A Protein Moonlighting Based Model to Reveal Transcription Factor Druggability for Prostate Carcinoma

Sagnik Sen*†, Ashmita Dey† and Ujjwal Maulik

Abstract

Background: Prostate cancer is the second leading cause of cancer-related death in men in the United States. Metastasis shows poor survival even though the recovery rate is high. In spite of numerous studies regarding prostate carcinoma, multiple questions are still unanswered. In this regards, gene regulatory network can uncover the mechanisms behind cancer progression, and metastasis. In the regulatory network, among the three prime molecular members, transcription factors and microRNAs are known for their regulatory activity where transcription factor can target both microRNAs and genes. Under a feed forward loop, transcription factors can be a good druggable candidate. However, due to the dynamic nature of transcription factors, designing an appropriate drug becomes challenging.

Result: We have proposed a computational model to study the uncertainty of transcription factors and suggest the appropriate cellular conditions for drug targeting. We have selected feed-forward loops depending on the shared list of the functional annotations among transcription factors, genes, and miRNAs. From the potential feed forward loop cores, six transcription factors were identified as druggable targets, which include AR, CEBPB, CREB1, ETS1, NFKB1, and RELA. The selected transcription factors have been investigated based on the evolutionary co-variance study, post-translation modifications, and disordered region identification. The structural unrest of the selected transcription factors can be observed from outcomes. Comparing the results of the aforementioned analyses, probable binding clefts have also been identified. transcription factors are known for their Protein Moonlighting properties, which provide unrelated multi-functionalities within the same or different subcellular localizations. Following that, we have identified such functions that are suitable for drug targeting. Finally, we have tried to identify membraneless organelles for providing more specificity to the proposed time and space theory.

Conclusion: 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 consider as one of the prime drug targets. The usual presence of TF at nucleus cannot explain its functional multiplicity whereas nucleus speckles can help to explain it. Finally, the combination of membranless phase separation and protein moonlighting has provided possible druggable period within the biological clock.

Keywords: Transcription Factors; Intrinsically Disorder Proteins; Protein Moonlighting; Liquid Liquid Phase Separation; Drug Target
1 Background

Globally, Prostate Cancer (PCa) is the second most common cancer in men and the first in the United States. According to recent statistics, the number of new prostate cancer cases will be approximately 191,930 in 2020 in the United States [1], and 33,330 deaths from this malignancy are expected. The diagnosis of PCa is done by Prostate-specific antigen (PSA) screening [2]. In men with local or regional disease, it can be successfully treated through surgery or radiation. In the case of metastatic PCa, androgen deprivation therapy (ADT) is used [3]. However, ADT has a low rate of success with aggressive PCa cases. Significant dysregulation of oncogenes or tumour suppressor genes and their interaction networks have vital roles in cancer development, including PCa [4]. In the gene-gene regulatory networks, perturbation of one regulator may lead to the dysregulation of a set of effector genes. Thus, regulators have a major contribution to the modulation of gene expression during tumorigenesis. Transcription Factors (TFs) and microRNAs are the key regulators at the transcriptional and post-transcriptional level, respectively. TFs are a class of proteins that can target and modulate the expression of protein-coding genes and miRNAs. miRNAs are small non-coding RNAs that may regulate the expression of protein-coding genes and TFs. The miRNAs and TFs jointly regulate the same protein-coding genes and can form a feed-forward loop (FFL) regulatory motifs in gene interaction networks. So, TFs and miRNAs in the FFLs can control the pathogenic progression of malignancies by affecting the pathways associated with cell proliferation, tumour suppression, and cell signalling pathways [5]. The TFs AR, STAT3, and ETS1 are among a few of the well-known regulators that can affect critical pathways during prostate cancer development [6, 7, 8].

TFs have been known for their protein moonlighting properties [9]. Protein Moonlighting is a term used for a class of proteins that are associated with completely unrelated functions. Therefore, TFs are naturally extremely uncertain and they can associate with non-transcription regulation activity at the same or diverse subcellular localization. Usually, Many TFs are not suitable as drug targets due to their lack of druggable pockets when folded but this may be a reason why TFs are not selecting as a potential drug target despite their ability to control the pathogenic progression of the malignancies. To address the problems associated with the uncertain nature of TFs, we have proposed a framework. We constructed a gene regulatory network (GRN) based on significantly expressed genes, their regulators, and sharing functional annotations. Next, the potential FFLs from the PCa associated GRN were selected. We then employed evolutionary coupling analysis, a disorder study to understand the venerable zone in the monomeric sequence of the PCa associated TFs. We subsequently mapped this sequence-based information to the structure of the TFs. Furthermore, the TFs aly analyzed based on the associated pathways and specified subcellular localization. To check for moonlighting functions, the TFs are re-analyzed using MoonProt [10] and MoonDB [11]. We identified significant associations between the TFs in the FFLs and drug response, which added additional biological relevance of the studied TFs in the context of PCa pathology.

2 Methodology

The experiment has few segregated stages. Firstly, the study begins with the identification of potential TF candidates. Subsequently, the structural study and systems
biology approaches have thoroughly been performed for the selected TFs. The experimental design of the work has been elaborately described below:

2.1 Identify responsible genes
We extracted 294 prostate tissue-specific genes from TissGDB database [12]. The database contains 2461 tissue-specific genes, curated for 22 tissue types that matched with 22 diverse cancer types. Among the prostate tissue-specific genes, 113 differentially expressed (DE) genes measured by TissGDB are considered to establish the networks (reported in Supplementary Table 1).

