Department of Bioscience and Nutrition
Karolinska Institutet, Stockholm, Sweden

EVOLUTION AND TRANSCRIPTIONAL REGULATION OF KINDLINS

Ammad Khan

Karolinska Institutet

Stockholm 2013

Published by Karolinska Institutet. Printed by Larserics Digital Print AB

ISBN 978-91-7549-095-3

DEDICATED TO BENAZIR BHUTTO AND MALALA YOUSAFZAI,
FOR THEIR COURAGE, BRAVERY AND COMMITMENT

The secret ingredient of our secret ingredient recipe is that there is no secret ingredient. You just have to believe.

Kung fu Panda, 2008

Abstract

Kindlins are relatively newly discovered focal adhesion proteins. The Kindlin family includes three highly conserved proteins: Kindlin-1, Kindlin-2 and Kindlin-3. All three Kindlins have been shown to interact with Integrin and this interaction plays important role in the functional properties of Kindlins. We here in explore for the first time the evolutionary history of these proteins. The phylogeny of the Kindlins suggests a single ancestral Kindlin protein present in the earliest metazoans e.g. sponges. This protein then underwent duplication events in insects and also experienced genome duplication in vertebrates, leading to the Kindlin family. A comparative study of the Kindlin paralogs showed that Kindlin-2 is the slowest evolving protein among the three family members and is evolving under strong purifying selection.

The FERM domain of each Kindlin is bipartite because of the unique insertion of PH domain, dividing the FERM domain into two domains. The FERM domain plays a key role in integrin activation. In one of our study, we tried to trace the evolutionary history of Kindlin FERM domain with respect to the FERM domain of other proteins. We showed that although Kindlin proteins have highly conserved domain among themselves, their FERM domain however is much less conserved when compared with the FERM domain containing proteins of B4.1 superfamily. In addition, we showed that the unique insertion of Pleckstrin homology (PH) like domain in Kindlin FERM domain have important evolutionary and hence functional consequences. We also traced the important residues in Kindlins by ranking them according to their evolutionary significance and discussed about the structure-function relationship of these ranks. We hypothesized that FERM domain originated from a proto-Talin protein in a unicellular or proto-multicellular organism and the advent of multicellularity was accompanied by a burst of FERM domain containing proteins (FDCPs) which supported the complex organization multicellularity requires.

Despite of having strong homology, Kindlins exhibit very contrasting expression profile where Kindlin-2 shows broad spectrum of expression while Kindlin-1 and Kindlin-3 are tissue specific and expressed predominantly in epithelial and hematopoietic system respectively. The expression pattern of Kindlins along with phylogenetic studies supports the subfunctionalization model of gene duplication. In one of our study, we tried to explore the evolutionary changes occurring in the regulatory regions of Kindlin genes. We found that, as with coding region, Kindlin-2 promoter is the most conserved promoter in Kindlin paralogs. We also showed that the conservation profile of Kindlin expression pattern in mammals very much go in line with the conservation of regulatory sequences. In addition, we made use of ENCODE histone modification data and showed that although a correlation is found between Kindlin expression pattern and extent of modification of H3K4me3, the conservation pattern of this epigenetic signature does not match much with the conservation of Kindlin expression. The *in silico* studies we performed on Kindlin promoters also provided us a platform to target the potential transcription binding sites and hence characterize these promoters functionally. In this direction, we studied the role of GLI1, SP1 and SRF in the transcriptional regulation of Kindlin-2.

LIST OF PUBLICATIONS

I. **KHAN, AA**, Janke A, Shimokawa T, Zhang H.

Phylogenetic analysis of kindlins suggests subfunctionalization of an ancestral unduplicated kindlin into three paralogs in vertebrates. Evolutionary Bioinformatics, 2011

2 **KHAN AA***, Raja H.

Evolution of FERM domain containing proteins and emergence of Kindlins (Manuscript)

* Equal contribution

3 **KHAN AA,** Strömblad S, Zhang H, Kumar V, Raja H.

A study of Kindlin family promoter evolution reveals the synergy between promoter evolution, protein evolution and evolution of transcription regulation (Manuscript).

4 Gao J, **KHAN AA**, Shimokawa T, Zhan J Strömblad S , Fang W and Zhang H.

A feedback regulation between Kindlin-2 and GLI1 in prostate cancer cells. FEBS LETT. 2013. (Epub Ahead of print)

5 **KHAN AA**, Shimokawa T, Strömblad S, Zhang H.

Functional characterization of human Kindlin-2 core promoter identifies a key role of SP1 in Kindlin-2 transcriptional regulation. Cell Mol Biol Lett. 2011 Dec;16(4):638-51

Table of contents

List of Abbreviations

B4.1	Band 4.1
BRE	TFIIB-Recognition Element
C.elegans	Caenorhabditis elegans
CGI	CpG islands
ChIP	chromatin immunoprecipitation
DCE	Downstream Core Element
DNMT	DNA methyltransferases
ECM	Extracellular matrix
ENCODE	Encyclopaedia of DNA elements
ERM	Ezrin, Radixin, Moesin
ETA	Evolutioanry trace analysis
FAK1	Focal adhesion Kinase1
FDCPs	FERM domain containing proteins
FERM	Four point one protein, Ezrin, Radixin, Moesin
GEFs	Guanine nucleotide exchange factors
GTPases	Guanosine triphosphatases
JAK3	Janus kinase3
KIND	Kindlin
KRIT	Krev interaction trapped
LAD III	Leukocyte adhesion deficiency III
LCR	Locus control regions
MIG2	Mitogen inducible gene 2
MTE	Motif Ten Element
MYO7	Myosin7
PH	Pleckstrin homology domain
PIC	Preinitiation complex

PIP2	Phosphatidylinositol 4,5-bisphosphate
PIP3	Phosphatidylinositol-3,4,5-triphosphate
PTB	Phosphotyrosine binding
RAP	Receptor-associated protein
RIAM	Rap1–GTP-interacting adapter molecule
TAF	TBP-associated factor
TBP	TATA-box-binding protein
TFBS	Transcription factor binding sites
TSS	Transcription start site

1 Introduction

1.1 Kindlins

The Kindlins are FERM domain containing proteins and represent a class of focal adhesion proteins implicated in integrin activation. There are three evolutionarily conserved members, Kindlin-1 (FERMT1, C20orf42, URP1; chromosome 20p12.3), Kindlin-2 (FERMT2, MIG2, PLEKHC1, UNC112, chromosome 14q22.1) and Kindlin-3 (FERMT3, UNC-112 related protein 2; chromosome 11q13.1) which share considerable sequence and structural similarities. Kindlin-1 shares 62% sequence similarity with Kindlin-2 and 49% with Kindlin-3. Kindlin-2 is the most conserved of all Kindlin proteins and might be the representative of the unduplicated ancient Kindlin protein [1]. The presence of approximately 300 amino acid long FERM (four point one protein, ezrin, radixin, moesin) domain is the structural as well as functional hallmark of Kindlins. A unique feature of Kindlins is that their FERM domain is disrupted due to the insertion of a PH domain in the F2 subdomain dividing it into two parts. The F3 sub domain of the FERM domains of all three Kindlins contains a crucial phosphotyrosine binding (PTB) fold resembling that of Talin. Functionally, Kindlins are primarily structural proteins involved in protein-protein interaction and lack catalytic domains. In addition to Integrin activation they have recently been associated with other functions too i.e. hedgehog signalling, genome stability and recruitment of β catenin into the nucleus [2-4].

1.1.1 Kindlin in diseases

Kindlins have been linked with many disease phenotypes too. For instance, Loss-of-function mutations in Kindlin-1 and Kindlin-3 cause Kindler syndrome and leukocyte adhesion deficiency-III syndrome, respectively. Kindler syndrome was in fact the first human genetic disorder clinically associated with Kindlins. It is caused by mutation of Kindlin-1 and is characterized by skin blistering, severe periodontitis and poililodermia [5, 6]. Kindlin-3 however has got more attention recently due to its involvement in a rare disease of hematopoietic system, leukocyte adhesion deficiency III (LAD-III). This disease features combined dysfunction of the β_1, β_2 and β_3 integrin families on platelets and leukocytes and is characterized by excessive bleeding and abnormal immune responses [7, 8]. Kindlin-2 has not yet been associated with any specific pathological phenotype. Embryonic knockout of Kindlin-2 proved embryonic lethal indicating the indispensability of this gene for the development of an organism [9]. Kindlins have also been implicated in many types of cancers. For instance de-regulation in Kindlin-1 and Kindlin-2 expression has been shown in different types of lung cancers, where both of the proteins are playing contrasting roles in cancer progression [10]. Similarly,

Kindlins have been shown to be implicated in skin cancer [11, 12], prostate cancer [2, 13], breast cancer [3, 14], gastric cancer [15, 16] and oral cancer [17].

1.1.2 Kindlin expression

A very distinct feature of Kindlins is the exhibition of extreme level of differential expression of all the three paralogs of Kindlins. For instance, Kindlin-1 is predominantly expressed in the epidermis and only weakly expressed in the dermis, while Kindlin-3 expression is restricted exclusively to hematopoietic tissues, where it is the dominant form of Kindlins expressed. On the other hand, Kindlin-2 is ubiquitously expressed in most parts of the body [18]. These differential expression patterns may in part be due to the very different promoter structures of these three paralogs where Kindlin-2 promoter is highly CpG rich and is regulated by SP1 transcription factor while the other two paralogs either lack CpG islands (in case of Kindlin-3 promoter) or don't have them in their regions upstream of their transcription start site i.e. in case of Kindlin-1 [19]. This differential expression of Kindlins also explains the distinctive phenotypes that result from the loss of different Kindlins. For instance, the Kindlin-2 homolog in *C. elegans*, UNC-112 is essential for embryonic development [20]. Similarly, the loss of Kindlin-2 in mice results in pre- implantation embryonic lethality and the knockdown of Kindlin-2 in zebrafish reveals a strong relationship between cardiac development and the function of Kindlin-2 [21]. Also, as Kindlin-3 expression is restricted in hematopoietic tissues, mice lacking Kindlin-3 show severe osteoporosis, haemorrhage and defects in the erythrocyte membrane skeleton, and die within one week after birth [22, 23].

1.1.3 Kindlins in Integrin activation

The function of many Integrins depends upon their activation. The Integrin activation is defined as a transformation from a low to a high-affinity state for their extracellular ligands [24]. This transformation is induced by various cellular agonists. The activation of Integrins in-turn leads to outside-in-signalling events that ultimately induce changes in cell shape and function. The cytoplasmic tails of Integrins act as the receptor for ligand binding. One of the most important and well-studied molecule involved in Integrin activation is Talin, a 270 kDA cytoskeletal protein [25-28]. Similar to Kindlins, Talin also consists of a FERM domain consisting of F1, F2 and F3 partitions [29]. It is the F3 subdomain of Talin FERM domain that induces Integrin activation by binding with its NPXY motif of Integrin beta3 cytoplasmic tail [30-33]. However, there is also a substantial evidence from different sort of studies i.e. mutational assays, in vivo experiments and theoretical considerations that the regions of the Integrin beta3 cytoplasmic tail that do not bind with Integrin cytoplasmic tail are also involved in Integrin activation. Numerous studies have been done to find the binding partners

for the [beta] cytoplasmic tail other than Talin [34]. Kindlin-1 and Kindlin-2 were among the important molecules which were found involved in Integrin activation through their binding with cytoplasmic tail of Integrin beta3 and in this way ultimately influence different cellular functions performed by Integrins e.g. cell migration, cell adhesion, cell spreading etc. [35, 36]. Both Kindlin-1 and Kindlin-2 have been found in focal adhesions and have since been considered as important focal adhesion molecules [36]. This binding of Kindlin does not take place at the Talin binding site on the proximal part of cytoplasmic tail of Integrin, but to the NXXY motif located on the distal part of Integrin β3 cytoplasmic tail. Although Kindlin itself have been shown to exhibit minimal ability to activate Integrins, it does synergize this activation when co-expressed with Talin head domain, which leads to the idea that Kindlins in fact are co-activators of Talin as far as Integrin activation is concerned [32, 37-40].

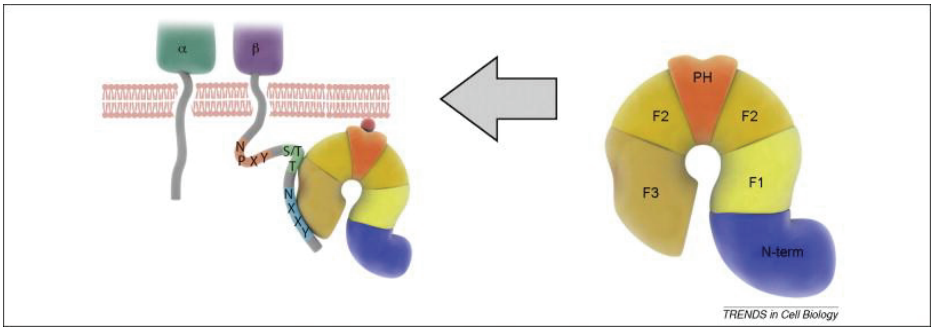

Figure 1 A simple schematic diagram showing binding of Kindlin FERM-F3 domain binding with NXXY motif cytoplasmic tail, an important step towards Integrin activation. (Meves et. al 2009, Trends in cell Biology. Taken with permissions from ScienceDirect)

1.2 Integrins

Integrins are obligate heterodimers containing two distinct chains, called α (alpha) and β (beta) subunits. There are eighteen α and eight β subunits that have been characterized, whereas the Drosophila genome encodes only five α and two β subunits, and *C.elegans* possess genes for two α subunits and one β. Both α and β subunits have small cytoplasmic domains and each can penetrate the plasma membrane. Integrin are the most important molecules which connect the ECM (extracellular matrix) to the actin cytoskeleton and hence provide the architectural framework to determine cell shape on one hand and to react to changes in the microenvironment on the other hand. The molecular organization of ECM around the cells is very much determined by Integrins [41, 42].

1.2.1 Integrin signalling

One of the most remarkable properties of Integrins is that in contrast to other transmembrane receptor proteins which exhibit one directional signalling, they elicit signals bidirectionally i.e. from the extracellular environment to the cell (outside-in-signalling) and from the interior of the cell to the extracellular environment (inside-out- signalling). Outside-in- signalling through Integrins takes place through ligands, which bind with the extracellular domain of Integrin, bringing about a conformational change allowing the transmission of signal into the cell. Integrin clustering may also occur at this stage. On the contrary, Inside-out-signalling originates duet to the recruitment of cytoplasmic proteins towards Integrins, activating the signalling cascade inside the cells, which ultimately leads to the activation of the Integrins. However, it should be kept in mind that this division into two distinct signalling entities may not be that black and white, instead both signalling events may occur simultaneously and reinforce each other [42-45].

