Psychiatry Research 333 (2024) 115720

Contents lists available at ScienceDirect

Psychiatry Research

journal homepage: www.elsevier.com/locate/psychres

Analysis of genetic diversity in patients with major psychiatric disorders versus healthy controls: A molecular-genetic study of 1698 subjects genotyped for 100 candidate genes (549 SNPs)

H.H. Stassen [a,*,1], S. Bachmann [b,c,d], R. Bridler [e], K. Cattapan [e,f], A.M. Hartmann [g], D. Rujescu [g], E. Seifritz [h], M. Weisbrod [i,j], Chr. Scharfetter [a,2]

[a] Institute for Response-Genetics, Department of Psychiatry, Psychotherapy and Psychosomatics, Psychiatric University Hospital, Zurich CH-8032, Switzerland
[b] Department of Psychiatry, Psychotherapy, and Psychosomatics, University of Halle, Halle D-06112, Germany
[c] Clienia AG, Psychiatric Hospital, Littenheid CH-9573, Switzerland
[d] Department of Psychiatry, Geneva University Hospitals, Thônex CH-1226, Switzerland
[e] Sanatorium Kilchberg, Kilchberg CH-8802, Switzerland
[f] University Hospital of Psychiatry and Psychotherapy, University of Bern, Bern CH-3012, Switzerland
[g] Clinical Division of General Psychiatry, Medical University of Vienna, Wien A-1090, Austria
[h] Department of Psychiatry, Psychotherapy and Psychosomatics, Psychiatric University Hospital, Zurich CH-8032, Switzerland
[i] Department of General Psychiatry, Center of Psychosocial Medicine, University of Heidelberg, Heidelberg D-69115, Germany
[j] SRH Hospital Karlsbad-Langensteinbach, Karlsbad D-76307, Germany

ARTICLE INFO

Keywords:
Vulnerability
Resilience
Gene vectors
Singular genes
Neural nets
Artificial intelligence
Classifiers
Schizophrenia
Depression
Bipolar illness
Schizoaffective disorders
Alzheimer's disease

ABSTRACT

Background: This study analyzed the extent to which irregularities in genetic diversity separate psychiatric patients from healthy controls.
Methods: Genetic diversity was quantified through multidimensional "gene vectors" assembled from 4 to 8 polymorphic SNPs located within each of 100 candidate genes. The number of different genotypic patterns observed per gene was called the gene's "diversity index".
Results: The diversity indices were found to be only weakly correlated with their constituent number of SNPs (20.5 % explained variance), thus suggesting that genetic diversity is an intrinsic gene property that has evolved over the course of evolution. Significant deviations from "normal" diversity values were found for (1) major depression; (2) Alzheimer's disease; and (3) schizoaffective disorders. Almost one third of the genes were correlated with each other, with correlations ranging from 0.0303 to 0.7245.
The central finding of this study was the discovery of "singular genes" characterized by distinctive genotypic patterns that appeared exclusively in patients but not in healthy controls. Neural Nets yielded nonlinear classifiers that correctly identified up to 90 % of patients. Overlaps between diagnostic subgroups on the genotype level suggested that (1) diagnoses-crossing vulnerabilities are likely involved in the pathogenesis of major psychiatric disorders; (2) clinically defined diagnoses may not constitute etiological entities.
Conclusion: Detailed analyses of the variation of genotypic patterns in genes along with the correlation between genes lead to nonlinear classifiers that enable very robust separation between psychiatric patients and healthy controls on the genotype level.

* Corresponding author.
E-mail address: k454910@bli.uzh.ch (H.H. Stassen).

[1] Present address: Institute for Response-Genetics, Psychiatric University Hospital (KPPP), University of Zurich, c/o Sanatorium Kilchberg, Alte Landstrasse 70, CH-8802 Kilchberg, Switzerland. Website: https://ifrg.ch.

[2] This work is dedicated to Professor Dr. Christian Scharfetter, a colleague and good friend over decades, with whom I planned and realized this project. Inconceivably for all of us, he passed away much too early after a short, serious illness. *Hans H. Stassen.*

https://doi.org/10.1016/j.psychres.2024.115720
Received 15 September 2023; Received in revised form 11 December 2023; Accepted 3 January 2024
Available online 10 January 2024

1. Background

There is little proven knowledge about etiology and pathogenesis of psychiatric disorders. Even after 50 years of modern psychiatry, (1) there are no causal treatment options; (2) it is not possible to reliably predict if and when a particular patient will respond to a particular treatment; and (3) in individual cases it is hardly possible to make any reliable prognosis.

As to the genetically predisposed factors postulated to be involved in the pathogenesis of psychiatric disorders, evidence clearly speaks against single causes, as psychiatric disorders aggregate in families, but do not segregate. In particular, psychiatric disorders do not follow simple Mendelian modes of inheritance. No homotypic diagnostic patterns are observed in families with multiple affected subjects. Typically, the clinical diagnoses of first and second degree relatives appear to be independent of the index case's primary diagnosis.

Most importantly, our studies of monozygotic (mz) twins discordant for schizophrenia disorders made it clear that genetically predisposed factors are not a sufficient condition for the development of psychiatric disorders (Braun et al., 2017). Rather, genetics seems to act in the sense of an unspecific "vulnerability", so that the unaffected co-twins of mz twins with schizophrenia may be at an increased risk of developing psychiatric symptoms, but can still do very well in daily life.

The pathogenesis of psychiatric disorders is further obscured by etiological heterogeneity, which suggests that multiple pathways can lead to the same clinical picture. *Eugen Bleuler*, the renowned father of "schizophrenia", already spoke of the "group of schizophrenias" to emphasize that "schizophrenia" does not represent an etiological entity (Bleuler, 1969). The most likely etiological scenario is a complex interplay between multiple, genetically predisposed endogenous factors and multiple exogenous factors that may induce the development of latent disorders by triggering the manifestation of clinically relevant symptoms. Among the exogenous factors, lifestyle, diet, consumption behavior, and physical activity play a prominent role. Inflammation appears to be another major exogenous constituent explaining some 15–25 % of the observed phenotypic variance (Stassen et al., 2021; Wang et al., 2022).

This project did not follow standard genotype-to-phenotype association methods that rely on "psychiatric diagnosis" as phenotype (Dennison et al., 2020; Legge et al., 2021; Levey et al., 2021; Howard et al., 2019; Gordovez and McMahon, 2020), but investigated the extent to which irregularities in genetic diversity might separate patients with major psychiatric disorders from healthy controls, where "genetic diversity" denotes the multitude of genotypic patterns observed with each gene.

Analyses of genetic diversity (GDAs) bring up the problem of hidden population stratification due to admixture of people with different ancestries (Berger et al., 2006; Price et al., 2010; Shi et al., 2021). We addressed this problem by (1) recruiting half of the healthy controls from the patients' unaffected first-degree relatives so that part of patients and controls shared their ancestry; and (2) developing a "natural" model of "biological ethnicity" through cluster analyses of 73 SNPs located within the *CLOCK* gene exhibiting distinctive adaptations of North-South and West-East specifics. Both methods yielded estimates of the amount of genotypic variance that is explainable by hidden population stratification.

The project relied on 100 candidate genes reported in the literature as "possibly" involved in the pathogenesis of psychiatric disorders, and whose genotypic patterns were assessed through 549 SNPs. However, we did not expect any of these genes to be directly linked to a psychiatric disorder, as this would otherwise have been found long ago. As we were interested in significant deviations from "normal" diversity values, as well as in setting up multidimensional genetic vector spaces that represent genetic diversity in a metric model (cf. Stassen et al., 2003), the main selection criterion for candidate genes was the utmost variation in genetic diversity across subjects. On this basis we searched for vulnerability and resilience genes by means of multi-layer Neural Nets (NNs) in combination with methods of Artificial Intelligence (AI). Specifically, we addressed the following questions:

(1) How to reproducibly quantify genetic diversity at a high resolution?
(2) Are there genes for which genetic diversity is reduced in male schizophrenia patients, given the fact that 80 % of male patients have no offspring?
(3) Are there vulnerability and resilience genes whose genotypic patterns can distinguish between psychiatric patients and healthy controls?
(4) To what extent do vulnerability and resilience genes correlate with each other, i.e. are there genotypic patterns that show up more than randomly with each other?

2. Methods

2.1. Data material

Data from patients and controls from five of our previous studies were (1) pooled, (2) coded in a standardized way, and (3) analyzed together. The study details can be found elsewhere (Stassen et al., 2007, 2011, 2021, 2022). Totally 1698 subjects were genotyped for 100 genes[1] and 549 specifically selected SNPs at a missing data rate < 5 % (96 autosomal genes; 1431 psychiatric patients; 267 healthy controls, of which 141 (52.8 %) were unaffected 1st degree relatives of the patients).

The patients had been recruited from the daily admissions at three university hospitals in Switzerland and Germany, and from the daily admissions at two private mental health treatment centers in Switzerland. Selection criterion was a suspected ICD-10 diagnosis of F20 (schizophrenia), F25 (schizoaffective disorders), F31 (bipolar illness), or F32/F33 (major depression). All patients had been informed about the goals of this research project and that they can discontinue participation at any time without giving reasons and without facing any disadvantages from this. All patients had signed a written informed consent.

Psychopathology was assessed by specifically trained interviewers: (1) previous history through the syndrome-oriented 63-item SADS Syndrome Check List SSCL-16 (Endicott and Spitzer, 1978); and (2) response to treatment over up to 5 weeks through either the 30-item Positive and Negative Syndrome Scale PANSS (Kay et al., 1987), or the 17/21-item Hamilton Depression Scale HAM-D (Hamilton, 1960). The study protocol also included the collection of blood samples for serum extraction and DNA isolation (Qiagen: QIAamp Blood Maxi Kit).