2.2 Establish the relation between gene, miRNA and TF
The miRTarBase database [13] includes experimentally verified 502 653 miRNA-target interactions of human. For the 113 genes, DE in the prostate cancer, we identified 1408 experimentally verified miRNA-target pairs from the miRTarBase database. We constructed a network based on the identified interaction pairs. The 113 DE-genes are further considered to identify the Transcription factors (TFs) from the experimentally verified TF-gene pairs available in the TRRUST database [14]. Moreover, the miRNAs from the established miRNA-gene interactions are pondered for constructing TF-miRNA pairs, where TFs regulating the miRNAs from TransmiR database [15]. We merged these three databases and determined TF-mediated gene regulation directly or indirectly through miRNAs. Modularity score [16] is calculated by using equation 1. This scoring system prioritized 6 TFs in the final merged network.

\[ M = \sum_{n=1}^{K} (e_{nn} - a_{n}^2) \]  

(1)

Where as \( e_{nn} \) represents the probability edge is in module \( n \) and \( a_{n} \) is the probability of a random edge that would fall into module \( n \).

To strengthen the established GRN, the pathway of each selected TFs, miRNAs and genes are considered from KEGG pathway. It is evidenced that they share a number of pathogenic and non-pathogenic pathways. The information of pathways is utilized to construct the final relational network among the three molecular regulators. The final combinations among TF-miRNA-gene are reported in Supplementary Table 4.

2.3 Direct Coupling Analysis
The fold of the native structure of a protein is maintained during the observation of slow changes that occurs to a protein sequence throughout the evolutionary time-frame [17]. Those amino acids remain unaltered are considered as conserved. The conserved residues are responsible for sustaining the structure and function of a specific protein. A small change or mutation in terms of size, shape or chemical composition of amino acids may lead to structure function disparity. Whereas, a compensatory change is occurred to a corresponding residue of the three-dimensional structure in order to balance the mutation of a particular position. This implies that in evolutionary time-frame the co-variation of amino acids is important to preserve or restore the structure-function paradigms. The residual dependencies
and co-variation effect are analyzed by using a statistical method known as direct coupling analysis or DCA [18]. In this model, Direct information (DI) score is calculated between two amino acids to measure the strength of direct relation. Equation 2. depicts how two sites of MSA is directly coupled.

\[
DI_{ij} = \sum AB \times p^{(dir)}_{ij} \times (A, B) \times \ln \frac{p^{(dir)}_{ij}(A, B)}{P_i(A)P_j(B)}
\]  

(2)

In the above equation, \(p^{(dir)}_{ij}\) represented as reweighted frequency of two residues. \(P_i(A)\) and \(P_j(B)\) are the individual probability at \(i\)th and \(j\)th residual position of a sequence. Whereas, the joint probability is defined as \(P_{ij}(A, B)\). Depending on the top DI scores, some co-varying evolutionary patches are selected for further study. Furthermore, to exhibit an effect, weighted network GDCA is constructed based on DI score. GDCA = \((V_{res}, EDI)\), here \(V_{res}\) is the set of nodes and EDI is the weighted edges, consists of DI score between the residues. The color modules have been established by using Girvan-Newman algorithm [19].

2.4 Disorder region of the common TFs
To understand the disorderness of these TFs, PONDR-vlxt [20] and IUPred-s databases [21] are used. These two web-based servers help to predict the disorder of the amino acids present in intrinsic proteins. However, PONDR is a meta-predictor, uses a consensus artificial neural network (ANN) prediction method, which was developed by combining the outputs of several individual disorder predictors and predict intrinsic disorder from an amino acid sequence. Similarly, IUPred presents a novel algorithm for predicting such regions from amino acid sequences by estimating their total pairwise inter-residue interaction energy, based on the assumption that IUP sequences do not fold due to their inability to form sufficient stabilizing inter-residue interactions. Optional to the prediction are built-in parameter sets optimized for predicting short or long disordered regions and structured domains. From these two servers, the score of amino acids are identified and the overlapped regions of the sequence are observed to understand the change of score due to disorderness.

2.5 Post Translational Modification
The act of a particular protein in eukaryotic cells is regulated by Post Translational Modification (PTM). This is a collective chemical modification of proteins translated from mRNA before becoming functional in multiple body cells. These modifications are responsible for the heterogeneity of proteins and also help in deploying similar proteins for different functions in different cells. Using PhosphoSitePlus database, the types of PTM are identified of the selected protein structures.

2.6 Calculate the stability score
The sequence is mapped to the structure of the respective TFs and compare the influential changes in sequence space that affect the structural behaviour of the TFs. In this regard, alascan mutagenesis technique is applied to the PDB structure using FoldX [22]. Alascan scanning method reveals the contribution of a specific residue to the stability or function of a given protein. Alanine is used due to its non-bulky,
chemically inert, methyl functional group that nevertheless mimics the secondary structure preferences that many of the other amino acids possess. Moreover, during this analysis, the quality assessment such as different dihedral bonds, torsion and van der Waal’s forces occurred to particular residues are also identified.

*Root Mean Square Fluctuation for Structural Ensemble*
Each of the TF has multiple solved multimeric structure. The cumulative residual square fluctuation can help to understand overall fluctuation trait of the structures. Root Mean Square Fluctuation (RMSF) is measure of the particle deviation. In RMSF, a mean over time is considered for a residue j at the current position and some reference position. The definition of the RMSF is given in Equation 4.

\[
RSMF_k = \left( \frac{1}{T} \sum_{m_n=1}^{T} \mod(r_i(t_j) - r_i^{re})^2 \right)^{0.5}
\]  

Where \( T \) is time over which the mean has been taken for reference position of the particle \( k \), \( r_i \). The RMSF has been observed based on the reference position of the particle \( j \) over time.

