Genotype-first approach to identify associations between *CDH1* germline variants and cancer phenotypes: a multicentre study by the European Reference Network on Genetic Tumour Risk Syndromes

José Garcia-Pelaez, Rita Barbosa-Matos, Silvana Lobo*, Alexandre Dias*, Luzia Garrido, Sérgio Castedo, Sónia Sousa, Hugo Pinheiro, Liliana Sousa, Rita Monteiro, Joaquin J Maqueda, Susana Fernandes, Fátima Carneiro, Nádia Pinto, Carolina Lemos, Carla Pinto, Manuel R Teixeira, Stefan Aretz, Svetlana Bajalica-Lagercrantz, Judith Balmaña, Ana Blatnik, Patrick R Benusiglio, Maud Blanluet, Vincent Bours, Hilde Brems, Joan Brunet, Daniele Calistri, Gabriel Capellá, Sergio Carrera, Chrystelle Colas, Karin Dahan, Robin de Putter, Camille Desseignés, Elena Domínguez-Garrido, Conceição Egas, D Gareth Evans, Damien Feret, Eleanor Fewings, Rebecca C Fitzgerald, Florence Coulet, María Garcia-Barcina, Maurizio Genuardi, Lisa Golmard, Karl Hackmann, Helen Hanson, Elke Holinski-Feder, Robert Hüneburg, Mateja Krajc, Kristina Lagerstedt-Robinson, Conxi Lázaro, Marjolijn J L Ligtenberg, Cristina Martínez-Bouzas, Sonia Merino, Geneviève Michils, Srdjan Novaković, Ana Patiño-García, Guglielmina Nadia Ranzani, Evelin Schröck, Inês Silva, Catarina Silveira, José L Soto, Isabel Spier, Verena Steinke-Lange, Gianluca Tedaldi, María-Isabel Tejada, Emma R Woodward, Marc Tischkowitz, Nicoline Hoogerbrugge, Carla Oliveira*

Summary

Background Truncating pathogenic or likely pathogenic variants of *CDH1* cause hereditary diffuse gastric cancer (HDGC), a tumour risk syndrome that predisposes carrier individuals to diffuse gastric and lobular breast cancer. Rare *CDH1* missense variants are often classified as variants of unknown significance. We conducted a genotype–phenotype analysis in families carrying rare *CDH1* variants, comparing cancer spectrum in carriers of pathogenic or likely pathogenic variants (PV/LPV; analysed jointly) or missense variants of unknown significance, assessing the frequency of families with lobular breast cancer among PV/LPV carrier families, and testing the performance of lobular breast cancer-expanded criteria for *CDH1* testing.

Methods This genotype-first study used retrospective diagnostic and clinical data from 854 carriers of 398 rare *CDH1* variants and 1021 relatives, irrespective of HDGC clinical criteria, from 29 institutions in ten member-countries of the European Reference Network on Tumour Risk Syndromes (ERN GENTURIS). Data were collected from Oct 1, 2018, to Sept 20, 2022. Variants were classified by molecular type and clinical actionability with the American College of Medical Genetics and Association for Molecular Pathology *CDH1* guidelines (version 2). Families were categorised by whether they fulfilled the 2015 and 2020 HDGC clinical criteria. Genotype–phenotype associations were analysed by Student's *t* test, Kruskal-Wallis, χ^2, and multivariable logistic regression models. Performance of HDGC clinical criteria sets were assessed with an equivalence test and Youden index, and the areas under the receiver operating characteristic curves were compared by *Z* test.

Findings From 1971 phenotypes (contributed by 854 probands and 1021 relatives aged 1–93 years), 460 had gastric and breast cancer histology available. *CDH1* truncating PV/LPVs occurred in 176 (21%) of 854 families and missense variants of unknown significance in 169 (20%) families. Multivariable logistic regression comparing phenotypes occurring in families carrying PV/LPVs or missense variants of unknown significance showed that lobular breast cancer had the greatest positive association with the presence of PV/LPVs (odds ratio 12·39 [95% CI 2·66–57·74], p=0·0014), followed by diffuse gastric cancer (8·00 [2·18–29·39], p=0·0017) and gastric cancer (7·81 [2·03–29·96], p=0·0027). 136 (77%) of 176 families carrying PV/LPVs fulfilled the 2015 HDGC criteria. Of the remaining 40 (23%) families, who did not fulfil the 2015 criteria, 11 fulfilled the 2020 HDGC criteria, and 18 had lobular breast cancer only or lobular breast cancer and gastric cancer, but did not meet the 2020 criteria. No specific *CDH1* variant was found to predispose individuals specifically to lobular breast cancer, although 12 (7%) of 176 PV/LPV carrier families had lobular breast cancer only. Addition of three new lobular breast cancer-centred criteria improved testing sensitivity while retaining high specificity. The probability of finding *CDH1* PV/LPVs in patients fulfilling the lobular breast cancer-expanded criteria, compared with the 2020 criteria, increased significantly (AUC 0·92 *vs* 0·88; *Z* score 3·54; p=0·0004).

Interpretation *CDH1* PV/LPVs were positively associated with HDGC-related phenotypes (lobular breast cancer, diffuse gastric cancer, and gastric cancer), and no evidence for a positive association with these phenotypes was found for *CDH1* missense variants of unknown significance. *CDH1* PV/LPVs occurred often in families with lobular breast cancer who did not fulfil the 2020 HDGC criteria, supporting the expansion of lobular breast cancer-centred criteria.

Lancet Oncol 2023; 24: 91–106

Published **Online** November 24, 2022 https://doi.org/10.1016/S1470-2045(22)00643-X

This online publication has been corrected. The corrected version first appeared at thelancet.com/oncology on January 3, 2022

*Contributed equally

Instituto de Investigação e Inovação em Saúde (i3S) (J Garcia-Pelaez MSc, R Barbosa-Matos MSc, S Lobo MSc, A Dias MSc, Prof S Castedo MD PhD, S Sousa BSc, H Pinheiro MD PhD, L Sousa MSc, C Lemos PhD, R Monteiro MSc, J J Maqueda MSc, Prof F Carneiro MD PhD, N Pinto PhD, Prof C Oliveira PhD), **Faculty of Medicine** (J Garcia-Pelaez, Prof S Castedo, S Fernandes PhD, Prof F Carneiro, Prof C Oliveira), **Institute of Biomedical Sciences Abel Salazar** (R Barbosa-Matos, S Lobo, C Lemos PhD, M R Teixeira MD PhD), **Center of Mathematics** (N Pinto), **and Institute for Molecular and Cell Biology** (C Lemos), **University of Porto, Porto, Portugal; Instituto de Patologia e Imunologia Molecular da Universidade do Porto, Porto, Portugal** (J Garcia-Pelaez, R Barbosa-Matos, S Lobo, A Dias, Prof S Castedo, S Sousa, H Pinheiro, L Sousa, R Monteiro, J J Maqueda, Prof F Carneiro, N Pinto, Prof C Oliveira)**; Centro**

Hospitalar Universitário São João, Porto, Portugal (L Garrido RN, Prof S Castedo, Prof F Carneiro)**; Serviço de Medicina Interna, Centro Hospitalar Tâmega e Sousa, Penafiel, Portugal** (H Pinheiro)**; Escola de Economia e Gestão, Universidade do Minho, Braga, Portugal** (L Sousa)**; Bioinf2Bio, Porto, Portugal** (J J Maqueda)**; Department of Laboratory Genetics, Portuguese Oncology Institute of Porto, Porto, Portugal** (C Pinto MSc, M R Teixeira)**; Porto Comprehensive Cancer Center Raquel Seruca, Porto, Portugal** (J Garcia-Pelaez, R Barbosa-Matos, S Lobo, A Dias, Prof S Castedo, S Sousa, H Pinheiro, L Sousa, C Lemos, Prof F Carneiro, N Pinto, C Pinto, M R Teixeira, Prof C Oliveira)**; Institute of Human Genetics, Medical Faculty, University of Bonn, Bonn, Germany** (Prof S Aretz MD PhD, I Spier MD)**; Department of Internal Medicine I** (R Hüneburg MD) **and National Center for Hereditary Tumor Syndromes** (Prof S Aretz, R Hüneburg, I Spier)**, University Hospital Bonn, Bonn, Germany; Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden** (S Bajalica-Lagercrantz MD PhD, K Lagerstedt-Robinson PhD)**; Department of Clinical Genetics, Cancer Genetic Unit** (S Bajalica-Lagercrantz, K Lagerstedt-Robinson) **and Cancer Theme** (S Bajalica-Lagercrantz)**, Karolinska University Hospital Solna, Stockholm, Sweden; Hospital Vall d'Hebron, Universitat Autonoma de Barcelona, Barcelona, Spain** (J Balmaña MD PhD)**; Department of Molecular Diagnostics** (Prof S Novaković PhD) **and Department of Clinical Cancer Genetics** (A Blatnik MD, M Krajc MD PhD)**, Institute of Oncology Ljubljana, Ljubljana, Slovenia; Medical Genetics Department, Pitié-Salpêtrière Hospital, AP-HP and Sorbonne University, Paris, France** (P R Benusiglio MD, C Desseignés MSc, F Coulet PharmD PhD)**; Service de Génétique Oncologique, Institut Curie, Paris, France** (L Golmard PharmD PhD, M Blanluet PharmD, C Colas MD PhD)**; Laboratory of**

Funding European Reference Network on Genetic Tumour Risk Syndromes, European Regional Development Fund, Fundação para a Ciência e a Tecnologia (Portugal), Cancer Research UK, and European Union's Horizon 2020 research and innovation programme.

Introduction

Hereditary diffuse gastric cancer (HDGC; OMIM number #137215) is an autosomal dominant tumour risk syndrome with increased predisposition to develop early onset diffuse gastric cancer and lobular breast cancer.[1–3] HDGC is mainly caused by germline single-nucleotide variants and copy-number variants in the E-cadherin (*CDH1*) gene (Ensembl gene ID ENSG00000039068; RefSeq ID NM_004360.4; Locus Reference Genomic ID LRG_301),[4,5] which are classified as pathogenic or likely pathogenic variants, according to the American College of Medical Genetics and Genomics and Association for Molecular Pathology *CDH1* variant curation guidelines.[6] The International Gastric Cancer Linkage Consortium (IGCLC) has published four sets of clinical criteria for *CDH1* testing to date (appendix p 1).[1–3,7] Although guidelines from 1999 supported *CDH1* genetic testing only in families with multiple cases of diffuse gastric cancer, the 2010 and 2015 guidelines widened testing to isolated patients with diffuse gastric cancer, aged 40 years or younger, and to individuals or families with both diffuse gastric cancer and lobular breast cancer before 50 years of age.[1,7] Phenotype-driven HDGC guidelines were superseded by genotype-driven guidelines in 2020, with clinical selection criteria being reformulated predominantly on the basis of actionable genetic test results (presence of a *CDH1* pathogenic or likely pathogenic variant).[2] The 2020 guideline proposed hereditary lobular breast cancer (HLBC) as a lobular breast cancer only predisposition syndrome, independent of HDGC (appendix p 1).[2,8]

In a large US cohort of probands (tested because of HDGC clinical ascertainment or multigene panel testing), individuals carrying *CDH1* pathogenic or likely pathogenic variants were calculated to have a lifetime risk of gastric cancer of 42% for men and 33% for women, whereas female breast cancer cumulative incidence was

Research in context

Evidence before this study

In 2018, we started a genotype-first study to define the cancer landscape in families carrying *CDH1* pathogenic or likely pathogenic variants and missense variants of unknown significance, assess the frequency of families with lobular breast cancer among families carrying pathogenic or likely pathogenic variants of *CDH1*, and test the performance of different sets of hereditary diffuse gastric cancer (HDGC) clinical criteria driving *CDH1* testing. For the latter, we considered different sets of HDGC criteria developed between 2015 and 2022 (2015, 2020, Yale—2022) for comparison with the lobular breast cancer-expanded criteria herein proposed. We searched PubMed for studies in English on "*CDH1* germline variants" and "diffuse gastric cancer" or "lobular breast cancer" or "LBC" or "hereditary diffuse gastric cancer" or "HDGC" from Jan 1, 2015 onwards (the year of publication of the HDGC policy review and the first rules for variant classification in the clinical context, by the American College of Medical Genetics and Genomics and Association for Molecular Pathology). References from relevant articles, HDGC guidelines papers before 2015, reviews, and previous meta-analyses were considered on a case-by-case basis to identify any additional or relevant study not retrieved by the PubMed search. In 2020, a new HDGC policy review highlighted the lack of genotype–phenotype correlations in *CDH1* variant carriers, which still vastly persists today, particularly regarding *CDH1*-related predisposition to lobular breast cancer, associations between missense *CDH1* variants of unknown significance and HDGC-related phenotypes, and *CDH1* variant-specific cancer predisposition. This knowledge gap provides the rationale for designing more accurate clinical criteria and for management of families with HDGC.

