Seminars in Immunology xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Seminars in Immunology

journal homepage: www.elsevier.com/locate/ysmim

Human genetics of infectious diseases: Unique insights into immunological redundancy

Jean-Laurent Casanova[a,b,c,d,e,*], Laurent Abel[a,c,d,*]

[a] St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, New York, NY, USA
[b] Howard Hughes Medical Institute, New York, NY, USA
[c] Laboratory of Human Genetics of Infectious Diseases, Necker Branch, Inserm U1163, Necker Hospital for Sick Children, Paris, France, EU
[d] Paris Descartes University, Imagine Institute, Paris, France, EU
[e] Pediatric Hematology and Immunology Unit, Necker Hospital for Sick Children, Paris, France, EU

ARTICLE INFO

Keywords:
Redundancy
Knockout
Loss of function
Primary infection
Inborn error of immunity
Natural selection

ABSTRACT

For almost any given human-tropic virus, bacterium, fungus, or parasite, the clinical outcome of primary infection is enormously variable, ranging from asymptomatic to lethal infection. This variability has long been thought to be largely determined by the germline genetics of the human host, and this is increasingly being demonstrated to be the case. The number and diversity of known inborn errors of immunity is continually increasing, and we focus here on autosomal and X-linked recessive traits underlying complete deficiencies of the encoded protein. Schematically, four types of infectious phenotype have been observed in individuals with such deficiencies, each providing information about the redundancy of the corresponding human gene, in terms of host defense in natural conditions. The lack of a protein can confer vulnerability to a broad range of microbes in most, if not all patients, through the disruption of a key immunological component. In such cases, the gene concerned is of *low redundancy*. However, the lack of a protein may also confer vulnerability to a narrow range of microbes, sometimes a single pathogen, and not necessarily in all patients. In such cases, the gene concerned is *highly redundant*. Conversely, the deficiency may be apparently neutral, conferring no detectable predisposition to infection in any individual. In such cases, the gene concerned is *completely redundant*. Finally, the lack of a protein may, paradoxically, be advantageous to the host, conferring resistance to one or more infections. In such cases, the gene is considered to display *beneficial redundancy*. These findings reflect the current state of evolution of humans and microbes, and should not be considered predictive of redundancy, or of a lack of redundancy, in the distant future. Nevertheless, these observations are of potential interest to present-day biologists testing immunological hypotheses experimentally and physicians managing patients with immunological or infectious conditions.

1. Introduction

For almost all human-tropic viruses, bacteria, fungi, and parasites, the clinical outcome of primary infection varies enormously, from asymptomatic to lethal infection. At one extreme, the most common infections with weakly pathogenic microbes can kill rare patients with the most severe forms of immunodeficiency. At the other extreme, rare and highly pathogenic microbes can be harmless in rare individuals with constitutive resistance [1]. In the vast majority of cases, the clinical outcome of infection is unexplained. This “infection enigma” is arguably the most important problem in the fields of immunology and microbiology [2]. There is also great interindividual variability in the course of secondary infection or reactivation from latency, but this should be treated as a separate problem, due to the strong influence of adaptive, somatic, immunological memory of the previous, primary infection [3,4]. Acquired immunity (i.e. somatic memory) may also play a key role in governing interindividual variability in the course of primary infection. For example, vaccination with an attenuated pathogen or infection with a related pathogen can prevent diseases caused by deadly pathogens, mimicking immunity to secondary or latent infection. Moreover, an acquired immunodeficiency of adaptive immunity can precipitate various severe primary infections. For example, therapeutic immunosuppression or HIV infection renders patients highly vulnerable to both common and “opportunistic” primary infections, the latter being defined as infections occurring preferentially or exclusively in patients with overt immunodeficiency. Host genetic

* Corresponding authors at: St. Giles Laboratory of Human Genetics of Infectious Diseases, Rockefeller Branch, The Rockefeller University, 1230 York Avenue, New York, NY, USA.
E-mail addresses: casanova@rockefeller.edu (J.-L. Casanova), laurent.abel@inserm.fr (L. Abel).

https://doi.org/10.1016/j.smim.2017.12.008
Received 15 November 2017; Accepted 13 December 2017

make-up is an additional mechanism contributing to interindividual variability in the response to primary infection [5]. Obviously, host germline genome can influence both the quality of adaptive immunity and the severity of acquired immunodeficiency, both of which play important roles in primary infection. However, the strongest effect of germline genetic variation is predicted to be that on innate and intrinsic immunity to infection [6]. In broader terms, inborn errors of host defense can affect innate or adaptive hematopoietic cells, and the immunity controlled by non-hematopoietic cells, through the secretion of molecules involved in host defense or cell-intrinsic immunity. These inborn errors can be immunologically overt or covert, and may underlie diverse common and opportunistic infections following primary contact with a microbe.

