Neurogenetics (2008) 9:101–108
DOI 10.1007/s10048-008-0119-3

ORIGINAL ARTICLE

Structural genomic variation in ischemic stroke

Mar Matarin · Javier Simon-Sanchez · Hon-Chung Fung · Sonja Scholz · J. Raphael Gibbs · Dena G. Hernandez · Cynthia Crews · Angela Britton · Fabienne Wavrant De Vrieze · Thomas G. Brott · Robert D. Brown Jr. · Bradford B. Worrall · Scott Silliman · L. Douglas Case · John A. Hardy · Stephen S. Rich · James F. Meschia · Andrew B. Singleton

Received: 23 October 2007 / Accepted: 14 January 2008 / Published online: 21 February 2008

Abstract Technological advances in molecular genetics allow rapid and sensitive identification of genomic copy number variants (CNVs). This, in turn, has sparked interest in the function such variation may play in disease. While a role for copy number mutations as a cause of Mendelian disorders is well established, it is unclear whether CNVs may affect risk for common complex disorders. We sought to investigate whether CNVs may modulate risk for ischemic stroke (IS) and to provide a catalog of CNVs in patients with this disorder by analyzing copy number metrics produced as a part of our previous genome-wide single-nucleotide polymorphism (SNP)-based association study of ischemic stroke in a North American white population. We examined CNVs in 263 patients with ischemic stroke (IS). Each identified CNV was compared with changes identified in 275 neurologically normal controls. Our analysis identified 247 CNVs, corresponding to 187 insertions (76%; 135 heterozygous; 25 homozygous duplications or triplications; 2 heterosomic) and 60 deletions (24%; 40 heterozygous deletions; 3 homozygous

Mar Matarin, Javier Simon-Sanchez, Hon-Chung Fung, Sonja Scholz and J Raphael Gibbs contributed equally to this article.

Electronic supplementary material The online version of this article (doi:10.1007/s10048-008-0119-3) contains supplementary material, which is available to authorized users.

M. Matarin · J. Simon-Sanchez · S. Scholz · D. G. Hernandez · A. Britton · A. B. Singleton
Molecular Genetics Section, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892, USA

J. Simon-Sanchez
Unitat de Genética Molecular, Departamento de Genómica y Proteómica, Instituto de Biomedicina de Valencia-CSIC, 46010 Valencia, Spain

H.-C. Fung · C. Crews · F. Wavrant De Vrieze · A. B. Singleton (✉)
Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, 35 Convent Dr, Bethesda, MD 20892, USA
e-mail: Singleta@mail.nih.gov

H.-C. Fung
Department of Neurology, Chang Gung Memorial Hospital and College of Medicine, Chang Gung University, Taipei, Taiwan

H.-C. Fung · J. A. Hardy
Reta Lila Weston Institute of Neurological Studies, Institute of Neurology, University College London, London, UK

J. R. Gibbs
Computational Biology Core, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892, USA

L. D. Case
Section on Biostatistics, Division of Public Health Sciences, Wake Forest University Health Sciences, Medical Center Boulevard, Winston-Salem, NC 27157-1063, USA

T. G. Brott · J. F. Meschia
Department of Neurology, Mayo Clinic, Jacksonville, FL 32256, USA

deletions; 14 heterosomic deletions). Most alterations (81%) were the same as, or overlapped with, previously reported CNVs. We report here the first genome-wide analysis of CNVs in IS patients. In summary, our study did not detect any common genomic structural variation unequivocally linked to IS, although we cannot exclude that smaller CNVs or CNVs in genomic regions poorly covered by this methodology may confer risk for IS. The application of genome-wide SNP arrays now facilitates the evaluation of structural changes through the entire genome as part of a genome-wide genetic association study.

Keywords Stroke · Cerebral ischemia · Genetics · Copy number variants

Introduction

Ischemic stroke (IS) is a heterogeneous multifactorial disorder. Studies in twins, families, and animal models provide substantial evidence for a genetic contribution to this disease [1, 2]. Some conditions where stroke occurs are inherited in a classical Mendelian pattern; for example, mutations in *NOTCH3* underlie cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) [3], and mutation of *BRI* causes familial British dementia, an amyloid angiopathy, with white matter lesions without hemorrhage [4].

Identifying individual causative mutations when stroke is the prominent feature remains problematic due to the complexity of genetic and environmental risk factors in this disease. While numerous genetic risk loci have been reported for IS, the majority of these have not been replicated. The most prominent exception is the reported association between *PDE4D* variants and ischemic stroke, first identified using an extended family based approach [5]. This work has been confirmed by several groups [6–13], although the variants associated at this locus are not consistent across populations [14].

