Received: 19 December 2018 | Revised: 5 March 2019 | Accepted: 18 March 2019

DOI: 10.1002/gepi.22204

RESEARCH ARTICLE

WILEY Genetic Epidemiology

OFFICIAL JOURNAL INTERNATIONAL GENETIC EPIDEMIOLOGY SOCIETY www.geneticepi.org

On the differences between mega- and meta-imputation and analysis exemplified on the genetics of age-related macular degeneration

Mathias Gorski[1] | **Felix Günther**[1,2] | **Thomas W. Winkler**[1] | **Bernhard H. F. Weber**[3] | **Iris M. Heid**[1]

[1]Department of Genetic Epidemiology, University of Regensburg, Regensburg, Germany

[2]Statistical Consulting Unit StaBLab, Department of Statistics, LMU Munich, München, Germany

[3]Institute of Human Genetics, University of Regensburg, Regensburg, Germany

Correspondence
Iris M. Heid, Department of Genetic Epidemiology, University of Regensburg, Franz-Josef-Strauß-Allee 11, 93053 Regensburg, Germany.
Email: iris.heid@klinik.uni-regensburg.de

Funding information
Deutsche Forschungsgemeinschaft, Grant/Award Number: DFG HE3690/5-1; National Institutes of Health; Advancing genomics through the AMD Genomics Consortium, Grant/Award Numbers: RES511967, NIH-2017 R01

Abstract

While current genome-wide association analyses often rely on meta-analysis of study-specific summary statistics, individual participant data (IPD) from multiple studies increase options for modeling. When multistudy IPD is available, however, it is unclear whether this data is to be imputed and modeled across all participants (mega-imputation and mega-analysis) or study-specifically (meta-imputation and meta-analysis). Here, we investigated different approaches toward imputation and analysis using 52,189 subjects from 25 studies of the International Age-related Macular Degeneration (AMD) Genomics Consortium including, 16,144 AMD cases and 17,832 controls for association analysis.

From 27,448,454 genetic variants after 1,000-Genomes-based imputation, mega-imputation yielded ~400,000 more variants with high imputation quality (mostly rare variants) compared to meta-imputation. For AMD signal detection ($P < 5 \times 10^{-8}$) in mega-imputed data, most loci were detected with mega-analysis without adjusting for study membership (40 loci, including 34 known); we considered these loci genuine, since genetic effects and *P*-values were comparable across analyses. In meta-imputed data, we found 31 additional signals, mostly near chromosome tails or reference panel gaps, which disappeared after accounting for interaction of whole-genome amplification (WGA) with study membership or after excluding studies with WGA-participants.

For signal detection with multistudy IPD, we recommend mega-imputation and mega-analysis, with meta-imputation followed by meta-analysis being a computationally appealing alternative.

KEYWORDS
age-related macular degeneration, genome-wide association studies, Genotype-imputation, mega-analysis, meta-analysis

[Correction updated after online publication dated 5 October 2018: "*P* value" has been changed to "*P*-value" throughout; "x" has been replaced with the multi symbol "×" in Model IV; "6" in "L1-L6" has been made as nonbold in Figure 1 caption; and values in Table 2 have been aligned.]

1 | INTRODUCTION

Genome-wide association studies (GWAS) are one of the most successful approaches to identify the genetic make-up of complex diseases and disease phenotypes (GWAS catalogue; MacArthur et al., 2017). GWAS require huge number of subjects usually obtained by combining multiple studies. The common genetic consortia approach is the pooling of study-specific summary statistics for each of the millions of genetic variants from multiple GWAS in meta-analyses (GWAMA; Klarin et al., 2018; Scott et al., 2017; Turcot et al., 2018). For this, the study analyst conducts quality control, imputation of untyped variants, and GWAS analysis and, then, provides summary statistics to the meta-analyst. The meta-analyst combines the per-variant summary statistics across studies, usually using a fixed-effect meta-analysis. To increase coverage and power, the number of studies and participants per study as well as the number of variants per chip or reference panel are constantly increasing, with current GWAMA efforts now often including >100 studies with up to 1,000,000 study participants (Visscher et al., 2017). As a consequence, this requires each of the hundred study analysts to repeat imputation and analyses to incorporate the information from larger reference panels and more complex analyses plans; the meta-analysts need to conduct quality control of an ever increasing number of study files with some, but limited options to detect analysis errors (Winkler et al., 2014).

Alternatively, individual participant data (IPD) of GWAS is gathered into one large data set, which enables joint imputation ("mega-imputation") and joint association analysis ("mega-analysis"; Fritsche et al., 2016). Certainly, having IPD of GWAS allows for all kinds of imputation and association approaches including study-specific imputation ("meta-imputation") and study-specific association analysis ("meta-analysis"), which mimics the current common consortia approach. Collecting IPD of GWAS for numerous studies is a logistical effort, since privacy issues of study participants have to be certified and study investigators to be persuaded. In contrast, there is a current trend toward study-specific GWAS data as IPD in research data bases that facilitates the collection of large-scale multistudy IPD (dbGaP; Rich et al., 2016; European Genome-phenome Archive; Lappalainen et al., 2015). Still, even when IPD is readily available, the processing of large-scale genome-wide IPD is a substantial computational burden: when mega-imputation bares no advantage over meta-imputation, the computational burden of imputing across all studies' data jointly can be spared; when, additionally, mega-analysis has no substantial advantage over meta-analysis, there is little motivation to gather IPD or—when IPD is available—to take on the extra computational effort of conducting mega-imputation and mega-analysis.

While there is previous work comparing mega-analyses versus meta-analysis highlighting mathematical equivalence of "mega-models" (i.e., one model applied across studies) with fixed effect meta-analysis of study-specific effects (Lin & Zeng, 2010a, 2010b), there is no published work comparing mega-imputed versus meta-imputed data. It is unclear to what extent these two imputation approaches yield differences in terms of imputation quality. It is also unclear whether one imputation approach is better in combination with one or the other association approach: the meta-imputation might yield better results when followed by meta-analysis, but it is also possible that a mega-model adjusting for study membership handles this in a more powerful way. The questions are whether there is a gain from having IPD compared with summary statistics based GWAMA and, when multistudy IPD is available, how this data is to be imputed and modeled: across all participants (mega-imputation and mega-analysis) or study-specifically (meta-imputation and meta-analysis).

We thus set out to investigate different approaches toward imputation (mega-imputation vs. meta-imputation) and analysis (mega-analysis vs. meta-analysis). One challenge in comparing different imputation approaches is the fact that simulated genome-wide genotype data is never realistic given the complexity in linkage disequilibrium and covariate structure. For this reason, imputation is often investigated using real data, but real data for an objective like ours needed to be large, from different studies, and available as IPD. We utilized the International Age-related Macular Degeneration Genomics Consortium (IAMDGC) data on the genetics of age-related macular degeneration (AMD; Fritsche et al., 2016), which is a data set available to us as IPD from 25 studies with a total of 52,189 individuals (Section 2.1). Our IPD is unique, since DNA was gathered and genotyped centrally with the same chip; the chip contained GWAS and exome content enabling the analysis also of rare variants. We imputed the data with multiple approaches (Sections 2.2 and 3.1) and evaluated the resulting mega- and meta-imputed data sets for their association based on 16,144 advanced AMD cases and 17,832 controls, which were the unrelated IAMDGC participants with European ancestry and a clear advanced AMD or no-AMD status. We applied five different association models implementing different levels of accounting for between-study differences, including a model that fully ignored study membership (as in the original analysis; Fritsche et al., 2016) and a model that mimicked a meta-analysis of study-specific summary statistics (the usual current GWAMA approach, Section 2.3). The genetics of advanced AMD and this data is ideal due to the strong genetic component of advanced AMD—with 34 genome-wide significant loci having been identified in this IAMDGC data in the original analysis (Fritsche et al., 2016).

Therefore, the IAMDGC data gave us the opportunity to explore the ability of each imputation/model combination to detect disease signals in a real setting.

Another challenge when comparing different imputation approaches is the fact that it is not fully straightforward how to compare different types of imputed data due to its high dimensionality. We defined a set of criteria of what made one imputed data set better than the other: we judged the resulting imputed data sets by the number of variants that were at high imputation quality (Sections 2.2 and 3.1) and, for each imputation/model combination, by the number of variants that yielded reliable results ("analyzable variants," Sections 2.4 and 3.2), by the degree of inflation/deflation in the empirical *P*-value distribution under the null hypothesis (Sections 2.5 and 3.3), and the number of true and false detected disease loci (Sections 2.6, 3.4–3.7, and Figure S1).

2 | MATERIALS AND METHODS

2.1 | The IAMGDC data set

To compare the impact of different imputation approaches on real data, we utilized the 52,189 study participants from the IAMDGC (Table S1). Data and quality control procedures had been described in detail elsewhere (Fritsche et al., 2016). Briefly, DNA of each study participant across 26 studies had been gathered. These included four studies where DNA had been derived from Whole Genome Amplification (WGA). The samples of all 52,189 individuals had been genotyped centrally using a custom-modified HumanCoreExome array (Illumina, Inc., San Diego, CA) yielding 508,740 genotyped autosomal variants after quality control (call rate ≥98.5%, Hardy-Weinberg $P > 10^{-6}$). Ethnicity, relatedness and the first two genetic principal components (genoPC1 and genoPC2) had been derived from this genetic data. The full IAMDGC data included individuals with advanced AMD, early AMD, and no AMD, related individuals and individuals from non-European ancestry. For our association analyses, we focused on the unrelated participants of European ancestry with advanced AMD or no AMD as in the original analysis yielding 16,144 advanced AMD cases and 17,832 controls, including 2,188 subjects from four studies with WGA as DNA source (Table S1). These study participants spread across 26 studies, but two studies (NHS_HPF and Rotterdam) contained only few participants with advanced AMD or no AMD, respectively; thus, we combined these two studies resulting in 25 studies for our methodological investigation here to enable a meta-analysis, while the data was originally designed for a mega-imputation and mega-analysis approach.

2.2 | Mega- versus meta-imputation and imputation quality

For the imputation of the genotypes of untyped variants for the 52,189 participants, we first phased the data to yield study haplotypes and then estimated the alleles of untyped variants using the 1000G Phase I v3 cosmopolitan reference panel (Web Resources). We used the 508,740 genotyped autosomal variants for the phasing step and 345,274 of these genotyped variants overlapping with the reference panel (Web Resources) for the estimation step. We applied two imputation approaches genome-wide: (a) a "mega-imputation," consisting of the phasing of all subjects jointly and the estimation of untyped variants for all subjects jointly, and (b) a "meta-imputation," where the data of each of the 25 studies was phased and estimated separately and the resulting study-specifically imputed data was joined afterwards. We utilized *shapeit.v2.r727.linux.x64* for the phasing (Delaneau, Marchini, & Zagury, 2011; 200 states, window size 2.5 Mb) and *minimac-omp_2013_7_17* (Web Resources) for the estimating step (1,115 2.5 Mb-chunks, 200 states, 500 kb overlapping regions).

