

RESEARCH ARTICLE

Extracting Primary Open-Angle Glaucoma from Electronic Medical Records for Genetic Association Studies

Nicole A. Restrepo¹, Eric Farber-Eger¹, Robert Goodloe¹, Jonathan L. Haines², Dana C. Crawford^{2*}

1 Center for Human Genetics Research, Vanderbilt University, Nashville, Tennessee, United States of America, **2** Department of Epidemiology & Biostatistics, Institute for Computational Biology, Case Western Reserve University, Cleveland, Ohio, United States of America

* dana.crawford@case.edu



OPEN ACCESS

Citation: Restrepo NA, Farber-Eger E, Goodloe R, Haines JL, Crawford DC (2015) Extracting Primary Open-Angle Glaucoma from Electronic Medical Records for Genetic Association Studies. PLoS ONE 10(6): e0127817. doi:10.1371/journal.pone.0127817

Academic Editor: Monica M. Jablonski, The University of Tennessee Health Science Center, UNITED STATES

Received: July 15, 2014

Accepted: April 20, 2015

Published: June 10, 2015

Copyright: © 2015 Restrepo et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The Database of Genotypes and Phenotypes (dbGaP) accession ID associated with the EAGLE BioVU dataset is PAGE-BioVu phs000559.

Funding: This work was supported by National Institutes of Health U01 HG004798 and its ARRA supplements. The dataset(s) used for the analyses described were obtained from Vanderbilt University Medical Center's BioVU which is supported by institutional funding and the National Center for Research Resources, Grant UL1 RR024975-01 (now at the National Center for Advancing Translational

Abstract

Electronic medical records (EMRs) are being widely implemented for use in genetic and genomic studies. As a phenotypic rich resource, EMRs provide researchers with the opportunity to identify disease cohorts and perform genotype-phenotype association studies. The Epidemiologic Architecture for Genes Linked to Environment (EAGLE) study, as part of the Population Architecture using Genomics and Epidemiology (PAGE) I study, has genotyped more than 15,000 individuals of diverse genetic ancestry in BioVU, the Vanderbilt University Medical Center's biorepository linked to a de-identified version of the EMR (EAGLE BioVU). Here we develop and deploy an algorithm utilizing data mining techniques to identify primary open-angle glaucoma (POAG) in African Americans from EAGLE BioVU for genetic association studies. The algorithm described here was designed using a combination of diagnostic codes, current procedural terminology billing codes, and free text searches to identify POAG status in situations where gold-standard digital photography cannot be accessed. The case algorithm identified 267 potential POAG subjects but underperformed after manual review with a positive predictive value of 51.6% and an accuracy of 76.3%. The control algorithm identified controls with a negative predictive value of 98.3%. Although the case algorithm requires more downstream manual review for use in large-scale studies, it provides a basis by which to extract a specific clinical subtype of glaucoma from EMRs in the absence of digital photographs.

Introduction

The use of electronic medical records (EMR) in biomedical research has gained traction in recent years, driven in part by the large-scale collection of biospecimens associated with EMRs. When linked to extensive biobanks, these large, data-dense resources can be utilized in large-scale genetic and genomic studies[1]. Compared to traditional epidemiological cohorts, which are expensive and can take years to decades to collect, EMRs are a relatively cost efficient

Sciences, Grant 2 UL1 TR000445-06). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: Co-author Dana C. Crawford is a PLOS ONE Editorial Board member. This membership does not alter the authors' adherence to all the PLOS ONE policies on sharing data and materials.

alternative with years of medical information immediately accessible. These medical records can be interrogated for a range of phenotypes often not available or included in epidemiological cohorts. Also, large sample sizes can be acquired from the development of phenotype algorithms standardized for use across multiple institutions[2,3].

One such EMR resource is the Vanderbilt University Medical Center (VUMC) Synthetic Derivative (SD). The SD is a de-identified version of Vanderbilt's EMR and contains inpatient and outpatient medical records collected at VUMC and affiliated clinics. Patient records consist of both structured (e.g., billing codes, procedure codes, laboratory values) and unstructured (e.g., clinical free text) data. To date, the Vanderbilt EMR contains over 2.2 million records with each record containing on average 6.5 years of medical history and an average of eight prescriptions. The SD is linked with VUMC's DNA repository known as BioVU[4]. These DNA samples are extracted from discarded blood samples collected from outpatient clinical laboratories. The SD in conjunction with BioVU has been used recently in several genetic association studies to identify genetic variants associated with a range of phenotypes such as multiple sclerosis[5], type 2 diabetes[6], cancer[7,8], electrocardiographic traits[9–11], pharmacogenomic-related outcomes[12–14], clinical quantitative traits[15–20], and hypothyroidism[21]. As part of the Population Architecture using Genomics and Epidemiology (PAGE) I study[22], we as the Epidemiologic Architecture for Genes Linked to Environment (EAGLE) study accessed ~15,000 DNA samples from non-European descent individuals in BioVU (EAGLE BioVU) to perform genetic association studies in diverse populations[23].

Here we introduce an algorithm to extract primary open-angle glaucoma (POAG) in EAGLE BioVU to identify cases and controls among African Americans ($n = 11,521$) for genetic association studies. POAG is a clinical subtype of glaucoma, a heterogeneous group of eye diseases characterized by chronic degeneration of the optic nerve and gradual vision-loss. As the second leading cause of blindness in the United States[24], glaucoma is a leading cause of vision disability. The most common form of this disease is POAG which disproportionately afflicts African Americans. The prevalence of POAG in African Americans is nearly double that observed in European-descent populations[25–27] with rates as high as 5.6% vs. 1.7%, respectively, in individuals over the age of 40 years[28].

A major challenge associated with use of the Vanderbilt EMR for POAG research is that the specialty eye clinic (the Vanderbilt Eye Institute or VEI) currently lacks an interface with structured data fields to digitally upload test results and forms into the broader EMR. Due to this lack of structured data fields and difficulty in de-identifying photographs, test results and digital photographs are not readily available for research use in the SD. Given these limitations, we sought to develop a phenotype algorithm using searchable and parsable elements available in the SD. With this goal in mind, we have 1) developed and implemented a data-mining algorithm to classify individuals as POAG cases or controls; and 2) manually verified the case/control status of individuals to evaluate the algorithm's performance. Overall, the algorithm identified 267 cases and 4813 controls among African Americans from EAGLE BioVU. After manual review we determined that the algorithm only had a positive predictive value of 51.6% to identify definite POAG cases with an accuracy of 76.3%. Controls were identified with a negative predictive value of 98.3%. Despite the limited positive predictive value of the algorithm in defining POAG case status, our strategy overall provides a starting point in the absence of digital photographs (i.e., the gold standard) to identify definitive cases of POAG from amongst over two million clinical records.

Results

Algorithm development

The full description of the POAG algorithm is available in **Methods**. Briefly, we developed an algorithm designed to identify POAG cases and controls from EAGLE BioVU[23], a subset of records and DNA samples from the larger Vanderbilt SD and BioVU, using a combination of International Classification of Diseases (ICD-9) diagnostic codes for glaucoma, Current Procedural Terminology (CPT) billing codes for ophthalmology/general clinic, and free text searches (Fig 1). Digital photographs taken over the course of multiple visits are the gold standard for diagnosing ocular diseases such as POAG; however these are not readily available in the SD for research as the VEI currently uses a manual/paper process for evaluating and recording the results of a patient's vision exams. Most of these forms are scanned and uploaded into the EMR as portable document format (PDF) forms. Because it is difficult to de-identify and parse PDF forms, this POAG case/control algorithm targeted structured and easily parsed unstructured data and was designed to maximize the number of individuals eligible for manual review to confirm case/control status.

