

## ORIGINAL ARTICLE

## OTTO: a new strategy to extract mental disease-relevant combinations of GWAS hits from individuals

H Ehrenreich<sup>1,2</sup>, M Mitjans<sup>1</sup>, S Van der Auwera<sup>3</sup>, TP Centeno<sup>4</sup>, M Begemann<sup>1</sup>, HJ Grabe<sup>3</sup>, S Bonn<sup>2,4</sup> and K-A Nave<sup>2,5</sup>

Despite high heritability of schizophrenia, genome-wide association studies (GWAS) have not yet revealed distinct combinations of single-nucleotide polymorphisms (SNPs), relevant for mental disease-related, quantifiable behavioral phenotypes. Here we propose an individual-based model to use genome-wide significant markers for extracting first genetic signatures of such behavioral continua. 'OTTO' (old Germanic = heritage) marks an individual characterized by a prominent phenotype and a particular load of phenotype-associated risk SNPs derived from GWAS that likely contributed to the development of his personal mental illness. This load of risk SNPs is shared by a small squad of 'similar' scattered under the genetically and phenotypically extremely heterogeneous umbrella of a schizophrenia end point diagnosis and to a variable degree also by healthy subjects. In a discovery sample of >1000 deeply phenotyped schizophrenia patients and several independent replication samples, including the general population, a gradual increase in the severity of 'OTTO's phenotype' expression is observed with an increasing share of 'OTTO's risk SNPs', as exemplified here by autistic and affective phenotypes. These data suggest a model in which the genetic contribution to dimensional behavioral traits can be extracted from combinations of GWAS SNPs derived from individuals with prominent phenotypes. Even though still in the 'model phase' owing to a world-wide lack of sufficiently powered, deeply phenotyped replication samples, the OTTO approach constitutes a conceptually novel strategy to delineate biological subcategories of mental diseases starting from GWAS findings and individual subjects.

*Molecular Psychiatry* (2018) **23**, 476–486; doi:10.1038/mp.2016.208; published online 6 December 2016

## INTRODUCTION

Building on the gigantic progress made on genome analyses, more and more genome-wide association studies (GWAS) on mental disease were conducted in recent years, some of them including >100 000 individuals.<sup>1,2</sup> Although the information obtained from these huge endeavors is of fundamental interest for population genetics, it has not been possible yet to elucidate genetic roots of neuropsychiatric disease-relevant behavioral phenotypes. Even though the so far largest GWAS on schizophrenia extracted 108 genome-wide significant loci,<sup>2</sup> their roles for individuals suffering from mental disease have remained largely obscure. The loci in aggregate are not suitable to predict any disease-relevant phenotypes apart from the schizophrenia diagnosis.<sup>3</sup> Although perhaps disappointing, this fact is readily explained by the excessive and multiplex heterogeneity of individuals gathering under the end point diagnoses of highly complex diseases ('mixed bags'; that is, every schizophrenic subject having another combination of genetic and environmental reasons to be schizophrenic), not to mention the clinical diagnostic uncertainty and total lack of objective measures.

Polygenic risk scores, considering in a stepwise manner the less significant single-nucleotide polymorphisms (SNPs) underneath the Bonferroni threshold or complex pathway analyses based on the identified risk genotypes, raised hopes to make more sense of the GWAS finds but the value of these scores turned out to be rather limited for complex disease and their reproducibility often questionable e.g.<sup>4–6</sup> In contrast, by exploring distinct SNPs derived

from the schizophrenia GWAS<sup>2</sup> information, general principles of interest for normal and abnormal brain development, contributing to disease pathophysiology, may be extracted.<sup>7</sup> Indisputably, GWAS methodology can yield interesting results for specific mechanistic insight upon analysis of narrow and well-defined phenotypes.<sup>8–10</sup>

The present paper proposes an entirely different way to use schizophrenia GWAS hits for understanding at least some behavioral aspects of mental illness: an approach beginning with individuals. We named this approach the 'OTTO project', where 'OTTO' (old Germanic = heritage) stands for an individual, male or female, with a particular load of genetic factors (here: GWAS hits) that—together with his personally experienced environmental input—contributed to the development of his discrete mental illness. Ideally, OTTO's illness is characterized by a particularly strong behavioral phenotype, which facilitates the here introduced procedure. We hypothesize that OTTO has some 'similar', that is, individuals who share with him not only the clinical end point diagnosis (schizophrenia) but, most importantly, his set of genetic (GWAS-derived) risk factors and possibly his discernible phenotype. Each OTTO together with his similars constitutes a subgroup, which we termed for this approach 'squad', whose members are characterized by a certain mental disease-relevant genotype–phenotype constellation. Exploring OTTO to uncover this constellation will, according to our hypothesis, open new avenues in the search for biological subgroups of mental diseases.

<sup>1</sup>Clinical Neuroscience, Max Planck Institute of Experimental Medicine, Göttingen, Germany; <sup>2</sup>DFG Center for Nanoscale Microscopy and Molecular Physiology of the Brain (CNMPB), Göttingen, Germany; <sup>3</sup>Department of Psychiatry and Psychotherapy, University Medicine, Greifswald, Germany; <sup>4</sup>Computational Systems Biology, German Center for Neurodegenerative Diseases, Göttingen, Germany and <sup>5</sup>Neurogenetics, Max Planck Institute of Experimental Medicine, Göttingen, Germany. Correspondence: Professor H Ehrenreich, Clinical Neuroscience, Max Planck Institute of Experimental Medicine, Hermann-Rein-Str.3, Göttingen 37075, Germany. E-mail: ehrenreich@em.mpg.de

Received 25 June 2016; revised 8 September 2016; accepted 7 October 2016; published online 6 December 2016

The OTTO approach detailed here to provide proof of concept has been grounded on two important assumptions: (1) The latest schizophrenia GWAS hits of the PGC (Psychiatric Genomics Consortium)<sup>2</sup> are pertinent for mental disease. (2) At least in some cases, these GWAS hits alone will assist in defining or approximating mental disease-relevant genotype-phenotype associations. At the same time, we recognize several constraints of our OTTO approach as currently implemented: (i) The biological significance of most SNPs derived from GWAS is not yet clear. Many SNPs are intergenic or intronic and thus still lack any viable path to mechanistic insight. (ii) The GRAS (Göttingen Research Association for Schizophrenia) data set we use to evaluate OTTO is very large compared with other deeply phenotyped samples, yet it is relatively small for genetic studies. (iii) Replication by other laboratories will pose a challenge owing to the present uniqueness of the GRAS-type deep phenotyping database and may require the use of phenotype proxies as exemplified here. (iv) A homozygous risk assumption is made to facilitate our newly applied procedures (purposely skipping potential heterozygous risk at this point and accepting loss of power).

To sum up, the present study has been designed to raise in the first part the awareness of the enormous heterogeneity of the schizophrenia GWAS subject matrix, which urgently calls for alternative strategies to unravel biological disease subgroups. Building on this ground in the second and third part, the 'OTTO project' as an individual-based approach to categorize GWAS finds in mental illness is introduced: Several variations of the OTTO idea are presented and translated into first data sets, providing examples of promising phenotype-to-genotype or genotype-to-phenotype approaches, all starting from individuals. Genotypes (SNPs) extracted in this way from GWAS findings apparently contribute in aggregate to autistic or affective behavioral continua. Most importantly, their accumulation yields encouraging signals of replicability in independent samples, an absolutely essential prerequisite for this kind of approach.

## MATERIALS AND METHODS

### Subjects

All parts of the present study comply with the Helsinki Declaration and were approved by the Ethics Committees of the Universities of Göttingen and Greifswald, Germany or of collaborating centers. All subjects and/or their authorized legal representatives gave written informed consent.

**GRAS schizophrenia patients.** The GRAS data collection has been initiated in 2005 and consists of meanwhile 1106 genotyped and deeply phenotyped patients, diagnosed with schizophrenia or schizoaffective disorder according to Diagnostic and Statistical Manual of Mental Disorders Fourth Edition, Text Revision (DSM-IV-TR),<sup>11</sup> recruited in 23 collaborating centers across Germany.<sup>12,13</sup>

**GRAS healthy controls.** Healthy controls were voluntary blood donors (total:  $N=1259$ ) from the Department of Transfusion Medicine, University Medical Center Göttingen, Germany, who widely fulfill health criteria, assessed by pre-donation screening.<sup>12</sup>

**GRAS disease controls.** The extended GRAS data collection of mental disease controls (here:  $N=65$ ) consists of male and female patients diagnosed with psychiatric disorders other than schizophrenia according to DSM-IV-TR.<sup>11</sup>

**ASD subjects.** Starting in 2011, the GRAS data collection has been expanded to include autism spectrum disorder (ASD) subjects (here:  $N=81$  males). Prerequisites for the present study were male gender, Intelligent Quotient  $\geq 75$  and expressive language skills allowing semistructured interviewing.<sup>14</sup>

**General population sample.** The replication sample of the general population comprises 2400 subjects from a follow-up (SHIP-LEGEND) of

the Study of Health in Pomerania (SHIP), conducted in North-East Germany. SHIP investigates common risk factors, subclinical disorders and manifest diseases in the general population.<sup>15</sup>

### Phenotyping

Phenotyping of all subjects included in the present study has been described in great detail.<sup>3,12-16</sup> Only subjects with data available on all respective items were included in the composite score-based analyses, explaining the slightly variable  $N$  numbers.