Added value of this study

To our knowledge, this is the first multicentre genotype–phenotype study and the largest dataset ever studied of individuals carrying rare *CDH1* variants and their relatives. It highlights specificities in *CDH1* variant-type related clinical phenotypes and supports the association of HDGC-related cancers with the presence of *CDH1* pathogenic or likely pathogenic variants, but did not find an association of HDGC-related cancers with the presence of *CDH1* missense variants of unknown significance. The study also demonstrates that families not fulfilling HDGC clinical criteria, but carrying a *CDH1* PV/LPVs, are mainly lobular breast cancer-enriched, supporting an expansion of current HDGC criteria to include additional lobular breast cancer-centred criteria.

Implications of all the available evidence

Given the lack of evidence supporting a positive association between the presence of *CDH1* germline missense variants and HDGC-related cancers, carriers of such variants are unlikely to be predisposed to HDGC. The recurrent identification of pathogenic or likely pathogenic variants in patients and families with lobular breast cancer, with or without gastric cancer, supports the expansion of clinical criteria for *CDH1* testing in this setting. The observations herein reported will contribute to clarify the context for clinical management of families carrying *CDH1* germline variants.

estimated at 55% at age 80 years.[9] These cumulative incidences are considerably lower for gastric cancer and higher for breast cancer compared with cohorts strictly fulfilling HDGC criteria and those from other world regions.[10–13] Given the high risk of cancer in carriers of *CDH1* pathogenic or likely pathogenic variants, current guidelines recommend intensive gastric and breast surveillance, prophylactic gastrectomy in men and women, and optional mastectomy in women, after a *CDH1* pathogenic or likely pathogenic variant is found in a family or individual with relevant history of cancer in these organs.[2]

Many *CDH1* pathogenic or likely pathogenic variants, found by multigene panel testing, occur in breast cancer patients not meeting the clinical criteria for HDGC.[12,14] Additionally, a 2022 cohort study of patients with cancer showed that *CDH1* germline pathogenic or likely pathogenic variants are enriched in patients with diffuse gastric cancer and lobular breast cancer.[15] The high number of patients with breast cancer carrying *CDH1* pathogenic or likely pathogenic variants in the USA led Lerner and colleagues[16] to propose the use of clinical criteria for hereditary breast and ovarian cancer, in addition to the modified and less restrictive 2020 criteria for HDGC (also known as Yale criteria), to select patients for *CDH1* testing. Although this approach increased the sensitivity of the criteria, no data were provided regarding their specificity or positive predictive value.[16]

CDH1 variants of unknown significance persist as a clinical challenge because of scarce genotype–phenotype data, and missense variants, the largest class of coding variants of unknown significance, have been claimed to preferentially predispose individuals to HLBC.[8] In parallel, endoscopic evidence of gastric cancer was found in more than 90% of families with HLBC, who carry *CDH1* pathogenic or likely pathogenic variants[17], raising questions about whether HLBC is an HDGC-independent entity. Although colorectal cancer has been considered as a possible *CDH1*-associated phenotype in some studies,[15,18] others claim it is rare among carriers of *CDH1* pathogenic or likely pathogenic variants.[9]

Genotype–phenotype association studies, particularly genotype-first studies, are expected to answer some of the abovementioned challenges. In this study, we aimed to conduct a genotype–phenotype analysis in carriers of rare *CDH1* variants and their relatives, to compare the cancer landscape in families carrying pathogenic or likely pathogenic variants (considered together for the purpose of our analyses) or missense variants of unknown significance, to assess the frequency of families with lobular breast cancer among families carrying pathogenic or likely pathogenic variants, and to test the performance of novel lobular breast cancer-expanded criteria for the selection of patients for *CDH1* testing. We used the collaborative environment of the European Reference Network on Genetic Tumour Risk Syndromes (ERN GENTURIS) to run a multicentre genotype-first study to identify associations between rare *CDH1* germline variants and cancer phenotypes, in the largest dataset of *CDH1* rare variant carriers and their relatives, the first and most detailed to date, to the best of our knowledge.

Methods

Participants and data collection

Data collection started on Oct 1, 2018, and ended on Sept 20, 2022. *CDH1* variant-related data were obtained from diagnostic laboratories across Europe (29 institutions in ten European countries; figure 1A; appendix pp 2–4) and were the result of genetic testing with either *CDH1*-targeted sequencing, or panel, whole-exome, or whole-genome sequencing. Only variants in the *CDH1* locus were collected for the purpose of the current study. The presence of a rare single-nucleotide variant or copy-number variant in *CDH1*, herein classified as any variant with frequency below 0·01 in gnomAD, in a proband irrespective of clinical diagnosis or HDGC suspicion, drove the collection of aggregated molecular and clinical information from carrier probands and their relatives (unconfirmed carriers). The term unconfirmed carrier refers to two different scenarios: affected relatives who were not submitted for carrier testing, particularly in the case of variants of unknown significance and likely benign or benign variants; and affected relatives from families carrying pathogenic or likely pathogenic variants who were generally submitted for carrier testing, but for whom the carrier status is generally unknown to this study. The cohort's general clinical features (appendix pp 2–3), information on participating institutions (appendix p 4), and molecular and clinical information indexed to probands from each family (appendix pp 5–46) were collected.

All patients signed an informed consent form for germline testing, either mentioning future research related to their susceptibility syndrome or allowing sample biobanking for future research. In most institutions, approval by ethics committees was waived, as this is a retrospective study; in the remaining institutions, local ethics approval was granted. The whole project received appraisal N19/CECRI/2022 from the Committee for Ethical and Responsible Conduct of Research of the study's leading institution, Instituto de Investigação e Inovação em Saúde, University of Porto, Portugal.

Clinical and variant data curation

Family history was revisited and compliance with the 2015 criteria,[1] 2020 criteria,[2] or Yale criteria[16] for HDGC (appendix pp 5–46) was registered. For the purpose of comparative analysis, a family with HDGC was defined as any family fulfilling the 2015 clinical criteria for HDGC. Phenotypes were collected from probands and relatives from families carrying *CDH1* germline variants.

Human Genetics, GIGA Institute, University of Liège, Liège, Belgium (V Bours MD PhD); **Center of Genetics, University Hospital, Liège, Belgium** (V Bours); **Department of Human Genetics, University of Leuven, Leuven, Belgium** (H Brems PhD, G Michils PhD); **Hereditary Cancer Programme, Catalan Institute of Oncology, Bellvitge Institute for Biomedical Research and Girona Biomedical Research Institute, Barcelona-Girona, Spain** (J Brunet MD PhD); **Laboratorio di Bioscienze, IRCCS Istituto Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori", Meldola, Italy** (D Calistri PhD, G Tedaldi PhD); **Hereditary Cancer Program, Catalan Institute of Oncology, Bellvitge Institute for Biomedical Research, Barcelona, Spain** (G Capellá MD PhD, C Lázaro PhD); **Centro de Investigación Biomédica en Red de Cáncer (CIBERONC), Madrid, Spain** (G Capellá, C Lázaro); **Oncology Service, Biocruces Bizkaia Health Research Institute, Cruces University Hospital, Cruces-Barakaldo, Bizkaia, Spain** (S Carrera MD); **Center of Human Genetics, IPG, Gosselies, Belgium** (K Dahan MD PhD, D Feret MSc); **Clinical Genetics Department, University Hospital of Ghent, Ghent, Belgium** (R de Putter MD); **Molecular Diagnostics Laboratory, Fundación Rioja Salud, Logroño, Spain** (E Domínguez-Garrido PhD); **CNC—Center for Neuroscience and Cell Biology, University of Coimbra, Coimbra, Portugal** (C Egas PhD); **Division of Evolution and Genomic Sciences, University of Manchester, Manchester, UK** (Prof D G Evans MD, E R Woodward MD PhD); **Manchester Centre for Genomic Medicine, Manchester, UK** (Prof D G Evans, E R Woodward); **Department of Medical Genetics, National Institute for Health Research Cambridge Biomedical Research Centre** (E Fewings PhD, Prof M Tischkowitz MD PhD), **and Early Cancer Institute** (Prof R C Fitzgerald MD), **University of Cambridge, Cambridge, UK; Genetics Unit, Biocruces Bizkaia Health**

Research Institute, Basurto University Hospital, Bilbao, Bizkaia, Spain (M Garcia-Barcina MD PhD, S Merino MD); Sezione di Medicina Genomica, Dipartimento di Scienze della Vita e Salute Pubblica, Università Cattolica del Sacro Cuore, Rome, Italy (Prof M Genuardi MD); UOC Genetica Medica, Dipartimento di Scienze di Laboratorio e Infettivologiche, Fondazione Policlinico Universitario A Gemelli IRCCS, Rome, Italy (Prof M Genuardi); Institute for Clinical Genetics, University Hospital Carl Gustav Carus and Faculty of Medicine, Technische Universität Dresden, Dresden, Germany (K Hackmann PhD, Prof E Schröck MD); National Center for Tumor Diseases, Dresden, Germany: German Cancer Research Center, Heidelberg, Germany (K Hackmann, Prof E Schröck); Helmholtz-Zentrum Dresden-Rossendorf, Dresden, Germany (K Hackmann, Prof E Schröck); German Cancer Consortium, Dresden, Germany (K Hackmann, Prof E Schröck); Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany (Prof E Schröck); SouthWest Thames Regional Genetics Service, St George's University Hospitals NHS Foundation Trust, London, UK (H Hanson MD); Medizinische Klinik und Poliklinik IV, Klinikum der Universität München, Munich, Germany (Prof E Holinski-Feder MD, V Steinke-Lange MD); Medizinisch Genetisches Zentrum, Munich, Germany (Prof E Holinski-Feder, V Steinke-Lange); Department of Human Genetics, Radboud University Medical Center, Nijmegen, Netherlands (Prof M J L Ligtenberg MD, Prof N Hoogerbrugge MD PhD); Department of Pathology (Prof M J L Ligtenberg) and Radboud Institute of Molecular Life Sciences (Prof M J L Ligtenberg, Prof N Hoogerbrugge), Radboud University Medical Center, Nijmegen, Netherlands; Genetics Service, Biocruces Bizkaia Health Research Institute, Cruces University Hospital, Cruces-Barakaldo, Bizkaia, Spain

Some individuals contributed more than one phenotype to the analysis if they had more than one phenotype. Information on age of cancer onset and gastric and breast cancer histology was collected (appendix pp 2–3).

Diffuse gastric cancer, lobular breast cancer, and gastric cancer were considered HDGC-related phenotypes. "Other" refers to all phenotypes that differed from HDGC-related phenotypes, breast cancer of unknown histotype (herein referred to as "non-specified breast cancer" or "breast cancer"), ductal breast cancer, ovarian cancer, and colorectal cancer (appendix p 3).

In the context of families, "gastric cancer only" refers to families enriched in gastric cancer, without breast cancer of any type, and could include cancers other than gastric or breast cancer; "breast cancer only" refers to families enriched in breast cancer, without gastric cancer of any type, but could include cancers other than gastric or breast cancer; "gastric cancer and breast cancer" refers to families with both gastric cancer and breast cancer, and could also include cancers other than gastric or breast cancer. "Other cancer" refers to families presenting exclusively cancers other than gastric or breast cancer.

Variant nomenclature was standardised according to the Human Genome Variation Society (HGVS) guidelines.[19] Variants were categorised into groups: coding truncating (including nonsense, frameshift, start loss, large deletion, large duplication, and splice-site variants); coding non-truncating (including missense, in-frame insertion, and in-frame exon skipping variants); non-coding (regulatory; including intronic large deletions, 5□ untranslated region single-nucleotide variants, 3□ untranslated region single-nucleotide variants, and intronic single-nucleotide variants); and synonymous (referring to single-nucleotide variants unrelated to splicing), according to the Ensembl Variant Effect Predictor (VEP;[20] appendix p 47). Variants located until within two bp from the start or end of exons and within five bp from the start or end of introns were analysed with NetGene2 version 2.4 splice-site prediction software,[21] and classified as splice-site variants if a positive impact on splicing was predicted. Recurrence was considered if the same variant occurred in three or more families. With use of the American College of Medical Genetics and Genomics and Association for Molecular Pathology *CDH1* guidelines (version 2), all variants were classified for their actionability in HDGC as a pathogenic or likely pathogenic (herein analysed jointly and referred to in short as PV/LPV), a variant of unknown significance, or a likely benign or benign variant (herein analysed jointly and referred to in short as LBV/BV; figure 1A),[6] and were confirmed with ClinGen *CDH1* Variant Curation Expert Panel updates. In case a patient presented two *CDH1* germline variants, the most damaging variant was considered for genotype–phenotype analysis (appendix p 48). In the current study, only variants classified as either pathogenic or likely pathogenic were considered actionable, as their identification in probands triggers cascade testing in relatives and risk reduction measures in carrier probands and carrier relatives. Because the 2020 guidelines for HDGC recommend at least 2 years of gastric or breast surveillance, or both, for probands carrying a variant of unknown significance, but not for relatives, this type of variant was considered non-actionable in this study when assessing the performance of different HDGC clinical criteria.

