Supplementary material

1. **Epi25 Consortium**

    1.1. Acknowledgments

2. **Supplementary Methods**

    2.1. Study cohorts

    2.2. Controls

        2.2.1. GPC controls

        2.2.2. Finrisk controls

        2.2.3. IBD cases and controls

    2.3. Methods

    2.4. PheWAS

3. **Supplementary Figures**

    3.1. Supplementary Fig. 1 Principle component analysis of cases and controls combined with the 1000 Genome Populations.

    3.2. Supplementary Fig. 2: CNV-load per individual

    3.3. Supplementary Fig. 3: Epilepsy phenotype residual distribution

    3.4. Supplementary Fig. 4: Global duplication burden in four epilepsy phenotypes

    3.5. Supplementary Fig. 5: Meta-Analysis

    3.6. Supplementary Fig. 6: Fraction of LFE patients with or without pathogenic CNVs among different lesion types

    3.7. Supplementary Fig. 7: GGE and NAFE genes form disease modules in the human interactome.

4. **Supplementary Tables**

4.1. Supplementary Table 1: Dataset summary

4.2. Supplementary Table 2: Phenotype Analysis

4.3. Supplementary Table 3: Summary of all findings obtained in this study

1. Epi25 Consortium

A full list of authors and affiliations

Epi25 sequencing, analysis, and project management at the Broad Institute

Yen-Chen Anne Feng[1,2,3,4], Daniel P. Howrigan[1,3,4], Liam E. Abbott[1,3,4], Katherine Tashman[1,3,4], Felecia Cerrato[3], Dennis Lal[4,5], Claire Churchhouse[1,3,4], Namrata Gupta[3], Benjamin M. Neale[1,3,4]

Epi25 executive committee

Samuel F. Berkovic[6], Holger Lerche[7], David B. Goldstein[8], Daniel H. Lowenstein[9]

Epi25 strategy, phenotyping, analysis, informatics, and project management committees

Samuel F. Berkovic[6], Holger Lerche[7], David B. Goldstein[8], Daniel H. Lowenstein[9], Gianpiero L. Cavalleri[10,11], Patrick Cossette[12], Chris Cotsapas[13], Peter De Jonghe[14,15,16], Tracy Dixon-Salazar[17], Renzo Guerrini[18], Hakon Hakonarson[19], Erin L. Heinzen[8], Ingo Helbig[20,21,22], Patrick Kwan[23,24], Anthony G. Marson[25], Slavé Petrovski[24,26], Sitharthan Kamalakaran[8], Sanjay M. Sisodiya[27], Randy Stewart[28], Sarah Weckhuysen[14,15,16], Chantal Depondt[29], Dennis

J. Dlugos[19], Ingrid E. Scheffer[6,30], Pasquale Striano[31], Catharine Freyer[9], Roland Krause[32], Patrick May[32], Kevin McKenna[9], Brigid M. Regan[6], Susannah T. Bellows[6], Costin Leu[4,5,27]

Authors from individual Epi25 cohorts:

Australia: Melbourne (AUSAUS)

Samuel F. Berkovic[6], Ingrid E. Scheffer[6,30], Brigid M. Regan[6], Caitlin A. Bennett[6], Susannah T. Bellows[6], Esther M.C. Johns[6], Alexandra Macdonald[6], Hannah Shilling[6], Rosemary Burgess[6], Dorien Weckhuysen[6], Melanie Bahlo[33,34]

Australia: Royal Melbourne (AUSRMB)

Terence J. O'Brien[23,24], Patrick Kwan[23,24], Slavé Petrovski[24,26], Marian Todaro[23,24]

Belgium: Antwerp (BELATW)

Sarah Weckhuysen[14,15,16], Hannah Stamberger[14,15,16], Peter De Jonghe[14,15,16]

Belgium: Brussels (BELULB)

Chantal Depondt[29]

Canada: Andrade (CANUTN)

Danielle M. Andrade[35,36], Tara R. Sadoway[36], Kelly Mo[36]

Switzerland: Bern (CHEUBB)

Heinz Krestel[37], Sabina Gallati[38]

Cyprus (CYPCYP)

Savvas S. Papacostas[39], Ioanna Kousiappa[39], George A. Tanteles[40]

Czech Republic: Prague (CZEMTH)

Katalin Šterbová[41], Markéta Vlcková[42], Lucie Sedlácková[41], Petra Laššuthová[41]

Germany: Frankfurt/Marburg (DEUPUM)

Karl Martin Klein[43,44], Felix Rosenow[43,44], Philipp S. Reif[43,44], Susanne Knake[44]

Germany: Bonn (DEUUKB)

Wolfram S. Kunz[45,46], Gábor Zsurka[45,46], Christian E. Elger[46], Jürgen Bauer[46], Michael Rademacher[46]

Germany: Kiel (DEUUKL)

Ingo Helbig[20,21,22], Karl Martin Klein[43,44], Manuela Pendziwiat[21], Hiltrud Muhle[21], Annika Rademacher[21], Andreas van Baalen[21], Sarah von Spiczak[21], Ulrich Stephani[21], Zaid Afawi[47],

Amos D. Korczyn[48], Moien Kanaan[49], Christina Canavati[49], Gerhard Kurlemann[50], Karen Müller-Schlüter[51], Gerhard Kluger[52,53], Martin Häusler[54], Ilan Blatt[48,55]

Germany: Leipzig (DEUULG)

Johannes R. Lemke[56], Ilona Krey[56]

Germany: Tuebingen (DEUUTB)

Holger Lerche[7], Yvonne G. Weber[7,57], Stefan Wolking[7], Felicitas Becker[7,58], Christian Hengsbach[7], Sarah Rau[7], Ana F. Maisch[7], Bernhard J. Steinhoff[59], Andreas Schulze-Bonhage[60], Susanne Schubert-Bast[61], Herbert Schreiber[62], Ingo Borggräfe[63], Christoph J. Schankin[64], Thomas Mayer[65], Rudolf Korinthenberg[66], Knut Brockmann[67], Gerhard Kurlemann[50], Dieter Dennig[68], Rene Madeleyn[69]

Finland: Kuopio (FINKPH)

Reetta Kälviäinen[70], Pia Auvinen[70], Anni Saarela[70]

Finland: Helsinki (FINUVH)

Tarja Linnankivi[71], Anna-Elina Lehesjoki[72]

Wales: Swansea (GBRSWU)

Mark I. Rees[73,74], Seo-Kyung Chung[73,74], William O. Pickrell[73], Robert Powell[73,75]

UK: UCL (GBRUCL)

Sanjay M. Sisodiya[27], Natascha Schneider[27], Simona Balestrini[27], Sara Zagaglia[27], Vera Braatz[27]

UK: Imperial/Liverpool (GBRUNL)

Anthony G. Marson[25], Michael R. Johnson[76], Pauls Auce[77], Graeme J. Sills[78]

Hong Kong (HKGHKK)

Patrick Kwan[23,24,79], Larry W. Baum[80,81,82], Pak C. Sham[80,81,82], Stacey S. Cherny[83], Colin H.T. Lui[84]

Croatia (HRVUZG)

Nina Barišic[85]

Ireland: Dublin (IRLRCI)

Gianpiero L. Cavalleri[10,11], Norman Delanty[10,11], Colin P. Doherty[86,11], Arif Shukralla[87], Mark McCormack[10], Hany El-Naggar[87,11]

Italy: Milan (ITAICB)

Laura Canafoglia[88], Silvana Franceschetti[88], Barbara Castellotti[89], Tiziana Granata[90]

Italy: Genova(ITAIGI)

Pasquale Striano[31], Federico Zara[91], Michele Iacomino[91], Francesca Madia[91], Maria Stella Vari[31], Maria Margherita Mancardi[91], Vincenzo Salpietro[31]

Italy: Bologna (ITAUBG)

Francesca Bisulli[92,93], Paolo Tinuper[92,93], Laura Licchetta[92,93], Tommaso Pippucci[94], Carlotta Stipa[92,93], Lorenzo Muccioli[92,93], Raffaella Minardi[92]

Italy: Catanzaro (ITAUMC)

Antonio Gambardella[95], Angelo Labate[95], Grazia Annesi[96], Lorella Manna[96], Monica Gagliardi[96]

Italy: Florence (ITAUMR)

Renzo Guerrini[18], Elena Parrini[18], Davide Mei[18], Annalisa Vetro[18], Claudia Bianchini[18], Martino Montomoli[18], Viola Doccini[18], Carla Marini[18]

Japan: RIKEN Institute (JPNRKI)

Toshimitsu Suzuki[97], Yushi Inoue[98], Kazuhiro Yamakawa[97]

