Genetic Imbalance Is Associated With Functional Outcome After Ischemic Stroke

Dorothea Pfeiffer, MD*; Bowang Chen, PhD*; Kristina Schlicht, PhD*; Philip Ginsbach, MSc; Sherine Abboud, MD, PhD; Anna Bersano, MD, PhD; Steve Bevan, PhD, BSc; Tobias Brandt, MD, PhD; Valeria Caso, MD, PhD; Stéphanie Debette, MD, PhD; Philipp Erhart, MD; Sandra Freitag-Wolf, PhD; Giacomo Giacalone, MD; Armin J. Grau, MD, PhD; Eyad Hayani, MD; Christina Jern, MD, PhD; Jordi Jiménez-Conde, MD, PhD; Manja Kloss, MD; Michael Krawczak, PhD; Jin-Moo Lee, MD, PhD; Robin Lemmens, MD, PhD; Didier Leys, MD, PhD; Christoph Lichy, MD; Jane M. Maguire, RN, PhD; Juan J. Martin, MD; Antti J. Metso, MD, PhD; Tiina M. Metso, MD, PhD; Braxton D. Mitchell, PhD; Alessandro Pezzini, MD; Jonathan Rosand, MD, MSc; Natalia S. Rost, MD, MPH; Martin Stenman, MD; Turgut Tatlisumak, MD, PhD; Vincent Thijs, MD, PhD; Emmanuel Touzé, MD, PhD; Christopher Traenka, MD; Inge Werner; Daniel Woo, MD, MSc; Elisabetta Del Zotto, MD, PhD; Stefan T. Engelter, MD; Steven J. Kittner, MD, MPH; John W. Cole, MD, MS; Caspar Grond-Ginsbach, PhD; Philippe A. Lyrer, MD; Arne Lindgren, MD, PhD; on behalf of CADISP; GISCOME; SiGN studies; and ISGC

Background and Purpose—We sought to explore the effect of genetic imbalance on functional outcome after ischemic stroke (IS).

Methods—Copy number variation was identified in high-density single-nucleotide polymorphism microarray data of IS patients from the CADISP (Cervical Artery Dissection and Ischemic Stroke Patients) and SiGN (Stroke Genetics Network)/GISCOME (Genetics of Ischaemic Stroke Functional Outcome) networks. Genetic imbalance, defined as total number of protein-coding genes affected by copy number variations in an individual, was compared between patients with favorable (modified Rankin Scale score of 0–2) and unfavorable (modified Rankin Scale score of ≥3) outcome after

Received April 25, 2018; final revision received November 26, 2018; accepted December 5, 2018.

From the Department of Neurology, Heidelberg University Hospital, Germany (D.P., T.B., E.H., M. Kloss, I.W., C.G.-G.); Department of Biology, Southern University of Science and Technology, Shenzhen, China (B.C.); Institute of Medical Informatics and Statistics, Kiel University, University Hospital Schleswig-Holstein Campus Kiel, Germany (K.S., S.F.-W., M. Krawczak); School of Informatics, University of Edinburgh, United Kingdom (P.G.); Laboratory of Experimental Neurology, Université Libre de Bruxelles, Brussels, Belgium (S.A.); Cerebrovascular Unit IRCCS Foundation C. Besta Neurological Institute, Milan, Italy (A.B.); School of Life Science, University of Lincoln, United Kingdom (S.B.); Suva/Swiss National Accident Insurance Fund, Lucerne, Switzerland (T.B.); Stroke Unit, Perugia University Hospital, Italy (V.C.); Inserm, Bordeaux Population Health Research Center, UMR 1219, University of Bordeaux, France (S.D.); Department of Neurology, Bordeaux University Hospital, France (S.D.); Department of Vascular and Endovascular Surgery, University Hospital Heidelberg, Germany (P.E.); Department of Neurology, San Raffaele University Hospital, Milan, Italy (G.G.); Department of Neurology, Klinikum Ludwigshafen, Germany (A.J.G.); The Sahlgrenska Academy, University of Gothenburg, Sweden (C.J.); Sahlgrenska University Hospital, Sweden (C.J.); IMIM-Parc de Salut Mar, Universitat Autonoma de Barcelona, Spain (J.J.-C.); Department of Neurology, Washington University School of Medicine, St Louis, MO (J.-M.L.); Department of Neurosciences, Experimental Neurology, and Leuven Brain Institute, KU Leuven, University of Leuven, Belgium (R.L.); VIB, Center for Brain & Disease Research, Laboratory of Neurobiology, Leuven, Belgium (R.L.); Department of Neurology, University Hospitals Leuven, Belgium (R.L.); Department of Neurology, University of Lille, France (D.L.); Department of Neurology, Klinikum Memmingen, Germany (C.L.); Faculty of Health, University of Technology Sydney, Australia (J.M.M.); Hunter Medical Research Institute, Priority Research Centre for Stroke and Traumatic Brain Injury, University of Newcastle, Australia (J.M.M.); Department of Neurology, Sanatorio Allende, Cordoba, Argentina (J.J.M.); Department of Neurology, Helsinki University Central Hospital, Finland (A.J.M., T.M.M., T.T.); Division of Endocrinology, Diabetes and Nutrition, Department of Medicine, University of Maryland School of Medicine, Baltimore (B.D.M.); Geriatrics Research and Education Clinical Center, Baltimore Veterans Administration Medical Center, Baltimore, MD (B.D.M.); Department of Clinical and Experimental Sciences, Neurology Clinic, University of Brescia, Italy (A.P., E.D.Z.); Center for Genomic Medicine (J.R.) and Department of Neurology (N.S.R.), Massachusetts General Hospital, Boston; Department of Clinical Sciences Lund, Neurology, Lund University, Sweden (M.S., A.L.); Department of Neurology and Rehabilitation Medicine, Skåne University Hospital, Lund, Sweden (M.S., A.L.); Institute of Neuroscience and Physiology, The Sahlgrenska Academy, University of Gothenburg, Sweden (T.T.); Department of Neurology, Sahlgrenska University Hospital, Gothenburg, Sweden (T.T.); Stroke Division, Florey Institute of Neuroscience and Mental Health, University of Melbourne, Heidelberg, Victoria, Australia (V.T.); Department of Neurology, Austin Health, Heidelberg, Victoria, Australia (V.T.); Paris Descartes University, INSERM UMR S894, Department of Neurology, Sainte-Anne Hospital, Paris, France (E.T.); Normandie Université, Université Caen-Normandie, Inserm U1237, CHU Côte de Nacre, Service de Neurologie, Caen, France (E.T.); Department of Neurology and Stroke Center, University Hospital Basel, Switzerland (C.T., S.T.E., P.A.L.); Department of Neurology and Rehabilitation Medicine, University of Cincinnati, OH (D.W.); Neurorehabilitation Unit, University of Basel, Switzerland (S.T.E.); University Center for Medicine of Aging, Felix Platter Hospital, Basel, Switzerland (S.T.E.); Department of Neurology, Veterans Affairs Medical Center, Baltimore, MD (S.J.K., J.W.C.); and Department of Neurology University of Maryland School of Medicine, Baltimore (S.J.K., J.W.C.).

