Characterization of the Protein Encoded by KIAA0319 - a Dyslexia Candidate Gene

CHAN, Hoi Ling

A Thesis Submitted in Partial Fulfilment
of the Requirements for the Degree of
Doctor of Philosophy
in
Molecular Biotechnology

The Chinese University of Hong Kong
April 2010

Thesis Committee

Professor Stephen K.W. Tsui (Chair)
Professor Mary M. Y. Waye (Thesis Supervisor)
Professor Edwin H.Y. Chan (Committee Member)
Professor Robert Z. Qi (External Examiner)

Abstract of thesis entitled:

Characterization of the Protein Encoded by KIAA0319 - a Dyslexia Candidate Gene

Submitted by Chan, Hoi Ling for the degree of Doctor of Philosophy in Molecular Biotechnology at The Chinese University of Hong Kong (January 2009)

Developmental Dyslexia (DD) refers to a reading disorder affecting individuals that possess otherwise normal intelligence. Having demonstrated by familial and twin studies, genetic factors are found to be of major significance to DD development. A strong dyslexia susceptibility gene KIAA0319 (K), of which crucial role in DD had been revealed by various linkage and association studies, was found to have 40% reduction in expression in the DD risk haplotype. Besides, both up- and down-regulation of K would result in impaired neuronal migration in rat. Despite the undoubtedly strong linkage of K to DD, biological and molecular knowledge of K is still lacking. Consequently, how K plays its role in DD remains unclear. To address this question, investigations of human K protein and its interactions in molecular level were performed. K protein is a large transmembrane protein which consists of four main parts, including the N-terminus of K which has a MANSC domain downstream of the signal peptide, a large cluster of five PKD domains in the middle of the protein sequence, a Cysteine -rich C6 region together with a transmembrane domain which had been demonstrated to be critical for forming K protein homodimer, and the only cytoplasmic C-terminus of K. Having shown that no gross effect on gene expression at both mRNA and protein level was found with overexpressing K by DNA microarray and two-dimensional gel electrophoresis, protein interactions involving K were targeted for investigation. Towards this goal, a

monoclonal antibody against K was raised, which is capable for recognizing native full-length K proteins in immunoblotting, indirect immunofluorescence staining, as well as in immunoprecipitation. A novel K interaction partner protein KIAA0319-Like (KL), which is a homologous protein of K with high sequence similarity (59%), has been found and confirmed by co-immunoprecipitation. No interaction was shown for truncation mutants of Cysteine-rich C6 region in either K or KL proteins, cuing that the interaction of K and KL at C6 region is a mimic of K homodimer, and led to a hypothesis that the function of K is regulated by KL, which serves as a molecular control of neuronal migration by regulating the formation of K dimer. Another known interaction partner of K protein, the μ–subunit of Adaptor protein 2 complex (AP2M1) which binds to cytoplasmic C-terminus of K (55% similarity to that of KL), was found to have similar binding behaviour towards K as well as KL by co-immunoprecipitation and molecular docking. In addition to AP2M1, two adaptor proteins FEM and SH2 were also confirmed to be interacting with cytoplasmic C-terminus of K, suggested that cytoplasmic region of K is responsible for interactions of downstream cellular pathways. Interaction of K with adaptor proteins also suggested that K might be a membrane receptor that mediates signalling *via* various adapter proteins. The N-terminus of K protein which has the least sequence similarity to KL (31%) is hence thought to confer to the specificity of the receptor and is critical to the function of K in DD.

中文摘要

發展閱讀困難（DD）是一種學習障礙，患者擁有正常智力但閱讀能力和寫作能力卻與常人有較大差距。家族和孿生研究顯示，遺傳因素是 DD 的主要病因。各種各樣的疾病與標識基因之相關性研究顯露了一個最佳的閱讀困難候選基因 KIAA0319（K）在發展閱讀困難扮演一個關鍵的角色。研究並且發現 K 的表現量在 DD 風險單體型中減少了 40%。另外，增加或減少 K 的表現量會造成大鼠神經細胞的遷移改變。儘管 K 對 DD 非常重要，其生物和分子知識仍然缺乏，所以 K 在 DD 扮演的角色仍然不明。因此，我們進行了人的 K 蛋白質和它在分子水平的蛋白質交互作用的實驗。 K 蛋白質是包括四主要部分的膜蛋白，包括在 N 端（有信號肽和 MANSC 結構域），在蛋白質序列中間有 5 個 PKD 結構域，一個對形成 K 二聚體非常關鍵的富有半胱氨酸的地區（C6 region 和跨細胞膜結構域）以及 K 唯一的細胞質 C 端。由 DNA 微陣列和二維膠凝體電泳法實驗顯示增加 K 的表現量對基因表達在 mRNA 和蛋白質水平上沒有顯著的影響，因此我們瞄準調查 K 的蛋白質交互作用。我們建立了一個會分泌單株抗體的 hybridoma 細胞系，那些單株抗體可在免疫印跡，間接免疫螢光法及免疫沉澱辨識人的 K 蛋白質。我們藉由免疫共沉澱證實了一個新的 K 交互作用夥伴蛋白 KL，是與 K 蛋白有高序列相似性（59%）的同源蛋白質。 刪除突變體（truncation mutants）顯示 K 與 KL 的 蛋白質交互作用發生在富有半胱氨酸的 C6 region，暗示這交互作用是在模仿由兩個 K 單體分子形成的 K 二聚體。這導致了 K 的作用是由 KL 調控的假說：KL 通過調控 K 二聚體的形成來控制神經細胞的遷移。另外，免疫共沉澱及對接顯示 K 蛋白質的另一個交互作用夥伴，AP2M1，對 K 及 KL 有相似的結合作用。除 AP2M1 之外，兩個接頭蛋白 FEM 和 SH2 也被證實與 K 的細胞質 C 端結合，使人聯想到 K 的細胞質區域大概負責下游通路的交互作用。從 K 與接頭蛋白的交互作用可以推測 K 也許是受體蛋白質（receptor），通過各種各樣的接頭蛋白傳遞分子訊息。而與 KL 有最少序列相似性的 K 蛋白質 N 端（31%）決定 K 受體蛋白質的特異性，因此對 K 在 DD 的作用至關重要。

Acknowledgements

I would like to gratefully acknowledge the enthusiastic supervision of Professor Mary M. Y. Waye during this work. I would not be able to complete this research without her invaluable suggestions, patience, kindness and support from the initial to the final level. I want to express my gratitude to Professor Stephen K.W. Tsui for inspiration and the technical discussions on my project, especially after my graduate seminars when I was desperate for suggestions. Thanks for my thesis assessment committee members for spending their precious time on reading this thesis and giving valuable comments. They are Professor Stephen K.W. Tsui, Professor Edwin H.Y. Chan and Professor Robert Z. Qi in The Hong Kong University of Science & Technology. I would also like to thank Prof Chui Yiu Loon for kindly providing the isotyping antibodies used in this study.

I am grateful to all my friends in Mong Man Wai Building Rm. 610, which include Miss Sheila Li, Miss Fion Fong, Miss Cecelia Leung, Dr Patrick Law, Mr. Cadmon Lim, Miss Venus Yeung, Dr Winnie Poon, Dr So Yeung, Miss Candy Li, Dr Ming Poon, Mr Kit Wong and Dr Patrick Ng for their encouragement, technical advices and effort in making a friendly and warm research environment. I especially thank Ceci, Fion and Sheila, for their generosity and emotional support that have kept me working. Dr Patrick Law and Mr. Cadmon Lim are thanked for numerous stimulating discussions, helps in improving my research techniques and general advice. In particular I would like to acknowledge the help of Miss Venus Yeung for her support and assistance - at all times. I would also like to thank Dr Winnie Poon and Dr Ming Poon for their support. Besides,

I would like to show my gratitude to Miss Pui Shan Tse, Miss Kiwi Lam and Miss Yo Yeung for the willingness to help throughout my study.

I thank all my dearest friends for being the surrogate family on top of their company and encouragements in days of tension and hardness. I specially thank Eric Chan for his support to tackle all types of technical problem in the project, the frequent technical discussions and help with experimental setup. Eric is especially thanked for his emotional support, understanding, endless patience, attention, care and encouragements when it was most required - plus the sharing of laughter, tears and memories.

I am forever indebted to my family members for their care and support. I dedicate my thesis to my mother who has provided unconditional love and financial support throughout my undergraduate and postgraduate studies.

Lastly, I offer my regards and blessings to all of those who supported me in any respect during the completion of the project.

Vivian Chan

Table of Contents

CHAPTER 3 Study of gene expression profile of cells overexpressing KIAA0319

CHAPTER 4 Establishment and characterization of specific anti-KIAA0319 antibodies

CHAPTER 5 Interaction of KIAA0319 and its homologue KIAA0319-Like

Abbreviations

Abbreviation	Full name
1st D	first dimension
2D gel	Two-dimensional gel electrophoresis
2nd D	second dimension
Ab	Antibody
AP2	adaptor protein 2 complex
AP2M1	μ-subunit of the Adaptor protein 2 complex
BLAST	Basic Local Alignment Search Tool
bp	base pair(s)
BSA	Bovine Serum Albumin
C6	region of the protein included 6 cysteine residues
CA3	Cornu Ammonis region 3
CBB	Coomassie Brilliant Blue
cDNA	complementary DNA
CO2	carbon dioxide
Co-IP	Co-immunoprecipitation
CP	cortical plate
Ct	threshold cycle
Cys	Cysteines
Da	Dalton
DAPI	4',6-diamidino-2-phenylindole
DD	Developmental Dyslexia
ddCt	delta-delta Ct
DEPC	Diethylpyrocarbonate
DMEM	Dulbeccos Modified Eagle's Medium
DMSO	Dimethyl Sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleoside triphosphate
DTT	dithiothreitol
E. coli	*Escherichia coli*
EDAC	N-(3-Dimethylaminopropyl)-N-ethylcarbodiimide hydrochloride
EGFP	enhanced green fluorescence protein
ELISA	Enzyme-linked immunosorbent assay
ER	endoplasmic reticulum
EtBr	ethidium bromide
FBS	Foetal bovine serum
FBS	Fetal bovine serum
FEM	Human homologue of the Caenorhabditis elegans sex determination

	fem1 protein
G4	G4 monoclonal antibody
GFP	green fluorescence protein
GFP	Green Fluorescent Protein
GST	Glutathione-S-transferase
HA	hemagglutinin
HAT	Hypoxanthine aminopterin thymidine
HAT-1	hepatocyte growth factor activator inhibitor-1
HEK	human embryonic kidney
HGFA	hepatocyte growth factor activator
HGPRT	Hypoxanthine guanine phosphoribosyl transferase
HPLC	High performance liquid chromatography
HRP	horse radish peroxidase
HT	hypoxanthine thymidine
IAA	iodoacetamide
IEF	isoelectric focusing
IF	immunofluorescence
IgG	Immunoglobulin G
IHC	immunohistochemistry
IP	immunoprecipitation
IPG	Immobilized pH gradient
JAK	Janus tyrosine kinase
K	KIAA0319
kDa	kilo Dalton
KL	KIAA0319-Like
KLH	keyhole limpet hemocyanin
LB	Luria broth
Mab	Monoclonal antibody
MANSC	Motif at N terminus with seven Cysteines
mB	mouse B polyserum
MCS	multiple cloning site
MOPS	3-(N'-Morpholino)propanesulfonic acid
mRNA	messenger RNA
MTT	3-(4,5-dimethylthiazol-2-yl)-diphenyl tetrazolium bromide
MTT	3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide
MW	molecular weight
NCBI	National Center for Biotechnology Information
NF	neurofilament
NGF	nerve growth factor
NHS	N-Hydroxysuccinimide
NTC	no template control

OD	optical density
OD	optical density
OPA	one-phor-all (buffer)
ORF	open reading frame
ORF	open reading frame
p124	peptide 124
p663	peptide 663
Pab	Polyclonal antibody
PBGD	porphobilinogen deaminase
PBS	Phosphate buffered saline
PCR	polymerase chain reaction
PEG	Polyethylene glycol
PI	propidium iodide
pI	isoelectric points
PKD	polycystic kidney disease
PNGase F	Peptide:N-Glycosydase F
PNPP	p-Nitrophenyl phosphate
PS	penicillin streptomycin
PVDF	polyvinylidence fluoride
RNA	ribonucleic acid
RNase	ribonuclease
RQ	Relative quantitations
RT-PCR	reverse transcription polymerase chain reaction
SDS	sodium dodecyl sulphate
SDS-PAGE	SDS-polyacrylamide gel electrophoresis
SH2	SH2B adaptor protein 1
shRNA	small hairpin RNA
SNPs	single nucleotide polymorphisms
TAE	Tris-acetate-EDTA
TBST	tris-buffered saline-Tween
TEMED	N,N,N',N',-tetramethylethylenediamine
TGN	trans- Golgi network
TM	transmembrane
UV	ultra violet
VZ	ventricular zone
WB	Western Blot
Y2H	yeast two-hybrid

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1. Developmental dyslexia associates with brain abnormalities

Developmental dyslexia (DD), also known as specific reading disabilities, is of unknown neurobiological cause that remains one of the most prevalent types of learning disabilities (Lyon et al. 2003). It is a specific language-based disorder of constitutional origin characterized by difficulties with accurate and/or fluent word recognition, as well as poor spelling and word-decoding abilities unexplainable by any kind of deficit in general intelligence, learning opportunity, general motivation or sensory acuity (Lyon et al. 2003; Dennis et al. 2009; Cope et al. 2005; Chan et al. 2007; Couto et al. 2008). Several cognitive and perceptual changes appear to associate with DD including changes in short-term memory, occulomotor skills, visuo-spatial abilities; sensory processing, semantic encoding, integration of letter and speech sounds, and phonological processing (Gabel et al. 2009).

DD is one of the most common neurobehavioral disorders affecting children, with prevalence rates up to 5-17.5% in school-age children worldwide (Dennis et al. 2009; Paracchini et al. 2008) and about 9.7-12.6% of children in Hong Kong (Chan 2008; Chan et al. 2007). It is a phenotypically complex developmental disorder with gender inequality. A male to female gender ratio of 2:1 was found in a research of Dutch schoolchildren (Kovel et al. 2004) and in HK (Chan et al. 2007).

Subtle structural anomalies were found in dyslexic brains, which provided evidence that although independent of intelligence levels, some children's learning problems are brain based. Diffusion tensor and functional magnetic resonance imaging demonstrated that there are reduced activation of cortical regions in dyslexics when engage in reading-related tasks (Silani et al. 2005; Perfetti et al. 2008; Shaywitz 1998; Meng et al. 2005) and unusual deactivations in left posterior temporal/inferior parietal areas (Habib Michel 2000). Also, neuroimaging studies have revealed altered density of grey and white matter of specific brain regions, such as the left middle and inferior temporal gyri and the left arcuate fasciculus in dyslexics (Silani et al. 2005; Perfetti et al. 2008), which provided evidence of an imperfectly functioning system in dyslexics.

1.2. Genetic factors are of major aetiological significance of developmental dyslexia

A range of neurobiological investigations has shown the disruption of neural systems for reading in DD that cross languages and cultures, showing that environmental factors may play a role. Dyslexic readers of English have decreased gray-matter volume in posterior brain systems whilst Chinese dyslexic show reduced gray matter volume in a left middle frontal gyrus region that previously shown to be important for Chinese reading and writing, suggested that there may be a different structural basis for reading impairment in cultures with alphabetic and logographic languages (Perfetti et al. 2008).

Besides environmental factors, multiple family and sibling studies have demonstrated that genetic factors are of major aetiological significance for development of DD (Kovel et al. 2004), with an incidence of up to 50% of children of dyslexic parents and 50% of siblings of dyslexics are affected (Dennis et al. 2009; Couto et al. 2008; Fisher & Francks 2006; Gabel et al. 2009). The observation that a trait is highly inherited is not enough to establish that it is genetic, since its familial nature may instead be due to shared environment. Twin studies, however, have demonstrated a strong genetic component to this family association with a concordance rate of 68% in monozygotic twins versus 38% in dizygotic twins (Dennis et al. 2009). A significantly higher concordance in monozygotic twins implies that the trait has a genetic component, with estimates of heritability ranging from 44% to 75%, (Meng et al. 2005). DD is highly familial and heritable, with heritability comparable to other complex diseases, and therefore most likely results from the interactions between multiple genetic and environmental risk factors (Paracchini et al. 2006; Gibson & Gruen 2008).

1.3. KIAA0319 is involved in the development of reading disability

1.3.1. KIAA0319 is a strong DD candidate gene

Linkage analysis with single nucleotide polymorphisms (SNPs), tracing the lineage of marker alleles and comparing it with the lineage of DD, have successfully identified several loci that are likely to harbour susceptibility genes for dyslexia, which include chromosomes 1,2, 3, 6, 11, 13, 15 and 18 (Paracchini et al. 2006; Couto et al. 2009). Amongst them, 1p34–p36, 2p, 6p21.3, and 15q21 have been frequently replicated

(Gabel et al. 2009; Paracchini et al. 2006). Numerous candidate susceptibility genes were identified with genetic association studies, these are DYX1C1, ROBO1, DCDC2, MRPL19, C2ORF3, GRIN2B, SLC2A3 and KIAA0319 (Abbreviated as K throughout this thesis) (Cope et al. 2005; Dennis et al. 2009; Gibson & Gruen 2008; Ludwig et al. 2009; Roeske et al. 2009). Some of these linkage findings have been replicated in independent samples, providing stronger evidence to support these genes as dyslexia susceptibility genes (Paracchini et al. 2006). Many of the candidate genes such as DYX1C1, DCDC2 and K are involved in global brain development processes such as neuronal migration and axonal guidance (Wang et al. 2006; Paracchini et al. 2006; Peschansky et al. 2009; Dennis et al. 2009; Levecque et al. 2009; Velayos-baeza et al. 2008).

The locus DYX2, on chromosome 6p22, is one of the most consistently identified candidate regions by linkage studies (Couto et al. 2009; Dennis et al. 2009; Paracchini et al. 2006; Ludwig et al. 2008), which harbour the two DD susceptibility genes DCDC2 and K. Several studies using SNPs selected within brain- expressed genes has supported K being a strong candidate gene for Dyslexia (Couto et al. 2009; Luciano et al. 2007; Francks et al. 2004; Cope et al. 2005; Harold et al. 2006). In two large samples of nuclear families from the UK and the US, significant associations with DD was found in a 77 kb region of high inter-marker linkage disequilibrium which harbours the first four exons of K, all of TTRAP and the region immediately upstream of THEM2 (Francks et al. 2004; Dennis et al. 2009). Then, study using a dense set of SNPs further identifies a risk haplotype in the same region (Cope et al. 2005; Francks et al. 2004), and the

association was replicated by an independent study (Paracchini et al. 2008). The SNPs with the strongest association with DD was found to be located near the first exon of K (Harold et al. 2006) In the DD risk haplotype, the expression of K was shown to be reduced by 40% while other genes in the locus remain unaffected (Harold et al. 2006; Paracchini et al. 2006). More recently, the DD risk haplotype has been shown to create an aberrant binding site for the transcriptional silencer OCT-1, which confers the reduced expression of K (Dennis et al. 2009). Altogether, these results indicated K as a susceptibility gene for DD. Furthermore, altered K expression leads to impaired neuronal migration in rat (Paracchini et al. 2006). Interestingly, neuroanatomical examination revealed increases in the incidence of a focal neocortical abnormality in DD brains which occur as a consequence of disrupted neuronal migration (Gabel et al. 2009). The discovery suggests a connection between neuronal migration in the neocortex and DD. At present, all of the DD candidate genes that have their functions studied were shown to be involved in neuronal migration (Peschansky et al. 2009). Accordingly, K's role in neuronal migration strongly supports that K is a DD susceptibility gene.

In summary, K is a strong DD susceptibility gene with supports from several independent association and functional studies. The identification of DD candidate genes has opened the way for analysis that can focus on its molecular and cellular mechanisms as well as presented a new opportunity to test hypotheses with respect to potential causes of DD.

1.3.2. KIAA0319 involved in neuronal migration

Several studies have demonstrated that K appears to influence the migration of developing neurons during early embryogenesis. Embryonic transfection with shRNA targeted against the rat homolog of K disrupts neuronal migration by cell autonomous and non-cell autonomous mechanisms during the development of the neocortex (Paracchini et al. 2006; Peschansky et al. 2009). The impairment in migration was rescued by cotransfecting the shRNA with a plasmid encoding rat K, demonstrated that the migration impairment following transfection of shRNA is indeed specific to K (Paracchini et al. 2006). The disrupted neurons appeared to lose their normal radial association with radial glial fibers, suggesting K may function in neuron to radial glia adhesion (Gabel et al. 2009; Velayos-baeza et al. 2008). Moreover, down regulation of K also results in hypertrophy of apical dendrites (Peschansky et al. 2009). In addition to knockdown of K, K overexpression alters the position of neurons within cortical lamina, with transfected neurons migrating to positions below their expected location (Peschansky et al. 2009), which demonstrated that altered expression of K would disrupt the normal migration of neurons.

Neuronal migration is a key step in the development of the neocortex, by which newly generated neurons travel from their origin to different parts of the developing brain along the radial glial fibre. The lost of this relationship when the expression of K is altered suggested that K could mediate the appropriate adhesion between neurons and the glial fibres during neuronal migration (Paracchini et al. 2006), probably *via* the PKD domains.

In addition to K, two additional DD candidate genes DCDC2 on 6p and ROBO1 on chromosome 3, have been shown to be involved in neuronal migration and axonal path finding (Meng et al. 2005; Hannula-jouppi et al. 2005). Interestingly, altered neuronal migration, including molecular layer ectopias, laminar dysplasia, and occasional focal microgyria, has been demonstrated in the brains of DD based on post-mortem anatomical studies (Peschansky et al. 2009; Dennis et al. 2009), which strengthened a hypothesis that impaired neuronal migration in development causes a predisposition to DD(Gabel et al. 2009). Given the important role of K in neuronal migration and its strong association to DD, further study of the mechanism of how K might affect neuronal migration is required for better understanding of DD.

1.3.3. KIAA0319 expressed highly in brain

In addition to the functional studies, it has been reported that K exhibits a relatively specific spatial-temporal expression pattern in the developing human and mouse cerebral neocortex (Paracchini et al. 2006). Microarray analysis of different human tissue has shown that the expression of K in fetal brain is five times over the average K expression (Figure 1-1) (Su et al. 2004) (http://biogps.gnf.org), showing that K is highly expressed in developing brain. *In situ* hybridization on sections of an early fetal human brain also revealed intense expression of K in the developing cerebral neocortex and ganglionic eminence (Paracchini et al. 2006). Expression was prominent in the developing cortical plate (CP) and in the ventricular zone (VZ) at this stage of development where neuronal migration is well underway in the fetal neocortex (Paracchini et al. 2006). Besides,

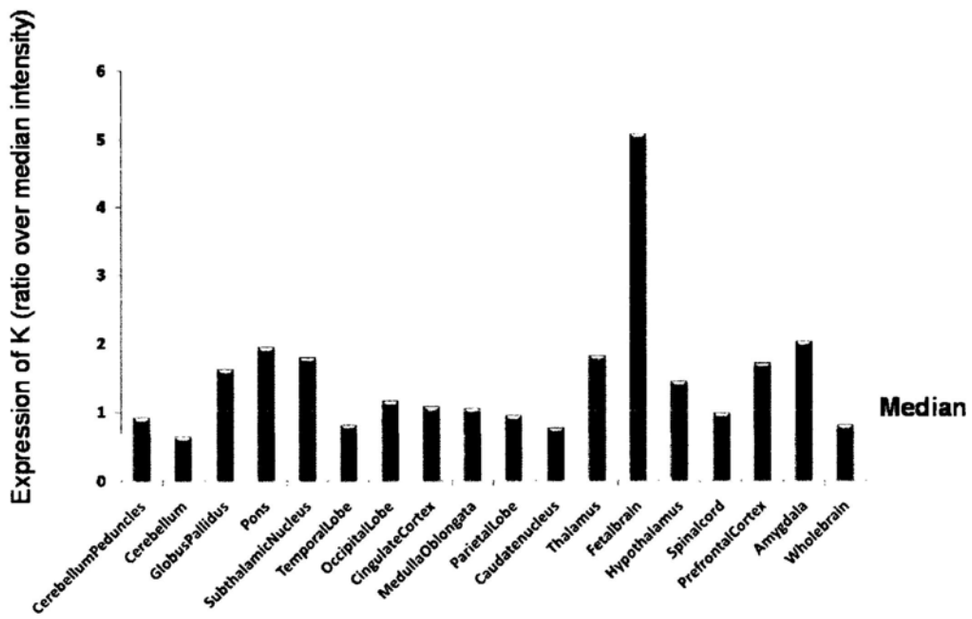


Figure 1-1. RNA expression pattern of K in brain tissue by microarray. Microarray were performed with different tissues (Su et al. 2004; http://biogps.gnf.org). Expression of K across different brain tissue using Affymetrix reporter: 206017_at were presented as a ratio of the intensity over median intensity. Expression of K in fetal brain was 5-fold above the average.

In situ hybridization in prenatal rat brains revealed that K was highly expressed in a regionally distinct manner with evidence of K expression in migrating neurons (Peschansky et al. 2009). There was increased expression of K in the VZ, intermediate zone, CP, striatum, hippocampus, and generalized moderate expression in the brain stem at all ages examined. Expression of K in the CP was shown to be increased over time. In addition, at 18th-19th day of gestation, there was increased expression of K in the mitral cell layer of the olfactory bulb (Peschansky et al. 2009). The high and specific expression of K in developing brain is consistent to the fact that it plays a role in neuronal migration during brain development (Paracchini et al. 2006).

In addition to the developing brain, K has been reported to exhibit a relatively specific expression in human adult brain, specifically in the cerebellum, superior parietal cortex, primary visual cortex, occipital cortex, amygdala, ganglionic eminence, mesencephalon and hippocampus, with much weaker expression in a restricted number of other tissues (Dennis et al. 2009; Meng et al. 2005). *In situ* hybridization has shown that in the adult mouse brain, expression of K was detected in the Cornu Ammonis region 3 (CA3) of the hippocampus and dentate gyrus, and in the Purkinje cell layer of the adult mouse cerebellum (Paracchini et al. 2006). The expression of K in mature neurons after migration suggested that it may have important functions in processes that would also affect learning (Gabel et al. 2009). It is therefore vital to investigate K more deeply. At present, most studies conducted have been focused on K's effect on neuronal migration and brain development. However, the mechanism underline is still unknown. For this reason, it is also very interesting to study K's function at the molecular level.

1.4. Multiple domains protein KIAA0319 involved in different functions

To study K's function at molecular level, apart from knowing its expression level, it is vital to gain more information on its protein structure. Genetically, the K gene is located in 6p22.3-p22.2 which contains 21 exons. The K open reading frame (ORF) consists of a coding region of 3,216 bp with several splice variants (Paracchini et al. 2007). Transcription of the 3 kb K mRNA is controlled by its own promoter which is regulated by the transcriptional silencer OCT-1 (Dennis et al. 2009). The K protein contains three main isoforms, the full length isoform is a type I plasma membrane protein that is consistent with its proposed function in neuronal migration (Velayos-baeza et al. 2008). The variant B lacks exon 19 while variant C lacks exons 19 and 20, both encode proteins lacking the C-terminal transmembrane domain (Velayos-baeza et al. 2008). These non-membranal isoforms could potentially have different functions than the full-length protein. While variant C was found to be localised to endoplasmic reticulum (ER), splice variant B was found to be secreted, which could indicate a role in signalling (Levecque et al. 2009; Velayos-baeza et al. 2008). Also, the secreted variant is consistent with a possible role in the significant non-cell autonomous component of migration disruption. (Peschansky et al. 2009).

The full length K protein, without the first 20 amino acids that contain the signal peptide, comprises 1,052 amino acids. K protein is a large trans-membrane protein which consists of four main parts, including the N-terminus of K which has a seven Cysteines

(MANSC) domain downstream of the signal peptide, a large cluster of five polycystic kidney disease homology domains (PKD) in the middle of the protein sequence, a Cysteine-rich C6 region together with a trans-membrane domain which had been demonstrated critical for forming K protein homo-dimer, and the only cytoplasmic C-terminus of K. (Figure 1-2A). The precise function of this protein and its involvement in neuronal migration are still unclear (Levecque et al. 2009). The MANSC domain is present in animal membrane and extracellular proteins, and has been postulated to bind to serine protease domains of HGFA and matriptase (Guo et al. 2004). PKD domain have been shown to form homophilic interactions and to mediate cell-cell adhesion in polycystin-1 (Velayos-baeza et al. 2008; Silberberg et al. 2005), imply that K might be involved in cell adhesion or cell-cell interaction (Velayos-baeza et al. 2008; Levecque et al. 2009). The multi Cysteine residues in the protein result in its poor solubility, which lead to great difficulty in its purification and the three-dimensional structure is therefore still unknown.

The domains predicted in K protein are not unique. K's homologue gene, KIAA0319-Like (KL) located in chromosome 1p34.2, encodes KL protein which has a similar domain arrangement compared with that of K (Figure 1-2B). Sequence alignment using similarity matrix BLOSUM62 (BioEdit) show that KL has 59% overall similarity and 45.9% amino acid identity with K, with only 6% of the coding sequence lying in gaps (Couto et al. 2008). Moreover, the PKD regions have 74.8% similarity and 62.4% identity. Given the high similarity with K and that KL is also expressed in the

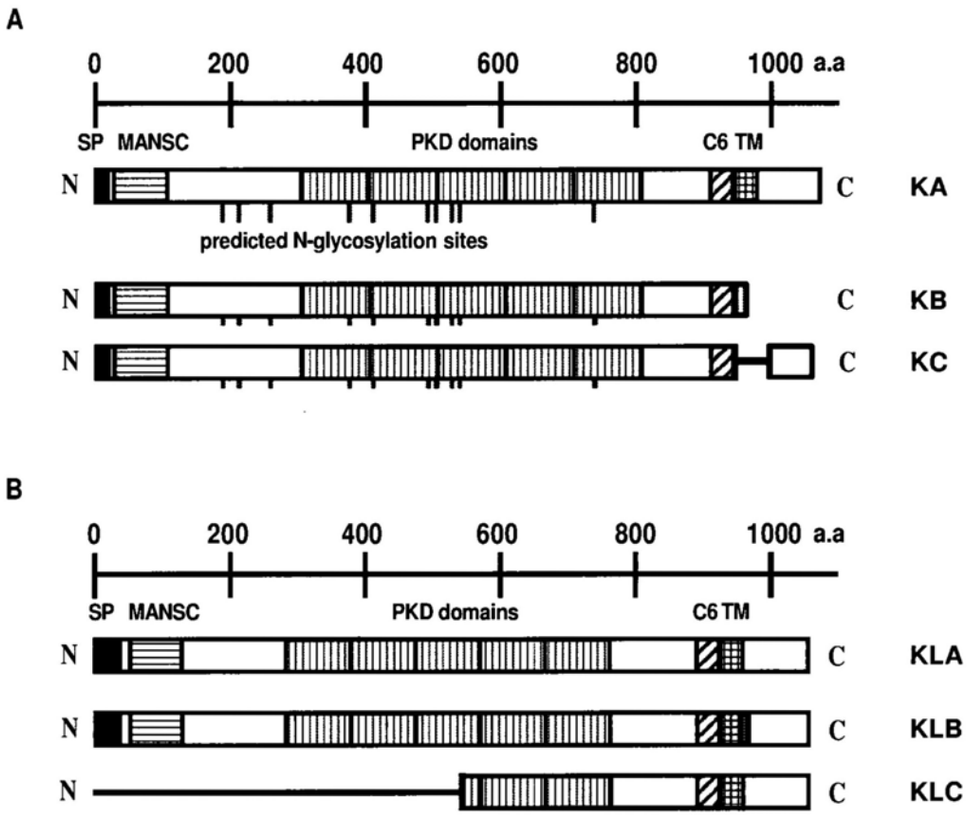


Figure 1-2. Domain arrangement of the human KIAA0319 and KIAA0319-Like proteins. **(A)** Graphical representation of the full length human KIAA0319 protein isoforms (KA, KB and KC). Regions or domains detected by computational analysis are shown: signal peptide (SP), MANSC (motif at N terminus with seven cysteines), Polycystic kidney disease domain (PKD 1 to 5), C6 domains (contains six conserved cysteine residues) and transmembrane domain (TM). Vertical lines below the main bar represent predicted N-glycosylation sites. **(B)** Graphical representation of the full length human KIAA0319-Like protein isoforms (KLA, KLB and KLC)

brain as well as other tissues (Couto et al. 2008), KL remains an interesting target for study, which might provide more information on the domain functions.

Four bands of full length K proteins were detected by western blot with sizes larger than the expected 115.8 kDa and two large bands approximately twice the size of the small ones, which indicated post-translational modification and dimerization (Levecque et al. 2009; Velayos-baeza et al. 2008). Indeed, it has been demonstrated that the central region including the PKD domains is N-glycosylated while a short fragment N-terminal to the PKD domains contains mucin-type O-glycosylation (Levecque et al. 2009; Velayos-baeza et al. 2008). Also, co-immunoprecipitation of differently tagged K proteins supports the formation of dimers by K (Levecque et al. 2009; Velayos-baeza et al. 2008). Collectively, the four bands detected for full length K proteins correspond to ER-glycosylated monomer (~150-170 kDa), fully glycosylated monomer (~180-210 kDa), ER-glycosylated dimer (~280-320 kDa), and fully glycosylated dimer (~350-400 kDa) (Velayos-baeza et al. 2008).

Glycosylation, an enzyme-directed site-specific process that attaches polysaccharide chains to the target proteins or lipids, is the most common and most complex form of posttranslational modification, which would play important roles in developmental processes (Haltiwanger & Lowe 2004; Wopereis et al. 2006). A large fraction of membrane-associated and secreted proteins in every eukaryote cell are N-glycosylated and such post-translational modification participate in many key biological processes including cell adhesion, molecular trafficking and clearance, protein folding, stability,

receptor activation, signal transduction, and endocytosis (Ohtsubo & Marth 2006). Similarly, mucin-type O-glycan is the most common form of O-glycosylation in humans, which have been found to function in protein structure and stability, immunity, receptor-mediated signalling, nonspecific protein interactions, modulation of the activity of enzymes and signalling molecules, and protein expression and processing (Wopereis et al. 2006). The function that glycosylation may have in the KIAA0319 protein is still unknown, but participation in cell-cell interaction or in cell adhesion is plausible.

It has been demonstrated that the Cysteine-containing domains are involved in the dimerization of K protein. Truncation studies have demonstrated that K exists mainly as homodimers when the transmembrane (TM) regions as well as the Cysteine (Cys) -rich C6 domain are present (Velayos-baeza et al. 2008). The monomeric form became more abundant when C6 domain is deleted, even the MANSC domain is still presence. In addition, the presence of the C6 domain and MANSC domain not only triggers dimerization but multimerization of the proteins (Velayos-baeza et al. 2008). Studies of mutating the Cysteine residue to other amino acids provide further support for the Cys residues involvement in dimerization (Velayos-baeza et al. 2008). The signal of the dimeric forms is particularly strong when the Cys residues in the TM domain are not changed. When these residues in the TM domain have been mutated, additional mutation of the Cys residues in the C6 domain decreases the level of dimerization while mutations in the MANSC domain do not have the same effect (Velayos-baeza et al. 2008). The formation of dimers depended on the presence of Cys-rich regions in the proteins suggests that several Cys residues are involved in disulphide bond formation between

the two molecules of the KIAA0319 dimer. Such dimeric conformation would most probably be required by some of the functions of K (Velayos-baeza et al. 2008).

The predicted domains must be essential to K's function and it is thus interesting to know which domain is significant to DD susceptibility. Study at the molecular level has provided important information. A yeast two hybrid screening has identified one protein interactor of K, the μ-subunit of the Adaptor protein 2 complex (AP2M1), which is involved in the clathrin-mediated endocytic pathway (Levecque et al. 2009). AP2M1 recognise the tyrosine- based YTIL motif at C-terminus of K in which the tyrosine-995 was critical for this interaction and is required for the subsequent internalization of K via a clathrin-dependent pathway (Levecque et al. 2009). Plasma membrane proteins, including cell-surface receptors, are tightly regulated to traffic between different subcellular compartments in response to specific signals, which play an important role in their sorting and regulation of their function. It is possible that the endocytosis of K is important to its function and the disruption of this internalization of K would be the reason for DD predisposition.

Another molecular study involved a library scale yeast two hybrid screening, which has identified two potential interacting partners for K, namely SH2B adaptor protein 1 (SH2) and human homologue of the Caenorhabditis elegans sex determination fem1 protein (FEM; Nakayama et al. 2002). Both SH2 and FEM are adaptor proteins expressed in brain and it is possible that K may interact with these proteins and therefore involved in their pathway. Molecular studies have provided important clues for how K may associate

to DD and therefore it is essential to continue the study at the molecular level.

1.5. Hypothesis and aims of the research

DD is one of the most prevalent types of learning disabilities. Impaired neuronal migration is unquestionably significant to DD. By understanding DD more fully, the development of genetic test for dyslexia susceptibility at an early age can be beneficial. Consequently special care to promote reading skills in children at risk of developing reading disorder can be provided at a critical stage when the brain is developing at an early age.

Structural brain differences in cortex and thalamus suggested DD may result from abnormal neuronal migration during fetal development. The condition being significantly heritable implies that genetic differences affecting early brain development caused DD. The studies showing K as a strong candidate gene involved in neuronal migration (Paracchini et al. 2006; Luciano et al. 2007; Francks et al. 2004; Cope et al. 2005; Harold et al. 2006), may provide an answer of how genetic factors play a critical role in DD predisposition. K as a multiple domain protein involved in different functions, the precise mechanism of how K affects brain development and thus connects to genetic susceptibility to DD is worthwhile of further studying at the molecular level.

CHAPTER 2

Research Method

2.1. General techniques

2.1.1. Cloning

2.1.1.1. Amplification of human KIAA0319 gene

Pairs of primers were designed flanking the full length human KIAA0319 open reading frame (ORF), the forward primer encompassing the ATG translation initiation codon and upstream sequences of K cDNA, and the reverse primer encompassing the translation stop codon and 17-21 bp downstream. Restriction sites and 5' end clamp sequences (TAGGGC or TATGCA) were added to each primer as listed in Table 2-1. 1-base pair or 2-base pair insertion was introduced to the forward primer making the fusion protein in frame with the N-terminal Tag of the vectors. Polymerase chain reaction (PCR) was performed in a 50 µl mixture with 200 nM each of the each of the primers, 1x *Pfu* buffer (10 mM KCl, 10 mM $(NH_4)_2SO_4$, 20 mM Tris-Cl (pH 8.75 at 25°C), 2 mM $MgSO_4$, 0.1% Triton X-100, 100 µg/ml BSA), 200 µM dNTP mixture (Promega), 2.5 Units *Pfu* polymerase (Promega), sterile nano pure water, and together with 1 µl diluted plasmid template. PCR reaction was performed using PTC-200 DNA Engine (Peltier Thermal Controller) by denaturing at 94°C for 2 minutes, followed by 35 cycles of amplification at 94°C for 30 seconds, 57°C for 30 seconds, 72°C for 5 minutes, with a final extension at 72°C for 10 minutes. The PCR product was analyzed on a 1 %

Table 2-1. List of primers used for subcloning KIAA0319.
Subcloning primers were designed based on the sequence of KIAA0319 (GenBank Accession Number: NM_014809). Each of the primers included a 5' end clamp sequences (TAGGGC or TATGCA) shown in **bold**, restriction sites indicated by underline, extra bases indicated by dots and initiating or stop codon to make the sequence in frame to the reading sequence of the vector.

Construct	Primer Name	Sequence 5'-->3'	Restriction Sites
Kmyc	KpcmvF	**TAGGGC**GAATTCATATGGCGCCC CCCACAGGTGT	*Eco*RI
	KpcmvR	**TAGGGC**GTCGACTTATCTGTCCTTT GAGCAAT	*Sal*I
Kpcdna	KpcdnaF	**TAGGGC**GAATTCGCCGCCGCCATG GCGCCCCCCACAG	*Eco*RI
	KpcdnaR	**TAGGGC**TCTAGATTATCTGTCCTTT GAGCAAT	*Xba*I
Kgfp	KgfpF	**TAGGGC**GAATTCAATGGCGCCCCCC ACAGGTGT	*Eco*RI
	KgfpR	**TAGGGC**GTCGACTTATCTGTCCTTT GAGCAAT	*Sal*I
KΔN myc	KΔN mycF	**TATGCA**GAATTCTGATGCCTGTTCA GAGGCCTGCA	*Eco*RI
	KΔN mycR	**TAGGGC**GTCGACTTATCTGTCCTTT GAGCAAT	*Sal*I
KΔNgfp	KΔNgfpF	**TATGCA**GAATTCCATGCCTGTTCAG AGGCCTGCA	*Eco*RI
	KΔNgfpR	**TAGGGC**GTCGACTTATCTGTCCTTT GAGCAAT	*Sal*I
KΔCmyc	KΔCmycF	**TAGGGC**GAATTCATATGGCGCCC CCCACAGGTGT	*Eco*RI
	KΔCmycR	**TATGCA**GTCGACTTAGTCAGCCTTC TCCTTTGA	*Sal*I
KΔCgfp	KΔCgfpF	**TAGGGC**GAATTCAATGGCGCCCCCC ACAGGTGT	*Eco*RI
	KΔCgfpR	**TATGCA**GTCGACTTAGTCAGCCTTC TCCTTTGA	*Sal*I
KΔTMmyc	KΔTMmycF	**TAGGGC**GAATTCATATGGCGCCC CCCACAGGTGT	*Eco*RI
	KΔTMmycR	**TATGCA**GTCGACTTAACAGTTGCTC TCT	*Sal*I
KΔTMgfp	KΔTMgfpF	**TAGGGC**GAATTCAATGGCGCCCCCC ACAGGTGT	*Eco*RI
	KΔTMgfpR	**TATGCA**GTCGACTTAACAGTTGCTC TCT	*Sal*I

TAE agarose (Invitrogen) gel with 0.7 μg/ml ethidium bromide (Invitrogen). 1 kb DNA Ladder Markers (Invitrogen) was included as ladder. The gel was visualized and photographed under ultra violet (UV) light, using a UVitec DNA gel machine.

2.1.1.2. Purification of PCR products

Electrophoresis separates the PCR products according to the sizes. The desired PCR products were visualized on a 2UV™ Transilluminator and then excised out and weighed. Purification of PCR product from the gel was performed using QIAquick® DNA Extraction Kit (QIAGEN). The excised gels were dissolved in 3 gel weight/volumes of QC buffer at 55°C for 10 minutes, with occasional shaking. The mixture was then loaded into a QIAquick spin column and put on a 2 ml collection tube. The column was then subjected to centrifugation at 13,000 r.p.m. for 1 minute. The flow through in the collection tube was discarded. The DNA bound to the column membrane was then washed once with 700 μl PE buffer. The column was centrifuged at 13,000 r.p.m. for 1 minute. After disposing of the flow through, the tube was re-centrifuged at 13,000 r.p.m. for 2 minutes, followed by 3 minutes air-drying to remove the residual ethanol in the buffer. The DNA was eluted from the membrane by applying 30 μl 65°C sterile nano pure water and incubated at room temperature for 2 minutes. The purified DNA products were obtained by centrifugation at 13,000 r.p.m. for 2 minutes.

2.1.1.3. Restriction enzyme digestion

Restriction enzyme digestion of the purified PCR product was performed in a 20 μl reaction mixture containing One-Phor-All PLUS buffer (Amersham Biosciences), 12

Units each of the restriction enzyme, 10 µl purified PCR product, and sterile nano pure water. Enzymatic digestion with single or double enzymatic reaction was performed in OPA^+ buffer at 1X or 2X concentration according to the restriction enzymes used. The reaction mixture was incubated at 37°C for 3 hours for restriction digestion followed by heat termination at 85°C for 20 minutes. Restriction digestion of the vector was performed in parallel using the same restriction enzymes. Single digestions with either one of the restriction enzymes were included as control for complete digestion. The digestion PCR products and vector were analyzed on a 1% TAE agarose gel. The target bands were excised out and subjected to PCR purification.

2.1.1.4. Ligation of gene products with vector

Ligation of the purified and digested gene products into vector was performed in a 20 µl reaction mixture containing 1X ligation buffer, 10 ng digested vector, 30 ng digested DNA insert, 5 units of T4 DNA ligase (New England Biolabs) and sterile nano pure water. A control reaction was performed in parallel by replacing the DNA insert with water. The reaction mixture was incubated at 16°C for 16 hours for ligation.

