A new method for identifying causal genes of schizophrenia and anti-tuberculosis drug-induced hepatotoxicity

Tao Huang1,2, Cheng-Lin Liu1, Lin-Lin Li3, Mei-Hong Cai1, Wen-Zhong Chen4, Yi-Feng Xu1,4, Paul F. O’Reilly5, Lei Cai1,4 & Lin He1,4

Schizophrenia (SCZ) may cause tuberculosis, the treatments for which can induce anti-tuberculosis drug-induced hepatotoxicity (ATDH) and SCZ-like disorders. To date, the causal genes of both SCZ and ATDH are unknown. To identify them, we proposed a new network-based method by integrating network random walk with restart algorithm, gene set enrichment analysis, and hypergeometric test; using this method, we identified 500 common causal genes. For gene validation, we created a regularly updated online database ATDH-SCZgenes and conducted a systematic meta-analysis of the association of each gene with either disease. Till now, only GSTM1 and GSTT1 have been well studied with respect to both diseases; and a total of 23 high-quality association studies were collected for the current meta-analysis validation. Finally, the GSTM1 present genotype was confirmed to be significantly associated with both ATDH (Odds Ratio (OR): 0.71, 95% confidence interval (CI): 0.56–0.90, P = 0.005) and SCZ (OR: 0.78, 95% CI: 0.66–0.92, P = 0.004) according to the random-effect model. Furthermore, these significant results were supported by “moderate” evidence according to the Venice criteria. Our findings indicate that GSTM1 may be a causal gene of both ATDH and SCZ, although further validation pertaining to other genes, such as CYP2E1 or DRD2, is necessary.

Schizophrenia (SCZ) is a severe, disabling, and chronic mental disorder that affects approximately 1% of the general population1. Recent research has identified a potential link between SCZ and an increased risk of contacting tuberculosis (TB)2, a devastating infectious disease that remains the major leading cause of death worldwide3. The most commonly used anti-TB drugs, although effective4, can induce anti-tuberculosis drug-induced hepatotoxicity (ATDH), the most frequent and serious side effect of TB therapy. ATDH, also known as anti-tuberculosis drug-induced liver injury (ATDILI)5, is caused by the drugs’ reactive metabolites rather than direct toxicity and can impede scheduled treatment, leading to increased complications morbidity and reduced treatment compliance6. Moreover, anti-TB drugs can also induce or cause relapse of psychiatric disorders, including SCZ-like disorders7–9.

Few previous studies have addressed the potential relationship between ATDH and SCZ, which belong to two different disease categories. Although we previously proposed that GST genes represent a link between ATDILI and SCZ10, the molecular basis underlying the connection between ATDH and SCZ is currently unclear. Therefore, identification of common causal genes underlying both diseases would greatly benefit our understanding of the relationship between ATDH and SCZ. In addition, knowledge about this link would also be beneficial to people at risk of developing SCZ and/or TB, particularly in the context of anti-TB drug therapy.

1Bio-X Institutes, Key Laboratory for the Genetics of Developmental and Neuropsychiatric Disorders (Ministry of Education), Shanghai Key Laboratory of Psychotic Disorders (No. 13dz2260500), Shanghai Jiaotong University, Shanghai 200030, China. 2Institute of Health Sciences, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China. 3School of Life Sciences and Biotechnology, Shanghai Jiaotong University, Shanghai, 200240, China. 4Shanghai Mental Health Center, Shanghai Jiaotong University, Shanghai, 200240, China. 5MRC SGDP Centre, Institute of Psychiatry, Psychology and Neuroscience, King’s College London, London WC2R 2LS, United Kingdom. Correspondence and requests for materials should be addressed to L.C. (email: cailei2010@126.com) or L.H. (email: helinhe123@yeah.net)
Table 1. The shared key causal genes for ATDH and SCZ.

| Genes   | Number of neighbors | Number of neighbors that are common disease | FDR corrected P   |
|---------|---------------------|--------------------------------------------|------------------|
| GSTM1   | 48                  | 33                                         | 5.61E-22         |
| CYP2E1  | 70                  | 36                                         | 4.54E-18         |
| GSTT1   | 13                  | 11                                         | 3.87E-09         |

The identification of all genes related to both ATDH and SCZ and subsequent evaluation of their biological roles in vitro and in vivo comprise one possible strategy for elucidating their molecular mechanisms. However, this approach represents an insurmountable challenge for the large-scale identification of disease-related genes. Therefore, we propose a new network-based pipeline for analyzing the connection between these two diseases and prioritizing the identification of potential common key causal genes that influence the development of these two diseases by integrating network random walk with restart (RWR) algorithm, gene set enrichment analysis (GSEA), and hypergeometric test.

