Human Molecular Genetics, 2019, Vol. 00, No. 00 1–9

doi: 10.1093/hmg/ddz002
Advance Access Publication Date: 4 January 2019
General Article

GENERAL ARTICLE

Mutations in *MAP3K1* that cause 46,XY disorders of sex development disrupt distinct structural domains in the protein

Adam Chamberlin[1], Robert Huether[1,†], Aline Z. Machado[2], Michael Groden[3], Hsiao-Mei Liu[1], Kinnari Upadhyay[3], O. Vivian[3], Nathalia L. Gomes[2], Antonio M. Lerario[4], Mirian Y. Nishi[2], Elaine M.F. Costa[2], Berenice Mendonca[2], Sorahia Domenice[2], Jacqueline Velasco[5], Johnny Loke[3] and Harry Ostrer[3,*]

[1]Ambry Genetics, Aliso Viejo, CA 92656, USA, [2]Division of Endocrinology, Hormone and Molecular Genetics Laboratory (LIM/42), Hospital das Clinicas, University of Sao Paulo Medical School, Avenida Dr. Eneas de C Aguiar 155, 2 andar Bloco 6, 05403-900 São Paulo, SP, Brazil, [3]Department of Pathology, Albert Einstein College of Medicine, Bronx, NY 10461, USA, [4]Division of Metabolism, Endocrinology and Diabetes, Department of Internal Medicine, University of Michigan, Ann Arbor, MI 48109, USA and [5]Pediatric Endocrinology, Marshfield Clinic, Marshfield, WI 54449, USA

*To whom correspondence should be addressed at: Department of Pathology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Ullmann 817, Bronx, NY 10461, USA. Tel: (718) 430 8605; Fax: (718) 430 2623; Email: harry.ostrer@einstein.yu.edu

Abstract

Missense mutations in the gene, *MAP3K1*, are a common cause of 46,XY gonadal dysgenesis, accounting for 15–20% of cases [Ostrer, 2014, Disorders of sex development (DSDs): an update. *J. Clin. Endocrinol. Metab.*, **99**, 1503–1509]. Functional studies demonstrated that all of these mutations cause a protein gain-of-function that alters co-factor binding and increases phosphorylation of the downstream MAP kinase pathway targets, MAPK11, MAP3K and MAPK1. This dysregulation of the MAP kinase pathway results in increased CTNNB1, increased expression of WNT4 and FOXL2 and decreased expression of SRY and SOX9. Unique and recurrent pathogenic mutations cluster in three semi-contiguous domains outside the kinase region of the protein, a newly identified N-terminal domain that shares homology with the Guanine Exchange Factor (residues Met164 to Glu231), a Plant HomeoDomain (residues Met442 to Trp495) and an ARMadillo repeat domain (residues Met566 to Glu862). Despite the presence of the mutation clusters and clinical data, there exists a dearth of mechanistic insights behind the development imbalance. In this paper, we use structural modeling and functional data of these mutations to understand alterations of the MAP3K1 protein and the effects on protein folding, binding and downstream target phosphorylation. We show that these mutations have differential effects on protein binding depending on the domains in which they occur. These mutations increase the binding of the RHOA, MAP3K4 and FRAT1 proteins and generally decrease the binding of RAC1. Thus, pathologies in MAP3K1 disrupt the balance between the pro-kinase activities of the RHOA and MAP3K4 binding partners and the inhibitory activity of RAC1.

[†]Present address: Tempus Labs, Chicago, IL 60654, USA.

Received: October 23, 2018. **Revised:** December 19, 2018. **Accepted:** December 31, 2018

Introduction

Disorders of sex development are common, affecting ~1:2000 live births (2). Among these conditions is 46,XY gonadal dysgenesis, also termed 'Swyer syndrome', for which mutations in genes involving transcription factors, signaling molecules and signal transduction molecules in the testes-determining genetic pathway are often identified (3,4). Normal male development includes the *SRY* (NM_003140.2) and *MAP3K4* (NM_001291958.1) genes that promote testis determination through the upregulation of SOX9 (NM_000346.3) via a feed-forward loop involving *FGF9* (NM_002010.2) (1). In turn, several of these genes [*SOX9*, *AXIN1* and *GSK3B* (NM_001146156.1)] could create a block to ovarian development by destabilizing CTNNB1 (β-catenin; NM_001098209.1). In lymphoblastoid cells (LCLs), reduced phosphorylation of MAPK11 (p38:NM_002751.6), MAP3K (ERK1:NM_001040056.2) and MAPK1 (ERK2:NM_138957.3) and reduced FOXL2 (NM_023067.3) expression upregulate the expression of SOX9 and the resulting feed-forward loop to block the ovarian development (1,5,6). In contrast, hyperphosphorylated MAPK11 (p38), MAP3K (ERK1) and MAPK1 (ERK2) and expression of AXIN1 (via destabilized GSK3β) and FRAT1 (NM_005479.3) result in the stabilization of CTNNB1 (β-catenin) and the upregulation of its downstream targets, FOXL2 and FST (NM_006350.3). Several studies have demonstrated that missense mutations in *MAP3K1* (NM_005921.1) cause a gain-of-function that increased the binding of protein co-factors, such as RHOA (NM_001313941.1), which result in an increase in phosphorylation of the downstream targets, MAPK11 (p38), MAP3K (ERK1) and MAPK1 (ERK2) (7,8). The result on the MAP kinase pathway is tilting the balance away from testes-promoting factors with increased activity of CTNNB1 (β-catenin), increased expression of WNT4 (NM_030761.4) and FOXL2 and decreased expression of SRY and SOX9 leading to abnormal gonad development. The mutations reported in these studies occurred in exons 2, 3, 13 and 14 of this 20 exon gene and were not directly mapped to the specific domains within the protein. Over time, the number of known unique mutations has grown, including the additional mutations reported here (9). These mutations tend to cluster into a previously unidentified N-terminal domain that shares homology with a Guanine Exchange Factor (GEF—residues Met164 to Glu231) and the ARMadillo repeat domain (ARM—residues Met566 to Glu862). Two other singleton mutations were observed, one in the Plant HomeoDomain (PHD) domain and the other before the N-terminus of the GEF domain. Whereas the presence of the mutation clusters and clinical data associated with them is known, the mechanistic causes behind the development imbalance were previously unknown. Here we provide evidence for a mechanistic explanation of how the missense alterations in these domains result in over-activation of MAP3K1 kinase signaling and a disruption of the testis development pathway.

