PNAS RESEARCH ARTICLE | EVOLUTION

# Evolutionary history of MEK1 illuminates the nature of deleterious mutations

Ekaterina P. Andrianova[a,b], Robert A. Marmion[c], Stanislav Y. Shvartsman[c,d,e,1], and Igor B. Zhulin[a,b,1]

Edited by Benjamin Neel, NYU Langone Health, New York, NY; received March 13, 2023; accepted July 24, 2023

**Mutations in signal transduction pathways lead to various diseases including cancers. MEK1 kinase, encoded by the human *MAP2K1* gene, is one of the central components of the MAPK pathway and more than a hundred somatic mutations in the *MAP2K1* gene were identified in various tumors. Germline mutations deregulating MEK1 also lead to congenital abnormalities, such as the cardiofaciocutaneous syndrome and arteriovenous malformation. Evaluating variants associated with a disease is a challenge, and computational genomic approaches aid in this process. Establishing evolutionary history of a gene improves computational prediction of disease-causing mutations; however, the evolutionary history of MEK1 is not well understood. Here, by revealing a precise evolutionary history of MEK1, we construct a well-defined dataset of MEK1 metazoan orthologs, which provides sufficient depth to distinguish between conserved and variable amino acid positions. We matched known and predicted disease-causing and benign mutations to evolutionary changes observed in corresponding amino acid positions and found that all known and many suspected disease-causing mutations are evolutionarily intolerable. We selected several variants that cannot be unambiguously assessed by automated prediction tools but that are confidently identified as "damaging" by our approach, for experimental validation in *Drosophila*. In all cases, evolutionary intolerant variants caused increased mortality and severe defects in fruit fly embryos confirming their damaging nature. We anticipate that our analysis will serve as a blueprint to help evaluate known and novel missense variants in MEK1 and that our approach will contribute to improving automated tools for disease-associated variant interpretation.**

serine–threonine kinase | damaging mutations | *Drosophila* | variants of unknown significance

## Significance

Genome sequencing has significantly improved diagnosis and treatment of genetic diseases and cancers; however, it produces many variants that cannot be easily interpreted. Automated bioinformatics tools can help predict functional consequences of variants of uncertain significance, but their accuracy is low. Here, by tracing precise evolutionary history of each amino acid position in MEK1 kinase (mutations in which cause neurodegenerative diseases and cancer), we can establish whether variants seen in humans are evolutionarily tolerant. Using published data and experiments, we show that evolutionarily tolerable variants in MEK1 are benign, whereas intolerable substitutions are damaging. Our approach will help in diagnostics of MEK1-associated diseases, it is generalizable to many other disease-associated genes, and it can help improve automated predictors.

In the last few decades, gene sequencing has demonstrated its great potential for diagnostics of inherited disorders and common pathologies in clinical practices. Genetic testing is broadly and routinely utilized by researchers and clinicians to identify genetic variations related to the disease and in many cases helps in choosing the therapeutic strategy. However, only a small fraction of reported variants is tested experimentally to establish causality and most newly discovered coding variations are neither described in other individuals nor studied in cellular or animal models. The growing sequencing data have led to massive accumulation of variants of uncertain significance (VUS): Currently, over 500,000 of reported variants are classified as a VUS in ClinVar, the largest public archive on human genetic variation and phenotypes (1). The lack of experimental data on VUS forces clinicians to rely on predictions using computational tools, such as PolyPhen-2 (2) and SIFT (3); however, the accuracy of these tools is relatively low. For example, a study comparing predictions for mutation outcome made by SIFT and PolyPhen-2 with the existing experimental data, found that false negatives comprised 43% and 33%, respectively (4). Similarly, a large comparative study assessed the performance of ten prediction tools and found that specificities of variant interpretation were 63 to 67% for SIFT and 73 to 75% for PolyPhen-2 (5). In some cases, well-studied pathogenic mutations were falsely declared to be benign by both tools (6). One of the possible reasons for the erroneous predictions is the inability of the automated algorithms to exclude functionally unrelated homologs from the multiple sequence alignment (MSA) used for variant interpretation. The majority of disease genes have been duplicated in their evolutionary history, and the role of paralogs in disease has been a subject of debate (7). For example, some studies suggested that in most cases, only one copy of the gene is associated with a disease (8), whereas others proposed that disease genes have functionally redundant paralogs (9). Therefore, functional roles of paralogs should be taken into account for variant interpretation. Previously, we have proposed an approach for improving variant prediction for genes from paralogous families, where only one copy of a gene is associated with a disease (6). As a test case, we used the human NPC1 gene, mutations in which lead to a devastating disorder—the Niemann-Pick type C disease (10). We specifically selected this gene because mutations in this gene cause severe abnormalities and death (often in childhood). More than 150 NPC1 disease–causative mutations

Author contributions: E.P.A., S.Y.S., and I.B.Z. designed research; E.P.A. and R.A.M. performed research; E.P.A., R.A.M., S.Y.S., and I.B.Z. analyzed data; and E.P.A., R.A.M., S.Y.S., and IB.Z. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.



[1]To whom correspondence may be addressed. Email: stas@princeton.edu or jouline.1@osu.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2304184120/-/DCSupplemental.

Published August 14, 2023.

are reported in the literature, and more than 20 NPC1 alleles found in high frequencies in adult human populations are likely to be benign. By contrast, inactivating mutations in the NPC1 paralog NPC1L1 do not cause a disease (11). We showed that removal of NPC1L1 and distant homologs from consideration improves variant interpretation. We assessed sensitivity, specificity, false discovery rate, accuracy, F1 score, and Matthews correlation coefficient in comparison of our approach with automated tools PolyPhen-2, SIFT, and PROVEAN (6). We achieved ~10% improvement in sensitivity, while causing a relatively low cost in specificity, and our approach outperformed other tools in terms of the overall quality, as measured by the Matthews correlation coefficient (6). We attribute such drastic improvement in sensitivity to the fact that our approach eliminates the false evolutionary signals introduced by functionally diverged paralogs and remote homologs that are included in the analysis by other tools. For example, all three automated tools predicted four known NPC1 pathogenic mutations as benign, most likely because the same substitutions are found in NPC1L1 sequences that are included in their MSA sets (6). Thus, we argue that understanding the gene's history and selecting only well-defined orthologs for variant interpretation is a critical step in improving evolutionary-based prediction methods for genes with functionally nonredundant paralogs.

The mitogen-activated protein kinase kinase 1 (MKK1, MEK1 or MAP2K1) is a member of the MEK paralogous group within the vast kinase family and the centerpiece of Ras/MAPK pathway, which plays a vital role in the regulation of cell growth and proliferation. Dysregulation of the Ras/MAPK cascade through activating mutations results in the abnormal signaling, which consequently leads to tumorigenesis and developmental disorders (12–14). Early approaches to simultaneous sequencing of multiple genes of Ras/MAPK signaling pathways have yielded hundreds of mutations in MEK1 possibly associated with RASopathies and cancer. For example, germline MEK1 mutations have been reported in patients with cardiofaciocutaneous (CFC) syndrome (15–17), and somatic mutations have been identified in melanoma (18–20), lung cancer (21, 22), gastric cancer (23), colon carcinoma (24), ovarian cancer (25), hairy cell leukemia (26), Langerhans cell histiocytosis (LCH), and the non-LCH (27–29). Due to the central role of MEK1 in the Ras/MAPK pathway signaling, it is a common target for highly selective inhibitors, such as Cobimetinib and Trametinib, that are used in cancer treatment (30–34). MEK inhibition has shown to be very successful in treating various cancers and is now also gaining recognition for treating RASopathies (35).

Seven MEK paralogs are encoded in the human genome (Fig. 1). A previous study identified duplication events and the relationship between the MEK paralogs (36); however, due to a small number of genomes available at the time, an updated analysis of this family is needed. MEK paralogs function in different pathways and have different roles in disease. For example, MEK1 interacts with Raf and ERK, whereas MEK3 interacts with MLK and p38. Aberrations in these pathways lead to different phenotypes and diseases. MEK1 is associated with RASopathies (12, 13), and melanoma (18, 19), whereas MEK3 has been implicated in inflammatory diseases such as rheumatoid arthritis (37) and also in acute lymphoblastic leukemia (38). Thus, we should exclude MEK3 from consideration, when assessing amino acid residues specific to MEK1 function. Even when paralogs participate in the same pathway, their settings can be different, leading to different phenotypes upon their malfunction. Both MEK1 and MEK2 (its most recent paralog) act within the same pathway and interact with the same upstream and downstream partners. However, MEK2 loss can be compensated for by MEK1, but not vice versa (39). Thus, even these close paralogs are not equal contributors, and MEK2 should be excluded from the analysis of MEK1. Here, we reveal a precise and comprehensive evolutionary history of the MEK family and confidently identify MEK1 orthologs in hundreds of metazoan genomes. We generated their curated MSAs suitable for interpretation of genetic variants in humans. This allowed us to demonstrate that all known and the vast majority of suspected disease-causing mutations in MEK1 are evolutionarily intolerable. Finally, we selected several variants that could not be unambiguously interpreted by automated prediction tools for experimental validation and show that computationally predicted evolutionary