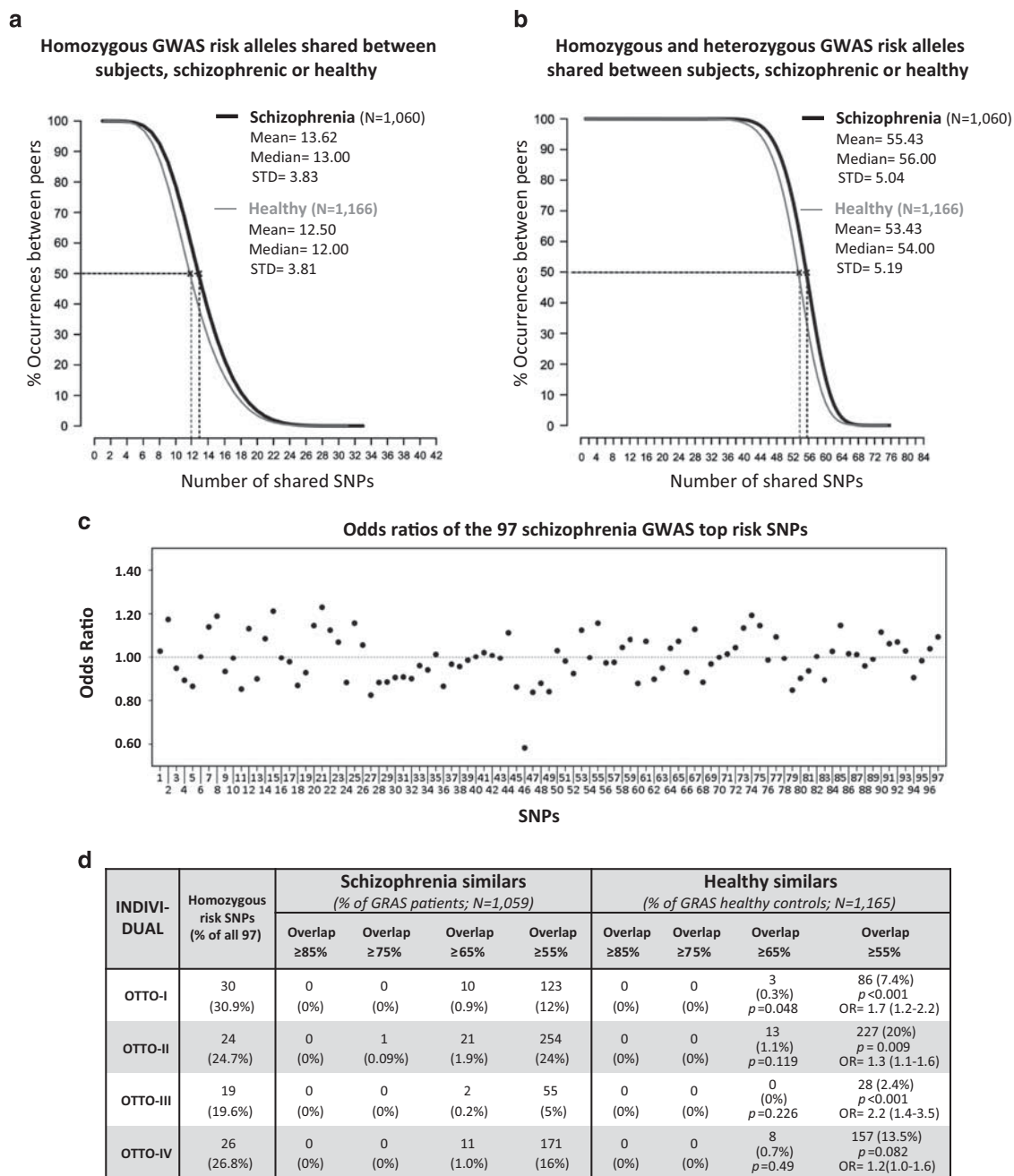
**Positive and Negative Syndrome Scale Autism Severity Score.** The Positive and Negative Syndrome Scale (PANSS) Autism Severity Score (PAUSS) is a cross-validated measure capturing the continuous nature of autistic behaviors.<sup>14</sup> It represents the sum of six items of the negative subscale and two items of the general subscale of the PANSS.<sup>17</sup>

**Affective Composite Score.** Depression, anxiety (PANSS general psychopathology subscale items 6 and 2; range 1-7),<sup>17</sup> suicidality (score 0-3: sum of dichotomous items suicidal ideation, suicide plans and suicide attempts) and Brief Symptom Inventory items paranoid ideation, depressivity, somatic concern and anxiousness<sup>18</sup> were integrated into the Affective Composite Score (ACS), calculated as sum of the single item values after Blom transformation of data.<sup>19</sup> Current depressive symptoms were assessed in the disease controls and the general population sample (SHIP) using the Beck Depression Inventory (BDI), which is a 21-item self-report questionnaire with high reliability and validity.<sup>20</sup> Self-reported chronic stress was measured in the general population with the 12-item Trier Inventory for the Assessment of Chronic Stress (TICS).<sup>21</sup>

### Genotyping

Genotyping has been performed using a semicustom Axiom myDesign genotyping array (Affymetrix, Santa Clara, CA, USA), based on a CEU marker backbone including 518 722 SNPs and a custom marker set including 102 537 SNPs. The array was designed using the Axiom Design Center (www.affymetrix.com), applying diverse selection criteria. Genotyping was performed by Affymetrix on a GeneTitan platform. SNP call rate  $>97\%$  and several other quality-control steps were applied. These steps were performed using either the Genotyping Console (GTC) software (Affymetrix) or R. In a subsequent step, markers in X, Y and mitochondrial chromosomes and those with Hardy-Weinberg equilibrium  $P < 1 \times 10^{-10}$  were removed, leaving 589 921 SNPs available for analyses.<sup>3,16,22</sup> Of the 1106 GRAS patients, 46 individuals had to be excluded for further analysis owing to relatedness, genotyping problems and genetic population outlier status (based on 10 principle components) ( $N=1060$ ; 709 males and 351 females; mean age  $39.47 \pm 12.64$  years; range 17-79). Similarly, of the 1259 healthy controls, 93 individuals had to be excluded for further analysis for the same reasons ( $N=1166$ ; 731 males and 435 females; mean age  $37.49 \pm 13.27$  years; range 18-69). For the present analyses, we selected those SNPs directly genotyped in our GRAS data set with genome-wide significance ( $P < 5 \times 10^{-8}$ ) in the most recent schizophrenia GWAS.<sup>2</sup> Pruning based on linkage disequilibrium  $r^2 \leq 0.05$  as implemented in PLINK<sup>23</sup> was carried out. This procedure left 97 GWAS-derived SNPs ready for analyses. Genotyping of the disease controls (for the present study:  $N=65$  with BDI available; 38 males and 27 females; mean age  $48.10 \pm 14.57$  years; range = 22-80) and the autism sample (for the present study, high-functioning male ASD subjects:  $N=81$ ; mean age  $31.78 \pm 10.83$  years; range = 16-63) was carried out analogously. The general population sample (from SHIP;  $N=2400$ ; mean age  $55.89 \pm 14.27$  years; range = 29-89) was genotyped using the Affymetrix Genome-Wide Human SNP array 6.0. Hybridization of genomic DNA was carried out in accordance with the manufacturer's standard recommendations. The overall genotyping efficiency was 98.6%. Imputation of genotypes in the SHIP cohort was performed with the software IMPUTE v0.5.0 against the 1000 Genomes (phase 1v3) reference panel using 869 224 genotyped SNPs.<sup>15</sup> Of the 4 SNPs in the schizoaffective model, 3 were imputed with a high imputation quality of  $>0.97$ .

**OTTO I-IV.** From the GRAS database, four schizophrenic subjects, referred to as OTTOs I-IV, with widely contrasting risk genotype constellation, that is, with lowest possible homozygous risk SNP overlap among each other, were selected (Figure 1).