For each of the two imputed data sets, each variant's imputation quality was computed as the ratio of the observed variance of allele dosages across individuals compared to the expected variance of allele dosages given the observed allele frequency (RSQ, Web Resources; Li, Willer, Ding, Scheet, & Abecasis, 2010). For the mega-imputed data, we obtained the variant's RSQ across all individuals ("mega-RSQ" derived from *minimac*); for the meta-imputed data, we derived a formula to compute a variant's pooled RSQ from study-specific minor allele frequencies (MAFs) and RSQs, where each study-specific RSQ is weighted by the number of study subjects and the study-specific MAF ("meta-RSQ," Supporting Information Note 1). Study specific MAFs and RSQs were obtained from *minimac*. A variant's MAF was computed across all 52,189 subjects in the mega-imputed data ("mega-MAF") and as a pooled MAF weighted by the number of subjects per study from study-specific MAFs in the meta-imputed data ("meta-MAF"). To quantify differences between imputation approaches, we compared the number of variants yielded by each of the imputation approaches within categories of RSQ and MAF: $RSQ < 0.3$ (low quality), $0.3 \leq RSQ < 0.8$ (medium quality), $0.8 \leq RSQ$ (high quality); $MAF = 0$ (monomorphic), $0 < MAF < 0.01$ (rare), $0.01 \leq MAF \leq 0.05$ (less frequent), and $0.05 < MAF$ (common).

To further differentiate effects of the phasing from the estimation step, we conducted an extended imputation experiment, where we did not only conduct mega-imputation and meta-imputation in the sense of

mega-phasing/mega-estimation and meta-phasing/meta-estimation, but also the hybrid approaches mega-phasing/meta-estimation and meta-phasing/mega-estimation. We applied these four imputation approaches in an example region on chromosome 5 (chr5:37.5-40 M) and compared differences in per-variant imputation quality between these four data sets overall and by MAF- and RSQ-categories. In a sensitivity analysis, we added the reference panel in the phasing step to conduct an extra imputation experiment for chromosome 1. For this, we restricted the genetic variants in the IAMDGC data to those available in the reference panel (29,741 variants from 47,359 genotyped variants on chromosome 1) and utilized the cosmopolitan 1000G Phase I version 3 reference panel for both phasing and estimation).

2.3 | Association analyses

For the mega-imputed and the meta-imputed data (here again: referring to the mega-phased/mega-estimated or meta-phased/meta-estimated data, respectively), we conducted association analyses comparing persons with advanced AMD versus no AMD (16,144 advanced AMD cases, 17,832 controls from 25 studies; Table S1). We applied five logistic regression models: a Model I that fully ignored study membership, Models II–IV that adjusted for increased levels of between-study differences, and a Model V being a fixed effect meta-analysis of study-specific summary statistics. In Models I–IV, we applied one logistic regression model across all study participants ("mega-models"), which requires IPD, while Model V treats the studies separately mimicking the current GWAMA approach ("meta-model").

The five models to estimate the genetic effect of each variant j, β_j, are described in the following. For the ith individual, we denote the imputed "observed" dosage as G_{ij} and the observed binary AMD outcome as Y_i; f is the logistic function and correspondingly f^{-1} the logit function; $\text{Prob}(Y_i{=}1|X)$ the probability of AMD given the covariate vector X. First, we applied a model across all participants ignoring study membership, adjusting for the first two genetic principal components, genoPC1_i and genoPC2_i, and DNA source $\text{WGA}_i \in \{0,1\}$),

$$f^{-1}(\text{Prob}(Y_i = 1|X)) = \alpha + \beta_j G_{ij} + \gamma \text{WGA}_i + \delta \text{genoPC1}_i + \varepsilon \text{genoPC2}_i. \quad \text{(Model I)}$$

Second, we added study membership ($K = 25$ studies) via 24 dummy variables, study_{ik} (1 if individual i is from study $k = 1...24$, and 0 else, Regensburg study as reference), which corresponds to estimating study-specific intercepts (intercept of Regensburg being α and $\alpha + \theta_k$ for each other study k),

$$f^{-1}(\text{Prob}(Y_i = 1|X)) = \alpha + \beta_j G_{ij} + \gamma \text{WGA}_i + \delta \text{genoPC1}_i + \varepsilon \text{genoPC2}_i + \sum_{k=1}^{K-1} \theta_k \text{study}_{ik}. \quad \text{(Model II)}$$

Third, we added interaction between study ($M = 4$ studies including participants with WGA as DNA source) and DNA source, via three interaction terms, $\text{WGA}_i \times \text{study}_{im}$ (1 for individual i with WGA as DNA source and if individual i is from WGA study $m = 1...3$, Columbia study as reference), which corresponds to estimating study-specific WGA effects (WGA effect of Columbia being γ and $\gamma + \vartheta_m$ for study m),

$$f^{-1}(\text{Prob}(Y_i = 1|X)) = \alpha + \beta_j G_{ij} + \gamma \text{WGA}_i + \delta \text{genoPC1}_i + \varepsilon \text{genoPC2}_i + \sum_{k=1}^{K-1} \theta_k \text{study}_{ik} + \sum_{m=1}^{K-1} \vartheta_m \text{WGA}_i \times \text{study}_{im}. \quad \text{(Model III)}$$

Forth, we added interaction between study ($K = 25$ studies) and two PCs via 48 interaction terms, $\text{study}_{ik} \times \text{genoPC1}_i$ and $\text{study}_{ik} \times \text{genoPC2}_i$ (1 if individual i is from study $k = 1...24$, Regensburg study as reference), which corresponds to estimating study-specific effects of principal components (PC1 effect of Regensburg being δ and $\delta + \tau_k$ for study k, PC2 effect of Regensburg being ε and $\varepsilon + \varphi_k$ for study k),

$$f^{-1}(\text{Prob}(Y_i = 1|X)) = \alpha + \beta_j G_{ij} + \gamma \text{WGA}_i + \delta \text{genoPC1}_i + \varepsilon \text{genoPC2}_i + \sum_{k=1}^{K-1} \theta_k \text{study}_{ik} + \sum_{m=1}^{M-1} \vartheta_m \text{study}_{im} \times \text{WGA}_i + \sum_{k=1}^{K-1} \tau_k \text{study}_{ik} \times \text{genoPC1}_i + \sum_{k=1}^{K-1} \varphi_k \text{study}_{ik} \times \text{genoPC2}_i. \quad \text{(Model IV)}$$

Fifth, we applied one model per study (25 models, $K = 25$), adjusting for genoPC1, genoPC2, and DNA source (like Model I, but now per study),

$$f^{-1}(\text{Prob}(Y_i = 1|X))_k = \alpha_k + \beta_{kj} G_{ij} + \gamma_k \text{WGA}_i + \delta_k \text{genoPC1}_i + \varepsilon_k \text{genoPC2}_i. \quad \text{(Model V)}$$

we meta-analyzed study-specific effects using the inverse-variance weighted fixed effects model, $\beta_j = \sum_{k=1}^{K} \beta_{kj} w_{kj} / \sum_{k=1}^{K} w_{kj}$, with being the inverse variance estimate for the estimated study-specific β_{kj}, as implemented in METAL (Willer, Li, & Abecasis, 2010; Web Resources).

All Models I–V assumed equal genetic effects across studies (or that the estimated average effect across studies is a reasonable effect; Rice, Higgins, & Lumley, 2017). We estimated and tested single-variant association based on Firth bias-corrected logistic regression and corresponding likelihood ratio tests (Firth, 1993), as in the original analysis, using EPACTS (Web Resources). The Firth-bias corrected test is recommended for binary trait genetic studies that include rare variants using EPACTS (number of maximum iterations for model fitting set to 150).

Different modeling approaches might work differently for the mega- or the meta-imputation approach: we thus compared results of each imputation/model combination with regard to the number of analyzable variants (see below), the *P*-value distribution under the null, and the ability to detect true disease signals.

2.4 | Number of analyzable variants

For GWAS, valid statistics are needed for as many variants as possible, to avoid false-positives and to not miss disease signals. Rare variants can lead to complete or quasi-complete separation yielding unstable, and thus unreliable, test-statistics due to the nonconvergence of the numerical optimization. We expect that the meta-model analyzing each variant by study (Model V) will have more issues with nonconvergence for rare variants than the mega-models (Model I–IV). It is also possible that models including more study-specific covariates will have more issues with rare variants than the model with the fewest covariates (Model I). There are different GWAMA approaches of how to filter variants with association results from logistic regression to ensure reliable estimates; we defined a variant as "analyzable" for a specific imputation/model (i.e., yielding trustworthy association test-statistics), if it had a MAC > 5, nonmissing effects/standard errors (*SE*)/*P*-values, and $SE < 10$ across all studies combined (for mega-models, Models I–IV) or per study (for meta-model, Model V), in analogy as described before (Mahajan et al., 2018). A higher number of analyzable variants in one imputation/model combination compared to others can be considered an advantage. We quantified the number of analyzable variants in each imputation/model combination on the example of chromosome 5.

2.5 | *P*-value distribution under the null hypothesis and genomic correction factor

In genome-wide screens for disease loci, we aim to avoid false positives. Therefore, we want to avoid inflation of test statistics under the null hypothesis, which can stem from uncontrolled confounding. Thus, the model should include covariates as much as necessary to account for confounder, but only as many as necessary to ensure sufficient power. One way to investigate the performance of a model is the inspection of the *P*-value distribution under the null in simulation experiments. When genome-wide variant data is available, a more realistic experiment for genome-wide results under the null is the empirical *P*-value distribution across variants excluding known disease signals.

Therefore, we explored whether any of the imputation/model combinations exhibited inflated/deflated statistics for all variants analyzable in all five models (see Section 2.4) on the example of chromosome 5 excluding known AMD loci (*C9* and *PRLR*/*SPEF2*, see Table S2). We visually inspected the *P*-value distribution in Quantile-quantile-plots and quantified inflation/deflation of the chi-squared statistics by the ratio of the median empirically observed chi-squared test statistics to the expected median ("lambda factors"), in analogy to the factor for genomic control (GC) correction (Devlin & Roeder, 1999).

2.6 | AMD signal detection

The genetics of (advanced) AMD is well-established with 34 AMD loci having shown genome-wide significance ($P < 5 \times 10^{-8}$) in the IAMDGC data in the original imputation/model (mega-imputation, Model I; Fritsche et al., 2016). We considered an imputation/model combination as advantageous when it identified more AMD loci than another combination, while at the same time making sure that the additionally identified loci were no artifacts. We postulated that a genome-wide significant locus was likely an artifact, when it appeared in one imputation/model result, but showed no effect or a strongly attenuated effect size in the others. Vice versa, we considered a genome-wide significant locus as a likely true locus, when effect sizes were similar across the different imputation/model approaches.