Network-based analyses, which assume that neighboring genes have similar effects, have demonstrated great promise for the identification of causal factors and key driver genes associated with a particular disease. However, in our proposed novel network-based analysis pipeline to identify candidate key causal genes of both ATDH and SCZ, we first applied RWR to expand the known ATDH- and SCZ-related genes, for which little overlap had previously been detected. RWR simulates a random walker on the network starting from seed genes and moving toward randomly chosen interacting neighbors at each step. After walking many steps, the probability of moving to each node on the network tends to stabilize, and novel candidate disease genes can be identified. A GSEA like method was then adopted to find the peak of a running sum curve, which indicates the stop point for RWR expansion. RWR expansion identified many common genes shared by the expanded ATDH and SCZ gene lists. Subsequently, a hypergeometric test was used to test all genes on the network to determine whether neighboring genes were also significantly enriched among the genes common to both ATDH and SCZ. A positive result indicated a key gene driver that could affect both ATDH- and SCZ-related genes.

Additionally, we conducted a field synopsis or systemic meta-analysis of published studies to assess whether these key gene drivers represent causal genes common to both ATDH and SCZ by validating the associations between gene polymorphisms and ATDH or SCZ. Systemic meta-analysis is a reliable approach that involves pooling both statistically significant and non-significant results from individual studies to generate a more precise conclusion. In the end, 500 candidate genes were detected as common causal ones for both ATDH and SCZ using our new method. Among these genes, the glutathione S-transferase Mu-1 gene (GSTM1) was validated as a causal factor of these two diseases. Furthermore, we continue to update the validation of other genes in our newly created online database ATDH-SCZgenes (http://www.bio-x.cn/ATDH-SCZgenes.html).

Results

Identification of causal genes for both ATDH and SCZ. Five known ATDH/ATDILI-related genes and 1,305 known SCZ-related genes were collected from GenBank. The human STRING network database was then used to map five and 1,079 genes were mapped on the network for ATDH and SCZ, respectively. Using these mapped genes, we predicted 3,045 and 1,458 possible ATDH- and SCZ-related genes, respectively, with potential effects on known disease-related genes. Among these, we identified 878 overlapping genes between the expanded ATDH and SCZ gene lists, which are shown in Supplementary Table S3. Furthermore, 500 genes with significant false discovery rate (FDR-) corrected hypergeometric test $P$ values ($< 10^{-8}$) were identified as causal factors for both ATDH and SCZ (Supplementary Table S4).

To validate these 500 genes as common causal factors for both ATDH and SCZ, we created a regularly updated online database ATDH-SCZgenes (http://www.bio-x.cn/ATDH-SCZgenes.html) and conducted a field synopsis/systemic meta-analysis to analyze associations of polymorphisms in each potential gene with ATDH or SCZ. During validation, GWAS data were firstly collected, and then, candidate gene association studies data were collected. To date, no GWAS for ATDH have been reported yet. Among these genes, only GSTM1, CYP2E1 and glutathione S-transferase theta-1 (GSTT1) ($P = 5.61E-22$, $4.54E-18$ and $3.87E-09$, respectively, Table 1) have been reported to associate with both ATDH and SCZ; all other genes were reported to associate with at most either disease alone. Genes with significant effects only on SCZ, which were obtained from the SzGene database or a systemic meta-analysis of genome wide association study (GWAS) data provided by Ricopili, are listed in Supplementary Table S5; in the future, association of these genes with ATDH will require testing.

Characteristics of the included studies. A flow diagram summarizing the study selection process is shown in Supplementary Fig. 1. A total of 33 and 24 potentially relevant studies regarding the association between CYP2E1/GSTM1/GSTT1 polymorphisms and the respective risks of ATDH and SCZ were identified after an initial screening based on the titles and/or abstracts of the candidate articles. After the second screening, 699 cases and 2,546 controls from fifteen studies of CYP2E1 and ATDH, 679 cases and 2,289 controls from fourteen studies of GSTM1 and ATDH, 592 cases and 2,569 controls from fourteen studied of GSTT1 and ATDH, one case-control study of CYP2E1 and SCZ, 1,469 cases and 1,605 controls from seven studies of GSTM1 and SCZ, and 936 cases and 971 controls from five studies of GSTT1 and SCZ were identified. The detailed characteristics of each study were listed in Supplementary Table S6. All studies confirmed the same complete loss of GSTM1 or GSTT1 mutation. Because the association studies involving CYP2E1 did not meet the fifth inclusion criterion (i.e.,
at least three studies regarding the association of each gene with either disease), studies involving this gene were omitted (detailed information about CYPE2E1 can be obtained from the online database). The genotype distributions of the cases and controls from all studies involving GSTM1 or GSTT1 in the context of ATDH and SCZ are presented in Tables 2 and 3, respectively. The null genotype refers to homozygous gene loss, which indicates a loss of gene function, and the present genotype includes both heterozygous gene loss and homozygous complete gene presence.

