MUC5b + environmental insult = IPF?

Introduction

Mucus, Mucins, MUC5b: structure, function and evolutionary importance

Mucus is an essential part of the innate immune system, considered to be universal within most phyla of both aquatic and teresterial meazoans. It plays a piviotal role in the prevention of disease by serving as an antimicrobial barier, it also has physiological functions including allowing the exchange of oxygen, carbon dioxide, nutirent and metabolites, lubricating surfaces and reducing damage due to sheer, reducing deydration of the epithelia and providing the polymeric matrix which enables ciliary-mucus particle transport. Mucus barriers are essential for the separation and protection of an organism from its external environment, and likely a prerequisite for the exclusion of bacteria from bobily tissues and evolution of gastrointestinal and respiratory tracts. The importance of mucus barriers is further underlined when one considers the energy investment continuous mucus production and release requires; for example, corals use mucus to trap particiles and transport them towards their mouths and the reef-building coral Acropora acuminata is thought to dedicate up to 40% of its daily net carbon fixation to this task alone.[@Bakshani2018] Mucins are a key component of mucus, they are highly evolutionary conserved large glycoproteins that date back around 600 million years to Nematostella vectensis, the starlet sea anemone, which is an early marine invertebrate. The earliest human mucin analgoue is found in Xenopus tropicalis, the African clawed frog, which evolved about 300 million years ago and mucins are the likely explanation for the observation that frogs show such great resistance to infection during disection and it has been shown that knockdown of mucin in the skin mucus barrier of Xenopus tropicalis tadpoles leads to susceptibility to infection by the opportunistic pathogen Aeromonas hydrophila.[@Dubaissi2018]

The mucin family is composed of proteins that contain tandom repeat structures with a high proportion of prolines, threonines, and serines; the PTS domain. It is further defined by extensive glycosylation of the PTS domain through N-Acetylgalactosamine O-linkages at the threonine and serine residues.[@Kufe2009] The resultant oligisaccharide chains and polymeric structure create the viscoeleastic properties of mucus which confer its barrier properties and play an important role in storage and secreation. [@Bakshani2018] Mucins are 50-90% carbohydrate and they are anionic because most of their terminal sugars contain carboxyl or sulphate groups. Mucin glycan helps to sequester pathogen by acting as a 'decoy' and providing sites for microbal adhesins to bind; for example, human salivary MUC5b interacts with streptococcal species, and patterns of glycosolation change during inflammation.[@Linden2008][@Jaramillo2018] Mucin barriers can be subverted by pathogens, strategies include production of enzymes to

degrade mucin core proteins and mucin carbohydrates, and evolution of effective motility through mucus gels - many mucosal bacterial pathogens are flagellated for this reason. There is evidence that degradative enzymes are required for pathogeneisis in species such as Vibrio cholorae and that flagella are required for infectivity in species such as Helicobacter pylori.[@Linden2008] Intracellular gel-forming mucins are stored in a compact and condensed form in granules within mucus-serceting cells. They are stored in solution with a high concentration of calcium ions and protons which is thought to be necessary to mask the anionic charge and prevent electrostatic repulsion, upon secreation mucins expand 1000-3000 fold taking up water to form a gel as calcium is exchanged for sodium and the pH rises.[@Bakshani2018] One consequence of mucins being stored in such a compact form is that when they're released they can obstruct the airway which in mouse models appears necessary for the clearance of helminth infection[@Jaramillo2018] and may provide a clue to their evolution.

