REVIEW OF LITERATURE

Review of Literature

2.1) Cancer deaths and environmental carcinogens

Cancer is a broad term that comprises of various non-communicable diseases which are mainly caused due to abnormal cell growth. Cancer has been rated as the second deadliest disease in the world with approximately 9.6 million deaths in 2018 and an estimate of double in the number of deaths by 2040 (WHO, February 2019). It has been reported by World Health Organization that lung cancer is the most diagnosed cancer type (approximately 11.6% of all the cancer cases) and is a leading cause of deaths related to cancer (18.4% of total deaths) (WHO report on cancer, 2020). According to the report of National Institutes of Health, two-thirds of all the cancer cases are caused due to environmental factors. There are around 228 substances reported by NIH in its 10th Report on Carcinogens that cause cancer or are suspected to cause cancer (http://ntp-server.niehs.nih.gov). It is caused by the accumulation of mutations in multiple genes, a phenomenon termed as "oncogene addiction" by Bernard Weinstein (Weinstein and Joe 2006, 2008).

Polycyclic Aromatic Hydrocarbons (PAHs) is one such group of hydrocarbons which are responsible for causing cancer in various organs. These are generated due to incomplete combustion of organic matter and are composed of two or more fused rings

(Bauer *et al* 2018). Benzo[a]pyrene, a compound that falls in the category of PAHs has been placed in group 1 (definite carcinogens) by International Agency for Research on Cancer(IARC monographs volume 92) and has also been listed in the 14th edition of Report On Carcinogens by National Toxicology Program (NTP) as "reasonably anticipated as human carcinogen" (http://ntp.niehs.nih.gov/go/roc). NTP has reported that BaPis capable of causing cancer in more than eight species and can impact different types of tissues. BaP has local and systemic carcinogenic effects owing to various routes of exposure like through food or water, inhalation or by subcutaneous or intravenous injections. There are various sources of BaP present in the environment which makes it almost impossible to escape its impact (Bauer *et al* 2018). The sources include vehicle exhausts, industrial chimneys, grilled food etc.

BaP is a crystalline aromatic hydrocarbon having five fused benzene rings and is formed by incomplete combustion at high temperatures from 300° C to 600° C. BaP enters the cell using aryl hydrocarbon receptors and activates the xenobiotic response element. On enzymatic metabolism, BaP forms bulky adducts that hampers the synthesis of DNA by DNA polymerase activity (Dhasmana *et al.,* 2014). The DNA adducts formed lead to mismatching at the time of DNA replication and methylate the promoters and cause inheritable mutations in DNA which further leads to tumorigenesis (Moorthy *et al.,* 2015).BaP has been reported to activate Aryl Hydrocarbon Receptor (AhR) inducing oxidative stress and DNA adduction, ultimately leading to accumulation of mutations in the cells (Souza *et al.,* 2016).

Important biomolecular targets of BaP include the proteins involved in the cell cycle regulation and pathways leading to cancer. Some of the top protein targets which have been identified in the study are QSOX1, PTGS2, NOS2 and ESR1. Studies conducted by Knutsvik G *et al* in 2016 have reported that high expression of QSOX1 is an important factor in reduced survival of breast cancer patients and can be used as a potent biomarker against breast cancer (Knutsvik *et al.,* 2016). It's over expression facilitates tumor cell invasion and migration at the site of stroma-tumor interface (Lake \ and Faigel 2014). In solid tumors, PTGS2 (COX2) has been reported to modulate the proliferation of cell and apoptosis (Sobolewski *et al*., 2009). In epithelial cancer cell lines, PTGS2 inhibitors arrest the cell cycle by upregulating p21 (Toyoshima *et al*., 2002). It has pro- inflammatory effects which triggers the tumor progression and has been reported to play major roles in cigarette smoking related cancers which includes lung cancer, bladder and gastric cancer (Huang and Chen, 2011). An increased expression of NOS2 promotes tumor growth (Eyler *et al*., 2011) and is responsible for decreased survival in breast cancer (Ambs and Glynn, 2011). ESR1 also has roles in cell cycle progression and has prognostic role in various types of cancers (Thomas and Gustafsson, 2011, Hamilton *et al* 2016 and Gao *et al* 2019).