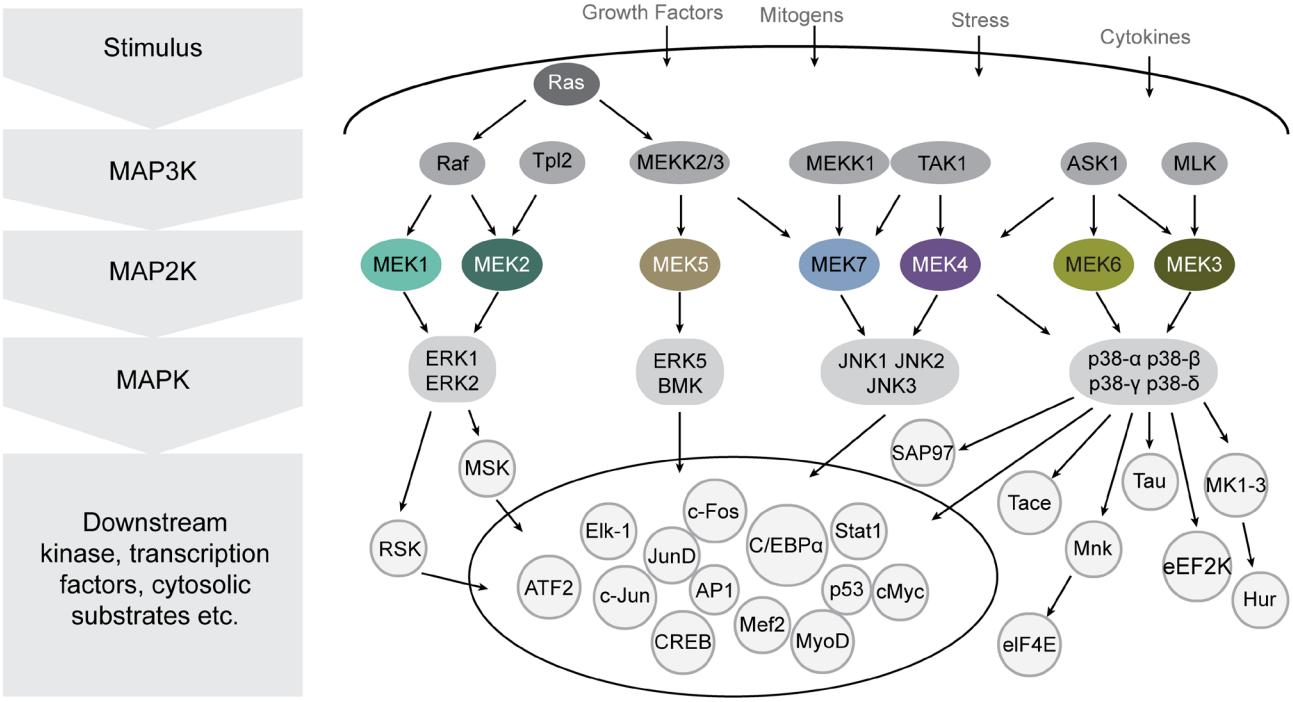


**Fig. 1.** Central role of MAP2K kinases in the MAPK/ERK pathway. Arrows indicate known interactions. Adapted from ref. 40.

intolerant variants cause increased lethality and severe defects in fruit fly embryos. Our computational evolutionary approach is applicable for interpretation of VUS in MEK1 and many other human genes, especially those from large paralogous families.

## Results

**Origins and Evolution of the MEK Family.** To reveal the precise evolutionary history of MEK, we first identified all MEK homologs and their closest relatives within the previously defined STE group of protein kinases (41). Available precomputed alignments of MEK-related homologs are either too broad (contain paralogs and distant homologs) or too narrow (contain too few orthologs). For example, Pfam and COG models cover the entire protein kinase superfamily, whereas the NCBI Conserved Domain database MEK1 alignment contains only a few sequences from vertebrates. Consequently, we used BLAST searches initiated with human MEK sequences against the selected genome dataset representing all major Eukarya lineages. All sequences collected in these searches were used to generate an MSA that was edited to correct misaligned regions, and the final MSA contained only regions corresponding to the kinase domain (Dataset S1). The alignment was then used for phylogenetic reconstruction by maximum likelihood and neighbor-joining methods (Fig. 2 and *SI Appendix*, Figs. S1 and S2). Both approaches yielded trees with similar topology, indicating the reliability of phylogenetic inference.

The phylogenetic reconstruction revealed three distinct clusters in the representative genome dataset (Fig. 2). The first cluster contained human MST1 (STK3), MST2 (STK4), MST4 (STK26), and YSK1 (STK25) kinases and had at least one sequence from each representative genome from our set. All seven human MEK sequences were found in the second cluster (termed MEK), which contained sequences only from Obazoa. The third cluster (termed MKK) contained sequences from all other major eukaryotic supergroups, including known MKK kinases from *Arabidopsis thaliana*. This cluster was not supported by appreciable bootstrap values; however, the Cluster of Orthologous Groups approach suggested that sequences from MKK cluster are more related to MEK1 than to any other MEK or MST sequences. Within the MEK cluster, sequences were further branching out to form three clades, each containing one or more human MEK proteins, in a good agreement with the early analysis (36). Clade I contained sequences from all Opisthokonta and included human MEK1 and MEK2. Clade II contained sequences from Metazoa, Choanoflagellida, and Filasteria and included human MEK5. Larger clade III was almost exclusively composed of sequences from Metazoa and included human MEK4, MEK7, MEK3, and MEK6. The bootstrap values supporting sequences from Choanoflagellida and Filasteria in clade III were weak and, therefore, to validate their orthologous relationships, we employed the Cluster of Orthologous Groups approach and assessed sequence and domain synapomorphies. As a result, we assigned each sequence to a respective MEK group and annotated each of the protein sequences from the representative genome set according to the *H. sapiens* MEK nomenclature. Phyletic distribution of MEK homologs (Fig. 3) also suggests that MEK1 is the ancestor of the entire MEK family, which is further supported by the fact that the MEK1 clade within Opisthokonta has the shortest average brunch length from the root (Fig. 2), indicating the least divergence, which is typical of the ancestor. Based on these observations and the presence of MEK homologs in Viridiplantae, Chromalveolata, Excavata, and Rhizaria, we further suggest that MEK1 was present in the last eukaryotic common ancestor and that its separation from MST-type serine/threonine kinases occurred prior to the emergence of major eukaryotic supergroups.

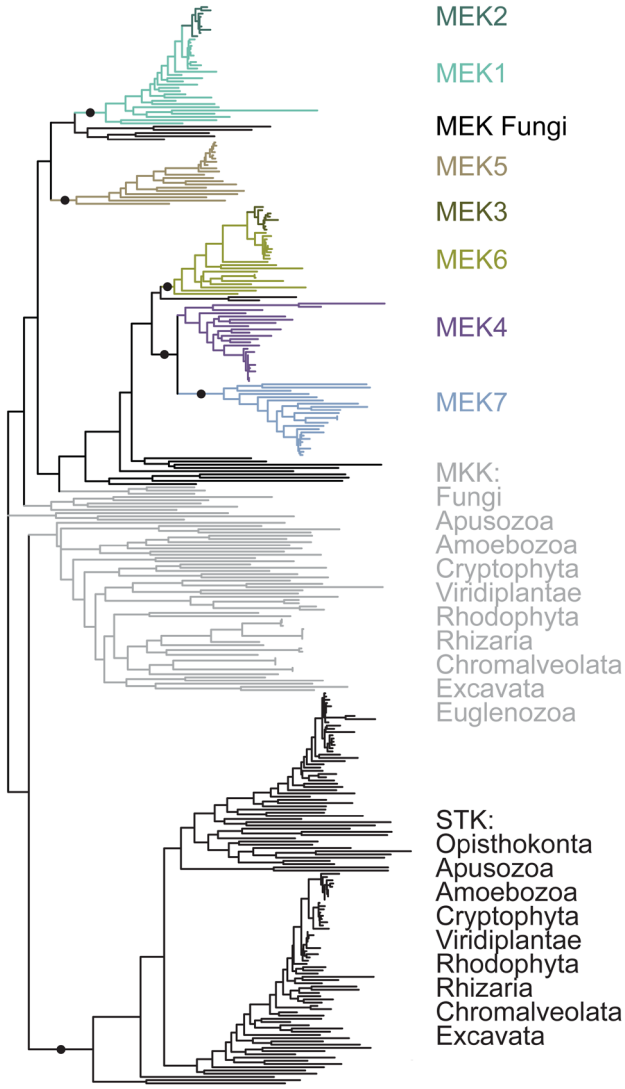


**Fig. 2.** Maximum likelihood phylogenetic tree of the MEK protein family and their closest homologs from the representative set of eukaryotic species. Bootstrap support values >50% are indicated as circles.

Several major duplication events have contributed to diversification of the MEK family in the lineage leading to humans. The first two duplication events resulting in the emergence of MEK5 and the common ancestor of MEK4/7/6/3 most likely have occurred in the last common ancestor of Holozoa, as both *Monosiga brevicollis* and *Capsaspora owczarzaki* have homologs that cluster with the corresponding clades. Another major duplication event has occurred at the root of Metazoan lineage, giving rise to MEK6, MEK7, and MEK4. Finally, with the emergence of vertebrates, the MEK family experienced an expansion again, with MEK2 and MEK3 diverging from MEK1 and MEK6, respectively.

