Received: 3 August 2018 | Revised: 9 November 2018 | Accepted: 11 December 2018

DOI: 10.1002/gepi.22189

RESEARCH ARTICLE

WILEY Genetic Epidemiology

OFFICIAL JOURNAL INTERNATIONAL GENETIC EPIDEMIOLOGY SOCIETY www.geneticepi.org

Spinning convincing stories for both true and false association signals

Richard J. Biedrzycki[1] | Ashley E. Sier[1,2]* | Dongjing Liu[1]* | Erika N. Dreikorn[1] | Daniel E. Weeks[1,3]

[1]Department of Human Genetics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, Pennsylvania

[2]Department of Epidemiology, University of Pittsburgh Graduate School of Public Health, Pittsburgh, Pennsylvania

[3]Department of Biostatistics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, Pennsylvania

Correspondence
Daniel E. Weeks, Department of Human Genetics, University of Pittsburgh Graduate School of Public Health, Public Health 3119, 130 DeSoto ST, Pittsburgh, PA 15261.
Email: weeks@pitt.edu

Funding information
Graduate School of Public Health, University of Pittsburgh

Abstract

When interpreting genome-wide association peaks, it is common to annotate each peak by searching for genes with plausible relationships to the trait. However, "all that glitters is not gold"—one might interpret apparent patterns in the data as plausible even when the peak is a false positive. Accordingly, we sought to see how human annotators interpreted association results containing a mixture of peaks from both the original trait and a genetically uncorrelated "synthetic" trait. Two of us prepared a mix of original and synthetic peaks of three significance categories from five different scans along with relevant literature search results and then we all annotated these regions. Three annotators also scored the strength of evidence connecting each peak to the scanned trait and the likelihood of further studying that region. While annotators found original peaks to have stronger evidence ($p_{\text{Bonferroni}} = 0.017$) and higher likelihood of further study ($p_{\text{Bonferroni}} = 0.006$) than synthetic peaks, annotators often made convincing connections between the synthetic peaks and the original trait, finding these connections 55% of the time. These results show that it is not difficult for annotators to make convincing connections between synthetic association signals and genes found in those regions.

KEYWORDS

association peaks, false positives, genome-wide association studies, literature review

1 | INTRODUCTION

We were once approached by a collaborator very excited about an association signal, stating that there is a gene in the region that is a perfect candidate for further follow-up and functional validation. However, further examination of the statistical results revealed that an error had been made in processing the data and the signal was entirely spurious. Such apparently convincing annotation of false-positive results can mislead and misdirect researchers, leading to wasted resources, time, and money. However, spinning a convincing story appears to be easy for any region in the genome, whether or not there is a true signal there. For example, Ioannidis, Tarone, and McLaughlin (2011) showed that 99% of early candidate genes identified for association with schizophrenia (SCZ) did not replicate using more powerful study designs. For the majority of these candidate gene studies, researchers were able, at the time of publication, to make convincing connections between these (likely) false positives and SCZ.

*These authors contributed equally to this study.

Our goal is to test if annotators of genome-wide association study (GWAS) results are able to make convincing connections between given traits and "synthetic peaks," loci wherein single-nucleotide polymorphisms (SNPs) were tested for traits genetically uncorrelated with the traits annotators were asked to annotate for. To do this, two of us created mixtures of "original" peaks and "synthetic" peaks for multiple GWAS scans using summary statistics from previous consortia studies, assembling detailed plots of the significantly associated regions and literature search results for genes and SNPs found in these regions. Using these materials as well as other resources as desired, we all, acting as annotators, attempted to make connections between the scanned trait and the associated regions. We expected annotators would be able to make convincing connections between the traits and the genes for both the "original" peaks and the "synthetic" peaks. As far as we are aware, this is the first study of its kind.