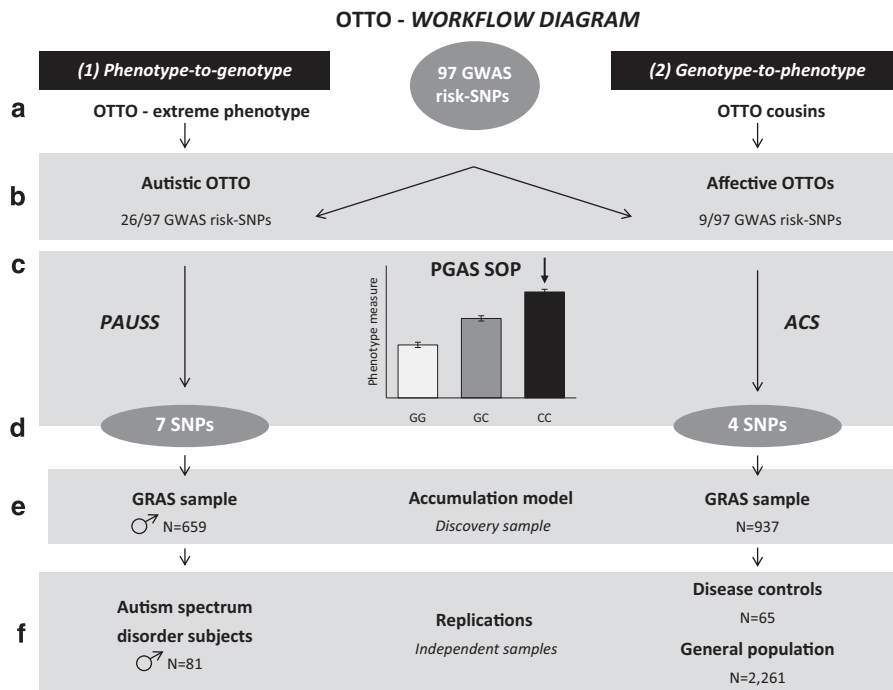


**Figure 1.** Enormous heterogeneity of the schizophrenia genome-wide association studies (GWAS) subject matrix and closeness of disease and health regarding GWAS hits. **(a, b)** Cumulative distributions of schizophrenia risk single-nucleotide polymorphisms (SNPs) in the GRAS (Göttingen Research Association for Schizophrenia) sample illustrate the genetic heterogeneity of GWAS on end point diagnosis. **(c)** Odds ratios of the 97 schizophrenia GWAS SNPs determined in 1060 schizophrenic/schizoaffective and 1166 healthy subjects of the GRAS data collection. **(d)** Presentation of 4 OTTOs with contrasting homozygous GWAS risk SNP constellation (compare Table 1) and their schizophrenia and healthy similars sharing  $\geq 85$ ,  $\geq 75$ ,  $\geq 65$  and  $\geq 55\%$  homozygous risk SNPs with the respective OTTO; *P*-values/odds ratio (95% confidence interval) from Fisher's exact test comparing schizophrenia with healthy similars of the  $\geq 65\%$  and  $\geq 55\%$  overlap squads.

'OTTO cousins'. From the raw genotype data (1106 schizophrenia patients), the full SNP data set was pruned using minor allele frequency ( $\geq 0.05$ ) and linkage disequilibrium between markers ( $r^2 \leq 0.05$ ) as inclusion criteria, leaving 33 311 markers ready for relatedness ascertainment. This was carried out with PLINK by calculating a genome-wide pairwise identity-by-descent estimation.<sup>23</sup> Six subject pairs with a PI-HAT value between 0.26 and 0.54 were identified that are normally excluded owing to relatedness (see above). In the 'cousin approach', however, 3 'OTTO cousin pairs' with a PI-HAT value of around 0.25 were further analyzed.

#### PGAS for SNP pruning and selection

All 97 GWAS risk SNPs are SNPs that 'survived' after linkage disequilibrium pruning ( $r^2 \leq 0.05$ ), an important prerequisite to exclude redundant information in the here pursued phenotype-based genetic association study (PGAS) approach. GWAS risk SNPs of OTTOs were individually screened by a PGAS approach using the adequate phenotype measure, here PAUSS or ACS. Single SNPs with a tendency of one homozygous genotype being associated with the phenotype of interest went into the accumulation models (standard



**Figure 2.** Individual-based approach to genome-wide association studies (GWAS) finds: OTTO workflow diagram to exemplify the applied novel procedures. **(a)** Selection of OTTOs based on (1) a phenotype-to-genotype approach starting from an individual with an extreme (here: autistic) phenotype and (2) a genotype-to-phenotype approach starting from cousin pairs with a shared severe (here: affective) phenotype. **(b)** Number of homozygous GWAS risk single-nucleotide polymorphisms (SNPs; of all 97 available after linkage disequilibrium pruning) carried by the autistic OTTO (26 SNPs) and the affective OTTOs (9 SNPs), respectively. **(c)** These GWAS risk SNPs are then individually screened for a tendency of association with the adequate phenotype measure (bar graph inset for illustration), here PAUSS (PANSS Autism Severity Score) and ACS (Affective Composite Score), respectively, using the GRAS (Göttingen Research Association for Schizophrenia) sample: PGAS (phenotype-based genetic association study) standard operating procedure (SOP). **(d)** Number of homozygous SNPs showing such tendency. **(e)** These SNPs go into the respective accumulation model and are evaluated in the discovery sample (GRAS). **(f)** Indispensable is the replication of any promising accumulation model in independent samples. In case of phenotype unavailability, a proxy phenotype has to be used.

operating procedure (SOP) explained in great detail in Stepniak *et al.*,<sup>16</sup> Figure 2).

#### Statistical analysis

Genotypic data were compiled recording the numbers of risk alleles contained in each of the 97 different SNPs. Distribution functions of shared SNPs were obtained by computing the number of shared SNPs for each possible pair of individuals (1060 schizophrenic and 1166 healthy individuals of the GRAS data collection). The counts representing the numbers of shared SNPs were used first to compute histograms of frequencies of shared SNPs and then to compute the reverse of the cumulative distribution function (CDF) of occurrences, where a reverse CDF is defined as  $1 - \text{CDF}$ ; 95% confidence intervals (CIs) computed through the Wald method along with a Kolmogorov–Smirnov test confirm that both distributions are indistinguishable from each other (Figures 1a and b). Mann–Whitney *U*-test or *t*-test were used for group comparisons, Jonckheere–Terpstra trend test was applied to test for trends. Spearman rank correlation coefficient was used to assess the strength of association between two non-parametric variables. Cronbach’s  $\alpha$  was calculated as a measure of internal consistency. For the ACS, rank-based Blom transformation<sup>19</sup> was applied to standardize all individual measures by transforming them into standard normally distributed surrogates prior to sum score computation. This maintained the order of the data but removed skewness from variable distributions. All *P*-values in discovery sample analyses are two-sided, in replications one-sided. Statistical analyses were performed with SPSS for windows version 17.0 (IBM-Deutschland, Munich, Germany) and STATA MP version 13.1 (StataCorp, College Station, TX, USA). Case–control analyses of SNP genotypes were performed using PLINK 1.07.<sup>23</sup> Fisher’s exact test was used to compare schizophrenia and healthy similars regarding the number of homozygous risk SNPs shared with the genetically defined OTTOs.

#### RESULTS

Raising the awareness of the enormous heterogeneity of the schizophrenia GWAS subject matrix and the closeness of health and disease

Before starting the OTTO approach, we aimed at documenting (i) the tremendous heterogeneity of the GWAS risk SNP distribution among schizophrenic individuals and controls, and (ii) the closeness of health and disease. For this, we explored the distribution of shared homozygous SNPs of the available 97 GWAS-derived risk SNPs<sup>2</sup> in individuals of the GRAS data collection, schizophrenic and healthy. As illustrated in Figure 1a, the number of any obtained combination of homozygous schizophrenia risk SNPs (of 97) shared among 50% of subjects is 13.0 for schizophrenic ( $N=1060$ ) and 12.0 for healthy subjects ( $N=1166$ ). Less than 10% of all GRAS subjects have any combination of 18 homozygous risk SNPs (of 97) in common. Including heterozygous in addition to homozygous risk SNPs, the numbers derived from any combination of the 97 risk SNPs reach 56.0 versus 54.0 shared genotypes between 50% of subjects, schizophrenic or healthy, respectively (Figure 1b). These data stress the enormous heterogeneity of the GWAS subject matrix used for identifying common schizophrenia risk alleles and the remarkable genetic similarity between health and disease regarding the GWAS hit (risk SNP) distribution.