**Building a Comprehensive Set of Well-Defined MEK1 Orthologs.** The list of MEK1 orthologs generated from the representative set of Eukarya genomes was used as a starting point and a reference, with the idea of its expansion by "filling the blanks," e.g., identifying MEK1 in genomes that were not represented in this dataset. We limited our searches to Metazoan genomes for several reasons. First, our phylogenetic analysis demonstrated that we could confidently distinguish MEK1 orthologs from paralogs in Metazoa. Second, more than two hundred metazoan genomes were available for

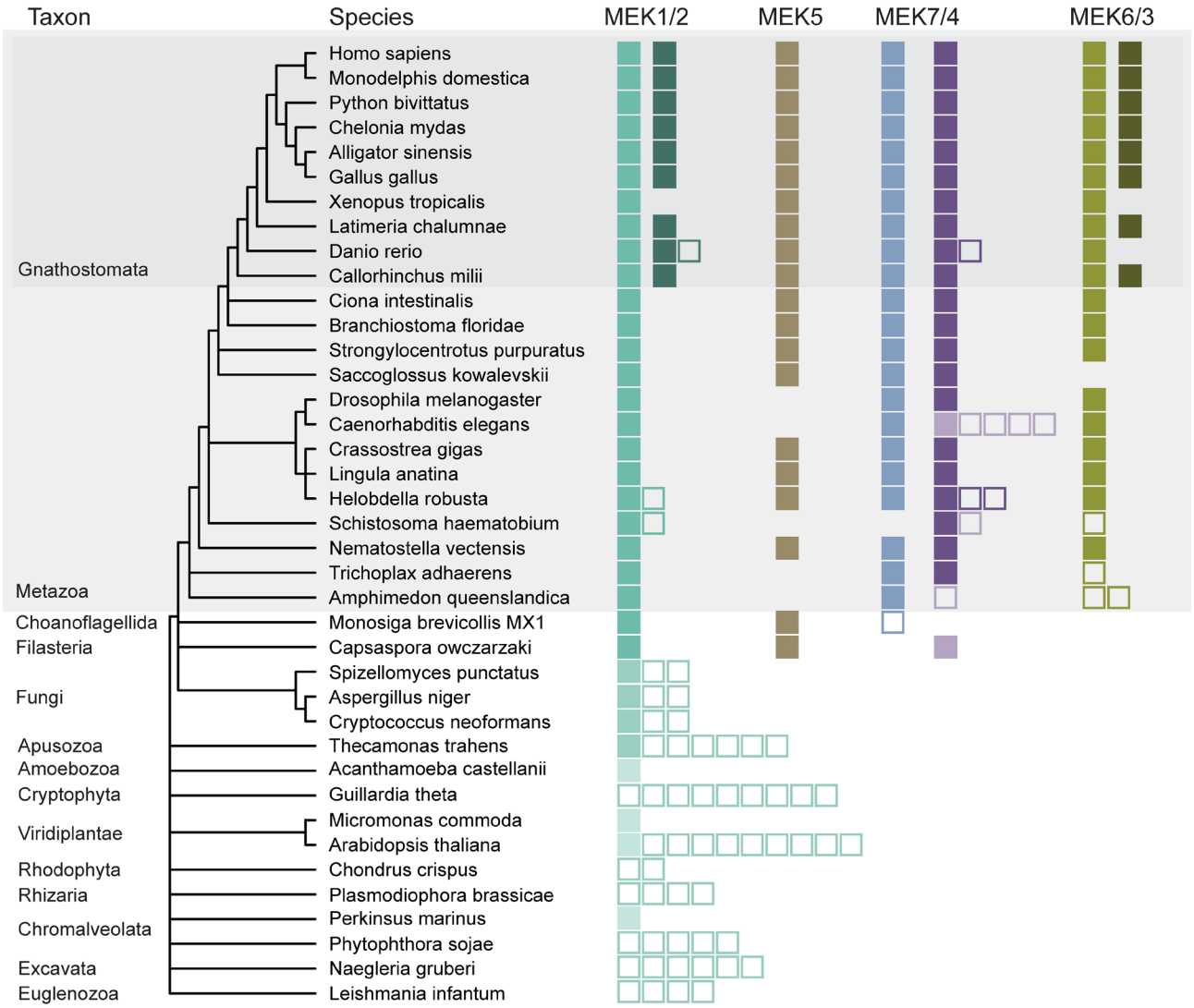


**Fig. 3.** Evolutionary history of the MEK family across the major eukaryotic lineages. Schematic eukaryotic tree of life was extracted from NCBI Taxonomy Browser. Each member of MEK family (MEK1 through MEK7) is indicated by a different color. Filled squares represent confidently assigned orthologs. Empty squares represent paralogs (placed next to the filled square of corresponding orthologs) or proteins that could not be confidently assigned to a specific MEK class.

analysis providing substantial depth. Finally, estimated time since Metazoan emergence, 635 Mya (42) should result in substantial sequence diversity, which is essential for productive sequence analysis and variant interpretation. To generate a comprehensive list of MEK1 orthologs, we searched all metazoan genomes available in RefSeq by BLAST using the human MEK1 sequence as a query. Best BLAST hits from each genome were collected and combined with MEK1 orthologs from the representative genome set generated in the previous step (Metazoa only). MEK2, MEK5, MEK3, MEK6, MEK4, and MEK7 sequences from the same representative set of metazoan genomes were added to this dataset to be used as outgroups in a downstream phylogenetic analysis and all full-length sequences were aligned. Neighbor-joining and maximum likelihood phylogenetic trees were built from the resulting MSA. Based on the phylogenetic inference, outgroups containing MEK2 through MEK7 and several paralogs were discarded, thus resulting in a master MSA containing more than 300 well-defined MEK1 orthologs from Metazoa (Dataset S2).

**Dataset of Confidently Assigned MEK1 Orthologs Provides a Necessary and Sufficient Evolutionary Depth.** Functionally important positions in proteins are conserved, whereas multiple substitutions observed in a group of orthologs indicate that protein tolerates changes in a given position without losing its function. An informative MSA should contain both conserved and variable positions, indicating that a given set of sequences has a sufficient evolutionary depth. To test our MEK1 ortholog MSA (Dataset S2), we analyzed 323 amino acid positions of the MEK1. We found that 66 amino acid positions (20% of the total analyzed residues) were invariable, indicating their critical importance (Dataset S3). Because no changes in these positions were tolerated during more than 600 Mya of MEK1 evolutionary history, any substitution in invariable positions should be considered deleterious. Among invariable positions, 27 were specifically conserved in MEK1 orthologs (variable in other MEK family members): These positions likely define MEK1-specific functions, such as interaction with its cognate partners within the MAPK signaling pathway. We also identified 28 positions, where a single substitution event has occurred (either in a single organism or in a common ancestor of a closely related group). Considering possibilities of a sequence error or a compensatory mutation, confident predictions for such positions should not be made. Finally, most positions (~70%) were variable (>2 independent substitution events during the evolutionary history). Fig. 4 shows examples of both invariable and highly variable positions in MEK1. For example, valine in position 82 is invariable in the entire dataset of more than 300 sequences,


**Fig. 4.** Conserved and variable positions in MEK1 proteins from various species. Small subset of more than 300 MAP2K1 orthologs is shown. Positions highlighted in black are invariable in the entire dataset of MEK1 orthologs. Variable positions are shown in teal.

but the neighboring position 83 (phenylalanine in human MEK1) is highly variable: It can be occupied not only by a large hydrophobic residue, such as tryptophan and leucine, but also by small polar residues, such as serine, threonine, and asparagine. Hence, we conclude that orthologs had enough time to diverge, while preserving key positions, and therefore the generated MEK1 ortholog MSA has necessary and sufficient evolutionary depth to evaluate consequences of missense mutations in the human MEK1 gene. For example, any change in position V82 in humans (including changes to any similar aliphatic amino acid) would be classified as evolutionarily intolerable and thus damaging. On the other hand, mutation F83S would be classified as tolerable and thus likely benign, even though a large aromatic residue is substituted with a small polar one.

**Functionally Important Residues Are Conserved in MEK1 Ortholog MSA.** Functionally important amino acid residues in proteins, such as catalytic sites, are well conserved. Thus, to validate our MEK1 ortholog MSA, we examined conservation of functionally characterized amino acid residues in MEK1 that are available in the literature (*SI Appendix*, Table S1). As expected, amino acid residues that are required for the kinase catalytic activity or define its key structural properties are conserved not only in the MEK1 ortholog MSA (Dataset S2), but also in the entire MEK homolog MSA (Dataset S1). For example, catalytic core residues K97, D190, and D208 (43), ATP-phosphate binding loop residues G75, G77, and G80 22177953 (43), and catalytic loop residues R189 and K192 (43) are invariable in both datasets. Amino acid residues comprising protein–protein interaction interfaces are usually conserved, but not as strictly as active or catalytic sites. It is worth noting that L314, which participates in MEK1-BRAF and MEK1-KSR1 complex formation (44, 45) is invariable in MEK1 orthologs, thus highlighting its functional importance. Another functionally important residue, E138, forms a salt bridge with BRAF R462 in the heterodimer complex (45). We observe alternations in this position in some MEK1 orthologs (Dataset S2): however, all changes are to polar residues only (aspartate, asparagine, and serine). Thus, interaction between these MEK1 and BRAF residues is preserved: either as a salt bridge or as a weaker hydrogen bond. Taken together, these observations further support the suitability of our MEK1 ortholog MSA for the variant interpretation.

**Known Pathogenic Mutations in MEK1 Are Evolutionarily Intolerable.** We further validated our MEK1 ortholog MSA against available experimental and clinical data. All known disease-causing MEK1 mutations were identified in the literature and are listed in Table 1. Satisfactorily, we found that all experimentally confirmed pathogenic mutations were predicted as such by our approach (Table 1). For example, in position 53, only one type of substitution (F53W, found exclusively in Nematoda) is observed in our MEK1 ortholog MSA (Dataset S2). Therefore, all other substitutions, including well-documented F53L and F53S mutations (15, 19) are evolutionarily intolerable. Another aromatic residue, Y130, had changed multiple times during MEK1 evolution, but the type of substitution was always the same—Y130F (Dataset S2). Thus, we can conclude that all other types of substitutions in this position, including a well-characterized, disease-causing Y130C mutation (15) are intolerable. Evolutionarily intolerable substitutions reduce fitness and thus should be defined as "damaging," "deleterious," and when the change in protein function leads to a disease, "disease-causing."

**Likely Benign Variants in MEK1 Are Evolutionarily Tolerable.** MEK1 alleles found in high frequencies in human populations (very recent evolutionary changes) are likely to be benign. Therefore, to use it as a negative control for our interpretation of missense mutations, we compiled a list of MEK1 missense variants with frequency >5 (Table 2) from more than 140,000 human exomes and genomes available in the Genome Aggregation Database (gnomAD, ref. 56). Remarkably, all these variants were found to be evolutionarily tolerable and therefore predicted as likely benign. For example, the A390T variant is found at high frequency in human cohorts, and we identified the same variant in other mammals (*Odobenus rosmarus divergens* and *Sorex araneus*). Furthermore, other substitutions for small amino acids in this position can be seen throughout the evolutionary history of MEK1 (e.g., valine, serine, glycine), reinforcing the notion that the observed A390T variant in humans is likely benign.

