

# **Understanding molecular mechanisms and predicting phenotypic effects of pathogenic tubulin mutations**

Thomas Attard<sup>1</sup>, Julie P.I. Welburn<sup>1\*</sup>, Joseph Marsh<sup>2\*</sup>

<sup>1</sup>Wellcome Trust Centre for Cell Biology, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3BF, Scotland, UK

<sup>2</sup>MRC Human Genetics Unit, Institute of Genetics & Cancer, University of Edinburgh, Edinburgh EH4 2XU, UK

\*Corresponding authors: [julie.welburn@ed.ac.uk](mailto:julie.welburn@ed.ac.uk), [joseph.marsh@ed.ac.uk](mailto:joseph.marsh@ed.ac.uk)

## Abstract

Cells rely heavily on microtubules for several processes, including cell division and molecular trafficking. Mutations in the different tubulin- $\alpha$  and - $\beta$  proteins that comprise microtubules have been associated with various diseases and are often dominant, sporadic and congenital. While the earliest reported tubulin mutations affect neurodevelopment, mutations are also associated with other disorders such as bleeding disorders and infertility. We performed a systematic survey of tubulin mutations across all isotypes in order to improve our understanding of how they cause disease, and increase our ability to predict their phenotypic effects. Both protein structural analyses and computational variant effect predictors were very limited in their utility for differentiating between pathogenic and benign mutations. This was even worse for those genes associated with non-neurodevelopmental disorders. We selected tubulin- $\alpha$  and - $\beta$  disease mutations that were most poorly predicted for experimental characterisation. These mutants co-localise to the mitotic spindle in HeLa cells, suggesting they may exert dominant-negative effects by altering microtubule properties. Our results show that tubulin mutations represent a blind spot for current computational approaches, being much more poorly predicted than mutations in most human disease genes. We suggest that this is likely due to their strong association with dominant-negative and gain-of-function mechanisms.

## Introduction

Microtubules are polarised cytoskeletal filaments essential in several cellular processes, ranging from cell division to signalling and transport. They assemble into axons, cilia, and the mitotic spindle, while also providing tracks for microtubule-associated proteins (MAPs) and motors for molecular trafficking (1, 2).

Microtubules self-assemble from tubulin- $\alpha$  and - $\beta$  heterodimers, with the dynamics of their assembly and disassembly being integral to their function (3). Tubulin- $\alpha$  and - $\beta$  are ubiquitous in eukaryotes, while related proteins in the FtsZ family show similarity in sequence, structure and function in archaea and bacteria (4, 5). Nine tubulin- $\alpha$  and ten tubulin- $\beta$  genes have been identified in humans, originating from evolutionary gene duplication events (6). In the tubulin field, these tubulin paralogues are referred to as isotypes. Between tubulin- $\alpha$  and - $\beta$ , conservation in sequence and structure is high, especially at interfaces stabilising the heterodimer, and at contacts between tubulin heterodimers across (lateral) and along (longitudinal) protofilaments (7). While tubulin- $\alpha$  and - $\beta$  both bind to GTP, only tubulin- $\beta$  can hydrolyse it to GDP, with residues on these binding sites amongst the most conserved (8). GTP hydrolysis enables distinct conformations that mediate the dimer's ability to be incorporated into microtubules (9). The C-terminal region makes up the outer surface of the microtubule and so contributes to most of the interactions with MAPs (10). Furthermore, many differences in amino acid sequences between isotypes occur in this region (11) and in the unstructured, highly negative tail (12). Tubulin- $\gamma$ ,  $\delta$  and  $\epsilon$  are more divergent in sequence than tubulin- $\alpha$  and - $\beta$  and are involved in the basal bodies of centrioles, rather than being self-assembled into dynamic polymers (13, 14).

A wide range of genetic disorders – called 'tubulinopathies' – have now been attributed to tubulin mutations. Over 225 pathogenic mutations in human tubulin isotypes have been reported (see Table 1). These findings highlight the importance of understanding tubulin function in different cell types. The first reported tubulin mutations associated with pathogenic phenotypes were found in TUB $\alpha$ 1A, TUB $\beta$ 2B, TUB $\beta$ 3, and TUB $\beta$ 4A and resulted in neurodevelopmental defects (15). Mutations in the  $\gamma$ -tubulin isotype TUB $\gamma$ 1 – which is necessary for microtubule nucleation – also cause a similar neurodevelopmental disorder (16, 17). While mutations in TUB $\alpha$ 4A and TUB $\beta$ 4A have been linked to neurodegenerative disease (18, 19), phenotypes outside the nervous system are now emerging. These include TUB $\beta$ 1 mutations associated with bleeding disorders (20), a link between tubulin- $\alpha$  acetylation and reduced sperm motility (21), and TUB $\beta$ 8 mutations connected with female infertility due to incorrect meiotic spindle assembly (22).

Despite the large number of identified pathogenic tubulin mutations, our understanding of the molecular mechanisms by which these mutations cause disease remains limited. The large majority of pathogenic tubulin mutations involve missense changes (*i.e.* single amino acid residue substitutions) and have autosomal dominant inheritance. There are only a few known exceptions, including homozygous null and in-frame deletion mutations in TUB $\beta$ 8 that can cause an oocyte maturation defect (23), a dominant nonsense mutation that results in a slightly truncated TUB $\alpha$ 4A linked to the neurodegenerative disease amyotrophic lateral sclerosis (19), and a recessive intron deletion in TUB $\alpha$ 8 in polymicrogyria patients that interferes with splicing, producing shorter mRNAs that do not contain exon 2 (24).

The absence of any known protein null mutations in tubulins causing dominant disease is striking. This tells us that the molecular mechanism underlying dominant mutations cannot simply be haploinsufficiency, whereby disease is caused by a complete lack of functional protein produced from the mutant allele. One possibility is that disease is caused by a milder loss of function (*i.e.* hypomorphic mutations). For example, pathogenic TUB $\alpha$ 1A mutations have been reported to disrupt interactions of the nascent protein with tubulin chaperones, which impairs heterodimer formation, suggesting that disease can be caused by a partial loss-of-function (25). Alternatively, a pathogenic mutation could act via a non-loss-of-function mechanism, having a dominant-negative effect or causing a gain of function (26). This would typically be associated with the mutant protein retaining the ability to incorporate into microtubules, which has been observed for many pathogenic tubulin mutations (27-32). For dominant-negative mutations, the incorporation of mutant protein directly or indirectly disrupts the activity of the wild-type protein (26). In these cases, the mutant tubulin retains its ability to form a heterodimer and assemble into microtubules before consequently impacting function in some other way, *e.g.* by perturbing microtubule properties or disrupting interactions with MAPs. For instance, TUB $\beta$ 3 mutations that alter the charged surface of the microtubule prevent molecular motors from binding and thus have profound impacts on cellular transport (33). Changes in microtubule properties could also induce a gain of function, as has been proposed for the pathogenic T178M variant in TUB $\beta$ 2A and TUB $\beta$ 3, which has been reported to make microtubules more stable and cause altered microtubule growth dynamics (34).

With the increasing accessibility of sequencing data, novel tubulin variants are being continually discovered (35). Since it is impractical to test them all experimentally, there is a strong need for computational approaches to identify tubulin mutations most likely to be pathogenic. Many variant effect predictors (VEPs) have been developed in recent years, and some of these are now in widespread use to help identify mutations that potentially have clinical significance (36). However, the performance of these predictors can vary quite

dramatically across different proteins and, to our knowledge, there has been no systematic assessment of their performance on tubulin mutations specifically. Missense mutations that cause pathogenicity by dominant-negative or gain-of-function mechanisms tend to be poorly predicted by most currently available VEPs (37), which could potentially limit their applicability to tubulins. It is therefore important for us to understand to what extent we can rely on computational predictors when assessing tubulin mutations.

In this study, we have first performed a systematic survey of known pathogenic missense mutations across all human tubulins and analysed their positions within the three-dimensional structures of tubulin heterodimers. This approach has allowed us to look for patterns in mutations across isotypes, structural locations, and phenotypes, in an attempt to obtain insight into the likely molecular disease mechanisms. Next, we assessed the performance of several different VEPs in distinguishing between pathogenic and putatively benign missense variants, observing that the predictive performance of all tested methods is poor compared to most other proteins. Finally, we have selected pathogenic tubulin mutations that were poorly predicted by computational approaches for experimental characterisation and show that the mutant proteins are able to be incorporated into microtubules, consistent with a likely dominant-negative mechanism. Our work suggests that many tubulin pathogenic mutations that act via non-loss-of-function mechanisms cause pathogenic phenotypes that cannot be explained computationally using current methods that rely on sequence conservation or protein structure. Overall, this study highlights the need for a greater understanding of microtubule-protein interactions to understand the molecular mechanisms underlying tubulinopathies.

## Results

### Survey of tubulin missense mutations

First, we compiled as many previously identified pathogenic or likely pathogenic dominant tubulin missense mutations as possible, using online databases (38, 39) and extensive literature searching (Table S1). In addition, we also identified missense variants in tubulin genes observed across >140,000 people from the gnomAD v2.1 database (40). Given that the gnomAD dataset comprises mostly healthy individuals without severe genetic disorders, these variants are unlikely to cause dominant disease, and we therefore refer to them as "putatively benign". However, we acknowledge that some of these variants could have milder effects, variable penetrance, or be associated with late-onset disease. Table 1 shows the numbers of pathogenic and gnomAD missense variants for each tubulin isotype and the associated type of genetic disease. Somatic mutations in tubulins are also implicated in cancer development; however, these mutations are likely to provide a selective advantage to

cancer cells by providing resistance to chemotherapeutic drugs (41-43). Hence these mutations might obscure our results and have not been included in our study, although we have noted two isotypes with links to cancer in which no other disease-related mutations have been identified yet (Table 1). While pathogenic mutations occur throughout the tubulins, mutations in both tubulin- $\alpha$  and - $\beta$  show clustering towards the intermediate and C-terminal domains when shown in the context of the linear amino acid sequence (Figure S1).

We also considered the gene-level missense constraint metric (MCM) scores provided by the gnomAD database (Table 1). These are derived from a model based on sequence context to predict the number of expected variants present in a healthy population relative to the number of actual variants observed (44). They provide a metric for the tolerance of each isotype to missense variation, with higher values representing genes that are more intolerant to amino acid sequence changes. Interestingly, we observed high MCM scores for isotypes linked with neurodevelopmental disorders (TUB $\alpha$ 1A, TUB $\beta$ 2A, TUB $\beta$ 2B, TUB $\beta$ 3, TUB $\beta$ 4A and TUB $\beta$ 5). TUB $\beta$ 6 is the only exception, and has only one pathogenic mutation reported so far, causing congenital facial palsy. These scores contrast with the much lower scores observed for TUB $\beta$ 1 and TUB $\beta$ 8, associated with platelet defects (45-49) and female infertility (22, 23, 50-53), respectively, which suggest that they are much more tolerant to sequence variation. Overall, our analysis indicates there is a stronger sequence constraint in the human population for tubulin isotypes that contribute to neurodevelopment. This may be due to the selective pressure of the process, compared to tubulin isotypes expressed in cells that affect organism fitness to a lesser extent.