We evaluated AMD locus detection analyzing the 16,144 IAMDGC participants with advanced AMD and the 17,832 controls in each of the two imputed data sets (mega- or meta-imputed) and up to five association models (Models I–V). Furthermore, we explored extended imputation approaches and models (four imputation approaches x 5

models, see Section 2.2) in example regions and selected lead variants. We conducted all comparisons without (for comparability across imputation/models) and with GC-correction (for comparison with the original results).

For the GC correction, we computed—for each imputation/model combination—the same GC-factors as in the original work (Fritsche et al., 2016), which were based on all common genotyped variants that were in the reference panel and outside of the 34 AMD loci (Table S2). The GC-correction was applied for test statistics computed across all studies; no study-specific GC-correction was applied. When using METAL for the Model V analyses, the study-specific GC-correction was turned OFF for all analyses.

For this investigation of AMD signal detection, we included all analyzable (see Section 2.4) and well-imputed variants (RSQ $\geq$ 0.3 for common or less frequent variants, $\geq$0.8 for rare variants) as in the original work (Fritsche et al., 2016). We defined a locus as "associated with advanced AMD," when it contained at least one variant with genome-wide significance, and a "locus region" as the $\pm$500 kb region around the "lead variant" (i.e., variant with the smallest *P*-value), merging loci in case of overlapping regions. We visualized association *P*-values with Manhattan Plots (Winkler et al., 2015) or locuszoom plots (Pruim et al., 2010) and compared effect sizes across imputation/model combinations via forest plots.

3 | RESULTS

3.1 | Comparing the imputation quality between mega- and meta-imputation

First, we were interested in whether we obtained more variants with high imputation quality with the mega-imputation (mega-phasing/mega-estimation) or the meta-imputation (meta-phasing/meta-estimation). We thus compared the mega- and the meta-imputed data of the 52,189 IAMDGC subjects (Section 2). We had 27,448,454 imputed variants in both data sets, which is the exact number of reference panel variants. When computing the imputation quality as mega-RSQ or meta-RSQ (Section 2, Supporting Information Note 1), we found 9,878,708 versus 9,479,610 variants with high RSQ ($\geq$0.8) in the mega- or the meta-imputed data, respectively (Table 1). Thus, the mega-imputation yielded 399,098 more variants with high RSQ, most of them

TABLE 1 Frequency of mega- and meta-imputed variants by imputation quality

MAF category	RSQ category	No. of variants (%), mega-imputation	No. of variants (%), meta-imputation
Total	Low	6,703,742 (24.42%)	7,145,953 (26.03%)
	Medium	10,866,004 (39.59%)	10,822,891 (39.43%)
	High	9,878,708 (35.99%)	9,479,610 (34.54%)
	All	27,448,454 (100.00%)	27,448,454 (100.00%)
Monomorphic	Low	10,010 (0.04%)	767 (<0.01%)
	Medium	49 (<0.01%)	0 (0%)
	High	0 (0%)	0 (0%)
	All	10,059 (0.04%)	767 (<0.01%)
Rare	Low	6,257,084 (22.80%)	6,679,661 (24.34%)
	Medium	8,800,129 (32.06%)	8,730,179 (31.81%)
	High	2,921,222 (10.64%)	2,578,360 (9.39%)
	All	17,978,435 (65.50%)	17,988,200 (65.54%)
Less frequent	Low	293,253 (1.07%)	316,112 (1.15%)
	Medium	1,238,935 (4.51%)	1,247,128 (4.54%)
	High	1,381,942 (5.03%)	1,352,121 (4.93%)
	All	2,914,130 (10.61%)	2,915,361 (10.62%)
Common	Low	143,395 (0.52%)	149,413 (0.54%)
	Medium	826,891 (3.01%)	845,584 (3.08%)
	High	5,575,544 (20.31%)	5,549,129 (20.22%)
	All	6,545,830 (23.85%)	6,544,126 (23.84%)

Note. We applied the mega- and the meta-imputation genome-wide (No. of variants = 27,448,454 after imputation, No. of participants = 52,189). Shown are absolute and relative frequencies of variants by RSQ categories (low: RSQ < 0.3, medium: 0.3 $\leq$ RSQ < 0.8, high: 0.8 $\leq$ RSQ) and by MAF categories (monomorphic: MAF = 0, rare: 0 < MAF < 0.01, less frequent: 0.01 $\leq$ MAF $\leq$ 0.05, common: 0.05 < MAF). For mega-imputed variants, RSQ and MAF were obtained by minimac ("mega-RSQ" and "mega-MAF"), for meta-imputed variants they were computed as pooled RSQ and MAF from study-specific RSQ and MAF values ("meta-RSQ" and "meta-MAF," see Section 2.2).

being rare (342,862 variants, $0 < MAF < 0.01$). The meta-imputation yielded only 88% of the rare variants with high RSQ achieved by the mega-imputation (2,578,360 out of 2,921,222 rare variants). The number of less-frequent ($0.01 \leq MAF \leq 0.05$) and common variants ($MAF > 0.05$) with high RSQ was comparable between the two approaches, though slightly higher for the mega- compared to the meta-imputation (Table 1).

We were then interested in whether the gain in rare variants with high RSQ by mega- versus meta-imputation derived from the phasing or the estimation step. For this, we conducted an extra imputation experiment in a 2.5 MB region on chromosome 5 (chr5:37.5-40 Mb) extending to four hybrid imputation approaches: mega-phasing/mega-estimating, mega-phasing/meta-estimating, mega-phasing/meta-estimating, or meta-phasing/meta-estimating (Section 2). Overall, we obtained 26,469 imputed variants in this chromosome 5 region by each of the four approaches. We found the highest number of rare variants with high RSQ from the mega-phasing/mega-estimating, followed by the mega-phasing/meta-estimating, meta-phasing/mega-estimating, or meta-phasing/meta-estimating (3,586, 3,431, 3,126 or 3,027 variants, respectively; Table S3). We found slightly more common variants with high RSQ for the mega-mega-imputation compared to the other three approaches.

A quantitative comparison of RSQ values showed, for common and less frequent variants, similar median RSQ (Figure 1, D1–D4) across the four imputation approaches and little differences in pairwise comparisons (Figure 1, U1-U6). For rare variants, we found the highest median RSQ for mega-phasing/mega-estimating and decreasing for mega-phasing/meta-estimating, meta-phasing/mega-estimating, and meta-phasing/meta-estimating (Figure 1, D1–D4). We found larger differences also in the pairwise comparisons (Figure 1, L1–L6), predominantly when comparing different phasing approaches (Figure 1, L2, L5). Our results indicated that the gain in rare variants with high imputation quality by mega-imputation (mega-phasing and mega-estimation) compared to meta-imputation (meta-phasing and meta-estimation) was primarily driven by the mega-phasing with a smaller gain from the mega-estimation.

3.2 | Number of analyzable variants for different imputation/model approaches

We would deem an imputation/model combination as superior that yields more variants with reliable association statistics ("analyzable variants," Section 2) than other combinations, to ensure the best possible coverage of variants with association results and to avoid variants with unreliable results. Thus, we were interested in whether the mega-imputation (here again: mega-phasing and mega-estimating) yielded more analyzable variants compared to the meta-imputation (meta-phasing and meta-estimating) in dependency on the association model (Section 2). We counted the number of analyzable variants after applying each of the five association models on each of the two imputation data sets (2×5 combinations) on the example of chromosome 5 (1,808,081 imputed variants, analyzing 16,144 subjects with advanced AMD and 17,832 controls; Section 2). We found a similar proportion of analyzable variants across Models I–IV in both the mega-imputed and the meta-imputed data (~86%; Table 2). For Model V, we obtained substantially fewer analyzable variants in the mega- and the meta-imputed data (~60%) compared to the other models. This drop was due to fewer rare variants being analyzable with Model V compared to Model I–IV (Table 2). In a sensitivity analysis applying alternative filtering for variants to guarantee stable statistics ($|\beta| < 5$ instead of/ or additional to the filter of $SE < 10$), we found the same pattern. Our observation is in line with the statement by Lin and Zeng (2010a) that rare variants analyzed per study (like our Model V) have more likely invalid or unstable statistics (i.e., not being "analyzable" in our setting) than when analyzed across all studies (our Models I–IV).

3.3 | *P*-value distribution under the null for different imputation/model approaches

The empirical *P*-value distribution for all variants excluding known disease loci mimics a simulation experiment under the null hypothesis and can provide insights into inflation/deflation of tests statistics, which can point toward modeling issues. Therefore, we were interested in whether any imputation/model combination yielded association statistics that were inflated or deflated when excluding known AMD loci. Again, we analyzed the 16,144 cases and 17,832 controls for association with advanced AMD for each imputation/model combination (2×5 combinations). We investigated the *P*-value distribution for all analyzable variants (see Section 3.2) on the example of chromosome 5, excluding known AMD loci (*C9*, *PRLR*/*SPEF2*, Table S2).

When visualizing the results as QQ-Plots (Figure 2a–f), we found an excess of small *P*-values and 224 variants with $P < 5 \times 10^{-8}$ for meta-imputation/Model I, which was apparent for common and rare variants. This excess of small *P*-values and the genome-wide significance of the 224 variants disappeared when analyzing the meta-imputed data with Models II–V. We did not find this excess of small *P*-values for mega-imputation/Model I, which was not surprising, since this was the imputation/model approach in the original analysis identifying the two AMD loci on chromosome 5 (that were excluded here). The question

