Chapter 6
CNVs and the
NMDA Receptor Complex

6.1 NMDA Receptor Complex, Schizophrenia and Cognition

Molecules at the synapse, the specialized junction between neurons, play a fundamental role in learning and memory, neural development and neural degeneration (Nakanishi et al. 1998; Kandel 2000). Alterations to these synaptic proteins can result in different human cognitive and behavioral phenotypes, including a number of neuropsychiatric disorders (Laumonnier et al. 2007; Ramocki and Zoghbi 2008; Sudhof 2008). Within the synapse proteome, the NMDA Receptor Complex is of particular interest due to its central role in glutamatergic signalling for synaptic transmission in the brain. The NMDA receptor complex (NRC), also known as MAGUK (membrane-associated guanylate kinase) Associated Signalling Complex (MASC), encompasses a key set of synaptic proteins in the postsynaptic density (PSD) (Husi et al. 2000; Husi and Grant 2001; Collins et al. 2006), centered around the NMDA receptor (*NMDAR,* N-Methyl D-Aspartate ionotropic receptor) (composed of NR1 and NR2 subunits) and associated MAGUK scaffolding proteins (e.g. *PSD-95*, *PSD-93*, *SAP-102* and *SAP-97*) (Figure 6.1).

The NRC/MASC complex comprises of 186 proteins (Collins et al. 2006) (Appendix F). A number of them (43 out of 186) have demonstrated involvement in synaptic plasticity (Grant et al. 2005), the changes in synaptic properties in response to neuronal stimulation, which form the basis of learning and memory formation. 54 of these synaptic proteins were associated with one or more nervous system disorders (Grant et al. 2005). Unraveling the functional complexities of these components is critical for our understanding of diseases, behaviors and the learning processes (Pocklington et al. 2006).

Pertinent to the investigation of psychiatric disorders, NRC/MASC components have provided support for the glutamate hypothesis of schizophrenia (section 1.5).

Hypofunction of the NR1 subunit of NMDAR in mice, for instance, displayed behavioral abnormalities mimicking schizophrenia (Mohn et al. 1999). Recent studies on copy number variations in schizophrenia further implicate the role of NRC/MASC proteins in schizophrenia. Candidate genes involved in schizophrenia CNV loci, for example *DLG2* (Walsh et al. 2008), *ERBB4* (Walsh et al. 2008), *NRXN1* (Rujescu et al. 2008), *CHRNA7* (ISC 2008; Stefansson et al. 2008) *and CYBIP1* (Stefansson et al. 2008) (section 1.6 and Figure 1.8), all converge to a direct or indirect involvement in the NRC/MASC complexes.

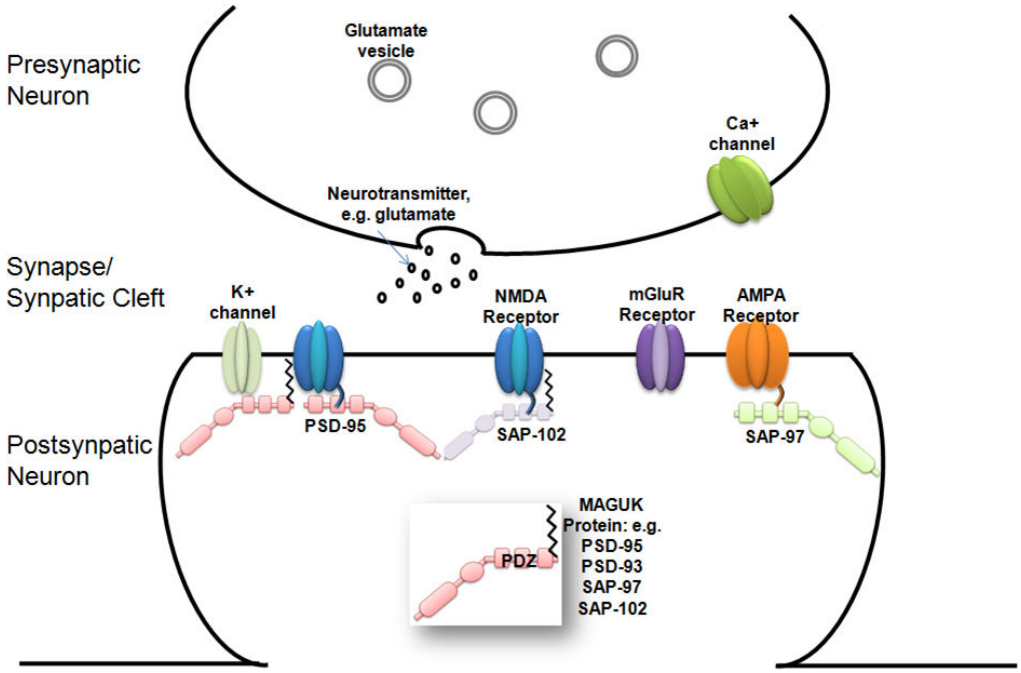

Figure 6.1 Schematic diagram of a glutamatergic excitatory synapse. The NMDA receptor (NMDAR) is an ionotropic glutamate receptor embedded in the postsynaptic membrane. Other glutamate receptors include the AMPA Receptor (alpha-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid receptor) and the mGluR receptor (metabotropic glutamate receptor). NMDAR and its associated proteins, such as the MAGUK proteins, were clustered at the postsynaptic neuron forming part of the postsynaptic density (PSD).

6.2 Copy Number Variation and the NMDA Receptor Complex

We assessed whether genes encoding NRC/MASC components are located at copy number variable loci. 186 NRC/MASC signalling complex components were mapped to the genomic coordinates of WGTP BAC clones for comparison against known CNVs in 269 normal HapMap individuals. 20 out of 186 (11%) MASC/NRC complex components were detected within CNV regions in one or more samples (herein refer as MASC CNVs) (Table 6.1, Figure 6.2).

The 20 MASC CNVs were characterized into 5 categories according to a previously defined classification system (Redon et al. 2006)[4]. Fifteen of them were classified as simple deletions, duplications or deletions with duplications. These CNVs were at relatively low frequency in the population (on average 1.4% of HapMap samples affected, corresponding to ~4 individuals out of 269).

The remaining MASC CNVs were complex or multiallelic, affecting the majority of the HapMap population. The most frequent variants were located near the genes *NSF* (N-ethylmaleamide-sensitive factor) (Figure 6.2q), *RPL13* (ribosomal protein L13) (Figure 6.2p) and *GRM5* (metatropic glutamate receptor 5) (Figure 6.2j). In particular, *NSF* was identified as a multi-allelic variant with at least 4 discrete CNV genotypes. Due to the interesting genomic architecture of this locus and the biological relevance of the *NSF* gene, we investigated the genomic region in more detail (see section 6.3).

[4] The classification system categorizes CNVs into 5 categories, namely: “deletions”, “duplications”, “deletions and duplications”, “multi-allelic” and “complex rearrangement”

Table 6.1 CNVs detected at 20 genes encoding NRC/MASC signalling complex components. The most frequent CNVs are highlighted in yellow, and the CNVs affecting core components of the MASC complex in blue.

Gene	Gene Description	Chr	Start	End	CNV Clones (% in HapMap)	Category
GNB1	guanine nucleotide binding protein (g protein), beta	1	1706590	1812355	Chr1tp-25A7(7.43%)	C
PKLR	pyruvate kinase, liver and rbc	1	153526254	153537843	Chr1tp-21C5(1.49%), Chr1tp-7H12(1.49%)	A
DLG1	**discs, large homolog 1**	**3**	**198255819**	**198509844**	**Chr3tp-20C11(13.38%)**	**E**
MOG	myelin oligodendrocyte glycoprotein	6	29732788	29748128	Chr6tp-22B5(0.37%), Chr6tp-1F1(0.37%)	A
YWHAG	tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, gamma polypeptide	7	75794053	75826252	Chr7tp-9C8(3.72%)	C
HRAS	v-ha-ras harvey rat sarcoma viral oncogene homolog	11	522242	525591	Chr11tp-17H2(0.37%)	B
SLC25A22	solute carrier family 25 (mitochondrial carrier: glutamate), member 22	11	780478	786221	Chr11tp-13G5(0.37%)	B
SLC1A2	solute carrier family 1 (glial high affinity glutamate transporter), member 2	11	35229329	35398186	Chr15tp-5C1(0.37%)	A
DLG2	**discs, large homolog 2, chapsyn-110**	**11**	**82843701**	**85015962**	**Chr11tp-13D9(0.37%), Chr11tp-18E5(0.37%)**	**A**
GRM5	**glutamate receptor, metabotropic 5**	**11**	**87881006**	**88438761**	**Chr11tp-5E4(4.83%), Chr11tp-11C4(13.01%), Chr11tp-3E11(13.38%)**	**E**
GAPDH	glyceraldehyde-3-phosphate dehydrogenase	12	6513918	6517797	Chr12tp-17B1(0.37%)	A
PDPK1	3-phosphoinositide dependent protein kinase-1	16	2527971	2593190	Chr16tp-11D7(2.6%)	C
GRIN2A	**glutamate receptor, ionotropic, n-methyl d-aspartate 2a**	**16**	**9754762**	**10184112**	**Chr16tp-12H9(0.74%)**	**C**
MAPK3	mitogen-activated protein kinase 3	16	30032928	30042131	Chr16tp-3F5(3.35%)	E
GOT2	glutamic-oxaloacetic transaminase 2, mitochondrial (aspartate aminotransferase 2)	16	57298538	57325747	Chr16tp-12E8(1.49%), Chr16tp-3F2(1.49%)	B
RPL13	**ribosomal protein l13**	**16**	**88154595**	**88158451**	**Chr16tp-1H4(21.56%)**	**E**
NSF	**n-ethylmaleimide-sensitive factor**	**17**	**42023391**	**42189997**	**Chr17tp-2F6(64.31%), Chr17tp-2F9(56.51%)**	**D**
DLGAP1	**discs, large (drosophila) homolog-associated protein 1**	**18**	**3486030**	**3870135**	**Chr18tp-8F6(0.37%), Chr18tp-3A2(0.74%), Chr18tp-6E9(0.74%)**	**A**
AKT2	v-akt murine thymoma viral oncogene homolog 2	19	45428064	45483105	Chr19tp-6A8(0.37%), Chr19tp-4C5(0.37%)	A
PLCB1	phospholipase c, beta 1 (phosphoinositide-specific)	20	8061296	8813547	Chr20tp-5C4(0.37%), Chr20tp-6E3(0.37%), Chr20tp-7D8(0.37%), Chr20tp-4C4(0.37%), Chr20tp-4B2(0.37%), Chr20tp-6F2(0.37%), Chr20tp-3B9(0.37%), Chr20tp-6B1(0.37%)	A

