From DEPT. OF LABORATORY MEDICINE, CLINICAL RESEARCH CENTER

Karolinska Institutet, Stockholm, Sweden

MOLECULAR DISSECTION OF B-LYMPHOCYTE SIGNALLING USING EXPRESSION PROFILING

Jessica M. Lindvall

Stockholm 2005

Molecular dissection of B-lymphocyte signalling using expression profiling

All previously published papers were reproduced with permission from the publisher.

Jessica M. Lindvall

ABSTRACT

Gene expression profiling and bioinformatics have emerged into playing a large and diverse role in many aspects of both the clinical and molecular research. These two approaches have in this thesis work been used in combination in order to shed light to the role of Bruton's tyrosine kinase (Btk) in B-lymphocyte development and signal transduction. When Btk is found to be defective and/or non-functional, e.g. due to mutations, these processes are disrupted, giving raise to the primary immunodeficiency disease X-linked agammaglobulinemia (XLA) in humans and X-linked immunodeficiency (*Xid*) in mice. Gene expression profiling has been used on both resting, primary splenic B-cells, with Btk-defective mice in comparison to a normal strain, as well as Epstein-Barr Virus (EBV) transformed peripheral B-lymphocytes from both XLA patients and healthy individuals. Differences as well as similarities in gene expression pattern not only within experiments, but also between B-cell lines compared to resting B-cells and whole primary splenic B-lymphocytes compared to purified transitional type 1 (T1) B-cell splenocytes, have been distinguished. Several potentially interesting genes, both of known and unknown character, have been discovered to be differentially expressed in a Btk dependent manner. Microsomal epoxide hydrolase (*mEH*), Ionized Ca^{2+} type 1 (*Iba1*) and *CD9* are three examples of genes found in the gene profile from Btk-defective mice. These genes have also been confirmed on protein level. Also, four Expressed Sequence Tags (ESTs) have been annotated and characterised by the use of bioinformatics tools. The gene expression profiling technology, in combination with bioinformatics and conventional biochemistry, has enabled the search for new target molecules in B-lymphocyte development and signalling in the context of Btk.

Jessica M. Lindvall

Till alla ni som gör mitt liv underbart...

Jessica M. Lindvall

LIST OF PUBLICATIONS

Islam T. C., **Lindvall J.**, Wennborg A., Brandén L. J., Rabbani H. and Smith C. I. E.
Expression profiling in transformed human B cells: influence of Btk mutations and comparison to B cell lymphomas using filter and oligonucleotide arrays.
Eur J Immunol. 2002 Apr;32(4):982-93.

Lindvall J. M., Blomberg K. E. M., Berglöf A., Yang Q., Smith C. I. E. and Islam T. C.
Gene expression profile of B cells from *Xid* mice and Btk knockout mice.
Eur J Immunol. 2004 Jul;34(7):1981-91.

Lindvall J. M., Blomberg K. E. M., Wennborg A. and Smith C. I. E.
Differential expression and molecular characterisation of *Lmo7*, *Myo1e*, *Sash1* and *Mcoln2* genes in Btk-defective B-cells.
Two first authors contributed equally to this work
Accepted in *Cellular Immunology* 2005

Lindvall J. M., Blomberg K. E. M., Berglöf A. and Smith C. I. E.
Distinct gene expression profile in Btk defective Transitional 1 B-lymphocytes.
Manuscript

Jessica M. Lindvall

CONTENTS

LIST OF ABBREVIATIONS

ANOVA	Analysis of variance
BCR	B-cell receptor
BLAST	Basic Local Alignment Tool
BLNK	B-cell linker protein
Btk	Bruton's tyrosine kinase
cDNA	complementary DNA
DAG	Diacylglycerol
dChip	DNA-Chip Analyser
DMT	Data Mining Tool
EBV	Epstein-Barr Virus
EST	Expressed Sequence Tag
GCOS	GeneChip® Operating Software
GO	Gene Ontology
GOTM	GOTree Machine
Iba1	Ionised Ca^{2+} type 1
Ig	Immunoglobulin
IP_3	$Ins(1,4,5)P_3$
ITAM	Immunoreceptor tyrosine-based activation motif
Lmo7	Lim domain only 7
LMP2A	Latent membrane protein 2A
LPS	Lipopolysaccharide
Mcoln2	Mucolipin2
mEH	microsomal Epoxide Hydrolase
MGED	Microarray Gene Expression Data
MIAME	Minimum Information About a Microarray Experiment
mRNA	Messenger RNA
Myo1e	Myosin 1e
NCBI	National Center for Biotechnology Information

NFκB	Nuclear Factor kappa B
PCR	Polymerase chain reaction
PH	Pleckstrin homology
PI3K	Phosphoinositide 3-kinase
PIP$_2$	Phosphoinositide PtdIns(4,5)P$_2$
PIP$_3$	Phosphatidylinositol-3,4,5-triphosphate
PKC	Protein Kinase C
PLC-γ2	Phospholipase C-γ2
PMA	Phorbol Myristate Acetate
Sash1	SAM and SH3 domain containing 1
SH	Src homology
SOM	Self-Organising Map
T1	Transitional type 1 B-cell
T2	Transitional type 2 B-cell
TH	Tec homology
WT	Wild type
Xid	X-linked immunodeficiency
XLA	X-linked agammaglobulinemia

1 INTRODUCTION

An intact B-cell development is of crucial importance for a fully functioning immune system. Many key molecules assure that this is sustained, since disruption can cause a range of diseases of varying severity. X-linked agammaglobulinemia (XLA) is a primary immunodeficiency disease caused by disruption in the B-cell differentiation path. Patients with XLA are lacking mature and antibody producing B-lymphocytes that are of great importance for a healthy immune system. The developmental block in XLA is caused by mutations in the gene Bruton's Tyrosine Kinase (*BTK*). The significance of a functional *BTK* gene in B-cell development is clear but its precise role is still largely unknown. By the use of gene expression profiling, attempts have been made to clarify the function of Btk in XLA and B-lymphocyte development.

1.1 B-LYMPHOCYTE DEVELOPMENT AND DIFFERENTIATION

Haematopoietic stem cells are thought to progress through an irreversible cascade of differentiation steps as they commit to a particular cell lineage. This is a very complex process that depends on a number of B-cell intrinsic factors that need to be expressed as well as extrinsic components found in the local environment at the time of commitment. B- and T- lymphocytes provide long-term immunological memory as well as highly specific effector functions in the immune defence to foreign antigens. For that purpose they carry antigen-specific receptors, the ligation of which regulates diverse responses such as proliferation, differentiation and cell death depending on their developmental stage and the microenvironment context. For B-cells, there are two major phases of development (Figure 1); an initial phase that is antigen independent and in which a diverse repertoire of antigen specific B-cells develops in the bone marrow. The second phase occurs when B-cells are stimulated by antigen to undergo clonal expansion in the peripheral lymphoid system.

The human body contains approximately 2×10^{12} lymphocytes, of which 5-15 % are B-cells. During B-lymphocyte development in the bone marrow (initial phase), each B-cell randomly rearranges its Immunoglobulin (Ig) gene segments to form one complete, productive heavy-

chain gene and one complete, productive light-chain gene. This is to ensure that all B-cell receptors (BCRs) on a single B-cell have the same antigen-binding site, whereas the total population of generated B-lymphocytes has a diverse repertoire of antigen-specific antibodies.
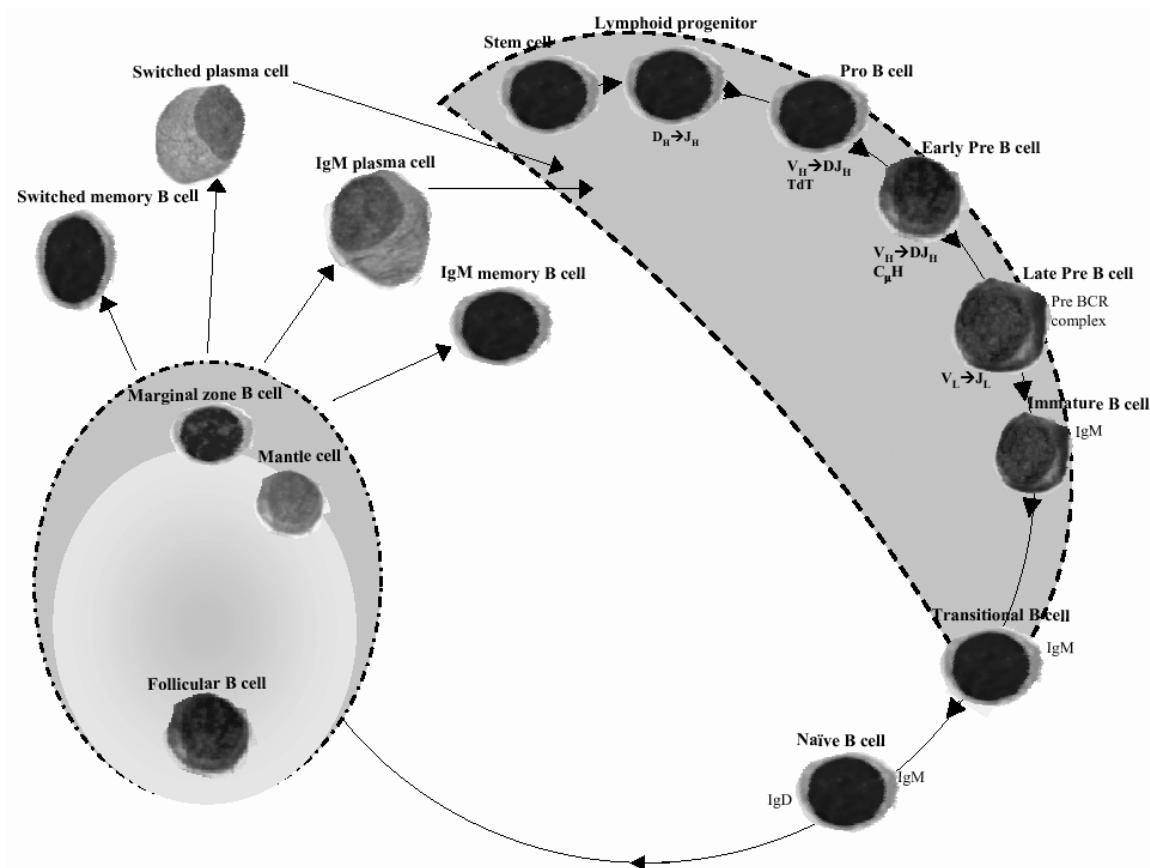
Figure 1. *Schematic overview of the B-cell development.*

During the second phase of B-lymphocyte development, taking place in the peripheral lymphoid organs such as the lymph nodes and the spleen, the immature B-cell that recognises self-antigen initiates strong BCR signalling resulting in receptor editing or apoptosis. The cells that survive this step are further selected for development into transitional B-cells (T1 and T2). Transitional B-lymphocytes leave the bone marrow and differentiate into mature circulating B-cells. These transitional B-lymphocytes must then pass through multiple checkpoints before entering the mature B-cell pool. Thus, before becoming mature B-lymphocytes, the transitional B-cells are

either selected to the follicle or the mantle zone. The follicular B lineage cells recirculate continuously and at the same time they are being enriched for specificities common to T-dependent antigens. Mantle zone B-cells, on the other hand, are static and enhanced for specificities common to T-independent antigens. The mature B-lymphocytes then circulate throughout the body and encounter foreign antigens. In addition to activating B-cell immune effector functions, signalling through the BCR also regulates B-cell development, survival and elimination of auto-reactive B-lymphocytes. Signal transduction through the BCR is an absolute requirement for the selection and development of B-cells at multiple checkpoints. The outcome of B-lymphocyte survival and differentiation or cell death depends on the maturation state of the B-cell, the level and duration of the BCR signal and also on signals from other receptors e.g. CD19 and CD40.

1.2   PRIMARY IMMUNODEFICIENCIES

The different B-lymphocyte developmental stages express various important signalling components, many of which when constitutively active, non-functional or absent cause disease. One of such diseases is the primary immunodeficiency disease XLA, which was discovered in 1952 by Dr. Ogden C. Bruton (Bruton, 1952). With the revealing of the underlying molecular basis for the disease, more than 100 other primary immunodeficiencies and their target genes have been described in the literature. These disorders involve components of both the adaptive and innate immune system, including T- and B-lymphocytes, phagocytic cells and various complement proteins. In general, the T-cell deficiencies are clinically more severe than those associated with antibody deficiency disorders. In the absence of treatment, those affected by defective cellular immunity, rarely survive beyond infancy or childhood. As with the T-cell immuno-deficiencies, the spectrum of antibody deficiency disorders is broad, ranging from the most severe type, with next to absent B-cells and Igs, to patients having selective isotype defects, to those with almost normal Ig levels in the serum.

The primary immunodeficiencies are characterised by defects of specific functions of the immune system, most often manifested by increased susceptibility to infections. The disorders, although rare, provide a wide spectrum of defects, where the antibody deficiencies represent the

most common types of primary immunodeficiencies in humans (Ochs et al., 1998). In B-cell deficiencies, the symptoms often do not appear until 9-12 months of age, since the patient has acquired protective levels of Ig from the mother. In addition to a high susceptibility to infections certain antibody deficiency disorders are associated with other features in their clinical presentation, e.g. autoimmunity and malignancies. During the last decade, many of the genes involved in antibody deficiencies have been identified and this has led to a better understanding and elucidation of the B-lymphocyte development and the differentiation pathways coupled to this.

1.2.1 XLA (X-linked agammaglobulinemia)

XLA patients present a vast number of distinct mutations in the *BTK* gene (Lindvall et al., 2005; Vihinen et al., 1997; Vihinen et al., 1995; Vihinen et al., 1999) all of which affects their B-cell development and differentiation (Campana et al., 1990; Noordzij et al., 2002). The XLA patients display a partial block between the pro-B- and pre-B-lymphocyte stage and also a complete block at the mature B-cell stage, leaving the patients without any antibody producing B-cells (Campana et al., 1990; Noordzij et al., 2002). Therefore the patients exhibit susceptibility to bacterial infections, due to virtually absent humoral immune responses (Lederman and Winkelstein, 1985; Rosen et al., 1984a; Rosen et al., 1984b). As XLA is inherited in an X-linked fashion, female carriers are found healthy but display non-random X-chromosome inactivation, which is limited to B-cells. For several years the existence of a genotype-phenotype correlation remained elusive. However, in a recent review it was demonstrated that there in fact exists such a correlation, since missense mutations are highly restricted in XLA with only certain residues being sensitive (Lindvall et al., 2005).

Approximately 1 in every 200,000 male is affected with XLA (Sideras and Smith, 1995). B-lineage cells in all organs are affected, resulting in a reduced size of lymph nodes and tonsils. The onset of infections typically occurs during the later part of the first year of life (Ochs and Smith, 1996; Sideras and Smith, 1995) and there is an almost total absence or a marked deficiency of serum Igs in the patients (Smith and Notarangelo, 1997). The commitment of haematopoietic stem cells to the B-cell lineage seems unimpaired, but the B-lymphocyte

development is arrested at the cytoplasmic μ^+ pre-B-cell stage (Campana et al., 1990; Noordzij et al., 2002). The percentage of circulating B-cells is reduced or extremely low. The severity of the block is variable, as some patients have near normal numbers of pre-B-cells. Nevertheless, these pre-B-lymphocytes are defective in their proliferative capacity, accounting for the paucity of mature B-cells in the periphery of these patients (Campana et al., 1990). However, pro-B-cell numbers in the bone marrow are normal or even increased (Campana et al., 1990; Noordzij et al., 2002). The defect is B-cell intrinsic not affecting T-lymphocytes and other lymphoid sub-populations, which appear in normal numbers (Smith and Notarangelo, 1997). Lymphoid tissues show absence of plasma cells, lymphoid follicles and germinal centers. The treatment for XLA remains the same as the one Dr. Bruton himself introduced in the 1950s, a combination of gammaglobulin substitution and antibiotics, which leaves most patients with few severe infections with this regular treatment.