### 2.7 Liquid Liquid Phase Separation
In this article, we are focusing on the multi-functionality of the TFs to establish the perfect time and space for drug targeting. Usually, Cellular organelles are mostly known for their membrane based segregated entities. TFs are mostly found in the nucleus (sometimes in mitochondria). Therefore, we have tried to observe the involvement of the membraneless organelles within the nucleus. Liquid Liquid Phase Separation (LLPS) has been observed to understand the aforementioned issue. LLPS is a method where the condensates help to separate the membraneless compartments from the homogeneous liquid. Here, we have used two databases i.e., LLPSDb \[23\] and PhasepDb \[24\] for LLPS study.

### 3 Results
In this section, the experimental outcomes are sequentially discussed. The outcomes are sub divided in three subparts i.e., PCa specific GRN; Sequential and structural study of the TFs and Functional study on TFs. The detail description is given below:

#### 3.1 TF-miRNA-gene regulatory network in Prostate Cancer
The objective of the work is to understand the influence of the TF during the regulation of gene expression directly or indirectly (through miRNA). In this regard, a two-fold analysis is performed. First, we established the relation among the genes, miRNAs, and TFs in the context of PCa. Second, an in-depth analysis of the sequence and structure of the selected TFs that have significant associations with pathology of PCa.

First, we built miRNA-gene and TF-gene networks for the differentially expressed genes responsible for prostate carcinoma. Subsequently, we identified TF-mediated miRNA regulation from the aforementioned TFs and miRNAs. To gain confidence,
we have only considered experimentally verified regulator-target associations. Finally, a network is constructed among the genes, miRNAs and TFs. This network heeds only those TFs responsible for both direct and indirect regulation of the FFL. As a result, 183 combinations among 8 genes, 39 miRNAs and 28 TFs are found from the final network (Supplementary Figure 1). In the network, genes, miRNAs and TFs are represented through a blue rectangle, yellow triangle and red oval respectively.

From 28 TFs we prioritized 6 highly responsible and influential TFs in prostate carcinoma based on their sharing of functional pathways (pathFA). We determined four pathFA those are shared by these 6 TFs. These pathFAs played important roles in cancer including prostate cancer. Furthermore, to understand the path through which the TFs are regulating the miRNAs and genes, the selected four pathFA (CAMP Signaling Pathway, Oxytocin Signaling Pathway, Pathway in Cancer and Prostate Cancer) are highlighted extremely. The association of the TFs with the pathFA are shown in FIG. 1 by using a circos plot, but the plot shows only three pathFA as the pathway in cancer is the mother of all the pathways involved in cancer formation.

In the circos plot, three pathways are shown, those are regulated by most of the 6 selected TFs. Laterly, these pathways are consider to understand through which path the TFs regulate the miRNA and genes present in the FFL. During the selection of the FFL have an impact on the three pathways 57 comibations are found. Those combinations contain 8 genes, 38 miRNAs and 6 TFs. Moreover, it is evidenced that these 6 TFs have a key role in the prostate cancer. In FIG.2, the final gene regulatory network is established. As the TFs are selected based on the functional perspective, therefore, it is expected that the final GRN is responsible for regulating the GRN core as well as a functional modification during prostate cancer. Interestingly, the TF AR has also been observed as differentially expressed at the RNA level in prostate cancer which is targeted by two TFs- ETS1 and CREB1 in two distinct FFL.

3.2 Discussion on Evolutionary Sequence Space, Post Translational Modifications, and Sequential Stability

From the final GRN, selected 6 TFs are analyzed to understand their regions involved in controlling the structural core throughout the natural adaptation. Evolutionary sequence space could detect more stringent domain following the natural adaptation trait of respective families. It is more global and robust than protein-specific information. In this regard, we have designed five DCA specific networks for six TFs. NFkB1 and RELA are sharing the same protein family. In FIG. 3, five networks are depending on the coupling propensity. As mentioned earlier, each module of the network can showcase the evolutionarily conserved intramolecular dependency. Therefore, residual co-variance can contemplate as keys for intra-molecular allostery. This shows how the selected residues can directly or partially control the structural stability while targeted by drug molecules. Besides, the co-variance can also represent the non-covalent residual interaction. So, the sequential frame can provide a set of precise influential residues. Applying the proposed sequential frame, we have identified five probable structural facets for AR, ETS1, CEBPB,
CREB1 and RELA. In terms of the prediction algorithm, AR shows the highest rate of disorder at specific residual positions.

Additionally, the miRNAs present in the FFL of the final GRN are considered to analyze their evolutionary trait during natural adaptation. Our main aim is to observe the influential TFs in prostate cancer. In this regard, miRNAs those are associated targeted by at least one TF are determined. To observe the evolutionary changes, miRNAs and their targeted genes are studied in white mice. As a result, 10 miRNAs are found for six TFs. These miRNAs also significantly regulated the genes in prostate carcinoma. Among the 101 miRNAs, 9 miRNAs are validated through a literature survey. Our results show, PTGS2 gene is targeted by all the influential TFs and also targeted by different miRNAs (participated in natural adaptation). This evidence supports the strong impact of the selected TFs in the modulation of the GRN during the disease.

