The use of oligonucleotide gene expression profiling to investigate a molecular classification of Common Variable Immunodeficiency

Tariq El-Shanawany

A thesis presented to Cardiff University
for the degree of Doctor of Medicine

2011

Contents

Declaration and Statements

Declaration

This work has not previously been accepted in substance for any degree and is not concurrently submitted in candidature for any degree.

Signed ……(candidate) Date ……

Statement 1

This thesis is being submitted in partial fulfilment of the requirements for the degree of MD.

Signed ……(candidate) Date ……

Statement 2

This thesis is the result of my own independent work/investigation, except where otherwise stated.

Other sources are acknowledged by explicit references.

Signed ……(candidate) Date ……

Statement 3

I hereby give consent for my thesis, if accepted, to be available for photocopying and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed ……(candidate) Date ……

Summary

Common variable immunodeficiency (CVID) is a primary antibody deficiency of unknown aetiology, the diagnosis of which is made by the exclusion of known causes of hypogammaglobulinaemia. In addition to recurrent and severe infections patients demonstrate a variable phenotype which may include other features such as granuloma and autoimmunity. Given the heterogeneous nature of this condition it appears likely that the label CVID encompasses a number of different conditions. The ability to classify CVID subgroups would be advantageous both clinically and from the research point of view. Accurate subgroup classification would allow the targeting of monitoring and treatment to those patients most at risk of complications. Furthermore, current research into the pathogenesis of CVID is hindered by the grouping of clinically and biologically distinct conditions together. To date, attempts at classification, such as by flow cytometry, have failed to accurately demarcate subgroups.

A total of 53 CVID patients were recruited to the study, and peripheral blood RNA was extracted, stored and analysed by gene expression microarray technology. The clinical and immunological data pertaining to these patients was gathered, analysed and used to allow bioinformatic analysis of the microarray data. The clinical data demonstrated a statistically significant tendency for some of the non-infective complications of CVID to cluster together, possibly suggesting a separate clinical subgroup. Flow cytometric analysis showed that in addition to previously described B and $CD4^+$ T cell phenotyping, $CD8^+$ phenotyping may be potentially useful and there was a correlation between decreased proportions of naïve $CD8^+$ T cells and the

presence of granulomatous disease. The analysis of the microarray data demonstrated a number of processes where there was differential gene expression between the clinical phenotypes, for example genes involved in the response to IL-1 in patients with granulomatous disease. Differential expression of genes involved in apoptosis was of particular interest and a consistent finding in the granuloma, autoimmunity and any complication subgroups.

Introduction

Immunoglobulin deficiency

Immunoglobulin functions

Immunoglobulins are glycoproteins produced by B cells and are found in membrane bound and soluble forms. Membrane bound immunoglobulin functions as a cell surface receptor activated by the binding of antigen specific for the particular antibody (Venkitaraman *et al.* 1991). Soluble immunoglobulin is the major secretory product of terminally differentiated B cells and is an important effector molecule central to the adaptive immune response against a number of pathogens, and in particular bacterial infection.

The terms immunoglobulin and antibody are usually used interchangeably, though immunoglobulin refers to the glycoprotein structure and antibody to the functional ability to bind antigen. Immunoglobulins are large molecules with a molecular weight of approximately 150kDa and are composed of two heavy chains of 50kDa and two light chains of 25kDa. Disulphide bonds link the light chains to the heavy chains and the two heavy chains together. Figure 1 shows a diagrammatic representation of the immunoglobulin molecule. Functionally, the molecule is divided into two Fab regions, the variable region of which binds antigen and determines specificity, and an Fc region which interacts with other components of the immune system and determines effector function.

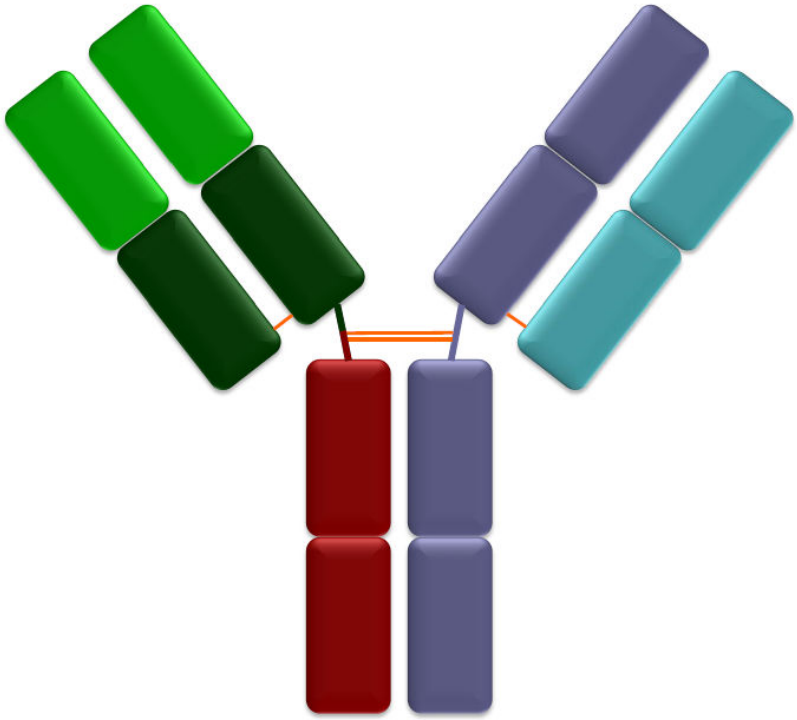


Figure 1: Diagrammatic representation of an IgG molecule. The right hand side demonstrates the heavy chain (purple) and light chain (blue). The left hand side demonstrates the Fc region (red) and Fab region (green), with the variable region highlighted (light green). Disulphide bonds are shown in orange.

Different antibody isotypes arise through class switching among genes for different heavy chain constant regions (Kraal *et al.* 1982). The isotypes are immunoglobulin M (IgM), IgD, IgG, IgA and IgE, and some isotypes have several subclasses: IgG (IgG_1, IgG_2, IgG_3 and IgG_4) and IgA (IgA_1 and IgA_2). Isotypic differences are functionally relevant; different Fc portions of the constant regions bind to a different range of Fc receptors and/or complement, thereby mediating different functional responses (Davies and Metzger 1983). Idiotypic differences refer to the different antigen binding specificities of antibodies and arise through differences in the variable regions, initially through the rearrangement of V-region genes (combinatorial and junction diversity) and subsequently through somatic hypermutation of the rearranged V genes (Early *et al.* 1980; Neuberger and Milstein 1995).

Antibodies protect the host from pathogens and their toxins in a number of ways. Simply by binding to their target they prevent the pathogen or toxin from functioning normally, for example through blocking access to cell surface receptors. This is described as neutralisation and is particularly of use against toxins and pathogens which require access to intracellular compartments to replicate.

The other functions of antibody require interaction between antigen/antibody complexes and other components of the immune system. IgM, IgG_3, IgG_1 and to a lesser degree IgG_2 are all able to activate the classical complement pathway through activation of C1 when bound in antibody/antigen complexes (Cooper 1985). The ensuing complement cascade results in the formation of peptide mediators of inflammation (the anaphylotoxins: C5a>C3a>C4a), opsonisation of the target (C3b and C4b) and formation of the membrane attack complex (C5b-9). IgA_1 is a weak activator of the complement system via the alternative pathway (Hiemstra *et al.* 1987). Antibodies also direct other functions of the immune system via interactions with cell surface Fc receptors (FcR) (Heyman 2000). IgG_1, IgG_3, IgA, and IgM, directly opsonise their targets through binding to macrophage and phagocyte Fc receptors such as FcγRI, FcγRIIA, FcαRI and FcμR. IgG subclasses 1 and 3 are also able to target NK killing via FcγRIII (antibody dependent cell-mediated cytotoxicity). IgE antibodies bind to FcεRI on mast cells, eosinophils and basophils, cross-linking of which leads to cell activation and degranulation. Table 1 summarises the distribution and functional significance of Fc receptors.

Receptor	FcγRI	FcγRIIA	FcγRIIB1&2	FcγRIII	FcαRI	Fcα/μR	FcεRI
Highest affinity Ig	IgG_1	IgG_1	IgG_1	IgG_1	IgA_1 and IgA_2	IgM	IgE
Other Ig	IgG_3, IgG_4, IgG_2	IgG_3, IgG_2, IgG_4	IgG_3, IgG_4, IgG_2	IgG_3		IgA	
Function	uptake, stimulation, respiratory burst, induction of killing	uptake, granule release (eosinophils)	inhibition of stimulation	Induction of killing	uptake, induction of killing	uptake	granule release, activation
Distribution	macrophages, neutrophils, eosinphils, DCs	macrophages, neutrophils, eosinophils, platelets, Langerhans' cells	FcγRIIB1: B cells, mast cells FcγRIIB2: macrophages, neutrophils, eosinophils	NK cells, eosinphils, macrophages, neutrophils, mast cells	macrophages, neutrophils, eosinphils	macrophages, B cells	mast cells, eosinphils, basophils

Table 1: Fc receptor distribution and function. DCs: dendritic cells.

Antibodies have a wide range of immunological functions and stimulate both the innate and adaptive arms of the immune response. IgE is important in the response against parasitic infection, while the other isotypes are central in the response to bacterial and some viral infections. IgG is distributed across all compartments of the body, has the highest serum concentration, longest half-life, shows the widest range of functions, and having undergone affinity maturation, is usually of the highest affinity for the target antigen. As a result, antibody deficiency is most severe when IgG production is affected. Similarly, replacement of IgG alone leads to a good clinical response in panhypogammaglobulinaemia.

Clinical consequences of antibody deficiency

Antibody deficiency primarily results in susceptibility to bacterial infection, though a number of factors influence disease severity. In selective IgA deficiency (IgAD), levels of IgG and IgM are normal and often there are no clinical features. There is however an association with autoimmune thyroid disease and coeliac disease. Unlike

other antibody deficiencies, IgAD is common with around 1:500 of the population affected (Lilic and Sewell 2001).

Deficiency of IgG with failure to produce antibody to test vaccination (for example to tetanus, Haemophilus influenzae type B and pneumococcus) is associated with severe disease. Infections may be recurrent, persistent and unusually severe, and episodes of pneumonia, bronchitis, sinusitis, conjunctivitis and otitis media frequently occur. Due to the rarity of primary antibody deficiencies, diagnosis is often delayed by many years (Oksenhendler *et al.* 2008). Chronic disease, especially if inadequately treated, results in end organ damage such as bronchiectasis. Patients with antibody deficiencies also suffer with non-specific features such as tiredness, lethargy, failure to thrive (in children) and arthralgia (Wood 2009).

Other features of antibody deficiency are associated with particular diseases. Those associated with common variable immunodeficiency (CVID) are detailed later. There tends not to be an increased susceptibility to viral pathogens in uncomplicated antibody deficiency, though enterovirus can be a particular problem in X-linked agammaglobulinaemia (XLA) (Halliday *et al.* 2003). Where there is an associated T cell defect, such as in hyper IgM disease (HIGM) resulting from defects in CD40L or CD40, there is also susceptibility to fungal, protozoal and viral infections and an increased risk of autoimmune disease (Notarangelo *et al.* 2007).

Causes of primary antibody deficiency

Secondary antibody deficiencies are much more common than primary antibody deficiencies, and causes include haematological malignancy, protein-losing

nephropathy or gastroenteropathy and iatrogenic causes (Chinen and Shearer 2010). Though some patients with secondary antibody deficiency require immunoglobulin replacement therapy, the sequelae of secondary antibody deficiency tend not to be as severe as primary antibody deficiency, probably as underlying immune function and antibody production is normal. For the remainder of this thesis, when discussing antibody deficiency, only primary antibody deficiency shall be considered.

Primary antibody deficiency can result from a "block" at any point between B cell progenitor development in the bone marrow to formation of a plasma cell capable of producing secretory antibody. While there are also B cell defects in severe combined immunodeficiencies (SCID) and combined immunodeficiencies (CID), the presentation of these conditions is quite different given the T cell defects with/without NK cell defects which are also present. Primary antibody deficiencies refer to those conditions where there is primarily a B cell defect.

The first antibody deficiency described was XLA, which arises due to mutation of the Btk gene encoded on the X chromosome (Anderson *et al.* 1996; Bruton 1952). Bruton tyrosine kinase (Btk) is required for B cell development, and in XLA there is an arrest between the pre-B and immature B cell stage resulting in an absence (usually complete) of peripheral B cells and agammaglobulinaemia. Conditions which prevent B cell development in the bone marrow include defects of molecules involved in the formation of the pre-B receptor (e.g. Vpreβ) and signal transduction of the B cell receptor (e.g. Igα, Igβ, Btk) (Conley *et al.* 2009).

Antibody deficiencies still occur where peripheral B cells are present. In transient hypogammaglobulinaemia of infancy (THI), numbers of B cells are normal, but there is believed to be a delay in development of antibody production capability. This emphasises and prolongs the physiological nadir of antibody levels at the age of 6 months which occurs between the loss of maternal IgG and the infants own production of immunoglobulin. In all cases, antibody production recovers spontaneously and usually by the age of 2 years.

In the class switching defects, also known as hyper-IgM syndromes (HIGM), a block at the class switching stage results in low/absent levels of IgG, IgA and IgE. Despite the name, IgM is not necessarily raised in HIGM. HIGM types 1 and 3 result from deficiency of CD40L (X-linked) and CD40 (autosomal recessive) respectively and there is also an associated, relatively mild T cell defect due to the loss of CD40-CD40L signalling. HIGM types 2, 4 and 5 are all autosomal recessive, have a pure B cell defect, and result from defects in activation-induced cytidine deaminase (AID), an unknown molecular cause and uracil-DNA glycosylase (UNG) respectively. In NFκB essential modulator (NEMO) deficiency there are variable defects of class switching, T cell function and an associated ectodermal dysplasia (Notarangelo *et al.* 2006).

Specific antibody deficiencies (SPAD) are characterised by normal B cell numbers, normal total IgG but low levels of specific IgG antibodies, usually against polysaccharide antigens, with a failure to respond to vaccination. In contrast to the defects of B cell development and class switching which are present from birth, SPAD may develop at any age. The cause is unknown, but is suspected to be polygenic possibly with environmental trigger (s). There are similarities between

SPAD and CVID, including the development of disease later in life. The presence of SPAD or CVID increases the chance of other family members having one of these conditions. SPAD and CVID remain one of the major areas where the genetics is poorly understood. CVID will be discussed in more detail in the following section. Table 2 summarises the causes of hypogammaglobulinaemia.

Pre B cell	Immature B cell	Class switching	Specific antibody production	Delay in antibody production (THI)	CVID
λ5 Vpreβ Cμ Igα Igβ	Btk BLNK	CD40L AID CD40 UNG NEMO	Polygenic +/- environmental influence	Polygenic +/- environmental influence	Polygenic +/- environmental influence

Table 2: Summary of known molecular causes of hypogammaglobulinaemia primarily due to B cell defect. BLNK: B cell linker protein.

CVID

Diagnosis and Pathogenesis

In contrast to the single gene defects described above, understanding of the pathogenesis of CVID is currently limited. CVID is largely a diagnosis of exclusion and single-gene defects, malignancy, drug reactions and other causes of secondary hypogammaglobulinaemia all need to be ruled out. See Table 3 for the diagnostic criteria of CVID (Conley *et al.* 1999; de Vries 2006). Unfortunately the diagnosis of CVID is often delayed, and during this delay irreversible end organ damage such as bronchiectasis may result. Features which should prompt investigation for an immunodeficiency are a family history, recurrent infections, unusually severe

infections and infections with unusual organisms (Cunningham-Rundles and Bodian 1999; de Vries 2006).

Probable CVID (all must be met)	**Possible CVID (all must be met)**	**Differential diagnosis (must be excluded)**
A decrease of more than 2 standard deviations below the mean for age in IgG and at least one of IgA or IgM	A decrease of more than 2 standard deviations below the mean for age in at least one of IgG, IgA or IgM	Genetic Disorders: Ataxia Telangiectasia, Autosomal forms of SCID, Hyper IgM Immunodeficiency, Transcobalamin II deficiency and hypogammaglobulinemia, X-linked agammaglobulinemia, X-linked lymphoproliferative disorder (EBV associated), X-linked SCID, Some metabolic disorders, Chromosomal Anomalies, Chromosome 18q- Syndrome, Monosomy 22 Trisomy 8, Trisomy 21
Greater than 2 years of age (i.e. THI excluded)	Greater than 2 years of age (i.e. THI excluded)	Drug Induced: Antimalarial agents, Captopril, Carbamazepine, Glucocorticoids, Fenclofenac, Gold salts, Penicillamine, Phenytoin, Sulfasalazine
Poor responses to vaccinations and/or absent isohemagglutinins	Poor responses to vaccinations and/or absent isohemagglutinins	Infectious Diseases: HIV, Congenital Rubella, Congenital infection with CMV, Congenital infection with Toxoplasma gondii, Epstein-Barr Virus
Defined causes of hypogammaglobulinemia have been excluded	Defined causes of hypogammaglobulinemia have been excluded	Malignancy: Chronic Lymphocytic Leukemia, Immunodeficiency with Thymoma, Non Hodgkin's lymphoma, B cell malignancy
		Systemic Disorders: Immunodeficiency caused by hypercatabolism of immunoglobulin, Immunodeficiency caused by excessive loss of immunoglobulins (renal loss, gastrointestinal loss, severe burns, lymphangiectasia)

Table 3: Diagnostic criteria for CVID. Adapted from (Conley *et al.* 1999).

CVID represents a spectrum of disease, and it has been suggested that some patients with CVID have a significant T cell defect and should therefore be classified separately as having late onset combined immunodeficiency (LOCID) (Malphettes *et*

al. 2009). The authors suggest that patients are defined as suffering with LOCID if they have suffered with an infection indicative of a severe deficiency of cell-mediated immunity and/or have a $CD4^+$ T cell count of <200 cells/µl. However, this does not take into account the fact that some CVID patients require immunosuppressive therapy and that susceptibility to such infections in these patients may be iatrogenic. Also, opportunistic infections occurred in CVID patients with both normal and low $CD4^+$ T cell counts, suggesting that this one marker is not sufficient to classify a separate disease. Furthermore, the significance of a low $CD4^+$ T cell count without a history of an opportunistic infection is unclear.

CVID is likely to represent a heterogeneous group of disorders which lead to antibody deficiency. Many of the cases of CVID are believed to be multifactorial in aetiology and may also include an environmental trigger or triggers, which fits with the observation that disease onset is often in adulthood (Cunningham-Rundles and Bodian 1999; Quinti *et al.* 2007). Some genetic causes/predisposing factors of CVID have been described:

ICOS (inducible T-cell co-stimulator) is transiently expressed on activated T cells and is a member of the CD28 family (Hutloff *et al.* 1999). The ligand, B7RP-1, is a member of the B7 family and is constitutively expressed on B cells. In murine knockout models there is a failure to produce antibody serotypes other than IgM and IgG_3 (McAdam *et al.* 2001). To date, nine patients with homozygous mutations in ICOS have been described, and all have antibody deficiency with a similar phenotype, and no autoimmune disease (Grimbacher *et al.* 2003; Salzer *et al.* 2004). All nine

patients share the same mutation (1815bp deletion leading to a premature stop codon at position 595), suggesting a common founder effect.

CD19 mutations also result in autosomal recessive hypogammaglobulinaemia (Kanegane *et al.* 2007; van Zelm *et al.* 2006). A number of different mutations have been described, though the total number of patients remains small (six). Numbers of circulating B cells were normal when measured by CD20, but numbers of memory B cells were reduced. None of the patients with CD19 mutations suffered with autoimmune phenomena. CD19 is normally expressed throughout B cell differentiation and constitutes part of the B cell co-receptor along with CD21 (complement receptor 2) and CD81 and acts to amplify intracellular signalling following B cell receptor ligation (Fujimoto *et al.* 2000).

TACI (transmembrane activator and calcium-modulating cyclophilin) is expressed on B cells and activated T cells. It interacts with B-cell activating factor (BAFF) and a proliferation inducing ligand (APRIL) present on B cells and influences B cell survival, differentiation and class switching (Mackay *et al.* 2003). In mouse models, TACI deficiency leads to a severe phenotype with abnormal antibody responses, polyclonal B cell expansion and systemic autoimmune disease (Seshasayee *et al.* 2003; Yan *et al.* 2001). However, in humans the role of TACI in disease is much less clear, with common polymorphisms such as c.81G>A, c.831T>C, c.752C>T and c.291T>G present in the healthy population. While up to 10% of CVID patients have been found to carry TACI polymorphisms, population studies have demonstrated that around 2% of healthy controls carry the same polymorphisms (Castigli *et al.* 2007; Pan-Hammarstrom *et al.* 2007; Salzer *et al.* 2005). As the prevalence of CVID is

estimated to be 1/25,000, this means that over 90% of people who are heterozygous or homozygous for one of the TACI polymorphisms do not have CVID (or do not have a clinical phenotype severe enough to result in a referral for immunological workup). CVID patients with TACI polymorphisms appear to be more likely to suffer from autoimmune disease and splenomegaly (Salzer *et al*. 2005; Waldrep *et al.* 2009). Therefore while TACI polymorphisms are not disease causing mutations, they may have a role as susceptibility and/or disease modifying genes.

B cell activating factor receptor (BAFF-R) is structurally similar to TACI, and as its name suggests, binds to BAFF (Boschetti *et al.* 2005). BAFF-R/BAFF interactions are important for B cell survival and homeostasis (Schiemann *et al.* 2001). Homozygous deletion of BAFF-R has been described in one family with two affected siblings, both of whom had low levels of IgG and IgM, though unusually for CVID had normal levels of IgA (Eibel *et al.* 2009).

Polymorphisms in the mismatch repair gene Msh5 encoded in the major histocompatibility complex (MHC) class III region have been described. Msh5 is involved in DNA repair following class switch recombination. CVID patients with Msh5 polymorphisms have abnormalities of their switch regions. However, Msh5 polymorphisms do not appear to be significantly more common in CVID compared to the general population (Sekine *et al.* 2007).

Despite the discovery of the above predisposing genetic factors, these only account for a minority of CVID cases. In most patients with CVID the aetiology remains unknown. To date, investigation of other candidate genes such as APRIL, BAFF and

BCMA have not identified disease causing mutations (Losi *et al.* 2006; Losi *et al.* 2005; Salzer *et al.* 2008). Most cases of CVID are sporadic, though the risk is increased where there is a family history of CVID, specific antibody deficiency, IgAD and/or THI. It could be argued that the cases of hereditary antibody deficiency such as ICOS and CD19 deficiency represent different diseases which have now been identified, and that these defined conditions should no longer be classed as CVID.

Infective complications and management

As might be expected from a collection of heterogeneous conditions, the clinical spectrum of CVID is variable. However, the most common presenting feature of CVID is recurrent and/or severe infections particularly bronchitis, sinusitis and/or otitis. The infectious complications are similar to those of other antibody deficiencies described above (Cunningham-Rundles and Bodian 1999; Oksenhendler *et al*. 2008). The most common organisms causing respiratory infections are pneumococcus and H. influenzae, and gastrointestinal infections are giardia, salmonella and campylobacter. Though meningitis is more frequent in CVID patients than the general population, it remains a rare complication. The most frequent causative organisms of meningitis in CVID are pneumococcus, N meningitidis and H influenzae. The infectious complications of CVID are summarised in Table 4. The mainstay for the prevention and treatment of infective complications is immunoglobulin replacement therapy and the use of antibiotics. Antibiotics may be used long term at a prophylactic dose and at high dose to treat infective exacerbations as they arise.

Infectious Complications	
Very Frequent (70-98% of patients)	Bronchitis
	Sinusitis
	Otitis
	Pneumonia
Common (up to 30% of patients)	Gastrointestinal infection
Less frequent (up to 10% of patients)	Recurrent herpes zoster
	Profuse warts (papillomavirus)
	Mycoplasma
Rare (<2% of patients)	Osteomyelitis
	Skin abcesses
	Septic arthritis
	Meningitis
	Encephalitis
	Chronic mucocutaneous candidiasis
	Visceral CMV infection
	Recurrent parotitis
	Pneumocystis jirovecii infection

Table 4: Infectious complications of CVID (Cunningham-Rundles and Bodian 1999; Oksenhendler *et al*. 2008; Quinti *et al*. 2007).

Polyclonal human normal immunoglobulin is used in CVID (and other primary antibody deficiencies) and significantly reduces the risk of infection. Existing formulations replace IgG only, and are a pooled blood product produced by the fractionation of plasma with up to 10,000 donors per batch. Individual donations are screened for HIV-1, HIV-2 and Hepatitis C virus antibodies and also for Hepatitis B surface antigen and alanine aminotransferase (ALT). Mini-pools (usually of 512 plasma donations) are also screened by PCR before being allowed to proceed to the fractionation process. A number of methods are in use by different manufacturers, including pH dependent precipitation, DEA sephadex and ion-exchange chromatography.

In addition to the checks on plasma used for fractionation, there are a number of other steps to ensure safety from the potential transmission of infectious particles. The fractionation process itself reduces by many log-fold any viruses which are present in the initial donation. For example cold ethanol fractionation has been shown to reduce the number of HIV particles by approximately 10^{10} (Henin *et al.* 1988). Increasingly, manufacturers are adding further antiviral steps following fractionation such as solvent/detergent removal (especially effective against enveloped viruses), pasteurisation and/or nano-filtration. Although the risk of prion disease transmission by immunoglobulin infusion is theoretically possible the anti-viral steps currently in place have been demonstrated to also be effective for the removal of prion particles (Reichl *et al.* 2002).

Every batch of pooled immunoglobulin is also subjected to a battery of tests, both to ensure safety and to introduce some standardisation of what is by its nature a variable product. These include checking that not less than 95% of the product is IgG, the distribution of IgG subclasses, that Fc function is normal, the levels of pre-kallikrein activator and the levels of haemagglutinins. Also assessed are the levels of specific IgG against a range of infections (including diphtheria, measles and polio) which must meet minimum titres.

However, it is accepted that there are significant batch-to-batch variations in many different antibody titres (Galama *et al.* 1997; Lamari *et al.* 1999). It is very difficult to control for this, although legislation is now in place for certain antibody specificities, for example anti-D (Thorpe *et al.* 2006). There are also likely to be manufacturer-to-manufacturer variations, as well as differences in plasma source with effects on

repertoire dependent on endemic diseases and vaccination status. Immunoglobulin products are therefore not identical, and a summary of human normal immunoglobulin products currently available in the UK is summarised in Table 5.

The major constituent of pooled immunoglobulin is IgG. The specificity repertoire of this IgG includes pathogens and superantigens, and also antibodies against immunoregulatory molecules such as cytokines, TCR, CD4, CD95 and a number of integrins and anti-idiotype antibodies (Takei *et al.* 1993). There are other immunologically active proteins present, including cytokines, soluble cytokine receptors, CD4 and MHC II and potentially the stabilising agents (mainly sugars) used (Lam *et al.* 1993).

Product Name	Admin route	Manufacturer name	Formulation	% Ig	IgA content (mg/ml)	Carbohydrate stabiliser
Flebogamma 5% DIF	IV	Grifols	Liquid	5	<0.05	Sorbitol
Flebogamma 10% DIF	IV	Grifols	Liquid	10	<0.05	Sorbitol
Gammagard SD	IV	Baxter	Lyophilised powder	5	<0.003	Glucose Glycine
Gammaplex	IV	BPL	Liquid	5	<0.01	D-sorbitol Glycine
Intratect	IV	Biotest	Liquid	5	<2	Glycine
Kiovig	IV	Baxter	Liquid	10	<0.14	Glycine
Octagam 5%	IV	Octapharma	Liquid	5	≤0.2	Maltose
Octagam 10%	IV	Octapharma	Liquid	10	≤0.4	Maltose
Privigen	IV	CSL Behring	Liquid	10	<0.025	Proline
Vigam Liquid	IV	BPL	Liquid	5	≤0.1	Sucrose Glycine
Gammanorm	SC	Octapharma	Liquid	16.5	≤0.0825	Glycine
Hizentra	SC	CSL Behring	Liquid	20	≤0.05	Proline
Subcuvia	SC	Baxter	Liquid	16	4.8	Glycine
Subgam	SC	BPL	Liquid	16	<3.2	Glycine
Vivaglobin	SC	CSL Behring	Liquid	16	≤1.7	Glycine

Table 5: Human normal immunoglobulin products currently available in the UK. Admin route: administration route, % Ig: percentage of immunoglobulin in product when made up for administration, IV: intravenous, SC: subcutaneous, BPL: Bio Products Laboratory, CSL: Commonwealth Serum Laboratories.

Human normal immunoglobulin is a blood product and the informed consent of patients should include an explanation of this and the very small risk of the transmission of an as yet unidentified pathogen (Boschetti *et al*. 2005). The largest problem was historically with transmission of hepatitis C virus (HCV) (Weiland *et al.* 1986) , but with appropriate screening, transmission of known viruses should be avoidable, and there have been no cases of HCV transmission since the identification of the causative agent. Furthermore, manufacturers continue to improve the safety profile with each new generation of products having additional antiviral steps in the production process (e.g. nanofiltration). The risk of prion disease transmission is currently topical, though in practice there are no known cases of transmission by immunoglobulin, even in patients who have received batches from donors who subsequently developed Creutzfeldt-Jakob disease (CJD) (El-Shanawany *et al.* 2009). A prospective study has been established with the particular purpose of following up all such patients. A previous prospective study of 13,508 intravenous infusions in 459 patients demonstrated a good safety profile of IVIG (Brennan *et al.* 2003). There are no reported cases of HIV transmission by immunoglobulin.

Immunoglobulin replacement therapy can be administered via a number of routes. Intramuscular immunoglobulin was first used in 1952 (Bruton 1952), but its use was limited by pain and the volume which could be administered, and it often proved difficult to provide sufficient replacement of immunoglobulin. The development of intravenous immunoglobulin preparations (IVIG) allowed the infusion of much greater volumes and could maintain physiological blood levels of immunoglobulin using doses of around 0.4g/kg/month. In addition immunoglobulin could be given as high dose (hdIVIG) as an immunomodulatory treatment for example for the treatment

of autoimmune and inflammatory diseases (El-Shanawany and Jolles 2007). Subsequently with the demonstration that immunoglobulin could be administered subcutaneously (SCIG) using portable syringe drivers this route has become increasingly popular and provides a feasible means for home therapy with the patients trained to infuse themselves (Gardulf 2007). IVIG can deliver larger volumes and is usually administered every 3 to 4 weeks, whereas SCIG is administered weekly. SCIG results in more stable IgG concentrations which is more physiological than the peaks and troughs associated with IVIG. SCIG has been demonstrated to be at least as efficacious as IVIG for immunoglobulin replacement (Chapel *et al.* 2000). Recently the development of hyaluronidase facilitated SCIG (fSCIG) has allowed the administration of high dose immunoglobulin via the subcutaneous route (Frost 2007; Melanned *et al.* 2008) and the first patient to receive fSCIG as part of a home therapy package has now been reported (Knight *et al.* 2010).

Non-infective complications

In contrast with defects predominantly restricted to B cells, such as XLA, patients with CVID can also suffer with a number of non-infectious complications due to T cell abnormalities and immune dysregulation. In particular, autoimmune disease, granulomatous disease and malignancy can complicate CVID.

Around 1 in 4 patients with CVID develop autoimmune disease. The most common are autoimmune cytopenias, especially immune thrombocytopenia (ITP) and autoimmune haemolytic anaemia (AHA), which can co-exist (Evan's syndrome) (Michel *et al.* 2004). Autoimmune thrombocytopenia in CVID can be challenging to

treat with steroids, and may require hdIVIG, splenectomy or anti-CD20 monoclonal antibody (rituximab) (El-Shanawany *et al.* 2007). Autoimmune neutropenia is less common, but can also occur (Wang and Cunningham-Rundles 2005).

Though less frequently than the cytopenias, there is also an increased risk of other autoimmune disease in CVID, including rheumatoid arthritis, juvenile rheumatoid arthritis, sicca syndrome, inflammatory bowel disease, primary biliary cirrhosis, alopecia, pernicious anaemia, autoimmune thyroid disease, nephritic syndrome, systemic lupus erythematosus and vascultides (Chapel *et al.* 2008; Cunningham-Rundles and Bodian 1999; Oksenhendler *et al*. 2008; Quinti *et al*. 2007). Management of autoimmune disease in CVID is complicated by the fact that immunosuppression is likely to worsen the underlying immunodeficiency. Decisions on the aggressiveness of treatment need to balance the conflicting needs of treating the autoimmune condition with preserving immunocompetence in an already immunodeficient individual.

Anti-IgA autoantibodies can develop, though are usually only found in patients with undetectable rather than reduced levels of IgA (Lilic and Sewell 2001). Some patients with these autoantibodies may have infusion reactions to IVIG (Ferreira *et al.* 1988), though the frequency with which this happens is debated. Current assays measure IgG anti-IgA but the correlation between the presence of these antibodies and reactions to IVIG is no longer as certain as was once thought (Lilic and Sewell 2001). High titre anti-IgA have the best association with adverse reactions. Patients with anti-IgA have been successfully treated with IVIG. It has also been argued that anaphylactic reactions to the small quantities of IgA in IVIG are likely to be due to IgE anti-IgA

antibodies, however this remains controversial and there is not a reliable and validated assay currently in use in diagnostic laboratories (Bjorkander *et al.* 1987).

There is a higher incidence of a number of malignancies in CVID patients compared to the general population. There is particular susceptibility to haematological malignancy, with Non-Hodgkin's and other lymphomas reported in up to 8% of CVID patients. Patients with CVID are estimated to have a greater than 400-fold risk of Non-Hodkin's lymphoma compared to the age-adjusted expected incidence. There is also a less marked, though statistically significant, increased risk of adenocarcinoma of the stomach. Other reported malignancies such as squamous cell carcinoma, melanoma and other adenocarcinomas occur with too low a frequency to be able to determine whether there is an increased prevalence or whether these are chance occurrences (Cunningham-Rundles and Bodian 1999; Quinti *et al*. 2007).

Persistent lymphadenopathy (without underlying haematological malignancy) is another feature which can be found in CVID patients. This can occur both with and without granulomatous disease, and the underlying cause is unclear. Some authors group patients with either one, or both, of persistent lymphadenopathy and granulomatous disease into the same subgroup described as polyclonal lymphoproliferation (Chapel *et al*. 2008; Chapel *et al.* 2012). However, other authors consider these two features as separate phenomena (Piqueras *et al.* 2003; Warnatz *et al.* 2002; Warnatz and Schlesier 2008; Wehr *et al.* 2008), and indeed have identified different risk markers for the two conditions.

Rates of granulomatous disease in CVID vary from around 8-22% depending on the study (Ardeniz and Cunningham-Rundles 2009; Morimoto and Routes 2005). The most common organ to be affected are the lungs, where there may also be an associated lymphoid infiltration (Park and Levinson 2010). However, granuloma can be found at many sites, including liver, lymph nodes, spleen, brain, skin and bowel. Developing autoimmune disease or granulomatous disease puts the patient in a group at increased risk of developing the other, and around 50% of patients with granulomatous disease have autoimmune disease (Mechanic *et al.* 1997). The small numbers make it difficult to confirm the suspicion that patients with granulomatous disease suffer with increased mortality compared to other CVID patients, but there is certainly increased morbidity. Patients with granulomatous-lymphocytic interstitial lung disease have been shown to have shortened survival (Bates *et al.* 2004). There are no clinical trials to guide the treatment of granulomatous disease, and as with autoimmune disease the aim is to strike a balance between treating the condition without worsening the underlying immunodeficiency. Case reports have shown improvement with the use of corticosteroids, other immunosuppressive agents such as cyclosporine, azathioprine and methotrexate and more recently with anti-TNFα therapy (infliximab) (Fakhouri *et al.* 2001; Hatab and Ballas 2005; Thatayatikom *et al.* 2005).

The granulomas are usually non-caseating and similar in appearance to the lesions in sarcoidosis. Some patients with CVID are initially diagnosed with sarcoidosis which can lead to further delay in the recognition of CVID. The pathogenesis of granulomatous disease in CVID is unclear. An infectious trigger has been proposed, and while in some cases toxoplasma gondii or human herpes virus 8 (HHV8) have

been identified, in most patients no infectious agent is found (Mrusek *et al.* 2004; Wheat *et al.* 2005). In some cases there is an increase in the Th1 cytokines IL-12 and IFN-γ, which has been suggested as being involved in the pathogensis (Cambronero *et al.* 2000). Other possible markers are an inverted $CD4^+/CD8^+$ ratio and reduced T cell mitogen responses (Mechanic *et al*. 1997), though more recent studies have not found such differences between CVID patients with and without granuloma (Ardeniz and Cunningham-Rundles 2009). The best marker for risk of granulomatous disease currently available is B cell memory phenotyping, which is discussed in more detail in the next section (Piqueras *et al*. 2003; Warnatz *et al*. 2002; Wehr *et al*. 2008).

Attempts at classification

Patients with CVID are all hypogammaglobulinaemic, but otherwise have a highly variable phenotype. While some patients improve dramatically on immunoglobulin replacement therapy leading essentially normal lives, others suffer with a combination of recurrent infections, progressive lung damage, autoimmune disease and/or granulomatous disease. Numerous attempts have been made to classify patients into pathogenetic and clinically relevant subgroups. Though CVID probably results from complex defects in B-T cell interactions, the end result is disruption of B cell function resulting in hypogammaglobulinaemia. Therefore, the investigation of defects of B cell function is an appealing approach. One method (Bryant classification) looked at the ability of B cells from CVID patients to secrete IgM and IgG *in vitro* following stimulation with IL-2 and anti-μ (Bryant *et al.* 1990). While it was possible to subclassify CVID into groups where the patient's B cells could not secrete IgM or IgG, could only secrete IgM or could secrete both isotypes (called groups A, B and C respectively), this *in vitro* classification did not translate into a meaningful clinical

classification. Later, the ability of B cells to secrete IgA using this *in vitro* system was also examined, but did not improve the ability of the assay to differentiate clinical CVID subgroups (Scott *et al.* 1994). Furthermore, it proved difficult to set up robust functional B cell assays which satisfied the stringent quality control in place in clinical laboratories, limiting the amount of research possible with this approach. The most successful methods of classification to date are those based on flow cytometric phenotyping of peripheral blood B cell differentiation.

Figure 2 summarises B cell development from the bone marrow immature B cell stage onwards. Immature B cells emigrate into peripheral blood and become transitional B cells, initially T1 transitional B cells with a $IgM^{hi}IgD^{lo}CD21^{lo}CD38^{hi}$ surface phenotype which mature into T2 cells ($IgM^{hi}IgD^{hi}CD21^{hi}CD38^{hi}$). The most immature peripheral B cells can be identified by $CD21^{lo}$ expression, and the transitional B cell compartment (both T1 and T2) identified by $IgM^{hi}CD38^{hi}$ expression (Casola 2007; Wehr *et al*. 2008). Subsequent to antigen exposure and appropriate co-stimulation, B cells develop into either marginal zone ($IgM^{hi}IgD^{lo}CD21^{int}CD38^{-}CD27^{+}$) or follicular ($IgM^{lo}IgD^{hi}CD21^{int}CD38^{int}CD27^{-}$) B cells. Marginal zone B cells develop into plasmablasts and then plasma cells. Plasma cells are not circulating and only very limited numbers of plasmablasts can be detected in the peripheral blood. Plasmablasts can be distinguished from other B cell subsets by their very high expression of CD27. Follicular B cells develop into memory B cells. Following class switching, surface expression of IgM and IgD is lost (and the new isotype class expressed). Therefore, as B cells develop from follicular to non-class switched memory to class switched memory B cells the phenotype changes from $CD27^{-}IgD^{+}$ to $CD27^{+}IgD^{+}$ to $CD27^{+}IgD^{-}$ (Sanz *et al.* 2008; Warnatz and Schlesier 2008).

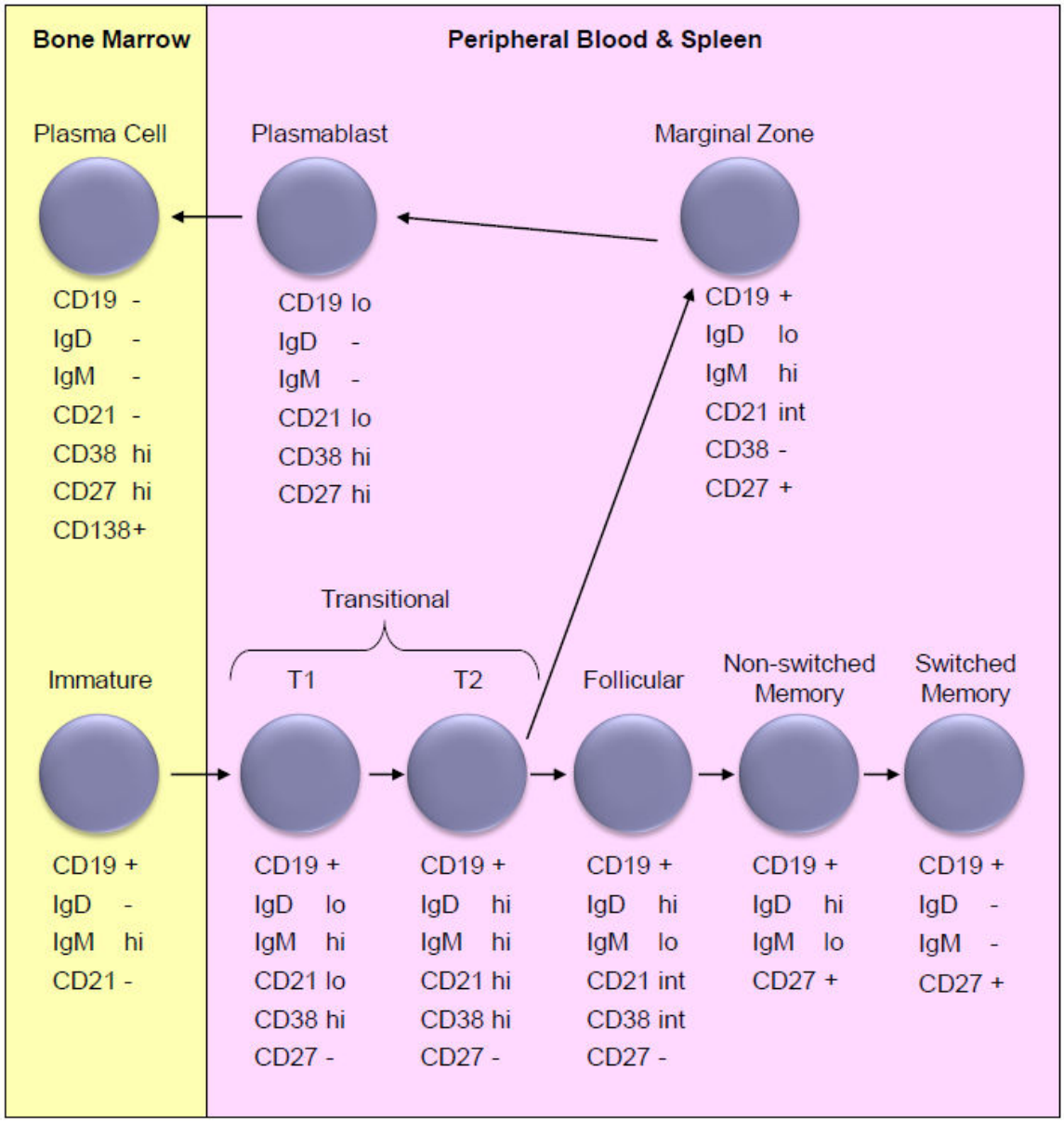


Figure 2: Peripheral B cell development and surface phenotypes.

Differences in the expression of surface immunoglobulin isotypes on B cells from patients with CVID have been described for some time (Siegal *et al.* 1971). More recently, the use of CD27 as a phenotypic marker of memory B cells, allowed further study of B cell memory phenotypes (Agematsu *et al.* 2002; Brouet *et al.* 2000). Further work and correlation with clinical phenotypes has led to the development of 3 classification methods for CVID: Freiburg, Paris and EUROclass (Piqueras *et al.* 2003; Warnatz *et al.* 2002; Wehr *et al.* 2008) which are summarised in Figure 3.

Memory B cell phenotyping has proved useful in the paediatric CVID population in addition to adults (Yong *et al.* 2010).

The Frieburg classification system (Warnatz *et al*. 2002) is based on a study of 30 well characterised CVID patients and 22 healthy controls and examines the percentage of B cell phenotypes as a percentage of all lymphocytes. Therefore, it is only applicable to individuals with >1% of B cells as a percentage of lymphocytes. The first differentiation is based on the percentage of class switched $CD27^{+}IgM^{-}IgD^{-}$ memory B cells. All healthy controls had >0.5% (of lymphocytes) of this subpopulation. CVID patients were classified into group I (<0.5%) and group II (≥0.5%). Group I was further differentiated into group Ia where there was an increased proportion of $CD21^{-}$ immature B cells (>20% of B cells), and group Ib where there was a normal proportion of $CD21^{-}$ B cells. CVID patients falling into group Ia were more likely to have splenomegaly and autoimmune cytopenias. In group Ia there were 10 patients, 100% of whom had splenomegaly and 60% had autoimmune disease. Group Ib consisted of 13 patients, 42% had splenomegaly and 46% had autoimmune disease. There were 7 patients in group II, and the percentage with splenomegaly was 14% and autoimmune disease 43%. While this approach can give an indicator of the risk of non-infectious complications in CVID, there is considerable overlap between the groups. A further limitation of this study is that CVID patients with granulomatous disease were excluded as there were only 2 such patients in the initial cohort.

The Paris classification system (Piqueras *et al*. 2003) divides CVID patients into three groups (MB0, MB1 and MB2) defined by the percentage of different B cell

subpopulations from the total B cell population. This study of 57 patients (and 20 healthy controls) included 11 patients with granulomatous disease. MB0 showed a reduced percentage of total memory B cells (<11% $CD27^+$ B cells of total CD19 B cells), MB1 having normal total memory B cells, but decreased class switched memory B cells (>11% $CD27^+$ and <8% $IgD^-IgM^-CD27^+$ B cells) and MB2 having a normal number of total memory B cells and class switched memory B cells. There was also no difference in rates of infectious complications between the groups. Group MB0 was associated with splenomegaly, lymphoid proliferation and granulomatous disease, group MB1 with splenomegaly and group MB2 having the lowest incidence of non-infectious complications. Rates of autoimmune disease were similar between groups MB0 and MB1 but higher than in MB2. However, as with the Freiburg classification, there remains overlap between the groups with splenomegaly, granulomatous disease and autoimmune disease occurring in every group, though at different frequency. Interestingly, analysis of somatic hypermutation in 4 patients with a severe defect of memory B cells was normal, suggesting that in these cases the block in B cell differentiation occurred between somatic hypermutation taking place and the formation of memory B cells (Piqueras *et al*. 2003).

The most recent classification system, EUROclass (Wehr *et al*. 2008) resulted from a study of 303 patients with a diagnosis of CVID. This study used both the Freiburg and Paris methods to investigate the patients and also looked at further parameters such as $CD38^{hi}IgM^{hi}$ transitional B cells. This larger study found that neither of the existing methods differentiated between patients with and without autoimmune disease at a statistically significant level, nor were any B cell phenotype features found which could do so. Different cut-offs were determined to better distinguish between patient

groups likely to suffer with granulomatous disease, lymphadenopathy and splenomegaly. Patients with <1% of B cells (as a percentage of total lymphocytes) are excluded from the classification. The following percentages all relate to percentage of B cells. The first step is to divide into two groups SmB^- and SmB^+ according to whether there are ≤2% or >2% class switched memory B cells respectively. The SmB^- group is further subdivided dependent on the numbers of transitional B cells. Tr^{hi} are those with ≥9% transitional B cells while Tr^{lo} have <9%. Both SmB^- and SmB^+ groups are then analysed for the proportion of $CD21^{lo}$ B cells. Those with ≥10% $CD21^{lo}$ B cells are defined as $CD21^{lo}$, and <10% $CD21^{lo}$ B cells $CD21^{norm}$. This nomenclature is potentially confusing; $CD21^{norm}$ refers to the group of patients where there is a normal percentage of $CD21^{lo}$ immature B cells, whereas the $CD21^{lo}$ group has an expanded population of $CD21^{lo}$ B cells. SmB^- patients are more likely to have splenomegaly and granuloma, Tr^{hi} patients are more likely to have lymphadenopathy and $CD21^{lo}$ patients are more likely to have splenomegaly and granuloma. While EUROclass improves on the previous classification methods, there still remains overlap between the groups. For example comparing the SmB^- and SmB^+ groups, the rates of splenomegaly are 52% and 24%, and the rates of granulomatous disease 17% and 4% respectively.

