Georgia State University

ScholarWorks @ Georgia State University

Psychology Faculty Publications

Department of Psychology

4-2013

Increased CNV-Region Deletions in Mild Cognitive Impairment (MCI) and Alzheimer's Disease (AD) Subjects in the ADNI Sample

Guia Guffanti
Columbia University, guffant@nyspi.columbia.edu

Federica Torri
University of California, Irvine, ftorri@uci.edu

Jerod Rasmussen
University of California, Irvine, rasmissj@uci.edu

Andrew P. Clark
University of Southern California, clarkap@usc.edu

Anita Lakatos
University of California, Irvine, alakatos@uci.edu

See next page for additional authors

Follow this and additional works at: https://scholarworks.gsu.edu/psych_facpub

Part of the Psychology Commons

Recommended Citation

Authors

Guia Guffanti, Federica Torri, Jerod Rasmussen, Andrew P. Clark, Anita Lakatos, Jessica Turner, James H. Fallon, Andew Saykin, Michael Weiner, Marquis P. Vawter, James A. Knowles, Steven G. Potkin, and Fabio Macciardi

Review

Increased CNV-Region deletions in mild cognitive impairment (MCI) and Alzheimer's disease (AD) subjects in the ADNI sample

Guia Guffanti [a,1], Federica Torri [b,1], Jerod Rasmussen [b], Andrew P. Clark [c], Anita Lakatos [b],
Jessica A. Turner [d], James H. Fallon [b], Andrew J. Saykin [e],
Michael Weiner [f] ADNI the Alzheimer's Disease Neuroimaging Initiative [2], Marquis P. Vawter [b,g],
James A. Knowles [c], Steven G. Potkin [b], Fabio Macciardi [b,h,i,*]

[a] Department of Psychiatry, Division of Child and Adolescent Psychiatry, Columbia University/NYSPI, New York, NY 10032, USA
[b] Department of Psychiatry and Human Behavior, University of California, Irvine, CA 92617, USA
[c] Department of Psychiatry and the Zilkha Neurogenetic Institute, University of Southern California, Los Angeles, CA 90033, USA
[d] Mind Research Network, Albuquerque, NM 87106, USA
[e] Department of Radiology Imaging Sciences, Indiana University School of Medicine, Indianapolis IN, 46206, USA
[f] Department of Radiology, Psychiatry and Neurology, University of California San Francisco, San Francisco, CA 94121, USA
[g] Functional Genomics Laboratory, Department of Psychiatry And Human Behavior, School of Medicine, University of California, Irvine, CA 92697, USA
[h] Center for Epigenetics and Metabolism, School of Medicine, University of California, Irvine, CA 92617, USA
[i] Dept of Pharmacological and Biomolecular Sciences, University of Milan, Italy

ARTICLE INFO

ABSTRACT

We investigated the genome-wide distribution of CNVs in the Alzheimer's disease (AD) Neuroimaging Initiative (ADNI) sample (146 with AD, 313 with Mild Cognitive Impairment (MCI), and 181 controls).
Comparison of single CNVs between cases (MCI and AD) and controls shows overrepresentation of large heterozygous deletions in cases (p-value < 0.0001). The analysis of CNV-Regions identifies 44 copy number variable loci of heterozygous deletions, with more CNV-Regions among affected than controls (p = 0.005). Seven of the 44 CNV-Regions are nominally significant for association with cognitive impairment. We validated and confirmed our main findings with genome re-sequencing of selected patients and controls. The functional pathway analysis of the genes putatively affected by deletions of CNV-Regions reveals enrichment of genes implicated in axonal guidance, cell–cell adhesion, neuronal morphogenesis and differentiation. Our findings support the role of CNVs in AD, and suggest an association between large deletions and the development of cognitive impairment

© 2013 Elsevier Inc. All rights reserved.

Contents

1. Introduction

Copy number variations (CNVs) represent an important source of genetic diversity that can affect biological functions. With the advent of genome-wide tools, CNV mapping on a genomic scale has proven to be crucially relevant to integrate the information provided by SNP microarray technologies in studies of complex traits [1]. The recent availability of high-throughput genotyping technologies has propelled an increasing number of genome-wide association studies (GWAS) of Alzheimer's disease (AD) producing a progressive number of candidate genes (http://www.alzgene.org/). Nevertheless, the risk factors identified in AD patients collectively explain a relatively modest amount of the total genetic risk for the disease. CNV may be an important piece to complete the puzzle of the "missing heritability" [2] in the study of cognitive impairment and AD. Recent studies supported a substantial role of chromosomal structural variations in the pathogenesis of several neurological disorders [3]. This evidence is consistent with the hypothesis that genetic susceptibility to late onset complex disorders such as Alzheimer's disease (AD), Amyotrophic Lateral Sclerosis (ALS) and Parkinson's disease (PD) depends on dosage-sensitive loci directly affected by copy number variations [4]. CNV position and dose-related effects may have a strong indirect influence on gene expression, as supported by recent studies on CNV's global role in shaping the human transcriptome [5]. These effects may be further enhanced by the increased action of environmental age-related factors [6].

Three recent studies screened for an effect of CNVs in three different samples of late-onset AD [7–9] using both whole genome and candidate gene approaches. Despite the lack of statistically significant results, several genes were deemed interesting for further investigations since they were modestly enriched in CNVs in cases compared to controls. Whole genome SNP-microarray CNV studies may generally fail to reach statistical significance in case–control study designs because of the rarity of the events. A significant methodological challenge is identifying a common pattern of variation across different subjects when structural events have different sizes. Large structural events partially overlapping with previously identified smaller CNVs in the general population are usually not considered novel, but the effect of a very large CNV may be totally different from those of much smaller CNVs. To address this concern, Redon et al. [10] introduced the concept of loci-encompassing CNV-Regions. CNV-Regions can provide a more realistic representation of the distribution of structural rearrangements across subjects than single copy number events [11]. The effect size of a CNV may explain only a small proportion of the whole genetic variability underlying common disorders, since rarely a single CNV has a high frequency in the general population [12]. On the contrary, a combination of different CNVs in different individuals, can alter biological functions in the same important pathway and result in a much larger effect, although identifying the correct pattern of jointly acting CNVs is complex. In this context the proximity of CNVs to functional genetic variants and regulatory elements is likely to be critical. Recent papers have described how CNVs are strongly involved in controlling overall gene expression [5]: their genome-wide distribution is non-random, and is strongly correlated with genomic features like exons, segmental duplications, repetitive elements (e.g. Alu elements) [1] and microRNAs [13]. Interestingly, CNVs can shape the expression level of a gene not only by gene dosage, but also by inducing a variety of epigenetic modifications some of which can act on genes located more than half a megabase from the physical location of the CNV [5,14].

A further challenge is developing a copy number detection algorithm that is not biased toward the specific probe-content of different available commercial array platforms [15]. Different algorithms and parameter optimization can lead to substantial differences in CNV detection [16,17]. Another critical aspect of a SNP microarray approach is the difficulty in resolving the breakpoints of CNVs, whose assignment is based on estimates driven from array intensities that are dependent on SNPs density [18]. To address these limitations, we opted for a strategy that integrates the SNP microarray information with deep sequencing validation of the most promising results [15,19,20].

In this study, we present the genome-wide distribution of Copy Number Variations (CNVs) and Copy-Number-Variable Regions (CNV-Regions) using SNP microarrays in the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort for which we have previously described genome-wide association analyses of individual SNPs [21,22]. We looked for duplications and deletions using intensity data from SNP microarrays in the ADNI cohort composed by 640 subjects, namely 181 healthy controls, 313 mild cognitive impairment (MCI) and 146 Alzheimer's disease (AD) patients. The affected cases (e.g. MCI and AD) include individuals with mild to moderate cognitive impairment. We investigated the association of CNVs and CNV-Regions with cognitive impairment as expressed in MCI and late-onset AD patients. We created CNV-Region profiles to investigate the potential consequences of these structural variations on gene function using a pathway-based framework. Finally, we confirmed our CNV results comparing individual calls with a different algorithm and validated our main findings by re-sequencing our CNV-Regions in selected patients and controls.

2. Results

2.1. Quality control

After removing subjects that did not pass the quality score threshold for the raw intensity data (QS > 0.2), the ADNI sample contained 640 subjects subdivided in 146 AD patients, 313 MCI patients and 181 matched healthy controls. The Nexus CNV calling algorithm identified 11,058 CNVs from the raw data across the whole sample. In 85 subjects we identified and excluded from the analysis 373 "overlapping" CNVs that can be grouped in 3 categories: 1) 74 segments called as a contiguous deletion and duplication (33 subjects), 2) 89 segments in which heterozygous and homozygous deletions could not be distinguished

(40 cases), 3) 1 case where a heterozygous duplication and a homozygous duplication could not be distinguished. There were 208 segments with contiguous heterozygous deletions (92 cases) that were analyzed as 92 deletions after merging two or more segments for each of the 92 subjects. This latter phenomenon seems to be due to slight differences in the intensities data that are detected by the Nexus CNV-calling algorithm as different events. We performed this merge procedure since the "call fragmentation" issue is known to possibly affect the assessment of large CNVs [16]. In total, 116 segments of the group of 208 were removed from the final list of CNVs along with the CNVs listed in categories 1 (74 segments), 2 (89 segments), and 3 (2 segments), leaving 10,777 CNVs.

2.2. Genome-wide map of CNVs detected in the ADNI sample

Our first aim was to evaluate the frequencies of CNVs detected in the ADNI population defined by deletions, duplications, size and diagnosis as reported in Table 2. Of the 10,777 CNVs, we identified 8443 heterozygous deletions, 806 homozygous deletions, and 1528 duplications, including 5 multi-allelic variants (i.e. high copy gain). The median size of CNVs was 230 kb, ranging from 2.5 kb to 72 Mb. Homozygous deletions tended to be smaller than heterozygous deletions and were similar in size and number in patients and controls, as were duplications.

