HGG Advances

COMMENTARY We're right here

Toward reporting standards for the pathogenicity of variant combinations involved in multilocus/oligogenic diseases

Sofia Papadimitriou,[1,2,3,9,*] Barbara Gravel,[1,2,3] Charlotte Nachtegael,[1,2] Elfride De Baere,[4] Bart Loeys,[5] Miikka Vikkula,[6] Guillaume Smits,[1,7,8] and Tom Lenaerts[1,2,3,*]

Summary

Although standards and guidelines for the interpretation of variants identified in genes that cause Mendelian disorders have been developed, this is not the case for more complex genetic models including variant combinations in multiple genes. During a large curation process conducted on 318 research articles presenting oligogenic variant combinations, we encountered several recurring issues concerning their proper reporting and pathogenicity assessment. These mainly concern the absence of strong evidence that refutes a monogenic model and the lack of a proper genetic and functional assessment of the joint effect of the involved variants. With the increasing accumulation of such cases, it has become essential to develop standards and guidelines on how these oligogenic/multilocus variant combinations should be interpreted, validated, and reported in order to provide high-quality data and supporting evidence to the scientific community.

Introduction

The analysis of human genetic variation in relation to genetic disease has led to the discovery of patterns more complex than the notion of "one gene-one disease phenotype," promoting the acceptance that a disease can be caused or modulated by the interaction of a small set of genes through epistatic mechanisms and multilocus/oligogenic effects. Such genetic disease models come in different flavors: the variants may need to be present simultaneously to cause the disease, while for monogenic plus modifier cases, a single variant is sufficient to lead to the disease phenotype, with additional variants acting as modifiers of the severity of the disease symptoms or penetrance (age of onset).[1] Cases including the simultaneous presence of multiple Mendelian diseases leading to overlapping phenotypes can also be interesting in this context.[2] Oligogenic models, even just those involving two genes (i.e., digenic or bilocus), remain hard to detect and validate due to the fact that the variants involved can be quite common in the general population, can have a smaller individual effect on the gene function, and can be located in genes that are not necessarily known to biologically interact or already known to be involved in the same disease.[3,4] Nevertheless, data on oligogenic variant combinations has started to accumulate in the scientific literature over the last years (Figure 1) and could now be of great help to discover useful patterns and better understand the genetic architecture of these diseases.

Data related to digenic diseases, the simplest form of oligogenicity, was first collected in the Digenic Diseases Database (DIDA).[6] This database enabled the development of a new generation of predictive tools targeting combinations of variants linked to disease.[7–9] With the constant increase of new data and the emergence of more complex cases, the recently developed Oligogenic Diseases Database (OLIDA; https://olida.ibsquare.be/)[5] moves beyond the digenic limits imposed on DIDA, serving as a data repository on all types of oligogenic combinations linked to disease phenotypes. The database further provides an original curation (or confidence) score for each oligogenic combination based on the quality of evidence that supports their association with disease. Indeed, while a single reference repository on oligogenic cases is essential, the implication of the variants and genes in the associated disease needs to be properly evaluated in order for it to be useful to the medical genetics community.

During the curation of the research articles for OLIDA, several recurring issues were encountered. Although the published articles report the potential existence of oligogenic models, adequate genetic evidence excluding a monogenic model is often missing (Figure 1). Additionally, in order to better understand the mechanisms

[1]Interuniversity Institute of Bioinformatics in Brussels, Université Libre de Bruxelles-Vrije Universiteit Brussel, 1050 Brussels, Belgium; [2]Machine Learning Group, Université Libre de Bruxelles, 1050 Brussels, Belgium; [3]Artificial Intelligence Laboratory, Vrije Universiteit Brussel, 1050 Brussels, Belgium; [4]Center for Medical Genetics, Ghent University Hospital, Department of Biomolecular Medicine, Ghent University, 9000 Ghent, Belgium; [5]Center for Medical Genetics, Antwerp University Hospital/University of Antwerp, 2650 Antwerp, Belgium; [6]Human Molecular Genetics, de Duve Institute, UCLouvain, Brussels, Belgium; [7]Center of Human Genetics, Hôpital Erasme, Université Libre de Bruxelles, 1070 Brussels, Belgium; [8]Hôpital Universitaire des Enfants Reine Fabiola, Université Libre de Bruxelles, 1020 Brussels, Belgium
[9]Lead contact
*Correspondence: sofia.papadimitriou@ulb.be (S.P.), tom.lenaerts@ulb.be (T.L.)
https://doi.org/10.1016/j.xhgg.2022.100165.

