Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Genetic susceptibility to infectious diseases: Current status and future perspectives from genome-wide approaches

Alessandra Mozzi¹, Chiara Pontremoli¹, Manuela Sironi* 

Bioinformatics, Scientific Institute IRCCS E.MEDEA, 23842 Bosisio Parini, Italy

ARTICLE INFO

Keywords:
GWAS
Infectious disease
Response to treatment or vaccine

ABSTRACT

Genome-wide association studies (GWASs) have been widely applied to identify genetic factors that affect complex diseases or traits. Presently, the GWAS Catalog includes > 2800 human studies. Of these, only a minority have investigated the susceptibility to infectious diseases or the response to therapies for the treatment or prevention of infections. Despite their limited application in the field, GWASs have provided valuable insights by pinpointing associations to both innate and adaptive immune response loci, as well as novel unexpected risk factors for infection susceptibility. Herein, we discuss some issues and caveats of GWASs for infectious diseases, we review the most recent findings ensuing from these studies, and we provide a brief summary of selected GWASs for infections in non-human mammals. We conclude that, although the general trend in the field of complex traits is to shift from GWAS to next-generation sequencing, important knowledge on infectious disease-related traits can be still gained by GWASs, especially for those conditions that have never been investigated using this approach. We suggest that future studies will benefit from the leveraging of information from the host's and pathogen's genomes, as well as from the exploration of models that incorporate heterogeneity across populations and phenotypes. Interactions within HLA genes or among HLA variants and polymorphisms located outside the major histocompatibility complex may also play an important role in shaping the susceptibility and response to invading pathogens.

1. General overview

Genome-wide association studies (GWASs) are based on the screening of many genomes to find genetic variants associated with a trait or disease. Both dichotomous and quantitative traits can be analyzed in GWASs, and the genomes can be those of unrelated subjects or derive from samples with familiar structure (e.g., parent-affected child trios). Because GWASs analyze variants distributed throughout the genome, they are unbiased with respect to prior biological knowledge, offering the potential to identify novel variants and genes associated to the trait of interest.

The first small-scale GWAS results were performed in 2005–2006 and investigated age-related macular degeneration, a common disease in elderly populations worldwide (Dewan et al., 2006; Klein et al., 2005). After those early efforts, the Wellcome Trust Case Control Consortium large-scale study in 2007 (Wellcome Trust Case Control Consortium, 2007) gave way to the GWAS era. In recent years, hundreds of GWASs have been published. Their results are stored in a dedicated database, the GWAS Catalog, presently hosted by the EMBL-EBI (https://www.ebi.ac.uk/gwas/) (MacArthur et al., 2017). As of April 24th 2017, the catalog comprised 2880 studies and > 9700 unique variant-trait associations. This figure refers to associations with a p value lower than or equal to $5 \times 10^{-8}$, which is usually considered as the threshold for genome-wide significance.

We mined the catalog to obtain a global overview of the application of GWAS in the field of infectious diseases. We classified studies and associations in broad categories depending on the disease or trait they analyzed. The category “infectious disease” was intended in a very broad sense: for instance, it included virus-associated cancers, specific phenotypes associated with infections (e.g., mood disorders in prin orient...
GWAS evolution

Table S1. Table that detected at least one associated variant at genome-wide significance, whereas a list of all studies is available as Supplementary material for infectious diseases, we review the most recent replication (Kennedy et al., 2012). Overall, 4%, 0.66%, and 0.31% of all studies in the catalog were classified in the categories “infectious disease”, “response to treatment for infectious disease”, and “response to vaccine”, respectively (Fig. 1). In terms of unique variant-trait associations, percentages were as follows: 2.5% for “infectious disease”, 0.18% for “response to treatment for infectious disease”, and 0.66% for “response to vaccine” (Fig. 1). Based on the relative number of studies and associations, it is clear that some traits/diseases have attracted huge efforts that provided consequent reward (e.g., inflammatory/autoimmune disorders, metabolic traits and phenotypes), whereas others have been deeply investigated with comparatively modest success (e.g., neurologic/neuropsychiatric/cognitive diseases and traits, cancers). The reasons for these differences are manifold and outside the scope of this Review, which focuses on infections and related traits. Based on the percentages given above and on the comparison with other traits, we conclude that, despite the huge burden that infections pose for human health, relatively few GWASs for infectious diseases have been performed and these have attained a moderate success. Apparently, the best performance was obtained by GWASs for response to vaccination, with 9 studies that reported 64 associations. However, most of these associations will require validation, as they derive from a relatively small-scale study with no replication (Kennedy et al., 2012).

Sheer numbers, though, are not the only indicator of success. Important results were obtained by GWASs, both by identifying or refining associations in genes that were a priori expected to modulate infection susceptibility and by pointing out unexpected genes and pathways. Herein we discuss some issues associated with GWASs for infectious diseases, we review the most recent findings ensuing from these studies, and we provide a brief summary of selected GWASs in non-human mammals. Table 1 summarizes the results of human studies that detected at least one associated variant at genome-wide significance, whereas a list of all studies is available as Supplementary Table S1.

2. GWAS evolution

GWASs rely on linkage disequilibrium between genotyped SNPs and causal variants (often ungenotyped). The number of variants on the SNP chips that are used for GWASs has increased sensibly since the development of this technology (from ~500,000 in the initial studies to >4,000,000 in the latest platforms). Moreover, the availability of an increasing number of fully sequenced genomes, together with the development of efficient imputation methods, has definitely expanded the power of GWASs to detect association with variants in the low range of the frequency spectrum (reviewed in (Visser et al., 2017)). However, the overall power of a GWAS is also critically dependent on the sample size (reviewed in (Visser et al., 2017) and see below) and in most cases power remains low for variants with a frequency below 1% (reviewed in (Visser et al., 2017)).

Notably, though, important developments have been achieved in terms of imputation of variants within the Major Histocompatibility Complex (MHC), with clear relevance for the investigation of infectious (and immuno-mediated) diseases. As recently reviewed elsewhere (Matzaraki et al., 2017), the development of large reference panels (at least for Europeans and Asian populations) and the use of advanced imputation methods allowed the fine mapping of human leukocyte antigen (HLA) variants as susceptibility factors for some infectious diseases (e.g., HIV-1, HCV, and HBV infection) and many more autoimmune conditions (Matzaraki et al., 2017).

Recently, an imputation algorithm for typing KIR (killer-cell immunoglobulin-like receptor) gene copy number has also been developed and applied to data from European populations (Vukcevic et al., 2015). However, KIR variability extends beyond copy number variation and next-generation sequencing technologies were proposed to allow high-resolution genotyping of KIR loci (Maniangou et al., 2017; Norman et al., 2016).

Finally, we mention that in 2009 a SNP chip (Immunochip) providing high coverage of immune-related genes, including HLA and KIR loci, came into use to fine-map and validate association signals for immune-mediated diseases (Cortes and Brown, 2011). The application of the Immunochip in the field of infectious diseases has been very limited, with the notable exception of a study that investigated the genetic susceptibility to candidaemia and identified three risk variants (Kumar et al., 2014).

3. Issues and caveats of GWASs for infectious diseases and related traits

Some considerations apply to GWASs, whether they investigate...
| Pathogen/disease                     | Year | Population | Associated loci | Trait description                                    | References                          |
|-------------------------------------|------|------------|-----------------|------------------------------------------------------|-------------------------------------|
| **Viruses**                         |      |            |                 |                                                      |                                     |
| Dengue virus                         | 2011 | Asian ancestry | MICB, PLCE1 | Dengue shock syndrome (DSS) susceptibility            | (Khor et al., 2011)                 |
| Diarrhoeal disease                  | 2016 | European ancestry | NTN5, SEC1P, FUT2 | Susceptibility to diarrhoeal disease in young children (1 and 2 years) | (Rustamante et al., 2016) |
| **Epstein-Barr virus (EBV)**        | 2013 | Hispanic ancestry | HLA-DRB1, HLA-DQB1 | EBV seroreactivity measured by Epstein-Barr virus nuclear antigen 1 (anti-EBNA-1) titer | (Rubicz et al., 2013) |
| **Hepatitis B virus (HBV)**         | 2009 | Asian ancestry | HLA-DPB1       | Chronic hepatitis B (CHB) susceptibility              | (Kamatani et al., 2009)            |
|                                     | 2010 | Asian ancestry | UBE4B, KIF1B, PGD | HBV-related hepatocellular carcinoma (HCC) susceptibility | (Zhang et al., 2010)              |
|                                     | 2011 | Asian ancestry | HLA-DPA1, HLA-DPB1, HLA-DQB1, HLA-DQB2 | CHB susceptibility                                  | (Mbarek et al., 2011) |
|                                     | 2011 | Asian ancestry | GRN2A          | HBV progression                                      | (Liu et al., 2011)                 |
|                                     | 2012 | Asian ancestry | HLA-DPA1, HLA-DPB1 | HBV clearance                                       | (Nishida et al., 2012)             |
|                                     | 2012 | Asian ancestry | GRIK1, HLA-DRB1, HLA-DQAI | HBV-related HCC                                    | (Li et al., 2012)                 |
|                                     | 2013 | Asian ancestry | STA4, HLA-DQ   | HBV-related HCC                                      | (Jiang et al., 2013)               |
|                                     | 2013 | Asian ancestry | HLA-DP, HLA-DQ, EBMT2, TCF19 | CHB susceptibility                                  | (Kim et al., 2013)                |
|                                     | 2013 | Asian ancestry | HLA-DQB2, HLA-C, UBE2L3 | CHB susceptibility                                  | (Ju et al., 2013)                 |
|                                     | 2014 | Asian ancestry | HLA-DQA2, HLA-DQB2, HLA-DPB1, HLA-DPA3 | CHB in males                                       | (Chang et al., 2014)              |
|                                     | 2015 | Asian ancestry | HLA-C, CFB, NOTCH4, HLA-DOA, CD40, HLA-DQB1, HLA-DQB2, HLA-DPA1 | CHB susceptibility                                  | (Jiang et al., 2015)             |
|                                     | 2016 | Asian ancestry | rs70009921 ( locus 8p21.3) | CHB susceptibility                                  | (Li et al., 2016)                 |
|                                     | 2017 | Asian ancestry | HLA-DR          | HCV-related acute-on-chronic liver failure (ACLF) susceptibility | (Tan et al., 2017)                |
| **Hepatitis C virus (HCV)**         | 2010 | European ancestry | IFNL2 (previously known as IL28A), IFNL3/IFNL4, IFNL1 (previously known as IL29) | Chronic hepatitis C (CHC) susceptibility | (Rauch et al., 2010)              |
|                                     | 2011 | Asian ancestry | HLA-DQ, HLA-DR, MICA | HCV-related HCC                                    | (Kumar et al., 2011)              |
|                                     | 2011 | Asian ancestry | DEPDC5         | HCV-related HCC                                      | (Miki et al., 2011)               |
|                                     | 2012 | European ancestry | RNF7, MERTK | Liver fibrosis progression related to HCV infection | (Patin et al., 2012)             |
|                                     | 2013 | Asian ancestry | C6orf10, BTN2L2, BTN2L2-HLA-DRA | CHC-induced liver cirrhosis (LC)                     | (Jicabe et al., 2013)             |
|                                     | 2013 | European, African and Afro-Caribbean ancestry | IFNL4, HLA-DQ | Spontaneous resolution of HCV infection               | (Duggal et al., 2013)             |
|                                     | 2013 | Asian ancestry | HLA-DQB1, HLA-DQA1 | CHC susceptibility                                  | (Miki et al., 2013)               |

