Large-scale composite hypothesis testing for omics analyses.

Annaïg De Walsche[1,2]✉, **Franck Gauthier**[2], **Alain Charcosset**[2], **and Tristan Mary-Huard**[1,2]

[1]MIA Paris-Saclay, INRAE, AgroParisTech, Université Paris-Saclay 91120 Palaiseau, France
[2]Génétique Quantitative et Evolution - Le Moulon, INRAE, CNRS, AgroParisTech, Université Paris-Saclay, 91190 Gif-sur-Yvette, France

Abstract
Composite Hypothesis Testing (CHT) based on summary statistics has become a popular strategy to assess the effect of a same marker (or gene) jointly across multiple traits or at different omics levels. Although significant efforts have been made to develop efficient CHT procedures, most approaches face scalability constraints in terms of the number of traits/omics and markers to handle, or fail to account for potential correlations across traits efficiently. Methods relying on mixture models partially circumvent these limitations, but do not provide proper p-values, hampering the use of classical multiple testing procedures, graphical representations (e.g. Manhattan or QQ plots) and limiting their comparison with alternative approaches.
We introduce the qch_copula approach that combines the mixture model approach with a copula function to account for dependencies across traits/omics. The method comes with a p-value that is consistently defined for any composite hypothesis to be tested. By significantly reducing the memory size burden of the EM algorithm during inference, the method scales to the analysis of several (up to 20) traits and $10^5 - 10^6$ markers. We conducted a comprehensive benchmark study, comparing our approach with 6 state-of-the-art methods recently developed. The qch_copula procedure efficiently controls Type I error rate and yields substantial gain in detecting various patterns of joint associations. The interest of the method is further illustrated by the joint analysis of 14 association studies to detect pleiotropic regions involved in psychiatric disorders.
The proposed method is implemented in the R package `qch`, available on CRAN.

meta-analysis | composite hypothesis testing | mixture model | data integration | pleiotropy
Correspondence: ***annaig.de-walsche@inrae.fr***

Introduction

Consider a study where the goal is to assess the joint effect of a specific drug treatment in two different tissues. One aims at defining a test procedure that rejects hypothesis H_0 «the drug has no joint effect», when both hypotheses H_0^1 «the drug has no effect on tissue 1» and H_0^2 «the drug has no effect on tissue 2» are false. This corresponds to a particular case of composite hypothesis testing where the composite H_0 hypothesis to be tested is $H_0^1 \cup H_0^2$. A popular strategy to perform Composite Hypothesis Testing (CHT) is to combine the test statistics and/or p-values derived for each of the marginal hypotheses H_0^1 and H_0^2 into a single summary statistics. While the p-values related to H_0^1 and H_0^2 can be obtained using common statistical procedures, providing a suitable summary statistic along with a valid rejection rule (that ensures the control of the false positive rate at the required nominal level) for the test of the composite H_0 hypothesis is not straightforward. The particular CHT problem of the form $H_0^1 \cup H_0^2$ was addressed as soon as the early 80s, with the seminal work of (1), and its generalization to the case of testing $H_0^1 \cup \ldots \cup H_0^Q$ with $Q \geq 2$ has been investigated by (2).

In the context of large omics studies CHT has rapidly become a standard procedure. In human genetics, it has been successfully used for mediation analysis (3–6) or for the detection of genomic regions that are associated with two diseases, e.g. Prostate Cancer and Type 2 Diabete (7). In these omics applications CHT is typically performed at the gene or marker level, resulting in a large number of simultaneously tested composite hypotheses. A first consequence is the need for CHT methods to explicitly provide p-values in order to be combined with multiple testing correction procedures. A second consequence is the opportunity to exploit the large amount of available data to model and infer the relationship between the 2 series of p-values, and to account for this relationship in the CHT procedure to better control Type I error rate (7).

More recently, CHT has also been used in the context of integrative genomics to jointly analyse multiple omics traits, such as DNA methylation, copy number variation and gene expression (8, 9). When applied to such studies, CHT typically requires the analysis of $Q > 2$ series of p-values (i.e. one serie per omics dataset) which led to the development of dedicated procedures. Most existing methods are based on a same model where the Q-uplet of p-values associated to each gene/marker is distributed as a multivariate mixture where each of the 2^Q components corresponds to a specific combination of H_0^q and H_1^q states (8–10). While versatile, the mixture approach suffers from several limitations. Although type I error rate can be evaluated through local or multi-class FDR in a mixture model setting (11–13), the approach does not yield p-values, limiting the range of multiple testing procedure to be applied, the use of sanity checks such as QQ-plots or histograms, and the comparison with alternative approaches. Moreover, the number of components of the mixture model restricts the application of method to $Q = 3$ (9) to ≈ 10 (8, 10). Based on the same mixture model approach, we propose a new method for CHT called qch_copula that explicitly accounts for the dependence structure across p-value series through a copula function (14). The new procedure comes with several important features. First we show

how a consistent p-value can be defined from the mixture model approach and show the connection between adaptive Benjamini-Hochberg FDR control (15) applied to these p-values and a local FDR control (11, 12) directly applied to the posteriors obtained from the mixture model. Second, we provide computational shortcuts that significantly alleviate the memory burden of EM inference procedure of the approach, extending its application to series of p-values as large as $n = 10^5/10^6$ genes/markers and $Q = 20$. Third, we show that the coupling of the non-parametric approach of (10) to estimate the marginal H_1^q distributions with the copula modelling for dependency yields excellent performance in terms of detection accuracy and Type I error rate control. Importantly, we provide the first benchmark study for the comparison of six recently developed CHT methods (5–10). It is shown that half of the methods do not correctly control for Type I error rate when the p-value series are correlated.

Materials and methods

In this section, we define the composite hypothesis (Section Composite hypothesis), then introduce our new method for CHT based on a mixture model approach (Section Model) and describe the inference (Section Inference) and the testing procedure (Section Testing composite hypothesis). Section Computational shortcuts specifies the algorithmic optimisations to alleviate the memory burden of inference procedure. The simulation framework used to evaluate the performance of our method and and its competitors is detailed in Section Simulation framework.

Composite hypothesis. Assume a collection of n items (e.g. genes or SNP) have been tested for their effects in Q conditions (e.g. traits, tissues or environments). We denote by H_0^q (resp. H_1^q) the null (resp. alternative) hypothesis corresponding to test q ($1 \leq q \leq Q$) and consider the set $\mathcal{C} := \{0,1\}^Q$ of all possible combinations of null and alternative hypotheses among Q. For a given configuration $c := (c_1, ..., c_Q) \in \mathcal{C}$, the joint hypothesis $\mathcal{H}^c$ is defined as

$$\mathcal{H}^c := \left(\bigcap_{q:c_q=0} H_0^q \right) \cap \left(\bigcap_{q:c_q=1} H_1^q \right)$$

Considering two complementary subsets $\mathcal{C}_0$ and $\mathcal{C}_1$ of $\mathcal{C}$, we define the composite null and alternative hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$ as

$$\mathcal{H}_0 := \bigcup_{c \in \mathcal{C}_0} \mathcal{H}^c, \quad \mathcal{H}_1 := \bigcup_{c \in \mathcal{C}_1} \mathcal{H}^c$$

We aim at testing $\mathcal{H}_0$ versus $\mathcal{H}_1$ for each of the n items.

Model. Let P_i^q be the p-value obtained for test q on item i, and $Z_i^q = -\Phi^{-1}(P_i^q)$ its negative probit transform, where Φ stands for the standard Gaussian cdf. We note $Z_i := (Z_i^1, ..., Z_i^Q)$ the z-score profile of item i.
Each item i is associated with a vector $L_i := (L_i^1, ..., L_i^Q) \in \mathcal{C}$, where L_i^q is the binary variable being 0 if H_{0i}^q holds and 1 if H_{1i}^q holds. Assuming that the items are independent, each z-score profile arises from a mixture model with 2^Q components defined as follows:

$$Z_i \sim \sum_{c \in \mathcal{C}} w_c \psi^c. \tag{1}$$

with ψ^c the distribution of Z_i conditionally on $L_i = c$ and $w_c := \Pr\{L_i = c\}$.
The mixture model (1) involves 2^Q multivariate distributions ψ^c that need to be estimated, which may become quite cumbersome when Q becomes large. To reduce the computational burden, we assume that all cumulative distributions of ψ^c have the following form:

$$\Psi_\theta^c(Z_i) = C_\theta(F_{c_1}^1(Z_i^1), ..., F_{c_q}^q(Z_i^q), ..., F_{c_Q}^Q(Z_i^Q))$$

where F_0^q (resp. F_1^q) is the cumulative distribution of Z_i^q conditional on $L_i^q = 0$ (resp. $L_i^q = 1$) and C_θ is a copula of parameter θ accounting for the dependency structure between the Q z-scores. This yields:

$$\psi_\theta^c(Z_i) = c_\theta\left(F_{c_1}^1(Z_i^1), ..., F_{c_Q}^Q(Z_i^Q)\right) \prod_{q:c_q=0} f_0^q(Z_i^q) \prod_{q:c_q=1} f_1^q(Z_i^q) \tag{2}$$

where f_0^q (resp. f_1^q) is the distribution of Z_i^q conditional on $L_i^q = 0$ (resp. $L_i^q = 1$). Note that in this last expression one has $F_0^q = \Phi$ for all q, so that inferring the emission distributions of the mixture boils down to the estimation of Q univariate cumulative distributions F_1^q and parameter θ.
In practice one needs to choose the form of the copula distribution. Here we considered Gaussian copula as it allows the specification of different correlations for each pair of variables. The density function of the Gaussian copula is given by:

$$c_\theta(u) = |\theta|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\xi_u^T(\theta^{-1} - I)\xi_u\right)$$
$$\text{where } \xi_u = (\Phi^{-1}(u_1), ..., \Phi^{-1}(u_Q)) \quad \forall u \in [0,1]^Q$$

Here θ is the $(Q \times Q)$ correlation matrix associated to the Gaussian copula. Note that setting $\theta = I_Q$ corresponds to the qch method of (10).

