

UC San Diego

UC San Diego Previously Published Works

Title

NeuroX, a fast and efficient genotyping platform for investigation of neurodegenerative diseases

Permalink

<https://escholarship.org/uc/item/22z464wc>

Journal

Neurobiology of Aging, 36(3)

ISSN

0197-4580

Authors

Nalls, Mike A
Bras, Jose
Hernandez, Dena G
[et al.](#)

Publication Date

2015-03-01

DOI

10.1016/j.neurobiolaging.2014.07.028

Peer reviewed

Published in final edited form as:

Neurobiol Aging. 2015 March ; 36(3): 1605.e7–1605.12. doi:10.1016/j.neurobiolaging.2014.07.028.

NeuroX, a fast and efficient genotyping platform for investigation of neurodegenerative diseases

Mike A. Nalls, Jose Bras, Dena G. Hernandez, Margaux F. Keller, Elisa Majounie, Alan E. Renton, Mohamad Saad, Iris Jansen, Rita Guerreiro, Steven Lubbe, Vincent Plagnol, J. Raphael Gibbs, Claudia Schulte, Nathan Pankratz, Margaret Sutherland, Lars Bertram, Christina M. Lill, Anita L. DeStefano, Tatiana Faroud, Nicholas Eriksson, Joyce Y. Tung, Connor Edsall, Noah Nichols, Janet Brooks, Sampath Arepalli, Hannah Pliner, Chris Letson, Peter Heutink, Maria Martinez, Thomas Gasser, Bryan J. Traynor, Nick Wood, John Hardy, and Andrew B. Singleton* on behalf of the International Parkinson's Disease Genomics Consortium (IPDGC) and the Parkinson's Disease meta-analysis consortium

Abstract

Our objective was to design a genotyping platform that would allow rapid genetic characterization of samples in the context of genetic mutations and risk factors associated with common neurodegenerative diseases. The platform needed to be relatively affordable, rapid to deploy, and use a common and accessible technology. Central to this project, we wanted to make the content of the platform open to any investigator without restriction. In designing this array we prioritized a number of types of genetic variability for inclusion, such as known risk alleles, disease-causing mutations, putative risk alleles, and other functionally important variants. The array was primarily designed to allow rapid screening of samples for disease-causing mutations and large population studies of risk factors. Notably, an explicit aim was to make this array widely available to facilitate data sharing across and within diseases. The resulting array, NeuroX, is a remarkably cost and time effective solution for high-quality genotyping. NeuroX comprises a backbone of standard Illumina exome content of approximately 240,000 variants, and over 24,000 custom content variants focusing on neurologic diseases. Data are generated at approximately \$50–\$60 per sample using a 12-sample format chip and regular Infinium infrastructure; thus, genotyping is rapid and accessible to many investigators. Here, we describe the design of NeuroX, discuss the utility of NeuroX in the analyses of rare and common risk variants, and present quality control metrics and a brief primer for the analysis of NeuroX derived data.

© 2014 Published by Elsevier Inc.

*Corresponding author at: Laboratory of Neurogenetics, National Institute on Aging, 35 Lincoln Drive, Bethesda, MD, USA. Tel.: +301 451 6079; fax: +301 451 5466. singleta@mail.nih.gov (A.B. Singleton) .

Disclosure statement

The authors declare they have no conflicts of interest, financial, or otherwise, related to the present work.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.neurobiolaging.2014.07.028>.

For full consortia membership, affiliations, and funding, please see the consortia section in the Supplementary text.

Keywords

Genotyping; Methods; Genetics; Neurodegeneration; Parkinson's; Meta-analysis; Imputation

1. Introduction

The availability of economical custom content additions to genome-wide or exome-wide genotyping arrays has permitted the development of tailored arrays for both genetic discovery and replication efforts. In the last few years, it has become evident that in the second wave of genome wide association (GWA) investigators sought to investigate variants below the threshold of genome-wide significance and fine map extant signals. Such an effort requires large-scale replication efforts involving the assay of large numbers of samples, and the interrogation of a very large number of candidate variants. A fairly inefficient approach to this problem was being used, where investigators representing single disease research groups pursued replication in isolation of other efforts both within and across diseases. In 2011, the National Institute of Neurological Disorders and Stroke convened a meeting that included investigators researching myriad common neurodegenerative diseases with the intent of identifying a more efficient solution. This meeting involved representatives from genetics groups leading GWA in Alzheimer's disease (AD), Parkinson's disease (PD), amyotrophic lateral sclerosis (ALS), multiple sclerosis, and frontotemporal dementia (FTD), among others. There was a broad consensus that the design of an accessible array that could type variants of interest for all major neurodegenerative diseases would be of great utility; such an array had the potential to benefit from an economy of scale, to reduce cost by allowing easy sharing of controls and allow direct comparison of genetic data across diseases. In response to that consensus, we modified the design of an array originally intended to serve as a replication assay for a large PD meta-analysis to include a wide variety of content relevant to the broader neurodegenerative disease research community. Here, we describe the content and use of this array, called NeuroX.