(continued on next page)
| Pathogen/disease | Year | Population | Associated loci | Trait description | References |
|------------------|------|------------|-----------------|------------------|------------|
| **Asian ancestry** | 2014 | European ancestry | HLA-DQA1, HLA-DRB1, NOTCH4 | Susceptibility to mixed cryoglobulinemia related to CHC | (Zignego et al., 2014) |
| **Japanese ancestry** | 2017 | Asian ancestry | TLL1 | HCC development after HCV eradication | (Matsuura et al., 2017) |
| **HCV and HIV co-infection** | 2016 | European ancestry | OXTR, MAP1LC3B1P1 | Liver fibrosis progression | (Ulveling et al., 2016) |
| **Varicella Zoster virus (VZV)** | 2015 | European, African, Hispanic ancestry | HCP5, HLA-B, DHFR | Herpes zoster susceptibility | (Crosslin et al., 2015) |
| **Human Immunodeficiency Virus (HIV)** | 2007 | European ancestry | HCP5, HLA-CHIV-1 viral load at set point | HIV-1 viral load at set point | (Fellay et al., 2007) |
| | 2009 | European ancestry | HCP5 | Progression to AIDS | (Limou et al., 2009) |
| | 2009 | European ancestry | HCP5, HLA-B, HLA-C | HIV-1 viral load at set point | (Fellay et al., 2009) |
| | 2010 | European ancestry | RYR3 | Susceptibility to atherosclerosis | (Shrestha et al., 2010) |
| | 2010 | European ancestry | HLA-C, MICA, HLA-B, HCP5, PSORS1C3 | HIV-1 viral load control | (International HIV Controllers Study et al., 2010) |
| | 2011 | European ancestry | PARD3B | Progression to AIDS | (Troyer et al., 2011) |
| | 2015 | European ancestry | AGKRI for Neutrophil count in HIV-infection; UGT1A, UGT1A1, UGT1A3, UGT1A4, UGT1A5, UGT1A6, UGT1A7, UGT1A8, UGT1A9, UGT1A10, MROH2A for Total bilirubin levels in HIV-infection | Multiple phenotypes measured by standard clinical laboratory practice | (Moore et al., 2015) |
| | 2015 | European ancestry | MICA, CCR2, HLA-B, HLA-C | HIV-1 viral load at set point | (McLaren et al., 2015) |
| **Human Papillomavirus (HPV)** | 2011 | European and Hispanic ancestry | HLA-DQB1 | HPV seroconversion | (Chen et al., 2011) |
| | 2013 | Asian ancestry | EXOC1, HLA-DPB2, GSDMB, rs9277952 (locus:6p21.32) | HPV-related cervical cancer | (Shi et al., 2013) |
| | 2013 | European ancestry | MICA, HLA-DRB1, HLA-DQA1, HLA | Susceptibility to cervical cancer. | (Chen et al., 2013) |
| | 2016 | European ancestry | rs73730372 (locus:6p21.32) | Susceptibility to cervical cancer. | (Chen et al., 2016) |
| **Influenza A virus (H1N1)** | 2015 | European ancestry | See Supplementary Table S1 | Susceptibility to influenza A (H1N1) infection and disease severity | (Garcia-Extebarria et al., 2015) |
| **Bacteria** | 2013 | European ancestry | TLR-1, HSPA6/FCGR2A/B | H. pylori seroprevalence defined using anti-H. pylori serum IgG antibody titers | (Mayerle et al., 2013) |
| **Helicobacter pylori** | 2010 | African ancestry | rs4331426 (locus 18q11.2) | Tuberculosis susceptibility | (Thye et al., 2010) |
| **Mycobacterium tuberculosis** | 2012 | African, Asian and European ancestry | WTI | Tuberculosis susceptibility | (Thye et al., 2012) |
| Pathogen/disease | Year | Population | Associated loci | Trait description | References |
|-----------------|------|------------|-----------------|------------------|------------|
| Leishmania donovani | 2014 | African ancestry and other | SMOC1, WT1 | Tuberculosis susceptibility | (Chimusa et al., 2014) |
|                  |      | The Gambia, Ghana, Malawi, South Africa | | | |
| Leishmania infantum chagasi | 2010 | African ancestry | ASAP1 | Tuberculosis susceptibility | (Curtis et al., 2015) |
| Porphyromonas gingivalis | 2016 | Subgingival bacterial biofilm | IL-12 | Tuberculosis resistance in HIV-positive individuals | (Sobota et al., 2016) |
| Streptococcus mutans | 2010 | European ancestry | GLT5D1 | Periodontitis susceptibility | (Schaefer et al., 2010) |
| Neisseria meningitidis | 2014 | African ancestry | CLEC19A, TRA, GGTA2P, TM9SF2, IFI16, RBBM3, C1QTNF7, TSNARE, HPVC1, SLC1A5A, PKP2, SNRPN | Periodontal disease-related phenotypes | (Öffenbacher et al., 2016) |
| Mycobacterium leprae | 2014 | European ancestry | MHC, HLA-DRB1-HLA-DQA1 | Visceral leishmaniasis susceptibility | (LeishGEN Consortium et al., 2013) |
| Salmonella typhi | 2014 | Asian and Hispanic ancestry | HLA-DRB1 | Enteric fever susceptibility | (Dunstan et al., 2014) |
| Staphylococcus aureus | 2015 | African ancestry | HLA-DRB1 | Susceptibility to meningococcal disease | (Davila et al., 2016) |
| Plasmodium falciparum | 2013 | Asian and Hispanic ancestry | HLA-DRB1-HLA-DQA1 | Severe malaria susceptibility | (Jallow et al., 2009) |
| Pathogen/disease | Year | Population | Associated loci | Trait description | References |
|------------------|------|------------|-----------------|------------------|------------|
| **African ancestry**<br>Discovery: The Gambia | 2012 | African ancestry | HBB, ABO, ATP2B4, MARVELD3 | Severe malaria susceptibility | (Timmann et al., 2012) |
| | | Replication: The Gambia and Ghana | | | |
| **African ancestry**<br>Discovery: The Gambia | 2013 | African ancestry | HBB, ABO | Severe malaria susceptibility | (Benn et al., 2013) |
| | | Replication: The Gambia and Ghana | | | |
| **African ancestry**<br>Discovery: Kenya, Malawi, The Gambia | 2015 | African ancestry | HBB, ABO, ATP2B4, FREM3/GYPE region | Severe malaria susceptibility | (Malaria Genomic Epidemiology Network et al., 2015) |
| | | Replication: Malawi, United Republic of Tanzania, Cameroon, Burkina Faso, The Gambia, Ghana, Mali | | | |
| **Prion (Creutzfeldt-Jakob disease)**<br>Discovery: UK | 2009 | European and Oceanian ancestry | PRNP | Susceptibility to variant CJD | (Mead et al., 2009) |
| | | Replication: UK and Papua New Guinea | | | |
| **Prion (Creutzfeldt-Jakob disease)**<br>Discovery: UK | 2011 | European ancestry | PRNP, MTMR7, NPAS2 | Susceptibility to variant CJD | (Sanchez-Juan et al., 2012) |
| | | Replication: France, UK | | | |
| **Prion (Creutzfeldt-Jakob disease)**<br>Discovery: Germany, Netherlands, UK | 2015 | European ancestry | PRNP, GRM8 | Susceptibility to sporadic CJD | (Sanchez-Juan et al., 2015) |
| | | Replication: Australia, Austria, France, Germany, Netherlands, Italy, Spain | | | |
| **GWAS for response to treatment for infectious diseases**<br>**Viruses**<br>Hepatitis C virus (HCV) | 2009 | European, African (African American or Afro-Caribbean) and Hispanic ancestry (NR) | IFNL3/IFNL4 (European Ancestry) | Response to antiviral treatment with pegylated interferon-alpha combined with ribavirin (pegIFNα/ribavirin) | (Ge et al., 2009) |
| | | Discovery: Australia | | | |
| | | Replication: Germany, Italy, UK, Australia | | | |
| **Viruses**<br>Hepatitis C virus (HCV) | 2009 | Asian ancestry | IFNL3/IFNL4 | Response to antiviral treatment with (pegIFNα/ribavirin) | (Tanaka et al., 2009) |
| | | Discovery: Japan | | | |
| | | Replication: Japan | | | |
| **Viruses**<br>Hepatitis C virus (HCV) | 2009 | European ancestry | IFNL3, IFNL3/IFNL4 | Response to antiviral treatment with (pegIFNα/ribavirin) | (Suppiah et al., 2009) |
| | | Discovery: Australia | | | |
| | | Replication: Germany, Italy, UK, Australia | | | |
| **Viruses**<br>Hepatitis C virus (HCV) | 2010 | European ancestry | IFNL2, IFNL3/IFNL4, IFNL1 | Response to antiviral treatment with (pegIFNα/ribavirin) | (Rauch et al., 2010) |
| | | Discovery: Germany, Switzerland | | | |
| **Viruses**<br>Hepatitis C virus (HCV) | 2010 | European, African American and Hispanic ancestry U.S. | ITPA | Susceptibility to treatment-related anaemia | (Fellay et al., 2010) |
| **Viruses**<br>Hepatitis C virus (HCV) | 2010 | Asian ancestry | ITPA | Susceptibility to treatment-related anaemia | (Ochi et al., 2010) |
| | | Discovery: Japan | | | |
| | | Replication: Japan | | | |
| **Viruses**<br>Hepatitis C virus (HCV) | 2011 | Asian ancestry | IFNL3/IFNL4 | HCV treatment response | (Ochi et al., 2011) |
| | | Discovery: Taiwan, Japan | | | |
| **Viruses**<br>Hepatitis C virus (HCV) | 2011 | Asian ancestry | DDRGK1, ITPA for Hb level and platelets count | Susceptibility to pegIFNα-induced thrombocytopenia and to RBV-induced anaemia | (Tanaka et al., 2011) |
| | | Discovery: Japan | | | |
| | | Replication: NR | | | |
| **Viruses**<br>Hepatitis C virus (HCV) | 2012 | European and African (African-American and Afro-Caribbean) and Hispanic ancestry U.S. | ITPA | Susceptibility to thrombocytopenia, neutropenia, and leukopenia related to pegIFNα treatment | (Thompson et al., 2012) |
| | | Discovery: Hispanic, European, and African (African American or Afro-Caribbean) (NR) | | | |
| **Viruses**<br>Hepatitis C virus (HCV) | 2012 | Hispanic, European, and African (African American or Afro-Caribbean) (NR) | IFNL3/IFNL4 | Lipid levels during and after HCV infection treatment and the interactions with sustained viral response (SVR) | (Clark et al., 2012) |
| **Human Immunodeficiency Virus (HIV)** | 2011 | Asian ancestry | CCHCR1 | Susceptibility to Nevirapine-induced rash | (Chantarangsu et al., 2011) |
| | | Discovery: Thailand | | | |
| | | Replication: Thailand | | | |
| **Human Immunodeficiency Virus (HIV)** | 2014 | European, Hispanic, African (African American or Afro-Caribbean) ancestry U.S. | RSPQ4 (hispanic - Grade ≥ 2); IL2RA, CXCL12, rs9501753 (locus:6p25.3), PTPRB (European ancestry - Grade ≥ 3); TANK, HDAC9, rs801350 | Susceptibility to peripheral neuropathy related to d4T/ddl-containing treatment | (Leger et al., 2014) |
infectious diseases or other conditions/traits. Although the sample size that is necessary to achieve an adequate statistical power depends on the genetic architecture of the analyzed trait, GWAS typically require large samples of individuals. This is because thousands (millions) of variants are simultaneously analyzed, making it difficult to disentangle true associations from false positives. Ideally, GWASs should enroll subjects of shared and homogeneous ethnical ancestry, to avoid issues of hidden relatedness, they replicated a previously reported association at the HLA-DQA1 locus and performed transethnic fine mapping of the association signals (Table 1). The authors also found a positive correlation between San (an African ethnic group) and hidden relatedness, they replicated a previously reported association at the HLA-DQA1 locus and performed transethnic fine mapping of the association signals (Table 1). The authors also found a positive correlation between San (an African ethnic group)

### 3.1. Sample size, population structure, and environmental confounders

Large populations samples are often difficult to recruit in developing countries, due to limited resources, inadequate health facilities, and inefficient screening or diagnostic procedures. In these countries, however, the burden imposed by infectious diseases is highest (Jones et al., 2008; Mabey et al., 2004). Also, some populations in developing areas have complex demographic histories. This is the case of African populations, that are genetically highly diverse and display a limited extension of linkage disequilibrium (LD) compared to non-Africans (International HapMap Consortium et al., 2007). For instance, one of the first GWASs for infectious diseases genotyped individuals in West Africa to identify susceptibility variants for severe malaria (Jallow et al., 2009). The authors found considerable population stratification, which was corrected for using principal components analysis (PCA); signals of association at known malaria resistance loci (e.g. HBB) were difficult to retrieve due to weak LD between causal variants (e.g., HbS locus) and tag SNPs (Jallow et al., 2009). Populations from South and Central America also display relevant population structure due to recent admixture (Bryc et al., 2010). In a GWAS for visceral Leishmaniasis that included families from Brazil, ancestry differences and close relationships were efficiently corrected for using a linear mixed model (LeishGEN Consortium et al., 2013). These models are an alternative to those based on PCA and rely on the incorporation of genetic relatedness between individuals directly in the statistical model (see (Eubshunthornwattana et al., 2014; Hayes, 2013; Yang et al., 2014) for reviews and comparison among different linear mixed model methods).

