# Clinical, Genetic and Molecular Aspects of Membranous Nephropathy

Horia Constantin Stănescu

University College London

PhD thesis

I, Horia Constantin Stanescu confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

## Acknowledgements

It is a great pleasure to thank the many people who made this thesis possible.

First of all, I would like to express my gratitude to my Ph. D. supervisor, Professor Robert Kleta, for his constant support, patience and friendship.

My thanks to my friend and mentor, Dr. Mauricio Arcos Burgos, who led my first steps in gene hunting.

I would like to thank Professor Nicholas Wood, my secondary supervisor, for his support and good advice.

My thanks to Professor Nina Scribanu, Professor William Gahl and Professor Maximilian Muenke, whose support allowed me to pursue my scientific path.

My thanks to all the friends and colleagues with whom I worked in our group: Sabina Domene, Guida Landoure, Riko Klootwijk, Anselm Zdebik, Detlef Bockenhauer, whose attitude and ideas stimulated me all along. A special mention amongst them to Alan Medlar for tea, walks, help with computer programming, editing and proofreading of the thesis.

My thanks to Mike Erdos, Mike Hubank and Mario Milani, for our interesting discussions about science and for their open and friendly advice.

I am grateful to Adina Bucur and Catalin Voinescu for their generous help with the computer programming and their patient and constant assistance. I am also grateful to Alex Vintilescu for helping me with the computer assisted graphic design.

My special thanks to Waldi Horer. His genuine passion for science has influenced my attitude towards research in a crucial way.

My thanks to Violeta Vintilescu, for her personal support and advice, and for the patience to proofread my English.

None of this could have been possible without my parents, Elena Lucia and Aurel Stanescu, to whom I express my warmest thanks.

Finally, I wish to thank all those whom I met along this path and whom I have not named here, but who made my pursuit interesting and meaningful.

Abstract

Membranous Nephropathy (MN) is one of the leading causes of end-stage renal disease (ESRD). MN is an autoimmune disease in which autoantibodies target antigens at the level of the glomerular basement membrane. The nature of these antibodies and the reason why they develop are not fully understood.

One of the strategies towards a better understanding of the disorder is genetic analysis, of which two approaches have been attempted: linkage mapping, based on a family suggestive for X-linked transmission of the MN trait; and whole genome association mapping, based on three case-control cohorts. The first cohort (335 cases and ethnically matched controls from the UK) was genotyped using SNP markers and analysed in an exploratory study which led to the identification of two highly significant loci of association. Two cohorts (146 biopsy proven MN cases and ethnically matched controls from the Dutch research group in Nijmegen and 75 biopsy proven cases and ethnically matched controls from the French research group in Paris) were used to successfully replicate the results.

The two loci which we identified and independently confirmed are located on chromosome 2 and on chromosome 6.

The chromosome 2 locus includes the PLA2R gene, confirming the hypothesis of Beck et al. which identified PLA2R as a key antigen in idiopathic MN by using an immunological approach [1].

The chromosome 6 locus lies within the extended Human Leukocyte Antigene (HLA) system locus, with the highest significance for association reached by alleles of HLA-DQA1.

Our results suggest that the susceptibility to membranous nephropathy is associated to genetic variants at the level of both PLA2R1 and HLA loci. The causative variants could be some of the polymorphisms captured by the genotyping array which was analysed or, more likely variants (single nucleotide or copy number variant type) situated nearby (and therefore in linkage disequilibrium).

TABLE OF CONTENTS

# LIST OF FIGURES

## LIST OF CLASSIFICATIONS



## LIST OF TABLES

# 1 Membranous nephropathy

## *1.1 Historical introduction*

> "Superficially, it might be said that the function of the kidneys is to make urine, but in a more considered view one can say that the kidneys make the stuff of philosophy itself"
>
> Homer W Smith

Since early history, urine, being readily accessible to investigation, has offered an opportunity for insights in the understanding of states of health and disease [2]. Its quantitative (*e.g.* polyuria – diabetes; Ebers Papyrus [3]) or qualitative (*e.g.* sweet urine – diabetes; Willis [4], possibly Ayurveda [5]), variations have always been considered to be important signs of disease.

Hippocrates was the first to describe the association of 'foaming urine' (which nowadays is considered a sign of proteinuria [6]) and renal disease in the context of the humoral theory of disease:

> "When bubbles settle on the surface of the urine, they indicate disease of the kidneys, and that the complaint will be protracted."
>
> (Aphorisms: VII, 34 [7])

Renal dropsy (which is a contraction of hy*dropsy*, oedema of renal cause) was described by Saliceto (Gulielmus de) in the XIII$^{th}$ century [8]. The first clear description of proteinuria (precipitation of urine with heat or acetic acid) happened in the XVII$^{th}$ century (Dekkers Frederik, 1673) [9]. Around the same period (1674) the association of oedema with proteinuria was noted by Domenico Cotugno / Cotunnius [10].

It was Richard Bright who defined the relation between dropsy, coagulable urine and renal disease [11], combination which was called *morbus brightii* (Bright's disease). The variable pathologic changes of the kidneys which became evident at *postmortem* investigations suggested many underlying causes for this syndrome.

The division of Bright's disease in two broad groups, based on pathological findings, led to the conclusion that the disorders of the kidney can be of inflammatory cause (nephritis) – in which glomerular lesions were predominant; as well as of non-inflammatory nature (nephrosis) – in which the tubular lesions

were considered to be prevailing. The concept of nephrosis has been introduced in 1905 by Friedrich von Muller specifically to convey the distinction between inflammatory and noninflammatory ('degenerative') pathologic processes of the kidney [12]. It survives today only in the name of the nephrotic syndrome [13].
The term nephrotic syndrome started to be used in the 1920-30s and was defined by Bradley and Tyson [14] as being:

> "…one of the most striking phenomena of renal disease. The combination of gross edema, hypoproteinemia, hypercholesterolemia, lipidemia and heavy proteinuria, in the absence of congestive heart failure, is unique and easily recognised."

A proportion of patients with nephrotic syndrome were noticed to have minimal glomerular changes, or even normal glomeruli. The cases with non-significant glomerular changes (with no cardiovascular causes of the nephrotic syndrome apparent *i.e.* no hypertension or cardiac hypertrophy) were grouped by Munk (1913) under the term lipoid nephrosis [15]. Due to the lack of glomerular morphological changes it was initially thought that 'nephrosis' was a predominantly tubular disorder.
E. T. Bell (1929) used Mallory's anilin blue (Heidenhain, azan-carmine) connective tissue stain (McGregor's technique [16]) to delineate changes in the glomerular basement membrane in some of the cases of lipoid nephrosis *latu senso* [17]. He suggested in 1938 a clinical pathological classification of glomerulopathies (dividing them in acute versus subacute and chronic) and described the membranous type of glomerular lesion under the name diffuse (as opposed to focal) membranous glomerulitis / glomerulonephritis (in opposition to the proliferative form) [18]. Later he included this entity together with lipoid nephrosis *strictu senso* (what is now referred to as minimal change disease), lobular glomerulonephritis (today's membranoproliferative glomerulonephritis) and chronic glomerulonephritis (today's focal and segmental glomerulosclerosis) within the Ellis type II 'nephritis' clinical category [19] (which is characterized by an insidious onset, with marked albuminuria and oedema) [20].
Anecdotally: the first mentioning in PubMed of the term membranous glomerulonephritis dates from 1955 [21].
In 1957 David Jones applied the special staining method that now bears his name (periodic acid-silver methenamine) to kidney biopsy specimens, thus delineating the specific / defining features of the anatomopathological lesion [22]. The

alternative name ‘membranous nephropathy’ was suggested: the term ‘pathy’ as opposed to ‘nephritis’ reflects the fact that the morphological changes at the level of the glomerulus are not associated with inflammatory changes [23].

The techniques of immunofluorescence [24] and electron microscopy [25], applied to the kidney in the 1950s-1960s, allowed histopathological observation at increased resolution, allowing for further refinement of the classification of glomerulopathies according to the identified morphological patterns [26].

Electron microscopy also permitted a better understanding of the ultrastructure and pathology of the glomerular basal membrane, although the existence of this important structure was known from the days of optical microscopy (the first to introduce the term basal membrane in 1857 were Todd and Bowman [27]).

## *1.2 Definition / Etymology*

Membranous nephropathy (*nephros* means kidney in Greek and *patein* suggests it being a disorder of non-inflammatory nature) is a glomerular disease, the term membranous referring to the morphological aspect of the glomerular basement membrane in this disorder [28]. Membranous nephropathy is therefore a pathologic diagnosis.

## *1.3 Histology*

The conspicuous target structure in membranous nephropathy, *i.e.* the glomerular basement membrane, is a central element of the glomerular filtration barrier. The glomerular filtration barrier is a selectively permeable structure consisting of three layers (from blood to urine): glomerular (fenestrated) endothelium, glomerular basement membrane and the interdigitated foot processes of podocytes (glomerular visceral epithelial cells which are interconnected by a multiprotein structure called slit diaphragm), see Figure 1.

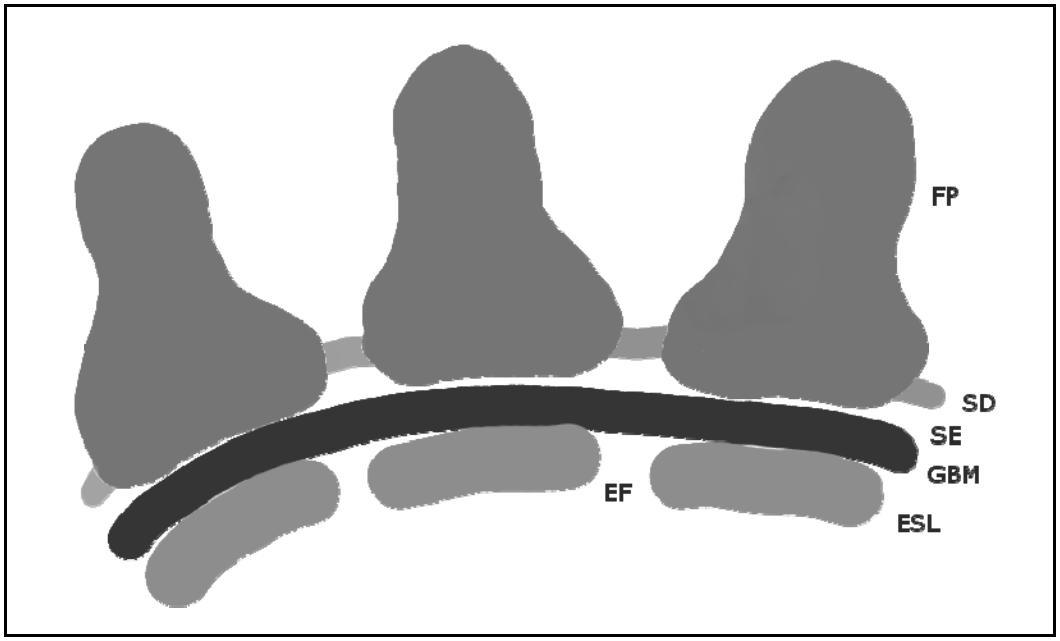


**Figure 1 Schematic ultrastructure of the glomerular filtration barrier**
FP – (podocyte) foot processes, SE – subepithelial space, SD – slit diaphragm, GBM – glomerular basement membrane, ESL – endothelial surface layer, EF – endothelial *fenestra*.

Alterations of the glomerular filtration barrier lead to the loss of selectivity of the structure, with passage of macromolecules in the urine, leading to the clinical manifestation which is the nephrotic syndrome [29].

## *1.4 Histopathology*

Membranous nephropathy is characterized histopathologically by three components [20]: the thickening of the glomerular basement membrane [30], finely granular immune deposits in the subepithelial space at the base of the foot processes of the podocytes (considered a pathognomonic aspect of membranous nephropathy) [31] and the presence of immunoglobulin G (IgG) in the aforementioned electrondense deposits [32].

The thickening of the glomerular basal membrane is thought to be the result of the interplay between deposition of immune deposits and synthesis of additional matrix material.

Through immunofluorescence microscopy one can visualise the immune deposits themselves by using stainings for total immunoglobulin G (IgG) or IgG4 whereas newly synthesized matrix material can be highlighted by the methenamine silver

/ Periodic acid-Schiff (PAS) staining for basement membranes (Jones' stain) [22].

Morphologically one can distinguish the four stages of membranous nephropathy that were defined by Ehrenreich and Churg [33,34]:

1. The first (early) stage shows only minimal changes: glomeruli appear almost normal by light microscopy, the capillary lumens are dilated, the thickening of the capillary wall is not detectable (this feature becomes visible by silver staining). Mesangial expansion and proliferation are usually absent. Immunofluorescence shows finely granular or pseudo-linear IgG deposits along the glomerular capillary wall. Electronomicroscopically small layers of electrondense sharply demarcated immune deposits with focal and segmental distribution can be observed on the epithelial face of the glomerular basement membrane. Projections of the basement membrane ('spikes') are small or absent. Podocyte foot processes show effacement are a constant feature of this stage.
2. In the second (fully developed) stage the glomerular enlargement and diffuse thickening of the capillary wall become visible on light microscopy. Mesangial expansion and proliferation are present in some cases. By silver staining, the extensions of the glomerular basement membrane appear black, being argyrophilic and therefore opaque to light transmission. These 'spike-like indentations' are the ones first described by David Jones in 1957 [22]. The interposing immune deposits appear lucent under light microscopy. Transmission Electron Microscopy (TEM) shows the image reversed (in a photographically negative manner): one can observe large, numerous, sharply demarcated electrondense immune deposits separated by the 'spike' like extrusions which have the same electron density as the basement membrane (representing the newly synthesized extracellular matrix material that has been added to the glomerular basement membrane). The spikes have the tendency to encircle the immune deposits. Foot process effacement is diffuse.
3. In the third (advanced) stage the more severe thickening of the glomerular basement membrane is visible by light microscopy. The capillary wall has a mottled appearance. Some cases present with mesangial proliferation and

sclerosis. By silver staining immune deposits appear like holes in the glomerular basement membrane (being translucent in contrast to the surrounding opaque argyrophilic material). By electron microscopy large and diffuse deposits, all along the glomerular basement membrane, appear as being engulfed by the extracellular matrix spikes. Their appearance is highly variable: coarsely or finely granular, with their electron density being greater, equal or less than that of the basement membrane material. The basement membrane appears thick and irregular.

4. In the fourth (late) stage the deposits are be incorporated in glomerular basement membrane, which is thickened by the apposition of newly synthesized matrix. Glomerulosclerosis (segmental or global) becomes evident by light microscopy.

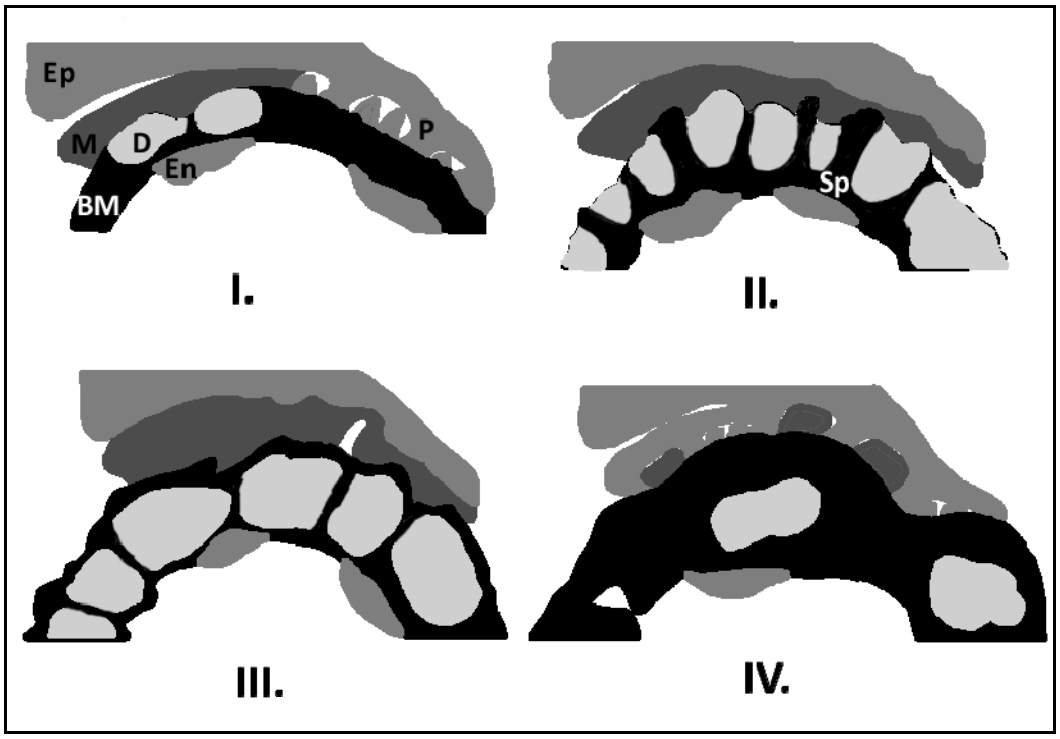


**Figure 2 Pathological stages of membranous nephropathy**
**Electron microscopically aspects of the glomerular basement membrane in the four pathological stages of membranous nephropathy; BM = basement membrane ; D = (*intra* BM) deposit ; M = medium electron-dense linear subepithelial deposit ; En = endothelium ; Ep = epithelium ; P = podocyte ; Sp = spike [35].**

The pathological aspects observed in the renal interstitium (macrophagic infiltration, overproduction of matrix by fibroblasts, eventually leading to tubulointerstitial fibrosis) are thought to be secondary – due to the toxic effect

proteinuria has on the tubular and interstitial cells (through plasma iron and proteins like the C5b-9 complex, lipid derived chemotactic factors, etc.).

Based on the pathology some variants of membranous nephropathy have been described such as:

- Membranous nephropathy with spherules [36],
- Segmental membranous nephropathy [37] (as opposed to the 'usual' form which is global [38]),
- Childhood membranous nephropathy with anti-tubular basement membrane nephritis [39].

In some instances secondary forms of membranous nephropathy (mostly lupus nephritis) show features that distinguish them from the idiopathic ones. Idiopathic membranous nephropathy is exclusively subepithelial whereas lupus nephritis presents with mesangial and subendothelial immune deposits. Lupus and malignancy associated membranous nephropathies also present with different immunoglobulin classes (predominantly IgG2 and IgG3 as opposed to the predominant IgG4 class which predominates in the idiopathic form). At the electronomicroscopical resolution level, lupus nephropathies sometimes present with tubuloreticular inclusions in the glomerular endothelium.

## *1.5 Clinical aspects*

Membranous nephropathy has an insidious onset, usually with (subclinical) proteinuria which leads later to peripheral oedema. It can be diagnosed at any age but its age distribution shows two peaks at 30 to 40 and 50 to 60 years. Males are more often affected than females (with the sex ratio being 2-3:1).

Of all the individuals affected by membranous nephropathy, 20% stay asymptomatic with non-nephrotic levels of proteinuria (< 3.5 g/day) and 80% become clinically manifest, presenting a full blown nephrotic syndrome with:

- proteinuria (> 3.5 g/day)
- hypoalbuminemia
- hyperlipidemia
- fluid retention (oedema)

Microscopic hematuria is present in about half of the adult patients afflicted with membranous nephropathy, being considered a marker of active disease. Macroscopic hematuria is very unusual.
Hypertension is also an uncommon finding in membranous nephropathy (30%).
It is estimated that about 40% of membranous nephropathy patients are at increased risk for thromboembolic phenomena (renal vein thrombosis).

## *1.6 Epidemiology*

Membranous nephropathy is a relatively rare disease [40] despite the fact of it being the main cause of idiopathic nephrotic syndrome in adults (about one third of all cases of nephrotic syndrome, even more so – up to half in older white adults [28]).
Children are rarely affected by membranous nephropathy: 2 - 15% (average of 5%) of pediatric patients with nephrotic syndrome show lesions reminiscent of membranous nephropathy at renal biopsy.
A certain degree of geographic variability has also been reported, with a higher proportion of membranous nephropathy determining nephrotic syndrome in Greece and Macedonia and a lower proportion in the United Kingdom [28].

## *1.7 Etiology*

The deposition of immune complexes on the epithelial side of the glomerular basement membrane appears to be triggered by a number of agents in genetically susceptible individuals.
In one third of the cases membranous nephropathy is considered to be secondary to (or has at least been associated with) infections (hepatitis B virus, malaria, syphilis, schistosomiasis etc.), drugs (non-steroid antiinflammatory drugs, penicillamine, gold salts etc.), toxins, chronical immunological disorders (class V lupus nephritis, rheumatoid arthritis, autoimmune thyroid disorders, Sjogren's

syndrome etc. [40]), allografts (kidney, bone marrow [41]) and malignancies. Even though potential antigens have been identified in secondary membranous nephropathy there is no definitive proof of their pathogenicity.

In about two thirds of the cases of membranous nephropathy the etiology remains unknown – this subset of membranous nephropathy cases being referred to as idiopathic membranous nephropathy - IMN).

The current paradigm asserts that idiopathic membranous nephropathy is autoimmune in nature. This hypothesis stems from the identification, in 1976, of possible antigens in what was then called 'immune complex nephritis' [42,43]. Autoantigens have been described for a rare form of congenital membranous nephropathy (neutral endopeptidase, NEP, initial case report [44], more cases [45]) as well as for the common adult form of IMN (M-type phospholipase A2 receptor, PLA2R) [1].

## *1.8 Pathogeny*

The first one to suggest that renal diseases can be caused by immune reactions was Schick in 1907 [46]. The autoimmune hypothesis in renal pathology gathered renewed support when in 1949 Lange *et al.* were able to demonstrate the presence of antibodies against kidney tissue emulsion in a group of patients with chronic glomerulonephritis [47].

### 1.8.1 Heymann nephritis

A pathophysiological model for human membranous nephropathy was established on the basis of an experimental rat model which was developed in the '50s by Walter Heyman (and has therefore been called Heymann nephritis) [48]. Heymann nephritis is induced by the immunization of rats with a mixture of antigens derived from rat proximal tubular brush border. These rats develop proteinuria with the histological features of human membranous nephropathy [49].

The triggering antigen in Heymann nephritis is a complex: HNAC (Heymann Nephritis Antigen Complex) [50,51] formed by megalin (gp330, LRP2 – Low density lipoprotein-related protein 2) and RAP (Receptor Associated Protein, LRPAP1 – Low density lipoprotein receptor-related protein associated protein 1). Megalin, the first protein to be described as a pathogenic antigen of Heymann

nephritis [52], is an endocytic receptor of the low-density lipoprotein receptor family which is located in the coated pits of the soles of rat podocyte foot processes and in the proximal tubular brush border [53]. Megalin and RAP form multimeric complexes in the endoplasmic reticulum, where RAP acts as a chaperone to assist the folding of megalin [54], complexes which are then trafficked to the cell surface [55]. At least five pathogenic epitopes, thought to lead to the formation of immune deposits, have been identified on the Heyman Nephritis Antigen Complex. One of them is located on RAP [56], the other four on megalin [57].

### 1.8.2 Experimental membranous nephropathy pathomechanisms

It is believed that the antigen is capped by autoantibodies in the coated pits on the soles of the podocyte. Subsequently the antibody-antigen complexes are shed on the glomerular basement membrane where they become stabilized in the *lamina rara externa* by cross-linking, leading to the (*in situ*) formation of subepithelial deposits [58].

The next step in the pathogenetic scenario is the lesion of glomerular epithelial cells, which is mediated by the complement membrane attack complex (MAC) [59]. Kerjaschki *et al.* showed that in Heymann Nephritis antibodies activated complement leading to C5b-9 (MAC) formation, its insertion in the podocyte [60] and assembly in the plasma membrane. Once MAC are assembled in the plasma membrane they can lead to the disruption of its permeable selectivity through formation of transmembrane channels or the rearrangement of membrane lipids with loss of membrane integrity [61]. As opposed to erythrocytes, nucleated cells (including glomerular epithelial cells) require higher doses of C5b-9 in order to be lysed, possibly through their ability to eliminate the complement channels / lesions and through a better osmoregulatory capacity [62]. Lower doses of MAC induces a sublethal (sublytic) response of nucleated cells [63] which appears to be due to the activation of specific signalling pathways. These pathogenic signals (calcium influx, transactivation of receptor tyrosine kinases, activation of phospholipase A2 with its consequences on arachidonic acid / prostanoid metabolism and induction of ER stress response, transcription factors, reactive oxygen species, enzymes, growth factors and

signals for proliferations, extracellular matrix) may lead to disruptions of the structure or functions of the targeted cells [61].

The autoimmune mechanism in which podocyte-membrane proteins are the autoantigenic targets was confirmed in different animal models of experimental membranous nephropathy: in mice aminopeptidase A [64] and in rabbits dipeptidyl peptidase IV and enkephalinase (neutral endopeptidase) [65] were shown to be the respective targets.

### 1.8.3 Human pathomechanisms

#### 1.8.3.1 Immune deposit formation

One important question is whether the immune deposits on the epithelial surface of the glomerular basement membrane are circulating immune complexes that were dissociated on the endothelial side, then passed through the membrane and reassociated on the epithelial side (due to their small size and cationic charge) or whether the immune deposits were formed locally (*in situ* complex formation) by antibodies binding circulating antigens that have been previously 'planted' on the endothelial surface or by antibodies binding intrinsic target antigens. In the case of the Heymann nephritis the latter hypothesis (antibodies binding *in situ* to an intrinsic target antigen) has been proven through experiments involving isolated perfused rat kidneys [66], but human megalin (glycoprotein 330, LRP2 low-density lipoprotein - related protein 2) has not been detected in subepithelial immune deposits in patients with membranous nephropathy, furthermore it appears not to be expressed in human podocytes [67].

#### 1.8.3.2 Glomerular injury

As in the case of the experimental nephritis of the rat, the lesion of the podocytes is thought to be mediated by the C5b-9 (MAC, membrane attack complex) multimer. Even though IgG4, the major component of the immune deposits in idiopathic membranous nephropathy, is believed to be anti-inflammatory not pro-inflammatory (not binding and therefore not activating complement) [68], complement somehow gets activated at the site of the lesion. The C5 component of complement is cleaved, C5a (a chemotactic factor) is supposedly flushed

downstream by the filtration forces at the glomerular barrier, therefore not moving back across the glomerular basement membrane to attract inflammatory cells from the circulation, thus explaining the paucicellularity of the lesion. The C5b fragment together with the C6 complement component form the so-called 'lipophilic complex' which inserts itself in the membrane of the podocyte. Once in the membrane, the lipophilic complex recruits the fractions C7, C8 and C9 to form the membrane attack complex / pore forming complex. The podocyte endocytes the pore forming complexes and transports them intracellularly in multivesicular bodies, to be later extruded on the urinary side. In the process, even though the podocytes are not lysed they are activated, leading to a cascade of signal transduction pathways which further lead to the production of potentially nephritogenic molecules: oxidants, proteases, cytokines, growth factors, vasoactive molecules and extracellular matrix. The increased expression of TGF (Transforming Growth Factors; TGFβ2, TGFβ3) and TGF receptors by the podocytes leads to the increased synthesis of matrix molecules which leads to the thickening of the glomerular basement membrane between the immune deposits (forming the characteristic argyrophilic 'spikes', later to become 'clubs', that are made visible through silver methenamine staining). In time the extracellular matrix synthesised *de novo* grows around and ultimately embeds the initial immune deposits. The podocyte number decreases through apoptosis, the remaining cells detach due to the expansion of the glomerular volume secondary to the increased intraglomerular pressure. Some areas of the glomerular basement membrane are denuded because of the limited ability of podocytes to proliferate (cyclin kinase inhibitors seem to be overexpressed leading to cell cycle arrest).

#### 1.8.3.3 Interstitium

It has been suggested that the interstitium suffers secondary in glomerulopathies, the mechanism being common to all unselective proteinuric disorders. Unselectivelly filtered through the glomerular barrier, plasma proteins (C5b-9 complex), lipid derived chemotactic factors and iron may have a toxic effect on tubular cells. This toxic effect exerts itself through macrophagic infiltration and overproduction of matrix by fibroblasts, leading eventually to tubulointerstitial fibrosis.

### 1.8.3.4 Alloimmune antenatal MN mechanism

Debiec *et al.* [44] described a family in which the mother, deficient in neutral endopeptidase (NEP), was immunized anti NEP during a miscarriage (the father being NEP positive). Circulating anti-NEP antibodies crossed the placenta in a subsequent pregnancy, leading to alloimmune antenatal membranous nephropathy. The clearance of the maternal anti-NEP antibodies led to the resolving of the disease in the child. It has been observed that in order for proteinuria to occur IgG1 anti-NEP antibodies are necessary: children of mothers with only the IgG4 subclass of anti-NEP antibodies are not affected [45].

#### *Neutral endopeptidase (NEP)*

The gene encoding *NEP* is located on chromosome 3q25.2 [69], its approved name (HGNC – HUGO Gene Nomenclature Committee [70]) is membrane metallo-endopeptidase (*MME*), alternative names are: *CD10*, and *CALLA* (Common Acute Lymphoblastic Leukemia Antigen). The product of the gene is also called neprilysin (renal endopeptidase, kidney 'enkephalinase', enzyme code E.C.3.4.24.11). It is a 100-kD type II transmembrane glycoprotein which is highly expressed in kidney (in the brush border of proximal tubules and on the glomerular epithelium). The enzyme is a neutral endopeptidase that cleaves peptides at the amino side of hydrophobic residues, thereby inactivating peptide hormones (e.g. glucagon, enkephalins, substance P, neurotensin, oxytocin, and bradykinin). It is worth mentioning that NEP has been shown to have an inflammation modulating effect [71].

### 1.8.3.5 Postnatal autoimmune mechanisms

#### *M-type phospholipase A2 receptor (PLA2R)*

Anti-PLA2R (M-type phospholipase A2 receptor) antibodies have been found in an initial report by Beck *et al.* (2009) [1] in the sera of 75% of idiopathic membranous nephropathy patients but not in that of secondary membranous nephropathy, other glomerular or autoimmune disorders or normal controls [40]. Anti-PLA2R antibodies only detect PLA2R in its non-reduced state, suggesting the dependence of the epitope(s) conformation on disulfide bonds. This also

might explain why the antigen has not been found before, despite the fact that similar techniques have been used [40].

PLA2R1 is a transmembrane protein member of the mannose receptor family [72]. The mammalian mannose receptor family is a member of the C-type Lectin superfamily [73], one of the families of membrane bound pattern recognition receptors [74].

**Pattern recognition receptors**

Pattern recognition receptors are important elements in the mechanism of innate immunity. The 'first line' immunological cells that encounter potentially threatening pathogens, professional antigen-presenting cells (APCs: macrophages, dendritic cells and B cells) are equipped with receptors which role is to recognize foreign antigens. Whilst B cells rely on a set of somatically variable receptors (surface immunoglobulins) to discriminate between a large variety of possible epitopes, macrophages and dendritic cells are reduced to fixed germline-encoded receptors to recognize simpler patterns of antigenic determinants. These fixed receptors are complement, Fc receptors and a large array of pattern recognition receptors (PRRs) [75]. PRRs recognize pathogen-associated molecular patterns (PAMPs) – highly conserved structures present in large groups of microorganisms (peptidoglycan PGN, lipopolysaccharides LPS, lipoteichoic acid LTA). PRRs can be classified in soluble / humoral (mannan-binding lectin MBL, collectins, ficolins, pentraxins and complement) and cell-bound receptors. The cell-bound receptors are either intracellular (NOD-like receptors NLRs, RIG-I-like receptors RLRs and cytosolic DNA receptors) or transmembrane (Toll-like receptors and C-type lectin receptors) [76].

Glycans (mono-, oligo- or polysaccharides and their conjugates, present at cell surfaces, incorporated in extracellular matrix or secreted - attached to glycoproteins) serve structural as well as functional roles (cell-cell and cell-matrix recognition, signalling etc.). Lectins or lectin receptors are glycan-binding proteins (which do not transform the saccharides and which have a nonimmune origin [77]) which participate to the non-structural roles of sugars. Their sugar-binding activity can be ascribed to modules referred to as carbohydrate-recognition domains (CRDs). Using the arrangement of CRDs as a

differentiating factor one can distinguish many families of lectin receptors, which have different cellular localizations [78], see Classification 1.

| | | | |
|---|---|---|---|
| IC | cytoplasmatic | calnexin (ER) | |
| | | M-type (ER) | |
| | | F-box (cytoplasmic) | |
| | | Selectins | L-type (ER, ERGIC, Golgi) |
| | | | P-type (secretory pathway) |
| | | | E-type (?) |
| | ctplsm & mmbr | R-type (Golgi, mmbr) | |
| | mmbr | I-type lectins / siglecs (mmbr) | |
| EC & IC | ctplsm | S-type lectins / galectins (ctplsm, EC) | |
| | mmbr | ficolins (mmbr, EC) | |
| | | C-type including collectins (mmbr, EC) | |
| | | interlectins (mmbr, EC) | |
| | | SUEL-related lectins (mucus, mmbr) [79] | |
| EC | chitinase-like lectins (EC) | | |
| | F-type (EC) | | |
| | collectins (EC?) | | |
| | lectin-like pentraxins [80] | | |

**Classification (1) Lectin (receptor) families**
Cellular localization: IC = intracellular; EC = extracellular; ER = endoplasmic reticulum; ERGIC = endoplasmic reticulum golgi intermediate compartment; ctplsm = cytoplasmatic ; mmbr = membrane.

Arguably the most important role of lectins is that of recognition molecules within the immune system [81].

*C-type lectin receptors*

C-type lectin receptors are a superfamily of proteins encountered only in pluricellular animals (metazoans). The domain which characterizes the superfamily is the C-type lectin-like domain (CTLD). CTLDs are characterized by a double loop (‘loop-in-a-loop’) which is stabilized by two disulfide bridges at its base. Vertebrate proteins that contain CTDLs (CTLDcps = CTDL containing

proteins) are classified based on their overall domain structure in 17 groups (families) [73] see Classification 2.

I –lecticans,
II – the ASGR group,
III – collectins,
IV – selectins,
V – NK receptors,
VI – the macrophage mannose receptor group,
VII – Reg proteins,
VIII – the chondrolectin group,
IX – the tetranectin group,
X – polycystin 1,
XI – attractin,
XII – EMBP,
XIII – DGCR2,
XIV – the thrombomodulin group,
XV – Bimlec,
XVI –SEEC,
XVII – CBCP

**Classification (2) of vertebrate proteins containing C-type lectin domains**
ASGR = AsialoGlycoprotein Receptor; Reg = regenerating; EMBP = Eosinophil Major Basic Protein; DGCR = DiGeorge syndrom Critical Region; Bimlec = uncharacterized lectin (which had a descriptive name assigned); SEEC = group of soluble proteins containig SCP, EGF (twice) and CTLD domains; CBCP = Calxβ and CTLD containing Protein; CTLD = C-type lectin-like domain.

C-type lectins have been shown to be expressed by antigen presenting cells (APCs) and to play a role in the innate immune system [82].

*The (macrophage) mannose receptor family*

The (macrophage) mannose receptor family, the VI$^{th}$ structural group of CTLDcps, encompasses, besides PLA2R, the mannose receptor (MRC1), the endocytic receptor Endo180 (formerly named so, now known as MRC2 – mannose receptor C type2), DEC205 (its accepted Human Gene Name

Committee name being LY75 - lymphocyte antigen 75) and the avian (yolk sac) Ig Y receptor (which is PLA1R) [72].

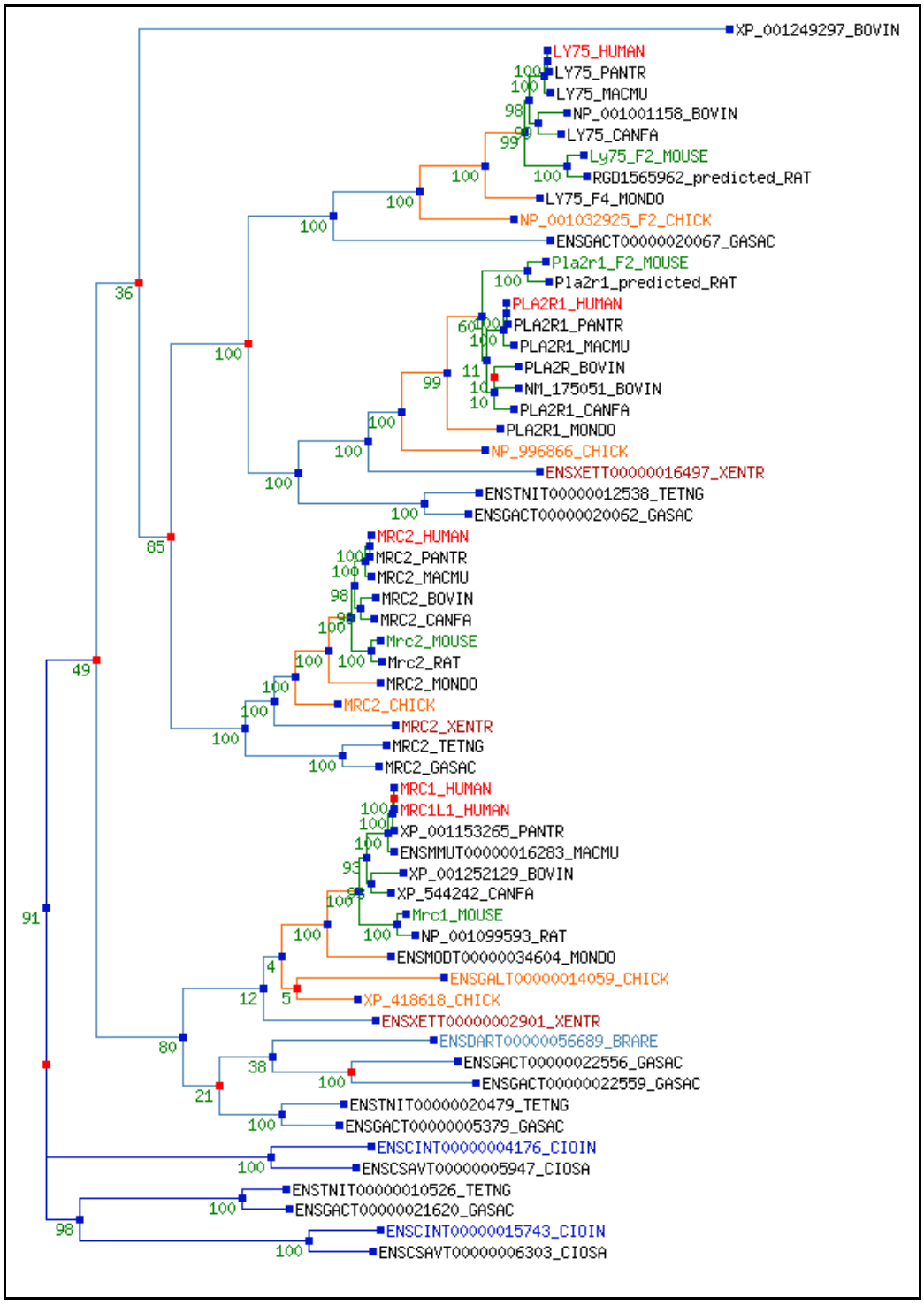


**Figure 3 Mannose receptor gene family**
**Phylogenetic tree obtained with TreeFam [83,84]. BOVIN = *Bos taurus*; PANTR = *Pan troglodytes*; MACMU = *Macaca mulatta*; CANFA = *Canis familiaris*; MONDO = *Monodelphis domestica*; CHICK = *Gallus gallus*; GASAC = *Gasterosteus aculeatus*; XENTR = *Xenopus tropicalis*; TETNG = *Tetraodon nigroviridis*; BRARE = *Danio rerio*; CIOIN = *Ciona intestinalis*; CIOSA = *Ciona savignyi*.**

All the members of the family are type I transmembrane proteins with a large extracellular segment which is arranged in a similar way: at the N-terminal a cystine rich domain (CysR), followed towards the membrane surface by a fibronectin type II domain (FNII) and 8 (respectively 10 in the case of DEC205) C-type lectin like domains (CTLDs) / carbohydrate recognition domains (CRDs). It has been shown for MR and Endo180 that only one of the CTLDs retains lectin activity (CTLD2 in the case of Endo180 and CTLD4 in the case of MR). Different conditions could alter the availability of the binding sites and thus the function of these proteins. Single particle electron microscopy studies have suggested a model for the three-dimensional conformation of members of the mannose receptor family. In this model the N-terminal end is 'bent' towards the 'active/functional CTLD of the receptor, thus generating a globular head. The compact folded conformation seems to occur under certain conditions (changes in pH, association with ligand binding etc.), this appearing to be a general property of the family [85].

The residues required for calcium-dependent sugar binding in the mannose receptor are not conserved in the PLA2R CTLDs, denoting the lack of calcium dependent lectin activity of the receptor [72].

The intracellular domain contains a motif which has been originally identified in the LDL (low density lipoprotein) receptor – to mediate coated-pit mediated endocytosis (Asn-Pro-Xxx-Tyr) [86].

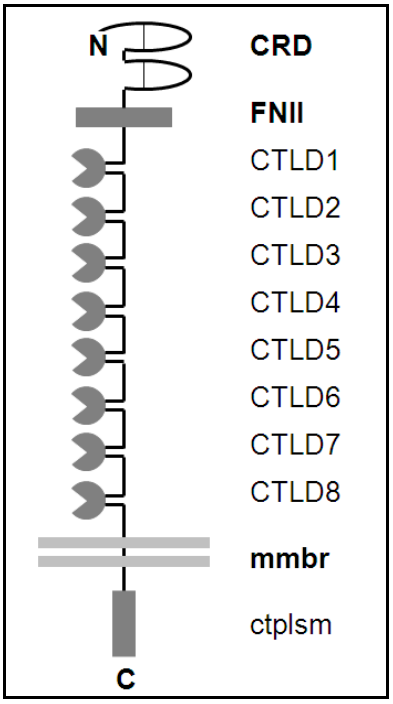


**Figure 4 Domain structure of Phospholipase A2 Receptor (PLA2R1)**
N – extracellular N terminal end ; CRD - cystine rich domain (thin vertical lines: S-S bounds); FNII - fibronectin type II domain; CTLD - C-type lectin like domains; mmbr – transmembrane domain; ctplsm – intracellular domain; C – intracellular C terminal end.

## Phospholipases A2

Phospholipases A2 (PLA2) are a very diverse family (superfamily) of lipolytic esterases (mammals having least 19 distinct members [87]). The enzymatic function of PLA2 implies hydrolization at the *sn*-2 position of membrane phospholipids, which leads *in vivo* to the release of lysophospholipids and polyunsaturated fatty acids which are further transformed in bioactive eicosanoids with various biological functions [88].

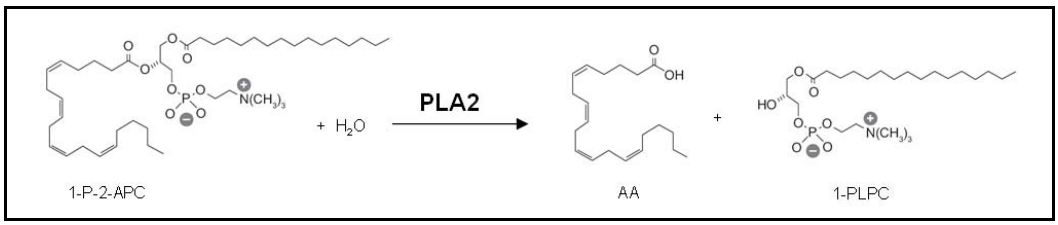


**Figure 5 Hydrolysis reaction catalyzed by Phospholipase A2 (PLA2)**
Modified after [89]; 1-P-2-APC = 1-Palmitoyl-2-arachidonyl-phosphatidylcholine; AA = Arachidonic acid; 1-PLPC = 1-Palmitoyl-lysophosphatidylcholine.

One can distinguish secreted (sPLA2s) and intracellular PLA2s (cytosolic cPLA2s, calcium-independent and platelet activating factor (PAF) acetyl hydrolase / oxidized lipid lipoprotein associated) [90]. First to be characterized were the toxic secreted PLA2s in the venom of bees [91] and snakes [92]. It is worth noting that PLA2 beside having an enzymatic / toxic action, is the major allergenic component of bee venom [93].

*Secreted Phospholipase A2 (sPLA2)*

The human genome contains ten sPLA2 genes. In order to perform its biochemical function hydrophilic sPLA2 has to adsorb directly to the hydrophobic substrate membrane interface. Pancreatic sPLA2s have a role in dietary phospholipid digestion, cPLA2s have a role in eicosanoid generation through arachidonic acid release. The functions of many of the sPLA2 proteins are unknown, it is not clear if they modulate eicosanoid production, some of their functions may even be unrelated to the enzymatic capabilities as suggested by the fact that group XIIB sPLA2-like lacks lipolytic activity - due to a polymorphism at the catalytic residue. Group IIA sPLA2 accumulate to high levels during inflammation. A large body of evidence suggests that sPLA2 and cPLA2 cooperate in releasing arachidonic acid. The molecular mechanism of this cooperation is yet to be described, although it has been suggested that sPLA2 action on phospholipids takes place – in the secretory compartment, before the enzyme leaves the cell. It is still a matter of debate whether the (probably eicosanoid mediated) effect of sPLA2 is proinflammatory or antiinflammatory [94].

A better established role for sPLA2s is the one in host defence through the innate immunity component of the immune system. The enzyme seems to be expressed at locations where it can strategically contribute to anti-microorganism defence – potential ports of entry (tears, seminal plasma, Paneth cells and intestinal lumen, inflammatory exudates, bronchoalveolar lavage, serum, prostatic epithelial cells and their secretions etc). Potential production of soluble PLA2R in the human kidneys has been suggested, based on the existence of an alternatively spliced transcript at this level [95].

The bactericidal activity of sPLA2 against Gram-negative bacteria necessitates the presence of serum permeability increasing protein. sPLA2s are not able by

themselves to penetrate the lipopolysaccharide outer layer of the envelope of Gram-negative bacteria. In order to be penetrated, the envelope has to be previously disrupted by the permeability increasing protein or the MAC of complement [96]. The anti Gram-positive bactericidal activity of sPLA2 is independent of serum factors. It was demonstrated in the context Paneth cells secretion (cryptidins and sPLA2) which contributes to the low content in microorganisms of the small intestine, despite their repeated oral entry [97]. As a matter of fact it has been proven that sPLA2 is the principal bactericidal component against Gram-positive bacteria in human tears [98].

The mechanism of the bactericidal effect of sPLA2 is based on the electrostatic attraction between the basic enzyme and the acidic components of the surface of Gram-positive bacteria (lipoteichoic acids) and subsequent hydrolization of the phospholipid components of the bacterial membrane [99].

Given that sPLA2s in the venom of insects and snakes are very similar to mammalian sPLAs it is likely that the latter would be toxic if left unchecked. One of the mechanisms to control sPLA2 function is its inhibition and degradation *via* binding to specific binding proteins [89]. Two types of receptors were identified (using as a 'bait' a neurotoxic sPLA2 from the snake venom) [100]. N-type (Neuronal) receptors are diverse in nature and bind to neurotoxic sPLA2s. Their structure and functions are not known and it has been suggested that the nomenclature 'receptor' would be inappropriate for the binding proteins of neurotoxic sPLA2s [95].

The M-type (Muscular) receptor is PLA2R1. PLA2R1 was first cloned from rabbit [101]. It is upregulated by inflammatory stimuli [102]. Mammalian sPLA2s are divided in two groups with respect to their affinity to PLA2R1: some bind to it with low nanomolar affinity, some do not bind at all [103]. The ones that bind are inhibited by the binding. The receptor has two isoforms, a soluble one and a membrane bound one. The membrane bound receptor internalizes after binding the enzyme, leading to its degradation. Therefore a role has been suggested for the receptor in controlling the activity of sPLA2 through its clearance after it has been secreted [100]. Other proteins of the C-type lectin superfamily have been shown to bind and inhibit sPLA2s, most remarkably snakes seem to be inhibiting the potential self toxic effect of their own venom through Plasmin alpha Inhibitors (PLI-α; which are homologous to C-type

lectins) [104]. The diversity of sPLA2s binding proteins suggests that sPLA2s do not only function through their enzymatic but also through mechanisms independent of their PLA2 activity [105].

*Catalytically independent functions of sPLA2*

In the context of renal pathology, two interesting catalytically independent functions of sPLA2 are worth mentioning: Mukherjee *et al.* described how uteroglobin, inhibiting sPLA2 activity and thus decreasing the level of lysophosphatidic acid (LPA), may prevent the activation of integrins which enhance tissue deposition of fibronectin (FN) in the glomeruli of mice (in a model of familial FN-deposit glomerulopathy) [106]; Birts *et al.* [107] described cell debris removal facilitation by GIIA human PLA2 (PLA2GIIA).

*Debris removal*

PLA2GIIA was first described in the synovial fluid of patients affected by rheumatoid arthritis [108]. It has been shown to rise during sepsis, to be associated with inflammatory disease, being recognized as an acute phase protein [109]. It exerts an antibacterial function due to its high affinity for bacterial membrane anionic phospholipids but has a very low affinity for the plasma membrane of normal host cells (due to the zwitterionic nature of their outer phospholipid monolayer). Only the cell membranes of cells that have been activated during the immune response and of cells that initiated apoptosis are hydrolyzed [110], it has been therefore suggested that the enzyme plays a role in the clearance of damaged cells [111]. The mechanism of the biological response induced by the binding of sPLA2GIIA to the cell surface is not well understood. sPLA2GIIA does not have high affinity cell surface receptors, it has been shown to bind to surface heparan sulfate proteoglycan (HSPG). Binding does not require enzyme catalysis and is followed by energy dependent uptake and trafficking of sPLA2GIIA to the cell nucleus. The result is significant swelling of the target cells, characteristic of macropinocytosis, which may play a role in the clearance of pathological anionic cell debris. By doing so sPLA2GIIA seems to have a resolving, primary antiinflammatory role in the course of the inflammatory response [107].

*Secretoglobin / Uteroglobin (UGB)*

The many synonyms of secretoglobin family 1A member 1 (SCGB1A1) [112] (uteroglobin, blastokinin, uteroglobin-like antigen / protein, Clara cell secretory protein / CCSP, polychlorinated biphenyl-binding protein etc. [113]), reflect either localisation or functional aspects clarified by each of the naming authorities. SCGB1A1 is a secreted protein inducible by steroids [114] which has been deemed one of the most powerful endogenous anti-inflammatory and immune-modulating ('natural immunosuppressor' [115]) agents [116], containing a representative peptide of a class called antiinflammins [117]. It is produced by the secretory epithelia of organs that communicate with the external environment [116]. It is expressed in the uterus (hence its name), thymus, pituitary gland, the respiratory tract, the gastrointestinal tract (including the pancreas), the mammary gland, the prostate and seminal vesicles [118]. It also has been shown to be present in the blood and in urine (urine protein-1 [119]) although it seems not to be synthesized in the kidney [106]. The surface structures of (rabbit) uteroglobin and that of sPLA2 are virtually superimposable [120]. Uteroglobin is a strong inhibitor of sPLA2 activity [121]. Uteroglobin has also been shown to be a sex dependent marker or renal (tubular or glomerular) dysfunction even though its gene (SCGB1 A1, secretoglobin, family 1A, member1) is encoded on an autosome (chromosome 11q12.3) [122]. Mutations of uteroglobin in mice leads to severe renal disease due to massive deposition of fibronectin (FN1) at the level of the glomeruli [123]. Extracellular matrix assembly through fibronectin self-aggregation (fibrillogenesis) relies on the binding of fibronectin (and activation of) receptor-like integrin heterodimers (particularly $\alpha 1\beta 5$ integrin) [124]. This process has been shown to be promoted by lysophosphatidic acid, which is generated by PLA2. It has been shown in mice that uteroglobin is able to prevent fibronectin deposition at the level of the glomerulus by counteracting fibronectin self-aggregation through its high affinity binding of fibronectin, which forms fibronectin-uteroglobin heteromers [106]. In humans, familial glomerular nephritis with fibronectin deposits (GFND, [OMIM 601894]) [125] seems to have a similar underlying pathogenetic mechanism,

even though mutations of fibronectin [126] and not of uteroglobin [127] have been implied in its etiology.

- Amyloid
- Non-amyloid - Ig-derived
  - cryoglobulinemia (type 2)
  - MIDD
  - GP in monoclonal γ-pathies
  - SLE (Systemic *Lupus Erythematosus*)
  - Immunotactoid fibrillary GP
- non-immune complex mediated / [128]
  - fibronectin GP
  - diabetic GP

**Classification (3) of fibrillary glomerulopathies**
MIDD - Monoclonal Immune Deposition Disease; GP - glomerulopathy.

Besides its role in non-immuno complex mediated diseases, uteroglobin seems to be implicated as a protective factor in IgA nephropathy (IGAN) at least in mice [129], the same role in humans being debated with arguments for [130] and against [131]. The mechanism of this protective effect has been suggested to imply UGB competing with IgA for the heparin-binding site of FN thus preventing the FN-IgA complex formation, a normal process [132] which seems to be enhanced in IGAN [133].

The mechanisms of sPLA2 receptor interaction at the cell membrane are little known. One of the better characterized sPLA-2 binding proteins is the M-type receptor or PLA2R1.

### M-type PLA2 Receptor

The M-type PLA2 receptor exists in two isoforms: membrane bound and secreted [134]. At least a part of the secreted isoform seems to be produced by cleaving of the membrane bound receptor by metalloproteinases (similarly to the mannose receptor [135]). PLA2R has been shown to be expressed in rats in the kidney (glomerular mesangial cells[136]), in the lung (on alveolar type II epithelial cells), in the vessels (rat vascular smooth muscle cells and vascular

endothelial cells), in the cartilage (synovial cells and chondrocytes) on a subset of splenic lymphocytes and on neutrophils [136]. The expression of PLA2R in humans has been documented for lung, kidney (it has been shown to be abundant on the podocytes of normal human kidneys [1]) and pancreas [137].

The contact between sPLA2 and its receptor takes place at the opening of the catalytic site of sPLA2. It has been shown that blockers of the binding of sPLA2s to the M-type receptor inhibits sPLA2 activity [138]. On the receptor part, the domains responsible for binding are the CRD-like domains [139], CRD3 to CRD5 being sufficient in conferring sPLA2-IB binding activity. As mentioned before PLA2R has no calcium dependent lectin activity, therefore sPLA2 binding to the receptor is calcium independent. *N*-linked sugar residues at the 15 potential *N*-glycosylation sites of PLA2R seem to be essential for the optimum recognition of sPLA2-IB (in mouse and bovine cells) [139].

PLA2s which have been shown in mice to bind to PLA2R are part of groups IB, IIA (with lower affinity: 6-10X lower [103]) and X. (apparently conserved positions of 16 cysteine residues account for this binding specificity). PLA2R ligands (group IB and X. sPLA2) are synthesized as inactive pro-enzymes, with a propeptide sequence at the $NH_2$ terminals [140]. The conversion from the pro-form to the mature form through proteolytic cleavage of the propeptide seems to be essential for optimum binding to the receptor [141]. The activation of the ligands is catalyzed by proteolytic enzymes (trypsin, plasmin and proprotein convertases) [142].

There is a high sequence identity of the PLA2 receptors among various mammals (rabbit, mouse, bovine and human) [143]. Human ligands of PLA2R are not well defined. PLA2R binds sPLA2 group IB (sPLA2-IB and PLA2R being also co-expressed in human in the pancreas, lung, and kidney), apparently doesn't bind group IIA and it is not known whether it binds group X sPLAs.

The receptor appears to be involved in the regulation of sPLA2 function, in both senses (stimulatory and inhibitory). The inhibitory function seems to be elicited through endocytosis internalization of its ligands by the membrane bound isoform as well as through inhibitory binding by the circulating soluble form. Group X sPLA2 enzymes possess very potent hydrolizing activity towards phosphatidylcholine, against which the protective mechanism of clearance *via* PLA2R mediated endocytosis seems to be particularly important.

The stimulatory function seems to be a biological consequence of PLA2R occupation by ligands, which leads to induction of signal transduction systems. As opposed to other enzyme ligands (*e.g.* proteinase activated receptors), where the enzymatic activity is required for ligand recognition and induction of the biological function, the enzymatic activity of sPLA2-IB is suppressed by binding to PLA2R.

Some of the effects induced by the PLA2-IB / PLA2R system are: growth promotion / proliferation, chemokinetic migration / cell invasion into the extracellular matrix, secretagogue effects in endocrine organs (release of progesterone in the *corpora lutea* of the rat, co-secretion of insulin in the rat pancreatic islets); effects mediated through production of lipid mediators (thromboxane, arachidonic acid, prostaglandins): contractile responses (guinea pig lung parenchyma) [95].

The source of the soluble isoform of PLA2R is somewhat obscure. It has been proposed to be the liver, and its production by the human kidney was also suggested by an alternatively processed transcript [137] – but this has not been confirmed.

Other members of the C-type lectin superfamily (*e.g.* SP-A surfactant protein A [144]) seem to be candidates for sPLA2 binding proteins through similar CRDs.

PLA2R has been shown to be activated by C5b-9 [145-149]. Its activation is supposed to lead to the expression of stress proteins [150] which are important in the glomerular epithelial cell lesion mechanism [151,152].

**Anti-PLA2R antibodies**

Anti-PLA2R antibodies in membranous nephropathy are of the IgG immunoglobulins subclass 4 [153,154]. They have been shown to be highly specific (100%) and sensitive (>75%) for idiopathic membranous nephropathy. Possible reasons for the sensitivity not reaching 100% are:

- technical issues of the immunoblot assay
- heterogeneity - additional antigens (*vide infra*)

- misclassification of patients with nephrotic idiopathic membranous nephropathy (as opposed to them having an unrecognized secondary cause)
- absence of immunologic disease activity at the time of the sampling

The last possibility is considered by Beck *et al.* to be the leading hypothesis of due to the fact that anti-PLA2R antibodies have been shown to be present in patients initially nephrotic due to idiopathic membranous nephropathy, absent in many of the cases of patients undergoing remission and present again in the cases of relapse.

### *Other glomerular antigens*

Other targets of the autoantibodies in idiopathic membranous nephropathy, such as intracellular glomerular proteins [155] have been hypothesized. It has also been suggested that low-levels of anti-NEP antibodies are needed for the pathogenesis in adults with IMN [156].

Recently two autoantigens (AKR1B1 and SOD2) which are deemed possible targets in IMN have been identified through a combined proteomic approach by Prunotto *et al.* [157]. Autoantibodies from sera of patients with membranous nephropathy have been blotted on two-dimensional (2D) electrophoresis gels which were used to separate human podocyte membrane protein extracts. The proteins which were bound by autoantibodies were afterwards characterized by MALDI-TOF (matrix-assisted laser desorption/ionization time of flight) mass spectrometry. The initial screen yielded a set of proteins that had been targeted by autoantibodies in MN sera: SOD2, AR, alpha-enolase, secernin-1, ubiquitin hydrolase and Ser/Thr protein phosphatase 2A. AR and SOD2 were later confirmed by Western blot, and their renal expression was established by immunofluorescence (IF) and immune electron microscopy (IEM). Prunotto *et al.* suggest that neo-expression (neither of them are expressed in glomeruli under physiological conditions – but might be induced under oxidative stress conditions) of these two autoantigens might be a continuation mechanism of the inaugural event which is represented by deposition of anti-NEP, anti-PLA2R etc. antibodies.

**Aldose reductase**

Aldose reductase(AR) / aldo-keto reductase family 1 member B1 (AKR1B1) is a member of the aldo-keto reductase family, which includes more than 40 members. This particular member catalyzes NADPH-dependent reduction of a number of aldehydes, including the aldehyde form of glucose into sorbitol (being implicated in the development of diabetic complications) and the aldehyde phospholipids (generated through the oxidation of unsaturated fatty acids) [158].

**Manganese superoxide dismutase**

Superoxid dismutase 3 (SOD2) is a member of the iron / manganese superoxid dismutase family, important component of the endogenous antioxidant cell protective system. Manganese SOD2 catalyses the toxic superoxide by-products of oxidative phosphorylation to hydrogen peroxide and diatomic oxygen.

### 1.8.4 Similar / related disorders

#### 1.8.4.1 Association of MN with other autoimmune disorders

Rare (sometimes just one case report in the medical literature) associations of membranous nephropathy with immunological [159] and autoimmune disorders have been reported: rheumatoid arthritis [160], celiac disease [161], ankylosing spondilitis [162], bullous pemphigoid [163], scleroderma [164], Crohn's disease [165], Sjogren syndrome [166], autoimmune thyroiditis [167].

#### 1.8.4.2 Basement Membrane Pathology

A specific role of basement membrane lesions in human pathology was first established in 1990, with the identification of collagen mutations in Alport syndrome [168,169].

*Renal basement membrane pathology*

Pathological entities in which the basement membrane is affected have been classified in genetically mediated and acquired basement membrane disorders [170].

Genetically mediated disorders are considered to be mainly structural defects, a proportion of the acquired basement membrane disorders are diseases with an (auto)immune component (in which the immune response is oriented against the

basement membrane) and there is an overlap category (in which the structural defects / variants of components the basement membrane could lead to the loss of tolerance and immune response oriented against themselves).

**Structural defects**

The major protein components of all basement membranes are laminin, type IV collagen, entactin / nidogen and sulfated proteoglycans [171], the main proteoglycan of the glomerular basement membrane is perlecan [172]. Mutations of laminin lead to Pierson syndrome (microcoria – congenital nephrotic syndrome; [OMIM #609049]). Mutations in type IV collagen lead to: Alport syndrome [173,174] (which is geneticaly heterogeneous, with all forms resulting from collagen type IV mutations and three described transmission modes: X linked [OMIM #301050] 85%, autosomal recessive [OMIM #203780] 15% and autosomal dominant [OMIM #104200] rare), thin basement membrane syndrome (TBMD, Benign Familial Hematuria BFH [OMIM #141200], clinically similar to Alport syndrome but milder) and Hereditary Angiopathy with Nephropathy, Aneurysms and Muscle Cramps syndrome (HANAC, [OMIM #611773]). Mutations of entactin / nidogen have not been associated to any phenotype and perlecan has not been shown to lead to a kidney phenotype (its mutations are associated with Silverman Handmaker [OMIM #224410] and Schwartz Jampel syndrome [OMIM #255800] type dwarfisms).

Mutations of the transcription factor LMX1B (LIM homeodomain 1B) lead to nail-patella syndrome [OMIM #161200] through the dysregulation of the synthesis of glomerular basement membrane collagen [175].

**Autoimmune Diseases targeting Basement Membranes**

The targets of autoimmune responses against basement membranes [176] include the skin basement membrane – epidermis junction (Bullous Pemphigoid – BP, epidermolysis bullosa – EB), the lung basement membrane and the glomerular basement membrane (GBM) – both affected in the case of Goodpasture Syndrome (GPS, also called Anti GBM disease; an autoimmune disease of lung and kidney, [OMIM 233450]) [177].

*Anti GBM disease (Goodpasture Syndrome)*

In the model of Goodpasture Syndrome [178] host factors have been theoretized to be immune response genes. This has been suggested by familial instances of the disease [179], including a case report of a pair of identical twins that were exposed to gasoline and turpentine for two weeks and five days before the onset of the disorder [180].

MHC class II locus has been deemed to be an important modulator of the immune response in GPS. Associations have been published between the inflammatory response of GPS in humans and HLA-DR and DQ alleles [181,182].

Due to limited sample sizes and strong linkage disequilibrium between DR and DQ alleles in humans it has been difficult to differentiate the attribution of susceptibility between the two regions. In a study in mice the susceptibility to anti-GBM disease was linked to the Aβ/Aα region in the H-$2^s$ strain which corresponds to the human HLA-DQ region [183].

Anti GBM disease has been shown in rare cases to overlap with MN (17 cases in the literature review by Patel *et al.*) [184]. This rare immune related glomerular condition might suggest the existence of a (partial) common pathogenic pathway between the two disorders. One possibility mentioned by Patel *et al.* is the altering of the glomerular basement membrane by the intra- and epimembranous immune complexes of membranous nephropathy followed by release of glomerular basement material (normal or altered) which may lead to the development of an anti-GBM response.

### 1.8.4.3 Autoimmune kidney diseases

The anatomical – clinical classification of kidney disorders (*vide infra*) is unsatisfactory from an etiological pathogenetical point of view. The group of primarily autoimmune kidney diseases includes anti-GBM disease (Goodpasture's syndrome), IgA nephritis (IGAN), membranous nephropathy and membranoproliferative glomerulonephritis [185]. Two of them where not treated in the context of basement membrane pathology and are mentioned for

completeness and possible parallels / overlaps of their respective pathogenic mechanisms.

### *IGAN*

IgA nephropathy, also called Berger's disease [186], has been shown to be an autoimmune disorder in which the autoantigen is the abnormally glycosylated IgA class 1 antibody [187]. Genetic studies [188]: a genome wide association analysis showed a strong signal of association in the MHC region on chromosome 6p, consistent with the autoimmune hypothesis [189].
Some evidence from the literature may suggest the existence of UGB (uteroglobin), FN (fibronectin) and IgA connections [130,132,133,190]. A review of the medical literature also reveals that membranous nephropathy and IgA nephropathy do not tend to overlap, on the contrary membranous nephropathy seems to coincide with instances of IgA deficiency [191].

### *Membranoproliferative glomerulonephritis*

Membranoproliferative glomerulonephritis [192] is an class of inflammatory glomerulopathies (hence the name glomerulonephritis) which are grouped together on the basis of similar pathologic glomerular changes suggesting common pathogenetic mechanisms [193]. It is characterised by glomerular mesangial expansion (due to both increased matrix and cellularity; hence the alternative name mesangioproliferative) and thickening of the capillary walls, due to accumulation of (subendothelial and intramembranous) immune deposits, cellular interposition and new basement membrane synthesis [194]. It is further split in two [195] (possibly three [196]) morphological subclasses:

I. classical (with subendothelial immune complex deposits, in which the classical complement pathway is activated),
II. rare (it is even disputed whether it is a real instance of membranoproliferative glomerulonephritis [197]) - dense deposit disease (with intra-membranous dense deposits, associated with C3NeF – C3 nephritic factor: a serum immunoglobulin which stabilizes C3bBb – C3 convertase leading to the persistent activation of the alternative complement pathway),

III. considered a subtype of class I (type I + additional subepithelial deposits)

The etiology of membranoproliferative glomerulonephritis has not been completely elucidated, but a strong link of former 'idiopathic' forms to hepatitis C virus (HCV) infection cryoglobulinemia was established [198]. The genetic component / predisposition can be inferred from familial occurrences [199,200]. Mesangioproliferative glomerulonephritis has also been associated with infected atrioventricular shunts, complement component deficiencies (C2, C3 and factor H) and partial lipodystrophy.

## *1.9 Kidney disorder classification*

Anatomically, membranous nephropathy is a glomerular (as opposed to tubular or interstitial) kidney disorder. Glomerular disorders are encompassed under the XIV$^{th}$ chapter (Diseases of the genitourinary system, categories N00 - N99) of the International Classification of Diseases version 10 (2007) [201], see classification 4.

N00 Acute nephritic syndrome

N01 Rapidly progressive nephritic syndrome

N02 Recurrent and persistent haematuria

N03 Chronic nephritic syndrome

N04 Nephrotic syndrome

N05 Unspecified nephritic syndrome

N06 Isolated proteinuria with specified morphological lesion

N07 Hereditary nephropathy, not elsewhere classified

N08 Glomerular disorders in diseases classified elsewhere.

.0 Minor glomerular abnormality (minimal change lesion)

.1 Focal and segmental glomerular lesions

.2 Diffuse membranous glomerulonephritis

.3 Diffuse mesangial proliferative glomerulonephritis

.4 Diffuse endocapillary proliferative glomerulonephritis

.5 Diffuse mesangiocapillary glomerulonephritis –
Membranoproliferative glomerulonephritis (types 1 and 3, NOS)

.6 Dense deposit disease - Membranoproliferative glomerulonephritis 2.

.7 Diffuse crescentic glomerulonephritis –
Extracapillary glomerulonephritis

.8 Other - Proliferative glomerulonephritis NOS

.9 Unspecified

**Classification (4) of glomerular diseases**
Glomerular diseases (with the exception of hypertensive renal disease) are classified, according to the World Health Organisation (WHO) International Classification of Diseases (ICD) criteria, in 8 three letter categories (N00 – N08; relating to clinical syndromes) according to mixed predominantly clinical (hematuria, proteinuria, nephritic syndrome, nephrotic syndrome) / temporal / evolutive (acute, rapidly progressive, recurrent, persistent, chronic) criteria: Each of the aforementioned categories have possible (fourth character) morphological subdivisions (which should not be used unless specifically identified – *e.g.* by renal biopsy or autopsy). NOS = not otherwise specified.

## *1.10 Genetics*

There are several arguments suggesting a strong genetic component in the etiology of MN. Familial aggregation has been described in the literature, and it is known that the variability of the immune molecules modulating the (auto)immune response is genetically encoded – as shown in the aforementioned anti-GBM paradigm.

Suggestive for the genetic control of the susceptibility for membranous nephropathy are also the numerous accounts of association of this disease status with major histocompatibility complex / human leucocyte antigen (immunologic and genetic) markers.

More or less successful small scale association studies have also been attempted based on a candidate gene approach.

### 1.10.1 Familial clustering

Taking in account the *caveat* that familial aggregation can be often considered a result of shared environment [202], there is still an argument for it suggesting a possible genetic susceptibility. Proof for familial aggregation [202-205], familial aggregation in the context of a geographical isolate [206], occurrence of membranous nephropathy in the case of siblings [207,208], HLA identical siblings [209,210] (in one case with different outcomes of the disease) [211], occurrence in the cases of twins [212,213], a possible syndromic (*ie* associated with deafness) variant of membranous nephropathy in the case of twins [214], and instances suggesting a possible mode of inheritance (X-linked) [215,216] have all been documented in the medical literature.

### 1.10.2 MN Human Leukocyte Antigen (HLA) Associations

#### 1.10.2.1 The human HLA complex

*History*

The existence of a locus controlling the delineation of self from non-self was first confirmed in mice (by P A Gorer and G D Snell in 1948 [217]). The molecular complex encoded at this locus was named the major histocompatibility complex

(MHC) [218]. The first human element of the MHC was described by Jean Dausset in 1958 (investigating sera from recipients of multiple transfusions) and was called MAC and later renamed HLA-A2 (Human Leucocyte Antigen) [219]. Soon association between HLA determined susceptibility and neoplasias were shown in mice (for the mouse HLA analogue H-2 in 1964 [220]) and in 1997 also in human patients. Noteworthy on the subject of glomerular pathology, the first non-neoplasic diseases shown to be associated to particular HLA antigens were glomerulonephritis (1969 [221]) and systemic lupus erythematosus (1971 [222]).

### *Structure and function*

HLA molecules present antigenic peptides to T lymphocytes in order to stimulate them during the adaptive immune response. All nucleated cells express HLA class I molecules, which present intracellular antigens (***e.g.*** viral antigens, after viruses entered the cell) to CD8 positive cytotoxic T lymphocytes (CTLs). HLA class II molecules stimulate CD4 positive helper T lymphocytes. These are activated by specialized antigen-presenting cells (APCs: macrophages, dendritic cells, B lymphocytes etc), which are the only cells expressing HLA class II molecules. Polymorphisms at the protein level were explored by serological typing methods until, with the discovery of PCR it became easier for the variability to be assessed by molecular typing at the DNA level. The initial naming of HLA 'entities' referring to serological types, was adapted to accommodate the nomenclature of molecularly defined alleles [223].

### *The human HLA genomic locus*

The human leucocyte antigen (HLA) genes are the most polymorphic genes in humans. They are stretching a genomic super-locus [224] of 4 Mb [225] (or, in the extended HLA – MHC locus 6 Mb [226]) on chromosome 6p21. Even though it is one of the best defined loci in the human genome, the HLA locus is also considered to be one of the most gene dense and complex ones [224]. Therefore it is difficult to distinguish causal variants from variants associated with it via linkage disequilibrium. Added to that, there are also technical difficulties ***e.g.*** the inability to amplify HLA sequence segments via PCR [227].

*Genetics of the R and Q components of the HLA-D region*

The classical D antigen HLA region (which encodes the subunits for the major MHC class II antigens: DP, DQ and DR) is located on the chromosome 6p21.3. DR and DQ loci are in situated in close vicinity, further away from the DP locus, see Figure 6.

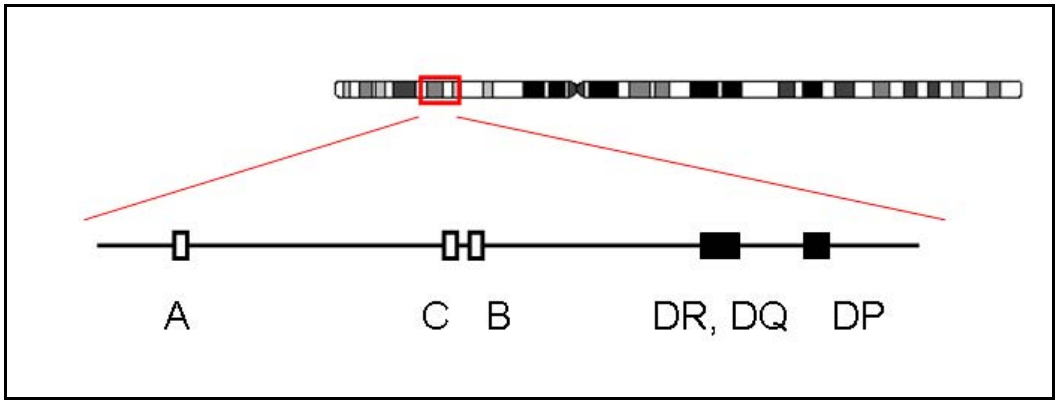


**Figure 6 HLA region on human chromosome 6**
Banded ideogram: chromosome 6; red box: extended HLA / MHC region; white boxes: class I antigen loci; black boxes: class II antigen loci.

The HLA-D locus is an extended haplotype encoding a 'superphenotype'. Two loci (DR and DQ, which are closely linked) contribute to the expression of this superantigen. Both DR and DQ complexes are heterodimeric (the units being alpha and beta) cell surface receptors.

### The HLA-DR locus

The HLA-DR locus was defined at the seventh Histocompatibility Testing Workshop (1977) and refined at the VIII$^{th}$ Histocompatibility Testing Workshop (1980) [228]. The DRA locus encodes the DRα chain of the HLA-DR dimer, the DRβ chain is encoded in a more complex way: although there are four possible loci, no more than three are present in any individual, no more than two on a single chromosome. The HLA-DRB1 locus is ubiquitous and very variable, HLA-DRB3, HLA-DRB4 and HLA-DRB5 have variable presence and variable variability in humans.

### The HLA-DQ locus

The DQα chain is encoded at the *DQA1* locus and the DQβ subunit is encoded by the *DQB1* gene. *DQB1* is more variable than *DQA1*.

*HLA-DQA1*

The HLA-DQ locus encompasses two sub-loci (A and B) which encode an heterodimeric HLA class II molecule. The two polypeptidic chains (monomers) which polymerize to form the dimer are called alpha (DQA, encoded on locus A) and beta (DQB, encoded on locus B). The monomers are both membrane anchored and form, by their coming together in the class II molecule its so called antigen presenting groove, see Figure 7.

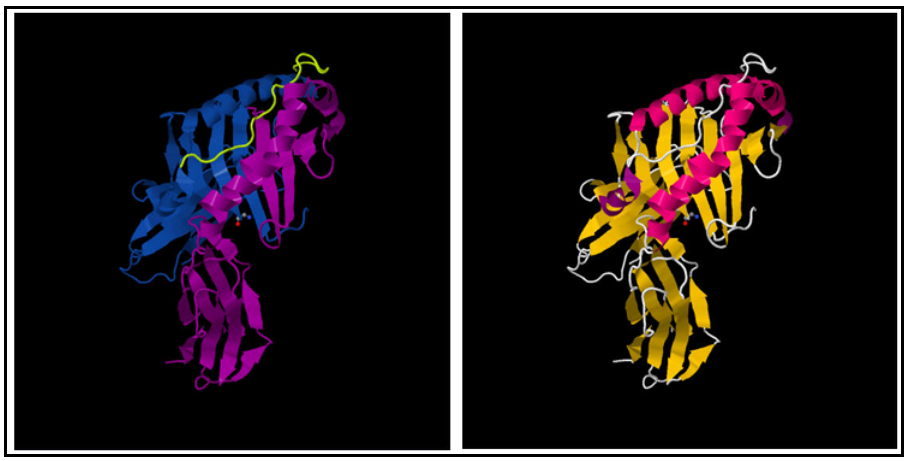

**Figure 7 HLA-DQ dimer structure**
Left: HLA-DQA1 (blue) / HLA-DB1 (magenta) dimer presenting an antigenic polypeptide (hypocretin,green) in the antigen presenting groove (PDB 1UVQ [229]; viewed in Jmol 12.0 [230]). Right: color coded map of the secondary structure of the same molecular complex. Colour key: alpha helix: red; beta sheet: yellow.

MHC class II molecules have a central role in the immune response. They are expressed in antigen presenting cells (APCs): macrophages, dendritic cells, B-lymphocytes etc. MHC class II molecules present the epitope peptides, which are 9 – 15 amino acid long and obtained by processing of extracellular antigenic proteins within APCs [231].
The majority of the polymorphisms within MHC class II molecules are located (on both the alpha and the beta chains) within the antigen presenting groove, being important in determining its peptide binding specificities. Sequence variants in HLA-DQA1 can thus change the shape / conformation of its antigen presenting groove and therefore its specificity for the immunogen (epitope) presentation.

Membranous nephropathy associations with MHC markers (initially serologic, later genetic) have been specifically investigated since the 1970's.

#### 1.10.2.2 Serotyping

The initial studies were performed using serological markers (by microlymphocytotoxicity – mlc – serotyping [232] of defined antigens). Associations to HLA-DR3 [233-235], HLA-DR3 and MT2 [236], HLA-DR3 and B8 [237-239], DR2 and MT1 (in the case of two Japanese [240,241]) and BF [242] were found. Later genetic markers were used (by genotyping of STS markers situated on the short arm of chromosome 6 which contains the HLA region). A mixed (serotyping and genotyping) approach led to contradictory results, identifying no association between membranous nephropathy and HLA-DR3 (DR-DQ) [243]. Different loci encode the molecules which correspond to the DR3 serotype.

#### 1.10.2.3 Genotyping

Genotyping identified associations between membranous nephropathy and DQA1 [244], DRB1 (and DQA1 – DQB1) [245], DR3 and DR5 [246], HLA-DR2 (in another Japanese study) [247], not C4 (considered to be a possible candidate gene, being associated with the DR3 haplotype in Caucasoids and DR2 haplotype in Japanese) in a Japanese cohort [248], Transporter associated with

Antigen Presentation (TAP1; located between HLA-DP and HLA-DQ genes, TAP dimers form a dimeric complex in the membrane of the endoplasmic reticulum, contributing to the ATP-dependent transport of processed epitope peptides from the cytosol to the endoplasmic reticulum lumen [249,250]; allegedly not in linkage disequilibrium [251] with HLA-DR3) [252], TAP1 and HLA-DM (the products of which participate in the process of loading of the epitope peptide on class II molecules [253,254]) [255] and Tumour Necrosis Factor (TNF; cytokine polymorphisms are considered good candidates in explaining the clinical variability of IMN) [256,257].

*Observations*

HLA-B8 or DR3 being very rare in Japanese the possibility has been raised of a IMG susceptibility gene being in population specific linkage disequilibrium with HLA-B8-DR3 in European Caucasoids respectively HLA-DR2 in Japanese [247].

Limitations of the typing methods made it difficult to assess whether the association of IMN with A1, B8, D33 (extended HLA haplotype in European Caucasoids) was due to an HLA-DR or -DQ locus, or both, or to another locus in linkage disequilibrium to HLA class II [258].

### 1.10.3 Small scale associations

After the association of membranous nephropathy to the HLA locus (and DR3/DR2 haplotype) was established, the search started for the causative gene(s) at this locus. Studied candidates were, as mentioned complement factor B (BF; human properdin factor, one of the components of the alternate complement pathway), complement factor 4 (C4), TAP1 and TNF.

Based on a similar strategy, non HLA association studies started to be driven by the selection of candidate genes, based on the then understanding of the structure and function of the glomerulus.

The results were contradictory, small scale association studies (but not genome wide association studies) are known to be plagued by false positive results [259].

#### 1.10.3.1 Nephrin

Nephrin (NPHS1) was considered to be a candidate based on its role in the structure of the glomerulus, which was discovered based on the fact that it was found to be associated with the congenital nephrotic syndrome of the Finnish type in genetic studies [260]. One single nucleotide polymorphism (SNP; rs437168) in NPHS1 was found to be associated (the frequency of the G allele was found to be significantly higher in the patient group when compared with the control group: $P = 0.014 < 0.05$; OR = 1.82; 95% Confidence Interval = 1.2 – 12.94) with the susceptibility to membranous nephropathy in a small (132 biopsy diagnosed patients) Taiwanese cohort [261].

