Mutation analysis of patients with Neurodegenerative disorders using NeuroX array

Mahdi Ghani\textsuperscript{a}, Anthony E Lang\textsuperscript{a,b,c}, Lorne Zinman\textsuperscript{d,c}, Benedetta Nacmias\textsuperscript{e}, Sandro Sorbi\textsuperscript{e}, Maria Carmela Tartaglia\textsuperscript{a,c}, Ezequiel I Surace\textsuperscript{1}, Christine Sato\textsuperscript{a}, Danielle Moreno\textsuperscript{a}, Zhengrui Xi\textsuperscript{a}, Rachel Hung\textsuperscript{a}, Mike A Nalls\textsuperscript{g}, Andrew Singleton\textsuperscript{a}, Peter St George-Hyslop\textsuperscript{a,c,h,*}, and Ekaterina Rogaeva\textsuperscript{a,c,*}

\textsuperscript{a}Tanz Centre for Research in Neurodegenerative Diseases, University of Toronto, 60 Leonard Street, Toronto, Ontario, Canada M5T 2S8
\textsuperscript{b}Morton and Gloria Shulman Movement Disorders Center and the Edmond J. Safra Program in Parkinson’s Disease, Toronto Western Hospital; Division of Neurology, Department of Medicine, University of Toronto, Ontario, Canada
\textsuperscript{c}Department of Medicine, Division of Neurology, University of Toronto, 1 King’s College Circle, Toronto, Ontario, Canada M5S 1A8
\textsuperscript{d}Sunnybrook Health Sciences Centre, 2075 Bayview Ave., Toronto, Ontario, Canada M4N 3M5
\textsuperscript{e}Department of Neuroscience, Psychology, Drug Research and Child Health (NEUROFARBA), University of Florence, Florence, Italy
\textsuperscript{f}Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Laboratorio de Biología Molecular, Instituto de Investigaciones Neurológicas Dr. Raúl Carrea (FLENI), Buenos Aires, Argentina
\textsuperscript{g}Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD 20892, USA
\textsuperscript{h}Cambridge Institute for Medical Research, and the Department of Clinical Neurosciences, University of Cambridge, Hills Road, Cambridge, UK CB2 0XY

Abstract

Genetic analyses of patients with neurodegenerative disorders have identified multiple genes that need to be investigated for the presence of damaging variants. However, mutation analysis by Sanger sequencing is costly and time consuming. We tested the utility of a recently designed semi-
custom genome-wide array (NeuroX; Illumina, Inc) tailored to study neurodegenerative diseases (e.g. mutation screening). We investigated 192 patients with four different neurodegenerative disorders for the presence of rare damaging variations in 77 genes implicated in these diseases. Several causative mutations were identified and confirmed by Sanger sequencing including PSEN1 p.M233T responsible for Alzheimer’s disease in a large Italian family, as well as SOD1 p.A4V and p.I113T in patients with Amyotrophic Lateral Sclerosis. In total, we identified 78 potentially damaging rare variants (frequency <1%), including ABCA7 p.L400V in a family with Alzheimer’s disease and LRRK2 p.R1514Q in 6 out 98 patients with Parkinson’s Disease (6.1%). In conclusion, NeuroX appears to be helpful for rapid and accurate mutation screening, although further development may be still required to improve some current caveats.

Keywords
NeuroX; mutation; neurodegenerative disorder

1. Introduction

Alzheimer’s disease (AD) together with other common neurodegenerative disorders, such as Parkinson’s disease (PD), Frontotemporal Dementia (FTD) and Amyotrophic Lateral Sclerosis (ALS) have major socio-economic effects in different societies. The world-wide prevalence of dementia doubles every 20 years and the health-care cost for people with dementia is increasing almost exponentially, costing up to US$604 billion in 2010 (World Health Organization, 2012). Knowledge of genetic factors causing or modifying risk of neurodegenerative disorders could help manage the challenges of aging populations.

