Chapter 4

Building and analyzing transcriptional regulatory network centered on the metastases suppressor NME2 in breast and lung cancers
4.1 Introduction

Biological processes are driven forward by spatial and temporal changes in regulatory state, that is, control by sets of transcription factors (TFs), the underlying control process is often described as a transcriptional regulatory cascade or network. TFs ability to regulate large number of genes that belongs to different functional classes makes them global transcription regulators. They control a complex regulatory cascade not only by directly controlling the expression of specific genes, but also indirectly regulating various cellular pathways by controlling local regulators. Recently, several studies used different approaches to characterize such regulators in various processes. Notably, identification of transcriptional regulatory network maintaining pluripotency of embryonic stem cells (Kim et al., 2008), transcriptional network for mesenchymal transformation of brain (Carro et al., 2010), unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses (Amit et al., 2009).

Nucleosome positioning along chromatin affects accessibility of the genomic DNA in vivo. Nucleosomes may bind to some genomic regions tightly and prevent transcription factors from approaching their sites. Alternatively, strategically positioned nucleosomes can promote long-range DNA bending and allow distal enhancers to interact with the transcriptional machinery (Lee et al., 2004; Sekinger et al., 2005). Crystal structures show that each nucleosome contains 147 base-pairs (bp) of DNA tightly wrapped around an octamer of H2A, H2B, H3 and H4 histone proteins (Harp et al., 2000). The linker DNA between two neighboring nucleosomes is 70 bp in higher eukaryotes (Williams et al., 1986). A high resolution map of nucleosome has been generated for human promoters by subjecting chromatin to
micrococcal nuclease (MNase) and detecting the undigested DNA with tiling arrays (Ozsolak et al., 2007). Recently Zhao and colleagues generated a genome-wide nucleosome map using MNase digestion followed by deep sequencing (MNase-Seq) (Schones et al., 2008).

TFs can influence nucleosome positioning by competing with nucleosomes for access to DNA. Recent study shows that nucleosomes occupy TF sites and must be repositioned to enable TFs to access their sites (Adkins and Tyler, 2006). It appears that TFs recruit chromatin modifier which alter the chromatin structure of promoters a way that exposes TF binding sequences (Lemon and Tjian, 2000). Previously, Lomvardas S. et. al. showed that a nucleosome obstructing transcription from the IFN-beta promoter slides in vivo in response to virus infection, thus exposing the previously masked TATA box and the initiation site, a requirement for transcriptional activation. (Lomvardas and Thanos, 2001). Recently, He H.H. et. al., showed androgen treatment dismisses a central nucleosome present at androgen receptor binding sites that is flanked by a pair of marked nucleosomes (He et al., 2010).

To test the interplay between chromatin (nucleosomes) and transcription factor binding, and the notion that chromatin guides regulatory binding to induce a combination of events that ultimately produce complex functional changes, we selected the biological complexity in progression of lung cancer. Following selection of a regulatory factor that was expected to impact progression of metastases in lungs from a large scale computational screen, we identified in vivo binding sites of the factor, nucleosome positions and transcriptome profiles for both metastatic and non-metastatic lung cancer cells (Figure 4.1). Results from a combined analysis of these data were compared with
expression profiles obtained from advanced/early lung cancer patients. Finally, we created a highly metastatic modified cell line to validate the findings.

Figure 4.1 Schema used to decipher interplay between nucleosome and transcription factor binding sites in lung cancer: Following selection of a regulatory factor that was expected to impact progression of metastases in lungs from a large scale computational screen, we identified in vivo binding sites of the factor, nucleosome positions and transcriptome profiles for both metastatic and non-metastatic lung cancer cells.

4.2 Results

4.2.1 Transcription network for NME2

Transcriptome profiles of 93 clinically certified early/advanced lung adenocarcinoma samples were screened for 1722 regulatory factors. We found 199 of these to be differentially expressed between early versus advanced stages of lung tumors (P<0.05; FDR<5%; Figure 4.2). Using ARACNe (Margolin et al., 2006) (to infer transcription factor (TF)-target interactions across 93 transcriptomes, see Methods) 6974 putative transcriptional interactions were detected among the 199 factors. A
Chapter 4: Building and analyzing transcriptional regulatory network centered on the metastases suppressor NME2 in breast and lung cancers

ranking based on number of interactions showed that an anti-metastatic factor, non-metastatic 2 (NME2) was connected to 54 of the 198 other differentially expressed TFs including established master regulators like ATF4, SMAD5 and FOXO1 and their targets (Figures 4.3). NME2 is a member of the NM23 (or NME) family of metastases suppressor genes and its low expression has been studied in advancement of oral and breast carcinoma (Baba et al., 1995; Bhujwalla et al., 1999; Fukuda et al., 1996; Miyazaki et al., 1999). In line we found NME2 was down-regulated in advanced lung carcinoma patients. Next, we made a modified cell line using lung adenocarcinoma A549 cells by depleting NME2 levels, transcriptome profile obtained from these cells was >70% concordant with the advanced lung carcinoma profiles supporting anti-metastatic function of NME2 (Figure 4.4). Based on this, A549 cells and MDA-MB-231 cells were used for further molecular studies.

