scientific reports

OPEN

Structural insights and evaluation of the potential impact of missense variants on the interactions of SLIT2 with ROBO1/4 in cancer progression

Debmalya Sengupta[1], Gairika Bhattacharya[1,3], Sayak Ganguli[2✉] & Mainak Sengupta[1✉]

The cognate interaction of ROBO1/4 with its ligand SLIT2 is known to be involved in lung cancer progression. However, the precise role of genetic variants, disrupting the molecular interactions is less understood. All cancer-associated missense variants of ROBO1/4 and SLIT2 from COSMIC were screened for their pathogenicity. Homology modelling was done in Modeller 9.17, followed by molecular simulation in GROMACS. Rigid docking was performed for the cognate partners in PatchDock with refinement in HADDOCK server. Post-docking alterations in conformational, stoichiometric, as well as structural parameters, were assessed. The disruptive variants were ranked using a weighted scoring scheme. In silico prioritisation of 825 variants revealed 379 to be potentially pathogenic out of which, about 12% of the variants, i.e. ROBO1 (14), ROBO4 (8), and SLIT2 (23) altered the cognate docking. Six variants of ROBO1 and 5 variants of ROBO4 were identified as "high disruptors" of interactions with SLIT2 wild type. Likewise, 17 and 13 variants of SLIT2 were found to be "high disruptors" of its interaction with ROBO1 and ROBO4, respectively. Our study is the first report on the impact of cancer-associated missense variants on ROBO1/4 and SLIT2 interactions that might be the drivers of lung cancer progression.

Cancer morbidity is mainly attributed to metastasis[1,2], where the invasion and migration of tumour cells are critical phenomena. However, the pathways that regulate metastasis are still unclear, which acts as a deterrent for an introduction of therapeutic modalities. Recent researches indicated that the SLIT-ROBO pathway also plays a critical role in the modulation of chemokine activation and cellular migration of different lineages[3–6]. The SLIT family of proteins docks to Roundabout (ROBO) receptors and modulate morphogenesis[7], leukocyte chemotaxis[4], and most importantly, metastasis and angiogenesis[8,9]. SLIT2 is the most studied member of the SLIT family that has been found to exert a tumour suppressive activity, mainly by binding to the ROBO1 receptor[10]. Strong evidence suggests that SLIT/ROBO signalling requires heparan sulfate proteoglycans (HSPGs) as a co-receptor[11–13]. HSPGs consist of a large proteoglycan core protein decorated with heparan sulfate chains, which stabilises the interaction between ROBO1 and SLIT2[11–13]. The recognition of *SLIT2* by *ROBO1* is mediated through proper interaction of the respective domains, such as *SLIT2.D2*[14] and *ROBO1.IG1* that, again, is facilitated by heparan sulfate units[11–13]. Less is known regarding ROBO4 and SLIT2 interaction.

The SLIT/ROBO signalling pathway has anti-migratory and pro-apoptotic properties in many cancers[15], such as Small Cell Lung Cancer (SCLC). These tumour-suppressive effects could be weakened by mutations[16] within the cognate partners and its mediators. Robo1 knockout mice exhibited pulmonary dysmorphogenesis that resulted in embryonic death in most cases, while the surviving ones showed bronchial hyperplasia[17]. The SLIT2-ROBO1 pathway was found to counteract HGF-mediated tumour cell migration by inhibiting CDC-42 and stimulating RAC-1[7]. Recent studies have indicated the possible anticancer attributes of SLIT2-ROBO1 signalling[18] pathway in tumour cell migration and angiogenesis in lung cancer[19], breast cancer[20], gastric cancer[21], and colorectal cancer[22]. For instance, the *SLIT2* gene is inactivated in multiple cancers[23–25], including lung cancer[26], often

[1]Department of Genetics, University of Calcutta, University College of Science (UCSTA), 35, Ballygunge Circular Road, Kolkata 700 019, India. [2]Department of Biotechnology, St. Xavier's College (Autonomous), 30, Mother Teresa Sarani, Kolkata 700 016, India. [3]Cactus Communications, Mumbai, India. ✉email: sayakbif@yahoo.com; sengupta.mainak@gmail.com

as a result of promoter hypermethylation, mutations (frequency of occurrence ~ 10%) and LOH[26,27]. Recent investigations have shown the role of Myo9b-RhoA in mediating the inhibitory effect of SLIT2 on lung cancer cell migration[28].

Similarly, the downregulation of USP33 (a ubiquitin ligase enzyme) reduces the stability of the ROBO1 receptor in lung cancer cells[29] and breast cancer cell migration[30]. Interestingly, ROBO1 was found to be a specific serum biomarker in SCLC[31]. Thus, it is well-established that the *SLIT2-ROBO1* signalling pathway is tumour suppressive in many cancers, including lung cancer[26,28,29]. The anti-angiogenic effect of ROBO4 is also well documented[8,32]. Increased ROBO4 expression has been correlated with increased overall survival in early-stage non-small cell lung cancer[33]. However, the role of the genetic variants on the interaction of SLIT2/ROBO1/4, potentially altering the anti-angiogenic pathway and overall tumour suppressive function, is yet to be elucidated.

This work aims at the in silico prioritisation, and pathogenicity assessment of the missense variants of SLIT2/ROBO1/4 reported in lung cancer cases, based on sequence- and structure-based parameters followed by docking analyses to determine the interaction dynamics. We hypothesise that the variants would be responsible for the disruption of effective receptor-ligand docking and could result in the alteration of its tumour suppressive function within the pulmonary cells leading to lung cancer metastasis. These variants either individually or in combination[34] could act as docking disruptors by inducing conformational changes, like fluctuations in the secondary structural elements, alterations in solvent accessible surface area or solvation free energy, number of hydrogen bonds, number of salt bridges, partial unfolding or displacement of helices, any altered local/global flexibility and docking pose as compared to the wild-type proteins. It is worth mentioning that in this work, the missense variants reported in other cancers were also taken into consideration because some of them could also be found in lung cancer with the incorporation of more sample data using advanced sequencing technologies. For instance, p.R119Q (Arg119Glu) is reported in both colorectal cancer and lung cancer. Likewise, more variants reported in other cancers could be found in lung cancer with further studies.

Methods

Data mining.
Somatic variants of *ROBO1*, *ROBO4*, and *SLIT2* were curated from the "Catalogue Of Somatic Mutations In Cancer" (COSMIC, release v73; http://cancer.sanger.ac.uk/cosmic) database[35]. The canonical sequences for ROBO1 (UniProt sequence ID: Q9Y6N7-1), ROBO4 (UniProt sequence ID: Q8WZ75-1) and SLIT2 (UniProt sequence ID: O94813-1) were curated from UniProt (https://www.uniprot.org/)[36,37] and the secondary structures were predicted using PSIPRED (Position Specific Iterated BLAST based secondary structure PREDiction) (http://bioinf.cs.ucl.ac.uk/psipred/) server[38].

Identification of the degree of disorder in the proteins.
The IUPred2A (Intrinsically Unstructured protein Prediction 2A) (https://iupred2a.elte.hu/)[39] analysis was done to understand the presence or absence of disordered or intrinsically unstructured regions in the context of the amino acid sequences. The presence of Intrinsically Unstructured Regions (IUR) tends to add a degree of randomness to the native conformation of the protein structures[40].

Variant prioritisation by sequence-specific parameters.
The variants obtained from data mining were subjected to an in silico prioritisation pipeline for their pathogenicity determination by sequence-based and structure-based analysis (Fig. 1). In Silico analysis of all the variants (categorised as "*lung cancer-associated missense variants*" and "*missense variants associated with other cancers*") were done by PROVEAN (Protein Variation Effect Analyzer) (http://provean.jcvi.org/index.php)[41], which predicted whether amino acid substitutions or indels are "deleterious" or "Neutral". The PROVEAN web server also includes the algorithm of SIFT (Sorting Intolerant From Tolerant)[42] that predicted whether a variant is "damaging" (SIFT Score ≤ 0.05) or "tolerated" (SIFT Score > 0.05) based on sequence homology and the physical properties of amino acids. The same set of variants were analysed in *PolyPhen2* (Polymorphism Phenotyping v2)[43] (http://genetics.bwh.harvard.edu/pph2/) that predicted the functional significance of an amino acid substitution as "Probably damaging", "Possibly damaging" or "Benign". The entire set of variants were further analysed in SNPs&GO (Single Nucleotide Polymorphisms & Gene Ontology) (https://snps-and-go.biocomp.unibo.it/snps-and-go/index.html)[44], which predicted whether the variant is associated with "Disease" or not. Further, iMUTANT 2.0 (http://folding.biofold.org/i-mutant/i-mutant2.0.html)[45] predicted whether the variants "increases" or "decreases" the stability of the protein structures calculated from the input protein sequence. Finally, the variants were analysed in FATHMM (Functional Analysis Through Hidden Markov Models) (http://fathmm.biocompute.org.uk/)[46], which predicted whether a protein variant is "cancer-associated" or a "passenger" variant by combining sequence conservation within hidden Markov models (HMMs). The scoring scheme followed for sequence-based variant prioritisation is depicted in (Table 1). A weighted score was applied to the outcomes of each prediction from the softwares. More weightage was given to the outcome, depicting the disruptive/damaging nature of the mutations and was assigned with higher scores while the benign/neutral or tolerated mutations were given less weightage with lower scores. For example, p.R119Q (ROBO1.IG1) is a "*deleterious*" mutation according to PROVEAN and was assigned with score 1 while p.D930N (SLIT2) is a "*neutral*" mutation and was assigned with score 0. The cumulative score for each variant was calculated for the sequence-based analysis, and the variants with the cumulative score above the median value of the distribution of cumulative scores of all the variants were selected for further structure-based variant prioritisation.

Generation of Homology models for the wild type proteins.
Complete crystal structure of the proteins was not available in the Protein Data Bank. Therefore, homology models were generated to study the alterations in the structural dynamics of ROBO1/SLIT2 and ROBO4/SLIT2 interactions based on variants. Once the
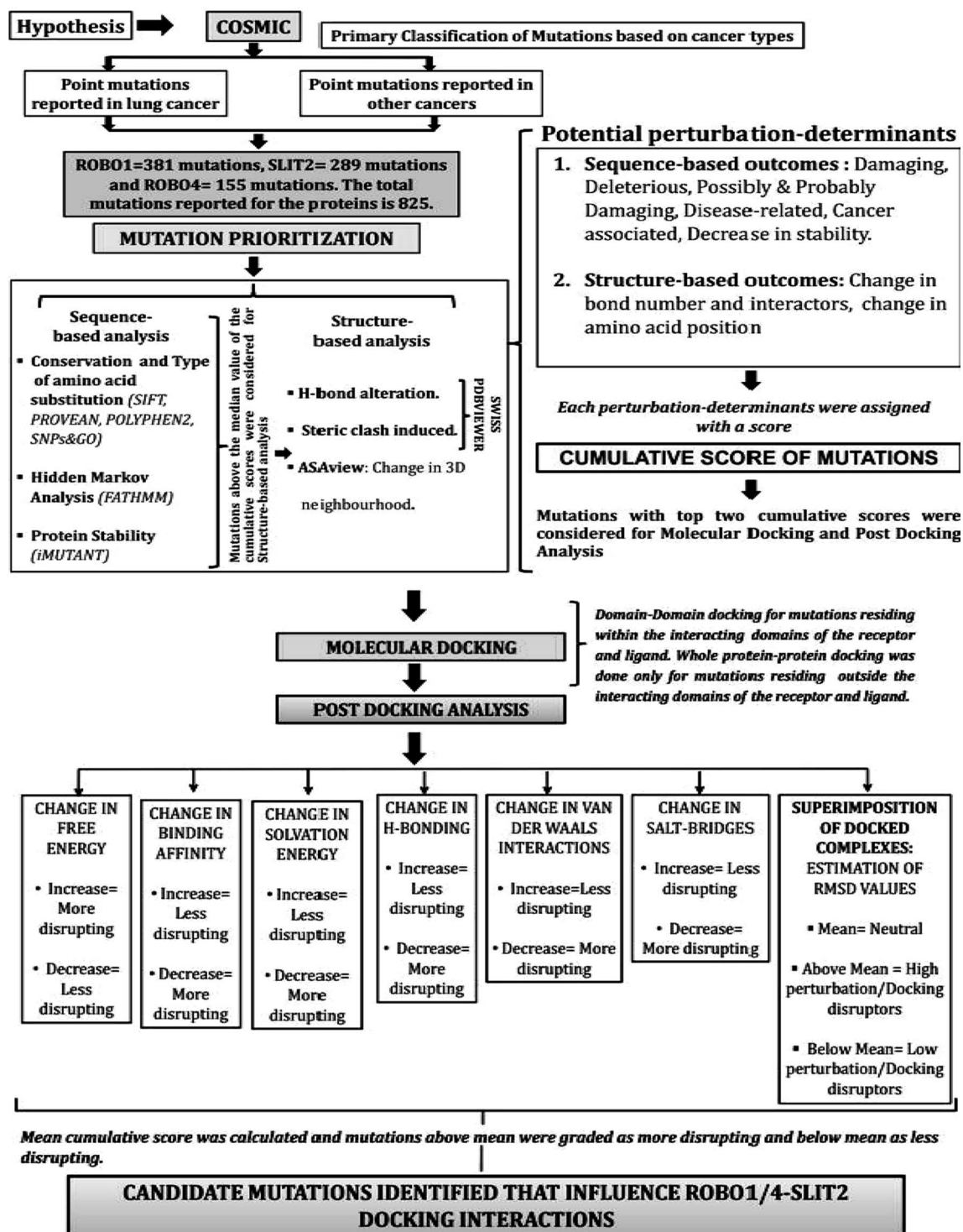
Figure 1. Schema of analysis. The work includes data mining followed by variant prioritisation and molecular docking for identification of candidate molecular targets as potential disruptors of ROBO/SLIT signalling mediated by ROBO1/SLIT2 and ROBO4/SLIT2.

