6 (0.81%)	1 (1.59%)		
South-Asian	6 (3.70%)	48 (6.48%)	1 (1.59%)		
Other	15 (9.26%)	36 (4.86%)	3 (4.76%)		
Age at diagnosis of lung disease (years)	39 [31;51]	51 [38;66]	14 [2;48]	<0.001	997
Age at transplantation or death (years)	52 [43;61]	67 [53;75]	64 [1;71]	<0.001	302
WHO functional class				0.277	918
I	2 (1.24%)	15 (2.11%)	2 (4.44%)		
II	32 (19.9%)	144 (20.2%)	10 (22.2%)		
III	96 (59.6%)	466 (65.4%)	27 (60.0%)		
IV	31 (19.3%)	87 (12.2%)	6 (13.3%)		
Exercise test					
Distance (meters)	350 [276;420]	330 [210;410]	371 [308;422]	0.028	810
Pre-test sats	96.0 [94.0;98.0]	96.0 [93.0;97.0]	97.5 [95.5;98.0]	0.006	738
Post-test sats	94.0 [89.0;96.8]	91.0 [85.0;95.0]	95.5 [86.5;97.0]	0.001	683
Lung function (%pred)					
FEV1	91.0 [79.0;100]	85.0 [72.3;96.0]	82.0 [70.0;98.0]	0.001	719
FVC	99.7 (17.2)	93.2 (19.4)	88.0 (17.7)	<0.001	703
KCO	83.4 [74.2;96.5]	68.0 [49.0;83.0]	72.0 [59.8;89.8]	<0.001	516

TLC	96.0 [89.0;106]	94.0 [85.0;104]	104 [98.0;110]	0.044	511
Haemodynamics					
mPAP(mmHg)	57.0 [52.0;66.8]	52.0 [42.0;61.0]	60.5 [48.2;82.2]	<0.001	916
mPAWP(mmHg)	10.0 [7.00;12.0]	9.00 [7.00;12.0]	9.00 [7.00;11.0]	0.841	822
PVR(WU)	14.5 [10.8;20.4]	10.3 [7.06;13.9]	12.8 [8.25;16.2]	<0.001	767
CO(L/min)	3.30 [2.69;3.94]	4.04 [3.25;5.10]	3.65 [3.09;4.61]	<0.001	852
CI(L/min/m^2)	1.90 [1.51;2.23]	2.30 [1.80;2.80]	2.63 [1.98;3.20]	<0.001	503
Vasoresponders	1 (1.28%)	51 (17.7%)	6 (10.2%)	0.001	425
Increased BNP (>50 pg·mL−1) or NT-proBNP (>300 pg·mL−1)	34 (97.1%)	140 (79.1%)	5 (71.4%)	0.011	219

Abbreviations: % pred, percentage of predicted value; BNP, brain natriuretic peptide; CI, cardiac index; CO, cardiac output; FEV1; forced expiratory volume in one second; FVC, forced vital capacity; mPAP, mean pulmonary artery pressure; KCO, carbon monoxide transfer coefficient; mPAWP, mean pulmonary artery wedge pressure; NT-pro-BNP, N-terminal pro-brain natriuretic peptide; PCH, pulmonary capillary hemangiomatosis; PVOD, pulmonary veno-occlusive disease; PVR, pulmonary vascular resistance; TLC, total lung capacity; WHO, World Health Organization.

Table 2. Radiological features of Computed Tomography (CT) of the chest in pulmonary arterial hypertension (PAH) cases analyzed as part of the imaging sub-study.

	BMPR2 **(n=34)**	**No causal variant identified (n=143)**	***TBX4*** **(n=8)**	**p value**	**Available (total n)**
Vascular features:					
Axial MPA diameter (mm)	34.2 [30.9;37.4]	34.6 [31.4;38.3]	34.0 [30.7;38.5]	0.674	185
PA to Ao ratio	1.23 [1.12;1.40]	1.14 [0.99;1.29]	1.18 [0.94;1.32]	0.083	185
Neovascularity	7 (23.3%)	10 (7.69%)	-	0.048	167
Lung & mediastinal features:					
Ground-glass opacities	18 (52.9%)	62 (43.4%)	-	0.016	185
*Centrilobular pattern				0.088	185
None	17 (50.0%)	90 (62.9%)	8 (100%)		
Subtle	4 (11.8%)	21 (14.7%)	-		
Present	13 (38.2%)	32 (22.4%)	-		
**Non-specific mosaic pattern				0.441	180
None	31 (91.2%)	130 (90.9%)	3 (100%)		
Subtle	-	7 (4.90%)	-		
Present	3 (8.82%)	6 (4.20%)	-		
Fibrosis				1.000	185
None	34 (100%)	137 (95.8%)	8 (100%)		

Present	-	6 (4.2%)	-		
Pleural effusion					
None	33 (97.1%)	132 (92.3%)	8 (100%)	1.000	185
Subtle	-	3 (2.10%)	-		
Present	1 (2.94%)	8 (5.59%)	-		
Interlobular septal thickening				0.157	179
None	27 (87.1%)	122 (87.1%)	5 (62.5%)		
Subtle	3 (9.68%)	9 (6.43%)	1 (12.5%)		
Present	1 (3.23%)	9 (6.43%)	2 (25.0%)		
Mediastinal lymphadenopathy				0.721	185
None	29 (85.3%)	122 (85.3%)	6 (75.0%)		
Present	5 (14.7%)	21 (14.7%)	2 (25.0%)		
Airway/acinar features:	11 (32.4%)	48 (33.6%)	7 (87.5%)	0.009	185
Bronchial wall thickening				0.051	179
None	26 (83.9%)	126 (90.0%)	6 (75.0%)		
Subtle	3 (9.68%)	11 (7.86%)	-		
Present	2 (6.45%)	3 (2.14%)	2 (25.0%)		
Emphysema				0.026	185
None	32 (94.1%)	120 (83.9%)	5 (62.5%)		
Subtle	-	13 (9.09%)	-		

Present	2 (5.88%)	10 (6.99%)	3 (37.5%)		
Air trapping				0.040	185
None	28 (82.4%)	121 (84.6%)	4 (50.0%)		
Subtle	4 (11.8%)	7 (4.90%)	1 (12.5%)		
Present	2 (5.88%)	15 (10.5%)	3 (37.5%)		
Suspected PVOD:	1 (3.03%)	8 (5.59%)	1 (12.5%)	0.500	184

Abbreviations: Ao, aortic; MPA, main pulmonary artery; PA, pulmonary artery; PVOD, pulmonary veno-occlusive disease.

Table 3. Clinical features of individuals heterozygous for *TBX4* gain-of-function missense variants.

Variant	Demographics	Primary lung phenotype & age at diagnosis	Details of lung disease & co-morbidities	Skeletal features	Follow-up	Pedigree information	ACMG classification & population frequency (gnomAD exomes)
c.104C>T; p.Ala35Val	*Dutch PAH patient cohort (40)*	PAH adult-onset	N/A	N/A	Deceased (unavailable details)	N/A	Likely Benign 0.009
c.432G>T; p.Met144Ile	Female, White Hispanic (22)	IPAH 28 years	Clinical diagnosis of possible PVOD due to radiological findings and reduced diffusion capacity. Histopathological findings (explanted lung tissue): typical PAH features, no evidence of PVOD or ILD. Unknown smoking status.	Absent	Clinical deterioration with bilateral lung transplantation 9 years following diagnosis	Maternally inherited variant, PAH ruled out at 67 years	Pathogenic 3.98E-06
c.432G>T; p.Met144Ile	Male, White Hispanic (22)	PVOD 62 years	Radiological findings typical of PVOD. Signs of heart failure , severe respiratory insufficiency. Not eligible for lung transplant due to advanced age and comorbidities. Unknown smoking status.	Absent	Deceased (26 months following diagnosis)	No known family history of PAH/SPS. Deceased parents prior to PAH diagnosis. Genetic counseling provided to the rest of the family (declined genetic testing)	Pathogenic 3.98E-06

c.652G>A; p.Val218Met	Female, White European (24)	chILD 5 months	Lung biopsy at 7 years: diffuse alveolar simplification, moderate thickened PA muscular wall, and PA fibrointimal proliferation. PFO, microcephaly, hearing loss.	Short stature, no SPS features	10 years: chILD, moderate pulmonary hypertension	Parents not tested	VUS 1.23E-04
c.743G>T; p.Gly248Val	Dutch family (16, 40)	-	Family members screened for PAH by echocardiography with none fulfilling diagnostic criteria (unavailable details)	Classical SPS phenotype, including patellar and pelvic anomalies	N/A	3-generation pedigree (family A) (16)	Pathogenic Not Reported
c.809T>G; p.Ile270Ser	Female, White European (37)	IPAH 57 years	Functional class III at diagnosis. Asthma, HTN, obesity, sleep apnoea. Never smoked.	N/A	Alive (1-year follow-up)	N/A	Likely Pathogenic 4.00E-06
c.1070C>T; p.Ala357Val	Female, White British (3)	IPAH 81 years	Functional class III at diagnosis. Radiological findings included minor bronchial wall thickening, no evidence of PVOD except mediastinal nodes, no interstitial lung disease. Hypothyroidism, IHD, T2DM, HTN, breast cancer. Never smoked.	N/A	Died in hospice 7 months following diagnosis (progressive heart failure)	N/A	VUS 1.99E-05

c.1102C>T; p.Arg368Cys	Female, African-Caribbean (3)	IPAH 39 years	Functional class III at diagnosis. Radiological findings are typical of PAH, no lung disease, PFO, hypothyroidism, chronic subdural hematoma. Unknown smoking status.	N/A	Deceased (50 years)	N/A	VUS 5.58E-05
c.1592A>G; p.Gln531Arg	Dutch family (16, 40)	-	Family members screened for PAH by echocardiography with none fulfilling diagnostic criteria (unavailable details)	Classical SPS phenotype, including patellar and pelvic anomalies	N/A	3-generation pedigree (family C) (16)	Likely Pathogenic Not Reported

Abbreviations: ACMG, American College of Medical Genetics and Genomics; chILD, childhood-onset interstitial lung disease; HTN, systemic hypertension; ILD, interstitial lung disease; IHD, ischaemic heart disease; IPAH, idiopathic pulmonary arterial hypertension; N/A, not available; PA, pulmonary artery; PAH, pulmonary arterial hypertension; PFO, patent foramen ovale; PVOD, pulmonary veno-occlusive disease; SPS, small patella syndrome; T2DM, Type 2 diabetes mellitus; VUS, variant of uncertain clinical significance.