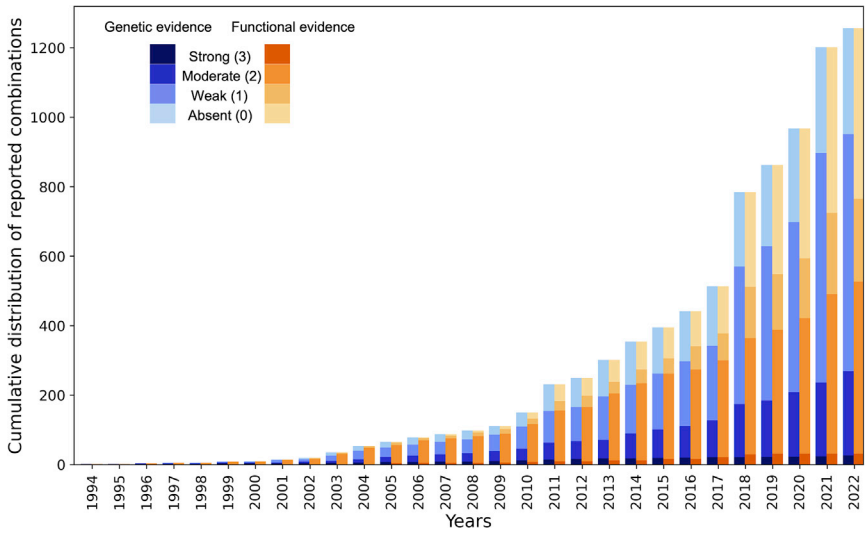

Figure 1. Cumulative number of reported oligogenic combinations per year and distribution of the associated genetic and functional evidence based on the data collected in OLIDA and extracted from articles published between January 1994 until February 2022
The color gradient represents the confidence score (as "absent," "weak," "moderate," or "strong" and the corresponding score number from 0 to 3 in parentheses) for the associated genetic (in blue) and functional (in orange) evidence linked to each oligogenic variant combination in OLIDA. A higher score is always associated with stronger evidence for the pathogenic association of an oligogenic variant combination with the described disease phenotype. Detailed information on how these scores are assigned can be found in the corresponding article of OLIDA.[5]

leading to disease, a proper functional assessment of how the variants and genes act synergistically to cause the phenotype is also important. However, functional experiments are not regularly performed, and even elaboration on the potential biological mechanisms leading to disease through the consultation of public databases or *in silico* tools is frequently lacking (Figure 1). In general, a lack of specificity has been observed in oligogenic articles that concerns both genetic and functional evidence, as many studies implicate only gene panels in their analyses and/or limited functional testing of specific oligogenic combinations. These issues tend to be even more profound for higher-order oligogenic combinations, which include more than two genes (Figure 2). Finally, it was noticed that many articles do not sufficiently report variant information, making the collection and evaluation of data tedious.

Although standards and guidelines for the interpretation and reporting of single sequence variants implicated in Mendelian diseases have been developed,[10,11] linking combinations of variants in different genes to a disease phenotype requires different considerations and is currently not subjected to any standards and guidelines. Based on the curation experience obtained while creating OLIDA, this commentary aims to initiate a discussion on how to properly report and assess the pathogenicity of oligogenic sequence variant combinations underlying genetic diseases by introducing a first set of recommendations. In our opinion, articles reporting on true oligogenic patients should address two main essential points: (1) the rejection of a monogenic model for the studied disease phenotype with the demonstration of sufficient genetic evidence that there is no coincidental presence of the associated variants in the patient, something that appears to be the most common limitation that was observed in literature, especially for monogenic plus modifier scenarios, and (2) that the evidence of a patho-mechanistic implication between an oligogenic combination and the studied phenotype is supported both genetically, with the use of family pedigrees and cohorts/variant databases, and functionally with adequate *in vivo*/*in vitro* experiments at the gene and variant level guided by *in silico* analyses.

Recommendations on reporting genetic evidence for oligogenic models

Studies investigating the association between genetic variants and disease typically do so by sequencing affected and unaffected individuals within a family (segregation studies) or within large cohorts (cohort studies). For oligogenic models, the main point of these analyses is to find evidence that proves that the co-occurrence of the variants, i.e., an oligogenic model, is compatible with the phenotype observed. In order to do that, authors should pay attention to the design of their study to ensure that they can statistically validate their observations either by confirming their results using several unrelated independent families or by dividing their cohorts into discovery and validation cohorts. Moreover, authors should carefully assess the number of individuals that they need to sequence to obtain sufficient statistical power under an oligogenic scenario and which reference population databases to use for variant interpretation (Table 1). Under different gene susceptibility and variant penetrance models, the necessary number of included cases/controls to obtain sufficient statistical power (more than 80%) for the discovery of a gene-gene association can vary from 300 to 500,000 individuals.[12,13] It is also important to broaden the search for variants beyond the existing gene panels since this could mask the presence of modifier variants in genes not already known to be involved in the disease. Ideally, a genome-wide sequencing analysis should be performed to increase as much as possible the specificity of the study. Furthermore, strict monogenic filtering