This image has been removed by the author for copyright reasons

Figure 3: Summary of Freiburg, Paris and EUROclass classifications. Reproduced from (Warnatz and Schlesier 2008).

(a) The Freiburg classification first discriminates CVID patients on the basis of class switched $CD19^{+}CD27^{+}IgM^{-}IgD^{-}$ memory B cells: patients with class switched B cells below 0.4% of lymphocytes are assigned to type I, patients with class switched B cells equal or above 0.4% are classified to type II. Patients of type I are subclassified on the basis of B cells expressing low amounts of CD21: patients exhibiting equal or more than 20% $CD21^{lo}$ cells out of $CD19^{+}$ B cells are grouped into type Ia and patients with less than 20% $CD21^{lo}$ B cells are classified to type Ib (representative dot plots are shown). (b) The Paris classification groups patients according to the percentage of memory B cell within $CD19^{+}$ B cells: patients with less than 11% of total $CD27^{+}$ cells of $CD19^{+}$ B cells are classified in group MB0, patients with more than 11% of $CD27^{+}$ B cells but less than 8% of class switched B cells are grouped into MB1 and all patients who do not fulfill one of these criteria are assigned to group MB2. (c) The EUROClass classification discriminates patients with more than 1% $CD19^{+}$ B cells of lymphocytes (group B^{+}) and equal or less than 1% (group B^{-}). The B^{+} group is subdivided into patients with more than 2% class switched memory $CD27^{+}IgM/IgD^{-}$ B cells within the $CD19^{+}$ B cell compartment (group smB^{+}) and patients with equal or less than 2% class switched cells (group smB^{-}). The smB^{-} group can be further subdivided based on the expansion of transitional B cells (part A): patients with equal or more than 9% of transitional B cells of total $CD19^{+}$ B cells are assigned to group Tr^{hi} and patients with less than 9% transitional B cells are classified into Tr^{norm}. As shown in part B, the groups smB^{+} and smB^{-} both are further subdivided into groups $CD21^{norm}$ (less than 10% $CD21^{lo}$ B cells of $CD19^{+}$ B cells) or $CD21^{lo}$ (equal or more than 10% $CD21^{lo}$ B cells). Representative flow cytometric plots of relevant B cell subpopulations are shown.

The utility of T cell phenotyping for the characterisation of CVID patients has also been investigated. A study of 38 patients suggested a link between increased $CD8^+CD57^+$ and $CD8^+HLA\text{-}DR^+$ T cells and splenomegaly. The presence of increased $CD8^+CD57^+$ and $CD8^+HLA\text{-}DR^+$ were associated with a CD4:CD8 ratio of ≤0.9, which may represent a straightforward method of identifying this group (Wright *et al.* 1990). A subsequent study of 50 patients showed an expansion of activated T cells corresponded with impaired memory B-cell differentiation, and was more frequent in patients with splenomegaly and granulomatous disease (Viallard *et al.* 2006). However, analysis of activated T cells has not to date been able to reliably classify CVID patients into different subgroups, and has not been shown to improve or add to the information available from B cell phenotyping.

The Rome classification system was derived from an investigation of 60 patients and utilises peripheral CD4 naïve T cell ($CD4^+CD45RA^+CD62L^+$) percentages to define patients belonging to the lower, intermediate and upper tertiles into groups I, II and III respectively (Giovannetti *et al.* 2007). Group I has <15% naïve CD4 T cells of total CD4 T cells, group II 16-29% and group III >30%. Patients in group I were more likely to have a severe B cell classification, whereas group III were more likely to have a normal B cell phenotype. Clinically there was correlation with splenomegaly with 95%, 70% and 35% of patients having splenomegaly in groups I, II and III respectively. The size of splenomegaly also varied with patients in group I tending to have more significant splenomegaly, and group III having the smallest degree of splenomegaly on average. Interestingly, at 68%, splenomegaly rates in this cohort were significantly higher than in other reports. CD4 and CD8 TCR repertoires were also examined by flow cytometry and spectratyping. Disrupted TCR repertoires of

both CD4 and CD8 T cells were found in patients with a contracted naïve CD4 T cell population.

Other studies of T cell phenotypes are also small, but have suggested a number of correlations between: low T_{reg} numbers and splenomegaly, though no link with granulomatous disease, autoimmunity or the development of bronchiectasis (Fevang *et al.* 2007), low numbers of naïve T cells and splenomegaly (Livaditi *et al.* 2007), low numbers of naïve T cells and low numbers of switched memory B cells (Giovannetti *et al*. 2007) and increased $CD4^+$ activation markers (increase in CD29, HLA-DR, CD45RO, decrease in CD27, CD62L, CD45RA) and Freiburg group Ia (Vlkova *et al.* 2006). One proposal is that combining B and T cell phenotyping may allow better characterisation of CVID patients (Moratto *et al.* 2006). Table 6 summarises the clinical associations with currently defined classification systems.

Group	Association (s)
Freiburg	
1a	**Highest risk of splenomegaly and autoimmune disease**
1b	**Intermediate risk of splenomegaly and autoimmune disease**
2	**Lowest risk of splenomegaly and autoimmune disease**
Paris	
MB0	**Increased risk of splenomegaly, lymphadenopathy and granuloma**
MB1	**Increased risk of splenomegaly**
MB2	**Lowest complication rates**
EUROclass	
SmB-	**Higher rates of splenomegaly and granuloma (compared to SmB+)**
Tr (hi)	**Higher rate of lymphadenopathy (compared to Tr (norm))**
CD21 (lo)	**Higher rates of splenomegaly and granuloma (compared to CD21 (norm))**
Rome	
T1	**Highest risk of splenomegaly**
T2	**Intermediate risk of splenomegaly**
T3	**Lowest risk of splenomegaly**

Table 6: Summary of clinical associations with different flow cytometric classifications.

Efforts continue to attempt to better classify CVID patients. B and T cell classification systems have thus far not shown any significant ability to change patient management, and the current methods described above are of limited utility. The ability to reliably classify CVID patients would be extremely useful, from both the clinical and research points of view. The determination of accurate predictive markers would allow appropriate monitoring and treatment to be targeted to the patients where it is most required. Furthermore, differentiation of CVID into definitive subgroups would enable research which is currently frustrated by the grouping of clinically and biologically distinct conditions together. One potential means of classifying CVID patients is through microarray gene expression technology.

Microarrays

Microarray technology

A microarray consists of a 2-dimensional array of small quantities of biological material placed with an accuracy measured in microns on a solid substrate such as a glass slide. Microarray technology allows multiplex testing with millions of probes and requires only a small amount of sample material. Examples of biological material which can be used as a probe on microarrays include oligonucleotides, proteins, antibodies and carbohydrates. This project utilised gene expression microarrays using oligonucleotide probes to determine mRNA transcript expression patterns.

Gene expression microarrays are based on a principle developed by Southern, namely that a labelled nucleic acid probe can be used to interrogate other nucleic acid molecules that are fixed to a solid support (Southern 1975). Northern blotting applies

this principle to the detection of mRNA, and is best suited for analysing a limited number of transcripts. Microarrays reverse the principle so that the probe is immobilised on the microarray slide, and the target sample is labelled and hybridised to the fixed probes (Schena *et al.* 1995). Probes are attached to the slide at high density. Each feature consists of hundreds of thousands of identical oligonucleotides present on a spot of diameter of less than 200 microns or on a microbead. Many thousands of features of differing oligonucleotide probes are arranged in an array on the slide. Sample mRNA is converted to labelled cDNA or cRNA (depending on the system) and hybridised to the probes on the microarray. The slide is then scanned using a detection system that quantifies the level of hybridisation of labelled target to each probe.

There are a number of different methods of producing gene expression arrays. Spotted oligonucleotide arrays use synthetic oligonucleotides usually of 20-80 nucleotides long. Long synthetic oligonucleotides have been shown to be as sensitive and specific as cDNA clones (Wang *et al.* 2003). The oligonucleotides are usually prepared into 384 well plates, and are then precisely spotted onto the glass microarray slide using a robotic arrayer, with tight environmental control of temperature and humidity. Robotic arrayers use precision engineered spotting pins, which may be solid or split (Figure 4). Split pins have the advantage of holding a reservoir of oligonucleotide in buffer and so can print multiple spots or features whereas solid pins can only print one spot before having to return to collect more DNA from the well. Advantages of spotted arrays are the precise control over the composition of each array, that they are easily customised and can be manufactured in house.

This image has been removed by the author for copyright reasons

Figure 4: Light microscopy of (a) solid spotting pin and (b) split spotting pin. (Arrayit Corporation, Sunnyvale, CA, USA).

An alternative approach is to synthesise the oligonucleotide probe base by base directly on the solid phase substrate. This is the approach taken by Affymetrix, where photolithography is used to direct each round of synthesis so that the base is added to the correct feature on a quartz substrate. Each nucleotide added to the oligonucleotide has a photo-chemically removable protective group on its 5' position to prevent the addition of further nucleotides. Localised photo-deprotection is achieved by the projection of light through a mask which directs each base addition to the correct features. Each step requires a different mask, which are expensive to produce, but once made can be used to manufacture a large number of identical arrays. These arrays are designed and synthesised purely on information from gene databases and do not require physical intermediates such as clones, PCR products or cDNAs. This technology is suited to the production of large numbers of "standard" arrays that can be widely used in many applications. The disadvantages of *in situ* synthesised oligonucleotide arrays are the cost, difficulties in customisation and that they require large quantities of starting mRNA (2-10μg) for hybridisation.

The above arrays are manufactured with features printed/produced in predefined locations. Illumina BeadChip arrays are based on synthesised oligonucleotides attached to microbeads which are then used to produce microarrays using a random self-assembly mechanism (Gunderson *et al.* 2004). Hundreds of thousands of identical oligonucleotides are attached to each microbead. The first 29 bases act as an address label and the next 50 bases act as the prove (Figure 5). The microbeads are pooled and then added over the surface of slides which have been etched with arrays of microwells of approximately 3 microns depth. The microbeads spontaneously assemble into the wells in a random order, and are held in place through a combination of Van der Waals forces and hydrostatic interactions. There is an excess of microwells (>1.6 million) such that each bead/oligonucleotide would be expected to be represented >30 times per array which provides internal technical repeats within each array. The next step is to sequentially hybridise each slide using the address region to build a map of which oligonucleotide is in each position (Gunderson *et al*. 2004). This process confirms that every oligonucleotide is present with an adequate number of repeats and also confirms the hybridisation performance of the address region of each bead. The information from the address regions is used to decode/map each slide, and when combined with the data from hybridisation of the target sample to the probe region provides a gene expression profile. In addition, there are a number of control beads which bind to RNA samples spiked into experimental sample to confirm labelling success, hybridisation stringency and signal generation. The BeadChips require small amounts of starting mRNA for each array (100ng), and have been shown to produce comparable data to other gene expression microarray methods (Barnes *et al.* 2005; Du *et al.* 2009).

This image has been removed by the author for copyright reasons

Figure 5: Diagrammatic representation of a microbead. Hundreds of thousands of identical oligonucleotides are attached to each bead. The oligonucleotides consist of an address region and a probe region complementary to the cRNA of interest. (Illumina Inc., San Diego, CA, USA).

Applications of Gene Expression Microarrays

Gene expression profiling has numerous applications and can be used to improve our understanding of biological processes. For example, determining changes in gene expression following drug administration contributes to the understanding of mechanism of action (Butte *et al.* 2000; Gerhold *et al.* 2002). Microarrays can identify gene expression patterns which are necessary for an efficient immune response against particular pathogens, such as the relative over-expression of genes encoding for granzymes and successful suppression of HTLV-1 (Vine *et al.* 2004). In tumour biology, the profiling of cell lines over-expressing oncogenes (Harkin *et al.* 1999) or treated with growth factors (Fambrough *et al.* 1999) can help with the determination of oncogenic mechanisms, explore downstream signalling pathways and identify possible therapeutic targets. Gene expression patterns can also be used as a means of providing a molecular phenotype. Applied to specific cancers previously recognised as identical by clinical and laboratory classifications, they have allowed further subdivisions which have yielded additional diagnostic and prognostic information. Examples include breast cancer (Perou *et al.* 2000; Sorlie *et al.* 2001), leukaemias (Golub *et al.* 1999) and lymphomas (Alizadeh *et al.* 2000), and gene expression microarrays have proved powerful tools for subclass discovery and prediction.

To date, there has been limited application of gene expression microarrays in CVID. High throughput genotyping of genomic DNA by genome-wide association study (GWAS) and copy number variations (CNVs) has been applied in a study of 363 CVID patients (Orange *et al.* 2011). This identified strong associations with CVID and polymorphisms in the MHC region, polymorphisms in a disintergrin and metalloproteinase (ADAM) genes and duplications of origin of recognition complex 4L (ORC4L). A number of single nucleotide polymorphisms (SNPs) led to an increased risk of various complications, and while some appear to be chance associations, a number are of potential interest. One example is sorting nexin 31 (SNX31) which may be involved in protein trafficking and also required for CD28 mediated T cell co-stimulation. Polymorphism of SNX31 was associated with organ specific autoimmunity (OSAI). The data from this study was able to allow prediction of the presence of CVID, and the diversity of genetic findings supports the hypothesis that CVID is a collection of heterogeneous conditions. However, this study of genomic DNA was not able to differentiate between the different complications and clinical subgroups within CVID. The examination of genomic DNA investigates the presence of pre-existing genetic predispositions, whereas gene expression examines the current transcriptional profile and is influenced by currently active cellular processes.

One study which examined gene expression investigated the expression of 5197 genes in anti-CD3 and anti-CD28 stimulated PBMC from one CVID patient compared to a healthy control and showed changes in the expression of immunoglobulin genes, and in genes involved in the Th1 pathway (Qin *et al.* 2003). While the authors suggested that skewing towards an excessive Th1 response may lead to decreased B cell help by

Th2 cells in CVID, analysis of more than one patient is required. Another study examined gene expression in sorted T cells from 6 CVID patients who had been selected on the basis of having poor T cell proliferative responses (Holm *et al.* 2004). Compared to controls, differential gene expression was noted in a number of potentially relevant ontogeny terms including cell cycle, apoptosis, signal transduction, cell-cell communication, adhesion/migration, TCRγδ and cytotoxic T cell. The main finding was of a predominance of $CCR7^-$ T effector/tissue homing cells in CVID patients with poor T cell proliferation. However, there was no investigation of correlation of the findings with clinical phenotype. The most recent study looked at the expression of 406 genes on a targeted array between 10 controls and 10 Rome group I patients and 10 Rome group III patients (Giovannetti *et al*. 2007). They concluded that there were discreet genetic profiles between the CVID subgroups, and in particular CD9 and adhesion molecules. However, these genes were identified by whether they were more than twofold up- or down-regulated compared to the controls, without any statistical analysis nor correction for multiple testing.

Indications for present research

The clinical care of CVID patients and research into the pathogenesis of the condition (s) are complicated by the inability to properly classify this heterogeneous disease. Prediction of patients at high risk of granulomatous disease, autoimmunity and splenomegaly would allow closer clinical monitoring of those who require it, and less intensive monitoring of those at low risk. If the risk of bronchiectasis could be stratified, then those at low risk could be spared radiation exposure from repeated computer tomography (CT) investigations, while management strategies could be put into place for those at high risk. At the moment, aetiological research is made

problematic by the heterogeneity of CVID and the fact that this “catch-all” diagnosis seems likely to include a number of different conditions. Better differentiation of CVID would enable more effective research into what are likely to be different pathogenic mechanisms in the different groups.

While progress has been made with flow cytometric techniques to differentiate CVID subgroups, these remain of limited utility. Microarrays allow the investigation of the expression of the whole genome, and are not limited for example by the number of channels on a flow cytometer. It is therefore easier to examine multiple parameters by microarray techniques, where both known and potentially unknown contributory factors can be examined. While the analysis of sorted cells (for example T cells and B cells) may provide information regarding alterations in gene pathways, such an approach would miss potentially useful classification criteria, and would be difficult to implement as a routine assay in a clinical laboratory. As well as detecting disease specific transcriptional differences, the investigation of the transcriptome of whole blood can also detect altered frequencies of normal blood cell types and the presence of abnormal cell types, and therefore this was the approach taken in this project. In conjunction with bioinformatic techniques, microarrays represent a potentially powerful tool in complex conditions such as CVID. The ideal scenario would be the identification of a relatively small number of genes, the expression of which could be used for subgroup prediction. This information could then be used to develop a targeted mini/midi-plex array for this purpose, or may inform an alternative flow cytometric strategy for classification.

Hypothesis

The hypothesis under investigation is that examination of gene expression through microarray technology will enable the identification of a subset of genes which would allow classification of CVID patients into clinically distinct subgroups. These subgroups are more likely to be aetiologically distinct than current classification methods allow.

Aims

- Request the fully informed consent of patients in the primary antibody deficiency cohort (including CVID, XLA and other causes) attending the Welsh National Immunology Service at University Hospital Wales

- Establish a biobank of stored RNA from those who consent to participate and confirm the quality of RNA extracted

- Collate the clinical and immunological data from participants to enable correlation of data generated by the analysis of RNA using microarray gene expression technology

- Examine RNA expression through microarray technology and subject results to bioinformatic analysis

- Determine whether the differential expression of a limited number of markers allows accurate subgroup classification

Materials and Methods

Study subjects

Primary antibody deficient patients attending the Welsh National Immunology Service at University Hospital Wales were recruited to the study. Copies of the patient information sheet and consent form can be found in appendices 2 and 3 respectively. Local research ethical committee (LREC) approval was granted for the study. Patients were provided with the information sheet in advance and then given the opportunity to ask a member of the Clinical Immunology team any questions arising as part of the consent process. Patients with CVID were identified as per agreed international diagnostic criteria (Conley *et al*. 1999).

Complications were identified through a combination of clinical, laboratory and radiological identifiers. Bronchiectasis was confirmed by the presence of characteristic features on high resolution computer tomography (HRCT). Autoimmune disease was defined as per diagnostic criteria for the particular condition. Granulomatous disease was defined by the presence of granuloma on histology, or in the situation where radiological investigations had identified lesions consistent with granuloma in the absence of other underlying cause. Persistent lymphadenopathy was defined as ≥2 enlarged lymph nodes of >1cm in diameter on CT or on examination and present for >1 year.

Flow Cytometry

All flow cytometry was performed on the same Beckman Coulter FC 500 flow cytometer and results analysed using CXP software (Beckman Coulter). Daily quality control was performed with Flow-Check and Flow-Set Fluorospheres (Beckman Coulter, 6605359, 737664, 737663, 6607007) and monthly with Immuno-Brite and Immuno-trol standards (Beckman Coulter, 6603473, 6607077, 6607098) as per manufacturers instructions. Lymphocytes were gated using CD45 versus side scatter.

Lymphocyte phenotyping was performed by staining 50µl of EDTA blood with 5µl CD45-FITC/CD4-PE/CD8-ECD/CD3-PC5 (tube 1) (Beckman Coulter, 6607013) and 5µl CD45-FITC/CD56-PE/CD19-ECD/CD3-PC5 (tube 2) (Beckman Coulter, 6607073) incubated at room temperature (18-23ºC) and protected from light for 10 minutes. Red cell lysis was performed by incubation with 500µl of Versalyse (Beckman Coulter, A09777), for 10 minutes at room temperature and protected from light. 50µl of Flow-Count Fluorospheres (Beckman Coulter, 7547053) added to each tube was used for enumeration and the data acquired immediately.

Naïve T cells were assessed using the following antibody panel optimised using negative and isotype controls: CD4-PE, CD3-PC5 and CD45RA-ECD (Beckman Coulter, A07751, A07749, IM2711U) with either CD27-FITC (Serotec, MCA755F) or CD62L-FITC (Coulter, IM1231U). 5µl of each antibody were added to 50µl EDTA blood which was processed as for lymphocyte phenotyping with the omission of the enumeration step. If immediate acquisition was not possible, the samples were washed twice in PBS (Oxoid, BR14a) (300g, 5 minutes, room temperature) and re-suspended in 500µl PBS/1% paraformaldehyde (Sigma, P-6148).

Memory B cell phenotyping was performed parallel to lymphocyte phenotyping. Staining was performed with CD19-PC7 (Beckman Coulter, IM3628), CD27-FITC (Serotec, MCA755F), CD21-PE (BD Pharmingen, A07778), CD38-FITC (Beckman Coulter, A07778), IgM-Cy5 (Jackson ImmunoResearch, 109-175-129) and IgD-PE (Southern Biotech, 2030-09). PBS/0.1% Sodium Azide (Sigma, S8032) was used to dilute CD27-FITC (1:5), IgD-PE (1:40) and IgM-CY5 (1:5). Antibody mixes B1 (20µl CD27-FITC, 20µl IgD-PE, 4µIgM-CY5 and 10µl CD19-PC7) and B2 (20µl CD21-PE, 20µl CD38-FITC, 4µl IgM-CY5 and 10µl CD19-PC7) were prepared. 1ml of EDTA blood was washed three times in 3mls PBS/Azide (300g, 5 minutes, room temperature), followed by re-suspension in 1ml of PBS/0.1%BSA (Sigma, A7906-100G). Two tubes were prepared and 100µl of washed blood was incubated with 15µl of B1 and B2 for 20 minutes at room temperature protected from light, followed by the addition of 1ml of Versalyse and a further 20 minute incubation. The samples were washed twice in PBS/Azide and fixed in PBS/1% paraformaldehyde and immediately acquired.

RNA Extraction, Storage and Integrity Assessment

Venous blood from study participants was collected into PAXgene blood RNA tubes (PreAnalytiX, 762165) and stored at room temperature overnight. RNA was extracted using the PAXgene Blood RNA kit v2 (PreAnalytiX, 762174) as per manufacturers instructions. The concentration of extracted RNA was determined by spectrophotometry (Nanodrop Technologies, ND-1000) and then logged, banked and stored at -70ºC. Prior to microarray analysis, the quality of RNA was confirmed by electropherogram (Agilent Technologies, 2100 Bioanalyser).

Production of spotted microarrays

Targeted arrays were produced using 1593 oligonucleotide probes (Qiagen Operon) prepared in 5 x 384 microarray well plates (Genetix, XGE05080) at a final concentration of 40µM in 30µl 1X Nexterion spot buffer (Schott, 1066029). Spotting pins were cleaned by 10 minutes sonication in 2% pin cleaning solution (Genetix, K2505), followed by washing with de-ionised water. Oligonucleotides were spotted onto epoxysilane coated glass slides (Schott, Nexterion Slide E, 1064016) using a QArray robotic spotter (Genetix), and then stored overnight in the dark at a humidity of 70-80%. Scanning was performed with a GenePix4000B microarray scanner (Molecular Devices) and images analysed with GenePix Pro 6.1 (Molecular Devices).

Synthesis of labelling cDNA and hybridisation to spotted arrays

Superscript plus direct labelling kit core module (Invitrogen, 46-6345), nucleotide module (Invitrogen, 46-6347) and low elution cDNA purification module (Invitrogen, 46-6346) were used as per manufacturer's instructions with 5-10µg RNA to produce cDNA labelled with either Alexa Fluor 555 or 647. Incorporation of labelled oligonucleotide was assessed by spectrophotometry (Nanodrop Technologies, ND-1000). Each hybridisation required 20pmol of dye in 3µl volume to which 28µl of Nexterion hybridisation buffer (Schott, 1066075) and 1 µl COT1-DNA (Invitrogen, 15279-011) was added and then incubated (95ºC, 3 minutes). The spotted arrays were washed and blocked according to manufacturer’s instructions, and 28µl of prepared solution was hybridised to the array (65ºC, 16 hours) in a hybridisation cassette (ArrayIt, AHCA). Post hybridisation washes were performed with saline sodium citrate buffer (SSC) (Eppendorf, 0032 006.655) at 2X and 0.2X concentration as per manufacturer’s instructions. The slides were dried under centrifugation (200g, room

temperature, 5minutes). Scanning was performed with a GenePix4000B microarray scanner (Molecular Devices) and images analysed with GenePix Pro 6.1 (Molecular Devices).

Synthesis of labelled cRNA and hybridisation to BeadChip gene expression array

cDNA was synthesised from 100ng extracted RNA and transcribed to produce biotinylated cRNA using the Illumina TotalPrep RNA amplification kit (Ambio, IL1791). Hybridisation to HumanRef-8 v3.0 Expression BeadChip (Illumina, BD-102-0203) was performed as follows. 750ng of cRNA in a volume of 15µl was heated (65ºC, 5 minutes) and allowed to cool before application to the array and incubated (58ºC, 16 hours) in a BeadChip Hyb Chamber (Illumina, 210948, 222682, 210930). Washes and labelling with streptavidin-Cy3 (Amersham Biosciences, PA43001) were performed as per manufacturer's instructions, and slides dried under centrifugation (275g, room temperature, 4 minutes). Scanning was performed with a BeadArray Reader (Illumina).

Statistics

Fisher's exact test, Pearson's product-moment correlation coefficient, Spearman's correlation, linear regression and Mann Whitney test were performed using Prism version 3.0 (GraphPad Software Inc.). Alpha values of 0.05 and 0.01 were used.

Microarray data analysis

The performance of quality controls (biological, sample labelling, hybridisation and negative controls) were checked with GenomeStudio version 2010.1 (Illumina) prior to analysis of the data.

The microarray data was quantile normalised using GenomeStudio version 2010.1 (Illumina) and then log transformed and exported using the Partek Gene Expression Report Plug-in version 2.16.0.0 (Partek). Subsequent analyses including hierarchical clustering, principal component analysis, batch effect removal, two sample t-test, multiple test corrections and class prediction were performed using Partek Genomic Suite version 6.3 (Partek).

Class discovery analysis was performed by Peter Giles, Central Biotechnology Services, Cardiff University, utilising ConsensusClusterPlus (Bioconductor Project) (Monti *et al.* 2003; Wilkerson and Hayes 2010).

Biological Analysis

Biological analysis of the microarray data, including functional annotation, pathway analysis and gene enrichment was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 (accessed at http://david.abcc.ncifcrf.gov/home.jsp) (Dennis *et al.* 2003; Huang *et al.* 2009) , with the background gene list Human Ref-8_V3_0_R2_11282963_A as a comparator.

Results Chapter 1

Clinical features of study subjects

Study subjects

There was the capacity to perform microarray analysis on 64 samples (see Results Chapter 3). In addition to the 53 CVID patients recruited to the study, a number of other patient samples were also analysed. Two patients with specific antibody deficiency (with normal total immunoglobulin levels) were included as this condition probably forms part of the spectrum of primary antibody deficiencies. Sample 70 came from a patient who was hypogammaglobulinaemic but did not meet the criteria in Table 3 for a diagnosis of CVID. Five well characterised XLA patients with proven deficiency of btk were also included as these patients may provide a useful comparison point and the pathogenesis of their condition is well understood. Thus a total of 61 patients, all of whom were of Caucasian descent and were all established on immunoglobulin replacement therapy were studied. Table 7 provides a summary of the clinical features of these patients

Though prior data has shown technical variability to be smaller than biological variability in gene expression studies in humans (Cheung *et al.* 2003), the use of technical replicates remains potentially useful and the following 3 replicates were included:

1) 80X – a second aliquot identical to sample 80
2) 80T – a sample from patient 80 taken at a different point in time (3 months later)
3) 96X – a second aliquot identical to sample 96

Patient ID	Sex	Age	Dx	Other info	Gran	Autoimmunity	S'ctomy	S'megaly	PL	Bronch	prior to dx	worsened on Ig	Smoker (pack years)
01	F	70	CVID	MZL (2006)	N	Y (ITP)	N	N	Y	Y	Y	N	N
03	F	51	CVID		N	N	N	N	N	Y	Y	-	N
04	F	69	CVID		N	Y (temporal arteritis)	N	N	N	Y	-	-	N
07	F	46	CVID		N	N	N	N	N	Y	N	Y	N
08	M	60	CVID		N	N	N	N	N	Y	-	Y	N
09	M	71	CVID		N	N	N	N	N	Y	Y	-	N
10	F	38	CVID		N	N	N	N	N	N	-	-	N
11	M	45	CVID		Y	Y (AHA)	Y	Y	Y	Y	-	-	N
13	M	67	CVID		Y	N	Y	Y	N	N	-	-	N
14	M	35	CVID		N	Y (Crohns)	N	N	N	Y	-	-	Y (10)
23	M	44	CVID		N	Y (UC)	N	Y	N	N	-	-	N
26	M	42	XLA	recurrent ca colon	N	N	N	N	N	N	-	-	N
27	M	50	CVID		N	N	N	N	N	Y	-	-	N
28	F	55	CVID		Y	N	Y	Y	Y	Y	-	-	N
29	F	61	CVID		N	N	N	N	N	Y	Y	-	N
30	F	32	CVID		Y	Y (ITP & neutropenia)	Y	Y	N	N	-	-	N
31	M	49	CVID		N	N	N	N	N	N	-	-	Y (40)
32	M	47	CVID		N	N	N	N	N	Y	Y	N	Y (20)
34	M	56	CVID		N	N	N	N	N	N	-	-	N
35	F	47	CVID		N	N	N	N	N	Y	Y	N	Y (9)
36	F	64	SPAD		N	N	N	N	N	N	-	-	Y (40)
39	M	16	XLA		N	N	N	N	N	Y	N	Y	N
40	F	32	SPAD		N	N	N	N	N	N	-	-	N
42	M	26	CVID		N	N	N	N	N	N	-	-	N
43	M	45	CVID		N	N	N	N	N	N	-	-	N
44	M	54	CVID		N	Y (ITP)	Y	Y	N	N	-	-	N

Patient ID	Sex	Age	Dx	Other info	Gran	Autoimmunity	S'ctomy	S'megaly	PL	Bronch	prior to dx	worsened on Ig	Smoker (pack years)
45	F	65	CVID	CLL (2006), haemochromatosis	N	N	N	N	N	N	-	-	N
46	F	34	CVID		N	Y (ITP)	Y	Y	N	Y	-	-	Y (18)
49	F	75	CVID		N	N	N	N	N	N	-	-	N
50	F	30	CVID	breast ca (2008)	N	N	N	N	N	Y	-	-	N
51	F	57	CVID		N	Y (pernicious anaemia)	N	N	N	Y	-	-	N
52	F	44	CVID		N	N	N	Y	N	Y	-	-	N
53	M	45	CVID		N	Y (AHA)	Y	Y	N	Y	-	Y	Y (?)
56	F	32	CVID	mitochondrial myopathy	N	N	N	N	N	N	-	-	N
58	M	47	CVID		N	N	N	N	N	Y	-	-	N
59	M	27	CVID		N	N	N	N	N	Y	Y	N	Y (7)
61	M	45	CVID		N	N	N	N	N	Y	-	N	N
63	M	39	CVID		N	N	N	N	N	Y	-	N	N
64	M	37	XLA		N	N	N	N	N	Y	N	Y	N
66	M	74	CVID	developed CLL 2 years after RNA extraction	N	N	N	N	N	Y	Y	N	Y (21)
67	F	74	CVID		Y	Y (ITP and neutropenia)	Y	Y	N	Y	-	N	N
68	M	52	CVID	x2 PE's	N	N	N	Y	N	Y	-	-	Y (34)
69	F	49	CVID		N	Y (ITP)	N	Y	Y	N	-	-	N
70	F	41	hypog		N	Y (ITP)	Y	Y	N	Y	Y	N	Y (30)
72	M	28	XLA		N	N	N	N	N	Y	Y	N	N
73	M	24	XLA		N	N	N	N	N	N	-	-	N
74	F	69	CVID		N	N	N	N	N	N	-	-	Y (18)
77	F	51	CVID		N	N	N	N	Y	N	-	-	N

Patient ID	Sex	Age	Dx	Other info	Gran	Autoimmunity	S'ctomy	S'megaly	PL	Bronch	prior to dx	worsened on Ig	Smoker (pack years)
78	F	75	CVID		N	Y (ITP)	N	N	N	Y	-	-	Y (35)
80	M	53	CVID		N	N	N	Y	N	N	-	-	N
82	M	17	CVID		N	Y (eczema)	N	N	N	N	-	-	N
84	F	26	CVID		N	N	N	N	N	N	-	-	N
85	M	18	CVID		N	N	N	N	N	Y	-	-	Y (3)
87	M	46	CVID		N	N	N	N	N	N	-	-	N
88	F	18	CVID		N	N	N	N	Y	N	-	-	N
89	M	40	CVID	Gilbert's syndrome	N	N	N	N	N	N	-	-	N
90	M	32	CVID		N	N	N	N	N	N	-	-	N
91	F	62	CVID		N	N	N	N	N	N	N	Y	N
94	M	39	CVID	Idiopathic T cell lymphopenia	Y	N	N	Y	Y	Y	Y	-	N
95	M	75	CVID		N	N	N	N	N	Y	-	-	Y (15)
96	F	60	CVID		N	N	N	N	N	Y	Y	-	Y (32)

Table 7: A summary of the patients included in the microarray analysis. Dx: diagnosis, Gran: granulomatous disease, S'ctomy: splenectomy, S'megaly: splenomegaly, PL: persistent lymphadenopathy, Bronch: bronchiectasis, prior to dx: was bronchiectasis present prior to diagnosis, worsened on Ig: did bronchiectasis worsen or develop while on immunoglobulin replacement therapy, -: information not available from notes, ITP: immune thrombocytopenia, AHA: autoimmune haemolytic anaemia, MZL: marginal zone lymphoma, ca = carcinoma, CLL: chronic lymphocytic leukaemia, PE: pulmonary embolism, UC: ulcerative colitis.

XLA patients

Five XLA patients were recruited to the study, with an age range of 16-42 and an average age of 29.4. All five patients have been confirmed to be negative for btk protein expression by flow cytometry, have absent B cells and were agammaglobulinaemic prior to commencement of immunoglobulin replacement therapy. All the patients are Caucasian (and all are male as expected for an X-linked condition).

One of these patients (patient ID 26) has suffered with multiple episodes of carcinoma of the colon at multiple different sites (2003, 2006 and 2010) which has now metastasised. He has received surgery and is continuing to receive chemotherapy. There is no family history of colon cancer, and polyposis syndromes such as Peutz-Jegher's syndrome and familial adenomatous polyposis have been excluded. There are reports of an increased risk of colorectal carcinoma in XLA (Brosens *et al.* 2008; van der Meer *et al.* 1993). The presence of metastatic carcinoma of the colon in addition to receiving chemotherapy may be expected to change the patient's profile on microarray analysis and so data from this particular patient will have to be interpreted in light of the clinical situation.

Three of the five patients have or have had bronchiectasis. Patient 39 developed bronchiectasis prior to being diagnosed and started on treatment at the age of 18 months. However, his bronchiectasis has continued to worsen despite immunoglobulin therapy, suggesting the importance of early intervention and preventing the development of bronchiectasis in the first place. Patient 64 developed bronchiectasis despite being diagnosed at birth and commenced on immunoglobulin

replacement therapy (due to a family history of XLA). However, his trough IgG levels have been persistently low, and on recent pharmacokinetic studies a short IgG half life of less than 7 days has been found. Normal IgG half life is in the order of 19-21 days (Eibl and Wedgwood 1989). Possible causes for the reduced half life (resulting in low trough levels) are currently under investigation. The third patient (patient ID 72) was diagnosed and started on treatment at the age of 3 years, and had already developed bronchiectasis. The bronchiectasis was anatomically limited and following left lower lobectomy at the age of 11, he has been clinically well and there is no evidence of bronchiectasis elsewhere on HRCT. None of the five XLA patients have granuloma, autoimmunity, splenomegaly or persistent lymphadenopathy, and these complications are not associated with XLA.

CVID Cohort

The 53 CVID patients consisted of 25 females and 28 males, with an age range of 17-75 and average age of 48.8. Two patients had $CD4^+$ T cell counts of <200 cells/µl. However, there was no history of opportunistic infection(s) consistent with a defect of cell-mediated immunity and therefore these two patients were included in the CVID cohort.

Six patients (3 male and 3 female) had granulomatous disease, a complication rate of 11.3%. This compares to published rates which vary from around 8 to 22% (Ardeniz and Cunningham-Rundles 2009; Morimoto and Routes 2005). None of these 6 patients have required specific treatment (such as steroids or other immune suppressants) for their granulomatous disease.

14 CVID patients (6 male and 8 female) in the cohort had one or more autoimmune conditions, a rate of 26.4% which is similar to previously published rates of around 20% (Cunningham-Rundles 2008). The commonest autoimmune complication was ITP, with 7 patients affected (15.1% of all CVID patients). Two of the patients with ITP also had neutropenia. There were 6 females and 1 male with ITP, which contrasts with the published literature where there is not a female preponderance. There were 2 cases of AHA, and other autoimmune conditions were pernicious anaemia, temporal arteritis, eczema, Crohn's disease and UC (all 1 patient per condition). Table 8 summarises the autoimmune conditions.

Type	Number (% of CVID patients)	Male	Female
ITP	7 (13.2)	1	6
Autoimmune neutropenia	2 (3.8)	0	2
AHA	2 (3.8)	2	0
Pernicious anaemia	1 (1.9)	0	1
Temporal arteritis	1 (1.9)	0	1
Eczema	1 (1.9)	1	0
Crohn's disease	1 (1.9)	1	0
UC	1 (1.9)	1	0

Table 8: Autoimmune conditions in the CVID cohort. There were 16 conditions in 14 patients.

Splenomegaly was or had been present in 14 patients (8 male and 6 female), a rate of 26.4%. Splenectomy had been performed on 8 patients (4 male and 4 female). The most common indication was for treatment of autoimmune disease not controlled by steroids or other treatments, though 3 patients had splenectomy performed for management of abdominal discomfort due to massive splenomegaly.

There is known to be an increased risk of haematological malignancies in CVID patients, and of the 53 patients in this cohort 1 developed marginal zone lymphoma (MZL) and 2 developed chronic lymphocytic leukaemia (CLL). There was also 1 case

of carcinoma of the breast. While there is a statistically significant increased risk of adenocarcinoma of the stomach, other solid tumours occur with too low a frequency to determine whether there is an increased risk in CVID (Cunningham-Rundles and Bodian 1999; Quinti *et al*. 2007).

Bronchiectasis, as defined by HRCT, was a common complication with 30 patients (16 male and 14 female) affected. In total 56.6% of CVID patients had bronchiectasis. Unsurprisingly, being a smoker or ex-smoker was a significant co-factor in the development of bronchiectasis. 85.7% of CVID patients with a smoking history had bronchiectasis compared to 46.2% of non-smokers. Information on the timing of development of bronchiectasis is available for 13 of the CVID patients. 11 patients developed bronchiectasis prior to diagnosis with only 2 patients developing bronchiectasis after treatment for CVID had been initiated. For 13 patients it was possible to determine whether bronchiectasis had worsened while they were on immunoglobulin replacement therapy by using a combination of lung function test results, CT scans and clinical information such as exercise tolerance and daily sputum production. Though 4 patients showed deterioration while on immunoglobulin, 9 patients remained stable. The above emphasises the importance of avoiding diagnostic delay in patients with CVID and the role of immunoglobulin replacement therapy to help prevent end organ lung damage.

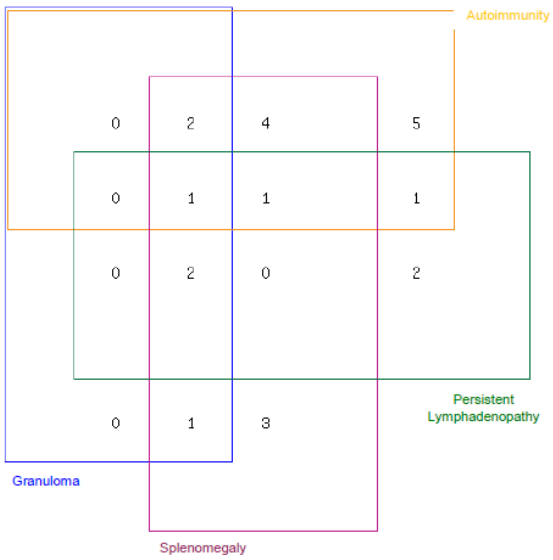
The presence of one of granuloma, autoimmunity, splenomegaly or persistent lymphadenopathy appears to increase the risk of the other complications being present. If there was no association between these complications, you would expect, for example, the complication rates of autoimmunity, splenomegaly and

lymphadenopathy to be the same in CVID patients with and without granulomatous disease. This is not the case, and the same holds true when comparing patients with and without autoimmunity, with and without splenomegaly and with and without lymphadenopathy (Figure 6). Using a null hypothesis that complication rates are not higher if another complication is present, the following associations reach statistical significance using Fisher's exact test for categorical variables: granuloma with splenomegaly, granuloma with lymphadenopathy, autoimmunity with splenomegaly, splenomegaly with granuloma, splenomegaly with autoimmunity and lymphadenopathy with granuloma (Table 9). It is possible that there are other associations present, but that the low number of patients with granulomatous disease and persistent lymphadenopathy (n=6 and n=7 respectively) meant that they did not reach statistical significance.

There is no increased risk of these complications associated with the presence of bronchiectasis or vice versa. Risk factors for bronchiectasis include recurrent infections, smoking history, ineffective immunoglobulin replacement, poor patient compliance with treatment and other potential factors. The development of bronchiectasis appears to be distinct from the other complications, which is perhaps unsurprising given that bronchiectasis represents end organ damage as a consequence of recurrent infections, rather than a distinct disease process. The independence of bronchiectasis from the other complications acts as a negative control and reinforces the finding that there is an increase in risk of granuloma, autoimmunity, splenomegaly and lymphadenopathy if one of these conditions is already present. This association suggests that the development of granulomatous disease, autoimmunity, splenomegaly and lymphadenopathy may share common mechanisms of pathogenesis.

	Granuloma	Autoimmunity	Splenomegaly	Lymphadenopathy
Granuloma	-	0.181	0.0001	0.025
Autoimmunity	0.181	-	0.0045	0.264
Splenomegaly	0.0001	0.0045	-	0.070
Lymphadenopathy	0.025	0.264	0.070	-
Bronchiectasis	0.47	0.36	0.61	0.65

Table 9: p values by Fisher's exact test using the null hypothesis that complication rates are not higher if there is another complication present. The following associations reach statistical significance of $p<0.05$: granuloma with splenomegaly, granuloma with lymphadenopathy, autoimmunity with splenomegaly, splenomegaly with granuloma, splenomegaly with autoimmunity and lymphadenopathy with granuloma.



Group	Number
All Granuloma	6
All Autoimmunity	14
All Splenomegaly	14
All Lymphadenopathy	7
Granuloma only	0
Autoimmunity only	5
Splenomegaly only	3
Lymphadenopathy only	2
Granuloma + Autoimmunity	0
Granuloma + Splenomegaly	1
Autoimmunity + Splenomegaly	4
Granuloma + Lymphadenopathy	0
Autoimmunity + Lymphadenopathy	1
Splenomegaly + Lymphadenopathy	0
Granuloma + Autoimmunity + Splenomegaly	2
Granuloma + Autoimmunity + Lymphadenopathy	0
Granuloma + Splenomegaly + Lymphadenopathy	2
Autoimmunity + Splenomegaly + Lymphadenopathy	1
Granuloma + Autoimmunity + Splenomegaly + Lymphadenopathy	1

Figure 6a: A Venn diagram summarising the distribution of granuloma, autoimmunity, splenomegaly and persistent lymphadenopathy seen in the cohort.

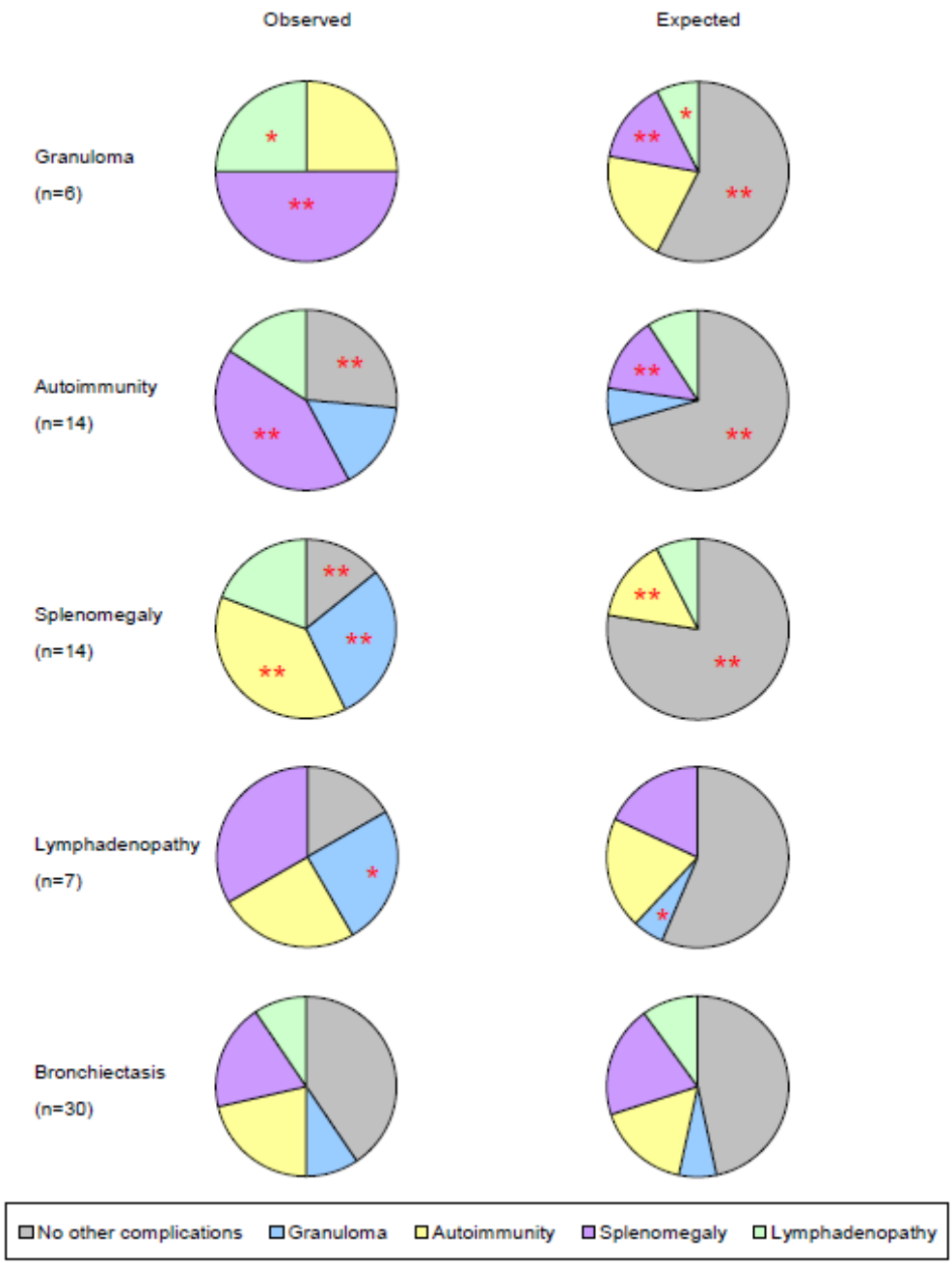


Figure 6b: The observed rate of other complications in patients with granuloma, autoimmunity, splenomegaly and persistent lymphadenopathy are contrasted with those which would be expected if the complication rates were the same whether other complications were present or not. *p<0.05 **p<0.01

Results Chapter 2

Immunological features and classification of CVID patients

B and T cell classification of the CVID cohort

Peripheral B cell phenotyping was performed on the 42 of 53 CVID patients, and patients classified by the Freiburg, Paris and EUROclass methods (as described in Figure 3) with examples shown in Figure 7.