2.1.1.5. Preparation of chemically competent bacterial cells *E. coli* strain DH5α

Preparation of chemically competent bacterial cells for transformation was based on the Hanahan method (Hanahan 1983) which took about three days. On day one, Luria broth (LB) agar (USB) plate without antibiotics was prepared as a solid platform for bacterial growth. *E. coli* strain DH5α was streaked onto the LB plate and incubated overnight at

37°C to obtain single colonies. On day two, a single colony of *E. coli* was inoculated in 10 ml LB medium (USB) and the culture was incubated at 37°C for 16-20 hours with shaking at 250 r.p.m. to obtain a starter culture. On day three, 5 ml overnight culture was sub-cultured into 250 ml LB medium and further incubated and shaken for about 2-3 hours until its absorbance at wavelength (λ) 600 nm reached 0.4-0.6. Bacterial cells were harvested by centrifugation at 6,000 r.p.m. at 4°C for 10 minutes. The supernatant was decanted and the cell pellet was resuspended in 83 ml (1/3 of the original culture volume) RF1 buffer (100 mM RbCl, 50 mM $MnCl_2$ $4H_2O$, 35 mM K acetate, 10 mM $CaCl_2$ $2H_2O$, 15% Glycerol, pH 5.8) by pipetting up and down. The suspension was incubated in ice for 15 minutes before centrifugation at 6,000 r.p.m. at 4°C for 10 minutes to harvest the cells. The supernatant was discarded and the cell pellet was resuspended in 20 ml (1/25 of original culture volume) RF2 buffer (10mM MOPS, 10 mM RbCl, 75 mM $CaCl_2$ $2H_2O$, 15% Glycerol, pH 6.8) by pipetting up and down. The suspension was incubated in ice for further 15 minutes and then aliquot to 200 µl fractions into pre-chilled 1.5 ml microcentrifuge tubes. The aliquots were frozen immediately in liquid nitrogen and then stored at -80°C.

2.1.1.6. Transformation of the ligation product into competent cells

The frozen competent cells *E. coli* strain DH5α cells were thawed on ice for 10 minutes. Ligation mixture was then mixed with 200 µl competent cells and incubated on ice for 20 minutes. The cells were heat shocked at 42°C for 2 minutes and then chilled immediately on ice for 10 minutes. To the cell mixture, 800 µl LB medium was added, followed by incubation at 37°C for 1 hour with shaking at 250 r.p.m. The cells were

centrifuged down at 9,000 r.p.m. for 3 minutes and resuspended in 150 µl LB medium. All the cell suspension was spread onto LB agar plate containing appropriate antibiotics and incubated at 37°C for 16 hours in an inverted position.

2.1.1.7. Diagnostic PCR for confirmation of successful ligation

Single colonies of the transformants were picked up from the LB- agar plates and seeded onto another LB- agar plate (containing appropriate antibiotics) grid to assign a number to each colony. Some of the cells were picked for PCR insert checking and the plate were incubated at 37°C for further 16 hours. PCR mixture was prepared using Ducan DNA *Taq* polymerase (*Thermus aquaticus*)in a 10 µl mixture with 1X $MgCl_2$ free PCR buffer (10 mM Tris-HCl (pH 9.0), 50mM KCl), 200 µM dNTP mixture (Promega), 200 µM $MgCl_2$, 2.5 units Ducan *Taq* polymerase , 200 nM of gene specific forward primer and vector specific reverse primer and sterile nano pure water. The reaction was performed on a PTC-200 DNA Engine (Peltier Thermal Controller) by denaturing at 94°C for 2 minutes, followed by 25 cycles of amplification at 94°C for 30 seconds, 57°C for 30 seconds, 72°C for 5 minute, with a final extension at 72°C for 10 minutes. The PCR products were analyzed on a 1% TAE agarose gel.

2.1.1.8. Small scale preparation of bacterial plasmid DNA

The clones that gave a desirable PCR product were inoculated in 5 ml LB medium containing appropriate antibiotics and incubated at 37°C overnight with shaking at 250 r.p.m. The overnight culture was harvested by centrifugation at 6,000 xg for 5 minutes. The plasmid DNA was isolated by using Rapid Plasmid Miniprep System (QIAGEN)

with recommended procedures of the kit. The supernatant was discarded and the cell pellet was resuspended in 250 µl 4°C Cell Suspension Buffer by vortexing. Then 250 µl Cell Lysis Buffer was added to the cell suspension and mixed by inverting the tube up and down for five times. Lysis was performed by incubating at room temperature for 5 minutes. To the lysis mixture, 350 µl Neutralization Buffer was added and mixed immediately by inverting up and down the tubes. The suspension was centrifuged at 13,000 r.p.m. for 10 minutes to remove the cell residues. The clear supernatant was loaded into a spin cartridge that was put into a 2 ml collection tube and centrifuged at 13,000 r.p.m. for 1 minute. The flow through was discarded and 700 µl Wash Buffer that contained ethanol was added to the spin cartridge and centrifuged at 13,000 r.p.m. After disposing of the flow through, the tube was recentrifuged at 13,000 r.p.m. for 2 minutes and then air-dried for 3 minutes to remove the residual buffer and ethanol. The plasmid DNA was eluted with 50 µl 65°C sterile nano pure water.

2.1.1.9. DNA sequencing of the cloned plasmid DNA

DNA sequencing for the cloned plasmid DNA was performed by the commercial sequencing service TechDragon Company with a vector specific forward primer and gene specific internal primers.

2.1.1.10. Large scale preparation of target recombinant plasmid DNA

After verification of the insert sequences by DNA sequencing, clones with interesting inserts were picked up from the streaked plates and inoculated into 150 ml LB medium containing appropriate antibiotics. The culture was incubated at 37°C overnight with

shaking at 250 r.p.m. The overnight culture was harvested by centrifugation at 6,000 x g at 4°C for 10 minutes. The plasmid DNA was isolated by using HiSpeed Plasmid Midi Kit (QIAGEN). The supernatant was discarded and the pellet was resuspended in 6 ml resuspension Buffer P1 by vigorous vortexing. Then 6 ml lysis Buffer P2 was added and mixed gently. The mixture was incubated at room temperature for 5 minutes cell lysis. After cell lysis, 6 ml neutralization Buffer P3 was added to the lysate and mixed gently immediately. The neutralized lysate was then transferred into the barrel of a QIAfilter™ Midi Cartridge that was closed by a cap, and allowed to incubate at room temperature for 10 minutes. During the incubation, a Hispeed™ Midi Tip was equilibrated with 4 ml Buffer QBT. After incubation, the neutralized lysate was filtered into the equilibrated Hispeed™ Midi Tip with the aid of a plunger. The lysate was allowed to enter the resin of the Midi Tip by gravity flow. The Midi Tip was then washed with 20 ml Buffer QC by gravity flow. The bound DNA was eluted with 5 ml Buffer QF. To the eluted DNA, 3.5 ml isopropanol was added to precipitate the DNA. The content was mixed and incubated at room temperature for 5 minutes. During the incubation, a QIAprecipatator™ Midi Module was fixed to a 20 ml syringe. The eluate/isopropanol mixture was transferred into the syringe and filtered into the Midi Module with a matched plunger. The QIAprecipatator™ Midi Module was removed from the syringe outlet nozzle before pulling out the plunger. The QIAprecipatator™ was reattached to the syringe and 2 ml 70% ethanol was applied to the syringe. The 70% ethanol was filtered through the QIAprecipatator™ by inserting the plunger. The bounded DNA was air dried to remove the residual ethanol by pressing air through the QIAprecipatator™. The QIAprecipatator™ was reattached to a 5 ml syringe and 750 µl 65°C sterile nano pure

water was applied. The matched plunger was inserted and pressed to transfer the eluted DNA into a 1.5 ml microcentrifuge tube. The quantity of plasmid DNA and protein was measured by nanodrop (Thermo) spectrophotometer at 260 nm and 280 nm respectively. The quality of the plasmid DNA was analyzed on a 1% TAE agarose gel.

2.1.2. Cell Culture

2.1.2.1. Culture medium

The media used in this study were purchased in powder form with the same preparation method suggested by the supplier. Dulbecco's modified Eagle's medium (DMEM), consisting phenol red, 4 mM L-glutamine and 4.5 g/L glucose. Liquid medium of DMEM was made by dissolving packs of DMEM powder in nano pure water at 1 pack to 1 liter ratio. 3.7g /L sodium bicarbonate was added and the pH was adjusted to 7.2 (0.2 below desire pH value 7.4). The medium was sterilized by filtering with 0.22 μM bottle-top filter (Millipore) and stored at 4°C. Addition of 10% (v/v) fetal bovine serum (PBS) (HyClone) and 1% (v/v) Penicillin-streptomycin (PS) (10,000 units/ml, HyClone) afterwards gained the complete DMEM medium.

2.1.2.2. Cell lines

The cell lines used in this study were cultured in a 37°C, 5% carbon dioxide incubator with humidified atmosphere. Cells were passed at 1:5 ratio every 3-4 days. For each passage, the spent medium was discarded and cells were washed twice with 1X phosphate buffer Saline (PBS) before trypsinization with 0.25% trypsin-EDTA solution (Invitrogen). The trypsinized cells were rinsed out with culture medium and subjected to

centrifugation at 1,000 xg for 3 minutes to remove the dead cells debris. The cell pellet was then resuspended in complete medium and cultured in desired culture flasks (IWAKI).

Human Kidney cell line HEK 293 (ATCC number CRL-1573) and HEK 293FT (supplied with a lentiviral kit from Invitrogen cat. no. K4975-00) exhibited adherent characteristics and were maintained in complete DMEM.

Human liver cell line WRL68 (ATCC number CL-48) exhibited adherent characteristics and was maintained in Complete DMEM.

Human neuroblastoma cell line SHSY-5Y (ATCC number CRL-2266) and human glioblastoma cell line u87MG (ATCC number HTB-14) exhibited adherent characteristics and were maintained in 1:1 DMEM and Ham's F12 Medium (Gibco) supplemented with 10% FBS and 1% PS (Complete F12)

Mouse myeloma cell line NS0 (European Collection of Animal Cell Cultures no: 85110503; Keen & Steward 1995) exhibited suspension characteristics and was maintained in complete DMEM along with 8-azaguanine (Sigma) to maintain the HGPRT minus phenotype.

Mouse hybridoma cells were grown in complete DMEM containing either Hypoxanthine Aminopterin Thymidine (HAT, Gibco) or hypoxanthine thymidine (HT, Gibco)

supplement. After selection was completed, the hybridoma cells were maintained in complete DMEM with double amount of FBS.

2.1.2.3. Freezing and thawing cells

To make frozen stock for long term storage of cells, cells were resuspended in complete medium and mixed with an equal amount of 20% DMSO in complete medium (Freezing mix) at a concentration of 1.0 $x10^7$ cells/ml. Following the vials being placed in a container that achieved a decrease of 1°C per minute inside a –80 °C freezer overnight, the cells were allocated in liquid nitrogen for long term storage.

To recover cells from nitrogen, cells were resuspended in complete DMEM media and centrifuged at 1,600 rpm for 10 minutes to remove DMSO in the freezing mix. They were then resuspended in complete DMEM in T75 cell culture flask. Media was refreshed every few days as for other cells.

2.1.3. DNA Transfection

2.1.3.1. Transfection reagents

Lipofectamine™ Reagent and Plus™ Reagent were purchased from Invitrogen Life Technology Ltd. Lipofectamine™ Reagent is a liposome formulation suitable for transfection of DNA into cultured eukaryotic cells. Plus™ Reagent is an enhancing reagent that can increase the transfection activity of Lipofectamine™ Reagent.

2.1.3.2. Small scale transfection of 6-well plate

Cells were seeded onto a 6-well plate the day before transfection at density that gives 60% confluent. Transfection of plasmid-DNA was performed by using Lipofectamine™ Reagent (Invitrogen) in DMEM without FBS or antibiotics according to the manufacturer's protocol. On the day of transfection, 0.8 μg plasmid were diluted in 100 μl serum-free DMEM. The diluted plasmid was then mixed with 6 μl Plus™ Reagent (Invitrogen) and allowed to incubate at room temperature for 15 minutes. Meanwhile, 4 μl Lipofectamine™ was diluted by 100 μl serum-free medium. After incubation, the DNA-Plus pre-complex and the diluted Lipofectamine™ reagent were mixed together and further incubated for 15 minutes at room temperature. During the incubation, the complete culture medium on the 6-well plate was replaced by 0.8 ml serum-free plain DMEM. The DNA-Plus-Lipofectamine complexes were added to the cells and incubated at 37°C, 5% CO_2 for 5 hours. After 5 hours incubation, the serum-free medium with the transfection mixture was replaced by 2 ml 10% serum medium and allowed to grow.

2.1.3.3. Large scale transfection of 100 mm dish

Cells were seeded onto a 100mm dish the day before transfection at density that gives 60% confluent. Transfection of plasmid-DNA was performed by using Lipofectamine™ Reagent (Invitrogen) and Plus™ Reagent (Invitrogen) as described by scaling up the plasmids to 6.0 μg.

2.1.4. Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was performed to separate the cellular proteins according to their electrophoretic mobility which was determined by the molecular weight. SDS-polyacrylamide gel was set to 8% -12% depending on the size of the target protein using 40% PlusOne Ready IEF solution 19:1 (Amersham Biosciences). Samples were mixed with 6x Laemmli buffer (125 mM Tris·HCl, pH 6.7, 6% SDS, 20% glycerol, 10% 2-mercaptoethanol, and 0.1% bromphenol blue) and subjected to boiling for 10 minutes before loading into the gel. Pre-stained SeeBlue rainbow marker (Invitrogen; 6 μl) or Novex Sharp Pre-stained Protein Standard (Invitrogen; 10μl) was loaded in parallel as a standard molecular weight marker. The gel was run at 150-180 volts for 1-1.5 hours.

2.1.5. Western Blot (WB)

After electrophoresis, the stacking gel was removed and the running gel was briefly rinsed with reverse osmosis water. The running gel was then equilibrated in transfer buffer (25 mM Tris, 192 mM glycine, 15% methanol) for 3 minutes. During the equilibration, a piece of Immobilon^{TM}-P Transfer Membrane was fixed with 100% methanol for 15 seconds and then equilibrated in transfer buffer. Two sets of 3 pieces of Whatman chromatography papers (Kent, UK) were soaked in transfer buffer. Transfer was performed with the running gel and the transfer membrane placed in between two sets of 3 pieces of chromatography papers. Air bubbles were excluded using a roller. The electroblot transfer was performed at 15 volts for 80 minutes on a Trans-blot$^{®}$SD semi-dry transfer cell (BioRad).

After transfer, the membrane was incubated in PBS, 0.1% Tween 20 (Sigma), 5% Blotting-Grade Blocker Non-Fat Dry Milk powder (BioRad; blocking buffer) with constant gentle shaking for an hour to block the non-specific binding sites of antibody. Appropriate primary antibody was added directly to the blotting solution and incubated with constant shaking at 4°C overnight. After primary antibody binding, the membrane was briefly rinsed with PBS 0.1%, Tween 20 (Sigma) (washing buffer) and then washed for three times with 10 minutes each with constant shaking. Appropriate secondary antibody was added to blocking buffer and incubated for 1 hour at room temperature with gentle shaking. After the antibody binding, the membrane was washed for three times with 10 minutes each with constant shaking.

Luminal detection solution was prepared by mixing equal amount Western Lightning™ Chemiluminescence Reagent Plus Enhancing reagent and Oxidizing reagent (Perkin Elmer). The membrane was briefly rinsed in washing buffer and then briefly dried by touching the edge of the membrane against a paper towel. The membrane was then immersed in an inverted position in the luminal detection solution and allowed to stand for 1 minute. The membrane was then covered by a transparent film and placed in a film cassette. A sheet of Super RX X-ray film (Fujifilm) was placed into the film cassette in a dark room and allowed to expose for 1 second to 2 minutes depending on the antibodies used. The film was developed with a Kodak film developer. To remove bound antibody, some blots were incubated with stripping buffer (62.5 mM Tris-HCl, pH 6.8, 100 mM β-mercaptoethanol, and 1% SDS) at 50°C for 30 minutes and blocked with blocking

buffer for 1 hour at room temperature. The blot can then used to re-probe other antibody. To correct for any loading artifact, Blots were re-probed with anti-β-actin IgG (Sigma). Alternatively, as loading control, the presence of rabbit or mouse heavy chain IgG in the immunoprecipitate samples were compared.

2.1.6. Generation of polyclonal antibodies (Pab)

2.1.6.1. Peptide synthesis

Synthetic peptide P124 (GSPSGIWGDSPEDI, corresponding to the amino acid residues 124-137 of human K) and p663 (VRGPSAVEMENIDK , corresponding to the amino acid 663-676 of human K) were synthesized using the standard method of solid-phase peptide synthesis (SBS genetech). Synthetic peptides were analyzed for purity by reverse-phase HPLC. Electrospray mass spectroscopy was used to verify the structure and purity of each peptide.

2.1.6.2. Peptide conjugation

To effectively elicit an antibody response and develop immunological memory, synthetic peptides P124 (GSPSGIWGDSPEDI - corresponding to the amino acids 124-137 of human K) were chemically conjugated to strongly immunogenic carrier proteins, keyhole limpet hemocyanin (KLH; Pierce) or BSA (Sigma), for immunization or ELISA, respectively. The carrier proteins were first dissolved in PBS at a final concentration of 5 mg/ml. 1mg of peptide were added to 1ml of carrier protein solution and sonicated in water bath for 30 minutes. 5mg/ml of N-Hydroxysuccinimide (NHS) (Sigma) and 10 mg/ml of N-(3-Dimethylaminopropyl)-N-ethylcarbodiimide hydrochloride (EDAC;

Sigma) were added and the mixtures were mixed at room temperature for 30 minutes. Desalting of the proteins was performed with disposable PD-10 desalting columns packed with Sephadex™ G-25 Medium (GE healthcare) following manufacturer instructions. Cloudy eluent collected were aliquoted and stored at -20°C.

2.1.6.3. Immunizations

Pab were generated in Balb/c mice (female, adult) by multiple immunizations. Mouse bloods were taken before any immunizations (preserum). Mice were immunized by inoculated intramuscularly with 30 ug of peptide P124 coupled to the carrier protein KLH in 100 ul Freund's complete adjuvant (sigma). Boosting inoculations were administered 3, 6, and 9 weeks after the initial inoculation with 15 ug of peptide P124 conjugated to KLH in Freund's incomplete adjuvant. The protein-adjuvant emulsion was always prepared within 1 h of inoculation into the animals. Bleeding was performed after the third inoculations.

To isolate serum, the blood was incubated at 4°C overnight to clot and centrifuged for 10 min at 2,000 xg at 4°C to separate the liquid phase from the clotted components. The cleared serum was collected and stored at −20°C until subjected to analysis. The antisera were tested by ELISA on peptide-coated plates or western blotting

2.1.6.4. Enzyme linked Immunosorbent Analysis (ELISA)

96-well culture plates were coated with 50ul per well of peptide P124 coupled to the

carrier protein BSA 4ug/ml in 100 mM carbonate-bicarbonate buffer, pH 9.6 (Sigma) and incubated overnight at 4°C (0.2 ug peptide /well). The plates were washed three times with washing buffer and blocked at 37°C for 1 hour with 2% BSA and 0.1% Tween 20 (Sigma) in PBS. After three washes with washing buffer, primary antibody (Pab diluted 1:1,000 or undiluted hybridoma supernatants) were added to the plates and incubated for 1 hour at 37°C. For a background and as a negative control PBS was used. The plates were washed five times and goat anti-mouse IgG (Fc specific) conjugated with Alkaline Phosphatase diluted 1:30,000 was added to the plates and incubated for 1 hour at 37°C. The plates were then washed five times and developed with 4 mg/ml p-Nitrophenyl phosphate (PNPP; Sigma) in 10% diethanolamine buffer (Sigma). The plates were analyzed with a microplate reader (BIO-RAD, model 3550) at 405 nm with wavelength correction set at 620 nm. Samples were considered positive if their optical density (OD) exceeded three times the negative control reading.

2.1.7. Synthetic peptide based neutralization

To determine which band or staining is specific, an immunizing peptide blocking experiment was performed. Peptide neutralization was performed by incubating 1 µl of antibody with 4ug of blocking peptide that corresponds to the epitope recognized by the antibody. Following incubation for 60 min at room temperature, the antibody/peptide mixtures were diluted into 4 ml blocking buffer and western blotting was performed as described.

2.1.8. Immunofluorescence (IF)

Cells were seeded on glass coverslips in a 6-well plate. Transfections were performed if necessary. The following day, cells were fixed in 4% paraformaldehyde, and permeabilized with 0.5% Triton X-100. Cells were then incubated for 2 h at room temperature with IF blocking buffer (PBS, 0.1% Triton X-100, 10% goat serum) and stained with appropriate primary antibody for 4°C overnight. At the end of the incubation time, cells were washed three times with PBS and incubated with appropriate secondary antibody for 1 hour at room temperature. Then, cells were washed three times with PBS and incubated with DAPI or Hoechst for 10 minutes at room temperature. Finally, cells were washed three times with PBS and prepared for microscopy observation.

2.1.9. Immunoprecipitation (IP) and Co-immunoprecipitation (Co-IP)

To perform IP and Co-IP, cells were washed in phosphate buffered Saline (PBS) and collected by scraping. Cells were then lysed in K-lysis buffer (50 mM Tris-HCL (pH 8), 150 mM NaCl, 1% Triton-X100) in the presence of protease inhibitors (Complete; Roche). The cleared cell lysates were incubated with appropriate primary antibody for 2 hours at 4°C. The immunocomplexes were absorbed onto 50 µl of Protein G Sepharose 4 Fast Flow beads (GE Healthcare) for overnight at 4°C. The beads were then washed three to five times in K-IP buffer (50 mM Tris-HCl (pH 7.8), 150 mM NaCl, 1 mM EDTA) in the presence of protease inhibitors (Complete; Roche) and incubated in 6x Laemmli buffer for 5 min at 95°C. The proteins were subjected to 8% SDS-PAGE

followed by either coomassie staining or WB. In the case of Co-IP, the immunoprecipitated complexes were analyzed by WB using corresponding antibodies.

2.2. Chapter 1 Methods

2.2.1. Computational analyses:

The deduced K proteins were analyzed *in silico* with a number of tools available at the ExPASy Molecular Biology Server (http://www.expasy.org/), including InterProScan (http://www.ebi.ac.uk/InterProScan) for domain/motif search and SignalP (http://www.cbs.dtu.dk/services/SignalP), NetNGlyc (http://www.cbs.dtu.dk/services/NetNGlyc), NetOGlyc (http://www.cbs.dtu.dk/services/NetOGlyc), and SOSUI (http://www.bp.nuap.nagoya-u.ac.jp/sosui) for prediction of signal peptide sequence, N-glycosylation, O-glycosylation, and transmembrane (TM) regions, respectively.

2.3. Chapter 3 Methods

2.3.1. Subcloning of full length KIAA0319 into myc expression vector

Subcloning of full length human K into pCMV-myc vector was performed with primers KpcmvF / KpcmvR as listed in Table 2-1. A 2-base pair insertion was introduced to the forward primer making the fusion protein in frame with the N-terminal c-myc tag of the pCMV-myc vector (Clontech). PCR was performed using 1 μl diluted hg00378 plasmid template obtained from Kazusa DNA research institute, which contain the full-length

human KIAA0319 cDNA in pBluescript II SK(+) vector. The PCR product and the vector were digested with *EcoRI* (New England Biolabs) and *SalI* (New England Biolabs). The transformed *E.coli* were spread onto LB agar plate containing ampicillin (100 μg/ml) (Amersham Biosciences). Diagnostic PCR were performed with KpcmvF and pCMV-myc vector specific reverse primer.

2.3.2. *In vitro* site-directed mutagenesis

To make the sequence corresponds to the RNA reference sequence available in the database (GenBank Accession Number: NM_014809), mutations at position 424, 469 and 1836 were introduced by PCR using the QuikChange™ site-directed mutagenesis kit from Stratagene. Primers with the desired mutation in the middle were designed according to the guideline in the kit, as listed in Table 2-2. A reaction mixture of 50 μl was set up which contained 1X reaction buffer, 125 ng of forward primer, 125 ng reverse primer, 1 μl dNTP mix (Promega), 2.5 units of *Pfu Ultra* HF DNA polymerase and 20 ng of DNA template. The mixtures were then subjected to a thermal cycling of DNA denaturation at 95°C for 30 seconds, followed by 16 cycles of 95°C for 30 seconds, 55°C for 30 seconds and 68°C for 10 minutes. The reaction was cooled down to <37°C when completed. The nonmutated parental dsDNA templates were then digested by *DpnI* restriction enzyme which selectively digested the methylated or semi-methylated templates. The reaction was performed by directly adding 10 units of *Dpn I* to the amplification reaction mix and incubating at 37°C for 3 hours. Transformation into the DH5α competent cells were performed by adding 10 μl *Dpn I*-treated DNA to 100 μl of

Table 2-2. List of primers for site-directed mutagenesis.
Primers for site-directed mutagenesis were designed as instructed by the kit. The forward and reverse mutagenic primers contained the desired mutations and anneal to the same sequence on opposite strands of the plasmid. The mutated nucleotides were marked in red and the affected codons shown in upper case and in bold.

Mutation (DNA)	Mutation (AA)	Primer Name	Sequence 5'-->3'
C=> A	Proline=> Threonine	424F	ggatatcagaaaggacttg**ACC**tttctaggcaaagattgg
		424R	ccaatctttgcctagaaa**GGT**caagtcctttctgatatcc
G=> T	Alanine=> Serine	469F	ggagatgtctgagtac**TCA**gatgactaccggg
		469R	cccggtagtcatc**TGA**gtactcagacatctcc
A=> G	Valine=> Valine	1836F	ggcaacagtctactgct**GTG**gtgactgtgattgtcc
		1836R	ggacaatcacagtcac**CAC**agcagtagactgttgcc

the competent cells. Colonies from the mutation sample plates were picked up and cultured. Sequencing for the target of interest was performed after obtaining the plasmid using Rapid Plasmid Miniprep System (QIAGEN). The mutated clone Kmyc was used as template for subcloning KIAA0319 into other vectors.

2.3.3. Gene expression pattern study (RNA level)

2.3.3.1. Primer design

Primers for real-time PCR were designed using ProbeFinder software v2.45 (Roche). The primers were designed to flank 150-300 base pairs of the target gene sequence with an annealing temperature of 58-60°C.

2.3.3.2. Cell transfection and RNA extraction

SHSY5Y cells were transfected with Kmyc or pCMV-myc vector. At 24 hours post-transfection, total RNA from SHSY5Y cells was isolated by TRIzol® reagent (Invitrogen). Cells were washed with sterilized 1X PBS twice and 4 ml of TRIzol® reagent was added to each dish to lyse the cells. After 3 minutes incubation with TRIzol® reagent at room temperature, the cells were made to detach from the plate by gently tapping the edge of the plate. The cell lysate was transferred into a 15 ml centrifuge tube. Then 1.2 ml ice-cool chloroform was added to the cell lysate and the mixture was shaken vigorously. The mixture was then centrifuged at 14,000 x g at 4°C for 20 minutes. After centrifugation, 3 layers: a lower red phenol-chloroform layer, a white interphase layer and a colourless upper aqueous layer were obtained. The upper aqueous layer containing RNA was transferred to four new 1.5 ml microcentrifuge tubes

and equal volume of 70% ethanol was added to increase the binding stability of RNA to the column membrane. The mixture was then transferred to an RNeasy spin column (QIAGEN) placed in a 2 ml collection tube. The column was then centrifuged at 9,000 x g for 15 seconds at 4°C. After centrifugation, the flow through was discarded. Then 700 µl Buffer RW1 was added to the RNeasy spin column and centrifuged at 9,000 x g for 15 seconds to wash the spin column. The flow through was removed completely and 500 µl Buffer RPE was added to the RNeasy spin column. The column was centrifuged at 9,000 x g for 15 seconds at 4°C to wash the column again. After discarding the flow through, the column was washed again with 500 µl Buffer RPE by centrifuge at 9,000 x g for 15 seconds. The column was then put into a new 2 ml collection tube and centrifuge at 9,000 x g for 15 seconds to remove the residual buffer. Finally, the column was placed on a new collection tube and the RNA sample was eluted with 30 µl of DEPC-treated water. DNase I was performed prior to reverse transcription in one 10 µl reaction with 1 ug RNA, according to the instruction of supplier. The total RNA yield was quantified by spectrophotometric analysis (NanoDrop Technology, Cambridge, UK) and the quality was verified by gel electrophoresis

2.3.3.3. Reverse Transcription for first strand complementary DNA (cDNA)

Reverse transcription for first strand cDNA synthesis was carried out using the High-Capacity cDNA Transcription Kits (Applied Biosystems). A 10 µl mixture was set up on ice with 2 µl of 10X RT Random Primers, 0.8 µl of 100 mM dNTP mix (Promega), 1 µl of MultiScribe™ Reverse Transcriptase and 1 µl of RNase Inhibitor and DEPC-treated water. To each 10 µl of the reaction mixture, 1 µg RNA in 10 µl volume

was added and mixed well. Reverse transcription was carried out in a thermal cycler with 25°C for 10 minutes, 37°C for 2 hours and terminated with incubation at 85°C for 1 minute. Then1 µl of RNase H (Applied Biosystems) was added and incubated at 37°C for 20 minutes. The first strand cDNA product was stored at -20°C.

2.3.3.4. Quantitative real-time PCR

For the comparison of mRNA expression between K overexpressed and non K overexpressed samples, Quantitative real-time PCR with Power SYBR® Green PCR Master Mix (Applied Biosystems) was performed. SYBR Green was a dye that can be bound to the minor groove of double stranded DNA which exhibited little fluorescence in solution but strong fluorescence when bound to DNA. When SYBR Green binds to the DNA, the intensity of fluorescent emissions increased in proportion, giving to an amplification curve from which the expression level of specific gene can be compared in separated samples.

Real-time PCR was performed in triplicate in a 20 µl reaction mixture of 1 µl of cDNA sample (about 30 ng), 2.5 µM forward and reverse primer mix, 10 µl Power SYBR® Green PCR Master Mix (Applied Biosystems) and sterile nano pure water. A separate mixture for human porphobilinogen deaminase (PBGD) was included as an internal control for each trial. The mixtures were loaded onto a 96-well PCR plate (ABI Prism, Applied Biosystems) and sealed. The PCR was performed on an ABI Fast 7500 real-time PCR machine with standard profile of 1 cycle of 50°C for 2 minutes, 95°C for 10 minutes, 40 cycles of 95°C for 15 seconds, 60°C for 1 minute. The relative amount of

the gene transcripts was determined using the cycle threshold (Ct) method with autoCt settings using Fast 7500 SDS program. After normalization with the endogenous PBGD control, the average critical Ct was used for calculating relative quantification by the fold change.

2.3.3.5. Microarray

In this study, 1 μg of total RNA for each sample were sent to Agilent for performing Competitively Hybridized (Two-Color) Microarrays. Briefly, The labeled cRNA was hybridized to 44K Human Whole-genome 60-mer oligo-chips (G4112F, Agilent Technologies, Palo Alto, CA, USA), which contained 41094 sequence-verified, oligonucleotide probes mapped to 40090 distinct genes. The image data were read and the intensity of the fluorescent images was quantified. The results were normalized by subtracting local background fluorescence with quality control passed using Feature Extraction version 10.5 (Agilent).

The microarray data obtained from Agilent was then analyzed by GeneSpring (Agilent) using 014850 as design id. The differentially expressed genes identified by microarray assay were validated by real-time PCR.

2.3.3.6. Analysis of real time PCR data

Real-time PCR analyzes the relative abundance of PCR products during the exponential phase, in which reagents are not limited. It is possible to make the PCR amplification efficiency close to 100% in the exponential phases of PCR reactions, if the PCR

conditions, primer characteristics, template purity, and amplicon lengths are optimal. Since the PCR product quantity in the exponential phase correlates with the initial template abundance, this enables us to compare initial abundance of template. In the real-time PCR data processing, fluorescence signal was plotted against cycle number on a logarithmic scale in which a baseline and a threshold can be set for further analysis. The PCR cycle number at which the SYBR green signal (and therefore cDNA) crosses the threshold is called the Ct. In this study, the delta-delta Ct (ddCt) method was used to quantitate alterations in mRNA levels. The dCt for each gene is calculated by subtracting the Ct number of target sample from that of endogenous gene PBGD. Then, the difference between dCt values of K overexpressed sample and vector transfected sample were calculated (ddCt). Relative quantitations (RQ) of the mRNA expression were represented as (efficiency) ^{ddCt}. Given that the amount of DNA theoretically doubles with every cycle of PCR, assuming the efficiency of amplification is 100%, RQ = 2 ^{ddCt}.

delta Ct in K overexpressed sample = Ct of target gene – Ct of PBGD (**a**)

delta Ct in myc vector reference = Ct of target gene – Ct of PBGD (**b**)

delta delta Ct = (**a**)-(**b**)

RQ= Efficiency $^{-\text{delta delta Ct}}$

RQ of K overexpressed sample divided by RQ of vector overexpressed samples represent the changed in relative gene expression affected by K overexpression (RQ ratio), value >1 indicate up regulation while value <1 indicate down regulation. The errors of each Ct were used to calculate the error for the RQ. This ddCt method is an

approximation method that assumes minimal correction for the standard gene, or that standard and target have similar efficiencies.

2.3.4. Establishment of polyclonal stable HEK293 cell line expressing KIAA0319

2.3.4.1. Cloning of Kpcdna

Subcloning of full length human KIAA0319 into pcDNA3.1 vector was performed using primers KpcdnaF / KpcdnaR as listed in Table 2-1. *EcoR*I and *Xba*I restriction sites were included to facilitate restriction enzyme digestion. Kozak Consensus sequence (G/ANNATGG) for the context of translation initiation codon is useful to optimize the gene expression in eukaryotic cells (Kozak, 2002). Hence GCCATGGCG was used in forward primer for amplifying K gene in the subcloning procedure. PCR was performed using 1 μl diluted Kmyc plasmid template. The transformed *E. coli* were spread onto LB agar plate containing ampicillin (100 μg/ml) (Amersham Biosciences). Diagnostic PCR were performed with KpcdnaF and KpcdnaR reverse primer. Plasmid constructs were checked by sequencing the entire insert.

2.3.4.2. Transfection of K-pcdna vector into HEK 293 cells

Kpcdna and pcDNA3.1 vectors were linearized in a 10 μl reaction mixture containing 7 μl of plasmid, 1 μl of 1x Buffer 3 (New England Biolabs), 10 units of *Bgl*II restriction enzyme and sterile nano pure water. The reaction mixture was incubated at 37°C for 3 hours for restriction digestion. HEK293 cells were transfected with digested and purified Kpcdna or pcDNA3.1 vectors.

2.3.4.3. Selection of the transfected HEK293 by Geneticin

Determination of Geneticin selection dosage

HEK293 cells were seeded onto a 96-well plate with $2x10^4$ per well. At 24 hours after seeding cells, cells were treated with neomycin analogue, Geneticin® Selective Antibiotic (Invitrogen) at concentration of 0, 100, 200, 400, 600, 800 and 1,000 μg/ml in complete DMEM. The diluted Geneticin was freshly prepared at each of the experiments. At 48 hours post-treatment, MTT proliferation assay was performed to determine the cell viability.

MTT assay is a quantitative colorimetric assay for the determination of cell survival and proliferation. It employs tetrazolium salts, 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide (MTT) as a substrate in measuring the activities of dehydrogenase enzymes in viable cells (Pagliacci et al. 1993). In viable cells, the tetrazolium ring of the pale yellow MTT will be cleaved to purple formazan products. The absorbance of the purple product colour could be measured with an ELISA plate reader which served as an indirect way to determine the viability of cell population.

Each well was washed with 100 μl 1X PBS twice before adding 100 μl MTT (USB Corporation) solution (3 mg/ml in PBS). The plates were then incubated for another 2 hours at 37°C, 5% CO_2. When the incubation was finished, 50 μl 10% SDS solution was added to each well to lyse the cells. The plate was gently shaken to mix the solutions. The plate was then placed in a 70°C oven for 15 min to dry up the SDS buffer. At the

end of incubation, 100 µl DMSO was added to each well. The plate was left at room temperature for 10 minutes to allow the complete dissolution of the formazan products. The absorbance at 540 nm was measured using a microplate reader (BioRad, model 3550).

Selection of stable clones

At 48 hours post-transfection with Kpcdna or pcDNA 3.1, cells were selected using Geneticin (50 mg/ml, $C_{20}H_{40}N_4O_{10}\cdot 2H_2SO_4$; Invitrogen). For the selection process, 800 µg/ml Geneticin was added to the culture medium. The selection process was continued for 2 weeks until no untransfected cells left in the control well. After 2 weeks of selection, the concentration of Geneticin was kept at 400 µg/ml to maintain selective pressure.

2.3.4.4. Determination of clones with high expression of K

Western blot

To determine the expression of K in the Geneticin-resistant stable clones, cells were rinsed with PBS and homogenized in lysis buffer (50 mM Tris-HCL pH 8, 150 mM NaCl, 1% Triton-X100) in the presence of protease inhibitors (Complete; Roche). The extracted protein lysate was analyzed by WB. SDS-polyacrylamide gel was set to 8% due to the large molecular weight of K. 10 µl Novex Sharp Pre-stained Protein Standard (Invitrogen) was loaded as a standard molecular weight marker. Mouse anti-K Monoclonal antibody G4 (see Section 2.4.3) diluted 1:500 was used as primary antibody. Goat anti-mouse IgG conjugated to HRP (1:10,000 dilution; DakoCytomation) used as

secondary antibody

Freezing polyclonal stable HEK293 cell line expressing KIAA0319

Clones with high expression of K were amplified and multiple frozen stocks were made.

2.3.5. Gene expression pattern study (protein level)

2.3.5.1. Examination of K expression in two dimensional gel electrophoresis (2D gel) samples.

To determine the expression of K in the 2D gel samples, HEK cells stably transfected with Kpcdna or pcdna vector were lysed in 2D lysis buffer (8M urea, 4% CHAPS, 2% Pharmalyte 3-10, 60 mM DTT containing protease inhibitor) on ice for 30 minutes. The precipitate in the suspension was removed by centrifugation at 13 krpm for 10 minutes at 4°C. The supernatants were assayed for K expression by WB using Mab G4 diluted 1:500 (see Section 2.4.3). Goat anti-mouse IgG conjugated to HRP (1:10,000 dilution) (DakoCytomation) was used as secondary antibody.

2.3.5.2. First dimension Isoelectric focusing (IEF)

A sample aliquot containing 4 mg of proteins were diluted in 2D rehydration solution (8M urea, 2% CHAPS, 0.5% Pharmalyte 3-10, 20 mM DTT, 0.002% bromophenol blue), making a final volume of 250 ul and then directly applied to a 13cm IEF strip having immobilized pH gradients (IPG) of 3-10 (GE Healthcare, immobilize Drystrip pH 3-10, 13 cm). Rehydration was achieved at 50 uA for 12 hours at 20°C using the Ettan IPGphor 3 system (GE Healthcare). Subsequently IEF was carried out for 60 minutes at

500 V, 60 minutes at 1,000 V and 2 to 2.5 hours at 8,000 V with a step-and-hold gradient until a total of 8,500 volt-hours had been achieved.

2.3.5.3. Second dimension SDS gel electrophoresis

IPG strips were then washed with distilled water and then equilibrated by rocking for 20 minutes at room temperature in 10ml SDS equilibration buffer (2% SDS, 50 mM Tris-HCL pH8.8, 6M urea, 30% (v/v) glycerol, 0.002% bromophenol blue) containing 10 mg/ml DTT. Then the gel was then equilibrated in 10 ml equilibration solution containing 250mg IAA for 20 minutes. Strips were then wash in SDS running buffer and placed on the top surface of the second dimension gel which was a 12% SDS polyacrylamide gel. 10ul Novex Sharp Pre-stained Protein Standard (Invitrogen) (10 µl) were applied onto small pieces of chromatography paper and inserted next to each strip on the top of each gel, after which the strips and markers were sealed with 0.5% agarose in SDS running buffer. The second dimension separation of proteins by molecular mass was achieved at a constant 200 V with cooling by running water. The gel was rinsed with distilled water twice and protein spots visualized by either silver stain or Colloidal Coomassie G-250 Staining (Dyballa & Metzger 2009)

2.3.5.4. Silver staining

Silver staining was performed using PlusOne™ Silver Staining Kit, Protein (GE Healthcare) according to the manufacturer instructions.

2.3.5.5. Colloidal Coomassie G-250 Staining

Colloidal Coomassie G-250 staining was performed using Coomassie staining solution (5% aluminium sulfate-(14-18)-hydrate, 10% ethanol, 2% orthophosphoric acid (85%), 0.02% (w/v) Coomassie Brilliant Blue (CBB) G-250) (Dyballa & Metzger 2009). The order of preparation was maintained to obtain high sensitivity as in the reference. Firstly, the aluminium sulfate (sigma) was dissolved in Milli-Q water. Thereafter, ethanol and CBB G-250 (Bio-Rad) was added. Orthophosphoric acid was added after the mixture was completely dissolved. After second dimension separation, the gels are rinsed three times with milli-Q water for 20 minutes each. Then, the gels were incubated well covered with the Coomassie solution by agitation on a shaker overnight. After the staining procedure, the Coomassie solution was removed and the gels were rinsed twice with Milli-Q water.

2.3.5.6. Computer analyses for 2D gel

A record of the positions of the visible protein spots on each gel was made by scanning the stained gel with the software ImageMaster Labscan v3.00 (Amersham Biosciences, UK). The gel images were then analyzed using ImageMaster 2D Platinum Software Version 5.0. (Amersham Biosciences, UK). Two quantitative measures are available for each protein spot: "Intensity" being the highest calibrated pixel intensity in the protein spot and "Area" representing protein spot's area in mm^2. In order to analyse gel similarities between samples, a 2D scatter plot of spot volume was done on common protein spots for the Intensity and Area [% volume (spot volume = spot area x spot intensity)] (Yin et al. 2009). A "spot-by-spot" quantitative comparison between two

treatments may be summarised as 2D scatter plots characterised by a linear fit ($y = ax + b$) and a correlation coefficient, r (Salekdeh et al. 2002). For a high correlation (similarity) between the expression of the common protein spots of the two samples, r and a should be close to unity and b close to zero (Holzmuller et al. 2008).

2.4. Chapter 4 Methods

2.4.1. Generation of polyclonal antibodies (Pab)

2.4.1.1. Subcloning of full length KIAA0319 into peGFPC1 expression vector.

N-terminally GFP tagged full length KIAA0319 expression constructs were generated using appropriate primers KgfpF / KgfpR as listed in Table 2-1 PCR was performed using Kmyc as template. A 1-base pair insertion was introduced to the forward primer making the fusion protein in frame with the GFP tag of the vector peGFPC1 (Clontech). The PCR products were cloned into the peGFPC1 plasmid using *EcoR*I and *Sal*I sites. The transformed *E.coli* were spread onto LB agar plate containing kanamycin (50 μg/ml) (Invitrogen). Diagnostic PCR were performed with K-pcmv-F and pcmv-myc specific reverse primer.

2.4.1.2. Immunizations and Screening by ELISA

Polyclonal antibodies were generated in mice using standard techniques. In brief, Balb/c mice were immunized with peptide P124 coupled to the carrier protein KLH. After multiple immunizations, mice blood was taken and the polyclonal antisera (Pab m9) analyzed by ELISA for Immunoglobulin G (IgG) recognising the peptide.

2.4.1.3. Recognition of K proteins by Pab using WB

To verify if the polyclonal antibody m9 detected the expressed K protein, WB was performed using Kpcdna protein as antigen. Pab m9 diluted 1:1,000 was used as primary antibody. Goat anti-mouse IgG conjugated to horse radish peroxidase (HRP) (1:3,000 dilution; DakoCytomation) was used as the secondary antibody.

2.4.1.4. Removal of N-glycosylation by PNGase F

WRL68 cells transfected with Kmyc were incubated in K lysis buffer (50 mM Tris-HCL pH8, 150 mM NaCl, 1% Triton-X100) containing proteinase inhibitor cocktail (Roche) on ice for 30 minutes. The lysates were then denatured with 1/10 volume of 10X glycoprotein denaturation buffer (New England Biolabs) and boiled at 100°C for 10 minutes. NP40 was added to a final concentration of 1% and 1/10 volume of 10X G7 reaction buffer (New England Biolabs) was added. The sample was split into half and 10 units of PNGase F (New England Biolabs) was added to one of the solutions. Both solutions (PNGase F treated and control) were incubated at 37 degrees for 1 hour and then subjected to WB analysis. Rabbit anti-myc Pab (SC789) 1:200 and Goat anti-rabbit IgG conjugated to HRP (1:3,000 dilution) (DakoCytomation) were used as primary and secondary antibody respectively.

2.4.2. Characterization of Pab

2.4.2.1. Pab Specificity to P124

To examine whether Pab is specific to K, synthetic peptide based neutralization was performed using P124 as a blocking peptide. Kmyc transfected or untransfected WRL68 cell lysate were used as antigen for WB. Pab m9 1:6,000 with or without blocking peptides, m9 preserum 1:6,000 or rabbit anti-myc polyclonal antibody (SC789) 1:200 were used as primary antibody. Goat anti-mouse IgG conjugated to HRP (1:3,000 dilution; DakoCytomation) and Goat anti-rabbit IgG conjugated to HRP (1:3,000 dilution; DakoCytomation) were used as secondary antibodies. The blot was re-probed with anti-β-actin IgG (Sigma) at a dilution of 1:10,000.