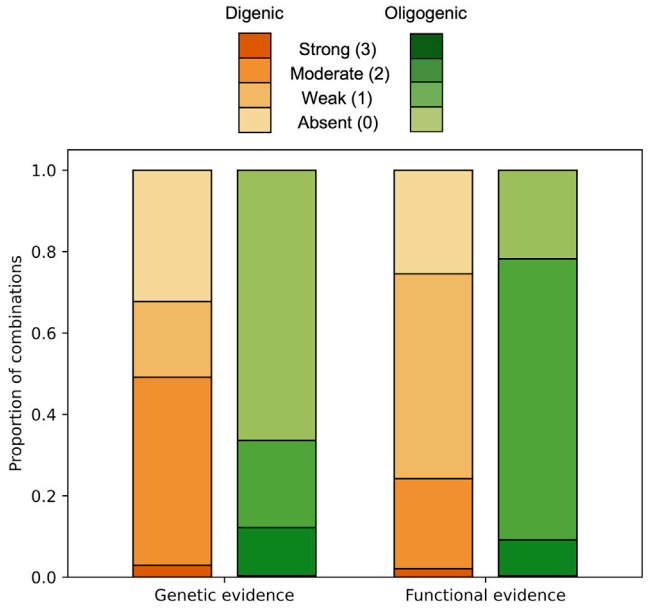

Figure 2. Proportion of the associated genetic and functional evidence scores for the digenic (in orange) and higher order (in green) oligogenic variant combinations present in OLIDA

The color gradient represents the confidence score (as "absent," "weak," "moderate," or "strong" and the corresponding score number from 0 to 3 in parentheses). A higher score is always associated with stronger evidence for the pathogenic association of an oligogenic variant combination with the described disease phenotype. Detailed information on how these scores are assigned can be found in the corresponding article of OLIDA.[5]

criteria that are commonly applied in oligogenic articles should be re-evaluated as they can lead to false-negative results, taking into account the fact that when studying multilocus models common variants can also be considered as candidates since they might not cause the phenotype individually but could contribute to it when in the presence of additional variants. For example, in the family case study of Zullo et al.,[14] variants with minor allele frequencies (MAFs) >0.2 were deemed very likely to affect the disease severity in patients with long QT syndrome, with further support from functional studies. Ideally, the allele frequency threshold to be chosen should also consider the prevalence and mode of the disease of interest, as, for example, when a monogenic plus modifier scenario is suspected, the MAF threshold can be more lenient for the detection of modifiers compared with a true oligogenic model.

The main question that needs to be addressed regarding the genetic evidence related to oligogenic models is how the phenotype of the individual with the variant combination differs from the phenotype of individuals not harboring the combination or harboring only a subset of the involved variants. The latter is particularly important not only for individual variants located in different genes but also for cases where the genes harbor homozygous or heterozygous compound variants, as proper evidence should show that these biallelic events in a single gene are not sufficient to cause the observed phenotype. The controversy following the triallelic cases of Bardet-Biedl syndrome (BBS [MIM: 209900, 600151, 605231, 615981, 615982, 615983, 615984, 615985, 615986, 615987, 615988, 615989, 615990, 615991, 615992, 615993, 615994, 615995, 615996, 617119, and 617406]), which is considered one of the most famous examples of a digenic disease, highlights the need for the provision of extensive genetic evidence with a large series of patient pedigrees coupled with functional evidence. Although the pathogenic evidence of some triallelic BBS combinations reported in OLIDA is promising, other reported cases need further support, as the required oligogenic evidence needs to demonstrate that the biallelic events in a single gene are not sufficient to cause the disease phenotype, with individuals carrying only the mutations in one gene being healthy or having less severe symptoms.

For segregation studies, addressing this question implies that both affected and unaffected individuals in—ideally several—unrelated families need to be sequenced in order to show that each variant in itself, and subset of variants in themselves, has/have a smaller or no effect on the phenotype (Figure 3). For the monogenic plus modifier scenario, we should clearly observe that the presence of the modifier gene(s) leads to the more severe phenotype of the patient compared with individuals carrying only the variant at the primary gene, while for the true oligogenic scenario, we should observe that individuals carrying either variant are unaffected. The presence and absence of variants should therefore be put in relation to the phenotype and should support the involvement of all related variant subsets in order to clearly reject a monogenic cause of the disease phenotype (Table 1). It is thus crucial that both the genotype and the phenotype of all sequenced individuals are thoroughly described. For instance, in the case study by Olivé et al.[22] on myopathy (MIM: 115197), the thorough description of phenotypic and genotypic features of eight individuals in the pedigree clearly rejects monogenicity by showing that each individual harboring variants in only one gene was unaffected. In the case where many relatives are not available for segregation analysis, the variants for which information is missing can be studied in individuals from control cohorts.

For cohort studies, searching for an oligogenic model should imply that both the frequency of the individual variants and the frequency of the oligogenic combination itself should be assessed in the control population,

Table 1. A first set of main recommendations on proper reporting on oligogenic variant combinations and provision of adequate genetic and functional evidence for their pathogenicity

Type of evidence	Type of analysis	Main goal	Recommendations on reporting	Recommendations on showing oligogenic evidence
Genetic evidence	sequencing and general reporting	proper oligogenic combination reporting	• explicit information on the sequencing procedure • explicit information on all variant and gene filtering steps (e.g., MAF, variant effect, gene panel) • comprehensive genetic reporting of sequence variants: genomic coordinates with the genome version and reference/alternative alleles, transcript information, cDNA and protein changes • complete genetic reporting of CNVs: starting and ending coordinates with genome version, explicit variation pattern and number of copies (if applicable)[15] • comprehensive reporting of the observed variant combinations per individual in a main table • consultation in OLIDA for the presence of the variant combination and its previous report • consultation in variant disease databases for genetic studies involving the individual variants (e.g., OLIDA, ClinVar, DisGeNet, OMIM[5,16–18])	• leniency on the MAF of variants is recommended, as oligogenic variants, especially those with a modifier effect, can be present individually in the control population at $>1\%$ frequency • preferably a whole-exome/genome analysis to avoid bias on current knowledge and increase specificity