Figure 4.2 Transcriptome profiles of 93 clinically certified early/advanced lung adenocarcinoma samples were screened for 1722 regulatory factors. Heat map (upper panel) and GSEA (lower panel) representation for 199 differentially expressed TFs between early versus advance cancer stage.
Chapter 4: Building and analyzing transcriptional regulatory network centered on the metastases suppressor NME2 in breast and lung cancers

4.2.1.1 ChIP-seq for NME2

From above observations it was likely that NME2 attenuates metastatic progression of lung cancers, therefore we first looked at the effect of NME2 expression on A549 cells. Analysis of >47,000 transcripts following NME2 expression revealed 781 genes as up regulated and 898 genes with decreased expression compared to A549 cells before NME2 expression ($P<0.005$). In order to further check whether cell-based gene expression induced by NME2 correlated with clinical samples we compared the expression pattern obtained from cells with transcriptomes determined from two independent lung cancer clinical studies ($n=93$ and $n=60$) grouped as advanced (III and IV) versus early stage (I and II).
Figure 4.4 NME2 depletion results in gene expression changes characteristic of tumors with metastatic proclivity. Expression change after depletion of NME2 in A549 cells relates closely to advanced lung cancers; heat map of normalized expression profile of 101 genes (126 probes) across 93 clinical datasets grouped stage-wise (P<0.01).

As expected, transcriptome profile following increased NME2 expression in A549 cells was positively correlated to that obtained from early stage clinical samples (P < 0.05) and not the advanced stage (for analysis we subtracted the profile of early from advanced samples, and then compared this with the changes on NME2 expression (Figure 4.5). Together, these observations indicated NME2 as a factor that could negatively influence progression of lung adenocarcinoma and also supported use of A549 cells as a cell-based model experiments.

Next, we used chromatin immunoprecipitation followed by de-novo parallel sequencing (ChIP-seq) to map NME2 occupancy in A549 cells. As NME2 levels appeared to be associated with early stages of lung adenocarcinoma (Figure 4.5), it
was possible that NME2 occupancy that was specific to cells with relatively increased NME2 would be important for reduced metastasis.

**Figure 4.5 Transcriptome profile following increased NME2 expression in A549 cells was positively correlated to that obtained from early stage clinical samples**

GSEA analyses of two gene sets representing NME2-repressed (left) and NME2-activated (right) genes in lung cancer samples grouped into early versus advance stage. NES, normalized enrichment score; p = nominal p value; FDR = false discovery rate.

To test this ChIP-seq was performed in replicate for both conditions, before and after expressing NME2. Out of roughly 7 million sequence reads, >80% uniquely aligned to the reference human genome in each case (Figure 4.6A). We identified the ChIP-seq reads that were enriched for NME2 relative to IgG for construction of peaks (see Methods) and found 2005 and 11017 peaks before and after NME2 expression, respectively (Figure 4.6B); chromosome-wise distribution of the peaks was largely similar, though enriched in case of NME2-expression as expected (Figure 4.6C). Introns and exons displayed contrasting features in terms of peak numbers with more than one-third (33%) and hundredth (1%) of ChIP-seq peaks localized to introns and exons respectively. Largely similar distribution was found in peaks before and after NME2 expression (Figure 4.6D).
Chapter 4: Building and analyzing transcriptional regulatory network centered on the metastases suppressor NME2 in breast and lung cancers

Figure 4.6 Genome-wide mapping of NME2 in A549 cells using ChIP-sequencing
(A) ChIP-Seq reads for NME2 were mapped on human reference genome and uniquely mapped reads were retained for further analysis; percentage of uniquely aligned reads was >80% in most cases. (B) Representative examples showing distribution of reads on chromosome 21. Upper panel represents reads obtained from cells before expressing NME2 (having endogenous NME2 levels), reads from ChIP-seq performed with cells after expressing NME2 are shown in lower panel. (C-D) Distribution of NME2 binding sites before and after expression across: different human chromosomes (C); genomic locations TSS (transcription start site), within gene and downstream of TES (transcription end site) (D).

A 12-mer consensus motif identified using Gibbs motif sampler (Thompson et al., 2007) was present in >70% of the ChIP-seq peaks (Figure 4.7A). We noted that the motif found from ChIP-seq was similar to the one reported earlier from analysis of a single promoter (Postel et al., 2000). As expected in more than 50% of the cases, the NME2 motif occurred within 50 bp of the centre of peak (Figure 4.7B). Furthermore, using position weight matrices for transcription factor binding sites (TFBS) reported in TRANSFAC database we found six transcription factors: FAC1, PAX4, PAX8, CDP CR3 and ZF5 to be significantly enriched within NME2 binding regions (Figure 4.7C).
Chapter 4: Building and analyzing transcriptional regulatory network centered on the metastases suppressor NME2 in breast and lung cancers

Figure 4.7 De-novo Motif finding and Transcription factor binding site enrichment in before and after NME2 expressed peaks A) A 12-mer consensus motif identified using Gibbs sampler B) Motif distribution in NME2 peaks C) Transcription factor binding site enriched in before (upper panel) and after (lower panel) NME2 expressed peaks.