* CNV categories: *A*: deletions, *B*: duplications, *C*: deletions and duplications, *D*: multi-allelic and *E*: complex re-arrangement.

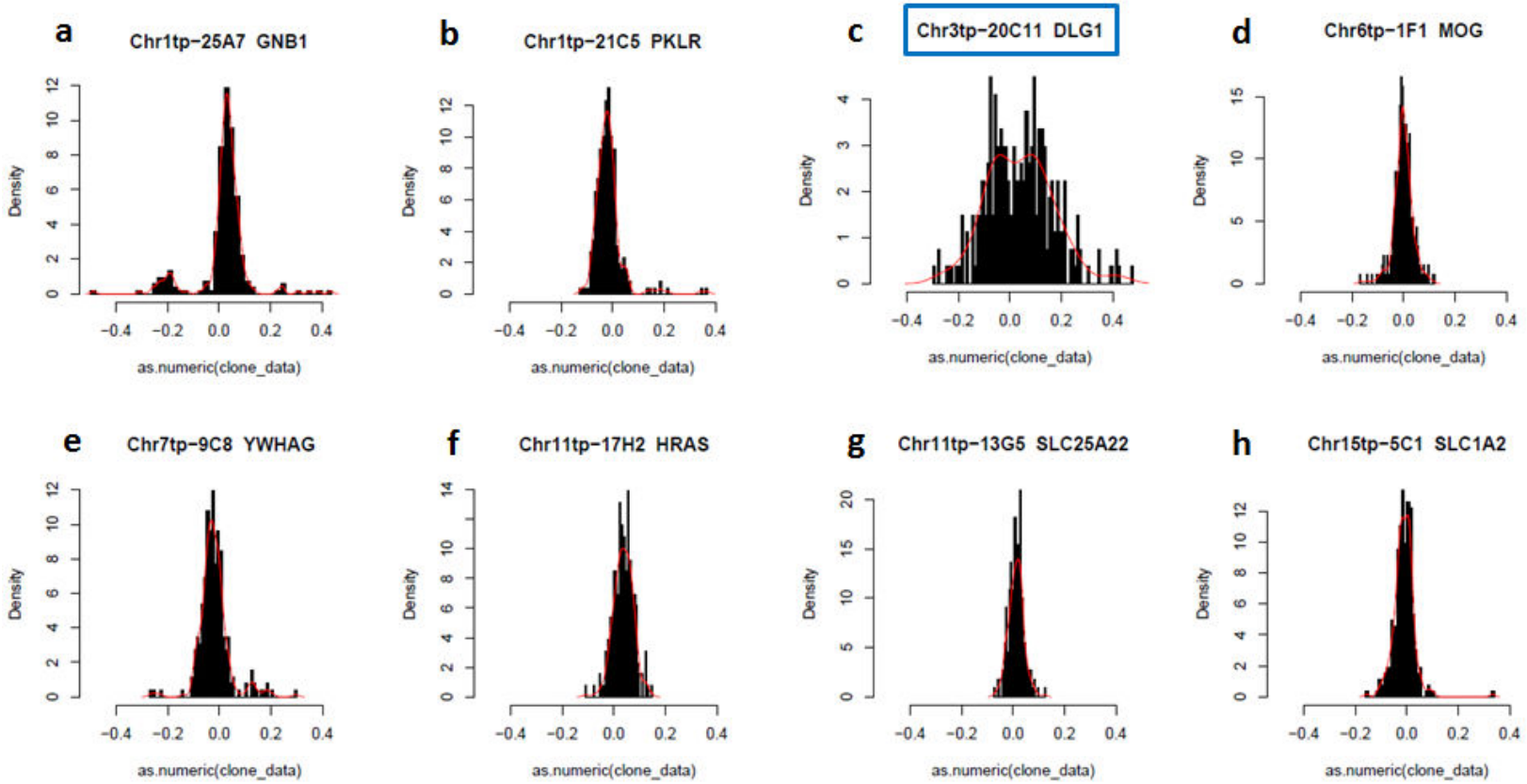

Figure 6.2 (to be continued)

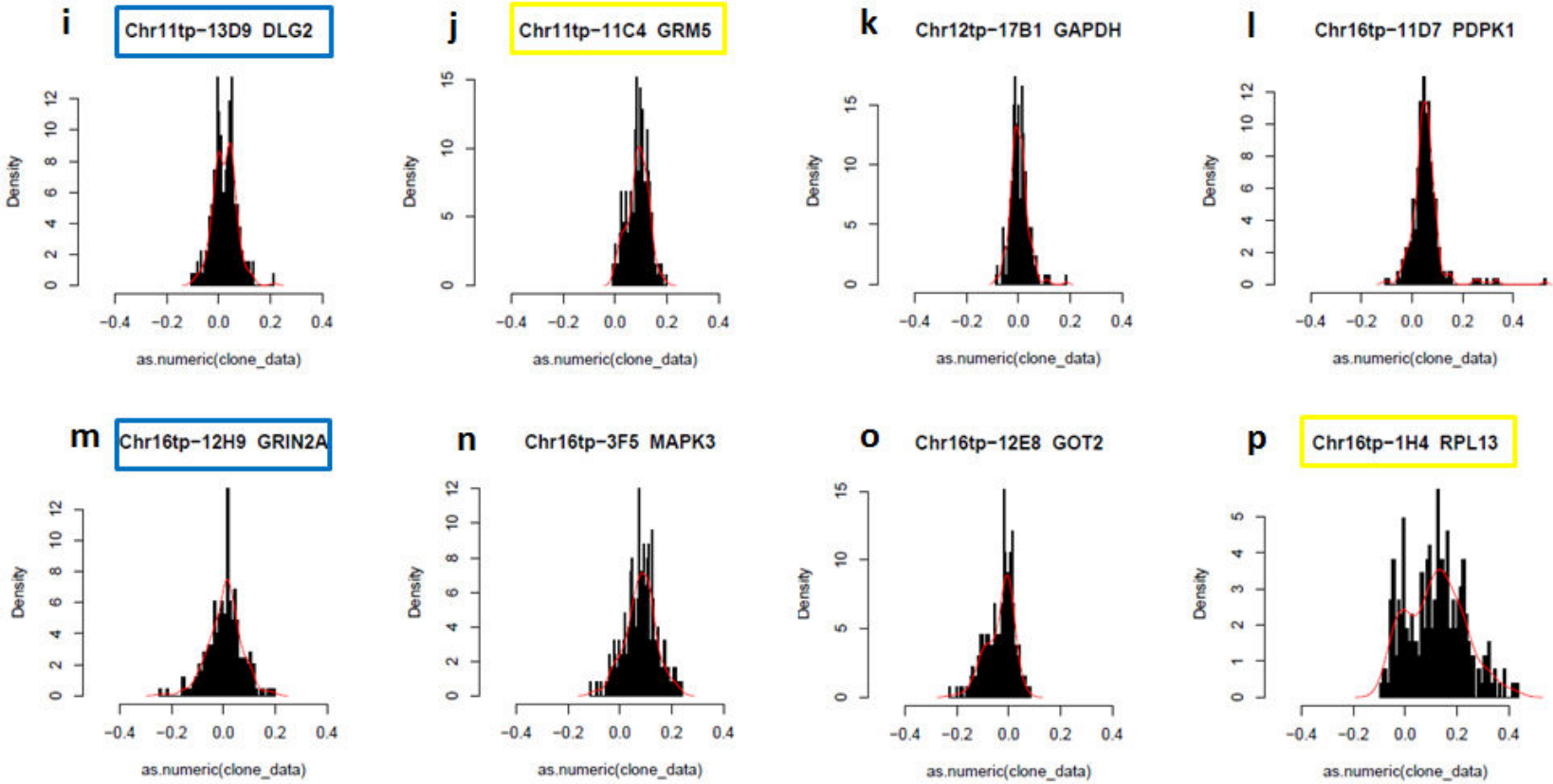

Figure 6.2 (to be continued)