1.2.2 *Xid* (X-linked immunodeficiency)

Mice also have a B lineage-specific defect originally described in the CBA/N strain (Scher et al., 1975a; Scher et al., 1975b). These *Xid* mice have a missense mutation at a conserved arginine residue within the Btk PH domain, an R28C mutation (Rawlings et al., 1993; Thomas et al., 1993). The disease in mice is milder than its human counterpart XLA where only a partial block between the pre-B- and mature B-lymphocyte stage is found (Hendriks et al., 1996; Khan et al., 1995). Thus, the mice exhibit a leaky effect of mature B-cells although they are Btk defective. *Xid* mice have half to a third the number of conventional follicular B-cells (B2 cells) of normal wild-type mice, and a severe reduction of the B1 subpopulations that predominates the peritoneum (Hardy et al., 1984). Analysis of the immune defects in *Xid* show that IgM and IgG3 subfractions of antibodies are clearly reduced (Wicker and Scher, 1986). B-lymphocytes from these Btk defective mice show a high rate of spontaneous apoptosis *ex vivo* and they are unresponsive to activation by mitogenic anti-Igs as well as thymus-independent type II antigens (Klaus et al., 1997; Mosier et al., 1977; Scher et al., 1976; Scher et al., 1975b; Sieckmann et al., 1978).

1.3 BTK (BRUTON'S TYROSINE KINASE)

1.3.1 The *BTK* gene

In 1993, two independent groups simultaneously identified the *BTK* gene (Tsukada et al., 1993; Vetrie et al., 1993). The mouse *Btk* gene was cloned concurrently as its human counterpart and the amino acid sequence was found to be 98.3 % identical to the human sequence (Rawlings et al., 1993). Also, the genomic organisation was established to be alike in the two species (Sideras et al., 1994). It was shown that the *BTK* gene encodes the cytoplasmic protein kinase affected in XLA in humans and *Xid* in mice. *BTK* has been mapped to the long arm of the X-chromosome in the region Xq21.3-22 and is composed of 19 exons ranging from 55 to 560 bases, spanning a genomic region of about 38 kb (Figure 2). The first exon contains a 5' untranslated region whereas the following eighteen exons codes for the Btk protein. Exon 19 is the longest enclosing the end of the catalytic region of the protein and also encompasses a 3' untranslated sequence (Sideras et al., 1994; Smith et al., 1998; Smith and Notarangelo, 1997).
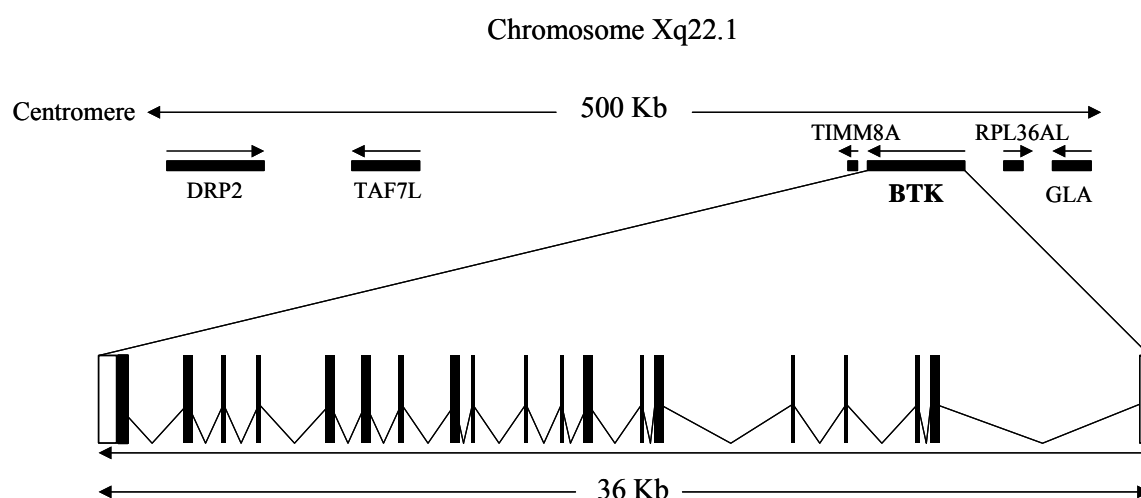

Figure 2. *The genomic composition of BTK.*

1.3.2 Cell biology of Btk

Btk belongs to the Tec family kinases, which also includes Txk, Bmx, Tec and Itk as its members (Lucas et al., 2003; Mano, 1999; Qiu and Kung, 2000). The Tec family kinases are

characterised by the presence of a Pleckstrin homology (PH) domain followed by a Tec homology (TH) domain N-terminal from the following Src homology (SH) domains [reviewed in (Smith et al., 2001)].

Btk is, from the N-terminal, built up by a PH domain, a TH domain followed by three SH domains, SH3, SH2 and SH1 the latter also called the kinase or catalytic domain (Figure 3). The Btk protein comprises 659 amino acids with a molecular weight of 77 kDa (Rawlings et al., 1993; Tsukada et al., 1993; Vetrie et al., 1993). Btk is present in all haematopoietic cells, but is selectively down-regulated in plasma cells and T-lymphocytes (Smith et al., 1994). Expression of Btk has also been detected at all stages of B-lymphocyte development (de Weers et al., 1993).

1.3.3 Btk in signalling

Various proteins, both *in vitro* and *in vivo*, have been identified to interact with the different domains of Btk (Figure 3). The different interaction partners vary from cytosolic proteins to nuclear transcription factors, equipping the Btk protein with a diversity of functions and placing it as a component of multiple signalling pathways where the BCR pathway is the most eluci-dated. Although in the last decade relatively much has been learned about events related to BCR signalling, the precise role of Btk in this pathway remains poorly understood. Nevertheless, there is now overwhelming evidence indicating that expression of Btk is indeed critical for several key steps in the life cycle of B-lineage cells including proliferation, development, diff-erentiation, survival and apoptosis (Gauld et al., 2002; Islam and Smith, 2000; Satterthwaite et al., 1998). Following antigen engagement of the BCR, tyrosine residues assembled into the so-called immunoreceptor tyrosine-based activation motif (ITAM) in the cytoplasmic tails of Igα/Igβ heterodimers are phosphorylated by members of the Src family kinases, creating docking sites for non-receptor tyrosine kinases and adaptor proteins (Borst et al., 1993; Burkhardt et al., 1991; Cambier et al., 1994; Hutchcroft et al., 1992; Reth, 1989; Reth, 1992; Reth and Wienands, 1997). In addition, as a result of phosphoinositide 3-kinase (PI3K) acti-vation, phosphatidylinositol-3,4,5-triphosphate (PIP$_3$) levels in the inner surface of the plasma membrane surge leading to the PH domain-mediated plasma membrane translocation of Btk

(Mohamed et al., 1999; Nore et al., 2000; Salim et al., 1996; Scharenberg et al., 1998; Varnai et al., 1999).

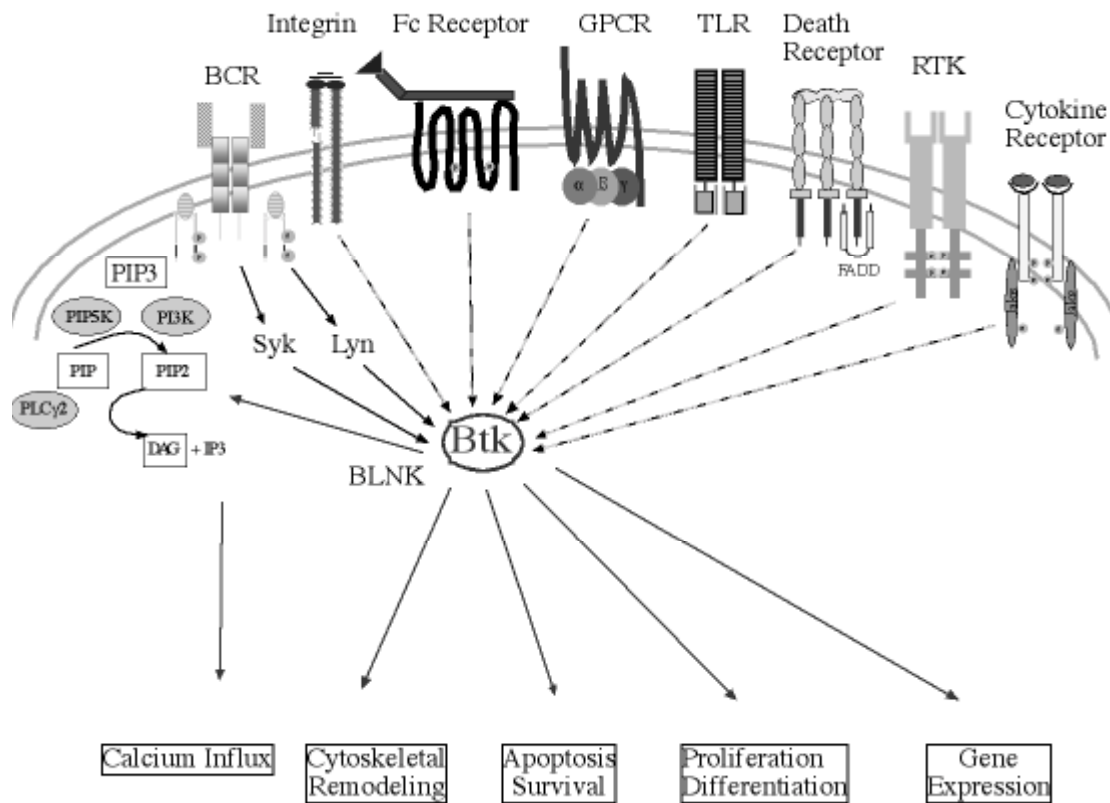| *In vivo*	BAM11 PKCbI/II, μ Fas	Gqα	Gqα		Fas
	PIP3	TH	pY223		pY551
	PH	**Btk** | **PRR**	**SH3**	**SH2**	**Kinase (SH1)**
In vitro	Gβγ BAP-135 (TFII-I) PKCθ Fas F-actin BAM11 GRK2	Gβγ Gqα Gqα12 Fyn Hck Lyn	EWS c-Cbl Gqα WASP Sab Sam68 Vav WASP	BLNK (SLP-65, BASH)	Gβγ Fas Caveolin-1

In vivo Btk protein-protein interactions: Analysis of full-length Btk. Specific interacting domain unknown.

TLR8, TLR4, MyD88, Mal, AKT(PKB), Bright, Caveolin-1

Immunol Rev. 2005 Feb;203:200-15. Lindvall JM *et. al.*

Figure 3. *The modular structure of Btk.* Schematic presentation of the different domains and their interacting molecules both *in vivo* and *in vitro*.

In addition to the BCR, activation of Btk can also occur following the stimulation of a diverse array of cell surface receptors (Figure 4), the common denominator of which is the generation of PIP$_3$ (Nore et al., 2000). The first event occurs at the plasma membrane, where members of the Src family kinase phosphorylate Btk at a key tyrosine residue in the activation loop of the kinase domain (Y551). This is followed by an autophosphorylation event at the Y223 residue in the SH3 domain (Park et al., 1996; Rawlings et al., 1996). Secondly, phosphorylated Btk recruits the adapter B-cell linker protein (BLNK also known as SLP-65 or BASH) together with phospholipase C-γ2 (PLC-γ2) to the plasma membrane bringing them in close proximity to Syk where tyrosine phosphorylation of both proteins occurs (Rodriguez et al., 2001).

Figure 4. *Schematic overview of known Btk signalling processes.*

Thirdly, phosphorylation and subsequent activation of PLC-$\gamma2$ leads to the hydrolysis of the phosphoinositide PtdIns(4,5)P$_2$ (PIP$_2$) to produce Ins(1,4,5)P$_3$ (IP$_3$) and diacylglycerol (DAG) causing calcium mobilisation and Protein kinase C (PKC) activation, respectively (reviewed in (Kurosaki, 1999)).

Accordingly, cells that are impaired or deficient in Btk activation have been shown to virtually lack BCR-dependent intracellular calcium mobilisation (Fluckiger et al., 1998; Genevier and Callard, 1997; Takata and Kurosaki, 1996). Plasma membrane signalling of Btk is believed to occur predominantly in the cholesterol/sphingolipid-rich environment — the lipid rafts and/or caveolae where other potential signal transducing molecules also gather to form the so-called

Btk-signalosome, which is composed of the Btk protein together with BLNK and PLC-γ2 (Fruman et al., 2000; Guo et al., 2000; Vargas et al., 2002). Notably, Btk signalling at the plasma membrane also involves events in the cytoplasm and the nucleus (e.g. (Antony et al., 2003; Baba et al., 1999; Backesjo et al., 2002; Bajpai et al., 2000; Lindvall and Islam, 2002; Petro et al., 2000; Rajaiya et al., 2005; Webb et al., 2000). To this end, signal transduction involving Btk has been shown to be critical in the activation of the transcription regulator nuclear factor (NF)κB following engagement of the BCR and Lipopolysaccharide (LPS) stimulation of the Toll-like receptors (Bajpai et al., 2000; Horwood et al., 2003; Petro et al., 2000). Moreover, additional transcription factors have been shown to be direct targets of Btk further reiterating its role in gene expression and regulation (Kikuchi et al., 2000; Novina et al., 1999; Rajaiya et al., 2005; Webb et al., 2000; Yang and Desiderio, 1997). Our lab and others have also shown that through direct modulation of the activity of small GTPases, Btk can induce robust cytoskeletal remodelling in stimulated cells (Inabe et al., 2002; Nore et al., 2000).

2 THE MICROARRAY TECHNOLOGY

The wealth of genomic information that were brought along after the completion of sequencing the human genome in 2001 (Lander et al., 2001; Venter et al., 2001) enables researchers to study the expression and function of every gene in the human body. In 2004 the International Human Genome Sequencing Consortium estimated the number of protein-coding genes present in the human genome to be 20,000-25,000 (Consortium, 2004). The question now arising is: How does one ever begin to simultaneously analyse this many genes in one experiment? One answer is by the use of the microarray technology. The use of gene expression profiling was first published in 1995 (Schena et al., 1995) and has since rapidly grown into a state-of-the-art technique used by a variety of different research areas. A microarray can be thought of as a miniaturised gene hybridisation– or detection– array. Individual microarray assays are measured in microns where the range of such an element can be as small as 5μm and up to 200μm depending on the type of array. Each element contains a DNA sequence from one gene and is used to measure the expression of its corresponding messenger RNA (mRNA) in a sample. There are now many different types of microarray platforms to choose from when considering a gene expression profiling project and two of the most common approaches are in brief described below.

2.1 COMPLEMENTARY DNA ARRAYS

One of the most widely used microarray technologies is the complementary DNA (cDNA) arrays where full-length cDNA clones are spotted either on filters or on specially coated glass slides. The full-length cDNA clones are helpful in enabling stringent hybridisation conditions and lowering cross-hybridisation of unrelated genes, although closely related gene families will still be able to anneal to some extent. The cDNA spotted arrays requires simultaneous analysis of two biological samples, a test and a control, each labelled with a different fluorescent dye (Cy3 (green) and Cy5 (red), respectively). Advantages using the spotted cDNA arrays are several, one of them being the low cost in comparison to the commercial Affymetrix high-density short oligonucleotide array technology.

2.2   SHORT OLIGONUCLEOTIDE ARRAYS

The short oligonucleotide array technology has been commercially available for several years and was first reported in early 1990 (Fodor et al., 1991). With complex eukaryotic genomes, there might be problems with cross-hybridisation of the oligonucleotides to unrelated probes due to the short length of the target sequence. The use of several oligonucleotides per gene or transcript helps to eliminate this problem. A major challenge for oligonucleotide arrays is the need for adequate design of gene-specific oligonucleotides with optimal sensitivity. Some oligo-nucleotide arrays, e.g. the GeneChips[®] from Affymetrix (described below), use a technology that takes advantage of the hybridisation specificity when using short oligonucleotides. It also makes use of, in contrast to the spotted cDNA arrays, the so-called "one sample, one chip" approach. However, the short oligonucleotide arrays have been criticised for a lack of sensitivity due to that a single capture probe is not always sufficient to distinguish the expression of a particular gene.