*Drs Pfeiffer, Chen, and Schlicht contributed equally.

Presented in part at the International Stroke Genetics Consortium Workshops, Milan, Italy, November 3, 2016 and Kyoto, Japan, April 12, 2018.

The online-only Data Supplement is available with this article at https://www.ahajournals.org/doi/suppl/10.1161/STROKEAHA.118.021856.

Correspondence to Arne Lindgren, MD, PhD, Department of Neurology, Skåne University Hospital, SE-22185, Lund, Sweden. Email arne.lindgren@med.lu.se

3 months. Subgroup analyses were confined to patients with imbalance affecting ohnologs—a class of dose-sensitive genes, or to those with imbalance not affecting ohnologs. The association of imbalance with outcome was analyzed by logistic regression analysis, adjusted for age, sex, stroke subtype, stroke severity, and ancestry.

Results—The study sample comprised 816 CADISP patients (age 44.2±10.3 years) and 2498 SiGN/GISCOME patients (age 67.7±14.2 years). Outcome was unfavorable in 122 CADISP and 889 SiGN/GISCOME patients. Multivariate logistic regression analysis revealed that increased genetic imbalance was associated with less favorable outcome in both samples (CADISP: *P*=0.0007; odds ratio=0.89; 95% CI, 0.82–0.95 and SiGN/GISCOME: *P*=0.0036; odds ratio=0.94; 95% CI, 0.91–0.98). The association was independent of age, sex, stroke severity on admission, stroke subtype, and ancestry. On subgroup analysis, imbalance affecting ohnologs was associated with outcome (CADISP: odds ratio=0.88; 95% CI, 0.80–0.95 and SiGN/GISCOME: odds ratio=0.93; 95% CI, 0.89–0.98) whereas imbalance without ohnologs lacked such an association.

Conclusions—Increased genetic imbalance was associated with poorer functional outcome after IS in both study populations. Subgroup analysis revealed that this association was driven by presence of ohnologs in the respective copy number variations, suggesting a causal role of the deleterious effects of genetic imbalance. **(*Stroke*. 2019;50:298-304. DOI: 10.1161/STROKEAHA.118.021856.)**

Key Words: DNA copy number variations ■ genetics ■ polymorphism, single nucleotide ■ prognosis ■ stroke

Stroke is a major cause of disability and death in adults. Although a substantial proportion of stroke risk still remains unexplained, a genetic component is supported by family studies and genome-wide association studies (GWAS).[1] In a minority of patients, stroke may be explained by rare Mendelian mutations but usually the disease results from complex patterns of modifiable and nonmodifiable risk factors, including multiple genetic variants of small effect. The advent of high-throughput genotyping has led to discovery of new genes related to complex forms of stroke.[2] In some small studies, outcome after stroke was associated with common alleles of a few candidate genes (*BDNF*, *GPIIIa*, and *COX2*).[3] However, large genome-wide searches of factors predicting outcome after ischemic stroke (IS) are still pending.

Structural genomic variation such as copy number variation (CNV) is increasingly recognized playing a role in many pathological conditions, including vascular diseases.[4–6] CNVs are widespread in the human genome and can be identified by means of single-nucleotide polymorphism (SNP) microarray platforms used for GWAS. However, the clinical interpretation of CNVs is challenging. Large CNVs (>500 kb) with low population frequency (<1%) are more likely to be deleterious than frequent CNVs of small size. Moreover, the gene content of a CNV appears to matter more than its mere physical length, particularly the total number of genes within the CNV and the function of these genes (eg, protein-coding versus noncoding, dosage-sensitivity versus dosage-insensitivity of gene products, number of interaction partners of the encoded proteins).[7,8] The concept of genetic imbalance, defined as the total number of protein-coding genes affected by CNVs in an individual, was introduced for the analysis of highly heterogeneous and complex phenotypes, including mental retardation, schizophrenia, or autism spectrum disorder.[9] These complex phenotypes have been related to quantitative variation in genomic content across different chromosomal regions, rather than sequence variation at specific candidate loci. Because outcome after ischemic stroke is a highly complex phenotype that depends on a variety of factors, including age, sex, stroke severity, frailty, occurrence of complications or new strokes, comorbidities, and socio-economic conditions, it appears worthwhile to explore the potential of genetic imbalance as an additional outcome predictor after IS.

