Identification of hub genes in common cancers of women in India and targeting for the search of anticancer agent from Punica granatum phytoconstituent using interaction network analysis and virtual screening

Anamika Mishra, Viswajit Mulpuru and Nidhi Mishra

Department of Applied Science, Indian Institute of Information Technology Allahabad, Prayagraj, India

Communicated by Ramaswamy H. Sarma

ABSTRACT

Cancer is one of the most dreadful diseases across the globe, with the advancement in this field a great advent has been achieved in treating cancer by various therapies like chemotherapy, radiation therapy, hormone therapy, gene therapy, and many more but also the most serious concern associated with the available treatments are the toxicities or the side effects linked to them, apart from this the treatment of many malignancies are still not available, because of these such issues, tremendous research is still going on in the whole world to find a better and more potent treatment option for cancer. Cancer develops due to the synergistic effects of both genetic and epigenetic factors. The mutations that change the normal functioning of the genes are responsible for cancer. Various genes are associated with cancers; many genes are commonly found to be mutated in diverse cancer types. In the present work, the genetic co-relation among the top five common cancers in Indian women has tried to be established, after that the identification of the hub gene was carried out with the use of CytoHubba module of Cytoscape. The hub gene product signaling pathway was then targeted for molecular docking with phytoconstituents of Punica granatum while the stability of the docked protein and ligand complex was validated through Molecular Dynamics Simulation.

INTRODUCTION

Cancer a disease characterized by uncontrolled proliferation of cells, metastasis, and invasion is known to affect a large population across the whole world; both genetic and environmental carcinogens (epigenetic factors) are the causative agent of cancer, cancer is a disease that not only affects the physical health of an individual rather it affects the social and emotional aspects of the life of an individual (Ataollahi et al., 2015; Blackadar, 2016). The five most common cancers in women in India are breast, oral, cervical, lung, and gastric (Cancer Statistics, n.d.).

Breast cancer is the most commonly occurring malignancies in the population of women and it is among the most commonly prevailing cancer of all types in the whole world, breast cancer at the molecular level shows heterogeneity implying that various molecular factors such as P53, ERBB2, EGFR, ESR1, MYC, etc., collectively contribute to the development of breast cancer (Harbeck, 2019). Breast cancer has been categorized into three major categories namely hormone receptor-positive/ERBB2 negative (HR+/ERBB2-), ERBB2+, and triple-negative (Waks & Winer, 2019). The various therapies available for breast cancer are radiation therapy, surgical removal, hormonal therapy, chemotherapy, and immunotherapy (Harbeck, 2019).

Oral cancer is the term used collectively for the cancer of the oral cavity and pharynx, it is a neoplasm malignant in nature, and it is one of the most commonly occurring cancers globally. Factors like smoking, alcohol consumption, UV radiation, human papillomavirus (HPV) are some of the risk factors for oral cancer. It is a multifactorial disease where many genetic factors contribute to the development of oral cancer, some of these genetic factors are TP53, NOTCH1, EGFR, CDKN2A, STAT3, Cyclin D1, Rb, VEGF, PDGF, IL8, TGF-b, etc. Despite extensive research in this field survival rate is quite low in this cancer (Cervino et al., 2019; Kumar et al., 2016; Rivera, 2015).

Cervical cancer is one of the commonly occurring malignancies among women. The prevalence of cervical cancer is more in low to middle-income countries (e.g. in African countries, Latin America, and Asian countries) (Balasubramaniam et al., 2019; Srivastava et al., 2014). Various risk factors for cervical cancer are poor personal and sexual hygiene, smoking, nutritional deficiency, sexual relations with multiple partners, sexual activity at an early age, and infection with Human papillomavirus (Akram Husain & Ramakrishnan, 2016; Balasubramaniam et al., 2019). The papillomavirus-encoded proteins affect the normal functioning of host proteins that functions to prevent tumor formation. E6 and E7 are major oncoproteins encoded by HPV, they are
overexpressed in the host, and they affect the normal functioning of tumor suppressor genes like p21, p27, p16, p53, pRb, and CDKs. Some other molecules that contribute to the development of cervical cancer are VEGFA, ERBB2, MCM proteins, TNF, EGFR, IL10, PIK3CA, and PTEN, etc. The molecular markers of cervical cancer are CDKN2A, Ki-67, c-FLIP, etc. The molecular markers of cervical cancer are CDKN2A, Ki-67, c-FLIP, etc. (Akram Husain & Ramakrishnan, 2016; Balasubramaniam et al., 2019).