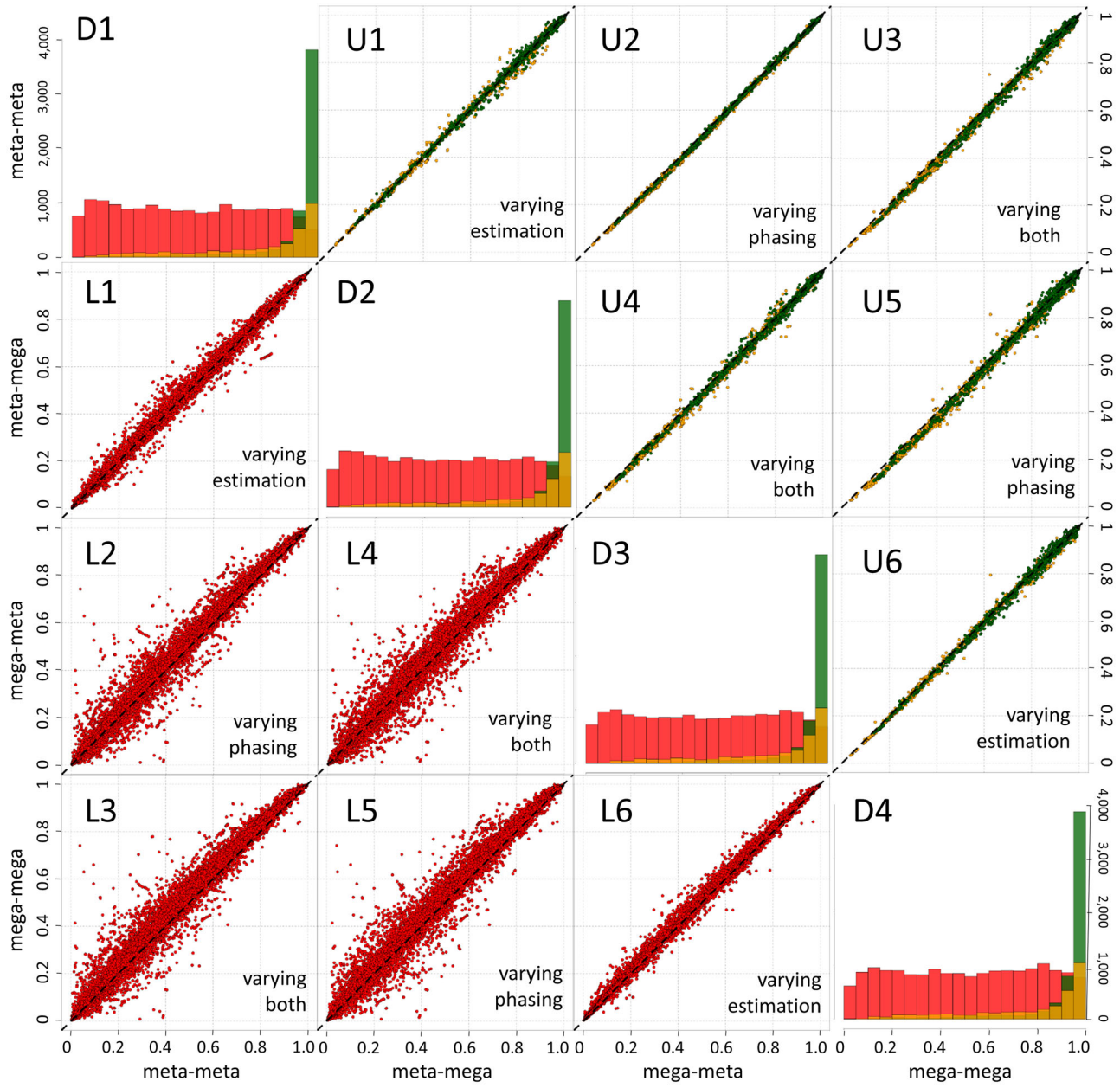

FIGURE 1 Imputation quality from four different imputation approaches on the example of one chunk on chromosome 5. We applied four different imputation approaches by conducting the two steps of imputation (phasing, estimating of untyped variants) across all studies ("mega") or by study ("meta"): mega-phasing/mega-estimating (mega–mega), mega-phasing/meta-estimating (mega–meta), meta-phasing/mega-estimating (meta–meta), meta-phasing/meta-estimating (meta–meta). This was conducted on the example of chunk chr5:37.5 M–40 Mb (#variants = 26,469). Shown is the distribution of RSQs (D1–D4) and the pairwise comparison of RSQ values between two imputation approaches (upper triangle U1–U6: common variants, MAF < 0.05, green, #variants = 5,730 and less frequent variants, 0.01 < MAF < = 0.05, orange, #variants = 3,305; lower triangle L1–L6: rare variants, 0 < MAF < 0.01, red, #variants = 17,434). The median imputation quality was 0.96 for common and less frequent variants in all four imputed data sets and 0.51, 0.50, 0.48, and 0.47 for the mega–mega, mega–meta, meta–mega, and meta–meta imputed rare variants, respectively. Variants are categorized based on the MAF computed on the mega-phased and mega-imputed variants

arose whether these "newly identified" AMD variants with meta-imputation/Model I were true loci or artifacts. When quantifying the inflation/deflation of statistics by lambda factors (Figure 2a,b) we found that λ factors were slightly above unity for Models I–IV, decreasing toward unity with increasing number of covariates in the model, for both imputation approaches (mega-imputation: $\lambda = 1.083$–1.047; meta-imputation: 1.117–1.047). This was found for the common variants and to some extent also for the rare variants (Figure 2c–f).

TABLE 2 Frequency of mega- and meta-imputed variants that were analyzable with Model I–V on the example of chromosome 5

MAF	#Overall (100%)	Model I	II	III	IV	V
Mega-imputed						
All	1,808,081	1,551,442 (85.81%)	1,549,774 (85.71%)	1,549,495 (85.70%)	1,549,400 (85.69%)	1,077,905 (59.62%)
Rare	1,215,401	958,793 (78.89%)	957,128 (78.75%)	956,849 (78.73%)	956,754 (78.72%)	485,371 (39.94%)
Less frequent	177,603	177,577 (99.99%)	177,574 (99.98%)	177,574 (99.98%)	177,574 (99.98%)	177,472 (99.93%)
Common	415,077	415,072 (100.00%)	415,072 (100.00%)	415,072 (100.00%)	415,072 (100.00%)	415,062 (100.00%)
Meta-imputed						
All	1,808,081	1,559,275 (86.24%)	1,557,904 (86.16%)	1,557,634 (86.15%)	1,557,510 (86.14%)	1,083,272 (59.91%)
Rare	1,215,432	966,654 (79.53%)	965,289 (79.42%)	965,018 (79.40%)	964,894 (79.39%)	490,764 (40.38%)
Less frequent	177,622	177,599 (99.99%)	177,593 (99.98%)	177,594 (99.98%)	177,594 (99.98%)	177,493 (99.93%)
Common	415,027	415,022 (100.00%)	415,022 (100.00%)	415,022 (100.00%)	415,022 (100.00%)	415,015 (100.00%)

Note. Shown are absolute and relative frequencies of mega- and meta-imputed variants overall (52,189 subjects) and that were analyzable (MAC > 5, nonmissing effects/*SE* and *P*-values and *SE* < 10) for Model I–V (16,144 advanced AMD cases and 17,832 controls) on the example of chromosome 5. Results are shown for all variants and by MAF categories (rare: $0 \leq \text{MAF} < 0.01$, less frequent: $0.01 \leq \text{MAF} \leq 0.05$; common: $0.05 < \text{MAF}$). Mega- and meta-imputed variants were categorized in all models by the MAF obtained from the Mega/Model I and Meta/Model I analysis, respectively. Proportions in parentheses are given relative to the number of variants per MAF category (second column, "#overall")

For Model V, which is a meta-analysis of study-specific effects, we found λ factors below unity for both mega- and meta-imputed data (Figure 2a,b, $\lambda = 0.874$ or 0.874, respectively). This was specific to rare variants (Figure 2e,f, $\lambda = 0.724$ or 0.729, respectively). The tendency for larger *P*-values compared to those expected assuming χ^2 distributed test statistics was also seen in the QQ-plots for rare variants (Figure 2e,f). The λs below unity for rare variants was not a finding specific to the meta-analyzed Firth logistic regression estimates, but was also found when meta-analyzing estimates from standard logistic regression ($\lambda = 0.83$). This can be attributed to unmet distributional assumptions (i.e., χ^2 distribution) for rare variants due to low MACs (as low as 5) in single studies.

3.4 | On the impact of different models on AMD signal detection using mega-imputed data

We consider an imputation/model approach as advantageous, when it enables the detection of more true AMD loci with the same data set than the other imputation/model approaches. A model with an increased level of accounting for covariates has less power, but also less probability to be miss-specified. Thus, we compared the results of the association analysis that mimics the original approach (Fritsche et al., 2016) (mega-imputation/Model I) with results from Model II–V and asked the question whether we would have identified the same loci, additional true or false loci. Again, we utilized the 16,144 AMD cases and 17,832 controls and all analyzable (Section 3.2) and well-imputed variants (RSQ ≥ 0.3 for common and less frequent, ≥0.8 for rare variants). Since GC-factors vary between imputation/model combinations disrupting comparability, we compared the following results without GC correction. We say that "the same locus is detected by two different models," when each model yields at least one genome-wide significant variant and the lead variant from one model resides in the locus region of the other. When re-evaluating mega-imputation/Model I, we obtained 40 loci (Figure 3a), including the 34 loci from the original analysis (Fritsche et al., 2016) and six additional loci. These six loci were also genome-wide significant in the original analysis without GC-correction, but had not been published due to the focus on GC-corrected results in the original work (Table S4).

Next, we evaluated the impact of the alternative models (Model II–V) compared to Model I on AMD signal detection in the mega-imputed data. We detected