Another issue often related to population structure is that of environmental confounders (Vilhjalmsson and Nordborg, 2013). A well-known example in the field of non-communicable diseases relates to type 2 diabetes (T2D) susceptibility. Several studies for T2D indicated that the proportion of Native American genetic ancestry is associated with lower socioeconomic status in admixed Latino populations from North and South America (Chakraborty et al., 1986; Florez et al., 2009; Martinez-Marignac et al., 2007; Parra et al., 2004). This effect is partially responsible for the generally higher incidence of T2D in Latinos. This exemplifies how the combined effect of genetic admixture and environmental factors has the potential to affect genetic associations. Indeed, poor socioeconomic conditions represent a risk factor for several infectious diseases (Baker et al., 2012; Braveman, 2011; Franco-Paredes et al., 2007; May, 2007; Menegus, 2010). Thus, possible environmental confounders must be accounted for to avoid spurious associations or loss of statistical power. However, even in the presence of admixture and environmental confounders, associations can be retrieved with confidence. In a GWAS for tuberculosis (TB), Chimusa and coworkers analyzed admixed South African Coloured case-control cohorts (Chimusa et al., 2014). After accounting for population stratification and hidden relatedness, they replicated a previously reported association at the WT1 gene (Thye et al., 2012) and performed transethnic fine mapping of the association signals (Table 1).

| Pathogen/disease | Year | Population | Associated loci | Trait description | References |
|------------------|------|------------|-----------------|-------------------|------------|
| Measles, Mumps and Rubella vaccine | 2014 | European ancestry U.S. | ACO1, PITRD | Immune responses to rubella vaccination | (Kennedy et al., 2014) |
| | 2014 | European ancestry Discovery: Denmark Replication: Denmark | IFI44L, rs1318653 (locus:1q32.2), CD46, ANO3, SCN1A, rs11105468 (locus:12q21.33), SCN2A for febrile seizures; IFI44L, SCN1A, ANO3, CD46, CD34 for febrile seizures MMR vaccine-related; ANO3, SCN1A for febrile seizures MMR vaccine-unrelated | Susceptibility to general and measles, mumps and rubella (MMR) vaccine-related febrile seizures | (Penny et al., 2014) |
| Hepatitis B vaccine | 2011 | Asian ancestry Discovery: Indonesia Replication: Indonesia | HLA-DR, HLA-DPB1 | Immune response to HBV vaccine | (Ovsyannikova et al., 2012) |
| | 2014 | Asian ancestry Discovery: China Replication: China | HLA-DRB1, BTLN2, HLA-DRB1, C6orf10 | Immune response to HBV vaccine | (Ovsyannikova et al., 2012) |
| Smallpox vaccine | 2012 | Hispanic, European and African (African-American and Afro-Caribbean) ancestry U.S. | MKE, rs10503951 (locus: 8p12), GPR12, ZDH2, SPRE1 (African ancestry); PCDH15, PRKCIQ (Hispanic ancestry) | Immune response to smallpox vaccine | (Kennedy et al., 2012) |
| | 2012 | European and African (African-American and Afro-Caribbean) ancestry U.S. | See Supplementary Table S1 | Cytokine responses to smallpox vaccine | (Kennedy et al., 2012) |

* Articles not in the GWAS catalog (as of April 24th, 2017).
ancestry proportion and TB status. Although complete information was available for a minority of individuals, no correlation was detected between socioeconomic status and ancestry components, leading the authors to suggest that the association between San ancestry and TB is not merely explained by differences in socioeconomic conditions (Chimusa et al., 2014).

3.2. Phenotype heterogeneity and case/control misclassification

Perhaps, the two most important points that characterize GWAS for infectious diseases compared to studies of non-communicable conditions are phenotypic heterogeneity and case-control misclassification. An important source of heterogeneity in GWASs for infectious diseases derives from genetic variation in the pathogen. One of the best known examples of how the genotype of the infectious agent can interact with the host genome refers to hepatitis C virus. Among the most significant findings of GWASs for infectious diseases was the identification in 2009–2010 of variants at the IFNL3/IFNL4 loci (previously known as IL28B) that associate with spontaneous clearance of HCV infection and with response to interferon-α/ribavirin therapy (Ge et al., 2009; Rauch et al., 2010; Suppiah et al., 2009; Tanaka et al., 2009) (Table 1). This finding set the basis for a personalized treatment of HCV infection (Matsuura et al., 2014). However, the effect of the IFNL3/IFNL4 variants on response to therapy is much stronger for patients infected with HCV genotypes 1 and 4 than for those infected with genotypes 2, 3, and 6 (Rauch et al., 2010; Akkarathamrongsin et al., 2014). Interestingly, interferon-α/ribavirin therapy is less effective for patients infected with genotypes 1 and 4 compared to the other genotypes (European Association for the Study of the Liver, 2011). Thus, the prognostic value of the IFNL3/IFNL4 variants is particularly relevant for patients infected with HCV genotypes that are poorly responsive to this treatment (Rauch et al., 2010). These observations well exemplify how important the interaction of the host and pathogen genomes can be. Very recently, a genome-to-genome GWAS for HCV infection was performed, shedding more light into the interaction between human IFNL4/IFNL3 variants and HCV genetic diversity (Ansari et al., 2017) (see Conclusions and perspectives section).

Another notable example of how pathogen heterogeneity can interact with the host genotype refers to TB susceptibility. The human Mycobacterium tuberculosis complex (MTBC) consists of several major phylogenetic lineages whose names reflect their association with geographic areas (Brites and Gagneux, 2015). Candidate gene approaches have shown that host genetic variants can modulate susceptibility to TB caused by specific MTBC strains (Herb et al., 2008; Caws et al., 2008; Intemann et al., 2009; van Crevel et al., 2009; Salie et al., 2014). For instance, a variant in TLR2 was specifically associated with an increased risk of TB caused by the Beijing strain in a Vietnamese population, whereas two SLC11A1 polymorphisms were found to be significantly more common in patients having tuberculosis caused by M. tuberculosis Beijing genotype strains than in patients carrying other genotypes (Caws et al., 2008; van Crevel et al., 2009). In populations from Ghana, an IRGM polymorphism was found to protect against disease caused by the Euro-American lineage (Intemann et al., 2009), and polymorphisms in ALOX5 were more strongly associated with TB caused by M. africanum West African 2 strain (Herb et al., 2008). Furthermore, associations of specific HLA class I alleles and disease caused by the Beijing, LAM, LCC and Quebec strains, as well as by the Euro-American or East Asian lineages, were found in a South African population (Salie et al., 2014).

Therefore, the genetic heterogeneity of MTBC may limit the reproducibility of GWASs, as different cohorts may indeed include patients with distinct TB epidemiologies and with TB caused by different M. tuberculosis strains.

Finally, misclassification of cases or controls is an issue for several association studies. In the case of infections, a major problem is the difficulty to assess exposure in the control population. For instance, it is sometimes impossible to determine whether individuals recruited as uninfected controls (e.g., based on serology) are seronegative because they were exposed but did not acquire the infection or because they were never or rarely exposed to the pathogen. Cases are also sometimes problematic to define. For instance, seropositivity can result from different routes of exposure, in turn associated with distinct probabilities of infection, and possibly modulated by diverse genetic factors. Moreover, the probability of infection may depend on several factors, including the genetic characteristics of the pathogen, the infectiousness of the transmitting individual, the general health status at the moment of exposure (this is clearly important for nosocomial infections, for instance).

Some of these issues are evident in a GWAS for susceptibility to HIV-1 infection. McLaren and coworkers combined information from multiple studies to obtain case-control cohorts of > 6300 HIV-1 positive cases and 7200 general population HIV-1 negative controls (McLaren et al., 2013). Cases derived from different studies and were infected via diverse routes (i.e., parenteral and sexual). No information was available for controls on HIV-1 exposure (e.g., at risk behavior or possible parenteral exposure). The study failed to detect association with HIV infection susceptibility other than the known CCR5Δ32 variant. An association signal at the HLA-B/HLA-C genes was not confirmed after correcting for frailty (survival) bias (McLaren et al., 2013). These results led to the conclusion that genetic variants that modulate HIV-1 acquisition are either rare or have small effects. However, as the authors note, the study design has a high potential for misclassification of both cases and controls (McLaren et al., 2013). Indeed, cases may include subjects with low HIV-1 susceptibility who were infected due to very high exposure (e.g. via multiple blood transfusions), whereas controls may comprise subjects who never exposed themselves to the virus. Because studies on high-risk populations indicated that the proportion of subjects who are naturally resistant to HIV-1 is around 20% (Plummer et al., 1999; Fowke et al., 1996), it can be hypothesized that as many as 80% of controls would seroconvert if exposed. McLaren and coworkers estimated that the sample size of their GWAS provided 80% power for variants with 0.1 allele frequency and genotype relative risk (GRR) of 1.3. Their calculations also show that if ~30%, ~60%, or ~80% of cases were misclassified as controls, GRRs of 2, 3, or 4 would be respectively necessary to achieve the same power (McLaren et al., 2013). In fact, the authors did detect association for the CCR5Δ32 variant, indicating that polymorphisms with strong effect could be identified.

Another situation that may potentially lead to classification biases is reinfection. For instance, HCV reinfection is relatively common in high-risk groups (e.g., drug users) (Grebelly et al., 2009; Grebelly et al., 2012; Midgard et al., 2016). If individuals are classified as persistently infected (chronic hepatitis C, CHC) based on viral loads measurements with relatively long intervals between tests, cases of clearance and reinfection can potentially be classified as persistent infections (Grebelly et al., 2009; Grebelly et al., 2012).

4. GWAS for viral infections

4.1. HIV-1

In 2007, Fellay and coworkers performed the first GWAS for an infectious disease by investigating viral set-point in HIV-1 infection (Fellay et al., 2007). This pilot study paved the way for several other GWASs. Indeed, the two signals detected within the HLA class I gene region were since confirmed by several studies (Fellay et al., 2009; Limou et al., 2009; International HIV Controllers Study et al., 2010) (Table 1, Fig. 2). Additional GWASs focused on HIV-1 viral load, progression to AIDS, and susceptibility to infection (Table 1). Their results have been extensively reviewed elsewhere (van Manen et al., 2012). In recent years, GWASs for HIV-1 infection have also focused on understanding the host determinants of response to standard and
experimental therapies, as well as on detecting variants responsible for susceptibility to treatment-related adverse effects (Table 1 and Supplementary Table S1) (Chantarangsu et al., 2011; Leger et al., 2014; Wanga et al., 2015) (Table 1 and Supplementary Table S1).