Rare causal mutations are highly penetrant, while common variants with >5% minor allele frequency (MAF), mainly assessed by Genome Wide Association Studies (GWASs), only modestly influence disease risk. For instance, analyses of autosomal dominant families with early-onset AD (<65 years) have found causal mutations in APP, PSEN1 and PSEN2; while GWASs identified or confirmed >20 loci associated with late-onset AD (APOE, ABCA7, BIN1, CASS4, CD2AP, CD33, CELF1, CLU, CR1, DSG2, EPHA1, FERTM2, HLA-DRB1, HLA-DRB5, INPP5D, MEF2C, MS4A4A, MS4A6A, NME8, PICALM, PTK2B, SLC24A4/RIN3, ZCWPW1 and SORL1) (Ghani and Rogaeva, 2014). Importantly, GWAS loci could harbour rare damaging variants with a greater risk-effect vs. common variants within the same gene. For example, both rare coding and common non-coding CLU variants are independently associated with AD (Bettens, et al., 2012). However, the role of copy number variations (CNVs) and recessive inheritance (e.g. estimated based on runs of homozygosity) need further exploration (Ghani, et al., 2012, Ghani, et al., 2013).

The genetics of other neurodegenerative disorders is also complex. GWAS and family-studies have implicated at least 24 loci for PD (ACMSD, ATP13A2, BST1, FBXO7, FGF20, GAK, GBA, GIGYF2, GPNMB, HIP1R, HLA-DRB5, LAMP3, LRRK2, MAPT, PARK2, PARK7, PINK1, PLA2G6, SNCA, STBD1, STK39, STX1B, SYT11 and VPS35) (International Parkinson's Disease Genomics Consortium (IPDGC); Wellcome Trust Case Control Consortium 2 (WTCCC2), 2011,Nalls, et al., 2011,Singleton, et al., 2013,Valente, et
al., 2004, van Duijn, et al., 2001). Similarly, ALS and FTD might be explained by mutations in several often overlapping genes (Hardy and Rogaeva, 2013). To date 26 genes have been identified for FTD and/or ALS (ALS2, ANG, ATXN2, C9orf72, CHMP2B, DAO, EWSR1, FIG4, FUS, GRN, HNRNPA1, HNRNPA2B1, MAPT, OPTN, PFN1, SETX, SIGMAR1, SOD1, SPG11, SQSTM1, TAF15, TARDBP, TMEM106B, UBQLN2, VAPB and VCP) (Ling, et al., 2013).

There are many similarities that connect different neurodegenerative disorders. For instance, rare TREM2 variants (e.g. p.R47H) are associated with AD (Guerreiro, et al., 2013), while the p.R47H has been reported as a risk for PD, FTD and ALS (Benitez and Cruchaga, 2013, Cady, et al., 2014, Rayaprolu, et al., 2013). Hence, the overlap between different genes implicated in neurodegenerative disorders has to be systematically explored using novel genotyping technologies that could be crucial for the fast, reliable and concurrent screening of known genes, instead of costly and time-consuming gene-by-gene Sanger sequencing (Guerreiro, et al., 2014).

Here, we tested the utility of a recently designed semi-custom genome-wide array (NeuroX; Illumina, Inc) tailored to investigate neurodegenerative diseases. It has the standard content of the Illumina Exome BeadChip (~240,000 coding variants) and ~24,000 custom variants, including reported mutations causing neurodegenerative disorders, as well as risk variants found by GWASs (Nalls, et al., 2014). Hence, NeuroX is suitable for both mutation screening and association studies (Kara, et al., 2014). We evaluated 192 patients affected by AD, PD, FTD or ALS and catalogued the coding variations detected in a subset of 77 genes implicated in these disorders.

2. Method

2.1 Human Samples

Informed consent was obtained from all participants in accordance with the respective ethical review board. Study participants were mainly Canadian of North European origin, recruited from hospitals specialized in neurodegenerative disorders. Cases with known pathological mutations were excluded from the study as described previously (Xi, et al., 2012), with the exception of cases (n=9) that were used as internal controls to evaluate the reliability of the NeuroX array.