Inference. The model parameters are the distributions $f_1^1, ..., f_1^Q$, the copula parameter θ and proportions w_c. Following (8) and (10), the inference procedure is split into two steps:

1. Get an estimate of each alternative distribution f_1^q,
2. Estimate proportions w_c and the copula parameter θ using an EM algorithm, after plug-in the estimates $\hat{f}_1^q$.

Marginal distributions. Combining Model (1) with the definition of the ψ_c Eq. (2), one has

$$Z_i^q \sim \pi_0^q f_0^q + (1 - \pi_0^q) f_1^q, \tag{3}$$

where f_0^q is the distribution of Z_i^q conditional on $L_i^q = 0$ and f_1^q the distribution conditional on $L_i^q = 1$. Since Z_i^q is defined

as the negative probit transform of a p-value, one has $f_0^q = \phi$ for all q's, with ϕ the standard Gaussian density function.
The null proportions can be directly derived by applying the following estimator:

$$\hat{\pi}_0^q = [n(1-\lambda)]^{-1}|\{i : P_i^q > \lambda\}|,$$

where $\lambda \in [0,1]$ is a tuning parameter that can be determined through bootstrap (16). Once the proportions of mixture Eq. (3) are known (or plugged-in after estimation), the alternative distribution f_1^q can be inferred through a non-parametric procedure using a kernel method (12). Both procedures (for the π_0^q and f_1^q estimations) are known to be consistent.

Configuration proportions and copula parameter. The kernel estimates $\hat{f}_1^q$ can be plugged directly into the mixture model, so that the only quantities to infer are the copula parameter θ and the weights w_c. Inference can be efficiently performed using a standard EM. Denoting $u_{ic} = (F_{c_1}(z_i^1), ..., F_{c_Q}(z_i^Q))^T$ and $\xi_{ic} = (\Phi^{-1}(u_{ic}^1), ..., \Phi^{-1}(u_{ic}^Q))^T$ one obtains the following explicit update equations (see Suppl. Mat. section 1 for details):

$$\text{E step:} \qquad \hat{\tau}_{ic} = \widehat{\Pr}\{L_i = c|Z_i;\hat{\theta}\} = \frac{\hat{w}_c \psi_{\hat{\theta}}^c(Z_i)}{\sum_{c' \in \mathcal{C}} \hat{w}_{c'} \psi_{\hat{\theta}}^{c'}(Z_i)}$$

$$\text{M step:} \qquad \hat{w}_c = \frac{1}{n}\sum_i \hat{\tau}_{ic}$$

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^n \sum_{c \in C} \hat{\tau}_{ic}(\xi_{ic}\xi_{ic}^T)$$

Testing composite hypothesis. The existing multivariate methods (8–10) are not strictly valid test procedures since they do not provide a p-value. We present here how consistent p-values can be defined from the mixture model approach.
Let c be the (unknown) configuration of the item under consideration, and $\mathcal{C}_0$ and $\mathcal{C}_1$ be any two complementary subsets of $\mathcal{C}$, and consider testing

$$\mathcal{H}_0 : \{c \in \mathcal{C}_0\} \quad \text{vs} \quad \mathcal{H}_1 : \{c \in \mathcal{C}_1\}$$

The mixture model 1 can be written in the following form:

$$Z \sim W_0\psi^0 + W_1\psi^1 := \psi$$

where $W_0 = \sum_{c \in \mathcal{C}_0} w_c$, $\psi^0 := W_0^{-1}\sum_{c \in \mathcal{C}_0} w_c\psi_c$ (resp. for W_1 and ψ^1). We consider the posterior

$$\begin{aligned}\tau(z) &= \Pr\{L \in \mathcal{C}_1 | Z = z\} \\ &= \sum_{c \in \mathcal{C}_1} \Pr\{L = c|Z = z\} = \frac{W_1\psi^1(z)}{W_0\psi^0(z) + W_1\psi^1(z)}\end{aligned}$$

as a test statistic. The corresponding p-value is then:

$$\begin{aligned}\text{pval(z)} &= \Pr_{\mathcal{H}_0}\{\tau(Z^*) > \tau(z)\} \\ &= \int_{\mathbb{R}^Q} \mathbb{1}_{\{\tau(z^*) \geq \tau(z)\}}\psi^0(Z^*)dz^* \\ &= \frac{1}{W_0}\int_{\mathbb{R}^Q} \mathbb{1}_{\{\tau(z^*) \geq \tau(z)\}}(1-\tau(z^*))\psi(z^*)dz^*\end{aligned}$$

where the last equality comes from the definition of τ. In practice, estimates of posteriors

$$\hat{\tau}_i = \frac{\widehat{W}_1\psi_{\hat{\theta}}^1(z_i)}{\widehat{W}_0\psi_{\hat{\theta}}^0(z_i) + \widehat{W}_1\psi_{\hat{\theta}}^1(z_i)}$$

can be computed for each item i. Using these estimates and approximating the integral by its empirical counterpart, one gets:

$$\widehat{\text{pval}}_i = \frac{1}{n\widehat{W}_0}\sum_{j=1}^n \mathbb{1}_{\{\hat{\tau}_j > \hat{\tau}_i\}}(1-\hat{\tau}_j)$$

Having access to p-values provides several advantages over existing methods that rely on posteriors only, from combining the testing procedure with any correction method for multiple testing to checking the p-value distribution for quality control and providing graphical displays such as Vulcano or Manhattan plots.
The connection between correcting the p-values $\widehat{\text{pval}}_i$ for multiple testing using the adaptive Benjamini Hochberg procedure (15) and the estimation of lFDR from the posteriors that are provided in (11) and (12) is thoroughly established in Suppl. Mat. Section 2.

Computational shortcuts. When jointly analyzing a large number of conditions, the number of configurations $|\mathcal{C}| = 2^Q$ increases exponentially with the number of conditions Q and the memory burden of the procedure can be prohibitive. A classical EM implementation requires the computation and the storage of the matrix $T = (\tau_{ic})$ in step (E), which becomes cumbersome whenever n and/or Q are large: assuming e.g. $n = 10^5$ and $Q = 15$, T requires 26 Go of storage. In step (M), the update of the weights w_c at the iteration $(t+1)$ is given by

$$\hat{w}_c^{(t+1)} = \frac{1}{n}\sum_i \hat{\tau}_{ic} = \frac{1}{n}\sum_i \frac{\hat{w}_c^{(t)}\psi_{\hat{\theta}^{(t)}}^c(Z_i)}{S_i^{(t)}}$$

with $S_i^{(t)} = \sum_{c \in \mathcal{C}} \hat{w}_c^{(t)}\psi_{\hat{\theta}^{(t)}}^c(Z_i)$. To reduce the memory burden, we propose to compute and save only the quantities $S_i^{(t)}, 1 \leq i \leq n$ in step (E). In step (M), the $\hat{w}_c^{(t)}\psi_{\hat{\theta}^{(t)}}^c(Z_i)$ are calculated on the fly for updating the weights $\hat{w}_c^{(t+1)}$. The same procedure is used to update the copula parameter $\hat{\theta}^{(t+1)}$ (See Suppl. Mat. Section 3 for details). This reduces the algorithm memory footprint from $O(n \times 2^Q)$ to $O(n + 2^Q)$, i.e. a reduction from 26 Go to 1 Mo of storage in our previous example, opening the way to applications where $n = 10^5 : 10^6$

and with up to $Q = 20$ conditions. As the shortcut involves the need to compute the $\psi^c_{\hat{\theta}^{(t)}}(Z_i)$ for $i = 1, ..., n$ and $c \in \mathcal{C}$ both to compute the $S_i^{(t)}$ in step (E) and to update the weights $\hat{w}_c^{(t+1)}$ and the copula parameter $\hat{\theta}^{(t+1)}$ in step (M), it results in an increase of the algorithm's complexity. In practice parallelization may be used to circumvent this problem, as implemented in the `qch` package.