### **Protein structural context is of limited utility for explaining tubulin mutation pathogenicity**

Next, we mapped all variants to three-dimensional structures of tubulins in the Protein Data Bank (PDB). We classified each mutation-carrying residue from every tubulin isotype based upon its structural location: whether it occurs at an intermolecular interface, on the protein surface, or buried in the protein interior. We used a hierarchy to classify structural locations (Figure S2). First, given that GTP binding (and hydrolysis for tubulin- $\beta$ ) is essential, we classified residues on the GTP nucleotide binding site as 'GTP binding'. Other tubulin mutations occurring at intermolecular interfaces were classified into three categories depending on where they occur in the model structures described in Table S2: at the intradimer interface, at an interdimeric interface (which we refer to as microtubule contacts, as they support lateral and longitudinal interactions with microtubules), or at interfaces with MAPs. Remaining mutations not at interface positions were then classified depending on whether they were on the protein interior or surface according to relative solvent accessible

surface area (54). Finally, we labelled any residues that were not present in any ordered part of any experimentally derived tubulin structures as 'outside of structure'. These typically occurred in the disordered C-terminal region, or at positions of divergence between the sequences of mapped structures and the tubulin isotype in question. Notably, while several gnomAD variants occurred at 'outside of structure' positions, this was true for only a single pathogenic mutation: R307H in TUB $\beta$ 1. Interestingly, all other tubulin- $\beta$  isotypes have a histidine at this position, explaining why the residue could not be mapped within our pipeline, as the wild-type amino acid residue needs to be present in the structure.

First, we used this classification system to assess whether tubulin mutations are enriched at particular locations. Despite the large dataset size, no significant differences in the locations of pathogenic or putatively benign gnomAD mutations were found across tubulin- $\alpha$  isotypes collectively (Figure 1A) or in any tubulin- $\alpha$  isotype individually (Figure 1B). This result was surprising given that the link between TUB $\alpha$ 1A and disease remains the most well-established of all tubulin isotypes, with 67 pathogenic variants linked with various neurodevelopmental defects (35). Observations were similar for TUB $\alpha$ 4A, the only other tubulin- $\alpha$  isotype with dominant pathogenic mutations (Figure 1B) (19). Furthermore, pathogenic mutations in both isotypes showed no obvious patterns or spatial clustering when visualised within the protein structure (Figure 2).

In contrast to tubulin- $\alpha$ , we do observe a significant enrichment of pathogenic tubulin- $\beta$  mutations relative to gnomAD variants at GTP binding residues and at lateral or longitudinal (*i.e.* interdimeric) microtubule contacts (Figure 1A), especially in TUB $\beta$ 2B and TUB $\beta$ 4A (Figure 1B). Mutations in these isotypes have been extensively linked to neuronal defects, and can be observed to form spatial clusters on tubulin structures (Figure 2). In contrast, putatively benign gnomAD variants are significantly more likely to occur on the protein surface or at 'outside of structure' positions (Figure 1B). TUB $\beta$ 3 and TUB $\beta$ 5 are also linked to similar disease phenotypes, but do not show any notable differences in structural location (15, 28, 55-57).

Interestingly, we found no significant differences in structural location in tubulin- $\beta$  isotypes associated with diseases outside the neuronal system (Figure 1B), nor did pathogenic mutations show any clear patterns or clustering at the three-dimensional structure level (Figure 2). We did, however, note that many TUB $\beta$ 1 variants occurred at 'outside of structure' residues (Figure 1B). This is due primarily to the greater sequence divergence between TUB $\beta$ 1 and experimentally determined tubulin structures, as any positions where the TUB $\beta$ 1 sequence is different than the aligned positions within the available structures will remain unmapped using our approach.

The centrosome-localised TUB $\gamma$ 1 has also been associated with neurodevelopmental disease (16, 17). Using the crystal structure of TUB $\gamma$ 1 containing two molecules per asymmetric unit (PDB ID: 3CB2) (58), we observed that 4 out of 8 pathogenic mutations identified so far occur at interfaces responsible for lateral interactions between TUB $\gamma$ 1 monomers (Figures 1 and 2, labelled in magenta). We suspect these interactions to occur between adjacent TUB $\gamma$ 1 chains within  $\gamma$ -TURC (59). Their distribution was not significantly different to gnomAD variants, although we note the small sample size.

### **Most pathogenic tubulin mutations are not highly disruptive to protein structure**

Next, we considered the predicted structural perturbations of pathogenic and putatively benign gnomAD missense variants using FoldX (60). This outputs a  $\Delta\Delta G$  value, in units of kcal/mol, with positive values indicating that a mutation is likely to destabilise protein structure and negative values indicating predicted stabilisation. Previous work has shown that computationally predicted  $\Delta\Delta G$  values can sometimes show considerable utility for the identification of pathogenic missense mutations, and for understanding likely molecular disease mechanisms (61).

Interestingly, we observe no significant differences between the  $\Delta\Delta G$  values of pathogenic and gnomAD missense variants for tubulin- $\alpha$ , - $\beta$  or - $\gamma$  (Figure 3A). Of the individual isotypes, only TUB $\beta$ 2B shows significantly higher  $\Delta\Delta G$  values for pathogenic mutations ( $p = 0.01$ ), although this would not remain significant when accounting for multiple testing (Figure 3B). We initially used  $\Delta\Delta G$  values that only consider the structural impact of variants on the monomer alone, as they are more consistent between structures. However, we also observed very similar results using the full  $\Delta\Delta G$  values calculated using the entire complex, including intermolecular interactions, as well as when using absolute  $\Delta\Delta G$  values (Figure S3).

These results suggest that the structural destabilisation is not a primary molecular disease mechanism underlying pathogenic tubulin mutations, and that considering structural impact is not particularly useful for differentiating between pathogenic and benign tubulin variants. Notably, this aligns with recent work showing that the predicted effects on protein stability tend to be much milder in gain-of-function and dominant-negative mutations than for pathogenic mutations associated with a loss of function (37), supporting the idea that most pathogenic tubulin mutations are due to non-loss-of-function effects.



## **Variant effect predictors show poor performance in discrimination between pathogenic and putatively benign tubulin mutations**

Next, we assessed the abilities of 25 different VEPs to distinguish between pathogenic and putatively benign tubulin missense mutations. A complete set of predictions for every tubulin mutation from all VEPs is provided in Table S3. To compare the performance of different VEPs, we first used a metric of predictor performance, known as the receiver operating characteristic (ROC) area under the curve (AUC) generated from each VEP over the entire dataset of tubulin missense variants (Figure 4A). Overall, the VEPs performed very poorly. Most had overall AUCs below 0.6, with the top-performing predictor, REVEL, having an AUC of only 0.68. In contrast, a recent study using a very similar methodology found that many VEPs had overall AUCs above 0.8, e.g. REVEL had an AUC of 0.9 for haploinsufficient disease genes, 0.85 for genes associated with a gain-of-function, and 0.83 for genes associated with dominant-negative effects (37). Thus, even considering that VEPs tend to do worse for non-loss-of-function mutations, the performance we observe here for tubulin mutations is strikingly poor.

It is important to note that our analysis is likely to overstate the predictive power of some VEPs. Supervised machine learning approaches underpin most VEPs, and typically use datasets of known pathogenic and benign variants for training. Since some VEPs are likely to have been trained using some of the tubulin mutations in our evaluation, their performance has a strong possibility of being overstated. This problem is particularly acute for metapredictors, including the top-performing methods in our analysis, REVEL and M-CAP, which combine supervised learning with multiple other predictors as inputs. In contrast, predictors based upon unsupervised machine approaches and those utilising empirical calculations should be free from this bias. Therefore, given the performance of the unsupervised predictor DeepSequence, ranking third overall, we likely consider it to be the most reliable predictor of tubulin mutation pathogenicity, consistent with its top-ranking performance in a recent study (36). However, even DeepSequence only achieves an AUC of 0.63 for tubulin mutations here, compared to well over 0.8 for all disease-associated proteins tested in that study.

Next, we compared the AUCs calculated for individual isotypes across all VEPs, considering isotypes with at least 10 pathogenic mutations. We found that pathogenic mutations in the tubulin- $\beta$  isotypes TUB $\beta$ 2B, TUB $\beta$ 3 and TUB $\beta$ 4A, which are all associated with neurodevelopmental diseases, were predicted better compared to TUB $\beta$ 8, associated with oocyte maturation defects (Figure 4B). Therefore, we classified tubulins into two groups based upon observed disease phenotypes: neurodevelopmental and other (as classified in

Table 1). We observed a significantly better performance on isotypes linked to neurodevelopmental diseases (Figure 5A). Interestingly, when we compare the performance on neurodevelopmental vs other disorders across the individual VEPs (Figure 5B), we observe that metapredictors and supervised VEPs, like REVEL, M-CAP and VEST4, outperformed all other VEPs on isotypes linked to neurodevelopmental disease but showed a drastic decrease in performance on other phenotypes. In contrast, unsupervised DeepSequence shows very similar performance between the two groups. This strongly suggests that certain VEPs have likely been overfitted in their training against the neurodevelopmental mutations.

Given its overall performance in our analyses and unsupervised nature, we currently recommend DeepSequence for predicting the effects of tubulin mutations, although we emphasise that its predictive utility is still relatively limited. Therefore, we have produced DeepSequence predictions for every possible amino acid substitution across most tubulin isotypes and have provided them as a resource (Table S4). We have also calculated optimal thresholds for DeepSequence using the closest point to the top left corner of our ROC curves. Based upon this, we suggest that DeepSequence scores lower than -5.89 are likely to be pathogenic in isotypes linked with neurodevelopmental phenotypes, and lower than -4.83 for isotypes linked with other phenotypes.

### **The most poorly predicted tubulin pathogenic mutations likely act *via* non-loss-of-function mechanisms**

To further examine the contributing factors behind the limited performance of VEPs in detecting pathogenic tubulin mutations, we sought to identify the pathogenic mutations predicted most poorly by current approaches. To do this, we developed a ranking method to examine mutations across all the VEPs used in this study. First, all predictions of gnomAD and pathogenic tubulin mutations were amalgamated. We then transformed all scores from individual VEPs to be on the same scale for comparison. Next, for each VEP, cumulative distribution ranks were computed for normalisation within a window partition across the combined dataset. Finally, for each mutation, these normalised ranks were averaged across all VEPs to generate a new metric we termed mean cumulative distribution (MCD), provided in Table S5. Essentially the MCD provides a single value for each mutation representing how damaging it is predicted to be across all of the VEPs we used in our analysis, ranging from 0 for mildest to 1 for most disruptive. MCD replicated the results we observed for the VEP analysis, with the most significant differences between pathogenic and gnomAD values being observed in the same isotypes (TUB $\beta$ 2B, TUB $\beta$ 3, TUB $\beta$ 4A) that showed higher ROC values (Figure S4).

