

ARTICLE

# Refining genome-wide linkage intervals using a meta-analysis of genome-wide association studies identifies loci influencing personality dimensions

Najaf Amin<sup>1,18</sup>, Jouke-Jan Hottenga<sup>2,18</sup>, Narelle K Hansell<sup>3,18</sup>, A Cecile JW Janssens<sup>4</sup>, Marleen HM de Moor<sup>2</sup>, Pamela AF Madden<sup>5</sup>, Irina V Zorkoltseva<sup>6</sup>, Brenda W Penninx<sup>7,8,9</sup>, Antonio Terracciano<sup>10</sup>, Manuela Uda<sup>11</sup>, Toshiko Tanaka<sup>10</sup>, Tonu Esko<sup>12</sup>, Anu Realo<sup>12</sup>, Luigi Ferrucci<sup>10</sup>, Michelle Luciano<sup>13</sup>, Gail Davies<sup>13</sup>, Andres Metspalu<sup>12</sup>, Goncalo R Abecasis<sup>14</sup>, Ian J Deary<sup>13</sup>, Katri Raikonen<sup>15</sup>, Laura J Bierut<sup>5</sup>, Paul T Costa<sup>10</sup>, Viatcheslav Saviouk<sup>2</sup>, Gu Zhu<sup>3</sup>, Anatoly V Kirichenko<sup>6</sup>, Aaron Isaacs<sup>1</sup>, Yurii S Aulchenko<sup>1</sup>, Gonneke Willemssen<sup>2</sup>, Andrew C Heath<sup>5</sup>, Michele L Pergadia<sup>5</sup>, Sarah E Medland<sup>3</sup>, Tatiana I Axenovich<sup>6</sup>, Eco de Geus<sup>2</sup>, Grant W Montgomery<sup>3</sup>, Margaret J Wright<sup>3</sup>, Ben A Oostra<sup>1,16</sup>, Nicholas G Martin<sup>3,18</sup>, Dorret I Boomsma<sup>2,18</sup> and Cornelia M van Duijn<sup>\*,1,17,18</sup>

Personality traits are complex phenotypes related to psychosomatic health. Individually, various gene finding methods have not achieved much success in finding genetic variants associated with personality traits. We performed a meta-analysis of four genome-wide linkage scans ( $N = 6149$  subjects) of five basic personality traits assessed with the NEO Five-Factor Inventory. We compared the significant regions from the meta-analysis of linkage scans with the results of a meta-analysis of genome-wide association studies (GWAS) ( $N \sim 17\,000$ ). We found significant evidence of linkage of neuroticism to chromosome 3p14 (rs1490265,  $LOD = 4.67$ ) and to chromosome 19q13 (rs628604,  $LOD = 3.55$ ); of extraversion to 14q32 (ATGG002,  $LOD = 3.3$ ); and of agreeableness to 3p25 (rs709160,  $LOD = 3.67$ ) and to two adjacent regions on chromosome 15, including 15q13 (rs970408,  $LOD = 4.07$ ) and 15q14 (rs1055356,  $LOD = 3.52$ ) in the individual scans. In the meta-analysis, we found strong evidence of linkage of extraversion to 4q34, 9q34, 10q24 and 11q22, openness to 2p25, 3q26, 9p21, 11q24, 15q26 and 19q13 and agreeableness to 4q34 and 19p13. Significant evidence of association in the GWAS was detected between openness and rs677035 at 11q24 ( $P$ -value =  $2.6 \times 10^{-06}$ , *KCNJ1*). The findings of our linkage meta-analysis and those of the GWAS suggest that 11q24 is a susceptible locus for openness, with *KCNJ1* as the possible candidate gene.

*European Journal of Human Genetics* advance online publication, 5 December 2012; doi:10.1038/ejhg.2012.263

**Keywords:** personality; *KCNJ1*; NEO; linkage; GSMA

## INTRODUCTION

Human personality is a compound of complex traits that are associated with several psychiatric,<sup>1–16</sup> and somatic disorders.<sup>17–26</sup> Despite high heritability ranging from 33 to 60%, the understanding of the genetic origins of personality trait variation is extremely limited. Where linkage analyses have identified several large regions, very few overlap.<sup>27</sup> Candidate gene studies have their own concerns in that the findings have generally not been replicated.<sup>27</sup> A recent genome-wide association study (GWAS), including up to 18 000 individuals, yielded only a few loci that attained genome-wide significance.<sup>28</sup>

The general success of GWAS in gene discovery and failure to replicate most of the significant linkage peaks for complex traits and diseases shifted interest towards GWAS. Association studies have benefitted from precise estimation of the locus, reproducibility of the findings and the availability of large population-based cohorts, which has led to powerful studies for gene discovery as well as replication using meta-analyses. However, the GWAS have not been very successful in finding genes of complex psychiatric/behavioral traits. The ‘case of the missing heritability’<sup>29</sup> has led to the view that common traits may be driven by relatively rare variants.<sup>30</sup> Despite the fact that rare variants are weakly tagged by the

<sup>1</sup>Unit of Genetic Epidemiology, Department of Epidemiology and Biostatistics, Erasmus University Medical Center, Rotterdam, The Netherlands; <sup>2</sup>Department of Biological Psychology, VU University Amsterdam, Amsterdam, The Netherlands; <sup>3</sup>Queensland Institute of Medical Research, Brisbane, QLD, Australia; <sup>4</sup>Department of Epidemiology, Erasmus University Medical Center, Rotterdam, The Netherlands; <sup>5</sup>Department of Psychiatry, Washington University School of Medicine, St Louis, MO, USA; <sup>6</sup>Institute of Cytology and Genetics, Russian Academy of Science, Novosibirsk, Russia; <sup>7</sup>Department of Psychiatry, University Medical Center Groningen, Groningen, The Netherlands; <sup>8</sup>Departments of Clinical Psychology and Psychiatry, Leiden University, Leiden, The Netherlands; <sup>9</sup>Department of Psychiatry, EMGO+ Institute, Neuroscience Campus Amsterdam, VU University Medical Center Amsterdam, Amsterdam, The Netherlands; <sup>10</sup>National Institute on Aging, NIH, Baltimore, MD, USA; <sup>11</sup>Istituto di Neurogenetica e Neurofarmacologia, CNR, Monserrato, Cagliari, Italy; <sup>12</sup>Estonian Genome Project, University of Tartu and Estonian Biocentre, Tartu, Estonia; <sup>13</sup>Department of Psychology, Centre for Cognitive Ageing and Cognitive Epidemiology, The University of Edinburgh, Edinburgh, UK; <sup>14</sup>Department of Biostatistics, Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA; <sup>15</sup>Department of Psychology, University of Helsinki, Helsinki, Finland; <sup>16</sup>Department of Clinical Genetics, Erasmus University Medical Center, Rotterdam, The Netherlands; <sup>17</sup>Centre of Medical Systems Biology, Leiden, Netherlands Consortium on Health Aging and National Genomics Initiative, the Hague, The Netherlands

<sup>18</sup>These authors contributed equally to this work.