The number of CNVs per individual did not differ significantly between cases and healthy controls. We observed a slightly higher number of the total copy number events in MCI subjects (mean #-of-calls = 18.1) compared to either AD patients (mean #-of-calls = 15.5) or healthy controls (mean #-of-calls = 15.6) (see Table 2 for details).

Although the differences in the total number of CNVs was not significant, we observed a greater number of copy number events with size >450 kb in AD and MCI, as well a larger mean size of these events: AD (mean size = 590 kb), MCI (mean size = 908 kb) versus healthy controls (mean size = 526 kb). When analyzed by CNV type, heterozygous deletions were significantly larger in MCI and AD patients than in healthy controls (χ^2 = 136.92, p-value < 0.0001). Fig. 1 compares the genomic distributions of heterozygous deletions larger than 450 kb in cases and healthy controls in an adapted version of the typical Manhattan plot, where instead of the p-values per SNP by chromosome we used the CNV size per subject by chromosome.

A Q–Q (quantile–quantile) plot revealed departures from the expected distribution of CNV size values (Fig. 2). The distribution of CNV size differed among the three diagnostic groups. In particular, the MCI CNV size significantly deviates from the reference line starting from a value of approximately 450 kb. The distribution of CNV size of the AD patients deviates from the normality starting at a value of

approximately 4 Mb. The departure from the reference line increases as the CNV size increases for both MCI and AD cases but not healthy controls, with a greater departure for MCI. The deviations in the Q–Q plot are primarily driven by the distribution of heterozygous deletion events (Fig. 3).

2.3. Identification of CNV-Regions with rare large deletions

Comparing the CNVs detected in our sample with those previously reported in reference control populations, like the Toronto's Database of Genomic Variants (DGV), requires consideration of the size of the CNV. For example, very large deletions are likely to encompass regions that harbor smaller copy number events reported in control reference populations and the biological consequences of larger deletions can be different and more functionally relevant. Consequently, we carefully evaluated the size and location of CNVs in order to distinguish potentially pathogenic from benign CNVs, as discussed by Conrad et al. [23]. CNVs are common among healthy individuals, and the strategy to identify potentially pathogenic candidates is key. Thus, following Conrad et al. [23], we consider CNVs as potentially pathogenic if the overlapping segments are at least 30% larger than those reported in the reference controls (Fig. 4) and are shared by four or more patients (at least 1% of our sample). Fig. 4 shows an example of the relationship between CNVs described in the DGV reference population and the CNV-Regions observed in our sample on chromosome 2p16.3-p16.2. We used the intensity data from the SNP microarray to estimate the breakpoint of each CNV (Conrad et al. [23]). Applying these criteria we identified 44 independent heterozygous deleted CNV-Regions (see Table S1).

To provide a first level of confirmation of these 44 CNV-Regions, we used an alternative algorithm for CNV identification from single nucleotide polymorphism arrays, i.e. PennCNV [24]. Although Nexus and PennCNV provide very different CNV calls both in terms of overall number and size [18], we opted for PennCNV since it is currently the most widely used algorithm. As expected, we found that 43% of deletions and 55% of duplications identified by Nexus were also called by PennCNV. Then, we focused specifically on CNV calls by PennCNV overlapping with the previously defined 44 CNV-Regions. We found that 42 over 44 CNV-Regions harbored CNVs detected by PennCNV, 29 of those with deletions larger than 450 kb (Table S2). As expected, PennCNV calls had a smaller size, but the overall coverage of deletions within each CNV-Region confirms the pattern of disruption potentially impacting the same genes and regulatory elements. Despite the differences between the algorithms, these data confirm the presence of regions of putative altered functional activity within the boundaries of the identified CNV-Regions.

2.4. CNV-Regions association analysis

To better investigate the increased number of heterozygous deletions evidenced by the Q–Q plot we evaluated the distribution of the 44 CNV–Regions with a logistic regression approach, since these regions may have more explanatory power than CNVs alone. Overall, the number of affected subjects with more than one CNV-Region deletion is significantly greater in cases than controls (χ^2 = 14.79, p-value = 0.005), with the highest contribution due to a larger proportion of MCI subjects having more than 5 CNV-Regions than healthy controls (Table 3). To further unravel the role of each single CNV-Region in determining the strength of the overall association, we tested each CNV-Region for association with cognitive impairment. We observed that seven of the forty-four CNV-Regions have nominal p-value < 0.05: CNV-Region 7, CNV-Region 14, CNV-Region 23, CNV-Region 28, CNV-Region 38, CNV-Region 48 and CNV-Region 70. These large deletions were present only in cases (MCI and AD patients) with the exception of one healthy control (CNV-Regions 28 and 70).

Table 1
Clinical and demographic characteristics of healthy controls, MCI and mild AD subjects at baseline assessment.

	Control	MCI	AD
# subjects	229	398	193
Mean Age	75.93 ± 5.0	74.83 ± 7.48	75.33 ± 7.47
Gender (male/female)	119/110	257/141	101/91
Smoker/non-smoker	84/145	163/235	75/118
Handedness (right/left)	211/18	362/36	181/12
Ethnicity (Hispanic/non-Hispanic/unknown)	2/224/3	14/380/4	4/187/2
Race (American Indian or Alaskan Native/Asian/African American/White/more than one race)	0/3/16/210/0	1/9/15/372/1	0/2/8/181/2
Mean years of education	16.04 ± 2.86	15.67 ± 3.04	14.70 ± 3.1
MMSE	29.11 ± 0.99	27.03 ± 1.77	23.32 ± 2.06
CDR global (0–0.5–1) # subjects in each category and mean values	229–0–0 0.0	2–396–0 0.49 ± 0.035	0–100–93 0.74 ± 0.25
ADAS-cog	6.20 ± 2.91	11.49 ± 4.42	18.36 ± 6.67
APOE (ε2/ε3/ε4)	38/354/66	28/509/259	9/214/163

Table 2
Characteristics of CNVs detected using microarray data in the ADNI sample.

| CNV size ranges		Homozygous deletions				Heterozygous deletions				Duplications				High copy gain			
		CTRL	MCI	AD	Total	CTRL	MCI	AD	Total	CTRL	MCI	AD	Total	CTRL	MCI	AD	Total
0–20 kb	Number of CNV detected	41	96	33	170	46	72	33	151	2	18	6	26	1	1	–	2
	Mean CNV size kb (SD)	11,382 (4807)	10,522 (4206)	11,130 (3797)	10,847 (4277)	14,560 (3116)	14,249 (3220)	14,235 (3534)	14,341 (3241)	15,439 (2099)	15,496 (2880)	16,244 (3004)	15,664 (2780)	16,054 (0)	16,054 (0)	–	16,054
20–100 kb	Number of CNV detected	120	170	77	367	468	777	369	1614	91	199	84	374	0	1	–	1
	Mean CNV size kb (SD)	45,912 (20,436)	47,134 (21,647)	44,442 (20,671)	46,169 (21,023)	62,431 (22,446)	63,201 (21,922)	62,080 (21,978)	62,721 (22,079)	58,457 (18,978)	58,685 (22,608)	59,282 (22,812)	58,764 (21,774)	–	46,259 (0)	–	46,259
100 kb–450 kb	Number of CNV detected	77	127	64	268	1092	2084	915	4091	212	380	168	760	2	–	–	2
	Mean CNV size kb (SD)	153,417 (47,905)	153,644 (40,175)	156,993 (40,218)	154,379 (42,412)	255,886 (116,761)	263,490 (117,022)	262,689 (116,642)	261,281 (116,884)	242,418 (106,611)	248,233 (103,951)	238,330 (102,066)	244,422 (104,231)	148,958 (32,608)		–	148,958 (32,608)
450 kb–2 Mb	Number of CNV detected	0	1	0	1	473	1001	432	1906	99	162	89	350	–	–	–	–
	Mean CNV size kb (SD)	–	527,566 (0)	–	527,566 (0)	873,373 (353,832)	913,183 (386,963)	875,095 (356,450)	894,671 (372,491)	857,426 (306,587)	881,915 (344,363)	866,698 (316,327)	871,119 (326,257)	–	–	–	–
2 Mb–10 Mb	Number of CNV detected	–	–	–	–	100	445	72	617	1	9	3	13	–	–	–	–
	Mean CNV size kb (SD)	–	–	–	–	3,487,306 (1,616,961)	4,223,484 (1,933,931)	3,641,255 (1,812,325)	4,036,226 (1,893,946)	2,268,253 (0)	2,673,911 (539,851)	2,948,826 (642,019)	2,706,148 (542,653)	–	–	–	–
>10 Mb	Number of CNV detected	–	–	–	–	6	46	12	64	–	–	–	–	–	–	–	–
	Mean CNV size kb (SD)	–	–	–	–	13,032,082 (3,996,942)	13,562,461 (9,103,163)	14,729,801 (5,023,062)	13,731,614 (8,069,805)	–	–	–	–	–	–	–	–
All	Number of CNV detected	238	394	174	806	2185	4425	1833	8443	405	768	350	1523	3	2		5
	Mean CNV size kb (SD)	74,744 (63,843)	73,764 (67,772)	79,522 (66,631)	75,297 (66,344)	526,015 (1,065,102)	907,721 (2,088,887)	589,584 (1,458,033)	739,869 (1,752,991)	355,300 (356,042)	355,757 (431,596)	374,569 (431,851)	359,958 (412,733)	104,656 (80,121)	31,156 (21,358)		75,256 (70,317)
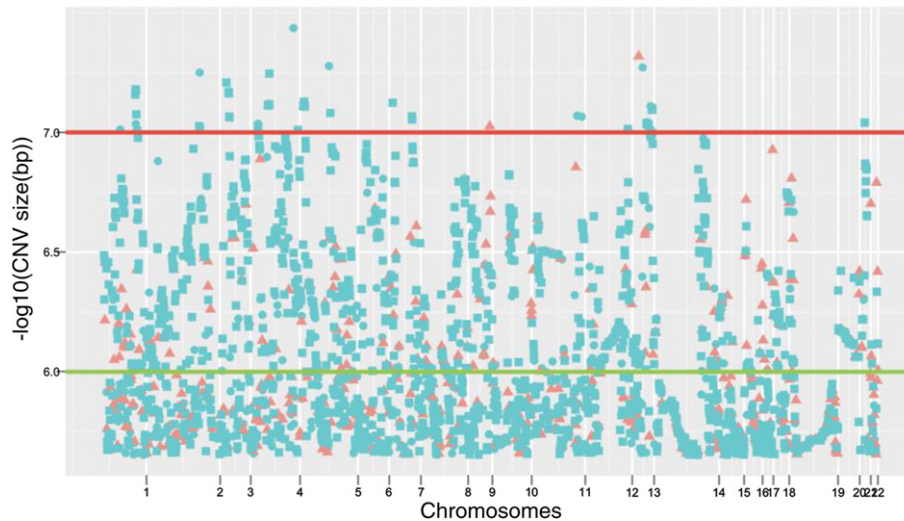
Fig. 1. Manhattan plot of the CNV size distribution for heterozygous deletions >450 kb in the ADNI sample. Each dot represents the CNV size in logarithmic scale of each deletion for each chromosome. The distance between ordered CNVs is relative and does not accurately depict the physical location of the deletions on the genome. Blue circles and squares represent CNVs belonging to MCI and AD subjects respectively. Red triangles represent CNVs belonging to healthy control subjects. The horizontal lines display two different size levels: green line for size 1 Mb and red line for size 10 Mb.