#### 1.10.3.2 NO enzymes

The NO (nitric oxide metabolism) enzymes were considered meaningful candidates based on their physiological role in vasomotility and their role in the pathophysiology of progressive renal disorders. The association between polymorphisms in NO enzymes and diabetic nephropathy has been queried thoroughly [262,263]. A study employing SNP markers in ACE-D (Angiotensin Converting Enzyme D) and ecNOS (endothelial cell Nitric Oxide Synthase) [264] found no significant associations to the membranous nephropathy trait.

#### 1.10.3.3 Plasminogen activator inhibitor-1

Plasminogen activator inhibitor-1 (PAI1) was deemed interesting based on its role in renal fibrosis. A polymorphism within this gene was found to be associated ($P$ =0.026) with disease progression (renal deterioration) in patients with membranous nephropathy [265].

#### 1.10.3.4 Transient receptor potential cation channel 6

Transient receptor potential cation channel 6 (TRPC6) was considered a potential candidate due to its interaction with nephrin and its proven role in the etiopathogenesis of focal segmental glomerulosclerosis (FSGS) [266]. In a targeted association study, markers within the *TRPC6* gene were not found to be associated to membranous nephropathy [267].

#### 1.10.3.5 β3 integrin subunit

Integrins have been implicated in the pathogenesis of glomerulonephritis [268]. A study assessing markers within the gene encoding the β3 integrin subunit established no association to the membranous nephropathy trait [269].

#### 1.10.3.6 Vascular Endothelial Growth Factor

Expression of the Vascular Permeability Factor (VPF) / Vascular Endothelial Growth Factor (VEGF) in the human normal kidney is highest at the level of visceral epithelial layer of the glomerulus, suggesting a functional role at that level [270]. Its expression has been shown to be altered in glomerular diseases (*i.e.* decreased in sclerotic areas), including in the case of membranous nephropathy, suggesting a role in the progression of glomerulopathies [271,272]. Its association with chronic kidney disease, of which a subset is documented MN has been assessed. An association of a marker with the progression towards chronic kidney disease (CKD) has been documented ( $P = 0.009$) [273].

#### 1.10.3.7 Phospholipase A2 Receptor

Following the immunological investigations of Beck *et al.*, which showed that autoantibodies in a majority of membranous nephropathy cases are oriented against Phospholipase A2 Receptor 1 (PLA2R1), there have been studies investigating the genetic association between polymorphisms of the gene and the respective trait. Two targeted association studies, one Taiwanese [261] and one Korean [274], both confirmed associations of markers in PLA2R1 to the membranous nephropathy phenotype.

# 2 Methods of research

## *2.1 Genetic dissection*

### 2.1.1 Genetics

Genetics (term derived from genetic, adjective from the Greek *genesis* - *Γενεσιζ* birth, origin, creation, generation [275]) is the science of heredity. It is generally accepted that its use in the present sense of the term was suggested for the first time by W. Bateson (in a letter to A. Sedgwick, dated April 18, 1905 [276]). It is worth mentioning that J. B. S. Haldane raised the historical point of Titus Lucretius Caro referring to *genitalia corpora*:

> *"quippe ubi non essent genitalia corpora cuique*
> *qui possit mater rebus consistere certa?"*
>
> ("For if there were no factors governing birth,
> how could we tell who anyone's mother was?")
>
> (*De rerum natura, liber primus* 1.167-68 [277])

as being immuable entities which we could consider the equivalent of today's genes [278] (although, arguably, conjuring them into existence in a way that resorts to circular argumentation: proof of the assertion [*genitalia corpora* existing] resting on the assertion itself [because they exist they explain maternal resemblance] [279]).

### 2.1.2 Theories of heredity

W. Johansson emphasized the distinction between two types of the heredity theory, in his terms 'transmission conception' *vs.* the modern 'genotype conception':

> 'The *personal qualities* of any individual organism do not at all cause the qualities of its offspring; but the qualities of both ancestor and descendant are in quite the same manner determined by the nature of the "sexual substances" – i.e., the gametes – from which they have developed. Personal qualities are *then the reactions of the gametes* joining to form a zygote; but the nature of the gametes is not determined by the personal qualities of the parents or ancestors in question.'
>
> (Johansson, 1911, p130)

#### 2.1.2.1 Transmission conception

The transmission conception (the generally accepted pre-Mendel paradigm, embraced by 'biometricians', and consistent with Darwin's 'pangenesis' [280]) was based on a 'blending' heredity model in which the qualities of the parents were blended together in the zygote and which traces its historical roots to Hippocrates (the Hippocratic theory of heredity states that different parts of the body secrete substances which are gathered in the gonads [281]). The blending / transmission theory posed two problems, one of which was pointed out by one of the most pertinent critics of Darwin, Fleeming Jenkin [282,283]: continuous blending of different characters down the flow of generations would lead to a diminishing of the initial differences between them, with the result of vanishing variability. But variability is a fundamental prerequisite of the theory of evolution, therefore the problem of heredity was from the very beginning situated at the core of the most important debate in biology.

The second problem of the blending heredity theory refers to the fact that it supports better the Lamarckian [284] evolutionary theory (transmission of acquired traits) than it does the Darwinian (natural selection from a pool of existing variants). It is perhaps worth mentioning that the Lamarckian evolutionary theory seems to gain new favour in the light of newly discovered epigenetic mechanisms [285].

#### 2.1.2.2 Genotype conception

The 'genotype conception', initiated by Galton and Weissman, has at its centre point the concept of the discrete unit ('element' encoding a 'Merkmal') of heredity that was advocated by Mendel [286].

*Genotype as a set of discrete units*

The definition of the genotype concept evolved from the blending (Darwin) towards the discrete type of inheritance/heredity unit 'element' (Mendel). This discrete unit was later to be called 'gene' by Johanssen [287]. The understanding of the discrete nature of inheritance marked the beginning of a debate pertaining to the question of how discrete units can explain continuous traits / phenotypes.

This split the scientific establishment into 'Mendelians' (Bateson etc.) and 'Biometricians' (Pearson etc.) [288].

What seemed initially to be a strong argument against the theory relying on a discrete nature of hereditary elements was its apparent difficulty to explain quantitative (continuous) traits. R. A. Fisher's attempt to solve this difficulty (which led to the polygenes / Fisher-Falconer model – in which many discrete genes cumulatively explain a continuous phenotype [289]) was the starting point of the elaboration of the new evolutionary synthesis.

*Modern synthesis*

The debate was only solved through the elaboration of the modern synthesis of the evolutionary theory in the framework of the theoretical foundations of population genetics. The foundations of this synthetic evolutionary theory was laid in the 1920s and 1930s, when R. A. Fisher [290,291], J. B. S. Haldane and S. Wright developed mathematical theoretical accounts on natural selection as a genetic process [292]. The real synthesis, a Darwinian fusion of genetics (experimental genetics and theoretical population genetics) and natural history (with taxonomy at its core), was fashioned in a book written in 1937 by T. Dobzhansky (Genetics and the Origin of Species) [292,293]. The fact that E. Mayr could have challenged the mathematical school of population genetics (Fisher, Wright and Haldane) to having created what he called 'beanbag genetics' [278] shows how important Dobzhansky's 'biological translation' of the theory was.

The "unit factor" ('in the gamete') was named *gene* by W. Johanssen, who also introduced the terms *genotype* and *phenotype* [294], marking the difference between the potential for a trait and the trait itself [295].

### 2.1.3 Definition of terms

Albeit fundamental concepts of biology, the terms gene, genotype, genome, phenotypic character, phenotype, phenome are fluid, ill defined, used in different (sometimes contradictory) senses by different authors – appear to be in constant evolution / redefinition [296].

#### 2.1.3.1 Gene

The concept of the ‘gene’ itself has a ‘historical burden’ [297]– being defined differently in the ‘classical’ period of genetics (that is in relation to the trait it encodes for; the ‘instrumental gene’ [295]) as opposed to the recent ‘molecular’ period (which is in relation to its underlying physical entity *i.e.* stretch of DNA; the ‘material gene’ [295]).

#### 2.1.3.2 Genotype

The initial definition of genotype [298] referred to ‘the totality of genes in a gamete’ – closer to what would be today called a genome [299]. The evolution of the term led to its actual meaning – which is the ‘genetic composition’ of an individual (*lato sensu*), in practice used as genetic status at a given locus (which is *stricto sensu* a partial genotype).

#### 2.1.3.3 Phenotype

In a similar way, the concept of the general phenotype (better termed ‘phenome’ as defined by Lewontin in 1992 [296] ‘the actual physical manifestation of the organism, including its morphology, physiology and behavior.’) can be viewed as the totality of partial phenotypes [298].

In practice phenotype is usually used in the sense of what is strictly the partial phenotype / phenotypic character (defined by Suzuki *et al.* as ‘ the form taken by some character … in a specific individual’ or ‘the detectable outward manifestation of a specific genotype’ [296]).

The added difficulty in defining the phenotype derives from what has been called the ‘blueprint’ concept, in which the phenotype ‘…is the manifestation … of the interaction of this information [genotype] with … the environment in the broadest sense – that enable the blueprint to be realized.’ (Futuyuma [296]).

The blueprint concept has also been criticised for its not taking in account of the process of development [300], which is another layer of complexity to be added to the understanding of the already complicated genotype – phenotype relationship.

### 2.1.4 Genotype Phenotype Association

To dissect the genetic component of a trait means to explore the link / association between its phenotype and genotype [300]. We could think of it in the larger context of the theoretical task for (population) genetics: to set up a complete theory as a set af laws that predictably map genotypes to a phenotype space and phenotypes to a genotype space [301].

#### 2.1.4.1 Many to many mapping

The genotype phenotype map [302] represents not one-to-one but rather many-to-many nonlinear relationships (one genotype many phenotypes: pleiotropy; one phenotype many genotypes: genetic heterogeneity) [298].
The one-to-one simplified version of mapping is an instance of reductionistic approach in science which is opposed (philosophically, epistemiologically) by the holistic approach (which, at the limit, argues that mapping is just a theoretical construct, all relations being many to many) [303,304].

*Genotype / phenotype duality?*

In a metaphorical way one might suggest that genotype and phenotype are two facets of a fundamental hereditary unit in the same way elementary particules have been theorized to have a dual (corpuscular and wave) nature by L. De Broglie.

The 'pathogenetic tree' metaphor used by Rimoin *et al.* is also helpful in understanding the genotype – phenotype interrelation in the case of Mendelian (but also complex 'Galtonian') conditions. In this comparison the mutation is at (or veri close to! s*ic*) the level of the genotype and everything more remote represents successive layers of phenotype at increasing complexity (in the comparison: the root represents the mutation, the trunk represents the aberrant gene product, the branches are the pathogenetic pathways – different branches for different organ and tissue specific pathology – and the leaves are the detectable phenotypic features) [305].

### 2.1.4.2 Two steps (paradigm shifts) enabling genetic mapping

*The one trait – one chromosome paradigm (Mendel – Boveri – Sutton)*

A premise to genetic mapping was the establishing of the fact that the heredity 'particles' (factors, Mendel) are based on chromosomes. This is due to T. Boveri, whose experimental program launched in 1889 led to the demonstration that nuclear rather than cytoplasmic factors determine the character of the developing embryo [306,307], and to W. S. Sutton (who, in 1902, observed the parallelism between Mendel's laws and the segregation and assortment of grasshopper chromosomes during meiosis [308]).

In 1911 E. B. Wilson mapped the first (human, even mammal) gene (the one specific for colorblindness) to a specific chromosome (chromosome X): as 'revealed by the cytological evidence; for what I have called the "homogametic" and the "digametic" sexes obviously correspond respectively to the homozygous and the heterozygous conditions.' (p. 260) [309].

*The many traits on one chromosome paradigm (synteny, linkage)*

The fact that there are more traits than chromosomes leads to the possibility of linkage (see below). Linkage is an apparent contradiction of Mendel's law of independent segregation, which would only apply strictly if there would be a one-to-one correspondence between traits and chromosomes. Serendipitously, Mendel studied traits of the green pea (*Pisum sativum*) that were situated on different chromosomes [310]. In 1905, W. Bateson, E. Saunders and R. Punett, whilst studying the sweat pea (*Lathyrus odoratus*), observed the phenomenon of non compliance with Mendel's second law (due to the fact that the respective traits were on the same chromosome – as was to become clear much later), phenomenon which at that time they called coupling (or connection, as opposed to independent segregation which they termed repulsion) [311]. The occurrence of gene loci on the same chromosome has been named 'synteny' [312].

The Sutton-Boveri chromosomal theory was further refined by T. H. Morgan who was able to show (based on *Drosophila melanogaster* experiments) that 'genetic units' lie on linear chromosomes like 'beads on a string' [313].
This led to the practical possibility of mapping. A.H. Sturtevant proved the possibility to determine the *quasi* physical (that is genetic) location of a gene, establishing the succession of genes on the *Drosophila* chromosome, and introducing the concept of genetic distance as a measure of recombination probability between two loci. Genetic sistance is measured in [c]M ([centi]Morgans), in honor of Sturtevant's mentor [314].

## *2.2 Mapping (by linkage)*

The practical possibility to map traits to genotypes by measuring recombinations between markers in pedigrees was thus established.

### 2.2.1 Markers

The 'genetic units' assessed by Sturtevant were genetic markers. They were phenotypic proxies for genotypes. It became clear that the effect of a genotype associated to a trait can be assessed through the observation of an indirect entity associated to the genotype proper. Historically, the initial 'markers' were thus phenotypic markers (*e.g.* Sturtevant's construction of the *Drosophila* genetic map using phenotypic markers such as eye color etc. [314]).
A (genetic) marker is thus defined as a measurable variable which can provide information about allelic variation at a given locus [315].
In the light of the genotype-phenotype duality metaphor, it is understandable that early phenotypes had to be discrete in order to be mapped to discrete genotypes. The concept of a discrete phenotype has its historical roots with F. Galton [316], who believed that large discontinuous variations are at the driving force of evolution, as opposed to his cousin C. Darwin, who thought small – yet continuous – variations are the ones who enter the 'ratchet' of selection to lead to evolution [317].
W. Bateson, K. Pearson and W.F.R. Weldon were all students of Garrod. Bateson became one of the founders of the 'Mendelian' (discontinuous variation) school of thought in evolutionism whereas Pearson and Weldon founded the 'biometric' (continuous variation) school [318].

The collaborations of Bateson with clinicians, notedly E. Nettleship and A. Garrod led to phenotypes such as alkaptonuria to be mentioned as an example of human Mendelian recessive inheritance by Bateson, and might have been important in the crystallization of the chemical variant – genetic unit concept by Garrod [319]. Other discrete human (clinical) phenotypes ('inborn errors of metabolism') studied by Garrod and which paved the way of his cementing of the concept of chemical individuality were: albinism, cystinuria and pentosuria [320].

#### 2.2.1.1 Phenotypic / morphologic markers

The aforementioned traits / disorders were obvious phenotypic markers due to them being visible (albinism: white skin, alkaptonuria: black urine etc). Sex was also an obvious, and early studied 'phenotypical' marker, used as such to elucidate its own heredity [321].

One of the early clinical (microscopic, pathologic / morphological) phenotypical markers was thalassemia. The study of its transmission and geographical distribution led to J.B.S. Haldane's evolutionary explanation of the hererozygote advantage in the case of malaria susceptibility [322].

#### 2.2.1.2 Serological markers

*ABO*

The discovery of the blood groups by Landsteiner coincided with the year which is considered to be the moment of rediscovery of Mendelism [323]. At the core of this discovery was (again) the crucial question about (quantifiable *ergo* discrete) intraspecific variability:

> '… proteins in various animals … are different…' [do] '… individuals within a species show similar … differences.' [324]?

Landsteiner and Richter probably already realized the possibility that blood groups are genetically determined, Landsteiner and Levine suggesting in a later publication that they might be used in assessing paternity cases [325]. In 1910, von Dungern and Hirzfeld, studying 72 families with 102 children, clearly stated the A and B agglutinogens of the red blood cells reflected Mendelian factors (they suggested two independent dominant alleles) [326].

It was for Bernstein to later demonstrate the correct hypothesis (multiple alleles at a single locus) [326]. Thus it has been argued, due not to the least to the quantifiable discrete characteristics of its phenotype, the ABO system locus (of which we now know it is situated on chromosome 9q34 [327]) became the first *bona* fide human marker locus [328]: due to the one-to-one correspondence of phenotype and genotype in the case of blood groups, it soon became evident that serotyping can become a proxy for genotyping of the respective locus.

Ludwig and Hanka Hirzfeld used the ABO marker in a study associating different blood groups to different ethnic groups (establishing that blood type A appeared to be more common in northern and central Europe and blood type B more common in eastern Europe, Africa and Asia) [329].

It was R. A. Fisher who suggested that the use of serological markers in the quest for detecting gene products…

> "… is going to lead to a greater advance (…) in the problems of human genetics than can be expected from any further work on biometrical or genealogical lines." [330,331]

### *HLA / MHC serology*

The study of Human Leucocyte Antigens started with experiments pertaining to tumor biology in mice (showing that the susceptibility of transplantable tumors is inheritable *ergo* genetically determined; Little [332] and Tyzzer, 1916 [333]).

At the suggestion of J. B. S. Haldane, in the 1930s, P. A. Gorer was studying the possible association of resistance factors against growth of allogenic tumors with blood groups. Thus he was able to discover the first antigen responsible for tumor rejection [334]. P. Medawar was able to show in the 1940s that rejection of allogenic skin grafts (host *versus* graft) has an immune mechanism [335]. In 1948 Gorer and the geneticist G. D. Snell [217] published a combined study of tumor antigenicity and genetics in mice, showing linkage between the gene encoding the histocompatibility factor discovered by Gorer (antigen II) and a marker called Fu (fused). The locus close to the Fu gene became known as histocompatibility locus II (2), in short H-2 [336]. In 1971 it became clear that the complexity of H-2 derives not only from the existance of many alleles but also of more than one locus [337,338], the H-2 locus became therefore the H-2 complex, or MHC (Major Histocompatibility Complex) in the mouse.

The human equivalent of the MHC, the HLA complex was discovered at the antigenic level in sera from multitransfused patients or multiparous women which reacted with leucocytes from different tested individuals (Dausset *et al.* [339,340] , Van Rood *et al.* [341] and Payne *et al.* [342] in 1958).

The combinatorial conundrum of the HLA complex: many loci each presenting with many alleles, started to be dissected immediately. The discovery, in 1962, of the first diallelic system – at the first locus (to be called locus HLA-B with the alleles 4a and 4b – which were later renamed Bw4 and Bw6) by analysing the reaction patterns of 60 sera from multiparous women with leucocytes from 100 donors, being another example for the early use of computer analysis in biology [343]. The second locus (to be called HLA-A) was suggested by Bodmer and Payne in 1966 [344] and confirmed in 1968 [345], and the third one (eventually to become HLA-C) was to be firmly established in 1973 [346]. To further deal with the increasing amount of complexity, International Histocompatibility Workshops (IHW) were established where techniques and results were discussed between investigators in the field. The third IHW in 1967 focused mainly on the study of families and the genetics of HLA antigens, establishing for the first time that most of the specificities were encoded on a single chromosomal region [347]. It was also the time and place R. Ceppellini introduced the term 'haplotype' [348].

Bach *et al.* suggested the existance of two classes of antigenic determinants in mice: serologically defined (SD) and lymphocyte defined (LD) [349]. This classification was later (1977, Klein) to be renamed class I (classical A, B and C and non-classical, later discovered G, E and F) *versus* class II, with LD being class II, its locus as HLA-D. The subtypes of the HLA-D antigenic determinants were identified as DR (initially DRw = D Related workshop, nomenclatures of the antigenic determinants were established by the Nomenclature Committee of the IHWS), DQ and DP.

In the 1970s the HLA complex was localized on the $6^{th}$ chromosome (1974) [350] on the short arm (6p21-22; 1977) [351].

The first reported association of a HLA serologic marker (then called 'leucocytic phenotype' – as if to further underline the switch from clearly phenotypic markers towards serologic / molecular ones!) with a disease (in this case Hodgkin's lymphoma) was reported in 1967 by Amiel at the third IHW [352].

This first association was weak: HLA-4C in patients 51% *versus* healthy individuals 27% [347].

*Polymorphism*

The concept of polymorphism is another instance in which a universally acceptable definition is hard to give. Polymorphisms are thus defined as common variants, where common has been set at an empirical threshold of 1% prevalence in a population [353]. It is maybe worth noticing that this threshold has been chosen mainly relating to the size of samples of human populations which were available in 1971, and has not been revised ever since.

### 2.2.1.3 Molecular markers

The first true molecular markers [315] were variants of enzymes (which used to be called allozymes, a contraction of the terms allelic [variants of] enzymes) distinguishable on gel electrophoresis due to differences (size, charge, conformation etc.) caused by amino-acid substitutions. Variability has been evaluated for *Drosophila* [354,355] and man [356] in 1966. The surprising results of more than expected within population polymorphism led to the hypothesis of (most) mutations being functionally neutral: Kimura's theory of neutral molecular evolution [357].

### 2.2.1.4 (True) genetic markers (DNA molecular)

In a strict sense, genetic markers are genetic material (DNA) markers. The discovery of DNA by Miescher [358] preceded the understanding of the fact that it was the vehicle of genetic information (by the momentous experiments of Griffith [359], Avery *et al.* [360], Hershey and Chase [361]).
The publication of the structure of DNA in 1953 by Watson and Crick, based on X-ray diffraction data by Franklin (on 'sodium thymonucleate') [362] and Wilkins [363], contains the now very famous understatement:

> "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material." [364]

which hints toward the importance of the discovery which is considered to mark the beginning of the era of molecular biology, and which was later expanded in a follow up paper [365]. This conceptual step was followed by a pragmatic step:

the ability to manipulate the 'semantophoretic molecules' (*semantides*) [366], giving raise to the era of genetic engineering based on recombinant technology. In order to achieve this, two breakthroughs were further needed: the first was the discovery, understanding of the mechanisms and ability to use the enzymes which modify nucleic acids. The isolation of DNA polymerase was published by A. Kornberg in 1958 [367,368], that of DNA ligase by Weiss and Richardson in 1968 [369] and the road towards the use of sequence specific restriction enzymes as 'workhorses' of molecular biology was set by a succession of theoretical and experimental developments which started with the observation that phage DNA is fragmented upon introduction in bacteria and ended with proving the usefulness of a combination of restriction enzymes and gel electrophoresis in the physical mapping of the SV40 viral genome [370]. The second had to do with superseding a technological bottleneck encountered by molecular biology due to the fact that cell based DNA amplification (cloning) is very elaborate [371]. This second difficulty was surpassed with the invention of the Polymerase Chain Reaction (PCR), which led to the possibility of cell free DNA amplification [372].

### *Restriction Fragment Length Polymorphisms*

From the point of view of genetic DNA marker development, restriction enzymes made it possible to distinguish between single nucleotide polymorphisms that were changing the sequence of one of their cutting sites. The difference between lengths of cut *versus* uncut fragments are evidenced through electrophoresis, the variants being called Restriction Fragment Length Polymorphisms (in short RFLPs) [373]. The first association study which was performed using genetic markers in 1989, was based on RFLP technology (markers) at the last step in localizing the gene (CFTR) mutated in cystic fibrosis (CF) [374].

### *Minisatellites*

Minisatelites are a class of genetic polymorphisms that has only been of limited use as markers (due mainly to their non-random distribution and the technically challenging procedure needed to assess them) is the class of minisatellites [375]. Minisatellites are tandem repeats of short (6 to 1k bp) motifs which show length polymorphism (spanning 0.5 to several kb). There is some similarity between the

consensus sequence of minisatellites and a viral (phage) sequence (GCTGTGG). They have a distribution in the human genome which shows a high bias toward chromosome ends (telomeres) [376].

### *Microsatellites*

Much more often used (more uniformly distributed across the genome, easy to assess based on a PCR assay) were microsatellites [377], also called Sequence Tagged Site (STS) type markers or Simple Sequence Repeat Polymorphisms (SSRPs). Microsatellites are tandem repeats, being composed of very short motifs (1 to 6 bp) which show length polymorphism (therefore are also part of the class of Variable Number of Tandem Repeat VNTR elements) [378].

One can note that DNA polymorphisms / variants / markers can be divided in two classess: quantitative variants (structural variants, repeat variation: minisatellites, microsatellites) *versus* strictly qualitatively variants (sequence variants proper: RFLPs, SNPs – transversions translations) [379].

### *SNPs*

At the beginning of the use of the name SNP (Single Nucleotide Polymorphism) there has been an 'unfortunate heterogeneity in the use of the term'. Initially SNPs were called only the biallelic sequence variants which were genotyped through high-throughput techniques. In time, the evolution of the term has led to the tendency to call (almost) all sequence variants SNPs [315].
Conceptually RFLPs are SNPs. Jeffrey's initial estimate 'one in 100 bp polymorphic' (estimate of human heterozygosity based on RFLPs) [380] was with hindsight very accurate. The number of common variants (primarily SNPs) identified through the efforts of the Human Genome Project, the SNP Consortium and the International HapMap Project was about 10 million [381]. The number of validated unique identifiers (#rs) in the NCBI database (dbSNP [382]; Build 132) is 19,727,605 [383].

The use of molecular markers in (linkage) mapping has been advocated in the early 1980s in a seminal paper by Botstein *et al.* [384].

### 2.2.2 Linkage mapping in humans

Mapping was proved to be easily applicable in model animals with short generation times and many offspring (*e.g. Drosophila, Mus* etc.), where mating experiments could be designed according to the question to be addressed.

In humans, the first gene to be assigned to an autosome (chromosome 1) was the gene encoding the Duffy blood group. This was based on the observation that the Duffy blood group trait co-segregates in the studied family with a cytological abnormality (alteration in the coiling structure, visible cytogenetically) on the long arm – which was the marker of chromosome 1 [385]. But this was a rather rare serendipitous occurrence.

Due to the impossibility to design mating experiments in man (as opposed to experimental organisms), special methods of analysis for pedigree data had to be developed for human linkage studies [386].

### 2.2.3 Linkage (definition)

The concept of 'linkage' (the term being used in genetics in the sense of Morgan and Lynch who wrote an article in 1912 entitled "The linkage of two factors in *Drosophila* that are not sex-linked" [387]) is best defined as the non independent segregation of two genetic loci [388].

### 2.2.4 Scoring procedures

There is a distinction between estimation of recombination frequencies (which requires the exact specification of a trait model) and linkage detection (which does not require this specification, but is more powerful in the presence of it) [389]. Bernstein (1931) applied for the first time scoring procedures to human linkage, showing that each family can be assigned a score whose sum provide a test of the null hypothesys (given the body of data is sufficiently large for the distribution of the score to be nearly normal) [326].
A development of Bernstein's method led to the first study of human linkage, published by Bell and Haldane in 1937 (which established the linkage of haemophilia and color blindness, on chromosome X) [390].

By 1953 there were three apparent methods available for the detection of linkage in human pedigrees [391]: The sib-pair method of Penrose (applicable to pairs of sisters and brothers when the partents' genotypes are unknown) [392,393]; the method of efficient scores (or weighted u-scores, by Fisher and Finney; applicable to nuclear families *i.e.* families consisting of parents and their children as opposed to extended families) and the method of likelihood ratio (introduced by Haldane and Smith in 1947; the most suitable for extended families). Smith showed that all methods are ultimately different forms of the likelihood / probability ratio test (nonsequential) [394].

### 2.2.4.1 Sib pair method

This is a nonparametric approach (does not depend on the assignment of the mode of inheritance). Sib pairs are categorized according to concordance / discordance for trait and marker. There is a natural excess of concordance over discordance even in the absence of linkage, but linkage between the trait locus and the marker locus will lead to additional excess concordants added to natural excess. The additional excess can be quantified as $\delta$ and can be used as a a test for linkage:

$$\delta = (\theta - 1/2)^2 > 0 \qquad 1.$$

The sib-pair method of L. Penrose was used by J. Mohr (in 1951) to establish the linkage between the Lutheran blood group and the Lewis (AB0 secretor) phenotype [395].

### 2.2.4.2 u-scores method

Linkage is considered to be significant when $\lambda > \lambda_0[\varepsilon]$ 2.

Where:

$\lambda_0[\varepsilon]$ is the significance point corresponding to level $\varepsilon$ and

$\lambda$ - the efficient score is the weighted combination $\lambda = \Sigma a_i x_i$

$x_i$ – observations

$C_i \subset a_i$ – class of alternatives to the null hypothesis (=1)

#### 2.2.4.3 Likelihood ratio

Haldane and Smith proposed as a test for linkage (λ) the likelihood ratio of the probability for recombination fraction (θ) to its value when θ = 0.5

$$\lambda = \theta / \theta(½) \qquad 3.$$

### 2.2.5 LOD score

C. A. B. Smith (in 1953 [391]) and N. E. Morton (in 1955 [396]) introduced the LOD (Logarithm of ODds) score method to genetic analysis. The key contribution of Morton was to consider the probabilities of type 1 (false rejection of the null hypothesis) and type 2 (failure to reject the null hypothesis) errors of the procedure. This was based on the principles of sequential hypothesis testing of Wald [397] and made the guidelines that Morton suggested for the analysis of linkage data easy to generalize / implement. By suggesting a value for the statistical significance of linkage (the now famous LOD > 3) one can add (hence the sequencial aspect) individuals / families (of an assumed homogenous phenotype) to the analysis until that threshold is met.

### 2.2.6 Computer methods

Allegedly the first use of a computer to solve a biological question was the tabulation of a solution for a differential equation in a paper by R. A. Fisher in 1950 [398]. Already in 1955, a vacuum-tube based predecessor of the IBM 650 computer was being used to tabulate lod scores for nuclear families (for N. Morton's paper which introduced his implementation of the LOD score). Based on Smith's 1959 formulation of likelihood [399], J. H. Renwick and J. Schulze developed software (written in machine language, therefore not transportable and implemented in Baltimore on an IBM 7094 computer) which was able to calculate LOD scores for large pedigrees [400].

The first linkage software to become widely available: LIPED (Likelihood in PEDigrees) [401] was developed by Ott to study a large Alaskan kindred with familial hypercholesterolemia [402] (1974). It was written in FORTRAN and computes two point LOD scores for simple pedigrees. It is based on pedigree likelihoods as computed through the Elston-Stewart (peeling) algorithm [403].

A series of popular linkage analysis programs followed, of which we mention LINKAGE (Lathrop 1984) [404], GENEHUNTER (the first software to perform exact multipoint linkage calculations for general families with many markers; Kruglyak *et al.* 1996) [405], SIMWALK (Sobel and Lange) [406] and ALLEGRO (Gudbjartsson *et al.* 2000) [407].

### 2.2.7 Positional cloning *vs.* functional cloning

Genetic mapping based on linkage of the trait with genetic markers has been called positional cloning or reverse genetics. This is in contrast to functional cloning (forward genetics) and by similarity with the reverse engineering approach [408].

Functional cloning is defined as being the identification of genes based on pre-existing biological information about the defect (disease → function → gene → map; *e.g.* purification of the normal protein product of the gene, antibodies against the protein product etc). The limiting factor of this approach is the fact that such functional insight for the vast majority of genetic disorders is non existent. By contrast, positional cloning starts with the mapping of the trait to a chromosome / region of interest which is then successively narrowed, leading eventually to the identification of the responsible gene which can then be studied functionally (disease → map → gene → function) [409].

Positional cloning has been a very successful approach, but one has to pay attention to the *caveat* of a possible (deliberate) ascertainment bias when using it: families are studied that show extreme phenotypes (fully penetrant, early onset etc). The understanding of the whole spectrum of alleles is therefore deemed to be incomplete [410].

#### 2.2.7.1 Genetic mapping

At the core of positional cloning lies genetic mapping. Genetic mapping is the evaluation of relative positions of loci on chromosomes inferred from frequencies of recombinations that are taking place between them. The measure unit for the genetic distance is the Morgan (M; usually the measure being in the range of centiMorgans - cM).

Physical mapping refers to establishing the absolute (physical) positions of loci on chromosomes and measuring the distance between them. The measurement

unit for physical distance is the base (basepair – bp; usually in the range of kilobases – kb).

Genetic mapping can be further split conceptually into linkage mapping (linkage) and linkage disequilibrium mapping (association) [411].

### 2.2.8 Linkage disequilibrium

An 'unfortunate term' [412] (as linkage equilibrium is suggestive for, yet represents no linkage between independent loci) introduced by Lewontin and Kojima in 1960 [413], linkage disequilibrium (LD) is defined as the tendency for alleles at two different loci to exhibit a statistical association [414]. Two aspects can contribute to linkage disequilibrium: (more often) the two loci are linked or the two alleles are under (common) evolutionary pressure [413]. Linkage disequilibrium can be measured by two variables (D' and $r^2$).

#### 2.2.8.1 D'

For one pair of loci (a, b) with two alleles representing the haplotype AB

$$D_{AB} = p_{AB} - p_A p_B$$

Where:

$D_{AB}$- coefficient of LD between the two alleles at the given loci

$p_{AB}$- haplotype frequency

$p_A p_B$ – product of frequencies of the alleles

$D' = D / D_{max}$ ($D'$ = 1 at perfect disequilibrium)

#### 2.2.8.2 $r^2$

$r^2$ - is a correlation coefficient: 1 represents perfect correlation and 0 no correlation at all.

$$r^2 = \frac{D^2}{p_A(1-p_A)p_B(1-p_B)}$$

### 2.2.8.3 Heatmap

A common form of representing measures of linkage disequilibrium [415] are heat map triangles, in which the measure of disequilibrium between two points is represented as a shade of colour at the intersection of the perpendiculars on the cathetes from the physical coordonates of the respective points plotted on the hypotenuse of a right-angled triangle, in a similar way that distances are represented on geographical maps, see Figure 8.

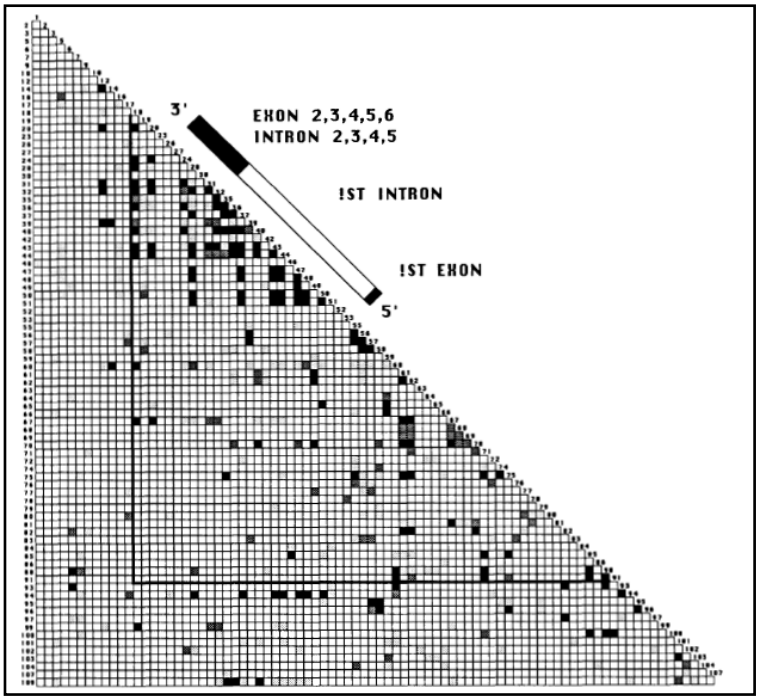


**Figure 8 Example of Linkage Disequilibrium (LD) heatmap**
From [415]; white – non significant; clear gray – significant at 5% level, dark gray – significant at 1% level, black – significant at 0.5% level.

## *2.3 Association mapping*

### 2.3.1 Background

Association was introduced as a mathematical / statistical concept by Yule [416] – applied to discrete variables, in distinction from correlation (which was defined in its biological / anthropometric acception by F. Galton [417] and applies to continuous variables).

The epidemiologic / genetic meaning of the term refers to the variable that measures (*lato sensu*) the co-occurrence of two (genetically encoded)

characteristics in a population, more often (*stricto sensu*) the relation between the trait / disease and its causal factor. It is understood that causal factors can be external (environmental) or internal (endogeneous).

The first association studies (in a very broad sense of the term) asking the question with regard to environmental factors are ancient. A passage in the bible (Daniel 1:11-16) can be interpreted as one of the first 'protocols of diet and health' [418]:

> "Prove thy servants, I beseech thee, ten days; and let them give us pulse to eat, and water to drink.
> Then let our countenances be looked upon before thee, and the countenance of the children that eat of the portion of the king's meat: and as thou seest, deal with thy servants.
> So he consented to them in this matter, and proved them ten days.
> And at the end of ten days their countenances appeared fairer and fatter in flesh than all the children which did eat the portion of the king's meat."
>
> (King James Version, Dan. 1:12-15)

A better documented instance is the experiment performed by the Scottish ship surgeon James Lind in 1746 to prove the link between citrus fruit and scurvy. It is generally considered to be the first proto-association study performed for medical purposes and was published in 1753 [419].

Association of traits (including diseases) to internal factors amounts to the aforementioned phenotype - genotype correspondence mapping.

The first classic association studies where testing links between blood groups as indirect ('proxy', phenotypic; *vide supra*) markers and different pathological entities. Weak but consistent associations could thus be established between the ABO system and traits involving the digestive tract (the first association thus described being the one between ABO blood groups – stated by the authors to be an indicator of 'genetic differences' - and gastric cancer [420]).
The evolution of markers from phenotypic to serologic and then molecular has been mentioned before. With the advent of molecular genetic (DNA) markers it became possible to directly investigate associations between the traits (phenotypes) and genotypes.

### 2.3.2 Overview of association studies

Association studies can be categorized in several types (after Balding [421]).

#### 2.3.2.1 Based on the phenotype assessed

Depending on the phenotype one can distinguish between qualitative traits, which can be dichotomous (*e.g.* case-control) and categorical (either ordered *e.g.* mild, moderate, severe; or unordered, *e.g.* distinct disease subtypes / categories); and quantitative, *i.e.* continuous traits.

#### 2.3.2.2 Based on the genotype

Depending on the genotype one can assess association one SNP at a time (allelic or genotypic) or combining SNPs (haplotype analysis).

#### 2.3.2.3 Based on the scope of the study

Association can be local, including candidate locus (candidate gene or candidate polymorphism) studies and fine mapping studies; or comprehensive *i.e.* genome wide.

### 2.3.3 The estimation of association significance

Different tests of association can be applied to assess significance in an association study for a dichotomous trait (case-control phenotype) [421].

#### 2.3.3.1 Single SNP analysis

*Allelic (allele count) test*

The contingency table for the allelic test is a 2X2 matrix (contingency table) [422] containing the counts of the two possible alleles per SNP (A and B) among cases and controls.

| | A | B |
|---|---|---|
| cases | | |
| controls | | |

There are many tests of significance [423], of which better known are the general tests: Pearson's $\chi^2$ test [424] and Fisher's exact test [421,425].

**Pearson's $\chi^2$ test**

Is used to obtain a p-value to reject the null hypothesis that the proportions in the rows and columns of a *m* x *n* unordered contingency table.

The formula used is:

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

Where:

$X^2$ = Pearson's cumulative test statistic

$O_i$ = observed frequency

$E_i$ = theoretical (expected) frequency

$n$ = number of cells in contingency table

The formula of the expected frequency is:

$$E_i = \frac{N}{n}$$

Where:

$N$ = number of observations

$n$ = number of cells

$X^2$ transformation to a p-value is based on the fact that $X^2$ approximates a chi-squared distribution with (m-1)(n-1) degrees of freedom (df). In the case of the 2 x 2 contingency table (2-1)(2-1) = 1 df.

The p-value is the probability of observing the test statistic at least as extreme (as $X^2$) in a chi-squared distribution.

In our case (where $X = X^2$):

$$p(X > x)$$

The probability distribution of X is completely characterized by the cumulative distribution function (CDF) [426].
The CDF gives the probability of a real value X to be less than (or equal to) x.

$$F_X(x) = p(X \leq x)$$

The two posibilities ($X > x$ respectively $X \leq x$) being mutually exclusive:

$$p(X > x) = 1 - p(X \leq x)$$

The p-value will therefore be the substraction of the CDF value from one.

$$p(X > x) = 1 - F_X(x)$$

Where, for discrete random variables X:

$$F_X(x) = \sum_{t \leq X} f(t)$$

The empirical threshold of statistical significance has been set at $p \leq 0.05$.

**Fisher's exact test**

Fisher criticized Pearson's approach to 'simple laboratory data' (in a characteristically 'Fisherian', yet not completely fair way) in that:

> 'Not only does it take a cannon to shoot a sparrow, but it misses the sparrow!'

The problem attacked by Fisher was mainly that of the size of the dataset, his hypergeometrical test being amenable to address small samples, which Pearson previously had scoffed about [427]:

> 'Only naughty brewers deal in small samples'

The formula of the Fisher exact test of nonrandom associations between two categorical variables (assessed in a 2x*r* contingency table) is:

$$p = \frac{(r_1!r_2!)(c_1!c_2!...c_r!)}{x_{11}!x_{12}!...x_{2r}!N!}$$

Where:

$c$ = column

$r$ = row

$x_{cr}$ = $x_{cr}$ the element on column $c$ row $r$

$c_r$ = column total

$r_c$ = row total

$N$= total elements

*Genotypic test*

The contingency table for the allelic test is a 2 x 3 matrix containing the counts of the three possible genotypes per SNP (aa, aA and AA) among cases and controls.

| | aa | aA | AA |
|---|---|---|---|
| cases | | | |
| controls | | | |

**Armitage test**

Under the assumption of an additive contribution of individual SNPs to the phenotype (aa < aA < AA; with aA being intermediate), the prefered test becomes the more conservative proportion trend test (Cochran Armitage). The principle of the Armitage test is the testing of the hypothesis zero slope for a line fitting the estimates for genotypic risk. For the ordered (by number of minor alleles in the genotype: zero, one or two) case / control contingency table:

| | aa | aA | AA |
|---|---|---|---|
| cases | | | |
| controls | | | |

The statistic for the Armitage test is:

$$z2 = \frac{b^2}{p_{case}(1-p_{case})}\sum(n_{0i}+n_{1i})(s_i-\bar{s})^2$$

Where:

$$p_{case} = \frac{n_{10}+n_{11}+n_{12}}{N}$$

$$\mathrm{p}_{1\mathrm{i}} = \frac{n_{1i}}{(n_{0i}+n_{1i})}$$

$$\bar{s} = \frac{\sum(n_{0i}+n_{1i})s_i}{N}$$

$$b = \frac{\sum(n_{0i}+n_{1i})(p_{1i}-p_{case})(s_i-\bar{s})}{\sum(n_{0i}+n_{1i})(s_i-\bar{s})^2}$$

prediction equation for (ordinary) least-squares fit:

$$\hat{p}_{1i} = p_{1i} + b(s_i - \bar{s})^2$$

For:

$n_{10}$ = count for cases with 0 minor alleles

$n_{11}$ = count for cases with 1 minor alleles

$n_{12}$ = count for cases with 2 minor alleles

$n_{00}$ = count for controls with 0 minor alleles

$n_{01}$ = count for controls with 1 minor alleles

$n_{02}$ = count for controls with 2 minor alleles

$N$ = total count

#### 2.3.3.2 Multiple SNPs

Analysing one SNP at a time ignores the block-structure of linkage disequilibrium (LD) in the human genome: regions of relatively low recombination frequency (so called haplotype blocks, in short haploblocks) are separated by 'hotspots' of recombination [428].

*Linkage disequilibrium blocks*

It has been shown that the human genome is structured in regions protected from cross-overs (so called recombination cold spots) separated by regions of intense recombinatorial activity (so called recombination hot spots). The regions of low recombinatorial activity are defined by loci in the respective stretches segregating together (being linked, therefore in linkage disequilibrium). These regions are therefore also called linkage disequilibrium blocks. In the European population linkage disequilibrium blocks (or haplotype blocks, in short haploblocks) have been shown to stretch in average over 30 kb.
Association studies detect markers that are in linkage disequilibrium with causative polymorphisms (or, in extremely lucky cases, the causative polymorphism is the marker for which association has been detected).

Information from neighboring markers can be captured by two different approaches: one which has been called marker (SNP) based (and is based on logistic regression) and a haplotype based one [429]. Importantly, haplotype based approaches can capture the combined effects of *cis*-acting causal variants which are tightly linked [430]. The fundamental problem of haplotype based methods is the extreme technical difficulty of observing haplotypes as such (short of single chromosome typing, spermatide DNA typing and other such tedious techniques), which leads to the necessity to infer them (so called phase inferrence). This introduces a certain level of uncertainty to the analysis, which is indirectly proportional with the level of linkage disequilibrium between markers [421].

### 2.3.4 Scope of the study classification of GWA studies

#### 2.3.4.1 Candidate gene association studies

Genetic markers started to be used for association studies involving candidate genes [431]. Case-control gene association studies though have been shown to be biased towards a high false-positive rate; which has been often attributed to stratification (though this has not been proven). More likely this bias is due to the low prior probability of association for any given gene. It is suggestive then that HLA – autoimmune disorder association studies show less false positive bias (HLA associations have been repeatedly confirmed), possibly reflecting the higher prior probability for this complex of loci when compared to other loci tested throughout the genome [432].

#### 2.3.4.2 Comprehensive association scans

Risch & Merikangas suggested the strategical change of the mapping method from linkage to association [433]. It became only a matter of technology evolving for the first genome wide association studies to be performed.

*Known genes association scan*

An intermediate strategy was the querying of the association of traits with functional SNP markers, genotyped by a PCR based high throughput method [434], in known genes. The first such study, by Ozaki *et al.*, used 92,788 SNPs to establish the association between LTA (Lymphotoxin Alpha) and myocardial infarction [435]. The second study assessed the association of rheumatoid arthritis with SNPs in known genes, its results being published stepwise, as they were revealed [436,437].

*Genome wide proper*

The first successful proper genome wide association study was published by Klein *et al.* in 2005, establishing a polymorphism in complement factor H as a cause for age related macular degeneration [438]. By GWAS proper we understand a study in which markers are (*quasi*) evenly distributed across the genome, with no regard to the positions of known genes.

### 2.3.5 Problems

Two problems have to be dealt with when performing an association study: population stratification and (for genome wide studies) multiple testing.

#### 2.3.5.1 Population Stratification / Structure

Population stratification (initially called population structure by S. Wright [439]) is defined as the difference between allele frequencies of cases and controls which are not due to the investigated trait but are determined by systematic ancestry differences. Population stratification has been shown to influence the results of case control association studies [440]. We are aware of the propensity of latent population stratification to inflate the type 1 error rate, generating false positive signals for association in variants informative for substructure [441,442]. The classic example of spurious association due to population stratificationon is the study of type II diabetes in Pima Indians. Pima Indians have higher rates of diabetes than individuals of Caucasoid ancestry. However the Pima population contains individuals with higher degrees of Caucasoid ancestry (with accordingly lower diabetes sucseptibility). Loci at higher frequency in Pima *versus* Caucasoids were hence spuriously associated with the disease [443]. The effect of population stratification is directly proportional to the number of samples used in the study). [444]

##### *Genomic control*

The careful choice of controls cannot completely alleviate the risk of hidden stratification (hidden or 'cryptic' stratification being stratification that is difficult to detect but which may still be significant in genetic terms) [445].
One of the methods suggested to correct for 'cryptic' stratification is called genomic control [446]. The association statistics are adjusted at each marker by dividing it by a uniform factor called genomic inflation factor (lambda), the assumption being that stratification artificially 'inflates' the probability of association.

##### *Structured association*

The limitations of the genomic control method are due to the differences in population frequencies of different markers (some markers have strong

differences between different populations, in this case the applied uniform genomic control inflation factor might be insufficient; other markers vary less markedly between populations, in that case applying the uniform genomic inflation factor is going to result in a loss of power for the association statistic).

Another method to deal with population stratification is structured association. Information from unlinked genotypes is used to assign samples to discrete subpopulation clusters. The software which implements the structured association algorithm is called STRUCTURE [445]. However the method cannot be applied to genome-wide association studies because of the computational cost of large numbers of markers. To deal with this problem a method called EIGENSTRAT [441] has been suggested, in which the discrete clusters of subpopulation are obtained (as in STRUCTURE), the difference being that principal components are used as synthetic variables. Synthetic variables are linear combinations of original variables which are summarized in such a way (by multivariate techniques – such as principal component analysis, PCA) as to minimize the amount of information lost [447].

### 2.3.6 Issues of Quality Control

#### 2.3.6.1 Hardy-Weinberg equilibrium

The Hardy-[448]-Weinberg equilibrium criterion states that in panmictic Mendelian populations gene frequencies are stable [449]. *Panmixia* pertains to the quality of the population to reproduce through random mating (*i.e.* each male of the population has an equal chance of mating with each female); a Mendelian population is named as such because the basic Mendelian laws of heredity are followed in its interbreeding individuals.

Deviations from Hardy-Weinberg equilibrium nullify the hypotheses of *panmixia* or evolutionary stasis. This is the case when the studied population is inbred, stratified, or when one deals with genotyping errors. It has therefore been suggested as a quality control measure in populational studies [421].

The suggested values ($P$) at which one considers deviation to be significant (and are used as a thresholds for the Hardy-Weinberg quality control step) are either $10^{-3}$ or $10^{-4}$ [421]. It has to be noted that deviations from Hardy-Weinberg

equilibrium can also be a sign of real association (in the case of deletions, duplications, mutations under selective pressure etc.).

### 2.3.7 Separating false-positive from true-positive associations

#### 2.3.7.1 Multiple testing correction

Statements based on statistical evidence are uncertain. This is usually dealt with by specifying a maximum probability threshold at which the null hypothesis will be rejected even though it is true (type I error) as the level of statistical significance. When more than one hypothesis is tested simultaneously (which is referred to in multiple testing) each test has its own type I error probability. The possibility that type I errors are commited increases therefore with the number of hypotheses / tests / comparisons (this increase is called 'inflation of the alpha level' – where alpha is the set significance level). The rate of Errors Per Comparison is usually referred to as PCE [450].

There are many methods that deal with the multiple testing problem, each adapted to different situations. One method that is considered to be suitable to address the problem in the case of association studies and therefore has been most commonly used herein is the unweighted Bonferroni correction.
The simple Bonferroni method takes the form:

$$\text{reject } H_i \text{ if: } p_i < \alpha[PF] = \frac{\alpha[PT]}{C}$$

Where:

$\alpha[PT]$ = alpha per test

$\alpha[PF]$ = alpha per family of tests

$C$ = number of tests (Comparisons)

It has been shown that without correction for multiple comparisons (conventional level of significance $P < 0.05$) an association study using 1 million SNPs will show 50,000 'associations' to the trait, almost all being false positives (due to chance alone) [451].

#### 2.3.7.2 Replication

Many genotype-phenotype associations reported in the era of candidate gene association studies have been shown to be spurious. This was done by the means of the replication study, which purpose is to evaluate a finding from a previous (exploratory) study. The same approach has been suggested in order to provide credibility for the validity of the findings in exploratory studies in the case of GWA studies [452]. Ideally identified associations have to be replicated in independent population studies.

#### 2.3.7.3 Examination of functional implications experimentally

Linkage and association studies identify genomic locations in which it is assumed polymporphisms causative for the studied trait are harbored. The causative link and ultimate proof of association is only provided by functional studies elucidating the effects of polymorphisms with functional implication (identified by the sequencing of the region of interest) on gene expression or the transcribed product. The functional studies usually used include studies of morphology, biochemistry, immunology and / or physiology at the cellular level, tissue level or organism level in biologic material from humans or from animal models (which have been usually genetically manipulated *e.g.* knock-out or knock-in models [451]).

### 2.3.8 Imputation of untyped genotypes

To increase the power of the studies performed on single platforms, to uniformize the datasets and not to lose genotypes that have not been genotyped across different platforms the possibility exists to impute missing genotypes based on the HapMap dataset. Existing software for imputation are PLINK, MACH, BEAGLE, IMPUTE [453].

# 3 Findings

## *3.1 MN Linkage mapping*

### 3.1.1 Family

We performed linkage analyis studying the largest reported family affected by membranous nephropathy to date, described in 2008 by Bockenhauer *et al.* [215]. Of the 34 individuals in the pedigree spanning five generations, 19 were investigated and 4 were shown to be affected.

### 3.1.2 Phenotype

In the case of membranous nephropathy, we considered the phenotype to be binary: affected *versus* nonaffected. Given that membranous nephropathy is a histopathologically defined entity the clinical diagnosis had to be confirmed through renal biopsy.

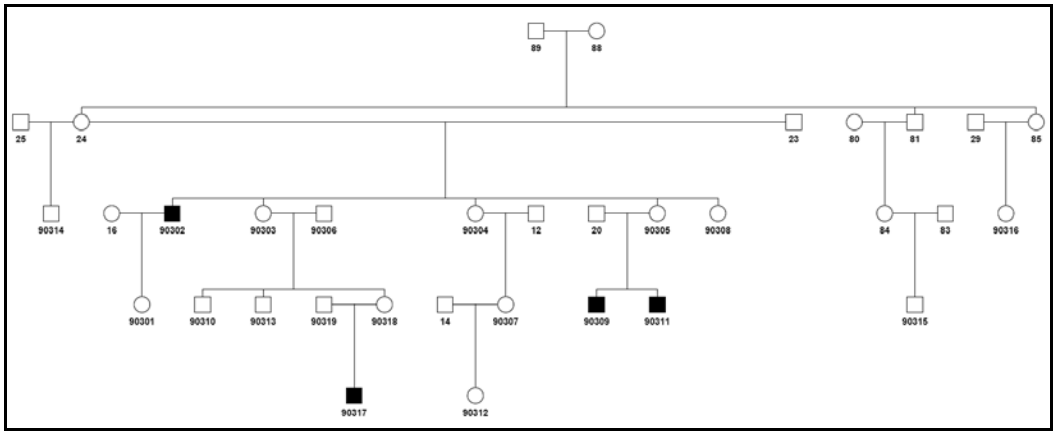


**Figure 9 Pedigree of the family reported by Bockenhauer *et al.***
Squares: males; circles: females; white: unaffected; black: affected.

### 3.1.3 Genetic component

Considering the fact that membranous nephropathy is a rare diagnosis in childhood (estimated incidence of less than 1 per million in the child population) it was assumed that the disease had a common basis in the family (the alternative hypothesis – of all individuals being independently affected by chance – having a probability of less than $10^{-24} = (10^{-6})^4$). An additional argument in favor of the

genetic *versus* environmental cause hypothesis was the fact that the affected individuals were living in three different households, without any observed increased incidence of membranous nephropathy in their vicinities.

### 3.1.4 Pathological – Clinical description

The patients are all males, they presented at different ages (between 1 and 67 years) with nephrotic-range proteinuria (greater than 3.5 g/1.73 m2/24 hours in adults [454]). In all of them kidney biopsies were performed which showed the typical pattern of membranous nephropathy (granular deposition of IgG, 'spike and dome' basement membrane proliferation). There were however some aspects consistent more with membranoproliferative glomerulonephritis: mesangial proliferation (in one of the individuals) and 'double contours' (in two of the individuals). Causes of secondary membranous nephropathy were searched but not identified: serum complement fractions C3 and C4 were not depressed, immunoserology was shown to be negative for anti-nuclear and anti-double-stranded DNA antibodies (positive in lupus nephropathy), no markers of chronic infections were found (hepatitis, HIV).

### 3.1.5 Serology

Serological studies were performed on sera from two of the affected individuals and their mothers to assess for antiglomerular antibodies. In the youngest (onset of nephrotic syndrome - unresponsive to treatment with corticosteroids - with microscopic haematuria at 1 year of age) affected member of the family there was positive immunofluorescence suggesting that foeto-maternal alloimmunization might have been a triggering factor in this case. The antigen remained elusive (it did not recognize specific proteins in the lysates of positive specimens on Western blot analysis).

### 3.1.6 Inheritance pattern

The segregation study of the pedigree (the fact that all affected individuals are males and connected through the maternal lineage) suggests an X-linked inheritance pattern. Other possibilities that cannot be discarded are an autosomal dominant inheritance model with incomplete penetrance or a mitochondrial inheritance model.

### 3.1.7 Linkage Analysis

Linkage analysis was performed as follows:

#### 3.1.7.1 Genotyping

Blood samples for DNA extraction were collected from 19 individuals - including all 4 affected individuals (identified as 90301 to 90319 in the pedigree). Genotyping was performed using the highly informative DeCode polymorphic STS (Sequence Tagged Site [455]) marker set of 2000 polymorphic markers [456], which covers the human genome at an average marker density of approximately 2 cM. The di- tri- or tetra microsatellite genetic markers were amplified using multiplex PCR with fluorescently labeled primers. The lengths of the markers were resolved by capillary electrophoresis in an automated process. In the DeCode set, chromosome X is covered by 86 STS polymorphic markers, the coverage for chromosome X being 193 cM (approximate genetic length of chromosome X in the DeCode genetic map) / 86 (number of markers) = 2.24 cM. Figure 8.

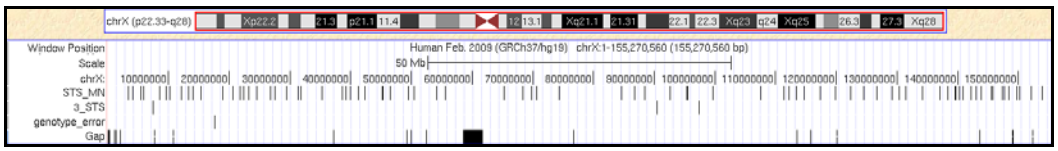


**Figure 10 Chromosome X markers used in linkage analysis**
The figure shows almost uniform coverage of chromosome X with 83 markers. Marker names were enquired on the UniSTS.aliases file accessed through the FTP server of the UniSTS [457] database at NCBI (accessed at 23/05/2010, last modified 21/05/2010). STS positions, corresponding to the human February 2009 (GRCh37/hg19) genome assembly, were obtained through the Table Browser data retrieval tool [458] of the UCSC Genome Browser [459]. For three DeCode (DXS*) type identifiers (DXS6786, DXS6799 and DXS9895) there were no aliases present in the UniSTS.aliases file that could be used for the UCSC query. Their physical positions were retrieved through individual search of the UniSTS database and are plotted on a different track in the figure. The plots on chromosome X were obtained using the Browser Extensible Data format (BED; tab separated; chromosome, start physical position, end physical position, name) custom track annotation tool on the UCSC Genome Browser. Also shown on the figure are the sequencing gaps corresponding to the human February 2009 (GRCh37/hg19) genome assembly. The largest gap corresponds to the centromere.

#### 3.1.7.2 Quality check: Pedcheck

The genotypes obtained were checked for Mendelian inconsistencies using Jeff O'Connell's software, Pedcheck [460]. Marker DXS1195 (see genotype_error custom track in Figure 10) was found to be inconsistent with individual 90309

(genotyped 3/3 at this locus) being the child of father 20 (genotyped 0/0 at DXS1195) and mother 90305 (genotyped 4/4 at DXS1195). We assumed this has been a genotyping error and re-coded the genotypes 0/0 both for individuals 90309 and 90305.

#### 3.1.7.3 Quality check: GRR

The structure of the pedigree (possible errors in the assignment of relative status) was checked using Goncalo Abecasis' software, GRR (Graphical Representation of Relationship errors) [461].

#### 3.1.7.4 Exact algorithm: Genehunter

Linkage was assessed first by using a modified version of the exact peeling algorithm of Elston and Stewart [462] as implemented by Lander and Kruglyak in the second version of the software GENEHUNTER [463]. GENEHUNTER performs an exact multipoint analysis. It assesses linkage model free (nonparametric) and model based (parametric). The computational time needed by GENEHUNTER increases lineary with the number of markers and exponentially with the number of individuals to be analyzed in one family. For the analysis of the X linked transmission model the X linked version of GENEHUNTER (XGG v. 1.3) was used.
Due to computation time and memory constraints not all individuals of the family can be analysed using the GENEHUNTER package.

#### 3.1.7.5 Merlin / Minx

To assess the X linked transmission model of inheritance the MINX (MERLIN in X) module ('unpublished but believed to be correct' [464]) of the MERLIN software [465] was used. The input for MINX/MERLIN was edited (LINKAGE format [404] to MERLIN format) using the data-handling software Mega2 [466].

#### 3.1.7.6 Haplotype analysis: Haplopainter

Haplotypes were reconstructed with GENEHUNTER X and MERLIN/MINX and visualized using HaploPainter [467].

### 3.1.7.7 Analysis (GENEHUNTER X)

We used a transmission model assuming an X linked dominant inheritance model, complete penetrance and a disease allele frequency of 0.001. Encoded in the datafile (DAT, which is the file describing the model of inheritance and the markers) input of the GENEHUNTER software in the LINKAGE format as follows:

```
0.9990 0.0010 # frequency of the major allele - frequency
of the minor allele
1 # number of traits
0.0000 0.0000 1.0000 # penetrance for females aa, aA, AA
0.0000 1.0000 # penetrance for males a0, A0
```

Result

Due to computation time and memory constraints, individuals 90308, 90314, 90312, 90301, 90316 and 90315 have been dropped from the analysis. No linkage peak was observed using this scenario, see figure 12.

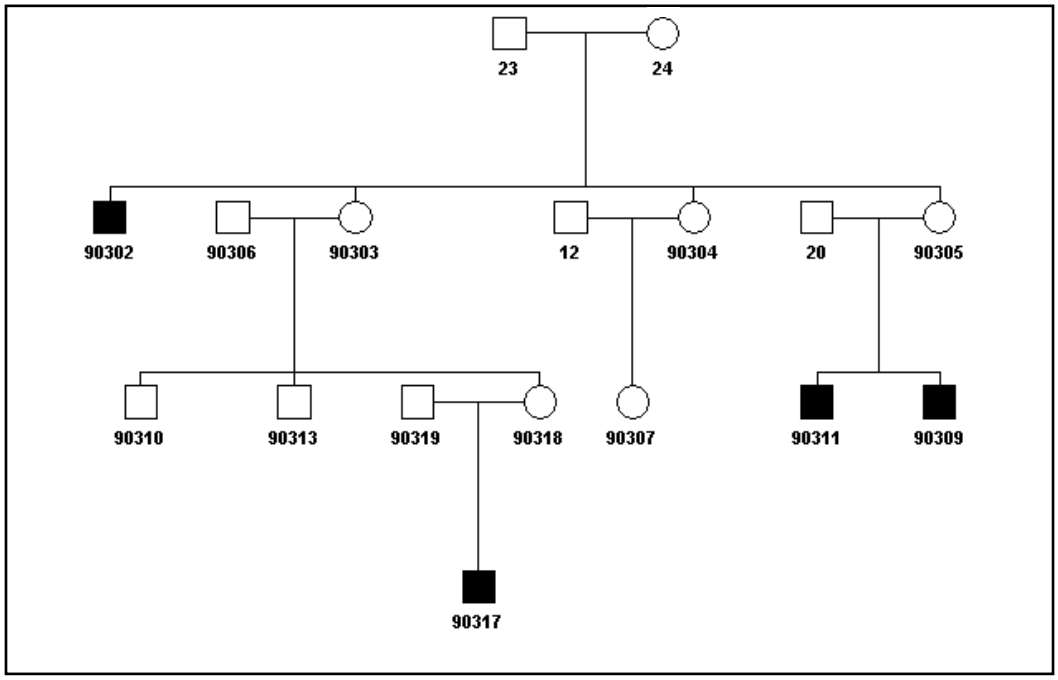


**Figure 11 Trimmed pedigree**
Used for linkage analysis under the constraints of GENEHUNTER. Squares: males; circles: females; white: unaffected; black: affected.

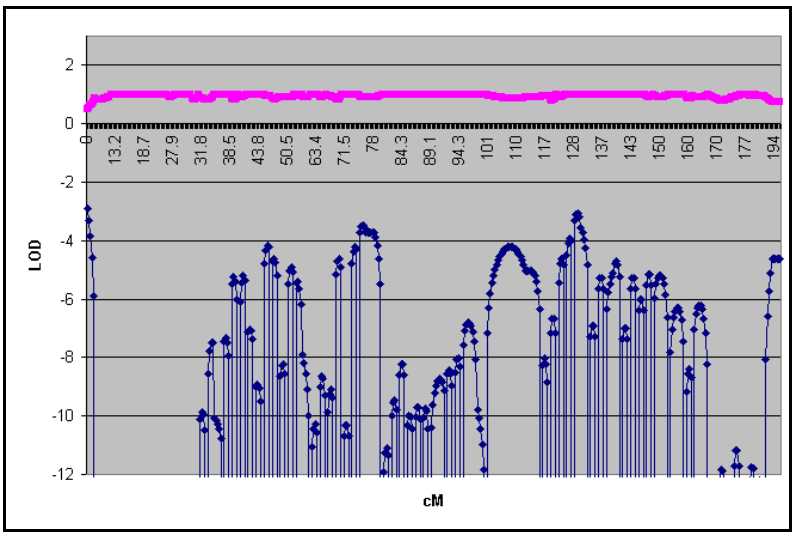


**Figure 12 LOD score curve for the first linkage analysis scenario**
Linkage analysis performed on the GENEHUNTER trimmed pedigree. The y axis quantifies the LOD score, the x axis the genetic distance (in cM – centimorgans, for chromosome X), the magenta line represents the information content, the blue line the LOD score curve.

### 3.1.7.8 'Affected only' analysis

Considering the possibility of incomplete penetrance, the analysis was redone using the affected-only scenario (the unaffected individuals were recorded as "phenotype unknown") as suggested by Lander [468], see Figure 13.

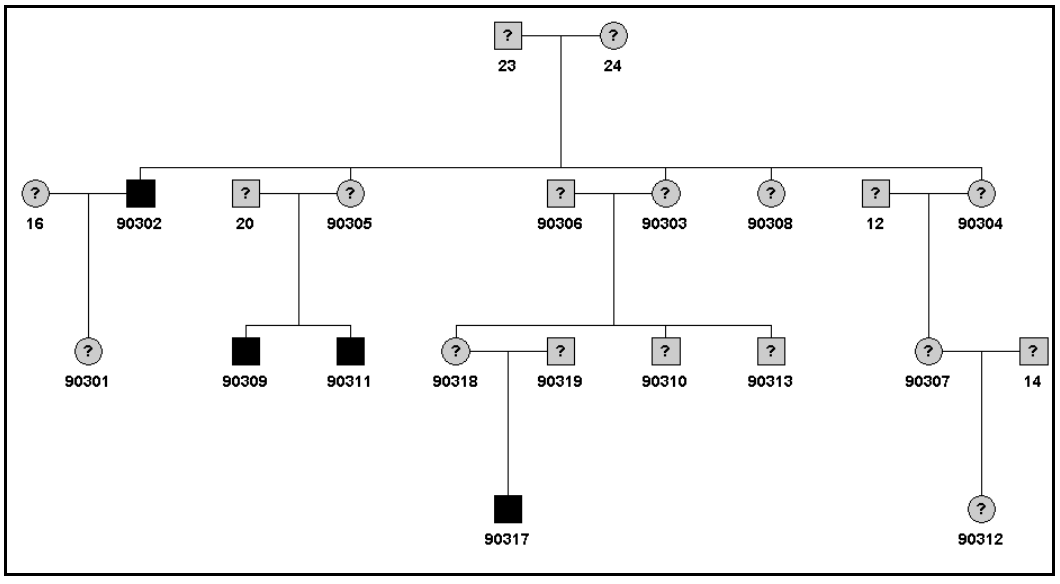


**Figure 13 Pedigree used for the affected-only linkage scenario**
Trimmed pedigree as used for the GENEHUNTER analysis in the phenotype unknown scenario (phenotype unknown is represented by question marks – colored gray; affected individuals: black, unaffected ones: white).

The LOD score curve for the affected-only scenario shows a positive peak (LOD ~ 1.2) between 110 and 120 cM. Table ; Figure 14.

**Table 1 LOD score affected-only linkage scenario**

| cM | LOD score | Information content |
|---|---|---|
| 118.14 | -1.20286 | 0.965445 |
| 118.46 | 0.996628 | 0.844138 |
| 118.77 | 1.171701 | 0.805222 |
| 119.09 | 1.171645 | 0.811299 |
| 119.41 | 0.99646 | 0.862723 |
| 119.73 | -1.20286 | 0.998735 |

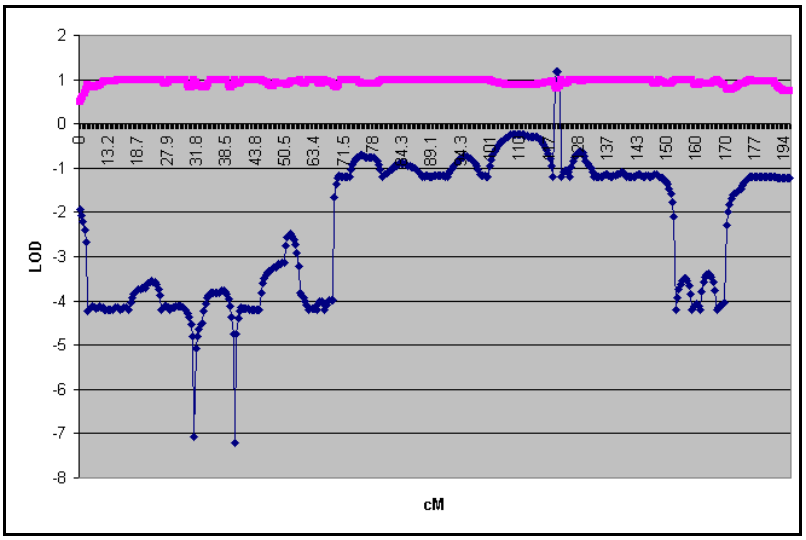


**Figure 14 LOD score curve for the affected-only linkage analysis scenario**
The y axis quantifies the LOD score, the x axis the genetic distance (in cM – centimorgans, for chromosome X), the magenta line represents the information content, the blue line the LOD curve.

The region of linkage has been defined between the markers where the inflection of the curve from negative to positive multipoint LOD scores takes place, see Table 2.

**Table 2 markers defining the linked region**
Affected-only linkage analysis scenario; physical coordonates (bp) on chromosome X according to the Human Feb. 2009 (GRCh37/hg19) assembly; genetic distances according to the deCODE map.

| marker names | chrX start | chrX end | cM |
|---|---|---|---|
| DXS8064 (AFMB319YA5) | 117272895 | 117272895 | 119.68 |
| DXS1001 (AFM248WE5) | 119836526 | 119836526 | 123.76 |

### 3.1.8 Haplotype reconstruction

Haplotype reconstruction was performed using the MINX software.

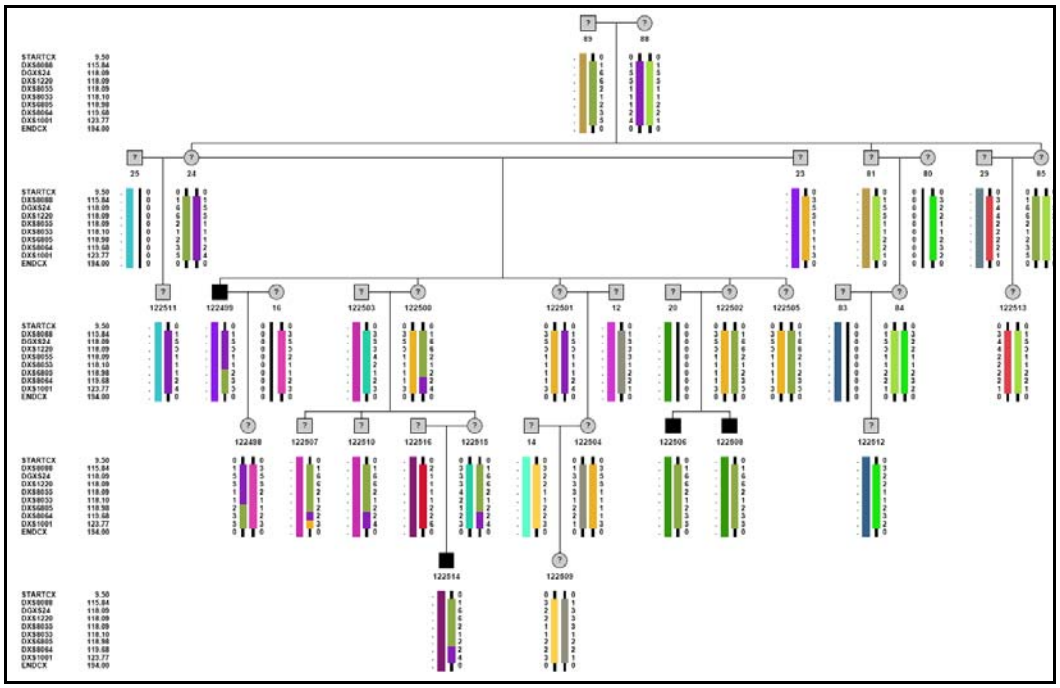

**Figure 15 Haplotype reconstruction on chr X**
The affected haplotype is represented in olive-green, the two recombinations establishing the upper and the lower limit of the region of interest are observed in individuals III.2 and V.1.

The region of interest has been redefined based on the markers immediately outside of the observed meiotic events (which represent the limits of the possible physical positions of the respective events): DXS8053 - DXS8064. Table 3; Figure 15.

**Table 3 markers defining the affected haplotype**
Affected-only scenario haplotype reconstruction; physical coordonates (bp) on chromosome X according to the Human Feb. 2009 (GRCh37/hg19) assembly; genetic distances according to the deCODE map.

| marker names | chrX start | chrX end | cM |
|---|---|---|---|
| DXS8053 (HSB291XA5) | 115564734 | 115565136 | 114.85 |
| cM DXS8064 (AFMB319YA5) | 117272895 | 117273206 | 119.68 |

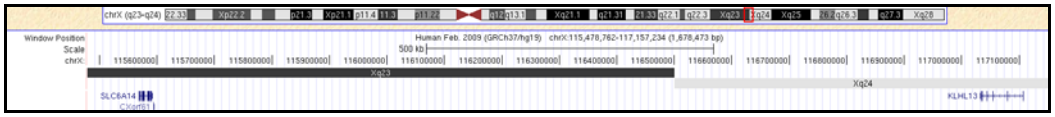

**Figure 16 genes in region of interest on chromosome X**
Representation of the region of interest on the UCSC Genome browser (human February 2009 GRCh37/hg19 genome assembly). Shown are the known genes (as defined by UCSC [469]). From right to left: SLC6A14, Cxorf61 and KLHL13.

The newly defined region of interest spans 1,678,473 bp and includes 3 known genes (as defined by UCSC [469]): SLC6A14 (SoLute Carrier family 6 member A14), Cxorf61 (chromosome X open reading frame 61) and KLHL13 (Kelch Like family member 13).

Given that membranous nephropathy has a strictly localized visible lesion (in the kidney at the level of the glomerulus), one can assume (as a working hypothesis) that the molecule(s) important in the pathogenesis of the disorder are going to be expressed at that particular location. To address this we performed *in silico* expression analysis on the set of genes which are encoded in the region of interest.

### 3.1.9 *In silico* expression analysis

We compared the expression levels of the known genes in the region of interest using the filtering capabilities of the UCSC Gene Sorter [470] program (part of the UCSC Genome Browser). Expression levels of human and mouse transcripts accessed through Gene Sorter were provided by Novartis GenAtlas 2. The publicly available expression profiles (normal human transcriptome; under the BioGPS – Gene Portal Hub [471]) of messenger RNA from a diverse array of normal human tissues and cell lines were assessed by hybridization on high-density custom oligonucleotide array chips targeting predicted genes. Biotinylated and fragmented cDNA synthesized based on total RNA from 79 human tissues was hybridized to the Affymetrix HG-U133A and custom GNF1H microarray chips targeting a set of 44,775 nonredundant human transcript probes compiled from Refseq, Celera and Ensembl (corresponding to 33,698 unique human genes – according to the January 2004 annotation; accounting for multiple probe sets interrogating single genes split transcripts). The arrays were scanned, the resulting images were analyzed by the M AS5 algorithm [472] and

the results were normalized by using global median scaling (to correct for global multiplicative effects, assuming that different sets of intensities, *i.e.* different slides, differ by a constant – global – factor [473]). The experimental procedure being 'the equivalent of multiple tissues Northern blots for thousands of genes at once' [474,475].

The levels of expression are calculated for each gene as $\log_2$ (tissue/reference) and the values range from -5.0 to 5.0 (0 being the median expression for the group of tissues, *i.e.* brain tissues, fetal brain included in one expression group).

The only gene expressed in the kidney according to the *in silico* expression dataset is *KLHL13*, at the relative level of 0.955.

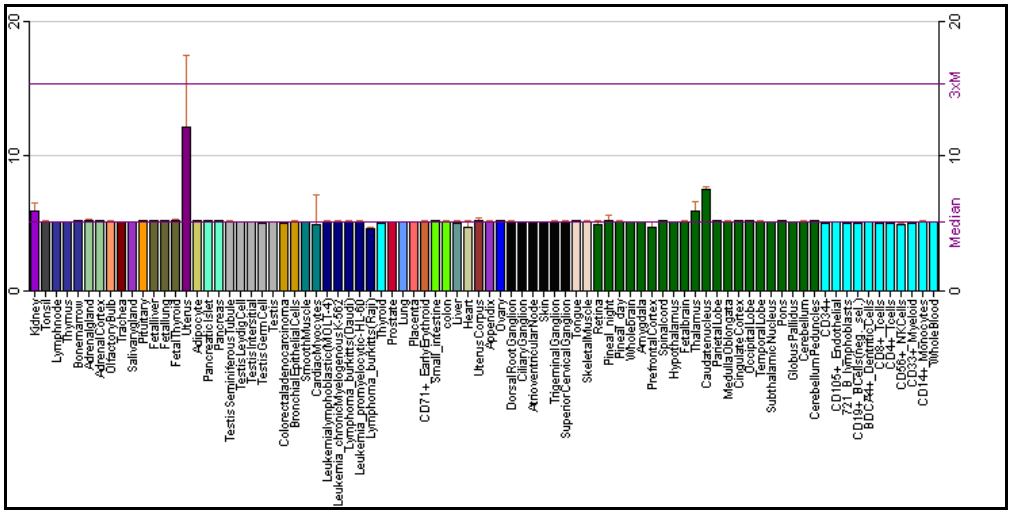


**Figure 17 Expression levels of *KLHL13***
Expression levels of *KLHL13*: x-axis different cells / tissues; y-axis normalized expression levels (BioGPS database, accessed 10/01/2011 [471]).

### 3.1.10 KLHL13

*KLHL13* (Kelch Like protein 13) is a member of the BTB (Bric-a-brac-Tramtrack-Broad complex)-Kelch protein family. It is a substrate specific adaptor for the Cul3-based E3 ligase, interacting with Cul3 (Cullin3). Cul3-BTB complexes are ubiquitin ligases [476]. It has been suggested that the Cul3/KLHL9/KLHL13 E3 ligase coordinates mitotic progression (completion of cytokinesis) [477].

## *3.2 MN Association mapping*

Given the paucity of families where there is a clear Mendelian inheritance pattern for the membranous nephropathy trait we considered applying the case-control association method to discover genes with an impact on the etiopathology of the disorder.

### 3.2.1 Exploratory (UK) analysis

#### 3.2.1.1 Ethical approval

IRB (Institutional Review Board) approval was given by the Oxford Multicentre Research Ethics Committee (Oxford, UK).

#### 3.2.1.2 Selection of cases

The cases that have been selected for the study are 336 patients which have been assessed and diagnosed by the UK Membranous Nephropathy Consortium. The diagnosis has been confirmed in all of them by kidney biopsy. As expected more males than females are affected: 231 males *versus* 105 females (gender ratio: 2.2 / 1). The samples have been provided by the MRC/Kidney Research UK National DNA Bank for Glomerulonephritis.

No manoeuvres have been attempted to enrich for specific disease-predisposing alleles in the case cohort (*e.g.* by selecting individuals that are affected to the extremes of the clinical spectrum, that present with an early age of onset or by using familial cases [478]).

#### 3.2.1.3 Genotyping

DNA was extracted from whole blood samples from patients, collected after informed written consent was given by each participant. Genotyping was performed on a HumanCNV370-Quadv3 BeadChip (Illumina, Inc.) platform (deCODE genetics, Iceland).

The choice of genotyping chips for the cases was based pragmatically on the availability of historical control genotypes assayed with the same platform, and the presence of copy number variation data for these chips.

*Genotyping technology*

The BeadChip genotyping technology is based on the Infinium HD Assay which is a Whole Genome Genotyping Single Base Extension (WGG-SBE) assay [479], an improvement of the Allele Specific Primer Extension (ASPE) assay [480,481]. This technology relies on an initial whole genome amplification step [482], performed on 200 – 400 ng genomic DNA followed by fragmentation of the DNA and then by the hybridization of the sequences adjacent to the SNP query sites on 50 bases long single probe oligonucleotides that have been designed appropriately. The hybridization is designed to stop one base before the interrogated base (which is the SNP marker proper). In the next step fluorescently labeled nucleotides are specifically incorporated according to the template through enzymatic single-base extension, see Figure 18. The color and intensity of the fluorescent signal are detected on an imaging system (Illumina iScan).