Figures

Figure 1

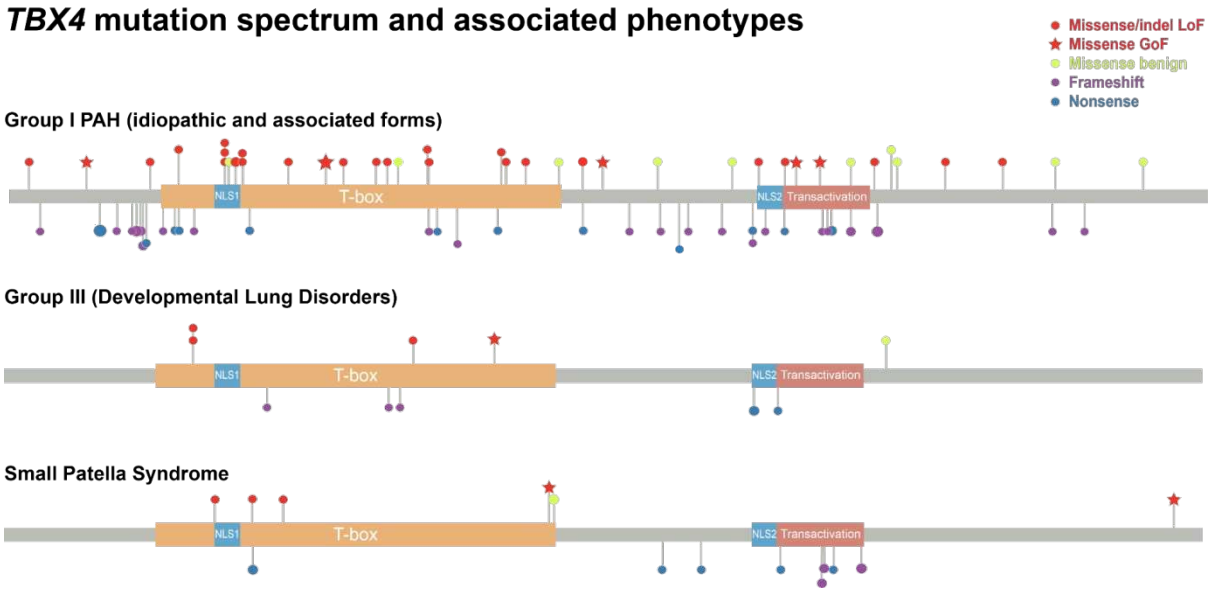


Figure 2

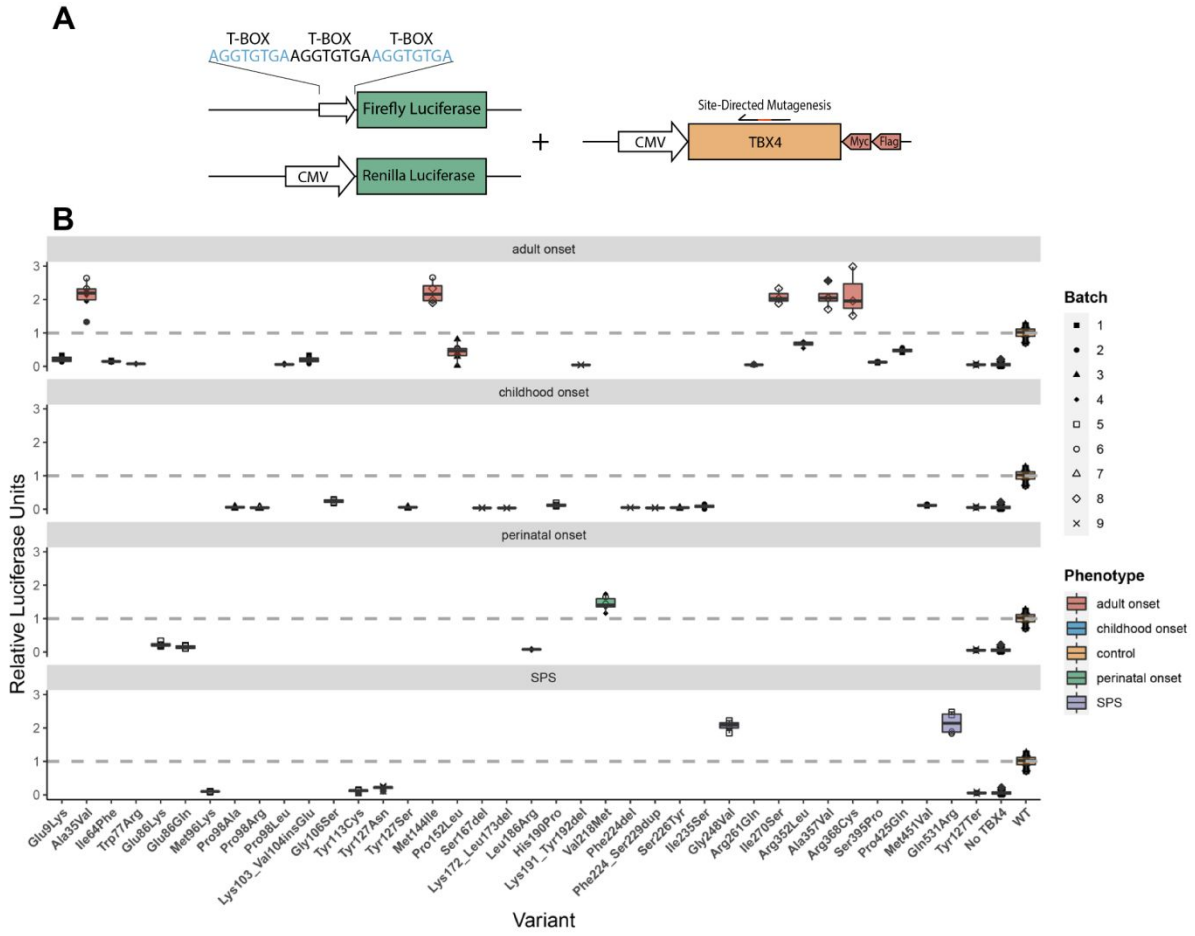

A
T-BOX
T-BOX
T-BOX
AGGTGTGAAGGTGTGAAGGTGTGA
Firefly Luciferase
CMV
Renilla Luciferase
+
Site-Directed Mutagenesis
CMV
TBX4
Myc
Flag
B
adult onset
childhood onset
perinatal onset
SPS
Relative Luciferase Units
Glu9Lys
Ala35Val
Ile64Phe
Trp77Arg
Glu86Lys
Glu86Gln
Met96Lys
Pro98Ala
Pro98Arg
Pro98Leu
Lys103_Val104insGlu
Gly106Ser
Tyr113Cys
Tyr127Asn
Tyr127Ser
Met144Ile
Pro152Leu
Ser167del
Lys172_Leu173del
Leu186Arg
His190Pro
Lys191_Tyr192del
Val218Met
Phe224del
Phe224_Ser229dup
Ser226Tyr
Ile235Ser
Gly248Val
Arg261Gln
Ile270Ser
Arg352Leu
Ala357Val
Arg368Cys
Ser395Pro
Pro425Gln
Met451Val
Gln531Arg
Tyr127Ter
No TBX4
WT
Variant
Batch
1
2
3
4
5
6
7
8
9
Phenotype
adult onset
childhood onset
control
perinatal onset
SPS

Figure 3

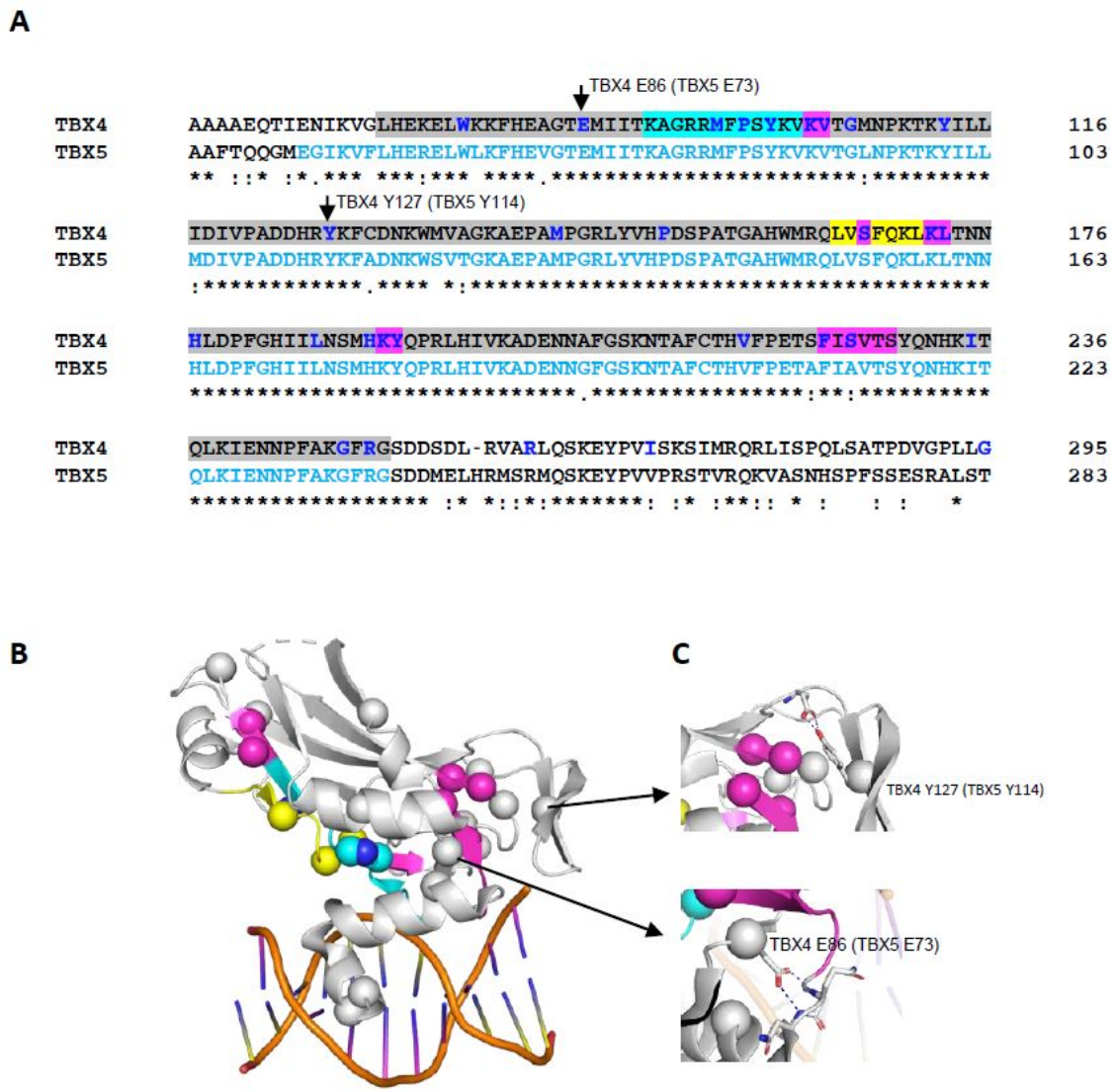

A
TBX4 E86 (TBX5 E73)
TBX4 AAAAEQTIENIKVGLHEKELWKKFHEAGTEMIITKAGRRMFPSYKVKVTGMNPKTKYILL 116
TBX5 AAFTQQGMEGIKVFLHERELWLKFHEVGTEMIITKAGRRMFPSYKVKVTGLNPKTKYILL 103
TBX4 Y127 (TBX5 Y114)
TBX4 IDIVPADDHRYKFCDNKWMVAGKAEPAMPGRLYVHPDSPATGAHWMRQLVSFQKLKLTNN 176
TBX5 MDIVPADDHRYKFADNKWSVTGKAEPAMPGRLYVHPDSPATGAHWMRQLVSFQKLKLTNN 163
TBX4 HLDPFGHIILNSMHKYQPRLHIVKADENNAFGSKNTAFCTHVFPETSFISVTSYQNHKIT 236
TBX5 HLDPFGHIILNSMHKYQPRLHIVKADENNGFGSKNTAFCTHVFPETAFIAVTSYQNHKIT 223
TBX4 QLKIENNPFAKGFRGSDDSDL-RVARLQSKEYPVISKSIMRQRLISPQLSATPDVGPLLG 295
TBX5 QLKIENNPFAKGFRGSDDMELHRMSRMQSKEYPVVPRSTVRQKVASNHSPFSSESRALST 283
B
C
TBX4 Y127 (TBX5 Y114)
TBX4 E86 (TBX5 E73)