A minimum baseline score of at least 21 on the general psychopathology PANSS-G Scale (primary "F2x.x" diagnoses), or of at least 15 on the HAM-D17 Scale (primary "F3x.x" diagnoses), was required at entry into study. The definitive diagnoses were decided by consensus of two experienced senior psychiatrists, with unclear cases being assigned to the residual group "other diagnoses". The late-onset Alzheimer's disease patients [24 males and 51 females of European ancestry; ages 78.3 ± 5.4 years; age-of-onset at 71.9 ± 4.9 years (range 65–85 years)] came from the NIHM (DNA and DSM-4 diagnoses).

The healthy control subjects were recruited either through advertising, or from the patients' unaffected first-degree relatives. Eligibility criteria for the healthy controls were: (1) European descent; (2) between 20 and 70 years of age, males and females; (3) native German speaker; and (4) no history of psychiatric disorders. All control subjects filled out the 63-item Zurich Health Questionnaire "ZHQ" (Kuny and Stassen, 1988). Using ZHQ data, we assigned subjects with a negative history of «consumption behavior», «psychosomatic disturbances», or «impaired mental health», to the residual diagnostic subgroup "other diagnoses". The information on ancestry was based on self-reports only. Details of

[3] For details see supplementary Tables 1, 2.

Table 1
The «Zurich Molecular-Genetic Study of Psychiatric Vulnerability» encompasses 2008 patients hospitalized for major psychiatric disorders along with 464 healthy controls. For this project, 1698 subjects were genotyped for 100 specifically selected genes and 549 polymorphic SNPs located within these genes. Ages are given in years.

Zurich Study of Genetic Diversity in Psychiatry					
	Diagnosis	Sample	Males	Females	Ages
Major Depression	ICD-10: F32/ F33	596	187	409	47.9 ± 13.0
Bipolar Disorders	ICD-10: F31	134	63	71	43.8 ± 14.1
Schizoaffective Disorders	ICD-10: F25	62	35	27	43.4 ± 14.2
Schizophrenia	ICD-10: F20	363	206	157	37.2 ± 11.9
Alzheimer‘s Disease	DSM-4: 290.0	75	24	51	78.3 ± 5.4
Other Diagnoses	n/a	201	84	117	48.8 ± 15.2
Healthy Controls	n/a	267	143	124	48.4 ± 20.3
total		1,698	742	956	

the sample composition are given in Table 1.

Genotyping was performed using the iPLEX assay on the MassARRAY MALDI-TOF mass spectrometer "Sequenom" (Oeth et al., 2009), multiplexed with 40+ separate loci per reaction. This method is based on single base extension (SBE) of SNP specific primers using mass modified ddNTPs. In addition, SBE primer length was used to ensure unambiguous resolution of SNP and alleles. Quality criteria were a sample call rate >80 %, SNP call rate >95 %, and genotypes of CEU Trios in accordance with HapMap database >98 %.

2.2. *Quantifying genetic diversity*

The estimation of the genetic diversity associated with the 100 candidate genes relied on "gene vectors" which were assembled per gene from the genotypes of 4–8 polymorphic SNPs located within each gene. As a SNP can exhibit three different expressions regardless of allele definition, a base-3 system[2] was used to construct gene vectors:

$$\text{“gene vector''}: v_i^{(j)}$$

$$= \sum_{k=1}^{m(j)} s_{ik}^{(j)} 3^{k-1} \qquad \begin{array}{ll} i = 1,2,\cdots N & \text{subjects} \\ j = 1,2,\cdots M & \text{genes} \\ s_{ik}^{(j)} \in \{0,1,2\} & \text{SNPs} \\ m(j) & \text{number of SNPs in the } j-th \text{ gene} \end{array}$$

With *m* SNPs, a total of 3^m different genotypic patterns would be theoretically possible per gene. However, no more than half of the theoretically possible patterns were actually found among the 1698 subjects of this project, due to SNP correlations. As a rule of thumb, an average of 100 different genotypic patterns is expected for a 10-dimensional gene vector made up of five SNPs. Thus, gene vectors assess genetic diversity at an adequate resolution, as plenty of "variation" means plenty of "information". The number of different genotypic patterns per gene was referred to as the gene's "diversity index".

As genetic diversity depends on sample size, we generated a set of calibration data by drawing 32 random samples of equal size from the total sample ($n = 1{,}698$) for each gene, and for 24 sample sizes in steps of 50 between 50 and 1200. By averaging across the 32 random samples, we obtained 100*24 normative distributions for the 100 candidate genes, covering all sample sizes of the diagnostic subgroups within this project. As an estimate of the correlation between two genes j_1 and j_2, we used the maximum frequency among the combinations of genotypic patterns of gene j_1 with gene j_2, divided by the sample size.

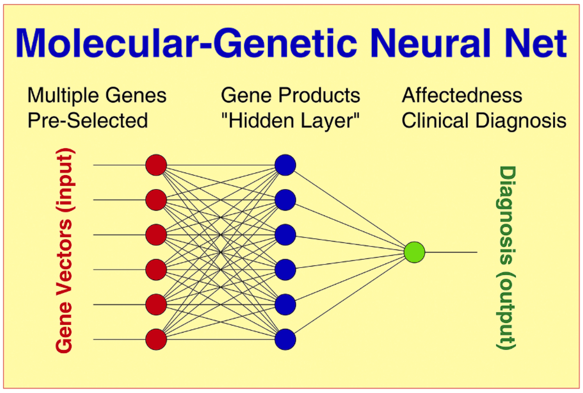

Figure 1. Principal schema of a multilayer Neural Net (NN) where clinical diagnosis (output) results from multiple gene vectors (input) connected to each other by complex interactions via one or more "hidden" layer(s). The NN algorithm iteratively constructs a model that is simultaneously fitted to the observed data of all patients. The achievable goodness of fit depends on the information included, the quality of underlying data, and the number of intermediate layers implemented to model nonlinear interactions.

2.3. *Neural nets and artificial intelligence*

Nonlinear Neural Nets (NN) connect the "neurons" of the input layer (the subjects' gene vectors) with the "neurons" of the output layer (the subjects' psychiatric diagnoses) via "hidden" layers. Our goal was to construct NN models that correctly classified all 1698 subjects in terms of psychiatric diagnoses through their gene vectors. NN connections were realized by (1) weight matrices; and (2) model fitting algorithms minimizing an error function in the weight space ("goodness of fit"). The most popular model fitting strategy, the backpropagation algorithm (Hecht-Nielsen, 1989), looks for the minimum of the error function using the method of gradient descent. The achievable precision of the model essentially depends on the information included, the quality of underlying data, and the number of intermediate layers implemented to model nonlinear interactions (Fig. 1).

Results derived through standard NN approaches, which use 80 % of samples for training and the remaining 20 % for testing tend to be over-optimistic and prone to spurious, non-reproducible results. By contrast, the ***k***-fold cross-validation approach splits the data into ***k*** roughly equal parts, using ***k***-1 partitions for training, while one partition is used for testing. This process is repeated until each partition has served as a testing set, so that ***k*** estimates of prediction errors are generated. The resulting prediction errors are approximately unbiased for the "true" error for sufficiently large ***k*** (***k*** ≈ 10 is a typical value in practice). In consequence, we relied on the ***k***-fold cross-validation strategy with ***k*** = 10 throughout the entire project and applied the well-proven "random walk" strategy in order to distinguish between local and global minima. In this way, spurious and non-reproducible results were effectively eliminated.

We used Artificial Intelligence (AI) methods in order to refine the initial, tentative weights of genes. Genes that showed high diversity indices but little variation across all population subsets were weighted lower than genes that displayed large variation, as little variation means little information and, therefore, a small contribution to discrimination.

The genes' "informativeness" in terms of genetic diversity was estimated by drawing random samples with sample sizes ranging from 3 % to 30 % of the total sample along with tens of thousands of iterations. The algorithm detected subsets of test persons with marked deviations

[4] A base-4 system would make the genotypic patterns much easier for people to read, but at the cost of 25% more memory and a 25% higher computational load.

from “normal” diversity indices. In parallel, the algorithm also looked for genotypic patterns that tended to show up exclusively in patients but not in controls.

2.4. Quantifying biological ethnicity

While GWAS are focusing on global, genome-wide “genetic ancestry” to estimate the probability of a subject to belong to a certain “ancestry group”, the concept of “biological ethnicity” has its focus on hidden population stratifications that arise locally within chromosome segments. The CLOCK gene was chosen because it likely shows distinctive adaptations to typical “North-South” and “West-East” diurnal and seasonal patterns that might not only give rise to population stratification, but might also lead to disruptions of the body’s internal clock (hypothesized to be linked to depression). We used five gene vectors by subdividing the gene into five segments, each with 15 SNPs, along with cluster analyses for the detection of “natural” subgroups inherent in the gene vectors. A principal component analysis prior to cluster analysis eliminated the correlations between the five gene vectors.

2.5. Statistical analyses

We used the *Statistical Analysis Software SAS/STAT 9.4* by SAS Institute Inc. (PROCs: *TTEST, GLM, ACECLUS*, CLUSTER, *FASTCLUS*, MODECLUS, *VARCLUS, PRINCOMP*, and *FACTOR*) along with *PROC HPNEURAL* from *SAS Enterprise Miner 15.1* for Neural Net analyses, complemented by NN and AI programs developed by our institute.