2.5. CNVs validation with next generation sequencing

To provide an initial validation of our findings, we used low-coverage whole genome re-sequencing to confirm the detection of CNVs in the seven CNV-Regions associated with cognitive impairment. We selected six subjects: two patients that presented CNVs in almost all the seven CNV-Regions, two healthy controls who did not report CNV calls (negative controls), and two additional patients who presented two and three CNVs, both called by Nexus and PennCNV. As reported in the 1000 genome study [25], low-coverage whole genome re-sequencing proved to be valid for CNV identification and provided higher resolution compared with microarrays. With sequencing, we confirmed the CNV-Regions, as well as corroborating the reliability of the longer Nexus algorithm. On average, 87.3% of the sequencing reads passed quality filters, and approximately 85% align to the reference sequence (see Supplementary Table S3 for further details). Ninety-nine percent of the bases of the aligned reads match the reference with an average error rate of 0.006. The average coverage genome wide was ~1.0×. Alignment metrics are reported in Supplementary Table S3.

Of the two sequenced patients with 7 CNV-Regions with nominally significant p-values, one had CNV calls in 5 of the 7 CNV-Regions that were identified by PennCNV. The second patient had CNV calls in all the seven CNV-Regions, 5 of which were identified by PennCNV. The remaining two CNV-Regions that were not called by PennCNV have nonetheless been validated by NGS (for details on CNV calls boundaries across Nexus and the two validation methods see Table S5). As expected, the two healthy subjects did not present any NGS-based CNV in the seven CNV-Regions.

Of the other two patients selected for having CNVs called by both Nexus and PennCNV, one presented two CNV calls on chromosomes 3 and 14. On chromosome 3, Nexus called a CNV of 146 kb (35,991,950–36,138,541) and PennCNV a CNV of 118 kb (36,027,069–36,145,613). Both segments were cross-validated by the NGS-based CNV of 921 kb (35,394,001–36,316,000) with additional refinement of the boundaries of the actual deletion. On chromosome 14 Nexus called a CNV of 711 kb (21,308,832–22,020,471) and PennCNV called a CNV of 70 kb (21,701,518–21,771,960). The NGS-based CNV size of 855 kb confirms the deletion at the locus and is in accord with Nexus overall size of the deletion. On chromosomes 5, 6, and 22: Nexus called three CNVs with sizes of 206 kb, 72 kb, and 199 kb, respectively. The CNVs called by PennCNV had sizes of 108 kb, 60 kb, and 2 kb, respectively.

The sequencing cross-validated CNVs on chromosomes 5, and 6 with sizes of 72 kb and 68 kb, but not on chromosome 22.

2.6. CNV-Region genes: functional classification (DAVID)

We analyzed the 231 genes that map within the boundaries of our CNV-Regions with DAVID, to identify related biological processes and molecular functions enriched in our dataset. The gene functional classification analysis identified 10 clusters of genes that share annotation terms for the same biological function. Among them only four have enrichments score > 1.3, a value equivalent to a p-value = 0.05, generally considered an adequate significance threshold for annotation enrichment analyses. Table 4 presents the results for the top four gene functional clusters along with the top functional annotation terms reported for each cluster and their enrichment fold value relative to the entire human genome (DAVID Background: Homo Sapiens). The functional analysis shows enrichment in our data for four different gene families: genes that encode for trans-membrane proteins, immunoglobulin-like domains found in several diverse protein families, the semaphorin protein family and leucine-repeat-rich (LRR) proteins.
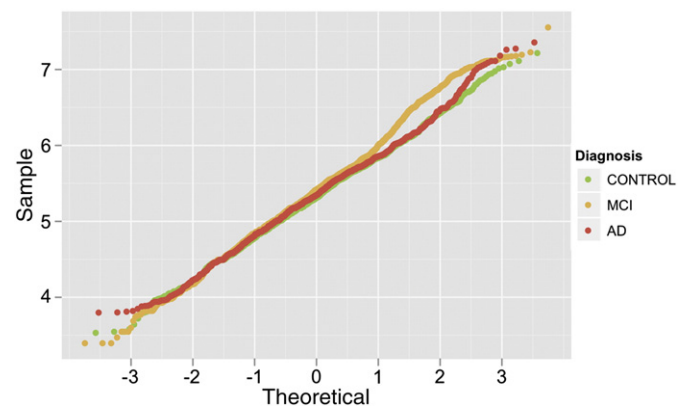

Fig. 2. Q–Q plot of the size distribution of CNVs detected in the ADNI sample by diagnostic group (healthy controls in green, MCI in orange and AD in red). The size of the CNVs are in logarithmic scale, for example the value = 4 in the table corresponds to CNVs with size = 10,000 bp, or the value = 7 to CNVs size equal to 10,000,000 bp.
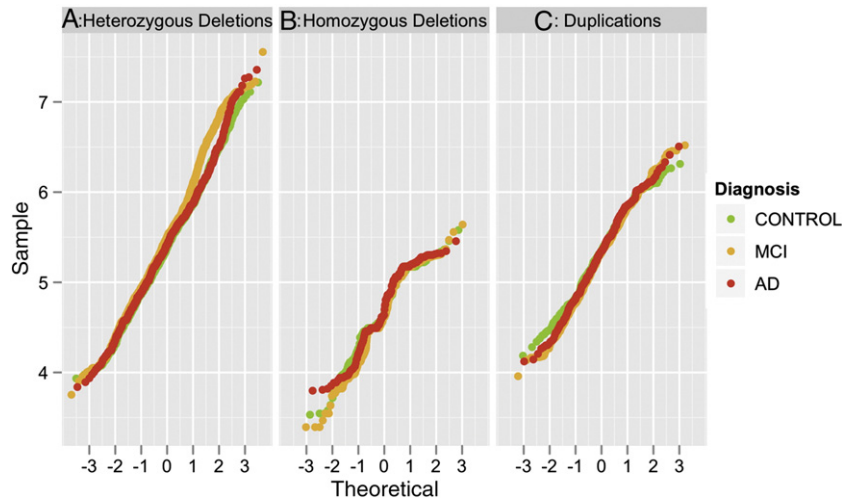
Fig. 3. Q–Q plot of the size distribution of CNVs detected in the ADNI sample represented by CNV type (panel A, heterozygous deletions; panel B, homozygous deletions; panel C, duplications) and diagnostic group (healthy controls in green, MCI in orange and AD in red). The size of the CNVs are in logarithmic scale, for example the value = 4 in the table corresponds to CNVs with size = 10,000 bp, or the value = 7 to CNVs size equal to 10,000,000 bp.