Normal human airway mucus is a hydrogel composed of approximately 98% water, 0.9% salt, 0.8% globular proteins, and 0.3% high-molecular-weight mucin polymers.[@Boucher2019] Mucin hypersecreation may increase the concentration of solids up to 15% resulting in viscous elastic mucus that is not easily cleared.[@Fahy2010] 17 genes encode mucins in the human genome of which the gene products of seven are secreted and the remainder are membrane bound. Five of the secreted mucins have terminal cysteine rich domains that can form disulfide bonds resulting in polymers that impart the properties of a gel. MUC5AC and MUC5B, two secreated gel-forming mucins, are strongly expressed in the human respartory tract. MUC5AC is predominantly expressed in the conducting airways and MUC5B is predominantly expressed in the respiratory airways (muc5b is also expressed in salivary glands, cervix, gallbladder, seminal fluid, and middle ear epithelium). Secreted mucins are large glycoproteins (up to 3x10ˆ6 D per monomer), ranking among the largest molecules encoded in mammalian genomes, and their expression induces and requires an endoplasmic reticulum stress response.[@Dickey2017] Mucin production and secreation are regulated by distinct mechanisms. Production is highly regulated at transriptional level. The ErbB family of proteins contains four receptor tyrosine kinases, structurally related to the epidermal growth factor receptor (EGFR), its first discovered member. ErbB-receptor signaling appears important for MUC5AC produciton since inhibition blocks MUC5AC up-regulation by diverse stimuli. Interleukin-13 (IL-13) is a cytokine secreted by T helper type 2 (Th2) cells, CD4 cells, Natural killer T cell, Mast cell, Basophil cells, Eosinophil cells and Nuocyte cells. IL-13 is a central regulator in IgE synthesis, goblet cell hyperplasia, mucus hypersecretion, airway hyperresponsiveness, fibrosis and chitinase up-regulation. It is a mediator of allergic inflammation and different diseases including asthma. IL-13 appears important because it increases MUC5AC expression (IL-1 beta appears to be an important stimulus for MUC5b expression[@Jaramillo2018]). Basal levels of production and secreation of MUC5AC and MUC5B change as part of an allergic response. The production of MUC5AC can increase 40-200 times as high as normal levels in humans with similar findings in mice, MUC5B

increases more modestly, 3 to 10 times in mice. The most important stimulus for secretion appears to be ATP which acts on apical membrane purinergenic ($P2Y_2$) receptors. Once secreted mucus gel is propelled in a proximal direction towards the mouth, by ciliary beating as part of the mucociliary escalator, where is expectorated or swallowed. [@Fahy2010]

MUC5b rs3570950 and respiratory disease

Expression and localisation of MUC5AC and MUC5B is different in patients with lung disease compared with health controls. MUC5AC expression is increased in asthma for example, while MUC5B expression is increased in COPD[@Kesimer2017] and IPF. In COPD MUC5b expression occurs in more proximal airways, whereas in IPF it localised to the bronchiole.[@Helling2017] MUC5b appears to be particularly important in IPF.

The gain of function promotor variant rs5270590, 3.5 kb upstream of the mucin 5b (MUC5B) transcriptional start site, is the strongest identified risk factor (genetic or otherwise) for the development of either sporadic or familial IPF. The largest study to date (1616 non-hispanic white patients with fibrotic interstitial pneumonias and 4683 controls) estimated that the odds of developing pulmonary fibrosis for those with one copy of the risk allele were 4.5 times (95% CI: 3.9, 5.2) the odds of those with no copies and that the odds for those with two copies are 20.2 times those with no copies (95% CI: 15.2–27.0).[@Fingerlin2013] The strength of association is substantially higher than for most other common risk variants for complex disease with the exception of the human leukocyte angigen (HLA) region for some autoimmune diseases such as type-1 diabetes mellitus and systemic lupus erythmaotsis which have OR greater than 10. The association between rs35705950 has been replicated in 3 genome wide association studies (GWAS) and a total of 10 independant cohorts including a mexican cohort and two asian cohorts and is thought to account for about a third of IPF cases.[@Evans2016] However, penetrance is low with up to 20% of non-hispanic whites having a least one copy of the variant yet IPF occuring only rarely. The rs35705950 variant is a G-to-T transversion that occurs in an area of the MUC5B 5' flanking region, a region which has characteristics of being an enhancer subject to epigentic control via dna methylation and histone modification.[@Helling2017] An enhancer is a sequence of DNA that functions to enhance transcription. A promoter is a sequence of DNA that initiates the process of transcription. A promoter has to be close to the gene that is being transcribed while an enhancer does not need to be close to the gene of interest. Publically available data through the Encyclopedia of DNA Elements (ENCODE) suggest MUC5b promotor site is a complex area of the genome with many transcriptional factors showing evidence of binding.[@Selman2006] In other words MUC5b expression likely a function of genetic and non-genetic factors.[@Evans2016] In addition to IPF, rs35705950 has been found to be positively associated with interstitial lung abnormalities (ILA), chronic hypersensitivity pneumonitis (CHP), rheumatoid arthritis asso-