secondary structures were obtained from PSIPRED[38], the sequences were subjected to BLAST analyses (https://blast.ncbi.nlm.nih.gov/Blast.cgi)[47,48] using the Protein Data Bank (https://www.rcsb.org/)[49]as the target database, to identify suitable homologous templates. The results showed four suitable templates each for ROBO1 (2V9Q, 5NOI, 2EO9, and 3WIH), ROBO4 (2V9Q, 2VRA, 3PUC, and 2VR9) and SLIT2 (2V70, 2V9T, 2V9S, and 2WFH). Each of the proteins and their corresponding interaction domains was modelled separately by comparative homology modelling using Modeller 9.17[50], and the modelled structures were then subjected to molecular simulation for 40 ns using a GROMOS96 43a2 (GROningen MOlecular Simulation) force field in GROMACS

Webservers	Predictive outcome for the mutants	Score
(A) Sequence-based analysis		
Provean	Neutral	0
	Deleterious	1
SIFT	Tolerated	0
	Damaging	1
Polyphen2	Benign	0
	Possibly Damaging	1
	Probably Damaging	2
SNPs&GO	Neutral	0
	Disease-related	1
iMUTANT 2.0	Increase Stability	0
	Decrease Stability	1
FATHMM	Passenger Variants	0
	Cancer-associated Variants	1
(B) Structure-based analysis		
H-bond	No Change compared to Wild Type (in any of the features like Bond Number, Bond Length, Interacting residues)	0
	Change compared to Wild Type (in any of the features like Bond Number, Bond Length, Interacting residues)	1
Steric clash	No Change compared to Wild Type	0
	Change compared to Wild Type	1
Asaview	The mutant residue has no positional alteration compared to the wild type	0
	The mutant residue shifts its position compared to the wild type	1

Table 1. The scoring scheme for the prioritisation of the variants of ROBO1, ROBO4, and SLIT2. Each of the predictive outcomes is assessed by comparing with the wild type protein sequence or structure.

(GROningen MAChine for Chemical Simulation)[51,52]. Following the simulation, the structures were analysed for torsional/ conformational stability from the Ramachandran plots of the models using the MolProbity server[53]. The Q–MEAN score verified the integrity and validity of the 3D structures of the generated homology models.

Generation of Mutant Structures from wild type homology models. The variants prioritised by the sequence-based analysis were inserted within the wild type homology models of ROBO1, ROBO4 and SLIT2 by using the MUTATE tool of DeepView v4.1.0[54]. The mutant proteins were subjected to the simulation as mentioned earlier conditions and validation protocol to generate the final mutated homology models of the proteins.

Structural analyses of the mutant models. The mutant models were assessed for alterations in the pattern of hydrogen bond interactions (in terms of the interacting residues), bond length and also in steric clashes between residues in comparison to their corresponding wild type structures (Fig. 1). Deviation of the loco-regional status of the mutable amino acid residues of mutant ROBO1, ROBO4, and SLIT2 compared to their wild types were analysed in terms of the Accessible Surface Area (ASA) or Solvent Accessibility (SA) in a webserver called ASAview (http://ccbb.jnu.ac.in/shandar/servers/asaview/)[55]. The scoring scheme followed for assigning a score to each outcome of the structure-based variant prioritisation is depicted in (Table 1).

A cumulative score for each variant (i.e. variants selected after sequence-based analysis) was calculated, and the variants with top two cumulative scores were selected for the subsequent docking analysis. The selection criteria were kept similar for variants residing within the docking domains and outside the docking domains, belonging to both lung cancer and non-lung cancer datasets.

After the prioritisation of the variants, alterations in potential ligand-binding hotspots, due to variants in ROBO1, ROBO4 and SLIT2 proteins were analysed in DoGSiteScorer (https://proteins.plus/) webserver[56]. This analysis was not used for prioritisation of the variants, but to assess the change in the respective structural attributes based on the variants finally selected for docking. The following parameters were assessed;

- *Drug score* The average drug score of all the pockets for both wild types and mutants was calculated. The mean drug score of wild type structures was annotated as 'Neutral'. The mean drug score of the mutants equal to the corresponding wild type was also annotated as 'Neutral'. Annotation 'Favourable' was assigned if the mean drug score of the mutant is more than its wild type. If the mean drug score of the mutants is less than the corresponding wild type, annotation 'Unfavourable' was assigned. This parameter depicts ligand (i.e., drugs, small molecules) binding capability of the structure, which could be considered as therapeutic targets.
- *The number of ligand-binding pockets* The number of ligand-binding pockets in the wild type structure was considered as the reference and was annotated as 'Neutral'. Gain in pockets upon variant is considered as 'Increase' while the loss of pockets upon variant was termed as 'Decrease'.

Parameters	Predictive outcome for the mutants	Weighted score
(A) Post docking analysis		
Changes in free energy (ΔG)	No change compared to wild type complex	0
	More negative compared to wild type complex	1
	Less negative compared to wild type complex	2
Changes in binding affinity ($\Delta\Delta G_{Bind}$)	No change compared to wild type complex	0
	More negative compared to wild type complex	1
	Less negative compared to wild type complex	2
Gain in solvation energy ($\Delta^i G$)	No change compared to Wild Type complex	0
	More negative compared to wild type complex	1
	Less negative compared to wild type complex	2
Change in h-bond number	No change compared to wild type complex	0
	Increased compared to wild type complex	1
	Decreased compared to wild type complex	2
Change in van der waals interactions	No change compared to wild type complex	0
	Increased compared to wild type complex	1
	Decreased compared to wild type complex	2
Change in the number of salt-bridges	No Change Compared To Wild Type complex	0
	Increased compared to wild type complex	1
	Decreased compared to wild type complex	2
Super-imposition (mean RMSD)	RMSD of the mutant docked complex is equal to the mean RMSD	0
	RMSD of the mutant docked complex is lesser than the mean RMSD	1
	RMSD of the mutant docked complex is higher than the mean RMSD	2

Table 2. Weighted scoring scheme for the docked complexes for the variants of ROBO1, ROBO4, and SLIT2. Each of the predictive outcomes is assessed by comparing with the wild type to wild type docked complex.

Molecular docking. Rigid docking of the whole proteins and also the interacting domains of the same was done using PatchDock (https://bioinfo3d.cs.tau.ac.il/PatchDock/) server[57]. The free energy of interaction of the wild type (native) and mutant complexes in the form of atomic contact energy based on shape complementarity principles was obtained from PatchDock and validated by using HADDOCK (High Ambiguity Driven protein–protein DOCKing) (https://haddock.science.uu.nl/) server[58,59]. The molecular docking was elaborately performed as a wild type to wild type and mutant to wild type proteins/domains. The best docking pose was selected based on the global energy of the docked complex and the biological relevance concerning the dynamics of the interaction between the membrane-bound receptor to its extracellular ligand.

Analyses of docked complexes. The mutant-to-wild type complexes were then analysed for their residue-specific interactions using PDBsum (http://www.ebi.ac.uk/thornton-srv/databases/pdbsum/Generate.html) server[60] and were compared with wild type-to-wild type docked complexes for changes in the bond number.

Change in the binding affinity ($\Delta\Delta G_{Bind}$ kcal/mol) due to variants were also checked for mutant-to-wild type complexes about wild type-to-wild type complexes in a web server named BeAtMuSiC v1.0 (prediction of Binding Affinity Changes upon MutationS) (http://babylone.ulb.ac.be/beatmusic/index.php)[61].

Similarly, the change in free energy (ΔG kcal/mol) due to variants was also checked for mutant-to-wild type complexes about wild type-to-wild type complexes during docking. Further, gain in solvation energy upon complex formation, for mutant-to-wild type docked complexes about wild type-to-wild type docked complexes, were assessed in PDBePISA (Protein Data Bank in Europe Proteins, Interfaces, Structures and Assemblies) webserver (http://www.ebi.ac.uk /pdbe/pisa/)[62].

A weighted scoring scheme was followed for assigning a score to each outcome of the post-docking analysis (Table 2).

Analyses of structural distortion. The mutant-to-wild type complexes were superimposed upon wild type-to-wild type complexes using *Matchmaker* tool of UCSF-Chimera 1.14rc[63] where the two docked complexes were subjected to pairwise alignment per residue to identify non-alignment of the superimposed structures at the residue per se or for the overall conformation of the interacting partners. The degree of conformational distortion of the mutant complex compared to the wild type was assessed by the magnitude of the RMSD (Root Mean Square Deviation) values. The mean RMSD value for all the complexes was estimated, and the scoring of the variants was done (Table 2).

Cumulative score and grading of variants from docked complex analysis. Following the weighted scoring scheme as depicted in (Table 2), a cumulative score was estimated for each docked complex by merely summing the individual weighted score of each of the parameters analysed post docking. It was followed by the calculation of the mean of the cumulative score distribution of all mutant complexes for the respective docking categories,

i.e. ROBO1.IG1:SLIT2.D2, ROBO4.IG1-2:SLIT2.D2, ROBO1:SLIT2, and ROBO4:SLIT2. The cumulative score of the wild type:wild type docked complexes was denoted as '0'. A variant was graded as "More disrupting" when its cumulative score was more than the mean of the cumulative scores of all the mutant complexes of a specific docking category and "Less disrupting" when its cumulative score was less than the mean of the cumulative scores of all the mutant complex for a specific docking category as mentioned above.

Results

Curation and categorisation of somatic variants.

Systematic mining of the COSMIC[35] database (release v73) for somatic missense variants revealed 381 variants in ROBO1, 289 variants in SLIT2 and 155 variants in ROBO4. Out of 825 variants; 34 in ROBO1, 57 in SLIT2, and 34 in ROBO4 were reported to be associated with lung cancer. Thus, these were categorised as *lung cancer dataset*. In comparison, the other 255 variants in ROBO1, 324 variants in SLIT2 and 121 variants in ROBO4 were considered as *non-lung cancer dataset*.

Prioritisation of variants on sequence-based and structure-based parameters.

Prioritisation of the variants was done according to our predefined pipeline, as shown in (Fig. 1). The variants were analysed in PROVEAN[41] that predicted nearly 50% of ROBO1, 29% of ROBO4, and 63% of SLIT2 variants of lung cancer dataset to be "deleterious". Similarly, 40% of ROBO1, 18% of ROBO4, and 56% of SLIT2 variants of non-lung cancer dataset were predicted to be "deleterious" by PROVEAN.

Similarly, analysis in SIFT[42] revealed about 65% of ROBO1, 47% of ROBO4, and 60% of SLIT2 variants of lung cancer dataset to be "damaging". About 61% of ROBO1, 45% of ROBO4, and 61% of SLIT2 variants of non-lung cancer dataset were predicted to be "damaging" by SIFT.