We also performed a functional cluster analysis annotation, which allows including even those genes that might not exert identical biological function but that are likely to co-function in the context of the same biological network. This analysis identified 7 of the 71 total clusters above the significant threshold of ES = 1.3 (Supplementary Table S6). The overall biological function of the two top clusters, Clusters 1 and 2 (ES = 2.03 and 1.71 respectively) is associated with genes implicated in glycosylation, the posttranslational modification process of integral membrane proteins like glycoproteins. Cluster 3 relates to axon guidance, neurogenesis and differentiation and includes the genes coding for the semaphorin protein family (e.g. *SEMA3A*, *-3C*, *-3D*, *-3E*) and the axon guidance receptors (e.g. *ROBO1* and *ROBO2*) as well as transcription factor involved in the differentiation of retinal ganglion cells (e.g. *ATOH7*). Interestingly, the two most significant annotation terms (p-value < 0.01 corrected for multiple testing) both belong to Cluster 3: the UniProt classification of the superfamily of semaphorin proteins, obtained from the Protein Information Resource SuperFamily (PIRSF) site, and the axon guidance pathway as described by KEGG [26]. Cluster 4 is enriched for annotation terms related to proteins containing immunoglobulin-like domains and includes genes coding for neural cell adhesion molecules (e.g., *NCAM2* and *JAM2*) as well as genes that regulate cell surface interactions during nervous system development (e.g. *CNTN5*). Cluster 5 relates to biological processes implicated in the modulation of the assembly that allows the fusion of transport vesicles and the plasma membrane in the cytoplasm (e.g. *STXBP5*) and in synaptic vesicle trafficking to cytoplasmic vesicles (e.g. *PLCO*). Cluster 6 reveals the enrichment of annotation terms related to neuron development, differentiation and more specifically neuron projection morphogenesis and axonogenesis. Besides the above mentioned semaphorin protein family and axon guidance receptors, the biological functions enriched in this cluster are associated with genes such as *SLITRK-1*, *-5* and *-6*, responsible for enhancing neuronal dendrite outgrowth, and *FOXG1*, a transcription repression factor which plays an important role in brain development. Finally, cluster 7 relates to cell–cell adhesion and includes genes coding for cadherins, glycoproteins involved in Ca^{2+}-mediated cell–cell adhesion (e.g. *CDH7*, *CDH19*, and *PCDH7*) as well as for proteins implicated in cellular migration (e.g. *ROBO1* and *ROBO2*).
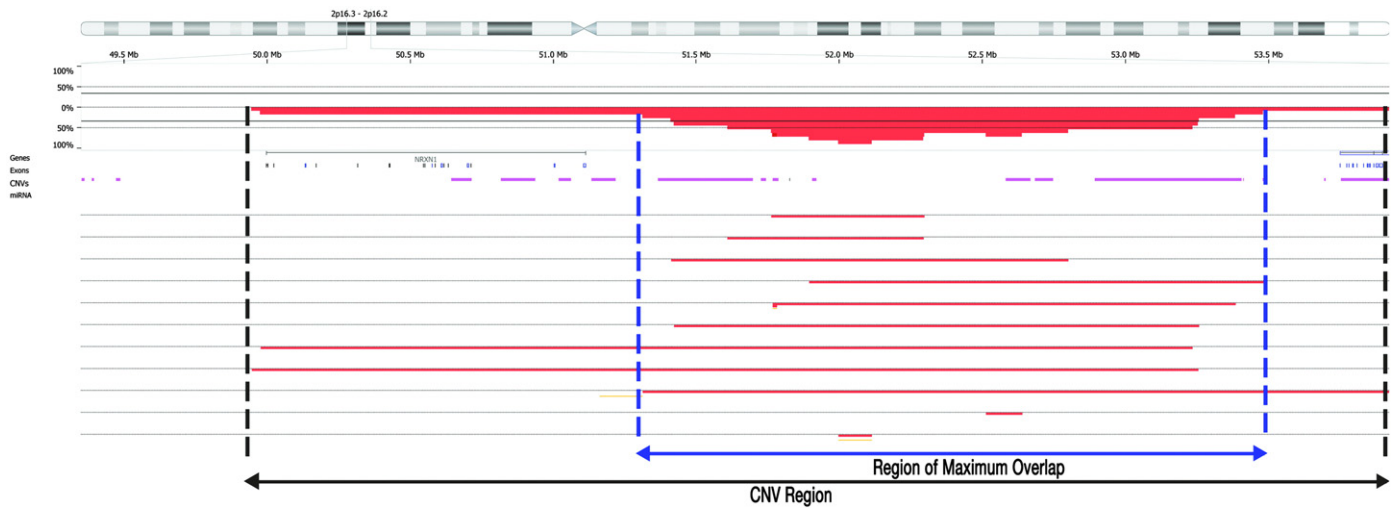

Fig. 4. CNV detection and CNV-Regions (Copy-Number-Variable Regions) boundaries. The physical map of chromosome 2p16.3-p16.2 is depicted. The percentages of duplications and deletions size are indicated by horizontal percentage reference line for the region. The CNVs from the reference population DGV are indicated by horizontal line (purple). Each horizontal line indicates a study subject and the red indicates the extent of the deletion. The vertical blue lines indicate the area of maximum overlap (95%) of deletions for the study subjects. The black vertical lines indicate the Copy-Number-Variable Regions (CNV-Regions), independent CNVs overlapping at the same copy-number variable loci.

Table 3
Co-occurrence of large deletions at more than 5 CNV-Regions.

Diagnostic group		Number of Individuals with CNV-Regions			
		No CNVRs	1–5 CNVRs	>5 CNVRs	Total
Healthy (n = 181)	Observed frequency	162 (89.5%)	18 (9.9%)	1 (0.84%)	181
	Expected frequency	147.9 (81.7%)	29.4 (16.24%)	3.7 (2%)	
MCI (n = 313)	Observed frequency	239 (76.36%)	64 (20.44%)	10 (3.2%)	313
	Expected frequency	255.8 (81.72%)	50.9 (16.26%)	6.4 (2.04%)	
AD (n = 146)	Observed frequency	122 (83.56%)	22 (7.03%)	2 (1.37%)	146
	Expected frequency	119.3 (81.71%)	23.7 (16.23%)	3 (2.05%)	
Total		523	104	13	

Pearson chi^2 (4) = 14.79 p-value = 0.005

3. Discussion

CNV mapping has proven to be highly relevant in increasing our understanding of genetic susceptibility to complex traits. We focused our investigations on CNV-Regions rather than CNVs, mostly because CNV-Regions are more likely to capture the extent of loci disrupted by deletions or duplications. Based on Redon el al. [10] a CNV-Region is defined as the union of juxtaposed or overlapping unique CNVs all of which may impact the same biological function. The assessment of the enrichment of biological signatures of the functional sequences that fall within a CNVR (i.e. genes, regulatory elements, non-coding RNAs) may inform on the functional genomic impact of the CNVs. Recently, several studies have attempted to study the functional impact of CNVs by integrating CNVs and gene expression data by reconstructing the functional CNV-Region networks [5,27,28]. The overall number of subjects with CNV-Regions is low, a finding compatible with CNV-Regions being rare events. We hypothesize that multiple rare CNV-Regions characterize a profile that confer susceptibility to cognitive impairment by acting synergistically in combination with other genetic, epigenetic or environmental factors. We found a statistically significant overrepresentation of subjects with more than one CNV-Region in affected cases compared to healthy controls and we found that very large deletions >450 kb are associated with MCI and/or AD. Recent studies supported the role of the overall load of large deletions in many complex traits [29,30]. Our findings are consistent in particular with the hypothesis that links large copy number events and disease susceptibility as previously described in the context of large CNV studies of schizophrenia, autism and other psychiatric disorders [31,32]

In addition, several studies provided evidence for a substantial role of chromosomal structural variations in the pathogenesis of neurological disorders [3,33–35]. For AD, three papers [7–9] reported the results of an initial CNVs analysis with both genome-wide and candidate gene strategies, evaluating the association of single CNVs to AD and MCI. Heinzen and colleagues [7] reported a duplication in the CHRNA7 gene, although not significantly overrepresented in cases, and the possibility of large heterozygous deletions in cases. Swaminathan and colleagues identified some potential candidate genes enriched in CNVs in cases (CSMD1, SLC35F2, HNRNPCL1 as well as the candidate gene CHRFAM7A), although none met the conventional significance (p-value < .05) after correction for multiple testing. Recently, the same authors replicated these findings in an independent sample from the NIA-LOAD/NCRAD Family Study, with the identification of a new candidate gene (IMMP2L) possibly involved in AD susceptibility [9]. These previous observations of single CNVs enriched in MCI and AD subjects within candidate genes may have failed to reach statistical significance in a traditional case–control study design because of the rarity of the events. Our focus on CNV-Regions shifts the attention from a single event to the region of overlap of events characterized by different sizes, but likely affecting the same underlying functional biology of the deleted or disrupted gene(s). This is consistent with the biological plausibility of the previous findings, despite the lack of statistical significance. Furthermore, the interpretation of the clinical significance of single CNVs, especially for small events <500 kb, is challenging since their pathogenicity is modulated by many factors. We also applied ab initio a stringent filtering procedure to ultimately pull out a set of rare candidate CNVs, excluding all the events quite common throughout the healthy population [36]. Our approach should be seen as a complementary methodology to single CNVs that leverages the power of genome-wide CNV-Region profiling to overcome the limitation of incomplete penetrance and variable expressivity of single CNVs

Our findings of increased number and size of heterozygous deletions associated with late-onset cognitive impairment are consistent with the neurobiology of late-onset diseases. The presence of CNVs in a coding region can alter the abundance of the corresponding transcripts affecting the amount of protein product that may influence cell differentiation [3]. Excessive protein production may lead to age-dependent protein misfolding with implied disruption of protein transport, mitochondrial dysfunction and apoptosis [37]. On the other hand, also CNVs present in the vicinity of genes may influence their expression through a variety of epigenetic mechanisms [5]. An advantage of CNVs and particularly CNV-Regions is that they identify structural changes within DNA that have the potential to affect gene function. To further elucidate the clinical relevance of our CNV-Regions, we analyzed the gene functions or pathways these CNV-Regions might affect using DAVID. The deletions within these CNV-Regions occur in genes implicated in the biological pathways of axonal guidance, neuronal morphogenesis and differentiation, cell–cell adhesion and glycoprotein glycosilation [38].

We established the reliability of our CNV calls using two different algorithms implemented in Nexus and PennCNV, and confirmed our

Table 4
Functional gene cluster classifications using the Database for Annotation, Visualization and Integrated Discovery (DAVID) for enrichment scores (ES) > 1.3.

