Fibrosis: ultimate and proximate causes

Victor J. Thannickal, Yong Zhou, Amit Gaggar, and Steven R. Duncan

Fibrotic disorders account for an increasing burden of disease-associated morbidity and mortality worldwide. Although numerous risk factors have been recognized, the etiologies of many of these clinical syndromes have not been identified, and they are often termed idiopathic or cryptogenic. Here, we provide an evolutionary perspective on fibrosis aimed at elucidating its etiopathogenesis. By asking the ultimate question of “why” this process evolved in multicellular organisms, we hope to uncover proximate explanations for “how” it causes disease in humans. We posit that physiological fibrosis-like reactions evolved as an essential process in host defense against pathogens and in normal wound healing. Based on this premise, we reason that pathological fibrosis is related to one or more of the following: unidentified infectious or noninfectious antigens, autoimmunity, impaired regenerative responses, and the antagonistically pleiotropic action of genes involved in wound healing or development. The importance of genetic susceptibility, epigenetics, aging, and the modern-day environment are highlighted. Consideration of both ultimate and proximate causation goes beyond philosophical cogitations, as it will better inform pathobiological mechanisms of disease and aid in the prevention and treatment of fibrotic diseases.

Imagine not being able to distinguish the real cause from that without which the cause would not be able to act as a cause. It is what the majority appear to do, like people groping in the dark; they call it a cause, thus giving it a name that does not belong to it.

—Plato, Phaedo 99

Introduction

Fibrotic disorders represent an increasing cause of morbidity and mortality worldwide. Since fibrosis is a predominant feature of the pathology of a wide range of diseases across multiple organ systems, fibrotic disorders have been estimated to contribute to an estimated 45% of all-cause mortality in the United States (1). Despite this tremendous clinical impact and advances in our understanding of the pathogenesis of fibrotic processes, effective therapies are limited, and the vast majority of clinical trials have failed. Here, we address the problem of fibrosis from an evolutionary perspective with the hope that this will better inform therapeutic approaches to these complex and intractable disorders. In essence, asking the ultimate teleological question of “why” this process evolved in multicellular organisms may provide us with a more proximate explanation as to “how” it causes or contributes to disease in humans. Understanding both ultimate and proximate causation goes beyond philosophical ideas or theories, as it will better inform pathobiological mechanisms of disease and aid in the prevention and treatment of fibrotic diseases.

“Why”: fibrosis as an evolutionarily conserved adaptive process

A consistent finding among all the fibrotic disorders in humans is the uniform loss of tissue architecture and progressive loss of organ function. It is not immediately obvious why fibrosis might provide any selective advantage, either survival or reproductive, to the organism. However, the tissue morphologic features that can broadly be defined as fibrosis — accumulation of extracellular matrix (ECM) and loss of cellular homeostasis — are seen in normal physiological responses. Importantly, the cellular and molecular mechanisms that participate in physiological responses are also much the same as those identified in pathological contexts. Two evolutionarily conserved adaptive or protective functions for physiological fibrosis are (a) host defense against pathogens and (b) wound healing in response to epithelial injury.

Host defense. In many ways, fibrosis may be viewed as a component of the innate and adaptive immune response to certain pathogens. Indeed, many of the cells and mediators that constitute classical host responses to infections are seen in fibrotic disorders (2). Thus, a more expanded view of innate and adaptive immunity incorporates the concept that formation of a fibrotic scar may serve a beneficial role to the host by preventing pathogen invasion and/or spread, in addition to the pathogen killing by classical cells of the innate immune response, such as macrophages and neutrophils. Viewed in this way, cells such as fibroblasts and myofibroblasts, which are essential to fibrosis, participate in host defense by containing pathogens in order to more effectively kill them, or as a fail-safe mechanism to prevent their spread when killing fails to completely eliminate them. There are several examples of a fibrotic scar that forms a barrier to prevent pathogen spread: the most common are the thick fibrotic walls surrounding bacterial abscesses and the fibrotic rim around granulomas that contain mycobacterial pathogens. In these examples, an excessive burden of microbes (as with aspiration pneumonia in an alcoholic) or the virulence of a particular microbe (as with Mycobacterium tuberculosis) may not allow killing and eradication of the invading pathogen, and survival of the host organism is dependent on containment of the pathogen by preventing its spread or dissemination.

Conflict of interest: The authors have declared that no conflict of interest exists.

