Comparing methods for performing trans-ethnic meta-analysis of genome-wide association studies

Xu Wang1, Hui-Xiang Chua2, Peng Chen1, Rick Twee-Hee Ong1, Xueling Sim6, Weihua Zhang7, Fumihiko Takeuchi8, Xuanyao Liu1,3, Chiea-Chuen Khor9, Wan-Ting Tay10, Ching-Yu Cheng1,4,10,11, Chen Suo1, Jianjun Liu8, Tin Aung4,10, Kee-Seng Chia1, Jaspal S. Kooner7, John C. Chambers7, Tien-Yin Wong4,10,11, E-Shyong Tai1, Norihiro Kato8 and Yik-Ying Teo1,2,3,5,9,*

1Saw Swee Hock School of Public Health, 2Department of Statistics and Applied Probability, 3NUS Graduate School for Integrative Science and Engineering, 4Department of Ophthalmology, and 5Life Sciences Institute, National University of Singapore, Singapore, Singapore, 6Department of Biostatistics, University of Michigan, Ann Arbor, USA, 7Faculty of Medicine, Imperial College, London, UK, 8Department of Gene Diagnostics and Therapeutics, National Center for Global Health and Medicine, Tokyo, Japan, 9Genome Institute of Singapore, Agency for Science, Technology and Research, Singapore, Singapore, 10Singapore Eye Research Institute, Singapore National Eye Center, Singapore, Singapore, and 11Duke-NUS Graduate Medical School, Singapore, Singapore

Received October 4, 2012; Revised January 25, 2013; Accepted February 6, 2013

Genome-wide association studies (GWASs) have discovered thousands of variants that are associated with human health and disease. Whilst early GWASs have primarily focused on genetically homogeneous populations of European, East Asian and South Asian ancestries, the next-generation genome-wide surveys are starting to pool studies from ethnically diverse populations within a single meta-analysis. However, classical epidemiological strategies for meta-analyses that assume fixed- or random-effects may not be the most suitable approaches to combine GWAS findings as these either confer low statistical power or identify mostly loci where the variants carry homogeneous effect sizes that are present in most of the studies. In a trans-ethnic meta-analysis, it is likely that some genetic loci will exhibit heterogeneous effect sizes across the populations. This may be due to differences in study designs, differences arising from the interactions with other genetic variants, or genuine biological differences attributed to environmental, dietary or lifestyle factors that modulate the influence of the genes. Here we compare different strategies for meta-analyzing GWAS across genetically diverse populations, where we intentionally vary the effect sizes present across the different populations. We subsequently applied the methods that yielded the highest statistical power to a trans-ethnic meta-analysis of seven GWAS in type 2 diabetes, and showed that these methods identified bona fide associations that would otherwise have been missed by the classical strategies.

INTRODUCTION

Genome-wide association studies (GWASs) have seen unprecedented successes at discovering novel genetic variants that influence the severity of different health outcomes in humans (1–3). While many of the early findings have been made in homogeneous populations of European ancestry (4–8), recent reports have unveiled novel genetic associations discovered from populations of African (9–12), East Asian (13–15) and South Asian (16) ancestries. Most of the identified variants individually contribute a modest effect to the health outcome. To discern the association signals from the statistical noise that is inadvertently present from querying more than a million variants, these successful efforts typically

*To whom correspondence should be addressed at: Department of Statistics and Applied Probability, Blk S16, Level 7, 6 Science Drive 2, Faculty of Science, National University of Singapore, Singapore 117546. Tel: +65 65162760; Fax: +65 68723919; Email: statyy@nus.edu.sg

© The Author 2013. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oup.com
meta-analyze several GWAS that have been performed in samples of similar ancestry. This increases the sample sizes while minimizing genetic heterogeneity across the study samples. The natural progression is to extend such meta-analyses to include samples from as many studies as possible, which can stem from different heterogeneous populations in the world.

When used to perform such global meta-analyses, classical statistical approaches that assume either fixed- or random-effects at each single-nucleotide polymorphism (SNP) require the same SNP to be present across all the studies in order to achieve the higher statistical power conferred by the joint analyses. In addition, fixed-effects models assume that the same SNP has to exhibit similar degree of association with the outcome of interest, in terms of the effect sizes, across most, if not all, of the studies. On the first requirement, it is common for the studies to be performed on different genotyping technologies, given the variety of commercial microarrays that differ in SNP density and placement. However, sophisticated and well-calibrated imputation procedures like IMPUTE (17) and MACH (18) have allowed the SNP contents of different studies to be harmonized to the same resolution, with the use of reference data from the HapMap (19) or the 1000 Genomes (20).