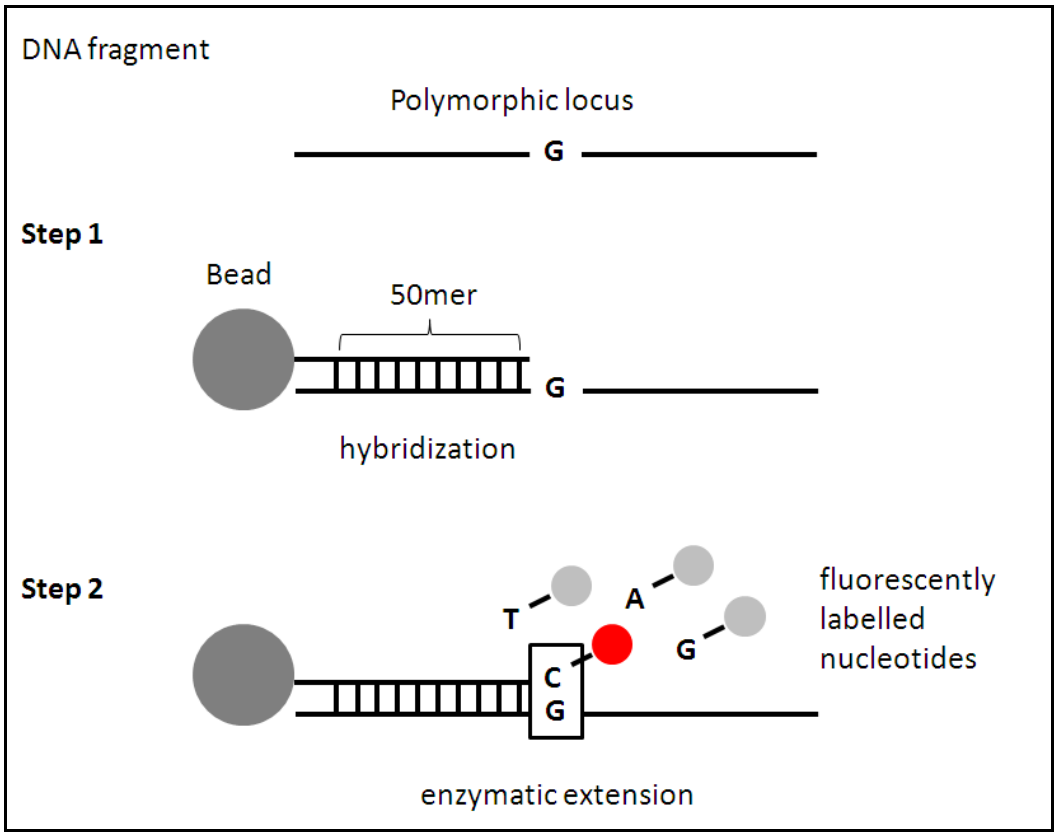


**Figure 18 Infinium genotyping assay**
Starting from amplified and fragmented genomic DNA; step 1: hybridization of unlabeled DNA fragments to 50mer oligonucleotide on a bead in the array; step 2: enzymatic single base extension (with fluorescently labeled nucleotide).

*Genotyping data*

Raw data (intensity data) was normalized using the Bead Studio analysis software (Illumina, Inc.). The normalization step used the raw data from the array itself plus five essential steps including outlier removal, background estimation, rotational estimation, shear estimation and scaling estimation. Eight CEPH samples (four pairs of duplicates) were used as controls. The reproducibility rate for the CEPH duplicates was greater than 99% per duplicate pair (average reproducibility rate 99.55%).

Coverage [483,484]: 373,397 markers were interrogated per BeadChip (344,678 SNPs, representing 92.31% and 28,286 CNV – Copy Number Variant - markers, representing 7.57%).

The definition of a successful sample implies a call rate that was greater than 99 % for SNPs and greater than 95 % for CNV markers. Lower quality data samples have call rates between 0.98 and 0.99. According to these definitions, 313 of the 336 samples where considered successful (93.15%) and 20/336 (5.95%) samples were considered lower quality data. The average call rate was 99.61%. One sample was identified as being pooled (*i.e.* contaminated whith DNA from different samples).

The call frequencies were 99.67% for SNPs (343,554 out of the released 344,678 SNPs on the chip) and 97.67% for CNVs (22,403 out of the released 28,286 CNV markers on the chip).

Genotypes were encoded according to the 'Forward' definition (same as dbSNP, see below).

*SNP encoding*

To perform an association analysis one theoretically needs just to distinguish between the two possible alleles for each given SNP (called A and B). The problem arises when comparing different populations the naming of A and B

might have been swapped. Calling them 'major' versus 'minor' allele (according to their respective frequences) is not a solution given that what is a major allele (more frequent) in one population might be minor (less frequent) in a different one. Ideally one would then want to designate the alleles according to the nucleotide present at that locus in the genotype (A, T, C or G).

The nucleotide type encoding of SNPs with regard to strand designation (sense versus antisense) and/or orientation (forward versus reverse) presents theoretical challenges due to the 'work in progress' nature of successive assemblies of the human genome reference sequence, where a clone can be assigned different orientations therefore the coding strand being changed (*e.g.* a sequence can be assembled into the forward / coding / sense strand in the previous version of the human genome assembly, but being corrected to the reverse strand in the next assembly).

In practice there are three ways of describing a SNP: forward (5' to 3') / reverse (3' to 5') depending on the way the SNP was first (historical) submitted to the database (as applied by dbSNP [382]), same (*i.e.* forward) strand 5' to 3' (as encoded in HapMap), and top / bottom (which is the proprietary solution to the encoding problem suggested by Illumina [471,485]). The most commonly used coding scheme is the forward encoding, which assumes that the orientation of the strand is known at the given position. Given that the human genome is relatively well annotated, the orientation of the strand being usually (but still not always!) known, one considers the sense (coding, forward) strand as being the reference on which to call the two possible alleles of each SNP.

### *Controls*

The control cohort for this study consists of 349 individuals that are assumed to be unaffected by membranous nephropathy. The genotypic data has been obtained for British controls from the population-based birth cohort (thus being a so called 'historical' control type cohort). We were aware of a possible misclassification bias, which would result from the impossibility to exclude latent diagnoses of membranous nephropathy in the control cohort. The consequences of misclassification bias are known to be modest unless the trait is common, but membranous nephropathy is not common therefore risk for selection bias is minimal. The Wellcome Trust Case Control Consortium

(WTCC) assuaged concerns about the damaging potential of the misclassification bias in a study [486].

The control cohort consists of 108 males and 241 females (with a gender ratio of ~ 1:2.4). DNA was extracted from blood and genotyped on an Infinium humanHap300 (Illumina) platform.

Coverage: 317,503 markers were interrogated per BeadChip (only Single Nucleotide Polymorphism SNP type markers, no Copy Number Variation CNV markers).

A successful sample was considered a DNA sample for which the call rate was greater than 99 % for SNPs. According to this definition 306 of the genotyped 349 samples where successful (87.67%). 43/349 (12.33%) samples were considered lower quality data (call rates between 0.98 and 0.99). The overall average call rate for the cases cohort was 99.51%.

#### 3.2.1.4 Data management and analysis

*Data management and description*

Files were stored as text files (.txt) and were manipulated using shell commands in Cygwin (a Linux like environment emulator [487]) and Python [488] scripting. The files encode matrices in which the columns are the genotyped markers and the rows represent the genotyped individuals. The data cells contain the genotypes of the genotype for the given marker in the given individual (A_B where A and B are the two alleles encoded as nucleotides in the forward convention). The first two columns of the matrix encode in a binary mode (according to the LINKAGE format [489]) the minimal phenotype of the individuals (gender: 0 male, 1 female; and affection status: 0 unaffected, 1 affected), see Table 4.

**Table 4 example of a case control matrix**

| | gender | status | rs12345 | rs123456 | etc |
|---|---|---|---|---|---|
| B0100603 | 0 | 0 | C_C | G_G | … |
| B0100604 | 0 | 1 | C_T | A_G | … |
| B0100605 | 1 | 0 | T_T | ?_? | … |
| etc | … | … | … | … | … |

Quality control and analysis were performed using the SNP & Variation Suite v7.3.1 (Golden Helix, Bozeman, MT, www.goldenhelix.com).

*Case-control matrix overlap and informative markers*

The intersection of 351,507 markers (cases) with 317,503 markers (controls) yielded 305,491 common markers. A common matrix was created by appending the two files, that is dropping all columns (markers) that were not overlapping.
From the 305,491 markers present in the common (case and control) matrix 305,452 markers are informative, having two alleles (D and d) and 39 markers proved to be uninformative, having only one allele (DD or dd) in all genotyped individuals.

### 3.2.1.5 Pre-QC Analysis

In order to evaluate the effect of filtering analyses were performed prior and after each step of the quality control procedure. The first genotypic association test computed was the basic allelic test (BAT), where the amount of minor versus major alleles (D *versus* d) are compared in cases versus controls (the affection 'status': affected *vs.* unaffected being the dependent variable). The statistic used was Pearson's Chi-squared test.

*Q-Q plot*

The expected chi-squared values were also computed in order to generate a quantile-quantile (Q-Q) plot, which visually shows how far the observed values deviate from the expected values [490].

Where there is perfect concordance between the observed values and the expected ones, the intersection will be plotted on the X = Y line on the graph (where the observed values are projected on the X axis and the expected values are projected on the Y axis).

The (slope intercept) equation of the line is:

$$y = mx + b$$
$$f(x) = y$$

substituting in the slope of m = 1 and Y - intercept b = 0 one would get the line at an 45$^{o}$ angle for the ideal case (of perfect concordance).

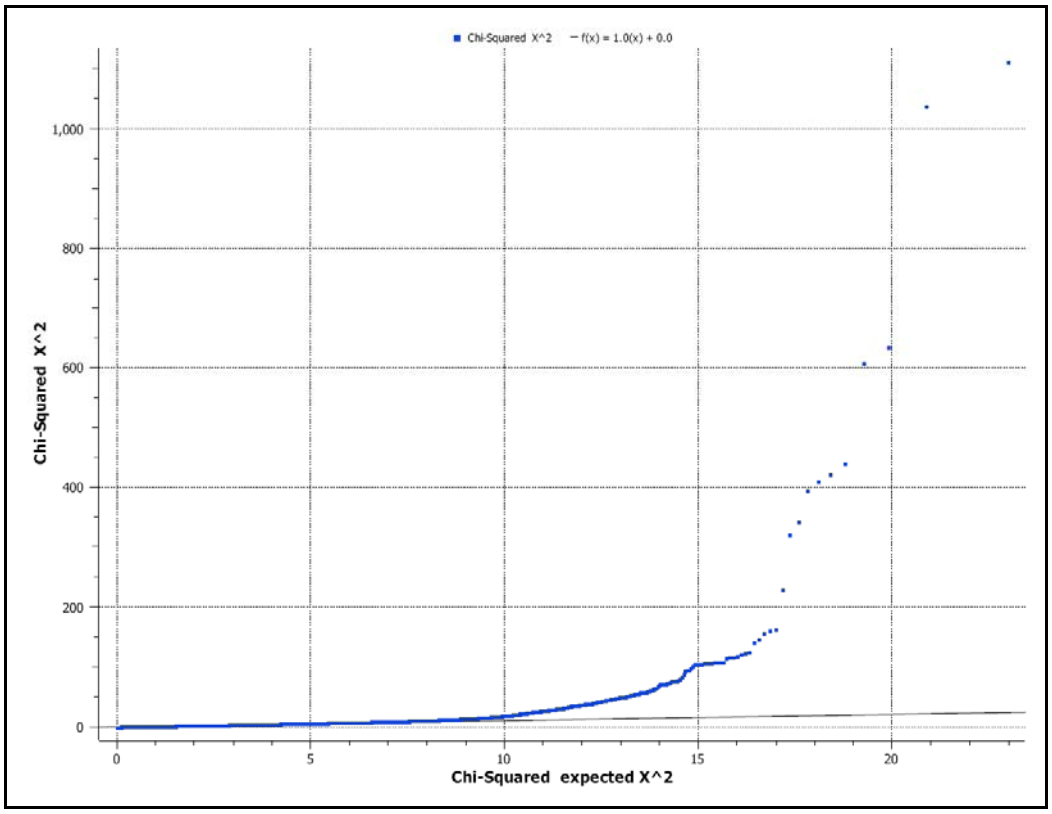


**Figure 19 Q-Q-plot for the UK dataset prior to quality control filtering**
Chi-squared test of the basic allelic analysis (D *versus* d) comparing minor *versus* major alleles in cases *versus* controls. Blue = Cartesian intersection point between observed and expected chi-squared values for each marker; black = concordance line corresponding to the null hypothesis (X = Y).

The overall adherence (over most of the range; see Figure 19) of observed points to the null hypothesis concordance line is an indicator for the absence of systematic sources of error (spurious association). The deviation from the null hypothesis concordance line observed for real datasets is explained by either spurious or real association.

### *Stratification*

As previously noted, the effect of population stratification is directly proportional to the number of samples used in the study. The number of samples in the present study being relatively small (684) the expected influence of stratification on the results is also small. To further minimize the possible influence of the

stratification effect, cases and controls were selected from the same population (British).

**Principal Components Analysis**

The extent of the stratification problem was studied by means of principal component analysis analysis (PCA; EIGENSTRAT implemented in SNP & Variation Suite v7.3.1). We calcuated the eigenvalues for each individual in the datasets studied for the assessment of membranous nephropathy association (British, French and Dutch) comparatively with a collection of four populations from three continents (Africa, Asia and Europe) which were included in the HapMap project [491] dataset.

The four populations (YRI, JPT, CHB and CEU) assessed in the HapMap project are considered to be illustrative for the genetic diversity of human populations. YRI is the shorthand notation for the Yoruba community from Ibadan, Nigeria on the African continent. The individuals genotyped have identified themselves as having four Yoruba grandparents. They cannot be considered in any way representative for the entire African continent, not even for West Africa, given the intricacies of multiple populations sharing that geographical areal. JPT stands for Japanese Tokyo, Asia. The people in this particular subset have been recruited from the metropolitan area of Tokyo, themselves or their ancestors coming from different parts of Japan. They are therefore considered to be representative of the majority population in Japan. CHB stands for Han Chinese in Beijing, China, Asia. The provenience of the genotyped individuals is the residential community at Beijing Normal University. They are self-identified as having at least three (out of four) Han Chinese grandparents. This subset can also not be considered representative of the entire Han Chinese population. CEU represents a a population of Utah Residents of Northern and Western European ancestry, which were collected in the 1980s by the Centre d'Etude du Polymorphisme Humain (CEPH). Because at that time the importance of the correct assessment of the origin of individuals included in the sample was underappreciated it is not very clear how accurately this reflects the patterns of genetic variation in the respective European area. The term Caucasoid is to be avoided due to its potentially racistic overtone and due to the fact that technically

Caucasoids are just the inhabitants of the region encompassed between the Black Sea and the Caspian Sea.

Up to 100 principal components were computed with EIGENSTRAT, using the additive model as the genetic model for principal component analysis (the additive model represents dd > Dd > DD).
The graphical representation of eigenvalues uses a scatter plot type called scree plot, which derives its name from the concave upvard image of the homonimous accumulation of rock fragments at the basis of a steep cliff.

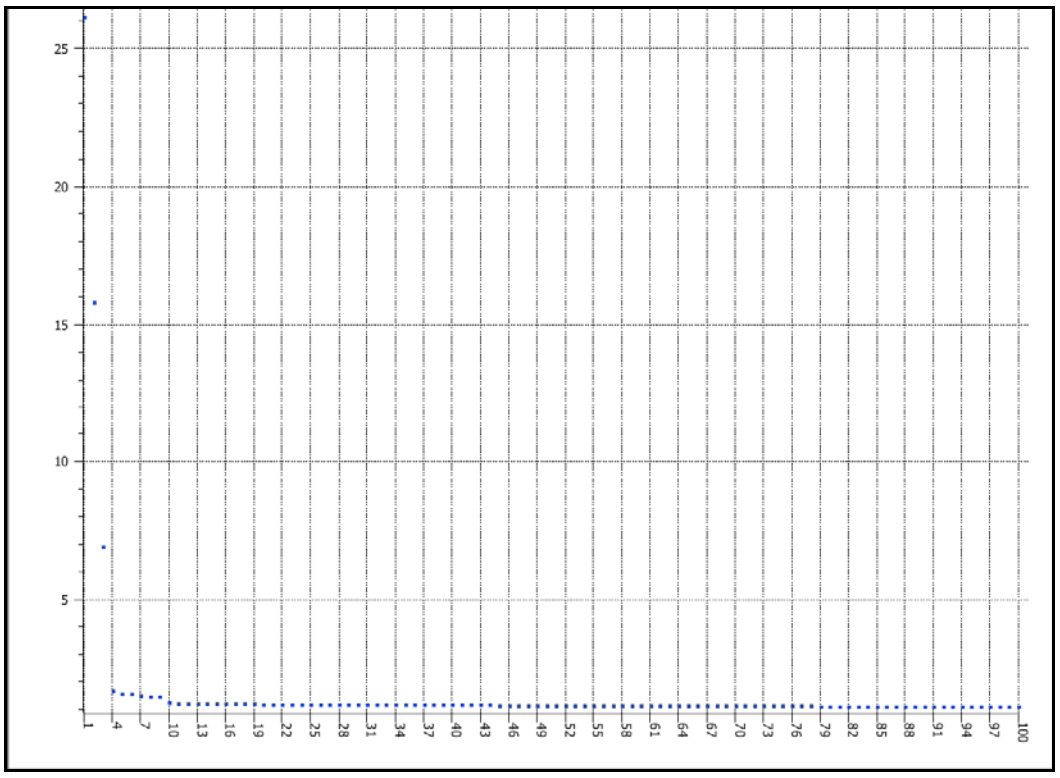


**Figure 20 Scree plot of the PC eigenvalues for six populations**
X axis: computed 100 principal components, Y axis: eigenvalues.

Comparing the eigenvalues for the known stratified dataset (six populations) with our initial case control set (British population) we get an estimate for the differences in stratification: the first 5 principal components computed for the stratified dataset being an order of magnitude higher than the first five principal components computed for the assessed population (British case and control set). Compare Figure 20 with Figure 21 and see Table 5.

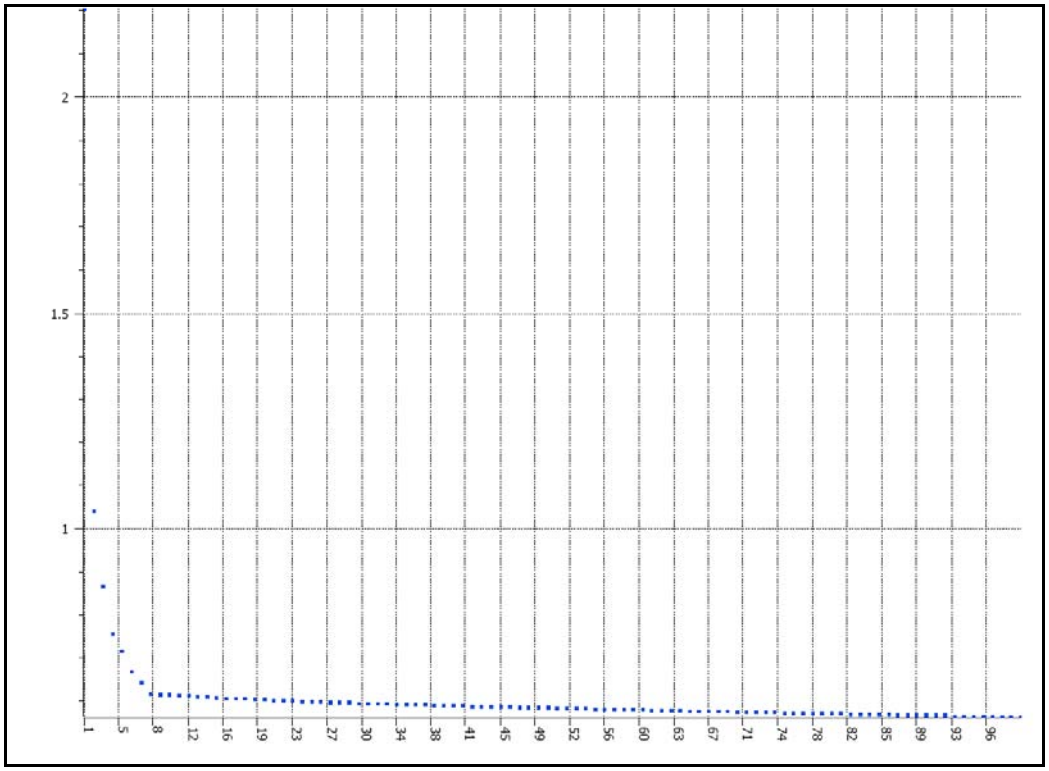


**Figure 21 Scree plot of PC eigenvalues for British cases versus British controls**
X axis: computed 100 principal components, Y axis: eigenvalues.

**Table 5 Eigenvalues of principal components (PC): UK *vs.* calibration set**
Comparison of the eigenvalues for the first five principal components (PC) in the case of the calibration dataset (including individuals of European, African and Asian origin) with the eigenvalues of the principal components for the UK dataset used in the present study.

| Eigenvalue of PC | control dataset | UK dataset |
|---|---|---|
| 1 | 26.16 | 2.20 |
| 2 | 15.83 | 1.04 |
| 3 | 6.92 | 0.87 |
| 4 | 1.69 | 0.76 |
| 5 | 1.58 | 0.72 |

We also compared the scattering of the eigenvalues for the individuals of the merged (six populations) dataset in a graphical manner. The distances between clouds of aggregation (individuals aggregate based on their geographic origin) reflect the amount of stratification observed. The first two principal components for each of the six populations were plotted in the following orthogonal projection (scatter plot), see Figure 22.

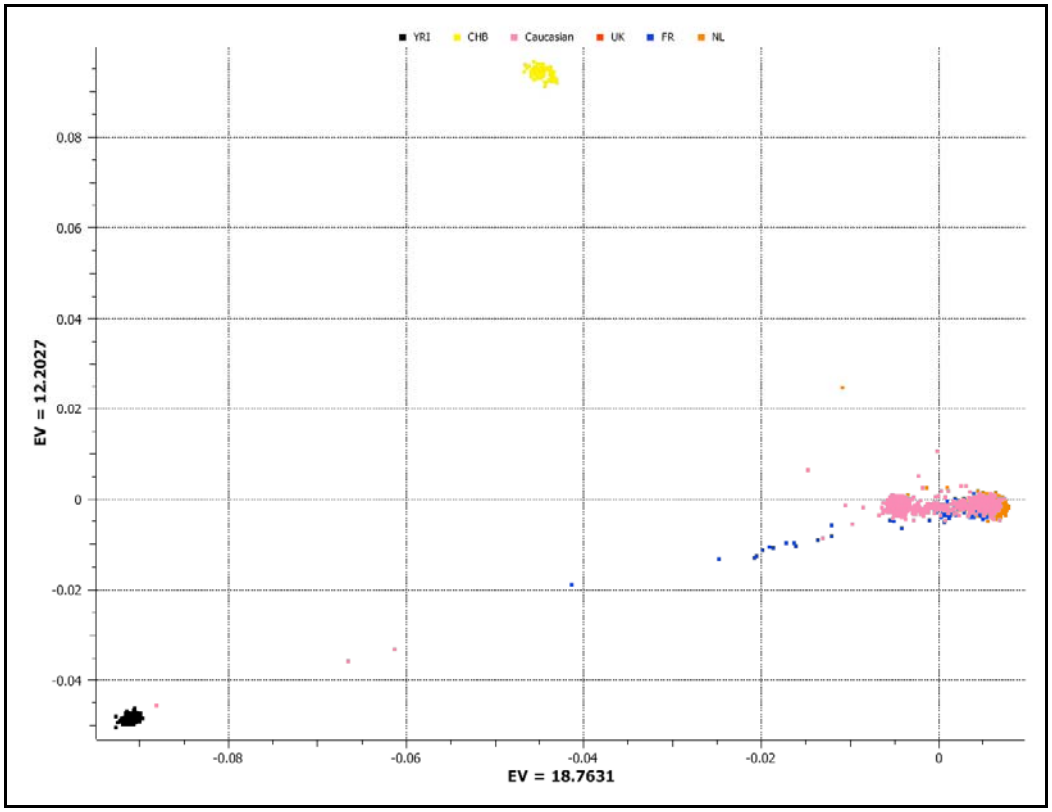


**Figure 22 'Calibrating' PC scatter plot**
EV = eigenvalue; on the abscisse the first principal component; on the ordinate the second principal component; YRI = Yoruban (in black); CHB = Han Chinese (in yellow); Caucasian = Caucasoid HapMap population (in pink); UK = British dataset (in red); FR = French dataset (in blue); NL = Dutch dataset (in orange). One observes the overall overlap of the three studied European populations with the Caucasoid HapMap dataset.

A similar plotting, only including the relevant area, was performed just for the British dataset, overlapping cases and controls to assess the differences between the scattering / grouping of cases *versus* controls (the two first principal components which also were plotted on the orthogonal projection), Figure 23.

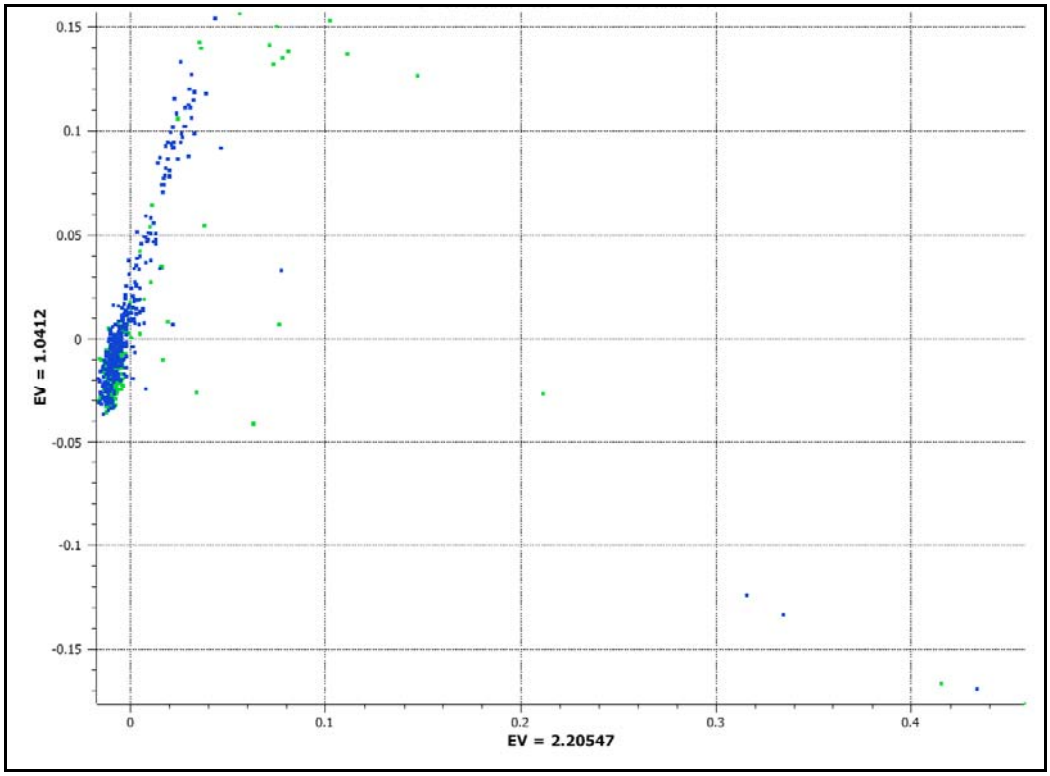


**Figure 23 UK case control PC scatterplot**
British dataset: EV = eigenvalue; on the abscisse the first principal component; on the ordinate the second principal component; cases (green); controls (blue) XY scatter plots.

To better visualize the overlap of the scattering / grouping of British cases *versus* British controls the two subsets have also been plotted separately, compare Figure 24 with Figure 25.

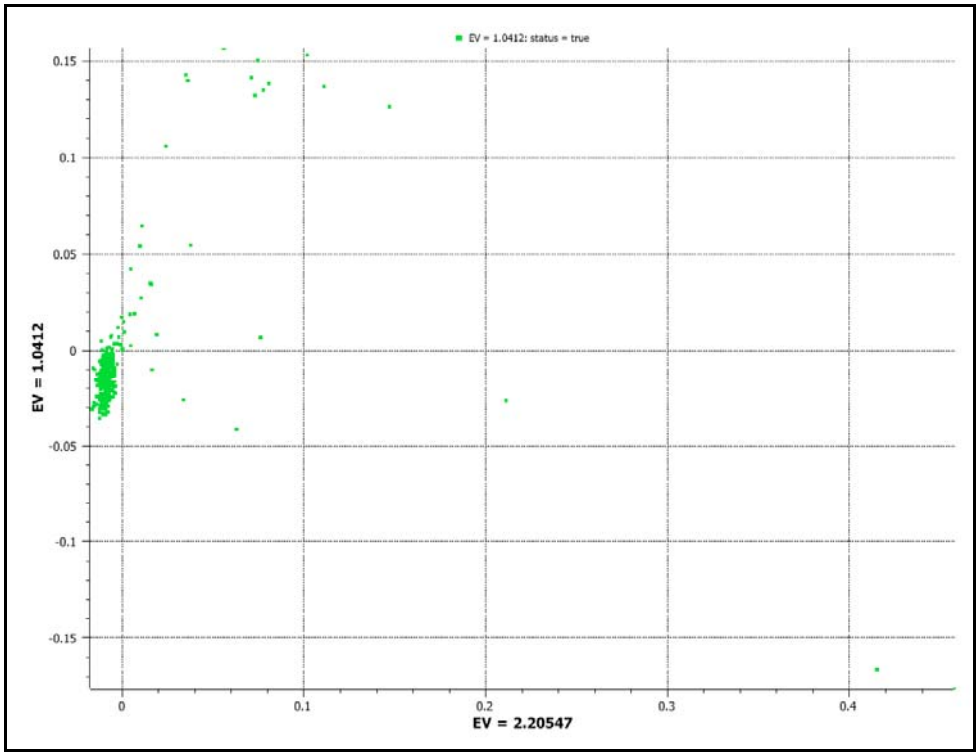


**Figure 24 UK cases PC scatterplot**
British cases dataset: EV = eigenvalue; on the abscisse the first principal component; on the ordinate the second principal component; cases (green) XY scatter plot.

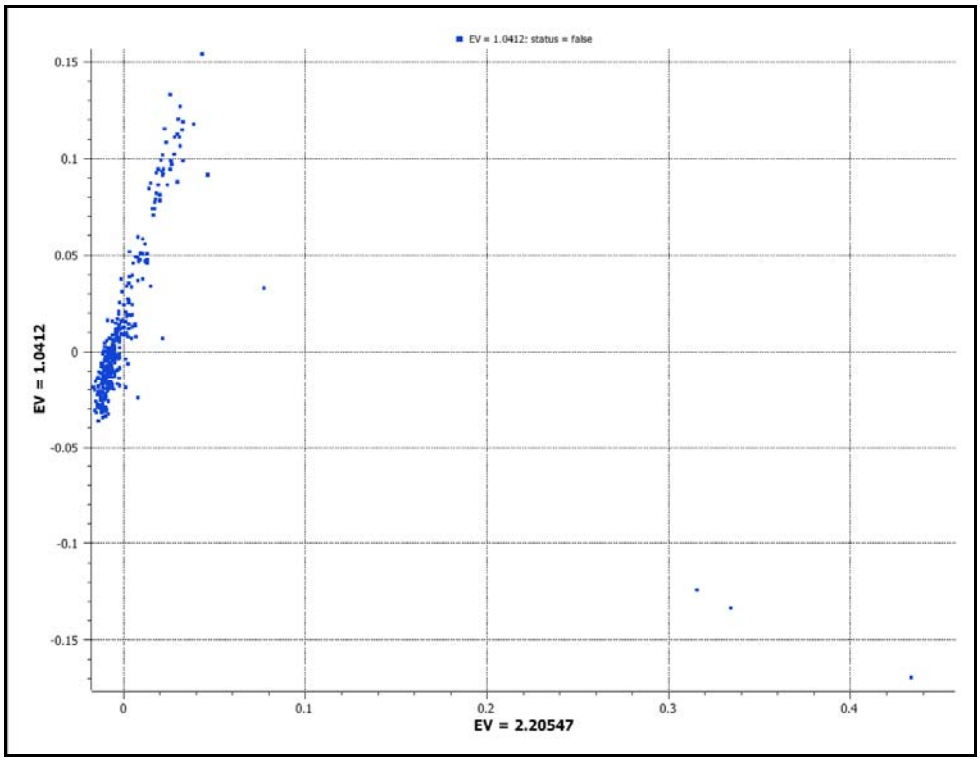


**Figure 25 UK controls PC scatterplot**
British controls: EV = eigenvalue; on the abscisse the first principal component; on the ordinate the second principal component; controls (blue) XY scatter plot.

### Genomic control

To control for possible stratification we applied the method of genomic control (*vide supra*), where the genomic control inflation factor ($\lambda$) for the UK dataset was found to be 1.16.

## *Results*

### Multiple testing

We applied the Bonferroni correction for multiple tests, dividing the conservative linkage significance level $P = 0.001$ (a more stringent significance level than the usual 0.05, the equivalent of LOD = 3, as suggested by Smith in 1953 [391]) by N, the number of tests which is the number of informative markers, in our case 305,452.

### Associated SNPs

After Bonferroni correction, 190 markers were shown to be associated with the trait ('status') above a significance level of Chi-Squared $-\log_{10}P$ of 8.48 ($P < 3.27 \times 10^{-9}$). When further correcting for stratification (applying the lambda – genomic inflation correction, that is dividing $P$ / lambda = 1.16 thus the corrected $-\log_{10}P$ becoming 8.54), 188 markers were still associated with the trait.

### Manhattan plot

The results were graphically presented in form of a Manhattan plot, which is a type of scatterplot named after the famous New York skyline. The first time a Manhattan plot was used to represent GWAS results was the Klein *et al.* macular degeneration paper of 2005 [438].

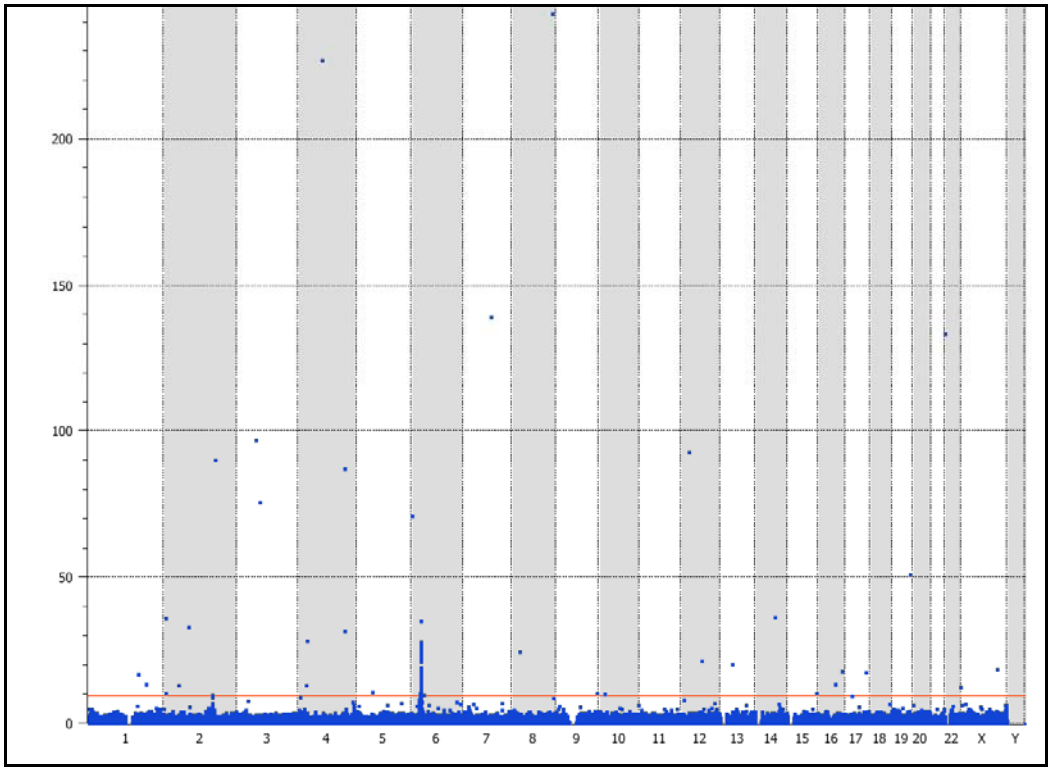


**Figure 26 BAT Manhattan plot, UK dataset prior to QC filtering**
Chi-squared test of the basic allelic test (BAT) comparing minor *versus* major alleles (D *versus* d) in cases *versus* controls. The level of significance (corrected for multiple comparisons and for genomic inflation) is represented by the red line at $-\log_{10}P$ of 8.54. (Y axis = $-\log_{10}P$, X axis = physical positions of the markers corresponding to the human February 2009 GRCh37/hg19 genome assembly, the chromosomes being plotted successively and shaded alternatively for better visualisation).

### 3.2.1.6 Quality Control

*Genotype quality control (QC)*

**Together (on merged case-control matrix)**

The first quality control procedures were initially performed as suggested [492] under the assumption that the genotyping was carried out under similar conditions (*i.e.* at the same time, in the same laboratory, ideally in the same batch etc) for cases and controls together. For the UK cohort the use of historical controls countradicted this strict assumption. We call the matrix that undervent the first quality control experiment the *common* (case and control) matrix. That means the two initial matrices (cases and controls) have been merged (the rows representing cases have been appended / added to the rows representing controls).

*First step: call rate per marker*

The quality control of the measured call rate per individual is a proxy for the quality of DNA genotyped. Some of the markers might present intrinsic problems (poor hybridization due to design, local nucleotide composition etc) resulting in a low call rate at a given locus. To assess the influence of these 'bad' markers we look at the minimal call rate per marker was measured, the results being as follows (for the complete case-control UK dataset): 109,927 markers were completely genotyped (call rate = 100%), 169,470 markers were genotyped at a call rate ≥ 99%; as plotted in the following histogram (Figure 27).

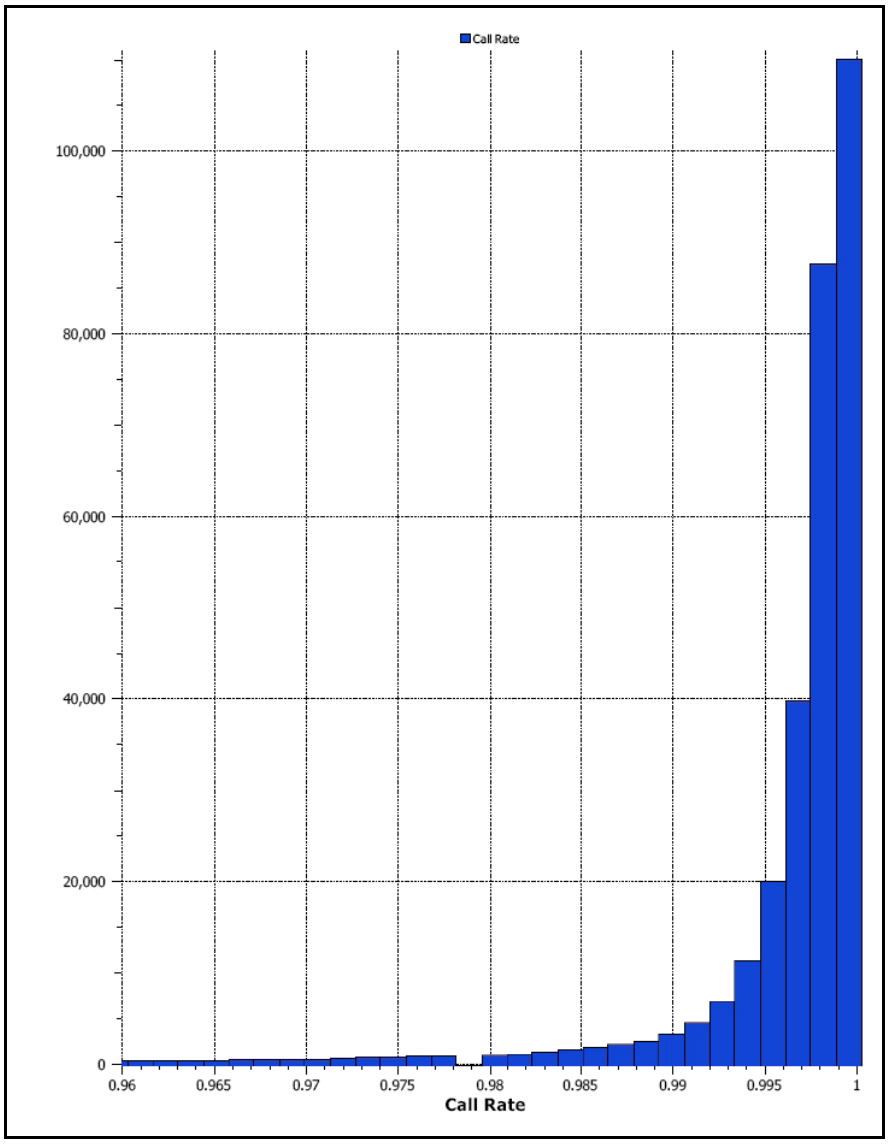


**Figure 27 Distribution of call rates per marker for the complete UK dataset**
X axis: call rate, Y axis: marker number.

To assess the differences between DNA quality of cases *versus* controls we plotted the same histogram for the two separate matrices and evaluated the overlay by using different colours (Figure 28).

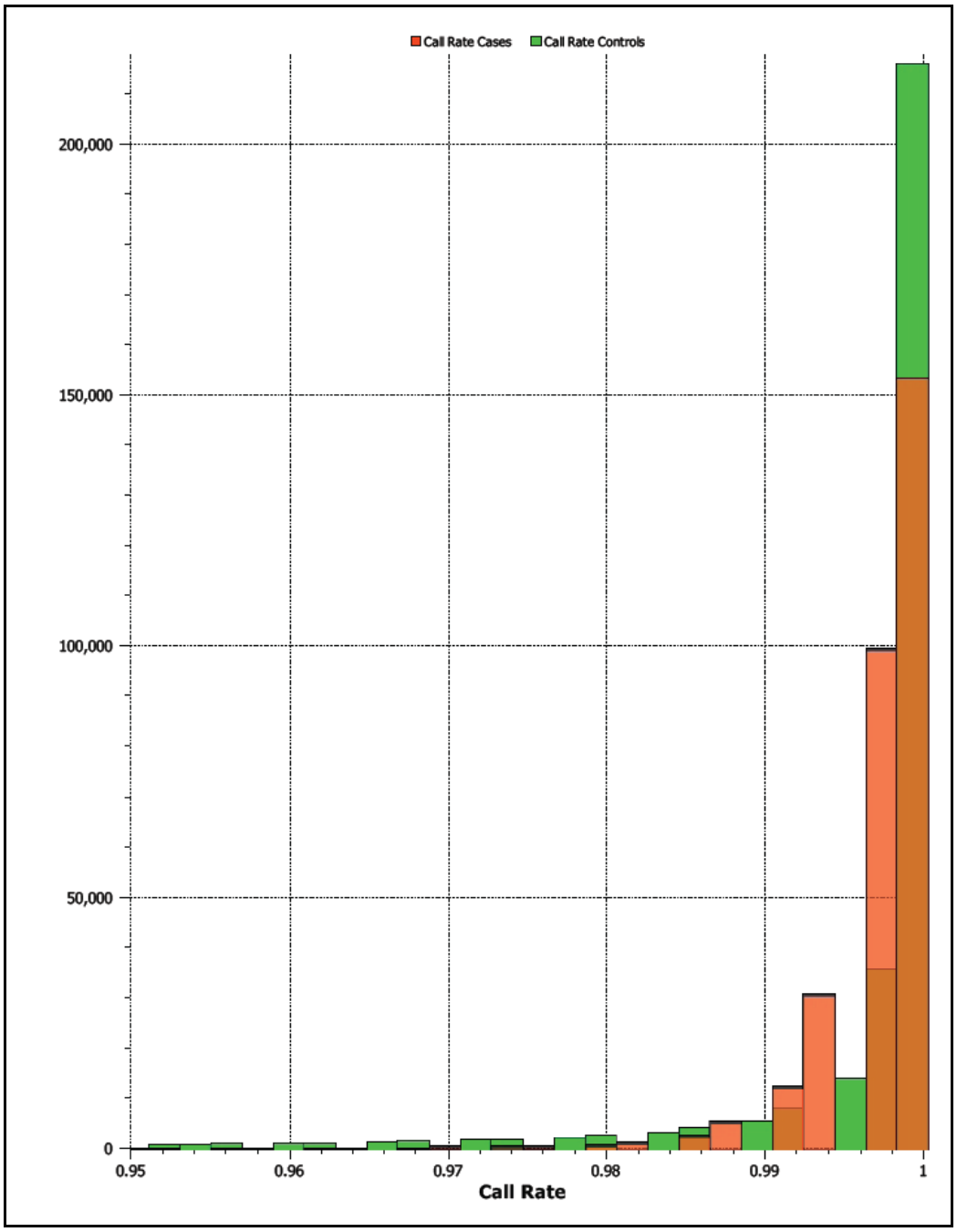


**Figure 28 Distribution of call rates per marker, UK dataset**
Cases (red) *versus* controls (green). Brown represents the overlap of the two instances. The genotyping quality of controls is shown to be better than that of cases.

The conclusion is that there is an overal good concordance between the two distinct datasets, yet the genotyping quality of controls is shown to be better than that of cases.

We went on to perform different analyses using the markers that have been called at different stringencies: 100% (maximum stringency), 96% (low stringency) and 99% (high stringency).

*Call rate per marker = 100% (maximum stringency)*

From 305,452 informative markers 109,909 (35.98 %) have a call rate of 100%.

The basic allelic test (BAT) was computed using the Pearson's Chi-squared test. The expected chi-squared values were computed and the quantile-quantile (Q-Q) plot was generated (see Figure 29).

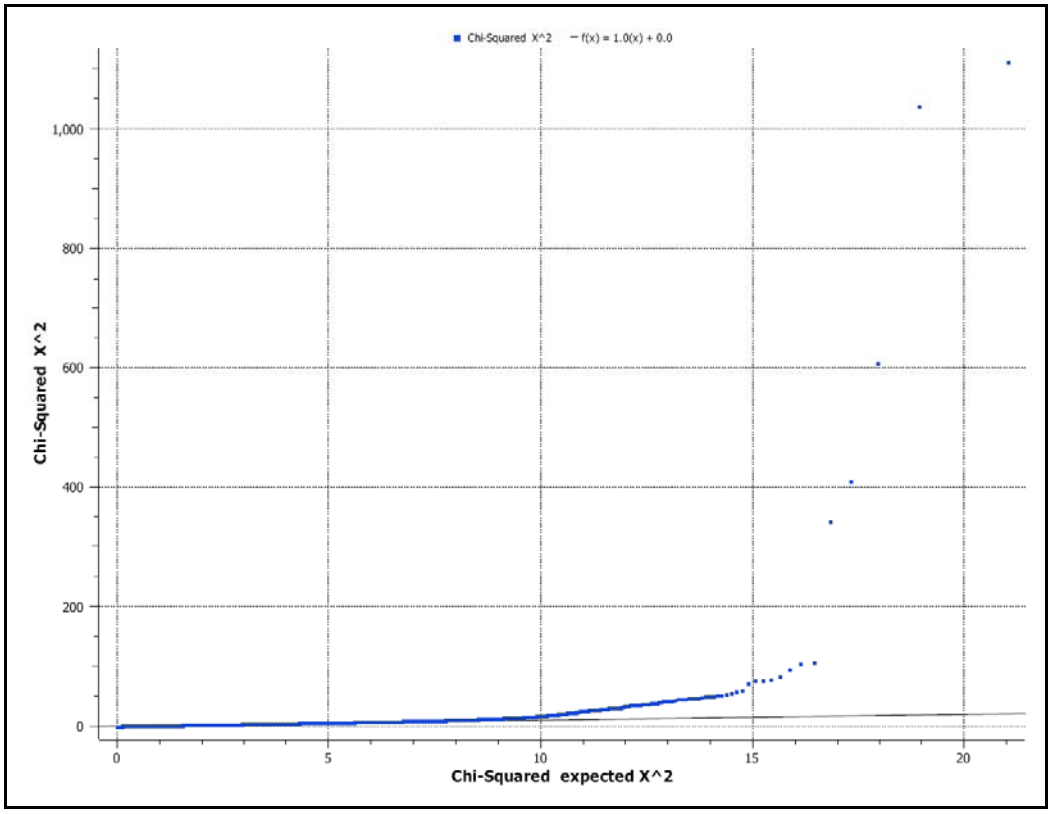


**Figure 29 Q-Q-plot for the UK dataset, maximum stringency scenario**
Q-Q plot for the Chi-squared test of the basic allelic analysis (D *versus* d) comparing minor *versus* major alleles in cases *versus* controls for the UK dataset after quality control filtering for 100% call rate per marker (maximum stringency).

The genomic inflation factor (λ) was found to be 1.14 (down 0.02 from no quality control).

Results

The Bonferroni correction for multiple tests (P = 0.001 / N = 109,909) yields the significance level of Chi-Squared $-\log_{10}P$ of 8.04 ($P < 9.09 \times 10^{-9}$). At this

significance level 60 markers were shown to be associated with the dependent variable ('status'). This is a substantial (138 markers, 69.69%) reduction from the 198 associated markers in the non quality control scenario. When correcting for stratification (P / lambda = 1.14 thus the corrected $-\log_{10}P$ becoming 8.09) there is no further reduction of the initial 60 associated markers.

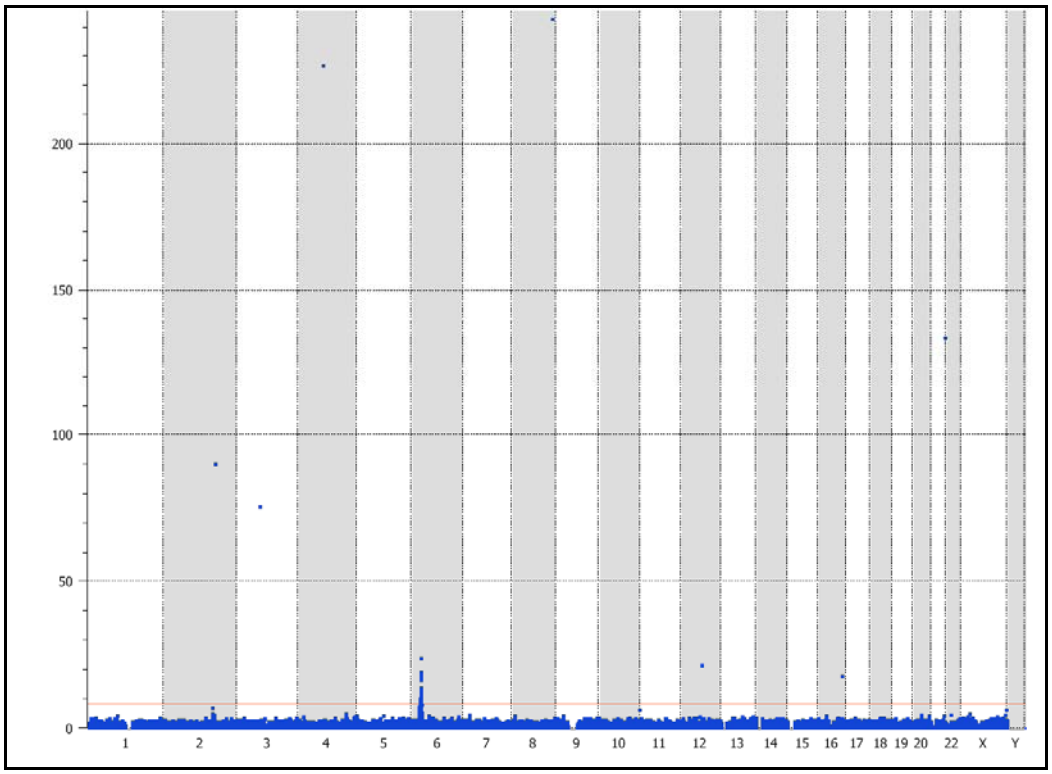


**Figure 30 BAT Manhattan plot, UK dataset, maximum stringency scenario**
Manhattan plot for the Chi-squared test of the basic allelic analysis (BAT) comparing minor *versus* major alleles (D *versus* d) in cases *versus* controls for the UK dataset after quality control filtering for 100% call rate per marker (maximum stringency). The level of significance (corrected for multiple comparisons and for genomic inflation) is represented by the red line at $-\log_{10}P$ of 8.09. (Y axis = $-\log_{10}P$, X axis = physical positions of the markers corresponding to the human February 2009 GRCh37/hg19 genome assembly, the chromosomes being plotted successively and shaded alternatively for better visualisation).

*Call rate per marker = 96% (low stringency)*

From 305,452 informative markers 298,178 (97.61 %) have a call rate of greater than 96 %. The basic allelic test (BAT) was computed using the Pearson's Chi-squared test statistic. The significance threshold $P = 0.001$ was corrected for multiple comparisons ($N = 298{,}178$) and for genomic inflation (lambda = 1.15) at $-\log_{10}P = 8.53$. At this threshold 159 markers were found to be associated with the trait (as compared to the 60 associated markers for the maximum stringency

analysis). This represents a reduction from the non quality control scenario (198 associated markers) of 39 markers (19.69%).

*Call rate per marker = 99% (high stringency)*

To optimize the threshold a conservative intermediary scenario at 99% call rate per genotype was assessed. From the initial 305,452 informative markers 26,094 did not satisfy the condition, 279,361 (91.45 %) did. The basic allelic test (BAT) was computed using the Pearson's Chi-squared test statistic. The significance threshold $-\log_{10}$ became 8.5 for $P = 0.001$ corrected for multiple comparisons ($N = 279,361$) and for the genomic inflation factor (lambda = 1.14. At this threshold 145 markers were found to be associated with the trait (a value situated, as expected, between the 60 associated markers for the maximum stringency analysis and the 159 associated markers of thelow stringency analysis, closer to the latter). The reduction from the non quality control scenario (198 associated markers) is 53 markers (26.77%).

*Second step: minor allele frequency (MAF)*

The minor allele defined as being the allele with the lowest frequency at a given locus. The reasons why it is suggested to drop markers with minor allele frequencies less than a given threshold are threefold: rare alleles are likely to be genotyping errors, when using the Chi-squared test one might overestimate the significance of differences for small counts, and low frequency minor allele markers have a low power to detect association. One can add to that the fact that multiple testing is going to be reduced by reducing the number of markers.

To evaluate the distribution of minor allele frequencies in cases, controls and the common case-control matrix we plotted the number of markers for each minor allele frequency in the following histograms, observing similar distribution patterns for all three matrices, Figure 31.

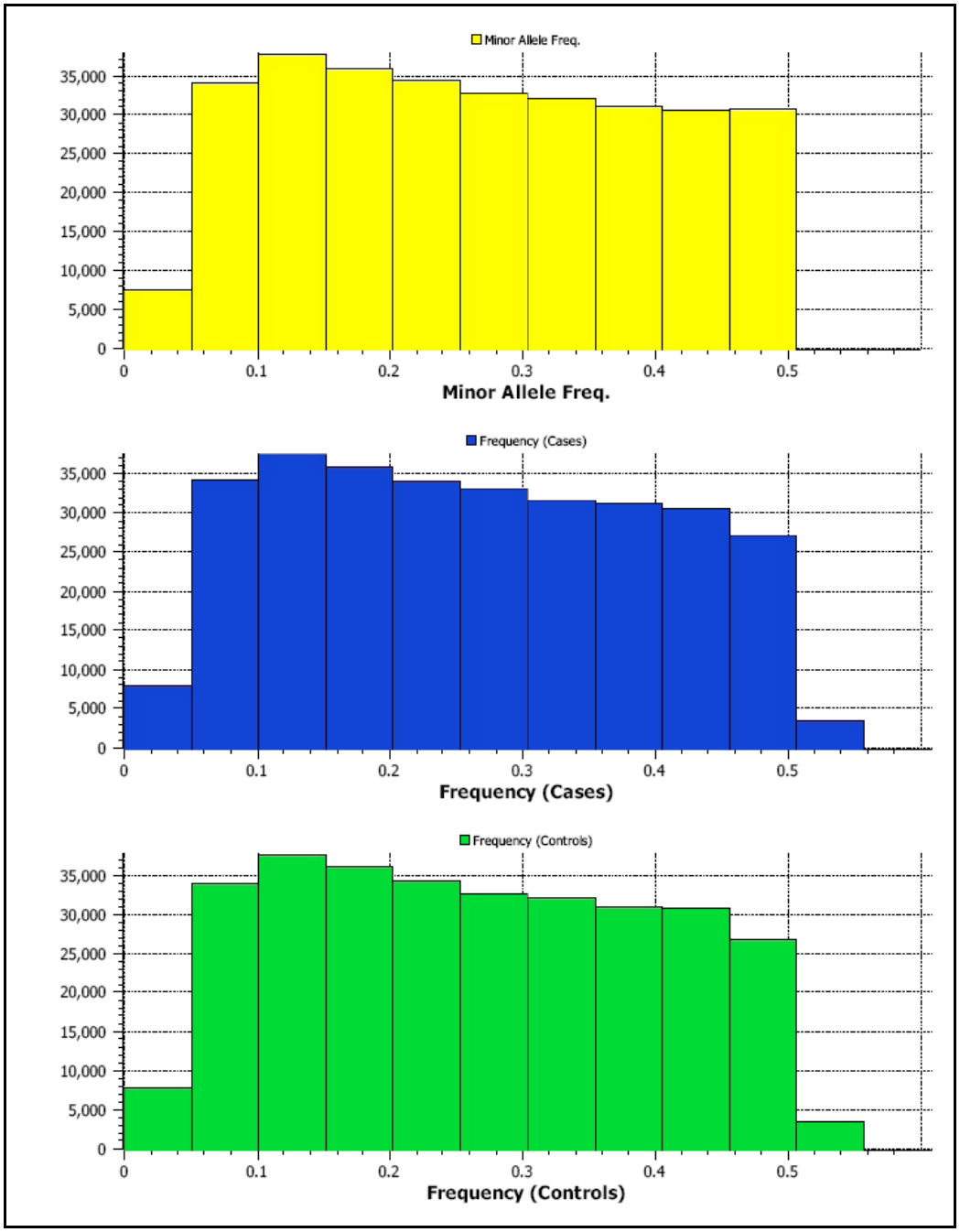


**Figure 31 Distributions of minor allele frequencies, UK dataset**
Histograms of the minor allele frequencies (MAFs) for the whole UK study set (yellow), for the cases (blue) and for the control (green). Note the quasi-similarity of cases *versus* controls.

In order to assess the influence of dropping markers with different minor allele frequencies two scenarios were compared: a minimal one in which all markers with a minor allele frequency lesser than 0.01 (1% being the arbitrary threshold for the definition of a polymorphic locus – therefore we exclude all rare variants; rare variants being defined as any alleles rarer than this) and a maximal scenario in which the threshold for minimal allele frequency exclusion was set at 0.1.

*Minimal drop MAF < 0.01 (exclusion of rare variants)*

According to the 0.01 MAF definition 179 markers were found to be rare variants. None of them was amongst the ones (188 markers in the previous – no quality control - analysis) found to be associated with the trait at *P* corrected for multiple comparisons (Bonferroni) and for genomic inflation (lambda) $-\log_{10}P$ being 8.54. The genomic control inflation factor (lambda) was found to be 1.16 for Chi-Squared. Thus at this threshold MAF filtering is not influencing the results (except for minimally reducing the number of comparisons with 179).

*Maximal drop MAF < 0.1*

At the maximal threshold of 0.1 (10% MAF) 41,469 markers were dropped of which 8 were associated with the trait in the previous analysis (180 *versus* 188 at $-\log_{10}P$ 8.54), 182 when the threshold is lowered according to the new number of multiple comparisons ($0.001/264{,}022/1.16 = 3.2 \times 10^{-9}$, $-\log_{10}P$ being 8.48).

*Medium drop MAF < 0.05*

To optimize the number of markers that can be dropped as non informative due to the low minor allele frequency in order to reduce the number of comparisons, a MAF threshold between 0.01 (minimal drop) and 0.1 (maximal drop) was chosen conservatively at 0.05 (5%). At the medium threshold of 0.05 (5%) 7,534 markers were dropped, 3 of which being associated with the trait in the previous analysis (185 *versus* 188 at $-\log_{10}P$ 8.54). Given that the reduction in the number of comparisons (with 7,534 leading to $-\log_{10}P = 7.54$) is obtained at the cost of only 3 associated markers being filtered out, the decision was reached to use the minimal drop threshold for the analysis.

*Third step: Hardy-Weinberg equilibrium (HWE) filtering*

Associated markers can be in disequilibrium due to *e.g.* them reflecting a copy number variant, CNV. To minimize the risk of *bona fide* associations to be excluded due to HWE filtering one can adopt two strategies: either apply the HWE filter only on the *controls* matrix (thus not excluding CNV markers or their SNP proxies that are overrepresented in the *cases* matrix); alternatively one can apply the strictest filtering criterion (cases and controls together), checking the

excluded markers for biological relevance *a posteriori*. We first explored the latter approach.

*Distribution of HWE P*

Hardy-Weinberg equilibrium *P* values were plotted for all the markers in the UK dataset (cases and controls together) before any other quality filtering measure, Figure 32.

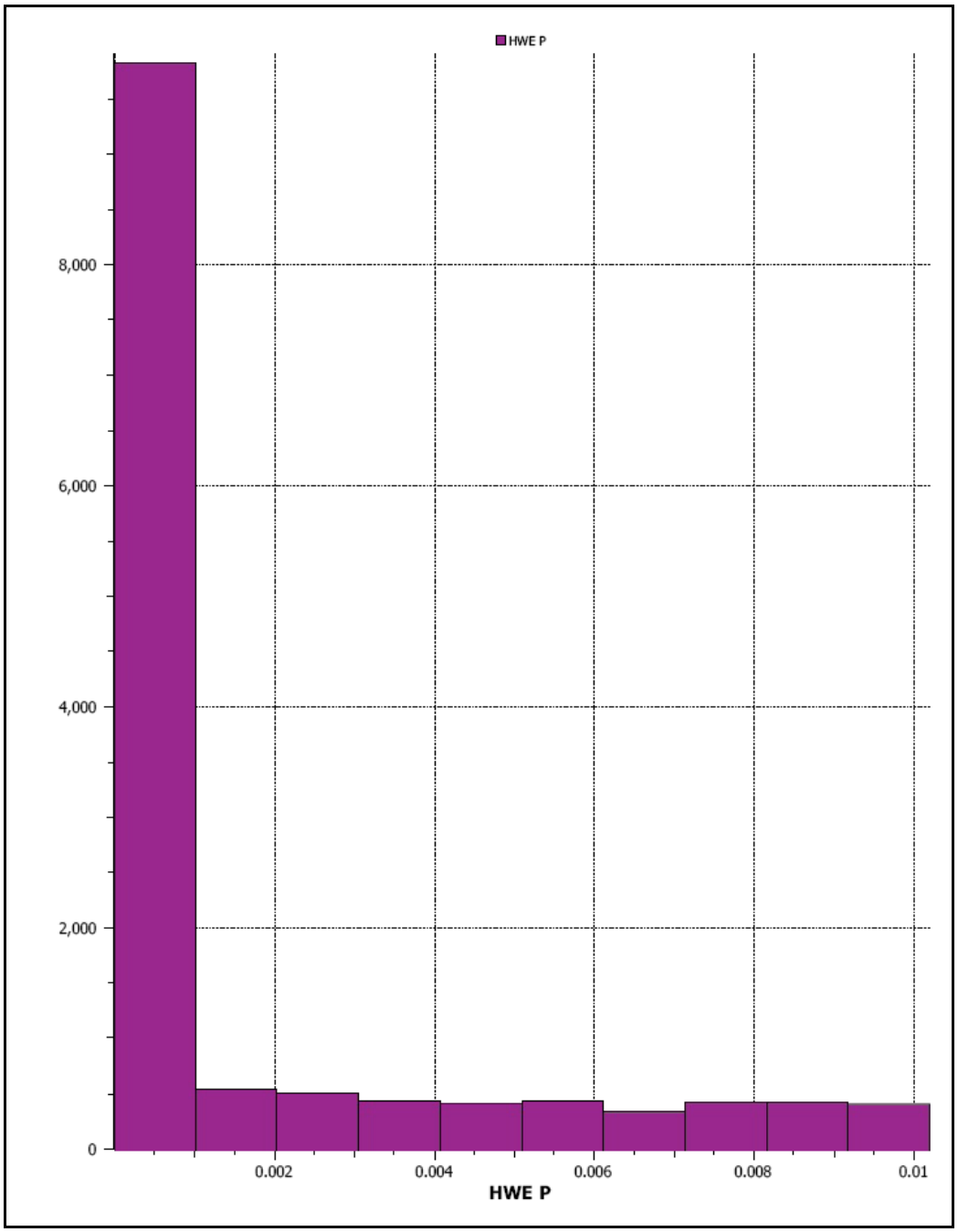


**Figure 32 Distribution of Hardy-Weinberg Equilibrium *P* values, UK dataset**
X axis: Hardy-Weinberg Equilibrium *P* values; Y axis: number of markers.

*Low HWE ($P = 10^{-4}$)*

Hardy-Weinberg equilibrium filtering was based on all samples (as opposed to cases only or controls only). At the low threshold of the *P* value for Hardy-Weinberg equilibrium $10^{-4}$, 9,060 markers were dropped of which 17 were associated with the trait in the non filtered analysis (171 *versus* 188 at $-\log_{10}P$ 8.54). The genomic inflation factor for this analysis is 1.15 for chi-squared. When the threshold is lowered according to the new number of multiple comparisons (0.001 / 296,329 / 1.147 = $2.94 \times 10^{-9}$, $-\log_{10}P$ being 8.53) one finds 172 associated markers.

*High HWE ($P = 10^{-3}$)*

Hardy-Weinberg equilibrium filtering was based on all samples (cases and controls in the same matrix). At the high threshold of the *P* value for HWE $10^{-3}$, 9,816 markers were dropped of which 18 were associated with the trait in the non filtered analysis (170 *versus* 188 at $-\log_{10}P$ 8.54). The genomic inflation factor for this analysis is 1.15 for chi-squared. When the threshold is lowered according to the new number of multiple comparisons (0.001 / 295,636 / 1.15 = $2.95 \times 10^{-9}$, $-\log_{10}P$ being 8.53) one finds 171 associated markers. One marker when compared to the low *P* value scenario (171 high *P versus* 172 low *P*).

*Extreme HWE ($P = 10^{-2}$)*

Hardy-Weinberg equilibrium filtering was performed on all samples (cases and controls together). At the high threshold of the P value for HWE $10^{-4}$, 13.614 markers were dropped of which 20 were associated with the trait in the non filtered analysis (168 *versus* 188 at $-\log_{10}$P 8.54). The genomic inflation factor for this analysis is 1.15. When the threshold is lowered according to the new number of multiple comparisons (0.001 / 291,838 / 1.146 = $2.99 \times 10^{-9}$, $-\log_{10}P$ being 8.52) one finds 169 associated markers (*versus* 171 for high HWE-*P* respectively 172 low HWE-*P*).

**Table 6 Markers excluded based on HWE QC filtering**
Columns: low (HWE *P* threshold at 0.0001): 17 markers excluded; medium (HWE *P* threshold 0.001): 18 markers excluded; extreme (HWE *P* threshold 0.01) 20 markers excluded. The excluded markers are represented in red.

| # | $-\log_{10}P$ | marker | chromosome | position | Low | High | Extreme |
|---|---|---|---|---|---|---|---|
| 1 | 242.8945505 | rs1036819 | chr8 | 135611945 | | | |
| 2 | 227.0322739 | rs1455311 | chr4 | 79964587 | | | |
| 3 | 139.5238956 | rs17400329 | chr7 | 90595695 | | | |
| 4 | 133.5400546 | rs2236479 | chr21 | 46919132 | | | |
| 5 | 96.9885534 | rs6781958 | chr3 | 60724390 | | | |
| 6 | 93.03622721 | rs1975920 | chr12 | 27735986 | | | |
| 7 | 90.25535638 | rs1549343 | chr2 | 171093785 | | | |
| 8 | 87.24764677 | rs2406233 | chr4 | 154034573 | | | |
| 9 | 75.73985376 | rs1447826 | chr3 | 74718758 | | | |
| 10 | 70.93781263 | rs2745566 | chr6 | 1558640 | | | |
| 11 | 51.25158476 | rs3745902 | chr19 | 55378008 | | | |
| 12 | 36.47536951 | rs1959034 | chr14 | 64336479 | | | |
| 13 | 36.25093906 | rs3021387 | chr2 | 8492923 | | | |
| 14 | 35.28596423 | rs2187668 | chr6 | 32605884 | | | |
| 15 | 33.11544641 | rs4393782 | chr2 | 82499876 | | | |
| 16 | 31.89819588 | rs361147 | chr4 | 152890563 | | | |
| 17 | 28.30104237 | rs9291561 | chr4 | 31228884 | | | |
| 18 | 27.91840586 | rs1150752 | chr6 | 32064726 | | | |
| … | | | | | | | |
| 38 | 21.93796759 | rs3129963 | chr6 | 32380208 | | | |
| 39 | 21.70563827 | rs7137203 | chr12 | 69548159 | | | |
| 40 | 21.42088638 | rs7750641 | chr6 | 31129310 | | | |
| 41 | 20.32632332 | rs4943552 | chr13 | 38504142 | | | |
| 42 | 19.25962881 | rs7762279 | chr6 | 32755290 | | | |
| 43 | 18.50487324 | rs7057244 | chrX | 121820344 | | | |
| 44 | 18.1424215 | rs1480380 | chr6 | 32913246 | | | |
| 45 | 17.97273658 | rs2857595 | chr6 | 31568469 | | | |
| 46 | 17.8204442 | rs2221705 | chr16 | 79362411 | | | |
| … | | | | | | | |
| 113 | 10.80170435 | rs2844657 | chr6 | 30829522 | | | |
| 114 | 10.80009518 | rs3923330 | chr5 | 53727092 | | | |
| 115 | 10.78783856 | rs2395174 | chr6 | 32404878 | | | |
| … | | | | | | | |
| 170 | 8.878890452 | rs2076537 | chr6 | 32317635 | | | |
| 171 | 8.855174724 | rs3911702 | chr8 | 137751040 | | | |
| 172 | 8.824183304 | rs1233579 | chr6 | 28712663 | | | |
| … | | | | | | | |
| 188 | 8.553166436 | rs7745603 | chr6 | 27090404 | | | |

*Filtered markers check*

Given that deviations from Hardy-Weinberg equilibrium can reflect true association, the markers that are filtered out were carefully assessed and annotated, with particular attention being paid to the presence of possible deletions or duplications. The name of each of the 20 markers, its physical position and the name of the gene it coincides with are given in the columns of Table 7.

**Table 7 Genes close to HWE filtered markers**
The presence of genes in the vicinity (1000 base pairs upstream and 1000 base pairs downstream) of the marker has been assessed (Annotation was performed in the UCSC human genome browser on the human February 2009 GRCh37/hg19 genome assembly).

| marker | chromosome | position | gene |
|---|---|---|---|
| rs1549343 | chr2 | 171093785 | *MYO3B* |
| rs3021387 | chr2 | 8492923 | none |
| rs1447826 | chr3 | 74718758 | none |
| rs6781958 | chr3 | 60724390 | *FHIT* |
| rs1455311 | chr4 | 79964587 | *BC036848* |
| rs2406233 | chr4 | 154034573 | none |
| rs361147 | chr4 | 152890563 | 100bp_*BC035179* |
| rs9291561 | chr4 | 31228884 | none |
| rs3923330 | chr5 | 53727092 | none |
| rs2745566 | chr6 | 1558640 | none |
| rs2857595 | chr6 | 31568469 | none |
| rs7762279 | chr6 | 32755290 | none |
| rs17400329 | chr7 | 90595695 | *PFTK1* |
| rs1036819 | chr8 | 135611945 | *ZFAT* / *ZFATAS* |
| rs3911702 | chr8 | 137751040 | none |
| rs1975920 | chr12 | 27735986 | *PPFIBP1* |
| rs7137203 | chr12 | 69548159 | none |
| rs1959034 | chr14 | 64336479 | *SYNE2* |
| rs3745902 | chr19 | 55378008 | *KIR3DL2*_duplication |
| rs2236479 | chr21 | 46919132 | *COL18A1* |

The most interesting occurrence (given that HW disequilibrium can reflect associated copy number variants, CNV) is that of marker rs3745902 on chromosome 19 (chr19:55378007-55378008) within the KIR3DL2 (killer cell immunoglobulin-like receptor, three) gene duplication.

KIR3DL2

*KIR3DL2* is a gene which lies within human killer cell immunoglobulin-like receptors (KIR) super locus [493], also called leukocyte receptor complex (LRC), a stretch of 1Mb sequence on chromosome 19q13.4. The KIR family of proteins are transmembrane glycoproteins classified by their number of extracellular immunoglobulin domains (2D and 3D) and by the length of their cytoplasmic domain (L - Long or S - Short). The long cytoplasmatic domain contains an Immune Tyrosine-based Inhibitory Motif (ITIM), therefore the L type KIRs transduce inhibitory signals upon ligand binding (whereas S type KIRs transduce activatory signals). KIRs are expressed by natural killer (NK) cells and subsets of T lymphocytes. There are two sets of NK inhibitory and activating receptor genes [494] and they belong either to the Ig superfamily (as does the KIR family, implicitly *KIR3DL2*) or to the C-type lectin superfamily (as does PLA2R).

### 3.2.1.7 Combining filtering criteria

Having gotten a feeling for how each filtering step does influence the outcome of the analysis for the dataset, different scenarios were tried combining different filtering thresholds for two cases: quality control on the *combined* (cases and controls together) matrix and quality control on two separated matrices.

*Cases and controls together QC filtering (together scenario)*

**Together low stringency scenario (call rate per marker > 96%)**

In this scenario, the parameters used for the combined quality control filtering of genotype (by marker) were: a call rate lower than 0.96; a minor allele frequency (MAF) lower than 0.01; a Hardy-Weinberg Equilibrium (HWE) *P* lower than 0.001. The filtering threshold thus defined was not passed by 16,827 markers (5.5 %; of the informative markers 305,452 = 100%; 39 markers were found to be monoallelic, therefore noninformative, from the total number of 305,491 markers). The markers that were filtered out are distributed as follows: 7,275 filtered due to the call rate; 179 filtered due to the minor allele frequency; 9,816 filtered due to the HWE *P*. *cave*: some markers did not pass more than one of the

filtering criteria, therefore 7275 + 179 + 9816 = 17,270 which is greater than the total number of markers (16,827).

*Population Stratification Detection*

As in the previous (non QC analyses), stratification was dealt with by the method of genomic control (structured association), the inflation factor (λ) computed for the given filtering conditions was found to be 1.14.

*Q-Q plots*

We used the Q-Q plot as a visual estimate of the quality of our dataset. The actual association *P* was plotted against the expected *P*. For good quality data, the plot follows the line y = x. Not only poor quality markers, but also those that are associated (true positives) will depart from y = x. The true positive associated markers will aggregate at the upper right of the plot [495] (Figure 33).

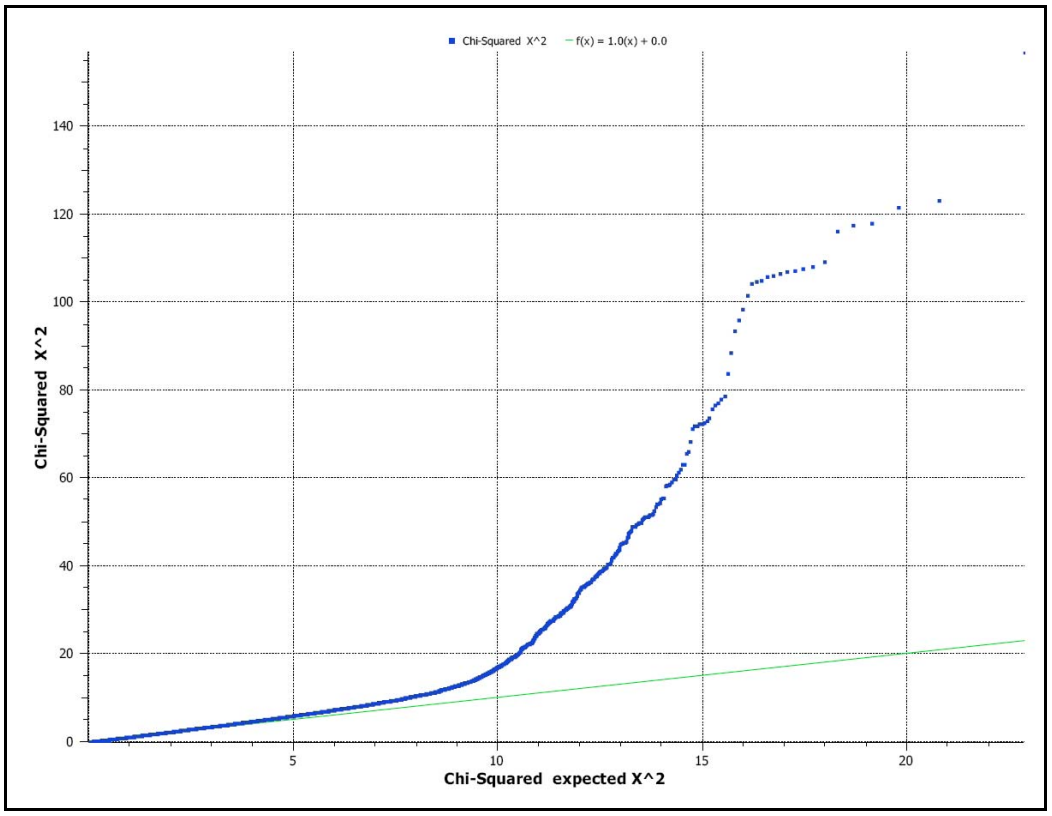


**Figure 33 Q-Q-plot for the UK dataset, together low stringency scenario**
Q-Q plot for the Chi-squared test of the basic allelic analysis (D *versus* d) plotting the actual association $-\log_{10}P$ to the expected $-\log_{10}P$ in cases *versus* controls for the UK dataset after the low stringency (call rate > 96%, minor allele frequency > 0.01, Hardy-Weinberg equilibrium $P > 0.001$) quality control filtering (288,664 markers, of which 288,625 informative).

Given that large regions of high significance (the Major Histocompatibility / Human Leucocyte Antigene HLA region) have a large influence on the Q-Q plot it has been suggested to exclude the markers of this region and to compare the plot thus obtained to the previous one. The HLA region has been defined according to Horton *et al.* [226] (*i.e.* a region of approximately 7.7 Mb on chromosome 6:25,652,429-33,421,465 - hg19; delimited by SCGN – secretagogin - on the telomeric end and SYNGAP - Synaptic Ras GTPase activating protein - towards the centromere). The 1,376 markers falling within the region have been excluded (Figure 34).