Lung cancer causes a large number of deaths globally; two main types of lung cancer are small-cell lung cancer and non-small cell lung cancer this has been further divided into three categories squamous cell carcinoma, adenocarcinoma, and substantial cell lung cancer. Smoking is the main risk factor for lung cancer, while some other risk factor for lung cancer is air pollutants like radon gas, asbestos, passive smoking. The genetic factor for the development of lung cancer is p53, EGFR, RRM1, ERCC1, etc. (Houssein, 2019; Jaggi, 2017).

In the present work genetic co-relation among these five cancers was tried to establish, by generating a Protein-protein interaction (PPI) network on the STRING database to find out the most crucial gene among all using the CytoHubba module of Cytoscape. After the identification of the most crucial gene, and validating the function of the gene, we perform docking analysis to look for a potential molecule from the phytoconstituents of Punica granatum that would target the signaling pathway of identified hub gene hence can be considered as an anticancer agent. The stability of the docked complex of Punica granatum phytoconstituent with the inhibitor of our gene of interest was established through Molecular Dynamics Simulation which was performed on GROMACS.

**Material and methods**

1. **Collections of gene data from NCBI**: The genes which are involved in the development of the five most common cancers among women in India namely breast, oral, cervical, lung and gastric were downloaded from NCBI and the genes were sorted according to gene weight. Total 4380 genes were associated with breast cancer, 329 genes were associated with oral cancer, 3091 genes were associated with lung cancer, 1151 genes were associated with cervical cancer and 2600 genes were associated with gastric cancer.

2. **Finding of common genes**: The genes which were common among the five above said cancers were identified by generating a Venn diagram on an online tool available at [http://bioinformatics.psb.ugent.be/cgi-bin/ls/article/Venn/venn.html](http://bioinformatics.psb.ugent.be/cgi-bin/ls/article/Venn/venn.html). A total of 144 genes were turned out to be common among the above mentioned cancers.

3. **Construction of PPI and search for the hub gene**: The common 144 genes were used to construct a PPI network on Cytoscape; the Hub genes were identified using the CytoHubba tool of Cytoscape.

4. **Docking analysis**: The identified hub gene was then considered for molecular docking to search for a potential candidate that could be used in cancer treatment. The docking analysis was performed on the GLIDE module of Schrodinger. The phytoconstituents of Punica granatum were used as the ligands for the docking.

5. **Molecular Dynamics Simulation**: The stability of the docked protein and ligand complex was further validated through molecular dynamics simulation, the simulation was performed on GROMACS version 2018.2 for 100 ns (van der Spoel et al., 2005) to validate the stability of the top 2 ligand-protein complex, with CHARMM36 all-atom force field (Huang & MacKerell, 2013) while the online platform of SwissParam web server (Zoete et al., 2011) was used for generating the topology of ligands as CHARMM36 force field do not have the force field parameters for ligands. Before combining the topology files of protein and ligand or solvated or minimized or equilibrated the GROMACS compatible files of protein and ligands were generated with Inhouse ad hoc scripts. TIP3P explicit model of water molecules was used for solvating protein-ligand complexes, while system neutralization was done using Cl- and Na+ ions. The minimization step was done using the steepest descent algorithm; the step was run till the maximum force is less than 10.0kJ/mol. The NVT and NPT ensembles were used for the equilibration of the system with a simulation time of 100ps for each. The process of equilibration was performed at 1 bar pressure and 300K temperature, after that MDS of 100ns was performed using leapfrog algorithm with the integration time step of 2fs, and trajectory was saved at every 30ps.

6. **Trajectory analysis**: The GROMACS analysis utilities were used for the analysis of molecular dynamics trajectories to compute the results. The Root Mean Square Deviation (RMSD) between the initial and simulated structure was calculated to depict the stability of the protein complex while the RMSF was calculated to determine the rigidity of the secondary structure.
Result and discussion

A total of 144 genes were common among the top 5 commonly occurring cancers in Indian Women (Figure 1) as determined through the construction of the Venn diagram.

A PPI network was generated using these 144 common genes, for which the genes uploaded on STRING Database, proteins interact with each other to carry out various cellular processes, this database tells the association among the proteins based on their functionality (Szklarczyk et al., 2015), the network was then further analyzed on Cytoscape which is a very popular freely available platform to analyze various network, it has many associated plug-ins such as MCODE, GeneMANIA, NetworkAnalyzer, BINGO, CytoHubba, etc. to analyze the various parameters of a network (Saito et al., 2012; Shannon et al., 2003).

These plugins can score and rank the nodes by network features, but most of these plugins do not include important and recently developed methods. Further, owing to the fact that the biological networks are heterogeneous, it is important to analyze a biological network using more than one analytical method for reliable validation.

