Chapter 1

Introduction: A case for studying transcription factors in cancer metastasis and development of drug resistance in cancer
Chapter 1: Introduction: A case for studying transcription factors in cancer metastasis and development of drug resistance in cancer

1.1 Why transcription factors in lung and breast cancer metastasis

Alteration of gene regulation is a characteristic feature of disease conditions such as cancer. Genetic anomalies within cancer cells co-operate with extra tumoral microenvironment to generate a transcriptional program unique to a tumor type (Joyce and Pollard, 2009). Predictably, this must involve differential expression and binding of a select group of transcription factors within each tumor type, therefore giving identity to the cancer transcriptome. Metastases or the dissemination and growth of primary tumor to distant sites in body are the most lethal aspects of cancer and account for approximately 90% deaths due to cancer (Gupta and Massague, 2006). Among all known cancers, lung cancer is a leading cause of death in men and women worldwide and accounts for one fourth of cancer deaths (Loberg et al., 2007). Among women, breast cancer is one of the leading killers. Primary tumors in lung frequently metastasize to brain, bone, adrenal gland and liver (Nguyen et al., 2009) Primary breast tumors metastasize to lung, brain and bone among other organs (Nguyen et al., 2009). However, the regulators of metastatic dissemination are largely unknown.

The present work is divided into primarily three sections: i) identification of transcription factors and there coordination in regulation of genes that change during cancer progression in lung and breast cancer; ii) identification of potential master regulators and regulatory networks involved in lung and breast cancer metastasis and iii) test inter-play of transcription factor binding site and nucleosome positions that may be important in metastasis.

This introductory chapter builds a case for studying transcription factors in cancer metastasis and development of drug resistance in cancer.
Chapter 1: Introduction: A case for studying transcription factors in cancer metastasis and development of drug resistance in cancer

1.2 Overview of cancer

Normal cells are delicately in sync with their environment and respond to external signals via tightly regulated signaling pathways that either trigger or repress growth. In order for a cell to undergo mitosis, a growth-promoting signal from the extracellular environment initiates a cascade of events that results in activation of genes that stimulate cell division. With few exceptions, most of the cell populations within an adult organism are terminally differentiated and no longer proliferate. Cancer arises when a cell, for a variety of reasons, escapes the normal constraints placed on its growth and begins to divide in an unregulated fashion.

Many factors are known to be involved in transformation of a normal cell to the cancer cell. (i) Mutation of growth regulatory genes in cancer- Mutation in molecules that regulate signaling pathways which stimulate growth, resulting in a constant `on’ signal to the cell. (ii) Cell Death- Mutation in molecules that regulate process of apoptosis. (iii) Cell-Cell Interaction- loss of cellular adhesion molecules (CAMs) and cadherins molecules. (iv) Angiogenesis-Extensive vascularization. (v) Migration and Metastasis- Spread of cancer cells to other tissue or organ. Figure 1.1 depicts how a cell breaks away from the primary tumor and generates a secondary neoplasm within the organism.

Cancers can be classified in various ways. Most common cancers of adult are carcinomas, representing cancer derived from epithelial cells. Leukemia’s and lymphomas are derived from blood-forming cells and lymphoid cells, respectively. Sarcomas are derived from mesenchymal tissues. Melanomas are derived from melanocytes, and retinoblastomas, and neuroblastomas, and glioblastoma are derived from retina, neurons and glia, respectively.
Chapter 1: Introduction: A case for studying transcription factors in cancer metastasis and development of drug resistance in cancer

Figure 1.1 Metastatic progression Individual cells within the primary tumor upregulate production of specific proteases, which gives the cell the ability to degrade ECM or basement membrane. Cells then break away from the primary tumor and begin to migrate. Migrating cells adhere to connective tissue and invade dermal tissue. Cells thus enter the vascular system by migrating between endothelial cells and moving through blood vessels. Lastly, tumor cells reach secondary sites where conditions are favorable for their continued growth.