\*Correspondence: Professor CM van Duijn, Unit of Genetic Epidemiology, Department of Epidemiology and Biostatistics, Erasmus Medical Center, Drive. Molewaterplein 50, PO Box 1738, Rotterdam 3015 GE, The Netherlands. Tel: +31 10 7043394; Fax: +31 10 7044675; E-mail: c.vanduijn@erasmusmc.nl

Received 5 September 2011; revised 21 September 2012; accepted 26 October 2012

common single-nucleotide polymorphisms (SNPs) present in the current arrays,<sup>30,31</sup> association signals in GWAS at loci including rare variants have been seen in particular for lipids.<sup>29</sup> Also, two recent papers argued that common variants may tag more rare ones from a theoretical perspective,<sup>30,31</sup> implying that linkage and association analyses should be able to identify the same loci. For complex traits, it was suggested to use genome-wide linkage and association analyses in two steps to maximize power.<sup>32</sup> Linkage may point towards the regions in the GWAS to look for possible associated variants, reducing the number of tests and relaxing the significance threshold, thus resulting in a higher chance of finding genetic variants.

In this study, we aimed to discover genetic variants, both rare and common, that affect personality traits, including neuroticism, extraversion, openness, agreeableness and conscientiousness. We first perform a quantitative trait linkage analysis in four independent cohorts – assuming that this would pick signals at loci containing large effects with relatively rare (population-specific) variants, and then combine these in a meta-analysis – assuming that the meta-analysis would pick linkage signals at loci harboring genes with relatively moderate to small effects. Finally, we compare the findings of our meta-analysis of linkage scans with those of the largest meta-analysis of GWAS of the same traits (which included the linkage samples examined)<sup>33</sup> to localize the possible linked gene. This is by far the largest linkage study conducted for the NEO personality traits combined with the largest GWAS of these traits conducted to date.

## MATERIALS AND METHODS

We performed a meta-analysis of linkage studies that used the NEO Five Factor Inventory (NEO-FFI) or the Revised NEO personality inventory (NEO-PI-R)<sup>34</sup> to assess the five basic personality traits including: (1) neuroticism – a trait that refers to the tendency to experience negative emotions; (2) extraversion – a measure of sociability, positive emotions and action; (3) openness – a measure of intellectual curiosity and preference for variety; (4) agreeableness – a measure of altruism, cooperation and harmony; and (5) conscientiousness – a measure of an individual's tendency to plan, organize and direct his impulses.<sup>34</sup> This five-factor model is a hierarchical model where each of these five traits is defined by six underlying facets assessed by the NEO-PI-R. These five traits, also known as the Big Five, are considered universal,<sup>35</sup> stable in adulthood<sup>36</sup> and orthogonal, but correlations appear possibly due to self-report.<sup>37</sup> Women score generally higher on neuroticism and agreeableness<sup>38–40</sup> compared with men. The NEO-FFI consists of 60 items, 12 for each trait, whereas the NEO-PI-R consists of 240 questions inclusive of the ones in the NEO-FFI. This study included linkage scans from four independent populations, including the Erasmus Rucphen Family (ERF), the Netherlands Twin Register (NTR), the Australian Adolescent sample (QIMR\_adolescent) and the Australian adult sample (QIMR\_adult) (Table 1). ERF and NTR used the NEO-FFI for personality assessment, whereas the Australian samples were assessed with either NEO-FFI or NEO-PI-R; however, for this study the 60 items of the NEO-FFI was used for the final assessment. Within each cohort the scores were considered invalid if an individual answered fewer than nine questions, otherwise the missing data were imputed by taking the individual's average score for the valid items of that dimension. The descriptive statistics of the study samples are provided in the Supplementary Table 1. Variance component (VC) linkage analysis, adjusted for age and gender, was performed in each cohort with MERLIN (Center for Statistical Genetics, University of Michigan, Ann Arbor, MI, USA) (Table 1). A power analysis for a quantitative trait locus (QTL) explaining 1, 5 and 10% of the trait variance was also performed in each cohort with POLY (Chen and Abecasis)<sup>41</sup> at fixed type 1 error rates of 1 and 5%.

### Erasmus Rucphen Family

The study sample consisted of 2657 individuals who participated in ERF study.<sup>42</sup> The study population essentially consists of one extended family spanning over 23 generations and including more than 23 000 individuals descending from 20 related couples who lived in the region between 1850 and

1900. All descendants were ascertained and descendants of 18 years and older were invited to participate. Spouses were invited only for family members who had children of 18 years and older.

For all participants, genomic DNA was extracted from peripheral venous blood utilizing the salting out method.<sup>43</sup> For genome-wide linkage analysis, genotyping was performed using the Illumina 6K linkage panel that includes markers distributed evenly across the human genome (median distance between markers is 301 kb). Of the 6000 single-nucleotide polymorphisms (SNPs), we used 5250, after quality control and excluding X-chromosomal SNPs. The genotyping was performed at the Center National de Génotypage in France according to the standard protocol.

Multipoint VC linkage analysis was performed using MERLIN v.1.0.1 software<sup>44</sup> for all the five personality traits. Marker allele frequencies were estimated from the data. The pedigrees were split on non-overlapping fragments of no more than 18 bits with the help of two programs: Jenti<sup>45</sup> and PedSTR.<sup>46</sup> Three sets of subpedigrees were obtained with different parameters with the help of these programs, which were then analyzed separately. These three sets differed one from another by number, size and structure of pedigree fragments. However, they demonstrated similar profiles of LODs for all analyzed traits (Spearman's correlations varied from 0.6 ( $P < 0.001$ ) to 0.8 ( $P < 0.001$ )), which allowed us to use the maximum of three values of LOD for each marker locus.<sup>47</sup> In accordance with Bonferroni correction, suggestive and significant thresholds were estimated as 2.34 and 3.75, respectively. The analysis was based on 2244 genotyped and phenotyped persons from ERF.

### Netherlands Twin Register

The NTR<sup>48</sup> sample consists of 711 families with 3412 non-clone individuals (1438 founders, 1870 females) with an average of 4.8 subjects per family. In all, 282 of these families have both founders genotyped and 138 families had one genotyped founder. In addition, there are 290 nuclear families with no genotyped founders, and one extended pedigree with four founders without genotypes. Autosomal genomes had 757 markers spaced at an average of 4.76 cM (range 0.0–20.59 cM), with average heterozygosity of 0.76. Founders had the genetic data for 446 autosomal microsatellites. NEO measures together with the age and gender information were available for a total of 1507 subjects (998 non-founder; 509 non-founders) from 409 families with genetic data. Of 409 families, 270 had two phenotyped siblings, 113 had three, 19 had four, 1 family had five, 4 families had six and 2 families had seven phenotyped siblings, resulting in a total of 835 sibling pairs. In addition, this sample also features 83 phenotyped MZ clones.