Lithuania (LTUUHK)

Birute Tumiene[99,100], Ruta Mameniskiene[127], Algirdas Utkus[127], Ruta Praninskiene[127], Jurgita Grikiniene[127], Ruta Samaitiene[127]

New Zealand: Otago (NZLUTO)

Lynette G. Sadleir[101], Chontelle King[101], Emily Mountier[101]

Turkey: Bogazici (TURBZU)

S. Hande Caglayan[102], Mutluay Arslan[103], Zuhal Yapici[104], Uluc Yis[105], Pinar Topaloglu[104], Bulent Kara[106], Dilsad Turkdogan[107], Asli Gundogdu-Eken[102]

Turkey: Istanbul (TURIBU)

Nerses Bebek[108,109], Sibel Ugur-Iseri[109], Betül Baykan[108], Baris Salman[109], Garen Haryanyan[108], Emrah Yücesan[110], Yesim Kesim[108], Çigdem Özkara[111]

USA: BCH (USABCH)

Beth R. Sheidley[112,113], Catherine Shain[112,113], Annapurna Poduri[112,113]

USA: Philadelphia/CHOP (USACHP) and Philadelphia/Rowan (USACRW)

Russell J. Buono[114,115,19], Thomas N. Ferraro[114,22], Michael R. Sperling[115], Dennis J. Dlugos[19,22], Warren Lo[116], Michael Privitera[117], Jacqueline A. French[118], Patrick Cossette[12], Steven Schachter[119], Hakon Hakonarson[19]

USA: EPGP (USAEGP)

Daniel H. Lowenstein[9], Ruben I. Kuzniecky[120], Dennis J. Dlugos[19,22], Orrin Devinsky[118]

USA: NYU HEP (USAHEP)

Daniel H. Lowenstein[9], Ruben I. Kuzniecky[120], Jacqueline A. French[118], Manu Hegde[9]

USA: Penn/CHOP (USAUPN)

Ingo Helbig[20,22], Pouya Khankhanian[121,122], Katherine L. Helbig[20], Colin A. Ellis[122]

Epi25 control cohort

Gianfranco Spalletta[123,124], Fabrizio Piras[123], Federica Piras[123], Tommaso Gili[125,123], Valentina Ciullo[123,126]

Affiliations

1 Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA

[2] Psychiatric & Neurodevelopmental Genetics Unit, Department of Psychiatry, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA

[3] Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, 7 Cambridge Center, Cambridge, MA 02142, USA

[4] Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA

[5] Genomic Medicine Institute, Cleveland Clinic, Cleveland, OH 44195, USA

[6] Epilepsy Research Centre, Department of Medicine, University of Melbourne, Victoria, Australia

[7] Department of Neurology and Epileptology, Hertie Institute for Clinical Brain Research, University of Tübingen, 72076 Tübingen, Germany

[8] Institute for Genomic Medicine, Columbia University, New York, NY 10032, USA

[9] Department of Neurology, University of California, San Francisco, CA 94110, USA

[10] The Department of Molecular and Cellular Therapeutics, The Royal College of Surgeons in Ireland, Dublin, Ireland

[11] The FutureNeuro Research Centre, Ireland

[12] University of Montreal, Montreal, QC H3T 1J4, Canada

[13] School of Medicine, Yale University, New Haven, CT 06510, USA

[14] Neurogenetics Group, University of Antwerp, Belgium

[15] Laboratory of Neurogenetics, Institute Born-Bunge, University of Antwerp, Belgium

[16] Division of Neurology, Antwerp University Hospital, Antwerp, Belgium

[17] LGS Foundation, NY 11716, USA

[18] Pediatric Neurology, Neurogenetics and Neurobiology Unit and Laboratories, Children's Hospital A. Meyer, University of Florence, Italy

[19] The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA

[20] Division of Neurology, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA

[21] Department of Neuropediatrics, Christian-Albrechts-University of Kiel, 24105 Kiel, Germany

[22] Perelman School of Medicine, University of Pennsylvania, PA 19104, USA

[23] Department of Neuroscience, Central Clinical School, Monash University, Alfred Hospital, Melbourne, Australia

[24] Departments of Medicine and Neurology, University of Melbourne, Royal Melbourne Hospital, Parkville, Australia

[25] Department of Molecular and Clinical Pharmacology, University of Liverpool, Liverpool, UK

[26] Centre for Genomics Research, Precision Medicine and Genomics, IMED Biotech Unit, AstraZeneca, Cambridge, UK

[27] Department of Clinical and Experimental Epilepsy, UCL Queen Square Institute of Neurology, London, UK and Chalfont Centre for Epilepsy, Chalfont St Peter, UK

[28] National Institute of Neurological Disorders and Stroke, MD 20852, USA

[29] Department of Neurology, Université Libre de Bruxelles, Brussels, Belgium

[30] Department of Paediatrics, University of Melbourne, Royal Children's Hospital, Florey and Murdoch Children's Research Institutes, Melbourne, Australia

[31] Pediatric Neurology and Muscular Diseases Unit, Department of Neurosciences, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health, University of Genoa, IRCCS "G. Gaslini" Institute, Genova, Italy

[32] Luxembourg Centre for Systems Biomedicine, University Luxembourg, Esch-sur-Alzette, Luxembourg

[33] Population Health and Immunity Division, the Walter and Eliza Hall Institute of Medical Research, Parkville 3052, VIC, Australia

[34] Department of Medical Biology, The University of Melbourne, Melbourne 3010, VIC, Australia

[35] Department of Neurology, Toronto Western Hospital, Toronto, ON M5T 2S8, Canada

[36] University Health Network, University of Toronto, Toronto, ON, Canada

[37] Departments of Neurology and BioMedical Research, Bern University Hospital and University of Bern, Bern, Switzerland

[38] Institute of Human Genetics, Bern University Hospital, Bern, Switzerland

[39] Neurology Clinic B, The Cyprus Institute of Neurology and Genetics, 2370 Nicosia, Cyprus

[40] Department of Clinical Genetics, The Cyprus Institute of Neurology and Genetics, 2370 Nicosia, Cyprus

[41] Department of Paediatric Neurology, 2nd Faculty of Medicine, Charles University and Motol Hospital, Prague, Czech Republic

[42] Department of Biology and Medical Genetics, 2nd Faculty of Medicine, Charles University and Motol Hospital, Prague, Czech Republic

[43] Epilepsy Center Frankfurt Rhine-Main, Center of Neurology and Neurosurgery, Goethe University Frankfurt and LOEWE Center for Personalized Translational Epilepsy Research (CePTER) Frankfurt, Germany

[44] Epilepsy Center Hessen-Marburg, Department of Neurology, Philipps University Marburg, Marburg, Germany

[45] Institute of Experimental Epileptology and Cognition Research, University Bonn, 53127 Bonn, Germany

[46] Department of Epileptology, University Bonn, 53127 Bonn, Germany

[47] Sackler School of Medicine, Tel-Aviv University, Ramat Aviv, Israel

[48] Tel-Aviv University Sackler Faculty of Medicine, Ramat Aviv 69978, Israel

[49] Hereditary Research Lab, Bethlehem University, Bethlehem, Palestine

[50] Department of Neuropediatrics, Westfälische Wilhelms-University, Münster, Germany

[51] Epilepsy Center for Children, University Hospital Neuruppin, Brandenburg Medical School, Neuruppin, Germany

[52] Neuropediatric Clinic and Clinic for Neurorehabilitation, Epilepsy Center for Children and Adolescents, Vogtareuth, Germany

[53] Research Institute Rehabilitation / Transition / Palliation, PMU Salzburg, Austria

[54] Division of Neuropediatrics and Social Pediatrics, Department of Pediatrics, University Hospital, RWTH Aachen, Aachen, Germany

[55] Department of Neurology, Sheba Medical Center, Ramat Gan, Israel

[56] Institute of Human Genetics, Leipzig, Germany

[57] Department of Neurosurgery, University of Tübingen, Germany

[58] RKU-University Neurology Clinic of Ulm, Ulm, Germany

[59] Kork Epilepsy Center, Kehl-Kork, Germany

[60] Epilepsy Center, University of Freiburg, Freiburg im Breisgau, Germany

[61] Section Neuropediatrics and Inborn Errors of Metabolism, University Children's Hospital, Heidelberg, Germany

[62] Neurological Practice Center & Neuropoint Patient Academy, Ulm, Germany

[63] Department of Pediatric Neurology and Developmental Medicine, LMU Munich, Munich, Germany

[64] Department of Neurology, University of Munich Hospital-Großhadern, Munich, Germany

[65] Saxonian Epilepsy Center Radeberg, Radeberg, Germany

[66] Division of Neuropediatrics and Muscular Disorders, University Hospital Freiburg, Freiburg, Germany