PolyPhen2[43] predicted about 15% of ROBO1, 15% of ROBO4, and 23% of SLIT2 variants of lung cancer dataset to be "possibly damaging". About 13% of ROBO1, 13% of ROBO4, and 19% of SLIT2 variants of non-lung cancer dataset were predicted to be "possibly damaging" by *PolyPhen2*. Similarly, about 53% of ROBO1, 50% of ROBO4, and 46% of SLIT2 variants of lung cancer dataset to be "possibly damaging", while 45% of ROBO1, 50% of ROBO4, and 43% of SLIT2 variants of non-lung cancer dataset were predicted to be "probably damaging" by *PolyPhen2*. SNPs&GO[44] predicted about 35% of ROBO1, 35% of ROBO4, and 32% of SLIT2 variants of lung cancer dataset to be "Disease-related". About 25% of ROBO1, 25% of ROBO4, and 35% of SLIT2 variants of non-lung cancer dataset were predicted to be "Disease-related" by SNPs&GO.

Interestingly, iMUTANT 2.0 revealed 91% of ROBO1, 79% of ROBO4, and 84% of SLIT2 variants of lung cancer dataset to cause a "decrease" in protein stability. About 89% of ROBO1, 76% of ROBO4, and 83% of SLIT2 variants of non-lung cancer dataset were predicted to cause a "decrease" in protein stability.

Amino acid sequence-based predictions from FATHMM[46] using hidden Markov Models revealed 0% of ROBO1, 6% of ROBO4, and 0% of SLIT2 variants of lung cancer dataset to be associated with some form of "cancer". About, 0.4% of ROBO1, 0.8% of ROBO4, and 2% of SLIT2 variants of non-lung cancer dataset were predicted to be associated with "cancer" by FATHMM.

The cumulative score of all sequence-based parameters per variant was calculated (Table 1). The variants with a cumulative score > 3 (median value) of the distribution of the cumulative scores, were selected for further structure-based analysis. Thus, from the lung cancer dataset, 18 variants of ROBO1, 29 variants of SLIT2 and 13 variants of ROBO4 were selected for structure-based analysis. From the non-lung cancer dataset, 113 variants of ROBO1, 134 variants of SLIT2 and 41 variants of ROBO4 were selected for structure-based analysis. The detailed results of the final set of prioritised variants are summarised in the following table (Table 3). The list of all variants of the three proteins with sequence-based scores is summarised in (Supplementary Information, Table S1).

The generation of mutant homology models as described in the Materials and Methods section and structural analysis of the mutants versus their respective wild types revealed variant-induced alteration in local hydrogen bond number and bond length among the interacting amino acid residues for the variants; 10 in ROBO1, 10 in ROBO4 and 14 in SLIT2 from lung cancer dataset. Similarly, for non-lung cancer dataset, the variants conferring change in local hydrogen bond number and bond length are 50 in ROBO1, 31 in ROBO4, and 89 in SLIT2. Steric clash-induced destabilisations, primarily found in variants involving bulky amino acids were 7 in ROBO1, 5 in ROBO4 and 9 in SLIT2 from the lung cancer dataset. In comparison to, 9 variants in ROBO1, variants in ROBO4, and 34 variants from the non-lung cancer dataset were found. The representative images of variant-induced changes for both the changes in H-bond and steric clashes compared to wild types are depicted in (Supplementary Information, Fig. S1). The detailed results of variant-specific alterations in local H-bond (number, length, interacting residues) and steric clash in the protein structures for the final set of prioritised variants are summarised in the following tables (Table 3, Supplementary Information, Table S2).

Analysis of the mutant structures of all the interacting domains of the 3 proteins in ASAview revealed 1 variant of SLIT2.D2 exhibit an inter-orbit shift in the position of the mutant amino acid residue compared to wild type structure in the lung cancer dataset. Similarly, in the non-lung cancer dataset, 5 variants of ROBO1.IG1, 7 variants of SLIT2.D2 and 2 variants of ROBO4.IG1-2 domains show an inter-orbit shift of mutant amino acid compared to wild types. However, 2 variants of ROBO1.IG1, 2 variants of SLIT2.D2 and 1 variant of ROBO4.IG1-2 domains in lung cancer dataset and 6 variants in SLIT2.D2 in the non-lung cancer dataset exhibits an intra-orbit amino acid position shift compared to wild types.

Similarly, for outside the docking domains, 3 variants of ROBO1 and 2 variants of ROBO4 show inter-orbital shift of mutant amino acid residue compared to wild type residue in the lung cancer dataset. In the non-lung cancer dataset, 1 variant of ROBO1 show inter-orbital shift of mutant amino acid residue compared to wild type residue. Only 1 variant of SLIT2 of lung cancer dataset show intra-orbital shift of mutant amino acid residue compared to wild type residue. The remaining variants do not have any impact on the change in the amino acid position compared to the wild type within the 3D protein structure. The representative images of the spiral plots

Genes	Variants	PROVEAN	SIFT	Polyphen2	SNPs&GO	iMUTANT 2.0	FATHMM	H-bond change	Steric clash change	ASA	Total score
(A) Lung cancer dataset											
(I) Variants within docking domain											
ROBO1	p.R119Q	1	1	2	1	1	0	1	0	1	8
ROBO1	p.K103N	1	1	2	1	1	0	1	0	1	8
SLIT2	p.D435V	1	1	2	1	1	0	1	1	1	9
SLIT2	p.R306L	1	1	2	1	1	0	1	1	1	9
SLIT2	p.Y323H	1	0	2	1	1	0	1	1	1	8
ROBO4	p.G102C	1	1	2	0	1	0	1	1	1	8
ROBO4	p.C207S	1	1	2	1	1	1	1	0	0	8
(II) Variants outside docking domain											
ROBO1	p.R635S	1	1	2	1	1	0	1	1	1	9
ROBO1	p.D312V	1	1	2	1	1	0	1	1	1	9
ROBO1	p.R306I	1	1	2	1	1	0	1	1	1	9
SLIT2	p.P54R	1	1	2	1	1	0	1	1	1	9
SLIT2	p.P665S	1	1	2	1	1	0	1	1	0	8
ROBO4	p.G829W	1	1	2	1	1	0	1	1	1	9
ROBO4	p.W367C	1	1	2	1	1	1	1	0	1	9
(B) Non-lung cancer dataset											
(I) Variants within docking domain											
ROBO1	p.I78S	1	1	2	1	1	0	1	0	1	8
ROBO1	p.G82V	1	1	2	1	1	0	1	0	1	8
ROBO1	p.P123L	1	1	2	1	1	0	1	1	0	8
ROBO1	p.R131C	1	1	2	1	1	0	1	0	1	8
ROBO1	p.G143R	1	1	2	1	1	0	1	0	1	8
ROBO1	p.G154E	1	1	2	1	1	0	1	0	1	8
SLIT2	p.C436F	1	1	2	1	1	0	1	1	1	9
(B) Non-lung cancer dataset											
(I) Variants within docking domain											
SLIT2	p.R388W	1	1	2	1	1	0	1	1	1	9
SLIT2	p.R456H	1	1	2	1	1	0	1	1	1	9
SLIT2	p.N430K	1	1	2	1	1	0	1	1	1	9
SLIT2	p.R420W	1	1	2	1	1	0	0	1	1	8
SLIT2	p.L350V	1	1	2	1	1	0	0	1	1	8
SLIT2	p.P431L	1	1	2	1	1	0	1	1	1	9
SLIT2	p.L331P	1	1	2	1	1	0	1	1	0	8
SLIT2	p.L398F	1	1	2	1	1	0	1	1	0	8
SLIT2	p.L371P	1	1	2	1	0	0	1	1	1	8
SLIT2	p.S352L	1	1	2	1	1	0	1	1	1	9
SLIT2	p.R462C	1	1	2	1	1	0	1	1	1	9
SLIT2	p.A340T	1	1	2	0	1	0	1	1	1	8
SLIT2	p.P322L	1	1	2	0	1	0	1	1	1	8
SLIT2	p.R461C	1	1	2	1	1	0	1	0	1	8
ROBO4	p.G86E	1	1	2	1	0	0	1	1	1	8
ROBO4	p.C207F	1	1	2	1	1	1	0	1	1	9
(II) Variants outside docking domain											
ROBO1	p.P176L	1	1	2	1	1	0	1	1	1	9
ROBO1	p.G337V	1	1	2	1	1	0	1	1	0	8
ROBO1	p.C476Y	1	1	2	1	1	0	1	1	0	8
SLIT2	p.C215Y	1	1	2	1	0	1	1	1	0	8
SLIT2	p.G576R	1	1	2	1	1	0	1	1	0	8
SLIT2	p.S636C	1	1	2	1	1	0	1	0	1	8
ROBO4	p.G425V	1	1	2	1	1	0	1	1	0	8
ROBO4	p.G876V	1	1	2	1	1	0	1	1	0	8

Table 3. The final set of in silico prioritised variants by sequence-based and structure-based parameters.

are depicted in (Supplementary Information, Fig. S2), and the results of the final set of prioritised variants with their respective scores are summarised in (Table 3).

Finally, for the interacting domains, the prioritisation analysis revealed 2 variants of ROBO1, 2 variants of ROBO4 and 3 variants of SLIT2 in lung cancer dataset and 6 variants of ROBO1, 2 variants of ROBO4 and 15 variants of SLIT2 from the non-lung cancer dataset as depicted in (Table 3). These variants were further analysed for their effect on the domain-domain interactions. Similarly, for full protein–protein interaction involving variants within the non-interacting regions of the proteins; 3 variants of ROBO1, 2 variants of SLIT2 and 2 variants of ROBO4 from lung cancer dataset and 3 variants of ROBO1, 3 variants of SLIT2, and 2 variants of ROBO4 were prioritised for further docking and post-docking analyses. The results of the final set of prioritised variants are depicted in (Table 3).

Structure validation and analysis. The sequence-based analysis of the wild type models for residue-specific IURs from IuPred2A indicates that 35.5% (38 disordered/69 ordered) residues of ROBO1.IG1, 59.6% (112 disordered/76 ordered) residues of ROBO4.IG1-2, and 0% (0 disordered/210 ordered) residues of SLIT2. D2 to be intrinsically unstructured (Supplementary Information, Fig. S2, Table S2). Similarly, 51.3% (847 disordered/804 ordered) residues of ROBO1, 45.4% (457 disordered/550 ordered) residues of ROBO4, and 0.005% (07 disordered/1522 ordered) residues of SLIT2 whole protein models, to be intrinsically unstructured (Supplementary Information, Fig. S3, Table S2). These analyses have led to the rationalisation of the Ramachandran plots and Q-mean validation outcomes, which indicate a high level of structural dynamism in protein conformation related to their function. The unstructured residues identified in the analyses may contribute towards the flexibility of the protein backbone and adjoining side chains to allow interaction with their cognate partners. It may also explain the role of HSPGs as a stabiliser of the interaction between ROBOs and SLITs.

The wild types and prioritised mutants of the generated homology models were assessed for their conformational stability in 3D by their respective Ramachandran plots and Q-mean (https:// swissmodel.expasy.org/qmean/) scores. The individual Q-mean scores of all the wild types and the mutants of both partial interacting domains and full protein models were analysed and are graphically represented in (Supplementary Information, Fig. S4, Table S3). Further, analysis of conformational stability by Ramachandran plots of the representative top-scoring variants of the interacting domains and whole protein models along with their respective wild types, are depicted in (Supplementary Information, Fig. S5). The percentage of amino acid residues in the allowed regions and the percentage of amino acids favoured in the Ramachandran plot assessed the stability of the protein structures.

Further, analysis of the effect of variants on the ligand/small molecule/drug binding capability of the proteins in DoGSiteScorer and the changes in the number of ligand binding pockets upon mutations is depicted in (Supplementary Information, Table S4). Representative images of the impact of variants on drug/ligand binding pockets in the proteins are depicted in (Supplementary Information, Fig. S6).

Molecular docking and post-docking analysis. Mutant-to-wild type molecular docking of the receptors (ROBO1 and ROBO4) and ligand (SLIT2) was performed to assess the effect of variants on the protein–protein interaction. Several parameters that influence the docking efficiency, such as '*Free energy change*'[64], '*Change in Binding affinity*', '*Gain in solvation energy*', '*Change in H-bond number*', '*Change in Van der Waals interactions*', '*Change in the number of salt bridges*' were considered. The analysis revealed that all the mutant ROBO1.IG1-SLIT2.D2 docked complexes have increased free energy change compared to the respective wild type. One variant in ROBO4.IG1-2 (G102C) docked to SLIT2.D2 (WT), and 1 variant in SLIT2.D2 docked to ROBO4.IG1-2 (WT) caused a decrease in free energy upon docking as compared to the wild type-wild type docked complexes. The results for both lung cancer and non-lung cancer datasets are summarised in the following table (Table 4).