Underlined are these facts by the overall low odds ratios (ORs) known for all 97 SNPs in the PGC schizophrenia GWAS data set<sup>2</sup> and reproduced here in the GRAS sample (Figure 1c), confirming poor class separation between healthy and schizophrenic GRAS



**Table 1.** Overview of the 97 schizophrenia GWAS risk SNPs and their distribution among OTTOs

SNP <sup>a</sup>	PGC schizophrenia GWAS results					GRAS	Homozygous schizophrenia GWAS risk SNP pattern of OTTOs										
	SNP	Chr.	Chr. position (hg19)	Band	Risk allele	Allele freq.	I	II	III	IV	Autistic	A1	A2	B1	B2	C1	C2
1	rs13596	1	8421203	p36.23	T	0.331	X										
2	rs1143702	1	44086831	p34.2	C	0.348	X		X						X		
3	rs1416266	1	73892824	p31.1	C	0.529						X					X
4	rs4271249	1	98398619	p21.3	T	0.593	X			X		X		X			
5	rs12138231	1	150115398	q21.2	A	0.826	X		X	X		X		X	X	X	X
6	rs10803138	1	243555219	q43	G	0.738	X			X <sup>b</sup>		X		X		X	X
7	rs11682175	2	57987593	p16.1	C	0.484			X								
8	rs16825349	2	146425531	q22.3	G	0.170										X	
9	rs4380187	2	185811940	q32.1	A	0.559			X					X	X		
10	rs4685	2	198257795	q33.1	T	0.321	X		X								
11	rs6704641	2	200164252	q33.1	A	0.822			X	X		X <sup>c</sup>	X <sup>c</sup>	X	X	X	X
12	rs3769472	2	201193373	q33.1	A	0.335											
13	rs10498163	2	225434445	q36.2	A	0.671						X <sup>c</sup>	X <sup>c</sup>			X	
14	rs10205801	2	233597196	q37.1	G	0.475			X								X
15	rs11714012	3	2499384	p26.3	A	0.281									X		X
16	rs13314421	3	36962388	p22.2	G	0.477											
17	rs13083798	3	52649748	p21.1	A	0.505	X								X	X	
18	rs696520	3	136056861	q22.3	T	0.579			X						X		
19	rs1011927	3	180636744	q26.33	A	0.788		X		X		X		X	X		
20	rs13107325	4	103188709	q24	T	0.070											
21	rs1106568	4	176861301	q34.2	G	0.255			X				X				
22	rs35804423	5	45354799	p12	A	0.182								X			
23	rs13169711	5	60660401	q12.1	A	0.503		X		X							X
24	rs16867571	5	88743219	q14.3	A	0.784		X					X	X			X
25	rs3798149	5	137722822	q31.2	C	0.265			X		X		X		X		X
26	rs177001	5	152610561	q33.1	G	0.505			X	X		X					
27	rs1028885	6	25434518	p22.2	A	0.784	X		X	X				X			X
28	rs9379780	6	25707171	p22.2	A	0.642			X								
29	rs651578	6	25873883	p22.2	T	0.778	X		X	X	X					X	X
30	rs198811	6	26128446	p22.2	T	0.633	X		X	X <sup>b</sup>						X	X
31	rs66972160	6	26359306	p22.2	T	0.617						X		X			X
32	rs2076029	6	26390830	p22.2	T	0.596						X		X		X	X
33	rs7768814	6	26892036	p22.2	A	0.628	X		X			X				X	
34	rs9379968	6	27244590	p22.1	G	0.541						X					
35	rs2205829	6	27480227	p22.1	G	0.462										X	
36	rs2056925	6	27690905	p22.1	A	0.770					X	X		X		X	X
37	rs203878	6	28048996	p22.1	T	0.657					X					X	X
38	rs1225715	6	28113373	p22.1	A	0.660					X <sup>b</sup>					X	X
39	rs1736895	6	28219826	p22.1	G	0.631					X					X	X
40	rs1052215	6	28348158	p22.1	G	0.643		X			X						X
41	rs916403	6	28408258	p22.1	A	0.581		X			X <sup>b</sup>						X
42	rs385492	6	29649547	p22.1	T	0.478			X			X					
43	rs3132565	6	31102964	p21.33	A	0.536		X			X						X
44	rs614549	6	31840625	p21.33	G	0.410	X									X	
45	rs72851052	6	32502249	p21.32	C	0.595		X	X	X		X					X
46	rs11969547	6	33160906	p21.32	G	0.965	X	X	X	X	X <sup>b</sup>	X	X	X	X	X	X
47	rs9461856	6	33395199	p21.32	A	0.525			X		X <sup>b</sup>			X	X		
48	rs971215	6	84277037	q14.2	C	0.598	X									X	
49	rs579536	6	96438499	q16.1	G	0.927	X	X	X	X					X		X
50	rs9398171	6	108983527	q21	C	0.295									X		
51	rs10243920	7	1950385	p22.3	G	0.596			X	X						X	X
52	rs13230321	7	86459347	q21.12	A	0.842		X	X			X <sup>c</sup>	X <sup>c</sup>		X	X	X
53	rs2192932	7	104653265	q22.3	A	0.371		X									
54	rs253376	7	110939653	q31.1	C	0.547		X	X	X							
55	rs7801375	7	131567263	q32.3	G	0.839	X	X	X	X		X	X		X	X	X
56	rs7779548	7	137074540	q33	G	0.670		X	X	X						X	
57	rs13261217	8	4183057	p23.2	G	0.202											
58	rs6471803	8	60667231	q12.1	G	0.424											
59	rs7815859	8	111476741	q23.2	A	0.212						X					
60	rs7822538	8	143323109	q24.3	T	0.597		X	X				X	X	X		
61	rs1319017	9	84736303	q21.32	A	0.318											
62	rs12784387	10	18804689	p12.31	T	0.760	X				X				X		
63	rs619824	10	104581288	q24.32	C	0.559	X		X			X					X
64	rs11027846	11	24394815	p14.3	C	0.500	X										
65	rs7127529	11	46582072	p11.2	C	0.184				X							
66	rs17601612	11	113317745	q23.2	G	0.637		X								X	X
67	rs7927176	11	123395864	q24.1	G	0.328		X									
68	rs12293670	11	124612932	q24.2	A	0.670	X					X			X	X	

**Table 1.** (Continued)

SNP <sup>a</sup>	PGC schizophrenia GWAS results					GRAS	Homozygous schizophrenia GWAS risk SNP pattern of OTTOs										
	SNP	Chr.	Chr. position (hg19)	Band	Risk allele	Allele freq.	I	II	III	IV	Autistic	A1	A2	B1	B2	C1	C2
69	rs11222406	11	130814093	q25	A	0.506		X	X					X	X		
70	rs11223648	11	133840412	q25	C	0.770	X	X		X		X	X	X	X	X	X
71	rs2370413	12	2354870	p13.33	C	0.468	X					X	X		X		
72	rs2239063	12	2511831	p13.33	A	0.711	X		X			X <sup>c</sup>	X <sup>c</sup>			X	X
73	rs4766428	12	110723245	q24.11	T	0.483		X						X			
74	rs61388686	12	123810873	q24.31	G	0.354			X			X					
75	rs2068012	14	30190316	q12	C	0.248											
76	rs36345	14	72389021	q24.2	C	0.558	X				X						
77	rs11628332	14	104241141	q32.33	T	0.424		X	X					X		X	
78	rs4635285	15	61838313	q22.2	T	0.541	X		X					X		X	
79	rs578776	15	78888400	q25.1	G	0.742	X		X	X		X					X
80	rs748455	15	85149575	q25.2	T	0.693		X			X			X			X
81	rs1894401	15	91429042	q26.1	A	0.527				X				X			
82	rs7191183	16	9900057	p13.2	C	0.321					X	X	X				X
83	rs3814883	16	29994922	p11.2	C	0.564	X										X
84	rs6498914	16	63699425	q21	T	0.485						X					
85	rs8044995	16	68189340	q22.1	A	0.184									X		
86	rs364584	17	2211416	p13.3	A	0.633		X			X <sup>b</sup>			X		X	
87	rs12150369	17	17911014	p11.2	A	0.337											
88	rs4801131	18	52752700	q21.2	C	0.606	X					X				X	
89	rs17596974	18	53166557	q21.2	T	0.044											
90	rs9636107	18	53200117	q21.2	G	0.480						X					X
91	rs9945732	18	53593297	q21.2	G	0.936	X	X		X	X	X	X	X	X	X	X
92	rs2916076	19	19545990	p13.11	C	0.358											
93	rs1004046	19	30992639	q12	C	0.278						X					
94	rs6028163	20	37477034	q11.23	C	0.646	X							X	X	X	X
95	rs132570	22	39976128	q13.1	C	0.484			X						X		
96	rs926914	22	41418154	q13.2	T	0.265			X			X		X			
97	rs5758653	22	42613485	q13.2	T	0.455				X				X			

Abbreviations: GRAS, Göttingen Research Association for Schizophrenia; GWAS, genome-wide association studies; PANSS, Positive and Negative Syndrome Scale; PGC, Psychiatric Genomics Consortium; SNP, single-nucleotide polymorphism. <sup>a</sup>Refer to SNP numbers in Figure 1c. <sup>b</sup>SNPs included in the accumulation model associated with PANSS Autism Severity Score, PAUSS (Figure 3). <sup>c</sup>SNPs included in the accumulation model associated with the Affective Composite Score, ACS (Figure 4)

subjects. To verify whether these data might be separable upon use of other methods, we applied a projection technique that is linear and incorporates class information, the so-called linear discriminant analysis,<sup>24</sup> and a highly non-linear tree-based method, the random forest classifier.<sup>25,26</sup> Results of both approaches again highlight that schizophrenic and healthy GRAS individuals cannot be separated using the genotypic data (all *P*-values > 0.25) (Supplementary-Figures S1 and S2).