Next, we used the MCD scores to identify pathogenic tubulin mutations that were most poorly predicted by the VEPs (Table 2). These mutations generally involve substitutions between chemically similar amino acids. In particular, mutations from valine to isoleucine at position 353 in three different isotypes – TUB $\alpha$ 1A, TUB $\beta$ 5 and TUB $\beta$ 8 – are among the worst predicted, with the variant in TUB $\beta$ 8 having the lowest MCD score of any pathogenic tubulin mutation. Phenotypically, in TUB $\alpha$ 1A and TUB $\beta$ 5, V353I is related to the malformations of cortical development (15), while the mutation is associated with female infertility in TUB $\beta$ 8, although whether the mutation is causing the issue is not shown (50).

Consequently, we experimentally tested whether mutations in the 353 position for TUB $\alpha$ 1 and TUB $\beta$ 8 enabled tubulin incorporation into microtubules. As V353I introduces a mild physicochemical change, as with most of the poorly predicted pathogenic tubulin mutations, we also tested a more disruptive mutation in both isotypes where the hydrophobic valine is changed to a positively charged arginine. V353R has to our knowledge not been observed in humans (nor is it possible with only a single nucleotide substitution), but it allows us to compare an apparently mild substitution to a more disruptive substitution at the same position. We transiently transfected the fluorescently tagged tubulin mutants into HeLa cells and analysed whether they were incorporated into the mitotic spindle, where microtubules can be clearly observed and where spindle architecture relies on MAPs. N-terminally tagged mCherry-TUB $\alpha$ 1 incorporated into microtubules with the aid of a linker; however mCherry-TUB $\beta$ 8 did not. We therefore tested a C-terminally tagged TUB $\beta$ 8-GFP, which did successfully incorporate.

Both wild-type TUB $\alpha$ 1 and TUB $\beta$ 8 isotypes strongly incorporated into the mitotic spindle in HeLa cells, co-localising with tubulin (Figures 6A-B and S5). The V353I mutants also showed incorporation that was not significantly different from wild type for both isotypes. This is also consistent with previous work showing that V353I in TUB $\beta$ 5 can still form a heterodimer and be incorporated into microtubules (28). Interestingly, however, the V353R mutant showed severe reduction in incorporation for TUB $\alpha$ 1, but no significant effect for TUB $\beta$ 8. We also tested whether the tubulin mutations affected mitotic spindle morphology and measured the mitotic spindle length (Figure 6C). Notably, while TUB $\alpha$ 1 V353I did not show reduced incorporation into microtubules, it did lead to significantly shorter spindles, suggesting the mutation may ultimately interfere with correct microtubule dynamics or interactions with microtubule-associated proteins. Other mutants showed no significant difference from wild type. We also confirmed that spindle aspect ratios were similar between wild-type and mutant tubulin constructs, indicating spindle scaling is conserved (Figure S6).

Our results clearly suggest a possible molecular mechanism underlying genetic disease in the TUB $\alpha$ 1A V353I mutation: the mutant can incorporate into microtubules along with wild-type protein, but appears to have a dominant-negative effect on spindle organisation of microtubules. In contrast, the V353R mutation, which we predict to be more severe at a protein structural level, shows reduced incorporation and thus has no effect on spindle morphology. This mirrors the recent observation that dominant-negative and gain-of-function mutations tend to be more structurally mild than loss-of-function mutations (37).

None of the tested TUB $\beta$ 8 mutations show any phenotypic effect. It is interesting to note that V353 residues in tubulin- $\alpha$  and - $\beta$  are not at precisely equivalent positions in a sequence alignment or in the structures of tubulin- $\alpha$  and - $\beta$  (Figure S7). V353 is a surface residue in tubulin- $\alpha$ , close to the longitudinal interdimeric interface, whereas in tubulin- $\beta$ , it is more buried in the protein interior, but close to the intradimer interface, and thus it is unsurprising that mutations at V353 could have different molecular effects in each isotype. It is possible that the molecular effects of V353 mutations in TUB $\beta$ 8 are too subtle to detect in our experimental system. Alternatively, the evidence that V353I is a true pathogenic mutation in TUB $\beta$ 8 is quite limited. It was observed in a heterozygous state in a single patient with infertility, but it is unknown if either of her parents also had the mutation (50). Thus, another possible explanation for the poor computational prediction of TUB $\beta$ 8 V353I is that it is not really pathogenic.

## Discussion

Tubulinopathies have previously been associated with defects in cortical brain development (15, 57, 62-65). However, several pathogenic tubulin mutations reported in recent years are not linked to neurodevelopmental diseases, and thus the phenotypical profile of 'tubulinopathies' needs redefining. As the mechanisms by which tubulin mutations give rise to these disease phenotypes are generally unknown, the extensive repertoire of newly identified mutations now at our disposal provides an opportunity to study pathogenicity globally, across the tubulin family.

The strong conservation in tubulin sequence and structure across isotypes (12), and the lethality of tubulin mutations in yeast (66) might lead to the assumption that many pathogenic mutations would cause protein instability and consequently a loss of function. However, this appears not to be the case, with predictions of changes in protein stability showing very little ability to discriminate between pathogenic and putatively benign mutations, with the exception of a slight trend observed for TUB $\beta$ 2B. There is some previous evidence for a loss-of-function mechanism underlying some pathogenic tubulin mutations, *e.g.* the R264C mutation in TUB $\alpha$ 1A perturbs chaperone binding and reduces the extent of heterodimer

assembly, which has been suggested to account for its pathogenic effects (25). However, it appears that the overall role of destabilisation in driving tubulin pathogenicity must be very limited.

Tubulin mutations can act via dominant-negative or gain-of-function mechanisms, where the mutant protein incorporates into the microtubule lattice and instead disrupts its function. These effects have been reported in tubulins before and include mutations in TUB $\alpha$ 1A (27) and others across different tubulin- $\beta$  isotypes (28-32). Once they are incorporated into microtubules, these mutations may alter biophysical microtubule properties or disrupt interactions with MAPs. The latter model includes interactions with non-motor MAPs, like NuMA and PRC1 (31, 32), and the impairment of motor protein binding and, subsequently, cellular transport (33). Multiple copies of mutant tubulin incorporated into microtubule can compound a mutation's effect. Also, some tissues express multiple tubulin- $\beta$  isotypes, especially in the developing brain (67). Here, redundancy may mitigate any loss-of-function in a specific isotype, but cells struggle to cope with dominant-negative or gain-of-function effects.

The classic definition of the dominant-negative effect involves the mutant protein disrupting the activity of the wild-type protein, either directly or indirectly, whereas the phenotypic effects of gain-of-function mutations are due to the mutant protein doing something different than the wild-type protein (26). For most proteins that form relatively small complexes, classification of a mutation as being associated with to either mechanism should usually be straightforward. However, given the large size of microtubule assemblies and their diverse functional roles, such classification may be ambiguous for tubulins. It depends on whether disease is due to reduced microtubule function (dominant negative), or whether altered microtubule properties are driving disease (gain of function). While here we have referred to TUB $\alpha$ 1 V353I, which incorporates into microtubules and reduces spindle length, as dominant negative, it is possible that its pathogenic effects might be better described as gain of function.

The pathogenic mutations predicted most poorly tend to involve substitutions between similar amino acids, with hydrophobicity and charges largely maintained between wild-type and mutant residues. This observation likely reflects the fact that most predictors incorporate some manner of amino acid substitution matrix, thus predicting more minor effects for variants that do not change a residue's charge or hydrophobic state. In contrast to more drastic mutations that tend to be better predicted, these mutations may be less detrimental to folding and microtubule assembly. Indeed, this difference is likely to explain why we observed a

weaker co-localisation to the mitotic spindle in the TUB $\alpha$ 1 V353R mutant compared to V353I.

In this study, the performance of supervised predictors, and especially metapredictors like REVEL and M-CAP, highlights the issue of circularity present in these types of VEPs, which are typically trained using published pathogenic and benign variants (68). There is a strong likelihood that some of these mutations are in training sets for these predictors. These biases would lead metapredictors and supervised methods to predict mutations in isotypes linked with neurodevelopmental diseases as pathogenic correctly, while mutations in other isotypes would be predicted poorly. This may explain why metapredictors performed so much worse on such mutations, especially when compared to unsupervised methods like DeepSequence and PROVEAN.

We judge DeepSequence to be the best performing VEP, considering consistency across isotypes and phenotypes. Unlike supervised methods, it does not use a labelled training set. DeepSequence can also make predictions based on the evolutionary conservation of entire sequences at once as opposed to no more than a few sites of interest, setting it apart from many other VEPs (64). Benchmarking several VEPs against deep mutational scanning data also highlighted DeepSequence as the best pathogenicity predictor (36). We, therefore, recommend using DeepSequence over metapredictors like REVEL and M-CAP to achieve more accurate predictions of tubulin variant pathogenicity.

This study presents the tubulin family as a blind spot for current phenotype predictors, especially in identifying mutations that result in non-neurodevelopmental pathogenic phenotypes. Inherent predictor biases due to training sets introduce circularity and could contribute to this. While effects other than loss of function appear crucial, the diversity of pathogenic phenotypes necessitates further study to comprehend the molecular mechanisms behind their pathogenicity fully. Understanding these mechanisms and what sets them apart from other proteins would be the key to improving the poor performance of current phenotype predictors in tubulin that we observed.

For non-neurodevelopmental phenotypes, pathogenic effects initially appeared to be isotype-specific, but this could be linked to where these isotypes are expressed and to what level, or other unknown factors. For example, pathogenic mutations in TUB $\beta$ 8 are associated to female infertility, but males carrying these mutations would presumably present a healthy phenotype (22). Importantly, many tubulin mutations are currently reported as linked to a disease or phenotype, but the evidence supporting causality is limited. It will be important to demonstrate whether the mutation is causative of pathogenesis or a consequence of other changes to cell homeostasis or genome stability. A pragmatic experimental approach is

therefore essential to define the mechanism of pathogenesis associated with tubulin mutations. It will then be possible to develop strategies to treat tubulin-related disorders using a personalised medicine approach.

## Methods

### Mutation datasets

A thorough literature search was first performed to identify as many pathogenic or likely pathogenic missense mutations across the tubulin family as possible. While the recently released tubulin mutation database (39) – as well as others such as OMIM (69) and ClinVar (38) – were excellent starting points, mutations from other publications (especially ones reported recently) were included (all detailed in Table 1). The literature search also uncovered a few other mutations that were incompletely dominant, homozygous, or of an unknown heritability, but these were omitted from the study. Any mutations described as *de novo* were assumed to be heterozygous and hence also dominant. As the underlying mechanisms with pathogenic mutations leading to cancer are likely to be more complex (and distinct), somatic mutations linked with such phenotypes were excluded. A set of putatively benign missense mutations for each isotype was retrieved from gnomAD v2.1 (40). We excluded mutations present in both pathogenic and gnomAD datasets and considered only isotypes with more than one pathogenic mutation.