Manual Review and algorithm performance

The SD records of all EAGLE BioVU POAG cases and a random sample of controls identified by the algorithm were manually reviewed by a single investigator to verify POAG case/control status. Positive predictive value (PPV), negative predictive value (NPV), and accuracy were then calculated to evaluate the algorithm's performance. Of the 267 individuals identified as cases by the algorithm, 138 were determined to be definite cases based on manual review of records retrieved from the SD. The records accessed for verification of case status were primarily composed of but not limited to surgical reports, optometry, and ophthalmology clinic notes. Other data that were taken into account were medication lists, general and specialty clinic reports, clinical communications, and problems lists. An individual was classified as a definite case if he/she met one of two criteria: 1) a written diagnosis by a Vanderbilt ophthalmologist/optometrist as pertained to a patient's exact clinical sub-type of glaucoma (An example of this is given in Fig 2) and 2) the patient's medical record contained *all* of the following: at least two independent mentions of the POAG ICD-9 code (365.11), a glaucoma medication (S1 Table), and a surgical procedure for treatment of POAG complications as identified by a surgical report. Surgical procedures for treatment of POAG may include argon laser trabeculoplasty, selective laser trabeculoplasty, MicroPulse laser trabeculoplasty, and Ex-Press Mini Glaucoma Shunt. Previous studies which successfully developed phenotype algorithms with high PPV for ocular traits found that CPT codes for surgical procedures were sufficient for positively identifying individuals with cataracts[29]. We classified individuals whose records were positive for glaucoma case status but lacked sufficient clinical records to determine the sub-type of glaucoma as "potential cases." The criteria for potential cases includes *all* of the following: at least one mention of POAG ICD-9 (365.11), an ophthalmology/fundus CPT code (i.e. 92012, 92014, or 92250), glaucoma medication, and a text mention of "POAG" or "glaucoma" in a clinic note or problems history. Of the records reviewed, we identified 67 potential cases. Sixty-two individuals were determined to be false positives. Upon examination of the records, 28 of the 62 false positives were found to be diagnosed with another clinical sub-type of glaucoma, such as uveitic (n = 6), chronic angle closure (n = 11), pediatric (n = 2), neovascular (n = 6), or glaucoma steroid responder (n = 3). Reviews of other false positives lacked sufficient records or communications to determine glaucoma/POAG status. Ambiguous notes included those that state that

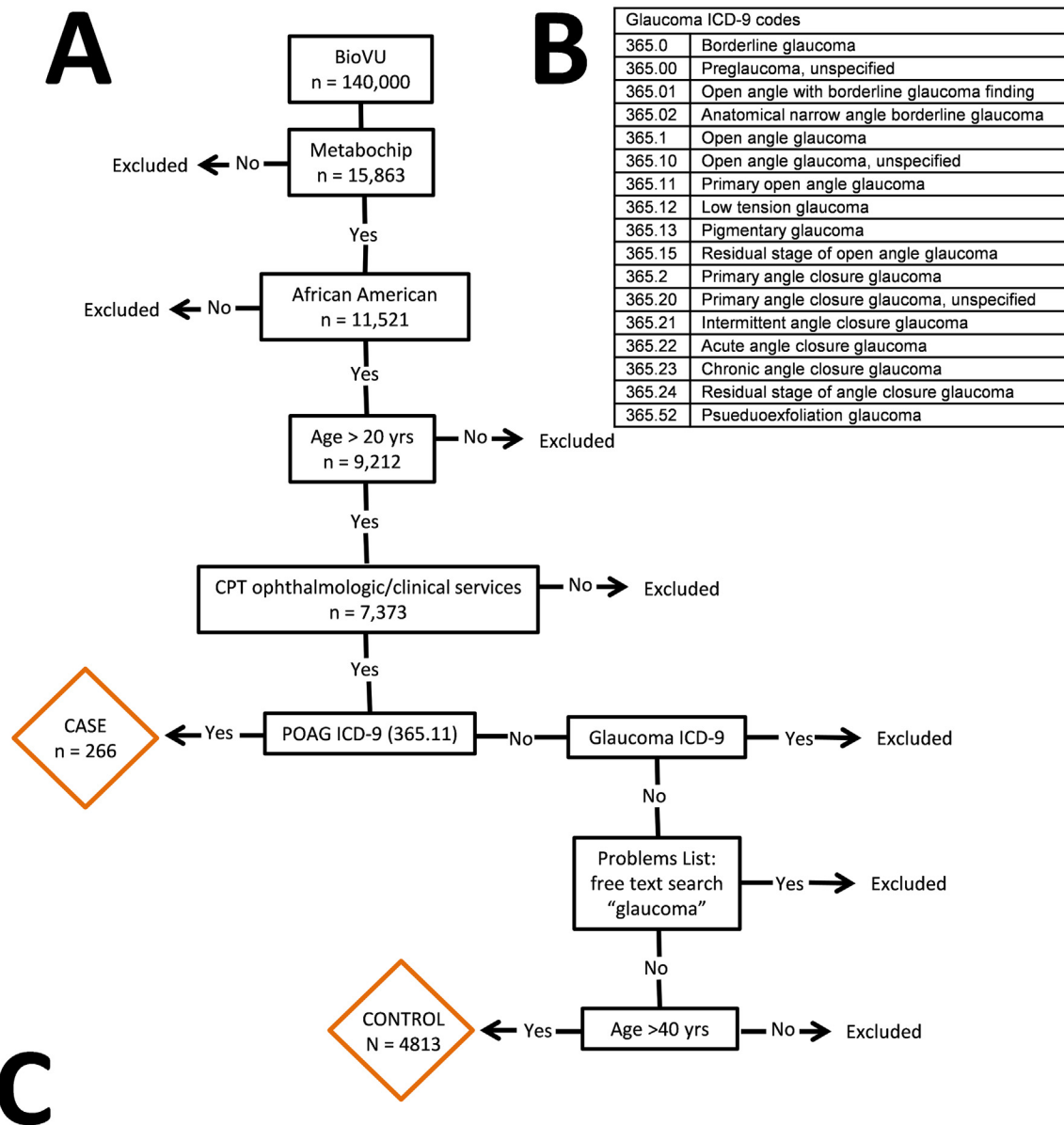


Fig 1. Criteria and decision tree used to classify individuals in EAGLE BioVU as POAG cases or controls. (A) Flow diagram of phenotype algorithm for POAG cases and controls. (B) List of glaucoma ICD-9 codes (C) List of CPT codes for ophthalmology and general clinic procedures.