In this effort, summary statistics for the largest available genome-wide association studies (GWAS) were mined to nominate known and candidate loci tagging risk for AD, FTD, multiple system atrophy (MSA), myasthenia gravis, Charcot Marie Tooth, progressive supranuclear palsy (PSP), ALS, and PD. Where available, putative risk variants identified by exome sequencing of familial and population based samples, as well as those derived from literature review for the previously mentioned diseases, were also included on the array. We also performed a systematic literature and database search for all mutations known to cause neurologic disease. Technical redundancies and reliable proxies were used for priority SNPs to guarantee quality genotyping calls produced by the array. This custom Neuro content includes over 24,000 neurodegenerative-focused variants; this custom library can be added to many off-the-shelf Illumina Infinium products; however, here we describe the use of this library when added to Illumina's Infinium HumanExome BeadChip, a product we have named NeuroX. Thus NeuroX includes full exome sequencing based variability standard to the Illumina HumanExome array v1.1 (242,901 variants) and neurologic and neurodegenerative disease focused content (24,706 variants). In addition to the ability to add the custom Neuro library to other illumina genotyping arrays, it is also relatively easy to add

new custom variants should the need arise. In this article however, we describe the initial version of the NeuroX array comprising the base exome and existing custom content.

From its inception, the NeuroX array was designed to be a rapid and cost effective solution for high quality genotype data. The current cost of the array is approximately \$57 per sample and genotyping of thousands of samples per week is achievable in most core laboratories, with this estimate including reagents but excluding labor and previous Illumina infrastructure costs. It is also notable that we are in the process of making a large amount of NeuroX data publicly available (dbgap address pending).

2. Methods

2.1. Array design

The custom content available on the NeuroX array was taken from 3 primary sources: large-scale GWAS, high throughput sequencing of families and cohort studies, and literature searches to identify risk factors and disease-causing mutations.

For GWAS based data sets we mined participant level data, when available, for diseases such as PD, ALS, FTD, and myasthenia gravis, including both published and unpublished data sets (ALSGEN Consortium et al., 2013; Chiò et al., 2009; Do et al., 2011; International Parkinson Disease Genomics Consortium et al., 2011; International Parkinson's Disease Genomics Consortium (IPDGC) and Wellcome Trust Case Control Consortium 2 (WTCCC2), 2011; Lill et al., 2012; Mok et al., 2012). Participant level GWAS data for AD and PSP were not available to our group at the time of chip design, so publicly available GWA loci for these diseases were included (1000 Genomes Project Consortium et al., 2012; Höglinger et al., 2011; Hollingworth et al., 2011; Lambert et al., 2009). Genome-wide significant loci from diseases of interest were included with either multiple proxies for the top SNP at every locus, or technical replicates, if proxies were not available. We have included up to 5 variants per significant locus. Loci were defined as any SNP reaching a genome-wide significant p -value and correlated at $r^2 < 0.50$ with any other significant SNPs within 250 kilobases for each disease of interest. All analyses were derived from at least 1000 Genomes level SNP coverage and used participant level data from the 1000 Genomes project to nominate proxies when possible. In addition, locus tagging SNPs were included to allow for the identification of new loci in larger sample series. For all SNPs associated in GWAS with diseases of interest that reached candidate p -values of 1×10^{-4} or stronger, additional haplotype-tagging SNPs were placed on the NeuroX array, in an attempt to facilitate future genotype imputation efforts. Tagging SNPs were selected based on an r^2 in 1000 Genomes samples at less than 0.50 with any other SNPs meeting the same p -value threshold within a 250 kilobase window, allowing for regional assessments of genetic variability. Whenever possible, GWAS based SNPs that were not the most significant within the locus were replaced by a proxy meeting the previously mentioned criteria if array design scores for the probe associated with that SNP failed (quality less than 0.80 and no array validation), as a means of only using higher quality SNPs on the NeuroX array. This led to the successful inclusion of almost 16,000 GWAS-derived variants or GWAS-related variants across multiple disease sources.

Sequence-based data generated by pilot studies within our consortia (both exome and genome sequencing) were mined to nominate rare and coding variants for inclusion on the NeuroX array. These data come from familial and cohort studies looking into AD, MSA, FTD, Charcot Marie Tooth, MSA, PSP, ALS, and PD. Cohort-derived sequence-based data were inclusive of any rare and coding variants at a frequency of less than 5% in the population from which the pilot data were collected. For data extracted from family-based sequencing studies, variants were filtered and only those not appearing in the 1000 Genomes Project and the NHLBI Exome Sequencing Project database were included (1000 Genomes Project Consortium et al., 2012; NHLBI GO Exome Sequencing Project, 2011). This led to the successful inclusion of 7485 rare sequence-based variants.