Gene functional classification			Top 3 associated annotation terms		
Gene cluster	N. of genes	Enrichment score	Annotation category	Annotation term	Enrichment fold
1	37	1.91	UP_SEQ_FEATURE	transmembrane region	3.79
			SP_PIR_KEYWORDS	transmembrane	3.76
			SP_PIR_KEYWORDS	membrane	2.99
2	8	1.78	SP_PIR_KEYWORDS	Immunoglobulin domain	35.81
			INTERPRO	IPR007110:Immunoglobulin-like	29.10
			INTERPRO	IPR003598:Immunoglobulin subtype 2	60.95
3	4	1.70	UP_SEQ_FEATURE	domain:Sema	462.41
			PIR_SUPERFAMILY	PIRSF005526:semaphorin	396.21
			INTERPRO	IPR001627:Semaphorin/CD100 antigen	403.04
4	6	1.56	UP_SEQ_FEATURE	repeat:LRR 10	136.13
			UP_SEQ_FEATURE	repeat:LRR 9	117.98
			UP_SEQ_FEATURE	repeat:LRR 8	106.90

most promising findings by assessing CNV consensus calls and CNV-Region boundaries using a sequencing strategy. A NGS-based CNV detection approach provides the highest sensitivity currently available and allows refining the CNV boundaries, as well as identifying events that cannot be detected by the most sensitive array technologies [39,40]. While deep-coverage (≥25×) whole-genome sequencing costs are significantly dropping, low-coverage sequencing (1–6× base coverage) is still the most feasible option [41]. Low and high sequencing both provide data comparable to CGH-based (comparative genomics hybridization) data [42]. The better resolution of NGS-based CNV detection identifies additional CNVs in the AD and in MCI patients compared to SNP microarray calls, strengthening our results. An increasing number of algorithms that interrogate deep sequencing data for CNV discovery are becoming available, although there is not yet a consensus on a "gold standard" method and analysis strategy. A weak point common to many NGS-based discovery approaches is the requirement for a paired reference sequenced genome since they were originally developed for the detection of cancer CNVs where the paired reference genome for a tumor is the related normal tissue [43]. There is no current consensus yet on the criteria to build a reference genome for CNV calling for complex diseases. Here, we opted to use ERDS, a read-depth NGS-based CNV discovery approach that relies on depth-of-coverage (DOC) (i.e. the density of reads mapping to the region) and detects changes in copy number by comparing the observed DOC within a sliding window of the genome to a reference genome [44–46]. This is currently the only sequencing-based CNV discovery method that allows for the accurate prediction of absolute copy numbers [44,47]. The greatest advantage of ERDS, and DOC algorithms in general, is the ability to detect broader range of CNV events, with best reliability for large events and better breakpoint resolution [48].

The overall high rate of consensus calls across approaches, both cross-algorithm reliabilities using PennCNV and NGS-based confirmation, supports the accuracy of Nexus CNV detection algorithm [49,50]. Recently, Dellinger et al. reported that Nexus may be affected by an overall of CNVs, especially with more relaxed analysis parameters [18]. Our comparative analysis however supports Nexus reliability, provided that the analysis is set up with conservative parameters to find a good trade-off between sensitivity and specificity in CNVs detection.

Our findings of large deletions suggest a link between chromosomal structural alterations and the development of MCI and AD. The CNV-Region strategy captures much more realistic information than the simple description of a catalog of single CNV events, enhancing the genomic relevance of CNVs into a more clinical translational perspective [38]. The higher prevalence in AD and MCI subjects of large deletions in genes involved in neurodevelopment and brain functions makes them good candidates for the definition of predictive profiles for the disease evolution as possible indicators of progressive brain dysfunction. The putative clinical significance of the large deletions reported here is based on several factors including the lack of complete overlapping with benign losses spanning these genomic regions (as reported in DGV), gene function and in most cases tissue specific expression. Although these structural alterations warrant future molecular investigations to fully understand their functional role, it is possible that disruption of the gene regulatory networks is the final common pathway [51]. We are aware that CNV analysis of DNA from blood has limitations, including the difficulty of directly studying the consequence of the identified CNV-Regions on the neuronal activity, particularly in the heterozygous state. There is considerable evidence about the presence of CNVs in normal brains [52,53], although the extent of these structural variations is unknown as is their role in brain functioning. Future research plans include the analysis of post-mortem brain from healthy, MCI and AD subjects to verify the neuronal presence of the CNV-Regions we detected, and possibly their effect on overall brain gene expression and pathways. It is very likely that the dysfunction of specific neuronal pathways underlying AD and MCI depends on additional genetic and/or epigenetic mechanisms to manifest a particular phenotype. Studies of characteristics that

are quantitative, change over time, or vary across clinical disorders like cognitive impairment, offer a great opportunity to deepen our understanding of the role of genetic variation on human behavior and diseases. This can perhaps be best accomplished by determining the correspondence between 'dimensional' phenotypic and genomic variation data.

4. Material and methods

4.1. Ethics statement

Study subjects gave written informed consent at the time of enrollment for imaging and genetic sample collection and completed clinical symptom assessments approved by each participating sites' Institutional Review Board (IRB).

4.2. ADNI

Data used in the preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu\ADNI). For up-to-date information see www.adni-info.org.

4.3. Participants

All subjects were part of the ADNI longitudinal multi-site observational study that included AD, mild cognitive impairment (MCI), and elderly individuals with normal cognition. All subjects have been assessed with clinical and cognitive measures at the time of collection, including ADAS-Cog, CDR-SB and MMSE, MRI and PET scans (FDG and 11C PIB) and blood and CNS biomarkers. Brain imaging, biological samples, and clinical assessments are longitudinally collected for a target of 200 healthy controls, 400 MCI, and 200 AD subjects. All AD patients included in the study are sporadic cases of mild AD that met NINCDS/ADRDA criteria for probable AD [54–56], between the ages of 55–90, with an MMSE score of 20–26 inclusive and having an MRI consistent with the diagnosis of AD (Table 1). Further details about the inclusion and exclusion criteria can be found in the ADNI protocols [57]. For CNV whole-genome screening analysis, we downloaded the entire ADNI genotyping dataset publicly available at the following link: http://www.loni.ucla.edu/ADNI/Data/. Following CNV quality control measures (described below), data from a total of 146 AD, 313 MCI subjects and 181 healthy controls were included in the analysis based on the diagnostic information collected at baseline.

4.4. Genotyping

The ADNI sample was genotyped using the Human-610 Quad BeadChip with a total of 620,901 markers, including 21,890 intensity-only probes specifically designed to improve CNV detection and 27,635 additional probes in SNPs desert genomic regions to enrich CNV coverage. DNA collection, genotyping and relative quality control analysis details are provided in Potkin et al. [21]. The intensity data were analyzed with Illumina GenomeStudio and are publicly available on the LONI website (www.loni.ucla.edu/ADNI/Data). For the purpose of this study we confined the analysis to autosomal chromosomes.

4.5. CNV segmentation algorithm

We used Nexus v5 (Biodiscovery Inc., El Segundo, CA, USA) to produce CNV calls. Nexus calling algorithm SNPrank Segmentation is based on the Circular Binary Segmentation model [58]. To detect CNVs and allelic ratio anomalies, it relies on the normalized measure of the total signal intensity for the two alleles of a SNP, defined as the Log R ratio (LRR) and the normalized measure of the allelic intensity ratio of the two alleles, defined as the B allele frequencies (BAF). Both LRR and BAF were computed from the array intensity data with Illumina Genome Studio v1.0.2 software using the manufacturer's cluster file as

a reference [21]. SNPrank Segmentation classifies three types of CNVs: 1) CN gain, corresponding to duplications (CN = 3 or 4 copies); 2) CN Loss, corresponding to single copy deletions (CN = 1 copy); 3) homozygous copy loss corresponding to a complete deletion event (CN = 0). We used standard calling parameters recommended by the manufacturer for Illumina array data: 1) minimum number of probes per segment of 5, 2) max contiguous probe spacing of 1000 kbp, and 3) significance threshold of p-value = 1×10^{-6} for CNV calling. The combination of these measures is particularly indicated for the detection of CNV in regions scarcely covered, while simultaneously accounting for the risk of false positive calls (e.g. centromeric and telomeric CNV, or CNV calls merely due to a background waviness of the LRR that involves few adjacent probes). We also ran PennCNV to evaluate CNV calls that had a minimum of three SNPs on autosomes in addition to standard default quality control parameters for CNV calling.

4.6. Statistical analysis

We used Nexus a specific pre-processing quality control parameter (QCscore) to check for the quality of the individuals' intensity raw data to minimize the risk of false positive CNV calls and we excluded samples exceeding QCscore of 0.2, a threshold value empirically determined for Illumina array data by the manufacturer. We removed overlapping CNVs within the same subject blind to diagnostic group where the overlapping refers to either a situation where more than one deletion or duplication were called for the same subject in the same region, or the same event was specified twice for the same subject using PLINK software package [59] (http://pngu.mgh.harvard.edu/purcell/plink), release v1.07. Proportion of overlap concordance of Nexus CNV calls with the PennCNV calls using PLINK was also calculated to provide validation across different CNV algorithms. We performed descriptive statistics of single CNV distributions by size and type using chi-square and Fisher exact test. To identify copy number variable loci (i.e. CNV-Regions for partially overlapping CNVs), we used the "union overlap" tool of PLINK that uses the ratio of the number of base-pairs intersected between different CNVs and the length of the CNV as a denominator in calculating the proportion overlap. The "union overlap" tool allowed to both 1) select segments that were 95% overlapping between them in a region with boundaries defined by our own data and 2) exclude segments that were overlapping for more than 70% with previously described CNVs. To investigate the effect of CNV-Regions on cognitive impairment, we tested for association with diagnoses and MMSE score using a logistic regression model. All descriptive and association analyses of single CNVs and CNV-Regions were performed with STATA11 (StataCorp Stata Statistical Software: Release11. 200X. College Station, TX: StataCorp LP).

4.7. Bioinformatics and in-silico functional pathway databases

We used the Database for Annotation, Visualization and Integrated Discovery (DAVID), release 6.7 [60] to screen the genes that appeared to be affected by CNV-Region variants. DAVID classifies the genes into functional groups based on annotation similarity criteria. DAVID calculates an enrichment score that, relative to the ADNI dataset, ranks the relevance of the annotation terms that describe the genes included in a functional cluster. To be included in the final list, we required a gene to harbor a copy number variant in at least 3 subjects and to map in the region of maximum overlapping within the boundaries of a CNV-Region. The list of genes affected by CNV-Region variants we submitted for the gene functional analysis had 231 entries. We used the fuzzy clustering algorithm implemented in the "gene functional classification" tool of DAVID to classify functionally related genes into groups based on co-occurrence of "motifs" underlying shared biological modules. Then, we refined our results with the "functional annotation clustering" analysis to cluster the annotation terms associated with the genes in our list. For the analysis of annotation enrichment, we set options to the default "medium" stringency criteria values provided by DAVID.