Results

Mutations in MAP3K1 associated with 46,XY gonadal dysgenesis cluster in the ARM, PHD and a putative GEF and are anticipated to have destabilizing effects

In the current study, we examined 12 different germline mutations in the *MAP3K1* gene expressed in LCLs, which caused abnormal developmental phenotypes, either reported previously or reported here for the first time (Fig. 1). These mutations have the characteristic of being either missense variants (p.Pro153Leu, p.Leu189Arg, p.Leu189Pro, p.Leu189Gln, p.Leu447Trp, p.Leu587His, p.Gly616Arg, p.Leu639Pro, p.Thr657Arg, p.Cys691Arg and p.Leu764Arg) or familial splice acceptor site variants (c.634-8T>A and c.2180-2A>G) that have been previously shown by RNA sequencing to cause in-frame insertions (p.Val211-Val212insIleGln) or deletions (p.Gly727-Ile761del), respectively (7). Samples were not available for functional studies of p.Leu189Gln, p.Leu447Trp, p.Leu587His and p.Leu764Arg. The p.Leu587His mutation co-segregated with the phenotype for three affected individuals in one family suggesting pathogenicity and correlating with the moderately disruptive folding energy (2.34; Supplementary Material, Table S1). The additional variant p.Leu189Gln occurred at a site previously reported as pathogenic for the p.Leu189Pro and p.Leu189Arg mutations and was also reported previously in an affected individual by Ambry Genetics (9). The p.Leu189Arg, p.Leu189Pro, p.Leu189Gln and p.Val211-Val212insIleGln mutations presented here mapped onto a GEF domain (residues Met164 to Glu231), whereas the p.Pro153Leu mutation was located N-terminal to this domain. The p.Leu447Trp mutation has been previously identified in the PHD domain between residues Met442 to Trp495 (10,11), whereas p.Leu587His, p.Gly616Arg, p.Leu639Pro, p.Thr657Arg, p.Cys691Arg and p.Leu764Arg are located in the ARM domain (residues Met566 to Glu862). Additional variants included the hypomorphic allele, p.Lys246Glu and those that were predicted to be benign or likely benign. These variants were C-terminal to the GEF (p.Gln237Arg, p.Ser245Phe and p.Lys246Glu), in the SWI2/SNF2 and MuDR (SWIM) domain between residues Glu303 and His393 (p.Pro370Ser) and C-terminal to the ARM domain (p.Val873Ile, p.Val889Leu, p.Asp806Asn, p.Val906Ile, p.Ser939Cys and p.Met955Thr) (Supplementary Material, Table S1).

Mutations in the GEF domain. GEFs are needed in a variety of signal transduction pathways and functions to facilitate the release of nucleotides from their binding sites (12). Therefore, whereas this particular domain had not been identified previously in MAP3K1, presumably because of a low sequence homology, it was identified through the use of Hidden Markov methods implemented in HHBlits (13) [PDB: 2MNI (14)]. The GEF activity matches well to both the function of MAP3K1 and with the clustering of mutations in this region (Fig. 1). The putative interface of RAC1/RHOA with the GEF domain is based on homologous structures [PDB: 1FOE (15), PDB: 2YIN (16) and PDB: 4YON (17)] that place the p.Val211-Val212insIleGln mutation at the interface with the RHOA and RAC1 proteins (Fig. 2A). The insertion of side chains with the p.Val211-Val212insIleGln mutation introduces steric effects that impact the interface. The variants p.Leu189Arg/Pro/Gln impact the local backbone of the GEF region (Fig. 2B). In support of this being a GEF, we performed a structural alignment of the putative GEF domain on MAP3K1 with the ATPase domain in HSP70, a known nucleotide-binding protein [PDB: 3IUC (18)]. Residues in subdomain IIB of HSP70 that form part of the ADP-binding domain (Lys273, Leu288, Glu293 and Arg297) align with those in the GEF of MAP3K1 (Lys175, Leu189, Glu200 and Arg204; Fig. 3), which supports functional mimicry. Furthermore, a search of the ExAC database revealed that there are no variants that reside at these sites, indicating that the positions are conserved and intolerant to alteration (19). Taken together, these observations point to the likelihood of a GEF domain in this region.

GEFs activate Rho proteins, such as RHOA and RAC1, by catalyzing the exchange of GDP for GTP and are known to be involved with signaling pathways transduced by MAP kinases (20–22). The mutations p.Val211-Val212insIleGln and

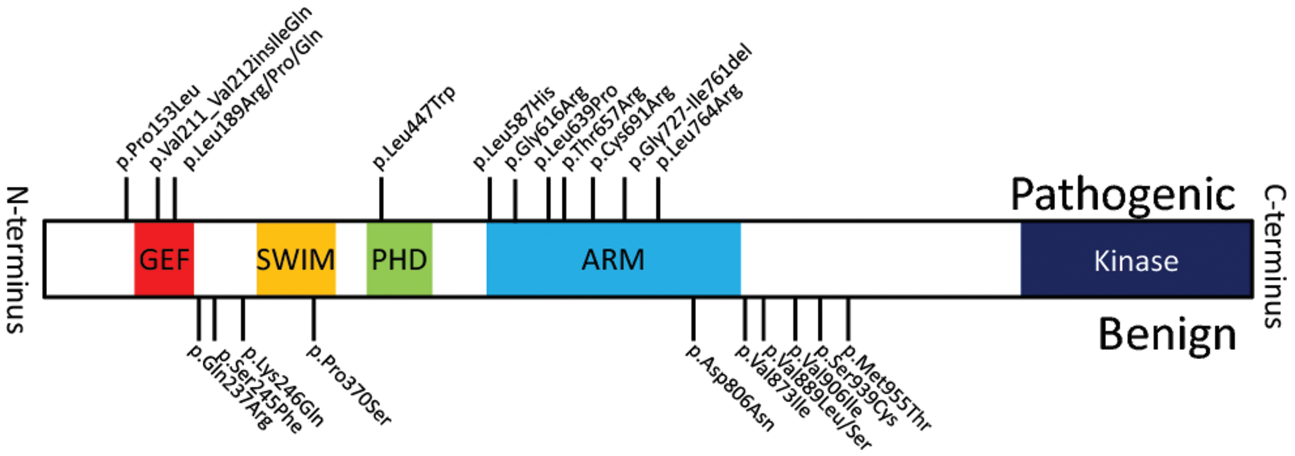

Figure 1. The distribution of known mutations cluster around a putative GEF domain, a PHD and an ARM, while the benign variants were distributed elsewhere through MAP3K1.

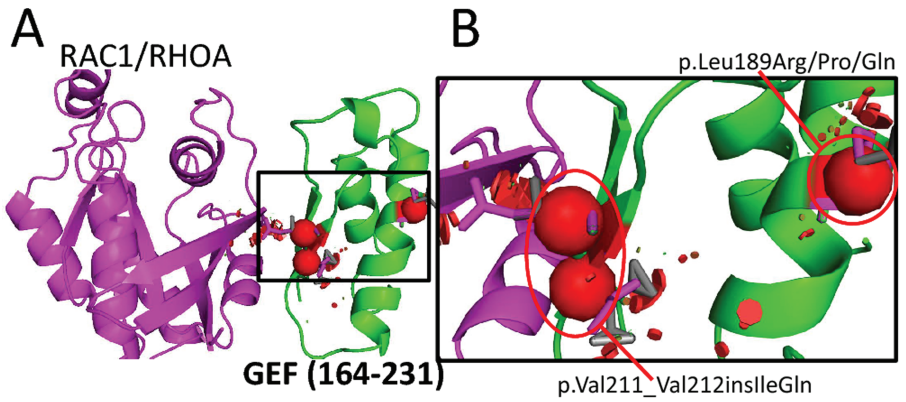

Figure 2. **(A)** The putative binding interface between the GEF domain of MAP3K1 (green) and alteration sites (red), with RAC1 or RHOA (magenta) [PDB: 2MNI (14)]. **(B)** The wild-type side chains (gray sticks) and local disruption by the structure of the side chains of the variants, p.Val211_Val212insIleGln, p.Leu189Arg/Pro/Gln, (magenta sticks) and their steric clashes (red dots).