An extensive systematic review of published literature was performed to include variants known to be involved in neurologic or neurodegenerative diseases for nomination onto the array. Briefly, we performed PubMed searches using the gene name and the word “mutation” as search parameters to identify articles describing mutations. The genes searched for were: *ABCA7*, *ACE*, *APOE*, *APP*, *ATP13A2*, *BACE1*, *CHMP2B*, *CLCN6*, *CLN3*, *CLN5*, *CLN6*, *CLN8*, *CSF1R*, *CST3*, *CTSD*, *DNAJC5*, *ECE2*, *FBXO7*, *FUS*, *GBA*, *GLA*, *GLB1*, *GRN*, *GUSB*, *HEXA*, *HEXB*, *LRRK2*, *MAPT*, *MFSN8*, *NEU1*, *NOTCH3*, *NPC1*, *NPC2*, *PANK2*, *PARK2*, *PARK7*, *PINK1*, *PLA2G6*, *PPT1*, *PSAP*, *PSEN1*, *PSEN2*, *SGSH*, *SNCA*, *SORL1*, *SPTLC1*, *TARDBP*, *TPP1*, *TREM2*, *TYROBP*, *VCP*, and *VPS35*. We complemented this search by including all the variants in the Parkinson Disease Mutation Database and the Alzheimer Disease & Frontotemporal Dementia Mutation Database (Cruts et al., 2012). In addition, updated GWAS loci for any traits meeting $p < 1 \times 10^{-8}$ in NHGRI GWAS catalog that were not already on the basic exome content were added to the array if the probe design score for that SNP was >0.8 (Hindorf et al., 2009a, 2009b). Also, as part of this phase of array design, special attention was paid to the APOE region with 34 variants being dedicated to genotyping of the canonical epsilon-4 compound genotype. This led to the successful inclusion of 1322 variants. For ALS, we also mined variants from a number of databases with several aims. To identify new mutation carriers, we collected from HGMD and ALSOD all mutations in common (C9orf72 excluding repeat expansions, FUS, MATR3, OPTN, SOD1, SPG11, TARDBP, UBQLN2, and VCP) and rare ALS genes (*ALS2*, *ANG*, *CHMP2B*, *DCTN1*, *FIG4*, *SETX*, *TAF15*, and *VAPB*). To identify association signals in and around known ALS genes, we mined 1000 Genome data to identify all multiethnic variants with minor allele frequency (MAF) >0.01 located in common ALS gene bodies ± 100 kb. We then used Plink to identify haplotype-tagging SNPs ($r^2 > 0.50$). For the ALS/FTD linked gene C9orf72, we mined variants located within the 242 kb Finnish 42-SNP haplotype and ± 20 kb (Laaksovirta et al., 2010, p. 21). To fine map exonic variation in known ALS genes, we mined 1000 Genome data to identify all multiethnic exonic variants with MAF >0 in common ALS genes.

2.2. Array genotyping

For the pilot analysis used to generate the data presented here, approximately 14,000 samples were genotyped and multiple calling methods tested. Samples tested were derived from a number of sources including DNA from whole blood, EBV transformed lymphocytes, and brain tissue. Genotyping was executed as per the manufacturer's protocol

(Illumina, Inc). Our genotype calling workflow used a publicly available cluster file for the exome array standard content, which we modified to maximize variant calling for the NeuroX custom content. This was accomplished using a combination of the Illumina GenomeStudio automated clustering algorithm, with manual inspection and modification for the subset of the clusters not included in Grove et al. (2013). As part of the array design process, we excluded a number of variants based on low design quality scores, which allowed us to retain diverse content and maximize the number of successfully typed variants.

In addition, we imputed a random subset of 1000 European ancestry unrelated individuals from a larger Parkinson's disease GWAS study that passed quality control after being genotyped on the NeuroX array (Nalls et al., 2014) using the default settings of MiniMac (Howie et al., 2012). Nonpalindromic SNPs passing quality control and overlapping with those included in the reference haplotypes (1000 Genomes Phase 1 Alpha Freeze version 3, multiethnic panel) were used for imputation (1000 Genomes Project Consortium et al., 2012). This allowed us to densely impute higher variant coverage into regions of interest related to neurologic diseases GWAS based on the currently available content on the array. In addition, we show that the NeuroX array can be used for basic quality control similar to standard GWAS, such as gender checking (evaluating concordance between self-reported and genetically determined genders as part of quality control) or estimating continental ancestry based on applying principal components analyses to common tagging SNPs (Supplementary Figs. 1 and 2).

3. Results and conclusions

Both NeuroX custom content and the standard HumanExome based content show that most of the SNPs across the MAF spectrum have GenTrain scores >0.7 , suggesting high quality genotype clusters are readily available (Fig. 1). As expected, lower MAFs are associated with slightly lower genotype cluster qualities ($p < 0.001$ from linear regression models across MAF strata in Fig. 1 comparing trends in GenTrain as MAF changes). GenTrain scores tend to be only marginally lower for the custom content, which is not entirely surprising, given that genotypes for these variants were clustered and called on a reference of approximately 14,000 samples as opposed to approximately 60,000 samples used to generate the reference cluster file used to call genotypes for the standard exome content.

Custom content on the NeuroX array spans 2236 megabases of the autosome, only slightly less than the approximately 2600 covered by early GWAS arrays on which many previous studies of neurodegenerative disease were based (Nalls et al., 2009). Mean per megabase coverage of the custom content is 10.754 variants per megabase, with a maximum of over 600 variants of interest for fine mapping of particular regions, with a comparative bias toward nonexonic and GWAS-derived variants (Table 1, Fig. 2). The maximum coverage for the NeuroX custom content occurring in regions of interest up to over 5-fold the depth of the standard content in the same region. In comparison, the standard content covers 2703 autosomal megabases at an average of 87.842 variants per megabase, with maximum coverage of certain exonic regions surpassing 1000 variants per megabase. The inclusion of tag SNPs within the GWAS-derived custom content in conjunction with standard content

variants have facilitated the successful imputation of over 1.2 million SNPs (imputation quality >0.30). Imputed variants from the NeuroX array cover 2703 megabases and average 478.400 variants per megabase with a maximum coverage up to over 16,500 variants per megabase in some regions of interest.