We found 411 and 2151 peaks, before and after NME2 expression respectively, in regions near transcription start sites (+/- 10 kb); the proportion of sites within 1 kb of the start site were found to be enriched relative to other regions on NME2 expression (Figure 4.8A). Based on the peaks found within 10 kb of transcription start sites it was possible that NME2 regulates expression of 396 and 1927 genes in A549 and NME2-expressed A549 cells, respectively (Figure 4.8B). Of these, 1834 genes were unique to the NME2-expressed condition. Pathway enrichment analysis of these genes revealed MAPK signaling, Wnt signaling, focal adhesion as enriched pathways which were also implicated in cancer metastasis (Figure 4.8B). Additionally, we checked the expression pattern of these genes obtained in cellular A549 model within lung cancer clinical samples (n=67) grouped as early (stage I) and advanced (stage III).
Figure 4.8 **Genes bound by NME2 in ‘NME2 expressed’ condition control multiple aspects of cancer cell function**  
A) Distribution of NME2 binding sites before and after expression across ± 10 kb of TSS.  
B) NME2 binding sites obtained in gene regulatory regions (+/- 10 kb) before and after NME2-expression in A549 cells. KEGG pathway analysis and expression correlation with lung cancer samples for genes bound by NME2 in expressed condition only.  
C) Gene expression signature of 450 genes differentially expressed on NME2 expression in A549 cells is significantly correlated with expression changes found in early lung cancers relative to advanced stages in two independent clinical studies. Enriched NME2 binding was found within 10 kb of TSS in all the 450 genes in cells with high levels of NME2 supporting role of NME2 occupancy in regulation of these genes.

We found that 1373 genes (2690 probes) showed a pattern of differential expression in lung cancer clinical samples early versus advance stage (Figure 4.8B).

Next we asked how many of these genes were differentially expressed on expressing NME2 in cells. On comparing with the transcriptome profiles, we found 450 genes that had NME2 occupancy only under NME2-expressed cells and were also differentially up or down-regulated in NME2-expressing A549 cells (relative to cells where NME2 levels where not enriched). Together this suggested that NME2 may
directly induce/repress 450 genes in cells when present in relatively higher levels. Following this we sought to ask whether expression of the 450 genes could distinguish between early versus late stage lung cancers obtained from clinical samples. For this analysis we used two different clinical studies and first constructed the gene signature representing the early stage relative to late stages in each study. Figure 4.8C compares the ‘early stage-specific’ signature for 450 genes uniquely expressed in NME2-expressing A549 cells – we found >70% concordance (p<0.05) supporting that NME2 induction in cells largely represented the expression changes observed in early stages of lung cancer.

As noted earlier all the 450 genes also had NME2 occupancy within 10 kb of their transcription start sites, this is shown in the right frame of Figure 4.8C to illustrate NME2-occupancy and its relationship with gene expression changes. Together, these results suggest that NME2 binding and consequent expression of genes is likely to attenuate lung cancer progression.

Next to examine the role of NME2 as ‘master regulator’ of cancer progression in other cancers we selected breast cancer as a case study. Moreover, few earlier studies implicate NME2 in breast metastasis suppression for example, a study by McDermott et al where they analyzed MDA-MB-231 cells following over-expression of NME2 and showed suppression in motility, invasiveness, and anchorage independent growth (McDermott et al., 2008). It is noteworthy in this context that transcription regulation by NME2 was never studied in breast cancer metastasis. Therefore, we built a binding site repository of NME2 in MDA-MB-231 breast cancer cells following transient over-expression of NME2.

We used ChIP-seq to map NME2 occupancy in MDA-MB-231 cells. ChIP-seq was performed in replicate, roughly 3 million sequence reads were generated out of that
>80% uniquely aligned to the reference human genome. We identified the ChIP-seq reads that were enriched for NME2 relative to IgG for construction of peaks (see Methods) and found 2387 peaks (represented on 224 genes) for NME2. Interestingly, gene ontology analyses of NME2 bound genes from MDA-MB-231 cells presents similar biological process and cellular pathway as observed in A549. Biological processes enriched were cell adhesion, cell junction and cadherin. While TGF-beta, Wnt signaling, focal adhesion pathways were enriched in pathway analysis. These process and pathways were implicated in cancer metastasis, therefore, these findings supports the possibility that NME2 performs its metastasis suppressor function in lung and breast cancer by controlling the above processes, and pathways.