3. Results

3.1. Diversity Index

In this Central European population of 1698 subjects, the diversity indices of the chosen candidate genes ranged from 18 (CYP2C19) to 476 (GPR39), with a mean value of 109.4 ± 82.8. Of the 681 SNPs originally genotyped, we had to exclude 190 SNPs (19.4 %) from subsequent analyses due to a missing data rate that was too high. To avoid possible biases caused by the varying numbers of SNPs within each gene, our plan was to weight genes reciprocally to the constituent number of their SNPs. Contrary to expectations, however, the diversity indices were found to be only weakly correlated with their constituent number of SNPs. In fact, a generalized linear regression model GLM explained no more than 20.5 % of the observed variance, whereas the combined factors chromosome, gene size, and gene position explained 6.63 %. It had therefore to be assumed that genetic diversity, as estimated by diversity indices, is an intrinsic gene property that has evolved over the course of evolution. An illustrative example is given in Fig. 2, where large differences showed up in the comparison between CYP2J2 (diversity index: 69) and SLC6A6 (diversity index: 182), although 5 SNPs were involved in both genes. Given these facts, weighting genes reciprocally to the number of their SNPs appeared to obscure this important gene property, thus being clearly counterproductive. The distribution of the ensemble of diversity indices exhibited two peaks (diversity indices around 70 and 170), along with seven genes exhibiting a diversity index above 250 (Fig. 2).

The 100*24 normative calibration curves, covering all 100 candidate genes and population sizes of this project, displayed a very robust behavior with respect to scattering and, when regarded as a function of sample size, with respect to continuity (Table 2).

This robustness is shown in Fig. 3 where the diversity indices of the two genes *CYP2J2* and *SCL6A6* are plotted for sample sizes between 50 and 1700 in steps of 50. The differences between the two curves regarding shape and steepness indicate different gene types, as *CYP2J2* belongs to the left gene group in Fig. 1 (distribution peak around 70), and *SLC6A6* to the middle gene group (distribution peak around 170).

The validity of the normative calibration curves was verified by comparing males ($n = 742$) with females ($n = 956$) in terms of the diversity indices of 96 autosomal genes. Virtually no differences showed up for any of the genes after correction for sample size. Thus, the amount of variance of between-population differences that was explainable by population size could be reduced to less than 10 %. This enabled us to accurately adjust comparisons between samples in terms of the sample

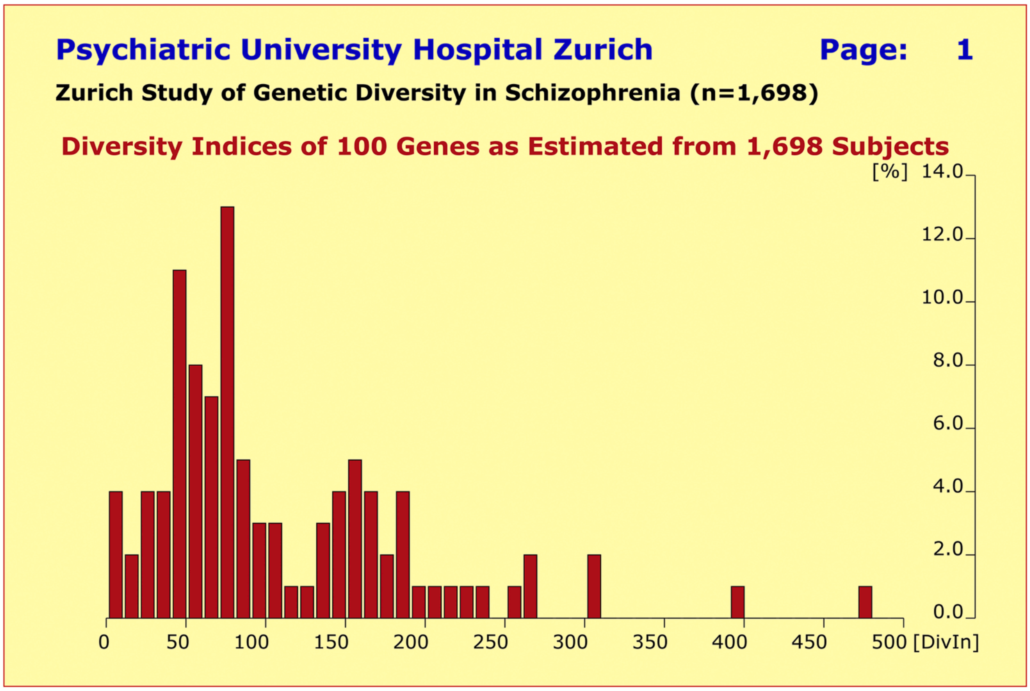

Figure 2. Distribution of the diversity indices of 100 genes as observed in 1,698 Central European subjects (including a small number of U.S. Americans). The diversity index ranged from 18 (*CYP2C19*) to 476 (*GPR39*) with a mean value of 109.4 ± 82.8. The distribution revealed two peaks (diversity indices around 70 and 170), along with 7 genes exhibiting a diversity index above 250. These results may indicate different types of genes.

Table 2
Expected values regarding diversity indices for 10 genes and sample sizes ranging from 100 to 1,000. Due to the well-behaved characteristics of the underlying calibration curves, simple linear interpolation between the sampling points is sufficient to calculate indices for intermediate sample sizes.

Diversity Indices as a Function of Sample Size										
Sample Size:	100	200	300	400	500	600	700	800	900	1000
PRDM2	46.3	68.7	82.4	94.9	104.3	113.0	121.0	127.3	134.9	139.5
OPRD1	30.8	41.6	48.5	55.8	58.9	65.0	69.0	72.8	74.2	77.3
GRIK3	59.2	87.1	105.6	120.1	130.6	141.7	148.7	157.7	165.5	171.1
CYP4B1	23.8	34.5	42.7	47.9	53.8	56.7	60.8	64.4	66.5	69.9
CYP2J2	22.2	29.8	35.1	40.0	42.3	46.5	48.7	52.0	53.4	56.5
ADAR	16.1	21.0	24.1	26.0	28.3	29.9	31.7	32.8	34.2	35.2
APOB	34.8	51.2	60.6	67.5	76.0	82.5	86.1	91.8	95.7	99.3
POMC	17.8	22.5	25.5	28.4	30.5	31.3	33.8	35.2	35.5	37.6
CYP1B1	23.5	31.9	39.4	43.4	47.7	51.9	54.0	59.2	62.2	63.2
GPR39	82.8	145.6	191.4	233.4	265.4	294.7	317.6	343.2	362.6	381.6

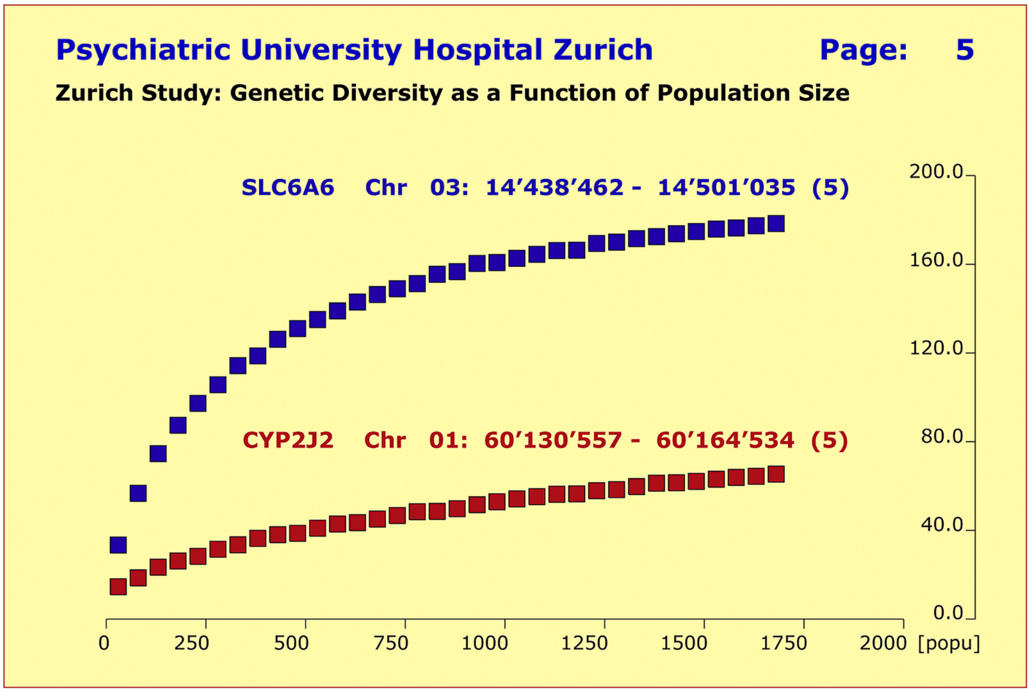

Figure 3. Diversity index as a function of sample size, with sample sizes ranging from 50 to 1,700. Lower half: gene *CYP2J2* on chromosome 1 with **diversity index=69**. *CYP2J2* belongs to the left group of genes in Fig. 1. Upper half: gene SLC6A6 on chromosome 3 with **diversity index=182**. SLC6A6 belongs to the middle group of genes in Fig. 1. The diversity index was determined for both genes from 5 SNPs each. This demonstrates that the diversity index is an intrinsic gene property and only weakly linked to the number of SNPs. All genetic analyses relied on a genetic-physical map derived from *Ensembl* Build 105 of September 25, 2021.

sizes involved. Differences derived through single gene comparisons were generally smaller than expected and did not survive Bonferroni corrections. As an alternative, we relied on the diversity indices of the total sample as reference, computed the differences between total sample and the diagnostic subgroup of interest, and created a “total score” by summing up the differences over the 100 genes. Subsequent t-tests yielded several significant differences: (1) a significant reduction in genetic diversity ($p < 0.0001$) for patients with major depression ($n = 596$); (2) a significant reduction in genetic diversity ($p < 0.0001$) for patients with Alzheimer’s disease ($n = 75$); and (3) a significant increase in genetic diversity ($p < 0.0001$) for patients with schizoaffective disorders ($n = 64$). It is important to note that population size did NOT explain the above deviations, as the deviations pointed in opposite directions for patients with Alzheimer’s disease ($n = 75$) compared to patients with schizoaffective disorders ($n = 64$). The observed deviations were related to a small number of genes, while the majority of genes showed no differences. The hypothesis of a reduction in genetic diversity among male patients with schizophrenia could not be confirmed ($p = 0.0693$).