The relative frequency of phenotypes, age of onset, and compliance with HDGC criteria were used in different comparisons involving families carrying different types of *CDH1* variants; families bearing recurrent missense variants of unknown significance or bearing recurrent truncating PV/LPVs; and families with gastric cancer only, breast cancer only, or gastric cancer and breast cancer.

The frequency of families carrying PV/LPVs among those complying with different HDGC criteria sets was used to evaluate the pick-up rate and performance of the 2015 and 2020 HDGC criteria, Yale criteria, and the lobular breast cancer-expanded criteria.

Statistical analysis

Continuous variables are presented as mean (SD). Phenotypes without an associated age of onset were excluded from statistical analyses. Multiple comparisons of age of disease onset in carrier probands and their relatives and different types of variants (PV/LPV, missense variant of unknown significance, or missense LBV/BV), were done using a Kruskal-Wallis test with Bonferroni correction. Categorical variables (phenotypes, compliance with HDGC criteria, and variant classification) are presented as count (%) and were compared with use of χ^2 tests. This approach was used to compare variant classes between families fulfilling or not fulfilling clinical criteria, and for the analysis of phenotypic distributions in families carrying different types of variants. Age of onset per phenotype was compared with use of an independent-samples Student's t test. Multivariable logistic regression models using the enter method, adjusting for phenotypes and age of onset to remove sampling bias in terms of age of onset, were used to identify independent predictors for distinct variant types (PV/LPVs, and missense variant of unknown significance). To evaluate the likelihood of each set of criteria producing false positive and false negative results, we calculated the sensitivity, specificity, positive predictive value, and negative predictive value, with corresponding 95% CIs, using R (version 4.1.3) with the library epiR (version 2.0.50),[23] whereas the Youden index (J) used to estimate the criteria's discriminating power, and the corresponding 95% CIs, were computed with the library ThresholdROC.[24] The term "lobular breast cancer-expanded criteria" represents a combination of all the 2020 HDGC criteria plus the new lobular-centred criteria herein proposed. To compare performance in terms of true positive or true negative detection for different sets of clinical criteria (2015 and 2020 HDGC criteria, Yale

criteria, and lobular breast cancer-expanded criteria), equivalence tests and corresponding 95% CIs were computed in R by directly implementing formulae. The equivalence limit (δ) was set to 0·12.[25] Areas under the receiver operating characteristic curves (AUCs) from different sets of clinical HDGC criteria (2015 and 2020 HDGC criteria, Yale criteria, and lobular breast cancer-expanded criteria) were compared against each other with *Z* tests, using the IBM SPSS Statistics for Windows (version 28.0.0). For a 95% CI, a *Z* score of 1·96 was used. Statistical significance was set at p<0·05 for all statistical tests.

Role of the funding source

The funders of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

398 different rare *CDH1* germline variants were reported in 854 families (figure 1A; appendix pp 5–46). From 854 carrier probands and 1021 relatives, age of onset was available for 1495 patients' phenotypes and ranged from 1–93 years. Altogether, 1971 phenotypes were collected, of which 460 were gastric and breast cancers with available histology (figure 1A; appendix pp 2–3).

Among the 854 families, coding truncating variants occurred in 184 (22%), missense variants in 316 (37%), and other variant types in 354 families (41%; appendix p 47). In terms of variant clinical actionability,[6] 176 (21%) families carried PV/LPVs (100 different truncating variants); 169 (20%) families carried missense variants of unknown significance (109 different missense variants); 147 (17%) families carried missense LBV/BVs (29 different missense variants); 182 (21%) families carried variants of unknown significance other than missense (122 different variants); and 180 (21%) families carried LBV/BVs other than missense (38 different variants) (appendix pp 47, 49–51). Probands carrying missense variants of unknown significance or missense LBV/BVs, and their relatives, were all significantly older than probands carrying PV/LPVs and their relatives (p<0·05 in all comparisons), whereas probands carrying missense variants of unknown significance and missense LBV/BVs and their relatives had similar ages of onset (appendix p 52).

CDH1 PV/LPVs included both single-nucleotide variants and copy-number variants and were distributed across the whole *CDH1* locus, targeting all exons, as were all other variant types (appendix pp 49–51). Only 15 PV/LPVs (11 single-nucleotide variants and four copy-number variants) were recurrent, and were found in 81 (46%) of 176 families (appendix p 49). The most frequent PV/LPV single-nucleotide variant, which occurred in 13 families, was the c.1901C>T (p.Ala634ProfsTer7) splice-site variant, and the most frequent copy-number variant, occurring in ten families, was the exon 1–2 deletion NC_000016.10:g.(?_68737292)_(68738411_?)del (appendix p 49).

182 (21%) families fulfilled the 2015 clinical criteria for HDGC[1] and 672 (79%) did not. The cohort of families fulfilling the 2015 criteria was significantly enriched for PV/LPVs compared with the cohort of families not fulfilling the criteria (136 [75%] of 182 *vs* 40 [6%] of 672; p<0·0001; figure 1B; appendix p 53). By contrast, missense variants of unknown significance were enriched in the cohort of families not fulfilling the criteria (155 [23%] of 672; these 155 families are part of 159 carrying coding non-truncating variants in families not fulfilling criteria; figure 1B; appendix p 54) compared with the cohort of families that fulfilled the criteria (14 families carrying coding non-truncating variants [8%] of 182 families fulfilling criteria; p<0·0001; figure 1B; appendix p 54).

Early onset diffuse gastric cancer, lobular breast cancer, and gastric cancer, the classic HDGC-related phenotypes, were the most frequent among 176 PV/LPV carrier families, accounting for 459 (73%) of 631 phenotypes in probands and relatives (figure 2A; appendix p 55). HDGC-related phenotypes were repeatedly observed only among families bearing recurrent PV/LPVs, but not among families bearing recurrent missense variants of unknown significance (appendix p 56).

Among 176 families carrying PV/LPVs, 362 (92%) of 392 diffuse gastric cancer or gastric cancer cases occurred in families fulfilling the 2015 criteria for HDGC, whereas only 32 (48%) of 67 cases of lobular breast cancer occurred in families fulfilling the criteria (figure 2A; appendix p 55). Among families carrying missense variants of unknown significance, 32 (65%) of 49 diffuse gastric cancer or gastric cancer cases occurred in families fulfilling the HDGC criteria, and among families carrying missense LBV/BVs, this proportion was 18 (38%) of 48. Lobular breast cancers were extremely rare among families carrying missense variants of unknown significance or LBV/BVs (figure 2B; appendix pp 55, 57).

Among families carrying PV/LPVs, 50 (56%) of 90 non-specified breast cancer cases (mainly in relatives) occurred in families fulfilling the criteria (figure 2A; appendix pp 55, 58). By contrast, among families carrying missense variants of unknown significance, only one (<1%) of 219 cases of non-specified breast cancer occurred in families fulfilling the HDGC criteria, and among families carrying missense LBV/BVs, this proportion was three (2%) of 150 cases (figure 2B; appendix pp 55, 57–58).

Overall, in families fulfilling HDGC criteria, the frequency of either gastric cancers altogether or non-specified breast cancer in relatives was significantly higher in carriers of PV/LPVs than in carriers of missense variants of unknown significance or carriers of LBV/BVs (p<0·0001 in all comparisons; figure 2A, B; appendix pp 55, 57–58).

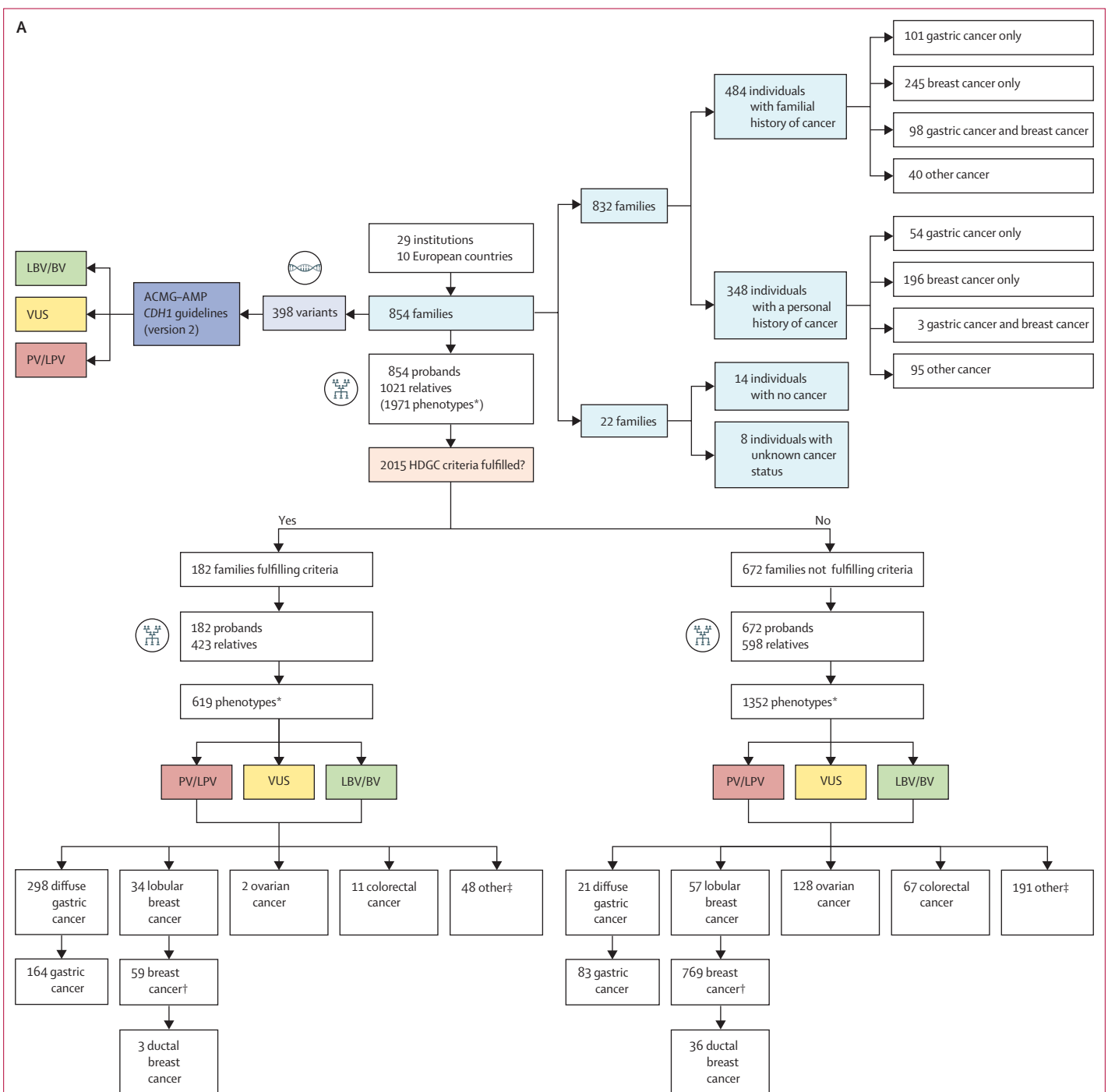
Multivariate logistic regression was used to evaluate the likelihood of occurrence of HDGC-related cancers in PV/LPV carrier families as compared with their occurrence in families carrying missense variants