4.2.2 Interplay between chromatin and transcription factor binding in progression of lung cancer

Next, we sought to understand the altered state of chromatin particularly in regulatory regions as a result of NME2 expression and how this influenced NME2 occupancy and consequent gene expression changes. With this notion, first we determined nucleosome positions in A549 cells before and after expressing NME2. Mono-nucleosomes were isolated by MNase digestion and in order to focus on regulatory interactions we mapped the nucleosome positions to putative promoter regions (-7.5 kb to 2.5 kb of transcription start sites) using tiled microarrays (Figure 4.9A representative figure of KCNC2 promoter). We found 157634 and 162570 nucleosomes in low- and high-NME2 cells respectively. As expected, promoter-nucleosome occupancy resulted in largely decreased expression of corresponding genes in respective cells ($R^2 = -0.73; P < 0.05$ before NME2 expression and $R^2 = -0.80; P < 0.05$ after NME2 expression; Figure 4.9B). In A549 cell with low-
Chapter 4: Building and analyzing transcriptional regulatory network centered on the metastases suppressor NME2 in breast and lung cancers

NME2, distribution of nucleosomes across promoter for many genes showed phasing which is consistent with pattern observed by previous studies for nucleosome distribution in human and yeast (ref) (Figure 4.9C). Next, to check relationship between nucleosome distribution and expression state of gene. We selected expressed (ON) and unexpressed genes (OFF) as top and bottom 10 percent of genes from the gene list sorted by expression level in low-NME2 A549 cells. Interestingly, we found Nucleosome Free Region (NFR) in OFF genes was less open compare to NFR in ON genes (Figure 4.9D). Additionally, shifting was also observed in + 2, + 3 and + 4 nucleosome between OFF versus ON genes (Figure 4.9E).

Figure 4.9 Nucleosome positions before and after expressing NME2 in A549 cells
A) UCSC browser representation of nucleosome positions with corresponding probe intensity on chromosome 12. Detailed view of the three loci showing the probe intensity and identified nucleosomes. B) Promoter-nucleosome occupancy resulted in largely decreased expression of corresponding genes in respective cells before NME2 expression (left panel) and after NME2 expression (right panel). C) For each indicated nucleosome type (-1, +1, +2, +3) we plot the distribution in cells before NME2 expression. D) Different distribution of Nucleosome Free Region (NFR) length among genes expressed and unexpressed before NME2 expression. E) Shifting was observed in +1, +2, +3 and + 4 nucleosome between OFF versus ON genes.
About 16% (18024) of nucleosomes were repositioned and as expected a large portion of genes containing these repositioned nucleosome were also differentially expressed (Figures 4.10A). Maximum repositioning of nucleosomes after NME2 expression was observed near TSS (Figure 4.10A upper panel). Gene ontology analysis for the genes that have nucleosome repositioning after NME2 expression revealed MAPK signaling, Wnt signaling, Focal adhesion as top enriched pathways. (Figure 4.10B).

Figure 4.10 Genes with repositioned nucleosomes after NME2 expression was also differentially expressed Distribution of repositioned nucleosomes with respect to TSS (Left panel), maximum repositioning of nucleosomes after NME2 expression was observed near TSS (Figure 4.10A upper panel). Expression profile of genes with repositioned nucleosome in before and after NME2 expressed A549 (Right panel) B) Gene ontology analysis for the genes that have nucleosome repositioning after NME2 expression.
4.2.2.1 NME2 target sites required for metastasis suppression are nucleosome-occupied in aggressive cancer cells

It is commonly understood that a substantial proportion of transcriptional interactions, typically association of regulatory factors to DNA is largely determined by chromatin, in other words, availability of transcription factor binding sites may be contingent upon positioned nucleosomes. To directly test this, we resorted to a combined analysis of ChIP-seq occupancy as well as nucleosome positions determined from A549 cells before and after NME2 expression, and posed a specific question: whether NME2 target sites were relatively unavailable due to positioned nucleosomes in metastatic cells (with relatively low NME2 expression) (Figure 4.11A). First, a higher frequency of positioned nucleosomes was noted in the vicinity of NME2 sites in cells that had relatively low NME2 expression. Out of 3956 NME2 target sites (-7.5 to + 2.5 kb of transcription start site) found in NME2-expressing cells, 1257 (~31%) present on 1119 putative promoters, were found to either overlap or were within 300 bases of a nucleosome in cells with relatively low NME2 (Fig 4.11B). This indicated that many of the NME2 binding sites were excluded by positioned nucleosomes in A549 cells. Conversely, nucleosome mapping in NME2-expressed conditions confirmed that roughly 70% out of 1257 NME2 binding sites were nucleosome-free, which involved repositioning of 870 nucleosomes (Figure 4.11C). Consistent with this, 1119 genes with repositioned nucleosomes among the two conditions showed change in gene expression as expected with concomitant increase in NME2 level (Figure 4.11D).
Figure 4.11 NME2 target sites were relatively unavailable due to positioned nucleosomes in metastatic cells (before NME2 expression) A) Schematic representation of possible relationship between nucleosome positions before and after NME2 expression, nucleosome positioning may result in reduced accessibility of NME2 on binding sites in metastatic condition. B) Distribution of distance between nucleosome position and NME2 binding site before NME2 expression shown by bar graph and relative occurrence of nucleosome (before vs. after NME2 expression) across NME2 binding site shown by line graph. C) Comparison of distance to nearest nucleosome from NME2 binding site in cells before and after NME2 expression. D) Altered nucleosome positions on expressing NME2 and corresponding gene expression change (right panel).