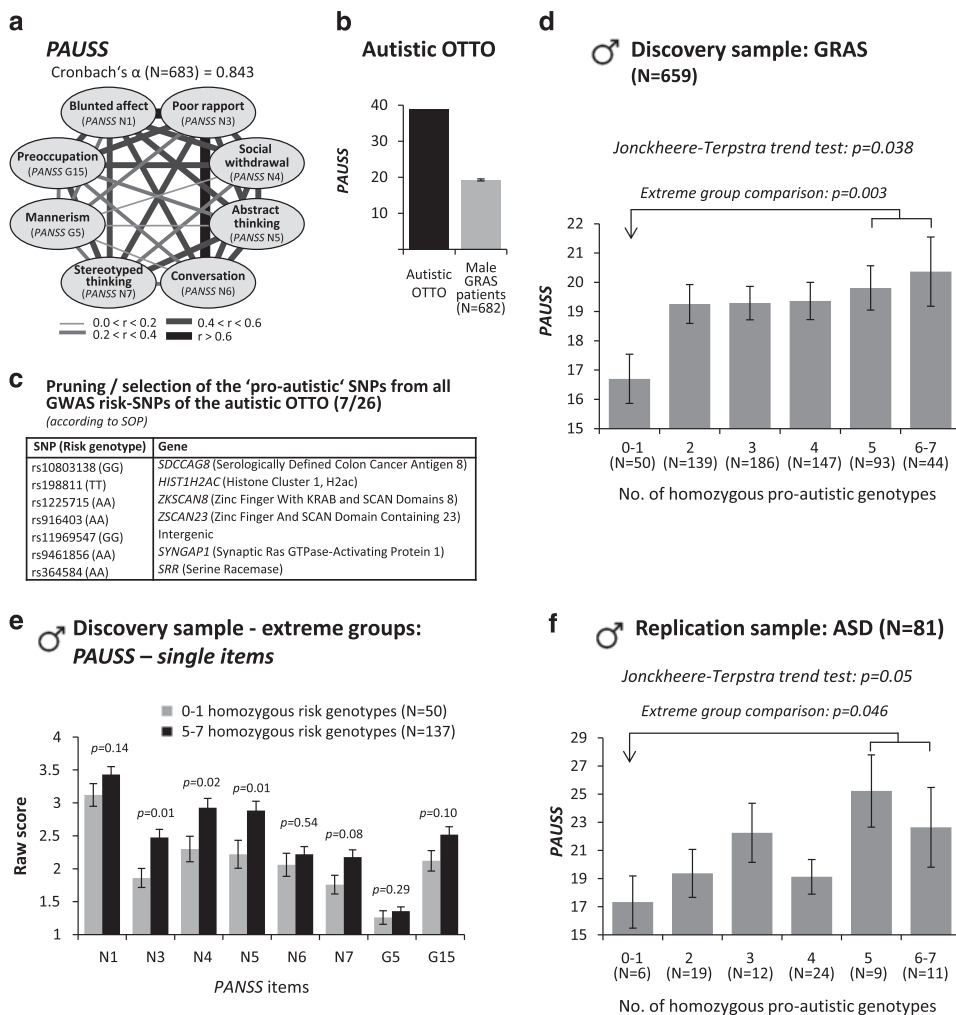
Risk SNP overlaps of four contrasting individual OTTOs with their respective schizophrenic and healthy similars in the GRAS database

We now moved on to a first approach based on individual OTTOs. For this, we chose four schizophrenic GRAS subjects with widely contrasting risk genotype constellation. In other words, we selected four OTTOs (I–IV) with lowest possible homozygous risk SNP overlap among each other. As shown in Figure 1d and Table 1, these 4 OTTOs carry between 19 and 30 homozygous risk genotypes each. Searching for schizophrenia similars of all 4 OTTOs by stepwise lowering the risk SNP overlap threshold (starting from ≥ 85%), only 1 other individual was found among 1059 GRAS schizophrenic patients with ≥ 75% overlap, in this case with OTTO-II; lowering to an overlap of ≥ 65% resulted in 2–21 schizophrenic individuals, similar to OTTOs I–IV. Extracting healthy similars, the situation up to an overlap in risk SNPs of ≥ 65% is broadly comparable. Only OTTO-I similars are less frequently found in healthy as compared with schizophrenic subjects (*P* = 0.048; OR = 3.7; 95% CI: 1.01–13.46), but a tendency is seen

for all OTTOs. At an overlap of ≥ 55%, distinctly more similars are found, with 3/4 healthy squads (those of OTTO-I, OTTO-II and OTTO-III) being smaller than the schizophrenic squads. This finding may be of potential interest with respect to genetic schizophrenia subgroups but needs replication and further refinement in larger samples. The small numbers of OTTO I–IV similars in the GRAS sample with ≥ 65% shared GWAS risk SNPs makes the vast heterogeneity of the GWAS subject matrix, as based on end point diagnoses, again very obvious. A genotype-to-phenotype approach with OTTOs I–IV as originally planned was therefore not deemed reasonable with our sample.

Phenotype-to-genotype approach: the autistic OTTO

We thus performed a reverse approach, that is, a direct phenotype-to-genotype approach, starting with an individual exhibiting an extreme phenotype that is not expected to be as prominently affected by antipsychotic medication as, for instance, the severity of positive symptoms. For this, a male individual of the GRAS sample was selected, characterized by an unusually severe autistic phenotype as determined by PAUSS (Figures 2 and 3a and b).<sup>14</sup> In fact, previous work underscores the convergence of schizophrenia-negative symptoms and autistic phenotypes. PAUSS evolved as a dimensional measure capturing the continuous nature of autistic behaviors.<sup>14</sup> Schizophrenia patients predominantly suffering from negative symptoms obtain high scores also on the Autism Diagnostic Observation Schedule.<sup>27</sup> Importantly, the here selected patient had been diagnosed with ASD in childhood (that is, before any antipsychotic medication)



**Figure 3.** Example of a 'bottom-up' phenotype-to-genotype approach starting from a schizophrenia patient with an extreme phenotype, the 'autistic OTTO'. (a) Quantification of autistic symptoms using the PAUSS (PANSS Autism Severity Score). Note the high intercorrelation of PAUSS items and the high internal consistency of the scale (Spearman rank correlation coefficients; Cronbach's  $\alpha$ ). (b) Illustration of the extreme phenotype of the autistic OTTO compared with the mean PAUSS ( $\pm$  s.e.m.) of male schizophrenia patients (GRAS (Göttingen Research Association for Schizophrenia)). (c) List of 7 selected 'pro-autistic' single-nucleotide polymorphisms (SNPs) of the 26 genome-wide association studies (GWAS) risk SNPs carried in homozygosity by the autistic OTTO. Pruning/selection was carried out according to their phenotypical relevance following the standard operating procedure described in the Materials and methods section. (d) An accumulation of the seven risk SNPs is associated with PAUSS in the discovery sample; two-sided  $P$ -values for Jonckheere–Terpstra trend test and Mann–Whitney  $U$ -test of extreme group comparison (data corrected by linear regression analysis for antipsychotic medication using chlorpromazine equivalent doses); mean  $\pm$  s.e.m. (e) Separate presentation of extreme group results for all PAUSS items shows a consistent tendency of higher values upon higher genetic risk. (f) Effect of risk SNP accumulation on PAUSS and extreme group contrast are replicated in an independent sample of individuals diagnosed with autism spectrum disorder (ASD); one-sided  $P$ -values for Jonckheere–Terpstra trend test and Mann–Whitney  $U$ -test of extreme group comparison; mean  $\pm$  s.e.m.