The contributions of genes to a given phenotype must not be purely additive for a given sample. If, for example, the contribution of one gene G1 to a given phenotype is 15 %, and the contribution of a second gene G2 is 10 %, then the joint contribution of G1 and G2 does not necessarily have to be 25 %, but can be considerably smaller, that is, just 20 %, or so. This is due to the fact that some genotypic pattern p1 from G1 may be linked to a genotypic pattern p2 from G2, such that p1 alone has the same contribution to the phenotype as p1 and p2 together (“redundancy”). The “correlation” between G1 and G2 is a measure of inherent redundancy.

Almost one third of the genes under investigation showed such correlations, ranging from $r = 0.0303$ (*GRIK3/TNF*) to $r = 0.7245$ (*CYP3A5/CYP3A7*), with a mean correlation of 0.1027 ± 0.1025 for the patients with schizophrenia disorders ($n = 363$); of 0.1069 ± 0.1020 for the patients with major depression ($n = 596$); and of 0.1053 ± 0.1021 for the healthy controls ($n = 267$). With $r \leq 0.105$ ($p \approx 0.010$), more than half of the empirically found correlations originated from smaller subsets among the patients of the diagnostic subgroups. The observed correlations were of interest in the context of the envisaged NN analyses, as the NN method evaluates interactions between genes. By contrast, the observed correlations did not provide insights into the biological function or relevance of such interactions.

3.2. Singular genes

The distributions of the genotypic patterns showed no substantial differences between healthy controls and the patients of 5 diagnostic subgroups (Fig. 4a,b,c), with the only exception of several genes among the Alzheimer's patients (Fig. 4c). Although comparisons of single genotypic patterns occasionally reached statistical significance outlasting Bonferroni corrections, the phenotypic variance explained by this was very small and non-additive.

AI-controlled analyses revealed several genes that appeared to be illness-specific, as they exhibited genotypic patterns that showed up exclusively in patients but not in healthy controls. For example, 33.9 % of the schizophrenia group showed genotypic patterns of gene *GPR39* which were completely absent in healthy controls. Similarly, 33.0 % of depressed patients showed genotypic patterns of gene *GRIA1*; 21.8 % of bipolar patients showed genotypic patterns of gene *STAT1*; 25.8 % of schizoaffective patients showed genotypic patterns of gene *ABCB1*; and 18.7 % of Alzheimer's patients showed genotypic patterns of gene

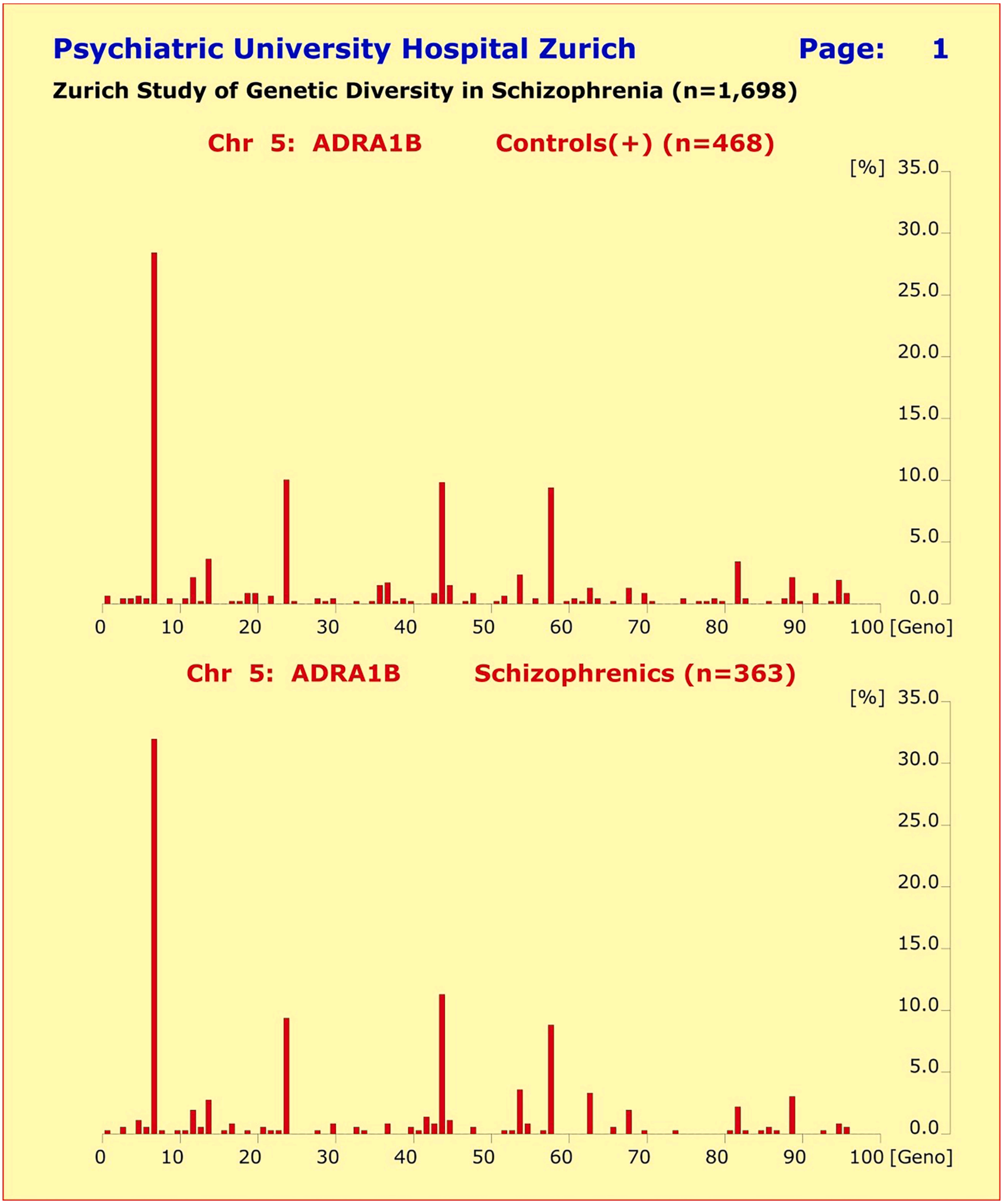

Figure 4a. . The distributions of the genotypic patterns of the genes under study showed no substantial differences between healthy controls (n=468, upper half) and the patients of 4 diagnostic subgroups under study. Shown here is the diagnostic subgroup of "Schizophrenia" (n=363, lower half).

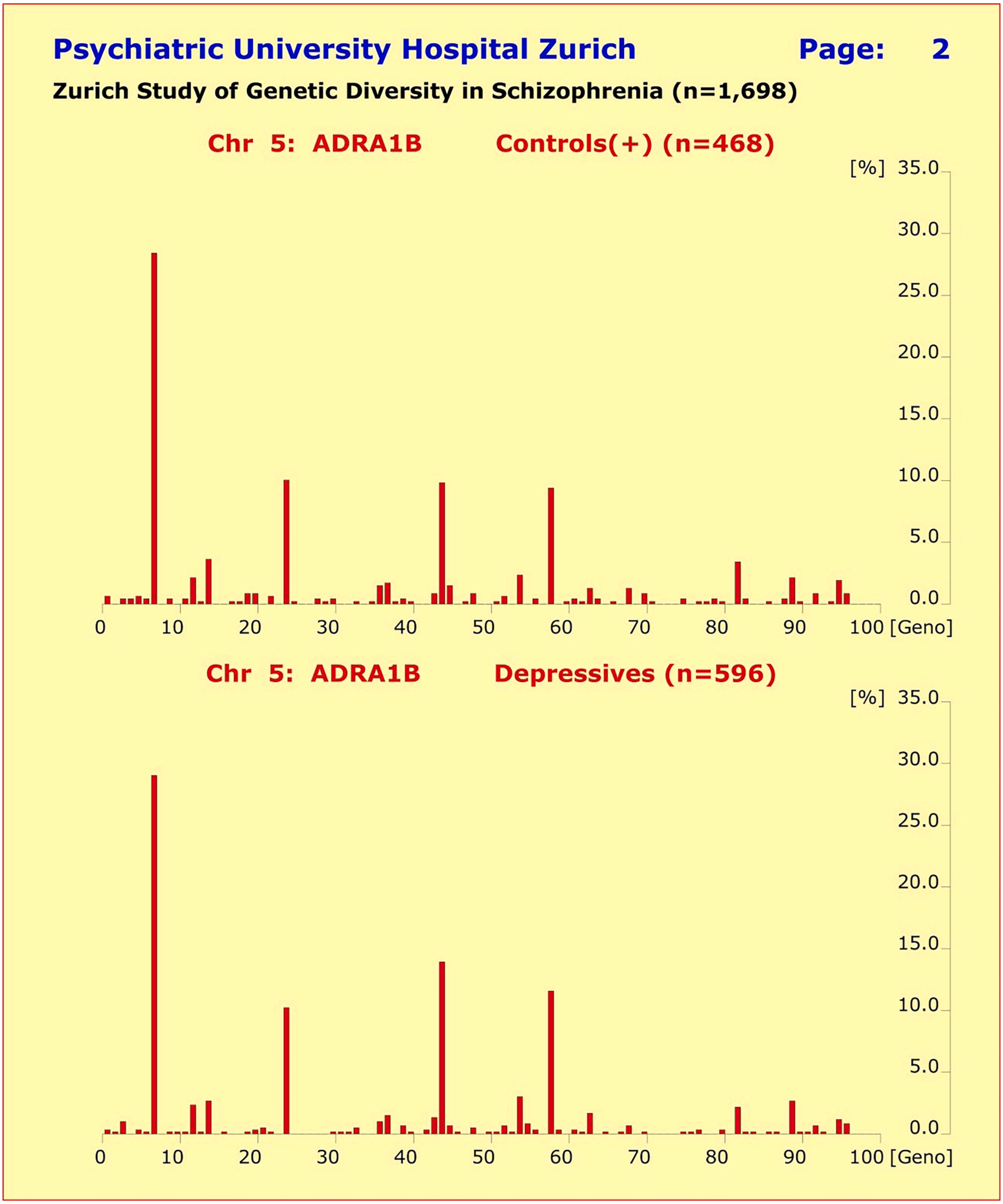

Figure 4b. . The distributions of the genotypic patterns of the genes under study showed no substantial differences between healthy controls (n=468, upper half) and the patients of 4 diagnostic subgroups under study. Displayed here is the diagnostic subgroup of “Depression” (n=596, lower half).