p.Leu189Arg/Pro/Gln all lie in the structured portions of the GEF domain (Fig. 2B). The mutation p.Val211-Val212insIleGln results in an insertion of a larger isoleucine at position Val211. Structurally, Val211 is buried in a β-sheet facing into an α-helix containing the putative nucleotide-binding region. The large side chain of isoleucine will likely disrupt packing of the surrounding residues resulting in a significant disruption of the domain (7.25 kcal/mol). For comparison, the lower threshold for known disrupting pathogenic variants has been shown to be ∼2.50 kcal/mol (23,24). Similarly, the mutations p.Leu189Arg/Pro/Gln lie within 8 Å of the putative nucleotide-binding site and at a potential internal interface between the MAP3K1 domains (Fig. 3). p.Leu189Arg/Pro/Gln mutations are destabilizing (1.07, 1.15 and 3.89 kcal/mol, respectively) and lie near a structurally sensitive region. The substituted side chains act to disrupt the backbone structure (p.Leu189Pro) or insert a charged/polar side chain into a buried pocket (p.Leu189Arg/Gln) that would disrupt both local folding and the adjacent binding pocket of the nucleotide (Figs 2B and 3). From the review of the structure and the homologous interdomain interactions of HSP70, it is likely that the alterations at Leu189 will have a significantly larger impact on protein function than the destabilization energies would suggest.

The GEF domain is targeted by the two competing proteins, RHOA and RAC1. Experimental binding affinities between the RAC1 and MAP3K1 proteins and the RHOA and MAP3K1 are altered by mutations in the GEF domain, p.Leu189Arg/Pro and p.Val211-Val212insIleGln (Fig. 4). The binding of RHOA was increased for the N-terminal p.Pro153Leu ($P = 5 \times 10^{-9}$), and the binding of RAC1 was decreased for the N-terminus and the GEF domain p.Pro153Leu/Arg/Pro mutations ($P = 0.001$, $P = 3 \times 10^{-9}$ and $P = 6 \times 10^{-6}$).

Molecular dynamics studies of the contact energies between the residues of the GEF domain with RAC1 and RHOA were performed based upon the same starting interfaces (15–17). Despite starting from the same configuration, the contact energies between the GEF domain and RHOA were consistently more favorable than the comparable contact energies for RAC1 (Supplementary Material, Figs S1 and S2). These findings indicate that RHOA likely exhibits a competitive advantage over RAC1 for binding at the GEF site. The contact energies between the residues p.Val211 and p.Val212 of MAP3K1 and RHOA show small to moderate effect (0.00 to −2.00 kcal/mol), whereas RAC1 does not appear to interact with these residues (Supplementary Material, Figs S1 and S2). The alteration of p.Val212Gln may stabilize the binding of both of these proteins by introducing a more favorable polar side chain. Conversely, the RAC1 binding affinity is decreased by both p.Leu189Arg and p.Leu189Pro ($P = 3 \times 10^{-9}$ and $P = 6 \times 10^{-5}$, respectively) and significantly destabilizes (1.15 and 3.89 kcal/mol) the structure of the GEF

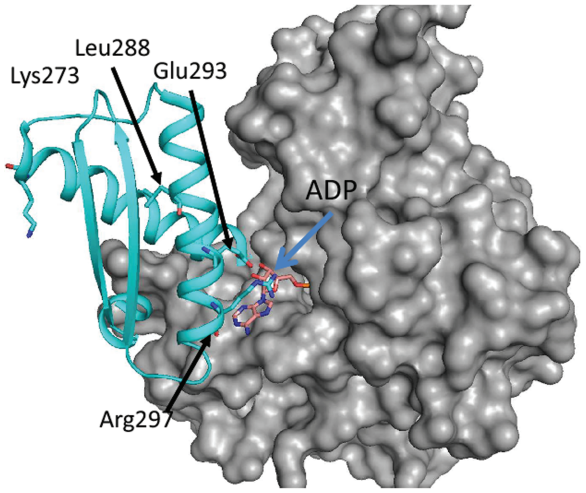

Figure 3. The homologous GEF domain of the HSP70 protein (cyan), the targeted ADP docked to the GEF domain (pink) and the functionally conserved residues, Lys273, Leu288, Glu293, Lys296 and Arg297, between MAP3K1 and HSP70 (sticks) [PDB: 3IUC (18)] and a putative interacting domain within HSP70 that could not be generated in MAP3K1 because of low sequence homology that completes the nucleotide binding pocket (gray surface).

domain (Fig. 2A; Supplementary Material, Fig. S1). The binding affinity of RHOA to MAP3K4 (via AXIN1) is increased for these variants (Fig. 4).

Mutations in the PHD domain. The PHD is a zinc finger domain that is observed in MAP3K1 and absent among the other MAP kinases (25). The PHD domain participates in the ubiquitination of MAP3K (ERK1), MAPK1 (ERK2) and JUN (NM_002228.3) resulting in degradation of the MAP3K (ERK1) and MAPK1 (ERK2) proteins (10,11,26). The mutation, p.Leu447Trp, lies within this domain and disrupts the residue Cys446, an amino acid that coordinates the zinc (Zn) ion, and is involved in binding to the ubiquitin E2 ligase partner (Fig. 5). Folding energy calculations show that p.Leu447Trp should be mildly stabilizing (−0.13). This apparent stabilization is a result of the substitution of a hydrophobic residue (Leu) with a polar residue (Trp) on the water-exposed surface (Supplementary Material, Table S1). However, the folding energy does not account for the impact of the variant on protein–protein binding. Rather, the substitution of a large side chain in p.Leu447Trp into the interface disrupts either the binding of the E2 protein or of the Zn ion to the PHD domain.

Mutations in the ARM domain. ARM repeats are a structural element commonly involved in protein–protein interactions, including those with AXIN1 (27,28). Furthermore, the interaction of ARM-type domains and related GTP proteins, like RHOA, have been observed in multiple structures, such as the binding of importin with RAN [PDB: 1WA5 (29) and 1IBR (30)]. The majority of the pathogenic variants in this region that are reported in this study or in the literature (p.Leu587His, p.Gly616Arg, p.Thr657Arg, p.Cys691Arg and p.Leu764Arg) lie buried or partially buried in the core of the repeats and are weakly to strongly disruptive (0.01–7.60 kcal/mol; Supplementary Material, Table S1). These findings suggest that disruption of the protein fold is the most common pathogenic mechanism. Variant p.Gly616Arg, however, is predicted to lie on one of the loops and may potentially disrupt function by disrupting a protein-binding interface. Conversely, none of the characterized pathogenic variants appear on the surface, where they could disrupt the interface with a binding partner. However, to investigate the binding we generated a putative model of the protein-binding complex based upon the known structures of CTNNB1 (β-catenin) with AXIN1 [PDB: 1QZ7 (28)] (Fig. 6A). The insertion of side chains in the variants introduce steric effects that impact the local folding of the ARM domain, and the larger deletion removes an entire repeat of the ARM domain (Fig. 6B). Putative models of the ARM domain with RAC1 and RHOA were also generated using the RAN-Importin complexes as starting models [PDB: 1WA5 (29) and 1IBR (30)]. Individual studies of these different starting ARM–RHOA interfaces was performed using molecular dynamics and revealed strong contacts (<−5.00 kcal/mol) between RHOA and ARM residues Arg697 to Glu708 and Arg763 to Glu862, respectively. The binding affinity profile of p.Gly727-Ile761del with the target proteins, RAC1, MAP3K4 and RHOA, supports that the presence of the variants impacts the protein function (Fig. 4). Functionally, an increase in the binding affinity of MAP3K1 for RHOA in the p.Gly727-Ile761del mutation ($P = 6 \times 10^{-9}$) was observed and is consistent with removing unfavorable interactions and generating a smaller pocket—allowing RHOA to maintain more favorable contacts simultaneously (Supplementary Material, Fig. S1). Conversely, RAC1 has significant interactions with some of the deleted residues and interacts weakly with the residues that remain after the deletion. Consequently, its binding affinity to mutant p.Gly727-Ile761del MAP3K1 is reduced (Figs 4 and 6; Supplementary Material, Fig. S2).