1.2.1 Genetic basis of cancer

More than 300 genes were shown to play a role in development of cancer (The Cancer Genome Project, 2011). Earlier, based on role of cancer genes in development of cancer there were two main classes-oncogenes and tumor suppressor genes. Now a third class, a sub-type of tumor suppressor genes called the caretaker genes are also recognized (Vogelstein and Kinzler, 2004).
Oncogenes-Genes who have the ability to cause cancer, in other words, gens that can induce normal cells to grow out of control and become cancer cells are called oncogenes. These are generally produced in the mutated form or expressed in higher levels in tumor cells. In normal cells these genes are therefore termed as proto-oncogenes, i.e., a normal gene that can become an oncogene. Proto-oncogenes normally control cell division rate and the degree of differentiation. Due to mutation in proto-oncogenes (usually through amplification, translocation or point mutation) the mutant form (oncogene) becomes permanently ‘turned on’ or activated. This subsequently causes quick cell division resulting in formation of a tumor. Only one of the two alleles of a proto-oncogene needs to be overactive in order to have an oncogenic effect. An example of an oncogene is BCR-ABL. This is formed by a translocation between sections of human chromosomes 9 and 22, resulting in what is known as the Philadelphia chromosome. The resulting Bcr-Abl protein interferes with signaling pathways within cells and causes the over-production of white blood cells, leading to chronic myeloid leukaemia (CML) (Konopka et al., 1985). Similarly, B-RAF, HER2, MYC, RAS and VEGF have been extensively studied and have been established as oncogenes across many cancer types.

Tumor suppressor genes-Tumor suppressor genes are involved in many inherent cellular processes including regulation of cell division, DNA repair, and apoptosis. Their inactivation promotes uncontrolled cell growth, which can lead to cancer development. In this class several genes were characterized such as APC, CDKN2A, RB, TP53 and VHL across many cancer types. TP53 gene is the most frequently inactivated gene in human cancers. It normally encodes a pivotal transcription factor
that puts a brake on abnormal cell growth and triggers apoptosis in cells that have sustained DNA damage.

**Caretaker genes**—These are a class of tumor suppressor genes but, as they do not regulate cell division, their inactivation is not directly responsible for cancer development. Many caretaker genes (like *ATM, BRCA1/2, MLH1, MSH2 and XPA*) also known as ‘stability genes’ (Vogelstein and Kinzler, 2004), are involved in recognizing and repairing DNA damage. Their inactivation causes genetic instability and leads to increase in the mutation rate of all genes, including oncogenes and tumor suppressor genes (Vogelstein and Kinzler, 2004).

### 1.2.2 Cancer Therapy

Currently therapies are using a combination of cytotoxic agents for killing cancer cells. While many of these therapies are specific to a particular cancer type; little or no affect is shown against other cancer types and moreover, often recurrence is found within a few years. As any single drug is ineffective against cancer some cancer cells survive and cause reappearance of outgrowth. Therefore, multiple drugs are necessary, but this multi-targeting is limited by toxicity to normal cells. Some of the widely used drug categories are the following:

i) **Antagonists of Growth Factors**—This category drugs target sex hormones which promote growth of cancer cells (breast, ovarian and prostate). Tamoxifen is one such drug which targets sex hormone oestrogen.

ii) **Blocking S Phase**—Many drugs are used to inhibit DNA synthesis by designing a small molecule structurally similar to metabolic compounds required in the synthesis of DNA and cell survival. As examples, fluorouracil is structurally
very similar to uracil, which is needed for DNA synthesis, and methotrexate is an analogue of the vitamin folic acid, also essential for DNA synthesis.

iii) DNA-damaging Agents-Drugs that damage DNA are effective in cancer treatment as cancer cells have increased growth rates relative to normal cells. In addition normal cells have effective DNA damage repair mechanism unlike cancer cells. Cytoxan and Cisplatin are widely used drugs in this category.

iv) Mitotic Inhibitors-Several drugs upset the mitotic mechanism in cancer cells. These include taxol from the yew tree and alkaloid toxins from the vinca plant.

Despite the development of new targeted anticancer therapies, there is no treatment that is 100% effective against cancer. This problem is compounded many folds by the incidence of cancer metastasis in addition to the development of drug resistance in cancer cells. Additional knowledge about these mechanisms may help to design strategies for better treatment and management of cancer.

1.3 Cancer Metastasis

Cancer is a term used for diseases in which cells without regard for normal growth limits aggressively proliferate and are able to invade surrounding and adjoining tissues. The process by which cancer cells spread from the site of origin to distant locations within the body is called metastasis. It has been reported that metastasis is responsible for as much as 90% of cancer-associated mortality (Gupta and Massague, 2006). Thus, in addition to early diagnosis, understanding the mechanisms of metastasis and halting its course are requisite for effective cancer therapy.