SCL6A1, all of which being completely absent in healthy controls. Because of their distinctive characteristics, these genes were termed “singular genes”. For each diagnostic subgroup, we found some 13–30 singular genes whose genotypic patterns appeared exclusively in at least 10 % of patients but not in healthy controls.

Most of the singular genes had higher than average diversity indices. The number of singular genes did not depend on sample size: (1) a total of 29 singular genes were found in the subgroup of schizophrenia patients ($n = 363$), virtually identical with the 28 singular genes observed in the subgroup of bipolar patients ($n = 134$); whereas (2) just 24 singular genes showed up in the subgroup of depressive patients ($n = 596$), compared to the 33 singular genes found in the much smaller subgroup of schizoaffective patients ($n = 62$) (Table 3).

Even though the diagnostic groups had singular genes in common, the singular genes differed from diagnostic subgroup to diagnostic subgroup in terms of genotypic patterns and intrinsic weights. It was even possible to identify a set of singular genes specific to the differences between schizophrenia and MDD patients. By contrast, we were not successful in finding health-specific “resilience genes”, i.e. genes with genotypic patterns observed in significant numbers among healthy controls but not in patients.

Extending the control group by those 201 cases who did not meet the

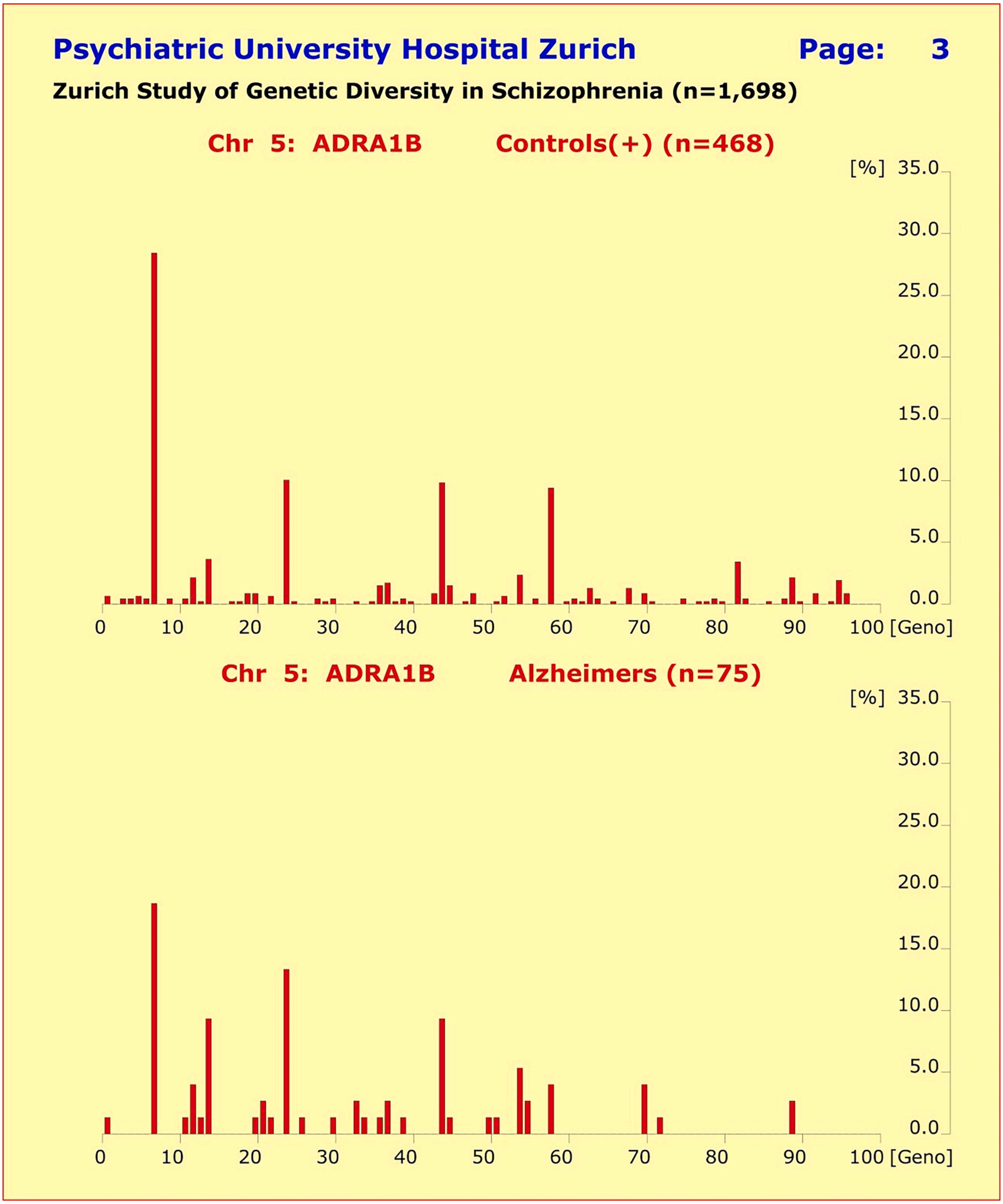

Figure 4c. . The distributions of the genotypic patterns of the genes under study showed no substantial differences between healthy controls (n=468, upper half) and the patients of 4 diagnostic subgroups under study. By contrast, the distribution of the Alzheimer's subgroup (n=75, lower half) exhibited significant deviations from all other ones.

criteria of major psychiatric disorders ("Controls(+)"; $n = 468$), and rerunning the AI-controlled analyses left the results essentially unchanged. Only the number of singular genes reaching significance dropped somewhat (Table 3).

3.3. Neural net analyses

Augmented by the structure-generating a priori knowledge of singular genes, the NN analyses achieved good steady-state results when comparing the diagnostic subgroups with healthy controls. For the subgroups of patients with schizophrenia disorders, major depression, bipolar illness, and schizoaffective disorders, the NN algorithm yielded a rate of about 90 % correctly classified patients along with a 10 % subset of patients labeled as "unknown" (Table 4). This in contrast to (1) the subgroup of patients suffering from Alzheimer's disease which performed with 80 % correctly classified subjects slightly worse; and (2) the conglomerate subgroup of patients "other diagnoses" where the optimization terminated with 40 % of subjects classified as "unknown" (39.8 % false-negative error rate).

The construction of classifiers that separate patients with

Table 3
«Singular genes» denote illness-specific genes for which genotypic patterns inherent in these genes show up exclusively in patients, but not in healthy controls. For each diagnostic subgroup, we found some 13-30 singular genes with frequencies between 10.0 % and 36.4 %. Weakening the clear-cut definition of "healthiness" for the control population (n = 267) by extending it with the 201 patients of our sample without severe psychiatric diagnoses (n = 468) left the results essentially unchanged. Only the number of singular genes reaching significance dropped somewhat in each diagnostic subgroup.

Patients with Major Psychiatric Disorders versus Healthy Controls					
		Controls: n = 267		Controls(+): n = 468	
Diagnosis	Patients	Singular Genes	Percentages	Singular Genes	Percentages
Schizophrenia	363	29	10.2 %–33.9 %	15	10.2 %–32.0 %
Major Depression	596	24	10.3 %–35.2 %	15	10.6 %–29.7 %
Bipolar Disorders	134	28	10.0 %–36.4 %	20	10.4 %–32.1 %
Schizoaffective Disorders	62	33	10.4 %–32.6 %	30	11.3 %–30.6 %
Alzheimer‘s Disease	75	18	10.2 %–23.1 %	13	10.7 %–24.0 %

Table 4
For four target populations, we found in comparisons with health controls a rate of about 90 % correctly classified patients along with a 10 % subgroup labeled as "unknown". The only exception was the subgroup of patients with "Alzheimer's disease" where apparently one or more genes of relevance were missing in the selection of candidate genes.

Neural Net Analysis: Classification of Patients			
Target Population	Control Population	correct	false-positive
Depressives (n = 596)	Healthy Controls (n = 267)	88.6 %	0.0 %
Schizophrenics (n = 363)	Healthy Controls (n = 267)	89.8 %	0.0 %
Bipolars (n = 134)	Healthy Controls (n = 267)	89.6 %	0.0 %
Schizoaffectives (n = 62)	Healthy Controls (n = 267)	90.3 %	0.0 %
Alzheimer's (n = 75)	Healthy Controls (n = 267)	80.0 %	0.0 %
Bipolars (n = 134)	Schizophrenics (n = 363)	81.3 %	4.6 %
Depressives (n = 596)	Schizophrenics (n = 363)	81.7 %	4.9 %
Schizoaffectives (n = 62)	Schizophrenics (n = 363)	80.1 %	5.0 %

schizophrenia disorders from patients with (1) bipolar illness; (2) major depression; or (3) schizoaffective disorders was somewhat less successful, with false-negative error rates of 20 %. In particular, the NN constraint of a clinically desirable false-positive error rate of 0 % could not be upheld and had to be raised to 5 % to achieve useful results. All this indicated considerable genetic overlaps between the diagnostic subgroups in the range of 20 %–25 % (Alzheimer's disease: 15 %). In other words, there were patients with similar vulnerability profiles who have been assigned to different diagnostic categories. Conversely, there was an average of 10 % of patients for whom the vulnerability models derived by NN analyses did not fit at all (Alzheimer's disease: 20 %).