Simulation framework. We evaluated and compared the performance of our method with existing methods in various simulated scenarios. Hereafter, we will refer to items and conditions as "markers" and "traits", mimicking a multi-trait genome-wide association study. Each dataset consists of a $n \times Q$ matrix of z-scores generated as follows. Markers are clustered into $K+1$ clusters. Each cluster is characterized by a configuration c, i.e. all markers of cluster k share the same configuration $c_k = (c_{k1}, ..., c_{kQ})$. One has

$$Z_i \sim \mathcal{N}(2c_k, \Sigma_\rho),$$

where Σ_ρ is a $Q \times Q$ correlation matrix where all covariances between traits are set to ρ. Note that $\mu_{iq} = 0$ if H_0^q holds for item i, and $\mu_{iq} = 2$ otherwise. Assuming that matrix Σ_ρ is common to all clusters, a given cluster k is characterized by its associated number of markers and associated configuration vector. We considered different values for Q (2,8,16), two scenarios: $sparse$ or $dense$, and two levels of correlation ($\rho = 0.3, 0.5$). In the $sparse$ scenario, a large proportion (i.e. ≥ 0.9) of the markers were associated with none of the traits, whereas in the dense case, half of the markers were associated with at least one trait. For each setting (i.e., a combination of a number of traits, a correlation level and a scenario), simulations were repeated 25 times. In total, $3 \times 2 \times 2 = 12$ settings were investigated (See Table 4 in Suppl. Mat. for details). All simulations were performed using $n = 10^5$.

We provide the details for generating one dataset in Setting 1: $Q = 2$, $\rho = 0.3$, $sparse$, and the other settings described in Suppl. Mat. Section 4. Twenty clusters of 50 markers associated with a single randomly chosen trait were simulated. Similarly, twenty clusters of 50 markers associated with the two traits were simulated. The process was repeated to generate ten clusters of 100 markers associated with one trait and ten clusters of 100 markers associated with two traits. A last cluster containing all remaining markers corresponding to the complete null configuration, i.e. $c = (0, 0)$, was generated. This last cluster included 96,000 markers, i.e., 96% of the total number of markers.

The data simulation process corresponds to the model described by Equations 1 and 2 with the marginal alternative distributions f_1^q being the Gaussian distribution $\mathcal{N}(2, 1)$ and the copula parameter θ being Σ_ρ. As a result, our simulation framework may be beneficial to the qch_copula, IMIX (9) and Primo methods (8) in terms of detection power, as all these methods rely on the same mixture model Equation 1. However, all methods should ensure a control of Type I error rate at the nominal level.

Method	Sparse	Dense
DACT_Efron	0 (0)	0 (0)
DACT_JC	0 (0)	**0.08 (0.017)**
DACT_NO	0.032 (0.03)	0.014 (0.015)
HDMT	**0.255 (0.018)**	**0.129 (0.008)**
PLACO	**0.148 (0.091)**	0.058 (0.027)
qch	**0.256 (0.018)**	**0.145 (0.008)**
IMIX	**0.595 (0.128)**	**0.222 (0.094)**
Primo	0.046 (0.021)	**0.079 (0.01)**
qch_copula	0.045 (0.021)	0.057 (0.012)

Table 1. False discovery rate of the procedures for the settings 1 and 2 ($Q = 2$,$\rho = 0.3$).
For each procedure the FDR (averaged over 25 runs) are displayed. Numbers in brackets correspond to standard errors. False discovery rate above 0.05 are in boldface..

Results

Simulations. The methodology presented in Section and hereafter called qch_copula was compared to six different methods recently proposed: DACT (5), HDMT (6), PLACO (7), IMIX (9) Primo (8) and qch (10). The first three methods are based on the specific design of a test statistic that combines two p-values for each item and cannot handle a number of conditions Q larger than 2. IMIX, Primo and qch handle the case $Q \geq 3$ and rely on distribution mixtures similar to the one of Equation 1. Furthermore, PLACO, IMIX, Primo and qch_copula explicitly account for dependencies across p-values series. We first compared methods regarding Type I error (T1E) control. The methods that achieved control of T1E at the requested level were further compared in terms of detection power.

Type I error rate control. In settings 1-4, $n = 10^5$ markers were tested for $Q = 2$ traits. The goal was to identify markers associated with both traits. As Primo only provides an estimate of the FDR at a given threshold, we relied on the equivalence between lFDR estimation and p-values based on posteriors proven in Section to derive p-values for each marker for this method. FDR control at a nominal level of 5% was then applied through the adaptive BH procedure to obtain the final list of candidate markers per method. The results are presented in Table 1 for the case $\rho = 0.3$. Significant departures from the nominal T1E level were observed for most methods. As expected, qch_copula represents a significant improvement upon the initial qch method (which assumes conditional independence) in presence of correlations, illustrating the interest in taking dependence into account. More generally, DACT, Primo and qch_copula are the only procedures to successfully control the FDR level whatever the scenario. Similar results were observed when the correlation between studies was high ($\rho = 0.5$, See Table 5 in Suppl. Mat.).

We further investigated TIE control in settings 5-8 and 9-12, where the number of traits was $Q = 8$ and $Q = 16$, respectively, to identify markers associated with at least 2, 4 or 8 traits. Only Primo and qch_copula were considered, as the other methods can only handle cases where $Q = 2$. Although

IMIX could theoretically handle cases with larger values of Q, in practice, the computational burden is prohibitive, and the associated R package does not handle cases where $Q > 3$. The results are displayed in Table 2 for the case $\rho = 0.3$ (settings 5-6 and 9-10) and Table 6 in Suppl. Mat. for $\rho = 0.5$ (settings 7-8 and 11-12). In settings 5-8 ($Q = 8$), both Primo and qch_copula effectively controlled for T1E. Applied to settings 9-12 ($Q = 16$), Primo either ran out of memory or resulted in prohibitive computational cost (>24h). Our method efficiently controlled for T1E in the dense scenario. In the sparse scenario, one could observe slightly inflated T1E, the actual risk being lower than 8% in most cases.

Detection power. The results of the power detection analysis are detailed for settings 5-6 and 9-10, where the number of traits was $Q = 8$ and $Q = 16$, respectively, and $\rho = 0.3$ in Table 2. Overall, qch_copula and Primo yielded high power whatever the tested composite hypothesis. However, qch_copula exhibited a substantial power improvement - up to 45% - over Primo, regardless of the simulation setting. Similar results were observed when the correlation between traits was high ($\rho = 0.5$, settings 7-8 and 11-12, see Table 6 in Suppl. Mat.) and when $Q = 2$ (settings 1-4, not shown). The average computational time associated with qch_copula for $Q = 16$ and $n = 10^5$ was 78 minutes for the model fitting step and ≈1mn for each tested composite hypothesis. qch_copula was run on a computing platform using one thread and 3.225 Go of RAM memory allocated under Ubuntu Linux OS.

Application to the detection of pleiotropic regions. To illustrate our method, we performed a comprehensive analysis of 14 psychiatric disorders using data derived from genetic association studies conducted by the Psychiatric Genomics Consortium, see Section 7 in Suppl. Mat. for details. The initial data consists of p-values obtained for $n = 6,267,062$ to $12,438,502$ single nucleic polymorphisms (SNPs) for the different studies. The analysis aims to identify pleiotropic genomic regions simultaneously associated with multiple disorders. In the initial analysis of (17), the authors applied the MAGMA method (18) to aggregate p-values at the gene level before running their analysis, drastically reducing the size of the data to 26,024 genes. Aggregated data were analysed using PLACO (7) to detect genes associated with two or more disorders, focusing on the top genes detected in at least eight disorders. As the number of traits allowed by PLACO is limited to 2, all pairwise analyses of 2 among the 14 traits were performed. For a given gene, the list of associated disorders corresponded to those found significant in at least one pairwise analysis.

The initial approach suffers from several technical limitations. Firstly, while the aggregation step reduces the computational burden of the CHT procedure, it may lead to significantly less accurate outcomes. This is because some regions (e.g. intergenic regions or genes not listed in the annotation list) may not be represented, but also because the detection resolution is now limited to the gene level. Second, the use of a pairwise approach may yield ambiguous or inconsistent results. For example, gene TMX2 was detected in the ADHD-AN and the AN-SCZ analyses but not in the ADHD-SCZ analysis. Similarly, gene MIR2113, one of the two genes "that are identified in 10 psychiatric disorders", is declared significant in only 26 out of the 45 PLACO pairwise comparisons involving the ten candidate disorders, with Anorexia Nervosa (AN) being identified in only 1 out of its 13 associated pairwise comparisons. Finally, note that aggregating the results of all pairwise analyses does not come with any statistical guarantee regarding false positive control for the final gene list.