Application of genome-wide single-nucleotide polymorphism (SNP) assays in the context of case-control, rather than family-based studies, holds considerable promise in the identification of novel susceptibility loci, even in complex genetic disorders. SNP-based genome-wide association studies have proven successful in diseases with high-risk conferring alleles, such as variability in *CFH* in age-related macular degeneration and for lower risk alleles in type II diabetes (odds ratio of ~1.1 to 1.4), albeit with extremely large cohorts [15–17]. In addition to providing SNP association data, the genotyping assay also generates metrics that allow detection of copy number variants (CNVs), i.e., segments of the genome that may have been deleted or duplicated [18].

The CNV distribution in the human genome and its potential association with risk for complex human disease have been the center of much discussion [19]. Large deletions and duplications have previously been associated with rare familial diseases [20, 21], in addition to common traits such as α-thalassemia [22], and color blindness [23]. We have recently performed a pilot study of genome-wide association in a North American cohort of ischemic stroke patients and controls [24]. We have now analyzed these data specifically to assess the role of CNVs in an attempt to determine whether structural genomic variation may contribute to risk for IS.

Materials and methods

Subject collection

The Ischemic Stroke Genetics Study (ISGS) supplied all the stroke samples for the current study. ISGS is a prospective five-center North American case-control study. The protocol for ISGS has been reported previously [25]. For the stroke cohort, all cases had recent (within 30 days) first-ever IS confirmed by history, physical examination, and head imaging (CT or MRI). Stroke was defined according to the World Health Organization (WHO) definition [26]. Iatrogenic, septic embolic, vasospastic, and vasculitic stroke cases were excluded.

A single neurologist rater (RDB) classified ischemic strokes according to the Trial of Org 10172 in Acute Stroke Treatment (TOAST) [27], Oxfordshire [28], and Baltimore [29] criteria based on medical record review. Video-certified examiners assessed neurological impairment using the National Institutes of Health (NIH) Stroke Scale [30]. Functional status was assessed using the Barthel Index [31], Oxford Handicap Scale [32], and the Glasgow Outcome Score [33].

R. D. Brown Jr.
Department of Neurology, Mayo Clinic,
Rochester, MN 55905, USA

B. B. Worrall
Departments of Neurology and Public Health Sciences,
University of Virginia,
Charlottsville, VA 22908, USA

S. Silliman
Department of Neurology,
University of Florida College of Medicine,
Jacksonville, FL 32256, USA

S. S. Rich · A. B. Singleton
Center for Public Health Genomics, University of Virginia,
Charlottsville, VA 22908, USA

The control cohort used here and the identification of structural alterations in this cohort have been previously described [18, 34]. Neurologically normal subjects were briefly derived from three panels (*NDPT002, NDPT006*, and *NDPT009*) containing DNA from 275 unrelated individuals from North America and one replicate sample (133 males and 142 females). Each panel contains DNA from 92 unrelated individuals without a history of Alzheimer disease, amyotrophic lateral sclerosis, ataxia, autism, bipolar disorder, brain aneurysm, dementia, dystonia, Parkinson's disease, or stroke. None had any first-degree relative with a known primary neurological disorder, and the mean age of participants at sample collection was 68 years, ranging from 55 to 88 years (for more details, see http://ccr.coriell.org/ninds/catalog/panel/).

Sample preparation

All individuals provided written consent for the genetic analysis. Epstein–Barr virus (EBV) immortalization was performed as previously described [35, 36]. DNA for the experiments was extracted from the EBV-immortalized lymphocyte cell lines (LCL); as we have previously shown, these LCLs remain highly faithful to the genotype of source [18].

Genotyping

All samples were assayed with the Illumina Infinium Human-1 and HumanHap300 SNP chips (Illumina Inc, San Diego, CA, USA). The Human-1 product assays 109,365 gene-centric SNPs and the HumanHap300 product assays 317,511 haplotype tagging SNPs derived from phase I of the International HapMap Project (www.hapmap.org). There are 18,073 SNPs in common between the two arrays; thus, the assays combined provide data on 408,803 unique SNPs. Any assay with a call rate below 95% was repeated on a fresh DNA aliquot; if the call rate persisted below 95% the sample was excluded from further analysis.

CNV detection

Data were analyzed using BeadStudio v2.1.10.0 (Illumina Inc., San Diego, CA). Two metrics were visualized using this tool: B allele frequency and log *R* ratio.