Mutations in MAP3K1 associated with 46,XY gonadal dysgenesis have differential effects on phosphorylation of downstream targets and protein binding depending on the domains in which they occur

Previously, we showed that mutations in *MAP3K1* associated with 46,XY gonadal dysgenesis result in markedly increased phosphorylation of MAPK11 (p38), MAP3K (ERK1) and MAPK1 (ERK2) (7). Here we show that the magnitude of the phosphorylation, as determined by the ratio of phosphorylation to total protein, varied for mutations in different domains and for MAP3K (ERK1) and MAPK1 (ERK2) reflected a balance between phosphorylation and degradation (Fig. 7). Aside from p.Lys246Glu, none of the benign variants showed differences from the others in these phosphorylation assays. The variant p.Lys246Glu showed a small increase in phosphorylation in MAPK11 (p38), but not MAP3K (ERK1) and MAPK1 (ERK2), which is an effect that we characterized previously as hypomorphic (7). The highest phospho-p38/total p38 was observed for the ARM domain p.Gly727-Ile761del, followed by the N-terminal p.Pro153Leu mutations ($P = 2.6 \times 10^{-8}$ and $P = 2.4 \times 10^{-8}$, respectively). The highest p-ERK/total ERK was observed for the ARM domain p.Gly616Arg and p.Gly727-Ile761del followed by GEF domain p.Leu189Pro mutations ($P = 2.0 \times 10^{-10}$, $P = 3.7 \times 10^{-10}$ and $P = 3.4 \times 10^{-8}$, respectively). We also observed that the amount of total ERK was reduced for all mutants versus benign variants (Supplementary Material, Fig. S3), indicating a possible increase in E2 enzyme activity for these mutations, because E2 enzyme is known to bind to the PHD domain of MAP3K1 that may lead to an increased ubiquitination and degradation of MAP3K (ERK1) and MAPK1 (ERK2) (11). Despite individual domain effects, there was a strong correlation between the phosphorylation of MAPK11(p38) with the phosphorylation MAP3K (ERK1) and

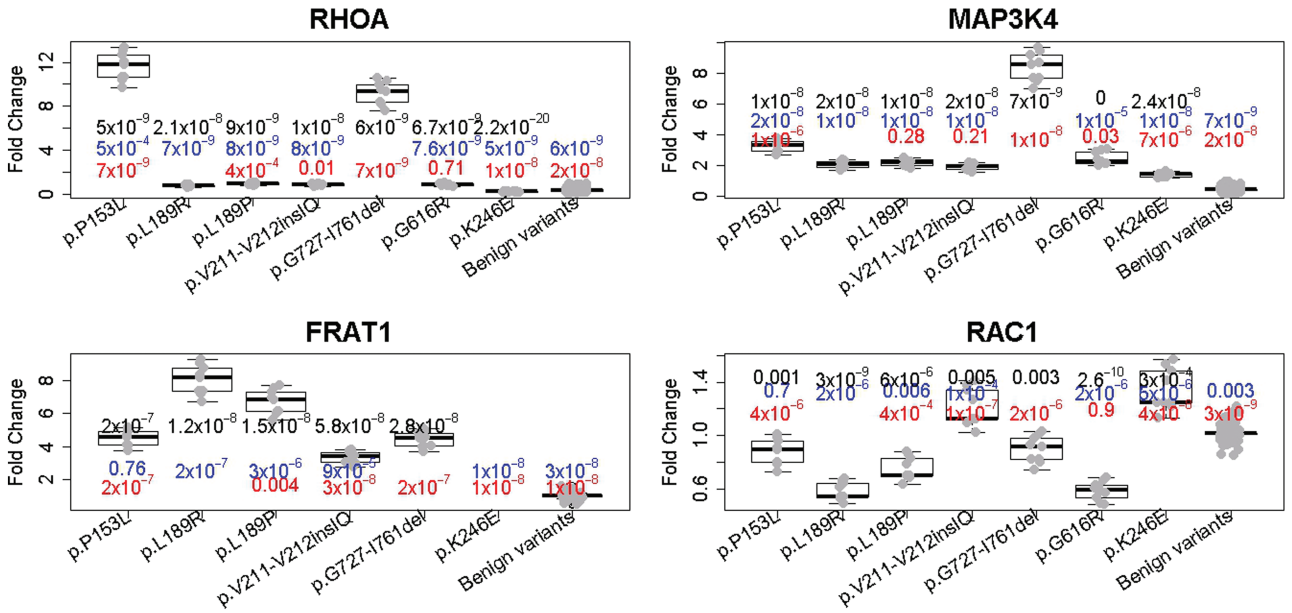

Figure 4. The binding of MAP3K1-interacting proteins varies by mutation site. Boxplots of flow variant, co-immunoprecipitation assays of MAP3K1 for each variant are shown with interacting proteins: (A) RHOA, (B) MAP3K4, (C) FRAT1 and (D) RAC1. The *P*-values above each variant indicate significance of overlap between the variant and p.Leu189Arg (red), p.Gly727-Ile761del (blue) and the benign variants (black) using the Mann–Whitney test.

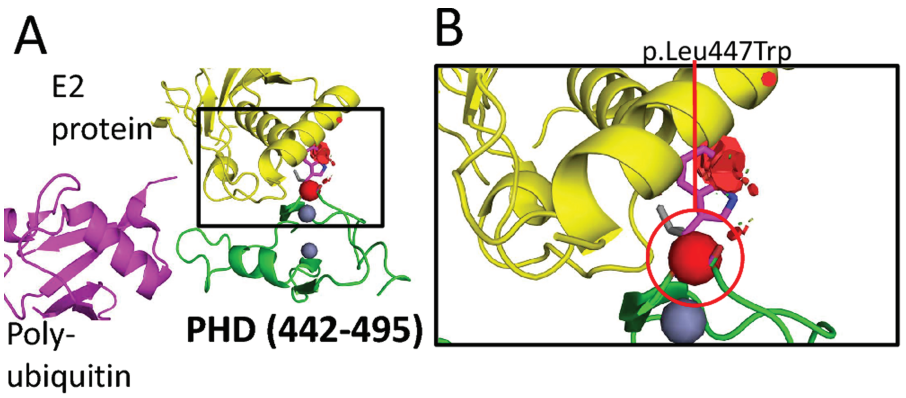

Figure 5. **(A)** The interface of the PHD domain (green) based on homology to the known complex of RFN31-UBE2D2-ubiquitin [PDB: 2LOB (37)] with the partner homologous E2 protein (yellow) and polyubiquitin (magenta) [PDB: 5EDV (39)]. **(B)** The wild-type side chain (gray sticks) and disruption of the interface by the side chains of the variant, p.Leu447Trp, (magenta sticks) and its steric clashes (red dots).

