Mendelian Diseases and Inflammatory Bowel Disease - data-mining for genetic risk and disease associated confounders

Holm H. Uhlig

Translational Gastroenterology Unit, Experimental Medicine, University of Oxford, John Radcliffe Hospital Oxford, OX3 9DU, UK. Phone: 0044 1865 8 57963
holm.uhlig@ndm.ox.ac.uk

Abstract
Healthcare databases are an excellent tool to study the epidemiology of Mendelian diseases. Analysis of healthcare utilization data by Han et al. identified robust associations between a large spectrum of Mendelian disorders and inflammatory bowel disease (IBD). In addition to likely causative associations between IBD and primary immunodeficiencies, multiple confounders need to be considered. Potential confounder are inflammation induced acquired disorders, *de novo* IBD caused by mycophenolate mofetil immunosuppression related to solid organ transplantation and *de novo* IBD caused by reconstructive surgery. The longitudinal analysis of healthcare utilization in patients with genetic diagnosis is required to resolve cause-effect-relationships, separate and identify confounders and investigate the causative gene networks that drive Mendelian disorder associated IBD.

Inflammatory Bowel Disease is increasingly recognized as a spectrum of phenotypic and genetically diverse disorders. Discovery in the field of IBD genetics has emerged from two different directions. On one side there are association studies in which tens of thousands of patients with IBD have been analyzed for common variants of typically low effects and low-frequency variants of intermediate effect in IBD risk [1, 2]. The other avenue is to investigate patients with Mendelian disorders that present with IBD and to study rare and ultra-rare genetic variants of high functional effect, even starting conceptually from single patients (n of one studies) [3, 4]. Since Mendelian diseases individually affect only a very small proportion of the population, it is by nature challenging to ascertain statistical associations and probe for causality. The availability of population based routine healthcare utilization data is a potential way to study the over 6800 different rare diseases and identify associations with phenotypes, pharmaco-utilization, response to procedures or health care costs [5-7]. In the United States and the European Union disorders with a prevalence of less than 64 or less than 50 per 100 000 people are considered as rare (but due to the large number of diseases overall 6-8% of the population are affected). Previously regarded as "orphans" of medicine those disorders are now the focus of intensive interest due to the recognized needs of the patients, the opportunities of improved genetic diagnostics, better understanding of the functional mechanisms and the potential to provide targeted personalized treatments.

Using healthcare databases that included 111 Mio individuals, Blair et al. investigated associations between Mendelian disorders with multiple polygenic disorders such as IBD, coeliac disease or rheumatoid arthritis [8]. This analysis indicated the potential of healthcare utilization data to inform the genetic architecture of inflammatory and autoimmune disorders by investigating joint genetic networks. In this issue of *Inflammatory Bowel Diseases* Han et al. replicate and adopt this strategy by investigating healthcare data of the Optum database that contains data of 55 Million patients with at least one International Classification of Diseases code including 183,855 Crohn's disease and 177,039 Ulcerative colitis patients [9]. 50 groups of Mendelian disorder-associated ICD9 diagnosis codes were significantly associated with Crohn's disease and Ulcerative colitis. Since the associations found by Blair et al. and Han et al. largely replicate with a very robust effect size, these data provide an excellent and thought-provoking basis for analysis of causative associations and confounder.

Several associations between Mendelian diseases and IBD identified by Han et al. are related to primary immune defects. About one third of the associated diagnosis codes is linked to at least one causative or likely-causative gene and effectively the majority of previously described Mendelian disease-associated genes are potentially associated [10]. This includes IBD associated Mendelian immunodeficiencies such as leucocyte defects, severe combined immunodeficiency, chronic granulomatous disease (*CYBB, CYBA, NCF1, NCF2, NCF4*), disorders of aromatic amino acid metabolism (the ICD9 code includes Hermansky Pudlak Syndrome caused by *HPS1, HPS4,* and *HPS6* defects into this group due to the albinism), Glycogenosis (Glycogenosis 1B caused by defects in *SLC37A4*), or hypogammaglobulinemia (including *BTK* defects). The epidemiological studies also identified Mendelian associations with IBD that have been postulated based on weak evidence (such as in case of neurofibromatosis [11] or Familial Mediterranean Fever [12]) and have not found associations with other Mendelian disorders (such as cystic fibrosis). A group of neurodegenerative diseases including cerebral degeneration due to generalized lipidoses, degenerative diseases of the basal ganglia or Huntington's disease are associated. This might hint towards defective autophagy in the neurodegenerative disorders such as previously shown in Niemann Pick type C that is associated with IBD [13]. However it is clear that even among the strongest associations many disorders lack a well-defined direct colitogenic pathomechanism. It is very likely that diverse confounders contribute to those associations. Those associations might be

driven by inflammation induced acquired pathology, colitogenic effects of solid organ transplant-associated immunosuppression and Mendelian disorder-associated reconstructive surgery. Far from being false positive associations or noise, some confounders themselves are experiments of nature that provide insight into the pathogenesis.