(*Continued on next page*)

Table 1. ***Continued***

Type of evidence	Type of analysis	Main goal	Recommendations on reporting	Recommendations on showing oligogenic evidence
Familial evidence	Pedigree	proof that the segregation of variants is consistent with an oligogenic genetic model	• clear description of the phenotype and genotype of the proband and relatives with a family tree • involvement of first- and second-degree relatives • segregation analysis of unaffected and (if applicable) less severely affected relatives • ideally, study of minimum 2 independent families • explicit ethnicity declaration • explicit consanguinity or endogamy declaration	• in general: demonstration of a clear variant segregation that is in agreement with the studied phenotype and the unaffected/milder affected relatives • true oligogenic scenario: relatives carrying either variant are unaffected • monogenic plus modifier scenario: relatives carrying the variant at the primary gene have milder symptoms/different age of onset/different subphenotype and neither variant alone is linked to the observed phenotype of the proband • if no clear segregation, authors should elaborate on their findings • if applicable (e.g., variant combination previously reported in OLIDA), elaboration on the agreement of the current findings with previous studies

(*Continued on next page*)

Table 1. ***Continued***

Type of evidence	Type of analysis	Main goal	Recommendations on reporting	Recommendations on showing oligogenic evidence
Statistical evidence	cohort/statistical	proof that the variant combination is absent in ethnically matched control individuals not having the disease phenotype	• clear description of the phenotype and genotype of probands and control individuals • ideally an in-house control cohort with ethnically matched individuals and assessed using the same sequencing technology • necessity of discovery and independent validation cases and controls cohorts • search for the presence of the variants together and individually in the control cohort(s) • search for the presence of variants in large population databases, first as a combination (at minimum in the 1000 Genomes Project[19]) and individually (e.g., gnomAD[20]) in the ethnically matching subpopulation • consultation of the ACMG criteria on the statistical pathogenicity for each variant • explicit ethnicity declaration for both case and control cohorts	• in general: demonstration that control individuals do not carry the variant combination and, if they carry a subset of the variants involved, that they do not exhibit the same disease phenotype as the patient • true oligogenic scenario: control individuals carrying either variant are unaffected • monogenic plus modifier scenario: control individuals carrying the variant at the primary gene have milder symptoms/different age of onset/different subphenotype and neither variant alone is linked to the observed phenotype of the proband • if applicable, elaboration on the agreement of the current findings with previous studies

(*Continued on next page*)

Table 1. ***Continued***

Type of evidence	Type of analysis	Main goal	Recommendations on reporting	Recommendations on showing oligogenic evidence
Functional evidence	gene level	proof that the genes (proteins) act together to cause the described phenotype	• *in vivo* or *in vitro* combinatorial and single knockout or dosage experiments • consultation of biological databases for known relations among the genes (proteins) (e.g., PPI and pathway databases, expression data) • if applicable, report on the biological role of the involved genes (proteins) and their link to the studied phenotype	• in general: demonstration of an epistatic effect of the involved genes (proteins) compared with their individual effects • true oligogenic scenario: the knockout/lower dosage of both genes is required to cause the studied phenotype • monogenic plus modifier scenario: primary gene knockout/lower dosage causes milder phenotype/different subphenotype and neither gene alone leads to the observed phenotype of the proband • if possible, further elaboration on the biological mechanisms underlying this epistatic effect derived from previous knowledge
	variant level	proof that the variants have an impact together on the described phenotype	• *in vitro* or *in vivo* combinatorial and single mutant experiments using the corresponding variants • *in silico* analysis of the combinatorial effect of the variant combination (e.g., oligogenic or monogenic pathogenicity predictors[7] and 3D modeling[21]) and consultation of the ACMG criteria on the functional pathogenicity for each variant • search in variant disease databases for functional studies involving either the variant combination or the individual variants (e.g., OLIDA, ClinVar, DisGeNet, OMIM[5,16–18])	• in general: demonstration of a joint pathogenic effect of the variant combination compared with their individual effects and the effects of any subcombinations • true oligogenic scenario: both variants are required to cause the observed phenotype • monogenic plus modifier scenario: the variant at the primary gene leads to milder symptoms/different age of onset/different subphenotype and neither variant alone leads to the observed phenotype of the proband • if possible, further elaboration of the biological mechanism in which the variants are causing the disease phenotype

The "true oligogenic" scenario corresponds to the case where all variants need to be present simultaneously to show any disease symptoms, while in the "monogenic plus modifier" scenario, the variant at the primary gene can still cause milder symptoms or a different sub-phenotype. MAF, minor allele frequency; SNV, single-nucleotide variant; indel, insertion or deletion; CNV, copy-number variant; PPI, protein-protein interaction.