PTMs occurring at the amino acids play a crucial role in protein’s function, physicochemical properties, conformation, stability and molecular interactions in response to developmental signals or environmental stimuli. In this order, in FIG. 4 influential subtypes of PTMs for the 6 TFs are shown extensively. The subtypes are represented by using different colours. Red, green yellow and black defines acetylation, ubiquitylation, phosphorylation and other respectively. Rate of phosphorylation is high in almost every TF except RELA. In AR, three more PTM subtypes viz., acetylation [25], ubiquitination and methylation [26] are well distributed throughout the sequence. Similarly, ETS1 consists of equal number ubiquitination [17] and acetylation sites. Interestingly, CEBPB has dimethylation sites [26]. Unlike other TFs, RELA has O-GlcNAc sites [27, 28] along with small amount acetylation and phosphorylation sites. Almost every sub-type has implications on prostate cancer where O-GlcNAc sites are participating in cell proliferation.

After PTM site identification, in sequence space, we have initially studied the flanking regions applying the disordered region prediction algorithms shown in FIG. 5. The prediction has provided protein-specific unstructured as well as expected changing cores. In this regard, intersecting disordered domains from two databases such as IUPred and PONDR are selected as disordered regions. The overlapping regions depict the fluctuations of the amino acids present at that particular region and due to the Ramachandran angles, they are highly responsible for disorderedness of the TFs. Moreover, it is known that the disorder region occurs at the PTM sites mostly and the scores of amino acids support the result of PTM also.

Besides PTM and disorder region identification, Alanine mutagenesis is used to interpret the contribution of a single residue towards the stability of a protein. Due to non-bulky, chemically inert, methyl functional group that nevertheless mimics the secondary structure preferences alanine is used to check the stability. In FIG. 6, graphs depict the stability score after interchanging the residues with alanine. Similarly, quality assessment of selected proteins is performed and the van der Waals clashes occurred to particular residues shown with colours and the details is reported in the Supplementary Table 2.

Furthermore, TRRUST database is used to unveil the association of diseases including prostate carcinoma with the six TFs we identified here. These TFs and their linked diseases are summarized in Supplementary Table 3. Top five diseases of each
TF are sorted depending on their p-values and a network based on that is shown in FIG. 7. The network diagram depicts the connection between TFs and diseases. Among these two diseases such as malignant neoplasm of breast and prostate are targeted by five TFs excluding CEBPB with high p-values.

3.3 Functional Annotations and Relation with Other Diseases

These TFs are found responsible for the prostate cancer, but to understand their reaction on the pathways, the protein interaction partners are identified. In this regard, STRING database (https://string-db.org/) is used and subsequently, the localization of the pathways, those are controlled by the TFs along with their partners are identified from Reactome pathway database. The Reactome Pathway (pathreactome) shows that TFs present in a single cellular localization involves two distinct processes. The TFs, protein interaction partners with their pathreactome for a particular localization are reported in TABLE 1. From this result, it is expected that these TFs may have a moonlighting property as they perform distinct functions in a same localization. Due to this property proteins can execute more than one function. Proteins gather this property through evolution. To validate the moonlighting property of the TFs both MoonProt and MoonDB databases are utilized. We found that AR and ETS1 are present in the database as a moonlighting protein, whereas for other four TFs Basic Local Alignment Search Tool (BLAST) is used to identify the highest matching with the queried sequence. We found that the highest matching sequence shows moonlighting property and the functions are performed by the queried sequence too. Though the searched proteins are not present in the two databases these also show moonlighting properties. This property of TFs along with the curated databases and processes are reported in TABLE 2.

We have tried to provide suggestive drugs for the selected TFs. In this regard, the respective drugs have been chosen based on the differential activity of the TFs. From the literature, AR is highly associated with Prostate Cancer where the under expression of the protein is responsible for the disease. Similarly, the two proteins from NFkB family are overexpressed during PCa [29]. NFkB family proteins are regulating the AR gene expression. Also, CEBPB is sharing the same level of expression with NFkB family proteins [30]. CEBPB is modulating the metastatic genes in Prostate Cell. In TABLE 4, we have prepared one such list where TFs and their corresponding expression and drugs are mentioned. In the case of AR, only top five agonist drugs have been considered. Only for ETS1, we cannot detect the appropriate drugs. All the information has been taken from DrugBank [31].