One variant of SLIT2.D2 (S352L) docked with ROBO1.IG1 (WT) and 1 variant of ROBO4.IG1-2 (G102C) docked with SLIT2.D2 (WT) show an increase in binding affinity of the interacting proteins compared to the respective wild type docked complexes, which depicts the docking stabilising effect of the variants. The remaining 8 ROBO1.IG1 mutants docked with SLIT2.D2 (WT), 17 SLIT2.D2 mutants docked with ROBO1.IG1 (WT), 3 ROBO4.IG1-2 mutants docked with SLIT2.D2 (WT), and 18 SLIT2.D2 mutants docked with ROBO4.IG1-2 (WT) show a decrease in binding affinity compared to the respective wild type-wild type docked complexes, signifying the docking disruption effect of the variants. The results for both lung cancer and non-lung cancer datasets are summarised in the following table (Table 4).

Analysis over ePISA webserver revealed 8 ROBO1.IG1 mutants docked with SLIT2.D2 (WT), 18 SLIT2. D2 mutants docked with ROBO1.IG1 (WT), 1 ROBO4.IG1-2 mutants docked with SLIT2.D2 (WT), and 12 SLIT2.D2 mutants docked with ROBO4.IG1-2 (WT) to show a decrease in solvation energy while 2 ROBO4. IG1-2 mutants docked with SLIT2.D2 (WT), and 6 SLIT2.D2 mutants docked with ROBO4.IG1-2 (WT) show an increase in solvation energy as compared to their respective wild type docked complexes of ROBO1.IG1/SLIT2.D2 and ROBO4.IG1-2/SLIT2.D2. Only 1 ROBO4.IG1-2 mutant docked with SLIT2.D2 (WT) shows no change in solvation energy as compared to their respective wild type docked complexes. The results for both lung cancer and non-lung cancer datasets are summarised in the following table (Table 4). Analysis of the alteration in bond numbers revealed 2 ROBO1.IG1 mutants docked with SLIT2.D2 (WT), 4 SLIT2.D2 mutants docked with ROBO1.IG1 (WT), 1 ROBO4.IG1-2 mutants docked with SLIT2.D2 (WT), and 7 SLIT2.D2 mutants docked with ROBO4.IG1-2 (WT) to show a decrease in H-bond number while 2 ROBO1.IG1 mutants docked with SLIT2. D2 (WT), 12 SLIT2.D2 mutants docked with ROBO1.IG1 (WT), and 5 SLIT2.D2 mutants docked with ROBO4. IG1-2 (WT) to show an increase in H-bond number as compared to their respective wild type docked complexes. Similarly, the change in the number of van der Waals interactions and salt bridges were assessed. The alterations in the number of bonded and non-bonded interactions in mutant to wild type docked complexes, collectively

Protein–Protein Interactions	Changes in free energy (ΔG)	Changes in Binding affinity (ΔΔG$_{Bind}$)	Gain in Solvation energy (Δ^{i}G)	H-bonds (N)	Van der Waals interactions (N)	Salt-bridges (N)	Super-imposition (RMSD)	Total Score (TS)	Predictive outcome
(A) Domain-domain docking interactions									
(I) ROBO1.IG1 + SLIT2.D2									
p.Y323H + p.ROBO1.IG1(WT)	Increase:2	Decrease:2	Decrease:2	Decrease:2	Increase:1	Decrease:2	2	13	More disrupting (TS > Mean TS)
p.G82V + p.SLIT2.D2(WT)	Increase:2	Decrease:2	Decrease:2	Decrease:2	Increase:1	Decrease:2	1	12	More disrupting (TS > Mean TS)
p.C436F + p.ROBO1.IG1(WT)	Increase:2	Decrease:2	Decrease:2	Increase:1	Increase:1	Decrease:2	2	12	More disrupting (TS > Mean TS)
p.R388W + p.ROBO1.IG1(WT)	Increase:2	Decrease:2	Decrease:2	Increase:1	Increase:1	Decrease:2	2	12	More disrupting (TS > Mean TS)
p.R461C + p.ROBO1.IG1(WT)	Increase:2	Decrease:2	Decrease:2	Increase:1	Increase:1	Decrease:2	2	12	More disrupting (TS > Mean TS)
p.L350V + p.ROBO1.IG1(WT)	Increase:2	Decrease:2	Decrease:2	Increase:1	Increase:1	Decrease:2	2	12	More disrupting (TS > Mean TS)
p.L331P + p.ROBO1.IG1(WT)	Increase:2	Decrease:2	Decrease:2	Decrease:2	Increase:1	Decrease:2	1	12	More disrupting (TS > Mean TS)
p.L398F + p.ROBO1.IG1(WT)	Increase:2	Decrease:2	Decrease:2	Decrease:2	Increase:1	Increase:1	2	12	More disrupting (TS > Mean TS)
p.G154E + p.SLIT2.D2(WT)	Increase:2	Decrease:2	Decrease:2	Decrease:2	Increase:1	Increase:1	1	11	More disrupting (TS > Mean TS)
p.R306L + p.ROBO1.IG1(WT)	Increase:2	Decrease:2	Decrease:2	Increase:1	Increase:1	Decrease:2	1	11	More disrupting (TS > Mean TS)
p.R456H + p.ROBO1.IG1(WT)	Increase:2	Decrease:2	Decrease:2	Increase:1	Increase:1	Increase:1	2	11	More disrupting (TS > Mean TS)
p.N430K + p.ROBO1.IG1(WT)	Increase:2	Decrease:2	Decrease:2	Increase:1	Increase:1	Decrease:2	1	11	More disrupting (TS > Mean TS)
p.R420W + p.ROBO1.IG1(WT)	Increase:2	Decrease:2	Decrease:2	Increase:1	Increase:1	Decrease:2	1	11	More disrupting (TS > Mean TS)
p.P431L + p.ROBO1.IG1(WT)	Increase:2	Decrease:2	Decrease:2	Increase:1	Increase:1	Increase:1	2	11	More disrupting (TS > Mean TS)
p.L371P + p.ROBO1.IG1(WT)	Increase:2	Decrease:2	Decrease:2	Decrease:2	Increase:1	Neutral:0	2	11	More disrupting (TS > Mean TS)
p.A340T + p.ROBO1.IG1(WT)	Increase:2	Decrease:2	Decrease:2	Increase:1	Increase:1	Decrease:2	1	11	More disrupting (TS > Mean TS)
Mean = 10.88									
p.R119Q + p.SLIT2.D2(WT)	Increase:2	Decrease:2	Decrease:2	Neutral:0	Increase:1	Increase:1	2	10	Less disrupting (TS < Mean TS)
p.I78S + p.SLIT2.D2(WT)	Increase:2	Decrease:2	Decrease:2	Increase:1	Increase:1	Neutral:0	2	10	Less disrupting (TS < Mean TS)
p.P123L + p.SLIT2.D2(WT)	Increase:2	Decrease:2	Decrease:2	Increase:1	Increase:1	Increase:1	1	10	Less disrupting (TS < Mean TS)
p.R131C + p.SLIT2.D2(WT)	Increase:2	Decrease:2	Decrease:2	Neutral:0	Increase:1	Decrease:2	1	10	Less disrupting (TS < Mean TS)
p.D435V + p.ROBO1.IG1(WT)	Increase:2	Decrease:2	Decrease:2	Neutral:0	Increase:1	Decrease:2	1	10	Less disrupting (TS < Mean TS)
p.S352L + p.ROBO1.IG1(WT)	Increase:2	Increase:1	Decrease:2	Increase:1	Increase:1	Increase:1	2	10	Less disrupting (TS < Mean TS)
p.R462C + p.ROBO1.IG1(WT)	Increase:2	Decrease:2	Decrease:2	Increase:1	Increase:1	Neutral:0	2	10	Less disrupting (TS < Mean TS)
p.P322L + p.ROBO1.IG1(WT)	Increase:2	Decrease:2	Decrease:2	Neutral:0	Increase:1	Decrease:2	1	10	Less disrupting (TS < Mean TS)
p.K103N + p.SLIT2.D2(WT)	Increase:2	Decrease:2	Decrease:2	Neutral:0	Increase:1	Increase:1	1	9	Less disrupting (TS < Mean TS)
p.G143R + p.SLIT2.D2(WT)	Increase:2	Decrease:2	Decrease:2	Neutral:0	Increase:1	Increase:1	1	9	Less disrupting (TS < Mean TS)
(II) ROBO4.IG1-2 + SLIT2.D2									
p.S352L + p.ROBO4.IG1-2(WT)	Increase:2	Decrease:2	Decrease:2	Decrease:2	Decrease:2	Decrease:2	2	14	More disrupting (TS > Mean TS)
p.R462C + p.ROBO4.IG1-2(WT)	Increase:2	Decrease:2	Decrease:2	Decrease:2	Decrease:2	Decrease:2	2	14	More disrupting (TS > Mean TS)
p.D435V + p.ROBO4.IG1-2(WT)	Increase:2	Decrease:2	Increase:1	Decrease:2	Decrease:2	Decrease:2	2	13	More disrupting (TS > Mean TS)
p.Y323H + p.ROBO4.IG1-2(WT)	Increase:2	Decrease:2	Increase:1	Decrease:2	Decrease:2	Decrease:2	2	13	More disrupting (TS > Mean TS)
p.R461C + p.ROBO4.IG1-2(WT)	Increase:2	Decrease:2	Increase:1	Decrease:2	Decrease:2	Decrease:2	2	13	More disrupting (TS > Mean TS)
Continued									