**MEK1 Missense Mutations Associated with RASopathies and Cancer Are Evolutionarily Intolerable.** Extensive search through COSMIC and ClinVar databases allowed us to retrieve all missense mutations in MEK1 that were reported to be associated with cancer and/or RASopathy. After excluding repetitive reports, we identified a total of 256 potentially pathogenic variants in the MEK1 protein (*SI Appendix*, Table S2). Among more than a hundred of disease-associated variants, seventeen are found in positions that are invariable (100% identity throughout the MEK1 MSA): Any change in any of these positions is evolutionarily intolerable and therefore likely to substantially alter MEK1 function. For example, two variants, G128D and G128V that are described in various cancers (57–60) occur in a position, where glycine is 100% conserved. Thus, not only the variants reported so far, but any other substitution of glycine in this position would be damaging. Many disease-associated mutations are found in positions that variable, but still well conserved. For example, R181M (leads to a loss of a positive charge) variant was described in thyroid carcinoma (61). Our analysis revealed that the only change in position 181 seen during MEK1 evolutionary history is R181K substitution, which maintains the charge and polarity. Thus, any change other than R181K should be classified as damaging.

**Table 1. Experimentally confirmed disease-causing missense mutations in MEK1 are evolutionarily intolerable**

| Mutation | Zygosity | Experimental data | Association | Evolutionary tolerance* |
|---|---|---|---|---|
| L42F | Heterozygous | Phenotypic changes of zebrafish embryos (46) | CFC (17) | No |
| R47Q | Homozygous | Stimulated ERK phosphorylation, transforming ability (23) | Cancer (23) | No |
| R49L | Heterozygous | Stimulated ERK phosphorylation, transforming ability (23) | Cancer (23) | No |
| F53S | Heterozygous | Stimulated ERK phosphorylation (15, 47)phenotypic changes of zebrafish embryos (46) | CFC (15) | No |
| F53L | Heterozygous | Stimulated ERK phosphorylation (19); phenotypic changes of zebrafish embryos (46) | Cancer (19) | No |
| T55P | ND | Stimulated ERK phosphorylation (47); phenotypic changes of zebrafish embryos (46) | CFC (48) | No |
| K57N | Heterozygous | Stimulated ERK phosphorylation (47)transforming ability (49) | Cancer (49) | No |
| D67N | Heterozygous | Stimulated ERK phosphorylation (25, 47); phenotypic changes of zebrafish embryos (46) | Cancer (25) | No |
| C121S | | Stimulated ERK phosphorylation (47, 50) | Cancer (50) | No |
| P124L | ND | Phenotypic changes of zebrafish embryos (46) | CFC (51) | No |
| P124Q | Heterozygous | Phenotypic changes of zebrafish embryos (46) | Costello syndrome (52) | No |
| P124S | Heterozygous | Stimulated ERK phosphorylation (19) | Cancer (19) | No |
| G128V | ND | Phenotypic changes of zebrafish embryos (46) | CFC (53) | No |
| Y130C | Heterozygous | Stimulated ERK phosphorylation (15, 47); phenotypic changes of zebrafish embryos (46) | CFC (15) | No |
| Y130H | Heterozygous | Phenotypic changes of zebrafish embryos (46) | CFC (17) | No |
| Y130N | ND | Phenotypic changes of zebrafish embryos (46) | CFC (54) | No |
| E203K | Heterozygous/ homozygous | Stimulated ERK phosphorylation (19, 47); phenotypic changes of zebrafish embryos (46) | Cancer (19) | No |
| E203Q | ND | Phenotypic changes of zebrafish embryos (46) | CFC (55) | No |
| I204T | Heterozygous | Stimulated ERK phosphorylation, transforming ability (23) | Cancer (23) | No |

*Evolutionary tolerance is assessed as the presence of a given variant in MEK1 ortholog MSA.

De novo substitution of leucine 42 to phenylalanine was found in an individual with a cardifaciocutaneous syndrome (17). Mutation L42F was also found in a desmoplastic melanoma sample (62). Position 42 is a part of nuclear export sequence (NES), and during evolution, it endured multiple substitutions from leucine to isoleucine and methionine (Dataset S2) presenting strong evidence that corresponding variants are likely benign in humans. Although phenylalanine is also a large hydrophobic residue, L42F substitution

**Table 2. High frequency MEK1 variants found in human cohorts**

| Variant | gnomAD version | Allele count | Allele frequency | Evolutionary changes* | Evolutionary tolerance |
|---|---|---|---|---|---|
| A19G | v3 | 8 | 5.6E-05 | Gly, >2 events | Yes |
| V93I | v2.1.1 | 11 | 3.9E-05 | Ile, >2 events | Yes |
| A283V | v3 | 40 | 2,8E-04 | Val, >2 events | Yes |
| P321S | v2.1.1 | 14 | 4.9E-05 | Ser, 1 event; multiple events of change to other polar aa | Yes |
| N345S | v2.1.1 | 7 | 2.8E-05 | Ser, 1 event; multiple events of change to other polar aa | Yes |
| A347T | v2.1.1 | 7 | 2.8E-05 | Thr, >2 events | Yes |
| G354H | v2.1.1 | 11 | 3.9E-05 | His, 1 event; multiple events of change to other aa | Yes |
| L375I | v2.1.1 | 11 | 4.4E-05 | Ile, >2 events | Yes |
| G380S | v2.1.1 | 5 | 2.0E-05 | Ser, 1 event; multiple events of change to other polar aa | Yes |
| A390T | v2.1.1 | 19 | 6.7E-05 | Thr, >2 events | Yes |
| V393I | v2.1.1 | 8 | 2.8E-05 | Ile, >2 events | Yes |

*Variants observed in MEK1 ortholog MSA. Substitutions that are seen in more than 2 events or 1 event and multiple changes to other amino acids at the same position are considered tolerable according to ref. 6.

is not seen in the evolutionary history of MEK1 and thus is likely damaging. Variants Q164K and Q164E were found in patients with CFCS (63) and cancer (64), respectively. Position 164 manifested greater variability. Specifically, Gln 164 had changed multiple times to other polar amino acids such as glutamic and aspartic acids, lysine, arginine, asparagine, and serine. Interestingly, MEK1 proteins in frogs and snakes contain lysine. Additionally, substitutions to a non-polar proline were found in several sequences from different clades, and substitutions to hydrophobic leucine, isoleucine, and valine were found once each. Thus, hypervariability of this position and the fact that the same substitutions are tolerated in other species suggest that Q164K and Q164E substitutions are likely benign.

Mutation T55P is associated with various RASopathies (48, 63). This position is hypervariable in the MEK1 ortholog set. Multiple changes to other small and large, polar and hydrophobic amino acids in this position occurred throughout MEK1 evolution. However, not a single MEK1 ortholog contains proline in this position. Position T55 is located within N-terminal helix A and T55P substitution likely leads to a structural damage as prolines cause kinks in alpha helices.

We compared results of our evolutionary analysis of cancer mutations with predictions by popular automated bioinformatics tools PolyPhen-2 (2) and SIFT (3). In four cases, our results disagree, and at least in two cases, we present additional evidence favoring our interpretation of the variants. Variants R47G, R47Q, P105S, and I103S were predicted to be benign by both automated tools (*SI Appendix*, Fig. S3), whereas our analysis suggested that all these variants are damaging, as no such replacements were found throughout the history of MEK1 (Dataset S2).

One of the extensively studied variants, F53L, is predicted as "tolerated" by SIFT. Over the last decade, it has been found in multiple cancer tissues (21, 26, 59, 65). F53L has been demonstrated to constitutively activate MEK1, indicated by the increased ERK phosphorylation in cell cultures (21, 66–68), and to severely affect embryonic lethality in *Drosophila* (46, 67). The result of F53L interpretation using our approach is in agreement with experimental data as it is found to be evolutionarily intolerable.

PolyPhen-2 prediction for variant R47G is "benign", with a score of 0.001. Our analysis suggests that it is "damaging" because no such substitution can be found in our MEK1 ortholog dataset (Dataset S2). Supporting our interpretation, variant R47G was found in two patients with Langerhans cell histiocytosis (69). Furthermore, position 47 is a part of helix A, formed by residues 43 to 61. Interaction of helix A with the kinase domain negatively regulates MEK1 kinase activity (70), thus mutations that disrupt this interaction likely result in constitutive ERK pathway activation. Substitutions in positions R47 and adjacent positions (48 and 49) were shown to affect MEK1 function (71). Substitution with alanine led to a sevenfold activation, and insertion of prolines yielded a constitutively active MEK1 (71). The behavior of these substitution mutants was attributed to α-helix disruption (70, 71). Thus, both clinical observation and experimental evidence suggest that R47G mutation is disease causing. PolyPhen-2 and SIFT predictions for variant R47Q are "benign" (with a score of 0.436) and "tolerated," respectively. No such substitution can be found in our dataset (Dataset S2), and therefore, we interpret it as damaging. Supporting our interpretation, the R47Q variant was first detected in breast cancer (invasive ductal carcinoma) cell line (23), and it was later identified in Langerhans cell histiocytosis (69, 72). The transforming ability of R47Q was confirmed by using a focus formation assay with mouse 3T3 fibroblasts, and activation potential was evaluated by the increase in the phosphorylation level of direct downstream kinases ERK1/2 (23). Similarly, our interpretation of variant P105S as "damaging" differs from that by automated tools: "benign" by PolyPhen-2, with a score of 0.418, and "tolerated" by SIFT. This variant was reported in patients with Langerhans cell histiocytosis (69), favoring our interpretation. It should be noted that evolutionary intolerance strongly suggests but does not confirm deleterious and disease-causing nature of mutations. Ultimately, experimental validation is necessary to validate these predictions.