**Association of GST polymorphisms with ATDH.** Evaluations of the associations between GSTM1/ GSTT1 polymorphisms and the risk of ATDH are summarized in Fig. 1. Significant heterogeneity in the effects of these polymorphisms was observed for GSTM1, but not for GSTT1 (\(P = 0.088\) and \(I^2 = 36\%, \ 95\% \ CI: 0-0.662\) for GSTM1, \(P = 0.12\) and \(I^2 = 32\%, \ 95\% \ CI: 0-0.61\) for GSTT1). Because fewer than 20 studies were included in the meta-analysis, the random-effect model was used for both GSTM1 and GSTT1. The frequencies of GSTM1/GSTT1 null genotypes among the cases were higher than those among the controls (cases vs. controls: 51.25% vs. 42.85% for GSTM1, and 33.61% vs. 31.88% for GSTT1). The GSTM1 present genotype was significantly associated with a decreased risk of ATDH (odd ratio (OR): 0.71, 95% CI: 0.56–0.90, \(P = 0.005\)) (Fig. 1A). No significant association between the GSTT1 present genotype and ATDH was observed (OR 0.83, 95% CI: 0.56–1.29, \(P = 0.38\); Fig. 1B).

**Association of GST polymorphisms with SCZ.** The combined results regarding the association between GSTM1/GSTT1 polymorphisms and the risk of SCZ are presented in Fig. 2. Significant effect heterogeneity was observed in relation to GSTT1, but not for GSTM1 (\(P = 0.27\) and \(I^2 = 21\%, \ 95\% \ CI: 0–0.642\) for GSTM1, \(P = 0.004\) and \(I^2 = 74\%, \ 95\% \ CI: 0.351–0.895\) for GSTT1). Again, the random-effect model was used for both genes because of the small number of studies. The GSTM1 null genotype and the GSTT1 present genotype were more frequent

### Table 2. Genotype distributions of GSTM1/T1 polymorphisms among ATDH as cases and non-ATDH as controls. The null genotype means homozygous loss of genes, and the present genotype includes heterozygous loss of genes and homozygous complete genes. \(P^*\) value for chi-square test of genotype distribution.

| First author, year | Ethnicity | Cases | Controls | \(P^*\) |
|-------------------|-----------|-------|----------|--------|
|                   |           | Null  | Present  | Total  | Null  | Present | Total  |        |
| **GSTM1 polymorphism** |          |       |          |        |       |         |        |        |
| Chatterjee, 2009  | Indian    | 25    | 26       | 51     | 49    | 51       | 100    | 0.998   |
| Gupta, 2013      | Indian    | 21    | 29       | 50     | 61    | 185      | 246    | 0.013   |
| Huang, 2007      | Chinese   | 42    | 21       | 63     | 29    | 34       | 63     | 0.02    |
| Kim, 2010        | Korea     | 26    | 31       | 57     | 104   | 86       | 190    | 0.226   |
| Leiro, 2008      | Spanish   | 12    | 23       | 35     | 25    | 35       | 60     | 0.477   |
| Liu, 2014        | Chinese   | 14    | 6        | 20     | 96    | 47       | 143    | 0.798   |
| Monteiro, 2012   | Brazilian | 21    | 38       | 59     | 34    | 84       | 118    | 0.358   |
| Rana, 2014       | Indian    | 19    | 36       | 55     | 42    | 203      | 245    | 0.004   |
| Roy, 2001        | Indian    | 17    | 16       | 33     | 8     | 25       | 33     | 0.022   |
| Singla, 2013     | Indian    | 10    | 7        | 17     | 165   | 226      | 391    | 0.175   |
| Sotsuka, 2011    | Japanese  | 12    | 8        | 20     | 50    | 42       | 92     | 0.645   |
| Tang, 2012       | Chinese   | 55    | 34       | 89     | 203   | 153      | 356    | 0.414   |
| Teixeira, 2011   | Brazilian | 11    | 15       | 26     | 61    | 80       | 141    | 0.928   |
| Wang, 2010       | Chinese   | 63    | 41       | 104    | 54    | 57       | 111    | 0.079   |
| **Total**        |           | 348   | 331      | 679    | 981   | 1308     | 2289   |        |
| **GSTT1 polymorphism** |          |       |          |        |       |         |        |        |
| Chatterjee, 2009  | Indian    | 3     | 48       | 51     | 3     | 97       | 100    | 0.391   |
| Gupta, 2013      | Indian    | 11    | 39       | 50     | 30    | 216      | 246    | 0.062   |
| Huang, 2007      | Chinese   | 24    | 39       | 63     | 25    | 38       | 63     | 0.855   |
| Kim, 2010        | Korea     | 34    | 23       | 57     | 103   | 87       | 190    | 0.469   |
| Leiro, 2008      | Spanish   | 17    | 18       | 35     | 16    | 44       | 60     | 0.031   |
| Liu, 2014        | Chinese   | 13    | 7        | 20     | 97    | 46       | 143    | 0.8     |
| Monteiro, 2012   | Brazilian | 11    | 48       | 59     | 28    | 90       | 118    | 0.442   |
| Rana, 2014       | Indian    | 14    | 41       | 55     | 79    | 164      | 245    | 0.308   |
| Roy, 2001        | Indian    | 5     | 28       | 33     | 1     | 32       | 33     | 0.087   |
| Singla, 2013     | Indian    | 8     | 9        | 17     | 102   | 289      | 391    | 0.056   |
| Sotsuka, 2011    | Japanese  | 7     | 13       | 20     | 40    | 52       | 92     | 0.486   |
| Tang, 2012       | Chinese   | 40    | 49       | 89     | 164   | 192      | 356    | 0.849   |
| Teixeira, 2011   | Brazilian | 4     | 22       | 26     | 27    | 114      | 141    | 0.65    |
| Wang, 2010       | Chinese   | 40    | 49       | 89     | 164   | 192      | 356    | 0.849   |
| **Total**        |           | 231   | 433      | 664    | 879   | 1653     | 2534   |        |
among cases than among controls (cases vs. controls: 56.71% vs. 51.15% for the \textit{GSTM1} null genotype, 31.2% vs. 38.62% for the \textit{GSTT1} null genotype). Statistically significant association were observed between the \textit{GSTM1} present genotype and a decreased risk of SCZ (OR 0.78, 95% CI: 0.66–0.92, \(P = 0.004\); Fig. 2A), and between the \textit{GSTT1} present genotype and SCZ (OR 1.37, 95% CI: 0.93–2.03, \(P = 0.11\); Fig. 2B).