Protein–Protein Interactions	Changes in free energy (ΔG)	Changes in Binding affinity (ΔΔG$_{Bind}$)	Gain in Solvation energy (Δ^{i}G)	H-bonds (N)	Van der Waals interactions (N)	Salt-bridges (N)	Super-imposition (RMSD)	Total Score (TS)	Predictive outcome
p.R420W + p.ROBO4.IG1-2(WT)	Increase:2	Decrease:2	Increase:1	Decrease:2	Decrease:2	Decrease:2	2	13	More disrupting (TS > Mean TS)
p.C436F + p.ROBO4.IG1-2(WT)	Increase:2	Decrease:2	Decrease:2	Increase:1	Increase:1	Decrease:2	2	12	More disrupting (TS > Mean TS)
p.R388W + p.ROBO4.IG1-2(WT)	Increase:2	Decrease:2	Decrease:2	Increase:1	Increase:1	Decrease:2	2	12	More disrupting (TS > Mean TS)
p.L371P + p.ROBO4.IG1-2(WT)	Increase:2	Decrease:2	Decrease:2	Increase:1	Increase:1	Decrease:2	2	12	More disrupting (TS > Mean TS)
p.C207S + p.SLIT2.D2(WT)	Increase:2	Decrease:2	Decrease:2	Decrease:2	Increase:1	Increase:1	2	12	More disrupting (TS > Mean TS)
p.P431L + p.ROBO4.IG1-2(WT)	Increase:2	Decrease:2	Decrease:2	Neutral:0	Increase:1	Decrease:2	2	11	More disrupting (TS > Mean TS)
p.A340T + p.ROBO4.IG1-2(WT)	Increase:2	Decrease:2	Decrease:2	Decrease:2	Decrease:2	Neutral:0	1	11	More disrupting (TS > Mean TS)
Mean = 10.90									
p.R306L + p.ROBO4.IG1-2(WT)	Increase:2	Decrease:2	Increase:1	Neutral:0	Increase:1	Decrease:2	2	10	Less disrupting (TS < Mean TS)
p.L350V + p.ROBO4.IG1-2(WT)	Increase:2	Decrease:2	Decrease:2	Increase:1	Decrease:2	Neutral:0	1	10	Less disrupting (TS < Mean TS)
p.L331P + p.ROBO4.IG1-2(WT)	Increase:2	Decrease:2	Decrease:2	Increase:1	Decrease:2	Neutral:0	1	10	Less disrupting (TS < Mean TS)
p.C207F + p.SLIT2.D2(WT)	Increase:2	Decrease:2	Increase:1	Neutral:0	Increase:1	Decrease:2	2	10	Less disrupting (TS < Mean TS)
p.G86E + p.SLIT2.D2(WT)	Increase:2	Decrease:2	Neutral:0	Neutral:0	Increase:1	Decrease:2	2	9	Less disrupting (TS < Mean TS)
p.N430K + p.ROBO4.IG1-2(WT)	Increase:2	Decrease:2	Decrease:2	Neutral:0	Decrease:2	Neutral:0	1	9	Less disrupting (TS < Mean TS)
p.P322L + p.ROBO4.IG1-2(WT)	Increase:2	Decrease:2	Decrease:2	Neutral:0	Decrease:2	Neutral:0	1	9	Less disrupting (TS < Mean TS)
p.R456H + p.ROBO4.IG1-2(WT)	Increase:2	Decrease:2	Increase:1	Neutral:0	Decrease:2	Neutral:0	1	8	Less disrupting (TS < Mean TS)
p.L398F + p.ROBO4.IG1-2(WT)	Decrease:1	Decrease:2	Decrease:2	Neutral:0	Increase:1	Increase:1	1	8	Less disrupting (TS < Mean TS)
p.G102C + p.SLIT2.D2(WT)	Decrease:1	Increase:1	Increase:1	Decrease:2	Increase:1	Neutral:0	1	7	Less disrupting (TS < Mean TS)
(B) Protein–protein docking interactions									
(III) ROBO1 + SLIT2									
p.P54R + p.ROBO1(WT)	Increase:2	Decrease:2	Increase:1	Decrease:2	Decrease:2	Increase:1	2	12	More disrupting (TS > Mean TS)
p.P665S + p.ROBO1(WT)	Increase:2	Decrease:2	Decrease:2	Decrease:2	Increase:1	Decrease:2	1	12	More disrupting (TS > Mean TS)
p.G576R + p.ROBO1 (WT)	Increase:2	Decrease:2	Increase:1	Decrease:2	Decrease:2	Decrease:2	1	12	More disrupting (TS > Mean TS)
p.R635S + p.SLIT2 (WT)	Increase:2	Decrease:2	Increase:1	Decrease:2	Increase:1	Decrease:2	1	11	More disrupting (TS > Mean TS)
p.D312V + p.SLIT2 (WT)	Increase:2	Increase:1	Decrease:2	Neutral : 0	Decrease:2	Decrease:2	2	11	More disrupting (TS > Mean TS)
p.P176L + p.SLIT2 (WT)	Increase:2	Decrease:2	Increase:1	Decrease:2	Increase:1	Decrease:2	1	11	More disrupting (TS > Mean TS)
p.C476Y + p.SLIT2 (WT)	Increase:2	Decrease:2	Increase:1	Increase:1	Decrease:2	Decrease:2	1	11	More disrupting (TS > Mean TS)
Mean = 10.64									
p.R306I + p.SLIT2 (WT)	Increase:2	Increase:1	Increase:1	Increase:1	Increase:1	Decrease:2	2	10	Less disrupting (TS < Mean TS)
p.S636C + p.ROBO1 (WT)	Increase:2	Increase:1	Increase:1	Decrease:2	Decrease:2	Neutral : 0	2	10	Less disrupting (TS < Mean TS)
p.G337V + p.SLIT2 (WT)	Increase:2	Increase:1	Increase:1	Neutral : 0	Increase:1	Decrease:2	2	9	Less disrupting (TS < Mean TS)
p.C215Y + p.ROBO1 (WT)	Increase:2	Increase:1	Increase:1	Increase:1	Increase:1	Increase:1	1	8	Less disrupting (TS < Mean TS)
(IV) ROBO4 + SLIT2									
p.W397C + p.SLIT2 (WT)	Increase:2	Decrease:2	Decrease:2	Increase:1	Increase:1	Increase:1	2	11	More disrupting (TS > Mean TS)
p.G425V + p.SLIT2(WT)	Increase: 2	Decrease: 2	Increase: 1	Decrease: 2	Decrease: 2	Neutral : 0	2	11	More disrupting (TS > Mean TS)
Continued									

Protein–Protein Interactions	Changes in free energy (ΔG)	Changes in Binding affinity (ΔΔG$_{Bind}$)	Gain in Solvation energy (Δ^{i}G)	H-bonds (N)	Van der Waals interactions (N)	Salt-bridges (N)	Super-imposition (RMSD)	Total Score (TS)	Predictive outcome
p.G876V + p.SLIT2(WT)	Increase: 2	Decrease: 2	Increase: 1	Decrease: 2	Decrease: 2	Neutral : 0	2	11	More disrupting (TS > Mean TS)
p.P665S + p.ROBO4(WT)	Increase:2	Decrease:2	Decrease:2	Increase:1	Increase:1	Increase:1	2	11	More disrupting (TS > Mean TS)
p.G829W + p.SLIT2 (WT)	Increase:2	Increase:1	Increase:1	Decrease:2	Decrease:2	Neutral : 0	2	10	More disrupting (TS > Mean TS)
p.G576R + p.ROBO4 (WT)	Increase:2	Decrease:2	Decrease:2	Neutral : 0	Increase:1	Increase:1	2	10	More disrupting (TS > Mean TS)
Mean = 9.222									
p.S636C + p.ROBO4 (WT)	Increase:2	Increase:1	Increase:1	Increase:1	Increase:1	Increase:1	2	9	Less disrupting (TS < Mean TS)
p.P54R + p.ROBO4(WT)	Increase:2	Decrease:2	Neutral : 0	Neutral : 0	Decrease:2	Neutral : 0	0	6	Less disrupting (TS < Mean TS)
p.C215Y + p.ROBO4 (WT)	Increase:2	Increase:1	Increase:1	Neutral : 0	Neutral : 0	Neutral : 0	0	4	Less disrupting (TS < Mean TS)

Table 4. Weighted score matrix for the analyses of docked complexes with the final gradation of the variants. An assessment of the protein–protein interaction influenced by somatic variants.

contribute to a more significant perturbation/disruption in the protein–protein interaction compared to their respective wild type docked complexes. The results of the analysis for both lung cancer and non-lung cancer datasets are depicted in (Table 4).

Finally, pairwise structural alignment of the mutant-to-wild type on wild type-to-wild type docked complexes followed by calculation of RMSD values of the alignment in UCSF-Chimera revealed true nature of the variants on structural distortions and alteration in the global conformation of the mutant complex and PPIs. Higher the RMSD value more significant is the deviation from the wild type conformation. Thus, to grade the variant, a mean score of the RMSD values of the docked complexes for ROBO1.IG1:SLIT2.D2 and ROBO4.IG1-2:SLIT2.D2 complexes were calculated separately. The RMSD value of each alignment was categorised into three groups, i.e. 'Neutral'—RMSD same as the mean value, 'High disruptor'—RMSD higher than the mean value and 'Low disruptors-RMSD lower than the mean value. Each category is depicted pictorially in the following figure (Fig. 2). The residue-specific interactions for each of the domain-domain docked complexes mentioned in Fig. 2 are summarised in Supplementary Information, Table S5.

Similarly, mutant to wild type whole protein–protein interaction analyses were done following the same protocol for the prioritised variants residing outside the interaction domains of the receptors and ligands. Out of 15 variants and 20 docked complexes, all were found to have increased free energy change compared to the respective wild types. A decrease in binding affinity was found for 12 docked complexes of 9 distinct variants, while the remaining showed an increase in binding affinity. Analysis over ePISA webserver revealed 5 docked complexes for 3 distinct variants to have decreased gain in solvation energy upon docking, 1 variant had a neutral effect on solvation energy. In comparison, 14 docked complexes with 11 distinct variants show an increase in solvation energy as compared to their respective wild type docked complexes of ROBO1.IG1:SLIT2.D2 and ROBO4.IG1-2:SLIT2.D2. About 8 variants each in the categories of interactions mentioned above cause a decrease in the number of hydrogen bonds, van der Waals interactions, and salt bridges, respectively that confers more significant perturbation/disruption in protein–protein interaction of the receptor and ligand. Following the same protocol, superimposition analysis was performed in UCSF-Chimera by a pairwise structural alignment that revealed 4 variants of ROBO1 and 3 variants of SLIT2 to have a more significant disruptive effect on ROBO1:SLIT2 docking interaction while 2 variants of ROBO1 and 2 variants of SLIT2 have a lesser disruptive effect on ROBO1:SLIT2 docking interaction.

Similarly, 4 variants of ROBO4 and 2 variants of SLIT2 have a more significant disruptive effect on ROBO4:SLIT2 docking interaction whereas 3 variants of SLIT2 have a lesser disruptive effect on ROBO4:SLIT2 docking interaction (Fig. 3). The residue-specific interactions for each of the protein–protein docked complexes mentioned in Fig. 3 are summarised in Supplementary Information, Table S6. The results for both lung cancer and non-lung cancer datasets are summarised (Table 4).

A cumulative weighted score of each post-docking analysis outcomes for each mutant was calculated, followed by the estimation of the mean score of a particular docking category (Table 1). The variant was graded as "more disruptive" when its cumulative score was above the mean and as "less disruptive" when its cumulative score was below mean tending towards '0′ (wild type) (Table 4 and Fig. 4).