MAPK1 (ERK2) (r^2 = 0.91 and $P = 3.3 \times 10^{-11}$, respectively) (Supplementary Material, Fig. S4).

Mutations in *MAP3K1* also increased binding of the MAP3K4 and FRAT1 proteins

Representative mutations for these domains were analyzed using flow variant analysis (FVA) and compared one to another and to benign variants for their binding affinities to other known protein co-factors, MAP3K4 (via AXIN1) and FRAT1 (Fig. 4) (5). None of the benign variants showed differences from each other in these binding assays. The hypomorphic p.Lys246Glu variant increased binding affinities to MAP3K4. For the others, the binding affinity varied according to the domains in which the mutations occur (Fig. 4). The highest binding of MAP3K4 (via AXIN) to MAP3K1 was observed for the N-terminal p.Pro153Leu and the ARM domain p.Gly727-Ile761del ($P = 1 \times 10^{-8}$ and $P = 7 \times 10^{-9}$, respectively), with significantly higher binding for the p.Gly727-Ile761del mutation compared with the p.Pro153Leu mutation ($P = 2 \times 10^{-8}$). The binding of FRAT1 was the highest for the GEF domain p.Leu189Arg and p.Leu189Pro mutations ($P = 1.2 \times 10^{-8}$ and $P = 1.5 \times 10^{-8}$, respectively). Although FRAT1 binding was increased for the GEF domain p.Val211-Val212insIleGln mutation, the magnitude of this effect was not as great as for the other two mutations in this domain.

Discussion

Germline gain-of-function mutations in the *MAP3K1* gene causing partial or complete gonadal dysgenesis favor the ovarian determining pathway could arise by overriding the testis-determining signal from an expressed wild-type *SRY* gene. The pathogenic mutations discussed in this study are clustered within specific functional domains and increase phosphorylation of MAPK11 (p38), MAP3K (ERK1) and MAPK1 (ERK2) and decrease total MAP3K (ERK1) and MAPK1 (ERK2).

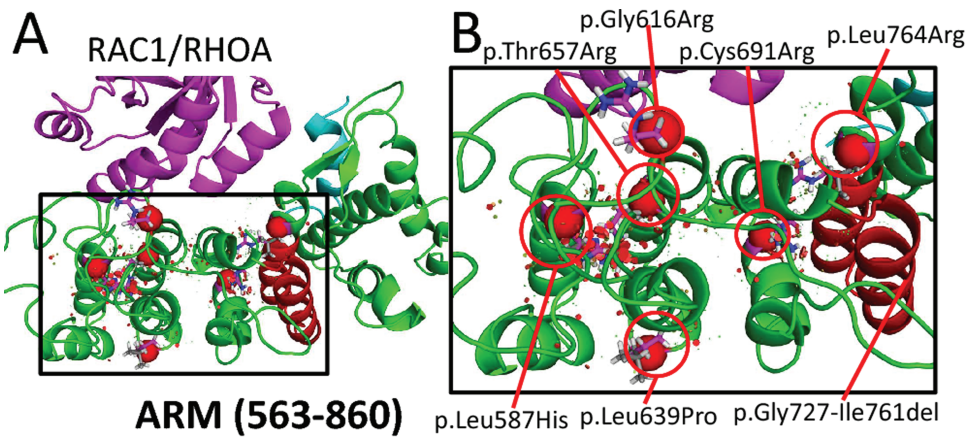

Figure 6. **(A)** The putative binding interface between the ARM domain of MAP3K1 (green) and altered sites (red spheres and red ribbon for the deletion) with RAC1 or RHOA (magenta) and AXIN1 (cyan) based on structural homology to CTNNB1 [β-catenin; PDB: 1QZ7 (28) and PDB: 4Y5J (41)]. **(B)** The wild-type side chains (gray sticks) and disruption by the side chains of the variants, p.Leu764Arg, p.Thr657Arg, p.Leu587His, p.Leu639Pro, p.Cys691Arg and p.Gly727-Ile761del, (magenta, sticks and red coloring on the ARM domain for the large deletion) and their steric clashes (red dots).

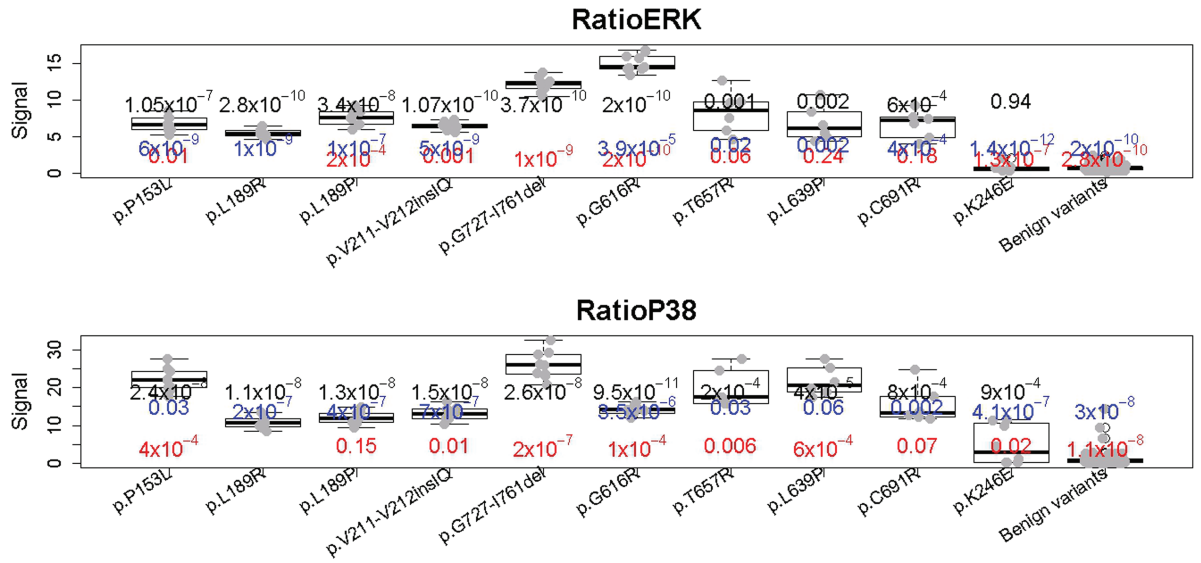

Figure 7. The phosphorylation of MAPK11(p38), MAP3K(ERK1) and MAPK1(ERK2) is increased in MAP3K1 mutations. Boxplots comparing RatioERK (Total phosphorylated MAP3K and MAPK1)/(Total MAP3K and MAPK1) and RatioP38 (Total phosphorylated MAPK11/Total MAPK11) measured by FVA for all MAP3K1 variants. The *P*-values above each variant indicate the degree of overlap between the results of the variant with p.Leu189Arg (red), p.Gly727-Ile761del (blue) and benign variants (black) using the Mann–Whitney test.