2.4.2.2. Pab Specificity to K and not KL

WRL 68 cell lines transfected with Kmyc, KLmyc or pCMV-myc vector were used as antigen for WB. The blot was probed with Pab m9 at dilution 1:3,000 or rabbit anti-KL polyclonal antiserum (H635) at dilution 1:1,000. Goat anti-mouse IgG conjugated to HRP (1:3,000 dilution) (DakoCytomation) and Goat anti-rabbit IgG conjugated to HRP (1:3,000 dilution) (DakoCytomation) were used as secondary antibodies. The blot was re-probed with anti-β-actin IgG (Sigma)

2.4.2.3. Pab detects K protein *in vivo*

For detection of the recombinant KIAA0319 protein, IF was performed. WRL68 cells were transfected with 2 μg of the Kmyc expression plasmid, using Lipofectamine and PLUS reagent (Invitrogen). Pab m9 (dilution 1:400) and SC789 (dilution 1:100) were

used as primary antibodies. Fluorescein (FITC) AffiniPure F(ab')2 Frag Donkey Anti-Rabbit IgG (H+L) (Jackson ImmunoResearch, USA) at a 1:200 dilution or Rhodamine (TRITC) AffiniPure Goat Anti-Mouse IgG (H+L) (Jackson ImmunoResearch :115025146) at 1:200 dilution were used as secondary antibodies. The nuclei of the cells were stained with 25 μg/ml Hoechst 33342 (Molecular Probes) for 5 minutes at room temperature.

For immunofluorescence of the endogenous K protein, SHSY-5Y cell line and WRL68 cell line were grown on glass coverslips. The next day, cells were fixed and permeabilized as mentioned above. Cells were stained with Pab m9 at a 1:400 dilution. Rhodamine (TRITC) AffiniPure Goat Anti-Mouse IgG (H+L) at 1:200 dilution or DyLight 488 AffiniPure Goat Anti-Mouse IgG (H+L) (JACKSON) at a 1:200 dilution were used as the secondary antibody. The nuclei of the cells were stained with 25 μg/ml Hoechst 33342 (Molecular Probes) or 20 μg/ml propidium iodide (PI) (Molecular Probes) .

2.4.2.4. Subcellular fractionations

The subcellular fractions of WRL68 cells overexpressing K protein were prepared by using a ProteoExtract subcellular proteome extraction kit (Calbiochem, San Diego, CA) according to the manufacturer's protocol. Extracted proteins were analyzed by WB using 10% polyacrylamide gel. Rabbit anti-myc Pab (SC789) 1:200 and Goat anti-rabbit IgG conjugated to HRP (1:3,000 dilution; DakoCytomation) were used as primary and secondary antibodies respectively. To monitor the purity of the fractions, histone H3,

GAPDH, Calnexin and Vimentin were assessed as nuclear, cytosolic, membrane/organelle and cytoskeletal specific markers, respectively.

2.4.3. Generation of monoclonal antibody (Mab)

The technique employed to produce Mab involved the fusion of mouse myeloma cells with B-lymphocytes that had been induced to produce specific antibodies. The outline to this approach is shown in Figure 2-1.

2.4.3.1. Immunization

Monoclonal antibodies were generated in mice by immunizations for generating polyclonal antibody. After repeated immunizations, mouse bloods were collected and subjected to ELISA and WB analysis. Once a ELISA and WB positive mouse was obtained, fusion was performed (as described in the following section). Three days before fusion, mouse was immunized with 15 ug of P124 diluted in PBS (booster immunization).

2.4.3.2. Fusion

We employ the chemical fusion technique with polyethylene glycol solution (PEG) (Sigma), which acts as a fusogen, for creation of hybridoma cells with mouse splenocytes according to the fusion protocol previously described (Köhler & Milstein 1975). The fusion partner cell line NS0 mouse myeloma cells (Keen & Steward 1995) was maintained in complete DMEM along with 8-azaguanine for one week. One day before fusion, NS0 were split to 2 x 10^5 cells /ml to ensure that the cell line is in the best

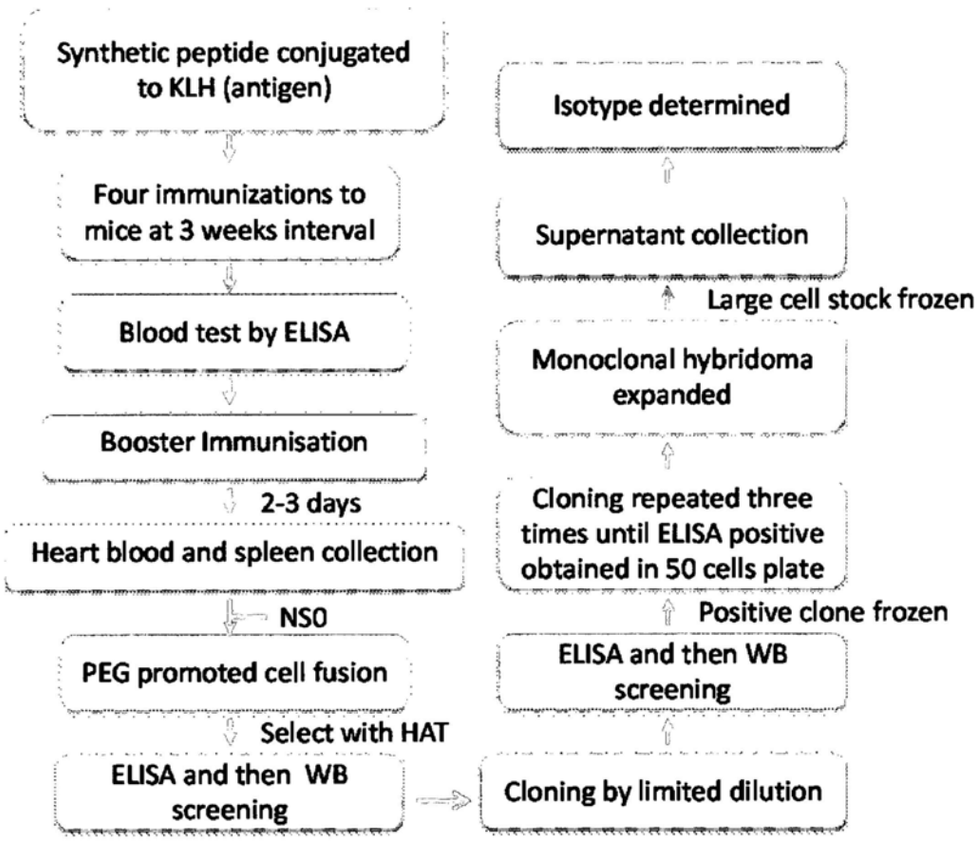


Figure 2-1. Flow diagram illustrating monoclonal antibody production by hybridoma approach.

possible condition prior to the fusion.

For production of hybridoma cells, mouse was sacrificed and splenocytes isolated by gently pushing the spleen through a mesh. Heart blood was also collected as polyclonal antibody (Pab mB). Splenocytes and the fusion partner cells were washed twice in DMEM and counted. Next, splenocytes and NS0 were mixed in 5:1 ratio and washed twice in serum free DMEM. The cell mixtures were centrifuge at 1,500 rpm for 5 minutes to pellet. Pre-warmed PEG was added to mixed cells in a drop-wise manner. The cell mixture was incubated with PEG for 1 minute with constant shaking. PEG was then diluted out of the cell mixture with addition of 4ml pre warmed DMEM in a drop-wise manner. 10ml pre warmed DMEM was added and the suspension incubated at 37 °C in 5% CO2 for five minutes. The hybridoma cells that were generated in this process were cultured in 96-well plates in complete DMEM at a density of $2x10^5$ cells/well. One day following the fusion, equal volume of two times concentration of Hypoxanthine Aminopterin Thymidine (HAT, Gibco) in complete medium was added, making the final concentration 1x HAT. This allowed the hybridoma cells to be distinguished from the non-fused cells (Figure 2-2). The non-fused myeloma cells, lacking an essential enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT) therefore died in the presence of aminopterin as the metabolism pathway of nucleotide production was blocked. The non-fused spleen cells do contain HGPRT enzyme but they perish due to a limited life span. Consequently only the fused hybridoma cells that acquired the HGPRT enzyme from spleen cells and immortality from myeloma cells are capable of survival and divide after addition of HAT media. To ensure that the cells were

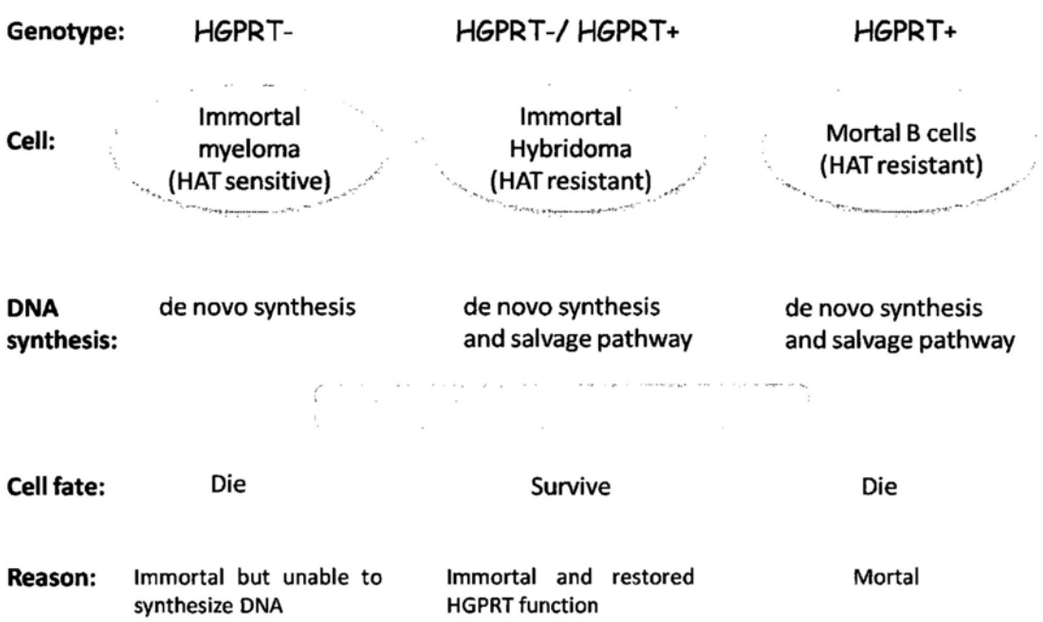


Figure 2-2. Cell fate upon HAT selection. Normal cells use *de novo* synthesis and salvage pathway for DNA synthesis. Immortal myeloma cells (NS0), deficient in the enzyme hypoxanthine-guanine phosphoribosyltransferase (HGPRT), cannot use the alternative salvage pathway and depends on *de novo* synthesis of nucleic acids. Aminopterin in HAT medium inhibit dihydrofolate reductase and therefore block DNA *de novo* synthesis. NS0 cells therefore die in the presence of aminopterin as they cannot synthesize nucleic acids. B cells survive HAT selection as they are HGPRT+ but they die due to limited life span. The only cells that survive are hybridoma, which acquire the immortality from NS0 and the HGPRT from B cells.

well nurtured, the medium was removed by aspiration, and replaced with freshly prepared media every few days. Aminopterin in HAT media slows down cell growth, therefore after selection of the fused cell was completed (approximately 2 weeks); the HAT media was gradually switched to hypoxanthine/ thymidine (HT; Gibco) media. After the pool of aminopterin in the cell was depleted, the media was eventually changed to complete DMEM with double amount of FBS to enhance cell growth

2.4.3.3. Screening by ELISA and WB

The screening of the supernatants for immunoglobulin production was performed after the hybridoma cells occupied approximately 1/4 of the well. ELISA coated with BSA conjugated P124 was used to screen hybridoma supernatants for IgG. Undiluted hybridoma supernatants were used as primary antibody. Mouse blood diluted in PBS (1:1,000) and hybridoma growth medium were included as positive and negative control respectively. ELISA positive clones were then subjected to WB analysis using Kpcdna or Kmyc overexpressed HEK cell lysate as antigen to verify if the supernatants secreted by the hybridoma cells detected the expressed protein. Hybridoma supernatants diluted 1:5 was used as primary antibody. Goat anti-mouse IgG conjugated to HRP (1:3,000 dilution) (DakoCytomation) were used as secondary antibody. The best hybridoma cell line for carrying out cloning could therefore be determined.

2.4.3.4. Cloning

Cloning was carried out to obtain a homogeneous cell clone secreting antibodies. Strong ELISA and western blot positive clones were sub-cloned by serial limiting dilution in

96-well plates at 3000, 500, and 50 cell-per-plate density in complete DMEM containing HT. The 0.5 cell-per-well limiting dilutions were performed twice to ensure that we generated clones. Thymus cells collected from 18-20 day old mice were used as feeder cells to plate out together with the hybridomas to support cell growth of single hybridoma cell. To ensure that the cells were well nurtured, the medium was removed by aspiration, and replaced with freshly prepared media every few days. Screening with ELISA and WB were performed after the hybridoma cells occupied approximately 1/4 of the wells. Undiluted hybridoma supernatants were used as primary antibody for ELISA. Kmyc overexpressed HEK293FT cell lysate were used as antigen for WB. Hybridoma supernatants diluted 1:5 was used as primary antibody. Goat anti-mouse IgG conjugated to HRP (1:3,000 dilution; DakoCytomation) was used as secondary antibody. Positive clones were frozen. Further cloning was required until an ELISA positive reading was obtained in all the wells of the 50 cells plate that shown equal or less than 30% of cell growth. In this case there was a high probability that the antibodies in the well would be monoclonal since the original cell concentration was less than a hybridoma cell per well. One further cloning can be performed to confirm that the hybridoma is fully cloned. Several frozen stocks were made for the monoclonal secreting hybridoma cells (Mab G4).

2.4.3.5. Freezing and thawing hybridoma cells

As a precaution, ELISA positive cells were frozen in each step of cloning. The fully cloned monoclonal cell lines were also frozen multiple times. Cells were resuspended in FBS and mixed with an equal amount of freezing mix. Occasionally, Cells were

recovered from nitrogen when experimental error occurred or unsuitable hybridoma cells were selected for cloning. Due to a low cell density when freezing, feeder cells were added to the recovered cells to improve cell growth. They were then plated together with feeder cells onto a 96 well plate.

2.4.3.6. Expand and harvest Mab

The fully cloned monoclonal cells were expanded into 24-well plates and then to T75 culture flask in complete DMEM. Once an appropriate amount of cells were generated, larger stock of cells were frozen in liquid nitrogen leaving some to grow for a period of days without addition of fresh media. The supernatant was next clarified by centrifugation at 4,000 rpm for 10 minutes to collect the Mab rich supernatants that subsequently were frozen at –20°C until needed.

2.4.3.7. Isotyping

The Mab G4's isotype was confirmed by ELISA. Primary antibody including Mab G4, mouse anti-gfp antibody , normal mouse IgG, Pab mB and 5% BSA were incubated with p124 coated ELISA plate at 4°C overnight. Isotype specific secondary antibody Goat anti-Mouse IgM – HRP (Invitrogen) and Goat anti-Mouse IgG (H+L) – HRP (Invitrogen) at dilution 1:1,000 were incubated with the plate at 37°C for 2 hours. Substrate PNPP was added and the plate incubated at 37°C for 30 minutes.

.

2.4.4. Characterization of Mab

2.4.4.1. Mab Specificity to P124

The specificity of Mab G4 to P124 was determined by synthetic peptide based neutralization using P124 as blocking peptide. Kpcdna or pcDNA 3.1 transfected HEK293FT cell lysates were used as antigen for WB. Mab G4 1:100 with or without blocking peptides, Pab mB at dilution 1:1,000, and mB preserum 1:1,000 were used as primary antibody. Goat anti-mouse IgG conjugated to HRP (1:3,000; DakoCytomation) was used as secondary antibody. All blots were re-probed with Mab G4 and anti-β-actin IgG (Sigma)

2.4.4.2. Mab Specificity to K and not KL

HEK293FT cell lines transfected with Kpcdna, pcDNA 3.1 vector, Kmyc, KLmyc, pCMV-myc vector, KLgfp and Kgfp were used as antigen for WB. The blot was probed with Mab G4 (1:100). Goat anti-mouse IgG conjugated to HRP (1:3,000; DakoCytomation) were used as secondary antibody. The blot was reprobed with mouse monoclonal IgG2a anti-GFP tag (B-2) antibody (1:5,000; Santa Cruz) and anti-β-actin IgG (1:10,000; Sigma).

2.4.4.3. Mab detects native K protein *in vivo*

For detection of the recombinant KIAA0319 protein, IF was performed. HEK293FT cells were transfected with 1 μg of the Kpcdna expression plasmid, using Lipofectamine and PLUS reagent (Invitrogen). Mab G4 at a 1:10 dilution with or without pre-incubation of blocking peptide P124 was used as primary antibody. Goat anti-mouse

IgG conjugated to Dylight 488 (JACKSON) at a 1:100 dilution was used as secondary antibody. The nuclei of the cells were stained with 20 μg/ml PI (Molecular Probes).

2.4.4.4. Mab capable for IP

To examine if the Mab can recognize native K protein and capable for IP analysis, human kidney HEK293FT cells were transfected with Kpcdna, Kgfpc1, pegfpc1 vector or pcDNA 3.1 vector. The cleared cell lysates were incubated with 100 ul of Mab G4 and 50 μl of Protein G Sepharose 4 Fast Flow beads (GE Healthcare) for overnight at 4°C. The presence of K protein in the immunocomplexes were either stained by Coomassie blue or analyzed by WB using anti-K Pab m9 (1:5000). Goat anti-mouse IgG conjugated to HRP (1:3,000; DakoCytomation) were used as secondary antibody

2.4.4.5. Mab production in serum free conditions

Mab was generated in serum free condition according the protocol described previously (Basalp et al. 2000). Briefly, 1 x 10^7 hybridomas growing in 10% FBS in T75 cell culture flask were passed in FBS-free media for two days, then replacement by 1% FCS and cultivation for two days, finally replacement by serum-free medium and cultivation for two days. For each step, 100 ml DMEM was used. The presence of anti-K antibody was verified by WB using HEK293 FT cells transfected with Kpcdna or pcDNA 3.1. Monoclonal supernatants harvest 1 to 6 at dilution 1:50 were used as primary antibody. Pab mB at dilution 1:1;000 was included as positive control. Goat anti-mouse IgG conjugated to HRP (1:3,000; DakoCytomation) were used as secondary antibody. Dye front was not run off.

2.5. Chapter 5 Methods

2.5.1. Interaction of K and KL

2.5.1.1. Co-IP of full length K and KLgfp

To investigate the binding of untagged K and KLgfp proteins, HEK293FT cells were transfected with 6 μg of Kpcdna and KLgfp. The cleared cell lysates were incubated with 50μl Mab G4 for 2 hours at 4°C. The immunocomplexes were absorbed onto 50 μl of Protein G Sepharose 4 Fast Flow beads (GE Healthcare) for overnight at 4°C. The resulting mixtures were analyzed by WB. The precipitated KL protein was detected using mouse monoclonal IgG2a anti-GFP tag (B-2) antibody at dilution 1:5,000 (Santa Cruz). Goat anti-mouse IgG conjugated to HRP (1:3,000 dilution) (DakoCytomation) was used as secondary antibody. The presence of K protein was detected by re-probe with Mab G4 (1:100). Mouse heavy chain IgG was used as loading control.

2.5.1.2. GFP does not detect KpcDNA

Kpcdna, pcDNA 3.1 vector, KLgfp or Kgfp transfected HEK293FT were used as antigen for WB. Mouse monoclonal IgG2a anti-GFP tag (B-2) antibody (Santa Cruz) at dilution 1:5000 was used as primary antibody. Goat anti-mouse IgG conjugated to HRP (1:3,000; DakoCytomation) was used as the secondary antibody. The blot was re-probe with Mab G4 at dilution 1:500 and anti-β-actin IgG (Sigma) at dilution 1:10,000.

2.5.1.3. Co-IP of full length Kgfp and KLmyc

HEK293FT cells were transfected with 6 μg of Kgfp and KLmyc. The cleared cell lysates were incubated with 5 μl of rabbit anti-KL antibody (H635) or normal rabbit IgG for 2 hours at 4°C. The immunocomplexes were absorbed onto 50 μl of Protein G Sepharose 4 Fast Flow beads (GE Healthcare) for overnight at 4°C. The resulting mixtures were analyzed by WB. The K protein was detected by Pab m9 at dilution 1:5,000. Goat anti-mouse IgG conjugated to HRP (1:3,000; DakoCytomation) was used as secondary antibody. The blot was re-probe with H635 at dilution 1:1,000 to detect KL and goat anti-rabbit IgG conjugated to HRP (1:3,000; DakoCytomation) was used as secondary antibody.

2.5.1.4. Co-IP of full length Kgfp and KLmyc with peptide based neutralization

HEK293FT cells were transfected with 6 μg of Kgfp and KLmyc. The cleared cell lysates were incubated with 5 ul mouse anti-GFP antibody for 2 hours at 4°C. The immunocomplexes were absorbed onto 50 μl of Protein G Sepharose 4 Fast Flow beads (GE Healthcare) for overnight at 4°C. The resulting mixtures were analyzed by WB. The KL proteins were detected by H635 (1:1,000 dilution) with or without KL blocking peptide. Goat anti-rabbit IgG conjugated to HRP (1:3,000 dilution; DakoCytomation) was used as secondary antibody. The membrane was reprobed with Pab m9 (1:5,000 dilution) for detection of K protein and goat anti-mouse IgG conjugated to HRP (1:3,000 dilution; DakoCytomation) was used as secondary antibody.

2.5.2. Co-localization of endogenous K and KL

For immunofluorescence of the endogenous K and KL protein, human glioblastoma-astrocytoma cells U87MG cells were seeded on glass coverslips in a 6-well plate. Cells were co-stained using a mouse Pab m9 (dilution 1:100) and anti-KL antibodies H635 (dilution 1:100) for 4°C overnight. DyLight 488 AffiniPure Goat Anti-Mouse IgG (H+L) (Jackson ImmunoResearch) and Rhodamine (TRITC) AffiniPure Goat Anti-Mouse IgG (H+L) (Jackson ImmunoResearch) (dilution 1:100) were used as secondary antibodies.

2.5.3. Interaction of K and KL deletion mutant

2.5.3.1. Construction of K and KL deletion mutants

K and KL deletion mutants were generated by PCR amplification using appropriate primers listed in Table 2-1 and Table 2-3. Namely, KΔN mycF / KΔN mycR, KΔNgfpF / KΔNgfpR, KΔCmycF / KΔCmycR, KΔCgfpF / KΔCgfpR, KΔTMmycF, KΔTMmycR, KΔTMgfpF / KΔTMgfpR, KLΔNmycF / KLΔNmycR, KLΔNgfpF / KLΔNgfpR, KLΔCmycF / KLΔCmycR, KLΔCgfpF / KLΔCgfpR, KLΔTMmycF / KLΔTMmycR and KLΔTMgfpF / KLΔTMgfpR. Full length Kmyc or KLgfp were used as PCR template. The PCR products were subcloned into the pCMV-myc plasmid (Clontech) and peGFPC1 (Clontech) using appropriate restriction enzymes including *EcoR*I, *Sal*I, *Not*I and *BamH*I. The transformed *E.coli* were spread onto LB agar plate containing Ampicillin (100 ug/ml) (Amersham Biosciences) or kanamycin (50 μg/ml) (Invitrogen) for pCMV-myc and peGFPC1 vectors respectively. The details of deleted amino acids in the deletion constructs were listed in Table 2-4.

Table 2-3. List of primers for subcloning KL

Subcloning primers were designed based on the sequence of KIAA0319-Like transcript variant 1 (GenBank Accession Number: NM_024874.3). Each of the primers included a 5' end clamp sequences (TAGGGC or TATGCA) shown in **bold**, restriction sites indicated by underline, extra bases indicated by dots and initiating or stop codon to make the sequence in frame to the reading sequence of the vector.

Construct	Primer Name	Sequence 5'-->3'	Restriction Sites
KLΔNmyc	KLΔNmycF	**TATGCA**GTCGACAATGTTCCAAACT GCAGATGAT	*Sal*I
	KLΔNmycR	**TATGCA**GCGGCCGCCTACAGGATCT CCTCCCG	*Not*I
KLΔNgfp	KLΔNgfpF	**TATGCA**GTCGACATGTTCCAAACTG CAGATGAT	*Sal*I
	KLΔNgfpR	**TATGCA**GGATCCCTACAGGATCTCC TCCCG	*Bam*HI
KLΔCmyc	KLΔCmycF	**TATGCA**GTCGACAATGGAGAAGAG GCTGGG	*Sal*I
	KLΔCmycR	**TATGCA**GCGGCCGCTTAGTCTGCCT TTTGCTTCCGCAGCT	*Not*I
KLΔCgfp	KLΔCgfpF	**TATGCA**GTCGACATGGAGAAGAGG CTGGG	*Sal*I
	KLΔCgfpR	**TATGCA**GGATCCTTAGTCTGCCTTT TGCTTCCGCAGCT	*Bam*HI
KLΔTMmyc	KLΔTMmycF	**TATGCA**GTCGACAATGGAGAAGAG GCTGGG	*Sal*I
	KLΔTMmycR	**TATGCA**GCGGCCGCTTAACAGTTGC TGTCTCC	*Not*I
KLΔTMgfp	KLΔTMgfpF	**TATGCA**GTCGACATGGAGAAGAGG CTGGG	*Sal*I
	KLΔTMgfpR	**TATGCA**GGATCCTTAACAGTTGCTG TCTCC	*Bam*HI

Table 2-4. Lists of K and KL deletion constructs used in p.130's western blots (Figure 5-5).

	Plasmid	Deleted aa	Short name
Kpcdna	pcDNA-3.1	--	
Kmyc	pcmv-myc	--	
Kgfp	peGFPC1	--	
KΔN myc	pcmv-myc	1-105	A
KΔNgfp	peGFPC1	1-105	
KΔCmyc	pcmv-myc	900-1072	C
KΔCgfp	peGFPC1	900-1072	
KΔTMmyc	pcmv-myc	953-1072	E
KΔTMgfp	peGFPC1	953-1072	
KLmyc	pcmv-myc	--	
KLgfp	peGFPC1	--	
KLΔNmyc	pcmv-myc	1-133	B
KLΔNgfp	peGFPC1	1-133	
KLΔCmyc	pcmv-myc	874-1049	D
KLΔCgfp	peGFPC1	874-1049	
KLΔTMmyc	pcmv-myc	927-1049	F
KLΔTMgfp	peGFPC1	927-1049	

2.5.3.1. Co-IP of deletion mutants

To investigate the binding site where KIAA0319 and KIAA0319-like proteins interact, human kidney HEK293FT cells were co-transfected with 6 μg of KLgfp, KΔN myc, KΔCmyc, Kgfp, KLΔNmyc, KLΔCmyc, KΔTMmyc, KLΔTMmyc or pegfpc1 vector. Co-IP was performed with 5ul of mouse anti-GFP antibody (GFP). The presence of K was detected by WB with Pab m9 (1:5000) or Mab G4 (1:100). KL was detected with H635 (1:1000) or mouse monoclonal IgG2a anti-GFP tag (B-2) antibody (1:5,000; Santa Cruz). Goat anti-mouse IgG conjugated to HRP (1:3,000; DakoCytomation) and goat anti-rabbit IgG conjugated to HRP (1:3,000; DakoCytomation) were used as secondary antibody

2.6. Chapter 6 Methods

2.6.1. Interaction of AP2M1 with full length K and KL

2.6.1.1. Cloning of AP2M1 into pCMV-myc expression vector

To generate N-terminally myc tagged full length human AP2M1 expression constructs, the ORF encoding AP2M1 were amplified using AP2M1mycF / AP2M1mycR (Table 2-5). cDNA generated in SHSY5Y cell line was used as template. The PCR products were cloned into the pCMV-myc plasmid (Clontech) using *SalI* and *NotI* sites, generating AP2M1myc construct. A 1-base pair insertion was introduced to the forward primer making the fusion protein in frame with the myc tag of the vector. The transformed *E. coli* were spread onto LB agar plate containing Ampicillin (100ug/ml) (Amersham Biosciences). Diagnostic PCR were performed with AP2M1mycF and

Table 2-5. List of primers for subcloning AP2M1, SH2 and FEM
Subcloning primers were designed based on the sequence of Adaptor-related protein complex 2 mu 1 subunit (GenBank Accession Number: NM_004068), SH2 (GenBank Accession Number: NM_015503.1) and FEM (GenBank Accession Number: NM_015322.3). Each of the primers included a 5' end clamp sequences (TAGGGC or TATGCA) shown in **bold**, restriction sites indicated by underline, extra bases indicated by dots and initiating or stop codon to make the sequence in frame to the reading sequence of the vector.

Construct	Primer Name	Sequence 5'-->3'	Restriction Sites
AP2M1myc	AP2M1mycF	**TATGCA**GTCGACAATGATTGGAGGCTT	*Sal*I
	AP2M1mycR	**TATGCA**GCGGCCGCCTAGCAGCGAGTTT	*Not*I
SH2myc	SH2mycF	**TATGCA**GAATTCTTATGAATGGTGCCCCTTCCCCAGAGGACG	*Eco*RI
	SH2mycR	**TATGCA**GTCGACTCATGGGAGGTGGTCGGTTACACAGTC	*Sal*I
FEMmyc	FEMmycF	**TATGCA**GAATTCTTATGGAGGGCCTGGCT GGCTA	*Eco*RI
	FEMmycR	**TATGCA**GTCGACTTAATGAAATCCAACAAACT	*Sal*I

pCMV-myc specific reverse primer.

2.6.1.1. Co-IP of K and AP2M1

HEK293FT were transfected with AP2M1myc, K-pcdna, K-GFP and pCMV-myc vector. Co-IP was performed with 30ul rabbit anti-myc polyclonal antibodies (SC789) for 2 hours at 4oC. The immunocomplexes were absorbed onto 50 µl of Protein G Sepharose 4 Fast Flow beads (GE Healthcare) for overnight at 4°C. The resulting mixtures were analyzed by WB. The K protein was detected by Mab G4 (1:100 dilution) or mouse monoclonal IgG2a anti-GFP tag (B-2) antibody (1:5000; Santa Cruz). Goat anti-mouse IgG conjugated to HRP (1:3,000; DakoCytomation) were used as secondary antibody

2.6.1.2. Co-IP of KL and AP2M1

HEK293FT were transfected with AP2M1myc, Kgfp , KLgfp and pCMV-myc vector. Co-IP was performed with 30ul rabbit anti-myc polyclonal antibodies (SC789) for 2 hours at 4°C. The immunocomplexes were absorbed onto 50 µl of Protein G Sepharose 4 Fast Flow beads (GE Healthcare) for overnight at 4°C. The resulting mixtures were analyzed by WB using mouse monoclonal IgG2a anti-GFP tag (B-2) antibody (1:5,000; Santa Cruz). Goat anti-mouse IgG conjugated to HRP (1:3,000; DakoCytomation) was used as secondary antibody.

2.6.1.3. Docking of K and KL peptides into the binding site of AP2M1

Docking software

AutoDock is a suite of automated docking tools designed to predict how small molecules, such as substrates or drug candidates, bind to a receptor of known 3D structure. The AutoDock version 4.2 was used for docking in this study (Morris et al. 1998), based on a hybrid search method that applied a Lamarckian genetic algorithm (Imberty et al. 2008). The genetic algorithm begins by generating a random population of individuals which uniformly explore the grid space, followed by a specified number of generation cycles, each one consisting of a mapping and fitness evaluation, a selection, a crossover, a mutation and an elitist selection. After this step, each generation cycle is followed by the local search. At the end of every docking, AutoDock reports the docked energy, the state variables, the coordinates of docked conformation and the estimated free energy of binding (Imberty et al. 2008).

Preparation of macromolecule

The atomic coordinates of AP2M1 (PDB code: 1H6E) was retrieved from Protein Data Bank. This structure was determined by X-ray crystallography to 3.6 Å resolutions. For the preparation of protein in docking simulation, water molecules and its original ligand peptide were removed. Polar hydrogen atoms were added to the protein, and the united atom Kollman charges were assigned for the protein.

Preparation of peptides

The three dimensional structures of peptides corresponding to 994-999 amino acid of K

(KYTILD) and 967-972 amino acid of KL (KYKILD) were drawn by using the chemical modeling software SKETCHER in CCP4 suit and designated as the ligands for the docking. These regions were selected as they are the only cytoplasmic AP2 binding motif. All hydrogen atoms were added and the united atom Gasteiger-Marsilli charges were added to the ligand. Atom types and bond types of the ligand molecule were inspected and corrected manually.

Docking parameters

The graphical user interface program AutoDock version 4.2 (Morris et al. 1998) was employed to prepare, run, and analyze the docking simulations to generate an ensemble of docked conformations for each ligand molecule.

The rigid roots of each ligand were defined automatically and amide bonds were made non-rotatable. AutoTors program was used to assign all rotatable dihedrals in the ligand. The numbers of rotatable bonds of the K and KL ligand peptide were 14 and 15 respectively. AutoDock requires pre-calculated three dimensional grid maps, one for each type of atom present in the ligand and it stores the interaction energy based on a macromolecular target using the AMBER force field. The size of the docking grid box (x, y, z) was assigned to be 56, 40, and 46 Å which was centered at the experimentally observed position of the ligand. The grid box was large enough to enclose the largest docked ligand of the entire test set. The grid spacing inside the docking box was set to be 0.375 Å. The default values were assigned for other parameters. In the study, the receptor (AP2M1) is held rigid while the ligand is free to rotate, translate and change

conformation during the docking application.

The genetic algorithm with local search (GALS) was applied to the model for the interaction/binding between the target macromolecule and the docked ligand. The population size was set to 300 and the individuals were initialized randomly. The maximum number of energy evaluation was set to 250,000; maximum number of generations was set to 27,000. The maximum number of top individual that automatically survived was set to one. The rates of gene mutation and crossover were set at 0.02 and 0.80 respectively. The local search frequency was set to 0.06. The default values were assigned for other parameters.

100 individual GALS runs were conducted to generate 100 docked conformations for each ligand. Cluster analysis was performed on the docked results using a root mean square (RMS) tolerance of 2.0 Å. The clusters were ranked from the averaged lowest energy group obtained for the cluster members to the highest. The analysis was carried out for the all docking clusters. The ligands were compared based on the best binding free energy obtained among all the docking runs. The conformation with the lowest final docked energy given by AutoDock was chosen.

2.6.2. Interaction of full length K with SH2 and FEM

2.6.2.1. Cloning of SH2 and FEM into pCMV-myc expression vector

To generate N-terminally myc tagged full length human SH2 and FEM expression constructs, the ORF encoding SH2 and FEM were amplified using primers SH2mycF /

SH2mycR and FEMmycF / FEMmycR (Table 2-5). A 2-base pair insertion was introduced to the forward primer making the fusion protein in frame with the myc tag of the vector. PCR was performed using 1 μl diluted fg04370 or hg00180s1 plasmid template obtained from Kazusa DNA research institute, which contains the full-length human SH2 or FEM cDNA respectively. The PCR products were cloned into the pCMV-myc plasmid (Clontech) using *Eco*RI and *Sal*I sites, generating SH2myc and FEMmyc constructs respectively. The transformed *E.coli* were spread onto LB agar plate containing Ampicillin (100 ug/ml; Amersham Biosciences). Diagnostic PCR were performed with SH2mycF or FEMmycF together with pCMV-myc specific reverse primer.

2.6.2.2. Co-IP of Kgfp and SH2myc or FEMmyc

HEK293FT were co-transfected with Kgfp, SH2myc, FEMmyc and pCMV-myc vector. Co-IP was performed with 30ul rabbit anti-myc polyclonal antibodies (SC789) for 2 hours at 4°C. The immunocomplexes were absorbed onto 50 μl of Protein G Sepharose 4 Fast Flow beads (GE Healthcare) for overnight at 4°C. The resulting mixtures were analyzed by WB. The K protein was detected by mouse monoclonal IgG2a anti-GFP tag (B-2) antibody (1:5,000; Santa Cruz). Goat anti-mouse IgG conjugated to HRP (1:3,000; DakoCytomation) were used as a secondary antibody.

CHAPTER 3

Study of gene expression profile of cells overexpressing KIAA0319

3.1 Expression level of K is critical to neuronal migration

DD is a one of the most common neurobehavioral disorders that greatly affect the reading ability in school age children. Numerous studies have been conducted in attempt to understand DD more fully and thus provide treatment and support to affected individuals. Independent linkage and association studies have demonstrated that K has a strong association with dyslexia (Couto et al. 2009; Luciano et al. 2007; Cope et al. 2005; Francks et al. 2004; Harold et al. 2006). Great advancement has been achieved with functional analyses of K using RNA interference in rat, which demonstrated that this gene plays an important role in neuronal migration during development (Paracchini et al. 2006). Nevertheless, the precise mechanism of how it might function in brain development and hence its role in dyslexia remains unknown. The SNPs that links K with dyslexia does not occur in the protein-coding sequence, therefore the functional DNA change(s) must be present within regulation region and affects the level of K protein (Dennis et al. 2009). Actually, studies show that in the dyslexia risk haplotype, the expression of K is reduced (Harold et al. 2006; Paracchini et al. 2006). RNAi study shows that K downregulation would lead to impaired neuronal migration in mice (Paracchini et al. 2006; Peschansky et al. 2009) while overexpression of K alters the

position of neurons within cortical lamina (Peschansky et al. 2009). This is the first proof that the level of K protein affects brain development and hence provides a link between K and dyslexia. Yet, how altered expression of K would affect neuronal migration remains mysterious. Indeed, only a small portion of animal proteins are sufficiently characterized, about 40% of the proteins encoded in eukaryotic genomes are proteins of unknown function. Their functional characterization remains one of the main challenges in modern biology.

3.2 Characterization of K up-regulation on gene expression

Altered K expression would leads to impaired neuronal migration, which leads to a hypothesis that K influence the proteins involved in neuronal migration either by regulating their expression or by interacting with them. Although K does not localize into nucleus or harbor any DNA binding domain, it possibly regulate gene expression by influencing signaling, which is consistent to the role of its secreted isoform. Also, it may function as a receptor that gives signal to downstream effectors and leads to changes in gene expression. Therefore, to determine if K would affect neuronal migration by modulating expression of genes and to associate K with biological processes of known function, gene expression analyses were desirable.

Gene expression is tightly regulated in all known life as it increases the versatility and adaptability of an organism by allowing the cell to express protein when needed. Several steps in the gene expression process may be modulated, including the transcription, RNA splicing, translation, and post-translational modification of a protein. Regulation of

transcription controls when transcription occurs and how much RNA is created. Post-transcriptional regulation is the control of gene expression at the RNA level, which include modulating the capping, splicing, addition of a Poly(A) Tail, the sequence-specific nuclear export rates and in several contexts sequestration of the RNA transcript. To assess if there are any difference in the effect of K on gene regulation at the mRNA level and the protein level, gene expression analyses were performed both at transcription and translation level. DNA microarray was performed in order to examine the effect of altered K expression on the transcript levels for numerous genes at once (expression profiling). To study whether or not the deregulation of K would affect gene expression profile at the protein level, overexpression of K followed by two-dimensional gel electrophoresis (2D gel) was conducted. Comparing the transcriptomes and proteinomes in K up-regulated samples would give us a clue on how K associated to neuronal migration and Dyslexia, either by regulating other genes expression or by interacting with other proteins.

3.2.1 Cloning of K into various mammalian expression vectors

In order to characterize K's functions, it is essential to overexpress K and obtain K proteins. To achieve this, full length human K cDNA were cloned into mammalian expression vector pCMV-myc, producing N-terminal myc tagged full length K protein. Full length Human KIAA0319 cDNA clone was obtained from Kazusa DNA research institute. Because the template has three nucleotides different from the NCBI data base, site-directed mutagenesis was performed to change it back to the NCBI reference sequence (NM_014809). The resulting N-terminal myc tagged full length K (Kmyc)

vector was used as template for subsequent subcloning of K, to overexpress K, to examine K's localization and to study its interactions with other proteins.

3.2.2 KIAA0319 does not lead to major effect of other genes' expression (mRNA level)

Oligonucleotide-based microarrays were used to conduct the comparative transcriptomic studies on the effect of K overexpression, Kmyc vector and pCMV-myc vector were used to overexpresses K protein in human neuroblastoma cell line SHSY5Y. One day following transfection, total RNA were extracted in the two samples. DNAse I treatment was used to remove genomic DNA contamination. Total RNA was converted to cDNA and the expression of K was analyzed by Real time PCR (Figure 3-1A).

After confirming K overexpression by Real time PCR, the RNA was sent to Agilent for performing DNA microarray using Agilent human whole genomic oligonucleotide microarrays (G4112F) with two colour labeling (n = 1).

The result obtained from Agilent were normalized and analyzed by GeneSpring (Agilent). Genes that were down regulated more than 4 fold or up regulated more than 2 fold in microarray were selected for validation by Real time PCR (Table 3-1). Among the down regulated genes, we found KL, the homologue of K. Although it was down regulated only 2.5 fold in microarray, given that it is K's homologue, it remains an interesting target and we also included it in our Real time PCR analysis. To perform Real time PCR, primers were designed using ProbeFinder software v2.45 (Roche) (Table 3-1).

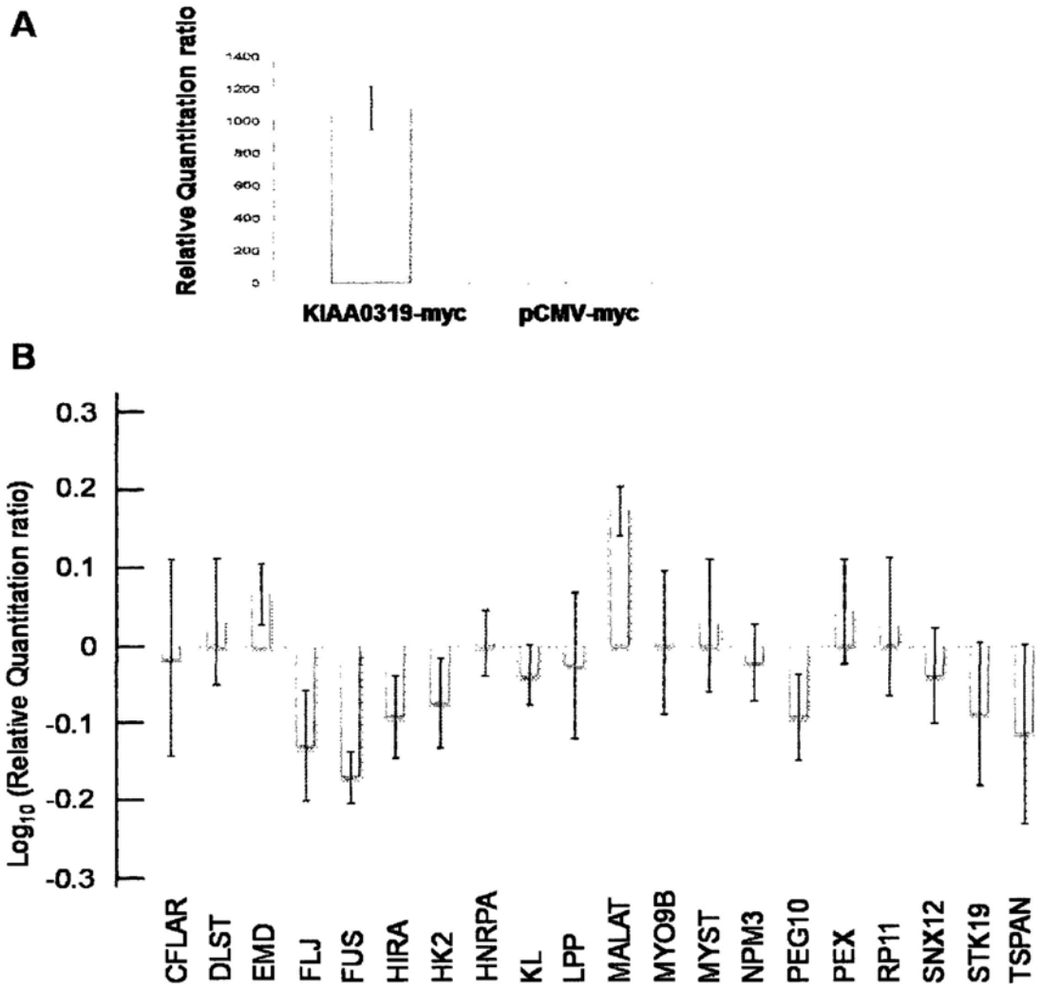


Figure 3-1. Quantification of relatively gene expression by real time PCR. (A) Expression of K in the microarray samples were analyzed by real-time PCR. Expression of endogenous gene PBGD was used for normalization. The results are presented with the values of the gene expression ratio's between Kmyc overexpressed vs. pCMV-myc vector transfected SHSY5Y cells. Comparing with SHSY5Y cell transfected with pCMV-myc vector, expression of K has increased for more then 1000 fold in Kmyc transfected cells. **(B)** A total of 20 differentially expressed genes identified by microarray assay were assayed for their relative expression between Kmyc transfected and pCMV-myc vector transfected SHSY5Y cells using real-time PCR. Expression of endogenous gene PBGD was used for normalization. The results are presented with the $\log_{10}$ values of the gene expression ratio's between Kmyc overexpressed vs. pCMV-myc vector transfected SHSY5Y cells. None of the genes has change in gene expression > 0.3 or <-0.3, corresponding to larger then two fold difference.