The genetic maps were obtained through the Rutgers University Map Interpolator.<sup>49</sup> The allele frequencies were estimated with the Mendel v.10.0 software package (Kenneth Lange, University of California, CA, USA).<sup>50</sup> VC linkage scan of the autosomal genome was conducted with MERLIN v.1.1.2.

### Queensland Institute of Medical Research, Australian Study Sample

NEO personality data (NEO-PI-R or NEO-FFI) were collected as part of two independent research streams – one focused on an adult sample (QIMR\_adult) and the other with an adolescent/young adult focus (QIMR\_adolescent). The QIMR\_adult data were collected as part of the Nicotine Addiction Genetics (NAG) study (2001–2005), which targeted families based on heavy smoking index cases identified in earlier interviews and questionnaires<sup>51</sup> and was itself part of the Interactive Research Project Grants (IRPG). This sample comprised 1349 genotyped individuals aged 21 to 85 years ( $M = 45.5 \pm 13.1$ ) from 519 families, and included 15 complete MZ pairs for whom data were averaged.

The QIMR\_adolescent data were collected from two population studies under the umbrella of the Brisbane Adolescent and Young Adult Twin Studies, specifically, studies of cognition (1996–ongoing)<sup>52</sup> and health and well-being (2002–2003),<sup>52</sup> and from a study of borderline personality disorder (2004–2006).<sup>53</sup> This sample comprised 1096 genotyped individuals aged 16 to 27 years ( $M = 19.4 \pm 2.7$ ), from 563 families, and included 127 complete MZ pairs for whom data were averaged.

Participants were typically Caucasian, predominantly Anglo-Celtic (ancestry outliers, identified using HapMap3 and GenomeEUtwin individuals as a reference panel, were excluded). Written, informed consent was obtained from all participants and from a parent or guardian for those aged under 18 years.

**Table 1** General features of independent genome scan

Cohort	ERF	NTR	QIMR_adolescents	QIMR_adults
Origin	Dutch	Dutch	Australian <sup>a</sup>	Australian <sup>a</sup>
Design	Extended families	sibpairs	sibpairs	sibpairs + extended families
Sample size (%women)	2244 (55.0)	1507 (61.6)	1096 (56.5)	1349 (55.7)
Total no. of markers used	5250	864	5479	5479
Marker type	Single nucleotide	Microsatellites	Single-nucleotide polymorphisms	Single-nucleotide polymorphisms
Analysis type	Variance components	Variance components	Variance components	Variance components
Analysis software	Merlin	Merlin	Merlin	Merlin
Personality assessment	NEO-FFI	NEO-FFI	NEO-PI-R/NEO-FFI	NEO-FFI

<sup>a</sup>Majority of the sample is Causasian, predominantly Anglo-Celtic (ancestry outliers, identified using HapMap3 and GenomeEUTwin individuals as a reference panel, were excluded).

Ethics approval was received from the institutional review boards appropriate to each study (QIMR and Washington University School of Medicine).

Genotyping was carried out using the Illumina (San Diego, CA, USA) 610K or 370K SNP platform and Illumina BeadStudio software, with 269 840 SNPs common to the subsamples passing QC (28% of the SNPs selected for linkage were from this set of SNPs).<sup>54</sup> Data were imputed using HapMap I + II (CEU, build 36 release 22) using the MACH software (Center for Statistical Genetics, Ann Arbor, MI, USA).<sup>54</sup> SNP selection for linkage analysis was matched as far as possible to the SNP set used with the ERF sample. The final selection contained 5479 SNPs, of which 5181 had a direct match with our genotyped or imputed data. For the remaining 298 SNPs, proxies based on linkage disequilibrium (>0.8), or position, were used. Multipoint VC linkage analysis was performed using MERLIN 1.1.2<sup>44</sup> in both samples.

### Meta-analysis of linkage scans

Results from individual genome-wide linkage scans were combined together in a meta-analysis using the Genome Search Meta-Analysis (GSMA).<sup>55,56</sup> The GSMA method divides the genome into several bins of equal width (chosen arbitrarily such that smallest chromosome should have at least two bins), and is robust to differences between studies in ascertainment schemes, marker maps and statistical methods used to detect linkage. Customarily a bin width of 30 cM, which gives 120 bins across the autosomal genome, is used. The bins are then ranked in descending order in each study, with the bin with the highest LOD score getting the highest rank. The ranks within a bin are summed across each study to get a summed rank (SR). The SR statistic is tested for significance using its distribution function or by simulation,<sup>55</sup> which gives the probability of observing a given SR for a bin ( $P_{SR}$ ).  $P_{SR}$  only gives the point-wise probability for the SR for a certain bin. A genome-wide interpretation of the results is obtained through the ordered statistic (OR), which determines the probability ( $P_{OR}$ ) of a given SR for a bin by chance when bins are assigned ranks randomly in multiple simulations.<sup>57</sup> Simulations show that a bin with a significant  $P_{SR}$  and a significant  $P_{OR}$  ( $P_{SR} < 0.05$  and  $P_{OR} < 0.05$ ) has a high probability of containing a true susceptibility locus.<sup>57</sup> For an individual bin, the genome-wide significance is defined as  $P_{SR} < 0.05/\text{number of bins}$  and suggestive as  $P_{SR} < 1/\text{number of bins}$ . For an individual scan, we used 3.3 as significant and 1.9 as suggestive linkage threshold.<sup>58</sup>

In this study, except for the NTR sample genome-wide linkage results were reported against SNP markers, which helped us map all the results directly to the base pair positions. For NTR, we mapped all the results to base pair positions by interpolation and using base pair positions from Rutgers map.<sup>49</sup> For each study, we divided the autosomal genome into 125 bins of width 25 million base pairs (mbp) each, giving a maximum of 10 bins on chromosome 1 and a minimum of 2 bins on chromosome 22. To evaluate the possibility of correlation between adjacent bins, we also performed the analysis using a bin width of 40 mbp. We performed weighted meta-analysis, where weights were calculated as the square root of the sample size in each study. A total of 10 000 permutations were performed to get the  $P_{SR}$  and  $P_{OR}$ . A bin was considered significant if (1) either the Bonferroni-corrected significance was achieved, that is,  $P_{SR} < 0.0004$  for a bin or (2) if both  $P_{SR}$  and  $P_{OR}$  were nominally significant for multiple bins.<sup>57</sup> Suggestive linkage threshold was set at  $P_{SR} < 0.008$ .

Heterogeneity testing was performed using the HEGESMA<sup>59</sup> software (School of Medicine, University of Thessaly, Thessaly, Greece).