The ability to determine the pathological relevance of a particular CNV is usually limited by sample size and lack of sufficient control data, particularly for low-frequency CNVs. One way to overcome these limitations would be to classify genetic imbalance as benign or (possibly) deleterious via the presence of ohnolog genes (ohnologs) within the regions of genetic imbalance. Ohnologs are a class of genes named after the Japanese-American geneticist Susumu Ohno. Ohnologs are supposed to be remnants of 2 complete genome duplications in early vertebrate evolution. They are overrepresented in pathogenic CNVs.[10–12] In the present study, we, therefore, also classified CNVs according to whether at least one ohnolog was among the protein-coding genes overlapping the region(s) of genetic imbalance.

Here we reanalyzed rare CNVs in patients from the CADISP study (Cervical Artery Dissection and Ischemic Stroke Patients), relating these variants, recently explored about their potential as genetic risk factors for cervical artery dissection (CeAD),[13] with outcome after IS because of CeAD or other causes. In an additional sample of microarray data from IS patients of various stroke subtypes, enrolled by the SiGN (Stroke Genetics Network)/GISCOME (Genetics of Ischaemic Stroke Functional Outcome) network,[14,15] CNV analysis was performed to validate any findings in the CADISP sample. We also classified CNVs as ohnolog-positive or ohnolog-negative and assessed whether the observed association between genetic imbalance and outcome was driven by ohnolog-positivity.

Methods

The data that support the findings of this study are available from the corresponding author on reasonable request.

CADISP Study Sample

A total of 983 patients with CeAD diagnosis, based on criteria widely accepted in the stroke community, were included in the CADISP study between 2004 and 2009.[16] In addition, 658 patients with IS attributable to causes other than CeAD (non-CeAD-patients) were enrolled. All patients were self-reported Europeans of white ancestry. DNA was genotyped with Illumina Human 610-Quad or Human 660W-Quad Bead Chips and analyzed in GWAS.[13,17] Rare CNVs had been analyzed before in 833 CADISP IS patients who were nondisabled before stroke (ie, premorbid modified Rankin Scale [mRS] score of 0) and with microarray data of sufficient quality and complete documentation of sex, age, stroke severity on admission available. Seventeen patients were excluded from the current study because of missing information on outcome. CNV data from the remaining 816 CADISP patients were analyzed about association with functional outcome.

GISCOME Study Sample

The SiGN/GISCOME study was described in detail elsewhere.[15] In short, GISCOME recruited 8831 IS patients with genotype and outcome data to examine the relationship between the 2. For the present study, SiGN/GISCOME individuals were included if they had been genotyped in the SiGN GWAS,[14] using the Illumina Omni 5M genotype platform, and if initial stroke severity information according to National Institutes of Health Stroke Scale and outcome data at 60 to 190 days according to mRS were available. For quality reasons, 165 cases with >20 CNV calls by PennCNV (n=127) and/or mosaicism (n=69) were excluded. Thus, of the 2663 cases with mRS, National Institutes of Health Stroke Scale, and TOAST (Trial of ORG 10172 in Acute Stroke Treatment) stroke sub classification/Causative Classification of Stroke system data available, from 6 different study centers, 2498 (93.8%) were included in the final SiGN/GISCOME sample (Table I in the online-only Data Supplement). Excluded and nonexcluded patients did not differ about baseline characteristics, IS subtype, stroke severity on admission, or functional outcome 3 months after stroke (Table II in the online-only Data Supplement).

CNV Analysis

Genetic imbalance was identified as described elsewhere.[5,6,18] Briefly, after PennCNV analysis of normalized GWAS microarray data to identify putative CNVs, and after rejection of low-quality results or small candidate variants (comprising <20 SNPs, for 610/660K microarrays and <100 SNPs, for 5M microarrays), all calls were validated by visual inspection after noise reduction (details in the online-only Data Supplement note on CNV validation). CNVs were classified as genic if they comprised the deletion of at least one coding exon, or a duplication that encompassed either an entire coding region or internal exons. In the CADISP population, only rare CNVs were analyzed (ie, <3 findings among the 3703 disease-free subjects in 2 high-quality CNV databases[6,8]). The CNV findings from the SiGN/GISCOME population were analyzed without frequency filtering. However, for consistency, an additional analysis was performed on rare CNVs only. The genetic imbalance level of an individual was defined as the total number of protein-coding genes affected by CNVs. Finally, we identified all strict ohnologs among the imbalanced genes according to the Ohnologs Browser (http://ohnologs.curie.fr/cgi-bin/BrowsePage.cgi) and categorized imbalance as either including or not-including at least one strict ohnolog.

Baseline Characteristics of Patients

The following clinical variables were included in the analysis: age, sex, ischemic stroke subtype (CADISP: CeAD versus non-CeAD; SiGN/GISCOME: TOAST stroke sub classification subtype, except Lund, Sweden, where IS subtype was determined according to the Causative Classification of Stroke system), and stroke severity on admission (assessed by National Institutes of Health Stroke Scale).

Outcome Evaluation

Functional outcome after 3 to 6 months in the CADISP sample and after 2 to 5 months in the SiGN/GISCOME sample was assessed by the mRS. Outcome was dichotomized for all analyses as favorable (mRS score of 0, 1, or 2) or unfavorable (mRS score of 3, 4, 5, or 6).

Principal Components Analysis

Patients with outlier positions in an ancestry-informative principal components analysis were removed in previous quality control steps. A second principal components analysis was performed on 50 000 randomly chosen SNPs to adjust the logistic regression models for ancestry. The 10 major principal components were used as potential confounders.