The B allele frequency is the theta value for an individual SNP corrected for cluster position. This parameter provides an estimate of the proportion of times an individual allele at each polymorphism was called A or B. In this setting, an individual who is homozygous for the B allele (BB genotype) would have a score close to 1, an individual homozygous for the A allele (AA) would have a score close to 0, and an individual who is heterozygous (AB) would have a score of approximately 0.5. Significant deviations from these figures in contiguous SNPs are indicative of a CNV. The log R ratio is defined as the log (base 2) ratio of the observed normalized *R* value for the SNP divided by the expected normalized *R* value for the SNPs theta value. The expected *R* value is calculated from the values theta and *R*, where *R* is the intensity of dye-labeled molecules that have hybridized to the beads on the array and theta is the ratio of signal at each polymorphism for beads recognizing an A allele to beads recognizing a B allele. The expected *R* value for any individual at any typed SNP is calculated using a large population of typed individuals. Therefore, the ratio of observed *R* to expected *R* in any individual at any SNP gives an indirect measure of genomic copy number. An *R* value above 1 is indicative of an increase in copy number, and an *R* value below 1 suggests a decrease (deletion) in copy number. While this metric exhibits a high level of variance for individual SNPs, it does provide a measure of copy number when log *R* ratio values for numerous contiguous SNPs are visualized.

Based on our own experience with this technology, we have established that the smallest copy number variation that can be reliably detected is ~50 kb [18, 37]. We evaluated both the log *R* ratio and the B allele frequency plots across the genome in all samples. Each identified CNV was compared with changes identified in our neurologically normal control population [18] and those published in the Database for Genomic Variants (http://projects.tcag.ca/variation/). We calculated that our study had good power (>90%) to detect rare variants conferring a genetic risk ratio of 3.75 or greater at an alpha of 0.05.

Results

Analysis of CNVs in the control cohort was previously performed by us in a manner identical to that described here [18] and deposited publicly (http://projects.tcag.ca/variation/). Our previous work has demonstrated that the majority (10 of 10 alterations examined) of simple copy number changes ≤1 Mb in size that were observed in LCLs are also apparent in the blood sample used for immortalization. In contrast, those copy number changes larger than 1 Mb (or present as heterosomic alterations) appear to represent artifacts of the LCL creation and culture process (10 of 12 alterations examined), or correspond to V(D)J-like recombination events (two of two alterations examined) [18].

Within the stroke cohort we identified a total of 231 CNVs that were simple deletions or duplications, corresponding to 185 insertions (80%) and 46 deletions (20%), ranging in size from 1.7 kb to 2.1 Mb (ESM-Tables 1 and 2). Most of the 231 simple CNVs have been previously reported in healthy individuals or overlap with previously

reported CNVs (ESM-Tables 1 and 2). Forty-five of the 231 simple CNVs (19.5%) are unique. Of these potential new sites of structural variation, only one genomic region, on chromosome 1, contained recurrent CNVs in three individuals with IS (*IS-14*, *IS-236* and *IS-553*). The three individuals showed an apparently identical duplication spanning the genes SPRY domain-containing SOCS box protein 1 (*SPSB1)* and hexose-6-phosphate dehydrogenase (*H6PD)*.

Because of the potential disease relevance of these alterations, we examined copy-number metrics at this locus in an additional 450 neurologically normal controls samples (*NDPT019, NDPT020, NDPT022, NDPT023*, *and NDPT024* from the NINDS neurogenetics repository at the Coriell Institute) using Illumina Infinium HumanHap550 SNP chips (unpublished data). These data showed the presence of CNVs at this locus in five of these samples (~1%).

We also identified 14 deletions and two duplications that were consistent with heterosomic copy number changes (those where not all of the cellular population examined carry the CNV) ranging from 30 kb to an entire chromosome (E-Table 3). Of the 14 heterosomic deletions, 50% spanned the immunoglobulin lambda gene cluster located at chromosome 22q11.22 and likely reflect normal V(D)J-type recombination [38]. Thus, in the IS patients, 146 of the 263 samples demonstrated some form of copy number variation. Forty-nine of these samples have more than one CNV, with a maximum of five within two samples (*IS-1* and *IS-55)*. Figure 1 shows the CNVs detected sorting by chromosome.

Discussion

There is increasing discussion of the impact structural genomic alterations are likely to have in common diseases [19]. Cataloging CNVs using a methodology that assesses this variation in a genome-wide manner is critical for the identification of disease-associated genetic variability. With recent technologies such as the high-density SNP-based assays used here, an abundance of genomic copy number variations have been reported, ranging from kilobases to megabases in size. Further, this variation is readily identified in apparently healthy individuals [18, 38-40].

We report here the first genome-wide analysis of CNVs in IS patients. The CNVs identified in the current study were widely distributed throughout the genome. The majority of CNVs were rare (or orphan) changes. While single sample to group comparisons using the current methodology may under-represent common CNVs, the observation that the majority of alterations are rare is consistent with previous reports. A clear limitation of a SNP-array based approach for CNV observation is based in coverage. Many regions of the genome are poorly covered with SNPs using these technologies and thus many CNVs will be missed, in addition SNPs were previously excluded from inclusion in such arrays based on apparent Mendelian errors within families such an exclusion would clearly lead to the removal of SNPs in CNV regions. While this issue would be less of a problem for rare or orphan CNVs, such an exclusion leads to an underestimation of common CNVs.