(C Martínez-Bouzas PhD, M I Tejada PhD); **Unidad de Medicina Genómica y Pediatría, Clínica Universidad de Navarra, Programa de Tumores Sólidos, Centro de Investigación Médica Aplicada, Instituto de Investigación Sanitaria de Navarra, Pamplona, Navarra, Spain** (Prof A Patiño-García PhD); **Department of Biology and Biotechnology, University of Pavia, Pavia, Italy** (Prof G N Ranzani BSc PhD); **GenoMed—Diagnósticos de Medicina Molecular, Lisbon, Portugal** (I Silva MSc, C Silveira MSc); **Molecular Genetics Laboratory, Elche University Hospital, Elche, Spain** (J L Soto BSc PhD); **European Reference Network on Genetic Tumour Risk Syndromes (ERN GENTURIS), Porto, Portugal** (Prof S Castedo, Prof F Carneiro, M R Teixeira, Prof C Oliveira); **ERN GENTURIS, Bonn, Germany** (Prof S Aretz, R Hüneburg, I Spier); **ERN GENTURIS, Stockholm, Sweden** (S Bajalica-Lagercrantz, K Lagerstedt-Robinson); **ERN GENTURIS, Barcelona, Spain** (J Balmaña, J Brunet, G Capellá, C Lázaro); **ERN GENTURIS, Ljubljana, Slovenia** (A Blatnik, M Krajc); **ERN GENTURIS, Liège, Belgium** (V Bours); **ERN GENTURIS, Paris, France** (C Colas); **ERN GENTURIS, Ghent, Belgium** (R de Putter); **ERN GENTURIS, Rome, Italy** (Prof M Genuardi); **ERN GENTURIS, Munich, Germany** (Prof E Holinski-Feder, V Steinke-Lange); **ERN GENTURIS, Nijmegen, Netherlands** (Prof M J L Ligtenberg, Prof N Hoogerbrugge); **ERN GENTURIS, Dresden, Germany** (Prof E Schröck)

Correspondence to:
Prof Carla Oliveira, Instituto de Investigação e Inovação em Saúde (i3S), University of Porto, Porto 4200-135, Portugal
carlaol@i3s.up.pt

See **Online** for appendix

of unknown significance (figure 2C). This analysis highlighted positive associations between the classic HDGC-related cancers (lobular breast cancer, diffuse gastric cancer, and gastric cancer) and the presence of PV/LPVs, in contrast to missense variants of unknown significance. Lobular breast cancer (a phenotype present in 47 (27%) of 176 PV/LPV carrier families and in four (2%) of 169 families carrying missense variants of unknown significance) had the greatest positive association (odds ratio [OR] 12·39 [95% CI 2·66–57·74], p=0·0014), followed by diffuse gastric cancer (8·00 [2·18–29·39], p=0·0017), and gastric cancer (7·81 [2·03–29·96], p=0·0027; figure 2C; appendix p 59). Cancers, such as non-specified breast cancer, ovarian



(Figure 1 continues on next page)

cancer, colorectal cancer, and all other cancers together, were not positively associated with the presence of a PV/LPV (figure 2C). The likelihood of ovarian cancer occurring in families carrying missense variants of unknown significance was significantly higher than in families carrying PV/LPVs (OR 0·12 [95% CI 0·02–0·71], p=0·019; figure 2C; appendix p 59). In the multivariate analysis, age of cancer onset did not differ with regard to cancer type or between families carrying PV/LPVs and missense variants of unknown significance (figure 2C; appendix p 59).

We next analysed the evolution of HDGC clinical criteria, considering the above genotype–phenotype associations. From 176 PV/LPV carrier families, 40 (23%) did not meet the 2015 criteria. Of these families, 11 (6%) fulfilled the 2020 HDGC criteria that considers eligible for *CDH1* testing isolated cases of diffuse gastric cancer in people younger than 50 years and those families with at least two cases of lobular breast cancer in family members younger than 50 years (appendix p 60). From the remaining 29 PV/LPV families who did not meet the 2020 HDGC criteria, only four families did not show breast cancer involvement (figure 3A; appendix p 61). 18 (62%) of 29 were families with lobular breast cancer with or without gastric cancer (figure 3B). Detailed analysis of family cancer history revealed that these 18 families presented at least one confirmed case of lobular breast cancer, either isolated before age 55 years (four of 18), in families with additional non-specified breast cancers and no gastric involvement (seven out of 18), or in families with history of non-specified breast cancer and gastric cancer (seven out of 18). This analysis resulted in a proposal of three new lobular breast cancer-centred criteria to be added to the 2020 HDGC criteria, and herein called lobular breast cancer-expanded criteria (figure 3B).

For 11 of 29 families who did not meet the 2020 HDGC criteria, representing 6% of 176 *CDH1* PV/LPV carrier families, no possible criteria could be envisioned due to very scarce or missing clinical information (figure 3B; appendix p 61).

We next stratified the full PV/LPV carrier cohort into family subgroups of gastric cancer only, gastric cancer and breast cancer, and breast cancer only, and plotted the relative phenotype frequency and distribution in each family subgroup. This approach allowed evaluation of the contribution of the lobular breast cancer-centred criteria in the different subgroups (figure 4; appendix pp 62–68).[1,2]

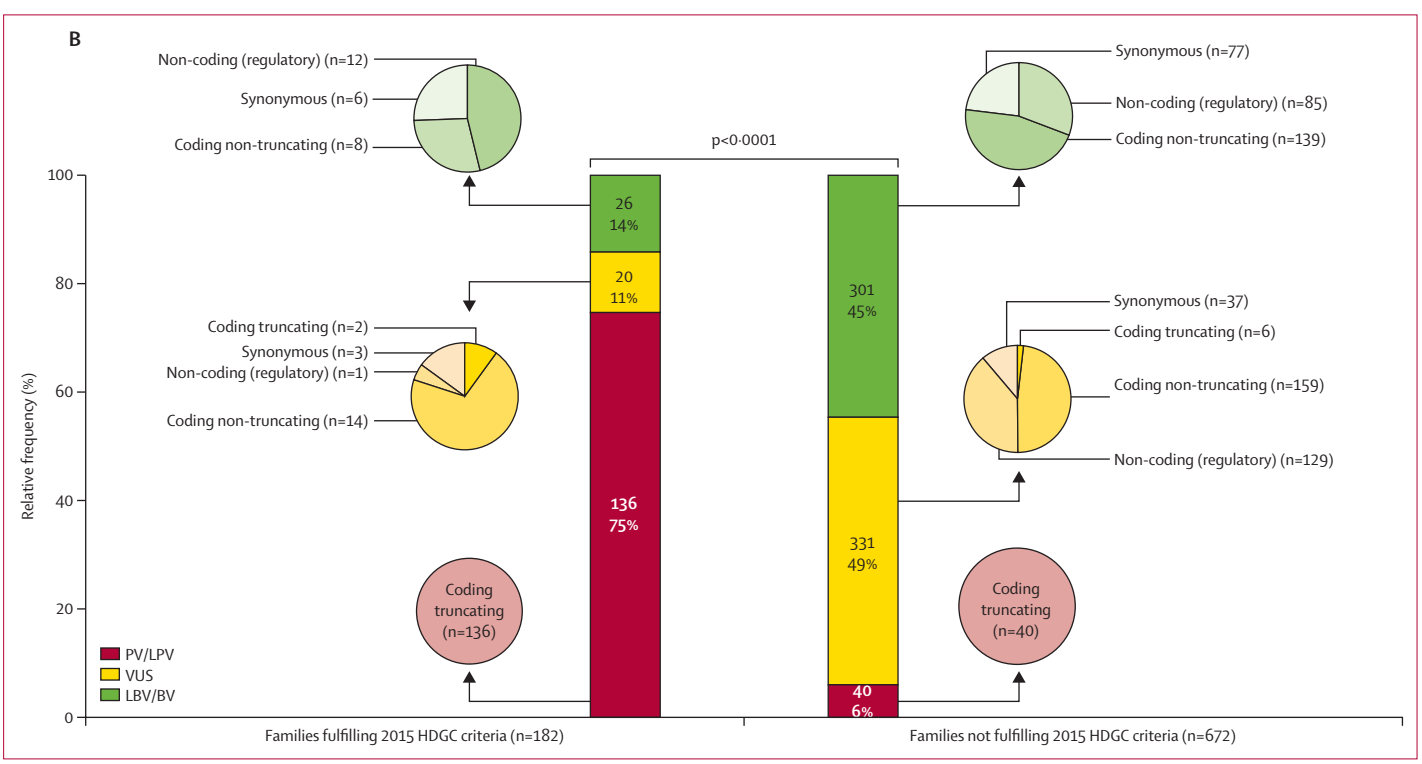
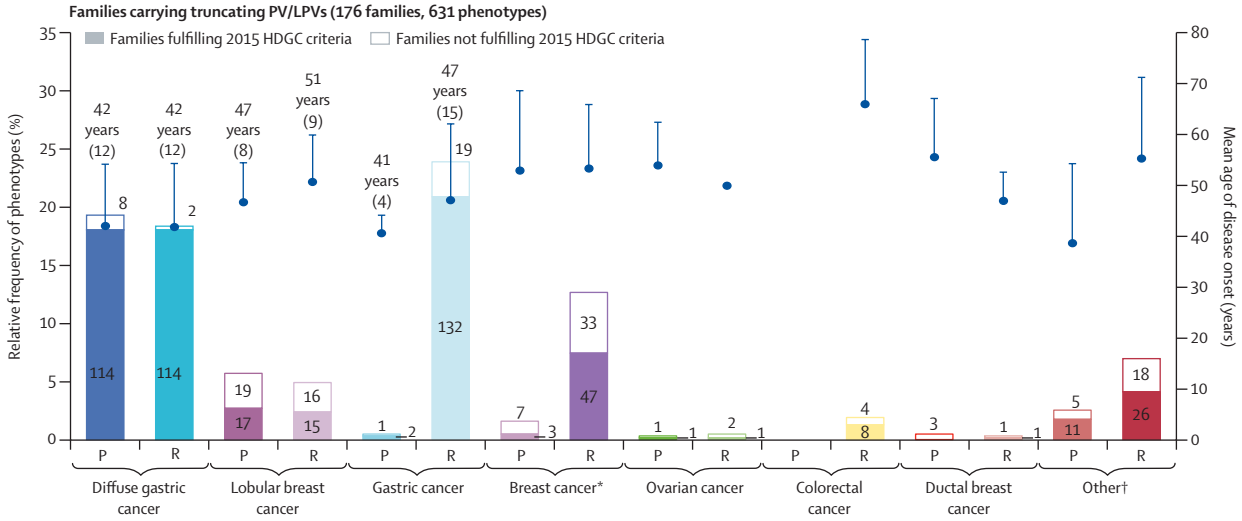


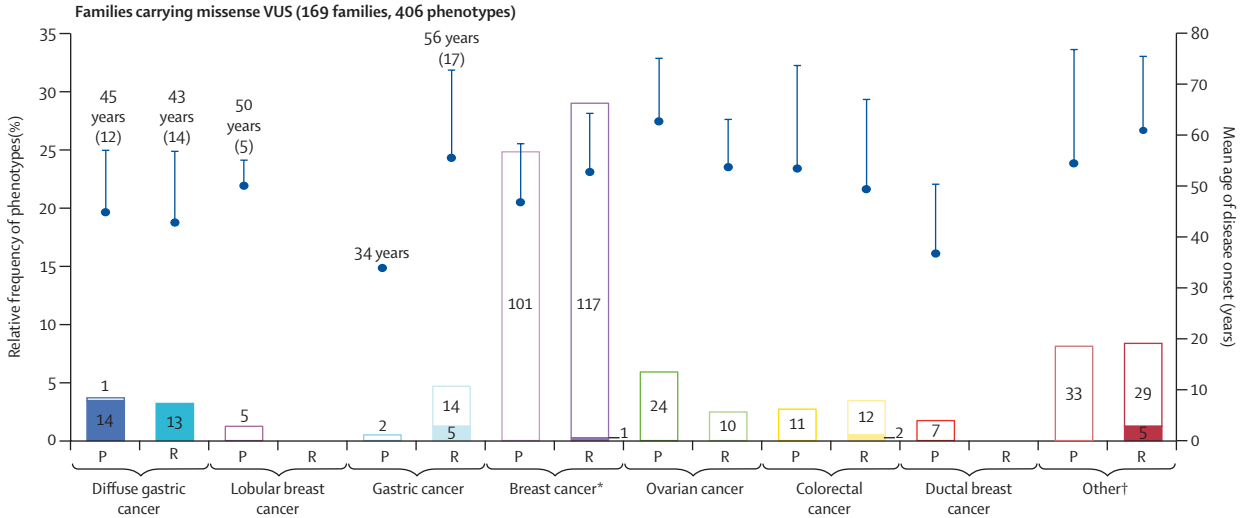
Figure 1: **Flowchart of the genotype–phenotype study and molecular and clinical features of the cohort of families with rare germline *CDH1* variants**
(A) Flowchart of the genotype–phenotype study. (B) Relative frequency of variant groups and clinical classification in families fulfilling or not fulfilling the 2015 HDGC criteria. A χ^2 test was used to compare the frequency of variant clinical classifications between families fulfilling and not fulfilling the 2015 HDGC criteria (appendix p 53). 14 of 14 coding non-truncating variants of unknown significance were missense variants occurring in families fulfilling the 2015 HDGC criteria. 155 of 159 coding non-truncating variants of unknown significance were missense variants occurring in families not fulfilling the 2015 HDGC criteria. Details on variant type and correspondence with clinical criteria and clinical classification are depicted in appendix (pp 47, 54). ACMG–AMP=American College of Medical Genetics and Genomics and Association for Molecular Pathology. HDGC=hereditary diffuse gastric cancer. LBV/BV=likely benign or benign variant. PV/LPV=pathogenic or likely pathogenic variant. VUS=variant of unknown significance. *Probands could contribute more than one phenotype. †Breast cancer of unknown histotype. ‡Other phenotypes not including gastric cancer, breast cancer, colorectal cancer, or ovarian cancer (detailed in appendix p 3).