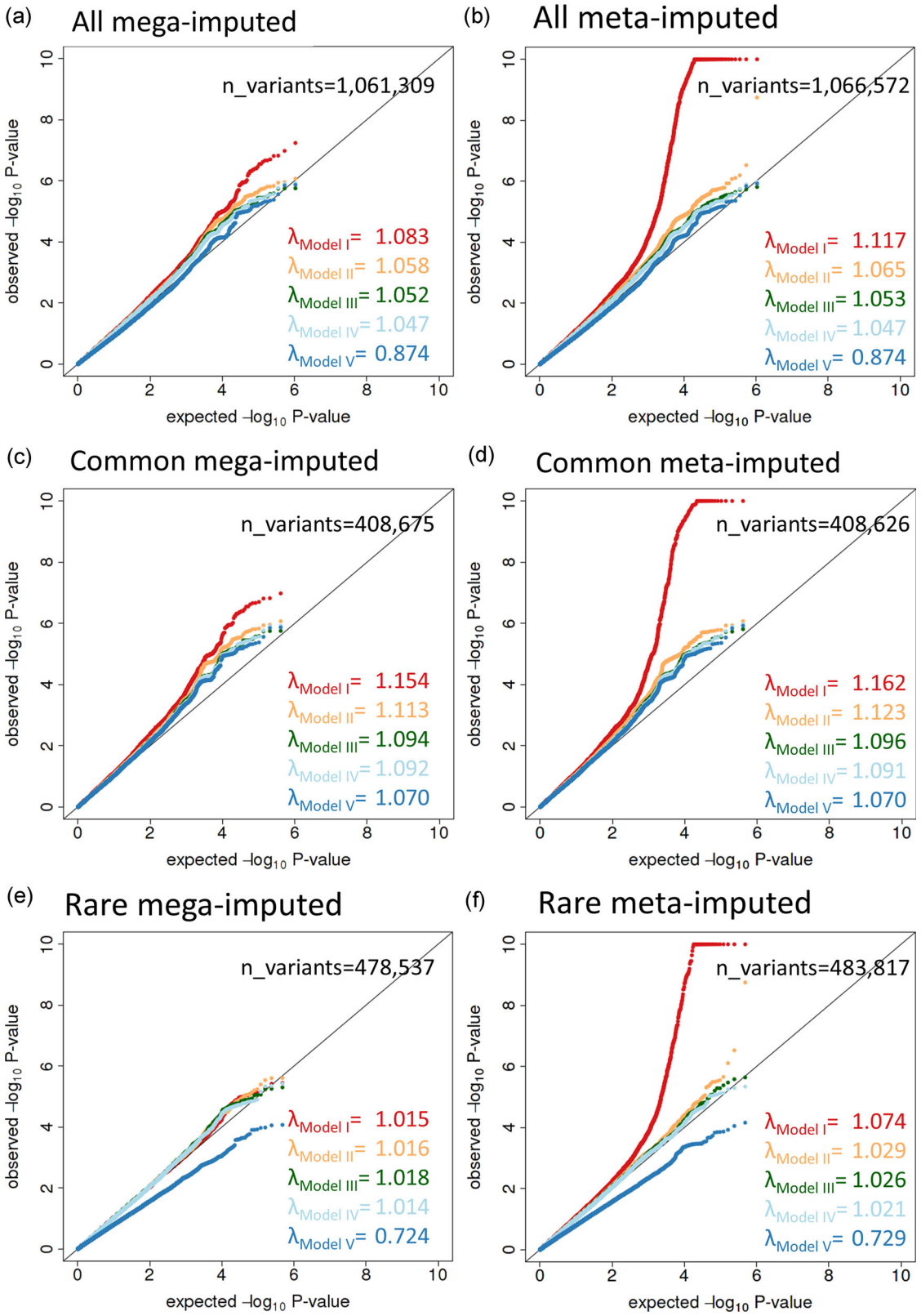

FIGURE 2 *P*-value distribution under the null for different imputation/model approaches on the example of chromosome 5. We analyzed all mega- and meta-imputed variants for association with advanced AMD (16,144 cases, 17,832 controls) with Models I–V (red, orange, green, light blue, and blue, respectively). Shown are observed vs. expected *P*-values and λ value for (a,b) all variants on chromosome 5 and (c,d) separately for the common (MAF > 0.05) or (e,f) rare variants (MAF < 0.01). AMD: age-related macular degeneration; MAF: minor allele frequency

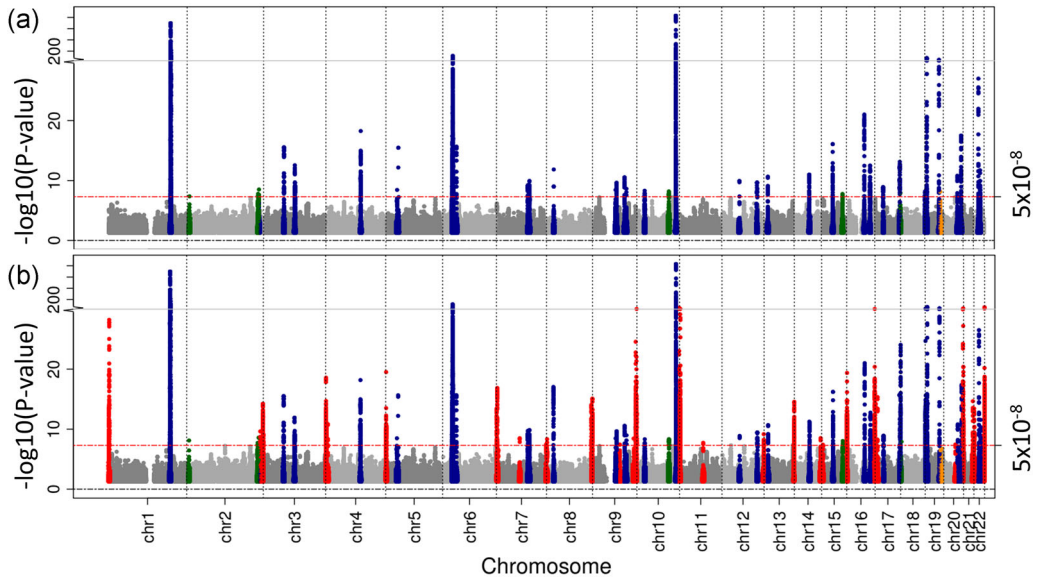

FIGURE 3 AMD signal detection for mega- and meta-imputed data analyzed with Model I. These results are from the association analysis for advanced AMD (16,144 cases and 17,832 controls) without GC-correction. Shown are genome-wide minus log10 *P*-values vs. genomic position for (a) the mega-imputation/Model I and (b) the meta-imputation/Model I. Color coded are the 34 loci identified in the original publication with GC-correction (Fritsche et al., 2016) (blue), the *NTN5* locus (orange, $P = 6.8 \times 10^{-9}$ in a, $P = 3.4 \times 10^{-7}$ in b), the additional five signals with $P < 5 \times 10^{-8}$ (green), and the 31 additional signals with $P < 5 \times 10^{-8}$ in b (red). AMD: age-related macular degeneration; GC: genomic control correction

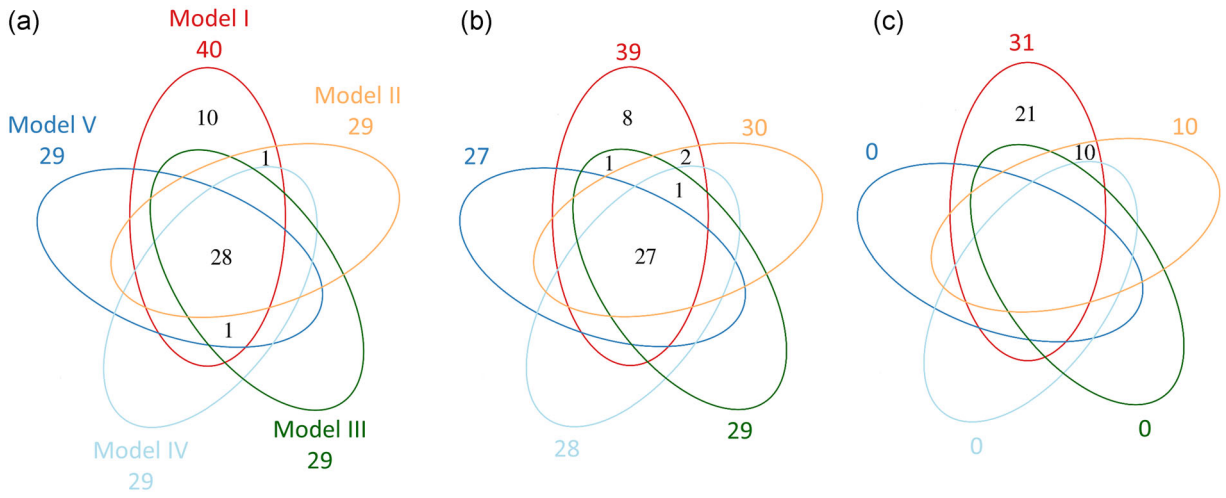

FIGURE 4 AMD locus detection in the mega- and meta-imputed data comparing Model I results with the other models. These results are from association analysis for advanced AMD (16,144 cases and 17,832 controls) without GC-correction. For the (a,b) 40 loci detected with $P < 5 \times 10^{-8}$ with mega-imputation/Model I and the (c) 31 additional loci detected with meta-imputation/Model I, we show the number of genome-wide significant loci (at least one variant with $P < 5 \times 10^{-8}$) detected by each of other models (Models II-V) (a) in the mega-imputed data (among the 40 loci), (b) in the meta-imputed data (among the 40 loci), and (c) in the meta-imputed data (among the 31 loci). AMD: age-related macular degeneration; GC: genomic control correction

29 genome-wide significant loci by each of the Models II–V (Figure 4a; Table S5). There were several interesting results: (a) all of the loci detected by Models II–V were among the 40 loci from Model I, making Model I the "most powerful model" if all 40 loci were genuine. (b) There were 28 loci that were identified by all five models, indicating that these loci were robust to different modeling and can be considered genuine ("safe loci"). This is supported by equal effect sizes across models (example for "safe" locus shown in Figure 5a; effect sizes