Statistical Analyses

In both studies, patients with favorable and unfavorable outcome were compared about sex, age, stroke subtype, stroke severity, genetic imbalance level, and country of recruitment by using χ^2 test, Student t test, or Mann-Whitney U test, as appropriate (univariate analysis). Logistic regression models were used to analyze the association between favorable outcome and genetic imbalance level (multivariate analysis). Results were expressed as odds ratios (OR) per gene affected by a CNV, with 95% CI. In the main analysis, the individual-specific number of protein-coding genes affected by CNVs (ie, the level of genetic imbalance) was included as an independent variable in the model, which was then adjusted for sex, age and the first 10 principal components as possible confounders, and included stroke subtype and stroke severity as additional, potentially relevant covariates. To investigate the impact of genetic imbalance in more detail, we also performed a subgroup analysis distinguishing between patients carrying, or not carrying, at least 1 CNV with an ohnolog. In another subgroup analysis, we compared patients with ohnolog-negative imbalance with the remaining patients.

To explore the validity of the statistical approach taken, and to cover different aspects of the association between outcome and genetic imbalance, we performed additional logistic regression analyses. First, we analyzed the association only between outcome and exceptionally large imbalances, defined as comprising 3 or more protein-coding genes (corresponding to the upper 5% of imbalances observed in our study), as compared with subjects without large imbalances. We also used propensity scores to cover more comprehensively the influence of potential confounders and adjusted the original analysis also for hypertension, diabetes mellitus, and smoking status for CADISP and, additionally, atrial fibrillation for SiGN/GISCOME. In the SiGN/GISCOME population, an analysis with rare variants only was performed as well. Detailed descriptions and summary of the results of the additional analyses are provided in Table III in the online-only Data Supplement. All models were evaluated by pseudo-R^2 values using the Cox and Snell method. Statistical analyses were performed with R and SPSS 19.0 statistics software packages.

Ethics

The study protocol was approved by relevant local authorities in all participating centers and complied with national regulations concerning ethics committee approval and informed consent.

Results

Figure I in the online-only Data Supplement illustrates the highly dispersed chromosomal localization of CNVs, both for patients with favorable (upper) and unfavorable outcome (lower) in both study populations. Most CNVs were rare. For example, the most frequent recurrent finding (on chromosome 22) was observed in 148 SiGN/GISCOME patients (5.9%). X-chromosomal imbalances were not analyzed in the SiGN/GISCOME population. In the CADISP sample, CNVs had been frequency filtered in an earlier investigation and only

Table 1. Predictors of Favorable Outcome After Stroke in the CADISP Cohort

Predictor	Outcome		OR	95% CI	*P* Value	
	Unfavorable (n=122)	Favorable (n=694)			Univariate	Multivariate
Female sex (n)*	40 (32.8)	297 (42.8)	1.35	0.79–2.36	0.046	0.28
Age (mean±SD)	46.4±8.5	43.8±10.6	0.97	0.95–0.99	0.002	0.040
CeAD cause (n)*	70 (57.4)	323 (46.5)	1.11	0.65–1.89	0.031	0.71
NIHSS (median)†	14 [0–40]	2 [0–24]	0.81	0.78–0.84	<0.001	<0.001
Imbalance (median/mean)†	0/0.94 [0–29]	0/0.51 [0–28]	0.89	0.82–0.95	0.24	0.001
Imbalance with ohnologs (median/mean)†	0/0.57 [0–29]	0/0.23 [0–28]	0.88	0.80–0.95	0.056	0.002
Imbalance without ohnologs (median/mean)†	0/0.30 [0–8]	0/0.25 [0–25]	0.93	0.80–1.18	0.94	0.42

The association between functional outcome and different types of genetic imbalance was assessed by multivariate logistic regression analysis (model 1: continuous genetic imbalance), each time adjusted for age, sex, and ancestry-derived principal components 1–10 as potential confounders, and including stroke subtype (CeAD vs non-CeAD) and stroke severity (NIHSS) as additional covariates. Univariate *P* values obtained by nonmodel-based methods: χ^2 test, Student *t* test, or Mann-Whitney *U* test, as appropriate. CADISP indicates Cervical Artery Dissections and Ischemic Stroke Patients; CeAD, cervical artery dissection; NIHSS, National Institutes of Health Stroke Scale; OR, model-adjusted odds ratio of favorable outcome.

*Percentage in brackets.

†Range in square brackets.

findings with a minor allele frequency ≤0.1% were available for the current study.

The CADISP population included 816 young IS patients (age 44.2±10.3 years). Univariate analyses revealed that patients with unfavorable outcome (n=122) were older than those with favorable outcome (mean age 46.4 versus 43.8 years; *P*=0.002), less often female (32.8% versus 42.8%, *P*=0.046), and had more severe strokes (Table 1). The mean genetic imbalance level of patients with unfavorable outcome was larger than for patients with favorable outcome, but the difference was not statistically significant (univariate *P*=0.24). Logistic regression analysis involving multiple outcome predictors revealed that the negative association between favorable outcome and genetic imbalance level was significant (*P*=0.001; OR=0.89; 95% CI, 0.82–0.95) and independent of stroke subtype, stroke severity, age, sex, and center of recruitment (Figure; Table 1). Adjustment of the outcome analysis with additional risk factors (hypertension, diabetes mellitus, and smoking status for CADISP and additionally also atrial fibrillation for SiGN/GISCOME) did not significantly affect the observed association between outcome and genetic imbalance (Table III in the online-only Data Supplement, Model 6).

For validation of the association between genetic imbalance and outcome, we analyzed 2498 IS patients from the SiGN/GISCOME population. Patients with unfavorable outcome in SiGN/GISCOME were also older (mean age 73.7 versus 64.3 years, *P*=0.002) and had more severe strokes but were more often female than patients with favorable outcome (51.6% versus 36.0%, *P*<0.001).

