MOKCa-3D Database: Functional and Structural Analysis of Missense Mutations in Cancer.

B. HAILE [1], A. CINAR [1], F. BANINI [1], K. ZHIKOVA [1], C. RICHARDSON [2], and F. PEARL [1,*]

[1] Bioinformatics Laboratory, Faculty of Science, Engineering and Medicine, University of Sussex, Brighton, England

[2] Section of Structural Biology, Institute of Cancer Research, London, England.

INTRODUCTION

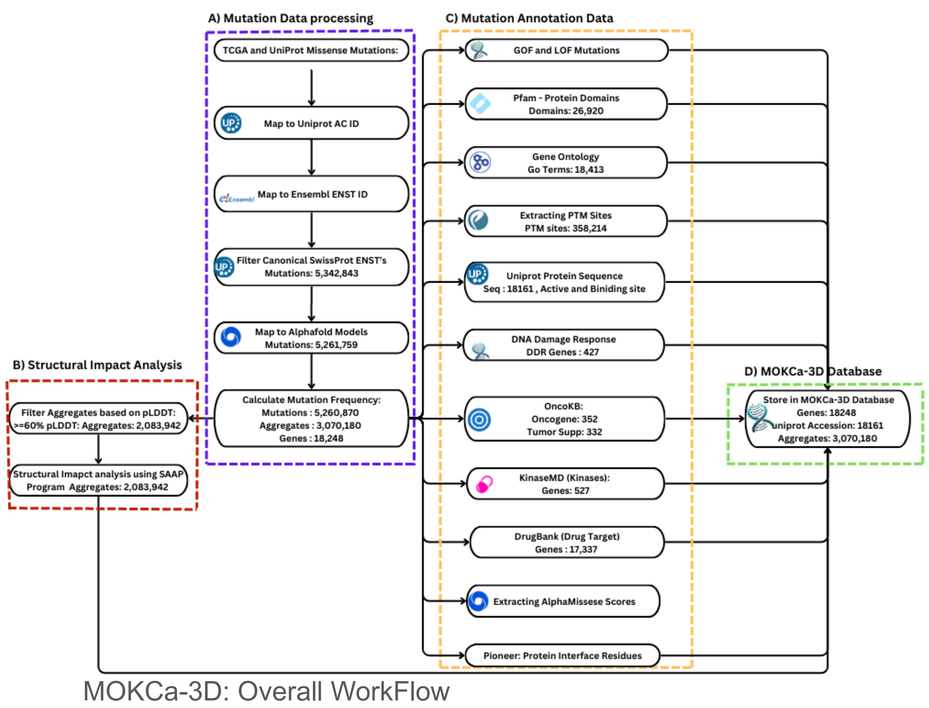
Determining the functional consequence of missense mutations acquired in the development of cancer is critical to the understanding of the evolution and the therapeutic vulnerabilities of an individual tumour. Over 3 millions unique missense mutations are hosted in MOKCa-3D. The results shown here are in whole or part based upon data generated by the TCGA[1] Research Network and UniProt [4]. The MOKCa-3D (https://bioinformaticslab.sussex.ac.uk/MOKCa-3D/) is a web-accessible database that contextualizes somatic missense mutations in cancer, incorporating functional and structural analyses to predict the mutation's impact on the protein function.

AIM

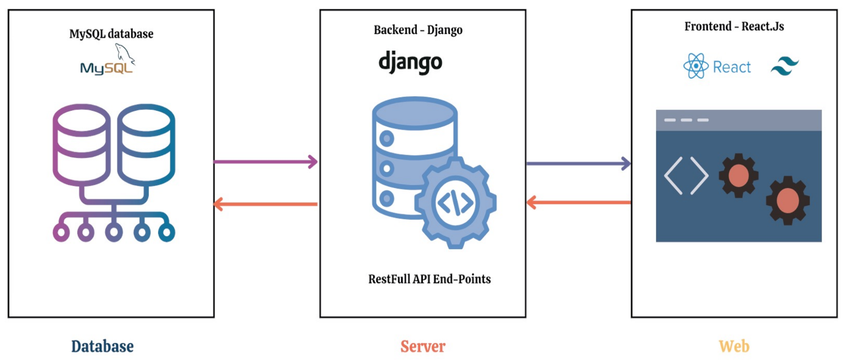
- To provide a comprehensive, accessible platform for exploring the functional and structural consequences of somatic missense mutations across cancer types.

METHOD

- Mutation processing pipeline and structural impact analysis using SAAP3 [2].
- Integration of functional annotations (GOF/LOF, pathogenicity).
- Relational MySQL database served through a Django RESTful API's and ReactJS frontend.



MOKCa-3D: Overall WorkFlow



MOKCa-3D Architectural Design, all mutation data and annotation data are stored in relational database and served via RESTful API (Django), with ReactJS frontend interface for user interaction.

RESULTS

The MOKCa-3D Web Interface

1.Gene Level Annotation

User can search for specific gene using the HGNC symbol, UniProt accession, or Ensembl ENSG ID, or browse the five gene class datasets using the dropdown list (oncogenes, tumour-suppressors, kinase, drug target, and DNA-damage response [3] genes) (Figure 1). Each gene page integrates information from external databases including UniProtKB [4], Ensembl [4], CanSar.ai [6], PhosPhositePlus [7], Amigo2 [8], and InterPro [9]. An interactive feature viewer displays the encoded protein sequence, annotated mutations, PTM-sites, protein interface residues [10], active sites, binding sites, and protein domain. The interactive graph allows users to inspect individual mutations and see the surrounding functional site annotations (Figure 2).