4 Discussion

The proposed work has three folds. Firstly, we have tried to identify most potential FFLs which are strongly responsible for the pathogenic progression of PCa, have been identified. Based on the sharing functional annotations among the molecular members from Figure 1., the truncated network has been formed (Shown in Figure 2.). The list of common sharing functional annotation has been given in the circos plot (Shown in Figure 3.). Interestingly, all of the KM Plots show a good affinity for PCa. This process helps to identify the potential TFs which can be targeted by suitable drugs. However, miRNAs targeting pattern can be random for the selected
species. To established the conservation of the study and to established the importance of the miRNAs in terms of PCa, an evolutionary study has been performed. Just to showcase that the relation between miRNAs and respective mRNAs are not random, we have checked with white mice as well. In the study, we have identified at least one FFL under from each of the regulator TFs. In Supplementary Table 4, the identified FFLs are marked in bold letters. All these studies help to establish the TFs and their corresponding FFLs which are immensely influential in PCa pathogenic progression. However, the prime objective of identifying the appropriate druggable session of the selected TFs is needed to be unveiled. In this regard, primarily we have studied the uncertainty of the TFs based on the evolutionary sequence-based study and also the physiological properties of each TFs in terms of PTM and disordered based information. In Figure 4., the regions of PTM have been shown for each of the TFs. Certainly, these regions have a higher propensity of not participating in folding. Though the disordered regions for the TFs have been observed based on two well-known prediction algorithms viz., PonDr-VLXT and IUPRED-s. Comparing the results from Figure 5. and Figure 6., the regions of interest in terms of instability have been selected. These analyses have completely relied on the physical and structural properties of those six TFs. The evolutionary sequence space conservation can reveal the residual dependencies in each protein sequence. Therefore, DCA is appropriate to detect the residues with most co-variance dependencies. More elaborately, DCA is an entropy-based tool where the entropy of coupling between two residues at a position has been identified. For the purpose, the residue-residue coupling has been taken in a disentangled manner. Subsequently, the scores have been curated following two specified thresholding. First, the top residue couples are taken based on a symmetric downfall of the score. Next, residue-residue coupling within 4 Armstrong residuals are opted out from the list. These curations have provided one final coupling list with peptide bond biases. Now, the information from Tables is extremely singular and monotonous. Therefore, weighted networks based on the tables have been formed, shown in Figure 3. These networks have been shown to depict the bigger picture of the coupling co-variations. Betweenness centrality based clustering is one such method where we have identified cores of unstable residual dependencies. The physical and structural properties shown in Figure 5. and Figure 6. have been compared with the network-based outcomes shown in Figure 3. These provide an unstable set of residues which can influence universally the whole structure of the TFs. Also, these bigger lists of unstable residues can clearly represent the rate of uncertainty of the TFs. Alanine mutagenesis is one of the techniques where the stability of each residue has been calculated in terms of their structural orchestration. Figure 7. shows the stability index based on alanine mutagenesis. Also, certain residues are mentioned where non-covalent clashes are observed for each of the protein structure. This information can assist to predict suitable binding sites. Comparing the aforementioned studies, the probable binding clefts for the selected TFs have been identified. NFkB1 and AR have very defined binding cleft Viz., golden and brown color modules respectively. For this TFs, we can identify non-covalent clashes within the color modules. For ETS1, color module yellow has the binding cleft. Similarly, CEBPB has binding clefts at cyan, grey and sea green module. For RELA and CREB1, the whole sequence space networks can
be considered as binding clefts. These binding zones are identified based on the average structural disorders at residues. To confirm the structural aspects, ensemble structural fluctuation of the solved structure have been observed (given at Supplementary Figure 2). The result of disorder ans fluctuation are similar. Therefore, it can be expected that the binding clefts can be utilized for drug targeting. However, the previous question still remained unanswered. From MoonDB and MoonPROT, we have identified the potential dual functionalities for five out of six selected TFs (except CEBPB) (Given in TABLE 1). In TABLE 2, each of the functions associated with the selected TFs has been listed mostly at two subcellular localizations i.e., Cytosol, and Nucleus. Comparing TABLE 1 and TABLE 2, we have identified some of the functions where deregulation has lead to PCa.

ARs are responsible for both the survival and function of the epithelial cells in the prostate. The receptors also regulate the growth of the tumour formed during the early stage of prostate cancer. It is found that Rho GDP Dissociation Inhibitor Alpha (GDIα) is a negative regulator of androgen receptor signalling pathway. Moreover, overexpressed GDIα downregulates the expression of AR at both the levels of mRNA and protein. This up and down-regulation of the AR genes are responsible for the abnormal balance between the growth and differentiation functions of the AR pathway during prostate cancer progression. However, different transcription factors are responsible to regulate the AR signalling pathway which directly or indirectly contributes to the progression of cancer. However, other selected TFs are also involved in a certain set of functional triggering which can partially modulate the activity of AR. Interestingly, these functions can be considered druggable selection based on the protein moonlighting properties. In the list, the associative activity of the ETS1 and AR has been shown where the modulation at ETS1 expression can alter the castrate-resistant response to AR antagonist treatment [32]. Oxidative stress signalling pathways play a vital role in the progression of castration-resistant PCa (CRPC). Transcription activity of the CREB1 can modulate the Oxidative stress signalling pathway by targeting the AR [33]. NFkB family transcription factor proteins are largely known for their association with the AR. In the study, we have identified two NFkB family proteins i.e., NFkb1 and RELA. In Zhang et al., implications of NFkB family proteins in regulations of the AR signalling pathway in PCa metastasis are broadly discussed. However, we have not found any evidence from literature where an association between CEBPB and AR has been established. Therefore, CEBPB is considered a novel marker. From TABLE 2, Signal Transduction has mentioned as potential functional subtypes, located at nucleoplasm. In terms of protein moonlighting properties, these can be an appropriate session for drug targeting. The main spatial division has shown prime subcellular localization i.e., cytosol and nucleoplasm which are separated through nuclear membrane. However, further sub-division within the nucleoplasm has been identified through LLPS study. In TABLE 3, the list of such localizations has been reported. Although, nucleus is a widely known subcellular localization for TFs, nucleus speckles, a type of the membraneless organelles, have been noticed as another localization for most of the TFs. Interestingly, the nucleus speckle is one such membraneless organelle where the activities at pre-mRNA phase have been performed. Therefore, the functional existences of TF at nuclear speckles can be avoided. Moreover, when the mRNA is
exiting from the nuclear speckles that phase may be considered as an appropriate session for drug targeting. In this regard, further in vivo validation is required.