4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK) is another compound that has been listed in the 14th edition of the Report on Carcinogens and has been placed in group 1 which is the category of "definite carcinogen". It is a tobacco specific compound and is mostly responsible for tumor growth in lungs (Zheng and Takano,

2011). It not only affects lungs but also is a causative agent for stomach, head and neck, pancreatic and liver cancers (Hecht 2002 and Haneef *et al*., 2014). NNK is a potent and hazardous compound found not only in cigarette smoke but also in other forms like snuffs and chewing tobacco and shows its effect not only on the active smokers but also on the non-smokers. Apart from cancer, NNK is also involved in causing significant implications in steatohepatitis (Zabala *et al*, 2015), tuberculosis (Pai *et al.*, 2007) and also in Alzhimers disease (Nunez, 2016).

On activation NNK and its metabolites promotes the formation of DNA adducts and hampers the cell proliferation processes, increases the growth of tumor cells through various receptor mediated processes (Xue *et al*., 2014). They induce methylation, pyridyl hydroxyl buylation and pyridyloxobutylation in nucleobases of DNA (Xue, 2014). It activates the nicotinic acetylcholine receptors and activates the pathways that leading to activations of various processes and pathways that facilitate the growth of tumors. NNK on exposure mutates the genes and alters the chromosomes and it also suppresses the immune system (Ge *et al*., 2015). Our studies published in 2019, shows that NNK has with approximately 550 biomolecular targets participating in various functions performed by cells. Among these 550 proteins, most of the proteins were involved in the cell cycle process (Anukriti.*et al*, 2019). In our study we found that CDK7 followed by CCNA1 (Cyclin A), CDKN1B, CASP8 have shown high binding affinities with NNK. CDK7 plays an important role in cell cycle progression. It activates CDK2 and CDK 1 during the S and the G2 phases (Schachter *et al.*, 2013 and

Larochelle *et al* 2007). CDK7 can phosphorylate all the important CDKs of cell cycle regulation, namely "CDK1, CDK2, CDK4 and CDK6" but only CDK1 and CDK2 have the capacity to phosphorylate CDK7 (Lolli and Johnson, 2007). CCNA1 (cyclin A) is one of the most important components of the cell cycle regulation which determines the fate of the cell. It is activated by two different CDKs in two different phases of the cell cycle namely S phase and M phase (Boer *et al*, 2008). Cyclin A in complex with CDK2 regulates the activation of Cdc25 and CDK1 (Mitra and Enders, 2004). It also promotes the progression of cell cycle to S Phase by interacting with MCM7 (Chibazakura *et al*., 2011). CDKN1B (p27 Kip1) is CDK inhibitor that controls the cell cycle progression at the G1 phase by inhibiting the complex formation of Cyclin D-CDK4/6 and Cyclin E-CDK2. This prevents cells from dividing at a fast rate. Hence it is also known as a tumor suppressor gene (Currier *et al.,* 2019). In various human cancers like lung cancer, prostrate, breast cancer etc, misfunctioning of p27 has been observed due to cytoplasmic mis-localization and degradation of p27 (Sun *et al.,* 2016).

2.2) Systems biology

Systems biology is an interdisciplinary field which is based on the anti-reductionist approach. Unlike molecular biology which focuses on a single molecule at a time, systems biology helps in getting a bird's eye view of whole system. It helps in understanding the involvement of each and every molecule in the system and their dynamics within the system. It also helps in gathering the information of the complex connectivity between all the components of the system (Barabasi and Oltvai, 2004).

Systems biology is based on the graph theory which is a branch of discrete mathematics. The first case study was done by Swiss mathematician Leonhard Euler in 1736, when he published the solution to the Königsberg bridge problem. Now current scenario has become the witness of the origin of a new movement of research in the study of complex networks. These networks can be physical objects in the euclidean space, like power grids, transportation networks, phone call networks, and internet web, social media networks, scientific authorship-citation networks, and also involved in biology and medicine, as neural networks or genetic, metabolic and protein interaction networks (PIN) (Boccalettiet *al* 2006). The ongoing trends of research are very progressive and dynamic, in the term of regular development of newer, efficient and high throughput techniques. The consequence of this progress is generation of huge scientific information on the pubmed and other scientific data warehouses. This huge information is enriching the pubmed like a precious treasure of scientific data and here the "system biology" can play a role as a key to this treasure. It helps in concatenating all the information present for any particular biological entity and its relations with other bio-entities and presents these complex binary interactions in the form of graphs or networks (Koutrouliet *al* 2020). This helps in proper visualization and better understanding of the clusters, integration, management of data, pattern recognitionetc (Mohyedinbonamet *al* 2014). With the help of graph theory, we can also find how robust is the network (Maslovand Sneppen, 2002) and also helps in understanding the evolution patterns (Babuet *al* 2004).