The goal of detecting pleiotropic SNPs exhibiting an effect in at least q disorders can be turned into a composite hypothesis testing procedure where the corresponding composite null hypothesis for SNP i is defined as follows:

$$\mathcal{H}_{0i} : \{\text{SNP } i \text{ is associated with at most } q-1 \text{ disorders}\}.$$

Following the initial analysis, we focused on $q \geq 8$ and performed CHT at a nominal type-I error rate $\alpha = 0.05$. We ran qch_copula on the 5,172,884 SNPs common to all analyses and compared the outcomes with those obtained using the PLACO pairwise analyses. The results are displayed in Table 3 (see Suppl. Mat. Fig. 2 for quality control of the p-values distribution and Fig. 3 for the Manhattan plots). Focusing on the composite hypothesis test with $q = 8$, our method identified 1,608 SNPs (see Fig. 1 for their description), corresponding to 28 distinct regions. The regions identified by qch_copula may include more than one gene. Out of the 38 candidate genes detected using the PLACO approach, 35 were also found by our method (i.e. at least one significant SNP within the gene was identified). The three remaining genes NEGR1, TMX2, and C11orf31 that were identified exclusively by PLACO had only 7, 4, and 3 p-values below 0.01 out of 14 obtained from the MAGMA analysis (for each gene), respectively. Since the goal was to detect genes associated with a minimum of eight disorders, one would expect the number of significant MAGMA p-values to be close to eight or higher. Additionally, qch_copula identified eight regions including more than 10 SNPs that were not detected by PLACO (see Suppl. Mat. Table 8 for their description). Three of these eight regions were not detected in the initial analysis because the annotation list contained no gene located in these regions. Among them, the top region identified by qch_copula, located on chromosome 5 contained 338 SNPs (see the corresponding peak on the Manhattan plot in Figure 1, left). Notably, 25 of these SNPs were detected as associated with 11 disorders (see the corresponding initial GWAS -log10(p-values) in Figure 1, right). This region shares an overlap with the gene RP11-6N13.1, which has been recurrently reported to be associated with psychiatric disorders (19–23), including ADHD, ASD, BIP, MDD, SCZ and TS, among others.

Conclusion

We have developed a novel approach called qch_copula for composite hypothesis testing based on a multivariate mixture model. Our method comes with a consistent p-value directly

Q	Scenario	Test	Primo		qch_copula	
			FDR	Power	FDR	Power
8	Dense	at least 2	0.008 (0.001)	0.143 (0.006)	0.051 (0.001)	0.609 (0.003)
		at least 4	0.005 (0.001)	0.126 (0.007)	0.044 (0.002)	0.534 (0.006)
		at least 8	0.048 (0.006)	0.125 (0.007)	0.05 (0.006)	0.186 (0.01)
	Sparse	at least 2	0.012 (0.004)	0.109 (0.008)	0.054 (0.004)	0.388 (0.007)
		at least 4	0.008 (0.004)	0.108 (0.009)	0.053 (0.004)	0.405 (0.014)
		at least 8	0.079 (0.019)	0.092 (0.009)	0.068 (0.023)	0.166 (0.019)
16	Dense	at least 2	-	-	0.052 (0.002)	0.646 (0.004)
		at least 4	-	-	0.057 (0.002)	0.678 (0.006)
		at least 8	-	-	0.057 (0.002)	0.659 (0.007)
	Sparse	at least 2	-	-	0.072 (0.004)	0.528 (0.008)
		at least 4	-	-	0.082 (0.004)	0.626 (0.01)
		at least 8	-	-	0.068 (0.006)	0.568 (0.018)

Table 2. Performance of the procedures Primo and qch_copula for the settings 5-6 ($Q = 8$) and 9-10 ($Q = 16$) with $\rho = 0.3$.
For each procedure the FDR and the detection power (averaged over 25 runs) are displayed. Numbers in brackets correspond to standard errors.

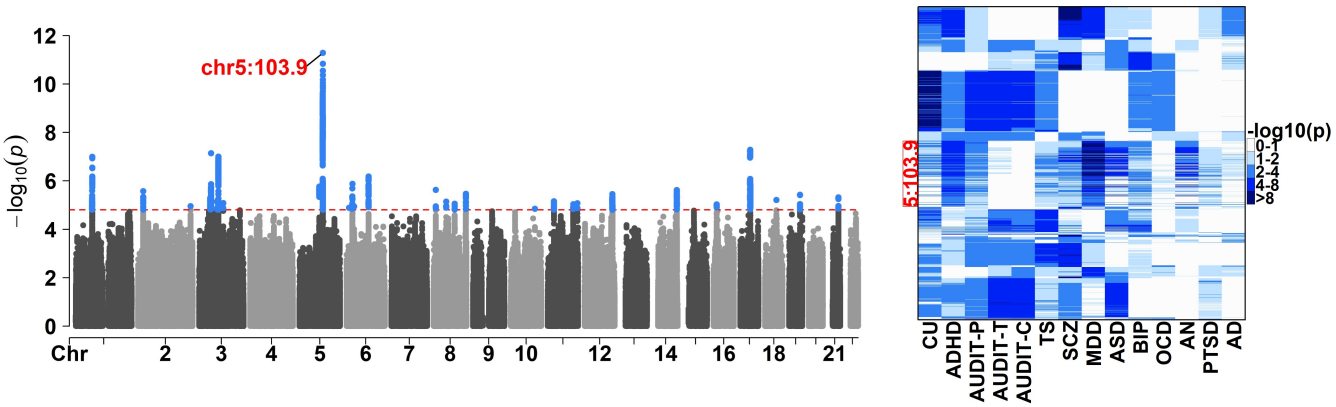

Fig. 1. Results of the composite hypothesis test identifying SNPs associated with at least eight disorders. **Left:** -log10(p-values) of the composite hypothesis test along the chromosomes. Significant SNPs are in blue. **Right:** Initial GWAS -log10(p-values) of the significant SNPs (in rows)

At least # of disorders	8	9	10	11	>11
PLACO	38	21	2	0	0
qch_copula	1,608	498	211	25	0

Table 3. Number of candidate genes and SNPs identified by PLACO and qch_copula, respectively, for different levels of pleiotropy (i.e., associated with at least the specified number of disorders).

defined from the mixture model approach, which, to the best of our knowledge, is the first of its kind. A key feature of qch_copula lies in its ability to model the emission distributions of the mixture components adaptively. Previous published methods rely on fully parametric models, assuming both the f_0^q and f_1^q distributions of Equation 3 to be Gaussian density functions (9), or (a mixture of) χ^2 density functions, after data transformation (8). While such modellings allow one to account for correlations across traits explicitly, they significantly constrain the shape of f_1^q. In comparison, qch_copula performs a non-parametric estimation of the alternative marginal distribution while still accounting for dependencies through the use of the copula function. This adaptive estimation procedure allows for a better fit of the data, resulting in an efficient control of Type I errors and improved detection performance. In practice, the correlation matrix is inferred jointly with the prior proportions within the M step of the EM algorithm, enhancing our ability to capture the dependence more accurately compared to methods like Primo and PLACO, where the correlation matrix is estimated upstream.

Our model assumes that the copula function is the same across all components, employing a single correlation matrix to model the dependence structure. While this may be seen as a limitation, it leads to practical benefits. Modelling the dependency componentwise, as in the IMIX procedure, can lead to a prohibitive computational burden and downgrade the results. This is because some components may be poorly represented in the dataset, compromising the inference of the associated correlation matrices. In contrast, qch_copula strikes a favourable balance between data fitting and computational efficiency in the inference procedure, offering a more practical and effective solution.

The procedure allows the testing of a broad range of composite hypotheses, as one can define any two complementary sets of configurations for $\mathcal{H}_0$ and $\mathcal{H}_1$. Note that a any series of composite hypotheses may be tested in parallel, since the testing step stands independently of the model inference step and does not require parameter re-estimation.

The approach relies exclusively on series of p-values, making it amenable to data integration of studies with diverse outcomes, including continuous, binary, and time-to-event measurements. The exclusive reliance on p-values is also particularly advantageous when direct access to the raw data is constrained due to ethical considerations or data confidentiality.

In conclusion, we strongly believe that the inherent flexibility of composite hypothesis testing has potential uses in various research settings, including multi-population or multi-environment association analyses in quantitative genetics and extended multi-omics data integration. The proposed methodology is implemented in the R package `qch`, available on CRAN.

ACKNOWLEDGEMENTS
The authors are grateful to the INRAE MIGALE bioinformatics facility (MIGALE, INRAE, 2020. Migale bioinformatics Facility, doi:10.15454/1.5572390655343293E12) for providing computing resources. The authors thank Psychiatric Genomics Consortium (PGC) for making the summary statistics of psychiatric disorders GWAS publicly available.

FUNDING
This work was partially supported by INRAE's metaprogram DIGIT-BI0 and by KWS. The funders had no role in the conceptualization, design, data collection, analysis, publication decision, or manuscript preparation.
No competing interest is declared.