After discussing the objectives, we have tried to provide suggestive drugs for the selected TFs. In this regard, the respective drugs have been chosen based on the differential activity of the TFs. More elaborately, each of the TFs has a specific expression level while triggering prostate cancer or associated elements. From the literature, we got the appropriate information about this. There are two categories of medicinal drugs i.e., agonist and antagonist. Therefore, if the disease pathogenicity increases due to the under expression of the molecules then an agonist drug works well on such TFs and vice versa. In TABLE 4, we have prepared one such list where TFs and their corresponding expression and drugs are mentioned. From the literature, it has been observed that the prime TF downregulation of AR is causing pathogenic progression of PCa. In the case of AR, only top five agonist drugs have been considered. Similarly, overexpressed NFkB family TFs (RELA, NFkB1) can modulate the hormonal activity of the AR. This combined process is modulating the tumorigenesis during PCa. Also, CREB1 is identified as overexpressed during PCa. Although the expression level of ETS1 has been verified from the literature, the suitable for the purpose is unknown. CEBPB is marked as a novel finding. All the information have been verified from the DrugBank database.

5 Conclusion
The study has aimed to unveil the dynamic nature of the TF like proteins. As per the outcomes, GRN has provided six potential TFs for PCa which are further studied. Interestingly, the evolutionary sequence level study has provided a clear understanding of the dynamic nature of the TF like proteins. These also provide probable binding cleft of the selected samples. The non-transcription activities have been ruled out based on the protein moonlighting properties and probable presence within nucleus speckles. The frame has provided a sequential set of experiments that have unveiled probable periods for TF drug targets within the biological clock. It can help to design the therapeutic perspective accordingly.

Abbreviations
PCa: Prostate Cancer TF: Transcription Factor microRNA: miRNA FFL: Feed Forward Loop GRN: Gene Regulatory Network PTM: Post Translational Modification PDB: Protein Data Bank RMSF: Root Mean Square Fluctuation LLPS: Liquid Liquid Phase Separation

Declarations
Not applicable.

Ethics approval and consent to participate
We used publicly available data for human and animal samples and cell line studies. No human and animals are directly involved.

Consent for publication
Not applicable.

Availability of data and material
All data generated or analyzed during this study are either taken from publicly available databases or included in this article.

Competing interests
The authors declare that they have no competing interests.
Author’s contributions
AD and SS have conceptualized and performed the experiments. AD, SS and UM have drafted the manuscript. The whole work has been supervised by UM.

Acknowledgements
We would like to thank DST-INSPIRE fellowship scheme for supporting the work of Ashmita Dey and Sagnik Sen.