Table 3-1. Genes that were differentially expressed in microarray when K was up-regulated. The following forward (F) and reverse (R) primers used in real-time PCR were designed using the ProbeFinder software (Roche)

Genes that were down regulated more than four fold in microarray					
ProbeName	GeneSymbol	Fold change	Forward primer	Reverse primer	Amplicon (bp)
NR_002819	MALAT1	6.2	ccctcaagagaacacaagaagtg	ggagaacaactcgcatcacc	62
NR_003246	FLJ40113	4.8	agcagcaagtaaaggagctga	ccaagctccttgtgggact	113
NM_032454	STK19	4.5	acctttggcttcagggactc	agcatctcggacggtgag	69
NM_001011724	RP11-78J21.1	4.3	ttcagtttgaccaaggctatga	tccccctgaatatatttttcctg	84
NM_003325	HIRA	4.3	aagcccccagagagcatt	gtcacttcattctccacctcaa	78
NM_001040152	PEG10	4.3	gaccccatccttcctgtctt	ttttcaaaacccgcttatttca	60
ENST00000374274	SNX12	4.3	cggcgctacagtgactttg	cccaggcagtggtggtact	78
NM_004960	FUS	4.2	ggccagagcagctattcttc	ggggagttgactgagttcca	67
NM_002136	HNRPA1	4.2	gtcagcttgctcctttctgc	gctcaaccctccaatgaaga	138
NM_000319	PEX5	4.1	tgaacgagccaagtcagcta	catcccgttttgccatct	62
NM_024874	KL	2.5	gagcacctcagcaccatacc	ggcagggttatctggacact	75

Genes that were up regulated more than two fold in microarray					
ProbeName	GeneSymbol	Fold change	Forward primer	Reverse primer	Amplicon (bp)
NM_014809	**KIAA0319**	**223.6**	tcggacggagccactaactc	gtgatcgtcactgctctggtttc	150
NM_006766	MYST3	3.5	ggcagtgcatcgagtgtaaa	cggtcacatgaatcacaaaag	73
NM_001933	DLST	2.8	tatgccagggaccaggttac	accaagtcatccttgcatacag	92
NM_005578	LPP	2.6	atactgcgagccctgctaca	atgggcttggaacacacatt	108
NM_006993	NPM3	2.4	aacccatgctcagtctgga	cagaaacatcattgctcatcg	126
NM_004145	MYO9B	2.4	caaccagcacatcttcaagc	tgttgtgccacgtgatcc	65
NM_000117	EMD	2.4	gacttcattcccagatgctga	tacatggggcgttcccta	102
AF009616	CFLAR	2.2	ctcaccgtccctgtacctg	caggagtgggcgttttctt	67
NM_005725	TSPAN2	2.2	gcaatgtgtgcttggatcat	cagtggttacttcagcagcaaat	94
NM_000189	HK2	2.2	tcccctgccaccagacta	tggacttgaatcccttggtc	66

To validate the microarray result, three independent cDNA samples of SHSY5Y transfected by either Kmyc or pCMV-myc vector were prepared as before. Expressions of the genes identified by microarray were assayed for their relative expression between samples with or without K overexpression using Real time PCR (n = 3). The endogenous gene PBGD was used as a loading control. We found that the results from Real time PCR cannot confirm that of microarray (Figure 3-1B). If the expression ratios of a gene across two samples are close to one, they are vulnerable to random shifts between up-regulated, not-regulated and down-regulated genes (Hornshøj et al. 2009). Therefore, differentially expressed genes were defined as at least a two-fold difference in relative gene expression or at least > 0.3 in Log RQ ratio. None of the 20 selected genes has > 0.3 Log RQ ratio and therefore cannot be confirmed to have significant changes gene expression in mRNA level. Due to limitations in the calculation method which results in the small number in dCt compared to the original Ct value, if the value of ddCt is very small (which, in our case, often < 0.1), then a small standard deviation in Ct value will leads to large error in RQ (as shown in the error bar in Figure 3-1B).

The result from Real time PCR showed that when KIAA0319 is up regulated for more than 1000 fold, no significant changes in gene expression were found amongst the 20 selected genes. The difference between the microarray data and the Real time PCR result may be due to the fact that microarray tends to predict more differentially expressed genes, some of which might be false positives (Hornshøj et al. 2009). It is because expression arrays have more ways to vary systematically than measures such as Real

time PCR. Because these 20 genes are the most differentially expressed genes in the microarray analysis, no further validation by Real time PCR were conducted. As a result, our study in microarray and real time PCR show that K does not significantly regulate other genes expression in mRNA level.

3.2.3 Establishment of polyclonal stable HEK293 cell line expressing K

In addition to mRNA level, we also examined the effect of K overexpression on protein level by 2D gel. Because transient transfection systems often produce highly variable results from one assay to the next, to improve the reproducibility of the 2D gel experiments, polyclonal stable cell line expressing K protein was desirable. Analysis showed that human embryonic kidney (HEK) 293 cells express the neurofilament (NF) subunits NF-L, NF-M, NF-H, and α-internexin as well as many other proteins typically only found in neurons (Shaw et al. 2002). The presence of these proteins in HEK293 cell line argues that it belongs to neuronal lineage. For this reason and because of its high efficiency of transfection and protein production, HEK293 was selected as a neuronal model for this study. Full length human KIAA0319 was subcloned into pcDNA3.1 vector that enable stable mammalian expression, which produces untagged K protein (Kpcdna). The resulting Kpcdna and pcDNA 3.1 empty vectors were transfected into HEK293 cells. After two weeks of selection with Geneticin, cell lines were frozen and the expression of K protein analyzed by Western Blot (WB; Figure 3-2). High levels of K proteins were detected in WB which demonstrates that a polyclonal stable cell line that constitutively expresses KIAA0319 was generated.

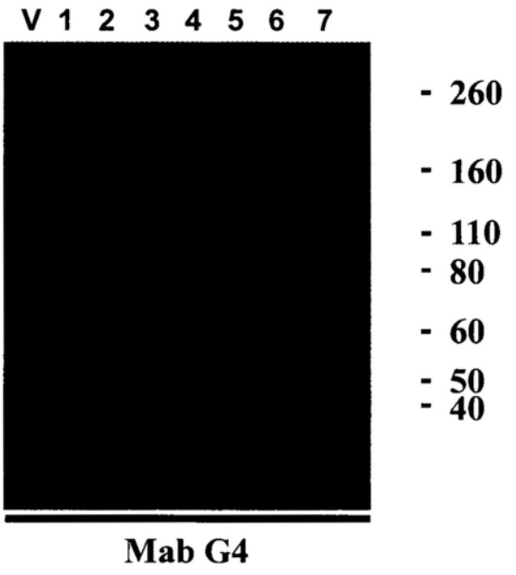


Figure 3-2. Expression of K in stable cell line. HEK293 cell lines stably transfected with pcDNA3.1 vector (V) or Kpcdna (clone 1-7) were analyzed by Western Blot with mouse anti-K monoclonal antibody (Mab G4) at dilution 1:500.

3.2.4 KIAA0319 does not affect other genes expression (protein level)

The high resolution of 2D gel separates thousands of proteins within a sample by their isoelectric points (pI) in the first dimension and their molecular weights in the second dimension (Black et al. 2009). It is highly sensitive and enables the simultaneous separation of a very large number of proteins and their individual quantitation. By the introduction of quantitative analyses, it was considered to be the best option for high-resolution profiling of global changes in cellular gene expression and protein modification (Bunai & Yamane 2005; Yuan & Desiderio 2005). To study the effect of K upregulation on protein expression profile, HEK293 cells stably transfected with Kpcdna or pcdna vector were lysed in 2D lysis buffer and the expression of K protein confirmed by WB analysis (Figure 3-3A). After confirming the expression of K protein, the total lysates were separated by IEF and SDS-PAGE (n = 2). The protein spots were visualized by either Colloidal Coomassie G-250 Staining (Dyballa & Metzger 2009) or silver staining. pI values were estimated and Novex molecular weight marker (Invitrogen) was used to assign molecular weight. The separation patterns of proteins obtained after 2D gel are shown in Figure 3-3B. The first observation is that these 2 samples do not display gross changes in protein expression. After careful examination of protein spots pattern, three pairs of spot seems to show visual difference in spot intensity. All three pairs of protein spots were picked for further analysis (Figure 3-3B for pair A, B and C). Possible identities of the three pairs were estimated by comparing with known protein spots in SWISS-2DPAGE database. Pair A, pI 7-7.2, MW 35-40, could be elongation factor Tu (*Homo sapiens*, P49411) or Immunoglobulin heavy chain gamma.

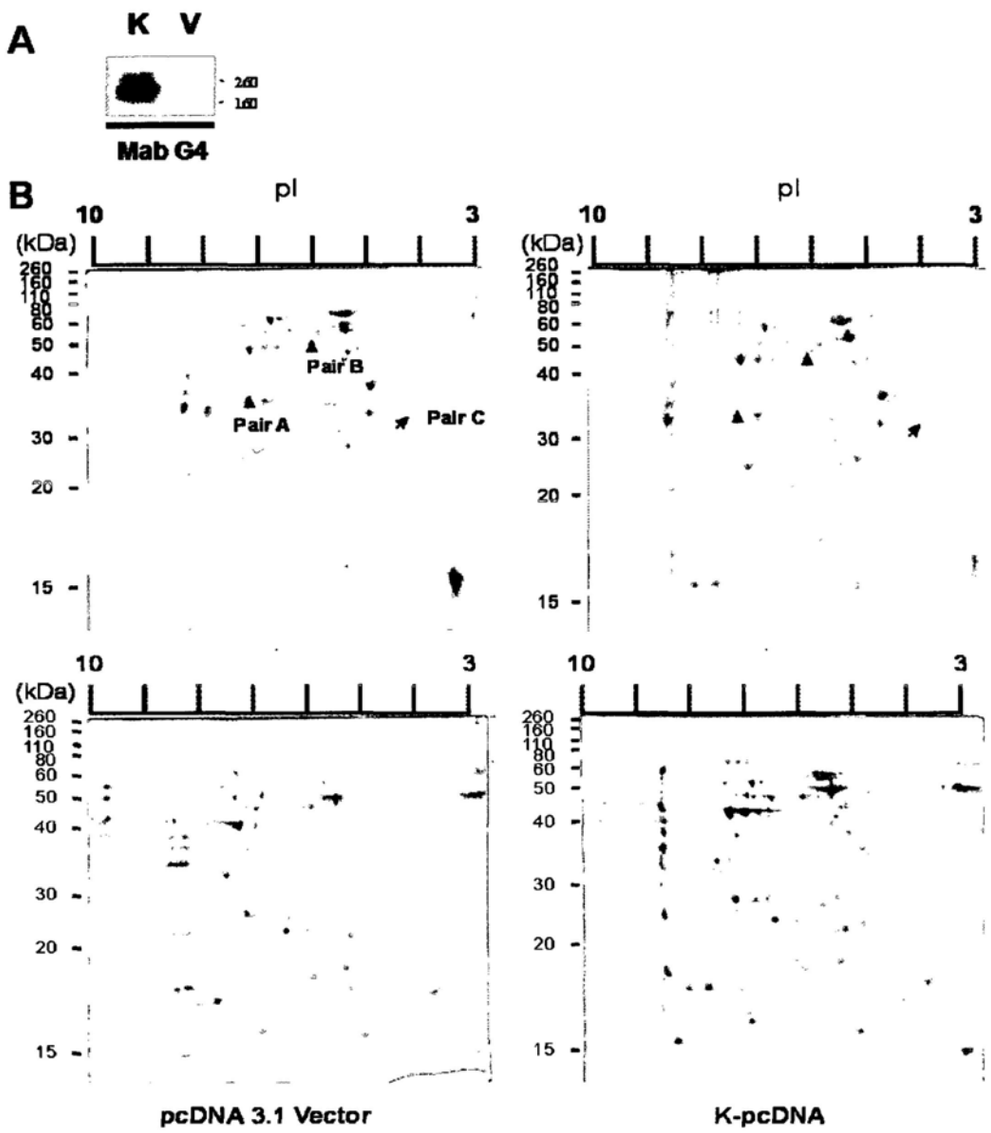


Figure 3-3. Comparison between the expression patterns of HEK293 cell lysate with or without K overexpression by using 2D gel electrophoresis. (A) HEK293 cell lines stably transfected with pcDNA3.1 vector (V) or Kpcdna (K) were lysed in 2d lysis buffer and analyzed by WB with mouse anti-K monoclonal antibody (Mab G4). **(B)** 2D samples analyzed by WB in (A) were subjected to 2D gel analysis. The first dimension was run on linear 13 cm IPG strip, pH 3-10. The second dimension was 12% SDS PAGE. The gels were stained with Coomassie brilliant blue (Top) or silver stain (Bottom). Three pairs of protein spots that seems to show minor changes in protein amount were indicated as Pair A, B and C by arrow.

(*Homo sapiens*, P99008). Pair B, pI=6, mw=50-55, could be Cellular tumor antigen p53(*Homo sapiens*, P04637). Pair C, pI=4, mw=32 , could be Apolipoprotein D (*Homo sapiens*, P05090). According to the protein spot intensity, pairs A and C were up regulated while pair B was down regulated. Before further analysis of these target proteins, we should confirm if they have been differentially expressed.

It is well known that 2D gels are often difficult to align and compare because of rubber-sheet distortions (i.e. local rotation, translation and magnification) with more distortion in some parts of the image than in other parts, making perfect overlay impossible (Lemkin 1997). Image sharpening and contrast enhancement may make them easier to compare, but some spots might disappear and new artefacts might appear if the grayscale values changed. Also, comparing gels this way is only good for a rough comparison and not to quantitate spots (Lemkin 1997). For this reason, to make a more accurate comparison, gels were analyzed by computer program which offer noise reduction, digital manipulation, warping, and enhancement. The 2D gel analysis software offer spot detection and matching algorithms which facilitate the extraction of statistically valid differences between groups of 2D gels. Protein spots were quantified base on spot intensity and area, which also allow normalization of protein loading as well as compensation of problem in first dimension (1^{st} D) or second dimension (2^{nd} D) gel running. Gels were scanned with ImageMaster Labscan v3.00 (Amersham Biosciences, UK) and spot detection performed using ImageMaster 2D platinum version 5.0 (Amersham Biosciences, UK). The gels were aligned, all the visible proteins spots were picked and more than 500 protein pairs were identified between K overexpressed sample

and control. The amount of proteins were quantified by percentage (%) volume (spot volume = spot area x spot intensity), which is proportional to protein amount. % intensity is a better measure of protein amount compared to spot intensity because it takes into accounts the total protein loading of a sample. Similarity, % volume is better then % intensity as the protein spots may distort and the shape changed due to difference in 1^{st} D or 2^{nd} D gel running.

The protein expression profiles were compared using scatter plot of % volume (Figure 3-4A). If the expression level of a protein is similar in the two samples, the spot should scatter around the correlation line, otherwise, if it was a quantitative differential expression, the spot should located far away from the correlation line and became outlier on the scatter plot (Holzmuller et al. 2008). The relationship between both samples for the area of common protein spots is significant with a high correlation coefficient of > 0.8, indicated high similarity between the expression of common protein spots between K overexpressed and control samples. In addition, as the slope of the scatter plot was close to 1, the intensity between samples does not need further normalization. From the scatter plot, as well as the low intensity spots, spots with high intensity also scatter around the linear regression line, showing that proteins expression have a positive correlation between samples. The lack of outlier demonstrated that the two samples do not display gross changes in protein expression. Pair A, B and C, which seems to show differential expressions by simple observation, located closed to the linear regression line, demonstrated that their expressions were similar between the two samples. Indeed, the magnified part of the gel image capture of the three pairs of protein spots

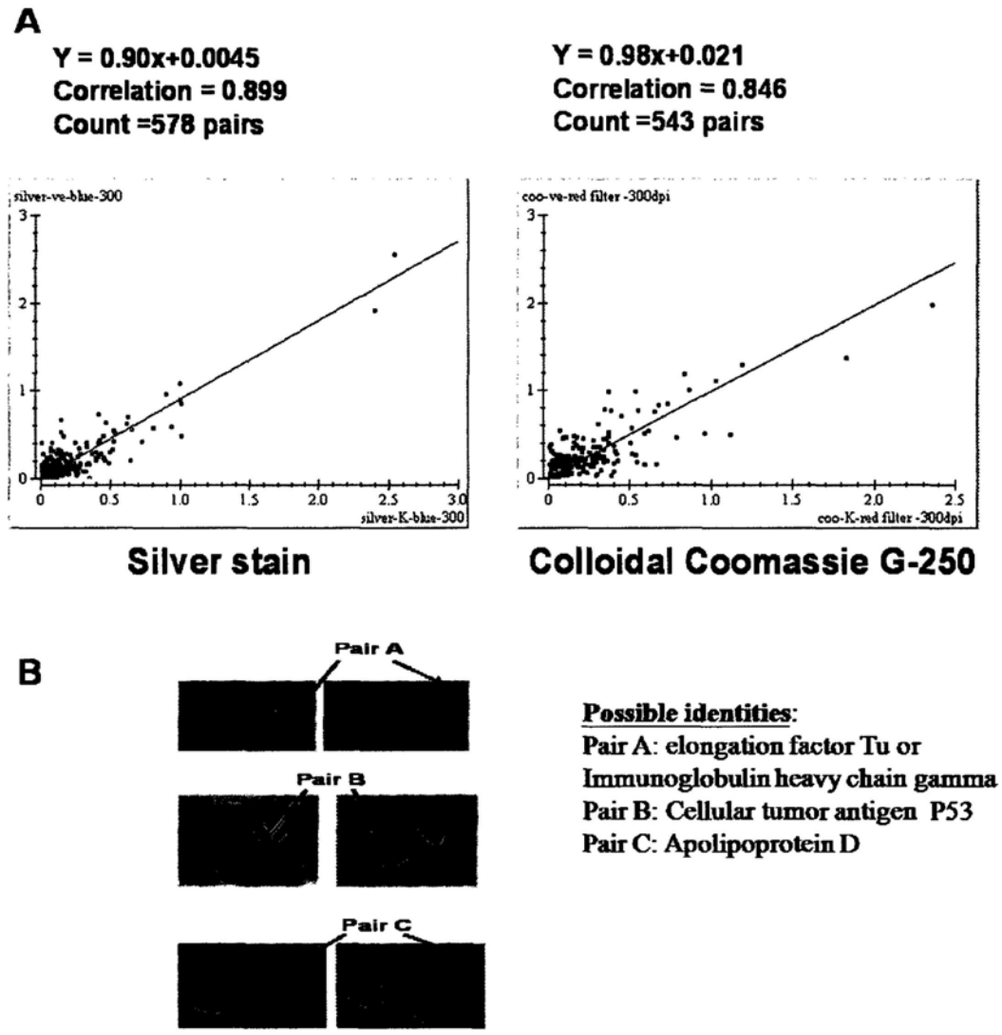


Figure 3-4. Analysis of the protein spots identified by 2D gel. (A) A 2D scatter plot was done on common protein spots for the percentage volume with the ImageMaster 2D Platinum Software Version 5.0. Pair A, B and C were represented by green dots. The protein spots scatter around the linear correlation line without outlier demonstrated that the two samples do not display observable changes in protein expression. **(B)** The magnified part of the gel image capture showing the protein spot pairs. Distortion and unequal loading was observed. Possible identities of the three protein pairs were estimated based on their pI and MW.

(Figure 3-4B) demonstrates some distortion in the gel which may result from problem in 1st D or 2nd D gel running. Unequal protein loading were also observed, which explain the visible difference in spot intensity and position. Therefore visible differences do not necessarily correlate to change in protein expressions.

All staining methods have strengthens and detection limit, proteins with low abundance might be missed. To observe as many spots as possible and thus observe most detectable differences, gels were stained by Colloidal Coomassie G-250 Staining (Dyballa & Metzger 2009) and Silver stain. While silver stain being very sensitive, it gives high background that is unfavourable to downstream analysis. This colloidal Coomassie G-250 staining gives lesser background and shows superior sensitivity that detects as low as 1 ng/band, which is superior to *Sypro Ruby* (a fluorescent stain) and comparable to silver stain (Dyballa & Metzger 2009). Furthermore, it is an endpoint method which offers a simple staining procedure and high quantitation. However, some proteins such as fetuin and mycin have weaker binding affinity for coomassie blue (Dyballa & Metzger 2009). Therefore, combining various sensitive staining methods would give higher sensitivity and provide a better representation of gene expression profile. No obvious changes in gene expression were found in both staining. As a result, the protein expression patterns were not grossly affected by K expression level and it was not an artefact introduced by staining methods.

Two-dimensional gel electrophoresis remains the single most effective method of general separation and analysis on a large scale with a small budget, which base on the

assumption that proteins are unlikely to be similar in two distinct properties (pI and MW). It can represent entire proteome and serve as a powerful tool for investigating global changes in cellular gene expression (Bunai & Yamane 2005). Nevertheless, this technique also has several limitations. Foremost is poor separation of very acidic, very basic proteins (range from pH 2.5-12), very small, very large proteins, highly hydrophobic proteins and nuclear proteins (Bunai & Yamane 2005). Protein solubilization by detergents is critical for fine 2D gel separation; however, no universal non-ionic or zwitterionic detergent can yet be applied to IEF for all proteins. Hydrophobic membrane proteins are relatively insoluble and many of them cannot be resolved by pH gradients during IEF (Bunai & Yamane 2005). This could be one of the reasons why the integral membrane protein K was not visible on the 2D gel and thus its expression in the 2D samples had to be confirmed by WB. Other techniques such as chromatography that are independent of 2D gel may be employed to resolve the problems associated with analyzing hydrophobic and alkaline proteins (Bunai & Yamane 2005).

Moreover, a high level of between gel reproducibility is essential for comparative proteomics. While 2D gels are not 100% reproducible, it was suggested that several replicates from a sample should be analyzed simultaneously to reduce any variability in the 2D gel system (Yuan & Desiderio 2005). Also, a large homogeneous gel should be used for improve gel resolution while a thicker gel would allows more tolerance in any background variation and a higher level of sample load.

In addition, the restriction of sample volume loaded on each IEF gel limit the detection of low abundance proteins. Besides, the high concentration of Urea in 2D buffer makes accurate protein quantification difficult. This result in unequal loading of proteins in the two samples which might also leads to artefacts (Figure 3-4). These make normalization and analysis with 2D gel analysis software essential.

Besides, although the expression of K protein was detected by WB using anti-K antibody, it was not visible on the 2D gel by coomassie and silver staining. One reason would be that the hydrophobic membrane protein K cannot be resolved by pH gradients during IEF. Another explanation is that although K was upregulated (as demonstrated by the presence of band in WB compared to control sample), its absolute protein amount remain very low and beyond the detection limit of the staining methods. Nevertheless, K was only down regulated for 40% in the DD risk haplotype which suggested that a small change in K expression level largely affect its function. Therefore, the increased in K's protein level in the 2D samples as demonstrated by WB should provide a fair comparison.

To achieve a comprehensive understanding of cellular proteins, the limitations of 2D gel should be overcome by adding other methods such as DNA micro array analysis, two-dimensional native and alternative types of chromatography that are independent of gels to separate individual protein components (Bunai & Yamane 2005). In this study, DNA microarray and 2D gel electrophoresis were performed and no obvious change of gene expression was observed in K overexpression. There might have some minor changes in gene expression, but were not detectable under these techniques.

Therefore, we continue to investigate the disruption of protein-protein interactions that might explain the association of DD predisposition to deregulation of K. In fact, proteins participate in extensive networks of protein-protein interactions inside the cells and abnormal protein-protein interactions have implications in a number of neurological disorders; include Creutzfeldt–Jakob and Alzheimer's diseases (Nazar Zaki 2008). Protein-protein interactions are of central importance for virtually every biological process in a living cell including signal transduction, cellular trafficking or protein modifications....etc. A preliminary step in understanding the functional and structure properties of a novel protein with unknown function is to determine which proteins interact with each other, thereby identifying the relevant biological pathways. Information about the interactions of K would therefore provide a molecular basis of K's function and improve our understanding of DD.

3.3 Conclusion

K expression level is critical to neuronal migration and hence brain development, which implies that its association to DD was due to either its deregulation which affects the expression of genes involved in neuronal migration, or that its deregulation disrupts one or more protein interactions. To determine if K affects neuronal migration by modulating expression of genes and to associate K with biological processes of known function, gene expression analyses was performed, both at the transcription and translation levels. Microarray identified differentially expressed genes were subjected to validation by Real time PCR and none of the 20 selected genes show larger then two-fold difference in

relative gene expression. The difference between the microarray data and the Real time result may due to the fact that microarray tends to predict more differentially expressed genes, some of which might be false positives. Therefore, the up regulation of K does not grossly affect gene expression at mRNA level. In parallel to microarray analysis, 2D gel electrophoresis was used to examine the effect of K overexpression on gene expression profiling at the protein level. A polyclonal stable HEK293 cell line expressing K protein was established accordingly. The separation patterns of proteins obtained after 2D-gel show no obvious differences in samples with or without K overexpression. All visible protein spots were picked and analyzed by software. More than 500 protein pairs were selected between the two protein-expression profiles and scatter plot of percentage volume found strong correlation of 0.87 ± 0.03 without any outliner, indicate the expression pattern between K overexpressed and control samples were similar. Staining with coomassie and silver stain also give comparable patterns. Taken together, no grossly differentially expressed genes were identified in our study, both at the mRNA and the protein level, when K was overexpressed. Accordingly, K's function might not be affecting other genes expression in an obvious manner. Its association to DD could therefore due to disruption of protein interactions. The presence of interaction networks are of central importance for virtually all levels of cellular function including architecture, metabolism and signalling, as well as the availability of cellular energy. Mapping of these interaction networks into functional units is thus of prime importance for systems biology. Information about these interactions would therefore provide a molecular basis of K's function and improve our understanding of DD.

CHAPTER 4

Establishment and characterization of specific anti-KIAA0319 antibodies

4.1 The prerequisites of specific anti-K antibody for studying protein-protein interaction

According to the previous section, it is clear that the expression level of K does not have any gross regulatory effect on other gene expression, neither at the transcription level, nor at the translation level. Very likely, the down regulated expression level of K associate to DD by disrupting one or more disease-related interactions with its partners. In fact, the presence of PKD domains imply that K might be involved in cell adhesion, cell-cell interactions or ligand binding (Bycroft et al. 1999), which largely depends on protein interactions. Being a novel protein with unknown function, K unfortunately has only very limited known interaction partners (Levecque et al. 2009). To characterize the correlation between K and DD, it would be ideal to fish out all possible interaction partners. Better understanding of DD, such as the mechanism behind and biological pathways involved, can also be cued from the identification of those interaction partners.

4.1.1 Methods of screening protein-protein interactions

There are a multitude of methods to screen for protein-protein interactions; each of the different approaches has its own strengths and weaknesses, especially with regard to the

sensitivity and specificity of the method. The yeast two-hybrid (Y2H) screening is a molecular biology technique used to investigate the interaction between artificial fusion proteins inside the nucleus of yeast. This approach can provide an important first hint for the identification of interaction partners *in vivo* in a relatively authentic cellular environment. It is one of the most frequently used methods for identify protein-protein interactions since it is comparatively inexpensive, do not need specialized large equipment and can be performed in any molecular biology laboratory with reasonable throughput. Experimental Y2H data have been a crucial part in establishing large synthetic human interactomes or to dissect mechanisms in human disease (Brückner et al. 2009). So far, a majority of published interactions have been detected using an Y2H screen. For instance, more than 5,600 protein interactions have been so far reported for yeast and about 6,000 for humans, establishing extensive protein interaction networks (Brückner et al. 2009). However, Y2H screening is not recommended for membrane protein because of their hydrophobic nature (Iyer et al. 2005). K is a type I plasma membrane protein (Velayos-baeza et al. 2008) and therefore should employ other screening methods. Besides, proteins requiring post-translational modifications such as glycosylation may not be detected in Y2H due to the lack of post-translational protein modifications in yeast. This approach also leads to a high number of false positive (and false negative) identifications because some proteins may interact when they are co-expressed in the yeast, although in reality they are never present in the same cell at the same time. Furthermore, Y2H only pick up direct interactions, components of a larger complex, which not necessarily all interact directly with each other can be missed.

Glutathione-S-transferase (GST) pull down screening is another invaluable tool for identifying previously unknown protein-protein interactions (Oyama et al. 2009; Shi & Andres 2005; Sanderson et al. 2008) It is an affinity purification method used to determine physical interaction between two or more proteins *in vitro*. GST pull down requires a purified and tagged protein (the bait) which will be used to capture a protein-binding partner (the prey). However, our attempt to purified GST tagged -K fragments in *E.coli* had failed perhaps because of the lack of post-translational modification in *E.coli* and/or the high Cysteine content in K that leads to proteins aggregates.

4.1.2 Immunoprecipitation screening for interacting partners

To get around with the expression obstacle, expression in mammalian system is desirable. Immunoprecipitation (IP) is small scale affinity purification of antigen using a specific antibody, which is considered to be the gold standard assay for screening protein-protein interactions *in vivo* (Nazar Zaki 2008; Grabowska et al. 2008; Liu et al. 2004). The protein of interest together with its interacting partners are isolated with specific antibody and the identities of the interacting partners are subsequently analyzed by gel electrophoresis and Mass spectrometry. In the case of Co-immunoprecipitation (Co-IP), where the identities of the interacting partners are known, the immunoprecipitated complex is analyzed by WB. Immunoprecipitation experiments reveal direct and indirect interactions. Thus, positive results may indicate that two proteins interact directly or may interact *via* a bridging protein. IP is a powerful technique that is used regularly by molecular biologists to analyze protein-protein

interactions and the interactions detected by this approach are considered to be real (Nazar Zaki 2008). One of the major technical problems with immunoprecipitation is that the specific antibody with high affinity towards the native protein is not always available. Many researchers therefore add tags onto either the C- or N- terminal end of the protein of interest and used anti-tag antibodies to pull down the protein. Myc tag, hemagglutinin (HA) tag, Green Fluorescent Protein (GFP) tag, GST tag and the FLAG-tag tag are frequently used tag. One of the major advantages with using tagged protein is that the same tag can be used on many different proteins and the same antibody is required each time. This saves the time and effort to produce protein specific antibodies, especially when many different proteins are involved in the research and that generating specific antibodies against every single protein are not possible.

4.1.3 The prerequisite of specific Anti-K antibody

While adding tags onto either the C- or N- terminal end of the protein may get around with problem of lacking the specific antibody, it raises some concerns regarding biological relevance because the tag itself may either obscure native interactions or introduce new and unnatural interactions. Consequently, in order to perform IP with untagged KIAA0319, antibody specifically against K was necessary. Specific antibodies including polyclonal antibodies (Pab) and monoclonal (Mab) antibodies are both useful for performing IP as well as many other immunoassays (Xiong et al. 2009; Vetrivel et al. 2008). These antibodies are typically produced by immunization of a mammal, such as mouse or rabbit. Pab are antibodies that are derived from different B cell lines and are simply purified from an immunized animal's serum. They are a mixture of

immunoglobulin molecules secreted against a specific antigen, each recognizing a different epitope. The amount of Pab obtained from each animal is limited and new immunization scheme is required if more antibody is desired in future. By contrast, Mab are monospecific antibodies that are identical because they are produced by one type of immune cell that are all clones of a single parent cell. The method of producing Mab involved fusing B cells from the immunized animal with immortal myeloma cells (Köhler & Milstein 1975), creating hybridomas which continually secret antibodies. The biggest advantage of using monoclonal antibodies over polyclonal antibodies is the specificity of their interactions. Because Mab only binds to a single epitope, it has less trouble with non-specific binding. Besides, after generating the Mab secreting hybridoma cell lines, it can provide unlimited source of Mab, saving the hassle of repeating the immunization scheme. Nevertheless, in spite of all the advantages, creating Mab secreting cell line is much more time consuming, more expensive and more labour intensive. Furthermore, not all Mab work well in detecting native protein because the single epitope recognised by the Mab might not be assessable. On the contrary, Pab recognise multiple epitopes and thus have a greater chance of detecting the native proteins and often with higher affinity. Therefore, to obtain specific anti-K antibody that is capable for performing IP and various assays, both Pab and Mab are desirable.

In this project, a mouse hybridoma cell line was constructed which can serve as an indefinite source of monoclonal antibody specific to K for the long term study of this protein. A specific anti-K polyclonal antibody was also produced. In addition to studying protein interactions by immunoprecipitation and co-immunoprecipitations, several

biochemical methods based on detection of complex antigen-antibody are made possible with the specific antibody. Such as detection and quantification of K by ELISA, identify K proteins according to their molecular weight by WB, examine K expression in tissue sections or to locate K proteins within cells with the assistance of a microscope by immunofluorescence and immunohistochemistry. A fully cloned monoclonal cell line and the antibodies produced could also be used for further research both academically and commercially.

4.2 Designing synthetic peptide for raising antibodies

Synthetic peptide are good for producing antibody because it can be generated in a highly purified form and are capable of targeting specific regions of the protein. The length of antigenic peptides are usually between 12 to 16 residues as these length are relatively easy to synthesize and are long enough to contain epitopes for generating antibody. Peptides capable to produce an immune response have to be antigenic and soluble under experimental conditions. Two sequences were chosen for making the synthetic peptides in this study, peptide P124 (GSPSGIWGDSPEDI, corresponding to the amino acid residues 124-137 of human K) and P663 (VRGPSAVEMENIDK, corresponding to the amino acid 663-676 of human K) (Figure 4-1A). It is useful to design more than one peptide, from different regions of the K protein, to increase the chance of raising an immune response in mice and producing a good antibody. In addition to hydrophobicity and secondary structure prediction, OptimumAntigenTM Design Tool (genscript) was used to analyze the Antigenicity of

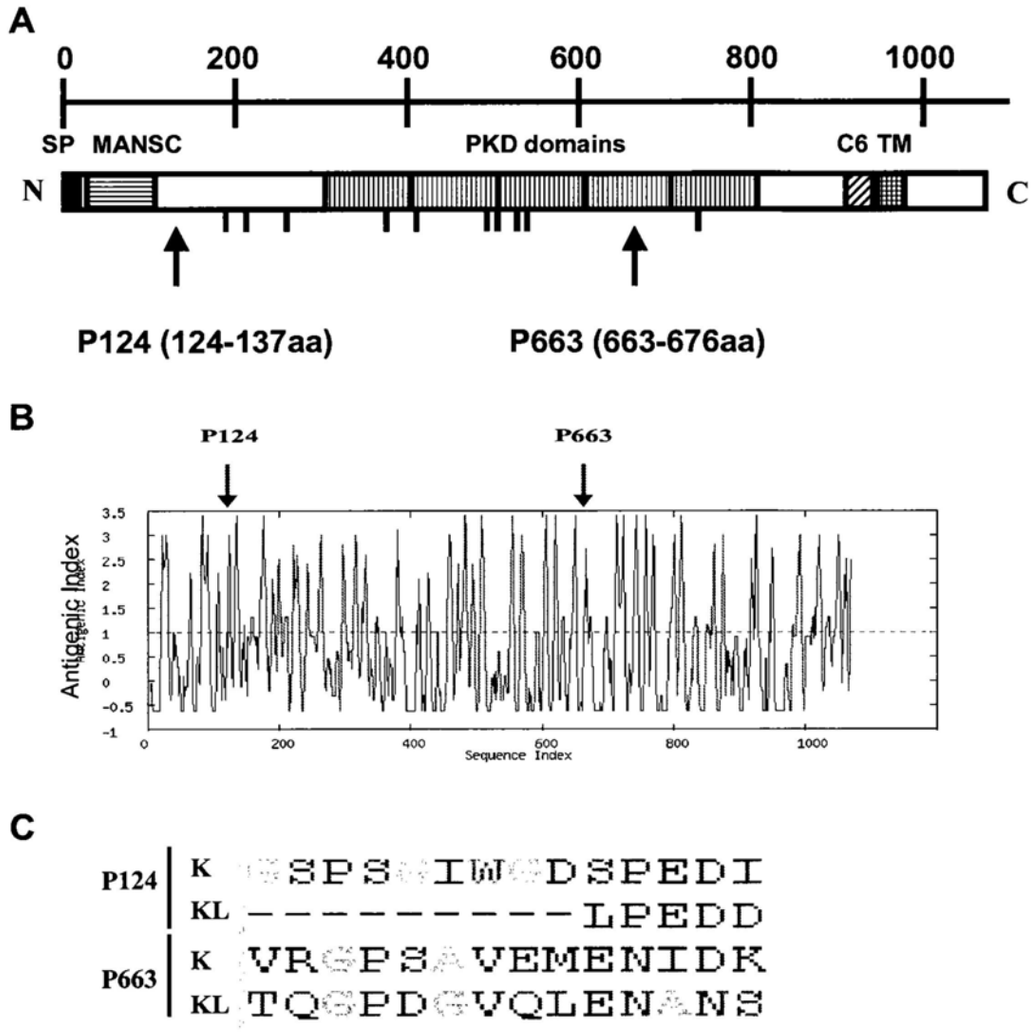


Figure 4-1. Synthetic peptides against human KIAA0319 protein. (A) Graphical representation of the full length human KIAA0319 protein (K). The two peptides, p124(GSPSGIWGDSPEDI) and p663(VRGPSAVEMENIDK), used to raise polyclonal and monoclonal antibodies are also indicated. **(B)** Antigenicity of KIAA0319 protein sequence are displayed as antigenicity plots. The antigenic regions are identified by OptimumAntigenTM Design Tool (genscript). **(C)** Protein sequence of full length K and KL were aligned. Alignment at regions of the two peptides were shown as indicated.

KIAA0319 protein (Figure 4-1B). Analysis by Antigen Profiler tool (openbiosystems) shows that the two peptides are very antigenic, with P124 score 3.8 and P663 score 2.6, which were classified as good antigen and excellent antigen respectively. Besides, these sequences are within the exposed portions of the membrane protein and hence more accessible to the antibody, which would therefore increase the probability that the antibody produced would work well in WB and immunoprecipitations. (Jacob & Unger 2007). A peptide that is within transmembrane region or region inside the cell would produce antibody that is less practical. A major concern about preparing sequence-specific antibodies is to avoid sequence that is common to a family of proteins. Because of the small size of a epitope that consist of 4-5 amino acids, cross activity problem may arise, which was solved by choosing region with low similarity to other proteins. Since K protein shared more than 45 % amino acid identity with KL, Basic Local Alignment Search Tool (BLAST) search and sequence alignment of K and KL has been performed to select peptide sequence that are specific to the K protein (Figure 4-1C). Another reason why we chose P124 is because it is located near the N-terminus of the K protein, where K and KL share the least similarity. Using Similarity Matrix BLOSUM62 (BioEdit), P124 is 21.4% identical between K and KL while P663 is 35.7% identical. Moreover, in designing the peptides, regions predicted to have posttranslational modifications such as N- and O-glycosylation as well as regions rich with probable disulfide bonds or modified residues are avoided.

The synthetic peptides were then conjugated to larger, immunogenic, carrier proteins which include KLH and BSA to increase immunogenicity and to provide T cell epitopes.

KLH conjugated peptides were used for immunization while BSA conjugated peptides were used for ELISA screening. Using differently conjugated peptides for immunization and ELISA screening avoid selection of an antibody that recognises the carrier proteins.

4.3 Generation of polyclonal antibodies

For quick production of specific anti-K antibody, polyclonal antibody against the synthetic peptide P124 conjugated to KLH was generated in balb/c mice. Animals were inoculated a total of five times, with 2- to 3- week intervals between inoculations. Antiserum isolated contains the polyclonal antibodies and were subjected to ELISA on plates coated with the P124 conjugated to BSA. The mouse 9 polyclonal antiserum (Pab m9) gave positive reading against P124-BSA in ELISA and was subjected to WB (Figure 4-2A). Results revealed that Pab m9 specifically recognized two bands of 170 kDa and 200 kDa. The molecular weight of K was larger than the expected size 120 kDa due to N- and O- glycosylation and the two bands corresponds to different glycosylated forms. To confirm this, total cell lysates were compared after treatment with or without PNGase F, which showed that the size of KIAA0319-myc protein is reduced when glycan moieties are removed (Figure 4-2B). This proved that K is N-glycosylated, which is consistent with the published data (Velayos-baeza et al. 2008). The size of Kmyc expressed in WRL68 was smaller than the untagged K expressed in HEK293FT (150 kDa and 170 kDa), which was possibly due to different glycosylation patterns produced in different cell lines (Lifely et al. 1995). The addition of myc tag as well as different culture conditions might also affect glycosylation of the protein. Taken together, the polyclonal antiserum m9 can detect denatured K protein on WB.

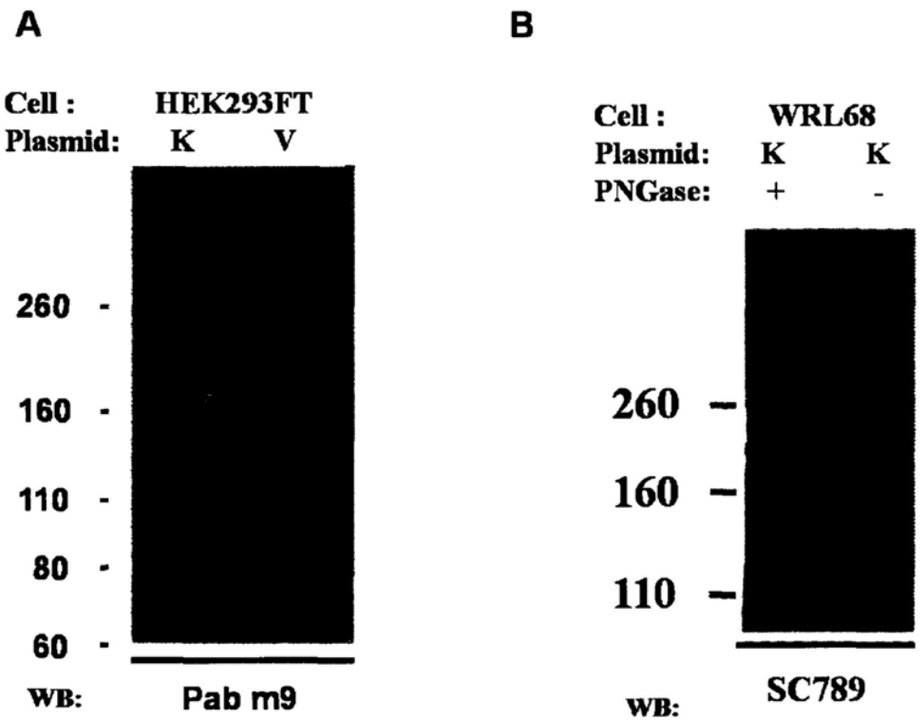


Figure 4-2. Western blot analysis of mouse 9 polyclonal antibody. (A) HEK293 FT cells were transfected with Kpcdna (K) or pcdna vector (V) as indicated. Total lysate were separated on a 8% SDS PAGE and transfer to PVDF membrane. The blot was probed with mouse 9 polyclonal antiserum at dilution 1:1000. **(B)** Western blot showing that KIAA0319 is N-glycosylated. WRL68 cell lines were transfected with Kmyc. Total cell lysates were compared after treatment with or without PNGase, an enzyme which removes nearly all N-glycosylations. The blot was probed with Rabbit anti-myc polyclonal antibody (SC789) at dilution 1:200. A shift in size was observed and indicated that KIAA0319 has N-linked glycosylation.

4.4 Demonstration of antibody specificity

4.4.1 Polyclonal antibody specifically recognize denatured K protein

As a specificity control, Pab m9 was incubated with its cognate peptide P124 in order to block specific binding (Figure 4-3A). Pab m9, but not mouse 9 pre-serum, detects a 170 kDa protein corresponding to the expected sizes of the K (Velayos-baeza et al. 2008). Rabbit anti-myc polyclonal antibody SC789 (santa cruz) react with this protein in the same experiment, which serve as a positive control. Pre-absorption of the Pab m9, with the original immunizing peptide sequence (P124) abolished staining of the bands corresponding to the expected sizes of the KIAA0319. Pre-absorption of the antibody, with irrelevant peptide (P663) shows no blocking effect. No band was observed in untransfected control. Therefore, Pab m9 is specific to the peptide P124 and would specifically recognize K protein.

4.4.2 Polyclonal antibody specific to K does not recognize KL

Because of the sequence similarity between K and KL, it is important to know whether the antibody against K would recognize KL. We evaluated the specificity of Pab m9 to K and KL in WRL68 cell lines separately transfected with myc tagged K and KL (KLmyc) (Figure 4-3B). WB analysis with Rabbit anti-KL polyclonal antibody (H635) demonstrated that KLmyc was expressed in the transfected cells. However, Pab m9 only detected the expected 170 kDa band whose molecular weight corresponded to the full-length K protein. Thus our results revealed that Pab m9 specifically recognized K and has no cross reactivity to KL.