The NetworkAnalyzer is a plugin that simply provides the topological details of a network including the number of nodes, edges and connected components, the network diameter, radius, density, centralization, heterogeneity, clustering coefficient, the characteristic path length, and degrees

Table 1. List of top 10 genes with DEGREE and MNC scoring methods.

| Rank | Name | Score |
|------|------|-------|
| 1    | TP53 | 107   |
| 2    | EGFR | 98    |
| 3    | AKT1 | 97    |
| 4    | MYC  | 97    |
| 5    | IL6  | 93    |
| 6    | VEGFA| 92    |
| 7    | CASP3| 85    |
| 8    | CCND1| 85    |
| 9    | CDH1 | 84    |
| 10   | TNF  | 84    |

Table 2. List of top 10 genes with CLOSENESS scoring method.

| Rank | Name | Score |
|------|------|-------|
| 1    | TP53 | 116.5 |
| 2    | EGFR | 112   |
| 3    | AKT1 | 111.5 |
| 4    | MYC  | 111.5 |
| 5    | IL6  | 109.5 |
| 6    | VEGFA| 108.8 |
| 7    | CASP3| 105.3 |
| 8    | CCND1| 105.3 |
| 9    | CDH1 | 105   |
| 10   | TNF  | 105   |

Table 3. List of top 10 genes with BETWEENNESS scoring method.

| Rank | Name | Score |
|------|------|-------|
| 1    | TP53 | 1265.752 |
| 2    | EGFR | 537.4 |
| 3    | MYC  | 529.5 |
| 4    | IL6  | 492.8 |
| 5    | AKT1 | 430   |
| 6    | ERBB2| 380   |
| 7    | TNF  | 363.3 |
| 8    | VEGFA| 295.8 |
| 9    | CDKN2A| 250.8 |
| 10   | CDH1 | 243.7 |

Table 4. Docking score of top 5 Punicagranatumgranatum phytoconstituents along with nutlin as the standard against MDM2 protein (PDB ID: 1RV1).

| S.no. | Name of the ligands | Docking Score |
|-------|---------------------|---------------|
| 1     | Casuarin            | −13.4         |
| 2     | Casuarinin          | −13.156       |
| 3     | 2-O-Galloylpunicalin| −12.574       |
| 4     | Corilagin           | −12.116       |
| 5     | Amurensin           | −11.62        |
| 6     | Nutlin (standard)   | −9.253        |
MCODE is a task specific plugin that identifies the cluster subnetwork with the highest connectivity from a complex network (Su et al., 2014; Udhaya Kumar et al., 2020). Similarly, BINGO (Biological Networks Gene Ontology tool) is a Cytoscape plugin that helps in the assessment of a biological network subgraph for Gene Ontology (GO) prediction (Maere et al., 2005). GeneMANIA is a plugin that identifies the most related genes to a query gene set using a guilt-by-association approach. This plugin uses a large database of 800 functional interaction networks from multiple organisms for predictions and each of the related genes is traceable to the source network used for the prediction.

CytoHubba is a plugin that makes biological network analysis easier by utilizing an extensive list of network features. CytoHubba includes several popular algorithms along with other novel network analysis algorithms. Cytohubba also optimizes the computations to drastically reduce the computation time required to perform all the eleven available analytical computations on a common desktop in a reasonable time. It also incorporates enhanced node retrieval to effectively extract subnetworks of interest from a complex biological network to aid researchers in searching and exploring biological subnetworks of interest/subdomains (Chin et al., 2014).

Considering the advantages of the CytoHubba over the other plugins, in the present study the CytoHubba module of Cytoscape is used to find the hub object of a network that is to find the most important nodes in a large network, was used to find out the top 10 most crucial genes of the network. CytoHubba has 11 scoring methods which are executed on the PPI network to find the hub nodes, these scoring methods fall in either of the two categories namely local method (which considers the relation between the node and its direct neighbor) or global method (it considers the relation between the node and whole network). Here we have considered two local (MNC and Degree) and two global methods (closeness and betweenness) to find out the hub...
nodes, in all four algorithms, TP53 protein tops the list as shown in tables (Tables 1–3) and figures (Figures 2–4).