Figure 4

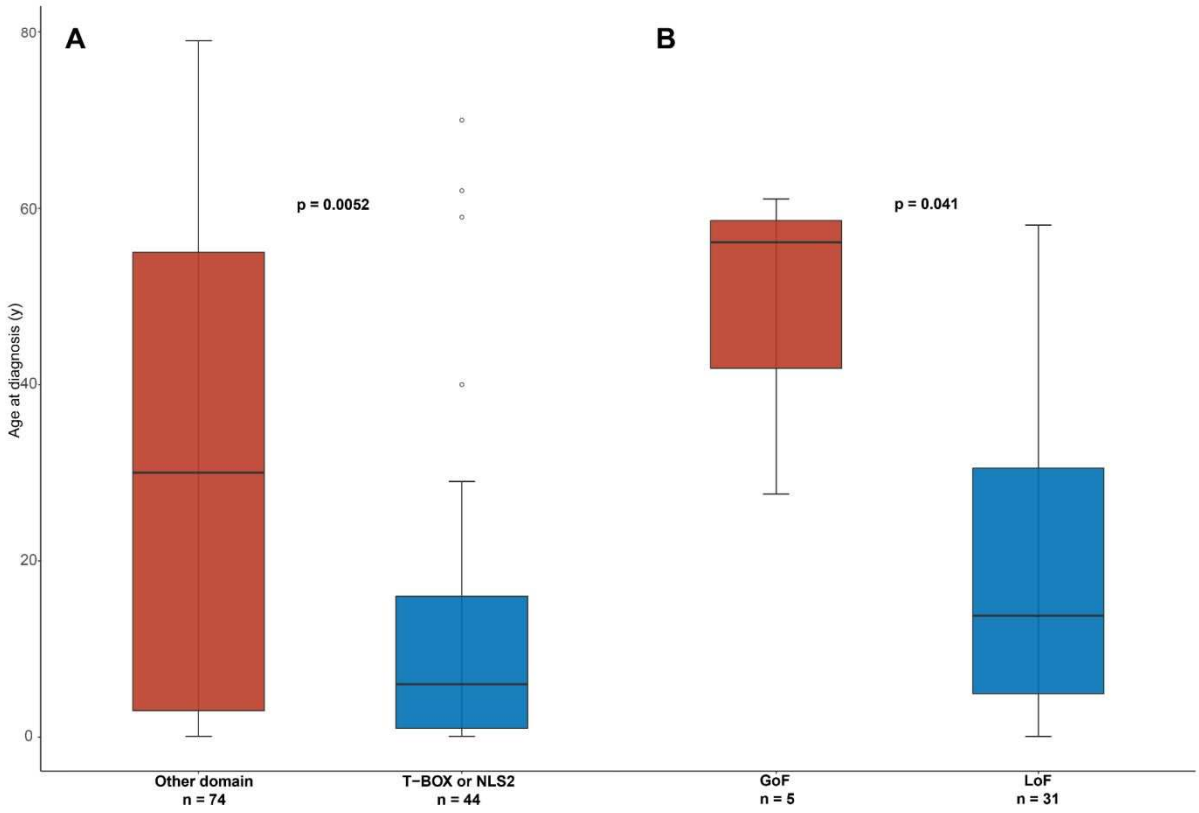
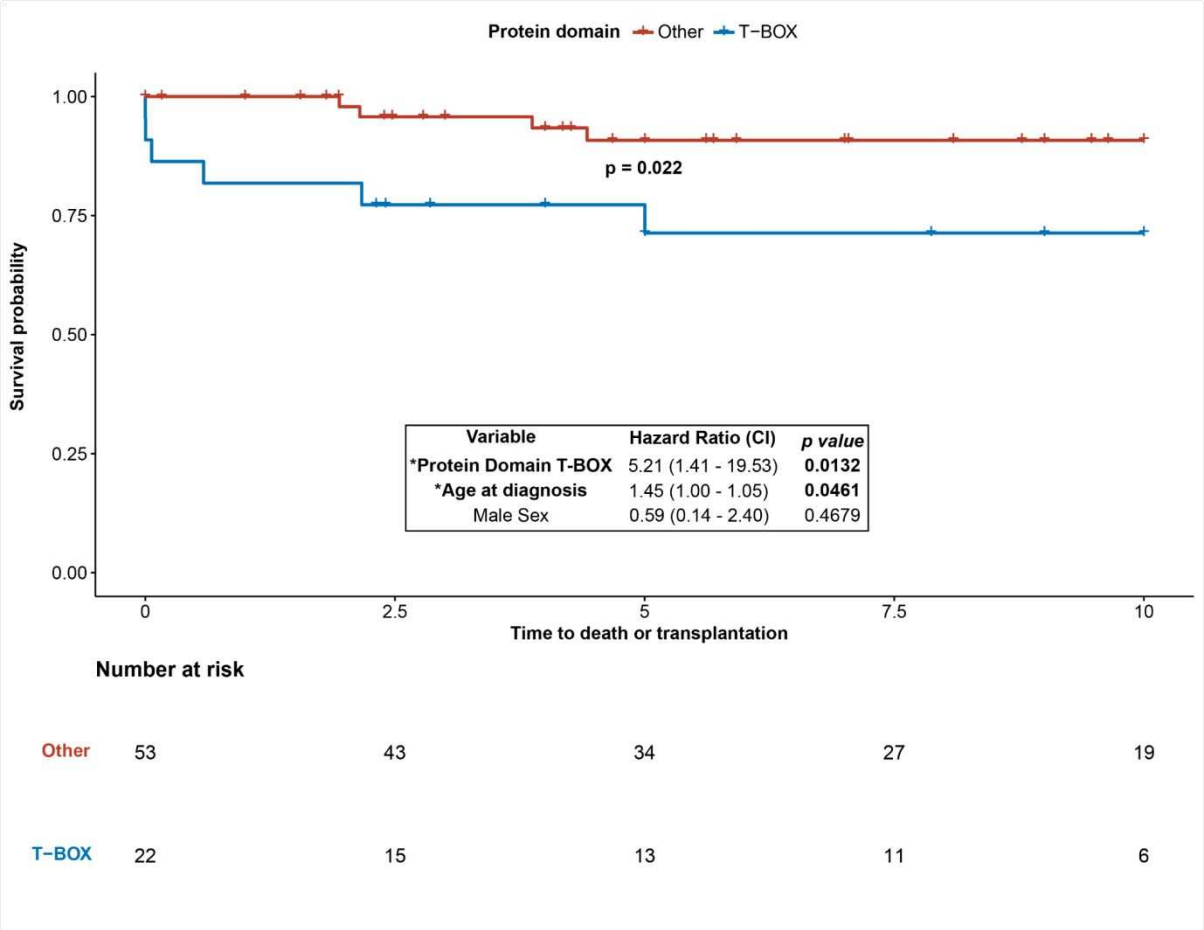

A
B
Age at diagnosis (y)
80
60
40
20
0
p = 0.0052
p = 0.041
Other domain
n = 74
T−BOX or NLS2
n = 44
GoF
n = 5
LoF
n = 31

Figure 5


Protein domain
Other
T−BOX
1.00
0.75
0.50
0.25
0.00
Survival probability
p = 0.022
Variable
Hazard Ratio (CI)
p value
*Protein Domain T-BOX
5.21 (1.41 - 19.53)
0.0132
*Age at diagnosis
1.45 (1.00 - 1.05)
0.0461
Male Sex
0.59 (0.14 - 2.40)
0.4679
0
2.5
5
7.5
10
Time to death or transplantation
Number at risk
Other
53
43
34
27
19
T−BOX
22
15
13
11
6

ONLINE DATA SUPPLEMENT

First Genotype-Phenotype Study in TBX4 Syndrome: Gain-of-Function Mutations Causative for Lung Disease

Matina Prapa[1,2,#],Mauro Lago-Docampo[3,4,#], Emilia M. Swietlik[1,5,6], David Montani[7], Mélanie Eyries[8], Marc Humbert[7], Carrie C.L. Welch[9], Wendy Chung[10], Rolf M.F. Berger[14], Ham Jan Bogaard[15], Olivier Danhaive[16,17], Pilar Escribano-Subías[18,19], Henning Gall[14], Barbara Girerd[7], Ignacio Hernandez-Gonzalez[20], Simon Holden[21], David Hunt[22], Samara M.A. Jansen[15], Wilhelmina Kerstjens-Frederikse[23], David Kiely[11,24], Pablo Lapunzina[25,26,27], John McDermott[28,29], Shahin Moledina[30], Joanna Pepke-Zaba[6], Gary J. Polwarth[6], Gwen Schotte[15], Jair Tenorio-Castaño[25,26,27], A.A. Roger Thompson[11,24], John Wharton[31], Stephen J. Wort[31], NIHR BioResource for Translational Research – Rare Diseases[32], National Cohort Study of Idiopathic and Heritable PAH[33], PAH Biobank Enrolling Centers' Investigators[33], Karyn Megy[1,5], Rutendo Mapeta[1,5], Carmen M. Treacy[1], Jennifer M Martin[1], Wei Li[1], Andrew J. Swift[11], Paul D. Upton[1], Nicholas W. Morrell[1,5,6,12,*], Stefan Gräf[1,12,13,*], Diana Valverde[3,4,*]

# Equal contribution

[1] Department of Medicine, University of Cambridge, Cambridge Biomedical Campus

[2] St George's University Hospitals NHS Foundation Trust

[3] CINBIO, Universidade de Vigo, 36310 Vigo, Spain

[4] Rare Diseases and Pediatric Medicine, Galicia Sur Health Research Institute (IIS Galicia Sur), SERGAS-UVIGO, 36312 Vigo, Spain

[5] Addenbrooke's Hospital NHS Foundation Trust, Cambridge Biomedical Campus

[6] Royal Papworth Hospital NHS Foundation Trust, Cambridge Biomedical Campus

[7] Université Paris-Sud, Faculté de Médecine, Université Paris-Saclay; AP-HP, Service de Pneumologie, Centre de référence de l'hypertension pulmonaire; INSERM UMR_S 999, Hôpital Bicêtre, Le Kremlin-Bicêtre

[8] Département de génétique, hôpital Pitié-Salpêtrière, Assistance Publique-Hôpitaux de Paris, and UMR_S 1166-ICAN, INSERM, UPMC Sorbonne Universités

[9] Columbia University Medical Center

[10] Department of Pediatrics, Columbia University

[11] Department of Infection, Immunity and Cardiovascular Disease, University of Sheffield

[12] NIHR BioResource for Translational Research, Cambridge Biomedical Campus

[13] Department of Haematology, University of Cambridge, Cambridge Biomedical Campus

[14] Centre for Congenital Heart Diseases, Pediatric Cardiology, Beatrix Children's Hospital, University Medical Center Groningen, University of Groningen, Groningen, the Netherlands.

[15] Department of Pulmonary Medicine, Amsterdam University Medical Centre, Vrije Universiteit Amsterdam, Amsterdam Cardiovascular Sciences, the Netherlands.

[16] Division of Neonatology, St-Luc University Hospital, Catholic University of Louvain, Brussels, Belgium.

[17] Department of Pediatrics, University of California San Francisco, San Francisco, CA, USA.

[18] Unidad Multidisciplinar de Hipertensión Pulmonar, Servicio de Cardiología, Hospital Universitario 12 de Octubre, 28041 Madrid, Spain

[19] CIBERCV, Centro de Investigación Biomédica en Red de Enfermedades Cardiovasculares, ISCIII, 28029 Madrid, Spain

[20] Department of Cardiology, Hospital Universitario Río Hortega, 47012 Valladolid, Spain

[21] Department of Clinical Genetics, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK.

[22] Wessex Clinical Genetics Service, Princess Anne Hospital, Southampton SO16 5YA, UK

[23] Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands.

[24] Sheffield Pulmonary Vascular Disease Unit, Royal Hallamshire Hospital

[25] Institute of Medical and Molecular Genetics (INGEMM)-IdiPAZ, Hospital Universitario La Paz-UAM, 28046 Madrid, Spain

[26] CIBERER, Centro de Investigación Biomédica en Red de Enfermedades Raras, ISCIII, 28029 Madrid, Spain

[27] ITHACA, European Reference Network on Rare Congenital Malformations and Rare Intellectual Disability, 1000 Brussels, Belgium

[28] Manchester Centre for Genomic Medicine, St Mary's Hospital, Manchester University NHS Foundation Trust, UK.

[29] Division of Evolution and Genomic Sciences, School of Biological Sciences, University of Manchester, UK.

[30] Great Ormond Street Hospital

[31] National Heart & Lung Institute, Imperial College London

[32] www.ipahcohort.com

[33] www.pahbiobank.org

*** Corresponding authors**

Nicholas W. Morrell (nwmorrel23@cam.ca.uk)

University of Cambridge, Department of Medicine

Box 157, Level 5, Addenbrooke's Hospital, Hills Road,

Cambridge, CB2 0QQ, United Kingdom

Tel: (+44) 1223 331666

Stefan Gräf (sg550@cam.ac.uk)

University of Cambridge, Department of Medicine

Addenbrooke's Hospital, Hills Road,

Cambridge, CB2 0QQ, United Kingdom

Diana Valverde (dianaval@uvigo.es)

CINBIO, Universidade de Vigo

Campus Universitario As Lagoas-Marcosende s/n

Vigo 36310, Spain

Tel: (+34) 986 811 953

I. Functional studies- expanded methods

- **A. eTable 1.** Cloning and sequencing primers.
- **B. eTable 2.** Site-directed mutagenesis primers.
- **C. eTable 3.** qPCR primers.

II. Variant assessment

A. Functional

1. **eFigure 1.** Validation of the overexpression levels of TBX4 at mRNA and protein levels.
2. **eFigure 2.** *TBX4* variants co-localize to the nuclei independently of their activity.
3. **eFigure 3.** TBX4 antibody validation for immunofluorescence.
4. **eFigure 6.** Minigene analysis of *TBX4* splice-site variants.
5. **eFigure 7.** Results of the luciferase assay for *TBX4* variants annotated as benign.
6. **eFigure 8.** Optimization of the luciferase assay.
7. **eFigure 9.** Loss of immunoreactivity of the p.Ile270Ser *TBX4* variant.

B. *In silico*

1. **eFigure 4.** Representation of the percentage number of variants with altered classification following functional assessment by luciferase assay.
2. **eFigure 5.** Comparison of the accuracy of different *in silico* prediction tools.

III. Clinical phenotype

A. Radiological sub-study

1. **eTable 4.** Reporting proforma for analysis of radiological features.

B. Histopathology

IV. Supplemental References

I. Expanded Methods

Reporter plasmids cloning

The FGF10 promoter and FGF10 “intronic island” were amplified using Phusion High Fidelity Polymerase (ThermoFisher). The FGF10 promoter was digested with XhoI + NheI and the FGF10 “intronic island” with SalI + NheI (NZYtech). Then, both were ligated for 1 hour at 22 °C following a 5:1 insert:vector ratio with pGL3 (Promega) and pmirGLO (Promega) respectively using a T4 ligase (Canvax).

The minipGL3 and minipmirGLO TBOX promoters were generated by annealing overlapping primers with x3 TBOX regions flanked by restriction enzyme sites. We annealed them by diluting 100 ng of each in 20 µL, increasing the temperature up to 95 °C for 3 minutes, and letting it cool on the bench to room temperature. Then, the products were digested with NheI + EcoRI/SalI (NZYtech) and ligated into pGL3 and pmirGLO.

For the bacterial transformation, we used 5 µl of the ligation product and 95 µl of NZYStar competent cells (NZYtech) following the manufacturer’s protocol. Transformants were screened by colony PCR using NZYTaq Green Mastermix (NZYtech) with the appropriate combination of sequencing primers from the eTable 1. We diluted the colonies in 5 µL of PBS and used 1 µL for the PCR in a total volume of 25 µL. PCR products were fractionated in a 2 % agarose gel and bands were visualized in a GelDoc EQ (BioRad). Finally, PCR bands were excised and purified prior to submission for Sanger sequencing in the *Centro de Apoio Científico Técnico á Investigación* (CACTI) of the University of Vigo.

Site-directed mutagenesis

We carried out site-directed mutagenesis for each of the variants using the primers listed in eTable 2 in the TBX4-MYC-DKK (Origene #RC217451) and the minigene constructs. The reaction was performed using NZYMut site-directed mutagenesis kit (NZYtech) following the manufacturer's protocol with slight modifications: elongation temperature was increased to 67 °C and all the reactions were supplemented with 5 % DMSO. To generate indels, we used Q5 site-directed mutagenesis kit (New England Biolabs) following the manufacturer's protocol.