Metastasis is an active process involving a number of genetic and/or epigenetic changes. As a result of these alterations cancer cells become capable of implanting in
foreign locations avoiding immune response and proliferate independently of growth factors or growth inhibitory signals. Virtually all cancers can form metastatic tumors. A detailed understanding of the molecular and cellular mechanisms behind each of these steps is imperative for insights into these complex processes. Metastasis is generally observed in later stages of cancer. The different pathological stages during metastasis include local invasion of a primary tumor, then migration to blood vessels or lymph nodes, and finally distal metastases. The process involves migration of a cancerous cell out of the original location, overcoming barriers to implantation in a foreign location, subsequently dividing in an uncontrolled mode, and/or metastasizing further. Gene expression profiling studies (Wang et al., 2004) suggest that metastatic potential is intrinsic to all tumor cells and that metastatic spread could be an early event in tumorigenesis. However, there are some evidences for the existence of a small population of cells within tumors, which exclusively are capable of metastasis, the so-called cancer stem cells. Whereas in most cancers the cell cycle checkpoint arrest is overcome by at least one transformative event early in cancer development in the traditional model of metastasis, a second genetic or epigenetic event is usually necessary for transition of a non-metastatic tumor to a metastatic one.

1.3.1 Metastatic Process

Cancer metastasis involves a series of interrelated steps, each of which can be rate limiting and failure in any state can shut down the entire metastatic process. Metastasis consists of (1) detachment of epithelial cells from the extracellular matrix (ECM), (2) survival within the bloodstream, and (3) growth at the metastatic site (Figure 1.2).
Chapter 1:  

Introduction: A case for studying transcription factors in cancer metastasis and development of drug resistance in cancer

Figure 1.2 Stages of metastasis (A) Tumor; (B) detachment of tumor cells from the ECM; (C) intravasation of tumor cells into the bloodstream; (D) transport of tumor cells through the bloodstream; (E) extravasation of tumor cells at distal site; and (F) growth of metastatic lesion

For prevention and cure of cancer it is of utmost importance to understand the biological basis of metastasis as it is now understood to be the lethal aspect of cancer. The existence of lymph node metastases in a cancer is a strong indicator of survival in patients as well as an indicator of whether other distal metastases will develop. In some cancers, lymph node metastases are better indicators of distant metastasis than in other cancers. For example, in head and neck cancer, a strong correlation was observed between lymph node metastases in the neck halves and survival rate of patients (Leemans et al., 1993). Moreover, only 7% of patients whose necks are free of lymph node metastases develop distant metastases. This observation seems to indicate that metastasis somehow involves infiltrating the lymphatic system and using it to migrate through the body. In other cancers such as breast cancer, however, 20-
30% of patients whose axillary lymph nodes are free of disease still develop distant metastases. Thus, there seems to be another distinct pathway of metastatic cell dissemination that is independent of the lymphatic system. This second pathway has been shown to utilize hematogenous routes in the vascular system. Even so, the presence of axillary lymph node metastasis continues to serve as a good indicator of whether the disease will spread (Braun et al., 2000).

The default fate of normal cells is apoptosis or programmed cell death but cancer cells in general are refractory to apoptotic signals and their mitotic division is independent of stimulation by growth factors. Therefore, development of tumor is directly correlated with resistance to apoptosis, and thereby it also increases metastatic potential of the tumor. However the reasons underlying this remain poorly understood. A possible reason might be that some cells within the tumor have ability to undergo mitosis independently of growth factors and become metastatic; alternatively, molecular events which give rise to metastatic growth also cause resistance to apoptosis. Therefore, it would be interesting to see the levels of pro-apoptotic/anti-apoptotic genes in metastatic tumor.

Recent development in cancer biology suggest a special group of cells within a solid tumor, cancer stem cells, attain increased ability to migrate and colonize secondary tissues. Additionally, small populations of cells in many malignant tumors, including acute myeloblastic leukemia (AML) (Lapidot et al., 1994), glioblastoma (Singh et al., 2003), small cell lung cancer (Krystal et al., 1996), non-small cell lung cancer (Kim et al., 2005), malignant melanoma (Klein et al., 2007), and breast cancer (Al-Hajj et al., 2003), display properties reminiscent of stem cells, the cell group that indefinitely retains the property of self-renewal by mitosis (Pardal et al., 2003). It is conceivable that migration of these cells could in principle lead to metastasis and
successful colonization at a distant site, and this may explain why successful primary metastasis is a relatively rare event. Strong evidence implicating these “cancer stem cells” in metastasis is provided by the observation of overlap between the genes and signaling pathways necessary for normal stem cell motility and those for metastatic cancer cells (Kucia et al., 2005).