ciated interstitial lung disease (RA-ILD), and myeloperoxidase-antineutrophil cytoplasmic antibody-associated vasculitis associated interstitial lung disease (AAV-ILD).[@Namba2019] It has also been found to not be associated with cutaenous systemic sclerois interstital lung disease (SSc-ILD), sarcoidosis, and myositis-ILD. [@Adegunsoye2019]

Potential role in IPF pathogenesis (and normal function inc make the point penetrance low need something else too e.g occ exposure and bring in recent review and coal dust) The rs5270590 variant is associated with a 34 fold increase in expression of MUC5b compared with wild type in healthy control populations and a 5 fold increase in patients with IPF (see figure 1).[@Evans2016] In IPF patients distal airway MUC5b is expressed preferentially, compared with MUC5Ac. MUC5b also expressed in honeycomb cysts, a defining characteristic of the usual interstitial pneumonia CT pattern typically seen in IPF.[@Seibold2013]

MUC5b expression (Evans 2016)

Proposed mechanisms for the role of the rs5270590 variant in the pathogenesis of IPF include:

1. excessive production of MUC5B by stem cells that attempt to regenerate injured bronchiolar and alvelar epithelium could disrupt normal development pathways and highjack normal repartive mechanisms of the distal lung resulting in fibroprolferation and honeycomb cyst formation.
2. excessive MUC5B production leads to reduced mucocilary function, retention of particles, and enhanced lung injury.
3. interaction between MUC5b and motile cilia since distinct cilium gene expression in IPF lung has been observed.
4. excessive MUC5b production inducing endoplasmic retiulum stress and the unfolded protein response.[@Evans2016]

Muc5b has been studied in mice. A Mub5b knockout mouse study found that muc5b is essential for mucociliary clearance, for controlling airway and middle ear infections, and maintaining immune homeostasis in the lungs. Knockout mice had airflow limitation and died from infection by multiple bacterial species, including Staphylococcus aureus.[@Roy2014] A trangenic muc5b mouse model of muc5b overexpression found that overexpression causes mucociliary dysfuction and ehnhances lung fibrosis on response to bleomycin.[@Hancock2018] Intriguingly, in recent bleomycin lung fibrosis model studies lung fibrosis was attenuated and mortality reduced in both germ-free mice and IL-17B deficient mice supporting the concept that fibrosis in response to epithelial injury is mediated by interaction of the immune system with microbiota.[@ODwyer2019][@Yang2019]