1.2.1.1 Integrin inside-out-signalling

One of the pivotal steps in inside-out-signalling of Integrins is the recruitment of Integrin binding proteins towards Integrins. Many important molecules are involved in this recruitment. Except of some cases, Rap, as subfamily of GTPases has emerged as a common activator responsible for inside-out integrin activating signalling pathway. This is done by the ability of Rap1to form a complex with the integrin activator Talin and hence targeting of Talin to β integrin cytoplasmic tail. Another molecule, RIAM also plays a critical role in recruiting Talin to plasma membrane [40]. However, Talin and Kindlins are two important molecules involved in inside out integrin signalling by activating its cytoplasmic tail. A detailed role of these molecules in Integrin activation is described in section 1.1.3.

1.2.1.2 Integrin Outside-in-signalling

Although Integrin activation increases the affinity of individual Integrins for ECM ligands, it is the increase in avidity of the Integrin interaction which is required by a cell to bind strongly with the ECM. This increase in avidity is brought about by the clustering of Integrins leading to the summing of thousands of weak interactions into a tightly bound adhesive unit. Firs, Integrins cluster into unstable structures called nascent adhesions which progress further into more mature and stable focal complexes called focal adhesions and finally into stalk like fibrillar adhesions. Once Integrins have been clustered together through their activation, they are then able to transmit vast array of signalling cascades, collectively referred to as outside-in-signalling. Important intracellular changes arising from this outside-in-signalling include increased tyrosine phosphorylation of specific substrates and an

increase in the concentration of lipid. The short-term changes arising from inside-out-signalling include cytoskeletal rearrangements that allow cells initiate migration via dynamic connections between Integrins and filamentous (F-) actin while long term changes include changes in gene expression and signalling pathways that affect the survival, growth, and differentiation of cells [44, 46, 47].

1.3 FERM domains

Domains are important modules present in proteins and are regarded as structural as well as functional units of proteins. Previously known as the B4.1 (band 4.1) homology and ERM domain, the FERM domain is named for the four proteins from which this domain was originally described: Band 4.1 (F), Ezrin (E), Radixin (R), and Moesin (M). The FERM domain is approximately 300 amino acids in length and is found in a number of cytoskeletal-associated proteins that are localized to the plasma membrane and cytoskeleton interface [48]. This domain is responsible for the PIP2 and PIP3 regulated binding of FERM domain containing proteins to the membrane, which provide the vital platform for the association of the cytoskeleton with the membrane by linking actin filaments to adhesion proteins. The FERM domains contain three lobes, with the N-terminal lobe resembling ubiquitin, the central lobe resembling acyl-CoA binding proteins and the C-terminal lobe having a structure similar to that of PTB/PH domains [48, 49]. The FERM domain is also thought to bind through its PH like fold to the adhesion proteins in a PIP2/PIP3-regulated fashion to provide a link between cytoskeletal signals and membrane dynamics. One exception to this is the FERM domain of Radixin where binding with phospholipids is not mediated by the PH-fold like subdomain of FERM domain but occurs at a cleft between two subdomains on a relatively flat face of the module [50].

1.3.1 Diversity in FERM domains

In human genome, there are around 50 distinct FERM domains encoded by over 30 genes. An alignment of FERM domains from Homo sapiens, Caenorhabditis elegans and Dictyostelium discoideum suggests three broad groups of FERM domain containing proteins. These can be defined by their predominant members, namely Kindlin and Talin, ERM proteins, guanine nucleotide exchange factors (GEFs), kinases and phosphatases, and Myosin and Krev interaction trapped (KRIT) proteins [49]. Although most mammalian FERM domain containing proteins fit into these three broad groups, the biological functions of these proteins vary widely and in most cases, the protein interacting partners that regulate their biological activities remain poorly understood. The presence of Talin in D. discoideum, a protozoan, represents the very ancient origin of FERM domains. As D. discoideum is a motile organism that is able to execute random and directional migration

efficiently, the presence of Talin in it is likely to represent the minimum set of FERM domains that are required to connect the actin cytoskeleton with membrane protrusion in order to generate and control the traction forces that are needed for cells to move. The larger set of FERM domain containing proteins in humans probably reflects the indispensability of these domains to meet the requirements of greater complexity of mammalian cells and multicellularity. Human FDCPs include certain kinases (for example, Focal adhesion kinase (FAK) and Janus kinases (JAKs)), Myosins (MYO7, MYO10 and MYO15), phosphatases (protein-tyrosine phosphatase 1E (PTPE1)), ERMs, Kindlins, Talins and a number of less well-characterized proteins.

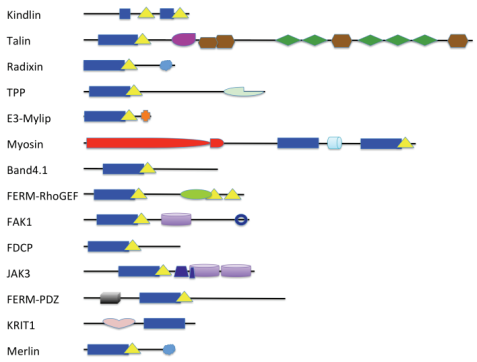

Figure 2 A schematic representation of FERM domain architecture in different FDCPs (See paper 2 for detail).

1.3.2 Kindlin FERM domain

Kindlin FERM domains is unique in the sense that its FERM domain is disrupted by a PH domain dividing the F2 subdomain into two parts [51]. No other FERM domain containing protein has disruption of its FERM domain by a Pleckstrin homology domain (PH domain). The presence of this PH domain in Kindlins seems to be an ancient phenomenon as it is present in the Kindlin FERM domain of most ancient origin i.e. in Amphimedon queenslandica, a very simple sponge. PH domain is a protein domain of approximately 120 amino acids that occurs in a wide range of proteins involved in intracellular signalling or as constituents of the cytoskeleton [52]. PH domain can bind Phosphatidylinositol lipids within cell membrane i.e. PIP3 and PIP2. Through interactions with these lipids, PH domains play a role in recruiting proteins to different membranes, and in this way targeting them to appropriate cellular compartments and also enabling them to interact with other components of the signal transduction pathways [53, 54]. The PH domain of Kindlins has a strong binding affinity

for PIP3 that makes it different from Talin that have strong binding affinity for PIP2. Binding of PIP3 with PH domain of Kindlins strengthens the binding of Kindlin-2 with plasma membrane which in turn facilitates Kindlin based Integrin activation [51].

1.4 Transcriptional regulation

The proper spatial and temporal expression of genes is indispensable for the accurate execution of biological processes such as development, proliferation, apoptosis, aging and differentiation. As a result, the deregulation of gene expression can often lead to disease. The completion of the human genome sequence and its annotation using computational and comparative genomic methods has led to the cataloguing of 20,000–25,000 protein-coding genes [55]. Key questions which post genomic era have posed, are related to understanding how these genes and their products function, as well as how their spatial and temporal expression patterns are established at both, the cellular as well as organismal level. Expression of each gene is governed by some regulatory elements and therefore, in order to understand the molecular mechanisms involved in the regulation of gene expression, it is important to identify the transcriptional regulatory elements associated with each gene. Also, to see the involvement of a gene in a pathological condition, it is important to understand how a gene is expressed and what players at molecular level are controlling the regulation of this expression. Thus, the functional elements in the genome not only involve the coding sequences translated into proteins but also the non-coding sequences involved in the regulation of gene expression. It is one of the main emerging challenges for genomics research to identify all functional elements in the genome, including those that regulate gene expression. The availability of the complete human genome sequence, in combination with genome-wide expression data and the advent of next 2^{nd} and 3^{rd} generation genomic sequencing [56-60] has tremendously facilitated the identification of regulatory elements controlling the gene expression. These databases can be used for individual genes and can serve as a starting point for studying the transcription regulation of human genes on a global scale. They provide information required to establish the spatial and temporal gene expression patterns and to elucidate the mechanism required for this establishment.

The expression of the genes encoding proteins (also called as class II genes) is regulated at different levels e.g. transcription initiation, transcription elongation, post transcriptional modifications and post translational modifications etc. However most of the transcriptional regulation occurs at the level of transcription initiation [61]. The initiation of transcription in Eukaryotes is started by RNA polymerase II. Eukaryotic genes transcribed by RNA polymerase II typically contain two distinct families of *cis*-acting regulatory DNA elements based upon their distance from the transcription starting site; (a) a promoter, which is composed of a core promoter and nearby proximal promoter,

and (b) distal regulatory elements, which can be enhancers, silencers, insulators, or locus control regions (LCR). These *cis*-acting transcriptional regulatory elements serve as recognition sites for *trans*-acting elements also known as transcription factors, which function either to enhance or repress the transcription [61].

1.4.1 Promoters and their role in Eukaryotic transcription

Promoters play important role in regulating the expression of a gene. They are located immediately upstream of the transcription start site and usually span 1-2kb promoter region. A promoter can be divided into two regions namely the core promoter region that starts from the transcription start site (TSS) and span 50-100bp while the other category is the proximal promoter that starts from immediate upstream of the core promoter [62]. The core promoter is responsible for the constitutive expression of a gene. It is harboured by transcription factor binding sites for general transcription factors i.e. TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH [63, 64]. The assembly of general transcription factors along with some mediators on the core promoter is termed as pre-initiation complex (PIC) [65, 66]. PIC directs the RNA polymerase Π to TSS and hence facilitates the initiation of transcription. The first step in PIC assembly is binding of TFIID, a multisubunit complex consisting of TATA-box-binding protein (TBP) and a set of tightly bound TBP-associated factors (TAFs). Once PIC is assembled, transcription then proceeds through a series of steps, including promoter melting, clearance, and escape, before a fully functional RNA polymerase II elongation complex is formed [67]. In the classical scheme, the core promoters have specific sequences known as TATA box. These TATA sequences facilitate the assembly of PIC by binding with TATA box binding protein. In addition to TATA box, the metazoan promoters are characterized by other regulatory elements e.g. Initiator element (Inr), Downstream Promoter Element (DPE), Downstream Core Element (DCE), TFIIB-Recognition Element (BRE), and Motif Ten Element (MTE) [62].

However, as more studies on the mechanism and dynamics of transcriptional regulation are coming out, it is becoming evident that classical scheme of promoter organization does not fit the new data originating from the latest research. For instance, it is now known that not all the genes have TATA boxes in their core promoters. In fact, TATA containing promoters represent only minority of promoters as most of the promoters lack any TATA box [68]. Additionally now we can divide promoters in two broad categories (1) those that are CpG rich and hence contain CpG islands (CGIs) and those that lack CpG islands. As a matter of fact, vertebrate genome differs markedly from invertebrates for having its major portion methylated at CpG dinucleotides. Because of this methylation, there is a general depletion in CpG dinucleotides in vertebrate genome giving rise to significant lowering of expected to observed ratio of CpG dinucleotides throughout the genome[69,

70]. However, the only exception to this pattern is the presence of CpG islands in the promoters of mammalian genes where a considerable majority of genes have CGIs in their promoters. Interestingly, all the housekeeping genes have CpG islands in their promoters while tissue specific genes usually lack CpG islands and instead contain TATA boxes in their promoters [68, 71]. Almost ubiquitous presence of CGIs in the constitutively active genes made the researchers to think that CGIs are permissive to transcription. Indeed, the latest advents in the tools for high throughput transcriptional analysis have shown that CGIs are not only resistant to methylation but also direct important transcription factors to promoters and make the transcription permissive by keeping the chromatin structure open [72-77].

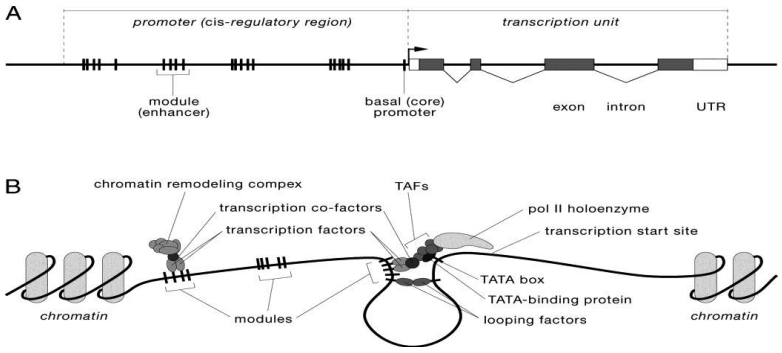

Figure 3 Promoter structure and function. (**A**) Organization of a generalized eukaryotic gene, showing the relative position of the transcription unit, basal promoter region (black box with bent arrow), and transcription factor binding sites (vertical bars). (B) General scheme of the operation of transcription through promoters and promoter interacting proteins. (Gregory et al, 2003, Mol Biol Evol taken with permission from Oxford university press)

1.4.2 The advent of ENCODE and role of epigenetics in transcriptional regulation

1.4.2.1 Epigenetics

Epigenetics is the study of heritable changes in gene expression or cellular phenotype, caused by mechanisms that involve changes other than genetic in nature i.e. mechanisms that does not involve changes in underlying DNA sequence. In eukaryotic cells, epigenetic modifications are encoded via

two primary modes which differ dramatically in their information content. First is the DNA methylation, the most well studied and characterized epigenetic modification, is binary in nature. Accuracy of DNA methylation pattern during development is extremely important for the survival of an organism and even minor disruptions of methylation density can be lethal during development. A broad range of developmental abnormalities and pathological conditions have been linked to abnormal methylation patterning [78, 79]. The second mode of epigenetic encoding is through modification of histones, the proteins that take part in the packaging of DNA in nucleosome. The nucleosome is the fundamental unit of the packaged DNA and it is composed of two copies of each of the four core histones (H3, H4, H2A and H2B) around which 146 base pairs of DNA are wrapped. Histones are evolutionarily highly conserved proteins, characterized by an accessible amino-terminal tail and a histone fold domain [80].