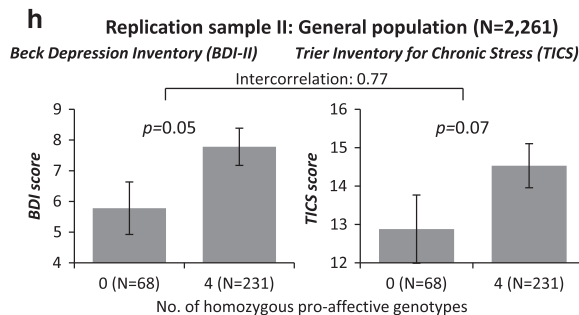
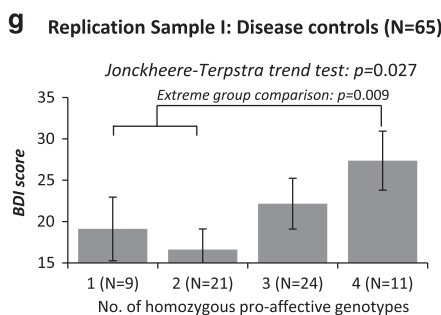
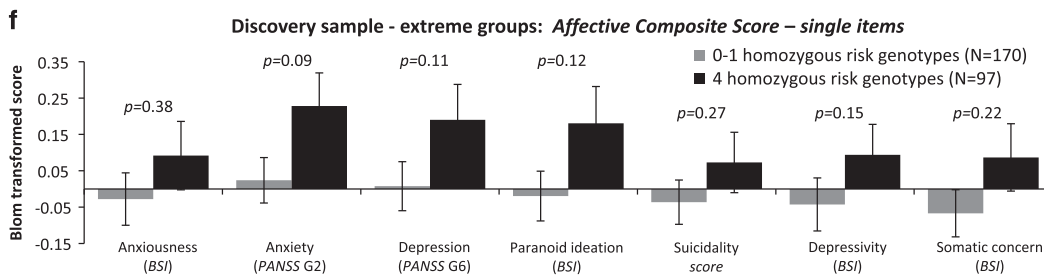
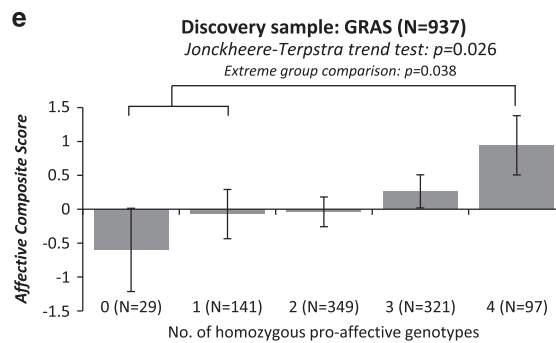
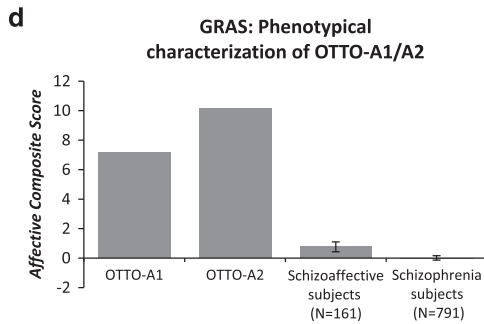
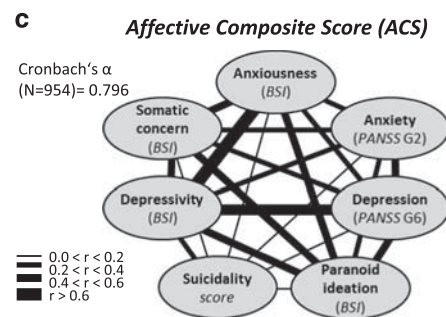
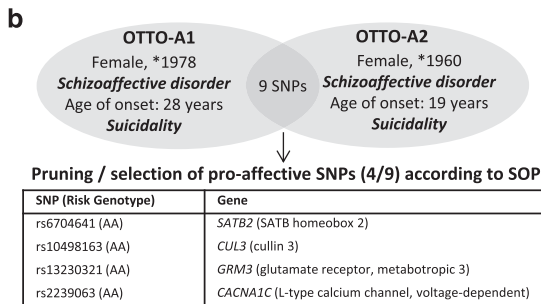
**Figure 4.** Example of a 'bottom-up' genotype-to-phenotype approach starting from schizoaffective OTTO cousin pairs. (a) Presentation of three OTTO cousin pairs in schizophrenic GRAS (Göttingen Research Association for Schizophrenia) subjects with shared homozygous risk alleles. Schizophrenia and healthy similars of each OTTO cousin pair sharing  $\geq 85\%$  and  $\geq 75\%$  homozygous risk single-nucleotide polymorphisms (SNPs) are shown; odds ratio (OR) and  $P$ -value from Fisher's exact test comparing respective squads of schizophrenia and healthy similars. (b) Schizoaffective OTTO-A1/A2 pair and list of the four selected 'pro-affective' SNPs out of the nine genome-wide association studies (GWAS) risk SNPs shared in homozygosity. Pruning/selection was performed according to standard operating procedure described in Materials and methods section. (c) Quantification of affective symptoms using the Affective Composite Score (ACS): High intercorrelation of ACS items and high internal scale consistency (Spearman rank correlation; Cronbach's  $\alpha$ ). (d) Extreme phenotypes of affective OTTOs compared with ACS of schizoaffective and schizophrenic GRAS patients (mean  $\pm$  s.e.m.). (e) Accumulation of the four risk SNPs associated with ACS in the discovery sample; two-sided  $P$ -values for Jonckheere–Terpstra trend test and Mann–Whitney  $U$ -test of extreme group comparison (data corrected by linear regression analysis for antipsychotic medication using chlorpromazine equivalent doses); mean  $\pm$  s.e.m. (f) Separate presentation of extreme group results for all ACS items shows a consistent tendency of higher values upon higher genetic risk. (g) Replication I: Effect of risk SNP accumulation on Beck Depression Inventory (BDI) as proxy of ACS and extreme group contrast are replicated in an independent sample of disease controls; one-sided  $P$ -values for Jonckheere–Terpstra trend and  $T$ -test of extreme group comparison; mean  $\pm$  s.e.m. (h) Replication II: BDI and Trier Inventory for the Assessment of Chronic Stress (TICS) as proxies of ACS for extreme group contrast in healthy controls (Study of Health in Pomerania); one-sided  $P$ -values for  $T$ -test of extreme group comparison; mean  $\pm$  s.e.m.

and was later switched to a schizophrenia diagnosis. Figure 3b illustrates the distinct phenotype of this 'autistic OTTO', contrasting him from the average of GRAS schizophrenic males. The autistic OTTO is homozygous for the risk allele at 26/97 GWAS

risk loci. To identify among the 26 those markers that likely co-modulate autistic features, we applied our PGAS SOP, using PAUSS in the male GRAS sample<sup>16</sup> (Figure 2). A total of seven homozygous risk SNPs, termed 'pro-autistic genotypes',

**a**

'OTTO cousin pairs'	Homozygous Risk SNPs (% of all 97)	Schizophrenia similars Overlap $\geq 85\%$ (% of GRAS patients)	Healthy similars Overlap $\geq 85\%$ (% of healthy controls)	OR (CI 95%) P-value	Schizophrenia similars Overlap $\geq 75\%$ (% of GRAS patients)	Healthy similars Overlap $\geq 75\%$ (% of healthy controls)	OR (CI 95%) P-value
OTTO-A1	9 (9.3%)	53 (5%)	56 (4.8%)	1.04 (0.7-1.5) 0.845	235 (22%)	247 (21%)	1.1 (0.9-1.3) 0.571
OTTO-A2	11 (11.3%)	34 (3.2%)	27 (2.3%)	1.4 (0.8-2.3) 0.242	135 (13%)	103 (9%)	1.5 (1.2-1.9) 0.003
OTTO-B1	18 (18.6%)	46 (4.3%)	26 (2.2%)	1.99 (1.2-3.2) 0.006	116 (11%)	99 (9%)	1.3 (1.0-1.8) 0.052
OTTO-B2							
OTTO-C1							
OTTO-C2							





individually showing a tendency of an association with PAUSS were extracted and included in the accumulation model (overview of the seven genes in Figure 3c). Interestingly, 5/7 identified 'pro-autistic' loci are within the extended major histocompatibility complex region (Table 1). An accumulation score, representing the sum of homozygous pro-autistic genotypes of the seven selected SNPs, was calculated for each male individual of discovery and replication samples. In the discovery sample, a significant trend of higher numbers of risk SNPs with higher PAUSS values emerged (Figure 3d). Extreme group comparison (0–1 versus 5–7 risk SNPs) was significant for PAUSS (Figure 3d) and revealed trends of higher values with higher-risk SNP numbers for all single PAUSS items (Figure 3e). As both the selection of PAUSS-associated SNPs of the autistic OTTO and the accumulation model were based on the male GRAS (that is, discovery) sample, 'circularity' of the procedure has to be admitted and the significant values obtained should thus be regarded as 'exploratory significance estimates'. Replicability is therefore the most crucial issue. Importantly, the discovery sample result was reproducible in a small independent sample of male individuals diagnosed with ASD (Figure 3f). As an important control, we screened all 71 (97–26) risk SNPs, which were absent (or heterozygous) in the autistic OTTO, for association with PAUSS using our PGAS SOP. We found 27/71 SNPs that showed the association pattern. Expectedly, their accumulation in the GRAS males yielded a highly significant Jonckheere–Terpstra trend ( $P=0.0000002$ ). However, replication in the ASD sample completely failed ( $P=0.919$ ), underlining the necessity to start from an individual (OTTO) for pruning of phenotype-irrelevant SNPs. These data encourage the use of a GWAS-derived risk SNP 'package' of a subject with extreme autistic traits for predicting the severity of autistic features in the sense of a behavioral continuum, as shown here in a schizophrenic and an independent ASD replication sample. Especially the replication is absolutely essential for this type of analysis as outlined above.