4.2. HBV and HCV

Susceptibility to Hepatitis B and Hepatitis C viruses (HBV and HCV) was also extensively investigated in GWASs. Worldwide, chronic HBV or HCV infections represent the major causes of progressive liver disease, including liver cirrhosis (LC) and hepatocellular carcinoma (HCC) (Shirvani-Dastgerdi et al., 2016). HBV predominantly affects Southeast Asia and the East Pacific, where it reaches the highest prevalence in the world (Custer et al., 2004). It is thus not surprising that most GWASs performed to date genotyped individuals of Asian ancestry (Table 1).

Genetic variations in HLA class II loci (HLA-DP and HLA-DQ) were significantly associated with susceptibility to chronic HBV infection and with HBV-related HCC (Mbarez et al., 2011; Kamatani et al., 2009; Nishida et al., 2012; Jiang et al., 2013; Jiang et al., 2015; Li et al., 2012; Kim et al., 2013; Hu et al., 2013; Chang et al., 2014; Tan et al., 2017) (Table 1). Interestingly, host genetic variants in HLA class II genes were also found in association with hepatitis B vaccine response (Png et al., 2011; Pan et al., 2014) (Table 1).

In 2015, Jiang et al. (2015) performed the largest GWAS on a Chinese population including about 18,400 among cases (chronically-infected) and controls (seronegative). The study confirmed most of the previously reported associations in HLA class II genes and detected a new variant, rs1883832 in CD40, strongly associated with HBV chronic infection (Table 1, Fig. 2). rs1883832 influences CD40 expression by affecting its translational efficiency (Jacobson et al., 2005). This association was confirmed, although not at genome-wide significant level, by a recent GWAS that analyzed chronically HBV-infected and spontaneously recovered subjects (Li et al., 2016). In this study, the authors also identified an intergenic variant at 8p21.3 which represents an expression quantitative trait locus (eQTL) for the INTS10 (integrator complex subunit 10) gene (Table 1). Interestingly, the authors demonstrated that INTS10 suppresses HBV replication in liver cells via IRF3, and confirmed decreased INTS10 protein levels in plasma samples of chronically HBV-infected patients compared to subjects who spontaneously cleared infection (Li et al., 2016).

Although these studies have opened up new horizons of research, analyses in populations of different ancestries are strictly required.

As mentioned above, a landmark discovery of GWAS for HCV infection was the identification of the IFNL3/IFNL4 association. In addition to this locus, variants in HLA class II loci were also found to affect viral clearance and disease progression in chronic HCV infection (Table 1, Fig. 2). In addition, polymorphisms in apoptosis-related genes such as RNF7, TULP1, and MERTK were associated with liver fibrosis, whereas variants in DEPDC5 and MICA were described as modulators of HCV-related hepatocellular carcinoma risk (Table 1, Fig. 2).

4.3. Other widespread viral infections

The GWA strategy was also exploited to investigate other infections caused by viral pathogens with worldwide diffusion. These include Human Papillomavirus (HPV), Varicella Zoster virus (VZV), and Influenza A virus (IAV). Overall, these works unveiled a limited number of associations, sometimes as a consequence of small sample sizes and, consequently, modest statistical power.

In the case of HPV, most GWASs for susceptibility to infection and progression to cervical carcinoma confirmed associations to HLA class II genes (Table 1, Fig. 2). It should be noted, however, that some of these studies did not directly test HPV infection status, nor HPV type, making it difficult to determine which trait these associations refer to. An interesting observation is that polymorphisms in the EXOC1 and GSDMB genes were associated with progression to cervical carcinoma (Shi et al., 2013) (Table 1). EXOC1 encodes a subunit of the exocyst complex involved in host innate immune response against DNA antigens (Ishikawa et al., 2009), whereas GSDMB codes for the cancer-associated gasdermin-like protein, which is highly expressed in cervical cancer cells (Sun et al., 2008). These data suggest that both immunological and carcinogenic factors contribute to the risk of cervical cancer development.

With respect to VZV, a GWAS was performed to investigate risk factors for herpes zoster (Crosdin et al., 2015). Although the study had limited power and potential for control misclassification (e.g., individuals who never suffered from primary infection with VZV are not at risk of virus reactivation), a significant association was detected in the MHC region where the non-coding HCP5 (HLA Complex P5) gene maps (Table 1). Intriguingly, variants in this region were also associated with delayed progression to AIDS (Table 1, Fig. 2). HCP5 derives from an endogenous retroviral element with sequence homology to the HIV-1 pol gene (Kulski and Dawkins, 1999) and is primarily expressed in immune system cells (Liu et al., 2008). The function of this noncoding gene and its role (if any) in restricting viral infection remain to be determined. In fact, a functional study showed that HCP5 does not restrict HIV-1 infection in vitro (Yoon et al., 2010), and refinement of the association signals within the HLA region indicated that the HCP5 variant associated with HIV-1 control is indeed a marker for the HLA-B*5701 allele (Fellay et al., 2009; Trachtenberg et al., 2009). Likewise, HCP5 variants that reached genome-wide significance in the VZV GWAS are in strong linkage disequilibrium with polymorphisms in the HLA-B gene region (Crosdin et al., 2015). It is thus possible that HCP5 has no effect on viral infection per se, but its variants tag specific HLA class I alleles/haplotypes.

As for IAV, a single study analyzed susceptibility to H1N1 virus infection (Garcia-Exeberria et al., 2015). The authors found no genome-wide significant associations when severe influenza cases were compared to mild cases. Conversely, several SNPs reached the significance threshold when severe and mild cases were compared to the general population (Table 1 and Supplementary Table S1). However, these associations were obtained on very small samples of patients and controls, and no replication cohort was analyzed.

Finally, we draw attention to a recent GWAS that searched for genetic variants associated with diarrhoeal episodes in young children (1–2 years old) (Bustamante et al., 2016). All subjects were recruited in developed countries, where the majority of diarrhoeal episodes are caused by viral infections (Wiegering et al., 2011). Despite the potential issues of phenotype heterogeneity, a significant association was found for variants in FUT2, a gene encoding an alpha (1,2)-fucosyltransferase which participates in the production of histo-blood group antigens (Table 1). In particular, the activity of FUT2 determines the expression of the ABO histo-blood group antigens on the gastrointestinal mucosa and in bodily secretions. Genetic diversity at the FUT2 gene is maintained in human populations by balancing selection (Koda et al., 2001; Fumagalli et al., 2009) and common FUT2 null alleles are present in many populations (Kelly et al., 1995; Koda et al., 1996; Liu et al., 1998). In homozgyotes, these alleles determine the “non-secretor” phenotype. In the GWAS for diarrhoeal episodes, non-secretors were found to be at lower risk compared to secretors, a result in line with candidate-gene studies that associated the non-secretor status with protection from Rotavirus and Norovirus infections (which represent the leading causes of diarrhoeal disease in developed regions) (Wiegering et al., 2011; Lindesmith et al., 2003; Carlsson et al., 2009; Imbert-Marcille et al., 2014; Thorven et al., 2005). In line with the balancing selection scenario, however, candidate-gene studies associated the non-secretor status with higher risk for infection by different pathogens, including Neisseria meningitidis, Streptococcus pneumoniae, and Candida spp. (Blackwell et al., 1986a; Blackwell et al., 1986b; Chaim et al., 1997). More recently, non-secretors were also found to be at higher risk to develop Crohn’s disease (Franke et al., 2010; McGovern et al., 2010). The molecular mechanism underlying the different
The predisposition to infection (and to Crohn's disease) of secretors and non-secretors most likely stems from the fact that both pathogenic and commensal microorganisms exploit oligosaccharides on the gastrointestinal mucosa for cytoadherence.

5. GWASs for bacterial infections

Studies on the genetic susceptibility to bacterial infections have investigated a wide variety of conditions, either resulting from specific
infections (e.g., tuberculosis, pneumococcal bacteremia) or deriving from more heterogeneous situations (e.g., periodontitis).

5.1. Specific infections

GWASs for infections caused by Neisseria meningitidis (Davila et al., 2010; Martinon-Torres et al., 2016), Helicobacter pylori (Mayerle et al., 2013), and Mycobacterium leprae (Zhang et al., 2009), strongly pointed to a role for risk variants in genes of the innate immune system. For H. pylori, variants in TLR1, known to be essential for innate immunity against bacterial infection, were identified as strongly associated with seroprevalence (Table 1, Fig. 2). For Neisseria meningitidis, two GWASs for meningitis and septicemia identified variants in the CFH gene region (Davila et al., 2016; Martinon-Torres et al., 2016) (Table 1, Fig. 2). CFH encodes complement factor H, a regulator of the complement system, which is bound by Neisseria meningitidis to escape complement mediated killing (Schneider et al., 2009).

In the case of leprosy, a single GWAS identified candidate susceptibility genes encoding proteins involved in innate immune response (NOD2, TNSSF15, and RIPK2) (Table 1, Fig. 2). Furthermore, in a GWAS for leprosy in the Chinese population, H. Liu et al. (2015) found an overlap between their hits and those for autoimmunity/inflammatory diseases. For example, association with CCDC88B, identified in the leprosy GWAS (Table 1), was also reported for primary biliary cirrhosis, sarcoidosis, and inflammatory bowel disease. Likewise, SNPs in TNSSF15, LRRK2, IL18RAP/IL1RL1, and LACC1/CCDC122 represent susceptibility loci for both leprosy and Crohn’s disease (Jostins et al., 2012; J.Z. Liu et al., 2015). This observation points to a shared immunological basis for infectious and inflammatory conditions (H. Liu et al., 2015). However, the study also showed that, whereas some shared loci (RIPK2 and LACC1/CCDC122) had concordant effects on the risk for leprosy and for inflammatory/autoimmune diseases, some others (e.g., IL18RAP/IL1RL1) showed opposite or discordant risk effects. These findings suggest a complex scenario and a delicate balance between response to infection and autoimmunity/chronic inflammation (Sironi and Clerici, 2010).

As expected, genes of the adaptive immune system also turned out to harbor polymorphisms associated with susceptibility to infections. Variants in the MHC region were identified as risk factors for leprosy (Zhang et al., 2009; H. Liu et al., 2015), Staphylococcus aureus infection (De Lorenze et al., 2016), enteric fever caused by Salmonella typhi (Dunstan et al., 2014), and TB (Sveinbjornsson et al., 2016) (Table 1, Fig. 2). For this latter infection, several GWASs have been performed (Table 1 and Supplementary Table 1). The WT1 association mentioned above was detected in two independent studies (Chimusa et al., 2014; Thye et al., 2012) and recently confirmed, although with a p value that did not reach genome-wide significance, in a family-based Moroccan discovery sample (Grant et al., 2016). The functional role played by WT1, a zinc-finger transcription factor, in TB susceptibility remains to be elucidated. An interesting possibility is that WT1 regulates cytokine expression or activation of the vitamin D receptor (Maurer et al., 2001).

Indeed, the role played by ASAP1, a gene associated to TB in a Russian discovery cohort and replicated in African populations (Table 1), is an independent predictor of pneumococcal bacteremia in febrile children (Kuppermann et al., 1998).

Thousands of lncRNA genes have been identified in the human genome (Atianand and Fitzgerald, 2014) and growing evidence suggests that these molecules are involved in gene regulation in different cell-types and tissues, including the immune system (Atianand and Fitzgerald, 2014). In 2013, using inter-crosses of mouse strains, Gomez and coworkers provided the first direct demonstration that a lncRNA (termed NeST) can modulate the outcome of infection (Gomez et al., 2013). In particular, NeST was shown to be responsible for the persistence of Thielier’s virus in the central nervous system, as well as for the clearance of Salmonella infection (Gomez et al., 2013). These data suggest that additional lncRNA await identification as important players in immune response and infectious disease susceptibility.

5.2. Caries, periodontitis, and acne

GWASs have also been applied to identify genetic determinants for very common human condition characterized by a strong interaction between environmental factors and the infecting commensal bacteria. These include caries, which is caused by a wide array of carbohydrate-fermenting bacteria, (Larsen and Fiehn, 2017), periodontitis, where an anaerobic bacterial biofilm participates in disease onset and progression (Larsen and Fiehn, 2017), as well as acne, which is characterized by overgrowth of Propionibacterium acnes.