The notion that some infections have a genetic basis dates back to the start of the 20th century. We previously reviewed the history and philosophy of this field [[10–13],5,7]. Briefly, plant, mammalian, and human geneticists proposed, at the start of the 20th century, that infectious diseases had a strong host genetic component. They documented this notion with elegant classical genetics studies. For example, a fungal infection of wheat was shown to segregate as a Mendelian trait as early as 1905. From the 1930s onward, different strains of laboratory animals were shown to have different outcomes when challenged with various infectious agents. Likewise, twin studies have shown that concordance for some infectious diseases is much greater for monozygotic than for dizygotic twins. Indeed, the idea that infectious diseases had a strong genetic component can be dated back to Pasteur's formulation of the germ theory itself. Pasteur proved that microbes were involved in disease development in his classic studies of two diseases of silkworm, *pébrine* and the *flacherie*. He convincingly made the case that *flacherie* also has an inherited element, in the sense that predisposition to the disease is transmitted, rather than the microbe being transmitted from the parents to the offspring [8,9]. Despite the publication of Mendel's discoveries in 1866, there was no real genetic theory in the 1865–1870 period, so Pasteur did not literally attribute *flacherie* to genetic variation. Nevertheless, the idea of a predisposition transmitted from the parents to the offspring was both proposed and documented. The first primary immunodeficiencies (PIDs) were described in the 1950s, at the beginning of the modern era of molecular genetics, with the description of molecules, cells, and mechanisms [10–13]. PIDs were then defined as Mendelian traits combining a rare immunological phenotype with a rare infectious phenotype. It was possible to detect such rare and artificial phenotypes thanks to the advent of antibiotics and the development of new immunological techniques facilitating the recognition of rare children with multiple, recurrent infections (who would previously have died during their first life-threatening infection) and of immunological defects (which would previously have been missed). A further major advance in elucidation of the emerging genetic architecture of infectious diseases was made in 1996, when life-threatening infectious diseases striking otherwise healthy individuals in the course of primary infection were shown to result from single-gene inborn errors of immunity that rarely display full penetrance [6,7]. Monogenic forms of resistance to infection were also characterized genetically during this period.

Attempts have been made to use the standard vocabulary of immunology to draw immunological conclusions from the observation of infectious phenotypes for categories of monogenic disorders [14,15]. However, this enterprise is somewhat risky. Indeed, host defense genes do not easily fall into immunology textbook dichotomies, such as innate vs. adaptive, hematopoietic vs. non-hematopoietic, or cell-intrinsic vs. cell-extrinsic immunity. A first problem with the innate/adaptive dichotomy is that it ignores key components of host defense, at least when the term "innate" is used to refer to leukocytes other than T and B cells. Indeed, cells other than leukocytes are involved in immunity. Each and every cell type of the body, not just tissue macrophages, which may be formed before hematopoiesis, contributes to host defense, by secreting molecules (e.g. hepatocytes secreting complement components) or through cell-intrinsic immunity (e.g. keratinocytes controlling papillomaviruses). More importantly, few, if any genes are expressed exclusively in any given cell type, rendering such binary classifications derived from cellular immunology fundamentally irrelevant to genetics, particularly for the innate/adaptive dichotomy. Only a handful of genes can be attributed purely to adaptive immunity (i.e. exclusively to T and B cells): those encoding the BCR and TCR chains. Likewise, very few host defense genes are not expressed at all in any T or B cells, not only because some of these cells, such as NKT and MAIT cells, which have a low level of TCR diversity, have "innate" features, but also more generally because T and B cells belong to the lymphoid lineage, which includes cells that do not express TCRs and BCRs, known as innate lymphoid cells (ILCs). The pattern of expression of human genes thus indicates that the vast majority of genes are expressed, to at least detectable levels, in at least some types of T and B cells and some other cells. For example, the most basic housekeeping genes can contribute to host defense, as illustrated by the immunological impact of inborn errors of metabolism (e.g. ADA and PNP deficiency). Last, but not least, adaptive immunity cannot be attributed solely to T and B cells, because the core players of the system, the α/β T cells, recognize antigens presented by HLA molecules on other cells, with $CD8^+$ T cells recognizing HLA-I molecules on nucleated cells and $CD4^+$ T cells recognizing HLA-II molecules on antigen-presenting cells (including B cells, which in early vertebrates also have phagocytic properties [16]). Deficiencies of HLA-I and HLA-II are therefore immunologically more "adaptive" than disorders of genes expressed both in T/B and other cells, even though $CD8^+$ and $CD4^+$ T cells are only indirectly afflicted by the lack of expression of HLA molecules. Moreover, the HLA molecules themselves are also not purely adaptive, because they play a key role in target recognition by NK cells.

By contrast, leaving these traditional notions of cellular immunology aside, human carriers of knockout mutations provide a unique opportunity to delineate the function of the corresponding genes in host defense more precisely. The number and diversity of inborn errors of immunity are increasing, and we will focus here on recessive traits causing a complete defect. We will not consider dominant disorders, partial defects, and gain-of-function mutations. With a mutation rate of 1.2×10^{-8} per nucleotide per generation, and seven billion people worldwide exposed to diverse microbes, the human population provides a unique resource for investigations of the impact of gene knockouts on host defense [17,18]. In particular, the study of humans with complete deficiencies of a particular gene can provide information about the function of the gene concerned in host defenses in natural conditions (*in natura*), *i.e.* its ecologically relevant and evolutionarily selected functions [18–20]. Since the identification of the first PID gene in 1985 [21], hemizygous and biallelic null mutations of various genes have been found to underlie multiple infections. Since the mid-1990s, complete deficiencies of other gene products have been found to underlie infections with a narrow range of microbes [7,22]. Also since the mid-1990s, biallelic null mutations of a few genes have been shown to confer protection against specific infections [7,22]. The last two decades have thus been marked by the discovery that monogenic lesions confer the protection or predisposition of otherwise healthy individuals to specific infections. With the more recent advent of large-scale next-generation sequencing (NGS), a fourth category of host defense genes was discovered, as several knockouts were found in humans with no overt infectious phenotype [23]. Some of these studies were conducted in specific contexts, such as bottle-necked [24,25] or consanguineous [26,27] populations, but the availability of large public databases, such as ExAC [28], and its extended version, gnomAD (http://gnomad.broadinstitute.org), has made it possible to search for common knockouts in the general population. These common knockouts can be defined as recessive defects caused by loss-of-function (LOF) variants with a minor allele frequency (MAF) $> 1\%$. Some of these knockouts may be associated with hitherto unknown resistances or predispositions to infection. Not all human genes involved in host defense fall into these