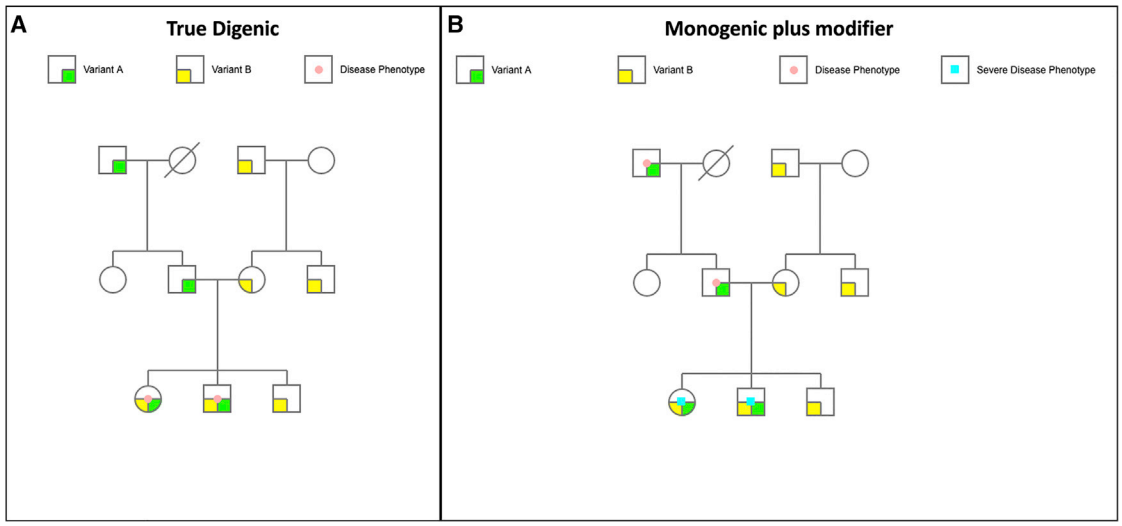

Figure 3. Examples of family pedigrees that demonstrate a clear genetic segregation for a digenic variant combination involving gene A and gene B

The digenic variant combination is associated with the disease phenotype under (A) the true digenic model, where the simultaneous presence of both variants in an individual are necessary for the development of the disease phenotype, and (B) the monogenic plus modifier model, where the variant at the second gene modifies the severity or age of onset of the symptoms caused by variant in the first gene. Variant A in the first gene is shown in green, variant B in the second gene is shown in yellow, and the individuals with the disease phenotype are shown with a red or blue dot.

the latter case being often missed in oligogenic papers. For the simplest case of a digenic disease, under a high digenic penetrance and/or high variant frequency scenario, it is estimated that 35,000 control samples could be sufficient to reach 80% statistical power, whereas control samples of sizes between 100,000 and 500,000 would be needed for digenic combinations of lower penetrance and/or involving rare variants[13]. Access to genome biobanks, such as the UK Biobank, can serve this purpose. However, to facilitate the process of properly assessing oligogenic combinations in clinical genetics, we believe that there is also an urgent need for large public control databases that can either offer the frequency of combinations of variants in a particular population or provide access to whole-genome sequences. For now, a search for the identified variant combinations in the 1000 Genomes Project,[19] which includes publicly available data from different ethnicities, already provides a first statistical indication, yet it is clearly underpowered, especially for rare and heterogeneous diseases (Table 1). Apart from a direct search in control populations, additional statistical analyses can also be conducted, such as demonstrating a stronger or earlier onset of the phenotype for individuals with oligogenic variants compared with individuals with monogenic variants[23] or by computing an odds ratio of finding the variant combination in an individual based on the single variants' allele frequencies in the population.[24]

Finally, it is important to stress that the identified variants should be thoroughly and accurately described by applying the Human Genome Variation Society (HGVS) recommendations on variant description nomenclature[25] and offer information at the genome, cDNA, and protein levels, when applicable, to avoid any ambiguity in their identification in future studies. For example, frameshift variants should not only be described at the protein level, as they could be caused by several different variants at the gDNA level. Furthermore, the identification number of the reference sequence should always be reported. Accuracy in the notation of these variants is essential for the data to be collected in repositories such as OLIDA and can help to retrieve already known information about their eventual involvement in genetic diseases from external databases.

Recommendations on reporting functional evidence for oligogenic models

Another limitation that we encountered in literature is the absence of adequate functional evidence implicating an oligogenic combination in a disease and a lack of elaboration on the biological mechanisms that are implicated for the genes and variants. While having undisputable genetic evidence can be a convincing argument for oligogenicity, high-quality functional evidence can bring an essential additional insight into the underlying molecular mechanism of such a model. This is particularly important in cases where undisputable genetic evidence is not available, as is the case for the majority of scientific papers that, at the same time, do not validate their results experimentally to compensate for the lack of sufficient genetic support (Figure 1). Even for individual variants, we have many times encountered the term "pathological" or "pathogenic" supported mostly by statistical evidence. This then leads to a cycle where authors claim the "pathogenicity" of oligogenic combinations without any individual or combinatorial functional evidence of pathogenicity.

In order to obtain functional oligogenic evidence of good quality, the focus should be on the causality of the multiple variants together for the studied disease phenotype and the demonstration of a synergistic or additive effect (i.e., proof that the effect of the combination is different than the variants' individual effects) using *in vivo* and/or *in vitro* experiments (Table 1). Nevertheless, in many papers, this information is either completely missing or the results are not sufficient to exclude a monogenic cause, as the authors often focus either on the individual effect of the involved genes and variants or only on the oligogenic combination itself without making comparisons to exclude a monogenic model.