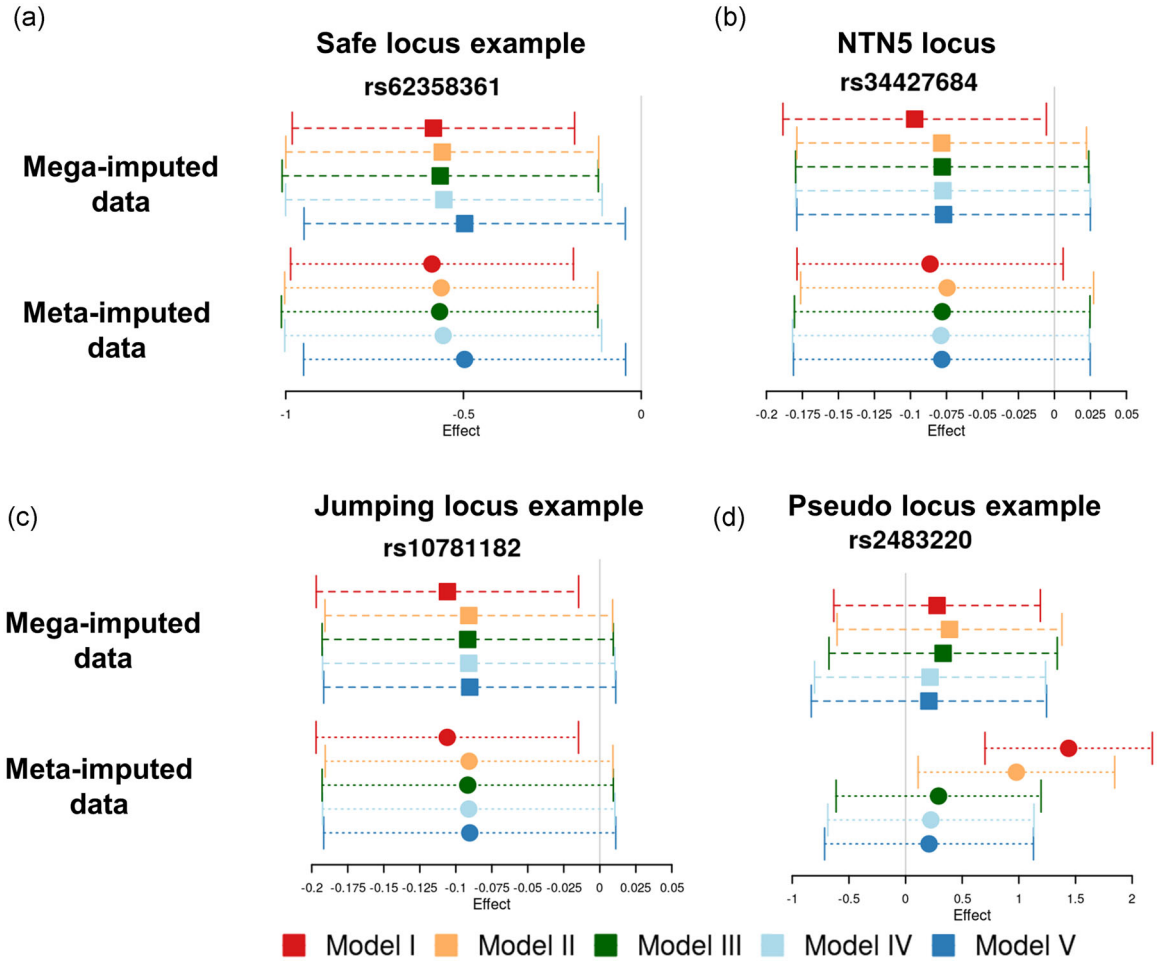

FIGURE 5 Effect sizes across imputation/model approaches on the example of four different types of variants. Results are shown from association analyses for advanced AMD (16,144 cases and 17,832 controls) without GC correction. Shown are effect sizes and confidence intervals ("genome-wide": effect ± 5.45 *SE*) for Models I–V (red, orange, green, light blue, and blue, respectively) illustrating example lead variants for four different types of loci: (a) a locus detected with genome-wide significance ($P < 5 \times 10^{-8}$) in all imputation/model approaches ("safe locus"), (b) the sole locus (*NTN5*) of the 40 that was detected with mega-imputation/Model I, but not with meta-imputation/Model I, (c) a locus that was detected by mega-imputation/Model I, but not with all models ("jumping locus"), and (d) a locus detected only with meta-imputation/Model I or Model II ("pseudo-signal"). AMD: age-related macular degeneration; GC: genomic control correction; *SE*: standard error

for all loci in Figure 3a). We also found a tendency toward larger (though still genome-wide significant) *P*-values and *SE* for models with higher levels of accounting for covariates, which mirrors the lower power for these models to detect these loci. (c) There were further 12 loci that were identified by at least one model, but not with each model ("jumping loci"). Of these, 10 loci were solely detected by Model I, two loci by Models I + II or Models I + III + IV + V (*KMT2E/SRPK2* and *MIR6130/RORB,* respectively, Figure 4a). We found similar *P*-values across all five models for all of these 12 loci (Table S5), just barely missing genome-wide significance for one or the other model, but always hitting genome-wide significance with Model I. We also found similar effect sizes across models (example for "jumping" locus in Figure 5c; effect sizes for all loci in Figure S3b). Therefore, we considered these 12 loci as genuine. (d) When looking more closely into the effect sizes across models and judging even smaller deviations, we found that Model II–V effect sizes were very stable across the four models for the 40 Model I lead variants, but Model I effect sizes differed slightly from these estimates for several of these 40 variants (Figure S3A,B, example variant Figure 5c). This might indicate a small bias from Model I from ignoring study membership due to the differing case–control ratios across the 25 IAMDGC studies (0.06–5.8), but this is difficult to judge conclusively from the observed data.

In summary, we found—for the mega-imputed data—that Model I yielded the most genome-wide significant loci compared to Models II–V. Since *P*-values and effect sizes were comparable across all models for all 40 lead

variants, we considered these signals as genuine and as detectable with genome-wide significance in Model I most likely due to the fewer parameters in the model compared to the other models.

3.5 | On the impact of meta-imputation compared to mega-imputation on AMD signal detection using model I

Mega-imputation of large-scale IPD, as conducted in the original IAMDGC analysis, is computationally much more challenging than a study-specific imputation (meta-imputation). We evaluated, on the example of our IAMDGC data, how the choice of meta-imputation rather than mega-imputation would have affected AMD signal detection (using Model I, 16,144 advanced AMD cases vs. 17,832 controls).

For meta-imputation/Model I, we identified 70 loci, including 39 of the 40 loci that we had detected with mega-imputation/Model I (Figure 3b): (a) For these 39 loci, we found similar effect sizes and *P*-values for meta-imputation/Model I and mega-imputation/Model I, with larger *P*-values (and larger confidence intervals) for the meta-imputed data (Table S6 and Figure S3). From this we conclude that mega-imputation had better power to detect these loci (given that we had already concluded that these loci were genuine, see above), but the meta-imputation would have been sufficient to detect these. (b) The one locus missed here (*NTN5* locus) showed a $P = 3.4 \times 10^{-7}$ comparable to mega-imputation/Model I ($P = 6.8 \times 10^{-9}$) and exhibited similar effect size (Figure 5b). We thus conclude that this locus was barely missed here, but genuine and detected in the mega-imputed data by chance or due to larger power. (c) For the 31 loci identified by meta-imputation/Model I "de novo" (i.e., not detected by mega-imputation/Model I), we found no immediate explanation: lead variants in these 31 loci had a wide range of MAF 0.0026–0.49) and effect sizes ($\beta = -2.47$ to 1.90; Table S7); so this was not specific to rare variants or small effects. Furthermore, lead variants were not characterized by low RSQ (RSQ ≥ 0.8 for 15 of these 31 variants). However, most of these 31 loci (28 of these) exhibited a striking pattern by residing at the tails of the chromosomes or near gaps in the reference panel (Table S7). We did not find any of the pseudosignals when analyzing the genotyped data without imputation (data not shown). We hypothesized that these 28 loci were artifacts due to insufficient LD information from flanking regions, which appeared as an issue in the meta-imputed, but not in the mega-imputed Model I analysis. The other three of the 31 were distant from tails or gaps and thus inconclusive with regard to why these signals appeared in meta-imputed/Model I, but not mega-imputed/Model I.

3.6 | On the impact of different models on AMD signal detection using meta-imputed data

We were interested in the impact of alternative Models II–V on the meta-imputed results: we asked the question whether the 70 loci detected by meta-imputation/Model I showed robust results across models and whether the one locus missed by meta-imputation/Model I compared to mega-imputation/Model I (*NTN5*) was detectable in any of the other models. We thus analyzed all variants in these 71 locus regions with Models II–V in the meta-imputed data.

We found the following: (a) among the 39 loci that were detected by both meta-imputation/Model I and mega-imputation/Model I, 27 loci were detected by all the other four models in the meta-imputed data (Figure 4b, Table S6). These 27 were among the 28 loci identified by all five models in the mega-imputed data. Thus, the detection of these 27 loci was independent of imputation or model approach ("safe loci," see Section 3.4). (b) The *NTN5* locus (the one missed by meta-imputation/Model I compared to mega-imputation/Model I) was not detected in any of the Models II–V in the meta-imputed data (smallest P in each of these models 3.02×10^{-5} to 6.46×10^{-5}). (c) Among the 31 loci detected by meta-imputation/Model I "de novo," none of these survived in Models III–V with genome-wide significance (i.e., $P \geq 5 \times 10^{-8}$) and the 31 signals showed little excess of small *P*-values (most signals with $P > 0.05/31 = 0.0016$, four signals with $P <= 0.0016$ consistently across Models III–V); 10 of the 31 still "survived" at genome-wide significance in Model II and there was an excess of low *P*-values also among the 21 other loci (13 of these 21 signals with $P <= 0.0016$), Figure 4c; Table S8). The effect sizes vanished or decreased substantially for Models III–V compared to Model I + II (example variant Figure 5d, effect sizes for all 31 variants, Figure S3C). This supports the notion that these loci were pseudo-signals from meta-imputation/Model I.

3.7 | Measurement error by meta-imputation as source for pseudosignals

It was interesting that the models accounting for between-study differences in WGA exposure (Models III–V, see Section 3.6) did not detect any of the 31 pseudosignals from meta-imputation/Model I at genome-wide significance. From the 25 studies in our data, four studies included study participants with WGA as DNA sources ("WGA-studies"). The proportion of participants with WGA varied between these studies 0.06–0.90); this proportion varied—within study—between cases and

controls (0.03 in cases vs. 0.06 in controls, 1.00 vs. 0.59, 0.55 vs. 1.00, and 0.32 vs. 0.90 in the 4 WGA-studies, Table S1), but not when joining the four studies (0.06 in cases, 0.06 in controls). Thus, WGA status and AMD were associated in each of the four WGA-studies with varying extent (OR of AMD for WGA vs. no WGA status = 0.53, 1.7, 0.55, 0.35).

Additionally, we have inferior DNA quality for WGA-participants (and thus in WGA-studies on average) compared to non-WGA-participants (non-WGA-studies) leading to a larger genotyping error and more missing genotypes. A variant with a larger proportion of missing genotypes is potentially subjected to larger imputation error, when genotype and LD information from flanking variants is insufficient. We thus expect a larger "measurement error" from imputation in allele dosages among WGA-studies versus non-WGA-studies. When inspecting the meta-imputed 31 pseudo-signal lead variants, we found different MAFs in the dosages among the four WGA studies, which also differed—more or less—from the MAF for non-WGA studies (e.g., rs79879353 MAF = 0.01–0.02 for non-WGA-studies and 0.01, 0.03, 0.08, and 0.32 in WGA studies; rs2289165 MAF = 0.21–0.26 in non-WGA-studies and 0.09, 0.15, 0.23, and 0.42 in WGA-studies). The latter was a genotyped variant not showing MAF differences in the genotypes (MAF 0.23–0.31 in non-WGA studies and 0.24, 0.26, 0.33, and 0.34 in WGA-studies), but differences in the proportion of missing genotypes (<0.01 of participants in each non-WGA-study, 0.08, 0.62, 0.83, 0.91 in WGA-studies). We did not find MAF differences for the mega-imputed pseudo-signal lead variants between WGA-studies and non-WGA-studies (MAF 0.01–0.04 for all studies for rs79879353 and 0.21–0.32 for rs2289165). The observed MAF differences in meta-imputed variants between WGA- and non-WGA-studies supported our hypothesis of a "measurement error" in allele dosages that differed within WGA-studies and compared to non-WGA-studies.

Why does this "measurement error" in allele dosages yield pseudosignals for AMD? When considering one of the WGA-studies, the measurement error induces an association between WGA status and dosages of some variants. The difference in AMD case-control ratios between WGA-participants and non-WGA-participants would yield a spurious association between dosages and AMD by confounding, when WGA-status was not accounted for in the model. When joining the four WGA-studies, we have an association of WGA-status and dosages with varying extent between WGA-studies and a varying association between WGA-status and AMD, which induces a spurious association between dosage and AMD when WGA-status is not accounted for study-specifically (Figure S4). This is consistent with our observation that most of the pseudosignals disappeared when including an interaction of study membership and WGA-status into the model (Models III–IV), while the mere inclusion of WGA-status as covariate in the mega-model was not always sufficient (Models I–II). The study-specific analysis including WGA-status also accounted effectively for this issue (Model V). We thus hypothesize that the majority of pseudosignals were derived by unaccounted confounding from varying imputation error and an unlucky situation of our strongly varying case-control ratios between WGA-studies.