Addressing the heterogeneity in both effect sizes and association signals across diverse populations is non-trivial. Assuming a causal variant is actually functional in all the different populations where the studies were conducted, there could still be several reasons underlying the heterogeneity of effect sizes detected in each of the study populations. First, the study designs are often not identical, and subtle variations in phenotype definitions or measurements across studies can be an inadvertent source of heterogeneity. Secondly, as the causal variant is seldom directly queried in a genetic association study, variations in the degree and pattern of linkage disequilibrium (LD) between an SNP and the causal variant across studies can also introduce heterogeneity in the observed effect sizes. Thirdly, the biological impact of a shared causal variant can also vary across populations due to its interactions with other functional variants in the genome that may be found at different frequencies in the different populations. Finally, nongenetic exposures of different study populations are unlikely to be similar, and it is possible that environmental and lifestyle factors can modify the impact of the genetic contribution (21), resulting in the same causal variant exerting a different influence to the health outcome across the different populations.

The point of global meta-analyses is to include as many studies as possible, agnostic of the population ancestry or genetic background of each study (12). Taking the formal threshold of genome-wide significance ($P < 5 \times 10^{-8}$) into account, the use of fixed-effects methods (FE) will be methodologically bounded to locate only genetic effects that are strongly exhibited in most of the studies with a similar effect size. Although random-effects methods (RE) are specifically designed to handle heterogeneity, they tend to rely on a conservative assumption that the effect sizes are different across studies even under the null hypothesis of no association (22). These standard epidemiological meta-analysis frameworks thus tend to overlook those signals that are either present in certain population clades only (which results in a significant down-weighting of the effect sizes towards the null of 0 by the populations not exhibiting the association), or where there is considerable heterogeneity in the effect sizes which increases the standard errors of the estimated effect sizes and thus dampening the statistical evidence of the pooled association signal.

To cope with heterogeneous effect sizes between studies, two new approaches for meta-analyzing GWAS data have recently been introduced. Han and Eskin developed an alternate random-effects model (RE-HE) that assumes a common true effect size of zero in all the studies under the null hypothesis and allows the effect sizes to vary among studies under the alternative hypothesis (22). By relaxing the conservative assumption of RE under the null hypothesis, RE-HE has been reported to be more powerful than standard random-effects models and yields higher statistical power than fixed-effects models in situations where there exist inter-study heterogeneity in effect sizes. The second method by Morris (MANTRA) was specifically designed to perform trans-ethnic meta-analysis (23). MANTRA adopts a Bayesian framework and assumes that studies from closely related populations are more likely to share a common true effect size, and the true effect size is allowed to vary across different population clades. When there exists a correlation between effect sizes and relatedness between populations, MANTRA has been reported to confer significantly higher power than both FE and RE.

Here we perform a comparison of the four strategies for meta-analyzing GWAS across genetically diverse populations to gauge the relative performance in terms of sensitivity and specificity. We achieve this through a series of simulations where we intentionally: (i) vary the effect sizes present across 10 populations in 5 different scenarios that mimicked different biological situations and (ii) vary the number of studies investigated between 10 and 30. By identifying the approaches that are robust to inter-study effect size heterogeneity, we subsequently performed a trans-ethnic meta-analysis of seven GWAS in type 2 diabetes and illustrate that these methods successfully identify bona fide associations that would otherwise have been missed by the classical FE and RE approaches.

**RESULTS**

**Power and false-positive rates**

We compared the performance of the two classical statistical methods for performing meta-analyses (FE, RE) with the two recently introduced strategies for trans-ethnic meta-analyses (RE-HE and MANTRA) using a series of simulations performed with HAPGEN using seed haplotypes from 10 HapMap Phase 3 populations (excluding the admixed population ASW). We simulated 3000 cases and 3000 controls for each of the 10 populations in triplicate, yielding a total of 30 studies in total and a possible sample size of 90 000 cases and 90 000 controls for the joint analysis of the 30 studies. In calculating the empirical false-positive rates, we simulated 300 000 SNPs in each of the 30 studies under the null hypothesis of no association (see Materials and
Methods for meta-analysis, we have defined statistical significance, by the author of MANTRA (23) was expected to correlate with the equivalent Bayes’ factor against the empirical false-positive rates (see Supplementary Material, Table S1 for the same comparison). As expected, all four approaches performed similarly in the ‘All populations’ scenario with power approaching 100%, due to the large sample size of the joint analysis of 30 studies. When we reduced the number of studies in the meta-analysis to 10 and 20, respectively, the power decreased across all four approaches (Supplementary Material, Fig. S3). In the remaining four scenarios where there existed heterogeneous effect sizes across the studies, RE-HE and MANTRA consistently outperformed both classical meta-analysis methods of FE and RE. In particular, in the ‘Out-of-Africa’ and ‘Europe and South Asia’ scenarios where only 70 and 40% of the studies, respectively, were expected to exhibit the association, we saw considerable gains in power by MANTRA compared with the rest of the methods.