Decreased numbers of CD4 naïve T cells are associated with abnormalities of both CD4 and CD8 TCR repertoires (Giovannetti *et al*. 2007). Therefore it is of interest to determine whether naïve CD8 and naïve CD4 T cell numbers correlate, and whether they add any further information regarding the classification of CVID patients. Naïve CD4 T cells were enumerated as shown in Figure 8. This was historical data gathered across 4 years. While the appearance of the flow cytometry plots suggests that the laser voltage was higher than necessary, the separate populations remain clear and with a distinct difference between patients with low proportions of naïve $CD4^+$ T cells and those with normal proportions.

The Rome classification describes identification of naïve CD4 T cells by a $CD45RA^+CD62L^+$ phenotype. CD27 positivity can also be used to identify naïve T cells (Kobata *et al.* 1994), and unlike CD62L is not rapidly shed from lymphocytes should they become activated during handling, or should there be a delay between venesection and the sample arriving in the laboratory. In a comparison of 27 sequential samples received for naïve T cell numbers (from all types of patients),

results from using CD27 correlated with CD62L with a Pearson product-moment correlation coefficient of r=0.966 (Table 10). For subsequent patients CD62L was no longer used, and $CD45RA^{+}CD27^{+}$ alone was used to identify naïve T cells. CD27 can also be used to identify naïve CD8 T cells. $CD3^{+}CD4^{-}$ T cells were equivalent in number to $CD3^{+}CD8^{+}$ T cells, and therefore $CD3^{+}CD4^{-}CD45RA^{+}CD27^{+}$ describes naïve CD8 T cells. This allows the enumeration of both CD4 and CD8 naïve T cells using a one tube method. Naïve T cells numbers were determined for 21 of the CVID patients.

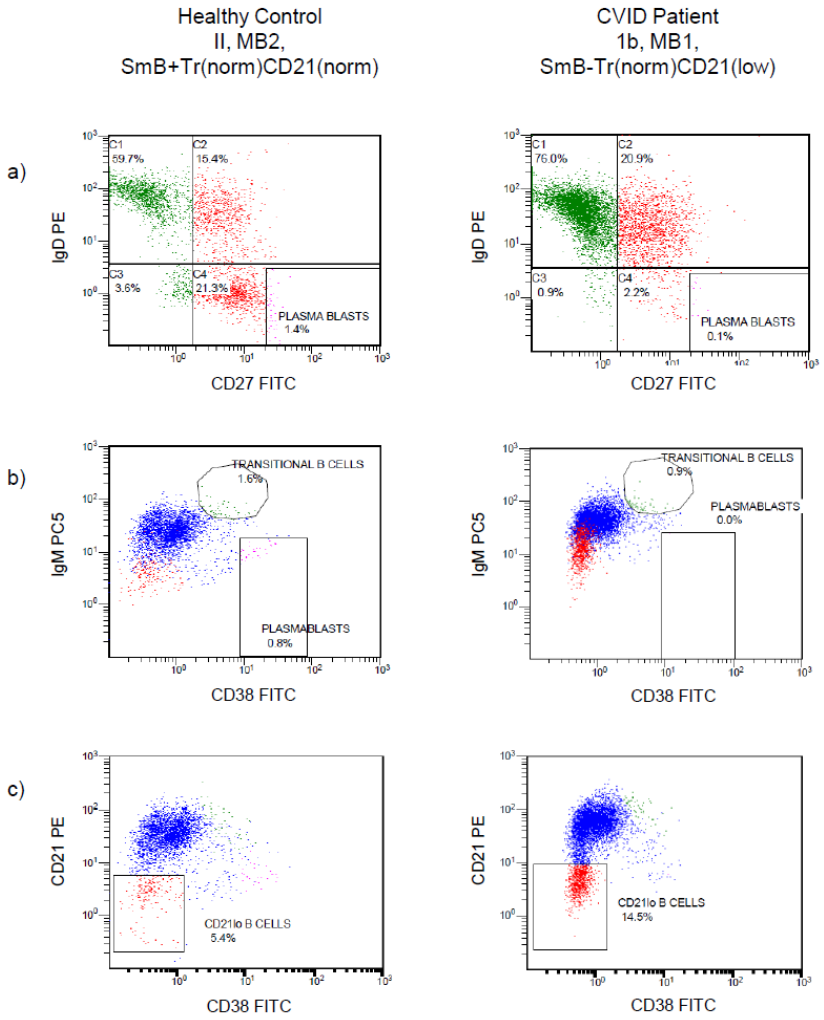


Figure 7: Peripheral B cell phenotypes by flow cytometry of a healthy control and a CVID patient. All plots are shown following gating on B cells using side scatter vs CD19. In scatterplots (a), C1=naïve B cells, C2=non-class switched memory B cells and C4=switched memory B cells.

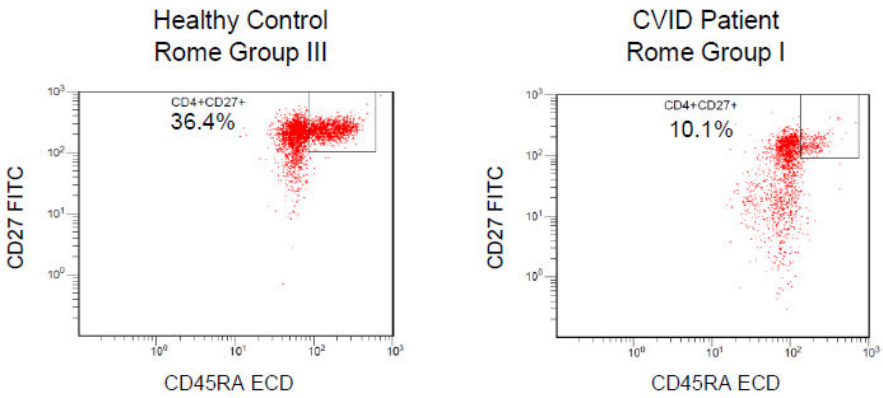


Figure 8: Enumeration of naïve CD4 T cells, following gating for $CD3^{+}CD4^{+}$ T cells. A healthy control and a Rome group I CVID patient are shown.

There was a statistically significant association between naïve $CD4^+$ and $CD8^+$ T cells using Spearman's correlation (non-parametric, one-tailed): for percentages of naïve $CD4^+$ and $CD8^+$ T cells (of total $CD4^+$ and $CD8^+$ T cells respectively) p=0.0014, and for absolute numbers p=0.029. Graphs of naïve $CD4^+$ and $CD8^+$ T cell percentages and absolute numbers are shown in Figure 9. The identification of a relationship between naïve $CD4^+$ and $CD8^+$ T cell levels suggests that in some CVID patients there may additionally be a defect of $CD8^+$ T cell function. Reports of visceral CMV and profuse warts secondary to papillomavirus in CVID patients (Cunningham-Rundles and Bodian 1999; Oksenhendler *et al*. 2008) implies that in addition to the expected susceptibility to bacterial infections as a result of the antibody deficiency, some patients also show increased susceptibility to viral infections, and that further investigation of $CD8^+$ T cells in such patients would be of relevance.

The results of B and T cell phenotyping are found in Table 11. There was no correlation between the severity of B cell phenotyping and severity of T cell phenotyping in these data (e.g. there was no over-representation of Freiburg 1a and/or under-representation of group 2 in the Rome T1 group). The clinical features relating to the different groups are discussed in the next section.

Sample	$CD45RA^{+}CD62L^{+}$ (% of $CD3^{+}CD4^{+}$)	$CD45RA^{+}CD27^{+}$ (% of $CD3^{+}$ $CD4^{+}$)
1	28.7	28.2
2	37.0	42.2
3	23.5	23.1
4	17.3	17.1
5	24.7	25.5
6	6.3	6.5
7	51.7	51.7
8	48.2	41.9
9	44.4	44.1
10	4.1	4.0
11	18.5	16.8
12	10.2	10.1
13	27.7	28.8
14	44.2	44.1
15	54.8	42.3
16	35.4	36.4
17	2.9	2.1
18	7.1	7.1
19	30.1	42.5
20	21.1	21.0
21	27.9	23.6
22	19.7	21.7
23	27.9	27.4
24	21.2	20.5
25	10.5	11.1
26	5.8	5.2
27	8.5	8.4

Table 10: Comparison of CD62L and CD27 for enumeration of naïve CD4 T cells demonstrates similar results. Pearson's r=0.966 (p<0.0001) with a slope of 0.934.

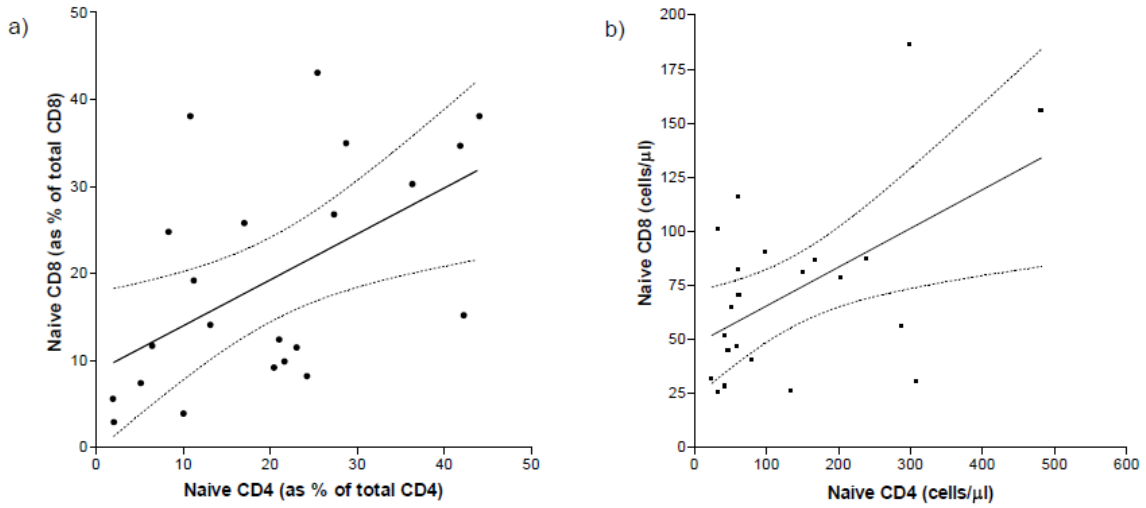


Figure 9: Naïve CD4 and CD8 T cells (a) as a percentage of total CD4 and CD8 T cells respectively and (b) as absolute numbers. Best fit lines by linear regression with 95% confidence interval are shown.

Patient ID	Dx	CD3 cells/µl	CD4 cells/µl	CD8 cells/µl	CD19 cells/µl	CD56 cells/µl	Freiburg	Paris	E UROclass SmB	Tr	CD21	Rome	Naïve CD4 %	cells/µl	Naïve CD8 %	cells/µl
Reference Range		800-2500	400-1500	200-1100	50-500	80-650	2	MB2	+	Norm	Norm	T3	*	*	*	*
1	CVID	1910	690	1170	530	80										
3	CVID	976	707	237	75	238	1b	MB1	-	Norm	Norm	T2	24.3	150.7	8.1	81.0
4	CVID	841	556	259	36	177	1b	MB2	+	Norm	Norm	T3	36.4	203.8	30.2	78.5
7	CVID	872	449	424	108	10	1b	MB1	-	Norm	Low	T1	5.2	23.9	7.3	31.4
8	CVID	797	293	519	181	102	2	MB1	+	Norm	Norm	T2	20.5	59.5	9.1	46.4
9	CVID	1127	571	499	234	129	1b	MB1	+	Norm	Norm	T2	21.7	80.3	9.8	40.2
10	CVID	931	692	221	91	44	1b	MB1	+	Norm	Norm					
11	CVID	4920	1626	3228	860	794						T1	2.1	33.8	2.8	100.8
13	CVID	1229	630	482	266	211	1b	MB1	-	Hi	Low					
14	CVID	2052	1446	539	2	117						T3	41.9	481.9	34.6	155.7
23	CVID	1101	613	426	286	121	1b	MB1	-	Norm	Norm					
27	CVID	2080	1040	960	0	50										
28	CVID	1922	1476	436	546	351	2	MB1	+	Norm	Norm	T1	2.0	33.2	5.5	25.3
29	CVID	1160	436	678	140	48	1b	MB1	-	Hi	Norm					
30	CVID	327	284	38	21	30										
31	CVID	1528	1140	400	120	190	2	MB2	+	Norm	Norm					
32	CVID	751	482	243	104	49	1b	MB1	-	Norm	Norm					
34	CVID	1132	501	575	326	74	2	MB1	+	Norm	Norm	T1	13.2	62.0	14	70
35	CVID	747	466	242	287	64	1b	MB1	-	Norm	Norm	T2	17.1	99.2	25.7	90.0
42	CVID	980	330	580	230	50										
43	CVID	1022	749	257	174	113	1b	MB1	-	Hi	Norm	T1	8.4	47.0	24.7	44.5
44	CVID	1384	228	1169	6	99										
45	CVID	1124	594	465	266	112	II	MB0	+	Norm	Low					
46	CVID	1089	827	250	300	80	1b	MB1	-	Hi	Norm	T2	28.8	239.0	34.9	87.3
49	CVID	492	441	38	79	49	1b	MB1	+	Norm	Norm					
50	CVID	991	672	310	66	76	1b	MB1	-	Norm	Low					
51	CVID	1317	648	471	308	280	II	MB1	+	Norm	Norm	T1	11.3	61.0	19.1	82.1
52	CVID	714	194	194	1	111						T1	10.9	52.3	38	64.6
53	CVID	5640	2240	3110	1280	150	1b	MB0	-	Norm	Norm					
56	CVID	1127	822	308	54	172	II	MB2	+	Norm	Norm					
58	CVID	297	172	119	18	55						T2	25.5	43.4	43.0	51.6

Patient ID	Dx	CD3 cells/µl	CD4 cells/µl	CD8 cells/µl	CD19 cells/µl	CD56 cells/µl	Freiburg	Paris	E UROclass SmB	Tr	CD21	Rome	Naïve CD4 %	cells/µl	Naïve CD8 %	cells/µl
59	CVID	986	609	300	139	156	1b	MB0	-	Norm	Norm					
61	CVID	1250	501	708	412	151	II	MB1	+	Norm	Norm					
63	CVID	1650	660	880	330	50										
66	CVID	2870	2087	797	670	370	II	MB2	+	Norm	Norm					
67	CVID	3062	1157	1893	381	106	1b	MB0	+	Hi	Norm					
68	CVID	1179	826	325	410	134	1b	MB1	-	Norm	Norm					
69	CVID	1376	905	462	117	31	1b	MB1	+	Norm	Norm					
74	CVID	938	734	203	175	211	1b	MB0	-	Norm	Norm	T3	42.3	308.8	15.1	30.2
77	CVID	1370	1131	236	47	138	1b	MB1	-	Norm	Low	T2	27.4	287.7	26.7	51.6
78	CVID	2876	1254	1617	273	127	II	MB2	+	Norm	Norm					
80	CVID	1402	640	757	124	65	1b	MB1	+	Norm	Norm	T2	23.1	168.6	11.4	86.6
82	CVID	1197	615	423	428	456	II	MB1	+	Hi	Norm					
84	CVID	1542	764	642	476	138	1b	MB2	+	Norm	Norm	T3	44.1	299.9	38.0	186.2
85	CVID	1366	809	519	400	100	1b	MB1	-	Norm	Norm					
87	CVID	1970	959	870	182	228	II	MB1	+	Norm	Low					
88	CVID	2519	1605	895	371	141	II	MB2	+	Norm	Norm					
89	CVID	1684	1158	380	243	172	1b	MB1	-	Norm	Norm					
90	CVID	1472	634	706	374	35	II	MB1	+	Norm	Norm					
91	CVID	822	549	224	45	246										
94	CVID	362	290	100	40	70	1a	MB1	-	Norm	Low					
95	CVID	1170	421	739	65	47	1b	MB1	-	Norm	Norm	T1	10.1	42.4	3.8	28.1
96	CVID	864	640	214	194	139	1b	MB1	-	Norm	Norm	T2	21.1	135.0	12.3	25.8
36	SPAD	1119	521	591	144	100	1b	MB1	-	Norm	Norm					
40	SPAD	1804	1154	644	311	163	1b	MB1	+	Norm	Norm					
70	hypo-gamma	2990	1600	1308	0	27						T1	6.5	61.1	11.6	116.0

Table 11: A summary of B and T cell phenotyping results. Dx: diagnosis, CD3: $CD3^+$ T cells, CD4: $CD3^+CD4^+$ T cells, CD8: $CD3^+CD8^+$ T cells, CD19:CD19 B cells, $CD56^+$: $CD3^-CD56^+$ NK cells, Naïve CD4: $CD3^+CD4^+CD27^+$ cells, as a percentage of CD4 cells and absolute numbers, Naïve CD8: $CD3^+CD4^-CD27^+$ cells, as a percentage of CD8 cells and absolute numbers. *Reference ranges for adult naïve T cell ranges are not available. (Reference ranges for 10-16 year olds are 151-507 cells/µl naïve CD4 T cells and 63-313 cells/µl naïve CD8 T cells.)

Correlation of immunological with clinical features

Analysis of naïve $CD8^+$ T cell parameters with the clinical features shown in Table 11 suggests that percentage of naïve $CD8^+$ T cells of total $CD8^+$ T cells is the most relevant parameter for classification. Relevant cutoffs were optimised to distinguish patients with and without granuloma, autoimmunity, splenomegaly and lymphadenopathy. Patients with <7%, 7-35% and >35% naïve $CD8^+$ T cells were classified into groups N8lo, N8int and N8hi respectively. Frequencies of clinical complications for each group in the Freiburg, Paris, EUROclass, Rome and Naïve CD8 classification schemes for this cohort were calculated and are shown in Table 12. (For reference, a summary of the reported clinical associations with the different classification schemes can be found in Table 6.)

Small numbers of patients fell into the most severe classifications of Freiburg and Paris which makes analysis of this data difficult. The application of EUROclass to this cohort shows increased rates of splenomegaly and granuloma in SmB- and CD21 (lo), which fits with the previously reported findings (Wehr *et al*. 2008). However, the difference in rates of splenomegaly and granuloma did not reach statistical significance, and with the Tr (hi) and Tr (norm) groups the reported association with lymphadenopathy was reversed. These findings emphasise that while between large groups the classification schemes may show differences, they are of less utility when applied to individual patients.

A similar picture emerges with the Rome naïve $CD4^+$ T cell classification, where splenomegaly rates are highest in T1 and lowest in T3 (the reported association), but again this did not reach statistical significance. The utility of naïve $CD8^+$ T cell to

classify CVID patients shows potential with rates of granulomatous disease, autoimmunity, splenomegaly and lymphadenopathy being highest in N8lo, intermediate in N8int and lowest in N8hi. However, only the difference in granuloma rates showed statistical significance (Fisher's exact test, p=0.012), and further evaluation of naïve CD8 T cells is required to better determine the presence of clinical associations and to determine the optimal cut off values for each subgroup.

Eighteen CVID patients (34.0%) were noted to have NK cell numbers below the reference range (80-650 cells/µl). By chance, you would expect 2.5 % of a randomly selected sample to be below the reference range. The significance of low NK cells is unclear, and there did not appear to be any particular clinical phenotype associated with this. However, the finding of possible immunological changes not related to B or T cells is of potential interest as such changes may be detected by an approach utilising microarray analysis of peripheral blood RNA expression.

Group	Cohort	Freiburg			Paris			EUROclass						Rome			Naïve CD8		
		1a	1b	2	MB0	MB1	MB2	SmB-	SmB+	Tr (hi)	Tr (norm)	CD21 (lo)	CD21 (norm)	T1	T2	T3	N8lo	N8int	N8hi
Total number	**53**	1	27	14	5	30	7	20	22	6	36	7	35	8	9	4	3	15	3
Granuloma % (n)	**11.3 (6)**	100 (1)	7.4 (2)	7.1 (1)	20 (1)	10 (3)	0 (0)	10 (2)	9.1 (2)	33.3 (2)	5.6 (2)	28.6 (2)	5.7 (2)	25 (2)	0 (0)	0 (0)	66.7 (2)	0 (0)	0 (0)
Autoimmunity % (n)	**26.4 (14)**	0 (0)	22.2 (6)	21.4 (3)	40 (2)	16.7 (5)	28.6 (2)	15 (3)	27.2 (6)	50 (3)	16.7 (6)	0 (0)	25.7 (9)	25 (2)	11.1 (1)	5 (2)	33.3 (1)	26.7 (4)	0 (0)
Splenomegaly % (n)	**26.4 (14)**	100 (1)	29.6 (8)	7.1 (1)	40 (2)	26.6 (8)	0 (0)	30 (6)	18.1 (4)	50 (3)	19.4 (7)	28.6 (2)	22.8 (8)	37.5 (3)	22.2 (2)	0 (0)	66.7 (2)	13.3 (2)	33.3 (1)
Lymphadenopathy % (n)	**13.2 (7)**	100 (1)	7.4 (2)	14.2 (2)	0 (0)	13.3 (4)	14.3 (1)	10 (2)	13.6 (3)	0 (0)	13.9 (5)	28.6 (2)	8.5 (3)	2 (2)	11.1 (1)	0 (0)	66.7 (2)	6.7 (1)	0 (0)
Bronchiectasis % (n)	**56.6 (30)**	100 (1)	59.3 (16)	42.9 (6)	60 (3)	56.7 (17)	42.9 (3)	70 (14)	40.9 (9)	50 (3)	55.6 (20)	42.9 (3)	57.1 (20)	75 (6)	77.8 (7)	50 (2)	100 (3)	66.7 (10)	66.7 (2)

Table 12: Frequency of clinical complications in different CVID subclasses defined by a variety of classification methods. Total number indicates the number of patients within each classification. % indicates the percentage of patients within the particular classification group who have the particular complication.

Eight patients (patient ID 4, 10, 14, 28, 52, 56, 80, 91), excluding the patients who developed CLL, have been observed to have falling numbers of B cells over time as shown in Figure 10. A further 6 patients had low B cell numbers below the reference range of 50-500 cells/µl (27, 30, 44, 58, 77, 94), but without sufficient historical data to determine whether their B cell numbers were previously normal in the past. All have been confirmed to express normal levels of btk protein, thereby excluding XLA. There was no obvious clinical difference with these patients, though the numbers may have been too small. Patients with low B cell frequency are excluded from the existing B cell phenotype classification schemes. Patients with low B cell numbers may represent another source of variability of peripheral blood RNA expression which may provide useful information in a microarray analysis of the CVID cohort.

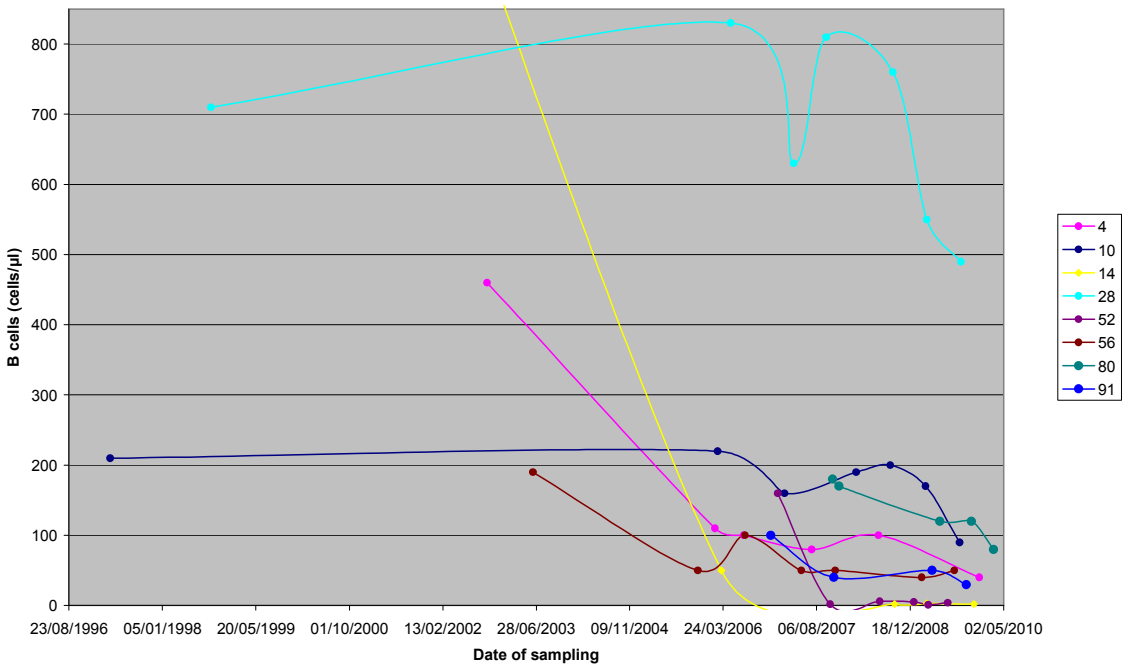


Figure 10: Graph showing serial B cell enumeration of patients with falling numbers of B cells (normal range 90-660 cellsµ/l).

Results Chapter 3

RNA extraction, gene expression profiling and quality control

RNA extraction and quality

RNA was extracted from the patients listed in Table 7, and the resultant biobank stored at -70ºC with the temperature logged and monitored. The concentration and integrity of the RNA for the 64 samples selected for microarray analysis was confirmed by electropherogram (performed by Central Biotechnology Services (CBS)), as shown in Table 13 and Figure 11. The RNA integrity number (RIN) gives a measurement of RNA degradation. The score ranges from 1 to 10, with 10 representing a perfect sample with no degradation. All the samples had a RIN of greater than 7, and ranged from 7.3 to 9.6.

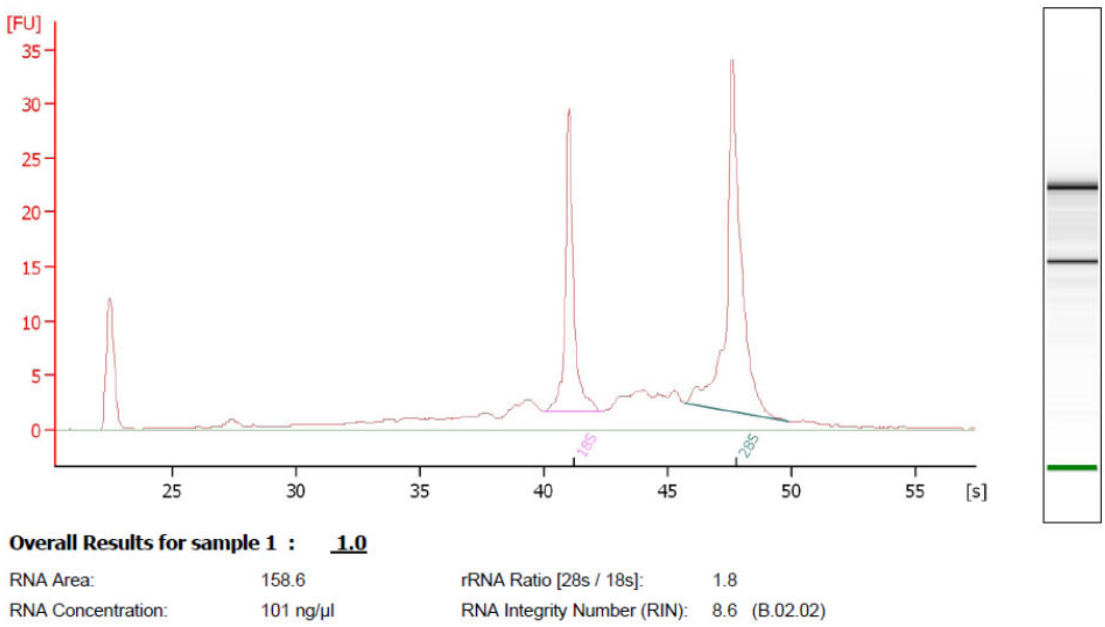


Figure 11: Electropherogram of sample 1 demonstrating good quality RNA with a 28S:18S ratio of 1.8, and a relatively flat baseline between the 5S and 18S peaks.

Sample ID	[RNA] (ng/µl)	RIN
1	101	8.6
3	62	7.4
4	147	8.7
7	27	8.7
8	25	8.6
9	57	9.0
10	68	9.2
11	126	9.4
13	67	9.1
14	52	9.2
23	25	8.6
26	39	8.3
27	132	9.1
28	56	9.3
29	98	7.5
30	22	9.1
31	109	9.5
32	37	7.8
34	55	9.5
35	95	8.8
36	185	8.9
39	150	9.1
40	80	9.5
42	168	9.2
43	70	7.8
44	45	8.9
45	233	8.6
46	215	9.0
49	54	8.5
50	47	9.0
51	112	8.2
52	82	8.2
53	84	8.4
56	24	8.3
58	49	8.5
59	35	9.0
61	87	7.3
63	50	9.4
64	217	8.0
66	90	8.3
67	28	9.6
68	63	9.3
69	148	7.8
70	22	9.1
72	55	9.2
73	37	8.7
74	97	8.4
77	67	7.4
78	91	8.6
80	62	9.1
80T	16	9.0
82	33	8.9
84	86	9.3
85	131	9.2
87	19	8.2
88	84	9.1
89	58	7.8
90	157	8.6
91	38	8.4
94	87	7.7
95	69	8.5
96	52	8.7

Table 13: Electropherogram results of samples selected for microarray analysis.

Printing of targeted microarray

The first steps in developing the targeted array were to select genes of interest and design thermoequivalent oligonucleotide probes specific for these genes. This was performed by Chris Mercer, and 1593 probes were manufactured (Operon Biotechnologies) and are listed in Appendix 4. The aim was to spot these probes to produce custom microarray slides.

A wash protocol to effectively clean the pins between printing different oligonucleotides was determined and tested. The wash protocol shown in Table 14 was used for subsequent print runs.

Step	Wash time (ms)	Dry time (ms)	Wait after (ms)
1	2000	0	500
2	2000	0	500
3	3000	0	500
4	3000	1500	500
5	0	1500	500
6	0	1500	1000
7	0	3000	0

Table 14: Pin wash protocol, using de-ionised water.

Optimisation of printing commenced using split pins as these are able to spot multiple times after each 'inking' and are therefore able to print microarrays significantly faster than solid pins. The split pins were able to consistently spot >45 times before needing to return to pick up further oligonucleotide solution. An ink time of 3000ms was sufficient to allow uptake of the solution.

Initial print runs produced arrays with variation in spot diameter, and with poor morphology of the larger diameter spots (Figure 12). A number of different

approaches to improve the arrays were tried. Blotting (printing several spots on a spare slide immediately after a re-ink, prior to spotting on the arrays) did not improve spot morphology or consistency. Double stamping, or immediately printing a second spot in the same position on the array, was also of no benefit. A stamp time of 150ms was used, and changes to stamp time made little difference to the quality of the final array. The humidity needed to be kept within 50 to 80%, and outside of this range the consistency of spotting was impaired. The depth adjustment determines how far beyond the initial contact between the pin and array slide the head holding the pins continues to travel. This is important when using split pins as changes in the amount of travel affect spot morphology; with too little travel a small or no spot is produced, but with too much travel a poor morphology spot often with a donut appearance is produced. Despite many attempts at adjusting the above parameters, it was not possible to achieve consistent results with the split pins. Even within the time frame of less than one day, the parameters needed for good quality spot morphology would vary. The position of each pin in the print head was recorded, and which position each pin would perform best in would vary. The optimal print depth adjustment was also particularly prone to variation. For this reason the decision was made to change to solid pins, which although slower to print with, are more consistent, and have less variables affecting their performance.

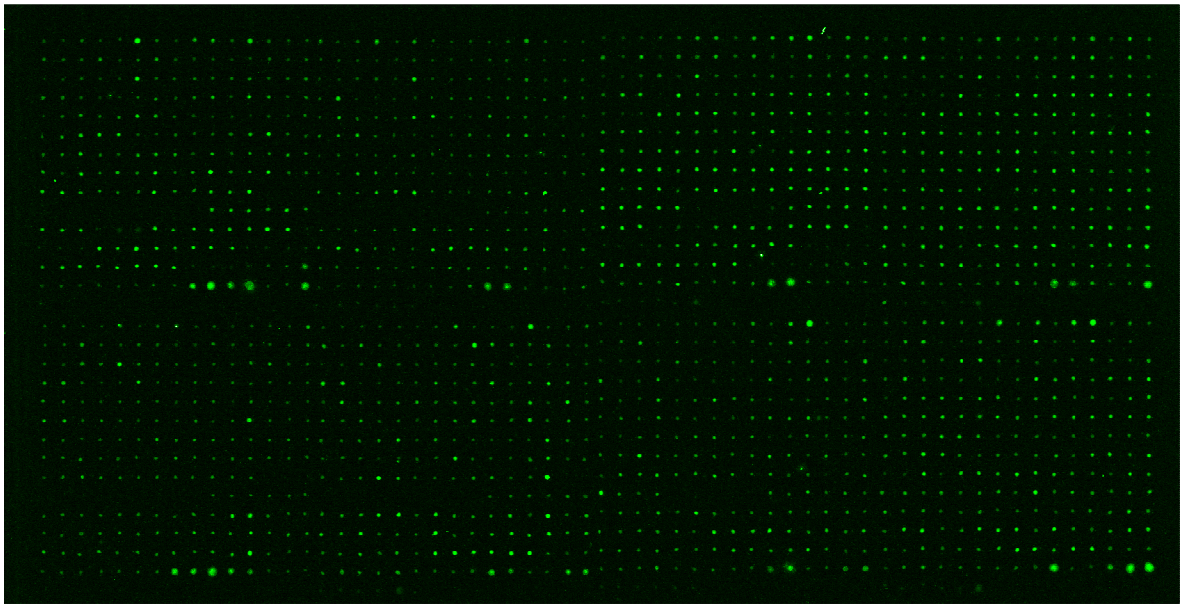


Figure 12: Example of early print run using split pins showing poor consistency between spots, and small spot diameter.

For solid pins the ink and stamp time are both zero, and there is 1 spot per ink. Consistent printing of food colouring (5µl in 200µl spot buffer) was easily achieved with a depth adjustment of 0 microns. These results were reproducible over time using the same settings (other than a change in the depth adjustment from -70 to 0 microns which was required following a failure of the printer necessitating repairs), thereby confirming the consistency of performance of the robotic printer. The depth adjustment was tested and confirmed across the printing area (64 slides). However, the printing of oligonucleotides remained poor, with variable size and morphology between oligonucleotide spots. It was observed that the same oligonucleotides printed consistently well or poorly between the print runs. There was the suspicion that the oligonucleotides used for printing may have degraded as these had been repeatedly freeze-thawed, but plating out fresh 384 well plates from stock oligonucleotides which had not been freeze-thawed did not improve the quality of printing.

The performance of different buffers with and without detergent, and with different concentrations of oligonucleotide was assessed. A comparison of saline sodium citrate (SSC) with and without detergent and Nexterion spot buffer with and without detergent was made. Oligonucleotide to detect IRAK1 mRNA was used for this test with concentrations of 10, 20, 40, 60 and 80µM for each buffer. The recommended oligonucleotide concentration of 40µM resulted in the highest quality spots regardless of the buffer used. Buffers with detergent (Tween 20) added outperformed those without detergent, and Nexterion spot buffer was superior to SSC. Further testing was performed using a range of detergents at a range of concentrations (Tween 20 at 0.0005%, 0.0025%, 0.005%, 0.025% and 0.05%, and Triton X-100 and sodium dodecyl sulfate (SDS) at 0.001%, 0.005%, 0.01%, 0.05%). For IRAK1, Tween 20 at 0.005% provided the largest spot (110µm diameter) with good morphology. While informative, this result has only optimised the printing of a single oligonucleotide probe. The buffer assessment was repeated with oligonucleotides chosen as they resulted in poor, medium and good morphology spots (CD84, ADAM5 and ICOS respectively). For ICOS, higher concentrations of detergent resulted in larger diameter spots which remained of good morphology. However, higher concentrations of detergent resulted in poor spot morphology for CD84 and ADAM5. Nexterion spot buffer with 0.01% Tween 20 appeared to be the best compromise, achieving spots with the greatest diameter while retaining acceptable morphology across the three oligonucleotides.

However, despite achieving good spot morphology for a limited number of oligonucleotides (Figure 13), when the above conditions were applied to all 1593 genes for a print run, it was not possible to print large numbers of oligonucleotides at

high quality (Figure 14). This suggests that the optimal print conditions vary between the oligonucleotide probes.

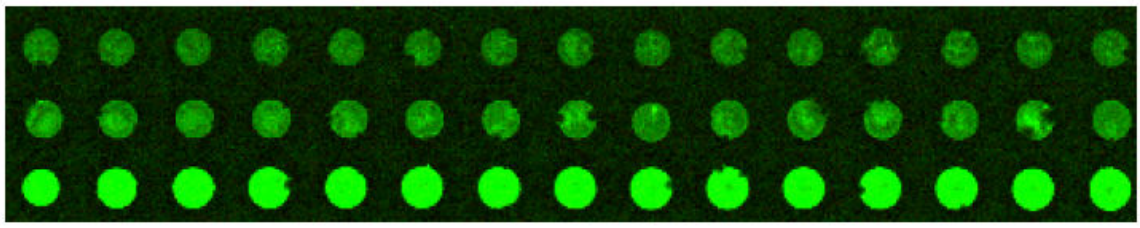


Figure 13: Printing of 3 oligonucleotids in same run (solid pins, depth adjustment 0 microns, Nexterion spot buffer with 0.01% Tween 20, [oligonucleotide]=40µM). Row 1: 15 repeats of CD84, row 2: ADAM5 and row 3: ICOS.

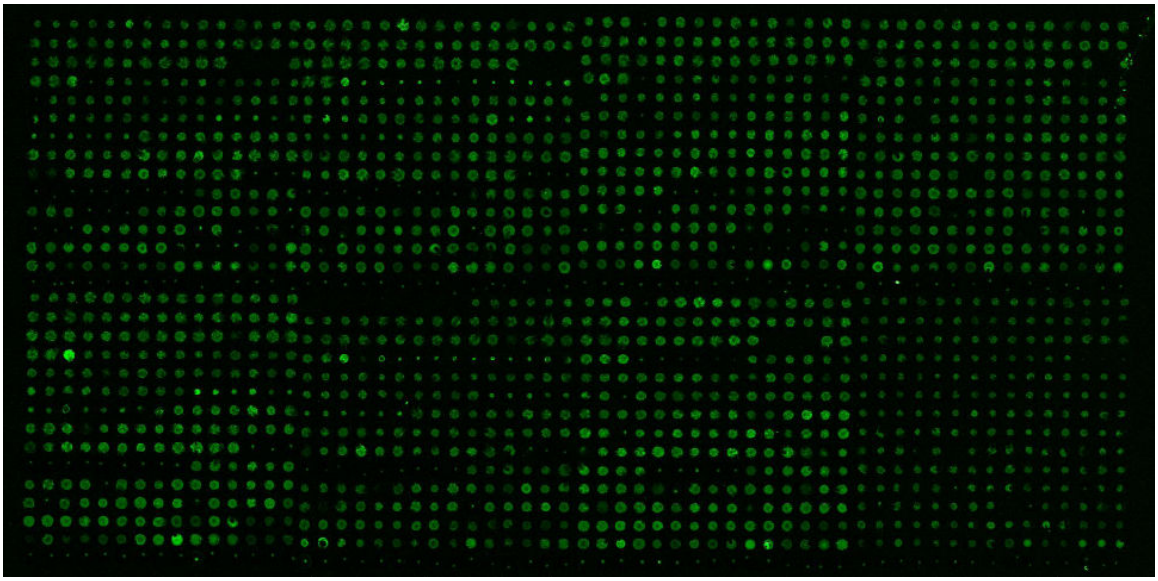


Figure 14: Example of spotted microarray (solid pins, depth adjustment 0 microns, Nexterion spot buffer with 0.01% Tween 20, [oligonucleotide]=40µM)

Given the technical issues experienced in producing in-house oligonucleotide manufactured arrays, the decision was made to use Illumina BeadChips to test the project hypotheses. The significantly higher number of genes represented on these chips (>24,000) has the additional benefit of allowing more robust normalisation of the data, thereby removing artefact due to experimental variation and increasing the likelihood of detecting differentially expressed genes between CVID phenotypes.

Although it proved difficult to print large oligonucleotide arrays using the QArray spotter, good quality printing was achieved for small numbers of oligonucleotides. This technology could potentially produce mini/midi-plex arrays with a limited number of oligonucleotide probes. If a limited number of genes that could differentiate CVID subgroups were identified, then such an array may prove a useful diagnostic tool.

Quality control of gene expression data generated by Illumina BeadChip

Gene expression of RNA extracted from the peripheral blood of the sixty-four patients detailed in Results Chapters 1 and 2 was analysed using the Illumina HumanRef-8 v3.0 Expression BeadChip. Each slide consisted of 8 arrays, A-H, and the samples were distributed across the slides as detailed in Appendix 5.

Prior to further analysis, a number of quality control (QC) measures were examined. Both sample dependent and sample independent control features were analysed. The sample independent controls utilise oligonucleotides spiked into the hybridisation solution and provide an independent check of the hybridisation, washing, staining and scanning steps. Included within the spike are different hybridisation controls which should return high, medium and low signals (Figure 15a). The stringency of hybridisation is examined by comparing perfect match (PM) probes with probes containing 2 mismatches (MM2). The PM signal should be higher than the MM2 signal, and poor returns from these probes would indicate a problem with specificity (Figure 15b). There are also negative control beads, which consist of probes with sequences which have no corresponding target in the genome, which provide a measure of the background level (Figure 15c). Sample dependent quality control is

confirmed by examination of housekeeping genes which are expected to be expressed in any cellular sample. The housekeeping genes are expected to produce a higher signal compared to the average for all genes, and provide a check for the quality of sample and labelling in addition to the steps from hybridisation onwards (Figure 15d). Both sample dependent and independent controls returned expected results across all the arrays.

Normalisation techniques are used in processing microarray data to resolve systematic errors and bias introduced by random experimental variation. The multi-step nature of analysis by microarrays leads to several potential sources of variation, which include sample preparation, RNA extraction, cDNA and cRNA generation, labelling of cRNA, array manufacture and hybridisation/washing efficiency. In the setting of whole genome expression analysis, most of the genes would be expected to be unaltered across the different samples. As a result the data is appropriate for quantile normalisation, which in addition to standardising the overall intensity of each array, also standardises the distribution of probe intensities. This is achieved by normalising the data so the highest intensity probes in each array have the same intensity, which is equal to the average intensity of the highest intensity probes across all the arrays in the set. This process is repeated with the next highest intensity probes and so forth (Bolstad *et al.* 2003). Quantile normalisation of the data was performed using GenomeStudio, and all subsequent analysis in Results Chapter 4 refers to the normalised data.

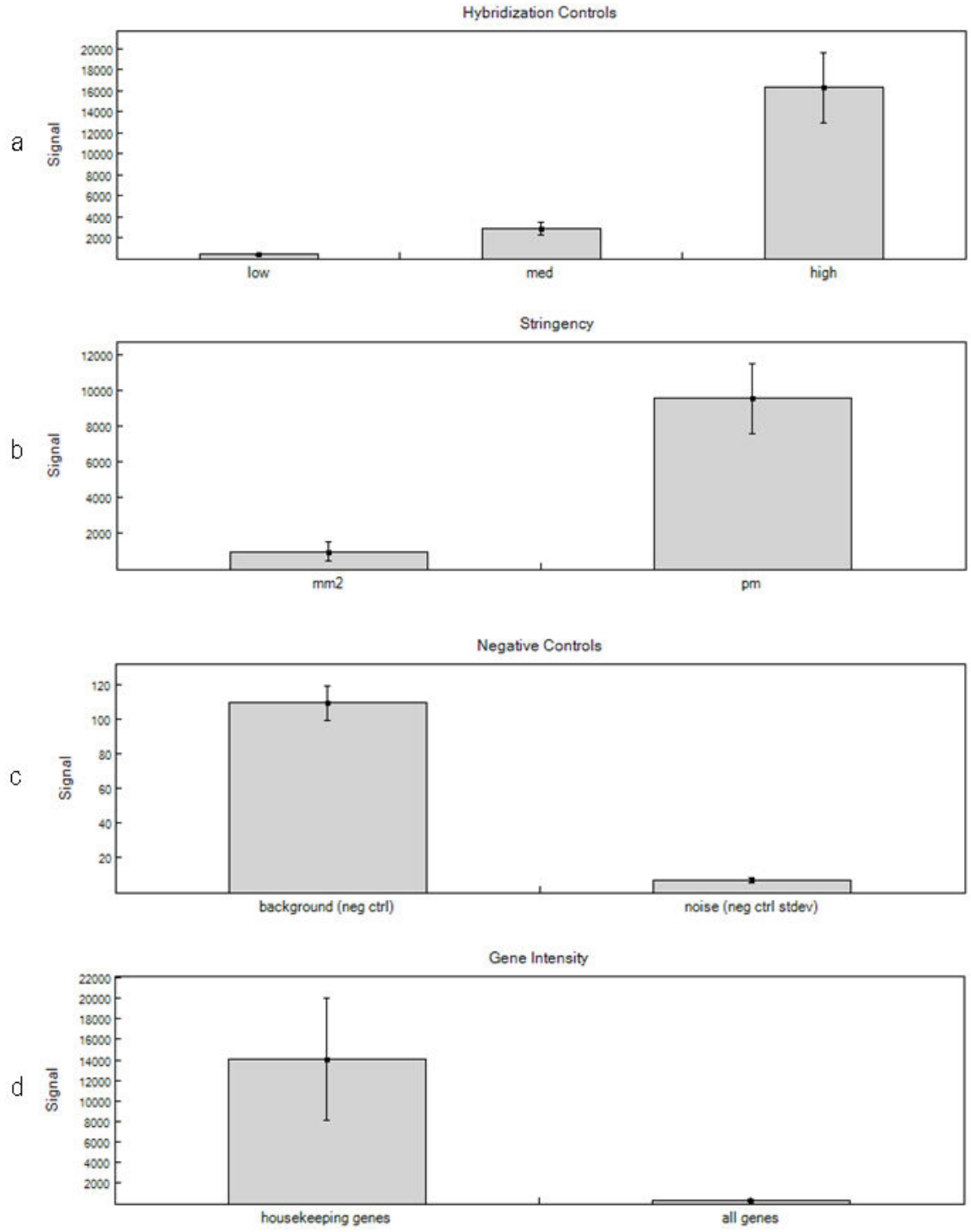


Figure 15: Summary of QC data from all microarrays. a) shows the signal from spiked hybridisation controls which should return low, medium and high signals. b) shows good stringency with low signal from probes containing 2 mismatches (MM2) compared to perfect match probes (PM). c) demonstrates low background levels. d) shows high level expression of housekeeping genes compared to the average for all genes.

Results Chapter 4

Analysis of peripheral blood gene expression

Principal component analysis

Microarray data is highly dimensional, which presents difficulties with visual representation of the data. Principal component analysis (PCA) is a linear transformation that converts x variables into y new variables, or principal components. The first principal component has as high a variance as possible thereby accounting for as much of the variability in the data as possible. Subsequent principal components must be uncorrelated, and have the highest variance possible within this constraint. The use of PCA to transform microarray data allows the possibility of viewing the microarray data in 3 dimensions while retaining as much variance in the visualised data as possible.

PCA of all 64 samples are shown in Figure 16. In this figure, as with all subsequent PCA scatter plots, the first principle component is on the x axis, second principle component on the y axis and the third principal component on the z axis. Fogging is used to indicate position on the z axis. Samples 80 and technical replicates (80X, an aliquot from an identical sample and 80T, a sample from patient 80 taken at a different time point) lie close together on the scatter plot, as do samples 96 and 96X. It was not possible to resolve the different clinical diagnostic categories using this approach. CVID, hypogammaglobulinaemia and SPAD can be considered as a spectrum of severity of idiopathic antibody deficiency. The XLA patients tend to lie

on the left of the scatterplot (i.e. are separated out by the first principle component) but still overlap with the other samples.