In an attempt to discover the variant that might be causing the linkage signals, we fine mapped the interesting regions from the meta-analysis of linkage scans with the results of a meta-analysis of GWAS of the NEO personality traits.<sup>28</sup> The meta-analysis of GWAS ( $n > 17\,000$ ) included the samples used in this study (and many others as well), but the marker sets used were different. Brief description of the studies included in the meta-analysis of GWAS is provided in Supplementary Table 2.

## RESULTS

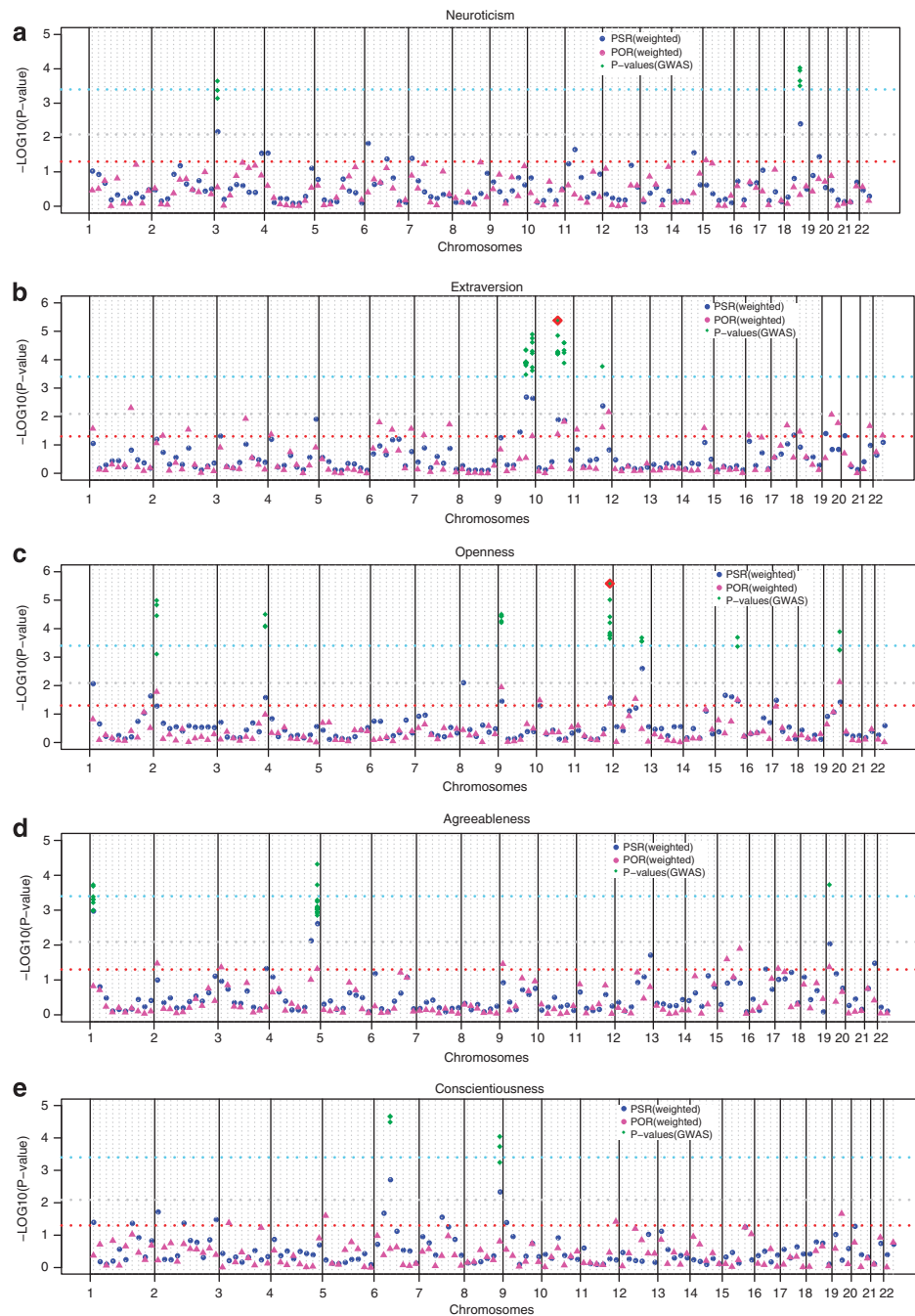
### Individual scans

Results of the power analysis are provided in the Supplementary Table 3. The results from individual genome-wide linkage scans are provided in Supplementary Figures 1–5 and significant and suggestive findings are summarized in Supplementary Table 4. Significant evidence of linkage was observed for neuroticism in the ERF study at chromosome 3p14 (rs1490265, LOD = 4.67) and at chromosome 19q13 (rs628604, LOD = 3.55) in the QIMR (adolescent) sample; for extraversion at 14q32 (ATGG002, LOD = 3.3) in the NTR sample; for agreeableness at 3p25 (rs709160, LOD = 3.67) and two adjacent regions on chromosome 15, including 15q13 (rs970408, LOD = 4.07) and 15q14 (rs1055356, LOD = 3.52) in the QIMR (adolescents) study. Considering the suggestive findings, there was an overlap at chromosome 2q14 between ERF (LOD = 2.1) and NTR (LOD = 2.03) for neuroticism and at 12q23 between ERF (LOD = 2.85) and NTR (LOD = 1.96) for openness (Supplementary Table 4).

### Meta-analysis

Genome-wide results of the meta-analysis are illustrated in Figure 1. Table 2 provides a summary of the bins with significant evidence of linkage. None of the bins crossed the Bonferroni-corrected genome-wide significance threshold. However, there were multiple bins for which both  $P_{SR}$  and  $P_{OR}$  were nominally significant (Table 2 and Figure 1).

For extraversion, five bins showed nominally significant  $P_{SR}$  and  $P_{OR}$ . These included bins 9.6, 11.5 and two adjacent bins on chromosome 10 (10.4 and 10.5). The linkage signals on the adjacent bins on chromosome 10 were being caused by the same peaks that extended over 40 cM (Supplementary Figure S2). When the bin width was increased to 40 mbp, the finding for the new bin on chromosome 10 covering the previous 10.4 bin and partially covering the 10.5 bin remained significant and the adjacent bin did not show linkage signals. When comparing the significant bins with the meta-analysis of GWAS, we identified clusters of SNPs with low  $P$ -values for the bin in chromosome 10 (rs7088779,  $P$ -value =  $4.2 \times 10^{-06}$ ) (Table 2 and Figure 1b). Rs7088779 was marginally significant after being corrected