1.3.2 Genetic basis of metastasis

Given the complexity of the metastatic process, genetic mechanisms behind it are likely to be complex. Studying genetic basis of metastasis provide better understanding about biological basis of the process and identification of molecular signature for better diagnosis, prognosis, and therapeutic intervention to restrict metastasis. A study by Ramaswamy et al. is a groundbreaking step in understanding the genetic basis of metastasis (Ramaswamy et al., 2003). In this study the authors measured genome-wide gene expression profiles of 12 samples of confirmed metastatic adenocarcinomas of diverse origins (breast, lung, prostate, uterine, and ovarian cancers) and compared them with those obtained from 64 confirmed non-metastasizing cancers of the same types. The comparison yielded a descriptor transcript set of 128 genes, of which 64 were overexpressed and 64 were underexpressed in the metastatic cancer samples. The descriptor gene set did not provide any obvious set of genes with related function. In fact, some genes that are underexpressed are unexpected (e.g., MLC2) and others that are overexpressed are of unknown significance in the context of metastasis (e.g., glucose phosphate isomerase). From this gene set, the authors derived a core gene expression signature with a refined set of 17 metastasis markers (8 overexpressed and 9 underexpressed). These 17 genes performed well as predictors of metastasis on other unrelated tumors with or without metastasis. This refined core set of genes also included several genes whose expression signatures defy simple logical expectations: e.g., actin
Chapter 1: Introduction: A case for studying transcription factors in cancer metastasis and development of drug resistance in cancer

gamma 2, myosin heavy chain 11, and myosin light chain kinase genes are underexpressed, whereas lamin B and type 1 collagens a1 and a2, which are known hallmarks of metastasis 39 overexpressed. Thus, despite the obvious utility of this core set of gene expression signature markers for more accurate prognosis, a biological understanding of their basis was not obvious.

A biologically insightful understanding of metastasis has come from identifying genes that suppress tumor metastasis using several different in vivo metastasis assays (Welch et al., 2000). Metastasis suppressor genes, or MSGs, are a special class of genes that are turned off in metastatic cells but, when re-expressed, inhibit metastasis without affecting tumorigenicity. At least 30 suppressor genes have been identified to date (Horak et al., 2008; Hurst and Welch, 2011; Yoshida et al., 2000), beginning with the discovery of NM23 in 1988 (Steeg, 2003). Other MSGs include NME1, a member of the nucleoside diphosphate kinase family of proteins implicated in cell cycle regulation, KISS1-a regulator of metalloproteases and a ligand of a G-protein-coupled receptor (Ohtaki et al., 2001), a mitogen activated protein kinase gene (MKK4), and BRMS1 which functions in gap junctions and reduces motility. Each of these genes provides interesting anchor to the spectrum of events thought to be responsible for distinct cellular stages of metastasis. In lung cancer, the invasion suppressor CRMP1 (collapsin response-mediator protein 1) has been identified as an invasion suppressor, but its efficacy as an MSG has only been demonstrated in vitro and thus has yet to be validated as a true MSG (Steeg, 2003).

1.3.3 Cellular program for metastasis: Epithelial-Mesenchymal Transition (EMT)

In cancer metastasis detachment of cancer cells from primary tumor is a first step; this is followed by entry into the bloodstream. Change in motility of the cancer cell is major factor which enables their separation from the primary tumor. Cells display two
major types of motility in early steps of metastasis. First and the most common version is epithelial mesenchymal transition (EMT)—the cell elongates, secretes extracellular enzymes to locally degrade the ECM, and migrates out. While TGF beta in cooperation with the Ras-GTPase signaling pathways can induce EMT, it is unclear whether cells that are so transformed indeed constitute metastatic cells (Oft et al., 2002). The other, more aggressive, type of motility is amoeboid translocation: the elongated cells take on a spherical morphology, and these spheroidal cells deform through pre-existing gaps in the ECM to disseminate into the bloodstream (Figure 1.3). There is also a third, rarer, form of cellular motility called collective migration that involves simultaneous mesenchymal motion of a cluster of cells.

![Diagram of types of motility](image)

**Figure 1.3 Types of motility in early steps of metastasis** (A) Mesenchymal motility, (1) Epithelial tumor cells in the ECM. (2) Elongation of tumor cell and degradation of ECM. (3) Degradation of ECM and mesenchymal migration. (B) Amoeboid motility, (1) Epithelial tumor cells in the ECM. (2) Deformation through pre-existing gaps in the ECM. (3) Amoeboid migration through cracks in the ECM.

Few studies suggest that certain transcriptional regulators may control entire sets of motility genes. For example, AP-1 transcription factor activity is correlated with
expression of cell motility genes (Ozanne et al., 2000). Twist, Six-1, and BRMS1, all transcriptional regulators, have also been implicated. Other genes involved in cell-cell signaling such as ErbB1, encoding epidermal growth factor receptor (EGFR), are implicated in cancer cell motility but not necessarily in growth of the primary tumor. Genes or proteins that are specifically implicated in cell motility in metastatic transformation are potential drug targets.