2 | METHODS

2.1 | Annotators

We (R.J.B. and D.E.W.) initially attempted to recruit annotators as coauthors, presenting this study at our fall 2017 departmental retreat to students, faculty, and colleagues. Interested annotators were provided with a handout that made them aware that for each trait, some of the peaks would be from a prior GWAS of a trait unrelated to the stated trait (Supporting Information S1 Text). Ultimately only three (A.E.S., D.L., and E.N.D.) submitted annotation; these three were completely blind to the source of the synthetic peaks. To increase the annotation pool, R.J.B. and D.E.W. also annotated scans. R.J.B was an M.S. in Human Genetics graduate student, D.E.W. is a Professor of Human Genetics and Biostatistics, and A.E.S was an M.P.H. in Epidemiology graduate student, D.L. is a Ph.D. in Human Genetics graduate student, and E.N.D. is a Ph.D. in Human Genetics graduate student.

2.2 | Selection of traits

As we were interested in whether human annotators could find convincing connections between synthetic peaks and scanned GWAS traits, we selected traits that were not genetically correlated with each other. Using genetic correlation data from Bulik-Sullivan et al. (2015), we created genetically uncorrelated trait pairs with correlation <0.10, producing the following pairs: SCZ and low-density lipoprotein (LDL); triglycerides (TG) and Crohn's disease (CD); high-density lipoprotein (HDL) and ulcerative colitis (UC); and coronary artery disease (CAD) and Alzheimer's disease (AD). We obtained GWAS summary statistics for these traits from the Psychiatric Genomics Consortium (PGC) for SCZ (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014), the Global Lipid Genetics Consortium for HDL, LDL, and TG (Willer et al., 2013), the Coronary Artery Disease (C4D) Genetics Consortium for CAD (The Coronary Artery Disease (C4D) Genetics Consortium, 2011), the International IBD Genetics Consortium for UC and CD (Liu et al., 2015), and the International Genomics of Alzheimer's Project for AD (Lambert et al., 2013). We randomly selected which trait would be used as the scan trait and which trait would provide synthetic peaks. In doing so, LDL, CD, and UC were chosen to provide these synthetic peaks. To examine if trait selection in each pair made a difference in what genes and regions were selected as having convincing connections, CAD and AD were chosen to provide each other's respective synthetic peaks, with two annotators annotating for (CAD, AD) and one annotating for (AD, CAD), where the first trait provided original peaks and the second trait provided synthetic peaks. Our hypothesis is that annotators would still make convincing connections between synthetic peaks and the scanned trait but the genes and regions they identify as noteworthy may be different for CAD and AD.

2.3 | Data cleaning and peak selection

We acquired summary statistics from the consortium websites and selected peaks with at least 500 kb separating each locus. Three peak significance categories were established: (a) "highly significant" for peaks with a p value less than 1×10^{-15}, (b) "moderately significant" for peaks with a p value between 5×10^{-8} and 1×10^{-15}, and (c) "suggestively significant" for peaks with a p value between 1×10^{-5} and 5×10^{-8}. We excluded any peak SNPs with an imputation INFO score less than 0.90, and we excluded peak SNPs that were indel variants. We excluded peaks with minor allele frequency less than 0.05 except for SCZ, where a cutoff of 0.10 was used instead to match the methods used by the PGC (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014). For "highly significant" and "moderately significant" peaks, we excluded peaks that were not reported in the original studies. We did not do this for "suggestively significant" peaks because it is quite likely that suggestively significant peaks were not highlighted in the original studies. For each trait pair, nine peaks were selected, with one of each type of significance for the original trait and two for each type of significance for the synthetic peaks, excluding (AD, CAD) which had eight as there was only one "moderately significant" CAD peak available to be used as a synthetic peak. As we were primarily interested in the annotation of synthetic peaks

and annotation time was a concern, we chose to have only one original trait peak for each significance category as opposed to two. Data cleaning and analysis were performed in R (R Core Team, 2017). We created regional association plots of each peak using LocusZoom with windows of 400 kb and using the European hg19 build from the 1,000 Genomes Project to provide linkage disequilibrium (LD) information (Pruim et al., 2010).