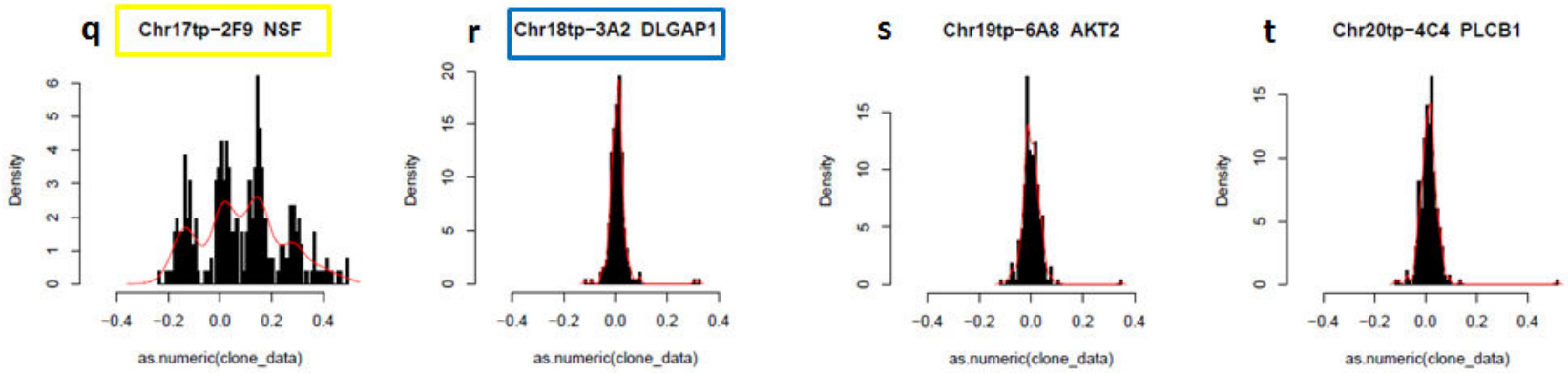

Figure 6.2 Log2ratio distributions for WGTP clones reporting CNV among 269 HapMap samples in 20 MASC regions. The most frequent CNVs were highlighted in yellow, and the CNVs affecting core components of the MASC complex in blue. *a: GNB1: ; b: PKLR; c: DLG1; d: MOG; e: YWHAG; f: HRAS; g: SLC25A22;h: SLC1A2; i: DLG2; j: GRM5; k: GAPDH; l: PDK1; m: GRIN2A; n: MAPK3; o: GOT2; p: RPL13; q: NSF; r: DLGAP1; s: AKT2; t: PLCB1*

The core NRC/MASC components, i.e. the NMDA receptor subunits and the closely associated MAGUK scaffolding proteins, were of particular interest for understanding cognition and diseases. Among the 20 MASC CNVs, 4 overlapped with the core NRC/MASC genes, detected as variants at *DLG1*, *DLG2*, *DLGAP1* and *GRIN2A* (Figure 6.3; Table 6.1 in yellow).

The CNV at *DLG1* (Discs, Large Homolog 1 (Drosophila)) showed a complex pattern, as displayed by a broad log2ratio distribution of clone *Chr3tp-20C11* (Figure 6.4a). At WGTP resolution, we were not able to resolve the precise CNV location and discrete copy number for individuals. The WGTP data was subsequently compared to the recently-released oligonucleotide array data (Genome Structural Variation Consortium 2008). We confirmed the presence of a frequent intronic CNV of ~5 kb within *DLG1*.

In addition, a duplication involving *DLG2* (Discs, Large Homolog 2 (Drosophila)) was detected in one individual in the WGTP platform. Oligo array data revealed additional CNVs in more individuals (deletions) (Figure 6.4b). *DLGAP1* (Discs Large-Associated Protein 1) was also identified within a larger duplication, affecting 3 consecutive BAC clones in the WGTP data (Figure 6.4c). A number of deletions were revealed by the oligo array data from the CNV Consortium.

Finally, *GRIN2A* (glutamate receptor, ionotropic, n-methyl d-aspartate 2a), encoding the NR2A subunit of the NMDA receptor, was detected with both duplications and deletions in the population by WGTP platform (Figure 6.4d), although no CNV was detected using higher-resolution oligonucleotide array.

In summary, copy number variants at multiple components of the NRC/MASC complex exist among normal, healthy individuals, despite the significance of these proteins in

normal brain functions such as synaptic plasticity and cognition. We further investigated the *NSF* CNV locus as it was the most frequent MASC CNV (section 6.3).

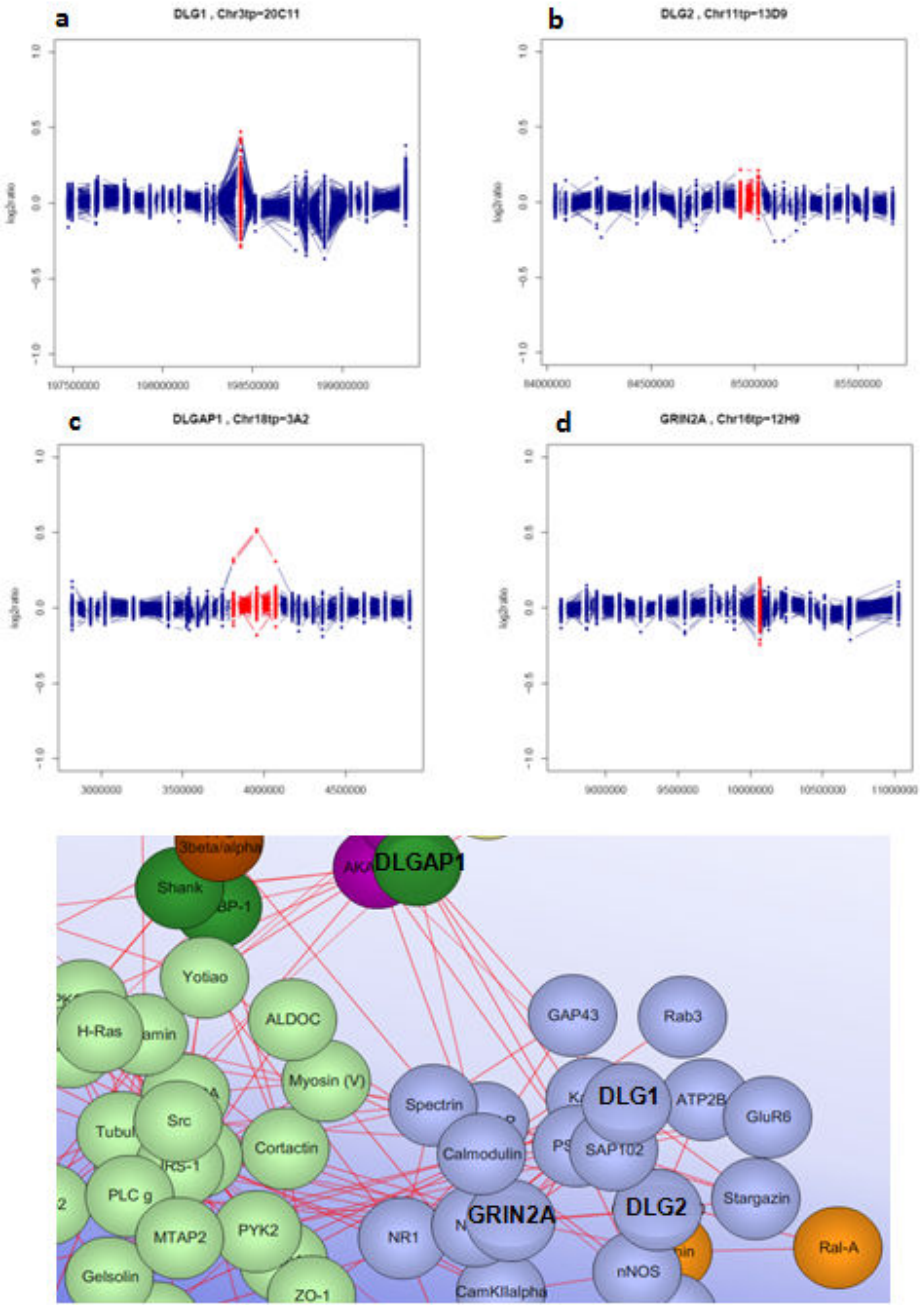

Figure 6.3 CNVs affecting core components of the NRC/MASC signalling complex. (a)-(d): WGTP array CGH profiles of the four CNVs in 269 HapMap individuals. (a) CNV at *DLG1* with complex pattern in the HapMap population b) Duplication at *DLG2* in one sample c) Large duplication at *DLGAP1* in 2 samples d) CNV at *GRIN2A* detected as both duplication and deletion (e) The four core NRC components within the network of NRC/MASC signalling complex (diagram modified from www.genes2cognition.com)

