Human gene copy number variation and infectious disease

Journal:	*Human Genetics*
Manuscript ID:	HumGen-14-0170.R1
Manuscript Type:	Review Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Hollox, EJ; University of Leicester, Hoh, Boon-Peng; Institute of Medical Molecular Biotechnology, Faculty of Medicine, Universiti Teknologi MARA, Sungai Buloh Campus, Jalan Hospital,
Keywords:	complex diseases, genome variability, genomic structure, genotyping, infections diseases - human genetics of, polymorphisms

University of Leicester

Department of Genetics
University of Leicester
University Road
Leicester LE1 7RH
UK

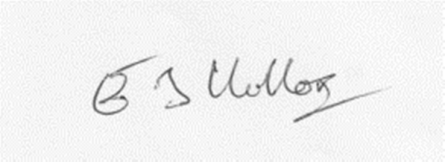
T +44 (0) 116 252 3374/3438
F +44 (0) 116 252 3378
E genetics@le.ac.uk
W http://www.le.ac.uk/genetics

Friday, 16 May 2014

Dear David

Please find attached our revised review manuscript entitled **"Human gene copy number variation and infectious disease"**. We have responded to the useful comments provided by the reviews, and have made some alterations to our review, visible under "track changes". We hope that we have addressed their comments satisfactorily, and that you now consider our review suitable for publication in *Human Genetics*.

With best wishes,



Ed Hollox, PhD, on behalf of both authors.

Human gene copy number variation and infectious disease

Edward J Hollox*, Department of Genetics, University of Leicester, Leicester, LE1 7RH UK

Boon-Peng Hoh, Institute of Medical Molecular Biotechnology, Faculty of Medicine, Universiti Teknologi MARA, Sungai Buloh Campus, Jalan Hospital, Sungai, Buloh, 47000, Malaysia

* Corresponding author:

Dr Ed Hollox

Department of Genetics,

University of Leicester

University Road,

Leicester LE1 7RH

Ejh33@le.ac.uk

(+44) 116 252 3407

Abstract

Variability in the susceptibility to infectious disease and its clinical manifestation can be determined by variation in the environment and by genetic variation in the pathogen and the host. Despite several successes based on candidate gene studies, defining the host variation affecting infectious disease has not been as successful as other multifactorial diseases. Both single nucleotide variation and copy number variation (CNV) in the host contribute to the host's susceptibility to infectious disease. In this review we focus on CNV, particularly on complex multiallelic CNV that is often not well characterised either directly by hybridisation methods or indirectly by analysis of genotypes and flanking single nucleotide variants. We summarise the well-known examples, such as alpha-globin deletion and susceptibility to severe malaria, as well as more recent controversies, such as the extensive CNV of the chemokine gene *CCL3L1* and HIV infection. We discuss the potential biological mechanisms that could underly any genetic association and reflect on the extensive complexity and functional variation generated by a combination of CNV and sequence variation, as illustrated by the Fc gamma receptor genes *FCGR3A*, *FCGR3B* and *FCGR2C*. We also highlight some understudied areas that might prove fruitful areas for further research.

Introduction

Infectious disease can be regarded as a complex multifactorial disease, with both genetic and environmental variation contributing to differing susceptibilities to infection, and differing effects of infection (Chapman and Hill 2012). Like any other multifactorial disease, the variation in who is infected has been divided into a genetic component and an environmental component, with heritability estimates equivalent to other multifactorial diseases (for example, heritability of malaria in Kenya is 0.25, equivalent to heritability of type 2 diabetes in Scandinavian countries (Mackinnon et al. 2005; Poulsen et al. 1999)). Like any other multifactorial disease, both rare and common genetic variation is likely to play a role. Genomewide association studies have been of limited success in identifying alleles that affect infectious disease susceptibility. We might infer from this that most genetic susceptibility is not determined by common alleles effectively assayed by these studies. Another complicating factor is that, at present, most GWASs of infectious disease are either using relatively small cohorts or by combining cohorts from different countries, and correcting for population differences statistically. The effect on the power to detect susceptibility variants, which may have different frequencies in regions with different patterns of LD in different populations, is unclear. For example, a large meta-analysis of over 5000 cases with severe malaria and almost 7000 controls established association, at genomewide significance, of only two loci already well-established in the literature *HBB* (encoding beta-globin) and *ABO* (encoding ABO blood group) (Band et al. 2013). Two types of variation, common copy number variation (CNV) and rare variants are not well tagged common alleles at SNPs, and are therefore not well assayed in current GWAS approaches ~~(Band et al. 2013)~~. Analyses of the effect of rare variants on infectious disease susceptibility have not yet been published, but there is increasing evidence that CNV has an important role.

CNV is simply different numbers of the same DNA sequence across different individuals, and includes simple deletion and duplications but also more complex multiallelic variation, with copy numbers ranging from 0 to 14, for example, as has been described for the *CCL3L1* gene (Aklillu et al. 2013; Walker et al. 2009). CNV can potentially affect phenotype in several different ways. Perhaps the simplest way is due to a gene dosage effect, where increased numbers of the same gene result in increased levels of mRNA and increased levels of protein. However, CNV can also create novel fusion genes, alter the distance of a gene from a regulatory element, or alter the number of protein-coding exons within a gene (figure 1). CNVs are just another form of variation, and subject to the same rules of population genetics as other variants. However, it is useful to distinguish CNVs into two categories which are based on mutational origin: non- recurrent and recurrent CNVs. Recurrent and non-recurrent CNVs are likely to have different mutation rates and different evolutionary trajectories.

Non-recurrent CNVs of any size can be generated by mechanisms such as non-homologous end-joining (NHEJ) or fork-stalling and template-switching (FosTes) ~~(NHEJ)~~, and, because they are often large, are more likely to affect genes and more likely to have an extremely deleterious phenotypic effect (Arlt et al. 2012). Large deletions, for example, will be rare in the population because negative selection acts to rapidly remove the deletion from the population. Recurrent infections are sometimes a symptom of multiple congenital abnormalities caused by a large chromosomal deletion. For example, patients with 22q11.2 deletion syndrome (OMIM #611867) often have recurrent infections due to low T-cell levels (McLean-Tooke et al. 2008), and patients with 16p12.2-p11.2 deletion syndrome (OMIM #613604) have recurrent ear infections (Okamoto et al. 2014).

There are also examples of rare small deletion alleles that segregate within a family, collectively known as Mendelian infectious susceptibility disease syndromes. For example, a small deletion (4 bp) in the *IFNGR1* gene leads to dominant susceptibility to mycobacterial infection (Jouanguy et al. 1999).

Recurrent multiallelic CNV is generated by non-allelic-homologous recombination (NAHR) between segmental duplications. These regions of segmental duplication are non-randomly distributed in the genome, and are particularly frequent in subtelomeric and pericentromeric regions, although they occur elsewhere in the genome (Bailey et al. 2002). These regions are difficult to assemble and often are associated with gaps in the reference genome assembly. The segmental duplications, sometimes in multiple copies, sponsor extensive NAHR and can harbour a large amount of CNV (Alkan et al. 2009; Redon et al. 2006). Importantly such regions can also harbour extensive sequence variation in the form of paralogous sequence variants (PSVs), which are differences between segmental duplications. Many of these segmental duplication-rich regions arose in the ancestor of great apes (Locke et al. 2003; Marques-Bonet et al. 2009; She et al. 2006) and diverged at a rate determined by a balance of new nucleotide substitution mutations occurring on either paralogue and the sequence-homogenising effect of gene conversion (Teshima and Innan 2004). Recurrent NAHR between these paralogues can shuffle these variants as well as generate CNV, and both copy number and sequence variation can contribute to diversity within these CNVs, which tend to be multiallelic and are sometimes associated with other polymorphic rearrangements such as inversions.

In this review we will focus on common multiallelic CNV where alleles are present at polymorphic frequencies within populations. Studies of complex multiallelic CNVs have shown that the mutation rate is high, several orders of magnitude higher than for single nucleotide polymorphisms, usually because of recurrent NAHR (Abu Bakar et al. 2009; Fu et al. 2010; Lam and Jeffreys 2006). This has two consequences: firstly CNVs can accumulate variation under mutation-drift balance resulting in a particular DNA sequence having a high level of standing variation and therefore a substrate for subsequent selection. Secondly, if the copy number allele is deleterious then, under mutation-selection balance, the strength of negative selection has to be stronger to remove a deleterious allele at a locus that has a higher mutation rate. If the negative selection is mild, or perhaps episodic, then a deleterious copy number allele might reach appreciable allele frequencies. This high mutation rate recurrently generating copy number alleles also can explain why multiallelic CNVs are not well tagged, or at least consistently tagged, by alleles at flanking SNPs. For an overview see (Locke et al. 2006) and for particular examples see (Hardwick et al. 2011; Hollox et al. 2009) but also see (Hardwick et al. 2014; Khan et al. 2013) as examples of multiallelic CNVs tagged by flanking SNPs.

In this review we aim to give an overview of the evidence that human host CNV affects susceptibility to infectious disease. Table 1 summarises the larger studies undertaken so far, and in the text we focus on the more well-established examples, as well as suggesting avenues for further research.

Host copy number variation and malaria

The most well-known example of copy number variation affecting infectious disease susceptibility is that of the α-globin genes *HBA1* and *HBA2*. α-globin is copy number variable, with most individuals having four copies per diploid genome, two copies of *HBA1* and two copies of *HBA2*. Homozygotes for α-globin deletion alleles (2 copies per diploid genome) have α^{+}-thalassemia, and individuals with one or two copies of duplications having no clinical phenotype (Harteveld and Higgs 2010). The

diploid copy number of α-globin can be related to distinct clinical features, as shown in figure 2, with a clear gene dosage effect: fewer genes, more severe symptoms.

The frequency of alleles with three *HBA* genes is low in most populations (<1%, although 5% in Greek Cypriots) (Goossens et al. 1980; Liu et al. 2000). However, the frequency of HBA deletions and alleles of even lower copy number are frequent in populations with endemic malaria, for example in non-malaria regions *HBA* deletion alleles are at a frequency of ~1%, while in malaria regions such as sub-Saharan Africa the frequency can reach 20%. This is due to the protective effects of the deletion allele against severe malaria (table 1, (May et al. 2007; Mockenhaupt et al. 2004; Williams et al. 2005). The α-globin locus is the only frequently polymorphic CNV where direct measures of mutation rate have been made using sperm. These suggest a mutation rate of $4.2x10^{-5}$ per sperm, which would predict a higher frequency than 1% in northern Europeans. This discrepancy is likely to be due to selection against the deletion allele in non-Europeans for a phenotype other than α^{+}thalassemia, which itself is often asymptomatic (Lam and Jeffreys 2006).

Two loci encoding receptors used by *Plasmodium falciparum* to gain entry into erythrocytes are also known to be copy number variable. The first, complement receptor 1 (*CR1*), shows copy number variation within the gene such that different alleles have different numbers of Long Homologous Repeats (LHRs), which encode 30 kDa extracellular domains involved in complement C3 binding (Dykman et al. 1983; Stoute 2011; Vik and Wong 1993). Four *CR1* alleles have been described with 3, 4, 5 or 6 LHR domains, with the 4 LHR domain allele (*CR1**F) being most frequent, at least in Europeans. The region of CR1 critical for *P.falciparum* erythrocyte invasion is within the N-terminal region of the protein, which is outside the copy number variable region of the gene (Park et al. 2014; Tham et al. 2010). However, the crystal structure of the receptor is not known and it is perhaps likely that higher numbers of the large LHR domains will interfere with the interaction with *P.falciparum*. Several SNPs within and surrounding *CR1*, including those responsible for the Knops blood group alleles (Moulds 2010), have been tested for association with various malaria clinical phenotypes, with inconsistent results (Thathy et al. 2005; Zimmerman et al. 2003). None has, to our knowledge, assessed the effect of *CR1* CNV.

The second locus includes the glycophorins *GYPA*, *GYPB*, and *GYPE* which are arranged as three tandem repeats of between 100-140kb in size with about 97% sequence identity. All three proteins are expressed on the surface of erythrocytes, and GYPA is a receptor for *P. falciparum* via erythrocyte-binding-antigen-175 (Duraisingh et al. 2003; Tolia et al. 2005). Extensive copy number variation of this region has been discovered in several genomewide surveys (table 2), but the nature and phenotypic consequences of this variation, particularly in terms of resistance to malaria, remain to be determined.