After detachment from primary tumor the next challenge faced by the migrating cancer cells is survival without normal matrix components and evade anoikis. EMT plays a significant role in survival of cancer cells in circulation. This is based on the finding that during EMT in breast cancer, levels of BCL2 increase, which in turn increase the metastatic potential of breast cancer epithelial cells by inhibiting matrix-degradation-induced apoptosis but does not affect primary tumor growth or cell motility (Martin and Leder, 2001; Pinkas et al., 2004).

The developmental signaling pathways Wnt, Notch, and Hedgehog have also been linked to EMT (Eccles and Welch, 2007). Wnt signaling is of particular interest as it has been associated with collective migration and is aberrantly activated in lung cancers (Mazieres et al., 2005). With regard to metastasis, loss of signaling by Wnt1, the first of the Wnt proteins to be discovered, has been shown to reduce the size of lung metastases in mice (You et al., 2006).

In a study to identify gene expression signatures associated with the propensity for metastasis, invasive breast cancer tumor cells were collected in vitro by virtue of their chemotactic ability and their mRNA expression levels were assayed in relation to their less invasive counterparts (Sahai, 2005). Genes associated with motility were most strikingly differentially regulated in invasive cells compared to those in non-
invasive cells. For example, coflin, Arp2/3 complex, and capping protein, all involved in lamellipod protrusion, extension, and tail retraction, were coordinately upregulated. By contrast, the ZBP1 gene is strongly downregulated in invasive breast carcinoma cells. ZBP1 binds to beta-actin mRNA and localizes the mRNA to the leading edge of cells.

Role of transcription factors in induction of EMT is now well appreciated. As several transcription factors like TCF3, TCF4 and TWIST1 have been shown as inducers of EMT by suppressing CDH1 (Ansieau et al., 2008). But a large scale study for transcription factors in EMT is not done till now.

1.4 Drug resistance in Cancer

Innate or acquired resistance of cancer cells is a serious obstacle in the chemotherapeutic treatment of cancer. Cancer cells develop drug resistance by dysregulation of various processes which include drug metabolism (decrease in drug activation or increase in drug inactivation), transport (decrease in drug influx or increased drug efflux), drug target alteration (increased target levels, altered function of target or decreased affinity of drug for target), DNA repair (increased repair of DNA damage or mutations in mismatch repair genes) and apoptosis (altered function and regulation that change the ability of tumor cells to execute programmed cell death) (Figure 1.4). Among these mechanisms, the best characterized mechanism is Multi-Drug Resistance (MDR) involving P-glycoprotein 170 (Pgp) (Kartner and Ling, 1989; Juliano and Ling, 1976; Ueda et al., 1999). Three types of MDR have been characterized till now: (i) classical Pgp-mediated MDR, in which the causal relationship between resistance and over-expression of Pgp has been clearly demonstrated in vitro; (ii) atypical MDR, which involves DNA topoisomerase interacting
drugs, and (iii) non-Pgp-mediated MDR, in which increased level of MRP, lung resistance protein (LRP) or the expression of other members of the ABC transporter gene family was observed (Gottesman and Pastan, 1993; Kuska, 1999). The US National Cancer Institute (NCI) has played a very important role in antitumor screening and testing. In 1985, the NCI performed a drug-screening programme on cancer by: (1) in vitro screening using an organ-site oriented approach, (2) using human solid tumor cell lines, (3) using small amounts of compounds for the primary screening, and (4) exploring natural products from diverse sources (e.g. plants, marine organisms and microbes).

**Figure 1.4 Deregulated processes in development of drug resistance** Image to summarize many of the ways in which cultured cancer cells have been shown to become resistant to cytotoxic anticancer drugs.

During 1990-1992, the NCI launched a panel of 60 human tumor cell lines for its drug screen programme. The main feature of the screening program is in vitro exposure (48 h) of a panel of 60 human tumor cell lines at five different concentrations with 10-fold serial dilution. Crude extracts were mostly evaluated at a single concentration. The
SRB (sulforhodamine B) protein stain assay was used to generate 60 dose effect curves for each compound. It should be noted from the mass-action law principle that 1000-fold dilution with five or more dose densities would be too broad to have smooth dose-effect curves, although it would provide some efficacy indications. This programme also provides: the development of the `mean graph' (Paull et al., 1989) to interpret the relative sensitivities of each of the tumor cell lines to an individual agent and the development of computer-based `Compare Program' (Paull et al., 1989) for recognition of these patterns of effectiveness.

1.4.1 Metastasis and Drug Resistance

The process of tumor metastasis is highly selective and consists of a series of sequential and unified steps (Fidler, 1990). Despite improvements in diagnosis, surgical techniques, patient care, and adjuvant therapies, most deaths from cancer are due to metastases that are resistant to conventional therapies. The major obstacle to effective treatment is the biological heterogeneity of tumor cells. Moreover, metastases can be located in lymph nodes and different organs, and the specific organ microenvironment influences the biological behavior of metastatic cells, including their response to systemic therapy (Fidler, 2002). One of the factors is the development of drug resistance phenotype in metastatic cancer cells (Dutour et al., 2007).