We investigated 192 subjects, including 98 PD, 43 ALS, 37 AD and 14 FTD patients prioritized for having an early age of onset and/or family history of neurodegenerative disease (Table 1). Most of the samples were unrelated (n=179), however for 13 familial patients we included a second affected family member to prioritise variants for segregation analysis. Family members of FLO25, FLO61 and FLO66 were genotyped by Sanger sequencing for the PSEN1, ABCA7 or TREM2 mutations detected by NeuroX (assays are available upon request).

2.2 Analysis of NeuroX array

The NeuroX data was loaded to GenomeStudio (Illumina Inc.). All markers were clustered using the default Gen Call threshold (0.15); 97.3% of markers had Gen Train score >0.7,
indicating high genotype quality. Genotypes were then converted to PLINK input files and allele frequencies were calculated. ANNOVAR annotation of the 267,151 single nucleotide polymorphisms (SNPs) on NeuroX revealed that this array can assess 230,921 exonic, 2,068 intronic- and 63 exonic-splicing SNPs (2 bp from an exon/intron boundary). The coding SNPs consisted of 213,892 non-synonymous, 9,639 synonymous, 5,242 stop-gain, and 219 stop-loss variations. The average call rate was 99.7% among all samples. Only one FTD sample was removed from the study due to poor DNA quality as indicated by the extensive chromosomal abnormalities revealed by cnvPartition.

Markers within 77 genes implicated in AD, PD, FTD and/or ALS (Supplemental Table 1) were extracted (n=4,441), including 1,945 non-synonymous, 163 synonymous, 95 stop-gain, 36 splicing variants and 1 stop-loss variant. Non-polymorphic markers (n=2,359) and markers missing genotypes in >5% of the samples (n=99) were excluded. We also removed non-coding (n=1,654), synonymous (n=43) and common variants with MAF>1% at the 1000Genomes (1000g2012apr_all) and Exome Variant Server (esp6500si_all) databases (n=138). Finally, variants overlapping segmental duplications were excluded due to possible genotyping error (Ghani, et al., 2013b). The remaining variants were filtered to those predicted to have a probable damaging effect on protein function, according to either PolyPhen (LJB2_PP2_HDIV_Pred) or SIFT (LJB2_SIFT) analyses implemented in ANNOVAR. Cases with damaging variants were extracted using PLINK.

3. Results

3.1 Analysis of internal controls

Five samples with heterozygous missense mutations previously detected by Sanger sequencing were all accurately recognized by NeuroX (Supplemental Figure 1a-d): an ALS case with SOD1 p.I114T (NeuroX_21:33039672), an AD case with PSEN1 p.A246E (NeuroX_rs63750526), a PD case with LRRK2 p.L1795F (NeuroX_rs111910483) and two PD cases with LRRK2 p.G2019S (exm994671). Also, a pair of previously reported monozygotic twins discordant for ALS (Xi Z, 2014) revealed identical genotypes for all NeuroX markers (n=267,607), except one in the TTC30B gene (exm244263). Sanger sequencing did not confirm this discrepancy, which was likely caused by a segmental duplication affecting the locus.

Four PD samples with known PARK2 CNVs were examined using cnvPartition and GenomeStudio (Supplemental Figure 2a-d). CnvPartition detected the heterozygous deletion affecting exons 3-6 in sample 6045 with CNV-values of 1 for 31 markers spanning the locus Chr6:162394342-162687778 (hg19); however it did not recognize the heterozygous duplications of exons 7 and 9 (sample 4932), exon 5 (sample 4033) or homozygous deletion of exon 2 (sample 6140). GenomeStudio also did not detect an increase in signal intensity for the duplication regions (likely due to the low density of NeuroX markers); however in sample 6140 a homozygous deletion is indicated by reduced signal intensity and an absence of genotype calls for 10 out of 15 NeuroX markers in exon 2. Notably, the 5 false-positive genotype calls included damaging variations (e.g. InDel Asp53Stop and p.Q34R), suggesting the need for careful validation of some NeuroX calls.
3.2 Discovery of new cases with Mendelian pathological mutations