As MANTRA clusters studies according to the genetic relatedness as measured by the allelic spectrum of the queried SNPs, we were keen to ensure that our simulations did not give MANTRA an unfair advantage by considering effect sizes that vary according to population clades. The fifth scenario introduced a common effect size in two Caucasian populations (CEU and TSI) and a Chinese population (CHD), and this was meant to mimic the situation where a shared environment triggered the genetic impact of an otherwise neutral locus. We observed that, in this scenario, all the methods had very low power <10%, although both MANTRA and RE-HE continue to yield higher power than FE and RE models with MANTRA providing the highest power at 6.1%.

### Application to T2D data

We applied the different meta-analysis approaches to combine the results from seven GWAS of type 2 diabetes in East, South-East and South Asian populations (Table 2). These included Chinese (SCES, SP2-1M, SP2-610), Malay (SIMES) and South Asian Indians (SINDI) from Singapore, Japanese from Tokyo (Japan) and South Asian Indians residing in London (LOLIPOP). A total of 1 663 404 autosomal SNPs were meta-analyzed, and the genomic control inflation factors of the FE and RE were 1.046 and 0.875, respectively. For MANTRA, we considered Bayes’ factor threshold of $10^5$ to define statistical significance (as recommended by the author), while for the remaining three approaches, we considered a $P$-value threshold of $7.9 \times 10^{-7}$.

All six association signals that were significant in the FE meta-analysis were similarly identified by RE-HE and MANTRA (Fig. 2, Table 3). These included the well-established T2D loci such as *CDKAL1*, *CDKN2A*, *KCNQ1* and *TCF7L2*. In

### Table 1: False-positive rate of FE, RE, RE-HE and MANTRA at thresholds of increasing significance

<table>
<thead>
<tr>
<th>Threshold</th>
<th>FE</th>
<th>RE</th>
<th>RE-HE</th>
<th>MANTRA</th>
<th>Threshold for MANTRA</th>
</tr>
</thead>
<tbody>
<tr>
<td>$5.00 \times 10^{-6}$</td>
<td>$4.97 \times 10^{-6}$</td>
<td>$4.11 \times 10^{-6}$</td>
<td>$4.97 \times 10^{-6}$</td>
<td>$1.35 \times 10^{-1}$</td>
<td>Bayes’ factor = $10^6$</td>
</tr>
<tr>
<td>$1.00 \times 10^{-5}$</td>
<td>$9.82 \times 10^{-5}$</td>
<td>$7.63 \times 10^{-5}$</td>
<td>$7.53 \times 10^{-5}$</td>
<td>$9.07 \times 10^{-3}$</td>
<td>Bayes’ factor = $10^3$</td>
</tr>
<tr>
<td>$1.00 \times 10^{-5}$</td>
<td>$9.81 \times 10^{-5}$</td>
<td>$7.13 \times 10^{-5}$</td>
<td>$8.26 \times 10^{-4}$</td>
<td>$7.49 \times 10^{-4}$</td>
<td>Bayes’ factor = $10^2$</td>
</tr>
<tr>
<td>$1.00 \times 10^{-5}$</td>
<td>$1.04 \times 10^{-5}$</td>
<td>$6.86 \times 10^{-5}$</td>
<td>$8.95 \times 10^{-5}$</td>
<td>$1.00 \times 10^{-5}$</td>
<td>Bayes’ factor = $10^1$</td>
</tr>
<tr>
<td>$1.00 \times 10^{-5}$</td>
<td>$2.98 \times 10^{-6}$</td>
<td>$0.00 \times 10^{-6}$</td>
<td>$8.95 \times 10^{-6}$</td>
<td>$1.00 \times 10^{-6}$</td>
<td>Bayes’ factor = $10^0$</td>
</tr>
<tr>
<td>$5.00 \times 10^{-6}$</td>
<td>$0.00 \times 10^{-6}$</td>
<td>$0.00 \times 10^{-6}$</td>
<td>$5.96 \times 10^{-6}$</td>
<td>$1.00 \times 10^{-6}$</td>
<td>Bayes’ factor = $10^0$</td>
</tr>
</tbody>
</table>