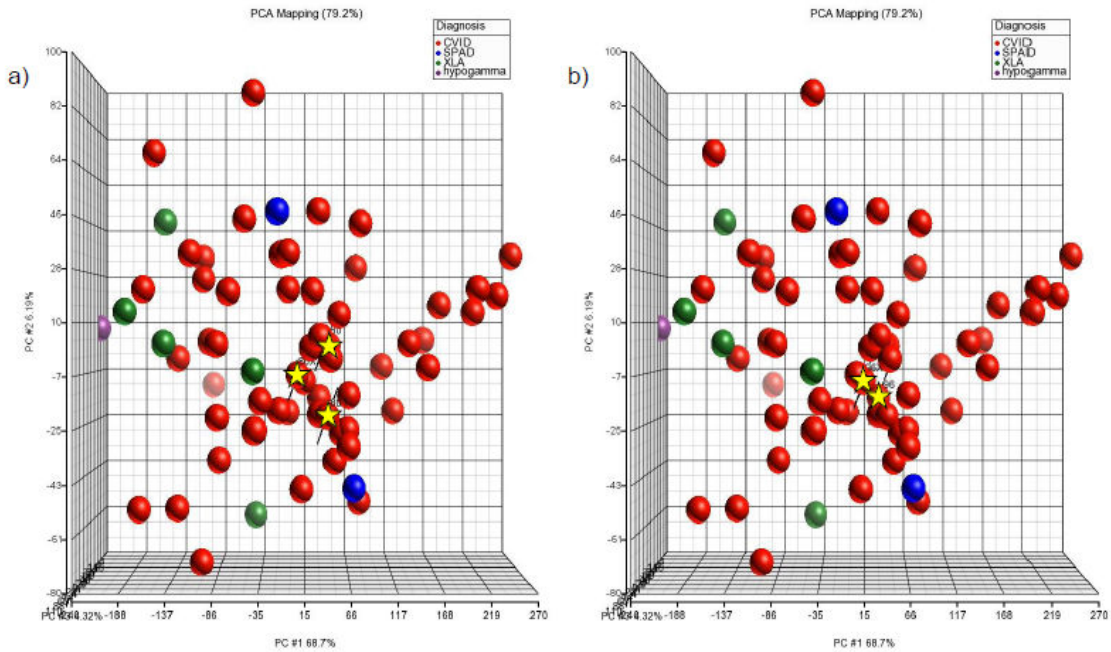


Figure 16: PCA of all 64 samples. Samples from patients with CVID are indicated in red, XLA in green, SPAD in blue and hypogammaglobulinaemia in purple. In scatter plot (a) samples 80, 80X and 80T are labelled with yellow stars, and in scatter plot (b) samples 96 and 96X are labelled.

Figure 17 shows the results of PCA of the 53 CVID patients as an initial look to determine whether those with different complications (granulomatous disease, autoimmunity, splenomegaly, lymphadenopathy and bronchiectasis) are easily distinguishable. PCA was not able to separate different CVID complications out into different groups. While PCA is a useful visualisation tool, it lacks the capability to detect subtle differences compared to techniques employed later in this chapter.

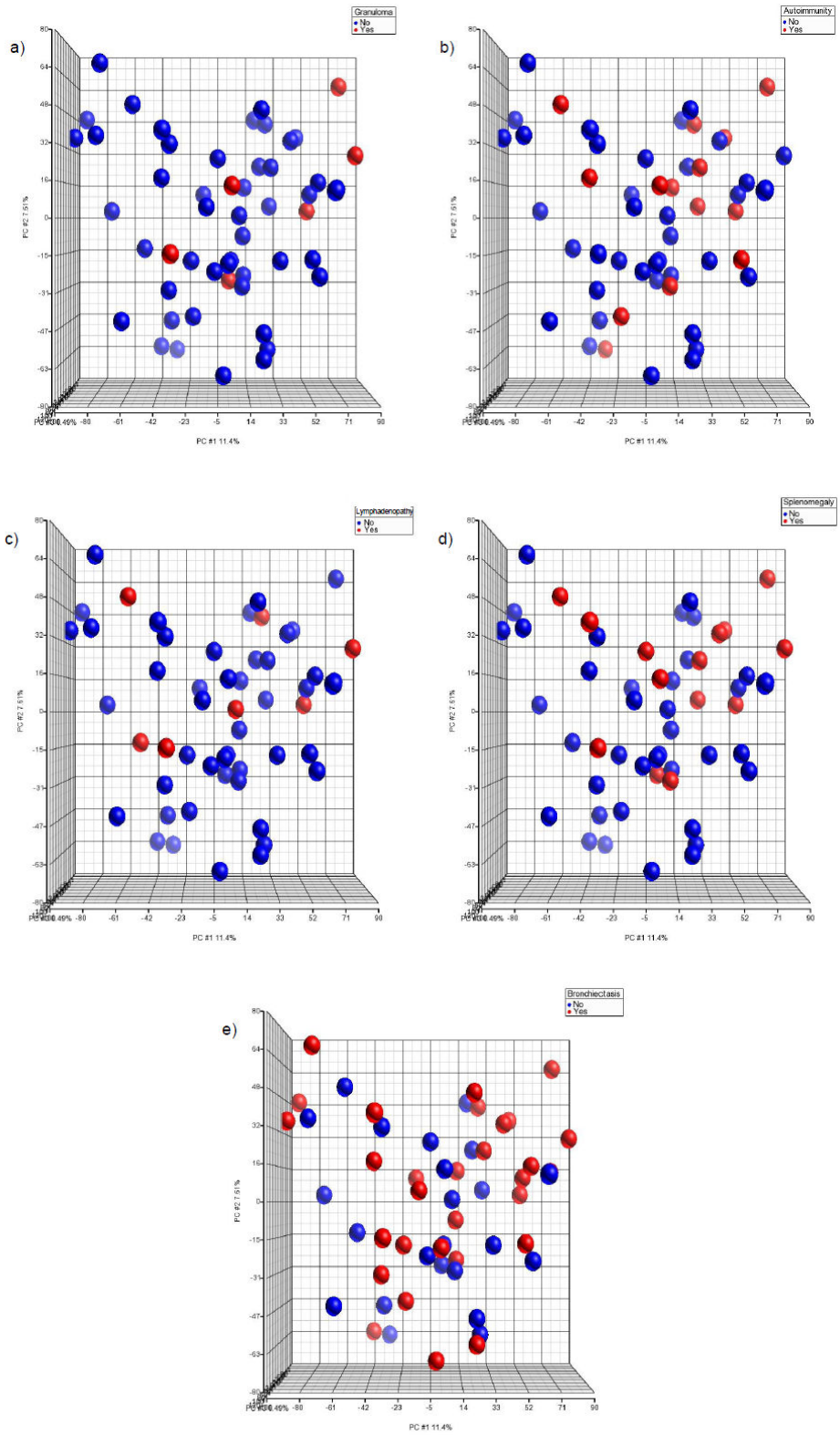


Figure 17: PCA of CVID patients with presence of (a) granuloma, (b) autoimmunity, (c) splenomegaly, (d) persistent lymphadenopathy and (e) bronchiectasis indicated in red.

Class discovery

Unsupervised class discovery potentially allows the discovery of unknown groups and is based on the microarray data alone without using any external information. Consensus clustering provides a method whereby the number and membership of possible clusters can be determined, and uses re-sampling to assess the stability or robustness of these clusters (Monti *et al*. 2003; Wilkerson and Hayes 2010). The method also provides an indication as to when the maximum number of stable groups has been determined, after which the addition of further groups is likely to be due to random picks rather than true structure within the data. The results of ConsensusClusterPlus analysis of the data (performed by Peter Giles, CBS) are shown in Figure 18. The consensus matrices in Figure 18 (a-d) show how often the samples cluster together. The consensus values range from 0 (white) where the samples never cluster together, to 1 (dark blue) where the samples always cluster together. The order of samples is the same along the top and side of the matrix, and determined by the results of the consensus clustering. Perfect clustering would result in distinct blue blocks along the diagonal. The consensus matrices are a visualisation tool to help assess the clusters' robustness, composition and number. The delta area plot shows relative increase in consensus for different numbers of clusters (k), and shows the change between k and k-1, i.e. shows the additional benefit (if any) of defining an additional cluster. The delta area plot is shown in Figure 18 (e), and each increase in the numbers of clusters over 4 leads to only a small incremental benefit. Furthermore, at 5 clusters and greater, patient samples which would be expected to cluster together were found in different groups. The technical replicates (80, 80X & 80T) remained in the same group when there were 2, 3 and 4 clusters, but at 5 clusters failed to group together.

Table 15 and Table 16 show the grouping of patients by consensus clustering alongside their clinical features and laboratory data. There were no particular clinical or laboratory features which were characteristic of any of the groups. Four of the XLA patients remained in the same group up to and including five clusters. The fifth XLA patient, who had metastatic carcinoma of the colon and had received chemotherapy, did not cluster with the other XLA patients. These results are similar to those from PCA, where the technical replicates are very close to each other, and four of the XLA patients were relatively close to each other, but not discreet from other patients. The use of another method of unsupervised learning, hierarchical Euclidean-distance-based clustering (Wiltgen and Tilz 2007), gave similar results (results not shown). Applying hierarchical clustering to only CVID patients did not allow the separation of different complications (granulomatous disease, autoimmunity, splenomegaly, persistent lymphadenopathy or bronchiectasis), which is also similar to the results from PCA. This is perhaps unsurprising as the differences between CVID patients may be more subtle than some of the other differences such as gender, age and other co-morbidities, and therefore difficult to detect by clustering based on the whole genome which would be expected to contain many differentially expressed genes only some of which would be related to the clinical features of interest. The next stage of analysis looked at clinically defined groups of interest to determine whether there were genes which were differentially expressed to a statistically significant level between these groups.

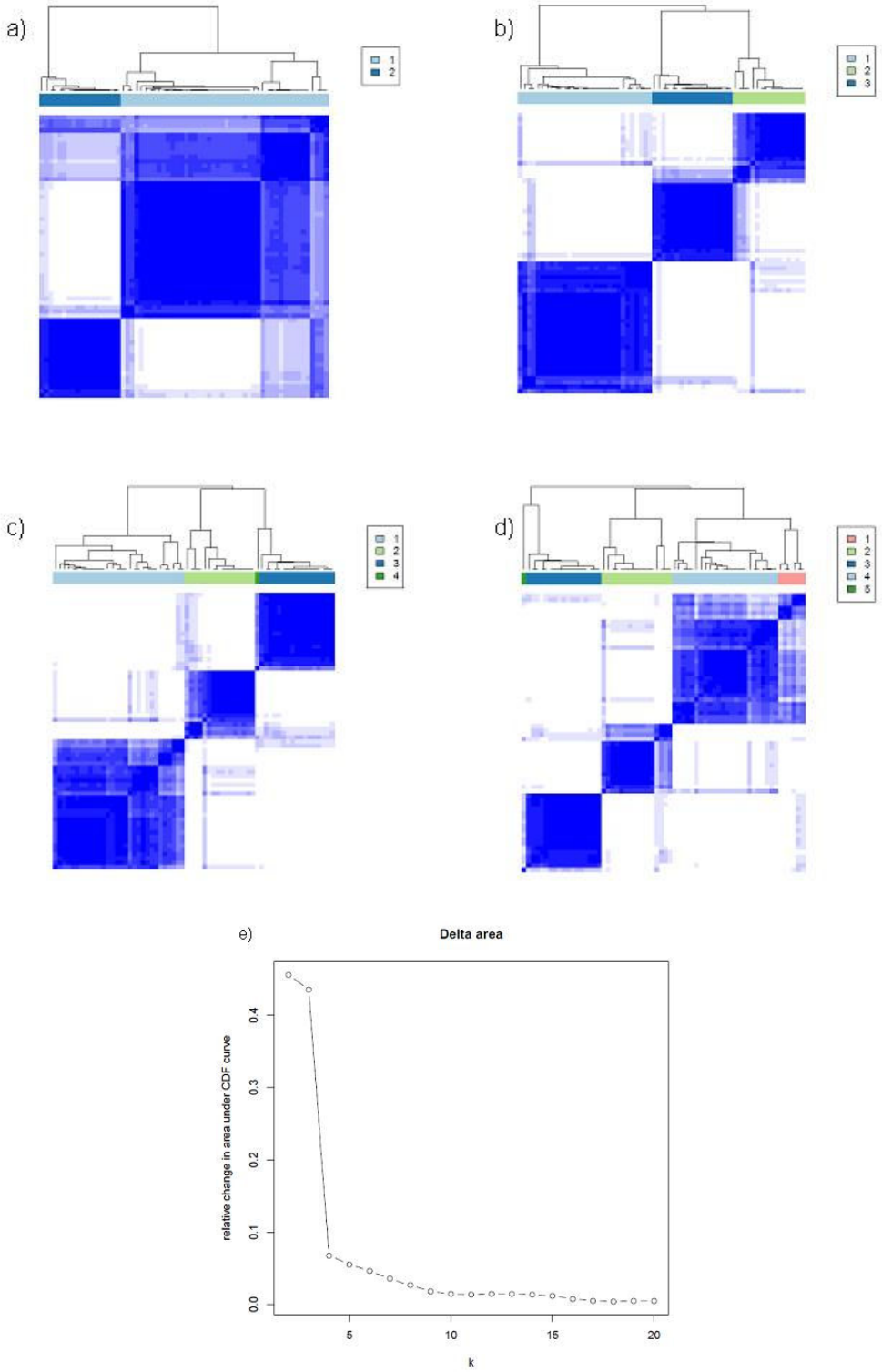


Figure 18: Results from consensus clustering analysis, with clustering into (a) 2, (b) 3, (c) 4 and (d) 5 groups. The consensus values range from 0 (white) where on resampling the samples never cluster together, to 1 (dark blue) where the samples always cluster together. The order of samples is the same along the top and side of the matrix, and determined by the results of the consensus clustering. Perfect clustering would result in distinct blue blocks along the diagonal. The delta area plot is also shown (e), and demonstrates the change in area under the cumulative distribution function (CDF) curve which allows determination of where consensus and cluster confidence is at a maximum. Increasing the number of clusters (k) beyond 4 results in only minimal change in the area under CDF curve.

ID	2 Clusters	3 Clusters	4 Clusters	5 Clusters	Sex	Age	Diagnosis	Other info	Gran uloma	Auto immunity	Spleno megaly	Lymphad enopathy	Bronch iectasis
1	1	1	1	1	F	70	CVID	Marginal zone lymphoma (2006)	N	Y	N	Y	Y
74	1	1	1	1	F	69	CVID		N	N	N	N	N
85	1	1	1	1	M	18	CVID		N	N	N	N	Y
90	1	1	1	1	M	32	CVID		N	N	N	N	N
80	1	1	1	1	M	53	CVID		N	N	Y	N	N
80X	1	1	1	1	M	53	CVID		N	N	Y	N	N
50	1	1	1	5	F	30	CVID	breast ca (2008)	N	N	N	N	Y
4	1	1	1	5	F	69	CVID		N	Y	N	N	Y
51	1	1	1	5	F	57	CVID		N	Y	N	N	Y
88	1	1	1	5	F	18	CVID		N	N	N	Y	N
94	1	1	1	5	M	39	CVID	T cell lymphopenia	Y	N	Y	Y	Y
29	1	1	1	5	F	61	CVID		N	N	N	N	Y
32	1	1	1	5	M	47	CVID		N	N	N	N	Y
45	1	1	1	5	F	65	CVID	developed CLL (2006), haemochromatosis	N	N	N	N	N
69	1	1	1	5	F	49	CVID		N	Y	Y	Y	N
77	1	1	1	5	F	51	CVID		N	N	N	Y	N
95	1	1	1	5	M	75	CVID		N	N	N	N	Y
34	1	1	1	5	M	56	CVID		N	N	N	N	N
36	1	1	1	5	F	64	SPAD		N	N	N	N	N
87	1	1	1	5	M	46	CVID		N	N	N	N	N
80T	1	1	1	5	M	53	CVID		N	N	Y	N	N
96X	1	1	1	5	F	60	CVID		N	N	N	N	Y
3	1	1	1	5	F	51	CVID		N	N	N	N	Y
52	1	1	1	5	F	44	CVID		N	N	Y	N	Y
89	1	1	1	5	M	40	CVID	Gilbert's syndrome	N	N	N	N	N
30	1	1	1	5	F	32	CVID		Y	Y	Y	N	N
35	1	1	1	5	F	47	CVID		N	N	N	N	Y
27	1	1	1	5	M	50	CVID		N	N	N	N	Y
42	1	1	1	5	M	26	CVID		N	N	N	N	N
96	1	1	1	5	F	60	CVID		N	N	N	N	Y
9	1	3	3	3	M	71	CVID		N	N	N	N	Y
13	1	3	3	3	M	67	CVID		Y	N	Y	N	N

Table 15 (part 1): Clinical features of patients grouped by consensus clustering analysis.

ID	2 Clusters	3 Clusters	4 Clusters	5 Clusters	Sex	Age	Diagnosis	Other info	Gran uloma	Auto immunity	Spleno megaly	Lymphad enopathy	Bronch iectasis
84	1	3	3	3	F	26	CVID		N	N	N	N	N
49	1	3	3	3	F	75	CVID		N	N	N	N	N
82	1	3	3	3	M	17	CVID		N	Y	N	N	N
91	1	3	3	3	F	62	CVID		N	N	N	N	N
23	1	3	3	3	M	44	CVID		N	Y	Y	N	N
44	1	3	3	3	M	54	CVID		N	Y	Y	N	N
46	1	3	3	3	F	34	CVID		N	Y	Y	N	Y
59	1	3	3	3	M	27	CVID		N	N	N	N	Y
10	1	3	3	3	F	38	CVID		N	N	N	N	N
58	1	3	3	3	M	47	CVID		N	N	N	N	Y
72	1	3	3	3	M	28	XLA		N	N	N	N	N
73	1	3	3	3	M	24	XLA		N	N	N	N	N
39	1	3	3	3	M	16	XLA		N	N	N	N	Y
64	1	3	3	3	M	37	XLA		N	N	N	N	Y
61	2	2	4	4	M	45	CVID		N	N	N	N	Y
40	2	2	2	2	F	32	SPAD		N	N	N	N	N
68	2	2	2	2	M	52	CVID	x2 PE's	N	N	Y	N	Y
14	2	2	2	2	M	35	CVID		N	Y	N	N	Y
67	2	2	2	2	F	74	CVID		Y	Y	Y	N	Y
78	2	2	2	2	F	75	CVID		N	Y	N	N	Y
70	2	2	2	2	F	41	hypogamma	IgG kappa paraprotein (3g/l)	N	Y	Y	N	Y
7	2	2	2	2	F	46	CVID		N	N	N	N	Y
8	2	2	2	2	M	60	CVID		N	N	N	N	Y
26	2	2	2	2	M	42	XLA	recurrent ca colon	N	N	N	N	N
43	2	2	2	2	M	45	CVID		N	N	N	N	N
53	2	2	2	2	M	45	CVID		N	Y	Y	N	Y
28	2	2	2	2	F	55	CVID		Y	N	Y	Y	Y
31	2	2	2	2	M	49	CVID		N	N	N	N	N
63	2	2	2	2	M	39	CVID		N	N	N	N	Y
56	2	2	2	2	F	32	CVID	mitochondrial myopathy	N	N	N	N	N
11	2	2	2	2	M	45	CVID		Y	Y	Y	Y	Y
66	2	2	2	2	M	74	CVID	developed CLL 2 years after RNA extraction	N	N	N	N	Y

Table 15 (part 2): Clinical features of patients grouped by consensus clustering analysis.

ID	2 Clusters	3 Clusters	4 Clusters	5 Clusters	Diagnosis	CD3+	CD4+	CD8+	CD19+	CD3-CD56+	Falling B cells?	Frei burg	Paris	EURO Class			Giovan netti
														SmB	Tr	CD21	
1	1	1	1	1	CVID	1910	690	1170	530	80	N						
74	1	1	1	1	CVID	938	734	203	175	211	N	1b	MB0	-	Norm	Norm	T3
85	1	1	1	1	CVID	1366	809	519	400	100	N	1b	MB1	-	Norm	Norm	
90	1	1	1	1	CVID	1472	634	706	374	35	N	II	MB1	Norm	Norm	Norm	
80	1	1	1	1	CVID	1402	640	757	124	65	Y	1b	MB1	Norm	Norm	Norm	T2
80X	1	1	1	1	CVID	1402	640	757	124	65	Y	1b	MB1	Norm	Norm	Norm	T2
50	1	1	1	5	CVID	991	672	310	66	76	N	1b	MB1	-	Norm	Low	
4	1	1	1	5	CVID	841	556	259	36	177	Y	1b	MB2	Norm	Norm	Norm	T3
51	1	1	1	5	CVID	1317	648	471	308	280	N	II	MB1	Norm	Norm	Norm	
88	1	1	1	5	CVID	2519	1605	895	371	141	N	II	MB2	Norm	Norm	Norm	
94	1	1	1	5	CVID	362	290	100	40	70	N	1a	MB1	-	Norm	Low	
29	1	1	1	5	CVID	1160	436	678	140	48	N	1b	MB1	-	Hi	Norm	
32	1	1	1	5	CVID	751	482	243	104	49	N	1b	MB1	-	Norm	Norm	
45	1	1	1	5	CVID	1124	594	465	266	112		II	MB0	Norm	Norm	Low	
69	1	1	1	5	CVID	1376	905	462	117	31	N	1b	MB1	Norm	Norm	Norm	
77	1	1	1	5	CVID	1370	1131	236	47	138	N	1b	MB1	-	Norm	Low	T2
95	1	1	1	5	CVID	1170	421	739	65	47	N	1b	MB1	-	Norm	Norm	T1
34	1	1	1	5	CVID	1132	501	575	326	74	N	2	MB1	Norm	Norm	Norm	T1
36	1	1	1	5	SPAD	1119	521	591	144	100	N	1b	MB1	-	Norm	Norm	
87	1	1	1	5	CVID	1970	959	870	182	228	N	II	MB1	Norm	Norm	Low	
80T	1	1	1	5	CVID	1402	640	757	124	65	Y	1b	MB1	Norm	Norm	Norm	T2
96X	1	1	1	5	CVID	864	640	214	194	139	N	1b	MB1	-	Norm	Norm	T2
3	1	1	1	5	CVID	976	707	237	75	238	N	1b	MB1	-	Norm	Norm	
52	1	1	1	5	CVID	714	194	194	1	111	Y						
89	1	1	1	5	CVID	1684	1158	380	243	172	N	1b	MB1	-	Norm	Norm	
30	1	1	1	5	CVID	327	284	38	21	30	N						
35	1	1	1	5	CVID	747	466	242	287	64	N	1b	MB1	-	Norm	Norm	T2
27	1	1	1	5	CVID	2080	1040	960	0	50	N						
42	1	1	1	5	CVID	980	330	580	230	50	N						
96	1	1	1	5	CVID	864	640	214	194	139	N	1b	MB1	-	Norm	Norm	T2
9	1	3	3	3	CVID	1127	571	499	234	129	N	1b	MB1	Norm	Norm	Norm	T2
13	1	3	3	3	CVID	1229	630	482	266	211	N	1b	MB1	-	Hi	Low	

Table 16 (part 1): Laboratory features of patients grouped by consensus clustering analysis

ID	2 Clusters	3 Clusters	4 Clusters	5 Clusters	Diagnosis	CD3+	CD4+	CD8+	CD19+	CD3-CD56+	Falling B cells?	Frei burg	Paris		EURO Class		Giovan netti
														SmB	Tr	CD21	
84	1	3	3	3	CVID	1542	764	642	476	138	N	1b	MB2	Norm	Norm	Norm	T3
49	1	3	3	3	CVID	492	441	38	79	49	N	1b	MB1	Norm	Norm	Norm	
82	1	3	3	3	CVID	1197	615	423	428	456	N	II	MB1	Norm	Hi	Norm	
91	1	3	3	3	CVID	822	549	224	45	246	Y						
23	1	3	3	3	CVID	1101	613	426	286	121	N	1b	MB1	-	Norm	Norm	
44	1	3	3	3	CVID	1384	228	1169	6	99	N						
46	1	3	3	3	CVID	1089	827	250	300	80	N	1b	MB1	-	Hi	Norm	T2
59	1	3	3	3	CVID	986	609	300	139	156	N	1b	MB0	-	Norm	Norm	
10	1	3	3	3	CVID	931	692	221	91	44	Y	1b	MB1	Norm	Norm	Norm	
58	1	3	3	3	CVID	297	172	119	18	55	N						T2
72	1	3	3	3	XLA	1040	230	800	1	50	N						
73	1	3	3	3	XLA	960	620	270	1	70	N						
39	1	3	3	3	XLA	2610	1200	1180	5	260							
64	1	3	3	3	XLA	1200	570	490	0	90							
61	2	2	4	4	CVID	1250	501	708	412	151	N	II	MB1	Norm	Norm	Norm	
40	2	2	2	2	SPAD	1804	1154	644	311	163	N	1b	MB1	Norm	Norm	Norm	
68	2	2	2	2	CVID	1179	826	325	410	134	N	1b	MB1	-	Norm	Norm	
14	2	2	2	2	CVID	2052	1446	539	2	117	Y						T3
67	2	2	2	2	CVID	3062	1157	1893	381	106	N	1b	MB0	Norm	Hi	Norm	
78	2	2	2	2	CVID	2876	1254	1617	273	127	N	II	MB2	Norm	Norm	Norm	
70	2	2	2	2	hypogamma	2990	1600	1308	0	27	N						T1
7	2	2	2	2	CVID	872	449	424	108	10	N	1b	MB1	-	Norm	Low	T1
8	2	2	2	2	CVID	797	293	519	181	102	N	2	MB1	Norm	Norm	Norm	T2
26	2	2	2	2	XLA	980	740	210	0	30							
43	2	2	2	2	CVID	1022	749	257	174	113	N	1b	MB1	-	Hi	Norm	
53	2	2	2	2	CVID	5640	2240	3110	1280	150	N	1b	MB0	-	Norm	Norm	
28	2	2	2	2	CVID	1922	1476	436	546	351	Y	2	MB1	Norm	Norm	Norm	
31	2	2	2	2	CVID	1528	1140	400	120	190	N	2	MB2	Norm	Norm	Norm	
63	2	2	2	2	CVID	1650	660	880	330	50	N						
56	2	2	2	2	CVID	1127	822	308	54	172	Y	II	MB2	Norm	Norm	Norm	
11	2	2	2	2	CVID	4920	1626	3228	860	794	N						T1
66	2	2	2	2	CVID	2870	2087	797	670	370		II	MB2	Norm	Norm	Norm	

Table 16 (part 2): Laboratory features of patients grouped by consensus clustering analysis.

Differential gene expression

To determine differentially expressed genes between different clinical groups, the first step was to remove the technical replicates from the data set and to include appropriate samples according to the question being asked. Therefore, for a comparison between XLA and CVID all XLA and CVID patients were included, and for complications of CVID only CVID patients were included in the analysis. The next step was to remove the effect of genes which were differentially expressed due to factors which were not of interest. Utilising batch effect remover within Partek, variance due to biological factors (gender and age) and due to experimental factors (microarray slide) were removed from the data. The batch effect remover applies a mixed model analysis of variance (ANOVA) to estimate the effects due to the selected variables, and adjusts the data to what it would be if all batches were equal (Ayroles and Gibson 2006). Following batch effect removal, statistical analysis of the resultant data for differentially expressed genes due to gender, age and microarray slide returned a p value of 1.0 for all genes. Differentially expressed genes between the two clinical groups of interest were determined using a 2 sample T test without assuming equal variance between the groups, and a false discovery rate (FDR) of <0.05.

FDR is a statistical method which allows correction for multiple comparisons, and is particularly suited to microarray experiments. As each microarray examines >24,000 genes, in effect >24,000 different experiments/comparisons have been performed between the study subjects. Traditional multiple comparison corrections (such as Bonferroni) work by using a p value which is calculated by dividing the original desired p value by the number of comparisons. While this approach works when there are a small number of comparisons, the large number of comparisons implicit with

microarrays makes this approach unfeasible. For example, applied to this project, the Bonferroni correction would result in using a maximum p value of 0.05/24,000 = 2.08×10^{-6}. Whereas the Bonferroni correction controls the chance of any false positive results, the FDR accepts that there will be false positive results, and instead controls the proportion of false positive results. For example, a FDR of <0.05 means that 5% of the results are expected to be type 1 errors, and therefore of a list of 100 differentially expressed genes, 5 of the genes would be false positives. While the trade off of using FDR is that some of the results will be false positives, the benefit is an increase in sensitivity for detecting differentially expressed genes.

The workflow is summarised in Figure 19. The following questions were asked: are there differentially expressed genes between XLA and CVID, and are there differentially expressed genes for the development of granuloma, autoimmunity, splenomegaly, persistent lymphadenopathy or bronchiectasis in CVID. Some authors have suggested that granulomatous disease and persistent lymphadenopathy should be combined under the heading 'polyclonal proliferation', and also that autoimmune disease should be separated into autoimmune cytopenias and OSAI (Orange *et al*. 2011). These grouping were also included in the analysis. A summary of the comparisons made are shown in Table 17.

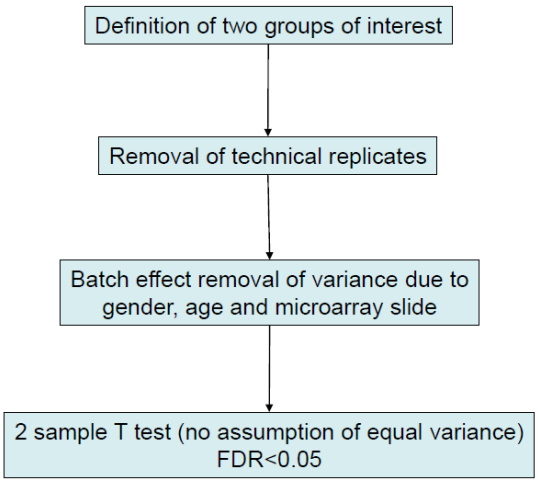


Figure 19: Workflow for determination of differentially expressed genes.

Comparison	Group 1	Group 2
1	XLA	CVID
2	CVID with granuloma	CVID without granuloma
3	CVID with autoimmunity	CVID without autoimmunity
4	CVID with splenomegaly	CVID without splenomegaly
5	CVID with persistent lymphadenopathy	CVID without persistent lymphadenopathy
6	CVID with bronchiectasis	CVID without bronchiectasis
7	CVID with any of granuloma, autoimmunity, splenomegaly and/or persistent lymphadenopathy	CVID without any of these complications
8	CVID with polyclonal proliferation	CVID without polyclonal proliferation
9	CVID with autoimmune cytopenia	CVID without autoimmune cytopenia
10	CVID with OSAI	CVID without OSAI

Table 17: Clinical groups between which differential gene expression was analysed utilising workflow in Figure 19.

45 differentially expressed genes were identified between XLA and CVID, and are shown below in Table 18 and Figure 20. Heatmaps are a method of graphically displaying the data, where values on a two-dimensional table are represented by colour. Clustering similar samples and similar genes near to each other on the axes increases the usability of the heatmap and assists visual interrogation of the data. Hierarchical clustering is performed by determining the Euclidean distance between the samples. The Euclidean distance between two data points on a graph is the distance in a straight line between them. The same principle holds whether the graph is 2D, 3D or, as in this case, highly dimensional. The Euclidean distance between two samples is represented on the dendrogram by height of the branch connecting the samples.

Unsurprisingly, many of the identified genes are those specific to B cells which are absent in XLA. Looking specifically at the expression of btk mRNA, this was lower in XLA patients but did not reach statistical significance. There are a number of possible explanations for this. There were only 5 XLA patients in the cohort and so this analysis may have lacked the statistical power to detect this difference. In addition, the lack of a functioning protein does not necessarily correlate with an absence of mRNA. Although the XLA patients have a mutation in their btk gene preventing the production of btk, there may still have been mRNA expression detectable by the probe used on the microarray slide.

Gene Symbol	Gene product	p-value	FDR
SPIB	transcription factor Spi-B	1.83E-12	3.37E-08
POU2AF1	POU domain class 2-associating factor 1	3.71E-09	3.41E-05
FCER2	FcεRII, CD23	2.61E-08	0.00013
LAMC1	laminin subunit gamma-1	2.82E-08	0.00013
NCAPD3	non-SMC condensin II complex subunit D3	1.96E-07	0.000723
FCRLA	Fc receptor-like A	3.92E-07	0.001202
IGLL1	Immunoglobulin lambda-like polypeptide 1, CD179B	5.26E-07	0.001383
SLX1B	SLX1 structure-specific endonuclease subunit homolog B	7.15E-07	0.001645
CXCR5	C-X-C chemokine receptor type 5	9.91E-07	0.002027
CD20	B-lymphocyte antigen CD20	1.46E-06	0.002691
NRTN	Neurturin	1.82E-06	0.003052
CD22	B-lymphocyte cell adhesion molecule	5.55E-06	0.008505
BGN	biglycan	9.42E-06	0.01283
PLEKHG1	pleckstrin homology domain-containing family G member 1	9.76E-06	0.01283
DUSP14	dual specificity phosphatase 14	1.08E-05	0.013265
SEC31A	protein transport protein Sec31A	1.19E-05	0.013279
TNFRSF13B	TACI	1.23E-05	0.013279
S100B	S100 calcium binding protein B	1.36E-05	0.013861
CNTNAP2	Contactin-associated protein-like 2	1.65E-05	0.015606
SUNC1	SUN domain-containing protein 3	1.78E-05	0.015606
LIMS1	LIM and senescent cell antigen-like-containing domain protein 1	1.78E-05	0.015606
TNPO1	transportin 1	1.96E-05	0.016434
TSPAN15	tetraspanin 15	2.31E-05	0.018441
ZNF785	zinc finger protein 785	3.36E-05	0.025789
TCF4	transcription factor 4	3.53E-05	0.026014
ANKRD10	ankyrin repeat domain 10	3.71E-05	0.026228
CD19	B-lymphocyte antigen CD19	4.06E-05	0.027667
PQLC2	PQ-loop repeat-containing protein 2	4.23E-05	0.02783
CD79A	Igα	4.63E-05	0.02879
TSPAN5	tetraspanin 5	4.69E-05	0.02879
ST6GALNAC4	ST6 sialyltransferase	5.28E-05	0.031338
DUS4L	dihydrouridine synthase 4-like	5.71E-05	0.032849
TRADD	TNFRSF1A-associated via death domain	6.26E-05	0.034836
THOC2	THO complex 2	6.44E-05	0.034836
ABI1	Abelson interactor 1	7.04E-05	0.035538
RBMS2	RNA binding motif, single stranded interacting protein 2	7.08E-05	0.035538
LOC402110	hypothetical locus 402110	7.15E-05	0.035538
SFPQ	splicing factor proline/glutamine-rich	7.37E-05	0.035683
MOSC2	MOSC domain-containing protein 2	8.10E-05	0.038211
GPR61	G protein-coupled receptor 61	9.07E-05	0.041714
CLASP2	cytoplasmic linker associated protein 2	0.000102	0.04596
BRD7	bromodomain-containing protein 7	0.000105	0.046213
FAM98C	family with sequence similarity 98	0.000114	0.046845
FBXL18	F-box and leucine-rich repeat protein 18	0.000114	0.046845
RNF113B	ring finger protein 113B	0.000115	0.046845

Table 18: List of differentially expressed genes between XLA and CVID.

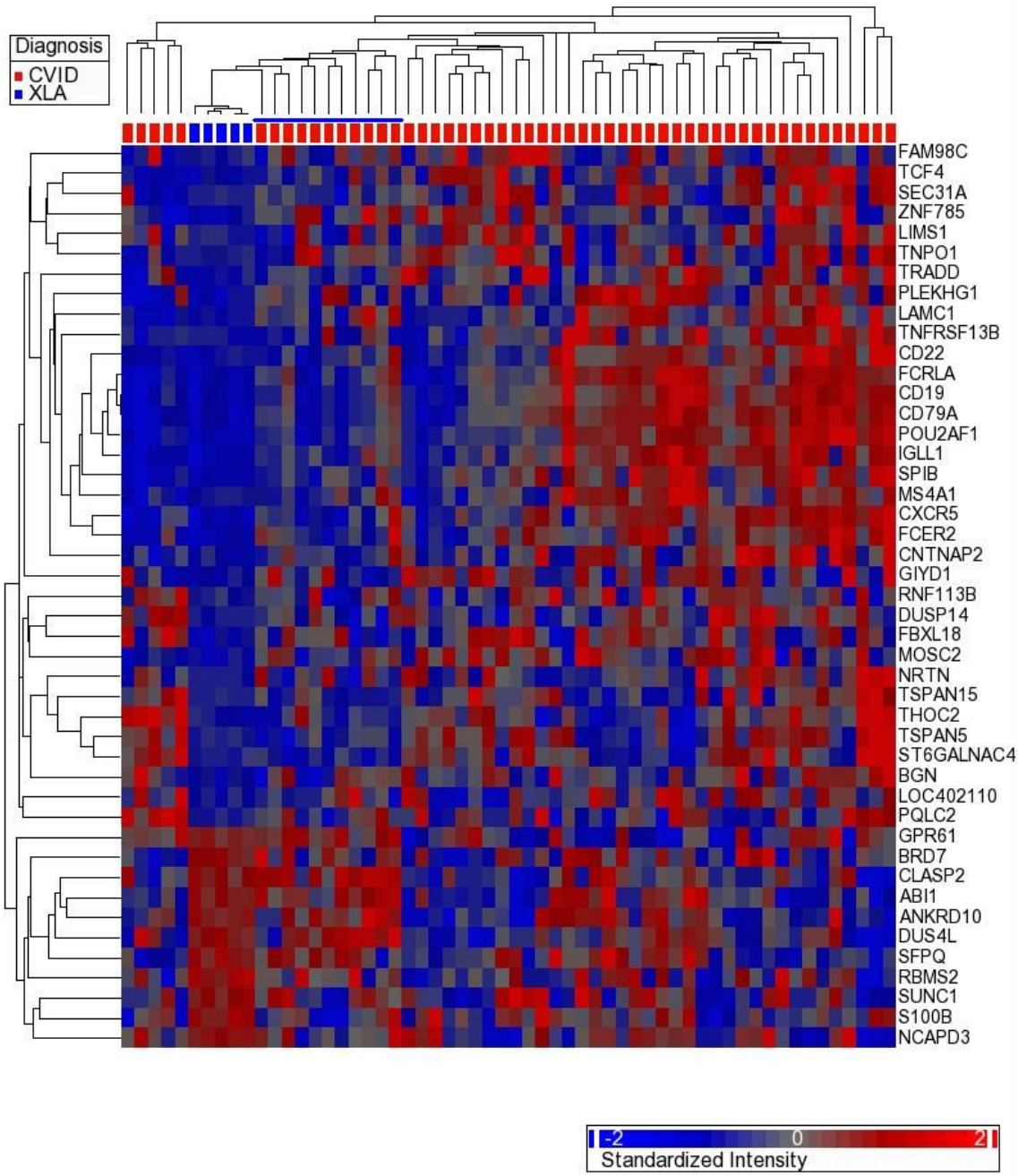


Figure 20: Heatmap of 45 genes differentially expressed between XLA and CVID. Individual patients are along the x axis and their diagnosis indicated by red (CVID) and blue (XLA). The 45 genes of interest are along the y axis, and the differential gene expression is indicated by colour and intensity. 11 CVID patients were relatively similar to the XLA patients and are marked by a blue line. These patients were: 30, 32, 49, 31, 27, 58, 44, 59, 10, 56 and 14.

The features of the 11 CVID patients who clustered closest to the XLA patients on the heatmap in Figure 20 were examined. These 11 CVID patients had a lower number of B cells compared to the other CVID patients as demonstrated in Figure 21. This in itself is perhaps expected, though of greater potential interest are the complication

rates in this subgroup: granuloma 0% (compared to 11.3% of CVID patients), autoimmunity 18.2% (compared to 26.4%), splenomegaly 9.1% (26.4%) and lymphadenopathy 9.1% (13.2%). The complication rates in these patients appear to be lower than in other CVID patients, though statistical significance was not reached. Patients with low B cells of <1% total lymphocytes were classified as B- in the EUROclass study and were not investigated any further (Wehr *et al*. 2008). However, these patients may represent a group less likely to develop complications.

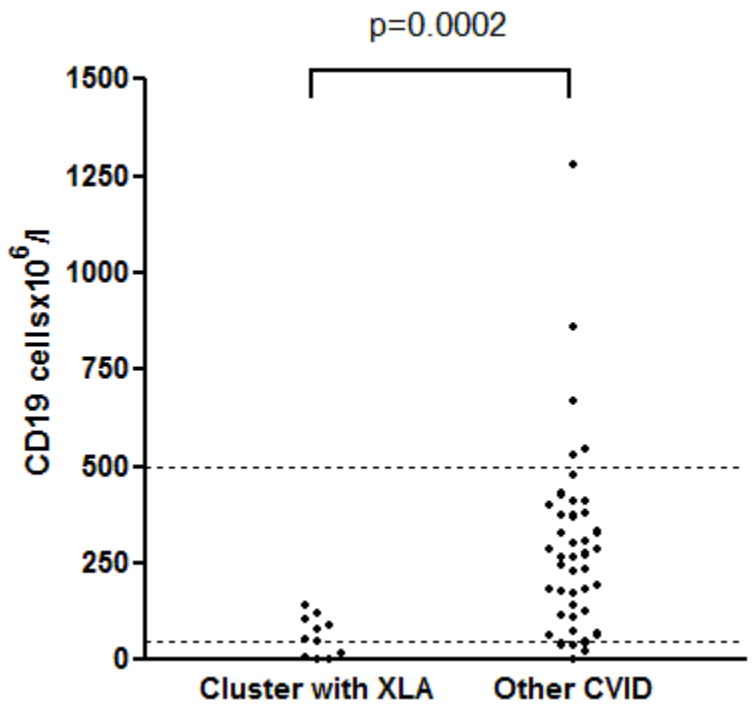


Figure 21: CD19 enumeration of CVID patients who clustered closest to XLA patients compared to other CVID patients. p=0.0002 by two-tailed Mann Whitney test. Reference range indicated by dotted lines (50-250 x10^6 cells/l).

Differentially expressed genes associated with the presence of granulomatous disease, splenomegaly and lymphadenopathy were also identified. No differentially expressed genes were identified using the workflow in Figure 19 for autoimmunity or bronchiectasis. Comparing CVID patients with any of granuloma, autoimmunity,

splenomegaly and/or lymphadenopathy with CVID patients without any complications did not demonstrate any differentially expressed genes between these groups (comparison 7 from Table 17). Examining the alternative classifications of CVID subgroups yielded potentially interesting findings. When granulomatous disease and persistent lymphadenopathy were combined, there were no differentially expressed genes despite this being a larger group. This suggests that these two conditions may have different underlying causes. Conversely, separating out autoimmune cytopenias from other autoimmune disease identified a number of genes which may be involved in this process.

The results are summarised in Table 19 which also shows the number of differentially expressed genes if the FDR rate is relaxed from 0.05 to 0.1, 0.25 and 0.5. While using a higher FDR rate would result in more type 1 errors being present within the gene list, it would also result in numerically larger gene lists containing more genes of interest. These larger gene lists may lead to improved clustering and prediction models and are investigated later in this chapter. Figure 22 through to Figure 33 show heatmaps generated from the differentially expressed genes.

Comparison	FDR<0.05	FDR<0.1	FDR<0.25	FDR<0.5
1 (XLA vs CVID)	45	65	137	332
2 (granuloma)	23	77	215	825
3 (autoimmunity)	0	0	16	1312
4 (splenomegaly)	43	135	851	3242
5 (lymphadenopathy)	4	8	32	79
6 (bronchiectasis)	0	0	0	0
7 (g, a, s and/or l)	0	0	0	196
8 (polyclonal proliferation)	0	0	52	1918
9 (autoimmune cytopenia)	41	209	1066	4189
10 (OSAI)	4	4	22	54

Table 19: Numbers of differentially expressed genes identified from the comparisons shown in Table 17 for different statistical stringencies. g, a, s and/or l: the presence of any of granuloma, autoimmunity, splenomegaly and/or lymphadenopathy

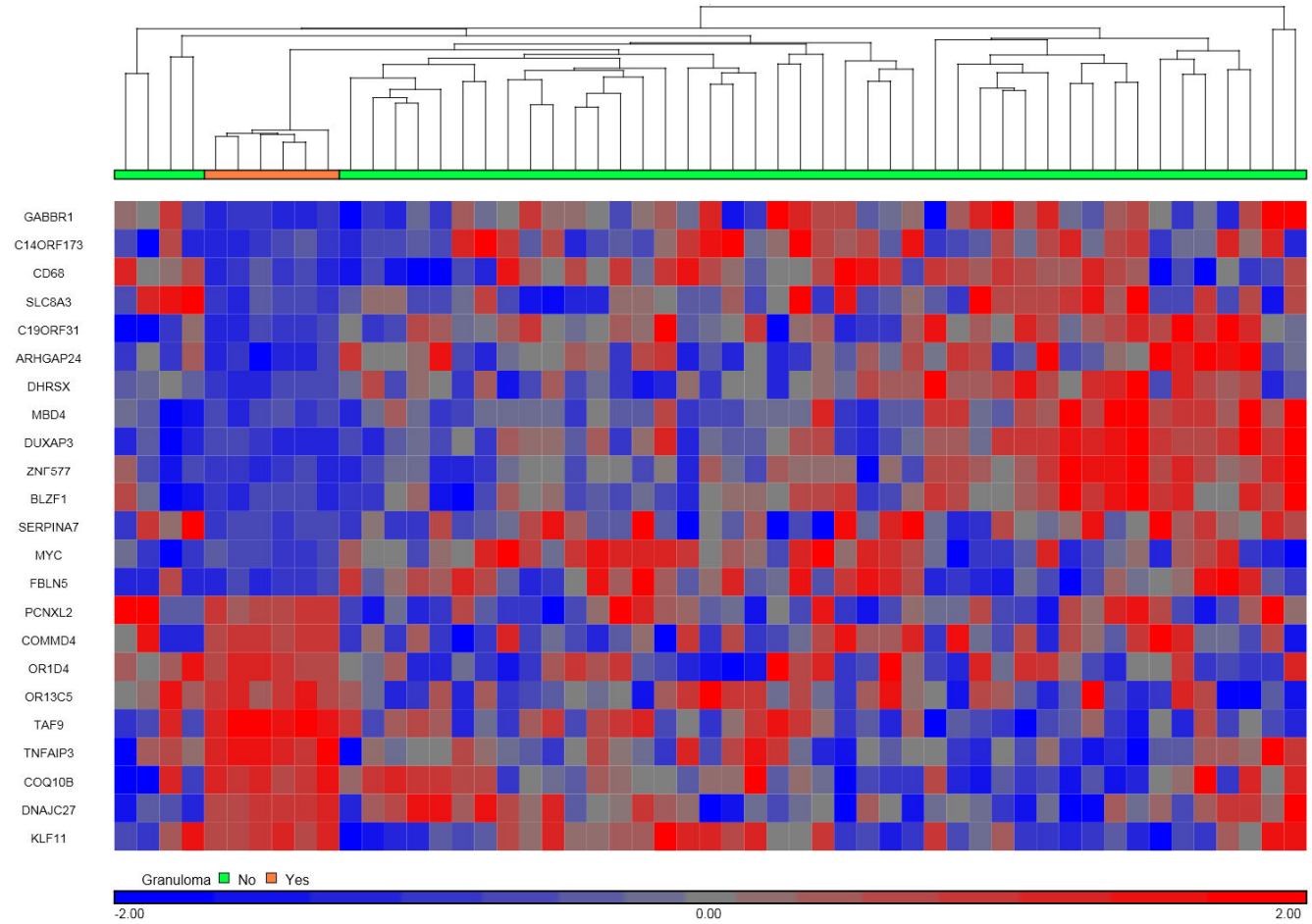


Figure 22: Heatmap of genes differentially expressed between CVID patients with and without granulomatous disease (FDR<0.05, 23 genes). Individual patients are along the x axis and the presence of granuloma indicated by orange (present) and blue (not present). The genes of interest are along the y axis, and the differential gene expression is indicated by colour and intensity.