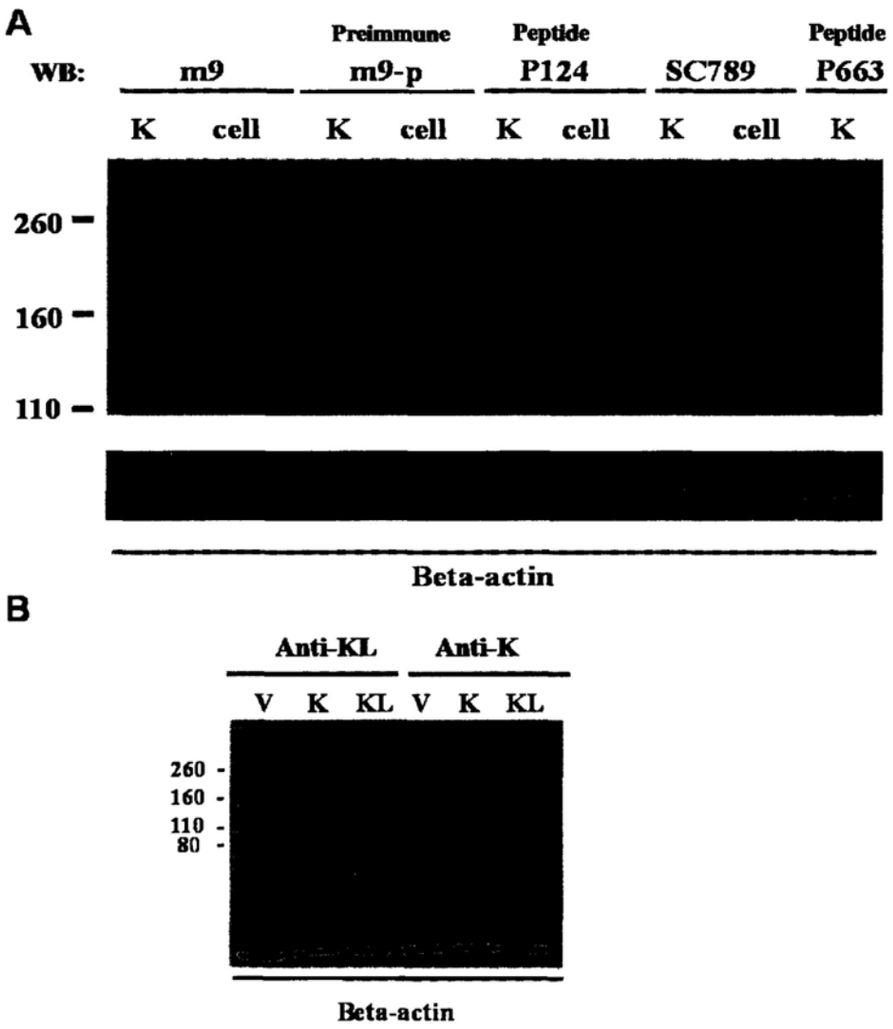


Figure 4-3. Characterization of mouse 9 polyclonal antibody. (A) Western blot showing that mouse 9 polyclonal antibody is specific to the peptide p124, which corresponds to KIAA0319. WRL 68 cell lines were either untransfected (cell) or transfected with Kmyc (K) and processed for WB analysis with specific antiserum mouse 9 at dilution 1:6000 (m9), mouse 9 preserum (m9-p), mouse 9 antiserum pre-incubated with peptide p124 (p124), mouse 9 antiserum pre-incubated with peptide p663 (p663) and rabbit anti-myc polyclonal antibody (SC789). **(B)** Western blot showing that mouse 9 polyclonal antibody is reacting specifically to Kmyc and not KLmyc. WRL 68 cell lines were transfected with Kmyc (K), KLmyc (KL) or pCMV-myc vector (V) as indicated and processed for WB. The blot was probed with mouse 9 polyclonal antiserum (anti-K) at dilution 1:3000 or rabbit anti-KL polyclonal antiserum (anti-KL) at dilution 1:1,000. Proper loading of sample were confirmed by probing the same blot with an anti-β-actin antibody.

4.5 Polyclonal antibody detect native K protein *in vivo*

Knowing that the polyclonal antiserum m9 recognize denatured K protein *in vitro*, it is also important to know whether it would detect native K protein *in vivo*. For this reason, Pab m9 were used to detect the cellular localization of K protein based on specificity characterized by WB analysis. To test whether Pab m9 is suitable for immunocytochemistry, Kmyc were overexpressed in human liver cell lines WRL68 and subjected to immunofluorescence (IF) with Pab m9 and anti-myc antibody SC789 (Figure 4-4A). The signal from SC789 and Pab m9 overlap and suggested that Pab m9 recognize the same Kmyc protein as that detected with the anti-myc antibody. Next, we used Pab m9 to detect the localization of endogenous K in human neuroblastoma cell line SHSY5Y (Figure 4-4B). Human liver cell line WRL68 that does not express endogenous K was included as the negative control. The polyclonal m9 stained SHSY5Y while no signal was observed from the control cell lines. To confirm that the signal comes from the specific primary antibody and not from non-specific binding of secondary antibody, IF was repeated with only the secondary antibody, but the primary antibody omitted by fluorescence confocal microscopy (Figure 4-4C). Using the same setting where the primary omitted sample was negative, staining with Pab m9 gives a high signal. Based on specificity characterized by WB analysis, the signal from Pab m9 was believed to be endogenous K protein. Taken together, the Pab m9 could detect native K protein *in vivo*.

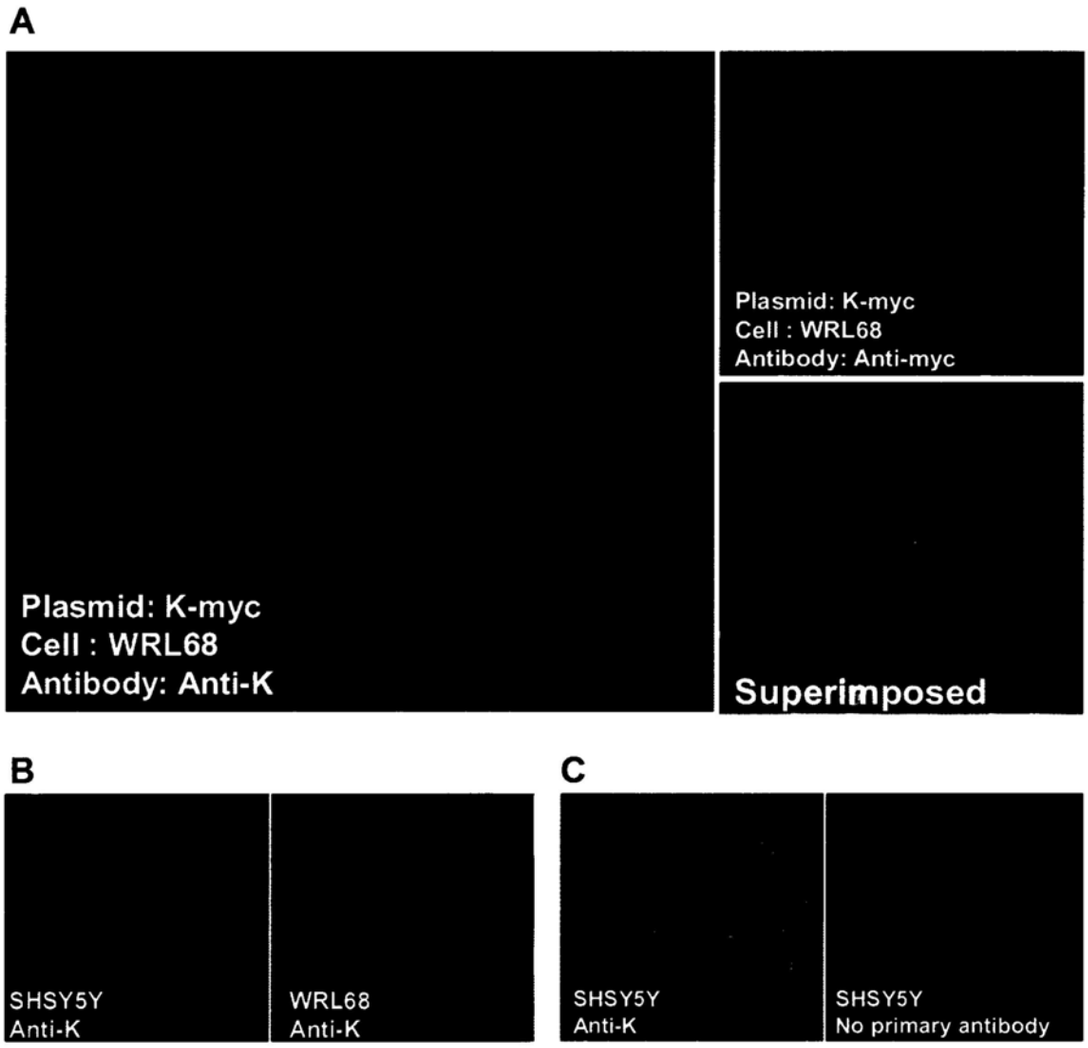


Figure 4-4. Immunofluorescence (IF) analysis of mouse 9 polyclonal antibody. (A) Kmyc were overexpressed in WRL68 cell lines. Cells were fixed and then probed with Pab m9 at a 1:400 dilution (anti-K) or SC789 at 1:100 dilution (anti-myc). Fluorescein (FITC) AffiniPure F(ab')2 Frag Donkey Anti-Rabbit IgG (H+L) at a 1:200 dilution or Rhodamine (TRITC) AffiniPure Goat Anti-Mouse IgG (H+L) at 1:200 dilution were used as secondary antibodies. The nuclei of the cells were stained with Hoechst. **(B)** Human SHSY5Y and WRL68 cells were fixed and then probed with Pab m9 at a 1:400 dilution. Cells were then probed with Rhodamine (TRITC) AffiniPure Goat Anti-Mouse IgG (H+L) at 1:200 dilution. The nuclei of the cells were stained with Hoechst. **(C)** SHSY5Y cells were fixed and then probed with Pab m9 (anti-K) at a 1:400 dilution. Cells were then probed with DyLight 488 AffiniPure Goat Anti-Mouse IgG (H+L) at a 1:200 dilution. The nuclei of the cells were stained with PI .

The results with Pab m9 in immunocytochemistry show that endogenous K protein is located within the cytoplasm of the cells. To apply this finding and to confirm K's subcellular localization, separate experiments using a cell fractionation kit (ProteoExtract® Subcellular Proteome Extraction Kit) was performed. The kit fractionates proteins taking advantage of the different solubilities of certain subcellular compartments in the four selected reagents, which demonstrated that Kmyc protein is extracted in the cytoskeletal protein fractions (Figure 4-5). Under an overexposed condition, Kmyc was also present in the membrane fraction. Four fractionation controls, histone H3, GAPDH, Calnexin and Vimentin were assessed as nuclear, cytosolic, membrane/organelle and cytoskeletal specific markers, respectively. The distinct localization of these marker proteins verified the purity of the fractions. In separate fractionation experiments, Kmyc protein expressed in HEK293FT cells were also extracted in the cytoskeletal protein fractions, which demonstrated that expression in different cell lines does not affect the localization of K protein. Taking these findings together, Kmyc is present in the cytoskeleton of the cells, which is consistent with the immunocytochemistry performed with Pab. Therefore, the Pab m9 can detect native K protein *in vivo*.

Results of immunocytochemistry and cell fractionation suggested that K has a cytoskeletal as well as membrane localization. In fact, it has been published that K is a membrane protein (Velayos-baeza et al. 2008) that follows the classical clathrin-mediated endocytosis pathway (Levecque et al. 2009). It is possible that after the membrane protein is internalized, it could be localized to the cytoskeleton. Or else, it

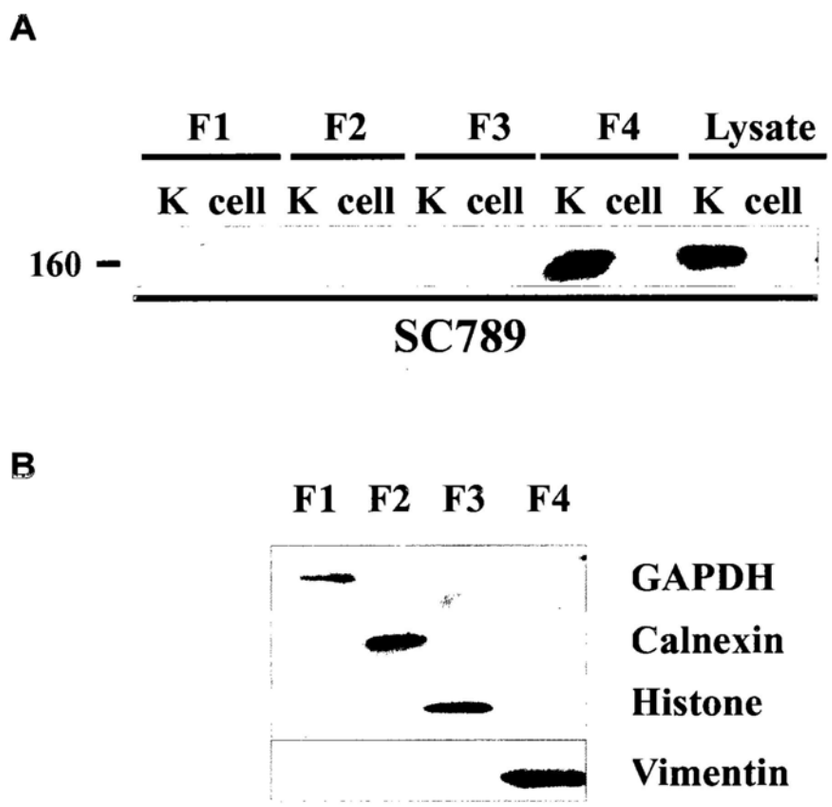


Figure 4-5 KIAA0319 localized to cytoskeletal matrix. (A) WRL 68 cells overexpressing Kmyc (K) and untransfected cells (cell) were submitted to subcellular fractionation in order to separate into nuclear, cytosolic, membrane and cytoskeletal matrix fractions, as described in Materials and Methods. Kmyc in the Cytosolic protein fraction 1 (F1), Membrane/organelle protein fraction 2 (F2), Nuclear protein fraction (F3), Cytoskeletal Matrix Protein Extraction (F4) as well as total cell lysate were detected by WB using SC789. **(B)** Successful fractionation was confirmed by the WB analysis of fraction marker molecules GAPDH, Calnexin, Histone and Vimentin in Kmyc fractions 1 to 4.

could localize at positions that the membrane is linked to the cytoplasmic cytoskeleton, for example, the focal adhesion. Moreover, since K have membrane isoforms and secreted isoforms, it is also possible that it has cytoskeletal isoforms. Further studies would be required in order to confirm these hypotheses.

As demonstrated by validations by peptide absorption, specificity against KL protein and Immunocytochemistry, a polyclonal antibody that is specific to K and capable for various immunoassays was produced. This Pab can be used to further characterize native or denatured K protein under *in vivo* as well as *in vitro* condition. However, due to the limited source of the Pab and because of the higher specificity of Mab, an anti-K Mab was desirable.

4.6 Generating monoclonal antibody

4.6.1 Monoclonal antibody production by hybridoma approach

A monoclonal antibody against the synthetic peptide P124 conjugated to KLH was generated in mice in an approach similar to generating polyclonal antiserum. Polyclonal antiserum was collected from mice after the third inoculation. Mouse B that was ELISA and WB positives (Figure 4-6B) was then selected for fusion and the hybridomas were plated on five 96 well plates in the presence of HAT selective medium. Six days after fusion, hybridoma supernatants were analyzed by ELISA. Fifty five wells were positive and all of them were passed into 24 well plates. At day 10, ELISA was repeated. Ten clones gave OD_{450} reading three times higher than the background level and were considered positives. The seven positive clones that gave highest reading – 9, 12, 13, 17,

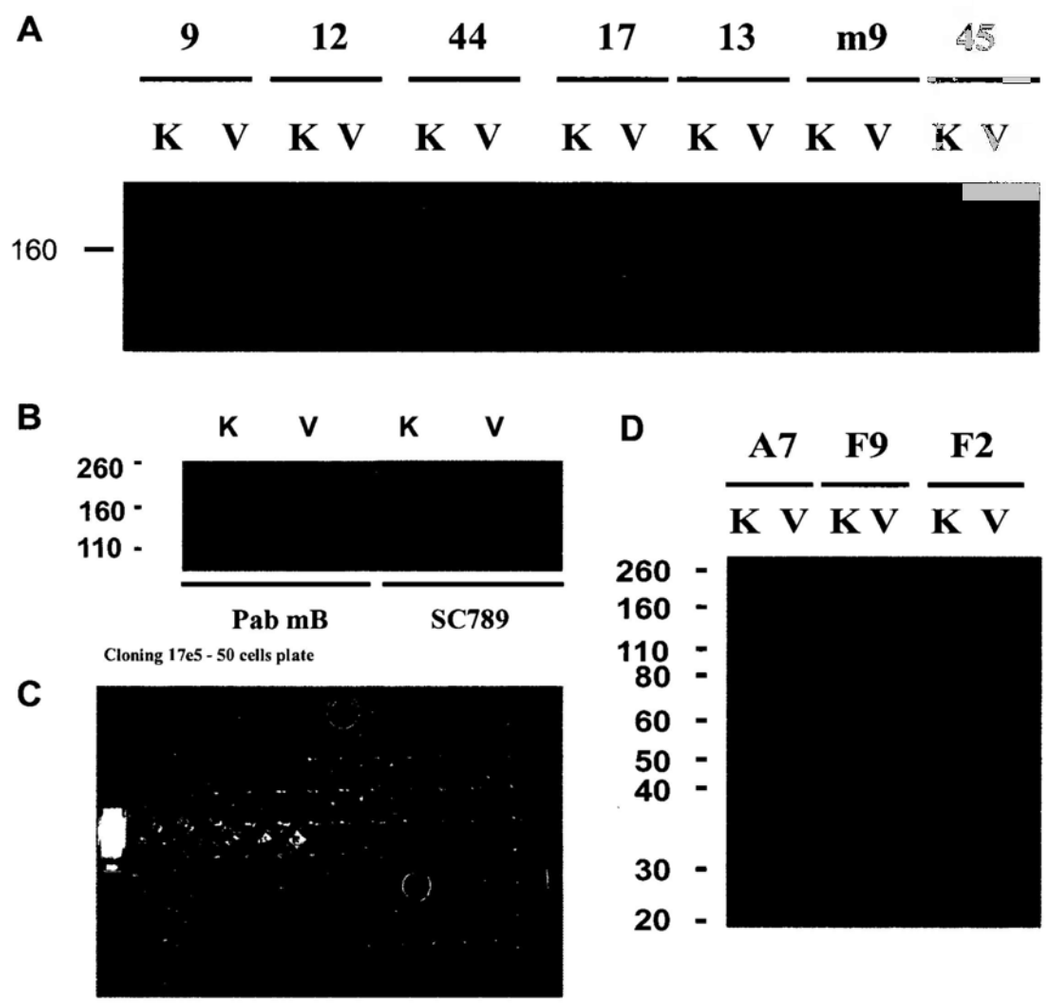


Figure 4-6. Generation of Monoclonal antibody. (A) HEK293 FT cells were transfected with Kmyc (K) or pCMV-myc vector (V) as indicated and used as antigen for WB analysis of Hybridoma supernatants. The blots were probed with hybridoma supernatants of clone 9, 23, 4, 17, 13 and 45. mouse 9 polyclonal antiserum at dilution 1:5000 (m9) were included as positive control. **(B)** WB showing that mouse B polyclonal antibody detect Kmyc. HEK293 FT cells were transfected with Kmyc (K) or pCMV-myc vector (V) as indicated and used as antigen for WB. The blots were probed with mouse B polyclonal antiserum at dilution 1:1000 (mB) or rabbit anti-myc polyclonal antibody at dilution 1:200 (SC789). (C) ELISA analysis of cloning 17e5 - 50 cells plate. Two strong ELISA positives 17E5A7 and 17E5F9 were circled. (D) WB analysis of monoclonal supernatants. HEK293 FT cells were transfected with untagged Kpcdna (K) or pcDNA3.1 vector (V) and processed for WB analysis with three days monoclonal supernatants 17E5A7 (A7), 17E5F9 (F9) or 17F2(F2). Dye front was not run off.

44 and 45 were frozen and analyzed by WB (Figure 4-6A). All clones were WB positives and clone 17 was selected for cloning into 50 cells/plate, 500 cells/plate and 3,000 cells/plate. ELISA was carried out after 7 days and 27 positives clones were frozen. The second cloning was carried out using the best ELISA positive clone, 17E5. After 10 days, ELISA was carried out for the 17E5 50 cells/plate (Figure 4-6C) and all positives clones were frozen. Since positives were obtained in 50 cells/plate, ELISA screening of the 3000 cell/plate was not done. Two strong ELISA positives 17E5A7 and 17E5F9 were analyzed by WB (Figure 4-6D) and selected for third round of cloning. The ELISA result of the 20 cells/plate of the third cloning demonstrated that all the wells that contained cells were positives, therefore the cell line was assumed to be purely cloned. The hybridomas of the third cloning were frozen more than once and the highest ELISA positive clone (17E5F9G4) was expanded to collect supernatant for further testing.

Full name of the monoclonal cell line: Fusion P124 mouse B 17E5(50)F9(20)G4.

Short name: Mab G4.

4.6.2 Monoclonal antibody is IgG isotype

Because the secondary antibodies used in ELISA and WB are anti-mouse IgG, the Mab G4 is most probably IgG. However, in some cases, the anti-mouse IgG antibody may cross react with IgM, therefore ELISA using isotype specific anti-mouse IgG or isotype specific anti-mouse IgM as secondary antibody was performed (Figure 4-7). The monoclonal antibody G4 was isotyped as IgG, which is a 150 kDa monomeric immunoglobulin that harbours two antigen binding sites.

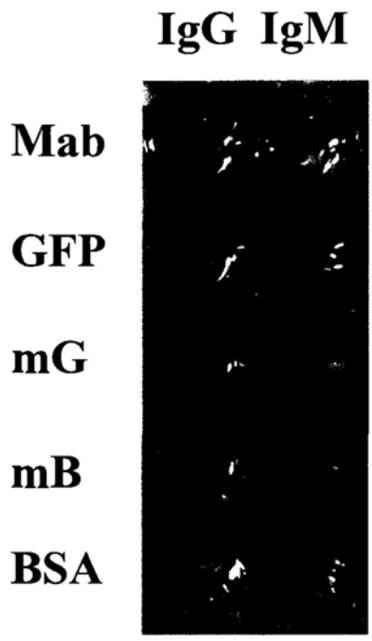


Figure 4-7. Isotyping of the monoclonal antibody. *0.1ug/well of peptide p124 conjugated to* BSA were coated onto 96-well plate in coating buffer. Primary antibody including monoclonal antibody G4 (mab), mouse anti-gfp antibody (GFP), normal mouse IgG (mG), mouse B antiserum (mB) and 5% BSA (BSA) were incubated with the plate at 4oC overnight. HRP conjugated isotype specific secondary antibodies goat anti-mouse IgG (IgG) or anti-mouse IgM (IgM) were added as indicated.

4.7 Demonstration of antibody specificity

4.7.1 Monoclonal antibody specifically recognize denatured K protein

To demonstrate that Mab G4 is specific to K, it was incubated with P124 in order to block specific binding (Figure 4-8A). Monoclonal supernatant, but not pre-serum, detects Kpcdna. Mouse B antiserum also detect the same band, which serves as positive control. Pre-absorption of the antibody, with the original immunizing peptide sequence (P124), prevented subsequent recognition of K. Pre-absorption of the antibody, with irrelevant peptide (P663) shows no blocking effect. All blots were reprobe with Mab G4, which shows that they contain Kpcdna. Therefore, Mab is specific to the peptide P124 and would specifically recognize K protein.

4.7.2 Monoclonal antibody is specific to K and does not recognize KL

Because of the similarity between K and KL, it is important to know whether the antibody would also recognize KL. We evaluated the specificity of Mab to K and KL in HEK293FT cell lines separately transfected with Kpcdna, Kmyc, KLmyc KLgfp and Kgfp (Figure 4-8B). WB analysis with mouse anti-GFP monoclonal antibody demonstrated that KLgfp was expressed but not detected by Mab G4. Results revealed that Mab G4 specifically recognized K and has no cross reactivity to KL.

4.8 Monoclonal antibody can detect native K protein *in vivo*

Knowing that the Mab G4 recognize denatured K protein *in vitro*, it is also important to test whether it would detect native K protein *in vivo*. For this reason, Mab G4 were

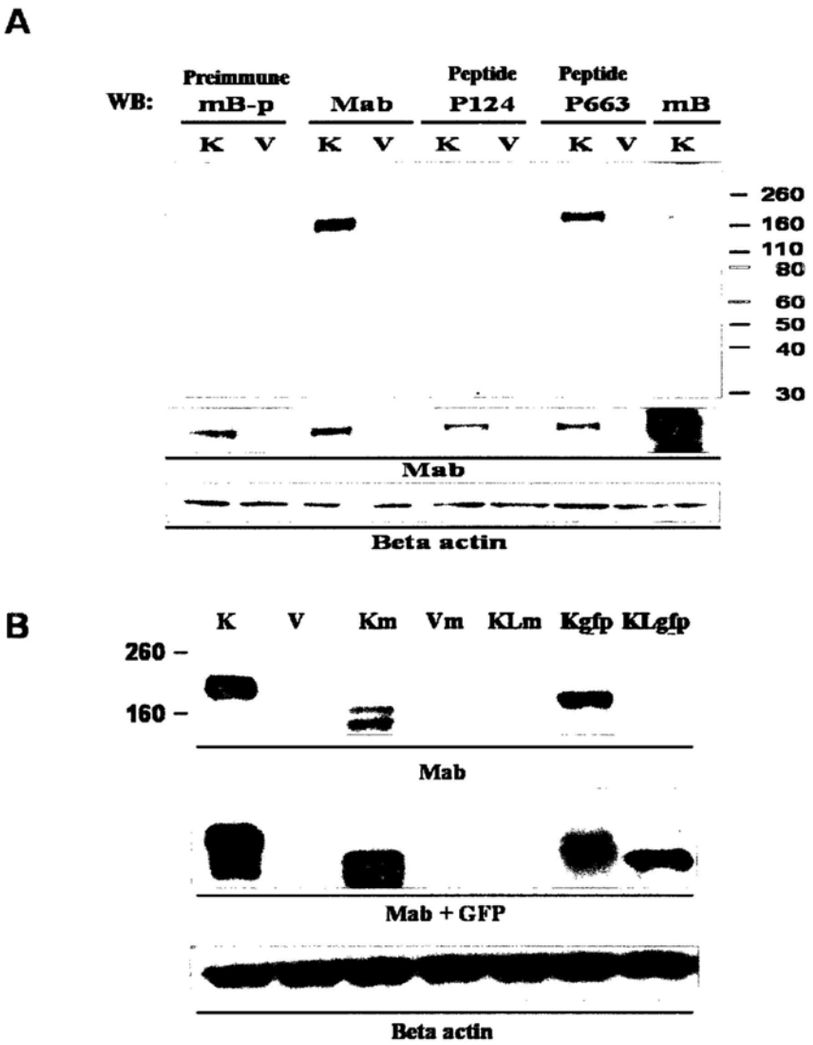


Figure 4-8. Characterization of Monoclonal antibody G4. (A) Western blot showing that mouse monoclonal antibody is specific to the peptide p124. HEK293FT cell lines were transfected with untagged Kpcdna (K) or pcDNA 3.1 vector (V) and processed for WB analysis with mouse B preserum (mB-p) at dilution 1:1000, mouse Mab G4 at dilution 1:100 (Mab), monoclonal antibody pre-incubated with peptide p124 (p124), monoclonal antibody pre-incubated with peptide p663 (p663) and mouse B antiserum (mB) at dilution 1:1000. **(B)** WB showing that Mab G4 recognise Kmyc, Kgfp and untagged Kpcdna but not KLgfp. Kpcdna (K), pcDNA 3.1 vector (V), kmyc (Km), KLmyc (KLm), pcmv-myc vector (Vm), KLgfp (KLg) and Kgfp (Kg) were overexpressed in HEK293FT cell lines and detected with mouse anti-KIAA0319 monoclonal antibody (Mab) and mouse anti-GFP antibody (GFP). The blots were reprobed with mouse anti-beta actin antibody (Beta actin).

used to detect the cellular localization of untagged K protein in HEK293FT cell line (Figure 4-9A). The Mab stained Kpcdna overexpressed cells while peptide absorption control pre-incubated with P124 give no signal under the same condition. Therefore Mab G4 can recognize native K proteins *in vivo*.

4.9 Monoclonal antibody capable for Immunoprecipitation

A specific Mab that would detect native K protein *in vivo* was generated. To employ this Mab for studying K's interacting partner, it is important to examine whether the Mab G4 can be used for IP. HEK293FT cell lines were separately transfected with Kpcdna or Kgfp and immunoprecipitated with Mab G4. The immunoprecipitated proteins were detected by Pab m9 (Figure 4-9B). In a separate experiment, the Kpcdna overexpressed lysate immunoprecipitated with Mab G4 were analyzed by SDS-PAGE and visualize by coomassie staining (Figure 4-9C). The presence of K proteins in both cases shows that the Mab specifically recognized full-length K proteins and is capable for IP analysis.

4.10 Harvesting Monoclonal antibody in serum free medium

To harvest a large amount of specific anti-K Mab for downstream applications, the Mab G4 was expanded in T150 cell culture flask and 50ml supernatants were collected every 3 days. The addition of serum used to culture hybridoma cells for Mab production has some disadvantages especially when the maximal antibody purity and strict quality are essential. The high protein content in FBS can be problematic for downstream applications, which often require additional processing steps for the purification of Mab from supernatants containing FBS (Basalp et al. 2000). Therefore, hybridomas G4 were

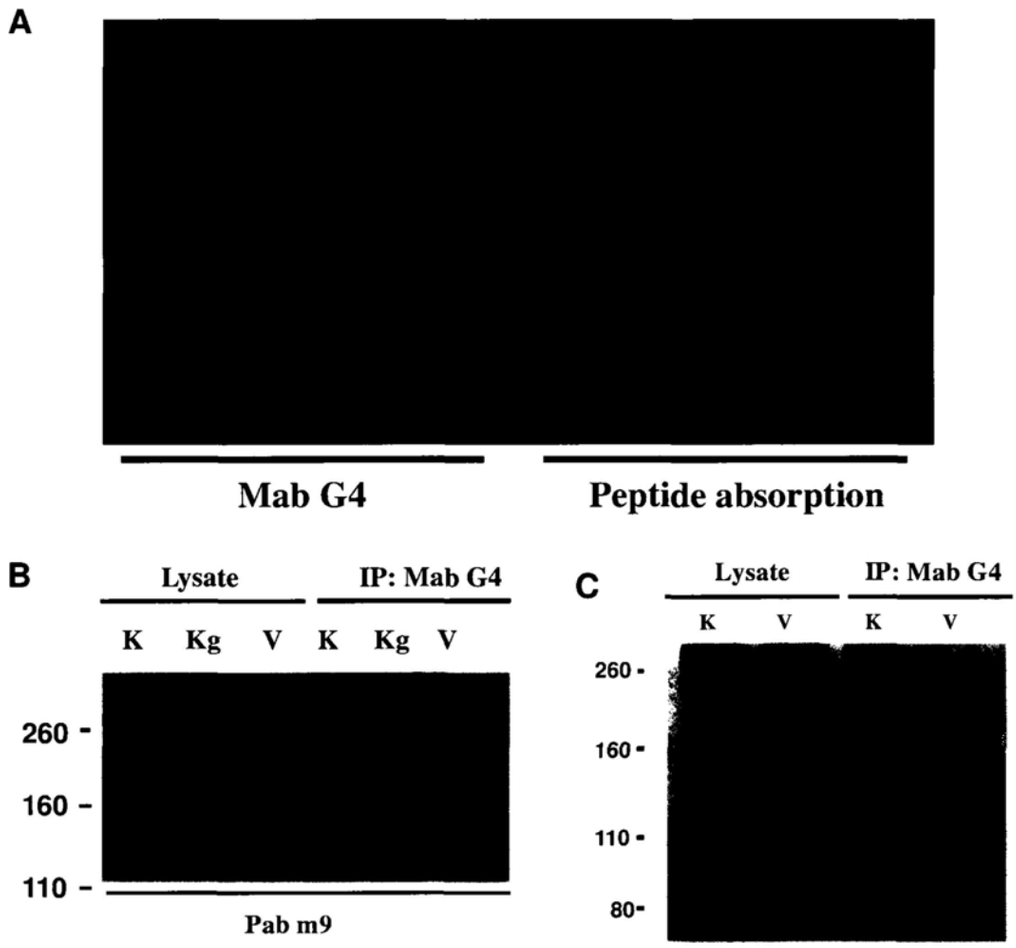


Figure 4-9. Monoclonal antibody G4 capable of detecting K protein in Immunofluorescence and Immunoprecipitation. **(A)** Immunofluorescence showing that mouse anti-KIAA0319 monoclonal antibody G4 is reacting specifically to KIAA0319 protein. HEK293 FT cells were transfected with untagged Kpcdna. Cells were then fixed and probed with Mab G4 at a 1:10 dilution or Mab G4 pre-incubated with peptide P124. Cells were then probed with goat anti-mouse immunoglobulin G conjugated to Dylight 488 at a 1:100 dilution. The nuclei of the cells were stained with PI. **(B)** Monoclonal antibody capable for Immunoprecipitation analysis. HEK293FT cell lines were transfected with Kpcdna (K), Kgfpc1 (Kg) or pegfpc1 vector (V) as indicated. Immunoprecipitation was performed with mouse anti-K monoclonal antibody (Mab G4) and the immunoprecipitated proteins were detected by mouse anti-K polyclonal antibody (Pab m9). **(C)** Immunoprecipitation was performed as in (B) and the immunoprecipitated proteins were visualize by Coomassie stain.

adapted to grow in serum free medium and harvest 5 was free of FBS. Nine harvests were collected which give a total of 450ml monoclonal antibodies. WB shows that Mab harvest 1-6 can detect untagged Kpcdna and no non-specific band was observed (Figure 4-10). Harvest 5 that was collected in the absence of serum also detects K, which can be further purified by protein G beads but was used as is in medium. Harvest 6 was collected after adding back serum and the strong band detected demonstrated that the cells were still healthy and actively producing antibody after the starving of FBS. Therefore, as demonstrated by validations by peptide absorption, specificity against KL protein, IF and IP, a Mab that would specifically bind denatured K protein as well as native K proteins was generated, which would be used in various immunoassays including IP.

4.11 Conclusion

In summary, in order to examine the interactions between K and other proteins, we generated specific Pab and Mab using synthetic peptides directed against the N-terminal of human K protein. The antibodies are specific to K in WB analysis and do not cross react with KL. Our results demonstrated that the Pab and the Mab detected denatured K protein as well as native K proteins under both *in vivo* and *in vitro* conditions, which exhibited the best performance in a variety of assays including immunoblotting and indirect immunofluorescence staining. Therefore, these antibodies likely recognized linear epitopes in the K protein and probably also work in immunohistochemistry. They are usable for subcellular localization assay such as examining the effect of cytoskeletal drugs on K's localization. In addition, the Mab were capable of recognition of native

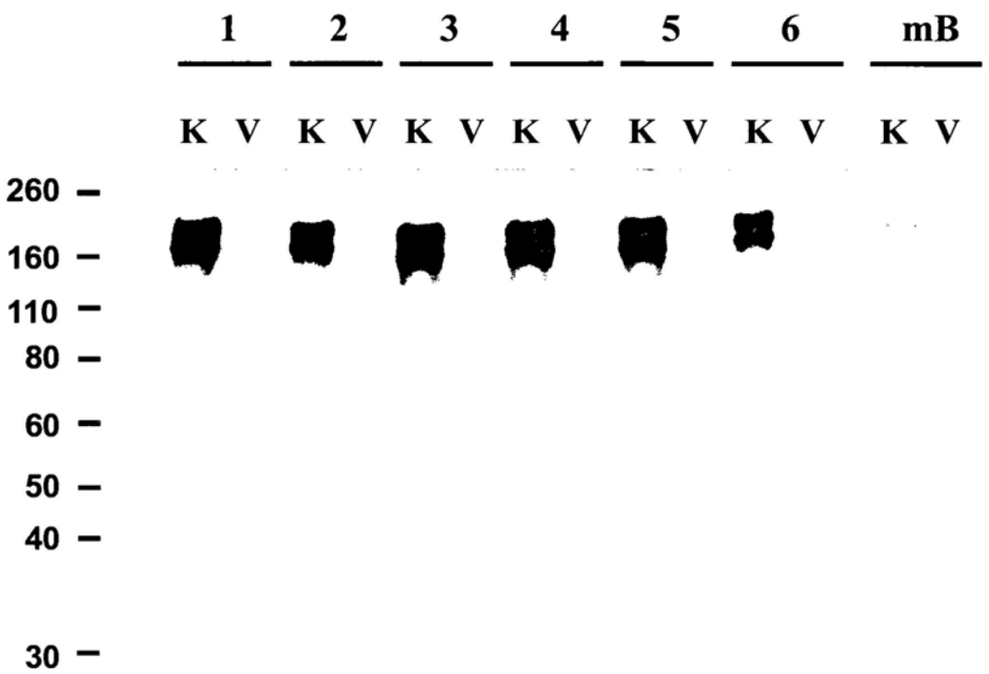


Figure 4-10. Western blot analysis of monoclonal supernatants. HEK293 FT cells were transfected with untagged Kpcdna (K) or pcDNA 3.1 vector (V) and processed for WB analysis with monoclonal supernatants harvest 1-6 at dilution 1:50. Mouse B polyclonal antiserum at dilution a:1000 (mB) was included as positive control. Dye front was not run off.

full-length K proteins by immunoprecipitation and can be use for studying the interactions of K with other proteins. Having established the specific anti-K Mab capable for performing various assays including the study of protein-protein interaction by co-IP, the characterization of how down regulation of K associated to DD is made possible. A fully cloned monoclonal cell line and the antibodies produced provide an ideal reagent for further investigation of the function of K proteins, which could also be used for further research both academically and commercially.

CHAPTER 5

Interaction of KIAA0319 and its homologue KIAA0319-Like

5.1 Study of K protein-protein interaction

A specific anti-K Mab that is capable for various assays is now available. Given the disruption of one or more protein-protein interaction is most likely the reason why K deregulation would lead to impaired brain development, it is vital to identify K's interaction partner. This information would give us a clue of its pathway and thus provide a link to its function.

5.1.1 K and its homologue KL are very similar

It is fundamental to study the protein-protein interaction that is significant to DD, for this reason, proteins that might be related to DD, such as KL that has been shown to have marginal association to DD (Couto et al. 2008), become interesting targets. K and its homologue KL both express in brain and therefore could be important for brain's function (Couto et al. 2008). The two proteins are very similar in protein coding, contain nearly identical domain structure and are heavily glycosylated. Both K and KL harbours a signal peptide and MANSC motif at their N-terminus, a central PKD domains region, a Cysteine-rich region (C6 motif and TM domain) and a short cytoplasmic region. We performed computational analyses on the human K and KL protein sequences in BioEdit

software (Hall 1999). Using Similarity Matrix BLOSUM62, the two proteins have about 45.9% overall amino acid identity and over 59% amino acids similarity. This high degree of similarity is concentrated at the central region and C-terminus, in which the central PKD region of K and KL share as high as 74.8% similarity and 62.4% amino acid identity. The C-terminus of the two proteins are also very similar, with the cytoplasmic regions of K and KL sharing 54.7% similarity (42% identity) and the Cysteine rich region (C6 and TM) sharing as high as 66.7% similarity (55.6% identity). The remaining N-terminus regions are 30.9 % similar (17.7% identity) while MANSC domain showing 36.6% similarity and 22% identity. With this high similarity between K and KL proteins, the author suggested that although consistent replications in independent samples are required to confirm its association to DD, KL might be related to DD when considered alongside the supporting evidence for K (Couto et al. 2008).

5.1.2 KL as a potential interacting partner of K

K is reported to interact with itself between its Cysteine rich domains (Velayos-baeza et al. 2008), with C6 and TM regions most critical for its dimerization. Given that K and KL are highly similar in protein coding as well as domain arrangement, it is very likely that they may have similar function and interact with each other. Because of the above reasons and the fact that a very specific anti-KIAA0319 monoclonal antibody has been raised by us, it is very interesting to study whether or not K and KL will interact with each other.

5.2 K interact with its homologue KL

5.2.1 Interaction between untagged K and GFP tagged KL

To study the possibility of complex formation between K and KL, mouse anti-K monoclonal antibody G4 was used to pull down untagged K in HEK 293FT cells overexpressing untagged K and KLgfp. The immunoprecipitated material was subjected to WB analysis using mouse anti-GFP antibody (Figure 5-1A). The presence of KLgfp in cells cotransfected with Kpcdna but not in control vector transfected cells demonstrated that KLgfp was immunoprecipitated with Kpcdna. Reprobe with mouse anti-K Mab G4 shows that K was immunoprecipitated. As a loading control, the presence of mouse heavy chain IgG in the immunoprecipitate samples were compared. Since both Kpcdna and KLgfp are larger then 190kDa, to show that the band detected by GFP antibody was not Kpcdna, HEK293FT cells expressing Kpcdna was subjected to WB with GFP antibody (Figure 5-1B). The blot was reprobed with mouse anti-GFP antibody as well as Mab G4 and the result shows that Kpcdna was expressed in the lysate but it was not detected by GFP antibody. Cell lysates expressing KLgfp or Kgfp were included as positive control showing that the GFP antibody was working. Because GFP antibody do not recognise Kpcdna, the band detected by GFP antibody in the co-immunoprecipitation experiment was KLgfp. Taken together, these experiments show that Kpcdna do interact with KLgfp.

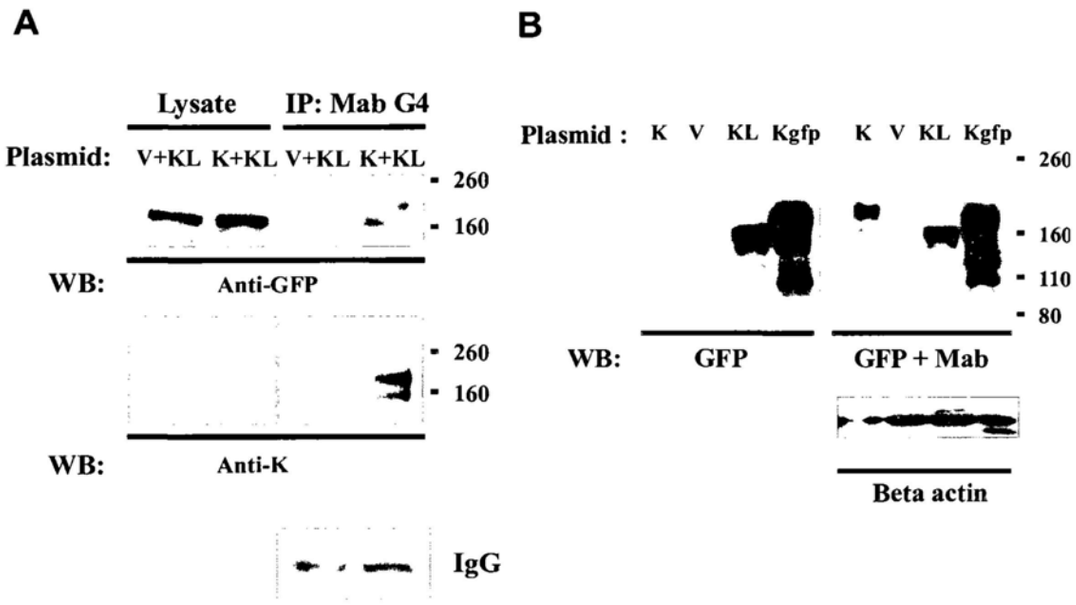


Figure 5-1. Co-immunoprecipitation of untagged KIAA0319 and GFP tagged KIAA0319-Like proteins. (A) HEK293-FT cell lines were co-transfected with pcDNA 3.1 vector (V), Kpcdna (K) or KLgfp (KL) as indicated. Co-IP was performed with Mab G4 and detected by western blot with mouse anti-GFP antibody. Total lysates were included to verify the expression of the tagged proteins. The Co-IP of KLgfp in cells co-transfected with K but not in control vector-transfected cells, indicates interaction of Kpcdna and KLgfp. Reprobe mouse Mab G4 (anti-K) shows that KIAA0319 was immunoprecipitated. As a loading control, the presence of mouse heavy chain IgG in the immunoprecipitate samples were compared. **(B)** Mouse anti-GFP antibody does not recognise Kpcdna protein. Kpcdna (K), pcDNA 3.1 vector (V), KLgfp (KL) and Kgfp (Kgfp) were overexpressed in HEK293FT cell lines and detected with mouse anti-GFP antibody (GFP) or Mab G4 (Mab).