A PSEN1 p.M233T mutation was detected in an Italian autosomal dominant FLO25 family with six members affected by AD (Figure 1a). The p.M233T substitution was found by heterozygosity of marker NeuroX_rs63751024 in two AD cases (8994 and 8995) (Supplemental Figure 1e). It was confirmed by Sanger sequencing, which revealed 8 carriers of the p.M233T among family members (Figure 1a). This pathogenic mutation was reported in several early onset AD families (Campion, et al., 1999, Guerreiro, et al., 2010, Kwok, et al., 1997, Park, et al., 2008, Raux, et al., 2005).

In addition to the internal control with the SOD1 p.I114T, we found the same mutation in patient 9108 with familial ALS (onset at age 50; disease duration 6 months). Sanger sequencing also detected the p.I114T in a currently unaffected sibling of patient 9108. Moreover, two unrelated familial ALS patients were heterozygous for SOD1 p.A5V (NeuroX_21:33032096); patient 9572 had onset at age 42 and a rapid ALS course (6 months), while the disease in patient 9432 was less severe with onset at age 58 and one year duration. Sanger sequencing confirmed the mutation and revealed the p.A5V in a currently asymptomatic sibling of patient 9432. The p.A5V is a common SOD1 mutation in the USA (also known as p.A4V, since the start methionine is post-translationally cleaved), while p.I114T is common in the UK (Battistini, et al., 2010, Niemann, et al., 2004).

3.4 Rare coding variants in 77 genes related to neurodegenerative disorders

A description of the variants identified in genes associated with four neurodegenerative disorders, their frequency in public databases, and predicted consequence on protein function are presented in Supplemental Table 2. In total, we detected 78 rare damaging SNPs observed in 98 out of 192 genotyped patients, including subjects with three (n=13) and two (n=22) concurrent variants. Of note, 211 InDel markers can be assessed by NeuroX, including 42 markers within the 77 genes of interest, of which only two were polymorphic, but not confirmed by Sanger sequencing.

3.5 Rare variants among PD patients

Several rare variants appeared to be relatively common (MAF >1%) in our PD cohort (Supplemental Table 2). For instance, the LRRK2 p.R1514Q variation (rs35507033 with MAF=0.2% in 1000Genomes) was heterozygous in 6 out of 98 unrelated PD samples (6.1%). In a previous study the p.R1514Q substitution was not significantly associated with sporadic late onset PD (>60 years). In datasets of different European origins, the MAF of p.R1514Q ranged from 1.0-3.3% in cases and 0-3.8% in controls (Toft, et al., 2007). The relatively high MAF of p.R1514Q in our study suggests that pathogenicity of the p.R1514Q substitution has to be further investigated in early onset PD dataset, since mean age at onset in our PD patients was only 41 years (Table 1).

We also identified several other very rare damaging variants, including TREM2 p.A105V in PD patient 5908 with age at onset of 46 (absent in 1000Genomes; MAF=0.022% at ESP); and APP p.E468K PD in patient 6892 with age at onset of 38 (MAF=0.074% at dbSNP138). Of note, AD mutations in APP are limited to duplications of the gene or missense mutations in exons 16 and 17, near the secretase processing sites (678-724 residues) (Ghani and
However, the E468K variant is located in the soluble secreted part of the APP protein (sAPPβ). Interestingly, sAPPβ is reported to regulate Transthyretin (TTR), a potential biomarker for preclinical diagnosis of PD (Arguelles, et al., 2010, Li, et al., 2010). Also, two rare variants in exon 26 of GIGYF2 (PD-associated gene) were detected in familial patient 6171 with onset at age 30 (p.H1165R) and patient 8563 with onset at age 39 (p.P1149T).