2.2.1   Affymetrix GeneChip[®] arrays

http://www.affymetrix.com/index.affx

The GeneChip[®] arrays (Lipshutz et al., 1999; Lockhart et al., 1996) are high-density short oligonucleotide arrays, which contain both known genes and so-called expressed sequence tags (ESTs). ESTs are short sub-sequences of expressed DNA and are synthesised using mRNA as a template. Short oligonucleotides are ordered by chemical synthesis (Chee et al., 1996; Fodor et al., 1991). The oligonucleotide probes are shorter than the DNA probes used in e.g. cDNA arrays where 100-2000 bases are used instead of 25mers in the Affymetrix GeneChip[®] tech-nology. The 25mer probe length confers high specificity and the use of multiple probes provides for high sensitivity and reproducibility. Eleven to twenty probe pairs are routinely used for each expression measurement. Affymetrix GeneChip[®] arrays have several advantages. One is the comparatively smaller feature size, permitting the assaying of larger numbers of transcripts in a single experiment. The advantages of designer probe sequences allow uniform hybridization behavior and the ability to distinguish closely related sequences, thereby enriching transcript

quantitation. Current densities of commercially available Affymetrix GeneChip® arrays allow
measurements of up to more than 56,000 probe sets on one single array for the human genome.

2.2.1.1 *The Affymetrix technology*

The Affymetrix microarray consists of a silicon slide with 25mer oligonucleotide sequences
"built" directly onto the surface using the same technology that revolutionised the micro-
processor industry — photolithographics (Fodor et al., 1991). In essence, the power of light is
harnessed to catalyse intermolecular reactions in building the oligonucleotide chains, thereby
producing the smallest feature size on the market (Lockhart et al., 1996). Chemical building
blocks (hydroxyl-protected deoxynucleosides) are incubated with the surface, and chemical
coupling occurs at those sites that have been illuminated in the preceding step. The following
step requires light to be directed to different regions of the substrate using a new mask, and the
chemical cycle is repeated. Random polynucleotides can thus be synthesised in a highly specific
manner at defined locations. Also, the GeneChip® array has multiple independent detector
probes built onto its surface for each targeted mRNA. This improves the signal-to-noise ratios
and the dynamic range of detection, as well as minimising cross-hybridisation effects. It also
helps to ensure the performance of the array. An additional redundancy comes from the use of
mismatch control probes identical to the perfect match partners with the exception of a single
base difference located in a central position. The mismatch probes serve as controls for specific
hybridisation and they facilitate the direct subtraction of background and cross-hybridisation
signals. Probes are selected from novel gene-specific sequences with optimal hybridisation be-
havior, biased toward the 3' end (to enhance information collection from samples using poly-A
extraction methods) and avoiding palindromes. If the probes are properly designed, each probe
will hybridise to mainly one specific sequence. In initiating the Affymetrix protocol, total RNA
is isolated and used as a template for double-stranded cDNA creation by a linear reverse
transcriptase polymerase chain reaction (PCR) using poly-T primers containing a T7 RNA
polymerase promoter sequence. The cDNA is then transcribed and labeled with biotin by in
vitro transcription using T7 RNA polymerase. The biotin-labelled RNA is then hybridised to
the array, which is stained with phycoerythrin-conjugated streptavidin prior to the scanning. A

grid is automatically laid over the array image and the intensities of each probe pair are used to make expression measurements. The general idea of the Affymetrix method (or short oligonucleotide arrays) is outlined in Figure 5.

Gene/mRNA sequences

Bioinformatics/
oligo design
algorithms

Perfect match ⎤ oligo
mismatch ⎦ set

Photolithographic
in situ synthesis

Labeling-hybridization/
washing

Laser

Scanning/data
analysis

Drug Discov Today. 2003 Feb 1;8(3):134-41. Barrett JC & Kawasaki ES
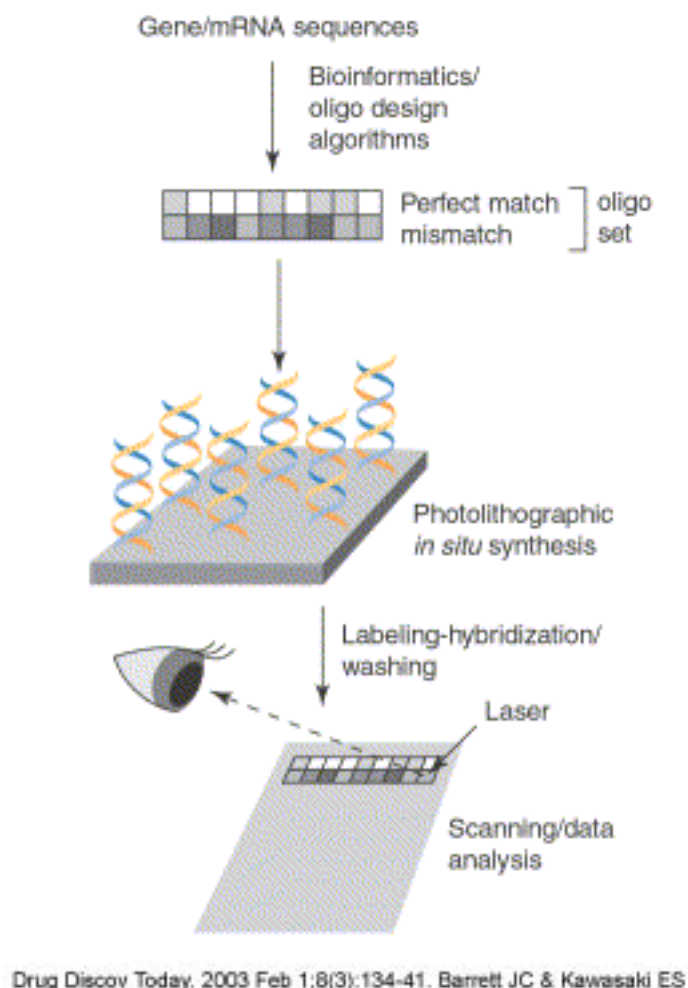
Figure 5. *Affymetrix GeneChip® oligonucleotide arrays.*

2.2.1.2 Affymetrix data analyses

Differences between perfect match and mismatch intensities as well as between perfect match and background intensities are used to calculate the Change p-value by the Wilcoxon's signed rank test. Also the level of photomultiplier saturation for each probe pair is evaluated where any

saturated probe cell is rejected from the analysis. The Change p-value ranges from 0.0 to 1.0 providing an estimate of the likelihood and direction of the change. For values close to 0.5, true change in the level of expression between the two RNA samples is unlikely, while values close to 0.0 indicate probability for an increase in gene expression level in the experimental array compared to the baseline array, whereas values close to 1.0 indicate likelihood for a decrease in gene expression level in the experimental array.

The Signal Log Ratio algorithm is used to estimate the measure and the direction of change of a transcript when two arrays are compared. Each probe pair on the experimental array is compared to the corresponding probe pair in the baseline array and the Signal Log Ratio is calculated using a One-step Tukey's Biweight method. This process eliminates differences due to different probe binding coefficients. The base 2 log scale is used, translating the Signal Log Ratio of 1.0 to a 2-fold increase in the expression level and of -1.0 to a 2-fold decrease. No change in the expression level is thus indicated by a Signal Log Ratio value 0 i.e. the transcript is found to have the ratio of 1 in the Comparative analysis.

2.3   EXPERIMENTAL DESIGN

Good experimental design in a microarray project requires the same principles and practices that are part of any scientific investigation. Appropriate controls are the foundation to any well-performed experiment. Both positive and negative controls can provide confidence in the results and even provide insight into the success or failure of the experimental protocol. Sufficient replicates should be planned to decrease experimental error and to provide statistical power. Another consideration is whether to pool samples or not. Pooling samples is often considered when RNA is in limited supply, or to minimise the effects of biological variation. Pooling is a debated issue since more errors can be introduced into the data. However, is it has recently been demonstrated that pooling RNA from an increased number of subjects can reduce the number of microarrays required, without any loss of precision (Kendziorski et al., 2003; Peng et al., 2003). Consideration regarding the experimental parameters is a must, so that whatever treatment, time, dose, individual, or tissue location is being studied, the results will be interpretable with

minimum number of confounders. Also, attention to the conventional but critical statistical and data analyses elements of a microarray experiment is needed in order to assure the very best outcome possible of the results.

2.4 SYSTEMS BIOLOGY

Although the scale of the transcriptomic experiments are impressive and researchers now have access to quantities of data that would have taken years to obtain in the past we now face the need to create means for interpreting the data. The construction of such methods requires a systematic approach that is built on concepts and techniques learned from other scientific disciplines such as mathematics, physics, stochastic processes etc. All of this knowledge wrapped up together with biology is what now is called systems biology. Systems biology takes advantage of computational methods to acquire and analyse interaction networks, such as gene expression pattern results, with the goal of mapping networks and finding patterns of regulation within them. Thus, although systems biology is dependent on gene expression profiling (as well as other high throughput technologies) for data generation, it encompasses the design and the use of analysis tools as well as the development of new ways to represent the data that are both meaningful and allow insight. For a more general overview and introduction on the topic system biology and bioinformatics see (Aderem, 2005; Ehrenberg et al., 2003; Vihinen, 2003). Another subject, as the volume of microarray data continues to expand, is the increasing need for accepted standard operating procedures and coordinated data deposition. The MGED (Microarray Gene Expression Data) Society exists to encourage MIAME (Minimum Information About a Microarray Experiment) compliance (Brazma et al., 2001) (guidelines can be found at http://www.mged.org/index.html), so that microarray data from different laboratories and platforms can be unambiguously interpreted and independently verified.

2.5 NORMALISATION

In order to compare one microarray to another, removing the systematic variation found in each array is necessary. Systematic differences can arise due to experimental factors such as sample

preparation, feature location, arrays, and fluorescent labelling efficiencies as well as various interactions of these effects. Thus, normalisation is a critical initial step to balance the individual signal intensities across the experimental factors while maintaining the class effect of interest. The normalisation procedure will affect different microarray experiments to different extents. Problems with the normalisation process exist and one of them is the recognition of the source of systematic bias. When normalising the data, there is a strong possibility that some of the biological information will be removed. The normalisation process should therefore be kept to a minimum.

2.6   DATA ANALYSIS

Data analysis is perhaps the most critical aspect of a microarray experiment and the least understood by the majority of biologists. Many different analytical approaches have been developed to achieve sensitivity in detecting gene changes while also providing a measure of statistical significance and likelihood of error. Measures of statistical significance can be made across multiple biological and technical replicate microarrays. The complexity of the data and statistical analysis requires the use of sophisticated visualisation– and analysis– software. The aim of a microarray experiment is usually to identify differentially expressed genes, with a measure of statistical significance. Most microarray experiments are designed with only one categorical factor (e.g. treatment or genotype), so the statistical analysis is often based on the t-test. Experiments with multiple categorical factors (e.g. genotype and time) require methods based on the analysis of variance (ANOVA). Once the data are appropriately normalised, it is common practice to consider a univariate testing problem for each gene and calculate t-statistics. The t-statistic tests the null hypothesis of equal mean expression levels in the two samples (e.g. treatment and control).

One of the challenges in interpreting microarray data is to group genes on the basis of similar regulation and function, or similar cellular state and biological phenotype. The idea is that genes that cluster together might contribute to the same biological pathway. Clustering is an exploratory method that can give valuable insight into the underlying relationships that are otherwise

not readily found in the data and thus can suggest interesting hypotheses concerning the relation-ships. Identification of unknown classes of co-regulated genes using whole-genome expression profiles [unsupervised learning e.g. cluster analysis using hierarchical clustering (Eisen et al., 1998)] or classification into known classes of functionally or structurally related genes (super-vised learning, discriminant analysis) are two common techniques used in microarray experi-ments. Supervised methods require the genes or conditions to be associated with labels that provide information about a pre-existing classification. This information comes from outside the microarray experiment and might include knowledge of gene function or regulation, disease subtype or tissue origin of a cell type. This classification information is used to drive the analysis of the microarray experimental results and thus (in a sense) 'supervises' the analysis. Unsupervised methods, however, require no additional information besides the expression data itself. They are geared towards the discovery of patterns in the data, unbiased by outside know-ledge. Not surprisingly, unsupervised methods are used for exploratory tasks, whereas super-vised approaches are used to address more direct questions (e.g. can the microarray data be used to predict accurate labels for new genes?).

2.7   DATA VALIDATION

The biomedical research community does not yet accept that microarray data can stand alone, without independent validation. Quantitative Real-Time PCR is commonly used to confirm mRNA levels, as it has higher sensitivity and lower RNA requirements than Northern blot. We and other have demonstrated that genes with relatively high expression and at least 2-fold reg-ulation are likely to be validated using Real-Time PCR technology (Islam et al., 2002; Lindvall et al., 2004; Rajeevan et al., 2001). Also, while mRNA reflects the functional state of the cell, it is the proteins, which ultimately carry out the instructions of the genome. Translation of mRNA into protein may be controlled independently of transcription and proteins may undergo post-translational modifications that alter their function. Therefore, to describe a biological event or system, gene expression data obtained by microarray analysis must be extended to the study of protein products e.g. by the use of Immunoblotting (see below). Defining the functions of

differentially expressed genes may be considered the ultimate validation of microarray data e.g. *in vitro* or *in vivo* experiments.

2.8 BIOLOGICAL APPLICATIONS

As the types and numbers of applications for microarray experiments are quite variable and constantly increasing, gene expression profiles utilised to monitor the expression level of genes in comparison between two conditions, remains one of the most widespread uses of the technology. The results from such studies could be observed as up- or down-regulation, or unchanged during particular conditions. The technology has also been used to identify novel target genes. On a basic scientific level, gene expression profiling is exercised to map the cellular, regional, or tissue-specific localisation of genes and their corresponding proteins. The microarray technology has been widely used in studies of lymphomas and solid tumours, drug response studies as well as basic research in immunology (Alizadeh et al., 2000; Clarke et al., 2004; Islam et al., 2003; Islam et al., 2002; Villeneuve and Parissenti, 2004). Other examples where microarrays have been used in order to study the transcriptome are: at the cellular level to map genes that are distinguished between different types of immune cells (Shaffer et al., 2001); at the sub-cellular level to map genes that encode membrane or cytosolic proteins (Ding et al., 2005; Fruman et al., 2000; Lindvall et al., 2004) and at the tissue region level to identify genes which are expressed in e.g. muscle, liver, or heart (Islam et al., 2003; Kim and Park, 2005; Lundholm et al., 2004; Nikawa et al., 2004; Sarang et al., 2003; Watanabe et al., 2005; Yanagawa et al., 2005). Because dozens, hundreds and sometimes even thousands of genes and proteins are involved in a given cellular activity, understanding how these genes work together and how these expression patterns overlap is critical for the knowledge on how genomes from different species function.

3 MATERIAL AND METHODS

3.1 EBV (EPSTEIN-BARR VIRUS) TRANSFORMATION

Epstein-Barr Virus (EBV) establishes a persistent latent infection in peripheral B-lymphocytes in humans and is associated with a variety of malignancies and proliferative disorders (Baumforth et al., 1999; Iwatsuki et al., 2000). Latent membrane protein 2A (LMP2A) is a viral protein expressed in latently infected B-cells. In EBV transformed B-lymphocytes grown *in vitro*, the LMP2A functions to block the BCR signal transduction, preventing normal calcium flux and the accumulation of tyrosine-phosphorylated proteins following BCR cross-linking (Longnecker and Miller, 1996). B-lymphocytes infected with EBV *in vitro* are immortalised and subsequently termed lymphoblastoid cell lines (Levitt et al., 1984). EBV will infect any resting B-cell, driving it out of the resting state to become an activated proliferating lymphoblast. Haematopoietic experimental studies often make use of EBV transformed B-lymphocytes and has also provided us with a tool in order to study the gene expression profile in XLA patient B-cells. EBV transformed peripheral blood B-lymphocytes were used from three independent controls and two XLA patients with known *BTK* mutations **(Paper I)**.