They include the p.Gly616Arg variant that has been previously reported as pathogenic on the basis of its co-segregation in a family with 46,XY gonadal dysgenesis (31) and is now confirmed by functional studies. Variants outside these domains do not tend to produce gain-of-function effects. Multiple binding partners are affected by variants in domains, without a specific one-to-one correlation between a domain and a binding partner. Previous truncation mutation analyses mapped the RHOA-binding domain to the N-terminal region of MAP3K1 (32). In the current study, we showed that mutations in the GEF and ARM domains affected binding affinities to both RHOA and RAC1, a partner that is known to counter the effects of RHOA on MAP3K1 (32). The greatest decrease for RAC1 binding was observed for mutations in the GEF suggesting a greater sensitivity or influence of GEF on RAC1 binding than on RHOA binding. MAP3K1 kinase activity represents a balance between the pro-kinase activities of the RHOA and MAP3K4 binding partners and the inhibitory activity of RAC1.

The individual contributions of different competing proteins at the domains of MAP3K1 discussed here best explains the complex binding affinity profiles observed in MAP3K1 (Fig. 4). RAC1, RHOA and AXIN1 are three of the many different potentially competing proteins at these interfaces, with AXIN1 acting as an intermediary for MAP3K4 binding (33,34). At first glance, the binding affinity profiles (Fig. 4) between the ARM domain of MAP3K1 and RAC1, RHOA and MAP3K4 (AXIN1) seem to be explained by the competition between RAC1 and RHOA and MAP3K4 (AXIN1) for a binding site at the surface. The deletion p.Gly727-Ile761del decreases RAC1 binding and increases the binding of RHOA and MAP3K4 (most likely through the binding of AXIN1), which is indicated by the contact energies (Fig. 6). These findings agree with a previous report that showed through truncation studies the binding of RHOA to the ARM domain (32). Conversely, there is a dearth of literature indicating that RAC1 binds the ARM domain. However, molecular dynamics simulations suggest a slightly more complex mechanism for the

pathology of p.Gly727-Ile761del; rather than simply deleting the binding pocket for RAC1, it strengthens the binding of RHOA. Because p.Gly727-Ile761del removes a single repeat from the ARM domain, it generates a more effective pocket for RHOA binding (Fig. 4). It is also known that CTNNB1 (β-catenin) binds through an ARM domain to AXIN1 and that it competes with MAP3K1 for AXIN1 binding (7,28). However, the contact energies of AXIN1 with MAP3K1, which were computed using molecular dynamics, were less clear and more variable than those with RAC1 and RHOA.

Reviewing the affinities of variants in the GEF domain, we find that disruptive variants in this region slightly increase RHOA binding affinity and markedly increase MAP3K4 (AXIN1) binding affinity (Fig. 4). Both the p.Leu189Gln and p.Leu189Pro mutations increase MAP3K4 (AXIN1) and RHOA binding (i.e. near the putative nucleotide binding site of the GEF domain), whereas they decrease the binding of RAC1. This behavior is similar to that observed in the mutation p.Gly727-Ile761del of the ARM1 domain. The p.Val211-Val212insIleGln variant increases the binding of all three proteins, RAC1, RHOA and MAP3K4 (AXIN1). This raises the question of whether RAC1 and RHOA bind at different interfaces near the GEF domain or bind competitively at the same position. The contact energies for RAC1 binding to the GEF domain indicate that the overall binding of MAP3K1 with RHOA is stronger than that with RAC1 at the putative interaction interface. The increased binding of MAP3K1 p.Val211-Val212insIleGln to both RAC1 and RHOA may be because of replacement of Val212 with a polar residue generates favorable contacts with both proteins with the activating effect of RHOA overwhelming the deactivating effect of RAC1 and thus disrupting the testes determining pathway. There was no identifiable homolog for an interaction of AXIN1 with a GEF domain and so it may be that MAP3K4 does not interact with the GEF domain and the increase in affinity is because of a decrease in competitive binding from RAC1 (Supplementary Material, Fig. S2). This hypothesis is supported by the observation that decreases in RAC1 binding correlate with increases in the binding of MAP3K4 (AXIN1) and RHOA (Fig. 4) (7).

These findings show that benign and pathogenic mutations in MAP3K1 observed in cases of 46,XY gonadal dysgenesis tend to be confined to specific regions of the protein, an observation that could be tested by editing these variants into the germline of mice. The net effect of increased binding of the pro-kinase RHOA and MAP3K4 proteins and decreased binding of the inhibitory RAC1 protein result in the gains-of-function observed with these mutations. These structure-functional observations improve understanding about the biological basis for tipping the developmental pathway from testicular to ovarian and facilitate annotation of future mutations observed in the *MAP3K1* genes of individuals with the clinical phenotype of 46,XY gonadal dysgenesis.

Materials and Methods

Analysis of cases

The cases include those newly reported here as well as those reported previously (summarized in Supplementary Material, Table S1). Cell culture, digital cell Western FVA and co-immunoprecipitation Western FVA were performed as described previously (7). In those studies, we showed that we obtained comparable results when testing germline variants in LCLs and transfected NT2/D1 cells. Here the analysis is limited to LCLs. The mean of nine replicates was calculated for each individual sample in each assay and presented by boxplots. Mann–Whitney tests were performed to determine whether the differences in the individual assays between cell lines and groups (controls, all mutants, GEF domain mutants and ARM domain mutants) were different for the several groups. This study was approved by the Albert Einstein College of Medicine Institutional Review Board, and appropriate informed consent was obtained from all human subjects.

Molecular modeling

The sequence of MAP3K1 was taken from the Uniprot database (35). The structures of MAP3K1 were built from homologous structures in the PDB database (36). The GEF homology model was built from protein Q4D059 in *Trypanosoma cruzi* [PDB: 2MNI (14)]. There is strong structural homology between the subdomain IIB in the ATPase domain of HSP70 and this N-terminal domain in MAP3K1, and so a model of the complex was derived from the HSP70-ADP complex [PDB: 3IUC (18)]. The PHD domain is homologous to the zinc finger domain of E3 ubiquitin-protein ligase [PRJ1—PDB: 2L0B (37)] and to Ring Box protein [RBX1––3DPL (38)]. A model of the structure of the PHD domain to potential homologous E2-ubiquitin partners was built using the homologous structure of RFN31-UBE2D2-ubiquitin [PDB: 5EDV (39)]. The ARM domain was identified as an ARM fold by Interpro (40). Investigation of the domain with HHBlits and Psi-BLAST confirmed that there were a number of potential homology models, all of moderate to low homology. The structure was built on homology to MSPS from *Drosophila melanogaster* [PDB: 4Y5J (41)]. While it exhibits only moderate sequence identity, ~30%, the structural characteristics of these regions are well conserved. Structurally, it resembles CTNNB1 (β-catenin) and may compete with CTNNB1 (β-catenin) for AXIN1 binding (7). We built the putative complex using CTNNB1 (β-catenin) complexed with AXIN1 [PDB: 1QZ7 (28)], as well as the putative interface of RHOA with ARM using importin with RAN [PDB: 1WA5 (29) and 1IBR (30)].