doi:10.1371/journal.pone.0127817.g001

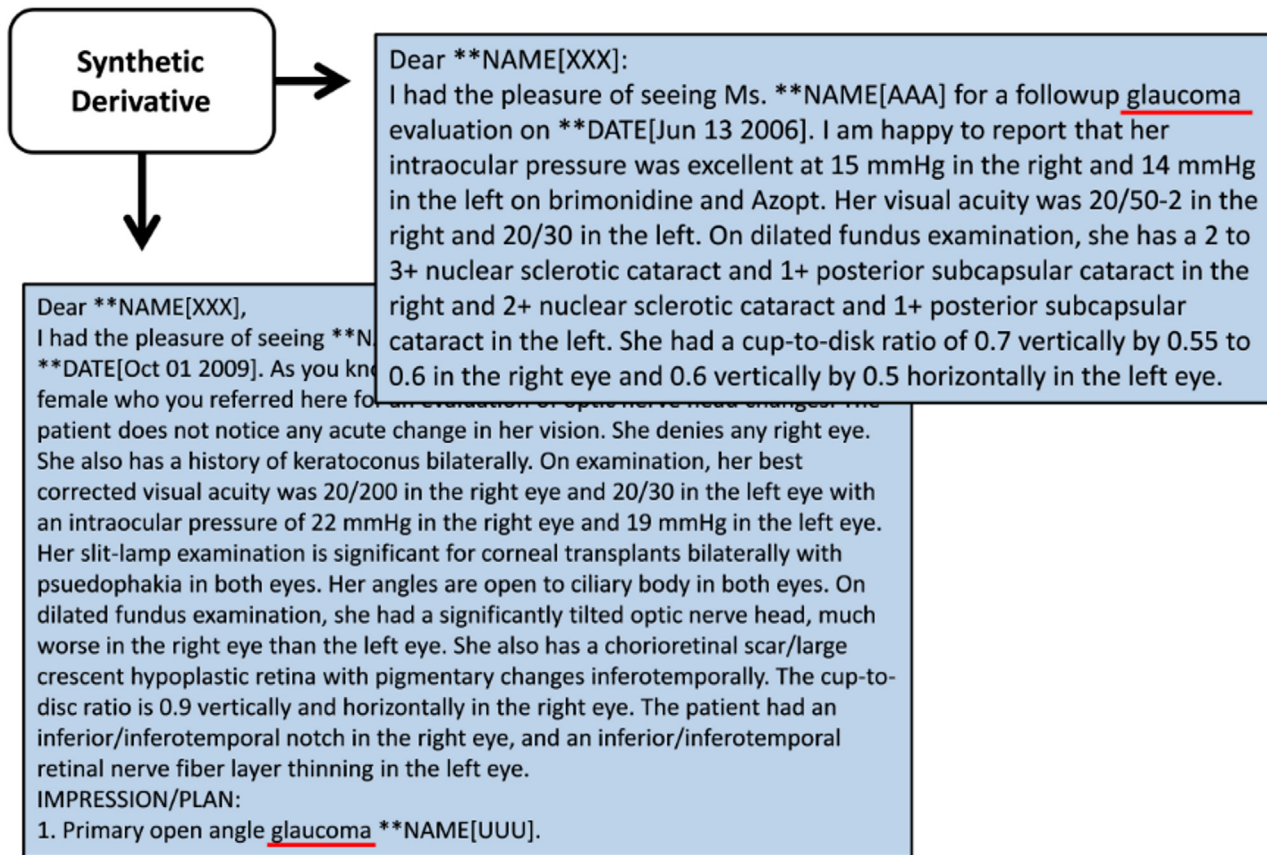


Fig 2. De-identified clinic notes extracted from the Synthetic Derivative. De-identified letters from the Vanderbilt Eye Institute’s ophthalmologists and optometrists extracted from the Synthetic Derivative. These letters represent the primary type of record used to verify POAG case status in EAGLE BioVU. It is an example of a definite case which has the specific clinical sub-type of glaucoma clearly stated and of a potential case which only includes a more general glaucoma diagnosis statement.

doi:10.1371/journal.pone.0127817.g002

a patient had “advanced glaucoma,” “ocular hypertension,” or was a “glaucoma suspect.” These individuals were excluded from case status as we could not reasonably verify a diagnosis.

For the POAG case algorithm, we determined that the PPV for definite cases was 51.6% with an accuracy of 76.3% (Table 1). When including potential cases, PPV was 76.7% with an accuracy of 83.1% (Table 1). We manually reviewed the SD medical records of 300 randomly selected controls identified by the algorithm to calculate NPV. Of the 300 individuals identified as controls with this algorithm (Fig 1), five showed evidence of glaucoma at the time of review

Table 1. Evaluation of primary open-angle glaucoma phenotype algorithm in African Americans from EAGLE BioVU.

	Sample Size	Manually reviewed	PPV	NPV	Accuracy
Cases	267	267	-	-	-
-Definite		138	51.6%	-	76.3%
-Potential		67	76.7%	-	83.1%
Controls	4813	300	-	98.3%	-

Definite cases were individuals whose POAG status could be determined with high likelihood. Potential cases were individuals whose medical records lacked sufficient information to make a definitive decision. Potential case results were calculated by including both potential and definite case numbers.

doi:10.1371/journal.pone.0127817.t001

or else had mention of glaucoma in their records. As an additional review function, we performed a free-text search in all available documents for mention of the following words and abbreviations: glaucoma, fundus, oph, ophth, and vision. The shorthand identifiers such as fundus, oph, ophth [i.e., funduscopy eye exam, ophthalmology clinic, and oph (common abbreviation for ophthalmology)] are used throughout the clinical records especially within communications across clinics. This function was performed to ensure that algorithm-identified controls without qualifying ICD-9 or CPT codes also did not have evidence of POAG in the free clinical text. Performance of the POAG control algorithm was found to have a NPV of 98.3% (Table 1).

Cup-to-disc ratio (CDR)

As mentioned, ocular phenotypes are difficult to extract from the VUMC EMR due to the lack of structured data fields with the VEI EMR interface. Extracting quantitative phenotypes such as cup-to-disc ratios, the ratio of optic cup diameter to optic disc diameter, proved particularly challenging as there is not a uniform field or manner in which ophthalmologists report these data. We have found, however, that these data can be extracted via expression matching from referral letters and clinic notes sent between ophthalmologists and clinicians in the process of patient care (Fig 2). The following key word phrases were included in the search in reference to the right eye (O.D.) and left eye (O.S.): CDR, cup-to-disc ratio, cup-to-disk, and cup-to-disk ratio. Upon manual review of the results it became apparent that a more sophisticated approach was necessary to differentiate whether the numbers recorded were for the horizontal optic cup ratio, the vertical optic cup ratio, or the quotient of the horizontal-to-vertical CDR which ophthalmologists report interchangeably. When multiple measures of CDR were available for a patient, the most recent measurement was taken.

A total of 132 POAG case records were positive for a CDR key word phrase, and all were manually reviewed for CDR. Only 9 of the 132 records (7%) were missing a measure for CDR. The median values for CDR in this study are 0.7 (SD 0.22) in the right eyes and 0.7 (SD 0.23) in the left eyes (Fig 3).

POAG study characteristics for African Americans

Of the 11,521 African Americans, 9,441 were over the age of 20 years and considered for downstream genetic association studies of POAG. Of these adults, 267 were identified as POAG

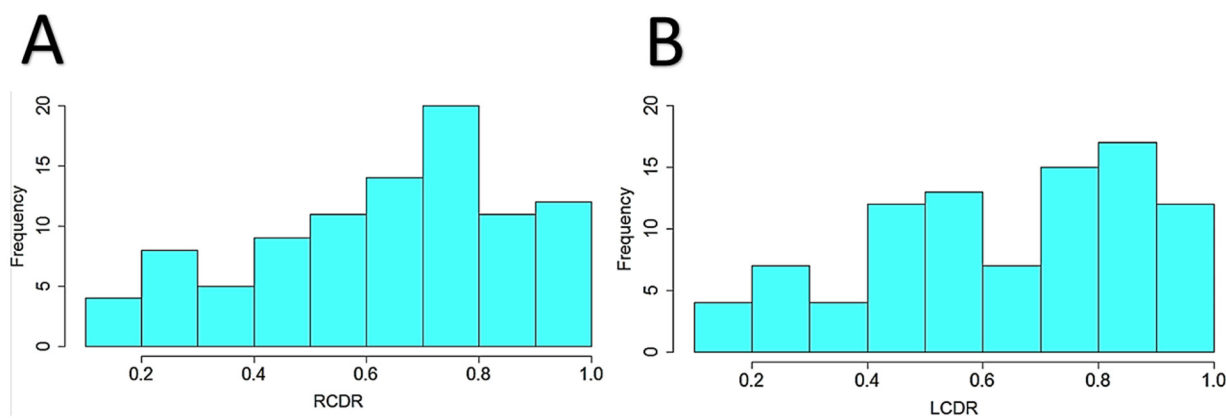


Fig 3. Histogram plots of the distribution of cup-to-disk (CDR) ratios of POAG cases from the Synthetic Derivative. Distribution of CDR for the right eyes (A) and the left eyes (B) of POAG cases from numbers that were manually extracted from the SD.

doi:10.1371/journal.pone.0127817.g003

Table 2. Study population characteristics of POAG definite cases and controls among African Americans in EAGLE BioVU.