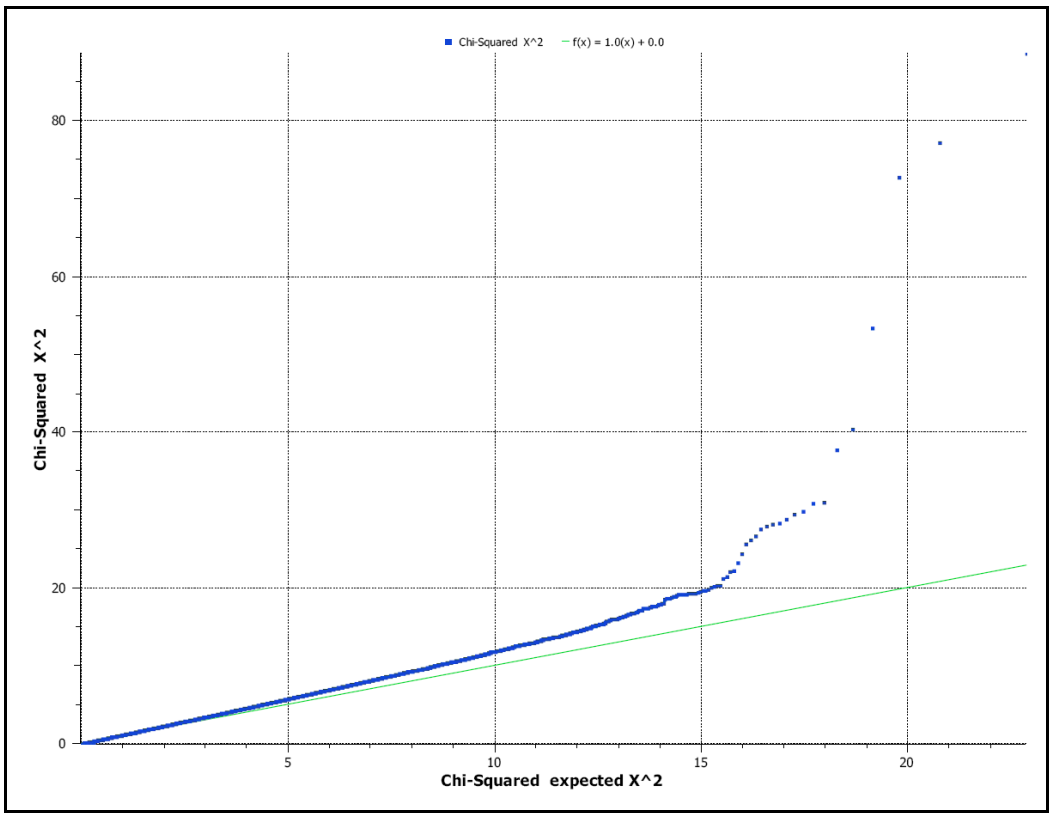


**Figure 34 Q-Q-plot: UK dataset, together low stringency scenario, no xMHC**
Q-Q plot for the Chi-squared test of the basic allelic analysis (D *versus* d) plot of the actual association $-\log_{10}P$ to the expected $-\log_{10}P$ in cases *versus* controls for the UK dataset after the low stringency (call rate > 96%, minor allele frequency > 0.01, Hardy-Weinberg equilibrium $P > 0.001$) quality control filtering; with (1,376) SNP markers in the MHC region excluded (287,288 markers).

To assess the influence of probably true positive associated markers on the Q-Q plot, the first 15 associated markers (at a $-\log_{10}P$ value of 6.8) have also been excluded from the dataset prior to plotting (Figure 35).

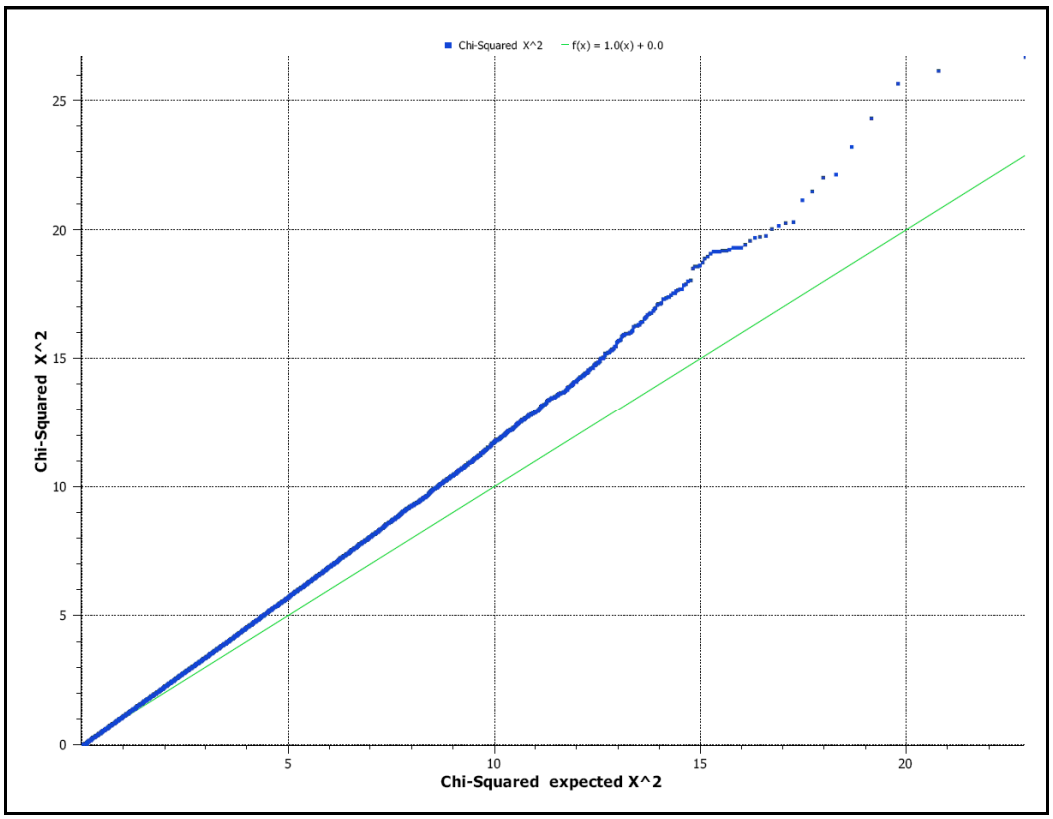


**Figure 35 Q-Q-plot: UK dataset, together low stringency, no xMHC, no assoc.**
Q-Q plot for the Chi-squared test of the basic allelic analysis (D *versus* d) plot of the actual association $-\log_{10}P$ to the expected $-\log_{10}P$ in cases *versus* controls for the UK dataset after the low stringency (call rate > 96, minor allele frequency > 0.01, Hardy-Weinberg equilibrium P > 0.001) quality control filtering; with 1,376 SNP markers in the MHC region and 15 associated SNPs (at a $-\log_{10}P$ value of 6.8) excluded.

*Basic Allelic Test (BAT)*

The basic allelic test (D *versus* d) for genotype association was performed for the UK cohort (cases and controls, 684 individuals) for the 288,625 SNP informative markers that passed the aforementioned quality control criteria (call rate > 96%, minor allele frequency > 0.01, Hardy-Weinberg equilibrium $P > 0.001$). The chosen test statistic (for the 2X2 contingency table) was the Chi-squared test. The genomic inflation factor for chi squared was found to be 1.14. The threshold for genome-wide significance corrected for multiple comparisons (Bonferroni correction) and for the genomic inflation factor (lambda = 1.14) was set at

$P$ = 0.001 / 288,625 / 1.142 = 3.03 x$10^{-9}$, that is -$\log_{10}P$ = 8.518. 151 markers were found to be associated with the trait (affected by membranous nephropathy) at this genome-wide significance threshold (compared with 188 for the non quality control analysis scenario).

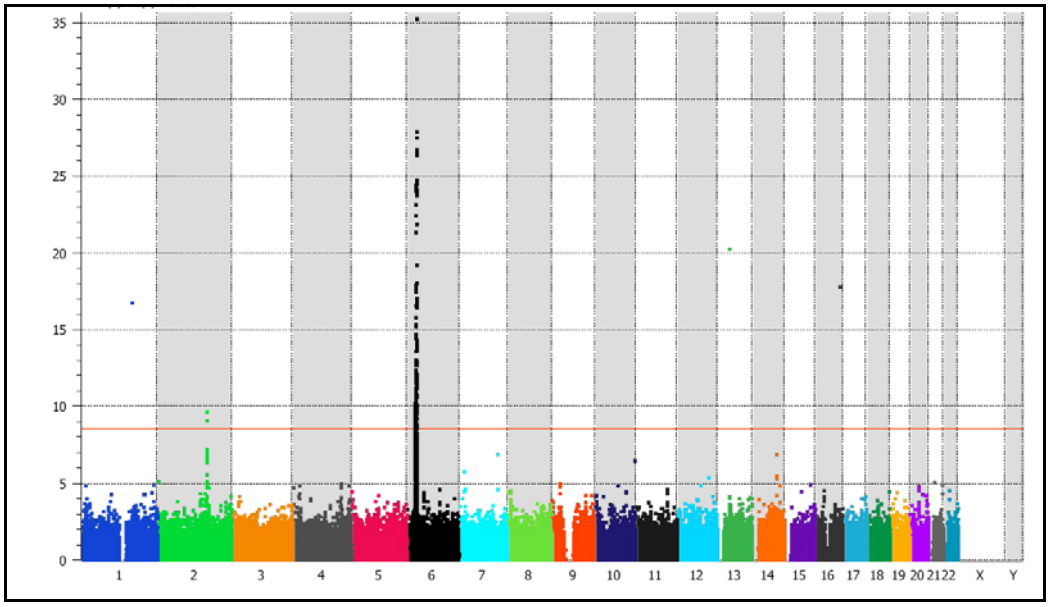


**Figure 36 BAT Manhattan plot, UK dataset, together low stringency scenario**
Manhattan plot for the Chi-squared test of the basic allelic test (BAT) comparing minor *versus* major alleles (D *versus* d) in cases *versus* controls for the UK dataset after quality control filtering for call rate > 96%, minor allele frequency > 0.01, Hardy-Weinberg equilibrium P > 0.001. The level of significance (corrected for multiple comparisons and for genomic inflation) is represented by the red line at -$\log_{10}$ = 8.518. (Y axis = -$\log_{10}P$, X axis = physical positions of the markers corresponding to the human February 2009 GRCh37/hg19 genome assembly, the chromosomes being plotted successively, shaded and coloured in different colours for better visualisation).

### Together high stringency scenario (call rate per marker > 99%)

Of the questioned 305,491 SNP markers, 305,452 markers proved informative (with two alleles) and 39 markers uninformative (with one allele). The high stringency scenario quality assurance filtering parameters are: call rate > 99%; minor allele frequency (MAF) < 0.01; Hardy-Weinberg Equilibrium $P$ value (HWE $P$) < 0.001.

34,896 (11.42 %; of all 305,491 the markers) markers did not pass the set filtering threshold (26,094 due to the call rate; 179 because of the minor allele frequency MAF and 9816 due to the Hardy-Weinberg equilibrium $P$). Some markers did not pass more than one of the filtering criteria, therefore the number

of filtered SNP markers is smaller than the sum of markers filtered on unique conditions (26,094 + 179 + 9816 = 36,089 > 34,896).

*Tests*

A number of genotype association tests were compared for the high stringency quality scenario: the basic allelic test (compares D with d); the genotypic test (compares DD vs dd vs Dd); the additive model (considers dd > Dd > DD); the dominant model (compares DD, Dd vs. dd) and the recessive model (compares DD vs. Dd, dd).

*Stratification*

The genomic inflation factor (λ) for the 270,595 used SNP markers that passed the quality control threshold was found to be 1.13.

*QQ plot*

As expected, the slope of the QQ plot slightly decreases with the increased stringency of the filtering, suggesting a better quality of the data. See Figure 37 compared with Figure 38.

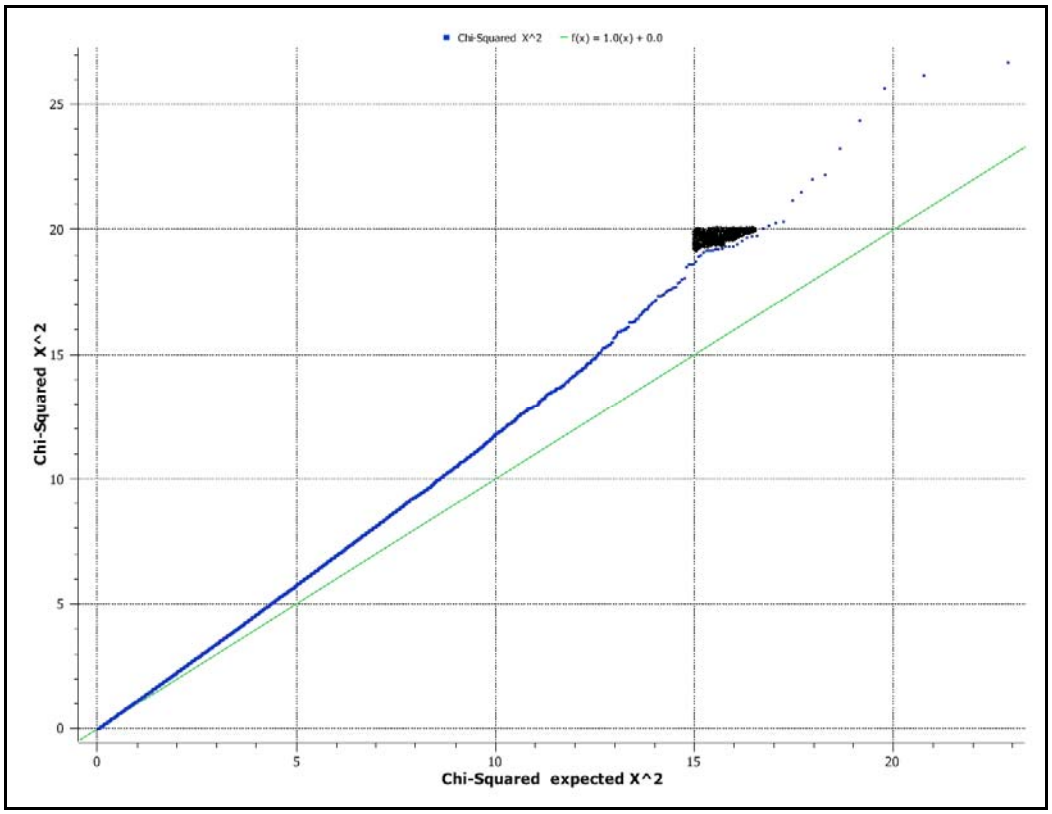


**Figure 37 Q-Q-plot UK, together low stringency scenario, no xMHC, no assoc.**
Q-Q plot for the UK dataset filtered at a call rate (CR) < 96. Blue: plot; green line: x = y.

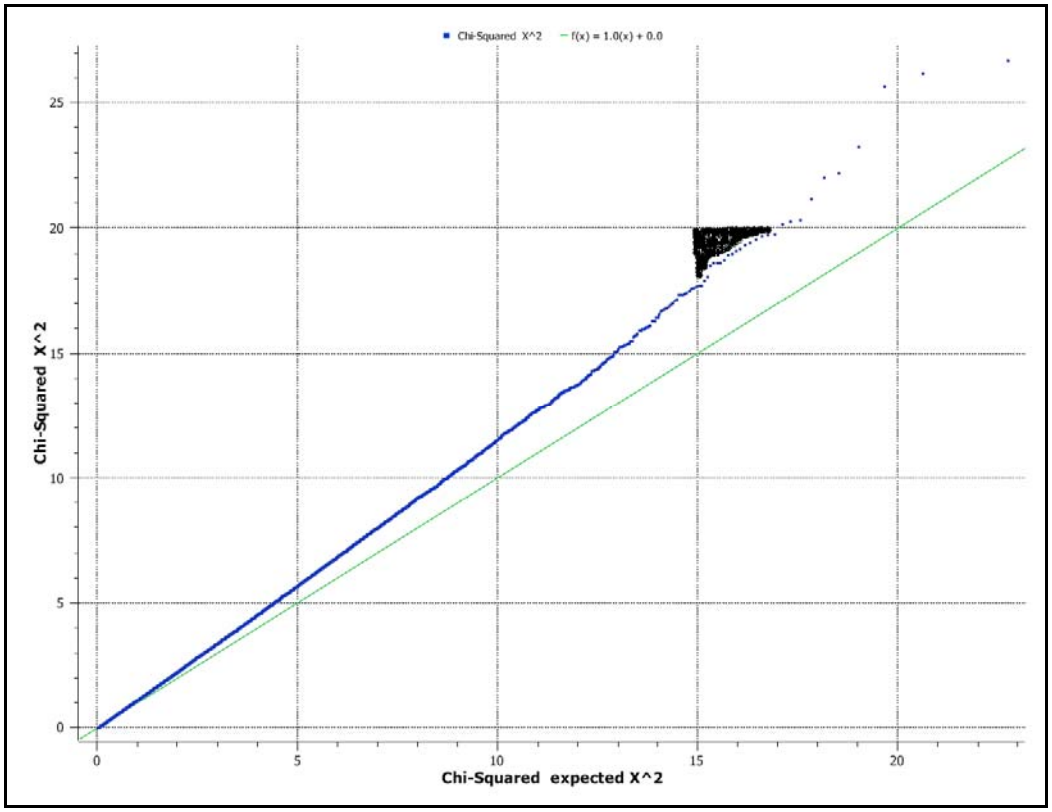


**Figure 38 Q-Q-plot UK, together high stringency scenario, no xMHC, no assoc.**
Q-Q plot of the UK dataset filtered at a call rate (CR) < 99. Increasing area of the black triangle suggests decreasing slope of the plot (blue line). Green line: x = y.

*Basic allelic test*

The chosen test statistic (for the 2X2 contingency table) was the chi-squared test. The threshold for genome-wide significance corrected for multiple comparisons (Bonferroni correction) was set at $P$ = 0.001 / 270,595 / 1.134 = 3.26 x$10^{-9}$; -$\log_{10}P$ = 8.487.

137 markers were found to be associated with the trait (affected by membranous nephropathy) at the genome-wide significance threshold.

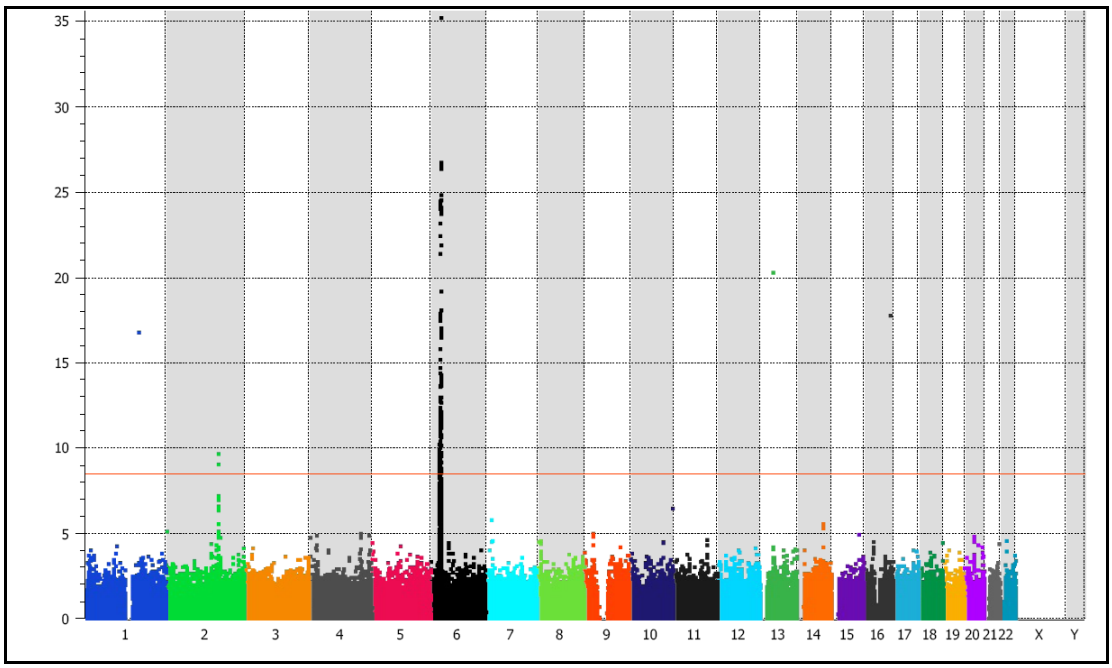


**Figure 39 BAT Manhattan plot, UK dataset, together high stringency scenario**
Manhattan plot for the Chi-squared test of the basic allelic test (BAT) comparing minor *versus* major alleles (D *versus* d) in cases *versus* controls for the UK dataset after quality control filtering for the high stringency scenario: call rate > 99%, minor allele frequency > 0.01, Hardy-Weinberg equilibrium $P > 0.001$. The level of significance (corrected for multiple comparisons and for genomic inflation) is represented by the red line at $-\log_{10}P = 8.487$. (Y axis = $-\log_{10}P$, X axis = physical positions of the markers corresponding to the human February 2009 GRCh37/hg19 genome assembly, the chromosomes being plotted successively, shaded and coloured in different colours for better visualisation).

Compared with the no quality filtering scenario this represents a drop in the number of markers of 27%; 137 of 188), compared to the low stringency quality filtering scenario it represents a drop in the number of markers of 9%; 137 of 151), see Table 8.

**Table 8 Comparison of associated SNPs for three QC scenarios**

Scenarios: no quality control, low stringency (call rate > 96%, minor allele frequency > 0.01, Hardy-Weinberg equilibrium $P > 0.001$) and high stringency quality control (call rate > 99%, minor allele frequency > 0.01, Hardy-Weinberg equilibrium $P > 0.001$); chr – chromosome; position – physical position; markers are ordered according to the descending $-\log_{10}P$ value; 188 –associated markers under the no quality control scenario; 151 – associated markers under the low stringency control scenario; 137 – associated markers under the high stringency control scenario.

| chr | position | $-\log_{10}P$ | 188 | 151 | 137 |
|---|---|---|---|---|---|
| 8 | 135681127 | 242.8945505 | rs1036819 | | |
| 4 | 80183611 | 227.0322739 | rs1455311 | | |
| 7 | 90433631 | 139.5238956 | rs17400329 | | |
| 21 | 45743560 | 133.5400546 | rs2236479 | | |
| 3 | 60699430 | 96.9885534 | rs6781958 | | |
| 12 | 27627253 | 93.03622721 | rs1975920 | | |
| 2 | 170802031 | 90.25535638 | rs1549343 | | |
| 4 | 154254023 | 87.24764677 | rs2406233 | | |
| 3 | 74801448 | 75.73985376 | rs1447826 | | |
| 6 | 1503639 | 70.93781263 | rs2745566 | | |
| 19 | 60069820 | 51.25158476 | rs3745902 | | |
| 14 | 63406232 | 36.47536951 | rs1959034 | | |
| 2 | 8410374 | 36.25093906 | rs3021387 | | |
| 6 | 32713862 | 35.28596423 | rs2187668 | | |
| 2 | 82353387 | 33.11544641 | rs4393782 | | |
| 4 | 153110013 | 31.89819588 | rs361147 | | |
| 4 | 30837982 | 28.30104237 | rs9291561 | | |
| 6 | 32172704 | 27.91840586 | rs1150752 | | |
| 6 | 32276749 | 27.54215129 | rs3134942 | | |
| 6 | 32369230 | 26.78993622 | rs7775397 | | |
| 6 | 32774504 | 26.66399161 | rs1794282 | | |
| 6 | 32500855 | 26.39264916 | rs3135353 | | |
| 6 | 31420305 | 26.17015264 | rs3134792 | | |
| 6 | 32698903 | 24.86631667 | rs3129763 | | |
| 8 | 26640531 | 24.86557187 | rs6993922 | | |
| 6 | 31978305 | 24.59074253 | rs558702 | | |
| 6 | 31709822 | 24.51763026 | rs3115663 | | |
| 6 | 31829012 | 24.41344778 | rs3131379 | | |
| 6 | 31711749 | 24.34017935 | rs9267522 | | |
| 6 | 31740113 | 24.27564526 | rs3130618 | | |
| 6 | 32280971 | 24.17224318 | rs3131296 | | |
| 6 | 31728499 | 24.09988556 | rs3117582 | | |
| 6 | 32471193 | 23.94354742 | rs1980493 | | |
| 6 | 32048876 | 23.89062354 | rs389884 | | |
| 6 | 32026839 | 23.79432354 | rs1270942 | | |
| 6 | 31556955 | 23.20241306 | rs3099844 | | |
| 6 | 31166319 | 22.51337062 | rs3130544 | | |
| 6 | 32488186 | 21.93796759 | rs3129963 | | |
| 12 | 67834426 | 21.70563827 | rs7137203 | | |
| 6 | 31237289 | 21.42088638 | rs7750641 | | |
| 13 | 37402142 | 20.32632332 | rs4943552 | | |
| 6 | 32863268 | 19.25962881 | rs7762279 | | |
| X | 121648025 | 18.50487324 | rs7057244 | | |
| 6 | 33021224 | 18.1424215 | rs1480380 | | |

| | | | |
|---|---|---|---|
| 6 | 31676448 | 17.97273658 | rs2857595 |
| 16 | 77919912 | 17.8204442 | rs2221705 |
| 6 | 31028103 | 17.71904934 | rs3132580 |
| 17 | 67919451 | 17.65103876 | rs3744311 |
| 6 | 31209653 | 17.5118627 | rs3130564 |
| 6 | 32269344 | 17.05444751 | rs204991 |
| 6 | 32447625 | 16.89806951 | rs2050189 |
| 1 | 162803852 | 16.82989189 | rs12039194 |
| 6 | 32407800 | 16.75801292 | rs1003878 |
| 6 | 32269408 | 16.73899681 | rs204990 |
| 6 | 31189184 | 16.64697203 | rs2233956 |
| 6 | 32444744 | 16.63777495 | rs3129939 |
| 6 | 31959213 | 16.49981308 | rs652888 |
| 6 | 31463297 | 15.87812204 | rs2596560 |
| 6 | 31351863 | 15.3756805 | rs3130696 |
| 6 | 30889981 | 15.25937021 | rs886424 |
| 6 | 31230479 | 14.72784754 | rs130065 |
| 6 | 31580699 | 14.71700661 | rs2855812 |
| 6 | 31648763 | 14.46854754 | rs1041981 |
| 6 | 32446673 | 14.33930723 | rs3129943 |
| 6 | 32242488 | 14.19429955 | rs3096697 |
| 6 | 32497626 | 13.99686776 | rs3135363 |
| 6 | 32262976 | 13.98526138 | rs204994 |
| 6 | 31586965 | 13.90303718 | rs2246618 |
| 6 | 32254470 | 13.82089148 | rs3134945 |
| 6 | 32772436 | 13.75766099 | rs2647012 |
| 6 | 32866372 | 13.68810007 | rs7758736 |
| 6 | 31710946 | 13.67704067 | rs1046089 |
| 6 | 31623778 | 13.61617115 | rs2071591 |
| 1 | 190284246 | 13.6070483 | rs12755225 |
| 16 | 53836712 | 13.47712779 | rs30960 |
| 2 | 48065223 | 13.15638305 | rs4953534 |
| 4 | 27051158 | 13.04332035 | rs2110177 |
| 6 | 31380240 | 13.04004135 | rs6457374 |
| 6 | 32381743 | 13.02140446 | rs6457536 |
| 6 | 32514320 | 12.99867819 | rs3129871 |
| 6 | 30760760 | 12.79692907 | rs9262143 |
| 6 | 31001920 | 12.76307723 | rs4678 |
| 6 | 32217957 | 12.73912427 | rs204999 |
| X | 3143733 | 12.58474102 | rs5983083 |
| 6 | 31796196 | 12.4180618 | rs805294 |
| 6 | 30441484 | 12.25766993 | rs3094054 |
| 6 | 31472686 | 12.20873767 | rs2251396 |
| 6 | 31248026 | 12.20838758 | rs1265159 |
| 6 | 32891063 | 12.13616118 | rs7383287 |
| 6 | 31197610 | 12.08188775 | rs3095314 |
| 6 | 32263559 | 12.08163442 | rs204993 |
| 6 | 32853021 | 12.0723072 | rs9276644 |
| 6 | 31041773 | 12.04319694 | rs3095089 |
| 6 | 31425326 | 12.0044511 | rs2156875 |
| 6 | 32030233 | 11.93785913 | rs630379 |
| 6 | 31525170 | 11.79222532 | rs3128982 |
| 6 | 32036993 | 11.77333305 | rs437179 |
| 6 | 32055439 | 11.77333305 | rs389883 |

| | | | |
|---|---|---|---|
| 6 | 31850569 | 11.73122654 | rs707928 |
| 6 | 31705732 | 11.71535927 | rs2242660 |
| 6 | 32352605 | 11.59886568 | rs3130340 |
| 6 | 32353805 | 11.59886568 | rs3115553 |
| 6 | 32368328 | 11.59886568 | rs6935269 |
| 6 | 30652380 | 11.57972368 | rs3132610 |
| 6 | 32786977 | 11.38272257 | rs9275572 |
| 6 | 32512837 | 11.35693255 | rs2395173 |
| 6 | 30387109 | 11.28067526 | rs3130380 |
| 6 | 32867426 | 11.22902126 | rs3948793 |
| 6 | 30435818 | 11.12350718 | rs3130350 |
| 6 | 30339203 | 11.05297605 | rs3094073 |
| 6 | 31589084 | 10.87399973 | rs2516400 |
| 6 | 30932511 | 10.80170435 | rs2844659 |
| 6 | 30937501 | 10.80170435 | rs2844657 |
| 5 | 53762849 | 10.80009518 | rs3923330 |
| 6 | 32512856 | 10.78783856 | rs2395174 |
| 6 | 32873429 | 10.75282627 | rs1383261 |
| 6 | 31724345 | 10.73990431 | rs805303 |
| 6 | 32888701 | 10.69374287 | rs11244 |
| 6 | 31307187 | 10.5663369 | rs3130473 |
| 2 | 7161487 | 10.52492845 | rs6722663 |
| 9 | 135267303 | 10.45449186 | rs2285487 |
| 6 | 31497763 | 10.44205369 | rs2844511 |
| 15 | 94338085 | 10.42378661 | rs4319717 |
| 6 | 30140501 | 10.40575405 | rs8321 |
| 6 | 30315474 | 10.35851841 | rs2844773 |
| 10 | 19969519 | 10.25057899 | rs1326986 |
| 6 | 28311279 | 10.24336538 | rs9380069 |
| 6 | 30146626 | 10.24336538 | rs9261290 |
| 6 | 32488240 | 10.1792513 | rs6932542 |
| 6 | 31080844 | 10.16616096 | rs1634718 |
| 6 | 31126525 | 10.04722579 | rs2523864 |
| 6 | 29139121 | 10.04219655 | rs3117143 |
| 6 | 41293408 | 9.968739987 | rs1968871 |
| 6 | 28106237 | 9.964254535 | rs149990 |
| 6 | 32509195 | 9.923210128 | rs3135338 |
| 6 | 32262263 | 9.778048388 | rs204995 |
| 6 | 32522251 | 9.751208115 | rs3129890 |
| 6 | 31118179 | 9.717477279 | rs2508015 |
| 6 | 31902571 | 9.715886123 | rs2763979 |
| 6 | 28251097 | 9.712295883 | rs9380064 |
| 2 | 160625743 | 9.68155809 | rs4664308 |
| 6 | 30053928 | 9.577058487 | rs3132685 |
| 6 | 32789255 | 9.559744783 | rs2858331 |
| 6 | 28350773 | 9.510488147 | rs1419183 |
| 6 | 28317081 | 9.509599612 | rs9295768 |
| 6 | 28318667 | 9.509599612 | rs967005 |
| 6 | 31174988 | 9.473996104 | rs2517403 |
| 6 | 27918605 | 9.377362194 | rs175597 |
| 17 | 22337222 | 9.377213207 | rs11657089 |
| 6 | 30805426 | 9.371116024 | rs3094127 |
| 6 | 31183460 | 9.351798196 | rs2844635 |
| 6 | 31233756 | 9.33706266 | rs720465 |

| | | | |
|---|---|---|---|
| 6 | 31429190 | 9.297097892 | rs2596501 |
| 6 | 31213392 | 9.240754412 | rs1265099 |
| 6 | 30188253 | 9.232746147 | rs2517598 |
| 6 | 32759095 | 9.228489045 | rs9275141 |
| 6 | 27642507 | 9.196001638 | rs10484399 |
| 6 | 31439808 | 9.176612019 | rs2523554 |
| 6 | 31443433 | 9.155420758 | rs2844573 |
| 4 | 6443243 | 9.14040739 | rs2240268 |
| 6 | 28738670 | 9.115691387 | rs13194504 |
| 2 | 160778828 | 9.087908524 | rs1870102 |
| 6 | 27945246 | 9.068176039 | rs200953 |
| 6 | 27923618 | 9.064970066 | rs13194781 |
| 6 | 31189968 | 9.015447581 | rs3130977 |
| 6 | 31168141 | 8.964154138 | rs2517452 |
| 6 | 30576770 | 8.948358807 | rs3131115 |
| 6 | 27599278 | 8.942354142 | rs6904596 |
| 6 | 28365356 | 8.886294168 | rs13211507 |
| 6 | 32425613 | 8.878890452 | rs2076537 |
| 8 | 137820222 | 8.855174724 | rs3911702 |
| 6 | 28820642 | 8.824183304 | rs1233579 |
| 6 | 31444229 | 8.795878124 | rs2523535 |
| 6 | 31253899 | 8.776689747 | rs887464 |
| 6 | 29088686 | 8.765442152 | rs3130893 |
| 6 | 27339129 | 8.753000081 | rs6938200 |
| 6 | 32771829 | 8.739961597 | rs6457617 |
| 6 | 28122576 | 8.686128447 | rs149900 |
| 6 | 28281198 | 8.685204466 | rs1233708 |
| 6 | 29450754 | 8.678537191 | rs3749971 |
| 6 | 31183354 | 8.65241017 | rs1064191 |
| 6 | 28407666 | 8.626034003 | rs853676 |
| 6 | 31968316 | 8.605229099 | rs535586 |
| 6 | 31972283 | 8.605229099 | rs659445 |
| 6 | 29062272 | 8.595432246 | rs3129791 |
| 6 | 29645203 | 8.573750486 | rs1235162 |
| 6 | 27814159 | 8.55317383 | rs9468213 |
| 6 | 27198383 | 8.553166436 | rs7745603 |

*Genotypic test*

Applying the Chi-square statistic test to the genotypic test (DD *versus* dd *versus* Dd comparison) for the 270,595 SNP markers which passed the high stringency quality control filtering criteria 107 were found to be associated with the trait at $-\log_{10}P$ 8.463 (corrected for multiple comparisons and for genomic inflation factor, $\lambda = 1.07$).

*Additive test*

Applying the Cochran Armitage statistic method to the additive model (which considers the effects on the trait to be increasing with the presence of none, one

or two alleles: dd > Dd > DD) for the 270,595 SNP markers which passed the high stringency quality control filtering criteria, 137 were found to be associated with the trait at $-\log_{10}P$ 8.485 (corrected for multiple comparisons and for genomic inflation factor, $\lambda = 1.13$).

*Dominant model*

Applying the Chi-square statistic test to the dominant model (which compares the effects of DD, Dd *versus* dd) for the 270,595 SNP markers which passed the high stringency quality control filtering criteria 104 of them were found to be associated with the trait at $-\log_{10}P$ 8.477 (corrected for multiple comparisons and for genomic inflation factor, $\lambda = 1.11$).

*Recessive model*

Applying the Chi-square statistic test to the recessive model (which compares DD *versus* Dd and dd) for the 270,595 SNP markers which passed the high stringency quality control filtering criteria only 14 were found to be associated with the trait at $-\log_{10}P$ 8.474 (corrected for multiple comparisons and for genomic inflation factor, $\lambda = 1.10$).

**Table 9 Comparison of associated SNPs for five association tests**

Comparison of $-\log_{10}P$ values for the strenght of association between the trait and different markers under different models: basic allelic test BAT, genotypic test GT, additive model AM, dominant model DM and recessive model RM.

| Marker | Chr | Position | BAT | GT | AM | DM | RM |
|---|---|---|---|---|---|---|---|
| rs2187668 | 6 | 32713862 | 35.286 | 34.09 | 33.879 | 33.827 | 9.803 |
| rs7775397 | 6 | 32369230 | 26.79 | 25.865 | 25.934 | 26.19 | - |
| rs1794282 | 6 | 32774504 | 26.664 | 25.655 | 25.754 | 25.967 | - |
| rs3135353 | 6 | 32500855 | 26.393 | 25.428 | 25.507 | 25.585 | - |
| rs3129763 | 6 | 32698903 | 24.866 | 25.858 | 25.87 | 24.819 | 9.232 |
| rs558702 | 6 | 31978305 | 24.591 | 23.502 | 23.814 | 23.745 | - |
| rs3115663 | 6 | 31709822 | 24.518 | 22.825 | 23.924 | 20.329 | 11.148 |
| rs3131379 | 6 | 31829012 | 24.413 | 23.375 | 23.618 | 23.716 | - |
| rs9267522 | 6 | 31711749 | 24.34 | 22.638 | 23.737 | 20.14 | 11.108 |
| rs3130618 | 6 | 31740113 | 24.276 | 22.539 | 23.641 | 20.007 | 11.148 |
| rs3131296 | 6 | 32280971 | 24.172 | 22.801 | 23.29 | 22.734 | - |
| rs3117582 | 6 | 31728499 | 24.1 | 23.127 | 23.434 | 23.416 | - |
| rs1980493 | 6 | 32471193 | 23.944 | 22.872 | 22.872 | 23.137 | - |
| rs389884 | 6 | 32048876 | 23.891 | 22.79 | 23.204 | 22.962 | - |
| rs1270942 | 6 | 32026839 | 23.794 | 22.645 | 23.077 | 22.802 | - |
| rs3099844 | 6 | 31556955 | 23.202 | 22.353 | 22.856 | 22.432 | - |
| rs3130544 | 6 | 31166319 | 22.513 | 22.086 | 22.829 | 21.853 | - |
| rs3129963 | 6 | 32488186 | 21.938 | 21.091 | 21.83 | 20.489 | - |
| rs7750641 | 6 | 31237289 | 21.421 | 20.812 | 21.594 | 20.501 | - |
| rs4943552 | 13 | 37402142 | 20.326 | 17.304 | 18.338 | 12.621 | 13.304 |
| rs7762279 | 6 | 32863268 | 19.26 | 18.356 | 18.756 | 18.919 | - |
| rs1480380 | 6 | 33021224 | 18.142 | 16.517 | 17.219 | 17.017 | - |
| rs2857595 | 6 | 31676448 | 17.973 | 18.096 | 17.942 | 18.641 | - |
| rs2221705 | 16 | 77919912 | 17.82 | 17.034 | 17.41 | 18.085 | - |
| rs3132580 | 6 | 31028103 | 17.719 | 16.497 | 17.467 | 15.956 | - |
| rs3130564 | 6 | 31209653 | 17.512 | 17.424 | 17.955 | 17.163 | - |
| rs204991 | 6 | 32269344 | 17.054 | 15.795 | 16.469 | 15.509 | - |
| rs2050189 | 6 | 32447625 | 16.898 | 15.947 | 16.898 | 14.765 | - |
| rs12039194 | 1 | 162803852 | 16.83 | 15.438 | 15.494 | 16.468 | - |
| rs1003878 | 6 | 32407800 | 16.758 | 15.949 | 16.708 | 15.174 | - |
| rs204990 | 6 | 32269408 | 16.739 | 15.422 | 16.273 | 14.776 | - |
| rs3129939 | 6 | 32444744 | 16.638 | 16.152 | 16.863 | 15.977 | - |
| rs652888 | 6 | 31959213 | 16.5 | 14.783 | 15.787 | 12.339 | 9.214 |
| rs2596560 | 6 | 31463297 | 15.878 | 15.184 | 15.533 | 15.24 | - |
| rs886424 | 6 | 30889981 | 15.259 | 14.355 | 15.34 | 13.98 | - |
| rs2855812 | 6 | 31580699 | 14.717 | 13.267 | 13.971 | 12.896 | - |
| rs1041981 | 6 | 31648763 | 14.469 | 14.145 | 14.768 | 13.172 | - |
| rs3129943 | 6 | 32446673 | 14.339 | 13.315 | 14.154 | 12.431 | - |
| rs3096697 | 6 | 32242488 | 14.194 | 13.078 | 14.054 | 11.103 | - |
| rs3135363 | 6 | 32497626 | 13.997 | 14.288 | 14.674 | 14.121 | - |
| rs204994 | 6 | 32262976 | 13.985 | 12.908 | 13.851 | 11.884 | - |
| rs3134945 | 6 | 32254470 | 13.821 | 12.885 | 13.869 | 11.11 | - |
| rs2647012 | 6 | 32772436 | 13.758 | 12.826 | 13.762 | 10.334 | 8.527 |
| rs7758736 | 6 | 32866372 | 13.688 | 12.967 | 13.375 | 13.5 | - |
| rs1046089 | 6 | 31710946 | 13.677 | 13.548 | 13.858 | 13.07 | - |
| rs2071591 | 6 | 31623778 | 13.616 | 13.374 | 14.086 | 12.285 | - |
| rs6457374 | 6 | 31380240 | 13.04 | 12.596 | 13.049 | 12.677 | - |
| rs3129871 | 6 | 32514320 | 12.999 | 12.43 | 13.412 | 11.46 | - |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| rs9262143 | 6 | 30760760 | 12.797 | 12.179 | 13.078 | 12.381 | - |
| rs4678 | 6 | 31001920 | 12.763 | 12.055 | 12.961 | 11.632 | - |
| rs204999 | 6 | 32217957 | 12.739 | 11.83 | 12.636 | 8.528 | 8.988 |
| rs805294 | 6 | 31796196 | 12.418 | 11.249 | 12.037 | 10.162 | - |
| rs2251396 | 6 | 31472686 | 12.209 | 13.287 | 12.584 | 13.959 | - |
| rs1265159 | 6 | 31248026 | 12.208 | 11.1 | 12.051 | 9.949 | - |
| rs7383287 | 6 | 32891063 | 12.136 | 11.199 | 11.961 | 11.086 | - |
| rs3095314 | 6 | 31197610 | 12.082 | 11.431 | 12.355 | - | 9.064 |
| rs204993 | 6 | 32263559 | 12.082 | 11.001 | 11.899 | 9.982 | - |
| rs9276644 | 6 | 32853021 | 12.072 | 10.794 | 11.613 | 9.736 | - |
| rs3095089 | 6 | 31041773 | 12.043 | 11.438 | 12.344 | 11.128 | - |
| rs2156875 | 6 | 31425326 | 12.004 | 10.805 | 11.63 | - | 8.546 |
| rs630379 | 6 | 32030233 | 11.938 | 10.753 | 11.58 | 10.087 | - |
| rs3128982 | 6 | 31525170 | 11.792 | 11.28 | 12.162 | 10.108 | - |
| rs437179 | 6 | 32036993 | 11.773 | 10.571 | 11.381 | 9.96 | - |
| rs389883 | 6 | 32055439 | 11.773 | 10.571 | 11.381 | 9.96 | - |
| rs707928 | 6 | 31850569 | 11.731 | 10.922 | 11.485 | 10.471 | - |
| rs2242660 | 6 | 31705732 | 11.715 | 11.671 | 11.841 | 11.492 | - |
| rs3130340 | 6 | 32352605 | 11.599 | 10.811 | 11.752 | 9.241 | - |
| rs3115553 | 6 | 32353805 | 11.599 | 10.811 | 11.752 | 9.241 | - |
| rs6935269 | 6 | 32368328 | 11.599 | 10.811 | 11.752 | 9.241 | - |
| rs3132610 | 6 | 30652380 | 11.58 | 10.969 | 11.905 | 11.015 | - |
| rs9275572 | 6 | 32786977 | 11.383 | 10.561 | 11.51 | - | - |
| rs2395173 | 6 | 32512837 | 11.357 | 11.117 | 12.078 | 10.369 | - |
| rs3130380 | 6 | 30387109 | 11.281 | 10.925 | 11.659 | 11.581 | - |
| rs3948793 | 6 | 32867426 | 11.229 | 10.045 | 10.844 | 9.147 | - |
| rs3130350 | 6 | 30435818 | 11.124 | 10.909 | 11.587 | 11.639 | - |
| rs3094073 | 6 | 30339203 | 11.053 | 10.696 | 11.378 | 11.256 | - |
| rs2844659 | 6 | 30932511 | 10.802 | 10.162 | 10.991 | 10.099 | - |
| rs2844657 | 6 | 30937501 | 10.802 | 10.162 | 10.991 | 10.099 | - |
| rs2395174 | 6 | 32512856 | 10.788 | 9.981 | 10.827 | - | - |
| rs1383261 | 6 | 32873429 | 10.753 | 9.293 | 10.108 | 8.557 | - |
| rs805303 | 6 | 31724345 | 10.74 | 10.61 | 10.996 | 10.319 | - |
| rs11244 | 6 | 32888701 | 10.694 | 9.417 | 10.131 | 9.262 | - |
| rs2844511 | 6 | 31497763 | 10.442 | 10.43 | 10.53 | - | 10.212 |
| rs8321 | 6 | 30140501 | 10.406 | 9.977 | 10.701 | 10.663 | - |
| rs2844773 | 6 | 30315474 | 10.359 | 9.911 | 10.658 | 10.403 | - |
| rs9380069 | 6 | 28311279 | 10.243 | 9.901 | 10.752 | 10.444 | - |
| rs9261290 | 6 | 30146626 | 10.243 | 9.807 | 10.547 | 10.477 | - |
| rs6932542 | 6 | 32488240 | 10.179 | 10.404 | 10.276 | - | 10.059 |
| rs1634718 | 6 | 31080844 | 10.166 | 9.476 | 10.399 | - | - |
| rs2523864 | 6 | 31126525 | 10.047 | 9.019 | 9.936 | - | - |
| rs149990 | 6 | 28106237 | 9.964 | 9.16 | 10.025 | 9.544 | - |
| rs3129890 | 6 | 32522251 | 9.751 | 9.169 | 9.829 | 9.205 | - |
| rs2508015 | 6 | 31118179 | 9.717 | 8.634 | 9.515 | - | - |
| rs2763979 | 6 | 31902571 | 9.716 | 8.587 | 9.351 | - | - |
| rs9380064 | 6 | 28251097 | 9.712 | 9.163 | 10.047 | 9.574 | - |
| rs4664308 | 2 | 160625743 | 9.682 | - | 9.083 | - | - |
| rs3132685 | 6 | 30053928 | 9.577 | 8.953 | 9.807 | 9.389 | - |
| rs2858331 | 6 | 32789255 | 9.56 | 8.658 | 9.566 | - | - |
| rs1419183 | 6 | 28350773 | 9.51 | 8.968 | 9.832 | 9.381 | - |
| rs9295768 | 6 | 28317081 | 9.51 | 9.077 | 9.886 | 9.473 | - |
| rs967005 | 6 | 28318667 | 9.51 | 9.077 | 9.886 | 9.473 | - |
| rs2517403 | 6 | 31174988 | 9.474 | 8.606 | 9.496 | - | - |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| rs175597 | 6 | 27918605 | 9.377 | - | 9.376 | 8.686 | - |
| rs3094127 | 6 | 30805426 | 9.371 | 8.775 | 9.685 | - | - |
| rs2844635 | 6 | 31183460 | 9.352 | 8.499 | 9.384 | - | - |
| rs720465 | 6 | 31233756 | 9.337 | - | 9.226 | - | - |
| rs2596501 | 6 | 31429190 | 9.297 | - | 8.941 | - | - |
| rs1265099 | 6 | 31213392 | 9.241 | 8.828 | 9.267 | - | - |
| rs2517598 | 6 | 30188253 | 9.233 | 9.028 | 9.671 | 9.628 | - |
| rs9275141 | 6 | 32759095 | 9.228 | - | 9.266 | - | - |
| rs10484399 | 6 | 27642507 | 9.196 | 8.73 | 9.574 | 8.683 | - |
| rs2523554 | 6 | 31439808 | 9.177 | - | 9.307 | - | - |
| rs13194504 | 6 | 28738670 | 9.116 | 8.745 | 9.559 | 9.459 | - |
| rs1870102 | 2 | 160778828 | 9.088 | - | - | - | - |
| rs200953 | 6 | 27945246 | 9.068 | - | 9.114 | - | - |
| rs13194781 | 6 | 27923618 | 9.065 | 8.707 | 9.558 | 8.75 | - |
| rs2517452 | 6 | 31168141 | 8.964 | - | 8.825 | - | - |
| rs3131115 | 6 | 30576770 | 8.948 | - | 8.91 | - | - |
| rs6904596 | 6 | 27599278 | 8.942 | - | 9.109 | - | - |
| rs13211507 | 6 | 28365356 | 8.886 | - | 9.385 | 9.029 | - |
| rs2076537 | 6 | 32425613 | 8.879 | - | 8.884 | - | - |
| rs1233579 | 6 | 28820642 | 8.824 | - | 9.294 | 9.16 | - |
| rs2523535 | 6 | 31444229 | 8.796 | - | 9.145 | - | - |
| rs887464 | 6 | 31253899 | 8.777 | - | 8.903 | - | - |
| rs3130893 | 6 | 29088686 | 8.765 | - | 9.234 | 9.099 | - |
| rs6457617 | 6 | 32771829 | 8.74 | - | 8.76 | - | - |
| rs1233708 | 6 | 28281198 | 8.685 | - | 8.57 | - | - |
| rs3749971 | 6 | 29450754 | 8.679 | - | 9.12 | 9.175 | - |
| rs1064191 | 6 | 31183354 | 8.652 | - | 8.556 | - | - |
| rs853676 | 6 | 28407666 | 8.626 | - | 8.853 | - | - |
| rs535586 | 6 | 31968316 | 8.605 | - | 8.496 | - | - |
| rs659445 | 6 | 31972283 | 8.605 | - | 8.496 | - | - |
| rs3129791 | 6 | 29062272 | 8.595 | - | 9.065 | 8.909 | - |
| rs1235162 | 6 | 29645203 | 8.574 | - | 8.795 | 8.814 | - |
| rs7745603 | 6 | 27198383 | 8.553 | - | - | - | - |
| rs3131093 | 6 | 28945416 | 8.537 | - | 9.006 | 8.849 | - |
| rs2746150 | 6 | 29550680 | 8.508 | - | 8.927 | 8.966 | - |
| rs4324798 | 6 | 28884096 | - | - | 8.87 | 8.702 | - |
| rs2523989 | 6 | 30186254 | - | - | 8.679 | - | - |
| rs2523987 | 6 | 30187972 | - | - | - | 8.643 | - |
| rs2253907 | 6 | 31444849 | - | - | - | - | 8.545 |

Nine of the markers (rs4943552 on chromosome 13; rs2187668, rs3129763, rs3115663, rs9267522, rs3130618, rs652888, rs2647012 and rs204999 on chromosome 6) are associated with the trait in all five scenarios, 88 are associated with the trait in 4 of the five scenarios, 20 of the markers are associated in 3 of the five, 19 in two and 5 only in one of the five scenarios.

The associated SNPs are distributed on five chromosomes as follows: one SNP on chromosome 1 (rs12039194), see Figure 38; two SNPs on chromosome 2 (rs4664308 and rs1870102), see Figure 39; 132 SNPs on chromosome 6 (all of which in the HLA region), see Figure 40; one SNP on chromosome 13 (rs4943552); see Figure 41; and one SNP on chromosome 16 (rs2221705), see Figure 42.

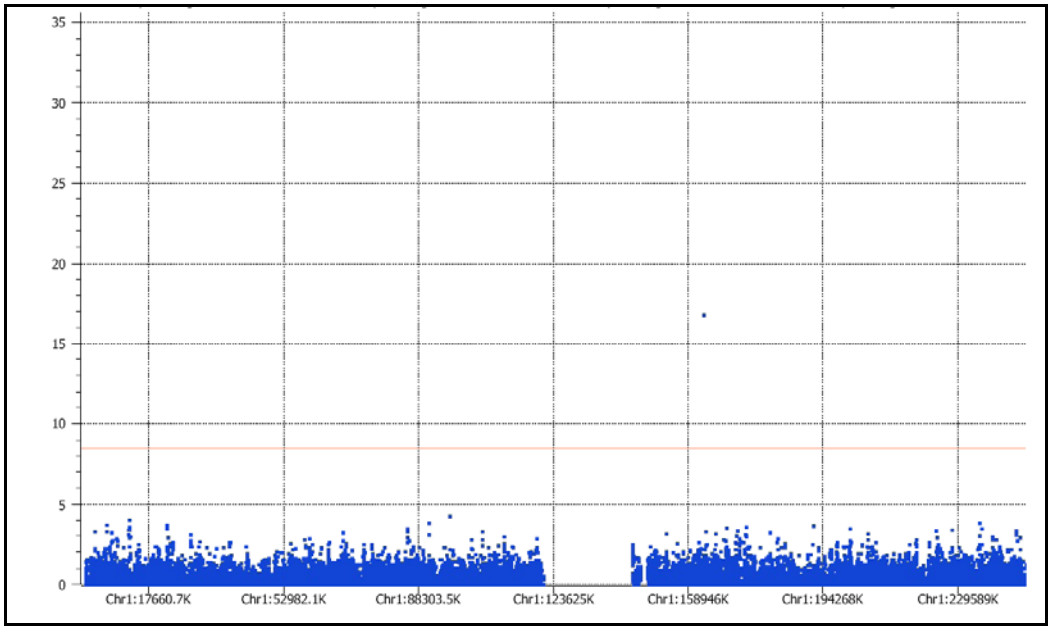


**Figure 40 BAT Manhattan plot, UK, together high stringency, chromosome 1**
Manhattan plot zoomed in on chromosome 1 representing the results of the basic allelic test (BAT) of association for the 'together' high stringency quality control scenario (together meaning appended before cleaning; call rate > 99%, minor allele frequency > 0.01, Hardy-Weinberg equilibrium $P > 0.001$). The significance line (red) is drawn at $-\log_{10}P$, Bonferroni and genomic inflation factor lambda corrected y = 8.487. The associated (at $-\log_{10}P = 15.438$) SNP marker at chr1 164537228 (February 2009 assembly GRCh37 / hg19) is rs12039194.

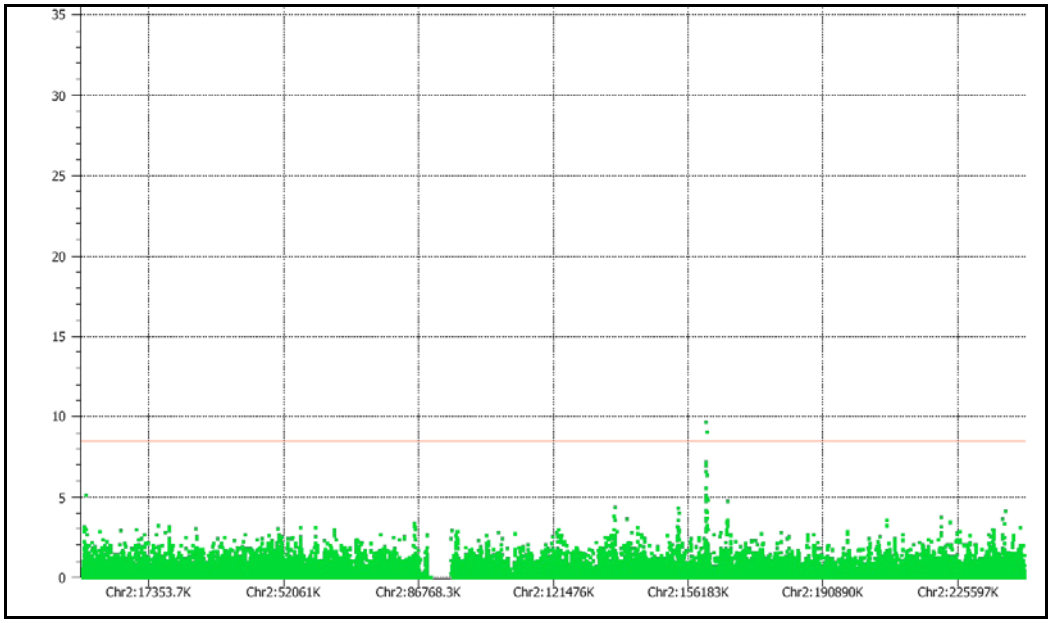


**Figure 41 BAT Manhattan plot, UK, together high stringency, chromosome 2**
Manhattan plot zoomed in on chromosome 2 representing the results of the basic allelic test (BAT) of association for the 'together' high stringency quality control scenario (together meaning appended before cleaning; call rate > 99%, minor allele frequency > 0.01, Hardy-Weinberg equilibrium $P > 0.001$). The significance line (red) is drawn at $-\log_{10}P$, Bonferroni and genomic inflation factor lambda corrected y = 8.487.

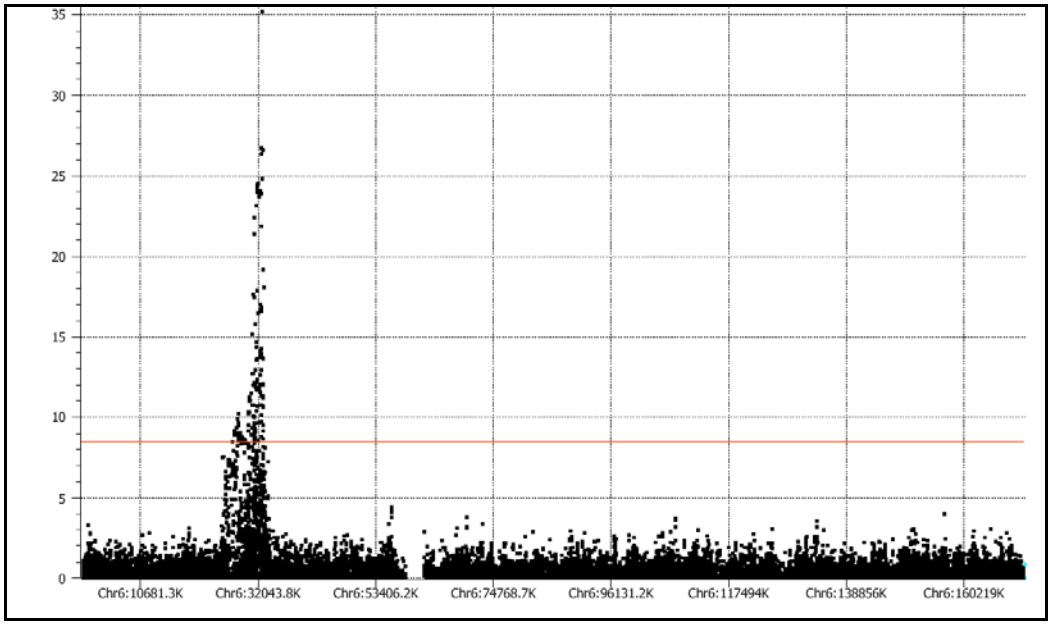


**Figure 42 BAT Manhattan plot, UK, together high stringency, chromosome 6**
Manhattan plot zoomed in on chromosome 6 representing the results of the basic allelic test (BAT) of association for the 'together' high stringency quality control scenario (together meaning appended before cleaning; call rate > 99%, minor allele frequency > 0.01, Hardy-Weinberg equilibrium P > 0.001). The significance line (red) is drawn at $-\log_{10}P$, Bonferroni and genomic inflation factor lambda corrected y = 8.487.

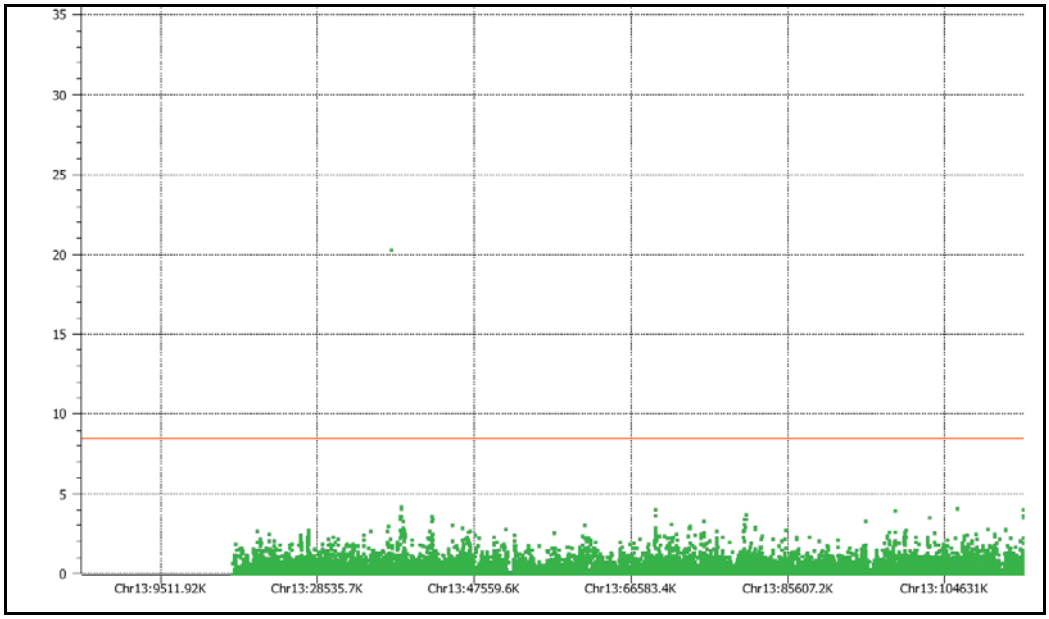


**Figure 43 BAT Manhattan plot, UK, together high stringency, chromosome 13**
Manhattan plot zoomed in on chromosome 13 representing the results of the basic allelic test (BAT) of association for the 'together' high stringency quality control scenario (together meaning appended before cleaning; call rate > 99%, minor allele frequency > 0.01, Hardy-Weinberg equilibrium $P > 0.001$). The significance line (red) is drawn at $-\log_{10}P$, Bonferroni and genomic inflation factor lambda corrected y = 8.487.

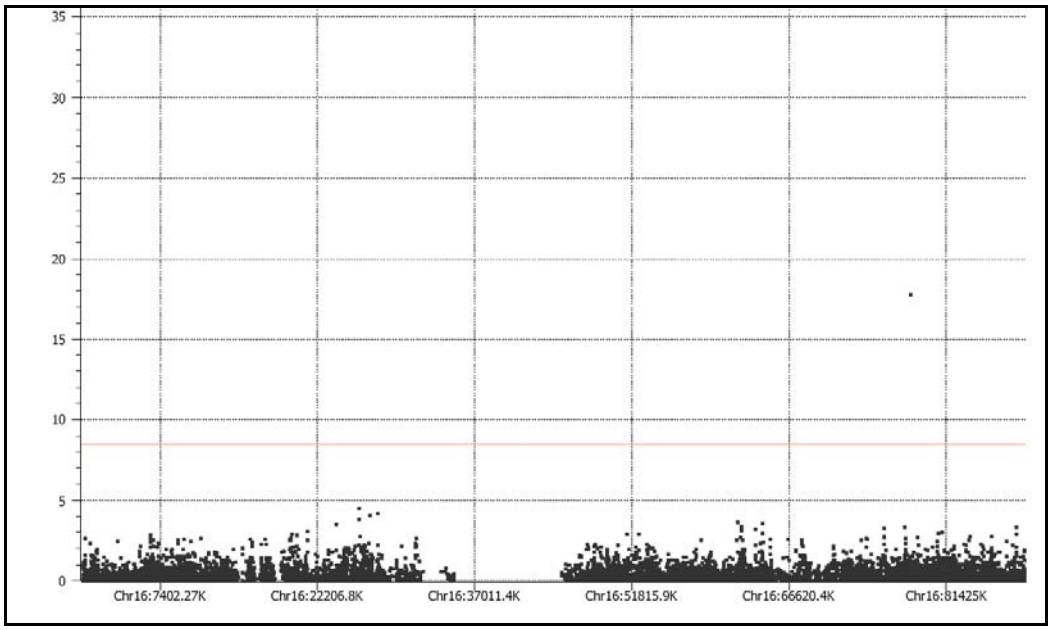


**Figure 44 BAT Manhattan plot, UK, together high stringency, chromosome 16**
Manhattan plot zoomed in on chromosome 16 representing the results of the basic allelic test (BAT) of association for the 'together' high stringency quality control scenario (together meaning appended before cleaning; call rate > 99%, minor allele frequency > 0.01, Hardy-Weinberg equilibrium P > 0.001). The significance line (red) is drawn at $-\log_{10}P$, Bonferroni and genomic inflation factor lambda corrected y = 8.487.

*Description of the regions of interest*

Six regions in the human genome were assessed for elements of interest in the light of the association of SNPs in the vicinity.

*Chromosome 1*

The SNP marker defining the region of interest on chromosome 1 is rs12039194 (physical position 164537228 bp on the February 2009 assembly GRCh37 / hg19). It is a coding (+) strand G > A/G type single nucleotide variant, a substitution (translation), the ancestral allele being G (present on the positive strand in the chimpanzee *Pan paniscus* genome, on the negative strand in the orangutan *Pongo pygmaeus* genome, and on the positive strand in the macaque *Macacus rhesus* genome). It has been validated by cluster, by frequency and by HapMap. Its average heterozygosity is 0.325 +/- 0.238. The weight quality assignment by dbSNP of rs12039194 is 1 (where weight can be 0, 1, 2, 3 or 10 with 0 and 10 being excluded from the dataset and 1 being the highest quality alignment). It lies within an intron of the *PBX1* gene.

PBX1

Pre-B-cell leukemia homeobox 1 is encoded on the plus (+) / coding strand of chromosome 1, spans 292,244 base pairs of genomic sequence, contains nine exons (all of them coding). The protein encoded by *PBX1* has 430 amino acids and is a transcription factor of the TALE (Three Aminoacid Loop Extension) / PBX1 homeobox family [496]. It is located in the cell nucleus where it forms heterodimers with MEIS1 (Myeloid Ecotropic viral Integration Site 1 homologue). The PBX1-MEIS complex is known to act as a cofactor to various transcription factors (most importantly Hox) [497] by penetrating repressive chromatin to mark specific genes for activation or inactivation [498]. The complex has selective DNA binding properties, its target DNA binding sequence on the genome being a bipartite consensus composed of the 5' PBX (TGAT) and 3' MEIS (TGACAG) half sites [499]. The DNA binding domain of PBX1 is a homeobox 1 DNA binding domain. PBX1 is expressed in all human tissues, with the exception of the cells of the B and T lineage. It has been suggested that PBX1 has a role in steroidogenesis (might be therefore important in sexual

development and / or differentiation) and even kidney development ([500] and iHOP - Information Hyperlinked Over Proteins, [501]).

*Chromosome 2*

The region of interest on chromosome 2 is driven by two SNP markers that passed the threshold for association (rs4664308 at $-\log_{10}P = 9.681$ and rs1870102 at $-\log_{10}P = 9.088$) but the morphology of the peak (aggregation of 16 markers at levels below the threshold but with a $-\log_{10}P$ score greater than 2 in a *quasi* Gaussian distribution around the marker showing strongest association, *i.e.* rs4664308) also strongly suggests a signal of association.

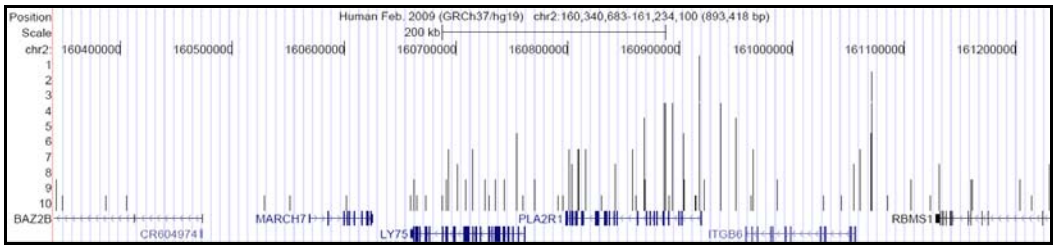


**Figure 45 Association peaks peri-PLA2R1**
Custom track plotting the $-\log_{10}$ Chi-square scores of the 86 SNPs in the peri-PLA2R1 region on chromosome 2, showing the two association peaks (at rs4664308, in the first intron of PLA2R1 and at rs1870102, upstream of *ITGB6*) and the morphology of the quasi Gaussian 'signal build up' around them.

SNP rs4664308 (physical position 160917497 bp on the February 2009 assembly GRCh37 / hg19) is a coding (+) strand A > A/G type single nucleotide variant, a substitution (translation), the ancestral allele being A (present on the positive strand of the chimpanzee *Pan paniscus* genome, on the positive strand of the orangutan *Pongo pygmaeus* genome, and on the positive strand of the macaque *Macacus rhesus* genome). It has been validated by cluster, by frequency, by HapMap and by 1000 genomes [502]. Its average heterozygosity is 0.391 +/- 0.207. The weight quality assignment by dbSNP of rs4664308 is 1. According to the functional classification, it is an intronic SNP, lying within an intron of the *PLA2R1* gene.

PLA2R1

Phospholipase A2 receptor 1 isoform 1 precursor is encoded on the minus (-) / noncoding strand of chromosome 2, spans 121,110 base pairs of genomic

sequence, contains 30 exons (all of them coding). The protein encoded by PLA2R1 has 1463 amino acids and is a receptor for secretory phospholipases A2 (PLA2s). It is a transmembrane protein with one transmembrane pass alpha helix domain and has been previously described in detail (*vide supra*).

Marker rs1870102 (physical position 161070582 bp on the February 2009 assembly GRCh37 / hg19) is a antisense (-) strand T > C/T type single nucleotide variant, a substitution (translation), the ancestral (reference) allele being T (present on the negative strand of the chimpanzee *Pan paniscus* genome, the negative strand of the orangutan *Pongo pygmaeus* genome, and the negative strand of the macaque *Macacus rhesus* genome). It has been validated by cluster, by frequency, by the 2 hit / 2 allele method (*i.e.* each allele has been observed in at least two chromosomes), by HapMap and by 1000 genomes. Its average heterozygosity is 0.376 +/- 0.216. The weight quality assignment by dbSNP of rs1870102 is 1. It is a SNP of unknown function, which is localized upstream of the ITGB6 gene.

ITGB6

Integrin beta 6 precursor is encoded on the minus (-) / noncoding strand of chromosome 2, spans 98,357 base pairs of genomic sequence, contains 15 exons (all of them coding). The protein encoded by ITGB6 has 788 amino acids and is a member of the integrin beta chain family - the beta subunits of the integrin proteins. It is believed (by similarity) to be a single-pass type I membrane protein. It has been shown to activate latent transforming growth factor (TGF)-β1 in knock-out mice, thus playing a role in the homeostasis of phospholipids and collectins, which has apparent implications for the understanding of the innate immune response and remodelling of the extracellular matrix / fibrotic mechanisms [503]. It has also been shown that ITGB6 is induced in endothelial cells by cytomegalovirus (CMV) infection, and leads, through activation of TGF-β1, to increased collagen production [504].

*Chromosome 6*

The 132 associated SNPs (at $-\log_{10}P > 8.487$) cluster on chromosome 6 in the 6p22.2-21.32 region, which corresponds to the extended HLA – MHC locus of

6 Mb [226]. For the purposes of this analysis the extended HLA locus is considered to be delimited at the telomeric end by the 5' extremity of *SCGN* (secretagogin precursor; at chr6 25,760,408 - hg18 assembly- respectively chr6 25,652,429 -GRCh37 / hg19 assembly) and at the centromeric end by the 3' extremity of *SYNGAP1* (synaptic Ras GTPase activating protein 1; at chr6 33,529,444 – hg18 assembly – respectively chr6 33,421,465 – GRCh37 / hg19 assembly). This region is covered by 1274 SNPs (between rs7747078 and rs2247385; at chr6:25,775,918-33,529,555 on the hg18 assembly) and contains 299 RefSeq genes (curated and non-redundant as defined by NCBI Reference Sequence [505]; hg18 assembly), Figure 46.

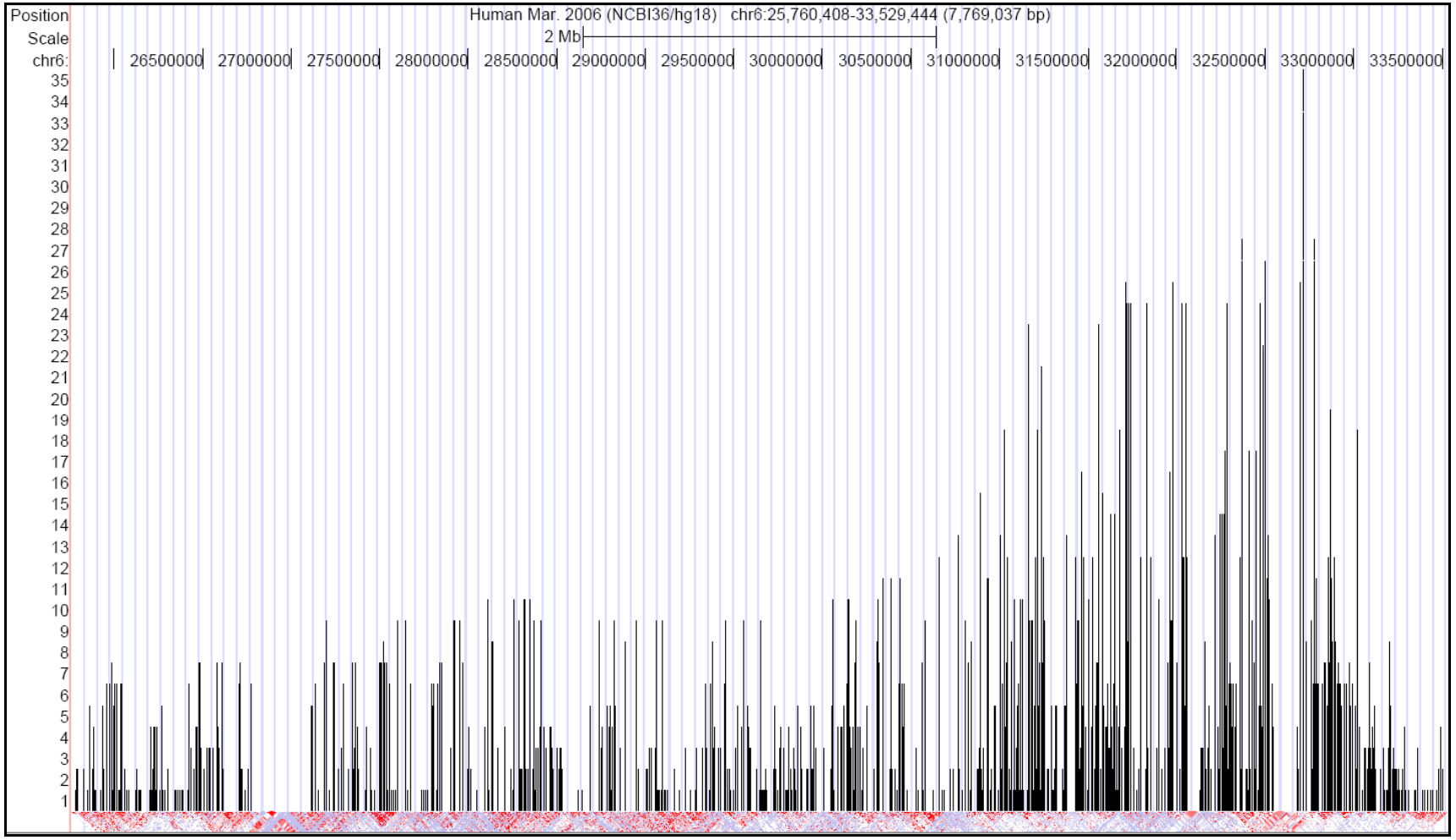


**Figure 46 Extended HLA region on chromosome 6**
Custom track plotting the $-\log_{10}$ Chi square scores (shown on the y axis, from 1 to 35) of the 1274 SNPs in the respective region. Linkage disequilibrium structure (HapMap Linkage Disequilibrium – Phase II – from phased genotypes, LOD LD values, inverted triangle display for the European CEPH CEU population) is shown underneath (hg18 assembly).

Of the 132 associated SNPs, 63 overlap known RefSeq gene regions (respectively 46 genes extended to include 500 bp upstream and 500 bp downstream), see Table 10.

**Table 10 RefSeq genes in the region of interest on chromosome 6**
46 RefSeq genes corresponding to the 63 SNPs associated to the membranous nephropathy trait in the region of interest on chromosome 6 (HLA superlocus).

| gene | associated SNP | $-\log_{10}P$ | position |
|---|---|---|---|
| ZNF187 | rs1419183 | 9.510488 | intron |
| PGBD1 | rs13211507 | 8.886294 | intron |
| ZNF323 | rs853676 | 8.626034 | intron |
| OR12D3 | rs3749971 | 8.678537 | CDS |
| HCG9 | rs3132685 | 9.577058 | intron |
| ZNRD1 | rs8321 | 10.40575 | 3'UTR |
| RNF39 | rs9261290 | 10.24337 | 3'UTR |
| TRIM31 | rs2517598 | 9.232746 | CDS |
| HLA-L | rs3094073 | 11.05298 | CDS |
| HCG18 | rs3130380 | 11.28068 | intron |
| ABCF1 | rs3132610 | 11.57972 | intron |
| KIAA1949 | rs9262143 | 12.79693 | CDS |
| FLOT1 | rs3094127 | 9.371116 | intron |
| VARS2 | rs4678 | 12.76308 | CDS |
| DPCR1 | rs3132580 | 17.71905 | CDS |
| PSORS1C1 | rs3095314 | 12.08189 | intron |
| | rs3130564 | 17.51186 | intron |
| | rs1265099 | 9.240754 | intron |
| PSORS1C2 | -//- | -//- | 3'UTR |
| CCHCR1 | rs720465 | 9.337063 | intron |
| TCF19 | rs7750641 | 21.42089 | CDS |
| PSORS1C3 | rs887464 | 8.77669 | upstream |
| HLA-B | rs2596501 | 9.297098 | downstream |
| MICB | rs2855812 | 14.71701 | intron |
| NFKBIL1 | rs2071591 | 13.61617 | intron |
| LTA | rs1041981 | 14.46855 | CDS |
| BAT2 | rs2242660 | 11.71536 | intron |
| | rs3115663 | 24.51763 | intron |
| | rs1046089 | 13.67704 | CDS |
| | rs9267522 | 24.34018 | CDS |
| BAT3 | rs805303 | 10.7399 | intron |
| | rs3117582 | 24.09989 | upstream |
| BAT4 | rs3130618 | 24.27565 | CDS |
| LY6G6C | rs805294 | 12.41806 | intron |
| MSH5 | rs3131379 | 24.41345 | intron |
| C6orf27 | rs707928 | 11.73123 | intron |
| EHMT2 | rs652888 | 16.49981 | intron |
| | rs535586 | 8.605229 | CDS |
| | rs659445 | 8.605229 | intron |
| C2 | rs558702 | 24.59074 | intron |
| CFB | rs1270942 | 23.79432 | intron |

| | | | |
|---|---|---|---|
| RDBP | rs630379 | 11.93786 | intron |
| SKIV2L | rs437179 | 11.77333 | CDS |
| STK19 | rs389884 | 23.89062 | intron |
| | rs389883 | 11.77333 | intron |
| EGFL8 | rs3096697 | 14.1943 | CDS |
| RNF5 | rs3134945 | 13.82089 | intron |
| RNF5P1 | -//- | -//- | intron |
| PBX2 | rs204994 | 13.98526 | intron |
| | rs204993 | 12.08163 | intron |
| GPSM3 | rs204991 | 17.05445 | intron |
| | rs204990 | 16.739 | intron |
| NOTCH4 | rs3131296 | 24.17224 | intron |
| C6orf10 | rs6935269 | 11.59887 | downstream |
| | rs7775397 | 26.78994 | CDS |
| | rs1003878 | 16.75801 | intron |
| | rs2076537 | 8.87889 | intron |
| | rs3129939 | 16.63777 | intron |
| | rs3129943 | 14.33931 | intron |
| | rs2050189 | 16.89807 | upstream |
| BTNL2 | rs1980493 | 23.94355 | intron |
| HLA-DQA1 | rs2187668 | 35.28596 | intron |
| HLA-DOB | rs11244 | 10.69374 | 3'UTR |
| | rs7383287 | 12.13616 | CDS |

HLA-DQA1

The SNP with the highest P value (rs2187668, $-\log_{10}P = 35.28596$) is located within the first intron of HLA-DQA1, which is a 5 exon (4 of which coding, the last one part of the 3'UTR untranslated region) gene spanning 6,246 base pairs of genomic DNA at chr6:32,713,161-32,719,407 (hg18 genome assembly).

Testis Specific Basic Protein (*TSBP*)

TSBP (previously called C6orf10) is the gene which overlaps with most associated SNPs (rs6935269, rs7775397, rs1003878, rs2076537, rs3129939, rs3129943 and rs2050189). It is a 23 exon (all coding) gene which spans 187,160 base pairs genomic DNA on the minus (-) strand of the GRCh37 / hg19 assembly of the human genome at chr6:32260476-32447634.

TSBP – HLA-DQA1 region

The two genes are flanking a variable region, containing a published deletion at chromosome 6p21.32 (one of the 935 deletions detected by HapMap Phase I data genotype analysis; present in an Yoruban individual NA19202 and spanning 168,339 base pairs on the hg18 genome assembly) [506] and a published copy

number variant of unknown direction (not determined whether gain or loss; one of 1447 copy number variable regions detected by HapMap Phase II data analysis by SNP and BAC microarray) [507], Figure 47.