As a proof of concept, we accurately tag known rare variants in neurodegenerative disease. For example, the p.G2019S mutation in *LRRK2* (rs34637584) was confirmed to be completely concordant for over 1000 samples genotyped using the NeuroX array that were also assayed via taqman genotyping (Paisán-Ruíz et al., 2004). *APOE* genotypes were extracted for a subset of over 2500 NeuroX assayed samples overlapping with a previous study based on targeted genotyping (Federoff et al., 2012) with only 93% accuracy to tag the APOE epsilon-4 haplotype associated with Alzheimer's risk. This haplotype is made of 2 SNPs rs7412 and rs429358. The discordance of APOE haplotypes between NeuroX and TaqMan genotyping was entirely driven by discordance at rs429358, with complete concordance at rs7412. Notably, we have identified rs429358 as a low quality variant on NeuroX. Rs7412 is of acceptable quality, with greater than 99% genotype concordance across 5 technical replicates, a success rate mirrored at most of the redundant sites across the array.

The data presented in this article unequivocally show that the NeuroX array is a powerful and reliable tool for the investigation of genetic factors associated with neurodegenerative disorders. Although not designed with clinical diagnosis in mind, we believe it will serve as a powerful analytic tool for research purposes and investigation of disease mechanisms. We have shown not only that the content of the array is useful in assaying both rare risk variants and common variability for use in future studies but also highly valuable in investigating known risk loci in more detail. We fully expect this array to become a starting point to the genetic analysis of neurodegenerative disorders, given its relevant and up-to-date genotyping content as well as its low cost. This custom array is being treated as an on-going venture and is currently being adapted to newer genotyping platforms outside of the standard exome array content described here and tuned for better accuracy and higher quality content, whereas still maintaining compatibility with current offerings. Additionally, the fact that virtually all samples derived from subjects with these disorders may be screened on the same platform will provide researchers with tremendous power to perform not only analysis of a single phenotype but also to compare different disease entities for overlaps or significant differences.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported in part by the Intramural Research Program of the National Institute on Aging, National Institutes of Health, Department of Health and Human Services (project numbers Z01-AG000949-02, under human subjects protocol 2003-077), and by the Wellcome Trust and/or MRC Joint Call in Neurodegeneration award (WT089698) to the UK Parkinson's Disease Consortium (UKPDC) whose members are from the UCL/Institute of Neurology, the University of Sheffield, and the MRC Protein Phosphorylation Unit at the University of Dundee. Additional funding information is provided in the consortium information section included as a Supplementary

online appendix. Special thanks to Megan L. Grove-Gaona, Jerome Rotter, Eric Boerwinkle, and Christopher O'Donnell on behalf of the CHARGE consortium for their helpful advice on aspects of this project. The authors thank the NHLBI GO Exome Sequencing Project and its ongoing studies which produced and provided exome variant calls for comparison: the Lung GO Sequencing Project (HL-102923), the WHI Sequencing Project (HL-102924), the Broad GO Sequencing Project (HL-102925), the Seattle GO Sequencing Project (HL-102926), and the Heart GO Sequencing Project (HL-103010). They thank and acknowledge all who made this research possible. This study used the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD (<http://biowulf.nih.gov>), and DNA panels, samples, and clinical data from the National Institute of Neurological Disorders and Stroke Human Genetics Resource Center DNA and Cell Line Repository. People who contributed samples are acknowledged in descriptions of every panel on the repository website. They thank the French Parkinson's Disease Genetics Study Group: Y Agid, M Anheim, A-M Bonnet, M Borg, A Brice, E Broussolle, J-C Corvol, P Damier, A Destée, A Dürr, F Durif, S Klebe, E Lohmann, M Martinez, P Pollak, O Rascol, F Tison, C Tranchant, M Vêrin, F Viallet, and M Vidailhet. They also thank the members of the French 3C Consortium: A Alperovitch, C Berr, C Tzourio, and P Amouyel for allowing us to use part of the 3C cohort, and D Zelenika for support in generating the genome-wide molecular data. They thank P Tienari (Molecular Neurology Programme, Biomedicum, University of Helsinki), T Peuralinna (Department of Neurology, Helsinki University Central Hospital), L Myllykangas (Folkhalsan Institute of Genetics and Department of Pathology, University of Helsinki), and R Sulkava (Department of Public Health and General Practice Division of Geriatrics, University of Eastern Finland) for the Finnish controls (Vantaa85+ GWAS data).