1.4.2 Gene signature for Drug Resistance

Cell lines were extensively used in the identification of drug resistance mechanisms. While cell lines differ in important ways from tumor tissue, the tumors themselves are often inaccessible for experimentation, so cell lines represent convenient experimental
Chapter 1: Introduction: A case for studying transcription factors in cancer metastasis and development of drug resistance in cancer

models. Genomic profiling has been used to correlate chemosensitivity patterns with baseline gene expression in a large number of untreated cell lines. Several studies have identified correlation between gene expression and resistance using the NCI60. The NCI60 is a panel of 60 cancer cell lines derived from a variety of tissues and organs and extensively studied by the National Cancer Institute’s Developmental Therapeutics Program (DTP). The DTP measured sensitivity of these 60 cell lines to common chemotherapeutics (http://dtp.nci.nih.gov). Unsupervised gene-expression clustering on NCI60, identify five large clusters in sensitive group of 118 drugs that correspond to mechanisms of action: DNA and DNA/RNA metabolites, tubulin inhibitors, DNA-damaging agents, topoisomerase 1 inhibitors, and topoisomerase 2 inhibitors (Scherf et al., 2000). These authors then identified gene-drug pairs where specific expression levels were found to be associated with drug sensitivity (e.g., 5-FU sensitivity was negatively correlated with dihydropyrimidine dehydrogenase (DPYD), the rate-limiting enzyme in endogenous uracil and thymidine catabolism and in exogenous 5-FU catabolism). Such drug-gene pairs were interpreted as proof of principle that baseline gene expression signatures can indict important pathways in chemoresistance.

In a complementary supervised approach, (Staunton et al., 2001) used a subset of NCI60 lines to develop a gene expression-based predictor of chemosensitivity. Using the sensitivity and resistance data for 232 chemical compounds and gene expression levels of 6,817 genes, they also observed similar mechanisms of action reported with the unsupervised approach. They then performed supervised analysis to search for genes that correlated with the clusters of drugs. This study identified genes that were significantly positively or negatively correlated with at least two of four anti-
metabolite drugs (for example 5-FU, carmofur, tegafur, and doxifluridine). Some of the gene expression patterns were shared by multiple mechanistic classes of drugs, but many were also unique to specific mechanisms. Finally, these authors demonstrated that gene expression-based classifiers of sensitivity or resistance gave improved prediction of sensitivity or resistance relative to what would be expected by chance. Both studies used genome-wide expression profiles of NCI60 data without selecting the specific genes in advance. This is the most common approach in microarray studies, particularly with the commercial availability of whole genome arrays that represent the vast majority of genes in the human genome. However, it is also possible to use the same signature-based approach with a more focused gene set. An example of this approach is afforded by the studies of Szakacs et al. (Szakacs et al., 2004), who focused on 48 known human ABC transporters and used real-time RT-PCR to obtain quantitative measures of gene expression. Each of these studies has the advantage of being able to efficiently study many different chemotherapeutics at the same time because only untreated profiles were used. The expression responses to treatment were not identified. Molecular profiling in response to treatment can provide more useful information about mechanisms of drug resistance. In view of that, experiments using a panel of breast cancer cell lines treated with two different chemotherapeutics, doxorubicin and 5-fluorouracil, showed that toxicant- and or mechanism-specific signature was identifiable (Troester et al., 2004a), but the predominant signature was that of general stress (Troester et al., 2004b). Thus in addition to characterizing how baseline characteristics affect chemosensitivity, improved understanding of how baseline characteristics of
individual cell lines predict stress responses may help to provide mechanistic insights into drug response. This gains further support from cell line models of luminal breast tumors, which have different signaling responses to chemotherapy treatment than cell line models of basal-like breast tumors (Troester et al., 2004b).

Tumor heterogeneity is a challenge in choosing representative cell lines for drug response studies. To address this issue, Neve et al. have assembled a collection of breast cancer cell lines for studying functionally distinct breast cancer subtypes (Neve et al., 2006). Interestingly, authors found heterogeneity in copy number and gene expression that was similar in cell lines and in primary tumors. They also showed that cell lines can also be classified based on expression profiles into tumor subtypes such as luminal and basal-like.