[67] University Children's Hospital, Göttingen, Germany

[68] Private Neurological Practice, Stuttgart, Germany

[69] Department of Pediatrics, Filderklinik, Filderstadt, Germany

[70] Neurocenter, Kuopio University Hospital, Kuopio Finland and Institute of Clinical Medicine, University of Eastern Finland, Finland

[71] Child Neurology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland

[72] Medicum, University of Helsinki, Helsinki, Finland and Folkhälsan Research Center, Helsinki, Finland

[73] Neurology Research Group, Swansea University Medical School, Swansea University SA2 8PP, UK

[74] Faculty of Medicine and Health, University of Sydney, Sydney, Australia

[75] Department of Neurology, Morriston Hospital, Abertawe Bro Morgannwg Health University Board, Swansea, UK

[76] Division of Brain Sciences, Imperial College London, London, UK

[77] Department of Neurology, Walton Centre NHS Foundation Trust, Liverpool, UK

[78] School of Life Sciences, University of Glasgow, Glasgow, UK

[79] Department of Medicine and Therapeutics, Chinese University of Hong Kong, Hong Kong, China

[80] The State Key Laboratory of Brain and Cognitive Sciences, University of Hong Kong, Hong Kong, China

[81] Centre for Genomic Sciences, University of Hong Kong, Hong Kong, China

[82] Department of Psychiatry, University of Hong Kong, Hong Kong, China

[83] Department of Epidemiology and Preventive Medicine and Department of Anatomy and Anthropology, Sackler Faculty of Medicine, Tel Aviv University, Israel

[84] Department of Medicine, Tseung Kwan O Hospital, Hong Kong, China

[85] Department of Pediatric University Hospital centre Zagreb, Croatia

[86] Neurology Department, St. James Hospital, Dublin, Ireland

[87] The Department of Neurology, Beaumont Hospital, Dublin, Ireland

[88] Neurophysiopathology, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy

[89] Unit of Genetics of Neurodegenerative and Metabolic Diseases, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy

[90] Department of Pediatric Neuroscience, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy

[91] Laboratory of Neurogenetics, IRCCS "G. Gaslini" Institute, Genova, Italy

[92] IRCCS, Institute of Neurological Sciences of Bologna, Bologna, Italy

[93] Department of Biomedical and Neuromotor Sciences, University of Bologna, Bologna, Italy

[94] Medical Genetics Unit, Polyclinic Sant'Orsola-Malpighi University Hospital, Bologna, Italy

[95] Institute of Neurology, Department of Medical and Surgical Sciences, University "Magna Graecia", Catanzaro, Italy

[96] Institute of Molecular Bioimaging and Physiology, CNR, Section of Germaneto, Catanzaro, Italy

[97] Laboratory for Neurogenetics, RIKEN Center for Brain Science, Saitama, Japan

[98] National Epilepsy Center, Shizuoka Institute of Epilepsy and Neurological Disorder, Shizuoka, Japan

[99] Institute of Biomedical Sciences, Faculty of Medicine, Vilnius University, Vilnius, Lithuania

[100] Centre for Medical Genetics, Vilnius University Hospital Santaros Klinikos, Vilnius, Lithuania

[101] Department of Paediatrics and Child Health, University of Otago, Wellington, New Zealand

[102] Department of Molecular Biology and Genetics, Bogaziçi University, Istanbul, Turkey

[103] Department of Child Neurology, Gulhane Education and Research Hospital, Health Sciences University, Ankara, Turkey

[104] Department of Child Neurology, Istanbul Faculty of Medicine, Istanbul University, Istanbul, Turkey

[105] Department of Child Neurology, Medical School, Dokuz Eylul University, Izmir, Turkey

[106] Department of Child Neurology, Medical School, Kocaeli University, Kocaeli, Turkey

[107] Department of Child Neurology, Medical School, Marmara University, Istanbul, Turkey

[108] Department of Neurology, Istanbul Faculty of Medicine, Istanbul University, Istanbul, Turkey

[109] Department of Genetics, Aziz Sancar Institute of Experimental Medicine, Istanbul University, Istanbul, Turkey

[110] Bezmialem Vakif University, Institute of Life Sciences and Biotechnology, Istanbul, Turkey

[111] Department of Neurology, Faculty of Medicine, Cerrahpasa University Istanbul, Istanbul, Turkey

[112] Epilepsy Genetics Program, Department of Neurology, Boston Children's Hospital, Boston, MA 02115, USA

[113] Department of Neurology, Harvard Medical School, Boston, MA 02115, USA

[114] Cooper Medical School of Rowan University, Camden, NJ 08103, USA

[115] Thomas Jefferson University, Philadelphia, PA 19107, USA

[116] Nationwide Children's Hospital, Columbus, OH 43205, USA

[117] University of Cincinnati, Cincinnati, OH 45220, USA

[118] Department of Neurology, New York University/Langone Health, New York, NY 10016, USA

[119] Beth Israel Deaconess/Harvard, Boston, MA 02115, USA

[120] Department of Neurology, Hofstra-Northwell Medical School, New York, NY 11549, USA

[121] Center for Neuro-engineering and Therapeutics, University of Pennsylvania, Philadelphia, PA 19104, USA

[122] Department of Neurology, Hospital of University of Pennsylvania, Philadelphia, PA 19104, USA

[123] Neuropsychiatry Laboratory, IRCCS Santa Lucia Foundation, Rome, Italy

[124] Division of Neuropsychiatry, Menninger Department of Psychiatry and Behavioral Sciences, Baylor College of Medicine, Houston, TX, USA

[125] IMT School for Advanced Studies Lucca, Lucca, Italy

[126] Department of Neurosciences, Psychology, Drug Research and Child Health, University of Florence, Florence, Italy

[127] Vilnius University, Faculty of Medicine, Institute of Clinical Medicine, Vilnius, Lithuania

1.1 Acknowledgments

We gratefully thank the Epi25 principal investigators, local staff from individual cohorts, and all the epilepsy patients who participated in the study for making possible this global collaboration and resource to advance epilepsy genetics research. This work is part of the Centers for Common Disease Genomics (CCDG) program, funded by the National Human Genome Research Institute (NHGRI) and the National Heart, Lung, and Blood Institute (NHLBI). CCDG-funded Epi25 research activities at the Broad Institute, including genomic data generation in the Broad Genomics Platform, are supported by NHGRI grant UM1 HG008895 (PIs: Eric Lander, Stacey Gabriel, Mark Daly, Sekar Kathiresan). The content is

solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

We are grateful for support from the Epilepsy Society, UK, and the Muir Maxwell Trust. Part of this work was undertaken at University College London Hospitals, which received a proportion of funding from the NIHR Biomedical Research Centres Funding Scheme. The Vanderbilt University Medical Center's biobank (BioVU) was supported by institutional funding and by the Vanderbilt CTSA grant UL1 TR000445 from NCATS/NIH. This research has been conducted using the UK Biobank Resource under Application Number '35124'. This work involved the use of the Enterprise Research Infrastructure & Services (ERIS) at Partners HealthCare. We thank the Partners HealthCare Biobank for providing genomic and health information data. The in-house project on inflammatory bowel disease without reported epilepsy was generated as part of the The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) IBD Genetics Consortium (IBDGC) supported by The Helmsley Charitable Trust and the Centers for Common Disease Genomes Program (NHGRI). DNA samples were obtained from the following collections: The Lunenfeld-Tanenbaum Research Institute Mount Sinai Hospital (PI: Mark Silverberg), The University of Pittsburgh School of Medicine (PI: Richard Duerr), The Emory University School of Medicine (PI: Subra Kugathasan), The Johns Hopkins Hospital (PI: Steven Brant), The Icahn School of Medicine at Mount Sinai (PI: Judy Cho), The Washington University School of Medicine (PI: Rodney Newberry), The University of Miami Miller School of Medicine (PI: Maria Abreu, Jake McCauley), and Cedars Sinai (PI: Dermot McGovern, Stephan Targan).