3.2 B-CELL SPLENOCYTES

Splenic B-cells were purified from age and sex matched *Xid*/CBA (obtained from Charles River Laboratories, Sweden), Btk KO/CBA (Btk KO/CBA were created by back-crossing Btk KO/SW129 mice (Khan et al., 1995) on CBA background strain for nine generations) and normal CBA mice (wild type) **(Paper II)**. B-cell splenocytes were also collected from 1-week-old CBA mice (wild type), *Xid*/CBA mice and Btk KO/CBA mice. In the spleens of one-week-old mice, all B-lymphocytes are of T1 origin (Loder et al., 1999). For the CBA control group an average of seven mice were used, whereas for the *Xid* and Btk KO groups nine respectively fifteen mice were taken in order to obtain 100 ng total RNA for the microarray experiments. CD19+ (PE-conjugated rat anti-mouse-CD19 antibodies) B-cells were sorted in order to obtain the T1 B-cell population. The purity of the cells was >92% **(Paper IV)**.

3.3   RNA PREPARATION

EBV transformed B-cells were cultivated under normal culture conditions and RNA was prepared using the Atlas^{TM} RNA labelling kit from Clontech as at the time was recommended by the supplier. Polyadenylated (Poly A$^+$) mRNA was prepared from 40 µg of RNA with the Clontech kit. This RNA preparation method was used for all RNA extractions prior to both the filter arrays and the Affymetrix GeneChip$^®$ hybridisations **(Paper I)**. For the hybridisations of the mouse U74Av2 GeneChip$^®$ arrays, RNA from whole splenic B-cells was prepared using the TRIzol method as directed by Gibco. A total of 10 µg total RNA was used in the hybridisations of the mouse U74Av2 GeneChip$^®$ arrays **(Paper II)**. For the Real-Time PCR experiments, total RNA from splenic B-cells (both unstimulated and stimulated) was extracted with the Qiagen RNeasy Mini kit. Total RNA (100 ng) from the T1 B-cells was extracted using the Qiagen RNeasy Mini kit and hybridised to the Mouse 430 2.0 GeneChip$^®$ arrays **(Paper IV)**.

3.4   IMMUNOBLOTTING

Immunoblotting is a technique for analysing and identifying proteins. The proteins are separated by electrophoresis in a polyacrylamide gel, and then transferred (blotted) onto a nitrocellulose membrane, where they bind in the same pattern as they formed in the gel. The antigen is overlaid first with a primary antibody, and then with a secondary antibody (anti-immunoglobulin or protein A) labelled with an enzyme (in our case; horseradish peroxidase). The horseradish peroxidase conjugated antibody can easily be detected with a chemiluminescence substrate, producing luminescence (photon) signals that can be detected by a phosphoimager instrument or with photography film. This technique was used as a conformational method in **Paper II**.

3.5   FLOW CYTOMETRY

Flow cytometry enables subpopulations of lymphocytes to be identified and quantitated by utilising monoclonal antibodies to various cell surface antigens. We have measured the quantity

of CD9 protein in total splenic B-cells from immunocompetent control mice and *Xid* mice **(Paper II)**.

3.6 QUANTITATIVE REAL-TIME PCR

Real-Time PCR is a widely used method for gene quantitation. It has a large dynamic range and has tremendous sensitivity as well as it can be highly sequence specific with little or no post-amplification processing. We have used the TaqMan chemistry together with the Comparative C_t ($2^{-\Delta\Delta Ct}$) method in order to calculate the changes in gene expression as a relative fold change difference between the experimental sample and the calibrator **(Paper II)**. The ribosomal RNA 18S has been used as the endogenous control.

3.7 FILTER ARRAYS

Clontech Atlas^{TM} Human Hematology/Immunology arrays, containing 588 genes, were hybridised to ^{32}P labelled cDNA probes prepared from EBV transformed B-cells from both healthy individuals and XLA patients. The hybridisations were performed according to the manufacturer's instructions. The filters were exposed to films overnight at -70°C. The filters were stripped after the initial screening in order to perform a duplicate experiment. This was carried out with independent RNA preparations from the same EBV transformed B-cell lines. The signals were identified by phosphoimaging and quantified using the Clontech Atlas^{TM} image software version 1.0 **(Paper I)**.

3.8 GENECHIP® ARRAYS

Five Affymetrix GeneChip® human U95Av2 arrays were used together with the Affymetrix Microarray Suite 4.0 and Data Mining Tool (DMT) version 3.0 in order to compare one XLA patient with two controls. The samples were independent RNA preparations from the same cell lines used on the Clontech Atlas™ filter arrays (one XLA patient and one of the controls) **(Paper I)**. Six mouse U74Av2 GeneChip® arrays, three in the initial round of experiment and

three in the duplicate experiment, were run on whole splenic B-cells from wild type (CBA), *Xid* and Btk KO mice **(Paper II)**. These data were analysed using Affymetrix Microarray Suite version 4.0 and DMT version 3.0. Duplicate preparations of T1 B-cell splenocytes, in total six arrays, from wild type, *Xid* and Btk KO mice were hybridised to Affymetrix Mouse 430 2.0 GeneChip® arrays and initially analysed using the Affymetrix GeneChip® Operating Software (GCOS) version 1.2 **(Paper IV)**. Throughout every experiment scaling, with a factor of 100, was used as the normalisation method of choice **(Papers I, II, IV)**. When scaling is applied, the intensity of the probe sets from experimental and baseline arrays is scaled to the same, user-defined target intensity.

3.9   STATISTICAL ANALYSIS

Due to few samples (as well as limited statistical knowledge at the time), a ratio-based analysis, comparative analysis was used in the two first studies **(Paper I and II)**. For the Clontech Atlas™ filter array and U95Av2 GeneChip® result, respectively, fold differences above ≥ 1.9 were considered significant **(Paper I)**. An Absolute– (single chip analysis) and a Comparative– analysis (ratio-based analysis, test vs. control) were performed in **Paper II** in order to find the genes intrinsic to the Btk defect. A fold difference of >2 and <-2 in the wild type compared to either *Xid* or Btk KO was considered to be a significant change. A fixed ratio of >4 and <-4 was considered significant when comparing the difference between the *Xid* and Btk KO mice **(Paper II)**. As the initial analysis to which of the three mouse groups (CBA, *Xid* and Btk KO) significantly differed from one another the use of the one-way ANOVA test, with mouse strain as the factor to consider, together with the Tukey-Kramer test, was applied. Also, in order to evaluate if the duplicate experiments differed significantly from each other, a t-test and a correlation analysis were used **(Paper IV)**.

3.10 GENE EXPRESSION ANALYSIS TOOLS

To assess the biological relevance in a gene expression profile, you need to rely on the results produced by the analysis tools of your choice. The microarray analysis tools aim to sort out the

biological information in the data and this is done by various algorithms implemented in the different softwares. Below is a short description of each tool that has been used throughout the projects.

3.10.1 DMT (Data Mining Tool)

The DMT software is an Affymetrix commercial analysis program, which provides a variety of tools for filtering and sorting the GeneChip® array data. Key features include pair wise statistical analysis (Mann-Whitney and t-test analysis), clustering algorithms (Self-Organising Maps (SOM) as well as a Pearson's Correlation Coefficient method) and annotation integration. We have been using this software in **Paper I** and **II**.

3.10.2 dChip (DNA-Chip Analyser)

http://biosun1.harvard.edu/complab/dchip/

The dChip software is a program, which offers several analysis methods (Li and Wong, 2001; Schadt et al., 2001). The model-based approach allows probe-level analysis on multiple arrays. Numerous gene expression studies, including one of ours (**Paper IV**), have made use of this program in the data analysis (Falt et al., 2005; Tsapogas et al., 2003).

3.10.3 GeneCluster 2.0

http://www.broad.mit.edu/cancer/software/genecluster2/gc2.html

GeneCluster 2.0 is a software package for analysing gene expression and other so-called bio-array data (Tamayo et al., 1999). This tool enables the user to utilise a variety of methods that is presented within the software. It includes unsupervised learning methods (hierarchical clustering, SOM, principal component analysis) and supervised machine learning methods suitable for classification and prediction applications (weighted voting and support vector machine). These methods are improved by a set of statistical tools that support the estimation of statistical

significance of observations (e.g., permutation testing). This analysing tool was used in **Paper IV**.

3.11 GENERAL BIOINFORMATIC TOOLS

When gene lists of interest have been extracted from the microarray analysis, gene annotation and relationship within the transcript lists is attempted to be found using bioinformatic tools. The Internet provides the infrastructure for accessing and sharing biological information, and has been decisive in the development of functional genomics and bioinformatics. I have been using various bioinformatic tools available through the use of the Internet. Depending on the biological question underlying the specific project a somewhat diverse approach has been applied in the papers. Overall, the databases from National Center for Biotechnology Information (NCBI) (Maglott et al., 2005; Wheeler et al., 2005) and Ensembl (Hubbard et al., 2005) have been widely used for the annotation process in each project covered by this thesis. In **Paper III** the two databases have been used in parallel due to the possibility of variations in the assembly of the genome and possible local divergences related to this could be found among the databases. As both of these databases are very large in what they are offering their users, I will not even try to cover all of their contents in this thesis. Below, only a short description for each used tool or database is described.

3.11.1 NCBI (National Center for Biotechnology Information)

http://www.ncbi.nlm.nih.gov/

NCBI creates public databases (which are freely available through the world-wide-web), conducts research in computational biology, develops software tools for analysing genome data and disseminates biomedical information. The center supports and distributes a variety of databases for the medical and scientific communities. In this thesis the Basic Local Alignment Search Tool (BLAST) search engine has especially been used in order to characterise and annotate unknown transcripts found in Paper II **(Paper III)**. BLAST finds regions of local similarity between sequences and the program evaluates either nucleotide or protein sequences to

sequence databases and calculates the statistical significance of the matches found. NCBI also provides links to several other useful biological databases and search-engines such as PubMed and LocusLink.

3.11.2 Ensembl

http://www.ensembl.org/

Ensembl annotates known genes and predicts novel genes, with functional annotation from the InterPro (Mulder et al., 2003), the Online Mendelian Inheritance in Man (OMIM) disease database (Antonarakis and McKusick, 2000), protein and gene family databases as well as SAGE expression databases. The Ensembl website provides a variety of alternative views of the data such as *Mapview* (relationship between cytogenetic bands and the genome sequence), *Geneview* (genes with their transcripts and gene structures) and *Proteinview* (individual Ensembl translations with functional annotations from InterPro). Similarity searching is also integrated into the Ensembl website with the use of BlastView. This program provides an integrated platform for sequence similarity searches against Ensembl databases and has especially been used in **Paper III**.

3.11.3 NetAffx

http://www.affymetrix.com/analysis/index.affx

NetAffx is a tool offered through the Affymetrix web site providing details regarding the probe sets used on the GeneChip®, and to describe them (where possible) in a functional context. The database is easy accessible (as well as free) for the user. The information supplied for each probe id falls into two categories. The first is static information, which provides the details regarding target sequence and describes what the probe was designed to interrogate. The other category is annotations, which refer to the information about the representative sequence for a probe set. The annotations include information from the NCBI database (Maglott et al., 2005; Wheeler et al., 2005), different model organism databases (Chen et al., 2005; de la Cruz et al., 2005; Eppig et al., 2005; FlyBaseConsortium, 2003), SwissProt (Boeckmann et al., 2003) etc.

Also, where possible, further annotations are made use of in order to put the probe sets into a more functional framework e.g. utilising GO terms (discussed below) (Ashburner et al., 2000). This database has been used throughout the whole thesis work.

3.11.4 GO (Gene Ontology) Consortium

http://www.geneontology.org/

The Gene Ontology^{TM} (GO) Consortium (Ashburner et al., 2000) aims to produce a dynamic controlled vocabulary that can be applied to all organisms even as knowledge of gene and protein roles in cells is accumulating and changing. An Ontology comprises a set of well-defined terms with well-defined relationships. Data can be annotated to varying levels depending on the amount and completeness of available information. This flexibility allows the users to narrow or widen the focus of queries. An Ontology can therefore be a vital tool enabling researchers to turn data into knowledge. The consortium is constructed of three essentially non-overlapping parts covering, the process to which the genes or proteins contribute (*Biological process*), function they fulfil (*Molecular function*) and their location within cells (*Cellular component*). This relationship between these three parts is one-to-many, reflecting the biological reality that a particular protein may function in several processes. Use of GO in analysis of experimental data from high-throughput methods enables integration of biological background data in a controlled manner. The database has been used throughout all thesis work.

3.11.5 GOTM (GOTree Machine)

http://genereg.ornl.gov/gotm/

GOTM is a web-based platform (Zhang et al., 2004) using GO hierarchies (Ashburner et al., 2000) for the interpretation of e.g. microarray data. Each of the three parts of GO (*Biological process*, *Molecular function* and *Cellular component*) is a Directed Acyclic Graph where nodes are GO terms describing a particular fraction of any of the three different parts. GOTM converts gene symbols, Affymetrix probe set Ids, Unigene IDs, Swiss-Prot IDs or Ensembl IDs to LocusIDs. A hierarchical GOTree structure is then generated, which is based on GO anno-

tations for the LocusIDs. The enriched GO categories are presented in a flat view format, a sub-tree view format and a Directed Acyclic Graph view format. This tool has been used in **Paper IV**.

4 AIMS OF THE THESIS

The overall aim was to study the role of Btk in B-lymphocyte development by the use of the gene expression profiling technique and by those means identify potential Btk targets in XLA and *Xid*, respectively. In detail, the aims were:

To study the influence of *BTK* mutations, using both Clontech Atlas™ filter arrays and Affymetrix (human U95Av2) GeneChip® high-density oligonucleotide arrays **(Paper I).**

To study the use of functional annotation and validation of expression data by comparison of microarray analyses **(Paper I)**.

To study the transcriptome, using Affymetrix (mouse U74Av2) GeneChip® arrays of the whole splenic B-lymphocyte population from immunocompetent control (CBA), *Xid* and Btk KO mice and to compare the differences in gene pattern between the different mouse groups **(Paper II)**.

By the use of an *in silico* approach, using different bioinformatics tools, annotate four potential Btk target genes found in Paper II **(Paper III)**.

By the use of Affymetrix (mouse 430 2.0) GeneChip® arrays, study if the Btk defect already manifests itself at the T1 B-lymphocyte stage **(Paper IV)**.

5 ETHICS

All studies were performed in accordance with approved ethical permissions. Dnr: 144-01 **(Paper I)**, Dnr: S4-99 **(Paper II)**, Dnr: S94-03 **(Paper III and IV)**.

6 PAPER I

Islam T. C., **Lindvall J.**, Wennborg A., Brandén L. J., Rabbani H. and Smith C. I. E.
Expression profiling in transformed human B cells: influence of Btk mutations and comparison to B cell lymphomas using filter and oligonucleotide arrays.
Eur J Immunol. 2002 Apr;32(4):982-93.

A decade ago, gene expression profiling made its debut and has since contributed to the understanding of both the clinical– and molecular– side of biology. The microarray techniques have emerged into powerful tools in order to study large set of transcripts and have allowed for comparative analysis on a high-through-put level. Gene expression profiling has developed into playing a versatile role in biology where researches have focused on studying cancer-related expression profiles using Lymphochip microarrays (Alizadeh et al., 2000) as well as investigating the underlying unknown mechanisms for a particular immunodeficiency disease using cDNA arrays (Reth and Wienands, 1997). Many other examples are found throughout the literature where different biological areas are studied as well as a range of microarray technologies are applied. We have in this study used both filter arrays and Affymetrix high-density short oligonucleotide arrays in order to investigate the influence that *BTK* mutations have on the expression profile in XLA patients. We also compared the gene expression profile found in our study to a previously published lymphoma expression profile done by the use of the Lymphochip microarrays (Alizadeh et al., 2000). The Lymphochip is a cDNA microarray especially suited for the analysis of lymphoid cells. In addition to this, the influence of EBV-transformation on peripheral B-lymphocytes from blood has been examined.