Recently, Holbeck et al., profiled the expression of the 48 human nuclear receptors (NRs) by quantitative RT-PCR in 51 human cancer cell lines of the NCI60 collection derived from nine different tissues, and identified druggable targets that may be pharmacologically manipulated for potential therapeutic intervention (Holbeck et al., 2010). Similarly, using transcriptomic and metabolomic measurements from the NCI60 cell line panel relationships between biological pathways and drug response was explored (Cavill et al., 2011).

1.5 Transcription factors in cancer

Transcription factors (TFs) are conventionally defined by their ability to bind specific DNA sequences and regulate transcription. TFs play very important role in several
biological processes, as they are often identified as key metabolic or developmental regulators. They can perform their function in many ways, for example, stabilization or blocking the binding of RNA polymerase to DNA or interaction with other TFs and/or chromatin proteins. Dysfunctions of TFs have been observed in many diseases.

Three main groups of transcription factors are known to be important in human cancer. The first to be recognized were the steroid receptors as oestrogen receptors in breast cancer and androgen receptors in prostate cancer. Anti-oestrogen and anti-androgen compounds such as tamoxifen and bicalutamide have been in clinical use for many years.

The second group of transcription factors that were recognized to have a role in cancer are resident nuclear proteins, which are activated by serine kinase cascades (Brivanlou and Darnell, Jr., 2002). In this category JUN was found as an oncogene in late 1980’s. First, the retrovirus (v-jun) and then the normal cellular form, c-JUN was identified. Thereafter; transcription activity of JUN was identified based on its similarity in a DNA-binding domain to an already identified yeast transcription factor, GCN4 (Vogt et al., 1987). This highly significant discovery provided the first concrete example of involvement of any TF in cancer other than steroid receptors. Subsequent work showed that c-JUN was phosphorylated on serine 63 and serine 73 following activation of signaling cascades, which made c-JUN maximally effective in stimulating transcription (Derijard et al., 1994). Additionally, the Fos, ErbA and Myc were identified as oncogenes, as high expression of these proteins was found in many cancer types. On the other hand similar boost to our understanding of growth
regulation in normal cells has also emerged from studies of the anti-oncogene proteins. Thus, studies on the Rb-1 gene, which was originally identified on the basis of its inactivation in retinoblastomas, have led to an understanding of its key role in regulating the balance between cellular growth and differentiation. Similarly, work on p53, which was originally identified as a protein interacting with the product of the SV40 large T oncogene, has led to the identification of its key role as the so-called ‘guardian of the genome’.

The third group of transcription factors that have oncogenic potential, and the most recently recognized, are latent cytoplasmic factors, activation of which is normally triggered by receptor-ligand interaction at the cell surface (Brivanlou and Darnell, Jr., 2002). The latent cytoplasmic proteins can be activated directly by tyrosine or serine kinases at the cell surface, or by various different cytoplasmic biochemical events that also feature kinases (some regulated by Ca\(^2+\) flux) or specifically regulated proteolysis. Despite their dissimilarity in detail, these pathways are all similar in that a protein–protein interaction at the cell surface triggers cytoplasmic events, which result in delivery to the nucleus of a protein that increases transcription through interaction with one or more of the many proteins that affect the initiation of transcription (Figure. 1.5).

These studies suggest that targeting transcription factor could be effective approach for cancer treatment. This suggests that screening for transcription factors in cancer and associated problems (metastasis and drug resistance) may be of interest.
Introduction: A case for studying transcription factors in cancer metastasis and development of drug resistance in cancer

Figure 1.5 Generalized signaling pathway The human genome contains information for several thousand cell-surface protein receptors and ligands for these receptors (in general, small secreted proteins). The diagram summarizes key reactions. Latent transcription factors might be activated at the cell membrane by either tyrosine (STATs) or serine (SMADs) phosphorylation. They then bind to the transport proteins called importins and enter the nucleus to participate in regulated gene transcription. In some pathways, the kinase enters the nucleus and effects an activation of resident nuclear-transcription factors. The active factors (in combination with DNA) bind other transcription factors (forming an enhanceosome) that attract co-activators (this includes several dozen possible proteins), leading to the final step of attracting the actual RNA synthesis machinery that includes RNA polymerase II (Pol II) and general transcription factors (GTFs).

1.6 Transcription regulatory network in cancer

Analyzing transcription regulatory network in cancer is an active field of research. The main building blocks for transcriptional regulatory networks are transcription factors and their target genes. The goal to build such network is to “explain” observed expression
patterns in cancer transformation. Such studies would provide us a network of contributory regulatory factors which explain the observed expression changes.