For dental caries a variety of loci were identified via GWASs. Studies in adults have yielded significant and “suggestive” associations within or near genes with roles in tooth development and host defense. Among these we mention LYZL2, encoding a bacteriolytic agent (Shaffer et al., 2013), NAMPT, which is involved in periodontal healing (Morrison et al., 2016), and BMP7, a tooth development gene (Morrison et al., 2016) (Table 1). To date, none of the dental caries loci identified in GWASs in adults have been followed up in fine-mapping or replication studies. Importantly, given the complexity of caries etiology and the spectrum of contributing environmental factors, some authors speculated that the effects of some genetic variants may vary across different populations (Morrison et al., 2016).
Although several GWASs were conducted for periodontitis, only few of them successfully identified implicated loci (Table 1 and Supplementary Table S1). Among them, the first GWAS for periodontitis (Schafer et al., 2010) studied the aggressive type of disease and identified an association with a marker in the GLT6D1 gene (encoding a glycosyltransferase) (Table 1). Functional experiments suggested that reduced GATA3 binding affinity to the GLT6D1 locus could be a component of the pathophysiology of periodontitis.

Subsequently, Offenbacher et al. (2016) used a promising approach that combined clinical phenotypes, biological intermediates of microbial burden, and measures of local inflammation to derive periodontal complex traits (PCTs). PCTs were carried forward to GWAS to identify PCT-associated loci among European American adult participants (Table 1). The authors found several significant signals in loci that included genes involved in immune response and epithelial barrier function. However, candidate loci did not associate with current clinically determined periodontal disease categories upon replication.

Finally, GWASs were also used to find associations between genetic variants and severe acne susceptibility. The top signals identified in a Chinese population included a SNP within the promoter region of DDB2 (a novel androgen receptor-interacting protein), as well as variants in SELLP/SELP:SELE gene cluster (Table 1). These selectins have important roles in regulating cutaneous inflammation (He et al., 2014) (Fig. 2). Instead, a GWAS conducted in a UK population identified genes linked to the TGFβ cell signaling pathway and to skin homeostasis, namely OVO1, PST and TGB2 (Navarin et al., 2014) (Table 1).

6. GWASs for parasitic and prion diseases

Parasitic infections, either caused by protozoa or by helminths, affect millions of people worldwide. For many parasitic diseases a heritable component has been demonstrated (Verra et al., 2009; Williams-Bangero et al., 2011; Dold and Holland, 2011; Choi et al., 2003). However very few GWAS were conducted; their results are summarized in Table 1 and were reviewed elsewhere (Mangano and Modiano, 2014). We add herein that a recent GWAS explored the genetic susceptibility to malaria (Malaria Genomic Epidemiology Network et al., 2015). Beyond confirming loci previously associated with severe malaria (HBB, ABO, ATP2B4) (Jallow et al., 2009; Timmann et al., 2012; Band et al., 2013), this study identified additional variants on chromosome 4, between FREG3 and a cluster of three glycoprotein genes (GYPE, GYPB, and GYPV) (Table 1). Glycoprophins are sialoglycoproteins abundantly expressed on the erythrocyte membrane, where they play a functional role in invasion by Plasmodium falciparum. In a recent follow-up study, the authors identified an array of large copy number variants (CNVs) affecting the GYPV and GYPB genes (Leffler et al., 2017). One of the identified CNVs (DUP4) is associated with resistance to severe malaria and explains the GWAS association signal (Malaria Genomic Epidemiology Network et al., 2015).

The most common form of human prion disease in humans is Creutzfeldt-Jacob disease (CJD), which is classified as sporadic, acquired (iatrogenic or variant) or familial (Iwasaki, 2017). Variants in PRP, encoding the prion protein (PrP), are the best known genetic factor for susceptibility to CJD (Owen et al., 1990; Mead et al., 2009) (Table 1). Nevertheless, associations to other genomic loci (MTMR7 and NPPS2 for variant CJD (Sanchez-Juan et al., 2012), GRMS for sporadic CJD (Sanchez-Juan et al., 2015)) were recently reported (Table 1).

MTMR7 is specifically expressed in the central nervous system (CNS) and is involved in the phosphatidylinositol pathway, whereas NPPS2 encodes a transcription factor implicated in neuronal function. Interestingly, the CJD susceptibility variant in NPPS2 is in strong linkage disequilibrium with an SNP that regulates PLCDD3 in trans (Sanchez-Juan et al., 2012). Because PLCDD3 encodes a catabolic enzyme of the phosphatidylinositol pathway, these data implicate this pathway in the susceptibility to variant CJD. As for GRMS, it encodes a member of the metabotropic glutamate receptor family. Other members of this family are involved in cellular signal transduction triggered by PrP (Um et al., 2013). Thus, these data highlight the ability of GWASs to pinpoint molecular disease pathways.

7. GWAS in non-human mammals

In comparison to human GWASs, association studies in domestic and laboratory animals offer advantages and disadvantages. Model animals, for instance, can be infected with a known dose of a genetically homogeneous pathogen, removing part of the variability associated to exposure routes and load in humans, as well as to the genetic heterogeneity of the infecting pathogen. Due to their peculiar demographic histories, both model and domestic animals often display extensive linkage disequilibrium and genetic homogeneity, at least within breeds or strains (Alhaddad et al., 2013; Badke et al., 2012; Boyko et al., 2010; Flint and Eskin, 2012; Lindblad-Toh et al., 2005). This implies that GWASs can be carried out using much fewer markers and samples than required in human studies. The other side of the coin is that the often complex breeding strategies pose challenges related to population structure, cryptic relatedness, and extensive selective sweeps resulting from artificial selection. These factors need to be accurately corrected for, to avoid spurious associations and loss of statistical power. Popular methods to account for these effects include the use of PCA to explicitly model ancestry contributions (Price et al., 2006) and the application of mixed model association methods (see (Hayes, 2013; Yang et al., 2014; Flint and Eskin, 2012) for review).

7.1. Mouse studies

In the field of mouse genetics, an important contribution to the mapping of complex traits came from the Collaborative Cross (CC), a collaborative effort aimed at providing the scientific community with a large panel of recombinant inbred mouse strains derived from genetically diverse founders (Churchill et al., 2004; Maurizio and Ferris, 2017). As founders are both classic inbred strains and wild-derived strains, the resulting mouse panel has higher genetic diversity and less population structure than other mouse-based resources. Within the CC framework, an ancestry-based approach was shown to be superior to marker-based methods for mapping QTLs (Aylor et al., 2011). The utility of this approach in the field of infectious diseases was initially demonstrated by a genome-wide scan of incipient CC lines for susceptibility to Aspergillus fumigatus, a pathogen that recapitulates in mice the signs of human aspergillosis (Durrant et al., 2011). The authors found that CC lines were heterogeneous in terms of susceptibility to A. fumigatus infection, measured as survival days after infection, and the broad-sense heritability of this phenotype was 0.78. The genome-wide scan identified several QTLs, including one on chromosome 8 that contained the Irf2 gene and another one on chromosome 10 covering cytokine (IL20ra, IL22ra) and interferon (Ifng1) receptors (Durrant et al., 2011). Incipient CC lines were then used to investigate the genetic susceptibility to IAV infection by recording several IAV-induced phenotypes, including virus replication, airway inflammation, weight loss, and pulmonary edema. Interestingly, a highly significant QTL on chromosome 16 (Hr1, Host response to Influenza), was identified (Ferris et al., 2013). Hr1 explained a considerable proportion of variation in several phenotypes and encompassed a genomic region where the Mx1 gene maps. Because Mx1 is a well-known IAV resistance gene (Staeheli et al., 1988), these findings provide a nice validation of the CC approach. Additional QTLs were mapped in the study and suggested that variants in Ifi16 and/or Nox4 (Hr12) modulate IAV-induced weight loss (Ferris et al., 2013). More recently, incipient CC lines were used to identify variants in Trim55 associated with vascular cuffing after infection with a mouse adapted SARS-CoV strain (Gralinski et al., 2015) and QTLs for survival to Klebsiella pneumoniae infection (Vered et al., 2014). Importantly, CC mice have been used to create improved models of EBOV (Ebola virus)
and WNV (West Nile virus) infection (Graham et al., 2016; Graham et al., 2015; Rasmussen et al., 2014) and mouse populations showing diverse susceptibility to M. tuberculosis and Pseudomonas aeruginosa (Lore et al., 2015; Smith et al., 2016). These models awaited genome-wide scans to unveil the genetic determinants of infection susceptibility and severity.

7.2. Associations in dogs

Dogs have also proven useful models for the mapping of variants associated with infectious disease phenotypes. Dog breeds represent genetic isolates deriving from a few founder individuals and subject to strong artificial selection. Thus, breeds differ in several phenotypes. One of these is granulomatous colitis, which is caused by mucosally invasive Escherichia coli (Simpson et al., 2006) and is described in boxers and bulldogs only (Craven et al., 2011; Manchester et al., 2013).

Within a larger project that assessed several traits and diseases, Hayward and coworkers performed a within-breed GWAS for granulomatous colitis and identified a strong association signal within a genomic region where several members of the SLAM (signaling lymphocyte activation molecule) gene family map (Hayward et al., 2016). Interestingly, variants in the corresponding human genomic region have been associated with the susceptibility to Crohn's disease and ulcerative colitis (Franke et al., 2010; J.Z. Liu et al., 2015; Barrett et al., 2008). These data indicate that, both in dogs and in humans, genes within this region contribute to the maintenance of intestinal immune homeostasis in the presence of commensal or pathogenic bacteria.

7.3. GWASs in cattle and swine

In cows, GWASs have been widely applied to the field of infectious disease susceptibility. This is clearly motivated by the fact that infections cause a major economic burden in the cattle industry worldwide. In this context, GWASs and other genomic approaches are regarded as powerful strategies to develop breeding programs aimed at the generation of less susceptible livestock (Raszek et al., 2016).

One of the most investigated traits was Johne's disease, a chronic gastrointestinal tract disease caused by Mycobacterium avium subspecies paratuberculosis. Several GWASs have been performed, mostly on naturally infected Holsteins and Jerseys (common breeds of dairy cattle): associated variants with small effect were identified in several analyses, but reproducibility among studies was extremely low, resulting in no confidently associated marker (Alpay et al., 2014; Kirkpatrick et al., 2011; Minozzi et al., 2010; Neibergs et al., 2016; Pant et al., 2010; Settles et al., 2009; Van Huizen et al., 2012; Zare et al., 2014). Although the reasons for low across-study consistency are likely different criteria to classify phenotypes and different statistical procedures, these results suggest that loci with a major effect on Johne's disease do not exist, at least in these breeds. Likewise, GWASs for susceptibility to bovine tuberculosis (caused by Mycobacterium bovis) yielded inconsistent results (Bermingham et al., 2014; Finlay et al., 2012; Kassahun et al., 2015; Richardson et al., 2016).

Efforts to identify susceptibility alleles for other infectious diseases detected small-effect loci with poor replicability (reviewed in Raszek et al., 2016)). To date, the greatest success was obtained for mastitis, which is commonly caused by bacterial infections; regions spanning the DCK, SLCA4A4, and EDN3 genes were detected in at least two studies (Kanazawa et al., 1989; Sahana et al., 2013; Sodeland et al., 2011; Wu et al., 2015). The functional role of these genes in mastitis and response to invading bacteria remains to be evaluated.

As is the case for cattle, infections represent major economic problems for the swine industry, with one viral pathogen, porcine reproductive and respiratory syndrome virus (PRRSV) accounting for substantial burden (Dekkers et al., 2017). Several GWASs for PRRSV susceptibility converged to identify a major locus on chromosome 4 (reviewed in Dekkers et al., 2017)). Fine mapping of the QTL region and a further functional genomics work identified a variant in the GBP5 gene as likely causal (Dekkers et al., 2017). GBP5 encodes an interferon-inducible guanosine triphosphatase which plays central roles in cell-intrinsic immunity. In humans, GBP5 acts as a restriction factor for HIV (Krapff et al., 2016). For the purpose of swine breeding, this result is relevant as a variation at GBP5 can be used for marker-assisted selection (Dekkers et al., 2017).