A



Number of cases																	Overall
Total	122	116	36	31	3	151	10	80	2	3	0	12	3	2	16	44	631
Fulfilling criteria	114	114	17	15	2	132	3	47	1	1	0	8	0	1	11	26	492
Not fulfilling criteria	8	2	19	16	1	19	7	33	1	2	0	4	3	1	5	18	139
With available age of onset	114	110	33	29	3	122	9	67	2	1	0	10	3	2	6	31	542

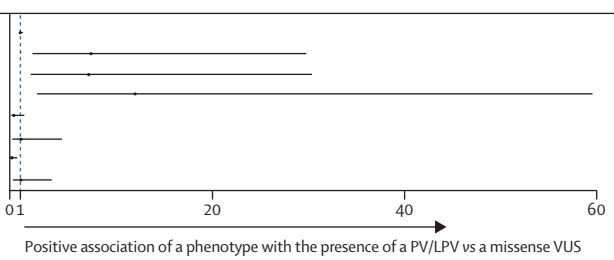
B



Number of cases																	Overall
Total	15	13	5	0	2	19	101	118	24	10	11	14	7	0	33	34	406
Fulfilling criteria	14	13	0	0	0	5	0	1	0	0	0	2	0	0	0	5	40
Not fulfilling criteria	1	0	5	0	2	14	101	117	24	10	11	12	7	0	33	29	366
With available age of onset	15	13	5	0	1	15	93	101	19	5	7	2	5	0	17	16	314

C

	Prevalence of variant among individuals with phenotype		Number of individuals without phenotype	Odds ratio (95% CI)	p value
	PV/LPV	Missense VUS			
Age (older *vs* younger; n=856)	..	..		1·00 (0·99–1·01)	0·97
Diffuse gastric cancer	224/252 (89%)	28/252 (11%)	604	8·00 (2·18–29·39)	**0·0017**
Gastric cancer	125/141 (89%)	16/141 (11%)	715	7·81 (2·03–29·96)	**0·0027**
Lobular breast cancer	62/67 (93%)	5/67 (7%)	789	12·39 (2·66–57·74)	**0·0014**
Breast cancer	76/270 (28%)	194/270 (72%)	586	0·39 (0·11–1·39)	0·15
Colorectal cancer	10/19 (53%)	9/19 (47%)	837	1·11 (0·24–5·19)	0·90
Ovarian cancer	3/27 (11%)	24/27 (89%)	829	0·12 (0·02–0·71)	0·019
Other	37/70 (53%)	33/70 (47%)	786	1·12 (0·29–4·24)	0·87



174 of 176 families fit into the following subgroups: 91 families with gastric cancer only (representing 52% of all families carrying PV/LPVs) and 255 phenotypes; 67 families with gastric cancer and breast cancer (representing 38% of all families carrying PV/LPVs) and 331 phenotypes, and; 16 families with breast cancer only (representing 9% of all families carrying PV/LPVs) and 43 phenotypes (figure 4; appendix pp 62–68). In families with gastric cancer only, 230 (90%)of 255 phenotypes are diffuse gastric cancer and gastric cancer. Of the 91 of families with gastric cancer only, 89 (98%) fulfilled the 2015 or 2020 HDGC criteria (figure 4A; appendix pp 62–68).[2] In families with gastric cancer and breast cancer, 162 (49%) of 331 phenotypes were diffuse gastric cancer and gastric cancer, 48 (15%) were lobular breast cancer, and 78 (24%) were non-specified breast cancer. 56 (84%) of 67 families fulfilled either the 2015 or 2020 HDGC criteria, and lobular breast cancer-centred criteria contributed additional seven (10%) of 67 families (figure 4B; appendix pp 62–68). From the 16 families with breast cancer only (figure 4C), 12 (7% of 176 families carrying PV/LPVs) presented lobular breast cancer only. In these 16 families, 19 (44%) of 43 phenotypes were lobular breast cancer, and 12 (28%) unspecified breast cancer. Two (13%) of 16 families fulfilled either the 2015 or 2020 HDGC criteria, and lobular breast cancer-centred criteria contributed additional 11 (69%) of 16 families (figure 4; appendix pp 62–68).

We next analysed whether specific PV/LPVs were more frequent or exclusive of subgroups of gastric cancer only, gastric cancer and breast cancer, or breast cancer only families (appendix pp 64–69). A single variant (c.1565+2dup) was recurrent and consistently associated with gastric cancer only families, and no variant was particularly associated with gastric cancer and breast cancer families (appendix pp 64–69). Variants occurring exclusively in families with breast cancer only occurred in a single family each (appendix pp 64–69).

Figure 2: **Distribution of phenotypes and cancer average age of onset, in probands and relatives from germline *CDH1* variant carrier families, and multivariable logistic regression models for the association of cancer phenotypes with PV/LPV and missense-variant of unknown significance**
Relative frequency of phenotypes, phenotype distribution, number of cases, and average age of onset in probands and relatives in families carrying PVs or LPVs molecularly classified as truncating (A) and in families carrying variants of unknown significance molecularly classified as missense (B) (data further detailed in appendix p 55). The left axis shows the relative frequency (%) of each phenotype out of the total number of phenotypes. Mean age of onset (SD) is shown for HDGC-associated phenotypes. Detailed data on the number of cases and average age of onset for all phenotypes are provided in the appendix (p 55). (C) Forest plot showing the results of a multivariable logistic regression model for the association of each phenotype with the presence of a PV/LPV *vs* a missense VUS in 856 patients with available age of disease onset and clinical phenotype. Probands could contribute more than one phenotype. HDGC=hereditary diffuse gastric cancer. PV/LPV=pathogenic or likely pathogenic variant. P=probands. R=relatives. VUS=variants of unknown significance. *Breast cancer of unknown histotype. †Other phenotypes not including gastric cancer, breast cancer, colorectal cancer, or ovarian cancer (detailed in appendix p 3).

We next plotted a summary of the 854 families, bearing variants classified according to the American College of Medical Genetics and Genomics and Association for Molecular Pathology *CDH1* variant classification (version 2), against the 2020 HDGC or the lobular breast cancer-expanded criteria to graphically represent improvement of clinical criteria in identifying PV/LPV carrier families (figure 5A). Further, the performance of different sets of HDGC criteria (2015, 2020, Yale, lobular breast cancer-expanded) was tested (figure 5B). For these tests, we considered PV/LPVs as actionable, and variants of unknown significance or LBV/BVs as non-actionable (appendix p 70).

The lobular breast cancer-expanded set of criteria showed better performance than the 2020 HDGC criteria (AUC 0·92 *vs* 0·88; *Z* score 3·54; p=0·0004), as well as better specificity (0·90 [95% CI 0·88–0·93] *vs* 0·34 [0·30–0·37]), and superiority in true negative detection (p<0·0001) compared with the Yale criteria[16] (specificity 0·34 [95% CI 0·30–0·37]; figure 5B, C; appendix pp 70–72), and equivalent true negative detection to that of the 2015 and 2020 HDGC criteria (p=0·059 and p=0·21, respectively; figure 5C). The lobular breast cancer-expanded criteria also showed better sensitivity (0·94 [95% CI 0·89–0·97]) than the 2015 HDGC criteria (0·77 [0·70-0·83] and the 2020 HDGC criteria (0·84 [0·77–0·89]), with superior true positive detection (p<0·0001 and p=0·0025, respectively), while maintaining high specificity (figure 5C; appendix pp 70–72).

Discussion

Our study shows that *CDH1* PV/LPVs are positively associated with HDGC-related phenotypes (lobular breast cancer, diffuse gastric cancer, and gastric cancer), and found no evidence of a positive association of these phenotypes with *CDH1* missense variants of unknown significance. As *CDH1* PV/LPVs occurred often in families with lobular breast cancer who did not meet the 2020 HDGC criteria, this study supports the expansion of HDGC criteria towards inclusion of additional lobular breast cancer-centred criteria.

The identification of *CDH1* pathogenic or likely pathogenic variants in patients with a relevant cancer history results in the application of clinical recommendations, including prophylactic removal of target organs in at-risk asymptomatic carriers. Therefore, it is crucial to know which *CDH1* variants are actionable and cause disease, whether variant classification is accurate, which organs are prone to cancer development in carriers, and if there are other histopathological cancer features specific to a particular *CDH1* variant, besides diffuse gastric cancer and lobular breast cancer. Answering these questions will improve clinical guidelines and ascertainment for genetic testing. Given that HDGC is a very rare tumour risk syndrome, ERN GENTURIS led a European multicentre study using a genotype-first