The tumor suppressor gene TP53 encodes the protein p53, which is a very important protein inside our cellular system. It performs various important functions such as promoting cell cycle arrest in response to cellular stress; controlling the process of cell death (apoptosis) by acting as a transcription factor for the bax gene; checks DNA damages, in consequence to that it either leads to cell cycle arrest or cell death; angiogenesis and cell differentiation, etc. to prevent the cancer development hence known as Guardian of Genome (Nag et al., 2013; Suzuki & Matsubara, 2011), it has reported that p53 is inactivated in most of the cancer cases, indicating the importance of p53, being a tumor suppressor gene mutation in TP53 gene enhances the chance of occurrence of cancer. P53 being a transcription factor is activated in response to a various stimulus, its concentration is regulated by another protein known as murine double minute 2 (MDM2), a ubiquitin ligase, they work in a negative feedback manner where the increased concentration of p53 induces the activation of MDM2 which in turn leads to the destruction of p53 by ubiquitin ligase action (Nag et al., 2013).

The importance of p53 have made it a prime target of search in the development of cancer therapy, the researchers tried to do so in different ways, such as by reactivating the mutant p53 by inhibiting MDM2 which is a ubiquitin ligase that regulates the p53 concentration by degrading it, in cancer therapy, it is targeted to increase the concentration of p53 by inhibiting MDM2, nutlin and MI-219 are the inhibitors MDM2, p53 based gene therapy and immunotherapy are the two other mechanisms (Suzuki & Matsubara, 2011).

In our work, we have tried to find the inhibitor of MDM2 from the phytoconstituents of Punica granatum through in silico ways in the search of potential and cost-effective alternative molecule in the field of TP53 based cancer therapy. The molecular docking has been performed to identify the potential inhibitor from the phytoconstituents of Punica granatum later on the stability of the top docking score molecule with MDM2 was checked through molecular dynamics simulation.

Since the last few years, various types of docking tools were developed such as DOCK (Kuntz et al., 1982), AutoDock (Österberg et al., 2002), AutoDock Vina (Trott & Olson, 2009), FlexX (Rarey et al., 1996), Glide (Ra, 2004), and GOLD (Jones et al., 1997). Even though all these tools aid in receptor-ligand docking, many of these tools differ in their ligand placement approach such as FlexX using incremental construction approaches, DOCK using shape-based algorithms, GOLD incorporating genetic algorithms, and Glide utilizing systematic search techniques. As stated, unlike other docking programs, Glide utilizes a systematic search technique to perform a complete systematic conformational, orientational, and positional space search of the docked ligand using molecular force field algorithms before further refining the results using Monte Carlo sampling.

Owing to the advantages of the Glide, molecular docking is performed using the Glide module which is a part of the Schrödinger suite. The crystal structure of the MDM2 protein (PDB ID: 1RV1) was downloaded from the PDB database, the protein was prepared in the Protein Preparation wizard of maestro Schrodinger, the ligands, i.e. phytoconstituents of Punica granatum were searched and downloaded from Pubchem Database then Prepared in the ligprep module of maestro. The prepared protein and ligands were then used for molecular docking.

The phytoconstituents of Punica granatum have shown good results against MDM2 as evident from the docking
scores (Table 4), their docking scores are much better than the standard which was Nutlin here, the ligand interaction diagram of the top two ligands along with the standard nutlin is provided in Figure 5, Casuariin and Casuariinin are making four hydrogen bonds along with other interaction while standard, i.e. nutlin is having only one hydrogen bonding with MDM2 protein.

Further, the stability of these top 2 ligands against MDM2 was determined through Molecular Dynamics Simulation (Figures 6 and 7). While there are few well known molecular dynamics simulation packages, such as AMBER (Case et al., 2005), CHARMM (Brooks et al., 1983), Desmond (Bowers, 2006), GROMACS (Pronk et al., 2013), LAMMPS (Plimpton, 1995), and NAMD (Phillips et al., 2005), LAMMPS and GROMACS are free and open source software while the other tools are licensed under either academic or commercial licenses. LAMMPS is an open source tool that is majorly used in modelling soft, solid-state materials and coarse-grain systems while GROMACS is one of the most widely used free and open-source molecular dynamics tools used to simulate biomolecules. In this study, GROMACS is used for molecular dynamics simulations of the protein-ligand complex.

The graphs obtained after MDS shows that RMSD and RMSF graphs of casuariinin are more stable than Casuariin. Casuariinin shows a good docking score and stable interaction as validated through MDS analysis which makes it a better option to be used against MDM2 for developing cancer treatment.

Conclusion

The genetic correlation among the top 5 commonly occurring cancer was established here using the computational approach, using the common genes the PPI network was generated and analyzed on Cytoscape, it was found that p53 is the hub gene which is known to act as a transcription factor known to regulate various important cellular events. The interaction of p53 and MDM2 has been targeted for developing p53 based cancer treatment. In search of that the Punica granatum phytoconstituents were docked against them, the phytoconstituents show promising results, the stability of the best-docked complexes were further analyzed by MDS, it was observed that casuariinin shows good docking score and even stable interaction with the protein that creates a chance to be used as a potential anticancer agent and could be tested against the cell lines of above-mentioned cancers.