References

- 1000 Genomes Project Consortium. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012; 491:56–65. [PubMed: 23128226]
- ALSGEN Consortium; Ahmeti KB, Ajroud-Driss S, Al-Chalabi A, Andersen PM, Armstrong J, Birve A, Blauw HM, Brown RH, Bruijn L, Chen W, Chio A, Comeau MC, Cronin S, Diekstra FP, Soraya Gkazi A, Glass JD, Grab JD, Groen EJ, Haines JL, Hardiman O, Heller S, Huang J, Hung W-Y, ITALSGEN consortium. Jaworski JM, Jones A, Khan H, Landers JE, Langefeld CD, Leigh PN, Marion MC, McLaughlin RL, Meininger V, Melki J, Miller JW, Mora G, Pericak-Vance MA, Rampersaud E, Robberecht W, Russell LP, Salachas F, Saris CG, Shatunov A, Shaw CE, Siddique N, Siddique T, Smith BN, Sufit R, Topp S, Traynor BJ, Vance C, van Damme P, van den Berg LH, van Es MA, van Vught PW, Veldink JH, Yang Y, Zheng JG. Age of onset of amyotrophic lateral sclerosis is modulated by a locus on 1p34.1. *Neurobiol. Aging*. 2013; 34:357.e7–357.e19. [PubMed: 22959728]
- Chiò A, Schymick JC, Restagno G, Scholz SW, Lombardo F, Lai S-L, Mora G, Fung H-C, Britton A, Arepalli S, Gibbs JR, Nalls M, Berger S, Kwee LC, Oddone EZ, Ding J, Crews C, Rafferty I, Washecka N, Hernandez D, Ferrucci L, Bandinelli S, Guralnik J, Macciardi F, Torri F, Lupoli S, Chanock SJ, Thomas G, Hunter DJ, Gieger C, Wichmann HE, Calvo A, Mutani R, Battistini S, Giannini F, Caponnetto C, Mancardi GL, La Bella V, Valentino F, Monsurro MR, Tedeschi G, Marinou K, Sabatelli M, Conte A, Mandrioli J, Sola P, Salvi F, Bartolomei I, Siciliano G, Carlesi C, Orrell RW, Talbot K, Simmons Z, Connor J, Piro EP, Dunkley T, Stephan DA, Kasperaviciute D, Fisher EM, Jabonka S, Sendtner M, Beck M, Bruijn L, Rothstein J, Schmidt S, Singleton A, Hardy J, Traynor BJ. A two-stage genome-wide association study of sporadic amyotrophic lateral sclerosis. *Hum. Mol. Genet.* 2009; 18:1524–1532. [PubMed: 19193627]
- Cruts M, Theuns J, Van Broeckhoven C. Locus-specific mutation databases for neurodegenerative brain diseases. *Hum. Mutat.* 2012; 33:1340–1344. [PubMed: 22581678]
- Do CB, Tung JY, Dorfman E, Kiefer AK, Drabant EM, Francke U, Mountain JL, Goldman SM, Tanner CM, Langston JW, Wojcicki A, Eriksson N. Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease. *Plos Genet.* 2011; 7:e1002141. [PubMed: 21738487]
- Federoff M, Jimenez-Rolando B, Nalls MA, Singleton AB. A large study reveals no association between APOE and Parkinson's disease. *Neurobiol. Dis.* 2012; 46:389–392. [PubMed: 22349451]
- Grove ML, Yu B, Cochran BJ, Haritunians T, Bis JC, Taylor KD, Hansen M, Borecki IB, Cupples LA, Fornage M, Gudnason V, Harris TB, Kathiresan S, Kraaij R, Launer LJ, Levy D, Liu Y, Mosley T, Peloso GM, Psaty BM, Rich SS, Rivadeneira F, Siscovick DS, Smith AV, Uitterlinden A, van Duijn CM, Wilson JG, O'Donnell CJ, Rotter JI, Boerwinkle E. Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium. *PloS One*. 2013; 8:e68095. [PubMed: 23874508]