The leftmost column is the threshold for the false-positive of FE, RE and RE-HE approaches, and the rightmost column is the threshold for the Bayesian approach. The $P$-values and the Bayes’ factors are assessed at the causal variants.
addition, RE-HE successfully located the HNF4A locus with stronger evidence (from $5.87 \times 10^{-6}$ in FE to $3.26 \times 10^{-7}$ in RE-HE). MANTRA similarly located HNF4A and further identified two more loci that did not achieve the definition of statistical significance by the other three methods. The reason that RE-HE and MANTRA performed better than FE at HNF4A was because of heterogeneous effect sizes at the index SNP rs4812829, where studies in East Asian populations (Japan, SCES, SP2-610, SP2-1M) exhibited strong evidence of T2D with odds ratios around 1.1; The South Asian Indians (SINDI and LOLIPOP) exhibited stronger evidence of 1.2; but SIMES exhibited a protective effect of odds ratio at 0.9 (Fig. 3).

The two loci that have been additionally identified by MANTRA include PIEZO2 and C6orf57. Among them, C6orf57 has previously been reported by Sim et al. (24), although we emphasize that our finding here does not constitute an independent validation as four of the seven GWAS (SP2-610, SP2-1M, SIMES, SINDI) we have used are from the report by Sim et al.

**DISCUSSION**

It is increasingly common to combine genome-wide scans of the same outcome that have been performed in genetically diverse populations. We have examined the statistical power and false-positive rates of four approaches for performing trans-ethnic meta-analysis of GWAS, and found the two recently introduced approaches by Morris (MANTRA) and Han and Eskin (RE-HE) are robust to heterogeneous effect sizes at a genuinely associated genetic locus that conventional epidemiological models assuming fixed- or random-effects will attenuate the overall signal, particularly when some studies fail to carry the association due to different study designs, environmental modification of genetic influence or the presence of diversity in the genetic architecture of the different populations. As with the fixed-effects model, the false-positive rates displayed by RE-HE are calibrated against the definition of statistical significance. While we cannot directly evaluate the expected false-positive rates for MANTRA as it relies on the use of Bayes’ factors, we observe that at the recommended Bayes’ factor threshold of $10^2$, the false-positive rate out of 300,000 simulations under the null remained at zero. These methods are readily available as two computationally efficient stand-alone linux software ‘MANTRA’ and ‘metasoft’, respectively.

Inter-study heterogeneity can pose a more serious challenge in meta-analyses of genetic association studies, compared with traditional nongenetic epidemiological meta-analyses. The additional complexity due to diverse genetic architectures of the different ancestries which affect SNP tagging efficiency, as well as the poorly understood implications of environmental modifiers in modulating the influence of the functional gene variants, imply that the observed effect sizes across trans-ethnic populations are unlikely to be homogeneous all the time. Current genome-wide meta-analyses have fundamentally relied on using fixed-effects models to locate genetic associations, which have significant limitations. First, does this mean that these meta-analyses may be consigned to identify only the variants that carry homogeneous effect sizes that are present in most, if not all, of the studies? Secondly, given the tendency to ‘regress towards the mean’ when estimating the pooled effect size, is it likely that the reported effect sizes from meta-analyses will be under-estimates of the true effect sizes, thus inadvertently discounting the amount of heritability explained in certain populations? We stress that this should not be confused with the ‘winner’s curse’ where the observed effect size of the first study may be an over-estimate of the true effect size. Importantly, classic genome-wide meta-analyses using fixed-effects models may be missing out on explaining more of the genetic heritability because fundamentally this may be an inappropriate tool to locate the genetic variants? Given the stringent criteria that have been imposed on GWAS reports, having a few studies exhibit null associations in the meta-analysis (due to reasons as discussed above) may just tilt the balance for some variants to be dropped from the follow-up replication experiments.

One drawback of MANTRA is the lack of proper effect size estimation for the meta-analysis. While it provides an estimate of the effect size at every SNP for each study and an overall Bayes’ factor as the combined association signal, it does not
estimate a combined effect size from the joint analysis of all the studies even in the absence of heterogeneity across studies. In theory, one can use the posterior probabilities generated from MANTRA to partition the studies into different clades, and a clade-specific effect size can be estimated manually. Under the assumption of no effect size

![Figure 2](https://example.com/figure2.png)

**Figure 2.** Manhattan plots from the FE, RE-HE and MANTRA. Manhattan plots of the T2D association signals from the fixed-effects (top panel) and random-effects by Han and Eskin (middle panel) and the Bayesian approach (bottom panel) meta-analyses of seven GWASs from East, South-East and South Asian populations. The horizontal line in the first two panel represents the P-value at $7.9 \times 10^{-7}$, and the line in the third panel represents Bayes’ factor at $10^5$. The names of the nearest gene of the significant loci are provided in all the plots.