Wound healing. The ability to survive wounding in the wild must have strongly influenced the survival of more complex multi-cellular metazoa. After cutaneous wounding, the formation of a fibrotic scar may benefit survival by several mechanisms, including prevention of excessive blood loss, barrier function against the entry of microbes, and formation of a provisional matrix to facilitate regeneration of damaged epithelium. Persistent fibrosis almost always accompanies incomplete or ineffective regeneration across multiple species (3). Interestingly, mammals appear to have lost the remarkable regenerative capacity of certain other species, such as amphibians (4). The reason for this is not clear; however, it is possible that such capacity was not lost, but rather favored in an apparent trade-off during mammalian evolution. Why would fibrosis be favored over regeneration? It could be argued that the bioenergetic investment in complete regeneration of a limb or an organ was too great for complex organisms; thus, limited resources were directed toward creating a fibrotic patchwork to preserve structural integrity that allowed the organism to live, despite loss of functional or reserve capacity. Another evolutionary trade-off may have been to limit the attendant risk of oncogenic transformation of a relatively small number of tissue-resident stem cells and/or progenitors that would be required to undergo a large number of cell divisions, an argument that is particularly relevant to larger, long-lived mammals that seem to have adopted a primarily stem cell-based mechanism for tissue regeneration. This contrasts with the remarkable regenerative capacity observed with limb amputation in amphibians, which rely on dedifferentiation of somatic cells at the wound margins to initiate regeneration without scarring or cancer (5).

“How”: emergence of pathological fibrosis

An evolutionary perspective of fibrosis may aid in understanding not only why, but also how fibrosis develops in pathological contexts. Following from the evolutionary reasoning provided above, a pathological or persistent fibrotic response may be related to (a) the persistence of a hitherto-unidentified pathogen or antigen; (b) the development of autoimmune reactions; (c) an impaired regenerative response that may relate to genetic, epigenetic, or age-related alterations; or (d) the antagonistic, pleiotropic action of selected genes involved in development and wound healing.

The first two relate to the host defense function of physiological fibrosis, and the last two to wound healing (Figure 1).

Role of infectious or noninfectious antigens. Numerous viral and parasitic infections have been identified as causal agents in human fibrotic disorders, in particular liver fibrosis (6); other organ systems such as the heart may also be involved, as exemplified by chronic infection with Trypanosoma cruzi, a well-known cause of cardiac fibrosis and cardiomyopathy (Chagas disease) (7). There appears to be an association of latent or other viral infections among subsets of patients with a progressive and fatal form of lung fibrosis known as idiopathic pulmonary fibrosis (IPF) (8). To date, our understanding of the link between infections and IPF is largely limited to studies showing increased prevalence of latent viruses in the lungs of patients with this disease compared with control populations, or among patients with acute (and typically severe) IPF exacerbations relative to those with stable pulmonary disease (9). Experimental murine models with γ-herpes viruses implicate a potential role for viral infections in the progression and/or exacerbation of lung fibrosis (10, 11). Viruses and other intracellular pathogens can cause apoptosis of alveolar epithelial cells, which is implicated in the pathogenesis of lung fibrosis (12). However, at present, it is unclear to what extent lung fibrosis that may be associated with viral or other infective agents is due to direct effects, rather than a secondary consequence of an injurious immune response provoked by these infections (13). Persistent viral or other infections that evoke immune responses may also predispose for the development of secondary autoimmunity (see below).

The cellular and molecular mechanisms that were hardwired to generate fibrotic responses to infectious pathogens may have been co-opted to respond to noninfectious stressors in our modern environment. Thus, environmental factors such as air pollution may be sufficient to trigger fibrotic responses in the genetically susceptible host. Indeed, noninfectious etiologies have clearly been identified in nonidiopathic causes of liver fibrosis (e.g., alcohol) and lung fibrosis (e.g., asbestosis). Put another way, it is possible that sufficient evolutionary time has not elapsed to allow the host to discriminate between infectious and noninfectious challenges of the present day.

Autoimmunity. Autoimmunity is defined by the presence of specific receptor-mediated lymphocyte responses to distinct, par-

Figure 1. Evolution of fibrosis. In most multicellular organisms, fibrosis-like reactions serve a useful purpose by limiting pathogen invasion and promoting wound healing. Pathological fibrosis may result from the persistence of extrinsic (unidentified) infections and/or antigens or of intrinsic neoantigens that evoke chronic autoimmune reactions. Impaired tissue regenerative capacity and/or the action of antagonistically pleiotropic genes may trigger perpetual wound healing responses that cause progressive fibrosis. These pathological processes may be instigated by genetic or epigenetic alterations in reparative and immune cells in the setting of modern environmental challenges and in an aging population.
ticular autologous peptides (14). Benign low-level autoimmune to self-proteins is probably near ubiquitous and may be critical for maintaining immunological homeostasis. In some cases, however, self-reactive immune responses become profoundly dysregulated, and injurious or even life-threatening to the host.