### Protein structural analyses

Three-dimensional models of tubulin structures were taken from the protein data bank (PDB) on 2020.05.27, using the first biological assembly from each entry to represent its quaternary structure. We searched for polypeptide chains with at least 70% sequence identity to human tubulin isotype over a stretch of at least 50 amino acid residues. This allowed structural analysis using related structures for some isotypes for which structures are not available. Importantly, we only considered those residues where the amino acid in the structure, as well as both adjacent residues, are the same as in the sequence of the isotype of interest. Unless otherwise stated, all structures that satisfy this threshold were considered. A hierarchy was implemented to determine where a residue occurred on the tubulin structure concerning its function (Figure S2). First, we looked through all structures to highlight residues interacting with GTP (or its analogues). Then, model structures were used (Table S2) to determine whether the remaining residues occur on interfaces between the corresponding subunit in the tubulin dimer or with MAPs and other tubulin dimers in the microtubule in that specified order. Residues not present on any interface were mapped to the highest resolution structure within the dataset. These mappings were used to classify

residues according to their location, separating residues on the surface (relative solvent accessible surface area >0.25) from ones at the protein interior. Once every residue was labelled, their distribution was compared between mutations in our pathogenic and gnomAD datasets. Finally, all mutations in tubulin- $\alpha$  or - $\beta$  were considered before moving on to specific isotypes with at least 5 identified pathogenic missense mutations, also using PyMOL (70) to confirm that these patterns could also be observed qualitatively. To facilitate further analysis of tubulin mutations, we have provided a list of all amino acids in tubulin- $\alpha$  and - $\beta$  that are involved in protein-protein or protein-ligand interactions (Table S6).

Structural predictions were then obtained from FoldX (60) for pathogenic and gnomAD variants. Each mutation was mapped onto the highest resolution structure from the dataset above. Here, the  $\Delta\Delta G$  values were calculated considering the monomer only, as well as the whole biological assembly, in the same manner as previously described (37).

### **Computational variant effect predictions**

Scores from VEPs were generated using the same pipeline as described previously (36). Though most of the methods used failed to produce scores for the entire combined dataset, the percentage of missing scores from each predictor did not exceed 7.8%, except for S3D-PROF and SNPs&GO-3D, which required PDB structures as inputs. For S3D-PROF and SNPs&GO-3D, structures with the closest sequence identity were used, which have the following PDB IDs: 5iyz (TUB $\alpha$ 1A), 5lxt (TUB $\alpha$ 4A), 4tv9 (TUB $\beta$ 1), 5nm5 (TUB $\beta$ 2A and TUB $\beta$ 8), 4i4t (TUB $\beta$ 2B), 4lnu (TUB $\beta$ 3), 5jqg (TUB $\beta$ 4A and TUB $\beta$ 4B) and 4zol (TUB $\beta$ 5).

For plotting ROC curves, mutations from the pathogenic dataset were labelled as true positives, with gnomAD variants marked as true negatives. The plots themselves, as well as the AUC, were calculated using the PRROC package in R. Due to predictors PROVEAN, DeepSequence, SIFT4G and BLOSUM operating via inverse metrics (i.e. variants predicted to be more damaging have lower scores), all scores were multiplied by -1 before plotting curves and calculating AUC values. Finally, ROC AUC values were calculated globally across the whole dataset for each VEP first before generating specific scores of each VEP for individual isotypes.

To calculate the MCD rankings, the datasets containing prediction scores for gnomAD and pathogenic mutations were combined. Then, the cumulative distribution function was applied to the collective scores obtained from each predictor using the `cume_dist` function from the DPLYR package on a per-VEP basis. Each value represents the proportion of scores that are less than or equal to the one obtained for that particular mutation from a specific VEP. Then, the cumulative distribution scores from all VEPs were averaged for each mutation to generate the MCD score.





## Acknowledgements

We thank Benjamin Livesey for help with generating the VEP outputs. This work was supported by the Medical Research Council via a Precision Medicine Doctoral Training Programme studentship to TA and a Career Development Award (MR/M02122X/1) to JAM. JAM is a Lister Institute Research Fellow. JPIW is supported by a Wellcome Trust Senior Research Fellowship (207430).

## References

1. Brouhard GJ, Rice LM. Microtubule dynamics: an interplay of biochemistry and mechanics. *Nat Rev Mol Cell Biol.* 2018;19(7):451-63.
2. Janke C, Magiera MM. The tubulin code and its role in controlling microtubule properties and functions. *Nat Rev Mol Cell Biol.* 2020;21(6):307-26.
3. Mitchison T, Kirschner M. Dynamic Instability of Microtubule Growth. *Nature.* 1984;312(5991):237-42.
4. Battaje RR, Panda D. Lessons from bacterial homolog of tubulin, FtsZ for microtubule dynamics. *Endocr Relat Cancer.* 2017;24(9):T1-T21.
5. Dyer N. Tubulin and its prokaryotic homologue FtsZ: a structural and functional comparison. *Sci Prog.* 2009;92(Pt 2):113-37.
6. Findeisen P, Muhlhausen S, Dempewolf S, Hertzog J, Zietlow A, Carlomagno T, et al. Six subgroups and extensive recent duplications characterize the evolution of the eukaryotic tubulin protein family. *Genome Biol Evol.* 2014;6(9):2274-88.
7. Lowe J, Li H, Downing KH, Nogales E. Refined structure of alpha beta-tubulin at 3.5 Å resolution. *J Mol Biol.* 2001;313(5):1045-57.
8. Kobayashi T. Dephosphorylation of tubulin-bound guanosine triphosphate during microtubule assembly. *J Biochem.* 1975;77(6):1193-7.
9. Mitchison TJ. Localization of an Exchangeable Gtp-Binding Site at the Plus End of Microtubules. *Science.* 1993;261(5124):1044-7.
10. Wang Z, Sheetz MP. The C-terminus of tubulin increases cytoplasmic dynein and kinesin processivity. *Biophys J.* 2000;78(4):1955-64.
11. Burns RG, Surridge C. Analysis of beta-tubulin sequences reveals highly conserved, coordinated amino acid substitutions. Evidence that these 'hot spots' are directly involved in the conformational change required for dynamic instability. *FEBS Lett.* 1990;271(1-2):1-8.
12. Chakraborti S, Natarajan K, Curiel J, Janke C, Liu J. The emerging role of the tubulin code: From the tubulin molecule to neuronal function and disease. *Cytoskeleton.* 2016;73(10):521-50.
13. Kollman JM, Merdes A, Mourey L, Agard DA. Microtubule nucleation by gamma-tubulin complexes. *Nat Rev Mol Cell Biol.* 2011;12(11):709-21.
14. Stathatos GG, Dunleavy JEM, Zenker J, O'Bryan MK. Delta and epsilon tubulin in mammalian development. *Trends in Cell Biology.* 2021;31(9):774-87.

15. Bahi-Buisson N, Poirier K, Fourniol F, Saillour Y, Valence S, Lebrun N, et al. The wide spectrum of tubulinopathies: what are the key features for the diagnosis? *Brain*. 2014;137:1676-700.
16. Brock S, Stouffs K, Scalais E, D'Hooghe M, Keymolen K, Guerrini R, et al. Tubulinopathies continued: refining the phenotypic spectrum associated with variants in TUBG1. *Eur J Hum Genet*. 2018;26(8):1132-42.
17. Poirier K, Lebrun N, Broix L, Tian G, Saillour Y, Boscheron C, et al. Mutations in TUBG1, DYNC1H1, KIF5C and KIF2A cause malformations of cortical development and microcephaly. *Nat Genet*. 2013;45(6):639+.
18. Luscan R, Mechaussier S, Paul A, Tian GL, Gerard X, Defoort-Dellhemmes S, et al. Mutations in TUBB4B Cause a Distinctive Sensorineural Disease. *Am J Hum Genet*. 2017;101(6):1006-12.
19. Smith BN, Ticozzi N, Fallini C, Gkazi AS, Topp S, Kenna KP, et al. Exome-wide Rare Variant Analysis Identifies TUBA4A Mutations Associated with Familial ALS. *Neuron*. 2014;84(2):324-31.
20. Stoupa A, Adam F, Kariyawasam D, Strassel C, Gawade S, Szinnai G, et al. TUBB1 mutations cause thyroid dysgenesis associated with abnormal platelet physiology. *Embo Mol Med*. 2018;10(12).
21. Bhagwat S, Dalvi V, Chandrasekhar D, Matthew T, Acharya K, Gajbhiye R, et al. Acetylated alpha-tubulin is reduced in individuals with poor sperm motility. *Fertil Steril*. 2014;101(1):95-104 e3.
22. Chen BB, Wang WJ, Peng XD, Jiang HF, Zhang SZ, Li D, et al. The comprehensive mutational and phenotypic spectrum of TUBB8 in female infertility. *Eur J Hum Genet*. 2019;27(2):300-7.
23. Feng RZ, Yan Z, Li B, Yu M, Sang Q, Tian GL, et al. Mutations in TUBB8 cause a multiplicity of phenotypes in human oocytes and early embryos. *J Med Genet*. 2016;53(10):662-71.
24. Abdollahi MR, Morrison E, Sirey T, Molnar Z, Hayward BE, Carr IM, et al. Mutation of the Variant alpha-Tubulin TUBA8 Results in Polymicrogyria with Optic Nerve Hypoplasia. *Am J Hum Genet*. 2009;85(5):737-44.
25. Tian G, Kong XP, Jaglin XH, Chelly J, Keays D, Cowan NJ. A pachygyria-causing alpha-tubulin mutation results in inefficient cycling with CCT and a deficient interaction with TBCB. *Mol Biol Cell*. 2008;19(3):1152-61.
26. Backwell L, Marsh JA. Diverse Molecular Mechanisms Underlying Pathogenic Protein Mutations: Beyond the Loss-of-Function Paradigm. *Annual Review of Genomics and Human Genetics*. 2022;23(1):null.
27. Aiken J, Buscaglia G, Aiken AS, Moore JK, Bates EA. Tubulin mutations in brain development disorders: Why haploinsufficiency does not explain TUBA1A tubulinopathies. *Cytoskeleton (Hoboken)*. 2020;77(3-4):40-54.
28. Breuss M, Heng JIT, Poirier K, Tian GL, Jaglin XH, Qu ZD, et al. Mutations in the beta-Tubulin Gene TUBB5 Cause Microcephaly with Structural Brain Abnormalities. *Cell Rep*. 2012;2(6):1554-62.
29. Cederquist GY, Luchniak A, Tischfield MA, Peeva M, Song Y, Menezes MP, et al. An inherited TUBB2B mutation alters a kinesin-binding site and causes polymicrogyria, CFEOM and axon dysinnervation. *Hum Mol Genet*. 2012;21(26):5484-99.
30. Cushion TD, Paciorkowski AR, Pilz DT, Mullins JGL, Seltzer LE, Marion RW, et al. De Novo Mutations in the Beta-Tubulin Gene TUBB2A Cause Simplified Gyral Patterning and Infantile-Onset Epilepsy. *Am J Hum Genet*. 2014;94(4):634-41.