An effort should be made to provide functional evidence both at the gene (including transcript and/or protein) and variant levels (Table 1). Addressing the pathogenicity of an oligogenic combination at the gene level is very important in understanding how they act synergistically to cause the disease. This can be achieved experimentally by, e.g., conducting *in vitro* or *in vivo* dosage or combinatorial gene knockout experiments, which should always be coupled with the respective single-gene experiments for comparison. The goal of any functional assessment trying to support an oligogenic model should be to demonstrate an epistatic effect of the involved genes, either by leading to the studied disease phenotype for a true oligogenic scenario or by affecting its severity, for the monogenic plus modifier scenario. The comparison with their individual effects is important in order to discard a monogenic model. Nevertheless, the variants involved in a specific oligogenic combination may not functionally be responsible for the disease even if there is proof that the genes act synergistically. Therefore, a similar synergistic assessment at the variant level is also needed with, e.g., experiments using *in vivo* models, cultures of patient cells, expression studies, and co-immunoprecipitation and binding assays (Table 1). An example of a nicely founded oligogenic functional assessment is described in the paper of Wang et al. on assessing a digenic etiology for Müllerian aplasia (MIM: 158330).[26] The authors compared double heterozygous, single knockout, and wild-type mouse models, by introducing the variants of the genes *GEN1* (MIM: 612449) and *WNT9B* (MIM: 602864) that are supposed to be involved in an oligogenic combination and demonstrated a clear synergistic effect for the double heterozygous mice on the uterus. Another example of a paper providing good functional evidence for a higher-order oligogenic case is that of Gifford et al., who described an oligogenic combination involving the genes *MRTFB* (MIM: 609463), *MYH7* (MIM: 160760), and *NKX2-5* (MIM: 600584) linked with left ventricular noncompaction (LVNC [MIM: 604169]).[27] Their work demonstrated that this variant combination produced the LVNC phenotype in mice and had a more detrimental effect compared with mice carrying individual and double heterozygous mutations.

We acknowledge that the conduction of functional experiments is often not possible due to lack of adequate *in vitro*/*in vivo* models or a lack of resources and equipment, especially in a clinical context. Nevertheless, with the advent of numerous computational tools in the field, it is possible to provide as common practice, and a bare minimum, *in silico* evidence of the pathogenic potential of a variant combination, even though such information is not adequate by itself to provide a diagnosis. This can be done by using combinatorial and monogenic pathogenicity predictors or tools assessing the 3D structure of proteins and how the involved variants can affect their interaction (Table 1). Several resources can also be used to better understand and interpret *in silico* the involvement of the genes in the disease, such as Gene Ontology (and Gene Ontology enrichment analyses), gene interaction databases, and biological pathway databases. *In silico* tools could help reduce the problem of lack of specificity in oligogenic studies by allowing us to consider many possible scenarios first and limiting the amount of oligogenic combinations that need to be tested functionally. Nevertheless, we would like to stress that functional experiments are the only means to provide definite proof of synergy among the involved genes and variants.

Conclusions and final remarks

By sharing these general considerations that were gathered during the curation process of OLIDA, we would like to open the discussion on the improvement of data quality in the field of oligogenic/multilocus disease research. We define a first set of standards and guidelines on how to properly report the evidence supporting oligogenic/multilocus models underlying genetic diseases. These recommendations are open to further assessment and debate, but as the number of publications identifying oligogenic causes to disease is increasing rapidly (Figure 1), initiating this discussion is imperative. We believe that it is also crucial that clear definitions of the different oligogenic cases, especially when modifier genes are involved, are put into place so that standardized and more specialized criteria can be defined for each type. Developing such concrete standards can also lead to the refutation of previously published oligogenic associations when these do not conform with the presence of at least the minimum required pathogenic synergistic evidence. Papers refuting oligogenic models after the presence of additional evidence have already emerged.[28] A standardized evaluation of reported oligogenic combinations can be done with the creation, for example, of a ClinGen Variant Combination Curation expert panel that will ensure the standards for their proper reporting and flag or refute combinations linked with insufficient evidence

based on, e.g., confidence scores like those in OLIDA.[5]

Proper reporting and assessment of oligogenic variant combinations contributes to the medical genetics field not only by consolidating the evidence supporting an oligogenic model but also by decreasing the amount of work of biocurators by, e.g., facilitating the addition of the data in public repositories such as OLIDA. Tools such as PubReCheck[29] can be used to make papers machine readable and can help increase their discoverability and interpretability with text-mining techniques, supporting the work of biocuration[29,30] and promoting the findings to other experts, such as the curation expert panels of ClinGen.[31]

Data and code availability

The source data for Figures 1 and 2 in the commentary are available at the OLIDA website (https://olida.ibsquare.be/).