As XLA patients have a blockage in their B-cell development, resulting in a lack of mature antibody secreting B-lymphocytes, patient material is not easily accessible. Therefore, peripheral blood lymphocytes are EBV-transformed and by these means a stable B-cell line is attained. Cell lines from three wild type (healthy individuals) and two XLA patients were established and studied using microarrays. WT1 (wild type 1), WT2 (wild type 2) and XLA1 (patient 1) were subjected to the Clontech Atlas™ Human Hematology/Immunology cDNA arrays and 55 genes

were found to be overlapping (expressed) in all three samples. In the comparative analysis between the XLA1 versus WT1 and XLA1 versus WT2, where an average of the two experiments were made, 11 genes were found to be differentially expressed by a factor of 1.9 or more. In order to extend these results; the study was expanded to include a set of Affymetrix GeneChips®. An independent wild type (WT3), WT1 and an additional patient (XLA2) was subjected to the Human U95Av2 arrays and 391 genes were found to be differentially expressed in at least four of the six comparisons made.

Biological variations in gene expression patterns have previously been examined (Cheung et al., 2003; Cheung and Spielman, 2002; Oettinger et al., 1990). Therefore we tried to determine the genes expressed in B-cells by comparing a duplicate microarray experiment using independent RNA preparations (referred to as H1 and H2 in Paper I). We found 52 genes out of 588 to be expressed with a (strong) similar signal. These genes represented a broad spectrum of cellular character. Thus, our findings demonstrate that gene expression profiling is an applicable technique for investigating the influence of *BTK* mutations in EBV-transformed B-lymphocytes from XLA patients. Also, the biological transcriptomic variation in a sample can be identified by the use of expression profiling.

A comparison to a previously published lymphoma expression profile (Alizadeh et al., 2000) was made in order to obtain a more comprehensive functional classification of the genes and transcripts being detected as well as differentially expressed in the XLA cell lines. Alizadeh *et. al.* used a Lymphochip containing approximately 18,000 clones to obtain expression clusters from 96 different lymphoma samples. The comparison with our gene expression profile from the EBV-transformed B-cell lines with the Lymphochip study was done by mapping the accession numbers to clusters in the UniGene database. The same approach was applied to the Atlas™ array study and the Lymphochip study and revealed predominance in the Alizadeh *et. al.* characterised Pan B-cell cluster. This provided (by the time of the study) a measure of validation of the different hybridisation technologies. The comparison between the Affymetrix GeneChip® high-density oligonucleotide arrays and the Lymphochip study gave a more extensive view in the clustering possibly due to the fact that the Affymetrix U95Av2 arrays contains

more than 12,000 transcripts and that these are short oligonucleotides and not represent full length clones as the cDNA Lymphochip does. Therefore, the use of functional annotation and validation of expression data by comparison of microarray analysis can be used as a basis for further biological data mining.

7 PAPER II

Lindvall J. M., Blomberg K. E. M., Berglöf A., Yang Q., Smith C. I. E. and Islam T. C.
Gene expression profile of B cells from *Xid* mice and Btk knockout mice.
Eur J Immunol. 2004 Jul;34(7):1981-91.

As a complement as well as a further investigation regarding the influence Btk has on the transcriptional level (Paper I), we analysed primary splenic B-cells from *Xid* and Btk KO mice and compared those with normal splenic B-cells from healthy mice bred on the same genetic background. The *Xid* mice have a missense mutation altering a conserved arginine residue within the Btk PH domain (R28C) (Rawlings et al., 1993; Thomas et al., 1993) whereas the Btk KO mice totally lack the *Btk* gene (Hendriks et al., 1996; Kerner et al., 1995; Khan et al., 1995). Even though the *Xid* mouse produces Btk protein (while to a lesser extent than the normal) it is non-functional in the sense of signal transduction in B-cells and function in the B-lymphocyte developmental pathway. The Btk KO mouse on the other hand is not translating the Btk protein at all.

By the use of Affymetrix mouse U74Av2 GeneChip® arrays, containing more than 12,000 mouse transcripts, we investigated the gene expression profile from healthy (CBA), *Xid* and Btk KO mice. The experiment was done in duplicate. A gene-average of 49 % was expressed in all samples and 4515 transcripts (corresponding to 36 % of all genes found on the array) were expressed in duplicate experiment. Of these 4515 transcripts, 38 probe-sets were found to be differentially expressed in the two Btk defective strains compared to wild type. Although it should be emphasised that the majority of genes behaved similar in the two Btk defective strains (maybe due to phenotype similarity), the expression of seven genes were found to differ with a factor of two or more between the *Xid* and Btk KO mouse.

One of the potentially most interesting findings in this study was the pronounced down-regulation (with a factor of 27) of the gene, microsomal epoxide hydrolase (mEH). As the paper describes, this gene could be of potential interest since the epileptic drug phenytoin, which can

cause IgA deficiency, is metabolised by mEH. Deficient detoxification of endogenous meta-bolites could potentially impair B-lymphocyte development. Irrespective of any potential mechanism of action, a direct causal relationship was excluded by the study of mice with a defective gene for mEH, since they did not show an *Xid* phenotype. However, the promoter of this gene is of great interest, since it seems to be highly regulated by Btk signalling. Another gene of potential interest was the Ionized Ca^{2+} type 1 gene (Iba1), which by *in silico* analysis was found to carry the binding site for the transcription factor PU.1 in its promoter. PU.1 is known to regulate Btk (Muller et al., 1996) but is not found here to be either down- or up-regulated and therefore the underlying mechanisms for the enhanced Iba1 expression is not known. On the other hand, another PU.1 related member, Spi-C, is increased in both Xid and Btk KO mice.

In order to get more insight into the biological behaviour of some of the genes found to be differentially expressed in the Btk defective cells, we also conducted quantitative Real-Time PCR on splenic B-cells stimulated with anti-IgM (a BCR activating agent), Phorbol Myristate Acetate (PMA)+ionomycin (which activates the PKC pathway) or LPS (which triggers the Toll-like receptor 4). Ten genes from our array experiment and an additional five genes from another microarray study (Fruman et al., 2000) were selected for the Real-Time PCR experiment. The five additional genes were not found to be differentially expressed in our experiment but were chosen, since they were found to be regulated in response to anti-IgM stimulation in normal splenic B-lymphocytes. Even though significant differences, i.e. resulting in an up- or down-regulation, were found between normal B-cells and Btk defective, the expression patterns were highly related in the three mouse strains. In general, the anti-IgM and PMA+ionomycin stimulations exhibited more similar responses compared to the LPS stimulus. Thus, the study reveals differences on the transcriptome level between both the Btk defective mouse models compared to the normal control as well as few, but significant, differences between the *Xid* versus the Btk KO mice.

8 PAPER III

Lindvall J. M., Blomberg K. E. M., Wennborg A. and Smith C. I. E.

Differential expression and molecular characterisation of *Lmo7*, *Myo1e*, *Sash1* and *Mcoln2* genes in Btk-defective B-cells.

Many of the ESTs are not (yet) linked to any characterised gene. In a microarray experiment, especially where oligonucleotides have been used, there are many target sequences that are of EST character. The ESTs have been widely used for gene discovery (Boguski et al., 1994), which extends beyond mammals (Ajioka et al., 1998). ESTs are small pieces of a DNA sequence (usually 200 to 500 nucleotides long) that are generated by sequencing either one or both ends of an expressed gene (in the Affymetrix GeneChip$^{\circledR}$ arrays they are usually from the 3' end). Because these ESTs (as well as the represented full-length genes) used on the Affymetrix GeneChip$^{\circledR}$ are generated from the 3' end of a transcript, they are likely to fall within non-coding, or untranslated regions. Therefore they tend to exhibit less cross-species conservation than do coding sequences. The idea of ESTs is to sequence bits of DNA that represent genes expressed in certain cells, tissues, or organs from different organisms and use these "tags" to fish a gene out of a portion of chromosomal DNA by matching base pairs. ESTs are generated by the sequencing of cDNA.

In Paper II we reported on a set of mouse genes that were found to be differentially expressed between Btk defective B-lymphocytes compared to normal healthy controls. Among the genes were four highly regulated (with a factor of more than 4-fold) ESTs. These four ESTs were in this study annotated and further characterised by the use of *in silico* tools. Since the essential information known about each EST was its probe target sequence provided by Affymetrix, this was extracted as an unbiased starting point for a gene identification effort. Initial BLAST searches were conducted against the whole mouse genome and/or EST databases using both the Ensembl and the NCBI databases. The databases were run in parallel due to their variations in

the assembly of the genome and possible local divergences related to this. For transcript A and B, referred to as Lim domain only 7 (*Lmo7*) and Myosin 1e (*Myo1e*), respectively, where no overlap with any known gene was found, we successfully elongated the existing gene (located in close proximity to the target sequence) by the use of the sequence assembly program, CAP3. CAP3 is used to accumulate a putative full-length sequence from shorter sequence stretches. Target sequences for Gene C and D were perfectly mapped to the genes SAM and SH3 domain containing 1 (*Sash1*) and Mucolipin2 (*Mcoln2*), respectively, on the initial BLAST search. Further annotations and characterisation of the identified sequences were done by the use of several protein databases and alignment programs.

In addition to this, data mining on two additional B-cell expression data sets was bioinformatically challenged in order to further investigate the four transcripts expression pattern. Neither of the predicted transcripts was found to be expressed at the pro-B-cell stage. A bit surprisingly is that neither *Myo1e* nor *Sash1* were found to be expressed at any of the four B-cell stages investigated (pro-B-, pre-B, mature B-cell and plasma cell). A possible explanation could be that the investigated samples in the particular array study are of cell line origin and therefore differ from the primary cells investigated in Paper II. *Lmo7* was found to be expressed only at the mature B-cell stage whereas *Mcoln2* was found at the pre-B-cell stage and forward. We also looked at the expression pattern for the four transcripts from our T1 B-cell microarray data (Paper IV) and found similar regulation of the transcripts found in the whole B-cell splenocyte population (Paper II).

To biologically characterise these four genes, semi-quantitative RT-PCR following stimulations with anti-IgM or PMA+ionomycin were conducted on whole splenic B-lymphocytes from Btk KO and wild type CBA mice as well as in a range of unstimulated mouse cell lines. Also, to ascertain that the different cell lines investigated in fact expressed the *in silico* predicted full-length transcripts, PCR amplification over several exons was performed. The results indicated that we indeed identified the correct gene product.

9 PAPER IV

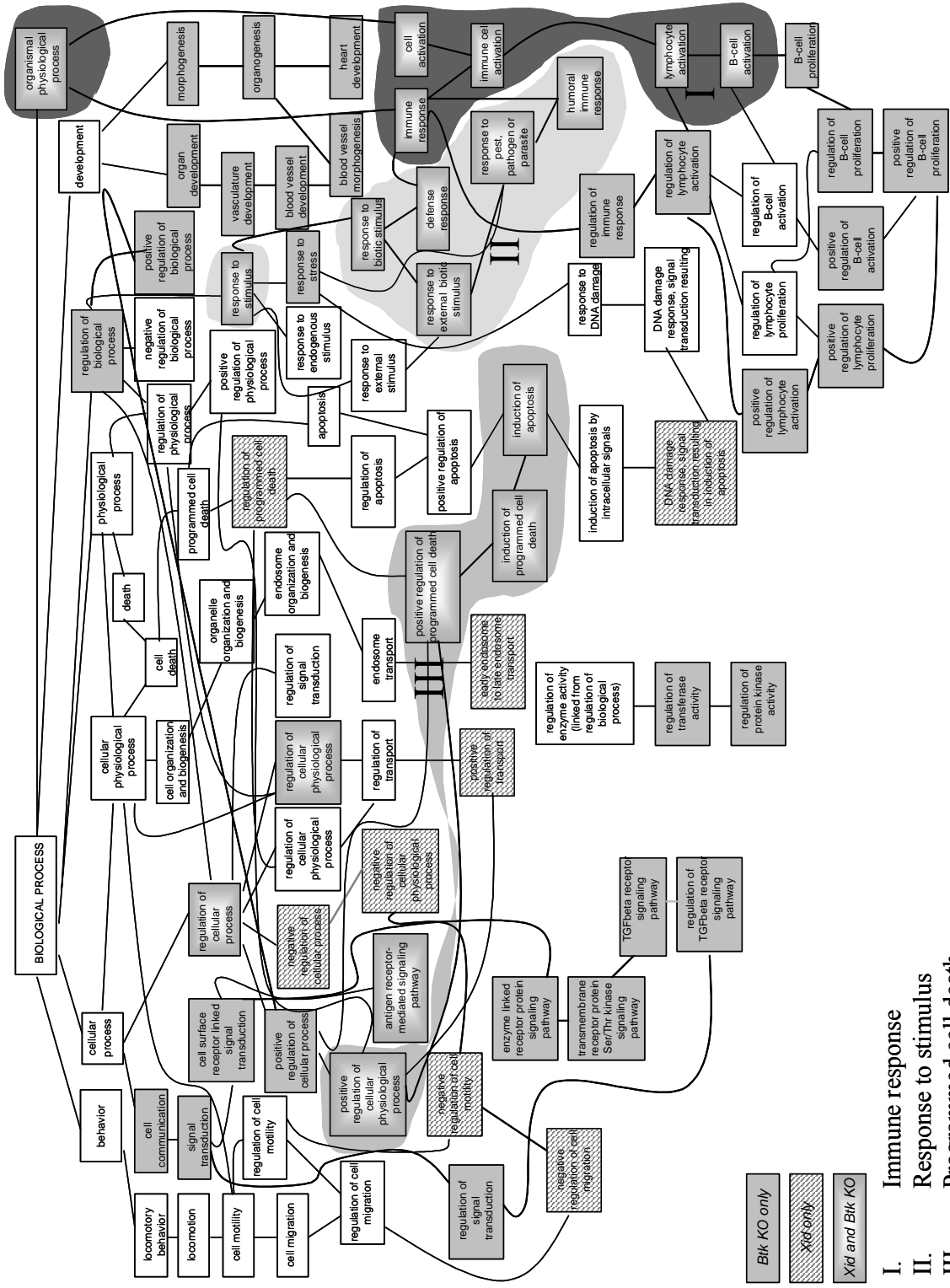**Lindvall J. M.**, Blomberg K. E. M., Berglöf A. and Smith C. I. E.
Distinct gene expression profile in Btk defective Transitional 1 B-lymphocytes.
Manuscript

Since *Btk* is expressed at all stages of the B-lymphocyte development (de Weers et al., 1993), the possibility that consequences of Btk signalling defects also are manifested at an earlier stage of B-cell development in either bone marrow or at the transitional stage exist. This possibility was investigated here, where we have purified T1 splenic B-cells from one-week-old *Xid* and Btk KO mice and compared to healthy wild-type CBA mice using gene expression profiling. There have been studies examining this possibility previously, but these have been restricted to investigating surface markers on respective B-cell stage as well as counting cell numbers and have therefore not examined the transcriptional pattern. Also, these studies have only examined the *Xid* mouse compared to healthy controls. As a consequence, the Btk KO mouse has not been considered having a different outcome than the *Xid* mouse. The results from these studies have placed the B-lymphocyte developmental block at the T2 stage in *Xid* mice (Loder et al., 1999; Petro et al., 2002; Su and Rawlings, 2002). T1 B-cells express a distinct set of surface markers and are located in the bone marrow, blood and spleen.

We took the advantage of the Affymetrix Mouse Genome 430 2.0 GeneChip$^{®}$ arrays covering the full mouse genome and a duplicate of each mouse group was run. We realise that a duplicate run is not sufficient for any statistical analysis, but since there are at least seven mice per group we believe to have a normal distribution. Also, replicates of this first set of experiments are under way. The initial statistical analysis using a one-way ANOVA test, with mouse strain as the factor to consider, together with the Tukey-Kramer test, revealed a significant difference between the *Xid* and Btk KO mouse group as well as between the Btk KO and CBA strain. On the other hand, and a bit surprisingly, no significant difference was found between the *Xid* and CBA mouse group. In order to exclude that this is due to the limited number of animals analysed so far, new sets of mice are currently being studied. When performing the comparative

Figure 6. *GOTM result.* Enriched Gene Ontology categories overlapping the two Btk defective mouse strains.