6.3 CNV at 17q21 near N-ethylmaleamide-Sensitive Factor (NSF)

6.3.1 Introduction to the 17q21 locus near *NSF*

We detected a multi-allelic variant near *NSF*, the most frequent CNV in the NRC/MASC CNV screen. The *NSF* CNV region is highly variable with a complex genomic architecture and association with mental illness, prompting us to further investigate the region. Structural variants in the region have been identified, including a well-studied common inversion of ~900 kb (Stefansson et al. 2005), resulting in two designated alleles H1 and H2. The H2 allele was suggested to be under positive selection. The inversion encompasses several genes involved in brain development including *NSF* (N-ethylmaleamide-sensitive factor, a membrane fusion protein), *CRHR1* (corticotrophin releasing hormone receptor), *IMP5* (intramembrane protease 5, a presenilin homologue), *MAPT* (microtubule associated protein tau) and *STH* (saitohin, an intronless gene within MAPT) (Stefansson et al. 2005). Among these, *NSF* is involved in the NRC/MASC complex, and *MAPT* is a component of the broader postsynaptic density (PSD) complex (Collins et al. 2006).

The H1/H2 inversion suppresses recombination, leading to divergence of the two alleles. The H2 haplotype is shown to exist only in populations with history of European admixture (Evans et al. 2004; Stefansson et al. 2005). Overall, H2 shows little genetic variation (Hardy et al. 2005). In contrast, many H1 variants have been described, with the most common reported variants being single nucleotide polymorphisms (SNPs) or small insertions/deletions that modify *MAPT* (tau) expression or splicing mechanisms (Rademakers et al. 2005; Caffrey et al. 2006; Caffrey et al. 2007; Myers et al. 2007; Hayesmoore et al. 2008).

The H1 & H2 inversion has been described in the context of several neurodegenerative diseases associations. The H2 haplotype was reported to have a protective role in a number of brain diseases, including progressive supranuclear palsy and corticobasal degeneration (de Silva et al. 2001; Pastor et al. 2004; Pittman et al. 2005; Rademakers et al. 2005), Parkinson's disease (Goris et al. 2007; Zabetian et al. 2007) and frontotemporal dementia (Borroni et al. 2005; Ghidoni et al. 2006) (more comprehensive reviews in (Rademakers et al. 2004; Kalinderi et al. 2008)). H2 was also demonstrated to have higher expression of a variant of *MAPT* mRNA transcript known as 4R, where the fine balance of 4R transcript with its 3R counterpart is critical in certain neurodegenerative diseases (Kalinderi et al. 2008).

More recently the 17q21 locus has been shown to be involved in a newly defined genomic disorder. A recurrent microdeletion of the genomic locus was detected in patients with learning disability (Koolen et al. 2006; Sharp et al. 2006; Shaw-Smith et al. 2006). The critical deletion region (~420 kb) was similar in all patients identified. In deletion carriers where parental DNA is available, the deletions were shown to be *de novo*. Moreover, for all patients identified so far, a H2 parental background is a necessary factor for the deletion to occur (Koolen et al. 2008). Clinical cases of mental retardation patients with the reciprocal duplication of the microdeletion syndrome have been detected (Kirchhoff et al. 2007), demonstrating the genomic instability of the region.

6.3.2 Known Genomic Structure of 17q21

The structure of the 17q21 genomic interval has been extensively described by Cruts *et al.* and Pittman *et al.* (Pittman et al. 2004; Cruts et al. 2005). There are three main LCR modules in the region, each with various subunits of duplication blocks (Figure 6.4). These LCRs flank both the H1/H2 inversion and the microdeletion syndrome breakpoints (Stefansson et al. 2005; Koolen et al. 2006; Sharp et al. 2006; Shaw-Smith et al. 2006). Specific segmental duplicons were attributed to the genomic instability of the region. H1 and its sub-haplotypes, together with the inverted H2 allele, differ in terms of the number and the orientation of the duplication blocks. Copy number polymorphisms within H1 at the *NSF* locus, defining several H1 subclades, have previously been reported (Stefansson et al. 2005). The 17q21 locus is a case in point for characterizing the relationship between segmental duplications, copy number changes, genomic instability and their phenotypic consequence.

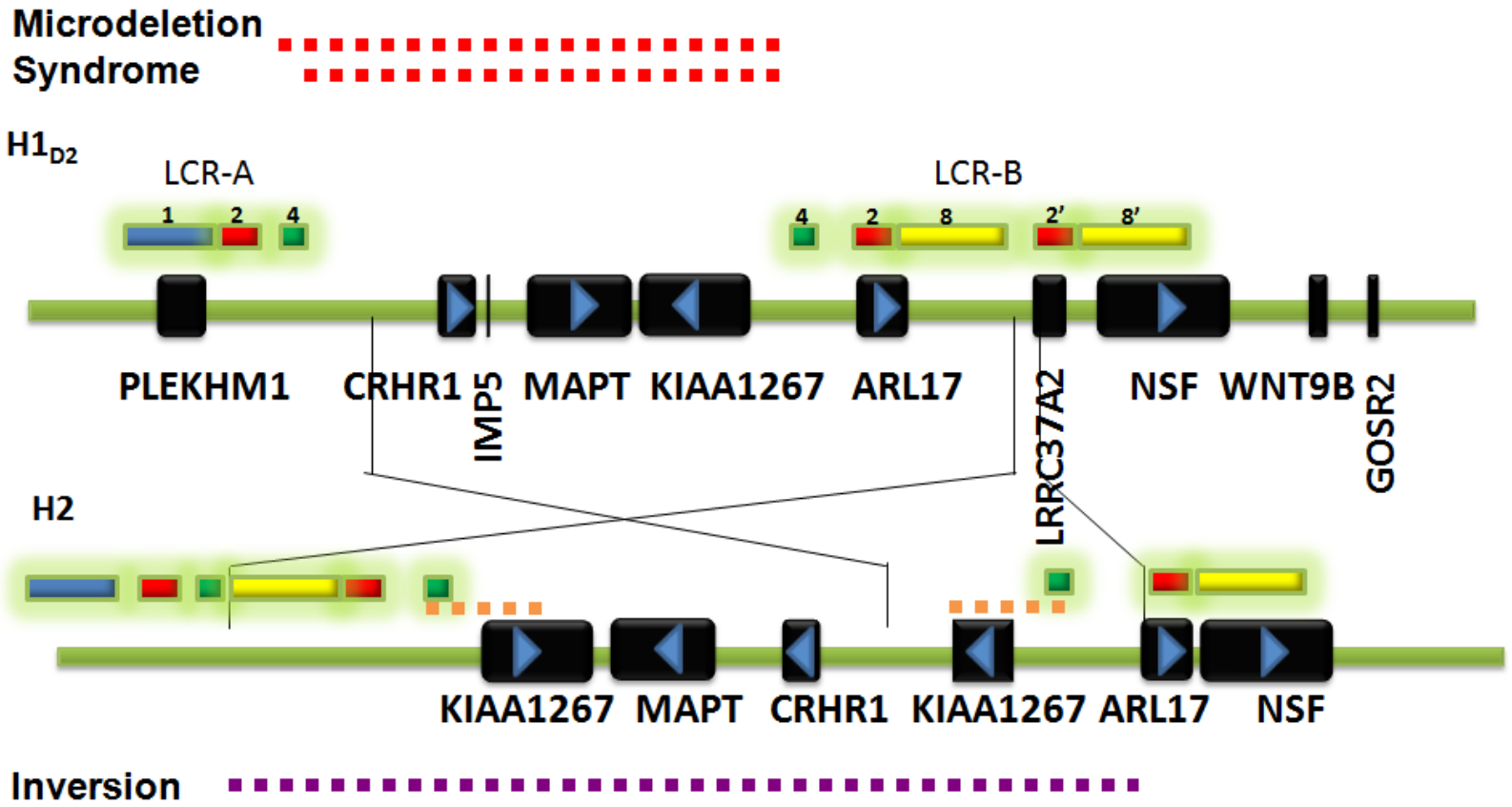

Figure 6.4 Schematic representation of the 17q21 locus. The H1-D2 allele (as in the reference assembly, UCSC genome browser) and the inverted H2 allele are mapped with low copy repeats (LCR) and major segmental duplication blocks. Estimated locations of the H1/H2 inversion and the 500 kb microdeletion syndrome with learning disability is shown as dotted purple and red lines respectively.