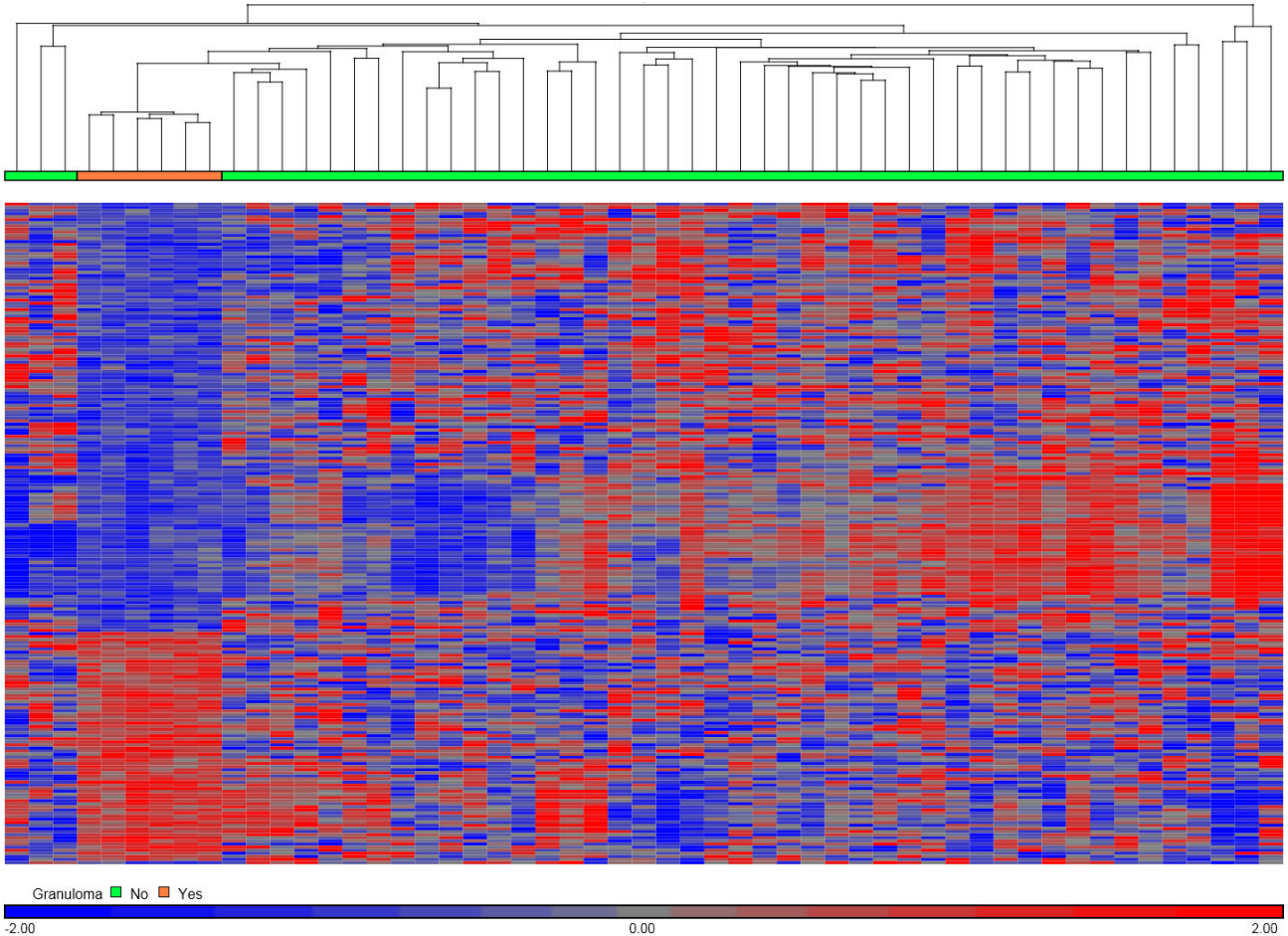


Figure 23: Heatmap of genes differentially expressed between CVID patients with and without granulomatous disease (FDR<0.25, 215 genes). Individual patients are along the x axis and the presence of granuloma indicated by orange (present) and blue (not present). The genes of interest are along the y axis, and the differential gene expression is indicated by colour and intensity.

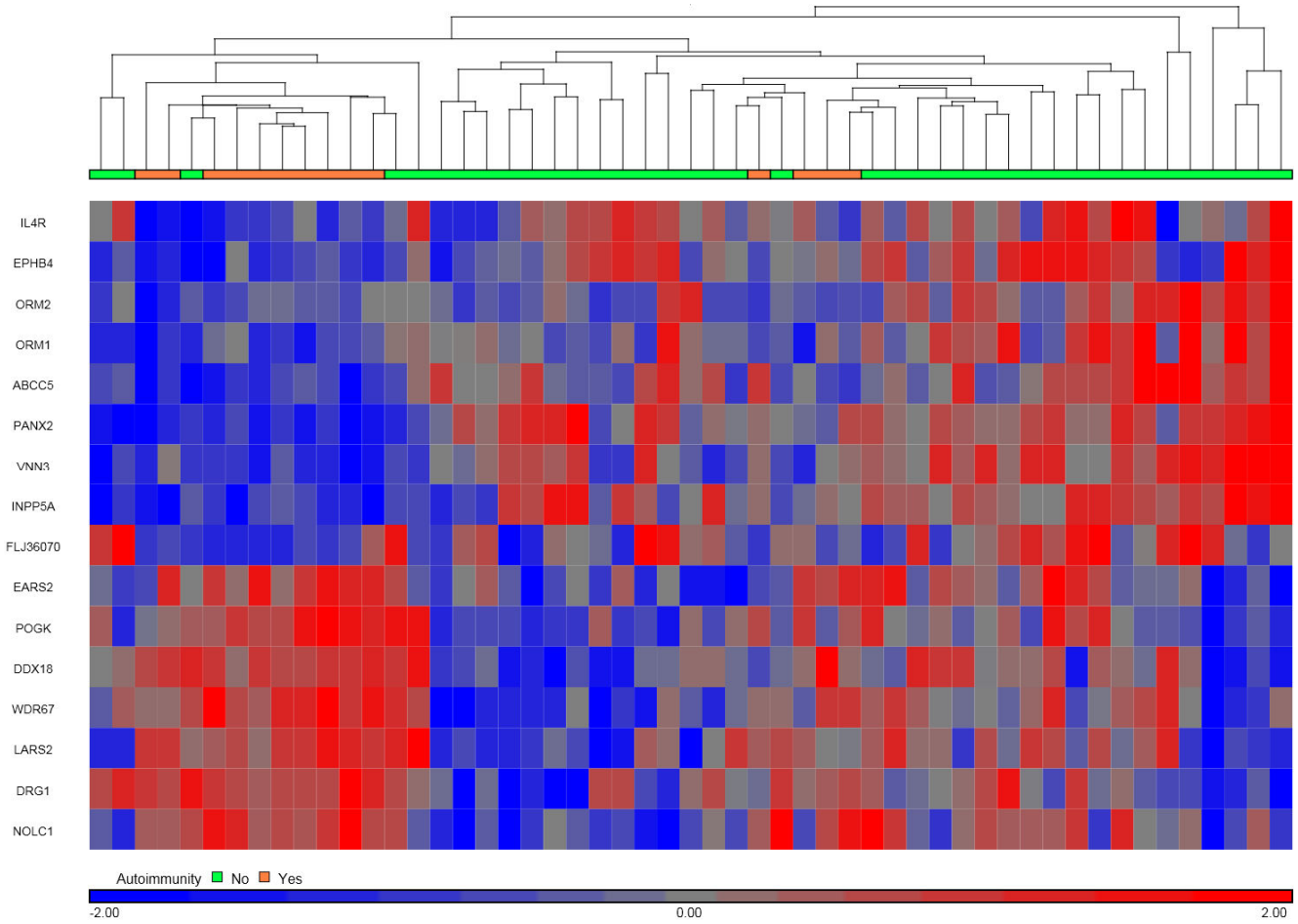


Figure 24: Heatmap of genes differentially expressed between CVID patients with and without autoimmune disease (FDR<0.25, 16 genes). Individual patients are along the x axis and the presence of autoimmunity indicated by orange (present) and blue (not present). The genes of interest are along the y axis, and the differential gene expression is indicated by colour and intensity.

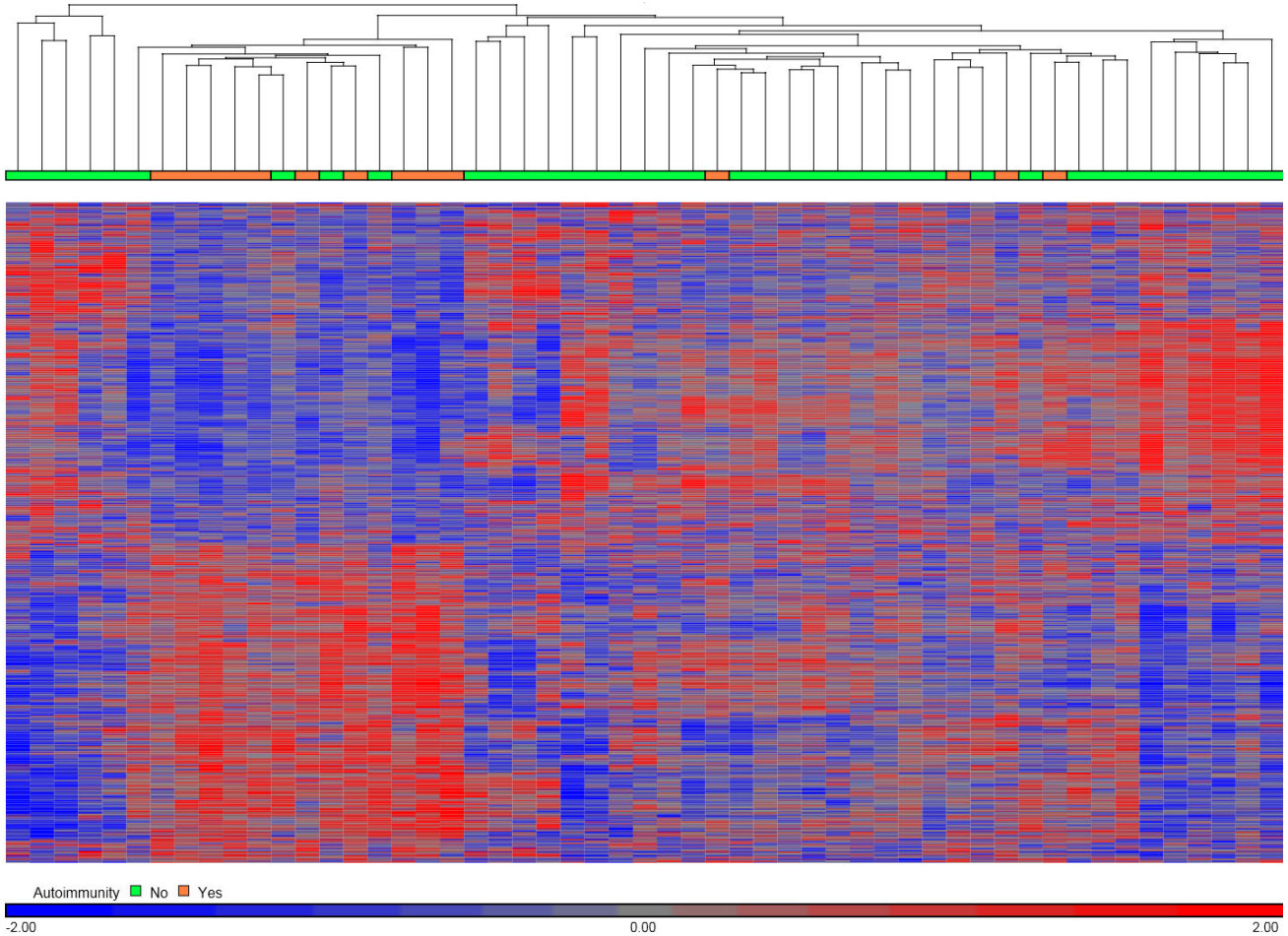


Figure 25: Heatmap of genes differentially expressed between CVID patients with and without autoimmune disease (FDR<0.5, 1312 genes). Individual patients are along the x axis and the presence of autoimmunity indicated by orange (present) and blue (not present). The genes of interest are along the y axis, and the differential gene expression is indicated by colour and intensity.

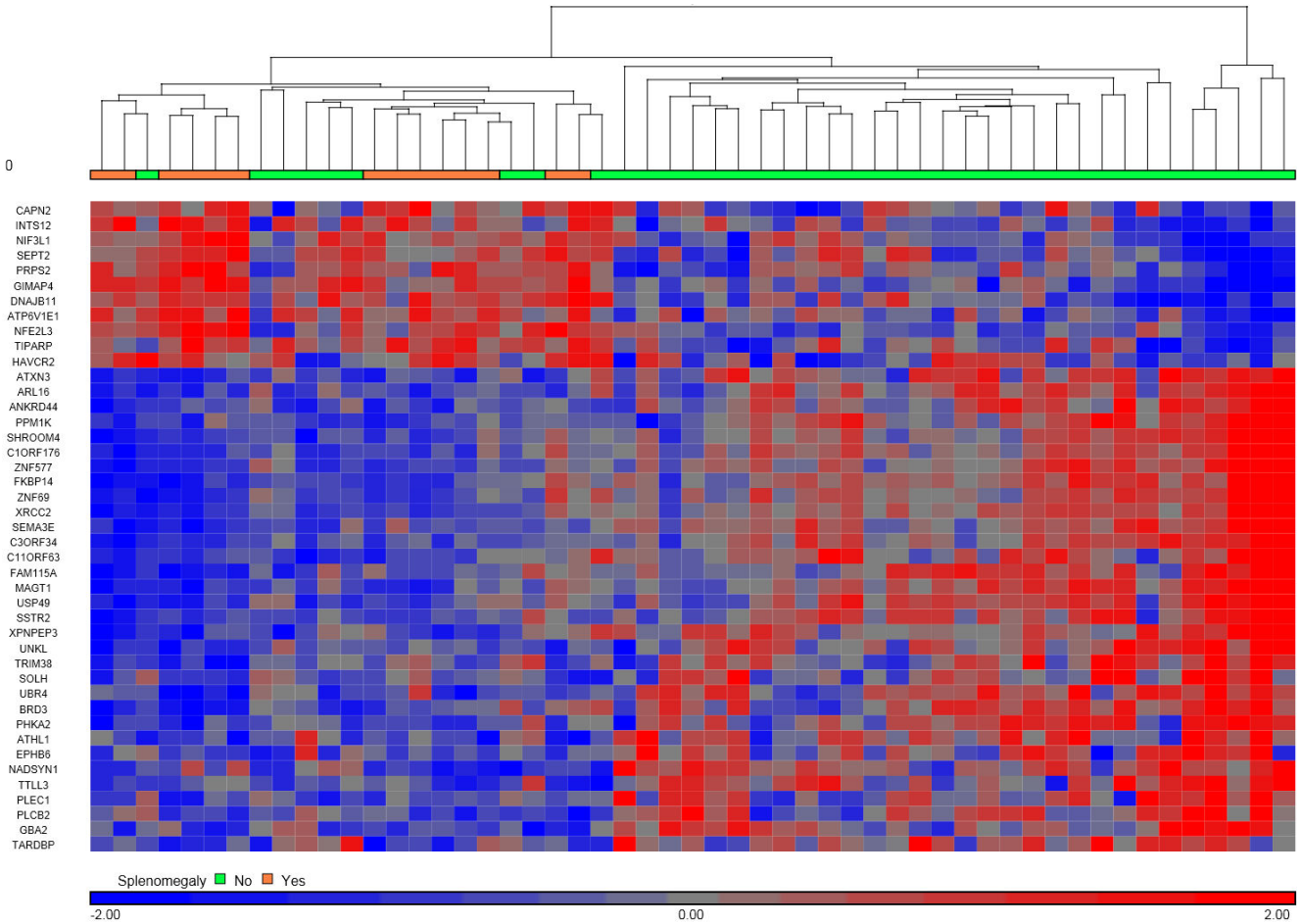


Figure 26: Heatmap of genes differentially expressed between CVID patients with and without splenomegaly (FDR<0.05, 43 genes). Individual patients are along the x axis and the presence of splenomegaly indicated by orange (present) and blue (not present). The genes of interest are along the y axis, and the differential gene expression is indicated by colour and intensity.

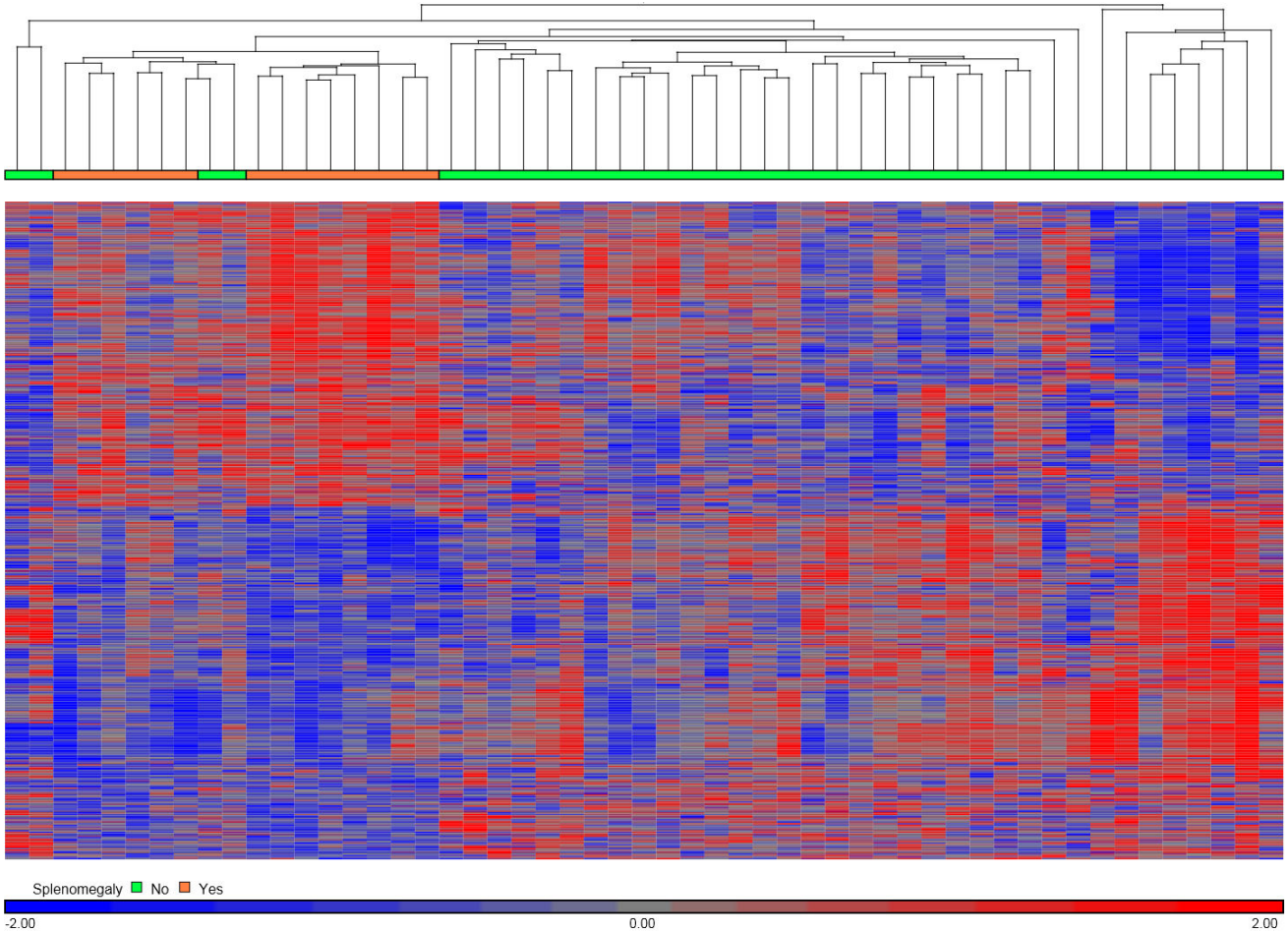


Figure 27: Heatmap of genes differentially expressed between CVID patients with and without splenomegaly (FDR<0.25, 851 genes). Individual patients are along the x axis and the presence of splenomegaly indicated by orange (present) and blue (not present). The genes of interest are along the y axis, and the differential gene expression is indicated by colour and intensity.

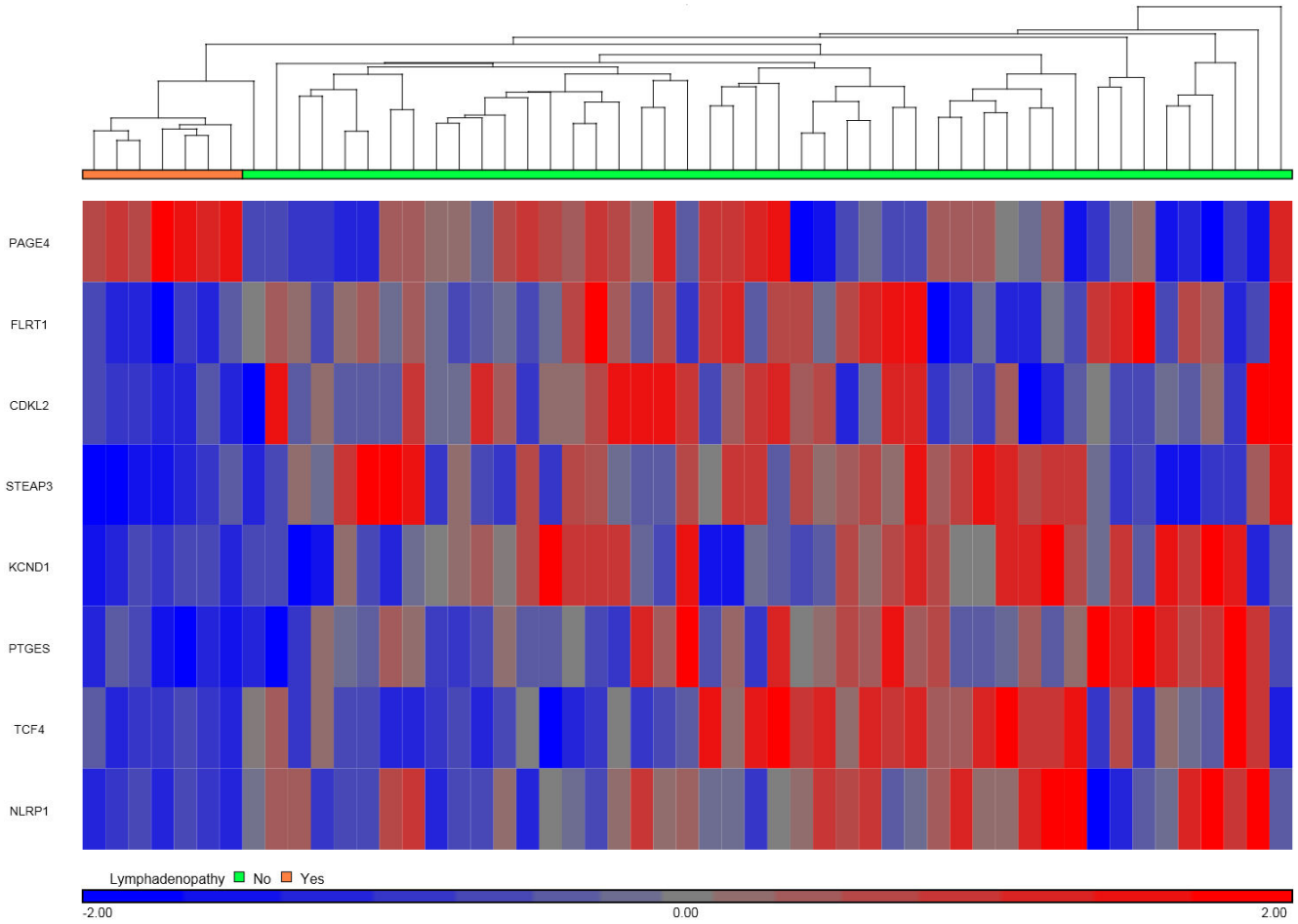


Figure 28: Heatmap of genes differentially expressed between CVID patients with and without persistent lymphadenopathy (FDR<0.1, 8 genes). Individual patients are along the x axis and the presence of lymphadenopathy indicated by orange (present) and blue (not present). The genes of interest are along the y axis, and the differential gene expression is indicated by colour and intensity.

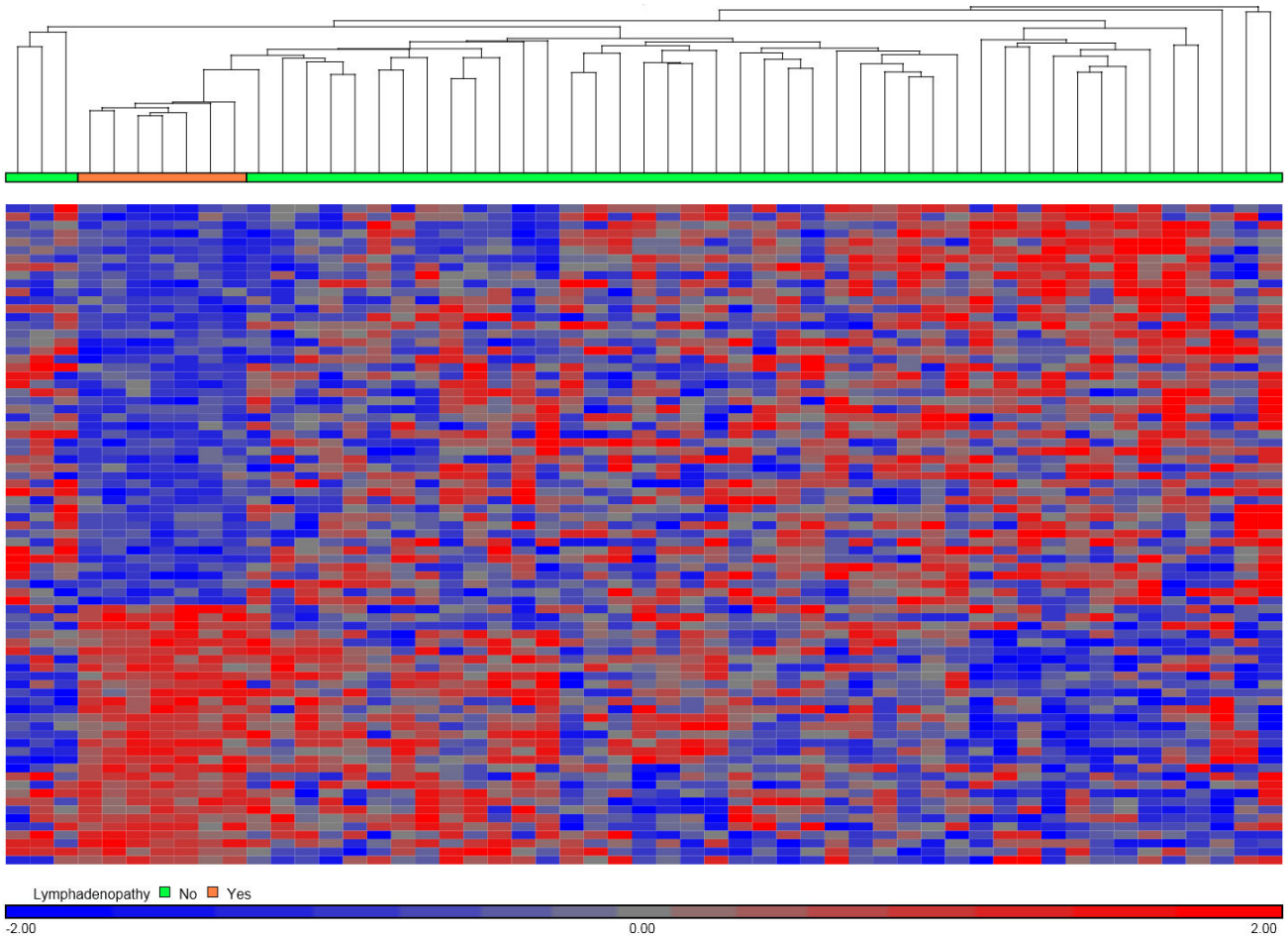


Figure 29: Heatmap of genes differentially expressed between CVID patients with and without persistent lymphadenopathy (FDR<0.5, 79 genes). Individual patients are along the x axis and the presence of lymphadenopathy indicated by orange (present) and blue (not present). The genes of interest are along the y axis, and the differential gene expression is indicated by colour and intensity.

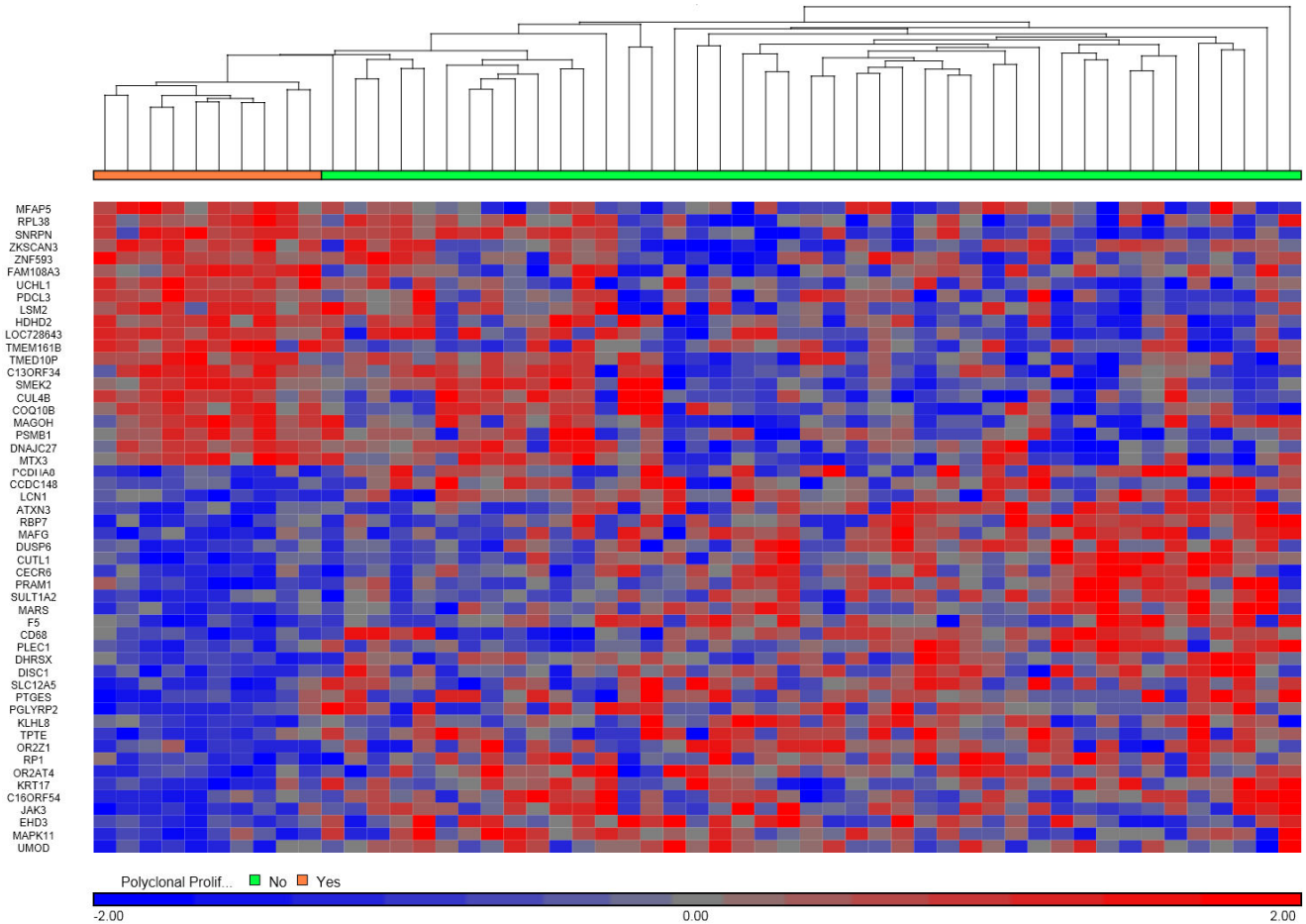


Figure 30: Heatmap of genes differentially expressed between CVID patients with and without polyclonal proliferation (FDR<0.25, 52 genes). Individual patients are along the x axis and the presence of polyclonal proliferation indicated by orange (present) and blue (not present). The genes of interest are along the y axis, and the differential gene expression is indicated by colour and intensity.

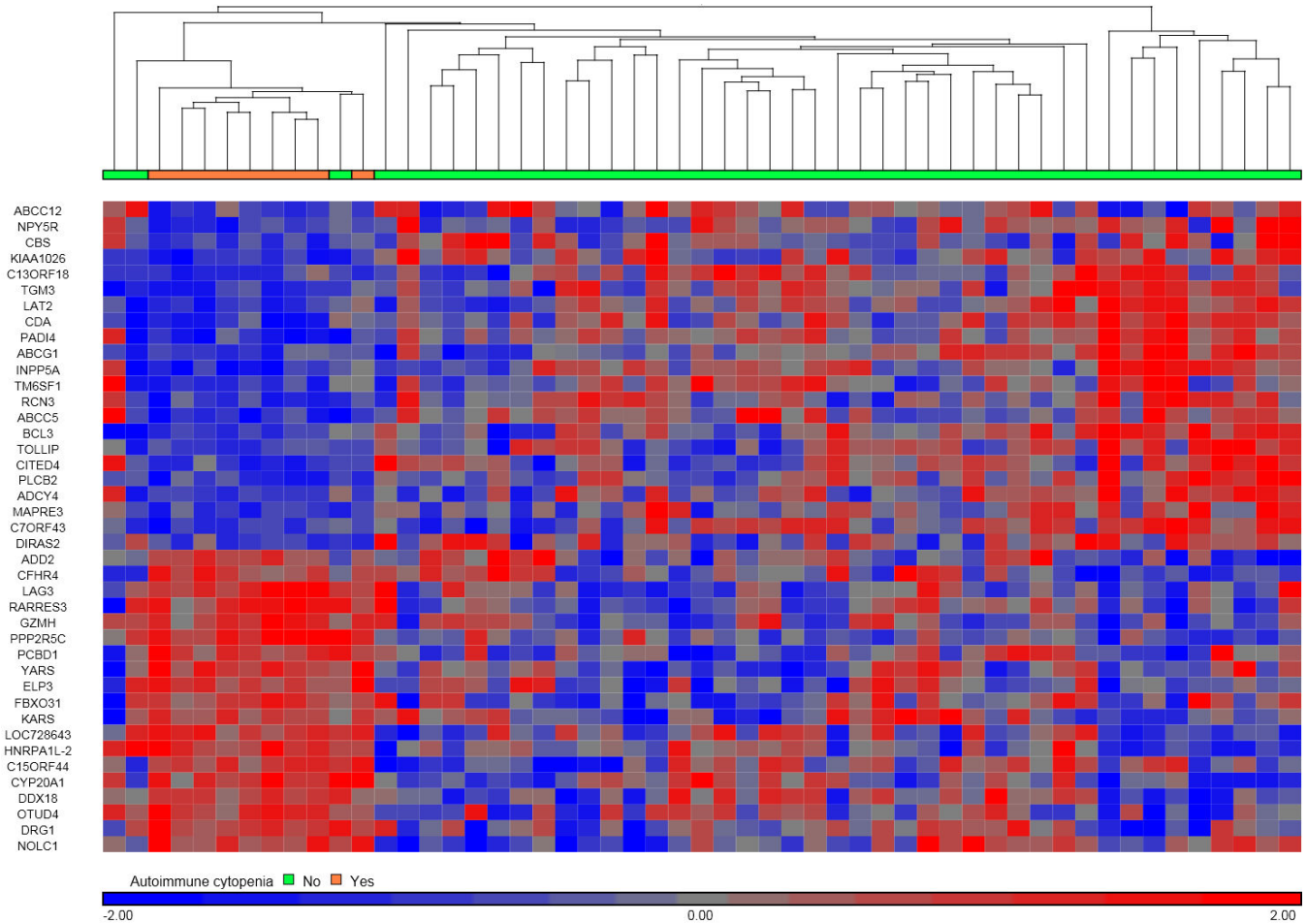


Figure 31: Heatmap of genes differentially expressed between CVID patients with and without autoimmune cytopenia (FDR<0.05, 41 genes). Individual patients are along the x axis and the presence of autoimmune cytopenia indicated by orange (present) and blue (not present). The genes of interest are along the y axis, and the differential gene expression is indicated by colour and intensity.

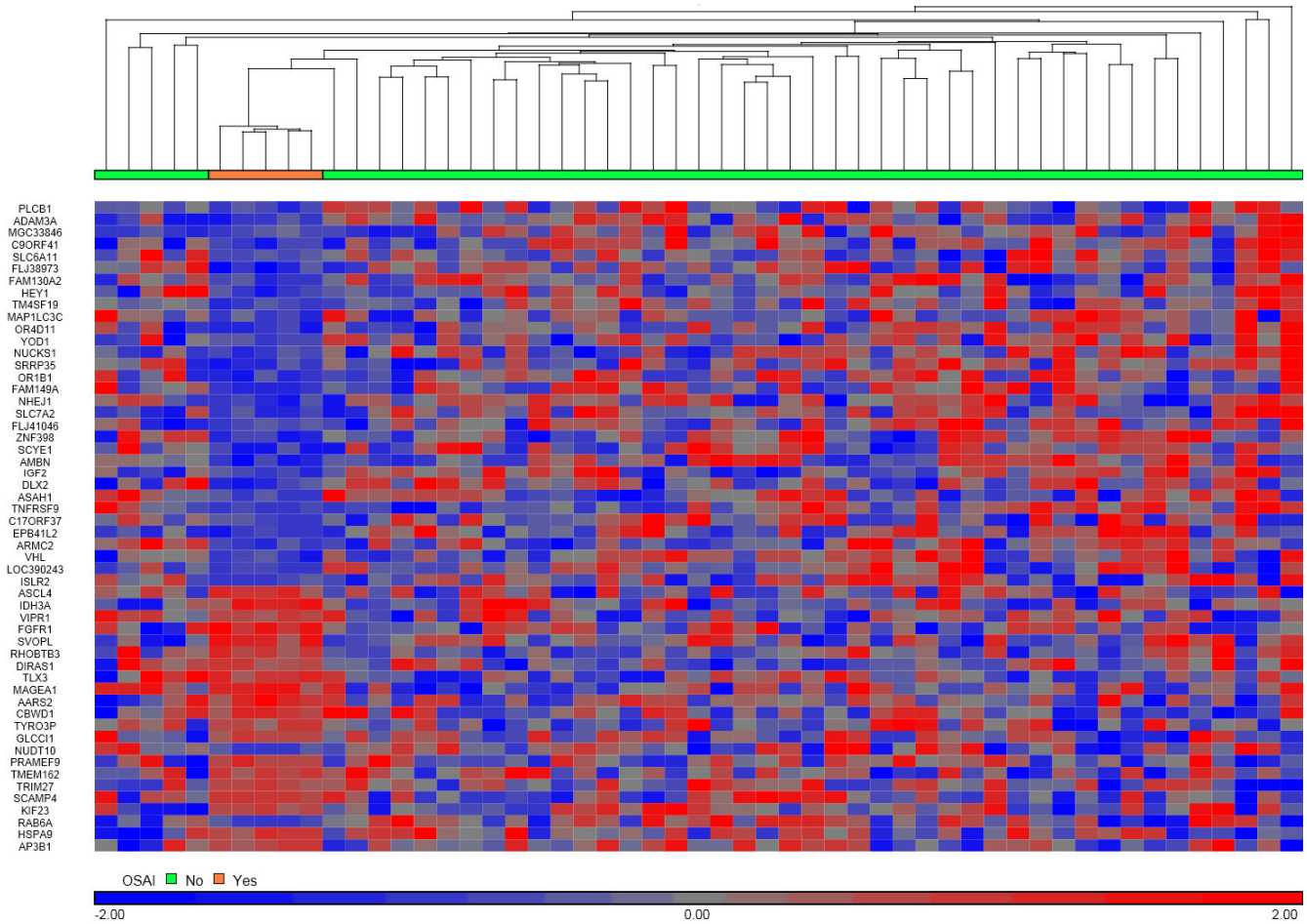


Figure 32: Heatmap of genes differentially expressed between CVID patients with and without OSAI (FDR<0.5, 54 genes). Individual patients are along the x axis and the presence of OSAI indicated by orange (present) and blue (not present). The genes of interest are along the y axis, and the differential gene expression is indicated by colour and intensity.

The heatmaps show the ability to cluster patients with various complications of CVID together. Figure 22 and Figure 23 show that patients with granulomatous disease cluster closely together when utilising differentially expressed genes. Utilising fewer genes with a more stringent FDR appears to lead to greater separation and this was investigated further when attempting to create class prediction models. A similar picture is seen in Figure 28 and Figure 29 for patients with persistent lymphadenopathy. The heatmaps for splenomegaly and lymphadenopathy show that it is also possible to cluster these patients together, though it is not possible to define a clear cut group as with the granulomatous disease and persistent lymphadenopathy groups. Figure 33 shows that three main groups can be identified; one with a high rate of complications, one with a low rate, and a final group with no complications.

The gene lists generated are of potential interest in investigating the pathogenetic mechanisms underlying the development of these complications, and are further explored in the biological analysis section (page 118). The ability of these gene lists to differentiate patients with and without different complications is also of interest. However, in the above figures, all the samples were used to generate the list of differentially expressed genes. The ability to class predict needs to be validated using an independent data set which was not used to generate either the gene list or the prediction model.

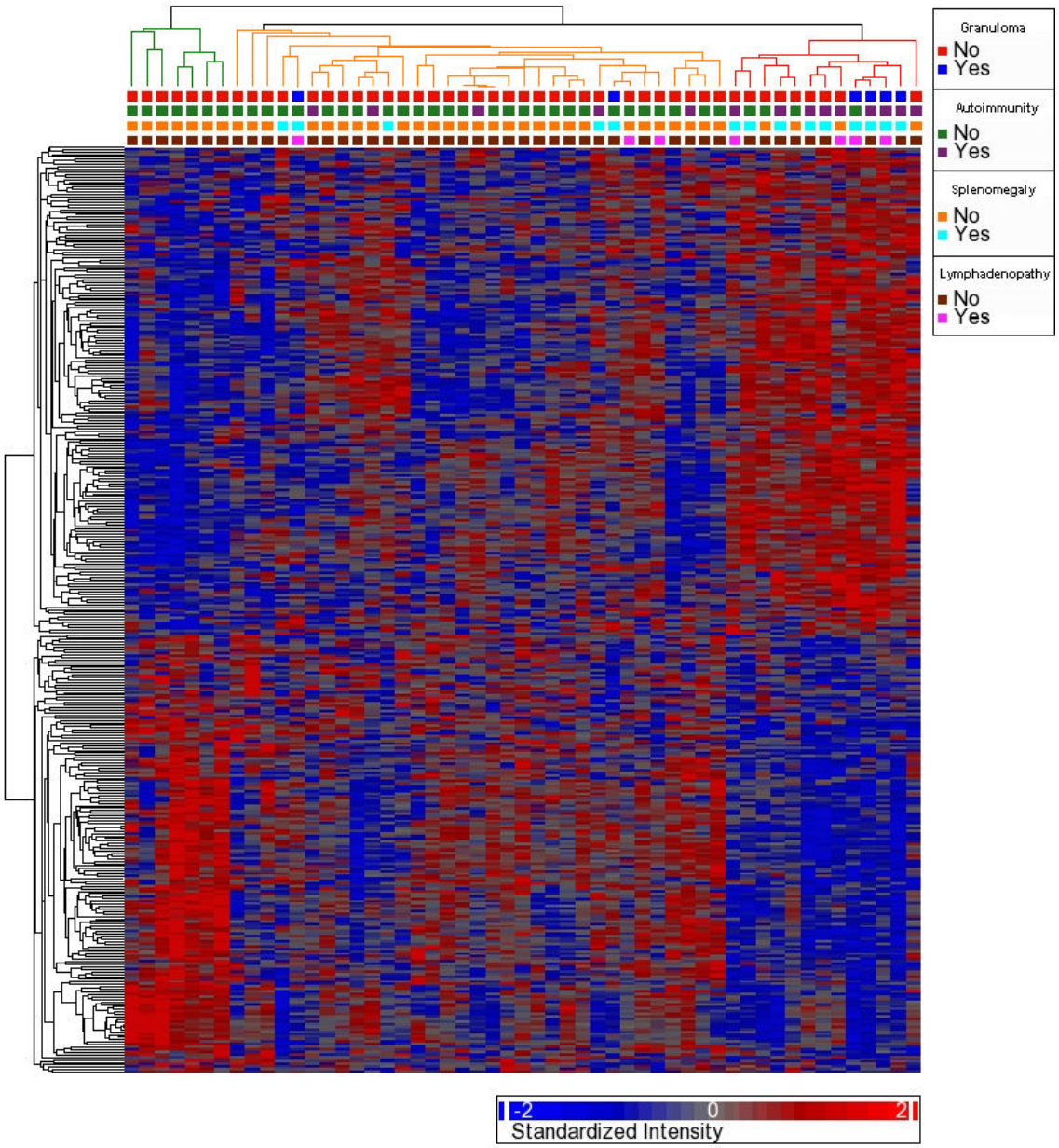


Figure 33: Heatmap of genes differentially expressed between CVID patients with any of granuloma, autoimmunity, splenomegaly and/or persistent lymphadenopathy and CVID patients with none of these complications (FDR<0.5, 196 genes). The three main clusters are highlighted in green, yellow and red.

Class Prediction

A number of different methods exist to produce prediction models. A comparison of support vector machine (SVM) and K nearest neighbour (KNN) was performed using the ability to predict gender. Both methods start by plotting the training data in multidimensional space. SVM attempts to separate the microarrays into areas, in this example into areas for males and for females. With 2D data, a line would be used to separate the data, 3D data a plane, and in higher dimensions a multidimensional hyperplane. KNN uses the nearest neighbour in terms of Euclidean distance to predict the properties of the test subject. 4 male and 4 female patients were removed from the data, and used to test the predictive power of the models. The SVM model only correctly classified 4 of the 8 test patients. KNN was able to correctly classify the gender of all 8 test patients. The ability to predict gender would be expected to be a relatively undemanding task of a prediction model, and KNN was used for the following analyses.

Table 20 shows the performance of KNN using all genes represented on the microarray, as determined by full leave-one-out validation. The correct rate shows the absolute numbers which were correct, but this figure is sensitive to prevalence. For example, a prediction model for granuloma which classified every single patient as not having granuloma would clearly be a poor model, but when applied to this group of patients would correctly predict all the patients without granuloma (48/53 patients) and have a correct rate of 90.6%. The normalised correct rate (the average of percentage correct from the predicted positive group and percentage correct from predicted negative group) allows for unbalanced data and is therefore a better indicator of the performance of a predictive test.

Complication	Correct rate	Normalised correct rate
Granuloma	84.9	62.4
Autoimmunity	56.6	40.8
Splenomegaly	73.58	61.54
Lymphadenopathy	73.6	48.5
Bronchiectasis	52.8	52.3
G,A,S and/or L	45.3	44.7

Table 20: Correct and normalised correct rates of KNN using all genes represented on the microarray as determined by full leave-one-out validation. G,A,S and/or L: the presence of any of granuloma, autoimmunity, splenomegaly and/or lymphadenopathy. The normalised correct rate indicates what the correct rate would be if the rates if those with and without the complication where the same (eg if 50% had autoimmunity and 50% did not have autoimmunity) and corrects for unbalanced proportions in the groups being tested.