It is recognizable that the list of significant Mendelian associations includes several syndromes and metabolic disorders that are treated by liver or kidney transplantation. Disorders of copper metabolism (Wilson's disease), defects of urea cycle metabolism, forms of glycogenosis, congenital syndromal abnormalities in particular Alagille syndrome (caused by JAG1 and NOTCH2 mutations), or polycystic kidney disease (such as PKD1, PKD2, and PKHD1 defects) fall into these groups. This list is far from complete since additional groups such as the aforementioned aromatic amino acid metabolism group also contain metabolic disorders (such as several forms of tyrosinemia). To reverse metabolic toxicity or to exchange a terminally damaged liver or kidney, solid organ transplantations are performed in these disorders. Immunosuppressant medication such as mycophenolate mofetil is used to avoid transplant rejection. In particular mycophenolate mofetil causes *de novo* appearance of inflammatory bowel disease–like colitis. Among symptomatic solid organ transplantation patients who underwent colonoscopy, mycophenolate mofetil-associated colitis was identified in 9% of patients [14]. Consistent with endoscopic findings of erythema, erosions and ulcers, histologically acute colitis was identified in 50% and inflammatory bowel disease-like histology was found in 36% of patients who underwent colonoscopy [14]. The colitogenic mechanism of mycophenolate mofetil might involve epithelial toxicity as indicated by increased epithelial apoptosis [15].

Another rare but informative example for a Mendelian confounder association is related to surgery. Genetically female patients with congenital adrenal hyperplasia (ICD9 classified as adrenogenital syndrome) are born with vaginal defects due to defective 21-hydroxylase deficiency causing an excess of androgens. Reconstructive surgery allows the formation of a neovagina using an intestinal autologous implant [16]. In a substantial number of patients the neovagina becomes inflamed and the „spreading" of the inflammatory process can cause a *de novo* IBD in some patients [17]. It is likely that bacterial dysbiosis in the neovagina causes a form of diversion colitis. Similarly, in mice ectopic dysbiotic colonisation can drive intestinal inflammation [18]. The above examples highlight that specialized surgical procedures or medications that relate to the treatment of individual disorders (but are not genotype specific itself) can account for the association to IBD.

Additional associations highlight the role of chronic inflammation. For instance, it is interesting that long QT syndrome is highly associated with IBD. Prolonged QT intervals in the electrocardiogram have been noted in several studies [19]. Those results therefore support an emerging role of inflammation in acquired QT prolongation [20].

A challenge with analyzing health care utilization datasets is that many classification codes are not specific. It is important to remember that the ninth revision of the ICD code was essentially developed in 1975 by the World Health Organization. Although modified, this 40-year-old code is neither a gene centered nor a molecular pathway centered classification. There are no distinct ICD-9 codes for very relevant disease groups such as „IBD unclassified", an IBD presentation more common in infants and young children. The ICD-9 code is not designed for a gene level analysis and highly relevant Mendelian disorders such as IL10 signaling defects [4] cannot be specified at all. An updated classification approach is emerging with the ICD11 revision due in 2018 (https://icd.who.int/dev11), allowing a more up to date classification of many Mendelian disease groups. However, this beta-version still lacks important molecular disease subgroups such as IL10 signaling defects as part of infantile onset IBD. A molecular pathway analysis based on pure ICD code associations will therefore be heavily biased. Many patients encompassed in the selected diagnosis codes will not have a Mendelian diagnosis, there are a variable

number of genes that actually contribute to each of the ICD-9 codes and there are multiple confounder effects. This suggests that a longitudinal analysis based on genetic stratification will be essential to identify cause and effect relationships.