References
1. Siegel, R.L., Miller, K.D., Jemal, A.: CA: A Cancer Journal for Clinicians. Cancer statistics (2019)
2. Patil, T., Bernard, B.: Complications of Androgen Deprivation Therapy in Men With Prostate Cancer. Oncology 32(9), 470–474 (2018)
3. Moore, D., d.M. Simoes, R., Dehmer, M., Streib, F.E.: Prostate Cancer Gene Regulatory Network Inferred from RNA-Seq Data. Current Genomics 20(1), 38–48 (2019)
4. Shitivelman, E., Beer, T.M., Evans, C.P.: Molecular pathways and targets in prostate cancer. Oncotarget 5(17), 7217–7259 (2014)
5. Baumgart, S.J., Nevedomskaya, E., Haendler, B.: Dysregulated Transcriptional Control in Prostate Cancer. International Journal of Molecular Sciences 20(2883), 1–20 (2019)
6. Chai, E.Z.P., Shanmugam, M.K., Arifso, F., et al: Targeting transcription factor STAT3 for cancer prevention and treatment. Pharmacology & Therapeutics 162, 86–97 (2016)
7. Lambert, M., Jambon, S., Depauw, S., David-Cordonnier, M.H.: Targeting Transcription Factors for Cancer Treatment. Molecules 23(6), 1479 (2018)
8. Jeffery, C.J.: What is Protein Moonlighting and Why is it Important? Molecules (2017)
9. Jassal, B., Matthews, L., Viteri, G.: The reactome pathway knowledgebase. Nucleic Acids Research 48(D1), 498–503 (2018)
10. Chen, C., Zabad, S., Liu, H., Wang, W., Jeffery, C.: MoonProt 2.0: an expansion and update of the moonlighting proteins database. Nucleic Acids Research 46(D1), 640–644 (2018)
11. Ribeiro, D.M., Briere, G., Bely, B., Spinelli, L., Brun, C.: MoonDB 2.0: an updated database of extreme multifunctional and moonlighting proteins. Nucleic Acids Research 47(D1), 398–402 (2019)
12. Kim, P., Park, A., Han, G., Sun, H., Jia, P., Zhao, Z.: TissGDB: tissue-specific gene database in cancer. Nucleic Acids Research 46(D1), 1031–1038 (2018)
13. Chou, C.H., Shrestha, S., Yang, C.D.: miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. Nucleic Acids Research 46(D1), 296–302 (2018)
14. Han, H., Cho, J.W., Lee, S., et al: TRRUST v2: an expanded reference database of human and mouse transcriptional regulatory interactions. Nucleic Acids Research 46(D1), 380–386 (2018)
15. Tong, Z., Cui, Q., Wang, J., Zhou, Y.: TransmiR v2.0: an updated transcription-factor-microRNA regulation database. Nucleic Acids Research 47(D1), 253–258 (2019)
16. Newman, M.E.J.: Modularity and community structure in networks. Proceedings of the National Academy of Sciences of the United States of America 103(23), 8577–8582 (2006)
17. Marcos, F., Pagnani, A., Lunt, B.: Direct-coupling analysis of residue coevolution captures native contacts across many protein families. Proceedings of the National Academy of Sciences of the United States of America 108(49), 1293–1301 (2011)
18. Marcos, F., Pagnani, A., Lunt, B.: Direct-coupling analysis of residue coevolution captures native contacts across many protein families. Proceedings of the National Academy of Sciences of the United States of America 108(49), 1293–1301 (2011)
19. Girvan, M., Newman, M.E.J.: Community structure in social and biological networks. Proc Natl Acad Sci U S A 99, 7821–7826 (2002)
20. Xue, B., Dunbrack, R.L., Williams, R.W., Dunker, A.K., Uversky, V.N.: Pondr-fit: A meta-predictor of intrinsically disordered amino acids. Biochim Biophys Acta 1804(4), 996–1010 (2010)
21. Dosztanyi, Z.: Prediction of protein disorder based on iupred. Protein Science 17(1), 331–340 (2018)
22. Schymkowitz, J., Borg, J., Stricher, F., Nys, R., Rousseau, F., Serrano, L.: The foldx web server: an online force field. Nucleic Acids Research 33, 382–388 (2005)
23. Li, Q., Peng, X., Li, Y., Tang, W., Zhu, J., Huang, J., Qi, Y., Zhang, Z.: Llpsdb: a database of proteins undergoing liquid–liquid phase separation in vitro. Nucleic Acids Research 48(D1), 320–327 (2020)
24. You, K., Huang, Q., Yu, C., et al: Phasepdb: a database of liquid–liquid phase separation related proteins. Nucleic Acids Research 48(D1), 354–359 (2020)
25. Fu, M., Ray, M., Wang, G.: Acetylation of androgen receptor enhances coactivator binding and promotes prostate cancer cell growth. Molecular and Cellular Biology 23(23), 8563–8575 (2003)
26. Kinoshita, H., Shi, Y., Sandefur, C., Meisner, L.F., Chang, C., Choon, A., Reznikoff, C.R., Bova, G.S., Friedl, A., Jarrard, D.F.: Methylation of the androgen receptor minimal promoter silences transcription in human prostate cancer. Cancer Research 160(13), 3623–3630 (2000)
27. Pang, J., Yang, Y.W., Huang, Y.: P110γ inhibition reduces histone h3k4 di-methylation in prostate cancer. Prostate 77(3), 299–308 (2017)
28. Iksilonen, H.M., Gorad, S.S., Duveau, D.Y.: Inhibition of o-glcnac transferase activity reprograms prostate cancer cell metabolism. Oncotarget 7(11), 12464–12476 (2016)
29. Zhang, L., Altuwaijri, S., Deng, F., et al: NF-kappab regulates androgen receptor expression and prostate cancer growth. Am J Pathol 175(2), 489–499 (2009)
30. Zhang, L., Altuwaijri, S., Deng, F., et al: Transitionally regulated c/ebpα isoform expression upregulates metastatic genes in hormone-independent prostate cancer cells. Prostate 68(12), 1362–1371 (2008)
31. Wishart, D.S., Feunang, Y.D., Go, A.C., et al: Drugbank 5.0: a major update to the drugbank database for 2018. Nucleic Acids Research 46(D1), 1074–1082 (2018)
32. Smith, A.M., Findlay, V.J., Bandurraga, S.G., Griffin, E.K., Spruill, L.S., Liu, A., Golshayan, A.R., Turner, D.P.: Ets1 transcriptional activity is increased in advanced prostate cancer and promotes the castrate-resistant
Figure 1  The circos plot to represent the association of the selected TFs with the shared pathways. Four pathways are found common among the selected molecular regulators including prostate cancer. The color and different size of the ribbon shows the type of relation depending on the p-value of the pathway and its corresponding TFs. The circus plot is used for better visualization of the tabular data. Here, the ribbons are connected between the TFs and their sharing pathway. The extent of the association is represented through a thickness of the ribbon and this thickness is based on the p-values of the pathways for each TFs. Ribbons touch the segment of the inner circle define the row value whereas the ribbons do not touch the segments. Moreover, segmentation provides the absolute scale of the specific region of interaction between TFs and pathFA. The percentage of the outer circle indicates the overall total of each segment respectively.

Figure 2  The final established gene regulatory network depending on the shared pathways of the three molecular regulators. Here red, blue and yellow represent TFs, miRNAs and genes.

Figure 3  A weighted network $G_{DCA}$ and corresponding color modules based on over all residual co-variation from DI score.

Figure 4  The representation of the residues having post transnational modification for each selected TFs.

Figure 5  The overlapped areas obtained from PONDR and IUPred databases. Result from PONDR and IUPred is represented with orange and grey color respectively.

Figure 6  The stability score of the amino acids of six TFs along with their bonds occurred in a particular residue. Torsion, Van der Waals, Chi and Cis-bond are represented with color yellow, green, red and orange respectively.