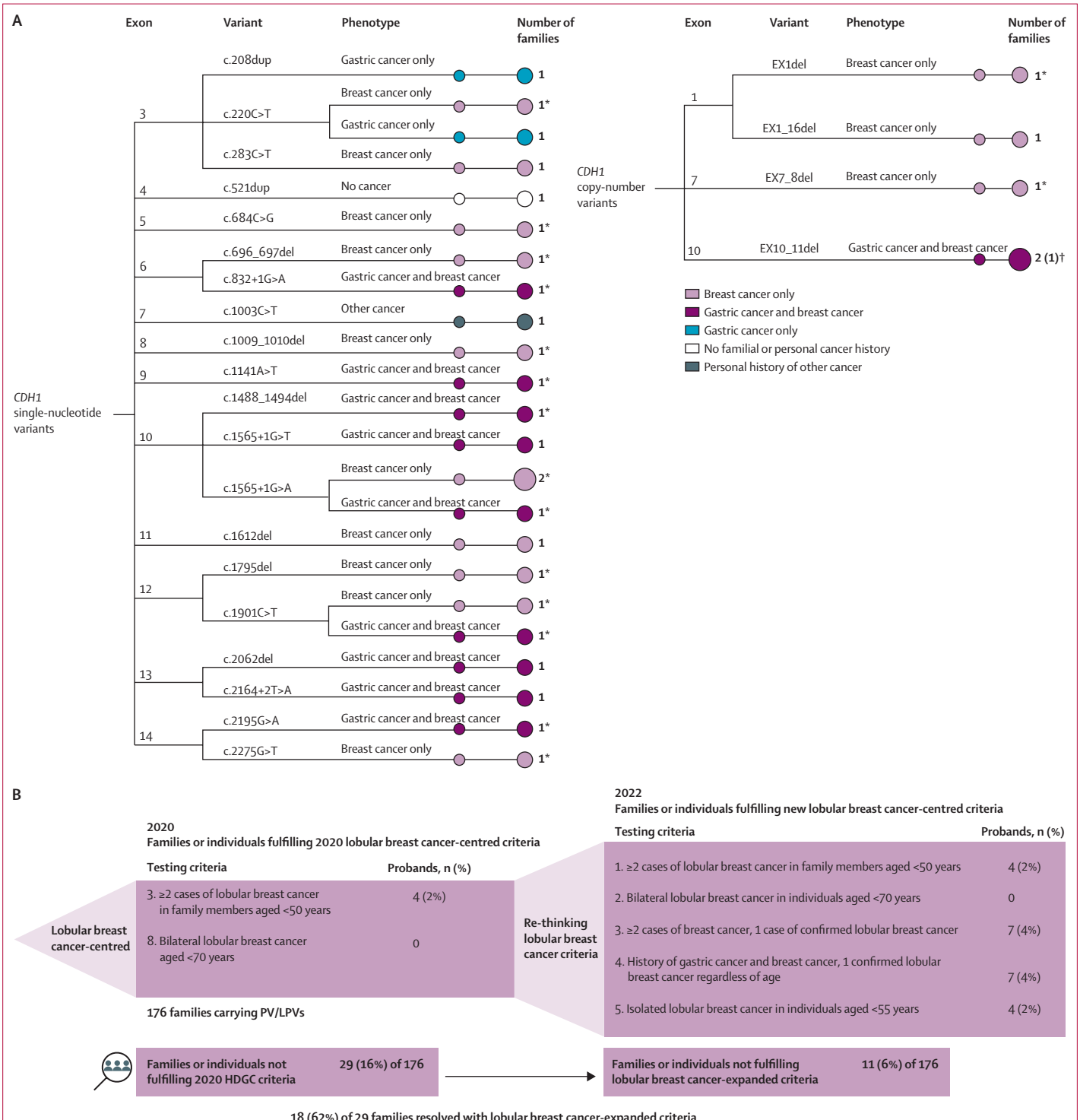


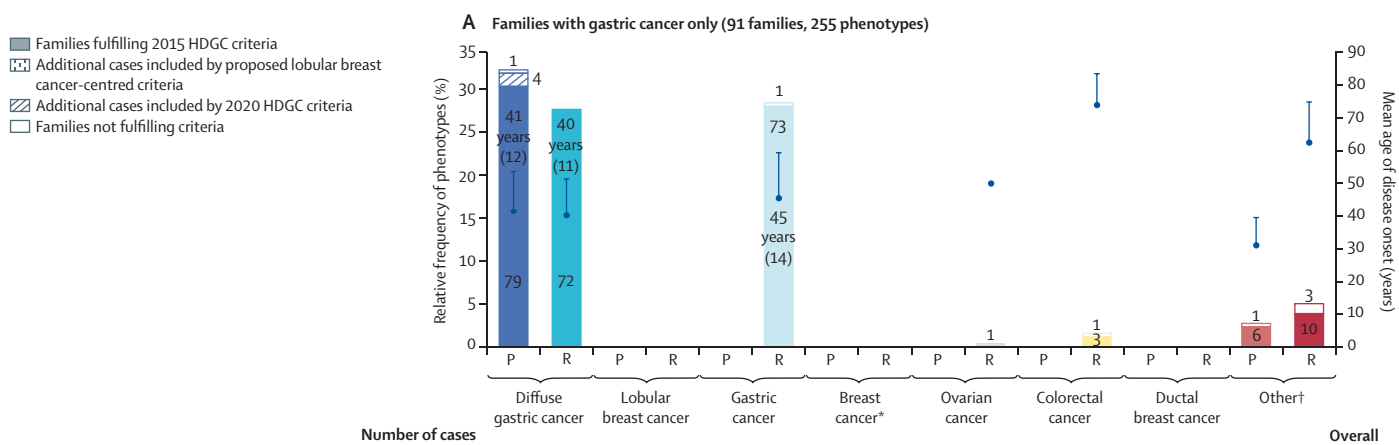
Figure 3: ***CDH1*** **PV/LPV carrier families and lobular breast cancer-related criteria**

(A) *CDH1* single-nucleotide and copy-number variants occurring in PV/LPV carrier families not fulfilling the 2020 HDGC criteria. PV/LPVs found in the cohort are shown by *CDH1* exon location. The size of each circle is proportional to the number of families found to carry a particular variant. (B) Proposed lobular breast cancer-expanded criteria and proportion of the 176 PV/LPV carrier families explained by these criteria. Lobular breast cancer-centred criteria 3, 4, and 5 are proposed as an addition to the 2020 HDGC criteria. dup=duplication. EX=exon. del=deletion. HDGC=hereditary diffuse gastric cancer. PV/LPV=pathogenic or likely pathogenic variant. *Families resolved by lobular breast cancer-centred criteria. †One of two families is resolved by the lobular breast cancer-centred criteria.

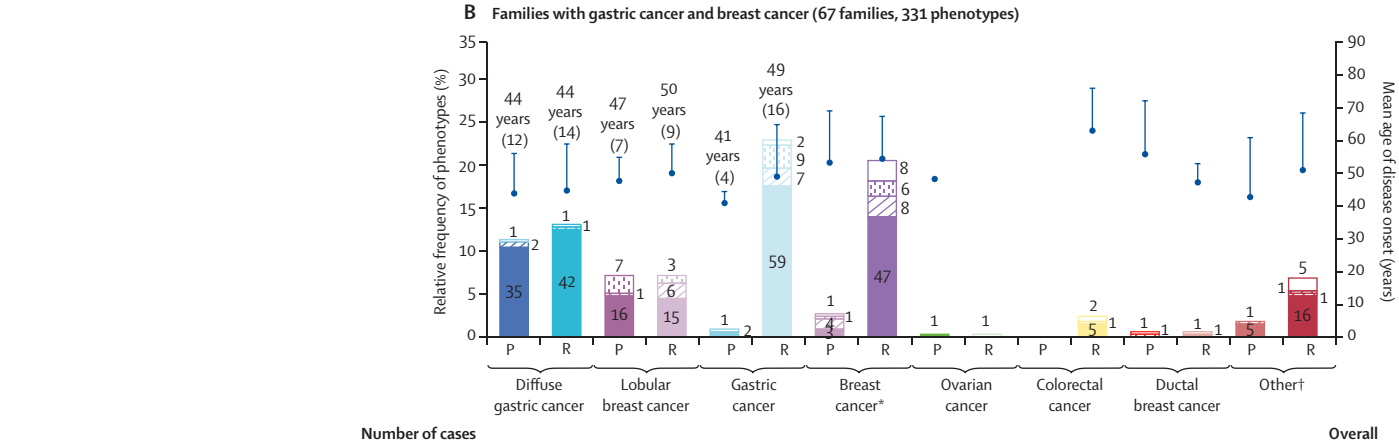
approach to address these questions, using the largest series to date (to our knowledge) of clinically characterised rare *CDH1* variant carriers and their relatives, who were tested irrespective of fulfilling clinical criteria.

Our data support germline truncating pathogenic or likely pathogenic variants as being major risk factors for HDGC-related cancers, and do not provide evidence of such an association for missense variants of unknown significance. These data confirm previous findings,[2,6,9] while challenging the role of most *CDH1* missense variants of unknown significance as being causative for HDGC.[8,27,28] Indeed, the limited number of missense variants, that were initially deposited in VEP and ClinVar as pathogenic or likely pathogenic variants, have later been shown to affect splicing, generating premature truncation.[11,29] We should not, however, discard the possibility that some purely missense variants can cause HDGC. Large databases and increasingly robust clinical and functional studies are needed to show whether some missense variants of unknown significance could cause HDGC.[6]

Our results consolidate the spectrum of cancers associated with *CDH1* PV/LPVs, which is enriched for early onset, histologically confirmed, diffuse gastric cancer and lobular breast cancer in probands, and gastric cancer in relatives, and occurs mainly in families fulfilling the 2020 HDGC clinical criteria.[2] Furthermore, to the best of our knowledge, we showed for the first time that lobular breast cancer had the greatest positive association with *CDH1* PV/LPVs, followed by

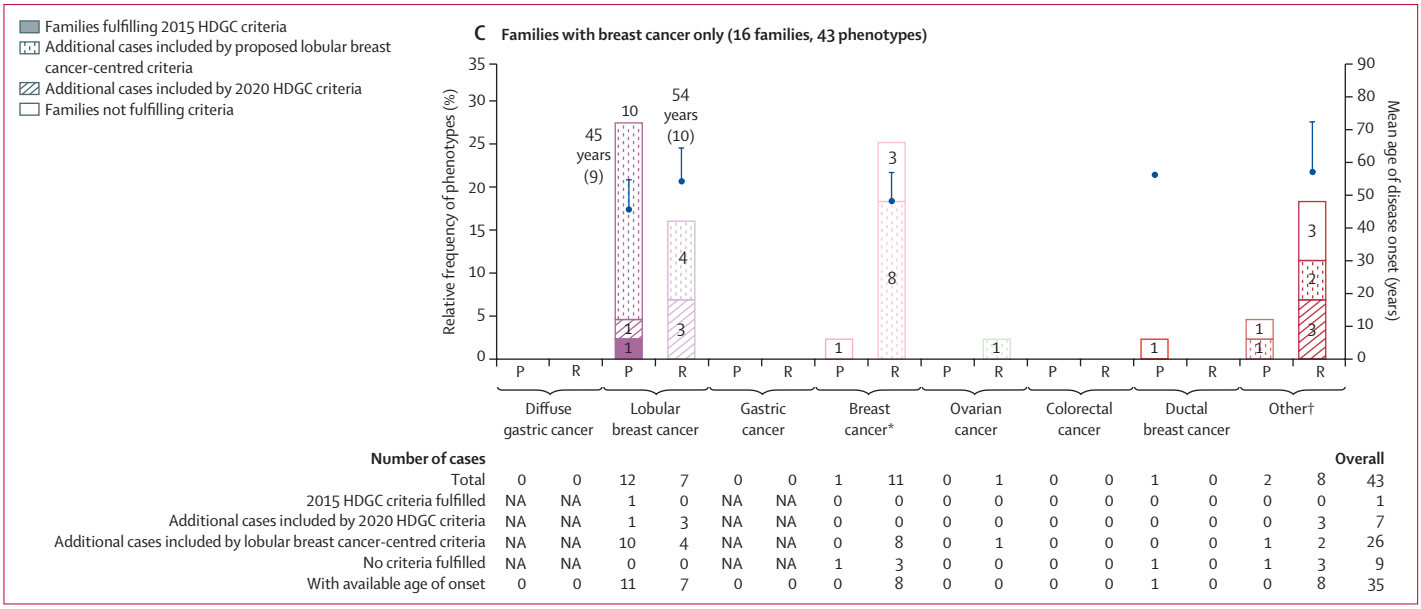


Number of cases	Diffuse gastric cancer		Lobular breast cancer		Gastric cancer		Breast cancer*		Ovarian cancer		Colorectal cancer		Ductal breast cancer		Other†		Overall
	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	
Total	84	72	0	0	0	74	0	0	0	1	0	4	0	0	7	13	255
2015 HDGC criteria fulfilled	79	72	NA	NA	0	73	NA	NA	0	1	0	3	NA	NA	6	10	244
Additional cases included by 2020 HDGC criteria	4	0	NA	NA	0	0	NA	NA	0	0	0	0	NA	NA	0	0	4
Additional cases included by lobular breast cancer-centred criteria	NA	NA	NA	NA	0	0	NA	NA	0	0	0	0	NA	NA	0	0	0
No criteria fulfilled	1	0	NA	NA	0	1	NA	NA	0	0	0	1	NA	NA	1	3	7
With available age of onset	80	67	0	0	0	57	0	0	0	1	0	3	0	0	2	8	218



Number of cases	Diffuse gastric cancer		Lobular breast cancer		Gastric cancer		Breast cancer*		Ovarian cancer		Colorectal cancer		Ductal breast cancer		Other†		Overall
	P	R	P	R	P	R	P	R	P	R	P	R	P	R	P	R	
Total	38	44	24	24	3	77	9	69	1	1	0	8	2	2	6	23	331
2015 HDGC criteria fulfilled	35	42	16	15	2	59	3	47	1	0	0	5	0	1	5	16	247
Additional cases included by 2020 HDGC criteria	2	1	1	6	0	7	4	8	0	0	0	0	0	0	1	1	31
Additional cases included by lobular breast cancer-centred criteria	0	0	7	3	0	9	1	6	0	0	0	1	1	0	0	1	29
No criteria fulfilled	1	1	0	0	1	2	1	8	0	1	0	2	1	1	0	5	24
With available age of onset	34	43	22	22	3	65	9	59	1	0	0	7	2	2	4	15	288

(Figure 4 continues on next page)



Number of cases																	Overall
Total	0	0	12	7	0	0	1	11	0	1	0	0	1	0	2	8	43
2015 HDGC criteria fulfilled	NA	NA	1	0	NA	NA	0	0	0	0	0	0	0	0	0	0	1
Additional cases included by 2020 HDGC criteria	NA	NA	1	3	NA	NA	0	0	0	0	0	0	0	0	0	3	7
Additional cases included by lobular breast cancer-centred criteria	NA	NA	10	4	NA	NA	0	8	0	1	0	0	0	0	1	2	26
No criteria fulfilled	NA	NA	0	0	NA	NA	1	3	0	0	0	0	1	0	1	3	9
With available age of onset	0	0	11	7	0	0	0	8	0	0	0	0	1	0	0	8	35

Figure 4: **Distribution of phenotypes in probands and relatives from families carrying germline *CDH1* PV/LPVs and having gastric cancer only, gastric cancer and breast cancer, and breast cancer only, with reference to HDGC clinical criteria**

Relative frequency of phenotypes, phenotype distribution, number of cases, and average age of onset of probands and relatives in families carrying PV/LPVs and having gastric cancer only, gastric cancer and breast cancer, or breast cancer only (data further detailed in appendix pp 62–63). The left axis shows the relative frequency (%) of each phenotype out of the total number of phenotypes. Mean age of onset (SD) is shown for HDGC-associated phenotypes. Detailed data on number of cases and average age of onset for other phenotypes are provided in appendix (pp 62–63). Probands could contribute more than one phenotype. HDGC=hereditary diffuse gastric cancer. PV/LPV=pathogenic or likely pathogenic variant. P=probands. R=relatives. VUS=variants of unknown significance. *Breast cancer of unknown histotype. †Other phenotypes not including gastric cancer, breast cancer, colorectal cancer, or ovarian cancer (detailed in appendix p 3).

diffuse gastric cancer and gastric cancer. These findings contribute to our understanding of the relationship between *CDH1* PV/LPVs and diffuse gastric cancer and lobular breast cancer in probands, and gastric cancer in relatives.[9,12,30] Other cancer types—including ductal breast cancer, ovarian cancer, and colorectal cancer—were extremely rare in PV/LPV carrier families, showing no evidence of association with the presence of PV/LPVs. This finding supports the exclusion of these cancer types from the spectrum of cancers associated with *CDH1* PV/LPVs.[31] Altogether, these genotype–phenotype data support clinical recommendations for intensive surveillance and prophylactic surgery of stomach and breast only in carriers of *CDH1* pathogenic or likely pathogenic variants.[2] The apparent association between the occurrence of ovarian cancer in families carrying missense variants of unknown significance merits further research, which is beyond the scope of the current study.