Acknowledgements

The research work was carried out in the laboratory of the Chemistry, Department of Applied Sciences, Indian Institute of Information Technology Allahabad, Prayagraj-India. The Molecular Dynamic simulation results reported in this work were performed on the Central Computing Facility of IIITA, Prayagraj.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Cancer Statistics. (n.d.). India Against Cancer. Retrieved August 18, 2021 from http://cancerindia.org.in/cancer-statistics/.

Akram Husain, R. S., & Ramakrishnan, V. (2016). A review of risk factors in the development of cervical malignancy. Cancer Science & Research: Open Access, 3(1), 1–1. https://doi.org/10.15226/csroa.2016.00123

Assenov, Y., Ramirez, F., Schelhorn, S.-E., Lengauer, T., & Albrecht, M. (2008). Computing topological parameters of biological networks. Bioinformatics (Oxford, England), 24(2), 282–284. https://doi.org/10.1093/BIOINFORMATICS/BTM554

Ataollahi, M., Sharifi, J., Paknahad, M., & Paknahad, A. (2015). Breast cancer and associated factors: A review. Journal of Medicine and Life, 8(4), 6. Retrieved August 18, 2021. pmc/articles/PMC5319297/.

Balasubramaniam, S. D., Balakrishnan, V., Oon, C. E., & Kaur, G. (2019). Key molecular events in cervical cancer development. Medicine, 55(7), 384. https://doi.org/10.3390/MEDICINA55070384

Blackadar, C. B. (2016). Historical review of the causes of cancer. World Journal of Clinical Oncology, 7(1), 54–86. https://doi.org/10.5306/WJCO.V7I1.54

Bowers, K. J. (2006). Scalable algorithms for molecular dynamics simulations on commodity clusters [Paper presentation]. Proceedings of the 2006 ACM/IEEE Conference on Supercomputing, SC’06, https://doi.org/10.1145/1188455.1188454

Brooks, B. R., Brucoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., & Karplus, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. Journal of Computational Chemistry, 4(2), 187–217. https://doi.org/10.1002/jcc.540040211

Case, D. A., Cheatham, T. E., Darden, T., Gohlke, H., Luo, R., Merz, K. M., Onufriev, A., Simmerling, C., Wang, B., & Woods, R. J. (2005). The Amber biomolecular simulation programs. Journal of Computational Chemistry, 26(16), 1668–1683. https://doi.org/10.1002/JCC.20290

Cervino, G., Fiorillo, L., Herford, A. S., Romeo, U., Bianchi, A., Crimi, S., D’Amico, C., De Stefano, R., Troiano, G., Santoro, R., Laino, L., Laino, G., & Cicciu, M. (2019). Molecular biomarkers related to oral carcinoma: Clinical trial outcome evaluation in a literature review. Disease Markers, 2019, 8040361. https://doi.org/10.1155/2019/8040361

Chin, C.-H., Chen, S.-H., Wu, H.-H., Ho, C.-W., Ko, M.-T., & Lin, C.-Y. (2014). CytoHubba: Identifying hub objects and sub-networks from complex interactome. BMC Systems Biology, 8 (4), S11-7. https://doi.org/10.1186/1752-0509-8-S4-S11

Harbeck, N. (2019). Breast cancer. Nature Reviews Disease Primers, 5(1), 1–31. https://doi.org/10.1038/s41572-019-0111-2

Houssein, N. (2019). Factors affecting lung cancer survival: A literature review. Asian Oncology Research Journal, 2(1), 1–9. https://www.journalaorj.com/index.php/AORJ/article/view/30094.

Huang, J., & Mackerell, A. D. (2013). CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. Journal of Computational Chemistry, 34(25), 2135–2145. https://doi.org/10.1002/JCC.23354

Jaggi, P. (2017). A review article on lung cancer diagnosis & treatment. Research & Reviews: Journal of Medical and Health Sciences, 6(2), 1–9.

Jones, G., Willett, P., Glen, R. C., Leach, A. R., & Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. Journal of Molecular Biology, 267(3), 727–748. https://doi.org/10.1006/JMBI.1996.0897

Kumar, R., Nanavati, T., Modi, G., & Dobariya, C. (2016). Oral cancer: Etiology and risk factors: A review. Journal of Cancer Research and Therapeutics, 12(2), 458–463. https://doi.org/10.4103/0973-1482.186696

Kuntz, I. D., Blaney, J. M., Oatley, S. J., Langridge, R., & Ferrin, T. E. (1982). A geometric approach to macromolecule-ligand interactions. Journal