Discussion

Metastasis is the process of tumour cells migrating to secondary sites through blood or lymph vessels. In lung cancer, SLIT-ROBO pathway inhibits tumour cell metastasis through SLIT/ROBO/ Myo9b/RhoA signalling axis[28]. Myosin 9b (Myo9b) is a RhoGAP protein, which modulates protrusion of lamellipodia and retraction of the tail by blocking RhoA activation in migrating immune cells[28]. Myo9b is overexpressed in lung cancer tissue, and it interacts with ROBO1-ICD through its RhoGAP domain, which suppresses the Myo9b RhoGAP activity[28]. Thus, the regular interaction of SLIT2 with ROBO1 suppresses Myo9b leading to inhibition of metastasis[28].
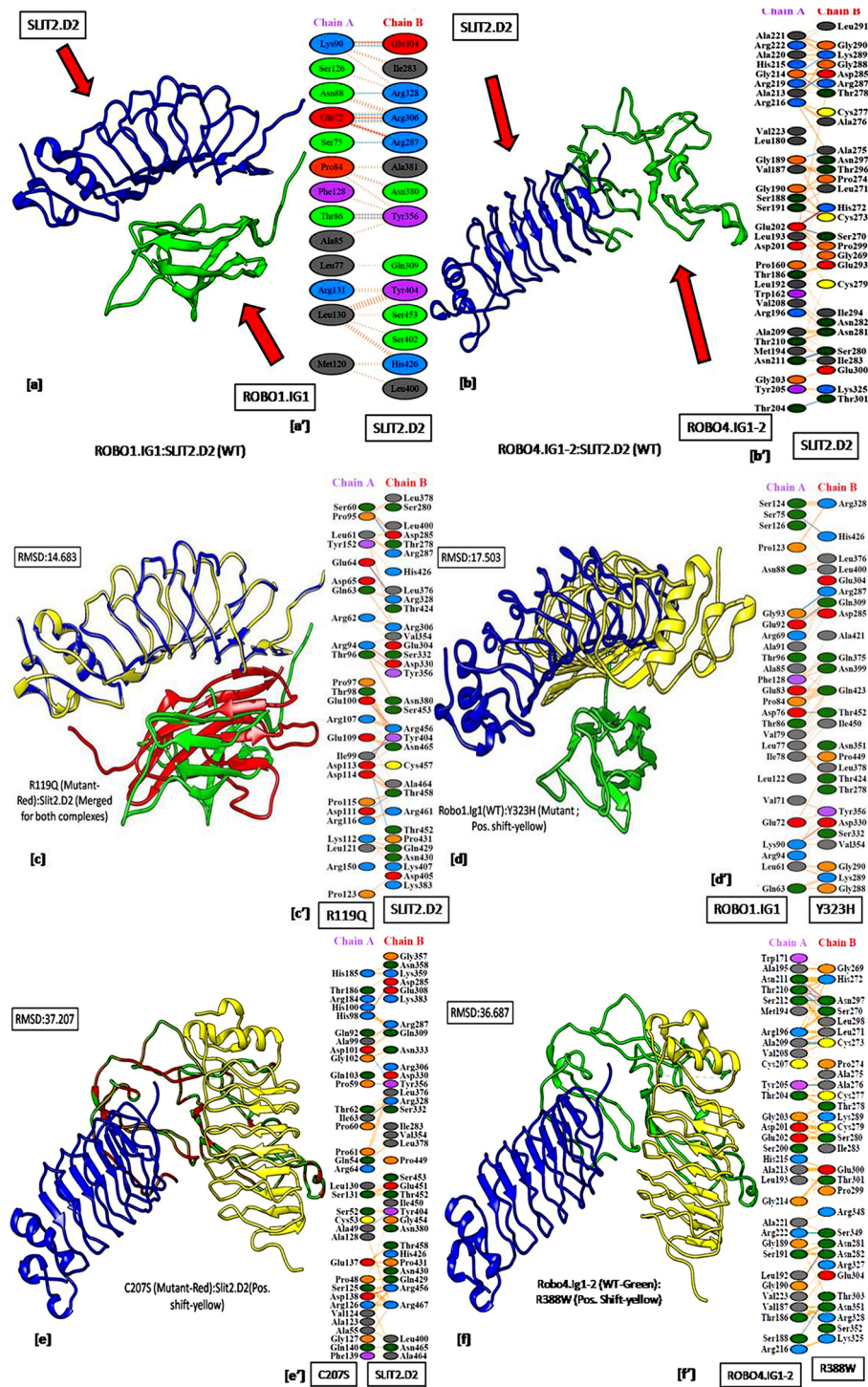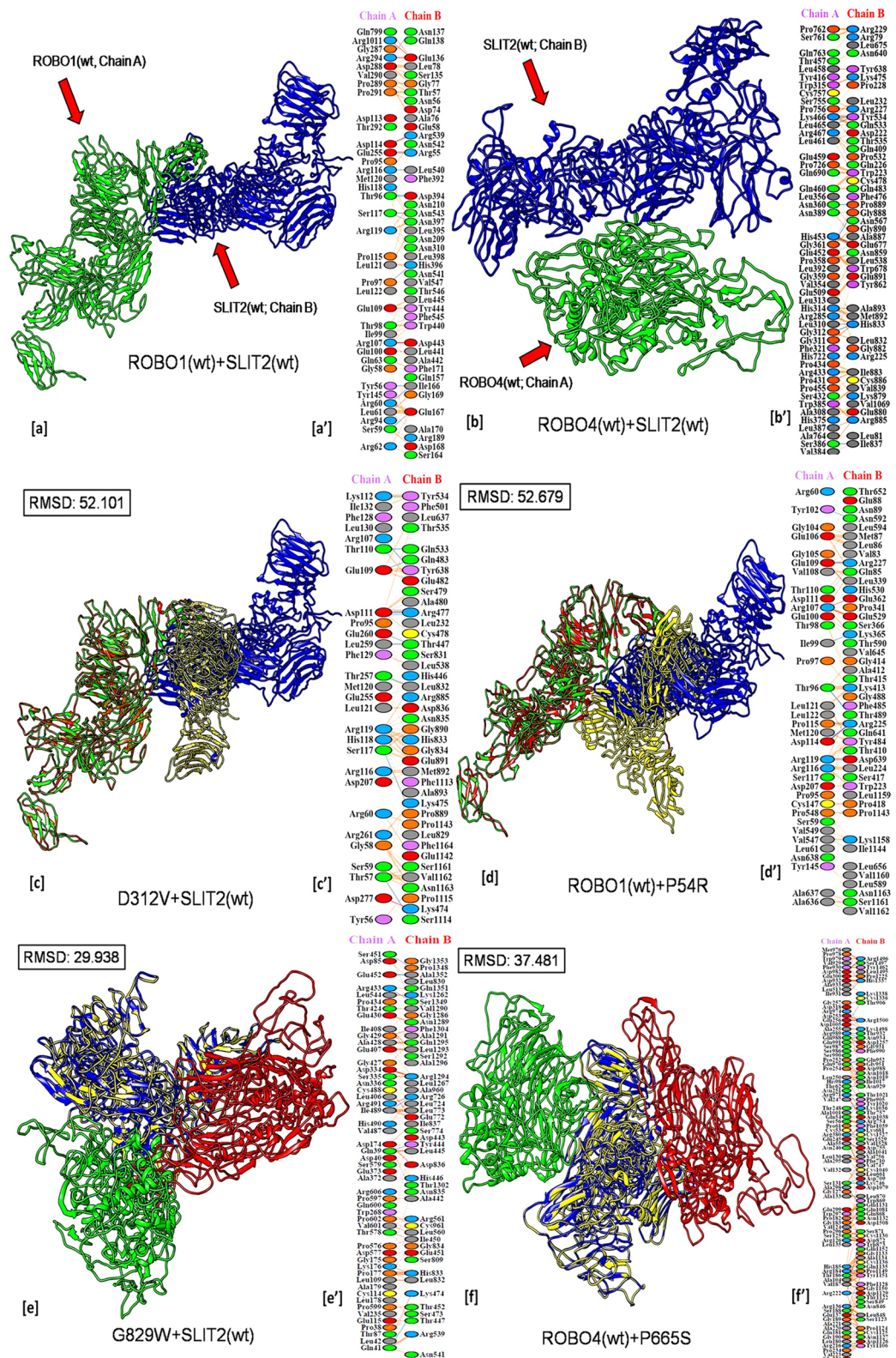
Figure 2. Domain-Domain interaction of wild type-to-wild type and mutant-to-wild type structures. (**a**) ROBO1.IG1(wt):SLIT2.D2(wt) docked complex, (**a′**) ROBO1.IG1(wt): SLIT2.D2(wt) interaction plot; (**b**) ROBO4.IG1-2(wt):SLIT2.D2(wt) docked complex, (**b′**) ROBO4.IG1-2(wt):SLIT2.D2(wt) interaction plot; (**c**) R119Q:SLIT2.D2(wt) docked complex, (**c′**) R119Q: SLIT2.D2(wt) interaction plot; (**d**) ROBO1.IG1(wt):Y323H docked complex, (**d′**) ROBO1.IG1(wt): Y323H interaction plot; (**e**) C207S:SLIT2.D2(wt) docked complex, (**c′**) C207S: SLIT2.D2(wt) interaction plot; (**f**) ROBO4.IG1-2(wt):R388W docked complex, (**b′**) ROBO4.IG1-2(wt):R388W interaction plot; Red depicts ROBO.IG1 and ROBO4.IG1-2 of mutant complex, Green depicts wild type ROBO1.IG1 and ROBO4.IG1-2, Blue depicts wild type SLIT2.D2, and Yellow depicts SLIT2.D2 of the mutant complex.

Figure 3. Protein–Protein interaction of wild type-to-wild type and mutant-to-wild type structures. (**a**) ROBO1(wt):SLIT2(wt) docked complex, (**a′**) ROBO1(wt):SLIT2(wt) interaction plot; (**b**) ROBO4(wt):SLIT2(wt) docked complex, (**b′**) ROBO4(wt):SLIT2(wt) interaction plot; (**c**) D312V:SLIT2(wt) docked complex, (**c′**) D312V: SLIT2(wt) interaction plot; (**d**) ROBO1(wt):P54R docked complex, (**d′**) ROBO1(wt): P54R interaction plot; (**e**) G829W:SLIT2(wt) docked complex, (**e′**) G829W: SLIT2(wt) interaction plot; (**f**) ROBO4(wt):P665S docked complex, (**b′**) ROBO4(wt):P665S interaction plot; Red depicts ROBO1 and ROBO4 of mutant complex, Green depicts wild type ROBO1 and ROBO4, Blue depicts wild type SLIT2 and Yellow depicts SLIT2 of mutant complex.
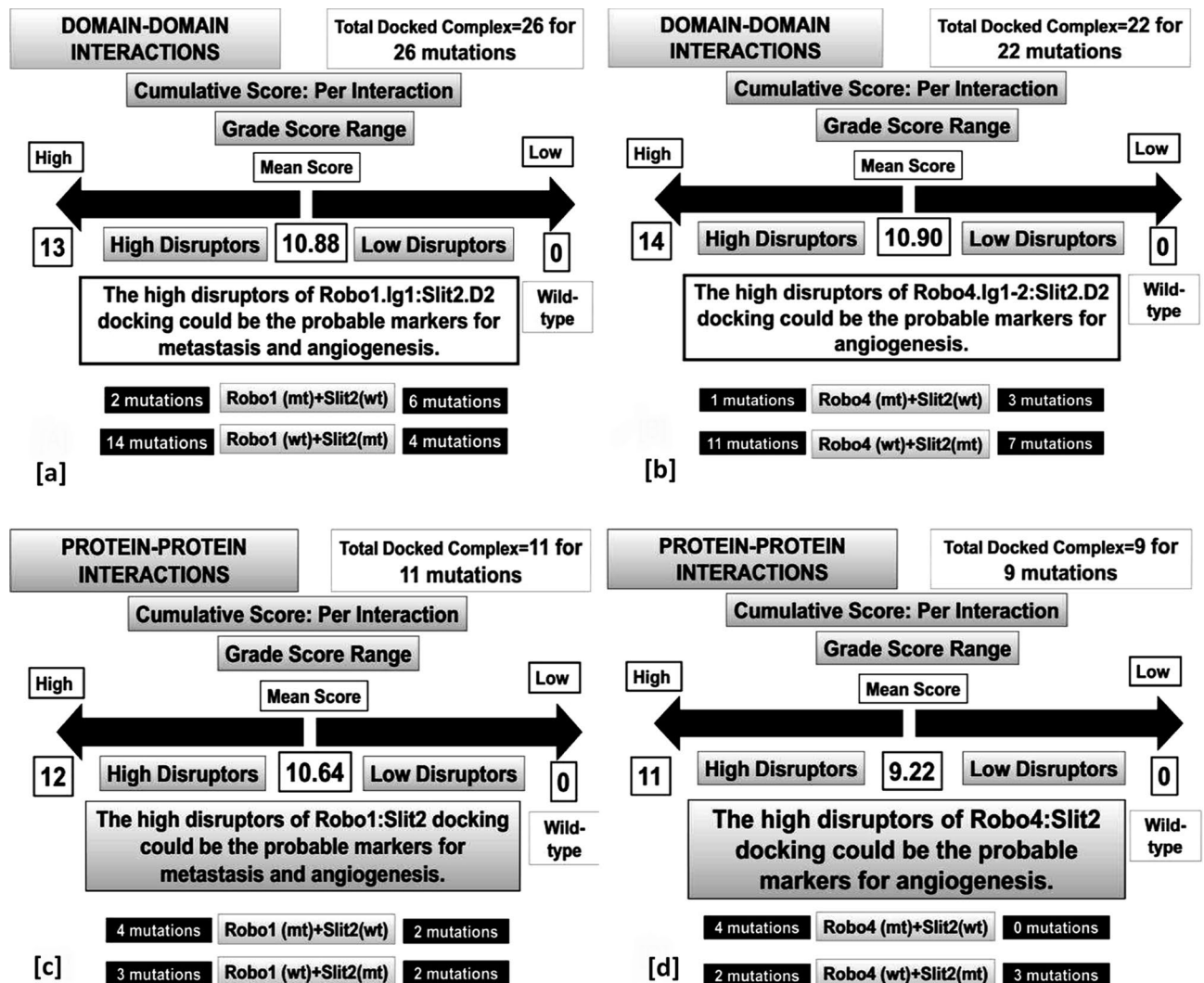
Figure 4. Grading of the variants for their potential impact on the interaction of the receptors to its cognate ligand. The assessment of the deviations between the interaction of wild type-to-wild type and mutant-to-wild type docked complexes. (**a**) ROBO1.IG1:SLIT2.D2 interaction, (**b**) ROBO4.IG1-2:SLIT2.D2 interaction, (**c**) ROBO1:SLIT2 interaction, and (**d**) ROBO4:SLIT2 interaction. The categorisation and distribution of the variants compared to wild types based on their cumulative scores were depicted.

Tumour progression and survival depends on its ability to induce angiogenesis to get access to oxygen and nutrients for metabolism. ROBO1 and ROBO4 were found to express in the vascular endothelial cells and may have some role in tumour angiogenesis, but very little is known to date. Differential expression of SLIT/ROBO has been found in various phenotypes of endothelial cells where ROBO1 induces long and thin actin fibres, whereas ROBO4 was found to induce short and thick actin fibres along with the membrane ruffles[65]. SLIT2 interacts, with its cognate receptor ROBO1 in the human umbilical vein endothelial cells (HUVECs) and tumour-associated endothelial cells[9]. Studies have shown that the ROBO4 receptor imparts inhibitory signal to endothelial migration, tube formation, and vascular permeability through its interaction with SLIT2, suggesting its negative regulatory role in angiogenesis in HUVEC cells[66–68] and an animal model for ocular angiogenesis[69].