Acknowledgments

This work has been conducted thanks to the following funding: Fonds de la Recherche Scientifique - FNRS (F.R.S.-FNRS) Fund for Research Training in Industry and Agriculture (40008622 to B.G.); F.R.S.-FNRS Postdoctoral Researcher Fellowship (40005602 to S.P.); F.R.S.-FNRS Projet de Recherche (35276964 to T.L.); F.R.S.-FNRS FRFS-WELBIO grant (WELBIO-CR-2019C-06 to M.V.); European Regional Development Fund and Innoviris (Brussels Public Organisation for Research and Innovation) within the framework of the Operational Program 2014–2020 ERDF-2020 project ICITY-RDI.-BRU (27.002.53.01.4524 to T.L.); Innoviris (Brussels Public Organisation for Research and Innovation) Joint R&D project Genome4Brussels (2020 RDIR 55b to T.L. and S.P.); European Research Council for the Consolidator Grant (Genomia – ERC-COG-2017-771945 to B.L.); Service Public de Wallonie Recherche by DIGITALWALLONIA4.AI (2010235—ARIAC to C.N. and T.L.); Research Foundation—Flanders (F.W.O.) Infrastructure project associated with ELIXIR Belgium (I002819N to T.L.); European Joint Programme on Rare Diseases Solve-RET project (EJPRD19-234 to E.D.B.); Research Foundation—Flanders (F.W.O.) Senior Clinical Investigator grant (1802220N to E.D.B.); and the European Reference Network – Eye Diseases Framework Partnership Agreement (739534-ERN-EYE to E.D.B.). Furthermore, the authors thank the Foundation 101 Genomes (f101g.org) for a fruitful collaboration, creative exchange and scientific support.

Declaration of interests

The authors declare no competing interests.

Web resources

OLIDA, https://olida.ibsquare.be/
OMIM, https://www.omim.org/
UK Biobank, https://www.ukbiobank.ac.uk/

References

1. Schäffer, A.A. (2013). Digenic inheritance in medical genetics. J. Med. Genet. *50*, 641–652.
2. Posey, J.E., Harel, T., Liu, P., Rosenfeld, J.A., James, R.A., Coban Akdemir, Z.H., Walkiewicz, M., Bi, W., Xiao, R., Ding, Y., et al. (2017). Resolution of disease phenotypes resulting from multilocus genomic variation. N. Engl. J. Med. *376*, 21–31.
3. Kousi, M., and Katsanis, N. (2015). Genetic modifiers and oligogenic inheritance. Cold Spring Harb. Perspect. Med. *5*, a017145.
4. Cordell, H.J. (2002). Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. Hum. Mol. Genet. *11*, 2463–2468.
5. Nachtegael, C., Gravel, B., Dillen, A., Smits, G., Nowé, A., Papadimitriou, S., and Lenaerts, T. (2022). Scaling up oligogenic diseases research with OLIDA: the oligogenic diseases database. Database *2022*, baac023.
6. Gazzo, A.M., Daneels, D., Cilia, E., Bonduelle, M., Abramowicz, M., Van Dooren, S., Smits, G., and Lenaerts, T. (2016). DIDA: a curated and annotated digenic diseases database. Nucleic Acids Res. *44*, D900–D907.
7. Papadimitriou, S., Gazzo, A., Versbraegen, N., Nachtegael, C., Aerts, J., Moreau, Y., Van Dooren, S., Nowé, A., Smits, G., and Lenaerts, T. (2019). Predicting disease-causing variant combinations. Proc. Natl. Acad. Sci. USA *116*, 11878–11887.
8. Renaux, A., Papadimitriou, S., Versbraegen, N., Nachtegael, C., Boutry, S., Nowé, A., Smits, G., and Lenaerts, T. (2019). ORVAL: a novel platform for the prediction and exploration of disease-causing oligogenic variant combinations. Nucleic Acids Res. *47*, W93–W98.
9. Versbraegen, N., Fouché, A., Nachtegael, C., Papadimitriou, S., Gazzo, A., Smits, G., and Lenaerts, T. (2019). Using game theory and decision decomposition to effectively discern and characterise bi-locus diseases. Artif. Intell. Med. *99*, 101690.
10. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., Grody, W.W., Hegde, M., Lyon, E., Spector, E., et al. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the association for molecular pathology. Genet. Med. *17*, 405–424.
11. MacArthur, D.G., Manolio, T.A., Dimmock, D.P., Rehm, H.L., Shendure, J., Abecasis, G.R., Adams, D.R., Altman, R.B., Antonarakis, S.E., Ashley, E.A., et al. (2014). Guidelines for investigating causality of sequence variants in human disease. Nature *508*, 469–476.
12. Gauderman, W.J. (2002). Sample size requirements for association studies of gene-gene interaction. Am. J. Epidemiol. *155*, 478–484.
13. Moreno-Ruiz, N., Genomics England Research Consortium, Lao, O., Aróstegui, J.I., Laayouni, H., and Casals, F. (2022). Assessing the digenic model in rare disorders using population sequencing data. Eur. J. Hum. Genet. *30*, 1439–1443.
14. Zullo, A., Frisso, G., Detta, N., Sarubbi, B., Romeo, E., Cordella, A., Vanoye, C.G., Calabrò, R., George, A.L., and Salvatore, F. (2017). Allelic complexity in long QT syndrome: a family-case study. Int. J. Mol. Sci. *18*, 1633.
15. Riggs, E.R., Andersen, E.F., Cherry, A.M., Kantarci, S., Kearney, H., Patel, A., Raca, G., Ritter, D.I., South, S.T., Thorland, E.C., et al. (2020). Technical standards for the interpretation and reporting of constitutional copy-number variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics (ACMG) and the Clinical Genome Resource (ClinGen). Genet. Med. *22*, 245–257.
16. Landrum, M.J., Lee, J.M., Benson, M., Brown, G.R., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Jang, W., et al. (2018). ClinVar: improving