**Experimental Validation of Selected Variants of Unknown Significance.** To validate our approach experimentally, we selected three variants of unknown significance—F53Y, Q56P, and G128D—for which there was no clear consensus between ten most widely used automated prediction tools (Fig. 5). The F53Y variant was reported as a rare melanoma mutation (73). The G128D variant was found in Langerhans cell histiocytosis (27). Q56P was originally found in rat fibroblasts as well as in human gastric and lung cancers (23, 49, 74). All three variants were identified as "evolutionary intolerable" and therefore "damaging" by our approach, although their conservation patterns were quite different. While position 128 is 100% invariable, position 53 had one change event to a similar aromatic amino acid, and position 56 is variable.

First, the selected mutations were tested using a misexpression assay in *Drosophila* embryos, as previously described (46). Previously, we found that mutations that cause lethality of *Drosophila* embryos belong to the gain-of-function type mutations (46). Gain-of-function mutations allow for signaling, even in the absence of upstream pathway activation, as well as a more rapid response to activation by RAF even in the absence of necessary adapter proteins (82–84). All three mutations significantly increased embryonic lethality (Fig. 6*A*), which indicated that all three variants, F53Y, Q56P, and G128D, are leading to MEK1 activation.

Next, we thoroughly examined cuticles of the wild-type embryos and embryos carrying mutated MEK1. Normal embryonic cuticles have a segmented trunk with 8 denticle belts, together with head and tail segments. In the first RAS signaling event of embryogenesis, the activated ligand Trunk is only present in the termini of the embryo, carving out the head and tail segments. We used misexpression by GAL4/UAS to drive mutant MEK1 expression in the early embryo. Mutant embryos without GAL4 had wild-type cuticles (Fig. 6 *B*, *Top* and Fig. 6*C*). Embryos with all three mutations in MEK1 developed with erased trunk fates (i.e., denticle belts) in the middle of the cuticle, an area which normally lacks RAS signaling. Additionally, the early occurrence of ectopic signaling was leading to the induction of a negative feedback loop, which reduced signaling in the termini, leading to a loss of head structures (Fig. 6 *B*, *Bottom* and Fig. 6 *D*–*F*). The same phenotype was previously reported for other gain-of function

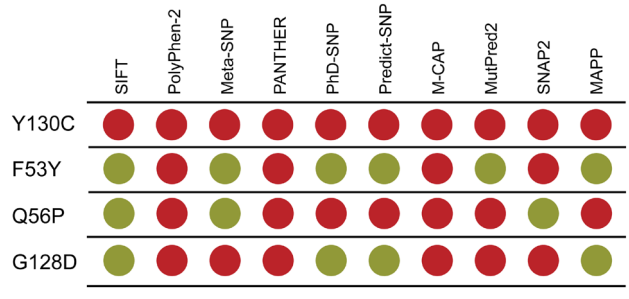


**Fig. 5.** Prediction of functional consequences for selected MEK1 variants of unknown significance by different automated tools (2, 3, 75–81). The green circle represents "benign" and red circle represents "pathogenic."

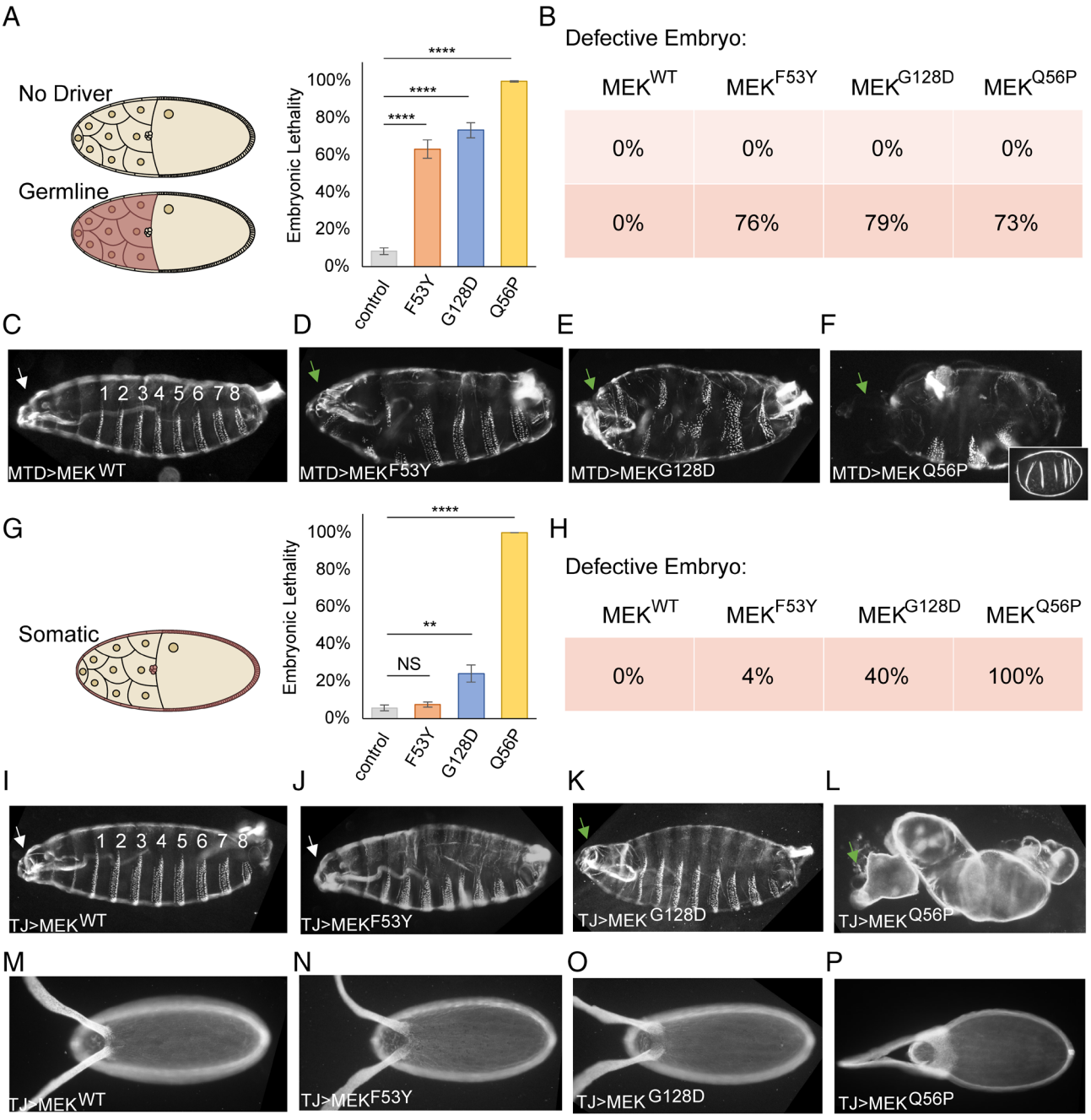


**Fig. 6.** Activating mutations F53Y, G128D, and Q56P in MEK1 cause developmental defects. (*A*) Lethality of all three mutations is increased when expressed under control of maternal driver (MTD-gal4). The control is overexpression of $mek^{WT}$ (n = 1,476). $mek^{F53Y}$ (n = 2,596); $mek^{G128D}$ (n = 1,742); $mek^{Q56P}$ (n = 727). (*B*) Defects in embryo patterning are absent without the Gal4 (*Top*). Defects in embryo patterning are frequent in all three mutants, when under control of a maternal germline driver (MTD-Gal4) (*Bottom*, red). (*C*) Normal embryonic cuticle with the 8 denticle belts shown as well as a normal head (white arrow). $mek^{WT}$ is phenotypically normal (n = 46). (*D–F*) Representative embryo cuticles show defects in the maternal patterning system. Head pattering is affected by all three mutants (green arrow). (*D*) $mek^{F53Y}$ cuticles have disrupted heads as well as fused or missing denticle belts (n = 83). (*E*) $mek^{G128D}$ cuticles have missing heads as well as fused or missing denticle belts (n = 24). (*F*) $mek^{Q56P}$ cuticles have missing heads as well as fused or missing denticle belts as well as a failure to retract the germband (n = 61). Additionally, 61% of embryos fail to form a cuticle (*Inset*). (*G*) Lethality of all three mutations is increased when expressed under control of the somatic follicle cell driver (TJ-gal4). The control is overexpression of $mek^{WT}$ (n = 2,606). $mek^{F53Y}$ (n = 1,167); $mek^{G128D}$ (n = 1,351); $mek^{Q56P}$ (n = 223). (*H*) A graded effect is seen in the embryo cuticles of the three mutants, when under control of a maternal somatic driver (TJ-Gal4) (red). (*I*) Normally, embryos have 8 segments, visualized by the presence of 8 denticle belts on the ventral side of the embryo. $mek^{WT}$ is phenotypically normal (n = 13). (*J–L*) Severity increases between the three mutants, which are dorsalized in a graded manner. Dorsalized eggshells result in dorsalized embryos with malformed heads (green arrows), as evidenced in the mutants $mek^{G128D}$ and $mek^{Q56P}$. (*J*) $mek^{F53Y}$ is phenotypically normal (n = 48). (*K*) $mek^{G128D}$ is weakly dorsalized (n = 43). (*L*) $mek^{Q56P}$ is very strongly dorsalized and has no detectable denticle belts (n = 30). (*M*) Normal eggshells show normal spacing of the respiratory filaments (n = 18). (*N–P*) Dorsalized embryos are caused by dorsalized eggshells. (*N*) F53Y eggshells are phenotypically normal (n = 14). (*O*) G128D eggshells are weakly dorsalized, as evidenced by the increase in the spacing between the two respiratory appendages (n = 12). (*P*) Q56P eggshells are more strongly dorsalized (n = 16).

MEK1 mutants (82, 83). Notably, ~60% of embryos expressing Q56P MEK1 failed to form any cuticle.

Next, we turned to the maternal patterning of the eggshell/egg. This patterning event sets up the dorso-ventral (DV) axis of the eggshell and embryo via EGFR signaling in somatic cells (85–87). The DV patterning of the eggshell is a sensitive system to assay defects in RAS signaling (88). Somatic cells with MEK1 gain-of-function mutants are expected to develop a dorsalized eggshell/embryo.