Host copy number variation and HIV

Genetic variation plays an important role in several clinical phenotypes related to infection by human immunodeficiency virus-1 (HIV-1), such as susceptibility to infection, time from infection to development of AIDS, and viral load levels at clinical latency (known as set point viral load) (Shea et al. 2013). A canonical example of this, and indeed a key example of a common variant having a strong effect on disease, is the CCR5**Δ**32 allele. This 32bp polymorphic deletion disrupts the reading frame of CCR5, the principal co-receptor for HIV, and has been associated with protection from HIV infection and a longer time from infection to AIDS (Carrington et al. 1999).

The first complex CNV to be associated with HIV-1 susceptibility involved the genes *CCL3L1* and *CCL4L1*. A large study on both horizontal transfer (adult-adult) and vertical transfer (mother-child) of HIV-1 suggested that higher copy number of *CCL3L1* led to protection against HIV-1 infection, and a lower copy number ~~let~~ led to susceptibility to HIV-1 infection (Gonzalez et al. 2005). The association was attractive, since *CCL3L1* and *CCL4L1* encode MIP-1-alpha and MIP1-beta, both ligands for CCR5, and could possibly compete with HIV-1 in occupying the CCR5 receptor. However, attempts to replicate the data have been inconsistent, with several large studies failing to replicate the data (table 1)(Bhattacharya et al. 2009; Lee et al. 2010). A recent meta-analysis of nine studies supported an association of lower *CCL3L1* with susceptibility to HIV (Liu et al. 2010), but this study did not critically analyse the quality of the published data used in the meta-analysis.

The difficulty in establishing a link between *CCL3L1* copy number and HIV-1 infection status is due to two related issues. Firstly, and perhaps most importantly, is the importance of the most accurate and precise estimates of diploid copy number for each individual in a cohort. This is important because we generally expect genetic effects to be small, and they are likely to be much smaller than effects due to noise or systematic bias of an assay. Noisy assays will generally lead to false negative results, because the small effect is swamped by random noise. Systematic bias in a study, either due to population stratification effects or technical biases (for example, assaying controls and cases in separate experiments), will increase the false positive rate as such biases will be interpreted as a real genetic effect. The strengths and weakness of different assays for *CCL3L1* copy number have been extensively discussed elsewhere (Cantsilieris and White 2013), and, importantly, these issues apply to all other multiallelic loci discussed in this review. Most studies show that, of high-throughput PCR-based assays, a form of quantitative PCR called the paralogue ratio test (PRT, (Armour et al. 2007; Walker et al. 2009), Figure 3~~a~~b) performs best, with digital droplet PCR also producing some promising results ((Hindson et al. 2011), Figure ~~3b~~3a).

The second issue is the poor understanding we have of the exact structure of the variation underlying the *CCL3L1* CNV. The consensus is that *CCL3L1* and *CCL4L1* are on a single copy number variable unit of around 90kb, which is tandemly repeated. This is supported by complete concordance of *CCL3L1* and *CCL4L1* copy number, as measured by PRT in Europeans (Walker et al. 2009). Such a structure is supported by fibre-FISH analysis of one European and one Yoruba individual (Perry et al. 2008). However, both fibre-FISH and PRT suggest a more complex structure in other sub-Saharan Africans (Aklillu et al. 2013). In a Yoruba trio, a copy number allele comprised of tandem copies of repeat units, some of which were larger and carrying the neighbouring *TBC1D3* gene. An alternative assembly of this region in provided by the Genome Reference Consortium in GRC38 supports this observation. In Ethiopians, PRT suggested that rare alleles in the population showed discordance between *CCL3L1* and *CCL4L1* copy number, reflecting further complexity, but unfortunately cells were not available from these individuals for fibre-FISH analysis (Aklillu et al. 2013).

An important clinical phenotype of HIV infection is the response to antiretroviral drugs, particularly given the implementation of highly-active antiretroviral therapy (HAART) to high-prevalence areas in Africa. Immune reconstitution, as it is called, following initiation of HAART varies between different people both in speed of CD4 recovery and final CD4 cell levels. Copy number of *CCL3L1*, together with a particular *CCR5* haplotype, was shown to affect immune reconstitution in European-Americans and African-Americans (Ahuja et al. 2008). However, a large study on therapy-naive

patients in Africa showed that *CCL3L1* affected immune reconstitution, but in the opposite way to the previous study, with individuals of low *CCL3L1* copy number showing an increased speed of CD4+ recovery (Aklillu et al. 2013). Therefore the role and possible mechanism of *CCL3L1* in immune reconstitution remains unclear. Furthermore, the role of copy number variation at genes involved in the metabolism of drugs (for example CYP2D6, (Bertilsson et al. 1993; Lundqvist et al. 1999)) in immune reconstitution has not been examined, and may be an interesting and important avenue to explore.

A similar paradox has been uncovered for a potential association of beta-defensin copy number and HIV-1 viral load and immune reconstitution. Beta-defensins are small peptides that have been shown to have antimicrobial and antiviral properties, including anti-HIV activity. A cluster of beta-defensins are copy number variable as a block, with copy number commonly ranging from 2-8 copies (Hollox et al. 2008a). We might expect that higher copy number of an antiviral peptide would result in a lower HIV viral load and improved immune reconstitution, but the opposite appears to be the case (Hardwick et al. 2012). We can reconcile this result with the fact that several beta-defensins, hBD-2 (encoded by DEFB4 gene) in particular, have chemoattractant activities to a variety of cells including dendritic cells (Yang et al. 1999) and Th-17 cells (Ghannam et al. 2011). This might allow HIV-1 to establish infection foci more effectively in individuals with higher copy number, and therefore higher levels of hBD-2 protein. Furthermore, unlike the chemoattractant activity of hBD2 (typically at 25-100 ng/ml), the anti-HIV-1 activity of hBD2 is only at unphysiological concentrations (> 4µg/ml) (Quiñones-Mateu et al. 2003; Sun et al. 2006), in contrast to physiological levels in serum and vaginal fluid of less than 10ng/ml in healthy and infected conditions and <150ng/ml in serum from inflammatory disease patients (Jansen et al. 2009; Jiang et al. 2012; Mitchell et al. 2013). This casts doubt on the role of hBD2 as an antiviral molecule in vivo.

The best-established role of CNV in HIV-1 infection remains that of Killer-cell immunoglobulin like receptor gene family (KIRs), members of which are expressed by natural killer cells and encoded by a gene cluster on chromosome 19 (Middleton and Gonzelez 2010). They bind to the major histocompatibility complex (MHC) class I ligands on the surface of target cells, and mediate either an inhibitory or activatory response by NK cells depending on the exact nature of the KIR. There are two main haplotypes of the KIR gene cluster, termed A and B, characterised by different complements of KIR genes. Within the B haplotype, there is further extensive variation in gene content and copy number, the functional consequences of which remain unclear. There has been a particular focus on the *KIR3DS1* gene which binds the MHC molecule encoded by the HLA Bw6 antigen (encoded by particular alleles of the HLA-B locus). Presence of the a particular KIR gene, *KIR3DS1*, in combination with a particular mismatched allele at HLA-B (Bw4-80I) significantly slows progression to AIDS. Furthermore, this effect is dependent on the copy number of *KIR3DS1*, with increased copies (in combination with a HLA-Bw4-80I allele) show lower levels of HIV-1 viral load in clinical latency phase (Alter et al. 2007; Bashirova et al. 2011; Pelak et al. 2011b). This is an interesting example of epistasis, where the effect of one allele is dependent on the presence of another allele at a physically unlinked locus. It may be the case that similar epistatic effects between copy number variable ligands and receptors exist.

Host copy number variation and other infectious diseases

HIV and malaria are the infectious diseases about which most research has been done, and most is known. However, other studies of host CNV and infectious disease have shown interesting results, and we discuss some below.

Haptoglobin (*HP*), a gene encoding an abundant acute-phase glycoprotein in the plasma, is one of the earliest blood serum proteins identified with copy number variation (CNV), carrying two alleles namely Hp1 and Hp2, spanning a 1.7kb intragenic duplication region in such a way that the Hp2 allele encodes a longer peptide chain than the Hp1 allele (Maeda et al. 1986; Maeda et al. 1983). HP shares ~94% sequence similarity with a neighbouring gene, haptoglobin-related protein (HPR,(Hardwick et al. 2014; Maeda et al. 1986)). The two genes originated in a triplication in Old World Monkeys resulting in three genes *HP*, *HPR* and *HPP* (haptoglobin primate)(McEvoy and Maeda 1988). All three genes are observed in extant Old World Monkeys and apes, except humans, who have lost the HPP gene after divergence from chimpanzees. In humans, HPR shows CNV between 2 and 4 copies per diploid genome, although an early study suggested rare copy numbers as high as 7 per diploid genome. The tandemly-arranged copies have been derived from a non-allelic homologous recombination event (NAHR) 1kb 3' to the genes so that the new copies are essentially identical to *HPR*, and are not an *HP-HPR* fusion gene (Maeda et al. 1986).

Haptoglobin protein (Hp) binds to the free haemoglobin (Hb) released by lysis of erythrocytes as a result of acute infection. The Hp-Hb complex is then cleared by binding to the macrophage scavenger receptor CD163, followed by endocytosis. Like Hp, haptoglobin related protein (Hpr) binds with free heme, but the Hpr-Hb complex binds to ApoL1 instead, and forms part of the trypanolytic lytic factor (TLF) providing innate immunity against trypanosomes (Nielsen and Moestrup 2009; Vanhollebeke et al. 2008). It is therefore conceivable that both *HP* and *HPR* gene copy number variations have experienced selection pressure in response to pathogens (Iskow et al. 2012; McDermid and Prentice 2006). The Hp2 allele has been suggested to protect against severe malaria (Atkinson et al. 2006) ~~(Atkinson et al., 2007)~~, but this has not been supported (Aucan et al. 2002; Cox et al. 2007), and there is no correlation between allele frequency and malarial endemicity, unlike alleles at other genes that are protective against malaria, such as the sickle-cell haemoglobin allele (Hardwick et al. 2014). There is suggestive evidence that duplication of the *HPR* gene is protective against human African trypanosomiasis (HAT), and the population distribution of the *HPR* duplication mirrors the distribution of *Trypanosoma* ~~*brucie*~~ *brucei gambiense*, which causes HAT. However, the duplication allele is at modest frequency (10%) even in HAT endemic areas and there is no evidence for recent direction selection (Hardwick et al. 2014; Rodriguez et al. 2012).

Fcγ receptors are the cell-surface receptors for immunoglobulin G (IgG), which is the most abundant Ig class in serum, constituting more than 75% of circulating immunoglobulin complex. They are expressed on various leucocytes, show extensive CNV and are categorized into three main classes: FcγRI, FcγRII, and FcγRIII (Nimmerjahn and Ravetch 2008).

FcγRI is a high affinity Fcγ receptor for monomeric IgG, with three highly similar genes (>95% identity) assembled in the human reference genome (*FCGR1A*, *FCGR1B* and *FCGR1C*) (Ernst et al. 1992; Maresco et al. 1996; van der Poel et al. 2011). The copy number of these genes has increased in the human lineage, and show copy number variation between individuals (Sudmant et al. 2010). All three genes are within the pericentromeric region of chromosome 1, and the human-specific copy number increase is likely to be associated with a pericentromeric inversion that distinguishes

human chromosome 1 from chimpanzee chromosome 1 (Maresco et al. 1998). The functional differences between the variants remain unclear: *FCGR1A* encodes a membrane-bound high-affinity IgG receptor, and *FCGR1B* and *FCGR1C* are predicted to encode soluble forms of the receptor, without a membrane spanning region. However, only *FCGR1A* has been shown to be definitely functional (van Vugt et al. 1999). Studies of knockout mice show that FcγRI protects against *Bordetella pertussis*, the causative agent of whooping cough (Ioan-Facsinay et al. 2002), and the entry of *E.coli* K1 into macrophages in neonatal meningitis is mediated by FcγRI (Mittal et al. 2010). Some studies have suggested the involvement of FcγRI in the pathogenesis of dengue fever (Chawla et al. 2013; Rodrigo et al. 2009; Rodrigo et al. 2006), and there is convincing evidence that FcγR1 mediates antibody-based protection against *Plasmodium falciparum* malaria (McIntosh et al. 2007). However, although shown to be a human specific CNV, with multiple copies, because the genes are within a complex pericentromeric region of chromosome 1 characterisation of the variation is challenging and nature of the CNV, and its relationship with disease, remains unknown.