Real world networks are of many types like weighted, unweighted, rooted, directed, nondirected etc. Weighted networks have weights assigned to their nodes or edges which are proportional to the capacity or intensity of interactions among various bio-entities within a network (Barrat*et al* 2004, Barthelemy*et al.* 2005) while an unweighted network doesn't have weights attached to the nodes or edges. Weights are the numerical values assigned to nodes/edges based in their relationship with the adjoining node. Directed graphs helps in knowing the direction of the flow of information and have a one-way relationship between nodes. Transcriptional regulatory graphs are example of directed graphs. These are also known as digraphs (Mason and Verwoerd, 2007, Diestel 2000). Undirected graphs are those in which there is a continuous to and from exchange of information takes place and the vertices have a mutual relationship and all the edges are bidirectional (Diestel 2000; Pavlopoulos*et al*, 2011). In real world, all such types of network are present which are being extensively studied based on the requirements of the hypotheses. In biological sciences extensive researches based on networks are going on to understand the pathogenesis of various diseases, drug discoveries, in understanding various cellular processes, neural networks etc., to name a few.

Protein- protein interaction networks are the ones which depict the physical relationship between different proteins. They help in providing information on how the proteins are interacting and coordinating with each other enabling processes to occur

inside our cells. It helps in proving a comprehensive overview of protein interaction patterns and thus enabling us to study biological processes and their related cellular components. A network consists of nodes and edges where nodes depict the proteins and the edges depict the interaction between those proteins. Edges can be directed as in case of metabolic networks or gene regulatory networks, undirected as in Protein- Protein Interaction networks and weighted. With the help of graph theory, networks are analyzed based on different criteria like node degree, shortest path length, scale free network, clustering coefficient, closeness centrality, betweenness and bottleneck to name a few (Antoniou and Tsompa, 2008, Streib and Dehmer2011).

Node degree is a primary and important criterion for the analysis of the network. It determines the connectivity of any node with other subsequent nodes. Node degree distribution plays an important role in analyzing any scale free network. It provides information on the structural aspects of the graph (Baraba´si and Albert, 1999). In the study of connectomics, degree distribution helps in finding if the network contains hubs or not (Bullmore*et al.*, 2016). This provides an opportunity of randomly selecting a node with degree k and is represented by p(k). Degree distribution gives information related to the scale free nature of the network along with the robustness of the network. It depends on the average degree and is independent of the size of the network (Barabasi, 2016).

Scale free network property is mostly rare but has a widespread use in the study of properties of networks. A network is said to be scale free if some nodes of the network with degree "n" follow the power law distribution $n^{-\alpha}$, where α should be greater than 1 (Broido and Clauset, 2019). Node degree distribution graph helps in understanding if the network under study is scale free or not. A scale free network has many nodes with low degree and few nodes with higher degrees (Batada*et al.,* 2006). These nodes which have higher degrees are also known as hubs and play an important role in maintaining the integrity of a network. Deletion of hub proteins shows lethal effects on the architecture of the network (Kang *et al.*, 2010). Hub proteins are divided into two categories namely date and party hubs. Date hubs act like connectors between different clusters and are usually referred to as dynamic hubs (Ekman *et al.*, 2006). Date hubs are considered to interact with their partners at different times and locations. In a study published by Kim et al in 2006 stated that date hubs can be considered as single interface hubs (Kim, 2006). Party hubs are static hubs and are considered to be the central parts of functional modules. They interact with their partners simultaneously and in a much constrained manner as opposed to date hubs (Changet al., 2013 and Ekmanet al., 2006).

Clustering Coefficient, also known as transitivity, is another important topological property of a network that helps in finding the efficiency of nodes in making clusters. Nodes that are densely connected have high clustering coefficients. Such nodes hold importance in a network as they tend to reflect the protein complexes and

functional modules present within the network. Mathematically it is defined as the ratio of the actual number of interacting edges and the total possible interacting edges (Masuda *et al.,* 2018). Clustering coefficient scores ranges from 0 to 1. 0 clustering coefficient indicates that the node is not connected with any other node while a clustering coefficient score of 1 is the largest score depicting the maximum number of connections (Mulder *et al.*, 2014).