2.4 | Literature review and annotation of significant hits

We performed automated literature review searches of PubMed, PubMed Central, and Google Scholar using the R packages "RISmed," "rvest," and "data.table" (Dowle, 2017; Kovalchik, 2017; Wickham, 2016). Due to web-scraping limitations, up to 1,000, 20, and 10 results, respectively, from each source, were obtained and recorded in spreadsheets. For each peak, our search query was the name of the peak SNP and the scanned trait, for example, "rs4393438 AND schizophrenia." We also queried the name of genes within ±50 kb of the SNP or genes with at least one marker in LD with the peak SNP with an $r^2 > 0.2$ and the scanned trait, for example, "RASA3 AND schizophrenia."

The five authors, acting as annotators, used the provided literature search results and LocusZoom plots to annotate each peak within a scan as having convincing connections with the scanned trait or not and record the results within provided summary sheets. Annotators were free to carry out additional literature and database searches on their own. The provided summary sheets stated the scanned trait and were divided up into multiple pages containing LocusZoom plots of each selected peak region centered on the most significant SNP as well as space for the annotators to write a paragraph describing their annotation results. The instructions provided to the annotators can be seen in Supporting Information S1 Text. Three annotators were also asked to rate the strength of evidence for association with the scanned trait of a particular peak on a scale of 0–3 with 0 being no evidence and 3 being very strong evidence. They were also asked to rate the likelihood of further study of those particular peaks on a scale of 0–3 with 0 being no likelihood and 3 being a very high likelihood.

2.5 | Analysis of annotation

Using the summary sheets filled out by the annotators, we recorded whether the annotators found a peak to have convincing connections with the scanned trait. To do this, we read the filled summary sheets and determined whether the annotators found connections based on what they had written. If the annotator explicitly stated that they found a connection, this was recorded as a convincing connection being made. If the annotator explicitly stated that they did not find a connection, this was recorded as no connection being made. If the annotator was not explicit in their response, the content of their report was more closely examined. If the annotator mentioned previous associations between the trait and loci, this was recorded as a convincing connection. As only three of the annotators were asked to rate the strength of evidence and likelihood of further study of each peak, we did not use these responses to determine whether a convincing connection was made or not.

To determine if there was an association between a peak being annotated as having convincing connections with the scanned trait and original/synthetic peak status, we performed Fisher's exact test for each individual annotator and Pearson's chi-squared test for all annotators combined. We did the same for peak significance category. To determine if there was an association between peak significance category and strength of evidence of association and likelihood of further study, we performed the Kruskal–Wallis test for the three annotators who were asked to answer this and all three of these annotators combined. We also performed Wilcoxon rank-sum tests for original/synthetic peak type and strength of evidence and likelihood of further study. We then performed polytomous multinomial logistic regression to assess the effect of original/synthetic peak status and peak significance category on the strength of evidence and the likelihood of further study. For both peak significance category and original/synthetic peak status, a rating of 0 was used as the reference.

3 | RESULTS

All five authors, acting as annotators, used the provided literature search results as well as any other desired sources to annotate whether genes at associated loci

TABLE 1 Traits annotated by each annotator, with "+" indicating they annotated this trait, and "–" indicating they did not

	Scanned Trait (original, synthetic)				
Annotator	(SCZ, LDL)	(TG, CD)	(HDL, UC)	(CAD, AD)	(AD, CAD)
1 (R.J.B.)	+	+	+	–	+
2 (E.N.D.)	+	+	–	–	–
3 (A.E.S.)	+	+	+	+	–
4 (D.L.)	+	+	+	+	–
5 (D.E.W.)	+	+	–	–	–

Note. AD: Alzheimer's disease; CAD: coronary artery disease; CD: Crohn's disease; HDL: high-density lipoprotein; LDL: low-density lipoprotein; SCZ: schizophrenia; TG: triglycerides; UC: ulcerative colitis.