I.   Immune response
II.   Response to stimulus
III. Programmed cell death

Btk KO only

xid-only

Xid and Btk KO

analysis for *Xid* versus CBA and Btk KO versus CBA, we find 432 and 207 transcripts being differentially regulated, respectively. A GO analysis of the outcome was performed using GOTM, identifying categories over-represented among the transcripts belonging to either the *Xid* versus CBA or Btk KO versus CBA comparison. Figure 6 illustrates the results from the GO functional analysis. Amongst the groups overlapping between the two comparisons are *Immune response*, *Programmed cell death* and *Response to stimulus*.

In order to identify transcripts correlated with a particular class distinction the software program GeneCluster 2.0 was used. Interestingly, mEH was found amongst the genes distinguishing *Xid* from CBA. mEH was not found in the analysis between Btk KO and CBA but amongst the forty predicted transcripts, ten overlapped with the marker analysis done for *Xid* and CBA. Overall the distinction between the Btk defective mouse strains and normal control was very clear indicating that there are large differences between them. Applying un-supervised hierarchical clustering on the six GeneChips® (two arrays per mouse group) divided the mouse groups according to their strains giving us confidence in the duplicate experiment. The clustering also confirmed the statistical finding that *Xid* is more similar to the CBA strain compared to the Btk KO. Although, differences between all three mouse strains were seen (Figure 7).
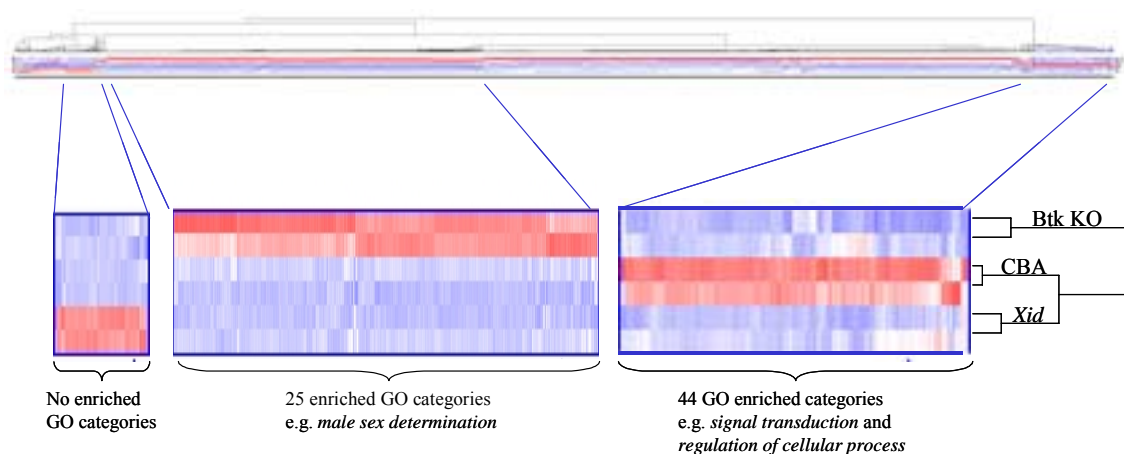

No enriched GO categories

25 enriched GO categories
e.g. *male sex determination*

44 GO enriched categories
e.g. *signal transduction* and *regulation of cellular process*

Btk KO

CBA

Xid

Figure 7. *Hierarchical clustering using dChip.* Highlighted are groups of transcripts identified by being characteristically ordered into a specific mouse strain. The different mouse groups are identified having their own typical gene pattern.

10 GENERAL DISCUSSION

Signal transduction is a fundamental process that takes place in all living organisms and understanding how this event occurs at the cellular level is of vital importance to essentially any field of biomedicine. There are several major steps involved in the deciphering of the signalling pathways: (a) Which molecules are involved in signalling? (b) Who talks to whom? i.e. making sense of the molecular interactions in a context particular way. (c) Where are the signalling events taking place? e.g. when a resting cell becomes activated. The challenge lies in reconstructing signalling modules and networks evoked in particular response to a single input as well as correlating the signalling response to multiple different cellular inputs. In addition to this there is also the need for interpretation of cross-talk between signalling modules in response to single and multiple inputs. One possible way of answering these various types of questions is by the use of gene expression profiling in combination with statistics, bioinformatics and traditional biochemical methods. Gene expression profiling has emerged to be a popular and useful tool, it and has thus provided a wealth of information waiting to be fully understood. Microarray experiments are still very expensive for most academic laboratories (although the price has been reduced over the years) and this has forced research laboratories to decrease the number of experiments and replicates in the project. But as the amount of gene expression data increases, as more and more laboratories are making use of this technology, and become publicly available there is the possibility to gather other data sets and together with your own results do meta-analysis. Caution should be taken in a meta-analysis attempt, since cross-platform– as well as within-platform– analysis is not always as straightforward as it seems. Awareness regarding the probe design on the array, RNA preparation, hybridisation method, and normalisation process amongst other should be taken into consideration. Despite all this, I believe that meta-analysis is required for the future work in genomics.

As other research areas (e.g. different areas of lymphoma research) have successfully used the microarray technology in order to distinguish between the normal and the "sick" state of a cell, the XLA defect does not seem to have such a clear and distinctive effect to the transcriptomic pattern i.e. the normal and defective state can not be separated just by looking at the whole

expression profile. Instead important target genes needs to be "fished out" from the large set of data perhaps not by exercising clustering methods, but rather by the use of specialised *in silico* tools. The number of bioinformatic tools and databases are constantly increasing and the large amount of choices that the researcher then face when starting to analyse the data is enormous. Therefore, the hunt for the perfect bioinformatic tool sometimes feels like "a-needle-in-a-haystack" story. *In silico* studies have been useful for the identification of potentially important target genes in the Btk regulation. Having said this, I believe that one of the major pit-falls is to trust the *in silico* data too much. After all, the data drawn from various databases might not be correct since there is a problem with re-annotation and updating of results in many databases. Most of the information one can extract is based on "wet-lab" results from published work. But as we all know, all data might not reflect the reality and false data/information of this kind might very well be present in the database. "Wet-lab" is of course (still) very much needed. Whatever predictions one can do using the computer has to be confirmed biologically. Figure 8 reflects my view on how "wet-lab" and "dry-lab" research "go hand-in-hand".

This "wet-lab"–"dry-lab" dilemma is reflected in Paper III where I started the project with a "dry-lab" approach and quickly realised that I needed to both confirm the findings biologically as well as extend the biological questions by functionality ("wet-lab"). Therefore, other experiments are essential to address the normal role of these factors in B-cells as well as the potential role they play together with Btk. One line of approach, in order for a more accurate reflection of the XLA patients, is the use of a Btk/Tec double KO mouse model. Since these mice lack mature B-cells and therefore share a similar phenotype to the XLA patients such a study could be a useful complement to the studies done in this thesis. However, a potential caveat in the analysis of Btk deficiency is the fact that this abnormality causes impaired differentiation. Thus, apart from the initial signalling defect, once the impairment has manifested itself it will also secondarily affect developmentally regulated genes. Another characteristic to the transcriptomic comparisons made, is the large differences seen when analysing established cell lines in comparison to the primary stage cells. This might not be that unexpected, since cell lines after
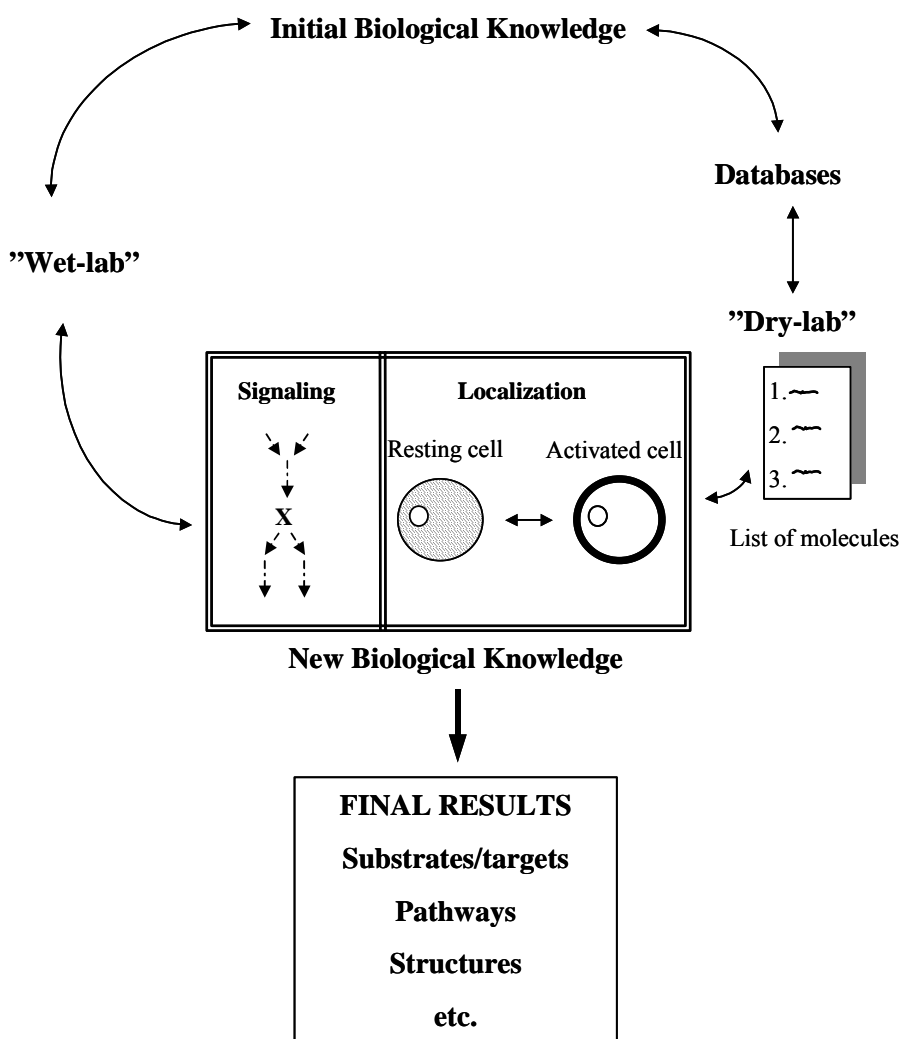
Brief Bioinform. 2003 Dec;4(4):315-24. Lindvall JM *et. al.*

Figure 8. *Schematic representation of the in silico signal transduction integration of information leading to the biological understanding of queried data.*

all are constantly dividing and activated by the surrounding medium, which then of course is reflected in the expression profile. Also, in our research, since the *Xid* and Btk KO mice have a milder phenotype compared to the XLA patients, this could perhaps explain the modest effect on the transcriptional level found in the expression study for the XLA B-cell lines.

Having discussed all of this, I have found tremendous potential in the results generated by this technology. Side by side with biochemistry, molecular biology and the advent of bioinformatics, statistical analytical methods and systems biology, together have significantly changed the platform of mainstream knowledge of biology and medicine. This evolution has set biological informations to unlimited horizons and openings for the future understanding of the function of our genes.

11 CLOSING REMARKS

With the use of gene expression profiling and applied bioinformatics, the transcriptional regulation of Btk has been studied. The thesis projects cover studies done on both XLA patient material and B-lymphocytes from *Xid* and Btk KO mouse models. Thus, both the aspect of stimulated EBV-transformed B-cells from humans and primary resting splenic B-lymphocytes from immunocompetent mice has been considered. Attempts to answer questions such as: How does the transcriptional pattern change when having a Btk defect? or Can we find the transcriptional targets of Btk?. Questions have been answered, but new ones have been brought into light and these needs to be solved with further experiments. For instance as we now have data from both cell lines and primary resting cells, comparison analyses should be conducted in an ordered fashion. This could potentially give more insight to the role of Btk. The findings from my thesis studies have (hopefully) contributed to the further knowledge of Btk as a key regulator of the B-lymphocyte development.

12 POPULÄRVETENSKAPLIG SAMMANFATTNING

Till min familj, vänner samt andra som under alla år inte riktigt har förstått vad det är jag har arbetat med. Jag hoppas att denna sammanfattning får det att klarna lite...

Olika sjukdomar samt yttre (och inre) påverkan kan förändra vår arvsmassa (gener) på andra sätt än vad som normalt ska ske. Kunskap om vad som förändras, är till stor hjälp vid t.ex. läkemedelsutveckling och diagnostik av olika sjukdomar. Vita blodkroppar, som är livsviktiga för vår antikroppsproduktion, innehåller många livsviktiga komponenter/molekyler som kan bli förstörda pga. sjukdomar eller förändringar i arvsmassan. X-länkad agammaglobulinemia (XLA) är en sådan sjukdom där genen för ett enzym kallat Brutons Tyrosin Kinas (*BTK*) är förändrad. XLA drabbar bara pojkar, som har en X och en Y kromosom. Kvinnor har två X kromosomer och drabbas då inte av XLA då en frisk X kromosom uppväger den andra sjuka. Sjukdomen kan gå i arv från mamma till gossebarn. Patienter med XLA lider av infektioner då antikroppar inte produceras som skydd. Hur Btk fungerar är därför av stor betydelse och det är det här som jag har studerat under mina doktorandår.

Genexpressionsprofilering är en teknik som ger möjlighet till att studera hela arvsmassan i ett försök på en och samma gång. På så sätt kan man undersöka den påverkan av arvsmassan som t.ex. förändringar av *BTK* resulterar i. Jag har i min avhandling använt mig av denna teknik för att kunna jämföra skillnaderna mellan XLA patienter och normala, friska individer. Den här metoden tillsammans med s.k. bioinformatik där man använder sig av olika biologiska databaser har gett mig möjligheten att utforska funktionen av *BTK* i vita blodkroppar. Musmodeller är ett annat mycket användbart redskap när man vill undersöka orsakerna till en sjukdom. Eftersom arvsmassan mellan möss och människor är väldigt lika så är musmodeller till stor hjälp när man vill undersöka orsakerna till en viss sjukdom. *Xid* är musens motsvarigheten till XLA men med en mildare sjukdomsbild. *Xid* möss har tillsammans med Btk Knock Out möss, där *Btk* genen har blivit borttagen (utknockad) använts i forskningen kring XLA.

13 ACKNOWLEDGEMENTS

This work has been performed at the Department of Laboratory Medicine, Clinical Research Center as well as at the Department of Biosciences at Novum. I am indebted to many persons who all made this thesis possible. First I would like to express my sincere gratitude to my supervisors

Professor **Edvard Smith**, a wonderful supervisor and mentor. Your broad knowledge has provided me with a mind beyond signal transduction and gene expression. Thank you for giving me the freedom to develop into an independent scientist. Also, thank you for making me laugh when it has felt like "five steps back, two steps forward" (even if you sometimes do have a somewhat strange sense of humour ; o)

Dr. **Tahmina Islam**, a warm and excellent mentor, for accepting me as a student in her projects and leading me through the tough first years of research. Thank you for introducing me to the field of bioinformatics and gene expression profiling. I will always be grateful…

and also all my past and present co-workers, especially

The whole of **MCG group** and then particularly **Lotta Asplund** for being "min solstråle på lab" and dear friend also outside of work. **Anna Berglöf** for wonderful little times, especially in Keystone. Good luck with your "pensel". **Emelie Blomberg**, my "side-kick" in the projects and for helping me to remember the details. **Magnus Bäckesjö** for being such a helpful and super nice person. **Abdalla (Abdi) Mohammed** for your wonderful support and providing me with "wienerbröd" when needed. **Beston Nore** for believing in me as a scientist. **Leonardo Vargas** for being a wonderful friend and for all the nice discussions and "fika-stunder" at Sandy´s. "Ler- och långhalm", **Mathias Svahn** and **Oscar Simonson**, for just being you.

and

mormor, som alltid trodde på mig. Jag saknar dig så otroligt mycket. **Morfar**, du skulle ha varit stolt över mig. Hoppas ni sitter på ett moln och ser mig...

mamma, du är min hjälte och förebild. **Pappa**, du har alltid fått mig att se mina möjligheter istället för mina brister. Tack för ert stöd och er kärlek, Jag älskar er sååååå mycket,

lillebror (Christer), för att du är den underbaraste bror som finns och som alltid finns där för sin "yster" när hon behöver det,

Peter, mitt livs kärlek, tack för att du hjälper mig att hitta lättare och mer praktiska lösningar när jag tenderar att komplicera saker och ting. Vad gjorde jag utan dig…

14 REFERENCES

Aderem, A. (2005). Systems biology: its practice and challenges. Cell *121*, 511-513.