#### Genotype-to-phenotype approach using cousin pairs: the affective OTTOs

We next reasoned that another way to define OTTOs—via additional risk genotype pruning—could be achieved by using pairs of subjects who share a certain proportion of their genome as well as the end point diagnosis. This should yield better information on true risk genotypes. We thus employed real genetic cousins, all diagnosed with schizophrenia, as OTTO cousin pairs for a genotype-to-phenotype approach (Figures 2 and 4a). Those SNPs for which the schizophrenia risk allele was present in homozygosity in both members of each pair were extracted (9, 11 and 18 SNPs for pairs A1/2, B1/2 and C1/2, respectively) and used to define 'OTTO cousin similars'. As for the autistic OTTO, schizophrenic and healthy similars were identified according to the amount of shared risk SNPs with each OTTO cousin pair, that is,  $\geq 85\%$  or  $\geq 75\%$  homozygous risk SNPs. As shown in Figure 4a, the number of individuals with risk SNP overlap  $\geq 85\%$  reaches values (34, 46 and 53) that can be used for a first exploratory genotype-to-phenotype approach. For two of the OTTO cousin pairs, even significant differences between numbers of schizophrenic and healthy similars are seen, potentially indicating signals of genetic schizophrenia subgroups. As an example of our genotype-to-phenotype approach, the OTTO-A1/2 pair (9 SNPs, 53 similars with  $\geq 85\%$  overlap) is further pursued here owing to a remarkably prominent, shared phenotype of these cousins.

To investigate the potential symptomatic relevance of the SNPs overlapping between OTTO-A1 and OTTO-A2, we first evaluated the clinical syndrome of the two female cousins based on the GRAS examination book<sup>13</sup> and all medical reports and discharge letters. OTTO-A1 was first diagnosed in 1998 with bipolar disorder,

depressive episode with psychotic symptoms, and later switched to the diagnosis schizoaffective disorder according to DSM-IV-TR.<sup>11</sup> She had attempted to commit suicide once. Her older cousin, OTTO-A2, was, in 1978, also first diagnosed with bipolar disorder, manic episode without overt psychotic symptoms, and received the diagnosis of schizoaffective disorder in 2006. OTTO-A2 had 7 × attempted to commit suicide. Thus schizoaffective and bipolar symptoms as well as suicidality emerged as overlapping clinical characteristics of the two cousins (Figure 4b). We therefore generated, based on the most prominent symptoms of OTTO-A1/2, an ACS, as described in the Methods section (Figure 4c). Applying the ACS, 4/9 risk SNPs of OTTO-A1/2 were selected using our PGAS SOP (Figure 4b). As illustrated in Figure 4d, the phenotype (ACS) of the two cousins, compared with the mean ACS of all schizoaffective and schizophrenia subjects of GRAS, stuck out, featuring again an extreme phenotype as suitable discovery matrix. We next explored a step-wise accumulation of the four selected ACS-associated risk SNPs in the GRAS discovery sample and obtained 'exploratory significance estimates' for overall trend as well as for extreme group comparison (Figure 4e). Again, extreme groups of 0–1 versus 4 homozygous risk SNPs tended to differ in all ACS single items (Figure 4f). GRAS disease controls and individuals of the general population served as replication samples I and II. For both, the BDI<sup>20</sup> had to be used as ACS proxy (owing to missing ACS information). In the small replication sample I, both trend test and extreme group comparison were significant (Figure 4g). Even in healthy individuals of SHIP (replication sample II), extreme groups were still borderline significant regarding BDI and the highly inter-correlated ( $r=0.77$ ) TICS<sup>21</sup> (Figure 4h). Again, control screening of the 88 (97–9) risk SNPs, which were absent (or heterozygous) in the affective OTTOs, for association with ACS using our PGAS SOP, led to 36 SNPs with the association pattern. Even though their accumulation in the GRAS sample resulted in a highly significant Jonckheere–Terpstra trend ( $P=0.00018$ ) owing to the above-mentioned circularity, replication in the disease controls failed completely ( $P=0.973$ ). Thus both individual-based approaches, the 'extreme autistic OTTO' and the 'affective OTTO cousin pair', led to small but replicable signals of phenotype–genotype associations.

## DISCUSSION

The present work has been designed as a first individual-based approach to categorize GWAS finds in mental illness. Referred to as the 'OTTO project', it follows a bottom-up approach to define biological subgroups of complex mental disease. The OTTO project builds on deeply phenotyped schizophrenic individuals who all carry a certain number of GWAS risk genotypes (as do all healthy individuals) and searches for genetic (GWAS risk SNPs) as well as phenotypic similars among subjects with the same end point diagnosis. The first part of this work was intended to exemplify the huge heterogeneity of the schizophrenia GWAS subject matrix. This heterogeneity urgently calls for alternative strategies to unravel biological disease subgroups or mental disease-relevant behavioral phenotypes. The second and third part translated the OTTO idea to first data and showed examples of promising phenotype-to-genotype and genotype-to-phenotype approaches, starting from individuals.

In fact, applying the OTTO procedure, one can now begin to delineate the contribution of normal genotypes (here: GWAS risk SNPs) to (disease) phenotypes.<sup>28</sup> Many behavioral traits constitute a continuum as exemplified in the present study for autistic (PAUSS) or affective symptoms (ACS). Health and disease genotypes are remarkably close, illustrated here by the cumulative distribution of shared GWAS risk SNPs. Also 'disease borders' according to classification systems commence to vanish: A recent multivariate sibling study of the Swedish population, starting

from >3 million individuals, reported on a general genetic factor of psychopathology, indicating that common psychiatric disorders share the same genetic origin.<sup>29</sup> Moreover, a proportion of GWAS risk SNPs may not even be disease phenotype related, as suggested by the necessity of pruning individual risk SNP 'packages' (to yield 7/26 or 4/9) for overt phenotype relevance. They may mark differences between schizophrenia cases and controls that are not associated with schizophrenia at all (for example, cigarette consumption or social status). Clearly, the mechanistic (molecular–cellular) contribution of OTTO's identified risk SNPs to the phenotype still remains to be elucidated.

Nevertheless, first hints toward biological (genetic) roots of behavioral features or of mental disease-relevant phenotypes may be obtained using our model approach. Such hints have been demonstrated here for an autistic phenotype deduced from an individual OTTO with extreme autistic features and an affective phenotype, derived from an OTTO cousin pair. Even though the signals are still small, they were reproducible in independent replications, including a general population sample. Ideally, biological subgroups derived from OTTOs may be used for GWAS against healthy controls or against samples with crude end point diagnosis (schizophrenia): The better the biological subgroup definition, the greater the chance to harvest more specific, disease-relevant genetic (GWAS) hits. This will, however, require deep phenotyping of large numbers of individuals to allow PGAS.<sup>16,28</sup> Recently, also the term 'phenome-wide association studies' was coined for exploiting readouts of phenotypic data from electronic health records, which may be very helpful but likely too crude for neuropsychiatric diseases.<sup>30</sup>

Not to forget is the tremendous influence of the environment, mediated by, for example, epigenetic modulation of the inherited substrate.<sup>3</sup> This suggests novel scientific endeavors for estimating the environmental impact on individuals (OTTOs) by 'deep phenotyping' also of the environment. For influences such as urbanicity, air pollution or constant noise exposure, deep environmental phenotyping may be realizable. First approaches in this direction, considering individual stress load, are already underway.<sup>31</sup> Deep phenotyping of individuals and environment to complement and start making sense of genetic (GWAS) information is ultimately inevitable but extremely labor intense and has neither been appreciated nor well funded in recent years. Based on the numbers of deeply phenotyped subjects extending to at least 5–10 times the GRAS collection, the OTTO model will yield many discernible disease-relevant phenotypes and ultimately deliver a superior subject matrix for GWAS.

