Literature Review | Methods for combining genotyping data from different arrays

Jessica R. Shaw | August 29, 2018

Method	Pros	Cons
Union imputation Impute genotypes using the combination of all variants typed on one or both arrays as input.	NA	Not a valid approach (Johnson et al., 2013; Xie et al., 2014). Results in spurious associations in the absence of a genotyping bias, even when used to combine very similar arrays (Johnson et al., 2013).
Intersection imputation Impute genotypes using variants present on both arrays as input.	Eliminates bias caused by unbalanced genotyping (Johnson et al., 2013; Xie et al., 2014). Can maintain satisfactory overall imputation quality with as few as 30% of total variants shared by arrays (Xie et al., 2014). Commonly used in case/control studies where cases and controls were genotyped on different platforms (Iperen et al., 2017).	Intractable if the number of variants represented on both arrays is too small (Verma et al., 2014). Reduces statistical power for downstream analyses (Xie et al., 2014). Discards high quality, directly measured genotypes for variants that were not typed on both arrays. Does not exclude the possibility of differential genotyping error rates, which may lead to differential imputation error rates.
Imputation before combination Impute genotypes separately from each array, then combine imputed datasets.	Maintains the intended 'backbone' of each array, consisting of variants strategically selected to inform genome-wide inference (Verma et al., 2014). Generates more genotypes of good quality than the intersection imputation method (Iperen et al., 2017). Standard method used by large consortia (Iperen et al., 2017).	Subject to differential imputation errors, both for variants that were imputed from both arrays and for variants that were directly genotyped on one array and imputed from the other (Sinott et al., 2012). Can lead to Type II errors (Beecham et al., 2010; Sinnott et al., 2012).

Literature Review | Methods for combining genotyping data from different arrays

Method	Pros	Cons
"Replacement strategy" Use the directly measured genotype for any SNP that was typed on at least one array and imputed with a high imputation quality score (IQS>0.9). The IQS is a quality metric that accounts for differential rates of chance concordance by minor allele frequency (Lin et al., 2010).	Eliminates most spurious associations (Xie et al., 2014). Improves statistical power compared to the intersection strategy (Xie et al., 2014).	Has not been used in any published studies (zero citations of Xie et al., 2014).
"Correction strategy" Combine measured and imputed genotypes for SNPs available on one or more of all arrays.	Yields a valid statistical test (Xie et al., 2014). Improves statistical power compared to the intersection strategy (Xie et al., 2014).	Has not been used in any published studies (zero citations of Xie et al., 2014).

References

Abecasis, G. (n.d.) Combining data from different genotyping platforms. *Center for Statistical Genetics at the University of Michigan*. https://www.genome.gov/pages/about/od/opg/gainkickoffworkshop/abecasis-gaincombiningdata.pdf

Iperen, E. P. A. van, Hovingh, G. K., Asselbergs, F. W., & Zwinderman, A. H. (2017). Extending the use of GWAS data by combining data from different genetic platforms. *PLOS ONE*, *12*(2), e0172082. https://doi.org/10.1371/journal.pone.0172082

Johnson, E. O., Hancock, D. B., Levy, J. L., Gaddis, N. C., Saccone, N. L., Bierut, L. J., & Page, G. P. (2013). Imputation across genotyping arrays for genome-wide association studies: assessment of bias and a correction strategy. *Human Genetics*, *132*(5), 509–522. https://doi.org/10.1007/s00439-013-1266-7

Sinnott, J. A., & Kraft, P. (2012). Artifact due to differential error when cases and controls are imputed from different platforms. *Human Genetics*, *131*(1), 111–119. https://doi.org/10.1007/s00439-011-1054-1

Verma, S. S., de Andrade, M., Tromp, G., Kuivaniemi, H., Pugh, E., Namjou-Khales, B., … Ritchie, M. D. (2014). Imputation and quality control steps for combining multiple genome-wide datasets. *Frontiers in Genetics*, *5*. https://doi.org/10.3389/fgene.2014.00370

Xie, Y., Hancock, D. B., Johnson, E. O., & Rice, J. P. (2014). Two Adjustment Strategies for Imputation across Genotyping Arrays. *Human Heredity*, *78*(2), 73–80. https://doi.org/10.1159/000363337