In addition, we detected two potentially damaging PARK2 variants: p.A46T in patient 7932 (age at onset of 30) and p.R253C in patient 7031 (age at onset of 41). Both patients had a family history of PD, however heterozygous mutations in this recessive PD gene have questionable pathological significance (Marras, et al., 2010). It is possible that disease manifestation in such cases is influenced by concurrent damaging variants. For instance, in PD patient 4033 with known heterozygous duplication of PARK2 exon 5 (age of onset 40) we also detected MAPT p.P140S, ABCA7 p.E316K and TREM2 p.H157Y. However, the risk associated with several variants is challenging to address. A large dataset would be required to determine whether genetic interactions manifest specific cellular/neuropathological phenotypes or lead to an earlier age of onset.

### 3.6 Rare variants among AD patients

We detected several mutations in genes reported by AD GWAS (ABCA7, ZCWPW1, SORL1, DSG2 and TREM2) (Lambert, et al., 2013). Two AD cases from Italian FLO61 family had an extremely rare heterozygous ABCA7 p.L400V variant absent in the ESP and 1000Genomes databases. Sanger sequencing confirmed the variant and detected another carrier that was unaffected at age 48 (Figure 1b). Both patients were also heterozygous for a rare ZCWPW1 p.H351R variant. Furthermore, two AD affected siblings of the Canadian TOR146 family (3515 and 3592), revealed the SORL1 p.H1813Q variant (MAF=0.003 in 1000Genomes), which was recently reported to be associated with AD in the analysis of probands from Caribbean Hispanic families (Vardarajan, et al., 2014). Notably, the same individuals in TOR146 are carriers of the DSG2 p.V56M variant.

First cousins with AD from Italian FLO66 family (8136 and 8145) were heterozygous for the extremely rare TREM2 p.D87N variant (MAF=0.074% in dbSNP138) affecting a residue conserved in all mammals. Sanger sequencing revealed that the p.D87N does not completely segregate with AD, since it was found in 4 out of 5 affected family members and in a subject unaffected by AD at age 80 (Figure 1c). However, p.D87N is likely a pathological variant with incomplete penetrance, as it was previously observed in 6 out of 1091 unrelated AD cases but not in 1105 controls (p=0.02) (Guerreiro, et al., 2013).

### 3.7 Rare variants among ALS patients

Heterozygous ATXN2 variants were detected in a familial ALS patient 9655 (p.M986V) and in an autopsy confirmed TDP43-positive ALS case 8861 (p.P954S). In another TDP43-positive ALS case 8851, we identified an extremely rare heterozygous p.D543N substitution at a highly conserved residue of ALS2, a known recessive gene causing the Juvenile form of ALS (Hadano, et al., 2001). However, the pathological significance of heterozygous ALS2 mutations remain to be investigated.
In ALS patient 8805 (age of onset 55) we observed two rare heterozygous variants (MAPT p.T17M and SETX p.C1554G). Currently it is unclear whether both variants contribute to the phenotype of this patient. Thus far, only three heterozygous SETX mutations were implicated in the Juvenile form of ALS (p.T3I, p.L389S and p.R2136H) (Chen, et al., 2004), and MAPT has not been associated with ALS.

4. Discussion

NeuroX is quite useful in the rapid screening of known mutations associated with neurodegenerative diseases. We identified several mutations reported to cause AD, PD or ALS. For instance, we detected the PSEN1 p.M233T mutation responsible for AD in a large family, which could be of clinical value, since carriers of such mutations might be enrolled in the longitudinal clinical AD trial of Dominantly Inherited Alzheimer Network (DIAN; http://www.dian-info.org/).

NeuroX allows the reliable concurrent investigation of rare coding variants in multiple genes for each subject, which is crucial for large studies investigating genes that could interact in the same network. For instance, rare variants in two AD-related genes (ABCA7 p.L400V and ZCWPW1 p.H351R) were observed in two related AD patients. NeuroX was also useful in detecting extremely rare variants. For example, we identified the very rare ALS2 p.D543N variant in an ALS patient. Several homozygous deletions in ALS2 were described in a young onset form of ALS (MIM #606352). Two affected sisters with Infantile-onset Ascending Hereditary Spastic Paralysis are the only cases described with a homozygous ALS2 missense mutation (p.C156Y) affecting a conserved residue in the Regulator of Chromatin Condensation (RCC1)-like domain and leading to down-regulation of the mutant protein (Eymard-Pierre, et al., 2006). Notably, the p.D543N mutation is also located in the same domain and its carrier needs to be investigated for the level of ALS2 protein and compound heterozygosity of ALS2.