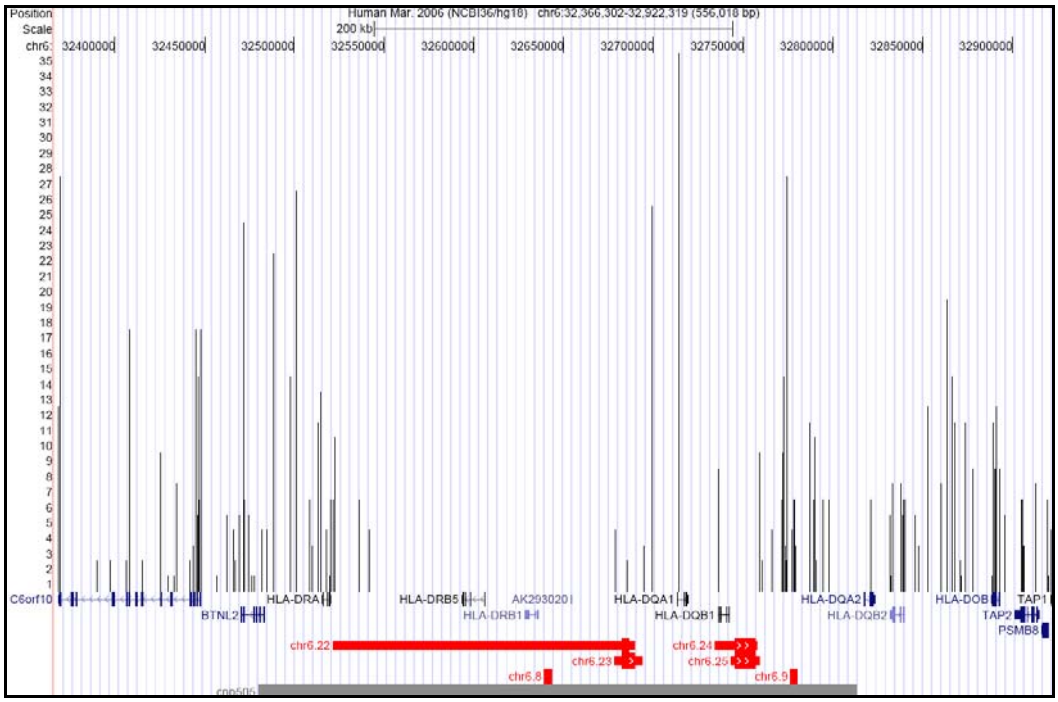


**Figure 47 TSBP – HLA-DQA1 region**
Representation of the region of interest on the UCSC Genome browser (human March 2008 hg18 genome assembly). Shown are the the log -10 chi square scores of the SNPs covering the region after stringent quality control (call rate > 99% etc.), known genes (as defined by UCSC) and structural variation (in red known deletions as described by Conrad *et al.* [506], in gray the gain or loss - unknown direction - variation described by Redon *et al.* [507]).

*Chromosome 13*

The associated SNP marker defining the region of interest on chromosome 13 is rs4943552 (physical position 38504142 bp on the February 2009 assembly GRCh37 / hg19). It is a coding / sense (+) strand T > C/T type single nucleotide variant, a substitution (translation), the ancestral allele being T (present on the positive strand of the chimpanzee *Pan paniscus* genome, on the negative strand of the orangutan *Pongo pygmaeus* genome and on the positive strand of the macaque *Macacus rhesus* genome). It has been validated by cluster, by frequency, by 2 hit / 2 allele, by HapMap and by 1000 genomes. Its average heterozygosity is 0.278 +/- 0.248. The weight quality assignment by dbSNP of

rs4943552 is 1. Its functional type is unknown, it lies at 60 kb upstream of the *TRPC4* gene.

TRPC4

The Transient Receptor Potential cation channel, subfamily C, member 4 gene spans 233,165 base pairs on the negative (antisense) strand of chromosome 13q13.3. It has 11 exons, 10 of which are coding, the first one transcribed in a fragment of the 5' UTR (untranslated region). Its product is the short transient receptor potential 4, a protein predicted to be 804 amino acids long.

*Chromosome 16*

The SNP marker associated to membranous nephropathy on chromosome 16 is rs2221705 (physical position 79362411 bp on the February 2009 assembly GRCh37 / hg19). It is a coding / sense (+) strand T > C/T type single nucleotide variant, a substitution (translation), the ancestral allele being T (present on the positive strand of the chimpanzee *Pan paniscus* genome, the negative strand of the orangutan *Pongo pygmaeus* genome and the positive strand of the macaque *Macacus rhesus* genome). It has been validated by cluster, by frequency, by 2 hit 2 allele, by HapMap and by 1000 genomes. Its average heterozygosity is 0.157 +/- 0.232. The weight quality assignment by dbSNP of rs2221705 is 1. Its functional type is unknown, it is located downstream from the *WWOX* gene.

WWOX

WW domain-containing oxidoreductase, is a nine exon (all coding) gene that spans 1,113,013 base pairs of genomic DNA on the positive (coding) strand (at chr16:79,238,846-79,246,564 on the February 2009 GRCh37 / hg19 assembly). It encodes a 414 amino acid protein which contains two WW domains and a SRD (short chain dehydrogenase / reductase) domain therefore belonging to the SDR family. Its expression is highest in hormonally regulated tissues (testis, ovary, prostate). The presence of the SRD domain and expression pattern suggests a role for this protein in steroid metabolism. Through similarity with its mouse homologue it has also been suggested that WWOX is important in tumor necrosis factor alpha induced apoptosis.

*Cases separated from controls QC filtering (separated scenario)*

**Separated high stringency scenario (CR > 99)**

The merging of the two initialy separated datasets (cases and controls) 'dilutes' the proportion of false calls, *i.e.* a small yet significant proportion of genotyping errors for one of the separated datasets might lose its significance when the threshold is applied for the merged dataset. To alleviate this pitfall the results of quality control performed on the common matrix was compared with quality control outcomes on the two datasets filtered separatedly are as follows:

*Cases*

A matrix with 335 rows (UK cases) and 351,515 columns (351,507 assessed markers and 8 columns of phenotypic descriptors) was created. The phenotypic descriptors are encoded as a binary variable (where 0 = no and 1 = yes) as follows: column 1: Ethnicity=YRI? (Yoruban), column 2: Ethnicity=CHB? (Han Chinese), column 3: Ethnicity=Caucasoid?, column 4: Ethnicity=UK?, column 5: Ethnicity=FR?, column 6: Ethnicity=NL?, column 7: gender (0 = male, 1 = female), column 8: status (0 = unaffected, 1 = affected). From the initial 351,507 assessed markers, 338,680 markers were found to have two alleles, 6,174 markers presented one allele and 6,653 markers had missing genotypes (no alleles). Quality assurance genotype filtering by marker (parameters) was performed with the following parameters: call rate < 0.99; minor allele frequency (MAF) < 0.01; Hardy-Weinberg Equilibrium *P* (HWE *P*) < 0.001.

39,920 (11.36 %; from all markers 351,507 = 100%) markers did not pass the filtering threshold (22,413 due to the call rate limitation at 99%; 8,826 due to the MAF limitation; 11,422 due to the HWE *P* condition; *cave*: some markers did not pass more than one of the filtering criteria (therefore: 22,413 + 8,826 + 11,422 = 42,661 > 39,920).

*Controls*

Similarly a matrix with 349 rows (UK controls) and 317,511 columns (317,503 assessed markers and 8 columns of phenotypic descriptors) was used. From the

assessed makers 317,419 markers were found to have two alleles and 84 markers were found to be uninformative (having just one allele).
Quality assurance genotype filtering by marker (parameters) was performed with the following parameters: call rate < 0.99; minor allele frequency (MAF) < 0.01; Hardy-Weinberg Equilibrium (HWE *P*) < 0.001.

44,818 (14.12%; from the total number of markers 317,503 = 100%) markers did not pass the filtering threshold (35,878 due to the call rate limit; 235 due to the MAF limit; 10528 due to the HWE *P* limit; *cave*: some markers did not pass more than one of the filtering criteria, therefore 35,878 + 235 + 10528 = 46,641 > 44,818).

*Case-control matrix overlap*

Pre QC there were 338,680 informative markers (cases) and 317,419 informative markers (controls). After QC the informative markers from the controls matrix that passed the filtering criteria (272,685) were intersected with the informative cases markers which passed the filter (311,587). This operation yielded 255,264 informative markers that are common for the two datasets (cases and controls; 75.46 % out of 338,680 = 100% cases and 80,51% out of 317,419 = 100% controls).

*Second HWE cleaning step*

Through the merging of the case and control cohorts, the structure of the population was changed, therefore the Hardy-Weinberg Equilibrium filtering was reiterated on the common case-control matrix. Acordingly 366 markers were filtered out (for a *P* value of the HWE < 0.001). The number of markers passing both cleaning steps remained thus 254,898.

*Population structure*

The estimate for stratification (genomic inflation factor, $\lambda$) is for the case control combined matrix 1.13 (which is close to 1, therefore not suggestive for stratification). The genotype principal component analysis was performed computing up to 100 components using the additive genetic model for PCA to detect stratification. The values of the Eigenvalues thus computed drop

exponentially from 2.22 towards 0.6, suggesting that most of the differences can be captured within the first two principal components, Figure 19. These first two components (with the eigenvalues – EV – of 2.22 and 1.04) were shown on the pre-append (separated) two step quality control filtering UK dataset orthogonal projection in Figure 21.

*QQ plot*

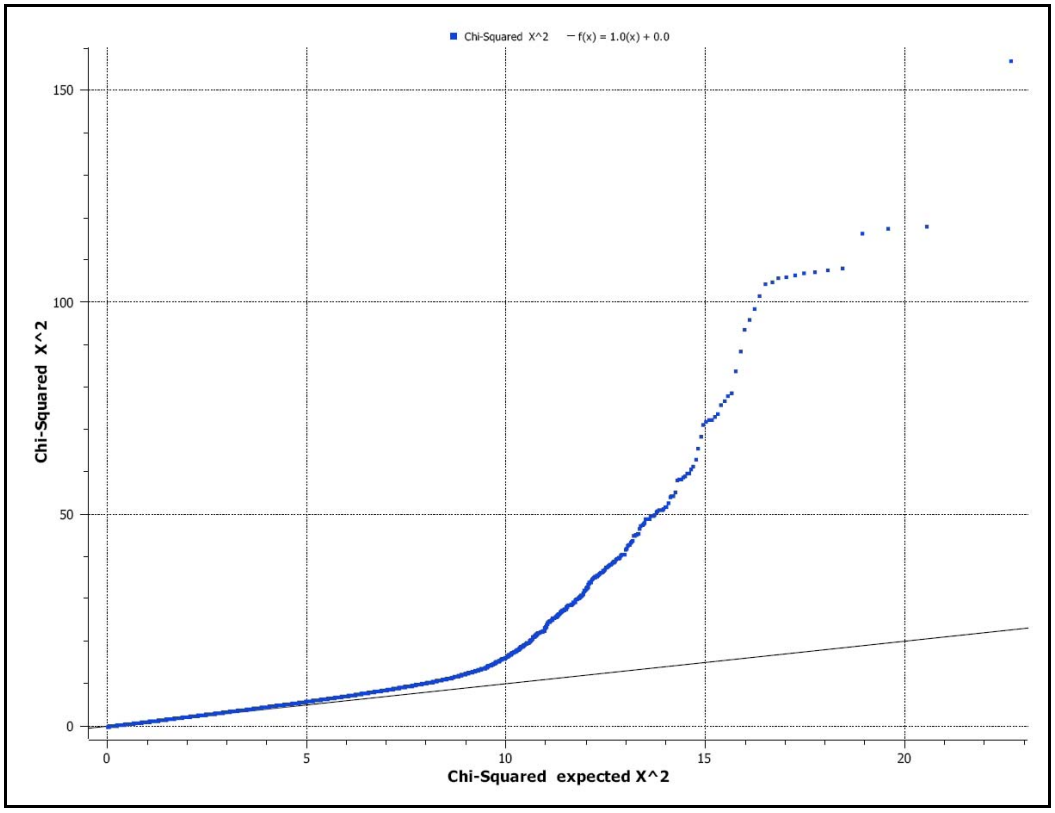


**Figure 48 Q-Q plot: UK dataset, separated high stringency**
Q-Q plot for the Chi-squared test of the basic allelic test (D *versus* d) plotting the actual association $-\log_{10}P$ to the expected $-\log_{10}P$ in cases *versus* controls for the UK dataset after pre-append (separated) two step quality control filtering for call rate > 99%, minor allele frequency MAF > 0.01, Hardy-Weinberg equilibrium HWE $P > 0.001$.

*Results*

*Basic Allelic Test*

All 254,898 common informative markers filtered through two quality control steps (pre append and post append) were used for the analysis. The threshold for genome-wide significance corrected for multiple comparisons (Bonferroni correction) was set at $P_1$ = 0.001 / 254,898 and then further adjusted for the genomic inflation factor $P = P_1 / 1.13 = 3.462 \times 10^{-9}$; $-\log_{10}P = 8.46$.

Basic allelic test (D *versus* d) the chosen test statistic (for the 2X2 contingency table) was the Chi-squared test. 124 markers were found to be associated with the trait (MN) at this genome-wide significance threshold, Figure 49.

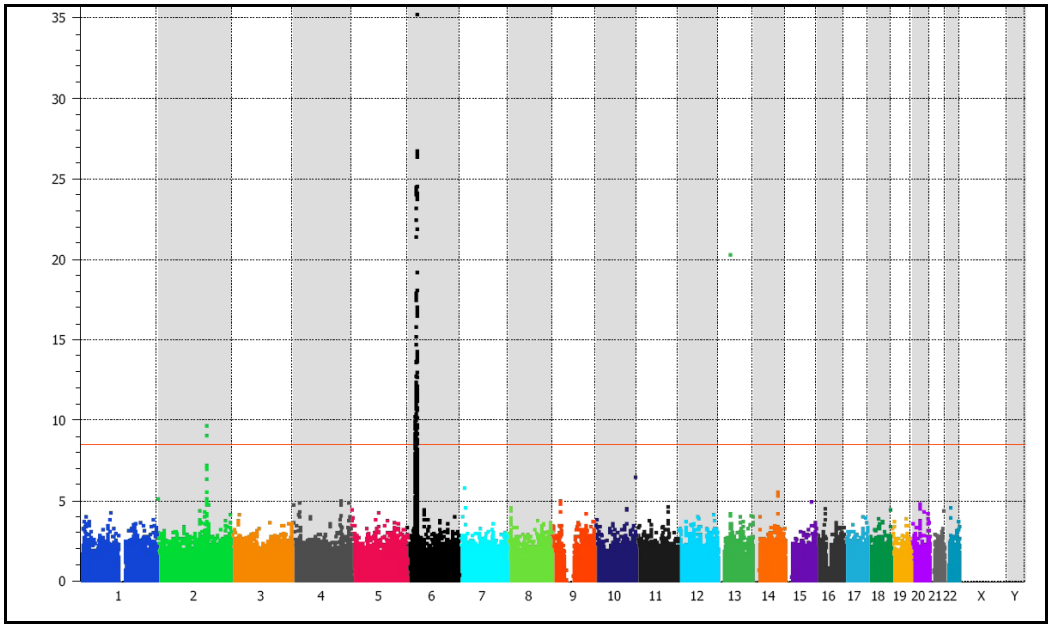


**Figure 49 BAT Manhattan plot, UK dataset, separated high stringency scenario**
Manhattan plot for the Chi-squared test of the basic allelic test (BAT) comparing minor *versus* major alleles (D *versus* d) in cases *versus* controls for the UK dataset after pre-append (separated) two step quality control filtering for call rate > 99%, minor allele frequency > 0.01, Hardy-Weinberg equilibrium P > 0.001. The level of significance (corrected for multiple comparisons and for genomic inflation) is represented by the red line at $-\log_{10} = 8.46$. (Y axis = $-\log_{10}P$, X axis = physical positions of the markers corresponding to the human February 2009 GRCh37/hg19 genome assembly, the chromosomes being plotted successively, shaded and coloured in different colours for better visualisation).

*Genotype Association Test*

When performing the Genotypic association test (DD *versus* dd *versus* Dd; the chosen test statistic being the Chi-squared test, with a threshold for genome-wide significance corrected for multiple comparisons and the genomic inflation factor ($P_1$ = 0.001 / 254,898 / 1.07 = 3.666 x$10^{-9}$; -$\log_{10}P$ = 8.43) 101 markers were found to be associated with the trait (MN) at the genome-wide significance threshold.

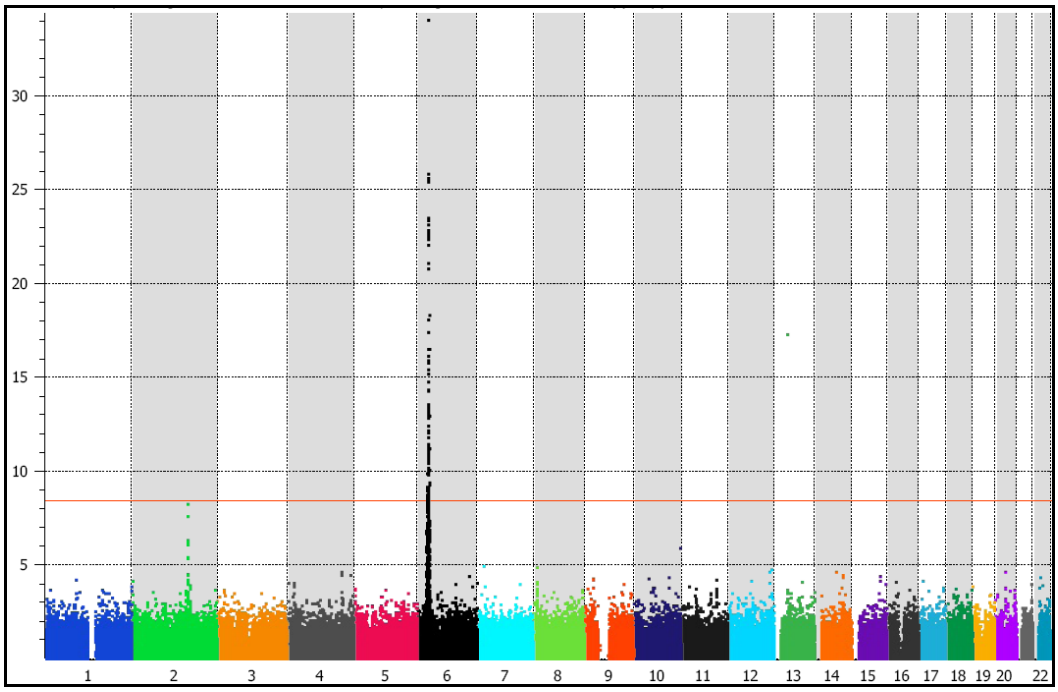


**Figure 50 GAT Manhattan plot, UK dataset, separated high stringency scenario**
Manhattan plot for the Chi-squared test of the genotype association test (GAT) comparing each of three possible genotype configurations (dd, dD, DD) in cases *versus* controls for the UK dataset after pre-append (separated) two step quality control filtering for call rate > 99%, minor allele frequency > 0.01, Hardy-Weinberg equilibrium P > 0.001. The level of significance (corrected for multiple comparisons and for genomic inflation) is represented by the red line at -$\log_{10}$ = 8.43. (Y axis = -$\log_{10}P$, X axis = physical positions of the markers corresponding to the human February 2009 GRCh37/hg19 genome assembly, the chromosomes being plotted successively, shaded and coloured in different colours for better visualisation).

*Additive Model*

The association test on an additive model (dd > Dd > DD; the chosen statistic being the Correlation Trend test, with a threshold for genome-wide significance corrected for multiple comparisons and for stratification set at $P$ = 0.001 / 254,898 / 1.13 = 3.47 x$10^{-9}$; -$\log_{10}P$ = 8.46) yielded 124 markers associated with the trait (MN) at the genome-wide significance threshold.

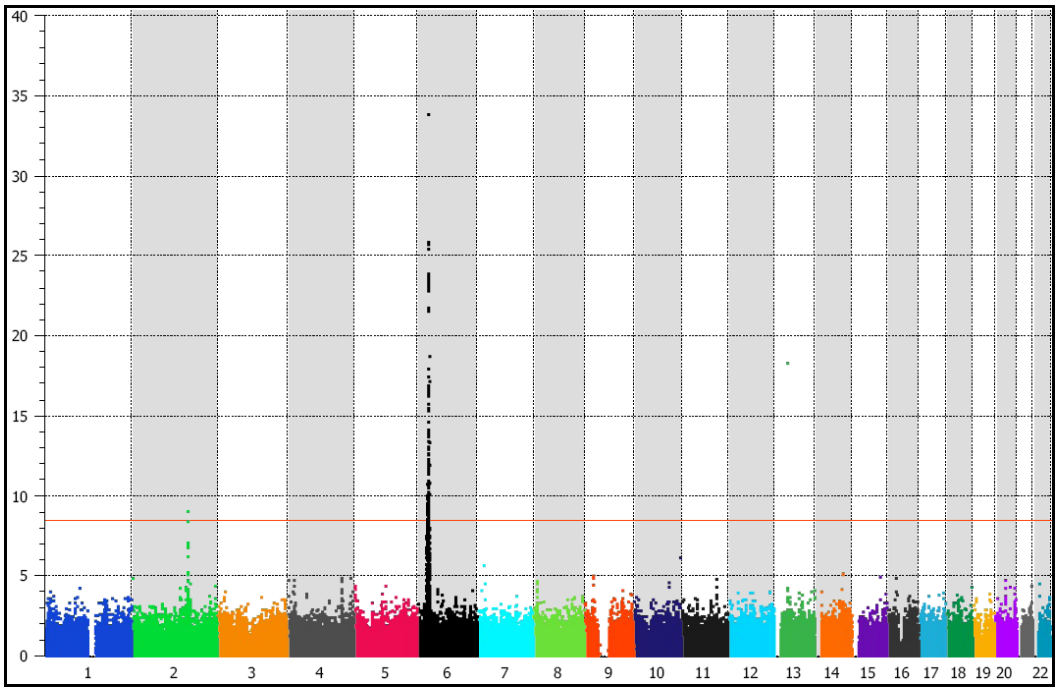


**Figure 51 AM Manhattan plot, UK dataset, separated high stringency scenario**
Manhattan plot for the Chi-squared test of the additive model (AM) test comparing the three possible genotype configurations (under the assumption of an increasing effect of the causative allele on the phenotype: dd > dD > DD) in cases *versus* controls for the UK dataset after pre-append (separated) two step quality control filtering for call rate > 99%, minor allele frequency > 0.01, Hardy-Weinberg equilibrium P > 0.001. The level of significance (corrected for multiple comparisons and for genomic inflation) is represented by the red line at -$\log_{10}$ = 8.46. (Y axis = -$\log_{10}P$, X axis = physical positions of the markers corresponding to the human February 2009 GRCh37/hg19 genome assembly, the chromosomes being plotted successively, shaded and coloured in different colours for better visualisation).

*Dominant Model*

The association test on an dominant model (DD, Dd *versus* dd; the chosen test statistic being Chi-squared, with a threshold for genome-wide significance corrected for multiple comparisons set at $P = 0.001 / 254{,}898 / 1.11 = 3.53 \times 10^{-9}$; $-\log_{10}P = 8.45$) yielded 95 markers associated with the trait (MN) at the genome-wide significance threshold.

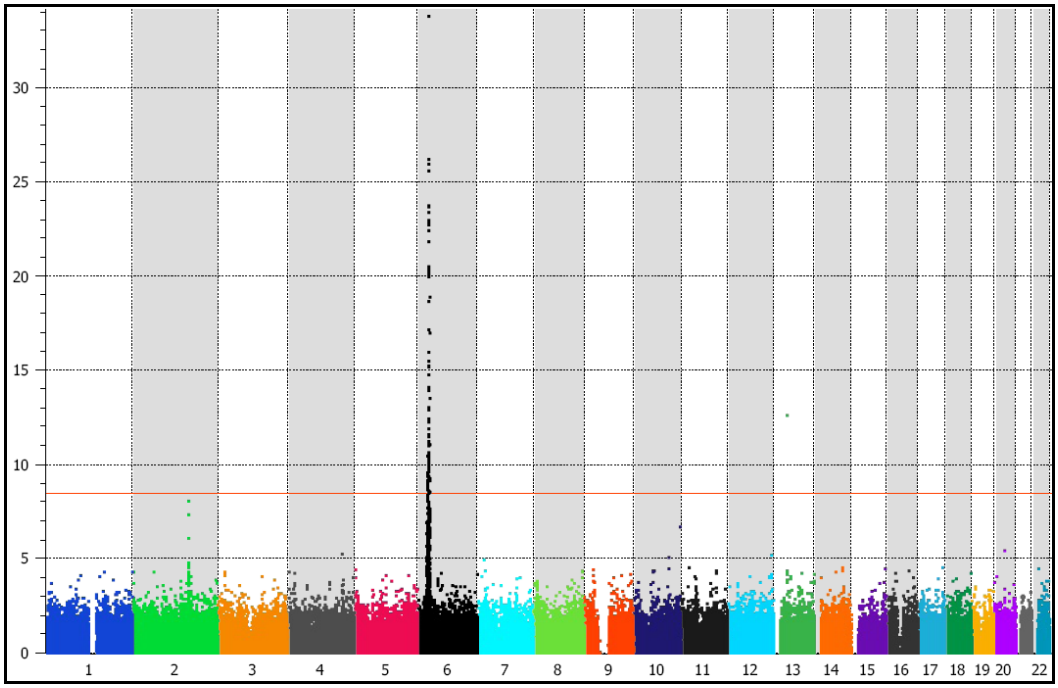


**Figure 52 DM Manhattan plot, UK dataset, separated high stringency scenario**
Manhattan plot for the Chi-squared test of the dominant model (DM) test assumes a binary type influence on the phenotype given the presence or absence of causative allele (dd *versus* dD and DD) in cases *versus* controls for the UK dataset after pre-append (separated) two step quality control filtering for call rate > 99%, minor allele frequency > 0.01, Hardy-Weinberg equilibrium P > 0.001. The level of significance (corrected for multiple comparisons and for genomic inflation) is represented by the red line at $-\log_{10} = 8.45$. (Y axis = $-\log_{10}P$, X axis = physical positions of the markers corresponding to the human February 2009 GRCh37/hg19 genome assembly, the chromosomes being plotted successively, shaded and coloured in different colours for better visualisation).

*Recessive Model*

The association test on an recessive model (DD vs. Dd, dd; the chosen test statistic being Chi-squared, with a threshold for genome-wide significance corrected for multiple comparisons set at $P = 0.001 / 254{,}898 / 1.09 = 3.59 \times 10^{-9}$; $-\log_{10}P = 8.44$) yielded only 12 markers associated with the trait (MN) at the genome-wide significance threshold.

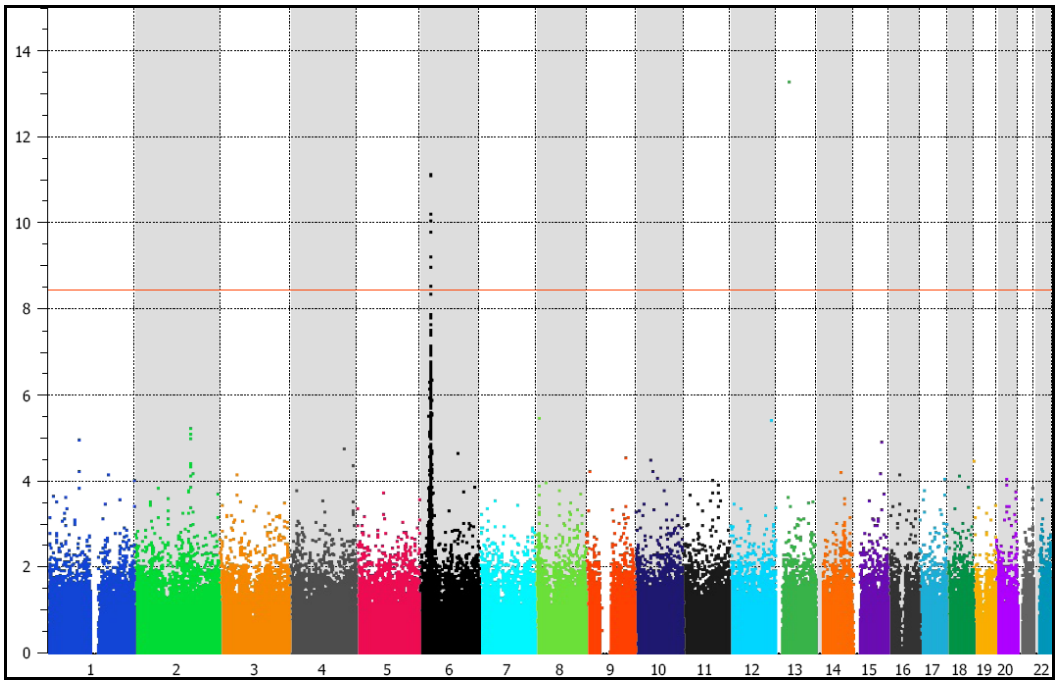


**Figure 53 RM Manhattan plot, UK dataset, separated high stringency scenario**
Manhattan plot for the Chi-squared test of the recessive model (RM) test assumes a binary type influence on the phenotype where only the homozygous presence of the causative allele (both alleles genotype) determines the phenotype (dd and dD *versus* DD) in cases *versus* controls for the UK dataset after pre-append (separated) two step quality control filtering for call rate > 99%, minor allele frequency > 0.01, Hardy-Weinberg equilibrium P > 0.001. The level of significance (corrected for multiple comparisons and for genomic inflation) is represented by the red line at $-\log_{10} = 8.44$. (Y axis = $-\log_{10}P$, X axis = physical positions of the markers corresponding to the human February 2009 GRCh37/hg19 genome assembly, the chromosomes being plotted successively, shaded and coloured in different colours for better visualisation).

*Haplotype analysis*

To assess visualy the haplotype block structure of the regions of interest, linkage disequilibrium was plottet graphically in triangular heatmap plots, the the colour of the point at the intersection of the projection of the lines originating in the queried markers reflecting a measure of LD (either $r^2$ or D'). The complexity of the HLA region on chromosome 6 is reflected in its pattern of increased linkage disequilibrium, see Figures 54 and 55.

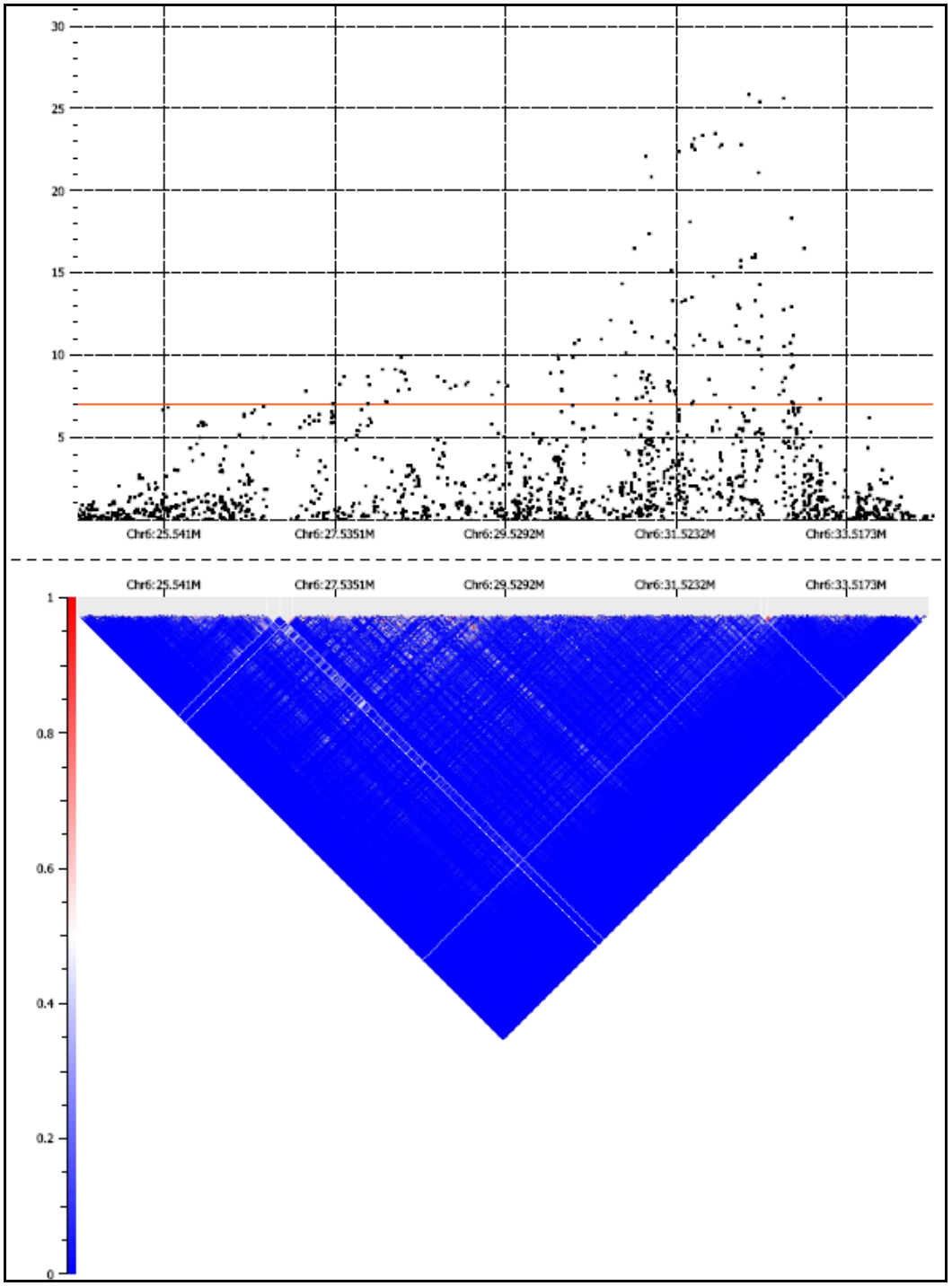


**Figure 54 LD plot, chromosome 6 (HLA region)**
UK case control association dataset; X axis: physical position; Y axis: association values $-\log_{10}P$; inverted triangle: heatmap of linkage disequilibrium ($r^2$) color coded blue (0) to red (1) according to the legend.

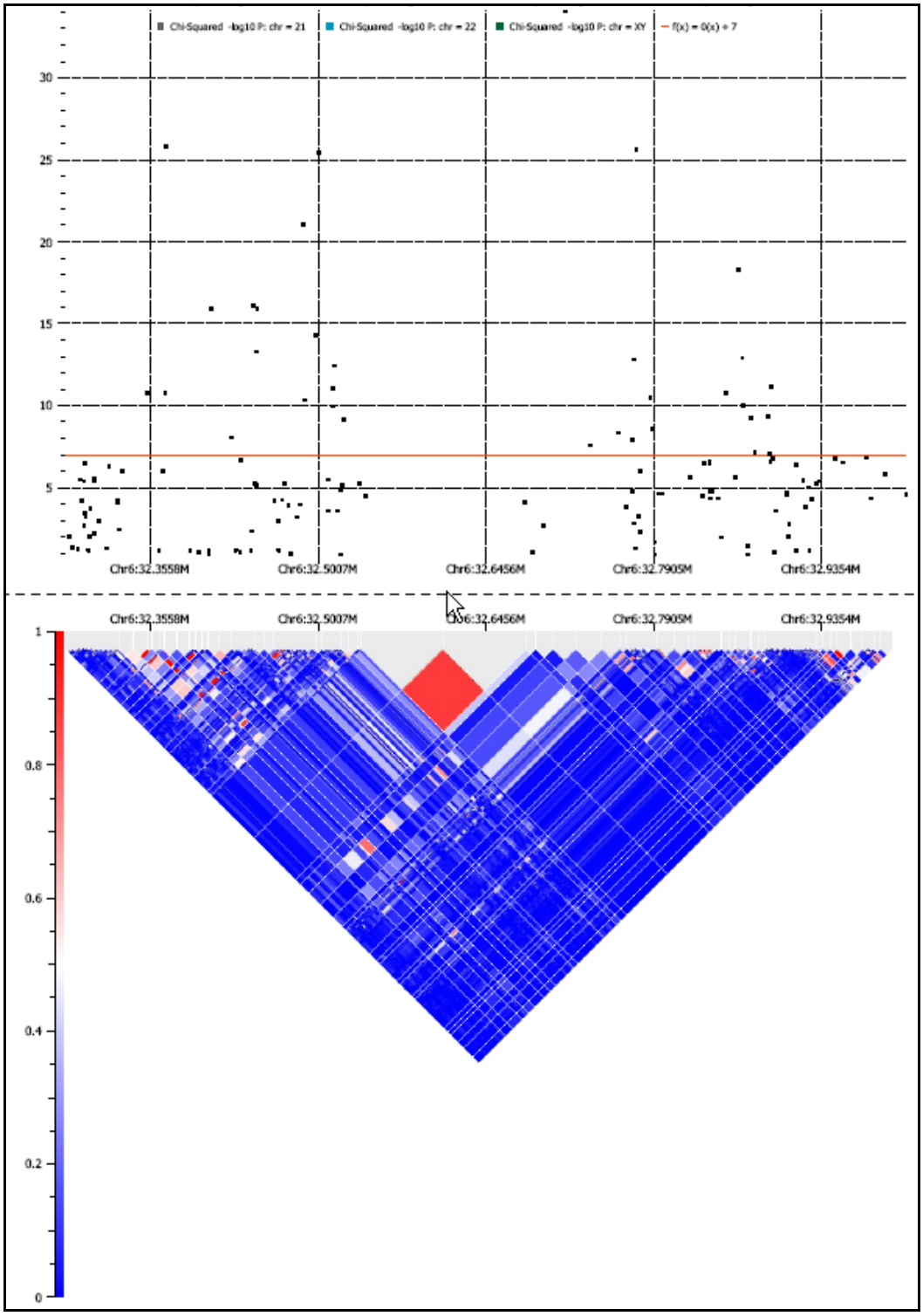


**Figure 55 LD plot, peri-PLA2R1 region**
Chromosome 6 (HLA region) linkage disequilibrium (LD) plot zoom in. Y axis: association values $-\log_{10}P$; inverted triangle: heatmap of linkage disequilibrium ($r^2$) color coded blue (0) to red (1) according to the legend.

The haplotype association test treats different haploblocks as combined markers.

*Haplotype block definition*

For our analysis, haplotype (LD) blocks were computed in SNP & Variation Suite v7.3.1 according to their definition by Gabriel *et al.* [428] (in order to minimize historical recombination). LD blocks were considered to have a D' at an upper confidence bound > 0.98 and a lower confidence bound > 0.7 at 98% confidence. Pairs were rejected with an upper bound < 0.9. The minimal allele frequency of markers assessed in order to define blocks was set at 0.05. Blocks were limited to less than 30 markers per block with a length lower than 160 kb. According to these criteria, 18,393 blocks including 42,595 markers were defined.

*Haplotype association test*

The Chi-squared test statistics significance threshold ($P = 0.001$) corrected for multiple comparisons (in this case the number of blocks: 18,393) was set at $-\log_{10}P = 7.26$. Association was computed per block and per haplotype (haplotypes were estimated according to the expectation-maximisation – EM – algorithm [508]; the frequency threshold for haplotype estimation was set at 0.01, EM convergence tolerance was set at 0.0001 and the maximum number of EM iterations was set at 50) see Figures 56 and 57.

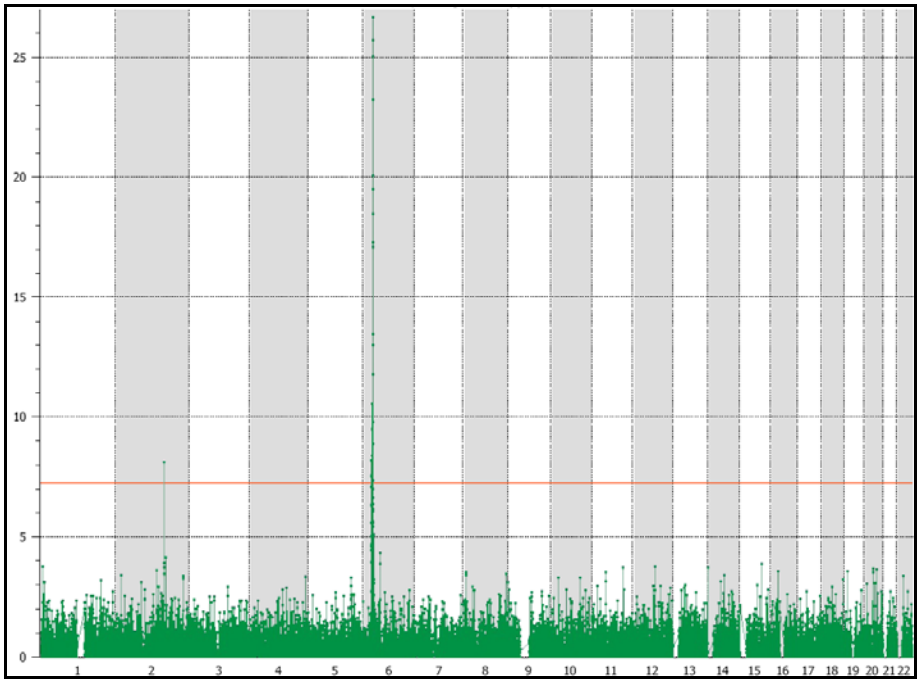


**Figure 56 HAT per block Manhattan plot, UK, separated high stringency scenario**
Results of the haplotype association test (HAT) per block analysis performed on the UK dataset. The level of significance (corrected for multiple comparisons) is represented by the red line at $-\log_{10} = 7.26$. (Y axis = $-\log_{10}P$, X axis = physical positions of the markers corresponding to the human February 2009 GRCh37/hg19 genome assembly, the chromosomes have been plotted successively, and shaded alternatively for better visualisation.

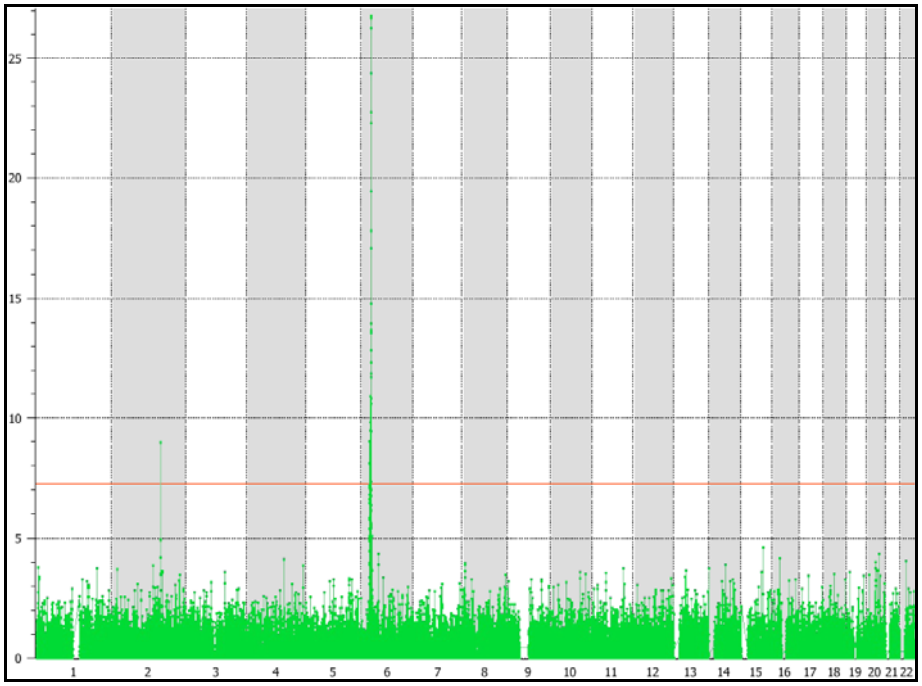


**Figure 57 HAT per haplotype Manhattan plot, UK, separated high stringency**
Results of the haplotype association test (HAT) per haplotype analysis performed on the UK dataset. The level of significance (corrected for multiple comparisons) is represented by the red line at $-\log_{10} = 7.26$. (Y axis = $-\log_{10}P$, X axis = physical positions of the markers corresponding to the human February 2009 GRCh37/hg19 genome assembly, the chromosomes have been plotted successively, and shaded alternatively for better visualisation.

**Separated permissive scenario (CR > 90)**

As mentioned before, merging of the two initialy separated datasets (cases and controls) 'dilutes' the proportion of markers which should be filtered out. Conversely, using a high call rate threshold for each separated dataset when performing a non-merged analysis, might become excessive. Therefore we tried a more permissive scenario filtering for call rates of under 90% (CR < 0.9). The permissivity for call rate was balanced through the use of a higher stringency criterion for the Hardy-Weinberg Equlibrium (HWE $P < 0.01$). The minor allele frequency was filtered as the 1% threshold as per the definition of polymorphisms (MAF < 0.01). We performed three separated steps of quality control for the separated scenario: first filtering step on the cases matrix followed by a second filtering step on the controls matrix and a third filtering step on the merged cases and controls matrix.

*Cases*

The initial case matrix with 335 rows (UK cases) and 351,515 columns (351,507 assessed markers and 8 columns of previously described phenotypic descriptors) was used. Of the assessed 338,680 markers are bialellic *ergo* informative. From the total of 351,507 markers 30,532 (8.67%) markers did not pass the filtering steps. This is similar to the 39,920 markers (11.36%) filtered in the separated high stringency scenario. The informative markers from the cases matrix which passed the filter were in the end 320,975.

*Controls*

The control matrix with 349 rows (UK controls) and 317,511 columns (317,503 assessed markers and 8 columns of phenotypic descriptors) was used. From the assessed makers 317,419 markers were found to be informative (biallelic).
19,959 (6.28% of the informative) markers did not pass the filtering procedure. This compares favorably to the 44,818 markers (14.12%) which were filtered in the separated high stringency scenario. The informative markers from the controls matrix that passed the filtering criteria were finally 297,543.

*Case-control matrix overlap*

After the quality filtering, the informative filtered markers from the controls matrix were intersected with the informative filtered markers from the cases matrix. This operation yielded 282,997 informative (biallelic) markers that are common for the two datasets (cases and controls; 83.56% out of the initial 338,680 = 100% informative markers for cases and 89.15% out of the intitial 317,419 = 100% informative markers for controls).

*Second HWE cleaning step*

Through the merging of the case and control cohorts, the structure of the population was changed, therefore the filtering (call rate < 0.9, MAF < 0.01, HWE P < 0.01) was reiterated on the common case-control matrix. Acordingly 1,988 markers were filtered out (for a *P* value of the HWE < 0.01). The number of markers passing both cleaning steps remained thus 281,009.

*Population structure*

The estimate for stratification (genomic inflation factor, $\lambda$) is for the case control combined matrix 1.15 (which is close to 1, therefore not suggestive for stratification).

*Results*

*Significance threshold*

The significance level that we adopted for our analyses ($P = 0.001$) was corrected for multiple testing by dividing it with the number of informative markers (N = 281,009). Correction for stratification is obtained by dividing the new significance value by $\lambda$ (the genomic inflation factor; $\lambda = 1.15$). The corrected $P = 0.001 / 281{,}009 / 1.15 = 3.1 \times 10^{-9}$; $-\log_{10}P = 8.51$.

*QQ plot*

The QQ plots, comparing the dataset with and omitting the SNPs within the extended HLA complex, show that the most extreme dispersion is in the HLA region, which is probably due to at least one association in this region and strong linkage disequilibrium within the associated region, Figure 58.

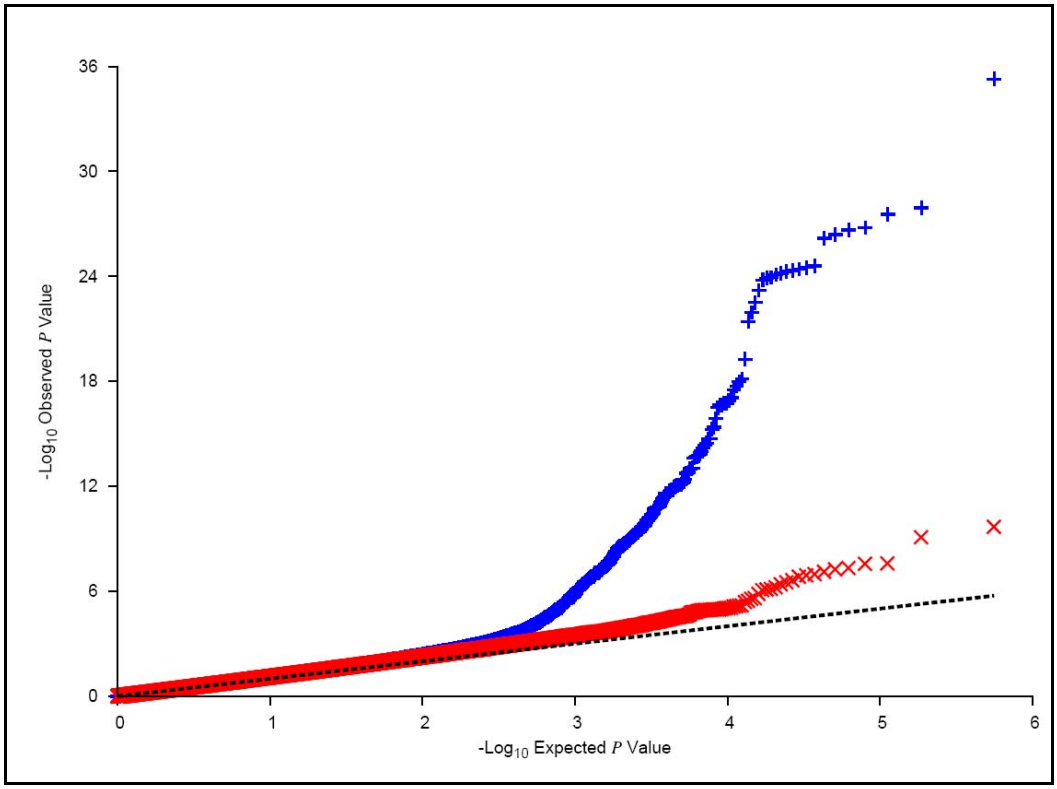


**Figure 58 QQ plots, UK, separated permissive scenario**
Quantile–quantile plot for observed (y axis) versus expected (x axis) -$\log_{10}P$ values comparing the impact of the extended MHC region (blue: all SNPs tested; red: SNPs within the extended MHC region excluded; dotted black line: null distribution).

*Basic allelic test*

At this corrected significance threshold, 146 markers were associated with the trait of which two lie within the PLA2R1 / ITGB6 region on chromosome 2 (rs4664308 – within PLA2R1 and rs1870102 – ITGB6), and 144 lie within the extended MHC region on chromosome 6.

### 3.2.1.8 Comparison of the different filtering scenarios

We used the basic allelic test (BAT; the simplest and most robust test) in order to compare the positive association hits for the different filtering scenarios that we employed (no filtering, stepwise filtering based on the call rate, stepwise filtering based on the minor allele frequency, stepwise filtering based on Hardy-Weinberg equilibrium; combined filtering performed on the case-control matrix together, combined filtering performed on cases and control matrices separatedly at different stringencies), Table 11.

**Table 11 Comparison of the effects of different filtering scenarios**

CR: call rate, MAF: minor allele frequency, HWE: Hardy-Weinberg Equilibrium, λ: genomic inflation factor

| | stringency | CR | MAF | HWE | markers | λ | $-\log_{10}P$ | associations |
|---|---|---|---|---|---|---|---|---|
| NO FILTERING | - | - | | | 305,452 | 1.16 | 8.54 | 188 |
| STEPWISE CR | low | 96 | - | - | 298,178 | 1.15 | 8.53 | 159 |
| | high | 99 | - | - | 279,361 | 1.14 | 8.5 | 145 |
| | max | 100 | - | - | 109,909 | 1.14 | 8.09 | 60 |
| STEPWISE MAF | min | - | 0.01 | - | 305,273 | 1.16 | 8.54 | 188 |
| | med | - | 0.05 | - | 297,918 | 1.16 | 7.54 | 185 |
| | max | - | 0.1 | - | 263,983 | 1.17 | 8.48 | 180 |
| STEPWISE HWE | low | - | - | 0.0001 | 296,392 | 1.15 | 8.53 | 172 |
| | high | - | - | 0.001 | 295,636 | 1.15 | 8.53 | 171 |
| | extreme | - | - | 0.01 | 291,838 | 1.15 | 8.52 | 168 |
| COMBINED together | low | 96 | 0.01 | 0.001 | 288,664 | 1.14 | 8.52 | 151 |
| | high | 99 | 0.01 | 0.001 | 270595 | 1.13 | 8.49 | 137 |
| COMBINED separated | high | 99 | 0.01 | 0.001 | 254,898 | 1.13 | 8.46 | 124 |
| | permissive | 90 | 0.01 | 0.01 | 281,009 | 1.15 | 8.51 | 146 |

We chose the permissive scenario for replication, after weighing the number of markers lost during the filtering with the overall cleanliness of the data (assessed by number of isolated – putative false positive association signals in regions other than *PLA2R1* on chromosome 2 and extended HLA on chromosome 6).

### 3.2.2 Replication

#### 3.2.2.1 French

*Selection*

**Cases**

75 kidney biopsy confirmed membranous nephropathy cases of French origin were selected (courtesy of Professor Pierre Ronco), Paris. Of these 58 were males and 17 females, the resulting gender ratio being 1: 3.41. No maneouvers to enrich the cohort for specific disease-predisposing alleles were performed.

**Controls**

157 French controls (clinically not affected by membranous nephropathy) were assessed, of which 66 were males and 91 females (leading to a gender ratio of 1: 0.72).

*Genotyping*

Both cases and controls were genotyped on a Illumina Human370k duo v1 platform at the same time in the same laboratory (Genepole). 339,845 SNPs were successfully genotyped.

*SNP encoding and mapping*

SNPs encoded as *Forward* (corresponding to dbSNP convention).

The Illumina370Map marker map was used. 333,616 markers were successfully mapped, 6229 markers from the genotyping chip were not present in the map file.

The markers not present in the Illumina map file were checked in dbSNP130 (accessed through the UCSC genome browser table function, build version hg18). 6 markers (rs2662733, rs28625410, rs34131647, rs35588459, rs4215, rs4292042) were not found in db SNP 130 (were dropped from dbSNP 128. Of

the remaining markers 22 were not uniquely mapped on hg18 (19 duplicates, 2 triplicates, 1 occured five times). Overall 6,201 markers were added to the map, increasing the number of successfully mapped markers to 339,817.

*Quality Control*

**Together permissive**

The quality control procedure was performed on the common (case and control) matrix [492] (assuming the genotyping has been performed at the same time for cases and controls together). For the French replication cohort this was indeed the case. From the total of 339,845 genotyped markers, 333,817 markers had two alleles, 6,028 markers were monoallellic, therefore uninformative. The criteria for quality assurance genotype filtering by marker were the ones established for the permissive scenario, that is: call rate CR < 90%; minor allele frequency (MAF) < 0.01; Hardy-Weinberg Equilibrium (HWE P) < 0.01.
24,796 (7.43 %; genotyped markers 333,817 = 100%) markers did not pass the filtering threshold.

**QQ plot**

The QQ plots (Figure 59) comparing the dataset with and omitting the SNPs within the extended HLA complex, show – similarly to the UK exploratory scenario, but with a lesser general dispersion, probably due to the fact that cases and controls were genotyped under more similar conditions (Figure 58) - that the most extreme dispersion is within the HLA region.

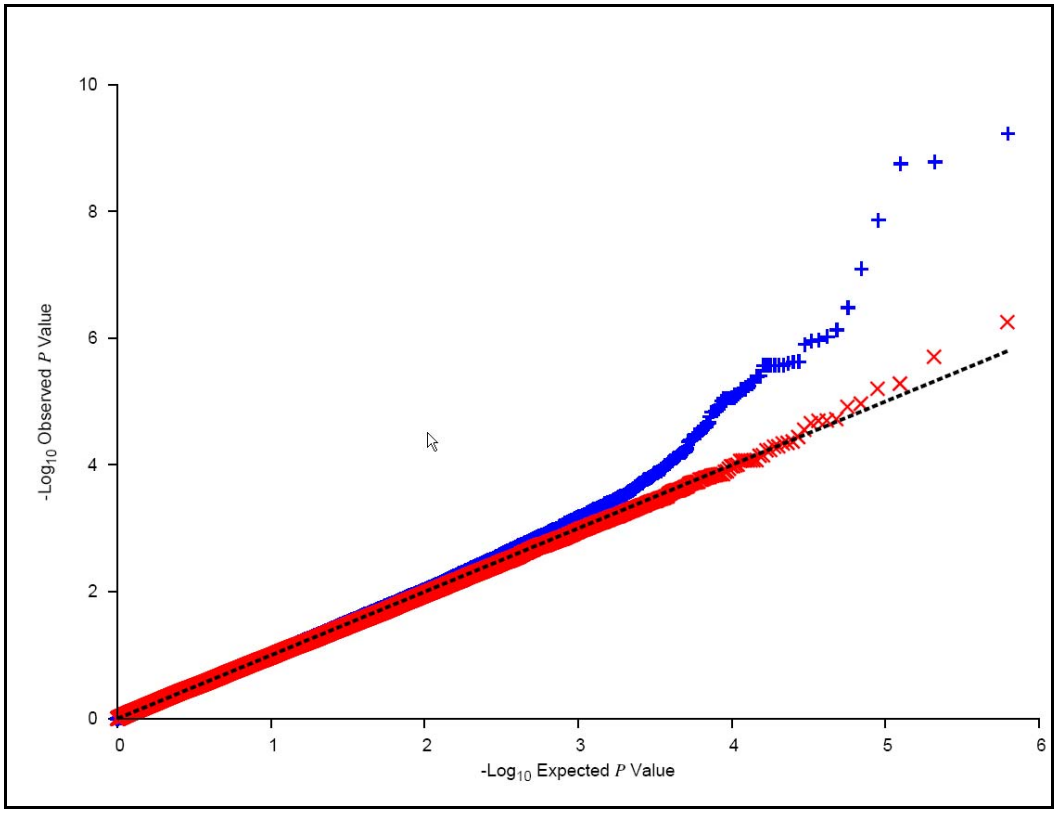


**Figure 59 QQ plots, FR replication, separated permissive scenario**
Quantile–quantile plot for observed (y axis) versus expected (x axis) -$\log_{10}P$ values comparing the impact of the extended MHC region (blue: all SNPs tested; red: SNPs within the extended MHC region excluded; dotted black line: null distribution).

## Basic Allelic Test

All 315,049 markers that passed the quality control were used for the analysis with following parameters: the first genetic model assessed was the basic allelic test (D vs. d); missing values were not used; input data was not corrected for stratification (the inflation factor λ found for chi-squared was 1.02, close to 1 therefore not suggestive for stratification).

The chosen statistic (of the 2X2 contingency table) for the basic allelic test was the Chi-squared test. The threshold for genome-wide significance corrected for multiple comparisons (Bonferroni correction) was set at $P = 0.001 / 315{,}049 = 3.17 \times 10^{-9}$; further corrected for stratification $1.6 \times 10^{-7} / 1.02 = 3.11 \times 10^{-9}$; $-\log_{10} = 8.5$.

3 markers (rs9272192, rs9273327, rs2187668) were found to be associated with the trait (MN) at the genome-wide significance threshold, see Figure 60.

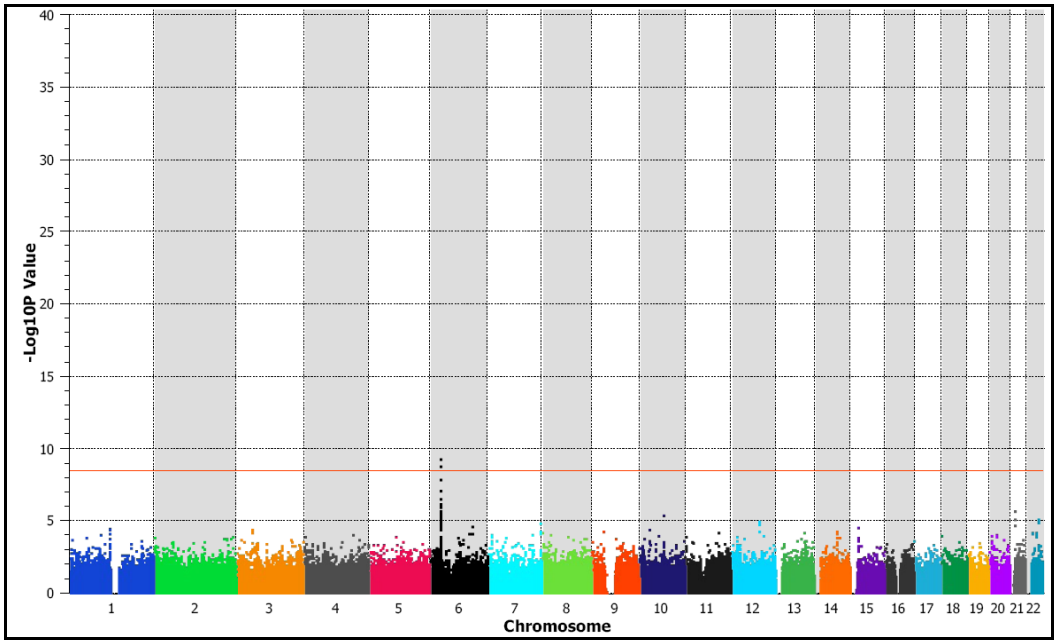


**Figure 60 BAT Manhattan plot, French replication dataset**
Manhattan plot for the Chi-squared test of the basic allelic test (BAT) comparing minor *versus* major alleles (D *versus* d) in cases *versus* controls for the French dataset. The level of significance (corrected for multiple comparisons and for genomic inflation) is represented by the red line at $-\log_{10}P = 8.5$. (Y axis = $-\log_{10}P$, X axis = physical positions of the markers corresponding to the human February 2009 GRCh37/hg19 genome assembly, the chromosomes being plotted successively, shaded alternatively for better visualisation).

The region of interest encompassed by the associated markers encompasses the HLA-DQA1 gene as seen in Figure 61.

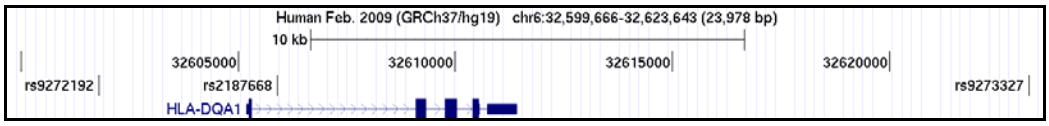


**Figure 61 Region of association, French replication dataset**
Shown are the three associated markers for the French dataset (rs9272192, rs9273327, rs2187668) centered by HLA-DQA1 in the extended MHC region on chromosome 6 (physical positions of the markers corresponds to the human February 2009 GRCh37/hg19 genome assembly).

## Haplotype analysis

### *Haploblock definition*

Haplotype blocks (Haploblocks) were computed in SNP & Variation Suite v7.3.1 according to the definition by Gabriel *et al.* [428]. As a measure of linkage disequilibrium defining each block D' was chosen at an upper confidence bound > 0.98 and a lower confidence bound > 0.7 (98% confidence). Pairs were

rejected with an upper bound < 0.9. The markers assessed in order to define blocks were chosen with a minor allele frequency of > 0.05. Blocks were limited to less than 30 markers per block and a length lower than 160 kb. Adhering to these criteria, 25,378 blocks including 59,807 markers were defined. Haplotypes were estimated according to the expectation-maximisation – EM – algorithm; the frequency threshold for haplotype estimation was set at 0.01, EM convergence tolerance was set at 0.0001, the maximum number of EM iterations was set at 50.

*Haplotype Association Test*

The Chi-squared test statistics was calculated per block and per haplotype, the significance threshold ($P$ = 0.001) corrected for multiple comparisons (corresponding to the number of blocks: 25,378) was set at $-\log_{10}P = 7.4$, see Figures 62 and 63.

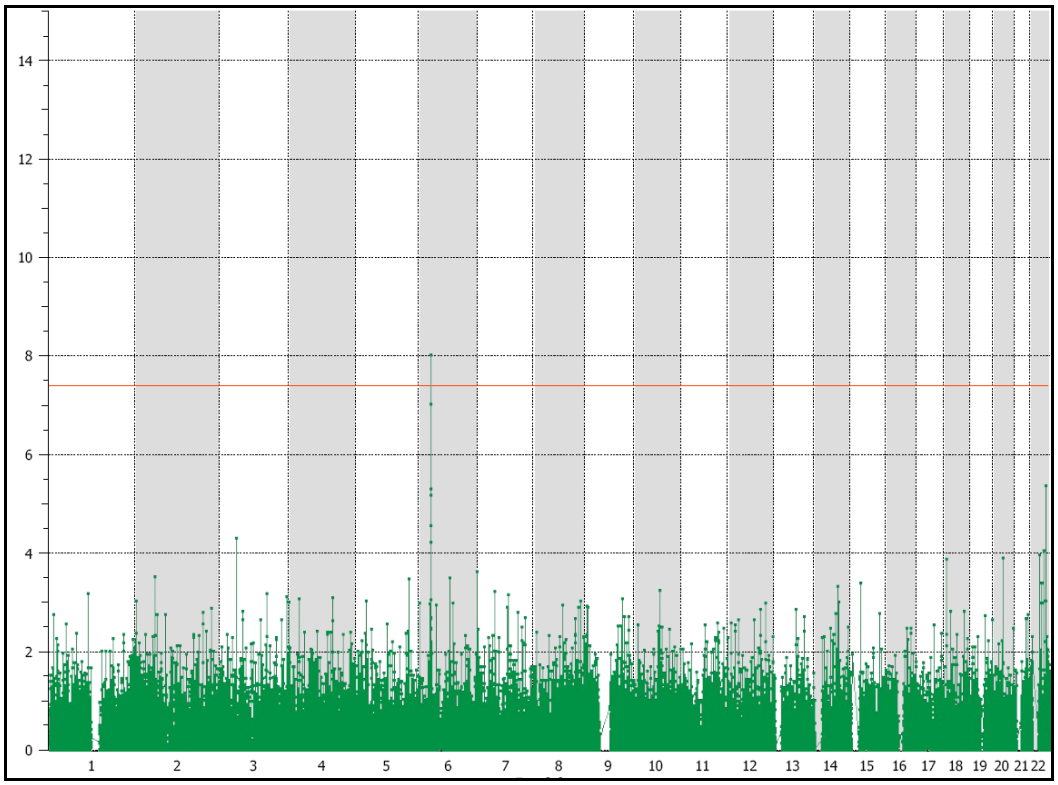


**Figure 62 HAT per block Manhattan plot, FR replication dataset**
Results of the haplotype association test (HAT) per block analysis. The level of significance (corrected for multiple comparisons) is represented by the red line at $-\log_{10} = 7.4$. (Y axis = -log10P, X axis = physical positions of the markers corresponding to the human February 2009 GRCh37/hg19 genome assembly, the chromosomes have been plotted successively, and shaded alternatively for better visualisation.

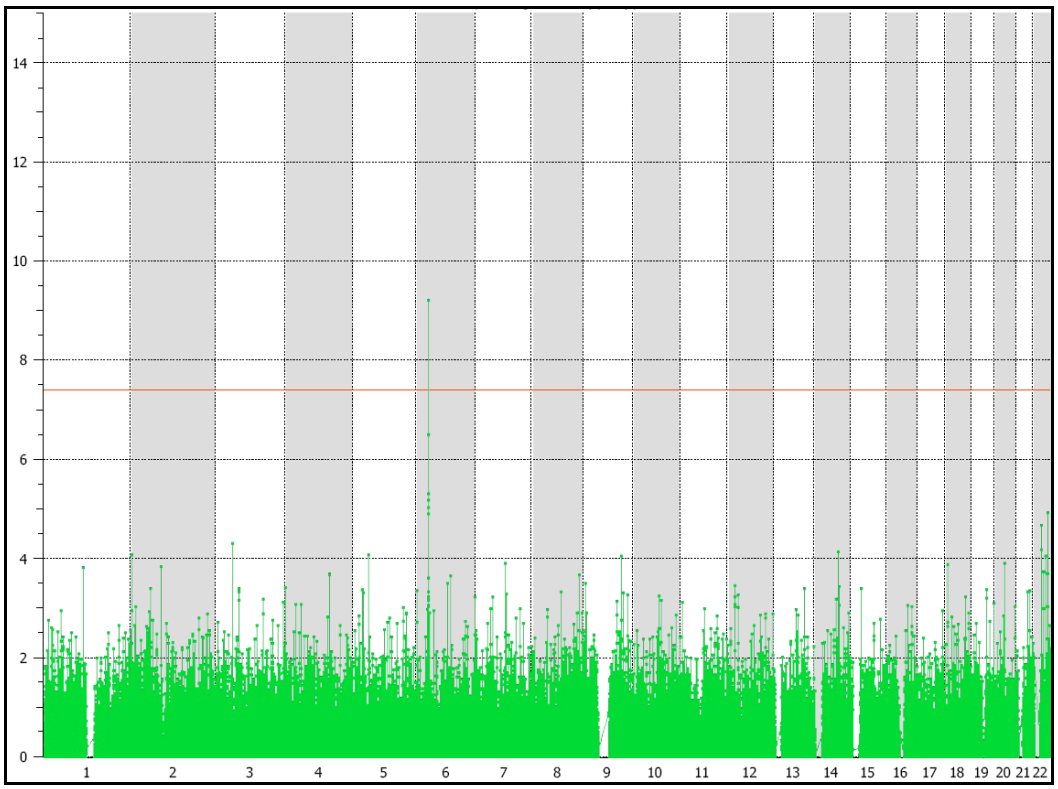


**Figure 63 HAT per haplotype Manhattan plot, FR replication dataset**
Results of the haplotype association test (HAT) per haplotype analysis. The level of significance (corrected for multiple comparisons) is represented by the red line at $-\log_{10} = 7.4$. (Y axis = -log10P, X axis = physical positions of the markers corresponding to the human February 2009 GRCh37/hg19 genome assembly, the chromosomes have been plotted successively, and shaded alternatively for better visualisation.

#### 3.2.2.2 Dutch

*Selection*

**Cases**

146 kidney biopsy confirmed Dutch cases (courtesy of Professor Jack Wetzels) were genotyped, of which 109 males and 37 females which results in a gender ratio of male : female ~ 3 / 1. No manoeuvres have been attempted to enrich the cohort for disease-predisposing alleles bearing individuals. The cases were genotyped on a Illumina Human 370 Quad v3 platform (UCL genotyping core; Dr. Mike Hubank). The same chip type was chosen for the genotyping of the Dutch cases (second replication study) as the one which was used for the initial, exploratory study (UK).

**Controls**

1832 controls from the Netherlands (906 males, 926 females; sex ratio ~ 1 / 1) genotyped for 370,404 SNPs were provided by the Dutch group. Dutch controls were genotyped by deCODE genetics, Iceland, using the HumanCNV370-Duo SNP chip. They are part of the Nijmegen Biomedical Study, a population-based survey conducted by the Department of Epidemiology and Biostatistics and the Department of Clinical Chemistry of the Radboud University Nijmegen Medical Centre [509]. This historical control dataset was rendered in PLINK [510] format. Reading the file in PLINK identified 30,038 heterozygous haploid genotypes, the minor allele frequency (MAF) was 0 for 23,981 SNPs, the total genotyping rate for the remaining SNPs was 0.93. No frequency or genotyping pruning was performed in PLINK. Reformatting of the binary file was performed in PLINK, separated textfiles were obtained formatted as in the PEDFILE linkage format:

```
FAM ID FATHER_ID MOTHER_ID SEX AFFECTION_STATUS MARKERS
```

and associated map file:

```
CHROMOSOME, SNP_ID, ?, PHYSICAL_POSITION
```

which were combined in a case matrix (similar to the case-control matrix shown in Table 4) to be analysed in the Helix SNP & Variation Suite v7.3.1.

*Quality control*

The same stringency quality control scenario as for the exploratory analysis and French replication has been applied for the Dutch dataset.

**Combined Separated Permissive Scenario**

Given that this is a setting of a historical control type dataset (genotyping performed at different timepoints for cases and for controls) we performed the first quality control steps before merging the two matrices and then added a second filtering step (as has been previously described for the UK dataset).

*Cases matrix*

373,397 SNPs (including CNV markers) were genotyped and successfully mapped with the Illumina 370k quad v3 map: 338,258 markers with two alleles, 6,824 markers with one allele (*ergo* noninformative), 28,288 markers with no alleles (all missing; of which 21,890 CNV markers). Quality assurance parameters for the genotype filtering by marker: call rate CR < 0.90; minor allele frequency (MAF) < 0.01; Hardy-Weinberg Equilibrium (HWE *P*) < 0.01.

44,837 (13,25 % of the informative markers 338,258 = 100%) markers did not pass the filtering threshold.

*Controls matrix*

370,404 SNPs including CNV markers were genotyped. Of which 340,836 markers are informative (biallelic), 5,587 markers are monoallelic, 23,981 markers are missing (of which 21,890 dedicated CNV type markers). The same quality assurance parameters as the ones for the cases were used (call rate < 90%; minor allele frequency (MAF) < 0.01; Hardy-Weinberg Equilibrium (HWE *P*) < 0.01).

20,465 (6 %; informative markers 340,836 = 100%) markers did not pass the filtering threshold.

*Together HWE < 0.01*

293,421 case matrix markers intersected with 320,371 control matrix markers yielded 283,512 common case control matrix markers). Applying the Hardy-

Weinberg Equilibrium filtering condition to the combined case / control matrix (HWE *P*) < 0.01 on the merged (case and controls) matrix led to 1,072 markers being inactivated.

*QQ plot*

The QQ plots comparing the Dutch dataset in two scenarios: with and omitting the SNPs within the extended HLA complex, showing the most extreme dispersion within the HLA region, Figure 64

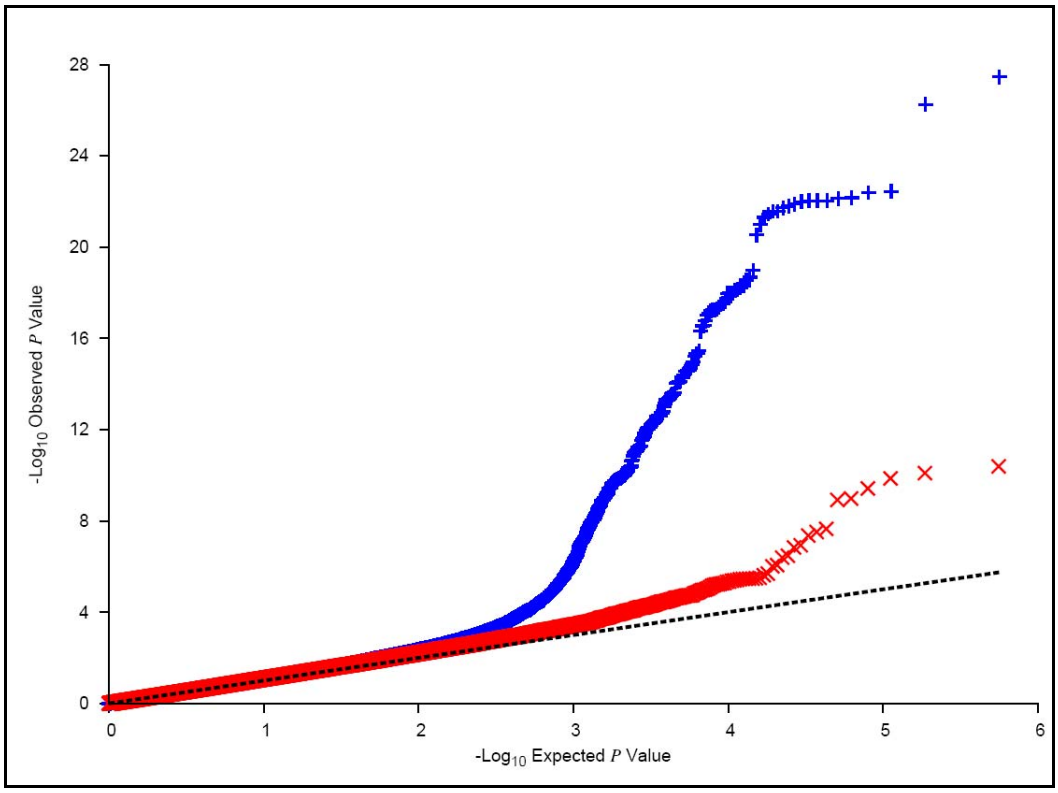


**Figure 64 QQ plots, NL replication, separated permissive scenario**
Quantile–quantile plot for observed (y axis) versus expected (x axis) -log10P values comparing the impact of the extended MHC region (blue: all SNPs tested; red: SNPs within the extended MHC region excluded; dotted black line: null distribution).

## *Basic Allelic Test*

All 282,440 of the informative markers that passed the quality control were used for the analysis with following parameters: the first genetic model assessed was the basic allelic test (D *versus* d); missing values were not used; input data was not corrected for stratification (the genomic inflation factor $\lambda = 1.12$ was close to 1 hence not suggestive for stratification); the statistic method used was the robust

basic allelic test (D *versus* d), the test statistic used being Chi-Squared with Bonferroni (multiple testing for 282,440 markers) correction and genomic inflation factor (λ = 1.12) adjustment ($P$ = 0.001 / 282,440 / 1.12 = 3.16 x$10^{-9}$; -$\log_{10}P$ = 8.5).

197 markers were found to be associated with the trait (IMN), see Figure 65.

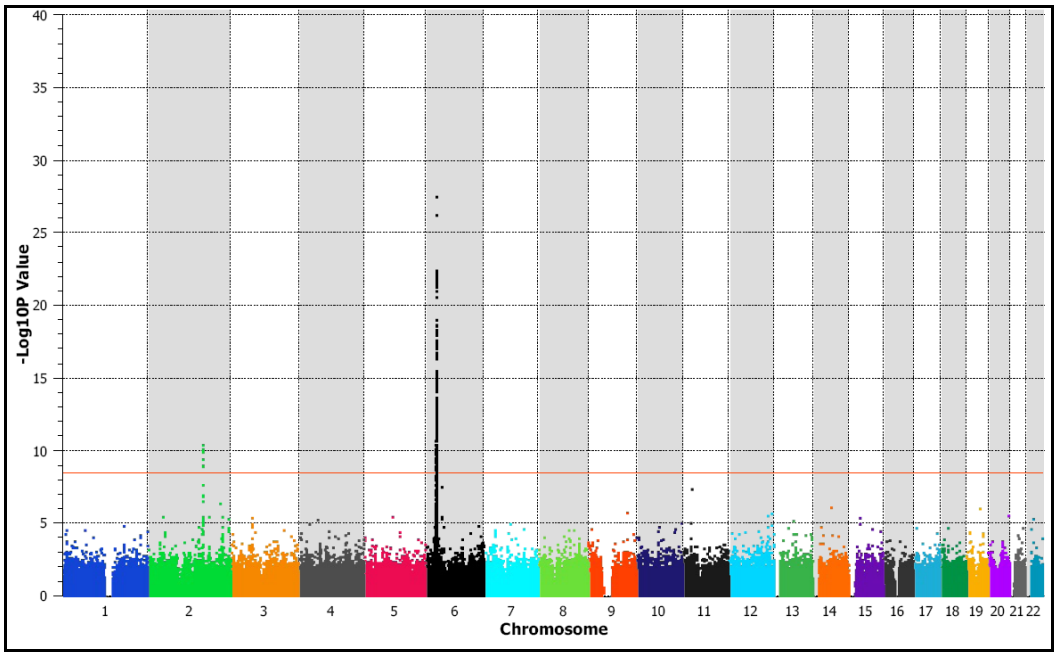


**Figure 65 BAT Manhattan plot, Dutch replication dataset**
Manhattan plot for the Chi-squared test of the basic allelic test (BAT) comparing minor *versus* major alleles (D *versus* d) in cases *versus* controls for the Dutch dataset. The level of significance (corrected for multiple comparisons and for genomic inflation) is represented by the red line at -$\log_{10}P$ = 8.5. (Y axis = -$\log_{10}P$, X axis = physical positions of the markers corresponding to the human February 2009 GRCh37/hg19 genome assembly, the chromosomes being plotted successively, shaded alternatively for better visualisation).

## *Haplotype analysis*

### Haploblock definition

Haplotype blocks (Haploblocks) were computed according to the definition by Gabriel et al. [426]. D' was set at an upper confidence bound > 0.98 and a lower confidence bound > 0.7 (98% confidence); pairs were rejected at an upper bound < 0.9. The minor allele frequency of assessed markers (less than 30 per block) was set at > 0.05. The maximum length of a block was set at 160 kb.

21,409 blocks including 56,807 markers were defined. Haplotypes were estimated according to the expectation-maximisation – EM – algorithm; the frequency threshold for haplotype estimation was set at 0.01, EM convergence tolerance was set at 0.0001, the maximum number of EM iterations was set at 50.

**Haplotype Association Test**

The Chi-squared test statistics was calculated per block and per haplotype, the significance threshold ($P$ = 0.001) was corrected for multiple comparisons (corresponding to the number of blocks: 21,409) was set at $-\log_{10}P = 7.33$, see Figures 66 and 67.