The accuracy of the current platform for identifying CNVs was previously evaluated [18, 41]. The EBV immortalization process and clonal nature of LCL culture has been shown to lead to structural genomic variation that is not detectable in the source tissue used for immortalization. In our previous work, concordance rates between DNA derived from LCLs and DNA extracted from source tissue were 100% for CNVs ≤1 MB and 17% for CNVs >1 MB (excluding apparent CNVs resulting from V(D)J type recombination) [18]. Given this observation, we concentrated analyses on CNVs less than that 1 Mb in size. Of the 45 CNVs identified that did not overlap with previously identified CNVs (Table 1), only 1 is recurrent in more than one IS sample; a duplication across *SPSB1* and *H6PD* identified in 3 individuals. However the presence of similar

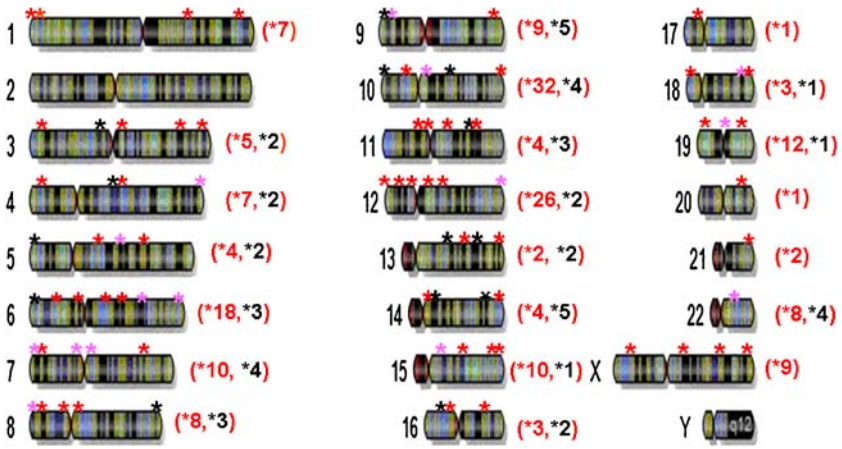

Fig. 1 Simple CNVs detected sorting by chromosome in 263 patients with ischemic stroke. The red and the black stars indicate regions with insertions and deletions respectively. Regions where we found insertions and deletions are showed with violet starts. Numbers in parenthesis are the total number of CNVs (insertions in red, deletions in black) found in each chromosome. Chromosome Y was excluded from the analysis

Table 1 Structural changes found in ischemic stroke that have not been previously reported in healthy controls

Sample	Type CNP	CHR	Location in Chr	Start bp	End bp	Size bp	startSNP	endSNP	Genes
IS-553	heterozygous duplication	1	1p36.22	9,243,800	9,309,900	66,100	rs2268169	rs4908842	H6PD, SPSB1
IS-14	heterozygous duplication	1	1p36.22	9,246,500	9,335,000	88,500	rs6688832	rs1299374	H6PD, SPSB1
IS-236	heterozygous duplication	1	1p36.22	9,246,500	9,336,000	89,500	rs6688832	rs11121384	H6PD, SPSB1
IS-55	heterozygous duplication	1	1q25.1	173,728,000	173,984,000	256,000	rs859395	rs546785	
IS-1	heterozygous duplication	1	1q42	226,872,000	226,995,000	123,000	rs765070	rs127557	FTHL2, RHOU
IS-1298	heterozygous duplication	3	3q26.1	163,361,400	163,421,900	60,500	rs2089739	rs10513581	
IS-174	homozygous duplication	3	3q27.1	184,938,000	185,228,000	290,000	rs262977	rs939335	YEATS2, MAP6D1, PARL, LOC391598, LOC647265
IS-1155	heterozygous duplication	3	3q12.2	101,837,000	101,916,300	79,300	rs591728	rs1144122	GPR128, TFG
IS-437	heterozygous duplication	4	4p15.33	14,308,000	14,516,000	208,000	rs4698280	rs6815830	
IS-1024	heterozygous duplication	4	4q23	100,903,800	100,966,200	62,400	rs716556	rs9307241	DAPP1
IS-472	*heterozygous deletion*	*4*	*4q21.21*	*81,604,000*	*82,138,000*	*534,000*	*rs9942234*	*rs2868079*	*C4orf22*
IS-168	heterozygous duplication	5	5q13.3 and 5q13	75,963,000	76,129,000	166,000	rs410262	rs458591	IQGAP2, F2R
IS-295	heterozygous duplication	5	5q23.1	120,960,000	121,059,000	99,000	rs7734808	rs7732370	
IS-141	heterozygous duplication	6	6q11.2	62,042,000	62,094,400	52,400	rs840061	rs1591548	
IS-615	heterozygous duplication	6	6q16	96,663,300	96,713,300	50,000	rs7768089	rs6571086	FUT9
IS-424	heterozygous duplication	6	6q21	124,750,188	124,907,081	156,893	rs332610	rs7770974	TCBA1
IS-323	heterozygous duplication	6	6q26	161,565,000	161,770,000	205,000	rs3778229	rs6455728	AGPAT4, PARK2
IS-551	*heterozygous deletion*	*6*	*6q21*	*113,575,000*	*114,025,000*	*450,000*	*rs2502389*	*rs7755236*	*LOC643884, LOC728590*
IS-568	heterozygous duplication	7	7p22	8,174,000	8,470,000	296,000	rs11764012	rs2040891	ICA1, NXPH1
IS-42	heterozygous duplication	7	7q31.32	122,818,668	123,545,119	726,451	rs991980	rs2215348	FLJ35834,NDUFA5, ASB15, LOC442721, WASL, HYALP1, HYAL4, SPAM1, LOC730130
IS-1031	heterozygous duplication	8	8p23.3	1,082,000	1,295,000	213,000	rs10046782	rs12541553	
IS-86	heterozygous duplication	8	8p21.2	25,511,200	25,543,900	32,700	rs7812975	rs4424264	
IS-386	heterozygous duplication	8	8p11.1	43,260,000	43,911,000	651,000	rs10958798	rs7000815	POTE8, LOC728563
IS-568	*heterozygous deletion*	*9*	*9p22*	*16,949,000*	*17,061,000*	*112,000*	*rs4961430*	*rs2129626*	
IS-90	heterozygous duplication	9	9p22	17,588,300	17,623,200	34,900	rs3808772	rs2208496	SH3GL2