The association between genetic imbalance and outcome was replicated in the SiGN/GISCOME cohort (*P*=0.004; OR=0.94; 95% CI, 0.91–0.98; Figure; Table 2). The variables included in the regression models explained 41% and 37% of the outcome variance in CADISP and SiGN/GISCOME, respectively (Cox and Snell pseudo-R2). Genetic imbalance was not associated with age (Spearman correlation coefficients: CADISP: ρ=−0.061, *P*=0.08; and SiGN/GISCOME: ρ=−0.020, *P*=0.32). Stroke subtype was not significantly associated with genetic imbalance (Kruskal-Wallis test; *P*=0.69). When the logistic regression analysis was stratified by TOAST stroke sub classification/Causative Classification of Stroke subtype, imbalance was significantly associated with stroke outcome in the Cardio-Embolic stroke subgroup (n=830) but not in the subgroups with large vessel disease, small vessel disease, or stroke of other or undetermined causes (Table IV in the online-only Data Supplement). The additional models focusing on large imbalances only (CADISP *P*<0.001 and SiGN/GISCOME *P*=0.017) or using propensity scores (CADISP *P*=0.002 and SiGN/GISCOME *P*=0.012) also yielded significant associations between imbalance level and outcome in both study groups (Table III in the online-only Data Supplement). An additional analysis with rare variants only, performed in the SiGN/GISCOME population, yielded similar results (OR=0.94; 95% CI, 0.91–0.98; *P*=0.006).

Table V in the online-only Data Supplement provides an overview of the number of CNVs in patients with respective outcome. No significant associations of individual CNVs with stroke outcome were found after Bonferroni corrections, although 2 CNVs (a duplication in chromosome 7p encompassing the whole Aryl hydrocarbon receptor [*AHR*] gene and a duplication in chromosome 17 including *GPR142, RPL38, TTYH2, DNAI2, CD300A, GPRC5C, KIF19, GPR142, BTBD17, CD300LB, CD300E, CD300C,* and *CD300LD*) were nominally more common (Fisher exact test *P*=0.001, without correction for multiple testing) in patients with unfavorable outcome (Figure II in the online-only Data Supplement).

We also assessed the effect of ohnolog-positivity (Figure; Tables 1 and 2). In both cohorts, confining the analysis to ohnolog-positive CNVs replicated the overall imbalance-outcome association (CADISP *P*=0.002, OR=0.88; 95% CI, 0.80–0.95 and SiGN/GISCOME *P*=0.002, OR=0.93; 95% CI, 0.89–0.98) whereas analysis of ohnolog-negative CNVs

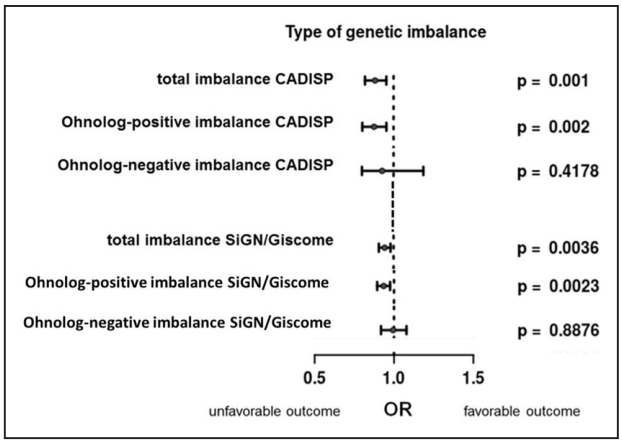

Figure. Odds ratios (OR) for favorable outcome after stroke for different types of genetic imbalance. The association between outcome and different types of genetic imbalance was assessed by logistic regression analysis, adjusted for age, sex, and the first 10 ancestry-derived principal components as potential confounders, and including stroke subtype and stroke severity (National Institutes of Health Stroke Scale) as additional covariates. CADISP indicates Cervical Artery Dissections and Ischemic Stroke Patients; GISCOME, Genetics of Ischaemic Stroke Functional Outcome; and SiGN, Stroke Genetics Network.

imbalance did not reveal such an association (CADISP P=0.42 and SiGN/GISCOME P=0.89).

Discussion

The present explorative study of genetic imbalance and outcome in patients with IS yielded the following key findings: (1) the risk of unfavorable outcome increased with the number of protein-coding genes involved in genetic imbalance; (2) the observed association between genetic imbalance and outcome was independent of age, sex, stroke subtype, and stroke severity; and (3) the association between outcome and ohnolog-positive, putatively pathogenic imbalance was statistically significant, whereas that with ohnolog-negative imbalance was not.

The choice of microarray platform, CNV detection algorithm, filtering strategy, and CNV validation method may all affect the results of genetic imbalance studies.[8,19,20] We previously tested different algorithms for CNV detection, including PennCNV, QuantiSNP, Birdsuite, and software from the Affymetrix Genotyping Console.[5] In our clinical DNA samples of variable quality, all detection algorithms were underperforming, which motivated us to systematically study false CNV detection and noise in SNP microarray data and develop new protocols for high throughput CNV validation after noise reduction.[18] Extensive manual curation of CNV findings was performed in the present study, even after the exclusion of low-quality samples. Many CNV findings were recurrent and strongly overlapped with CNVs reported in public databases. Moreover, some of the CNV findings from the CADISP samples have been validated elsewhere, using independent molecular methods, including quantitative polymerase chain reaction and sequencing of breakpoint-joining polymerase chain reaction fragments.[6]

The above notwithstanding, CNV detection in SNP microarray data is challenging because these platforms were not primarily designed for CNV detection. Also, regions with segmental duplications are rich in CNV, but these regions are underrepresented in microarray platforms and hard to cover by next generation sequencing. These shortcomings may explain why some common CNVs with potential impact on IS[21,22] were not identified in the current study. Moreover, SNPs within regions of segmental duplication or common CNVs are likely to violate Hardy Weinberg equilibrium[23,24] and may, therefore, be rare among microarray platform probe sets.