This European cohort had a predominance of families with gastric cancer only and gastric cancer and breast cancer over families with breast cancer only, which represented 9% of all *CDH1* PV/LPV carrier families. By contrast, similar cohorts in the USA reported breast cancer only (presumably HLBC) in more than 35% of families carrying *CDH1* pathogenic or likely pathogenic variants.[9,12] Our observation that a large fraction (38%) of PV/LPV carrier families have both gastric cancer and breast cancer are further supported by additional data from the USA, which showed occult diffuse gastric cancer in risk-reduction gastrectomy specimens from 93·8% of individuals with *CDH1* pathogenic or likely pathogenic variants in presumed HLBC families.[17]

27% of *CDH1* PV/LPV carrier families, herein described, had at least one histologically confirmed case of lobular breast cancer, with or without gastric cancer (sometimes unconfirmed diffuse gastric cancer), highlighting the need (as reinforced by others)[32] to widen the 2020 HDGC criteria.[2] To maximise the diagnosis of *CDH1* PV/LPV carriers, we propose three novel lobular breast cancer-centred criteria to drive *CDH1* testing in patients with histologically confirmed lobular breast cancer with or without a history of gastric cancer, representing most clinical contexts among carriers of *CDH1* PV/LPVs who do not meet the 2020 HDGC criteria.[2] This widening needs to assure that the criteria remain sensitive enough to identify as many pathogenic or likely pathogenic variant carriers as possible, with genetic testing being done preferentially in individuals who are most likely to carry such variants. Lerner and colleagues[16] followed a similar rationale and proposed testing for *CDH1* variants in everyone fulfilling the modified 2020 HDGC criteria and the Hereditary Breast and Ovarian Cancer criteria (Yale criteria), reporting an increased sensitivity and simplified system. Although a high sensitivity was attained, it was accompanied by a substantial decrease in specificity and Youden index (appendix p 70). Using the Yale criteria to select patients for *CDH1* testing will lead to

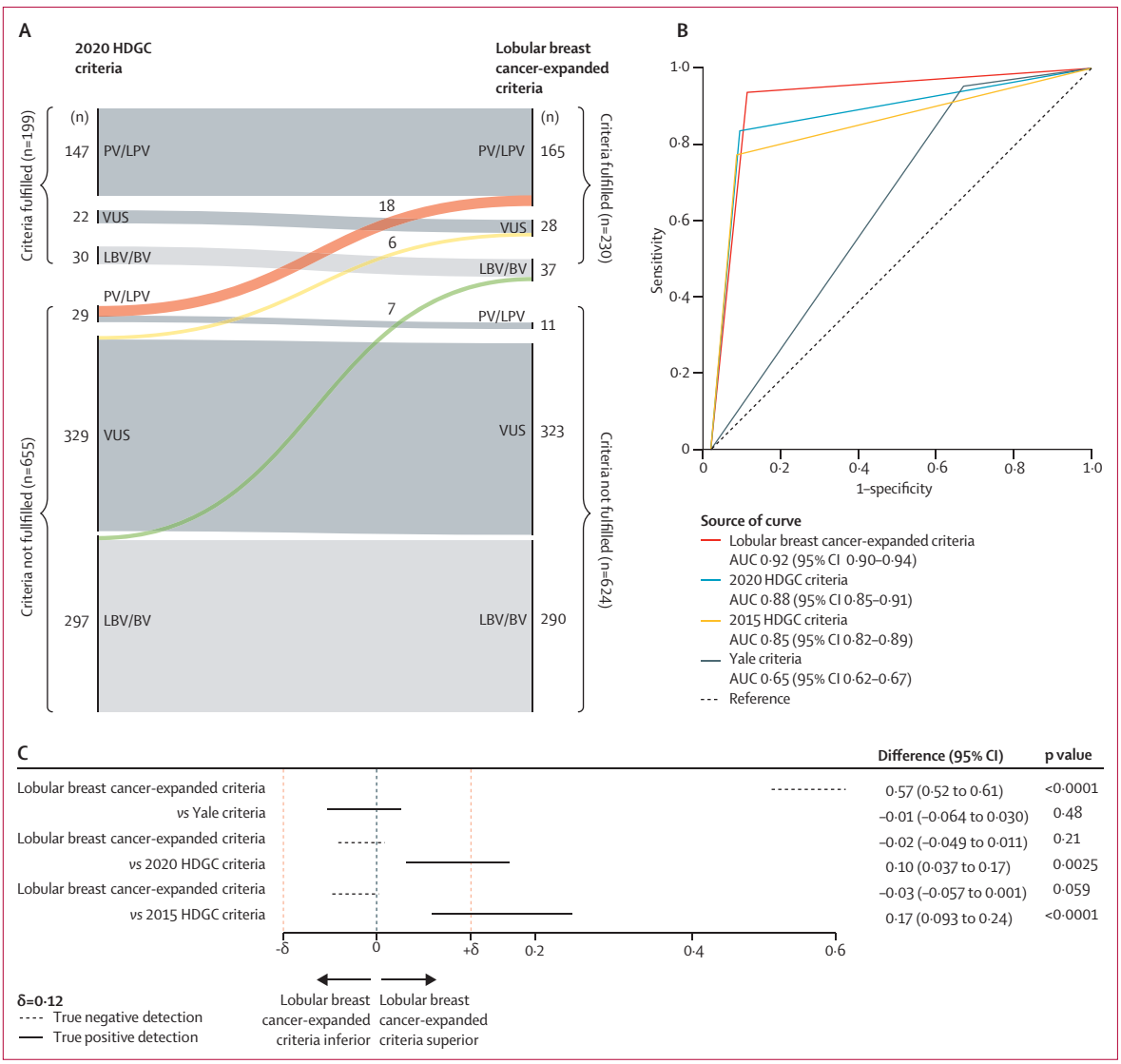


***Figure 5:* *CDH1* variant classification and distribution according to compliance with 2020 HDGC clinical criteria, and performance evaluation of HDGC clinical criteria sets**

(A) Alluvial plot (created with the visualisation platform developed by Mauri and colleagues)[26] displaying the number of families carrying *CDH1* variants and fulfilling 2020 HDGC criteria or lobular breast cancer-expanded criteria. Variants are categorised according to the American College of Medical Genetics and Genomics and Association for Molecular Pathology (version 2). (B) Receiver operating characteristic analysis comparing the AUCs for four different clinical criteria sets (lobular breast cancer-expanded *vs* 2020 HDGC criteria Z=3·54, AUC difference 0·042 [95% CI 0·02–0·06], p=0·0004). (C) Forest plot representing the lobular breast cancer-expanded clinical criteria equivalence test. The equivalence limit (δ) was set to 0·12. HDGC=hereditary diffuse gastric cancer. LBV/BV=likely benign or benign variant. PV/LPV=pathogenic or likely pathogenic variant. VUS=variants of unknown significance.

an excess of patients with breast cancer being tested with a very low pickup rate, representing a waste of resources. Contrarily, the lobular breast cancer-expanded criteria proposed in this Article showed superior performance to that of previous criteria, and are expected to identify a larger number of carriers of *CDH1* pathogenic or likely pathogenic variants with high specificity and sensitivity.

Despite the established *CDH1*-related causality in lobular breast cancer, HLBC as an entity remains controversial since its definition in 2020.[2,8,12,17] Our genotype–phenotype analysis does not provide evidence that specific *CDH1* PV/LPVs predispose individuals differentially to HDGC (families with gastric cancer only or gastric cancer and breast cancer) or to HLBC (families with lobular breast cancer only), as suggested previously, particularly in the HLBC context.[8,17,27,28] These data are supported by Gamble and colleagues, who studied 31 families tested because of suspicion of HLBC, and carrying 19 different *CDH1* pathogenic or likely pathogenic variants.[17] These authors conclude that a proportion of families with HLBC who initially did not report history of gastric cancer exhibited high rates of

occult signet ring cell gastric cancer in a prospective study. Additionally, our literature review showed that most *CDH1* PV/LPVs reported in families with breast cancer only have been reported in families with HDGC elsewhere (appendix pp 73–77).[17] Alternatively, HLBC and HDGC might differentially occur in pathogenic or likely pathogenic variant carrier families, if they originate from particular geographical regions and due to exposure to particular environmental factors or the presence of modifier genetic events,[33–35] which might favour clinical expression of diffuse gastric cancer only or lobular breast cancer only, or both (appendix pp 73–77). Additionally, these modifier genes or environmental factors might prevent the progression of intramucosal gastric signet ring cell carcinoma,[35] leading to familial presentations of lobular breast cancer only, explaining HLBC in some instances. In such cases, *CDH1*-associated HLBC might exist as a HDGC-independent clinical entity, according to the 2020 guidelines paper.[2] In any case, as available data indicate that *CDH1* pathogenic or likely pathogenic variants predispose individuals to both diffuse gastric cancer and lobular breast cancer, clinical management should address both organs.

An additional factor that might constitute a confounder is variant classification. Six *CDH1* missense variants earlier described as related to HLBC[8,17,27,28] were revisited and reclassified in this study as LBV/BVs or variants of unknown significance, in accordance with American College of Medical Genetics and Genomics and Association for Molecular Pathology *CDH1* guidelines (appendix pp 73–77). This example highlights the need of accurate variant classification for correct evaluation of disease-causal factors, and adequate disease management. Clinical classification is challenging, mainly for newly discovered variants with missing clinical data.[6,32] This is one of the reasons that led us to share as many clinical details of the families analysed as possible (appendix pp 5–46).

A limitation of this study is the fact that Sweden, the Netherlands, and the UK contributed 83 (10%) of 854 families tested specifically for *CDH1* due to clinical suspicion of HDGC, as opposed to all other countries that contributed data irrespective of HDGC clinical suspicion, as requested in the frame of the study design. Additionally, as a multicentre study, we recognise a possible information bias inherent in the use of retrospective data from multiple institutions, with restricted and non-harmonised phenotypic descriptions. Furthermore, assumption of *CDH1*-positive carrier status in clinically affected relatives, and lack of confirmed histology in many cases of gastric cancer and breast cancer among relatives, are limitations. We also recognise the possibility of overfitting in the present study, and, to our knowledge, a cohort with the same level of detail is currently not available for cross-validation purposes. This is the most detailed description of clinical data from rare *CDH1* variant carriers published to date, to the best of our knowledge, which we believe minimises the study's limitations.

In summary, this is the first multicentre genotype-first study to identify associations between rare *CDH1* germline variants and cancer phenotypes, using the largest dataset of rare *CDH1* variant carriers and their relatives, to the best of our knowledge, completed under the umbrella of ERN GENTURIS. The study highlights associations between a specific category of rare *CDH1* germline variants and HDGC specific phenotypes, and proposes novel clinical criteria to trigger *CDH1* germline testing. Altogether, these findings will contribute to clarify the context for clinical management in families carrying *CDH1* germline variants.

Contributors

JG-P and CO conceptualised the study and did the formal analysis. JG-P, RB-M, SL, AD, and CO contributed to investigation and methodology. RB-M, JG-P, SL, AD, and CO contributed to figure design. JG-P and SL contributed to classification of variants using American College of Medical Genetics and Genomics and Association for Molecular Pathology *CDH1* guidelines. AD, NP, and CLe contributed to the statistical analysis. LS contributed to data management. RM contributed to the Variant Effect Predictor analysis. JJ-M contributed to the *CDH1* gene models. LGa, SCas, SS, HP, SF, FC, CP, MRT, SA, SB-L, JBa, AB, PRB, MB, VB, HB, JBr, DC, GC, SCar, CC, KD, RdP, CD, ED-G, CE, DGE, DF, EF, RCF, CF, MG-B, MG, LGo, KH, HH, EH-F, RH, MK, KL-R, CLa, MJLL, CM-B, SM, GM, SN, AP-G, GNR, ES, ISi, CS, JLS, ISp, VS-L, GT, M-IT, ERW, MT, NH, and CO contributed to data collection. JG-P and CO wrote the original draft of the manuscript. All authors contributed to reviewing and editing the manuscript. CO contributed to supervision, project administration, and funding acquisition. JG-P, RB-M, SL, AD, LGa, and CO directly accessed and verified the underlying raw data in all research articles needed for this submission. All authors had access to the data in the study, critically revised the manuscript for important intellectual content, and have agreed to the submitted version of the results in accordance with the underlying data.