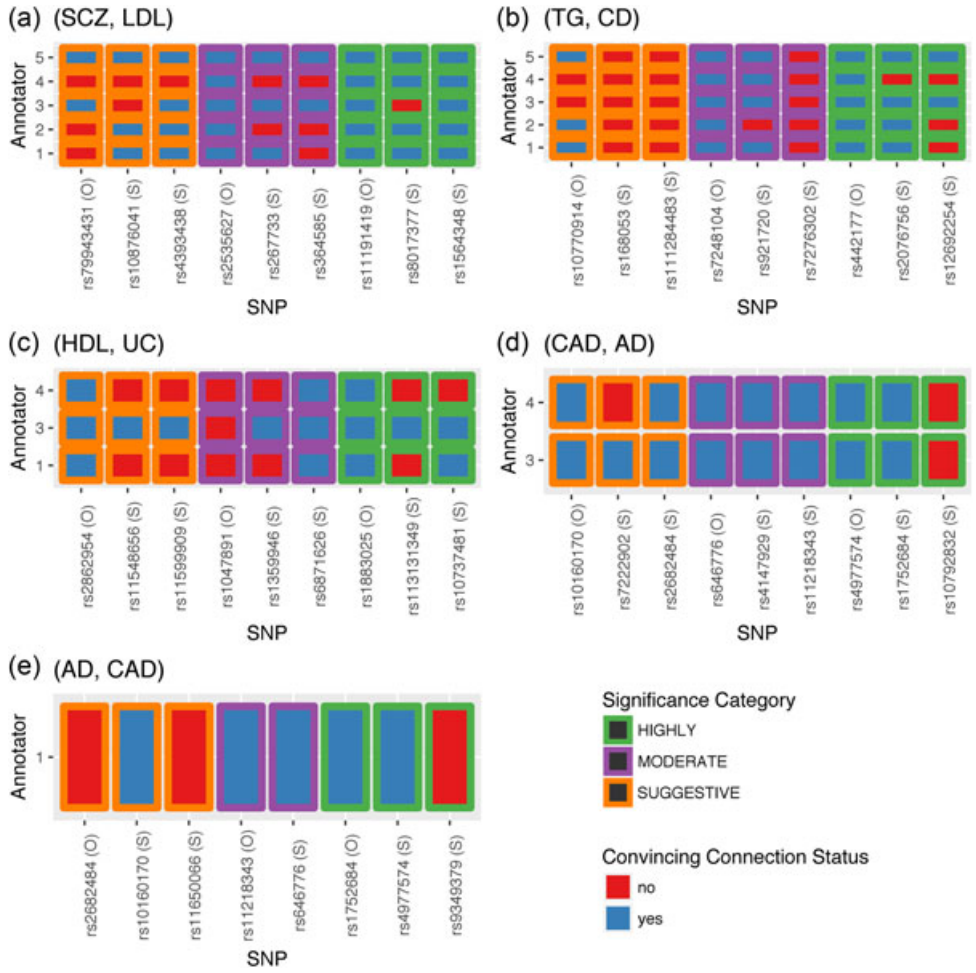

FIGURE 1 Annotation status of SNPs for showing evidence for association divided by annotator. (a) Schizophrenia and LDL, (b) triglycerides and Crohn's disease, (c), HDL and ulcerative colitis, (d) coronary artery disease and Alzheimer's disease, and (e) Alzheimer's disease and coronary artery disease. "O" indicates an original peak, and "S" indicates a synthetic peak. HDL: high-density lipoprotein; LDL: low-density lipoprotein; SNPs: single-nucleotide polymorphisms

identified through GWAS had convincing connections to the scanned trait. The traits each author annotated can be seen in Table 1; due to limited time, annotators 2 and 5 only annotated two of the scans.

Each scan consisted of nine peaks (except AD which had eight) divided into three peak significance categories: suggestively significant, moderately significant, or highly significant. Each category contained three peaks: one "original" peak using summary statistics from a consortium studying the scanned trait, and two "synthetic" peaks using summary statistics from a consortium studying a genetically uncorrelated trait, except AD which had only one moderately significant synthetic peak in its scan.

How each peak was annotated can be seen in Figure 1a-e. Recall that for each significance level, we have one original and two synthetic peaks, except in the (AD, CAD) scan (Figure 1e), where there was only one moderately significant synthetic peak available. As significance category increased, there is an increase in the number of convincing connections made, with all of the annotators finding convincing connections for all of the highly significant original peaks and all but one of the moderately significant original peaks. Many of the synthetic peaks were also found to have convincing connections by the annotators.