5.2.2 Interaction between GFP tagged K and myc tagged KL

Proven that untagged K interacts with GFP tagged KL, we would like to further investigate this interaction. For consistency of the experiments and to show that N-terminal tag of K does not disrupt this protein-protein interaction, Co-IP was repeated with myc tagged KL and GFP tagged K. Cells co-transfected with Kgfp and KLmyc were immunoprecipitated with an rabbit anti-KL antibody H635 or with normal rabbit IgG as a control, and the immunoprecipitated fractions were subjected to WB with anti-K Pab m9. K was found to co-immunoprecipitate with KL in KL-myc co-transfected cells Co-IP with H635 but not in control cells Co-IP with normal rabbit IgG (Figure 5-2A). The blot was stripped and reprobed with H635, showing that KL was immunoprecipitated. Our group has demonstrated that KLmyc has three differently glycosylated forms, and the size were 140 kDa, 160 kDa and 190 kDa. The size of KL immunoprecipitated with K was larger then the size of KL in cell lysate, which corresponds to KL glycosylated form a. To further confirm the interactions and to confirm that the band was KL and not due to non-specific binding, reciprocal pull down was performed. Cells co-transfected with Kgfp and KLmyc were immunoprecipitated with mouse anti-GFP monoclonal antibody. The immunoprecipitated complex was analyzed by WB with either H635 or H635 preincubated with KL peptide (Figure 5-2B). The absence of bands in the peptide absorption control shows that the bands detected by H635 were KLmyc. This experiment demonstrated that KLmyc was pulled down by Kgfp and not by GFP vector, thus confirming the interaction. The blot was then stripped and re-probed with Pab m9, showing that Kgfp was immunoprecipitated in both cases.

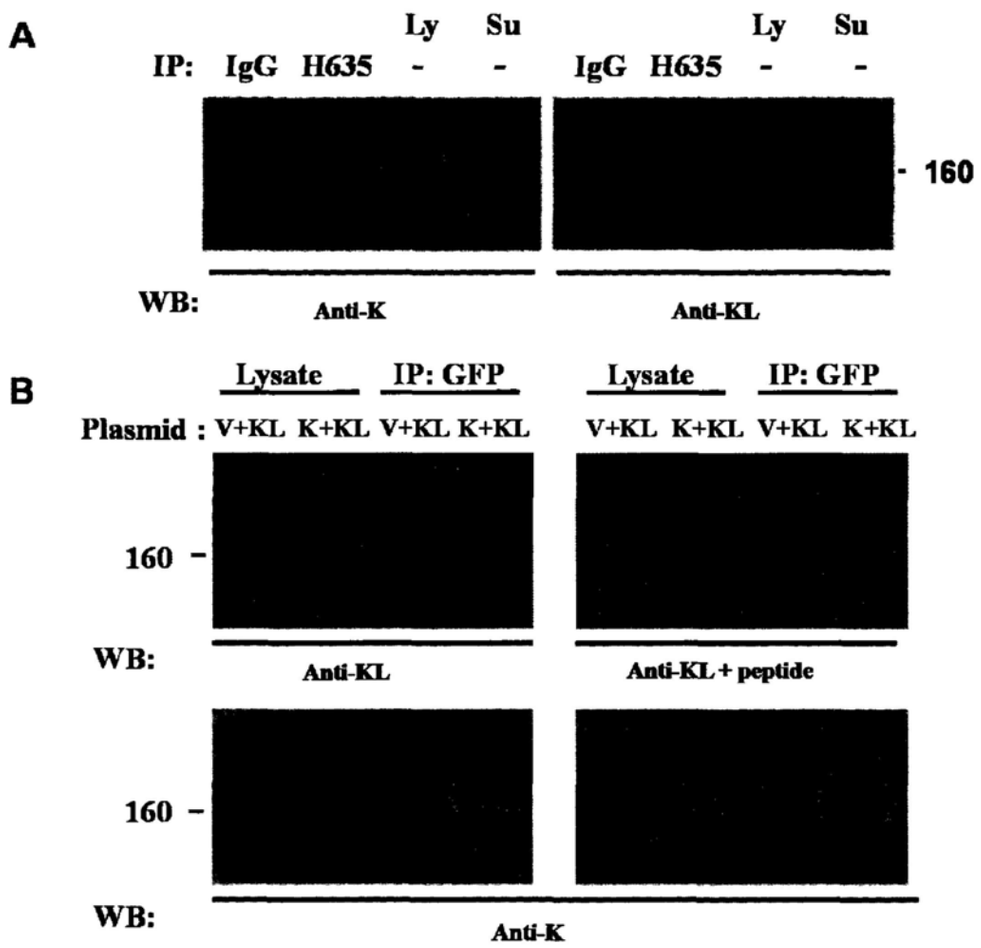


Figure 5-2. Co-immunoprecipitation of GFP tagged KIAA0319 and myc tagged KIAA0319-Like proteins. (A) HEK293-FT cell lines were co-transfected with KLmyc and Kgfp. Co-IP was performed with rabbit anti-KL antibody (H635) or normal rabbit IgG (IgG). WB with Pab m9 shows that Kgfp was coimmunoprecipitated in KL-myc co-transfected cells co-IP with rabbit anti-KL antibody (H635) but not in control cells Co-IP with normal rabbit IgG (IgG), indicates interaction of KIAA0319 and KL. Reprobe H635 shows that KL was successfully pulled down by H635 but not by rabbit IgG. Total lysate (Ly) and precleard supernatant (Su) was included to shows that KLmyc and Kgfp were expressed in both samples. **(B)** Reciprocal Co-IP with peptide absorption control. HEK293-FT cell lines were co-transfected with peGFPC1 vector (V), Kgfp(K) or KLmyc (KL) as indicated. Co-IP of KIAA0319-gfp(K) was performed with mouse anti-GFP antibody and detected by WB with either rabbit anti-KL antibody (H635) or H635 pre incubated with KL peptide. Total lysates were included to verify the expression of the tagged proteins The Co-IP of KL in cells co-transfected with K but not in control vector-transfected cells, indicates interaction of KIAA0319 and KL. The specificity of the band of KL was test by peptide absorption control (H635+peptide). Reprobe mouse anti-K Pab (m9) shows that KIAA0319-GFP was immunoprecipitated.

The results with N-terminal tagged proteins and untagged protein both demonstrated that K and KL do interact with each other. Taken together, our results demonstrated that the N-terminal tag does not affect interaction of K to KL.

5.3 Endogenous K and KL colocalized in glioblastoma cells

Colocalization experiments in cell culture can give information about the spatiotemporal dynamics of the protein-protein interactions. K interacts with KL, and therefore these two proteins should reside on the same intracellular compartment. Therefore, we first tested whether endogenous K and KL are colocalized in glioblastoma cells. Human glioblastoma-astrocytoma cells (U87MG) were co-stained with the anti-K Pab m9 and the anti-KL antibodies H635 (Figure 5-3). While K protein is mainly localized in the cell membrane with some in the cytoplasm, KL is localized throughout the cell in a more diffused pattern and is mainly localized in the nucleus. The nuclear localization of KL is consistent with the results reported for another peptide-purified rabbit polyclonal monospecific antibody for KL as reported in the Protein Atlas, although in that database they have remarked that their results has only low validation score: http://www.proteinatlas.org/gene_info.php?antibody_id=11169&g_no=ENSG00000142687) . There is a partial overlap between their localization and that endogenous K and KL proteins are co-localized in the peripheral region and the neurites of U87-MG. These could be the sites of their interactions, which further support our results that they do interact.

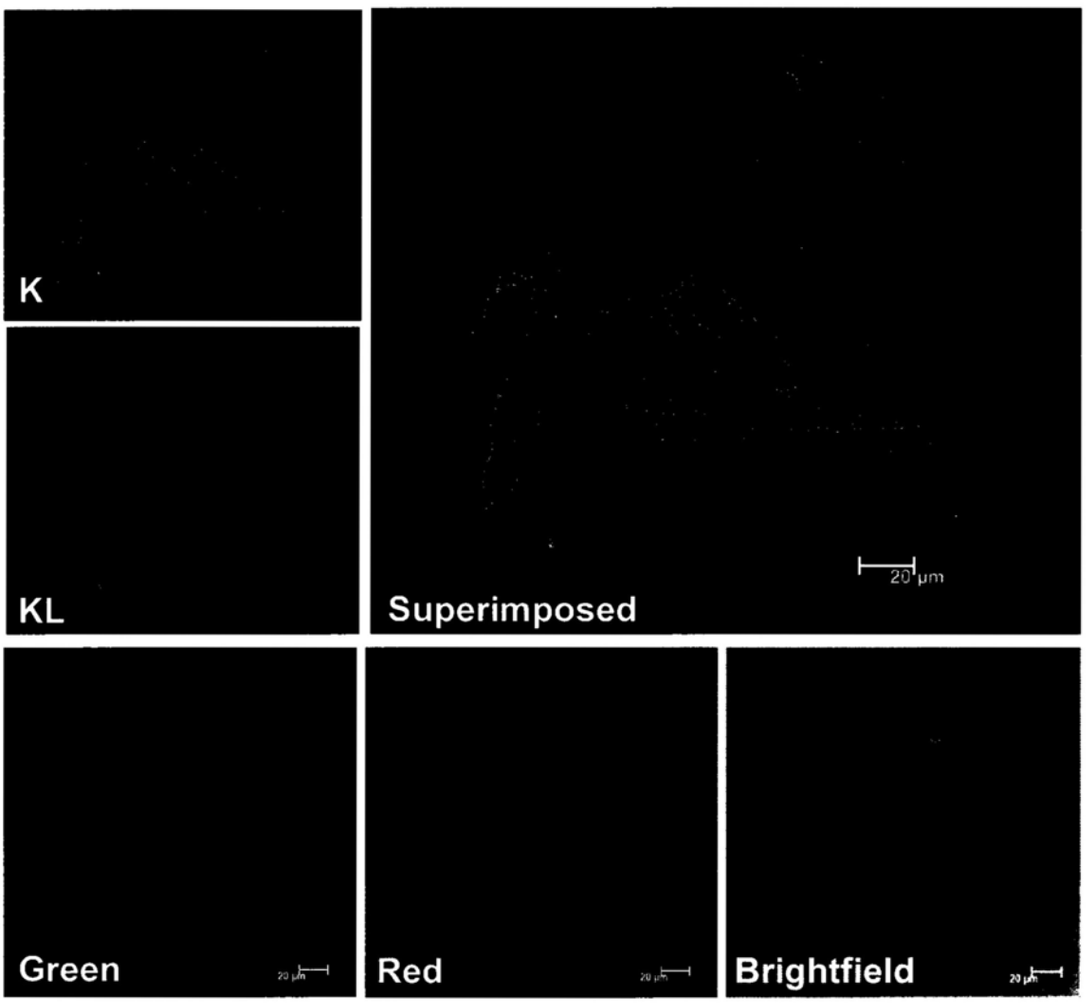


Figure 5-3. Immunofluorescence staining pattern of endogenous K and KL. Human glioblastoma-astrocytoma cell line (U87MG) were fixed and then probed with Pab m9 (K) at a 1:100 dilution or Pab H635 (KL) at a 1:100 dilution. Cells were then washed and probed with DyLight 488 AffiniPure Goat Anti-Mouse IgG (H+L) (Jackson ImmunoResearch) and Rhodamine (TRITC) AffiniPure Donkey Anti Rabbit IgG (H+L) (Jackson ImmunoResearch, USA.). No obvious staining was observed in all control slides treated with the above procedure but in the absence of primary antibodies.

5.4 K and KL interact at Cys rich region

It is clear that K interacts with KL, but the region which they interact is still unknown. For this reason, truncation mutants were created to analyse the effect of removing specific domains or regions within the K and KL protein on this interactions. As mentioned in introduction, K forms homodimers at its Cys rich C-terminus. This region has a high similarity between K and KL and could possibly be the site where they interact. Nevertheless, other possibility cannot be excluded. Therefore, K and KL proteins were roughly divided into 3 parts: the N-terminus including MANSC domain, the central PKD region and the C-terminus contains C6, TM and the cytoplasmic region. Accordingly, two deletion constructs were created, which include the N-deleted construct that contains the central PKD region as well as the cytoplasmic region and the C-deleted construct that contains the MANSC domain as well as the central PKD region (Figure 5-4). The deletion constructs were myc or GFP tagged at the N-terminus and the normal reading frame was always preserved.

Co-IP of myc tagged N-deleted mutants of K (KΔNmyc) or myc tagged C-deleted mutants of K (KΔCmyc) with GFP tagged full length KL shows that KLgfp interacts with KΔNmyc and not KΔCmyc. Reprobe with H635 shows that full length KLgfp was immunoprecipitated in both cases (Figure 5-5A). Reciprocal experiment of full length Kgfp with KL deleted mutants (KLΔNmyc and KLΔCmyc) also show that full length K interact with KLΔNmyc but not KLΔCmyc. Reprobing with Pab m9 shows that full length Kgfp was immunoprecipitated in both cases (Figure 5-5B). These results

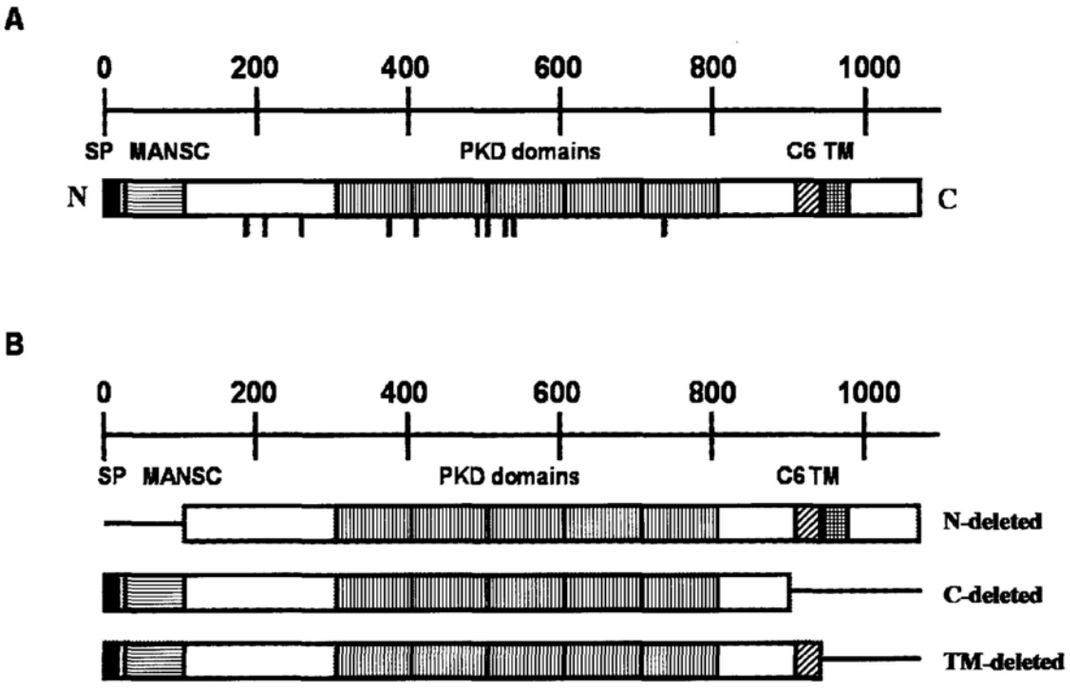


Figure 5-4. The human KIAA0319 protein and deletion proteins. (A) Graphical representation of the full length human KIAA0319 protein (K). **(B)** Representation of the deletion proteins used in this work, including N-deleted mutant, C-deleted mutant, TM-deleted mutant, all cloned into pCMV-myc and peGFP-C1 vectors. The deleted regions of the protein are represented by black lines. Details of deleted amino acids are indicated in Table 2-4.

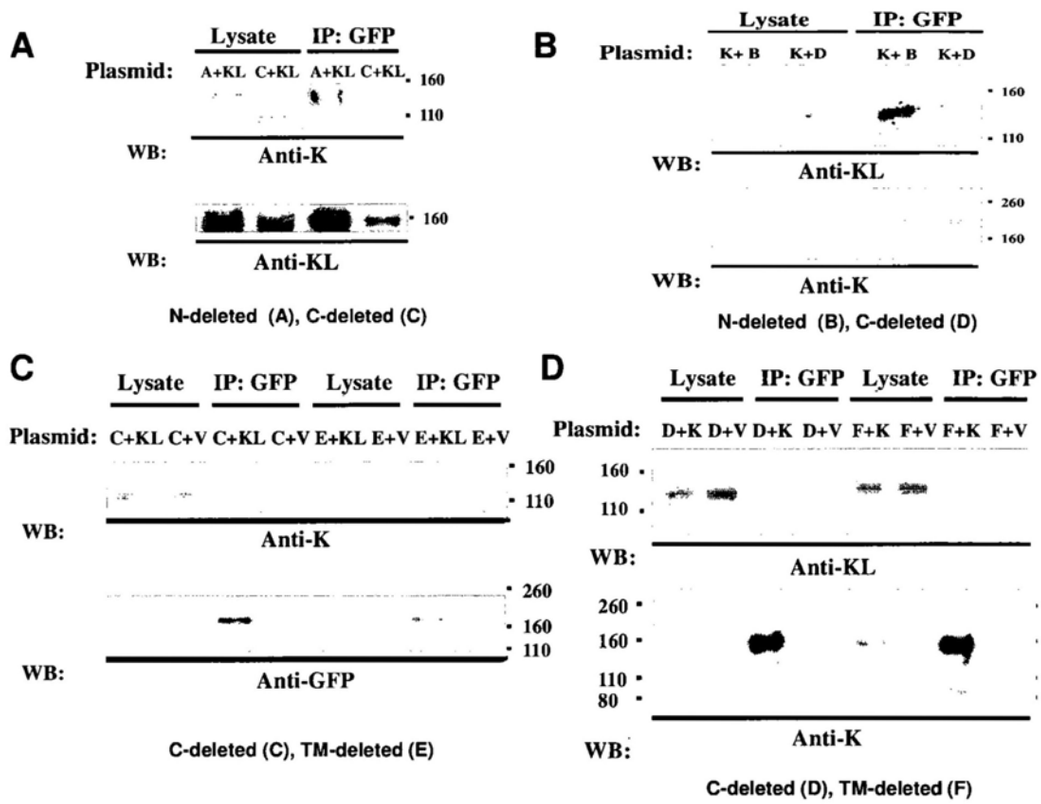


Figure 5-5. Interaction of K and KL deletion mutants. (A) Co-IP of full length KL and K deletion mutants. HEK293FT cell lines were co-transfected with N-deleted Kmyc (A) , KLgfp (KL) and C-deleted Kmyc (C) as indicated. Co-IP was performed with anti-GFP antibody (GFP) and detected by WB with Pab m9. Total lysates were included to verify the expression of the tagged proteins. KLgfp pull down N-deleted Kmyc but not C-deleted Kmyc, indicates that KLgfp interacts with C-terminus of K protein. Reprobe H635 shows that KL were immunoprecipitated in both cases. **(B)** Co-IP of full length K and KL deletion mutants. HEK293FT cell lines were co-transfected with Kgfp (K), N-deleted KLmyc (B) and C-deleted KLmyc (D) as indicated. Co-IP was performed with anti-GFP antibody (GFP) and detected by WB with H635. Total lysates were included to verify the expression of the tagged proteins. The Co-IP of N-deleted KLmyc but not C-deleted KLmyc with Kgfp indicates that Kgfp interact with C-terminus of KL protein. Reprobe Pab m9 shows that Kgfp was pull down in both cases. **(C)** Co-IP of full length KL and TM-deletion mutants of K. HEK293FT cell lines were co-transfected with KLgfp (KL), pegfpc1 vector (V), C-deleted Kmyc (C) and TM-deleted Kmyc (E) as indicated. Co-IP was performed with anti-GFP antibody (GFP) and detected by WB with Mab G4. The Co-IP of TM-deleted Kmyc and not C-deleted Kmyc by GFP antibody, indicates interaction of KLgfp with the C6 domain of K. **(D)** Co-IP of full length K and TM-deletion mutants of KL. HEK293FT cell lines were co-transfected with Kgfp (K), pegfpc1 vector (V), C-deleted KLmyc (D) and TM-deleted KLmyc (F) as indicated. Co-IP was performed with anti-GFP antibody (GFP) and detected by WB with H635. The Co-IP of TM-deleted KLmyc and not C-deleted KLmyc by GFP antibody, indicates interaction of Kgfp with the C6 domain of KL.

therefore confirmed that K and KL interact at their C-terminus.

K and KL interact at their C-termini, which are still very large regions including an extracellular region, transmembrane domain and a cytoplasmic region. In order to investigate which of these regions they interact, these regions are added back to the C-terminus deleted construct one by one. The TM-deleted constructs are created consequently, in which C6 is present and the region C-terminal to it is deleted (Figure 5-4B). Co-IP was performed with full length KLgfp and TM-deleted K mutant (KΔTMmyc). While KLgfp were immunoprecipitated in both cases, KLgfp pull down KΔTMmyc but not KΔCmyc (Figure 5-5C). In the reciprocal experiment with full length Kgfp and TM-deleted KL mutant (KLΔTMmyc), Kgfp pull down KLΔTMmyc but not KLΔCmyc (Figure 5-5D). Taken together, K and KL do interact when C6 is present, therefore C6 is the site where they interact. Nevertheless, because no deletion mutants with C6 deleted while TM region present was constructed, we cannot exclude the possibility that the Cys rich TM region may also be important for this interaction. In fact, the TM region of K and KL are also Cys rich, and in addition to C6, the TM region is also important for dimerization of K (Velayos-baeza et al. 2008). Therefore, TM is very likely also required for the interaction between K and KL. Taken together, Cys rich regions are very important for K and KL's interaction, consequently this interaction most likely involve disulphur bonding between several Cys residues in this region between K and KL, which is consistent with the report of how K dimerizes (Velayos-baeza et al. 2008).

5.5 K interact with KL due to sequence similarity

We have demonstrated that K and KL interact at the Cysteine rich region, which is the site involved in K's dimerization (Velayos-baeza et al. 2008). In fact, comparison of K and KL showed that their sequence and structure are very similar in this region (66.7% similar and 55.6% identical) with 9 out of 10 Cys residues conserved. Consequently the binding between K and KL is mimicking the binding of K with itself, due to their high similarity at the Cys rich region and its ability to dimerize.

5.6 Discussion

K and KL are highly similar, and they can form a heterodimer according to our results. The similarity between K and KL are surprisingly high. Apart from the highly similar protein sequence and structure, they have comparable domain arrangement and are heavily glycosylated. Furthermore, they form homodimers with itself as well as heterodimer at Cysteine rich region. Certainly, due to KL's amino acid sequence similarity with K and the ability to form heterodimer with K, it is possible that KL have similar function as K and thus link to DD *via* K. Functional study of K shows that alteration of K expression disrupt neuronal migration (Peschansky et al. 2009; Paracchini et al. 2006), and because neuronal migration greatly affect DD, K's role in neuronal migration is most probably why K would leads to DD predisposition. K and KL are both expressed strongly in the brain (Couto et al. 2008). If the two proteins share identical function in the neuronal migration pathway because of their high similarity, then in the migration experiment where K was knockdown by shRNA, as long as KL's high endogenous level is not affected, it should be able to compensate for the reduced

level of K. However, the impaired neuronal migration caused by down regulation of K was not rescued by endogenous KL level. In an unlikely event the shRNA directed against K could also down regulated endogenous KL owing to the sequence similarity between K and KL, then the impaired neuronal migration may caused by down regulation of both K and KL. Nevertheless, this impairment is rescued by K overexpression alone and KL overexpression was not necessary, which verified that the migration deficit is K specific (Peschansky et al. 2009). Accordingly, two possibilities arise. The level of endogenous KL was either not affected by the shRNA directed against K, or that the decreased level of KL does not significantly affect neuronal migration. If it was the former case that KL's level was not affected, then K's down regulation alone can lead to impaired neuronal migration and it proves that the high endogenous level of KL in brain could not compensate the effect caused by down regulation of K. Otherwise, if it was the latter case that KL was also down regulated, then the rescue experiment shows that K's expression alone is sufficient to mask the effect of KL's down regulation. Consequently, KL is either not critical to neuronal migration or that unlike K, it does not function in promoting neuronal migration. In any case, K and KL are undoubtedly different in terms of their role in neuronal migration. Presently, all of the candidate dyslexia susceptibility genes whose functions have been explored have been found to be involved in neuronal migration (Peschansky et al. 2009; Velayos-baeza et al. 2008). The difference in their role in neuronal migration explains why K might be more important to DD susceptibility.

Moreover, while K has restricted expression in brain with expression in fetal brain 5

times higher then its average expressions, KL do not display this restricted expression. KL is expressed in brain but its expression in fetal brain is lower then its average expression. (Su et al. 2004; Figure 5-6).

Besides the low fetal brain expression and the lack of a functional assay that can prove the role of KL in brain development, whereas K being a strong dyslexia candidate gene (rs9461045, P = 0.0001) (Dennis et al. 2009) with supports from functional studies (Peschansky et al. 2009), KL has only show marginal association (rs7523017; P = 0.042) that was not yet subjected to correction by multiple studies (Couto et al. 2008). Furthermore, no independent study has reported the association of KL to dyslexia. Indeed, our group's unpublished data shows that KL has no association to dyslexia in our sample (Lim et al, personal communication). Moreover, the paper also suggested that the causal variants which show marginal association to DD may not necessarily lies in KL, but could also mean association with neighbouring brain expressed genes that lie within the same region (Couto et al. 2008). Altogether, KL has only marginal association with DD and that strong and more consistent evidence for association of KL with DD is required before putting a linkage between KL and DD. Nevertheless, because of its high similarity to K, KL might be related to DD when considered alongside with the strong evidence of K in DD (Couto et al. 2008), and this high similarity is also the reason why they interact. Velayos-baeza and coworkers have suggested that some of the K functions would most probably require a dimer conformation of K (Velayos-baeza et al. 2008), therefore the K-KL interaction that mimics K dimerization might be important to the functions of K in neuronal migration or DD.

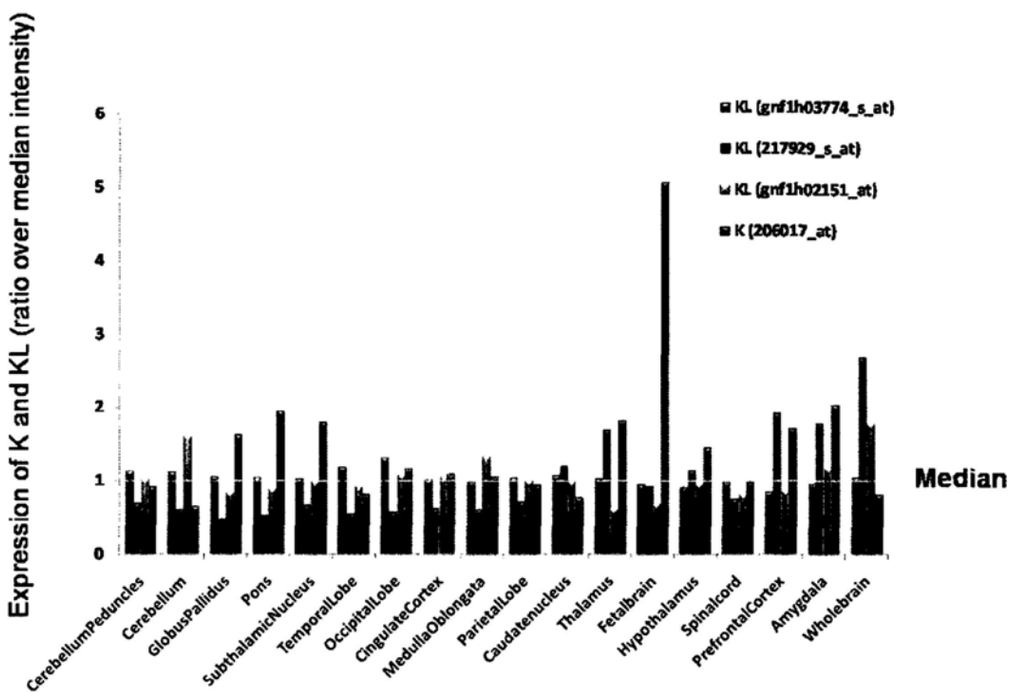


Figure 5-6. RNA expression pattern of K and KL in brain tissue by microarray. Microarray were performed with different tissues (Su et al. 2004; http://biogps.gnf.org). Expression of K and KL across different brain tissue were analyzed using different reporters as indicated. The gene expressions were presented as a ratio of the intensity over median intensity. Expression of K in fetal brain was 5-fold above the its average expression while expression of KL in fetal brain was lower then its average expression.

5.7 Conclusion

In conclusion, K and its homologue KL are very similar in their protein coding sequence and domain structure. Consequently, KL is an interesting target for studying K's function and also a potential interacting target. Co-IP experiment with untagged K and KLgfp have shown that K and KL do interact. Using Kgfp and KLmyc, similar results were obtained, confirming the interactions. Peptide absorption show that band that was recognised by anti-KL polyclonal antibodies was indeed KL and that Kgfp interact with KLmyc. We can deduce from our fusion tagged recombinant experiments that the N-terminal tag does not disrupt protein-protein interaction in this system. Immunofluorensence experiments have show that endogenous K and KL proteins are co-localized in the peripheral region and the neurites of human glioblastoma-astrocytoma cells, which could be the sites of their interactions. Using deletion mutants, the Cys rich region at the C-terminus was shown to be the critical region for the interaction of K and KL. Due to the high similarity between K and KL, it was not surprising that K and KL interact at the domain where K dimerizes, and this heterodimer conformation may be important for the functions of K. To enrich the knowledge of the relationship between K protein and DD, apart from the interaction of K with KL, interactions with other modifying factors would be an interesting topic worth studying.

CHAPTER 6

Protein-protein interactions of KIAA0319

6.1 K interacts with Adaptor protein 2 complex and follows the clathrin mediated endocytosis pathway

Selection for other interactions that are specific and important to DD is desirable and essential to understand K's function in brain development. Recently, K was shown to interact with the μ-subunit of the Adaptor protein 2 complex (AP2M1) (Levecque et al. 2009). Adaptor protein 2 complex (AP2) are clathrin-associated, heterotetrameric adaptor protein complexes, which mediate protein-sorting events at plasma membrane (Nakatsu & Ohno 2003; Lindwasser et al. 2008; Owen et al. 2004). Many plasma membrane proteins are internalized in response to specific signals, which is thought to regulate their functions (Yap et al. 2007). Clathrin mediated endocytosis is the best characterized route of protein internalization which internalize many cell surface receptors and integral membrane proteins (Levecque et al. 2009). Clathrin-coated vesicles belong to a major class of transport vesicles which are found associated with the cell's limiting membrane, the trans-Golgi network (TGN), and on some endosomes, as well as on the vesicular/tubular structures derived from them (Owen et al. 2004). Clathrin is a coat protein that serves as a mechanical scaffold but is itself unable to bind directly to membrane components. Sorting of cargo into the clathrin scaffold is mediated by clathrin adaptor proteins which recognize signals at the cytoplasmic region of the

cargo (Owen et al. 2004). AP2 complex is the main clathrin adaptor protein that regulates receptor-mediated endocytosis. A number of plasma membrane proteins including the receptors, transporters, adhesion molecules and virus products are known to be internalized in an AP2-dependent fashion (Nakatsu & Ohno 2003). AP2 complex composed of two large subunits (α, β2, 110–130 kDa), a medium subunit (μ2, approximately 50 kDa), and a small subunit (σ2, 15–20 kDa) (Robinson 2004; Owen et al. 2004). The most characterized sorting signals recognized by the μ2 subunit of AP2 include the tyrosine-based motif and dileucine based motifs (Nakatsu & Ohno 2003). Among them, the Yxxϕ motif (where X is any amino acid and ϕ is a bulky hydrophobic residue: leucine, isoleucine, methionine, phenylalanine, valine) is the main tyrosine based signal involved in protein sorting at plasma membrane (Nakatsu & Ohno 2003; Owen et al. 2004; Levecque et al. 2009; Ohno et al. 1995).

It was shown that the AP2M1 binding motif YTIL is located at the C-terminus of K (Levecque et al. 2009). As a membrane protein, the trafficking of K is essential for its function. By interacting with the μ2 subunit of AP2, the integral membrane protein K is internalized back into the cell *via* the clathrin mediated endocytosis pathway (Levecque et al. 2009). Therefore, it is very likely that the deregulation of K endocytosis is significant to DD and would allow us to gain further insight into its role in DD.

6.2 Interactions of full length K and AP2M1

The interaction of K and AP2M1 was identified in an Y2H screening using the cytoplasmic region of K as bait and verified with *in vitro* GST pull down experiments

using GST tagged K fragment (Levecque et al. 2009). To further confirm that full length K would bind to full length AP2M1 *in vivo*, HEL293FT cell lines coexpressed AP2M1myc and either Kgfp or Kpcdna were immunoprecipitated with rabbit anti-myc antibody SC789. The successful pull down of Kgfp and Kpcdna in the presence of AP2M1 but not in control vector transfected cells showed that AP2M1-myc does pull down Kgfp and Kpcdna (Figure 6-1). The results of tagged K and untagged K were consistent, which demonstrated that N-terminal tag of myc or gfp do not interfere with the K and AP2M1 interactions. Together with the previous experiments showing that tagged or untagged K both interact with KL, we can safely conclude that the N-terminal tagging of K does not obstruct protein interactions in our system. For the consistency of the experiments, N-terminal tagged construct were used for following Co-IP experiments.

These results confirmed that full length K and full length AP2M1 do interact *in vivo,* which further support K internalization by the clathrin mediated endocytosis pathway. The internalization of K is very important for its functions and therefore this interaction of K and AP2M1 could be significant to DD predisposition.

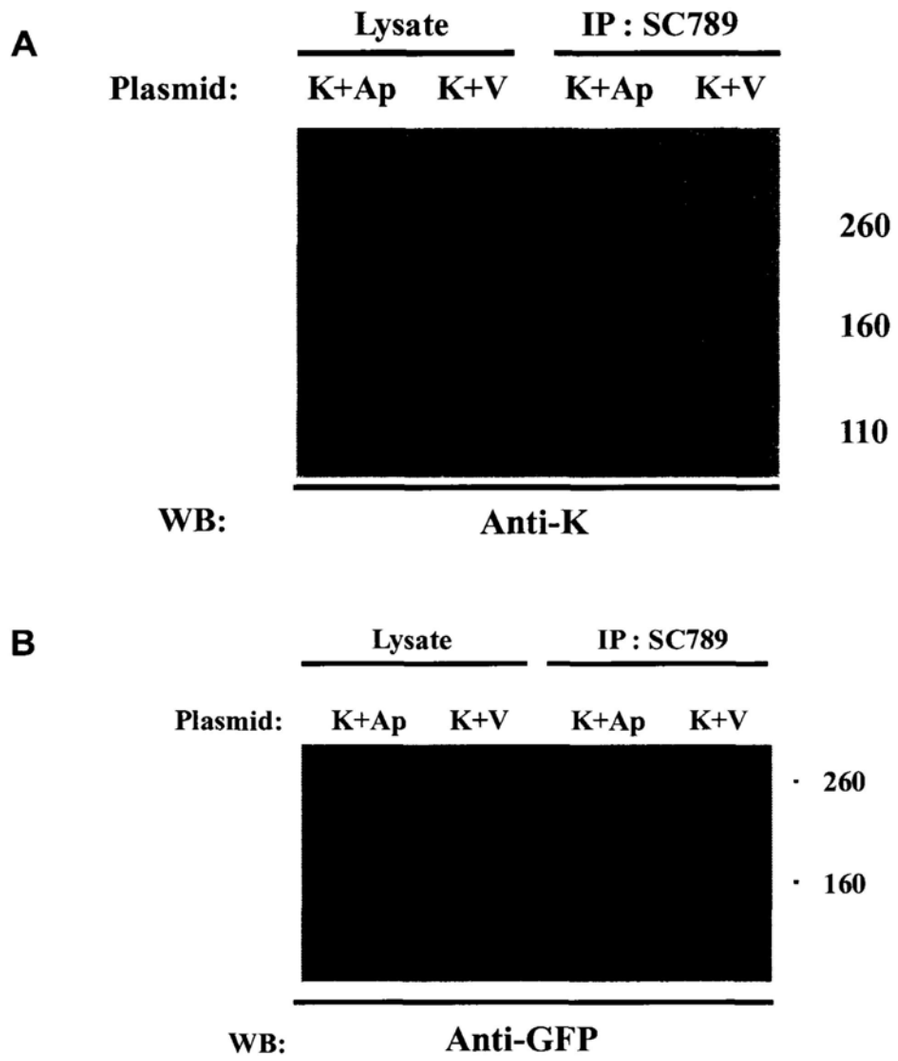


Figure 6-1. Interaction of K and AP2M1. (A) Co-IP of Kpcdna and AP2M1. HEK293FT cell lines were co-transfected with Kpcdna (K), AP2M1myc (Ap) and pCMV-myc vector (V) as indicated. Co-IP was performed with 30ul rabbit anti-myc polyclonal antibodies (SC789) and detected by mouse anti-K monoclonal antibody (Mab G4). The Co-IP of Kpcdna in cells cotransfected with AP2M1 but not in control vector-transfected cells, indicate interactions of AP2M1myc and Kpcdna. **(B)** Co-IP of Kgfp and AP2M1. HEK293FT cell lines were co-transfected with Kgfp (K), AP2M1myc (Ap) and pCMV-myc vector (V) as indicated. Co-IP was performed with 30ul rabbit anti-myc polyclonal antibodies (SC789) and detected by mouse anti-K monoclonal antibody (Mab G4). The Co-IP of Kgfp in cells cotransfected with AP2M1 but not in control vector-transfected cells, indicate interactions of AP2M1myc and Kgfp.

6.3 Interactions of KL and AP2M1

K and AP2M1 binds *in vivo*, which could be important for K's function in brain development. AP2M1 is widely expressed among different tissue with expression in whole brain, amygdala and prefrontal cortex three times over the median expression levels (Su et al. 2004). Although AP2 has not been found to be associated to DD, the interaction of K and AP2M1 could be important to DD. Given that the C-terminus of K and KL share a high similarity, it is interesting to know whether KL will also interact with K's interaction partners AP2M1 and follows the same pathway of trafficking. For this reason, study of the interaction between AP2M1 and KL, a gene that is highly similar to K and yet only weakly associated to DD, could give us a clue about the relationship between K, KL and AP2M1.

To investigate whether KL binds to AP2M1, Co-IP with KLgfp and AP2M1myc using rabbit anti-myc polyclonal antibody SC789 was performed (Figure 6-2A). Immunoprecipitated KLgfp was observed in cell cotransfected with AP2M1myc but not in control vector transfected cells, consequently KL and AP2M1 do interact with each other. GFP tagged KL were used in the experiments, and the N-terminal GFP tagged do not hinder interactions with AP2M1 was demonstrated in previous Co-IP experiments performed with untagged K and GFP tagged K. As a result, KL interacts with AP2M1 and it is essential to find out which region does they interact. The search of AP2M1 binding motif Yxxϕ in K and KL reveal several possible binding sites (Table 6-1).

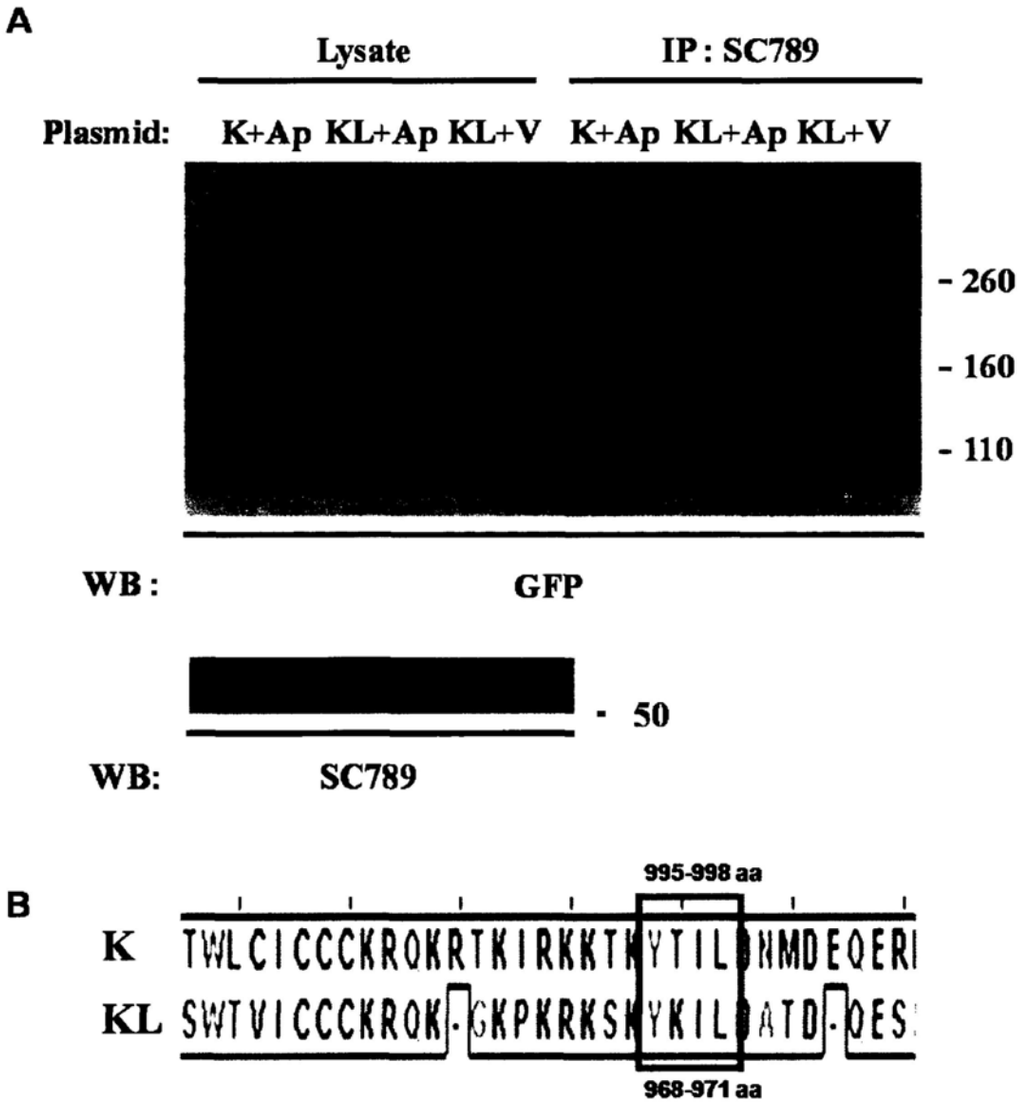


Figure 6-2. Interaction of AP2M1 and KL. (A) HEK293FT cell lines were co-transfected with Kgfp (K), KLgfp (KL), AP2M1myc (Ap) and pCMV-myc vector (V) as indicated. Co-IP was performed with rabbit anti-myc polyclonal antibodies (SC789) and analyzd by WB using mouse anti-GFP antibody (GFP). The Co-IP of KLgfp in cells cotransfected with AP2M1 but not in control vector-transfected cells, indicate interactions of AP2M1myc and KLgfp. Co-IP of Kgfp and AP2M1 was included as positive control. Reprobe with SC789 shows that AP2 was expressed. **(B)** Sequence alignment of K and KL protein at AP2M1 binding site. Protein sequence of full length K (GenBank Accession Number: NM_014809) and KL (GenBank Accession Number: NM_024874.3) were aligned. Alignment at regions of AP2M1 binding sites were indicated by black box.

Table 6-1. List of predicted AP2M1 binding motif (Yxxϕ) in K and KL proteins. All tyrosine residues within K and KL protein sequence were identified. Tyrosine residues that are followed by hydrophobic amino acids are predicted as potential AP2M1 binding motif. Sites that are at the cytoplasmic region are in bold.

K's AP2M1 possible binding motif: (Transmembrane region = 953-983aa)	KL's AP2M1 possible binding motif: (Transmembrane region = 927-957aa)
99 YLTF 102	32 YLFY 35
160 YREL 163	35 YTCF 38
182 YTDW 185	287 YATP 290
527 YPPV 530	303 YPVI 306
584 YLHL 587	393 YVNV 396
752 YLWI 755	501 YPPV 504
958 YVTV 961	932 YVII 935
995 YTIL 998	**968 YKIL 971**

Given that AP2M1 binds only to the sorting signal present in the cytoplasmic domains of protein cargo (Levecque et al. 2009), YKIL AP2M1 binding motif at 968-971 amino acids of KL was the only one predicted binding site that could possibly recognize by AP2M1. Aligning the protein sequence of K and KL reveal that K also posses a AP2M1 binding motif YTIL at 995-998 amino acids (Figure 6-2B), which is the sorting signal recognize by AP2M1 (Levecque et al. 2009). Therefore, it is clear that the YKIL AP2M1 binding motif at 968-971 amino acids of KL was the site recognized by AP2M1.