FCGR2 and *FCGR3* genes encode low-affinity Fcγ receptors, FcγRII and FcγRIIII respectively, which bind to IgG-antigen immune complexes, and initiate either inhibitory or activatory signalling responses within the cell, depending on the type of receptor engaged (Willcocks et al. 2009). The human reference genome assembly shows two copies of the *FCGR3* gene, termed *FCGR3A* and *FCGR3B*, and three copies of the *FCGR2* gene, *FCGR2A*, *FCGR2B* and *FCGR2C*. *FCGR3A* and *FCGR3B* are distinguished by a premature stop mutation in *FCGR3B* which results in a truncated FcγrIIIB receptor without a transmembrane domain, attached to the membrane by GPI anchor (Ravetch and Perussia 1989). The receptors encoded by *FCGR3A* and *FCGR3B* are functionally distinct, with FCGR3B expressed primarily by neutrophils, and *FCGR3A* expressed in natural killer cells, dendritic cells, monocytes and macrophages . The two *FCGR3* genes are on a 82kb segmental duplication which shares ~98% sequence identity. *FCGR2A* and *FCGR2B* are at either end of the segmental duplications, with *FCGR2C* a fusion gene of *FCGR2A* and *FCGR2B*, spanning the two segmental duplications (figure 4). The segmental duplication is copy number variable, with deletions and duplications at appreciable frequency in different populations (Aitman et al. 2006; Hollox et al. 2009)(figure 4), resulting in CNV for *FCGR3B*, *FCGR3A* and *FCGR2C*, but not *FCGR2A* nor *FCGR2B* (Breunis et al. 2009; Reilly et al. 1994). There is a gene dosage effect for *FCGR3B*, where expression levels on the cell surface reflect gene copy number (Willcocks et al. 2008). Deletions of *FCGR3B* also alters the expression pattern of *FCGR2B*, which itself is not CNV, causing it to be expressed on natural killer cells ((Mueller et al. 2012; van der Heijden et al. 2012), figure 4 example 5)). Furthermore, because of the particular structure of the CNV, the copy number of *FCGR2C* varies in concert with *FCGR3B* and *FCGR3A* ((Machado et al. 2012), figure 4). Because CNV of *FCGR3B* is associated with altered cell expression of *FCGR2B* and CNV of *FCGR2C*, it is difficult to determine the cause of an observed association between *FCGR3B* copy number and disease. Sequence variation within paralogues adds another layer of complexity. For example, *FCGR2C* is a polymorphic pseudogene, (Gln57Stop, (Ernst et al. 2002; Metes et al. 1998), Figure 4 example 1) and most copies, at least in Europe, are non-functional; also, extensive variation within *FCGR3B* alters its affinity for different IgG classes ((Ory et al. 1989a; Ory et al. 1989b; Salmon et al. 1990), Figure 4 example 2).

The low affinity Fcγ receptors trigger a number of immuno-regulatory functions, including degranulation, phagocytosis and regulation of antibody production. GWASs have shown association of SNPs within *FCGR2A* with susceptibility to *Helicobacter pylori* and Kawasaki disease (Khor et al. 2011; Mayerle et al. 2013). However, at present it is unclear whether these associations represent

variation within *FCGR2A* itself, sequence variation within the neighbouring CNV or the CNV itself. Such associations are difficult to disentangle, particularly because of the unclear relationship between copy number variation between flanking SNPs and CNVs. In general, such complex multiallelic CNVs are not well tagged by flanking SNP alleles (Hollox et al. 2009; Locke et al. 2006), but, given a large enough dataset and a particular CNV mutational history, certain CNV alleles may be partially tagged by flanking SNP alleles or SNP haplotypes.

Both Fcγ receptors regions contain interesting functional candidate genes for infections that undergo antibody-dependent enhancement, such as dengue virus (Littaua et al. 1990). During dengue infection, the host immune response triggers a mechanism called antibody dependent enhancement (ADE), whereby a heterotypic antibody (one from previous infection of a different dengue serotype) binds to the virus from the secondary infection but does not neutralize it. The virus then enters the cell like a "Trojan Horse" – forming the virus-antibody immunoglobulin complex, which enter the cell via binding to the Fcγ receptors (Chan et al. 2011; Nimmerjahn and Lux 2014), therefore creates a phenomenon called "cytokine storm" which ultimately could result in severe form of dengue. Consequently, CNV of Fcγ receptor gene clusters may have a significant impact on dengue severity, possibly via its alteration of gene dosage.

Host CNV and Autoimmune Disease

Many of the immunity genes that show CNV have been investigated in the context of susceptibility to autoimmune diseases (Olsson and Holmdahl 2012; Schaschl et al. 2009). Perhaps the best-established is the CNV of complement C4 within the MHC region at chromosome 6p21.32. A tandemly-arranged 33kb segmental duplication with >99% sequence identity carries the paralogous genes C4A and C4B, and both genes vary in copy number independently of each other (Belt et al. 1985; Chung et al. 2002; Dykman et al. 1983; Yu et al. 1986). Lower copy number of C4 has ~~become~~ been identified as a risk factor to several autoimmune diseases, particularly systemic lupus erythematous (SLE) (Yang et al. 2007).

Other examples of CNVs discussed in this review have also been associated with autoimmune and inflammatory diseases, but results have not been conclusive and often not reproducible. For example, *CCL3L1* has become a candidate gene of interest in many autoimmune diseases, including rheumatoid arthritis, SLE and asthma, but these results have not been replicated and doubts have been raised on the accuracy of the methods used for typing this CNV (Carpenter et al. 2011; Lee et al. 2011; Mamtani et al. 2008; McKinney et al. 2008; Nordang et al. 2012). A similar story exists for Fcγ receptor gene CNV, although there is more convincing evidence of an association of low *FCGR3B* copy number with rheumatoid arthritis (Graf et al. 2012; Marques et al. 2010; McKinney and Merriman 2012; Robinson et al. 2012), and with SLE (Aitman et al. 2006). The only large-scale, replicated CNV study so far is the association of high beta-defensin copy number with psoriasis (Hollox et al. 2008b; Stuart et al. 2012).

Identification of CNVs affecting the susceptibility to autoimmune and inflammatory diseases- informs study on CNVs and infectious diseases, but we might expect the same loci to be involved. The relationship between genetic variation in infectious disease susceptibility and autoimmune/inflammatory disease is most clear in the MHC region where, for example, the

susceptibility allele for inflammatory disease ankylosing spondylitis HLA-B*27 is associated with delayed progression to AIDS in HIV-infected patients (Kaslow et al. 1996; Schlosstein et al. 1973). In recent GWASs -an overlap between loci containing susceptibility alleles for Crohn's disease and loci affecting response to pathogens has been identified (Cagliani et al. 2013; Jostins et al. 2012). Together this points to a fine balance between a strong and appropriate immune response to challenge by a pathogen and an excessive and inappropriate response leading to autoimmune disease (Figure 5). Genetic variation at immune genes will alter the fulcrum position of this balance and lead to susceptibility to infectious or inflammatory diseases. It should be remembered that the genes encoding the immune response have evolved in conditions where co-morbidity, particularly with helminths, was the norm. One hypothesis is that removal of helminth infectious burden in modern populations has resulted in a immune system prone to autoimmune disease (Sironi and Clerici 2010). We would predict that new, derived alleles at immunity genes would be protective against autoimmune disease and susceptible to infectious disease and old, ancestral alleles susceptible to autoimmune disease and protective against infectious disease (Di Rienzo and Hudson 2005). Support from this model comes from analysis of sequence variation within the *FCGR2/3* CNV, where the ancestral HNA1a and *FCGR2C**Gln57 variants are associated with helminth diversity across human populations (Machado et al. 2012), and have been associated with susceptibility to ideopathic pulmonary fibrosis (Bournazos et al. 2010) and the haematological autoimmune disease idiopathic thrombocytopenic purpura idiopathic respectively (Breunis et al. 2008). There is some suggestive evidence that *FCGR2C**Gln57 is protective against tuberculosis co-infection in an HIV-infected Ethiopian cohort, but this may be confounded by *FCGR3B* copy number, and the causative allele is unclear (Machado et al. 2013).

Formatted: Font: (Default) Calibri, Italic, Font color: Custom Color(RGB(0,0,10))

Summary and Challenges

Several studies have associated complex CNV with susceptibility to infectious disease, or another clinical parameter related to infectious disease, such as progression (table 1). Taken together, these point to a potentially important role of common CNV in determining an individual's susceptibility and response to infectious disease. Currently, complex CNV studies are focused on a candidate gene approach, which are inherently biased and often yield false positive results. Even for candidate genes, accurately typing a single complex CNV is not straightforward. Real-time quantitative PCR approaches, in particular, are generally not robust enough for accurate, precise copy number calling, and many studies accept results from these assays uncritically, without internal controls or validation. Other more robust methods include the paralogue ratio test (PRT) and digital droplet PCR but each have strengths and limitations in terms of cost per sample, amount of DNA used and ease of assay design. In our studies, PRT is cost effective, with assay-design made more straightforward by an online database (Veal et al. 2013), but extensive validation of each assay, and ideally multiple assays for the same CNV, is still required.

The unbiased assessment of genetic variation using genomewide approaches would be more likely to yield robust reproducible associations, in theory. However, unlike genomewide SNP genotyping, which is very reliable, hybridisation data from SNP chips often cannot reliably type complex copy number variation because of inherent noisiness of SNP hybridisation data and systematic differences between cohorts and batches that are poorly understood. Similar issues affect array comparative genomic hybridisation (Pinto et al. 2011). The great hope is sequence read-depth mapping from new sequencing technologies, which can yield sequence data and copy number data based on the

number of reads that map to a given region on the reference genome (Krumm et al. 2012). However, this assumes the sequence is actually present in the reference genome, requires high-coverage whole genome data, and is computationally intensive.

We might expect such complex CNV loci to be under selection if different copy number alleles affected the susceptibility to infectious disease. For non-CNV regions, methods for testing for the action of selection from population diversity data are well-developed, because of the use of the infinite site model, which is an accurate mutation model for single nucleotide polymorphisms , allowing for the development of a robust neutral null model (Nielsen 2005). For CNV loci, this is not the case, and most analyses have relied on population differentiation statistics as an indicator of selection (Perry et al. 2007; Redon et al. 2006). Instead of the infinite-sites model, appropriate null-models of variation in copy number might be based on the stepwise-mutation model (SMM), which is much used for modelling the neutral behaviour of microsatellites because it can allow for recurrent mutations, but the SMM assumes all copies are equivalent, therefore ignoring sequence differences between copies. Some work has been published using coalescent theory to model sequence variation in a region where a duplicated copy is polymorphic in the population, and there are therefore just two CNV alleles: one-copy alleles and two-copy alleles (Teshima and Innan 2012; Thornton 2007). A neutral model combining a mutational model of DNA sequence within a CNV together with a mutational model of the CNV itself will be a real advance and a very useful analytical tool, and could use either coalescent simulation or forward simulation approaches.

We would also like to call for the continued support of sample collection, particularly in large studies of the epidemiology of infectious disease where the role of host variation can often be overlooked. Both host variation, pathogen variation and environmental variation should be studied together rather than assembling the pieces after the research has been done. We would also support internationalisation of research – no country has a monopoly on methods or expertise, and restricting sharing of data or resources impedes our understanding, to the detriment not only of the researchers but of individuals with the disease. Nevertheless, we anticipate an exciting future for studying host complex CNV and infectious disease susceptibility: the combination of technical challenge, biological interest and clinical importance is what makes the field of complex CNV so exciting.

Acknowledgements

B.P.H. thanks the Ministry of Long Term Research Grant Scheme (LRGS) for Infectious Disease, 2011-14, Ministry of Higher Education, Malaysia [600-RMI/LRGS 5/3 (3/2011)]. E.J.H. thanks colleagues for support, and Razan Abujaber and Ezgi Kucikkilic for comments on the manuscript.

Figure Legends

Figure 1 **The different possible mechanisms of CNV affecting phenotype**

a) Gene dosage effect. The CNV affects an entire gene, altering the number of copies of the full gene and altering the total amount of mRNA and protein encoded by the gene. Example: beta-defensins (Hollox et al. 2003; Jansen et al. 2009)

b) Position effect. The CNV affects the distance between a regulatory element and the gene. The regulatory element can be either an enhancer or a repressor, or affect tissue specificity. Example: HoxD in mouse (Montavon et al. 2012).

c) Fusion gene. A deletion caused by unequal crossing over between the two copies of DNA sequence results in a fusion gene. If the genes on copy 1 and copy 2 are identical, this will have ~~an~~ no effect, but if they have diverged in sequence or regulation the new fusion gene may have novel effects. Example: butyrophilin-like genes (Aigner et al. 2013).

d) Extra protein coding domains. Variation in the number of tandem repeats of coding exons within a gene alters the number of functional protein domains, and final size of the protein. Example: ~~Complement~~ complement receptor 1 (Wong et al. 1989).