Expression of mutant MEK1 under control of a somatic driver significantly increased embryonic lethality (Fig. 6*G*). The fraction of abnormal cuticles and lethality agreed with the severity of embryo dorsalization. (Fig. 6 *H*–*L*). Somatic expression of MEK1 Q56P variant resulted in cuticles that are strongly dorsalized, phenocopying embryos that are null for dorsal protein (89). The eggshell patterning by these mutants should also be disrupted. Normally, the respiratory tubes on the dorsal side of the eggshell are positioned by an intermediate level of RAS signaling, established by EGFR ligand localized to dorso-anterior and result in a precisely patterned eggshell (Fig. 6*M*) (88). Increased RAS signaling forces these tubes to be positioned further apart in a graded manner (86). The MEK1 mutants created eggshells that are increasingly dorsalized by mutants of greater strength (Fig. 6 *N*–*P*). Taken together, the expression of F53Y, Q56P, and G128D MEK1 mutants in two different tissues shows that these variants are pathogenic with different levels of severity.

## Discussion

The ever-increasing number of genome sequencing studies revealed dozens of variations in MEK1 sequence and many more are likely to be discovered. However, only a small portion of these mutations has been proven to affect MEK1 activity, thus leading to developmental diseases and cancers, whereas most variants remain VUS. The functional consequences of mutations may be verified experimentally, for example, by testing the kinase phosphorylation activity in vitro (66) or analyzing morphological effects in zebrafish or fruit flies (46). These methods present indisputable advantages; however, they are not feasible for the analysis of hundreds of genetic variants. Facing these challenges, researchers and clinicians usually turn to automated prediction tools, such as PolyPhen-2 (2), SIFT (3), and others. Automated sequences searches employed in these tools and other similar methods cannot identify possible duplication events in the gene history; hence, MSAs that are used for variant interpretation usually include both orthologs and paralogs of the gene of interest, which leads to false negatives, when paralogs have distinct functions and different roles in disease.

The presence of seven human MEK proteins increases the chances of including paralogs into the analysis, and inclusion of even one or two sequences of functionally unrelated paralogs may lead to erroneous interpretation (6) To avoid this, we have established a precise evolutionary history of the MEK family and identified all duplication events. A phylogenetic study of the MEK family conducted in 1999 (36) determined fundamental relationships between the family members and suggested that its diversification in fungal and animal lineages involved a series of gene duplication events. However, due to a small number of sequences available at that time, this analysis was limited in scope and a detailed evolutionary history of this pathway remained unexplored. In the past decade, the number of sequenced eukaryotic genomes increased dramatically; however, their distribution is not uniform: Some lineages are well represented, but for many others, representation is sparse. Such imbalance could easily lead to a biased perspective on evolutionary history. Therefore, for our initial analysis, we have selected 39 well-assembled high-quality genomes from as many eukaryotic lineages as possible. The resulting phyletic distribution of MEK homologs across these 39 organisms (Fig. 3) agreed with the previous finding that MEK duplications occurred independently in metazoa and fungi (36), and it shows that various independent MEK duplications happened in all other major eukaryotic supergroups. Based on these results, only metazoan organisms were included into the final dataset for variant interpretation. Subsequently, we collected all available MEK1 orthologs from Metazoa. constructed their MSA and proved that it had a sufficient evolutionary depth for an adequate analysis of observed variants. Next, we collected the data on known functionally and structurally important amino acid residues and showed they are conserved in our dataset. Finally, we showed that according to our MSA, all experimentally tested disease-causing mutations (positive control) are evolutionarily intolerable, whereas all high-frequency MEK1 variants seen in human cohorts (negative control) are evolutionarily tolerable. Thus, we concluded that the resulting MEK1 MSA (Dataset S2) was of high quality and suitable for interpretation of VUS.

Using the MEK1 MSA, we analyzed all missense mutations in MEK1 reported in COSMIC and ClinVar databases and confidently assigned most of them as "tolerable" or "intolerable" based on their presence or absence in the MEK1 MSA. Upon comparison of our conclusions with predictions made by SIFT and PolyPhen-2, we found many conflicting interpretations. For example, several mutations that were assigned as intolerable or uncertain by our approach were predicted to be benign by the automated tools. For example, even some of the most extensively studied variants, F53L and K57E (46, 66), which are firmly established to cause major MEK1 function alterations, are predicted as tolerable by SIFT. We analyzed the MEK1 MSA used by PolyPhen-2 and detected not only paralogous sequences (namely MEK2, MEK6, and MEK4), but also unrelated kinases from fungi, that carried queried variants and thus led to erroneous interpretation. Based on our analysis, we selected F53Y, Q56P, and G128D MEK1 mutations to test them experimentally using Drosophila embryos. These three mutations were found in different cancer tissues (27, 73, 74). By our approach each of these mutations was concluded to be evolutionarily intolerable and therefore damaging; however, automated tools failed to generate a consensus for the outcome (Fig. 5). The misexpression array suggests that all three substitutions lead to the constitutive activation of MEK1 kinase, and because of the MEK1 activity, the number of the lethal or defective embryos increased significantly in each case. Although the severity of each mutation is different, with Q56P being the most severe one, the data obtained from two different patterning systems expressing F53Y, Q56P, and G128D MEK1 mutants demonstrated that all these variants were pathogenic.

We could not reliably interpret the evolutionary tolerability for a small set of MEK1 variants (~13%), due to their rare occurrence in available metazoan genomes. This apparent weakness of our approach will be diminished as more genomes become available. It is also important to stress that many variants are weak in their impact and evolutionary fitness, requiring evaluation at a higher level of resolution (90). On the other hand, the ambiguity of current interpretations for such mutations makes them a perfect target for experimental validation. Nature had millions of years to experiment with MEK1, and results of many successful substitutions are recorded in the genomes of the living species. Here, we demonstrate that by careful analysis of these data, we can improve assessment of novel variants in MEK1 and, potentially, in many other human proteins, especially those from large paralogous families that are common to signal transduction. Our approach requires careful selection of orthologous sequences and avoiding any paralogs, which presents a known problem for automated methods (7). Although our approach cannot be fully automated at this time, we hope that results obtained with different types of paralogous families will lead to generalizations that can be implemented in automated tools for disease-associated variant interpretation.

## Materials and Methods

**Defining Evolution of the MEK Family.** To identify closely related proteins, human MEK1 protein sequence (NP_002746.1) was used as query search against the human genome using BLASTP tool. Human MEK1 protein sequence was then queried against a subset of individual genomes of Eukaryote species in NCBI Refseq or nonredundant protein databases (species are listed on Fig. 3). The $10^{-35}$ E-value cutoff and coverage of more than 75% parameters were used for each of the separate BLASTP searches. All collected sequences were used to generate a MSA using MAFFT v7 L-INS-i algorithm (91). The MSA was manually trimmed to include only the kinase domain. Domain composition was revealed using CDvist tool (92). Partial and redundant sequences were removed from the alignment. Sequence alignment editing was performed in Jalview (93). We used two approaches for phylogenetic relationship reconstruction, maximum likelihood and neighbor joining methods with 1,000 bootstrap replications. Phylogenetic inference was conducted using with MEGAX package (94). Phylogenetic trees were analyzed and visualized using iTOL tool (95).

**Orthology Assignment and MEK1 Orthologs Master Set.** Orthologs and paralogs were distinguished using several approaches i) by identifying reciprocal best BLAST hits between protein sequences of two organisms; ii) domain architecture of the group of proteins; iii) and confirmed by the presence of monophyletic clades of each group of the orthologs in the phylogenetic trees.

For the broader panel of MEK1 sequences, human MEK1 protein sequence was used to query the RefSeq database of metazoan species using BLAST. Retrieved sequences were aligned with MAFFT v7.154b E-INS-i (91). MEK2, MEK5, MEK3/6, and MEK4/7 sequences from representative species set, assigned earlier, were added to the analysis as outgroups. Neighbor joining and maximum likelihood phylogenetic trees were generated with 500 bootstrap replications with MEGAX (94). The outgroups on the tree were not considered to be MEK1 true orthologs and were discarded from the master MSA of MEK1 ortholog sequences set.

**Interpretation of Variants.** ClinVar, COSMIC (96), and literature were searched to identify missense mutations that were reported to be associated with cancer or any type of RASopathy. The Genome Aggregation Database (gnomAD, ref. 56) database was searched to identify MEK1 variants that occur frequently in human populations.

To assign the significance of each of the mutations, we used SAVER algorithm (6). Briefly, this approach focuses on i) making a set of sequences containing only functional orthologs, which requires the establishment of evolutionary history of a protein, and ii) counting how many times the substitution occurred independently during evolutionary history of the protein, rather than utilizing the abundance of the variant in MSA, to neutralize bias in the alignment created by the inclusion of sequences that are very closely related to each other.

**Experimental Procedures.**

***Fly stocks.*** Fly stocks were maintained under standard conditions and crosses were performed at 25 °C. MTD-GAL4, UAS-MEK$^{WT}$, and TJ-GAL4 flies were described previously (46, 97, 98). cDNA for Dsor1, the Drosophila homologue of MEK, was cloned into an intermediate vector, pMiniT, and mutated by PCR to generate the disease variant alleles (New England Biolabs #E0554S). The utilized mutagenic primers are as follows: F53Y- ATCAAGATGTaCCTCAGCCAGAAGG, G128D-CACATTGTCGaTTTCTACGGC, and Q56P- TTCCTCAGCCcGAAGGAGAAG. The UAS-MEK variants were cloned into pTIGER and integrated into the attP40 site using the ΦC31-based integration system as described previously (82).

***Cuticle phenotyping.*** Embryos were dechorionated with 50% bleach after being aged for 22 h as previously reported (83, 99). Dechorionated embryos were shaken in methanol and heptane (1:1) and incubated overnight in a media containing lactic acid and Hoyer's media (1:1) at 65 °C. Embryos were imaged in darkfield on a Nikon Eclipse Ni.

