

LETTER TO THE EDITOR

Open Access



What can we learn from common variants associated with unexpected phenotypes in rare genetic diseases?

Jeanette Erdmann*

Abstract

The purpose of this article is to stimulate discussion about whether a phenome-wide association study is a suitable tool for uncovering late-onset risks in patients with monogenic disorders that are not yet fully recognized because the life expectancy of people with such conditions has only recently extended, and they now reach older ages when they may develop additional complications.

Keywords: Collagen VI congenital muscular dystrophy, Col VI-CMD, PheWAS, Late-onset risk, Unexpected phenotypes

I am well aware that the following analysis has weaknesses and that the results should not be regarded as a definite statement about the late-onset risk for diverticular disease in Col VI-CMD.

My interest is based on having, after almost 45 years without knowing what is causing my slow but ongoing progressive neuromuscular condition, diagnosed myself as a carrier of a pathogenic variant in the *COL6A2* gene, leading to collagen VI congenital muscular dystrophy (Col VI-CMD), using next-generation sequencing and modern information technology [1].

Col VI-CMD is primarily caused by variants in three collagen VI genes, *COL6A1*, *COL6A2*, and *COL6A3* [2, 3], and much less frequently by variants in *COL12A1* [4].

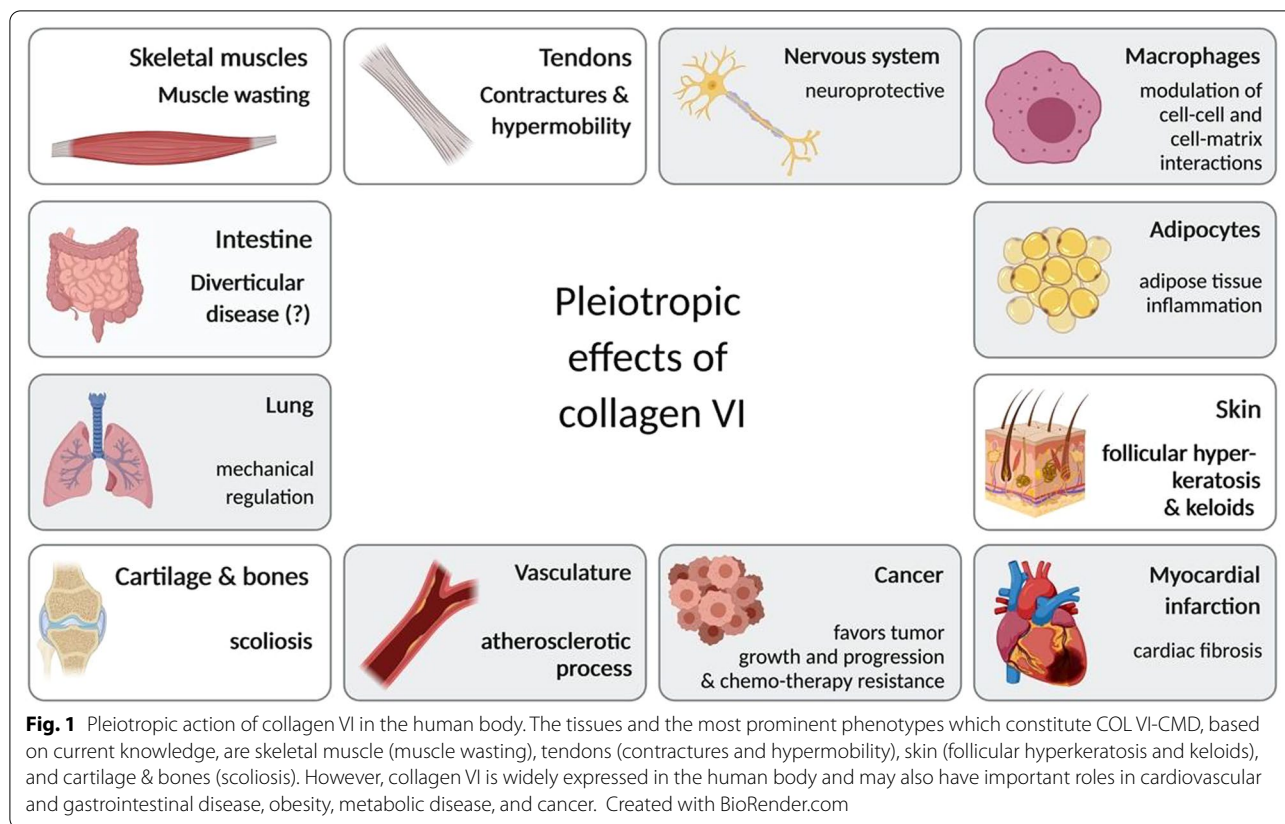
The focus of the clinical course of patients with Col VI-CMD is mostly on the primary pathological phenotype of (slow) progressive muscle weakness, contractures, and hyperflexibility, and respiratory impairment due to exhausted respiratory muscles [5]; however, since collagen VI functions as part of the extracellular matrix [5], it