6.3.3 Array CGH Data Reveals Two Major CNVs at 17q21

The MASC CNV screen provided evidence for copy number variations near the *NSF* gene at 17q21. Here we present a detailed molecular and cytogenetic study on the genomic structure of the chr17q21 locus. Analysis of the WGTP array data on 269 HapMap samples from three ethnicities[5] delineated two major blocks of CNV: CNV_{NSF} (Figure 6.5, right) and $CNV_{KIAA1267}$ (Figure 6.5, left) (with reference to the genes involved).

6.3.3.1 $CNV_{NSF:}$ Copy number Variant at 5' end of NSF

CNV_{NSF} (~41.7 Mb to 42.2 Mb NCBI36) is part of the low copy repeat LCR A (segmental duplicon subunits 2 and 8) (Cruts et al. 2005). It encompasses the truncated duplication of exons 1-13 of the gene *NSF* (N-ethylmaleamide-sensitive factor). This CNV has been demonstrated previously through sequence analysis in the Icelandic population (Steffanson et al), and has led to the definition of 3 H1 sub-alleles designated as ***H1-D0-1*** or ***H1-D1*** (both having 1 copy of NSF), ***H1-D2*** (2 copies of NSF) and ***H1-D3*** (3 copies of NSF).

Bivariate clustering of the log2ratio of two clones overlapping *NSF* (*Chr17tp-2F9* and *Chr17tp-12B1*) confirmed the multi-allelic nature of CNV_{NSF}, but revealed higher complexity than previously expected. Given that the reference DNA used for array hybridization has a three copies of the sequence (demonstrated by FISH experiments), samples from Yoruba are estimated to have diploid copy numbers varying from 2 to 5 (4 distinct genotype signals), while samples from Asia carry diploid copy numbers varying from 2 to 7 (6 distinct genotype signals) (Figure 6.6). Data from the European samples is less easy to interpret, due to homology of BAC sequences to the second CNV block

[5] The three populations are CEU: Samples from Utah with European ancestry; YRI: Samples from Yoruba with African ancestry; JPT+CHB: Samples from Japanese and Chinese with Asian ancestry

$CNV_{KIAA1267}$, but they also seem to display diploid copy numbers varying from 2 to 7. Combining the evidence, a maximum of 7 copies of the truncated *NSF* (diploid state) was identified, suggesting the existence of a ***H1-D4*** allele with 4 copies of 5' *NSF* (haploid state). The presence of such allele was confirmed in a CEU sample by FISH experiments (section 6.3.3).

6.3.3.2 $CNV_{KIAA1267:}$ Copy number Variant at KIAA1267

$CNV_{KIAA1267}$ (left, ~41.5 Mb to 41.7 Mb), the second major CNV block, overlaps with subunit 4 of LCR-A (Cruts et al. 2005), and a brain-expressed gene *KIAA1267* (Oliveira et al. 2004). The existence of structural variation at the region has previously been implicated in the studies of H1 and H2 inversion haplotypes. The H2 inverted allele and the H1-D1 allele (a variant of H1) were suggested to have overlapping but not identical duplications at this region (Stefansson et al. 2005).

Clustering the log2ratio of BAC clones at $CNV_{KIAA1267}$ (*Chr17tp-2G12* and *Chr17tp-3D7*) generated 3 distinct clusters in the CEU samples, corresponding to 2, 3 and 4 copies of $CNV_{KIAA1267}$. In Yoruba or Asian samples, on the other hand, the CNV is non-variant. In conclusion, the higher copy number of $CNV_{KIAA1267}$ is European specific.

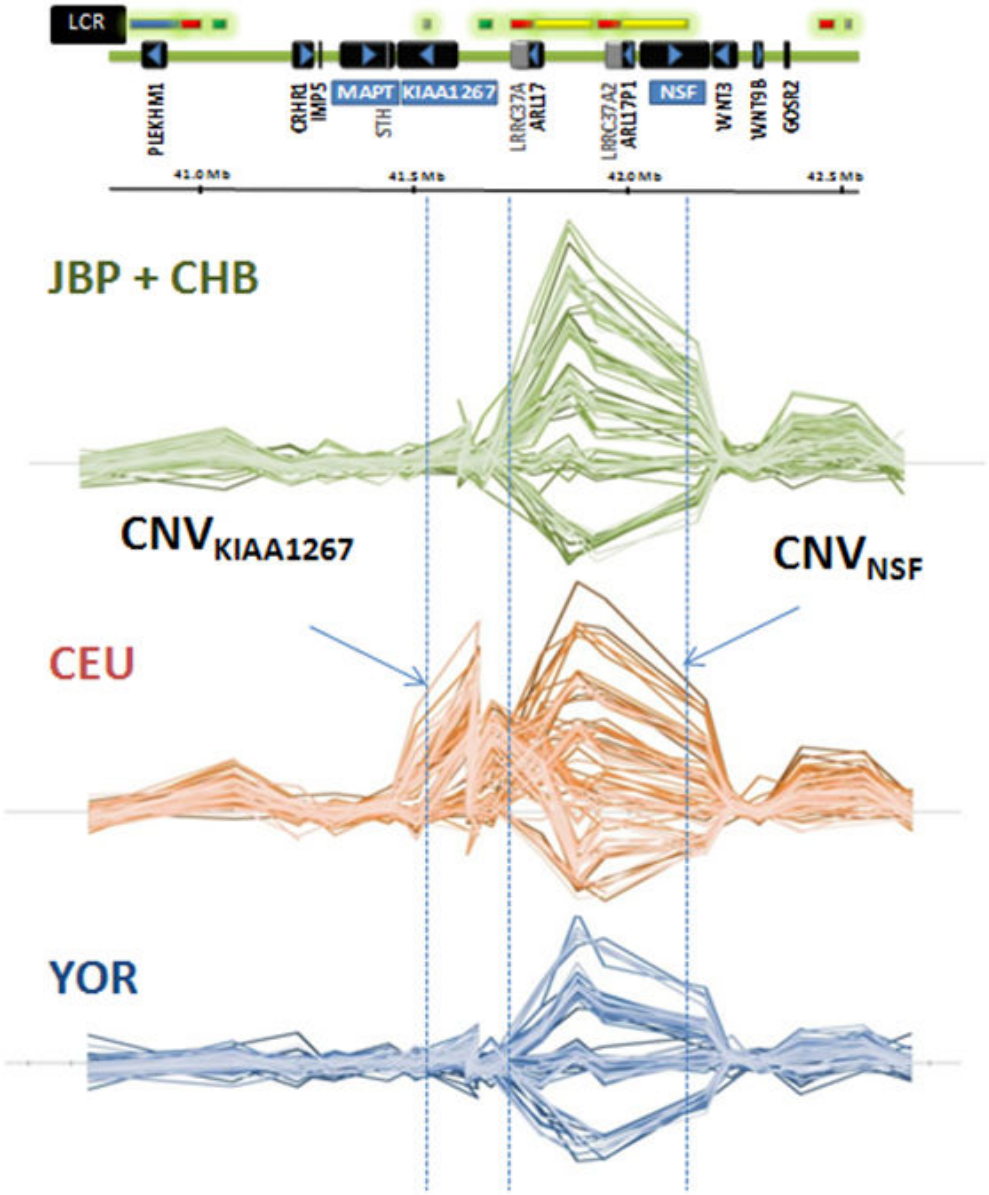

Figure 6.5 Array CGH genomic profiles of the 3 HapMap ethnic groups at chr17q21. Top panel: Schematic representation of the region chromosome 17 40.8 Mb- 42.5 Mb, with locations of genes and low copy repeats for reference. Bottom panel: Genomic profiles from WGTP data on samples of Asian ancestry (JPT+CHB), samples of European ancestry (CEU) and samples of African ancestry (YOR). Two major blocks of CNVs are indicated in the diagram (*left*: $CNV_{KIAA1267}$; right: CNV_{NSF}).