Table 1 (continued)

Sample	Type CNP	CHR	Location in Chr	Start bp	End bp	Size bp	startSNP	endSNP	Genes
IS-552	*heterozygous deletion*	*9*	*9p24*	*9,465,000*	*9,563,000*	*98,000*	*rs16929560*	*rs638230*	
IS-389	homozygous duplication or triplication	10	10p12	25,999,000	26,066,800	67,800	rs6482510	rs2263000	
IS-158	heterozygous duplication	11	11p12	39,007,000	39,120,000	113,000	rs1599459	rs7120739	
ND3843	heterozygous duplication	11	11q22.3	107,262,000	107,444,000	182,000	rs2847190	rs7104942	CUL5,RAB39, LOC643949
IS-145	heterozygous duplication	13	13q21.32	67,676,000	67,798,000	122,000	rs6562520	rs9564447	
IS-337	*heterozygous deletion*	*13*	*13q21.1*	*54,036,000*	*54,422,000*	*386,000*	*rs9563238*	*rs9536919*	
IS-1134	*heterozygous deletion*	*13*	*13q31.1*	*85,599,004*	*85,842,380*	*243,376*	*rs1413455*	*rs9585170*	
IS-13	heterozygous duplication	15	15q21.3	53,302,000	53,546,000	244,000	rs17238192	rs600753	RAB27A, PIGB, CCPG1, MIRN628, DYX1C1, LOC729120
IS-55	homozygous duplication or triplication	15	15q24–q25	83,799,700	83,875,900	76,200	rs12443473	rs2344436	AKAP13
IS-55	heterozygous duplication	15	15q25.3	88,651,000	88,800,000	149,000	rs6496649	rs3213649	GABARAPL3, MGC75360, IQGAP1
IS-276	heterozygous duplication	18	18p11.2	7,803,000	8,013,000	210,000	rs364418	rs2156236	PTPRM
IS-13	heterozygous duplication	18	18q23	72,553,700	72,598,400	44,700	rs4635418	rs4890931	
IS-276	heterozygous duplication	19	19q13.42	61,175,000	61,284,000	109,000	rs306487	rs6509981	NALP8, NALP5, LOC729982
IS-266	heterozygous duplication	19	19q13.43	62,695,000	62,888,000	193,000	rs2074071	rs752150	ZNF419, MGC4728, ZNF549, ZNF550, ZNF416, ZIK1, ZNF530, ZNF134, ZNF211, ZSCAN4, ZNF551
IS-1132	heterozygous duplication	20	20q13.2	51,262,600	51,307,100	44,500	rs1296032	rs6091653	TSHZ2
IS-462	heterozygous duplication	21	21q22.11	34,417,000	36,526,000	2,109,000	rs4816519	rs2070515	LOC728778, LOC728556, RPS9P1, CBR3, DOPEY2, RPL3P1
IS-24	heterozygous duplication	X	Xq13	75,202,600	75,274,600	72,000	rs2213481	rs4892427	

CNVs over this locus in 5 of an additional 460 controls suggests that this variant in not a risk factor for IS. The remaining CNVs may be of importance in the pathobiology of IS; however, given the low frequency of each individual alteration, screening of these variants in a very large cohort (1000's of cases and controls) would be required to make any unequivocal conclusions.