For validation we selected six genes and compared the nucleosome occupancy and NME2 binding using qPCR. As expected, we detected low nucleosome occupancy and high NME2 occupancy in cells with NME2 expression as compared to that in cells before NME2 was expressed (Figure 4e; Primer information provided as Appendix 5).
Figure 4.12 Quantitative real time PCR validation for NME2 binding sites and nucleosome positioning: Quantitative real time PCR analysis for NME2 binding site and nucleosome positioning (right panel) for 6 genes before versus after NME2 expressing cells; UCSC browser representation for these genes (right panel). Relative fold change is shown on x-axis. Experiments were performed in triplicate; error bars are for standard deviation (* and ** represents $P < 0.05$ and $P < 0.01$ respectively).
4.2.2.2 Cells with increased metastatic potential confirm nucleosome occupancy of NME2 target sites

The above findings suggested that in about 70% of cases nucleosome repositioning between the two states (after versus before NME2 expression) exposes binding sites for NME2 occupancy and this correlates with the gene expression changes that negatively impact cancer progression. Therefore we reasoned that in addition to NME2-expressed cells, A549 cells ablated for NME2 would provide a reasonable model to test nucleosome positioning and NME2 occupancy in a contrasting condition. Moreover, in keeping with fact that our initial analysis revealed NME2 as a metastasis suppressor, it was also possible that NME2-depleted cells would have higher metastatic potential relative to A549 cells alone. Based on these, an NME2-depleted A549 line was made. We first checked and results confirmed enhanced metastatic potential in vivo using immunocompromised mice where enhanced colonization of the lung was noted in NME2-depleted A549 cells relative to A549 cells. Thereafter, we determined nucleosome positions in the NME2-depleted cells. Interestingly, all the 1257 NME2 binding sites that were found to be either occupied or in vicinity of nucleosomes in A549 cells were also found to be nucleosome-occupied in the NME2-depleted condition (Figure 4.13A). To illustrate this point Figure 4.13B shows two promoters where NME2 binding was excluded (due to nucleosome positioning) in A549 and NME2-depleted A549 cells but were available following NME2 expression.

Following this we asked whether expression of genes where NME2 target sites were nucleosome-occupied in the NME2-depleted cells correlated with clinical cases of advanced lung carcinoma, as would be expected from cells that seeded higher metastasis. We found 148 genes where the NME2 target site was not available due to
nucleosome occupancy in NME2-depleted cells. As expected, expression of the 148 genes had significant negative correlation with the early stage transcriptome profiles obtained from 93 patients (Figure 4.14 right panel). In contrast, these 148 genes had NME2 occupancy (and were nucleosome-free) in NME2-expressing cells. Moreover, we also found that expression of the same 148 genes when examined in the profile obtained from NME2-expressing non metastatic cells, were found to positively correlate with the early stage profiles (Figure 4.14 left panel).

Figure 4.13 Cells with increased metastatic potential confirm nucleosome occupancy of NME2 target sites A) Nucleosome positions are identical before and after silencing NME2 in A549 cells – nucleosome positions from nearest NME2 binding site are shown. B) Representative examples of two promoters where NME2 binding was excluded (due to nucleosome positioning) in cells with relatively low or silenced NME2 but were available on expressing NME2.
Together, this suggested that NME2-occupancy following nucleosome repositioning could significantly influence the metastatic/non-metastatic state of the A549 cells.

Figure 4.14 Expression of the 148 genes in transcriptome profiles obtained from 93 patients Genetically NME2 expressing cells match with early stage cancer cells. Whereas, NME2 silenced cells match with advance stage lung samples. Heat map of corresponding gene expression (upper panel) and correlation graph (lower panel).

4.3 Materials and Methods

4.3.1 Cell culture

A549 cells were procured from ATCC and MDAMB231 cells were obtained from NCCS, Pune. Cells were cultured in Dulbecco’s Modified Eagle’s Medium (DMEM) (Sigma, USA) supplemented with 10% FBS and 5% CO₂ at 37°C. A549 cells with stably depleted NM23H2 were cultured in DMEM containing puromycin (1mg/ml) to maintain selection pressure.
4.3.2 Sample preparation for NME2 ChIP-seq

ChIP assays were performed as per the protocol provided by Upstate Biotechnology with modifications as suggested in Fast ChIP protocol. Nearly $1 \times 10^5$ cells were plated to achieve a confluency of 80%. They were then transfected with pcDNA3-NME2-Myc tag construct and its control pcDNA3-Myc tag construct. ChIP was performed after 48hrs of transfection using antibody against Myc tag (From Sigma, monoclonal Anti-c-Myc-Cy3, Clone 9E10).

Briefly, cells were fixed with 1% formaldehyde for 10 min and lysed. Chromatin was sheared to an average size of 500 bp using a Misonix 3000 sonicator. Twenty-five percent of lysate was used to isolate input chromatin using phenol–chloroform and ethanol precipitation. Lysate was pre-cleared using protein-A sepharose beads, and ChIP was performed using 2 mg of the respective antibody incubated overnight at 4°C. Immune complexes were collected using herring sperm DNA-saturated protein-A Sepharose and washed extensively. Chelex-100 MYC-tagged NME2 was quantified, and 10 ng from each sample was taken for end repair using Illumina sample preparation kit. Samples were purified using PCR purification kit (Qiagen, Germany). Thereafter ‘A’ base was added to the samples’ 3’end using Illumina sample prepration kit. Then flow-cell primer specific adapters were ligated to the ChIP DNA fragments and samples were further purified by MinElute columns. Size selection was done after adapter ligation using 2% agarose gel. Gel extraction columns (Qiagen) were used to purify DNA fragments ranging between 150 and 350 bases. These eluted samples were then purified using MinElute columns and were then amplified for 18 cycles to enrich adapter-ligated DNA fragments. After PCR purification and elution, the DNA was quantified using Picogreen.
method and then 3.5 picomoles of each sample was sequenced (36 cycles) on GAII (Illumina, USA) according to manufacturer’s protocol.