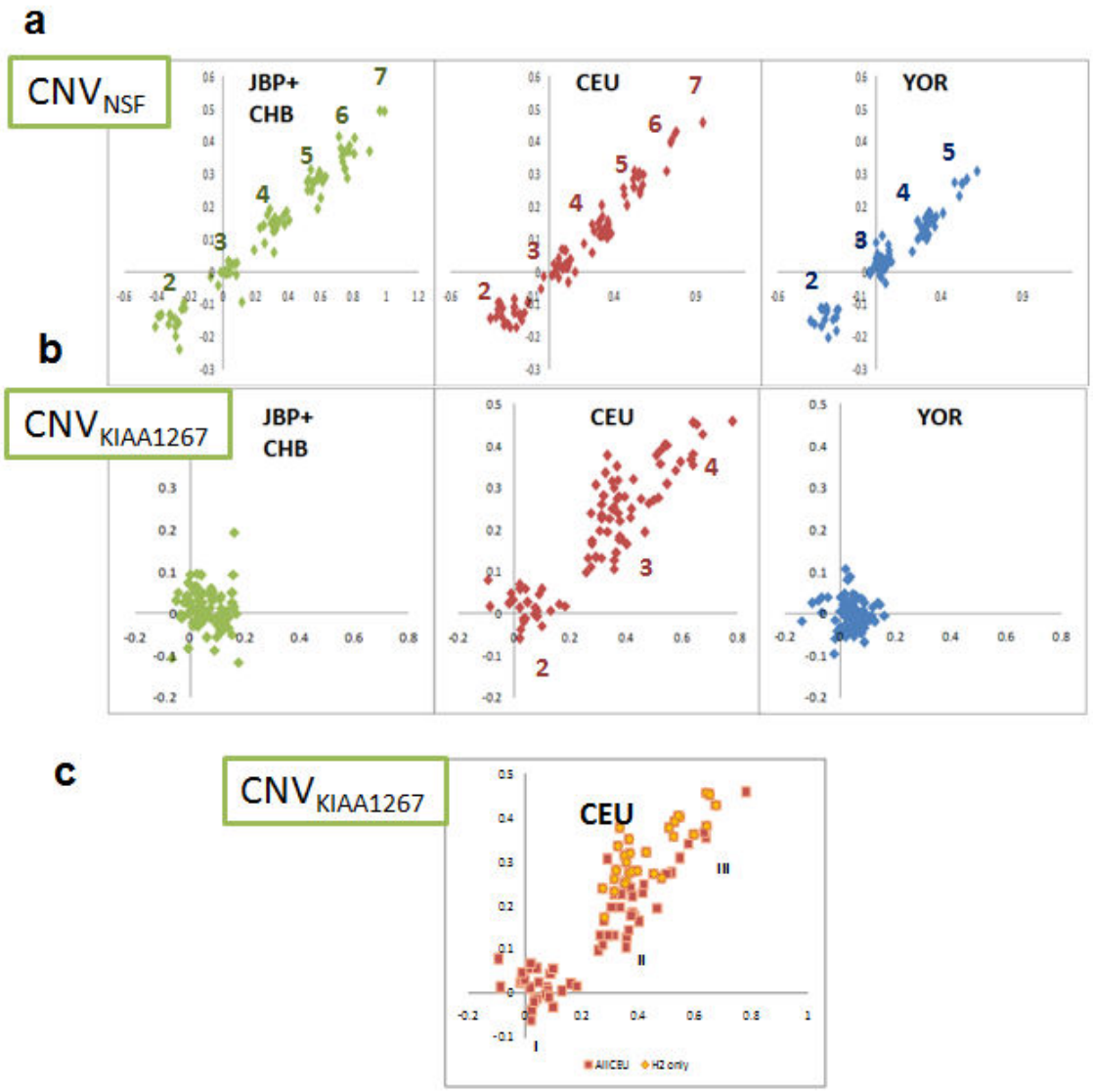

Figure 6.6 Population differentiation at two major CNV blocks at 17q21 Bivariate clustering of log2ratio from WGTP data demonstrates population bias at CNV_{NSF} and $CNV_{KIAA1267}$. **a)** *CNV_{NSF}: 2-7 copies (diploid) in JPT+CHB and CEU; 2-5 copies (diploid) in YOR. **b)** **$CNV_{KIAA1267}$: 2-4 copies in CEU; non-variant in JPT+CHB or YOR (2 diploid copies in each individual). **c)** H2 carriers (orange) are mapped exclusively to the higher copy number clusters (II and III) of the CEU diagram representing $CNV_{KIAA1267}$. The remaining samples at cluster II and III belongs to an H1 allele variant, H1-D1.

* Bivariate clustering of log2ratio from *Chr17tp-14B1* (x-axis) and *Chr17tp-2F9* (y-axis)
** Bivariate clustering of log2ratio from *Chr17tp-2G12* (x-axis) and *Chr17tp-3D7* (y-axis)
JPT+CHB: samples from Japan and China with Asian ancestry; CEU: samples from Utah with European ancestry; YOR: samples from Yoruba with African ancestry

6.3.4 Resolving CNVKIAA1267 with SNP and High Resolution Oligo Array Data

The WGTP data demonstrated a European-specific pattern at $CNV_{KIAA1267}$. Coincidentally, H2 from the common inversion was also found in the European population or population with European admixture only (Evans et al. 2004; Stefansson et al. 2005). To explore a relationship between $CNV_{KIAA1267}$ and H2, we integrated the HapMap WGTP CNV results with corresponding SNP data (IHMC 2005). SNP rs9468, a surrogate marker for the H1/H2 inversion, was used to categorize samples into H1 or H2 carriers. Samples carrying one or more H2 alleles were plotted on the CEU diagram of $CNV_{KIAA1267}$ (Figure 6.6c). These H2 carriers (orange) all mapped to higher copy numbers of $CNV_{KIAA1267}$ (cluster II & III).

However, $CNV_{KIAA1267}$ is only partially explained by H2. Non-H2 variants were also observed at the high-copy-number clusters of $CNV_{KIAA1267}$. This suggests heterogeneity at the CNV locus. High resolution oligonucleotide from HapMap individuals (Genome Structural Variation Consortium 2008) confirms at least two types of duplications at the $CNV_{KIAA1267}$ locus (Figure 6.7). The smaller duplication was explained by H2 carriers. The larger CNV would involve the H1-D1 variant, an allele defined by previous literature as having a duplication overlapping the region. Like H2, this H1-D1 allele is absent in non-European populations (see Figure 6.6b, diagrams for JBP+CHB/YOR).

This is the first demonstration of a non-H2 variant being European specific, and it raises questions of the evolutionary relationship between this H1-D1 variant and the H1/H2 inversion.

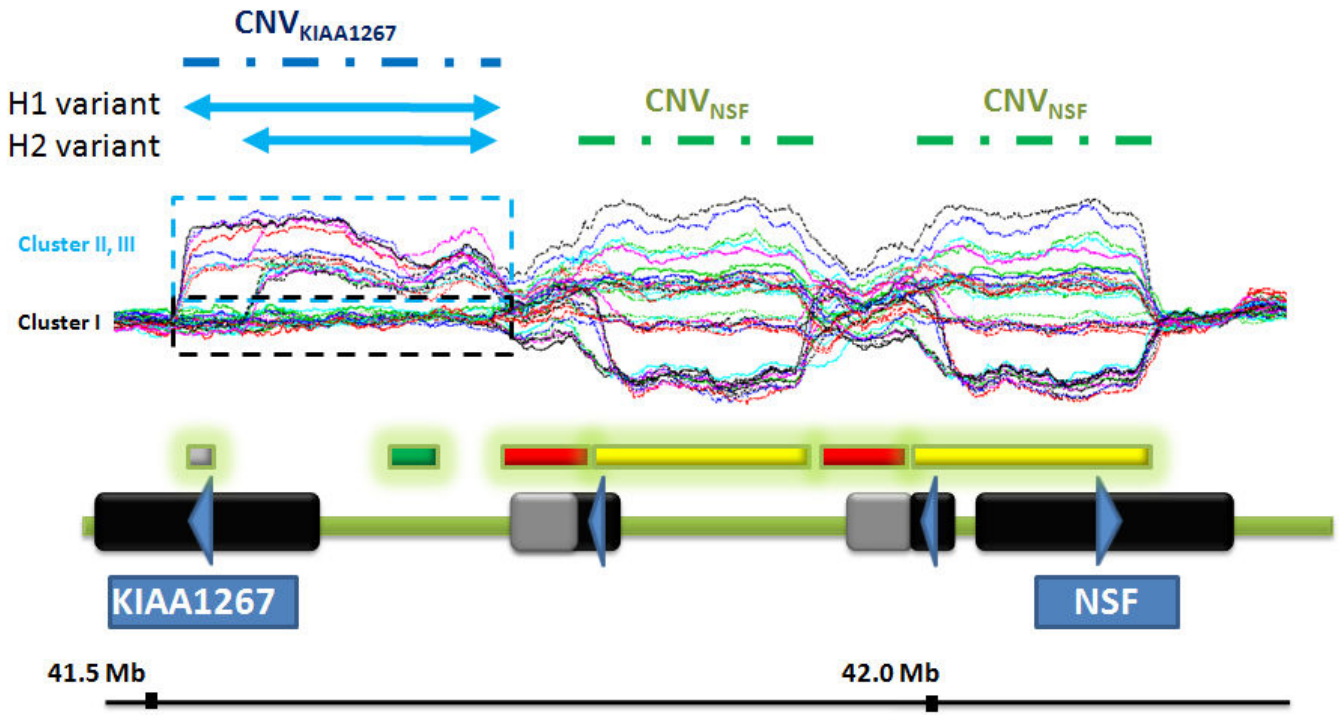

Figure 6.7 High resolution oligo array CGH profiles for 20 individuals at chr17 41.5 Mb- 42.2 Mb. Top panel: Nimblegen oligonucleotide array data for 20 HapMap CEU samples resolved the two CNV blocks: CNV_{NSF} and $CNV_{KIAA1267}$. $CNV_{KIAA1267}$ is heterogeneous, comprising of a larger variant (corresponding to H1-D1 carriers) and a smaller variant (corresponding to H2 carriers). Bottom panel: Schematic representation of the region with locations of genes and low copy repeats for reference.