Nevertheless, NeuroX has some limitations similar to other genome-wide arrays. For instance, the APOE alleles defined by SNPs rs429358 and rs7412 are not reliably detected (e.g. NeuroX markers for rs429358 were non-polymorphic in our study), since the APOE locus has high GC-content and its genotyping requires special conditions (Addya, et al., 1997). Also, NeuroX cannot evaluate repeat expansions (e.g. in C9orf72) known to be associated with several neurological disorders. In addition, CNVs, composing >12% of the human genome (Stankiewicz and Lupski, 2010) and implicated in many neuropsychiatric disorders (Cook and Scherer, 2008), are not reliably identified by NeuroX. However, the enriched probe density of NeuroX at a few loci (e.g. SNCA) could allow CNVs to be estimated based on signal intensities.

Furthermore, NeuroX contains only 211 InDel markers, while several million InDels are found throughout the human genome and many of them map to functionally important sites (Mullaney, et al., 2010). Also, careful consideration may be required when a NeuroX InDel marker appears polymorphic. For instance, an InDel marker in GRN (NeuroX_rs63750768) was found to be heterozygous in several of our samples but not confirmed by Sanger sequencing, because this 4-bp InDel variation overlaps a common SNP (rs25646) that
interferes with the result of the InDel marker. Moreover, SNP probes on this array might occasionally have design problems producing misleading results. For example, the marker NeuroX_15:44856873 (rs80338869) was wrongly reported to be a premature stop codon in SPG11, however rs80338869 has three alleles and the array results reflect the common synonymous minor allele (MAF ~2%).

In summary, NeuroX provides rapid and accurate detection of coding variations within genes associated with neurodegenerative disorders, except a few loci and could be the first step in mutation analysis. For clinical diagnostic purposes, it can be supplemented with standard accredited genotyping methods to address its limitations.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by grants from the Canadian Institutes of Health Research, Ontario Research Fund, the W. Garfield Weston Foundation (ER, PSH); the Intramural Research Program of the National Institute on Aging, National Institutes of Health (NIH), part of the Department of Health and Human Services; project number ZO1 AG000958-11 (AS).

References


- Utility of novel genome-wide NeuroX array tailored to neurodegenerative diseases was tested.
- We evaluated 192 patients for damaging variations in neurodegenerative disease genes.
- 78 potentially damaging rare variants and causative mutations were identified.
- NeuroX is helpful for rapid and accurate mutation screening.
- Further development may be still required to improve some current NeuroX caveats.
Figure 1.
Diagrams of three Italian families affected by Alzheimer’s Disease, in which mutations were identified by NeuroX. Genotypes are shown below the ID number with the mutant allele highlighted in bold font. Where available age at the time of examination is shown in the upper right corner and age of onset is indicated below the ID number. The gender of family members is masked to protect patient confidentiality. a) FLO25 with the PSEN1 p.M233T mutation (T>C). b) FLO61 with the rare ABCA7 p.L400V variant (C>G). c) FLO66 with the TREM2 p.D87N (G>A).
Table 1

Samples characteristics in different disease categories

<table>
<thead>
<tr>
<th>Diagnosis</th>
<th>N cases</th>
<th>Mean Age at onset (SD) years</th>
<th>Familial cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>37</td>
<td>64 (14)</td>
<td>100%</td>
</tr>
<tr>
<td>ALS</td>
<td>42</td>
<td>55 (11.2)</td>
<td>57%</td>
</tr>
<tr>
<td>FTD</td>
<td>14</td>
<td>58 (10.5)</td>
<td>100%</td>
</tr>
<tr>
<td>PD</td>
<td>98</td>
<td>41 (7.9)</td>
<td>49%</td>
</tr>
</tbody>
</table>