The classifiers derived through NN analyses were composed of 6–10 genes: 4-5 core genes that were common to all classifiers, plus 2–5 accessory genes that depended on the target population (Table 5). The classifiers were non-unique. It was readily possible to exclude 1-2 genes (up to 3 genes) of an optimized classifier and re-run the NN analyses. This replaced the eliminated genes by other compatible genes, so that the modified classifiers achieved similar, only slightly reduced performances.

This redundancy inherent in the classifier genes was due to the correlations between these genes. For example, in the diagnostic subgroup of schizophrenia disorders (n = 363), gene *STAT1* was correlated with genes *CYP3A5, CYP3A7, CYP3A4, CYP1A1, CYP1A2, CYP2B6*, and *CYP2D7*, with correlation coefficients between 0.1377 and 0.2287. And gene *STAT4* was correlated with genes *CYP3A5, CYP3A7*, and *CYP2B6*, with correlation coefficients ranging from 0.1240 to 0.1405, while gene *CYP27A1* was correlated with genes *CYP3A5, CYP3A7, CYP3A4*, and *SLC4A3*, with correlation coefficients between 0.3636 and 0.5840. The results of the other diagnostic subgroups were similar. The virtually ubiquitous interconnectedness of genes was very complex and could not be broken down in a straightforward manner.

Table 5
Classifier genes have been identified by the NN algorithm as contributing to the separation between the diagnostic subgroups and healthy controls. All genetic analyses relied on a genetic-physical map derived from *Ensembl* Build 105 of September 25, 2021.

Classifier Genes				
Gene	Chr	Position	SNPs	Gene function
GRIK3	1	[37′053′646, 37′273′310]	6	Glutamate Ionotropic Receptor Kainate Type Subunit 3
GPR39	2	[132′891′234, 133′110′278]	7	G Protein-Coupled Receptor 39
STAT1	2	[191′543′841, 191′587′662]	6	Signal Transducer And Activator Of Transcription 1
STAT4	2	[191′605′785, 191′726′016]	5	Signal Transducer And Activator Of Transcription 4
SLC6A1	3	[11′007′829, 11′055′169]	6	Solute Carrier Family 6 Member 1
SLC6A6	3	[14′438′462, 14′501′035]	5	Solute Carrier Family 6 Member 6
GRID2	4	[93′455′427, 94′666′318]	5	Glutamate Ionotropic Receptor Delta Type Subunit 2
GRIA1	5	[152′848′630, 153′166′430]	7	Glutamate Ionotropic Receptor AMPA Type Subunit 1
GABRR1	6	[89′949′881, 89′986′927]	6	Gamma-Aminobutyric Acid Type A Receptor Subunit Rho 1
ABCB1	7	[86′976′581, 87′140′076]	8	ATP Binding Cassette Subfamily B Member 1
ADAM22	7	[87′391′265, 87′597′026]	8	ADAM Metallopeptidase Domain 22
NRG1	8	[31′593′683, 32′572′900]	5	Neuregulin 1
CRH	8	[67′248′418, 67′261′291]	7	Corticotropin Releasing Hormone
DBH	9	[135′490′024, 135′513′490]	6	Dopamine Beta-Hydroxylase
GRIA4	11	[104′987′234, 105′355′300]	6	Glutamate Ionotropic Receptor AMPA Type Subunit 4
NCAM	11	[112′576′708, 112′650′283]	6	Neural cell adhesion molecule
GRIK4	11	[120′102′346, 120′358′590]	5	Glutamate Ionotropic Receptor Kainate Type Subunit 4
GABRA5	15	[24′684′562, 24′774′457]	6	Gamma-Aminobutyric Acid Type A Receptor Subunit Alpha 5

3.4. Mental health

Reversing the methodological approach of "separating patients from healthy controls" to "separating healthy controls from patients" by means of NN classifiers did not lead to a useful operationalization of "mental health". Although NN analyses based on healthy controls (n = 267) as target population and patients (n = 1,431) as control population yielded a list of genes with genotypic patterns that occurred only in the target population, the contributions of these genotypic patterns to separation were generally small with no major contributor. Even with 18 genes, no more than 46 % of the healthy control subjects were correctly classified, while 54 % were labeled as “unknown”. Inclusion of further genes led to only marginal improvements, thus suggesting that genetic factors that strengthen resilience among patients and controls might not be detectable in this way.

3.5. Biological ethnicity

The diversity indices of the 5 CLOCK gene segments (made up of 73 SNPs) lay between 148 and 181, thus indicating a sufficient resolution of between-subject similarities and differences. The correlations between the gene segments were above average with values between 0.3609 and 0.4380, so that certain combinations of genotypic patterns across gene segments were more frequent than expected by chance, underlining the good utility of the CLOCK gene for modelling biological ethnicity[3].

Principal component analysis eliminated the correlation between the 5 CLOCK gene segments almost completely. The first two eigenvalues already explained 97.4 % of the observed variance, so that subsequent cluster analyses were carried out solely with the two corresponding eigenvectors. We found 3 clearly separated clusters, but these were unrelated to the status of affectedness and the patients’ clinical diagnoses.

4. Discussion

Unlike standard genotype-to-phenotype association methods with “psychiatric diagnosis” as phenotype (Horwitz et al., 2019; Unal-Aydin et al., 2021), this project explored the extent to which irregularities in genetic diversity separate patients with major psychiatric disorders from healthy controls. Specifically, we searched for distinct traces in the patients’ genotypic patterns caused by the genetic component of psychiatric disorders (Kendler, 2015; Smeland et al., 2020). Key elements were (1) the “gene vectors” assembled from 4–8 polymorphic SNPs located within genes and representing the genes’ distinctive “fingerprints”; (2) the genes’ diversity indices defined through the number of different genotypic patterns observed with each gene; and (3) the quantification of correlations between genes.

The gene vectors resulted from high-precision genotyping with very low missing data rates. As there was no need for statistical imputations (Marchini and Howie, 2010), the overall data quality met very high standards. The data analyses provided a body of quite convincing evidence that genetic diversity is most likely an intrinsic gene property that can be successfully quantified using “gene vectors” and “diversity indices”. As a direct consequence, any set of 4–8 sufficiently polymorphic SNPs located within genes can be expected to yield comparable estimates of genetic diversity. On the other hand, genetic diversity essentially depended on the population under investigation, in other words, on selection and number of subjects drawn from the population. To address the problem of sample size dependence, we constructed normative calibration curves per gene and for sample sizes in the range of 50–1200 by means of a comprehensive random sampling algorithm that systematically evaluated all diagnostic subgroups along with the healthy controls. Once the differences in sample size were compensated for in this way, the amount of variance of between-sample differences that was explainable by sample size could be reduced to less than 10 %.

The normative calibration curves displayed a very robust behavior with respect to scattering and, when regarded as a function of sample size, with respect to continuity. The validity of the normative calibration curves was verified by comparing males ($n = 742$) with females ($n = 956$) where no differences showed up after correction for sample size. Additional support came from the fact that “sample size” did not explain the deviations in genetic diversity from “normal” values as observed for patients with Alzheimer’s disease ($n = 75$) compared to patients with schizoaffective disorders ($n = 64$), as these deviations pointed in opposite directions. And most importantly, the distributions of diversity indices were found to be virtually identical for diagnostic subgroups of quite different size: for example, healthy controls ($n = 267$), depression ($n = 596$), and schizophrenia ($n = 363$) (Fig. 3a, b). This in contrast to the distribution derived from the Alzheimer’s subgroup ($n = 75$) (Fig. 3c). In consequence, the proposed method of approach apparently constituted a sound basis for high-resolution analyses of the variation of genotypic patterns in genes and the correlations between genes (McKinney et al., 2006; Boucher and Jenna, 2013; Moore et al., 2019).

The diversity indices of the diagnostic subgroups under investigation were not homogeneously distributed over the distribution of the total sample ($n = 1698$). Rather, significant deviations from “normal” diversity indices showed up for three diagnostic subgroups: (1) a significant decrease for major depression; (2) a significant decrease for Alzheimer’s disease; and (3) a significant increase for schizoaffective disorders. These deviations were related to a small number of genes, while the majority of genes showed no such differences. If the observed irregularities is a constituent of genetic vulnerability to psychiatric disorders, then the three diagnostic subgroups apparently follow etiologically different vulnerability pathways (Talarico et al., 2022), and schizoaffective depression is different from major depressive disorder, despite clinically similar symptoms.

Detailed analysis of the observed irregularities revealed the existence of singular genes, that is, illness-specific genes for which certain genotypic patterns showed up exclusively in patients, but not in healthy controls. For each of the diagnostic subgroups, we found between 13 and 30 singular genes, where the number was independent of the sample size. It is highly unlikely that the singular genes with their illness-specific characteristics are entirely due to methodological artifacts. It is equally unlikely that the singular genes were mainly the result of hidden population stratifications, since half of the healthy controls were unaffected 1st-degree relatives of the study patients, expected to share a major part of population stratification with their affected 1st degree relatives. Here it is important to note that the use of unaffected 1st-degree relatives as healthy controls leads to a reduction in separation between patients and controls, rather than to an inflation. Re-running the analyses without the unaffected 1st-degree relatives as controls left the configuration of singular genes virtually unchanged. By contrast, attempts to correct for population stratification by ancestry maps derived through principal component analyses were much less powerful (Gaspar and Breen, 2019). Given the distinctive characteristics of singular genes, NN analyses achieved steady-state results of 80 %–90 % correctly classified subjects when comparing diagnostic subgroups with healthy controls. The only exception with a 20 false-negative error rate was the subgroup of Alzheimer’s disease patients. Evidently, genes of critical relevance to Alzheimer’s disease were missing.