Figure 2 **Copy number of alpha globin (*HBA*) and different clinical phenotypes**

Formatted: Font: (Default) Calibri, Italic, Font color: Custom Color(RGB(0,0,10))

Different observed diploid copy numbers of HBA are shown in descending order, together with the schematic gene arrangement (dark blue representing alpha-1-globin and pale blue representing alpha-2-globin), and the blood disorder and infectious disease phenotypes of each copy number.

Figure 3 **Two robust PCR-based methods to measure CNV**

a) Digital droplet PCR. This approach uses two TaqMan assays to detect presence of CNV and/or reference loci in emulsion droplets after emulsion PCR amplification. The thousands of droplets are effectively miniaturised PCR reactions, with the genomic DNA at limiting dilution such that most droplets will not contain a DNA molecule that can be amplified by the CNV or reference primers.

b) Paralogue ratio test (PRT). This approach uses carefully designed primers that amplify a region within the CNV of interest but also at a diploid reference control locus. Such primers are often targeted to diverged repeat elements, and the products amplified from the reference and CNV need to be distinguished by size, so that they can be detected by capillary electrophoresis and quantified. In practice, at least two PRTs are designed per CNV locus, and positive controls of known copy number used to normalise for variation between experiments.

Figure 4 **Genetic structure of low-affinity Fc gamma receptor region**

A duplication of the FCGR3 gene, probably caused by NAHR between *FCGR2A* and *FCGR2B* led to two *FCGR3* genes, *FCGR3A* and *FCGR3B*, and a *FCGR2C* fusion gene. This has been the substrate for further complex copy number and sequence variation at this locus, examples indicated by different numbers. All these have been observed and published (cited in the text), with the exception of the structure of the known duplicated alleles (examples 6 and 7) which is predicted as the reciprocal of the deletion structure following NAHR.

Figure 5 A hypothetical model of the relationship between genetic variation, infectious disease and auto-immune disease

The relationship between infection and the immune response is shown as a balance, with genetic variation mediating the response by moving the position of the fulcrum of the balance.

Table 1 All studies – large study groups (n>250)

Note that, for several studies, e.g. Gonzalez et al, although one result is reported several cohorts were tested and analysed using different approaches.

Disease	**Locus**	**Clinical phenotype/study design**	**Technology**	**P value**	**Stength of association - odds ratio or B value (95% CI)**	**Sample size**	**Population**	**Result**	**Reference**
Malaria	Alpha-globin	Severe malaria cases / random controls	Multiplex PCR	0.013	0.73 (0.57-0.94)	655 cases, 648 controls	Kenyan	Reduced prevalence of malaria in α+ thalassemia heterozygotes	(Williams et al. 2005)
		Severe malaria cases / random controls (under 5 years old)	Multiplex PCR	0.04	0.74 (0.56-0.98)	261 cases, 1093 controls	Ghanaian	Reduced prevalence of malaria in α+ thalassemia heterozygotes	(Mockenhaupt et al. 2004)

		Severe malaria cases / random controls	Multiplex PCR	0.03	0.82 (0.69-0.96)	2591 cases, 2048 controls	Ghanaian	Reduced prevalence of malaria in α+ thalassemia heterozygotes	(May et al. 2007)
	CCL3L1	Malaria parasite load and number of clinical episodes	PRT	Not significant	Not applicable	922 individuals of known relatedness	Tanzanians	No association	(Carpenter et al. 2012)
HIV	~~b~~beta-defensin	Immune reconstitution after HAART	PRT	0.003	B=-24.68 CD4 cells/mm3 (--40.79 to -8.58) low/high copy number class	2250 observations	Ethiopians and Tanzanians	Low copy number associated with stronger response to treatment	(Hardwick et al. 2012)
		Viral load prior to HAART treatment	PRT	0.005	B= $3.5x10^4$ HIV copies/ml per beta-defensin copy ($1.05 x10^4$-$5.90 x10^4$)	563	Ethiopians and Tanzanians	High copy number associated with higher viral load	(Hardwick et al. 2012)

Formatted: Font: (Default) Calibri, Font color: Custom Color(RGB(0,0,10))

	CCL3L1	Immune reconstitution after HAART (chronic infection)	Real-time qPCR	0.0041	Not given	441 individuals	Americans of mixed ethnicity	Low copy number associated with weaker response to treatment	(Ahuja et al. 2008)
		Immune reconstitution after HAART (late infection)	PRT	0.012	Beta=-4.75 CD4 cells/mm3 per *CCL3L1* copy (-1.05, -8.46)	1692 observations	Ethiopians and Tanzanians	Low copy number associated with stronger response to treatment	(Aklillu et al. 2013)
		Viral load prior to HAART treatment	PRT	Not significant	Not applicable	656	Ethiopians and Tanzanians	No association	(Aklillu et al. 2013)
		HIV infection status	qPCR	5.4x10-6	3.79(2.13-6.73)	409 cases, 394 controls	African-Americans	2 or fewer copies associated with HIV-infection	(Gonzalez et al. 2005)
		HIV infection status	qPCR	Not significant	Not applicable	411	African-Americans and European-	No association	(Shao et al. 2007)

							Americans		
		Mother to infant HIV transmission	Real-time qPCR	0.004	Not given	314 mother-child pairs	South African black	Low copy number associated with HIV infection	(Kuhn et al. 2007)
		HIV infection status	Real-time qPCR	0.0004	Not given	95 cases, 205 controls	Japanese	2 or fewer copies associated with HIV-infection	(Nakajima et al. 2007)
		HIV viral load at set point	Real-time qPCR	Not significant	Not applicable	1042	European American	No association	(Urban et al. 2009)
		HIV viral load after resolution of initial infection	Real-time qPCR	Not significant	Not applicable	740	European American	No association	(Bhattacharya et al. 2009)
		Immune reconstitution after HAART	Real-time qPCR	Not significant	Not applicable	527	European American	No association	(Bhattacharya et al. 2009)
		HIV status in intravenous drug users, case-control	Real-time qPCR	0.006	2.04 (1.23-3.45)	374	Estonian	2 or fewer copies associated	(Huik et al. 2010)

		study						with HIV-infection	
		Infected/uninfected case control study	Real-time qPCR	Not significant	Not applicable	153 cases, 159 controls	Zimbabwean	No association	(Larsen et al. 2012)
	FCGR3A/FCGR3B	Immune reconstitution after HAART	PRT/REDVR	Not significant	Not applicable	1823 (individual measurements)	Ethiopians and Tanzanians	No association	(Machado et al. 2013)
		Viral load prior to HAART treatment	PRT/REDVR	Not significant	Not applicable	684	Ethiopians and Tanzanians	No association	(Machado et al. 2013)
	Effective KIR3DS1 (in combination with HLA-B allele)	Viral load set point,	Real-time qPCR	$4.2x10^{-6}$	Not given	1429	European	Increase in effective copy number lowers VL set point	(Pelak et al. 2011a)
Hepatitis C	*CCL3L1*	Absence/presence case control study	Real-time qPCR	0.02	1.54	254 cases, 210 controls	Germans of European ancestry	2 or fewer copies associated with Hepatitis C	(Grünhage et al. 2010)
Tuberculosis	*FCGR3A/FCGR3B*	Absence/presence TB in HIV-infected	PRT/REDVR	0.002	1.454 (1.148-	442 cases, 278 controls	Ethiopian	Lower mean FCGR3B copy	(Machado et al. 2013)

		patients case-control study			1.841)			number in cases	
	FCGR3A/FCGR3B	Absence/presence TB in HIV-infected patients case-control study	PRT/REDVR PRT	Not significant	Not applicable	145 cases, 202 controls	Tanzanian	No association	(Machado et al. 2013)
	CCL3L1	Absence/presence of diagnosed TB case-control study	PRT	Not significant	Not applicable	141 cases, 341 controls	!Xhosa ~~No association~~	~~!Xhosa~~ No association	(Carpenter et al. 2014)
				Not significant	Not applicable	621 cases, 511 controls	~~No~~ Peruvian ~~association~~	~~Peruvian~~No association	

Table 2 Identification of CNV in the glycophorin cluster from different genomewide studies

Method	Population	Chromosome 4 coordinates (hg18 assembly)	Reference
oligo aCGH	YRI 2 gain	144932472-144973708	(Conrad et al. 2009)
	YRI 2 loss	145132085-145238644	
	YRI 1 gain	145173415-145201955	
	YRI 1 loss	145219660-145229403	
BAC aCGH	6 Gains 11 Losses total in 1 CEU, 3CHB and 11 YRI	144705898-145489197	(Redon et al. 2006)
Fosmid end sequencing	YRI deletion (NA18507)	144993427-145265979	(Kidd et al. 2008)
	CHB insertion (NA18555)	145022821-145056523	
	CHB insertion (NA18555)	144866282-144899908	
	YRI insertion (NA19240)	144921716-144955874	
BAC aCGH	CHB 2 loss, YRI 2 loss	145097717-145279154	(Locke et al. 2006)
	CHB 1 gain, CEU 1 gain 1 loss, JPT 3 gain 3 loss, YRI 5 loss	144914358-145076420	
	CHB 1 gain	144877610-145043981	

HapMap samples were used in all studies: YRI, Yoruba from Ibadan, Nigeria; CHB, Chinese from Beijing, China; JPT, Japanese from Tokyo, Japan; CEU, European-Americans from Utah, USA.

References

Abu Bakar S, Hollox EJ, Armour JA (2009) Allelic recombination between distinct genomic locations generates copy number diversity in human beta-defensins. Proc Natl Acad Sci U S A 106: 853-8. doi: 10.1073/pnas.0809073106

Ahuja SK, Kulkarni H, Catano G, Agan BK, Camargo JF, He W, O'Connell RJ, Marconi VC, Delmar J, Eron J, Clark R, Frost S, Martin J, Ahuja S, Deeks S, Little S, Richman D, Hecht F, Dolan M et al.(2008) CCL3L1-CCR5 genotype influences durability of immune recovery during antiretroviral therapy of HIV-1–infected individuals. Nature medicine 14: 413-420.

Aigner J, Villatoro S, Rabionet R, Roquer J, Jiménez-Conde J, Martí E, Estivill X (2013) A common 56-kilobase deletion in a primate-specific segmental duplication creates a novel butyrophilin-like protein. BMC genetics 14: 61.

Aitman TJ, Dong R, Vyse TJ, Norsworthy PJ, Johnson MD, Smith J, Mangion J, Roberton-Lowe C, Marshall AJ, Petretto E, Hodges MD, Bhanghal G, Patel SG, Sheehan-Rooney K, Duda M, Cook PR, Evans DJ, Domin J, Flint J, Boyle JJ, Pusey CD, Cook HT (2006) Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. Nature 439: 851-855.

Aklillu E, Odenthal-Hesse L, Bowdrey J, Habtewold A, Ngaimisi E, Yimer G, Amogne W, Mugusi S, Minzi O, Makonnen E et al. (2013) CCL3L1 copy number, HIV load, and immune reconstitution in sub-Saharan Africans. BMC infectious diseases 13: 536.

Alkan C, Kidd JM, Marques-Bonet T, Aksay G, Antonacci F, Hormozdiari F, Kitzman JO, Baker C, Malig M, Mutlu O et al. (2009) Personalized copy number and segmental duplication maps using next-generation sequencing. Nature Genetics 41: 1061-1067.

Alter G, Martin MP, Teigen N, Carr WH, Suscovich TJ, Schneidewind A, Streeck H, Waring M, Meier A, Brander C et al. (2007) Differential natural killer cell–mediated inhibition of HIV-1 replication based on distinct KIR/HLA subtypes. The Journal of experimental medicine 204: 3027-3036.

Arlt MF, Wilson TE, Glover TW (2012) Replication stress and mechanisms of CNV formation. Current opinion in genetics & development 22: 204-210.

Armour JAL, Palla R, Zeeuwen PLJM, den Heijer M, Schalkwijk J, Hollox EJ (2007) Accurate, high-throughput typing of copy number variation using paralogue ratios from dispersed repeats. Nucleic acids research 35: e19-e19.

Atkinson SH, Rockett K, Sirugo G, Bejon PA, Fulford A, O'Connell MA, Bailey R, Kwiatkowski DP, Prentice AM (2006) Seasonal childhood anaemia in West Africa is associated with the haptoglobin 2-2 genotype. PLoS medicine 3: e172.