Autoimmune responses often develop as a secondary consequence of adaptive immune host defenses that were initially, and appropriately, directed against extrinsic antigens (e.g., microbes, viruses, and protein toxins or irritants) (14, 15). In certain conditions, particularly when the provocation is chronic and/or fulminating, and conditional on the presence of “permissive” genetic influences, these immune responses can lose their typically exquisite specificity for one or a small, finite number of foreign peptides, instead acquiring more generalized reactivity to other antigens, which may include self-peptides (e.g., epitope spread). In other cases, the extrinsic antigens share primary sequences or conformational homology with self-peptides that then become the collateral targets of misdirected, cross-reactive immune responses (e.g., epitope mimicry). The altered microbiome within damaged lungs (or chronic viral infections) can be highly antigenic, and thus conducive for the development of “secondary” autoimmunity. In addition, normally inert self-proteins, for which immune tolerance was established in early life, can later become immunogenic (e.g., neoantigens) if they are covalently modified by reactive chemicals present in tobacco smoke or by ROS or reactive nitrogen species produced by activated leukocytes.

The overwhelming preponderance of these secondary autoimmune responses appear to be clinically benign, but they are highly pathogenic in an important minority of cases, including carditis, nephritis, neurologic dysfunction associated with group A streptococcal and other microbial infections, paraneoplastic syndromes linked to malignancies, and myriad other tissue-specific autoimmune disorders (15–17). The cascade of injury processes evoked by autoimmune responses can include direct cytotoxicities by autoantibodies or autoreactive T cells; elaboration of numerous proinflammatory, vasoactive, and profibrotic mediators (including IL-4, IL-13, and TGF-β); and recruitment and/or activation of other immune effectors and mesenchymal cells (18–20). The end result of many clinically distinct autoimmune syndromes is tissue fibrosis affecting one or more organs (2).

The importance of recognizing autoimmune processes in the genesis of a fibrotic disorder centers on the potential that directed therapies to interrupt these particular immune processes may be much more effective and/or readily tolerated than global immunosuppressants. Autoimmune lung diseases can be notoriously refractory to treatment with glucocorticosteroids and other nonspecific immunosuppressants; however, modalities that reduce autoantibody production or physically remove autoantibodies may be more effective (19, 21–24).

Although an autoimmune basis is well appreciated in fibrosis associated with various connective tissue diseases, this has not been established as a pathogenic mechanism in the so-called idiopathic cases, including IPF (25). Interestingly, a recent study showed that high levels of circulating autoantibodies to heat shock protein 70 are associated with poor outcomes in IPF patients (26), and other reports describe clonally expanded and/or highly differentiated lymphocytes in these patients, as well as abnormally high levels of B cell trophic and chemotactic mediators that parallel findings in classic autoantibody diseases (e.g., rheumatoid arthritis, systemic lupus erythematosus, and Sjögren’s syndrome) (27–30). Nonetheless, the precise roles of autoantibodies and neoantigens in the development and/or progression of idiopathic fibrotic disorders, as well as in acute exacerbations of these diseases, deserve further study.

**Impaired regeneration.** Epithelial regeneration is a fundamental requirement for normal wound healing, and its impairment may lead to fibrosis. Genetic, epigenetic, and/or age-related alterations may confer susceptibility to fibrosis. Genetic studies in familial cases of IPF indicate a number of pathways related to protein folding and/or trafficking that may adversely affect the regenerative capacity of the epithelium (31). Epigenetic alterations that influence epithelial cell regeneration and fibroblast differentiation or fate may also predispose to fibrotic disease (32, 33).