1.4.2.1.1 DNA methylation

The mechanism by which CpG methylation is transmitted from parent to daughter cell is fairly well understood. DNA methyltransferases are the proteins that take part in the methylation of DNA. This methylation patterning is first established during embryonic development by the *de novo* DNA methyltransferases DNMT3A/B, and then is maintained in subsequent cell divisions by the maintenance methyltransferase DNMT1 [81]. As DNA methylation occurs in binary fashion, the specificity of DNMT1 for hemi-methylated CpG dinucleotides provides a mechanism whereby CpGs in the newly synthesized DNA strand are methylated based on the presence of methylation in the CpG dinucleotide in the complementary template strand [82, 83]. In this way, a mean is provided for the transmission of methylation during the life of an organism. However, hereditary transmission of methylation from parent to offspring is complicated by the fact that a wave of demethylation occurs during early embryogenesis [83, 84]

1.4.2.1.2 Histone modification

Although the field of histone modification is still in its infancy and hence a vast majority of these modifications remain poorly understood, there has been progress in recent years in understanding the roles that methylation and acetylation of histones play in transcriptional regulation [83, 84]. The mechanism of maintenance of Histone modifications is not as clear as is DNA methylation. In yeast, the interaction between the modified histones and the effector proteins provide a framework in which existing histone modifications recruit chromatin modifying enzymes interactions between modified histones and effector proteins are maintained through cell division [85].This interaction provides a framework whereby existing histone modifications become targets for chromatin modifying enzymes

which act in a localized area to modify histones following DNA synthesis and cellular division. After random segregation of the histones, each daughter chromosome is expected to inherit some modified histones enabling the local propagation of the modification state to newly deposited histones through the actions of effector proteins and thus, the modification state could then spread locally through the actions of effector proteins to newly deposited histones [86] .The case for heritability of histone modifications in multicellular organisms is much less clear. The most compelling evidence comes from studies of mitotic inheritance of lineage-specific gene expression patterns catalysed by Polycomb-group (PcG) and trithorax-group (trxG) histone modifying protein complexes [87]. In order to understand the function of histone modification in disease it is required to know the pattern of distribution of these modifications in the genome. This has been achieved by use of state of the art techniques coupled with the use of modification-specific histone antibodies that has revolutionized our ability to study chromatin structure on a genome-wide scale. For instance, the investigators have coupled chromatin immunoprecipitation (ChIP) assay using modification-specific histone antibodies with genomic microarrays (ChIP-chip) or high throughput sequencing (ChIP-Seq) in order to examine the distribution of modified histones throughout the mouse and human genome [88, 89].

1.4.2.2 The ENCODE project

The launching of the encyclopaedia of DNA elements (ENCODE) is an epoch making event in the history of molecular life sciences. The project aims to fully describe the list of the functional elements that make up the human genome. The accurate activation of these functional elements constitutes all the cellular as well as physiological diversity one sees in living organisms. The main difference between ENCODE project and other genome projects is that the ENCODE project does not only provides information on the human genomic sequences but goes far beyond i.e. it contains the data about the degree of DNA methylation and chemical modifications to histones that can influence the rate of transcription of DNA into RNA molecules [90]. In addition to short range promoter-transcription factor interactions, ENCODE also examines long-range chromatin interactions, such as looping of the DNA molecule to interact with a distal functional element, that alter the relative proximities of different chromosomal regions in three dimensions and also affect transcription [91]. To find the functional elements in whole human genome, most state of the art techniques were adopted. One such technique is to find DNAse 1 hypersensitive sites, the presence of which indicates the sequences that are accessible to transcription factors [92-94]. Another important section of ENCODE is to catalogue the sequences and quantities of RNA transcripts, from both non-coding and protein-coding regions. The ENCODE pilot project focused on just 1% of the genome and its results hinted that the list of human genes was incomplete [95]. Although there was scepticism about the

feasibility of scaling up the project to the entire genome and to many hundreds of cell types, recent advances in low-cost, rapid DNA-sequencing technology radically changed that view. Now the ENCODE consortium presents a menu of 1,640 genome-wide data sets prepared from 147 cell types [95]. One of the more remarkable and astounding findings of ENCODE project is that 80% of the human genome contains elements linked to biochemical functions, rejecting the widely held view that the human genome is mostly composed of 'junk DNA' [95]. Interestingly, these results show that many of the DNA elements which were previously correlated with certain diseases lie within or very near non-coding functional DNA elements, providing new leads for linking genetic variation and disease. The data also shows that that about 75% of the genome is transcribed at some point in some cells and that genes are highly interlaced with overlapping transcripts that are synthesized from both DNA strands. This has forced the researchers to rethink the definition of a gene as the minimum unit of heredity. It has long been considered that the regulation of a gene expression is controlled by the regulatory elements situated in its close proximity. However, with the new data coming out of ENCODE project, it is becoming clear that beyond the linear organization of genes and transcripts on chromosomes, there lies a more complex network of chromosome loops and twists through which the core promoter, proximal promoter and the distal enhancer or silencer elements can communicate with each other [91].

In addition to many triumphs ENCODE project entails, there still are some challenges which have to be tackled in order to make maximum of it. One of the major future challenges for ENCODE will be to capture the dynamic aspects of gene regulation. Most of the assays done in ENCODE provide a single snapshot of cellular regulatory events, whereas a time series capturing of multiple events will throw more light on how these events happen. Similarly, the data obtained from multiple batches of huge number of cells may be misleading as individual cells in a batch can exhibit different behaviour. Therefore along with the parallel real-time analysis of the multiple regulatory events controlling the gene expression, use of single cell analysis will help even more to decipher the mysteries of gene regulation. In addition, perhaps the biggest challenge at the moment is to analyse and translate the tremendous amount of data generated by systemic approaches adopted in ENCODE project, into meaningful and understandable format that can be used to understand different dynamics of transcriptional regulation [90, 93, 96, 97].

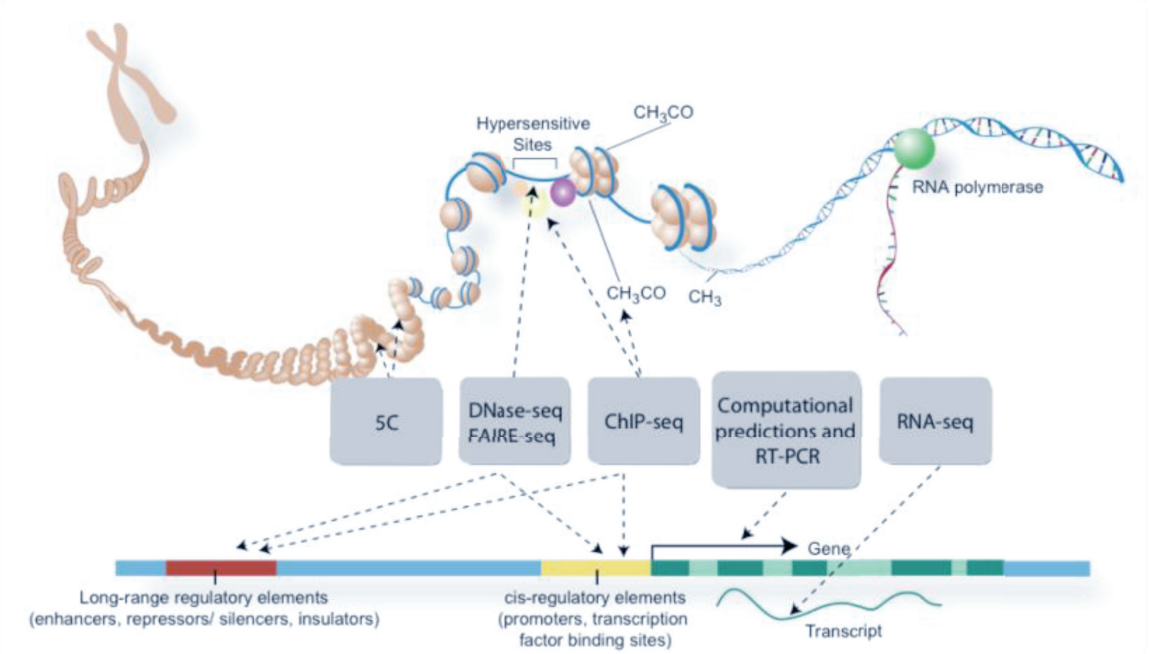

Figure 4 "ENCODE investigators employ a variety of assays and methods to identify functional elements. The discovery and annotation of gene elements is accomplished primarily by sequencing RNA from a diverse range of sources, comparative genomics, integrative bioinformatic methods, and human curation. Regulatory elements are typically investigated through DNA hypersensitivity assays, assays of DNA methylation, and chromatin immunoprecipitation (ChIP) of proteins that interact with DNA, including modified histones and transcription factors, followed by sequencing (ChIP-Seq)". (Description from University of California at Santa Cruz Credits: Darryl Leja (NHGRI), Ian Dunham (EBI)).

1.5 Metazoan Evolution

The transition from simple to complex life did not take place in one single quantum leap but instead occurred gradually. However this evolutionary voyage of living organisms can be petitioned in some key transitional events. The first major transition took place when cells without a well-defined nucleus attained nuclear membrane and hence gave rise to eukaryotic organization which gave rise to the

inception of complex cellular organization i.e. origin of membranous organelles. However for next two billion years, life on earth did not change appreciably and remained confined in single cell organization [98]. The second major transition took place some two billion years ago when the unicellular protozoans start clumping with each other and a sort of division of labour started to take place between them which ultimately lead to the appearance of multicellular organization. The

unfathomably long time taken for this step, from one cell to many celled organization itself reflect the gigantic effort nature had to put to attain this level of complexity in living organisms. The multicellular organisms in their very start diverged into two major domains of life, the plant kingdom and the metazoans or animal kingdom. The third major transition took place when metazons developed bilateral symmetry which gave rise to many highly developed features i.e. well defined plane of mirror symmetry dividing the body into left and right i.e. bilateral symmetry, triploblastic nature of germ layer because of the introduction of a third mesoderm. All these changes were accompanied with the advent of more complex organogenesis of the body, diverse cellular organization and more complex repertoire of cellular signalling, rapid locomotion, expansion in size and more evolved nervous system. Just before the Cambrian era, another major transition occurred when bilaterally symmetrical metazoans diverged into two major lineages namely protostomes and deutrostomes. In protsotomes the mouth develops from blastopore while in deutrostomes, blastopore becomes the anus and the mouth develops from a second opening on the dorsal surface of the animal. All the present day vertebrates are deutrostomes; signify the importance of this mega evolutionary event [99-101].

1.5.1 Molecular evolution: Evolution of proteins and regulatory elements

Classically, the concept of evolution is based on detailed comparisons of the morphological as well as physiological aspects of living organisms. However with advent of high throughput genome sequencing, it is the molecular tools dealing with the comparison of DNA and protein sequences of different organisms that are defining the foundations of our understanding of the evolution of living organisms. The process of natural selection that acts as the principle driving force for the evolutionary process to take place does not occur at morphological or physiological level but at molecular level. And the molecule that is directly acted upon by for forces of natural selection is DNA. It is the changes in DNA that provides the raw material for natural selection to act upon and hence evolutionary process to proceed. However DNA as we know is just a coding molecule which codes the information content and hence does not by itself make the structures that performs different functions. Therefore, the manifestation of natural selection can only be understood when we also take into account the other important molecule i.e. proteins. These are the proteins that are structural as

well as functional molecules giving the defining characters to organic life. The transformation of information content into physical entities in living organisms, the so called basic dogma of Biological sciences had been one of the greatest puzzles of Biology. Thanks to the advent of molecular Biology in 1950s that we now know that information content in DNA is translated into proteins which act as the physical entities to become different structures i.e. structural proteins or perform different functions i.e. enzymes, hormones, growth factors etc. The part of the DNA that encodes into proteins is called the coding region or the gene. As these are the genes that are translated into proteins, the study of the coding region of DNA has been considered the most important and fascinating part in the field of molecular evolution [102-104].

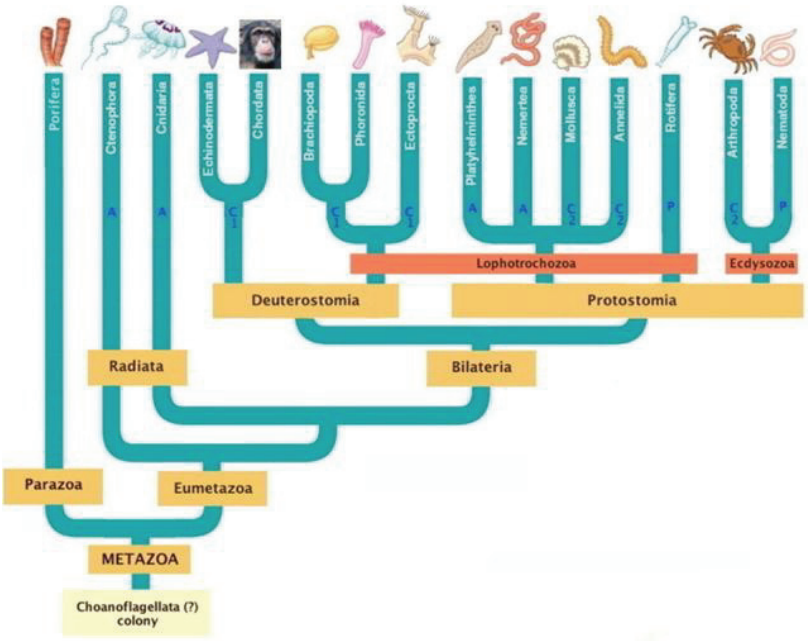

Figure 5 Schematic phylogenetic representation of Metazoan evolution. (Taken from the web (http://www.erectus.it/main.html with permission).

1.5.2 Evolution of transcriptional regulation

If we look at the evolutionary history of living organisms, we can see that there has been a gradual increase in the genetic tool kit these organisms possess, owing to the morphological as well as physiological diversity and complexity they exhibit. For instance, human genome contains 6 times more genes then the simplest prokaryotes and double to that of drosophila. However, this is not always true as plants have even higher number of genes then human [105]. Many of the species which may be related to distant taxa still show very high conservation in the coding regions of their genes i.e. genes of most of the vertebrates are highly conserved. For instance, genes of most of the vertebrates are highly conserved. Therefore the repertoire of morphological diversity we see in living organisms can not only be explained on the basis of changes occurring in coding region of the genes [100, 103]. The regulatory regions of the genomes are the parts that regulate the expression of the genes. If coding sequences are there to dictate the linear amino acid sequence in primary structure of protein, regulatory sequences dictate the special as well as temporal expression of a gene. In this way, regulatory motifs act as secondary codons and in the absence of these secondary codes, there will be no expression of proteins and hence no selection pressure which would ultimately lead to extinction of that gene. In addition, as these regulatory sequences regulate the temporal, spatial, quantitative and qualitative expression of a gene, the morphological as well as physiological diversification generated by this transcriptional control is enormous. However, unlike the evolution of protein coding regions, which has become a very mature field as it is aided by well-defined and thoroughly studied tools and techniques, the field of studying evolution of non-coding regions or regulatory regions is still in its infancy. Another hindrance is the much less functional constrain acted on non-coding regulatory elements as compared to coding regions. Because of this relaxed constrain, it is quite difficult to find and then characterize evolutionarily important regions or segments of the genome. Nevertheless, considerable amount of progress have been made in this field in last decade or so. One of the biggest developments in this direction is of that of ENCODE project that very much deals with the identification and characterization of importance of regulatory elements at genome scale. The ENCODE data reaffirm the previous holding that promoters play the most important role in controlling the gene expression and although certain level of diversity is found in promoter structures, most of them are evolving under negative functional constrain as compared to the neutral genomic sequences. Promoters do not exhibit homogenous rate of variation. Some promoters contain cluster of conserved elements or transcription factor binding sites and evolve under the effect of purifying selection while other contain rapidly evolving segments and evolve under positive selection pressure. The latter case is interesting as positive selection has been shown in the promoters of many closely related species. A study comparing 20 well-characterized regulatory regions in mammals showed that

approximately one-third of the binding sites in humans are probably not functional in rodents [59]. This relaxed constrain on promoters serve as the mean of novel transcriptional regulatory schemes for gene expression and hence enhancing the diversity in cellular as well as organismal phenotypes [102] [104, 106].