Aucan C, Walley AJ, Greenwood BM, Hill AV (2002) Haptoglobin genotypes are not associated with resistance to severe malaria in The Gambia. Transactions of the Royal Society of Tropical Medicine and Hygiene 96: 327-328.

Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE (2002) Recent segmental duplications in the human genome. Science 297: 1003-1007.

Band G, Le QS, Jostins L, Pirinen M, Kivinen K, Jallow M, Sisay-Joof F, Bojang K, Pinder M, Sirugo G (2013) Imputation-based meta-analysis of severe malaria in three African populations. PLoS genetics 9: e1003509.

Bashirova AA, Thomas R, Carrington M (2011) HLA/KIR restraint of HIV: surviving the fittest. Annual review of immunology 29: 295-317.

Belt KT, Yu CY, Carroll MC, Porter RR (1985) Polymorphism of human complement component C4. Immunogenetics 21: 173-180.

Bertilsson L, Dahl M, Sjöqvist F, Åberg-Wistedt A, Humble M, Johansson I, Lundqvist E, Ingelman-Sundberg M (1993) Molecular basis for rational megaprescribing in ultrarapid hydroxylators of debrisoquine. The Lancet 341: 63.

Bhattacharya T, Stanton J, Kim E-Y, Kunstman KJ, Phair JP, Jacobson LP, Wolinsky SM (2009) Ccl3l1 and hiv/aids susceptibility. Nature medicine 15: 1112-1115.

Bournazos S, Bournazou I, Murchison JT, Wallace WA, McFarlane P, Hirani N, Simpson AJ, Dransfield I, Hart SP (2010) Fcγ receptor IIIb (CD16b) polymorphisms are associated with susceptibility to idiopathic pulmonary fibrosis. Lung 188: 475-481.

Breunis WB, van Mirre E, Bruin M, Geissler J, de Boer M, Peters M, Roos D, de Haas M, Koene HR, Kuijpers TW (2008) Copy number variation of the activating FCGR2C gene predisposes to idiopathic thrombocytopenic purpura. Blood 111: 1029-1038.

Breunis WB, van Mirre E, Geissler J, Laddach N, Wolbink G, van der Schoot E, de Haas M, de Boer M, Roos D, Kuijpers TW (2009) Copy number variation at the FCGR locus includes FCGR3A, FCGR2C and FCGR3B but not FCGR2A and FCGR2B. Human Mutation 30: E640-E650.

Cagliani R, Pozzoli U, Forni D, Cassinotti A, Fumagalli M, Giani M, Fichera M, Lombardini M, Ardizzone S, Asselta R et al. (2013) Crohn's disease loci are common targets of protozoa-driven selection. Molecular biology and evolution 30: 1077-1087.

Cantsilieris S, White SJ (2013) Correlating multiallelic copy number polymorphisms with disease susceptibility. Human Mutation 34: 1-13.

Carpenter D, Färnert A, Rooth I, Armour JA, Shaw M-A (2012) < i> CCL3L1</i> copy number and susceptibility to malaria. Infection, Genetics and Evolution: 1147-1154.

Carpenter D, Taype C, Goulding J, Levin M, Eley B, Anderson S, Shaw M-A, Armour JA (2014) CCL3L1 copy number, CCR5 genotype and susceptibility to tuberculosis. BMC medical genetics 15: 5.

Carpenter D, Walker S, Prescott N, Schalkwijk J, Armour JA (2011) Accuracy and differential bias in copy number measurement of CCL3L1 in association studies with three auto-immune disorders. BMC genomics 12: 418.

Carrington M, Dean M, Martin MP, O'Brien SJ (1999) Genetics of HIV-1 infection: chemokine receptor CCR5 polymorphism and its consequences. Human molecular genetics 8: 1939-1945.

Chan KR, Zhang SL-X, Tan HC, Chan YK, Chow A, Lim APC, Vasudevan SG, Hanson BJ, Ooi EE et al. (2011) Ligation of Fc gamma receptor IIB inhibits antibody-dependent enhancement of dengue virus infection. Proceedings of the National Academy of Sciences 108: 12479-12484.

Chapman SJ, Hill AVS (2012) Human genetic susceptibility to infectious disease. Nat Rev Genet 13: 175-188.

Chawla T, Chan KR, Zhang SL, Tan HC, Lim AP, Hanson BJ, Ooi EE (2013) Dengue virus neutralization in cells expressing Fc gamma receptors. PLOS one 8: e65231.

Chung EK, Yang Y, Rennebohm RM, Lokki M-L, Higgins GC, Jones KN, Zhou B, Blanchong CA, Yu CY (2002) Genetic Sophistication of Human Complement Components< i> C4A</i> and< i> C4B</i> and< i> RP-C4-CYP21-TNX</i>(RCCX) Modules in the Major Histocompatibility Complex. The American Journal of Human Genetics 71: 823-837.

Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, Fitzgerald T, Hu M, Ihm C, Kristiansson K, Macarthur D, Macdonald J, Onyiah I, Pang A, Robson S, Stirrups K, Valsesia A, Walter K, Wei J, Consortium" WTCC, Tyler-Smith C, Carter N, Lee C, Scherer S, Hurles M (2009) Origins and functional impact of copy number variation in the human genome. Nature 464: 704-712.

Cox SE, Doherty C, Atkinson SH, Nweneka CV, Fulford AJ, Ghattas H, Rockett KA, Kwiatkowski DP, Prentice AM (2007) Haplotype association between haptoglobin (Hp2) and Hp promoter SNP (A-61C) may explain previous controversy of haptoglobin and malaria protection. PLOS one 2: e362.

Di Rienzo A, Hudson RR (2005) An evolutionary framework for common diseases: the ancestral-susceptibility model. Trends in genetics 21: 596-601.

Duraisingh MT, Maier AG, Triglia T, Cowman AF (2003) Erythrocyte-binding antigen 175 mediates invasion in Plasmodium falciparum utilizing sialic acid-dependent and-independent pathways. Proceedings of the National Academy of Sciences 100: 4796-4801.

Dykman TR, Cole JL, Iida K, Atkinson JP (1983) Polymorphism of human erythrocyte C3b/C4b receptor. Proceedings of the National Academy of Sciences 80: 1698-1702.

Ernst L, van de Winkel J, Chiu I, Anderson C (1992) Three genes for the human high affinity Fc receptor for IgG (Fc gamma RI) encode four distinct transcription products. Journal of Biological Chemistry 267: 15692-15700.

Ernst LK, Metes D, Herberman RB, Morel PA (2002) Allelic polymorphisms in the FcγRIIC gene can influence its function on normal human natural killer cells. Journal of molecular medicine 80: 248-257.

Fu W, Zhang F, Wang Y, Gu X, Jin L (2010) Identification of copy number variation hotspots in human populations. The American Journal of Human Genetics 87: 494-504.

Ghannam S, Dejou C, Pedretti N, Giot J-P, Dorgham K, Boukhaddaoui H, Deleuze V, Bernard F-X, Jorgensen C, Yssel H (2011) CCL20 and β-defensin-2 induce arrest of human Th17 cells on inflamed endothelium in vitro under flow conditions. The Journal of Immunology 186: 1411-1420.

Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs RJ, Freedman BI, Quinones MP, Bamshad MJ (2005) The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science 307: 1434-1440.

Goossens M, Dozy AM, Embury SH, Zachariades Z, Hadjiminas MG, Stamatoyannopoulos G, Kan YW (1980) Triplicated alpha-globin loci in humans. Proceedings of the National Academy of Sciences 77: 518-521.

Graf SW, Lester S, Nossent JC, Hill CL, Proudman S, Lee A, Rischmueller M (2012) Low copy number of the FCGR3B gene and rheumatoid arthritis: a case-control study and meta-analysis. Arthritis Research and Therapy 14.

Grünhage F, Nattermann J, Gressner OA, Wasmuth HE, Hellerbrand C, Sauerbruch T, Spengler U, Lammert F (2010) Lower copy numbers of the chemokine< i> CCL3L1</i> gene in patients with chronic hepatitis C. Journal of hepatology 52: 153-159.

Hardwick RJ, Amogne W, Mugusi S, Yimer G, Ngaimisi E, Habtewold A, Minzi O, Makonnen E, Janabi M, Machado LR, Viskaduraki M, Mugusi F, Aderaye G, Lindquist L, Hollox EJ, Aklillu E (2012) β-defensin Genomic Copy Number Is Associated With HIV Load and Immune Reconstitution in Sub-Saharan Africans. Journal of Infectious Diseases 206: 1012-1019.

Hardwick RJ, Machado LR, Zuccherato LW, Antolinos S, Xue Y, Shawa N, Gilman RH, Cabrera L, Berg DE, Tyler-Smith C et al. (2011) A worldwide analysis of beta-defensin copy number variation suggests recent selection of a high-expressing DEFB103 gene copy in East Asia. Human Mutation 32: 743-750.

Hardwick RJ, Ménard A, Sironi M, Milet J, Garcia A, Sese C, Yang F, Fu B, Courtin D, Hollox EJ (2014) Haptoglobin (HP) and Haptoglobin-related protein (HPR) copy number variation, natural selection, and trypanosomiasis. Human Genetics 133: 69-83.

Harteveld CL, Higgs DR (2010) Review α-thalassaemia. Orphanet J Rare Dis 5: 13.

Hindson BJ, Ness KD, Masquelier DA, Belgrader P, Heredia NJ, Makarewicz AJ, Bright IJ, Lucero MY, Hiddessen AL, Legler TC et al. (2011) High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. Analytical chemistry 83: 8604-8610.

Hollox EJ, Armour JA, Barber JC (2003) Extensive normal copy number variation of a β-defensin antimicrobial-gene cluster. The American Journal of Human Genetics 73: 591-600.

Hollox EJ, Barber JCK, Brookes AJ, Armour JAL (2008a) Defensins and the dynamic genome: what we can learn from structural variation at human chromosome band 8p23. 1. Genome Research 18: 1686-1697.

Hollox EJ, Detering JC, Dehnugara T (2009) An integrated approach for measuring copy number variation at the FCGR3 (CD16) locus. Human Mutation 30: 477-484.

Hollox EJ, Huffmeier U, Zeeuwen PL, Palla R, Lascorz J, Rodijk-Olthuis D, van de Kerkhof PC, Traupe H, de Jongh G, den Heijer M et al. (2008b) Psoriasis is associated with increased beta-defensin genomic copy number. Nature Genetics 40: 23.

Huik K, Sadam M, Karki T, Avi R, Krispin T, Paap P, Rüütel K, Uusküla A, Talu A, Abel-Ollo K et al. (2010) CCL3L1 copy number is a strong genetic determinant of HIV seropositivity in Caucasian intravenous drug users. Journal of Infectious Diseases 201: 730-739.

Ioan-Facsinay A, De Kimpe S, Hellwig S, Van Lent P, Hofhuis F, Van Ojik H, Sedlik C, Da Silveira S, Gerber J, De Jong Y et al. (2002) FcγRI (CD64) contributes substantially to severity of arthritis, hypersensitivity responses, and protection from bacterial infection. Immunity 16: 391-402.

Iskow RC, Gokcumen O, Lee C (2012) Exploring the role of copy number variants in human adaptation. Trends in genetics 28: 245-257.

Jansen PA, Rodijk-Olthuis D, Hollox EJ, Kamsteeg M, Tjabringa GS, de Jongh GJ, van Vlijmen-Willems IM, Bergboer JG, van Rossum MM, de Jong EM (2009) β-Defensin-2 protein is a serum biomarker for disease activity in psoriasis and reaches biologically relevant concentrations in lesional skin. PLOS one 4: e4725.

Jiang W, Ghosh SK, Flyckt R, Kalinowska M, Starks D, Jurevic R, Weinberg A, Lederman MM, Rodriguez B (2012) Bacterial colonization and beta defensins in the female genital tract in HIV infection. Current HIV research 10: 504.

Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, Lee JC, Schumm LP, Sharma Y, Anderson CA et al. (2012) Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 491: 119-124.

Jouanguy E, Lamhamedi-Cherradi S, Lammas D, Dorman SE, Fondanèche M-C, Dupuis S, Döffinger R, Altare F, Girdlestone J, Emile J-F et al. (1999) A human IFNGR1 small deletion hotspot associated with dominant susceptibility to mycobacterial infection. Nature Genetics 21: 370-378.