In summary, our study did not detect any common genomic structural variation unequivocally linked to IS. We cannot exclude the possibility that smaller CNVs or CNVs in genomic regions poorly covered by this methodology may confer risk for IS. The recent availability of higher density CNV-directed arrays from both Affymetrix and Illumina will increase the number of CNVs that can be

detected and thus may go some way toward addressing this question; however, because some studies have reported linkage disequilibrium (LD) between CNVs and proximal SNPs [42, 43], we would predict that our LD-based whole-genome association study [24] would have detected common CNVs linked to disease, whether detected directly or not in this study, by showing association at SNPs tagging such changes. The application of genome-wide SNP arrays now facilitates the evaluation of structural changes through the entire genome as part of a genome-wide genetic association study.

References

1. Hassan A, Markus HS (2000) Genetics and ischaemic stroke. Brain 123(Pt 9):1784–1812
2. Tournier-Lasserve E (2002) New players in the genetics of stroke. N Engl J Med 347:1711–1712
3. Joutel A, Corpechot C, Ducros A, Vahedi K, Chabriat H, Mouton P, Alamowitch S, Domenga V, Cecillion M, Marechal E, Maciazek J, Vayssiere C, Cruaud C, Cabanis EA, Ruchoux MM, Weissenbach J, Bach JF, Bousser MG, Tournier-Lasserve E (1996) Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia. Nature 383:707–710
4. Vidal R, Frangione B, Rostagno A, Mead S, Revesz T, Plant G, Ghiso J (1999) A stop-codon mutation in the BRI gene associated with familial British dementia. Nature 399:776–781
5. Gretarsdottir S, Thorleifsson G, Reynisdottir ST, Manolescu A, Jonsdottir S, Jonsdottir T, Gudmundsdottir T, Bjarnadottir SM, Einarsson OB, Gudjonsdottir HM, Hawkins M, Gudmundsson G, Gudmundsdottir H, Andrason H, Gudmundsdottir AS, Sigurdardottir M, Chou TT, Nahmias J, Goss S, Sveinbjornsdottir S, Valdimarsson EM, Jakobsson F, Agnarsson U, Gudnason V, Thorgeirsson G, Fingerle J, Gurney M, Gudbjartsson D, Frigge ML, Kong A, Stefansson K, Gulcher JR (2003) The gene encoding phosphodiesterase 4D confers risk of ischemic stroke. Nat Genet 35:131–138
6. Meschia JF, Brott TG, Brown RD Jr., Crook R, Worrall BB, Kissela B, Brown WM, Rich SS, Case LD, Evans EW, Hague S, Singleton A, Hardy J (2005) Phosphodiesterase 4D and 5-lipoxygenase activating protein in ischemic stroke. Ann Neurol 58:351–361
7. Nilsson-Ardnor S, Wiklund PG, Lindgren P, Nilsson AK, Janunger T, Escher SA, Hallbeck B, Stegmayr B, Asplund K, Holmberg D (2005) Linkage of ischemic stroke to the PDE4D region on 5q in a Swedish population. Stroke 36:1666–1671
8. van Rijn MJ, Slooter AJ, Schut AF, Isaacs A, Aulchenko YS, Snijders PJ, Kappelle LJ, van Swieten JC, Oostra BA, van Duijn CM (2005) Familial aggregation, the PDE4D gene, and ischemic stroke in a genetically isolated population. Neurology 65:1203–1209
9. Saleheen D, Bukhari S, Haider SR, Nazir A, Khanum S, Shafqat S, Anis MK, Frossard P (2005) Association of phosphodiesterase 4D gene with ischemic stroke in a Pakistani population. Stroke 36:2275–2277
10. Nakayama T, Asai S, Sato N, Soma M (2006) Genotype and haplotype association study of the STRK1 region on 5q12 among Japanese: a case-control study. Stroke 37:69–76
11. Woo D, Kaushal R, Kissela B, Sekar P, Wolujewicz M, Pal P, Alwell K, Haverbusch M, Ewing I, Miller R, Kleindorfer D, Flaherty M, Chakraborty R, Deka R, Broderick J (2006) Association of Phosphodiesterase 4D with ischemic stroke: a population-based case-control study. Stroke 37:371–376