<table>
<thead>
<tr>
<th>RSID</th>
<th>CHR</th>
<th>POS</th>
<th>Effect allele</th>
<th>Other allele</th>
<th>Log10(Bayes’ factor)</th>
<th>FE P-value</th>
<th>RE P-value</th>
<th>RE-HE P-value</th>
<th>Nearest GENE</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs9295474</td>
<td>6</td>
<td>20 760 696</td>
<td>G</td>
<td>C</td>
<td>6.78</td>
<td>5.90E-08</td>
<td>9.67E-04</td>
<td>3.22E-09</td>
<td>CDKAL1</td>
</tr>
<tr>
<td>rs1048886</td>
<td>6</td>
<td>71 345 910</td>
<td>G</td>
<td>A</td>
<td>6.21</td>
<td>5.25E-02</td>
<td>3.57E-01</td>
<td>2.11E-06</td>
<td>C6orf57</td>
</tr>
<tr>
<td>rs2383208</td>
<td>9</td>
<td>22 122 076</td>
<td>G</td>
<td>A</td>
<td>7.87</td>
<td>6.35E-10</td>
<td>9.60E-05</td>
<td>2.95E-10</td>
<td>CDKN2AB</td>
</tr>
<tr>
<td>rs7903146</td>
<td>10</td>
<td>114 748 339</td>
<td>T</td>
<td>C</td>
<td>6.84</td>
<td>6.23E-09</td>
<td>6.23E-09</td>
<td>1.23E-08</td>
<td>TCF7L2</td>
</tr>
<tr>
<td>rs2237896</td>
<td>11</td>
<td>2 815 016</td>
<td>G</td>
<td>A</td>
<td>9.22</td>
<td>2.83E-11</td>
<td>3.56E-08</td>
<td>5.23E-11</td>
<td>KCNQ1</td>
</tr>
<tr>
<td>rs11636554</td>
<td>15</td>
<td>75 395 802</td>
<td>G</td>
<td>A</td>
<td>6.18</td>
<td>3.94E-08</td>
<td>3.05E-06</td>
<td>6.66E-08</td>
<td>PEAK1</td>
</tr>
<tr>
<td>rs7178572</td>
<td>15</td>
<td>75 534 245</td>
<td>G</td>
<td>A</td>
<td>5.69</td>
<td>1.10E-07</td>
<td>4.12E-06</td>
<td>1.82E-07</td>
<td>HMG20A</td>
</tr>
<tr>
<td>rs676809</td>
<td>18</td>
<td>10 677 717</td>
<td>G</td>
<td>A</td>
<td>5.35</td>
<td>4.07E-01</td>
<td>6.18E-01</td>
<td>1.50E-05</td>
<td>PIEZO2</td>
</tr>
<tr>
<td>rs4812829</td>
<td>20</td>
<td>42 422 681</td>
<td>G</td>
<td>A</td>
<td>5.18</td>
<td>5.87E-06</td>
<td>3.03E-02</td>
<td>3.26E-07</td>
<td>HNF4A</td>
</tr>
</tbody>
</table>

The SNPs showing here are those with the highest association signal in each region. The regions that have been identified by FE, RE-HE and MANTRA are highlighted in grey in the respective columns. We have adopted a P-value threshold of $7.9 \times 10^{-7}$ as this is empirically equivalent to Bayes’ factor threshold of $10^5$. 

Table 3. SNPs exhibiting significant association signals ($P < 7.9 \times 10^{-7}$ in FE, RE-HE and Bayes’ factor $\geq 10^5$ in MANTRA) of seven type 2 diabetes GWASs.
heterogeneity in the null hypothesis, the RE-HE is more powerful than the naïve RE meta-analysis but it gives the same effect size estimate as the RE because the effect size estimation is performed only under the alternative hypothesis where both RE and RE-HE carry the same assumption. Under the assumption of no effect size heterogeneity in the null hypothesis, the RE-HE is more powerful than the naïve RE meta-analysis but it gives the same effect size estimate as the RE because the effect size estimation is performed only under the alternative hypothesis where both RE and RE-HE carry the same assumption.