6.4 The binding properties of AP2M1 to K and KL is alike

Owing to the high sequence similarity at cytoplasmic region, the binding properties of AP2M1 to K and KL is hypothesized to be the same. To test this hypothesis, *In silico* techniques offer attractive alternatives to experimental methods (X-ray crystallography and NMR; Ramsland et al. 2009). In particular, molecular docking could provide important information about AP2M1-peptide interactions when crystallization of C-terminus of K and KL are difficult. Molecular docking identifies binding modes of a ligand by flexing and orienting it in a protein active site and scoring them (Imberty et al. 2008) in order to achieve an optimized conformation for both the protein and ligand such that the free energy of the overall system is minimized (Ali et al. 2010). For predicting the peptide orientation in binding sites, flexible docking methods have to be used in order to account for the possible orientations of pendent groups (i.e. hydrogen bond network directed by hydroxyl/hydroxymethyl group orientation) (Imberty et al. 2008). A molecular docking study, using molecular mechanics calculations with AutoDock (Morris et al. 1998) was used to help predict the binding properties of the

complex crystal structure of AP2M1 (PDB ID: 1H6E) towards two peptides, corresponds to 994-999 amino acid of K (KYTILD) and 967-972 amino acid of KL (KYKILD) through the differences in the interaction energies and inclusion geometries. AutoDock is based on a semiempirical free energy force field, and it uses a Lamarckian Genetic Algorithm (GALS) (Kalliokoski et al. 2009).The program is widely used in docking studies and can also be used for virtual screening, provided that supercomputing resources are available.

Docking simulation of 100 runs of GALS was performed for a set of 100 ligands into the binding site of AP2M1. During the searching process, the receptor was regarded as rigid, while ligands were regarded as being flexible. The best docked conformation was determined as the one which having the lowest interaction energies of both docking energy and binding free energy among the 100 different poses generated. For the docking of the K and KL ligand peptides against AP2M1, the binding free energies for the best conformation were −1.97 kcal mol^{-1} and -0.62 kcal mol^{-1} respectively. In the best docked conformation (Figure 6-3A), the tyrosine in binding motifs of K and KL ligands formed hydrogen bonds to AP2M1 (Figure 6-3B and C) and leucine at the +3 position to the tyrosine has strong hydrophobic interaction with AP2M1 (Figure 6-3D). These interaction properties are the same in K and KL. In fact, this binding mode of YXXϕ to AP2M1 is also the same for other protein such as cytotoxic T-lympocyte protein 4 (CTLA-4), revealed from the complex crystal structure of AP2M1 with its binding motif peptide (PDB ID: 1H6E; Figure 6-3E). This observation suggested that

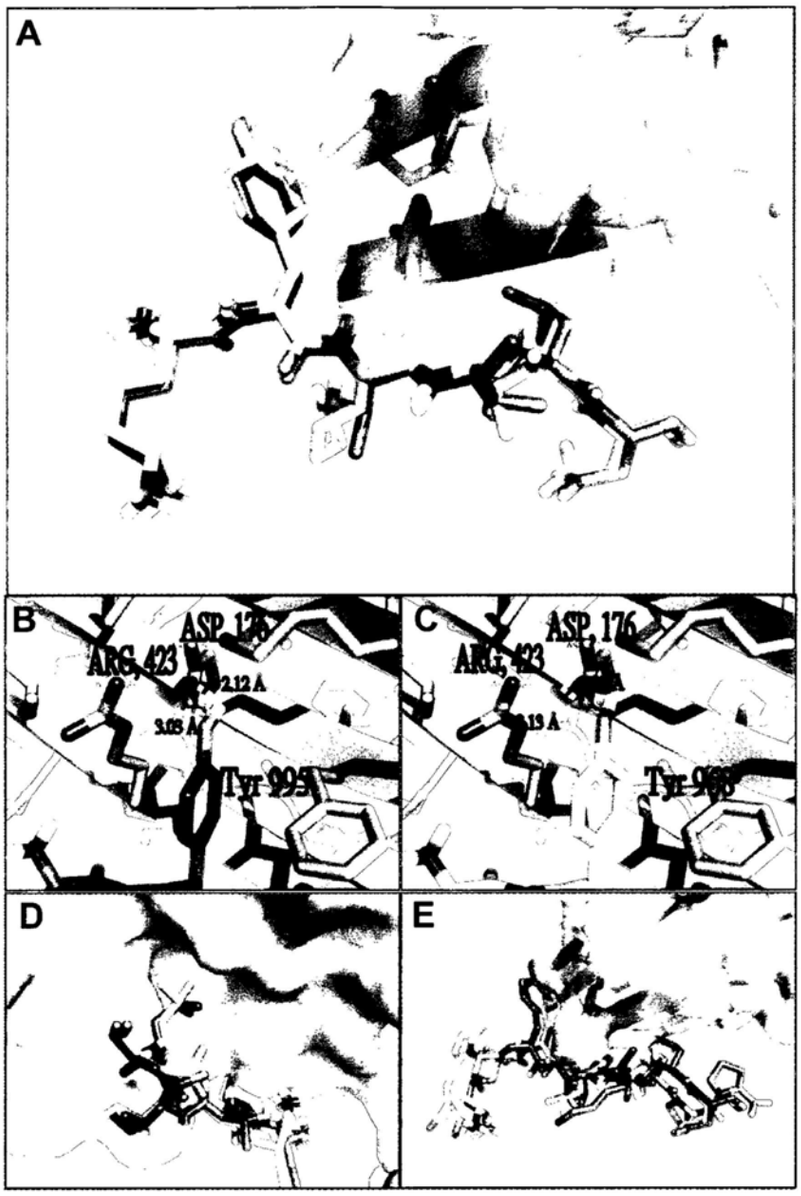


Figure 6-3. The best conformation of K and KL ligands docked into the binding site of AP2M1. (A) The docked ligands K and KL were exactly superimposed (shown in *sticks*, red for K ligand and yellow for KL ligand). **(B)** and **(C)** The hydrogen bonding (*yellow dotted lines*) between AP2M1 and the tyrosine in binding motifs of K (B) and KL (C). **(D)** In docked ligands of K and KL, the leucine at the +3 position to the tyrosine has strong hydrophobic interaction with AP2M1. **(E)** The docked ligands of K (red) and KL (yellow) as well as the original ligand peptide CTLA-4 (Cyan) binds to AP2M1 in similar way.

the binding of AP2M1 to K is not unique. Therefore, the interaction between K and AP2M1, although could still be important to DD *via* K trafficking, should not determine the significance of K in DD.

6.5 Interaction of full length K to two adaptor proteins

In addition to AP2M1, K has two potential interacting partners identified in a library scale yeast two hybrid screening (Nakayama et al. 2002), namely SH2B adaptor protein 1 (SH2) and human homologue of the Caenorhabditis elegans sex determination fem1 protein (FEM). SH2 is a member of a family of adaptor proteins that influences multiple ligand-activated receptor tyrosine kinases, including the receptors for nerve growth factor (NGF), insulin, IGF-I and the cytokine receptor-associated Janus tyrosine kinase (JAK) family kinases (Maures et al. 2009). It is a key enhancer of RET tyrosine kinase that has also been implicated in cell motility and regulation of the actin rearrangement (Maures et al. 2009; Rider et al. 2009; Donatello et al. 2007). SH2 is also required for NGF-induced neurite outgrowth and maintenance of the differentiated state (Maures et al. 2009). The other potential interacting partner, human FEM was first identified as a Fas- and tumor necrosis factor receptor (TNFR)- interacting protein that mediates apoptosis when over- expressed in MCF7 cells (Chan et al. 2000). It has been shown to interact with Cul2/Rbx1, and serve as a substrate recognition subunit in an E3 complex (Kamura et al., 2004). FEM is also an adaptor in replication stress-induced signaling that leads to the activation of Human checkpoint kinase 1 (Sun & Shieh 2009). Both SH2 and FEM express in brain and K possibly involved in their pathways, it is therefore interesting to confirm their interaction by means other than yeast two hybrid analysis.

To investigate if full length K would bind to full length SH2 and FEM, HEL293FT cell lines coexpressed Kgfp and either SH2myc or FEMmyc were immunoprecipitated with rabbit anti-myc antibody SC789. The successful pull down of Kgfp in the presence of SH2myc or FEMmyc but not in control vector transfected cells showed that SH2myc and FEMmyc does interact with Kgfp (Figure 6-4). Our data hence confirmed the interaction of full length K and the two adaptor proteins, SH2 and FEM. Although further co-localization studies and Co-IP with deletion mutants will be required to confirm their sites of interaction, given that SH2 is a cytoplasmic protein that anchored to plasma membrane (Maures et al. 2009) and FEM is a cytoplasmic protein that also has nuclear localization (Sun & Shieh 2009), it is very likely that these two proteins interact with the cytoplasmic region of K inside the cytoplasm.

The cytoplasmic region of K interacts with multiple adaptor proteins including AP2M1, SH2 and FEM, which should lead to different cellular functions. This finding provides us an insight that cytoplasmic region of K, like most of the cytoplasmic regions of plasma membrane proteins, is responsible for interactions of downstream pathways.

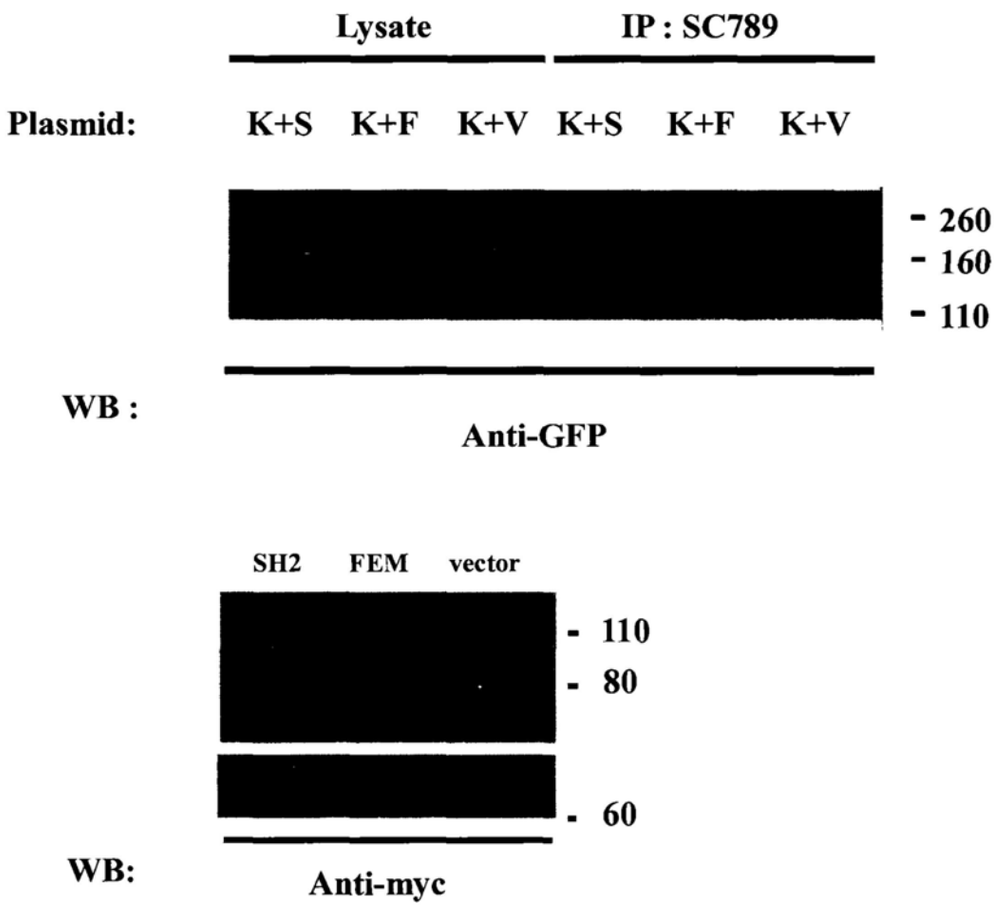


Figure 6-4. Co-IP of Kgfp with SH2 and FEM. HEK293FT cell lines were co-transfected with Kgfp (K), SH2myc (S), FEMmyc (F) and pCMV-myc vector (V) as indicated. Co-IP was performed with rabbit anti-myc polyclonal antibodies (SC789) and detected by mouse anti-GFP antibody (GFP). The Co-IP of Kgfp in cells cotransfected with SH2 or FEM but not in control vector-transfected cells, indicated Kgfp interacts with SH2 and FEM . Reprobe with rabbit anti-myc polyclonal antibodies (SC789) shows that both SH2 and FEM are expressed in the cell lysate.

6.6 Conclusion

K as a membrane protein, its cellular trafficking may be important for its function in DD. It has been demonstrated by Y2H and GST pull down that K interacts with AP2M1 and followed the clathrin mediated endocytosis pathway. To examine whether full length K interact with full length AP2M1, Co-IP was performed. AP2M1 binds both untagged and N-terminal tagged K proteins, which verify that K and AP2M1 interact *in vivo*. This interaction is clearly important to K's function and it is therefore interesting to know whether this interaction determine the importance of K in DD. Given the high similarity between K and KL and that K is more closely associated with DD, the possibility of interaction between KL and AP2M1 was investigated. Evidently, KL, the interacting partner of K, also binds AP2M1. Sequence alignment demonstrated that both K and KL binds to AP2M1 at their C-terminus. Our docking results also demonstrated that K and KL binds with the clathrin-associated heterotetrameric adaptor protein complex AP2 in a similar manner. Therefore, the interaction between K and AP2M1, although could still be important to DD *via* K trafficking, is not unique to K. We also confirmed by Co-IP the interaction of full length K to two adaptor proteins, SH2 and FEM, which most probably interact with the cytoplasmic region of K. Accordingly, our results suggest that the cytoplasmic region of K is most likely involved in the cellular downstream pathways.

CHAPTER 7

Discussion

Developmental dyslexia (DD) is a specific language-based disorder that has high prevalence rates (Dennis et al. 2009; Paracchini et al. 2008). Genetic factors are demonstrated to be of major aetiological significance by various studies. One of the strong DD candidate gene KIAA0319, which encodes a large transmembrane protein consisting 4 main parts in sequence, namely the N-terminus, the PKD domain cluster, the Cysteine rich region, and the cytoplasmic C-terminus, was found to be highly expressed in fetal brain and involved in neuronal migration (Paracchini et al. 2006), which is a putative cause for development of DD. However, owing to lack of biological and molecular knowledge to K protein, how it takes a role in DD remains unknown.

7.1 Overexpressing K had no observable effect on gene expression in both mRNA and protein level

Although the precise function of K is not yet clear, the deregulation of K was shown to be critical to abnormal development of brain (Paracchini et al. 2006), and strongly correlated DD. One of the straight forward hypothesises is K protein has a function of regulating other gene expressions, of which the alternation can lead to problems in brain development and in reading ability. To test this hypothesis, using human K protein as protein model, 2 complementary experiments using DNA microarray and 2D gel analysis were performed. DNA microarray using Agilent human whole genomic

oligonucleotide microarrays (G4112F) was done to fish out possible genes of which the expression is regulated by K protein. Twenty targets with the highest fold change in gene expression, ranging from 2 to 6 fold of both up- or down-regulation, were subjected to real time PCR analysis for confirmation, but none of them can be confirmed to have significant changes gene expression in mRNA level. On the other hand, protein-expression profiles with and without over-expressing K protein in HEK293 cells were compared by utilizing 2D gel analysis. After alignment and normalization of the protein spots, strong correlation of 0.87 ± 0.03 without any outliner was found between the two protein-expression profiles which consist of more than 500 protein spot pairs, indicating no observable protein expressions were regulated by over-expressing K protein. These observations clearly shown that K protein is likely not to be related to DD by regulating other gene expression, neither in mRNA nor in protein level, though we cannot rule out the possibility that the experiments did not have high enough sensitivity to pick up some less abundant mRNA or proteins that were regulated by K. Instead, K protein may associate with DD *via* other means such as affecting protein subcellular localization, or more common, disruption of disease-related interactions. In fact, abnormalities in protein-protein interactions have implications in a number of neurological disorders, including Creutzfeld-Jacob and Alzheimer's diseases (Nazar Zaki 2008). To enrich the knowledge of the relationship between K protein and DD, interactions of human K protein at different parts were investigated.

7.2 Cytoplasmic C-terminus of K is responsible for interactions of various cellular pathways

Human K protein is a large type I plasma membrane protein which consists of multiple domain and totally 1072 amino acids. Its 94-amino-acid cytoplasmic region at C-terminus is relatively short, yet can be crucial for protein functions. In a recent research Levecque and coworkers had demonstrated human K cytoplasmic region can interact *in vitro* with the μ-subunit of Adaptor protein 2 complex (AP2M1), of which the binding will lead to endocytosis by clathrin mediated endocytosis pathway, by GST pull down assay (Levecque et al. 2009). The *in vivo* interaction between full length K and AP2M1 has also been demonstrated by co-immunoprecipitation in my study. It is interesting to know if this interaction at cytoplasmic region determines the importance of K in DD. To address this question, this interaction is best to be compared with the interaction between AP2M1 and KL, which is a homologue protein to K with very high sequence similarity (59%) but only has marginal association to DD and has no concrete experimental proof to show its significance in DD (Couto et al. 2008). Like K, AP2M1 binding motif YXXϕ, where X and ϕ represent any and hydrophobic residue respectively, is also present in KL's cytoplasmic region, and its interaction to AP2M1 was confirmed by co-immunoprecipitation. Owing to the high sequence similarity at the cytoplasmic region, the binding properties of AP2M1 to K and KL is hypothesized to be the same. In fact, it is strongly supported by the result of docking the binding motifs of K and that of KL to AP2M1. The tyrosine in binding motifs of both protein formed hydrogen bonds to AP2M1 and leucine at the +3 position to the tyrosine has strong hydrophobic interaction with AP2M1. In fact, this binding mode of YXXϕ to AP2M1 is also the same for other

protein such as cytotoxic T-lympocyte protein 4 (CTLA-4), revealed from the complex crystal structure of AP2M1 with its binding motif peptide (PDB ID: 1H6E). This observation suggested that the binding of AP2M1 to K is not unique. On the contrary, it is a common type of binding which should not determine the significance of K in DD, otherwise KL, or even CTLA-4, should also be proved to have strong association to DD.

The binding of AP2 complex to K will lead to endocytosis and recycling of K (Levecque et al. 2009). This finding provides us an insight that cytoplasmic region of K is responsible for interactions of cellular downstream pathways such as K trafficking, instead of determining the importance of K in DD. In fact, most if not all of the cytoplasmic regions of plasma membrane proteins are involved in the cellular downstream pathways. Moreover, cytoplasmic regions of one protein can bind to several interacting proteins which lead to various pathways for different functions of the protein. For example, the cytoplasmic region of neural cell adhesion molecule L1 (L1CAM) can bind to AP2 complex for its recycling(Tyukhtenko et al. 2008), to cytoskeletal adaptor protein ankyrin for regulating L1 mediated adhesion and migration (Gil et al. 2003), to FERM-domain-containing proteins such as ezrin for involving in dynamic axonal growth processes and neuronal migration during development (Herron et al. 2009), as well as to RanBPM for regulating L1-mediated ERK activation (Cheng et al. 2005; Herron et al. 2009). Not surprisingly, K protein also has several interaction partners (Nakayama et al. 2002) which should also interact with its cytoplasmic region. In addition to AP2M1, FEM and SH2 were also confirmed to be interaction partners for K protein, which should lead to different cellular functions. Although the cytoplasmic

region does not directly determine the importance of K in DD, it is possible, probably with some molecular events at other part of protein (such as binding of a ligand in the extracellular part) happened in prior, it interacts with proteins which lead to some yet to be characterized cellular downstream pathways that favour development of DD.

7.3 Dimerization as a regulation of K's functions

Knowing that disruption of one or more disease related protein-protein interaction is the reason why K deregulation would lead to impaired brain development, and therefore discovery of K's interaction partners is desirable. For this reason, interactions between full-length human K protein and its homologue KL were investigated. As demonstrated by co-immunoprecipitation, both untagged K and N-terminal tagged K interacts with full length N-terminal tagged KL, confirming K and KL do interact and the interaction was not affected by N-terminal tag. A partial overlap between their subcellular localization demonstrated by immunocytochemistry provides further support of their interaction. No interaction was shown for truncation mutants of Cysteine-rich C6 region in either K or KL proteins, demonstrated that the Cys rich region is the site of their interaction. The same region was involved in K dimerization while the Cys residues are particularly important and probably involved in disulphide bond formation between two molecules of K (Velayos-baeza et al. 2008). K and KL interacts at this dimerization region of K and the fact that they share as high as 66.7% similarity (55.6% identical) in this region with 9 out of 10 Cys residues conserved implied that KL bind to K in a way similar to K binding to itself. Given that a dimer conformation of K would most probably be required by some of the K functions (Velayos-baeza et al. 2008), the K-KL interaction that mimic

K dimerization might be biologically relevant.

K is undoubtedly essential for neuronal migration (Paracchini et al. 2006; Peschansky et al. 2009) and its expression is highest in fetal brain when the bulk of neuronal migration in the mammalian brain occurs (Ghashghaei et al. 2007). Although neuronal migration is evident in regions of postnatal rodent brains including the cerebellum, hippocampus and rostral migratory stream, in contrast to non-human models, evidence for organized, long-distance migration of newly generated neurons in the adult human brain is lacking (Ghashghaei et al. 2007). Indeed, it has been published that neurons in the human cerebral neocortex are not generated in adulthood at detectable levels but are generated perinatally (Bhardwaj et al. 2006) and it is well accepted that the number of new neurons generated, the transitory amplifying progenitor cell population, as well as the needs of neurons migration decrease with increasing age (Ghashghaei et al. 2007).The expression of K also decreases in the postnatal brain (Su et al. 2004), which is consistent to its role in neuronal migration. In contrast, KL express slightly higher in adult brain when neuron migration is less essential (Su et al. 2004). The high expression of K in fetal brain when KL expression is not particularly high suggested that in fetal brain the K-K homodimer would be the predominant form as opposed to the K-KL heterodimer. While in adult brain when the expression of K is lower and KL expression increased, the K-KL heterodimer would be the dominant form. Because of the importance of K in neuronal migration during fetal brain development, the K homodimer is probably the functional form. The formation of K-KL heterodimer in adult brain when the needs of neurons migration decreased led to the hypothesis that in addition to regulation at K's protein

expression level, the function of K is regulated by KL, which serves as a molecular control of neuronal migration by regulating the formation of K dimer. This hypothesis is consistent with the fact that these two proteins have different roles in neuronal migration and could provide an explanation why high endogenous KL expression does not compensate for K's loss of function. To further investigate the role of K dimer in neuronal migration, constructs of K that have Cys residues mutated and therefore remains monomer can be introduced to rat brain and be examined if the mutant would rescue the impaired neuronal migration caused by down regulation of K in the same way that wild-type K would. It is also interesting to find out in addition to the decrease in K expression, whether dyslexics have increased level of KL expression as a compensatory reaction.

7.4 Interactions at PKD region

The central region of K harbours multiple PKD domains which may be involved in cellular adhesion, as PKD domains in polycistin–1 are necessary for adhesion of renal epithelial cells (Bycroft et al. 1999). Because cell adhesion may play a role in cell migration (Paracchini et al. 2006; Paracchini et al. 2007), this region of the protein may be important for K's function. Indeed, it has been speculated that the PKD domains of KIAA0319 could mediate the appropriate adhesion between neurons and the glial fibres during neuronal migration, possibly by interaction with molecules such as astrotactin and the neuregulin-Erb4 system, which have been shown previously to participate in the adhesion of migrating neurons to radial glia (Paracchini et al. 2006). Nevertheless, KL also posses this PKD region and this is the region where K and KL proteins share the

most similarity, with as high as 74.8% similarity and 62.4% amino acid identity. Given that KL has a different role compared to that of K possibly by serving as its regulator, interactions at PKD cluster, although could be very important to K's function, should not be unique to K or determine its significance in DD, unless prior molecular events happened at region where K and KL differ and result in differential downstream pathway that favour DD development.

7.5 N-terminus is critical to K's function

7.5.1 K and KL share least similarity at their N-terminus

K and KL is highly similar in majority of the protein and yet function differently, as a result, the region where K and KL differ would provide important information on its function and DD predisposition. While the PKD regions, the Cysteine rich region and the cytoplasmic region of K and KL share 75%, 66.7% and 54.7% similarity respectively, the N-terminus region of the proteins show the least similarity with only 30.9 % amino acid are similar (17.7% identity), which should be the region that is significant for K's function. Domain swapping experiments such as replacing the N-terminus of KL by the N-terminus of K followed by function assay on neuronal migration would be interesting, which would give us further information on the importance of this region on K's function. For the reason that deregulation of K leads to impaired neuronal migration due to the disruption in protein-protein interaction, the N-terminus of K may function by interacting with other proteins outside the cells. Identifying the interaction partners that binds to N-terminus of K would therefore provide a clue of how K leads to impaired migration and DD.

7.5.2 K may function as membrane receptor that mediate signalling *via* various adaptor proteins

Two proteins, SH2B adaptor protein 1 (SH2) and human homologue of the Caenorhabditis elegans sex determination fem1 protein (FEM), were identified by a large scale Y2H screening as potential interacting partners of K (Su et al. 2002). Our data with Co-IP confirmed that full length K interacts with these two proteins, most possibly at the cytoplasmic region of K. SH2 is a membrane protein expresses in brain and cerebellum, which mediates activation of various ligand-activated receptor tyrosine kinases and function in cytokine and growth factor receptor signalling (Maures et al. 2009). The other interacting partner FEM expresses in cerebellum and cerebral cortex. It was shown to be a caspase substrate that mediates apoptosis (Chan et al. 2000) and is involved in replication stress-induced signalling that leads to the activation of Human checkpoint kinase 1 (Sun & Shieh 2009). These two binding partners are both adapter proteins that lead to activation of kinases and mediate signalling through many receptors (Maures et al. 2009; Sun & Shieh 2009) and thus the binding of K to these partners led to hypothesis that K might be a membrane receptor that mediate signalling *via* various adapter proteins. The extracellular N-terminus of K, which shares the least similarity to KL, hence confers the specificity of the receptor and therefore explains why K is important to neuronal migration and KL the regulator. K harbours a MANSC domain at the N-terminus with only 36.6% similarity and 22% identity to that of KL, suggesting MANSC domain could be important for K's interaction to some other specific factors. MANSC domain containing proteins are mostly of unknown function and the precise

function of MANSC domain remains unclear. Yet, deletion experiments shows that deletion protein of hepatocyte growth factor activator inhibitor-1 (HAT-1) contains only MANSC domain could binds and inhibits hepatocyte growth factor activator (HGFA), suggested that MANSC play a part in complex formation and is capable of protein binding (Guo et al. 2004). Accordingly, to understand how K function in neuronal migration and leads to dyslexia, it is essential to study which proteins interact with the N-terminus of K, particularly at the MANSC domain.

7.6 Conclusion

In this study, we try to investigate how K links to neuronal migration and thus dyslexia. We first tested whether K is linked to neuronal migration by regulating gene expression or by interacting with other proteins. Microarray analysis and 2D gel electrophoresis of K overexpression shows that no observable protein expressions were regulated by over-expressing K protein. Therefore it might be the changed subcellular localization or more common, the disruption of protein-protein interactions that account for the migration deficit. To gain further knowledge of the relationship between K protein and DD by investigating interaction of K to other proteins, a monoclonal antibody (Mab) that is specific to K and capable for various immunoassays was established by using a synthetic peptide. Afterwards, by Co-IP and docking, we demonstrated that K and KL can bind to AP2M1 in the same manner and that the C-terminus of K is responsible for interactions of cellular downstream pathways. Studies of the interaction of K with its homologue KL by Co-IP and deletion mutants shows that they form heterodimers due to structural similarity, which leads to our hypothesis that KL act as a regulator of K. Consequently, the region that is significant in K's important association with DD was

hypothesized to be the N-terminal of K, which has least similarity with KL. We have also confirmed by Co-IP the interaction of K to two receptor adaptor proteins, FEM and SH2, which suggested that K might be membrane receptor and the N-terminus especially MANSC domain confers its specificity to its ligand. It is therefore essential to identify protein-protein interactions that happen at this part of K protein. This basic characterisation provides the information needed to advance in the functional analysis of the K protein, which would allow us to gain new insights on the biological mechanisms underlying the reading process.

7.7 Future aspects

The difference in protein sequence between K and KL at the N-terminus is the most interesting region that might give a clue to explain why K is important to dyslexia. Therefore, structural characterization of the N-terminus of K can be informative. Because the majority part of K and KL proteins are largely similar while their MANSC domains were least similar, crystallization of the MANSC domains of K and KL to find out how they differ structurally would provide important clues of why K promote neuronal migrations while KL act as regulator. For this reason, the N-terminus of K was cloned into *E.coli* expression vectors with or without tags, which include GST, MBP, His-SUMO and His tag. The proteins were expressed but found to be insoluble. Therefore optimization with different expression conditions is needed before we can attempt protein crystallization studies.

In addition to structural characterization, to understand how K deregulation would lead

to impaired migration by disruption of protein interaction, it would be interesting to conduct further investigations using the K Mab developed by us and to find proteins that interact with the N-terminus of K. If the Cysteine rich N-terminus of K can be expressed in mammalian cells, Immunoprecipitation using N-terminus of K is a possible way to screen for interacting partners. The identity of the possible interacting partners can then be examined with 2D gel and mass spectrometry. After finding the interacting partners of K that bind to its N-terminus, co-crystallization with the interaction partner can be performed. Alternatively, their interaction sites can be studied by docking and see if there's any difference between the interaction with K and KL.

In this study, the very specific anti-K Mab G4 was characterized, which is capable for a variety of immunoassays that may provide an ideal reagent for further investigation of the function of K proteins. At present, all studies on K's expression pattern in brain were conducted by in situ hybridisation using oligonucleotide probes, which only measure the expression of K at the mRNA level and any post transcriptional regulation would be ignored. Immunohistochemistry using the specific Mab could lead to a more accurate study on K protein's expression pattern in human brain. Besides, based on the high sensitivity and specificity of the Mab in ELISA screening, it may be valuable in developing a screening method for the level of K expression in dyslexic samples that is more objective, easier to carry out, more affordable to the general public and thus more convenient for massive screening.

There could be an enormous amount of work carried out to further studies on K 's

involvement in neuronal migration and DD association. The N-terminus is probably the most critical region for the functional role of K in DD that deserved further investigations. Some of the potential differences between K and KL were shown in this study, which would only be the beginning of a mechanistic understanding of DD. Further study will be required to describe in detail the downstream pathway following disturbance of the DD genes. Such knowledge may benefit the diagnosis of dyslexia in the future, which will enable dyslexic children to be diagnosed much earlier, accordingly treatment and training can be applied at critical periods and dyslexic children could be given appropriate support to improve their reading ability effectively.

References

Ali, H.I. et al., 2010. A comparative study of AutoDock and PMF scoring performances, and SAR of 2-substituted pyrazolotriazolopyrimidines and 4-substituted pyrazolopyrimidines as potent xanthine oxidase inhibitors. *J Comput Aided Mol Des*, 24, 57-75.

Basalp, A., Çirkoglu, B. & Bermek, E., 2000. Simple Production and Purification of Monoclonal Antibodies in Serum-Free Medium. *Turk J Biol*, 24, 189–196.

Bhardwaj, R.D. et al., 2006. Neocortical neurogenesis in humans is restricted to development. *PNAS*, 103(33), 12564-8.

Black, M.A. et al., 2009. A statistical model to identify differentially expressed proteins in 2D PAGE gels. *PLoS comput Biol*, 5(9), e1000509.

Brückner, A. et al., 2009. Yeast two-hybrid, a powerful tool for systems biology. *Int. J. Mol. Sci.*, 10(6), 2763-88.

Bunai, K. & Yamane, K., 2005. Effectiveness and limitation of two-dimensional gel electrophoresis in bacterial membrane protein proteomics and perspectives. *J. Chromatogr. B*, 815(1-2), 227-36.

Bycroft, M. et al., 1999. The structure of a PKD domain from polycystin-1: implications for polycystic kidney disease. *EMBO J.*, 18(2), 297-305.

Chan CW, 2008. Overview of Specific Learning Disabilities (SLD)/Dyslexia Developments Over the Last Decade in Hong Kong. *HK J Paediatr*, 13, 196-202.

Chan, D.W. et al., 2007. Prevalence, Gender Ratio and Gender Differences in Reading-Related Cognitive Abilities among Chinese Children with Dyslexia in Hong Kong. *Educational Studies*, 33(2), 249-265.

Chan, S.L. et al., 2000. The Caenorhabditis elegans sex determination protein FEM-1 is a CED-3 substrate that associates with CED-4 and mediates apoptosis in mammalian cells. *J. Biol. Chem.*, 275(24), 17925-8.

Cheng, L., Lemmon, S. & Lemmon, V., 2005. RanBPM is an L1-interacting protein that regulates L1-mediated mitogen-activated protein kinase activation. *J. Neurochem*, 94(4), 1102-10.

Cope, N. et al., 2005. Strong evidence that KIAA0319 on chromosome 6p is a susceptibility gene for developmental dyslexia. *Am. J. Hum. Genet.*, 76, 581-591.

Couto, J.M. et al., 2009. Association of attention-deficit/hyperactivity disorder with a candidate region for reading disabilities on chromosome 6p. *Biol psychiat*, 66(4), 368-75.

Couto, J.M. et al., 2008. The KIAA0319-like (KIAA0319L) gene on chromosome 1p34 as a candidate for reading disabilities. *J. Neurogenet*, 22(4), 295-313.

Dennis, M.Y. et al., 2009. A common variant associated with dyslexia reduces expression of the KIAA0319 gene. *PLoS Genet*, 5(3), e1000436.

Donatello, S. et al., 2007. SH2B1beta adaptor is a key enhancer of RET tyrosine kinase signaling. *Oncogene*, 26(45), 6546-59.

Dyballa, N. & Metzger, S., 2009. Fast and sensitive colloidal coomassie G-250 staining for proteins in polyacrylamide gels. *J Vis Exp*, (30), 2-5.

Fisher, S.E. & Francks, C., 2006. Genes, cognition and dyslexia: learning to read the genome. *Trends cogn sci*, 10(6), 250-7.

Francks, C. et al., 2004. A 77-kilobase region of chromosome 6p22.2 is associated with dyslexia in families from the United Kingdom and from the United States. *Am. J. Hum. Genet.*, 75, 1046-1058.

Gabel, L.a. et al., 2009. Progress towards a cellular neurobiology of reading disability. *Neurobiol. Dis*. [Epub ahead of print]

Ghashghaei, H.T., Lai, C. & Anton, E.S., 2007. Neuronal migration in the adult brain: are we there yet? *Nat. Rev. Neurosci*, 8(2), 141-51.

Gibson, C. & Gruen, J., 2008. The human lexinome: Genes of language and reading. *J Commun Disord.*, 41(5), 409-420.

Gil, O.D. et al., 2003. Ankyrin binding mediates L1CAM interactions with static components of the cytoskeleton and inhibits retrograde movement of L1CAM on the cell surface. *J. Cell Biol.*, 162(4), 719-30.

Grabowska, D. et al., 2008. Postnatal induction and localization of R7BP, a membrane-anchoring protein for RGS7 family-Gβ5 complexes in brain. *Neuroscience*, 151(4), 969-982.

Guo, J. et al., 2004. MANSC: a seven-cysteine-containing domain present in animal membrane and extracellular proteins. *Trends Biochem. Sci*, 29(4), 172-174.

Habib Michel, 2000. The neurological basis of developmental dyslexia: An overview and working hypothesis. *Brain*, 123(12), 2373-2399.

Hall, T., 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acid S*, 41, 95-98.

Haltiwanger, R.S. & Lowe, J.B., 2004. Role of glycosylation in development. *Annu. Rev. Biochem*, 73, 491-537.

Hanahan, D., 1983. Studies on transformation of Escherichia coli with plasmids. *J Mol Biol*, 166(4), 557-580.

Hannula-jouppi, K. et al., 2005. The axon guidance receptor gene ROBO1 is a candidate gene for developmental dyslexia. *PLoS Genet*, 1(4), e50.

Harold, D. et al., 2006. Further evidence that the KIAA0319 gene confers susceptibility to developmental dyslexia. *Mol. Psychiatry*, 11(12), 1085-91, 1061.

Herron, L.R. et al., 2009. The intracellular interactions of the L1 family of cell adhesion molecules. *Biochem. J.*, 419(3), 519-31.

Holzmuller, P. et al., 2008. Virulence and pathogenicity patterns of Trypanosoma brucei gambiense field isolates in experimentally infected mouse: differences in host immune response modulation by secretome and proteomics. *Microbes Infect.*, 10(1), 79-86.

Hornshøj, H. et al., 2009. Transcriptomic and proteomic profiling of two porcine tissues using high-throughput technologies. *BMC genomics*, 10, 30.

Imberty, A., Nurisso, a. & Kozmon, S., 2008. Comparison of docking methods for carbohydrate binding in calcium-dependent lectins and prediction of the carbohydrate binding mode to sea cucumber lectin CEL-III. *Mol. Simul.*, 34(4), 469-479.

Iyer, K. et al., 2005. Utilizing the Split-Ubiquitin Membrane Yeast Two-Hybrid System to Identify Protein-Protein Interactions of Integral Membrane Proteins. *Sci. STKE*, 2005(275), 13.

Jacob, E. & Unger, R., 2007. A tale of two tails: why are terminal residues of proteins exposed? *Bioinformatics*, 23(2), e225-e230.

Kalliokoski, T. et al., 2009. The effect of ligand-based tautomer and protomer prediction on structure-based virtual screening. *J. Chem. Inf. Model.*, 49(12), 2742-8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19928753.

Keen, M.J. & Steward, T.W., 1995. Adaptation of cholesterol-requiring NS0 mouse myeloma cells to high density growth in a fully defined protein-free and cholesterol-free culture medium. *Cytotechnology*, 17(3), 203-211.

Kovel, C.G. et al., 2004. Genomewide scan identifies susceptibility locus for dyslexia on Xq27 in an extended Dutch family. *J. Med. Genet*, 41(9), 652-657.

Köhler, G. & Milstein, C., 1975. Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature*, 256, 495-497.

Lemkin, P.F., 1997. Comparing two-dimensional electrophoretic gel images across the Internet. *Electrophoresis*, 18(3-4), 461-70.

Levecque, C. et al., 2009. The dyslexia-associated protein KIAA0319 interacts with adaptor protein 2 and follows the classical clathrin-mediated endocytosis pathway. *Am J Physiol Cell Physiol.*, 297(1), C160-8.

Lifely, M. et al., 1995. Glycosylation and biological activity of CAMPATH-1H expressed in different cell lines and grown under different culture conditions. *Glycobiology*, 5(8), 813-822.

Lindwasser, O.W. et al., 2008. A diacidic motif in human immunodeficiency virus type 1 Nef is a novel determinant of binding to AP-2. *J. Virol*, 82(3), 1166-74.

Liu, G. et al., 2004. Netrin requires focal adhesion kinase and Src family kinases for axon outgrowth and attraction. *Nat. Neurosci.*, 7(11), 1222.

Luciano, M. et al., 2007. A haplotype spanning KIAA0319 and TTRAP is associated with normal variation in reading and spelling ability. *Biol psychiat*, 62(7), 811-7.

Ludwig, K.U. et al., 2008. Investigation of interaction between DCDC2 and KIAA0319 in a large German dyslexia sample. *J. Neural Transm*, 115(11), 1587-9.

Ludwig, K.U. et al., 2009. Variation in GRIN2B contributes to weak performance in verbal short-term memory in children with dyslexia. *Am J Med Genet Part B.*

Lyon, G.R., Shaywitz, S.E. & Shaywitz, B.A., 2003. A Definition of Dyslexia. *Ann. Dyslexia*, 53(1).

Maures, T.J., Chen, L. & Carter-Su, C., 2009. Nucleocytoplasmic shuttling of the adapter protein SH2B1beta (SH2-Bbeta) is required for nerve growth factor (NGF)-dependent neurite outgrowth and enhancement of expression of a subset of NGF-responsive genes. *Mol Endocrinol*, 23(7), 1077-91.

Meng, H. et al., 2005. DCDC2 is associated with reading disability and modulates neuronal development in the brain. *PNAS*, 102(47), 17053-8.

Morris, G.M. et al., 1998. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem*, 19(14), 1639-1662.

Nakatsu, F. & Ohno, H., 2003. Adaptor protein complexes as the key regulators of protein sorting in the post-Golgi network. *Cell Struct Funct*, 28(5), 419-29.

Nakayama, M., Kikuno, R. & Ohara, O., 2002. Protein–Protein Interactions Between Large Proteins: Two-Hybrid Screening Using a Functionally Classified Library Composed of Long cDNAs. *Genome Research*, 12(11), 1773-1784.

Nazar Zaki, 2008. Prediction of Protein–Protein Interactions Using Pairwise Alignment and Inter-Domain Linker Region. *Engineering Letter*, 16(4).

Ohno, H. et al., 1995. Interaction of tyrosine-based sorting signals with clathrin-associated proteins. *Science*, 269(5232), 1872-5.

Ohtsubo, K. & Marth, J.D., 2006. Glycosylation in cellular mechanisms of health and disease. *Cell*, 126(5), 855-67.

Owen, D.J., Collins, B.M. & Evans, P.R., 2004. Adaptors for clathrin coats: structure and function. *Annu. Rev. Cell Dev. Biol.*, 20, 153-91.

Oyama, S. et al., 2009. Dysbindin-1, a schizophrenia-related protein, functionally interacts with the DNA- dependent protein kinase complex in an isoform-dependent manner. *PloS one*, 4(1), e4199.

Pagliacci, M.C. et al., 1993. Genistein inhibits tumour cell growth in vitro but enhances mitochondrial reduction of tetrazolium salts: a further pitfall in the use of the MTT assay for evaluating cell growth and survival. *Eur J Cancer*, 29A(11), 1573-7.

Paracchini, S. et al., 2007. Alternative splicing in the dyslexia-associated gene KIAA0319. *Mamm. Genome*, 18, 627-634.

Paracchini, S. et al., 2008. Association of the KIAA0319 dyslexia susceptibility gene with reading skills in the general population. *Am J psychiatry*, 165, 1576-1584.

Paracchini, S. et al., 2006. The chromosome 6p22 haplotype associated with dyslexia reduces the expression of KIAA0319, a novel gene involved in neuronal migration. *Hum. Mol. Genet.*, 15(10), 1659-1666.

Perfetti, C.A. et al., 2008. A structural-functional basis for dyslexia in the cortex of Chinese readers. *PNAS*, 105(14), 5561-6.

Peschansky, V.J. et al., 2009. The Effect of Variation in Expression of the Candidate Dyslexia Susceptibility Gene Homolog Kiaa0319 on Neuronal Migration and Dendritic Morphology in the Rat. *Cereb. Cortex*. [Epub ahead of print]

Ramsland, P.A. et al., 2009. Molecular docking of carbohydrate ligands to antibodies: structural validation against crystal structures. *J. Chem. Inf. Model.*, 49(12), 2749-60.

Rider, L. et al., 2009. Adapter protein SH2B1beta cross-links actin filaments and regulates actin cytoskeleton. *Mol Endocrinol*, 23(7), 1065-76.

Robinson, M.S., 2004. Adaptable adaptors for coated vesicles. *Trends cell biol*, 14(4), 167-74.

Roeske, D. et al., 2009. First genome-wide association scan on neurophysiological endophenotypes points to trans-regulation effects on SLC2A3 in dyslexic children. *Mol. Psychiatry*, 1-11.

Salekdeh, G. et al., 2002. A proteomic approach to analyzing drought-and salt-responsiveness in rice. *Field Crops Res.*, 76(2-3), 199–219.

Sanderson, C.M. et al., 2008. Spastin and atlastin, two proteins mutated in autosomal dominant hereditary spastic paraplegia, are binding partners. *Hum. Mol. Genet.*, 15(2), 307-318.

Shaw, G. et al., 2002. Preferential transformation of human neuronal cells by human adenoviruses and the origin of HEK 293 cells. *FASEB J.*, 16(8), 869-71.

Shaywitz, S.E., 1998. Dyslexia. *New Engl J Med*, 338(5), 307-312.

Shi, G. & Andres, D.A., 2005. Rit Contributes to Nerve Growth Factor-Induced Neuronal Differentiation via Activation of B-Raf–Extracellular Signal-Regulated Kinase and p38 Mitogen-Activated Protein Kinase Cascades. *Mol. Cell. Biol.*, 25(2), 830-846.

Silani, G. et al., 2005. Brain abnormalities underlying altered activation in dyslexia: a voxel based morphometry study. *Brain*, 128(10), 2453-61.

Silberberg, M. et al., 2005. Mispolarization of desmosomal proteins and altered intercellular adhesion in autosomal dominant polycystic kidney disease. *Am J Physiol Renal Physiol*, 288(6), 1153-63.

Su, A.I. et al., 2004. A gene atlas of the mouse and human protein-encoding transcriptomes. *PNAS*, 101(16), 6062-7.

Su, A.I. et al., 2002. Large-scale analysis of the human and mouse transcriptomes. *PNAS*, 99(7), 4465-70.

Sun, T. & Shieh, S., 2009. Human FEM1B is required for Rad9 recruitment and CHK1 activation in response to replication stress. *Oncogene*, 28(18), 1971-81.