31. Ti SC, Pamula MC, Howes SC, Duellberg C, Cade NI, Kleiner RE, et al. Mutations in Human Tubulin Proximal to the Kinesin-Binding Site Alter Dynamic Instability at Microtubule Plus- and Minus-Ends. *Dev Cell*. 2016;37(1):72-84.
32. Tischfield MA, Baris HN, Wu C, Rudolph G, Van Maldergem L, He W, et al. Human TUBB3 Mutations Perturb Microtubule Dynamics, Kinesin Interactions, and Axon Guidance. *Cell*. 2010;140(1):74-87.
33. Niwa S, Takahashi H, Hirokawa N. beta-Tubulin mutations that cause severe neuropathies disrupt axonal transport. *Embo J*. 2013;32(10):1352-64.
34. Park K, Hoff KJ, Wethekam L, Stence N, Saenz M, Moore JK. Kinetically Stabilizing Mutations in Beta Tubulins Create Iso-type-Specific Brain Malformations. *Front Cell Dev Biol*. 2021;9:765992.
35. Hebebrand M, Huffmeier U, Trollmann R, Hehr U, Uebe S, Ekici AB, et al. The mutational and phenotypic spectrum of TUBA1A-associated tubulinopathy. *Orphanet J Rare Dis*. 2019;14.
36. Livesey BJ, Marsh JA. Using deep mutational scanning to benchmark variant effect predictors and identify disease mutations. *Molecular Systems Biology*. 2020;16(7):e9380.
37. Gerasimavicius L, Livesey BJ, Marsh JA. Loss-of-function, gain-of-function and dominant-negative mutations have profoundly different effects on protein structure: implications for variant effect prediction. *bioRxiv*. 2021:2021.10.23.465554.
38. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. 2018;46(D1):D1062-D7.
39. Pham CL, Morrisette NS. The tubulin mutation database: A resource for the cytoskeleton community. *Cytoskeleton*. 2019;76(2):186-91.
40. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv*. 2019:531210.
41. Huzil JT, Chen K, Kurgan L, Tuszynski JA. The roles of beta-tubulin mutations and isotype expression in acquired drug resistance. *Cancer Inform*. 2007;3:159-81.
42. Nami B, Wang ZX. Genetics and Expression Profile of the Tubulin Gene Superfamily in Breast Cancer Subtypes and Its Relation to Taxane Resistance. *Cancers*. 2018;10(8).
43. Parker AL, Teo WS, McCarroll JA, Kavallaris M. An Emerging Role for Tubulin Iso-types in Modulating Cancer Biology and Chemotherapy Resistance. *Int J Mol Sci*. 2017;18(7).
44. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536(7616):285-91.
45. Burley K, Westbury SK, Mumford AD. TUBB1 variants and human platelet traits. *Platelets*. 2018;29(2):209-11.
46. Kunishima S, Kobayashi R, Itoh TJ, Hamaguchi M, Saito H. Mutation of the beta 1-tubulin gene associated with congenital macrothrombocytopenia affecting microtubule assembly. *Blood*. 2009;113(2):458-61.
47. Kunishima S, Nishimura S, Suzuki H, Imaizumi M, Saito H. TUBB1 mutation disrupting microtubule assembly impairs proplatelet formation and results in congenital macrothrombocytopenia. *Eur J Haematol*. 2014;92(4):276-82.

48. Lecine P, Italiano JE, Jr., Kim SW, Villeval JL, Shivdasani RA. Hematopoietic-specific beta 1 tubulin participates in a pathway of platelet biogenesis dependent on the transcription factor NF-E2. *Blood*. 2000;96(4):1366-73.
49. Wang D, Villasante A, Lewis SA, Cowan NJ. The mammalian beta-tubulin repertoire: hematopoietic expression of a novel, heterologous beta-tubulin isotype. *J Cell Biol*. 1986;103(5):1903-10.
50. Chen BB, Li B, Li D, Yan Z, Mao XY, Xu Y, et al. Novel mutations and structural deletions in TUBB8: expanding mutational and phenotypic spectrum of patients with arrest in oocyte maturation, fertilization or early embryonic development. *Hum Reprod*. 2017;32(2):457-64.
51. Feng RZ, Sang Q, Kuang YP, Sun XX, Yan Z, Zhang SZ, et al. Mutations in TUBB8 and Human Oocyte Meiotic Arrest. *New Engl J Med*. 2016;374(3):223-32.
52. Huang LL, Tong XH, Luo LH, Zheng SX, Jin RT, Fu YY, et al. Mutation analysis of the TUBB8 gene in nine infertile women with oocyte maturation arrest. *Reprod Biomed Online*. 2017;35(3):305-10.
53. Xiang JJ, Wang W, Qian CF, Xue JY, Wang T, Li HB, et al. Human oocyte maturation arrest caused by a novel missense mutation in TUBB8. *J Int Med Res*. 2018;46(9):3759-64.
54. Levy ED. A Simple Definition of Structural Regions in Proteins and Its Use in Analyzing Interface Evolution. *J Mol Biol*. 2010;403(4):660-70.
55. Cushion TD, Dobyns WB, Mullins JGL, Stoodley N, Chung SK, Fry AE, et al. Overlapping cortical malformations and mutations in TUBB2B and TUBA1A. *Brain*. 2013;136:536-48.
56. Erro R, Hersheson J, Ganos C, Mencacci NE, Stamelou M, Batla A, et al. H-ABC syndrome and DYT4: Variable expressivity or pleiotropy of TUBB4 mutations? *Movement Disord*. 2015;30(6):828-33.
57. Romaniello R, Arrigoni F, Panzeri E, Poretti A, Micalizzi A, Citterio A, et al. Tubulin-related cerebellar dysplasia: definition of a distinct pattern of cerebellar malformation (vol 27, pg 5080, 2017). *Eur Radiol*. 2017;27(12):5093-.
58. Rice LM, Montabana EA, Agard DA. The lattice as allosteric effector: Structural studies of alpha beta- and gamma-tubulin clarify the role of GTP in microtubule assembly. *P Natl Acad Sci USA*. 2008;105(14):5378-83.
59. Wieczorek M, Urnavicius L, Ti SC, Molloy KR, Chait BT, Kapoor TM. Asymmetric Molecular Architecture of the Human gamma-Tubulin Ring Complex. *Cell*. 2020;180(1):165-+.
60. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: A study of more than 1000 mutations. *J Mol Biol*. 2002;320(2):369-87.
61. Gerasimavicius L, Liu X, Marsh JA. Identification of pathogenic missense mutations using protein stability predictors. *Sci Rep*. 2020;10(1):15387.
62. Fallet-Bianco C, Laquerriere A, Poirier K, Razavi F, Guimiot F, Dias P, et al. Mutations in tubulin genes are frequent causes of various foetal malformations of cortical development including microlissencephaly. *Acta Neuropathol Com*. 2014;2.
63. Oegema R, Cushion TD, Phelps IG, Chung SK, Dempsey JC, Collins S, et al. Recognizable cerebellar dysplasia associated with mutations in multiple tubulin genes. *Hum Mol Genet*. 2015;24(18):5313-25.
64. Riesselman AJ, Ingraham JB, Marks DS. Deep generative models of genetic variation capture the effects of mutations. *Nat Methods*. 2018;15(10):816-+.

65. Romaniello R, Zucca C, Arrigoni F, Bonanni P, Panzeri E, Bassi MT, et al. Epilepsy in Tubulinopathy: Personal Series and Literature Review. *Cells*. 2019;8(7).
66. Schatz PJ, Solomon F, Botstein D. Isolation and Characterization of Conditional-Lethal Mutations in the Tub1 Alpha-Tubulin Gene of the Yeast *Saccharomyces-Cerevisiae*. *Genetics*. 1988;120(3):681-95.
67. Hausrat TJ, Radwitz J, Lombino FL, Breiden P, Kneussel M. Alpha- and beta-tubulin isoforms are differentially expressed during brain development. *Dev Neurobiol*. 2021;81(3):333-50.
68. Grimm DG, Azencott CA, Aicheler F, Gieraths U, MacArthur DG, Samocha KE, et al. The Evaluation of Tools Used to Predict the Impact of Missense Variants Is Hindered by Two Types of Circularity. *Hum Mutat*. 2015;36(5):513-23.
69. Online Mendelian Inheritance in Man, OMIM®. [Internet]. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD). 2019. Available from: <https://omim.org/>.
70. Schrodinger, LLC. The PyMOL Molecular Graphics System, Version 2.3.2. 2015.
71. Cheeseman IM, Desai A. A combined approach for the localization and tandem affinity purification of protein complexes from metazoans. *Science's STKE : signal transduction knowledge environment*. 2005;2005(266):p11.
72. Goldberg IG, Allan C, Burel J-M, Creager D, Falconi A, Hochheiser H, et al. The Open Microscopy Environment (OME) Data Model and XML file: open tools for informatics and quantitative analysis in biological imaging. *Genome Biology*. 2005;6(5):R47.
73. Rueden CT, Schindelin J, Hiner MC, DeZonia BE, Walter AE, Arena ET, et al. ImageJ2: ImageJ for the next generation of scientific image data. *BMC Bioinformatics*. 2017;18(1):529.
74. Aiken J, Buscaglia G, Bates EA, Moore JK. The alpha-Tubulin gene TUBA1A in Brain Development: A Key Ingredient in the Neuronal Isoform Blend. *J Dev Biol*. 2017;5(3).
75. Boissel S, Fallet-Bianco C, Chitayat D, Kremer V, Nassif C, Rypens F, et al. Genomic study of severe fetal anomalies and discovery of GREB1L mutations in renal agenesis. *Genet Med*. 2018;20(7):745-53.
76. Gardner JF, Cushion TD, Niotakis G, Olson HE, Grant PE, Scott RH, et al. Clinical and Functional Characterization of the Recurrent TUBA1A p.(Arg2His) Mutation. *Brain Sci*. 2018;8(8).
77. Keays DA, Tian G, Poirier K, Huang GJ, Siebold C, Cleak J, et al. Mutations in alpha-tubulin cause abnormal neuronal migration in mice and lissencephaly in humans. *Cell*. 2007;128(1):45-57.
78. Poirier K, Keays DA, Francis F, Saillour Y, Bahi N, Manouvrier S, et al. Large spectrum of lissencephaly and pachygyria phenotypes resulting from de novo missense mutations in tubulin alpha 1A (TUBA1A). *Hum Mutat*. 2007;28(11):1055-64.
79. Martello LA, Verdier-Pinard P, Shen HJ, He LF, Torres K, Orr GA, et al. Elevated levels of microtubule destabilizing factors in a Taxol-resistant/dependent A549 cell line with an alpha-tubulin mutation. *Cancer Res*. 2003;63(6):1207-13.
80. Poruchynsky MS, Kim JH, Nogales E, Annable T, Loganzo F, Greenberger LM, et al. Tumor cells resistant to a microtubule-depolymerizing hemiasterlin analogue, HTI-286, have mutations in alpha- or beta-tubulin and increased microtubule stability. *Biochemistry-U.S.* 2004;43(44):13944-54.
81. Yang CPH, Verdier-Pinard P, Wang F, Lippaine-Horvath E, He LF, Li DS, et al. A highly epothilone B-resistant A549 cell line with mutations in tubulin that confer drug dependence. *Mol Cancer Ther*. 2005;4(6):987-95.