***Embryonic lethality.*** Lethality was determined by setting up egg-lay cages for a period allowing for ~300 eggs to be oviposited onto grape juice agar. Plates were aged for greater than 30 h, and lethality was determined as the fraction of unhatched eggs. Replicates were conducted to allow for survey of over 1,000 eggs for each genotype.

**Data, Materials, and Software Availability.** All study data are included in the article and/or supporting information.

**ACKNOWLEDGMENTS.** This work was supported by NIH grants R35GM131760 (to I.B.Z.) and R01GM086537, R01GM141843, and R01GM134204 (to S.Y.S.). We thank Daniel Caffrey for helpful comments on the manuscript.

Author affiliations: [a]Department of Microbiology, The Ohio State University, Columbus, OH 43210; [b]Translational Data Analytics Institute, The Ohio State University, Columbus, OH 43210; [c]The Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, NJ 08544; [d]Department of Molecular Biology, Princeton University, Princeton, NJ 08544; and [e]Flatiron Institute, Simons Foundation, New York, NY 10010

1. A. Henrie *et al.*, ClinVar Miner: Demonstrating utility of a Web-based tool for viewing and filtering ClinVar data. *Hum. Mutat.* **39**, 1051–1060 (2018).
2. I. A. Adzhubei *et al.*, A method and server for predicting damaging missense mutations. *Nat. Methods* **7**, 248–249 (2010).
3. P. C. Ng, S. Henikoff, SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).
4. Y. M. Di, E. Chan, M. Q. Wei, J. P. Liu, S. F. Zhou, Prediction of deleterious non-synonymous single-nucleotide polymorphisms of human uridine diphosphate glucuronosyltransferase genes. *AAPS J.* **11**, 469–480 (2009).
5. A. Niroula, M. Vihinen, How good are pathogenicity predictors in detecting benign variants? *PLoS Comput. Biol.* **15**, e1006481 (2019).
6. O. Adebali, A. O. Reznik, D. S. Ory, I. B. Zhulin, Establishing the precise evolutionary history of a gene improves prediction of disease-causing missense mutations. *Genet Med.* **18**, 1029–1036 (2016).
7. Y. Moreau, L. C. Tranchevent, Computational tools for prioritizing candidate genes: Boosting disease gene discovery. *Nat. Rev.* **13**, 523–536 (2012).
8. J. E. Dickerson, D. L. Robertson, On the origins of Mendelian disease genes in man: The impact of gene duplication. *Mol. Biol. Evol.* **29**, 61–69 (2012).
9. W. H. Chen, X. M. Zhao, V. van Noort, P. Bork, Human monogenic disease genes have frequently functionally redundant paralogs. *PLoS Comput. Biol.* **9**, e1003073 (2013).
10. M. T. Vanier, Niemann-Pick disease type C. *Orphanet J. Rare Dis.* **5**, 16 (2010).
11. N. O. Stitziel *et al.*, Myocardial Infarction Genetics Consortium Investigators, Inactivating mutations in NPC1L1 and protection from coronary heart disease. *N. Engl. J. Med.* **371**, 2072–2082 (2014).
12. K. A. Rauen, The RASopathies. *Annu. Rev. Genomics Hum. Genet.* **14**, 355–369 (2013).
13. P. J. Smits, D. J. Konczyk, C. L. Sudduth, J. A. Goss, A. K. Greene, Endothelial MAP2K1 mutations in arteriovenous malformation activate the RAS/MAPK pathway. *Biochem. Biophys. Res. Commun.* **529**, 450–454 (2020).
14. L. B. Do Prado, C. Han, S. P. Oh, H. Su, Recent advances in basic research for brain arteriovenous malformation. *Int. J. Mol. Sci.* **20**, 5324 (2019).
15. P. Rodriguez-Viciana *et al.*, Germline mutations in genes within the MAPK pathway cause cardio-facio-cutaneous syndrome. *Science* **311**, 1287–1290 (2006).
16. Y. Aoki, T. Niihori, Y. Narumi, S. Kure, Y. Matsubara, The RAS/MAPK syndromes: Novel roles of the RAS pathway in human genetic disorders. *Hum. Mutat.* **29**, 992–1006 (2008).
17. M. L. Dentici *et al.*, Spectrum of MEK1 and MEK2 gene mutations in cardio-facio-cutaneous syndrome and genotype-phenotype correlations. *Eur. J. Hum. Genet.* **17**, 733–740 (2009).
18. K. Trunzer *et al.*, Pharmacodynamic effects and mechanisms of resistance to vemurafenib in patients with metastatic melanoma. *J. Clin. Oncol.* **31**, 1767–1774 (2013).
19. S. I. Nikolaev *et al.*, Exome sequencing identifies recurrent somatic MAP2K1 and MAP2K2 mutations in melanoma. *Nat. Genet* **44**, 133–139 (2011).
20. H. Shi *et al.*, Preexisting MEK1 exon 3 mutations in V600E/KBRAF melanomas do not confer resistance to BRAF inhibitors. *Cancer Discov.* **2**, 414–424 (2012).
21. M. E. Arcila *et al.*, MAP2K1 (MEK1) mutations define a distinct subset of lung adenocarcinoma associated with smoking. *Clin. Cancer Res.* **21**, 1935–1943 (2015).
22. M. Nicos *et al.*, Sensitive methods for screening of the MEK1 gene mutations in patients with central nervous system metastases of non-small cell lung cancer. *Clin. Transl. Oncol.* **18**, 1039–1043 (2016).
23. Y. L. Choi *et al.*, Oncogenic MAP2K1 mutations in human epithelial tumors. *Carcinogenesis* **33**, 956–961 (2012).
24. A. K. Murugan, J. Dong, J. Xie, M. Xing, MEK1 mutations, but not ERK2 mutations, occur in melanomas and colon carcinomas, but none in thyroid carcinomas. *Cell Cycle* **8**, 2122–2124 (2009).
25. A. L. Estep, C. Palmer, F. McCormick, K. A. Rauen, Mutation analysis of BRAF, MEK1 and MEK2 in 15 ovarian cancer cell lines: implications for therapy. *PLoS One* **2**, e1279 (2007).
26. J. J. Waterfall *et al.*, High prevalence of MAP2K1 mutations in variant and IGHV4-34-expressing hairy-cell leukemias. *Nat. Genet.* **46**, 8–10 (2014).
27. D. S. Nelson *et al.*, MAP2K1 and MAP3K1 mutations in Langerhans cell histiocytosis. *Genes Chromosomes Cancer* **54**, 361–368 (2015).
28. R. Chakraborty *et al.*, Mutually exclusive recurrent somatic mutations in MAP2K1 and BRAF support a central role for ERK activation in LCH pathogenesis. *Blood* **124**, 3007–3015 (2014).
29. E. L. Diamond *et al.*, Diverse and targetable kinase alterations drive histiocytic neoplasms. *Cancer Discov.* **6**, 154–165 (2016).
30. L. Musacchio *et al.*, MEK inhibitor as single agent in low grade serous ovarian and peritoneal cancer: A systematic review and meta-analysis. *Cancer Treat Rev.* **110**, 102458 (2022).
31. D. M. Gershenson *et al.*, Trametinib versus standard of care in patients with recurrent low-grade serous ovarian cancer (GOG 281/LOGS): An international, randomised, open-label, multicentre, phase 2/3 trial. *Lancet* **399**, 541–553 (2022).
32. D. Talloa *et al.*, BRAF and MEK targeted therapies in pediatric central nervous system tumors. *Cancers (Basel)* **14**, 4264 (2022).