Kaslow R, Carrington M, Apple R, Park L, Munoz A, Saah A, Goedert J, Winkler C, O'brien S, Rinaldo C (1996) Influence of combinations of human major histocompatibility complex genes on the course of HIV–1 infection. Nature medicine 2: 405-411.

Khan FF, Carpenter D, Mitchell L, Mansouri O, Black HA, Tyson J, Armour JA (2013) Accurate measurement of gene copy number for human alpha-defensin DEFA1A3. BMC genomics 14: 719.

Khor CC, Davila S, Breunis WB, Lee Y-C, Shimizu C, Wright VJ, Yeung RS, Tan DE, Sim KS, Wang JJ et al. (2011) Genome-wide association study identifies FCGR2A as a susceptibility locus for Kawasaki disease. Nature Genetics 43: 1241-1246.

Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, Graves T, Hansen N, Teague B, Alkan C, Antonacci F (2008) Mapping and sequencing of structural variation from eight human genomes. Nature 453: 56-64.

Krumm N, Sudmant PH, Ko A, O'Roak BJ, Malig M, Coe BP, Quinlan AR, Nickerson DA, Eichler EE (2012) Copy number variation detection and genotyping from exome sequence data. Genome Research 22: 1525-1532.

Kuhn L, Schramm DB, Donninger S, Meddows-Taylor S, Coovadia AH, Sherman GG, Gray GE, Tiemessen CT (2007) African infants' CCL3 gene copies influence perinatal HIV transmission in the absence of maternal nevirapine. AIDS (London, England) 21: 1753-1761.

Lam K-WG, Jeffreys AJ (2006) Processes of copy-number change in human DNA: the dynamics of α-globin gene deletion. Proceedings of the National Academy of Sciences 103: 8921-8927.

Larsen MH, Thørner LW, Zinyama R, Amstrup J, Kallestrup P, Gerstoft J, Gomo E, Erikstrup C, Ullum H et al. (2012) < i> CCL3L</i> gene copy number and survival in an HIV-1 infected Zimbabwean population. Infection, Genetics and Evolution: 1087-1093.

Lee EY, Yue FY, Jones RB, Lo C, Sheth P, Hyrcza MD, Kovacs C, Benko E, Kaul R, Ostrowski M (2010) The impact of CCL3L1 copy number in an HIV-1-infected white population. Aids 24: 1589-1591.

Lee H, Bae S, Choi BW, Choi JC, Yoon Y (2011) Copy number variation of< i> CCL3L1</i> influences asthma risk by modulating IL-10 expression. Clinica Chimica Acta 412: 2100-2104.

Littaua R, Kurane I, Ennis FA (1990) Human IgG Fc receptor II mediates antibody-dependent enhancement of dengue virus infection. The Journal of Immunology 144: 3183-3186.

Liu S, Yao L, Ding D, Zhu H (2010) CCL3L1 copy number variation and susceptibility to HIV-1 infection: a meta-analysis. PLOS one 5: e15778.

Liu Y, Old J, Miles K, Fisher C, Weatherall D, Clegg J (2000) Rapid detection of a-thalassaemia deletions and a-globin gene triplication by multiplex polymerase chain reactions. British journal of haematology 108: 295-299.

Locke DP, Segraves R, Carbone L, Archidiacono N, Albertson DG, Pinkel D, Eichler EE (2003) Large-scale variation among human and great ape genomes determined by array comparative genomic hybridization. Genome Research 13: 347-357.

Locke DP, Sharp AJ, McCarroll SA, McGrath SD, Newman TL, Cheng Z, Schwartz S, Albertson DG, Pinkel D, Altshuler DM, Eichler EE (2006) Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. American journal of human genetics 79: 275-290.

Lundqvist E, Johansson I, Ingelman-Sundberg M (1999) Genetic mechanisms for duplication and multiduplication of the human< i> CYP2D6</i> gene and methods for detection of duplicated< i> CYP2D6</i> genes. Gene 226: 327-338.

Machado LR, Bowdrey J, Ngaimisi E, Habtewold A, Minzi O, Makonnen E, Yimer G, Amogne W, Mugusi S, Janabi M (2013) Copy Number Variation of Fc Gamma Receptor Genes in HIV-Infected and HIV-Tuberculosis Co-Infected Individuals in Sub-Saharan Africa. PLOS one 8: e78165.

Machado LR, Hardwick RJ, Bowdrey J, Bogle H, Knowles TJ, Sironi M, Hollox EJ (2012) Evolutionary history of copy-number-variable locus for the low-affinity fcγ receptor: mutation rate, autoimmune disease, and the legacy of Helminth infection. The American Journal of Human Genetics.

Mackinnon MJ, Mwangi TW, Snow RW, Marsh K, Williams TN (2005) Heritability of malaria in Africa. PLoS medicine 2: e340.

Maeda N, McEvoy SM, Harris HF, Huisman T, Smithies O (1986) Polymorphisms in the human haptoglobin gene cluster: chromosomes with multiple haptoglobin-related (Hpr) genes. Proceedings of the National Academy of Sciences 83: 7395-7399.

Maeda N, Yang F, Barnett DR, Bowman BH, Smithies O (1983) Duplication within the haptoglobin Hp2 gene. Nature 309: 131-135.

Mamtani M, Rovin B, Brey R, Camargo JF, Kulkarni H, Herrera M, Correa P, Holliday S, Anaya J-M, Ahuja SK (2008) CCL3L1 gene-containing segmental duplications and polymorphisms in CCR5 affect risk of systemic lupus erythaematosus. Annals of the rheumatic diseases 67: 1076-1083.

Maresco D, Blue L, Culley L, Kimberly R, Anderson C, Theil K (1998) Localization of FCGR1 encoding Fcγ receptor class I in primates: molecular evidence for two pericentric inversions during the evolution of human chromosome 1. Cytogenetic and Genome Research 82: 71-74.

Maresco D, Chang E, Theil K, Francke U, Anderson C (1996) The three genes of the human FCGR1 gene family encoding FcγRI flank the centromere of chromosome 1 at 1p12 and 1q21. Cytogenetic and Genome Research 73: 157-163.

Marques-Bonet T, Kidd JM, Ventura M, Graves TA, Cheng Z, Hillier LW, Jiang Z, Baker C, Malfavon-Borja R, Fulton LA (2009) A burst of segmental duplications in the genome of the African great ape ancestor. Nature 457: 877-881.

Marques RB, Thabet MM, White SJ, Houwing-Duistermaat JJ, Bakker AM, Hendriks G-J, Zhernakova A, Huizinga TW, van der Helm-van AH, Toes RE et al. (2010) Genetic variation of the Fc gamma receptor 3B gene and association with rheumatoid arthritis. PLOS one 5: e13173.

May J, Evans JA, Timmann C, Ehmen C, Busch W, Thye T, Agbenyega T, Horstmann RD (2007) Hemoglobin variants and disease manifestations in severe falciparum malaria. Jama 297: 2220-2226.

Mayerle J, den Hoed CM, Schurmann C, Stolk L, Homuth G, Peters MJ, Capelle LG, Zimmermann K, Rivadeneira F, Gruska S et al. (2013) Identification of Genetic Loci Associated With Helicobacter pylori Serologic Status GWAS of H pylori Infection Susceptibility. Jama 309: 1912-1920.

McDermid J, Prentice A (2006) Iron and infection: effects of host iron status and the iron-regulatory genes haptoglobin and NRAMP1 (SLC11A1) on host-pathogen interactions in tuberculosis and HIV. Clinical Science 110: 503-524.

McEvoy SM, Maeda N (1988) Complex events in the evolution of the haptoglobin gene cluster in primates. Journal of Biological Chemistry 263: 15740-15747.

McIntosh RS, Shi J, Jennings RM, Chappel JC, de Koning-Ward TF, Smith T, Green J, van Egmond M, Leusen JH, Lazarou M (2007) The importance of human FcγRI in mediating protection to malaria. PLoS pathogens 3: e72.

McKinney C, Merriman ME, Chapman PT, Gow PJ, Harrison AA, Highton J, Jones PB, McLean L, O'Donnell JL, Pokorny V et al. (2008) Evidence for an influence of chemokine ligand 3-like 1 (CCL3L1) gene copy number on susceptibility to rheumatoid arthritis. Annals of the rheumatic diseases 67: 409-413.

McKinney C, Merriman TR (2012) Meta-analysis confirms a role for deletion in FCGR3B in autoimmune phenotypes. Human molecular genetics. 21: 2370-2376.

McLean-Tooke A, Barge D, Spickett GP, Gennery AR (2008) Immunologic defects in 22q11. 2 deletion syndrome. Journal of Allergy and Clinical Immunology 122: 362-367. e4.

Metes D, Ernst LK, Chambers WH, Sulica A, Herberman RB, Morel PA (1998) Expression of functional CD32 molecules on human NK cells is determined by an allelic polymorphism of the FcγRIIC gene. Blood 91: 2369.

Middleton D, Gonzelez F (2010) The extensive polymorphism of KIR genes. Immunology 129: 8-19.

Mitchell C, Gottsch ML, Liu C, Fredricks DN, Nelson DB (2013) Associations between vaginal bacteria and levels of vaginal defensins in pregnant women. American journal of obstetrics and gynecology 208: 132. e1-132. e7.

Mittal R, Sukumaran SK, Selvaraj SK, Wooster DG, Babu MM, Schreiber AD, Verbeek JS, Prasadarao NV (2010) Fcγ receptor I alpha chain (CD64) expression in macrophages is critical for the onset of meningitis by Escherichia coli K1. PLoS pathogens 6: e1001203.

Mockenhaupt FP, Ehrhardt S, Gellert S, Otchwemah RN, Dietz E, Anemana SD, Bienzle U (2004) α+-thalassemia protects African children from severe malaria. Blood 104: 2003-2006.

Montavon T, Thevenet L, Duboule D (2012) Impact of copy number variations (CNVs) on long-range gene regulation at the HoxD locus. Proceedings of the National Academy of Sciences 109: 20204-20211.

Moulds J (2010) The Knops blood-group system: a review. Immunohematology 26: 2-7.

Mueller M, Barros P, Witherden AS, Roberts AL, Zhang Z, Schaschl H, Yu C-Y, Hurles ME, Schaffner C, Floto RA et al. (2012) Genomic Pathology of SLE-Associated Copy-Number Variation at the< i> FCGR2C/FCGR3B/FCGR2B</i> Locus. The American Journal of Human Genetics.

Nakajima T, Ohtani H, Naruse T, Shibata H, Mimaya J-i, Terunuma H, Kimura A (2007) Copy number variations of CCL3L1 and long-term prognosis of HIV-1 infection in asymptomatic HIV-infected Japanese with hemophilia. Immunogenetics 59: 793-798.

Nielsen MJ, Moestrup SK (2009) Receptor targeting of hemoglobin mediated by the haptoglobins: roles beyond heme scavenging. Blood 114: 764-771.

Nielsen R (2005) Molecular signatures of natural selection. Annu. Rev. Genet. 39: 197-218.

Nimmerjahn F, Lux A (2014) LILR-B1 blocks activating FcγR signaling to allow antibody dependent enhancement of dengue virus infection. Proceedings of the National Academy of Sciences 111: 2404-2405.

Nimmerjahn F, Ravetch JV (2008) Fcγ receptors as regulators of immune responses. Nature Reviews Immunology 8: 34-47.

Nordang G, Carpenter D, Viken M, Kvien T, Armour J, Lie B (2012) Association analysis of the CCL3L1 copy number locus by paralogue ratio test in Norwegian rheumatoid arthritis patients and healthy controls. Genes and Immunity 13: 579-582.

Okamoto N, Fujii T, Tanaka J, Saito K, Matsui T, Harada N (2014) A clinical study of patients with pericentromeric deletion and duplication within 16p12. 2–p11. 2. American Journal of Medical Genetics Part A 164: 213-219.

Olsson LM, Holmdahl R (2012) Copy number variation in autoimmunity—importance hidden in complexity? European journal of immunology 42: 1969-1976.

Ory PA, Clark MR, Kwoh EE, Clarkson SB, Goldstein IM (1989a) Sequences of complementary DNAs that encode the NA1 and NA2 forms of Fc receptor III on human neutrophils. Journal of Clinical Investigation 84: 1688.

Ory PA, Goldstein IM, Kwoh EE, Clarkson SB (1989b) Characterization of polymorphic forms of Fc receptor III on human neutrophils. Journal of Clinical Investigation 83: 1676.