Bibliography

1. M. E. Sobel. Asymptotic confidence intervals for indirect effects in structural equation models. *Source: Sociological Methodology*, 13:290–312, 1982.
2. R. L. Berger. Multiparameter hypothesis testing and acceptance sampling. 24:295–300, 1982.
3. Y.-T. Huang. Joint significance tests for mediation effects of socioeconomic adversity on adiposity via epigenetics. 2018.
4. Y.-T. Huang. Genome-wide analyses of sparse mediation effects under composite null hypotheses. 2019.
5. Z. Liu et al. Large-scale hypothesis testing for causal mediation effects with applications in genome-wide epigenetic studies. *Journal of the American Statistical Association*, 2021.
6. J. Y. Dai, J. L. Stanford, and M. LeBlanc. A multiple-testing procedure for high-dimensional mediation hypotheses. *Journal of the American Statistical Association*, 117:198–213, 2022.
7. D. Ray and N. Chatterjee. A powerful method for pleiotropic analysis under composite null hypothesis identifies novel shared loci between type 2 diabetes and prostate cancer. *PLoS Genetics*, 16, 2020.
8. K. J. Gleason et al. Primo: Integration of multiple gwas and omics qtl summary statistics for elucidation of molecular mechanisms of trait-associated snps and detection of pleiotropy in complex traits. *Genome Biology*, 21:1–24, 2020.
9. Z. Wang and P. Wei. Imix: a multivariate mixture model approach to association analysis through multi-omics data integration. *Bioinformatics*, 36:5439–5447, 2021.
10. T. Mary-Huard, S. Das, I. Mukhopadhyay, and S. Robin. Querying multiple sets of p-values through composed hypothesis testing. *Bioinformatics*, 38:141–148, 12 2021.
11. G. J. McLachlan, R. W. Bean, and L. B. T. Jones. A simple implementation of a normal mixture approach to differential gene expression in multiclass microarrays. *Bioinformatics*, 22:1608–1615, 2006.
12. S. Robin, A. Bar-Hen, J. J. Daudin, and L. Pierre. A semi-parametric approach for mixture models: Application to local false discovery rate estimation. *Computational Statistics & Data Analysis*, 51:5483–5493, 2007.
13. T. Mary-Huard, V. Perduca, M.-L. Martin-Magniette, and G. Blanchard. Error rate control for classification rules in multiclass mixture models. *The international journal of biostatistics*, 18(2):381–396, 2022.
14. A. Sklar. Random variables, joint distribution functions, and copulas. *Kybernetika*, 9(6): 449–460, 1973.
15. Y. Benjamini and Y. Hochberg. On the adaptive control of the false discovery rate in multiple testing with independent statistics. *Journal of Educational and Behavioral Statistics*, 25: 60–83, 2000.
16. J. D. Storey, J. E. Taylor, and D. Siegmund. Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach. *J. R. Statist. Soc. B*, 66:187–205, 2004.
17. H. Lu et al. A comprehensive gene-centric pleiotropic association analysis for 14 psychiatric disorders with gwas summary statistics. *BMC Medicine*, 19, 2021.
18. C. A. de Leeuw, J. M. Mooij, T. Heskes, and D. Posthuma. Magma: Generalized gene-set analysis of gwas data. *PLoS Computational Biology*, 2015.
19. P. H. Lee et al. Genomic relationships, novel loci, and pleiotropic mechanisms across eight psychiatric disorders. *Cell*, 179:1469–1482.e11, 12 2019.
20. L. Xia, K. Xia, D. R. Weinberger, and F. Zhang. Common genetic variants shared among five major psychiatric disorders: a large-scale genome-wide combined analysis, 2019.
21. Y. S. Lin, C. C. Wang, and C. Y. Chen. Gwas meta-analysis reveals shared genes and biological pathways between major depressive disorder and insomnia. *Genes*, 12, 10 2021.
22. J. Ward et al. The genomic basis of mood instability: identification of 46 loci in 363,705 uk biobank participants, genetic correlation with psychiatric disorders, and association with gene expression and function. *Molecular Psychiatry*, 25:3091–3099, 11 2020.
23. V. Powell et al. Investigating regions of shared genetic variation in attention deficit/hyperactivity disorder and major depressive disorder: a gwas meta-analysis. *Scientific Reports*, 11, 12 2021.
24. H. Hult and F. Lindskog. Multivariate extremes, aggregation and dependence in elliptical distributions. 34:587–608, 2002.
25. L. Duncan et al. Significant locus and metabolic genetic correlations revealed in genome-wide association study of anorexia nervosa. *American Journal of Psychiatry*, 174:850–858, 2017.
26. T. Otowa et al. Meta-analysis of genome-wide association studies of anxiety disorders. *Molecular psychiatry*, 21:1391, 10 2016.
27. D. Demontis et al. Discovery of the first genome-wide significant risk loci for attention-deficit/hyperactivity disorder. *Nature genetics*, 51:63, 2019.
28. J. Grove et al. Identification of common genetic risk variants for autism spectrum disorder. *Nature genetics*, 51:431, 2019.
29. S. Sanchez-Roige et al. Genome-wide association study meta-analysis of the alcohol use disorders identification test (audit) in two population-based cohorts. *American Journal of Psychiatry*, 176:107–118, 2019.
30. E. A. Stahl et al. Genome-wide association study identifies 30 loci associated with bipolar disorder. *Nature genetics*, 51:793, 2019.
31. J. A. Pasman et al. Gwas of lifetime cannabis use reveals new risk loci, genetic overlap with psychiatric traits, and a causal influence of schizophrenia. *Nature neuroscience*, 21:1161, 2018.
32. N. R. Wray et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nature genetics*, 50:668, 2018.
33. P. D. Arnold et al. Revealing the complex genetic architecture of obsessive-compulsive disorder using meta-analysis. *Molecular Psychiatry*, 23:1181–1181, 2018.
34. C. M. Nievergelt et al. International meta-analysis of ptsd genome-wide association studies identifies sex- and ancestry-specific genetic risk loci. *Nature Communications 2019 10:1*, 10:1–16, 2019.
35. M. Lam et al. Comparative genetic architectures of schizophrenia in east asian and european populations. *Nature genetics*, 51:1670–1678, 2019.
36. D. Yu et al. Interrogating the genetic determinants of tourette's syndrome and other tic disorders through genome-wide association studies. *American Journal of Psychiatry*, 176: 217–227, 2019.

Supplementary Note 1: Estimation of the proportions w_c and the copula parameter θ using an EM algorithm.

Initialisation. The prior proportions are initialized using

$$\hat{w}_c = \prod_{q:c_q=0} \hat{\pi}_0^q \prod_{q:c_q=1} (1-\hat{\pi}_0^q)$$

where $\hat{\pi}_0^q$ corresponds to the marginal prior probability in model 3 estimated in the first step of the inference procedure. Parameter θ is initialized using the inversion of Kendall's tau estimator (24):

$$\hat{\theta}_{ij} = \sin \frac{\pi}{2} \tau_{ij}$$

where τ_{ij} is the Kendall's tau for the random vector $(Z_i, Z_j)^T$.

Maximum likelihood estimate for the Gaussian copula parameter. The complete likelihood of the observations $z_i = (z_i^1, ..., z_i^Q), 1 \leq i \leq n$ and the latent variables $l_{ic} = \mathbb{1}_{\{l_i=c\}}, 1 \leq i \leq n, c \in \mathcal{C}$ is given by:

$$\begin{aligned} L(w,\theta|z,l) &= \prod_{i=1}^{n} \prod_{c\in\mathcal{C}} [w_c \psi_\theta^c(z_i)]^{l_{ic}} \\ &= \prod_{i=1}^{n} \prod_{c\in\mathcal{C}} \left[w_c \left(\prod_{q:c_q=0} \phi(z_i^q) \prod_{q:c_q=1} f_1^q(z_i^q) \right) c_\theta(F_{c_1}(z_i 1^1), ..., F_{c_Q}(z_i^Q))) \right]^{l_{ic}} \end{aligned}$$

Define $Q\left(w,\theta|w^{(t)},\theta^{(t)}\right)$ as the expected value of the complete log-likelihood, with respect to the current conditional distribution of L given Z and the current estimates of the parameters $w^{(t)}$ and $\theta^{(t)}$:

$$Q(w,\theta|w^{(t)},\theta^{(t)}) = \mathbb{E}_{L|Z,w^{(t)},\theta^{(t)}} [\log L(w,\theta|z,l)]$$

We denote $u_{ic} = (F_{c_1}(z_i^1), ..., F_{c_Q}(z_i^Q))$ and $\xi_{ic} = (\Phi^{-1}(u_{ic}^1), ..., \Phi^{-1}(u_{ic}^Q)), 1 \leq i \leq n, c \in \mathcal{C}$. One has:

$$
\begin{aligned}
Q(w,\theta|w^{(t)},\theta^{(t)}) &= \sum_{i=1}^{n}\sum_{c\in\mathcal{C}}\hat{\tau}_{ic}\left[\log(w_c)+\sum_{q:c_q=0}\log(\phi(z_i^q))+\sum_{q:c_q=1}\log(f_1^q(z_i^q))+\log(c_\theta(u_{ic}))\right]\\
&= \sum_{i=1}^{n}\sum_{c\in\mathcal{C}}\hat{\tau}_{ic}\log(c_\theta(u_{ic}))+\text{cste}\\
&= \sum_{i=1}^{n}\sum_{c\in\mathcal{C}}\hat{\tau}_{ic}\log\left[|\theta|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\xi_{ic}^T(\theta^{-1}-I)\xi_{ic}\right)\right]+\text{cste}\\
&= \frac{n}{2}\log(|\theta^{-1}|)-\frac{1}{2}\sum_{i=1}^{n}\sum_{c\in\mathcal{C}}\hat{\tau}_{ic}(\xi_{ic}^T(\theta^{-1}-I)\xi_{ic})+\text{cste}\\
&= \frac{n}{2}\log(|\theta^{-1}|)-\frac{1}{2}\sum_{i=1}^{n}\sum_{c\in\mathcal{C}}\hat{\tau}_{ic}(\xi_{ic}^T\theta^{-1}\xi_{ic})+\text{cste}\\
&= \frac{n}{2}\log(|\theta^{-1}|)-\frac{1}{2}\sum_{i=1}^{n}\sum_{c\in\mathcal{C}}\hat{\tau}_{ic}\text{tr}\left(\xi_{ic}^T\theta^{-1}\xi_{ic}\right)+\text{cste}\\
&= \frac{n}{2}\log(|\theta^{-1}|)-\frac{1}{2}\sum_{i=1}^{n}\sum_{c\in\mathcal{C}}\hat{\tau}_{ic}\text{tr}\left(\xi_{ic}\xi_{ic}^T\theta^{-1}\right)+\text{cste}\\
&= \frac{n}{2}\log(|\theta^{-1}|)-\frac{1}{2}\text{tr}\left(\sum_{i=1}^{n}\sum_{c\in\mathcal{C}}\hat{\tau}_{ic}(\xi_{ic}\xi_{ic}^T)\theta^{-1}\right)+\text{cste}\\
&= \frac{n}{2}\log(|\Lambda|)-\frac{1}{2}\text{tr}\,(S\Lambda)+\text{cste}
\end{aligned}
$$

where $\hat{\tau}_{ic}=\frac{w_c^{(t)}\psi^c_{\theta^{(t)}}(Z_i)}{\sum_{c\in\mathcal{C}}w_c^{(t)}\psi^c_{\theta^{(t)}}(Z_i)}$, $\Lambda=\theta^{-1}$ and $S=\sum_{i=1}^n\sum_{c\in\mathcal{C}}\hat{\tau}_{ic}(\xi_{ic}\xi_{ic}^T)$.
Differentiating the log-likelihood with respect to $\Lambda=\theta^{-1}$ leads to

$$
\begin{aligned}
\frac{\delta}{\delta\Lambda}\log L(\theta|x) &= \frac{n}{2}\frac{\delta}{\delta\Lambda}\log(|\Lambda|)-\frac{1}{2}\frac{\delta}{\delta\Lambda}\text{tr}\,(S\Lambda)\\
&= \frac{n}{2}(\Lambda^{-1})^T-\frac{1}{2}S^T\\
&= \frac{n}{2}\theta^T-\frac{1}{2}\left(\sum_{i=1}^{n}\sum_{c\in C}\hat{\tau}_{ic}(\xi_{ic}\xi_{ic}^T)\right)^T
\end{aligned}
$$

Setting the derivative at 0, one obtains

$$
\frac{\delta}{\delta\Lambda}\log L(\theta|x)=0\Leftrightarrow\hat{\theta}=\frac{1}{n}\sum_{i=1}^{n}\sum_{c\in C}\hat{\tau}_{ic}(\xi_{ic}\xi_{ic}^T) \tag{4}
$$

Note that the explicit expression of $\widehat{\theta}$ obtained in Eq. (4) guarantees that the estimated matrix is symmetric, but not necessarily a positive-definite one with value 1 on the diagonal. In practice, one needs to check that $\hat{\theta}$ is properly a correlation matrix at each (M) step of the EM procedure, and whenever needed correct the initial estimate by i) setting all negative eigenvalues at 0 then ii) scaling the resulting covariance matrix into a correlation matrix.

Supplementary Note 2: Equivalence between Benjamini-Hochberg correction on p-values and the estimation of lFDR from the posteriors

In the classical setting of the test procedure, one can control the FDR at a nominal level of α by applying the Benjamini-Hochberg (BH) procedure. Assume n test were performed and let $p_{(1)} \leq ... \leq p_{(n)}$ be the set of ordered p-values. Define the value r such as

$$r = \arg\max_k \left\{ k \,/\, \frac{n_0 p_{(k)}}{k} \leq \alpha \right\}.$$

The set of rejected hypotheses is then $S_{BH} = \{i_{(1)}, ..., i_{(r)}\}$, where $i_{(k)}$ is the index associated to p-value $p_{(k)}$. Here n_0 is the number of items for which $\mathcal{H}_0$ holds, which is usually unknown. In practice, n_0 can be upper bounded by n or adaptively obtained from the data, which will be designated hereafter by the adaptive BH procedure(15).

Alternatively, in the setting of a mixture model with two components 0 and 1 with emission distributions ψ_0, ψ_1 and priors W_0, W_1, one can define the posterior

$$\tau(z) = \frac{W_1 \psi_1(z)}{W_0 \psi_0(z) + W_1 \psi_1(z)},$$

The local FDR can then be approximately controlled at a nominal level of α by applying the following empirical procedure (11, 12). Assume n items were generated under the previous mixture model and let $\tau_{(1)} \leq ... \leq \tau_{(n)}$ be the set of ordered posteriors. Define the value m such as

$$m = \arg\max_k \left\{ k \,/\, \frac{1}{k} \sum_{1=1}^{k} (1 - \tau_{(i)}) \leq \alpha \right\}.$$

The set of items classified into class 1 is then $S_{lFDR} = \{i_{(1)}, ..., i_{(m)}\}$, where $i_{(k)}$ is the index associated to posterior $\tau_{(k)}$.

The connection between the two approaches in the context of CHT is established by the next proposition:

Proposition 1: Consider n observations drawn according to mixture model (1) of Section 2.2, and let $\hat{\tau}_i$ and $\widehat{\text{pval}}_i$ be the posterior and p-value associated to item i, respectively. Consider the adaptive BH procedure where n_0 is estimated by $n\hat{W}_0$, where $\hat{W}_0 = \sum_{c \in \mathcal{C}_0} \hat{w}_c$. Then $S_{BH} = S_{lFDR}$.

Proof:

For a given item i, the p-value corresponding to the composite hypothesis test is constructed as follows:

$$\widehat{\text{pval}}_i = \frac{1}{nW_0} \sum_{j=1}^{n} \mathbb{1}_{\{\hat{\tau}_j > \hat{\tau}_i\}} (1 - \hat{\tau}_j) \tag{5}$$

Let $\widehat{\text{pval}}_{(1)} < ... < \widehat{\text{pval}}_{(n)}$ represent the ordered set of p-values, and $i_{(k)}$ is the index associated to p-value $\widehat{\text{pval}}_{(k)}$.

According to the adaptive BH procedure, the estimate of the False Discovery Rate when the set of rejected hypotheses is $S_{BH} = \{i_{(1)}, ..., i_{(k)}\}$ is given by:

$$\widehat{FDR}(k) = \frac{\hat{n}_0}{k} \widehat{\text{pval}}_{(k)} \tag{6}$$

with $\hat{n}_0$ an estimate of the number of true null hypotheses. Here, one can estimate the number of null hypotheses as $\hat{n}_0 = nW_0$. Hence, we have:

$$\begin{aligned}
\widehat{FDR}(k) &= \frac{nW_0}{k} \widehat{\text{pval}}_{(k)} \\
&= \frac{nW_0}{k} \frac{1}{nW_0} \sum_{j=1}^{n} \mathbb{1}_{\{\hat{\tau}_j > \hat{\tau}_k\}} (1 - \hat{\tau}_j) \\
&= \frac{1}{k} \sum_{j=1}^{n} \mathbb{1}_{\{\hat{\tau}_j > \hat{\tau}_k\}} (1 - \hat{\tau}_j)
\end{aligned}$$

By the definition of the p-value 5, $\widehat{\text{pval}}_i < \widehat{\text{pval}}_j \Leftrightarrow \hat{\tau}_j > \hat{\tau}_i$.

Consequently:

$$\begin{aligned}\widehat{FDR}(k) &= \frac{1}{k}\sum_{j=1}^{n}\mathbb{1}_{\{\widehat{\text{pval}}_j<\widehat{\text{pval}}_{(k)}\}}(1-\hat{\tau}_j)\\ &= 1-\frac{1}{k}\sum_{j=1}^{n}\mathbb{1}_{\{\hat{\tau}_j>\hat{\tau}_k\}}\hat{\tau}_j\\ &= \widehat{lFDR}(\hat{\tau}_k)\end{aligned}$$

where $\widehat{lFDR}(\hat{\tau}_k)$ is the estimate of the false discovery rate when thresholding at the level $\hat{\tau}_k$ defined by (11) and (12). Thus, applying the adaptive BH correction with $\hat{n}_0 = nW_0$ to p-value Eq. (5) is equivalent to using the local FDR rejection rule defined by (11) and (12). ■

Supplementary Note 3: Algorithm for the implicit EM

Here we present how the burden associated to the storage of the posterior matrix can be avoided in the EM algorithm. Quantities in red correspond to the ones that are actually stored in place of the posterior matrix for later use.

Algorithm 1 Proportions w_c and copula parameter θ inference using EM with memory handling

Inputs: f0Mat $\in \mathcal{M}_{n\times Q}$ containing the $f_0(Z_i^q)$.
f1Mat $\in \mathcal{M}_{n\times Q}$ containing the $\hat{f}_1^q(Z_i^q)$.
$\pi_0 \in \mathbb{R}^Q$ containing the proportions $\hat{\pi}_0^q$.