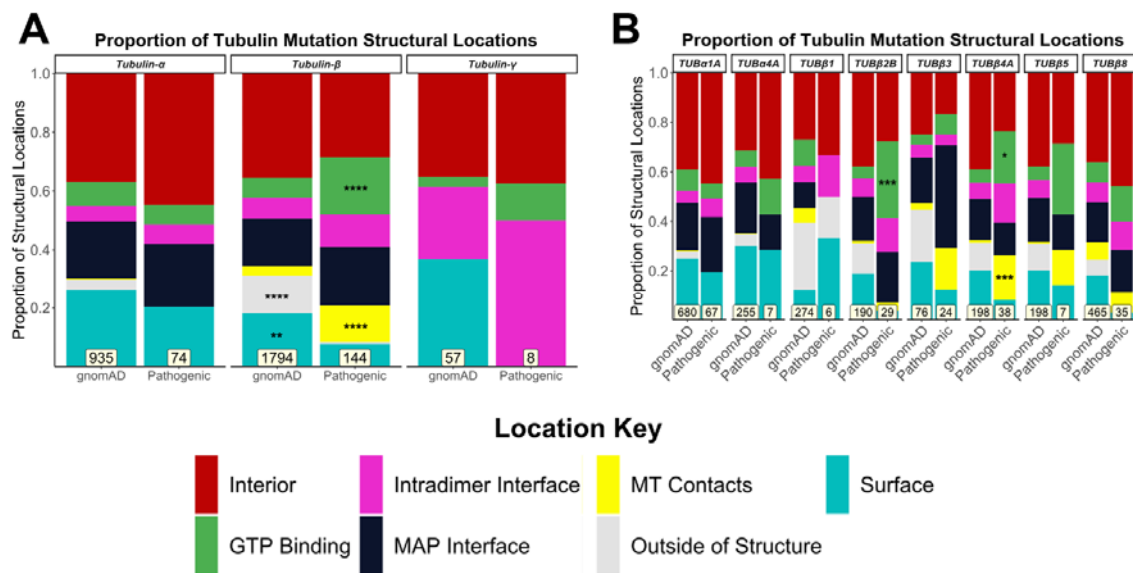
82. Wang J, Chen W, Wei WW, Lou JY. Oncogene TUBA1C promotes migration and proliferation in hepatocellular carcinoma and predicts a poor prognosis. *Oncotarget*. 2017;8(56):96215-24.
83. Alazami AM, Patel N, Shamseldin HE, Anazi S, Al-Dosari MS, Alzahrani F, et al. Accelerating Novel Candidate Gene Discovery in Neurogenetic Disorders via Whole-Exome Sequencing of Prescreened Multiplex Consanguineous Families. *Cell Rep*. 2015;10(2):148-61.
84. Sferra A, Fattori F, Rizza T, Flex E, Bellacchio E, Bruxelles A, et al. Defective kinesin binding of TUBB2A causes progressive spastic ataxia syndrome resembling saccinopathy. *Hum Mol Genet*. 2018;27(11):1892-904.
85. Amrom D, Tanyalcin I, Verhelst H, Deconinck N, Brouhard GJ, Decarie JC, et al. Polymicrogyria with dysmorphic basal ganglia? Think tubulin! *Clin Genet*. 2014;85(2):178-83.
86. Guerrini R, Mei D, Cordelli DM, Pucatti D, Franzoni E, Parrini E. Symmetric polymicrogyria and pachygyria associated with TUBB2B gene mutations. *Eur J Hum Genet*. 2012;20(9):995-8.
87. Jaglin XH, Poirier K, Saillour Y, Buhler E, Tian G, Bahi-Buisson N, et al. Mutations in the beta-tubulin gene TUBB2B result in asymmetrical polymicrogyria. *Nat Genet*. 2009;41(6):746-52.
88. MacKinnon S, Oystreck DT, Andrews C, Chan WM, Hunter DG, Engle EC. Diagnostic Distinctions and Genetic Analysis of Patients Diagnosed with Moebius Syndrome. *Ophthalmology*. 2014;121(7):1461-8.
89. Minoura I, Takazaki H, Ayukawa R, Saruta C, Hachikubo Y, Uchimura S, et al. Reversal of axonal growth defects in an extraocular fibrosis model by engineering the kinesin-microtubule interface. *Nat Commun*. 2016;7.
90. Arai-Ichinoi N, Uematsu M, Sato R, Suzuki T, Kudo H, Kikuchi A, et al. Genetic heterogeneity in 26 infants with a hypomyelinating leukodystrophy. *Hum Genet*. 2016;135(1):89-98.
91. Blumkin L, Halevy A, Ben-Ami-Raichman D, Dahari D, Haviv A, Sarit C, et al. Expansion of the spectrum of TUBB4A-related disorders: a new phenotype associated with a novel mutation in the TUBB4A gene. *Neurogenetics*. 2014;15(2):107-13.
92. Carvalho D, Santos S, Martins B, Marques FP. TUBB4A novel mutation reinforces the genotype-phenotype correlation of hypomyelination with atrophy of the basal ganglia and cerebellum. *Brain*. 2015;138.
93. Curiel J, Bey GR, Takanoashi A, Bugiani M, Fu XQ, Wolf NI, et al. TUBB4A mutations result in specific neuronal and oligodendrocytic defects that closely match clinically distinct phenotypes. *Hum Mol Genet*. 2017;26(22):4506-18.
94. Ferreira C, Poretti A, Cohen J, Hamosh A, Naidu S. Novel TUBB4A mutations and expansion of the neuroimaging phenotype of hypomyelination with atrophy of the basal ganglia and cerebellum (H-ABC). *Am J Med Genet A*. 2014;164(7):1802-7.
95. Hamilton EM, Polder E, Vanderver A, Naidu S, Schiffmann R, Fisher K, et al. Hypomyelination with atrophy of the basal ganglia and cerebellum: further delineation of the phenotype and genotype-phenotype correlation. *Brain*. 2014;137:1921-30.
96. Hersheson J, Mencacci NE, Davis M, MacDonald N, Trabzuni D, Ryten M, et al. Mutations in the autoregulatory domain of beta-tubulin 4a cause hereditary dystonia. *Ann Neurol*. 2013;73(4):546-53.
97. Isakov O, Lev D, Blumkin L, Celniker G, Leshinsky-Silver E, Shomron N. Crowdfunding Effort Identifies the Causative Mutation in a Patient with Nystagmus, Microcephaly, Dystonia and Hypomyelination. *J Genet Genomics*. 2015;42(2):79-81.

98. Kancheva D, Chamova T, Guergueltcheva V, Mitev V, Azmanov DN, Kalaydjieva L, et al. Mosaic dominant TUBB4A mutation in an inbred family with complicated hereditary spastic paraplegia. *Movement Disord*. 2015;30(6):854-8.
99. Lohmann K, Wilcox RA, Winkler S, Ramirez A, Rakovic A, Park JS, et al. Whispering dysphonia (DYT4 dystonia) is caused by a mutation in the TUBB4 gene. *Ann Neurol*. 2013;73(4):537-45.
100. Lu Y, Ondo Y, Shimojima K, Osaka H, Yamamoto T. A novel TUBB4A mutation G96R identified in a patient with hypomyelinating leukodystrophy onset beyond adolescence. *Hum Genome Var*. 2017;4:17035.
101. Miyatake S, Osaka H, Shiina M, Sasaki M, Takanashi J, Haginoya K, et al. Expanding the phenotypic spectrum of TUBB4A-associated hypomyelinating leukoencephalopathies. *Neurology*. 2014;82(24):2230-7.
102. Pizzino A, Pierson TM, Guo Y, Helman G, Fortini S, Guerrero K, et al. TUBB4A de novo mutations cause isolated hypomyelination. *Neurology*. 2014;83(10):898-902.
103. Purnell SM, Bleyl SB, Bonkowsky JL. Clinical exome sequencing identifies a novel TUBB4A mutation in a child with static hypomyelinating leukodystrophy. *Pediatr Neurol*. 2014;50(6):608-11.
104. Pyle A, Smertenko T, Bargiela D, Griffin H, Duff J, Appleton M, et al. Exome sequencing in undiagnosed inherited and sporadic ataxias. *Brain*. 2015;138:276-83.
105. Sagnelli A, Magri S, Farina L, Chiapparini L, Marotta G, Tonduti D, et al. Early-onset progressive spastic paraplegia caused by a novel TUBB4A mutation: brain MRI and FDG-PET findings. *J Neurol*. 2016;263(3):591-3.
106. Simons C, Wolf NI, McNeil N, Caldovic L, Devaney JM, Takanohashi A, et al. A de novo mutation in the beta-tubulin gene TUBB4A results in the leukoencephalopathy hypomyelination with atrophy of the basal ganglia and cerebellum. *Am J Hum Genet*. 2013;92(5):767-73.
107. Vanderver A, Simons C, Helman G, Crawford J, Wolf NI, Bernard G, et al. Whole Exome Sequencing in Patients with White Matter Abnormalities. *Ann Neurol*. 2016;79(6):1031-7.
108. Wang WW, Zhang HX, Wang XM, Patterson J, Winter P, Graham K, et al. Novel mutations involving beta I-, beta IIA-, or beta IVB-tubulin isotypes with functional resemblance to beta III-tubulin in breast cancer. *Protoplasma*. 2017;254(3):1163-73.
109. Bowling KM, Thompson ML, Amaral MD, Finnila CR, Hiatt SM, Engel KL, et al. Genomic diagnosis for children with intellectual disability and/or developmental delay. *Genome Med*. 2017;9.
110. Isrie M, Breuss M, Tian GL, Hansen AH, Cristofoli F, Morandell J, et al. Mutations in Either TUBB or MAPRE2 Cause Circumferential Skin Creases Kunze Type. *Am J Hum Genet*. 2015;97(6):790-800.
111. Leonard NJ. A second patient with MCA/MR syndrome with multiple circumferential skin creases. *Am J Med Genet*. 2002;112(1):91-4.
112. Madrigal I, Rabionet R, Alvarez-Mora MI, Sanchez A, Rodriguez-Revenga L, Estivill X, et al. Spectrum of clinical heterogeneity of beta-tubulin TUBB5 gene mutations. *Gene*. 2019;695:12-7.
113. Ulucan H, Koparir E, Koparir A, Karaca E, Emre R, Gezdirici A, et al. Circumferential skin folds and multiple anomalies: confirmation of a distinct autosomal recessive Michelin tire baby syndrome. *Clin Dysmorphol*. 2013;22(2):87-90.