The NN classifiers were not unique because significant correlations between the genes caused a certain amount of redundancy. Given this redundancy, it is unlikely that there is a direct causal link between singular genes and psychiatric disorders since then several genes would have to overlap in their causal effects. Rather, the observed illness-specific irregularities might be signs of a latent, cross-diagnosis vulnerability that makes it easier for exogenous factors to trigger the onset of psychiatric disorders, or to weaken the resilience of those affected. In fact, diagnosis-crossing vulnerabilities along with resilience factors may be involved in the pathogenesis of psychiatric disorders. The more so, as the genetic overlap between diagnostic subgroups seems to indicate that the clinically defined diagnoses do not represent biological entities. This is in line with clinical observations: (1) no homotypic diagnostic patterns are observed in families with multiple affected subjects; (2) there appears to be a continuum between affective and psychotic disorders (Stassen et al., 2006); (3) a majority of patients with a clinical diagnosis of schizophrenia also report major depressive symptoms; and (4) the clinical diagnoses of monozygotic twins who both developed a psychiatric disorder can be quite different even though they share the same genome (Braun et al., 2017).

[5] Biological ethnicity has its focus on population stratifications that arise locally within chromosome segments.

It is generally accepted that the body's immune system can be strengthened in a quite straightforward way: get enough sleep, control consumption behavior, care for a balanced diet, and do regular exercises. And best of all, this goes hand in hand with strengthening the body's robustness ("resilience") with regard to physical and mental health problems. The term "resilience" encompasses all those endogenous mechanisms that support and maintain health, thereby showing considerable between-subject differences (Braun et al., 2017).

As part of this project, we explored the idea that there might be an equivalent to the latent "vulnerability" concept revealed by our data. Specifically, we hoped that there might be some protective shield ("resilience") that compensates for the negative effects of vulnerability through a set of singular genes. Contrary to expectations, the data analyses did not readily lead to the envisaged results. In fact, what we experience as "resilience" may refer to something more fundamental, more comprehensive, and more complex compared to the narrowly defined "vulnerability", so that genetic factors that strengthen resilience among patients and controls might not be detectable in this way. In other words: while a useful vulnerability model appears to be well within reach, a comparable resilience model may not exist.

By construction, Genetic Diversity Analyses (GDAs) have a much higher resolution than single SNP approaches because of the great variability inherent in "gene vectors". Therefore, the results of GDAs and Genome-Wide Association Studies (GWAS) are only comparable to a limited extent, as signal detection differs not only quantitatively but also qualitatively. In particular, the "nearest gene" approach of GWAS complicates the cross-comparison of results excessively.

There are 100+ psychiatry-relevant GWAS (Chimusa and Defo, 2022). Reproducibility appears to be the central problem as every study finds something different, even when relying on pretty robust phenotypes like "response to treatment" (Allen and Bishop, 2019). Similarly inconsistent results come from GWAS using endophenotypes (Greenwood et al., 2016). Reproducibility can get compromised because (1) it is difficult to interpret associations: signals with strong associations may be "false-positives" while signals with weak associations may be "false-negatives"; and (2) the phenotypic variance explained by single SNPs is tiny and non-additive, so that GWAS require thousands of cases and controls (Dattani et al., 2022).

Schizophrenia GWAS: The most sophisticated approach to explaining associations detected by GWAS is by the "Schizophrenia Working Group" of the Psychiatric Genetics Consortium (Trubetskoy et al., 2022). This approach relies on a combination of fine-mapping, transcriptomic analysis, and functional genomic annotations. The authors reported 120 prioritized loci distributed over the entire genome (with the exception of chromosome 22): 88 intron-, 16 intergenic-, 4 missense-, 3 regulatory region-, 2 splice donor-, two 3 prime UTR-, two 5 prime UTR-, 1 non-coding transcript exon-, and 2 synonymous variants. The overlap with the GDA results was marginal.

Major depression GWAS: In their meta-analysis of seven cohorts, the "Major Depressive Disorder Working Group" of the Psychiatric Genomics Consortium (Howard et al., 2019) reported 44 prioritized loci distributed over 18 chromosomes: 27 intron-, 12 intergenic-, 2 regulatory region-, one 3 prime UTR-, and 2 non-coding transcript exon variants. The overlap with the GDA results was marginal.

Bipolar disorder GWAS: A review of 15 GWAS yielded a list of 67 loci distributed over 18 chromosomes (Li et al., 2022): 43 intron-, 12 intergenic-, three 3 prime UTR-, two 5 prime UTR-, 3 regulatory region-, 1 splice acceptor-, 1 synonymous-, and 2 non-coding transcript exon variants. The overlap with the GDA results was marginal.

Alzheimer's disease GWAS: In a review article of 3 recent GWAS in comparison to the meta-analysis carried out by the International Genomics of Alzheimer's Project (IGAP), the authors reported 77 loci distributed over 18 chromosomes (Andrews et al., 2020): 41 intron-, 14 intergenic-, 8 regulatory region-, three 3 prime UTR-, 5 missense-, 3 TF binding site-, 1 stop gained-, and 2 non-coding transcript exon variants. The newer GWAS were not independent of each other, yet produced some inconsistent outcomes. There was no overlap with the GDA results.

Excluding the Alzheimer's disease GWAS, there were 4 genes that received the strongest support from the cross comparisons: GRM1, GABBR2, GRIN2A, NRG1, and CACNA1C (did not reach significance in the GDA study). For methodological reasons, the poor overlap of results between GWAS and GDAs could be expected. Far less understandable are the inconsistencies between GWAS. Particularly disillusioning is the fact that GWAS results explain far less than 10 % of phenotypic variance. Therefore, the question arises whether GWAS are the most promising approach to psychiatric genetics.

Given their robustness, the results of this study can undoubtedly be replicated by independent patient samples. At first glance, existing GWAS with large samples of patients and controls appear to be a good basis for replicating our results. However, GWAS typically have relatively high error rates along with high percentages of missing data. This may not be a major problem in single SNP analyses, but can become an unmanageable obstacle in multivariate approaches (Dattani et al., 2022). Another problem arises from the fact that the SNPs of GWAS are fixed and cannot be freely chosen within genes as needed.

5. Conclusions

Multidimensional gene vectors enable high-resolution analyses of the genetic differences between patients and controls, which emerge from the variation of genotypic patterns in genes and from the correlations between genes.

The central finding of this study was the discovery of singular genes with their ability to separate patients from healthy controls. Even though singular genes do not establish a causal link to psychiatric disorders, they constitute clinically significant signs of latent vulnerabilities that make it easier for exogenous factors to trigger the onset of psychiatric disorders. Of particular interest is the genetic overlap between diagnostic subgroups as this indicates that clinically defined diagnoses may not represent etiological entities.

The proposed method of approach may have cleared the way to clinical applications that facilitate the early detection of latent psychiatric disorders among risk cases, so that early interventions can be started before clinically relevant symptoms develop.

6. Limitations

The majority of patients and controls came from Central Europe, so that the variation in biological ethnicity was modest. One must also assume that the classifiers constructed through this sample will not necessarily show the same good performance with ethnically different populations.

Ethics

The studies were approved by the local ethics committees of the Canton of Zurich, the Canton of Thurgau, the University of Heidelberg, and the University of Munich. All participants signed a written informed consent.

CRediT authorship contribution statement

H.H. Stassen: Funding acquisition, Supervision. **S. Bachmann:** Project administration, Supervision. **R. Bridler:** Investigation, Supervision. **K. Cattapan:** Investigation, Project administration. **A.M. Hartmann:** Formal analysis, Methodology. **D. Rujescu:** Investigation, Project administration, Supervision. **E. Seifritz:** Supervision, Validation, Writing – review & editing. **M. Weisbrod:** Investigation, Supervision. **Chr. Scharfetter:** Conceptualization, Funding acquisition, Investigation, Methodology.

Declaration of competing interest

The authors do not have any competing financial interests.

Acknowledgments

This work was funded in part by a grant of the Swiss National Science Foundation (32B0-119885).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.psychres.2024.115720.

References

Allen, J.D., Bishop, J.R., 2019. A systematic review of genome-wide association studies of antipsychotic response. Pharmacogenomics 20 (4), 291–306. https://doi.org/10.2217/pgs-2018-0163.

Andrews, S.J., Fulton-Howard, B., Goate, A., 2020. Interpretation of risk loci from genome-wide association studies of Alzheimer's disease. Lancet Neurol. 19 (4), 326–335. https://doi.org/10.1016/S1474-4422(19)30435-1.

Berger, M., Stassen, H.H., Köhler, K., Krane, V., Mönks, D., Wanner, C., et al., 2006. Hidden population substructures in an apparently homogeneous population bias association studies. Eur. J. Hum. Genet. 14, 236–244. https://doi.org/10.1038/sj.ejhg.5201546.

Bleuler, E., 1969. Dementia Praecox or the Group of Schizophrenias. Translated from German by Joseph Zinkin, International Universities Press Inc., 8th Edition, New York.

Boucher, B., Jenna, S., 2013. Genetic interaction networks: better understand to better predict. Front. Genet. 4, 290. https://doi.org/10.3389/fgene.2013.00290.

Braun, S., Bridler, R., Müller, N., Schwarz, M.J., Seifritz, E., Weisbrod, M., et al., 2017. Inflammatory processes and schizophrenia: two independent lines of evidence from a study of twins discordant and concordant for schizophrenic disorders. Eur. Arch. Psychiatry Clin. Neurosci. 267, 377–389. https://doi.org/10.1007/s00406-017-0792-z.