	Definite Cases > 20 yrs (SD)	Controls > 40 yrs (SD)
N	138	4813
Age at Diagnosis (years)	62.0 (12.0)	–
Age at Last Clinic (years)	–	54 (11.7)
Sex (% female)	63.7	60
Hypertensive (%)	55.1	46.6
BMI (kg/m ²)	30.1 (6.7)	30.1 (8.0)
Diastolic (mm/Hg)	74.5 (8.1)	80 (33.6)
Systolic (mm/Hg)	134.5 (14.1)	124 (26.2)
Cholesterol (mg/dL)	183 (40.6)	161 (65.2)
HDL (mg/dL)	52.5 (25.0)	53 (38.6)
LDL (mg/dL)	103 (42.9)	99 (50.7)
Triglycerides (mg/dL)	125 (76.3)	98 (67.8)

Median values were calculated for the following: Age at POAG diagnosis was determined by the date of when POAG ICD-9 (365.11) was first mentioned in the records. Age at last clinic visit (LCV) was taken as the date of the last CPT mentioned in the records for controls. An individual was classified as hypertensive if he/she met one of three criteria: systolic blood pressure > 140 mm/Hg, diastolic blood pressure > 90 mm/Hg, or on hypertension medications all within a two year window of when they were diagnosed with POAG in cases and a two year window of their LCV date for controls. Blood pressure (systolic and diastolic), lipids (total cholesterol, high-density cholesterol, low-density cholesterol, and triglycerides), and body mass index (height and weight) were calculated from labs or measurements within two years of POAG diagnosis or LCV. Abbreviations: standard deviation (SD)

doi:10.1371/journal.pone.0127817.t002

cases (2.82%) and 4,813 as POAG controls. As might be expected for an age-related ocular disease, the median age of cases was older than controls (62 versus 54 years; [Table 2](#)). More than half of the cases and controls were female; and approximately half were hypertensive. On average, both cases and controls were obese (median body mass index > 30.0 kg/m²).

Preliminary genetic association analysis results for *CDKN2B-AS1* region

In previous genetic association studies, the *CDKN2B-AS1* gene region has been associated with glaucoma clinical sub-types and quantitative glaucoma traits, such as CDR, POAG [30], and normal tension glaucoma in populations of European and Asian descent. *CDKN2B-AS1* was targeted for fine-mapping by the Illumina Metabochip, with 459 SNPs assayed in this region. In this preliminary analysis of African American adults from EAGLE BioVU after quality control, 135 definite POAG cases and 1,376 POAG controls ([S2 Table](#)) were tested for an association with 258 common variants (MAF > 5%) in the *CDKN2B-AS1* region using a logistic regression model adjusted for age and sex. Although some SNPs were nominally associated at $p < 0.05$ ([S1 Fig](#)), none passed a strict Bonferroni correction ($p < 0.0001$).

Discussion

We developed a phenotype algorithm to extract POAG case and control status using a combination of structured and unstructured data available for research in the Vanderbilt de-identified EMR. The algorithm was implemented for 11,521 African Americans in EAGLE BioVU. Of those patients over the age of twenty ($n = 9,441$), we identified 267 cases of POAG and 4,813 controls without POAG by use of phenotype algorithms. Upon manual review of all POAG cases identified by the algorithm, 138 were classified as definite and 67 as potential, with the remaining 62 classified as false positives due to misclassification or missing data ([S3 Table](#)). Manual review of 300 controls confirmed control status among individuals identified by the algorithm. Evaluation of the algorithm revealed high performance for POAG control

status (NPV = 98.3%) but low performance for POAG case status (PPV = 51.6% and accuracy = 76.3%).

Despite the low PPV of this POAG algorithm, we are encouraged that the algorithm has modest accuracy. This POAG algorithm was designed for an EMR environment lacking digital photographs for review in a research setting. We intentionally designed the algorithm to assist us in identifying all possible true POAG cases given that manual review of the EAGLE BioVU's almost 10,000 African American adults would be prohibitively time and resource consuming. An alternative strategy to triaging records for manual review is the requirement of a single ICD-9 code for POAG. Based on these relaxed criteria, we identified 309 African Americans in EAGLE BioVU whose EMR contained an ICD-9 code for POAG. After review of these additional individuals for definite POAG case status, we determined that all measures of performance were lower (PPV = 47.2% and accuracy = 72.4%) compared with the algorithm developed here.

As evident by this study, broad phenotyping of a disease cohort is well within the capabilities of an EMR, even one with limited access to clinical data, through the use of simple text-mining techniques incorporating pattern matching and structured data from the EMR. The algorithm designed here for POAG is stringent. The PPV (51.6%) of this POAG algorithm for definite case status is well below the threshold of 95% adopted by consortia such as the electronic Medical Records & Genomics (eMERGE) network[31] for use in large-scale genetic association studies[3,29]. However, the addition of potential cases substantially increases the PPV to 76.7%. If we had designed the algorithm to merely detect an individual with a "general" glaucoma classification, the PPV increased to 87.2%. Furthermore, we have developed a highly discriminatory algorithm (NPV 98.3%) that can identify ocular controls. In the development of phenotype algorithms there is the constant tradeoff between overly strict or vague criteria resulting in loss of cases or misclassification of subjects, both of which will lead to a loss in statistical power in downstream genetic association studies.

Although most clinics within a health care organization that maintains an EMR adopt the digital system, occasionally a clinic may be excluded from converting to an all-digital interface or choose to opt out. The reasons for these exclusions may vary, but the consequences of missing data ultimately impact studies. Evaluation of the algorithm was limited by the lack of available funduscopy images (i.e., the "gold standard") for validation, a limitation of this study. Missing data limited our ability to verify all POAG cases initially identified. Several cases identified here could only be classified as potential cases given the lack of sufficient clinical records to determine the sub-type of glaucoma. This limitation underscores the need for better implementation of EMRs across healthcare organizations for use in biomedical research.

Missing data can also introduce misclassification bias into studies. The control algorithm developed here was designed around the concept that an individual is free of POAG. However, without a complete medical work-up, there is the potential that a control is an undiagnosed case. Misclassification and potential ascertainment bias is also possible in case identification where cases are only those individuals who have been evaluated by a specialist and are therefore potentially extreme or overly symptomatic cases. A well-known barrier for individuals seeking medical attention is low socioeconomic status which disproportionately affects African Americans[32]. Additionally, at least one study suggests that the lifetime prevalence of undiagnosed glaucoma in African Americans is 50%[33]. This may in part explain the limited number of African American POAG cases ($n = 267$) representing only 2.82% of African Americans in EAGLE BioVU. Given the expected prevalence of POAG at 4–5% among African Americans [27,34], presumably the cases in EAGLE BioVU are only being diagnosed once vision loss becomes severe.