Annotators 1, 4, and 5 were asked to answer how strong they felt the evidence for association was at each peak as well as the likelihood of continuing to study that region for its role in the scanned trait. The counts of their responses by peak type and peak significance category can be seen in Figure 2 a,b. Synthetic peaks had wider distributions of strength of evidence and likelihood of further study scores than original peaks. For original peaks, both strength of evidence and likelihood of further study scores increased as significance category increased.

Annotators found convincing connections between 55% of synthetic peaks and 81% of original peaks and the scanned

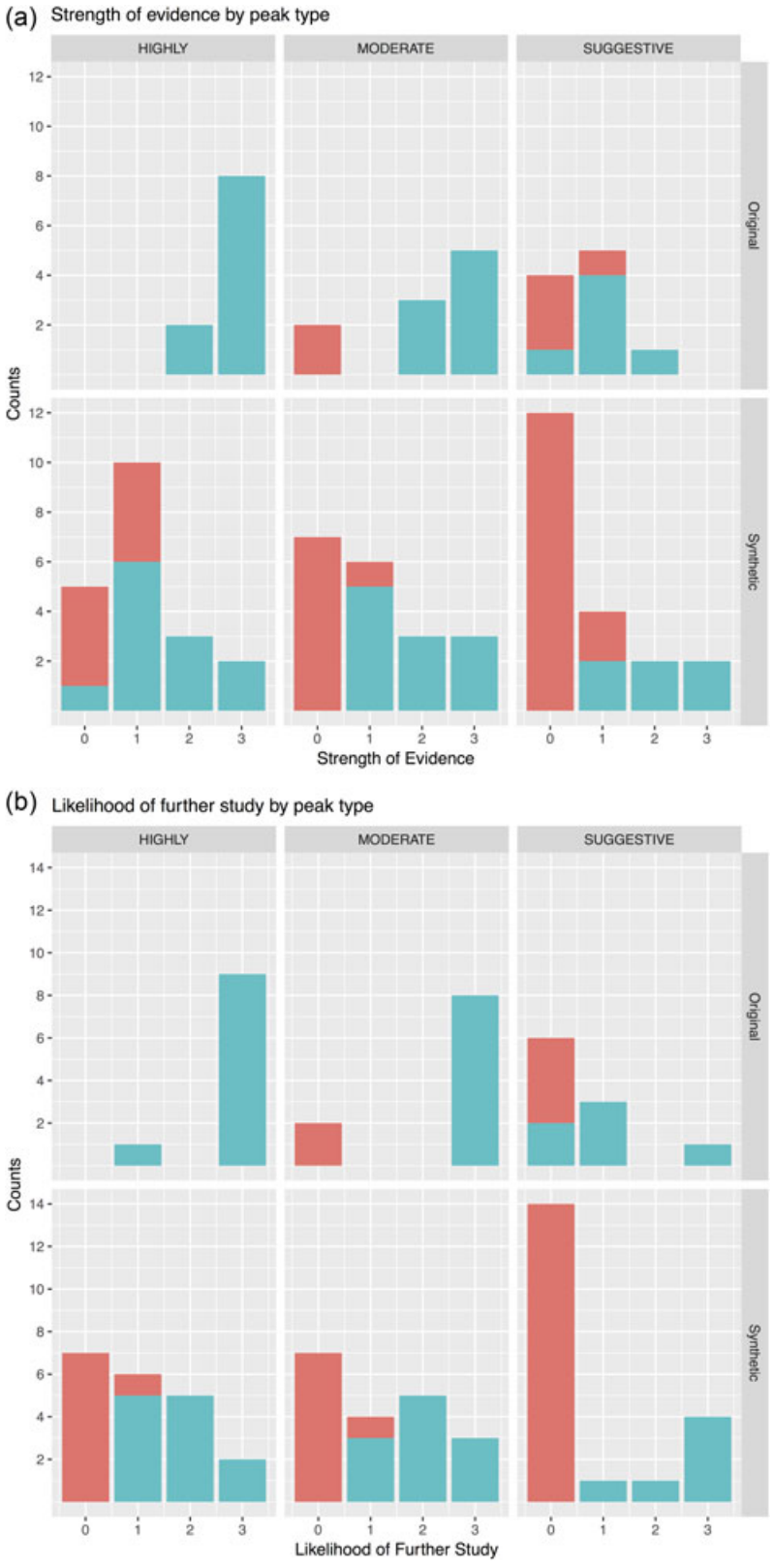

FIGURE 2 Strength of evidence and likelihood of further study of annotated peaks. (a) Strength of evidence and (b) likelihood of further study scores from Annotators 1, 4, and 5 by peak significance category, original or synthetic peak, and convincing connection status signified by red for no convincing connection made and teal for a convincing connection made

trait (Table 2A). We found that convincing connection status was significantly associated with original/synthetic peak status ($\chi^2 = 8.58$; $df = 1$; $p = 0.003$). We also found that convincing connection status was significantly associated with peak significance category ($\chi^2 = 8.99$; $df = 2$; $p = 0.011$) (Table 2B). As the peak significance category increased from suggestive to moderate to highly, convincing connections were made more frequently not only for original peaks

TABLE 2A Convincing connection status by peak type

Peak type	Convincing connection status *N* (%)	
	No	Yes
Synthetic	43 (45)	52 (55)
Original	9 (19)	39 (81)

$\chi^2 = 8.58$, $df = 1$, $p = 0.003$.

(increasing from 62% to 100%) but also for synthetic peaks (increasing from 41% to 66%) (Table 2C). We performed Fisher's exact tests of independence of convincing connection status and original/synthetic peak type for each annotator but none was significant, possibly from a lack of power due to low sample sizes.