Declaration of interests

DGE declares fees from Astrazeneca and Recursion. ERW declares grants from International Alliance for Cancer Early Detection, for which they are codirector of the research domain. GNR declares receipt of funding for study materials, medical writing, and article processing charges from Italian Ministry of Education (GNR). MJLL declares consulting fees (via the Radboud University Medical Center) from Merck Sharp & Dohme (MSD), AstraZeneca, Lilly, Janssen-Cilag, Illumina, GlaxoSmithKline. PRB declares fees from AstraZeneca, MSD, and Bristol Myers Squibb; and is a scientific committee member for the Geneticancer patients association. JBa declares fees from AstraZeneca, Lilly, and Pfizer. SA is a member of APC subVCEP of the InSiGHT/ClinGen Hereditary Colorectal Cancer/Polyposis Variant Curation Expert Panel; is an unpaid member of the German Gene Diagnostics Commission and speaker of the Centre for Hereditary Tumour Syndromes of the University of Bonn. RH declares grants from SLA Pharma and Janssen Pharmaceuticals; consulting fees from Janssen and One Two Therapeutics; equipment from Fujifilm; is the head of German Consortium for Familial Gastrointestinal Cancer; and is an unpaid advisory board member of the Lynch Syndrome advocacy Group and the Familial Polyposis Group. ES declares grants from NCT/DKTK Master. ES declares honoraria for presentations from AstraZeneca, Georg Thieme Verlag KG, and payment for expert testimony from Illumina; is a member of the board of directors of Deutsche Gesellschaft für HumanGenetik; an advisor for Dresden-concept Genome Center; and is board of directors president (paid) for LNS laboratoire National de Santé. RdP declares support for presentations (via his institution) from MSD and AstraZeneca. GC declares to receive funding for study materials, medical writing, article processing charges from the Spanish Ministry of Science and Innovation, the Instituto de Salud Carlos III CIBERONC, and the Government of Catalonia.; consulting fees from VCN Biosciences Synthetic Biologics; is the chair of the Council of the International

Society of Hereditary Gastrointestinal Tumours and the FUREGA Fundació Recerca en Gastroenterologia; and stock in Synthetic Biologics. CLa declares consulting fees and honoraria from AstraZeneca and MSD, and is a paid advisory board member for Illumina.

Data sharing

Data, aggregated and pseudo-anonymised at the source, on *CDH1* germline variant molecular and clinical classification, disease spectrum, family history, and age of disease onset, will be made available through a bulk submission to the ClinVar database and in the appendix (pp 5–46) of this Article. All participating institutions have confirmed their agreement in making the information available and public.

Acknowledgments

This research was supported by ERN GENTURIS (project ID 739547); ERN GENTURIS is partly cofunded by the EU within the framework of the Third Health Programme ERN-2016—Framework Partnership Agreement 2017–2021. This work has received funding from the European Union's Horizon 2020 research and innovation programme (grant number 779257, Solve-RD: H2020-SC1–2017-Single-Stage-RTD) via a PhD studentship awarded to JG-P. The work has also been supported by the European Regional Development Fund through the COMPETE 2020-Operacional Programme for Competitiveness and Internationalisation, Portugal 2020, Programa Operacional Regional do Norte under the project "The Porto Comprehensive Cancer Center" with the reference NORTE-01-0145-FEDER-072678—Consórcio PORTO.CCC—Porto.Comprehensive Cancer Center; the Portuguese national funds through Fundação para a Ciência e Tecnologia, Ministério da Ciência, Tecnologia e Inovação (project references PTDC/BTM-TEC/30164/2017 [3DChrome] and PTDC/BTM-TEC/6706/2020 [LEGOH]); the University of Porto through the Doctoral Programme in Biomedicine (Faculty of Medicine; awarded to JG-P), Doctoral Programme in Molecular and Cellular Biotechnology Applied to Health Sciences (ICBAS; awarded to RB-M and SL), Doctoral Programme In Molecular And Cell Biology (ICBAS; awarded to AD); the Manchester National Institute for Health and Care Research (NIHR) Biomedical Research Centre (IS-BRC-1215-20007), providing support to DGE and ERW; the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014), providing support to MT; Medical Research Council (RG84369) and Cancer Research UK (C14478/A12088; RG80420/A21047/A29225), providing support to RF; the Carlos III National Health Institute and Ministerio de Ciencia e Innovación, funded by the European Regional Development Fund (PI19/00553; PID2019-111254RB-I00) and CIBERONC (CB16/12/00234), providing support to CLa, GC, and JBr; the Government of Catalonia (Pla estratègic de recerca i innovació en salut [PERIS_MedPerCan and URDCat projects], 2017SGR1282 and 2017SGR496); and the Fundação para a Ciência e Tecnologia (UIDB/04539/2020, UIDP/04539/2020, LA/P/0058/2020, and POCI-01-0145-FEDER-022184) providing support to CE. We are grateful to the ERN GENTURIS National Coordinators for promoting data collection at the country level for the current project; Rui Magalhães (ICBAS, University of Porto, and Centro Hospitalar Universitário do Porto) for consultancy in statistics; and Víctor Treviño (Escuela de Medicina y Ciencias de la Salud del Tecnológico de Monterrey, Monterrey, Mexico) for the R script for gene models.

References

1 van der Post RS, Vogelaar IP, Carneiro F, et al. Hereditary diffuse gastric cancer: updated clinical guidelines with an emphasis on germline *CDH1* mutation carriers. *J Med Genet* 2015; **52:** 361–74.

2 Blair VR, McLeod M, Carneiro F, et al. Hereditary diffuse gastric cancer: updated clinical practice guidelines. *Lancet Oncol* 2020; **21:** e386–97.

3 Caldas C, Carneiro F, Lynch HT, et al. Familial gastric cancer: overview and guidelines for management. *J Med Genet* 1999; **36:** 873–80.

4 Guilford P, Hopkins J, Harraway J, et al. E-cadherin germline mutations in familial gastric cancer. *Nature* 1998; **392:** 402–05.

5 Oliveira C, Senz J, Kaurah P, et al. Germline *CDH1* deletions in hereditary diffuse gastric cancer families. *Hum Mol Genet* 2009; **18:** 1545–55.

6 Lee K, Krempely K, Roberts ME, et al. Specifications of the ACMG/AMP variant curation guidelines for the analysis of germline *CDH1* sequence variants. *Hum Mutat* 2018; **39:** 1553–68.

7 Fitzgerald RC, Hardwick R, Huntsman D, et al. Hereditary diffuse gastric cancer: updated consensus guidelines for clinical management and directions for future research. *J Med Genet* 2010; **47:** 436–44.

8 Corso G, Figueiredo J, La Vecchia C, et al. Hereditary lobular breast cancer with an emphasis on E-cadherin genetic defect. *J Med Genet* 2018; **55:** 431–41.

9 Roberts ME, Ranola JMO, Marshall ML, et al. Comparison of *CDH1* penetrance estimates in clinically ascertained families *vs* families ascertained for multiple gastric cancers. *JAMA Oncol* 2019; **5:** 1325–31.

10 Pharoah PD, Guilford P, Caldas C. Incidence of gastric cancer and breast cancer in *CDH1* (E-cadherin) mutation carriers from hereditary diffuse gastric cancer families. *Gastroenterology* 2001; **121:** 1348–53.

11 Kaurah P, MacMillan A, Boyd N, et al. Founder and recurrent *CDH1* mutations in families with hereditary diffuse gastric cancer. *JAMA* 2007; **297:** 2360–72.

12 Xicola RM, Li S, Rodriguez N, et al. Clinical features and cancer risk in families with pathogenic *CDH1* variants irrespective of clinical criteria. *J Med Genet* 2019; **56:** 838–43.

13 Hansford S, Kaurah P, Li-Chang H, et al. Hereditary diffuse gastric cancer syndrome: *CDH1* mutations and beyond. *JAMA Oncol* 2015; **1:** 23–32.

14 Lowstuter K, Espenschied CR, Sturgeon D, et al. Unexpected *CDH1* mutations identified on multigene panels pose clinical management challenges. *JCO Precis Oncol* 2017; **1:** 1–12.

15 Adib E, El Zarif T, Nassar AH, et al. *CDH1* germline variants are enriched in patients with colorectal cancer, gastric cancer, and breast cancer. *Br J Cancer* 2022; **126:** 797–803.

16 Lerner BA, Xicola RM, Rodriguez NJ, Karam R, Llor X. Simplified and more sensitive criteria for identifying individuals with pathogenic *CDH1* variants. *J Med Genet* 2022; published online Jan 25. https://doi.org/10.1136/jmedgenet-2021-108169.

17 Gamble LA, Rossi A, Fasaye GA, et al. Association between hereditary lobular breast cancer due to *CDH1* variants and gastric cancer risk. *JAMA Surg* 2022; **157:** 18–22.

18 Oliveira C, Bordin MC, Grehan N, et al. Screening E-cadherin in gastric cancer families reveals germline mutations only in hereditary diffuse gastric cancer kindred. *Hum Mutat* 2002; **19:** 510–17.

19 den Dunnen JT, Dalgleish R, Maglott DR, et al. HGVS recommendations for the description of sequence variants: 2016 update. *Hum Mutat* 2016; **37:** 564–69.

20 McLaren W, Gil L, Hunt SE, et al. The ensembl variant effect predictor. *Genome Biol* 2016; **17:** 122.

21 Hebsgaard SM, Korning PG, Tolstrup N, Engelbrecht J, Rouzé P, Brunak S. Splice site prediction in *Arabidopsis thaliana* pre-mRNA by combining local and global sequence information. *Nucleic Acids Res* 1996; **24:** 3439–52.

23 Stevenson M, Sergeant E. EpiR: tools for the analysis of epidemiological data. https://cran.r-project.org/web/packages/epiR/index.html (accessed Aug 25, 2022).

24 Perez-Jaume S, Skaltsa K, Pallarès N, Carrasco JL. ThresholdROC: optimum threshold estimation tools for continuous diagnostic tests in R. *J Stat Softw* 2017; **82:** 1–21.

25 Walker E, Nowacki AS. Understanding equivalence and noninferiority testing. *J Gen Intern Med* 2011; **26:** 192–96.

26 Mauri M, Elli T, Caviglia G, Uboldi G, Azzi M. RAWGraphs: a visualisation platform to create open outputs. 12th Biannual Conference on Italian SIGCHI Chapter; Sept 18–20, 2017 (abstr 28).

27 Girardi A, Magnoni F, Vicini E, et al. *CDH1* germline mutations in families with hereditary lobular breast cancer. *Eur J Cancer Prev* 2022; **31:** 274–78.

28 Corso G, Intra M, Trentin C, Veronesi P, Galimberti V. *CDH1* germline mutations and hereditary lobular breast cancer. *Fam Cancer* 2016; **15:** 215–19.

29 Barbosa-Matos R, Leal Silva R, Garrido L, et al. The *CDH1* c.1901C>T variant: a founder variant in the Portuguese population with severe impact in mRNA splicing. *Cancers (Basel)* 2021; **13:** 4464.

For **ClinVar** see https://www.ncbi.nlm.nih.gov/clinvar/

30 Lo W, Zhu B, Sabesan A, et al. Associations of *CDH1* germline variant location and cancer phenotype in families with hereditary diffuse gastric cancer (HDGC). *J Med Genet* 2019; **56:** 370–79.

31 Massari G, Magnoni F, Favia G, et al. Frequency of *CDH1* germline mutations in non-gastric cancers. *Cancers (Basel)* 2021; **13:** 2321.

32 Katona BW, Clark DF, Domchek SM. *CDH1* on multigene panel testing: look before you leap. *J Natl Cancer Inst* 2020; **112:** 330–34.

33 Zhang H, Ahearn TU, Lecarpentier J, et al. Genome-wide association study identifies 32 novel breast cancer susceptibility loci from overall and subtype-specific analyses. *Nat Genet* 2020; **52:** 572–81.

34 Tanikawa C, Kamatani Y, Toyoshima O, et al. Genome-wide association study identifies gastric cancer susceptibility loci at 12q24.11-12 and 20q11.21. *Cancer Sci* 2018; **109:** 4015–24.

35 Ellison-Loschmann L, Sporle A, Corbin M, et al. Risk of stomach cancer in Aotearoa/New Zealand: a Māori population based case-control study. *PLoS One* 2017; **12:** e0181581.