For the FINRISK population controls, we thank: Aarno Palotie[1,2,3,4,5]

Affiliations

[1] Analytic and Translational Genetics Unit, Department of Medicine, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA

[2] Psychiatric & Neurodevelopmental Genetics Unit, Department of Psychiatry, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, USA

[3] Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA

[4] Institute for Molecular Medicine Finland, University of Helsinki, Helsinki 00014, Finland

[5] Department of Neurology, Massachusetts General Hospital, Boston, MA 02114, USA

For the Genomic Psychiatry Cohort (GPC) healthy controls, we thank: Michele T. Pato[1], Carlos N. Pato[1], Evelyn J. Bromet[2], Celia Barreto Carvalho[3], Eric D. Achtyes[4], Maria Helena Azevedo[5], Roman Kotov[2], Douglas S. Lehrer[6], Dolores Malaspina[7], Stephen R. Marder[8], Helena Medeiros[1], Christopher P. Morley[9], Diana O. Perkins[10], Janet L Sobell[11], Peter F. Buckley[12], Fabio Macciardi[13], Mark H. Rapaport[14], James A. Knowles[1], Genomic Psychiatry Cohort (GPC) Consortium, Ayman H. Fanous[1,15], Steven A. McCarrol[16,17,18]

Affiliations

[1] Department of Psychiatry and Behavioral Sciences, SUNY Downstate Medical Center, Brooklyn, NY, USA

[2] Department of Psychiatry, Stony Brook University, Stony Brook, NY, USA

[3] Faculty of Social and Human Sciences, University of Azores, PT

[4] Cherry Health and Michigan State University College of Human Medicine, Grand Rapids, MI, USA

[5] Institute of Medical Psychology, Faculty of Medicine, University of Coimbra, Coimbra, PT

[6] Department of Psychiatry, Wright State University, Dayton, OH, USA

[7] Departments of Psychiatry, Genetics & Genomics, Icahn School of Medicine at Mount Sinai, NY, USA

[8] Semel Institute for Neuroscience at UCLA, Los Angeles, CA, USA

[9] Departments of Public Health and Preventive Medicine, Family Medicine, and Psychiatry and Behavioral Sciences, State University of New York, Upstate Medical University, Syracuse, NY, USA

[10] Department of Psychiatry, University of North Carolina, Chapel Hill, NC, USA

[11] Department of Psychiatry & Behavioral Sciences, University of Southern California, Los Angeles, CA, USA

[12] Department of Psychiatry, Virginia Commonwealth University School of Medicine, Richmond, VA, USA

[13] Department of Psychiatry and Human Behavior, University of California, Irvine, CA, USA

[14] Department of Psychiatry and Behavioral Sciences, Emory University, Atlanta, GA, USA

[15] Department of Psychiatry, Veterans Administration New York Harbor Healthcare System, Brooklyn, NY, USA

[16] Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, 7 Cambridge Center, Cambridge, MA 02142, USA

[17] Stanley Center for Psychiatric Research, Broad Institute of Harvard and MIT, Cambridge, MA 02142, USA

[18] Department of Genetics, Harvard Medical School, Boston, MA, 02115, USA

For the Inflammatory Bowl Disease Cohort (IBD) cases and controls, we thank: The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) IBD Genetics Consortium (IBDGC) supported by The Helmsley Charitable Trust and the Centers for

Common Disease Genomes Program (NHGRI).

DNA samples were obtained from the following collections: The Lunenfeld-Tanenbaum Research Institute Mount Sinai Hospital (PI: Mark Silverberg), The University of Pittsburgh School of Medicine (PI: Richard Duerr), The Emory University School of Medicine (PI: Subra Kugathasan), The Johns Hopkins Hospital (PI: Steven Brant), The Icahn School of Medicine at Mount Sinai (PI: Judy Cho), The Washington University School of Medicine (PI: Rodney Newberry), The University of Miami Miller School of Medicine (PI: Maria Abreu, Jake McCauley), and Cedars Sinai (PI: Dermot McGovern, Stephan Targan).

2. Supplementary Methods

2.1 Study cohorts

Patients or their legal guardians provided signed informed consent according to local IRB requirements. Samples had been collected over a 20-year period in some centers, so the consent forms reflected standards at the time of collection. Samples were only accepted if the consent did not exclude data sharing. For samples collected after January 25, 2015, consent forms required specific language according to the NIH Genomic Data Sharing Policy (https://osp.od.nih.gov/wp-content/uploads/NIH_GDS_Policy.pdf).

2.2 Controls

2.2.1. Genomic Psychiatry Cohort (GPC) controls

Brief description/reference:

Samples from the Genomic Psychiatry Cohort (GPC) that were contributed to Epi25 were a subset of the overall control participants with no personal or family history of schizophrenia or bipolar disorder. All the samples were genotyped on the GSA-MD v1.0 at the Broad Institute. A detailed description of the GPC can be found here: Pato et al. (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3729260/)

2.2.2. FINRISK controls

Brief description/reference:

FINRISK controls were genotyped on the GSA-MD v1.0 at the Broad Institute. A detailed description of the FINRISK samples can be found here: https://www.ncbi.nlm.nih.gov/pubmed/29165699

Acknowledgments:

The FINRISK controls were part of the FINRISK studies supported by THL (formerly KTL: National Public Health Institute) through budgetary funds from the government, with additional funding from institutions such as the Academy of Finland, the European Union, ministries and national and international foundations and societies to support specific research purposes. Genotyping of FINRISK controls was supported by the Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA.

2.2.3. IBD cases and controls

Brief description/reference:

Samples from the Inflammatory Bowl Disease Cohort (IBD) that were contributed to Epi25 were patients with inflammatory bowl disease. All the samples were genotyped on the GSA-MD v1.0 at the Broad Institute.

2.3 Methods

Intensity Sample QC:

Intensity-based QC was conducted to remove samples with low quality data based on the following empirically defined thresholds across three different metrics: Thresholds

for (1) waviness factor, (2) Log-R ratio standard deviation, and (3) B-allele frequency drift were calculated by taking the median +3x SD to determine outlying samples as performed in Huang et al. (Huang et al., 2017). Following intensity-based QC, all samples had a Log-R ratio standard deviation of < 0.25, absolute value of waviness factor < 0.04, and a B-allele frequency drift < 0.007.

Call Filtering and Delineation of Rare CNVs:

CNV calls were removed from the dataset if they spanned less than 20 markers, were less than 20Kb in length, had a SNP density < 0.0001 (amount of markers/length of CNV) or overlapped by more than 0.5 of their total length with regions known to generate artifacts in SNP-based detection of CNVs (Marshall et al., 2017). This included immunoglobulin domain regions, telomeric regions (defined as 500Kb from the chromosome ends), and centromeric regions (coordinates were provided by PennCNV for hg19). Further, we excluded CNVs overlapping > 80% of regions known to be recurrent copy number variations in the general population (11,732 CNVs from http://dgv.tcag.ca/dgv/app/home) for a part of the analyses (see "CNV Burden Analysis"). Additionally, all CNV calls spanning more than 20 markers and equal to or more than 1Mb in length were included in the analysis even if the SNP density was < 0.0001 (Huang et al., 2017; Marshall et al., 2017).

We assigned all CNV calls a specific frequency count using PLINK v.1.07 (Purcell et al., 2007), with the option --cnv-freq-method2 0.5. Here, the frequency count of an individual CNV is determined as 1 + the total number of CNVs overlap by at least 50% of its total length (in bp), irrespective of CNV type. We then filtered our callset for rare CNVs with a frequency of 186 or lower across all samples).

After CNV quality control, 11,826 of 17,992 (7,425 cases and 4,401 controls) QC-passed individuals had at least rare CNV.

Regression of Potential Confounds on Case-Control Ascertainment:

It is important to ensure that any bias in gender and ancestry does not drive spurious associations with epilepsy. To ensure the robustness of the analysis, CNV burden analyses included potential confounding variables as covariates in a logistic regression framework. Due to the number of tests run at breakpoint level association, we employed a step-wise logistic regression approach to allow for the inclusion of covariates in our case-control association, as previously described in Marshall and Howrigan et al. (Marshall et al., 2017), which we term the epilepsy residual phenotype. Covariates included sex for burden and breakpoint association analysis and the first ten ancestry principal components for breakpoint association analysis.

To calculate the epilepsy residual phenotype, we first fitted a logistic regression model of covariates to affection status, and then extracted the Pearson residual values for use in a quantitative association design for downstream analyses. Residual phenotype values in cases are all above zero, and controls below zero, and are plotted against overall Kb burden in Supplementary Fig. 3.

