1. **A.Dey**, S. Sen, U. Maulik. 2022. Study of transcription factor druggabilty for prostate cancer using structure information, gene regulatory networks and protein moonlighting. Briefings in Bioinformatics, 23(1): 1-13, PMID: **34849560**, DOI: 10.1093/bib/bbab465 .

    - The study has provided certain possibilities on TF-based therapeutics. The controlled dynamic nature of the TF may have enhanced the chances where TFs can be considered as one of the prime drug targets. Finally, the combination of membranless phase separation and protein moonlighting has provided possible druggable period within the biological clock. The proposed framework is applied to the identified Transcription factor as biomarkers of prostate cancer and their possible therapeutic solution.

2. S. Sen, **A. Dey**, S. Bandhyopadhyay, V. N. Uversky, U. Maulik. 2021. Understanding structural malleability of the SARS-CoV-2 proteins and relation to the comorbidities. Briefings in Bioinformatics, 22(6): 1-15, PMID: 34143202, DOI: 10.1093/bib/bbab232.
    - Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), a causative agent of the coronavirus disease (COVID-19), is a part of the β-Coronaviridae family. The virus contains five major protein classes viz., four structural proteins [nucleocapsid (N), membrane (M), envelop (E) and spike glycoprotein (S)] and replicase polyproteins (R), synthesized as two polyproteins (ORF1a and ORF1ab). Due to the severity of the pandemic, most of the SARS-CoV-2-related research is focused on finding therapeutic solutions. However, studies on the sequences and structure space throughout the evolutionary time frame of viral proteins are limited. Besides, the structural malleability of viral proteins can be directly or indirectly associated with the dysfunctionality of the host cell proteins. This dysfunctionality may lead to comorbidities during the infection and may continue at the post-infection stage. In this regard, we conduct the evolutionary sequence-structure analysis of the viral proteins to evaluate their malleability. Subsequently, intrinsic disorder propensities of these viral proteins have been studied to confirm that the short intrinsically disordered regions play an important role in enhancing the likelihood of the host proteins interacting with the viral proteins.

3. **A.Dey**, S. Sen, U. Maulik. 2021. Unveiling COVID-19-associated organ-specific cell types and cell-specific pathway cascade. Briefings in Bioinformatics, 22(2): 914-923, PMID: 32968798, PMCID: PMC7543283, DOI: 10.1093/bib/bbaa214.

    - The novel coronavirus or COVID-19 has first been found in Wuhan, China, and became pandemic. Angiotensin-converting enzyme 2 (ACE2) plays a key role in the host cells as a receptor of Spike-I Glycoprotein of COVID-19 which causes final infection. ACE2 is highly expressed in the bladder, ileum, kidney and liver, compared with ACE2 expression in the lung-specific pulmonary alveolar type II cells. In this

study, the single-cell RNAseq data of the five tissues from different humans are curated and cell types with high expressions of ACE2 are identified. Subsequently, protein–protein interaction networks have been established. From the network, potential biomarkers which can form functional hubs, are selected based on k-means network clustering. It is observed that angiotensin PPAR family proteins show important roles in functional hubs. To understand the functions of the potential markers, corresponding pathways have been researched thoroughly through the pathway semantic networks. Subsequently, the pathways have been ranked according to their influence and dependency in the network using PageRank algorithm. The outcomes show some important facts in terms of infection. Firstly, the renin-angiotensin system and PPAR signaling pathway can play a vital role for enhancing the infection after its intrusion through ACE2. Next, pathway networks consist of a few basic metabolic and influential pathways, e.g. insulin resistance. This information corroborates the fact that diabetic patients are more vulnerable to COVID-19 infection. Interestingly, the key regulators of the pathways are angiontensin and PPAR family proteins. Hence, angiotensin and PPAR family proteins can be considered as possible therapeutic targets.

4. **A.Dey**, S. Sen, Vladimir N Uversky, U. Maulik. 2021. Structural facets of POU2F1 in light of the functional annotations and sequence-structure patterns. Journal of Biomolecular Structure and Dynamics, 39(3): 1093-1105, PMID: 32081083, DOI: 10.1080/07391102.2020.1733092.

    - POU domain class 2 homebox 1 or POU2F1 is broadly known as an important transcription factor. Due to its association with different types of malignancies, POU2F1 became one of the key factors in pancancer analysis. However, in spite of considering this protein as a potential drug target, none of the drug targeting POU2F1 has been designed as of yet due to the extreme structural flexibility of this protein. In this article, we have proposed a three-level comprehensive framework for understanding the structural conservation and co-variation of POU2F1. First, a gene regulatory network based on the normal and pathological functions of POU2F1 has been created for better understanding the strong association between POU2F1 deregulation and cancers. After that, based on the evolutionary sequence space analysis, the comparative sequence dynamics of the protein members of POU domain family has been studied mostly between non-human and human species. Subsequently, the reciprocity effect of the residual co-variation has been identified through direct coupling analysis. Along with that, the structure of POU2F1 has been analyzed depending on quality assessment and normal mode-based structure network. Comparing the sequence and structure space information, the most significant set of residues viz., 3, 9, 13, 17, 20, 21, 28, 35, and 36 have been identified as structural facets for function. This study demonstrates that the structural malleability of POU2F1 serves as one of the prime reasons behind its functional multiplicity in terms of protein moonlighting.