Figure 7  Top five diseases, based on the p-values, of each TFs are selected to establish the network. Green colour nodes are diseases, connected with all four TFs (red oval nodes). Green colour nodes represent the diseases targeted by five out of six TFs.
### Table 1: Sharing common pathways with the nearest neighbour of the selected TFs from Protein-Protein Interaction having impact on Prostate Carcinoma from Reactome database.

| TF          | Nearest Neighbour | Localization | Pathway                                                                 |
|-------------|-------------------|--------------|------------------------------------------------------------------------|
| AR          | NCOA2             | Nucleoplasm  | Activated PKN1 stimulates transcription of AR regulated genes KLK2 and KLK3 |
|             | KDM1A             | Nucleoplasm  | Activated PKN1 stimulates transcription of AR regulated genes KLK2 and KLK3 |
|             | CCND1             | Nucleoplasm  | Transcriptional regulation by RUNX2                                    |
|             | CREBBP            | Nucleoplasm  | C-type lectin receptors, Cytosolic sensors of pathogen-associated DNA   |
|             | NFKBIA            | Cytosol      | TCR signaling, Signal by B Cell Receptor, NK1 activates NFkB by phosphorylation and activation of IKKs complex |
|             | IKBKB             | Cytosol      | p75 NTR receptor-mediated signalling, Signal by B Cell Receptor         |
|             | IKBKG             | Cytosol      | TCR signaling                                                           |
| RELA        | NFKBIB            | Cytosol      | Signal by B Cell Receptor, NK1 activates NFkB by phosphorylation and activation of IKKs complex |
|             | CHUK              | Nucleoplasm  | Cytosolic sensors of pathogen-associated DNA, C-type lectin receptors   |
|             | NFKB1             | Cytosol      | p75 NTR receptor-mediated signalling, Signal by B Cell Receptor, NK1 activates NFkB by phosphorylation and activation of IKKs complex |
|             | RELA              | Cytosol      | TCR signaling, Signal by B Cell Receptor, NK1 activates NFkB by phosphorylation and activation of IKKs complex |
|             | IKBKB             | Cytosol      | p75 NTR receptor-mediated signalling, Signal by B Cell Receptor         |
|             | IKBKG             | Cytosol      | TCR signaling                                                           |
| RELA        | NFKB1             | Cytosol      | Signal by B Cell Receptor, NK1 activates NFkB by phosphorylation and activation of IKKs complex |
|             | RELA              | Cytosol      | TCR signaling                                                           |
|             | NFKB1             | Cytosol      | Signal by B Cell Receptor, NK1 activates NFkB by phosphorylation and activation of IKKs complex |
|             | RELA              | Cytosol      | TCR signaling                                                           |
|             | NFKB1             | Cytosol      | Signal by B Cell Receptor, NK1 activates NFkB by phosphorylation and activation of IKKs complex |
|             | RELA              | Cytosol      | TCR signaling                                                           |
|             | NFKB1             | Cytosol      | Signal by B Cell Receptor, NK1 activates NFkB by phosphorylation and activation of IKKs complex |
|             | RELA              | Cytosol      | TCR signaling                                                           |
|             | NFKB1             | Cytosol      | Signal by B Cell Receptor, NK1 activates NFkB by phosphorylation and activation of IKKs complex |
|             | RELA              | Cytosol      | TCR signaling                                                           |
|             | NFKB1             | Cytosol      | Signal by B Cell Receptor, NK1 activates NFkB by phosphorylation and activation of IKKs complex |

### Table 2: The moonlighting function of the selected TFs from MoonProt and MoonDB.

| Name of the Protein | Function 1                  | Function 2                                      | Database Used | Search Process |
|--------------------|-----------------------------|-------------------------------------------------|---------------|----------------|
| AR                 | Signal Transduction         | Cellular nitrogen compound metabolic process    | moonDB        | Data repository |
| RELA               | Inositol phosphate metabolic process | Scaffold, binds protein kinase CK2, TCOF1, and upstream-binding factor (UBF) | moonProt      | BLAST          |
| NFKB1              | Carbohydrate metabolic process | Binds to HIF (hypoxia-inducible factor) protein and inhibits nuclear HIF action | moonProt      | BLAST          |
| ETS1               | Signal Transduction         | Positive regulation of transcription from RNA polymerase II promoter | moonDB        | Data Repository |
| CREB1              | ATF2 activating transcription factor bZIP family of transcription factors binds DNA as a dimer | Recruiting Mre11 to IR-induced foci (IRIF) in the DNA damage response, this function does not require DNA binding domain | moonProt      | BLAST          |
| LEBPS              | Not Known                   | Not Known                                       | Not Known     | Not Known      |
Table 3 The List of Alternative Cellular Localizations in Terms of Liquid Liquid Phase Separation

| TF    | Subcellular Localization1 | Subcellular Localization2 | Database |
|-------|---------------------------|---------------------------|----------|
| AR    | Nucleus                   | Cytoplasm                 | LLPSDb   |
| RELA  | Nucleus                   | Nucleus Speckles          | PhasepDb |
| NFkB1 | Nucleus                   | Nucleus Speckles          | PhasepDb |
| CREB1 | Nucleus                   | -                         | PhasepDb |
| CEBPB | Nucleus                   | Nucleus Speckles          | PhasepDb |

Table 4 The 6 TFs and their corresponding expression along with the drugs are reported from DrugBank.

| Transcription Factor | Types of expression | Drug                                                                 |
|----------------------|---------------------|----------------------------------------------------------------------|
| AR                   | under expression    | Oxandrolone, Testosterone, Nandeolone, Phenopropionate, Drostanolone |
| RELA                 | overexpression      | Dimethyl Fumarate                                                     |
| NFkB1                | overexpression      | SC-236                                                                |
| ETS1                 | overexpression      | Not known                                                             |
| CREB1                | overexpression      | Naloxome                                                              |
| CEBPB                | underexpression     | Quercetin                                                             |