access to variant interpretations and supporting evidence. Nucleic Acids Res. *46*, D1062–D1067.
17. Piñero, J., Ramírez-Anguita, J.M., Saüch-Pitarch, J., Ronzano, F., Centeno, E., Sanz, F., and Furlong, L.I. (2019). The DisGeNET knowledge platform for disease genomics: 2019 update. Nucleic Acids Res. *48*, D845–D855.
18. Amberger, J.S., Bocchini, C.A., Scott, A.F., and Hamosh, A. (2019). OMIM.org: leveraging knowledge across phenotype–gene relationships. Nucleic Acids Res. *47*, D1038–D1043.
19. 1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R. (2015). The 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature *526*, 68–74.
20. Karczewski, K.J., Francioli, L.C., Tiao, G., Cummings, B.B., Alföldi, J., Wang, Q., Collins, R.L., Laricchia, K.M., Ganna, A., Birnbaum, D.P., et al. (2020). The mutational constraint spectrum quantified from variation in 141, 456 humans. Nature *581*, 434–443.
21. Pahari, S., Li, G., Murthy, A.K., Liang, S., Fragoza, R., Yu, H., and Alexov, E. (2020). SAAMBE-3D: predicting effect of mutations on protein-protein interactions. Int. J. Mol. Sci. *21*, 2563.
22. Olivé, M., Abdul-Hussein, S., Oldfors, A., González-Costello, J., van der Ven, P.F.M., Fürst, D.O., González, L., Moreno, D., Torrejón-Escribano, B., Alió, J., et al. (2015). New cardiac and skeletal protein aggregate myopathy associated with combined MuRF1 and MuRF3 mutations. Hum. Mol. Genet. *24*, 6264.
23. Jiang, H., Liang, S., He, K., Hu, J., Xu, E., Lin, T., Meng, Y., Zhao, J., Ma, J., Gao, R., et al. (2020). Exome sequencing analysis identifies frequent oligogenic involvement and variants in adolescent idiopathic scoliosis. J. Med. Genet. *57*, 405–413.
24. Roman-Naranjo, P., Moleon, M.D.C., Aran, I., Escalera-Balsera, A., Soto-Varela, A., Bächinger, D., Gomez-Fiñana, M., Eckhard, A.H., and Lopez-Escamez, J.A. (2021). Rare coding variants involving MYO7A and other genes encoding stereocilia link proteins in familial meniere disease. Hear. Res. *409*, 108329.
25. den Dunnen, J.T., Dalgleish, R., Maglott, D.R., Hart, R.K., Greenblatt, M.S., McGowan-Jordan, J., Roux, A.-F., Smith, T., Antonarakis, S.E., and Taschner, P.E.M. (2016). HGVS recommendations for the description of sequence variants: 2016 update. Hum. Mutat. *37*, 564–569.
26. Wang, L., Zhang, Y., Fu, X., Dong, S., Tang, S., Zhang, N., Song, C., Yang, N., Zhang, L., Wang, H., et al. (2020). Joint utilization of genetic analysis and semi-cloning technology reveals a digenic etiology of Müllerian anomalies. Cell Res. *30*, 91–94.
27. Gifford, C.A., Ranade, S.S., Samarakoon, R., Salunga, H.T., de Soysa, T.Y., Huang, Y., Zhou, P., Elfenbein, A., Wyman, S.K., Bui, Y.K., et al. (2019). Oligogenic inheritance of a human heart disease involving a genetic modifier. Science *364*, 865–870.
28. Ołdak, M., Lechowicz, U., Pollak, A., Oziębło, D., and Skarżyński, H. (2019). Overinterpretation of high throughput sequencing data in medical genetics: first evidence against TMPRSS3/GJB2 digenic inheritance of hearing loss. J. Transl. Med. *17*, 1–6.
29. Leaman, R., Wei, C.-H., Allot, A., and Lu, Z. (2020). Ten tips for a text-mining-ready article: how to improve automated discoverability and interpretability. PLoS Biol. *18*, e3000716.
30. Singhal, A., Leaman, R., Catlett, N., Lemberger, T., McEntyre, J., Polson, S., Xenarios, I., Arighi, C., and Lu, Z. (2016). Pressing needs of biomedical text mining in biocuration and beyond: opportunities and challenges. Database *2016*, baw161.
31. Rehm, H.L., Berg, J.S., Brooks, L.D., Bustamante, C.D., Evans, J.P., Landrum, M.J., Ledbetter, D.H., Maglott, D.R., Martin, C.L., Nussbaum, R.L., et al. (2015). ClinGen–the clinical genome resource. N. Engl. J. Med. *372*, 2235–2242.