2.4 PheWAS

List of phenotypes tested in the combined epilepsy phenotype for big CNV (>2Mb) burden: Age of first seizure, Intellectual Disability, Autism, Pharmacoresistance, Psychosis, Abnormal Neurological Examination, Febrile Seizures, Generalized Seizures,

Myoclonic Seizures, Absence Seizures, Atypical Absence Seizures, Tonic Seizures, Atonic Seizures. Phenotypes tested only in focal epilepsy patients: Structural Epilepsy, Aura Seizures, Dyscognitive Seizures, Bilateral Seizures. Phenotypes tested only in epileptic encephalopathy patients: Developmental Delay, Regression, Movement Disorder, CNS Infection, Head Trauma, Status Epilepticus: convulsive, Status Epilepticus: non-convulsive, Spasms Seizures, Hemiclonic Seizures, Bilateral Clonic Seizures, Ohtahara syndrome, Early myoclonic encephalopathy, Early onset epileptic encephalopathy, Infantile onset epileptic encephalopathy, Epilepsy of infancy with migrating focal seizures, West syndrome/infantile spasms, Late-onset epileptic spasms, Lennox-Gastaut syndrome, Epilepsy with myoclonic atonic seizures, Dravet syndrome, Landau-Kleffner syndrome, Epileptic encephalopathy with continuous spike-and-wave during sleep, Febrile Infection Related Epilepsy Syndrome, Hemiconvulsion-Hemiplegia-Epilepsy, Nonsyndromic epileptic encephalopathy with focal seizure, Nonsyndromic epileptic encephalopathy with generalized seizures, Nonsyndromic epileptic encephalopathy with mixed or unclassified seizures.

2.5 Network analysis

Construction of the human protein-protein interactome:

To build a comprehensive human protein-protein interactome, we focused on high-quality protein-protein interactions (PPIs) with five types of experimental evidences: (1) Kinase-substrate interactions by literature-derived low-throughput and high-throughput experiments; (2) Binary PPIs tested by high-throughput yeast-two-hybrid (Y2H) systems (Rual *et al.*, 2005; Rolland *et al.*, 2014); (3) PPIs from protein three-dimensional (3D) structures; (4) Literature-curated PPIs identified by affinity purification followed by mass spectrometry, Y2H and by literature-derived low-throughput experiments; and (5) signaling networks supported by literature-derived low-throughput experiments. The

genes were mapped to their Entrez ID based on the NCBI database (Ncbi Resource Coordinators, 2016) as well as their official gene symbols based on GeneCards (http://www.genecards.org/). Duplicated PPIs and all computationally predicted data, such as evolutionary analysis, metabolic associations, and gene co-expression data, were deleted. The resulting updated human interactome used in this study includes 351,444 protein-protein interactions (PPIs) among 17,706 proteins. The detailed descriptions are provided in our recent studies (Cheng *et al.*, 2018; Cheng *et al.*, 2019a; Cheng *et al.*, 2019b; Smith *et al.*, 2019).

Disease module detection:

Using the human interactome, we examined the largest connected component of the gene lists of the four types of epilepsy phenotypes to see if the genes can form modules. To test the significance, a permutation test was performed for each phenotype gene list using randomly selected genes with similar degree distribution to the ones being examined and repeated 10,000 times. The largest connected components were visualized using Cytoscape 3.7.1. Hub genes were identified by high degree and betweenness centrality. Betweenness centrality indicates the fraction of all the shortest paths that pass through a specific node.

Network proximity analysis:

Network proximities for all pairs of gene lists were calculated using the closest method:

$$\langle d_{AB} \rangle = \frac{1}{||A|| + ||B||} \left(\sum_{a \in A} min_{b \in B}\, d(a,b) + \sum_{b \in B} min_{a \in A}\, d(a,b) \right)$$

where $d(a,b)$ is the shortest path between gene a and b from gene list A and B, respectively. A permutation test was conducted for each pair using two randomly selected

gene lists that have similar degree distributions to the ones being examined and repeated 10,000 times. Z-scores were calculated based on the permutation test.

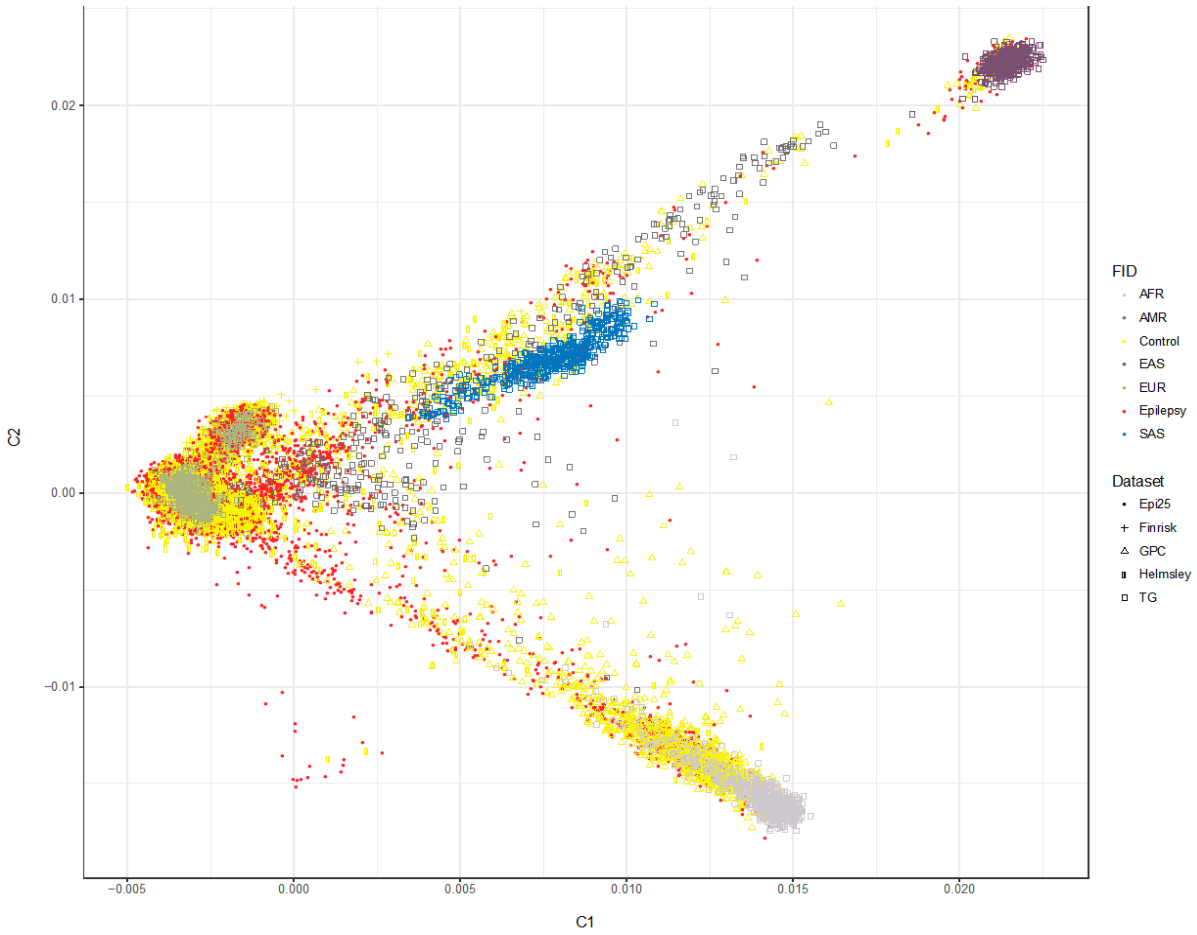
In addition, we also quantified the overlap between gene lists by calculating the overlap coefficient C and Jaccard index J as follows:

$$C = |A \cap B| / \min(|A|, |B|)$$

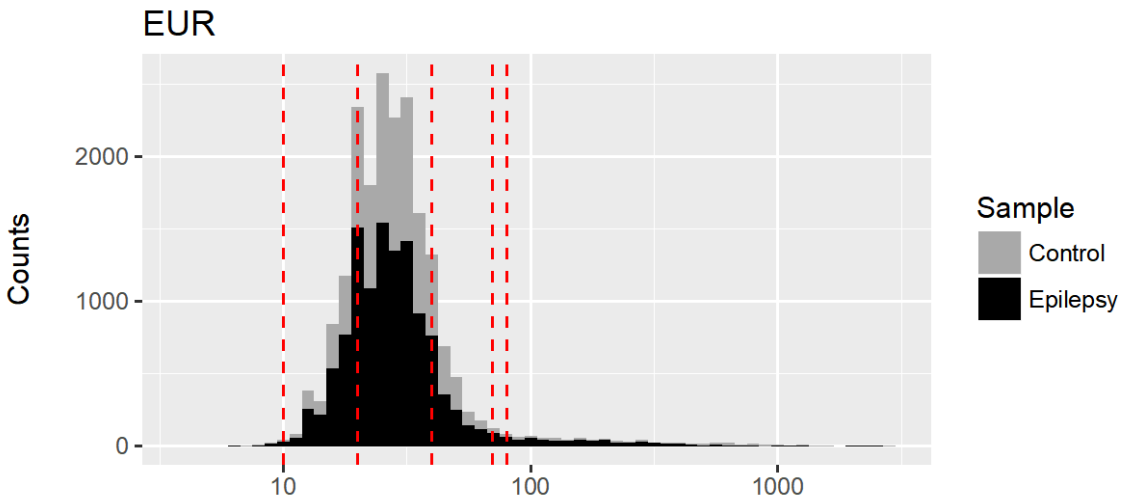
$$J = |A \cap B| / |A \cup B|$$

Both values range from 0 to 1. C can be interpreted as the fraction of the smaller set which is the subset of the larger set. When J and C equal 0, the two sets have no common genes. When $C = 1$, one set is the subset of the other set. When $J = 1$, the two sets of genes are identical.