12. Brophy VH, Ro SK, Rhees BK, Lui LY, Lee JM, Umblas N, Bentley LG, Li J, Cheng S, Browner WS, Erlich HA (2006) Association of phosphodiesterase 4D polymorphisms with ischemic stroke in a US population stratified by hypertension status. Stroke 37:1385–1390
13. Zee RY, Brophy VH, Cheng S, Hegener HH, Erlich HA, Ridker PM (2006) Polymorphisms of the phosphodiesterase 4D, cAMP-specific (PDE4D) gene and risk of ischemic stroke: a prospective, nested case-control evaluation. Stroke 37:2012–2017
14. Rosand J, Bayley N, Rost N, de Bakker PI (2006) Many hypotheses but no replication for the association between PDE4D and stroke. Nat Genet 38:1091–1092 author reply 1092–1093
15. Zeggini E, Weedon MN, Lindgren CM, Frayling TM, Elliott KS, Lango H, Timpson NJ, Perry JR, Rayner NW, Freathy RM, Barrett JC, Shields B, Morris AP, Ellard S, Groves CJ, Harries LW, Marchini JL, Owen KR, Knight B, Cardon LR, Walker M, Hitman GA, Morris AD, Doney AS, Burton PR, Clayton DG, Craddock N, Deloukas P, Duncanson A, Kwiatkowski DP, Ouwehand WH, Samani NJ, Todd JA, Donnelly P, Davison D, Easton D, Evans D, Leung HT, Spencer CC, Tobin MD, Attwood AP, Boorman JP, Cant B, Everson U, Hussey JM, Jolley JD, Knight AS, Koch K, Meech E, Nutland S, Prowse CV, Stevens HE, Taylor NC, Walters GR, Walker NM, Watkins NA, Winzer T, Jones RW, McArdle WL, Ring SM, Strachan DP, Pembrey M, Breen G, St Clair D, Caesar S, Gordon-Smith K, Jones L, Fraser C, Green EK, Grozeva D, Hamshere ML, Holmans PA, Jones IR, Kirov G, Moskvina V, Nikolov I, O'Donovan MC, Owen MJ, Collier DA, Elkin A, Farmer A, Williamson R, McGuffin P, Young AH, Ferrier IN, Ball SG, Balmforth AJ, Barrett JH, Bishop DT, Iles MM, Maqbool A, Yuldasheva N, Hall AS, Braund PS, Dixon RJ, Mangino M, Stevens S, Thompson JR, Bredin F, Tremelling M et al (2007) Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. Science 316:1336–1341
16. Saxena R, Voight BF, Lyssenko V, Burtt NP, de Bakker PI, Chen H, Roix JJ, Kathiresan S, Hirschhorn JN, Daly MJ, Hughes TE, Groop L, Altshuler D, Almgren P, Florez JC, Meyer J, Ardlie K, Bengtsson Bostrom K, Isomaa B, Lettre G, Lindblad U, Lyon HN, Melander O, Newton-Cheh C, Nilsson P, Orho-Melander M, Rastam L, Speliotes EK, Taskinen MR, Tuomi T, Guiducci C, Berglund A, Carlson J, Gianniny L, Hackett R, Hall L, Holmkvist J, Laurila E, Sjogren M, Sterner M, Surti A, Svensson M, Tewhey R, Blumenstiel B, Parkin M, Defelice M, Barry R, Brodeur W, Camarata J, Chia N, Fava M, Gibbons J, Handsaker B, Healy C, Nguyen K, Gates C, Sougnez C, Gage D, Nizzari M, Gabriel SB, Chirn GW, Ma Q, Parikh H, Richardson D, Ricke D, Purcell S (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 316:1331–1336
17. Scott LJ, Mohlke KL, Bonnycastle LL, Willer CJ, Li Y, Duren WL, Erdos MR, Stringham HM, Chines PS, Jackson AU, Prokunina-Olsson L, Ding CJ, Swift AJ, Narisu N, Hu T, Pruim R, Xiao R, Li XY, Conneely KN, Riebow NL, Sprau AG, Tong M, White PP, Hetrick KN, Barnhart MW, Bark CW, Goldstein JL, Watkins L, Xiang F, Saramies J, Buchanan TA, Watanabe RM, Valle TT, Kinnunen L, Abecasis GR, Pugh EW, Doheny KF, Bergman RN, Tuomilehto J, Collins FS, Boehnke M (2007) A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. Science 316:1341–1345
18. Simon-Sanchez J, Scholz S, Fung HC, Matarin M, Hernandez D, Gibbs JR, Britton A, de Vrieze FW, Peckham E, Gwinn-Hardy K, Crawley A, Keen JC, Nash J, Borgaonkar D, Hardy J, Singleton A (2007) Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals. Hum Mol Genet 16:1–14