To conclude, the here suggested model approach, starting with an 'extreme phenotype OTTO' and in particular with an 'OTTO cousin pair' may aid in revealing phenotypical contributions of a certain GWAS risk SNP constellation. We note, however, that while for OTTO and his squad, the identified risk SNPs are among the relevant genetic underpinnings of their phenotype, other individuals with autistic or affective traits may have totally different genetic reasons for showing this particular phenotype ('many ways lead to Rome'). Nevertheless, complementary to the top-down approach of GWAS in complex diseases, the OTTO project provides a bottom-up model exploiting GWAS results for the identification of distinct genetic roots of behavioral continua, as relevant for mental disease subcategories.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGMENTS

We are indebted to Professor Michael O'Donovan, MRC Centre for Neuropsychiatric Genetics and Genomics, School of Medicine, Cardiff University, UK for critical reading

of the manuscript and highly appreciated feedback and advice. This work was supported by the Max Planck Society, the Max Planck Förderstiftung, the DFG (CNMPB), EXTRABRAIN EU-FP7, the Niedersachsen-Research Network on Neuroinfectiology (N-RENNT), and EU-AIMS. The research of EU-AIMS receives support from the Innovative Medicines Initiative Joint Undertaking under grant agreement no. 115300, resources of which are composed of financial contribution from the European Union's Seventh Framework Programme (FP7/2007-2013), from the EFPIA companies and from Autism Speaks. SHIP is part of the Community Medicine Research net of the University of Greifswald, which is funded by the Federal Ministry of Education and Research (grant nos. 01ZZ9603, 01ZZ0103 and 01ZZ0403), the Ministry of Cultural Affairs and the Social Ministry of the Federal State of Mecklenburg-West Pomerania. Genome-wide data in SHIP have been supported by a joint grant from Siemens Healthcare, Erlangen and the Federal State of Mecklenburg-West Pomerania. We thank all subjects for participating in the study, and all the many colleagues who have contributed over the past decade to the GRAS data collection.

## AUTHOR CONTRIBUTIONS

HE and KAN conceptualized and HE planned, supervised and coordinated the project. MB recruited and examined GRAS patients as well as subjects of the extended GRAS collection. MM was instrumental for all database work. MM, SV and TP performed the PGAS study, including phenotypical and statistical analyses of discovery sample and replicate samples. HJG and SB contributed to the study design related to their expertise and to respective data interpretation. HE with the support of KAN wrote the paper. All authors read and approved the final version of the manuscript.

## REFERENCES

- 1 Mah S, Nelson MR, Delisi LE, Reneland RH, Markward N, James MR *et al*. Identification of the semaphorin receptor PLXNA2 as a candidate for susceptibility to schizophrenia. *Mol Psychiatry* 2006; **11**: 471–478.
- 2 Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* 2014; **511**: 421–427.
- 3 Stepniak B, Papiol S, Hammer C, Ramin A, Everts S, Hennig L *et al*. Accumulated environmental risk determining age at schizophrenia onset: a deep phenotyping-based study. *Lancet Psychiatry* 2014; **1**: 444–453.
- 4 Papiol S, Mitjans M, Assogna F, Piras F, Hammer C, Caltagirone C *et al*. Polygenic determinants of white matter volume derived from GWAS lack reproducibility in a replicate sample. *Transl Psychiatry* 2014; **4**: e362.
- 5 Terwisscha van Scheltinga AF, Bakker SC, van Haren NE, Derks EM, Buizer-Voskamp JE, Boos HB *et al*. Genetic schizophrenia risk variants jointly modulate total brain and white matter volume. *Biol Psychiatry* 2013; **73**: 525–531.
- 6 Van der Auwera S, Wittfeld K, Homuth G, Teumer A, Hegenscheid K, Grabe HJ. No association between polygenic risk for schizophrenia and brain volume in the general population. *Biol Psychiatry* 2015; **78**: E41–E42.
- 7 Sekar A, Bialas AR, de Rivera H, Davis A, Hammond TR, Kamitaki N *et al*. Schizophrenia risk from complex variation of complement component 4. *Nature* 2016; **530**: 177–183.
- 8 Hammer C, Begemann M, McLaren PJ, Bartha I, Michel A, Klose B *et al*. Amino acid variation in HLA class II proteins is a major determinant of humoral response to common viruses. *Am J Hum Genet* 2015; **97**: 738–743.
- 9 Hammer C, Wanitchakool P, Sirianant L, Papiol S, Monnheimer M, Faria D *et al*. A coding variant of ANO10, affecting volume regulation of macrophages, is associated with borrelia seropositivity. *Mol Med* 2015; **21**: 26–37.
- 10 Clausnitzer M, Dankel SN, Kim KH, Quon G, Meuleman W, Haugen C *et al*. FTO obesity variant circuitry and adipocyte browning in humans. *N Engl J Med* 2015; **373**: 895–907.
- 11 American Psychiatric Association. *Diagnostic and Statistical Manual of Mental Disorders*, 4th edn, text revision. American Psychiatric Press: Washington, DC, USA, 2000.
- 12 Begemann M, Grube S, Papiol S, Malzahn D, Krampe H, Ribbe K *et al*. Modification of cognitive performance in schizophrenia by complexin 2 gene polymorphisms. *Arch Gen Psychiatry* 2010; **67**: 879–888.
- 13 Ribbe K, Friedrichs H, Begemann M, Grube S, Papiol S, Kastner A *et al*. The cross-sectional GRAS sample: a comprehensive phenotypical data collection of schizophrenic patients. *BMC Psychiatry* 2010; **10**: 91.
- 14 Kastner A, Begemann M, Michel TM, Everts S, Stepniak B, Bach C *et al*. Autism beyond diagnostic categories: characterization of autistic phenotypes in schizophrenia. *BMC Psychiatry* 2015; **15**: 115.

- 15 Volzke H, Alte D, Schmidt CO, Radke D, Lorbeer R, Friedrich N *et al*. Cohort profile: the Study of Health in Pomerania. *Int J Epidemiol* 2011; **40**: 294–307.
- 16 Stepniak B, Kastner A, Poggi G, Mitjans M, Begemann M, Hartmann A *et al*. Accumulated common variants in the broader fragile X gene family modulate autistic phenotypes. *EMBO Mol Med* 2015; **7**: 1565–1579.
- 17 Kay SR, Fiszbein A, Opler LA. The Positive and Negative Syndrome Scale (PANSS) for schizophrenia. *Schizophr Bull* 1987; **13**: 261–276.
- 18 Franke G. *Brief Symptom Inventory (BSI)*. Beltz: Goettingen, Germany, 2000.
- 19 Blom G. *Statistical Estimates and Transformed Beta-Variables*. Wiley: New York, USA, 1958.
- 20 Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An inventory for measuring depression. *Arch Gen Psychiatry* 1961; **4**: 561–571.
- 21 Schulz P, Schlotz W, Becker P. Trierer Inventar zur Erfassung von chronischem Streß (TICS): skalenkonstruktion, teststatistische überprüfung und validierung der skala arbeitsüberlastung. *Diagnostica* 1999; **45**: 8–19.
- 22 Hammer C, Stepniak B, Schneider A, Papiol S, Tantra M, Begemann M *et al*. Neuropsychiatric disease relevance of circulating anti-NMDA receptor autoantibodies depends on blood-brain barrier integrity. *Mol Psychiatry* 2014; **19**: 1143–1149.
- 23 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D *et al*. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–575.
- 24 Fisher RA. The use of multiple measurements in taxonomic problems. *Ann Eugenics* 1936; **7**: 179–188.
- 25 Breiman L. Random forests. *Mach Learn* 2001; **45**: 5–32.
- 26 Liaw A, Wiener M. Classification and regression by random Forest. *R News* 2002; **2/3**: 18–22.
- 27 Bastiaansen JA, Meffert H, Hein S, Huizinga P, Ketelaars C, Pijnenborg M *et al*. Diagnosing autism spectrum disorders in adults: the use of Autism Diagnostic Observation Schedule (ADOS) module 4. *J Autism Dev Disord* 2011; **41**: 1256–1266.
- 28 Ehrenreich H, Nave KA. Phenotype-based genetic association studies (PGAS)-towards understanding the contribution of common genetic variants to schizophrenia subphenotypes. *Genes* 2014; **5**: 97–105.
- 29 Pettersson E, Larsson H, Lichtenstein P. Common psychiatric disorders share the same genetic origin: a multivariate sibling study of the Swedish population. *Mol Psychiatry* 2015; **21**: 717–721.
- 30 Bush WS, Oetjens MT, Crawford DC. Unravelling the human genome-phenome relationship using phenome-wide association studies. *Nat Rev Genet* 2016; **17**: 129–145.
- 31 Tost H, Champagne FA, Meyer-Lindenberg A. Environmental influence in the brain, human welfare and mental health. *Nat Neurosci* 2015; **18**: 1421–1431.



This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>

© The Author(s) 2018

Supplementary Information accompanies the paper on the Molecular Psychiatry website (<http://www.nature.com/mp>)