The mutant screening was undertaken by colony PCR as described in the previous section. Positive colonies were expanded in 3 mL of LB media supplemented with the appropriate antibiotic and then harvested to extract plasmid DNA using NZYminiprep (NZYtech). Plasmid DNA was quantified using a Nanodrop 3000 (ThermoFisher). PCR products were Sanger sequenced to confirm the *TBX4* mutations.

Cell culture

We cultured HeLa cells (ATCC) in DMEM (Corning) supplemented with 1 % streptomycin/penicillin and 10 % Fetal Bovine Serum (FBS).

Luciferase assay

For the luciferase assay, HeLa cells were grown in 24 well plates until achieving 80-90 % confluence. Transfection was carried out using Lipofectamine 3000 (ThermoFisher) following the manufacturer's protocol, after 24 hours we removed the transfection media and added full growth media. 24 hours later cells were washed

and harvested for analysis using the DualGlo® Luciferase Assay System (Promega) according to the manufacturer's instructions. The luciferase results were read in an EnVision 2100 multilabel reader (Perkin Elmer) using white half-area 96-well plates (Greiner). Firefly luciferase results were normalized against *Renilla*, and the data were represented as a fold change of the WT.

Optimization of the luciferase assay

To optimize the reporter to use and the quantity of TBX4-overexpression plasmids, we transfected cells in 96-well plates with 40 ng of a reporter (empty pGL3/pmirGLO, miniTBX-pGL3 or miniTBX-pmirGLO) and either 40/30/20/10/0 ng of the WT TBX4 overexpression plasmid or 40 ng of a predicted Pathogenic variant (p.Val103_Lys104insGlu) and a nonsense variant control (p.Tyr127Ter). All the conditions were scaled proportionally to conduct the experiments in 24-well plates.

Real-Time quantitative PCR

We cultured HeLa cells in 24-well plates until confluence, then transfected them with 100 ng of each of the variants (WT, p.Pro98Arg, p.Met144Ile, p.Arg250Trp, p.Arg261Gln, and p.Ile270Ser) using lipofectamine 3000 following the manufacturer's protocol. Fresh medium was added the next day and the cells were harvested 24 hours later. RNA extraction was carried out using the NZY Total RNA Isolation Kit (NZYtech) following the manufacturer's protocol. We used 100 ng of RNA for retrotranscription using NZY M-MuLV First-Strand cDNA synthesis kit (NZYtech). Real-Time quantitative PCR (qPCR) was carried out using PowerUp SYBR Green Master Mix (ThermoFisher), 1 µL of 1:10 cDNA dilution, and the primers shown in eTable 3. The reaction was performed using a total volume of 15

µL in a Step-One Plus Real-Time PCR system (ThermoFisher), cycling conditions were as follows: 50 °C for 2 minutes, 95 °C for 2 minutes, 40 cycles of 95 °C for 15 seconds and 30 seconds at 60 °C; followed by a melting curve.

To normalize the expression, we used the -ΔCT method using *YWHAZ* and *ALAS1* as reference genes.

Western blot

Hela cells were grown in 6-well plates until 80-90% confluence. Then, we transfected the cells with 500 ng of plasmid DNA for each of the chosen variants (WT, p.Pro98Arg, p.Met144Ile, p.Arg250Trp, p.Arg261Gln, and p.Ile270Ser) using lipofectamine 3000 (ThermoFisher) following the manufacturer's protocol. After 24 hours, we changed the media, and the following day we extracted protein for each of the variants using RIPA buffer supplemented with protease inhibitors (Sigma).

We prepared 20 µg of protein in Laemlii's sample buffer (BioRad) containing 5 % B-mercaptoethanol (Sigma) and heated it at 95 °C for 5 minutes. Proteins were separated by SDS-Page using a 12 % mini-Protean TGX precast gel (BioRad). After electrophoresis, we transferred the proteins to a PVDF membrane using a Trans-Blot Turbo Transfer pack (BioRad) in a Trans-Blot Turbo system (BioRad) for 7 minutes at 1.3 A. We then incubated the membrane for 1 hour in a blocking buffer composed of 5 % w/v nonfat milk in Tris-buffered saline (TBS) containing 0.2 % Tween20 (TBS-T). Immunoblotting was carried out by incubating the membrane at room temperature with a 1:1000 dilution of anti-Flag-HRP antibody (Abcam #ab49763) for 1 hour. We then washed the membrane with TBS and developed the blot using Clarity Western ECL substrate (BioRad).

We treated the membrane with Restore Plus Stripping Buffer (ThermoFisher) for 5 minutes, washed it in TBS, repeated the blocking step, and incubated it at 4 °C overnight with a 1:500 dilution of anti-TBX4 antibody (Sigma #AV33739). Then, we washed the membrane three times in blocking buffer and incubated with a 1/10000 dilution of Goat anti-Rabbit IgG (Abcam #ab205718) for 1 hour at room temperature before developing the blot in the same way as before. Imaging was carried out in a ChemiDoc™ (Bio-Rad) digital camera-based imaging system. We used total protein as loading control by staining the membrane with Coomassie Brilliant Blue.

Immunofluorescence

We cultured HeLa cells in 6 well plates with a 24x24 mm glass coverslip until 80-90 % confluence was achieved. Transfection was carried out with the same conditions and for the same variants as stated in the western blot protocol. Cells were washed 3 times in PBS before being fixed with 4 % formalin for 10 minutes at 37 ºC. After washing the cells six times in PBS, we proceeded to permeabilize them in PBS+BSA 1 % (w/v) containing 0.1 % Triton X-100 (v/v). Then, we blocked them in PBS + BSA 2 % (blocking buffer) for 1 hour at room temperature. We incubated the cells with the primary antibodies overnight in blocking buffer, washed three times in blocking buffer for 5 minutes, and incubated with the secondary antibodies and 4', 6-diamidino-2-phenylindole dihydrochloride (DAPI) (1 µg/mL) in blocking buffer for 1 hour in the dark. Finally, we washed the coverslips three times for 5 minutes in PBS and mounted them in ProLong Diamond Antifade Mountant (ThermoFisher). Images were acquired using a Leica DMI6000 inverted microscope with an integrated confocal module SP5 (Leica Microsystems). All the images were processed with ImageJ (v.1.8.0). For co-localization analysis, we used the EzColocalization plugin(1).

The following antibodies and dilutions were used: anti-TBX4 (Sigma #AV33739, 1:500), anti-pancadherin (abcam #ab22744, 1:500), Alexa Fluor 488-conjugated goat anti-rabbit (ThermoFisher #A-11008, 1:1000) and Alexa Fluor 594-conjugated goat anti-mouse (ThermoFisher #A-11005, 1:1000). To validate the TBX4 antibody, we compared the results with TBX4-GFP (Origene #RG217451) expression.

Minigene assay

We used Phusion High Fidelity Polymerase (ThermoFisher) to amplify 2 regions of the *TBX4* gene, one including exon 5 (TBX4_e5_pSPL3), and another including exons 6 and 7 (TBX4_e6_7_pSPL3) using the primers described in eTable 1. We purified the amplicons, digested them with EcoRI/XhoI + NheI (NZYtech), and cloned the products into the exon trapping p.SPL3 vector (ThermoFisher). Transformants were selected with ampicillin and confirmed by Sanger sequencing using colony PCR.

Mutants were generated by site-directed mutagenesis with the primers described in eTable 2, 2.5 µg of each of the variants were used for the transfection of HeLa cells with Lipofectamine 2000 (ThermoFisher) following the manufacturer's recommendations. After 48 hours, we extracted RNA and carried out an RT-PCR in the same way as stated in the qPCR protocol. We then used 2 µL of cDNA for a PCR using Phusion High Fidelity Polymerase (ThermoFisher), the primers SA2 (5'-ATCTCAGTGGTATTTGTGAGC-3') and SD6 (5'-TCTGAGTCACCTGGACAACC-3) with the following thermocycling conditions: 98 ºC for 3 minutes, 35 cycles of 10 seconds at 98 ºC, 30 seconds at 58 ºC and 30 seconds at 72 ºC, followed by a final extension at 72 ºC for 7 minutes. Finally, we separated the PCR products by gel

electrophoresis and imaged them. PCR products were submitted for Sanger sequencing to confirm the presence/absence of the exons.

Variant classification

We assessed variant pathogenicity according to the American College of Medical Genetics and Genomics (ACMG) guidelines (2) using the VarSome Clinical tool followed by manual curation. The BP1 criterion (missense variant in a gene for which primarily truncating variants are known to cause disease) was removed whenever automatically applied by VarSome; as illustrated by this study, both truncating and missense variants are causative for TBX4-associated disease phenotypes. BS2 (variant for highly penetrant condition seen in healthy individuals) was also removed as the penetrance of *TBX4* sequence variants is suspected to be low, albeit no accurate estimates exist to date. Instead, we used the BS1 criterion (allele frequency is greater than expected for the disorder) whenever applicable. We calculated the maximum tolerated allele frequency plausible for a *TBX4* pathogenic variant using the statistical framework developed by Whiffin et al. (3) available from DECIPHER. This was estimated at 5.00×10^{-8} using a dominant inheritance model with idiopathic pulmonary arterial hypertension (IPAH) prevalence of 1 in 1.000.000 (Orphanet), TBX4 lung disease penetrance of 0.5, and maximum allelic contribution of *TBX4* to the PAH genetic architecture of approximately 5% taking into account both pediatric- and adult-onset cohorts (4–7). Where the ethnicity of individual cases was known, population frequency data (gnomAD) were filtered accordingly. The PS3/BS3 criterion (+/- damaging effect on protein function or splicing) was applied to reclassify functionally assessed variants (supplementary data, xlsx).

Statistical analysis and data visualization

For the luciferase results, we first performed a Shapiro-Wilk test to assess normality, followed by one-way ANOVA with a Bonferroni correction for multiple comparisons using the ggpubr package in R. Box plots and dot plots were generated using the R package ggplot2 (8).

eTable 1. Cloning and sequencing primers.

Primer ID	Sequence 5' - 3'	Annealing T (°C)
FGF10promF_XhoI	AAACTCGAGGCACCAACATCCATAACTCC	66
FGF10promR_NheI	AAAGCTAGCCCAATATGGAGGTCAAACGC	66
FGF10_INT_NheI_R	AAAGCTAGCCGTGTAGAGGTGCTTTTAAAGTG	65
FGF10_INT_SalI_F	AAAGTCGACGCAAACCTCAAAAGAACAGC	65
miniTBX4_pmirGLO_F	CTAGCAGGTGTGAAGGTGTGAAGGTGTGAG	55
miniTBX4_pmirGLO_R	TCGACTCACACCTTCACACCTTCACACCTG	55
miniTBX4_pGL3_F	CTAGCAGGTGTGAAGGTGTGAAGGTGTGAG	55
miniTBX4_pGL3_R	AATTCTCACACCTTCACACCTTCACACCTG	55
TBX4_exon6-7_F	AAAGGATCCTCCCCAAGGAGGGTAGTGAG	58
TBX4_exon6-7_R	AAAGCTAGCGCCCAGCTGTGATCCCTAAC	58
TBX4_exon5_F	ACAGAATTCGCACCCTGGACTTTTGCTGA	58
TBX4_exon5_R	AAAGCTAGCCCTGTCTCTGAAGGCACCAG	58
pgl3_lucN_R	CCTTATGCAGTTGCTCTCC	55
EBV-rev	GTGGTTTGTCCAAACTCATC	55
T7_F	TAATACGACTCACTATAGGG	55
TBX4_3F	TGGAAGAAGTTCCACGAGGC	55
TBX4_3F_anti	GCCTCGTGGAACTTCTTCCA	55
TBX4_4F	AGTGATGACAGTGACCTGCG	55
TBX4_4R	GGGCTTCCAGATAGGATCGC	55
TBX4_2F	CAGCTATAGCGTGCAGACGA	55
TBX4_2F_anti	TCGTCTGCACGCTATAGCTG	55
TBX4_2R	CCGTCAGTCCAGTTCTCCAC	55

eTable 2. Site-directed mutagenesis primers.