has long been suspected that there are also late-onset disease risks, beyond progressive muscle weakness, such as a higher risk of aneurysms. Also, impairments of the cardiovascular system and intestinal tract are not excluded (Prof. Dr. med. Carsten Bönnemann, personal communication). The functions of collagen VI, so important in muscle disease, may also have implications for obesity, metabolic disease, and cancer in patients with Col VI-CMD (see [6, 7] for detailed reviews) (Fig. 1); however, this has yet to be systematically investigated, as there is currently no sufficiently comprehensive longitudinal registry for patients with this condition. Nevertheless, some unexpected phenotypes caused by rare genetic variants in *COL6A2* and *COL6A3* have been discovered in recent studies; for example, *COL6A2* defects in patients with myoclonus epilepsy [8] and *COL6A3* defects causing dystonia [9].

In general, patients with neuromuscular disorders have a significantly longer life expectancy today than they did a few decades ago, due to better care [10]. Hence, congenital neuromuscular diseases, such as Col VI-CMD or Duchenne muscular dystrophy [11], should now also be considered diseases of adulthood. Consequently, more public health interventions are needed

*Correspondence: jeanette.erdmann@uni-luebeck.de
Institute for Cardiogenetics, University of Lübeck, Ratzeburger Allee 160,
Building 67, 23562 Lübeck, Germany





to support such patients and their families as they pass from childhood into adult life. Hence, the early detection of late-onset disease risks, beyond the primary muscle disease, can be vital.

I am well aware of critical health issues that could be related to my condition; 10 years ago, I was severely ill, suffering from acute diverticulitis, a condition characterized by inflammation of one or more diverticula (bulges in the colon wall). In mild cases, diverticulitis can be cured with antibiotics, while in severe cases, surgery is the only therapeutic option. In my case, despite presenting with severe rectal bleeding, leading to fainting and repeated bouts of diverticulitis, my doctors decided not to consider surgery, rather treating me with high doses of antibiotics. This informed decision was made because of general caution regarding anesthesia in patients with neuromuscular disease, and my specific condition, which had required night-time non-invasive ventilation for almost 15 years, due to impaired lung function because of a severely exhausted diaphragm. Since we have decided against surgery, the problem of the diverticula is not really treated, but

has hovered over me, like the sword of Damocles, for the last decade, and will continue to do so for years to come.

In 2010, Denny and colleagues suggested the concept of phenome-wide association studies (PheWAS) by performing a “reverse genome wide association study (GWAS)”, thereby determining, for a given genotype, the range of associated clinical phenotypes [12]. This reverse genetic approach can provide novel insights not readily attainable by forward genetic strategies. PheWAS takes advantage of increasingly large sets of human genetic variation data, coupled with dense phenotypic information, to analyze genotype–phenotype associations [13]. In this way, it is possible to generate an almost complete picture of the pleiotropic effects of genetic variations and respective genes, where pleiotropy describes the phenomenon in which a gene influences two or more, seemingly unrelated, phenotypic traits [14]. Before PheWAS was conceptualized, pleiotropy was established through intensive phenotyping of relatively small disease cohorts and, most importantly, by functional studies in mice and human cell culture models. As just one example, genetic variants in

GJA1, which encodes connexin 43, cause oculodentodigital dysplasia (OMIM #164200), a rare condition characterized by a typical facial appearance and highly variable findings related to the eyes, teeth, and fingers [15].

Within the last decade, several large-scale biobanks have been established worldwide, often with genomic as well as comprehensive phenotypic data, with total enrollment in the largest biobanks surpassing 500,000 individuals [16]. A prime example of genotypic and phenotypic data made publicly available is the UK Biobank (UKBB). UKBB aims to improve the prevention, diagnosis, and treatment of a variety of serious and life-threatening diseases, including cancer, heart disease, stroke, diabetes, arthritis, osteoporosis, eye disease, depression, and dementia [17]. It tracks the health and well-being of 500,000 volunteers and provides health and genetic information to researchers from science and industry. This makes the UKBB the most comprehensive clinical and genetic data resource currently publicly available. Linking the PheWAS approach and UKBB data allows researchers to associate every single genetic variant with more than 3,000 phenotypes stored in the UKBB for each patient. UKBB data can be accessed through several platforms, including <http://pheweb.sph.umich.edu/>.