8. Conclusions and perspectives

As detailed above, GWASs for infectious diseases have provided important biological insight and, in some cases, the results have opened the way to personalized therapy (Booth et al., 2012). Huge gaps however remain, with a number of infections still not addressed by GWAS or similar approaches. For instance, hundreds of millions of people are infected by helminths (Hotez et al., 2008), but no GWAS has investigated the genetic susceptibility to these parasites even though linkage studies provided evidence for several QTls (reviewed in Mangano and Modiano, 2014)). Diarrhoea and meningitis also account for a heavy health burden, especially among children in developing countries (World Health Organization, http://apps.who.int/gho/data/node/home). Nonetheless, only two studies provided genome-wide scans for genetic associations to these conditions (Bustamante et al., 2016; Davila et al., 2010).

The general trend in the field of complex traits is to shift from GWAS to next-generation sequencing approaches (exome or genome sequencing). These latter provide several advantages over GWASs, which mainly derive from the possibility to identify rare or private risk variants. However, next-generation approaches require sample sizes even larger than GWASs (Auer et al., 2016; Zik et al., 2014) and the amount of phenotypic variance that is explained by rare penetrant variants may largely differ among traits. We consider that important knowledge on infectious disease susceptibility (and related traits) can be still gained by GWASs, especially for those conditions that have never been investigated using this approach.

A promising strategy for future GWASs will be to leverage information from the host’s and pathogen’s genomes. An interesting step forward in this direction came from a study that generated human genome-wide genotyping data of antiretroviral naive patients and almost complete HIV-1 genome sequences to systematically search for associations between host and virus variants (genome-to-genome scan) (Bartha et al., 2013). The authors found strong associations between SNPs that tag HLA class I alleles and viral mutations in CTL (cytotoxic T lymphocyte) epitopes. These results clearly highlight the selective pressure imposed by the host immune system on the viral genome. No signals were detected outside the HLA class I loci and, on the viral genome, most selected sites were located in Gag and Nef. However, the majority of host-associated HIV-1 mutations were found to have no or little effect on viral load, suggesting that the virus can compensate for selective pressure with little fitness cost (Bartha et al., 2013). More recently, Ansari et al. (2017) performed a genome-to-genome scans in patients chronically infected with HCV. As in the HIV-1 study, the authors found that the adaptive immune system exerts a selective pressure on the viral genome and drives the evolution of several positions across the HCV genome. Importantly, Ansari and coworkers also showed that the host genotype for a functional IFNL4 variant modulates viral load only when the individuals are infected by a virus that carries a specific amino acid residue (serine) at position 2414 in the NSSA protein. This result clearly indicates that the host and viral genomes interact to determine the control of infection (Ansari et al., 2017). The majority of patients recruited in the study were infected with genotype 3 and HCV is genetically heterogeneous, suggesting that other and possibly different interactions exist in patients infected by distinct HCV genotypes.

The functional polymorphism in IFNL4 is a dinucleotide variant (rs469415590, TT/AG) and the effect of the “favorable” TT allele is recessive (Prokunina-Olsson et al., 2013). This finding highlights the possible relevance of non-additive models in genetic associations. In
particular, overdominant models may be particularly worth exploring in the field of infectious diseases. Remarkable examples of heterozygote advantage include those described for HBB and G6PD in relation to malaria. Heterozygosity at HLA genes has also been shown to protect against different infections (reviewed in (Quintana-Murci, 2016)), and the same is true for a common variant in TIRAP, with heterozygotes protected from malaria, invasive pneumococcal disease, bacteremia, and TB (Khor et al., 2007). Heterozygosity for a common polymorphism in PRNP also confers relative resistance to prion diseases (Mead et al., 2003). Overdominance causes multiple alleles to be maintained at a locus via balancing selection (reviewed in (Quintana-Murci, 2016)). A well-known example of this phenomenon is described above for the FUT2 gene. The maintenance of multiple ABO histo-blood groups, which most likely resulted from the selective pressure exerted by different pathogens, is also due to balancing selection (reviewed in (Quintana-Murci, 2016)). Several works have indicated that targets of balancing selection in primate genomes have often evolved in response to pathogen-driven selective pressures (Leffler et al., 2013; Ferrer-Admettia et al., 2008; Azevedo et al., 2015; Fumagalli and Sironi, 2014), making overdominance an appealing model to test for resistance against infection. It should however be noted that the action of balancing selection does not necessarily imply that heterozygotes are protected against a specific disease. Epitomal in this respect is the case of FUT2 described above, as well as that of ABO: individuals with O histo-blood group are protected against severe malaria and cerebral malaria, but at higher risk of developing highly symptomatic chola infection (Malaria Genomic Epidemiology Network and Malaria Genomic Epidemiology Network, 2014; Cooling, 2015). Interestingly, large-scale studies that focused on specific variants involved in resistance to malaria (Malaria Genomic Epidemiology Network and Malaria Genomic Epidemiology Network, 2014; Clarke et al., 2017) indicated that G6PD deficiency also results in different susceptibility phenotypes. Thus, the level of G6PD activity is associated with decreased risk of cerebral malaria, but with increased risk of severe malarial anaemia (Malaria Genomic Epidemiology Network and Malaria Genomic Epidemiology Network, 2014; Clarke et al., 2017). Moreover, by using a Bayesian statistical framework that allows for heterogeneity of effects across populations and phenotypes, the authors showed that homozygotes for the derived allele of a variant in the 5' upstream region of CD40LG have significantly reduced risk of severe malaria in The Gambia, but significantly increased risk in Kenya (Malaria Genomic Epidemiology Network and Malaria Genomic Epidemiology Network, 2014). Finally, a significant epistatic interaction was noted between the Hbc variant and an SNP in ATP2B4, this latter encoding a major erythrocute calcium channel (Malaria Genomic Epidemiology Network and Malaria Genomic Epidemiology Network, 2014). Overall, these results indicate that models of genetic resistance/susceptibility are often complex and that statistical methods that incorporate heterogeneity across populations and phenotypes may improve the power and reliability of GWASs.

For infectious diseases, a particular role may be played by interactions within HLA genes or among HLA variants and polymorphisms located outside the major histocompatibility complex. One example of this latter scenario relates to polymorphic variants in HLA-C and in the MIR148A gene (for microRNA-148a, mir-148a). Kulkarni et al. (2011) identified a polymorphism in the 3'UTR of HLA-C which affects a miR-148a binding site and associates with HLA-C expression levels, as well as with HIV-1 control. More recently, the same authors showed that an insertion/deletion polymorphism flanking MIR148A modulates the expression of this microRNA and the level of HIV-1 control only in individuals carrying HLA-C alleles with an intact miR-148a binding site. The MIR148A variant has no effect on HIV-1 control among subjects who carry HLA-C alleles that do not bind mir-148a (Kulkarni et al., 2013). Situations similar to the one described by Kulkarni and coworkers may be common and may be particularly important for HLA class I and KIR loci, as they biologically interact as binding partners. For instance, epistatic interactions among specific KIR and HLA loci have been shown to modulate the progression to AIDS (Hancock et al., 2008; Gaudieri et al., 2005; Martin et al., 2007), as well as the spontaneous (Thons et al., 2017; Khakoo et al., 2004) and therapy-induced clearance of HCV infection (Ahlenstiel et al., 2008). Likewise, interactions within the MHC may be common and relevant in modulating infection-related phenotypes. Indeed, it was recently shown that non-additive and interaction effects within HLA loci are widespread and modulate the risk of different autoimmune diseases (Lenz et al., 2015).

Taking these possibilities into account would certainly benefit the discovery of novel genetic effects modulating infection susceptibility or progression.

Finally, we note that, as recently highlighted, association studies have been dramatically skewed in terms of population inclusion: a 2016 survey of GWASs in the Catalog indicated that 81% of analyzed samples were of European ancestry and 14% of Asian ancestry (Popejoy and Fullerton, 2016). All other populations remained severely underrepresented. This trend is somehow attenuated in the studies we reviewed herein, as populations of Asian origin were included in > 33% of studies (most of them on HBV, HCV, and mycobacterial infections), whereas 14% and 17% of GWASs analyzed at least one cohort of African or Hispanic/Latino ancestry, respectively. However, only two studies, both of them on prion diseases, included Pacific Islanders (from Papua New Guinea) and no GWAS recruited Australian aborigines or other native peoples. These percentages clearly highlight the need to extend GWAS analysis for infectious diseases to under-represented populations. This is required to ensure that the benefits of research are equally distributed, especially in light of the recent evidence of limited portability of GWAS results across populations (Martin et al., 2017).

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2017.09.028.

References

Ahlenstiel, G., Martin, M.P., Gao, X., Carrington, M., Rehermann, B., 2008. Distinct KIR/HLA compound genotypes affect the kinetics of human antiviral natural killer cell responses. J. Clin. Invest. 118, 1017–1026.

Akkarathamroongsin, S., Thong, V.D., Payungpoon, S., Poovorawan, K., Prapanuwattana, P., Poovorawan, Y., Tangjkivhangnich, P., 2014. IFNλ3 (288b) and IFNλ4 polymorphisms are associated with treatment response in Thai patients infected with HCV genotype 1, but not with genotypes 3 and 6. J. Med. Virol. 86, 1482–1490.

Albalad, H., Khan, R., Grahn, R.A., Gandolfi, B., Mullikin, J.C., Cole, S.A., Graftydd-Jones, T.J., Haggstrom, J., Lohi, H., Longieri, M., Lyons, L.A., 2013. Extent of linkage disequilibrium in the domestic cat, Felis silvestris catus, and its breeds. PLoS One 8, e52527.

Alpay, F., Zare, Y., Kamalludin, M.H., Huang, X., Shi, X., Shook, G.E., Collins, M.T., Kirkpatrick, B.W., 2014. Genome-wide association study of susceptibility to infection by Mycobacterium avium subspecies paratuberculosis in Holstein cattle. PLoS One 9, e10794.

Ansari, M.A., Pedergnana, V., L Ci P, Cagi, M., Deli, A., Bonsall, D., Chatrurvedi, N., Bartha, I., Smith, D., Nicholson, G., McVean, G., Trebes, A., Piazza, P., Fellay, J., Cooke, G., Foster, G.R., STOP-HCV Consortium, Hudson, E., McLachlan, J., Simmonds, P., Bowden, R., Klerman, P., Barnes, E., Spencer, C.C.A., 2017. Genome-to-genome analysis highlights the effect of the human innate and adaptive immune systems on the hepatitis C virus. Nat. Genet.

Aslanian, M.K., Fitzgerald, K.A., 2014. Long non-coding RNAs and control of gene expression in the immune system. Trends Mol. Med. 20, 623–631.

Auer, P.L., Reiner, A.P., Wang, G., Kang, H.M., Abecasis, G.R., Altshuler, D., Bamshad, M.J., Nickerson, D.A., Tracy, R.P., Rich, S.S., NHHBI GO Exome Sequencing Project, Leaf, S.M., 2016. Guidelines for large-scale sequence-based complex trait association studies: lessons learned from the NHHBI exome sequencing project. Am. J. Hum. Genet. 99, 791–801.