Our study is the first report that involves the characterisation of somatic variants of ROBO1.IG1, ROBO4.IG1-2 and SLIT2.D2 domains reported in various cancers including lung cancer, and their role in domain-domain interactions of the receptors (ROBO1.IG1 and ROBO4.IG1-2) and the ligand (SLIT2.D2).

The entire set of missense somatic variants was analysed using the tools mentioned in the materials and method segments, and the results were further categorised using a weighted scoring scheme. The method of variant prioritisation followed in this work conglomerates and assimilated predictions of multiple tools, thus reducing chances of biased misinterpretations. Our results showed that most of the variants were predicted to be "Passenger" in nature by the FATHMM algorithm for cancer driver identification. However, the same variants were predicted to be "Pathogenic" by the FATHMM algorithm for assessing pathogenicity of the coding variants. Thus, it indicates their potential role in the molecular interaction between the ROBO1/4 receptors with its cognate ligand SLIT2.

The Gromos ffG43a2 forcefield has been used previously to study the dynamics of Amyloidβ-Peptides[70] in which the unstructured regions of the peptides have been found to exhibit high mobility in conformational sampling space which allows them better chances of interactions with cognate partners. In our study, we have also found that the unstructured residues though very minimal in number exhibit higher mobility and thus contribute towards the interaction dynamics of the proteins, which is following the observations of a previously published report[71]. Further, structure-based analysis of the selected variants was done to assess their impact on conformational and structural changes within the protein models. The structural attributes analysed, includes a change in hydrogen bond and steric clash pattern and number between wild type and mutants[72–76]. Thus, it substantiates the local structural distortion upon variants that might be detrimental to the integrity and functionality of the proteins. Change in amino acid localisation within the respective domains correlating to its accessible surface area for solvent action depicts the severity of the variants on structure–function alterations. However, our analysis for variant-induced alteration in ligand binding potential of the domains revealed the impact of such variants on the global conformational distortion and modifications of the sub-regions/pockets within the protein structure for possible therapeutic interventions targeting the ROBO/SLIT pathway in cancer pathogenesis. The variants could create or ablate a ligand/drug-binding pocket, which further signifies the role of the variants as candidate molecular targets in lung cancer biology. Our weighted scoring and subsequent grading of the variants provided insight into their pathogenicity and probable disrupting effect on molecular docking. Post-docking comparative analysis revealed the disruptive effect of the variants on docking dynamics by comparing the change in free energy, binding affinity, gain in solvation energy, bond numbers between mutant and wild type docked complexes. Some variants were found to increase the stability of the docking that indicates their potential role in altering the potential of downstream signal transduction. The pairwise structural alignment of the mutant docked complex over wild type docked complex provided more profound insight into the variant-induced structural distortions that are related to the changes in free energy, binding affinity, solvation energy, and bond numbers. Recurrent variants (p.R119Q and p.R131C of ROBO1 and p.R462C, p.A340T, p.P322L, p.R461C, p.R388W, p.R420W, and p.G576R of SLIT2) found that implies the pathogenic significance of the variants across tumour types. Thus, it signifies the potential role of the variants in critical cancer regulatory pathways, such as metastasis and angiogenesis. Our observation mostly points out towards the inter-group amino acid substitutions with critical localisation within the loops and hinge regions of the protein structures to be responsible for its potential for docking disruption. Another aspect includes the differential surface electrostatic potential induced by the variants, like p.D312V, p.G154E, and p.D435V to be an essential determiner for their impact on molecular interaction between the cognate receptor and ligand. Often the amino acids belonging to the same group, viz. p.G82V, p.I78S, also imparts significant influence on the molecular interaction between the proteins, which could be due to their localisation with the protein structure and not because of the nature of the substitution. These variants could be the potential players for disruption in ROBO/SLIT interaction, leading to increased aggressiveness and metastasis in lung cancer.

Molecular dynamics simulation of the docked complexes for 40 ns using an explicit solvent system — TIP3P water model[77] in Gromos96 program[78] followed by coarse-grained simulations was done to evaluate their stability. Once the simulated complexes were generated, the same was analysed for their structural similarities with the original docked complexes, and then the interaction data were also analysed. Majority of structural comparisons yielded averaged RMSD values of < 3.0 Å, which is an acceptable range for the comparisons of protein–protein interaction complexes which are generated through docking and then subjected to simulation[79]. Apart from that, the residue-specific Root Mean Square Fluctuation Values (RMSF) (Supplementary Information, Table S7) were also generated for each of the complexes. All the results indicate that the original docked complexes, as well as the simulated complexes, were conformationally stable and thus, the interaction data and the residues and sites of interactions can be considered to be a good fit for understanding Robo-Slit complexes.

Differences in the mutational landscape exist across various cancer types among between the primary and metastatic cancers. Although there is an extensive commonality in the mutations between the primary and metastatic tumours, some are specific for metastasis. Mutation sites in TP53 vary between primary and metastatic sites in Colorectal cancer, Glioma, NSCLC, and Prostate Adenocarcinoma[80]. The approach of our work paves the way to predict if mutations in ROBO1/4 and SLIT2 have the potential to disrupt their cognate physical interaction that in turn, could regulate the crucial hallmarks of cancer viz. tumour spread and metastasis. Such mutations could be the potential driver or regulator mutations for metastasis and cancer progression. These mutations could serve as molecular targets to design effective novel anticancer therapeutic approaches by precision and personalised medicine. Further, investigations are needed to confirm the computational data in vitro and in vivo, which is beyond the scope of the work presented.

This study deals with the first-ever report on the structural insight of ROBO4.IG1-2 and SLIT2.D2 interaction in a computational model. The structure of ROBO4 is distinctively different from the other ROBO receptors with only two *Immunoglobulin-like* (Ig-like) domains and two fibronectin domains. The details of the molecular interaction of ROBO4.IG1-2 with SLIT2.D2 in our study could provide shreds of evidence for the ROBO4 mediated angiogenesis in lung cancer pathogenesis.

The results of the present study validate a well designed computational pipeline to detect the pathogenic variants conferring disruption to molecular docking of candidate proteins and reduce the exorbitant expenses in the experimental determination of causal variants. With the advent of high-throughput sequencing technologies, new somatic variants have been identified in lung cancer tissue that was previously grouped under other cancers. Therefore, we considered all the variants for in silico assessment of their pathogenicity on docking disruption, which can be further explored for any real-time evidence in metastasis and angiogenesis associated with lung cancer pathophysiology.

Data availability

All data generated or analysed during this study are included in this published article and its Supplementary files.

References

1. Park, K. S. *et al.* A crucial requirement for Hedgehog signaling in small cell lung cancer. *Nat. Med.* **17**, 1504–1508. https://doi.org/10.1038/nm.2473 (2011).
2. Valastyan, S. & Weinberg, R. A. Tumor metastasis: molecular insights and evolving paradigms. *Cell* **147**, 275–292. https://doi.org/10.1016/j.cell.2011.09.024 (2011).
3. Brantley-Sieders, D. M. *et al.* Angiocrine factors modulate tumor proliferation and motility through EphA2 repression of Slit2 tumor suppressor function in endothelium. *Cancer Res.* **71**, 976–987. https://doi.org/10.1158/0008-5472.CAN-10-3396 (2011).
4. Geutskens, S. B., Hordijk, P. L. & van Hennik, P. B. The chemorepellent Slit3 promotes monocyte migration. *J. Immunol.* **185**, 7691–7698. https://doi.org/10.4049/jimmunol.0903898 (2010).
5. Prasad, A., Qamri, Z., Wu, J. & Ganju, R. K. Slit-2/Robo-1 modulates the CXCL12/CXCR4-induced chemotaxis of T cells. *J. Leukoc. Biol.* **82**, 465–476. https://doi.org/10.1189/jlb.1106678 (2007).
6. Wu, J. Y. *et al.* The neuronal repellent Slit inhibits leukocyte chemotaxis induced by chemotactic factors. *Nature* **410**, 948–952. https://doi.org/10.1038/35073616 (2001).
7. Stella, M. C., Trusolino, L. & Comoglio, P. M. The Slit/Robo system suppresses hepatocyte growth factor-dependent invasion and morphogenesis. *Mol. Biol. Cell* **20**, 642–657. https://doi.org/10.1091/mbc.E08-03-0321 (2009).
8. Koch, A. W. *et al.* Robo4 maintains vessel integrity and inhibits angiogenesis by interacting with UNC5B. *Dev. Cell.* **20**, 33–46. https://doi.org/10.1016/j.devcel.2010.12.001 (2011).
9. Marlow, R. *et al.* Vascular Robo4 restricts proangiogenic VEGF signaling in breast. *Proc. Natl. Acad. Sci. U.S.A* **107**, 10520–10525. https://doi.org/10.1073/pnas.1001896107 (2010).
10. Feng, Y., Feng, L., Yu, D., Zou, J. & Huang, Z. srGAP1 mediates the migration inhibition effect of Slit2-Robo1 in colorectal cancer. *J. Exp. Clin. Cancer Res.* **35**, 191. https://doi.org/10.1186/s13046-016-0469-x (2016).
11. Fukuhara, N., Howitt, J. A., Hussain, S. A. & Hohenester, E. Structural and functional analysis of slit and heparin binding to immunoglobulin-like domains 1 and 2 of Drosophila Robo. *J. Biol. Chem.* **283**, 16226–16234. https://doi.org/10.1074/jbc.M800688200 (2008).
12. Hohenester, E. Structural insight into slit-robo signalling. *Biochem. Soc. Trans.* **36**, 251–256. https://doi.org/10.1042/BST0360251 (2008).
13. Morlot, C. *et al.* Structural insights into the Slit-Robo complex. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 14923–14928. https://doi.org/10.1073/pnas.0705310104 (2007).
14. Seiradake, E. *et al.* Structure and functional relevance of the Slit2 homodimerisation domain. *EMBO Rep.* **10**, 736–741. https://doi.org/10.1038/embor.2009.95 (2009).
15. Ballard, M. S. & Hinck, L. A roundabout way to cancer. *Adv. Cancer Res.* **114**, 187–235. https://doi.org/10.1016/B978-0-12-386503-8.00005-3 (2012).
16. Xu, F. *et al.* Whole-exome and targeted sequencing identify ROBO1 and ROBO2 mutations as progression-related drivers in myelodysplastic syndromes. *Nat. Commun.* **6**, 8806. https://doi.org/10.1038/ncomms9806 (2015).
17. Xian, J. *et al.* Inadequate lung development and bronchial hyperplasia in mice with a targeted deletion in the Dutt1/Robo1 gene. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 15062–15066. https://doi.org/10.1073/pnas.251407098 (2001).
18. Mehlen, P., Delloye-Bourgeois, C. & Chedotal, A. Novel roles for Slits and netrins: axon guidance cues as anticancer targets?. *Nat. Rev. Cancer* **11**, 188–197. https://doi.org/10.1038/nrc3005 (2011).
19. Nasarre, P., Potiron, V., Drabkin, H. & Roche, J. Guidance molecules in lung cancer. *Cell Adh. Migr.* **4**, 130–145 (2010).
20. Prasad, A., Paruchuri, V., Preet, A., Latif, F. & Ganju, R. K. Slit-2 induces a tumor-suppressive effect by regulating beta-catenin in breast cancer cells. *J. Biol. Chem.* **283**, 26624–26633. https://doi.org/10.1074/jbc.M800679200 (2008).
21. Huang, T., Kang, W., Cheng, A. S., Yu, J. & To, K. F. The emerging role of Slit-Robo pathway in gastric and other gastro intestinal cancers. *BMC Cancer* **15**, 950. https://doi.org/10.1186/s12885-015-1984-4 (2015).
22. Zhou, W. J. *et al.* Slit-Robo signaling induces malignant transformation through Hakai-mediated E-cadherin degradation during colorectal epithelial cell carcinogenesis. *Cell Res.* **21**, 609–626. https://doi.org/10.1038/cr.2011.17 (2011).
23. Dallol, A. *et al.* SLIT2, a human homologue of the Drosophila Slit2 gene, has tumor suppressor activity and is frequently inactivated in lung and breast cancers. *Cancer Res.* **62**, 5874–5880 (2002).
24. Kim, H. K. *et al.* Slit2 inhibits growth and metastasis of fibrosarcoma and squamous cell carcinoma. *Neoplasia* **10**, 1411–1420 (2008).
25. Yu, J. *et al.* The neuronal repellent SLIT2 is a target for repression by EZH2 in prostate cancer. *Oncogene* **29**, 5370–5380. https://doi.org/10.1038/onc.2010.269 (2010).
26. Tseng, R. C. *et al.* SLIT2 attenuation during lung cancer progression deregulates beta-catenin and E-cadherin and associates with poor prognosis. *Cancer Res.* **70**, 543–551. https://doi.org/10.1158/0008-5472.CAN-09-2084 (2010).
27. Maiti, G. P. *et al.* Frequent inactivation of SLIT2 and ROBO1 signaling in head and neck lesions: clinical and prognostic implications. *Oral Surg. Oral Med. Oral Pathol. Oral Radiol.* **119**, 202–212. https://doi.org/10.1016/j.oooo.2014.09.029 (2015).
28. Kong, R. *et al.* Myo9b is a key player in SLIT/ROBO-mediated lung tumor suppression. *J. Clin. Invest.* **125**, 4407–4420. https://doi.org/10.1172/JCI81673 (2015).
29. Wen, P. *et al.* USP33, a new player in lung cancer, mediates Slit-Robo signaling. *Protein Cell* **5**, 704–713. https://doi.org/10.1007/s13238-014-0070-z (2014).
30. Yuasa-Kawada, J., Kinoshita-Kawada, M., Rao, Y. & Wu, J. Y. Deubiquitinating enzyme USP33/VDU1 is required for Slit signaling in inhibiting breast cancer cell migration. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 14530–14535. https://doi.org/10.1073/pnas.0801262106 (2009).
31. Taguchi, A. *et al.* Lung cancer signatures in plasma based on proteome profiling of mouse tumor models. *Cancer Cell* **20**, 289–299. https://doi.org/10.1016/j.ccr.2011.08.007 (2011).
32. Park, K. W. *et al.* Robo4 is a vascular-specific receptor that inhibits endothelial migration. *Dev. Biol.* **261**, 251–267 (2003).
33. Pircher, A. *et al.* Favorable prognosis of operable non-small cell lung cancer (NSCLC) patients harboring an increased expression of tumor endothelial markers (TEMs). *Lung Cancer* **81**, 252–258. https://doi.org/10.1016/j.lungcan.2013.04.014 (2013).
34. Houlston, R. S. & Peto, J. The search for low-penetrance cancer susceptibility alleles. *Oncogene* **23**, 6471–6476. https://doi.org/10.1038/sj.onc.1207951 (2004).
35. Bamford, S. *et al.* The COSMIC (catalogue of somatic mutations in cancer) database and website. *Br. J. Cancer.* **91**, 355–358. https://doi.org/10.1038/sj.bjc.6601894 (2004).
36. Apweiler, R. *et al.* UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* **32**, D115-119. https://doi.org/10.1093/nar/gkh131 (2004).