6.3.5 Validating CNVNSF and CNVKIAA1267 by qPCR and FISH

Given the complexity and multi-allelic nature of the described copy number variations, array CGH data at Chr17q21 is best interpreted with independent platforms to provide quantitative and qualitative measurements in terms of the number, size, distance and orientation of CNVs.

First, quantitative real-time PCR (qPCR) was performed to validate the multi-allelic variants CNV_{NSF} and $CNV_{KIAA1267}$. Genotypes of CNV_{NSF} ranging from 2 to 6 copies (diploid) were validated, and genotypes of $CNV_{KIAA1267}$ from 2 to 4 copies (diploid) were confirmed (Figure 6.7).

To visualize the copy number and the orientation of rearrangement at the *NSF* locus, we performed fluorescent in situ hybridizations (FISH) using DNA fiber slides made from lymphoblastoid cell lines. Fosmid probes at the 5' duplicated (labeled with biotin-dUTP, green) and 3' non-duplicated ends (labeled with digoxigenin-dUTP, red) of the *NSF* CNV locus.

The haplotypes of different alleles from HapMap samples were visualized according to the number of green signals (Figure 6.8). Duplication of the 5' *NSF* is tandem, ranging from copy number of 1 to 4 (haploid) on a given allele. The presence of the ***H1-D4*** (4 copies of 5' *NSF*) as we described before is validated.

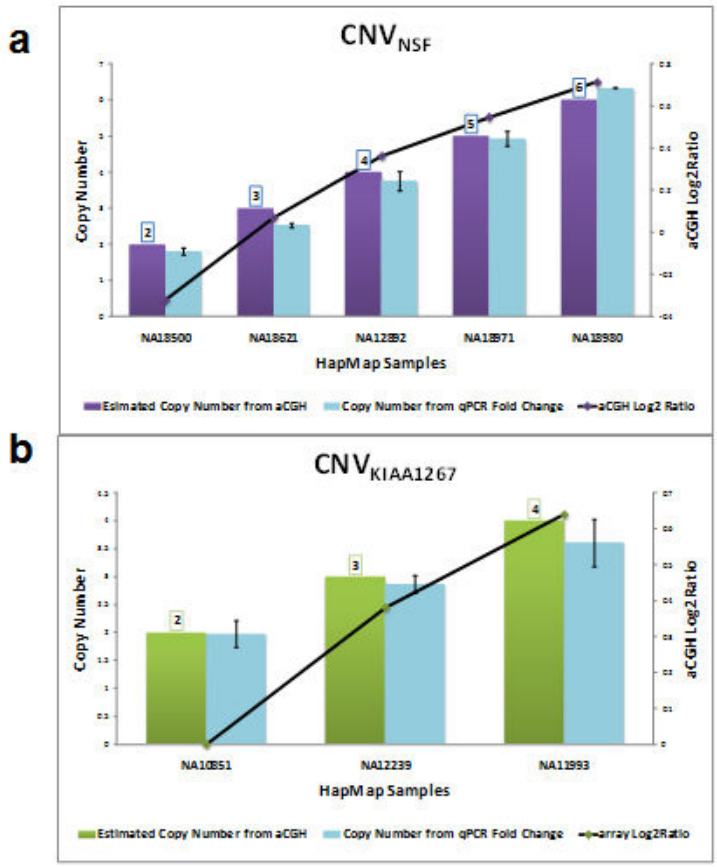

Figure 6.8 Quantitative PCR validation of $CNV_{KIAA1267}$ and CNV_{NSF}. a) CNV_{NSF} detected as 2 to 6 copies*. **b)** $CNV_{KIAA1267}$ detected as 2 to 4 copies*.
** diploid copy number*
blue: copy number calculated from qPCR; purple or green: expected copy number

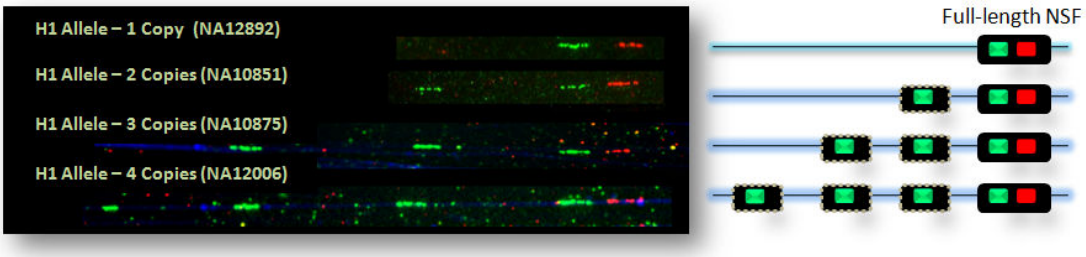

Figure 6.9 Fiber-FISH experiment to visualize copy number of CNV_{NSF}. A fosmid probe at the 5' *NSF* portion overlapping with CNV_{NSF} was labeled green and another fosmid at the 3' *NSF* invariant portion was labeled red. Alleles with tandem duplication from 1 to 4 copies* of CNV_{NSF} were observed
** haploid copy number*

6.3.6 Genotyping HapMap Individuals for 17q21 Structural Variants

Combining available genetic data (CNV and SNP) with the findings of various structural variations in the region, we genotyped all unrelated HapMap individuals for the 17q21 allele carried (children from the trios were removed from analysis). Table 6.2 demonstrates the trends of population differentiation. Horizontally across the table one sees a European specificity for both H2 and H1-D1 alleles, corresponding to higher copy numbers of $CNV_{KIAA1267.}$ Vertically down the table one sees an Asian and European bias for the highest copy numbers (6 or 7) of CNV_{NSF}.

Table 6.2 CNV genotypes for all unrelated HapMap individuals* at $CNV_{KIAA1267}$ and CNV_{NSF}.

CNV-NSF	H2 allele			H1-D1 allele			Other H1 Variants			Population Total		
Copy Number	ASIA	CEU	YOR	ASIA	CEU	YOR	ASIA	CEU	YOR	ASIA	CEU	YOR
2	0	0	0	0	15	0	30	15	20	30	30	20
3	0	11	0	0	12	0	34	11	58	34	34	58
4	0	4	0	0	4	0	50	22	36	50	30	36
5	0	7	0	0	0	0	34	9	6	34	16	6
6	0	2	0	0	0	0	26	4	0	26	6	0
7	0	0	0	0	0	0	6	0	0	6	0	0
Total	0	24	0	0	31	0	180	61	120	180	116	120

*90 ASIA samples, 58 CEU samples, 60 YOR samples

**offspring from the trios in CEU and YOR were removed from analysis so that only unrelated HapMap individuals were genotyped*

6.3.7 CNVNSF and CNVKIAA1267 in Schizophrenia Versus Control

In chapter 4.6 we introduced bivariate clustering of BAC clone log2ratio in the WGTP data to genotype CNV and test for disease association. The 17q21 region near *NSF* was one of the 31 regions across the genome with significant bias between schizophrenia patients and controls. We examined this region in the case and control cohorts in more detail with respect to CNV_{NSF} and $CNV_{KIAA1267}$.

We genotyped CNV_{NSF} and $CNV_{KIAA1267}$ using the same clones as in the HapMap analysis. Clusters were manually merged for CNV genotype counts. (Figure 6.9)

Both the Scottish SCZ and LBC cohorts demonstrate a full range of copy numbers for CNV_{NSF} (2-6+) and $CNV_{KIAA1267}$ (2-4), consistent with HapMap samples of European ancestry. For CNV_{NSF} there is a slight bias of patients at lower copy number but the difference is non-significant after clusters are manually merged (p value = 0.1931, chi-squared test). For $CNV_{KIAA1267}$ there is a nominally significant bias (p-value= 0.01689) of an association of low copy number with disease (Table 6.3). Nevertheless, the result is treated with caution since our case and control cohorts are modest in size and may therefore subject to sampling bias.