Park HJ, Guariento M, Maciejewski M, Hauhart R, Tham W-H, Cowman AF, Schmidt CQ, Mertens HD, Liszewski MK, Hourcade DE et al. (2014) Using mutagenesis and structural biology to map the binding site for the Plasmodium falciparum merozoite protein PfRh4 on the human immune adherence receptor. Journal of Biological Chemistry 289: 450-463.

Pelak K, Need AC, Fellay J, Shianna KV, Feng S, Urban TJ, Ge D, De Luca A, Martinez-Picado J, Wolinsky SM (2011a) Copy number variation of KIR genes influences HIV-1 control. PLoS biology 9: e1001208.

Pelak K, Need AC, Fellay J, Shianna KV, Feng S, Urban TJ, Ge D, De Luca A, Martinez-Picado J, Wolinsky SM, Martinson J, Jamieson B, Bream J, Martin M, Borrow P, Letvin N, McMichael A, Haynes B, Telenti A, Carrington M, Goldstein D, Alter G, Immunology" NCfHAV (2011b) Copy Number Variation of KIR Genes Influences HIV-1 Control. PLoS biology 9: e1001208.

Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villanea FA, Mountain JL, Misra R et al. (2007) Diet and the evolution of human amylase gene copy number variation. Nature Genetics 39: 1256-1260.

Perry GH, Yang F, Marques-Bonet T, Murphy C, Fitzgerald T, Lee AS, Hyland C, Stone AC, Hurles ME, Tyler-Smith C, Eichler EE, Carter NP, Lee C, Redon R (2008) Copy number variation and evolution in humans and chimpanzees. Genome Research 18: 1698-1710.

Pinto D, Darvishi K, Shi X, Rajan D, Rigler D, Fitzgerald T, Lionel AC, Thiruvahindrapuram B, MacDonald JR, Mills R (2011) Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. Nature biotechnology 29: 512-520.

Poulsen P, Kyvik KO, Vaag A, Beck-Nielsen H (1999) Heritability of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance–a population-based twin study. Diabetologia 42: 139-145.

Quiñones-Mateu ME, Lederman MM, Feng Z, Chakraborty B, Weber J, Rangel HR, Marotta ML, Mirza M, Jiang B, Kiser P (2003) Human epithelial [beta]-defensins 2 and 3 inhibit HIV-1 replication. Aids 17: F39-F48.

Ravetch J, Perussia B (1989) Alternative membrane forms of Fc gamma RIII (CD16) on human natural killer cells and neutrophils. Cell type-specific expression of two genes that differ in single nucleotide substitutions. The Journal of experimental medicine 170: 481.

Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W et al. (2006) Global variation in copy number in the human genome. Nature 444: 444-454.

Reilly AF, Surrey S, Rappaport EF, Schwartz E, McKenzie SE (1994) Variation in human fcgr2c gene copy number. Immunogenetics 40: 456-456.

Robinson JI, Carr IM, Cooper DL, Rashid LH, Martin SG, Emery P, Isaacs JD, Barton A, Wilson AG, Barrett JH (2012) Confirmation of association of FCGR3B but not FCGR3A copy number with susceptibility to autoantibody positive rheumatoid arthritis. Human Mutation 33: 741-749.

Rodrigo W, Block OK, Lane C, Sukupolvi-Petty S, Goncalvez AP, Johnson S, Diamond MS, Lai C-J, Rose RC, Jin X et al. (2009) Dengue virus neutralization is modulated by IgG antibody subclass and Fcγ receptor subtype. Virology 394: 175-182.

Rodrigo WSI, Jin X, Blackley SD, Rose RC, Schlesinger JJ (2006) Differential enhancement of dengue virus immune complex infectivity mediated by signaling-competent and signaling-incompetent human FcγRIA (CD64) or FcγRIIA (CD32). Journal of virology 80: 10128-10138.

Rodriguez S, Williams DM, Guthrie PA, McArdle WL, Smith GD, Evans DM, Gaunt TR, Day IN (2012) Molecular and population analysis of natural selection on the human haptoglobin duplication. Annals of human genetics 76: 352-362.

Salmon J, Edberg J, Kimberly R (1990) Fc gamma receptor III on human neutrophils. Allelic variants have functionally distinct capacities. Journal of Clinical Investigation 85: 1287.

Schaschl H, Aitman T, Vyse T (2009) Copy number variation in the human genome and its implication in autoimmunity. Clinical & Experimental Immunology 156: 12-16.

Schlosstein L, Terasaki PI, Bluestone R, Pearson CM (1973) High association of an HL-A antigen, W27, with ankylosing spondylitis. New England Journal of Medicine 288: 704-706.

Shao W, Tang J, Song W, Wang C, Li Y, Wilson C, Kaslow R (2007) CCL3L1 and CCL4L1: variable gene copy number in adolescents with and without human immunodeficiency virus type 1 (HIV-1) infection. Genes and Immunity 8: 224-231.

She X, Liu G, Ventura M, Zhao S, Misceo D, Roberto R, Cardone MF, Rocchi M, Green ED, Archidiacano N et al. (2006) A preliminary comparative analysis of primate segmental duplications shows elevated substitution rates and a great-ape expansion of intrachromosomal duplications. Genome Research 16: 576-583.

Shea PR, Shianna KV, Carrington M, Goldstein DB (2013) Host genetics of HIV acquisition and viral control. Annual review of medicine 64: 203-217.

Sironi M, Clerici M (2010) The hygiene hypothesis: an evolutionary perspective. Microbes and Infection 12: 421-427.

Stoute JA (2011) Complement receptor 1 and malaria. Cellular microbiology 13: 1441-1450.

Stuart PE, Hüffmeier U, Nair RP, Palla R, Tejasvi T, Schalkwijk J, Elder JT, Reis A, Armour JA (2012) Association of β-defensin copy number and psoriasis in three cohorts of European origin. Journal of Investigative Dermatology 132: 2407-2413.

Sudmant PH, Kitzman JO, Antonacci F, Alkan C, Malig M, Tsalenko A, Sampas N, Bruhn L, Shendure J, Eichler EE (2010) Diversity of human copy number variation and multicopy genes. Science 330: 641.

Sun L, DeMasi L, Lafferty M, Goicochea M, Lu W, Garzino-Demo A (2006) CCR6 mediates the intracellular HIV inhibitory activity of human beta-defensin 2. Retrovirology 3: S77.

Teshima KM, Innan H (2004) The effect of gene conversion on the divergence between duplicated genes. Genetics 166: 1553-1560.

Teshima KM, Innan H (2012) The coalescent with selection on copy number variants. Genetics 190: 1077-1086.

Tham W-H, Wilson DW, Lopaticki S, Schmidt CQ, Tetteh-Quarcoo PB, Barlow PN, Richard D, Corbin JE, Beeson JG, Cowman AF et al. (2010) Complement receptor 1 is the host erythrocyte receptor for Plasmodium falciparum PfRh4 invasion ligand. Proceedings of the National Academy of Sciences 107: 17327-17332.

Thathy V, Moulds JM, Guyah B, Otieno W, Stoute JA (2005) Complement receptor 1 polymorphisms associated with resistance to severe malaria in Kenya. Malaria journal 4: 54.

Thornton KR (2007) The neutral coalescent process for recent gene duplications and copy-number variants. Genetics 177: 987-1000.

Tolia NH, Enemark EJ, Sim B, Joshua-Tor L (2005) Structural Basis for the EBA-175 Erythrocyte Invasion Pathway of the Malaria Parasite< i> Plasmodium falciparum</i>. Cell 122: 183-193.

Urban TJ, Weintrob AC, Fellay J, Colombo S, Shianna KV, Gumbs C, Rotger M, Pelak K, Dang KK, Detels R, Martinson JJ, O'Brien SJ, Letvin NL, McMichael AJ, Haynes BF, Carrington M, Telenti

A, Michael NL, Goldstein DB (2009) CCL3L1 and HIV/AIDS susceptibility. Nat Med 15: 1110-1112. doi: http://www.nature.com/nm/journal/v15/n10/suppinfo/nm1009-1110_S1.html

van der Heijden J, Breunis WB, Geissler J, de Boer M, van den Berg TK, Kuijpers TW (2012) Phenotypic variation in IgG receptors by nonclassical FCGR2C alleles. The Journal of Immunology 188: 1318-1324.

van der Poel CE, Spaapen RM, van de Winkel JG, Leusen JH (2011) Functional characteristics of the high affinity IgG receptor, FcγRI. The Journal of Immunology 186: 2699-2704.

van Vugt MJ, Reefman E, Zeelenberg I, Boonen G, Leusen JH, van de Winkel JG (1999) The alternatively spliced CD64 transcript FcγRIb2 does not specify a surface-expressed isoform. European journal of immunology 29: 143-149.

Vanhollebeke B, De Muylder G, Nielsen MJ, Pays A, Tebabi P, Dieu M, Raes M, Moestrup SK, Pays E (2008) A haptoglobin-hemoglobin receptor conveys innate immunity to Trypanosoma brucei in humans. Science 320: 677-681.

Veal CD, Xu H, Reekie K, Free R, Hardwick RJ, McVey D, Brookes AJ, Hollox EJ, Talbot CJ (2013) Automated design of paralogue ratio test assays for the accurate and rapid typing of copy number variation. Bioinformatics 29: 1997-2003.

Vik DP, Wong WW (1993) Structure of the gene for the F allele of complement receptor type 1 and sequence of the coding region unique to the S allele. The Journal of Immunology 151: 6214-6224.

Walker S, Janyakhantikul S, Armour JA (2009) Multiplex Paralogue Ratio Tests for accurate measurement of multiallelic CNVs. Genomics 93: 98-103.

Willcocks LC, Lyons PA, Clatworthy MR, Robinson JI, Yang W, Newland SA, Plagnol V, McGovern NN, Condliffe AM, Chilvers ER et al. (2008) Copy number of FCGR3B, which is associated with systemic lupus erythematosus, correlates with protein expression and immune complex uptake. The Journal of experimental medicine 205: 1573-1582.

Willcocks LC, Smith KG, Clatworthy MR (2009) Low-affinity Fcγ receptors, autoimmunity and infection. Expert reviews in molecular medicine 11: e24.

Williams TN, Wambua S, Uyoga S, Macharia A, Mwacharo JK, Newton CR, Maitland K (2005) Both heterozygous and homozygous α+ thalassemias protect against severe and fatal Plasmodium falciparum malaria on the coast of Kenya. Blood 106: 368-371.

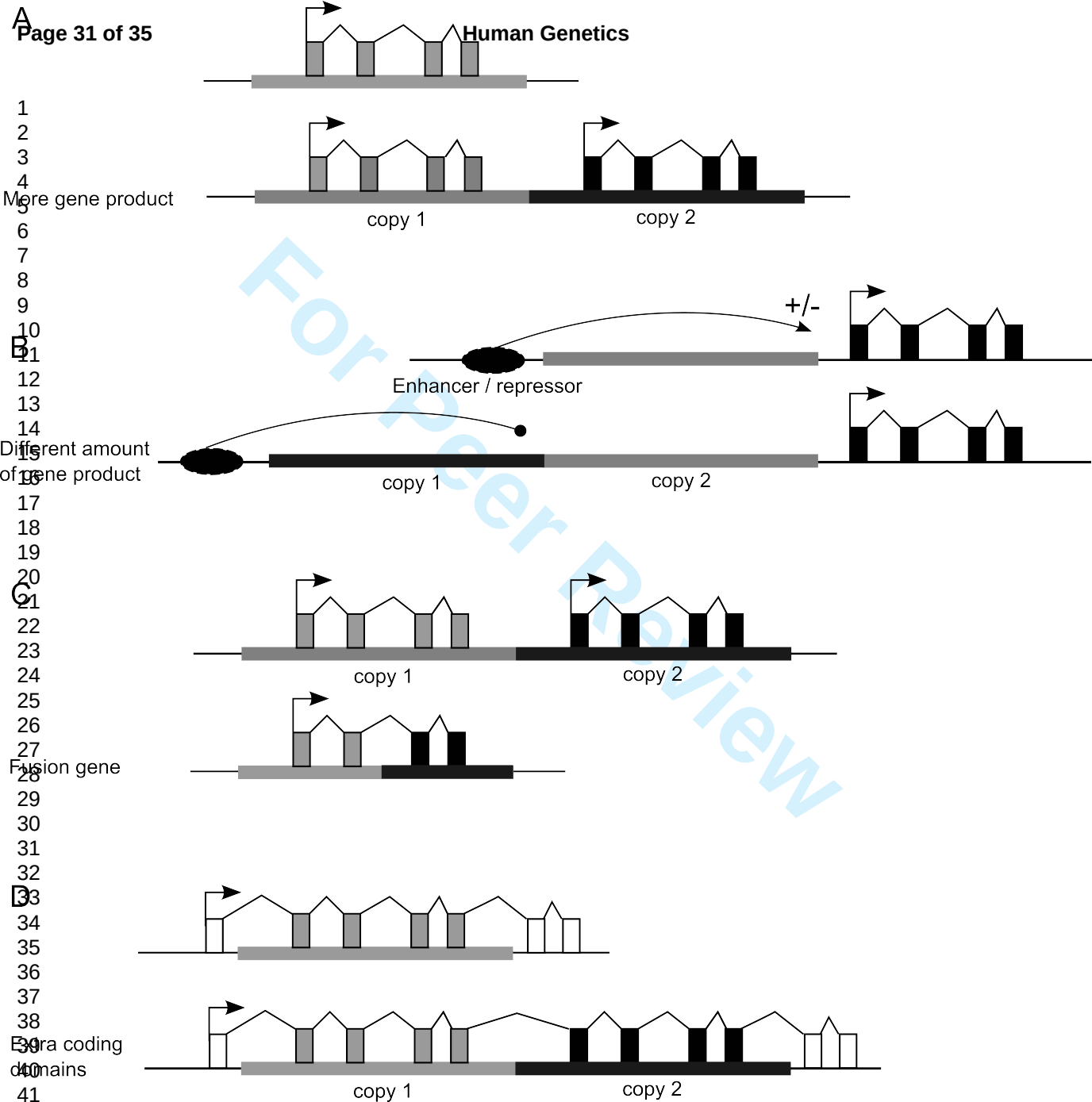
Wong W, Cahill J, Rosen M, Kennedy C, Bonaccio E, Morris M, Wilson J, Klickstein L, Fearon D (1989) Structure of the human CR1 gene. Molecular basis of the structural and quantitative polymorphisms and identification of a new CR1-like allele. The Journal of experimental medicine 169: 847-863.