The power of genomics-healthcare databases approaches has been described recently by identification of biological traits based on electronic patient records and genomic data [21]. Linking diagnosis codes with longitudinal healthcare utilization, surgical procedures, prescriptions, textual data such as histology and genetic diagnosis (up to the exome or genome sequencing variant level) supported by multi-dimensional pattern analysis using machine learning approaches and artificial intelligence is challenging but feasible ones ethical considerations of data linkage and informed consent are carefully considered. Healthcare utilization database approaches combined with availability of genomic data and biorepositories will overcome several classical problems of underpowered studies due to small patient numbers, lack of control groups, center and referral bias that currently affect many current studies in rare disorders.

HHU is supported by the Crohn's & Colitis Foundation of America (CCFA), by the The Leona M. and Harry B. Helmsley Charitable Trust as well as the Oxford Biomedical Research Center. There is no conflict of interest. HHU declares project collaboration with Eli Lilly, Regeneron and UCB Pharma and consultancy/speaker for UCB Pharma, Pfizer, AbbVie.

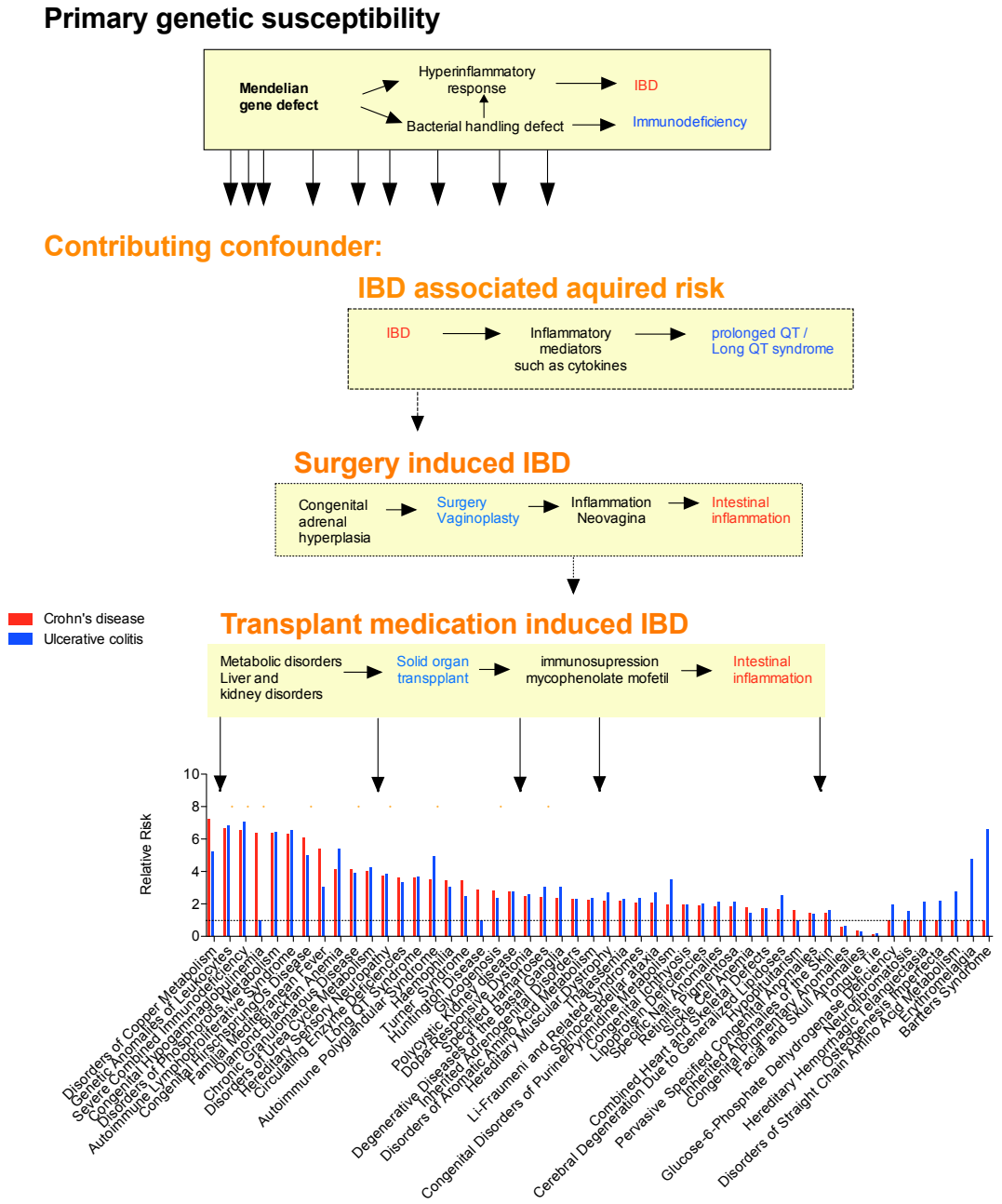

Figure: Mendelian disease associations and inflammatory bowel disease.
50 disease associations between Crohn's disease and ulcerative colitis identified by Han et al. [9]. Primary genetic susceptibility that drive those associations via immune dysregulation are indicated and potential confounder are shown.