infection/immunity

The frequency of the disease associated allele at rs35705950 exceeds 10% in European populations (https://www.ncbi.nlm.nih.gov/snp/rs35705950) but is less than 1% in African and East Asian populations. Clearly the rs35705950 variant is not subject to negative selection due to IPF risk since onset is well after the reproductive age begins[@Evans2016]; the variation in frequency observed is consistent with strong positive selection. The increased MUC5b expression in the airways associated with the rs35705950 variant may have conferred a survival advantage by providing protection against lung infection. [@Dickey2017][@Jaramillo2018] A relation between the rs35705950 variant, disease risk, and infection is also supported by the observation that in a prospective study of 65 IPF patients have higher bacterial loads than COPD and healthy controls and within IPF patients those with homozygous (TT) for variant had significantly lower bacterial loads (P=0.01), measured by 16S rRNA quantitative polymerase chain reaction of bronchoalveolar lavage samples. Within IPF those with higher bacterial loads were also at increased risk of death.[@Molyneaux2014] These finding are consistent with observation that the rs35705950 variant is associated with improved survival in IPF[@Peljto2013] and fewer acute respiratory disease events in the COPDGene cohort with interstitial features.[@Ash2018] However, these studies are vulnerable to index event bias, by which selection of subjects according to disease status creates biased associations if common causes of incidence and prognosis if not properly accounted for.[@Dudbridge2019] For example, it is known that the rs35705950 variant is associated with interstital lung abnormalities[@Hunninghake2013], since the diagnosis of IPF relies heavily on radiological appearances individuals with the variant might tend to be diagnosed earlier in the course of their disease giving the false impression, when comparing them to IPF patients without the disease variant that is associated with survival. Further support for the importance of infection in IPF provided by the observation that immunomodulatory therapies such as interferon gamma, ethanercept, prednisolone, azathioprine and N-acetylcysteine have failed to prolong survival in IPF[@WarheitNiemi2019] to prolong survival in IPF, from a small (N = 181) double blinded randomized controlled study which found reduced symptom burden and improved survival associated with cotrimoxazole[@Shulgina2013], as well as evidence from genetic and animal studies. IPF GWAS have identified single nucleotide variants associated with disease suceptibility in the Toll interacting protein (TOLLIP) gene, for example rs111521887. TOLLIP is an inhibatory adapator protein within Toll-like receptors (TLR) and part of the innate immune system recognising pathogen associated molecular patterns (PAMPs)[@Noth2013] and, intriguingly, in a mouse bleomycin lung fibrosis model the abscence of a microbiome protected against mortality.[@ODwyer2019]

inorganic occupational stimuli

While the frequency of the disease associated allele at rs35705950 exceeds 10% in European populations(https://www.ncbi.nlm.nih.gov/snp/rs35705950), its penetrance is low. the median prevalence of IPF for men and women in Europe is approximately 3.75 per 100000 for the period 2001-2013[@Marshall2018], other genetic and/or enviornmental factors must be at play. In addition to responding to PAMPs as outlined above the innate immune system also responds to damage-associated molecular patterns (DAMPs) which can result from inhalation of inorganic respirable toxins such as silica or asbestos.[@Dostert2008] Secreation of the inflammatory cytokine IL-1beta (which is also a stimulus for MUC5b expression) is elevated in alveolar macrophages of patients with ILD, including IPF, sarcoidosis, silicosis, RA-ILD, and asbestosis.[@Byrne2016][@Howrylak2017] Inflammasome are multiprotein intracellular complexes that detect pathogenic microorganisms (PAMPs) and sterile stressors (DAMPs). The NLRP3 (NOD-, LRR- and pyrin domain-containing protein 3) inflammasome is an intracellular sensor that detects a broad range of PAMPs and DAMPs leading to caspase 1-dependent release of the pro-inflammatory cytokines IL-1 beta and IL-18, as well as to gasdermin D-mediated pyroptotic cell death.[@Swanson2019] Interestingly the NLRP3 inflammasome appears to be implicated, albeit with differing activation patterns[@Lasithiotaki2016], in all of these conditions, interaction between smoking (a risk factor for IPF) and the NLRP3 inflammasome is recognised, and recent work has shown age-depedent suceptibility to pulmonary fibrosis in a bleomycin-induced lung injury mouse model.[@StoutDelgado2016] Occupational risk factors such as metal, wood, and stone dust exposure are well recognised in IPF, accounting for up to 8% of cases the basis of a meta-analysis of case-control data[@Blanc2019] and its likely that innate immune system activation via the NLRP3 inflammasome and other means by occupational exposures mediates this risk.

Conclusion

The apparently complex interplay between exposure to organic and inorganic respiratory toxins, the mucus barrier, respiratory epithelium and resident cells such as alveolar macrophages in idiopathic pulmonary fibrosis remains incompletely characterised but genetic, epigenetic, gene-expression, and epidemiological studies are beginning to fill in the gaps. Gene-environment interaction between the rs5270590 variant and occupational exposures would be expected if ? (where do abestos/silica/metal fibres enter lung?)