The age and gender compositions of the EAGLE BioVU definite POAG cases (Table 2) and controls differ when compared with other clinical and epidemiologic cohorts. More than half of EAGLE BioVU definite POAG cases (63.7%) and controls (60%) are female. In contrast, the International Consortium of African Ancestry Research in Glaucoma (ICAARE-Glaucoma; n = 2,150) study identified fewer female cases and controls[35]: African American cases (49.5%), African American controls (54.6%), Ghanaian cases (43.9%), and Ghanaian controls (57.3%). Also, the mean age at diagnosis for ICAARE-Glaucoma African American cases (57.0 years) is younger than EAGLE BioVU POAG cases (62.0 years) while the ICAARE-Glaucoma control group (59.4 years) is older compared with EAGLE BioVU POAG controls (54 years). These demographic differences are likely the result of differences in recruitment or ascertainment of African Americans within their respective communities, which can potentially introduce heterogeneity in downstream studies.

Analysis of *CDKN2B-AS1* in African Americans

This study did not find a significant association in the *CDKN2B-AS1* region for POAG in African Americans. These results are not entirely surprising given the differences in susceptibility and linkage disequilibrium between European and African-descent populations (S2 and S3 Figs). Most of the significant associations between this gene and POAG were discovered in European and Japanese populations (Table 3) [30,36–38]. Of these SNPs, two were available on the MetaboChip but were not found to be associated with POAG in this study. Our results are somewhat consistent with Liu et al [35] who investigated known European POAG risk loci in both an African American and Ghanaian population. Liu et al failed to replicate many of the European risk loci in their African populations, but they did identify one SNP (rs10120688) in *CDKN2B-AS1* significantly associated with POAG. This SNP is not available on the MetaboChip and as such we could not replicate this association in our dataset. *CDKN2B-AS1* rs1333049 was available for analysis in this study, and interestingly, our results (OR = 0.72; p = 0.06) were similar to that reported Liu et al for the same SNP (OR = 0.89; p = 0.07) [35].

Limitations and Conclusions

Our study has a number of limitations. The algorithm was developed at VUMC under the restrictions of a de-identified EMR, which limits access to certain data types, and the SD, which does not currently contain all pertinent medical records. As mentioned before the VEI does not utilize an electronic interface with structured data fields. Lack of these structured data fields hinders the extraction of ophthalmology exam results from the SD. Lack of exam results may prevent researchers from definitively ascertaining an individual’s ocular disease status. Due to

Table 3. Published index variants for the *CDKN2B-AS1* region associated with POAG or POAG associated trait and availability of these variants on the MetaboChip.

rs#	population	OR	p-value	Discovery study	Current Study	OR	p-value
rs7865618	Japanese	1.78	9x10 ⁻¹¹	Nakano et al[36]	not available	-	-
rs1063192	Japanese	1.33	5x10 ⁻¹⁰	Osman et al[37]	-	0.92	0.75
rs2157719	European American	1.45	2x10 ⁻¹⁸	Wiggs et al[38]	-	0.97	0.92
rs4977756	European	1.50	4.7x10 ⁻⁹	Burdon et al[30]	not available	-	-
rs10120688	African American	1.21	0.002	Liu et al[35]	not available	-	-

Shown are significant index variants which are listed on the NHGRI GWAS catalog and within PubMed. Included is the availability of the index variants on the MetaboChip and summary results for the current studies association analysis of African Americans with POAG in the *CDKN2B-AS1* region.

doi:10.1371/journal.pone.0127817.t003

these limitations it is unclear if our algorithm can be exported for use in the EMRs of other medical institutions. Additionally, one investigator performed the manual review preventing us from assessing intra- or inter-grader variability. This investigator was aware of the algorithm's determination of case and control status, which may have introduced bias. However, we modeled our algorithm development process after the eMERGE Phenotype Working Group workflow[2,39]. In the eMERGE Network, algorithm development and assessment are an iterative process. That is, content experts design the initial algorithm and deploy it. After a round of manual reviews and performance calculations, the algorithm is altered, re-deployed, and re-evaluated. There is potential bias in reviewing and assessing the performance of the algorithm when adopting the eMERGE workflow.

Despite the limitations in portability and data access, we were able to define primary open-angle glaucoma from the VUMC's SD. We have a diverse population connected to a depth of medical data, even if not all of it is easily searchable. Our study has made available more case counts for African Americans with ocular disease. And, BioVU continues to accrue samples as well as update the medical records associated with samples already collected; therefore, the accrual of additional cases is anticipated. The ability to extract ocular phenotypes from EMRs will provide researchers with previously unaccessed datasets to further advances in ocular genetic research and vision-loss prevention.

Methods

Ethics Statement

BioVU study subjects are not consented. DNA is collected from discarded blood samples remaining after routine clinical testing and is linked to de-identified medical records. According to the Vanderbilt Institutional Review Board (IRB) and the Federal Office of Human Research Protections provisions, the Vanderbilt protocol is considered nonhuman subjects research (The Code of Federal Regulations, 45 CFR 46.102 (f)). The IRB at Vanderbilt University approved this research.

Population

The Epidemiologic Architecture for Genes Linked to Environment (EAGLE) study, as part of the Population Architecture using Genomics and Epidemiology (PAGE) I study[22], accesses clinical and epidemiological collections with racially/ethnically diverse populations. These collections are used to perform and generalize genetic association studies for common human diseases, including common ocular diseases such as POAG. As part of PAGE I, EAGLE genotyped all DNA samples in BioVU from non-European-descent individuals as of 2011 (EAGLE BioVU; $n = 15,863$; Table 2)[23]. All DNA samples were genotyped on the Illumina MetaboChip by the Vanderbilt University Center for Human Genetics Research DNA Resources Core. The MetaboChip is a custom array designed for replication and fine mapping of genome-wide association study (GWAS)-identified variants for metabolic and cardiovascular traits [40]. The data described here will be available through the database of Genotypes and Phenotypes (dbGaP).

Development of Phenotype Algorithm

Initial screening criteria for study population. Individuals included for this POAG study were African American adults over the age of 20 years as of March 20, 2013. We excluded pediatric glaucoma cases because it is a separate condition caused by developmental issues prior to birth and/or by a very rare genetic mutation. The genetics of pediatric glaucoma are likely to be

fundamentally different from the genetics of glaucoma in older adults, and this genetic heterogeneity would result in lower power to detect genetic associations for adult POAG. The final inclusion criteria required that an individual's medical records include either one mention of a CPT code for ophthalmology or a CPT code for general clinic procedures (Fig 1).

Primary open-angle glaucoma cases. POAG cases were identified as individuals with at least one ICD-9 code for POAG (365.11). Cases were not excluded if they contained an additional ICD-9 code for another sub-type of glaucoma (list of glaucoma ICD-9 codes included in Fig 1). It is not uncommon for patients to experience different sub-types of glaucoma bilaterally or within the same eye which is also known as mixed- or combined- mechanism glaucoma. Also, over the course of a lifetime a patient may progress from one form to another such as closed-angle glaucoma to open-angle glaucoma.

Primary open-angle glaucoma controls. Controls were defined as individuals whose records were devoid of any glaucoma ICD-9 code. If the algorithm identified "glaucoma," "glaucome," "glocoma," "gloucoma," "gluacoma," "glucoma," or "glycoma" in a free text search of problems lists and clinical notes, the individual was excluded. We excluded individuals under the age of 40 years, as calculated from a given birth-date, to reduce contamination of controls with future cases.

Calculation of PPV, NPV, Sensitivity, Specificity, and Accuracy. The performance of the case and control algorithms were calculated as follows: PPV was calculated as the ratio of true positives (TP) over TP + false positive (FP) [$PPV = TP / (TP + FP)$]. A TP was an individual who was identified by the case algorithm as a case and was then confirmed by manual review to be a true case. A FP in turn was an individual identified as a case that was determined not to be a case during manual review. NPV is the ratio of true negative (TN) (i.e., a control who was confirmed as a control) over TN and FN [$NPV = TN / (TN + FN)$]. Lastly, Accuracy was calculated as the ratio of the sum of TP and TN over the sum of all positives and all negatives [$Accuracy = (TP + TN) / (Positives (TP + FP) + Negatives (TN + FN))$].