Primer ID	Sequence 5' - 3'	Annealing T (°C)
Glu9Lys_anti	GCCTCCTCGCTCTTGGACAGGCCCTTATCC	67
Glu9Lys_sense	GGATAAGGGCCTGTCCAAGAGCGAGGAGGC	67
Ala35Val_anti	GCTGAGGCCCGGCGCTACCAGCGCGGGCTCGG	67
Ala35Val_sense	CCGAGCCCGCGCTGGTAGCGCCGGGCCTCAGC	67
Ile64Phe_anti	CCTTGATGTTCTCGAAGGTCTGCTCCGCGG	67
Ile64Phe_sense	CCGCGGAGCAGACCTTCGAGAACATCAAGG	67
Trp77Arg_anti	TCGTGGAACTTCTTCCGGAGCTCCTTCTCATGC	67
Trp77Arg_sense	GCATGAGAAGGAGCTCCGGAAGAAGTTCCACGA	67
Glu86Gln_anti	AGTGATGATCATCTGGGTGCCCGCCTCG	67
Glu86Gln_sense	CGAGGCGGGCACCCAGATGATCATCACT	67
Glu86Lys_anti	CCTTAGTGATGATCATCTTGGTGCCCGCCTCGTG	67
Glu86Lys_sense	CACGAGGCGGGCACCAAGATGATCATCACTAAGG	67
Met96Lys_anti	GTAGCTGGGGAACTTCCTCCTGCCAGC	67
Met96Lys_sense	GCTGGCAGGAGGAAGTTCCCCAGCTAC	67
Pro98Ala_anti	TTACCTTGTAGCTGGCGAACATCCTCCTGCC	67
Pro98Ala_sense	GGCAGGAGGATGTTCGCCAGCTACAAGGTAA	67
Pro98Leu_anti	CTTTTACCTTGTAGCTGAGGAACATCCTCCTGCCA	67
Pro98Leu_sense	TGGCAGGAGGATGTTCCTCAGCTACAAGGTAAAAG	67
Pro98Arg_anti	CTTTTACCTTGTAGCTGCGGAACATCCTCCTGCC	67
Pro98Arg_sense	GGCAGGAGGATGTTCCGCAGCTACAAGGTAAAAG	67
Tyr100Cys_anti	CTGTGACTTTTACCTTGCAGCTGGGGAACATCCTC	67
Tyr100Cys_sense	GAGGATGTTCCCCAGCTGCAAGGTAAAAGTCACAG	67
Lys103_Val104ins_sense	CCCCAGCTACAAGGTAAAAGAAGTCACAGGCATGAACCCC	67

Lys103_Val104ins_anti	GGGGTTCATGCCTGTGACTTCTTTTACCTTGTAGCTGGGG	67
Gly106Ser_anti	TGGGGTTCATGCTTGTGACTTTTACCTTGTAGCTGG	67
Gly106Ser_sense	CCAGCTACAAGGTAAAAGTCACAAGCATGAACCCCA	67
Tyr113Cys_anti	TGTCAATCAGCAGGATACACTTGGTCTTGGGGTTC	67
Tyr113Cys_sense	GAACCCCAAGACCAAGTGTATCCTGCTGATTGACA	67
Tyr127Ser_anti	GTTGTCACAGAACTTGGAGCGATGGTCATCGGC	67
Tyr127Ser_sense	GCCGATGACCATCGCTCCAAGTTCTGTGACAAC	67
Tyr127Asn_sense	CTGCCGATGACCATCGCAACAAGTTCTGTGACAAC	67
Tyr127Asn_sense	GTTGTCACAGAACTTGTTGCGATGGTCATCGGCAG	67
Tyr127Ter_sense	GCCGATGACCATCGCTAGAAGTTCTGTGACAACAAATGG	67
Tyr127Ter_anti	CCATTTGTTGTCACAGAACTTCTAGCGATGGTCATCGGC	67
Met144Ile_anti	CAGCCTTCCTGGAATGGCTGGCTCAGC	67
Met144Ile_sense	GCTGAGCCAGCCATTCCAGGAAGGCTG	67
Pro152Leu_anti	GTGGCAGGAGAATCCAGGTGGACATACAGCC	67
Pro152Leu_sense	GGCTGTATGTCCACCTGGATTCTCCTGCCAC	67
His177Tyr_anti	CCAAAGGGGTCCAGGTAGTTGTTTGTCAGCTTCAGC	67
His177Tyr_sense	GCTGAAGCTGACAAACAACTACCTGGACCCCTTTGG	67
Leu186Arg_anti	GGTACTTGTGCATAGAGTTGCGGATGATATGGCCAAAGG	67
Leu186Arg_sense	CCTTTGGCCATATCATCCGCAACTCTATGCACAAGTACC	67
His190Pro_anti	CGCGGCTGGTACTTGGGCATAGAGTTGAGGA	67
His190Pro_sense	TCCTCAACTCTATGCCCAAGTACCAGCCGCG	67
Val218Met_anti	AGGTCTCTGGGAACATGTGGGTGCAGAAAGC	67

Val218Met_sense	GCTTTCTGCACCCACATGTTCCCAGAGACCT	67
Ser226Tyr_anti	GGTAGGAGGTCACATAGATGAAGGAGGTCTCTGGG	67
Ser226Tyr_sense	CCCAGAGACCTCCTTCATCTATGTGACCTCCTACC	67
Ile235Ser_anti	CTCAATTTTCAGCTGGGTGCTCTTGTGATTCTGGTAGG	67
Ile235Ser_sense	CCTACCAGAATCACAAGAGCACCCAGCTGAAAATTGAG	67
Gly248Val_anti	CACTGCCCCGGAATACCTTGGCAAAAGGG	67
Gly248Val_sense	CCCTTTTGCCAAGGTATTCCGGGGCAGTG	67
Arg250Trp_anti	CTGTCATCACTGCCCCAGAATCCCTTGGCAAAAG	67
Arg250Trp_sense	CTTTTGCCAAGGGATTCTGGGGCAGTGATGACAG	67
Arg250Gln_anti	CTGTCATCACTGCCCTGGAATCCCTTGGCAA	67
Arg250Gln_sense	TTGCCAAGGGATTCCAGGGCAGTGATGACAG	67
Arg261Gln_anti	CTTTGCTCTGCAGTTGGGCCACACGCAGG	67
Arg261Gln_sense	CCTGCGTGTGGCCCAACTGCAGAGCAAAG	67
Ile270Ser_anti	CTCATGATGCTTTTGGAACTCACGGGGTATTCTTTGC	67
Gly295Ala_anti	CCTGGTGGGTGGCGAGCAGGGGG	67
Gly295Ala_sense	CCCCCTGCTCGCCACCCACCAGG	67
Ile270Ser_sense	GCAAAGAATACCCCGTGAGTTCCAAAAGCATCATGAG	67
Asp329Tyr_anti	TAGAAGAGGCTTGAGTACCTCTGGGTGGGAAAG	67
Asp329Tyr_sense	CTTTCCCACCCAGAGGTACTCAAGCCTCTTCTA	67
Asp341His_anti	CGGGTACCGTGTCGTCTTTTCAGGCAGTG	67
Asp341His_sense	CACTGCCTGAAAAGACGACACGGTACCCG	67
Arg352Leu_anti	GGGCTTCCAGATAGGATAGCTTGCAAGGTAAGTCC	67
Arg352Leu_sense	GGACTTACCTTGCAAGCTATCCTATCTGGAAGCCC	67
Ala357Val_anti	CCCACCGAAGAGGGGACTTCCAGATAGGATC	67

Ala357Val_sense	GATCCTATCTGGAAGTCCCCTCTTCGGTGGG	67
Arg368Cys_anti	GAGGGGGGGAACAGAAATAGTGATCCTCCC	67
Arg368Cys_sense	GGGAGGATCACTATTTCTGTTCCCCCCCTC	67
Tyr382Ser_anti	CACCTCACTGCAGGAGGAGGGGCTCAG	67
Tyr382Ser_sense	CTGAGCCCCTCCTCCTGCAGTGAGGTG	67
Ser395Pro_anti	CGGGCCCTGAACCTGGGTACATACATGCTTCTC	67
Ser395Pro_sense	GAGAAGCATGTATGTACCCAGGTTCAGGGCCCG	67
Glu400Lys_anti	CCCCGGCAATCTTGGGCCCTGAACC	67
Glu400Lys_sense	GGTTCAGGGCCCAAGATTGCCGGGG	67
Gly403Arg_anti	CCCCAGACACCCTGGCAATCTCGGG	67
Gly403Arg_sense	CCCGAGATTGCCAGGGTGTCTGGGG	67
Pro425Gln_anti	GCTATAGCTGGTGTACTGCGACACTGAAGTCC	67
Pro425Gln_sense	GGACTTCAGTGTCGCAGTACACCAGCTATAGC	67
Met451Val_anti	CCGCGGCATCACGGTGGTGGCGG	67
Met451Val_sense	CCGCCACCACCGTGATGCCGCGG	67
Asn475His_anti	GAGACTGGGAGAGCTGATCGTAGACACTAAAGTGGG	67
Asn475His_sense	CCCACTTTAGTGTCTACGATCAGCTCTCCCAGTCTC	67
Glu515Lys_anti	GGTTTGAGAGTAGAGAAACTTATTGGCAGCATTTAGATGTGGC	67
Glu515Lys_sense	GCCACATCTAAATGCTGCCAATAAGTTTCTCTACTCTCAAACC	67
Gln531Arg_anti	CCATTCCTGAATGGTACGGTAAGGAAGATTCTCGG	67
Gln531Arg_sense	CCGAGAATCTTCCTTACCGTACCATTCAGGAATGG	67
702mut_anti	GCAGTGGGGCAGTGGCTGTATCTTGTGATTCTGGTAGG	67
702mut_sense	CCTACCAGAATCACAAGATACAGCCACTGCCCCACTGC	67
792-1_anti	CGGGGTATTCTTTGGTGAAGGGGTGGGGA	67

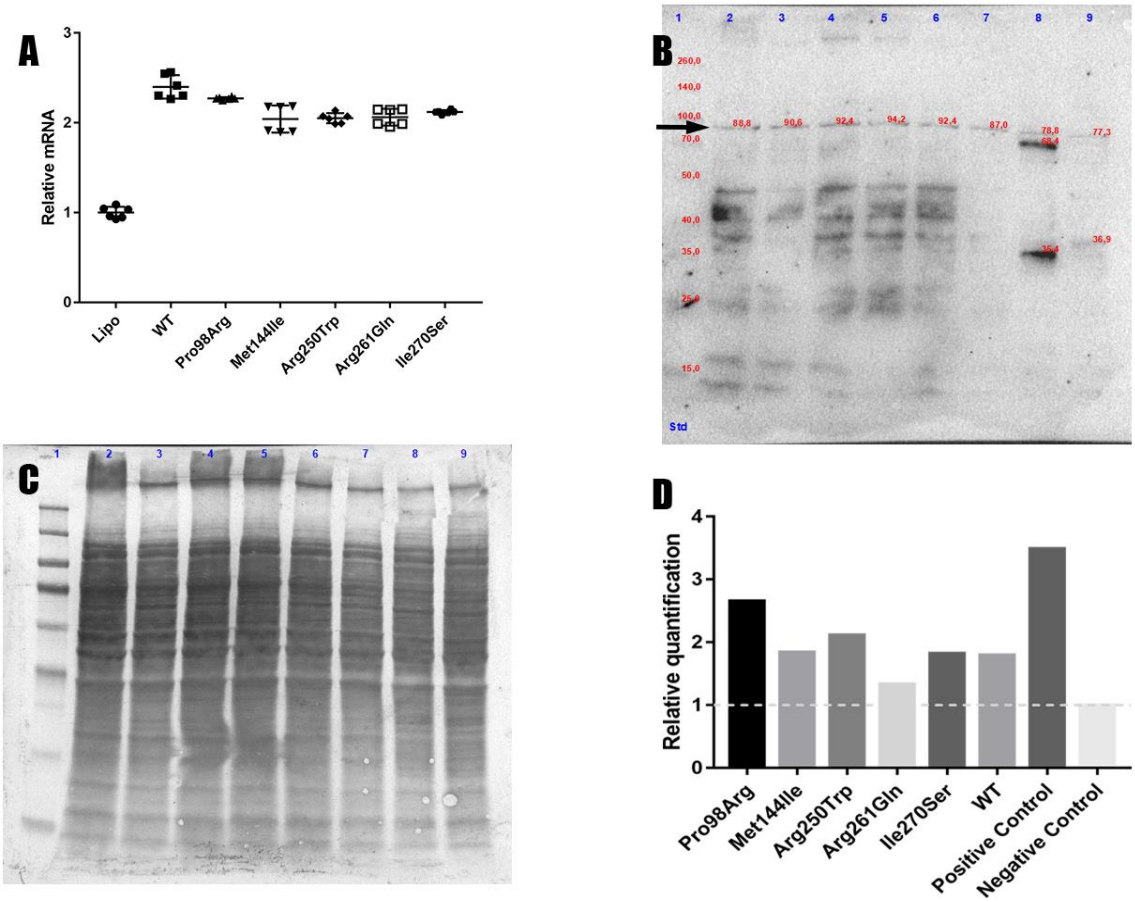
792-1_sense	TCCCCACCCCTTCACCAAAGAATACCCCG	67
1021+1GA_anti	GACCAGGAGAGCCCTATCTCGTCTTTTCAGGC	67
1021+1GA_sense	GCCTGAAAAGACGAGATAGGGCTCTCCTGGTC	67
Q5_Ser167del_F	CTTCCAGAAGCTGAAGCTGACAAAC	69
Q5_Ser167del_R	ACCAGCTGCCGCATCCAG	69
Q5_Lys172_Leu173del_F	GAAGCTGACAAACAACCACC	65
Q5_Lys172_Leu173del_R	TGGAAGGAGACCAGCTGC	65
Q5_Lys191_Tyr192del_F	CAGCCGCGGCTCCACATC	69
Q5_Lys191_Tyr192del_R	GTGCATAGAGTTGAGGATGATATGGC	69
Q5_Phe224del_F	ATCTCTGTGACCTCCTAC	60
Q5_Phe224del_R	GGAGGTCTCTGGGAACAC	60
Q5_Phe224_Ser229dup_F	cagagatgaAGGAGGTCACAGAGATGAAG	59
Q5_Phe224_Ser229dup_R	cagagatgaAGGAGGTCACAGAGATGAAG	59

eTable 3. qPCR primers.