### 4.3.3 High throughput sequencing using Illumina Genome Analyzer II

NM23 H2-ChIPed DNA from lung adenocarcinoma A549 cells and breast cancer MDA-MB-231 cells over expressing MYC-tagged NM23 H2 was processed to add adaptors, PCR amplified, and used for cluster generation and sequencing using Illumina Genome Analyzer II (GAII) sequencer as per manufacturer’s instructions. Reads were aligned (mapped) to the unmasked human reference genome (NCBI v36, hg18) using the Eland application (Illumina). Only uniquely mapped reads were retained. The sequences were then mapped to the reference human genome (NCBI Build 36, UCSC hg18). Only those that aligned uniquely to the reference genome were further considered for peak generation. Almost half of these reads mapped to single sites in the human genome. Sequence reads that mapped to multiple sites in the genome were removed from subsequent analysis. The location of each remaining unique sequence read in the genome was recorded. To accommodate variations relative to the reference genome, up to two mismatches were allowed.

### 4.3.4 Peak generation following ChIP-seq

The resulting sequence read distribution was processed with ChIP-Seq peak locator utility, CisGenome (Ji et al., 2008). CisGenome uses a conditional binomial model to identify regions in which the ChIP reads are significantly enriched relative to the control reads. A false discovery rate (FDR) of 10% was considered while predicting
NME2 target regions. In order to filter out low-quality sites we applied two post-processing options boundary refinement and single-strand filtering.

### 4.3.5 Mononucleosome preparation

A549 cells were grown in T-75 flasks till 80% confluency in DMEM. These were then trypsinized and pelleted into 15ml falcons and washed with ice cold 1X PBS. Five volumes, of the packed volume, of 1X hypotonic buffer were added and cells were homogenised and incubated on ice for 15min. IGEPAL was added to final concentration of 0.6%, lysate was than vortexed and centrifuged at 12000rpm. Pellet (nuclei) was resuspended in MNase digestion buffer. Nuclei suspension was incubated on ice. An 2µl of aliquot nuclei suspension was incubated with 498µl OD solution at room temperature for 20min, after which its OD was measured.1unit of MNase/OD was added to the suspension and incubated at 37ºC for 30min. 1/3 volume of stop buffer was added to end the digestion. Proteinase K was added in a concentration of 1mg/ml and kept in water bath overnight. Phenol chloroform purification followed by ethanol precipitation was done to pellet down the mononucleosomal DNA. This was then washed with 70% alcohol and dissolved in water. DNase treated cells were taken as control.

### 4.3.6 Sample preparation for promoter tiling array

Reaction mixture containing 3µg of DNA, 50mM Tris-Cl, 5mM MgCl₂ and 0.3µg/µl of random hexamers was incubated at 95ºC for 5min and chilled to 4ºC. 5µl dNTP (1.2mM dGTP, dCTP, dATP and 0.25mM dUTP) and 50Units of Klenow was incubated first at 22ºC for 10min, 37ºC for 30min and finally 95ºC for 5min and then
chilled as above. 50Units of Klenow was added again and mix was first kept at 22°C for 10min and then 37°C for 30min. Reaction was stopped by phenol chloroform followed by purification with Qiagen columns. Purified product was fragmented with UDG and APEI of whole transcript Terminal Labeling Kit by incubating at 37°C for 90min followed by 93°C for 2min and then chilling it to 4°C. Labeling was done by incubating 45µl of fragmented DNA with labeling reagent and TdT (terminal deoxynucleotidyl transferase), again provided in whole transcript Terminal Labelling Kit. The control DNase treated cells were also subjected to labeling reactions. These were then hybridized to Human Promoter Tiling array 1.0R (Affymetrix).

4.3.7 ChIP, nucleosome positioning validation PCRs

For performing chromatin immunoprecipitation (ChIP), A549 cells and MDA-MB-231 cells were transfected with PcDNA-NME2-MYC tag construct using Lipofectamine 2000 and 48h post transfection cells were harvested and chromatin immunoprecipitation was performed as described above.

For, validating individual nucleosome positions, quantitative real–time PCRs were performed using ChIPped DNA on ABI’s 7500 fast platform. Equal amount of ChIP DNA was taken from the test (from NME2 over expressing cells) and control (cells having endogenous NME2) samples. The results were analysed using comparative ct method.