Information on premorbid mRS, cardiovascular risk factors, and complications during acute hospitalization and early follow-up in the CADISP and SiGN/GISCOME studies were incomplete, which is a limitation of our study. Furthermore, we were unable to analyze mRS on a continuous scale since this information was missing for some centers. Stratification by type of unfavorable outcome, for

Table 2. Predictors of Favorable Outcome After Stroke in the SiGN/GISCOME Cohort

Predictor	Outcome		OR	95% CI	*P* Value	
	Unfavorable (n=889)	Favorable (n=1609)			Univariate	Multivariate
Female sex (n)*	459 (51.6)	580 (36.0)	8.11	0.75–96.33	<0.001	0.081
Age (mean±SD)	73.7±12.9	64.3±13.7	0.94	0.93–0.95	<0.001	<0.001
TOAST/CCS subtype	n.d.	n.d.	1.01	0.94–1.09	<0.001	0.71
NIHSS (median)†	7 [0–41]	3 [0–30]	0.82	0.80–0.84	<0.001	<0.001
Imbalance (median/mean)†	0/1.18 [0–48]	0/0.90 [0–27]	0.94	0.91–0.98	0.91	0.0036
Imbalance with ohnologs (median/mean)†	0/0.69 [0–48]	0/0.37 [0–27]	0.93	0.89–0.98	0.093	0.002
Imbalance without ohnologs (median/mean)†	0/0.49 [0–14]	0/0.53 [0–13]	0.99	0.92–1.07	0.19	0.89

The association between functional outcome and different types of genetic imbalance was assessed by multivariate logistic regression analysis (model 1: continuous genetic imbalance), adjusted for age, sex, and ancestry-derived principal components 1–10 as potential confounders, and including stroke subtype (TOAST/CCS) and stroke severity (NIHSS) as additional covariates. Univariate *P* values obtained by nonmodel-based methods: χ^2 test, Student *t* test, or Mann-Whitney *U* test, as appropriate. CCS indicates Causative Classification System; GISCOME, Genetics of Ischaemic Stroke Functional Outcome; n.d., not done; NIHSS, National Institutes of Health Stroke Scale; OR, model-adjusted odds ratio of favorable outcome; SiGN, Stroke Genetics Network; and TOAST, Trial of ORG 10172 in Acute Stroke Treatment stroke sub classification.

*Percentage in brackets.

†Range in square brackets.

example, mRS, National Institutes of Health Stroke Scale, or cognitive function, may improve the analysis of outcome predictors further, including genetic factors. Another possible drawback of our study may be that the 2 samples of IS patients differed with regard to some important aspects: The CADISP study population was younger, on average, than the SiGN/GISCOME population, which matters because age itself is a predictor of outcome. Nevertheless, genetic imbalance level remained associated with outcome in both study samples after adjustment for age.

The ischemic stroke subtypes of the 2 cohorts in our study were also notably different which may at least partially explain why the results in the 2 cohorts were not identical. A stronger association between outcome and genetic imbalance was seen in the CADISP cohort, and a subgroup analysis in SiGN/GISCOME revealed that the association was confined to patients with cardio-embolic stroke, an IS subgroup that was not specifically analyzed in CADISP. This notwithstanding, it cannot be excluded that CNV burden is also a more general determinant of human physiology, a concept supported by previous reports of a relationship with for example, common variable immunodeficiency,[25] neuropsychiatric,[26] or other anthropometric traits.[27] Thus, general mechanisms of recovery that help to avoid complications of stroke may be impaired in subjects with higher genetic imbalance, independent of individual stroke subtype.

Imbalance identification was performed using different microarray platforms, which resulted in different cutoffs for filtering of the CNV findings. By including only CNVs of >20 SNPs in CADISP and >100 SNPs in SiGN/GISCOME, we may have missed smaller variants, implying that we might even have underestimated the influence of genetic imbalance. Finally, for the CADISP sample, only rare CNVs were available whereas, in the SiGN/GISCOME population, we performed ab initio CNV identification without frequency filtering. However, an additional analysis of the SiGN/GISCOME data with frequency filtering for rare variants hardly changed the association observed between imbalance level and outcome.

Strengths of our study include well-characterized, large study samples and a consistent association between genetic imbalance level and outcome after IS observed in 2 independent study samples. Adjustment for age, sex, stroke severity, and ancestry suggests independence of these associations from potential confounders. In addition, our key finding was corroborated by subgroup analyses of ohnolog-positive and ohnolog-negative genetic imbalance. Ohnologs are likely to be dosage-sensitive genes as they are refractory to CNV and have rarely experienced small-scale duplication.[28] Consistent with this, ohnologs have been associated with disease and to be overrepresented in pathogenic CNVs.[11] Comparing ohnolog-positive and ohnolog-negative imbalances may further identify genes driving the pathogenicity of CNVs. Our finding that imbalance enriched in ohnologs is associated with worse outcome is in line with the hypothesis that dosage-sensitive ohnologs play a role in the pathogenicity of CNVs. Importantly, the high quality of CNV validation in the current study should have served to minimize the rate of false-positive CNV detection and, thus, to improve the reliability of the results. In fact, visual CNV inspection is a valuable complement to common CNV calling algorithms.

Our study also tried to capture the complex architecture of disease outcome. Common SNPs account only for a small percentage of phenotypic variation, and by investigating genetic imbalance, we added structural variants to the repertoire of potential predictors of outcome after IS, which may help to detect some of the missing heritability and be diagnostically useful.[29] A general concept of genetic imbalance, rather than imbalance of a specific genetic locus, was preferred in our study because of the complexity of the outcome phenotype after IS.