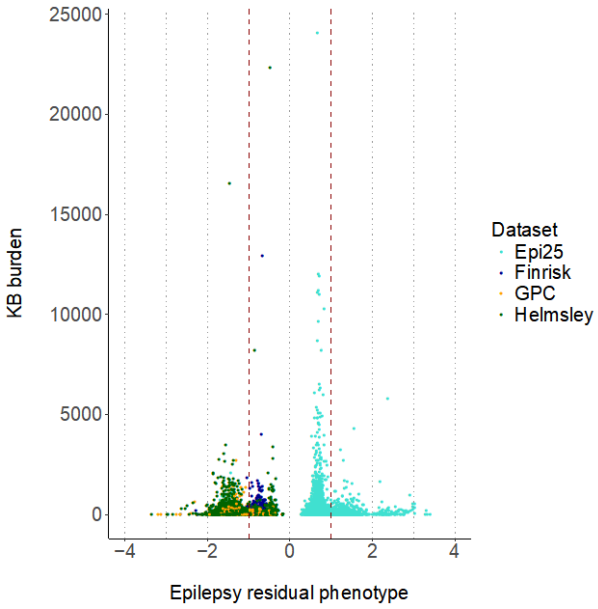
3. Supplementary Figures



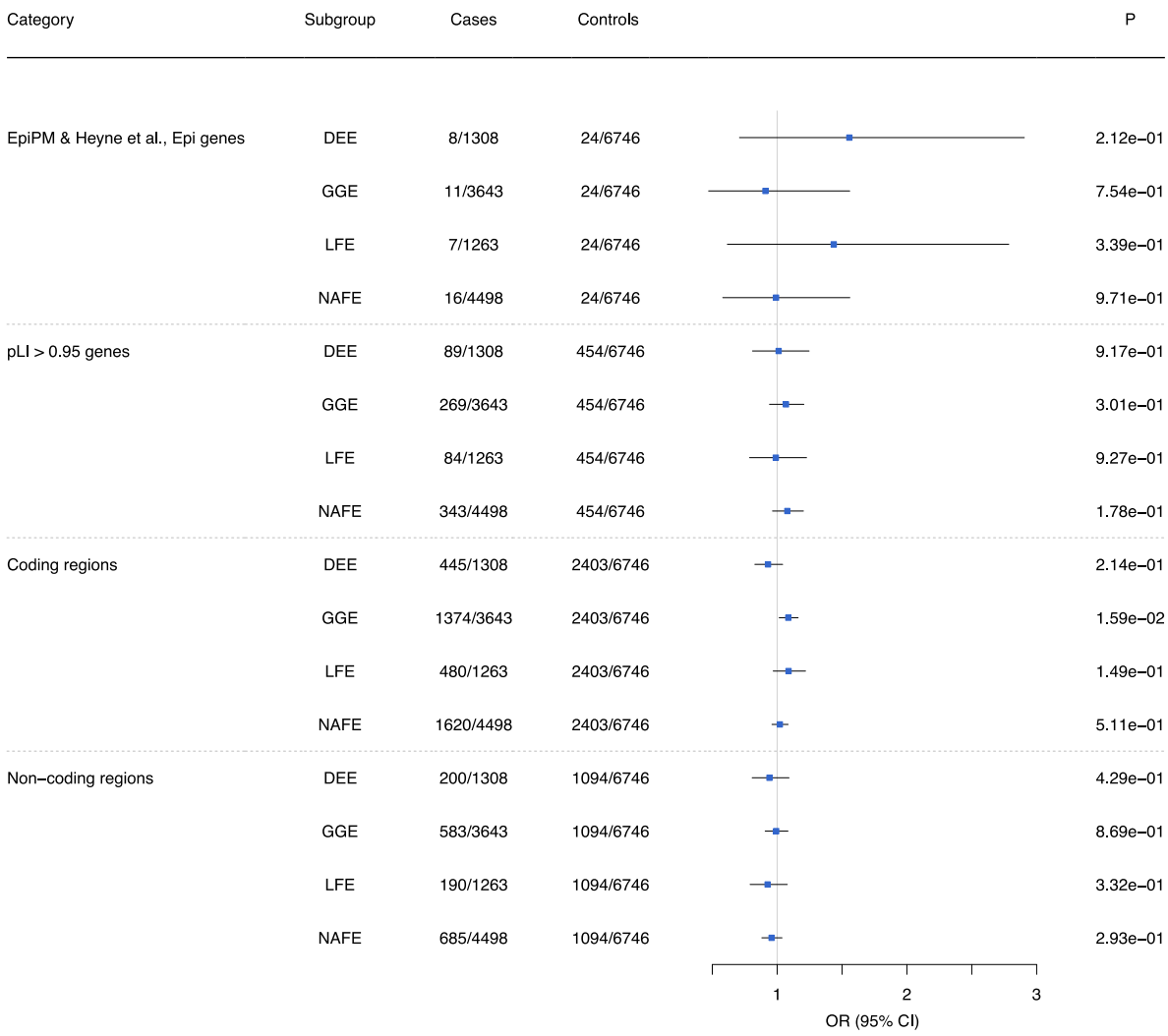
Supplementary Fig. 1 Ancestry principle component analysis of cases and controls combined with the 1000 Genome Populations. Different colors represent different populations with cases in red and controls in yellow. We removed most controls because of non-European ancestry. This affected in particular unaffected controls derived from the Genomic Psychiatry Cohort (GPC). The GPC cohort compromises controls from Caucasian, African American, Latino and other ethnicities. For more details on this cohort see (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3729260/). AFR = Africans; AMR = Americans; EAS = East Asians; EUR = Europeans; SAS = South Asians; TG = The Thousand Genome Project Population.



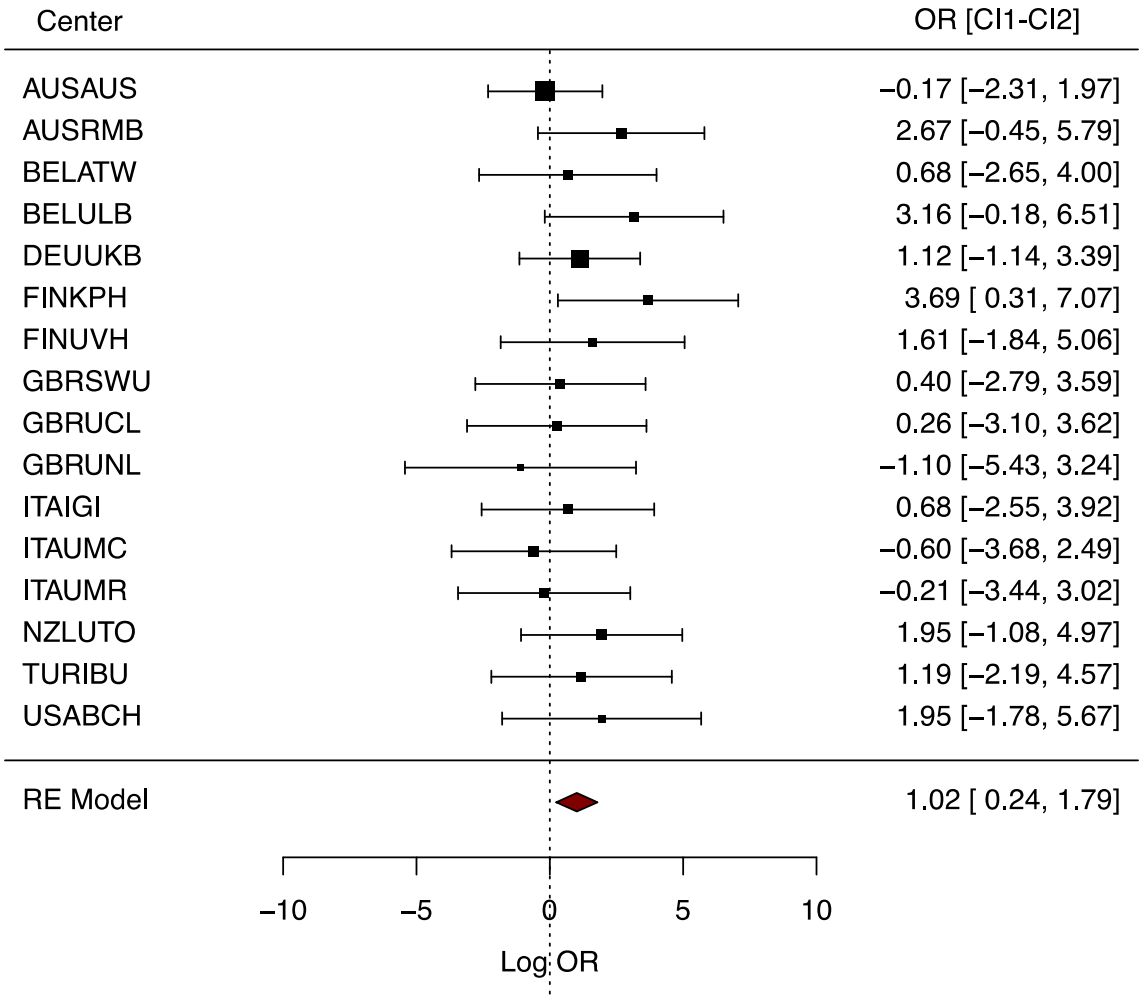
Supplementary Fig. 2 CNV-load per individual. CNV-load of cases (black) and controls (grey) with the number of CNVs per sample on the x-axis and the counts of samples with the specific CNV-load on the y-axis. Red dashed lines are at 10, 30, 60, 80, and 90 CNVs per individual.