Extraction and calculation of individual demographic elements. Demographic data and laboratory measurements were extracted and calculated with an emphasis for use in future case/control studies. For cases, age at POAG diagnosis was determined by the date in the records for the first mention of a POAG ICD-9 (365.11). For controls, age at last clinic visit (LCV) was taken as the date of the last CPT mentioned in the records. An individual, regardless of case/control status, was classified as a hypertensive if he/she met one of three criteria: 1) systolic blood pressure > 140 mm/Hg, 2) diastolic blood pressure > 90 mm/Hg, or 3) mention of a hypertension medication within a two-year window of when an individual was diagnosed with POAG for cases or within a two-year window of his/her LCV date for controls. Median values were calculated for the following laboratory measurements within a two-year window of an individual's POAG diagnosis or LCV for controls: blood pressure (systolic and diastolic), lipids (total cholesterol, high-density cholesterol, low-density cholesterol, and triglycerides), and body mass index (height and weight).

Genetic association analyses

We tested for an association between POAG and 258 SNPs in the *CDKN2B-AS1* region of chromosome 14 in African Americans. Individuals included in this analysis were those identified as "definite" POAG cases regardless of age and POAG controls over the age of 60 years. Individuals passed quality control protocol for cryptic relatedness (based on identify-by-descent). Initially 459 SNPs were available for study. Each SNP considered for analysis passed quality control standards for HWE ($p < 0.001$), genotyping efficiency (>95%), and minor allele frequency (>5%). After quality control, 258 SNPs were available for analyses. Analyses were

conducted using PLINKv1.07[41]. Each SNP was tested for an association using logistic regression assuming a log-additive genetic model adjusted by age and sex.

Supporting Information

S1 Table. List of glaucoma medications used in validation of primary open-angle glaucoma cases. Indications is a limited list of FDA approved uses as stated on the Drugs@FDA website as of October 4th, 2014: <http://www.accessdata.fda.gov/scripts/cder/drugsatfda/>. (DOCX)

S2 Table. Study population characteristics of POAG definite cases and controls over 60 years among African Americans in EAGLE BioVU. Median values were calculated for the following: Age at POAG diagnosis was determined by the date of when POAG ICD-9 (365.11) was first mentioned in the records. Age at last clinic visit (LCV) was taken as the date of the last CPT mentioned in the records for controls. An individual was classified as hypertensive if he/she met one of three criteria: systolic blood pressure > 140 mm/Hg, diastolic blood pressure > 90 mm/Hg, or on hypertension medications all within a two year window of when they were diagnosed with POAG in cases and a two year window of their LCV date for controls. Blood pressure (systolic and diastolic), lipids (total cholesterol, high-density cholesterol, low-density cholesterol, and triglycerides), and body mass index (height and weight) were calculated from labs or measurements within two years of POAG diagnosis or LCV. Abbreviations: standard deviation (SD) (DOCX)

S3 Table. Study population characteristics of total POAG cases and controls among African Americans in EAGLE BioVU. Median values were calculated for the following: Age was calculated from a given birth year. Age at POAG diagnosis was determined by the date of when POAG ICD-9 (365.11) was first mentioned in the records. Age at last clinic visit (LCV) was taken as the date of the last CPT mentioned in the records for controls. An individual was classified as hypertensive if he/she met one of three criteria: systolic blood pressure > 140 mm/Hg, diastolic blood pressure > 90 mm/Hg, or on hypertension medications all within a two year window of when they were diagnosed with POAG in cases and a two year window of their LCV date for controls. Blood pressure (systolic and diastolic), lipids (total cholesterol, high-density cholesterol, low-density cholesterol, and triglycerides), and body mass index (height and weight) were calculated from labs or measurements within two years of POAG diagnosis or LCV. Abbreviations: standard deviation (SD) (DOCX)

S1 Fig. Locus zoom plot of African American POAG association results for *CDKN2B-AS1*. Figure was generated using LocusZoom plot (<http://csg.sph.umich.edu/locuszoom/>) with no linkage disequilibrium calculations. Results are shown for the African American POAG association results in the *CDKN2B-AS1* region for the model adjusted by age and sex. (TIF)

S2 Fig. Locus Zoom plot of African American POAG association results for *CDKN2B-AS1* with LD calculations from 1000 Genomes YRI. Figure was generated using LocusZoom plot (<http://csg.sph.umich.edu/locuszoom/>) with linkage disequilibrium calculations from the hg18 1000 Genomes June 2010 YRI dataset. Results are shown for the African American POAG association results in the *CDKN2B-AS1* region for the model adjusted by age and sex. (TIF)

S3 Fig. Locus Zoom plot of African American POAG association results for *CDKN2B-AS1* with LD calculations from 1000 Genomes CEU. Figure was generated using LocusZoom plot (<http://csg.sph.umich.edu/locuszoom/>) with linkage disequilibrium calculations from the hg18 1000 Genomes June 2010 CEU dataset. Results are shown for the African American POAG association results in the *CDKN2B-AS1* region for the model adjusted by age and sex. (TIF)

Acknowledgments

The Vanderbilt University Center for Human Genetics Research, Computational Genomics Core provided computational and/or analytical support for this work.

Author Contributions

Conceived and designed the experiments: NAR DCC JLH. Performed the experiments: NAR EFE RG. Analyzed the data: NAR. Wrote the paper: NAR DCC JLH.