Along these lines, two interesting studies have been published very recently, both using PheWAS and data from large biobanks in the context of Mendelian diseases. First, Tcheandjieu and colleagues reported that the spectrum of associations of common and rare variants in genes involved in Mendelian diseases can be extended to individual phenotypes within the general population [18]. This study was based on four well-described syndromic diseases (Alagille, Marfan, DiGeorge, and Noonan syndromes) and PheWAS analysis of UKBB data, and show that specific phenotypes associated with these rare disease genes can also be identified in population-based data by PheWAS.

Even more interestingly, Park et al. [19] used a cohort of >11,000 unselected individuals from the Penn Medicine Biobank to identify associations of rare variants in the *LMNA* (Lamin A/C) gene with diverse phenotypes using a PheWAS approach. The authors demonstrated that pathogenic *LMNA* variants are an underdiagnosed cause of cardiomyopathy. Intriguingly, they also detected an unreported association between loss of function variants in *LMNA* and renal disease, a phenotype apparently unconnected with cardiomyopathy.

A very convenient way to access UKBB data, in addition to publicly available curated GWAS information, is at <https://atlas.ctglab.nl/PheWAS> [20]. This website hosts

a comprehensive database of publicly available GWAS summary statistics and results from GWAS of 600 traits from UK Biobank release 2. Here, users are able to both access original summary statistics and obtain a variety of results from pre-performed analyses, such as risk loci information, LD regression score [21], MAGMA [22], and multi GWAS comparisons [20].

Leveraging this rich data resource, I performed an exploratory gene-based PheWAS for *COL6A2*, with the aim of identifying potential late-onset risks in patients with Col VI-CMD. My hypothesis is that the association of common genetic variants in *COL6A2* with phenotypes deposited in publicly available GWAS datasets may reveal late-onset disease risks, which could inform future disease management. The results of the PheWAS for *COL6A2* over a broad range of phenotypes are presented in Fig. 2a, b.

The most significant finding is an association between the *COL6A2* gene and waist-hip ratio ($p=5.0e-09$) [23]. Interestingly, the second most significant genome wide hit was with diverticular disease ($p=2.4e-8$) [24] (Fig. 2a). Moreover, the association between rs12626197 and diverticular disease could be replicated using data from the FinnGen study (data freeze 3, spring 2019), consisting of 135,638 individuals (accessed November 2020 at <http://r3.finngen.fi/>) (Fig. 2b).

The association of common variants at the *COL6A2* gene locus with diverticular disease was further supported by publicly available gene and protein expression data. *COL6A2* is highly expressed in connective tissue and vasculature at both the RNA and protein levels, but also in colon and intestine (Fig. 2c, d).

To validate these findings, comprehensive patient registries, with a specific focus on secondary (late-onset) phenotypes, are required; however, in the absence of such registries, the link between COL6-CMD and the gut could be studied using animal models, for example, knockouts of *Col6a2* in zebrafish or mice.

In summary, this exploratory PheWAS appears to support the hypothesis that diverticular disease may be a late-onset risk for patients carrying *COL6A2* mutations leading to Col VI-CMD. However, association does not definitively establish a causal relationship between diverticulitis and genetic defects in *COL6A2*, since other genetic and environmental factors (e.g., reduced activity levels, diet, etc.) may contribute.

It is my intention to stimulate systematic studies of whether late-onset risks in monogenic disorders can be uncovered by PheWAS analysis.