Further studies are needed to evaluate the causes of the identified associations. The notion of genetic imbalance is fairly unspecific and indicates the loss or gain of any protein-coding gene. Future research should examine if stroke outcome is related to imbalance in specific, predefined sets of genes, for example, in genes associated with biological processes like inflammatory response or specific pathways like TGF-β (transforming growth factor β) receptor signaling. One theme could be if genetic imbalance causes a general loss of resilience and thus affects stroke recovery rates and susceptibility to other diseases. Another idea would be to focus on specific genes involved in disease-relevant pathways, such as inflammation, cell death, or neuronal recovery/plasticity. Future pathway-based CNV investigations might lead to identification of gene families relevant to outcome after IS and provide further insight into the mechanisms of stroke recovery.

Acknowledgments

We thank the patients who participated in this study and staff and participants at CADISP (Cervical Artery Dissection and Ischemic Stroke Patients) centers and SiGN/GISCOME (Genetics of Ischaemic Stroke Functional Outcome).

Sources of Funding

The CADISP study (Cervical Artery Dissection and Ischemic Stroke Patients) was supported by Inserm, Lille 2 University, Institut Pasteur de Lille and Lille University Hospital; the Contrat de Projet Etat-Region 2007 Région Nord-Pas-de-Calais; Centre National de Genotypage; Emil Aaltonen Foundation; Paavo Ilmari Ahvenainen Foundation; Helsinki University Central Hospital Research Fund; Academy of Finland; Helsinki University Medical Foundation; Päivikki and Sakari Sohlberg Foundation; Aarne Koskelo Foundation; Maire Taponen Foundation; Aarne and Aili Turunen Foundation; Lilly Foundation; Alfred Kordelin Foundation; Finnish Medical Foundation; Biomedicum Helsinki Foundation; Maud Kuistila Foundation; Orion-Farmos Research Foundation; Finnish Brain Foundation; Projet Hospitalier de Recherche Clinique Régional; Fondation de France; Génopôle de Lille; Adrinord; Basel Stroke-Funds; Käthe-Zingg-Schwichtenberg-Fonds of the Swiss Academy of Medical Sciences; the Swiss National Science Foundation (33CM30-124119; 33CM30-140340/1); the Swiss Heart Foundation; the Neurological Clinic and the University of Basel; and Fonds voor Wetenschappelijk Onderzoek (FWO) Vlaanderen (Research Foundation–Flanders). The SiGN study was funded by a cooperative agreement grant from the US National Institute of Neurological Disorders and Stroke, National Institutes of Health (U01 NS069208). Funding information for each collection in SiGN is reported in https://ars.els-cdn.com/content/image/1-s2.0-S1474442215003385-mmc1.pdf (pp 159–170). Dr Cole was partially supported by National Institutes of Health (NIH) grants R01-NS100178 and R01-NS105150,

the US Department of Veterans Affairs, the American Heart Association (AHA) Cardiovascular Genome-Phenome Study (Grant-15GPSPG23770000), and the AHA-Bayer Discovery Grant (Grant-17IBDG33700328). GISCOME: Dr Jern was funded by the Swedish Research Council (K2014-64X-14605-12-5), the Swedish state and Region Västra Götaland (ALFGBG-429981), the Swedish Heart and Lung Foundation (20130315), and the Swedish Stroke Association; Dr Lemmens was senior clinical investigator of FWO Flanders; Dr Lindgren was funded by Region Skåne, Lund University, the Swedish Heart and Lung Foundation, the Freemasons Lodge of Instruction Eos Lund, Skåne University Hospital, the Foundation of Färs&Frosta—one of Sparbanken Skåne's ownership Foundations, and the Swedish Stroke Association. Dr Freitag-Wolf was funded by the German Research Foundation (DFG) through the Cluster of Excellence No. 306 (Inflammation at Interfaces).

Disclosures

Dr Lindgren reports speech and seminar honoraria from Bayer and BMS/Pfizer and consulting for Bayer, AstraZeneca, Boehringer Ingelheim, BMS/Pfizer and Reneuron. The other authors report no conflicts.