Tyukhtenko, S. et al., 2008. Characterization of the neuron-specific L1-CAM cytoplasmic tail: naturally disordered in solution. It exercises different binding modes for different adaptor proteins. *Biochemistry*, 47, 4160-4168.

Velayos-baeza, A. et al., 2008. The dyslexia-associated gene KIAA0319 encodes highly N- and O-glycosylated plasma membrane and secreted isoforms. *Hum. Mol. Genet.*, 17(6), 859-871.

Vetrivel, K. et al., 2008. Localization and regional distribution of p23/TMP21 in the brain. *Neurobiol. Dis*, 32(1), 37–49.

Wang, Y. et al., 2006. DYX1C1 functions in neuronal migration in developing neocortex. *Neuroscience*, 143(2), 515-22.

Wopereis, S. et al., 2006. Mechanisms in protein O-glycan biosynthesis and clinical and molecular aspects of protein O-glycan biosynthesis defects: a review. *Clin. Chem.*, 52, 574-600.

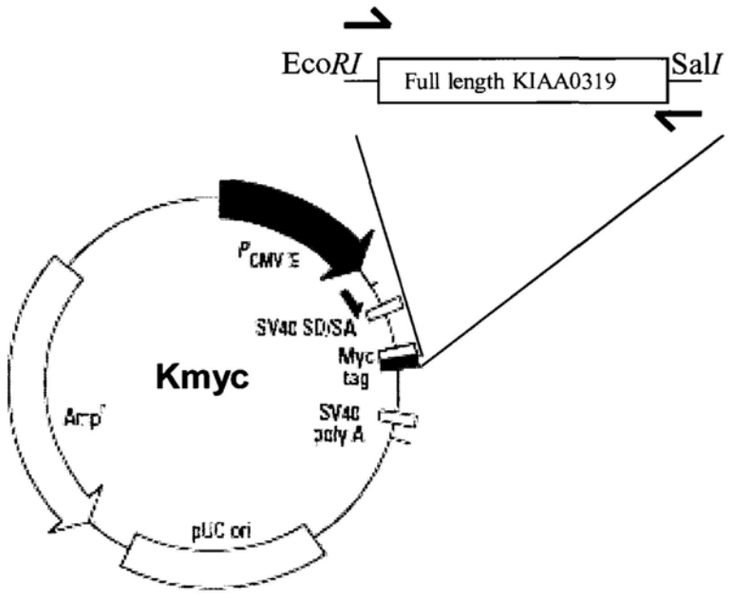
Xiong, H. et al., 2009. ABCG2 is up-regulated in Alzheimer's brain with cerebral amyloid angiopathy and may act as a gatekeeper at the blood-brain barrier for Aβ1-40 peptides. *J Neurosci*, 29(17), 5463-5475.

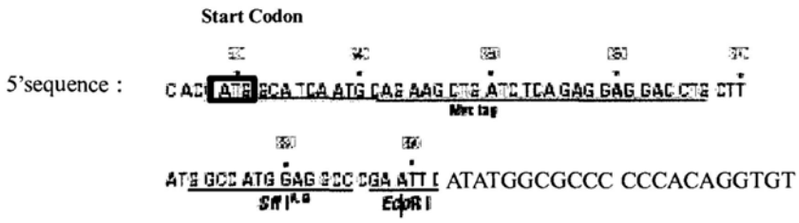
Yap, A.S., Crampton, M.S. & Hardin, J., 2007. Making and breaking contacts: the cellular biology of cadherin regulation. *Curr. Opin. Cell Biol*, 19(5), 508-14.

Yin, G. et al., 2009. Mitochondrial damage in the soybean seed axis during imbibition at chilling temperatures. *Plant Cell Physiol.*, 50(7), 1305-18.

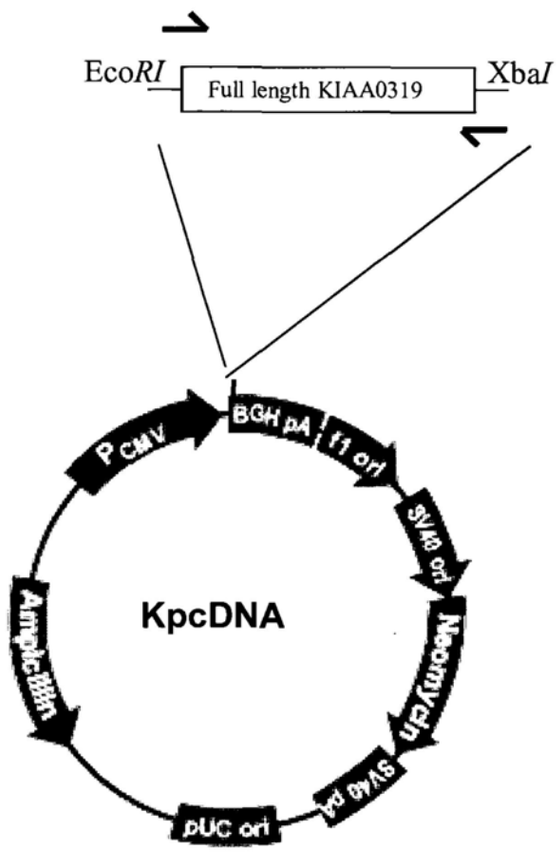
Yuan, X. & Desiderio, D.M., 2005. Proteomics analysis of human cerebrospinal fluid. *J. Chromatogr. B*, 815(1-2), 179-89.

Appendix





Schematic diagram of plasmids constructed
Construct Name: Kmyc
Encoding full length KIAA0319 (1-1072 aa) with N terminal myc tag
Primer sequence : **TAGGGC**GAATTCATATGGCGCCC CCCACAGGTGT
Restriction enzyme : Eco*RI* and Sal*I*
Construct used in Figures 3-1, 4-2, 4-3, 4-4, 4-5, 4-6, 4-8

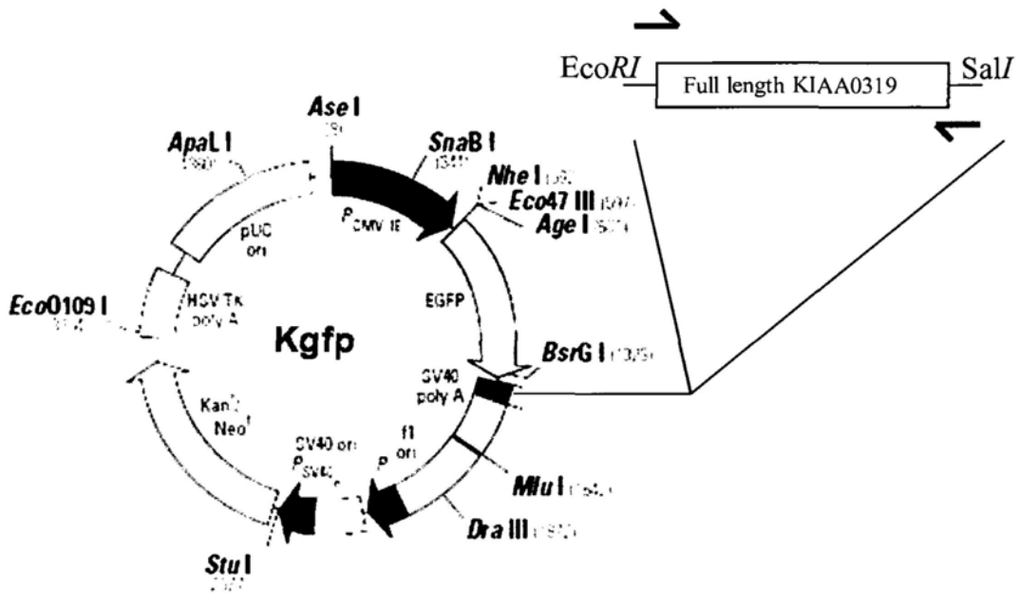


Start Codon

5'sequence : GAATTCGCCGCCGCCATGGCGCCCCCCACAG

Schematic diagram of plasmids constructed

Construct Name: KpcDNA
Encoding full length KIAA0319 (1-1072 aa) without tag
Primer sequence : **TAGGGC**GAATTCGCCGCCGCCATGGCGCCCCCCACAG
Restriction enzyme : Eco*RI* and Xba*I*
Construct used in Figures 3-2, 3-3, 4-2, 4-6, 4-8, 4-9, 4-10, 5-1, 6-1



Start Codon before EGFP

5'sequence : TAC AAG TCC GGA CTC AGA TCT CGA GCT CAA GCT TCG AAT TC AATGGCGCCCCCCACAGGTGT

EGFP
BspE I Bpl I Xho I Sac I Hind III EcoR I
Ecl136 II

Schematic diagram of plasmids constructed

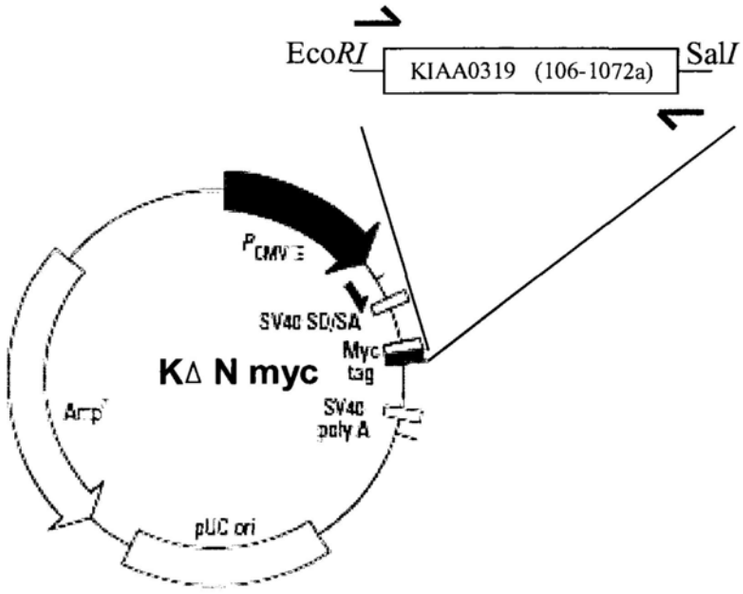
Construct Name: Kgfp

Encoding full length KIAA0319 (1-1072 aa) with N terminal GFP tag

Primer sequence : **TAGGGC**GAATTCAATGGCGCCCCCCACAGGTGT

Restriction enzyme : Eco*RI* and Sal*I*

Construct used in Figures 4-8, 5-1, 5-2, 5-5, 6-1, 6-2



Start Codon

5'sequence : CAC ATG GCA TCA ATG GAG AAG CTG ATC TCA GAG GAG GAC CTG CTT

Myc tag

ATG GCC ATG GAG GCC CGA ATT C TGATGCCTGTTCAGAGGCCTGCA

Sfi I[A,B] *EcoR I*

Schematic diagram of plasmids constructed

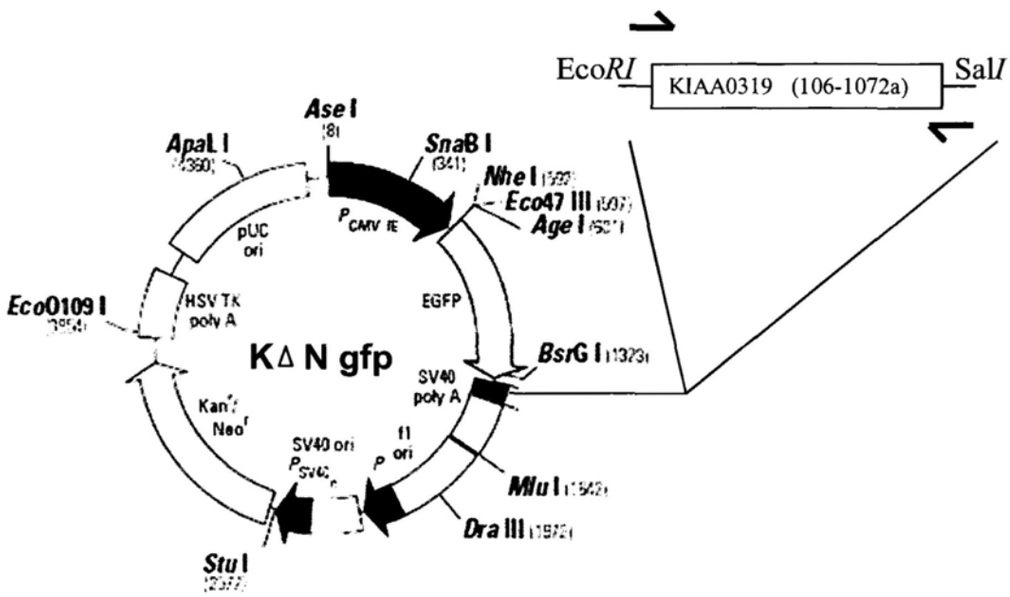
Construct Name: KΔ N myc

Encoding N-deleted KIAA0319 (106-1072 aa) with N terminal myc tag

Primer sequence : **TATGCA**GAATTCTGATGCCTGTTCAGAGGCCTGCA

Restriction enzyme : Eco*RI* and Sal*I*

Construct used in Figures 5-5



Start Codon before EGFP

5'sequence : TAC AAG TCC GGA CTC AGA TCT CGA GCT CAA GCT TCG AAT TC CATGCCTGTTCAGAGGCCTGCA

BspE I Bgl II Xho I Sac I Hind III EcoR I
Ecl136 II

Schematic diagram of plasmids constructed

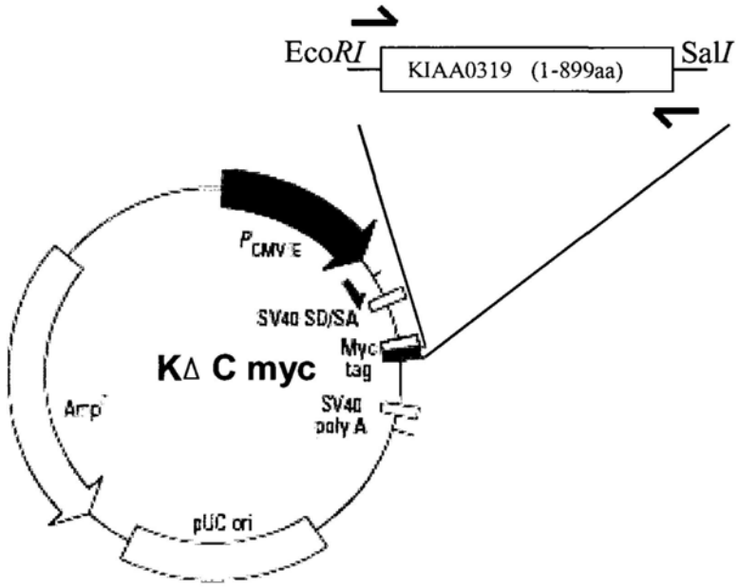
Construct Name: KΔ N gfp

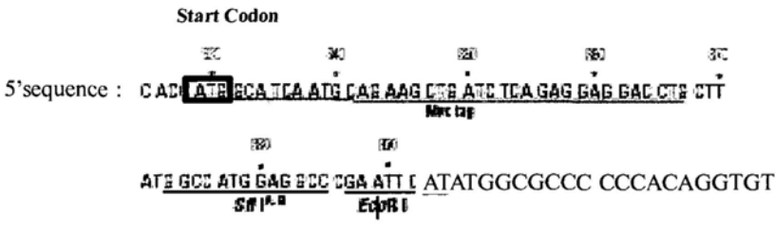
Encoding N-deleted KIAA0319 (106-1072 aa) with N terminal GFP tag

Primer sequence : **TATGCA**GAATTCCATGCCTGTTCAGAGGCCTGCA

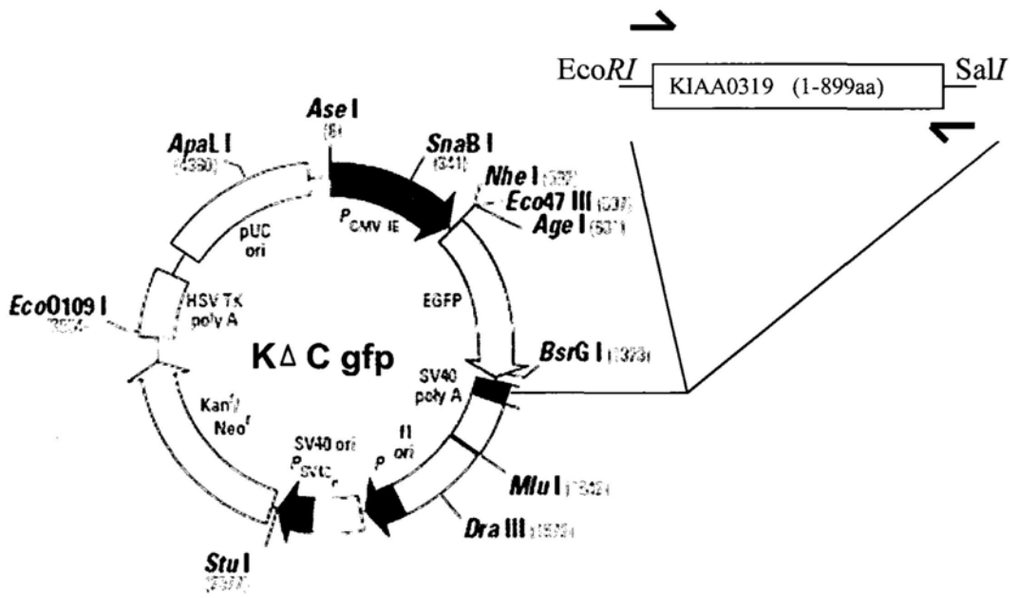
Restriction enzyme : Eco*RI* and Sal*I*

Construct not used in thesis





Schematic diagram of plasmids constructed
Construct Name: KΔ C myc
Encoding C-deleted KIAA0319 (1-899 aa) with N terminal myc tag
Primer sequence : **TAGGGC**GAATTCATATGGCGCCC CCCACAGGTGT
Restriction enzyme : Eco*RI* and Sal*I*
Construct used in Figures 5-5

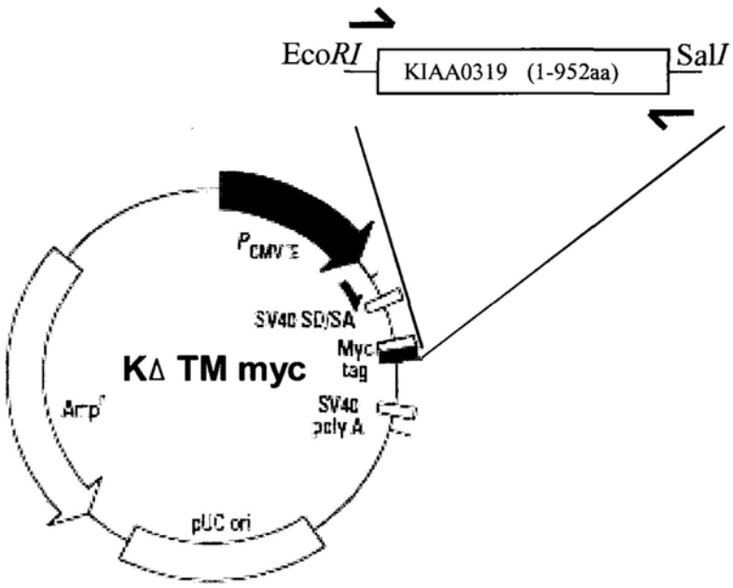


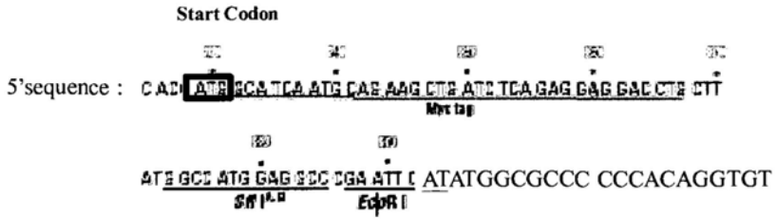
Start Codon before EGFP

5'sequence : TAC AAG TCC GGA CTC AGA TCT CGA GCT CAA GCT TCG AAT TC AATGGCGCCCCCCACAGGTGT

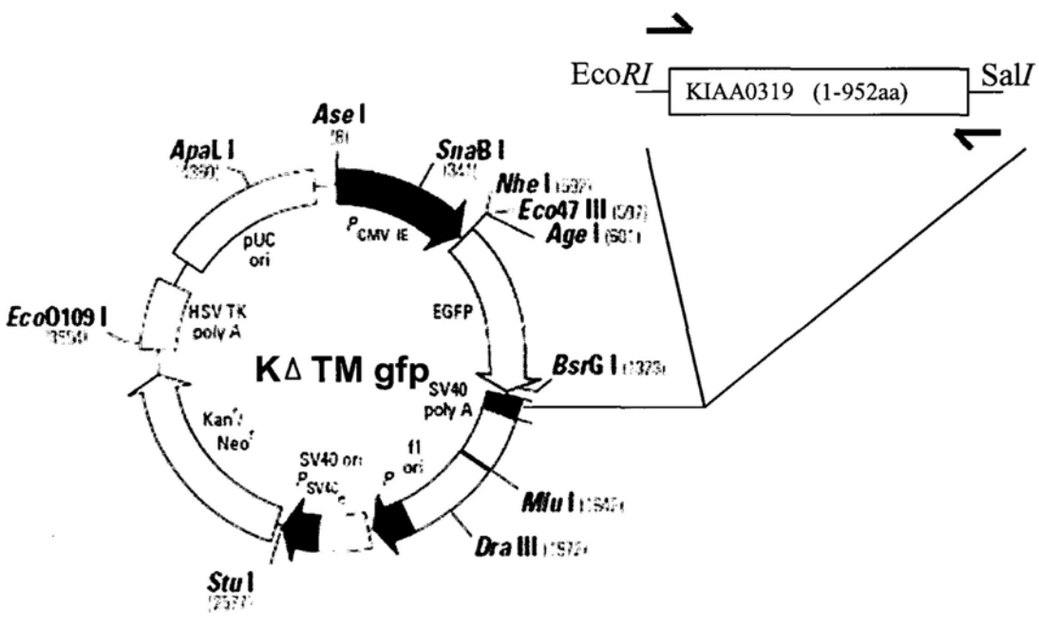
BspE I Bgl II Xho I Sac I Hind III EcoR I
Ecl136 II

Schematic diagram of plasmids constructed
Construct Name: KΔC gfp
Encoding C-deleted KIAA0319 (1-899 aa) with N terminal GFP tag
Primer sequence : **TAGGGC**GAATTCAATGGCGCCCCCCACAGGTGT
Restriction enzyme : Eco*RI* and Sal*I*
Construct not used in thesis





Schematic diagram of plasmids constructed
Construct Name: KΔ TM myc
Encoding TM-deleted KIAA0319 (1-952 aa) with N terminal myc tag
Primer sequence : **TAGGGC**GAATTCATATGGCGCCC CCCACAGGTGT
Restriction enzyme : Eco*RI* and Sal*I*
Construct used in Figures 5-5



Start Codon before EGFP

5'sequence : TAC AAG TCC GGA CTC AGA TCT CGA GCT CAA GCT TCG AAT TC AATGGCGCCCCCCACAGGTGT

BspE I Bgl II Xho I Sac I Hind III EcoR I
Ecl136 II

Schematic diagram of plasmids constructed

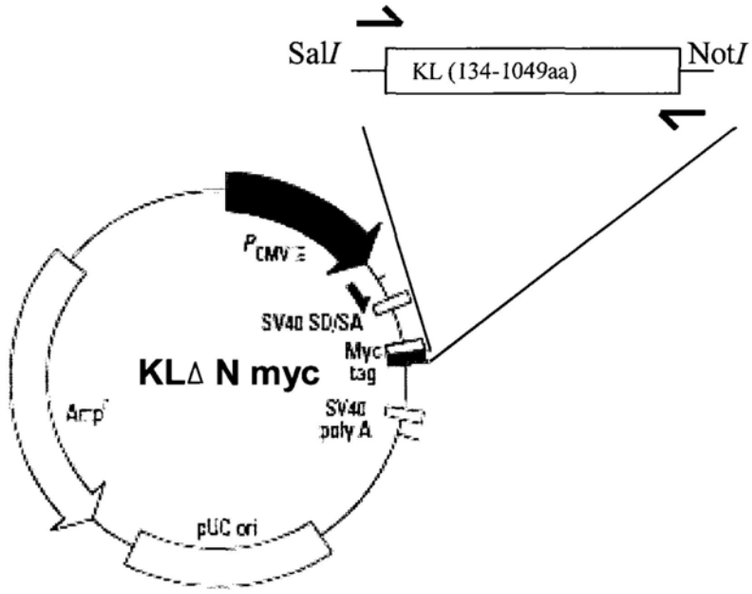
Construct Name: KΔTM gfp

Encoding TM-deleted KIAA0319 (1-952 aa) with N terminal GFP tag

Primer sequence : **TAGGGC**GAATTCAATGGCGCCCCCCACAGGTGT

Restriction enzyme : Eco*RI* and Sal*I*

Construct not used in thesis



Start Codon

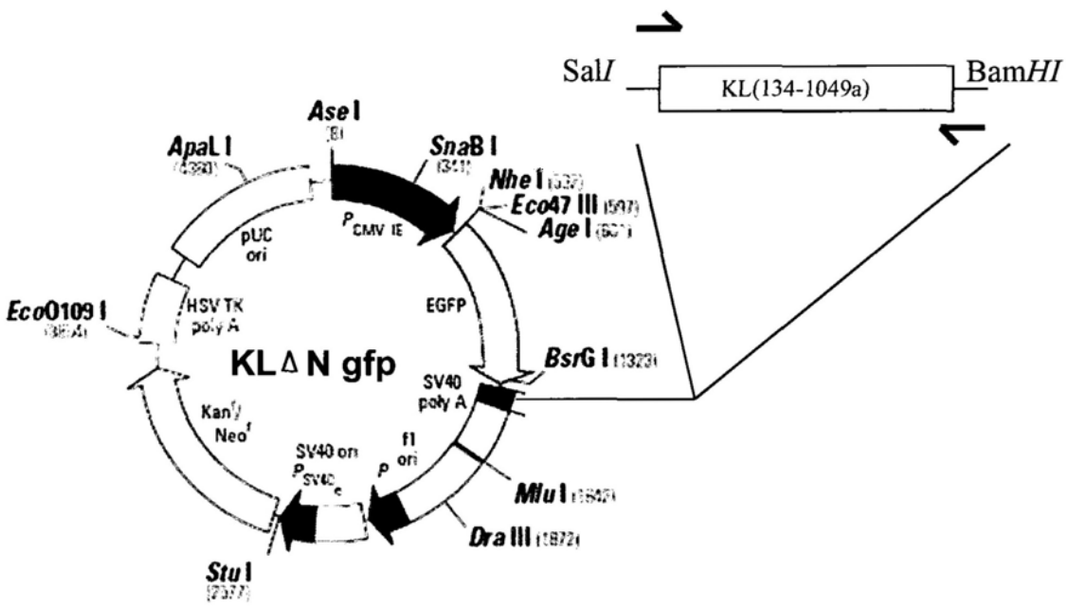
5'sequence : C AC ATG GCA TCA ATG GAG AAG CTG ATC TCA GAG GAG GAC CTG CTT

Myc tag

ATG GCC ATG GAG GCC CGA ATT CGG TCG AC AATGTTCCAAACTGCAGATGAT

Sfi I[A,B] *EcoR I* *Sal I*[B]

Schematic diagram of plasmids constructed
Construct Name: KLΔN myc
Encoding N-deleted KL (134-1049 aa) with N terminal myc tag
Primer sequence : **TATGCA**GTCGACAATGTTCCAAACTGCAGATGAT
Restriction enzyme : Sal*I* and Not*I*
Construct used in Figures 5-5



Start Codon before EGFP

5'sequence : EGFP TAC AAG TCC GGA CTC AGA TCT CGA GCT CAA GCT TCG AAT TCT GCA GTC GAC ATGTTCCAAACTGCAGATGAT

BspE I Bgl II Xho I Sac I Hind III EcoR I Pst I Sal I Acc I
Eco136 II

Schematic diagram of plasmids constructed

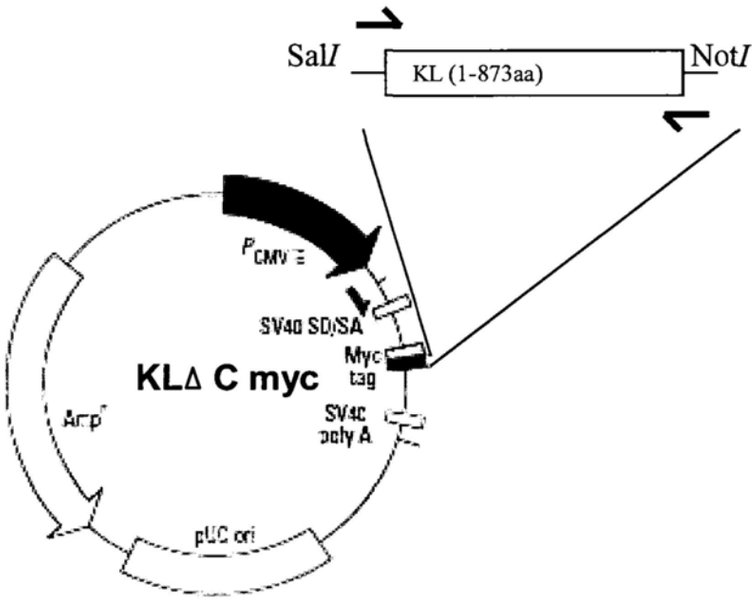
Construct Name: KLΔ N gfp

Encoding N-deleted KL (134-1049 aa) with N terminal GFP tag

Primer sequence : **TATGCA**GTCGACATGTTCCAAACTGCAGATGAT

Restriction enzyme : Sal*I* and Bam*HI*

Construct not used in thesis



Start Codon

5'sequence : C ACC ATG GCA TCA ATG GAG AAG CTG ATC TCA GAG GAG GAC CTG TTT

Myc tag

ATG GCC ATG GAG GCC GAA TTC GGG TCG AC AATGGAGAAGAGGCTGGG

NcoI EcoRI SalI

Schematic diagram of plasmids constructed

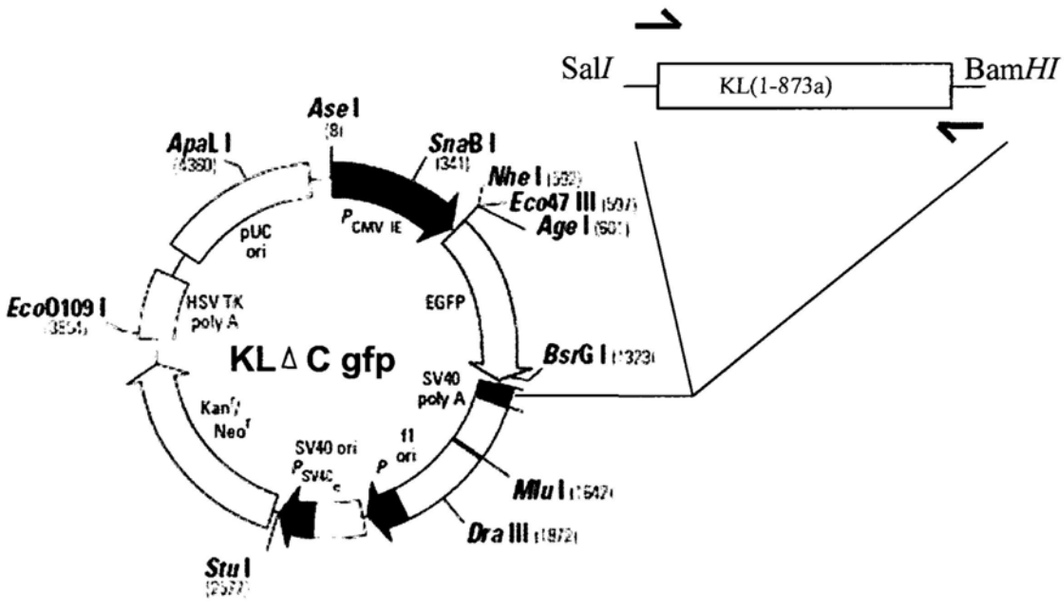
Construct Name: KLΔ C myc

Encoding C-deleted KL (1-873 aa) with N terminal myc tag

Primer sequence : **TATGCA**GTCGACAATGGAGAAGAGGCTGGG

Restriction enzyme : Sal*I* and Not*I*

Construct used in Figures 5-5



Start Codon before EGFP

5'sequence : EGTP TAC AAG TCC GGA CTC AGA TCT CGA GCT CAA GCT TCG AAT TCT GCA GTC GAC ATGGAGAAGAGGCTGGG

BspE I Bgl II XhoI Sac I EcoT136 II Hind III EcoR I Pst I SalI Acc I

Schematic diagram of plasmids constructed

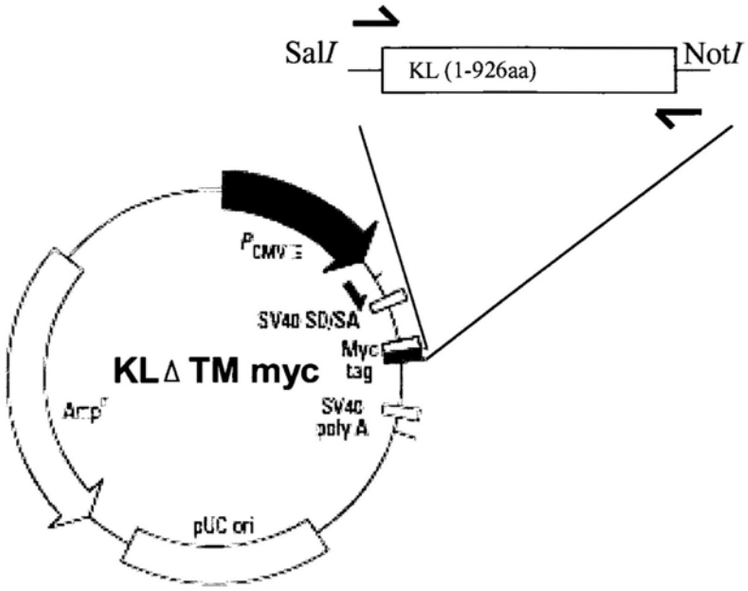
Construct Name: KLΔ C gfp

Encoding C-deleted KL (1-873 aa) with N terminal GFP tag

Primer sequence : **TATGCA**GTCGACATGGAGAAGAGGCTGGG

Restriction enzyme : Sal*I* and Bam*HI*

Construct not used in thesis



Start Codon

5'sequence : CACC ATG GCA TCA ATG GAG AAG CTG ATC TCA GAG GAG GAC CTG CTT

Myc tag

ATG GCC ATG GAG GCC CGA ATT CGG TCG AC AATGGAGAAGAGGCTGGG

Sfi I[A,B] *EcoR* I *Sal* I[B]

Schematic diagram of plasmids constructed

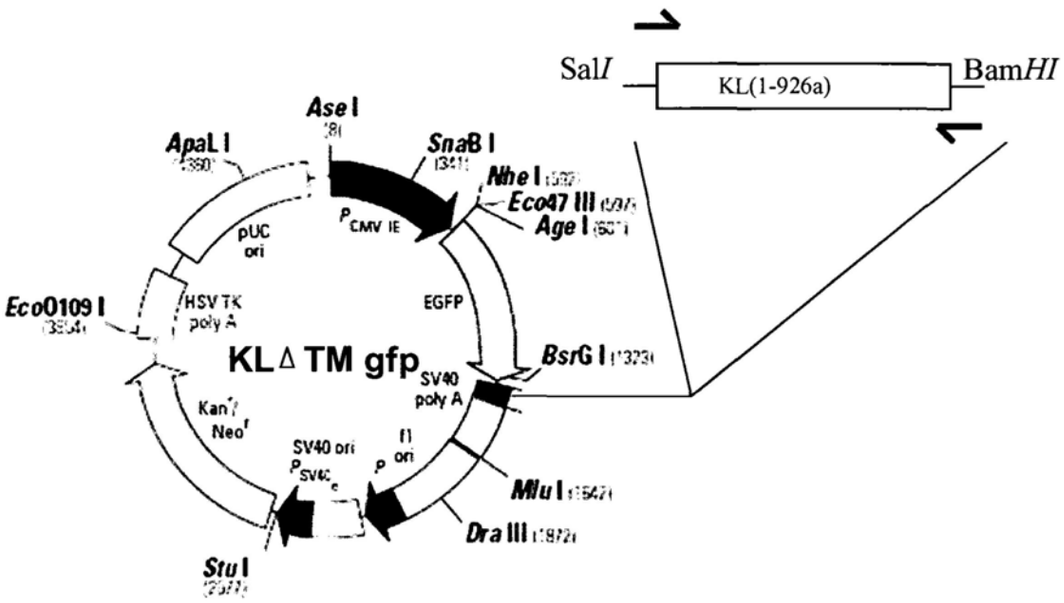
Construct Name: KLΔ TM myc

Encoding TM-deleted KL (1-926 aa) with N terminal myc tag

Primer sequence : **TATGCA**GTCGACAATGGAGAAGAGGCTGGG

Restriction enzyme : Sal*I* and Not*I*

Construct used in Figures 5-5



Start Codon before EGFP

5'sequence : EGFP TAC AAG TCC GGA CTC AGA TCT CGA GCT CAA GCT TCG AAT TCT GCA GTC GAC ATGGAGAAGAGGCTGGG

BspEI BglII XhoI SacI Eco136II HindIII EcoRI PstI SalI AccI

Schematic diagram of plasmids constructed

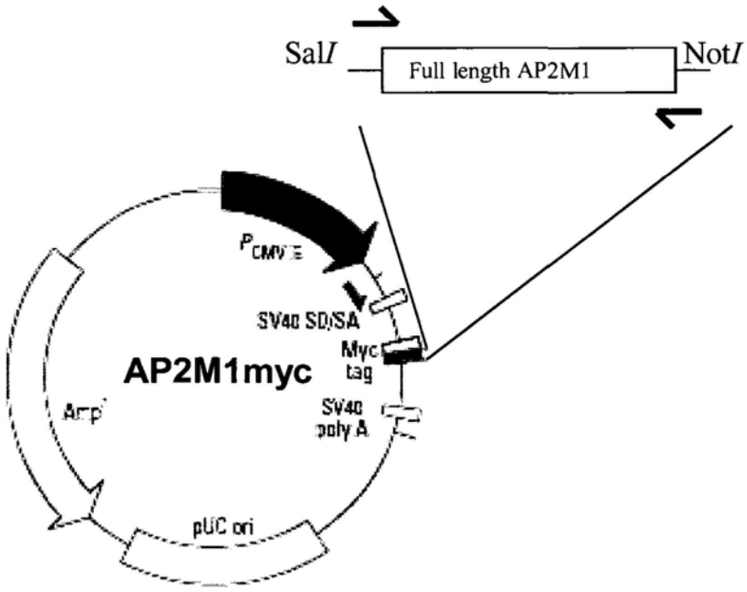
Construct Name: KLΔ TM gfp

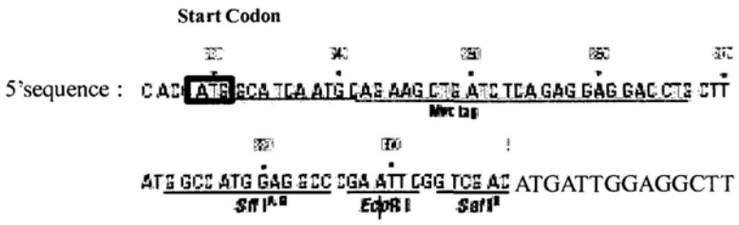
Encoding TM-deleted KL (1-926 aa) with N terminal GFP tag

Primer sequence : **TATGCA**GTCGACATGGAGAAGAGGCTGGG

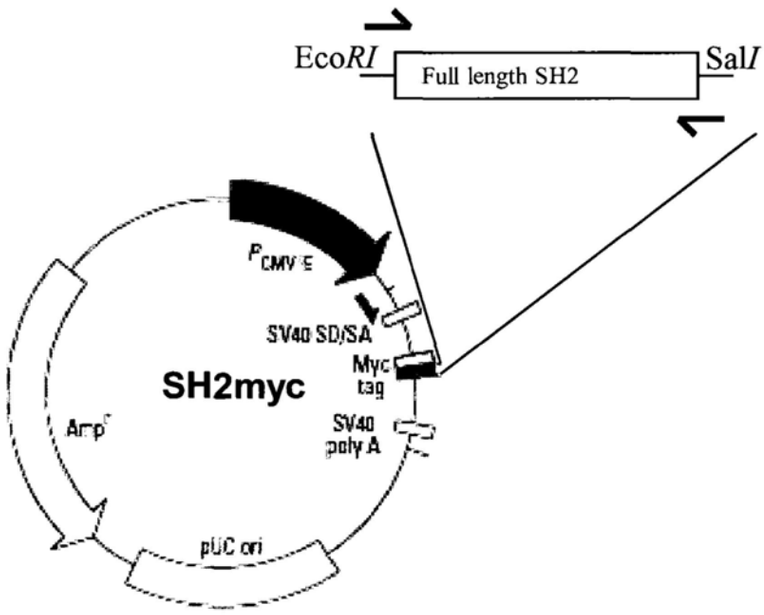
Restriction enzyme : Sal*I* and Bam*HI*

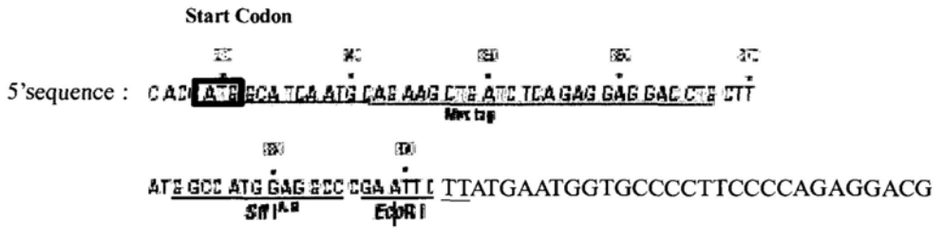
Construct not used in thesis



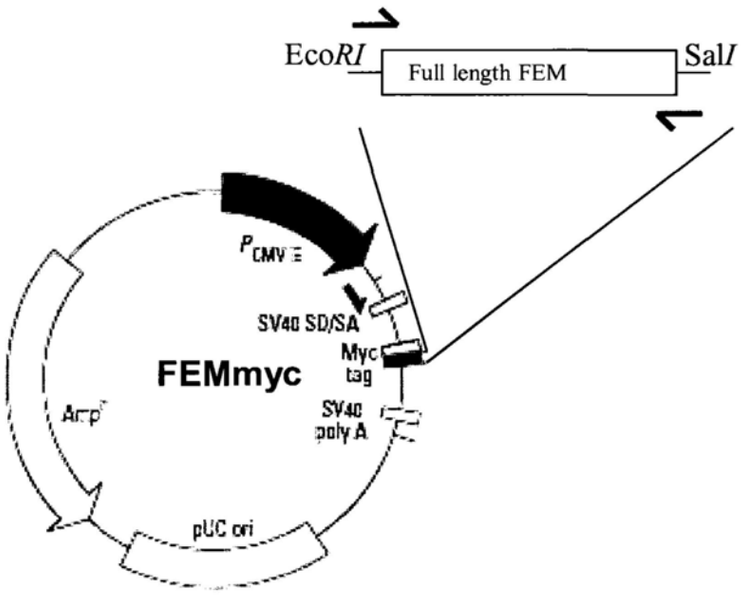


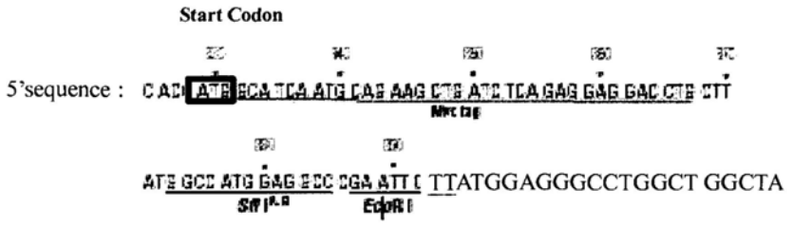
Schematic diagram of plasmids constructed
Construct Name: AP2M1myc
Encoding full length AP2M1 with N terminal myc tag
Primer sequence : **TATGCA**GTCGACAATGATTGGAGGCTT
Restriction enzyme : Sal*I* and Not*I*
Construct used in Figures 6-1, 6-2





Schematic diagram of plasmids constructed
Construct Name: SH2myc
Encoding full length SH2 with N terminal myc tag
Primer sequence : **TATGCA**GAATTCTTATGAATGGTGCCCCTTCCCCAGAGGACG
Restriction enzyme : Eco*RI* and Sal*I*
Construct used in Figures 6-4





Schematic diagram of plasmids constructed
Construct Name: FEMmyc
Encoding full length FEM with N terminal myc tag
Primer sequence : **TATGCA**GAATTCTTATGGAGGGCCTGGCT GGCTA
Restriction enzyme : Eco*RI* and Sal*I*
Construct used in Figures 6-4