2. Mutation Level Annotation

The user can select individual mutations from the gene-level page (Figure 2). This loads the mutation-level page that provides relevant functional annotation extracted from external databases (Figure 3). These include the predicted pathogenicity [11] of the mutation and, if available, a GOF/LOF assessment curated from literature with PubMed ID links to the relevant articles. Residue modifications, protein-interface residues, active sites, and binding sites, that occur at the position of the mutation or within 3 residues, are presented in a list. An interactive session loads the AlphaFold2 [12] model and highlights the mutated residue on the structure for further 3D exploration. The structural impact analysis of the mutation is presented in a highly summarized manner, highlighting 10 features from SAAP3 that detail how the mutation affects the structure and potentially impairs protein functionality

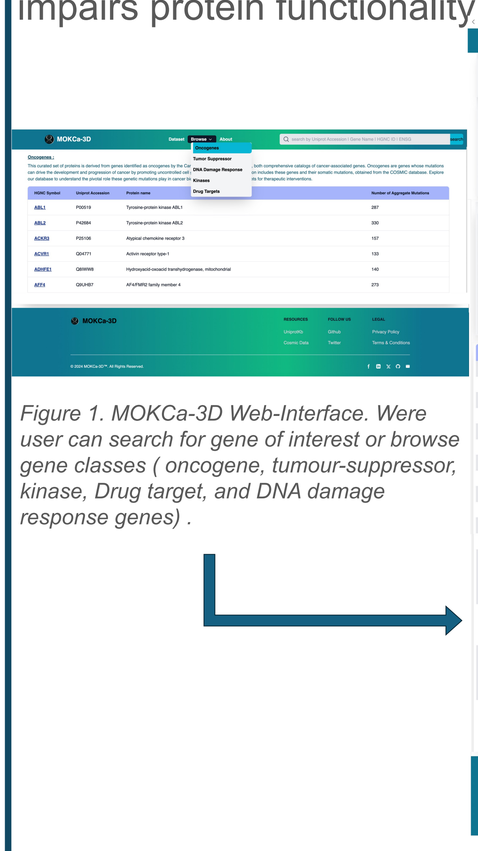


Figure 1. MOKCa-3D Web-Interface. Were user can search for gene of interest or browse gene classes (oncogene, tumour-suppressor, kinase, Drug target, and DNA damage response genes) .

Figure 2. Gene level page: user can search for gene of interest and retrieve all mutations present in more comprehensive way. Annotation data from UniProt, Ensembl, CanSar.ai, PhosPhositePlus, Gene Ontology, Pfam domains are presented.

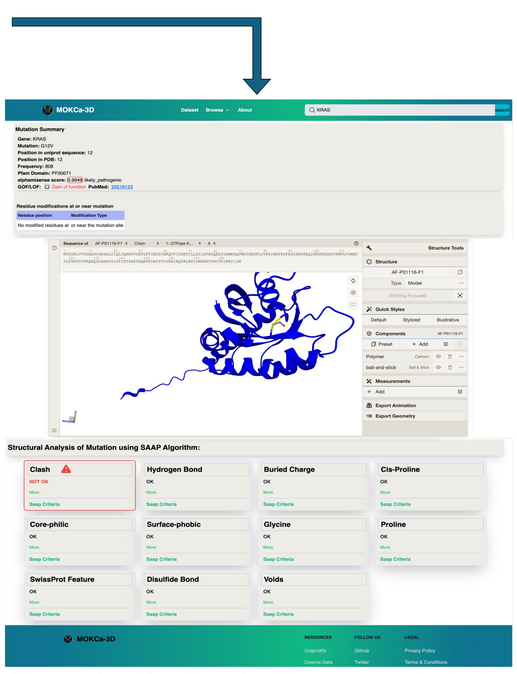


Figure 3. Mutation Level page: In-depth information about the functional consequence of the individual mutation and their structural impact result from SAAP3.

CONCLUSIONS

MOKCa-3D provides in-depth annotation and interpretation of cancer-associated missense mutations by integrating functional and structural insights. This supports molecular understanding of tumour progression and may help in creating personalised therapeutic strategies.

REFERENCES

1. The results shown here are in whole or part based upon data generated by the TCGA Research Network: https://www.cancer.gov/tcga.
2. Nouf S Al-Numair, A.C.M., The SAAP pipeline and database: tools to analyze the impact and predict the pathogenicity of mutations. BMC Genomics, 2013. 14.
3. Pearl, L.H., et al., Therapeutic opportunities within the DNA damage response. Nat Rev Cancer, 2015. 15(3): p. 166-80.
4. UniProt, C., UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Res, 2023. 51(D1): p. D523-D531.
5. Harrison, P.W., et al., Ensembl 2024. Nucleic Acids Res, 2024. 52(D1): p. D891-D899.
6. Gingrich, P.W., et al., canSAR 2024-an update to the public drug discovery knowledgebase. Nucleic Acids Res, 2025. 53(D1): p. D1287-D1294.
7. Hornbeck, P.V., et al., PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. Nucleic Acids Res, 2015. 43(Database issue): p. D512-20.
8. Carbon, S., et al., AmiGO: online access to ontology and annotation data. Bioinformatics, 2009. 25(2): p. 288-9.
9. Paysan-Lafosse, T., et al., InterPro in 2022. Nucleic Acids Res, 2023. 51(D1): p. D418-D427.
10. Xiong, D., et al., A structurally informed human protein-protein interactome reveals proteome-wide perturbations caused by disease mutations. Nat Biotechnol, 2024.
11. Cheng, J., et al., Accurate proteome-wide missense variant effect prediction with AlphaMissense. Science, 2023. 381(6664): p. eadg7492.
12. Varadi, M., et al., AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res, 2022. 50(D1): p. D439 D444.

CONTACT INFORMATION

Dr. Frances M G Pearl

Email: f.pearl@sussex.ac.uk

Tel: +44 (0) 1273876631