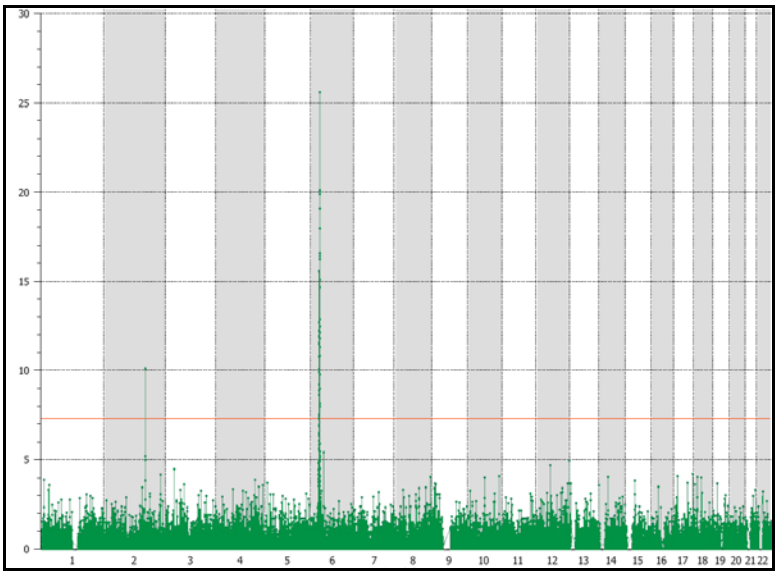


**Figure 66 HAT per block Manhattan plot, Dutch replication dataset**
Results of the haplotype association test (HAT) per block analysis. The level of significance (corrected for multiple comparisons) is represented by the red line at $-\log_{10} = 7.33$. (Y axis = $-\log_{10}P$, X axis = physical positions of the markers corresponding to the human February 2009 GRCh37/hg19 genome assembly, the chromosomes have been plotted successively, and shaded alternatively for better visualisation.

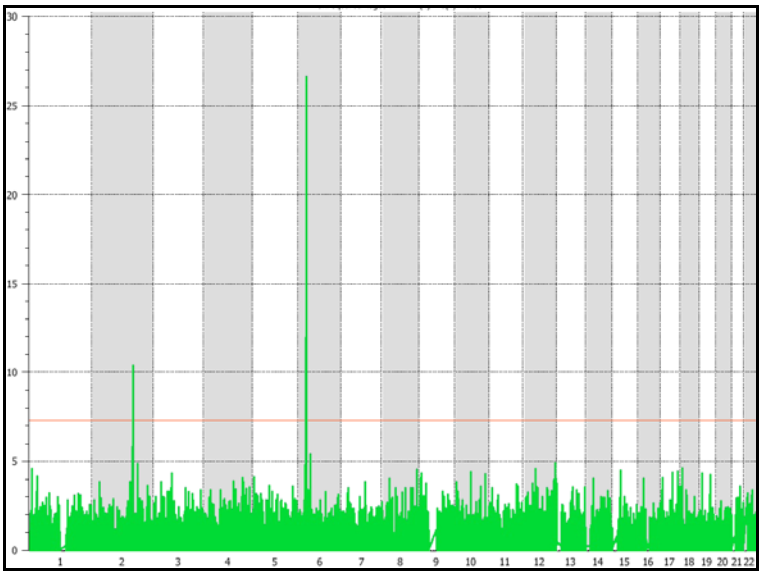


**Figure 67 HAT per haplotype Manhattan plot, Dutch replication dataset**
Results of the haplotype association test (HAT) per haplotype analysis. The level of significance (corrected for multiple comparisons) is represented by the red line at $-\log_{10} = 7.33$. (Y axis = $-\log_{10}P$, X axis = physical positions of the markers corresponding to the human February 2009 GRCh37/hg19 genome assembly, the chromosomes have been plotted successively, and shaded alternatively for better visualisation.

### 3.2.3 Joint GWAS

#### 3.2.3.1 Common matrix and QC

The common Caucasoid descent of all the cases and controls in the aforementioned studies (*i.e.* UK, French, Dutch) suggested the possibility of performing a joint analysis of all 556 cases and 2,338 controls. The marker matrices of UK cases, UK controls, Dutch cases and Dutch controls have been filtered separatedly, using the criteria (CR < 0.9, MAF < 0.01, HWE $P < 0.01$) of the separated permissive filtering scenario; the marker matrix of the French cases and controls has been filtered using the same criteria applied within the together permissive filtering scenario (having been genotyped together). The common matrix, obtained through the interesecting of all the filtered matrices, contains 244,860 common markers before the secondary (HWE $P < 0.01$) filtering step, which leads to further pruning of 2,036 markers, leading to a final (common, quality filtered) marker set of 242,824.

#### 3.2.3.2 Stratification

The genomic inflation factor (λ) was found to be 1.29.

#### 3.2.3.3 QQ Plots

QQ plots comparing the joint (UK, FR and NL) datasets, with and omitting the SNPs within the extended HLA complex, show the most extreme dispersion within the HLA region, Figure 68.

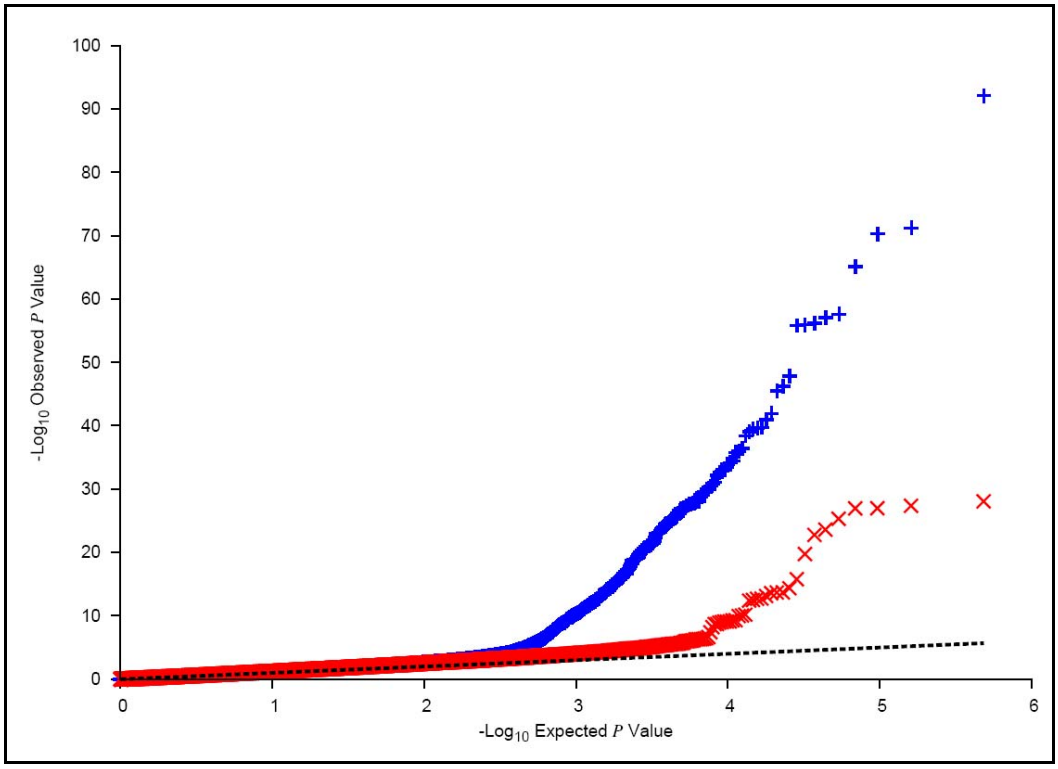


**Figure 68 QQ plots, joint analysis, separated permissive scenario**
Quantile–quantile plot for observed (y axis) versus expected (x axis) -log10P values comparing the impact of the extended MHC region (blue: all SNPs tested; red: SNPs within the extended MHC region excluded; dotted black line: null distribution).

### 3.2.3.4 Analysis

We performed the Basic Allelic Test, with the test statistic being Chi-squared, at a significance level corrected for multiple comparisons (242,824 markers) and genomic inflation factor ($\lambda$ = 1.29): $P$ = 0.001 / 242,824 / 1.29 = 3.19 x $10^{-9}$; $-\log_{10}P$ = 8.5.

### 3.2.3.5 Results

312 markers were found to be associated with the trait, the signal for association reaching significance level only at the two regions of interest previously described (*i.e.* the *PLA2R1* locus on chromosome 2 – 30 markers; and the extended HLA locus on chromosome 6 locus – 282 markers) see Figure 69.

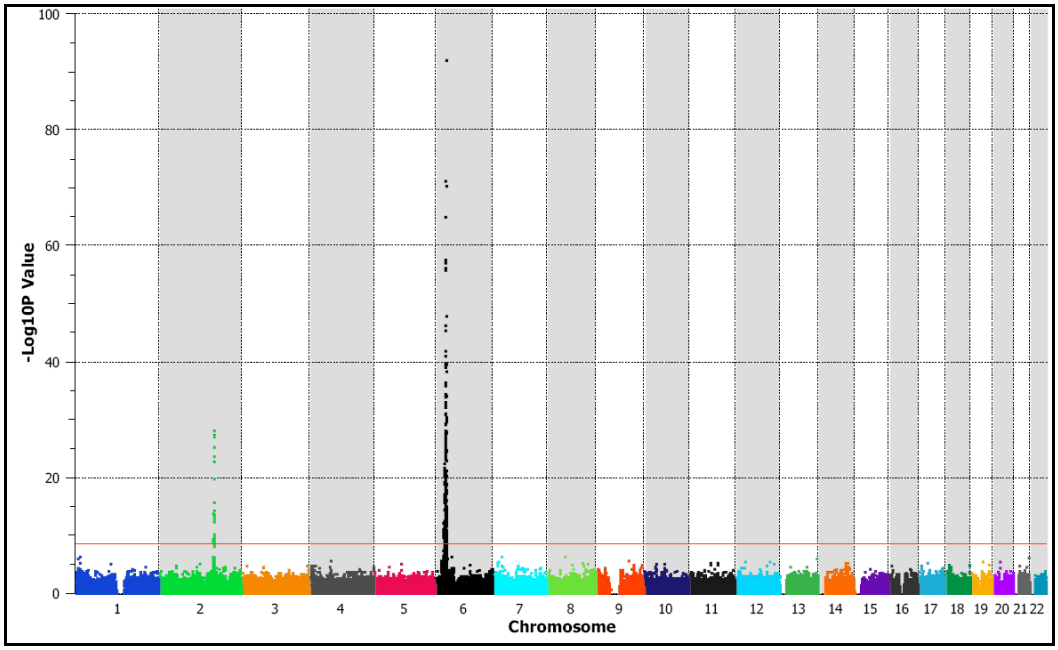


**Figure 69 BAT Manhattan plot, joint analysis**
Manhattan plot for the joint genome-wide association study (GWAS) of idiopathic membranous nephropathy (IMN). Joint analysis of genome-wide association study for 556 Caucasoid cases of IMN and 2338 Caucasoid controls using 242,824 SNPs. The horizontal red line indicates the genome-wide significance level for the basic allele test (upper plot) utilizing chi-square test after Bonferroni correction and adjustment for the genomic inflation factor ($-\log_{10}P < 8.5$). The x axis shows the physical position / chromosomes, the y axis shows the $-\log_{10}P$ significance levels. Note the highly significant associations with an allele on chromosome 6 (containing HLA-DQA1) and an allele on chromosome 2 (containing PLA2R1).

## 3.2.4 Comparison of the results

### 3.2.4.1 Chromosome 6

From the 282 significant associations between trait and marker for the joint analysis which are located within the extended HLA region on chromosome 6, 112 are positive association signals in the UK cohort, 68 in the Dutch cohort and only 1 in the French cohort, see Table 12.

**Table 12 Comparison of association signals on chromosome 6**
Association signals in the extended HLA region on chromosome 6, shown in red the ($-\log_{10}P$ values above the significance threshold ($-\log_{10}P > 8.5$); rs9267522 (at position 31711749) nonexistent in dbSNP 130 (being previously called rs11229) appeared only in dbSNP 131, hence no allele information for this marker in the table.

| Marker | Hg18 Position | dbSNP130 alleles | Locus Region Gene | F $-\log_{10}P$ | NL $-\log_{10}P$ | UK $-\log_{10}P$ | 3P $-\log_{10}P$ |
|---|---|---|---|---|---|---|---|
| rs7776065 | 25853721 | C/T | | 0.3706 | 3.9939 | 4.9934 | 10.283 |
| rs9885831 | 25858772 | A/C | | 0.4676 | 3.68 | 4.9653 | 9.7686 |
| rs1324087 | 25949387 | C/T | | 0.9333 | 2.0233 | 5.8477 | 11.04 |
| rs13198474 | 25982402 | A/G | SLC17A3 | 2.0635 | 3.7801 | 6.7085 | 12.114 |
| rs9467632 | 25997685 | A/G | | 1.8861 | 2.1353 | 6.2465 | 10.266 |
| rs7748167 | 26012631 | A/C | | 2.2469 | 2.3868 | 6.3758 | 11.113 |
| rs3734523 | 26033966 | C/T | SLC17A2 | 1.9161 | 2.4504 | 6.2955 | 10.515 |
| rs6940007 | 26039936 | A/C | | 1.9161 | 2.3604 | 6.2955 | 10.523 |
| rs3857546 | 26265741 | C/T | | 1.1344 | 3.4411 | 4.9231 | 12.222 |
| rs6923139 | 26421327 | C/T | | 1.7473 | 4.7525 | 6.099 | 16.554 |
| rs9379858 | 26475668 | C/T | BTN3A2 | 1.1275 | 2.7331 | 7.22 | 11.456 |
| rs9379859 | 26477528 | C/T | BTN3A2 | 1.1711 | 2.788 | 7.22 | 11.656 |
| rs12176317 | 26480765 | A/G | BTN3A2 | 1.1711 | 2.5377 | 7.0639 | 11.346 |
| rs3799380 | 26575161 | C/T | BTN2A1 | 0.4825 | 3.9338 | 7.3691 | 11.078 |
| rs2093169 | 26603078 | C/T | | 0.6153 | 3.3224 | 7.4443 | 9.7421 |
| rs9379897 | 26709505 | C/T | | 1.2094 | 3.8355 | 7.382 | 12.487 |
| rs7745603 | 27198383 | C/T | | 1.6885 | 3.4862 | 8.5532 | 10.306 |
| rs6913660 | 27199404 | A/C | | 1.6572 | 4.7186 | 7.0969 | 10.293 |
| rs13219354 | 27293643 | C/T | | 2.0207 | 5.5738 | 6.4362 | 12.54 |
| rs6938200 | 27339129 | A/G | | 1.3992 | 2.7063 | 8.753 | 9.2423 |
| rs7746199 | 27369303 | C/T | | 0.4855 | 3.447 | 4.006 | 10.826 |
| rs2076305 | 27483590 | G/T | | 0.6385 | 5.2314 | 7.8076 | 11.647 |
| rs6926142 | 27503743 | C/T | | 0.2664 | 4.0238 | 7.1763 | 8.6462 |
| rs1569566 | 27508010 | G/T | | 0.5688 | 4.694 | 7.0489 | 9.9575 |
| rs1883216 | 27533164 | A/C | ZNF184 | 0.5204 | 4.6696 | 6.8013 | 9.2255 |
| rs6904596 | 27599278 | A/G | | 1.5164 | 9.4505 | 8.9424 | 21.441 |
| rs10484399 | 27642507 | A/G | | 1.2639 | 9.9443 | 9.196 | 20.104 |
| rs10946940 | 27668566 | A/G | | 0.2839 | 4.3513 | 5.5708 | 8.9692 |
| rs9380010 | 27791551 | A/G | | 0.6188 | 3.2199 | 5.7046 | 10.209 |
| rs9295740 | 27797481 | A/G | | 1.4989 | 7.077 | 4.9644 | 15.729 |
| rs9468213 | 27814159 | A/G | | 0.3858 | 5.029 | 8.5532 | 10.193 |
| rs9283880 | 27823222 | A/C | | 0.6743 | 4.1942 | 7.3583 | 11.537 |
| rs7776351 | 27834710 | C/T | | 0.6177 | 4.2088 | 6.5877 | 11.565 |
| rs742047 | 27847359 | A/G | | 0.4446 | 5.1169 | 7.3739 | 11.182 |
| rs175597 | 27918605 | A/G | | 1.6826 | 8.0349 | 9.3774 | 20.98 |
| rs200997 | 27919794 | C/T | | 0.422 | 3.6758 | 6.3572 | 9.6978 |
| rs200991 | 27923473 | G/T | | 1.3031 | 4.218 | 7.8529 | 16.111 |
| rs13194781 | 27923618 | A/G | | 1.821 | 9.8488 | 9.065 | 20.757 |
| rs200951 | 27943909 | A/G | | 0.2098 | 2.9651 | 6.9215 | 9.7883 |
| rs200953 | 27945246 | C/T | | 1.9161 | 8.0349 | 9.0682 | 21.287 |
| rs200969 | 27967432 | A/G | | 0.3524 | 4.0121 | 6.8267 | 11.317 |
| rs203890 | 28130227 | C/T | | 0.4208 | 5.0106 | 7.2635 | 12.016 |
| rs9468276 | 28135667 | A/C | | 0.4208 | 5.8547 | 6.8496 | 12.243 |
| rs1233708 | 28281198 | A/G | | 0.1754 | 3.7221 | 8.6852 | 8.9672 |
| rs9380069 | 28311279 | A/G | | 0.7612 | 6.9437 | 10.243 | 17.073 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| rs9295768 | 28317081 | A/G | | 0.6221 | 5.6408 | 9.5096 | 14.494 |
| rs967005 | 28318667 | A/G | | 0.6221 | 5.8677 | 9.5096 | 14.564 |
| rs1419183 | 28350773 | G/T | ZNF187 | 0.8292 | 7.1478 | 9.5105 | 16.603 |
| rs13211507 | 28365356 | C/T | PGBD1 | 1.822 | 9.2152 | 8.8863 | 19.259 |
| rs853676 | 28407666 | A/G | ZNF323 | 0.4677 | 8.2542 | 8.626 | 16.612 |
| rs213228 | 28439231 | G/T | ZKSCAN3 | 0.2768 | 3.569 | 3.3782 | 8.9879 |
| rs13194504 | 28738670 | A/G | | 2.004 | 9.4365 | 9.1157 | 21.484 |
| rs6456834 | 28808331 | G/T | | 1.3115 | 6.2656 | 4.4396 | 16.555 |
| rs1233579 | 28820642 | A/G | | 2.296 | 10.088 | 8.8242 | 22.371 |
| rs1233585 | 28824324 | A/G | | 0.95 | 3.0704 | 4.7398 | 8.9699 |
| rs4324798 | 28884096 | A/G | | 2.004 | 7.247 | 8.4233 | 18.947 |
| rs209181 | 28900456 | C/T | | 0.8524 | 3.7326 | 3.4978 | 8.623 |
| rs3131093 | 28945416 | C/T | | 2.004 | 9.4873 | 8.5373 | 20.396 |
| rs3129791 | 29062272 | A/G | | 2.004 | 9.9477 | 8.5954 | 21.074 |
| rs3130893 | 29088686 | A/G | | 2.004 | 9.7775 | 8.7654 | 21.049 |
| rs3117143 | 29139121 | A/C | | 1.9772 | 10.367 | 10.042 | 21.784 |
| rs3130817 | 29287216 | G/T | | 1.0294 | 5.6655 | 7.0367 | 12.032 |
| rs3130827 | 29338662 | C/T | LOC651503 | 0.7832 | 2.541 | 6.1895 | 10.387 |
| rs1884123 | 29364399 | C/T | | 0.9088 | 2.5901 | 6.4598 | 10.793 |
| rs3117426 | 29379991 | C/T | | 1.2527 | 5.7202 | 7.8585 | 11.693 |
| rs6934993 | 29445479 | A/G | | 1.4695 | 3.4696 | 5.9243 | 11.337 |
| rs3749971 | 29450754 | C/T | OR12D3 | 2.3641 | 7.9768 | 8.6785 | 19.581 |
| rs2746150 | 29550680 | C/T | | 2.2818 | 8.8813 | 8.5081 | 20.363 |
| rs1235162 | 29645203 | C/T | | 2.6527 | 10.641 | 8.5738 | 23.434 |
| rs3129073 | 29723801 | C/T | | 2.5969 | 9.1772 | 4.626 | 14.267 |
| rs2535238 | 29753017 | G/T | ZFP57 | 2.0873 | 8.6813 | 2.8818 | 12.684 |
| rs2747430 | 29756485 | A/G | ZFP57 | 1.7247 | 5.3952 | 1.8882 | 9.6963 |
| rs2747457 | 29764396 | G/T | | 1.5592 | 7.6506 | 2.8945 | 11.751 |
| rs3131888 | 29771868 | A/G | | 0.7598 | 6.3377 | 0.334 | 8.9165 |
| rs3129055 | 29778240 | C/T | | 1.2893 | 7.1085 | 2.7555 | 10.681 |
| rs1610601 | 29808162 | G/T | AX747705 | 1.7486 | 7.2935 | 4.4511 | 12.631 |
| rs1633021 | 29854848 | A/G | | 1.5592 | 4.715 | 4.6694 | 11.683 |
| rs1736971 | 29884301 | A/C | | 1.3053 | 4.415 | 1.7644 | 9.6238 |
| rs1611133 | 29917361 | C/T | | 1.7141 | 5.9956 | 1.6283 | 12.938 |
| rs2734986 | 29926547 | A/G | | 2.2009 | 7.6196 | 4.8836 | 17.449 |
| rs2734985 | 29926641 | A/G | | 1.6006 | 4.9737 | 3.7588 | 14.965 |
| rs2517861 | 29929961 | A/G | | 1.3109 | 4.8343 | 3.1529 | 14.703 |
| rs1611710 | 29936895 | C/T | | 1.0459 | 3.1167 | 2.4418 | 8.6783 |
| rs2524005 | 30007656 | C/T | AK097625 | 1.4196 | 6.6557 | 2.8877 | 14.254 |
| rs3132685 | 30053928 | C/T | HCG9 | 2.7478 | 9.8333 | 9.5771 | 23.906 |
| rs259940 | 30119913 | C/T | | 0.8462 | 5.9664 | 4.719 | 9.709 |
| rs259919 | 30133482 | C/T | C6orf12 | 0.3783 | 6.0351 | 5.9001 | 10.775 |
| rs8321 | 30140501 | G/T | ZNRD1 | 2.1589 | 10.664 | 10.406 | 26.152 |
| rs1245371 | 30146331 | C/T | RNF39 | 0.9708 | 6.4095 | 4.7543 | 9.9776 |
| rs9261290 | 30146626 | C/T | RNF39 | 2.4238 | 11.587 | 10.243 | 27.076 |
| rs1264702 | 30173554 | A/G | | 1.6289 | 7.3517 | 1.7322 | 13.478 |
| rs916570 | 30174010 | C/T | | 1.6023 | 6.4868 | 1.4364 | 13.418 |
| rs1264695 | 30175871 | C/T | | 1.2708 | 6.4081 | 1.492 | 12.152 |
| rs1116221 | 30179309 | A/G | TRIM31 | 1.2797 | 7.0751 | 3.015 | 11.512 |
| rs2523989 | 30186254 | A/G | TRIM31 | 0.4056 | 7.3749 | 8.4011 | 19.083 |
| rs2523987 | 30187972 | G/T | TRIM31 | 0.6944 | 8.583 | 8.1736 | 20.823 |
| rs2517598 | 30188253 | C/T | TRIM31 | 0.4011 | 6.2418 | 9.2327 | 18.695 |
| rs2517592 | 30201116 | A/G | | 0.4266 | 7.6361 | 4.1248 | 15.369 |
| rs3129690 | 30304834 | C/T | | 1.3322 | 2.0916 | 7.5359 | 10.476 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| rs2844773 | 30315474 | G/T | | 3.0002 | 11.872 | 10.359 | 25.261 |
| rs1362089 | 30317781 | A/C | | 1.4428 | 2.5293 | 6.5166 | 9.3141 |
| rs3094073 | 30339203 | C/T | FLJ45422 | 2.9746 | 14.567 | 11.053 | 27.909 |
| rs3130380 | 30387109 | A/G | AK055657,CR624833 | 1.4913 | 12.742 | 11.281 | 29.155 |
| rs3094626 | 30431602 | A/G | | 0.5439 | 4.8971 | 6.196 | 13.406 |
| rs3130350 | 30435818 | G/T | | 1.0115 | 12.615 | 11.124 | 28.071 |
| rs1012411 | 30440534 | A/C | | 0.0981 | 4.4121 | 4.4912 | 12.03 |
| rs3094054 | 30441484 | A/C | | 0.6587 | 10.981 | 12.258 | 27.024 |
| rs2844745 | 30451682 | A/G | | 0.0204 | 2.5736 | 3.6733 | 8.9233 |
| rs3132622 | 30456653 | A/G | | 0.452 | 3.8623 | 5.6184 | 12.36 |
| rs3094694 | 30559883 | A/G | | 0.383 | 6.5398 | 6.5614 | 16.42 |
| rs3131115 | 30576770 | C/T | | 0.9388 | 5.6115 | 8.9484 | 15.235 |
| rs3132610 | 30652380 | C/T | ABCF1 | 0.7689 | 13.338 | 11.58 | 29.383 |
| rs1140809 | 30719655 | A/C | C6orf134 | 0.8696 | 5.3036 | 1.707 | 14.216 |
| rs1076828 | 30736231 | C/T | DHX16 | 0.0529 | 7.1707 | 2.4949 | 11.95 |
| rs9262143 | 30760760 | C/T | KIAA1949 | 1.0532 | 14.552 | 12.797 | 33.027 |
| rs1075496 | 30766218 | G/T | NRM | 0.3974 | 6.049 | 2.2449 | 13.233 |
| rs3094127 | 30805426 | C/T | FLOT1 | 1.3505 | 8.9215 | 9.3711 | 27.42 |
| rs2535319 | 30822458 | A/G | | 0.407 | 2.7068 | 7.2064 | 10.706 |
| rs3129975 | 30840033 | C/T | | 0.3796 | 6.8938 | 8.2598 | 18.833 |
| rs886424 | 30889981 | A/G | hCG_2038200 | 0.5642 | 13.026 | 15.259 | 35.995 |
| rs1264344 | 30908556 | A/G | | 0.9791 | 2.7504 | 2.5619 | 9.4865 |
| rs2844659 | 30932511 | A/G | | 1.0842 | 17.606 | 10.802 | 32.588 |
| rs2844657 | 30937501 | C/T | | 1.259 | 16.564 | 10.802 | 32.212 |
| rs1264323 | 30963886 | C/T | | 1.7171 | 8.6893 | 4.6079 | 16.185 |
| rs1049623 | 30972808 | A/G | DDR1 | 1.7171 | 8.6627 | 4.5206 | 16.095 |
| rs753725 | 30998850 | A/G | VARS2 | 0.2406 | 5.7645 | 2.8766 | 11.402 |
| rs4678 | 31001920 | C/T | VARS2 | 2.3385 | 15.22 | 12.763 | 33.948 |
| rs2532934 | 31002738 | C/T | | 0.2081 | 5.7715 | 2.7942 | 11.335 |
| rs3132571 | 31013292 | C/T | | 1.0132 | 9.599 | 5.6112 | 23.094 |
| rs3132580 | 31028103 | C/T | DPCR1 | 1.7086 | 11.662 | 17.719 | 39.534 |
| rs2240804 | 31028869 | C/T | DPCR1 | 0.8594 | 4.7147 | 0.9508 | 9.759 |
| rs3095089 | 31041773 | A/C | | 0.9891 | 12.571 | 12.043 | 30.878 |
| rs886403 | 31065597 | C/T | MUC21 | 0.3317 | 8.0382 | 7.9733 | 21.166 |
| rs1634718 | 31080844 | A/G | | 0.6832 | 9.0402 | 10.166 | 24.595 |
| rs2523898 | 31101512 | C/T | | 0.1139 | 5.6968 | 6.2234 | 10.324 |
| rs2844665 | 31114834 | A/G | | 0.6522 | 3.2197 | 4.6524 | 14.846 |
| rs2517552 | 31115569 | A/G | | 0.528 | 3.1448 | 4.9821 | 15.496 |
| rs2523864 | 31126525 | A/G | | 0.374 | 5.1812 | 10.047 | 24.994 |
| rs3130955 | 31162490 | A/C | | 1.0115 | 4.4701 | 5.0403 | 13.248 |
| rs3130544 | 31166319 | A/C | | 3.504 | 16.318 | 22.513 | 57.063 |
| rs2517448 | 31170646 | A/G | | 0.4536 | 3.3357 | 5.4021 | 15.913 |
| rs6457327 | 31182009 | A/C | | 0.5491 | 3.1726 | 5.2726 | 15.599 |
| rs1064191 | 31183354 | A/G | | 0.515 | 4.4337 | 8.6524 | 19.11 |
| rs2844635 | 31183460 | C/T | | 1.0668 | 8.3229 | 9.3518 | 24.907 |
| rs2233956 | 31189184 | A/G | | 3.4641 | 13.456 | 16.647 | 46.232 |
| rs3130981 | 31191792 | C/T | CDSN,PSORS1C1 | 0.3401 | 2.0781 | 4.1963 | 10.605 |
| rs3094212 | 31193749 | C/T | CDSN,PSORS1C1 | 0.4131 | 3.0885 | 5.3738 | 8.6289 |
| rs3095314 | 31197610 | A/G | PSORS1C1 | 0.5802 | 4.357 | 12.082 | 23.965 |
| rs3130564 | 31209653 | C/T | PSORS1C1 | 1.5942 | 10.403 | 17.512 | 41.953 |
| rs1265099 | 31213392 | C/T | PSORS1C2,PSORS1C1 | 1.4859 | 5.9567 | 9.2408 | 14.664 |
| rs3094663 | 31215066 | A/G | PSORS1C2,PSORS1C1 | 0.8917 | 0.9322 | 4.9394 | 10.559 |
| rs1265086 | 31217861 | A/C | | 0.9046 | 4.0696 | 5.2211 | 18.145 |
| rs746647 | 31222161 | C/T | CCHCR1 | 1.4275 | 8.1528 | 7.0604 | 23.187 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| rs2240064 | 31222552 | C/T | CCHCR1 | 1.0578 | 2.4834 | 5.4072 | 11.226 |
| rs1265112 | 31225998 | A/G | CCHCR1 | 1.4275 | 8.0999 | 7.0604 | 23.059 |
| rs3130453 | 31232828 | C/T | CCHCR1 | 0.9351 | 2.742 | 5.6616 | 13.651 |
| rs2239524 | 31233548 | G/T | | 0.8358 | 2.6043 | 3.9323 | 10.59 |
| rs720465 | 31233756 | G/T | | 0.9529 | 5.3099 | 9.3371 | 25.384 |
| rs7750641 | 31237289 | C/T | TCF19 | 3.5133 | 17.534 | 21.421 | 57.586 |
| rs1419881 | 31238572 | C/T | TCF19 | 1.098 | 2.2446 | 7.4291 | 14.641 |
| rs3130931 | 31242867 | A/G | POU5F1 | 0.5196 | 1.031 | 6.097 | 11.701 |
| rs1265159 | 31248026 | C/T | | 2.2573 | 9.9806 | 12.208 | 36.416 |
| rs3094188 | 31250224 | G/T | | 0.5955 | 3.3801 | 4.8408 | 14.13 |
| rs3868542 | 31253818 | A/G | | 0.5434 | 2.9514 | 3.6226 | 11.048 |
| rs887464 | 31253899 | A/G | | 1.4875 | 3.7839 | 8.7767 | 13.201 |
| rs3869109 | 31292175 | C/T | | 0.787 | 3.2466 | 5.1618 | 12.105 |
| rs3130685 | 31314185 | C/T | | 0.7707 | 3.4918 | 4.5248 | 13.468 |
| rs3095250 | 31316319 | A/G | | 0.6138 | 2.7772 | 5.1403 | 9.4846 |
| rs3130712 | 31317489 | C/T | | 0.5685 | 2.9138 | 5.6381 | 9.9645 |
| rs3130696 | 31351863 | A/G | | 1.1949 | 5.486 | 15.376 | 31.095 |
| rs2243868 | 31369255 | A/G | | 1.6376 | 4.6862 | 4.9776 | 14.198 |
| rs2524089 | 31374501 | A/C | | 1.7959 | 6.8793 | 4.8714 | 16.075 |
| rs9366778 | 31377152 | A/G | | 0.4704 | 3.3277 | 2.9961 | 10.15 |
| rs3873386 | 31381724 | A/G | | 1.0797 | 2.6775 | 2.0578 | 10.343 |
| rs3134792 | 31420305 | A/C | | 4.4733 | 27.465 | 26.17 | 71.195 |
| rs2156875 | 31425326 | A/G | | 0.5327 | 5.4475 | 12.004 | 14.098 |
| rs2596501 | 31429190 | A/G | | 1.4368 | 3.1407 | 9.2971 | 10.541 |
| rs2523608 | 31430538 | C/T | | 0.1993 | 3.7633 | 4.8535 | 13.025 |
| rs2523589 | 31435313 | A/C | | 2.4315 | 5.2704 | 6.059 | 15.301 |
| rs2523554 | 31439808 | A/G | | 2.6823 | 4.6519 | 9.1766 | 17.232 |
| rs2844575 | 31442924 | A/G | | 1.0904 | 4.8986 | 5.7741 | 12.05 |
| rs2844511 | 31497763 | C/T | | 1.7987 | 3.4477 | 10.442 | 13.731 |
| rs2596464 | 31520940 | A/G | | 1.4882 | 6.5352 | 7.9235 | 19.004 |
| rs3128982 | 31525170 | A/G | | 1.4131 | 7.1607 | 11.792 | 27.706 |
| rs2284178 | 31540104 | C/T | HCP5 | 2.6928 | 3.1855 | 4.5907 | 10.354 |
| rs2523674 | 31544768 | C/T | | 1.0365 | 6.101 | 6.1973 | 17.723 |
| rs2516424 | 31556294 | C/T | | 1.3953 | 4.8866 | 7.4155 | 15.825 |
| rs2516400 | 31589084 | C/T | | 1.5617 | 5.4749 | 10.874 | 18.912 |
| rs2516398 | 31589505 | A/C | | 1.38 | 2.8311 | 4.6629 | 8.5409 |
| rs2844494 | 31591394 | G/T | | 1.4769 | 3.6301 | 4.4251 | 9.4993 |
| rs3132454 | 31597623 | C/T | | 1.4632 | 2.2161 | 3.7258 | 9.7831 |
| rs2071591 | 31623778 | C/T | NFKBIL1 | 2.2342 | 7.0109 | 13.616 | 25.656 |
| rs1041981 | 31648763 | A/C | LTA | 2.2342 | 6.4028 | 14.469 | 25.52 |
| rs2857595 | 31676448 | C/T | | 2.3329 | 10.326 | 17.973 | 34.41 |
| rs2260000 | 31701455 | C/T | BAT2 | 1.1625 | 3.1182 | 4.3287 | 8.9515 |
| rs2242660 | 31705732 | C/T | BAT2 | 2.6802 | 8.9307 | 11.715 | 24.484 |
| rs3115663 | 31709822 | A/G | BAT2 | 5.0536 | 17.968 | 24.518 | 56.169 |
| rs1046089 | 31710946 | A/G | BAT2 | 2.7508 | 9.2599 | 13.677 | 27.589 |
| rs2261033 | 31711570 | C/T | BAT2 | 1.0578 | 6.4914 | 7.5908 | 18.409 |
| rs9267522 | 31711749 | -/- | BAT2 | 5.0536 | 17.992 | 24.34 | 55.959 |
| rs1077393 | 31718508 | C/T | BAT3 | 1.5347 | 4.5059 | 7.9876 | 13.883 |
| rs805262 | 31736712 | A/G | | 0.3943 | 2.8445 | 5.4315 | 9.2152 |
| rs3130618 | 31740113 | A/C | BAT4 | 5.0536 | 17.992 | 24.276 | 55.808 |
| rs707928 | 31850569 | C/T | C6orf27 | 1.8107 | 10.879 | 11.731 | 26.601 |
| rs2763979 | 31902571 | C/T | | 2.5165 | 8.5065 | 9.7159 | 24.101 |
| rs660550 | 31945256 | A/C | SLC44A4 | 0.7255 | 4.6255 | 6.8635 | 11.063 |
| rs644827 | 31946420 | C/T | SLC44A4 | 0.7255 | 4.5352 | 6.8635 | 11.135 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| rs494620 | 31946692 | A/G | SLC44A4 | 0.6362 | 4.4143 | 6.4153 | 10.276 |
| rs2242665 | 31947288 | A/G | SLC44A4 | 0.7255 | 4.9364 | 6.8635 | 11.38 |
| rs652888 | 31959213 | C/T | EHMT2 | 3.8863 | 12.284 | 16.5 | 45.483 |
| rs535586 | 31968316 | A/G | EHMT2 | 0.8114 | 4.8041 | 8.6052 | 16.631 |
| rs659445 | 31972283 | C/T | EHMT2 | 0.747 | 5.6782 | 8.6052 | 17.224 |
| rs2734335 | 32001923 | C/T | | 1.1059 | 3.36 | 6.9536 | 10.085 |
| rs2072633 | 32027557 | C/T | CFB | 0.2038 | 4.4137 | 5.9927 | 10.866 |
| rs630379 | 32030233 | G/T | RDBP | 1.5147 | 5.6013 | 11.938 | 20.72 |
| rs437179 | 32036993 | G/T | SKIV2L | 1.63 | 5.194 | 11.773 | 20.094 |
| rs389884 | 32048876 | C/T | | 5.5692 | 21.982 | 23.891 | 65.09 |
| rs389883 | 32055439 | A/C | | 1.7802 | 5.0261 | 11.773 | 20.626 |
| rs185819 | 32158045 | C/T | | 1.0249 | 3.4547 | 7.9711 | 12.097 |
| rs204999 | 32217957 | A/G | | 1.7823 | 12.172 | 12.739 | 39.045 |
| rs3096697 | 32242488 | C/T | EGFL8 | 3.0569 | 14.3 | 14.194 | 40.948 |
| rs2269423 | 32253685 | G/T | AGPAT1 | 1.4127 | 3.4381 | 4.9105 | 14.677 |
| rs3134945 | 32254470 | G/T | RNF5 | 2.6429 | 13.548 | 13.821 | 39.72 |
| rs204994 | 32262976 | A/G | PBX2 | 1.2401 | 12.064 | 13.985 | 35.807 |
| rs204993 | 32263559 | C/T | PBX2 | 0.8566 | 10.332 | 12.082 | 32.81 |
| rs2071277 | 32279661 | A/G | NOTCH4 | 2.7735 | 4.3283 | 5.2274 | 11.634 |
| rs394657 | 32295001 | A/G | NOTCH4 | 2.4139 | 5.9284 | 6.3815 | 16.084 |
| rs422951 | 32296361 | A/G | NOTCH4 | 1.5146 | 4.2236 | 4.519 | 10.539 |
| rs443198 | 32298384 | C/T | NOTCH4 | 1.1377 | 8.0363 | 5.6897 | 24.657 |
| rs9267873 | 32307330 | C/T | | 0.9679 | 7.664 | 2.5875 | 14.652 |
| rs411326 | 32319295 | A/G | | 1.3592 | 1.367 | 6.281 | 8.633 |
| rs3130320 | 32331236 | C/T | | 0.565 | 2.9526 | 6.4581 | 11.931 |
| rs3115553 | 32353805 | A/G | | 1.8352 | 9.7634 | 11.599 | 29.601 |
| rs926070 | 32365544 | C/T | | 2.7058 | 4.0703 | 6.7694 | 14.162 |
| rs6935269 | 32368328 | C/T | | 1.8352 | 9.7806 | 11.599 | 29.644 |
| rs7775397 | 32369230 | G/T | C6orf10 | 6.025 | 22.393 | 26.79 | 70.277 |
| rs6457536 | 32381743 | A/G | C6orf10 | 1.8549 | 9.9079 | 13.021 | 29.957 |
| rs2076537 | 32425613 | C/T | C6orf10 | 1.5671 | 4.959 | 8.8789 | 14.953 |
| rs2395150 | 32434023 | A/G | C6orf10 | 1.6601 | 3.6357 | 7.1133 | 13.522 |
| rs3129939 | 32444744 | A/G | C6orf10 | 2.2009 | 11.833 | 16.638 | 39.657 |
| rs3129943 | 32446673 | A/G | C6orf10 | 2.3347 | 7.3385 | 14.339 | 30.457 |
| rs2050190 | 32447054 | C/T | C6orf10 | 1.5976 | 8.5241 | 5.5981 | 18.903 |
| rs2050189 | 32447625 | A/G | C6orf10 | 1.4989 | 9.6428 | 16.898 | 34.181 |
| rs4424066 | 32462406 | A/G | | 1.6455 | 3.8071 | 4.6858 | 9.4172 |
| rs3817973 | 32469089 | A/G | | 1.6601 | 4.0874 | 4.7622 | 9.7493 |
| rs1980493 | 32471193 | A/G | BTNL2 | 3.3075 | 14.707 | 23.944 | 47.85 |
| rs2076530 | 32471794 | A/G | BTNL2 | 1.5732 | 4.5123 | 5.7953 | 10.985 |
| rs6932542 | 32488240 | A/G | | 2.0615 | 3.7305 | 10.179 | 13.7 |
| rs3135338 | 32509195 | A/G | | 2.3627 | 2.6091 | 9.9232 | 13.303 |
| rs2395174 | 32512856 | G/T | | 1.8288 | 5.2422 | 10.788 | 17.29 |
| rs2395182 | 32521295 | G/T | | 0.8205 | 3.5825 | 5.6435 | 9.6605 |
| rs3129890 | 32522251 | C/T | | 2.4025 | 9.842 | 9.7512 | 24.636 |
| rs9268832 | 32535767 | C/T | | 0.4327 | 3.3204 | 5.5041 | 8.768 |
| rs2187668 | 32713862 | A/G | HLA-DQA1 | 8.7471 | 26.252 | 35.286 | 92.097 |
| rs9275141 | 32759095 | G/T | | 0.9307 | 3.4623 | 9.2285 | 9.7872 |
| rs6457617 | 32771829 | C/T | | 0.2934 | 4.2025 | 8.74 | 12.25 |
| rs2647012 | 32772436 | A/G | | 1.411 | 5.353 | 13.758 | 18.764 |
| rs9275390 | 32777134 | C/T | | 0.5038 | 4.5102 | 3.8592 | 10.771 |
| rs9275572 | 32786977 | A/G | | 1.4146 | 5.7745 | 11.383 | 16.035 |
| rs2858331 | 32789255 | C/T | | 1.1781 | 1.5938 | 9.5597 | 9.5227 |
| rs12177980 | 32794062 | A/G | | 0.1538 | 2.6311 | 5.6477 | 8.7223 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| rs9461799 | 32797507 | C/T | | 0.1538 | 2.7697 | 5.6477 | 8.9401 |
| rs2301271 | 32833171 | C/T | HLA-DQB2 | 0.7209 | 1.6788 | 7.0465 | 8.8639 |
| rs7453920 | 32837990 | A/G | HLA-DQB2 | 0.7209 | 1.7135 | 7.0465 | 8.8536 |
| rs2051549 | 32838064 | C/T | HLA-DQB2 | 0.7604 | 2.0292 | 7.0717 | 9.2192 |
| rs2071550 | 32838918 | G/T | HLA-DQB2 | 0.1507 | 2.3812 | 5.0638 | 10.065 |
| rs9368741 | 32845485 | A/G | | 0.1507 | 2.3013 | 5.0638 | 10.023 |
| rs9276644 | 32853021 | C/T | | 2.2133 | 6.2554 | 12.072 | 22.966 |
| rs719654 | 32860117 | C/T | | 1.6019 | 1.289 | 6.6841 | 8.8607 |
| rs7758736 | 32866372 | A/G | | 2.6799 | 8.7077 | 13.688 | 27.832 |
| rs3948793 | 32867426 | C/T | | 1.9549 | 5.9739 | 11.229 | 21.162 |
| rs1383261 | 32873429 | A/G | | 2.5831 | 3.8469 | 10.753 | 14.868 |
| rs11244 | 32888702 | C/T | HLA-DOB | 3.4825 | 4.8119 | 10.694 | 17.877 |
| rs2856997 | 32889754 | G/T | HLA-DOB | 1.218 | 3.7591 | 7.8105 | 20.259 |
| rs2071474 | 32890560 | A/G | HLA-DOB | 1.6536 | 1.4819 | 7.5475 | 12.053 |
| rs7383287 | 32891064 | A/G | HLA-DOB | 1.9926 | 6.5058 | 12.136 | 27.816 |
| rs241424 | 32912912 | C/T | TAP2 | 2.8887 | 1.6293 | 6.727 | 10.024 |
| rs6924102 | 32919361 | A/G | PSMB8 | 0.6949 | 3.6305 | 6.1974 | 12.673 |
| rs6929078 | 32946642 | C/T | | 1.4111 | 2.3594 | 5.7007 | 9.8467 |
| rs7757767 | 32953851 | A/G | | 1.4111 | 2.3217 | 5.5148 | 9.7788 |
| rs241404 | 32973975 | A/G | AK057104 | 2.0394 | 1.773 | 6.5301 | 8.9604 |
| rs1480380 | 33021224 | C/T | | 4.3916 | 7.0487 | 18.142 | 38.393 |

### 3.2.4.2 Chromosome 2

Within the extended *PLA2R1* region on chromosome 2 there are 30 significant associations between trait and marker for the joint analysis, 2 of them are positive association signals in the UK cohort, 6 in the Dutch cohort and none in the French cohort (Table 13).

**Table 13 Comparison of association signals on chromosome 2**
Association signals in the extended *PLA2R1* region on chromosome 2, shown in red the ($-\log_{10}P$ values above the significance threshold ($-\log_{10}P > 8.5$).

| Marker | Hg18 Position | dbSNP130 alleles | LocusRegion Gene | F $-\log_{10}P$ | NL $-\log_{10}P$ | UK $-\log_{10}P$ | 3P $-\log_{10}P$ |
|---|---|---|---|---|---|---|---|
| rs2042772 | 160370304 | T/C | LY75 | 1.03 | 1.34 | 2.14 | 9.10 |
| rs6711770 | 160401007 | T/C | LY75 | 0.72 | 1.40 | 4.21 | 9.29 |
| rs10929956 | 160422861 | T/C | LY75 | 0.67 | 4.85 | 3.62 | 13.70 |
| rs2666986 | 160451017 | C/T | LY75 | 0.96 | 6.84 | 1.88 | 13.58 |
| rs7601374 | 160462296 | T/C | LY75 | 1.67 | 5.10 | 5.15 | 19.76 |
| rs12692575 | 160478277 | G/A | | 0.79 | 4.46 | 2.34 | 9.30 |
| rs1995950 | 160508874 | G/A | PLA2R1 | 0.87 | 2.94 | 3.58 | 9.00 |
| rs6432570 | 160516560 | T/C | PLA2R1 | 1.57 | 3.41 | 4.18 | 15.79 |
| rs3792161 | 160517469 | T/C | PLA2R1 | 1.00 | 2.95 | 3.69 | 9.10 |
| rs2715921 | 160550074 | G/A | PLA2R1 | 0.93 | 5.16 | 3.10 | 10.18 |
| rs1567537 | 160565814 | T/C | PLA2R1 | 0.52 | 5.22 | 3.62 | 10.05 |
| rs2667014 | 160576628 | A/G | PLA2R1 | 1.43 | 6.47 | 2.32 | 13.12 |
| rs883423 | 160587897 | C/T | PLA2R1 | 1.13 | 2.53 | 2.49 | 8.92 |
| rs3749117 | 160593688 | C/T | PLA2R1 | 2.48 | 10.10 | 7.11 | 27.37 |
| rs3792189 | 160594118 | C/A | PLA2R1 | 1.65 | 10.39 | 5.05 | 23.62 |
| rs3792192 | 160595121 | A/G | PLA2R1 | 2.31 | 9.87 | 7.25 | 26.96 |
| rs6722275 | 160601506 | T/C | PLA2R1 | 2.39 | 9.43 | 6.98 | 26.98 |
| rs877635 | 160611093 | A/C | PLA2R1 | 0.97 | 4.21 | 4.80 | 13.68 |
| rs4664308 | 160625743 | G/A | PLA2R1 | 2.29 | 8.98 | 9.68 | 28.06 |
| rs7598162 | 160643828 | T/C | | 0.28 | 5.47 | 6.63 | 12.68 |
| rs7573270 | 160657779 | C/T | | 0.28 | 5.35 | 6.40 | 12.41 |
| rs10497213 | 160671849 | T/C | ITGB6 | 2.23 | 7.65 | 6.83 | 22.79 |
| rs10497212 | 160672946 | C/A | ITGB6 | 0.12 | 3.56 | 3.73 | 8.70 |
| rs4580348 | 160694788 | G/A | ITGB6 | 1.71 | 2.60 | 1.89 | 8.70 |
| rs983916 | 160723312 | C/A | ITGB6 | 0.58 | 5.25 | 3.75 | 14.34 |
| rs4665164 | 160778324 | T/G | ITGB6 | 0.96 | 3.45 | 4.98 | 12.72 |
| rs1870102 | 160778828 | C/T | ITGB6 | 1.51 | 8.92 | 9.09 | 25.31 |
| rs3772069 | 160840008 | A/G | RBMS1 | 0.88 | 1.61 | 3.26 | 9.95 |
| rs12620861 | 160937905 | C/A | RBMS1 | 0.83 | 1.61 | 3.04 | 9.13 |
| rs7423892 | 161037890 | T/C | RBMS1 | 2.06 | 2.98 | 4.88 | 12.50 |

### 3.2.5 Linkage disequilibrium

#### 3.2.5.1 Chromosome 2

The marker on chromosome 2 with the second strongest association *P* value to IMN, rs1870102 (see Figure 70) is situated upstream of a gene (*ITGB6* – integrin beta 6 precursor) which is a neighbour of *PLA2R1*, a strong candidate for idiopathic membranous nephropathy. The situation of rs1870102 within a linkage disequilibrium block encompassing that gene has been checked for the HapMap Phase II release data on European (CEPH) samples (computed using the Haploview program [511], visualised in the UCSC genome browser as a track which has been created by D. Thomas; Figures 70 and 71) and on our own data (Figures 72, 73 and 74; computed in Haploview 4.2; according to the method of confidence intervals; the selected attributes being: confidence interval minima for strong linkage disequilibrium 0.9 - 0.6, upper confidence interval maximum for strong recombination 0.7, fraction of strong linkage disequilibrium in informative comparisons 0.7, markers with a minor allele frequency < 0.05 were excluded).

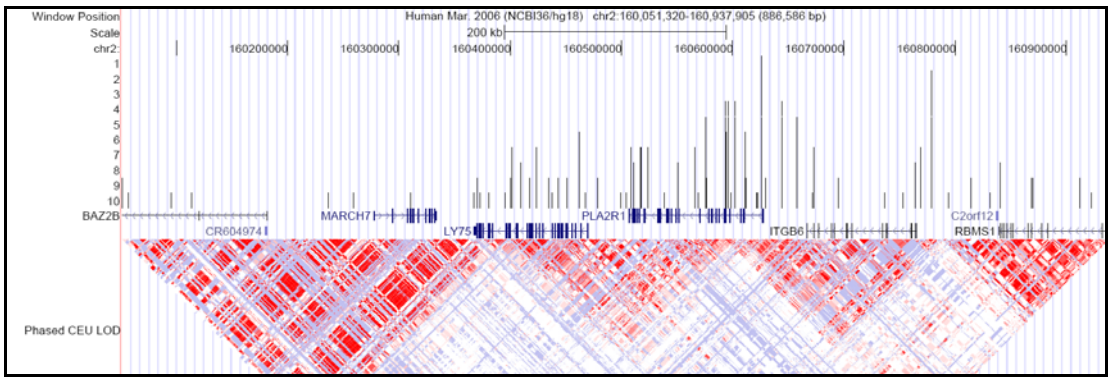


**Figure 70 LD blocks, HapMap CEU, in the region of interest on chromosome 2**
Inverted triangle display for the linkage disequilibrium (LD) structure of CEPH CEU population (HapMap Linkage Disequilibrium – Phase II – from phased genotypes, LOD LD values; hg18 assembly). Four haplotype blocks are perceivable.

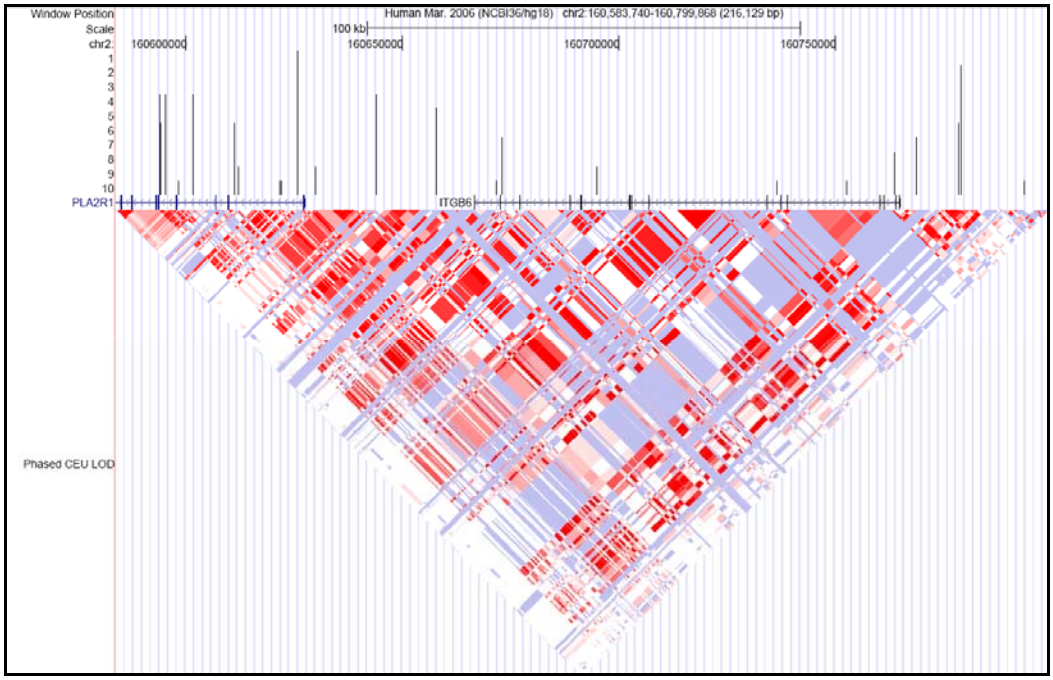


**Figure 71 LD block peri-PLA2R1, HapMap**
The third of the four linkage disequilibrium blocks in the region of interest previously shown on chromosome 2. It encompasses the two association peaks (in PLA2R1 and upstream of ITGB6). Inverted triangle display for the linkage disequilibrium (LD) structure of CEPH CEU population (HapMap Linkage Disequilibrium – Phase II – from phased genotypes, LOD LD values; hg18 assembly).

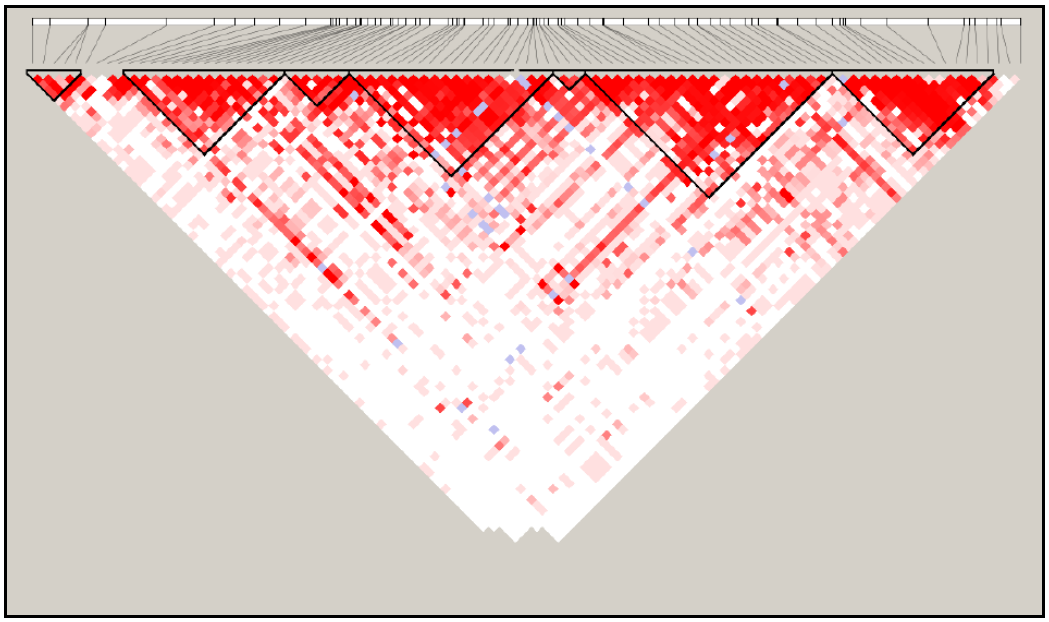

**Figure 72 LD plot, peri-PLA2R1, joint analysis cases**
Region coordonates chr2: 160032130 - 161140312 (rs964176 - rs13010247).

**Figure 73 LD plot, peri-PLA2R1, joint analysis controls**
Region coordonates chr2: 160032130 - 161140312 (rs964176 - rs13010247).

**Figure 74 LD plot, peri-PLA2R1, joint analysis cases and controls**
Region coordonates chr2: 160032130 - 161140312 (rs964176 - rs13010247).

It appears that rs1870102 is situated within an LD block including PLA2R1 in both the HapMap and our own dataset. The question arises whether the association signal detected at rs1870102 is due to a causative polymorphism affecting the closest gene (*ie* ITGB6) or reflects the strong linkage disequilibrium to a region where the causative polymorphism resides in PLA2R1.

The $r^2$ values for the most significantly associated markers around the locus of interest on chromosome 2 are shown (colour coded) in Figure 75, which was generated with LocusZoom version 1.1 [512].

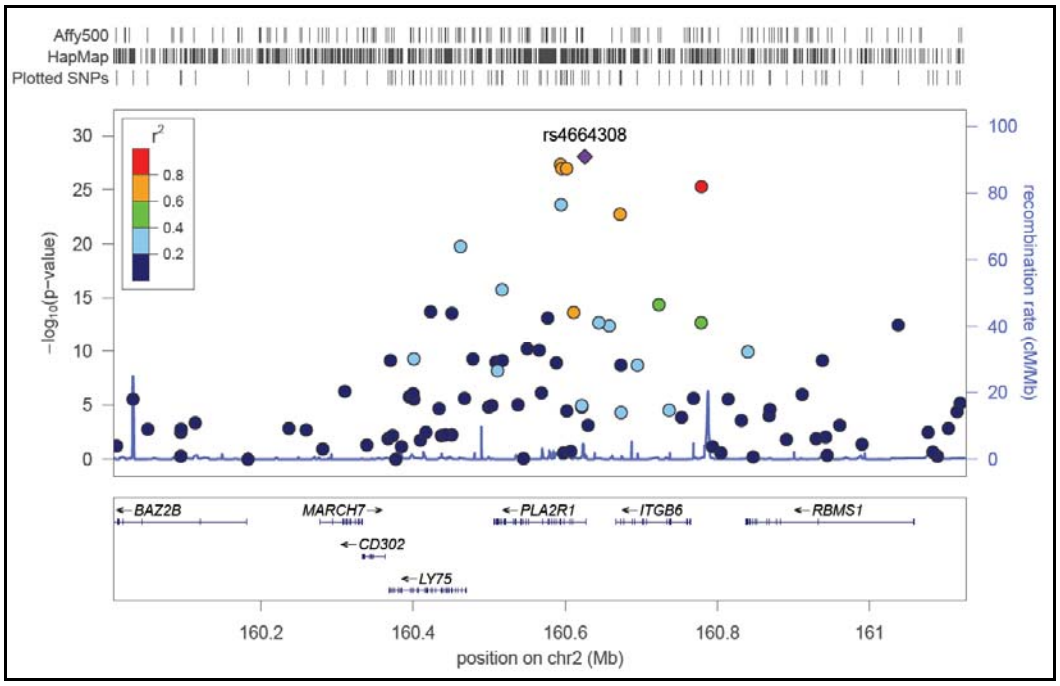


**Figure 75 LocusZoom plot of the region on chromosome 2**
Plot on Human Genome build hg18, range chr2:160006257-161127367; number of markers plotted: 93, reference SNP: rs4664308 ($P$ = 8.63 x $10^{-29}$). Y axis: significance level for association, left X axis: physical position, right X axis: recombination hotspot level (continuous blue curve along the lower margin of the graph); color coded: $r^2$ level. HapMap population used for LD computation: HapMap Phase II CEU. Upper margin of the graph: coverage of the region with Affy500 and HapMap markers *vs.* plotted markers.

### 3.2.5.2 Chromosome 6

Linkage disequilibrium has also been assessed for the extended MHC region on chromosome 6. We compared the LD block structure in cases (Figure 76) with controls (Figure 77) and case-control together (Figure 78). LD blocks were computed using Haploview 4.2).

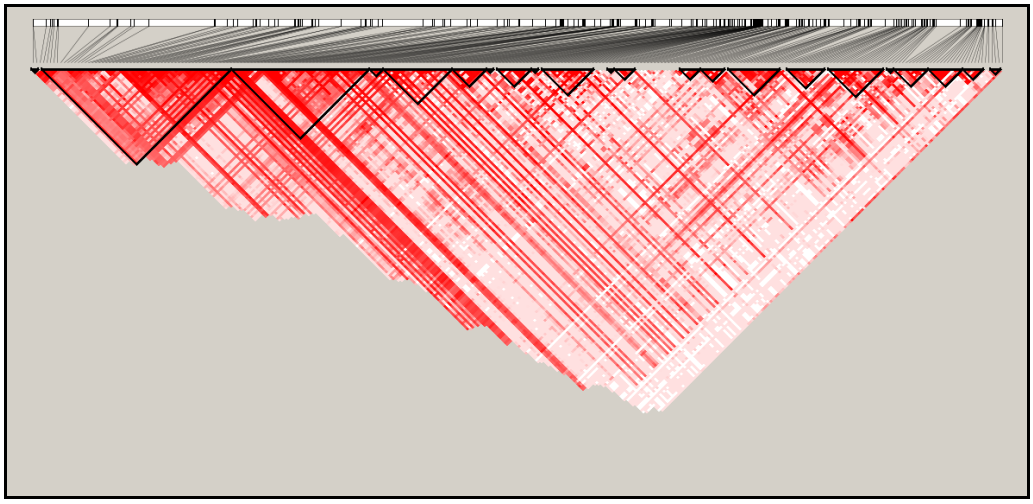

**Figure 76 LD plot, joint analysis, extended HLA region, cases**
Coordinates of the region assessed: chr6:25853721 - 33021224 (rs7776065 - rs1480380).

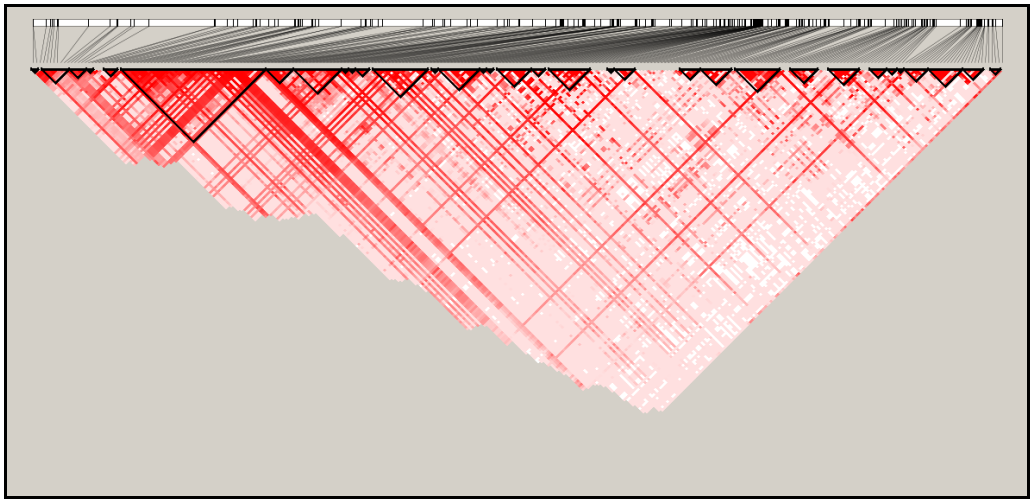

**Figure 77 LD plot, joint analysis, extended HLA region, controls**
Coordinates of the region assessed: chr6:25853721 - 33021224 (rs7776065 - rs1480380).

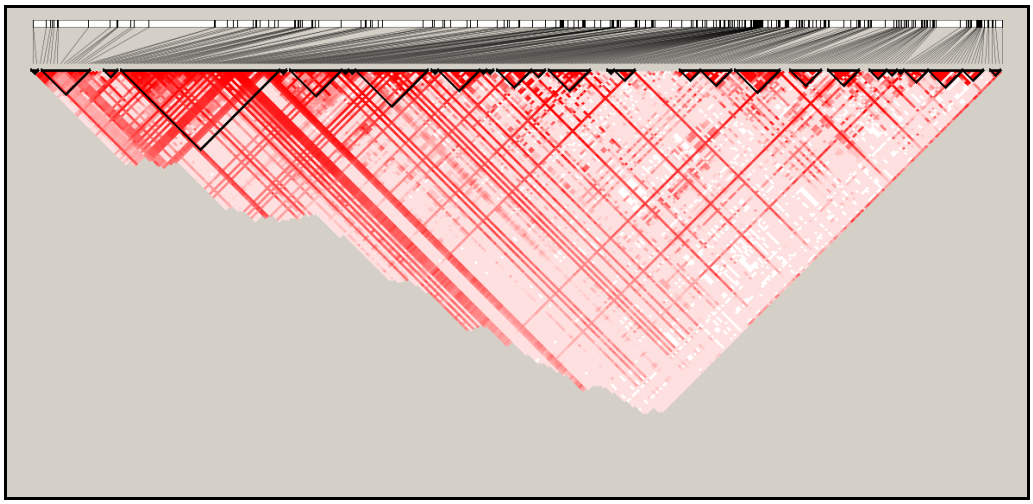

**Figure 78 LD plot, joint analysis, extended HLA region, cases and controls**
Coordinates of the region assessed: chr6:25853721 - 33021224 (rs7776065 - rs1480380).

The LD values ($r^2$) for the most significantly associated markers around the locus of interest on chromosome 6 (colour coded; generated with LocusZoom version 1.1), are shown in Figure 79.

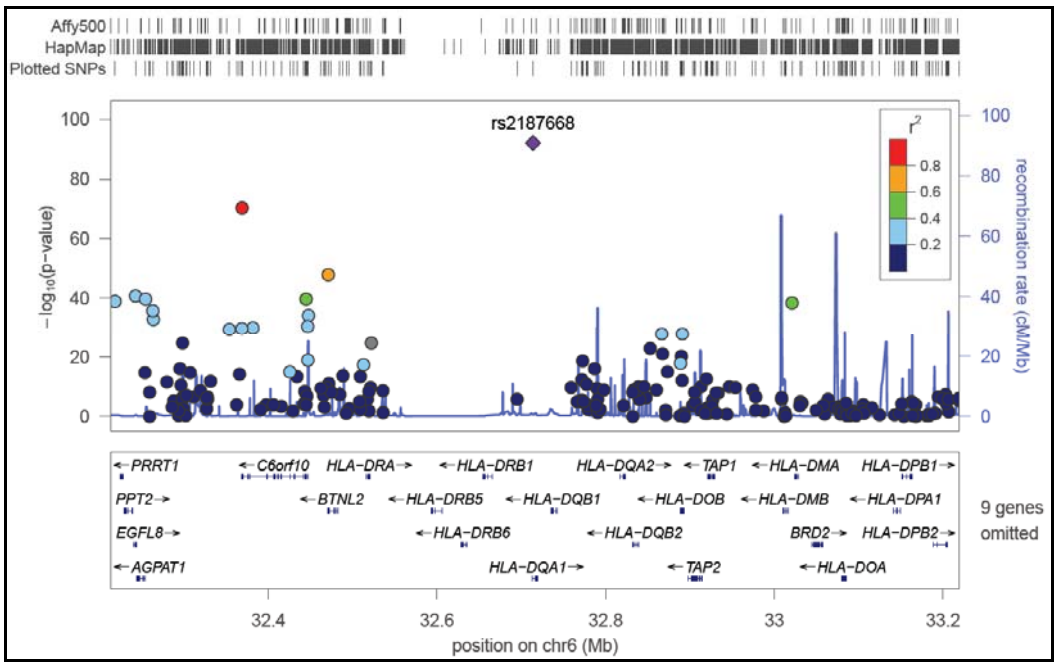


**Figure 79 LocusZoom plot of the region on chromosome 6**
Plot on Human Genome build hg18, range chr6:32213160-33219407, number of markers plotted: 209, reference marker: rs2187668 (*P* value 8 x $10^{-93}$); genes omitted: RNF5, RNF5P1, AGER, PBX2, GPSM3, NOTCH4, PSMB8, PSMB9, PPP1R2P1. Y axis: significance level for association, left X axis: physical position, right X axis: recombination hotspot level (continuous blue curve along the lower margin of the graph); color coded: $r^2$ level. HapMap population used for LD computation: HapMap Phase II CEU. Upper margin of the graph: coverage of the region with Affy500 and HapMap markers *vs.* plotted markers.

## 3.2.6 Epistasis

Statistical epistasis (non-additivity, deviation from additivity) is considered to be suggestive for biological epistasis (interaction among biomolecules). In the case of allele by allele epistasis for two loci there are 4 possible alleles: a, A, b, B. If two alleles (one at each locus: A, B) are causative, the effect of alleles on the phenotypes can be quantified, Table 14.

**Table 14 Epistasis**

| Alleles | 0 | A | B | AB |
|---|---|---|---|---|
| Quantified phenotype | 0 | x | y | x + y + e |

In the case of e = 0 the model is additive. If e > 0 one deals with synergistic epistasis, in the case of e < 0 with antagonistic epistasis.

Allele by allele epistasis was estimated using the logistic regression module implemented in PLINK for a binary trait (affected vs. unaffected by IMN). The allele dosage model to be fitted is:

$$Y \sim b0 + b1.A + b2.B + b3.AB + e$$

Not all combinations of SNPs were tested due to computationally constraints. We selected a subset of 13 significantly associated SNPs in the PLA2R1 region of interest and 281 SNPs significantly associated SNPs in the region of interest on chromosome 6 (extended HLA) for the combined analysis of the three assessed populations. Multiple pairwise correction for 3,653 comparisons yielded a significance threshold (for $P < 0.05$) of $-\log_{10}P = 4.86$. None of the values of the pairwise epistasis signals reached this significance level (Table 15).

**Table 15 Pairwise epistasis signals**

Pairwise epistasis signals for the allele by allele test combining 13 SNPs on chromosome 2 and 281 SNPs on chromosome 6. Shown are the results for $-\log_{10}P > 3$.

| SNP/chr2 | SNP/chr6 | -log10P |
|---|---|---|
| rs1870102 | rs6924102 | 4.48096 |
| rs4665164 | rs6924102 | 3.639217 |
| rs4664308 | rs1736971 | 3.570086 |
| rs12620861 | rs1633021 | 3.496482 |
| rs12620861 | rs1736971 | 3.408824 |
| rs6722275 | rs443198 | 3.300683 |
| rs4664308 | rs7383287 | 3.270916 |
| rs4664308 | rs1633021 | 3.258297 |
| rs4664308 | rs6924102 | 3.212469 |
| rs1870102 | rs7383287 | 3.211337 |
| rs4664308 | rs2844635 | 3.207678 |
| rs3792192 | rs443198 | 3.19942 |
| rs3749117 | rs6924102 | 3.172825 |
| rs7423892 | rs1736971 | 3.136083 |
| rs3749117 | rs443198 | 3.134659 |
| rs3772069 | rs1633021 | 3.128485 |
| rs4665164 | rs2395150 | 3.100727 |
| rs983916 | rs2516424 | 3.09114 |
| rs877635 | rs7383287 | 3.019542 |

### 3.2.7 Relative Risk / Odds Ratio

In statistical epidemiology, the relative risk (RR) refers to either the risk ratio or the odds ratio. The risk ratio refers to the proportion of the risk of an event (*i.e.* in cohort studies or clinical trials developing a disorder) occuring, between one two groups in which the risk is estimated. One of the two groups being exposed (to a potentially causal factor) the other being non-exposed (to the same factor). In case control studies the risk ratio cannot be obtained directly. A substitute can be calculated which is the odds ratio (OR), also named relative odds and approximate relative risk. OR and RR are considered interchangeable for practical purposes (especially when dealing with large samples and rare diseases, [513]).

$$OR = \frac{\frac{P_1}{1-P_1}}{\frac{P_0}{1-P_0}}$$

$$RR = \frac{OR}{(1-P_0)+P_0 \cdot OR}$$

Where:

RR = Relative Risk

OR = Odds Ratio

$P_0$ = incidence of the outcome of interest in the nonexposed group

$P_1$ = incidence of the outcome of interest in the exposed group

In the case of our study we assessed the risk of developing membranous nephropathy in the group bearing the predisposing genotype(s) *vs.* the same risk in the group bearing the protective genotype(s). The Odds Ratios for the most significantly associated alleles are given in Table 16.

**Table 16 Odds Ratios for most significantly associated alleles**
CI Confidence Intervals, MAF Minor allele frequency.

| | chr6: rs2187668 (HLA-DQA1) | chr2: rs4664308 (PLA2R1) |
|---|---|---|
| French GWAS | | |
| OR (95% CI) | 4.48 (2.68 – 7.50) | 1.87 (1.20 – 2.92) |
| MAF cases / MAF controls | 31.3% / 9.2% | 23.3% / 36.3% |
| p-value | 1.8x10-9 | 5.1x10-3 |
| Dutch GWAS | | |
| OR (95% CI) | 3.76 (2.92 – 4.86) | 2.27 (1.73 – 2.97) |
| MAF cases / MAF controls | 37.0% / 13.5% | 26.0% / 44.4% |
| p-value | 5.6x10-27 | 1.0x10-9 |
| British GWAS | | |
| OR (95% CI) | 5.33 (4.04 – 7.02) | 2.10 (1.67 – 2.64) |
| MAF cases / MAF controls | 41.9% / 11.9% | 25.3% / 41.6% |
| p-value | 5.2x10-36 | 2.1x10-10 |
| Joint GWAS | | |
| OR (95% CI) | 4.32 (3.73 – 5.01) | 2.28 (1.96 – 2.64) |
| MAF cases / MAF controls | 39.2% / 13.0% | 25.2% / 43.4% |
| p-value | 8.0x10-93 | 8.6x10-29 |

The combined risk of individuals homozygous for both risk alleles (SNP rs4664308 - PLA2R1 and SNP rs2187668 - HLA-DQA1) reaches an odds ratio (OR) of ~ 80 (additive based upon the combination of genotypes), which reaches a level which might even be considered diagnostic, as shown in Table 17.

**Table 17 Odds Ratios for the combined risk genotypes**
Odds Ratios for idiopathic membranous nephropathy computed depending on combined risk genotypes at at SNP rs4664308 (PLA2R1) and SNP rs2187668 (HLA-DQA1). Individuals homozygous for the low risk allele are considered to be the reference category. For each category, the OR of IMN is listed, followed by the 95% confidence interval, followed by the number of individuals / number of cases in that genotype category for the joint analysis.

| | | **SNP rs4664308** | | |
|---|---|---|---|---|
| **SNP rs2187668** | GG | GG<br>OR=1<br>(ref)<br>n=354/14 cases | GA<br>OR=2.22<br>(1.24-3.97)<br>n=944/79 cases | AA<br>OR=4.19<br>(2.36-7.46)<br>n=659/97 cases |
| | GA | OR=6.07<br>(3.01-12.27)<br>n=115/23 cases | OR=8.49<br>(4.73-15.22)<br>n=363/94cases | OR=25.43<br>(14.32-45.16)<br>n=348/178 cases |
| | AA | OR=20.24<br>(5.51-74.38)<br>n=11/5 cases | OR=31.03<br>(13.72-70.19)<br>n=41/23cases | OR=78.46<br>(34.55-178.17)<br>n=55/42 cases |

# 4 Discussion

## *4.1 Phenotype*

Membranous nephropathy is a disorder in which patients lose large amounts of protein through the urine, and the progression of a substantial proportion of those affected by it towards renal failure cannot be efficiently stopped. In order to treat this affliction, and most importantly to prevent the progression in the cases where it happens, we would ideally want to base our treatment on an etiological / pathogenetical basis. That, of course, would imply the understanding of the cause(s) and of the mechanism(s) of the disorder. Strong evidence has been gathered as to the autoimmune nature of the disease, which in turn is based on an important genetic component. The most important proportion of the genetic predisposition in the case of autoimmune disorders has been shown to be conferred by molecules encoded in the HLA region. HLA associated disorders are caused by a complex interplay of genes and environmental factors. It is generally considered that HLA variants are necessary but not sufficient for these diseases to develop [514]. An important and difficult question is which of the many polymorphic loci in the HLA region is causally involved (*ie* primarily associated) and which one is secondarily associated due to the strong linkage disequilibrium existing between genes at this super-locus [515]? This raises a catch-22-like problem: since in order to establish the primary associations, the knowledge of the mechanisms is needed and the reason to unveil associations is for them to suggest possible mechanisms.

## *4.2 Methods: Genetic mapping*

### 4.2.1 Linkage

Given the rarity of IMN, a classical Mendelian monogenic transmission mechanism was plausible. This was further supported by the existence of known families where the phenotype was segregating. At least one of these families suggested X-linked transmission. The X chromosome effect was further supported by the knowledge of the different prevalence of IMN in males versus females. The classical linkage analysis of the respective family yielded

inconclusive results. One possible explanation for this is that IMN is not 'simple' (monogenic Mendelian), but could theoretically follow a digenic (oligogenic) or even polygenic model of transmission. The real question is: 'how complex is IMN?'. The concepts of complexity and, implicitly, complex disorder are best treated in the context of the 'effect size / allele frequency determinants paradigm', which has been suggested by McCarthy *et al.* in 2008 [478], see Figure 80.

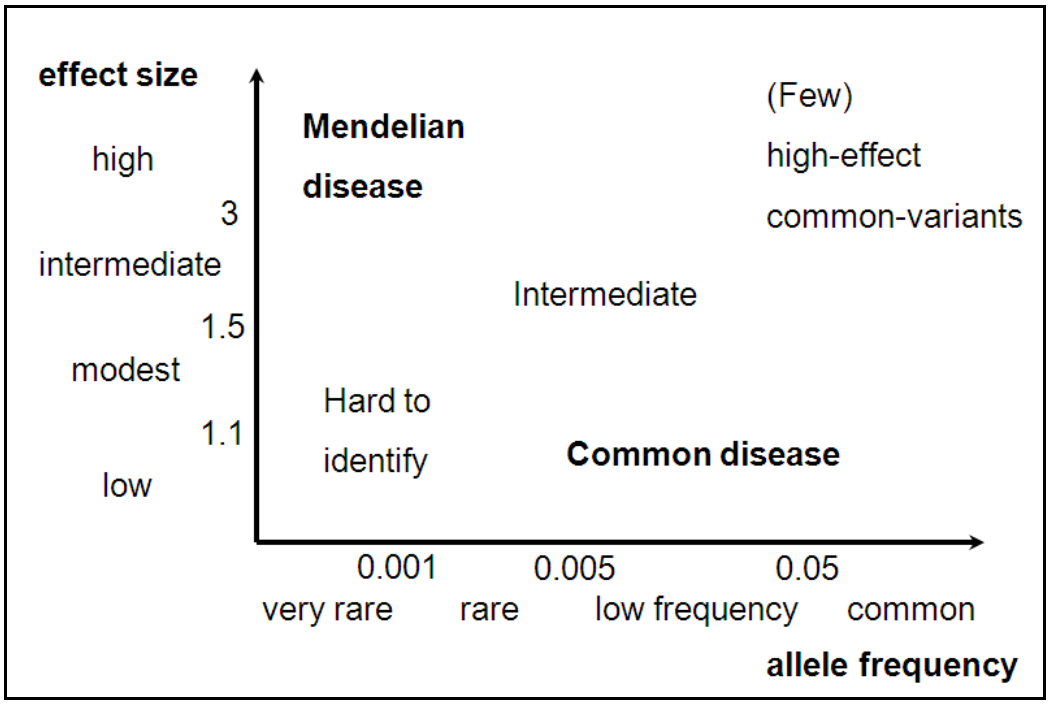


**Figure 80 Effect size – allele frequency paradigm graph**
Mendelian diseases: usually determined by rare (frequency: 0.001 – 0.005) alleles with high (> 3) effect size; are best mapped by linkage. Common diseases: considered to be determined by either common (frequency > 0.05) or rare (frequency: 0.001 – 0.005) alleles of (usually) low (< 1.1) effect size; are usually mapped by association.

This paradigm also suggests different strategies to be used in different gene mapping projects depending on the situation of the target in one of the classification sets. Rare disorders assessed are usually Mendelian monogenic and mapped *via* linkage (positional cloning), while complex disorders are generally considered to be caused by common alleles which contribute a small effect size to the phenotype and are best addressed *via* GWA studies. In order to adapt our mapping strategy, we should rephrase the question pertaining to the complexity of IMN as 'where is it situated within the effect size / allele frequency paradigm?' (Figure 80).