We followed up on our suspicion that the WGA-studies triggered pseudosignals by restricting the Model I analysis to the 21 non-WGA-studies (14,953 cases and 15,414 controls, Table S8) and found the following. (a) For 24 of 31 pseudo-signal lead variants, the association disappeared ($P > 0.05/31 = 0.0016$, one example in Figure S5A,B). We would thus deem these 24 pseudosignals as WGA-study-related. (b) It was interesting to look at the other seven of the 31 lead variants (one still genome-wide significant and six that showed some association $P <= 0.0016$): these seven included the three pseudosignals that did not reside at chromosome tails/gaps. We would deem these seven pseudosignals as related to a study-specific issue, but not necessarily to a WGA-study or study × WGA-interaction issue. Overall, our observations supported our hypothesis of an issue in the WGA-studies that triggered the majority of the 31 pseudosignals.

We also followed up on our observation that the pseudosignals appeared in meta-imputed (meta-phased/meta-estimated), but not in mega-imputed (mega-phased/mega-estimated) data by conducting an extra imputation experiment: based on mega- or meta-phased genome-wide haplotypes, we re-estimated all variants in the 31 "chunks" (i.e., 2.5 Mb region used for the estimation step) containing the 31 pseudo-signal lead variants across (a) all studies (mega-estimation) or (b) by study (meta-estimation). Then we extracted the 31 meta-imputation/ Model I lead variants in the 31 pseudosignals from the four imputed data sets. By this, we obtained allele dosages for all participants for each of the 31 pseudo-signal lead variants from four imputation approaches (meta-phased/meta-estimated, meta-phased/mega-estimated, mega-phased/meta-estimated, mega-phased/mega-estimated). We repeated association analyses for these 31 variants using the five models (4 × 5 results for each of the 31 variants). For most of these 31 variants, we found the same pattern in the meta-phased/mega-estimated data as in meta-phased/meta-estimated data: a genome-wide significant signal for Model I, but not for Models III–V (e.g., rs2483220 Model I $P = 2.14 \times 10^{-36}$, Model III–V P = 0.11 to 0.25; Table S9). Most pseudosignals vanished in the mega-phased/meta-estimated/Model I + II. From this, we

conclude that the issue derived from the meta-phasing rather than the meta-estimation.

We then investigated if including the reference haplotypes into the study-specific phasing eliminated the observed pseudosignals (on the example of the region around the chromosome 1 pseudo-signal, chr1:0-5 MB). We found still a genome-wide significant pseudo-signal at the start of the chromosome with Model I that was diminished with Model II and vanished with Models III–V. We concluded that this alternative phasing did not avoid the pseudo-signal generation.

In summary, we conclude that a majority of the 31 signals detected additionally with meta-imputation/Model I (not detected by mega-imputation/Model I) were pseudosignals from an issue in the WGA-studies including a WGA-related imputation error and the varying case–control ratio in WGA-studies.

3.8 | AMD genetics in IAMDGC data with alternative imputation/model approaches

When identifying genetic loci for advanced AMD is the primary objective, GC-corrected *P*-values are to be judged. While the original work had applied a sequential forward selection approach to detect independent variants and identified 34 loci (Fritsche et al., 2016), the same 34 loci were detected by our regular single-variant association search genome-wide in the same imputed data. Our mega-imputation/Model I approach with GC correction can be considered a repetition of the originally published imputation model (same software specifications). In our mega-imputation/Model I results with GC correction ($\lambda = 1.13$ as in the original publication), we detected 36 loci for advanced AMD including the 34 loci described before and two new loci (near *DNMP* and *NTN5,* both $P = 4.9 \times 10^{-8}$ in our analysis vs. 1.26×10^{-7} and 1.23×10^{-7} originally, Table S4). We attributed these two new loci to slight differences in allele dosages and thus a chance difference from reimputation (Figure S2 and Supporting Information Note 2).

When we applied Models II-V with GC correction in the mega-imputed data, we obtained 26, 27, 25, and 29 genome-wide significant loci, respectively (GC-factor = 1.10, 1.09, 1.09, and 1.01, respectively). All of these were included in the 36 loci (Table S5). The consistency of effect sizes across the models also holds for GC-corrected results (Section 3.3), since effect sizes do not change upon GC-correction.

4 | DISCUSSION

Here, we investigated whether mega-imputation or meta-imputation was preferable for genome-wide searches for disease loci across multiple studies, when the data is available as IPD. To exemplify this on real data, we utilized one of the largest case-control data sets on advanced AMD with genome-wide information, the IAMDGC data, on >50,000 participants from 25 studies. We found that the mega-imputation was superior to the meta-imputation with respect to a higher number of well-imputed variants, particularly for rare variants, and more genuine AMD loci detectable with genome-wide significance.

When comparing differently imputed genetic data sets, it is not fully straight forward as to how to compare these data sets given the high dimensionality of the data. We opted to count the number of variants imputed with high imputation quality and the number of variants that yielded trustworthy association statistics for a specific imputation/model combination ("analyzable variants"). We found that meta-imputation yielded only 88% of the rare variants at high imputation quality achieved by the mega-imputation. For the number of analyzable variants, we found that the choice of the model was more relevant than the choice of the imputation approach: only ~60% of variants in the reference panel were analyzable with a "meta-model" (Model V) compared to ~85% analyzable with a "mega-model" (i.e., one model applied to the full data, Models I–IV). This drop was predominantly due to rare variants: among the rare variants analyzable with mega-models, only 50% were analyzable with the meta-model. These results highlight the gain from mega-models to analyze rare variants independent of the imputation approach.

Another approach to compare differently imputed data sets is a comparison of disease signal detection. Among the overall 20 different approaches towards imputation and association modeling that we explored, we found the highest number of genuine AMD signals when analyzing the mega-imputed data with an association model fully ignoring study membership (Model I), namely 40 loci without GC-correction and 36 loci with GC-correction. The lead variants' effect estimates were stable across the five models, and also *P*-values were similar, but tended to increase for models with increasing level of accounting for covariates. Based on this observation, we considered these signals as genuine and Model I as a reasonable parsimonious model for a genome-wide screen in the mega-imputed IAMDGC data. However, Model I does not account for the varying case-control ratios across the 25 IAMDGC studies, and is therefore not an ideal model for this data. A model accounting for study membership (Model II), and thus for varying case-control ratios, would be a good alternative and a follow-up of findings with more complex models is advisable. In fact, the ideal logistic regression model to quantify a variant's effect for AMD is a model that includes all associated risk factors,

that is all other associated genetic variants and age (additionally to study membership).

When applying Model I on the meta-imputed data, we found 31 signals additional to those detected with Model I in the mega-imputed data, most of them located near tails of chromosomes or gaps in the reference panel. We considered these 31 signals as pseudosignals, since they vanished with other models. We attributed the majority of these signals to an issue in the WGA-studies, mostly from unaccounted confounding from WGA-related imputation error and the varying case-control ratios within the four WGA-studies. The imputation error was inconspicuous in the mega-imputed data, but re-enforced by meta-imputation, possibly due to the weaker information in flanking variants in the study-specific phasing. While we found this issue accounted for by modeling the interaction of study membership and WGA-status (Models III + IV) or by excluding WGA-studies, this requires the issue to be known. A more general recommendation would be a meta-analysis (Model V) for meta-imputed data.

There is no previously published work on the comparison of mega-imputation with meta-imputation. A large body of work on clinical trials highlights scenarios where mega-analysis and meta-analysis are equivalent (Mathew & Nordstrom, 1999; Olkin & Sampson, 1998). There is some work comparing mega- and meta-analysis for genetic variants: Lin and Zeng (2010a) analyzed 224 genotyped SNPs (4,792 subjects from two studies) with various models similar to our Models II–V. They demonstrated empirically little differences between mega- and meta-analyses and concluded that the differences observed in variants with low MAF (<3.1%) were due to the instability in effects estimates. The same authors showed, theoretically and empirically (30 genotyped variants, 3,135 cases of depression and 3,266 controls, four studies) that meta-analysis was statistically as efficient as mega-analysis evaluating different models comparable to our Models II, III, and V (Lin & Zeng, 2010b). Both publications were based on genotyped variants without aspects of imputation. Some previous work investigated aspects of imputation: in one study genotyped twice with different arrays (1,952 subjects; 32,903 or 14,071 SNPs on chromosome 22 with Illumina Omni 2.5 or Illumina 1 M, respectively) and imputed (411,376 variants, using 1000G), the impact of the different arrays on variant imputation quality and signal detection was explored (Xie, Hancock, Johnson, & Rice, 2014). In another investigation merging nine studies genotyped with different arrays (overall 51,000 subjects), the steps of mega-imputation including variant panel harmonization and quality control were described without association analysis (Verma et al., 2014). While these investigations address an important aspect of imputation across multiple studies, namely the issue of different genotyping arrays, they do not compare different imputation approaches or association models. In contrast, we focused on issues that were beyond diverse array technologies, for which our IAMDGC data was ideal, since all samples were genotyped on one array centrally. Another work applied meta-imputation, but not mega-imputation (131,880 HapMap-II-imputed variants on chromosome 16 and 18; 9,791 subjects from four studies) to search for gene-smoking interaction for systolic blood pressure; they compared mega-analysis to meta-analysis with different interaction modeling (Sung et al., 2014).

Our work thus represents the first investigation of mega-imputation compared to meta-imputation, which fills a current gap, and we investigate this systematically in combination with various association models. Furthermore, our investigation is substantially larger in the number of participants and the number of variants as a realistic scenario of GWAMA research. An alternative approach to explore structure in differently imputed data sets is the determination of genetic principal components and their inclusion as covariates in association models. We did not analyze this in full detail in our study, which can be considered a limitation. Our investigation was restricted to the analysis of unrelated samples, as in the original analysis (Fritsche et al., 2016), and on a binary outcome. Some of our findings apply also to continuous outcome: our conclusions on imputation quality, the equivalence of results from mega- compared to meta-analysis for common variants, and the potential confounding from study-specific differences in imputation error when the mean outcome differs between studies. Our work focused on analytical steps after genotype quality control: we had genotypes assessed centrally with the same chip array across all individuals and we kept the genotype quality control constant across our scenarios. This leaves issues from non-centralized genotyping, heterogeneous variant panels, and study-specific versus overall genotype quality control to future research.