1.5.2.1 Type of mutations affecting transcriptional regulation

Mutations in the regulatory region affecting the transcription of a gene can be of many types. (1) Small-scale, local mutations, small indels, point mutations, can modify, eliminate or generate new binding sites and can also affect transcription by altering the space between TFBS [107-111]. (2) Process of transposition can lead to insertion of new regulatory sequences into promoters generating new TFBS or can lead to the disruption of already present TFBS [112] (3) Through the process of retroposition many new promoters can be assembled [113, 114]. (4) Gene duplication may lead to the generation of two promoters one of which may acquire novel *cis*-elements due to the redundancy produced as a result of this duplication event and (5) Gene conversion is another important but less studied phenomenon that can spread regulatory elements within a gene family [109, 115-120].

1.5.2.2 Effect of mutations in regulatory sequences

The aforementioned changes occurring in noncoding DNA leads to several types of evolutionary change in gene expression (1) temporal changes or changes in timing of gene expression have been documented in many taxa [121-123]. (2) Differential expression of genes in different tissues or changes in spatial extent of gene expression. Spatial changes are of peculiar interest as they affect many important features directly i.e. body proportions, organ size and number and a great many other anatomical features [124-127] (3) Changes in the level of gene expression is another important evolutionary parameter that is affected by the changes associated regulatory sequences [128-130]. (4) The evolutionary changes in transcriptional responses to different environmental changes play important role in the evolution of polyphenism and contribute towards phenotypic plasticity [131, 132]. (5) With the advent of Microarray, many studies have shown that populations can harbour variation in which genes are expressed in a sex-specific manner [133] (6) The multicellular organisms also exhibit a peculiar feature of transcriptional regulation which certain genes are expressed in a succession certain phases of their life cycles, both spatially as well as temporally. For instance, a gene expression can be abandoned if the transcription factor controlling its expression is not expressed in that specific region or during a specific time point i.e. Several independent losses of patterning roles for homeodomain transcription factors in arthropods [134-139]. Conversely, promoter can acquire a new binding site giving rise to a new regulatory linkage, a process known as recruitment or co-option

[140-142]. In addition to the *cis* changes in promoter sequence of gene, the *trans* changes i.e. changes in the promoter sequence of transcription factors, can also affect the transcriptional regulation of gene of interest. Three classes of mutations can be given as an example here. (1) Mutations affecting the transcription regulation of an upstream transcription factor [98, 143-147]. (2) Mutations which affect the DNA-binding domain of an upstream transcription factor can also lead to change the expression profile that might have evolutionary consequences [148-151] (3) And finally, mutations affecting the protein-protein interacting domain of transcription factor which would affect their ability to bind with transcriptional activators and co-activators and hence change the expression profile of a gene which in turn might have evolutionary significant phenotypic effects [152-156].

In short, we can say that promoters and hence the transcriptional regulation of genes are evolving under different but distinct functional constrains. A promoter consists of nucleotide sequences which may or may not act as binding sites and hence the selection pressure on promoters very much depend upon the number and types of TFBS it harbours. Consequently promoters would obviously be less constrained than the coding regions and more than the neutral intronic regions. And this relatively relaxed conservation of regulatory sequences leads to the novelty and diversity in gene expression which ultimately contributes towards organismal evolution.

2 Methods and techniques

2.1 Phylogenetic analysis

The use of phylogenetic analysis is the most common and powerful routine towards understanding the evolutionary history of genes and proteins. Some of the most commonly used phylogenetic analysis includes sequence alignments (both global as well as local), BLAST, estimation of the rate of substitutions in different sequences, estimation of the selection pressure on genes/proteins, estimation of rate of indels, generating phylogenetic trees and phylogenetic shadowing for predicting TFBS in the regulatory regions etc. All these techniques and tools were used in different projects we worked on.

In order to identify residues (traces) which are important in FERM domain containing proteins **(Article 2)**, we employed evolutionary trace analysis (ETA) (http://mammoth.bcm.tmc.edu/). ETA divide the input sequence into two parts one known as the superclass which consists of all the sequences used as input, while the other group is called as subclass which consists of sequences which are only one of the component of the input sequences. Conserved residues in the former determine the globally important residues while the one conserved in subclass but not in superclass would define the class specific residues [157]. In our study, we made use of ETA to mainly see which are the class specific as well as superfamily specific residues in Kindlin FERM domains. In superclass analysis, the residues that were conserved whole along the tree were given the maximum rank, 1. Same rank was given to residues in the subclass if they were conserved in all the members of the subclass. The class specific residues were obtained by subtracting the highest ranked residues in subclass from the superclass. For structural analysis, we used *pyetv* tool, a modified vision of *pymol* employed in the ETA webserver [158]. For the evolutionary study of noncoding regions (**Article 3**), promoter sequences were aligned by using dialign alignment tool specialized for the alignment of non-coding sequences [159]. In order to estimate the extent of divergence that has taken place in mammalian promoters, we selected six mammalian species (in addition to human) two primates, two rodents and two laurasiatharians. K-estimator software was used to calculate the divergence (k) in Kindlin promoter sequences. K-estimator calculated the overall number of nucleotide substitutions per site by using different site correcting methods [160]. For detection of CGIs in Kindlin promoters in 5'upstream region, CGI searcher program was used. The threshold set for assigning CGIs to a gene was length: 500bp, CpG obs/exp frequency: 0.6 and finally, GC percentage: 55% [161, 162]. The site correcting method we used was Kumara-2p method employed in the program [163]. For estimation of histone modification on core promoter sequences, ENCODE project database employed in UCSC genome browser was used [88, 164-166].

2.2 Promoter cloning

To characterize Kindlin promoters, we cloned them by conventional cloning methods. Briefly, the putative promoter sequences were amplified by PCR using human genomic DNA. The product was cloned upstream of the *firefly* luciferase gene in pGL-4.21 (Promega, USA). Progressive 5′-end deletions of Kindlin-2 and Kindlin-1 promoters were performed by PCR amplification by using primer sets for these promoters. All constructs were isolated with Jetquick maxi prep plasmid extraction kit Genomed, USA) and confirmed by DNA sequencing (Macrogen, South Korea).

2.3 Transfection assay

Cells were plated in 24-well plates and incubated overnight. All the constructs for Kindlin-2 and Kindlin-1 promoters cloned in the luciferase vectors were transfected into cells with 0.6 µl Fugene-6 and 200 ng plasmid DNA per well, following the manufacturer's instructions. In all transfections, Renilla luciferase expression plasmid, pRL-SV40 was co-transfected as a transfection control (20 ng per well). The cells were cultured at 37C^{o} for 48 h. Luciferase activities were measured by the Dual Luciferase Reporter Assay system (Promega, USA). For serum treatment, 4 h before the transfection, cells were starved with 0% serum free medium and incubated for 24 h and then treated with 10% serum. Luciferase assay was done at different time intervals after the serum treatment.

2.4 Realtime PCR assay

In order to see the endogeneous expression of Kindlin mRNAs **(article 1)** effect of serum starvation on endogenous expression of Kindlin-2 mRNA **(article 3)** and the effect on Kindlin-2 on Gli1 expression **(article 4)**, real-time RT-PCR method was employed by using SYBR Green. Triplicate samples of each PCR mixture, each containing 4.7 µl of POWER SYBR Green PCR master mixture (Applied Biosystems), 0.3 µl of a 10 pmol/µl of Primer mixture, 0.3 µl of cDNA, and Water to a total volume of 1 µl were transferred into a 96-well plate on an ABI 7500 Fast Real Time PCR System (Applied Biosystems). The samples were initially incubated at 95°C for 3 min, fol- lowed by 45 cycles with 95°C for 15 s, 60°C for 60 s. Dissociation curves were generated after each PCR run to ensure that a single specific product was amplified. The results were analysed with the comparative Cycle threshold (Ct) method. For normalization, we used the expression level of βactin (ACTB) or GAPDH.

2.5 Western blot

Western blot assay was employed in different studies **(article 3, article 4 and article 5).** Briefly, Total protein was extracted from serum starved or serum treated 3t3 cells by cell lysis buffer

containing 20 mM Tris–HCl (pH 7.5), 1% Triton X-100, 100 mM NaCl, 0.5% Nonidet P-40 and protease inhibitor cocktail (Roche, Germany). 20 µg of protein per lane was resolved on 10% sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS–PAGE) and electro-blotted onto a PVDF membrane (Millipore). Blots were incubated with an affinity purified anti-Kindlin-2 polyclonal antibody (1:1000 dilution) or an anti-tubulin antibody (Santa Cruz). Following incubation with horseradish peroxidase (HRP)-conjugated anti-rabbit secondary antibody at 1:3000, detection was performed with a chemiluminescent substrate kit.

2.6 siRNA treatment

Pre-validated Kindlin-2 siRNA, GLI1 siRNA (article 4) and SP1 siRNA (article 5) and control siRNA were used for different knockdown assays. The siRNA transfection was done with Lipofectamine-2000 (Invitrogen). Briefly, cells were plated in 6-well plates and were transfected with SP1 siRNA (25 nM and 50 nM) or control siRNA (50 nM) on the following day. The medium was changed six hours after the transfection. Cells were harvested after 72 hours.

2.7 Chromatin in situ hybridization (ChIP)

In order to see the binding of SP1 with Kindlin-2 promoter, ChIP assays were performed for the transcription factor SP1. PC3 cells were fixed with 1% formaldehyde, quenched using 1.25 M glycine, resuspended in SDS lysis buffer and sonicated (10 s pulses×5 at 50% duty cycle, output 5, Diagenode sonicator) to generate chromatin sizes between 200 and 1000 bp. 20 µl of soluble DNA fraction was saved as input (10%). 5 µg of anti-SP1 (Santa Cruz) and mouse IgG (Millipore) antibodies were used for the ChIP assay. 2 µl of eluted DNA from ChIP and input reactions was used for PCR with primers -150F 5′ AGGGCAGCTCTGCGGGCGGCGAA -3′ and +450 ′ AGAGCCATGGCTCCTTCCTGCG -3′ using the program 95°C for 2 min (95°C for 30s, 60°C for 60s and 68°C for 1 min) 40 cycles and 68°C for 7 min. The products were resolved on 1% agarose and were confirmed for specificity by sequencing. Each ChIP assay was repeated twice.

3 Main objective of the study

The main aim of my PhD study was to elucidate the evolutionary history of Kindlins (both, the evolution of Kindlin proteins as well as the evolution of Kindlin transcriptional regulation) and characterize their transcriptional regulation.

3.1 Specific objectives

1- To elucidate the evolutionary history of individual Kindlin proteins
2- To elucidate the FERM domain of Kindlins with respect to the evolution of the FERM domain of other FERM domain containing proteins.
3- To elucidate the evolution of transcriptional regulation of Kindlins
4- To study the transcriptional regulation of Kindlin-2

4 Result and discussion

4.1 Evolution of Kindlins

A lot of work has been done on the functional and clinical aspects of Kindins. However, to best of our knowledge, not much work has been done on understanding the evolutionary history of Kindlins. Towards this end, we tried to explore the evolution of Kindlins. In our first study, we tried to explore the evolutionary history of individual Kindlin proteins. The phylogenetic tree for Kindlins suggested an interesting evolutionary history for Kindlin family proteins. The phylogeny exhibits a topology of the form (A)(BC) ie, Kindlin-1 and Kindlin-3 form a cluster while Kindlin-2 forms an out-group. It appears that in invertebrates before the divergences of arthropods, a single ancestral Kindlin had its ancient origin in earliest metazoan, like hydra. This single Kindlin gene copy underwent two duplication events in vertebrates that may have occurred together with two rounds of genome duplication. Regardless of the exact mechanism, these events gave rise to three Kindlin paralogs: Kindlin-1, Kindlin-2 and Kindlin-3. The three vertebrate Kindlin paralogs maintained in all subsequent vertebrate forms, probably due to the selection pressures that have promoted the diverse and complicated morphological and physiological properties of vertebrates. Comparison between non-synonymous and synonymous substitutions in orthologous Kindlin transcripts showed that although all the vertebrate Kindlins are evolving under purifying selection, Kindlin-2 is under much more constrain as compared to other Kindlin paralogs. This result is congruent with the experimental data on Kindlin expression we showed in this work. For instance, we showed that Kindlin-2 is a ubiquitously expressed gene playing its structural and functional roles in broad array of tissues. It has been indicated in various studies that ubiquitously expressed genes tend to evolve slowly compared to those with tissue specific expression. As noted previously, Kindlin-1 and Kindlin-3 are expressed in specific tissues (epithelial tissues and the hematopoietic system, respectively), and it is therefore commensurate that the evolutionary rate of these paralogs is much higher than that of Kindlin- 2. In support of this are Kindlin knock out studies for all three Kindlins which show that Kindlin-2 knockout mice die during early embryogenesis. In contrast to the relatively milder phenotypes of Kindlin-1 and Kindlin-3, these studies support the idea that Kindlin-2 is under tighter functional constraint than Kindlin 1 and 3. We also employed Tajima relative rate test to estimate the rate of divergence of Kindlin paralogs with respect to each other. This approach once again reaffirm the strong conservation of Kindlin-2 as compared to its two other counterparts. The distinct expression patterns of all three Kindlins in vertebrates thus supports the subfunctionalization model of gene duplication.31 It seems likely that that the function of the ancestral unduplicated Kindlin was subfunctionalized in vertebrates in part due to the divergence of Kindlin expression location. Thus,

while Kindlin-1 and Kindlin-3 are expressed exclusively in epithelial and hematopoietic tissues, respectively, Kindlin-2 – being the representative of ancestral Kindlin gene – is expressed in a variety of tissues, in a pattern less ubiquitous than the original unduplicated Kindlin gene. Ultimately, it seems that this subfunctionalization of Kindlin expression patterns may have provided a degree of selective advantage associated with the diversification of higher order functions performed by Integrins in various tissues.