Chimusa, E.R., Defo, J., 2022. Dissecting meta-analysis in GWAS Era: Bayesian framework for gene/subnetwork-specific meta-analysis. Front. Genet. 13, 838518 https://doi.org/10.3389/fgene.2022.838518.

Dattani, S., Howard, D.M., Lewis, C.M., Sham, P.C., 2022. Clarifying the causes of consistent and inconsistent findings in genetics. Genet. Epidemiol. 46 (7), 372–389. https://doi.org/10.1002/gepi.22459.

Dennison, C.A., Legge, S.E., Pardiñas, A.F., Walters, J.T.R., 2020. Genome-wide association studies in schizophrenia: recent advances, challenges and future perspective. Schizophr. Res. 217, 4–12. https://doi.org/10.1016/j.schres.2019.10.048.

Endicott, J., Spitzer, R.L., 1978. A diagnostic interview: the Schedule for Affective Disorders and Schizophrenia (SADS). Arch. Gen. Psychiatry 35, 837–844. https://doi.org/10.1001/archpsyc.1978.01770310043002.

Gaspar, H.A., Breen, G., 2019. Probabilistic ancestry maps: a method to assess and visualize population substructures in genetics. BMC Bioinfo. 20, 116. https://doi.org/10.1186/s12859-019-2680-1.

Gordovez, F.J.A., McMahon, J.L., 2020. The genetics of bipolar disorder. Mol. Psychiatry 25 (3), 544–559. https://doi.org/10.1038/s41380-019-0634-7.

Greenwood, T.A., Lazzeroni, L.C., Calkins, M.E., Freedman, R., Green, M.F., Gur, R.E., et al., 2016. Genetic assessment of additional endophenotypes from the consortium on the genetics of schizophrenia family study. Schizophr. Res. 170 (1), 30–40. https://doi.org/10.1016/j.schres.2015.11.008.

Hamilton, M., 1960. A rating scale for depression (HAM-D). J. Neurosurg. Psychiat. 23, 56–62. https://doi.org/10.1136/jnnp.23.1.56.

Hecht-Nielsen, R., 1989. Theory of backpropagation neural network. In: Proceedings of the International Joint Conference on Neural Networks, 1. IEEE, pp. 593–611.

Horwitz, T., Lam, K., Chen, Y., Xia, Y., Liu, C., 2019. A decade in psychiatric GWAS research. Mol. Psychiatry 24 (3), 378–389. https://doi.org/10.1038/s41380-018-0055-z.

Howard, D.M., Adams, M.J., Clarke, T.K., Hafferty, J.D., Gibson, J., Shirali, M., et al., 2019. Genome-wide meta-analysis of depression identifies 102 independent variants and highlights the importance of the prefrontal brain regions. Nat. Neurosci. 22 (3), 343–352. https://doi.org/10.1038/s41593-018-0326-7.

Kay, S.R., Fiszbein, A., Opler, L.A., 1987. The Positive and Negative Symptom Scale (PANSS) for Schizophrenia. Schiz. Bull. 13, 261–276. https://doi.org/10.1093/schbul/13.2.261.

Kendler, K.S., 2015. A joint history of the nature of genetic variation and the nature of schizophrenia. Mol.. Psychiatry 20 (1), 77–83. https://doi.org/10.1038/mp.2014.94.

Kuny, S., Stassen, H.H., 1988. The Zurich Health Questionnaire (ZQH) with 63 items assessing «regular exercises», «consumption behavior», «impaired physical health», «psychosomatic disturbances» and «mental health» in the general population. Psychiatric University Hospital Zurich, available in 6 major languages on request (https://ifrg.ch/instruments.php).

Legge, S.E., Santoro, M.L., Periyasamy, S., Okewole, A., Arsalan, A., Kowalec, K., 2021. Genetic architecture of schizophrenia: a review of major advancements. Psychol. Med. 51 (13), 2168–2177. https://doi.org/10.1017/S0033291720005334.

Levey, D.F., Stein, M.B., Wendt, F.R., Pathak, G.A., Zhou, H., Aslan, M., et al., 2021. Bi-ancestral depression GWAS in the Million Veteran Program and meta-analysis in >1.2 million individuals highlight new therapeutic directions. Nat. Neurosci 24 (7), 954–963. https://doi.org/10.1038/s41593-021-00860-2.

Li, M., Li, T., Xiao, X., Chen, J., Hu, Z., Fang, Y., 2022. Phenotypes, mechanisms and therapeutics: insights from bipolar disorder GWAS findings. Mol. Psychiatry 27 (7), 2927–2939. https://doi.org/10.1038/s41380-022-01523-9.

Marchini, J., Howie, B., 2010. Genotype imputation for genome-wide association studies. Nat. Rev. Genet. 11 (7), 499–511. https://doi.org/10.1038/nrg2796.

McKinney, B.A., Reif, D.M., Ritchie, M.D., Moore, J.H., 2006. Machine learning for detecting gene-gene interactions. Appl. Bioinform. 5 (2), 77–88. https://doi.org/10.2165/00822942-200605020-00002.

Moore, J.H., Mackay, T.F.C., Williams, S.M., 2019. Testing the assumptions of parametric linear models: the need for biological data mining in disciplines such as human genetics. BioData Min. 12, 6. https://doi.org/10.1186/s13040-019-0194-z.

Oeth, P., del Mistro, G., Marnellos, G., Shi, T., van den Boom, D., 2009. Qualitative and quantitative genotyping using single base primer extension coupled with matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MassARRAY). Methods Mol. Biol. 578, 307–343. https://doi.org/10.1007/978-1-60327-411-1_20.

Price, A.L., Zaitlen, N.A., Reich, D., Patterson, N., 2010. New approaches to population stratification in genome-wide association studies. Nat. Rev. Genet. 11 (7), 459–463. https://doi.org/10.1038/nrg2813.

Shi, H., Gazal, S., Kanai, M., Koch, E.M., Schoech, A.P., Siewert, K.M., et al., 2021. Population-specific causal disease effect sizes in functionally important regions impacted by selection. Nat. Commun. 12 (1), 1098. https://doi.org/10.1038/s41467-021-21286-1.

Smeland, O.B., Frei, O., Dale, A.M., Andreassen, O.A., 2020. The polygenic architecture of schizophrenia - rethinking pathogenesis and nosology. Nat. Rev. Neurol. 16 (7), 366–379. https://doi.org/10.1038/s41582-020-0364-0.

Stassen, H.H., Hoffmann, K., Scharfetter, C., 2003. Similarity by state/descent and genetic vector spaces: analysis of a longitudinal family study. In: Almasy L, Amos CI, Bailey-Wilson JE, Cantor RM, Jaquish CE, Martinez M, Neuman RJ, Olson JM, Palmer LJ, Rich SS, Spence MA, MacCluer JW (eds) genetic analysis workshop 13: Analysis of longitudinal family data for complex diseases and related risk factors. BMC Genet. 4 (S59), 1–6.

Stassen, H.H., Scharfetter, C., Angst, J., 2006. Functional psychoses —molecular-genetic evidence for a continuum. In: Marneros, A., Akiskal, H.S. (Eds.), The overlap of affective and schizophrenic spectra. Cambridge University Press, pp. 55–78.

Stassen, H.H., Bachmann, S., Bridler, R., Cattapan, K., Herzig, D., Schneeberger, A., et al., 2021. Inflammatory processes linked to major depression & schizophrenic disorders and the effects of polypharmacy in psychiatry: evidence from a longitudinal study of 279 patients under therapy. Eur. Arch. Psychiatry Clin. Neurosci. 271 (3), 507–520. https://doi.org/10.1007/s00406-020-01169-0.

Stassen, H.H., Angst, J., Hell, D., Scharfetter, C., Szegedi, A., 2007. Is there a common resilience mechanism underlying antidepressant drug response? Evidence from 2848 patients. J. Clin. Psychiatry 68 (8), 1195–1205. https://doi.org/10.4088/jcp.v68n0805.

Stassen, H.H., Anghelescu, I.G., Angst, J., Böker, H., Lötscher, K., Rujescu, D., et al., 2011. Predicting response to psychopharmacological treatment. Survey of recent results. Pharmacopsychiatry 44, 263–272. https://doi.org/10.1055/s-0031-1286290.

Stassen, H.H., Bachmann, S., Bridler, R., Cattapan, K., Herzig, D., Schneeberger, A., et al., 2022. Detailing the effects of polypharmacy in psychiatry: longitudinal study of 320 patients hospitalized for depression or Schizophrenia. Eur. Arch. Psychiatry Clin. Neurosci. 272 (4), 603–619. https://doi.org/10.1007/s00406-021-01358-5.

Talarico, F., Costa, G.O., Ota, V.K., Santoro, M.L., Noto, C., Gadelha, A., et al., 2022. Systems-level analysis of genetic variants reveals functional and spatiotemporal context in treatment-resistant Schizophrenia. Mol. Neurobiol. 59 (5), 3170–3182. https://doi.org/10.1007/s12035-022-02794-7.

Trubetskoy, V., et al., 2022. Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature 604 (7906), 502–508. https://doi.org/10.1038/s41586-022-04434-5.

Unal-Aydin, P., Aydin, O., Arslan, A., 2021. Genetic architecture of depression: where do we stand now? Adv. Exp. Med. Biol. 1305, 203–230. https://doi.org/10.1007/978-981-33-6044-0_12.

Wang, Y., Wei, J., Chen, T., Yang, X., Zhao, L., Wang, M., et al., 2022. A whole transcriptome analysis in peripheral blood suggests that energy metabolism and inflammation are involved in major depressive disorder. Front. Psychiatry 13, 907034. https://doi.org/10.3389/fpsyt.2022.907034.