5. S. Sen, **A. Dey**, U. Maulik. 2021. Studying the effect of alpha-synuclein and Parkinson's disease linked mutants on inter pathway connectivities. Scientific Reports, 11(16365): 1-14, PMID: 34381149, PMCID: PMC8358055, DOI: 10.1038/s41598-021-95889-5.

    - Parkinson's disease is a common neurodegenerative disease. The differential expression of alpha-synuclein within Lewy Bodies leads to this disease. Some missense mutations of alpha-synuclein may be resultant in functional aberrations. In this study, our objective is to verify the functional adaptation due to early and late-onset mutation which can trigger or control the rate of alpha-synuclein aggregation. In this regard, we have proposed a computational model to study the difference and similarities among the Wild type of alpha-synuclein and mutants i.e., A30P, A53T, G51D, E46K, and H50Q. Evolutionary sequence space analysis is also performed in this experiment. Subsequently, a comparative study has been performed between structural information and sequence space outcomes. The study shows the structural variability among the selected subtypes. This information assists inter pathway modeling due to mutational aberrations. Based on the structural variability, we have identified the protein–protein interaction partners for each protein that helps to increase the robustness of the inter-pathway connectivity. Finally, a few pathways have been identified from 12 semantic networks based on their association with mitochondrial dysfunction and dopaminergic pathways.

6. J. P. Sarkar, I. Saha, A. Lancucki, N. Ghosh, M. Wlasnowolski, G. Bokota, **A. Dey**, P. Lipinski, D. Plewczynski. 2020. Identification of miRNA Biomarkers for Diverse Cancer Types using Statistical Learning Methods at the Whole Genome Scale. Frontiers in genetics, 11(982): 1-25, PMID: 33281862, PMCID: PMC7691578, DOI: 10.3389/fgene.2020.00982.

    - Genome-wide analysis of miRNA molecules can reveal important information for understanding the biology of cancer. Typically, miRNAs are used as features in statistical learning methods to train learning models to predict cancer. This motivates us to propose a method that integrates clustering and classification techniques for diverse cancer types with survival analysis via regression to identify miRNAs that can potentially play a crucial role in the prediction of different types of tumors. Our method has two parts. The first part is a feature selection procedure, called the stochastic covariance evolutionary strategy with forward selection (SCES-FS), which is developed by integrating stochastic neighbor embedding (SNE), the covariance matrix adaptation evolutionary strategy (CMA-ES), and classifiers, with the primary objective of selecting biomarkers. SNE is used to reorder the features by performing an implicit clustering with highly correlated neighboring features. A subset of features is selected heuristically to perform multi-class classification for diverse cancer types. In the second part of our method, the most important features

identified in the first part are used to perform survival analysis via Cox regression, primarily to examine the effectiveness of the selected features. For this purpose, we have analyzed next generation sequencing data from The Cancer Genome Atlas in form of miRNA expression of 1,707 samples of 10 different cancer types and 333 normal samples. The SCES-FS method is compared with well-known feature selection methods, and it is found to perform better in multi-class classification for the 17 selected miRNAs, achieving an accuracy of 96%. Moreover, the biological significance of the selected miRNAs is demonstrated with the help of network analysis, expression analysis using hierarchical clustering, KEGG pathway analysis, GO enrichment analysis, and protein-protein interaction analysis. Overall, the results indicate that the 17 selected miRNAs are associated with many key cancer regulators, such as MYC, VEGFA, AKT1, CDKN1A, RHOA, and PTEN, through their targets. Therefore, the selected miRNAs can be regarded as putative biomarkers for 10 types of cancer.

7. S. Sen, **A. Dey**, S. Chowdhury, U. Maulik, K. Chattopadhyay. 2019. Understanding the evolutionary trend of intrinsically structural disorders in cancer relevant proteins as probed by Shannon entropy scoring and structure network analysis. BMC bioinformatics, 19(13): 231-242, PMID: 30717651, PMCID: PMC7394331, DOI: 10.1186/s12859-018-2552-0.