An initial search and build of homology models was performed using PHYRE (42). PHYRE performs an extensive search of available homology models using HHBlits (13), followed by fold library scanning informed by PSIPRED (43) and a final round of loop modeling on the remaining regions. Our subsequent investigation and follow-up on the homology models were built using the homology modeling module of ROSETTA (44). The sequence alignments were generated using HHBlits with the pdb70 data set (36) and confirmed using TCOFFEE (45,46). The initial build from the Hidden Markov Modelling (HMM) alignment was generated using a fixed backbone residue substitution with relaxation of the side chains. Second regions in the structural template that correspond to gaps in the query sequence are deleted, and the consequent breaks in the backbone are closed using loop modeling. Third, the gaps in the structural template relative to the query are filled in using loop modeling. Finally, the homology model is checked for any potential errors and breaks in the backbone and corrected with loop modeling. The energy calculations for the destabilization of both internal structure and the protein–protein binding were performed by using the FoldX program (47), and the lower energy bound for pathogenic variants that cause structural disruption of protein folding was 2.50 kcal/mol (23,24). The images and protein structural alignments were generated using PYMOL (The PyMOL Molecular Graphics System, Version 1.8; Schrödinger, LLC).

Molecular mechanical simulations on the protein–protein binding interfaces were performed using 0.15 molar NaCl solutions in water boxes generated using the Visual Molecular Dynamics program (VMD) (48). The simulations were performed using Not Another Molecular Dynamics program (NAMD) (49) with the CHARMM22 force field (50). Periodic boundary conditions were applied using Ewald summation (51). The time simulation was run for 100 ns with an initial 1 ns minimization round, and the steps were updated at 2 fs intervals. Clustering of the resultant trajectories to identify potentially representative configurations were performed using GROMACS (52). Identification and tabulation of residue contact energies was performed using the NAMD Energy module in VMD (48).

Conflict of Interest statement. A.C. and H.-M.L. are full-time employees of Ambry Genetics Corp., a CLIA approved clinical genetics testing laboratory. R.H. is a former full-time employee of Ambry Genetics Corp.

Funding

March of Dimes Birth Defects Foundation (6-FY14-366).

References

1. Ostrer, H. (2014) Disorders of sex development (DSDs): an update. *J. Clin. Endocrinol. Metab.*, **99**, 1503–1509.
2. Lee, P.A., Houk, C.P., Ahmed, S.F., Hughes, I.A. and International Consensus Conference on Intersex (2006) Consensus statement on management of intersex disorders. *Pediatrics*, **118**, e488–e500.
3. Arboleda, V.A., Sandberg, D.E. and Vilain, E. (2014) DSDs: genetics, underlying pathologies and psychosexual differentiation. *Nat. Rev. Endocrinol.*, **10**, 603–615.
4. Eggers, S. and Sinclair, A. (2012) Mammalian sex determination—insights from humans and mice. *Chromosome Res.*, **20**, 215–238.
5. Warr, N., Carre, G.A., Siggers, P., Faleato, J.V., Brixey, R., Pope, M., Bogani, D., Childers, M., Wells, S., Scudamore, C.L. *et al.* (2012) Gadd45gamma and Map3k4 interactions regulate mouse testis determination via p38 MAPK-mediated control of Sry expression. *Dev. Cell*, **23**, 1020–1031.
6. Gierl, M.S., Gruhn, W.H., von Seggern, A., Maltry, N. and Niehrs, C. (2012) GADD45G functions in male sex determination by promoting p38 signaling and Sry expression. *Dev. Cell*, **23**, 1032–1042.
7. Loke, J., Pearlman, A., Radi, O., Zuffardi, O., Giussani, U., Pallotta, R., Camerino, G. and Ostrer, H. (2014) Mutations in MAP3K1 tilt the balance from SOX9/FGF9 to WNT/beta-catenin signaling. *Hum. Mol. Genet.*, **23**, 1073–1083.
8. Loke, J. and Ostrer, H. (2012) Rapidly screening variants of uncertain significance in the MAP3K1 gene for phenotypic effects. *Clin. Genet.*, **81**, 272–277.
9. Granados, A., Alaniz, V.I., Mohnach, L., Barseghyan, H., Vilain, E., Ostrer, H., Quint, E.H., Chen, M. and Keegan, C.E. (2017) MAP3K1-related gonadal dysgenesis: six new cases and review of the literature. *Am. J. Med. Genet. C Semin. Med. Genet.*, **175**, 253–259.
10. Charlaftis, N., Suddason, T., Wu, X., Anwar, S., Karin, M. and Gallagher, E. (2014) The MEKK1 PHD ubiquitinates TAB1 to activate MAPKs in response to cytokines. *EMBO J.*, **33**, 2581–2596.
11. Lu, Z., Xu, S., Joazeiro, C., Cobb, M.H. and Hunter, T. (2002) The PHD domain of MEKK1 acts as an E3 ubiquitin ligase and mediates ubiquitination and degradation of ERK1/2. *Mol. Cell*, **9**, 945–956.
12. Chardin, P., Camonis, J.H., Gale, N.W., van Aelst, L., Schlessinger, J., Wigler, M.H. and Bar-Sagi, D. (1993) Human Sos1: a guanine nucleotide exchange factor for Ras that binds to GRB2. *Science*, **260**, 1338–1343.
13. Remmert, M., Biegert, A., Hauser, A. and Soding, J. (2011) HHblits: lightning-fast iterative protein sequence searching by HMM–HMM alignment. *Nat. Methods*, **9**, 173–175.
14. Lopez-Castilla, A., Pons, T. and Pires, J.R. (2015) NMR structure and dynamics of Q4D059, a kinetoplastid-specific and conserved protein from Trypanosoma cruzi. *J. Struct. Biol.*, **190**, 11–20.
15. Worthylake, D.K., Rossman, K.L. and Sondek, J. (2000) Crystal structure of Rac1 in complex with the guanine nucleotide exchange region of Tiam1. *Nature*, **408**, 682–688.
16. Kulkarni, K., Yang, J., Zhang, Z. and Barford, D. (2011) Multiple factors confer specific Cdc42 and Rac protein activation by dedicator of cytokinesis (DOCK) nucleotide exchange factors. *J. Biol. Chem.*, **286**, 25341–25351.
17. Lucato, C.M., Halls, M.L., Ooms, L.M., Liu, H.J., Mitchell, C.A., Whisstock, J.C. and Ellisdon, A.M. (2015) The phosphatidylinositol (3,4,5)-trisphosphate-dependent rac exchanger 1 ras-related C3 botulinum toxin substrate 1 (P-Rex1.Rac1) complex reveals the basis of Rac1 activation in breast cancer cells. *J. Biol. Chem.*, **290**, 20827–20840.
18. Wisniewska, M., Karlberg, T., Lehtio, L., Johansson, I., Kotenyova, T., Moche, M. and Schuler, H. (2010) Crystal structures of the ATPase domains of four human Hsp70 isoforms: HSPA1L/Hsp70-hom, HSPA2/Hsp70-2, HSPA6/Hsp70B', and HSPA5/BiP/GRP78. *PLoS One*, **5**, e8625.
19. Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B. *et al.* (2016) Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, **536**, 285–291.
20. Fanger, G.R., Johnson, N.L. and Johnson, G.L. (1997) MEK kinases are regulated by EGF and selectively interact with Rac/Cdc42. *EMBO J.*, **16**, 4961–4972.
21. Tong, L. and Tergaonkar, V. (2014) Rho protein GTPases and their interactions with NFκB: crossroads of inflammation and matrix biology. *Biosci. Rep.*, **34**, 283–295.
22. Kyriakis, J.M. and Avruch, J. (2012) Mammalian MAPK signal transduction pathways activated by stress and inflammation: a 10-year update. *Physiol. Rev.*, **92**, 689–737.
23. Bromberg, Y. and Rost, B. (2009) Correlating protein function and stability through the analysis of single amino acid substitutions. *BMC Bioinformatics*, **10** (Suppl. 8), S8.
24. Tokuriki, N., Stricher, F., Schymkowitz, J., Serrano, L. and Tawfik, D.S. (2007) The stability effects of protein mutations appear to be universally distributed. *J. Mol. Biol.*, **369**, 1318–1332.
25. Suddason, T. and Gallagher, E. (2015) A RING to rule them all? Insights into the Map3k1 PHD motif provide a new mechanistic understanding into the diverse roles of Map3k1. *Cell Death Differ.*, **22**, 540–548.
26. Xia, Y., Wang, J., Xu, S., Johnson, G.L., Hunter, T. and Lu, Z. (2007) MEKK1 mediates the ubiquitination and degradation of c-Jun in response to osmotic stress. *Mol. Cell. Biol.*, **27**, 510–517.
27. Groves, M.R. and Barford, D. (1999) Topological characteristics of helical repeat proteins. *Curr. Opin. Struct. Biol.*, **9**, 383–389.
28. Xing, Y., Clements, W.K., Kimelman, D. and Xu, W. (2003) Crystal structure of a beta-catenin/axin complex suggests a