37. UniProt, C. The universal protein resource (UniProt). *Nucleic Acids Res.* **36**, D190-195. https://doi.org/10.1093/nar/gkm895 (2008).
38. Buchan, D. W., Minneci, F., Nugent, T. C., Bryson, K. & Jones, D. T. Scalable web services for the PSIPRED protein analysis workbench. *Nucleic Acids Res.* **41**, W349-357. https://doi.org/10.1093/nar/gkt381 (2013).
39. Meszaros, B., Erdos, G. & Dosztanyi, Z. IUPred2A: context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Res.* **46**, W329–W337. https://doi.org/10.1093/nar/gky384 (2018).
40. Patthy, L. *Protein Evolution* (Blackwell Publishing, Hobken, 2008).
41. Choi, Y. & Chan, A. P. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* **31**, 2745–2747. https://doi.org/10.1093/bioinformatics/btv195 (2015).
42. Ng, P. C. & Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Res.* **31**, 3812–3814 (2003).
43. Adzhubei, I., Jordan, D. M. & Sunyaev, S. R. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr. Protoc. Hum. Genet.* **Chapter 7**, Unit7 20 (2013). https://doi.org/10.1002/0471142905.hg0720s76
44. Calabrese, R., Capriotti, E., Fariselli, P., Martelli, P. L. & Casadio, R. Functional annotations improve the predictive score of human disease-related mutations in proteins. *Hum. Mutat.* **30**, 1237–1244. https://doi.org/10.1002/humu.21047 (2009).
45. Capriotti, E., Fariselli, P. & Casadio, R. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. *Nucleic Acids Res.* **33**, W306–W310. https://doi.org/10.1093/nar/gki375 (2005).
46. Shihab, H. A. *et al.* Predicting the functional, molecular, and phenotypic consequences of amino acid substitutions using hidden Markov models. *Hum. Mutat.* **34**, 57–65. https://doi.org/10.1002/humu.22225 (2013).
47. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2 (1990).
48. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
49. Berman, H. M. *et al.* The protein data bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
50. Webb, B. & Sali, A. Comparative protein structure modeling using MODELLER. *Curr. Protoc. Bioinformatics* **47**(56), 1–32. https://doi.org/10.1002/0471250953.bi0506s47 (2014).
51. Berendsen, H. J. C., van der Spoel, D. & van Drunen, R. GROMACS—a message-passing parallel molecular-dynamics implementation. *Comput. Phys. Commun.* **91**, 43–56. https://doi.org/10.1016/0010-4655(95)00042-e (1995).
52. Lindahl, E., Hess, B. & Van Der Spoel, D. GROMACS 3.0: a package for molecular simulation and trajectory analysis. *J. Mol. Model.* **7**, 306–317. https://doi.org/10.1007/s008940100045 (2001).
53. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta. Crystallogr. D. Biol. Crystallogr.* **66**, 12–21. https://doi.org/10.1107/S0907444909042073 (2010).
54. Guex, N. & Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* **18**, 2714–2723. https://doi.org/10.1002/elps.1150181505 (1997).
55. Ahmad, S., Gromiha, M., Fawareh, H. & Sarai, A. ASAView: database and tool for solvent accessibility representation in proteins. *BMC Bioinform.* **5**, 51. https://doi.org/10.1186/1471-2105-5-51 (2004).
56. Volkamer, A., Kuhn, D., Rippmann, F. & Rarey, M. DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment. *Bioinformatics* **28**, 2074–2075. https://doi.org/10.1093/bioinformatics/bts310 (2012).
57. Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. & Wolfson, H. J. PatchDock and SymmDock: servers for rigid and symmetric docking. *Nucleic Acids Res.* **33**, W363-367. https://doi.org/10.1093/nar/gki481 (2005).
58. Dominguez, C., Boelens, R. & Bonvin, A. M. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.* **125**, 1731–1737. https://doi.org/10.1021/ja026939x (2003).
59. van Zundert, G. C. P. *et al.* The HADDOCK2.2 Web server: user-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.* **428**, 720–725. https://doi.org/10.1016/j.jmb.2015.09.014 (2016).
60. Laskowski, R. A., Jablonska, J., Pravda, L., Varekova, R. S. & Thornton, J. M. PDBsum: structural summaries of PDB entries. *Protein Sci.* **27**, 129–134. https://doi.org/10.1002/pro.3289 (2018).
61. Dehouck, Y., Kwasigroch, J. M., Rooman, M. & Gilis, D. BeAtMuSiC: prediction of changes in protein-protein binding affinity on mutations. *Nucleic Acids Res.* **41**, W333-339. https://doi.org/10.1093/nar/gkt450 (2013).
62. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797. https://doi.org/10.1016/j.jmb.2007.05.022 (2007).
63. Pettersen, E. F. *et al.* UCSF Chimera–a visualisation system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612. https://doi.org/10.1002/jcc.20084 (2004).
64. Naganathan, A. N., Doshi, U., Fung, A., Sadqi, M. & Munoz, V. Dynamics, energetics, and structure in protein folding. *Biochemistry* **45**, 8466–8475. https://doi.org/10.1021/bi060643c (2006).
65. Sheldon, H. *et al.* Active involvement of Robo1 and Robo4 in filopodia formation and endothelial cell motility mediated via WASP and other actin nucleation-promoting factors. *FASEB J.* **23**, 513–522. https://doi.org/10.1096/fj.07-098269 (2009).
66. Jones, C. A. *et al.* Robo4 stabilises the vascular network by inhibiting pathologic angiogenesis and endothelial hyperpermeability. *Nat. Med.* **14**, 448–453. https://doi.org/10.1038/nm1742 (2008).
67. Tian, R. *et al.* Investigation of the regulation of Roundabout4 by hypoxia-inducible factor-1alpha in microvascular endothelial cells. *Invest. Ophthalmol. Vis. Sci.* **56**, 2586–2594. https://doi.org/10.1167/iovs.14-14409 (2015).
68. Yang, Y. C. *et al.* The differential roles of Slit2-exon 15 splicing variants in angiogenesis and HUVEC permeability. *Angiogenesis* **18**, 301–312. https://doi.org/10.1007/s10456-015-9467-4 (2015).
69. Chen, H. *et al.* Slit-Robo signaling in ocular angiogenesis. *Adv. Exp. Med. Biol.* **664**, 457–463. https://doi.org/10.1007/978-1-4419-1399-9_52 (2010).
70. Olubiyi, O. O. & Strodel, B. Structures of the amyloid beta-peptides Abeta1-40 and Abeta1-42 as influenced by pH and a D-peptide. *J. Phys. Chem. B* **116**, 3280–3291. https://doi.org/10.1021/jp2076337 (2012).
71. Song, D., Luo, R. & Chen, H. F. The IDP-specific force field ff14IDPSFF improves the conformer sampling of intrinsically disordered proteins. *J. Chem. Inf. Model.* **57**, 1166–1178. https://doi.org/10.1021/acs.jcim.7b00135 (2017).
72. Chiappori, F., D'Ursi, P., Merelli, I., Milanesi, L. & Rovida, E. In silico saturation mutagenesis and docking screening for the analysis of protein-ligand interaction: the Endothelial Protein C Receptor case study. *BMC Bioinform.* **10**(Suppl 12), 33. https://doi.org/10.1186/1471-2105-10-S12-S3 (2009).
73. Heler, R., Bell, J. K. & Boland, L. M. Homology model and targeted mutagenesis identify critical residues for arachidonic acid inhibition of Kv4 channels. *Channels (Austin)* **7**, 74–84. https://doi.org/10.4161/chan.23453 (2013).
74. Jubb, H. C. *et al.* Mutations at protein-protein interfaces: small changes over big surfaces have large impacts on human health. *Prog. Biophys. Mol. Biol.* **128**, 3–13. https://doi.org/10.1016/j.pbiomolbio.2016.10.002 (2017).
75. Knape, K., Linder, T., Wolschann, P., Beyer, A. & Stary-Weinzinger, A. In silico analysis of conformational changes induced by mutation of aromatic binding residues: consequences for drug binding in the hERG K+ channel. *PLoS ONE* **6**, e28778. https://doi.org/10.1371/journal.pone.0028778 (2011).
76. Sonawane, P., Patel, K., Vishwakarma, R. K., Singh, S. & Khan, B. M. in Silico mutagenesis and docking studies of active site residues suggest altered substrate specificity and possible physiological role of Cinnamoyl CoA Reductase 1 (Ll-CCRH1). *Bioinformation* **9**, 224–232. https://doi.org/10.6026/97320630009224 (2013).
77. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935. https://doi.org/10.1063/1.445869 (1983).

nature research

78. Scott, W. R. P. *et al.* The GROMOS biomolecular simulation program package. *J. Phys. Chem. A* **103**, 3596–3607. https://doi.org/10.1021/jp984217f (1999).
79. Tsai, H. H., Tsai, C. J., Ma, B. & Nussinov, R. In silico protein design by combinatorial assembly of protein building blocks. *Protein Sci.* **13**, 2753–2765. https://doi.org/10.1110/ps.04774004 (2004).
80. Liu, G., Zhan, X., Dong, C. & Liu, L. Genomics alterations of metastatic and primary tissues across 15 cancer types. *Sci. Rep.* **7**, 13262. https://doi.org/10.1038/s41598-017-13650-3 (2017).

Acknowledgements

Author contributions

D.S. and M.S.: Conceptualisation. G.B., D.S.: Data Curation. S.G., D.S., and G.B.: Methodology, Software, and Investigation. M.S.: Project Administration. S.G. and M.S.: Resources. M.S. and S.G.: Supervision. S.G., M.S., G.B. and D.S.: Formal analysis and Validation. D.S.: Writing-Original draft. D.S., S.G. and M.S.: Writing-Reviewing and editing. S.G., M.S.: Supervision. All the authors have reviewed the manuscript and have agreed for publication.

Funding

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-020-78882-2.

Correspondence and requests for materials should be addressed to S.G. or M.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.