Shortest path lengthis one of the most frequently studied parameter in real world networks. This depicts the efficiency of a network in passing the information (Ye *et al* 2010). Low shortest path length score depicts that the nodes are closely related and are taking part in performing a function. Shortest path length score is usually high for randomly selected nodes in a network (Embar*et al* 2016). Average shortest path length informs about how easily a information navigates within a network (Ren *et al* 2018).

Closeness centralityor farness is another topological property of a network which tells how close is a node with other nodes(Golbeck2013) and how efficiently information passes from that node to other nodes. Its calculation is based on the shortest path length of the node and is expressed as "normalized inverse of the sum of the topological distances in the graph" (https://www.ebi.ac.uk/training/online/course/network-analysis-protein-interaction-data-introduction/building-and-analysing-ppins-1) depicting that any node with highest

closeness centrality score is nearest to all the nodes with shortest path length and passes on the information very efficiently and quickly.

Betweenness centrality was proposed independently by Anthonisse (1971) and Freeman (1977). It is also calculated on shortest path length score and the nodes with betweenness centrality are believed to control the flow of the information. It is the measure of fraction of shortest paths passing through a node. It plays a crucial role in determining the hub nodes (Bullmore*el al.,* 2016). Till date various types of betweenness centralities have been worked out on depending on the range of its applications.It is divided into two categories namely node betweenness centrality which helps in finding the central node of the network and edge betweennness centrality which assists in telling the fractions of paths through which maximum information is supposed to get passed between two nodes (Brandes2008 and Dietz*et al* 2010).

The concept of bottlenecks was first proposed in a study conducted by Yu H *et al* in 2007 on yeast protein-protein interaction networks. Bottleneck scores are of immense importance in finding the key connector proteins in a network. It is considered to be more significant than degree scores and is helpful in finding the dynamic components in any network (Yu *et al* 2007). Proteins with high bottlenecks are considered to be potential targets of any pathogen or drug and prove to be successful candidates in designing drugs against any disease (McDermott *et al.,* 2009). High betweenness scores,

in graph theory are usually considered to be the bottlenecks and they control the flow of information (Chen*et al.,* 2016).

Modulation and GO enrichment analysis are also an important part of analysis of data using systems biology approach. Modulation is a process of removal of noise from the network generated for any biological process. It helps in finding closely related nodes that come together to perform a specific function within the cell. A cluster is a group of genes that are co-expressed (Klipp*et al.,* 2009). To validate the results of modulation, GO enrichment is preferred. There are various algorithms which are used to find the clusters like MINE, MCL, CMC, MCODE, Cfinder, ClusterONEetc.(Xu*et al* 2013). GO enrichment helps in creating a better understanding of the functional profiles of genes present in any biological processes. It is a quantitative method which uses statistical approaches in finding functionally related genes/proteins for a particular pathway or disease (Huang *et al.,* 2009).

2.3) Bio-kinetics of human cell cycle

Mathematical models are the models developed to understand a system using mathematical language and concepts. These are not only generated for understanding the complex mechanisms of regulatory machineries but are also developed to understand the biophysics of cellular pathways, developmental patterns of organisms, dynamics of various disease progressions, neural systems and brain, to name a few (Deutsch*et al.,* 2007). The mathematical models developed for biological systems are applying one of

the three classic approaches namely ODE (Ordinary Differential equations) base modeling, Petri net based modeling and Boolean modeling (Ji *et al* 2017). For the biological networks like the signaling pathways, metabolic pathways, tumor growths or cell-cell interaction networks, ODE approach is the most suited one (Ingalls, 2012). With an increase in the computational capacity, ODE has been frequently used for finding the time dependent variations in a biological process (Ji *et al* 2017). ODE based modeling has been classified into three categories namely "law of mass action", "Michaelis and Menten kinetics" and "Hill reaction". Many scientists like Tyson JJ., Edelstein SJand Goldbeter A, to name a few, have developed mathematical models for various processes in eukaryotic cells (Tyson, 1991, Goldbeter 1991 and Edelstein *et al.,* 1996). Mathematical modeling of biological networks, on intracellular scale helps in understanding how the biomolecules are perturbed in response to external factors and drug molecules (Ji, *et al.,* 2017).Many repositories like Biomodel database has been developed that stores mathematic biomodels depicting biomolecules interacting dynamically at various scales with each other and creating a pathway.