**Figure 1** Results of the meta-analysis of the linkage scans. (a) For neuroticism, (b) for extraversion, (c) for openness, (d) for agreeableness and (e) for conscientiousness. Dots represent  $P_{SR}$  and triangles represent  $P_{OR}$ . The X axis shows the whole autosomal genome divided by solid vertical lines into chromosomes, which are further divided into bins by dotted gray lines. The Y axis shows the negative principal log of the  $P$ -values. The red horizontal dotted line represents the nominal threshold  $P$ -value=0.5, the gray dotted horizontal line represents the suggestive threshold  $P$ -value=0.008 and the light blue horizontal dotted line represents the Bonferroni-corrected threshold  $P$ -value=0.0004. A bin is considered significant if a circle for that bin surpasses the sky blue line or if both the circle and the triangle for a specific bin are above the dotted red line. The green diamonds represent the  $P$ -values of the SNPs from the results of the meta-analysis of the GWAS of the same traits falling in the bins of interest. The green diamonds with the red outline are the  $P$ -values significant/borderline significant after correcting for the number of total SNPs in the bin.

for the number of SNPs in the bin. This SNP is located between *CRTAC1* (cartilage acidic protein 1) and *C10orf28*, a region previously implicated in Alzheimer's disease.

For openness, five bins (9.1, 11.6, 15.4 and 19.3) were significant in that they showed both  $P_{SR}$  and  $P_{OR}$  < 0.05 (Table 2 and Figure 1c). A cluster of eight SNPs with very low  $P$ -values was identified from the

GWAS for the bin 11.6 (Figure 1c), which maps to 11q24 region. The most significant SNP (rs677035) showed a  $P$ -value of  $2.6 \times 10^{-06}$ , which passed the Bonferroni threshold. Rs677035 is an intergenic SNP located between *FLI1* and *KCNJ1*.

Two bins (4.8 and 19.1) showed nominally significant  $P_{SR}$  and  $P_{OR}$  for agreeableness, but were not supported by association (Table 2 and



**Table 2 Comparison of interesting regions of linkage meta-analysis with the personality GWAS meta-analysis**

Trait	Bin	Cyto	SR	OR	P-value	PHET	Approx. region (mbp)	SNP GWAS (N > 17 000)	Effect allele	$\beta$	MAF	Effective no. of SNPs in the region <sup>a</sup>	P-value (nominal)	Bonferroni threshold <sup>b</sup>	Candidate gene	Distance of the SNP to the candidate gene (kb) <sup>c</sup>
Extraversion	9.6	9q34	0.003	0.049	0.036	125–150	rs11103814	A	-0.35	0.33	2049	$1.3 \times 10^{-5}$	$4.2 \times 10^{-6}$	OLFMI, KIAA0649	204.8, 153.8	
	10.4	10q24	0.017	0.042	0.063	75–100	rs7922066	A	-0.44	0.18	2986	$4.2 \times 10^{-6}$	$2.1 \times 10^{-6}$	CRTAC1	Intron(NM_001206528.1)	
	10.5	10q24	0.018	0.015	0.042	100–125	rs7905091	T	0.47	0.13	3073	$2.5 \times 10^{-5}$	$2.1 \times 10^{-6}$	LOXL4, PYROXD2	18.1, 97.2	
Openness	11.5	11q22	0.005	0.024	0.073	100–125	rs12419614	A	1.65	0.03	3177	$1.1 \times 10^{-4}$	$2.0 \times 10^{-6}$	GRIA4	Intron(NM_000829.3)	
	2.1	2p25	0.067	0.016	0.186	0–25	rs11685318	A	-0.38	0.20	3056	$1.0 \times 10^{-5}$	$1.8 \times 10^{-6}$	PDIA6, KCNF1	40.4, 58.7	
Agreeableness	9.1	9p21	0.045	0.011	0.053	0–25	rs10966354	T	-0.30	0.37	3865	$3.1 \times 10^{-5}$	$1.4 \times 10^{-6}$	ELAVL2, TUSC1	810.7, 1039.6	
	11.6	11q24	0.034	0.043	0.625	125–150	rs677035	T	-0.56	0.08	1393	<b><math>2.6 \times 10^{-6}</math></b>	<b><math>4.4 \times 10^{-6}</math></b>	FLII, KCNJ1	13.9, 10.9	
	15.4	15q26	0.045	0.031	0.280	75–100	rs8034376	T	-0.47	0.11	3135	$2.0 \times 10^{-4}$	$2.0 \times 10^{-6}$	FAM169B, IGF1R	6.5, 128.6	
Agreeableness	19.3	19q13	0.048	0.007	0.165	50–75	rs2287826	A	-0.62	0.07	2106	$1.2 \times 10^{-4}$	$4.8 \times 10^{-6}$	LILRA4, LAIR1	12.8, 1.9	
	4.8	4q34	0.003	0.048	0.008	175–200	rs13113475	A	0.28	0.20	1959	$4.8 \times 10^{-5}$	$3.1 \times 10^{-6}$	SPCS3, VEGFC	228.8, 122.5	
	19.1	19p13	0.011	0.042	0.134	0–25	rs10413538	T	0.49	0.28	3209	$3.7 \times 10^{-4}$	$3.6 \times 10^{-6}$	ZNF826	Utr-5(NM_001039884.1)	

<sup>a</sup>P-values from GWAS significant after Bonferroni correction; in bold italics.

<sup>b</sup>Effective number of SNPs from the GWAS present in the bin.

<sup>c</sup>Bonferroni threshold based on the total number of SNPs (both genotyped and imputed) from the GWAS present in the bin.

<sup>d</sup>Positions are taken from dbSNP (<http://www.ncbi.nlm.nih.gov/snp>) and SCAN (<http://www.scandb.org/newinterface/about.html>) databases.

Figure 1d). Heterogeneity between studies was detected for bin 4.8. None of the bins for neuroticism or conscientiousness showed significant linkage (Figure 1e).

## DISCUSSION

We found significant evidence of linkage of neuroticism with chromosomes 3p14 (rs1490265, LOD = 4.67) and 19q13 (rs628604, LOD = 3.55); of extraversion with 14q32 (ATGG002, LOD = 3.3); and of agreeableness with 3p25 (rs709160, LOD = 3.67) and with two adjacent regions on chromosome 15, including 15q13 (rs970408, LOD = 4.07) and 15q14 (rs1055356, LOD = 3.52) in the individual scans. In the meta-analysis, we discovered one region on chromosome 11q24 significantly linked to openness, which was also strongly supported by the results of the largest GWAS of personality traits performed to date.