In summary, there is growing interest and need to perform global meta-analyses as different research groups extend their collaborations to include GWAS from multiple heterogeneous ethnicities. Having the right analytical tools for performing such trans-ethnic meta-analyses is vital, as conventional fixed-effects methods used in classical epidemiology will likely locate gene regions where the associations have already been uncovered in bulk of the studies. In such meta-analyses, there is almost no control of the experimental designs of the different studies, and this is likely to affect the ability to discover a genuine association. One convenient example is the FTO gene locus for T2D which was not detected in case–control designs that matched samples by obesity status [e.g. FUSION (25)], since the contribution of FTO to T2D was via increasing obesity propensity. The HNF4A locus provides another useful and validated example that an unguided fixed-effects meta-analysis of studies from diverse ethnicities can omit genuine biological signals. We have assessed the performance of four methods to perform trans-ethnic meta-analyses and showed that MANTRA and RE-HE are robust in accounting the various situations of effect size heterogeneity.

MATERIALS AND METHODS

Fixed-effects meta-analysis

FE assumes a common true effect size \( \mu \) for a particular allele at the SNP across all the studies, and the effect size of each study \( T_i \) is draw from a normal distribution with mean \( \mu \) and variance \( \sigma^2 \). Let \( v_i \) be the variance of the \( i \)th study in a meta-analysis of \( K \) studies (although \( v_i \) is an estimated value, it is treated as the true variance of study \( i \) in the meta-analysis), and let \( w_i = v_i^{-1} \) be the reciprocal of the variance, the inverse-variance-weighted effect-size estimator of the true effect size is

\[
\hat{\mu} = \frac{\sum_{i=1}^{K} w_i T_i}{\sum_{i=1}^{K} w_i}.
\]

The variance of \( \hat{\mu} \) is estimated as \( \hat{\sigma}^2 = \left( \sum_{i=1}^{K} w_i \right)^{-1} \). Under the null hypothesis that there is no association, the test statistic can be calculated as \( (\hat{\mu} - \mu)^2 / \hat{\sigma}^2 \), which follows a chi-square distribution with 1 degree of freedom.

Random-effects meta-analysis

RE assumes the true effect size for the \( i \)th study \( \theta_i \) is sampled from a normal distribution with mean \( \mu \) and variance \( \tau^2 \). The between-study variance \( \tau^2 \) is estimated by the method of moments (26). Define

\[
Q = \sum_{i=1}^{K} w_i (T_i - \hat{T}_F)^2
\]

and

\[
c = \sum_{i=1}^{K} w_i - \frac{\sum_{i=1}^{K} w_i^2}{\sum_{i=1}^{K} w_i}.
\]

The between-study variance is estimated as

\[
\hat{\tau}^2 = \frac{Q - (K - 1)}{c}
\]

when \( Q > (K - 1) \) or 0 when \( Q \leq (K - 1) \).

The inverse-variance-weighted effect size estimator is similar to that of the fixed-effect model but with the additional variance term accounted, as follows:

\[
\hat{\mu} = \frac{\sum_{i=1}^{K} (v_i + \hat{\tau}^2)^{-1} T_i}{\sum_{i=1}^{K} (v_i + \hat{\tau}^2)^{-1}}.
\]

Under the null hypothesis that there is no association, the test statistic can be calculated as \( (\hat{\mu} - \mu)^2 / \hat{\sigma}^2 \), which similarly follows a chi-square distribution with 1 degree of freedom.
As the random-effects meta-analysis accounts for additional variability between the studies, the procedure is generally more conservative than the fixed-effects meta-analysis.

The extent of inter-study heterogeneity can be assessed by comparing the test statistic $Q$ against a chi-square distribution with $K-1$ degrees of freedom, which tests the null hypothesis that there is no variability in the distribution of the true effect sizes.

Random-effects meta-analysis by Han and Eskin (RE-HE)

RE-HE assumes that the true effect sizes are different among studies under the alternative hypothesis that a particular SNP is associated with the phenotype of interest in all studies. However, in the absence of any evidence of association, the true effect size should be zero in all the studies, and RE-HE adopts a hybrid approach that assumes there is no effect size heterogeneity ($\mu = 0$, $\tau^2 = 0$) under the null hypothesis. Taking a likelihood approach, the likelihoods under the null and alternative hypotheses are, respectively

$$L_0 = \prod_{i=1}^{K} \frac{1}{\sqrt{2\pi v_i}} \exp\left(-\frac{T_i^2}{2v_i}\right)$$

$$L_1 = \prod_{i=1}^{K} \frac{1}{\sqrt{2\pi(v_i + \tau^2)}} \exp\left(-\frac{(T_i - \mu)^2}{2(v_i + \tau^2)}\right).$$