$\hat{w}_c^{(t)} \leftarrow \prod_{q:c_q=0} \pi_0^q \prod_{q:c_q=1}(1-\pi_0^q)$ ▷ initialisation
$\hat{\theta}^{(t)} \leftarrow$ Kendall's tau estimate

while $||\hat{w}_c^{(t)} - \hat{w}_c^{(t+1)}|| > \epsilon$ or $||\hat{\theta}^{(t)} - \hat{\theta}^{(t+1)}|| > \epsilon$ **do**
  Implicit step E:
  for i in $1:n$ **do**
    $S_i^{(t)} \leftarrow 0$
    for $c \in \mathcal{C}$ **do**
      $S_i^{(t)} \leftarrow S_i^{(t)} + \hat{w}_c^{(t)} \psi^c_{\hat{\theta}^{(t)}}(Z_i)$ ▷ computation of $\psi^c_{\hat{\theta}^{(t)}}(Z_i)$ on the fly
    end for
  end for

  Step M:
  for $c \in \mathcal{C}$ **do**
    $\hat{w}_c^{(t+1)} \leftarrow 0$
    for i in $1:n$ **do**
      $\hat{w}_c^{(t+1)} \leftarrow \hat{w}_c^{(t+1)} + \frac{\hat{w}_c^{(t)} \psi^c_{\hat{\theta}^{(t)}}(Z_i)}{S_i^{(t)}}$ ▷ computation of $\psi^c_{\hat{\theta}^{(t)}}(Z_i)$ on the fly
    end for
    $\hat{w}_c^{(t+1)} \leftarrow \frac{1}{n}\hat{w}_c^{(t+1)}$
  end for

  $\hat{\theta}^{(t+1)} \leftarrow 0$
  for i in $1:n$ **do**
    for $c \in \mathcal{C}$ **do**
      $\hat{\theta}^{(t+1)} \leftarrow \hat{\theta}^{(t+1)} + \hat{\tau}_{ic}(\xi_{ic}\xi_{ic}^T)$ ▷ computation of $\psi^c_{\hat{\theta}^{(t)}}(Z_i)$ on the fly
    end for
  end for
  $\frac{1}{n}\hat{\theta}^{(t+1)} \leftarrow \hat{\theta}^{(t+1)}$
  $t \leftarrow t+1$
end while

Outputs: $\hat{w}_c \in \mathbb{R}^{2^Q}$ containing the configuration proportions estimates.
$\hat{\theta}$ the estimate of the copula parameter.

Supplementary Note 4: Simulation framework

Setting	Q	Correlation	Scenario
1	2	0.3	sparse
2			dense
3		0.5	sparse
4			dense
5	8	0.3	sparse
6			dense
7		0.5	sparse
8			dense
9	16	0.3	sparse
10			dense
11		0.5	sparse
12			dense

Table 4. Description of the 12 settings considered for the simulation study.

Details for generating each setting. Each dataset consists of a $n \times Q$ matrix of Z-scores generated as follows. Markers are clustered into $K+1$ clusters. Each cluster is characterized by a configuration c, i.e. all markers of cluster k share the same configuration $c_k = (c_{k1}, ..., c_{kQ})$. One has

$$Z_i \sim \mathcal{N}(2c_k, \Sigma_\rho),$$

where Σ_ρ is a $Q \times Q$ correlation matrix where all covariances between traits are set to ρ. Assuming that matrix Σ_ρ is common to all clusters, a given cluster k is characterized by its associated number of markers and associated configuration vector. We considered different values for Q (2,8,16), two scenarios: $sparse$ or $dense$, and two levels of correlation ($\rho = 0.3, 0.5$). Denoting $n_{50,\tilde{q}}$ and $n_{100,\tilde{q}}$ the number of clusters of 50 and 100 markers respectively, which are associated with exactly $\tilde{q}$ randomly chosen traits, the details for generating a dataset for each simulation setting are presented below.

Settings 1 to 4: $Q = 2$***.***

- Sparse scenario: $n_{50,1} = n_{50,2} = 20$ and $n_{100,1} = n_{100,2} = 10$, such that $w_{00} = 0.96$, $w_{01} + w_{10} = 0.02$ and $w_{11} = 0.02$.

- Dense scenario: $n_{50,1} = 400$, $n_{100,1} = 200$ and $n_{50,2} = 100$, $n_{100,2} = 50$, such that $w_{00} = 0.5$, $w_{01} + w_{10} = 0.4$ and $w_{11} = 0.1$.

Settings 5 to 8: $Q = 8$***.***

- Sparse scenario: $n_{50,j} = 20$ and $n_{100,j} = 10$ for $j \in \{1,2,4,6,8\}$, such that $w_0 = 0.90$ for the markers being associated with none of the traits and $w_j = 0.02$ for the markers being associated with exactly $j \in \{1,2,4,6,8\}$ traits.

- Dense scenario: $n_{50,j} = 100$ and $n_{100,j} = 50$ for $j \in \{1,2,4,6,8\}$, such that $w_0 = 0.50$ for the markers being associated with none of the traits and $w_j = 0.1$ for the markers being associated with exactly $j \in \{1,2,4,6,8\}$ traits.

Settings 9 to 12: $Q = 16$***.***

- Sparse scenario: $n_{50,q} = 20$ and $n_{100,q} = 10$ for $q \in \{1,2,4\}$, $n_{50,\tilde{q}} = 10$ and $n_{100,\tilde{q}} = 5$ for $\tilde{q} \in \{6,8,14,16\}$, such that $w_0 = 0.90$ for the markers being associated with none of the traits, $w_q = 0.02$ for the markers being associated with exactly $q \in \{1,2,4\}$ traits, and $w_{\tilde{q}} = 0.01$ for the markers being associated with exactly $\tilde{q} \in \{6,8,14,16\}$ traits.

- Dense scenario: $n_{50,q} = 100$ and $n_{100,q} = 50$ for $q \in \{1,2,4\}$, $n_{50,\tilde{q}} = 50$ and $n_{100,\tilde{q}} = 25$ for $\tilde{q} \in \{6,8,14,16\}$, such that $w_0 = 0.5$ for the markers being associated with none of the traits, $w_q = 0.1$ for the markers being associated with exactly $q \in \{1,2,4\}$ traits, and $w_{\tilde{q}} = 0.05$ for the markers being associated with exactly $\tilde{q} \in \{6,8,14,16\}$ traits.

Supplementary Note 5: Evaluation of the FDR control for composite hypothesis testing of two analysis

Method	Sparse	Dense
DACT_Efron	0 (0)	0 (0)
DACT_JC	0 (0)	0.026 (0.018)
DACT_NO	**0.104 (0.033)**	0.055 (0.013)
HDMT	**0.41 (0.017)**	**0.374 (0.015)**
PLACO	**0.3 (0.083)**	0.052 (0.025)
QCH	**0.437 (0.015)**	**0.239 (0.006)**
IMIX	**0.871 (0.021)**	**0.295 (0.079)**
Primo	**0.081 (0.022)**	**0.112 (0.009)**
QCH_copula	0.057 (0.025)	0.061 (0.008)

Table 5. False discovery rate of the procedures for settings 3 and 4 ($Q = 2$, $\rho = 0.5$).
For each procedure the FDR (averaged over 25 runs) are displayed. Numbers in brackets correspond to standard errors. False discovery rate above 0.05 are in boldface.

Supplementary Note 6: Detection power in large-scale composite hypothesis testing

			Primo		qch_copula	
Q	Scenario	Test	FDR	Power	FDR	Power
8	Dense	at least 2	0.005 (0.001)	0.144 (0.004)	0.058 (0.002)	0.617 (0.004)
		at least 4	0.006 (0.001)	0.147 (0.005)	0.064 (0.003)	0.549 (0.005)
		all	0.075 (0.006)	0.178 (0.006)	0.104 (0.007)	0.28 (0.006)
	Sparse	at least 2	0.008 (0.003)	0.115 (0.01)	0.063 (0.005)	0.422 (0.007)
		at least 4	0.01 (0.004)	0.12 (0.011)	0.067 (0.007)	0.415 (0.011)
		all	0.142 (0.031)	0.09 (0.01)	0.104 (0.023)	0.119 (0.012)
16	Dense	at least 2	-	-	0.057 (0.002)	0.722 (0.004)
		at least 4	-	-	0.067 (0.002)	0.756 (0.004)
		at least 8	-	-	0.083 (0.002)	0.681 (0.005)
		all	-	-	0.124 (0.011)	0.23 (0.014)
	Sparse	at least 2	-	-	0.08 (0.007)	0.647 (0.007)
		at least 4	-	-	0.094 (0.008)	0.724 (0.009)
		at least 8	-	-	0.1 (0.008)	0.582 (0.014)
		all	-	-	0.172 (0.062)	0.096 (0.023)

Table 6. Performance of the procedures Primo and qch_copula for the settings 7-8 ($Q = 8$) and 11-12 ($Q = 16$) with $\rho = 0.5$.
For each procedure the FDR and the detection power (averaged over 25 runs) are displayed. Numbers in brackets correspond to standard errors.