Utilising a restricted gene list, generated from differential gene expression, may improve class prediction. Such an approach would focus on those genes likely to discriminate between the groups, and exclude genes which are unlikely to be differentially expressed. However, to accurately test class prediction, the gene list needs to be created from samples separate from the test group. For granulomatous disease, samples 13 (with granuloma), 14, 23 and 27 (without granuloma) were removed from the dataset prior to generation of a gene list of differentially expressed genes as per the workflow in Figure 19. For splenomegaly samples 9, 10, 14 and 27 (with splenomegaly), 11, 13, 23 and 28 (without splenomegaly) were held back as a test group, and for lymphadenopathy samples 11, 28 (with lymphadenopathy), 10, 13, 23 and 29 (without lymphadenopathy) were held back. The results are shown in Table 21.

Full leave-one-out validation performed with Partek to assess KNN using these gene lists showed very good normalised correct rates. However, this process tested the samples which had been used to generate the gene list. When applied to the test groups, class prediction performed less well. Consistent with the results in Table 20, attempts to predict granulomatous disease and splenomegaly were more successful. In

the granuloma test group 3 out of 4 patients were correctly classified, but there was a failure to correctly classify the patient with granulomatous disease. Increasing the number of genes by using the top x number of genes with the most significant p value did not help to improve prediction. With the splenomegaly test group, increasing the number of genes used led to sample 28 being correctly classified, but with sample 11 becoming incorrectly classified and therefore did not change the overall performance.

Complication	Granuloma						Splenomegaly						Lymphadenopathy					
No of genes (FDR)	90 (<0.05)		250		500		87 (<0.1)		250		500		62 (<0.5)		250		500	
Prediction of test samples green=correct red=incorrect	14	13	14	13	14	13	9	13	9	11	9	11	10	11	10	11	10	11
	23		23		23		10	23	10	13	10	13	29	13	29	13	29	13
	27		27		27		11	28	14	23	14	23		23		23		23
							14		27		27			28		28		28
							27		28		28							

Table 21: Results of KNN class prediction for granuloma, splenomegaly and lymphadenopathy. Three KNN models were created for each complication and the number of genes used indicated. The accuracy of class prediction was determined using the test group. Sample numbers of patients correctly classified are in green and those incorrectly classified are in red.

Class prediction has not been able to definitively classify complications using this dataset. Use of the whole genome gave poor results as might have been expected from PCA shown in Figure 17. Differentially expressed genes provided improved predictive ability, and there is the potential that with a larger patient set class prediction may be successful. The lists of differentially expressed genes are themselves of interest and may provide insight into the pathogenetic mechanisms of different CVID subgroups, warranting further analysis.

Biological analysis of differentially expressed genes

The Database for Annotation, Visualization and Integrated Discovery (DAVID) (http://david.abcc.ncifcrf.gov/home.jsp) provides an online bioinformatic tool to enable the extraction of biological features and meaning from lists of differentially expressed genes (Dennis *et al*. 2003). The process of enrichment highlights the biological annotations (such as gene ontology terms) which are statistically overrepresented within the gene list and allows the discovery of biological processes which may be involved in the disease process. For example if 5% of the genes in a gene list are involved in a particular process compared to 0.5% of genes in the background population (the list of all genes tested) then this suggests that this process may be involved in the pathogenesis of the condition (this example is equivalent to a fold enrichment of 10).

Enrichment has a stronger performance when lists of a reasonable size (e.g. ≥100) are used (Huang *et al*. 2009). Enrichment can be meaningfully applied to gene lists with high FDR rates as those genes which are in the list by random chance are unlikely to cluster in the same biological pathway, and are therefore likely to be removed by the process of enrichment. For the following biological analyses a gene list of 100 genes was used, provided the FDR rate was not above 0.5 (Table 22). Biological interpretation of bronchiectasis as a complication was not performed as there were no differentially expressed genes even at an FDR of 0.5. The biological analysis of genes differentially expressed in splenomegaly did not yield any statistically significant results. The p values in the results tables below (Table 23 to Table 28) represent the probability of the degree of fold enrichment happening by chance using the modified Fisher's exact test. This test is more stringent than Fisher's exact test and is also

known as the Expression Analysis Systematic Explorer (EASE) score (Hosack *et al.* 2003). Further information on the differentially expressed genes can be found in Appendix 6.

Complication	Length of list	Maximum FDR
Granuloma	100	0.13
Autoimmunity	100	0.36
Splenomegaly	100	0.09
Lymphadenopathy	79	0.5
G, A, S and/or L	100	0.46
Polyclonal Proliferation	100	0.32
Autoimmune cytopenia	100	0.072
OSAI	55	0.5

Table 22: Gene lists used for biological analysis. G, A, S and/or L: Any combination of granuloma, autoimmunity, splenomegaly and/or lymphadenopathy.

Granulomatous disease			
Gene Ontology (GO) Term	**p value**	**Fold Enrichment**	**Genes**
GO:0009314~response to radiation	0.0046	5.439607	SLC1A2, XRCC2, NR2F6, MBD4, MYC, C9ORF80
GO:0009628~response to abiotic stimulus	0.0147	3.452059	SLC1A2, XRCC2, NR2F6, MBD4, MYC, NFKBIL1, C9ORF80
GO:0043171~peptide catabolic process	0.0272	71.80282	ECE1, TPP1
GO:0008629~induction of apoptosis by intracellular signals	0.0354	9.972613	HIPK2, MBD4, MYC
GO:0007167~enzyme linked receptor protein signalling pathway	0.0368	3.215052	EPHB6, CREB1, HIPK2, TIPARP, SPTBN1, FGF1
GO:0042981~regulation of apoptosis	0.0674	2.042432	XRCC2, CREB1, ADAMTSL4, HIPK2, TAF9, MBD4, TNFAIP3, MYC, NFKBIL1
GO:0043067~regulation of programmed cell death	0.0706	2.021982	XRCC2, CREB1, ADAMTSL4, HIPK2, TAF9, MBD4, TNFAIP3, MYC, NFKBIL1
GO:0010941~regulation of cell death	0.0718	2.014418	XRCC2, CREB1, ADAMTSL4, HIPK2, TAF9, MBD4, TNFAIP3, MYC, NFKBIL1
GO:0008284~positive regulation of cell proliferation	0.0738	2.633355	HIPK2, MFGE8, VIPR1, IL21, FGF1, MYC
GO:0043523~regulation of neuron apoptosis	0.0859	6.050799	XRCC2, HIPK2, NFKBIL1
GO:0007166~cell surface receptor linked signal transduction	0.0866	1.553725	IRAK2, OPRK1, CREB1, GNA11, TIPARP, GABBR1, ITPKB, VIPR1, GPR1, OR1D4, OR13C5, EPHB6, HIPK2, SPTBN1, FGF1
GO:0008219~cell death	0.0871	2.060339	ECE1, TPP1, ADAMTSL4, HIPK2, KLF11, TNFAIP3, MYC, NFKBIL1
GO:0070555~response to interleukin-1	0.0894	21.11848	IRAK2, TAF9
GO:0016265~death	0.0897	2.045664	ECE1, TPP1, ADAMTSL4, HIPK2, KLF11, TNFAIP3, MYC, NFKBIL1
GO:0006915~apoptosis	0.0990	2.162735	ECE1, ADAMTSL4, HIPK2, KLF11, TNFAIP3, MYC, NFKBIL1

Table 23: Enriched functional annotation terms associated with granulomatous disease. p value indicates result from modified Fisher's exact test. Fold enrichment indicates the degree by which the biological annotation is overrepresented within the gene list.

Autoimmunity			
Gene Ontology Term	**p value**	**Fold Enrichment**	**Genes**
GO:0006260~DNA replication	0.0136	5.334422	RFC4, SUPT16H, PCNA, POLA2, MCM3
GO:0007049~cell cycle	0.0138	2.561088	ANAPC1, NOLC1, DUSP1, RGS2, SEH1L, GSK3B, CCNF, FBXO31, MCM3, CDC25B
GO:0042981~regulation of apoptosis	0.0184	2.44129	PLEKHF1, ARHGEF3, TNFRSF10B, DUSP1, GSK3B, DDX19B, APAF1, DAPK2, PRDX1, ADA
GO:0043067~regulation of programmed cell death	0.0195	2.416847	PLEKHF1, ARHGEF3, TNFRSF10B, DUSP1, GSK3B, DDX19B, APAF1, DAPK2, PRDX1, ADA
GO:0010941~regulation of cell death	0.0199	2.407806	PLEKHF1, ARHGEF3, TNFRSF10B, DUSP1, GSK3B, DDX19B, APAF1, DAPK2, PRDX1, ADA
GO:0000320~re-entry into mitotic cell cycle	0.0202	96.55303	GSK3B, CCNF
GO:0043065~positive regulation of apoptosis	0.0203	3.203181	PLEKHF1, ARHGEF3, TNFRSF10B, DUSP1, DDX19B, DAPK2, PRDX1
GO:0033554~cellular response to stress	0.0209	2.82422	RFC4, GSK3B, SUPT16H, PCNA, FBXO31, GML, PRDX1, EIF2B1
GO:0043068~positive regulation of programmed cell death	0.0209	3.18057	PLEKHF1, ARHGEF3, TNFRSF10B, DUSP1, DDX19B, DAPK2, PRDX1
GO:0043039~tRNA aminoacylation	0.0211	13.16632	LARS2, AARS2, EARS2
GO:0043038~amino acid activation	0.0211	13.16632	LARS2, AARS2, EARS2
GO:0006418~tRNA aminoacylation for protein translation	0.0211	13.16632	LARS2, AARS2, EARS2
GO:0010942~positive regulation of cell death	0.0213	3.165673	PLEKHF1, ARHGEF3, TNFRSF10B, DUSP1, DDX19B, DAPK2, PRDX1
GO:0006917~induction of apoptosis	0.0217	3.689925	PLEKHF1, ARHGEF3, TNFRSF10B, DDX19B, DAPK2, PRDX1
GO:0012502~induction of programmed cell death	0.0221	3.678211	PLEKHF1, ARHGEF3, TNFRSF10B, DDX19B, DAPK2, PRDX1
GO:0051336~regulation of hydrolase activity	0.0237	3.609459	C13ORF18, WDR67, TNFRSF10B, S1PR4, ABHD5, APAF1
GO:0000280~nuclear division	0.0243	4.470048	ANAPC1, NOLC1, SEH1L, CCNF, CDC25B
GO:0007067~mitosis	0.0243	4.470048	ANAPC1, NOLC1, SEH1L, CCNF, CDC25B
GO:0000087~M phase of mitotic cell cycle	0.0250	4.429038	ANAPC1, NOLC1, SEH1L, CCNF, CDC25B

Autoimmunity (cont)			
Gene Ontology Term	**p value**	**Fold Enrichment**	**Genes**
GO:0006999~nuclear pore organization	0.0252	77.24242	SEH1L, AHCTF1
GO:0006997~nucleus organization	0.0268	11.58636	SEH1L, AHCTF1, APAF1
O:0048285~organelle fission	0.0277	4.291246	ANAPC1, NOLC1, SEH1L, CCNF, CDC25B
GO:0042542~response to hydrogen peroxide	0.0319	10.53306	DUSP1, PRDX1, ADA
GO:0007242~intracellular signaling cascade	0.0457	1.882431	ARHGEF3, RFC4, TNFRSF10B, DUSP1, GSK3B, S1PR4, REM2, PCNA, PRKCH, FBXO31, DAPK2, GML
GO:0046700~heterocycle catabolic process	0.0534	7.935866	ATP5B, HAL, ADA
GO:0000302~response to reactive oxygen species	0.0547	7.828624	DUSP1, PRDX1, ADA
GO:0031571~G1 DNA damage checkpoint	0.0547	35.11019	FBXO31, GML
GO:0051345~positive regulation of hydrolase activity	0.0599	4.413853	TNFRSF10B, S1PR4, ABHD5, APAF1
GO:0051301~cell division	0.0604	3.329415	ANAPC1, SEH1L, CCNF, AHCTF1, CDC25B
GO:0006915~apoptosis	0.0750	2.326579	PLEKHF1, ARHGEF3, TNFRSF10B, APAF1, DAPK2, GML, PDCD1
GO:0006297~nucleotide-excision repair, DNA gap filling	0.0786	24.13826	RFC4, PCNA
GO:0012501~programmed cell death	0.0795	2.291089	PLEKHF1, ARHGEF3, TNFRSF10B, APAF1, DAPK2, GML, PDCD1
GO:0000279~M phase	0.0803	3.017282	ANAPC1, NOLC1, SEH1L, CCNF, CDC25B
GO:0010035~response to inorganic substance	0.0830	3.842907	DUSP1, DDX19B, PRDX1, ADA
GO:0031575~G1/S transition checkpoint	0.0833	22.71836	FBXO31, GML
GO:0046638~positive regulation of alpha-beta T cell differentiation	0.0973	19.31061	IL4R, ADA
GO:0006259~DNA metabolic process	0.0982	2.408807	RFC4, SUPT16H, PCNA, APAF1, POLA2, MCM3

Table 24: Enriched functional annotation terms associated with autoimmune disease. p value indicates result from modified Fisher's exact test. Fold enrichment indicates the degree by which the biological annotation is overrepresented within the gene list.

Lymphadenopathy			
Gene Ontology Term	**p value**	**Fold Enrichment**	**Genes**
GO:0006812~cation transport	0.0335	2.836996	STEAP3, CATSPER4, ATP13A1, KCND1, SLC24A4, SLC12A5, ATP7B
GO:0030001~metal ion transport	0.0538	2.87384	STEAP3, CATSPER4, KCND1, SLC24A4, SLC12A5, ATP7B
GO:0030641~regulation of cellular pH	0.0704	27.00212	MAFG, SLC12A5
GO:0030004~cellular monovalent inorganic cation homeostasis	0.0997	18.78408	MAFG, SLC12A5

Table 25: Enriched functional annotation terms associated with lymphadenopathy. p value indicates result from modified Fisher's exact test. Fold enrichment indicates the degree by which the biological annotation is overrepresented within the gene list.

Granuloma, Autoimmunity, Splenomegaly, and/or Lymphadenopathy			
Gene Ontology Term	**p value**	**Fold Enrichment**	**Genes**
GO:0008283~cell proliferation	0.0150	3.417152	NDE1, CD274, CD81, UCHL1, PCNA, TNFSF14, TNFRSF4
GO:0006917~induction of apoptosis	0.0159	3.992378	PLEKHF1, ARHGEF3, CFLAR, DDX19B, RIPK3, TNFSF14
GO:0012502~induction of programmed cell death	0.0161	3.979703	PLEKHF1, ARHGEF3, CFLAR, DDX19B, RIPK3, TNFSF14
GO:0008588~release of cytoplasmic sequestered NF-kappaB	0.0187	104.4672	NLRP12, TNFSF14
GO:0006289~nucleotide-excision repair	0.0258	11.82648	ATXN3, RFC4, PCNA
GO:0051249~regulation of lymphocyte activation	0.0308	5.763708	IL4R, CD274, TNFSF14, TNFRSF4
GO:0042981~regulation of apoptosis	0.0311	2.377256	PLEKHF1, ARHGEF3, CFLAR, XRCC2, DDX19B, RIPK3, NLRP12, TNFSF14, TNFRSF4
GO:0043067~regulation of programmed cell death	0.0327	2.353454	PLEKHF1, ARHGEF3, CFLAR, XRCC2, DDX19B, RIPK3, NLRP12, TNFSF14, TNFRSF4
GO:0007249~I-kappaB kinase/NF-kappaB cascade	0.0334	10.27546	RIPK3, NLRP12, TNFSF14
GO:0010941~regulation of cell death	0.0334	2.344651	PLEKHF1, ARHGEF3, CFLAR, XRCC2, DDX19B, RIPK3, NLRP12, TNFSF14, TNFRSF4
GO:0051222~positive regulation of protein transport	0.0396	9.355273	NLRP12, TNFSF14, TNFRSF4
GO:0002694~regulation of leukocyte activation	0.0407	5.158875	IL4R, CD274, TNFSF14, TNFRSF4
GO:0042346~positive regulation of NF-kappaB import into nucleus	0.0416	46.42987	NLRP12, TNFSF14
GO:0050710~negative regulation of cytokine secretion	0.0461	41.78689	NLRP12, TNFRSF4
GO:0050865~regulation of cell activation	0.0466	4.887355	IL4R, CD274, TNFSF14, TNFRSF4
GO:0043065~positive regulation of apoptosis	0.0480	2.970632	PLEKHF1, ARHGEF3, CFLAR, DDX19B, RIPK3, TNFSF14
GO:0043068~positive regulation of programmed cell death	0.0492	2.949662	PLEKHF1, ARHGEF3, CFLAR, DDX19B, RIPK3, TNFSF14
GO:0010942~positive regulation of cell death	0.0501	2.935847	PLEKHF1, ARHGEF3, CFLAR, DDX19B, RIPK3, TNFSF14
GO:0042993~positive regulation of transcription factor import into nucleus	0.0595	32.14376	NLRP12, TNFSF14
GO:0050000~chromosome localization	0.0684	27.85792	NDE1, CDCA5
GO:0051303~establishment of chromosome localization	0.0684	27.85792	NDE1, CDCA5

Granuloma, Autoimmunity, Splenomegaly, and/or Lymphadenopathy (cont)			
Gene Ontology Term	**p value**	**Fold Enrichment**	**Genes**
GO:0006297~nucleotide-excision repair, DNA gap filling	0.0728	26.1168	RFC4, PCNA
GO:0050709~negative regulation of protein secretion	0.0728	26.1168	NLRP12, TNFRSF4
GO:0006955~immune response	0.0736	2.332601	FCAR, S1PR4, IL4R, CD274, TNFSF14, HLA-DPB1, TNFRSF4
GO:0051251~positive regulation of lymphocyte activation	0.0750	6.529201	IL4R, TNFSF14, TNFRSF4
GO:0042345~regulation of NF-kappaB import into nucleus	0.0771	24.58052	NLRP12, TNFSF14
GO:0042307~positive regulation of protein import into nucleus	0.0771	24.58052	NLRP12, TNFSF14
GO:0002696~positive regulation of leukocyte activation	0.0861	6.026955	IL4R, TNFSF14, TNFRSF4
GO:0043122~regulation of I-kappaB kinase/NF-kappaB cascade	0.0875	5.969555	TRIM38, CFLAR, NLRP12
GO:0007243~protein kinase cascade	0.0905	2.885835	CD81, RIPK3, NLRP12, TNFSF14, MARK2
GO:0019932~second-messenger-mediated signalling	0.0931	3.64951	SSTR2, RFC4, S1PR4, PCNA
GO:0050867~positive regulation of cell activation	0.0932	5.750489	IL4R, TNFSF14, TNFRSF4
GO:0008624~induction of apoptosis by extracellular signals	0.0947	5.698212	ARHGEF3, CFLAR, RIPK3
GO:0046824~positive regulation of nucleocytoplasmic transport	0.0987	18.99404	NLRP12, TNFSF14
GO:0051223~regulation of protein transport	0.0990	5.546932	NLRP12, TNFSF14, TNFRSF4

Table 26: Enriched functional annotation terms associated with the development of any of granuloma, autoimmunity, splenomegaly, and/or lymphadenopathy (GASL). p value indicates result from modified Fisher's exact test. Fold enrichment indicates the degree by which the biological annotation is overrepresented within the gene list.

Polyclonal Proliferation			
Gene Ontology Term	**p value**	**Fold Enrichment**	**Genes**
GO:0008380~RNA splicing	0.014426132	4.10555138	SNRPN, MAGOH, USP39, LOC728643, LSM2, TTF2
GO:0070498~interleukin-1-mediated signaling pathway	0.02055606	95.1119403	IRAK2, IRAK3
GO:0002755~MyD88-dependent toll-like receptor signaling pathway	0.030677462	63.4079602	IRAK2, IRAK3
GO:0000375~RNA splicing, via transesterification reactions	0.043404233	5.039043195	MAGOH, USP39, LOC728643, LSM2
GO:0000398~nuclear mRNA splicing, via spliceosome	0.043404233	5.039043195	MAGOH, USP39, LOC728643, LSM2
GO:0000377~RNA splicing, via transesterification reactions with bulged adenosine as nucleophile	0.043404233	5.039043195	MAGOH, USP39, LOC728643, LSM2
GO:0019221~cytokine-mediated signaling pathway	0.048349105	8.392230026	IRAK2, IRAK3, JAK3
GO:0007229~integrin-mediated signaling pathway	0.049629758	8.270603504	PLEK, PRAM1, ITGAM
GO:0006396~RNA processing	0.057166539	2.498249839	SNRPN, MAGOH, USP39, LOC728643, LSM2, DIMT1L, TTF2
GO:0002224~toll-like receptor signaling pathway	0.065297807	29.2652124	IRAK2, IRAK3
GO:0030534~adult behavior	0.068777676	6.875561949	MAFG, CHD7, UCHL1
GO:0032088~negative regulation of NF-kappaB transcription factor activity	0.074963666	25.36318408	IRAK2, IRAK3
GO:0006397~mRNA processing	0.079508347	3.029042685	MAGOH, USP39, LOC728643, LSM2, TTF2
GO:0002221~pattern recognition receptor signaling pathway	0.079759612	23.77798507	IRAK2, IRAK3
GO:0030641~regulation of cellular pH	0.079759612	23.77798507	MAFG, SLC12A5
GO:0001959~regulation of cytokine-mediated signaling pathway	0.079759612	23.77798507	IRAK2, IRAK3
GO:0070555~response to interleukin-1	0.084531069	22.37928007	IRAK2, IRAK3
GO:0009057~macromolecule catabolic process	0.089103002	2.045418071	PSMB1, PGLYRP2, MAGOH, USP39, UCHL1, KLHL12, CUL4B, DNASE1L1
GO:0002758~innate immune response-activating signal transduction	0.089278158	21.13598673	IRAK2, IRAK3
GO:0002218~activation of innate immune response	0.089278158	21.13598673	IRAK2, IRAK3
GO:0044092~negative regulation of molecular function	0.091445612	2.882180009	IRAK2, IRAK3, PLEK, PSMB1, DUSP6

Table 27: Enriched functional annotation terms associated with polyclonal proliferation. p value indicates result from modified Fisher's exact test. Fold enrichment indicates the degree by which the biological annotation is overrepresented within the gene list.

Autoimmune Cytopenia			
Gene Ontology Term	**p value**	**Fold Enrichment**	**Genes**
GO:0002263~cell activation during immune response	0.001079801	19.39878234	LAT2, EOMES, BCL3, TLR4
GO:0002366~leukocyte activation during immune response	0.001079801	19.39878234	LAT2, EOMES, BCL3, TLR4
GO:0015721~bile acid and bile salt transport	0.001375399	52.37671233	MIP, AKR1C4, SLC10A2
GO:0002757~immune response-activating signal transduction	0.002964477	13.69325813	LAT2, LAX1, SH2B2, TLR4
GO:0002764~immune response-regulating signal transduction	0.003674943	12.69738481	LAT2, LAX1, SH2B2, TLR4
GO:0050778~positive regulation of immune response	0.008059288	6.235322896	LAT2, LAX1, SH2B2, TLR4, LAG3
GO:0048584~positive regulation of response to stimulus	0.009793983	4.534780288	LAT2, LAX1, SH2B2, TLR4, NPY5R, LAG3
GO:0002684~positive regulation of immune system process	0.009966599	4.515233821	LAT2, LAX1, IL4R, SH2B2, TLR4, LAG3
GO:0045321~leukocyte activation	0.01123403	4.382988479	LAT2, LAX1, TOLLIP, EOMES, BCL3, TLR4
GO:0042990~regulation of transcription factor import into nucleus	0.011579159	18.06093529	BCL3, TACC3, PRDX1
GO:0009593~detection of chemical stimulus	0.01399288	16.3677226	TLR4, PLCB2, ABCG1
GO:0002253~activation of immune response	0.014302657	7.759512938	LAT2, LAX1, SH2B2, TLR4
GO:0050851~antigen receptor-mediated signaling pathway	0.014841763	15.87173101	LAT2, LAX1, SH2B2
GO:0002429~immune response-activating cell surface receptor signaling pathway	0.020381705	13.42992624	LAT2, LAX1, SH2B2
GO:0051262~protein tetramerization	0.020381705	13.42992624	PCBD1, CDA, TGM3
GO:0001775~cell activation	0.020911757	3.741193738	LAT2, LAX1, TOLLIP, EOMES, BCL3, TLR4
GO:0050868~negative regulation of T cell activation	0.022392284	12.77480789	LAX1, IL4R, LAG3
GO:0002768~immune response-regulating cell surface receptor signaling pathway	0.023427168	12.47064579	LAT2, LAX1, SH2B2
GO:0042306~regulation of protein import into nucleus	0.024481453	12.18063077	BCL3, TACC3, PRDX1
GO:0043038~amino acid activation	0.02555489	11.90379826	YARS, EPRS, KARS

Autoimmune Cytopenia (cont)			
Gene Ontology Term	**p value**	**Fold Enrichment**	**Genes**
GO:0006418~tRNA aminoacylation for protein translation	0.02555489	11.90379826	YARS, EPRS, KARS
GO:0043039~tRNA aminoacylation	0.02555489	11.90379826	YARS, EPRS, KARS
GO:0002292~T cell differentiation during immune response	0.02793336	69.83561644	EOMES, BCL3
GO:0002287~alpha-beta T cell activation during immune response	0.02793336	69.83561644	EOMES, BCL3
GO:0002293~alpha-beta T cell differentiation during immune response	0.02793336	69.83561644	EOMES, BCL3
GO:0051016~barbed-end actin filament capping	0.02793336	69.83561644	TRIOBP, ADD2
GO:0015718~monocarboxylic acid transport	0.031200928	10.68912497	MIP, AKR1C4, SLC10A2
GO:0033157~regulation of intracellular protein transport	0.034803119	10.07244468	BCL3, TACC3, PRDX1
GO:0051250~negative regulation of lymphocyte activation	0.034803119	10.07244468	LAX1, IL4R, LAG3
GO:0002252~immune effector process	0.037879429	5.330963087	LAT2, LAX1, BCL3, PRDX1
GO:0002695~negative regulation of leukocyte activation	0.038558633	9.523038605	LAX1, IL4R, LAG3
GO:0046822~regulation of nucleocytoplasmic transport	0.0411445	9.1888969	BCL3, TACC3, PRDX1
GO:0030097~hemopoiesis	0.043235813	3.730535066	EPAS1, EOMES, BCL3, TACC3, ADD2
GO:0050866~negative regulation of cell activation	0.043794075	8.877408869	LAX1, IL4R, LAG3
GO:0045414~regulation of interleukin-8 biosynthetic process	0.049724827	38.79756469	BCL3, TLR4
GO:0070302~regulation of stress-activated protein kinase signaling pathway	0.052107995	8.057955743	TLR4, PRDX1, CBS
GO:0051098~regulation of binding	0.052920613	4.655707763	PCBD1, BCL3, TLR4, ADD2
GO:0042088~T-helper 1 type immune response	0.055096986	34.91780822	BCL3, TLR4
GO:0032872~regulation of stress-activated MAPK cascade	0.055096986	34.91780822	TLR4, PRDX1
GO:0042534~regulation of tumor necrosis factor biosynthetic process	0.055096986	34.91780822	BCL3, TLR4

Autoimmune Cytopenia (cont)			
Gene Ontology Term	**p value**	**Fold Enrichment**	**Genes**
GO:0048534~hemopoietic or lymphoid organ development	0.05800018	3.383508548	EPAS1, EOMES, BCL3, TACC3, ADD2
GO:0032386~regulation of intracellular transport	0.059430131	7.482387476	BCL3, TACC3, PRDX1
GO:0031349~positive regulation of defense response	0.060935126	7.377001736	TLR4, NPY5R, LAG3
GO:0042035~regulation of cytokine biosynthetic process	0.062453189	7.274543379	BCL3, TLR4, LAG3
GO:0006955~immune response	0.063119308	2.22761137	LAT2, LAX1, TOLLIP, IL4R, EOMES, BCL3, TLR4, PRDX1
GO:0042113~B cell activation	0.068652339	6.891672675	LAT2, LAX1, BCL3
GO:0002520~immune system development	0.069172565	3.185931407	EPAS1, EOMES, BCL3, TACC3, ADD2
GO:0002286~T cell activation during immune response	0.071034415	26.85985248	EOMES, BCL3
GO:0050865~regulation of cell activation	0.072389112	4.083954178	LAX1, IL4R, TLR4, LAG3
GO:0046632~alpha-beta T cell differentiation	0.076287754	24.94129159	EOMES, BCL3
GO:0002683~negative regulation of immune system process	0.076672907	6.466260781	LAX1, IL4R, LAG3
GO:0032677~regulation of interleukin-8 production	0.081511797	23.27853881	BCL3, TLR4
GO:0002443~leukocyte mediated immunity	0.083292671	6.161966156	LAT2, BCL3, PRDX1
GO:0007242~intracellular signaling cascade	0.083931552	1.70192404	ADCY4, LAT2, DIRAS2, LAX1, TOLLIP, CD81, BCL3, FBXO31, SH2B2, TLR4, PLCB2, RGL2
GO:0042345~regulation of NF-kappaB import into nucleus	0.091872632	20.53988719	BCL3, PRDX1
GO:0002285~lymphocyte activation during immune response	0.097009745	19.39878234	EOMES, BCL3
GO:0002275~myeloid cell activation during immune response	0.097009745	19.39878234	LAT2, TLR4
GO:0046631~alpha-beta T cell activation	0.097009745	19.39878234	EOMES, BCL3
GO:0046649~lymphocyte activation	0.099117753	3.563041655	LAT2, LAX1, EOMES, BCL3

Table 28: Enriched functional annotation terms associated with autoimmune cytopenia. p value indicates result from modified Fisher's exact test. Fold enrichment indicates the degree by which the biological annotation is overrepresented within the gene list.

The results of biological analysis of the data shows enrichment particularly of processes related to apoptosis and cell cycle in patients with granuloma or autoimmunity alone, and those with any combination of the four complications. With the analysis of the persistent lymphadenopathy gene list, no gene ontology terms which are biologically relevant to the development of lymphadenopathy were identified. This suggests that the terms identified in Table 25 are the result of chance clustering of genes rather than these processes being involved in the pathogenesis of lymphadenopathy. There was a similar situation for the OSAI gene list (data not shown). The OSAI and persistent lymphadenopathy gene lists were the weakest lists in statistical terms and so it is perhaps unsurprising that these lists failed to identify any processes of interest. These results illustrate the importance of ensuring that gene lists produced by high-throughput technologies make biological sense, and that their final interpretation should take into account both statistical significance and biological context.

The enrichment of terms related to apoptosis and control of proliferation is of potential biological relevance. Inherited defects of apoptosis are known to lead to autoimmune lymphoproliferative syndrome (ALPS) which shares some of the features of the complications of CVID (Fisher *et al.* 1995; Rensing-Ehl *et al.* 2010). Within the granuloma and polyclonal proliferation gene lists, genes involved in the response to IL-1 were also identified as being statistically significant which is consistent with the known role of IL-1 in inflammatory disease (Dinarello 2011). Examining the data from patients with at least one complication, genes involved in the NFκB pathway were identified and this also could be a finding of potentially interesting biological

relevance. The possible implications of these findings and potential avenues for further research will be examined further in the Discussion.

Discussion

Conclusions and indications for continued research

The cohort of CVID patients recruited for this study show similar clinical features to other cohorts reported in the literature (Table 29). Therefore, the cohort studied is likely to be a representative group of CVID patients. Statistically significant associations of the presence of granuloma with splenomegaly and lymphadenopathy, and splenomegaly with autoimmunity were found as summarised in Figure 6b. The association of different CVID complications with each other was highlighted in a multicentre study of 334 CVID patients, which showed lymphadenopathy associated with granuloma and hepatomegaly, and splenomegaly associated with granuloma, hepatomegaly, and lymphadenopathy (Chapel *et al*. 2008). The association of granulomatous disease, lymphadenopathy, splenomegaly and autoimmunity with each other suggests that there may be common pathogenetic mechanisms in the development of these complications. Accordingly, research into the pathogenesis of CVID should consider CVID subgroups as separate entities. In forming the clinical subgroups, complications arising from immune dysregulation are more likely to be important in separating out underlying pathogenetic mechanisms. Complications which result from infections (such as bronchiectasis) are influenced by many factors including those not related to the underlying disease such as length of delay in diagnosis and commencement of immunoglobulin replacement therapy and whether each infectious episode was promptly and adequately treated with antibiotics. Though in total there may be greater than two CVID subtypes, a first step to disentangling the heterogeneity of this condition would be to consider patients with and without

evidence of immune dysregulation as separate clinical phenotypes. Of course there may be further subgroups to be eventually delineated; for example splenomegaly often results from ITP (the commonest autoimmune complication in CVID) and so patients with autoimmunity and splenomegaly may represent a separate group from those with granulomatous disease and/or persistent lymphadenopathy.

Complication	Frequency in cohort (%)	Frequency in literature (%)
Granuloma	11.3	8-22
Autoimmunity	26.4	22
Splenomegaly	26.4	30
Lymphadenopathy	13.2	15

Table 29: Frequency of different complications in CVID patients within this cohort compared to cohorts reported in the literature (Ardeniz and Cunningham-Rundles 2009; Chapel *et al.* 2008; Cunningham-Rundles 2008; Cunningham-Rundles and Bodian 1999; Morimoto and Routes 2005)

The use of peripheral B cell phenotyping by the Freiburg, Paris and EUROclass methods gave similar results to published works (Piqueras *et al*. 2003; Warnatz *et al*. 2002; Wehr *et al*. 2008), and while they are able to risk stratify patients to a degree, these classification methods lack the power to accurately define the complications affecting an individual patient. Deficiency of class-switched memory B cells may represent an effect of the disease process, rather than being the fundamental cause of immune dysregulation which is more likely to be mediated by other immune cells.

The Rome classification utilises levels of naïve CD4 T cells to group different CVID patients. While it has been noted that decreased numbers of naïve CD4 T cells correspond with abnormalities of both CD4 and CD8 T cell repertoires (Giovannetti *et al*. 2007), there are no published investigations of naïve CD8 T cells in CVID. There was a statistically significant increase in the rate of granulomatous disease in patients with a low percentage of naïve $CD8^+$ T cells, and there also appeared to be higher

rates of autoimmunity, splenomegaly and lymphadenopathy. A previous study of 38 CVID patients identified that a low CD4:CD8 ratio of ≤0.9 was associated with an increased proportion of $CD8^{+}CD57^{+}$ T cells, and an increased proportion of $CD8^{+}$HLA-DR^{+} T cells. Although statistical significance was not reached, there was the suggestion that the patient group with low CD4:CD8 ratio had a higher rate of granulomas (Wright *et al*. 1990). A more recent investigation of $CD8^{+}$ phenotypes in 34 HLA-$A2^{+}$ CVID patients demonstrated that there were increased numbers of late differentiation stage $CD8^{+}$ T cells ($CCR7^{-}CD127^{-}CD57^{+}$PD-$1^{+}CD8^{+}$) in patients with granulomatous disease and patients with lymphadenopathy. While CMV infection in healthy individuals increases the proportion of highly differentiated $CD8^{+}$ T cells, the association between increased $CCR7^{-}CD127^{-}CD57^{+}$PD-$1^{+}CD8^{+}$ T cells and complications in CVID was independent of CMV status (Kuntz *et al.* 2011). Within the literature no infectious cause of granulomatous disease in CVID which has been reproduced across cohorts has been identified. The increased numbers of late differentiation stage $CD8^{+}$ T cells in granulomatous disease is consistent with the findings within this project as an increase in this population would be expected to lead to a decrease in naïve $CD8^{+}$ T cells as a proportion of total $CD8^{+}$ T cells. These findings suggest that further investigation of naïve $CD8^{+}$ T cells is warranted. In additional to the recognised role of B cells and $CD4^{+}$ T cells, $CD8^{+}$ T cells may also play a role in the development and/or phenotype of CVID.

A group of CVID patients have low numbers of B cells, and while this group has been excluded from previous classifications, it appears that they may represent an interesting group in their own right. Eight patients were identified in whom B cell numbers appeared to be declining over time. This suggests that in at least some CVID

patients with low B cell numbers, the B cell lymphopenia occurs subsequent to the development of antibody deficiency. The reason for the drop in B cell numbers is not clear, but possibilities include a decrease in bone marrow production or a loss of peripheral survival signals/factors. Another possibility is that an autoimmune phenomenon is occurring. This would appear to be possible, given that autoantibody mediated autoimmune cytopenias such as AHA and ITP are common in CVID. Low numbers of B cells are also common, affecting 10 of the cohort of 54 patients (18.5%). There are reports in the literature of production of anti-lymphocyte autoantibodies in systemic lupus erythematosus (SLE), some of which are specific for particular lymphocyte populations (Winfield and Mimura 1992). Investigation for the presence of anti-B cell autoantibodies in CVID would be of interest and comparison could be made between CVID patients with normal numbers of B cells, those with low but stable B cell numbers and those where there is a documented drop in B cell numbers. Some but not all CVID patients with low B cell numbers clustered with XLA patients as shown in Figure 20 and Figure 21. This group of CVID patients appeared to be clinically similar to XLA patients, and had lower rates of granuloma, autoimmunity, splenomegaly and lymphadenopathy compared to other CVID patients.

The investigation of differentially expressed genes between CVID patients with different complications showed that some degree of separation of the subgroups within the cohort was possible. Expanding the study across multiple centres may allow the identification of a set of genes, the expression of which could be used to reliably differentiate between CVID patients. While investigation of gene expression of sorted B and T cells may provide a targeted look at specific pathways within these cell populations, the methodology of such an approach would be difficult to translate

into a routine diagnostic assay, and moreover would potentially miss relevant information. A whole blood approach has been shown to be useful in the investigation of autoimmune diseases (Bauer *et al.* 2009; Pascual *et al.* 2010), and in addition to disease-specific transcriptional differences within cell type (s) can also detect differences due to the presence of unusual cell types and due to altered frequencies of normal cell types in the blood.

Differential gene expression between the CVID subgroups demonstrated a number of interesting findings which suggest possible pathogenetic mechanisms and promising avenues for further research. Other authors have taken the approach of investigating polymorphisms of genes of interest in CVID patients. An investigation of NOD2 polymorphisms in 285 CVID patients identified a polymorphism which was statistically significantly overrepresented in patients with autoimmunity or enteropathy. There was also a possible increase of this polymorphism in CVID patients with granuloma, though this did not reach statistical significance. Another NOD2 polymorphism was found to potentially protect from splenomegaly (Packwood *et al.* 2010). A study of immunoregulatory gene polymorphisms in 163 CVID patients found that IL-1α -889 c/c homozygous polymorphisms were overrepresented in the CVID patients, though did not identify a particular clinical CVID subgroup. Specific TNFα and IL-10 polymorphisms were associated with granulomatous disease (Mullighan *et al.* 1999). This demonstrates the value of investigating a disease using a number of different approaches. While differences in mRNA expression levels of IL-10 (p=0.003, FDR=0.252), TNFα (p=0.176, FDR=0.722) and NOD2 (p=0.964, FDR=0.994) in granulomatous disease were not demonstrated when multiple testing was corrected for, this does not exclude the possibility that functional differences due

to polymorphisms may play a role. Similarly, that the examined IL-1 α polymorphism was associated with CVID as a whole, rather than with the granulomatous disease subgroup, does not rule out a role of the IL-1 pathway in granuloma formation. A role for IL-1 is suggested by the finding in this project that there was statistically significant differential expression of genes related to the response to IL-1 in both the granuloma and polyclonal proliferation groups. There is longstanding evidence for involvement of IL-1 in granuloma formation (Kasahara *et al.* 1988), and furthermore endogenous IL-1 receptor antagonist (IL1-RA) can help restrict excess granuloma formation (Iizasa *et al.* 2005). Excessive IL-1 production and activity is associated with the autoinflammatory diseases, most of which respond to treatment with anakinra (recombinant IL1-RA). Blau syndrome is an autosomal dominantly inherited condition which arises due to mutations in nucleotide-binding oligomerisation domain containing 2 (NOD2), also known as caspase recruitment domain family member 15 (CARD15) (Henckaerts and Vermeire 2007). Blau syndrome is characterised by granulomatous inflammation, polyarthritis, uveitis, and dermatitis (Borzutzky *et al.* 2010), and responds to anakinra therapy (Dinarello 2011). Current treatments for granulomatous disease associated with CVID include steroids and other immune suppressants (such as hydroxychloroquine, cyclosporine, azathioprine and methotrexate), and anti-TNFα therapy has also been effective in some cases (Ardeniz and Cunningham-Rundles 2009). Further research on targeting the IL-1 pathway as a possible therapy for the treatment of CVID granulomatous disease could be considered, especially in the setting where other therapeutic interventions have failed to provide significant benefit.

Examination of the list of differentially expressed genes between CVID patients with any of granuloma, autoimmunity, splenomegaly and/or lymphadenopathy and CVID patients without these complications identified enrichment of genes involved in the nuclear factor kappa B (NFκB) signalling pathway. NFκB is a transcription factor responsible for mediating the response to a number of different stimuli (Li and Verma 2002). The signalling pathways activated by pro-inflammatory cytokines including IL-1 and TNFα, lipopolysaccharide (LPS) binding to toll like-receptor (TLR) 4 and CD14, and TCR ligation all converge and lead to NFκB activation. The variable phenotype of defective NFκB signalling due to NEMO deficiency, where there may be increased susceptibility to pyogenic bacteria, viruses and/or nonpathogenic mycobacterial infections (Uzel 2005), demonstrates the involvement of NFκB in the response to a number of different signals. Of particular interest is that TACI signalling is also mediated by NFκB (Gross *et al.* 2001), and that polymorphisms in TACI increase the risk of autoimmune disease and splenomegaly in CVID patients (Salzer *et al*. 2005; Waldrep *et al*. 2009). However, the prevalence of TACI polymorphisms in the general population emphasises the complexity and multifactorial nature of this condition, and suggests that the investigation of polymorphisms in other genes involved in pathways affecting NFκB activation would be of interest. Another notable function of NFκB is that it protects lymphocytes against TNFα induced apoptosis (Beg and Baltimore 1996; Jeremias *et al.* 1998).

The differential expression of genes involved in apoptosis and control of cell cycle was a consistent finding in the granuloma, autoimmunity and any complication subgroups. Defects in apoptosis lead to ALPS, and a number of gene mutations have been identified including FAS, FASL and caspases 8 and 10. As a result of the

apoptotic defect, patients suffer with lymphoproliferative disease, increased susceptibility to haematological malignancy and autoimmune disease, especially autoimmune cytopenia (s) (Fisher *et al*. 1995; Oliveira *et al.* 2010). An increase in $\alpha\beta TCR^{+}CD4^{-}CD8^{-}$ (double negative) T cells is a characteristic feature of ALPS, and this unusual population appears to result from a failure of apoptosis of $CD8^{+}$ T cells (Bristeau-Leprince *et al.* 2008). There are three ALPS subgroups; ALPS-defined genetic mutation (e.g. ALPS-FAS), ALPS-U where there is defective *in vitro* FAS mediated apoptosis and ALPS phenotype where apoptosis is normal.

There is considerable phenotypic overlap between ALPS and some CVID subgroups. While ALPS is usually associated with hypergammaglobulinaemia, this is not an absolute finding. A CVID patient with autoimmunity and persistent lymphadenopathy might be clinically indistinguishable from a hypogammaglobulinaemic ALPS patient. Studies of *in vitro* apoptosis do not form part of the diagnostic criteria for CVID and therefore it is possible that a number of CVID patients could be classified as ALPS-U. The diagnosis reached is somewhat arbitrary, though may depend on whether the main presenting complaint was due to immune dysregulation or hypogammaglobulinaemia. Further investigation of double negative T cells and apoptosis in CVID patients offers an interesting approach to potentially aid differentiation of this heterogeneous group. A study of 66 ALPS patients and 126 CVID patients illustrates some of these points (Rensing-Ehl *et al*. 2010). 5 of the 66 ALPS patients were hypogammaglobulinaemic and suffered with recurrent infections. Of the CVID patients, 31 patients had a combination of lymphoproliferation with raised double negative T cells. A sample of 10 of these 31 patients were studied for *in vitro* apoptosis which was found to be defective in 1 patient. However, this study did

not investigate whether the CVID patients with lymphoproliferation were distinguishable from other CVID patients on immunological grounds. Further investigation would be of interest and in particular examining whether there are differences in the percentage of double negative T cells between CVID patients, and whether these findings are able to differentiate CVID clinical phenotypes. Also, the determination of the frequency of apoptosis defects in CVID patients may be of significance and this group of patients may have a different underlying pathogenetic mechanism. The investigation of apoptosis in CVID patients is rarely considered, and the findings from differential gene expression and crossover in phenotype with ALPS suggest that this would be an interesting avenue to pursue, and that it may be appropriate to consider apoptosis normal and apoptosis abnormal CVID as different diseases.

CVID represents a grouping of heterogeneous conditions linked by a common finding of hypogammaglobulinaemia. The clinical characterisation of the Welsh cohort, and the statistically significant finding that a number of complications are associated together, suggests that hypogammaglobulinaemia without other complications and hypogammaglobulinaemia with complications may represent different disease processes. This may represent a first step towards separating out different subgroups, though the final classification would be expected to be somewhat more complex. A statistically significant increase in granuloma rates in patients with low numbers of naïve CD8 T cells was demonstrated which suggests that in addition to B and CD4 T cell phenotyping, the investigation of CD8 T cells may provide a further means of classifying CVID patients. However, it would appear that rather than refining classifications based on clinical phenotype or lymphocyte enumeration, moving

towards a molecular classification may offer a means of more clearly demarcating CVID subgroups. The consistent finding of differential expression of genes involved in apoptosis suggests that characterising CVID patients according to whether their lymphocytes are apoptosis normal or apoptosis abnormal may represent a further step along this route. Moving towards an accurate classification of CVID subgroups would help direct appropriate management of affected patients and also enable research into the variety of different conditions which are currently encompassed by the term CVID.

Acknowledgements

I would like to take this opportunity to thank my supervisors, Bernhard Moser and Stephen Jolles for their support, and also Adrian Heaps for his help and microarray expertise. I am also grateful to Lesley Jones and Claire Harris for the interesting and useful discussions during the annual appraisal process. I would like to acknowledge the help and support of CBS and in particular Megan Musson and Peter Giles. Thank you to Mo Moody and all in the Immunology diagnostic laboratory at University Hospital of Wales, without whom the diagnosis and monitoring of immunodeficient patients would not be possible, including Tony Bennett and Catherine Rowlands who performed the lymphocyte phenotyping. I appreciate the opportunity to conduct this research which was funded by a grant from the Primary Immunodeficiency Association (PiA), and I would also like to thank the patients who have kindly donated blood samples for research. I am grateful to the Immunology nurses Emily Carne, Hayley Arnold, Mary Neal, Emma Knight, Katrin Morris and Sophie Barber for their help in the consent process and collecting of blood samples, and the Immunology secretaries Bethan Lee and Lisa Thomas for tracking down and retrieving patient notes from the depths of NHS storage. Finally, I would especially like to thank my wife Sarah for all her support and for reviewing the thesis, my brother for his encouragement and mathematical discussions, and my parents and sister for their support and encouragement.