4.8. Graphics (Nexus and R-ggplot2)

All plots and histograms were created using the plotting system "ggplot2" of R (http://had.co.nz/ggplot2/). All images related to CNVs and CNV-Regions have been created within the graphical framework provided by Nexus v5.

4.9. Whole genome sequencing

After sonication of genomic DNA using a Covaris S2 to an average size of 250–300 bp, libraries were constructed manually using the Wellcome Trust protocol and reagents [61]. Each library was then sequenced to depth of ~4–7× coverage on a Illumina HiSeq2000 DNA sequencer using v5 kits, with pair-end reads of 75–100 bp flow cell. The Illumina pipeline (v1.7–1.8) was then used to convert digital images into base pair calls (with quality scores). The number of raw reads generated per patient is reported in Supplementary Table S3.

Sequencing reads were aligned to the NCBI36 human reference genome (Ensembl hg18 release 50: ftp://ftp.ensembl.org/pub/current/fasta/homo_sapiens/dna/) using BWA software v0.5.9 (http://bio-bwa.sourceforge.net/) [62]. We used the software package SAMtools v0.7.1 (http://samtools.sourceforge.net/) to generate the SAM/BAM format files after screening the alignment data for duplicate reads, sorting and indexing procedures. Summary Alignment metrics were calculated using the CollectAlignmentSummaryMetrics program implemented in the PICARD Java-based command-line software package v2.6.21 (http://picard.sourceforge.net/). The CNV calling was performed with "Estimation by Read Depth with Single Nucleotide Variants" (ERDS) software v1.02 (http://web.duke.edu/~mz34/erds.htm), a Hidden Markov Model (HMM) based approach that relies on depth-of-coverage (DOC) to infer the copy number state. It represents an extensions to the methods described in [63]. The algorithm has been described more in detail in [64]. A deletion was called when its average read depth was below 0.7 ∗ expected read depth (corresponding to copy number <1.4), and a duplication is called when its average read depth is above 1.3 ∗ expected read depth (corresponding to copy number >2.6). Expected read depth is calculated using the expectation maximization (EM) approach and corrected by GC bias. Sequence Variant Analyzer (SVA, http://www.svaproject.org) and Integrative Genome Viewer (IGV, http://www.broadinstitute.org/software/igv/) were used to visually inspect and annotate the CNV-Regions.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ygeno.2013.04.004.

References

[1] S.A. McCarroll, F.G. Kuruvilla, J.M. Korn, S. Cawley, J. Nemesh, A. Wysoker, M.H. Shapero, P.I. de Bakker, J.B. Maller, A. Kirby, A.L. Elliott, M. Parkin, E. Hubbell, T. Webster, R. Mei, J. Veitch, P.J. Collins, R. Handsaker, S. Lincoln, M. Nizzari, J. Blume, K.W. Jones, R. Rava, M.J. Daly, S.B. Gabriel, D. Altshuler, Integrated detection and population-genetic analysis of SNPs and copy number variation, Nat. Genet. 40 (2008) 1166–1174.

[2] T.A. Manolio, F.S. Collins, N.J. Cox, D.B. Goldstein, L.A. Hindorff, D.J. Hunter, M.I. McCarthy, E.M. Ramos, L.R. Cardon, A. Chakravarti, J.H. Cho, A.E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C.N. Rotimi, M. Slatkin, D. Valle, A.S. Whittemore, M. Boehnke, A.G. Clark, E.E. Eichler, G. Gibson, J.L. Haines, T.F. Mackay, S.A. McCarroll, P.M. Visscher, Finding the missing heritability of complex diseases, Nature 461 (2009) 747–753.

[3] B. Kalman, E. Vitale, Structural chromosomal variations in neurological diseases, Neurologist 15 (2009) 245–253.

[4] H.M. Blauw, J.H. Veldink, M.A. van Es, P.W. van Vught, C.G. Saris, B. van der Zwaag, L. Franke, J.P. Burbach, J.H. Wokke, R.A. Ophoff, L.H. van den Berg, Copy-number variation in sporadic amyotrophic lateral sclerosis: a genome-wide screen, Lancet Neurol. 7 (2008) 319–326.

[5] C.N. Henrichsen, E. Chaignat, A. Reymond, Copy number variants, diseases and gene expression, Hum. Mol. Genet. 18 (2009) R1–8.

[6] L. Feuk, A.R. Carson, S.W. Scherer, Structural variation in the human genome, Nat. Rev. Genet. 7 (2006) 85–97.

[7] E.L. Heinzen, A.C. Need, K.M. Hayden, O. Chiba-Falek, A.D. Roses, W.J. Strittmatter, J.R. Burke, C.M. Hulette, K.A. Welsh-Bohmer, D.B. Goldstein, Genome-wide scan of copy number variation in late-onset Alzheimer's disease, J. Alzheimers Dis. 19 (2010) 69–77.

[8] S. Swaminathan, S. Kim, L. Shen, S.L. Risacher, T. Foroud, N. Pankratz, S.G. Potkin, M.J. Huentelman, D.W. Craig, M.W. Weiner, A.J. Saykin, The Alzheimer's disease neuro-imaging initiative, genomic copy number analysis in Alzheimer's disease and mild cognitive impairment: an ADNI study, Int J Alzheimers Dis 2011 (2011) 729478.

[9] S. Swaminathan, L. Shen, S. Kim, M. Inlow, J.D. West, K.M. Faber, T. Foroud, R. Mayeux, A.J. Saykin, Analysis of copy number variation in Alzheimer's disease: the NIALOAD/NCRAD Family Study, Curr. Alzheimer Res. 9 (2012) 801–814.

[10] R. Redon, S. Ishikawa, K.R. Fitch, L. Feuk, G.H. Perry, T.D. Andrews, H. Fiegler, M.H. Shapero, A.R. Carson, W. Chen, E.K. Cho, S. Dallaire, J.L. Freeman, J.R. Gonzalez, M. Gratacos, J. Huang, D. Kalaitzopoulos, D. Komura, J.R. MacDonald, C.R. Marshall, R. Mei, L. Montgomery, K. Nishimura, K. Okamura, F. Shen, M.J. Somerville, J. Tchinda, A. Valsesia, C. Woodwark, F. Yang, J. Zhang, T. Zerjal, L. Armengol, D.F. Conrad, X. Estivill, C. Tyler-Smith, N.P. Carter, H. Aburatani, C. Lee, K.W. Jones, S.W. Scherer, M.E. Hurles, Global variation in copy number in the human genome, Nature 444 (2006) 444–454.

[11] C. Vogler, L. Gschwind, B. Rothlisberger, A. Huber, I. Filges, P. Miny, B. Auschra, A. Stetak, P. Demougin, V. Vukojevic, I.T. Kolassa, T. Elbert, D.J. de Quervain, A. Papassotiropoulos, Microarray-based maps of copy-number variant regions in European and sub-Saharan populations, PLoS One 5 (2010) e15246.

[12] M. Jakobsson, S.W. Scholz, P. Scheet, J.R. Gibbs, J.M. VanLiere, H.C. Fung, Z.A. Szpiech, J.H. Degnan, K. Wang, R. Guerreiro, J.M. Bras, J.C. Schymick, D.G. Hernandez, B.J. Traynor, J. Simon-Sanchez, M. Matarin, A. Britton, J. van de Leemput, I. Rafferty, M. Bucan, H.M. Cann, J.A. Hardy, N.A. Rosenberg, A.B. Singleton, Genotype, haplotype and copy-number variation in worldwide human populations, Nature 451 (2008) 998–1003.

[13] M. Marcinkowska, M. Szymanski, W.J. Krzyzosiak, P. Kozlowski, Copy number variation of microRNA genes in the human genome, BMC Genomics 12 (2011) 183.

[14] V. Guryev, K. Saar, T. Adamovic, M. Verheul, S.A. van Heesch, S. Cook, M. Pravenec, T. Aitman, H. Jacob, J.D. Shull, N. Hubner, E. Cuppen, Distribution and functional impact of DNA copy number variation in the rat, Nat. Genet. 40 (2008) 538–545.

[15] C. Alkan, B.P. Coe, E.E. Eichler, Genome structural variation discovery and genotyping, Nat. Rev. Genet. 12 (2011) 363–376.

[16] D. Pinto, K. Darvishi, X. Shi, D. Rajan, D. Rigler, T. Fitzgerald, A.C. Lionel, B. Thiruvahindrapuram, J.R. Macdonald, R. Mills, A. Prasad, K. Noonan, S. Gribble, E. Prigmore, P.K. Donahoe, R.S. Smith, J.H. Park, M.E. Hurles, N.P. Carter, C. Lee, S.W. Scherer, L. Feuk, Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants, Nat. Biotechnol. 29 (2011) 512–520.

[17] L. Winchester, C. Yau, J. Ragoussis, Comparing CNV detection methods for SNP arrays, Brief. Funct. Genomic. Proteomic. 8 (2009) 353–366.

[18] A.E. Dellinger, S.M. Saw, L.K. Goh, M. Seielstad, T.L. Young, Y.J. Li, Comparative analyses of seven algorithms for copy number variant identification from single nucleotide polymorphism arrays, Nucleic Acids Res. 38 (2010) e105.

[19] J.O. Korbel, A.E. Urban, J.P. Affourtit, B. Godwin, F. Grubert, J.F. Simons, P.M. Kim, D. Palejev, N.J. Carriero, L. Du, B.E. Taillon, Z. Chen, A. Tanzer, A.C. Saunders, J. Chi, F. Yang, N.P. Carter, M.E. Hurles, S.M. Weissman, T.T. Harkins, M.B. Gerstein, M. Egholm, M. Snyder, Paired-end mapping reveals extensive structural variation in the human genome, Science 318 (2007) 420–426.

[20] P. Medvedev, M. Stanciu, M. Brudno, Computational methods for discovering structural variation with next-generation sequencing, Nat. Methods 6 (2009) S13–20.