Our meta-analysis included 6149 individuals for multiple extended families and families with sibships. There are a number of methodological issues relevant for the interpretation of the findings. We used a physical map (base pair positions) and bin width of 25 mbp (which would roughly translate to 25 cM) as opposed to the traditional 30 cM bin width used in all previous studies. The choice of 25 mbp was made after taking into account the genetic maps of all four studies and specially the position of last reported markers on chromosome 22 (49 mbp), giving two bins of about equal size on the smallest chromosome as was required, thus avoiding manipulation of data. However, this selection could lead to a correlation between two adjacent bins. Interestingly, this correlation was observed only for extraversion in two pairs of adjacent bins on chromosomes 9 and 10. The size of the linkage peak on chromosome 9 extends to about 50 cM (Supplementary Figure S2), implying that choice of a bin width of more than 30 cM could not have removed this correlation. For chromosome 10, the significance remained even after increasing the bin width to 40 mbp, which suggests that our result on 10q24 for extraversion is consistent. The GSMA gave broad linkage regions spread over 25 mbp, but we made an effort to localize the susceptibility genes by using additional information from the meta-analysis of GWAS.

It is interesting to note that none of the significant regions from the individual scans showed any evidence of linkage in the meta-analysis. There may be several explanations: first, our significant findings of individual studies may just be false positives. This may be a possibility, but it is difficult to believe that all significant findings are false, as VC linkage analysis is usually robust in detecting linkage.<sup>60</sup> Second, this may be due to the differences in the power of the various studies (see Supplementary Table 3). For example, the reason why the linkage of neuroticism to 3p14 was detected in the ERF study only may be the size of the study, which was twice as large as any other included study. Third and more likely, is possible locus heterogeneity across populations. High locus heterogeneity, which results in inconsistent linkage peaks, is one characteristic of complex traits like personality. However, the heterogeneity analysis did not provide evidence of heterogeneity at this locus. Also, the rank-based test used for the meta-analysis of linkage scans is insensitive to the significance of a linkage peak in an individual study and is more adept at finding subtle linkage peaks present in all studies included in the meta-analysis. For instance, the meta-analysis ignores the overlap at chromosome 2q14 between ERF (LOD = 2.01) and NTR (LOD = 2.03) for neuroticism despite falling in the same bin. Similarly, two adjacent regions for openness at chromosome 11q25 for ERF (LOD = 2.05) and 11q24 for QIMR\_adolescent (LOD = 2.55) and chromosome 12q23 for ERF (LOD = 2.85) and 12q24 for NTR

(LOD = 1.96) that showed suggestive evidence of linkage were not picked up in the meta-analysis. Such results were also observed in studies that used parametric methods (Fisher's method) to meta-analyze genome-wide linkage scans.<sup>61</sup>

The most interesting finding of the meta-analysis is 11q24 that showed nominally significant  $P_{SR}$  and  $P_{OR}$  and also significant association in the GWAS after adjusting for multiple testing. The 11q24 region is implicated in mental retardation<sup>62</sup> and migraine.<sup>63</sup> The significant SNP in the GWAS rs677035 is located between *FLII* and *KCNJ1*. Previously, we have found linkage to other potassium channel genes including *KCNJ2*, *KCNJ6* and *KCNJ16*,<sup>27</sup> making *KCNJ1* the most interesting gene in the region. The other regions of interest include 10q24 for extraversion. The region showed both  $P_{SR}$  and  $P_{OR} < 0.05$ , and additionally showed borderline significant evidence of association in the meta-analysis of GWAS after adjusting for multiple testing. The region has been implicated in spinocerebellar ataxia,<sup>64</sup> AD and cognitive function.<sup>65,66</sup>

We cannot discard other regions that showed a high probability of containing true susceptibility loci in the linkage analyses as being false positives based on insignificant association results of the GWAS. Our multiple testing corrections for association within a linkage region were based on total number of SNPs in the region (rather than independent SNPs), which made the Bonferroni threshold very conservative. Secondly, the power of the linkage analysis and association analysis to detect a QTL depends on the effect size of the QTL and the linkage disequilibrium between the candidate loci,<sup>32</sup> as the former is more powerful to detect relatively rare large effect loci, while the latter is more powerful for finding frequent loci with subtle effects. The power of association analysis to detect a QTL declines rapidly with the decreasing degree of linkage disequilibrium between the QTL and the candidate locus.<sup>32</sup> It is also possible that because of arbitrary binning of the linkage peaks for the meta-analysis, which is not very precise, we might have completely missed the association signal.

The regions we have discovered with linkage are good candidates for sequencing exomes and regulatory regions and may unveil variants with moderate to large effects contributing to the make-up of human personality. The criteria for significant evidence for linkage in the meta-analyses were not reached in this study. The highest level of significance for linkage reached, translated into a high probability (of unknown size) to include a susceptibility locus. It is obvious that false negatives may occur because of locus heterogeneity using the current methods. To overcome these problems, we combined the data from the linkage analyses of the five basic NEO personality traits in four independent populations with those of the genome wide association analyses. Here we relied on classical association analysis and genetic imputations based on HapMap. This approach will be strengthened by imputing populations that underwent GWAS using the 1000 genomes, providing a greater resolution and a better coverage of rare variants.

## CONFLICT OF INTEREST

The other authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

**ERF:** The genotyping for the ERF study was supported by EUROSPAN (European Special Populations Research Network) and the European Commission FP6 STRP Grant (018947; LSHG-CT-2006-01947). The ERF study was further supported by grants from the Netherlands Organization for Scientific Research, Erasmus MC, the Center for Medical Systems Biology (CMSB) and the Netherlands Brain Foundation (HersenStichting Nederland).

We are grateful to all patients and their relatives, general practitioners and neurologists for their contributions, and to P Veraart for her help in genealogy, Jeannette Vergeer for the supervision of the laboratory work and P Snijders for his help in data collection.

**NTR:** We acknowledge financial support from the Netherlands Organization for Scientific Research (NWO): Grants 575-25-006, 480-04-004, 904-61-090; 904-61-193, 400-05-717 and Spinozapremie SPI 56-464-14192. MHMdeM is financially supported by ZonMW (Addiction) Grant number 311-60008. We further acknowledge financial support from the Center for Medical Systems Biology (NWO Genomics), Neuroscience Campus Amsterdam; EU/QLRT-2001-01254; NIMH R01 MH059160; Geestkracht program of ZonMW (10-000-1002); and matching funds from universities and mental health care institutes involved in NESDA. Genotyping was funded by the Genetic Association Information Network (GAIN) of the Foundation for the US National Institutes of Health, and analysis was supported by grants from GAIN and the NIMH (MH081802). Genotype data were obtained from dbGaP (<http://www.ncbi.nlm.nih.gov/dbgap>, accession number phs000020.v1.p1).