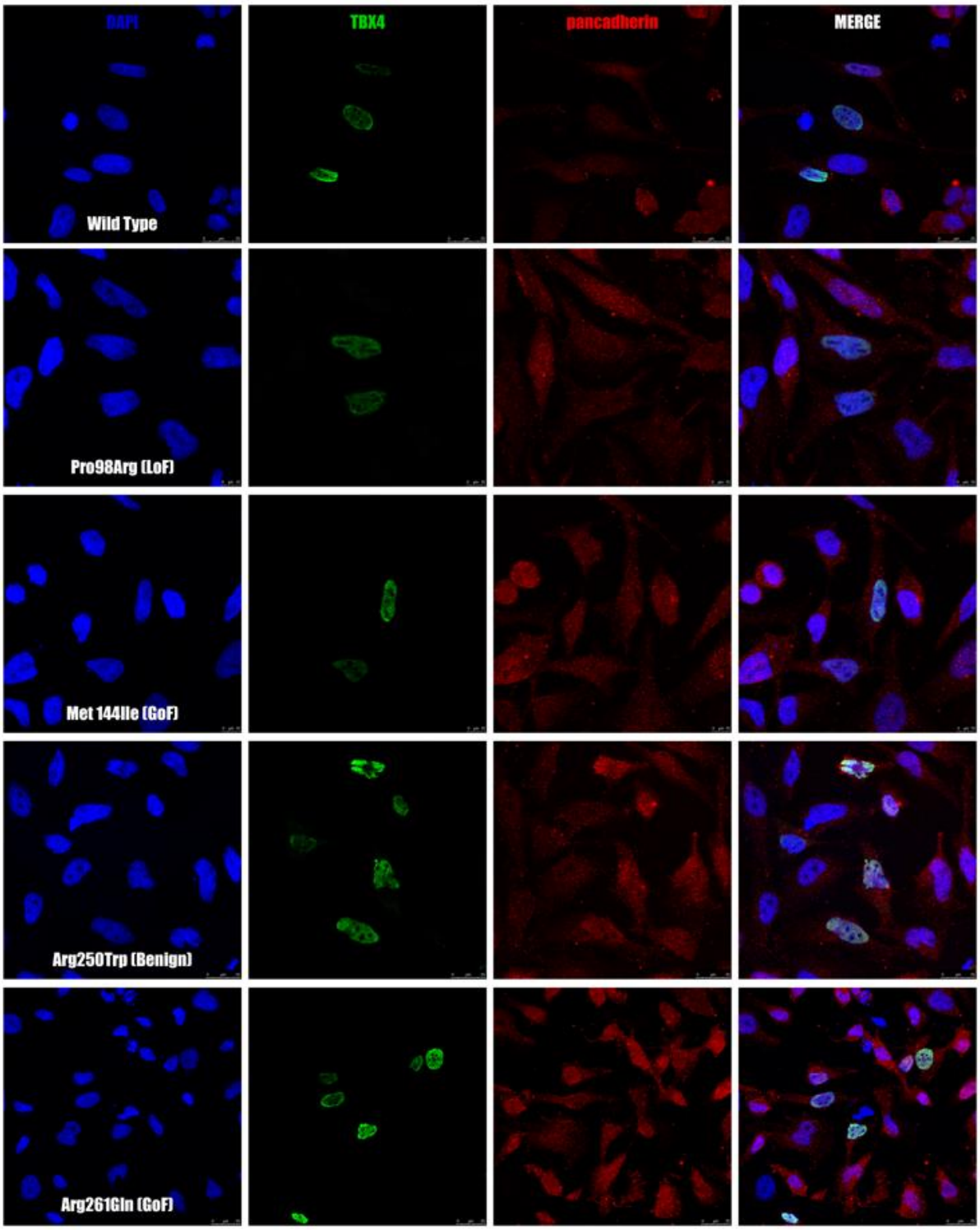
Primer ID	Sequence 5' - 3'	Annealing T (°C)
TBX4_qPCR_F	CTTTCCCACCCAGAGGGACT	60
TBX4_qPCR_R	GGGGCTTCCAGATAGGATCG	60
ALAS_qPCR_F	AGTGTGAAAACCGATGGAGG	60
ALAS_qPCR_R	CGATCATACTGAAAAGTGGAAACAG	60
YWHAZ_qPCR_F	ATGCAACCAACACATCCTATC	60
YWHAZ_qPCR_R	GCATTATTAGCGTGCTGTCTT	60

II. Variant assessment

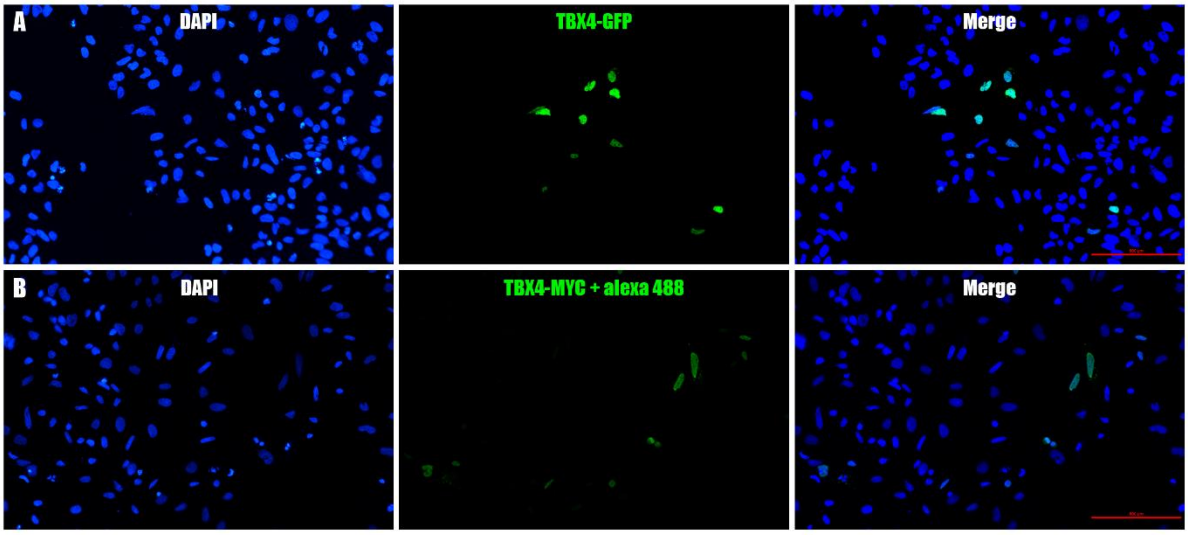
A. Functional



eFigure 1. Validation of the overexpression levels of TBX4 at mRNA and protein levels. A) qPCR results after transfection with the TBX4 constructs. *TBX4* mRNA expression of the wild-type and variants are 2-fold higher than the lipofectamine control after normalization. Data are mean ± standard deviation. B) FlagHRP western blot of HeLa cells after transfection under the same conditions as for the qPCR. All the variants show a band of 80-90kDa corresponding to TBX4 after double sumoylation. A 70 kDa band in the positive control may be TBX4 after with single sumoylation. C) Coomassie staining of the PVDF membrane used in the western blot. All the samples were loaded in similar levels. D) Quantification of the western blot using Coomassie staining as a loading control, each bar represents the quantification for the lane above it. The variant p.Pro98Arg (LoF) showed the highest level and p.Arg261Gln (GoF) the lowest with just 1.3 fold. Data are presented as band intensity/lane protein intensity.



eFigure 2. *TBX4* variants co-localize to the nuclei independently of their activity. A) Immunofluorescence for the different *TBX4* variants marked with: nuclei (blue/DAPI), anti-TBX4 (green/alexa488) and anti-pan-cadherin (red/alexa594). All the variants colocalize with DAPI (n = 2).

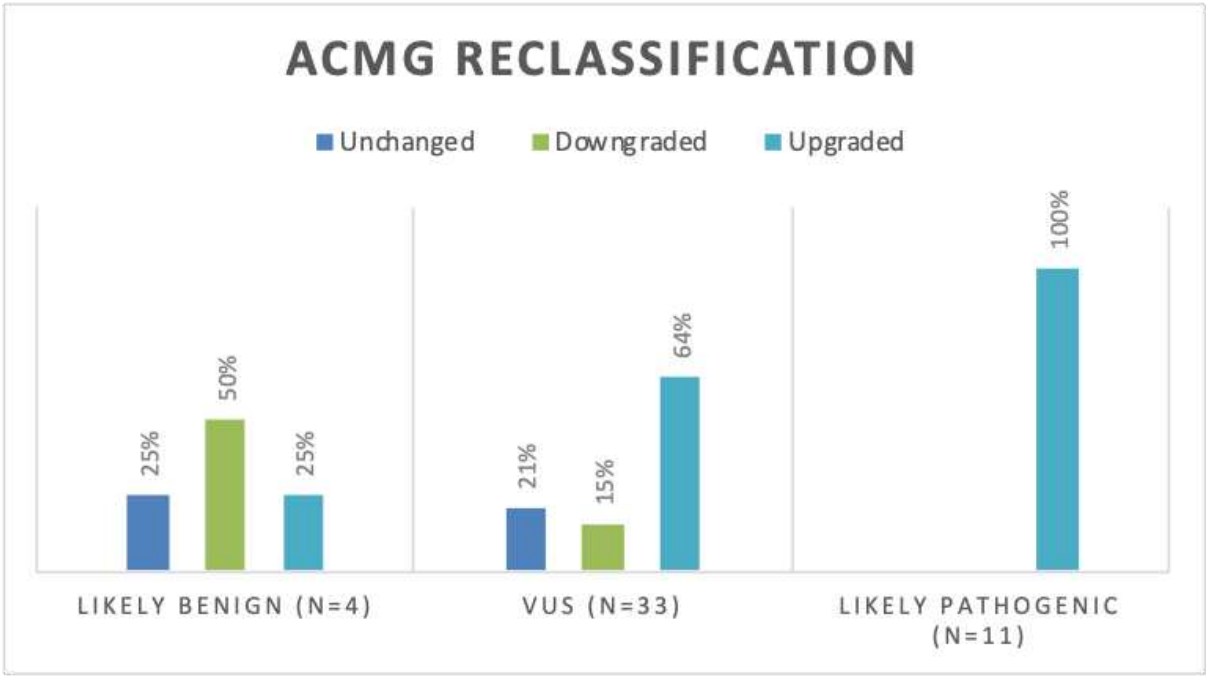


eFigure 3. TBX4 antibody validation for immunofluorescence. Cells were transfected with TBX4-GFP (A) and TBX4-MYC-HA (B). We imaged the cells in a fluorescence microscope to confirm that we had the same pattern between GFP and the antibody.

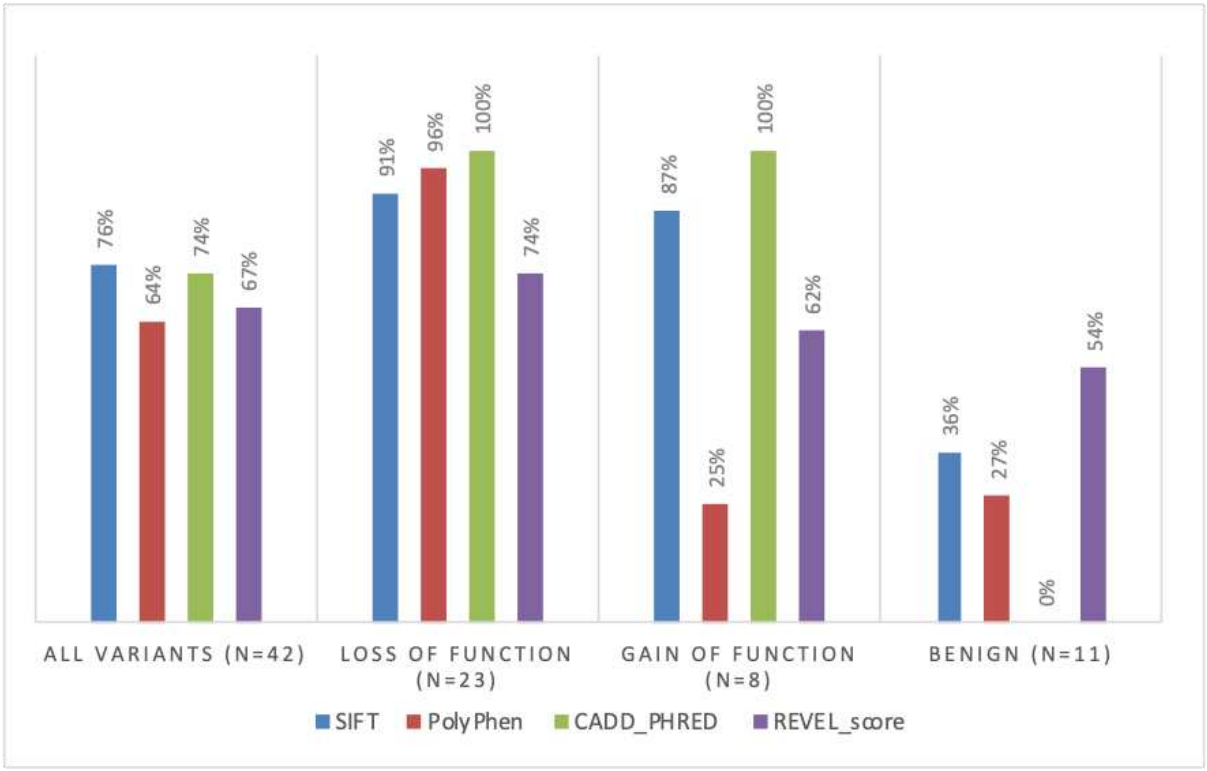
Minigene assay

Using hybrid minigenes encoding exon 5, we confirmed that c.702+1G>A induces the skipping of exon 5. The absence of this exon does not induce a frameshift, so the resulting transcript could yield a protein of 494 amino acids lacking 51 amino acids in the T-BOX domain, which taking into account our indel data should make the protein non-functional. The minigenes including exons 6 and 7 showed that the variant c.792-1G>C does not affect the correct processing of these *TBX4* exons, while the recurrent variant c.1021+1G>A induces a double exon skipping event where both exon 6 and 7 are skipped in the mutated constructs. No frameshift is induced in this double skipping, but the resulting protein would have a reduced length of 271 amino acids if not degraded.