References

1. Kohane IS. Using electronic health records to drive discovery in disease genomics. *Nat Rev Genet.* 2011; 12: 417–428. doi: [10.1038/nrg2999](https://doi.org/10.1038/nrg2999) PMID: [21587298](https://pubmed.ncbi.nlm.nih.gov/21587298/)
2. Newton KM, Peissig PL, Kho AN, Bielinski SJ, Berg RL, Choudhary V, et al. Validation of electronic medical record-based phenotyping algorithms: results and lessons learned from the eMERGE network. *Journal of the American Medical Informatics Association.* 2013; 20: e147–e154. doi: [10.1136/amiajnl-2012-000896](https://doi.org/10.1136/amiajnl-2012-000896) PMID: [23531748](https://pubmed.ncbi.nlm.nih.gov/23531748/)
3. Kho AN, Hayes MG, Rasmussen-Torvik L, Pacheco JA, Thompson WK, Armstrong LL, et al. Use of diverse electronic medical record systems to identify genetic risk for type 2 diabetes within a genome-wide association study. *Journal of the American Medical Informatics Association.* 2012; 19: 212–218. doi: [10.1136/amiajnl-2011-000439](https://doi.org/10.1136/amiajnl-2011-000439) PMID: [22101970](https://pubmed.ncbi.nlm.nih.gov/22101970/)
4. Roden DM, Pulley JM, Basford MA, Bernard GR, Clayton EW, Balsler JR, et al. Development of a Large-Scale De-Identified DNA Biobank to Enable Personalized Medicine. *Clin Pharmacol Ther.* 2008; 84: 362–369. doi: [10.1038/clpt.2008.89](https://doi.org/10.1038/clpt.2008.89) PMID: [18500243](https://pubmed.ncbi.nlm.nih.gov/18500243/)
5. Ritchie MD, Denny JC, Crawford DC, Ramirez AH, Weiner JB, Pulley JM, et al. Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. *Am J Hum Genet.* 2010; 86: 560–572. doi: [10.1016/j.ajhg.2010.03.003](https://doi.org/10.1016/j.ajhg.2010.03.003) PMID: [20362271](https://pubmed.ncbi.nlm.nih.gov/20362271/)
6. Ng MCY, Shriner D, Chen BH, Li J, Chen WM, Guo X, et al. Meta-Analysis of Genome-Wide Association Studies in African Americans Provides Insights into the Genetic Architecture of Type 2 Diabetes. *PLoS Genet.* 2014; 10: e1004517. doi: [10.1371/journal.pgen.1004517](https://doi.org/10.1371/journal.pgen.1004517) PMID: [25102180](https://pubmed.ncbi.nlm.nih.gov/25102180/)
7. Kocarnik JM, Park SL, Han J, Dumitrescu L, Cheng I, Wilkens LR, et al. Replication of Associations between GWAS SNPs and Melanoma Risk in the Population Architecture Using Genomics and Epidemiology (PAGE) Study. *J Invest Dermatol.* 2014; 134: 2049–2052. doi: [10.1038/jid.2014.53](https://doi.org/10.1038/jid.2014.53) PMID: [24480881](https://pubmed.ncbi.nlm.nih.gov/24480881/)
8. Cheng I, Kocarnik JM, Dumitrescu L, Lindor N, Chang-Claude J, Avery C et al (2013) Pleiotropic effects of genetic risk variants for other cancers on colorectal cancer risk: PAGE, GECCO, and CCFR Consortia.
9. Jeff JM, Ritchie MD, Denny JC, Kho AN, Ramirez AH, Crosslin D, et al. Generalization of variants identified by genome-wide association studies for electrocardiographic traits in African Americans. *Annals of Human Genetics.* 2013; 77: 321–332. doi: [10.1111/ahg.12023](https://doi.org/10.1111/ahg.12023) PMID: [23534349](https://pubmed.ncbi.nlm.nih.gov/23534349/)
10. Denny JC, Ritchie MD, Crawford DC, Schildcrout JS, Ramirez AH, Pulley JM, et al. Identification of Genomic Predictors of Atrioventricular Conduction. *Circulation.* 2010; 122: 2016–2021. doi: [10.1161/CIRCULATIONAHA.110.948828](https://doi.org/10.1161/CIRCULATIONAHA.110.948828) PMID: [21041692](https://pubmed.ncbi.nlm.nih.gov/21041692/)
11. Ritchie MD, Denny JC, Zuvich RL, Crawford DC, Schildcrout JS, Bastarache L, et al. Genome- and Phenome-Wide Analyses of Cardiac Conduction Identifies Markers of Arrhythmia Risk. *Circulation.* 2013; 127: 1377–1385. doi: [10.1161/CIRCULATIONAHA.112.000604](https://doi.org/10.1161/CIRCULATIONAHA.112.000604) PMID: [23463857](https://pubmed.ncbi.nlm.nih.gov/23463857/)
12. Oetjens M, Bush WS, Birdwell KA, Dilks HH, Bowton EA, Denny JC, et al. Utilization of an EMR-biorepository to identify the genetic predictors of calcineurin-inhibitor toxicity in heart transplant recipients. *Pac Symp Biocomput.* 2014; 2014: 253–264. PMID: [24297552](https://pubmed.ncbi.nlm.nih.gov/24297552/)

13. Ramirez AH, Shi Y, Schildcrout JS, Delaney JT, Xu H, Oetjens MT, et al. Predicting warfarin dosage in European and African Americans using DNA samples linked to an electronic health record. *Pharmacogenomics*. 2012; 13: 407–418.
14. Delaney JT, Ramirez AH, Bowton E, Pulley JM, Basford MA, Schildcrout JS, et al. Predicting Clopidogrel Response Using DNA Samples Linked to an Electronic Health Record. *Clin Pharmacol Ther*. 2012; 91: 257–263. doi: [10.1038/clpt.2011.221](https://doi.org/10.1038/clpt.2011.221) PMID: [22190063](https://pubmed.ncbi.nlm.nih.gov/22190063/)
15. Crosslin D, McDavid A, Weston N, Nelson S, Zheng X, Hart E, et al. Genetic variants associated with the white blood cell count in 13,923 subjects in the eMERGE Network. *Human Genetics*. 2012; 131: 639–652. doi: [10.1007/s00439-011-1103-9](https://doi.org/10.1007/s00439-011-1103-9) PMID: [22037903](https://pubmed.ncbi.nlm.nih.gov/22037903/)
16. Rasmussen-Torvik LJ, Pacheco JA, Wilke RA, Thompson WK, Ritchie MD, Kho AN, et al. High Density GWAS for LDL Cholesterol in African Americans Using Electronic Medical Records Reveals a Strong Protective Variant in APOE. *Clinical and Translational Science*. 2012; 5: 394–399. doi: [10.1111/j.1752-8062.2012.00446.x](https://doi.org/10.1111/j.1752-8062.2012.00446.x) PMID: [23067351](https://pubmed.ncbi.nlm.nih.gov/23067351/)
17. Crosslin DR, McDavid A, Weston N, Zheng X, Hart E, de Andrade M, et al. Genetic variation associated with circulating monocyte count in the eMERGE Network. *Human Molecular Genetics*. 2013; 22: 2119–2127. doi: [10.1093/hmg/ddt010](https://doi.org/10.1093/hmg/ddt010) PMID: [23314186](https://pubmed.ncbi.nlm.nih.gov/23314186/)
18. Ding K, de Andrade M, Manolio TA, Crawford DC, Rasmussen-Torvik LJ, Ritchie MD, et al. Genetic Variants That Confer Resistance to Malaria Are Associated with Red Blood Cell Traits in African-Americans: An Electronic Medical Record-based Genome-Wide Association Study. *G3: Genes|Genomes|Genetics*. 2013; 3: 1061–1068. doi: [10.1534/g3.113.006452](https://doi.org/10.1534/g3.113.006452) PMID: [23696099](https://pubmed.ncbi.nlm.nih.gov/23696099/)
19. Gong J, Schumacher F, Lim U, Hindorf L, Haessler J, Buyske S, et al. Fine Mapping and Identification of BMI Loci in African Americans. *The American Journal of Human Genetics*. 2013; 93: 661–671. doi: [10.1016/j.ajhg.2013.08.012](https://doi.org/10.1016/j.ajhg.2013.08.012) PMID: [24094743](https://pubmed.ncbi.nlm.nih.gov/24094743/)
20. Malinowski JR, Denny JC, Bielinski SJ, Basford MA, Bradford Y, Peissig PL, et al. Genetic Variants Associated with Serum Thyroid Stimulating Hormone (TSH) Levels in European Americans and African Americans from the eMERGE Network. *PLoS ONE*. 2014; 9: e111301. doi: [10.1371/journal.pone.0111301](https://doi.org/10.1371/journal.pone.0111301) PMID: [25436638](https://pubmed.ncbi.nlm.nih.gov/25436638/)
21. Denny JC, Crawford DC, Ritchie MD, Bielinski SJ, Basford MA, Bradford Y, et al. Variants near FOXE1 are associated with hypothyroidism and other thyroid conditions: using electronic medical records for genome- and phenome-wide studies. *Am J Hum Genet*. 2011; 89: 529–542. doi: [10.1016/j.ajhg.2011.09.008](https://doi.org/10.1016/j.ajhg.2011.09.008) PMID: [21981779](https://pubmed.ncbi.nlm.nih.gov/21981779/)
22. Matisse TC, Ambite JL, Buyske S, Carlson CS, Cole SA, Crawford DC, et al. The Next PAGE in Understanding Complex Traits: Design for the Analysis of Population Architecture Using Genetics and Epidemiology (PAGE) Study. *American Journal of Epidemiology*. 2011; 174: 849–859. doi: [10.1093/aje/kwr160](https://doi.org/10.1093/aje/kwr160) PMID: [21836165](https://pubmed.ncbi.nlm.nih.gov/21836165/)
23. Crawford DC, Goodloe R, Farber-Eger E, Boston J, Pendergrass SA, Haines JL, et al. Leveraging epidemiologic and clinical collections for genomic studies of complex traits. *Human Heredity*. In press.
24. Resnikoff S, Pascolini D, Etya'a'le D, Kocur I, Pararajasegaram R, Pokharel GP, et al. Global data on visual impairment in the year 2002. *Bull World Health Organ*. 2004; 82: 844–851. PMID: [15640920](https://pubmed.ncbi.nlm.nih.gov/15640920/)
25. Congdon N, O'Colmain B, Klaver CC, Klein R, Munoz B, Friedman DS, et al. Causes and prevalence of visual impairment among adults in the United States. *Arch Ophthalmol*. 2004; 122: 477–485. PMID: [15078664](https://pubmed.ncbi.nlm.nih.gov/15078664/)
26. Stein JD, Kim DS, Niziol LM, Talwar N, Nan B, Musch DC, et al. Differences in Rates of Glaucoma among Asian Americans and Other Racial Groups, and among Various Asian Ethnic Groups. *Ophthalmology*. 2011; 118: 1031–1037. doi: [10.1016/j.ophtha.2010.10.024](https://doi.org/10.1016/j.ophtha.2010.10.024) PMID: [21310489](https://pubmed.ncbi.nlm.nih.gov/21310489/)
27. Friedman DS, Wolfs RC, O'Colmain BJ, Klein BE, Taylor HR, West S, et al. Prevalence of open-angle glaucoma among adults in the United States. *Arch Ophthalmol*. 2004; 122: 532–538. PMID: [15078671](https://pubmed.ncbi.nlm.nih.gov/15078671/)
28. Vajaranant TS, Wu S, Torres M, Varma R. The Changing Face of Primary Open-Angle Glaucoma in the United States: Demographic and Geographic Changes From 2011 to 2050. *American Journal of Ophthalmology*. 2012; 154: 303–314. doi: [10.1016/j.ajo.2012.02.024](https://doi.org/10.1016/j.ajo.2012.02.024) PMID: [22541661](https://pubmed.ncbi.nlm.nih.gov/22541661/)
29. Peissig PL, Rasmussen LV, Berg RL, Linneman JG, McCarty CA, Waudby C, et al. Importance of multi-modal approaches to effectively identify cataract cases from electronic health records. *Journal of the American Medical Informatics Association*. 2012; 19: 225–234. doi: [10.1136/amiainl-2011-000456](https://doi.org/10.1136/amiainl-2011-000456) PMID: [22319176](https://pubmed.ncbi.nlm.nih.gov/22319176/)
30. Burdon KP, Macgregor S, Hewitt AW, Sharma S, Chidlow G, Mills RA, et al. Genome-wide association study identifies susceptibility loci for open angle glaucoma at TMCO1 and CDKN2B-AS1. *Nat Genet*. 2011; 43: 574–578. doi: [10.1038/ng.824](https://doi.org/10.1038/ng.824) PMID: [21532571](https://pubmed.ncbi.nlm.nih.gov/21532571/)
31. McCarty C, Chisholm R, Chute C, Kullo I, Jarvik G, Larson E, et al. The eMERGE Network: A consortium of biorepositories linked to electronic medical records data for conducting genomic studies. *BMC Medical Genomics*. 2011; 4: 13. doi: [10.1186/1755-8794-4-13](https://doi.org/10.1186/1755-8794-4-13) PMID: [21269473](https://pubmed.ncbi.nlm.nih.gov/21269473/)