4.2 Evolution of the FERM domain

Exploring the evolutionary history of domains is an important step towards understanding the structural and functional importance of proteins. Towards this direction, we started with elucidating the evolutionary history of the proteins containing FERM domains. Very little is written on the evolution of FERM domain containing proteins and whatever is written is written as a supplementary component of other major studies. In our previous work (see section 4.1), we have elucidated the evolution of Kindlin family of proteins. In the current project, we explored the evolution of the FERM domain of Kindlins with respect to other FERM domain containing proteins. We selected one paralog of all the FERM domain containing proteins. In this way, we generated the phylogeny of 14 proteins, across diverse range of species both from vertebrates and invertebrates, generating a comprehensive neighbour joining phylogenetic tree from 179 protein sequences in total. Some interesting results were obtained from this phylogeny. All the FERM domain containing proteins seems to be belonging to eukarya and none was found in the other domains of life i.e. archea or bacteria. In eukarya, FERM domain containing proteins went back mainly to metazoan with the only exception of Myosin and Talin that goes back to protozoa. In metazoan, we can see two classes of FERM domain containing proteins, those which went back to parazoa (sponges) and those which went back to Eumatzoa which in turn is divided into two groups: the proteins present in non bilateria i.e. Cnidaria and the other one to bilateria i.e. to fruity fly. None of the FERM domain containing protein was found in plants, another major group of Eukarya after metazoans. A thorough analysis of the FERM domain sequences of the FDCPs reflected some interesting insights in the evolution of these proteins. Firstly, a comparison of the Kindlin proteins sequences reconfirmed our previous finding that Kindlin-2 is the most conserved protein in Kindlin family while Kindlin-1 and Kindlin-3 are relatively more diverged. Kindlin FERM domain show the closet homology with domain of another important Integrin interacting protein, Talin1 (27%) followed by the FERM domain of Merlin (16%). Kindlin FERM domain shows least homology with (only 2%) JAK3. In terms of overall percentage similarity with other members of the FERM domain superfamily, Radixin and Merlin showed the maximum conservation (20%). Kindlin shows considerably less conservation (10%) while both the kinases in this group i.e. JAK3 and FAK1 showed the least overall conservation (7%). Not much variation was

seen in the orthologous sequences for each protein where all the proteins showed more than 80% conservation in within their respective orthologous sequences. The only exception to this was JAK3 which showed 68% conservation, much less than the other FERM domain super family members. Unlike any other FERM domain containing protein, the FERM domain of Kindlins has a unique disruption because of the insertion of a PH domain in it. Thus Kindlins in this way would have two PH domain modules one because of the insertion and the other one already present in its third domain (also known as FERM-C or F3 subdomain). Calculation of the substitution based distances in the FERM-M and FERM-C domains in orthologous Kindlins in Kindlin-1 and Kinlin-2 clearly showed that FERM-C subdomain have undergone much higher level of substitutions as compared to FERM-M. To confirm it further, we did the same analysis in Talin too where we could clearly see that FERM-C in Talin was much more conserved then FERM-M subdomain. The most prominent implication of this interesting evolutionary event is that Talin and Kindlin differs in terms of their binding affinities for two of the most important membrane phosphor lipids i.e. Phosphatidylinositol (4,5)-bisphosphate (PIP2) and Phosphatidylinositol (3,4,5)-trisphosphate (PIP3) where Kindlin-2 shows greater affinity to bind with PIP3 and Talin with PIP2. Evolutionary trace analysis of Kindlin FERM domains within themselves and with respect to other FDCPs singled out those traces which are important for individual Kindlins and which are important for the superclass as a whole. Within the Kindlin proteins, ETA shows that Kindlin-2 has the maximum number of class specific residues as compared to both Kindlin-1 and Kindlin-3. This is very much in accordance with our previous finding in which we showed that Kindlin-2 is the most conserved protein in Kindlin protein family. This can be seen from another angle too. In our previous work we had speculated that ubiquitous expression of Kindin-2 might be the major cause of extra functional constrains exerted on Kindlin-2. On the other hand, Kindlin-1 and Kindlin-3 being tissue specific in terms of their expression, should evolve faster than Kindlin-2. This fact is even more prominent in case of Kindlin-3 where we can see only 14 class specific residues as the expression of Kindlin-3 is exclusively restricted in hematopoietic system. Many residues enhanced by our ETA analysis have been cited in the literature as important and vital naturally occurring and experimentally engineered variants of the respective protections. This enhances the validity of ETA as tool to predict the important components of a protein which can be used for future mutation analysis. For instance we found many highly ranked superclass specific residues that have not yet been mutated and hence our results provide them as important candidates for the future mutation analysis in Kindlins and other FDCPs.

4.3 Evolution of the transcriptional regulation of Kindlins

In our previous study we have shown that Kindlin-2 is the most conserved Kindlin protein and speculated that this relatively conserved evolution of Kindlin-2 may be due to its ubiquitous expression pattern which contributes towards the subfunctionalization of ancient unduplicated invertebrate Kindlin into three Kindlin paralogs in vertebrates [1]. This laid down the foundation of our third project in which we tried to explore the evolution of transcriptional regulation of Kindlins. In addition to look at genomic sequence level, we also made use of ENCODE histone modification and RNA seq data and showed that pattern of histone modification was quite conserved in Kindlin promoters although it does not match very well with the conservation of expression profile of Kindlins. A study of the rate of substitution of mammalian Kindlin promoters gave a reflection of the extent of divergence in these promoters. It was quite clear from the results that Kindlin-2 showed the least divergence while Kindlin-1 and Kindlin-3 promoters showed relatively more divergence with Kindlin-3 being the most divergent of all three Kindlin promoters. For all the Kindlins, core promoter showed more conservation as compared to the full length promoter. Interestingly, the divergence pattern of Kindlin promoters very much followed the same pattern as the evolutionary pattern of Kindlin proteins, as shown in our previous work. Interestingly however, in contrast to Kindlin proteins which were more constrained then the neutral genomic sequences, Kindlin promoters seems to be evolving under positive selection. This is in accordance with other studies which show that promoter sequences tend to evolve under positive selection which contributes towards the diverse expression pattern of their relevant genes by allowing novel and expanded combinations of transcription factors to bind with the *Cis* elements harboured by these promoters. Almost similar trend was also seen when we analysed the rate of *indels* in these promoters. As Kindlins exhibit a peculiar pattern of expression i.e. Kindlin-2 being expressed in broad spectrum of tissues while Kindlin-1 and Kindlin-3 pre-dominantly expressed in epithelial and hematopoietic tissues, we attempted to explore the presence, localization and extent of CGIs in Kindlin promoters. Interestingly, Kindlin-2 showed CGIs in all the mammals we chose. No CGIs was seen in Kindlin-2 promoters in cold blooded vertebrates i.e. lizard and frog. Kindlin-1 on the other hand, showed a non-uniform pattern of CGIs across the species selected for this study i.e. the upstream region of Kindlin-1 promoter lacked CGIs in all the species while in only human and mouse CGIs was found in some part of the 5'end UTR region but not in the region upstream of TSS. In all the species, Kindlin-3 did not show any CGIs whatsoever, whether upstream or downstream of the TSS within the promoter region. As presence or absence of CpG islands have been associated with the expression of genes, the specific profile of CGIs in Kindlin promoters can be correlated with the expression profile they exhibit and therefore might have its evolutionary implication. We compared the RNA-seq based expression in different cell

lines for all the three Kindlins with the H3K4me3 modification pattern in the core promoter regions. It was interesting to find that both human and mouse showed a significant correlation between the expression of Kindlins in different cell lines and modification of H3K4me3 in these cell lines at 95% significance level. It is in accordance with the previous studies and data released by ENCODE that H3K4me3 is a reliable predictive marker for transcriptional activation. We then tried to know whether or not modification in H3K4me3 in Kindlin promoters is conserved in mouse and human. Interestingly, we saw significant conservation of histone modification in mouse versus human comparison. However, this high level of conservation of H3k4me3 signatures was not in line with the conservation pattern of expression of Kindlins. In Kindlins, we clearly saw difference in conservation of three Kindlin paralogs in mouse/human comparison which actually was more in line with the promoter divergence of these proteins. This mean that histone modification although are important for regulating the expression of genes, their role in the evolution of transcriptional regulation may not be that much direct and more rigorous efforts are needed to decipher this relationship.

4.4 Kindlin-2 and hedgehog signalling

Once we had elucidated the evolution of Kindlin transcriptional regulation, we next turned our attention towards the molecules that might play their role in transcriptional regulation of Kindlins. We have recently demonstrated that Kindlin-2 regulates the drug sensitivity of prostate cancer cells [13]. However, the mechanism remained to be elucidated. To this end, Hedgehog signalling has been known to play an important role in the regulation of prostate cancer progression. We then interested in scrutinizing the possible link between Kindlin-2 and the SHH effector GLI1 in prostate cancer cells. In order to see the effect of GLI1 on Kindlin-2, when we knocked down GLI1, level of Kindlin-2 expression increased showing the negative effect of GLI1 on Kindlin-2 expression. Ectopic expression of GLI1 in HKC cells having low endogenous expression of GLI1 also decreased Kindlin-2 level, both at protein and mRNA level. Next we detected whether the inhibition of Hedgehog signalling by cyclopamine, a specific inhibitor of Smoothened could mimic the effect of GLI1 in prostate cancer cells. Interestingly, we found that Kindlin-2 protein was increased 4 h after treatment, and reached to a peak at 8 h, after that the Kindlin-2 protein was gradually decreased both in DU-145 and PC-3 cells, suggesting that Kindlin-2 is an early response gene of Hedgehog signalling. Altogether, our finding suggests that Kindlin-2 is a new regulatory target of Hedgehog signalling. We also found that GLI binds with Kindlin-2 promoter and transient transfection of GLI1 leads to down regulation of Kindlin-2 promoter activity which further confirmed our finding that GLI1 affect Kidnlin-2 expression at transcriptional level. In order to see whether there exists a feedback loop between Kindlin-2 and GLI1, we knocked down Kindlin-2 using siRNA in PC-3 and found that GLI1

mRNA and protein were also reduced. Additionally, overexpression of Kindlin-2 also increased the mRNA and protein of GLI1 confirming the presence of a feedback loop between Kindlin-2 and GLI1. We also examined whether the feedback loop was functional in the presence of cyclopamine. In consistence with previous studies, GLI1 was found decreased by the inhibition of Smoothened using and the presence of Kindlin-2 significantly reduced the effect of cyclopamine on GLI1. It has been reported that activated glycogen synthase kinase 3β (GSK3β) phosphorylates GLI1 and induces its degradation. In this regards, we inhibited GSK3β activity and found a significant increase in GLI1 protein level. These results suggested that Kindlin-2 positively regulates GLI1, a mechanism that is independent of Smoothened but partially through inhibition of GSK3β without the exclusion of other mechanisms. It was known that cyclopamine can induce cell death in prostate cancer cells. In view of the above results, we wondered whether Kindlin-2 is involved in the drug sensitivity of prostate cancer cells to cyclopamine. Importantly, apoptosis induced by cyclopamine was indeed increased when Kindlin-2 was knocked down. In support of this result, Kindlin-2 knockdown increased ratio of Bax/Bcl-2 induced by cyclopamine in PC-3 cells, while the ratio was unchanged in Kindlin-2 knockdown cells without the presence of cyclopamine. In addition, the cell proliferation remained unchanged indicated by proliferating cell nuclear antigen (PCNA). We also showed that although Kindlin-2 shRNA alone has no effect on PC-3 cell viability, it could notably facilitate the cell death induced by cyclopamine suggesting a therapeutic window where knocking down of Kindlin-2 in combination with addition of cyclopamine could be more efficient way of killing prostate cancer cells.

4.5 Role of SP1 in transcriptional regulation of Kindlin-2

In our evolutionary studies, we had shown that Kindlin-2 is the most conserved Kindlin protein and speculated that specific ubiquitous expression pattern of Kindlin-2 might be one of the reason of this conserved evolution. We also showed that Kindlin-2 promoter is highly CpG rich and we know that CpG islands are present in all the ubiquitously expressed genes and hence have been associated with expression profile of the respective genes. This made us to explore the transcription factor that would target these CpG islands and hence could be the reason for ubiquitous expression of Kindlin-2. In order to characterize the Kindlin-2 promoter, serial deletions were made of 1600bp long Kindlin- 2 promoter and their activity was tested by luciferase assay in COS7 cells. Luciferase assays clearly showed that this 1600bp long Kindlin promoter is highly active promoter, a characteristic of the promoters of ubiquitously expressed genes. Additionally, we also found that major transcriptional activity of Kindlin-2 promoter is controlled by a small 39bp long fragment residing in -122 to -83 upstream regions with respect to Kindlin-2 transcription start site. It reflects the simplicity of Kindlin-2 promoter which can be regulated by a small fragment located very close to the transcription start

site. In order to the characterization of the Kindlin-2 core promoter, we carried out both *in silico* as well as experimental tests. *In silico* analysis of 300bp long region upstream of the Kindlin-2 TSS revealed that it is highly CpG rich region. Additionally, it contained many SP1 biding sites. Further assays with SP1 inhibitors and SP1 siRNAs clearly showed that both, at promoter level as well as transcriptional level, SP1 played its role in controlling Kindlin-2 expression. We further showed by ChIP assay that SP1 regulate Kindlin-2 expression by directly binding with Kindlin-2 core promoter. As SP1 itself is expressed ubiquitously, we can hypothesize that it is SP1 that is responsible for the ubiquitous expression of Kindlin-2. Additionally, absence of SP1 binding sites in Kindlin-1 and Kindlin-2 core promoters may be the reason of tissue specific expression of these two Kindlin paralogs.

5 Concluding remarks

Kindlins are relatively new entry into the club of Integrin interacting focal adhesion molecules. A lot of work has been done on the functional aspects of the Kindlin proteins. However two of the important aspects of Kindlins have not been given much importance yet. One is the evolutionary history of Kindlins and the second is their transcriptional regulation. This PhD project is a step towards this direction so that the gap present at the moment in our understanding of these two important aspects of Kindlins can to some extent be narrowed.

Elucidation of the evolutionary history of genes on one hand tells us about the different aspects of their evolution and on the other hand helps us to understand their function. In our first paper, we carried out a comprehensive study of phylogenetic history of Kindlins and found out that Kindlins have quite ancient origin with the most primitive representation in earliest metazons. We also showed that in Kindlin paralogs, Kindlin-2 is the most conserved of Kindlin proteins and hence it could be the representative of ancient unduplicated single Kindlin gene. FERM domains are the structural as well as functional hallmark of Kindlins and hence elucidation of their evolutionary history would help us in understanding many functional as well as structural aspects of Kindlins. And this is the main objective of our second paper. To best of our knowledge, it is the first most comprehensive study of evolution of FERM domains. Appearance of FERM domain containing proteins in Dictyostyleium means that these proteins might have played important role in transition from unicellular to multicellular organization. More rigorous analysis towards this direction may enlighten us more about the origin of multicellular organization and the role FERM domain played in this connection. In this work we showed the evolutionary implication of unique disruption of Kindlin FERM domain by PH domain. We showed that this might have physiological aspect too and exploring more rigorously towards this direction can be of great importance for the future research in the field of Kindlin proteins. The use of evolutionary trace analysis of FERM domains of different proteins enhanced our knowledge of important residues in FERM domain containing proteins in general and Kindlins in especial. This knowledge can be used as candidate traces for future mutation analysis which could be of great value to understand the mechanism of different functions these protein domains are performing i.e. cell adhesion, cell to cell contact, cell migration, Integrin inside- out- signalling and acting as adaptor between cell membrane and Actin cytoskeleton etc. Additionally, these important traces can also be considered as potential naturally occurring variants involved in different diseases and can be of great clinical importance.