Ajioka, J. W., Boothroyd, J. C., Brunk, B. P., Hehl, A., Hillier, L., Manger, I. D., Marra, M., Overton, G. C., Roos, D. S., Wan, K. L., *et al.* (1998). Gene discovery by EST sequencing in Toxoplasma gondii reveals sequences restricted to the Apicomplexa. Genome Res *8*, 18-28.

Alizadeh, A. A., Eisen, M. B., Davis, R. E., Ma, C., Lossos, I. S., Rosenwald, A., Boldrick, J. C., Sabet, H., Tran, T., Yu, X., *et al.* (2000). Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature *403*, 503-511.

Antonarakis, S. E., and McKusick, V. A. (2000). OMIM passes the 1,000-disease-gene mark. Nat Genet *25*, 11.

Antony, P., Petro, J. B., Carlesso, G., Shinners, N. P., Lowe, J., and Khan, W. N. (2003). B cell receptor directs the activation of NFAT and NF-kappaB via distinct molecular mechanisms. Exp Cell Res *291*, 11-24.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., *et al.* (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet *25*, 25-29.

Baba, Y., Nonoyama, S., Matsushita, M., Yamadori, T., Hashimoto, S., Imai, K., Arai, S., Kunikata, T., Kurimoto, M., Kurosaki, T., *et al.* (1999). Involvement of wiskott-aldrich syndrome protein in B-cell cytoplasmic tyrosine kinase pathway. Blood *93*, 2003-2012.

Backesjo, C. M., Vargas, L., Superti-Furga, G., and Smith, C. I. E. (2002). Phosphorylation of Bruton's tyrosine kinase by c-Abl. Biochem Biophys Res Commun *299*, 510-515.

Bajpai, U. D., Zhang, K., Teutsch, M., Sen, R., and Wortis, H. H. (2000). Bruton's tyrosine kinase links the B cell receptor to nuclear factor kappaB activation. J Exp Med *191*, 1735-1744.

Baumforth, K. R., Young, L. S., Flavell, K. J., Constandinou, C., and Murray, P. G. (1999). The Epstein-Barr virus and its association with human cancers. Mol Pathol *52*, 307-322.

Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., *et al.* (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res *31*, 365-370.

Boguski, M. S., Tolstoshev, C. M., and Bassett, D. E., Jr. (1994). Gene discovery in dbEST. Science *265*, 1993-1994.

Borst, J., Brouns, G. S., de Vries, E., Verschuren, M. C., Mason, D. Y., and van Dongen, J. J. (1993). Antigen receptors on T and B lymphocytes: parallels in organization and function. Immunol Rev *132*, 49-84.

Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C. A., Causton, H. C.*, et al.* (2001). Minimum information about a microarray experiment (MIAME)-toward standards for microarray data. Nat Genet *29*, 365-371.

Bruton, O. C. (1952). Agammaglobulinemia. Pediatrics *9*, 722-727.

Burkhardt, A. L., Brunswick, M., Bolen, J. B., and Mond, J. J. (1991). Anti-immunoglobulin stimulation of B lymphocytes activates src-related protein-tyrosine kinases. Proc Natl Acad Sci U S A *88*, 7410-7414.

Cambier, J. C., Pleiman, C. M., and Clark, M. R. (1994). Signal transduction by the B cell antigen receptor and its coreceptors. Annu Rev Immunol *12*, 457-486.

Campana, D., Farrant, J., Inamdar, N., Webster, A. D., and Janossy, G. (1990). Phenotypic features and proliferative activity of B cell progenitors in X-linked agammaglobulinemia. J Immunol *145*, 1675-1680.

Chee, M., Yang, R., Hubbell, E., Berno, A., Huang, X. C., Stern, D., Winkler, J., Lockhart, D. J., Morris, M. S., and Fodor, S. P. (1996). Accessing genetic information with high-density DNA arrays. Science *274*, 610-614.

Chen, N., Harris, T. W., Antoshechkin, I., Bastiani, C., Bieri, T., Blasiar, D., Bradnam, K., Canaran, P., Chan, J., Chen, C. K.*, et al.* (2005). WormBase: a comprehensive data resource for Caenorhabditis biology and genomics. Nucleic Acids Res *33*, D383-389.

Cheung, V. G., Conlin, L. K., Weber, T. M., Arcaro, M., Jen, K. Y., Morley, M., and Spielman, R. S. (2003). Natural variation in human gene expression assessed in lymphoblastoid cells. Nat Genet *33*, 422-425.

Cheung, V. G., and Spielman, R. S. (2002). The genetics of variation in gene expression. Nat Genet *32 Suppl*, 522-525.

Clarke, P. A., Poele, R. T., and Workman, P. (2004). Gene expression microarray technologies in the development of new therapeutic agents. Eur J Cancer *40*, 2560-2591.

Consortium, T. I. H. G. S. (2004). Finishing the euchromatic sequence of the human genome. Nature *431*, 931-945.

de la Cruz, N., Bromberg, S., Pasko, D., Shimoyama, M., Twigger, S., Chen, J., Chen, C. F., Fan, C., Foote, C., Gopinath, G. R.*, et al.* (2005). The Rat Genome Database (RGD): developments towards a phenome database. Nucleic Acids Res *33*, D485-491.

de Weers, M., Verschuren, M. C., Kraakman, M. E., Mensink, R. G., Schuurman, R. K., van Dongen, J. J., and Hendriks, R. W. (1993). The Bruton's tyrosine kinase gene is expressed throughout B cell differentiation, from early precursor B cell stages preceding immunoglobulin gene rearrangement up to mature B cell stages. Eur J Immunol *23*, 3109-3114.

Ding, S. Z., Torok, A. M., Smith, M. F., Jr., and Goldberg, J. B. (2005). Toll-like receptor 2-mediated gene expression in epithelial cells during Helicobacter pylori infection. Helicobacter *10*, 193-204.

Ehrenberg, M., Elf, J., Aurell, E., Sandberg, R., and Tegner, J. (2003). Systems biology is taking off. Genome Res *13*, 2377-2380.

Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A *95*, 14863-14868.

Eppig, J. T., Bult, C. J., Kadin, J. A., Richardson, J. E., Blake, J. A., Anagnostopoulos, A., Baldarelli, R. M., Baya, M., Beal, J. S., Bello, S. M., *et al.* (2005). The Mouse Genome Database (MGD): from genes to mice--a community resource for mouse biology. Nucleic Acids Res *33*, D471-475.

Falt, S., Merup, M., Gahrton, G., Lambert, B., and Wennborg, A. (2005). Identification of progression markers in B-CLL by gene expression profiling. Exp Hematol *33*, 883-893.

Fluckiger, A. C., Li, Z., Kato, R. M., Wahl, M. I., Ochs, H. D., Longnecker, R., Kinet, J. P., Witte, O. N., Scharenberg, A. M., and Rawlings, D. J. (1998). Btk/Tec kinases regulate sustained increases in intracellular Ca2+ following B-cell receptor activation. Embo J *17*, 1973-1985.

FlyBaseConsortium (2003). The FlyBase database of the Drosophila genome projects and community literature. Nucleic Acids Res *31*, 172-175.

Fodor, S. P., Read, J. L., Pirrung, M. C., Stryer, L., Lu, A. T., and Solas, D. (1991). Light-directed, spatially addressable parallel chemical synthesis. Science *251*, 767-773.

Fruman, D. A., Satterthwaite, A. B., and Witte, O. N. (2000). Xid-like phenotypes: a B cell signalosome takes shape. Immunity *13*, 1-3.

Gauld, S. B., Dal Porto, J. M., and Cambier, J. C. (2002). B cell antigen receptor signaling: roles in cell development and disease. Science *296*, 1641-1642.

Genevier, H. C., and Callard, R. E. (1997). Impaired Ca2+ mobilization by X-linked agammaglobulinaemia (XLA) B cells in response to ligation of the B cell receptor (BCR). Clin Exp Immunol *110*, 386-391.

Guo, B., Kato, R. M., Garcia-Lloret, M., Wahl, M. I., and Rawlings, D. J. (2000). Engagement of the human pre-B cell receptor generates a lipid raft-dependent calcium signaling complex. Immunity *13*, 243-253.

Hardy, R. R., Hayakawa, K., Parks, D. R., and Herzenberg, L. A. (1984). Murine B cell differentiation lineages. J Exp Med *159*, 1169-1188.

Hendriks, R. W., de Bruijn, M. F., Maas, A., Dingjan, G. M., Karis, A., and Grosveld, F. (1996). Inactivation of Btk by insertion of lacZ reveals defects in B cell development only past the pre-B cell stage. Embo J *15*, 4862-4872.

Horwood, N. J., Mahon, T., McDaid, J. P., Campbell, J., Mano, H., Brennan, F. M., Webster, D., and Foxwell, B. M. (2003). Bruton's tyrosine kinase is required for lipopolysaccharide-induced tumor necrosis factor alpha production. J Exp Med *197*, 1603-1611.

Hubbard, T., Andrews, D., Caccamo, M., Cameron, G., Chen, Y., Clamp, M., Clarke, L., Coates, G., Cox, T., Cunningham, F., *et al.* (2005). Ensembl 2005. Nucleic Acids Res *33 Database Issue*, D447-453.

Hutchcroft, J. E., Harrison, M. L., and Geahlen, R. L. (1992). Association of the 72-kDa protein-tyrosine kinase PTK72 with the B cell antigen receptor. J Biol Chem *267*, 8613-8619.

Inabe, K., Ishiai, M., Scharenberg, A. M., Freshney, N., Downward, J., and Kurosaki, T. (2002). Vav3 modulates B cell receptor responses by regulating phosphoinositide 3-kinase activation. J Exp Med *195*, 189-200.

Islam, T. C., Asplund, A. C., Lindvall, J. M., Nygren, L., Liden, J., Kimby, E., Christensson, B., Smith, C. I. E., and Sander, B. (2003). High level of cannabinoid receptor 1, absence of regulator of G protein signalling 13 and differential expression of Cyclin D1 in mantle cell lymphoma. Leukemia *17*, 1880-1890.

Islam, T. C., Lindvall, J., Wennborg, A., Branden, L. J., Rabbani, H., and Smith, C. I. E. (2002). Expression profiling in transformed human B cells: influence of Btk mutations and comparison to B cell lymphomas using filter and oligonucleotide arrays. Eur J Immunol *32*, 982-993.

Islam, T. C., and Smith, C. I. E. (2000). The cellular phenotype conditions Btk for cell survival or apoptosis signaling. Immunol Rev *178*, 49-63.

Iwatsuki, K., Xu, Z., Ohtsuka, M., and Kaneko, F. (2000). Cutaneous lymphoproliferative disorders associated with Epstein-Barr virus infection: a clinical overview. J Dermatol Sci *22*, 181-195.

Kendziorski, C. M., Zhang, Y., Lan, H., and Attie, A. D. (2003). The efficiency of pooling mRNA in microarray experiments. Biostatistics *4*, 465-477.

Kerner, J. D., Appleby, M. W., Mohr, R. N., Chien, S., Rawlings, D. J., Maliszewski, C. R., Witte, O. N., and Perlmutter, R. M. (1995). Impaired expansion of mouse B cell progenitors lacking Btk. Immunity *3*, 301-312.

Khan, W. N., Alt, F. W., Gerstein, R. M., Malynn, B. A., Larsson, I., Rathbun, G., Davidson, L., Muller, S., Kantor, A. B., Herzenberg, L. A., and et al. (1995). Defective B cell development and function in Btk-deficient mice. Immunity *3*, 283-299.

Kikuchi, Y., Hirano, M., Seto, M., and Takatsu, K. (2000). Identification and characterization of a molecule, BAM11, that associates with the pleckstrin homology domain of mouse Btk. Int Immunol *12*, 1397-1408.

Kim, S., and Park, Y. M. (2005). Specific gene expression patterns in liver cirrhosis. Biochem Biophys Res Commun *334*, 681-688.

Klaus, G. G., Holman, M., Johnson-Leger, C., Elgueta-Karstegl, C., and Atkins, C. (1997). A re-evaluation of the effects of X-linked immunodeficiency (xid) mutation on B cell differentiation and function in the mouse. Eur J Immunol *27*, 2749-2756.

Kurosaki, T. (1999). Genetic analysis of B cell antigen receptor signaling. Annu Rev Immunol *17*, 555-592.

Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001). Initial sequencing and analysis of the human genome. Nature *409*, 860-921.

Lederman, H. M., and Winkelstein, J. A. (1985). X-linked agammaglobulinemia: an analysis of 96 patients. Medicine (Baltimore) *64*, 145-156.

Levitt, D., Ochs, H., and Wedgwood, R. J. (1984). Epstein-Barr virus-induced lymphoblastoid cell lines derived from the peripheral blood of patients with X-linked agammaglobulinemia can secrete IgM. J Clin Immunol *4*, 143-150.

Li, C., and Wong, W. H. (2001). Model-based analysis of oligonucleotide arrays: expression index computation and outlier detection. Proc Natl Acad Sci U S A *98*, 31-36.

Lindvall, J., and Islam, T. C. (2002). Interaction of Btk and Akt in B cell signaling. Biochem Biophys Res Commun *293*, 1319-1326.

Lindvall, J. M., Blomberg, K. E. M., Berglof, A., Yang, Q., Smith, C. I. E., and Islam, T. C. (2004). Gene expression profile of B cells from Xid mice and Btk knockout mice. Eur J Immunol *34*, 1981-1991.

Lindvall, J. M., Blomberg, K. E. M., Väliaho, J., Vargas, L., Heinonen, J. E., Berglöf, A., Mohamed, A. J., Vihinen, M., and Smith, C. I. E. (2005). Bruton tyrosine kinase:cell biology,

sequence conservation, mutation spectrum, siRNA modifications and expression profiling. Immunol Rev *203*, 200-215.

Lipshutz, R. J., Fodor, S. P., Gingeras, T. R., and Lockhart, D. J. (1999). High density synthetic oligonucleotide arrays. Nat Genet *21*, 20-24.

Lockhart, D. J., Dong, H., Byrne, M. C., Follettie, M. T., Gallo, M. V., Chee, M. S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H., and Brown, E. L. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat Biotechnol *14*, 1675-1680.

Loder, F., Mutschler, B., Ray, R. J., Paige, C. J., Sideras, P., Torres, R., Lamers, M. C., and Carsetti, R. (1999). B cell development in the spleen takes place in discrete steps and is determined by the quality of B cell receptor-derived signals. J Exp Med *190*, 75-89.

Longnecker, R., and Miller, C. L. (1996). Regulation of Epstein-Barr virus latency by latent membrane protein 2. Trends Microbiol *4*, 38-42.

Lucas, J. A., Miller, A. T., Atherly, L. O., and Berg, L. J. (2003). The role of Tec family kinases in T cell development and function. Immunol Rev *191*, 119-138.

Lundholm, L., Moverare, S., Steffensen, K. R., Nilsson, M., Otsuki, M., Ohlsson, C., Gustafsson, J. A., and Dahlman-Wright, K. (2004). Gene expression profiling identifies liver X receptor alpha as an estrogen-regulated gene in mouse adipose tissue. J Mol Endocrinol *32*, 879-892.

Maglott, D., Ostell, J., Pruitt, K. D., and Tatusova, T. (2005). Entrez Gene: gene-centered information at NCBI. Nucleic Acids Res *33*, D54-58.

Mano, H. (1999). The Tec family protein-tyrosine kinases: a subset of kinases for a subset of signalings. Int J Hematol *69*, 6-12.

Mohamed, A. J., Nore, B. F., Christensson, B., and Smith, C. I. E. (1999). Signalling of Bruton's tyrosine kinase, Btk. Scand J Immunol *49*, 113-118.

Mosier, D. E., Zitron, I. M., Mond, J. J., Ahmed, A., Scher, I., and Paul, W. E. (1977). Surface immunoglobulin D as a functional receptor for a subclass of B lymphocytes. Immunol Rev *37*, 89-104.