    - Malignant diseases have become a threat for health care systems. A panoply of biological processes is involved as the cause of these diseases. To unveil the mechanistic details of these diseased states, we analyzed protein families relevant to these diseases. Our present study pivots around four apparently unrelated cancer types among which two are commonly occurring, viz. Prostate Cancer, Breast Cancer and two relatively less frequent viz. Acute Lymphoblastic Leukemia and Lymphoma. Eight protein families were found to have implications for these cancer types. Our results strikingly reveal that some of the proteins with implications in the cancerous cellular states were showing the structural organization disparate from the signature of the family it constitutes. The sequences were further mapped onto respective structures and compared with the entropic profile. The structures reveal that entropic scores were able to reveal the inherent structural bias of these proteins with quantitative precision, otherwise unseen from other analysis. Subsequently, the betweenness centrality scoring of each residue from the structure network models was resorted to explore the changes in dependencies on residue owing to structural disorder. These observations help to obtain the mechanistic changes resulting from the structural orchestration of protein structures. Finally, the hydropathy indexes were obtained to validate the sequence space observations using Shannon entropy and in-turn establishing the compatibility.

8. A. Dey, U. Maulik. 2022. Bioinformatics pipeline to unveil the heterogeneity of Glioblastoma Multiforme. Oral presentation: 2022 IEEE Calcutta Conference (CALCON), Kolkata, West Bengal, DOI: 10.1109/CALCON56258.2022.10060233.
    - Understanding cellular heterogeneity is a breakthrough in both the field of biology and medicine. Cells harboring from the same genome show functional disparity in various microenvironments. Here, cell-specific regulatory circuits help to reveal the mode of regulation of the biomarkers and the cause of abnormalities regarding the disease progression. Therefore, each signal during the regulation process is important and crucial to determine the heterogeneity of disease. Low resolution cell isolation techniques used previously, averaging the signals of each cell, are not feasible to reconstruct the gene regulatory network. Recently, advancement of single-cell RNA sequencing techniques enabled to capture the transcriptomic aspects of each cell. Though single-cell gene expression studies open new pathways to unveil the biological complexities, they are not sufficient to understand the cellular state under a disease condition. The local biological networks will further escalate the perception of the cellular heterogeneity more clearly. In this study, we established the local networks of the cell types responsible for one of the most aggressive cancers known as glioblastoma multiform. We identified the transcription factors that are responsible for regulating the mode of the cell-specific hub biomarkers and finally lead to disease progression. Moreover, the identified crucial transcription factors from the network are RELA, NFKB-family, STAT3, SP1, FOS and JUN. In the future, these transcription factors can be considered as a successful therapeutic target during designing precision medicine strategies.

9. A. Dey, U. Maulik. 2020. Identification of Cell-types based on the Pathway of Markers using Single-cell data. Oral presentation: 2020 IEEE Calcutta Conference (CALCON), Kolkata, West Bengal, DOI: 10.1109/CALCON49167.2020.9106539.
    - The advancement of single-cell sequencing technology has made it possible to describe high throughput and low-cost genome-wide sequencing. There are several methods that utilize the single-cell sequencing techniques to determine the cell types that construct a complex tissue. Clustering followed by dimensionality reduction are used to determine cell type. Moreover, statistical analysis is performed to identify the uniqueness of each cell type. In this study, traditional hierarchical clustering is performed to identify the cell types present in peripheral blood cells. This classification reveals the morphologically and phonetically different cells present in peripheral blood cells of a healthy donor. Each cell type contains a specific marker that defines the individual character of the cell type. Furthermore, these markers play an important role in biological pathways. The association of markers with pathway helps in understanding the cell-to-cell heterogeneity. During the study, cell markers of each cell type are further considered to analysis the associated pathways. The analysis shows how the change in gene expression contributes to pathway shift due to several biological causes. This information

provides new insight in the understanding of biology process from a single-cell perspective.

10. S. Sen, A. Dey, U. Maulik. 2018. Identifying potential hubs for kidney renal clear cell carcinoma from tf-mirna-gene regulatory networks. Oral presentation: 2018 IEEE Applied Signal Processing Conference (ASPCON), Kolkata, West Bengal, DOI: 10.1109/ASPCON.2018.8748806.
    - Kidney Renal Clear Cell Carcinoma (KIRC) is common kidney cancer and ranks 8 th among all other popular cancer types. Mostly adult humans are affected by this malignant disease. However, enough information has not been gathered for this special case of renal carcinoma. The experimentally validated data are also in a non-uniformed form which makes it more difficult to decode the relation among miRNAs, genes and TFs. To address this obstacle in this current study, miRNAs, genes and TFs those associated with KIRC and played a significant role are considered as individual elements and a framework is proposed to find the connection among each entity. From this proposed framework some genes are identified those are not only targeted by miRNAs or TFs separately but also gene regulation has an impact due to the passive effect of TFs. From this network gene regulation due to hsa-miR-146a-5p is indirectly controlled by those TFs also. For the further validation of these informative miRNAs and genes KEGG pathway and GO enrichment analysis are performed. In future study, this framework should be granted more attention in order to understand better prognosis about malignant diseases.