When it comes to evolution, not only coding regions are important but the evolution of regulatory regions also is of the same importance. Elucidation of transcriptional regulation of Kindlins is even

more important as Kindlins exhibit very distinct expression profile which is extremely contrasting with respect to each other. In our attempt to elucidate the evolution of transcriptional regulation of Kindlins, we have shown that there exists some sort of synergy between Kindlin protein evolution and the promoter evolution which in turn comply with the pattern of expression divergence we see in mammalian Kindlins. This gives rise to important questions which have evolutionary, functional as well as clinical relevance. For instance, we have shown that Kindlin-2 is ubiquitously expressed gene and we speculated that this ubiquitous expression could be the reason of more stringent evolution of Kidnlin-2 as compared to its two other counterpart paralogs. The synergy we showed between Kindlin protein evolution and promoter evolution thus may confirm our hypothesis that not only coding regions but also the regulatory regions shape the evolutionary fate of proteins. In the same project, we also showed relationship between Kindlin expression pattern and histone modification in Kindlin promoter and showed that although histone modification is conserved in mammals but its conservation profile did not comply with the evolutionary history of either the Kindlin proteins or Kindlin expression pattern. Therefore, evolution of the regulatory sequences might still be considered as the best candidate to explain the evolution of expression divergence as compared to evolution of histone modification pattern. This was also confirmed by our use of phylogenetic shadowing approach through which we found ultra-conserved SRF and SP1 binding sites and showed experimentally their importance in regulating Kindlin-2 expression. Finally, towards our surge for the elucidation of transcriptional regulation of Kindlins in terms of important transcription factors, we found that GLI1, an important Hedgehog signalling molecule and SP1 play important role in the regulation of Kindlin-2 expression. GLI1 story is interesting in the sense that it discovered a novel feedback loop between Kindlin-2 and Gli1 where GLI1 down regulates Kidnlin-2 and Kindlin-2 in turn up regulates GLI1 in prostate cancer cells. This is an important finding and has opened new avenues for studying the different aspects of Kindlin-2 as signalling molecule as it has traditionally been considered as a structural protein. Furthermore, existence of the aforementioned feedback loop in prostate cancer cells also indicates the potential therapeutic implication of our findings. SP1 story on the other hand shows the importance of binding sites of certain transcription factors in the promoter of a gene cannot only control the transcription of it but can also be responsible for the expression profile it exhibits i.e. ubiquitous expression in the case of Kindlin-2. It also reaffirms the potential role SP1 plays in controlling the transcription of housekeeping genes which contains CpG islands in their promoters.

Acknowledgements

I would like to thank all those who have been providing me all sorts of support and help during my PhD.

Accept my gratitude, **Hongquan Zhang** , my main supervisor for accepting me as your PhD student. One of the most moving things in your supervision was that you always welcomed new ideas and always encouraged me to drive my research in my own direction. It is pity that you had to move to China and I could not learn as much from you as any PhD student would wish. I am also thankful to you for showing your warmest hospitality and care when I was in China. It really made my visit pleasant.

Staffan Strömblad, my co-supervisor and director of the SST lab, I thank you too for accepting me as a student to work in your group under your co-supervision. Your eagerness and commitment to conduct research of the highest level was always a model to follow. Your ability to listen to every type of criticism and respond calmly is something always fascinated me and taught me that how we should behave in a professionally working environment.

Takashi Shimokawa, although you were my co-supervisor but you took care of my research even more than a main supervisor can. All the experimental techniques I learned during my PhD, I owe to you and hence a considerable portion of this thesis could have not been accomplished without your guidance and support. Those were perhaps the best days of my PhD when we would sit together in the reading room and talked for hours for our strategies to elucidate the transcriptional regulation of Kindlin-2 and Kindlin-1. I wish you had not gone to Japan so that I could work more under your guidance and hence produce more.

My gratitude to **Karin Dahlman-Wright**, the head of department of Biosciences and Nutrition for providing excellent research facilities in the department and taking the initiatives to make the research environment more interactive and conducive. I am also thankful to Karin for taking care of my administrative matters while my main supervisor was not here.

My utmost gratitude to **Lennart Nilsson**, the director of doctoral education, for being so caring and concerned about the affairs of PhD students. I must also express my personal thanks to Lennart for taking care of my educational as well as financial and administrative matters especially when I was facing scarce situations to carry out the experimental part of my research.

My kindest regards to **Higher Education Commission of Pakistan (HEC)** for rewarding me with PhD scholarship. Without the continuous support of HEC, I could not accomplish this achievement.

Axel Janke, accept my sincere regards for collaborating with me in my very first PhD project that lead to the publication of my first paper. That obviously was a special event for me. And thanks a lot for welcoming and hosting me in your lab where a considerable portion of my third project was accomplished. I am sure we will have more collaborative work in near future.

Raja Hashim Ali, your friendship was perhaps the best find of my PhD. Our collaboration leads to two manuscripts both of which are part of this thesis. But it is not about the number of manuscripts. It is more about the number of nights we spent on talking about our project, on the evolution of our proteins and on discussing about the deepest questions in this regard. But our round 5 is still not finished yet. You know!

There are certain people when you see them, you feel that they are borne to help, to facilitate, to take care of other people. Those who are by default, 911ians, the rescuers. **Helene Olofsson**, you are one of those rare people. I can only say that we were lucky to have you among us. In your absence, SST lab never looked the same. And thanks for all the support, help and guidance you provided to me over the last four years or so.

Marianne Van Dijk, although you came in SST lab when I was *almost* winding up my PhD, still we developed nice friendship which was refreshing on certain times and encouraging and useful on other times. Thanks for all the guidance and help you have been providing to me during the last phase of my PhD.

Tania Costa, Thanks for your help and guidance for the preparation of my seminars and presentations. I truly wish we should have interacted like this way before, for which I blame myself.

John Lock, it was always refreshing to talk with you and to listen to your scientific talk. I wish you to be even more successful scientist in future. Science needs people like you.

An Zhengwen (**Wenny**), Now when I am writing this, it reminds me all those days when you and me would stay in the lab till very late at night and talk about the hard times we both were facing. But that time have passed now. The sun is always there to rise.

Xiaowei Gong, Your experimental skills is something every researcher would love to have. But for me it was our long walks which I would remember more. So let us have another long walk before I leave. May be from Flemensberg to Märsta?

Sylvie Le Guyader Thanks for helping me with your nice comments and suggestion you gave me whole through my PhD. And of course for the nice refreshment you provide on every Wednesday through your great cookie club.

Sara Göransson, Your way of organizing your stuff was something I always tried to follow but would always fail by quite a margin. Your maturity in grabbing the crux behind the concept of a scientific question was another thing I would miss.

Hamdah Shafqat Abbasi, it was really great to have a sincere and caring friend like you. Best of luck with your family life and future career. Hope to see you back in Pakistan soon.

Zhilun it was always a treat to see the display of your skills during every lab meeting. Keep doing it.

Tina, the queen of western blot, how many late night hours we worked together in lab, with almost same sort of stress. But as you know, this stress is not everlasting. So keep moving.

Miao, what a treat it was to see someone who wanted to drive the things in her own way, who had her own ideas and who was so motivated to pursue her scientific career. Keep going!!!

Pablo Hernández, Your critical but useful comments have been very helpful for me. Thanks a lot for that!

Mehrdad Jafari, unfortunately we did not interact much, but whenever, it was quite refreshing.

Anderw Petersson, thanks for all the useful suggestions you have been giving me throughout my PhD.

Ulrich Berge, Jacobs Kowalewski, Gabriela Imreh and **Alexa Kiss** thanks for your nice comments, company and help.

Minna, Steve, Laure, and **Ylva**, the past fellows in SST lab, thanks for the comments and help you have been providing to me whenever I needed.

I wish very best of luck to all my SST lab fellows for their future endeavours.

I am thankful to all my friends with whom I worked during my stay in Beijing, China, especially to **Gao Jinchao** with whom I collaborated for Kindlin-Hhh work.

My gratitude to my friends with whom I worked during my stay in in Frankfurt, Germany, especially to **Vikas Kumar** with whom I collaborated for Kindlin promoter evolution work.

Peter Zaphiropoulos, thanks for your kind suggestions and guidance. **Anthony Wright,** thanks for giving very useful suggestions on my pre-dissertation seminar.

My kindest regards to the people from administration section especially **Monica Ahlberg**, **Lena Magnell**, **Marie Franzen**, and **Eva Nordlander** for providing help and assistance whenever I needed.

Rauf Awan, jolly and refreshing sweeper of our department, I would miss those heated dialogues I underwent with you on the political and social issues of Pakistan.

My thanks to my friends **Alamdar Hussain** and **Shareef Hasni** for providing me all sort of support during my whole PhD. To **Wei Liu** for providing his useful guidance since the very start of my PhD and to **Qie Kuang** for being such a nice friend. To my other friends including **Israr** and **Iram**, **Faryal** and **Faheem**, **Salam**, **Qaiser**, **Inez**, **Shahzad**, **Ivean**, **Irfan**, **Musthtaq** and **Asif** for their great friendship which made my stay in Sweden pleasant and memorable.

And last but never the least, my greatest gratitude to my mother and my late father without whose efforts, training and love, I could never accomplish this achievement.

References

1. Khan AA, Janke A, Shimokawa T, Zhang H: **Phylogenetic analysis of kindlins suggests subfunctionalization of an ancestral unduplicated kindlin into three paralogs in vertebrates**. *Evolutionary bioinformatics online* 2011, **7**:7-19.
2. Gao J, Khan AA, Shimokawa T, Zhan J, Stromblad S, Fang W, Zhang H: **A feedback regulation between Kindlin-2 and GLI1 in prostate cancer cells**. *FEBS letters* 2013.
3. Zhao T, Guan L, Yu Y, Pei X, Zhan J, Han L, Tang Y, Li F, Fang W, Zhang H: **Kindlin-2 promotes genome instability in breast cancer cells**. *Cancer letters* 2013, **330**(2):208-216.
4. Yu Y, Wu J, Wang Y, Zhao T, Ma B, Liu Y, Fang W, Zhu WG, Zhang H: **Kindlin 2 forms a transcriptional complex with beta-catenin and TCF4 to enhance Wnt signalling**. *EMBO reports* 2012, **13**(8):750-758.
5. Kindler T: **Congenital poikiloderma with traumatic bulla formation and progressive cutaneous atrophy**. *The British journal of dermatology* 1954, **66**(3):104-111.
6. Has C, Castiglia D, del Rio M, Diez MG, Piccinni E, Kiritsi D, Kohlhase J, Itin P, Martin L, Fischer J *et al*: **Kindler syndrome: extension of FERMT1 mutational spectrum and natural history**. *Human mutation* 2011, **32**(11):1204-1212.
7. Svensson L, Howarth K, McDowall A, Patzak I, Evans R, Ussar S, Moser M, Metin A, Fried M, Tomlinson I *et al*: **Leukocyte adhesion deficiency-III is caused by mutations in KINDLIN3 affecting integrin activation**. *Nat Med* 2009, **15**(3):306-312.
8. Wang H, Lim D, Rudd CE: **Immunopathologies linked to integrin signalling**. *Seminars in immunopathology* 2010, **32**(2):173-182.
9. Montanez E, Ussar S, Schifferer M, Bosl M, Zent R, Moser M, Fassler R: **Kindlin-2 controls bidirectional signaling of integrins**. *Genes & development* 2008, **22**(10):1325-1330.
10. Zhan J, Zhu X, Guo Y, Wang Y, Wang Y, Qiang G, Niu M, Hu J, Du J, Li Z *et al*: **Opposite role of Kindlin-1 and Kindlin-2 in lung cancers**. *PloS one* 2012, **7**(11):e50313.
11. Shi X, Wu C: **A suppressive role of mitogen inducible gene-2 in mesenchymal cancer cell invasion**. *Molecular cancer research : MCR* 2008, **6**(5):715-724.
12. An Z, Dobra K, Lock JG, Stromblad S, Hjerpe A, Zhang H: **Kindlin-2 is expressed in malignant mesothelioma and is required for tumor cell adhesion and migration**. *International journal of cancer Journal international du cancer* 2010, **127**(9):1999-2008.
13. Gong X, An Z, Wang Y, Guan L, Fang W, Stromblad S, Jiang Y, Zhang H: **Kindlin-2 controls sensitivity of prostate cancer cells to cisplatin-induced cell death**. *Cancer letters* 2010, **299**(1):54-62.
14. Sin S, Bonin F, Petit V, Meseure D, Lallemand F, Bieche I, Bellahcene A, Castronovo V, de Wever O, Gespach C *et al*: **Role of the focal adhesion protein kindlin-1 in breast cancer growth and lung metastasis**. *Journal of the National Cancer Institute* 2011, **103**(17):1323-1337.
15. Shen Z, Ye Y, Kauttu T, Seppanen H, Vainionpaa S, Wang S, Mustonen H, Puolakkainen P: **The novel focal adhesion gene kindlin-2 promotes the invasion of gastric cancer cells mediated by tumor-associated macrophages**. *Oncology reports* 2013, **29**(2):791-797.
16. Shen Z, Ye Y, Dong L, Vainionpaa S, Mustonen H, Puolakkainen P, Wang S: **Kindlin-2: a novel adhesion protein related to tumor invasion, lymph node metastasis, and patient outcome in gastric cancer**. *American journal of surgery* 2012, **203**(2):222-229.
17. Ratzinger S, Grassel S, Dowejko A, Reichert TE, Bauer RJ: **Induction of type XVI collagen expression facilitates proliferation of oral cancer cells**. *Matrix biology : journal of the International Society for Matrix Biology* 2011, **30**(2):118-125.