\section*{Sensitivity analyses and publication bias.}
A sensitivity analysis was conducted via sequential analysis after omitting one study at a time to assess the effects of individual studies on the overall meta-analysis estimate. When one study was excluded, the \(P\) values for overall effects ranged from 0.004 to 3.97E-5 and from 0.17 to 0.68, in the \textit{GSTM1/GSTT1} and ATDH fixed-effect model analyses, respectively; for the fixed-effect model analyses of \textit{GSTM1/GSTT1} and SCZ, the respective \(P\) values for overall effects ranged from 0.001 to 0.015 and from 0.005 to 0.32. These values indicate the stability of these analytical results.

Furthermore, Harbord’s test indicated no significant publication bias in the overall meta-analysis except for studies of the association between \textit{GSTM1} and SCZ (\(P = 0.56\) for \textit{GSTM1} vs. ATDH; \(P = 0.08\) for \textit{GSTT1} vs. ATDH; \(P = 0.0637\) for \textit{GSTM1} vs. SCZ; \(P = 0.91\) for \textit{GSTT1} vs. SCZ).

\section*{Credibility of meta-analysis results.}
The Power and Sample Size Program\(^{19}\) indicated that the total sample size had a power >90% to detect significant associations of the \textit{GSTM1} present genotype with ATDH and SCZ.
at ORs of 0.71 and 0.78, respectively. Furthermore, the respective false-positive reporting probabilities (FPRP) at a P value < 0.05 for the associations of GSTM1 present genotype with ATDH and SCZ were 0.088 and 0.063, with respective ORs of 0.6 and 0.7 (Supplementary Table S7).

Moreover, the strict inclusion criteria of this meta-analysis had addressed the genotyping quality. For the meta-analysis of ATDH studies, the nminor for the GSTM1 null genotype was 1,329, and a grade of A was given. The I² was 36% and a grade of B was given. After excluding a 2001 study by Roy20, a significant association remained between the GSTM1 present genotype and ATDH (P = 0.01), with a Harbord’s test P value of 0.86820. For the meta-analysis of SCZ studies, the nminor for the GSTM1 present genotype was 1,654, and a grade of A was given. A grade of A was also given for the I² of 21%. After excluding 2001 study by Harada21, a significant association remained between the GSTM1 present genotype and SCZ (P = 0.014), with a Harbord’s test P value of 0.138. According to the Venice criteria22,23, “moderate” cumulative evidence supported significant associations of the GSTM1 present genotype with both ATDH and SCZ.

Discussion

As noted previously, ATDH can impede TB treatment schedules and thereby increase complications morbidity6. Psychiatric disorders has also been reported to represent an additional adverse effect of anti-TB drugs5,7. SCZ is a severe psychiatric disorder, given that TB and SCZ are frequent co-morbid conditions 2,7, understanding the molecular basis for the relationship between ATDH and SCZ would not only facilitate personalized medicine by allowing physicians to identify potential exacerbation of SCZ and induction of ATDH patients, but could help to elucidate the molecular mechanisms common to both diseases. Previously, however, the relationship between ATDH and SCZ has been unclear, and knowledge about common biological determinants between these conditions has yet to emerge.

Previously, we proposed that GST genes might serve as a link between ATDILI and SCZ10. To provide a global perspective of the hidden molecular basis for the connection between ATDH and SCZ, we proposed a protein-protein interaction (PPI) network-based analysis pipeline that would prioritize possible key drivers that might affect both ATDH and SCZ by extending ATDH-related or SCZ-related gene sets to neighboring genes and identifying key causal genes that overlap in these extended gene sets. Although a direct overlap of known gene sets is the most intuitive way of exploring a genetic association of two diseases, such a direct overlap would fail to reflect the complexity of the intertwined regulation between these two diseases. Moreover, because the known disease gene sets are incomplete, the lacking genes could lead to a naive comparison. In contrast, in our proposed method, we qualitatively analyzed the potential nature of a gene as a key driver of two diseases using the hypergeometric test. Furthermore, identified candidate key drivers were ranked according to P value significance and network microenvironments (i.e., potential interaction neighbors). The ability to reconstruct potential causal signaling will facilitate further molecular biology studies.