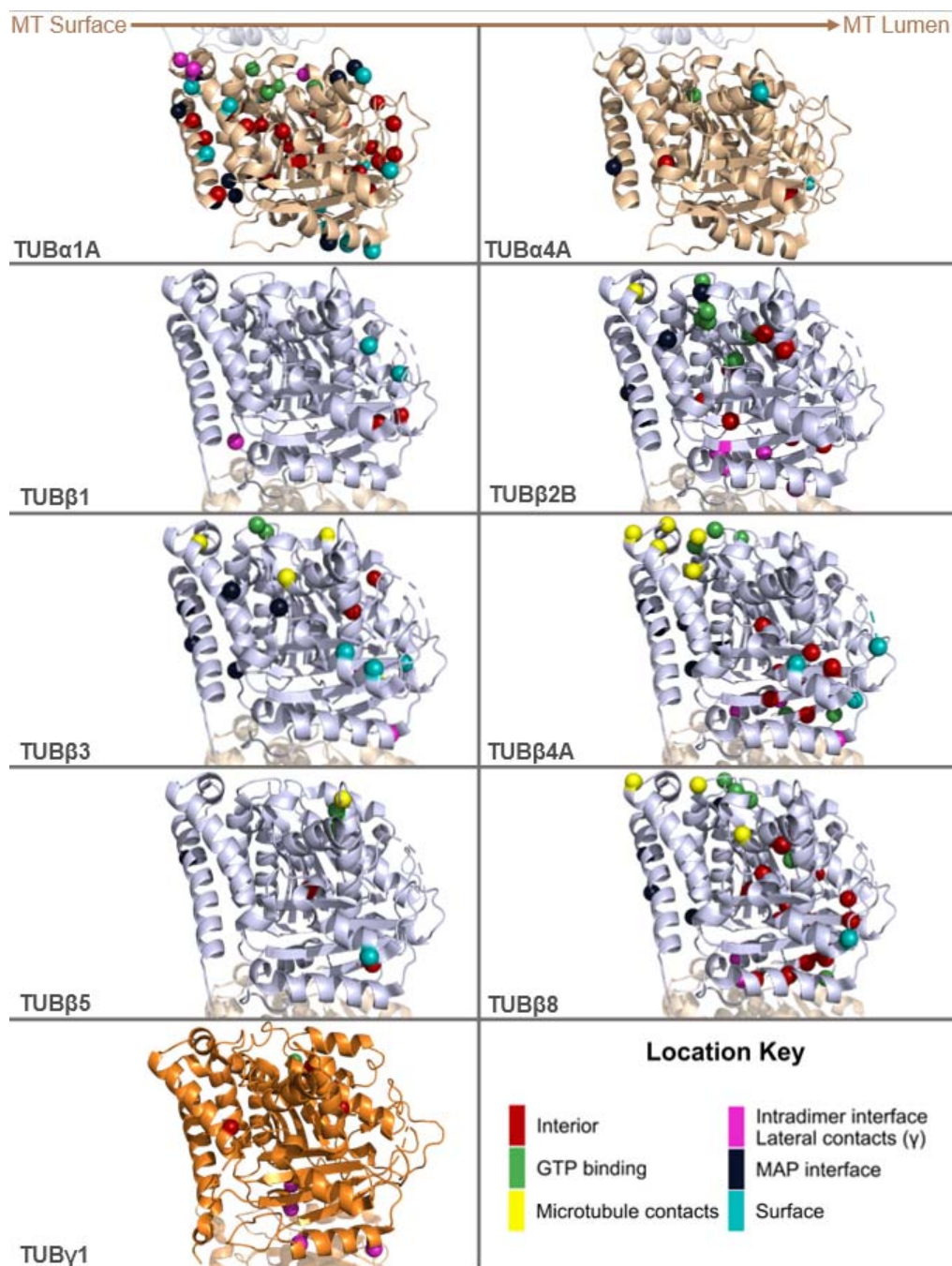
114. Wiszniewski W, Gawlinski P, Gambin T, Bekiesinska-Figatowska M, Obersztyn E, Antczak-Marach D, et al. Comprehensive genomic analysis of patients with disorders of cerebral cortical development. *Eur J Hum Genet.* 2018;26(8):1121-31.
115. Fazeli W, Herkenrath P, Stiller B, Neugebauer A, Fricke J, Lang-Roth R, et al. A TUBB6 mutation is associated with autosomal dominant non-progressive congenital facial palsy, bilateral ptosis and velopharyngeal dysfunction. *Hum Mol Genet.* 2017;26(20):4055-66.
116. Leask A, Stearns T. Expression of amino- and carboxyl-terminal gamma- and alpha-tubulin mutants in cultured epithelial cells. *J Biol Chem.* 1998;273(5):2661-8.

## Figures



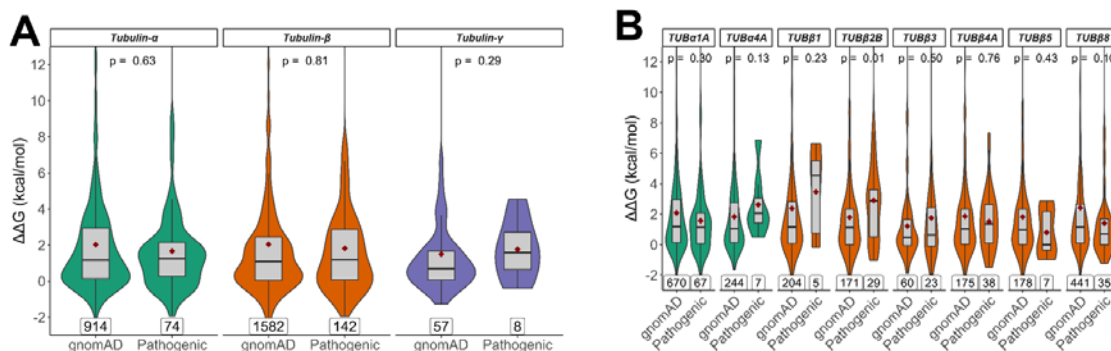
**Figure 1: Comparison of the distribution of structural locations between pathogenic and putatively benign tubulin variants.**

The proportion of pathogenic and putatively benign gnomAD mutations in each location type for tubulin- $\alpha$ ,  $\beta$  and  $\gamma$  globally (**A**) and for individual isotypes (**B**). Mutation totals for each group are shown at the bottom, and only isotypes with at least five pathogenic mutations were included. Fisher's exact tests were used to compare frequencies between gnomAD and pathogenic mutations considering each family and isotype separately. Asterisks indicate a location class with a significantly higher proportion of mutations compared to its corresponding group, where \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ , \*\*\*\*  $p < 0.0001$ .



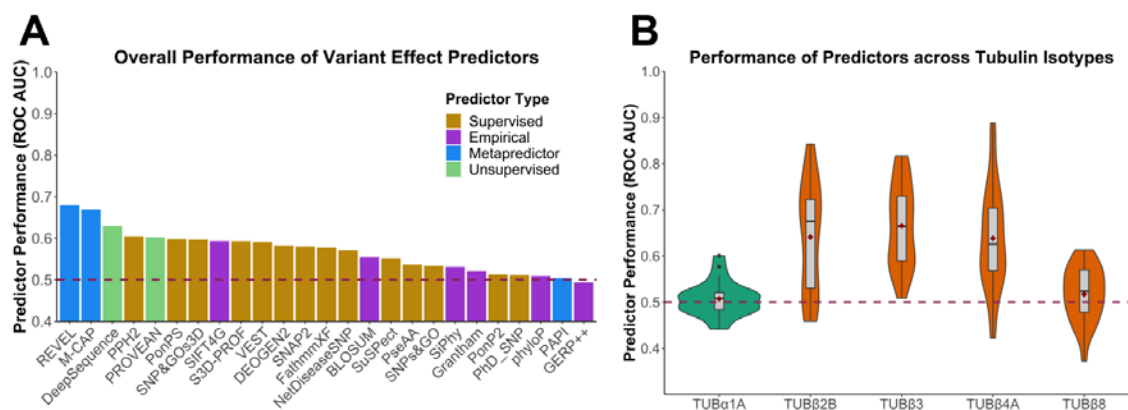
**Figure 2: Visualisation of pathogenic tubulin mutations on three-dimensional structures of tubulin heterodimers.**

Structures of tubulin isotypes with at least five identified pathogenic mutations. Coloured residues indicate pathogenic mutations according to their location. The position of the tubulin subunit in relation to the microtubule (MT) is noted at the top. Residues in red denote mutations buried in the protein interior, while light cyan residues are on the surface. Green residues indicate mutations occurring at GTP binding interfaces, while ones in magenta highlight residues on intradimer interfaces. Navy blue and yellow residues mark mutations on residues interacting with microtubule associating proteins (MAPs) and other tubulin dimers, respectively. PDB IDs: 6s8k (for all tubulin- $\alpha$  and  $\beta$  isotypes) and 3cb2 (for TUB $\gamma$ 1).



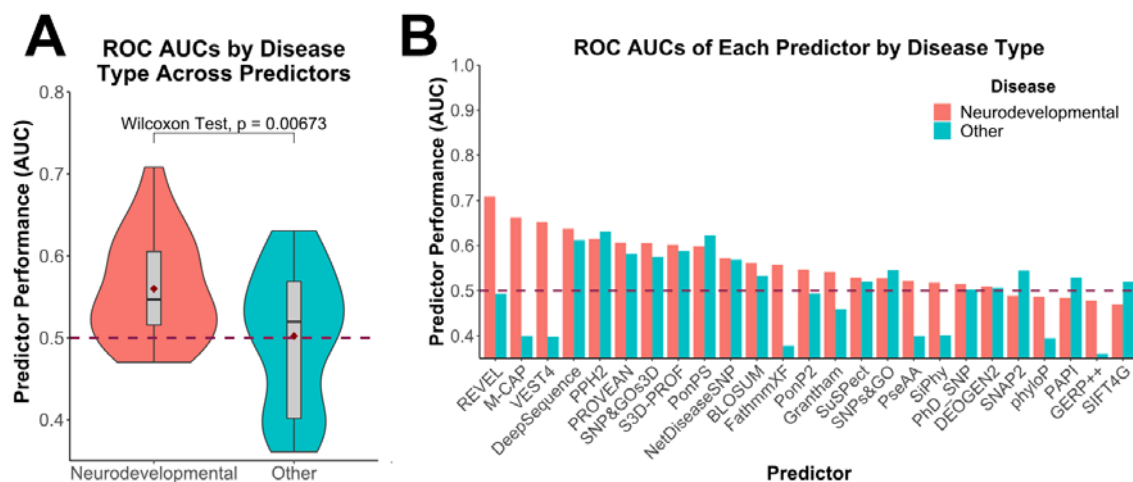
**Figure 3: Comparison of predicted changes in protein stability between pathogenic and putatively benign tubulin mutations.**

$\Delta\Delta G$  values representing the predicted change in free energy of folding were calculated with FoldX considering the structure of the monomeric subunit only. Scores are shown for tubulin- $\alpha$  and - $\beta$  families globally (**A**) and in isotypes with at least five identified pathogenic mutations (**B**). Maroon diamonds indicate the mean  $\Delta\Delta G$  values, and mutation totals for each group are also shown at the bottom. The p-values displayed were obtained via unpaired Wilcoxon tests.