18. Ussar S, Wang HV, Linder S, Fassler R, Moser M: **The Kindlins: subcellular localization and expression during murine development**. *Experimental cell research* 2006, **312**(16):3142-3151.
19. Khan AA, Shimokawa T, Stromblad S, Zhang H: **Functional characterization of human Kindlin-2 core promoter identifies a key role of SP1 in Kindlin-2 transcriptional regulation**. *Cellular & molecular biology letters* 2011, **16**(4):638-651.
20. Rogalski TM, Mullen GP, Gilbert MM, Williams BD, Moerman DG: **The UNC-112 gene in Caenorhabditis elegans encodes a novel component of cell-matrix adhesion structures required for integrin localization in the muscle cell membrane**. *The Journal of cell biology* 2000, **150**(1):253-264.
21. Dowling JJ, Gibbs E, Russell M, Goldman D, Minarcik J, Golden JA, Feldman EL: **Kindlin-2 is an essential component of intercalated discs and is required for vertebrate cardiac structure and function**. *Circulation research* 2008, **102**(4):423-431.
22. Kruger M, Moser M, Ussar S, Thievessen I, Luber CA, Forner F, Schmidt S, Zanivan S, Fassler R, Mann M: **SILAC mouse for quantitative proteomics uncovers kindlin-3 as an essential factor for red blood cell function**. *Cell* 2008, **134**(2):353-364.
23. Moser M, Nieswandt B, Ussar S, Pozgajova M, Fassler R: **Kindlin-3 is essential for integrin activation and platelet aggregation**. *Nat Med* 2008, **14**(3):325-330.
24. Berman AE, Kozlova NI: **Integrins: structure and functions**. *Membrane & cell biology* 2000, **13**(2):207-244.
25. Calderwood DA, Zent R, Grant R, Rees DJ, Hynes RO, Ginsberg MH: **The Talin head domain binds to integrin beta subunit cytoplasmic tails and regulates integrin activation**. *The Journal of biological chemistry* 1999, **274**(40):28071-28074.
26. Tadokoro S, Shattil SJ, Eto K, Tai V, Liddington RC, de Pereda JM, Ginsberg MH, Calderwood DA: **Talin binding to integrin beta tails: a final common step in integrin activation**. *Science* 2003, **302**(5642):103-106.
27. Petrich BG, Marchese P, Ruggeri ZM, Spiess S, Weichert RA, Ye F, Tiedt R, Skoda RC, Monkley SJ, Critchley DR *et al*: **Talin is required for integrin-mediated platelet function in hemostasis and thrombosis**. *The Journal of experimental medicine* 2007, **204**(13):3103-3111.
28. Nieswandt B, Moser M, Pleines I, Varga-Szabo D, Monkley S, Critchley D, Fassler R: **Loss of talin1 in platelets abrogates integrin activation, platelet aggregation, and thrombus formation in vitro and in vivo**. *The Journal of experimental medicine* 2007, **204**(13):3113-3118.
29. Critchley DR, Gingras AR: **Talin at a glance**. *Journal of cell science* 2008, **121**(Pt 9):1345-1347.
30. Legate KR, Fassler R: **Mechanisms that regulate adaptor binding to beta-integrin cytoplasmic tails**. *Journal of cell science* 2009, **122**(Pt 2):187-198.
31. Vinogradova O, Velyvis A, Velyviene A, Hu B, Haas T, Plow E, Qin J: **A structural mechanism of integrin alpha(IIb)beta(3) "inside-out" activation as regulated by its cytoplasmic face**. *Cell* 2002, **110**(5):587-597.
32. Wegener KL, Partridge AW, Han J, Pickford AR, Liddington RC, Ginsberg MH, Campbell ID: **Structural basis of integrin activation by talin**. *Cell* 2007, **128**(1):171-182.
33. Petrich BG, Fogelstrand P, Partridge AW, Yousefi N, Ablooglu AJ, Shattil SJ, Ginsberg MH: **The antithrombotic potential of selective blockade of talin-dependent integrin alpha IIb beta 3 (platelet GPIIb-IIIa) activation**. *The Journal of clinical investigation* 2007, **117**(8):2250-2259.
34. Chen YP, Djaffar I, Pidard D, Steiner B, Cieutat AM, Caen JP, Rosa JP: **Ser-752-->Pro mutation in the cytoplasmic domain of integrin beta 3 subunit and defective activation of platelet integrin alpha IIb beta 3 (glycoprotein IIb-IIIa) in a variant of Glanzmann thrombasthenia**.

Proceedings of the National Academy of Sciences of the United States of America 1992, **89**(21):10169-10173.

35. Kloeker S, Major MB, Calderwood DA, Ginsberg MH, Jones DA, Beckerle MC: **The Kindler syndrome protein is regulated by transforming growth factor-beta and involved in integrin-mediated adhesion**. *The Journal of biological chemistry* 2004, **279**(8):6824-6833.
36. Shi X, Ma YQ, Tu Y, Chen K, Wu S, Fukuda K, Qin J, Plow EF, Wu C: **The MIG-2/integrin interaction strengthens cell-matrix adhesion and modulates cell motility**. *The Journal of biological chemistry* 2007, **282**(28):20455-20466.
37. Ma YQ, Qin J, Wu C, Plow EF: **Kindlin-2 (Mig-2): a co-activator of beta3 integrins**. *The Journal of cell biology* 2008, **181**(3):439-446.
38. Malinin NL, Zhang L, Choi J, Ciocea A, Razorenova O, Ma YQ, Podrez EA, Tosi M, Lennon DP, Caplan AI *et al*: **A point mutation in KINDLIN3 ablates activation of three integrin subfamilies in humans**. *Nat Med* 2009, **15**(3):313-318.
39. Shattil SJ, Kim C, Ginsberg MH: **The final steps of integrin activation: the end game**. *Nature reviews Molecular cell biology* 2010, **11**(4):288-300.
40. Kahner BN, Kato H, Banno A, Ginsberg MH, Shattil SJ, Ye F: **Kindlins, integrin activation and the regulation of talin recruitment to alphaIIbbeta3**. *PloS one* 2012, **7**(3):e34056.
41. Lock JG, Wehrle-Haller B, Stromblad S: **Cell-matrix adhesion complexes: master control machinery of cell migration**. *Seminars in cancer biology* 2008, **18**(1):65-76.
42. Hynes RO: **Integrins: bidirectional, allosteric signaling machines**. *Cell* 2002, **110**(6):673-687.
43. Harburger DS, Calderwood DA: **Integrin signalling at a glance**. *Journal of cell science* 2009, **122**(Pt 2):159-163.
44. Shen B, Delaney MK, Du X: **Inside-out, outside-in, and inside-outside-in: G protein signaling in integrin-mediated cell adhesion, spreading, and retraction**. *Current opinion in cell biology* 2012, **24**(5):600-606.
45. Hu P, Luo BH: **Integrin bi-directional signaling across the plasma membrane**. *Journal of cellular physiology* 2013, **228**(2):306-312.
46. Schwartz MA, Schaller MD, Ginsberg MH: **Integrins: emerging paradigms of signal transduction**. *Annual review of cell and developmental biology* 1995, **11**:549-599.
47. Legate KR, Wickstrom SA, Fassler R: **Genetic and cell biological analysis of integrin outside-in signaling**. *Genes & development* 2009, **23**(4):397-418.
48. Chishti AH, Kim AC, Marfatia SM, Lutchman M, Hanspal M, Jindal H, Liu SC, Low PS, Rouleau GA, Mohandas N *et al*: **The FERM domain: a unique module involved in the linkage of cytoplasmic proteins to the membrane**. *Trends in biochemical sciences* 1998, **23**(8):281-282.
49. Frame MC, Patel H, Serrels B, Lietha D, Eck MJ: **The FERM domain: organizing the structure and function of FAK**. *Nature reviews Molecular cell biology* 2010, **11**(11):802-814.
50. Hamada K, Shimizu T, Matsui T, Tsukita S, Hakoshima T: **Structural basis of the membrane-targeting and unmasking mechanisms of the radixin FERM domain**. *The EMBO journal* 2000, **19**(17):4449-4462.
51. Liu Y, Zhu Y, Ye S, Zhang R: **Crystal structure of kindlin-2 PH domain reveals a conformational transition for its membrane anchoring and regulation of integrin activation**. *Protein & cell* 2012, **3**(6):434-440.
52. Haslam RJ, Koide HB, Hemmings BA: **Pleckstrin domain homology**. *Nature* 1993, **363**(6427):309-310.
53. Harlan JE, Hajduk PJ, Yoon HS, Fesik SW: **Pleckstrin homology domains bind to phosphatidylinositol-4,5-bisphosphate**. *Nature* 1994, **371**(6493):168-170.

54. Carpten JD, Faber AL, Horn C, Donoho GP, Briggs SL, Robbins CM, Hostetter G, Boguslawski S, Moses TY, Savage S *et al*: **A transforming mutation in the pleckstrin homology domain of AKT1 in cancer**. *Nature* 2007, **448**(7152):439-444.
55. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W *et al*: **Initial sequencing and analysis of the human genome**. *Nature* 2001, **409**(6822):860-921.
56. Liu MJ, Xie M, Mao J, Li H, Yan WH, Chen Y: **[Application of next-generation sequencing technology for genetic diagnosis of Duchenne muscular dystrophy]**. *Zhonghua yi xue yi chuan xue za zhi = Zhonghua yixue yichuanxue zazhi = Chinese journal of medical genetics* 2012, **29**(3):249-254.
57. Zhao J, Grant SF: **Advances in whole genome sequencing technology**. *Current pharmaceutical biotechnology* 2011, **12**(2):293-305.
58. Barrett T, Edgar R: **Gene expression omnibus: microarray data storage, submission, retrieval, and analysis**. *Methods in enzymology* 2006, **411**:352-369.
59. Dermitzakis ET, Clark AG: **Evolution of transcription factor binding sites in Mammalian gene regulatory regions: conservation and turnover**. *Molecular biology and evolution* 2002, **19**(7):1114-1121.
60. Parkinson H, Kapushesky M, Shojatalab M, Abeygunawardena N, Coulson R, Farne A, Holloway E, Kolesnykov N, Lilja P, Lukk M *et al*: **ArrayExpress--a public database of microarray experiments and gene expression profiles**. *Nucleic acids research* 2007, **35**(Database issue):D747-750.
61. Maston GA, Evans SK, Green MR: **Transcriptional regulatory elements in the human genome**. *Annual review of genomics and human genetics* 2006, **7**:29-59.
62. Juven-Gershon T, Hsu JY, Theisen JW, Kadonaga JT: **The RNA polymerase II core promoter - the gateway to transcription**. *Current opinion in cell biology* 2008, **20**(3):253-259.
63. Thomas MC, Chiang CM: **The general transcription machinery and general cofactors**. *Critical reviews in biochemistry and molecular biology* 2006, **41**(3):105-178.
64. Orphanides G, Lagrange T, Reinberg D: **The general transcription factors of RNA polymerase II**. *Genes & development* 1996, **10**(21):2657-2683.
65. Conaway JW, Florens L, Sato S, Tomomori-Sato C, Parmely TJ, Yao T, Swanson SK, Banks CA, Washburn MP, Conaway RC: **The mammalian Mediator complex**. *FEBS letters* 2005, **579**(4):904-908.
66. Malik S, Roeder RG: **Dynamic regulation of pol II transcription by the mammalian Mediator complex**. *Trends in biochemical sciences* 2005, **30**(5):256-263.
67. Hahn S: **Structure and mechanism of the RNA polymerase II transcription machinery**. *Nature structural & molecular biology* 2004, **11**(5):394-403.
68. Carninci P, Sandelin A, Lenhard B, Katayama S, Shimokawa K, Ponjavic J, Semple CA, Taylor MS, Engstrom PG, Frith MC *et al*: **Genome-wide analysis of mammalian promoter architecture and evolution**. *Nature genetics* 2006, **38**(6):626-635.
69. Coulondre C, Miller JH, Farabaugh PJ, Gilbert W: **Molecular basis of base substitution hotspots in Escherichia coli**. *Nature* 1978, **274**(5673):775-780.
70. Tweedie S, Charlton J, Clark V, Bird A: **Methylation of genomes and genes at the invertebrate-vertebrate boundary**. *Molecular and cellular biology* 1997, **17**(3):1469-1475.
71. Yamashita R, Suzuki Y, Sugano S, Nakai K: **Genome-wide analysis reveals strong correlation between CpG islands with nearby transcription start sites of genes and their tissue specificity**. *Gene* 2005, **350**(2):129-136.

72. Illingworth RS, Gruenewald-Schneider U, Webb S, Kerr AR, James KD, Turner DJ, Smith C, Harrison DJ, Andrews R, Bird AP: **Orphan CpG islands identify numerous conserved promoters in the mammalian genome**. *PLoS genetics*, **6**(9).
73. Guenther MG, Levine SS, Boyer LA, Jaenisch R, Young RA: **A chromatin landmark and transcription initiation at most promoters in human cells**. *Cell* 2007, **130**(1):77-88.
74. Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M: **An integrated encyclopedia of DNA elements in the human genome**. *Nature*, **489**(7414):57-74.
75. Weinmann AS, Yan PS, Oberley MJ, Huang TH, Farnham PJ: **Isolating human transcription factor targets by coupling chromatin immunoprecipitation and CpG island microarray analysis**. *Genes & development* 2002, **16**(2):235-244.
76. Fouse SD, Shen Y, Pellegrini M, Cole S, Meissner A, Van Neste L, Jaenisch R, Fan G: **Promoter CpG methylation contributes to ES cell gene regulation in parallel with Oct4/Nanog, PcG complex, and histone H3 K4/K27 trimethylation**. *Cell stem cell* 2008, **2**(2):160-169.
77. Weber M, Hellmann I, Stadler MB, Ramos L, Paabo S, Rebhan M, Schubeler D: **Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome**. *Nature genetics* 2007, **39**(4):457-466.
78. Schaefer CB, Ooi SK, Bestor TH, Bourc'his D: **Epigenetic decisions in mammalian germ cells**. *Science* 2007, **316**(5823):398-399.
79. Robertson KD: **DNA methylation and human disease**. *Nature reviews Genetics* 2005, **6**(8):597-610.
80. Kornberg RD, Lorch Y: **Twenty-five years of the nucleosome, fundamental particle of the eukaryote chromosome**. *Cell* 1999, **98**(3):285-294.
81. Reik W, Dean W, Walter J: **Epigenetic reprogramming in mammalian development**. *Science* 2001, **293**(5532):1089-1093.
82. Bird A: **DNA methylation patterns and epigenetic memory**. *Genes & development* 2002, **16**(1):6-21.
83. Goll MG, Bestor TH: **Eukaryotic cytosine methyltransferases**. *Annual review of biochemistry* 2005, **74**:481-514.
84. Morgan HD, Santos F, Green K, Dean W, Reik W: **Epigenetic reprogramming in mammals**. *Human molecular genetics* 2005, **14 Spec No 1**:R47-58.
85. Grewal SI, Moazed D: **Heterochromatin and epigenetic control of gene expression**. *Science* 2003, **301**(5634):798-802.
86. Goldberg AD, Allis CD, Bernstein E: **Epigenetics: a landscape takes shape**. *Cell* 2007, **128**(4):635-638.
87. Ringrose L, Paro R: **Epigenetic regulation of cellular memory by the Polycomb and Trithorax group proteins**. *Annual review of genetics* 2004, **38**:413-443.
88. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, Wang Z, Wei G, Chepelev I, Zhao K: **High-resolution profiling of histone methylations in the human genome**. *Cell* 2007, **129**(4):823-837.
89. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA *et al*: **Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome**. *Nature genetics* 2007, **39**(3):311-318.
90. Consortium EP, Dunham I, Kundaje A, Aldred SF, Collins PJ, Davis CA, Doyle F, Epstein CB, Frietze S, Harrow J *et al*: **An integrated encyclopedia of DNA elements in the human genome**. *Nature* 2012, **489**(7414):57-74.
91. Sanyal A, Lajoie BR, Jain G, Dekker J: **The long-range interaction landscape of gene promoters**. *Nature* 2012, **489**(7414):109-113.