Appendices

Appendix 1: Abbreviations

ADAM	a disintegrin and metallopeptidase
AHA	autoimmune haemolytic anaemia
AID	activation induced cytidine deaminase
ALPS	autoimmune lymphoproliferative syndrome
ALT	alanine aminotransferase
ANOVA	analysis of variance
BAFFR	B cell activating factor receptor
BLNK	B-cell linker
BLS	Bare lymphocyte syndrome
Btk	Bruton's tyrosine kinase
CARD15	caspase recruitment domain family member 15
CD	cluster of differentiation
CDF	cumulative distribution function
cDNA	complementary DNA
CGD	chronic granulomatous disease
CLL	chronic lymphocytic leukaemia
CMC	chronic mucocutaneous candidiasis
CML	chronic myeloid leukemia
CNV	copy number variations
cRNA	complementary RNA
CT	computer tomography

CVID	common variable immunodeficiency
Cy5	cyanine 5
DAVID	database for annotation, visualization and integrated discovery
DNA	deoxyribonucleic acid
EASE	expression analysis systematic explorer
ECD	phycoerythrin-Texas Red conjugate (energy coupled dye)
Fab	fragment, antigen binding (of immunoglobulin)
Fc	fragment, crystallisable (of immunoglobulin)
FcR	Fc receptor
FDR	false discovery rate
FITC	fluorescein isothiocyanate
FHL	familial haemophagocytic lymphohistiocytosis
GASL	granuloma, autoimmunity, splenomegaly &/or persistent lymphadenopathy
GO	gene ontology
GWAS	genome wide association study
HAE	hereditary angioedema
HCV	hepatitis C virus
HHV8	human herpesvirus 8
HIGE	hyper-IgE syndrome
HIGM	hyper-IgM syndrome
HIV	human immunodeficiency virus
HRCT	high resolution computer tomography
HSCT	haematopoietic stem cell transplant
HTLV-1	human T cell lymphotropic virus type 1
IBD	inflammatory bowel disease

ICOS	inducible T-cell co-stimulator
IFN	interferon
Ig	immunoglobulin
IgAD	IgA deficiency
IL	interleukin
IL-1RA	IL-1 receptor antagonist
IRAK1	interleukin-1 receptor-associated kinase 1
ITP	immune thrombocytopenia
JAK	Janus kinase
LOCID	late onset combined immunodeficiency
LREC	local research ethics committee
KNN	K nearest neighbour
LPS	lipopolysaccharide
MHC	major histocompatibility complex
mRNA	messenger RNA
MZL	marginal zone lymphoma
NEMO	NFκB essential modulator
NFκB	nuclear factor kappa-light-chain-enhancer of activated B cells
NK	natural killer
NOD2	nucleotide-binding oligomerisation domain containing 2
ORC4L	origin of recognition complex 4L
OSAI	organ specific autoimmunity
PBMC	peripheral blood mononuclear cell
PCA	principal component analysis
PE	pulmonary embolism

PE	phycoerythrin
PC7	phycoerythrin-cyanine7 conjugate
QC	quality control
RIN	RNA integrity number
RNA	ribonucleic acid
SCID	severe combined immunodeficiency
SDS	sodium dodecyl sulphate
SLE	systemic lupus erythematosus
SNP	single nucleotide polymorphism
SNX31	sorting nexin 31
SSC	saline sodium citrate
SVM	support vector machine
TACI	transmembrane activator and calcium-modulator and cyclophilin ligand interactor
TAP	the transporter associated with antigen presentation
TCR	T cell receptor
TLR	toll like-receptor
TNFα	tumour necrosis factor alpha
UC	ulcerative colitis.
UNG	uracil-DNA glycosylase
VCF	velocardiofacial syndrome
WAS	Wiskott Aldrich syndrome
XLA	X-linked agammaglobulinaemia

Appendix 2: Patient information sheet

Medical Biochemistry
& Immunology
☎ 029 2074 5814
🗎 029 2074 8383

Consultant:	Dr S R A Jolles MSc PhD MRCP MRCPath	
Specialist Registrar:	Dr Tariq El-Shanawany	
Clinical Nurse Specialists:	Emily Carne	
	Emma Knight	Tel: 029 2074 8380
Secretary:	Ms Lisa Thomas	

An investigation into the mechanisms of Primary Immunodeficiency (PID) using gene expression profiling

Introduction

The Clinical Immunology Team at the University Hospital of Wales conduct studies to help develop and improve patient care. To enable the department to do this we require blood samples from patients and healthy controls to investigate this further.

Invitation

You are being invited to take part in the research within the department. Before you decide, it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with friends, relatives and your doctors if you wish. Please do not hesitate to contact us using the details provided at the end if there is anything that is not clear or if you would like further information. Take time to decide whether or not you wish to take part.

What will actually happen?

The department plans to investigate the mechanisms of disease and associated complications in patients with Primary Immunodeficiency. It is necessary to compare the results from patients with healthy controls; therefore samples will be obtained in the same manner as from patients. **From willing and consented participants, a 20ml blood sample (approximately 4 teaspoons of blood) will be taken using only one needle.**

These samples will be taken at the same time as your usual bloods samples are taken (at your clinical appointment), so there is no need for another needle. You may be asked to give blood on several occasions in addition to your usual bloods.

Only the staff in the department working on the study will know details of participants.

Components of the blood, RNA and DNA extracted will be stored in the laboratory at the University Hospital of Wales and cell lines will be established. These may then be used for further studies investigating Primary Immunodeficiency in the future. The sample may only be used for studies investigating Primary Immunodeficiency and each study will have to gain independent ethical approval. By consenting to participate, your sample will be given as a gift and no payment shall be made for the sample or associated costs. You will retain no right to the sample given in the event of withdrawal from the research.

The purpose of the research

The purpose of this research is to investigate the circulating cells in patients with Primary Immunodeficiency and understand the mechanisms that cause patients to have particular symptoms. This may help in the accurate diagnosis of patients with Primary Immunodeficiency. Accurate diagnosis not only determines treatment choices but also has implications for prognosis, carrier testing and potentially antenatal diagnosis.

Why have I been Chosen?

You have been chosen because you are a patient with a confirmed diagnosis of Primary Immunodeficiency.

Do I have to take part?

It is up to you to decide whether or not to take part. If you do decide to take part you will be given this information sheet to keep and will be asked to sign a consent form, which you will also keep. If you decide to take part you are still free to withdraw at any time without giving a reason. This will not affect the standard of care you receive now or in the future.

What will happen to me if I take part?

You will be asked to donate a 20ml blood sample (approximately four teaspoons of blood) in addition to your usual blood samples. The blood will be taken from a vein in your arm using the same needle and should not take longer than two minutes. You may be asked to give a sample such as this on more than one clinic visit, up to a maximum of 300ml (approximately half a pint) in a year.

What is expected of me?

As a patient you will be sent/given the information form prior to your next visit. At your next clinic appointment one of the staff involved in the study will go through the information with you. If you are happy to take part in the study, you will be consented and a blood sample will be taken in addition to your routine blood samples.

What are the possible risks of taking part?

As a patient the blood taking procedure will happen through the needle that you have your routine samples taken from, so no need for extra needle insertions.

The possible risk of taking part is the same risk as having a normal blood test.

You may suffer with slight pain during the blood taking procedure and a small bruise may appear for a couple of days after sample collection.

What are the possible benefits of taking part?

The possible benefit of giving samples for research is to expand the existing knowledge and to improve patient care for patients with Primary Immunodeficiency.

What if new information becomes available?

Sometimes during the course of a research project, new information becomes available. In such an event, you will be informed in person or in writing. Sometimes the information will be published in medical journals, confidentiality will be maintained and no personal information will be published.

What if something goes wrong?

If you are harmed due to someone's negligence, then you may have grounds for a legal action. Regardless of this, if you wish to complain about any aspect of the way you have been approached or treated the normal National Health Service mechanisms are available to you.

Will my taking part in this study be kept confidential?

All information, which is collected, about you during the course of the research will be kept strictly confidential.

What will happen to the results of the research?

The results may be published in medical journals. The results may also be presented at national/international meetings for immunology specialists and local patient meetings. Personal details will remain confidential at all times.

Who is organizing the funding for the research?

The Primary Immunodeficiency Association and The Clinical Trials Fund at University Hospital of Wales have provided funding for this research.

Who has reviewed this research?

The research has been independently peer reviewed by the PIA and the committee of the Clinical Trials Fund. In addition the Research and Development department at the University Hospital of Wales and the local ethics committee in South East Wales have reviewed the research project.

Contacts for further information

If you have any queries or questions please contact the staff below for further information.

Dr. Stephen Jolles
Consultant Immunologist
Tel: 029 2074 5814
stephen.jolles@cardiffandvale.wales.nhs.uk

Sister Emily Carne
Immunology & Allergy Clinical Nurse Specialist
Tel: 029 2074 8380
emily.carne@cardiffandvale.wales.nhs.uk

THANK YOU FOR TAKING THE TIME TO READ THIS INFORMATION

The information from the research will be held by the Clinical Immunology Team at the University Hospital of Wales and will not be given out to any persons without personal details being removed or made anonymous, confidentiality will be strictly maintained at all times, according to the data protection act.

Appendix 3: Patient consent form

Medical Biochemistry
& Immunology
☎ 029 2074 5814
🗎 029 2074 8383

Consultant: Dr S R A Jolles MSc PhD MRCP MRCPath
Specialist Registrar: Dr Tariq El-Shanawany
Clinical Nurse Specialists: Emily Carne
Emma Knight Tel: 029 2074 8380
Secretary: Ms Lisa Thomas

An investigation into the mechanisms of Primary Immunodeficiency (PID) using gene expression profiling

Consent form for Patients

(This part is to be completed by the Patient) (please tick)

I confirm that:

1. I have read and understood the patient information leaflet ☐
   (Please take a copy home with you to keep)
2. I have had an opportunity to discuss the study and ask questions ☐
3. I have had a satisfactory answer to all of my questions ☐
4. I have received enough information about the study ☐
5. I understand that I am free to withdraw from the study:
   - At any time
   - Without having to give a reason
   - Without affecting my future medical care ☐
6. I have had sufficient time to come to my decision ☐
7. I understand that components of my blood will be stored for future clinical studies, relevant to my condition ☐
8. I understand that the blood sample is given as a gift, and that I will retain no right to that blood once donated ☐
9. I agree to take part in the study ☐

Participant (Please sign below and date your signature)

Signed: _____ Date: _____

Print name: _____

Investigator

Signed: _____ Date: _____

Print name: _____

Appendix 4: List of genes for printed microarray

Name	RefSeq
8D6A	NM_016579
A-1-microglobulin	NM_001633
A2M al 2 macroglo	NM_000014
ABCA6	NM_080284
ABCA7	NM_033308
ABCA9	NM_080283
ABCB1	NM_000927
ABCG1	NM_207630
ACAS2	NM_018677
ACP5	NM_001611
ACTB	NM_001101
ACTN1	NM_001102
ACTN4	NM_004924
ACVR2	NM_001616
ADA	NM_000022
ADAM10	NM_001110
ADAM11	NM_021612
ADAM12	NM_003474
ADAM15	NM_003815
ADAM17	NM_003183
ADAM18	NM_014237
ADAM18	NM_014237
ADAM19	NM_023038
ADAM2	NM_001464
ADAM20	NM_003814
ADAM21	NM_003813
ADAM22	NM_021723
ADAM22	NM_021723
ADAM23	NM_003812
ADAM28	NM_014265
ADAM29	NM_021780
ADAM29	NM_014269
ADAM30	NM_021794
ADAM33	NM_025220
ADAM3A	X89657
ADAM3A	NR_001569
ADAM5	NR_001448
ADAM5	NR_001448
ADAM5	XM_171190
ADAM6	NR_002224
ADAM7	NM_003817
ADAM8	NM_001109
ADAM9	NM_001005845
ADAMDEC1	NM_014479
ADAMDEC1	NM_014479
ADAMTS1	NM_006988
ADAMTS10	NM_030957
ADAMTS12	NM_030955
ADAMTS13	NM_139025
ADAMTS14	NM_139155
ADAMTS15	NM_139055
ADAMTS16	NM_139056
ADAMTS17	NM_139057
ADAMTS18	NM_199355
ADAMTS19	NM_133638
ADAMTS2	NM_014244
ADAMTS3	NM_014243
ADAMTS4	NM_005099
ADAMTS5	NM_007038
ADAMTS6	NM_014273
ADAMTS7	NM_014272
ADAMTS8	NM_007037
ADAMTS9	NM_182920
ADAMTSL1	NM_139264
ADPRT	NM_001618
AGC1	NM_013227
AGO3	AK027796
AGO4	NM_017629
AGPAT2	NM_006412
AID	AB040430
AIF1	NM_001623
AIM2	NM_004833
AIRE	NM_000658
ALCAM	R13558
ALEX3	NM_016607
AMN	NM_030943
ANG	AA682399

ANP32A	NM_006305
ANP32B	NM_006401
ANXA1	NM_000700
AOP/PRDX3	NM_006793
AOP2	NM_004905
APAF1	NM_013229
APAF1	NM_013229
APC	NM_000038
APC	NM_000038
APEX/REF-1	NM_001641
APEX2	NM_014481
APOB48R	NM_018690
APOL3	NM_014349
APS	NM_020979
ARC	NM_015193
ARF1	NM_001658
ARF3	NM_001659
ARG1	NM_000045
ARHE	NM_005168
ARHGAP1	NM_004308
ARHGDIA	BC106044
ARVCF	NM_001670
ATF1	NM_005171
ATF2	NM_001880
ATF3	NM_001674
ATF4	D90209
ATF5	NM_012068
ATF6	NM_007348
ATF7	NM_006856
ATK	X58957
ATM	NM_000051
ATOX1	NM_004045
ATP6A1	AF113129
ATP6B1	M25809
ATP6E	NM_001696
ATP6F	NM_004047
ATP6L	M62762
ATP6M8-9	NM_005765
ATP6S1	D16469
ATP6V0D1	NM_004691
ATP6V1C1	NM_001007254
ATP6V1F	NM_004231
ATP6V1f	NM_004231
ATP6V1G1	NM_004888
ATR	NM_001184
ATRX	NM_138271
AUH	NM_024758
AXIN1	NM_003502
AXIN2	NM_004655
B2M	AA670408
B7H1	NM_014143
B7H2	NM_015259
B7H3	NM_025240
BAD	NM_004322
BAG1	NM_004323
BAK1	NM_001188
BAK1	NM_001188
BAX	NM_004324
BAX	NM_138761
BBC3/PUMA	NM_014417
BBC6	AK023420
BCL10	NM_003921
BCL2	NM_000633
BCL3	NM_005178
BCL6	NM_001706
BCL6	NM_001706
BCL9	NM_004326
BCL-X/BCL2L1	NM_138578
BCoR	NM_017745
BID	NM_197967
BIK	NM_001197
BIM	NM_006538
BIRC2	NM_001166
BIRC2	NM_001166
BIRC3	NM_001165
BIRC3	NM_001165
BIRC5	NM_001168
BLIMP1	NM_001198

BLM	NM_000057
BLNK	NM_013314
BLR1	NM_032966
BLR1	NM_032966
BMK	NM_139033
BMP5	NM_021073
BMP6	NM_001718
BMP8	NM_181809
BMPR1A	NM_004329
BMPR2	NM_001204
BNIP3	NM_004052
BNIP3	NM_004052
BNIP3L	NM_004331
BPAG1	NM_001723
BRCA1	NM_007295
BRCA2	NM_000059
BSG	NM_198591
BSG	NM_198591
BTC	NM_001729
BTG1	NM_001731
BTG2	NM_006763
BY55	NM_007053
C/EBP	NM_004364
C-193	NM_014391
C1QA	NM_015991
C1QB	X03084
C1QBP	X75913
C1QR	NM_012072
C1R	X04701
C1S	NM_201442
C2	NM_000063
C3	K02765
C4A	NM_007293
C4B	NM_001002029
C4BPA	NM_000715
C4BPB	NM_000716
C5	NM_001735
C5R1	M62505
C6	NM_000065
C7	NM_000587
C8A	M16974
C8B	M16973
C8G	U08198
C9	NM_001737
CAPG	NM_001747
CAPON	NM_014697
CASK	NM_003688
CASK	NM_003688
CASP1	M87507
CASP10	NM_032977
CASP10	AF111345
CASP2	U13022
CASP3	NM_032991
CASP4	NM_001225
CASP5	NM_004347
CASP5	U28015
CASP6	NM_032992
CASP7	NM_033339
CASP8	NM_033356
CASP8AP2	NM_012115
CASP9	U60521
CAT	NM_001752
CCL1	NM_002981
CCL11	NM_002986
CCL14	NM_004166
CCL16	NM_004590
CCL17	D43767
CCL18	NM_002988
CCL19	NM_006274
CCL2	NM_002982
CCL20	NM_004591
CCL20	NM_004591
CCL21	NM_002989
CCL22	NM_002990
CCL23	NM_005064
CCL24	NM_002991
CCL25	NM_005624
CCL26	NM_006072
CCL28	NM_019846
CCL3	NM_002983
CCL4	NM_002984
CCL5	NM_002985
CCL7/MCP3	NM_006273
CCL8	NM_005623

CCNA1	NM_003914
CCNB1	NM_031966
CCNB2	NM_004701
CCNC	NM_005190
CCND1	NM_053056
CCND2	NM_001759
CCND3	NM_001760
CCNE2	NM_057749
CCNF	NM_001761
CCNG1	NM_004060
CCNG2	NM_004354
CCNH	NM_001239
CCNT1	NM_001240
CCNT2	NM_001241
CCR1	NM_001295
CCR2	NM_000647
CCR3	NM_001837
CCR4	NM_005508
CCR5	NM_000579
CCR6	NM_004367
CCR6	NM_004367
CCR7	NM_001838
CCR8	NM_005201
CCR8	NM_005201
CCR9	NM_031200
CCR9	NM_031200
CCRL2	NM_003965
CCS	NM_005125
CD16	X52645
CD16	AB025256
CD163	NM_004244
CD19	NM_001770
CD1A	NM_001763
CD1B	NM_001764
CD1C	NM_001765
CD1D	NM_001766
CD1E	NM_030893
CD2	NM_001767
CD20	NM_152866
CD207	AJ242859
CD209	NM_021155
CD22	NM_001771
CD24	NM_013230
CD244	NM_016382
CD28	NM_006139
CD2AP	NM_012120
CD33	NM_001772
CD33L1	NM_012211
CD34	NM_001773
CD37	NM_001774
CD39L1	NM_001246
CD39L1	NM_001246
CD3D	NM_000732
CD3E	NM_000733
CD3G	NM_000073
CD4	NM_000616
CD40 ligand	X67878
CD40L	NM_000074
CD43	J04536
CD44	NM_000610
CD45	Y00062
CD48	NM_001778
CD5/T1/Leu-1	NM_014207
CD53	NM_000560
CD55	M31516
CD58	NM_001779
CD59	NM_203329
CD5L	NM_005894
CD6	NM_006725
CD62L	NM_000655
CD64	X14356
CD68	NM_001251
CD69	NM_001781
CD7	NM_006137
CD72	NM_001782
CD74	NM_004355
CD79A	NM_001783
CD79B	NM_000626
CD8	NM_001768
CD8	NM_001768
CD80	NM_005191
CD84	NM_003874
CD86	NM_006889
CD88	Y10205

CD8B1	NM_004931
CD9	NM_001769
CD9	NM_001769
CD97	NM_001784
CDC10	NM_001788
CDC20	NM_001255
CDC25A	NM_201567
CDC25B	NM_004358
CDC25C	NM_022809
CDC37	NM_007065
CDC42	BC002711
CDC6	NM_001254
CDH1	NM_004360
CDH11	NM_033664
CDH13	NM_001257
CDH16	NM_004062
CDH17	NM_004063
CDH18	NM_004934
CDH3	NM_001793
CDH4	NM_001794
CDH5	NM_001795
CDH6	NM_004932
CDK2	NM_001798
CDK4	NM_000075
CDK5	NM_004935
CDKN1A	NM_078467
CDKN1A/p21	NM_000389
CDKN1B	NM_004064
CDKN1C	NM_000076
CDKN2A	NM_000077
CDKN2A	NM_000077
CDKN2B	NM_078487
CDKN2C	NM_078626
CDKN2D	NM_079421
CDKN3	NM_005192
CDS1	NM_001263
CEACAM7	NM_006890
CEB1	NM_016323
CEBPD	NM_005195
CEBPE	NM_001805
CENPE	NM_001813
CES1	NM_001266
CFB	X72875
CFD	M84526
CFH	NM_000186
CFI	J02770
CFLAR	NM_003879
CFP	M83652
CFTR	NM_000492
CGN	AF263462
CHEK1	NM_001274
CHEK2/RAD53	NM_007194
CHIT1	NM_003465
CIITA	X74301
CIZ1	NM_012127
CKN1	NM_000082
CKN1	NM_000082
CLASP1	NM_015282
CLDN1	NM_021101
CLDN10	NM_006984
CLDN11	BC013577
CLDN12	AL136770
CLDN14	AF314090
CLDN15	NM_014343
CLDN16	NM_006580
CLDN17	NM_012131
CLDN18	NM_016369
CLDN2	NM_020384
CLDN3	NM_001306
CLDN4	NM_001305
CLDN5	NM_003277
CLDN6	BC008934
CLDN8	BC020866
CLDN9	NM_020982
CLECSF6	NM_016184
CLECSF9	NM_014358
CLN2	NM_000391
CNK	NM_004073
COL8A2	NM_005202
CORO1A	NM_007074
CORO1C	NM_014325
COX5B	NM_001862
COX7	NM_001867
CP	NM_000096

CPM	NM_001005502
CR1	X14362
CR2	NM_001006658
CR2	NM_001877
CRACC	NM_021181
CRB1	NM_201253
CREB1	NM_134442
CREBBP	NM_004380
CRLF1	NM_004750
CRLF2	NM_022148
CROC4	NM_006365
CSF1	NM_000757
CSF2RA	NM_172247
CSF2RB	NM_000395
CSF3	NM_000759
CSF3	NM_000759
CSF3R	NM_156038
CST6	NM_001323
CTARCK (CCL27)	NM_006664
CTCF	NM_006565
CTD	NM_004715
CTL2	NM_020428
CTLA4	NM_005214
CTNNA1	D14705
CTNNB1	NM_001904
CTNND1	NM_001331
CTSB	NM_147780
CTSD	NM_001909
CTSG	NM_001911
CTSH	NM_004390
CTSL	NM_145918
CTSS	NM_004079
CTSZ	NM_001336
CTTN	NM_005231
CX3CL1	NM_002996
CX3CR1	NM_001337
CXCL1	NM_001511
CXCL1	NM_001511
CXCL10	NM_001565
CXCL11	NM_005409
CXCL12	NM_199168
CXCL13	NM_006419
CXCL14	NM_004887
CXCL16	NM_022059
CXCL2	NM_002089
CXCL3	NM_002090
CXCL5	NM_002994
CXCL6	NM_002993
CXCL9	NM_002416
CXCR2	NM_001557
CXCR3	NM_001504
CXCR4	NM_001008540
CXCR6	NM_006564
CYBA	M21186
CYBB	X04011
CYG	NM_004130
CYP24	NM_000782
CYP2J2	NM_000775
CYR61	NM_001554
CYSLTR1	NM_006639
CYSLTR2	NM_020377
DAD1	NM_001344
DAP	NM_004394
DAP3	NM_004632
DAPK1	NM_004938
DAPK3	NM_001348
DAXX	NM_001350
DCLRE1C	AJ296101
DCLRE1C	NM_022487
DDB1	NM_001923
DDB2	NM_000107
DDIT3/GADD153	NM_004083
DDT	NM_001355
DEFA1	NM_004084
DEFA1	BC112188
DEFA3	NM_005217
DEFA4	NM_001925
DEFB1	NM_005218
DEL1	U70312
DFFA	NM_213566
DGCR2	NM_005137
DICER1	NM_177438
DKK3	NM_015881
DLK1	NM_003836

DLL3	NM_016941
DNApolG	NM_002693
DNCL1	NM_006094
DR6	NM_014452
DSC1	NM_024421
DSC2	NM_024422
DSC2	NM_024422
DSC2	NM_024422
DSC3 (RT-PCR)	NM_001941
DSG1	NM_001942
DSG2	NM_001943
DSG3	NM_001944
DSP	NM_004415
DSTN	NM_006870
DTAF1	AB002331
DVL1	NM_004421
E2-EPF	NM_014501
E2-EPF	NM_014501
EBAF	NM_003240
EBI2	NM_004951
EBI3	NM_145659
EDN2	NM_001956
EEA1	NM_003566
EFNB1	NM_004429
EGFR	NM_005228
EGR1	NM_001964
EGR2	NM_000399
EGR3	NM_004430
EGR4	NM_001965
EIF2B5	NM_003907
EIF2C1	NM_012199
ELA2	NM_001972
ENG	NM_000118
ENTPD6	NM_001247
EPHA3/HEK4	NM_005233
EPHB4	NM_004444
EPHB6	NM_004445
EPOR	NM_000121
ERBB2	NM_004448
ERBB3	NM_001982
ERBIN	NM_018695
ERCC1	NM_202001
ERCC1	NM_202001
ERCC2	NM_000400
ERCC3	NM_000122
ERCC4	NM_005236
ERCC5	NM_000123
ERCC6	NM_000124
ERK1	NM_002746
ERK2	M84489
ERK3	X59727
EVPL	NM_001988
EXO1	NM_130398
EXOSC9	NM_005033
EXT1	NM_000127
EYA1	NM_000503
FADD	NM_003824
FANCG	NM_004629
FAS	M67454
FasL	U11821
FASTK	NM_025096
FCAR	X54150
FCER1A	NM_002001
FCER1G	M33195
FCER2	NM_002002
FCGR1A	NM_000566
FCGR2A	NM_021642
FcRIIb2	M31934
FDX1	M18003
FDXR	J03826
FEN1	NM_004111
FGF1	NM_000800
FGFR1	NM_023109
FGR	NM_005248
FHR-3	NM_021023
FIGF	NM_004469
FKBP1	NM_000801
FKBP4	NM_002014
FLJ11785	NM_021930
FLJ12287	NM_022367
FLT3LG	NM_001459
FOS	NM_005252
FOS	NM_005252
FosB	NM_006732

FOSL1	NM_005438
FOSL2	NM_005253
FOXO3A	NM_201559
FOXP3	NM_014009
FPN1	NM_014585
FPR1	NM_002029
FPRL1	NM_001462
FPRL2	NM_002030
FSBP	NM_006550
FST	NM_006350
FTL	NM_000146
FYN	NM_153047
FZD3	NM_017412
FZD4	NM_012193
G6PD	NM_000402
G6PT1	Y15409
GA17	NM_006360
GAD1/GAD67	NM_000817
GAD2	NM_000818
GADD45	NM_001924
GADD45	NM_001924
GADD45B	NM_015675
GADD45G	NM_006705
GAK	NM_005255
GAPD	NM_002046
GAS1	NM_002048
GATA1	NM_002049
GATA3	NM_001002295
GC	NM_000583
GDI1	NM_001493
GDI2	NM_001494
GFI1	NM_005263
GFRA2	NM_001495
GG2-1	NM_014350
GIP3	NM_002038
GJA1	NM_000165
GJB1	NM_000166
GM-CSF	NM_000758
GNB5	NM_016194
GNLY	NM_006433
GOSR1	NM_004871
GPLD1	L11702
GPR30	NM_001505
GPR65	NM_003608
GPX1	XM_497092
GR/NR3C1	NM_000176
GRB2	NM_002086
GRB7	NM_005310
GSK3A	NM_019884
GSTT1	NM_000853
GTF2A1	NM_015859
GTF2B	NM_001514
GTSE1/B99	NM_016426
GUSB	NM_000181
GZMA	NM_006144
GZMB	NM_004131
GZMK	NM_002104
H1FX	NM_006026
H2AFL	NM_003512
H2AFL	NM_003512
H3.3B	NM_005324
HB15	Z11697
HCK	NM_002110
HCS	NM_000411
HDAC1	NM_004964
HEK	M83941
Hemopexin	NM_000613
HGS	NM_004712
HIF1A	NM_181054
HIF1A	NM_001530
HIVEP2	X65644
HK1	NM_000188
HLA-A	NM_002116
HLA-DMB	U15085
HLA-DNA	NM_002119
HLA-DOB	NM_002120
HLA-DPB1	NM_002121
HLA-DQA	NM_020056
HLA-DQB	NM_002123
HLA-DRA	NM_019111
HLA-DRB1	BC108922
HLA-E	NM_005516
HLA-E	X56841
HLA-G	AB201550

HM74	NM_006018
HML2	NM_182906
HN1	NM_001002032
HOXA3	NM_030661
HOXD3	NM_006898
HPS1	NM_000195
HPS3	NM_032383
HRH1	NM_000861
HRH2	NM_022304
HSF1	NM_005526
HSP75	AF043254
HSPCA	NM_005348
HSPCA	NM_005348
HSPCB	NM_007355
HSPD1	NM_199440
HSPD1	NM_199440
HTR1A	NM_000524
HTR1B	NM_000863
HTR1D	NM_000864
HTR1E	NM_000865
HTR1F	NM_000866
HTR2A	NM_000621
HTR2B	NM_000867
HTR2C	NM_000868
HTR3A	NM_213621
HTR4	NM_199453
HTR5A	NM_024012
HUS1	NM_004507
HYAL2	NM_003773
HYOU1	NM_006389
ICAM1	NM_000201
ICAM2	NM_000873
ICAM4	NM_022377
ICAM5	NM_003259
ICAP-1A	NM_004763
ICOS	NM_012092
ICOSL	AF199028
Id2	NM_001007792
IDH1	NM_005896
IDH2	NM_002168
IDUA	NM_000203
IER3	NM_003897
IER3	NM_003897
IFI16	AF208043
IFI30	NM_006332
IFIT1	NM_001001887
IFIT2	NM_001547
IFNA1	NM_024013
IFNAR1	NM_000629
IFNB1	NM_002176
IFNG	NM_000619
IFNGR1	J03143
IFNGR2	NM_005534
IGBP1	NM_001551
IGF1	NM_000618
IGF1R	NM_000875
IGF2	NM_000612
IGF2R	NM_000876
IGFBP1	NM_000596
IGFBP2	NM_000597
IGFBP3	NM_000598
IGFBP5	NM_000599
IGFBP6	NM_002178
IGFBP7	NM_001553
IGHA2	L04540
IGHE	L00022
IGHG1	AJ294730
IGHG2	AJ294731
IGHG3	AJ390235
IGHG4	AJ294733
IGHM	X58529
IGJ	NM_144646
IGKC	X96754
IGLL1	M27749
IGSF1	NM_205833
IK	NM_006083
IKAP	NM_003640
IKBKB	AF080158
IKBKG	AF074382
IKKA/CHUK	NM_001278
IKKE	NM_014002
IKZF5	NM_022466
IL10	NM_000572
IL10RA	NM_001558

IL10RB	NM_000628
IL11	NM_000641
IL11RA	NM_004512
IL12A	NM_000882
IL12A	M65290
IL12A	NM_000882
IL12B	NM_002187
IL12RB2	NM_001559
IL13	NM_002188
IL14	L15344
IL-15	NM_172175
IL16	NM_004513
IL18	NM_001562
IL18R1	NM_003855
IL18RAP	NM_003853
IL19	NM_153758
IL1A	X02851
IL1B	NM_000576
IL1R1	NM_000877
IL1R2	NM_004633
IL1RAP	NM_002182
IL1RL1	NM_016232
IL2	X01586
IL20	NM_018724
IL21	NM_021803
IL21R	NM_021798
IL22	NM_020525
IL23	NM_016584
IL2RA	X01057
IL2RB	NM_000878
IL2RG	NM_000206
IL3	NM_000588
IL3RA	NM_002183
IL4	NM_000589
IL4R	NM_000418
IL5	NM_000879
IL5RA	M96652
IL6	NM_000600
IL6R	NM_000565
IL6ST	NM_175767
IL7	NM_000880
IL7R	M29696
IL8	NM_000584
IL8RA	NM_000634
IL9	NM_000590
IL9	NM_000590
IL9R	M84747
ILF1	NM_004514
ILF2	NM_004515
ILF3	NM_012218
ILK	NM_004517
ILT10	NR_003061
ILT11	NM_021250
ILT8	AF041262
IMP-1	NM_006546
IMP-2	NM_006548
IMPG1	NM_001563
IMPG2	NM_016247
INDO	NM_002164
INHBA	NM_002192
INHBB	NM_002193
iNOS	AF049656
INPPL1	NM_001567
INSR	NM_000208
IRAK1	NM_001569
IRAK2	NM_001570
IRAK3	NM_007199
IRAK4	NM_016123
IRC1/CMRF-35H	NM_007261
IRE-BP	M58511
IRF1	NM_002198
IRF3	NM_001571
IRF4	NM_002460
IRF5	NM_002200
IRF7	NM_004031
IRFR6	NM_006147
IRLB	X63417
ISG15	NM_005101
ISGF3G	NM_006084
ITCH	NM_031483
ITCH	AB056663
ITGA1	NM_181501
ITGA2	X17033
ITGA2B	NM_000419

ITGA3	NM_005501
ITGA4	L12002
ITGA6	NM_000210
ITGA7	NM_002206
ITGA8	NM_003638
ITGA9	NM_002207
ITGAE	NM_002208
ITGAL	NM_002209
ITGAL	NM_002209
ITGAM	NM_000632
ITGAX	NM_000887
ITGB1	NM_133376
ITGB2	M15395
ITGB3	NM_000212
ITGB4	NM_000213
ITGB4BP	NM_181468
ITGB5	NM_002213
ITGB6	NM_000888
ITGB7	NM_000889
ITGB8	NM_002214
ITGBL1	NM_004791
IVL	NM_005547
JAG1	NM_000214
JAG2	NM_002226
JAK1	NM_002227
JAK2	NM_004972
JAK3	U09607
JAM	AF172398
JNK1	L26318
JNK2	NM_002752
JNK3	NM_138981
Jun	NM_002228
JUNB	NM_002229
JunD	NM_005354
JUP	NM_021991
JUP	NM_021991
K1	NM_006121
K10	NM_000421
K13	NM_153490
K14	NM_000526
K6B	NM_005555
KCNK1/HOHO1	NM_002245
KIF23	NM_138555
KIF23	NM_138555
KIF3C	NM_002254
KIR2DL1	NM_014218
KIT	NM_000222
KLRB1	NM_002258
KLRC1	NM_002259
KLRC3	NM_002261
KLRD1	NM_007334
KNSL6/MCAK	NM_006845
KOC1	NM_006547
KRT 7	NM_005556
KRT17	NM_000422
KRT18	NM_199187
KRT18	NM_000224
KRT19	NM_002276
KRT5	NM_000424
KRT8	NM_002273
LAG3	NM_002286
LAIR1	NM_002287
LAIR2	NM_002288
LAMC2	NM_005562
LAMP1	NM_005561
LAMP2	NM_002294
LAMR1	AK055991
LAMR1 laminin R1	NM_002295
LAPTM5	NM_006762
LAT	NM_014387
LCK	NM_005356
LCP2	NM_005565
LEP	NM_000230
LEP16	NM_032563
LEPR	NM_017526
LEPR	NM_002303
LGALS3BP	NM_005567
LGMN	NM_005606
LIF	NM_002309
LIFR	NM_002310
LIG4	NM_002312
LILRA1	AF025529
LILRA2	NM_006866
LILRA2	NM_006866

LILRA3	NM_006865
LILRA4	NM_012276
LILRA4	NM_012276
LILRA5	NM_024317
LILRB1	NM_006669
LILRB1	AF283985
LILRB2	NM_005874
LILRB2	AF004231
LILRB3	NM_006864
LILRB4	NM_006847
LILRB4	NM_006847
LILRB5	NM_006840
LIN7A	NM_004664
LIPB	M74775
LNK	NM_005475
LOR (loricrin)	NM_000427
LOX	NM_002317
LPIN1	NM_145693
LRDD	NM_145887
LSP1	NM_002339
LST1	NM_007161
LTA	NM_000595
LTA4H	NM_000895
LTB	NM_009588
LTBR	NM_002342
LTC4S	NM_145867
LTF	NM_002343
LU	NM_005581
LY64	NM_005582
LY64	NM_005582
LY6E	NM_002346
LY75/DEC205	NM_002349
LY9	NM_002348
LYPLA1	NM_006330
LYPLA2	NM_007260
LYST	U70064
LYZ	NM_000239
M6PR	NM_002355
MADH1/SMAD1	NM_005900
MADH2	NM_005901
MADH3	NM_005902
MADH4	NM_005359
MADH5	NM_005903
MADH9	NM_005905
MAEA	NM_005882
MAGEH1	NM_014061
MAGI1	AF401656
MAGI-3	AF213259
Magmas	NM_004287
MAL	NM_022438
MANBA	NM_005908
MAP3K5	NM_005923
MAPK11	U53442
MAPK12	U66243
MAPK14	L35253
MAPK14 (p38)	NM_001315
MAPK6	NM_002748
MARCO	NM_006770
MARK1	NM_018650
MARK3	NM_002376
MARS	NM_004990
MBL2	NM_000242
MC1R	NM_002386
MCAM	NM_006500
MCL1	NM_021960
MCM6	NM_005915
MD2	NM_015364
MDM2	NM_006879
MEFF1	NM_003692
MGST1	NM_145764
MGST2	NM_002413
MGST3	NM_004528
MIF	NM_002415
MKK1	NM_002755
MKK2	L11285
MKK3	NM_002756
MKK4	NM_003010
MKK5	NM_002757
MKK6	NM_002758
MKK7	NM_145185
MKNK2	NM_199054
MLH1	NM_000249
MLL5	NM_018682
MLLT4	AB011399

MM-1 (prefoldin 5)	NM_002624
MMAA	NM_172250
MMAB	NM_052845
MMD	NM_012329
MMP1	NM_002421
MMP10	NM_002425
MMP11	NM_005940
MMP12	NM_002426
MMP13	NM_002427
MMP14 MT	NM_004995
MMP15 MT	NM_002428
MMP16 MT	NM_022564
MMP17 MT	NM_016155
MMP19	NM_022792
MMP2	NM_004530
MMP20	NM_004771
MMP21	NM_147191
MMP23A	NR_002946
MMP23B	NM_006983
MMP24	NM_006690
MMP25	NM_022468
MMP25	NM_022468
MMP26	NM_021801
MMP27	NM_022122
MMP28	NM_024302
MMP3	NM_002422
MMP7	NM_002423
MMP8	NM_002424
MMP9	NM_004994
MMP9	NM_004994
MMS19L	NM_022362
MOX2	NM_005944
MP1	AF201947
MPDZ	NM_003829
MPDZ	NM_003829
MPG	NM_002434
MPO	M19507
MPP6	NM_016447
MPS1	NM_001030
MPV17	NM_002437
MRC1	NM_002438
MRC2	NM_006039
MRE11A	NM_005591
MRE11A	NM_005590
MSH2	NM_000251
MSH2	NM_000251
MSH3	NM_002439
MSH6	NM_000179
MSH6	NM_000179
MSN	NM_002444
MSR1	D90187
MST1	NM_020998
MST1R	NM_002447
MT1X	X76717
MTCP1	NM_014221
MTF1	NM_005955
MTX1	NM_198883
MTX2	NM_001006635
MX2	NM_002463
MYBL1	X66087
MYC	NM_002467
MYCBP	NM_012333
MYD88	NM_002468
MYO5A	NM_000259
NAB1	NM_005966
NAB2	NM_005967
NAPG	U78107
NBEA	NM_015678
NBS1	NM_002485
NBS1	NM_002485
NBS1	NM_002485
NCAM	X55322
NCF1	M55067
NCF2	NM_000433
NCF4	NM_000631
NCOR2	NM_006312
NEDD5	NM_001008491
NEU1	NM_000434
NF2	NM_181827
NFAT5	NM_138713
NFATC1	NM_172388
NFKB1	NM_003998
NFKB2	NM_002502
NFKBIA	NM_020529

NFKBIE	U91616
NGFB	NM_002506
NIK	NM_003954
NINJ1	NM_004148
NIPP-1 like	NM_003491
NME1	NM_198175
NME2	NM_002512
NOD1/CARD4	NM_006092
NOS1	U31466
NOS2A	NM_000625
NOS3	NM_000603
NOTCH1	NM_017617
NOTCH2	NM_024408
NOTCH3	NM_000435
NOTCH4	NM_004557
NOX1	NM_007052
NOX3	NM_015718
NOX4	NM_016931
NOX5	NM_024505
NR2F6	NM_005234
NRAMP1	D50402
NRAMP2	L37347
NRGN	NM_006176
NSF	AF102846
NTRK1	NM_002529
OCLN	NM_002538
OLR1	NM_002543
ORM2	NM_000608
OSMR	NM_003999
P2RX4	NM_175567
P2RY2	NM_176072
P2Y6R	NM_004154
p300	NM_001429
p38D	NM_002754
p53	NM_000546
P53AIP1	NM_022112
PA26	NM_014454
PACE	NM_002569
PADI4	NM_012387
PAEP	NM_002571
PAR-3	NM_019619
PARD6A	NM_016948
PARD6B	AB044555
PARD6G	AB044556
PAX1	NM_006192
PAX3	NM_000438
PAX9	NM_006194
PCK2	NM_004563
PCNA	NM_182649
PDCD1	NM_005018
PDCD10	NM_007217
PDGFRB	NM_002609
PDL2	NM_025239
PEA15	NM_003768
PGDS	NM_000954
PGF	NM_002632
PGM3	NM_015599
PGR1	NM_033296
PHB	NM_002634
PI3 (prot inhib 3)	NM_002638
PI3K/PIK3C2B	NM_002646
PIAS2	NM_004671
PIAS2	NM_004671
PIG7	NM_004862
PIGR	NM_002644
PIM1	NM_002648
PIM2	NM_006875
PIN1	NM_006221
PIR51	NM_006479
PKC1	NM_002591
PKCb1	NM_212535
PKCm	NM_002742
PKCz	NM_002744
PKP1	NM_000299
PKP2	NM_001005242
PKP3	NM_007183
PKP4	NM_003628
PLAB	NM_004864
PLAT	NM_000931
PLAU	NM_002658
PLAUR	NM_001005377
PLCB2	NM_004573
PLCB3	NM_000932
PLCG1	M34667

PLCG2	NM_002661
PLN	NM_002667
PMCA4	NM_001001396
PMCA4	NM_001001396
PMS1	NM_000534
PMS2	NM_000535
PNP	X00737
POLH	NM_006502
POLR2A	NM_000937
POLR3F	U93869
POR1	NM_012402
POU2AF1	NM_006235
POU2F2	NM_003698
PPARBP	NM_004774
PPARG	NM_015869
PPIA	NM_021130
PPM1D/Wip1	NM_003620
PPT	NM_000310
PRAK	NM_003668
PRDX1	NM_181697
PRDX4	NM_006406
PRF1	NM_005041
PRG2	NM_002728
PRKACA	NM_002730
PRKCA	X52479
PRKCD	L07860
PRKCE	X65293
PRKCG	NM_002739
PRKCH	M55284
PRKCH	NM_006255
PRKCI	NM_002740
PRKCI	NM_002740
PRKCN	NM_005813
PRKCQ	NM_006257
PRKCQ	NM_006257
PRKDC	NM_006904
PRKDC	NM_006904
PRKR	NM_002759
PRKR	NM_002759
PRKRA	NM_003690
PRKRI	NM_006260
PRODH1	NM_016335
PROML1	NM_006017
PRPL2	X86019
PS1	NM_007318
PSCD2L (cytohesin 1)	NM_004762
PSEN2	NM_000447
PSG11	NM_002785
PSMD13	NM_002817
PSME1	NM_176783
PSTPIP1	NM_003978
PTGDR	NM_000953
PTGER4	NM_000958
PTGS1	NM_000962
PTGS2	NM_000963
PTGS2	NM_000963
PTMA	NM_002823
PTN	NM_002825
PTPN1	NM_002827
PTPNS1	NM_080792
PTPRCAP/LPAP	NM_005608
PTPRF (LAR)	NM_002840
PTPRJ	NM_002843
PTX-3	NM_002852
PURA	NM_005859
PVRL1	AF060231
PXMP3	NM_000318
RAB10	NM_016131
RAB11A	NM_004663
RAB11B	NM_004218
Rab13	NM_002870
RAB14	AF203689
RAB18	NM_021252
RAB2	NM_002865
RAB27A	NM_004580
RAB3A	NM_002866
RAB3B	NM_002867
RAB3C	NM_138453
RAB5A	NM_004162
RAB5B	NM_002868
RAB5C	NM_004583
RAB6	NM_002869
RAB7	NM_004637
RAB7L1	NM_003929

RAB9	NM_004251
RAB9B	NM_016370
RAB9P40	NM_005833
RABEP1	X91141
RAC1	NM_018890
RAC1	M29871
RACK17	NM_005806
RAD1	NM_133377
RAD17	NM_133341
RAD18	NM_020165
RAD21	NM_006265
RAD23A	NM_005053
RAD23B	NM_002874
RAD50	NM_005732
RAD51	NM_002875
RAD51C	NM_058216
RAD51L	NM_133630
RAD51L1	NM_133509
RAD52	NM_134424
RAD52B	NM_145654
RAD54B	NM_134434
RAD54L	NM_003579
RAD9	NM_004584
RAG1	NM_000448
RAG2	NM_000536
RANK	NM_003839
RAP1B	X08004
RARA	NM_000964
RARB	NM_016152
RARG	NM_000966
RDBP	NM_002904
RDX	NM_002906
REC8	NM_005132
REL	NM_002908
RELA	L19067
RELB	NM_006509
REM	NM_014012
REPRIMO	NM_019845
REV1L	NM_016316
REV1L	NM_016316
RFC2	NM_181471
RFNG	XM_113942
RFX5	NM_000449
RFXAP	NM_000538
RIG1	NM_022370
RIP5	NM_018840
RMRP	M29916
ROCK1	NM_005406
RPA3	NM_002947
RPL5	NM_000969
RPS20	AF113008
RPS4X	M22146
RRAD	NM_004165
RRM2B/p53R2	NM_015713
RSHL1	NM_030785
RSN	NM_002956
RXRA	NM_002957
RXRB	NM_021976
RXRG	NM_006917
SAP	AF072930
SAP18	NM_005870
SCAR	NM_003931
SCARA3	NM_016240
SCARB1	NM_005505
SCLY	NM_016510
SCYA13	NM_005408
SCYE1	NM_004757
SDC1	NM_002997
SDC2	NM_002998
SDC3	NM_014654
SDC4	NM_002999
SDCBP	NM_005625
SDCBP	NM_005625
SDCBP2	NM_080489
SDHC	NM_003001
SEC24B	NM_006323
SELE/ELAM1	NM_000450
SELP/CD62	NM_003005
SELPLG	NM_003006
SEPP1	NM_005410
SERPINB1	NM_030666
SERPINB2	NM_002575
SERPING1	M13690
SFN/14-3-3 sigma	NM_006142

SFRP5	NM_003015
SFTP1	NM_005411
SIGLECL1	NM_053003NM_033329
SKIV2L	NM_006929
SLAM	U33017
SLC11A1	NM_000578
SLC21A3	NM_134431
SLC35C1	NM_018389
SLC35C1	NM_018389
SLC4A2	NM_003040
SMAD6	U59914
SMAD7	AF010193
SMC6	NM_024624
SMO Smoothened hom	NM_005631
SNAPA	NM_003827
SNARE-2	NM_054022
SOCS1	NM_003745
SOCS2	NM_003877
SOCS2	NM_003877
SOCS3	NM_003955
SOCS3	NM_003955
SOD1	NM_000454
SOD2	NM_000636
SOD3	NM_003102
SORBS1	NM_006434
SP1	NM_138473
SP110	NM_080424
SP2	NM_003110
SP3	NM_003111
SP-B/SFTPB	NM_000542
SPINK5	NM_006846
SPP1	NM_000582
SPRR1B	NM_003125
SPRR3	NM_005416
SPS	NM_012247
SPS2	NM_012248
SRC	NM_198291
SREBF1	NM_001005291
SRF	NM_003131
SSBP1	NM_003143
STAT1	NM_007315
STAT2	NM_005419
STAT3	NM_139276
STAT4	NM_003151
STAT5A	NM_003152
STAT5B	NM_012448
STAT6	NM_003153
STK11	NM_000455
STX3	U32315
STX4	U07158
STX6	NM_005819
STX7	NM_003569
STXBP2	U63533
SWAP70	NM_015055
SYK	NM_003177
SYMPK	NM_004819
TAB1	NM_006116
TAC1	X54469
TAF7	NM_005642
TAF9	NM_003187
TAHCCP1	AB011182
TAK1	NM_003298
TA-NFKBH	NM_032721
TANK	NM_004180
TAP2	NM_000544
TAPBP	NM_003190
TA-PP2C	NM_139283
TARC (CCL17)	NM_002987
TBET	AF241243
TBX1 Di George Synd	NM_080647
TCF8	NM_030751
TCIRG1	NM_006053
TCRA	M12959
TCRB	X00437
TCRD	X73617
TCRG	AK056843
TERT	NM_003219
TF	NM_001063
TFAP2A	X52611
TFAP2B	X95694
TFAP2C	NM_003222
TFEB	NM_007162
TFR2	NM_003227
TFRC	BC001188