- Hindorff, L.; MacArthur, J.; Morales, J.; Junkins, H.; Hall, P.; Klemm, A.; Manolio, T. A Catalog of Published Genome-Wide Association Studies. Bethesda, MD, USA: 2009a.
- Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, Manolio TA. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl. Acad. Sci. U. S. A.* 2009b; 106:9362–9367. [PubMed: 19474294]
- Höglinger GU, Melhem NM, Dickson DW, Sleiman PMA, Wang L-S, Klei L, Rademakers R, de Silva R, Litvan I, Riley DE, van Swieten JC, Heutink P, Wszolek ZK, Uitti RJ, Vandrovцова J, Hurtig HI, Gross RG, Maetzler W, Goldwurm S, Tolosa E, Borroni B, Pastor P, PSP Genetics Study Group, Cantwell LB, Han MR, Dillman A, van der Brug MP, Gibbs JR, Cookson MR, Hernandez DG, Singleton AB, Farrer MJ, Yu C-E, Golbe LI, Revesz T, Hardy J, Lees AJ, Devlin B, Hakonarson H, Müller U, Schellenberg GD. Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. *Nat. Genet.* 2011; 43:699–705. [PubMed: 21685912]
- Hollingworth P, Harold D, Sims R, Gerrish A, Lambert J-C, Carrasquillo MM, Abraham R, Hamshere ML, Pahwa JS, Moskvina V, Dowzell K, Jones N, Stretton A, Thomas C, Richards A, Ivanov D, Widdowson C, Chapman J, Lovestone S, Powell J, Proitsi P, Lupton MK, Brayne C, Rubinsztein DC, Gill M, Lawlor B, Lynch A, Brown KS, Passmore PA, Craig D, McGuinness B, Todd S, Holmes C, Mann D, Smith AD, Beaumont H, Warden D, Wilcock G, Love S, Kehoe PG, Hooper NM, Vardy ERLC, Hardy J, Mead S, Fox NC, Rossor M, Collinge J, Maier W, Jessen F, Rütger E, Schürmann B, Heun R, Kölsch H, van den Bussche H, Heuser I, Kornhuber J, Wiltfang J, Dichgans M, Frölich L, Hampel H, Gallacher J, Hüll M, Rujescu D, Giegling I, Goate AM, Kauwe JSK, Cruchaga C, Nowotny P, Morris JC, Mayo K, Sleegers K, Bettens K, Engelborghs S, De Deyn PP, Van Broeckhoven C, Livingston G, Bass NJ, Gurling H, McQuillin A, Gwilliam R, Deloukas P, Al-Chalabi A, Shaw CE, Tsolaki M, Singleton AB, Guerreiro R, Mühleisen TW, Nöthen MM, Moebus S, Jöckel K-H, Klopp N, Wichmann H-E, Pankratz VS, Sando SB, Aasly JO, Barcikowska M, Wszolek ZK, Dickson DW, Graff-Radford NR, Petersen RC, Alzheimer's Disease Neuroimaging Initiative; van Duijn CM, Breteler MMB, Ikram MA, DeStefano AL, Fitzpatrick AL, Lopez O, Launer LJ, Seshadri S, CHARGE consortium; Berr C, Campion D, Epelbaum J, Dartigues J-F, Tzourio C, Alperovitch A, Lathrop M, EADII consortium. Feulner TM, Friedrich P, Riehle C, Krawczak M, Schreiber S, Mayhaus M, Nicolhaus S, Wagenpfeil S, Steinberg S, Stefansson H, Stefansson K, Snaedal J, Björnsson S, Jonsson PV, Chouraki V, Genier-Boley B, Hiltunen M, Soininen H, Combarros O, Zelenika D, Delepine M, Bullido MJ, Pasquier F, Mateo I, Frank-Garcia A, Porcellini E, Hanon O, Coto E, Alvarez V, Bosco P, Siciliano G, Mancuso M, Panza F, Solfrizzi V, Nacmias B, Sorbi S, Bossù P, Piccardi P, Arosio B, Annoni G, Seripa D, Pilotto A, Scarpini E, Galimberti D, Brice A, Hannequin D, Licastro F, Jones L, Holmans PA, Jonsson T, Riemenschneider M, Morgan K, Younkin SG, Owen MJ, O'Donovan M, Amouyel P, Williams J. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat. Genet.* 2011; 43:429–435. [PubMed: 21460840]
- Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat. Genet.* 2012; 44:955–959. [PubMed: 22820512]
- International HapMap 3 Consortium. Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, Peltonen L, Dermitzakis E, Schaffner SF, Yu F, Peltonen L, Dermitzakis E, Bonnen PE, Altshuler DM, Gibbs RA, de Bakker PIW, Deloukas P, Gabriel SB, Gwilliam R, Hunt S, Inouye M, Jia X, Palotie A, Parkin M, Whittaker P, Yu F, Chang K, Hawes A, Lewis LR, Ren Y, Wheeler D, Gibbs RA, Muzny DM, Barnes C, Darvishi K, Hurler M, Korn JM, Kristiansson K, Lee C, McCarroll SA, Nemesh J, Dermitzakis E, Keinan A, Montgomery SB, Pollack S, Price AL, Soranzo N, Bonnen PE, Gibbs RA, Gonzaga-Jauregui C, Keinan A, Price AL, Yu F, Anttila V, Brodeur W, Daly MJ, Leslie S, McVean G, Moutsianas L, Nguyen H, Schaffner SF, Zhang Q, Ghorji MJR, McGinnis R, McLaren W, Pollack S, Price AL, Schaffner SF, Takeuchi F, Grossman SR, Shlyakhter I, Hostetter EB, Sabeti PC, Adebamowo CA, Foster MW, Gordon DR, Licinio J, Manca MC, Marshall PA, Matsuda I, Ngare D, Wang VO, Reddy D, Rotimi CN, Royal CD, Sharp RR, Zeng C, Brooks LD, McEwen JE. Integrating common and rare genetic variation in diverse human populations. *Nature.* 2010; 467:52–58. [PubMed: 20811451]