In addition, identified key drivers can provide clues about therapeutic interventions that affect genes from both diseases. We note that EnrichNet use similar methods to determine the enrichment of one gene set into a particular pathway or other signatures24; these include employing information from the PPI network and extending the seed genes to neighboring gene using the RWR method. However, EnrichNet measures the significance of a relationship between a series of extended genes and a particular pathway according to different RWR distance cutoffs, whereas our method determines the best RWR distance cutoff. Specifically, we adopted the concept of the leading-edge subset used in GSEA, which is a sorted the neighbor gene lists based on RWR distances, and labeled the seed genes as positive and other genes as negative. Using our method, the peak at which the running sum maximally deviates from zero determines the best RWR distance cutoff; subsequently, we can determine the best extended genes for either disease and study the overlaps to identify candidate key causal genes using the hypergeometric test. Therefore, EnrichNet and our analysis pipeline use similar methods in different ways to address different problems.
In the present study, we used our novel pipeline analysis to identify 500 genes with a $P < 10^{-8}$ as possible causal genetic factors shared by ATDH and SCZ. Given the nature of ATDH, however, it is difficult to collect a sufficient number of patients with both ATDH and SCZ in the absence of other comorbidities. Because systemic meta-analysis is considered as a powerful tool for the identification of genes associated with a certain disease, we have created and regularly updated the online database ATDH-SCZgenes (http://www.bio-x.cn/ATDH-SCZgenes.html) to analyze and validate the association between candidate genes and both diseases.

Among the 500 evaluated candidate genes, to date, only $\text{GSTM1}$, $\text{CYP2E1}$ and $\text{GSTT1}$ were found to associate with both ATDH and SCZ; all others associated with neither or only one of the diseases. In the analysis of $\text{CYP2E1}$ and ATDH, a pooled OR of 1.2 (95% CI: 0.85–1.68) was determined for the rs2031920 ($-1053C>T$) polymorphism, whereas a fixed-effect model yielded a pooled OR of 1.3 (95% CI: 1.06–1.59). Only one previous study has evaluated and identified a positive association between $\text{CYP2E1}$ and SCZ. In contrast, the SzGene database and a systemic meta-analysis of SCZ GWAS data failed to corroborate that significant association, but consistently supported $\text{DRD2}$ to be significantly associated with SCZ. $\text{DRD2}$ has been reported as a prominent genetic risk factor for susceptibility to severe alcoholism, which is associated with liver damage. Further validation of the associations of these genes with ATDH and SCZ is needed in the future.

Through a systemic meta-analysis, we validated a significant association of the $\text{GSTM1}$ null genotype with increased risks of both ATDH and SCZ, and these significant results were supported by "moderate" evidence according to the Venice criteria, suggesting that $\text{GSTM1}$ may be a causal factor shared by both ATDH and SCZ.

ATDH has been widely suggested to be a Glutathione S-transferases ($\text{GSTs}$) related disease. Through conjugating glutathione with free radical scavengers and facilitating their elimination from the body to reduce potential toxicities of target substances, GSTs comprise a superfamily of detoxification enzymes that are encoded in two main genes: the $\text{GSTM1}$ gene on chromosome 1p13.3, which encodes for cytosolic GST class Mu 1 enzyme, and the $\text{GSTT1}$ gene on chromosome 22q11.2, which encodes for cytosolic GST class theta 1 enzyme. Both $\text{GSTM1}$ and $\text{GSTT1}$ may harbor a null mutation comprising a complete deletion of the respective gene via unequal homologous crossover, and homozygous null mutations can lead to a variable, tissue-specific loss of GSTs activity.

$\text{GSTM1}$ is mainly expressed in the liver and brain, a fact that supports the significant associations observed between this gene and both ATDH and SCZ. Furthermore, $\text{GSTM1}$ not only detoxifies the toxic metabolites of anti-TB drugs generated by CYP2E1 in the liver, but also catalyzes the conjugation of glutathione with aminochrome and dopa-o-quinone metabolites of oxidized dopamine in the brain. Reactive oxygen species are generated at high rates in the brain, and regulation of the growth and pruning of neurons is partly attributed to the redox mechanism that controls the balance between neuro destructive oxidants and neuro protective antioxidants. Therefore, GSTM1 inactivation due to the $\text{GSTM1}$ null genotype not only causes liver injury but also promotes the accumulation of neuro destructive oxidants and consequent development of SCZ.

Compared with our previous study, the present study exhibits the following improvements: (1) the current meta-analysis or field synopsis is more compressive and systematic and involves as many candidate genes as possible including $\text{GST}$ genes; (2) the current analysis is updated regularly using the online database ATDH-SCZgenes; and (3) more strict statistical methods were used in this analysis, including ORs instead of risk ratios and a threshold P value for publication bias of 0.1 instead of 0.05. However, this study also has some limits, including use of the STRING database, a functional association network without direction. The availability of a disease specific directed network might lead to more comprehensive and concrete conclusions using the current method.
Additionally, the significant association of \( \text{GSTM1} \) with SCZ should be interpreted with caution, as the \( P \) value for the publication bias test was 0.0637; this value represents the evidence of small-study effects with all studies included. However, our sensitivity analysis to assess the effects of individual studies on the overall meta-analysis estimate yielded \( P \) values for overall effects of 0.001–0.015 in the \( \text{GSTM1} \) and SCZ analysis, indicating the stability of positive results obtained with these analyses. Furthermore, till now, no GWASs on ATDH have been reported and the number of reported ATDH-associated genes is small; therefore, only \( \text{GSTM1} \) and \( \text{GSTT1} \) genes could be validated in the current systemic meta-analysis. Validation of other causal genes must be performed in future studies.