19. Lupski JR (2007) Structural variation in the human genome. N Engl J Med 356:1169–1171
20. Inoue K, Osaka H, Sugiyama N, Kawanishi C, Onishi H, Nezu A, Kimura K, Yamada Y, Kosaka K (1996) A duplicated PLP gene causing Pelizaeus–Merzbacher disease detected by comparative multiplex PCR. Am J Hum Genet 59:32–39
21. Singleton AB, Farrer M, Johnson J, Singleton A, Hague S, Kachergus J, Hulihan M, Peuralinna T, Dutra A, Nussbaum R, Lincoln S, Crawley A, Hanson M, Maraganore D, Adler C, Cookson MR, Muenter M, Baptista M, Miller D, Blancato J, Hardy J, Gwinn-Hardy K (2003) alpha-Synuclein locus triplication causes Parkinson's disease. Science 302:841
22. Lauer J, Shen CK, Maniatis T (1980) The chromosomal arrangement of human alpha-like globin genes: sequence homology and alpha-globin gene deletions. Cell 20:119–130
23. Vollrath D, Nathans J, Davis RW (1988) Tandem array of human visual pigment genes at Xq28. Science 240:1669–1672
24. Matarin M, Brown WM, Scholz S, Simon-Sanchez J, Fung HC, Hernandez D, Gibbs JR, De Vrieze FW, Crews C, Britton A, Langefeld CD, Brott TG, Brown RD Jr., Worrall BB, Frankel M, Silliman S, Case LD, Singleton A, Hardy JA, Rich SS, Meschia JF (2007) A genome-wide genotyping study in patients with ischaemic stroke: initial analysis and data release. Lancet Neurol 6:414–420
25. Meschia JF, Brott TG, Brown RD Jr., Crook RJ, Frankel M, Hardy J, Merino JG, Rich SS, Silliman S, Worrall BB (2003) The Ischemic Stroke Genetics Study (ISGS) Protocol. BMC Neurol 3:4
26. The World Health Organization MONICA (1988) Project (monitoring trends and determinants in cardiovascular disease): a major international collaboration. WHO MONICA Project Principal Investigators. J Clin Epidemiol 41:105–114
27. Adams HP Jr., Bendixen BH, Kappelle LJ, Biller J, Love BB, Gordon DL, Marsh EE 3rd (1993) Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment. Stroke 24:35–41
28. Bamford J, Sandercock P, Dennis M, Burn J, Warlow C (1991) Classification and natural history of clinically identifiable subtypes of cerebral infarction. Lancet 337:1521–1526
29. Johnson CJ, Kittner SJ, McCarter RJ, Sloan MA, Stern BJ, Buchholz D, Price TR (1995) Interrater reliability of an etiologic classification of ischemic stroke. Stroke 26:46–51
30. Lyden P, Brott T, Tilley B, Welch KM, Mascha EJ, Levine S, Haley EC, Grotta J, Marler J (1994) Improved reliability of the NIH Stroke Scale using video training. NINDS TPA Stroke Study Group. Stroke 25:2220–2226
31. Collin C, Wade DT, Davies S, Horne V (1988) The Barthel ADL Index: a reliability study. Int Disabil Stud 10:61–63
32. Bamford JM, Sandercock PA, Warlow CP, Slattery J (1989) Interobserver agreement for the assessment of handicap in stroke patients. Stroke 20:828
33. Jennett B, Bond M (1975) Assessment of outcome after severe brain damage. Lancet 1:480–484
34. Fung HC, Scholz S, Matarin M, Simon-Sanchez J, Hernandez D, Britton A, Gibbs JR, Langefeld C, Stiegert ML, Schymick J, Okun MS, Mandel RJ, Fernandez HH, Foote KD, Rodriguez RL, Peckham E, De Vrieze FW, Gwinn-Hardy K, Hardy JA, Singleton A (2006) Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data. Lancet Neurol 5:911–916
35. Miller G, Shope T, Lisco H, Stitt D, Lipman M (1972) Epstein–Barr virus: transformation, cytopathic changes, and viral antigens in squirrel monkey and marmoset leukocytes. Proc Natl Acad Sci U S A 69:383–387
36. Tumilowicz JJ, Gallick GE, East JL, Pathak S, Trentin JJ, Arlinghaus RB (1984) Presence of retrovirus in the B95-8 Epstein–Barr virus-producing cell line from different sources. In Vitro 20:486–492
37. Gibbs JR, Singleton A (2006) Application of genome-wide single nucleotide polymorphism typing: simple association and beyond. PLoS Genet 2:e150
38. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M (2004) Large-scale copy number polymorphism in the human genome. Science 305:525–528
39. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, Gonzalez JR, Gratacos M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME (2006) Global variation in copy number in the human genome. Nature 444:444–454
40. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C (2004) Detection of large-scale variation in the human genome. Nat Genet 36:949–951
41. Huang J, Wei W, Zhang J, Liu G, Bignell GR, Stratton MR, Futreal PA, Wooster R, Jones KW, Shapero MH (2004) Whole genome DNA copy number changes identified by high density oligonucleotide arrays. Hum Genomics 1:287–299
42. McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC, Dallaire S, Gabriel SB, Lee C, Daly MJ, Altshuler DM (2006) Common deletion polymorphisms in the human genome. Nat Genet 38:86–92
43. Locke DP, Sharp AJ, McCarroll SA, McGrath SD, Newman TL, Cheng Z, Schwartz S, Albertson DG, Pinkel D, Altshuler DM, Eichler EE (2006) Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. Am J Hum Genet 79:275–290