[21] S.G. Potkin, G. Guffanti, A. Lakatos, J.A. Turner, F. Kruggel, J.H. Fallon, A.J. Saykin, A. Orro, S. Lupoli, E. Salvi, M. Weiner, F. Macciardi, Hippocampal atrophy as a quantitative trait in a genome-wide association study identifying novel susceptibility genes for Alzheimer's disease, PLoS One 4 (2009) e6501.

[22] L. Shen, S. Kim, S.L. Risacher, K. Nho, S. Swaminathan, J.D. West, T. Foroud, N. Pankratz, J.H. Moore, C.D. Sloan, M.J. Huentelman, D.W. Craig, B.M. Dechairo, S.G. Potkin, C.R. Jack Jr., M.W. Weiner, A.J. Saykin, Whole genome association study of brain-wide imaging phenotypes for identifying quantitative trait loci in MCI and AD: A study of the ADNI cohort, Neuroimage 53 (2010) 1051–1063.

[23] D.F. Conrad, D. Pinto, R. Redon, L. Feuk, O. Gokcumen, Y. Zhang, J. Aerts, T.D. Andrews, C. Barnes, P. Campbell, T. Fitzgerald, M. Hu, C.H. Ihm, K. Kristiansson, D.G. Macarthur, J.R. Macdonald, I. Onyiah, A.W. Pang, S. Robson, K. Stirrups, A. Valsesia, K. Walter, J. Wei, C. Tyler-Smith, N.P. Carter, C. Lee, S.W. Scherer, M.E. Hurles, Origins and functional impact of copy number variation in the human genome, Nature 464 (2010) 704–712.

[24] K. Wang, M. Li, D. Hadley, R. Liu, J. Glessner, S.F. Grant, H. Hakonarson, M. Bucan, PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data, Genome Res. 17 (2007) 1665–1674.

[25] J. Wu, K.R. Grzeda, C. Stewart, F. Grubert, A.E. Urban, M.P. Snyder, G.T. Marth, Copy Number Variation detection from 1000 Genomes project exon capture sequencing data, BMC Bioinformatics 13 (2012) 305.

[26] M. Kanehisa, S. Goto, M. Furumichi, M. Tanabe, M. Hirakawa, KEGG for representation and analysis of molecular networks involving diseases and drugs, Nucleic Acids Res. 38 (2010) D355–360.

[27] B.E. Stranger, M.S. Forrest, M. Dunning, C.E. Ingle, C. Beazley, N. Thorne, R. Redon, C.P. Bird, A. de Grassi, C. Lee, C. Tyler-Smith, N. Carter, S.W. Scherer, S. Tavare, P. Deloukas, M.E. Hurles, E.T. Dermitzakis, Relative impact of nucleotide and copy number variation on gene expression phenotypes, Science 315 (2007) 848–853.

[28] C. Park, J. Ahn, Y. Yoon, S. Park, Identification of functional CNV region networks using a CNV-gene mapping algorithm in a genome-wide scale, Bioinformatics 28 (2012) 2045–2051.

[29] R.A. Yeo, S.W. Gangestad, C. Gasparovic, J. Liu, V.D. Calhoun, R.J. Thoma, A.R. Mayer, R. Kalyanam, K.E. Hutchison, Rare copy number deletions predict individual variation in human brain metabolite concentrations in individuals with alcohol use disorders, Biol. Psychiatry 70 (2011) 537–544.

[30] R.A. Yeo, S.W. Gangestad, J. Liu, V.D. Calhoun, K.E. Hutchison, Rare copy number deletions predict individual variation in intelligence, PLoS One 6 (2011) e16339.

[31] ISC, Rare chromosomal deletions and duplications increase risk of schizophrenia, Nature 455 (2008) 237–241.

[32] T. Walsh, J.M. McClellan, S.E. McCarthy, A.M. Addington, S.B. Pierce, G.M. Cooper, A.S. Nord, M. Kusenda, D. Malhotra, A. Bhandari, S.M. Stray, C.F. Rippey, P. Roccanova, V. Makarov, B. Lakshmi, R.L. Findling, L. Sikich, T. Stromberg, B. Merriman, N. Gogtay, P. Butler, K. Eckstrand, L. Noory, P. Gochman, R. Long, Z. Chen, S. Davis, C. Baker, E.E. Eichler, P.S. Meltzer, S.F. Nelson, A.B. Singleton, M.K. Lee, J.L. Rapoport, M.C. King, J. Sebat, Rare structural variants disrupt multiple genes in neurodevelopmental pathways in schizophrenia, Science 320 (2008) 539–543.

[33] S.R. Gilman, J. Chang, B. Xu, T.S. Bawa, J.A. Gogos, M. Karayiorgou, D. Vitkup, Diverse types of genetic variation converge on functional gene networks involved in schizophrenia, Nat. Neurosci. 15 (2012) 1723–1728.

[34] J.R. Lupski, Brain copy number variants and neuropsychiatric traits, Biol. Psychiatry 72 (2012) 617–619.

[35] S. Girirajan, J.A. Rosenfeld, B.P. Coe, S. Parikh, N. Friedman, A. Goldstein, R.A. Filipink, J.S. McConnell, B. Angle, W.S. Meschino, M.M. Nezarati, A. Asamoah, K.E. Jackson, G.C. Gowans, J.A. Martin, E.P. Carmany, D.W. Stockton, R.E. Schnur, L.S. Penney, D.M. Martin, S. Raskin, K. Leppig, H. Thiese, R. Smith, E. Aberg, D.M. Niyazov, L.F. Escobar, D. El-Khechen, K.D. Johnson, R.R. Lebel, K. Siefkas, S. Ball, N. Shur, M. McGuire, C.K. Brasington, J.E. Spence, L.S. Martin, C. Clericuzio, B.C. Ballif, L.G. Shaffer, E.E. Eichler, Phenotypic heterogeneity of genomic disorders and rare copy-number variants, N. Engl. J. Med. 367 (2012) 1321–1331.

[36] J. Sebat, B. Lakshmi, J. Troge, J. Alexander, J. Young, P. Lundin, S. Maner, H. Massa, M. Walker, M. Chi, N. Navin, R. Lucito, J. Healy, J. Hicks, K. Ye, A. Reiner, T.C. Gilliam, B. Trask, N. Patterson, A. Zetterberg, M. Wigler, Large-scale copy number polymorphism in the human genome, Science 305 (2004) 525–528.

[37] J.A. Lee, J.R. Lupski, Genomic rearrangements and gene copy-number alterations as a cause of nervous system disorders, Neuron 52 (2006) 103–121.

[38] E.H. Cook Jr., S.W. Scherer, Copy-number variations associated with neuropsychiatric conditions, Nature 455 (2008) 919–923.

[39] R.E. Mills, K. Walter, C. Stewart, R.E. Handsaker, K. Chen, C. Alkan, A. Abyzov, S.C. Yoon, K. Ye, R.K. Cheetham, A. Chinwalla, D.F. Conrad, Y. Fu, F. Grubert, I. Hajirasouliha, F. Hormozdiari, L.M. Iakoucheva, Z. Iqbal, S. Kang, J.M. Kidd, M.K. Konkel, J. Korn, E. Khurana, D. Kural, H.Y. Lam, J. Leng, R. Li, Y. Li, C.Y. Lin, R. Luo, X.J. Mu, J. Nemesh, H.E. Peckham, T. Rausch, A. Scally, X. Shi, M.P. Stromberg, A.M. Stutz, A.E. Urban, J.A. Walker, J. Wu, Y. Zhang, Z.D. Zhang, M.A. Batzer, L. Ding, G.T. Marth, G. McVean, J. Sebat, M. Snyder, J. Wang, E.E. Eichler, M.B. Gerstein, M.E. Hurles, C. Lee, S.A. McCarroll, J.O. Korbel, Mapping copy number variation by population-scale genome sequencing, Nature 470 (2011) 59–65.

[40] H. Park, J.I. Kim, Y.S. Ju, O. Gokcumen, R.E. Mills, S. Kim, S. Lee, D. Suh, D. Hong, H.P. Kang, Y.J. Yoo, J.Y. Shin, H.J. Kim, M. Yavartanoo, Y.W. Chang, J.S. Ha, W. Chong, G.R. Hwang, K. Darvishi, H. Kim, S.J. Yang, K.S. Yang, H. Kim, M.E. Hurles, S.W. Scherer, N.P. Carter, C. Tyler-Smith, C. Lee, J.S. Seo, Discovery of common Asian copy number variants using integrated high-resolution array CGH and massively parallel DNA sequencing, Nat. Genet. 42 (2010) 400–405.

[41] H.M. Wood, O. Belvedere, C. Conway, C. Daly, R. Chalkley, M. Bickerdike, C. McKinley, P. Egan, L. Ross, B. Hayward, J. Morgan, L. Davidson, K. MacLennan, T.K. Ong, K. Papagiannopoulos, I. Cook, D.J. Adams, G.R. Taylor, P. Rabbitts, Using next-generation sequencing for high resolution multiplex analysis of copy number variation from nanogram quantities of DNA from formalin-fixed paraffin-embedded specimens, Nucleic Acids Res. 38 (2010) e151.

[42] A. Abyzov, A.E. Urban, M. Snyder, M. Gerstein, CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing, Genome Res. 21 (2011) 974–984.

[43] P.J. Campbell, P.J. Stephens, E.D. Pleasance, S. O'Meara, H. Li, T. Santarius, L.A. Stebbings, C. Leroy, S. Edkins, C. Hardy, J.W. Teague, A. Menzies, I. Goodhead, D.J. Turner, C.M. Clee, M.A. Quail, A. Cox, C. Brown, R. Durbin, M.E. Hurles, P.A. Edwards, G.R. Bignell, M.R. Stratton, P.A. Futreal, Identification of somatically acquired rearrangements in cancer using genome-wide massively parallel paired-end sequencing, Nat. Genet. 40 (2008) 722–729.