33. A. Harder, MEK inhibitors–Novel targeted therapies of neurofibromatosis associated benign and malignant lesions. *Biomark Res.* **9**, 26 (2021).
34. S. Pasquali, A. V. Hadjinicolaou, V. Chiarion Sileni, C. R. Rossi, S. Mocellin, Systemic treatments for metastatic cutaneous melanoma. *Cochrane Database Syst. Rev.* **2**, CD011123 (2018).
35. C. Bergqvist, P. Wolkenstein, MEK inhibitors in RASopathies. *Curr. Opin. Oncol.* **33**, 110–119 (2021).
36. D. R. Caffrey, L. A. O'Neill, D. C. Shields, The evolution of the MAP kinase pathways: Coduplication of interacting proteins leads to new signaling cascades. *J. Mol. Evol.* **49**, 567–582 (1999).
37. M. Chabaud-Riou, G. S. Firestein, Expression and activation of mitogen-activated protein kinase kinases-3 and -6 in rheumatoid arthritis. *Am. J. Pathol.* **164**, 177–184 (2004).
38. Y. M. Ayala-Marin, A. H. Grant, G. Rodriguez, R. A. Kirken, Quadruple and truncated MEK3 mutants identified from acute lymphoblastic leukemia promote degradation and enhance proliferation. *Int. J. Mol. Sci.* **22**, 12210 (2021).
39. J. L. Bromberg-White, N. J. Andersen, N. S. Duesbery, MEK genomics in development and disease. *Brief. Funct. Genomics* **11**, 300–310 (2012).
40. G. Sabio, R. J. Davis, TNF and MAP kinase signalling pathways. *Semin. Immunol.* **26**, 237–245 (2014).
41. G. Manning, D. B. Whyte, R. Martinez, T. Hunter, S. Sudarsanam, The protein kinase complement of the human genome. *Science* **298**, 1912–1934 (2002).
42. A. Sebe-Pedros, B. M. Degnan, I. Ruiz-Trillo, The origin of Metazoa: A unicellular perspective. *Nat. Rev.* **18**, 498–512 (2017).
43. R. Roskoski Jr., MEK1/2 dual-specificity protein kinases: Structure and regulation. *Biochem. Biophys. Res. Commun.* **417**, 5–10 (2012).
44. D. F. Brennan *et al.*, A Raf-induced allosteric transition of KSR stimulates phosphorylation of MEK. *Nature* **472**, 366–369 (2011).
45. J. R. Haling *et al.*, Structure of the BRAF-MEK complex reveals a kinase activity independent role for BRAF in MAPK signaling. *Cancer Cell* **26**, 402–413 (2014).
46. G. A. Jindal *et al.*, In vivo severity ranking of Ras pathway mutations associated with developmental disorders. *Proc. Natl. Acad. Sci. U.S.A.* **114**, 510–515 (2017).
47. Y. Kubota *et al.*, Qualitative differences in disease-associated MEK mutants reveal molecular signatures and aberrant signaling-crosstalk in cancer. *Nat. Commun.* **13**, 4063 (2022).
48. C. Nava *et al.*, Cardio-facio-cutaneous and Noonan syndromes due to mutations in the RAS/MAPK signalling pathway: Genotype-phenotype relationships and overlap with Costello syndrome. *J. Med. Genet.* **44**, 763–771 (2007).
49. J. L. Marks *et al.*, Novel MEK1 mutation identified by mutational analysis of epidermal growth factor receptor signaling pathway genes in lung adenocarcinoma. *Cancer Res.* **68**, 5524–5528 (2008).
50. N. Wagle *et al.*, Dissecting therapeutic resistance to RAF inhibition in melanoma by tumor genomic profiling. *J. Clin. Oncol.* **29**, 3085–3096 (2011).
51. Y. Narumi *et al.*, Molecular and clinical characterization of cardio-facio-cutaneous (CFC) syndrome: Overlapping clinical manifestations with Costello syndrome. *Am. J. Med. Genet. A* **143A**, 799–807 (2007).
52. K. W. Gripp *et al.*, Further delineation of the phenotype resulting from BRAF or MEK1 germline mutations helps differentiate cardio-facio-cutaneous syndrome from Costello syndrome. *Am. J. Med. Genet. A* **143A**, 1472–1480 (2007).
53. A. L. Schulz *et al.*, Mutation and phenotypic spectrum in patients with cardio-facio-cutaneous and Costello syndrome. *Clin. Genet.* **73**, 62–70 (2008).
54. G. Yoon, J. Rosenberg, S. Blaser, K. A. Rauen, Neurological complications of cardio-facio-cutaneous syndrome. *Dev. Med. Child Neurol.* **49**, 894–899 (2007).
55. A. M. Nystrom *et al.*, Noonan and cardio-facio-cutaneous syndromes: Two clinically and genetically overlapping disorders. *J. Med. Genet.* **45**, 500–506 (2008).
56. K. J. Karczewski *et al.*, The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* **581**, 434–443 (2020).
57. E. F. Mason *et al.*, Detection of activating MAP2K1 mutations in atypical hairy cell leukemia and hairy cell leukemia variant. *Leuk. Lymphoma* **58**, 233–236 (2017).
58. X. S. Puente *et al.*, Non-coding recurrent mutations in chronic lymphocytic leukaemia. *Nature* **526**, 519–524 (2015).
59. D. A. Landau *et al.*, Mutations driving CLL and their evolution in progression and relapse. *Nature* **526**, 525–530 (2015).
60. R. Yaeger *et al.*, Clinical sequencing defines the genomic landscape of metastatic colorectal cancer. *Cancer Cell* **33**, 125–136.e3 (2018).
61. S. A. Forbes *et al.*, COSMIC: Exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res.* **43**, D805–811 (2015).
62. A. H. Shain *et al.*, Exome sequencing of desmoplastic melanoma identifies recurrent NFKBIE promoter mutations and diverse activating mutations in the MAPK pathway. *Nat. Genet.* **47**, 1194–1199 (2015).
63. M. Cizmarova *et al.*, New mutations associated with rasopathies in a central european population and genotype-phenotype correlations. *Ann. Hum. Genet.* **80**, 50–62 (2016).
64. A. Zehir *et al.*, Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat. Med.* **23**, 703–713 (2017).
65. J. A. Couto *et al.*, Somatic MAP2K1 mutations are associated with extracranial arteriovenous malformation. *Am. J. Hum. Genet* **100**, 546–554 (2017).
66. Y. Gao *et al.*, Allele-specific mechanisms of activation of MEK1 mutants determine their properties. *Cancer Discov.* **8**, 648–661 (2018).
67. G. A. Jindal *et al.*, How activating mutations affect MEK1 regulation and function. *J. Biol. Chem.* **292**, 18814–18820 (2017).
68. E. Kinoshita-Kikuta *et al.*, Increase in constitutively active MEK1 species by introduction of MEK1 mutations identified in cancers. *Biochim. Biophys. Acta Proteins Proteom.* **1867**, 62–70 (2019).
69. K. Zeng *et al.*, BRAFV600E and MAP2K1 mutations in Langerhans cell histiocytosis occur predominantly in children. *Hematol. Oncol.* **35**, 845–851 (2017).
70. T. O. Fischmann *et al.*, Crystal structures of MEK1 binary and ternary complexes with nucleotides and inhibitors. *Biochemistry* **48**, 2661–2674 (2009).
71. S. J. Mansour, J. M. Candia, J. E. Matsuura, M. C. Manning, N. G. Ahn, Interdependent domains controlling the enzymatic activity of mitogen-activated protein kinase kinase 1. *Biochemistry* **35**, 15529–15536 (1996).
72. N. A. Brown *et al.*, High prevalence of somatic MAP2K1 mutations in BRAF V600E-negative Langerhans cell histiocytosis. *Blood* **124**, 1655–1658 (2014).
73. E. M. Van Allen *et al.*, The genetic landscape of clinical resistance to RAF inhibition in metastatic melanoma. *Cancer Discov.* **4**, 94–109 (2014).
74. D. Bottorff, S. Stang, S. Agellon, J. C. Stone, RAS signalling is abnormal in a c-raf1 MEK1 double mutant. *Mol. Cell Biol.* **15**, 5113–5122 (1995).
75. H. Tang, P. D. Thomas, PANTHER-PSEP: Predicting disease-causing genetic variants using position-specific evolutionary preservation. *Bioinformatics* **32**, 2230–2232 (2016).
76. E. Capriotti, R. B. Altman, Y. Bromberg, Collective judgment predicts disease-associated single nucleotide variants. *BMC Genomics* **14 Suppl 3** (Suppl. 3), S2 (2013).
77. J. Bendl *et al.*, PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS Comput. Biol.* **10**, e1003440 (2014).
78. V. Pejaver *et al.*, Inferring the molecular and phenotypic impact of amino acid variants with MutPred2. *Nat. Commun.* **11**, 5918 (2020).
79. K. A. Jagadeesh *et al.*, M-CAP eliminates a majority of variants of uncertain significance in clinical exomes at high sensitivity. *Nat. Genet.* **48**, 1581–1586 (2016).
80. Y. Bromberg, B. Rost, SNAP: Predict effect of non-synonymous polymorphisms on function. *Nucleic Acids Res.* **35**, 3823–3835 (2007).
81. E. A. Stone, A. Sidow, Physicochemical constraint violation by missense substitutions mediates impairment of protein function and disease severity. *Genome Res.* **15**, 978–986 (2005).
82. R. A. Marmion *et al.*, Molecular mechanisms underlying cellular effects of human MEK1 mutations. *Mol. Biol. Cell* **32**, 974–983 (2021).
83. Y. Goyal *et al.*, Divergent effects of intrinsically active MEK variants on developmental Ras signaling. *Nat. Genet.* **49**, 465–469 (2017).
84. E. Yeung *et al.*, Inference of multisite phosphorylation rate constants and their modulation by pathogenic mutations. *Curr. Biol.* **30**, 877–882.e6 (2020).
85. F. S. Neuman-Silberberg, T. Schupbach, Dorsoventral axis formation in Drosophila depends on the correct dosage of the gene gurken. *Development* **120**, 2457–2463 (1994).
86. F. S. Neuman-Silberberg, T. Schupbach, The Drosophila dorsoventral patterning gene gurken produces a dorsally localized RNA and encodes a TGF alpha-like protein. *Cell* **75**, 165–174 (1993).
87. K. E. James, J. B. Dorman, C. A. Berg, Mosaic analyses reveal the function of Drosophila Ras in embryonic dorsoventral patterning and dorsal follicle cell morphogenesis. *Development* **129**, 2209–2222 (2002).
88. A. M. Queenan, A. Ghabrial, T. Schupbach, Ectopic activation of torpedo/Egfr, a Drosophila receptor tyrosine kinase, dorsalizes both the eggshell and the embryo. *Development* **124**, 3871–3880 (1997).
89. E. A. Drier, L. H. Huang, R. Steward, Nuclear import of the Drosophila Rel protein Dorsal is regulated by phosphorylation. *Genes Dev.* **13**, 556–568 (1999).
90. M. Miller, D. Vitale, P. C. Kahn, B. Rost, Y. Bromberg, funtrp: Identifying protein positions for variation driven functional tuning. *Nucleic Acids Res.* **47**, e142 (2019).
91. K. Katoh, J. Rozewicki, K. D. Yamada, MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **20**, 1160–1166 (2019).
92. O. Adebali, D. R. Ortega, I. B. Zhulin, CDvist: A webserver for identification and visualization of conserved domains in protein sequences. *Bioinformatics* **31**, 1475–1477 (2015).
93. A. M. Waterhouse, J. B. Procter, D. M. Martin, M. Clamp, G. J. Barton, Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
94. S. Kumar, G. Stecher, M. Li, C. Knyaz, K. Tamura, MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
95. I. Letunic, P. Bork, Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).
96. J. G. Tate *et al.*, COSMIC: The catalogue of somatic mutations in cancer. *Nucleic Acids Res.* **47**, D941–D947 (2019).
97. S. Mazzalupo, L. Cooley, Illuminating the role of caspases during Drosophila oogenesis. *Cell Death Differ.* **13**, 1950–1959 (2006).
98. D. Olivieri, M. M. Sykora, R. Sachidanandam, K. Mechtler, J. Brennecke, An in vivo RNAi assay identifies major genetic and cellular requirements for primary piRNA biogenesis in Drosophila. *EMBO J.* **29**, 3301–3317 (2010).
99. H. E. Johnson *et al.*, The spatiotemporal limits of developmental Erk signaling. *Dev. Cell* **40**, 185–192 (2017).