1. Huang, H., et al., *Fine-mapping inflammatory bowel disease loci to single-variant resolution.* Nature, 2017. **547**(7662): p. 173-178.

2. Luo, Y., et al., *Exploring the genetic architecture of inflammatory bowel disease by whole-genome sequencing identifies association at ADCY7.* Nat Genet, 2017. **49**(2): p. 186-192.
3. Casanova, J.L., et al., *Guidelines for genetic studies in single patients: lessons from primary immunodeficiencies.* The Journal of experimental medicine, 2014. **211**(11): p. 2137-49.
4. Glocker, E.O., et al., *Inflammatory bowel disease and mutations affecting the interleukin-10 receptor.* The New England journal of medicine, 2009. **361**(21): p. 2033-45.
5. Schneeweiss, S. and J. Avorn, *A review of uses of health care utilization databases for epidemiologic research on therapeutics.* J Clin Epidemiol, 2005. **58**(4): p. 323-37.
6. Kesselheim, A.S. and J.J. Gagne, *Strategies for postmarketing surveillance of drugs for rare diseases.* Clin Pharmacol Ther, 2014. **95**(3): p. 265-8.
7. Casey, J.A., et al., *Using Electronic Health Records for Population Health Research: A Review of Methods and Applications.* Annu Rev Public Health, 2016. **37**: p. 61-81.
8. Blair, D.R., et al., *A nondegenerate code of deleterious variants in Mendelian loci contributes to complex disease risk.* Cell, 2013. **155**(1): p. 70-80.
9. Han, L.e.a., *Mendelian Disease Associations Reveal Novel Insights into Inflammatory Bowel Disease.* Inflammatory Bowel Diseases, 2017.
10. Uhlig, H.H., et al., *The diagnostic approach to monogenic very early onset inflammatory bowel disease.* Gastroenterology, 2014. **147**(5): p. 990-1007 e3.
11. Adams, W., et al., *Concurrent Ulcerative Colitis and Neurofibromatosis Type 1: The Question of a Common Pathway.* Pediatrics, 2016. **137**(2): p. e20150973.
12. Beser, O.F., et al., *Association of inflammatory bowel disease with familial Mediterranean fever in Turkish children.* J Pediatr Gastroenterol Nutr, 2013. **56**(5): p. 498-502.
13. Schwerd, T., et al., *Impaired antibacterial autophagy links granulomatous intestinal inflammation in Niemann-Pick disease type C1 and XIAP deficiency with NOD2 variants in Crohn's disease.* Gut, 2017. **66**(6): p. 1060-1073.
14. Calmet, F.H., et al., *Endoscopic and histological features of mycophenolate mofetil colitis in patients after solid organ transplantation.* Ann Gastroenterol, 2015. **28**(3): p. 366-373.
15. Liapis, G., et al., *Histological spectrum of mycophenolate mofetil-related colitis: association with apoptosis.* Histopathology, 2013. **63**(5): p. 649-58.
16. Amies Oelschlager, A.M., A. Kirby, and L. Breech, *Evaluation and management of vaginoplasty complications.* Curr Opin Obstet Gynecol, 2017. **29**(5): p. 316-321.
17. Lima, M., et al., *Vaginal replacement in the pediatric age group: a 34-year experience of intestinal vaginoplasty in children and young girls.* J Pediatr Surg, 2010. **45**(10): p. 2087-91.
18. al., A.e., *Ectopic colonization of oral bacteria in the intestine drives TH1 cell induction and inflammation.* Science, 2017.
19. Yorulmaz, E., et al., *Prolonged QT dispersion in inflammatory bowel disease.* World J Gastroenterol, 2013. **19**(1): p. 65-71.

20. Lazzerini, P.E., P.L. Capecchi, and F. Laghi-Pasini, *Long QT Syndrome: An Emerging Role for Inflammation and Immunity.* Front Cardiovasc Med, 2015. **2**: p. 26.
21. Dewey, F.E., et al., *Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study.* Science, 2016. **354**(6319).