- International Parkinson Disease Genomics Consortium. Nalls MA, Plagnol V, Hernandez DG, Sharma M, Sheerin U-M, Saad M, Simón-Sánchez J, Schulte C, Lesage S, Sveinbjörnsdóttir S, Stefánsson K, Martínez M, Hardy J, Heutink P, Brice A, Gasser T, Singleton AB, Wood NW. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet*. 2011; 377:641–649. [PubMed: 21292315]
- International Parkinson's Disease Genomics Consortium (IPDGC) Wellcome Trust Case Control Consortium 2 (WTCCC2). A two-stage meta-analysis identifies several new loci for Parkinson's disease. *Plos Genet*. 2011; 7:e1002142. [PubMed: 21738488]
- Laaksovirta H, Peuralinna T, Schymick JC, Scholz SW, Lai S-L, Myllykangas L, Sulkava R, Jansson L, Hernandez DG, Gibbs JR, Nalls MA, Heckerman D, Tienari PJ, Traynor BJ. Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: a genome-wide association study. *Lancet Neurol*. 2010; 9:978–985. [PubMed: 20801718]
- Lambert J-C, Heath S, Even G, Campion D, Sleegers K, Hiltunen M, Combarros O, Zelenika D, Bullido MJ, Tavernier B, Letenneur L, Bettens K, Berr C, Pasquier F, Fiévet N, Barberger-Gateau P, Engelborghs S, De Deyn P, Mateo I, Franck A, Helisalmi S, Porcellini E, Hanon O, European Alzheimer's Disease Initiative Investigators, de Pancorbo MM, Lendon C, Dufouil C, Jaillard C, Leveillard T, Alvarez V, Bosco P, Mancuso M, Panza F, Nacmias B, Bossù P, Piccardi P, Annoni G, Seripa D, Galimberti D, Hannequin D, Licastro F, Soininen H, Ritchie K, Blanché H, Dartigues J-F, Tzourio C, Gut I, Van Broeckhoven C, Alperovitch A, Lathrop M, Amouyel P. Genome-wide association study identifies variants at *CLU* and *CR1* associated with Alzheimer's disease. *Nat. Genet*. 2009; 41:1094–1099. [PubMed: 19734903]
- Lill CM, Roehr JT, McQueen MB, Kavvoura FK, Bagade S, Schjeide B-MM, Schjeide LM, Meissner E, Zauft U, Allen NC, Liu T, Schilling M, Anderson KJ, Beecham G, Berg D, Biernacka JM, Brice A, DeStefano AL, Do CB, Eriksson N, Factor SA, Farrer MJ, Foroud T, Gasser T, Hamza T, Hardy JA, Heutink P, Hill-Burns EM, Klein C, Latourelle JC, Maraganore DM, Martin ER, Martínez M, Myers RH, Nalls MA, Pankratz N, Payami H, Satake W, Scott WK, Sharma M, Singleton AB, Stefánsson K, Toda T, Tung JY, Vance J, Wood NW, Zabetian CP, Young P, Tanzi RE, Khoury MJ, Zipp F, Lehrach H, Ioannidis JPA, Bertram L. Comprehensive research synopsis and systematic meta-analyses in Parkinson's disease genetics: the PDGene database. *Plos Genet*. 2012; 8:e1002548. [PubMed: 22438815]
- Mok K, Traynor BJ, Schymick J, Tienari PJ, Laaksovirta H, Peuralinna T, Myllykangas L, Chiò A, Shatunov A, Boeve BF, Boxer AL, DeJesus-Hernandez M, Mackenzie IR, Waite A, Williams N, Morris HR, Simón-Sánchez J, van Swieten JC, Heutink P, Restagno G, Mora G, Morrison KE, Shaw PJ, Rollinson PS, Al-Chalabi A, Rademakers R, Pickering-Brown S, Orrell RW, Nalls MA, Hardy J. Chromosome 9 ALS and FTD locus is probably derived from a single founder. *Neurobiol. Aging*. 2012; 33:209.e3–209.e8. [PubMed: 21925771]
- Nalls MA, Pankratz N, Lill CM, Hernandez D, Saad M, DeStefano AL, Kara E, Bras JM, Sharma M, Schulte C, Keller MF, Arepalli S, Letson C, Edsall C, Stefánsson H, Liu X, Pliner H, Lee J, Cheng R, IPDGC; PSG-PROGENI; 23andMe; NGRC; HIHG; AJDI; CHARGE; NABEC; UKBEC; GPDC; AGAG. Ikra MA, Ioannidis JPA, Hadjigeorgiou GM, Bis JC, Martínez M, Perlmutter J, Goate A, Marder K, Fiske B, Sutherland M, Xiromerisiou G, Myers RH, Clark LN, Stefánsson K, Hardy JA, Heutink P, Chen H, Wood NW, Houlden H, Payami H, Brice A, Scott WK, Gasser T, Bertram L, Eriksson N, Foroud T, Singleton AB. Large scale meta analysis of genome-wide association data in Parkinson's disease reveals 6 novel risk loci. 2014 In press.
- Nalls MA, Simon-Sanchez J, Gibbs JR, Paisan-Ruiz C, Bras JT, Tanaka T, Matarin M, Scholz S, Weitz C, Harris TB, Ferrucci L, Hardy J, Singleton AB. Measures of autozygosity in decline: globalization, urbanization, and its implications for medical genetics. *Plos Genet*. 2009; 5:e1000415. [PubMed: 19282984]
- NHLBI GO Exome Sequencing Project. Exome Variant Server. Seattle, WA: 2011.
- Paisán-Ruiz C, Jain S, Evans EW, Gilks WP, Simón J, van der Brug M, López de Munain A, Aparicio S, Gil AM, Khan N, Johnson J, Martínez JR, Nicholl D, Carrera IM, Pena AS, de Silva R, Lees A, Martí-Massó JF, Pérez-Tur J, Wood NW, Singleton AB. Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. *Neuron*. 2004; 44:595–600. [PubMed: 15541308]
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010; 38:e164. [PubMed: 20601685]