There is growing recognition that aging is a risk factor for fibrosis. For example, the diagnosis of IPF is rarely made before 50 years of age, and its incidence and prevalence increases sharply with advancing age (34). The biological mechanisms that account for IPF as an age-related disease are not well understood. Several of the recently proposed hallmarks of aging have been linked to the pathogenesis of IPF (35, 36). The switch from regeneration to fibrosis as the primary repair response may represent slow, stochastic changes in cells and/or tissues driven by the accumulation of cellular or molecular damage, as suggested by Kirkwood’s disposable soman theory of aging (37). Alternatively, aging has been theorized to be programmed by genetic and/or metabolic factors, loss of proteostasis, and/or impaired stress responsiveness at more or less predictable ages that may be species specific (38, 39). In either case, it remains to be determined whether genes and/or pathways that have been shown to modulate lifespan in model organisms are tenable therapeutically in specific age-related human fibrotic diseases such as IPF. This is an area of investigation that is likely to reap enormous reward, not only with regard to fibrosis, but for a growing number of diseases in which aging is a known risk factor.

**Antagonistic pleiotropy.** The concept of antagonistic pleiotropy was proposed by Williams in 1957 to explain senescence and aging (40); however, this concept may even more aptly apply to a number of complex, age-related diseases (41, 42). If a particular gene mediates beneficial effects in early life but exerts detrimental effects after reproductive age, there will be evolutionary pressure to conserve that gene despite its potential disease-causing effects with aging. We have proposed the concept of antagonistic pleiotropy to explain the role of the ROS-generating enzyme NADPH oxidase 4 (NOX4), which mediates myofibroblast differentiation and normal wound healing in young subjects, but promotes persistent fibrosis with aging (43–45). One explanation for the pleiotropic nature of NOX4 action is the lack of a counter-regulatory response involving NFE2-related factor 2 (NRF2) in aged animals; this NOX4/NRF2 imbalance leads to persistent fibrosis associated with the acquisition of a senescent and apoptosis-resistant myofibroblast phenotype in aged mice, whereas young mice demonstrate the capacity for fibrosis reversibility (44).

The antagonistic pleiotropic action of genes may also apply to numerous developmental genes and/or pathways that are reac-
tivated in age-related fibrotic disease (46). Perhaps this is best illustrated by the canonical Wnt/β-catenin pathway. This developmental pathway, which is essential for organogenesis and normal stem cell function, is upregulated in lungs of patients with IPF and other fibrotic disorders (47–49); importantly, targeting the Wnt/β-catenin pathway protects against fibrosis in preclinical animal models (48–50). The mechanisms that account for the detrimental roles of developmental genes and/or pathways in age-related diseases may involve the permissive effect of specific aging phenotypes, including loss of proteostasis, metabolic reprogramming, immune dysregulation, and alterations in stem cell niches.

Implications for fibrosis research

Can an evolutionary perspective of fibrosis better inform our investigational efforts? Based on the preceding discussion of ultimate causation, perhaps studies of the etiology of idiopathic fibrosis should focus on unidentified infectious agents, or alterations in the tissue microbiome. Alternatively (or additionally), is the modern-day environment to blame? Environmental challenges in an aging individual may be sufficient to trigger autoimmunity, areas that have received little investigative attention. Identification of specific environmental triggers will have major public health value by aiding in the development of preventive, rather than curative, strategies.

Similarly, understanding of human regenerative potential that is likely not lost, but rather suppressed or latent, will unveil new therapeutic opportunities; in particular, understanding mechanisms of dedifferentiation of somatic cells at sites of tissue injury deserve further study. Whether loss of cellular plasticity accounts for age-related propensity to fibrosis is currently unknown. From a clinical standpoint, the health benefits of halting the progression of fibrosis may be almost as important as reverse remodeling to achieve normal organ structure and function, a daunting but possibly attainable proposition. Important ECM-related signaling and epigenetic mechanisms that contribute to the persistence and progression of fibrosis are being uncovered (51, 52). Finally, the identification of antagonistically pleiotropic genes, many of which are linked to embryonic development and/or normal wound healing, may be particularly attractive for therapeutic targeting in chronic diseases of aging, including organ fibrosis.

Acknowledgments

Work in the authors’ laboratories relating to the content and/or concepts proposed herein is supported by NIH grants PO1 HL114470 and R01 AG046210 (to V.J. Thannickal), R01 HL102371 (to A. Gaggar), and R01 HL119660 (to S.R. Duncan) and by AHA Grant-in-aid 14GRNT20180023 (to Y. Zhou).

Address correspondence to: Victor J. Thannickal, Division of Pulmonary, Allergy, and Critical Care, Department of Medicine, University of Alabama at Birmingham, 1530 3rd Avenue South, THT 422, Birmingham, Alabama 35294-0006, USA. Phone: 205.975.6376; E-mail: vjthann@uab.edu.