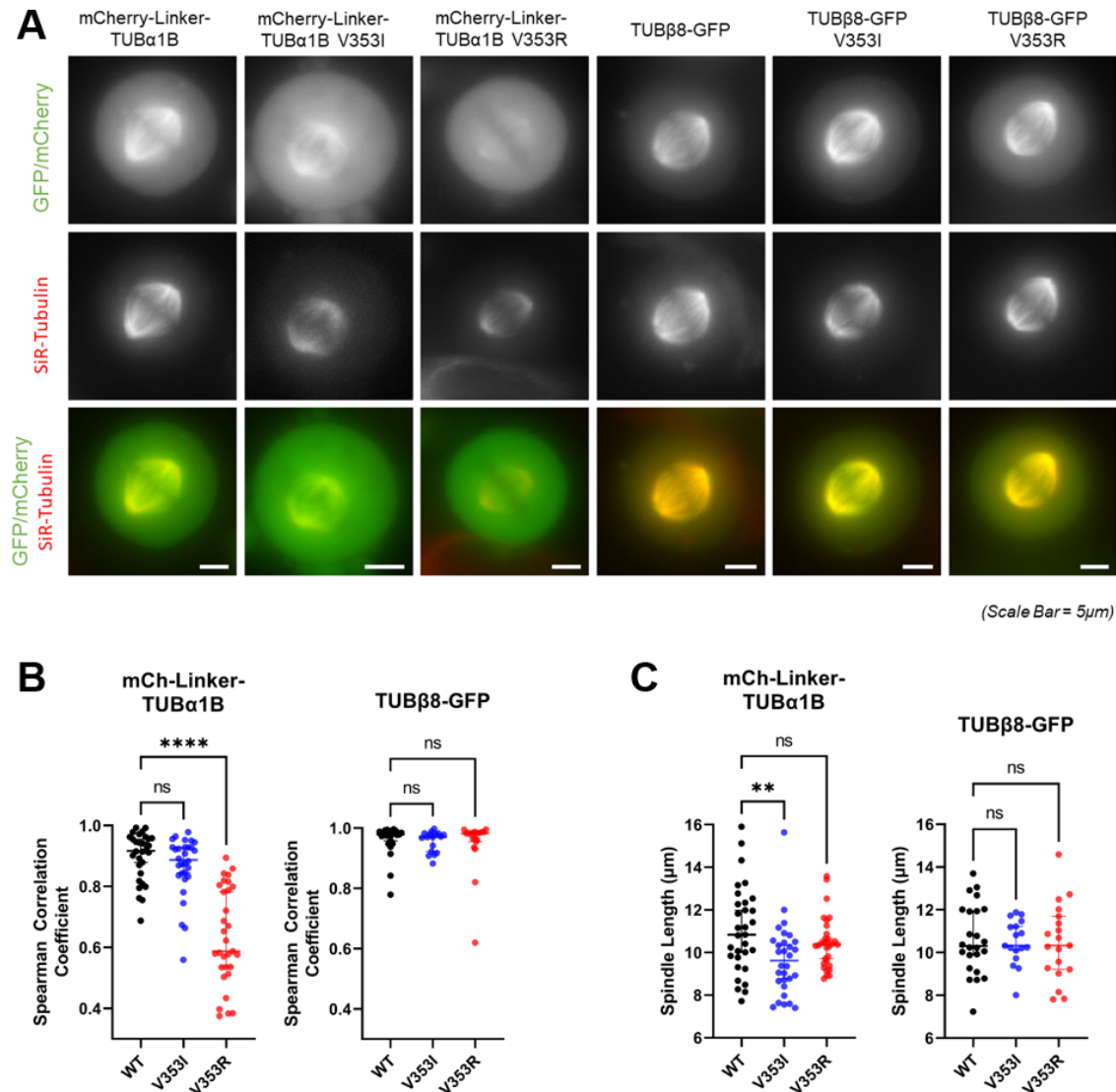
**Figure 4: Assessment of VEP performance for identification of pathogenic tubulin mutations.**

**(A)** ROC AUC values for each VEP across all tubulin- $\alpha$  and - $\beta$  isotypes with at least one identified pathogenic mutation, colour coded according to predictor category. **(B)** Distribution of ROC AUC values across all VEPs for isotypes with at least 10 identified pathogenic missense mutations. Dashed line indicates the performance of a random predictor. Maroon diamonds indicate the mean area.



**Figure 5: Comparison of VEP performance on pathogenic mutations in tubulin genes associated with neurodevelopmental vs other disease phenotypes.**

**(A)** Distribution of ROC AUC values across all VEPs for isotypes with mutations linked with neurodevelopmental or other disease phenotypes. The p-value stated was obtained via a paired Wilcoxon test. Maroon diamonds indicate the mean area. **(B)** ROC AUC values for each VEP in isotypes with mutations associated with neurodevelopmental or other disease phenotypes. Dashed line indicates the performance of a random predictor.



**Figure 6: Effects of V353 mutants on microtubules and mitotic spindles in cells.**

**(A)** Representative live-cell images of mitotic HeLa cells after being transfected with fluorescently tagged wild type and mutant tubulin- $\alpha$ 1B and tubulin- $\beta$ 8 constructs (green) and incubated with SiR-tubulin (red) **(B)** Wide linescans across the whole mitotic spindle were taken to measure fluorescence intensity of tubulin across the mitotic spindle compared to total tubulin. For each cell, two-sided Spearman correlations were calculated between both signal intensities. **(C)** Spindle length measurements for the cells transfected with fluorescently tagged wild-type and mutant tubulin- $\alpha$ 1B and tubulin- $\beta$ 8 and measured in **(B)**. For each construct, median values are displayed with 95% confidence intervals. Asterisks indicate Kruskal-Wallis test (with post-hoc Dunn) significance values, \*\*\*,  $P < 0.001$ ; \*\*\*\*,  $P < 0.0001$ . Bars, 5 $\mu$ m.

## Tables

**Table 1: Summary of all tubulin isotypes considered for this study.**

The number of dominant pathogenic mutations identified for each isotype is denoted, as well as their pathogenicity classification. Where necessary, additional comments about the pathogenicity type and references for all mutations are also included. MCM = missense constraint metric (Z-Score obtained from gnomAD); ALS = amyotrophic lateral sclerosis; H-ABC = hypomyelination with atrophy of basal ganglia and cerebellum

<i>Isotype</i>	<i># pathogenic</i>	<i># gnomAD</i>	<i>MCM</i>	<i>Phenotype classification</i>	<i>References and comments</i>
<i>TUBα1A</i>	67	680	5.584	neurodevelopmental	Various neuropathies (35, 74-78)
<i>TUBα1B</i>	0	696	5.412	cancer-only	Roles in resistance to chemotherapy (79-81)
<i>TUBα1C</i>	0	153	2.171	cancer-only	Identified as an oncogene (82)
<i>TUBα3D</i>	0	343	1.533	non-disease	
<i>TUBα3E</i>	0	531	-0.296	non-disease	Only 1 homozygous mutation known, which was excluded (83)
<i>TUBα4A</i>	7	255	3.303	other (neurodegenerative)	All linked with ALS (19). One deletion mutation also identified ( <i>excluded</i> )
<i>TUBα8</i>	0	201	0.631	non-disease	Homozygous 14bp deletion linked with polymicroglia (24) ( <i>excluded</i> )
<i>TUBβ1</i>	6	268	0.003	other (bleeding disorders)	Linked to platelet defects (45-47)
<i>TUBβ2A</i>	3	192	5.263	neurodevelopmental	Simplified gyral patterning, infantile-onset epilepsy and progressive spastic ataxia (30, 84)
<i>TUBβ2B</i>	29	190	5.120	neurodevelopmental	Various neuropathies (15, 29, 55, 57, 63, 85-87)
<i>TUBβ3</i>	24	76	4.579	neurodevelopmental	Various neuropathies, a few mutations associated with milder phenotypes (15, 32, 55, 57, 63, 65, 88, 89)
<i>TUBβ4A</i>	38	198	4.262	neurodevelopmental	Mostly H-ABC and Hypomyelination, some less severe and a few linked with Dystonia (56, 90-107)
<i>TUBβ4B</i>	2	201	4.498	other (sensory)	Linked heavily with cancer (108). Two mutations have been linked to a sensorineural neurodegenerative disease impairing vision (18)
<i>TUBβ5</i>	7	198	5.625	neurodevelopmental	Various neuropathies, mostly severe (28, 109-114)
<i>TUBβ6</i>	1	74	2.637	neurodevelopmental	Linked with non-progressive congenital facial palsy (115)
<i>TUBβ8</i>	35	464	1.860	other (female infertility)	Linked to oocyte maturation defects (22, 23, 50-53)
<i>TUBβ8B</i>	0	228	-2.372	non-disease	
<i>TUBγ1</i>	8	57	4.155	neurodevelopmental	All linked to cortical development malformations, one of unknown inheritance (16, 17) ( <i>excluded</i> ). Two deletions also cause a lack of centrosome localisation (116)
<i>TUBγ2</i>	0	113	2.443	non-disease	

**Table 2: The most poorly predicted pathogenic tubulin mutations.**

For each of the VEPs used in this study, cumulative distribution ranks were generated in a combined dataset of putatively benign and pathogenic tubulin variants. MCD scores for each variant were calculated by averaging the ranks from each predictor. Pathogenic variants with the lowest MCD scores are shown here, representing those most poorly predicted by VEPs.

Isotype	Mutation	MCD Score
TUB $\beta$ 8	V353I	0.188
TUB $\alpha$ 1A	I238V	0.215
TUB $\alpha$ 1A	I219V	0.215
TUB $\beta$ 8	I4L	0.221
TUB $\alpha$ 1A	D127E	0.245
TUB $\beta$ 8	A352S	0.247
TUB $\beta$ 1	R307H	0.253
TUB $\beta$ 8	M330I	0.280
TUB $\beta$ 5	V353I	0.280
TUB $\alpha$ 1A	S54N	0.300
TUB $\beta$ 8	R2K	0.311
TUB $\alpha$ 1A	A270S	0.326
TUB $\alpha$ 1A	V353I	0.335
TUB $\alpha$ 1A	I188L	0.341
TUB $\beta$ 3	M323V	0.342
TUB $\beta$ 8	F200L	0.352
TUB $\beta$ 8	V175M	0.354
TUB $\alpha$ 1A	I5L	0.357
TUB $\alpha$ 1A	V409I	0.371
TUB $\alpha$ 1A	R214H	0.372