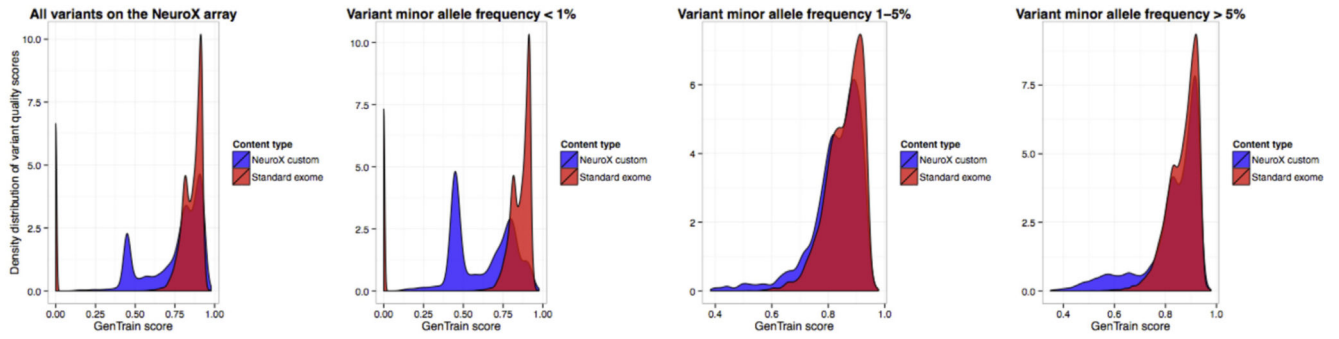


Fig. 1.

For both NeuroX custom content and the standard content included on the array, most of the SNPs across the minor allele frequency spectrum have GenTrain scores >0.7 , suggesting quality genotype clusters are readily available. Discrepancies across content type are partially because of genotype cluster method differences between the 2 sets of variants (custom and standard content) but also because of the inclusion of rare and difficult to genotype loci in the custom content of the array. Abbreviation: SNPs, single-nucleotide polymorphisms. (For interpretation of the references to color in this Figure, the reader is referred to the web version of this article.)

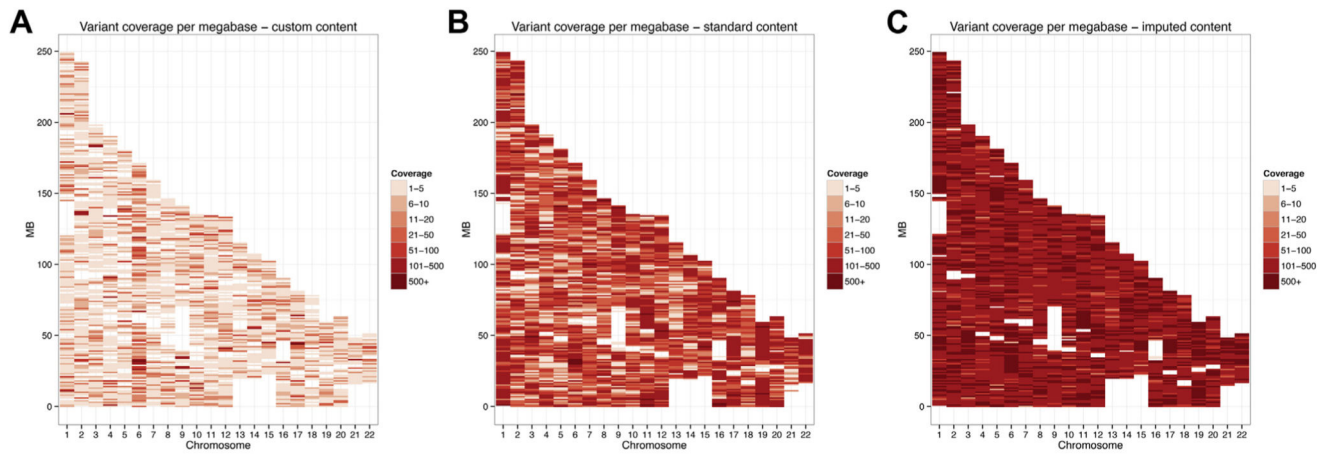


Fig. 2.

Autosomal variant coverage per megabase for different content classes. Panel A, custom content coverage; panel B, standard content coverage; and panel C, coverage for successfully imputed variants (imputation quality >0.30). (For interpretation of the references to color in this Figure, the reader is referred to the web version of this article.)

Table 1

Comparison of content across standard and custom content

Content type	Custom content	Standard content
Numebr of variants	24,706	242,901
Variants less than MAF 0.01 (%)	31.531	82.277
Variants less than MAF 0.05 (%)	40.047	86.078
Variants at MAF 0.05 to 0.50 (%)	59.953	13.922
Mean MAF	0.148	0.031
Exonic variants (%)	36.151	96.504
Nonsynonymous coding variants (% of exonic)	33.934	91.332

Data are based on clustering of over 14,000 Parkinson's disease cases and controls as described in (Nalls et al., 2014). All annotations from ANNOVAR (Wang et al., 2010). Key: MAF, minor allele frequency.