In summary, we provide a list of possible causal genetic factors associated with both ATDH and SCZ, and have identified a shared genetic basis of these two diseases. Furthermore, we have created and will regularly update the ATDH-SCZgenes online database to validate the association of each candidate gene with either disease. Finally, \( \text{GSTM1} \) was validated as a causal factor of both ATDH and SCZ, whereas other genes such as \( \text{CYP2E1} \) and \( \text{DRD2} \) will require further validation.

Methods

Ethics statement. The current research was performed in compliance with the Helsinki Declaration and was approved by the Bioethics committee of the Bio-X Institutes of Shanghai Jiaotong University. Informed consent was obtained from all subjects.

Network-based analysis. All known ATDH/ATDILI- and SCZ-related genes reported (including GWAS) to have significant effects on the relative disease by at least one study were collected using GenBank (http://www.ncbi.nlm.nih.gov/gene/), and are listed in Supplementary Table S1. These genes were used as seed genes in the subsequent analysis. To explore the causal factors of ATDH and SCZ, a novel network-based analysis pipeline was developed per the workflow shown in Fig. 4. First, established seed genes related to ATDH or SCZ were mapped onto the highest confidence human STRING network (version 9.1, confidence score > 0.900), which included a total of 8,823 genes34. These seed genes were subsequently expanded on the network using the RWR method35 as follows: a PPI network \( \text{G} = (\text{V}, \text{E}) \) comprised of a set of proteins \( \text{V} \) and a set of interactions \( \text{E} \) is represented by an \( n \times n \) adjacency matrix \( \text{A} \), where \( n \) is the number of proteins. The entries at row \( i \) and column \( j \) are set to 1 if...
protein $i$ interacts with protein $j$; otherwise they are set to 0. First, the adjacency matrix $A$ was normalized in a column-wise manner as follows

$$A_{(i,j)}' = \frac{A_{(i,j)}}{\sum_{k=1}^{n} A_{(i,k)}}$$

(1)

The random walker initiates at a set of seed genes, (e.g., known disease genes). The initial state $P_0$ can be formulated as a column vector

$$P_0 = [\psi_1 \psi_2 \cdots \psi_j \cdots \psi_m]^T$$

(2)

where $\psi_i$ is set to $1/m$ for $m$ seed genes and to 0 for the other genes on the network, and $n$ is the number of genes on the network. The random walker randomly visits the adjacent genes for every $t \rightarrow t+1$. The state probabilities $P_{t+1}$ at time $t+1$ are calculated as follows

$$P_{t+1} = (1-r)A^TP_t + rP_0$$

(3)

where $P_t$ represents the state probabilities at time $t$, $r$ is the restart probability (i.e., starting from the seed genes again), which was set to 0.7 as suggested by multiple previous studies. This process was repeated until a steady-state was reached; this was defined as a difference between two steps of $<1e\text{-}6$ according to previous studies.

After expansion, all genes on the network were assigned disease gene probability. To determine the boundaries of gene expansion, we developed a new idea from GSEA and calculated the running sum from the top to bottom of the ranked gene list. Specifically, when we expanded ATDH genes from the ATDH seed genes, all genes on the network were ranked according to the likelihood of being an ATDH-related gene. If we encountered a gene that was not an established SCZ seed gene, $\psi_i/\sqrt{N-G}$ was added to the running sum, where $N$ is the number of all network genes and $G$ is the number of known established SCZ seed genes; otherwise, $\psi_i/\sqrt{N-G}$ was added. Based on the running sum, a peak of network expansion from the ATDH seed genes was determined. This cutoff was then used to obtain a list for ATDH seed gene expansion. A list for SCZ seed gene expansion was generated similarly. Common genes between these two lists (i.e., overlapped genes) were studied because these might more robustly reflect the common genetic basis of the two diseases. Finally, we screened all possible causal gene factors by testing their neighbors for overlapped genes using the hypergeometric test $P$ value.

$$P = \sum_{k=m}^{n} \binom{M}{k} \binom{N-M}{n-k} \binom{N}{n}$$

(4)

where $N$ is the number of all network genes, $M$ is the number of common disease genes, $n$ is the number of neighbor genes, and $m$ is the number of neighbor genes that are common disease genes. To control the FDR, hypergeometric test $P$ values were adjusted according to the method published by Benjamini & Hochberg.

The analysis was conducted using R software version 3.1.2 and RWR code was obtained from RWOAG package (https://r-forge.r-project.org/R/?group_id=1126).