**QIMR:** We thank Marlene Grace and Ann Eldridge for sample collection; Anjali Henders, Megan Campbell, Lisa Bowdler, Steven Crooks and staff of the Molecular Epidemiology Laboratory for sample processing and preparation; Harry Beeby, David Smyth and Daniel Park for IT support. We acknowledge support from the Australian Research Council (A7960034, A79906588, A79801419, DP0212016 and DP0343921), Beyond Blue and the Borderline Personality Disorder Research Foundation. Genotyping was funded by the National Health and Medical Research Council (Medical Bioinformatics Genomics Proteomics Program, 389891). Further, we gratefully acknowledge Drs Dale R Nyholt and especially Scott Gordon for their substantial efforts involving the QC and preparation of the QIMR and NAG/IRPG GWA data sets from which SNPs were selected. Dr Nyholt also contributed 8% of the NAG/IRPG GWA cohort (NHMRC IDs 339462, 442981, 389938 and 496739). In addition, we thank him for helpful discussions regarding methodology.

**GWAS cohorts:** We also acknowledge the individual contributions of the cohorts included in the meta-analysis of genome-wide association scans. Paul T Costa Jr receives royalties from the NEO-FFI. Laura J Bierut is an inventor on the patent 'Markers for Addiction' (US 20070258898) covering the use of certain SNPs in determining the diagnosis, prognosis and treatment of addiction. Dr Bierut served as a consultant for Pfizer Inc. in 2008.

- 1 Bienvenu OJ, Hettema JM, Neale MC, Prescott CA, Kendler KS: Low extraversion and high neuroticism as indices of genetic and environmental risk for social phobia, agoraphobia, and animal phobia. *Am J Psychiatry* 2007; **164**: 1714–1721.
- 2 Cloninger CR, Svrakic DM, Przybeck TR: A psychological model of temperament and character. *Arch Gen Psychiatry* 1993; **50**: 975–990.
- 3 Cramer D: Type A behaviour pattern, extraversion, neuroticism and psychological distress. *Br J Med Psychol* 1991; **64** (Part 1): 73–83.
- 4 Crowe M, Andel R, Pedersen NL, Fratiglioni L, Gatz M: Personality and risk of cognitive impairment 25 years later. *Psychol Aging* 2006; **21**: 573–580.
- 5 Fanous AH, Neale MC, Aggen SH, Kendler KS: A longitudinal study of personality and major depression in a population-based sample of male twins. *Psychol Med* 2007; **37**: 1163–1172.
- 6 Goodwin RD, Castro M, Kovacs M: Major depression and allergy: does neuroticism explain the relationship? *Psychosom Med* 2006; **68**: 94–98.
- 7 Hettema JM, Neale MC, Myers JM, Prescott CA, Kendler KS: A population-based twin study of the relationship between neuroticism and internalizing disorders. *Am J Psychiatry* 2006; **163**: 857–864.
- 8 Kaasinen V, Nurmi E, Bergman J et al: Personality traits and brain dopaminergic function in Parkinson's disease. *Proc Natl Acad Sci USA* 2001; **98**: 13272–13277.
- 9 McCrae RR, Yang J, Costa Jr PT et al: Personality profiles and the prediction of categorical personality disorders. *J Pers* 2001; **69**: 155–174.
- 10 Menza M: The personality associated with Parkinson's disease. *Curr Psychiatry Rep* 2000; **2**: 421–426.
- 11 Sandahl C, Lindberg S, Bergman H: The relation between drinking habits and neuroticism and weak ego among male and female alcoholic patients. *Acta Psychiatr Scand* 1987; **75**: 500–508.
- 12 Van Os J, Jones PB: Neuroticism as a risk factor for schizophrenia. *Psychol Med* 2001; **31**: 1129–1134.
- 13 Wilson RS, Schneider JA, Arnold SE, Bienias JL, Bennett DA: Conscientiousness and the incidence of Alzheimer disease and mild cognitive impairment. *Arch Gen Psychiatry* 2007; **64**: 1204–1212.