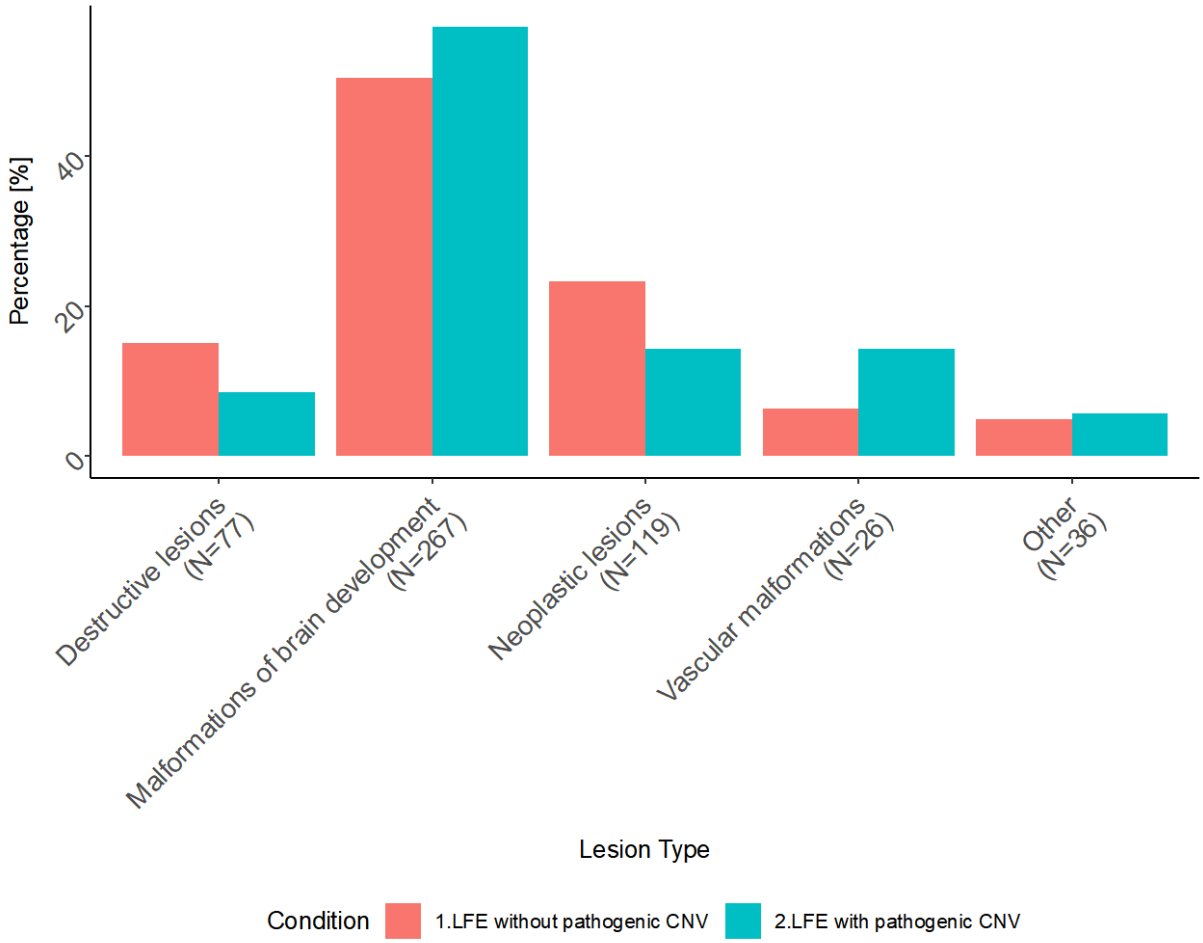
Supplementary Fig. 3 Epilepsy phenotype residual distribution. X-axis: Distribution of phenotype residual values after regressing case/control status on selected covariates. Plotted against overall deletion burden (Y-axis) to visually inspect if large residuals have an excess of deletion burden, which can lead to higher false positive associations. Epilepsy cases have positive residual values and controls negative residual values. Red lines show x-intercepts at -1 and 1, accordingly.



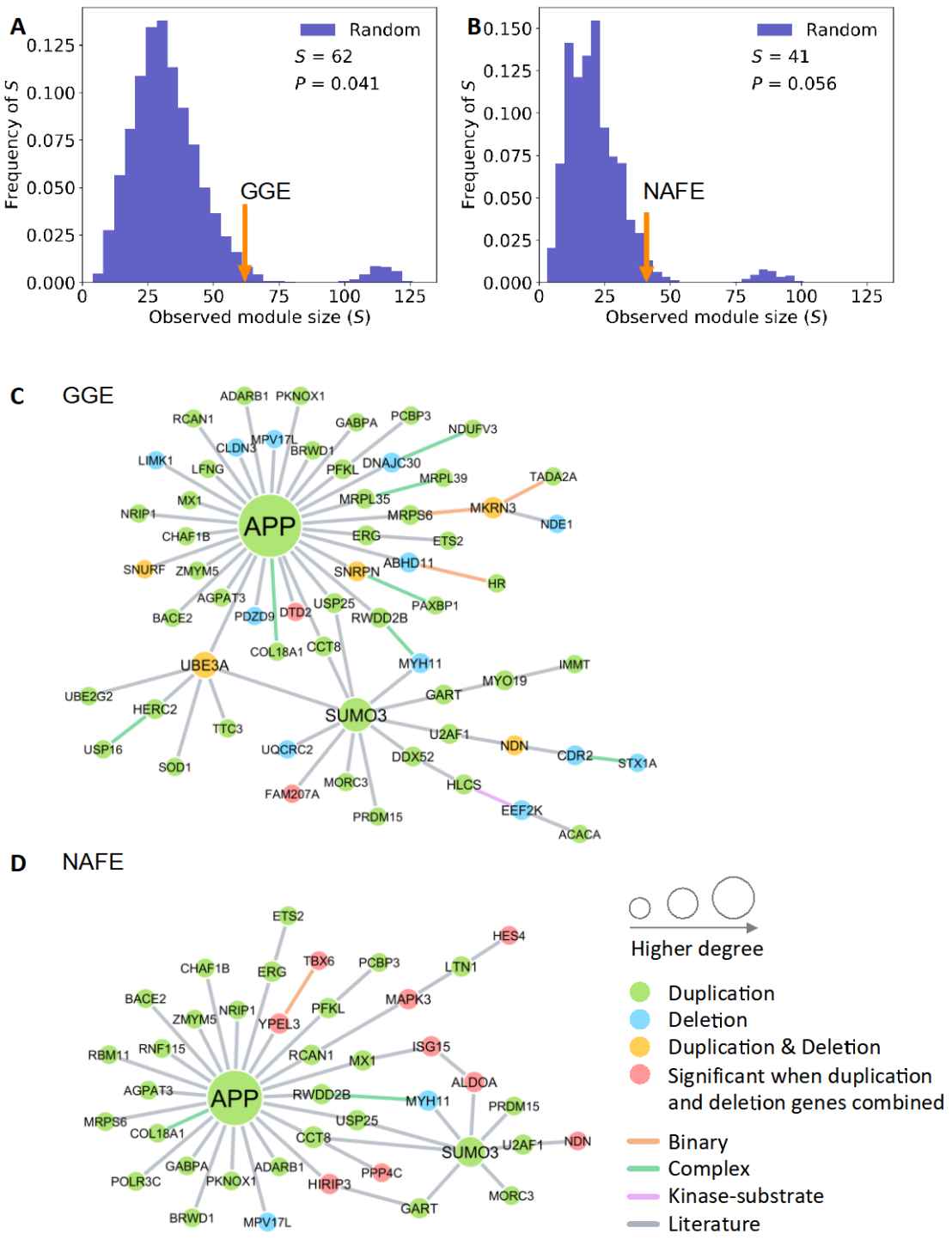
Supplementary Fig. 4 Global duplication burden across different gene sets and non-coding regions in five different epilepsy phenotypes. Rare (<= 1% frequency) duplication burden was elucidated for different gene lists (Category). Odds ratios (ORs) and p-values were calculated using a binomial regression for rare CNVs with sex as a covariate. CNV are defined as "coding" if 80% of their overall length overlaps a gene. *Significance p < 1.4e-3 after Bonferroni correction. 95% CIs are shown. DEE = Developmental and epileptic encephalopathies; GGE = Genetic generalized epilepsies; LFE = Lesional focal epilepsies; NAFE = Non-acquired focal epilepsies.