4.3.8 Primer design

Primers were designed against the obtained peak region using ABI’s Primer Express software (Primer information is provided as Appendix 5).
4.3.9 Gene expression microarrays

For expression profiling of over-expressed NME2, A549 cells were transfected with PcDNA-NME2-MYC tag construct (‘high NME2’) and PcDNA-MYC tag construct (‘low NME2’). RNA was isolated from the cells 48h after transfection using trizol method (InvitroGen) as per manufacturer’s protocol. Total RNA was processed to hybridize to Illumina HumanHT-12 v4 Expression BeadChip as per manufacturer’s instructions. Three biological replicates were averaged and data was analyzed using BeadStudio (P <0.005 of fold change).

For expression profiling of NME2 depleted, after 48 hours of transfection of siRNA duplexes, total RNA from A549 cells was isolated using TRI reagent (sigma) and cDNA synthesized using cDNA archive kit (Applied Biosystems, USA). For transcriptome analysis, total RNA was processed to hybridize to Affymetrix HG-U133 plus 2.0 GeneChip as per manufacturer’s instructions. Three biological replicates were averaged and significance analysis performed using GCOS (at least 1.2 fold, P <0.005 of fold change).

4.3.10 Meta-analysis of gene expression related to clinical lung tumor samples

Transcriptome data of lung tumors from Expression Project for Oncology (expO) database (GSE2109) were used. expO is hosted by International Genomics Consortium (IGC, USA, www.intgen.org) and provides gene expression datasets of clinically annotated sets of tumor samples. We grouped tumors either on the basis of clinical stage (I to IV; stage I and II termed as early, and stage III and IV termed as
advanced). Gene expressions across datasets were normalized using Z score transformation (Cheadle et al., 2003).

The significance of changes, $v$ (defined as relative fold change in this study) between expression pattern of clinical samples and NME2-depleted A549 cells was calculated using

$$v = s (1 - p)$$

Where ‘p’ represented a p-value calculated by performing a student’s t-test between the advanced and early stages of cancer samples. ‘s’ denoted the sign of the difference between the average values in these two sets considered above. ‘v’ indicated the extent to which a gene was up or down-regulated in the advanced cancer stage with maximal and minimal values of 1 and -1 respectively. ‘v’ for genes from the dataset of NME2-depleted A549 cells or NME2-overexpressed A549 was calculated similarly. Correlation coefficient was calculated between significance of changes in clinical samples and NM23 H2-depleted A549 cells.

### 4.3.11 Discovery of NME2-motif and transcription factor enrichment

Following peak generation by CisGenome, a 12-mer motif for NME2 binding was identified by Gibbs motif sampler at its default parameters. All peaks with binding enrichment of 2 fold or more were considered for motif discovery. A representation for the motif was made using weblogo utility (Crooks et al., 2004). The peak sequences were given as input into the Match tool to scan for enriched PWMs listed in TRANSFAC professional.
4.4 Discussion

4.4.1 NME2 as a key metastasis suppressor gene

NME2 is a member of the first discovered family of metastasis suppressor genes, NME. Reduced expression of NME2 in patient tumor samples has been connected to aggressive progression of prostate, ovarian, pituitary, head and neck squamous cancers (Thakur et al., 2011). Further, in vivo experiments using breast, oral squamous, colon carcinoma and melanoma cells have shown a definitive role of NME2 in metastasis suppression (Thakur et al., 2011). While these observations supported the metastasis suppressor activity of NME2; it remained unclear which molecular function of NME2 could mediate metastasis suppression. NME2 functions as a nucleoside diphosphate (NDK) as well as histidine protein kinase; however, the relationship between enzymatic activity and metastasis suppression is not well understood. Notably, of the multiple paralogs of NME2 in the human genome which are expressed ubiquitously, many retain nucleoside kinase activity (Lacombe et al., 2000). The biochemical redundancy therefore is likely to compensate for loss of enzymatic activity if NME2 kinase function were compromised in cancer. It is thus plausible that non-enzymatic regulatory roles of NME2 may be important in the context of metastasis suppression.

4.4.2 The gene expression regulatory functions of NME2

One of the earliest reports on the function of NME2 identified its role as a transcription factor for oncogene c-MYC (Postel et al., 1993). Since then several reports have supported its function as a regulator of genes involved in cancer progression supporting a candidature of regulatory role of NME2 in control of
metastasis. However, a couple of reports have raised doubts on NME2 in being a conventional transcription factor and two specific concerns emanated (Steeg et al., 2011). First, NME2 appeared to lack a modular protein structure typical of a transcription factor. Experiments with NM23 H2 fusion protein did not reveal functionally distinct DNA binding and trans-activation domains in yeast based domain swap experiments. Second, NME2 was observed to nonspecifically bind single stranded oligonucleotides raising questions on target specificity in isolated reports. In spite of the criticisms, a clear understanding of gene regulatory role of NME2 has slowly emerged. In brief, the association of NME2 with c-MYC promoter was extended to several cell types, and the nature of NME2 interaction with c-MYC promoter thoroughly investigated (Thakur et al., 2011). These analyses revealed NME2 association with a guanine-rich motif in c-MYC promoter. Importantly, NME2 was also observed at guanine-rich promoters of PDGF-A, cluster of differentiation 11 b (cd11b), myeloperoxidase (MPO), chemokine receptor 5 (CCR5) genes. Further, the biochemical analysis of amino acids of NME2 showed that DNA binding and kinase activity of NME2 are mediated by different amino acids, indicating a modularity of structure (Postel et al., 2002).