- 14 Rahman MA, Eysenck SB: Psychoticism and response to treatment in neurotic patients. *Behav Res Ther* 1978; **16**: 183–189.
- 15 Meins W, Dammast J: Do personality traits predict the occurrence of Alzheimer's disease? *Int J Geriatr Psychiatry* 2000; **15**: 120–124.
- 16 Distel MA, Trull TJ, Willemssen G *et al*: The five-factor model of personality and borderline personality disorder: a genetic analysis of comorbidity. *Biol Psychiatry* 2009; **66**: 1131–1138.
- 17 Goodwin RD, Stein MB: Peptic ulcer disease and neuroticism in the United States adult population. *Psychother Psychosom* 2003; **72**: 10–15.
- 18 Lichtenstein P, Pedersen NL, Plomin R, Faire UD, McClearn GE: Type A behavior pattern, related personality traits and self-reported coronary heart disease. *Person Individ Diff* 1989; **10**: 419–426.
- 19 Robertson DA, Ray J, Diamond I, Edwards JG: Personality profile and affective state of patients with inflammatory bowel disease. *Gut* 1989; **30**: 623–626.
- 20 Shipley BA, Weiss A, Der G, Taylor MD, Deary IJ: Neuroticism, extraversion, and mortality in the UK Health and Lifestyle Survey: a 21-year prospective cohort study. *Psychosom Med* 2007; **69**: 923–931.
- 21 Winstow DJ, Wakefield JA, Goldsmith WM: The relationship between personality, health symptoms and disease. *Person Individ Diff* 1990; **11**: 717–723.
- 22 Sutin AR, Costa Jr. PT, Uda M, Ferrucci L, Schlessinger D, Terracciano A: Personality and metabolic syndrome. *Age (Dordr)* 2010; **32**: 513–519.
- 23 Terracciano A, Sutin AR, McCrae RR *et al*: Facets of personality linked to underweight and overweight. *Psychosom Med* 2009; **71**: 682–689.
- 24 Sutin AR, Scuteri A, Lakatta EG *et al*: Trait antagonism and the progression of arterial thickening: women with antagonistic traits have similar carotid arterial thickness as men. *Hypertension* 2010; **56**: 617–622.
- 25 Sutin AR, Terracciano A, Deiana B *et al*: Cholesterol, triglycerides, and the Five-Factor Model of personality. *Biol Psychol* 2010; **84**: 186–191.
- 26 Sutin AR, Terracciano A, Deiana B *et al*: High neuroticism and low conscientiousness are associated with interleukin-6. *Psychol Med* 2010; **40**: 1485–1493.
- 27 Amin N, Schuur M, Gusareva ES *et al*: A genome-wide linkage study of individuals with high scores on NEO personality traits. *Mol Psychiatry* 2012; **17**: 1031–1041.
- 28 de Moor MHM, Costa PT, Terracciano A *et al*: Meta-analysis of genome-wide association studies for personality. *Mol Psychiatry* 2012; **17**: 337–349.
- 29 Maher B: Personal genomes: the case of the missing heritability. *Nature* 2008; **456**: 18–21.
- 30 Yang J, Benyamin B, McEvoy BP *et al*: Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* 2010; **42**: 565–569.
- 31 Dickson SP, Wang K, Krantz I, Hakonarson H, Goldstein DB: Rare variants create synthetic genome-wide associations. *PLoS Biol* 2010; **8**: e1000294.
- 32 Sham PC, Cherny SS, Purcell S, Hewitt JK: Power of linkage versus association analysis of quantitative traits, by use of variance-components models, for sibship data. *Am J Hum Genet* 2000; **66**: 1616–1630.
- 33 de Moor MH, Costa PT, Terracciano A *et al*: Meta-analysis of genome-wide association studies for personality. *Mol Psychiatry* 2012; **17**: 337–349.
- 34 Costa PT, McCrae RR: *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five Factor Inventory (NEO-FFI) Professional Manual*. Psychological Assessment Resources, 1992.
- 35 McCrae RR, Terracciano A: Personality Profiles of Cultures P. Universal features of personality traits from the observer's perspective: data from 50 cultures. *J Pers Soc Psychol* 2005; **88**: 547–561.
- 36 Terracciano A, Costa Jr PT, McCrae RR: Personality plasticity after age 30. *Pers Soc Psychol Bull* 2006; **32**: 999–1009.
- 37 Biesanz JC, West SG: Towards understanding assessments of the big five: multitrait-multimethod analyses of convergent and discriminant validity across measurement occasion and type of observer. *J Pers* 2004; **72**: 845–876.
- 38 Costa Jr PT, Terracciano A, McCrae RR: Gender differences in personality traits across cultures: robust and surprising findings. *J Pers Soc Psychol* 2001; **81**: 322–331.
- 39 Budaev SV: *Personality and Individual Differences*. Amsterdam: Elsevier, 1999; Vol. 26, pp 801–813.
- 40 Goodwin RD, Gotlib IH: Gender differences in depression: the role of personality factors. *Psychiatry Res* 2004; **126**: 135–142.
- 41 Chen WM, Abecasis GR: Estimating the power of variance component linkage analysis in large pedigrees. *Genet Epidemiol* 2006; **30**: 471–484.
- 42 Pardo LM, MacKay I, Oostra B, van Duijn CM, Aulchenko YS: The effect of genetic drift in a young genetically isolated population. *Ann Hum Genet* 2005; **69**: 288–295.
- 43 Miller SA, Dykes DD, Polesky HF: A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 1988; **16**: 1215.
- 44 Abecasis GR, Cherny SS, Cookson WO, Cardon LR: Merlin – rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 2002; **30**: 97–101.
- 45 Falchi M, Fuchsberger C: Jenti: an efficient tool for mining complex inbred genealogies. *Bioinformatics* 2008; **24**: 724–726.
- 46 Axenovich TI, Zorkoltseva IV, Liu F, Kirichenko AV, Aulchenko YS: Breaking loops in large complex pedigrees. *Hum Hered* 2008; **65**: 57–65.
- 47 Bellenguez C, Ober C, Bourgain C: A multiple splitting approach to linkage analysis in large pedigrees identifies a linkage to asthma on chromosome 12. *Genet Epidemiol* 2009; **33**: 207–216.
- 48 Boomsma DI, de Geus EJ, Vink JM *et al*: Netherlands Twin Register: from twins to twin families. *Twin Res Hum Genet* 2006; **9**: 849–857.
- 49 Matise TC, Chen F, Chen W *et al*: A second-generation combined linkage physical map of the human genome. *Genome Res* 2007; **17**: 1783–1786.
- 50 Lange K, Cantor R, Horvath S *et al*: Mendel version 4.0: A complete package for the exact genetic analysis of discrete traits in pedigree and population data sets. *Amer J Hum Genetics* 2001; **69**(Suppl): 504.
- 51 Pergadia ML, Agrawal A, Loukola A *et al*: Genetic linkage findings for DSM-IV nicotine withdrawal in two populations. *Am J Med Genet B* 2009; **150B**: 950–959.
- 52 Wright MJ, Martin NG: The Brisbane Adolescent Twin Study: outline of study methods and research projects. *Aust J Psychol* 2004; **56**: 65–78.
- 53 Distel MA, Trull TJ, Derom CA *et al*: Heritability of borderline personality disorder features is similar across three countries. *Psychol Med* 2007; **38**: 1219–1229.
- 54 Medland SE, Nyholt DR, Painter JN *et al*: Common variants in the trichohyalin gene are associated with straight hair in Europeans. *Am J Hum Genet* 2009; **85**: 750–755.
- 55 Pardi F, Levinson DF, Lewis CM: GSMA: software implementation of the genome search meta-analysis method. *Bioinformatics* 2005; **21**: 4430–4431.
- 56 Wise LH, Lanchbury JS, Lewis CM: Meta-analysis of genome searches. *Ann Hum Genet* 1999; **63**: 263–272.
- 57 Levinson DF, Levinson MD, Segurado R, Lewis CM: Genome scan meta-analysis of schizophrenia and bipolar disorder, part 1: methods and power analysis. *Am J Hum Genet* 2003; **73**: 17–33.
- 58 Lander E, Kruglyak L: Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* 1995; **11**: 241–247.
- 59 Zintzaras E, Ioannidis JP: HEGESMA: genome search meta-analysis and heterogeneity testing. *Bioinformatics* 2005; **21**: 3672–3673.
- 60 Shugart YY, O'Connell JR, Wilson AF: An evaluation of the variance components approach: type I error, power and size of the estimated effect. *Eur J Hum Genet* 2002; **10**: 133–136.
- 61 Province MA, Kardia SL, Ranade K *et al*: A meta-analysis of genome-wide linkage scans for hypertension: the National Heart, Lung and Blood Institute Family Blood Pressure Program. *Am J Hypertens* 2003; **16**: 144–147.
- 62 Bhalla K, Luo Y, Buchan T *et al*: Alterations in CDH15 and KIRREL3 in patients with mild to severe intellectual disability. *Am J Hum Genet* 2008; **83**: 703–713.
- 63 Cader ZM, Noble-Topham S, Dymont DA *et al*: Significant linkage to migraine with aura on chromosome 11q24. *Hum Mol Genet* 2003; **12**: 2511–2517.
- 64 Nikali K, Suomalainen A, Terwilliger J, Koskinen T, Weissenbach J, Peltonen L: Random search for shared chromosomal regions in four affected individuals: the assignment of a new hereditary ataxia locus. *Am J Hum Genet* 1995; **56**: 1088–1095.
- 65 Liu F, Arias-Vasquez A, Sleegers K *et al*: A genomewide screen for late-onset Alzheimer disease in a genetically isolated Dutch population. *Am J Hum Genet* 2007; **81**: 17–31.
- 66 Bertram L, Blacker D, Mullin K *et al*: Evidence for genetic linkage of Alzheimer's disease to chromosome 10q. *Science* 2000; **290**: 2302–2303.

Supplementary Information accompanies the paper on European Journal of Human Genetics website (<http://www.nature.com/ejhg>)