**Literature search.** The digital medical databases PubMed, Scopus, ISI Web of Science, EMBASE, Chinese National Knowledge Infrastructure (CNKI) and the Chinese BioMedical Literature Database were searched for studies with publication dates up to 03/31/2016 using the following keywords: ("anti-tuberculosis drug-induced liver injury", "anti-tuberculosis drug-induced hepatotoxicity", "ATDH" or "ATDILI") and ("Schizophrenia"), together with the full name or abbreviation of each candidate causal gene, including: "cytochrome P4502E1", "CYP2E1", "UDP-glucuronosyltransferase 1A6", "UGT1A6", "glutathione S-transferaseM1", "GSTM1", and "glutathione S-transferase T1" or "GSTM1". The references of retrieved articles were also reviewed to identify additional relevant literatures.

**Inclusion and exclusion criteria.** Articles included in the meta-analysis complied with the following criteria: (1) original case-control association studies based on randomly selected individuals; (2) Provision of complete genotype distribution data; (3) cases comprising TB patients with ATDH and controls comprising TB patients without ATDH or (4) cases comprising SCZ patients and controls comprising healthy subjects; and (5) at least three available studies regarding the association of each gene with either disease. Other studies, such as case-only studies, duplications, animal studies, comparisons of laboratory methods, editorials, and review articles were excluded. Studies were further evaluated using the Quality-Evaluation Score Sheet (version 2.0) (Supplementary Table S2): a score of 8–10 indicated a high quality study.

**Data extraction.** Data extraction was performed independently by two reviewers using a standardized protocol and reporting form. Discrepancies between the two reviewers were resolved by further discussion with a third party. For overlapping studies, the study with the larger sample size was retained for the meta-analysis. The recorded study characteristics included: (1) the first author's name, (2) publication year, (3) sample ethnicity, (4) control and case characteristics, (5) methods used for genotyping and (6) target genes.

**Statistical analysis.** The strengths of the associations between gene polymorphisms and the risk of ATDH or SCZ were measured using ORs with corresponding 95% CIs. Pooled ORs were calculated for null vs. present
Credibility of meta-analysis results. A power analysis was performed using the Power and Sample Size Program with $\alpha = 0.05$ as the level of significance; effects sizes were estimated from the meta-analyses$^{19}$. To assess the noteworthiness of an association, the FPRP$^{14}$ was estimated using a FPRP threshold of 0.2 and prior probabilities of 0.05–10$^{-4}$. Cumulative evidence for genetic associations of GSTM1 and GSTT1 present genotypes with ATDH and SCZ, respectively, were assessed according to the Venice interim criteria, which include the amount of evidence, replication of results and protection from bias$^{22}$. Regarding the amount of evidence, grades of A, B, and C were given for a noteworthiness of 100%2, 50–100%, and <50%, respectively. Regarding replication, grades A, B, and C were given for the percentages of variability in the effect estimates that were attributable to heterogeneity rather than sampling error, an $I^2 > 50\%$ was roughly considered to indicate substantial heterogeneity$^{50}$. This formula use Q as well as degrees of freedom$^{26,51}$. A sensitivity analysis, in which one study at a time was removed prior to analysis, was conducted to evaluate whether a single study would significantly affect the results. This analysis used a model other than the model used to calculate the pooled ORs. Harbord's test was used to test small-study effects among which publication bias might be a contributor, and a P value < 0.1 was considered representative of statistically significant publication bias$^{52}$. All statistical analyses were implemented in Review Manager 5.2 (The Nordic Cochrane Centre, Copenhagen, Denmark) and Stata version 11.2 (Stata Corporation, College Station, TX, USA).

Data accessibility. The datasets supporting this article have been uploaded as part of the Supplementary Material.