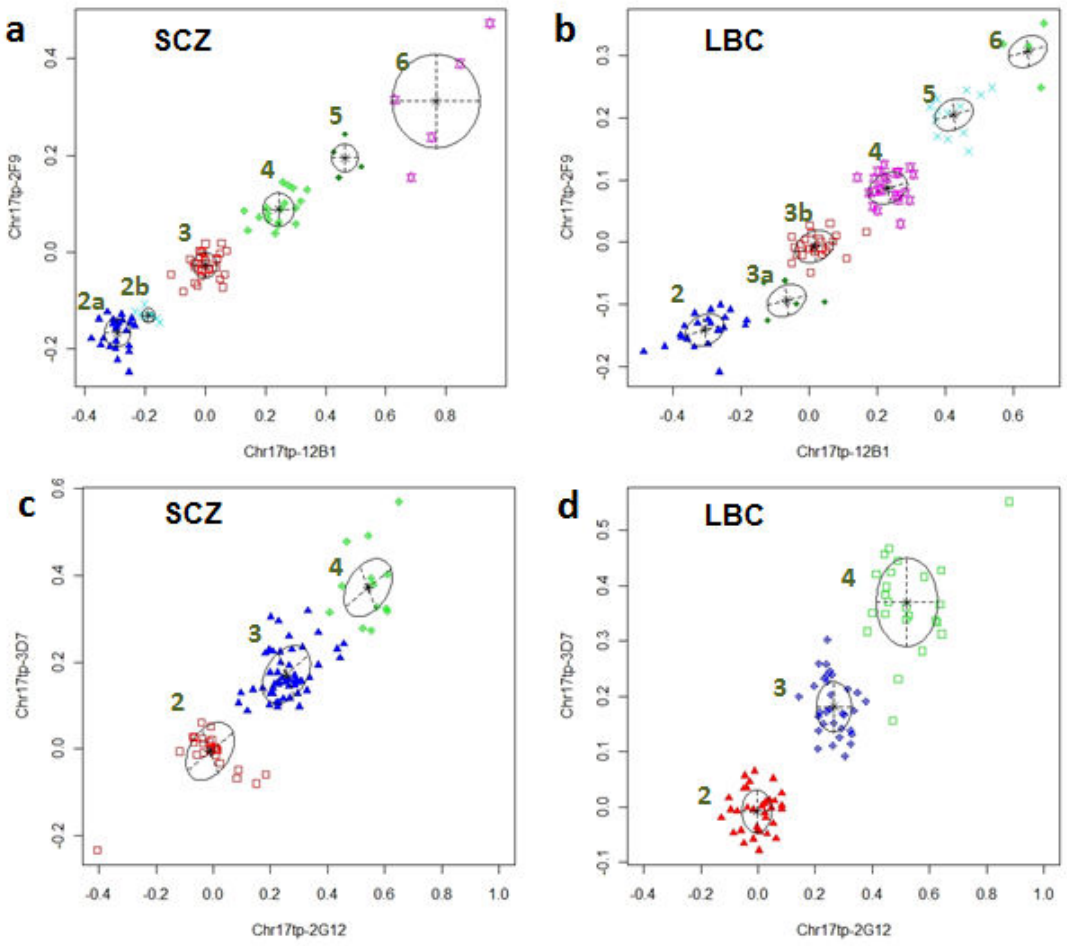

Figure 6.10 Comparing SCZ and LBC samples at $CNV_{KIAA1267}$ and CNV_{NSF}.
(a) & (b) CNV_{NSF} was detected as 2 to 6 copies in SCZ and LBC samples*.
(c) & **(d)** $CNV_{KIAA1267}$ was detected as 2 to 4 copies in SCZ and LBC samples**.
* Bivariate clustering of log2ratio from *Chr17tp-14B1* (x-axis) and *Chr17tp-2F9* (y-axis)
** Bivariate clustering of log2ratio from *Chr17tp-2G12* (x-axis) and *Chr17tp-3D7* (y-axis)

Table 6.3 SCZ and LBC genotype counts at $CNV_{KIAA1267}$ and CNV_{NSF}.

CNV-NSF Copy Number	SCZ	LBC	CNV-KIAA1267 Copy Number	SCZ	LBC
2	32	22	2	25	34
3	31	30	3	49	32
4	18	22	4	13	24
5+	9	17	5+	0	0
Total	90	91	Total	87	90

6.3.8 Evolutionary History of CNVNSF and CNVKIAA1267

To elucidate the relationships between CNVs and the more ancient segmental duplications in the 17q21 region, we classified segmental duplicons into ancestral or derived loci according to in silico predictions from Jiang *et al.* (Jiang et al. 2007) (Figure 6.10). Both CNV_{NSF} and $CNV_{KIAA1267}$ were relatively recent events, flanked by more ancient segmental duplicons (ancestral loci). Genomic regions near ancient segmental duplicons were suggested to be genetically more fragile, favoring evolution of the younger polymorphic CNVs in the complex region (Jiang et al. 2007). These ancestral SDs probably acted as predisposition factors for structural rearrangements in the region, leading to the formation of CNV_{NSF} and $CNV_{KIAA1267}$.

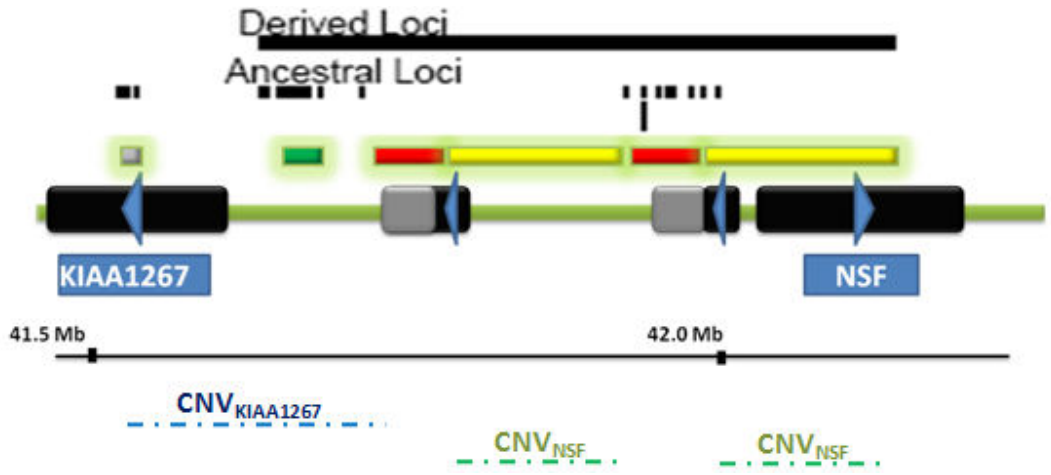

Figure 6.11 Locations of derived and ancestral loci of segmental duplications at 17q21. Locations of the two types of segmental duplications were depicted with respect to $CNV_{KIAA1267}$ and CNV_{NSF}.

6.4 Chapter Summary and Discussion

In this chapter we described a biological pathway approach to studying normal copy number variations in genes relevant to cognition and schizophrenia. Using NRC/MASC as a complex central to neurological functions, we assessed whether such components could be involved in copy number variations.

In the first part of this chapter we described 20 MASC CNVs detected in 269 apparently normal HapMap individuals. Four of these overlapped with core MASC components. Three were confirmed by high-resolution oligonucleotide array data. The presence of CNVs at such important synaptic proteins in healthy individuals suggests that (i) CNVs at these genes are benign and did not translate into pathogenic gene function (e.g. intronic CNVs may have no functional consequence); or (ii) neurological functions have high plasticity such that disrupting functions of one synaptic component could be compensated by other mechanisms in the network; or (iii) such CNVs did affect neurological functions, but the effect was not detrimental, perhaps contributing to variations in brain function, e.g. cognitive ability, among the population. The last hypothesis is particularly interesting for the genetics of neuroscience, and future studies investigating CNVs in human cognitive and behavior traits may reveal its validity.

On the other hand, investigating CNVs in such an important set of synaptic molecules provides us a baseline to evaluate existing and upcoming CNV screens of neurological diseases. A number of psychiatric disease CNV studies have reported variants at neuronal complexes as potential CNV candidates. A deletion at *DLG2*, for instance, was detected as a patient CNV in one genome-wide schizophrenia screen (Walsh et al. 2008). However, CNVs in the same gene were also identified in control samples in another schizophrenia study (ISC 2008), and in our analysis of normal HapMap

individuals. It is not surprising that disease-causing CNVs might exist in healthy controls, perhaps at lower frequencies, at slightly varied genomic locations, or with combinations of different interacting factors. Alternatively, such CNVs may be benign in both patients and controls. It is therefore important to characterize CNVs at key neuronal complexes in normal healthy individuals as a basis for disease comparison.

In the second part of the chapter, we described copy number variations near the *NSF* locus in HapMap samples from three ethnic groups. We revealed two major blocks of copy number variation, CNV_{NSF} and $CNV_{KIAA1267,}$ and further characterized them using quantitative real-time PCR, Fiber-FISH and oligo array CGH data. Interestingly, both CNVs showed population bias across the three ethnic groups examined. In particular, we revealed a H1 variant specific to the CEU samples, which is analogous to the European-specific H2 inverted allele. This H1 variant could represent an evolutionary remnant of the H1/H2 inversion, or a more recent structural rearrangement specific to the European population.