mechanism for the beta-catenin destruction complex. *Genes Dev.*, **17**, 2753–2764.
29. Matsuura, Y. and Stewart, M. (2004) Structural basis for the assembly of a nuclear export complex. *Nature*, **432**, 872–877.
30. Vetter, I.R., Arndt, A., Kutay, U., Gorlich, D. and Wittinghofer, A. (1999) Structural view of the Ran–Importin beta interaction at 2.3 A resolution. *Cell*, **97**, 635–646.
31. Pearlman, A., Loke, J., Le Caignec, C., White, S., Chin, L., Friedman, A., Warr, N., Willan, J., Brauer, D., Farmer, C. *et al.* (2010) Mutations in MAP3K1 cause 46,XY disorders of sex development and implicate a common signal transduction pathway in human testis determination. *Am. J. Hum. Genet.*, **87**, 898–904.
32. Gallagher, E.D., Gutowski, S., Sternweis, P.C. and Cobb, M.H. (2004) RhoA binds to the amino terminus of MEKK1 and regulates its kinase activity. *J. Biol. Chem.*, **279**, 1872–1877.
33. Luo, W., Ng, W.W., Jin, L.H., Ye, Z., Han, J. and Lin, S.C. (2003) Axin utilizes distinct regions for competitive MEKK1 and MEKK4 binding and JNK activation. *J. Biol. Chem.*, **278**, 37451–37458.
34. Zhang, Y., Neo, S.Y., Wang, X., Han, J. and Lin, S.C. (1999) Axin forms a complex with MEKK1 and activates c-Jun NH(2)-terminal kinase/stress-activated protein kinase through domains distinct from Wnt signaling. *J. Biol. Chem.*, **274**, 35247–35254.
35. UniProt Consortium (2013) Update on activities at the Universal Protein Resource (UniProt) in 2013. *Nucleic Acids Res.*, **41**, D43–D47.
36. Bourne, P.E., Addess, K.J., Bluhm, W.F., Chen, L., Deshpande, N., Feng, Z., Fleri, W., Green, R., Merino-Ott, J.C., Townsend-Merino, W. *et al.* (2004) The distribution and query systems of the RCSB Protein Data Bank. *Nucleic Acids Res.*, **32**, D223–D225.
37. Liu, G., Tong, S., Xiao, R., Acton, T.B., Everett, J.K. and Montelione, G.T. (2010) Solution NMR structure of zinc finger domain of E3 ubiquitin-protein ligase praja-1 from Homo sapiens, Northeast Structural Genomics Consortium (NESG) target HR4710B. PDB: 2L0B. (https://www.rcsb.org/structure/2L0B)
38. Duda, D.M., Borg, L.A., Scott, D.C., Hunt, H.W., Hammel, M. and Schulman, B.A. (2008) Structural insights into NEDD8 activation of cullin-RING ligases: conformational control of conjugation. *Cell*, **134**, 995–1006.
39. Lechtenberg, B.C., Rajput, A., Sanishvili, R., Dobaczewska, M.K., Ware, C.F., Mace, P.D. and Riedl, S.J. (2016) Structure of a HOIP/E2~ubiquitin complex reveals RBR E3 ligase mechanism and regulation. *Nature*, **529**, 546–550.
40. Finn, R.D., Attwood, T.K., Babbitt, P.C., Bateman, A., Bork, P., Bridge, A.J., Chang, H.Y., Dosztanyi, Z., El-Gebali, S., Fraser, M. *et al.* (2017) InterPro in 2017-beyond protein family and domain annotations. *Nucleic Acids Res.*, **45**, D190–D199.
41. Howard, A.E., Fox, J.C. and Slep, K.C. (2015) Drosophila melanogaster mini spindles TOG3 utilizes unique structural elements to promote domain stability and maintain a TOG1- and TOG2-like tubulin-binding surface. *J. Biol. Chem.*, **290**, 10149–10162.
42. Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. and Sternberg, M.J. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.*, **10**, 845–858.
43. Jones, D.T. (1999) Protein secondary structure prediction based on position-specific scoring matrices. *J. Mol. Biol.*, **292**, 195–202.
44. Kim, D.E., Chivian, D. and Baker, D. (2004) Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.*, **32**, W526–W531.
45. Erb, I. and Notredame, C. (2016) How should we measure proportionality on relative gene expression data? *Theory Biosci.*, **135**, 21–36.
46. Catuara-Solarz, S., Espinosa-Carrasco, J., Erb, I., Langohr, K., Gonzalez, J.R., Notredame, C. and Dierssen, M. (2016) Combined treatment with environmental enrichment and (−)-epigallocatechin-3-gallate ameliorates learning deficits and hippocampal alterations in a mouse model of Down syndrome. *eNeuro*, **3**, 1–18.
47. Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F. and Serrano, L. (2005) The FoldX web server: an online force field. *Nucleic Acids Res.*, **33**, W382–W388.
48. Humphrey, W., Dalke, A. and Schulten, K. (1996) VMD: visual molecular dynamics. *J. Mol. Graph.*, **14**, 33, 27–38, 38.
49. Phillips, J.C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R.D., Kale, L. and Schulten, K. (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem.*, **26**, 1781–1802.
50. Mackerell, A.D. Jr., Feig, M. and Brooks, C.L. III (2004) Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.*, **25**, 1400–1415.
51. York, D.M., Wlodawer, A., Pedersen, L.G. and Darden, T.A. (1994) Atomic-level accuracy in simulations of large protein crystals. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 8715–8718.
52. Hess, B., Kutzner, C., van der Spoel, D. and Lindahl, E. (2008) GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J. Chem. Theory Comput.*, **4**, 435–447.