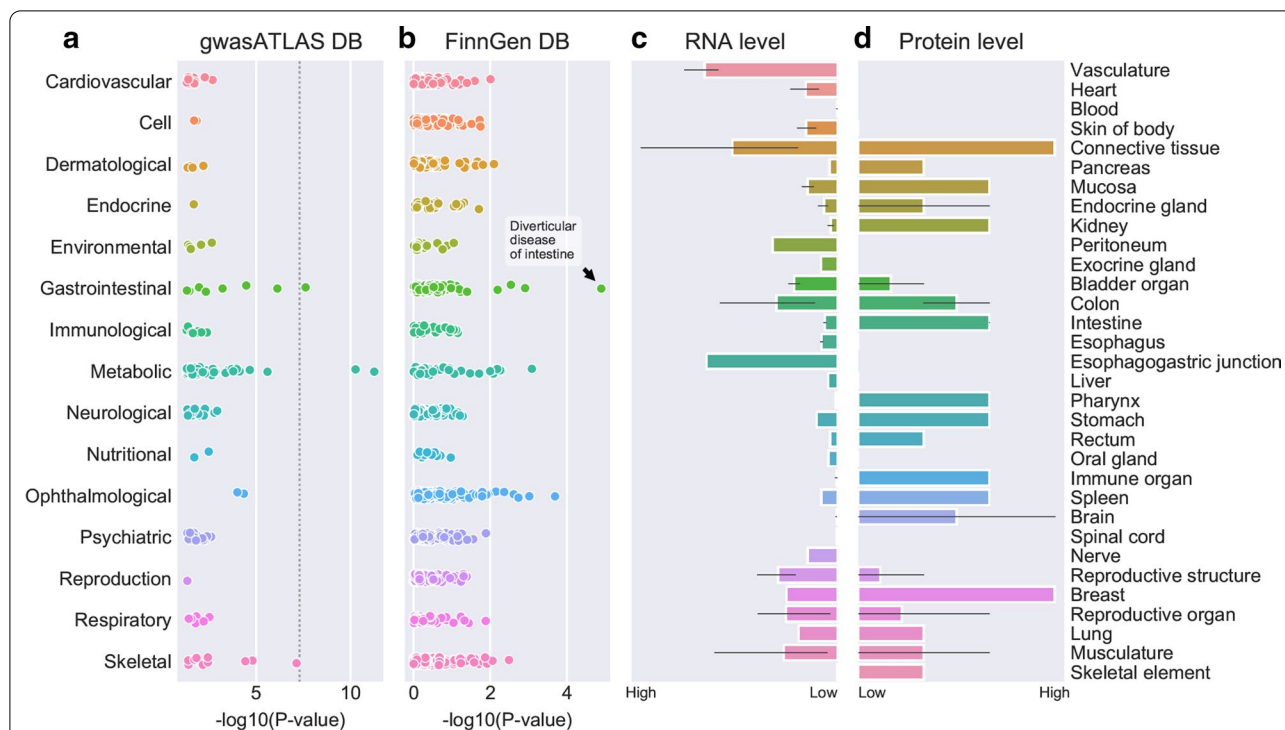


Fig. 2 a–d Results of screening for genetic associations between common variants tagging the *COL6A2* gene and a broad spectrum of phenotypes (**a, b**), as well as RNA (**c**) and protein (**d**) expression of collagen type VI, alpha 2. **a** Results of a *COL6A2* gene-based PheWAS. Plot showing association results for rs12626197, a common intronic variant tagging the *COL6A2* gene on chromosome 21, across all phenotypes in the gene atlas database (accessed July 2020). Phenotypes are clustered according to related diseases (e.g., cardiovascular diseases or gastrointestinal diseases). **b** Replication in the FinnGen study. Plot showing the association results for rs12626197 across all phenotypes (clustered as in panel **a**) in the FinnGen database (accessed July 2020). rs12626197 shows an association signal for diverticular disease, thereby replicating the finding from the gene atlas database shown in panel **a**. **c** RNA expression of *COL6A2*. Summary of *COL6A2* RNA expression in normal human tissue based on RNA-seq expression and data from the expression atlas (<https://www.targetvalidation.org/target/ENSG00000142173>, accessed November 2020). *COL6A2* is expressed in the colon and intestine, as well as other human tissues. **d** Protein expression of collagen type VI, alpha 2. Summary of protein expression in normal human tissue based on the human protein atlas (<http://www.proteinatlas.org/ENSG00000142173-COL6A2/tissue>, accessed November 2020). Collagen type VI, alpha 2 is expressed in the colon and intestine, as well as other human tissues

Abbreviations

Col VI-CMD: Collagen VI congenital muscular dystrophy; GWAS: Genome wide association study; PheWAS: Phenome-wide association study; UKBB: UK Biobank.

Acknowledgements

Thanks to Prof. Heribert Schunkert and Prof. Markus M. Nöthen for critical reading and discussion, and to Tobias Reinberger for providing Fig. 2. Thanks to the unknown reviewers for very constructive comments helping to improve the manuscript.

Authors' contributions

JE: concept, drafting. Author read and approved the final manuscript.

Funding

Funded by institutional budget.

Availability of data and materials

Not applicable.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

No competing interests.