The maximum likelihood estimates for $\mu$ and $\tau^2$ can be obtained by an iterative procedure, as suggested by Hardy and Thompson (27),

$$\hat{\mu}_{(n+1)} = \frac{\sum_{i=1}^{K} (T_i/(v_i + \hat{\tau}^2_{(n)}))}{\sum_{i=1}^{K} (1/(v_i + \hat{\tau}^2_{(n)}))}$$

$$\hat{\tau}^2_{(n+1)} = \frac{\sum_{i=1}^{K} ((T_i - \hat{\mu}_{(n+1)})^2 - v_i)/(v_i + \hat{\tau}^2_{(n)}))^2}{\sum_{i=1}^{K} (1/(v_i + \hat{\tau}^2_{(n)}))}.$$

The likelihood ratio test statistic is thus

$$S_{\text{New}} = -2\log(\lambda) = \sum_{i=1}^{K} \log\left(\frac{v_i}{v_i + \hat{\tau}^2}\right) + \sum_{i=1}^{K} \frac{T_i^2}{v_i} - \sum_{i=1}^{K} \frac{(T_i - \hat{\mu})^2}{v_i + \hat{\tau}^2}.$$

When the number of studies is large, the test statistic asymptotically follows a mixture of a chi-square distribution with 1 degree of freedom and a chi-square distribution with 2 degrees of freedom. When the number of studies is small, the tabulated $P$-values are provided by Han and Eskin, with a reasonable accuracy for up to $10^{-8}$. For more significant $P$-values, the asymptotic $P$-value corrected by the ratio between the asymptotic $P$-value and the true $P$-value estimated at $10^{-8}$ is used.

The effect-size estimate and its confidence interval in RE-HE are the same as those in RE. This is because RE-HE only modified the assumption under the null hypothesis. The effect size estimation is performed under the alternative hypothesis which is exactly the same as in RE.

Bayesian approach meta-analysis (MANTRA)

MANTRA assumes that studies from the same ethnic group are more homogeneous, thus they are likely to share the same effect size (denoted as population-specific effect $\beta_s$ for the $s$th population cluster). But effect sizes vary among different population clusters.

Let $M_0$ denote the null model of no associations and $M_1$ the alternative model in a Bayesian framework. Let $b$ be the observed effect size from each study and $s$ the respective standard deviation, the evidence of association can be assessed by the Bayes’ factor (28)

$$\Lambda = \frac{f(b_s|b|M_1)}{f(b_s|b|M_0)} = \frac{\int_{0}^{\infty} f(b_s|\theta_0|M_1) \alpha(\theta_0) d\theta_0}{\int_{0}^{\infty} f(b_s|\theta_0|M_0) \alpha(\theta_0) d\theta_0},$$

where $\theta_0$ is the model parameter including the population-specific effects $\beta_s$. The likelihood is

$$f(b_s|\theta) = f(b_s|\beta) = \prod_{i=1}^{K} f(b_{i,s}|\beta_i),$$

and

$$f(b_{i,s}|\beta_i) \propto \frac{1}{s_i} \exp\left[-\frac{(b_i - \beta_i)^2}{2s_i^2}\right].$$

Based on the assumption that allelic effects are likely to vary between broad ethnic groups rather than closely related populations, the studies are assigned into ‘ethnic clusters’ based on the average allele frequency similarity, measured by the $F_{ST}$ metric (29). The $\beta_s$ is then determined by the respective cluster effect.

As it is not possible to evaluate the marginal likelihood $f(b_s|s,M)$ directly, the joint posterior density of $f(\theta|b,s,M)$ is considered instead. Here $f(\theta|M)$ denote the prior density function of parameters under a specific model $M$. The posterior density is subsequently approximated using a Metropolis-Hastings Markov chain Monte Carlo (MCMC) algorithm. The details of the method can be found in the publication by the author (23).

Simulation set-up

Case–control data were simulated using the HAPGEN (30) program, with seed haplotypes from 10 HapMap 3 populations (excluding ASW) and population-averaged recombination rates from Phase 2 of HapMap (19). The effective population sizes used in the simulations are: (i) 11 418 for CEU, GIH, MEX and TSI; (ii) 14 269 for CHB, CHD and JPT; (iii) 17 469 for LWK, MKK and YRI. Only SNPs that are not on the Illumina 1 M BeadChip, with minor allele frequencies of at least 1% in all populations are chosen as the causal SNPs in the simulations. For each simulation, we generated 30 studies, where each of the 10 HapMap3 population is used to simulated three studies, with 3000 cases and 3000 controls in each study. In the simulations to calculate the false-positive rates, the allelic relative risk for every causal SNP in each population was set at 1.0 and the meta-analyses