4.4.3 Novel insights into NME2 association with DNA: Guanine-rich motifs, and nucleosomes

The results of this chapter provided novel insights into the gene regulatory roles of NME2 in the context of lung cancer progression. While previous studies were limited to few gene promoters, this study used an unbiased genomic strategy of ChIP coupled with high throughput DNA sequencing (ChIP-seq) to explore the genome wide
association of NME2. The genomic analyses revealed a large number of sites in DNA where NME2 associated with target chromatin. Most of these sites have not been reported before. Interestingly, analysis for consensus motif among target sites of NME2 revealed a 12-mer guanine rich motif. The motif was present in majority of NME2 target sites indicating NME2 association with DNA could be mediated through the G-rich motif. The G-rich motif discovered here bore surprising similarity with a previously reported motif for NME2 binding with DNA (Postel et al., 2000).

Apart from extending a catalogue of NME2 association sites in DNA, and uncovering the presence of a G-rich motif in target promoters, this study provided a new approach to understand how NME2 can associate with DNA. The binding of transcription factor to its cognate DNA sequence depends on the accessibility of the target sequence in the chromatin. As eukaryotic DNA is wrapped in nucleosomes, accessibility of the target sequence becomes a key step in transcriptional regulation. Thus, the target site must be made accessible so that transcription factor binding takes place. In view of this conceptual background, this study used tiling microarrays to explore the nucleosome occupancy of >20,000 human gene promoters. A combinatorial analysis of NME2 binding sites in genome, and nucleosome occupancy map of promoters yielded insights into chromatin environment of NME2 target sites. It became clear that nucleosome occupancy of NME2 target sites varied with a change in the level of NME2 in cancer cells. In a non-metastatic cell line with high level of NME2, most of the target sites with NME2 binding peak were nucleosomes-free. In comparison, in a metastatic cell line, with low level of NME2, only a fraction of total NME2 sites were occupied and the rest had well positioned nucleosomes. Taken together, it could be inferred that NME2 binding changed the nucleosome occupancy at the target promoter. This observation is in line with a recent
report that binding of a transcription factor displaces the nucleosome occupying its cognate consensus sequence of DNA bases (He et al., 2010). Taken together, the results described here suggested a change in nucleosome occupancy of target promoters by NME2 as has also been reported for well established transcription factors.

### 4.4.4 Metastasis suppression by NME2: gene signatures from cellular and clinical samples

Interestingly, this study found that many NME2 binding sites on promoters of genes with important roles in metastasis suppression were obstructed by nucleosome in metastatic cells. Analysis of genes regulated by NME2 showed that several signaling pathways such as focal adhesion, ECM-receptor interaction, MAP kinase and wnt were enriched. Interestingly, wnt signaling has recently been shown to be involved in lung cancer metastasis to brain (Nguyen et al., 2009). Further, the change in expression of NME2 altered the transcriptional program of A549 cells. It was found that the gene expression program of A549 cells depleted for NME2 related closely to gene expression signature of advanced and metastatic stages of lung cancer. In tune, transcriptome of A549 cells with high level of NME2 resembled closely to early stages of lung cancer. Taken together, these observations indicated the clinical relevance of altered expression of NME2 in metastasis progression.

### 4.4.5 Limitations of the present approach

There were two aspects of NME2-target promoter interactions which were not probed; did NME2 alter the nucleosome occupancy of non-promoter regions of genome, and how did NME2 binding to target sites lead to nucleosome depletion. As
a promoter tiling microarray was used to map the nucleosome occupancy, the status of nucleosomes outside the promoter regions could not be studied. Thus, it could not be ascertained whether NME2 dependent effects on nucleosome occupancy extended to non-promoter regions. It could be interesting to understand if the changes in nucleosome occupancy could be a global consequence of alterations in expression of NME2. Further, the molecular mechanisms of NME2 mediated changes in nucleosome occupancy were not studied. Previous results obtained in this laboratory showed that nucleosomes are depleted in guanine rich genomic regions that can potentially adopt a secondary structure: G-quadruplex (Halder et al., 2009). Notably, results obtained in this chapter suggested occurrence of a guanine rich motif in NME2 target promoters. Thus, it could be speculated that NME2 prefers binding to guanine rich sequences in DNA which intrinsically exclude nucleosomes. However, this conjecture needs to be verified experimentally.

4.4.6 Conclusions

This study used a combination of genomic approaches to explore the relationship between NME2 association with DNA and identified novel targets. It was also clear that NME2 binding to target sites was related to changes in nucleosome occupancy. The sites where NME2 interacted with target DNA were depleted for nucleosomes; this behavior is typical of a transcription factor. Interestingly, majority of NME2 target promoters contained a guanine-rich motif which is similar to a previously reported motif for NME2. Importantly, most of these association sites are novel and when studied in more detail would likely provide provocative insights into the process of metastasis control.