References

1. Falcone GJ, Malik R, Dichgans M, Rosand J. Current concepts and clinical applications of stroke genetics. *Lancet Neurol*. 2014;13:405–418. doi: 10.1016/S1474-4422(14)70029-8
2. Chauhan G, Debette S. Genetic risk factors for ischemic and hemorrhagic stroke. *Curr Cardiol Rep*. 2016;18:124. doi: 10.1007/s11886-016-0804-z
3. Lindgren A, Maguire J. Stroke recovery genetics. *Stroke*. 2016;47:2427–2434. doi: 10.1161/STROKEAHA.116.010648
4. Grond-Ginsbach C, Erhart P, Chen B, Kloss M, Engelter ST, Cole JW. Copy number variation and risk of stroke. *Stroke*. 2018;49:2549–2554. doi: 10.1161/STROKEAHA.118.020371
5. Grond-Ginsbach C, Chen B, Pjontek R, Wiest T, Jiang Y, Burwinkel B, et al. Copy number variation in patients with cervical artery dissection. *Eur J Hum Genet*. 2012;20:1295–1299. doi: 10.1038/ejhg.2012.82
6. Grond-Ginsbach C, Chen B, Krawczak M, Pjontek R, Ginsbach P, Jiang Y, et al; CADISP Group. Genetic imbalance in patients with cervical artery dissection. *Curr Genomics*. 2017;18:206–213. doi: 10.2174/1389202917666160805152627
7. Korbel JO, Kim PM, Chen X, Urban AE, Weissman S, Snyder M, et al. The current excitement about copy-number variation: how it relates to gene duplications and protein families. *Curr Opin Struct Biol*. 2008;18:366–374. doi: 10.1016/j.sbi.2008.02.005
8. Valsesia A, Macé A, Jacquemont S, Beckmann JS, Kutalik Z. The growing importance of CNVs: new insights for detection and clinical interpretation. *Front Genet*. 2013;4:92. doi: 10.3389/fgene.2013.00092
9. Ledbetter DH, Martin CL. Cryptic telomere imbalance: a 15-year update. *Am J Med Genet C Semin Med Genet*. 2007;145C:327–334. doi: 10.1002/ajmg.c.30149
10. Singh PP, Arora J, Isambert H. Identification of ohnolog genes originating from whole genome duplication in early vertebrates, based on synteny comparison across multiple genomes. *PLoS Comput Biol*. 2015;11:e1004394. doi: 10.1371/journal.pcbi.1004394
11. McLysaght A, Makino T, Grayton HM, Tropeano M, Mitchell KJ, Vassos E, et al. Ohnologs are overrepresented in pathogenic copy number mutations. *Proc Natl Acad Sci USA*. 2014;111:361–366. doi: 10.1073/pnas.1309324111
12. Rice AM, McLysaght A. Dosage sensitivity is a major determinant of human copy number variant pathogenicity. *Nat Commun*. 2017;8:14366. doi: 10.1038/ncomms14366
13. Debette S, Kamatani Y, Metso TM, Kloss M, Chauhan G, Engelter ST, et al; International Stroke Genetics Consortium; CADISP Group. Common variation in PHACTR1 is associated with susceptibility to cervical artery dissection. *Nat Genet*. 2015;47:78–83. doi: 10.1038/ng.3154
14. NINDS Stroke Genetics Network (SiGN); International Stroke Genetics Consortium (ISGC). Loci associated with ischaemic stroke and its subtypes (SiGN): a genome-wide association study. *Lancet Neurol*. 2016;15:174–184. doi: 10.1016/S1474-4422(15)00338-5
15. Maguire JM, Bevan S, Stanne TM, Lorenzen E, Fernandez-Cadenas I, Hankey GJ, et al. GISCOME – Genetics of Ischaemic Stroke Functional Outcome network: a protocol for an international multicentre genetic association study. *Eur Stroke J*. 2017;2:229–237.
16. Debette S, Metso TM, Pezzini A, Engelter ST, Leys D, Lyrer P, et al; CADISP-Group. CADISP-genetics: an International project searching for genetic risk factors of cervical artery dissections. *Int J Stroke*. 2009;4:224–230. doi: 10.1111/j.1747-4949.2009.00281.x
17. Cheng YC, Stanne TM, Giese AK, Ho WK, Traylor M, Amouyel P, et al; WTCCC-2 Consortium. Genome-wide association analysis of young-onset stroke identifies a locus on chromosome 10q25 near HABP2. *Stroke*. 2016;47:307–316. doi: 10.1161/STROKEAHA.115.011328
18. Ginsbach P, Chen B, Jiang Y, Engelter ST, Grond-Ginsbach C. Copy number studies in noisy samples. *Microarrays (Basel)*. 2013;2:284–303. doi: 10.3390/microarrays2040284
19. Pinto D, Darvishi K, Shi X, Rajan D, Rigler D, Fitzgerald T, et al. Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants. *Nat Biotechnol*. 2011;29:512–520. doi: 10.1038/nbt.1852
20. Wiszniewska J, Bi W, Shaw C, Stankiewicz P, Kang SH, Pursley AN, et al. Combined array CGH plus SNP genome analyses in a single assay for optimized clinical testing. *Eur J Hum Genet*. 2014;22:79–87. doi: 10.1038/ejhg.2013.77
21. Tiszlavicz Z, Somogyvári F, Szolnoki Z, Sztriha LK, Németh B, Vécsei L, et al. Genetic polymorphisms of human β-defensins in patients with ischemic stroke. *Acta Neurol Scand*. 2012;126:109–115. doi: 10.1111/j.1600-0404.2011.01613.x
22. Nørskov MS, Frikke-Schmidt R, Loft S, Sillesen H, Grande P, Nordestgaard BG, et al. Copy number variation in glutathione S-transferases M1 and T1 and ischemic vascular disease: four studies and meta-analyses. *Circ Cardiovasc Genet*. 2011;4:418–428. doi: 10.1161/CIRCGENETICS.111.959809
23. Graffelman J, Jain D, Weir B. A genome-wide study of Hardy-Weinberg equilibrium with next generation sequence data. *Hum Genet*. 2017;136:727–741. doi: 10.1007/s00439-017-1786-7
24. Chen B, Cole JW, Grond-Ginsbach C. Departure from Hardy Weinberg equilibrium and genotyping error. *Front Genet*. 2017;8:167. doi: 10.3389/fgene.2017.00167
25. Keller M, Glessner J, Resnick E, Perez E, Chapel H, Lucas M, et al. Burden of copy number variation in common variable immunodeficiency. *Clin Exp Immunol*. 2014;177:269–271. doi: 10.1111/cei.12255
26. Guyatt AL, Stergiakouli E, Martin J, Walters J, O'Donovan M, Owen M, et al. Association of copy number variation across the genome with neuropsychiatric traits in the general population. *Am J Med Genet B Neuropsychiatr Genet*. 2018;177:489–502. doi: 10.1002/ajmg.b.32637
27. Macé A, Tuke MA, Deelen P, Kristiansson K, Mattsson H, Nõukas M, et al. CNV-association meta-analysis in 191,161 European adults reveals new loci associated with anthropometric traits. *Nat Commun*. 2017;8:744. doi: 10.1038/s41467-017-00556-x
28. Makino T, McLysaght A. Ohnologs in the human genome are dosage balanced and frequently associated with disease. *Proc Natl Acad Sci USA*. 2010;107:9270–9274. doi: 10.1073/pnas.0914697107
29. Lupski JR, Belmont JW, Boerwinkle E, Gibbs RA. Clan genomics and the complex architecture of human disease. *Cell*. 2011;147:32–43. doi: 10.1016/j.cell.2011.09.008