### 4.2.2 Association

After the initial success of mapping the causative gene (rare variant) in an apparently simple rare monogenic disorder (age related macular degeneration) was considered to be a proof of principle [438], the most visible GWAS studies started focusing on common diseases, therefore falling in the low effect size common variant category. There is a debate which opposes the 'common allele common disorder' (CACD) and the 'rare alleles common disorder' (RACD) hypotheses. Our understanding of the age-related macular degeneration (AMD) lesson is that a common complex disorder (AMD *latu sensu* implies at least 7 known loci) can be addressed by 'splitting' the complex phenotype in simpler well defined sub-phenotypes (the successful association study found the gene associated with AMD4, CFH – complement factor H, based on a genome-wide screen of only 96 cases and 50 controls!), or alternatively by brute force (associations in diabetes started to become significant only after increasing the numbers of the cases and controls by extensive international collaboration, a 'lumping' approach). In defining a genetic entity, the approach can be either a 'splitting' or a 'lumping' one [516]. One could envision glomerulonephritis to be a 'lumped' entity, of which IMN would be a better defined (pathologically defined through renal biopsy) 'splinter'. Therefore one might expect relatively small numbers of cases versus controls to be sufficient to identify a hypothetical major influence locus. Our exploratory (UK) study proved this to be true, by employing 336 cases and 349 controls, with the resulting 2 loci of association: on chromosome 6, in the extended HLA region, and on chromosome 2, in the vicinity of *PLA2R1*. The main conclusion of our studies is that sequence variations within HLA and PLA2R1 may be partially responsible in determining the susceptibility for MN (in Caucasoids). The findings regarding the PLA2R1 locus support, in an unbiased manner (genome wide association studies being hypothesis free - therefore unbiased by *a priori* assumptions) and using a different approach, the finding by Beck *et al.* (whose results are based on serological – immunological technology) that PLA2R1 is an important autoantigen in patients with MN.

Our studies revealed a stronger association of the HLA locus with IMN than the one of PLA2R1, which suggests that the risk conferred by HLA-DQA1 alleles is higher than the risk conferred by PLA2R1 alleles. For the joint analysis the Odds Ratio for the HLA-DQA SNP on chr6 (rs2187668) is 4.32 (confidence interval: 3.73 – 5.01) compared with the Odds Ratio for the PLA2R1 SNP on chromosome 2 (rs4664308) which is 2.28 (confidence interval: 1.96 – 2.64).

#### 4.2.2.1 Problems of association studies

There are two classes of known important problems in interpreting association results, both deriving from the same question: is the result relevant? A positive association result can be a false positive due to chance or to a systematic bias introduced through the study design or the way the data were analysed. It can also be a real association in linkage disequilibrium with the true (causative) site or the causative site itself (which is the real target). This problem can be addressed in a statistical / epidemiological approach (to eliminate false positives) – by replicating the initial exploratory study, thus adding statistical relevance. The best way of addressing the causality problem is to look for biological proof / meaning of the results.

#### 4.2.2.2 Replication

Replication studies were performed according to the criteria (study independency, population and phenotype similarity, demonstration of similar effects and their magnitudes) suggested by the NCI-NHGRI Working Group on Replication in Association Studies [452] for establishing positive replication. The results of the initial exploratory study (UK) were replicated in two independent yet similar (historically, geographically and culturally) Caucasoid populations: French (analysing 75 cases and 157 controls) and Dutch (analysing 146 cases and 1832 controls). The French study replicated the locus on chromosome 6, the Dutch study both loci (on chromosomes 2 and 6).

The causality problem implies different aspects: linkage disequilibrium, combined independent signals and epistasis.

#### 4.2.2.3 Linkage Disequilibrium

Linkage disequilibrium was expected to be a problem at the level of the extended HLA locus on chromosome 6 due to the complex structure of this locus (genes in this complex are linked tightly due to the fact that recombination rarely occurs, possibly to protect some evolutionary advantageous grouping of gene variants on the same haplotypes). Establishing the primary HLA association is therefore notoriously difficult, which is further complicated by the possibility of a primary association with a particular combination of genes in the region (*ie* an extended haplotype); or by phenotype heterogeneity (in which different forms of the disorder could be associated with different HLA haplotypes). Empirical evidence can suggest primary association candidates: they are usually considered to have the strongest association signal and functionally they usually are variants of peptide presenting HLA molecules [517]; both conditions apply to our (strongest) association signal in HLA-DQA1.

Even at the locus on chromosome 2, there is at least a pair of markers which are in strong linkage disequilibrium, which could be either incidental or suggest a common regulatory element being important in the expression of functionally related genes in the genomic region (*vide infra*).

#### 4.2.2.4 Independent association signals

A general problem for association studies is the question regarding association signals in a region of interest: how many are due to linkage disequilibrium and how many may represent independent association signals? This is an even more acute problem when dealing with association studies for immune related traits. Many show associations within the HLA region [518], the association signal within this super-locus often showing a 'cloud' morphology. Much of the dispersion of the cloud can be explained by the tight linkage disequilibrium of the region, but there is also accumulating evidence for the contribution of combined independent signals.

In our studies, the presence of significant associations within the extended HLA locus, with SNPs in little linkage disequilibrium with the top signal, suggest the presence of independent risk variants.

#### 4.2.2.5 Epistasis

The problem of epistasis appears once we realize we are dealing with one phenotype (at least, if we do not take into account heterogeneity) and (at least) two loci. Epistasis and pleiotropy are recognized as explanations for the deviations from simple Mendelian ratios. If a simple disorder is (ideally) defined by a one-to-one relation between phenotype and genotype, complex disorders can be defined (in theory) by a many-to-many relationship. Many phenotypes would reflect phenotypic heterogeneity or pleiotropy, whilst many genotypes reflect genotypic heterogeneity or epistasis.

Epistasis (which etymologically means 'standing upon') has been defined statistically and biologically. In 1901, Bateson used the term for the first time to describe the masking effect of one gene upon another gene. Fisher in 1918 referred to 'epistacy', which would reflect deviations from additivity between the effect two loci have upon a phenotype. There are instances in which statistical epistasis does not correspond to biological epistasis (which is nowadays understood to reflect interactions of products of genes) and, conversely, biological interactions between products of genes can sometimes be not observed as statistical epistasis [519]. In the case of our study, the signal for statistical epistasis between chromosome 6 and chromosome 2 does not reach statistical significance (Table 15). This can be explained by the fact that the two loci are probably not encoding molecules which are part of a biological module of coevolved molecules. It seems to be more plausible that we are dealing with 'accidental epistasis' (immunology 'weapon / target'), in which there is no evolutionary gain from this interaction, therefore no statistical deviation from the additivity of the influences on the trait would be expected. Referring back to Figure 80, one can understand the spectrum of disease being flanked by the two extreme cases: on one hand unique detrimental mutations (the most detrimental being non viable) and, on the other hand, the genes whose function has been efficiently 'canalized' (canalization being defined as the mechanism by which gene functions are multiplied in complex organisms, making them redundant

*ergo* robust to single mutations [520], epistasis thus contributing to robustness [521]), to allow for flexibility and higher adaptability of the organism even when confronted with mutations at the respective loci. A subset, perhaps even the majority, of genes are situated in between these two extremes [522].

#### 4.2.2.6 Causation

The jump from statistically relevant association to inferring causation is an important and difficult one. Given that biological relationships can only be inferred from empirical observations, the problem can be stated as 'what constitutes sufficient evidence for causation?' [523]. There have been proposals to quantify clear criteria of causality, modelled on the Henle-Koch postulates which were considered the *sine qua non* for establishing a causal relation between an organism (pathogenic agent) and the disease it was supposed to cause [524,525], but the danger has been recognized early of inflexible rules becoming a hindrance rather than an aid in elucidating etiopathogeny (Rivers, 1937, commenting on the inadequacy of Koch's postulate applied to viral infections [526]). Therefore the consensus shifted towards using guidelines as opposed to formal rules. An important set of principles pertaining to the relation between association and causation (strength, consistency, specificity, temporality, biological gradient, plausibility, coherence, experiment, analogy) were stated by Hill in 1965 [527]. Our data adhere to the conditions of strength, due to our choosing of a conservative *P* value. Consistency has been ensured by successive replications following our exploratory study. Our replication studies can be considered then to be, within the theoretical framework of Lykken, conceptual replications, as opposed to operational replications [528]. The latter would strive to repeat the original study exactly whereas the former had the purpose to confirm the initially observed associations by using different methods *i.e.* genotyping method, phenotyping method, statistical procedure, data-analytic software *i.e.* PLINK). Specificity must not be overemphasized, given that we cannot assume a simple one-to-one relationship between cause and effect. Plausibility, coherence and analogy are addressed within the framework of immunogenetics, two exemplary cases, which illustrate a similar approach to the deciphering of molecular mechanisms being type I diabetes [529], with a prevalence of ~ 1-2/1000, and celiac disease, with a prevalence of ~ 1/100)

[530]. Diabetes and celiac disease turned out to imply a strong, major locus, HLA effect and added polygenic and environmental components. The environmental aspect is best understood in the case of celiac disease, where the disease inducing agent is known to be gluten. The very small incidence of IMN when compared to celiac disease suggests probable differences along with the similarities in the pathogenetic architecture of the two disorders.
The most important of the conditional principles suggested by Hill *i.e.* experimental evidence, will have to be addressed through further investigation.

## *4.3 Molecular dissection*

The highly significant associations between alleles of HLA-DQA1 and PLA2R1 and the IMN trait do not establish causality. They could, however, suggest a possible mechanism in which risk variants of immune system proteins and glomerular components can interact to establish a basis for the development of membranous nephropathy. We have reviewed the knowledge pertaining to PLA2R1 and HLA in order to delineate what is known, what is to be still found out about their biology – and whether we can suggest biologically relevant scenarios in which genetic variants at the two loci might be thought to influence the pathogenesis of IMN, in the hope to add to the conditions of biological plausibility and coherence, on the quest for causality.

### 4.3.1 Peri PLA2R1 locus

There are three genes comprised in the two linkage disequilibrium blocks which show association signals on chromosome 2: LY75, PLA2R1 and ITGB.

#### 4.3.1.1 PLA2 / PLA2R unit

PLA2R1 seems to be an obvious candidate as the causal molecule in the association region of interest on chromosome 2 since Beck and colleagues suggested, based on immunological methodology, that it is the target antigen in IMN [1]. PLA2R1 is likely to encode two isoforms: the proteins corresponding to the two alternative splicing products are predicted to be transmembrane and soluble. The membrane bound form is expressed at the level of the podocytes, whilst the soluble form, produced by the cleaving of the membrane bound form by metalloproteinases, is hypothesized to be secreted by the liver and by the

kidney. Their physiological function is not well defined. PLA2R1 is a member of the subgroup VI (mannose receptor family) of the C-type animal lectin superfamily. The members of this family (which also include CD-205 / LY75 and CD-206 / MMR) are antigen-uptake receptors which bind, internalize and deliver ligands to the endosomal compartment of antigen presenting cells, where the HLA class II type loading occurs, Figure 81. One can consider the antigen-uptake receptors 'interface molecules' linking the simple, innate, pattern recognition based immune response with the complex, specific adaptive immune response mediated by HLA molecules.

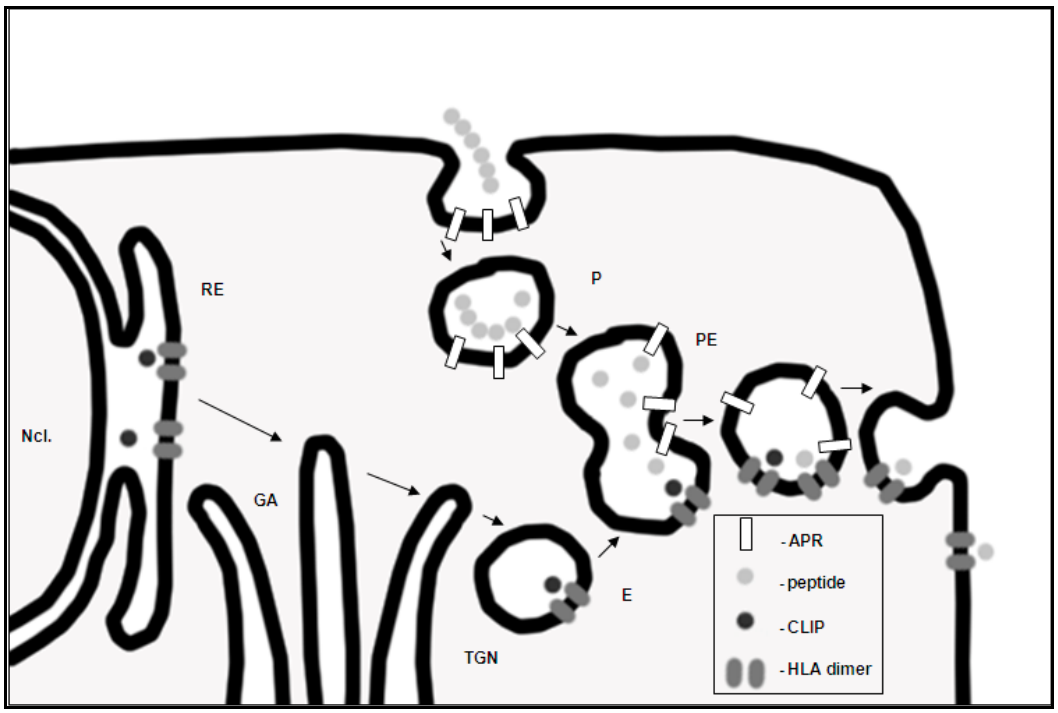


**Figure 81 Presentation of Exogenous Antigens**
Antigens are taken up through endocytosis mediated by antigen-uptake receptors (APR). The antigens are cleaved at the level of phagosomes (P). Phagosomes merge with endosomes (E) into endophagosomes (PE). The antigen peptides occupy the presentation groove of the HLA class II dimer which had been previously protected by the CLIP (Class II Inhibiting Protein) fragment. Ncl – nucleus, RE – endoplasmic reticulum, GA – Golgi apparatus, TGN – trans Golgi network.

Apart from its possible role as an antigen-uptake molecule, PLA2R can function through ligand binding, as classical receptors. The known ligands of PLA2R are synthesized as inactive pro-enzymes which are activated through proteolytic cleavage by proprotein convertases: secreted PLA2s. The discovery of PLA2 receptors expanded the concept of simple, enzymatic, PLA2 function into that of the more complex biological activity of the PLA2 / PLA2R functional unit. The

simple function of PLA2 is presumed to be primarily digestive and pertains to the enzymatic hydrolysis of fatty acid ester bond of phospholipids. The complex biological activity of PLA2 / PLA2R manifests itself through the binding of PLA2 to the cell surface PLA2R. The binding of the enzyme to its receptor has been hypothesized to be regulatory for the enzymatic function (in both ways: stimulatory and inhibitory). The inhibitory mechanisms are supposed to be based on ligand internalization by endocytosis (for the membrane bound isoform) and inhibitory binding – leading to the clearance of the active enzyme (for the circulating, soluble form). The stimulatory mechanisms are based on the activation of signal transduction signals through the occupying of the receptor by its ligand. The signal transduction can lead to a plethora of biological effects, mostly mediated by the products of lipid mediators (thromboxane, arachidonic acid, prostaglandins): growth promotion and proliferation, chemokinetic migration, secretagogue effects, contractile response etc.

sPLA2 are considered to be a class of proinflammatory mediators, stored within inflammatory cells and released in the extracellular space (plasma) upon activation (during local or systemic inflammation) [531]. It therefore follows that depending on the prevailing effect, the regulatory function of the binding by the receptor (PLA2R1) will lead to it having a proinflammatory (in the case of stronger activation) or anti-inflammatory effect (if, on the contrary, inhibition is stronger), Figure 82.

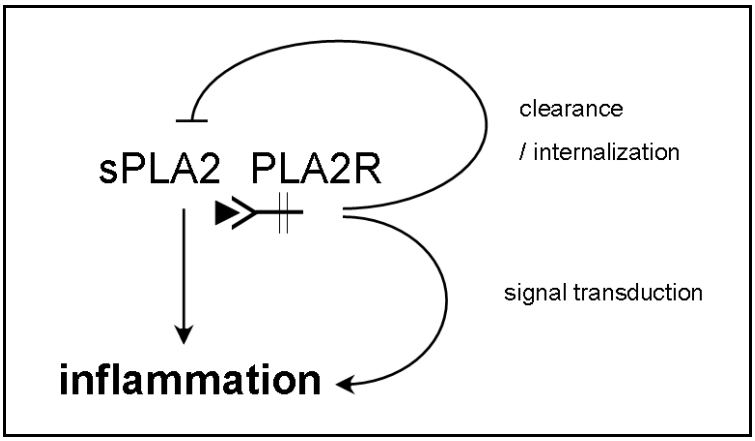


**Figure 82 Influence of the sPLA2/PLA2R functional unit on inflammation**
Upon dockage of the ligand (sPLA2) on its receptor (PLA2R) two antagonistic effects impacting on the inflammatory role of sPLA2 can occur: inhibitory, due to the clearance of the enzyme by its receptor-dependent endocytic internalization followed by digestion *ergo* inactivation; stimulatory, due to the activation of a stimulatory (eicosanoid) signal transduction pathway.

*Antigen PLA2R1*

The findings of our association studies also suggest another aspect of the pathogenesis of IMN: the formation of autoantibodies. PLA2R1 has been shown to be present not only on normal podocytes but also in the immune deposits of patients with idiopathic membranous nephropathy. Sequence variants causal for (or increasing susceptibility to) IMN could lead to conformational changes of the peptide or to changes to the pattern of antigen processing, PLA2R1 could be the antigenic protein, and the epitopes could be fragments of it. A similar mechanism has been shown in the case of the Goodpasture autoantigen [532].

A candidate to the title of causal polymorphism in this sense (possible conformation change resulting in pathogenic autoantigenicity) could be the methionine to valine (M292V) amino acid substitution (SNP rs3749117) which is associated with IMN in our study.

This appears to be even likelier in the light of the suggestion that sulfhydril (-*SH*) groups of the PLA2R1 extra cytoplasmic domains could be modified under the influence of environmental factors *i.e.* therapeutic or accidental exposure to heavy metals (gold, mercury etc. [533-535]), which is an important cause of secondary MN.

#### 4.3.1.2 LY75

It is worth noting that LY75 (Lymphocyte antigen 75), the gene situated telomeric from PLA2R1, is the human equivalent of DEC-205 (CD 205) [536] which is also a member of subgroup of the VI C-type animal lectin family. The tandem situation of the two genes (similarity) suggests the possibility of a past gene duplication event followed by divergent evolution. Theoretically, being an antigen-uptake molecule at the interface between innate and adaptive immune responses (like PLA2R1) LY75 could also be involved in the pathogenesis of IMN.

#### 4.3.1.3 ITGB6

The gene situated centromeric from PLA2R1 is ITGB6 (Integrin Beta 6 precursor). It could be functionally relevant to IMN through its role in the homeostasis of phospholipids and collectins, and in the remodelling of the extracellular matrix (it has been shown to lead to increased collagen production *via* activation of TGF-β1 when induced by viral infection in endothelial cells; [504]). It is worth mentioning that the association signal upstream of ITGB6 could be associated to a regulatory element (enhancer, silencer) which might participate in the co-regulation (expression) of genes downstream (ITGB6, PLA2R1, LY75).

### 4.3.2 HLA

The association of IMN with the HLA super-locus [224], demonstrated both serologically and through molecular typing has been reviewed by Powis [537]. One of the most polymorphic regions within HLA is DQ – DR, with DR being considered the most polymorphic class II HLA human locus. The IMGT / HLA database (part of the ImMunoGeneTics project [538]) documents to date (December 2010) 966 different DRB alleles, the HLA-DQB1 locus harboring 144 alleles and HLA-DQA1 35 alleles. Associations of HLA-DQA1 alleles with various immune related disorders have been published. In MN the culprit between the two loci seemed to be DQ, given the number of publications showing association of DQA1 with the phenotype [244,245,539]. Another argument why the HLA-DQ region appears to be important in defining the genetic basis for susceptibility for IMN is a study of the H2 region in mice, in which the Aβ / Aa region was found to be linked to susceptibility to anti-GBM disease (Goodpasture syndrome). The murine Aβ / Aα region corresponds to the human HLA-DQ region. This suggests a possible role for HLA-DQA1 alleles in immune related glomerular disorders [183].

Our study has managed to define the HLA-DQ – MN trait association with a better resolution of detail (we have been able to identify 282 SNP type markers within the HLA region which are associated with the MN trait. The strongest association was shown to be with HLA-DQA1 alleles).

Advances in genotyping technology allow for a higher resolution in resolving HLA alleles (the Illumina genotyping platform being designed with a high density of coverage over the complex HLA super-locus [226]). A possible reason which led to misleading results in the past is also the fact that HLA-DQA1 is not (as of December 2010) part of the routine HLA typing routine for renal transplantation (at least in the UK).

The presence of significant associations within the extended HLA locus, with SNPs in little linkage disequilibrium with the top signal, suggests the presence of independent risk variants. The distribution of the association signal (Figure 46) may suggest a possible functional link between the type II HLA molecules (HLA-DQA1) and type I HLA molecules**.** This has been suggested *e.g.* in the case of vitiligo [540].

Alternatively, this might be just the result of strong linkage disequilibrium between class I and class II molecules encoding genes due to co-evolution under similar evolutionary constraints of a functional / physiological / immunological unit ('immune response module').
Even though our findings are highly suggestive towards implying the importance of genetic variants within HLA-DQA1 in the pathogenesis of IMN, we cannot exclude the possibility of causal variants (in linkage disequilibrium with associated markers) being located in one of the genes close by.

## *4.4 Suggested pathogenic models*

### 4.4.1 Main pathogenic model

In trying to depict this vulnerability, we could use a 'shooting barrack' metaphor, in which the 'target' is the autoantigen at the level of the glomerular basement membrane (PLA2R1 etc), the 'bullets' are the autoantibodies which are fired by the 'gun' (which is the immune system) of which the 'trigger' is the HLA type II component (with its molecular representative in this particular case being HLA-DQA1). The question remains as to what the 'finger of the trigger' (the causative event) might be.

In a ‘key-lock’ type of compatibility, IMN specific sequence variants of PLA2R1 might lead to autoantigenic conformations, leading to their recognition by IMN specific HLA-DQA1 products, with their subsequent presentation by APCs to immune competent cells ($CD4^+$ T cells, B-lymphocytes) and the resulting triggering of autoantibodies, see Figure 83.

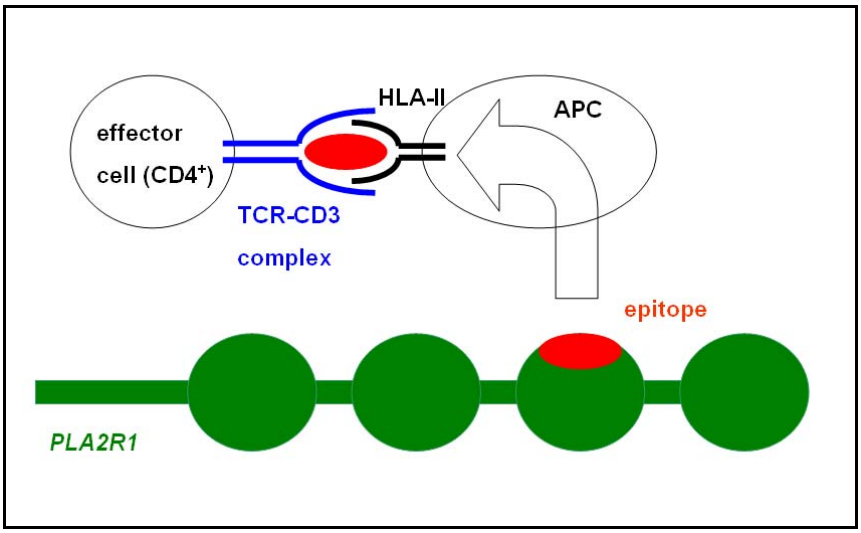


**Figure 83 Suggested schematic pathogenic mechanism**
TCR – T cell receptor; HLA-II - Human Leucocyte Antigen class II molecule (DQA-DQB dimer); APC – Antigen Presenting Cell; green – schematic structure of PLA2R1; red – hypothetic epitope; white arrow – endocytosis, processing of the antigen and presenting of the epitope on the surface of the antigen presenting cell.

The co-occurrence of risk alleles in PLA2R1 and HLA-DQA1 in an individual may lead to the possibility of a shortcut in the tightly regulated adaptive immune system and allow for a breach in self-tolerance, the immune attack leading to lesions at the level of the glomerular basement membrane (with the characteristic pathology of membranous nephropathy).

It is known that a proportion of IMN patients have a negative serology for anti PLA2R1 antibodies, which might be explained by the fact that HLA-DQA1 risk alleles can trigger autoantibodies against PLA2R1 or other targets. Two possible such autoantigenic targets have been published [157]. We did not identify significant associations (neither in the three separate population based studies nor in the pooled analysis) of the MN trait to markers within these genes (AKR1B1 on chromosome 7q33 and SOD on chromosome 6q25.3 respectively).

Other instances in which similar models of 'trigger and target' autoimmune mechanisms may apply have been published: in vitiligo, genetic markers have been shown to be associated with trigger variants in the HLA region and with target variants of TYR (tyrosinase - a previously well characterized autoantigen in vitiligo [541]).
Another lesson from previous HLA related genome wide association studies has been the hint that predisposition to autoimmune disorders might imply a complicated interplay of innate and adaptive immunity, which is based on variants encoded at the HLA super-locus (*e.g.* alopecia areata [542]).

### 4.4.2 Extended pathogenic model

The pathogenic mechanism of IMN can be modelled as a combination of two elements: an autoimmune aspect (the common element, including MN in the *genus proximum* of autoimmune disorders) and the reaction to the autoimmune attack, which includes the inflammatory response and the pathognomonic aspect of the basal membrane thickening (which is the defining *differentia specifica* of MN), see Figure 84.

#### 4.4.2.1 Autoimmune attack

PLA2R1 is an endocytosis receptor / pattern recognition like molecule. The role of pattern recognition molecules is to bind potential extracellular antigens. The pattern recognition process leads to endocytosis. The receptor bound antigens are processed in the early endocytic vesicles of the Antigen Presenting Cells: polypeptidic fragments are cleaved by peptidases. Some of these fragments (~ 10 amino acids) are extrinsic epitopes which will be taken up on the antigen presenting groove HLA type II molecules, in order to be then presented to the executor cells (T cells etc.) of the adaptive immune response system.

In the pathogenic context of IMN it seems that a part of PLA2R1 has its conformation altered as to be recognized as non self. The antigenic fragment of PLA2R1 (the intrinsic epitope) is processed (cleaved by peptidases) in the endocytic vesicle and be transferred on HLA type II triggering the production of

autoantibodies oriented against the autoantigenic portion of the extracellular chain of PLA2R1.

#### 4.4.2.2 Response to the autoimmune attack

Anti PLA2R1 autoantibodies can alter its function by blocking the binding of its ligands. In the case of sPLA2 this will have an effect on inflammation; in the case of collagen fragments [543] extracellular matrix turnover will be affected.

*Inflammation*

The most probable effect of the binding of sPLA2 by PLA2R1 is anti-inflammatory, due to the clearance of proinflammatory (Figure 82). Inhibition of this effect might lead to increased local inflammation with the subsequent lesions of podocytes (the long term vulnerability thereof being possibly the key to the evolution of the disorder towards resolution or towards progression) which will be manifest clinically as nephrotic syndrome.

*Extracellular matrix (ECM) / collagen turnover*

Both elements of the sPLA2/PLA2R1 functional unit have been shown to be important in extracellular membrane homeostasis: sPLA2 has been shown to activate the self aggregation of fibronectin (via the activation of integrin dimers through lysophosphatidic acid LPA); PLA2R1 has been shown to be able to bind collagen fragments. The reduced collagen binding capability of PLA2R1 due to autoantibody blocking of the binding site

*Importance of collagen uptake in MN*

Functional sequence variants having an effect on the collagen uptake function of PLA2R1 can, in theory, confer a ‘membranous type’ pathology to autoimmune lesions determined by autoantigens different from PLA2R1 itself (in a ‘two hit’ like hypothesis). This might be the case in type V lupus nephritis. The autoimmune attack (first step) could be determined by a lupus like autoantigen, whereas the membranous anatomopathological aspect of this particular subtype could be determined by a genetic variant of PLA2R1 which would impede collagen clearance.

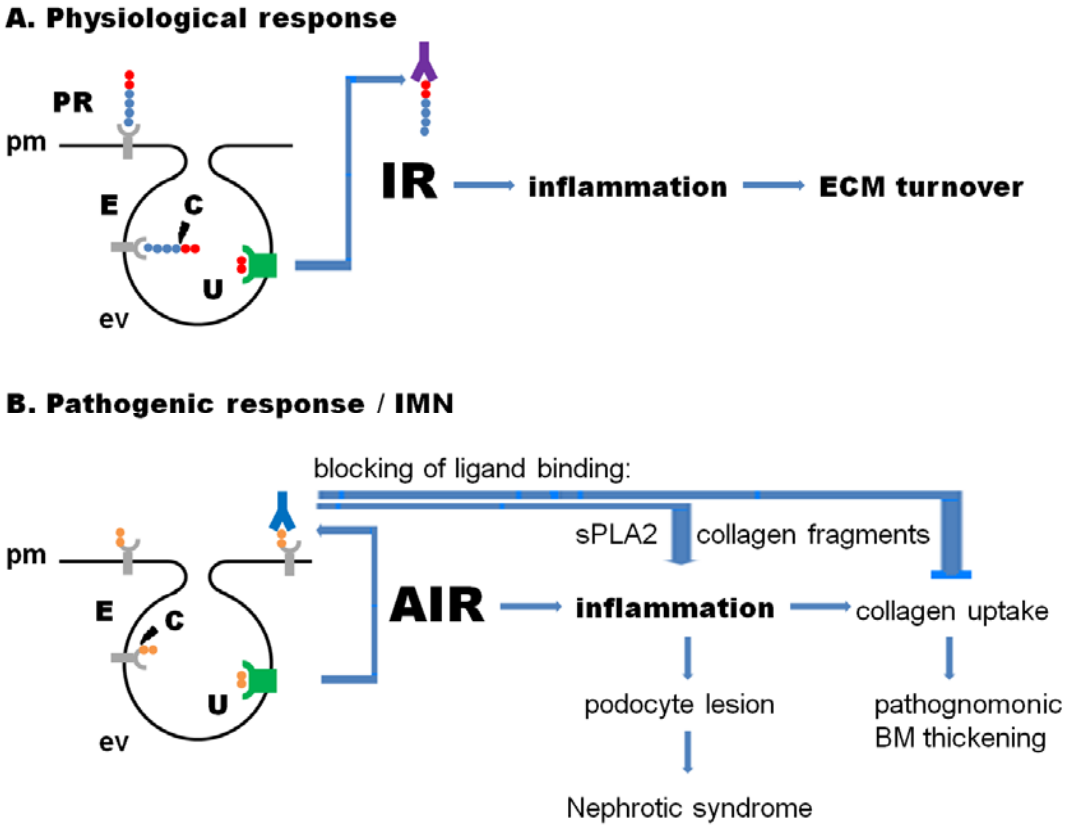


**Figure 84 Extended suggested pathogenic mechanism**
A. Physiological immune response at the level of the glomerulus; B. Pathogenic autoimmune response in the case of Membranous Nephropathy.
Blue and red circles – extrinsic antigen, red circles – extrinsic epitope, grey 'Y' shape – endocytosis receptor / pattern recognition molecule, green 'Y' shape – HLA type II receptor, purple 'Y' shape – Antibody, blue 'Y' shape - autoAntibody; pm – (podocytic) plasma membrane, ev – endocytic vesicle, PR – Pattern Recognition, E – Endocytosis, C – Cleavage (by peptidases), U – Uptake (by HLA type II), IR – Immune Response, AIR – Autoimmune Response, BM – basal membrane.

### 4.4.3 Alternative pathogenic models

A loss of function mutation of PLA2R1 could reduce its capacity to inhibit the antiinflamatory action of sPLA at the level of the glomerulus. Mutations of regulatory elements reducing the level of expression of PLA2R1 would lead to similar results.

Loss of function mutation of PLA2R1 could lead to its reduced capacity to clean up cellular detritus (immune complexes) in the case of local inflammation, resulting in the accumulation of extracellular material at the level of the glomerular basal membrane.

Gain of function mutations of PLA2R1 (or regulatory mutations resulting in its overexpression) could result directly in activation (or exacerbation) of a signal transduction pathway (autacoid factors, probably eicosanoids: arachidonic acid / prostaglandins) leading to pathologic results, possibly through the induction of ER stress responses which would result in glomerular epithelial cell lesions.

Mutations leading to the functional transformation of PLA2R1 from a molecule specialized mainly in interacting with sPLA2 into an antigen-uptake molecule (similar to DEC205, MR) could lead to the endocytosis of a polypeptidic fragment acting as an autoantibody.

Another possible pathogenetic mechanism was suggested by El Kossi *et al.* [544] who described a recipient of a kidney transplant in which the onset of *de novo* membranous nephropathy was associated with a donor specific alloantibody directed against HLA DQ7. In this scenario HLA type II molecular complexes are themselves the targets of antibodies as opposed to being triggers of the autoantibody synthesis. This model could be consistent with the modulation of the predisposition to autoimmune disorders by self HLA derived peptides [545].

## *4.5 Remaining questions*

It is not clear what the trigger of the autoimmune process is that ultimately leads to the deposition of immune complexes at the level of the basement membrane in the course of membranous nephropathy. A related question is why the onset of the process occurs usually late in life (with the exception of perinatal membranous nephropathy) and why the time of onset is so variable. As in the instances of many autoimmune disorders, infections have been suggested as

potential triggering factors since the first clinical descriptions of membranous nephropathy (mentioning of a history of upper respiratory tract disease one to three weeks before the onset of symptoms of the nephrotic syndrome [546]).
The delineation of the epitope (which part of the sequence of PLA2R is the 10-15 aminoacid long peptide that will be recognized as non-self by the immune system? Is the epitope native or is it 'exposed' in the course of time by triggering conditions? (the latter hypothesis being more consistent with a solution to the problem of the late / variable age of onset of the disease). The nature of this condition can be speculated upon (given also what is known from the etiology of secondary membranous nephropathy).
Metal exposure (environmental, accidental or therapeutic) has been involved in the pathogenesis of the nephrotic syndrome. In the literature there are descriptions of instances where glomerulonephridites seem to have been induced by exposure to: cadmium and nickel [547], bismuth [548] (used in arthritis therapy), gold [549] (also used in arthritis therapy), mercury [533-535] (used in lues therapy, skin lightening compounds etc.). More specifically, mercury has been suspected to be a trigger in the generation of anti-laminin antibodies [550].
Furthermore, smoking has been suggested as a cause of rheumatoid arthritis [551], hypothesis which has led to a model of the etiology of RA in which citrullination of peptides at the level of the lung of the smoker leads to the loss of tolerance and triggers an HLA restricted immune response against what has become an autoantigen [552]. This hypothesis found recent confirmation by gene environment interaction enquiries in the context of genome wide association studies [553].
Food antigens have been implied as a possible cause for 'some cases of human glomerulonephritis' based on anecdotal evidence [554].

What is the mechanism of the lesion: IgG4 does not lead to complement activation; could direct toxicity of the antibody (neutralizing enzymatic or non-enzymatic activity, stimulating receptor-related pathways) [156] be an explanation? Alternative lectin pathway activation [555,556]? What would then the role of MBL (Mannose binding lectin) be? Could the similarity between PLA2R MBL play a role? Can the immune complexes be completely uncoupled from the damage [557]?

The question remains as to what the mechanism of the gender influence in membranous nephropathy is. A related unanswered question is why the prevalence of membranous nephropathy is higher in males, which is the opposite from the usual higher prevalence of autoimmune disorders in females.

Other questions relate to the putative functional link between PLA2R, AKR1B1 and SOD2, if one exists. It has been shown that PLA2R can regulate senescence in a reactive oxygen species – DNA damage – p53 dependent manner [558]. Given that AKR1B1 and SOD2 are important elements of the reactive oxygen species protection mechanisms, 'pathogenic senescence' could be hypothesised to be this 'missing link'.

# 5 Conclusions

The most effective treatment for a disorder is the one that addresses its cause (etiologic treatment) or its mechanism (pathogenetic treatment) directly. This implies that effective treatment relies on the understanding of the usually complex interplay between genes and environment which makes up the specific phenotype (disease) dealt with. In a simplified model, health state is the result of an equilibrium between the challenges of the environment and the homeostatic mechanisms of the organism. These mechanisms can be conceptually dissected into 'unit steps of homeostasis' [559]. The variability of the unit steps (proteins), which allows for an optimal adaptation to the environment (in the case of homeostasis) is encoded at the genetic level. The genetic variants (alleles) which lead to disease susceptibility (the effect size of it being a gradient between all and nothing) can be identified by applying genetic mapping techniques.

The classical gene mapping approach to a rare disorder is linkage analysis. The power of this approach is limited by the small number of informative families available. Therefore there are usually no twin studies for the estimation of heritability, but the usefulness of heritability estimates is considered to be limited [560]. We proved in our study that a rare, well defined phenotype can be successfully mapped by applying the genome wide association method on relatively small cohorts of patients. We were able to show that in 556 (398 male) cases of IMN of Caucasoid descent, the disorder is strongly associated with risk alleles within the HLA locus in general and with HLA-DQA1 in particular, as well as with PLA2R1 alleles on chromosome 2. No other autosomal allele was significantly associated with IMN, within the technical limitations of our study (power and coverage).

The conceptual limitations of association studies pertain mostly to the issues of causality and confounders. Association is not proof for causality, but strength of association (quantified in high $-\log_{10}P$ values) and consistency thereof (replication in two Caucasoid populations – FR, NL – similar to the exploratory one – UK) are suggestive for it. Confidence in the strength of association is enhanced by the biological plausability of these findings. In our case, there are independent studies suggesting the importance of the PLA2R1 molecule and the

HLA region in the etiopathogeny of IMN. We suggest a pathogenic mechanism in which the antibody response in IMN is related to presentation of a "risk" PLA2R1 epitope by an "IMN" HLA-DQA1 allele.

Further confounders of the relation between association and causality that we encountered in our study are: the issue of independent associations, epistasis, environmental influences (including the internal environment differences which lead to different susceptibilities in males *versus* females), linkage disequilibrium and copy number variants.

Ultimately solving the etiopathogenical problem in the case of IMN will rely on further studies to assess sequence variations in the HLA-DQA1 and the PLA2R1 regions and ulterior functional experiments performed to assess the biological relevance of the identified variations.

# 6 Bibliography


[1] L. H. Beck, Jr., R. G. Bonegio, G. Lambeau, D. M. Beck, D. W. Powell, T. D. Cummins, J. B. Klein, and D. J. Salant, M-type phospholipase A2 receptor as target antigen in idiopathic membranous nephropathy, N. Engl. J Med., 361 (2009) 11-21.

[2] J. A. Armstrong, Urinalysis in Western culture: a brief history, Kidney Int., 71 (2007) 384-387.

[3] Sanders L.J., From Thebes to Toronto and the 21st Century: An Incredible Journey, Diabetes Spectrum, (2002).

[4] F. N. ALLAN, The writings of Thomas Willis, M.D.; diabetes three hundred years ago, Diabetes, 2 (1953) 74-77.

[5] L. L. FRANK, Diabetes mellitus in the texts of old Hindu medicine (Charaka, Susruta, Vagbhata), Am J Gastroenterol., 27 (1957) 76-95.

[6] C. J. Diskin, T. J. Stokes, L. M. Dansby, T. B. Carter, and L. Radcliff, Surface tension, proteinuria, and the urine bubbles of Hippocrates, Lancet, 355 (2000) 901-902.

[7] Hippocrates. Aphorisms. Classics MIT. <http://classics.mit.edu/Hippocrates/aphorisms.mb.txt> . 2009. 5-10-2010.
Ref Type: Electronic Citation

[8] V. Bonomini, C. Campieri, M. Zuccoli, and G. Cristofolini, Guilielmus de Saliceto and his contributions to renal medicine, Am J Nephrol., 17 (1997) 274-281.

[9] J. S. Cameron, Milk or albumin? The history of proteinuria before Richard Bright, Nephrol. Dial. Transplant., 18 (2003) 1281-1285.

[10] Garrison F H, Morton L T, and Norman J M, Diseases of the Genito-Urinary System: Kidney, in: Jeremy M.Norman (Ed.), Morton's Medical Bibliography, Scolar Press, 1991, pp. 649-657.

[11] J. S. Cameron, Bright's disease today: the pathogenesis and treatment of glomerulonephritis--I, Br. Med. J, 4 (1972) 87-90.

[12] J. Brodehl, [The changing meaning of the term nephrosis since Noeggerath (author's transl)], Klin. Padiatr., 189 (1977) 199-206.

[13] J. S. Cameron and J. Hicks, The origins and development of the concept of a "nephrotic syndrome", Am J Nephrol., 22 (2002) 240-247.

[14] S. E. BRADLEY and C. J. TYSON, The nephrotic syndrome, N. Engl. J Med., 238 (1948) 260-266.

[15] V. E. POLLAK, S. Rosen, C. L. PIRANI, R. C. MUEHRCKE, and R. M. KARK, Natural history of lipoid nephrosis and of membranous glomerulonephritis, Ann. Intern. Med, 69 (1968) 1171-1196.

[16] L. McGregor, The Finer Histology of the Normal Glomerulus, Am J Pathol., 5 (1929) 545-558.

[17] E. T. BELL, Lipoid Nephrosis, Am J Pathol., 5 (1929) 587-622.

[18] E. T. BELL, A clinical and pathological study of subacute and chronic glomerulonephritis, including lipoid nephrosis, Am J Pathol., 14 (1938) 691-736.

[19] A. Ellis, NATURAL HISTORY OF BRIGHT'S DISEASE : CLINICAL, HISTOLOGICAL AND EXPERIMENTAL OBSERVATIONS, The Lancet, 239 (1942) 1-7.

[20] R. J. Glassock, The pathogenesis of idiopathic membranous nephropathy: a 50-year odyssey, Am J Kidney Dis., 56 (2010) 157-167.

[21] H. E. MACMAHON, C. G. CHILD, III, and J. F. PATTERSON, Healed infectious hepatitis, portal hypertension, membranous glomerulonephritis, Bull. Tufts. N. Engl. Med. Cent., 1 (1955) 184-192.

[22] D. B. JONES, Nephrotic glomerulonephritis, Am J Pathol., 33 (1957) 313-329.

[23] B. R. Farrow and C. R. Huxtable, Membranous nephropathy and the nephrotic syndrome in the cat, J Comp Pathol., 81 (1971) 463-467.

[24] A. H. COONS and M. H. KAPLAN, Localization of antigen in tissue cells; improvements in a method for the detection of antigen by means of fluorescent antibody, J Exp. Med, 91 (1950) 1-13.

[25] D. C. PEASE and R. F. BAKER, Electron microscopy of the kidney, Am J Anat., 87 (1950) 349-389.

[26] T. Ehrenreich and J. Churg, Pathology of Membranous Nephropathy, Pathol. Annu., 3 (1968) 145-186.

[27] B. Todd R and Bowman W, The Physiological Anatomy and Physiology of Man, Blanchard and Lea, Philadelphia 1857.

[28] Couser W G, Membranous Nephropathy, in: Feehally J, Floege J, and Johnson R J (Eds.), Comprehensive Clinical Nephrology, Mosby Elsevier, 2007, pp. 231-242.

[29] E. Machuca, G. Benoit, and C. Antignac, Genetics of nephrotic syndrome: connecting molecular genetics to podocyte physiology, Hum Mol. Genet, 18 (2009) R185-R194.

[30] C. Ponticelli, Membranous nephropathy, J Nephrol., 20 (2007) 268-287.

[31] H. Z. MOVAT and D. D. McGREGOR, The fine structure of the glomerulus in membranous glomerulonephritis (lipoid nephrosis) in adults, Am J Clin. Pathol., 32 (1959) 109-127.

[32] R. C. MELLORS, L. G. ORTEGA, and H. R. HOLMAN, Role of gamma globulins in pathogenesis of renal lesions in systemic lupus erythematosus and chronic membranous glomerulonephritis, with an observation on the lupus erythematosus cell reaction, J Exp. Med, 106 (1957) 191-202.

[33] J. Churg and T. Ehrenreich, Membranous nephropathy, Perspect. Nephrol. Hypertens., 1 Pt 1 (1973) 443-448.

[34] F. Ferrario and M. P. Rastaldi, Histopathological atlas of renal diseases: membranous glomerulonephritis, J Nephrol., 17 (2004) 635-636.

[35] T. Ehrenreich, J. G. PORUSH, J. Churg, L. Garfinkel, S. Glabman, M. H. Goldstein, E. GRISHMAN, and S. L. YUNIS, Treatment of idiopathic membranous nephropathy, N Engl J Med, 295 (1976) 741-746.

[36] J. Kowalewska, K. D. Smith, K. L. Hudkins, A. Chang, A. B. Fogo, D. Houghton, D. Leslie, J. Aitchison, R. F. Nicosia, and C. E. Alpers, Membranous glomerulopathy with spherules: an uncommon variant with obscure pathogenesis, Am J Kidney Dis., 47 (2006) 983-992.

[37] M. Obana, K. Nakanishi, M. Sako, N. Yata, K. Nozu, R. Tanaka, K. Iijima, and N. Yoshikawa, Segmental membranous glomerulonephritis in children: comparison with global membranous glomerulonephritis, Clin. J Am Soc. Nephrol., 1 (2006) 723-729.

[38] Y. Segawa, S. Hisano, M. Matsushita, T. Fujita, S. Hirose, M. Takeshita, and H. Iwasaki, IgG subclasses and complement pathway in segmental and global membranous nephropathy, Pediatr. Nephrol., (2010).

[39] B. Ivanyi, I. Haszon, E. Endreffy, P. Szenohradszky, I. B. Petri, T. Kalmar, R. J. Butkowski, A. S. Charonis, and S. Turi, Childhood membranous nephropathy, circulating antibodies to the 58-kD TIN antigen, and anti-tubular basement membrane nephritis: an 11-year follow-up, Am J Kidney Dis., 32 (1998) 1068-1074.

[40] L. H. Beck, Jr. and D. J. Salant, Membranous nephropathy: recent travels and new roads ahead, Kidney Int., (2010).

[41] K. Brukamp, A. M. Doyle, R. D. Bloom, N. Bunin, J. E. Tomaszewski, and B. Cizman, Nephrotic syndrome after hematopoietic cell

transplantation: do glomerular lesions represent renal graft-versus-host disease?, Clin. J Am Soc. Nephrol., 1 (2006) 685-694.

[42] S. O'Regan, M. Smith, and K. N. Drummond, Antigens in human immune complex nephritis, Clin. Nephrol., 6 (1976) 417-421.

[43] F. Milgrom, W. A. Campbell, and G. A. Andres, Antigen in immune complex nephritis. V. Recovery and identification by gel precipitation, Immunology, 30 (1976) 277-280.

[44] H. Debiec, V. Guigonis, B. Mougenot, F. Decobert, J. P. Haymann, A. Bensman, G. Deschenes, and P. M. Ronco, Antenatal membranous glomerulonephritis due to anti-neutral endopeptidase antibodies, N. Engl. J Med., 346 (2002) 2053-2060.

[45] H. Debiec, J. Nauta, F. Coulet, B. M. van der, V. Guigonis, T. Schurmans, H. E. de, F. Soubrier, F. Janssen, and P. Ronco, Role of truncating mutations in MME gene in fetomaternal alloimmunisation and antenatal glomerulopathies, Lancet, 364 (2004) 1252-1259.

[46] Schick B., Die Nachkrankheiten des Scharlach, Jb. Kinderheilk., 65 (1907) 132-173.

[47] K. Lange, M. M. Gold, D. Weiner, and V. Simon, AUTOANTIBODIES IN HUMAN GLOMERULONEPHRITIS, J Clin. Invest, 28 (1949) 50-55.

[48] W. HEYMANN, D. B. HACKEL, S. HARWOOD, S. G. WILSON, and J. L. HUNTER, Production of nephrotic syndrome in rats by Freund's adjuvants and rat kidney suspensions, Proc. Soc. Exp. Biol. Med., 100 (1959) 660-664.

[49] D. Kerjaschki and T. J. Neale, Molecular mechanisms of glomerular injury in rat experimental membranous nephropathy (Heymann nephritis), J Am Soc. Nephrol., 7 (1996) 2518-2526.

[50] R. A. Orlando, D. Kerjaschki, H. Kurihara, D. Biemesderfer, and M. G. Farquhar, gp330 associates with a 44-kDa protein in the rat kidney to form the Heymann nephritis antigenic complex, Proc. Natl. Acad. Sci. U. S. A, 89 (1992) 6698-6702.

[51] M. G. Farquhar, D. Kerjaschki, M. Lundstrom, and R. A. Orlando, gp330 and RAP: the Heymann nephritis antigenic complex, Ann. N Y. Acad. Sci., 737 (1994) 96-113.

[52] D. Kerjaschki and M. G. Farquhar, The pathogenic antigen of Heymann nephritis is a membrane glycoprotein of the renal proximal tubule brush border, Proc. Natl. Acad. Sci. U. S. A, 79 (1982) 5557-5561.

[53] A. Saito, S. Pietromonaco, A. K. Loo, and M. G. Farquhar, Complete cloning and sequencing of rat gp330/"megalin," a distinctive member of

the low density lipoprotein receptor gene family, Proc. Natl. Acad. Sci. U. S. A, 91 (1994) 9725-9729.

[54] D. Biemesderfer, G. Dekan, P. S. Aronson, and M. G. Farquhar, Biosynthesis of the gp330/44-kDa Heymann nephritis antigenic complex: assembly takes place in the ER, Am J Physiol, 264 (1993) F1011-F1020.

[55] R. A. Orlando and M. G. Farquhar, Cellular trafficking of megalin (gp330) and the receptor associated protein (RAP), *The American Society for Cell Biology 34th annual meeting. San Francisco, California, December 10-14, 1994,* 1994, p. 187a.

[56] D. Kerjaschki, R. Ullrich, M. Exner, R. A. Orlando, and M. G. Farquhar, Induction of passive Heymann nephritis with antibodies specific for a synthetic peptide derived from the receptor-associated protein, J Exp. Med, 183 (1996) 2007-2015.

[57] H. Yamazaki, R. Ullrich, M. Exner, A. Saito, R. A. Orlando, D. Kerjaschki, and M. G. Farquhar, All four putative ligand-binding domains in megalin contain pathogenic epitopes capable of inducing passive Heymann nephritis, J Am Soc. Nephrol., 9 (1998) 1638-1644.

[58] D. Kerjaschki, A. Miettinen, and M. G. Farquhar, Initial events in the formation of immune deposits in passive Heymann nephritis. gp330-anti-gp330 immune complexes form in epithelial coated pits and rapidly become attached to the glomerular basement membrane, J Exp. Med, 166 (1987) 109-128.

[59] D. J. Salant, S. Belok, M. P. Madaio, and W. G. Couser, A new role for complement in experimental membranous nephropathy in rats, J Clin. Invest, 66 (1980) 1339-1350.

[60] D. Kerjaschki, M. Schulze, S. Binder, R. Kain, P. P. Ojha, M. Susani, R. Horvat, P. J. Baker, and W. G. Couser, Transcellular transport and membrane insertion of the C5b-9 membrane attack complex of complement by glomerular epithelial cells in experimental membranous nephropathy, J Immunol., 143 (1989) 546-552.

[61] A. V. Cybulsky, R. J. Quigg, and D. J. Salant, Experimental membranous nephropathy redux, Am J Physiol Renal Physiol, 289 (2005) F660-F671.

[62] A. V. Cybulsky, R. J. Quigg, and D. J. Salant, The membrane attack complex in complement-mediated glomerular epithelial cell injury: formation and stability of C5b-9 and C5b-7 in rat membranous nephropathy, J Immunol., 137 (1986) 1511-1516.

[63] B. P. Morgan, Effects of the membrane attack complex of complement on nucleated cells, Curr. Top. Microbiol. Immunol., 178 (1992) 115-140.

[64] K. J. Assmann, J. P. Van Son, H. B. Dijkman, and R. A. Koene, A nephritogenic rat monoclonal antibody to mouse aminopeptidase A.

Induction of massive albuminuria after a single intravenous injection, J Exp. Med, 175 (1992) 623-635.

[65] P. Ronco, L. Allegri, E. Brianti, F. Chatelet, E. H. Van Leer, and P. Verroust, Antigenic targets in epimembranous glomerulonephritis. Experimental data and potential application in human pathology, Appl. Pathol., 7 (1989) 85-98.

[66] W. G. Couser, D. R. Steinmuller, M. M. Stilmant, D. J. Salant, and L. M. Lowenstein, Experimental glomerulonephritis in the isolated perfused rat kidney, J Clin. Invest, 62 (1978) 1275-1287.

[67] P. Ronco and H. Debiec, Target antigens and nephritogenic antibodies in membranous nephropathy: of rats and men, Semin. Immunopathol., 29 (2007) 445-458.

[68] J. S. Van der Zee and R. C. Aalberse, The role of IgG in immediate-type hypersensitivity, Eur. Respir. J Suppl, 13 (1991) 91s-96s.

[69] R. Tran-Paterson, H. F. Willard, and M. Letarte, The common acute lymphoblastic leukemia antigen (neutral endopeptidase-3.4.24.11) gene is located on human chromosome 3, Cancer Genet Cytogenet., 42 (1989) 129-134.

[70] E. A. Bruford, M. J. Lush, M. W. Wright, T. P. Sneddon, S. Povey, and E. Birney, The HGNC Database in 2008: a resource for the human genome, Nucleic Acids Res., 36 (2008) D445-D448.

[71] M. A. Shipp, G. B. Stefano, S. N. Switzer, J. D. Griffin, and E. L. Reinherz, CD10 (CALLA)/neutral endopeptidase 24.11 modulates inflammatory peptide-induced changes in neutrophil morphology, migration, and adhesion proteins and is itself regulated by neutrophil activation, Blood, 78 (1991) 1834-1841.

[72] L. East and C. M. Isacke, The mannose receptor family, Biochim. Biophys. Acta, 1572 (2002) 364-386.

[73] A. N. Zelensky and J. E. Gready, The C-type lectin-like domain superfamily, FEBS J, 272 (2005) 6179-6217.

[74] Y. Kumagai and S. Akira, Identification and functions of pattern-recognition receptors, J Allergy Clin. Immunol., 125 (2010) 985-992.

[75] J. J. Garcia-Vallejo and K. Y. van, Endogenous ligands for C-type lectin receptors: the true regulators of immune homeostasis, Immunol. Rev., 230 (2009) 22-37.

[76] K. Rohmann, T. Tschernig, R. Pabst, T. Goldmann, and D. Dromann, Innate immunity in the human lung: pathogen recognition and lung disease, Cell Tissue Res., 343 (2010) 167-174.

[77] Varki A, Etzler ME, Cummings R.D., and Esko JD, Plant Lectins, in: Varki A, Cummings R, Esko J, Freeze H, Stanley P, Bertozzi CR, Hart G, and Etzler ME (Eds.), Essentials of Glycobiology, Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY), 2009.

[78] Drickamer K. A genomics resource for animal lectins. <http://www.imperial.ac.uk/research/animallectins/default.html> . 3-10-2006. Imperial College. 4-4-2011.
Ref Type: Electronic Citation

[79] H. Tateno, SUEL-related lectins, a lectin family widely distributed throughout organisms, Biosci. Biotechnol. Biochem., 74 (2010) 1141-1144.

[80] H. Gewurz, X. H. Zhang, and T. F. Lint, Structure and function of the pentraxins, Curr. Opin. Immunol., 7 (1995) 54-64.

[81] D. C. Kilpatrick, Animal lectins: a historical introduction and overview, Biochim. Biophys. Acta, 1572 (2002) 187-197.

[82] E. P. McGreal, L. Martinez-Pomares, and S. Gordon, Divergent roles for C-type lectins expressed by cells of the innate immune system, Mol. Immunol., 41 (2004) 1109-1121.

[83] H. Li, A. Coghlan, J. Ruan, L. J. Coin, J. K. Heriche, L. Osmotherly, R. Li, T. Liu, Z. Zhang, L. Bolund, G. K. Wong, W. Zheng, P. Dehal, J. Wang, and R. Durbin, TreeFam: a curated database of phylogenetic trees of animal gene families, Nucleic Acids Res., 34 (2006) D572-D580.

[84] J. Ruan, H. Li, Z. Chen, A. Coghlan, L. J. Coin, Y. Guo, J. K. Heriche, Y. Hu, K. Kristiansen, R. Li, T. Liu, A. Moses, J. Qin, S. Vang, A. J. Vilella, A. Ureta-Vidal, L. Bolund, J. Wang, and R. Durbin, TreeFam: 2008 Update, Nucleic Acids Res., 36 (2008) D735-D740.

[85] J. Boskovic, J. N. Arnold, R. Stilion, S. Gordon, R. B. Sim, A. Rivera-Calzada, D. Wienke, C. M. Isacke, L. Martinez-Pomares, and O. Llorca, Structural model for the mannose receptor family uncovered by electron microscopy of Endo180 and the mannose receptor, J Biol. Chem., 281 (2006) 8780-8787.

[86] K. Hanasaki, Mammalian phospholipase A2: phospholipase A2 receptor, Biol. Pharm. Bull., 27 (2004) 1165-1167.

[87] J. Balsinde, M. V. Winstead, and E. A. Dennis, Phospholipase A(2) regulation of arachidonic acid mobilization, FEBS Lett., 531 (2002) 2-6.

[88] J. E. Burke and E. A. Dennis, Phospholipase A2 structure/function, mechanism, and signaling, J Lipid Res., 50 Suppl (2009) S237-S242.

[89] G. Lambeau and M. H. Gelb, Biochemistry and physiology of mammalian secreted phospholipases A2, Annu. Rev. Biochem., 77 (2008) 495-520.

[90] R. H. Schaloske and E. A. Dennis, The phospholipase A2 superfamily and its group numbering system, Biochim. Biophys. Acta, 1761 (2006) 1246-1259.

[91] E. Habermann, Bee and wasp venoms, Science, 177 (1972) 314-322.

[92] C. Montecucco, J. M. Gutierrez, and B. Lomonte, Cellular pathology induced by snake venom phospholipase A2 myotoxins and neurotoxins: common aspects of their mechanisms of action, Cell Mol. Life Sci., 65 (2008) 2897-2912.

[93] A. K. Sobotka, R. M. Franklin, N. F. Adkinson, Jr., M. Valentine, H. Baer, and L. M. Lichtenstein, Allergy to insect stings. II. Phospholipase A: the major allergen in honeybee venom, J Allergy Clin. Immunol., 57 (1976) 29-40.

[94] M. Menschikowski, A. Hagelgans, and G. Siegert, Secretory phospholipase A2 of group IIA: is it an offensive or a defensive player during atherosclerosis and other inflammatory diseases?, Prostaglandins Other Lipid Mediat., 79 (2006) 1-33.

[95] K. Hanasaki and H. Arita, Phospholipase A2 receptor: a regulator of biological functions of secretory phospholipase A2, Prostaglandins Other Lipid Mediat., 68-69 (2002) 71-82.

[96] P. Elsbach, J. Weiss, R. C. Franson, S. Beckerdite-Quagliata, A. Schneider, and L. Harris, Separation and purification of a potent bactericidal/permeability-increasing protein and a closely associated phospholipase A2 from rabbit polymorphonuclear leukocytes. Observations on their relationship, J Biol. Chem., 254 (1979) 11000-11009.

[97] S. S. Harwig, L. Tan, X. D. Qu, Y. Cho, P. B. Eisenhauer, and R. I. Lehrer, Bactericidal properties of murine intestinal phospholipase A2, J Clin. Invest, 95 (1995) 603-610.

[98] X. D. Qu and R. I. Lehrer, Secretory phospholipase A2 is the principal bactericide for staphylococci and other gram-positive bacteria in human tears, Infect. Immun., 66 (1998) 2791-2797.

[99] T. Koprivnjak, A. Peschel, M. H. Gelb, N. S. Liang, and J. P. Weiss, Role of charge properties of bacterial envelope in bactericidal action of human group IIA phospholipase A2 against Staphylococcus aureus, J Biol. Chem., 277 (2002) 47636-47644.

[100] G. Lambeau and M. Lazdunski, Receptors for a growing family of secreted phospholipases A2, Trends Pharmacol. Sci., 20 (1999) 162-170.

[101] G. Lambeau, P. Ancian, J. Barhanin, and M. Lazdunski, Cloning and expression of a membrane receptor for secretory phospholipases A2, J Biol. Chem., 269 (1994) 1575-1578.

[102] S. Beck, G. Beck, T. Ostendorf, J. Floege, G. Lambeau, T. Nevalainen, H. H. Radeke, S. Gurrieri, U. Haas, B. Thorwart, J. Pfeilschifter, and M. Kaszkin, Upregulation of group IB secreted phospholipase A(2) and its M-type receptor in rat ANTI-THY-1 glomerulonephritis, Kidney Int., 70 (2006) 1251-1260.

[103] M. Rouault, C. C. Le, E. Boilard, F. Surrel, A. Singer, F. Ghomashchi, S. Bezzine, S. Scarzello, J. Bollinger, M. H. Gelb, and G. Lambeau, Recombinant production and properties of binding of the full set of mouse secreted phospholipases A2 to the mouse M-type receptor, Biochemistry, 46 (2007) 1647-1662.

[104] E. Valentin and G. Lambeau, What can venom phospholipases A(2) tell us about the functional diversity of mammalian secreted phospholipases A(2)?, Biochimie, 82 (2000) 815-831.

[105] E. Valentin and G. Lambeau, Increasing molecular diversity of secreted phospholipases A(2) and their receptors and binding proteins, Biochim. Biophys. Acta, 1488 (2000) 59-70.

[106] A. B. Mukherjee, G. C. Kundu, A. K. Mandal, N. Pattabiraman, C. J. Yuan, and Z. Zhang, Uteroglobin: physiological role in normal glomerular function uncovered by targeted disruption of the uteroglobin gene in mice, Am J Kidney Dis., 32 (1998) 1106-1120.

[107] C. N. Birts, C. H. Barton, and D. C. Wilton, A catalytically independent physiological function for human acute phase protein group IIA phospholipase A2: cellular uptake facilitates cell debris removal, J Biol. Chem., 283 (2008) 5034-5045.

[108] W. Pruzanski, P. Vadas, E. Stefanski, and M. B. Urowitz, Phospholipase A2 activity in sera and synovial fluids in rheumatoid arthritis and osteoarthritis. Its possible role as a proinflammatory enzyme, J Rheumatol., 12 (1985) 211-216.

[109] R. M. Crowl, T. J. Stoller, R. R. Conroy, and C. R. Stoner, Induction of phospholipase A2 gene expression in human hepatoma cells by mediators of the acute phase response, J Biol. Chem., 266 (1991) 2647-2651.

[110] G. Atsumi, M. Murakami, M. Tajima, S. Shimbara, N. Hara, and I. Kudo, The perturbed membrane of cells undergoing apoptosis is susceptible to type II secretory phospholipase A2 to liberate arachidonic acid, Biochim. Biophys. Acta, 1349 (1997) 43-54.

[111] C. E. Hack, G. J. Wolbink, C. Schalkwijk, H. Speijer, W. T. Hermens, and B. H. van den, A role for secretory phospholipase A2 and C-reactive protein in the removal of injured cells, Immunol. Today, 18 (1997) 111-115.

[112] R. S. Krishnan and J. C. Daniel, Jr., "Blastokinin": inducer and regulator of blastocyst development in the rabbit uterus, Science, 158 (1967) 490-492.

[113] J. Klug, H. M. Beier, A. Bernard, B. S. Chilton, T. P. Fleming, R. I. Lehrer, L. Miele, N. Pattabiraman, and G. Singh, Uteroglobin/Clara cell 10-kDa family of proteins: nomenclature committee report, Ann. N Y. Acad. Sci., 923 (2000) 348-354.

[114] H. M. Beier, Uteroglobin: a hormone-sensitive endometrial protein involved in blastocyst development, Biochim. Biophys. Acta, 160 (1968) 289-291.

[115] I. Dierynck, A. Bernard, H. Roels, and L. M. De, The human Clara cell protein: biochemical and biological characterisation of a natural immunosuppressor, Mult. Scler., 1 (1996) 385-387.

[116] A. B. Mukherjee, E. Cordella-Miele, T. Kikukawa, and L. Miele, Modulation of cellular response to antigens by uteroglobin and transglutaminase, Adv. Exp. Med Biol., 231 (1988) 135-152.

[117] L. Miele, E. Cordella-Miele, A. Facchiano, and A. B. Mukherjee, Novel anti-inflammatory peptides from the region of highest similarity between uteroglobin and lipocortin I, Nature, 335 (1988) 726-730.

[118] A. Peri, E. Cordella-Miele, L. Miele, and A. B. Mukherjee, Tissue-specific expression of the gene coding for human Clara cell 10-kD protein, a phospholipase A2-inhibitory protein, J Clin. Invest, 92 (1993) 2099-2109.

[119] P. J. Jackson, R. Turner, J. N. Keen, R. A. Brooksbank, and E. H. Cooper, Purification and partial amino acid sequence of human urine protein 1. Evidence for homology with rabbit uteroglobin, J Chromatogr., 452 (1988) 359-367.

[120] I. Morize, E. Surcouf, M. C. Vaney, Y. Epelboin, M. Buehner, F. Fridlansky, E. Milgrom, and J. P. Mornon, Refinement of the C222(1) crystal form of oxidized uteroglobin at 1.34 A resolution, J Mol. Biol., 194 (1987) 725-739.

[121] S. W. Levin, J. D. Butler, U. K. Schumacher, P. D. Wightman, and A. B. Mukherjee, Uteroglobin inhibits phospholipase A2 activity, Life Sci., 38 (1986) 1813-1819.

[122] A. M. Bernard, R. R. Lauwerys, A. Noel, B. Vandeleene, and A. Lambert, Urine protein 1: a sex-dependent marker of tubular or glomerular dysfunction, Clin. Chem., 35 (1989) 2141-2142.

[123] Z. Zhang, G. C. Kundu, C. J. Yuan, J. M. Ward, E. J. Lee, F. DeMayo, H. Westphal, and A. B. Mukherjee, Severe fibronectin-deposit renal glomerular disease in mice lacking uteroglobin, Science, 276 (1997) 1408-1412.

[124] E. Ruoslahti, Fibronectin and its receptors, Annu. Rev. Biochem., 57 (1988) 375-413.

[125] E. H. Strom, G. Banfi, R. Krapf, A. B. Abt, G. Mazzucco, G. Monga, F. Gloor, J. Neuweiler, R. Riess, P. Stosiek, and ., Glomerulopathy associated with predominant fibronectin deposits: a newly recognized hereditary disease, Kidney Int., 48 (1995) 163-170.

[126] F. Castelletti, R. Donadelli, F. Banterla, F. Hildebrandt, P. F. Zipfel, E. Bresin, E. Otto, C. Skerka, A. Renieri, M. Todeschini, J. Caprioli, R. M. Caruso, R. Artuso, G. Remuzzi, and M. Noris, Mutations in FN1 cause glomerulopathy with fibronectin deposits, Proc. Natl. Acad. Sci. U. S. A, 105 (2008) 2538-2543.

[127] M. Vollmer, R. Krapf, and F. Hildebrandt, Exclusion of the uteroglobin gene as a candidate for fibronectin glomerulopathy (GFND), Nephrol. Dial. Transplant., 13 (1998) 2417-2418.

[128] S. M. Korbet, M. M. Schwartz, and E. J. Lewis, The fibrillary glomerulopathies, Am J Kidney Dis., 23 (1994) 751-765.

[129] F. Zheng, G. C. Kundu, Z. Zhang, J. Ward, F. DeMayo, and A. B. Mukherjee, Uteroglobin is essential in preventing immunoglobulin A nephropathy in mice, Nat. Med, 5 (1999) 1018-1025.

[130] E. Menegatti, A. Nardacchione, M. Alpa, C. Agnes, D. Rossi, M. Chiara, V. Modena, L. M. Sena, and D. Roccatello, Polymorphism of the uteroglobin gene in systemic lupus erythematosus and IgA nephropathy, Lab Invest, 82 (2002) 543-546.

[131] R. Coppo, M. Chiesa, P. Cirina, L. Peruzzi, and A. Amore, In human IgA nephropathy uteroglobin does not play the role inferred from transgenic mice, Am J Kidney Dis., 40 (2002) 495-503.

[132] J. C. Davin, V. M. Li, J. Nagy, J. M. Foidart, J. B. Foidart, S. G. Barbagallo, M. Malaise, and P. Mahieu, Evidence that the interaction between circulating IgA and fibronectin is a normal process enhanced in primary IgA nephropathy, J Clin. Immunol., 11 (1991) 78-94.

[133] B. Chowdhury, Z. Zhang, and A. B. Mukherjee, Uteroglobin interacts with the heparin-binding site of fibronectin and prevents fibronectin-IgA complex formation found in IgA-nephropathy, FEBS Lett., 582 (2008) 611-615.

[134] K. K. Higashino, Y. Yokota, T. Ono, S. Kamitani, H. Arita, and K. Hanasaki, Identification of a soluble form phospholipase A2 receptor as a circulating endogenous inhibitor for secretory phospholipase A2, J Biol. Chem., 277 (2002) 13583-13588.

[135] L. Martinez-Pomares, J. A. Mahoney, R. Kaposzta, S. A. Linehan, P. D. Stahl, and S. Gordon, A functional soluble form of the murine mannose

receptor is produced by macrophages in vitro and is present in mouse serum, J Biol. Chem., 273 (1998) 23376-23380.

[136] O. Ohara, J. Ishizaki, and H. Arita, Structure and function of phospholipase A2 receptor, Prog. Lipid Res., 34 (1995) 117-138.

[137] P. Ancian, G. Lambeau, M. G. Mattei, and M. Lazdunski, The human 180-kDa receptor for secretory phospholipases A2. Molecular cloning, identification of a secreted soluble form, expression, and chromosomal localization, J Biol. Chem., 270 (1995) 8963-8970.

[138] E. Boilard, M. Rouault, F. Surrel, C. C. Le, S. Bezzine, A. Singer, M. H. Gelb, and G. Lambeau, Secreted phospholipase A2 inhibitors are also potent blockers of binding to the M-type receptor, Biochemistry, 45 (2006) 13203-13218.

[139] H. Fujita, K. Kawamoto, K. Hanasaki, and H. Arita, Glycosylation-dependent binding of pancreatic type I phospholipase A2 to its specific receptor, Biochem. Biophys. Res. Commun., 209 (1995) 293-299.

[140] K. Hanasaki, T. Ono, A. Saiga, Y. Morioka, M. Ikeda, K. Kawamoto, K. Higashino, K. Nakano, K. Yamada, J. Ishizaki, and H. Arita, Purified group X secretory phospholipase A(2) induced prominent release of arachidonic acid from human myeloid leukemia cells, J Biol. Chem., 274 (1999) 34203-34211.

[141] Y. Morioka, A. Saiga, Y. Yokota, N. Suzuki, M. Ikeda, T. Ono, K. Nakano, N. Fujii, J. Ishizaki, H. Arita, and K. Hanasaki, Mouse group X secretory phospholipase A2 induces a potent release of arachidonic acid from spleen cells and acts as a ligand for the phospholipase A2 receptor, Arch. Biochem. Biophys., 381 (2000) 31-42.

[142] T. Nakano, H. Fujita, N. Kikuchi, and H. Arita, Plasmin converts pro-form of group I phospholipase A2 into receptor binding, active forms, Biochem. Biophys. Res. Commun., 198 (1994) 10-15.

[143] K. Higashino, J. Ishizaki, J. Kishino, O. Ohara, and H. Arita, Structural comparison of phospholipase-A2-binding regions in phospholipase-A2 receptors from various mammals, Eur. J Biochem., 225 (1994) 375-382.

[144] A. B. Fisher, C. Dodia, A. Chander, M. F. Beers, and S. R. Bates, Inhibition of Trimeresurus flavoviridis phospholipase A2 by lung surfactant protein A (SP-A), Biochim. Biophys. Acta, 1211 (1994) 256-262.

[145] A. V. Cybulsky, D. J. Salant, R. J. Quigg, J. Badalamenti, and J. V. Bonventre, Complement C5b-9 complex activates phospholipases in glomerular epithelial cells, Am J Physiol, 257 (1989) F826-F836.

[146] M. Panesar, J. Papillon, A. J. McTavish, and A. V. Cybulsky, Activation of phospholipase A2 by complement C5b-9 in glomerular epithelial cells, J Immunol., 159 (1997) 3584-3594.

[147] A. V. Cybulsky, J. Papillon, and A. J. McTavish, Complement activates phospholipases and protein kinases in glomerular epithelial cells, Kidney Int., 54 (1998) 360-372.

[148] A. V. Cybulsky, T. Takano, J. Papillon, and A. J. McTavish, Complement-induced phospholipase A2 activation in experimental membranous nephropathy, Kidney Int., 57 (2000) 1052-1062.

[149] A. V. Cybulsky, J. C. Monge, J. Papillon, and A. J. McTavish, Complement C5b-9 activates cytosolic phospholipase A2 in glomerular epithelial cells, Am J Physiol, 269 (1995) F739-F749.

[150] A. V. Cybulsky, T. Takano, J. Papillon, A. Khadir, J. Liu, and H. Peng, Complement C5b-9 membrane attack complex increases expression of endoplasmic reticulum stress proteins in glomerular epithelial cells, J Biol. Chem., 277 (2002) 41342-41351.

[151] A. V. Cybulsky, T. Takano, J. Papillon, and K. Bijian, Role of the endoplasmic reticulum unfolded protein response in glomerular epithelial cell injury, J Biol. Chem., 280 (2005) 24396-24403.

[152] K. Bijian and A. V. Cybulsky, Stress proteins in glomerular epithelial cell injury, Contrib. Nephrol., 148 (2005) 8-20.

[153] H. Imai, K. Hamai, A. Komatsuda, H. Ohtani, and A. B. Miura, IgG subclasses in patients with membranoproliferative glomerulonephritis, membranous nephropathy, and lupus nephritis, Kidney Int., 51 (1997) 270-276.

[154] T. Doi, M. Mayumi, K. Kanatsu, F. Suehiro, and Y. Hamashima, Distribution of IgG subclasses in membranous nephropathy, Clin. Exp. Immunol., 58 (1984) 57-62.

[155] H. Wakui, H. Imai, A. Komatsuda, and A. B. Miura, Circulating antibodies against alpha-enolase in patients with primary membranous nephropathy (MN), Clin. Exp. Immunol., 118 (1999) 445-450.

[156] P. Ronco and H. Debiec, First Identification of an Antigen in Autoimmune Idiopathic Membranous Nephropathy: Toward Targeted Therapy?, Am J Kidney Dis., (2009).

[157] M. Prunotto, M. L. Carnevali, G. Candiano, C. Murtas, M. Bruschi, E. Corradini, A. Trivelli, A. Magnasco, A. Petretto, L. Santucci, S. Mattei, R. Gatti, F. Scolari, P. Kador, L. Allegri, and G. M. Ghiggeri, Autoimmunity in membranous nephropathy targets aldose reductase and SOD2, J Am Soc. Nephrol., 21 (2010) 507-519.

[158] S. Srivastava, M. Spite, J. O. Trent, M. B. West, Y. Ahmed, and A. Bhatnagar, Aldose reductase-catalyzed reduction of aldehyde phospholipids, J Biol. Chem., 279 (2004) 53395-53406.

[159] P. Ronco and H. Debiec, Pathophysiological lessons from rare associations of immunological disorders, Pediatr. Nephrol., 24 (2009) 3-8.

[160] N. Maruotti, A. Corrado, A. Gaudio, and F. P. Cantatore, Membranous nephropathy in rheumatoid arthritis: a case report, Clin. Exp. Rheumatol., 27 (2009) 840-842.

[161] D. Prasad, H. S. Khara, M. Gupta, and P. Sterman, Celiac disease associated membranous nephropathy - a rare cause or coincidence? A case report, Cases. J, 2 (2009) 7018.

[162] R. Gupta, A. Sharma, R. Arora, A. K. Dinda, A. Gupta, and S. C. Tiwari, Membranous glomerulonephritis in a patient with ankylosing spondylitis: a rare association, Clin. Exp. Nephrol., 13 (2009) 667-670.

[163] E. J. Soine, P. S. Nelson, and L. T. Nesbitt, Jr., A patient with coexistent bullous pemphigoid and primary membranous glomerulonephropathy, J Am Acad. Dermatol., 60 (2009) 1042-1044.

[164] R. G. Parry, J. C. Mason, and N. J. Marley, Membranous nephritis in a patient with scleroderma: a case report, Nephrol. Dial. Transplant., 11 (1996) 363-364.

[165] E. V. O'Loughlin, L. Robson, B. Scott, F. Alexander, and D. G. Gall, Membranous glomerulonephritis in a patient with Crohn's disease of the small bowel, J Pediatr. Gastroenterol. Nutr., 4 (1985) 135-139.

[166] K. Yoshida, J. Suzuki, K. Kume, S. Suzuki, M. Isome, K. Kato, and H. Suzuki, Membranous glomerulonephritis with primary Sjogren's syndrome detected by urine screening of school children, Clin. Nephrol., 45 (1996) 422.

[167] J. M. Pena Porta, I. J. Gonzalez, and C. V. de Vera Floristan, [Autoimmune thyroiditis, subclinical hypothyroidism, and nephrotic syndrome caused by membranous nephropathy], Nefrologia., 28 (2008) 572-573.

[168] J. C. Myers, T. A. Jones, E. R. Pohjolainen, A. S. Kadri, A. D. Goddard, D. Sheer, E. Solomon, and T. Pihlajaniemi, Molecular cloning of alpha 5(IV) collagen and assignment of the gene to the region of the X chromosome containing the Alport syndrome locus, Am J Hum Genet, 46 (1990) 1024-1033.

[169] D. F. Barker, S. L. Hostikka, J. Zhou, L. T. Chow, A. R. Oliphant, S. C. Gerken, M. C. Gregory, M. H. Skolnick, C. L. Atkin, and K. Tryggvason, Identification of mutations in the COL4A5 collagen gene in Alport syndrome, Science, 248 (1990) 1224-1227.

[170] A. Kefalides N and P. Borel J, Genetically Mediated and Acquired Basement Membrane Disorders, in: A. K. Nicholas (Ed.), Current Topics in Membranes

Basement Membranes: Cell and Molecular Biology, Academic Press, 2005, pp. 353-379.

[171] J. H. Miner, Renal basement membrane components, Kidney Int., 56 (1999) 2016-2024.

[172] C. E. Kashtan and Y. Segal, Genetic disorders of glomerular basement membranes, Nephron Clin. Pract., 118 (2011) c9-c18.

[173] S. T. Reeders, Molecular genetics of hereditary nephritis, Kidney Int., 42 (1992) 783-792.

[174] K. Tryggvason, J. Zhou, S. L. Hostikka, and T. B. Shows, Molecular genetics of Alport syndrome, Kidney Int., 43 (1993) 38-44.

[175] R. Morello, G. Zhou, S. D. Dreyer, S. J. Harvey, Y. Ninomiya, P. S. Thorner, J. H. Miner, W. Cole, A. Winterpacht, B. Zabel, K. C. Oberg, and B. Lee, Regulation of glomerular basement membrane collagen expression by LMX1B contributes to renal disease in nail patella syndrome, Nat. Genet, 27 (2001) 205-208.

[176] A. Nicholas and P. Jacques, Immune[hyphen (true graphic)]Mediated Diseases Involving Basement Membranes, in: A. K. Nicholas (Ed.), Current Topics in Membranes
Basement Membranes: Cell and Molecular Biology, Academic Press, 2005, pp. 323-352.

[177] A. T. Van and L. Bruckner-Tuderman, Basement membranes and human disease, Cell Tissue Res., 339 (2010) 167-188.

[178] R. J. Butkowski, J. P. Langeveld, J. Wieslander, J. Hamilton, and B. G. Hudson, Localization of the Goodpasture epitope to a novel chain of basement membrane collagen, J Biol. Chem., 262 (1987) 7874-7877.

[179] V. V. Gossain, A. R. Gerstein, and A. W. Janes, Goodpasture's syndrome: a familial occurrence, Am Rev. Respir. Dis., 105 (1972) 621-624.

[180] A. J. D'Apice, P. Kincaid-Smith, G. H. Becker, M. G. Loughhead, J. W. Freeman, and J. M. Sands, Goodpasture's syndrome in identical twins, Ann. Intern. Med., 88 (1978) 61-62.