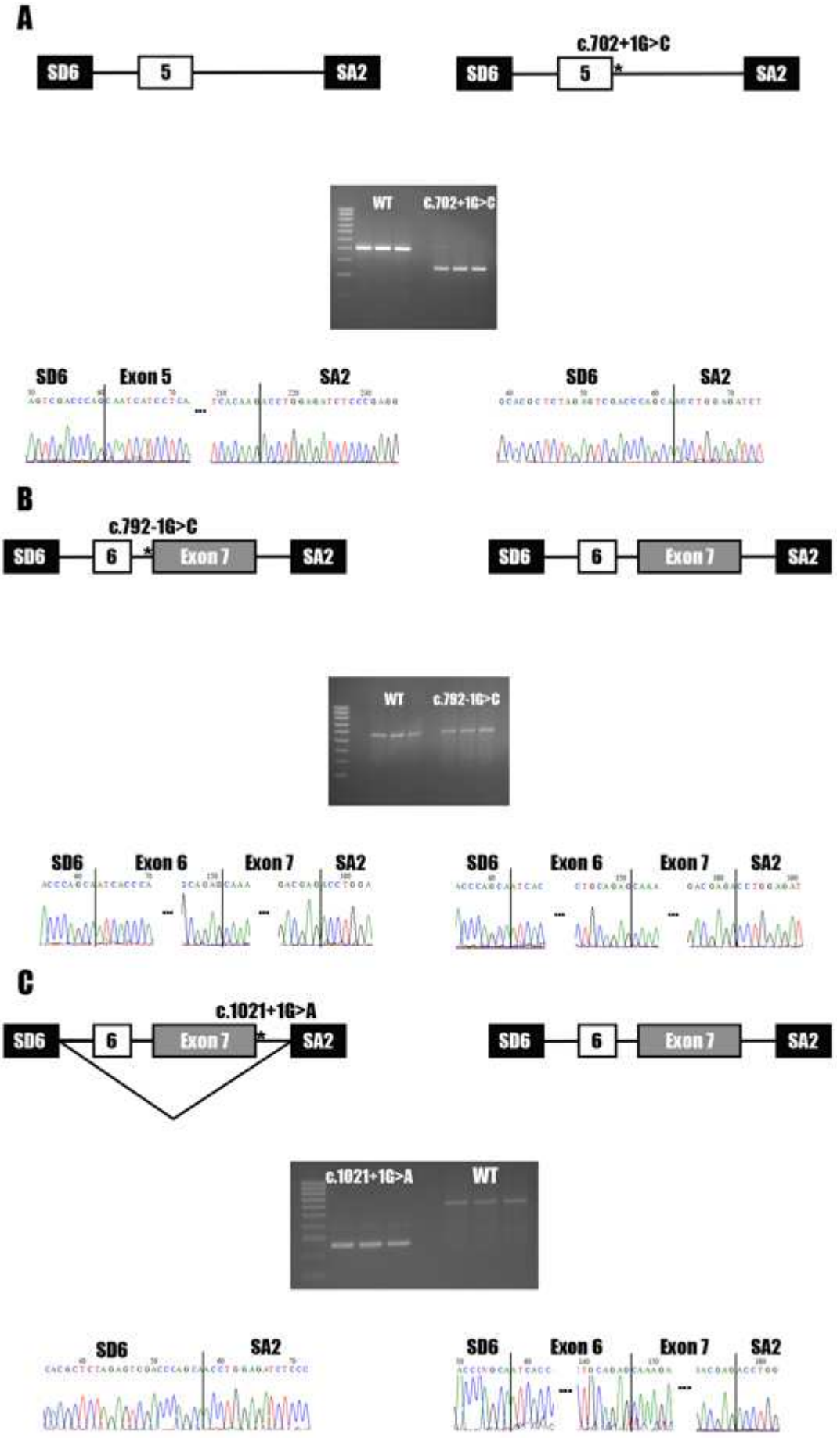
B. *In silico*



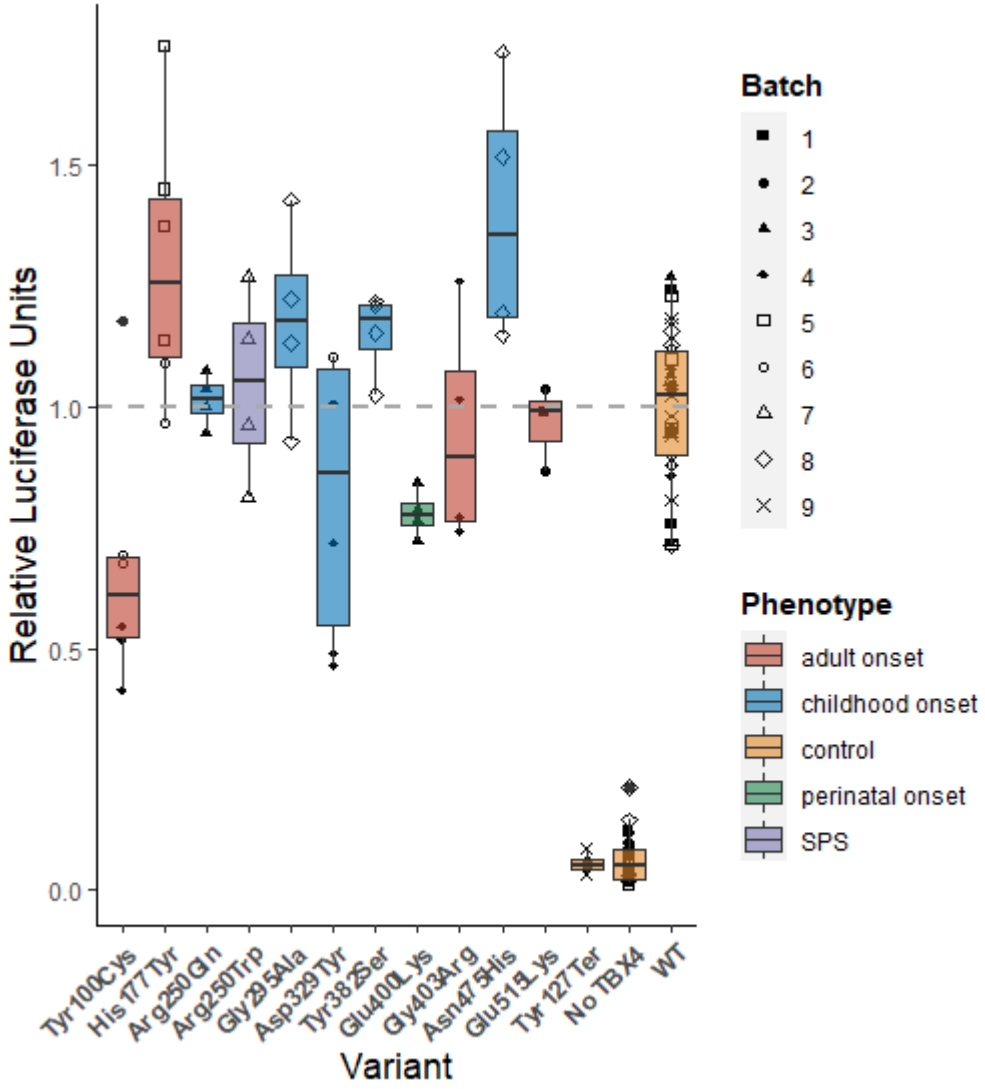
eFigure 4. Representation of the percentage number of variants with altered classification following functional assessment by luciferase assay. As per guidelines issued by the American College of Medical Genetics and Genomics (ACMG), interpretation of *TBX4* variants was amended following the application of the functional evidence criterion. Abbreviations: VUS, variant of uncertain significance.



eFigure 5. Comparison of the accuracy of different *in silico* prediction tools. Percentage number of functionally assessed *TBX4* variants (by luciferase assay) where pathogenicity was correctly predicted by SIFT (Sorting Intolerant From Toleran) (9), PolyPhen (Polymorphism Phenotyping) (10), CADD (Combined Annotation Dependent Depletion) (11), and REVEL (Rare Exome Variant Ensemble Learner) (12). Overall, the predictors classified correctly similar percentages of variants. When interpreting gain-of-function variants, the CADD score was the most reliable package with noted inconsistencies across different tools. For evidence supportive of pathogenicity, we applied the following criteria: SIFT ≠ tolerated; PolyPhen ≠ benign; CADD $\geq$ 15; REVEL $\geq$ 0.5.

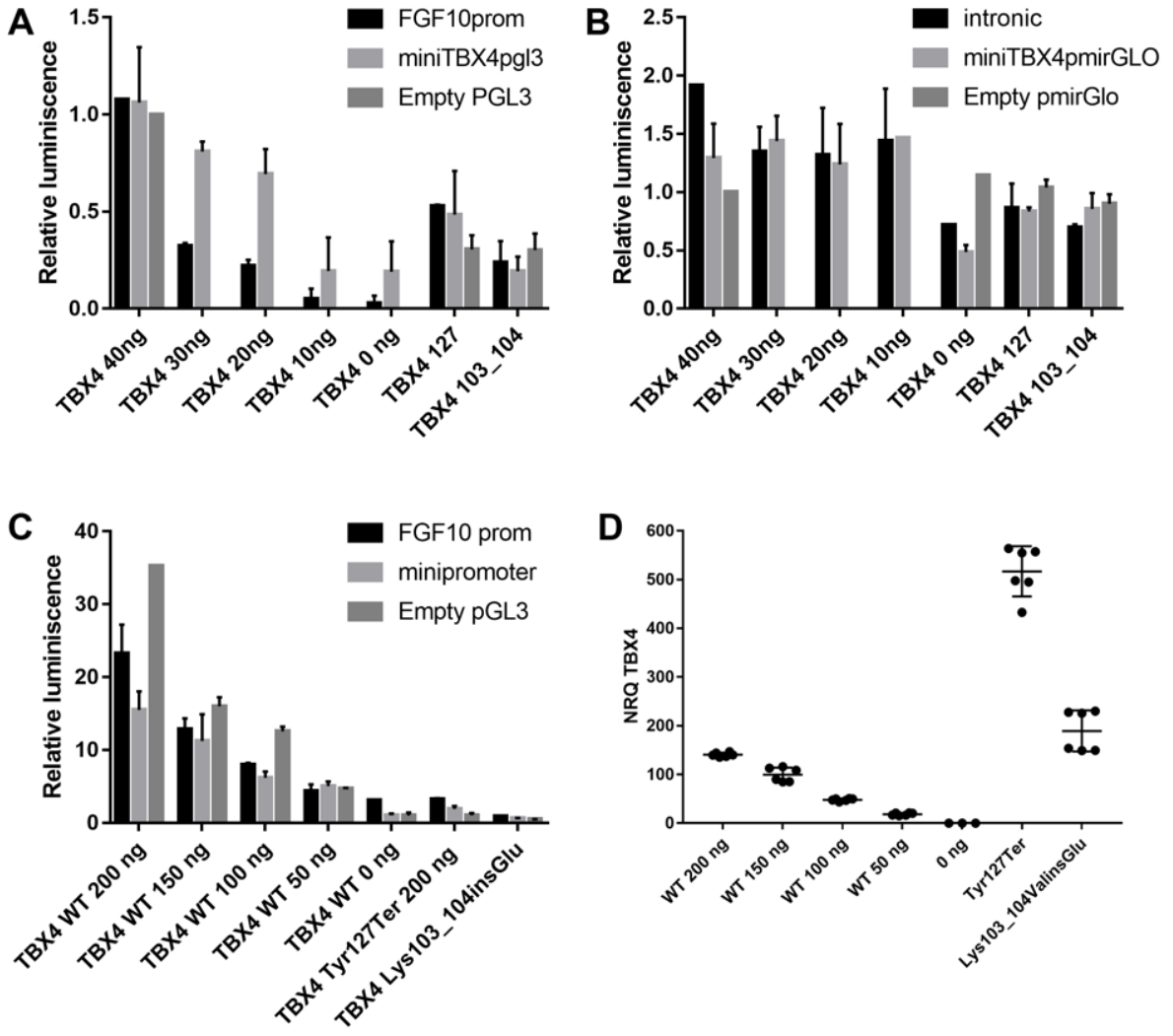

A
SD6
5
SA2
c.702+1G>C
WT
c.702+1G>C
SD6
Exon 5
SA2
SD6
SA2
B
c.792-1G>C
SD6
6
Exon 7
SA2
SD6
6
Exon 7
SA2
WT
c.792-1G>C
SD6
Exon 6
Exon 7
SA2
SD6
Exon 6
Exon 7
SA2
C
c.1021+1G>A
SD6
6
Exon 7
SA2
SD6
6
Exon 7
SA2
c.1021+1G>A
WT
SD6
SA2
SD6
Exon 6
Exon 7
SA2

eFigure 6. Minigene assay of *TBX4* splice-site variants. A) c.702+1G>C induces the skipping of exon 5. B) c.792-1G>C appears benign as splicing was identical to the wild-type. C) c.1021+1G>A induces a double exon skipping, as the minigene lost the exons 6 and 7 that were encoded.