Mulder, N. J., Apweiler, R., Attwood, T. K., Bairoch, A., Barrell, D., Bateman, A., Binns, D., Biswas, M., Bradley, P., Bork, P.*, et al.* (2003). The InterPro Database, 2003 brings increased coverage and new features. Nucleic Acids Res *31*, 315-318.

Muller, S., Sideras, P., Smith, C. I. E., and Xanthopoulos, K. G. (1996). Cell specific expression of human Bruton's agammaglobulinemia tyrosine kinase gene (Btk) is regulated by Sp1- and Spi-1/PU.1-family members. Oncogene *13*, 1955-1964.

Nikawa, T., Ishidoh, K., Hirasaka, K., Ishihara, I., Ikemoto, M., Kano, M., Kominami, E., Nonaka, I., Ogawa, T., Adams, G. R*., et al.* (2004). Skeletal muscle gene expression in space-flown rats. A mouse model of uterine leiomyosarcoma. Faseb J *18*, 522-524.

Noordzij, J. G., de Bruin-Versteeg, S., Comans-Bitter, W. M., Hartwig, N. G., Hendriks, R. W., de Groot, R., and van Dongen, J. J. (2002). Composition of precursor B-cell compartment in bone marrow from patients with X-linked agammaglobulinemia compared with healthy children. Pediatr Res *51*, 159-168.

Nore, B. F., Vargas, L., Mohamed, A. J., Branden, L. J., Backesjo, C. M., Islam, T. C., Mattsson, P. T., Hultenby, K., Christensson, B., and Smith, C. I. E. (2000). Redistribution of Bruton's tyrosine kinase by activation of phosphatidylinositol 3-kinase and Rho-family GTPases. Eur J Immunol *30*, 145-154.

Novina, C. D., Kumar, S., Bajpai, U., Cheriyath, V., Zhang, K., Pillai, S., Wortis, H. H., and Roy, A. L. (1999). Regulation of nuclear localization and transcriptional activity of TFII-I by Bruton's tyrosine kinase. Mol Cell Biol *19*, 5014-5024.

Ochs, H. D., and Smith, C. I. E. (1996). X-linked agammaglobulinemia. A clinical and molecular analysis. Medicine (Baltimore) *75*, 287-299.

Ochs, H. D., Smith, C. I. E., and Puck, J. M. (1998). Primary Immunodeficiency Diseases: A Molecular and Genetic Approach, Oxford University Press).

Oettinger, M. A., Schatz, D. G., Gorka, C., and Baltimore, D. (1990). RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. Science *248*, 1517-1523.

Park, H., Wahl, M. I., Afar, D. E., Turck, C. W., Rawlings, D. J., Tam, C., Scharenberg, A. M., Kinet, J. P., and Witte, O. N. (1996). Regulation of Btk function by a major autophosphorylation site within the SH3 domain. Immunity *4*, 515-525.

Peng, X., Wood, C. L., Blalock, E. M., Chen, K. C., Landfield, P. W., and Stromberg, A. J. (2003). Statistical implications of pooling RNA samples for microarray experiments. BMC Bioinformatics *4*, 26.

Petro, J. B., Gerstein, R. M., Lowe, J., Carter, R. S., Shinners, N., and Khan, W. N. (2002). Transitional type 1 and 2 B lymphocyte subsets are differentially responsive to antigen receptor signaling. J Biol Chem *277*, 48009-48019.

Petro, J. B., Rahman, S. M., Ballard, D. W., and Khan, W. N. (2000). Bruton's tyrosine kinase is required for activation of IkappaB kinase and nuclear factor kappaB in response to B cell receptor engagement. J Exp Med *191*, 1745-1754.

Qiu, Y., and Kung, H. J. (2000). Signaling network of the Btk family kinases. Oncogene *19*, 5651-5661.

Rajaiya, J., Hatfield, M., Nixon, J. C., Rawlings, D. J., and Webb, C. F. (2005). Bruton's tyrosine kinase regulates immunoglobulin promoter activation in association with the transcription factor Bright. Mol Cell Biol *25*, 2073-2084.

Rajeevan, M. S., Vernon, S. D., Taysavang, N., and Unger, E. R. (2001). Validation of array-based gene expression profiles by real-time (kinetic) RT-PCR. J Mol Diagn *3*, 26-31.

Rawlings, D. J., Saffran, D. C., Tsukada, S., Largaespada, D. A., Grimaldi, J. C., Cohen, L., Mohr, R. N., Bazan, J. F., Howard, M., Copeland, N. G., and et al. (1993). Mutation of unique region of Bruton's tyrosine kinase in immunodeficient XID mice. Science *261*, 358-361.

Rawlings, D. J., Scharenberg, A. M., Park, H., Wahl, M. I., Lin, S., Kato, R. M., Fluckiger, A. C., Witte, O. N., and Kinet, J. P. (1996). Activation of BTK by a phosphorylation mechanism initiated by SRC family kinases. Science *271*, 822-825.

Reth, M. (1989). Antigen receptor tail clue. Nature *338*, 383-384.

Reth, M. (1992). Antigen receptors on B lymphocytes. Annu Rev Immunol *10*, 97-121.

Reth, M., and Wienands, J. (1997). Initiation and processing of signals from the B cell antigen receptor. Annu Rev Immunol *15*, 453-479.

Rodriguez, R., Matsuda, M., Perisic, O., Bravo, J., Paul, A., Jones, N. P., Light, Y., Swann, K., Williams, R. L., and Katan, M. (2001). Tyrosine residues in phospholipase Cgamma 2 essential for the enzyme function in B-cell signaling. J Biol Chem *276*, 47982-47992.

Rosen, F. S., Cooper, M. D., and Wedgwood, R. J. (1984a). The primary immunodeficiencies (1). N Engl J Med *311*, 235-242.

Rosen, F. S., Cooper, M. D., and Wedgwood, R. J. (1984b). The primary immunodeficiencies. (2). N Engl J Med *311*, 300-310.

Salim, K., Bottomley, M. J., Querfurth, E., Zvelebil, M. J., Gout, I., Scaife, R., Margolis, R. L., Gigg, R., Smith, C. I. E., Driscoll, P. C., *et al.* (1996). Distinct specificity in the recognition of phosphoinositides by the pleckstrin homology domains of dynamin and Bruton's tyrosine kinase. Embo J *15*, 6241-6250.

Sarang, Z., Haig, Y., Hansson, A., Vondracek, M., Warngard, L., and Grafstrom, R. (2003). Microarray assessment of fibronectin, collagen and integrin expression and the role of fibronectin-collagen coating in the growth of normal, SV40 T-antigen-immortalised and malignant human oral keratinocytes. Altern Lab Anim *31*, 575-585.

Satterthwaite, A. B., Li, Z., and Witte, O. N. (1998). Btk function in B cell development and response. Semin Immunol *10*, 309-316.

Schadt, E. E., Li, C., Ellis, B., and Wong, W. H. (2001). Feature extraction and normalization algorithms for high-density oligonucleotide gene expression array data. J Cell Biochem Suppl *Suppl 37*, 120-125.

Scharenberg, A. M., El-Hillal, O., Fruman, D. A., Beitz, L. O., Li, Z., Lin, S., Gout, I., Cantley, L. C., Rawlings, D. J., and Kinet, J. P. (1998). Phosphatidylinositol-3,4,5-trisphosphate (PtdIns-3,4,5-P3)/Tec kinase-dependent calcium signaling pathway: a target for SHIP-mediated inhibitory signals. Embo J *17*, 1961-1972.

Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science *270*, 467-470.

Scher, I., Ahmed, A., Strong, D. M., Steinberg, A. D., and Paul, W. E. (1975a). X-linked B-lymphocyte immune defect in CBA/HN mice. I. Studies of the function and composition of spleen cells. J Exp Med *141*, 788-803.

Scher, I., Sharrow, S. O., and Paul, W. E. (1976). X-linked B-lymphocyte defect in CBA/N mice. III. Abnormal development of B-lymphocyte populations defined by their density of surface immunoglobulin. J Exp Med *144*, 507-518.

Scher, I., Steinberg, A. D., Berning, A. K., and Paul, W. E. (1975b). X-linked B-lymphocyte immune defect in CBA/N mice. II. Studies of the mechanisms underlying the immune defect. J Exp Med *142*, 637-650.

Shaffer, A. L., Rosenwald, A., Hurt, E. M., Giltnane, J. M., Lam, L. T., Pickeral, O. K., and Staudt, L. M. (2001). Signatures of the immune response. Immunity *15*, 375-385.

Sideras, P., Muller, S., Shiels, H., Jin, H., Khan, W. N., Nilsson, L., Parkinson, E., Thomas, J. D., Branden, L., Larsson, I., and et al. (1994). Genomic organization of mouse and human Bruton's agammaglobulinemia tyrosine kinase (Btk) loci. J Immunol *153*, 5607-5617.

Sideras, P., and Smith, C. I. E. (1995). Molecular and cellular aspects of X-linked agammaglobulinemia. Adv Immunol *59*, 135-223.

Sieckmann, D. G., Scher, I., Asofsky, R., Mosier, D. E., and Paul, W. E. (1978). Activation of mouse lymphocytes by anti-immunoglobulin. II. A thymus-independent response by a mature subset of B lymphocytes. J Exp Med *148*, 1628-1643.

Smith, C. I. E., Backesjo, C. M., Berglof, A., Branden, L. J., Islam, T., Mattsson, P. T., Mohamed, A. J., Muller, S., Nore, B., and Vihinen, M. (1998). X-linked agammaglobulinemia: lack of mature B lineage cells caused by mutations in the Btk kinase. Springer Semin Immunopathol *19*, 369-381.

Smith, C. I. E., Baskin, B., Humire-Greiff, P., Zhou, J. N., Olsson, P. G., Maniar, H. S., Kjellen, P., Lambris, J. D., Christensson, B., Hammarstrom, L., and et al. (1994). Expression of Bruton's

agammaglobulinemia tyrosine kinase gene, BTK, is selectively down-regulated in T lymphocytes and plasma cells. J Immunol *152*, 557-565.

Smith, C. I. E., Islam, T. C., Mattsson, P. T., Mohamed, A. J., Nore, B. F., and Vihinen, M. (2001). The Tec family of cytoplasmic tyrosine kinases: mammalian Btk, Bmx, Itk, Tec, Txk and homologs in other species. Bioessays *23*, 436-446.

Smith, C. I. E., and Notarangelo, L. D. (1997). Molecular basis for X-linked immunodeficiencies. Adv Genet *35*, 57-115.

Su, T. T., and Rawlings, D. J. (2002). Transitional B lymphocyte subsets operate as distinct checkpoints in murine splenic B cell development. J Immunol *168*, 2101-2110.

Takata, M., and Kurosaki, T. (1996). A role for Bruton's tyrosine kinase in B cell antigen receptor-mediated activation of phospholipase C-gamma 2. J Exp Med *184*, 31-40.

Tamayo, P., Slonim, D., Mesirov, J., Zhu, Q., Kitareewan, S., Dmitrovsky, E., Lander, E. S., and Golub, T. R. (1999). Interpreting patterns of gene expression with self-organizing maps: methods and application to hematopoietic differentiation. Proc Natl Acad Sci U S A *96*, 2907-2912.

Thomas, J. D., Sideras, P., Smith, C. I. E., Vorechovsky, I., Chapman, V., and Paul, W. E. (1993). Colocalization of X-linked agammaglobulinemia and X-linked immunodeficiency genes. Science *261*, 355-358.

Tsapogas, P., Breslin, T., Bilke, S., Lagergren, A., Mansson, R., Liberg, D., Peterson, C., and Sigvardsson, M. (2003). RNA analysis of B cell lines arrested at defined stages of differentiation allows for an approximation of gene expression patterns during B cell development. J Leukoc Biol *74*, 102-110.

Tsukada, S., Saffran, D. C., Rawlings, D. J., Parolini, O., Allen, R. C., Klisak, I., Sparkes, R. S., Kubagawa, H., Mohandas, T., Quan, S., and et al. (1993). Deficient expression of a B cell cytoplasmic tyrosine kinase in human X-linked agammaglobulinemia. Cell *72*, 279-290.

Vargas, L., Nore, B. F., Berglof, A., Heinonen, J. E., Mattsson, P. T., Smith, C. I. E., and Mohamed, A. J. (2002). Functional interaction of caveolin-1 with Bruton's tyrosine kinase and Bmx. J Biol Chem *277*, 9351-9357.

Varnai, P., Rother, K. I., and Balla, T. (1999). Phosphatidylinositol 3-kinase-dependent membrane association of the Bruton's tyrosine kinase pleckstrin homology domain visualized in single living cells. J Biol Chem *274*, 10983-10989.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A.*, et al.* (2001). The sequence of the human genome. Science *291*, 1304-1351.

Vetrie, D., Vorechovsky, I., Sideras, P., Holland, J., Davies, A., Flinter, F., Hammarstrom, L., Kinnon, C., Levinsky, R., Bobrow, M., and et al. (1993). The gene involved in X-linked agammaglobulinaemia is a member of the src family of protein-tyrosine kinases. Nature *361*, 226-233.

Vihinen, M. (2003). Signal transduction-related bioinformatics services. Brief Bioinform *4*, 325-331.

Vihinen, M., Belohradsky, B. H., Haire, R. N., Holinski-Feder, E., Kwan, S. P., Lappalainen, I., Lehvaslaiho, H., Lester, T., Meindl, A., Ochs, H. D.*, et al.* (1997). BTKbase, mutation database for X-linked agammaglobulinemia (XLA). Nucleic Acids Res *25*, 166-171.

Vihinen, M., Cooper, M. D., de Saint Basile, G., Fischer, A., Good, R. A., Hendriks, R. W., Kinnon, C., Kwan, S. P., Litman, G. W., Notarangelo, L. D., and et al. (1995). BTKbase: a database of XLA-causing mutations. International Study Group. Immunol Today *16*, 460-465.

Vihinen, M., Kwan, S. P., Lester, T., Ochs, H. D., Resnick, I., Valiaho, J., Conley, M. E., and Smith, C. I. E. (1999). Mutations of the human BTK gene coding for bruton tyrosine kinase in X-linked agammaglobulinemia. Hum Mutat *13*, 280-285.

Villeneuve, D. J., and Parissenti, A. M. (2004). The use of DNA microarrays to investigate the pharmacogenomics of drug response in living systems. Curr Top Med Chem *4*, 1329-1345.

Watanabe, A., Cornelison, R., and Hostetter, G. (2005). Tissue microarrays: applications in genomic research. Expert Rev Mol Diagn *5*, 171-181.

Webb, C. F., Yamashita, Y., Ayers, N., Evetts, S., Paulin, Y., Conley, M. E., and Smith, E. A. (2000). The transcription factor Bright associates with Bruton's tyrosine kinase, the defective protein in immunodeficiency disease. J Immunol *165*, 6956-6965.

Wheeler, D. L., Barrett, T., Benson, D. A., Bryant, S. H., Canese, K., Church, D. M., DiCuccio, M., Edgar, R., Federhen, S., Helmberg, W.*, et al.* (2005). Database resources of the National Center for Biotechnology Information. Nucleic Acids Res *33 Database Issue*, D39-45.

Wicker, L. S., and Scher, I. (1986). X-linked immune deficiency (xid) of CBA/N mice. Curr Top Microbiol Immunol *124*, 87-101.

Yanagawa, B., Taylor, L., Deisher, T. A., Ng, R., Schreiner, G. F., Triche, T. J., Yang, D., and McManus, B. M. (2005). Affymetrix oligonucleotide analysis of gene expression in the injured heart. Methods Mol Med *112*, 305-320.

Yang, W., and Desiderio, S. (1997). BAP-135, a target for Bruton's tyrosine kinase in response to B cell receptor engagement. Proc Natl Acad Sci U S A *94*, 604-609.

Zhang, B., Schmoyer, D., Kirov, S., and Snoddy, J. (2004). GOTree Machine (GOTM): a web-based platform for interpreting sets of interesting genes using Gene Ontology hierarchies. BMC Bioinformatics *5*, 16.