[181] B. Huey, K. McCormick, J. Capper, C. Ratliff, B. W. Colombe, M. R. Garovoy, and C. B. Wilson, Associations of HLA-DR and HLA-DQ types with anti-GBM nephritis by sequence-specific oligonucleotide probe hybridization, Kidney Int., 44 (1993) 307-312.

[182] A. J. Rees, D. K. Peters, N. Amos, K. I. Welsh, and J. R. Batchelor, The influence of HLA-linked genes on the severity of anti-GBM antibody-mediated nephritis, Kidney Int., 26 (1984) 445-450.

[183] R. Kalluri, T. M. Danoff, H. Okada, and E. G. Neilson, Susceptibility to anti-glomerular basement membrane disease and Goodpasture syndrome

is linked to MHC class II genes and the emergence of T cell-mediated immunity in mice, J Clin. Invest, 100 (1997) 2263-2275.

[184] D. Patel, N. Nivera, and A. R. Tunkel, Anti-glomerular basement membrane disease superimposed on membranous nephropathy: a case report and review of the literature, J Med Case. Reports., 4 (2010) 237.

[185] M. Segelmark and T. Hellmark, Autoimmune kidney diseases, Autoimmun. Rev., (2009).

[186] J. Berger and N. Hinglais, [Intercapillary deposits of IgA-IgG], J Urol. Nephrol. (Paris), 74 (1968) 694-695.

[187] J. Novak, B. A. Julian, M. Tomana, and J. Mestecky, IgA glycosylation and IgA immune complexes in the pathogenesis of IgA nephropathy, Semin. Nephrol., 28 (2008) 78-87.

[188] K. Kiryluk, B. A. Julian, R. J. Wyatt, F. Scolari, H. Zhang, J. Novak, and A. G. Gharavi, Genetic studies of IgA nephropathy: past, present, and future, Pediatr. Nephrol., 25 (2010) 2257-2268.

[189] J. Feehally, M. Farrall, A. Boland, D. P. Gale, I. Gut, S. Heath, A. Kumar, J. F. Peden, P. H. Maxwell, D. L. Morris, S. Padmanabhan, T. J. Vyse, A. Zawadzka, A. J. Rees, M. Lathrop, and P. J. Ratcliffe, HLA Has Strongest Association with IgA Nephropathy in Genome-Wide Analysis, J Am Soc. Nephrol., 21 (2010) 1791-1797.

[190] S. Pouria and S. J. Challacombe, Uteroglobin deficient mice-a novel animal model for IgA nephropathy?, Gut, 46 (2000) 452-453.

[191] J. B. Huang, W. C. Yang, C. C. Hu, A. H. Yang, and C. C. Lin, IgA deficiency with membranous glomerulonephritis: a case report and review, J Nephrol., 16 (2003) 154-158.

[192] W. G. MacCallum, Glomerular Changes in Nephritis, Southern Medical Journal, 27 (1934) 39-40.

[193] B. Alchi and D. Jayne, Membranoproliferative glomerulonephritis, Pediatr. Nephrol., 25 (2010) 1409-1418.

[194] K. D. Smith and C. E. Alpers, Pathogenic mechanisms in membranoproliferative glomerulonephritis, Curr. Opin. Nephrol. Hypertens., 14 (2005) 396-403.

[195] R. Habib, C. Kleinknecht, M. C. Gubler, and M. Levy, Idiopathic membranoproliferative glomerulonephritis in children. Report of 105 cases, Clin. Nephrol., 1 (1973) 194-214.

[196] C. F. Strife, P. T. McEnery, A. J. McAdams, and C. D. West, Membranoproliferative glomerulonephritis with disruption of the glomerular basement membrane, Clin. Nephrol., 7 (1977) 65-72.

[197] P. D. Walker, F. Ferrario, K. Joh, and S. M. Bonsib, Dense deposit disease is not a membranoproliferative glomerulonephritis, Mod. Pathol., 20 (2007) 605-616.

[198] C. E. Alpers and K. D. Smith, Cryoglobulinemia and renal disease, Curr. Opin. Nephrol. Hypertens., 17 (2008) 243-249.

[199] P. L. Berry, P. T. McEnery, A. J. McAdams, and C. D. West, Membranoproliferative glomerulonephritis in two sibships, Clin. Nephrol., 16 (1981) 101-106.

[200] P. R. Stutchfield, R. H. White, A. H. Cameron, R. A. Thompson, P. Mackintosh, and L. Wells, X-linked mesangiocapillary glomerulonephritis, Clin. Nephrol., 26 (1986) 150-156.

[201] WHO and DIMDI (German Institute of Medical Documentation and Information). International Statistical Classification of Diseases and Related Health Problems 10th Revision Version for 2007. <http://apps.who.int/classifications/apps/icd/icd10online/> . 2010. 17-10-2010.
Ref Type: Electronic Citation

[202] C. Muller, F. Alenabi, F. Chantrel, S. Muller, C. Trivin, and B. Faller, Familial membranous glomerulopathy, toxic exposure and/or genetic sensibility?, Clin. Nephrol., 70 (2008) 422-423.

[203] F. Scolari, A. Amoroso, S. Savoldi, E. Prati, P. Scaini, A. Manganoni, I. Borelli, G. Mazzola, L. Canale, G. Sacchi, and ., Familial occurrence of primary glomerulonephritis: evidence for a role of genetic factors, Nephrol. Dial. Transplant., 7 (1992) 587-596.

[204] F. Scolari, A. Amoroso, S. Savoldi, I. Borelli, B. Valzorio, E. Costantino, M. Bracchi, M. Usberti, E. Prati, and R. Maiorca, Familial membranous nephropathy, J Nephrol., 11 (1998) 35-39.

[205] C. D. Short, J. Feehally, R. Gokal, and N. P. Mallick, Familial membranous nephropathy, Br. Med. J (Clin. Res. Ed), 289 (1984) 1500.

[206] C. Izzi, S. Sanna-Cherchi, E. Prati, R. Belleri, A. Remedio, R. Tardanico, M. Foramitti, S. Guerini, B. F. Viola, E. Movilli, I. Beerman, R. Lifton, L. Leone, A. Gharavi, and F. Scolari, Familial aggregation of primary glomerulonephritis in an Italian population isolate: Valtrompia study, Kidney Int., 69 (2006) 1033-1040.

[207] M. Maccario, S. Segagni, E. Efficace, V. Piazza, F. Poggio, G. Villa, L. Picardi, G. Bovio, F. Galli, M. Nai, and ., Idiopathic membranous nephropathy in two siblings, Nephrol. Dial. Transplant., 10 (1995) 108-110.

[208] K. Sato, H. Oguchi, K. Hora, T. Furukawa, S. Furuta, H. Shigematsu, and S. Yoshizawa, Idiopathic membranous nephropathy in two brothers, Nephron, 46 (1987) 174-178.

[209] I. Elshihabi, C. I. Kaye, and A. Brzowski, Membranous nephropathy in two human leukocyte antigen-identical brothers, J Pediatr., 123 (1993) 940-942.

[210] D. Vasmant, K. Murnaghan, A. Bensman, J. Y. Muller, and B. Mougenot, Familial idiopathic membranous glomerulonephritis, Int. J Pediatr. Nephrol., 5 (1984) 193-196.

[211] L. Grcevska and M. Polenakovic, Idiopathic membranous nephropathy (IMN) in two HLA-identical brothers with different outcome of the disease, Clin. Nephrol., 52 (1999) 194-196.

[212] A. Guella, M. Akhtar, and P. Ronco, Idiopathic membranous nephropathy in identical twins, Am J Kidney Dis., 29 (1997) 115-118.

[213] A. Vangelista, R. Tazzari, and V. Bonomini, Idiopathic membranous nephropathy in 2 twin brothers, Nephron, 50 (1988) 79-80.

[214] M. Meroni, A. Volpi, M. Usberti, G. Battini, L. T. Tarelli, F. Giordano, and A. Sessa, Two brothers with idiopathic membranous nephropathy and familial sensorineural deafness, Am J Kidney Dis., 15 (1990) 269-272.

[215] D. Bockenhauer, H. Debiec, N. Sebire, M. Barratt, P. Warwicker, P. Ronco, and R. Kleta, Familial membranous nephropathy: an X-linked genetic susceptibility?, Nephron Clin. Pract., 108 (2008) c10-c15.

[216] A. H. Tay, E. C. Ren, B. Murugasu, S. K. Sim, P. H. Tan, A. H. Cohen, and H. K. Yap, Membranous nephropathy with anti-tubular basement membrane antibody may be X-linked, Pediatr. Nephrol., 14 (2000) 747-753.

[217] J. Klein, George Snell's first foray into the unexplored territory of the major histocompatibility complex, Genetics, 159 (2001) 435-439.

[218] W. F. Bodmer, The HLA system: introduction, Br. Med. Bull., 34 (1978) 213-216.

[219] D. Turner, The human leucocyte antigen (HLA) system, Vox Sang., 87 Suppl1 (2004) 87-90.

[220] F. Lilly, E. A. Boyse, and L. J. Old, GENETIC BASIS OF SUSCEPTIBILITY TO VIRAL LEUKAEMOGENESIS, Lancet, 2 (1964) 1207-1209.

[221] R. Patel, M. R. Mickey, and P. I. Terasaki, Leucocyte antigens and disease. I. Association of HL-A2 and chronic glomerulonephritis, Br. Med. J, 2 (1969) 424-426.

[222] F. C. Grumet, A. Coukell, J. G. Bodmer, W. F. Bodmer, and H. O. McDevitt, Histocompatibility (HL-A) antigens associated with systemic lupus erythematosus. A possible genetic predisposition to disease, N. Engl. J Med., 285 (1971) 193-196.

[223] Nomenclature Committee for Factors of the HLA System. HLA Nomenclature. <http://hla.alleles.org/nomenclature/index.html> . 2010. The Anthony Nolan Trust. 17-3-2010.
Ref Type: Electronic Citation

[224] T. Shiina, K. Hosomichi, H. Inoko, and J. K. Kulski, The HLA genomic loci map: expression, interaction, diversity and disease, J Hum Genet, 54 (2009) 15-39.

[225] Complete sequence and gene map of a human major histocompatibility complex. The MHC sequencing consortium, Nature, 401 (1999) 921-923.

[226] R. Horton, L. Wilming, V. Rand, R. C. Lovering, E. A. Bruford, V. K. Khodiyar, M. J. Lush, S. Povey, C. C. Talbot, Jr., M. W. Wright, H. M. Wain, J. Trowsdale, A. Ziegler, and S. Beck, Gene map of the extended human MHC, Nat. Rev. Genet, 5 (2004) 889-899.

[227] C. A. Stewart, R. Horton, R. J. Allcock, J. L. Ashurst, A. M. Atrazhev, P. Coggill, I. Dunham, S. Forbes, K. Halls, J. M. Howson, S. J. Humphray, S. Hunt, A. J. Mungall, K. Osoegawa, S. Palmer, A. N. Roberts, J. Rogers, S. Sims, Y. Wang, L. G. Wilming, J. F. Elliott, P. J. de Jong, S. Sawcer, J. A. Todd, J. Trowsdale, and S. Beck, Complete MHC haplotype sequencing for common disease gene mapping, Genome Res., 14 (2004) 1176-1187.

[228] International Histocompatibility Working Group. IHWG. <http://www.ihwg.org/about/history.html> . 2010. 17-3-2010.
Ref Type: Electronic Citation

[229] C. Siebold, B. E. Hansen, J. R. Wyer, K. Harlos, R. E. Esnouf, A. Svejgaard, J. I. Bell, J. L. Strominger, E. Y. Jones, and L. Fugger, Crystal structure of HLA-DQ0602 that protects against type 1 diabetes and confers strong susceptibility to narcolepsy, Proc. Natl. Acad. Sci. U. S. A, 101 (2004) 1999-2004.

[230] Jmol: an open-source Java viewer for chemical structures in 3D. <http://www.jmol.org/> . 30-7-2010. 4-4-2011.
Ref Type: Electronic Citation

[231] Zinkernagel R., Basic Principles of Immunology, in: Kayser F.H., Bienz K.A., Eckert J., and Zinkernagel R.M. (Eds.), Medical Microbiology, Thieme, Stuttgart, New York, 2005, pp. 43-144.

[232] K. K. Mittal, M. R. Mickey, D. P. Singal, and P. I. Terasaki, Serotyping for homotransplantation. 18. Refinement of microdroplet lymphocyte cytotoxicity test, Transplantation, 6 (1968) 913-927.

[233] P. T. Klouda, J. Manos, E. J. Acheson, P. A. Dyer, F. S. Goldby, R. Harris, W. Lawler, N. P. Mallick, and G. Williams, Strong association between idiopathic membranous nephropathy and HLA-DRW3, Lancet, 2 (1979) 770-771.

[234] J. C. le Petit, B. Laurent, and F. C. Berthoux, HLA-DR3 and idiopathic membranous nephritis (IMN) association, Tissue Antigens, 20 (1982) 227-228.

[235] C. C. Huang, Strong association of HLA-DR3 in Chinese patients with idiopathic membranous nephropathy, Tissue Antigens., 33 (1989) 425-426.

[236] G. A. Muller, C. Muller, G. Liebau, J. Kompf, H. Ising, and P. Wernet, Strong association of idiopathic membranous nephropathy (IMN) with HLA-DR 3 and MT-2 without involvement of HLA-B 18 and no association to BfF1, Tissue Antigens, 17 (1981) 332-337.

[237] F. C. Berthoux, B. Laurent, J. C. le Petit, C. Genin, F. Broutin, F. Touraine, A. A. Hassan, and A. Champailler, Immunogenetics and immunopathology of human primary membranous glomerulonephritis: HLA-A, B, DR antigens; functional activity of splenic macrophage Fc-receptors and peripheral blood T-lymphocyte subpopulations, Clin. Nephrol., 22 (1984) 15-20.

[238] H. U. Rashid, S. S. Papiha, B. Agroyannis, A. R. Morley, M. K. Ward, D. F. Roberts, and D. N. Kerr, The associations of HLA and other genetic markers with glomerulonephritis, Hum Genet., 63 (1983) 38-44.

[239] S. S. Papiha, S. K. Pareek, R. S. Rodger, A. R. Morley, R. Wilkinson, D. F. Roberts, and D. N. Kerr, HLA-A, B, DR and Bf allotypes in patients with idiopathic membranous nephropathy (IMN), Kidney Int., 31 (1987) 130-134.

[240] S. Tomura, H. Kashiwabara, H. Tuchida, H. Shishido, S. Sakurai, T. Miyajima, K. Tsuji, and J. Takeuchi, Strong association of idiopathic membranous nephropathy with HLA-DR2 and MT1 in Japanese, Nephron, 36 (1984) 242-245.

[241] Y. Hiki, Y. Kobayashi, I. Itoh, and N. Kashiwagi, Strong association of HLA-DR2 and MT1 with idiopathic membranous nephropathy in Japan, Kidney Int., 25 (1984) 953-957.

[242] H. Nishimukai, I. Nakanishi, H. Kitamura, and Y. Tamaki, Factor B subtypes in Japanese patients with IgA nephropathy and with idiopathic membranous nephropathy, Exp. Clin. Immunogenet., 5 (1988) 196-202.

[243] S. H. Sacks, A. Bushell, N. A. Rust, J. A. Karagiannis, D. P. Jewell, J. G. Ledingham, K. J. Wood, and A. J. McMichael, Functional and biochemical subtypes of the haplotype HLA-DR3 in patients with celiac disease or idiopathic membranous nephropathy, Hum Immunol., 20 (1987) 175-187.

[244] R. W. Vaughan, A. G. Demaine, and K. I. Welsh, A DQA1 allele is strongly associated with idiopathic membranous nephropathy, Tissue Antigens, 34 (1989) 261-269.

[245] R. W. Vaughan, M. R. Tighe, K. Boki, S. Alexoupolos, J. Papadakis, J. S. Lanchbury, K. I. Welsh, and D. G. Williams, An analysis of HLA class II gene polymorphism in British and Greek idiopathic membranous nephropathy patients, Eur. J Immunogenet., 22 (1995) 179-186.

[246] B. I. Freedman, B. J. Spray, G. M. Dunston, and E. R. Heise, HLA associations in end-stage renal disease due to membranous glomerulonephritis: HLA-DR3 associations with progressive renal injury. Southeastern Organ Procurement Foundation, Am J Kidney Dis., 23 (1994) 797-802.

[247] S. Ogahara, S. Naito, K. Abe, I. Michinaga, and K. Arakawa, Analysis of HLA class II genes in Japanese patients with idiopathic membranous glomerulonephritis, Kidney Int., 41 (1992) 175-182.

[248] S. H. Sacks, S. Nomura, C. Warner, S. Naito, S. Ogahara, R. Vaughan, and D. Briggs, Analysis of complement C4 loci in Caucasoids and Japanese with idiopathic membranous nephropathy, Kidney Int., 42 (1992) 882-887.

[249] J. J. Neefjes, F. Momburg, and G. J. Hammerling, Selective and ATP-dependent translocation of peptides by the MHC-encoded transporter, Science, 261 (1993) 769-771.

[250] M. J. Kleijmeer, A. Kelly, H. J. Geuze, J. W. Slot, A. Townsend, and J. Trowsdale, Location of MHC-encoded transporters in the endoplasmic reticulum and cis-Golgi, Nature, 357 (1992) 342-344.

[251] I. Djilali-Saiah, V. Benini, S. Daniel, R. Assan, J. F. Bach, and S. Caillat-Zucman, Linkage disequilibrium between HLA class II (DR, DQ, DP) and antigen processing (LMP, TAP, DM) genes of the major histocompatibility complex, Tissue Antigens, 48 (1996) 87-92.

[252] D. Chevrier, M. Giral, V. Braud, J. P. Soulillou, and J. D. Bignon, Membranous nephropathy and a TAP1 gene polymorphism, N Engl J Med, 331 (1994) 133-134.

[253] L. K. Denzin and P. Cresswell, HLA-DM induces CLIP dissociation from MHC class II alpha beta dimers and facilitates peptide loading, Cell, 82 (1995) 155-165.

[254] V. S. Sloan, P. Cameron, G. Porter, M. Gammon, M. Amaya, E. Mellins, and D. M. Zaller, Mediation by HLA-DM of dissociation of peptides from HLA-DR, Nature, 375 (1995) 802-806.

[255] D. Chevrier, M. Giral, R. Perrichot, D. Latinne, P. Coville, J. Y. Muller, J. P. Soulillou, and J. D. Bignon, Idiopathic and secondary membranous nephropathy and polymorphism at TAP1 and HLA-DMA loci, Tissue Antigens., 50 (1997) 164-169.

[256] D. Thibaudin, L. Thibaudin, P. Berthoux, C. Mariat, J. P. Filippis, B. Laurent, E. Alamartine, and F. Berthoux, TNFA2 and d2 alleles of the

tumor necrosis factor alpha gene polymorphism are associated with onset/occurrence of idiopathic membranous nephropathy, Kidney Int., 71 (2007) 431-437.

[257] C. Bantis, P. J. Heering, S. Aker, M. Siekierka, N. Kuhr, B. Grabensee, and K. Ivens, Tumor necrosis factor-alpha gene G-308A polymorphism is a risk factor for the development of membranous glomerulonephritis, Am J Nephrol., 26 (2006) 12-15.

[258] S. H. Sacks, C. Warner, R. D. Campbell, and I. Dunham, Molecular mapping of the HLA class II region in HLA-DR3 associated idiopathic membranous nephropathy, Kidney Int. Suppl., 39:S13-9. (1993) S13-S19.

[259] J. N. Hirschhorn, K. Lohmueller, E. Byrne, and K. Hirschhorn, A comprehensive review of genetic association studies, Genet Med, 4 (2002) 45-61.

[260] M. Kestila, U. Lenkkeri, M. Mannikko, J. Lamerdin, P. McCready, H. Putaala, V. Ruotsalainen, T. Morita, M. Nissinen, R. Herva, C. E. Kashtan, L. Peltonen, C. Holmberg, A. Olsen, and K. Tryggvason, Positionally cloned gene for a novel glomerular protein--nephrin--is mutated in congenital nephrotic syndrome, Mol. Cell, 1 (1998) 575-582.

[261] Y. H. Liu, C. H. Chen, S. Y. Chen, Y. J. Lin, W. L. Liao, C. H. Tsai, L. Wan, and F. J. Tsai, Association of phospholipase A2 receptor 1 polymorphisms with idiopathic membranous nephropathy in Chinese patients in Taiwan, Journal of Biomedical Science, 17 (2010) 81-89.

[262] E. Ritz and S. R. Orth, Nephropathy in patients with type 2 diabetes mellitus, N Engl J Med, 341 (1999) 1127-1133.

[263] H. Taniwaki, E. Ishimura, N. Matsumoto, M. Emoto, M. Inaba, and Y. Nishizawa, Relations between ACE gene and ecNOS gene polymorphisms and resistive index in type 2 diabetic patients with nephropathy, Diabetes Care, 24 (2001) 1653-1660.

[264] M. Burg, J. Menne, T. Ostendorf, V. Kliem, and J. Floege, Gene-polymorphisms of angiotensin converting enzyme and endothelial nitric oxide synthase in patients with primary glomerulonephritis, Clin. Nephrol., 48 (1997) 205-211.

[265] C. H. Chen, K. H. Shu, M. C. Wen, K. J. Chen, C. H. Cheng, J. D. Lian, M. J. Wu, T. M. Yu, and F. J. Tsai, Impact of plasminogen activator inhibitor-1 gene polymorphisms on primary membranous nephropathy, Nephrol. Dial. Transplant., 23 (2008) 3166-3173.

[266] N. Mukerji, T. V. Damodaran, and M. P. Winn, TRPC6 and FSGS: the latest TRP channelopathy, Biochim. Biophys. Acta, 1772 (2007) 859-868.

[267] W. C. Chen, S. Y. Chen, C. H. Chen, H. Y. Chen, Y. W. Lin, T. J. Ho, Y. C. Huang, J. L. Shen, F. J. Tsai, and Y. H. Chen, Lack of association

between transient receptor potential cation channel 6 polymorphisms and primary membranous glomerulonephritis, Ren Fail., 32 (2010) 666-672.

[268] S. Adler and H. R. Brady, Cell adhesion molecules and the glomerulopathies, Am J Med, 107 (1999) 371-386.

[269] C. Bantis, P. J. Heering, S. Aker, N. Kuhr, B. Grabensee, and K. Ivens, Influence of beta3 integrin gene Leu/Pro33 polymorphism on primary glomerulonephritis, Nephron Exp. Nephrol., 99 (2005) e33-e37.

[270] L. F. Brown, B. Berse, K. Tognazzi, E. J. Manseau, W. L. Van de, D. R. Senger, H. F. Dvorak, and S. Rosen, Vascular permeability factor mRNA and protein expression in human kidney, Kidney Int., 42 (1992) 1457-1461.

[271] K. Shulman, S. Rosen, K. Tognazzi, E. J. Manseau, and L. F. Brown, Expression of vascular permeability factor (VPF/VEGF) is altered in many glomerular diseases, J Am Soc. Nephrol., 7 (1996) 661-666.

[272] K. Doi, E. Noiri, and T. Fujita, Role of vascular endothelial growth factor in kidney disease, Curr. Vasc. Pharmacol., 8 (2010) 122-128.

[273] A. M. Summers, B. M. Coupes, M. F. Brennan, S. A. Ralph, C. D. Short, and P. E. Brenchley, VEGF -460 genotype plays an important role in progression to chronic kidney disease stage 5, Nephrol. Dial. Transplant., 20 (2005) 2427-2432.

[274] S. Kim, H. J. Chin, K. Y. Na, S. Kim, J. Oh, W. Chung, J. W. Noh, Y. K. Lee, J. T. Cho, E. K. Lee, and D. W. Chae, Single Nucleotide Polymorphisms in the Phospholipase A(2) Receptor Gene Are Associated with Genetic Susceptibility to Idiopathic Membranous Nephropathy, Nephron Clin. Pract., 117 (2010) c253-c258.

[275] Williams R, Genetic, Keywords: A Vocabulary of Culture and Society, Taylor & Francis, 1976, pp. 116-117.

[276] L. Darden, William Bateson and the promise of Mendelism, J Hist Biol., 10 (1977) 87-106.

[277] Titus Lucretius Carus. Titi Lucreti Cari de Rerum Natura Liber Primus. http://www.thelatinlibrary.com/lucretius/lucretius1.shtml> . 2011. 4-4-2011.
Ref Type: Electronic Citation

[278] J. B. HALDANE, A DEFENSE OF BEANBAG GENETICS, Perspect. Biol. Med, 7 (1964) 343-359.

[279] C. J. Classen, Poetry and Rhetoric in Lucretius, Transactions and Proceedings of the American Philological Association, 99 (1968) 77-118.

[280] G. Bergman, Pangenesis as a source of new genetic information. The history of a now disproven theory, Riv. Biol., 99 (2006) 425-443.

[281] G. P. Sykiotis, G. D. Kalliolias, and A. G. Papavassiliou, Hippocrates and genomic medicine, Arch. Med Res., 37 (2006) 181-183.

[282] Morris S W, Fleeming Jenkin and *The Origin of Species*: a reassessment, BJHS, 27 (1994) 313-343.

[283] Jenkin F, (Review of) 'The Origin of Species', The North British Review, 92 (1867) 277-318.

[284] R. W. Burkhardt, Lamarck, evolution and the politics of science, J Hist Biol., 3 (1970) 275-298.

[285] A. E. Handel and S. V. Ramagopalan, Is Lamarckian evolution relevant to medicine?, BMC. Med Genet, 11 (2010) 73.

[286] Mendel G., Versuche uber Pflanzen Hybriden, Verhandlungen des naturforschenden Vereins zu Brunn, 4 (1866) 3-47.

[287] Johannsen W., Elemente der exakten Erblichkeitslehre, Gustav Fischer, Jena 1909.

[288] Provine W.B., The Origins of Theoretical Population Genetics, University Of Chicago Press, Chicago 2001.

[289] M. Rasmuson, Review: polygenes revisited, HEREDITAS, 125 (1996) 1-9.

[290] Fisher R A, The Correlation between Relatives on the Supposition of Mendelian Inheritance, Transactions of the Royal Society of Edinburgh, 52 (1918) 399-433.

[291] Fisher R A, The genetical theory of natural selection, Clarendon Press, London 1930.

[292] F. J. Ayala and W. M. Fitch, Genetics and the origin of species: an introduction, Proc. Natl. Acad. Sci. U. S. A, 94 (1997) 7691-7697.

[293] Dobzhansky T G, Genetics and the Origin of Species, Columbia University Press, 1982.

[294] N. Roll-Hansen, Sources of Wilhelm Johannsen's genotype theory, J. Hist Biol., 42 (2009) 457-493.

[295] R. Falk, What is a gene?, Studies In History and Philosophy of Science Part A, 17 (1986) 133-173.

[296] M. Mahner and M. Kary, What exactly are genomes, genotypes and phenotypes? And what about phenomes?, J Theor. Biol., 186 (1997) 55-63.

[297] O. Nachtomy, A. Shavit, and Z. Yakhini, Gene expression and the concept of the phenotype, Stud. Hist Philos. Biol. Biomed. Sci., 38 (2007) 238-254.

[298] F. B. Churchill, William Johannsen and the Genotype Concept, Journal of the History of Biology, 7 (1974) 5-30.

[299] Lewontin R. The Genotype/Phenotype Distinction. <http://plato.stanford.edu/entries/genotype-phenotype/> . 23-1-2004. Stanford Encyclopedia of Phylosophy. 3-3-2010.
Ref Type: Electronic Citation

[300] M. Pigliucci, Genotype-phenotype mapping and the end of the 'genes as blueprint' metaphor, Philos. Trans. R. Soc. Lond B Biol. Sci., 365 (2010) 557-566.

[301] Lewontin R C, The Problem, The Genetic Basis of Evolutionary Change, Columbia University Press, New York and London, 1974, pp. 3-15.

[302] P. Alberch, From genes to phenotype: dynamical systems and evolvability, Genetica, 84 (1991) 5-11.

[303] P. Schuster, A beginning of the end of the holism versus reductionism debate?: Molecular biology goes cellular and organismic, Complexity, 13 (2007) 10-13.

[304] A. Zucker, Holism and reductionism: a view from genetics, J Med Philos., 6 (1981) 145-163.

[305] Murphy E.A. and Pyeritz R.E., Pathogenetics of disease, in: Rimoin D.L., Pyeritz R.E., Korf B.R., Connor J.M., and Emery A.E.H. (Eds.), Emery and Rimoin's Principles and Practice of Medical Genetics, Churchill Livingstone, 2002, pp. 439-455.

[306] T. Boveri, An Organism Produced Sexually without Characteristics of the Mother, The American Naturalist, 27 (1893) 222-232.

[307] M. D. Laubichler and E. H. Davidson, Boveri's long experiment: sea urchin merogones and the establishment of the role of nuclear chromosomes in development, Dev. Biol., 314 (2008) 1-11.

[308] Sutton WS, On the Morphology of the Chromosome Group in *Brachystola magna*, Biological Bulletin, 4 (1902) 24-39.

[309] E. B. Wilson, The Sex Chromosomes, Archiv fur Mikroskopische Anatomie, 77 II (1911) 249A-271A.

[310] S. Blixt, Why didn't Gregor Mendel find linkage?, Nature, 256 (1975) 206.

[311] Bateson W, Saunders E.R., and Punnett R.C., Experimental studies in the physiology of heredity, Reports to the Evolution Committee of the Royal Society, 2 (1905) 1,80-55,99.

[312] E. Passarge, B. Horsthemke, and R. A. Farber, Incorrect use of the term synteny, Nat. Genet, 23 (1999) 387.

[313] Morgan T.H., The Theory of the Gene, The American Naturalist, 51 (1917) 513-544.

[314] H. Sturtevant A, The linear arrangement of six sex-linked factors in Drosophila, as shown by their mode of association, Journal of Experimental Zoology, 14 (1913) 43-59.

[315] C. Schlotterer, The evolution of molecular markers--just a matter of fashion?, Nat. Rev. Genet, 5 (2004) 63-69.

[316] Galton F., Hereditary Genius: an Inquiry into its Laws and Consequences, MacMillan and co., London 1869.

[317] N. W. Gillham, Evolution by jumps: Francis Galton and William Bateson and the mechanism of evolutionary change, Genetics, 159 (2001) 1383-1392.

[318] A. R. Rushton, Nettleship, Pearson and Bateson: the biometric-Mendelian debate in a medical context, J Hist Med Allied Sci., 55 (2000) 134-157.

[319] P. S. Harper, William Bateson, human genetics and medicine, Hum Genet, 118 (2005) 141-151.

[320] A. E. Garrod, Inborn Errors of Metabolism, Frowde, London 1909.

[321] R. C. Punnett and W. Bateson, THE HEREDITY OF SEX, Science, 27 (1908) 785-787.

[322] Haldane J.B.S., The Rate of Mutations of Human Genes, HEREDITAS, Proceedings of the Eighth International Congress of Genetics (1949) 267-273.

[323] W. H. Schneider, The history of research on blood group genetics: initial discovery and diffusion, Hist Philos. Life Sci., 18 (1996) 277-303.

[324] K. Landsteiner, INDIVIDUAL DIFFERENCES IN HUMAN BLOOD, Science, 73 (1931) 403-409.

[325] K. Landsteiner and P. Levine, ON THE INHERITANCE OF AGGLUTINOGENS OF HUMAN BLOOD DEMONSTRABLE BY IMMUNE AGGLUTININS, J Exp. Med, 48 (1928) 731-749.

[326] J. F. Crow, Felix Bernstein and the first human marker locus, Genetics, 133 (1993) 4-7.

[327] F. Yamamoto, H. Clausen, T. White, J. Marken, and S. Hakomori, Molecular genetic basis of the histo-blood group ABO system, Nature, 345 (1990) 229-233.

[328] R. Owen, Karl Landsteiner and the first human marker locus, Genetics, 155 (2000) 995-998.

[329] M. Okroi and L. J. McCarthy, The original blood group pioneers: the Hirszfelds, Transfus. Med Rev., 24 (2010) 244-246.

[330] W. Bodmer and C. Bonilla, Common and rare variants in multifactorial susceptibility to common diseases, Nat. Genet, 40 (2008) 695-701.

[331] Provine W.B., The Origins of Theoretical Population Genetics, University Of Chicago Press, Chicago 1 A.D.

[332] J. F. Crow, C. C. Little, cancer and inbred mice, Genetics, 161 (2002) 1357-1361.

[333] C. C. Little and E. E. Tyzzer, Further experimental studies on the inheritance of susceptibility to a Transplantable tumor, Carcinoma (J. W. A.) of the Japanese waltzing Mouse, J Med Res., 33 (1916) 393-453.

[334] P. Gorer, The detection of a hereditary antigenic difference in the blood of mice by means of human group a serum, Journal of Genetics, 32 (1936) 17-31.

[335] P. B. Medawar, The behaviour and fate of skin autografts and skin homografts in rabbits: A report to the War Wounds Committee of the Medical Research Council, J Anat., 78 (1944) 176-199.

[336] P. A. Gorer, S. Lyman, and G. D. Snell, Studies on the Genetic and Antigenic Basis of Tumour Transplantation. Linkage between a Histocompatibility Gene and 'Fused' in Mice, Proceedings of the Royal Society of London. Series B, Biological Sciences, 135 (1948) 499-505.

[337] G. D. Snell, M. Cherry, and P. Demant, Evidence that H-2 private specificities can be arranged in two mutually exclusive systems possibly homologous with two subsystems of HL-A, Transplant. Proc., 3 (1971) 183-186.

[338] E. Thorsby, A tentative new model for the organization of the mouse H-2 histocompatibility system: two segregant series of antigens, Eur. J Immunol., 1 (1971) 57-59.

[339] J. DAUSSET, [Iso-leuko-antibodies.], Acta Haematol., 20 (1958) 156-166.

[340] Dausset J, This Week's Citation Classic, Current Contents, 15 (1983).

[341] J. J. VAN ROOD, J. G. EERNISSE, and L. A. van, Leucocyte antibodies in sera from pregnant women, Nature, 181 (1958) 1735-1736.

[342] R. PAYNE and M. R. ROLFS, Fetomaternal leukocyte incompatibility, J Clin. Invest, 37 (1958) 1756-1763.

[343] J. J. VAN ROOD and L. A. van, LEUKOCYTE GROUPING. A METHOD AND ITS APPLICATION, J Clin. Invest, 42 (1963) 1382-1390.

[344] W. Bodmer, J. Bodmer, S. Adler, R. Payne, and J. Bialek, GENETICS OF "4" AND "LA" HUMAN LEUKOCYTE GROUPS*, Annals of the New York Academy of Sciences, 129 (1966) 473-489.

[345] F. Kissmeyer-Nielsen, A. Svejgaard, and M. Hauge, Genetics of the human HL-A transplantation system, Nature, 219 (1968) 1116-1119.

[346] B. G. Solheim, A. Bratlie, L. Sandberg, L. Staub-Nielsen, and E. Thorsby, Further evidence of a third HL-A locus, Tissue Antigens, 3 (1973) 439-453.

[347] E. Thorsby, A short history of HLA, Tissue Antigens, 74 (2009) 101-116.

[348] W. F. Bodmer, In memoriam Ruggero Ceppellini 1917-1988, Immunogenetics, 29 (1989) 145-147.

[349] F. H. Bach, M. B. Widmer, M. L. Bach, and J. Klein, Serologically defined and lymphocyte-defined components of the major histocompatibility complex in the mouse, J Exp. Med, 136 (1972) 1430-1444.

[350] S. H. van, A. Westerveld, A. Hagemeijer, J. R. Mees, K. P. Meera, and O. B. Zaalberg, Human antigen and enzyme markers in man-Chinese hamster somatic cell hybrids: evidence for synteny between the HL-A, PGM3, ME1, and IPO-B loci, Proc. Natl. Acad. Sci. U. S. A, 71 (1974) 962-965.

[351] M. H. Breuning, van den Berg-Loonen EM, L. F. Bernini, J. B. Bijlsma, L. E. van, K. P. Meera, and L. E. Nijenhuis, Localization of HLA on the short arm of chromosome 6, Hum Genet, 37 (1977) 131-139.

[352] J. D. Zervas, I. W. Delamore, and M. C. Israels, Leucocyte phenotypes in Hodgkin's disease, Lancet, 2 (1970) 634-635.

[353] L. Cavalli-Sforza and Bodmer W.F., Transient and Balanced Polymorphisms, The Genetics of Human Populations, Dover Publications Inc., Mineola, New York, 1999, pp. 118-189.

[354] J. L. Hubby and R. C. Lewontin, A molecular approach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in Drosophila pseudoobscura, Genetics, 54 (1966) 577-594.

[355] R. C. Lewontin and J. L. Hubby, A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and

degree of heterozygosity in natural populations of Drosophila pseudoobscura, Genetics, 54 (1966) 595-609.

[356] H. Harris, Enzyme Polymorphisms in Man, Proceedings of the Royal Society of London. Series B, Biological Sciences, 164 (1966) 298-310.

[357] M. Kimura, Evolutionary rate at the molecular level, Nature, 217 (1968) 624-626.

[358] R. Dahm, Discovering DNA: Friedrich Miescher and the early years of nucleic acid research, Hum Genet, 122 (2008) 565-581.

[359] F. Griffith, The Significance of Pneumococcal Types, J Hyg. (Lond), 27 (1928) 113-159.

[360] O. T. Avery, C. M. Macleod, and M. McCarty, STUDIES ON THE CHEMICAL NATURE OF THE SUBSTANCE INDUCING TRANSFORMATION OF PNEUMOCOCCAL TYPES : INDUCTION OF TRANSFORMATION BY A DESOXYRIBONUCLEIC ACID FRACTION ISOLATED FROM PNEUMOCOCCUS TYPE III, J Exp. Med, 79 (1944) 137-158.

[361] A. D. HERSHEY and M. CHASE, Independent functions of viral protein and nucleic acid in growth of bacteriophage, J Gen. Physiol, 36 (1952) 39-56.

[362] R. E. FRANKLIN and R. G. GOSLING, Molecular configuration in sodium thymonucleate, Nature, 171 (1953) 740-741.

[363] M. H. WILKINS, A. R. STOKES, and H. R. WILSON, Molecular structure of deoxypentose nucleic acids, Nature, 171 (1953) 738-740.

[364] J. D. WATSON and F. H. CRICK, Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid, Nature, 171 (1953) 737-738.

[365] J. D. WATSON and F. H. CRICK, Genetical implications of the structure of deoxyribonucleic acid, Nature, 171 (1953) 964-967.

[366] E. Zuckerkandl and L. Pauling, Molecules as documents of evolutionary history, J Theor. Biol., 8 (1965) 357-366.

[367] I. R. Lehman, M. J. Bessman, E. S. Simms, and A. Kornberg, Enzymatic synthesis of deoxyribonucleic acid. I. Preparation of substrates and partial purification of an enzyme from Escherichia coli, J Biol. Chem., 233 (1958) 163-170.

[368] M. J. Bessman, I. R. Lehman, E. S. Simms, and A. Kornberg, Enzymatic synthesis of deoxyribonucleic acid. II. General properties of the reaction, J Biol. Chem., 233 (1958) 171-177.

[369] Kresge N, Simoni R.D., and Hill R.L., DNA Replication in Bacteriophage: the Work of Charles C. Richardson, J Biol. Chem., 282 (2007) e22-e23.

[370] R. J. Roberts, How restriction enzymes became the workhorses of molecular biology, Proc. Natl. Acad. Sci. U. S. A, 102 (2005) 5905-5908.

[371] S. N. Cohen, A. C. Chang, H. W. Boyer, and R. B. Helling, Construction of biologically functional bacterial plasmids in vitro, Proc. Natl. Acad. Sci. U. S. A, 70 (1973) 3240-3244.

[372] Rabinow Paul, Making PCR. A Story of Biotechnology, The University of Chicago Press, Chicago & London 1996.

[373] A. J. Jeffreys, DNA sequence variants in the G gamma-, A gamma-, delta- and beta-globin genes of man, Cell, 18 (1979) 1-10.

[374] B. Kerem, J. M. Rommens, J. A. Buchanan, D. Markiewicz, T. K. Cox, A. Chakravarti, M. Buchwald, and L. C. Tsui, Identification of the cystic fibrosis gene: genetic analysis, Science, 245 (1989) 1073-1080.

[375] A. J. Jeffreys, V. Wilson, and S. L. Thein, Hypervariable 'minisatellite' regions in human DNA, Nature, 314 (1985) 67-73.

[376] G. Vergnaud and F. Denoeud, Minisatellites: mutability and genome architecture, Genome Res., 10 (2000) 899-907.

[377] M. Litt and J. A. Luty, A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene, Am J Hum Genet, 44 (1989) 397-401.

[378] E. Buschiazzo and N. J. Gemmell, The rise, fall and renaissance of microsatellites in eukaryotic genomes, Bioessays, 28 (2006) 1040-1050.

[379] D. Housman, Human DNA polymorphism, N Engl J Med, 332 (1995) 318-320.

[380] D. N. Cooper, B. A. Smith, H. J. Cooke, S. Niemann, and J. Schmidtke, An estimate of unique DNA sequence heterozygosity in the human genome, Hum Genet, 69 (1985) 201-205.

[381] Integrating common and rare genetic variation in diverse human populations, Nature, 467 (2010) 52-58.

[382] S. T. Sherry, M. H. Ward, M. Kholodov, J. Baker, L. Phan, E. M. Smigielski, and K. Sirotkin, dbSNP: the NCBI database of genetic variation, Nucleic Acids Res., 29 (2001) 308-311.

[383] NCBI. dbSNP. <http://www.ncbi.nlm.nih.gov/projects/SNP/> . 23-9-2011. 4-4-0011.
Ref Type: Electronic Citation

[384] D. Botstein, R. L. White, M. Skolnick, and R. W. Davis, Construction of a genetic linkage map in man using restriction fragment length polymorphisms, Am J Hum Genet, 32 (1980) 314-331.

[385] R. P. Donahue, W. B. Bias, J. H. Renwick, and V. A. McKusick, Probable assignment of the Duffy blood group locus to chromosome 1 in man, Proc. Natl. Acad. Sci. U. S. A, 61 (1968) 949-955.

[386] V. A. McKusick, The morbid anatomy of the human genome: a review of gene mapping in clinical medicine (1), Medicine (Baltimore), 65 (1986) 1-33.

[387] T. H. Morgan and C. J. Lynch, The Linkage of Two Factors in Drosophila That Are Not Sex-Linked, Biological Bulletin, 23 (1912) 174-182.

[388] A. W. Edwards, Linkage methods in human genetics before the computer, Hum Genet, 118 (2005) 515-530.

[389] E. Thompson, Detection of human genetic linkage: foundations, Theor. Popul. Biol., 60 (2001) 203-213.

[390] J. Bell and J. B. S. Haldane, The Linkage between the Genes for Colour-Blindness and Haemophilia in Man, Proceedings of the Royal Society of London. Series B, Biological Sciences, 123 (1937) 119-150.

[391] C. A. B. Smith, The Detection of Linkage in Human Genetics, Journal of the Royal Statistical Society. Series B (Methodological), 15 (1953) 153-192.

[392] J. H. Edwards, Penrose and sib-pairs, Ann. Hum Genet, 62 (1998) 365-377.

[393] L. S. PENROSE, A further note on the sib-pair linkage method, Ann. Eugen., 13 (1946) 25-29.

[394] J. B. HALDANE and C. A. SMITH, A new estimate of the linkage between the genes for colourblindness and haemophilia in man, Ann. Eugen., 14 (1947) 10-31.

[395] J. MOHR, A search for linkage between the Lutheran blood group and other hereditary characters, Acta Pathol. Microbiol. Scand., 28 (1951) 207-210.

[396] N. E. Morton, Sequential tests for the detection of linkage, Am J Hum Genet, 7 (1955) 277-318.

[397] Wald A, Sequential Tests of Statistical Hypotheses, The Annals of Mathematical Statistics, 16 (1945) 117-186.

[398] A. W. Edwards, Statistical methods for evolutionary trees, Genetics, 183 (2009) 5-12.

[399] C. A. SMITH, Some comments on the statistical methods used in linkage investigations, Am J Hum Genet, 11 (1959) 289-304.

[400] Ott J, Computer Programs for Parametric Linkage Analysis, Analysis of Human Genetic Linkage, The Johns Hopkins University Press, Baltimore and London, 1999, pp. 185-191.

[401] J. Ott, Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies, Am J Hum Genet, 26 (1974) 588-597.

[402] J. Ott, H. G. Schrott, J. L. Goldstein, W. R. Hazzard, F. H. Allen, Jr., C. T. Falk, and A. G. Motulsky, Linkage studies in a large kindred with familial hypercholesterolemia, Am J Hum Genet, 26 (1974) 598-603.

[403] R. C. Elston and J. Stewart, A general model for the genetic analysis of pedigree data, Hum Hered., 21 (1971) 523-542.

[404] G. M. Lathrop and J. M. Lalouel, Easy calculations of lod scores and genetic risks on small computers, Am J Hum Genet, 36 (1984) 460-465.

[405] L. Kruglyak, M. J. Daly, M. P. Reeve-Daly, and E. S. Lander, Parametric and nonparametric linkage analysis: a unified multipoint approach, Am J Hum Genet, 58 (1996) 1347-1363.

[406] E. Sobel and K. Lange, Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics, Am J Hum Genet, 58 (1996) 1323-1337.

[407] D. F. Gudbjartsson, K. Jonasson, M. L. Frigge, and A. Kong, Allegro, a new computer program for multipoint linkage analysis, Nat. Genet, 25 (2000) 12-13.

[408] F. H. Ruddle, The William Allan Memorial Award address: Reverse genetics and beyond, Am J Hum Genet, 36 (1984) 944-953.

[409] F. S. Collins, Positional cloning: let's not call it reverse anymore, Nat. Genet, 1 (1992) 3-6.

[410] M. R. Pollak, Expanding the spectrum of NPHS1-associated disease, Kidney Int., 76 (2009) 1221-1223.

[411] Newton E Morton. History of genetic mapping. Lynn Jorde , Peter Little Mike Dunn Shankar Subramaniam. Encyclopedia of Genetics, Genomics, Proteomics and Bioinformatics Genomics. 2005.
Ref Type: Electronic Citation

[412] M. Slatkin, Linkage disequilibrium--understanding the evolutionary past and mapping the medical future, Nat. Rev. Genet., 9 (2008) 477-485.

[413] R. C. Lewontin and K. i. Kojima, The Evolutionary Dynamics of Complex Polymorphisms, Evolution, 14 (1960) 458-472.

[414] C. D. Langefeld and T. E. Fingerlin, Association methods in human genetics, Methods Mol. Biol., 404 (2007) 431-460.

[415] N. Miyashita and C. H. Langley, Molecular and phenotypic variation of the white locus region in Drosophila melanogaster, Genetics, 120 (1988) 199-212.

[416] Yule G.U., On the Association of Attributes in Statistics: With Illustrations from the Material of the Childhood Society, &c, Philos. Trans. R. Soc. Lond A Math. Phys. Sci., 194 (1900) 257-319.

[417] F. Galton, Co-Relations and Their Measurement, Chiefly from Anthropometric Data, Proceedings of the Royal Society of London, 45 (1888) 135-145.

[418] Gallin J.I., A Historical Perspective on Clinical Research, in: Gallin J.I. (Ed.), Principles and Practice of Clinical Research, Academic Press, 2002, pp. 1-11.

[419] K. J. Carpenter, A short history of nutritional science: part 1 (1785-1885), J Nutr., 133 (2003) 638-645.

[420] I. AIRD, H. H. BENTALL, and J. A. ROBERTS, A relationship between cancer of stomach and the ABO blood groups, Br. Med. J, 1 (1953) 799-801.

[421] D. J. Balding, A tutorial on statistical methods for population association studies, Nat. Rev. Genet, 7 (2006) 781-791.

[422] D. Liddell, Practical Tests of 2 X 2 Contingency Tables, Journal of the Royal Statistical Society. Series D (The Statistician), 25 (1976) 295-304.

[423] F. Yates, Test of Significance for 2 x 2 Contingency Tables, Journal of the Royal Statistical Society. Series A (General), 147 (1984) 426-463.

[424] R. L. Plackett, Karl Pearson and the Chi-Squared Test, International Statistical Review / Revue Internationale de Statistique, 51 (1983) 59-72.

[425] R. A. FISHER, On the Interpretation of chi2 from Contingency Tables, and the Calculation of P, Journal of the Royal Statistical Society, 85 (1922) 87-94.

[426] Shilgalis T.W., Cumulative Distribution Functions, Teaching Statistics, 10 (1988) 82-86.

[427] A. Agresti, A Survey of Exact Inference for Contingency Tables, Statistical Science, 7 (1992) 131-153.

[428] S. B. Gabriel, S. F. Schaffner, H. Nguyen, J. M. Moore, J. Roy, B. Blumenstiel, J. Higgins, M. DeFelice, A. Lochner, M. Faggart, S. N. Liu-Cordero, C. Rotimi, A. Adeyemo, R. Cooper, R. Ward, E. S. Lander, M.

J. Daly, and D. Altshuler, The structure of haplotype blocks in the human genome, Science, 296 (2002) 2225-2229.

[429] E. R. Waldron, J. C. Whittaker, and D. J. Balding, Fine mapping of disease genes via haplotype clustering, Genet Epidemiol., 30 (2006) 170-179.

[430] A. G. Clark, The role of haplotypes in candidate gene studies, Genet Epidemiol., 27 (2004) 321-333.

[431] H. K. Tabor, N. J. Risch, and R. M. Myers, Candidate-gene approaches for studying complex genetic traits: practical considerations, Nat. Rev. Genet., 3 (2002) 391-397.

[432] N. J. Risch, Searching for genetic determinants in the new millennium, Nature, 405 (2000) 847-856.

[433] N. Risch and K. Merikangas, The future of genetic studies of complex human diseases, Science, 273 (1996) 1516-1517.

[434] Y. Ohnishi, T. Tanaka, K. Ozaki, R. Yamada, H. Suzuki, and Y. Nakamura, A high-throughput SNP typing system for genome-wide association studies, J Hum Genet, 46 (2001) 471-477.

[435] K. Ozaki, Y. Ohnishi, A. Iida, A. Sekine, R. Yamada, T. Tsunoda, H. Sato, H. Sato, M. Hori, Y. Nakamura, and T. Tanaka, Functional SNPs in the lymphotoxin-alpha gene that are associated with susceptibility to myocardial infarction, Nat. Genet, 32 (2002) 650-654.

[436] S. Tokuhiro, R. Yamada, X. Chang, A. Suzuki, Y. Kochi, T. Sawada, M. Suzuki, M. Nagasaki, M. Ohtsuki, M. Ono, H. Furukawa, M. Nagashima, S. Yoshino, A. Mabuchi, A. Sekine, S. Saito, A. Takahashi, T. Tsunoda, Y. Nakamura, and K. Yamamoto, An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis, Nat. Genet, 35 (2003) 341-348.

[437] A. Suzuki, R. Yamada, X. Chang, S. Tokuhiro, T. Sawada, M. Suzuki, M. Nagasaki, M. Nakayama-Hamada, R. Kawaida, M. Ono, M. Ohtsuki, H. Furukawa, S. Yoshino, M. Yukioka, S. Tohma, T. Matsubara, S. Wakitani, R. Teshima, Y. Nishioka, A. Sekine, A. Iida, A. Takahashi, T. Tsunoda, Y. Nakamura, and K. Yamamoto, Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis, Nat. Genet, 34 (2003) 395-402.

[438] R. J. Klein, C. Zeiss, E. Y. Chew, J. Y. Tsai, R. S. Sackler, C. Haynes, A. K. Henning, J. P. SanGiovanni, S. M. Mane, S. T. Mayne, M. B. Bracken, F. L. Ferris, J. Ott, C. Barnstable, and J. Hoh, Complement factor H polymorphism in age-related macular degeneration, Science, 308 (2005) 385-389.

[439] S. WRIGHT, Genetical structure of populations, Nature, 166 (1950) 247-249.

[440] C. C. Li, Population subdivision with respect to multiple alleles, Ann. Hum Genet, 33 (1969) 23-29.

[441] A. L. Price, N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich, Principal components analysis corrects for stratification in genome-wide association studies, Nat. Genet., 38 (2006) 904-909.

[442] B. F. Voight and J. K. Pritchard, Confounding from cryptic relatedness in case-control association studies, PLoS. Genet., 1 (2005) e32.

[443] W. C. Knowler, R. C. Williams, D. J. Pettitt, and A. G. Steinberg, Gm3;5,13,14 and type 2 diabetes mellitus: an association in American Indians with genetic admixture, Am J Hum Genet, 43 (1988) 520-526.

[444] D. E. Reich and D. B. Goldstein, Detecting association in a case-control study while correcting for population stratification, Genet Epidemiol., 20 (2001) 4-16.

[445] J. K. Pritchard, M. Stephens, and P. Donnelly, Inference of population structure using multilocus genotype data, Genetics, 155 (2000) 945-959.

[446] B. Devlin, K. Roeder, and L. Wasserman, Genomic control, a new approach to genetic-based association studies, Theor. Popul. Biol., 60 (2001) 155-166.

[447] P. Menozzi, A. Piazza, and L. Cavalli-Sforza, Synthetic maps of human gene frequencies in Europeans, Science, 201 (1978) 786-792.

[448] G. H. Hardy, MENDELIAN PROPORTIONS IN A MIXED POPULATION, Science, 28 (1908) 49-50.

[449] E. O. Wilson and Bossert W.H., A Primer of Population Biology, Sinauer Associates, Inc., Sunderland, Massachussets 1971.

[450] Shaffer J P, Multiple Hypothesis Testing, Annu. Rev. Psychol., 46 (1995) 561-584.

[451] T. A. Pearson and T. A. Manolio, How to interpret a genome-wide association study, JAMA, 299 (2008) 1335-1344.

[452] S. J. Chanock, T. Manolio, M. Boehnke, E. Boerwinkle, D. J. Hunter, G. Thomas, J. N. Hirschhorn, G. Abecasis, D. Altshuler, J. E. Bailey-Wilson, L. D. Brooks, L. R. Cardon, M. Daly, P. Donnelly, J. F. Fraumeni, Jr., N. B. Freimer, D. S. Gerhard, C. Gunter, A. E. Guttmacher, M. S. Guyer, E. L. Harris, J. Hoh, R. Hoover, C. A. Kong, K. R. Merikangas, C. C. Morton, L. J. Palmer, E. G. Phimister, J. P. Rice, J. Roberts, C. Rotimi, M. A. Tucker, K. J. Vogan, S. Wacholder, E. M. Wijsman, D. M. Winn, and F. S. Collins, Replicating genotype-phenotype associations, Nature, 447 (2007) 655-660.

[453] D. Ellinghaus, S. Schreiber, A. Franke, and M. Nothnagel, Current software for genotype imputation, Hum Genomics, 3 (2009) 371-380.

[454] C. S. Wingo and W. L. Clapp, Proteinuria: potential causes and approach to evaluation, Am J Med Sci., 320 (2000) 188-194.

[455] M. Olson, L. Hood, C. Cantor, and D. Botstein, A common language for physical mapping of the human genome, Science, 245 (1989) 1434-1435.

[456] A. Kong, D. F. Gudbjartsson, J. Sainz, G. M. Jonsdottir, S. A. Gudjonsson, B. Richardsson, S. Sigurdardottir, J. Barnard, B. Hallbeck, G. Masson, A. Shlien, S. T. Palsson, M. L. Frigge, T. E. Thorgeirsson, J. R. Gulcher, and K. Stefansson, A high-resolution recombination map of the human genome, Nat. Genet, 31 (2002) 241-247.

[457] NCBI/NIH. UniSTS. <http://www.ncbi.nlm.nih.gov/unists> . 2011. 4-4-2011.
Ref Type: Electronic Citation

[458] D. Karolchik, A. S. Hinrichs, T. S. Furey, K. M. Roskin, C. W. Sugnet, D. Haussler, and W. J. Kent, The UCSC Table Browser data retrieval tool, Nucleic Acids Res., 32 (2004) D493-D496.

[459] W. J. Kent, C. W. Sugnet, T. S. Furey, K. M. Roskin, T. H. Pringle, A. M. Zahler, and D. Haussler, The human genome browser at UCSC, Genome Res., 12 (2002) 996-1006.

[460] J. R. O'Connell and D. E. Weeks, PedCheck: a program for identification of genotype incompatibilities in linkage analysis, Am. J. Hum. Genet., 63 (1998) 259-266.

[461] G. R. Abecasis, S. S. Cherny, W. O. Cookson, and L. R. Cardon, GRR: graphical representation of relationship errors, Bioinformatics, 17 (2001) 742-743.

[462] R. C. Elston and J. Stewart, A general model for the genetic analysis of pedigree data, Hum. Hered., 21 (1971) 523-542.

[463] L. Kruglyak, M. J. Daly, M. P. Reeve-Daly, and E. S. Lander, Parametric and nonparametric linkage analysis: a unified multipoint approach, Am. J. Hum. Genet., 58 (1996) 1347-1363.

[464] Abecasis, G. MINX. <http://www.sph.umich.edu/csg/abecasis/merlin/FAQ.html> . 2010. Center for Statistical Genetics, University of Michigan. 14-12-2010.
Ref Type: Electronic Citation

[465] G. R. Abecasis, S. S. Cherny, W. O. Cookson, and L. R. Cardon, Merlin--rapid analysis of dense genetic maps using sparse gene flow trees, Nat. Genet, 30 (2002) 97-101.

[466] N. Mukhopadhyay, L. Almasy, M. Schroeder, W. P. Mulvihill, and D. E. Weeks, Mega2: data-handling for facilitating genetic linkage and association analyses, Bioinformatics., 21 (2005) 2556-2557.

[467] H. Thiele and P. Nurnberg, HaploPainter: a tool for drawing pedigrees with complex haplotypes, Bioinformatics., 21 (2005) 1730-1732.

[468] E. S. Lander and N. J. Schork, Genetic dissection of complex traits, Science, 265 (1994) 2037-2048.

[469] F. Hsu, W. J. Kent, H. Clawson, R. M. Kuhn, M. Diekhans, and D. Haussler, The UCSC Known Genes, Bioinformatics, 22 (2006) 1036-1046.

[470] W. J. Kent, F. Hsu, D. Karolchik, R. M. Kuhn, H. Clawson, H. Trumbower, and D. Haussler, Exploring relationships and mining data with the UCSC Gene Sorter, Genome Res., 15 (2005) 737-741.

[471] GNF. BioGPS The Gene Portal Hub. <http://biogps.gnf.org/> . 2011. 4-4-2011.
Ref Type: Electronic Citation

[472] E. Hubbell, W. M. Liu, and R. Mei, Robust estimators for expression analysis, Bioinformatics, 18 (2002) 1585-1592.

[473] C. Steinhoff and M. Vingron, Normalization and quantification of differential expression in gene expression microarrays, Brief. Bioinform., 7 (2006) 166-177.

[474] A. I. Su, M. P. Cooke, K. A. Ching, Y. Hakak, J. R. Walker, T. Wiltshire, A. P. Orth, R. G. Vega, L. M. Sapinoso, A. Moqrich, A. Patapoutian, G. M. Hampton, P. G. Schultz, and J. B. Hogenesch, Large-scale analysis of the human and mouse transcriptomes, Proc. Natl. Acad. Sci. U. S. A, 99 (2002) 4465-4470.

[475] A. I. Su, T. Wiltshire, S. Batalov, H. Lapp, K. A. Ching, D. Block, J. Zhang, R. Soden, M. Hayakawa, G. Kreiman, M. P. Cooke, J. R. Walker, and J. B. Hogenesch, A gene atlas of the mouse and human protein-encoding transcriptomes, Proc. Natl. Acad. Sci. U. S. A, 101 (2004) 6062-6067.

[476] L. Pintard, A. Willems, and M. Peter, Cullin-based ubiquitin ligases: Cul3-BTB complexes join the family, EMBO J, 23 (2004) 1681-1687.

[477] I. Sumara, M. Quadroni, C. Frei, M. H. Olma, G. Sumara, R. Ricci, and M. Peter, A Cul3-based E3 ligase removes Aurora B from mitotic chromosomes, regulating mitotic progression and completion of cytokinesis in human cells, Dev. Cell, 12 (2007) 887-900.

[478] M. I. McCarthy, G. R. Abecasis, L. R. Cardon, D. B. Goldstein, J. Little, J. P. Ioannidis, and J. N. Hirschhorn, Genome-wide association studies

for complex traits: consensus, uncertainty and challenges, Nat. Rev. Genet., 9 (2008) 356-369.

[479] F. J. Steemers, W. Chang, G. Lee, D. L. Barker, R. Shen, and K. L. Gunderson, Whole-genome genotyping with the single-base extension assay, Nat. Methods, 3 (2006) 31-33.

[480] M. N. Kuppuswamy, J. W. Hoffmann, C. K. Kasper, S. G. Spitzer, S. L. Groce, and S. P. Bajaj, Single nucleotide primer extension to detect genetic diseases: experimental application to hemophilia B (factor IX) and cystic fibrosis genes, Proc. Natl. Acad. Sci. U. S. A, 88 (1991) 1143-1147.

[481] T. Pastinen, M. Raitio, K. Lindroos, P. Tainola, L. Peltonen, and A. C. Syvanen, A system for specific, high-throughput genotyping by allele-specific primer extension on microarrays, Genome Res., 10 (2000) 1031-1042.

[482] K. L. Gunderson, F. J. Steemers, G. Lee, L. G. Mendoza, and M. S. Chee, A genome-wide scalable SNP genotyping assay using microarray technology, Nat. Genet, 37 (2005) 549-554.

[483] I. Pe'er, P. I. de Bakker, J. Maller, R. Yelensky, D. Altshuler, and M. J. Daly, Evaluating and improving power in whole-genome association studies using fixed marker sets, Nat. Genet., 38 (2006) 663-667.

[484] J. C. Barrett and L. R. Cardon, Evaluating coverage of genome-wide association studies, Nat. Genet., 38 (2006) 659-662.

[485] Illumina, SNP Genotyping. "TOP/BOT" Strand and "A/B" Allele. <http://www.illumina.com/documents/products/technotes/technote_topbot.pdf> . 2006. 4-4-2011.
Ref Type: Electronic Citation

[486] Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls, Nature., 447 (2007) 661-678.

[487] Faylor C, Vinschen C, Selkowitz Y, Dessent B, and Korn D. Cygwin 1.7.7-1. <http://www.cygwin.com/> . 31-8-2010. 4-4-2011.
Ref Type: Electronic Citation

[488] van rossum G. Python. <http://www.python.org/> . 20-12-2010. Python Software Foundation. 4-4-2011.
Ref Type: Electronic Citation

[489] Terwilliger J.D. and Ott J., Handbook of Human Genetic Linkage, The Johns Hopkins University Press, Baltimore 1994.

[490] NIST/SEMATECH. e-Handbook of Statistical Methods. <http://www.itl.nist.gov/div898/handbook/eda/section3/qqplot.htm> . 23-6-2010. 7-1-2010.
Ref Type: Electronic Citation

[491] The International HapMap Project, Nature, 426 (2003) 789-796.

[492] Golden Helix. SNP & Variation Suite 7 Tutorial: SNP Analysis. <http://www.goldenhelix.com/SNP_Variation/tutorials/snp-with-svs7/introduction.html> . 2009.
Ref Type: Electronic Citation

[493] P. A. Gourraud, A. Meenagh, A. Cambon-Thomsen, and D. Middleton, Linkage disequilibrium organization of the human KIR superlocus: implications for KIR data analyses, Immunogenetics, 62 (2010) 729-740.

[494] J. Trowsdale, R. Barten, A. Haude, C. A. Stewart, S. Beck, and M. J. Wilson, The genomic context of natural killer receptor extended gene families, Immunol. Rev., 181 (2001) 20-38.

[495] Lambert C. Achieving Genome-Wide Success Part III: Quality Assurance and Data Prep in SNP and CNV Studies. <http://www.goldenhelix.com/Events/recordings/gws-3-quality-assurance/index.html> . 9-12-2009. Golden Helix. 31-5-2010.
Ref Type: Electronic Citation

[496] T. R. Burglin, Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals, Nucleic Acids Res., 25 (1997) 4173-4180.

[497] C. B. Moens and L. Selleri, Hox cofactors in vertebrate development, Dev. Biol., 291 (2006) 193-206.

[498] C. G. Sagerstrom, PbX marks the spot, Dev. Cell, 6 (2004) 737-738.

[499] M. Mojsin and M. Stevanovic, PBX1 and MEIS1 up-regulate SOX3 gene expression by direct interaction with a consensus binding site within the basal promoter region, Biochem. J, 425 (2010) 107-116.

[500] J. Yu, A. P. McMahon, and M. T. Valerius, Recent genetic studies of mouse kidney development, Curr. Opin. Genet Dev., 14 (2004) 550-557.

[501] R. Hoffmann and A. Valencia, A gene network for navigating the literature, Nat. Genet, 36 (2004) 664.

[502] 1000 Genomes - A Deep Catalog of Human Genetic Variation. <http://www.1000genomes.org/> . 2010. 1000 Genomes Organization. 16-12-2010.
Ref Type: Electronic Citation

[503] L. L. Koth, B. Alex, S. Hawgood, M. A. Nead, D. Sheppard, D. J. Erle, and D. G. Morris, Integrin beta6 mediates phospholipid and collectin homeostasis by activation of latent TGF-beta1, Am J Respir. Cell Mol. Biol., 37 (2007) 651-659.

[504] T. Tabata, H. Kawakatsu, E. Maidji, T. Sakai, K. Sakai, J. Fang-Hoover, M. Aiba, D. Sheppard, and L. Pereira, Induction of an epithelial integrin

alphavbeta6 in human cytomegalovirus-infected endothelial cells leads to activation of transforming growth factor-beta1 and increased collagen production, Am J Pathol., 172 (2008) 1127-1140.

[505] K. D. Pruitt, T. Tatusova, and D. R. Maglott, NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins, Nucleic Acids Res., 33 (2005) D501-D504.

[506] D. F. Conrad, T. D. Andrews, N. P. Carter, M. E. Hurles, and J. K. Pritchard, A high-resolution survey of deletion polymorphism in the human genome, Nat. Genet, 38 (2006) 75-81.

[507] R. Redon, S. Ishikawa, K. R. Fitch, L. Feuk, G. H. Perry, T. D. Andrews, H. Fiegler, M. H. Shapero, A. R. Carson, W. Chen, E. K. Cho, S. Dallaire, J. L. Freeman, J. R. Gonzalez, M. Gratacos, J. Huang, D. Kalaitzopoulos, D. Komura, J. R. MacDonald, C. R. Marshall, R. Mei, L. Montgomery, K. Nishimura, K. Okamura, F. Shen, M. J. Somerville, J. Tchinda, A. Valsesia, C. Woodwark, F. Yang, J. Zhang, T. Zerjal, J. Zhang, L. Armengol, D. F. Conrad, X. Estivill, C. Tyler-Smith, N. P. Carter, H. Aburatani, C. Lee, K. W. Jones, S. W. Scherer, and M. E. Hurles, Global variation in copy number in the human genome, Nature, 444 (2006) 444-454.

[508] A. P. Dempster, N. M. Laird, and D. B. Rubin, Maximum Likelihood from Incomplete Data via the EM Algorithm, Journal of the Royal Statistical Society. Series B (Methodological), 39 (1977) 1-38.

[509] J. F. Wetzels, L. A. Kiemeney, D. W. Swinkels, H. L. Willems, and H. M. den, Age- and gender-specific reference values of estimated GFR in Caucasians: the Nijmegen Biomedical Study, Kidney Int., 72 (2007) 632-637.

[510] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly, and P. C. Sham, PLINK: a tool set for whole-genome association and population-based linkage analyses, Am. J. Hum. Genet., 81 (2007) 559-575.

[511] J. C. Barrett, B. Fry, J. Maller, and M. J. Daly, Haploview: analysis and visualization of LD and haplotype maps, Bioinformatics, 21 (2005) 263-265.

[512] R. J. Pruim, R. P. Welch, S. Sanna, T. M. Teslovich, P. S. Chines, T. P. Gliedt, M. Boehnke, G. R. Abecasis, and C. J. Willer, LocusZoom: Regional visualization of genome-wide association scan results, Bioinformatics, 26 (2010) 2336-2337.

[513] J. Zhang and K. F. Yu, What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes, JAMA, 280 (1998) 1690-1691.

[514] E. Thorsby, Invited anniversary review: HLA associated diseases, Hum Immunol., 53 (1997) 1-11.

[515] I. Evseeva, K. K. Nicodemus, C. Bonilla, S. Tonks, and W. F. Bodmer, Linkage disequilibrium and age of HLA region SNPs in relation to classic HLA gene alleles within Europe, Eur. J Hum Genet, 18 (2010) 924-932.

[516] V. A. McKusick, On lumpers and splitters, or the nosology of genetic disease, Perspect. Biol. Med, 12 (1969) 298-312.

[517] E. Thorsby, HLA-Associated Disease Susceptibility - Which Genes Are Primarily Involved, The Immunologist, 3 (1995) 51-58.

[518] M. M. Fernando, C. R. Stevens, E. C. Walsh, P. L. De Jager, P. Goyette, R. M. Plenge, T. J. Vyse, and J. D. Rioux, Defining the role of the MHC in autoimmunity: a review and pooled analysis, PLoS. Genet, 4 (2008) e1000024.

[519] J. H. Moore and S. M. Williams, Traversing the conceptual divide between biological and statistical epistasis: systems biology and a more modern synthesis, Bioessays, 27 (2005) 637-646.

[520] D. C. Krakauer and J. B. Plotkin, Redundancy, antiredundancy, and the robustness of genomes, Proc. Natl. Acad. Sci. U. S. A, 99 (2002) 1405-1409.

[521] R. E. Lenski, C. Ofria, R. T. Pennock, and C. Adami, The evolutionary origin of complex features, Nature, 423 (2003) 139-144.

[522] K. I. Goh, M. E. Cusick, D. Valle, B. Childs, M. Vidal, and A. L. Barabasi, The human disease network, Proc. Natl. Acad. Sci. U. S. A, 104 (2007) 8685-8690.

[523] G. P. Page, V. George, R. C. Go, P. Z. Page, and D. B. Allison, "Are we there yet?": Deciding when one has demonstrated specific genetic causation in complex diseases and quantitative traits, Am J Hum Genet, 73 (2003) 711-719.

[524] A. S. Evans, Causation and disease: the Henle-Koch postulates revisited, Yale J Biol. Med, 49 (1976) 175-195.

[525] S. Falkow, Molecular Koch's Postulates Applied to Microbial Pathogenicity, Reviews of Infectious Diseases, 10 (1988) S274-S276.

[526] T. M. Rivers, Viruses and Koch's Postulates, J Bacteriol., 33 (1937) 1-12.

[527] A. B. HILL, THE ENVIRONMENT AND DISEASE: ASSOCIATION OR CAUSATION?, Proc. R. Soc. Med, 58 (1965) 295-300.

[528] D. T. Lykken, Statistical significance in psychological research, Psychol. Bull., 70 (1968) 151-159.

[529] F. Pociot, B. Akolkar, P. Concannon, H. A. Erlich, C. Julier, G. Morahan, C. R. Nierras, J. A. Todd, S. S. Rich, and J. Nerup, Genetics of type 1 diabetes: what's next?, Diabetes, 59 (2010) 1561-1571.

[530] A. Rubio-Tapia and J. A. Murray, Celiac disease, Curr. Opin. Gastroenterol., 26 (2010) 116-122.

[531] F. Granata, B. Balestrieri, A. Petraroli, G. Giannattasio, G. Marone, and M. Triggiani, Secretory phospholipases A2 as multivalent mediators of inflammatory and allergic disorders, Int. Arch. Allergy Immunol., 131 (2003) 153-163.

[532] V. Pedchenko, O. Bondar, A. B. Fogo, R. Vanacore, P. Voziyan, A. R. Kitching, J. Wieslander, C. Kashtan, D. B. Borza, E. G. Neilson, C. B. Wilson, and B. G. Hudson, Molecular architecture of the Goodpasture autoantigen in anti-GBM nephritis, N Engl J Med, 363 (2010) 343-354.

[533] L. Pelletier, F. Hirsch, J. Rossert, E. Druet, and P. Druet, Experimental mercury-induced glomerulonephritis, Springer Semin. Immunopathol., 9 (1987) 359-369.

[534] C. G. BECKER, E. L. BECKER, J. F. MAHER, and G. E. SCHREINER, Nephrotic syndrome after contact with mercury. A report of five cases, three after the use of ammoniated mercury ointment, Arch. Intern. Med, 110 (1962) 178-186.

[535] J. S. Cameron and J. R. Trounce, MEMBRANOUS GLOMERULONEPHRITIS AND THE NEPHROTIC SYNDROME APPEARING DURING MERSALYL THERAPY, Guys. Hosp. Rep., 114 (1965) 101-107.

[536] M. Kato, T. K. Neil, G. J. Clark, C. M. Morris, R. V. Sorg, and D. N. Hart, cDNA cloning of human DEC-205, a putative antigen-uptake receptor on dendritic cells, Immunogenetics, 47 (1998) 442-450.

[537] Powis SH, The genetics of glomerulonephritis and systemic disorders affecting the kidney, in: Flinter F, Maher E, and Saggar-Malik A (Eds.), The Genetics of Renal Disease, Oxford University Press, Oxford, 2003, pp. 417-454.

[538] M. P. Lefranc, V. Giudicelli, C. Ginestoux, J. Jabado-Michaloud, G. Folch, F. Bellahcene, Y. Wu, E. Gemrot, X. Brochet, J. Lane, L. Regnier, F. Ehrenmann, G. Lefranc, and P. Duroux, IMGT, the international ImMunoGeneTics information system, Nucleic Acids Res., 37 (2009) D1006-D1012.

[539] S. Ogahara, I. Michinaga, T. Hiratsuka, K. Abe, and S. Naito, [DNA typing of HLA-class II genes in idiopathic nephropathy], Nippon Rinsho., 50 (1992) 3072-3078.

[540] Y. Jin, S. A. Birlea, P. R. Fain, K. Gowan, S. L. Riccardi, P. J. Holland, C. M. Mailloux, A. J. D. Sufit, S. M. Hutton, A. madi-Myers, D. C.

Bennett, M. R. Wallace, W. T. McCormack, E. H. Kemp, D. J. Gawkrodger, A. P. Weetman, M. Picardo, G. Leone, A. Taieb, T. Jouary, K. Ezzedine, N. van Geel, J. Lambert, A. Overbeck, and R. A. Spritz, Variant of TYR and Autoimmunity Susceptibility Loci in Generalized Vitiligo, N Engl J Med, 362 (2010) 1686-1697.

[541] Y. Jin, S. A. Birlea, P. R. Fain, C. M. Mailloux, S. L. Riccardi, K. Gowan, P. J. Holland, D. C. Bennett, M. R. Wallace, W. T. McCormack, E. H. Kemp, D. J. Gawkrodger, A. P. Weetman, M. Picardo, G. Leone, A. Taieb, T. Jouary, K. Ezzedine, G. N. van, J. Lambert, A. Overbeck, and R. A. Spritz, Common variants in FOXP1 are associated with generalized vitiligo, Nat. Genet, 42 (2010) 576-578.

[542] L. Petukhova, M. Duvic, M. Hordinsky, D. Norris, V. Price, Y. Shimomura, H. Kim, P. Singh, A. Lee, W. V. Chen, K. C. Meyer, R. Paus, C. A. Jahoda, C. I. Amos, P. K. Gregersen, and A. M. Christiano, Genome-wide association study in alopecia areata implicates both innate and adaptive immunity, Nature, 466 (2010) 113-117.

[543] B. Leitinger and E. Hohenester, Mammalian collagen receptors, Matrix Biol., 26 (2007) 146-155.

[544] K. M. El, A. Harmer, J. Goodwin, B. Wagner, J. Shortland, C. Angel, and W. McKane, De novo membranous nephropathy associated with donor-specific alloantibody, Clin. Transplant., 22 (2008) 124-127.

[545] A. Snijders, D. G. Elferink, A. Geluk, A. L. van Der Zanden, K. Vos, G. M. Schreuder, F. C. Breedveld, R. R. de Vries, and E. H. Zanelli, An HLA-DRB1-derived peptide associated with protection against rheumatoid arthritis is naturally processed by human APCs, J Immunol., 166 (2001) 4987-4993.

[546] J. Churg, E. GRISHMAN, M. H. GOLSTEIN, S. L. YUNIS, and J. G. PORUSH, IDIOPATHIC NEPHROTIC SYNDROME IN ADULTS. A STUDY AND CLASSIFICATION BASED ON RENAL BIOPSIES, N Engl J Med, 272 (1965) 165-174.

[547] L. Friberg, Proteinuria and kidney injury among workmen exposed to cadmium and nickel dust; preliminary report, J Ind. Hyg. Toxicol., 30 (1948) 32-36.

[548] J. W. BEATTIE, Nephrotic syndrome following sodium bismuth tartrate therapy in rheumatoid arthritis, Ann. Rheum. Dis., 12 (1953) 144-146.

[549] J. C. Lee, M. DUSHKIN, E. J. EYRING, E. P. ENGLEMAN, and J. HOPPER, Jr., RENAL LESIONS ASSOCIATED WITH GOLD THERAPY; LIGHT AND ELECTRON MICROSCOPIC STUDIES, Arthritis Rheum., 8 (1965) 1-13.

[550] R. Lauwerys, A. Bernard, H. Roels, J. P. Buchet, J. P. Gennart, P. Mahieu, and J. M. Foidart, Anti-laminin antibodies in workers exposed to mercury vapour, Toxicol. Lett., 17 (1983) 113-116.

[551] M. P. Vessey, L. Villard-Mackintosh, and D. Yeates, Oral contraceptives, cigarette smoking and other factors in relation to arthritis, Contraception, 35 (1987) 457-464.

[552] L. Klareskog, P. Stolt, K. Lundberg, H. Kallberg, C. Bengtsson, J. Grunewald, J. Ronnelid, H. E. Harris, A. K. Ulfgren, S. Rantapaa-Dahlqvist, A. Eklund, L. Padyukov, and L. Alfredsson, A new model for an etiology of rheumatoid arthritis: smoking may trigger HLA-DR (shared epitope)-restricted immune reactions to autoantigens modified by citrullination, Arthritis Rheum., 54 (2006) 38-46.

[553] H. Kallberg, L. Padyukov, R. M. Plenge, J. Ronnelid, P. K. Gregersen, van der Helm-van Mil AH, R. E. Toes, T. W. Huizinga, L. Klareskog, and L. Alfredsson, Gene-gene and gene-environment interactions involving HLA-DRB1, PTPN22, and smoking in two subsets of rheumatoid arthritis, Am J Hum Genet, 80 (2007) 867-875.

[554] F. J. van der Woude, P. J. Hoedemaeker, G. M. van der, P. A. de Graeff, M. J. de, T. H. The, and G. K. van der Hem, Do food antigens play a role in the pathogenesis of some cases of human glomerulonephritis?, Clin. Exp. Immunol., 51 (1983) 587-594.

[555] N. Miura, Y. Mori, M. Yoshino, N. Suga, W. Kitagawa, H. Yamada, K. Nishikawa, and H. Imai, Membranous nephropathy (bubbling appearance and spike formation) without immunoglobulin deposition in a patient with systemic lupus erythematosus, Clin. Exp. Nephrol., 12 (2008) 475-478.

[556] S. Hisano, M. Matsushita, T. Fujita, M. Takeshita, and H. Iwasaki, Activation of the lectin complement pathway in post-streptococcal acute glomerulonephritis, Pathol. Int., 57 (2007) 351-357.

[557] R. Clynes, C. Dumitru, and J. V. Ravetch, Uncoupling of immune complex formation and kidney damage in autoimmune glomerulonephritis, Science, 279 (1998) 1052-1054.

[558] A. Augert, C. Payre, L. Y. de, J. Gil, G. Lambeau, and D. Bernard, The M-type receptor PLA2R regulates senescence through the p53 pathway, EMBO Rep., 10 (2009) 271-277.

[559] Childs B, Genetic Medicine - A Logic of Disease, The Johns Hopkins University Press, Baltimore & London 1999.

[560] S. P. Rose, Commentary: heritability estimates--long past their sell-by date, Int. J Epidemiol., 35 (2006) 525-527.

# 7 Publications that have arisen from the thesis

Stanescu HC*, Arcos-Burgos M*, Medlar A*, Bockenhauer D, Kottgen A, Dragomirescu L, Voinescu C, Patel N, Pearce K, Hubank M, Stephens HA, Laundy V, Padmanabhan S, Zawadzka A, Hofstra JM, Coenen MJ, den Heijer M, Kiemeney LA, Bacq-Daian D, Stengel B, Powis SH, Brenchley P, Feehally J, Rees AJ, Debiec H, Wetzels JF*, Ronco P*, Mathieson PW*, Kleta R*. Risk HLA-DQA1 and PLA(2)R1 alleles in idiopathic membranous nephropathy. N Engl J Med. 2011 Feb 17;364(7):616-26.

*contributed equally