References

1. Sullivan, P. F., Kendler, K. S. & Neale, M. C. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. Archives of general psychiatry 60, 1187–1192, doi: 10.1001/archpsyc.60.12.1187 (2003).
2. Kuo, S. C. et al. Incidence and outcome of newly-diagnosed tuberculosis in schizophrenics: a 12-year, nationwide, retrospective longitudinal study. BMC infectious diseases 13, 351, doi: 10.1186/1471-2334-13-351 (2013).
3. WHO. World Health Organization: Global tuberculosis report. (Date of access: 6/6/2015) (2012).
4. Blumberg, H. M. et al. American Thoracic Society/Centers for Disease Control and Prevention/Infectious Diseases Society of America: treatment of tuberculosis. American journal of respiratory and critical care medicine 167, 603–662, doi: 10.1164/rccm.167.4.603 (2003).
5. Nathwani, R. A. & Kaplowitz, N. Drug hepatotoxicity. Clinics in liver disease 10, 207–217, doi: 10.1016/j.cld.2006.05.009 (2006).
6. Gubay, B. E. et al. Side effects due to primary antituberculosis drugs during the initial phase of therapy in 1149 hospitalized patients for tuberculosis. Respiratory medicine 100, 1834–1842, doi: 10.1016/j.rmed.2006.01.014 (2006).
7. Doherty, A. M. et al. A review of the interplay between tuberculosis and mental health. General hospital psychiatry 35, 398–406, doi: 10.1016/j.genhosppsych.2013.03.018 (2013).
8. Wiedorn, W. S. & Ervin, F. Schizophreniclike psychotic reactions with administration of isoniazid. A.M.A. archives of neurology and psychiatry 72, 321–324, doi: 10.1001/archneurpsyc.1954.02330030055005 (1954).
9. Bernardino, M., Gatell, J. M. & Parellada, E. Acute exacerbation of chronic schizophrenia in a patient treated with antituberculosis drugs. The American journal of psychiatry 148, 1402, doi: 10.1176/ajp.148.10.1402 (1991).
10. Cai, L. et al. Meta-Analysis-Based Preliminary Exploration of the Connection between ATDII and Schizophrenia by GSTM1/T1 Gene Polymorphisms. PloS one 10, e0128643, doi: 10.1371/journal.pone.0128643 (2015).
11. Schadt, E. E. et al. An integrative genomics approach to infer causal associations between gene expression and disease. Nature genetics 37, 710–717, doi: 10.1038/ng1589 (2005).
12. Zhang, B. et al. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. Cell 153, 707–720, doi: 10.1016/j.cell.2013.03.030 (2013).
13. Chen, L. et al. A hybrid computational method for the discovery of novel reproduction-related genes. PloS one 10, e0117090, doi: 10.1371/journal.pone.0117090 (2015).
14. Cai, L. et al. Identification of genetic associations of SP110/MYBBP1A/RELA with pulmonary tuberculosis in the Chinese Han population. Human genetics 132, 265–273, doi: 10.1007/s00439-012-1244-5 (2013).
15. Cai, L., Yang, Y. H., He, L. & Chou, K. C. Modulation of Cytokine Network in the Comorbidity of Schizophrenia and Tuberculosis. Current topics in medicinal chemistry 16, 655–665, doi: 10.2174/1568026615666150819105033 (2016).
16. Liu, J. et al. Association of EGF rs4444903 and XPD rs13181 polymorphisms with cutaneous melanoma in Caucasians. Medicinal chemistry 11, 351–559, doi: 10.2174/1573460410666141224115516 (2015).
17. Allen, N. C. et al. Systematic meta-analyses and field synopsis of genetic association studies in schizophrenia: the SzGene database. Nature genetics 40, 827–834, doi: 10.1038/ng.1701 (2008).
18. Schizophrenia Working Group of the Psychiatric Genomics, C. Biological insights from 108 schizophrenia-associated genetic loci. Nature 511, 421–427, doi: 10.1038/nature13595 (2014).
19. Dupont, W. D. & Plummer, W. D. Jr. Power and sample size calculations. A review and computer program. Controlled clinical trials 11, 116–128, doi: 10.1016/0197-2456(90)90005-M (1990).
20. Roy, B. et al. Increased risk of antituberculosis drug-induced hepatotoxicity in individuals with glutathione S-transferase M1 ‘null’ mutation. Journal of gastroenterology and hepatology 16, 1033–1037, doi: 10.1046/j.1440-1746.2001.02585.x (2001).
21. Harada, S., Tachikawa, H. & Kawanishi, Y. Glutathione S-transferase M1 gene deletion may be associated with susceptibility to certain forms of schizophrenia. Biochemical and biophysical research communications 281, 267–271, doi: 10.1006.bbrc.2001.4347 (2001).
M.-H.C. Analyzed the data: T.H. and L.C. Contributed reagents/materials/analysis tools: L.-L.L., W.-Z.C., Y.-F.X. Conceived and designed the experiments: T.H., L.C. and L.H. Performed the experiments: T.H., C.-L.L., L.C. and L.H.

Author Contributions

(No. 12Z102050009).

Program (2010CB529600 and 2012CB910102), the Chinese National Science Foundation (No. 31101015 and 31270911).

We are grateful to Dr. Simone de Jong for her discussion and comments. This work was supported by the 973 Program (2010CB529600 and 2012CB910102), the Chinese National Science Foundation (No. 31101015 and 81421061) and the scientific research foundation for the returned overseas, the Chinese Ministry of Education (No. 12Z102050009).

Acknowledgements

We are grateful to Dr. Simone de Jong for her discussion and comments. This work was supported by the 973 Program (2010CB529600 and 2012CB910102), the Chinese National Science Foundation (No. 31101015 and 81421061) and the scientific research foundation for the returned overseas, the Chinese Ministry of Education (No. 12Z102050009).

Author Contributions

Conceived and designed the experiments: T.H., L.C. and L.H. Performed the experiments: T.H., C.-L.L., L.C. and M.-H.C. Analyzed the data: T.H. and L.C. Contributed reagents/materials/analysis tools: L.-L.L., W.-Z.C., Y.-F.X. and L.H. Wrote the paper: T.H., L.C. and P.F.O. All authors gave final approval for publication.
Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Huang, T. et al. A new method for identifying causal genes of schizophrenia and anti-tuberculosis drug-induced hepatotoxicity. Sci. Rep. 6, 32571; doi: 10.1038/srep32571 (2016).

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

© The Author(s) 2016