The models created for cell cycle of mammals are extremely complex as compared to those of lower eukaryotes. The cell cycles of higher organisms have much more components/ proteins involved as compared to lower eukaryotes. The regulatory pathways are highly stringent and are controlled by various checkpoints. Cell cycle progression depends on the catalytic activity of the CDKs, though the concentrations of CDKs usually remain constant through-out the cell division cycle. Changes are observed

in the concentrations of cyclins in a sequential and periodic manner as each cyclin appears in specific phases of cell cycle (Blagosklonny, 2001). The transition between different phases is a highly orchestered event and is controlled by different checkpoints. Any change in the proteins involved triggers the checkpoints and the progression of the cell cycle gets hampered. CDKs and cyclins play a central role in the cell cycle progression but apart from these there are many other proteins which play crucial roles in the proper division of cells and prevent cells from becoming cancerous. Various inhibitors like p21, p27 kip1 etc play an important role in keeping a check on the cell cycle progressions by inhibiting cyclin-CDK complexes and thus preventing the cells from becoming cancerous. By measuring the changes in the concentrations of such proteins, conclusions can be made about the behavior of system under a particular situation. It also helps in analyzing the sequential importance of all the components in the system as all the components do not carry equal weightage within the system/pathway (Fischer. 2008).

2.4) Nanoparticles as guard against environmental carcinogens

From past few decades, carbon nanoparticles have paved their ways towards medicines. They are still a subject of study because of their size dependant toxicity (Elhissi*et al.,* 2012). These are rolled graphene sheetswhich on being loaded with drugs (functionalized CNTs) have displayed less toxicity and are non-immunogenic (Bianco et al., 2005, Francisco-Marquez *et al* 2010 and Mohajeri*et al* 2018). Apart from graphenes, single walled carbon nanotubes and multiple walled carbon nanotubes are also of

immense interest as drug delivery systems for many diseases. These carbon nanoparticles are also being used as scavengers of pollutants (Galano, 2008). Studies have been made which prove that nanoparticles can be used as scavengers of environmental carcinogens. TiO_2 nanoparticles have been used in the biological systems as scavengers of BaP to prevent the cells from the hazardous impacts of BaP like cancer (Dhasmana *et al.,* 2014).

Carbon nanotubes are considered to be the main strength of nanotechnology and are often referred as wonder material (Raval*et al.,* 2018). They have extraordinary electronic, physical and chemical properties that make them suitable to be used in various fields like waste-water treatment, super capacitors, green nanocomposites etc (Ong *et al.,*2010).CNTs are also being used as scavenging agents of environmental pollutants. They have high surface area to volume ratio which enables high adsorption of pollutants on their surfaces. Further, the surfaces of CNTS have high free surface energy due to loss of surface atomic coordinates present in the bulk material.

Single walled carbon nanotubes have been successfully used as targeted drug delivery system. They have sp2 hybridized carbon atoms on their surfaces (Lanzani and Luer, 2011). SWCNT when immobilized on microporous ceramic filter was proposed to absorb E.coli and other pathogenic microorganisms from waste water (Estévez, 2008). These have also been conjugated with Human Serum Albumin (HSA) to deliver the drug for breast cancer treatment (Li *et al.,* 2017). Studies have been conducted by Binaco *et*

al in 2011, to make biodegradable and biocompatible nanoparticles to reduce its toxicity and making it suitable for medical purposes (Bianco *et al.,* 2011). Kagan V E *et al* also have stated that carboxylated SWCNTs are degraded by the neutrophiles by their myeloperoxidase activity (Kagan *et al.,* 2010).

Multi-walled carbon nanotubes are two or more concentric graphene hollow cylinders which are arranged coaxially. Like SWCNTS, MWCNTS also find their application in biosensors (Clausen *et al.,* 2012). Wang Et al has developed dopamine and serotonin biosensors using MWCNTS with the help of abrasive technique (Wang *et al.,* 2003). Studies conducted by Mali *et al* and Vashist *et al* separately in 2011 has reported high propensity of fuctionalized MWCNTS across cell membrane and low toxicity, making them potent for drug delivery (Mali *et al.,* 2011 and Vashist*et al.,* 2011).

Fullerenes are allotropic variety of carbon with manysignificant properties owing to its chemical structure and its relatively inert interior cage which facilitates in capturing and isolation of atomic and small molecular species (Rasovik, 2017). It is the first nanostructure made from carbon family and is being actively used as free radical scavenger and antioxidant. They are also being frequently used as drug delivery systems and as gene carriers. They very easily occupy the hydrophobic catalytic sites of HIV proteases, blocking its interaction with the substrate (Bakry*et al* 2007). It is also being used as photodynamic therapy for cancer (Rasovik, 2017).