Supplementary Note 7: Application to the detection of pleiotropic regions

Information of 14 psychiatric disorders analyzed in this study. We analyzed 14 psychiatric disorders obtained from the Psychiatric Genomics Consortium (PGC) [1]. The disorders studied are anorexia nervosa (AN), anxiety disorder (AD), attention-deficit/hyperactivity disorder (ADHD), autism spectrum disorder (ASD), alcohol use disorders (AUDIT-T alcohol use traits identification test based on total score, AUDIT-C alcohol use traits identification test based on consumption, AUDIT-P alcohol use traits identification test based on problematic consequences of drinking), bipolar disorder (BIP), cannabis use (CU[2]), major depression disorder (MDD), obsessive-compulsive disorder (OCD), post-traumatic stress disorder (PTSD), schizophrenia (SCZ) and Tourette's syndrome (TS). Table 7 details the information about each study. For all the disorders, we obtained their European-only summary statistics and performed the following quality control:
(i) excluded non-biallelic SNPs and those with strand-ambiguous alleles;
(ii) removed duplicated SNPs.

Traits	N(case/control)	PGC study	Reference
AN	14,477 (3,495/10,982)	an2017	(25)
AD	17,526 (5,761/11,765)	anx2016	(26)
ADHD	53,293 (19,099/34,194)	adhd2019	(27)
ASD	46,350 (18,381/27,969)	asd2019	(28)
AUDIT-T	121,604	sud2019-alcuse	(29)
AUDIT-C	121,604	sud2019-alcuse	(29)
AUDIT-P	121,604	sud2019-alcuse	(29)
BIP	51,710 (20,352/31,358)	bip2019	(30)
CU	184,765 (53,180/131,585)	-	(31)
MDD	480,359 (135,458/344,901)	mdd2018	(32)
OCD	9,725 (2,688/7,037)	ocd2018	(33)
PTSD	174,659 (23,212/151,447)	ptsd2019	(34)
SCZ	77,096 (33,640/43,456)	scz2019asi	(35)
TS	14,307 (4,819/9,488)	ts2019	(36)

Table 7. Summary information of 14 psychiatric disorders analyzed in this study.

[1] as available from the PGC website: https://pgc.unc.edu/for-researchers/download-results/
[2] available from the following website: https://drive.google.com/drive/folders/1uFBeUwE-0UA0YmLTVQ9Ym_oU2HkHPdhk

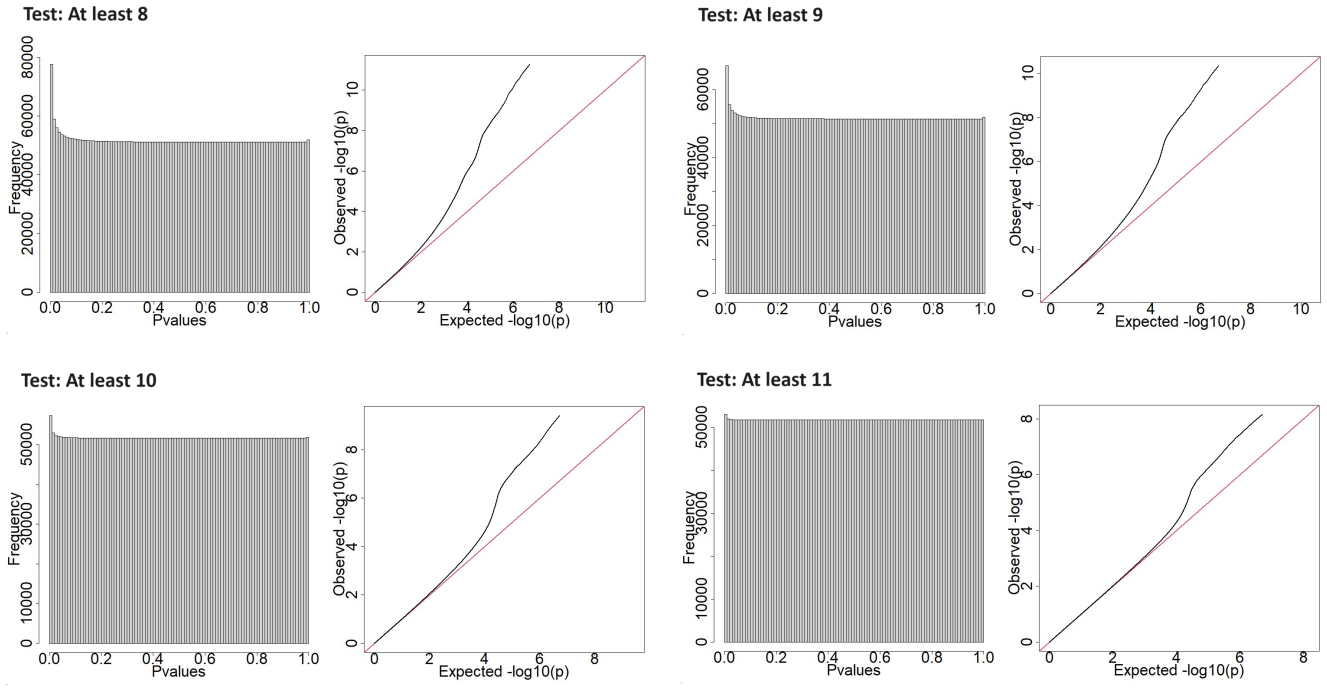

Fig. 2. Quality control of the p-values distribution corresponding of the different composite hypothesis tests. **Left:** Histogram of the qch_copula p-values. **Right:** Quantile-Quantile plot of the qch_copula p-values.

Chr	Start-Stop	Nb SNPs	Top SNP	Pval top SNP
1	73843468-73991792	156	rs11210259	1.013e-07
2	**22558876-22606275**	14	rs11688810	2.665e-06
3	52605005-53101580	90	rs11922961	7.273e-08
5	**103671867-104092037**	338	rs30266	5.162e-12
6	25975883-26339131	12	rs115810	4.773e-06
8	143309504-143352779	11	rs4458978	3.416e-06
11	**27634373-27679916**	13	rs11030084	6.959e-06
11	113285916-113305314	16	rs2734837	9.245e-06

Table 8. Summary information of the 8 regions associated with at least eight disorders specifically found by qch_copula.
The regions not included in the initial analysis due to the absence of any gene in the annotation list are denoted in boldface.

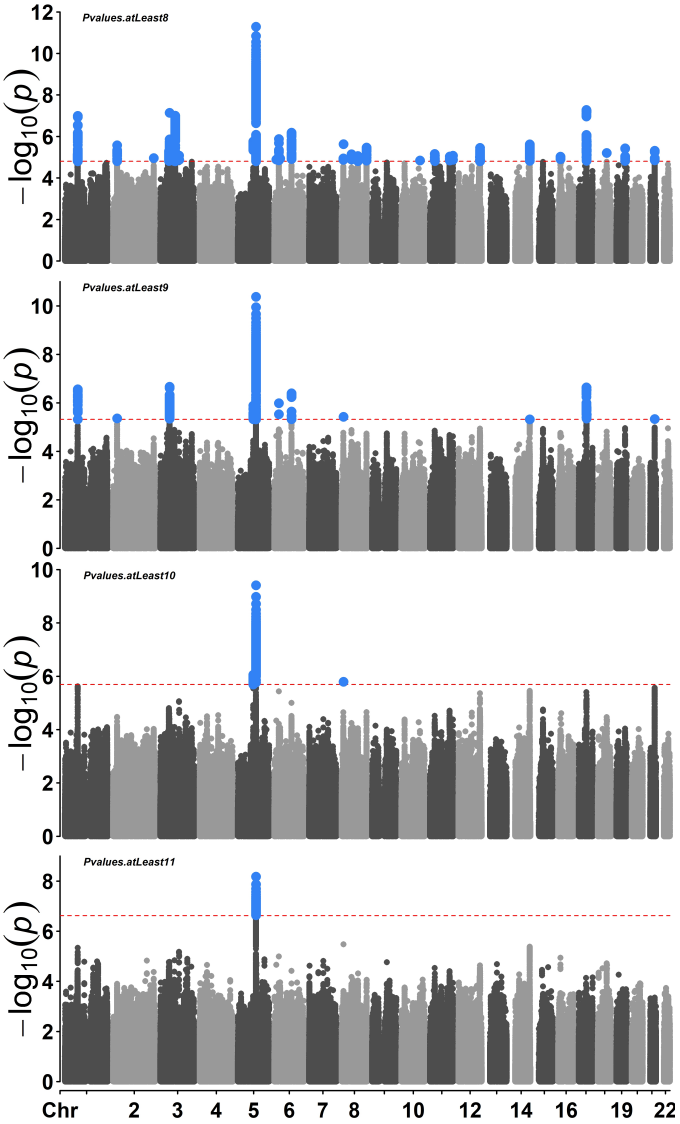

Fig. 3. -log10(p-values) of the different composite hypothesis tests along the chromosomes. Significant SNPs are in blue.