Aylor, D.L., Valdar, W., Foulds-Mathes, W., Bus, R.J., Verdugo, R.A., Baric, R.S., Ferris, M.T., Frelinger, J.A., Heise, M., Friedman, M.B., Gralinski, L.E., Bell, T.A., Didion, J.D., Hua, K., Nehrenberg, D.L., Powell, C.L., Steigerwalt, J., Xie, Y., Kelada, S.N., Collins, F.S., Yang, L.V., Schwartz, D.A., Branstetter, L.S., Cheder, E.J., Miller, D.R., Spees, J., Liu, E.Y., McMillan, L., Sarkar, A., Wang, J., Wang, W., Zhang, Q., Broman, K.W., Korstanje, R., Darrant, C., Mott, R., Rissi, I.A., Pomp, D., THREADGILL, D., de Villena, P.F., Churchill, G.A., 2013. Genetic analysis of complex traits in the emerging Collaborative Cross. Genome Res. 21, 1213–1222.

Azevedo, L., Serrano, C., Amorim, A., Cooper, D.N., 2015. Trans-species polymorphism in humans and the great apes is generally maintained by balancing selection that modulates the host immune response. Hum. Genomics 9 (21:015:0043:1).

Badke, Y.M., Bates, R.O., Ernst, C.W., Schwab, C., Steibel, J.P., 2017. Estimation of linkage disequilibrium in four US pig breeds. BMC Genomics 13 (24-2164-13-24).

Baker, M.G., Bernard, I.T., Kvakbib, A., Verrall, A., Zhang, J., Keall, M., Wilson, N., Hall, T., Howden-Chapman, P., 2012. Increasing incidence of serious infectious diseases and inequalities in New Zealand: a national epidemiological study. Lancet 379, 1112–1119.
A. Mozee et al.

Infection, Genetics and Evolution 66 (2018) 286–307

American admixture in the Mexican Americans of San Antonio, Texas. Genet. Epidemiol. 3. 435–454.

Gupta, P., Chakrabarty, R., Chen, S., Su, W.H., Wang, Y.C., Weng, C.C., Yu, C.J., Hsu, C.L., Hsieh, A.B., Chien, R.N., Chu, C.M., Tai, D.L. 2014. A genome-wide association study on chronic HBV infection and its clinical progression in male Han-Taiwanese. PLoS One 9, e97924.

Guan, S., Mushiroda, T., Mahasirimongkol, S., Kiertiburanakul, S., Sunagkunprasert, S., Manosuthi, W., Tantisittirat, W., Charoensiriyawattana, A., Sura, T., Takahashi, A., Kubo, M., Kamatan, N., Chantarat, W., Nakamura, Y., 2011. Genome-wide association study of resistance to nevirapine-induced rash. Clin. Infect. Dis. 53, 341–484.

Chen, D., McKay, J.D., Cliftord, G., Gaborieau, V., Chabrier, A., Waterbe, T., Zaridze, D., Lisowski, P., Fabbanza, E., Benvenu, J., Vanoy, J., Foretova, L., Matos, N., Stenzena-Damians, R., Marzo, P., Wider, C., Wunsch-Filho, V., Eluf-Neto, J., Fernandez Garrute, L., Matus, E., Zelenika, D., Boland, A., Boffetta, P., Pavlina, M., Lathrop, M., Brownan, P., 2011. Genome-wide association study of resistance to nevirapine-induced rash. Clin. Infect. Dis. 53, 341–484.

Chen, D., Joo-Pecirep, I., Hammar, J., Jansso, E., Sborn, G., Liel, F., Magnusson, P.K., McKay, J.D., Wiliander, E., Gyllensten, U., 2013. Genome-wide association study of susceptibility loci for cervical cancer. J. Natl. Cancer Inst. 105, 624–635.

Chen, D., Enroth, S., Hui, S., yang, W., Yu, M., Muen, L., Xu, Y., Sung, S., Yellen, G. 2016. Pooled analysis of genome-wide association studies of cervical intraepithelial neoplasia. J. Natl. Cancer Inst. 108, 947–957.

Chen, M.G., Ritchie, M.D., Verma, S.S., Armstrong, G., Stallings, S., Denny, J.C., Carroll, K.M., McCarty, C.A., Peissig, P.L., Doheny, K.F., Pugh, E., Kho, A., Pacheco, J., Hayes, M.H., 2011. Genome-wide association study identifies novel loci associated with resistance to nevirapine-induced rash. Genes, Chromosomes & Cancer 50, 87–99.

Chen, M.G., Ritchie, M.D., Verma, S.S., Armstrong, G., Stallings, S., Denny, J.C., Carroll, K.M., McCarty, C.A., Peissig, P.L., Doheny, K.F., Pugh, E., Kho, A., Pacheco, J., Hayes, M.H., 2011. Genome-wide association study identifies novel loci associated with resistance to nevirapine-induced rash. Genes, Chromosomes & Cancer 50, 87–99.

Chen, M.G., Ritchie, M.D., Verma, S.S., Armstrong, G., Stallings, S., Denny, J.C., Carroll, K.M., McCarty, C.A., Peissig, P.L., Doheny, K.F., Pugh, E., Kho, A., Pacheco, J., Hayes, M.H., 2011. Genome-wide association study identifies novel loci associated with resistance to nevirapine-induced rash. Genes, Chromosomes & Cancer 50, 87–99.

Chen, M.G., Ritchie, M.D., Verma, S.S., Armstrong, G., Stallings, S., Denny, J.C., Carroll, K.M., McCarty, C.A., Peissig, P.L., Doheny, K.F., Pugh, E., Kho, A., Pacheco, J., Hayes, M.H., 2011. Genome-wide association study identifies novel loci associated with resistance to nevirapine-induced rash. Genes, Chromosomes & Cancer 50, 87–99.

Chen, M.G., Ritchie, M.D., Verma, S.S., Armstrong, G., Stallings, S., Denny, J.C., Carroll, K.M., McCarty, C.A., Peissig, P.L., Doheny, K.F., Pugh, E., Kho, A., Pacheco, J., Hayes, M.H., 2011. Genome-wide association study identifies novel loci associated with resistance to nevirapine-induced rash. Genes, Chromosomes & Cancer 50, 87–99.
A. Mozzi et al.

Infection, Genetics and Evolution 66 (2018) 286–307

Stebbing, E., Liu, J.Z., Kopanitis, L., Ignatyeva, O., Balabanov, Y., Nikolayevsky, V., Baessmann, I., Thye, T., Meyer, C.G., Nurnberg, P., Horstmann, R.D., Delveski, B., F. Plagens, J. Barrett, J.C., 2015. Susceptibility to tuberculosis is associated with variants in the ASAP1 gene encoding a regulator of dendritic cell maturation. Nat. Genet. 47, 523–527.

Custer, B., Sullivan, S.D., Haak, T.K., Ioeje, U., Veening, D.L., Kowdley, K.V., 2014. The hepatitis C virus genome: what has DAA therapy taught us so far? J. Hepatol. 60, 89–98.

Davila, S., Wright, V.J., Khor, C.S., Sim, K.S., Bender, A., Breunis, W.B., Inwald, D., Nadal, S., Betts, H., Carrol, E.D., de Groot, R., Harrns, W.P., Hazleot, J., Eoments, M., Lim, C.C., Contri, S., Castagna, T.H., Castagna, M., Smith, V.W., Levin, K.M., M.L., International Meningococcal Genomics Consortium, 2010. Genome-wide asso- ciation study identifies variants in the CFH region associated with host susceptibility to meningocebal disease. Nat. Genet. 42, 772–776.

Deekos, C.J., Bowland, D.R., Luedke, J.K., Plante, G., 2017. Host genetics of response to porcine reproductive and respiratory syndrome in nursery pigs. Vet. Microbiol. 208, 11–20.

DeLorenze, G.N., Nelson, CL., Scott, W.K., Allen, A.S., Ray, G.T., Taqi, A.L., Queensberry Jr., P.M., Powder Jr, W.G., 2016. Polymorphisms in HLA class II genes are associated with susceptibility to Staphylococcus aureus infection in a white population. J Infect Dis 213, 816–823.

Develpoor, K., DeMartino, S., Zhang, S.S., Lui, D.T., Zhao, C., Tam, P.O., Chan, W.M., Lam, D.S., Smythe, M., Barnstable, C., Pang, C.P., Hob, J., 2006. HTRA1 promoter polymorphism in weight age-related macular degeneration. Science 314, 998–999.

Dold, C., Holland, C.V., 2011. Investigating the underlying mechanism of resistance to Ascaris infection. Microbes Infect. 13, 624–628.

Duggal, P., Thio, C.L., Wojcik, G.L., Duggal, P., Thio, C.L., Wojcik, G.L., Mangia, A., Latanich, R., Kim, A.Y., Flint, J., Eskin, E., 2012. Genome-wide association studies in mice. Nat. Rev. Genet. 13, 749–762.

Dewan, A., Liu, M., Hartman, S., Zhang, S.S., Liu, D.T., Zhao, C., Tam, P.O., Chan, W.M., Ferris, M.T., Aylor, D.L., Bottomly, D., Whitmore, A.C., Aicher, L.D., Bell, T.A., Bradel, F., Ferrer-Admetlla, A., Bosch, E., Sikora, M., Marques-Bonet, T., Ramirez-Soriano, A., Muller, D.R., Rosenweig, E., Valdar, W., Wang, J., Churchill, G.A., Threadgill, D.W., McWeeny, S.K., Kate, M.G., Pardo-Manuel de Villena, F., Baric, R.S., Heise, M.T., 2013. Modeling host genetic regulation of influenza pathogenesis in the collaborative cross. PLoS Pathog. 9, e1003196.

Finlay, E.K., Berry, D.P., Wickham, B., Gormley, E.P., Bradley, D.G., 2012. A genome wide association scan of bovine tuberculosis susceptibility in Holstein-Friesian dairy cattle. PLoS One 7, e40879.

Flint, J., Eskin, E., 2012. Genome-wide association studies in mice. Nat. Rev. Genet. 13, 807–817.

Floroz, J.C., Price, A.L., Campbell, D., Ribas, L., Parra, M.V., Yu, F., Dueque, C., Saxena, R., Gallego, N., Tello-Ruiz, M., Franco, L., Rodriguez-Torres, M., Villegas, A., Bedoya, G., Aguilar-Salinas, C.A., Tusie-Luna, M.T., Ruiz-Linares, A., Reich, D., 2009. Strong contributions of genetic factors to genetic correlations across 60 000 nucleotides in human: implications for admixture studies of type 2 diabetes. Diabetologia 52, 1528–1536.

Fowke, K.R., Nagelkerke, N.J., Kimani, J., Simonsen, J.N., Anzala, A.O., Iwau, J.J., MacDonald, K.S., Ngugi, E.N., Plummer, F.A., 1996. Resistance to HIV-1 infection among persistently seronegative prostitutes in Nairobi, Kenya. Lancet 347, 1348–1351.

Franco-Paredes, C., Jones, D., Rodriguez-Moraes, A.J., Santos-Preciado, J.I., 2007. Commentary: improving the health of neglected populations in Latin America. BMC Public Health 7, 111.

Franke, A., McGovern, D.P., Barrett, J.C., Wang, K., Radford-Smith, G.L., Ahmad, T., Lees, C.W., Balsamo, J., Lee, S., Roberts, A., Anderson, C.A., Bis, J.C., Bumpstead, S., Ellingham, D., Fenton, E., G. et al., 2009. Replication of genome-wide association studies for seronegativity in HIV-1 infection among persistently seronegative prostitutes in Nairobi, Kenya. Lancet 347, 1348–1351.
Khor, C.C., Chapman, S.J., Vannberg, F.O., Dunne, A., Murphy, C., Ling, E.Y., Frodsham, A., Krapp, C., Hotter, D., Gawanbacht, A., McLaren, P.J., Kluge, S.F., Sturzel, C.M., Mack, K., Kennedy, R.B., Ovsyannikova, I.G., Haralambieva, I.H., Lambert, N.D., Pankratz, V.S., Koda, Y., Tachida, H., Pang, H., Liu, Y., Soejima, M., Ghaderi, A.A., Takenaka, O., Kimura, Kassahun, Y., Mattiangeli, V., Hubert, C., Jeffreys, A.E., Rowlands, K., Rockett, K.A., Spencer, C.C.A., Kwickisch, D.P., Malaria Genomic Epidemiology Network, 2017. Resistance to malaria through variation at the chr1421 locus. Science 356.http://dx.doi.org/10.1126/science.aam6393. (Epub 2017 124).-756.