92. Crawford GE, Holt IE, Whittle J, Webb BD, Tai D, Davis S, Margulies EH, Chen Y, Bernat JA, Ginsburg D *et al*: **Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS)**. *Genome research* 2006, **16**(1):123-131.
93. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B *et al*: **The accessible chromatin landscape of the human genome**. *Nature* 2012, **489**(7414):75-82.
94. Neph S, Vierstra J, Stergachis AB, Reynolds AP, Haugen E, Vernot B, Thurman RE, John S, Sandstrom R, Johnson AK *et al*: **An expansive human regulatory lexicon encoded in transcription factor footprints**. *Nature* 2012, **489**(7414):83-90.
95. Consortium EP, Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET *et al*: **Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project**. *Nature* 2007, **447**(7146):799-816.
96. Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, Cheng C, Mu XJ, Khurana E, Rozowsky J, Alexander R *et al*: **Architecture of the human regulatory network derived from ENCODE data**. *Nature* 2012, **489**(7414):91-100.
97. Ecker JR, Bickmore WA, Barroso I, Pritchard JK, Gilad Y, Segal E: **Genomics: ENCODE explained**. *Nature* 2012, **489**(7414):52-55.
98. Alberts B: **Molecular biology of the cell**, 4th edn. New York: Garland Science; 2002.
99. Morris SC: **Early metazoan evolution: Reconciling paleontology and molecular biology**. *American Zoologist* 1998, **38**(6):867-877.
100. Morris SC: **Evolution: Bringing Molecules into the Fold**. *Cell* 2000, **100**(1):1-11.
101. Darwin C: **On the Origin of Species by means of Natural Selection, or the preservation of favoured races in the struggle for life**. London: John Murray; 1859.
102. Kimura M: **The neutral theory of molecular evolution**. Cambridge [Cambridgeshire] ; New York: Cambridge University Press; 1983.
103. Graur D, Li W-H, Li W-HFome: **Fundamentals of molecular evolution**, 2nd ed. / Dan Graur, Wen-Hsiung Li. edn. Sunderland, Mass.: Sinauer Associates; 2000.
104. Kimura M: **Evolutionary rate at the molecular level**. *Nature* 1968, **217**(5129):624-626.
105. Pertea M, Salzberg SL: **Between a chicken and a grape: estimating the number of human genes**. *Genome biology* 2010, **11**(5):206.
106. Graur D, Li W-H: **Fundamentals of molecular evolution**, 2nd edn. Sunderland, Mass.: Sinauer Associates; 2000.
107. Gonzalez P, Rao PV, Nunez SB, Zigler JS, Jr.: **Evidence for independent recruitment of zeta-crystallin/quinone reductase (CRYZ) as a crystallin in camelids and hystricomorph rodents**. *Molecular biology and evolution* 1995, **12**(5):773-781.
108. Shashikant CS, Kim CB, Borbely MA, Wang WC, Ruddle FH: **Comparative studies on mammalian Hoxc8 early enhancer sequence reveal a baleen whale-specific deletion of a cis-acting element**. *Proceedings of the National Academy of Sciences of the United States of America* 1998, **95**(26):15446-15451.
109. Rockman MV, Wray GA: **Abundant raw material for cis-regulatory evolution in humans**. *Molecular biology and evolution* 2002, **19**(11):1991-2004.
110. Streelman JT, Kocher TD: **Microsatellite variation associated with prolactin expression and growth of salt-challenged tilapia**. *Physiological genomics* 2002, **9**(1):1-4.
111. Ludwig MZ, Kreitman M: **Evolutionary dynamics of the enhancer region of even-skipped in Drosophila**. *Molecular biology and evolution* 1995, **12**(6):1002-1011.
112. Trefilov A, Berard J, Krawczak M, Schmidtke J: **Natal dispersal in rhesus macaques is related to serotonin transporter gene promoter variation**. *Behavior genetics* 2000, **30**(4):295-301.

113. Long M, Wang W, Zhang J: **Origin of new genes and source for N-terminal domain of the chimerical gene, jingwei, in Drosophila**. *Gene* 1999, **238**(1):135-141.
114. Wang W, Brunet FG, Nevo E, Long M: **Origin of sphinx, a young chimeric RNA gene in Drosophila melanogaster**. *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**(7):4448-4453.
115. Korneev S, O'Shea M: **Evolution of nitric oxide synthase regulatory genes by DNA inversion**. *Molecular biology and evolution* 2002, **19**(8):1228-1233.
116. Li X, Noll M: **Evolution of distinct developmental functions of three Drosophila genes by acquisition of different cis-regulatory regions**. *Nature* 1994, **367**(6458):83-87.
117. Gu Z, Nicolae D, Lu HH, Li WH: **Rapid divergence in expression between duplicate genes inferred from microarray data**. *Trends in genetics : TIG* 2002, **18**(12):609-613.
118. Stauber M, Prell A, Schmidt-Ott U: **A single Hox3 gene with composite bicoid and zerknullt expression characteristics in non-Cyclorrhaphan flies**. *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**(1):274-279.
119. Giordano M, Marchetti C, Chiorboli E, Bona G, Momigliano Richiardi P: **Evidence for gene conversion in the generation of extensive polymorphism in the promoter of the growth hormone gene**. *Human genetics* 1997, **100**(2):249-255.
120. Cereb N, Yang SY: **The regulatory complex of HLA class I promoters exhibits locus-specific conservation with limited allelic variation**. *Journal of immunology* 1994, **152**(8):3873-3883.
121. Swalla BJ, Jeffery WR: **Requirement of the Manx gene for expression of chordate features in a tailless ascidian larva**. *Science* 1996, **274**(5290):1205-1208.
122. Kim J, Kerr JQ, Min GS: **Molecular heterochrony in the early development of Drosophila**. *Proceedings of the National Academy of Sciences of the United States of America* 2000, **97**(1):212-216.
123. Skaer N, Pistillo D, Simpson P: **Transcriptional heterochrony of scute and changes in bristle pattern between two closely related species of blowfly**. *Developmental biology* 2002, **252**(1):31-45.
124. Schiff NM, Feng Y, Quine JA, Krasney PA, Cavener DR: **Evolution of the expression of the Gld gene in the reproductive tract of Drosophila**. *Molecular biology and evolution* 1992, **9**(6):1029-1049.
125. Abzhanov A, Kaufman TC: **Crustacean (malacostracan) Hox genes and the evolution of the arthropod trunk**. *Development* 2000, **127**(11):2239-2249.
126. Brunetti CR, Selegue JE, Monteiro A, French V, Brakefield PM, Carroll SB: **The generation and diversification of butterfly eyespot color patterns**. *Current biology : CB* 2001, **11**(20):1578-1585.
127. Scemama JL, Hunter M, McCallum J, Prince V, Stellwag E: **Evolutionary divergence of vertebrate Hoxb2 expression patterns and transcriptional regulatory loci**. *The Journal of experimental zoology* 2002, **294**(3):285-299.
128. Crawford DL, Segal JA, Barnett JL: **Evolutionary analysis of TATA-less proximal promoter function**. *Molecular biology and evolution* 1999, **16**(2):194-207.
129. Jin W, Riley RM, Wolfinger RD, White KP, Passador-Gurgel G, Gibson G: **The contributions of sex, genotype and age to transcriptional variance in Drosophila melanogaster**. *Nature genetics* 2001, **29**(4):389-395.
130. Schadt EE, Monks SA, Drake TA, Lusis AJ, Che N, Colinayo V, Ruff TG, Milligan SB, Lamb JR, Cavet G *et al*: **Genetics of gene expression surveyed in maize, mouse and man**. *Nature* 2003, **422**(6929):297-302.

131. Gerhart J, Kirschner M: **Cells, embryos, and evolution : toward a cellular and developmental understanding of phenotypic variation and evolutionary adaptability**. Malden, Mass. ; Oxford: Blackwell Science; 1997.
132. Carroll SB, Grenier JK, Weatherbee SD: **From DNA to diversity : molecular genetics and the evolution of animal design**, 2nd ed. edn. Oxford: Blackwell; 2005.
133. Kopp A, Duncan I, Godt D, Carroll SB: **Genetic control and evolution of sexually dimorphic characters in Drosophila**. *Nature* 2000, **408**(6812):553-559.
134. Duboule D, Wilkins AS: **The evolution of 'bricolage'**. *Trends in genetics : TIG* 1998, **14**(2):54-59.
135. Gibson G, Honeycutt E: **The evolution of developmental regulatory pathways**. *Current opinion in genetics & development* 2002, **12**(6):695-700.
136. Dawes R, Dawson I, Falciani F, Tear G, Akam M: **Dax, a locust Hox gene related to fushi-tarazu but showing no pair-rule expression**. *Development* 1994, **120**(6):1561-1572.
137. Falciani F, Hausdorf B, Schroder R, Akam M, Tautz D, Denell R, Brown S: **Class 3 Hox genes in insects and the origin of zen**. *Proceedings of the National Academy of Sciences of the United States of America* 1996, **93**(16):8479-8484.
138. Grbic M, Nagy LM, Carroll SB, Strand M: **Polyembryonic development: insect pattern formation in a cellularized environment**. *Development* 1996, **122**(3):795-804.
139. Mouchel-Vielh E, Blin M, Rigolot C, Deutsch JS: **Expression of a homologue of the fushi tarazu (ftz) gene in a cirripede crustacean**. *Evolution & development* 2002, **4**(2):76-85.
140. Lowe CJ, Wray GA: **Radical alterations in the roles of homeobox genes during echinoderm evolution**. *Nature* 1997, **389**(6652):718-721.
141. Saccone G, Peluso I, Artiaco D, Giordano E, Bopp D, Polito LC: **The Ceratitis capitata homologue of the Drosophila sex-determining gene sex-lethal is structurally conserved, but not sex-specifically regulated**. *Development* 1998, **125**(8):1495-1500.
142. Keys DN, Lewis DL, Selegue JE, Pearson BJ, Goodrich LV, Johnson RL, Gates J, Scott MP, Carroll SB: **Recruitment of a hedgehog regulatory circuit in butterfly eyespot evolution**. *Science* 1999, **283**(5401):532-534.
143. Burke AC, Nelson CE, Morgan BA, Tabin C: **Hox genes and the evolution of vertebrate axial morphology**. *Development* 1995, **121**(2):333-346.
144. Averof M, Patel NH: **Crustacean appendage evolution associated with changes in Hox gene expression**. *Nature* 1997, **388**(6643):682-686.
145. Stern DL: **A role of Ultrabithorax in morphological differences between Drosophila species**. *Nature* 1998, **396**(6710):463-466.
146. Beldade P, Brakefield PM, Long AD: **Contribution of Distal-less to quantitative variation in butterfly eyespots**. *Nature* 2002, **415**(6869):315-318.
147. Manzanares M, Wada H, Itasaki N, Trainor PA, Krumlauf R, Holland PW: **Conservation and elaboration of Hox gene regulation during evolution of the vertebrate head**. *Nature* 2000, **408**(6814):854-857.
148. Conlon FL, Fairclough L, Price BM, Casey ES, Smith JC: **Determinants of T box protein specificity**. *Development* 2001, **128**(19):3749-3758.
149. D'Elia AV, Tell G, Paron I, Pellizzari L, Lonigro R, Damante G: **Missense mutations of human homeoboxes: A review**. *Human mutation* 2001, **18**(5):361-374.
150. Brickman JM, Clements M, Tyrell R, McNay D, Woods K, Warner J, Stewart A, Beddington RS, Dattani M: **Molecular effects of novel mutations in Hesx1/HESX1 associated with human pituitary disorders**. *Development* 2001, **128**(24):5189-5199.

151. Lutz B, Lu HC, Eichele G, Miller D, Kaufman TC: **Rescue of Drosophila labial null mutant by the chicken ortholog Hoxb-1 demonstrates that the function of Hox genes is phylogenetically conserved**. *Genes & development* 1996, **10**(2):176-184.
152. Dawson SJ, Morris PJ, Latchman DS: **A single amino acid change converts an inhibitory transcription factor into an activator**. *The Journal of biological chemistry* 1996, **271**(20):11631-11633.
153. Galant R, Carroll SB: **Evolution of a transcriptional repression domain in an insect Hox protein**. *Nature* 2002, **415**(6874):910-913.
154. Ronshaugen M, McGinnis N, McGinnis W: **Hox protein mutation and macroevolution of the insect body plan**. *Nature* 2002, **415**(6874):914-917.
155. Avila S, Casero MC, Fernandez-Canton R, Sastre L: **Transactivation domains are not functionally conserved between vertebrate and invertebrate serum response factors**. *European journal of biochemistry / FEBS* 2002, **269**(15):3669-3677.
156. Enard W, Przeworski M, Fisher SE, Lai CS, Wiebe V, Kitano T, Monaco AP, Paabo S: **Molecular evolution of FOXP2, a gene involved in speech and language**. *Nature* 2002, **418**(6900):869-872.
157. Lichtarge O, Bourne HR, Cohen FE: **An evolutionary trace method defines binding surfaces common to protein families**. *Journal of molecular biology* 1996, **257**(2):342-358.
158. Lua RC, Lichtarge O: **PyETV: a PyMOL evolutionary trace viewer to analyze functional site predictions in protein complexes**. *Bioinformatics* 2010, **26**(23):2981-2982.
159. Brudno M, Steinkamp R, Morgenstern B: **The CHAOS/DIALIGN WWW server for multiple alignment of genomic sequences**. *Nucleic acids research* 2004, **32**(Web Server issue):W41-44.
160. Comeron JM: **K-Estimator: calculation of the number of nucleotide substitutions per site and the confidence intervals**. *Bioinformatics* 1999, **15**(9):763-764.
161. Takai D, Jones PA: **Comprehensive analysis of CpG islands in human chromosomes 21 and 22**. *Proceedings of the National Academy of Sciences of the United States of America* 2002, **99**(6):3740-3745.
162. Takai D, Jones PA: **The CpG island searcher: a new WWW resource**. *In silico biology* 2003, **3**(3):235-240.
163. Kimura M: **A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences**. *Journal of molecular evolution* 1980, **16**(2):111-120.
164. Bernstein BE, Kamal M, Lindblad-Toh K, Bekiranov S, Bailey DK, Huebert DJ, McMahon S, Karlsson EK, Kulbokas EJ, 3rd, Gingeras TR *et al*: **Genomic maps and comparative analysis of histone modifications in human and mouse**. *Cell* 2005, **120**(2):169-181.
165. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K *et al*: **A bivalent chromatin structure marks key developmental genes in embryonic stem cells**. *Cell* 2006, **125**(2):315-326.
166. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, Alvarez P, Brockman W, Kim TK, Koche RP *et al*: **Genome-wide maps of chromatin state in pluripotent and lineage-committed cells**. *Nature* 2007, **448**(7153):553-560.