[44] C. Alkan, J.M. Kidd, T. Marques-Bonet, G. Aksay, F. Antonacci, F. Hormozdiari, J.O. Kitzman, C. Baker, M. Malig, O. Mutlu, S.C. Sahinalp, R.A. Gibbs, E.E. Eichler, Personalized copy number and segmental duplication maps using next-generation sequencing, Nat. Genet. 41 (2009) 1061–1067.

[45] D.Y. Chiang, G. Getz, D.B. Jaffe, M.J. O'Kelly, X. Zhao, S.L. Carter, C. Russ, C. Nusbaum, M. Meyerson, E.S. Lander, High-resolution mapping of copy-number alterations with massively parallel sequencing, Nat. Methods 6 (2009) 99–103.

[46] S. Yoon, Z. Xuan, V. Makarov, K. Ye, J. Sebat, Sensitive and accurate detection of copy number variants using read depth of coverage, Genome Res. 19 (2009) 1586–1592.

[47] P.H. Sudmant, J.O. Kitzman, F. Antonacci, C. Alkan, M. Malig, A. Tsalenko, N. Sampas, L. Bruhn, J. Shendure, E.E. Eichler, Diversity of human copy number variation and multicopy genes, Science 330 (2010) 641–646.

[48] O. Harismendy, P.C. Ng, R.L. Strausberg, X. Wang, T.B. Stockwell, K.Y. Beeson, N.J. Schork, S.S. Murray, E.J. Topol, S. Levy, K.A. Frazer, Evaluation of next generation sequencing platforms for population targeted sequencing studies, Genome Biol. 10 (2009) R32.

[49] K. Darvishi, Application of Nexus copy number software for CNV detection and analysis, Curr. Protoc. Hum. Genet. (2010), Chapter 4, Unit 4 14 11–28.

[50] H. Matsuzaki, P.H. Wang, J. Hu, R. Rava, G.K. Fu, High resolution discovery and confirmation of copy number variants in 90 Yoruba Nigerians, Genome Biol. 10 (2009) R125.

[51] S.G. Potkin, F. Macciardi, G. Guffanti, J.H. Fallon, Q. Wang, J.A. Turner, A. Lakatos, M.F. Miles, A. Lander, M.P. Vawter, X. Xie, Identifying gene regulatory networks in schizophrenia, Neuroimage 53 (2010) 839–847.

[52] J.W. Westra, R.R. Rivera, D.M. Bushman, Y.C. Yung, S.E. Peterson, S. Barral, J. Chun, Neuronal DNA content variation (DCV) with regional and individual differences in the human brain, J. Comp. Neurol. 518 (2010) 3981–4000.

[53] S.K. Rehen, Y.C. Yung, M.P. McCreight, D. Kaushal, A.H. Yang, B.S. Almeida, M.A. Kingsbury, K.M. Cabral, M.J. McConnell, B. Anliker, M. Fontanoz, J. Chun, Constitutional aneuploidy in the normal human brain, J. Neurosci. 25 (2005) 2176–2180.

[54] B. Dubois, H.H. Feldman, C. Jacova, S.T. Dekosky, P. Barberger-Gateau, J. Cummings, A. Delacourte, D. Galasko, S. Gauthier, G. Jicha, K. Meguro, J. O'Brien, F. Pasquier, P. Robert, M. Rossor, S. Salloway, Y. Stern, P.J. Visser, P. Scheltens, Research criteria for the diagnosis of Alzheimer's disease: revising the NINCDS-ADRDA criteria, Lancet Neurol. 6 (2007) 734–746.

[55] G. McKhann, D. Drachman, M. Folstein, R. Katzman, D. Price, E.M. Stadlan, Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease, Neurology 34 (1984) 939–944.

[56] M. Storandt, E.A. Grant, J.P. Miller, J.C. Morris, Rates of progression in mild cognitive impairment and early Alzheimer's disease, Neurology 59 (2002) 1034–1041.

[57] S.G. Mueller, M.W. Weiner, L.J. Thal, R.C. Petersen, C.R. Jack, W. Jagust, J.Q. Trojanowski, A.W. Toga, L. Beckett, Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI), Alzheimers Dement. 1 (2005) 55–66.

[58] A.B. Olshen, E.S. Venkatraman, R. Lucito, M. Wigler, Circular binary segmentation for the analysis of array-based DNA copy number data, Biostatistics 5 (2004) 557–572.

[59] S. Purcell, B. Neale, K. Todd-Brown, L. Thomas, M.A. Ferreira, D. Bender, J. Maller, P. Sklar, P.I. de Bakker, M.J. Daly, P.C. Sham, PLINK: a tool set for whole genome association and population-based linkage analyses, Am. J. Hum. Genet. 81 (2007) 559–575.

[60] D.W. Huang, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources, Nat. Protoc. 4 (2009) 44–57.

[61] M.A. Quail, H. Swerdlow, D.J. Turner, Improved protocols for the illumina genome analyzer sequencing system, Curr. Protoc. Hum. Genet. (2009), Chapter 18 (2009) Unit 18 12.

[62] H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform, Bioinformatics 25 (2009) 1754–1760.

[63] D.R. Bentley, S. Balasubramanian, H.P. Swerdlow, G.P. Smith, J. Milton, C.G. Brown, K.P. Hall, D.J. Evers, C.L. Barnes, H.R. Bignell, J.M. Boutell, J. Bryant, R.J. Carter, R. Keira Cheetham, A.J. Cox, D.J. Ellis, M.R. Flatbush, N.A. Gormley, S.J. Humphray, L.J. Irving, M.S. Karbelashvili, S.M. Kirk, H. Li, X. Liu, K.S. Maisinger, L.J. Murray, B. Obradovic, T. Ost, M.L. Parkinson, M.R. Pratt, I.M. Rasolonjatovo, M.T. Reed, R. Rigatti, C. Rodighiero, M.T. Ross, A. Sabot, S.V. Sankar, A. Scally, G.P. Schroth, M.E. Smith, V.P. Smith, A. Spiridou, P.E. Torrance, S.S. Tzonev, E.H. Vermaas, K. Walter, X. Wu, L. Zhang, M.D. Alam, C. Anastasi, I.C. Aniebo, D.M. Bailey, I.R. Bancarz, S. Banerjee, S.G. Barbour, P.A. Baybayan, V.A. Benoit, K.F. Benson, C. Bevis, P.J. Black, A. Boodhun, J.S. Brennan, J.A. Bridgham, R.C. Brown, A.A. Brown, D.H. Buermann, A.A. Bundu, J.C. Burrows, N.P. Carter, N. Castillo, E.C.M. Chiara, S. Chang, R. Neil Cooley, N.R. Crake, O.O. Dada, K.D. Diakoumakos, B. Dominguez-Fernandez, D.J. Earnshaw, U.C. Egbujor, D.W. Elmore, S.S. Etchin, M.R. Ewan, M. Fedurco, L.J. Fraser, K.V. Fuentes Fajardo, W. Scott Furey, D. George, K.J. Gietzen, C.P. Goddard, G.S. Golda, P.A. Granieri, D.E. Green, D.L. Gustafson, N.F. Hansen, K. Harnish, C.D. Haudenschild, N.I. Heyer, M.M. Hims, J.T. Ho, A.M. Horgan, K. Hoschler, S. Hurwitz, D.V. Ivanov, M.Q. Johnson, T. James, T.A. Huw Jones, G.D. Kang, T.H. Kerelska, A.D. Kersey, I. Khrebtukova, A.P. Kindwall, Z. Kingsbury, P.I. Kokko-Gonzales, A. Kumar, M.A. Laurent, C.T. Lawley, S.E. Lee, X. Lee, A.K. Liao, J.A. Loch, M. Lok, S. Luo, R.M. Mammen, J.W. Martin, P.G. McCauley, P. McNitt, P. Mehta, K.W. Moon, J.W. Mullens, T. Newington, Z. Ning, B. Ling Ng, S.M. Novo, M.J. O'Neill, M.A. Osborne, A. Osnowski, O. Ostadan, L.L. Paraschos, L. Pickering, A.C. Pike, D. Chris Pinkard, D.P. Pliskin, J. Podhasky, V.J. Quijano, C. Raczy, V.H. Rae, S.R. Rawlings, A. Chiva Rodriguez, P.M. Roe, J. Rogers, M.C. Rogert Bacigalupo, N. Romanov, A. Romieu, R.K. Roth, N.J. Rourke, S.T. Ruediger, E. Rusman, R.M. Sanches-Kuiper, M.R. Schenker, J.M. Seoane, R.J. Shaw, M.K. Shiver, S.W. Short, N.L. Sizto, J.P. Sluis, M.A. Smith, J. Ernest Sohna Sohna, E.J. Spence, K. Stevens, N. Sutton, L. Szajkowski, C.L. Tregidgo, G. Turcatti, S. Vandevondele, Y. Verhovsky, S.M. Virk, S. Wakelin, G.C. Walcott, J. Wang, G.J. Worsley, J. Yan, L. Yau, M. Zuerlein, J.C. Mullikin, M.E. Hurles, N.J. McCooke, J.S. West, F.L. Oaks, P.L. Lundberg, D. Klenerman, R. Durbin, A.J. Smith, Accurate whole human genome sequencing using reversible terminator chemistry, Nature 456 (2008) 53–59.

[64] K. Pelak, K.V. Shianna, D. Ge, J.M. Maia, M. Zhu, J.P. Smith, E.T. Cirulli, J. Fellay, S.P. Dickson, C.E. Gumbs, E.L. Heinzen, A.C. Need, E.K. Ruzzo, A. Singh, C.R. Campbell, L.K. Hong, K.A. Lornsen, A.M. McKenzie, N.L. Sobreira, J.E. Hoover-Fong, J.D. Milner, R. Ottman, B.F. Haynes, J.J. Goedert, D.B. Goldstein, The characterization of twenty sequenced human genomes, PLoS Genet. 6 (2010).