A. Mozzi et al.

Infection, Evolution and Genetics 66 (2018) 286–307

sequence to human endogenous retroviruses HERV-L and HERV-16. Immunochemicals 49, 404–412.

Kemeny, C.J., Cato, N., Urabe, Y., Takahashi, A., Muruyama, R., Hosono, N., Otsuka, M., Tateishi, R., Omata, M., Nakagawa, H., Kose, K., Kaminami, K., Kubo, M., Nakamura, Y., Matsuda, K., Genome-wide association study identifies a susceptibility locus for HCV-induced hepatocellular carcinoma. Nat. Genet. 43, 455–458.

Kor, Y., Cheng, S.C., Smeds, M.J., Micropolymerasease-Boroulleis, E., Karjalainen, J., Franke, L., Wintho, S., Plantinga, T.S., van de Veerdonk, F.L., van der Meir, J.W.M., Joosten, L.A.B., Bochud, P.Y., Marchetti, O., Perel, J., Xaveri, D.J., Wijmenga, C., 2014. Immunosupp SNP array identifies novel genetic variants conferring susceptibility to candidiasis. Nat. Comm. 5, 4675.

Kuppermann, N., Fletcher, G.R., Daff, M.K., 1998. Predictors of occult pneumococcal bacteremia in young febrile children. Ann. Emerg. Med. 31, 679–684.

Larsen, T., Fiehn, N.E., 2017. Dental biofilm infections - an update. AMPS 125, 376–384.

Leffler, E.M., Gao, Z., Pfeifer, S., Segrel, L., Auton, A., Venn, O., Bowdton, R., Wall, D., Sell, G., Malaria susceptibility gene 1 (MS1) and 2 (MS2) follow the signature of ancient balancing selection shared between humans and chimpanzees. Science 339, 1578–1582.

Leffler, E.M., Band, G., Budy, G.B.J., Kivinen, K., Le, Q.S., Clarke, G.M., Bojany, K.A., Cono, J.-T., Moseley, P., England, P., 2015. Admixture mapping of tuberculosis susceptibility in an African-European hybrid cattle population. Front. Genet. 6, 210.

Kelly, R.J., Rosquier, S., Giorgi, D., Lennon, G.G., Lowe, J.B., 1995. Sequence and expression of a candidate gene for the human Secretor blood group alpha(1,2)fucosyltransferase gene (FUT2). Homozygosity for an enzyme-inactivating nonsense mutation commonly correlates with the non-secretor phenotype. J. Biol. Chem. 270, 4640–4649.

Kennedy, R.B., Ovsyannikova, I.G., Pankratz, V.S., Poland, G.A., 2014. Genome-wide analysis of polymorphisms associated with cytokine responses in smallpox vaccine recipients. Hum. Genet. 131, 1403–1421.

Kennedy, R.B., Ovsyannikova, I.G., Vierkant, R.A., Kelly, R.J., Rouquier, S., Giorgi, D., Lennon, G.G., Lowe, J.B., 2011. Association of polymorphisms in the HLA and HLA class II genes with tuberculosis and chronic hepatitis B in Asians. Nat. Genet. 41, 591–596.

Kamatani, Y., Wattanapokayakit, S., Ochi, H., Kawaguchi, T., Takahashi, A., Hosono, N., M. Suzuki, T., Kaminami, K., Nakamura, H., Puseman, A., Sura, T., Daigo, Y., Chibata, I., Kato, M., Kaminami, K., Nakamura, K., 2014. A genome-wide association study identifies variants in the HLA-DR locus associated with chronic hepatitis B in Asians. Nat. Genet. 41, 591–596.

Kikuchi, J.K., Dawkins, R.L., 1999. The P58 cytokine family gene in the MHC is related to human endogenous retroviruses HERV-L and HERV-16. Immunochemicals 49, 404–412.

Kemeny, C.J., Cato, N., Urabe, Y., Takahashi, A., Muruyama, R., Hosono, N., Otsuka, M., Tateishi, R., Omata, M., Nakagawa, H., Kose, K., Kaminami, K., Kubo, M., Nakamura, Y., Matsuda, K., Genome-wide association study identifies a susceptibility locus for HCV-induced hepatocellular carcinoma. Nat. Genet. 43, 455–458.

Kor, Y., Cheng, S.C., Smeds, M.J., Micropolymerasease-Boroulleis, E., Karjalainen, J., Franke, L., Wintho, S., Plantinga, T.S., van de Veerdonk, F.L., van der Meir, J.W.M., Joosten, L.A.B., Bochud, P.Y., Marchetti, O., Perel, J., Xaveri, D.J., Wijmenga, C., 2014. Immunosupp SNP array identifies novel genetic variants conferring susceptibility to candidiasis. Nat. Comm. 5, 4675.

Kuppermann, N., Fletcher, G.R., Daff, M.K., 1998. Predictors of occult pneumococcal bacteremia in young febrile children. Ann. Emerg. Med. 31, 679–684.

Leffler, E.M., Gao, Z., Pfeifer, S., Segrel, L., Auton, A., Venn, O., Bowdton, R., Wall, D., Sell, G., Malaria susceptibility gene 1 (MS1) and 2 (MS2) follow the signature of ancient balancing selection shared between humans and chimpanzees. Science 339, 1578–1582.

Leffler, E.M., Band, G., Budy, G.B.J., Kivinen, K., Le, Q.S., Clarke, G.M., Bojany, K.A., Cono, J.-T., Moseley, P., England, P., 2015. Admixture mapping of tuberculosis susceptibility in an African-European hybrid cattle population. Front. Genet. 6, 210.

Kelly, R.J., Rosquier, S., Giorgi, D., Lennon, G.G., Lowe, J.B., 1995. Sequence and expression of a candidate gene for the human Secretor blood group alpha(1,2)fucosyltransferase gene (FUT2). Homozygosity for an enzyme-inactivating nonsense mutation commonly correlates with the non-secretor phenotype. J. Biol. Chem. 270, 4640–4649.

Kennedy, R.B., Ovsyannikova, I.G., Pankratz, V.S., Poland, G.A., 2014. Genome-wide analysis of polymorphisms associated with cytokine responses in smallpox vaccine recipients. Hum. Genet. 131, 1403–1421.

Kennedy, R.B., Ovsyannikova, I.G., Vierkant, R.A., Kelly, R.J., Rouquier, S., Giorgi, D., Lennon, G.G., Lowe, J.B., 2011. Association of polymorphisms in the HLA and HLA class II genes with tuberculosis and chronic hepatitis B in Asians. Nat. Genet. 41, 591–596.

Kamatani, Y., Wattanapokayakit, S., Ochi, H., Kawaguchi, T., Takahashi, A., Hosono, N., M. Suzuki, T., Kaminami, K., Nakamura, H., Puseman, A., Sura, T., Daigo, Y., Chibata, I., Kato, M., Kaminami, K., Nakamura, K., 2014. A genome-wide association study identifies variants in the HLA-DR locus associated with chronic hepatitis B in Asians. Nat. Genet. 41, 591–596.

Kikuchi, J.K., Dawkins, R.L., 1999. The P58 cytokine family gene in the MHC is related to human endogenous retroviruses HERV-L and HERV-16. Immunochemicals 49, 404–412.

Kemeny, C.J., Cato, N., Urabe, Y., Takahashi, A., Muruyama, R., Hosono, N., Otsuka, M., Tateishi, R., Omata, M., Nakagawa, H., Kose, K., Kaminami, K., Kubo, M., Nakamura, Y., Matsuda, K., Genome-wide association study identifies a susceptibility locus for HCV-induced hepatocellular carcinoma. Nat. Genet. 43, 455–458.

Kor, Y., Cheng, S.C., Smeds, M.J., Micropolymerasease-Boroulleis, E., Karjalainen, J., Franke, L., Wintho, S., Plantinga, T.S., van de Veerdonk, F.L., van der Meir, J.W.M., Joosten, L.A.B., Bochud, P.Y., Marchetti, O., Perel, J., Xaveri, D.J., Wijmenga, C., 2014. Immunosupp SNP array identifies novel genetic variants conferring susceptibility to candidiasis. Nat. Comm. 5, 4675.
loci involved in antibiotic response to Mycobacterium avium spp. paratuberculosis in cattle. PLoS One 5, e11117.

Morera, A.L., Verma, S., Ornellas, S., Verma, S.S., Johnson, D.H., Dar, E.S., Gulick, R.M., Rahaib, R., Robbins, G.K., Ritchie, M.D., Haas, D.W. 2015. Phenome-wide association study relating pretreatment laboratory parameters with human genetic variants in AIDS clinical trials group protocols. Open Forum. Infect. Dis. 2, ofu0313.

Morera, A.L., Verma, S.S., Johnson, D.H., Dar, E.S., Gulick, R.M., Rahaib, R., Robbins, G.K., Ritchie, M.D., Haas, D.W. 2015. Phenome-wide association study relating pretreatment laboratory parameters with human genetic variants in AIDS clinical trials group protocols. Open Forum. Infect. Dis. 2, ofu0313.

Morera, A.L., Verma, S., Ornellas, S., Verma, S.S., Johnson, D.H., Dar, E.S., Gulick, R.M., Rahaib, R., Robbins, G.K., Ritchie, M.D., Haas, D.W. 2015. Phenome-wide association study relating pretreatment laboratory parameters with human genetic variants in AIDS clinical trials group protocols. Open Forum. Infect. Dis. 2, ofu0313.

Morera, A.L., Verma, S., Ornellas, S., Verma, S.S., Johnson, D.H., Dar, E.S., Gulick, R.M., Rahaib, R., Robbins, G.K., Ritchie, M.D., Haas, D.W. 2015. Phenome-wide association study relating pretreatment laboratory parameters with human genetic variants in AIDS clinical trials group protocols. Open Forum. Infect. Dis. 2, ofu0313.

Morera, A.L., Verma, S., Ornellas, S., Verma, S.S., Johnson, D.H., Dar, E.S., Gulick, R.M., Rahaib, R., Robbins, G.K., Ritchie, M.D., Haas, D.W. 2015. Phenome-wide association study relating pretreatment laboratory parameters with human genetic variants in AIDS clinical trials group protocols. Open Forum. Infect. Dis. 2, ofu0313.
Zhang, F., Liu, H., Chen, S., Low, H., Sun, L., Cui, Y., Chu, T., Li, Y., Fu, X., Yu, Y., Yu, G., Shi, B., Tian, H., Liu, D., Yu, X., Li, J., Lu, N., Bao, F., Yuan, C., Liu, J., Liu, H., Zhang, L., Sun, Y., Chen, M., Yang, Q., Yang, H., Yang, R., Zhang, L., Wang, Q., Liu, H., Zuo, F., Zhang, H., Khor, C.C., Hibberd, M.L., Yang, S., Liu, J., Zhang, X., 2011. Identification of two new loci at IL23R and RAB32 that influence susceptibility to leprosy. Nat. Genet. 43, 1247–1251.

Zignego, A.L., Wojcik, G.L., Cacoub, P., Visentini, M., Casato, M., Mangia, A., Latanich, R., Charles, E.D., Gragnani, L., Terrier, B., Piazzola, V., Dustin, L.B., Khakoo, S.I., Busch, M.P., Lauer, G.M., Kim, A.Y., Alric, L., Thomas, D.L., Duggal, P., 2014. Genome-wide association study of hepatitis C virus- and cryoglobulin-related vasculitis. Genes Immun. 15, 500–505.

Zuk, O., Schaffner, S.F., Samocha, K., Do, R., Hechter, E., Kathiresan, S., Daly, M.J., Neale, B.M., Sunyaev, S.R., Lander, E.S., 2014. Searching for missing heritability: designing rare variant association studies. Proc. Natl. Acad. Sci. U. S. A. 111, E455–64.