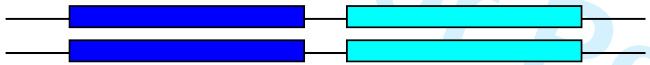
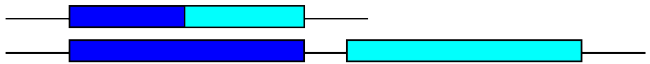
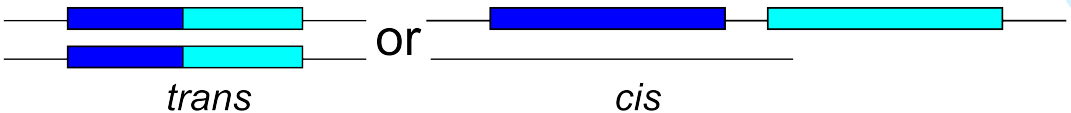
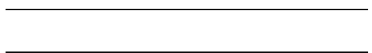
Yang D, Chertov O, Bykovskaia S, Chen Q, Buffo M, Shogan J, Anderson M, Schröder J, Wang J, Howard O et al. (1999) β-Defensins: linking innate and adaptive immunity through dendritic and T cell CCR6. Science 286: 525-528.

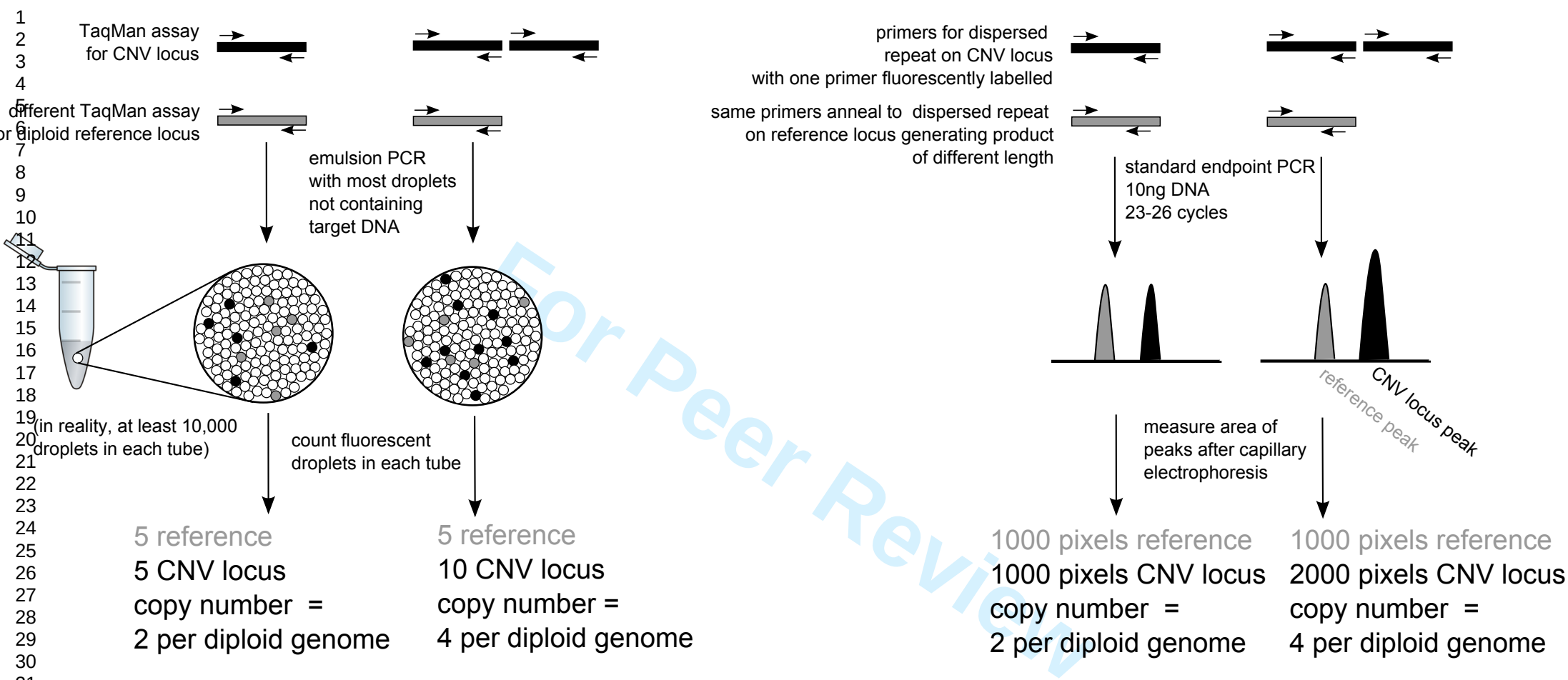
Yang Y, Chung EK, Wu YL, Savelli SL, Nagaraja HN, Zhou B, Hebert M, Jones KN, Shu Y, Kitzmiller K et al. (2007) Gene copy-number variation and associated polymorphisms of complement component C4 in human systemic lupus erythematosus (SLE): low copy number is a risk factor for and high copy number is a protective factor against SLE susceptibility in European Americans. The American Journal of Human Genetics 80: 1037-1054.

Yu CY, Belt KT, Giles CM, Campbell RD, Porter RR (1986) Structural basis of the polymorphism of human complement components C4A and C4B: gene size, reactivity and antigenicity. The EMBO journal 5: 2873.

Zimmerman P, Fitness J, Moulds J, McNamara D, Kasehagen L, Rowe JA, Hill A (2003) CR1 Knops blood group alleles are not associated with severe malaria in the Gambia. Genes and Immunity 4: 368-373.

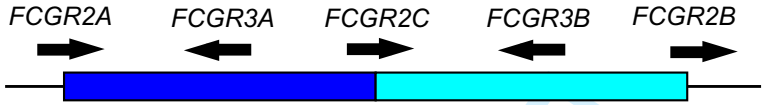


α-globin gene copy number	α-globin gene arrangement	Blood disorder phenotype	Infectious disease phenotype
5		normal	normal
4		normal	normal
3		normal	protection against severe malaria
2		mild anemia (α^{+}thallasemia)	protection against severe malaria
1		moderately severe hemolytic anemia	not known
0		hydrops fetalis	not applicable



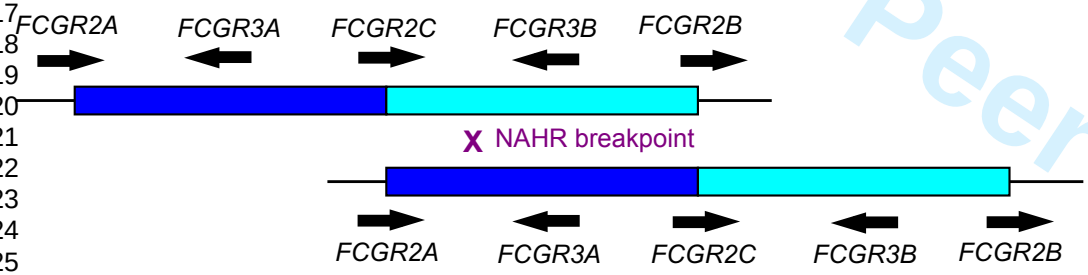
Ancestral gene region

duplication ~25 million years ago
and subsequent divergence



single nucleotide
variation 1,2

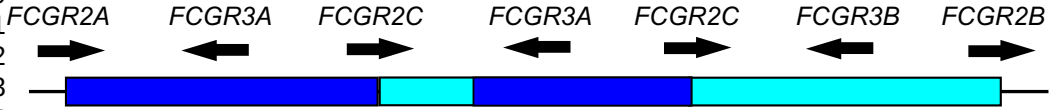
NAHR / unequal crossing over



duplicated alleles 6,7

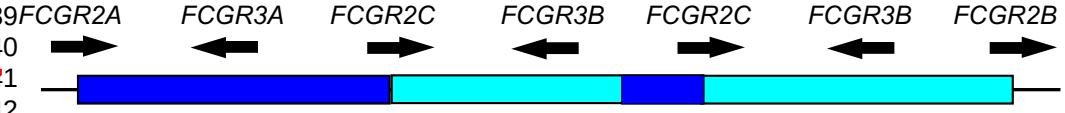
deleted alleles 3,4,5

6



NAHR resulting in duplication of genes with increase in gene expression
Example, *FCGR3A* and *FCGR2C* duplication (predicted).

7

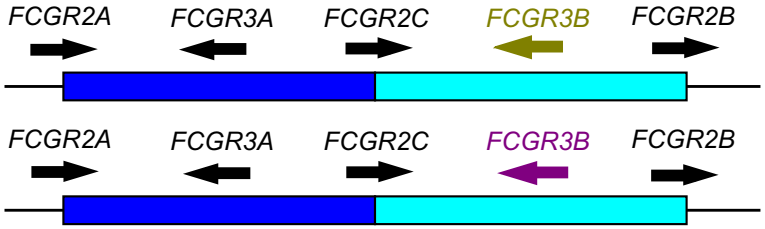


NAHR at alternative breakpoint resulting in duplication of genes
Example, *FCGR3B* and *FCGR2C* duplication (predicted).

1

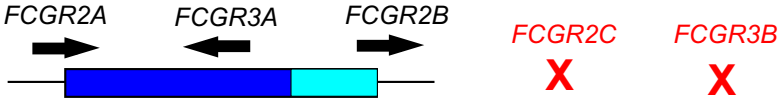
SNP causing a premature stop codon or splice site mutation
Example: *FCGR2C* polymorphisms rending it a pseudogene

2



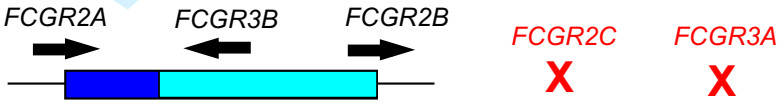
SNP generating allelic differences within a gene.
Example: *FCGR3B* polymorphisms causing amino acid changes and functional variation in IgG binding affinity.

3



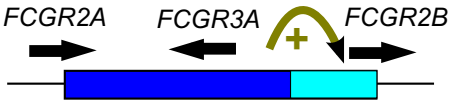
NAHR resulting in deletion of genes,
with loss of gene expression from those genes.
Example, *FCGR3B* and *FCGR2C* deletion resulting in complete loss of expression of these geen in homozygotes, and a gene dosage effect in heterozygotes

4



NAHR at alternative breakpoint.
Example, *FCGR3A* and *FCGR2C* deletion.

5



NAHR resulting in inappropriate expression of genes
Example, *FCGR2B* expression on NK cells from deleted alleles, probably due to increased proximity of NK-specific enhancer.