Our investigation has a particular strength due to the fact that our IAMDGC data was genotyped with one array centrally, which limits the sources of between-study differences. The only uncontrolled source of between-study difference in this genetic data was the difference in DNA isolation: four studies included study participants where DNA was derived via WGA. This uncontrolled source was then promptly the reason for most of the pseudosignals in the meta-imputed data. The fact that our data included WGA-participants can be considered a limitation, but also an opportunity for methodological investigation. It should be noted that the IAMDGC data was designed to conduct mega-imputation and mega-analysis and not for a study-specific approach. It is possible, that IPD from multiple studies, where each study is well designed with matched cases and controls,

would not feature such issues as the ones we observed. Therefore, our investigation might be considered a rather extreme example of unaccounted confounding; still, this example was able to highlight difficulties of meta-imputation and helped us understand this methodological issue in principle.

To what extent are our experiences from investigating the IAMDGC data relevant for future studies? The issue that some of our studies included participants with WGA is not ubiquitous to GWAS. However, it is always possible - and not necessarily known or observable—that there are study-specific issues of any kind: it might not be fully clear what these issues are, making it impossible to define an adequate mega-model. The inclusion of all possible study characteristics and their interaction with study membership into a mega-model is not an appealing alternative, due to the substantially increased computational burden. A mega-imputation followed by a mega-model ignoring study membership might be a reasonable parsimonious approach for genome-wide searches also for other scenarios with central genotyping on a single array. Still, such a simple model is not ideal and more fine-tuned models need to be considered, including a detailed exploration of potential sources of bias or confounding that will be specific to each situation. The question of how to impute and analyze large multistudy or multisite IPD is particularly timely given the numerous and large study data that becomes more and more available as IPD by open access research platforms (dbGaP, UKBB; Bycroft et al., 2018; Rich et al., 2016), the European initiative for open research data, Web Resources). We recommend mega-imputation and mega-analysis, with meta-imputation followed by meta-analysis being a computationally appealing alternative. For multisite studies still at the brink of DNA isolating and chipping, like the NAKO study from 18 sites across Germany (German National Cohort, 2014), our findings support the value of central DNA isolation and genotyping or a random assignment of participants to DNA isolation/genotyping site combined with appropriate modeling. With regard to the question whether it is worthwhile to gather IPD of GWAS rather than sticking to the current approach of gathering summary statistics, our results indicate that meta-imputation in combination with meta-analysis works well for common and less frequent variants. However, we also find a potential gain from IPD for rare variants and, of course, from more options towards exploring the data.

In summary, our comprehensive investigation of different imputation and modeling approaches for multistudy GWAS provides insights into the handling of such data that will facilitate future GWAS and help enhance the detection of disease signals.

5 | WEB RESOURCES

The URLs for data presented herein are as follows:

EPACTS: Association analysis tool: https://genome.sph.umich.edu/wiki/EPACTS;

GIANT ALL 1000G PhaseI v3 reference panel: http://www.sph.umich.edu/csg/abecasis/MaCH/download/1000G.2012-03-14.html;

METAL: Meta-analysis software: https://genome.sph.umich.edu/wiki/METAL_Documentation;

Imputation software minimac: https://genome.sph.umich.edu/wiki/Minimac;

European initiative for open research data: http://ec.europa.eu/research/openscience/index.cfm?pg=openaccess;

Imputation quality RSQ definition: http://genome.sph.umich.edu/wiki/Minimac3_Info_File;

ACKNOWLEDGMENTS

This study is funded by the project "Advancing genomics through the AMD Genomics Consortium" (NIH-2017 R01 EY RES511967 to Prof. Dr. Iris Heid) of the National Institutes of Health (NIH) and the project "On the gain of mega-analyses-versus meta-analyses for genome-wide association studies" (DFG HE3690/5-1 to Prof. Dr. Iris Heid) from the Deutsche Forschungsgemeinschaft (DFG). The authors thank the IAMDGC (http://amdgenetics.org). The list of consortium members reflects the author list of the previous publication by Fritsche et al. (2016), listed in the Supplement.

ORCID

Mathias Gorski http://orcid.org/0000-0002-9103-5860

REFERENCES

Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L. T., Sharp, K., ... Marchini, J. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature*, *562*(7726), 203–209. https://doi.org/10.1038/s41586-018-0579-z

Delaneau, O., Marchini, J., & Zagury, J. F. (2011). A linear complexity phasing method for thousands of genomes. *Nature Methods*, *9*(2), 179–181. https://doi.org/10.1038/nmeth.1785

Devlin, B., & Roeder, K. (1999). Genomic control for association studies. *Biometrics*, *55*(4), 997–1004.

Firth, D. (1993). Bias reduction of maximum likelihood estimates. *Biometrika*, *80*(1), 27–38.

Fritsche, L. G., Igl, W., Bailey, J. N., Grassmann, F., Sengupta, S., Bragg-Gresham, J. L., & Heid, I. M. (2016). A large genome-wide association study of age-related macular degeneration highlights

contributions of rare and common variants. *Nature Genetics*, *48*(2), 134–143. https://doi.org/10.1038/ng.3448

German National Cohort Consortium (2014). The German National Cohort: Aims, study design and organization. *European Journal of Epidemiology*, *29*(5), 371–382. https://doi.org/10.1007/s10654-014-9890-7

Klarin, D., Damrauer, S. M., Cho, K., Sun, Y. V., Teslovich, T. M., Honerlaw, J., & Assimes, T. L. (2018). Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. *Nature Genetics*, *50*(11), 1514–1523. https://doi.org/10.1038/s41588-018-0222-9

Lappalainen, I., Almeida-King, J., Kumanduri, V., Senf, A., Spalding, J. D., Ur-Rehman, S., & Flicek, P. (2015). The European Genome-phenome Archive of human data consented for biomedical research. *Nature Genetics*, *47*(7), 692–695. https://doi.org/10.1038/ng.3312

Li, Y., Willer, C. J., Ding, J., Scheet, P., & Abecasis, G. R. (2010). MaCH: Using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genetic Epidemiology*, *34*(8), 816–834. https://doi.org/10.1002/gepi.20533

Lin, D. Y., & Zeng, D. (2010a). Meta-analysis of genome-wide association studies: No efficiency gain in using individual participant data. *Genetic Epidemiology*, *34*(1), 60–66. https://doi.org/10.1002/gepi.20435

Lin, D. Y., & Zeng, D. (2010b). On the relative efficiency of using summary statistics versus individual-level data in meta-analysis. *Biometrika*, *97*(2), 321–332. https://doi.org/10.1093/biomet/asq006

MacArthur, J., Bowler, E., Cerezo, M., Gil, L., Hall, P., Hastings, E., & Parkinson, H. (2017). The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Research*, *45*(D1), D896–D901. https://doi.org/10.1093/nar/gkw1133

Mahajan, A., Taliun, D., Thurner, M., Robertson, N. R., Torres, J. M., Rayner, N. W., & McCarthy, M. I. (2018). Fine-mapping type 2 diabetes loci to single-variant resolution using high-density imputation and islet-specific epigenome maps. *Nature Genetics*, *50*(11), 1505–1513. https://doi.org/10.1038/s41588-018-0241-6

Mathew, T., & Nordstrom, K. (1999). On the equivalence of meta-analysis using literature and using individual patient data. *Biometrics*, *55*(4), 1221–1223.

Olkin, I., & Sampson, A. (1998). Comparison of meta-analysis versus analysis of variance of individual patient data. *Biometrics*, *54*(1), 317–322.

Pruim, R. J., Welch, R. P., Sanna, S., Teslovich, T. M., Chines, P. S., Gliedt, T. P., & Willer, C. J. (2010). LocusZoom: Regional visualization of genome-wide association scan results. *Bioinformatics*, *26*(18), 2336–2337. https://doi.org/10.1093/bioinformatics/btq419

Rice, K., Higgins, J.P.T., & Lumley, T. (2017). A re-evaluation of fixed effect(s) meta-analysis. *Journal of the Royal Statistical Society*, *181*(1), 205–227.

Rich, S. S., Wang, Z. Y., Sturcke, A., Ziyabari, L., Feolo, M., O'Donnell, C. J., & Psaty, B. M. (2016). Rapid evaluation of phenotypes, SNPs and results through the dbGaP CHARGE summary results site. *Nature Genetics*, *48*(7), 702–703. https://doi.org/10.1038/ng.3582

Scott, R. A., Scott, L. J., Magi, R., Marullo, L., Gaulton, K. J., Kaakinen, M., & Meta-analysis, C. (2017). An expanded genome-wide association study of type 2 diabetes in Europeans. *Diabetes*, *66*(11), 2888–2902. https://doi.org/10.2337/db16−1253

Sung, Y. J., Schwander, K., Arnett, D. K., Kardia, S. L., Rankinen, T., Bouchard, C., & Rao, D. C. (2014). An empirical comparison of meta-analysis and mega-analysis of individual participant data for identifying gene-environment interactions. *Genetic Epidemiology*, *38*(4), 369–378. https://doi.org/10.1002/gepi.21800

Turcot, V., Lu, Y., Highland, H. M., Schurmann, C., Justice, A. E., & Fine, R. S., Understanding Society Scientific Group. (2018). Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. *Nature Genetics*, *50*(1), 26–41. https://doi.org/10.1038/s41588-017-0011-x

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*, 370. https://doi.org/10.3389/fgene.2014.00370

Visscher, P. M., Wray, N. R., Zhang, Q., Sklar, P., McCarthy, M. I., Brown, M. A., & Yang, J. (2017). 10 Years of GWAS discovery: Biology, function, and translation. *American Journal of Human Genetics*, *101*(1), 5–22. https://doi.org/10.1016/j.ajhg.2017.06.005

Willer, C. J., Li, Y., & Abecasis, G. R. (2010). METAL: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics*, *26*(17), 2190–2191. https://doi.org/10.1093/bioinformatics/btq340

Winkler, T. W., Day, F. R., Croteau-Chonka, D. C., Wood, A. R., Locke, A. E., Magi, R. ... Genetic Investigation of Anthropometric Traits (GIANT) Consortium. (2014). Quality control and conduct of genome-wide association meta-analyses. *Nature Protocols*, *9*(5), 1192–1212. https://doi.org/10.1038/nprot.2014.071

Winkler, T. W., Kutalik, Z., Gorski, M., Lottaz, C., Kronenberg, F., & Heid, I. M. (2015). EasyStrata: Evaluation and visualization of stratified genome-wide association meta-analysis data. *Bioinformatics*, *31*(2), 259–261. https://doi.org/10.1093/bioinformatics/btu621

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

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Gorski M, Günther F, Winkler TW, Weber BHF, Heid IM. On the differences between mega- and meta-imputation and analysis exemplified on the genetics of age-related macular degeneration. *Genet. Epidemiol.* 2019;43:559–576. https://doi.org/10.1002/gepi.22204