Supplementary Fig. 5 Meta-Analysis. We performed a meta-analysis across centers which submitted samples to our study to rule out that the duplication enrichment in GGE patients with febrile seizures is due to a phenotyping bias from a single center. Twelve centers had a greater ratio of duplications in patients with GGE and febrile seizures compared to four centers in which we did not observe this enrichment. Overall, in the systematic meta-analysis we observe significant enrichment for the association of large duplications in patients with GGE and febrile seizures and can rule out a single center bias.



Supplementary Fig. 6 Fraction [%] of LFE patients with or without pathogenic CNVs among different lesion types. In a total of 525 patients with LFE and specific lesion information, in 35 patients large duplications were identified compared to 490 without. Across five major lesion types, patients with malformations of brain development carry with 57.41 % the highest fraction of pathogenic CNVs.



Supplementary Fig. 7 GGE and NAFE genes form disease modules in the human interactome. (A) Sixty-two genes from the 167 GGE genes form a disease module (p = 0.041, permutation test of 10,000 repeats). (B) Forty-one genes from the 129 NAFE genes from a disease module (p = 0.056, permutation test of 10,000 repeats). (C) and (D) – disease module

visualization for GGE (C) and NAFE (D). Node size indicates the node degree. Node color indicates whether the gene is in the duplication, the deletion, or both lists. Edge color indicate the type of evidence for the PPI in the interactome (for details see Supplementary Methods).

4. Supplementary Tables

	Epi25	Finrisk	GPC	Helmsley	**TOTAL**
Case	13,420	0	0	0	**13,420**
Control	302	1,569	5,415	5,571	**12,857**

Supplementary Table 1 Dataset summary. All subjects included in the CNV calling before quality control. Cases were taken from one single data set whereas controls were taken from four different data sets.

CNV Type	Epilepsy Type	Phenotype	P-value (*<3.1*10^{-4})	OR	2.5% CI - 97.5% CI
Duplication	GGE	Febrile Seizures	4.07*10^{-5*}	3.25	1.8-5.92
Duplication	FE	Structural Focal Epilepsy	2.33*10^{-4*}	2.72	1.57-4.56
Deletion	LFE	Dyscognitive Seizures	1.6*10^{-3}	0.05	0.001-0.45
Duplication	GGE	Absence Seizures	4.5*10^{-3}	3.23	1.34-9.45
Duplication	LFE	Bilateral Seizures	2.05*10^{-2}	7.51	1.18-314
Deletion	LFE	Bilateral Seizures	2.24*10^{-2}	0.094	0.002-0.96
Duplication	NAFE	Pharmacoresistance	4.64*10^{-2}	0.37	0.12-1.03

Supplementary Table 2 Phenome-wide association study (PheWAS). We tested the burden of big CNVs (>2Mb) in patients with the different epilepsy phenotypes. P-values and ORs were obtained using a two-sided Fisher's Exact Test. Multiple testing correction for 161 test results in a significant p-value < 3.1*10^{-4}. GGE = Genetic Generalized Epilepsy, FE = Focal Epilepsy (including NAFE and LFE), LFE = Lesional Focal Epilepsy, NAFE = Non-acquired Focal Epilepsy, CI = Confidence Interval.

	Likely pathogenic CNV burden	Enriched categories	Genome-wide enriched loci
Developmental and epileptic encephalopathies	1.72x	• >2Mb del • High pLI del	• 15q11.2-q13.1dup

Genetic generalized epilepsy	2.43x	• >2Mb del
		• 500kb−2mb del
		• Epi hotspots
		  • 15q13.3
		  • 16p.13.11
		• High pLI del
		• Coding del
			• 15q13.2-q13.3del
			• 16p13.11del

Lesional focal epilepsy	1.68x	• 16p.13.11	• 16p13.11del
Non-acquired focal epilepsy	1.37x	• Epi hotspots	n.s.

Supplementary Table 3 Summary of all findings obtained in this study. n.s. = not significant.

Z-score	DEE	GGE	LFE	NAFE
DEE	-	-1.19	-1.65	-1.43
GGE	-	-	-0.91	-1.49
LFE	-	-	-	-0.44
NAFE	-	-	-	-

P value	DEE	GGE	LFE	NAFE
DEE	-	0.090	0.046	0.063
GGE	-	-	0.163	0.062
LFE	-	-	-	0.291
NAFE	-	-	-	-

Supplementary Table 4 Network proximities of all pairs of epilepsy phenotypes. Network proximities were represented in z-scores with corresponding p-values based on permutation tests of 10,000 repeats. Three pairs of phenotypes showed small proximities: DEE and LFE (z-score = -1.65, p = 0.046), DEE and NAFE (z-score = -1.43, p = 0.063), and GGE and NAFE (z-score = -1.49, p = 0.062).

Overlap	DEE	GGE	LFE	NAFE
DEE	-	22	3	9
GGE	-	-	8	101
LFE	-	-	-	6
NAFE	-	-	-	-

J	DEE	GGE	LFE	NAFE
DEE	-	0.101	0.033	0.047
GGE	-	-	0.044	0.518
LFE	-	-	-	0.041
NAFE	-	-	-	-

C	DEE	GGE	LFE	NAFE
DEE	-	0.306	0.130	0.125
GGE	-	-	0.348	0.783
LFE	-	-	-	0.261
NAFE	-	-	-	-

Supplementary Table 5 Overlap between the genes of epilepsy phenotypes. The number of overlapping genes for all pairs of epilepsy gene lists are shown. Overlap coefficient C and Jaccard index J were also computed (see method) to show the fraction of overlapping genes and the similarity between the lists. GGE and NAFE showed a large overlap ($C = 0.783$) and similarity (J=0.518).

References

Cheng F, Desai RJ, Handy DE, Wang R, Schneeweiss S, Barabasi AL, et al. Network-based approach to prediction and population-based validation of in silico drug repurposing. Nat Commun 2018; 9: 2691.
Cheng F, Kovacs IA, Barabasi AL. Network-based prediction of drug combinations. Nat Commun 2019a; 10: 1197.
Cheng F, Lu W, Liu C, Fang J, Hou Y, Handy DE, et al. A genomewide positioning systems network algorithm for in silico drug repurposing. Nat Commun 2019b; 10: 3476.
Huang AY, Yu D, Davis LK, Sul JH, Tsetsos F, Ramensky V, et al. Rare copy number variants in NRXN1 and CNTN6 increase risk for tourette syndrome. Neuron 2017; 94: 1101–11 e7.
Marshall CR, Howrigan DP, Merico D, Thiruvahindrapuram B, Wu W, Greer DS, et al. Contribution of copy number variants to schizophrenia from a genome-wide study of 41,321 subjects. Nat Genet 2017; 49: 27–35.
Rolland T, Tasan M, Charloteaux B, Pevzner SJ, Zhong Q, Sahni N, et al. A proteome-scale map of the human interactome network. Cell 2014; 159: 1212–26.
Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, Li N, et al. Towards a proteome-scale map of the human protein-protein interaction network. Nature 2005; 437: 1173–8.
Smith IN, Thacker S, Seyfi M, Cheng F, Eng C. Conformational dynamics and allosteric regulation landscapes of germline PTEN mutations associated with autism compared to those associated with cancer. Am J Hum Genet 2019; 104: 861–78.