32. Crimmins EM, Hayward MD, Seeman TE (2004) Race/ethnicity, socioeconomic status, and health. In: Anderson NB, Bulatao RA, Cohen B, editors. *Critical Perspectives on Racial and Ethnic Differences in Health in Late Life*. Washington, DC: The National Academies Press.
33. Ladapo JA, Kymes SM, Ladapo JA, Nwosu VC, Pasquale LR. Projected clinical outcomes of glaucoma screening in African American individuals. *Arch Ophthalmol*. 2012; 130: 365–372. doi: [10.1001/archophthalmol.2011.1224](https://doi.org/10.1001/archophthalmol.2011.1224) PMID: [22411665](https://pubmed.ncbi.nlm.nih.gov/22411665/)
34. Friedman DS, Katz J, Bressler NM, Rahmani B, Tielsch JM. Racial differences in the prevalence of age-related macular degeneration: The Baltimore eye survey. *Ophthalmology*. 1999; 106: 1049–1055. PMID: [10366070](https://pubmed.ncbi.nlm.nih.gov/10366070/)
35. Liu Y, Hauser MA, Akafo SK, Qin X, Miura S, Gibson JR, et al. Investigation of Known Genetic Risk Factors for Primary Open Angle Glaucoma in Two Populations of African Ancestry. *Investigative Ophthalmology & Visual Science*. 2013; 54: 6248–6254.
36. Nakano M, Ikeda Y, Tokuda Y, Fuwa M, Omi N, Ueno M, et al. Common Variants in *CDKN2B-AS1* Associated with Optic-Nerve Vulnerability of Glaucoma Identified by Genome-Wide Association Studies in Japanese. *PLoS ONE*. 2012; 7: e33389. doi: [10.1371/journal.pone.0033389](https://doi.org/10.1371/journal.pone.0033389) PMID: [22428042](https://pubmed.ncbi.nlm.nih.gov/22428042/)
37. Osman W, Low SK, Takahashi A, Kubo M, Nakamura Y. A genome-wide association study in the Japanese population confirms 9p21 and 14q23 as susceptibility loci for primary open angle glaucoma. *Human Molecular Genetics*. 2012; 21: 2836–2842. doi: [10.1093/hmg/dds103](https://doi.org/10.1093/hmg/dds103) PMID: [22419738](https://pubmed.ncbi.nlm.nih.gov/22419738/)
38. Wiggs JL, Yaspan BL, Hauser MA, Kang JH, Allingham RR, Olson LM, et al. Common Variants at 9p21 and 8q22 Are Associated with Increased Susceptibility to Optic Nerve Degeneration in Glaucoma. *PLoS Genet*. 2012; 8: e1002654. doi: [10.1371/journal.pgen.1002654](https://doi.org/10.1371/journal.pgen.1002654) PMID: [22570617](https://pubmed.ncbi.nlm.nih.gov/22570617/)
39. Kho AN, Pacheco JA, Peissig PL, Rasmussen L, Newton KM, Weston N, et al. Electronic Medical Records for Genetic Research: Results of the eMERGE Consortium. *Science Translational Medicine*. 2011; 3: 79re1. doi: [10.1126/scitranslmed.3001807](https://doi.org/10.1126/scitranslmed.3001807) PMID: [21508311](https://pubmed.ncbi.nlm.nih.gov/21508311/)
40. Voight BF, Kang HM, Ding J, Palmer CD, Sidore C, Chines PS, et al. The MetaboChip, a Custom Genotyping Array for Genetic Studies of Metabolic, Cardiovascular, and Anthropometric Traits. *PLoS Genet*. 2012; 8: e1002793. doi: [10.1371/journal.pgen.1002793](https://doi.org/10.1371/journal.pgen.1002793) PMID: [22876189](https://pubmed.ncbi.nlm.nih.gov/22876189/)
41. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analysis. *Am J Hum Genet*. 2007; 81: 559–575. PMID: [17701901](https://pubmed.ncbi.nlm.nih.gov/17701901/)