In particular for yeast, this line of investigation has proven to be fruitful in revealing regulatory principles that are not detectable from the mere analysis of expression data (Janga et al., 2008). Early studies, for instance, could link changes in the structure of regulatory networks to the type of external stimuli and the corresponding transcriptional response (Luscombe et al., 2004). Such impressive interrogations were made possible by the systematic experimental mapping of yeast transcription factor binding sites using Chromatin-Immunoprecipitation on chip (ChIP-chip) experiments. Unfortunately, the systematic experimental charting of human transcription factor binding sites is still at a very early phase with experiments being limited to a small number of transcription factors and cell types. At present, many collections of transcription factor binding sites for humans thus rely considerably on in silico matching between promoter regions and position weighted matrices describing the consensus binding sites of transcription factors. Further difficulties in the construction of comprehensive regulatory networks are (1) a high number of false positive predictions of transcription factor binding sites based on simple sequence matching, (2) the choice of an adequate size of human promoter regions, (3) the combinatorial action of transcription factors within cis-regulatory modules and (4) the influence of the-generally unknown-chromatin structure on the accessibility of binding sites.

Despite these challenges, first effort has been undertaken to construct genomewide regulatory networks for cancer research. Notably, Kluger and colleagues examined the topological properties of regulatory networks to characterize gene deregulation during tumorigenesis (Tuck et al., 2006). For construction of a regulatory network,
they utilized a collection of transcription factors stored in the Transcription factor (TRANSFAC) database. Potential target genes were determined by position weighted matrices. The basal connectivity network was then intersected with co-expression of genes from different cancer microarray studies to obtain condition-specific regulatory networks. In the subsequent analysis, network features such as degree distributions were used to differentiate between diseased and healthy patient samples. Although no significant improvement of classification accuracy was achieved compared to conventional microarray analysis, the procedure offered some valuable insights in the potential causative mechanisms of gene deregulation. Most intriguingly, genes that discriminate best between disease conditions tend to be highly localized on the transcriptional network. It is important to note that the applied strategy implies that expression levels of transcription factors can be proxies for their activity states. However, this might neglect important post-translational modifications.

An impressive project, which can also serve as a prime example for integrative network biology, is the assembly and analysis of the B-cell interactome by Califano and co-workers. This model of the molecular network for B-cells not only includes transcriptional regulatory, but also protein-protein and modifying post-translational interactions derived from a variety of experimental and computational resources. In the study by Mani et al., a strategy was developed to scrutinize the B-cell interactome for dysregulated interactions in three distinct types of lymphoma (Mani et al., 2008). In contrast to conventional microarray analysis focusing on the differential regulation of genes, the loss or gain of correlation between interacting genes was analyzed. Remarkably, the examination of dysregulated interactions pointed more clearly to the set of known genetic lesions than simple differential gene expression did.
Furthermore, potential downstream effectors could be identified which would have been missed using gene expression alone. Notably, these results probably would not have been derived without the construction of a cell type-specific network.

1.7 A strategy to build transcriptional regulatory network in progression of cancer

Establishment of a primary tumor is seldom fatal; it is the malignant transformation of tumor entrusting capacity of metastatic dissemination to cells of the primary tumor and subsequent growth at secondary sites in the body that poses clinical challenges. Viewed in this way, cancer is an infectious disease within a single body; originating at a single site and subsequently spreading to multiple other organs. While the genetic changes including activation of oncogenes and inactivation of tumor suppressors serve to confer selective growth advantage to cancer cells at the site of the primary tumor, it seems, however, that metastatic reprogramming including dissemination and survival in transit may require additional set of genetic changes.

One of the most important features of malignancy is localized invasion of the basement membrane, survival in circulation, organ specific colonization, and establishment of metastatic growths. Such changes necessarily require conversion of the epithelial nature of cells to undifferentiated namely mesenchymal type. Collectively termed epithelial to mesenchymal transitions (EMT) these events draw parallels with large scale but developmentally timed and co-ordinate cell movements during gastrulation within the embryo. Adoption of this developmental process to explain steps in metastasis is now shown by several experimental studies across many cancer types (Polyak and Weinberg, 2009). Interestingly, a large scale study for
transcription factors in EMT has not been addressed. Therefore we aim to build a transcription regulatory network to identify genetic changes and regulatory factors which orchestrate these changes in EMT. As a case study we selected lung and breast cancer as they have highest mortality rate among men and women worldwide. Additionally, several studies have shown EMT in lung and breast cancer (Foubert et al., 2010; Hung et al., 2009; Jiang et al., 2011; Shih and Yang, 2011).

Our approach uses the following strategy: a) test an approach of building a transcription factor binding site (TFBS) network using predicted TFBS sites; b) build TFBS network for EMT in lung and breast cancer using available experimental dataset from public domain and c) identify master regulator of lung and breast cancer progression and determine their regulatory network. Details of implementation of each part of the strategy and results obtained therein are described as individual chapters.