Downloaded from https://academic.oup.com/hmg/article-abstract/22/11/2303/634133 by guest on 20 February 2019
were performed at the causal SNPs. The false-positive rates for FE, RE and RE-HE were calculated by counting the proportion of the causal SNPs which exhibited a meta-analysis P-value smaller than a threshold varies from $5 \times 10^{-2}$ to $5 \times 10^{-5}$. For MANTRA, the false-positive rate was calculated as the proportion of the causal SNPs which exhibited Bayes’ factor larger than a threshold varies from $10^0$ to $10^5$. To calculate statistical power, we considered five separate scenarios: (i) ‘All populations’, where all 30 studies share the same allelic relative risk of 1.1 at the causal SNP; (ii) ‘Out-of-Africa’, where all the non-African populations share the same allelic relative risk of 1.1 at the causal SNP, while the remaining African populations (9 studies from LWK, MKK and YRI) carry a null allelic relative risk of 1.0. The last two approaches risk of 1.1 at the causal SNP, while the remaining populations (9 studies from CHB, CHD and JPT) carry an allelic relative risk of 1.2 at the causal SNP while the European and South Asian populations carry an allelic relative risk of 1.1; (iii) ‘Europe and South Asia’, where only studies from CEU, GIH, MEX and TSI carry an allelic relative risk of 1.1, while the remaining 18 studies carry a null allelic relative risk of 1.0; (iv) ‘Effect size heterogeneity’, where the genetic effects are present only in non-African populations but the East Asian populations (9 studies from CHB, CHD and JPT) share an allelic relative risk of 1.2 at the causal SNP while the European and South Asian populations carry an allelic relative risk of 1.1; (v) ‘Environment modifier’, where the populations living in Europe and US (9 studies from CEU, GIH, MEX and TSI) share an allelic relative risk of 1.1 at the causal SNP, while the remaining populations carry a null allelic relative risk of 1.0. The last two approaches are meant to parallel the situation where different environmental exposures modify the influence of the genes on phenotype severity. The generated effect sizes are normally distributed about the respective means of 1.0, 1.1 and 1.2 in the different scenarios (Supplementary Material, Fig. S4).

The simulated datasets and the scripts to perform the four meta-analyses are available at http://www.statgen.nus.edu.sg/~trans-ethnic/.

Type 2 diabetes GWAS

We considered seven GWASs in type 2 diabetes (T2D) from Singapore, Japan and the UK that have previously been reported either individually or in meta-analyses (Table 1).

**Japan**: This consists of a T2D study across 931 cases and 1404 controls sampled from four regions in Tokyo that have been genotyped on the Illumina HumanHap 550 BeadChip (13); **SCES**: The Singapore Chinese Eye Study included 302 T2D cases and 1089 controls, out of a total of 1952 Chinese subjects that have been genotyped on the Illumina Human610-Quad BeadChip; **SIMES**: The Singapore Malay Eye Study genotyped 3280 Malay subjects in Singapore which included 794 T2D cases and 1240 controls genotyped on the Illumina Human610-Quad BeadChip (13,24); **SINDI**: The Singapore Indian Eye Study genotyped 3400 South Asian Indian subjects in Singapore, which included 977 T2D cases and 1169 controls genotyped on the Illumina Human610-Quad BeadChip (16,24); **SP2-1M**: The Singapore Prospective Study Program (SP2) which genotyped 5499 subjects, of which 928 T2D cases and 939 controls were genotyped on the Illumina HumanHap-1M BeadChip (13,24); **SP2-610**: Of the 5499 SP2 subjects, 1082 T2D cases and 1006 controls were genotyped on the Illumina Human610-Quad BeadChip (13,24);

**LOLIPOP**: This is a population-based cohort of South Asian samples that reside in West London and have all four grandparents born in the Indian subcontinent (which include India, Pakistan, Sri Lanka and Bangladesh), where 1783 T2D cases and 4773 controls were genotyped on the Illumina Human610-Quad BeadChip (16). All seven studies have been imputed using reference data from Phase 2 of the International HapMap Project, thus harmonizing the SNP content of these studies.

**SUPPLEMENTARY MATERIAL**

Supplementary Material is available at HMG online.

**ACKNOWLEDGEMENTS**

We thank the two anonymous reviewers for their constructive comments which have greatly improved the article.

**Conflict of Interest statement**: None declared.

**FUNDING**

This project acknowledges the support of the Saw Swee Hock School of Public Health from the National University of Singapore, National Medical Research Council (NMRC) 0796/2003, Singapore and the Biomedical Research Council (BMRC), 09/1/35/19/616, Singapore. Y.Y.T., C.P., and R.T.H.O. acknowledge support from the National Research Foundation, NRF-RF-2010-05, Singapore.

**REFERENCES**