TGFA	NM_003236
TGFB1	NM_000660
TGFB1	NM_000660
TGFB1I1	NM_015927
TGFB2	NM_003238
TGFB3	NM_003239
TGFBR1	NM_004612
TGFBR2	NM_003242
TGM1	NM_000359
THBS2	NM_003247
THO4	NM_005782
TIAF1	NM_004740
TIAL1	NM_003252
TIEG	NM_005655
TIEG2	NM_003597
TIF-1A	NM_015905
TIM-1	NM_012206
TIM-3	NM_032782
TIMP1	NM_003254
TIMP2	NM_003255
TIMP2	NM_003255
TIMP3	NM_000362
TIMP3	NM_000362
TIMP4	NM_003256
TIMP4	NM_003256
TIRAP	NM_052887
TIS11B	NM_004926
TIS11D	U07802
TJP1	NM_175610
TJP2	NM_004817
TJP3	NM_014428
TJP4	NM_080604
TLR1	NM_003263
TLR10	NM_030956
TLR2	NM_003264
TLR3	NM_003265
TLR4	NM_138556
TLR5	NM_003268
TLR6	NM_006068
TLR7	NM_016562
TLR8	NM_016610
TLR9	AF245704
TM6SF2	NM_023002
TM7SF1	NM_003272
TM9SF1	NM_006405
TNF	NM_000594
TNFAIP1	NM_021137
TNFAIP2	NM_006291
TNFAIP3	NM_006290
TNFAIP6	NM_007115
TNFRSF10A	NM_003844
TNFRSF10B	NM_147187
TNFRSF10C	NM_003841
TNFRSF11B	NM_002546
TNFRSF12	NM_148965
TNFRSF13B	NM_012452
TNFRSF14	NM_003820
TNFRSF17	NM_001192
TNFRSF18	AF117297
TNFRSF18	NM_004195
TNFRSF1A	NM_001065
TNFRSF1B	NM_001066
TNFRSF4	NM_003327
TNFRSF5	NM_001250
TNFRSF5	H98636
TNFRSF6	NM_000043
TNFRSF7	NM_001242
TNFRSF8	NM_152942
TNFRSF9	NM_001561
TNFSF10	NM_003810
TNFSF11	NM_033012
TNFSF12	NM_003809
TNFSF13	NM_003808
TNFSF13B/BLYS	NM_006573
TNFSF14	AF036581
TNFSF14	NM_003807
TNFSF14	NM_003807
TNFSF4	NM_003326
TNFSF7	NM_001252
TNFSF8	L09753
TNFSF9	NM_003811
TNR Tenascin-R	NM_003285
TNXB	NM_032470
TOP2A	NM_001067

TOP2B	NM_001068
TP53BPL	AA098867
TPS1	NM_003293
TPSB1	NM_003294
TRADD	NM_153425
TRAF1	U19261
TRAF2	NM_021138
TRAF3	NM_145725
TRAF4	NM_004295
TRAF5	NM_004619
TRAF6	NM_004620
TRAIL	U37518
TRAP1	NM_016292
TRAP1	NM_016292
TRAP3	U12597
TREX1/ATRIP	NM_033627
TRIF	NM_014261
TRIM	AJ224878
TRIP	NM_005879
TRIP6	NM_003302
TSC2	NM_000548
TSSC3	NM_003311
TTP	M63625
TYROBP	NM_198125
TYRP1	NM_000550
UBE2A	NM_003336
UBE2B	NM_003337
UBE2I	NM_003345
UBE3A	NM_130838
UBL1	NM_003352
UNG	NM_080911
UNG	NM_003362
UNG2	NM_021147
UNG2	NM_021147
UQCRC2	NM_003366
UVRAG	NM_003369
VAMP1	NM_016830
VAMP2/FSTL3	NM_005860
VAMP4	NM_201994
VAPA	NM_003574
VAV1	X16316
VAV2	NM_003371
VAV-3	NM_006113
VCAM1	NM_001078
VCL	NM_014000
VCL	NM_014000
VDAC1	NM_003374
VDR	NM_000376
VEGF	NM_003376
VEGFC	NM_005429
VPS45A	NM_007259
VTI1B	NM_006370
VTN	NM_000638
WASP	U12707
WEE1	NM_003390
WIF1	NM_007191
WNT5A	NM_003392
WRN	NM_000553
WSX-1	NM_004843
XBP1	NM_005080
XBP1	NM_005080
XCL1	NM_002995
XCL1	NM_002995
XCL2	NM_003175
XCR1	NM_005283
XCR1	NM_005283
XDH	U39487
XLKD1	NM_006691
XP5	NM_014357
XPA	NM_000380
XPC	NM_004628
XRCC1	NM_006297
XRCC2	NM_005431
XRCC3	NM_005432
XRCC4	NM_022550
XRCC4	NM_022550
XRCC5	NM_021141
XRCC5	NM_021141
YKT6	NM_006555
YME1L	NM_139312
YWHAE	NM_006761
YWHAG	AA432085
YWHAZ	NM_003406
YY1	M76541

YY1/E4TF1	NM_003403
ZAK	NM_016653
ZAP70	L05148
ZBP1	NM_030776
ZFP318	AK057103
ZMPSTE24	NM_005857
ZNFN1A1	NM_006060
ZNFN1A4	NM_022465

Appendix 5: Positions of samples on BeadChip arrays

Patient ID	Slide Barcode	Position
1	5050550033	A
9	5050550033	B
13	5050550033	C
40	5050550033	D
50	5050550033	E
61	5050550033	F
68	5050550033	G
84	5050550033	H
4	5050550060	A
14	5050550060	B
49	5050550060	C
51	5050550060	D
67	5050550060	E
78	5050550060	F
88	5050550060	G
94	5050550060	H
29	5083714003	A
32	5083714003	B
45	5083714003	C
69	5083714003	D
77	5083714003	E
82	5083714003	F
91	5083714003	G
95	5083714003	H
34	5083714028	A
36	5083714028	B
70	5083714028	C
74	5083714028	D
85	5083714028	E
87	5083714028	F
80T	5083714028	G
96X	5083714028	H
3	5262667008	A
7	5262667008	B
8	5262667008	C
23	5262667008	D
26	5262667008	E
43	5262667008	F
44	5262667008	G
53	5262667008	H
10	5262667025	A
30	5262667025	B
35	5262667025	C
56	5262667025	D
58	5262667025	E
72	5262667025	F
73	5262667025	G
80	5262667025	H
11	5262667051	A
27	5262667051	B
39	5262667051	C
42	5262667051	D
64	5262667051	E
66	5262667051	F
80X	5262667051	G
96	5262667051	H
28	5262667022	A
31	5262667022	B
46	5262667022	C
52	5262667022	D
59	5262667022	E
63	5262667022	F
89	5262667022	G
90	5262667022	H

Appendix 6:

Additional information regarding differentially expressed genes

Gene Symbol	Gene product	Disease Associations
AARS2	alanyl-tRNA synthetase 2, mitochondrial	
ABHD5	1-acylglycerol-3-phosphate O-acyltransferase abhydrolase domain-containing protein 5	ichthyosis, developmental delay and steatohepatitis (Chanarin-Dorfman syndrome)
ADA	adenosine deaminase	SCID
ADAMTSL4	ADAMTS-like protein 4	
AHCTF1	AT hook containing transcription factor 1	
ANAPC1	anaphase promoting complex subunit 1	
APAF1	apoptotic peptidase activating factor 1	lymphoma, melanoma, neuroblastoma
ARHGEF3	Rho guanine nucleotide exchange factor 3	
ATP13A1	ATPase type 13A1	
ATP5B	ATP synthase, H+ transporting mitochondrial F1 complex, beta subunit	
ATP7B	ATPase, Cu++ transporting, beta polypeptide	Wilson's disease
ATXN3	ataxin 3	spinocerebellar ataxia type 3
C13ORF18	C13orf18 chromosome 13 open reading frame 18	
C9ORF80	chromosome 9 open reading frame 80	
CATSPER4	cation channel, sperm associated 4	
CCNF	cyclin F	polyscystic kidney disease, breast cancer
CD274	B7 homolog 1 (B7-H1), programmed cell death 1 ligand 1 (PD-L1)	interacts with ICOS, response to viruses, IBD, lung cancer
CD81	target of the antiproliferative antibody 1 (TAPA-1), tetraspanin-28 (Tspan-28)	cell adhesion, tumour metastasis
CDC25B	cell division cycle 25 homolog B	gastric, pancreatic, ovarian and oesophageal cancers
CDCA5	cell division cycle associated 5	
CFLAR	CASP8 and FADD-like apoptosis regulator	Hodgkin's lymphoma
CREB1	cAMP responsive element binding protein 1	melanoma, colon cancer, response to viruses
DAPK2	death-associated protein kinase 2	
DDX19B	DEAD (Asp-Glu-Ala-As) box polypeptide 19B	
DUSP1	dual specificity phosphatase 1	ovarian and lung cancer
EARS2	glutamyl-tRNA synthetase 2, mitochondrial	
ECE1	endothelin converting enzyme 1	velocardiofacial syndrome (VCF)
EIF2B1	eukaryotic translation initiation factor 2B, subunit 1 alpha	leukoencephalopathy
EPHB6	EPH receptor B6	
FBXO31	F-box protein 31	breast cancer
FCAR	IgA Fc receptor, CD89	
FGF1	fibroblast growth factor 1	tumour growth and invasion
GABBR1	gamma-aminobutyric acid (GABA) B receptor, 1	
GML	glycosylphosphatidylinositol anchored molecule like protein	mucosa-associated lymphoid tissue lymphoma, other tumours
GNA11	guanine nucleotide binding protein, alpha 11	
GPR1	G protein-coupled receptor 1	response to HIV
GSK3B	glycogen synthase kinase 3 beta	neurodegenerative disease, prostate cancer, neuroblastoma
HAL	histidine ammonia-lyase	

HIPK2	homeodomain interacting protein kinase 2	breast, thyroid and colorectal cancer
HLA-DPB1	major histocompatibility complex, class II, DP beta 1	sarcoidosis, Grave's disease
IL21	interleukin-21	inflammatory bowel disease (Crohn's and UC)
IL4R	interleukin-4 receptor	mastocytosis
IRAK2	interleukin-1 receptor-associated kinase 2	
ITPKB	inositol 1,4,5-trisphosphate 3-kinase B	
KCND1	potassium voltage-gated channel, Shal-related subfamily, member 1	
KLF11	pruppel-like factor 11	pancreatic cancer
LARS2	leucyl-tRNA synthetase 2, mitochondrial	
MAFG	v-maf musculoaponeurotic fibrosarcoma oncogene homolog G	
MARK2	microtubule affinity-regulating kinase 2	melanoma, lung and breast cancer
MBD4	methyl-CpG binding domain protein 4	colorectal cancer
MCM3	minichromosome maintenance complex component 3	astrocytoma, leukaemia, lymphoma, solid organ tumours
MFGE8	milk fat globule-EGF factor 8 protein	
MYC	v-myc myelocytomatosis viral oncogene homolog	adenocarcinomas, lymphoma
NDE1	nudE nuclear distribution gene E homolog 1	CML, ovarian cancer
NFKBIL1	nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor-like 1	UC
NLRP12	NLR family, pyrin domain containing 12	response to mycobacterium
NOLC1	nucleolar and coiled-body phosphoprotein 1	acute phase response
NR2F6	nuclear receptor subfamily 2, group F, member 6	
OPRK1	opioid receptor, kappa 1	hyperalgesia
OR13C5	olfactory receptor, family 13, subfamily C, member 5	
OR1D4	olfactory receptor, family 1, subfamily D, member 4	
PCNA	proliferating cell nuclear antigen	tumourgenesis
PDCD1	programmed cell death 1	SLE, rheumatoid arthritis, lymphoma
PLEKHF1	leckstrin homology domain containing, family F member 1	
POLA2	polymerase (DNA directed), alpha 2	
PRDX1	peroxiredoxin 1	breast cancer
PRKCH	protein kinase C, eta	
REM2	RAS (RAD and GEM)-like GTP binding 2	
RFC4	replication factor C (activator 1) 4	
RGS2	regulator of G-protein signaling 2	leukaemia
RIPK3	receptor-interacting serine-threonine kinase 3	colon and lung cancer
S1PR4	sphingosine-1-phosphate receptor 4	
SEH1L	nucleoporin SEH1-like	
SLC12A5	solute carrier family 12 (potassium/chloride transporter), member 5	
SLC1A2	solute carrier family 1 (glial high affinity glutamate transporter), member 2	
SLC24A4	solute carrier family 24 (sodium/potassium/calcium exchanger), member 4	
SPTBN1	spectrin, beta, non-erythrocytic 1	neurofibromatosis 2, neuroblastoma
SSTR2	somatostatin receptor 2	pituitary adenoma, lung and breast carcinoma, lymphoma
STEAP3	STEAP family member 3	
SUPT16H	suppressor of Ty 16 homolog	
TAF9	RNA polymerase II, TATA box binding protein (TBP)-associated factor	
TIPARP	TCDD-inducible poly (ADP-ribose) polymerase	
TNFAIP3	tumor necrosis factor, alpha-induced protein 3	
TNFRSF10B	tumor necrosis factor receptor superfamily member 10b	rheumatoid arthritis
TNFRSF4	tumor necrosis factor receptor superfamily member 4, OX40	Grave's disease
TNFSF14	tumor necrosis factor (ligand) superfamily member 14	rheumatoid arthritis, melanoma

TPP1	tripeptidyl peptidase I	hereditary neurodegenerative disease
TRIM38	tripartite motif-containing 38	
UCHL1	ubiquitin carboxyl-terminal esterase L1	lymphoma
VIPR1	vasoactive intestinal peptide receptor 1	
WDR67	WD repeat domain 67	
XRCC2	X-ray repair complementing defective repair in Chinese hamster cells 2	breast, ovarian and skin cancer

Reference List

Agematsu, K., Futatani, T., Hokibara, S., Kobayashi, N., Takamoto, M., Tsukada, S., Suzuki, H. *et al.* (2002). Absence of memory B cells in patients with common variable immunodeficiency. *Clin Immunol* **103:**34-42.

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

Anderson, J. S., Teutsch, M., Dong, Z. and Wortis, H. H. (1996). An essential role for Bruton's tyrosine kinase in the regulation of B-cell apoptosis. *Proc Natl Acad Sci U S A* **93:**10966-10971.

Ardeniz, O. and Cunningham-Rundles, C. (2009). Granulomatous disease in common variable immunodeficiency. *Clin Immunol* **133:**198-207.

Ayroles, J. F. and Gibson, G. (2006). Analysis of variance of microarray data. *DNA Microarrays, Part B: Databases and Statistics* **411:**214-233.

Barnes, M., Freudenberg, J., Thompson, S., Aronow, B. and Pavlidis, P. (2005). Experimental comparison and cross-validation of the Affymetrix and Illumina gene expression analysis platforms. *Nucleic Acids Res* **33:**5914-5923.

Bates, C. A., Ellison, M. C., Lynch, D. A., Cool, C. D., Brown, K. K. and Routes, J. M. (2004). Granulomatous-lymphocytic lung disease shortens survival in common variable immunodeficiency. *J Allergy Clin Immunol* **114:**415-421.

Bauer, J. W., Bilgic, H. and Baechler, E. C. (2009). Gene-expression profiling in rheumatic disease: tools and therapeutic potential. *Nat Rev Rheumatol* **5:**257-265.

Beg, A. A. and Baltimore, D. (1996). An essential role for NF-kappa B in preventing TNF-alpha-induced cell death. *Science* **274:**782-784.

Bjorkander, J., Hammarstrom, L., Smith, C. I., Buckley, R. H., Cunningham-Rundles, C. and Hanson, L. A. (1987). Immunoglobulin prophylaxis in patients with antibody deficiency syndromes and anti-IgA antibodies. *J Clin Immunol* **7:**8-15.

Bolstad, B. M., Irizarry, R. A., Astrand, M. and Speed, T. P. (2003). A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19:**185-193.

Borzutzky, A., Fried, A., Chou, J., Bonilla, F. A., Kim, S. and Dedeoglu, F. (2010). NOD2-associated diseases: Bridging innate immunity and autoinflammation. *Clin Immunol* **134:**251-261.

Boschetti, N., Stucki, M., Spath, P. J. and Kempf, C. (2005). Virus safety of intravenous immunoglobulin: future challenges. *Clin Rev Allergy Immunol* **29:**333-344.

Brennan, V. M., Salome-Bentley, N. J. and Chapel, H. M. (2003). Prospective audit of adverse reactions occurring in 459 primary antibody-deficient patients receiving intravenous immunoglobulin. *Clin Exp Immunol* **133:**247-251.

Bristeau-Leprince, A., Mateo, V., Lim, A., Magerus-Chatinet, A., Solary, E., Fischer, A., Rieux-Laucat, F. *et al.* (2008). Human TCR alpha/beta (+) CD4 (-)CD8 (-) double-negative T cells in patients with autoimmune lymphoproliferative syndrome express restricted V beta TCR diversity and are clonally related to CD8 (+) T cells. *Journal of Immunology* **181:**440-448.

Brosens, L. A., Tytgat, K. M., Morsink, F. H., Sinke, R. J., Ten Berge, I. J., Giardiello, F. M., Offerhaus, G. J. *et al.* (2008). Multiple colorectal neoplasms in X-linked agammaglobulinemia. *Clin Gastroenterol Hepatol* **6:**115-119.

Brouet, J. C., Chedeville, A., Fermand, J. P. and Royer, B. (2000). Study of the B cell memory compartment in common variable immunodeficiency. *Eur J Immunol* **30:**2516-2520.

Bruton, O. C. (1952). Agammaglobulinemia. *Pediatrics* **9:**722-728.

Bryant, A., Calver, N. C., Toubi, E., Webster, A. D. and Farrant, J. (1990). Classification of patients with common variable immunodeficiency by B cell secretion of IgM and IgG in response to anti-IgM and interleukin-2. *Clin Immunol Immunopathol* **56:**239-248.

Butte, A. J., Tamayo, P., Slonim, D., Golub, T. R. and Kohane, I. S. (2000). Discovering functional relationships between RNA expression and chemotherapeutic susceptibility using relevance networks. *Proc Natl Acad Sci U S A* **97:**12182-12186.

Cambronero, R., Sewell, W. A., North, M. E., Webster, A. D. and Farrant, J. (2000). Up-regulation of IL-12 in monocytes: a fundamental defect in common variable immunodeficiency. *J Immunol* **164:**488-494.

Casola, S. (2007). Control of peripheral B-cell development. *Current Opinion in Immunology* **19:**143-149.

Castigli, E., Wilson, S., Garibyan, L., Rachid, R., Bonilla, F., Schneider, L., Morra, M. *et al.* (2007). Reexamining the role of TACI coding variants in common variable immunodeficiency and selective IgA deficiency. *Nat Genet* **39:**430-431.

Chapel, H., Lucas, M., Lee, M., Bjorkander, J., Webster, D., Grimbacher, B., Fieschi, C. *et al.* (2008). Common variable immunodeficiency disorders: division into distinct clinical phenotypes. *Blood* **112:**277-286.

Chapel, H., Lucas, M., Patel, S., Lee, M., Cunningham-Rundles, C., Resnick, E., Gerard, L. *et al.* (2012). Confirmation and improvement of criteria for clinical phenotyping in common variable immunodeficiency disorders in replicate cohorts. *J Allergy Clin Immunol*.

Chapel, H. M., Spickett, G. P., Ericson, D., Engl, W., Eibl, M. M. and Bjorkander, J. (2000). The comparison of the efficacy and safety of intravenous versus subcutaneous immunoglobulin replacement therapy. *J Clin Immunol* **20:**94-100.

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.

Chinen, J. and Shearer, W. T. (2010). Secondary immunodeficiencies, including HIV infection. *J Allergy Clin Immunol* **125:**S195-203.

Conley, M. E., Dobbs, A. K., Farmer, D. M., Kilic, S., Paris, K., Grigoriadou, S., Coustan-Smith, E. *et al.* (2009). Primary B cell immunodeficiencies: comparisons and contrasts. *Annu Rev Immunol* **27:**199-227.

Conley, M. E., Notarangelo, L. D. and Etzioni, A. (1999). Diagnostic criteria for primary immunodeficiencies. Representing PAGID (Pan-American Group for Immunodeficiency) and ESID (European Society for Immunodeficiencies). *Clin Immunol* **93:**190-197.

Cooper, N. R. (1985). The classical complement pathway: activation and regulation of the first complement component. *Adv Immunol* **37:**151-216.

Cunningham-Rundles, C. (2008). Autoimmune manifestations in common variable immunodeficiency. *J Clin Immunol* **28 Suppl 1:**S42-45.

Cunningham-Rundles, C. and Bodian, C. (1999). Common variable immunodeficiency: clinical and immunological features of 248 patients. *Clin Immunol* **92:**34-48.

Davies, D. R. and Metzger, H. (1983). Structural basis of antibody function. *Annu Rev Immunol* **1:**87-117.

de Vries, E. (2006). Patient-centred screening for primary immunodeficiency: a multi-stage diagnostic protocol designed for non-immunologists. *Clin Exp Immunol* **145:**204-214.

Dennis, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C. and Lempicki, R. A. (2003). DAVID: Database for annotation, visualization, and integrated discovery. *Genome Biology* **4:**-.

Dinarello, C. A. (2011). Interleukin-1 in the pathogenesis and treatment of inflammatory diseases. *Blood*.

Du, R., Tantisira, K., Carey, V., Bhattacharya, S., Metje, S., Kho, A. T., Klanderman, B. J. *et al.* (2009). Platform dependence of inference on gene-wise and gene-set involvement in human lung development. *BMC Bioinformatics* **10:**189.

Early, P., Huang, H., Davis, M., Calame, K. and Hood, L. (1980). An immunoglobulin heavy chain variable region gene is generated from three segments of DNA: VH, D and JH. *Cell* **19:**981-992.

Eibel, H., Warnatz, K., Salzer, U., Rizzi, M., Fischer, B., Gutenberger, S., Bohm, J. *et al.* (2009). B-cell activating factor receptor deficiency is associated with an adult-onset antibody deficiency syndrome in humans. *Proceedings of the National Academy of Sciences of the United States of America* **106:**13945-13950.

Eibl, M. M. and Wedgwood, R. J. (1989). Intravenous immunoglobulin: a review. *Immunodefic Rev* **1 Suppl:**1-42.

El-Shanawany, T. and Jolles, S. (2007). Intravenous immunoglobulin and autoimmune disease. *Ann N Y Acad Sci* **1110:**507-515.

El-Shanawany, T., Jolles, S., Unsworth, D. J. and Williams, P. (2009). A recipient of immunoglobulin from a donor who developed vCJD. *Vox Sang* **96:**270.

El-Shanawany, T. M., Williams, P. E. and Jolles, S. (2007). Response of refractory immune thrombocytopenic purpura in a patient with common variable immunodeficiency to treatment with rituximab. *J Clin Pathol* **60:**715-716.

Fakhouri, F., Robino, C., Lemaire, M., Droz, D., Noel, L. H., Knebelmann, B. and Lesavre, P. (2001). Granulomatous renal disease in a patient with common variable immunodeficiency. *Am J Kidney Dis* **38:**E7.

Fambrough, D., McClure, K., Kazlauskas, A. and Lander, E. S. (1999). Diverse signaling pathways activated by growth factor receptors induce broadly overlapping, rather than independent, sets of genes. *Cell* **97:**727-741.

Ferreira, A., Garcia Rodriguez, M. C., Lopez-Trascasa, M., Pascual Salcedo, D. and Fontan, G. (1988). Anti-IgA antibodies in selective IgA deficiency and in primary immunodeficient patients treated with gamma-globulin. *Clin Immunol Immunopathol* **47:**199-207.

Fevang, B., Yndestad, A., Sandberg, W. J., Holm, A. M., Muller, F., Aukrust, P. and Froland, S. S. (2007). Low numbers of regulatory T cells in common variable immunodeficiency: association with chronic inflammation in vivo. *Clin Exp Immunol* **147:**521-525.

Fisher, G. H., Rosenberg, F. J., Straus, S. E., Dale, J. K., Middelton, L. A., Lin, A. Y., Strober, W. *et al.* (1995). Dominant Interfering Fas Gene-Mutations Impair Apoptosis in a Human Autoimmune Lymphoproliferative Syndrome. *Cell* **81:**935-946.

Frost, G. I. (2007). Recombinant human hyaluronidase (rHuPH20): an enabling platform for subcutaneous drug and fluid administration. *Expert Opin Drug Deliv* **4:**427-440.

Fujimoto, M., Poe, J. C., Hasegawa, M. and Tedder, T. F. (2000). CD19 regulates intrinsic B lymphocyte signal transduction and activation through a novel mechanism of processive amplification. *Immunol Res* **22:**281-298.

Galama, J. M., Vogels, M. T., Jansen, G. H., Gielen, M. and Heessen, F. W. (1997). Antibodies against enteroviruses in intravenous Ig preparations: great variation in titres and poor correlation with the incidence of circulating serotypes. *J Med Virol* **53:**273-276.

Gardulf, A. (2007). Immunoglobulin treatment for primary antibody deficiencies: advantages of the subcutaneous route. *BioDrugs* **21:**105-116.

Gerhold, D. L., Jensen, R. V. and Gullans, S. R. (2002). Better therapeutics through microarrays. *Nat Genet* **32 Suppl:**547-551.

Giovannetti, A., Pierdominici, M., Mazzetta, F., Marziali, M., Renzi, C., Mileo, A. M., De Felice, M. *et al.* (2007). Unravelling the complexity of T cell abnormalities in common variable immunodeficiency. *J Immunol* **178:**3932-3943.

Golub, T. R., Slonim, D. K., Tamayo, P., Huard, C., Gaasenbeek, M., Mesirov, J. P., Coller, H. *et al.* (1999). Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* **286:**531-537.

Grimbacher, B., Hutloff, A., Schlesier, M., Glocker, E., Warnatz, K., Drager, R., Eibel, H. *et al.* (2003). Homozygous loss of ICOS is associated with adult-onset common variable immunodeficiency. *Nat Immunol* **4:**261-268.

Gross, J. A., Dillon, S. R., Mudri, S., Johnston, J., Littau, A., Roque, R., Rixon, M. *et al.* (2001). TACI-Ig neutralizes molecules critical for B cell development and autoimmune disease. impaired B cell maturation in mice lacking BLyS. *Immunity* **15:**289-302.

Gunderson, K. L., Kruglyak, S., Graige, M. S., Garcia, F., Kermani, B. G., Zhao, C., Che, D. *et al.* (2004). Decoding randomly ordered DNA arrays. *Genome Res* **14:**870-877.

Halliday, E., Winkelstein, J. and Webster, A. D. (2003). Enteroviral infections in primary immunodeficiency (PID): a survey of morbidity and mortality. *J Infect* **46:**1-8.

Harkin, D. P., Bean, J. M., Miklos, D., Song, Y. H., Truong, V. B., Englert, C., Christians, F. C. *et al.* (1999). Induction of GADD45 and JNK/SAPK-dependent apoptosis following inducible expression of BRCA1. *Cell* **97:**575-586.

Hatab, A. Z. and Ballas, Z. K. (2005). Caseating granulomatous disease in common variable immunodeficiency treated with infliximab. *J Allergy Clin Immunol* **116:**1161-1162.

Henckaerts, L. and Vermeire, S. (2007). NOD2/CARD15 disease associations other than Crohn's disease. *Inflammatory Bowel Diseases* **13:**235-241.

Henin, Y., Marechal, V., Barre-Sinoussi, F., Chermann, J. C. and Morgenthaler, J. J. (1988). Inactivation and partition of human immunodeficiency virus during Kistler and Nitschmann fractionation of human blood plasma. *Vox Sang* **54:**78-83.

Heyman, B. (2000). Regulation of antibody responses via antibodies, complement, and Fc receptors. *Annu Rev Immunol* **18:**709-737.

Hiemstra, P. S., Gorter, A., Stuurman, M. E., Van Es, L. A. and Daha, M. R. (1987). Activation of the alternative pathway of complement by human serum IgA. *Eur J Immunol* **17:**321-326.

Holm, A. M., Sivertsen, E. A., Tunheim, S. H., Haug, T., Bjerkeli, V., Yndestad, A., Aukrust, P. *et al.* (2004). Gene expression analysis of peripheral T cells in a subgroup of common variable immunodeficiency shows predominance of CCR7 (-) effector-memory T cells. *Clin Exp Immunol* **138:**278-289.

Hosack, D. A., Dennis, G., Sherman, B. T., Lane, H. C. and Lempicki, R. A. (2003). Identifying biological themes within lists of genes with EASE. *Genome Biology* **4:**-.

Huang, D. W., Sherman, B. T. and Lempicki, R. A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4:**44-57.

Hutloff, A., Dittrich, A. M., Beier, K. C., Eljaschewitsch, B., Kraft, R., Anagnostopoulos, I. and Kroczek, R. A. (1999). ICOS is an inducible T-cell co-stimulator structurally and functionally related to CD28. *Nature* **397:**263-266.

Iizasa, H., Yoneyama, H., Mukaida, N., Katakoka, Y., Naito, M., Yoshida, N., Nakashima, E. *et al.* (2005). Exacerbation of granuloma formation in IL-1 receptor antagonist-deficient mice with impaired dendritic cell maturation associated with Th2 cytokine production. *J Immunol* **174:**3273-3280.

Jeremias, I., Kupatt, C., Baumann, B., Herr, I., Wirth, T. and Debatin, K. M. (1998). Inhibition of nuclear factor kappa B activation attenuates apoptosis resistance in lymphoid cells. *Blood* **91:**4624-4631.

Kanegane, H., Agematsu, K., Futatani, T., Sira, M. M., Suga, K., Sekiguchi, T., van Zelm, M. C. *et al.* (2007). Novel mutations in a Japanese patient with CD19 deficiency. *Genes Immun* **8:**663-670.

Kasahara, K., Kobayashi, K., Shikama, Y., Yoneya, I., Soezima, K., Ide, H. and Takahashi, T. (1988). Direct evidence for granuloma-inducing activity of interleukin-1. Induction of experimental pulmonary granuloma formation in mice by interleukin-1-coupled beads. *Am J Pathol* **130:**629-638.

Knight, E., Carne, E., Novak, B., El-Shanawany, T., Williams, P., Pickersgill, T. and Jolles, S. (2010). Self-administered hyaluronidase-facilitated subcutaneous immunoglobulin home therapy in a patient with primary immunodeficiency. *J Clin Pathol*.

Kobata, T., Agematsu, K., Kameoka, J., Schlossman, S. F. and Morimoto, C. (1994). Cd27 Is a Signal-Transducing Molecule Involved in CD45RA (+) Naive T-Cell Costimulation. *Journal of Immunology* **153:**5422-5432.

Kraal, G., Weissman, I. L. and Butcher, E. C. (1982). Germinal centre B cells: antigen specificity and changes in heavy chain class expression. *Nature* **298:**377-379.

Kuntz, M., Goldacker, S., Blum, H. E., Pircher, H., Stampf, S., Peter, H. H., Thimme, R. *et al.* (2011). Analysis of bulk and virus-specific CD8+ T cells reveals advanced differentiation of CD8+ T cells in patients with common variable immunodeficiency. *Clin Immunol* **141:**177-186.

Lam, L., Whitsett, C. F., McNicholl, J. M., Hodge, T. W. and Hooper, J. (1993). Immunologically active proteins in intravenous immunoglobulin. *Lancet* **342:**678.

Lamari, F., Anastassiou, E. D., Tsegenidis, T., Dimitracopoulos, G. and Karamanos, N. K. (1999). An enzyme immunoassay to determine the levels of specific antibodies toward bacterial surface antigens in human immunoglobulin preparations and blood serum. *J Pharm Biomed Anal* **20:**913-920.

Li, Q. and Verma, I. M. (2002). NF-kappaB regulation in the immune system. *Nat Rev Immunol* **2:**725-734.

Lilic, D. and Sewell, W. A. (2001). IgA deficiency: what we should-or should not-be doing. *J Clin Pathol* **54:**337-338.

Livaditi, O., Giamarellos-Bourboulis, E. J., Kakkas, I., Kapsimali, V., Lymberi, P., Papastariades, C. and Douzinas, E. E. (2007). Grouping of patients with common variable immunodeficiency based on immunoglobulin biosynthesis: comparison with a classification system on CD4-naive cells. *Immunol Lett* **114:**103-109.

Losi, C. G., Salzer, U., Gatta, R., Lougaris, V., Cattaneo, G., Meini, A., Soresina, A. *et al.* (2006). Mutational analysis of human BLyS in patients with common variable immunodeficiency. *J Clin Immunol* **26:**396-399.

Losi, C. G., Silini, A., Fiorini, C., Soresina, A., Meini, A., Ferrari, S., Notarangelo, L. D. *et al.* (2005). Mutational analysis of human BAFF receptor TNFRSF13C (BAFF-R) in patients with common variable immunodeficiency. *J Clin Immunol* **25:**496-502.

Mackay, F., Schneider, P., Rennert, P. and Browning, J. (2003). BAFF AND APRIL: a tutorial on B cell survival. *Annu Rev Immunol* **21:**231-264.

Malphettes, M., Gérard, L., Carmagnat, M., Mouillot, G., Vince, N., Boutboul, D., Bérezné, A. *et al.* (2009). Late-onset combined immune deficiency: a subset of common variable immunodeficiency with severe T cell defect. *Clin Infect Dis* **49:**1329-1338.

McAdam, A. J., Greenwald, R. J., Levin, M. A., Chernova, T., Malenkovich, N., Ling, V., Freeman, G. J. *et al.* (2001). ICOS is critical for CD40-mediated antibody class switching. *Nature* **409:**102-105.

Mechanic, L. J., Dikman, S. and Cunningham-Rundles, C. (1997). Granulomatous disease in common variable immunodeficiency. *Ann Intern Med* **127:**613-617.

Melanned, I. R., Stein, M. R., Wasserman, R. L., Leibl, H., Engl, W., Yocum, R. C. and Schiff, R. I. (2008). Recombinant human hyaluronidase facilitates dispersion of subcutaneously administered Gammagard liquid and enables administration of a full monthly bose in a single site to patients with immunodeficiency diseases. *Journal of Allergy and Clinical Immunology* **121:**S83-S83.

Michel, M., Chanet, V., Galicier, L., Ruivard, M., Levy, Y., Hermine, O., Oksenhendler, E. *et al.* (2004). Autoimmune thrombocytopenic purpura and common variable immunodeficiency: analysis of 21 cases and review of the literature. *Medicine (Baltimore)* **83:**254-263.

Monti, S., Tamayo, P., Mesirov, J. and Golub, T. (2003). Consensus clustering: A resampling-based method for class discovery and visualization of gene expression microarray data. *Machine Learning* **52:**91-118.

Moratto, D., Gulino, A. V., Fontana, S., Mori, L., Pirovano, S., Soresina, A., Meini, A. *et al.* (2006). Combined decrease of defined B and T cell subsets in a group of common variable immunodeficiency patients. *Clin Immunol* **121:**203-214.

Morimoto, Y. and Routes, J. M. (2005). Granulomatous disease in common variable immunodeficiency. *Curr Allergy Asthma Rep* **5:**370-375.

Mrusek, S., Marx, A., Kummerle-Deschner, J., Tzaribachev, N., Enders, A., Riede, U. N., Warnatz, K. *et al.* (2004). Development of granulomatous common variable immunodeficiency subsequent to infection with Toxoplasma gondii. *Clin Exp Immunol* **137:**578-583.

Mullighan, C. G., Marshall, S. E., Bunce, M. and Welsh, K. I. (1999). Variation in immunoregulatory genes determines the clinical phenotype of common variable immunodeficiency. *Genes and Immunity* **1:**137-148.

Neuberger, M. S. and Milstein, C. (1995). Somatic hypermutation. *Curr Opin Immunol* **7:**248-254.

Notarangelo, L. D., Lanzi, G., Peron, S. and Durandy, A. (2006). Defects of class-switch recombination. *Journal of Allergy and Clinical Immunology* **117:**855-864.

Notarangelo, L. D., Lanzi, G., Toniati, P. and Giliani, S. (2007). Immunodeficiencies due to defects of class-switch recombination. *Immunol Res* **38:**68-77.

Oksenhendler, E., Gerard, L., Fieschi, C., Malphettes, M., Mouillot, G., Jaussaud, R., Viallard, J. F. *et al.* (2008). Infections in 252 patients with common variable immunodeficiency. *Clin Infect Dis* **46:**1547-1554.

Oliveira, J. B., Bleesing, J. J., Dianzani, U., Fleisher, T. A., Jaffe, E. S., Lenardo, M. J., Rieux-Laucat, F. *et al.* (2010). Revised diagnostic criteria and classification for the autoimmune lymphoproliferative syndrome (ALPS): report from the 2009 NIH International Workshop. *Blood* **116:**E35-E40.

Orange, J. S., Glessner, J. T., Resnick, E., Sullivan, K. E., Lucas, M., Ferry, B., Kim, C. E. *et al.* (2011). Genome-wide association identifies diverse causes of common variable immunodeficiency. *J Allergy Clin Immunol*.

Packwood, K., Drewe, E., Staples, E., Webster, D., Witte, T., Litzman, J., Egner, W. *et al.* (2010). NOD2 polymorphisms in clinical phenotypes of common variable immunodeficiency disorders. *Clin Exp Immunol* **161:**536-541.

Pan-Hammarstrom, Q., Salzer, U., Du, L., Bjorkander, J., Cunningham-Rundles, C., Nelson, D. L., Bacchelli, C. *et al.* (2007). Reexamining the role of TACI coding variants in common variable immunodeficiency and selective IgA deficiency. *Nat Genet* **39:**429-430.

Park, J. H. and Levinson, A. I. (2010). Granulomatous-lymphocytic interstitial lung disease (GLILD) in common variable immunodeficiency (CVID). *Clin Immunol* **134:**97-103.

Pascual, V., Chaussabel, D. and Banchereau, J. (2010). A genomic approach to human autoimmune diseases. *Annu Rev Immunol* **28:**535-571.

Perou, C. M., Sorlie, T., Eisen, M. B., van de Rijn, M., Jeffrey, S. S., Rees, C. A., Pollack, J. R. *et al.* (2000). Molecular portraits of human breast tumours. *Nature* **406:**747-752.

Piqueras, B., Lavenu-Bombled, C., Galicier, L., Bergeron-van der Cruyssen, F., Mouthon, L., Chevret, S., Debre, P. *et al.* (2003). Common variable immunodeficiency patient classification based on impaired B cell memory differentiation correlates with clinical aspects. *J Clin Immunol* **23:**385-400.

Qin, H., Yamada, M., Tian, L., Stewart, D. M., Gulino, A. V. and Nelson, D. L. (2003). Tracking gene expression in primary immunodeficiencies. *Curr Opin Allergy Clin Immunol* **3:**437-442.

Quinti, I., Soresina, A., Spadaro, G., Martino, S., Donnanno, S., Agostini, C., Claudio, P. *et al.* (2007). Long-term follow-up and outcome of a large cohort of patients with common variable immunodeficiency. *J Clin Immunol* **27:**308-316.

Reichl, H. E., Foster, P. R., Welch, A. G., Li, Q., MacGregor, I. R., Somerville, R. A., Fernie, K. *et al.* (2002). Studies on the removal of a bovine spongiform encephalopathy-derived agent by processes used in the manufacture of human immunoglobulin. *Vox Sanguinis* **83:**137-145.

Rensing-Ehl, A., Warnatz, K., Fuchs, S., Schlesier, M., Salzer, U., Draeger, R., Bondzio, I. *et al.* (2010). Clinical and immunological overlap between autoimmune

lymphoproliferative syndrome and common variable immunodeficiency. *Clinical Immunology* **137:**357-365.

Salzer, U., Chapel, H. M., Webster, A. D., Pan-Hammarstrom, Q., Schmitt-Graeff, A., Schlesier, M., Peter, H. H. *et al.* (2005). Mutations in TNFRSF13B encoding TACI are associated with common variable immunodeficiency in humans. *Nat Genet* **37:**820-828.

Salzer, U., Maul-Pavicic, A., Cunningham-Rundles, C., Urschel, S., Belohradsky, B. H., Litzman, J., Holm, A. *et al.* (2004). ICOS deficiency in patients with common variable immunodeficiency. *Clin Immunol* **113:**234-240.

Salzer, U., Neumann, C., Thiel, J., Woellner, C., Pan-Hammarstrom, Q., Lougaris, V., Hagena, T. *et al.* (2008). Screening of functional and positional candidate genes in families with common variable immunodeficiency. *BMC Immunol* **9:**3.

Sanz, I., Wei, C., Lee, F. E. and Anolik, J. (2008). Phenotypic and functional heterogeneity of human memory B cells. *Semin Immunol* **20:**67-82.

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.

Schiemann, B., Gommerman, J. L., Vora, K., Cachero, T. G., Shulga-Morskaya, S., Dobles, M., Frew, E. *et al.* (2001). An essential role for BAFF in the normal development of B cells through a BCMA-independent pathway. *Science* **293:**2111-2114.

Scott, L. J., Bryant, A., Webster, A. D. and Farrant, J. (1994). Failure in IgA secretion by surface IgA-positive B cells in common variable immunodeficiency (CVID). *Clin Exp Immunol* **95:**10-13.

Sekine, H., Ferreira, R. C., Pan-Hammarstrom, Q., Graham, R. R., Ziemba, B., de Vries, S. S., Liu, J. *et al.* (2007). Role for Msh5 in the regulation of Ig class switch recombination. *Proc Natl Acad Sci U S A* **104:**7193-7198.

Seshasayee, D., Valdez, P., Yan, M., Dixit, V. M., Tumas, D. and Grewal, I. S. (2003). Loss of TACI causes fatal lymphoproliferation and autoimmunity, establishing TACI as an inhibitory BLyS receptor. *Immunity* **18:**279-288.

Siegal, F. P., Pernis, B. and Kunkel, H. G. (1971). Lymphocytes in human immunodeficiency states: a study of membrane-associated immunoglobulins. *Eur J Immunol* **1:**482-486.

Sorlie, T., Perou, C. M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T. *et al.* (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* **98:**10869-10874.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J Mol Biol* **98:**503-517.

Takei, S., Arora, Y. K. and Walker, S. M. (1993). Intravenous immunoglobulin contains specific antibodies inhibitory to activation of T cells by staphylococcal toxin superantigens [see comment]. *J Clin Invest* **91:**602-607.

Thatayatikom, A., Thatayatikom, S. and White, A. J. (2005). Infliximab treatment for severe granulomatous disease in common variable immunodeficiency: a case report and review of the literature. *Ann Allergy Asthma Immunol* **95:**293-300.

Thorpe, S. J., Fox, B., Heath, A., Behr-Gross, M. E., Virata, M. L. and Yu, M. Y. (2006). International collaborative study to assess candidate reference preparations to control the level of anti-D in IVIG for use in Europe and the United States. *Biologicals* **34:**209-212.

Uzel, G. (2005). The range of defects associated with nuclear factor kappaB essential modulator. *Curr Opin Allergy Clin Immunol* **5:**513-518.

van der Meer, J. W., Weening, R. S., Schellekens, P. T., van Munster, I. P. and Nagengast, F. M. (1993). Colorectal cancer in patients with X-linked agammaglobulinaemia. *Lancet* **341:**1439-1440.

van Zelm, M. C., Reisli, I., van der Burg, M., Castano, D., van Noesel, C. J., van Tol, M. J., Woellner, C. *et al.* (2006). An antibody-deficiency syndrome due to mutations in the CD19 gene. *N Engl J Med* **354:**1901-1912.

Venkitaraman, A. R., Williams, G. T., Dariavach, P. and Neuberger, M. S. (1991). The B-cell antigen receptor of the five immunoglobulin classes. *Nature* **352:**777-781.

Viallard, J. F., Blanco, P., Andre, M., Etienne, G., Liferman, F., Neau, D., Vidal, E. *et al.* (2006). CD8+HLA-DR+ T lymphocytes are increased in common variable

immunodeficiency patients with impaired memory B-cell differentiation. *Clin Immunol* **119:**51-58.

Vine, A. M., Heaps, A. G., Kaftantzi, L., Mosley, A., Asquith, B., Witkover, A., Thompson, G. *et al.* (2004). The role of CTLs in persistent viral infection: cytolytic gene expression in CD8+ lymphocytes distinguishes between individuals with a high or low proviral load of human T cell lymphotropic virus type 1. *J Immunol* **173:**5121-5129.

Vlkova, M., Thon, V., Sarfyova, M., Blaha, L., Svobodnik, A., Lokaj, J. and Litzman, J. (2006). Age dependency and mutual relations in T and B lymphocyte abnormalities in common variable immunodeficiency patients. *Clin Exp Immunol* **143:**373-379.

Waldrep, M. L., Zhuang, Y. and Schroeder, H. W., Jr. (2009). Analysis of TACI mutations in CVID & RESPI patients who have inherited HLA B*44 or HLA*B8. *BMC Med Genet* **10:**100.

Wang, H. Y., Malek, R. L., Kwitek, A. E., Greene, A. S., Luu, T. V., Behbahani, B., Frank, B. *et al.* (2003). Assessing unmodified 70-mer oligonucleotide probe performance on glass-slide microarrays. *Genome Biol* **4:**R5.

Wang, J. and Cunningham-Rundles, C. (2005). Treatment and outcome of autoimmune hematologic disease in common variable immunodeficiency (CVID). *J Autoimmun* **25:**57-62.

Warnatz, K., Denz, A., Drager, R., Braun, M., Groth, C., Wolff-Vorbeck, G., Eibel, H. *et al.* (2002). Severe deficiency of switched memory B cells (CD27 (+)IgM (-)IgD (-)) in subgroups of patients with common variable immunodeficiency: a new approach to classify a heterogeneous disease. *Blood* **99:**1544-1551.

Warnatz, K. and Schlesier, M. (2008). Flowcytometric phenotyping of common variable immunodeficiency. *Cytometry Part B-Clinical Cytometry* **74B:**261-271.

Wehr, C., Kivioja, T., Schmitt, C., Ferry, B., Witte, T., Eren, E., Vlkova, M. *et al.* (2008). The EUROclass trial: defining subgroups in common variable immunodeficiency. *Blood* **111:**77-85.

Weiland, O., Mattsson, L. and Glaumann, H. (1986). Non-A, non-B hepatitis after intravenous gammaglobulin. *Lancet* **1:**976-977.

Wheat, W. H., Cool, C. D., Morimoto, Y., Rai, P. R., Kirkpatrick, C. H., Lindenbaum, B. A., Bates, C. A. *et al.* (2005). Possible role of human herpesvirus 8 in the lymphoproliferative disorders in common variable immunodeficiency. *J Exp Med* **202:**479-484.

Wilkerson, M. D. and Hayes, D. N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26:**1572-1573.

Wiltgen, M. and Tilz, G. P. (2007). DNA microarray analysis: Principles and clinical impact. *Hematology* **12:**271-287.

Winfield, J. B. and Mimura, T. (1992). Pathogenetic significance of anti-lymphocyte autoantibodies in systemic lupus erythematosus. *Clin Immunol Immunopathol* **63:**13-16.

Wood, P. (2009). Primary antibody deficiency syndromes. *Ann Clin Biochem* **46:**99-108.

Wright, J. J., Wagner, D. K., Blaese, R. M., Hagengruber, C., Waldmann, T. A. and Fleisher, T. A. (1990). Characterization of common variable immunodeficiency: identification of a subset of patients with distinctive immunophenotypic and clinical features. *Blood* **76:**2046-2051.

Yan, M., Wang, H., Chan, B., Roose-Girma, M., Erickson, S., Baker, T., Tumas, D. *et al.* (2001). Activation and accumulation of B cells in TACI-deficient mice. *Nat Immunol* **2:**638-643.

Yong, P. L., Orange, J. S. and Sullivan, K. E. (2010). Pediatric common variable immunodeficiency: Immunologic and phenotypic associations with switched memory B cells. *Pediatr Allergy Immunol*. **21:**852-858.