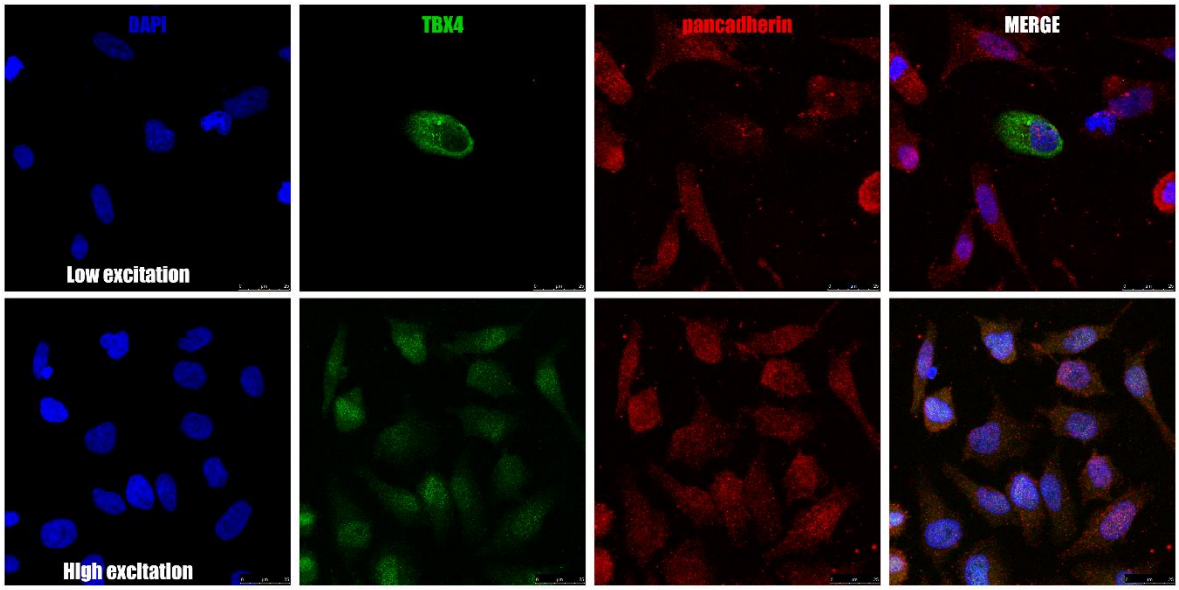
eFigure 7. Results of the luciferase assay for *TBX4* variants annotated as benign. We calculated the Firefly:Renilla ratio for each sample and normalized the data using the WT treatment, which was designated as 1. Data are shown as median ± interquartile ranges for at least 4 different replicate experiments. None of these

variants were statistically significant when compared with WT after One-Way ANOVA with a Bonferroni post-hoc correction.



eFigure 8. Optimization of the luciferase assay. A) Titration of TBX4 levels in the different constructs in a 96-well plate (FGF10 promoter, mini TBOX and empty pGL3), p.Tyr127Ter and p.Lys103_Val104insGlu were used as pathogenicity controls, 40 ng of each luciferase construct were used for all the conditions. The luciferase activities of each construct increased proportionally to the transfected amount of *TBX4*. B) Titration of TBX4 levels of the different constructs in a 96-well plate (FGF10 intronic island, mini TBOX and empty pGL3), p.Tyr127Ter and

p.Lys103_Val104insGlu were used as pathogenicity controls, 40 ng of each luciferase construct were used for all the conditions. None of the constructs responded to TBX4 in a stable dose-dependent manner. C) Titration of TBX4 levels in a 24-well plate, 200 ng of each luciferase construct were used for all the conditions. D) *TBX4* mRNA levels quantified by qPCR. *TBX4* mRNA levels increased proportionally as the ng used for the transfection, 200 ng of the two pathogenic variants were used. All the data are shown as mean ± SD (n = 2).



eFigure 9. Loss of immunoreactivity of the p.Ile270Ser *TBX4* variant. When we used low excitation settings we were able to detect small apoptotic bodies, while with high excitation we detected only autofluorescence (n=2).

III. Clinical

Radiological substudy

We obtained computed tomography (CT) images of the chest for 13 TBX4 cases recruited to the National Institute for Health Research BioResource–Rare Diseases (NBR) study; all scans were performed at the time of PAH diagnosis. Five of the above individuals were carriers of truncating (frameshift and nonsense) *TBX4* variants. Out of the remaining 8 cases with missense variants, only 3 were classed as likely pathogenic/pathogenic (taking into account our functional work) with the remaining 5 excluded from the radiological substudy (total TBX4 case n = 8). Findings were compared to CT diagnostic scans from PAH cases with *BMPR2* mutations (n = 34) and no mutations (n = 143) analyzed on the open-source software 4 Horos (Annapolis, MD USA) by the same cardiothoracic radiologist (AS), blinded to the underlying diagnosis, smoking status, and genotype. Radiological features were scored semi-quantitatively using a customized proforma (eTable 4).

eTable 4. Reporting proforma for analysis of radiological features.

Abbreviations: CTPA - Computerized Tomography Pulmonary Angiogram, HRCT - High-Resolution Computerized Tomography, GGO - ground-glass opacities, PVOD - Pulmonary veno-occlusive disease.

Parameter	**Response**
ID	character
Reader	character
CT scan date	date
Slice thickness	numeric
Number of slices	numeric

CTPA	done/not done
HRCT	done/not done
Expiratory CT	done/not done
Pleural effusion	Nil;Trace;Mild;Moderate;Severe
Subcutaneous oedema	Present, absent
Severity of GGO centrilobular pattern	None;Subtle;Present
Severity of GGO non-specific mosaic pattern	None;Subtle;Present
Distribution of GGO	C-central; U-upper; Z-zonal; D-diffuse
Pulmonary arteriovenous malformations	Present, absent
Largest bronchial artery size	Numeric (mm)
Mediastinal venous collaterals	Present, absent
Intralobular septal thickening	None;Subtle;Present
Mediastinal lymphadenopathy	None;Subtle;Present
Emphysema	None;Subtle;Present
Bronchial wall thickening	None;Subtle;Present
Fibrosis	None;Subtle;Present
Air trapping	None;Subtle;Present
Suspected PVOD	Yes, no

Histopathology

Lung histology was available for 17 previously published *TBX4* cases (supplementary data, xlsx). Alongside varying degrees of pulmonary vascular hypertensive changes, parenchymal abnormalities were present in all individuals. Alveolar features ranged from thin septa to diffuse dysplasia representing growth arrest at different stages of lung development. On the severe end of the spectrum,

out of 4 cases with lethal lung maldevelopment, 3 had mutations in the T-BOX domain including 2 loss-of-function missense variants affecting the same amino-acid position (p.Glu86Gln and p.Glu86Lys) and 1 protein-truncating variant (c.524_527del). The fourth case carried a paternally inherited *TBX4* missense variant at the C-terminus (c.1198G>A) shown to be benign by our functional analyses; this was an infant who died at 4 months due to progressively worsening PAH with respiratory failure. As reported by the authors, trio exome sequencing (proband and parents) also showed a *de novo* nonsense variant in the *CTNNB1* gene which may account for some of the observed phenotypic features, including microcephaly and muscle spasticity.

Interstitial fibrotic changes were described in 9 individuals including 2 cases with non-specific interstitial pneumonia (NSIP). All of the above were heterozygous for either missense loss-of-function or protein-truncating *TBX4* variants the majority of which (6/9) were located in the T-BOX domain, including the 2 NSIP cases. Bronchial abnormalities of varying degrees were reported in 8/17 individuals. A single patient with histological findings of pulmonary veno-occlusive disease (PVOD) harbored a missense variant (c.1145A>C) shown to be benign by our functional analyses. Another individual was heterozygous for a pathogenic missense variant (c.432G>T) inducing gain-of-function with reported typical findings of PAH in the explanted lung tissue alongside emphysematous changes.

IV. Supplemental References

1. Stauffer W, Sheng H, Lim HN. EzColocalization: An ImageJ plugin for visualizing and measuring colocalization in cells and organisms. *Scientific Reports*

2018;8:15764.

2. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine* 2015;17:405–424.
3. Whiffin N, Minikel E, Walsh R, O'Donnell-Luria AH, Karczewski K, Ing AY, Barton PJR, Funke B, Cook SA, MacArthur D, Ware JS. Using high-resolution variant frequencies to empower clinical genome interpretation. *Genet Med* 2017;19:1151–1158.
4. Eyries M, Montani D, Nadaud S, Girerd B, Levy M, Bourdin A, Trésorier R, Chaouat A, Cottin V, Sanfiorenzo C, Prevot G, Reynaud-Gaubert M, Dromer C, Houeijeh A, Nguyen K, Coulet F, Bonnet D, Humbert M, Soubrier F. Widening the landscape of heritable pulmonary hypertension mutations in paediatric and adult cases. *Eur Respir J* 2019;53:.
5. Zhu N, Gonzaga-Jauregui C, Welch CL, Ma L, Qi H, King AK, Krishnan U, Rosenzweig EB, Ivy DD, Austin ED, Hamid R, Nichols WC, Pauciulo MW, Lutz KA, Sawle A, Reid JG, Overton JD, Baras A, Dewey F, Shen Y, Chung WK. Exome Sequencing in Children With Pulmonary Arterial Hypertension Demonstrates Differences Compared With Adults. *Circ Genom Precis Med* 2018;11:e001887.
6. Gräf S, Haimel M, Bleda M, Hadinnapola C, Southgate L, Li W, Hodgson J, Liu B, Salmon RM, Southwood M, Machado RD, Martin JM, Treacy CM, Yates K, Daugherty LC, Shamardina O, Whitehorn D, Holden S, Aldred M, Bogaard HJ, Church C, Coghlan G, Condliffe R, Corris PA, Danesino C, Eyries M, Gall H, Ghio

S, Ghofrani H-AA, *et al.* Identification of rare sequence variation underlying heritable pulmonary arterial hypertension. *Nature Communications* 2018;9:1416.

7. Thoré P, Girerd B, Jaïs X, Savale L, Ghigna M-R, Eyries M, Levy M, Ovaert C, Servettaz A, Guillaumot A, Dauphin C, Chabanne C, Boiffard E, Cottin V, Perros F, Simonneau G, Sitbon O, Soubrier F, Bonnet D, Remy-Jardin M, Chaouat A, Humbert M, Montani D. Phenotype and outcome of pulmonary arterial hypertension patients carrying a TBX4 mutation. *Eur Respir J* 2020;55:1902340.
8. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*, 2nd ed. 2016. Cham: Springer International Publishing : Imprint: Springer; 2016. doi:10.1007/978-3-319-24277-4.
9. Sim NL, Kumar P, Hu J, Henikoff S, Schneider G, Ng PC. SIFT web server: Predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research* 2012;40:W452–W457.
10. Adzhubei I, Jordan DM, Sunyaev SR. Predicting functional effect of human missense mutations using PolyPhen-2. *Current Protocols in Human Genetics* 2013;Chapter 7:Unit7.20.
11. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Research* 2018;doi:10.1093/nar/gky1016.
12. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, McDonnell SK, Baheti S, Musolf A, Li Q, Holzinger E, Karyadi D, Cannon-Albright LA, Teerlink CC, Stanford JL, Isaacs WB, Xu J, Cooney KA, Lange EM, Schleutker J, Carpten JD, Powell IJ, Cussenot O, Cancel-Tassin G, Giles GG, MacInnis RJ, Maier C, Hsieh C-L, Wiklund F, Catalona WJ, Foulkes WD, *et al.* REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *American Journal of*

Human Genetics 2016;99:877.