Received: 24 September 2020 Accepted: 6 January 2021

Published online: 21 January 2021

References

- Erdmann J, Schunkert H. Forty-five years to diagnosis. *Neuromuscul Disord.* 2013;23(6):503–5.
- Jobsis GJ, Bolhuis PA, Boers JM, Baas F, Wolterman RA, Hensels GW, et al. Genetic localization of Bethlem myopathy. *Neurology.* 1996;46(3):779–82.
- Pan TC, Zhang RZ, Pericak-Vance MA, Tandan R, Fries T, Stajich JM, et al. Missense mutation in a von Willebrand factor type A domain of the alpha 3(VI) collagen gene (COL6A3) in a family with Bethlem myopathy. *Hum Mol Genet.* 1998;7(5):807–12.
- Hicks D, Farsani GT, Laval S, Collins J, Sarkozy A, Martoni E, et al. Mutations in the collagen XII gene define a new form of extracellular matrix-related myopathy. *Hum Mol Genet.* 2014;23(9):2353–63.
- Bönnemann CG. The collagen VI-related myopathies: muscle meets its matrix. *Nat Rev Neurol.* 2011;7(7):379–90.
- Chen P, Cescon M, Bonaldo P. Collagen VI in cancer and its biological mechanisms. *Trends Mol Med.* 2013;19(7):410–7.

7. Sun K, Park J, Kim M, Scherer PE. Endotrophin, a multifaceted player in metabolic dysregulation and cancer progression, is a predictive biomarker for the response to PPARgamma agonist treatment. *Diabetologia*. 2017;60(1):24–9.
8. Karkheiran S, Krebs CE, Makarov V, Nilipour Y, Hubert B, Darvish H, et al. Identification of COL6A2 mutations in progressive myoclonus epilepsy syndrome. *Hum Genet*. 2013;132(3):275–83.
9. Zech M, Lam DD, Francescato L, Schormair B, Salminen AV, Jochim A, et al. Recessive mutations in the alpha3 (VI) collagen gene COL6A3 cause early-onset isolated dystonia. *Am J Hum Genet*. 2015;96(6):883–93.
10. Landfeldt E, Thompson R, Sejersen T, McMillan HJ, Kirschner J, Lochmuller H. Life expectancy at birth in Duchenne muscular dystrophy: a systematic review and meta-analysis. *Eur J Epidemiol*. 2020;35:643–53.
11. Mercuri E, Bonnemann CG, Muntoni F. Muscular dystrophies. *Lancet*. 2019;394(10213):2025–38.
12. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene–disease associations. *Bioinformatics*. 2010;26(9):1205–10.
13. Roden DM. Phenome-wide association studies: a new method for functional genomics in humans. *J Physiol*. 2017;595(12):4109–15.
14. Cerrone M, Remme CA, Tadros R, Bezzina CR, Delmar M. Beyond the one gene-one disease paradigm: complex genetics and pleiotropy in inheritable cardiac disorders. *Circulation*. 2019;140(7):595–610.
15. Laird DW. Syndromic and non-syndromic disease-linked Cx43 mutations. *FEBS Lett*. 2014;588(8):1339–48.
16. Small AM, O'Donnell CJ, Damrauer SM. Large-scale genomic biobanks and cardiovascular disease. *Curr Cardiol Rep*. 2018;20(4):22.
17. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203–9.
18. Tcheandjieu C, Aguirre M, Gustafsson S, Saha P, Potiny P, Haendel M, et al. A phenome-wide association study of 26 mendelian genes reveals phenotypic expressivity of common and rare variants within the general population. *PLoS Genet*. 2020;16(11):e1008802.
19. Park J, Levin MG, Haggerty CM, Hartzel DN, Judy R, Kember RL, et al. A genome-first approach to aggregating rare genetic variants in LMNA for association with electronic health record phenotypes. *Genet Med*. 2020;22(1):102–11.
20. Watanabe K, Stringer S, Frei O, Umičević Mirkov M, de Leeuw C, Polderman TJC, et al. A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet*. 2019;51(9):1339–48.
21. Bulik-Sullivan BK, Loh PR, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium, et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291–5.
22. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol*. 2015;11(4):e1004219.
23. Pulit SL, Stoneman C, Morris AP, Wood AR, Glastonbury CA, Tyrrell J, et al. Meta-analysis of genome-wide association studies for body fat distribution in 694 649 individuals of European ancestry. *Hum Mol Genet*. 2019;28(1):166–74.
24. Schafmayer C, Harrison JW, Buch S, Lange C, Reichert MC, Hofer P, et al. Genome-wide association analysis of diverticular disease points towards neuromuscular, connective tissue and epithelial pathomechanisms. *Gut*. 2019;68(5):854–65.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