To assess if the peak type was associated with the strength of evidence and the likelihood of further study, we performed Wilcoxon rank-sum tests. We found that there was a significant difference between original and synthetic peaks for both strength of evidence ($U = 518.5$; $p < 0.001$) and likelihood of further study ($U = 543$; $p = 0.002$). After performing Kruskal–Wallis tests, we found that likelihood of further study was significantly associated with peak significance category ($H = 6.13$; $df = 2$; $p = 0.047$) while strength of evidence was not significantly associated ($H = 4.71$; $df = 2$; $p = 0.095$).

We then performed polytomous multinomial logistic regression to evaluate the effects of peak significance category and original/synthetic peak type on strength of evidence and likelihood of further study. To adjust for multiple testing, we used a Bonferroni correction for six tests. Annotators found original peaks to have significantly stronger evidence ($\beta = 2.01$; standard error $[SE] = 0.65$; $t = 3.07$; $p = 0.003$; $p_{\text{Bonf}} = 0.017$) and would be significantly more likely to continue studying original peaks compared with synthetic peaks ($\beta = 1.95$; $SE = 0.57$; $t = 3.40$; $p = 0.001$; $p_{\text{Bonf}} = 0.006$). Annotators found suggestively significant peaks to have significantly weaker evidence than highly significant peaks ($\beta = -2.77$; $SE = 0.93$; $t = -2.99$; $df = 79$; $p = 0.004$; $p_{\text{Bonf}} = 0.011$). They also were more likely to recommend highly significant peaks for further study than suggestively significant peaks, but none of the tests was significant after correction for multiple testing ($p > 0.05$).

TABLE 2B Convincing connection status by peak significance category

Peak significance category	Convincing connection status *N* (%)	
	No	Yes
Highly	11 (23)	37 (77)
Moderate	16 (34)	31 (66)
Suggestive	25 (52)	23 (48)

$\chi^2 = 8.99$, $df = 2$, $p = 0.011$.

TABLE 2C Convincing connection status by peak type and peak significance category

Peak significance category	Peak type	Convincing connection status *N* (%)	
		No	Yes
Highly	Original	0 (0)	16 (100)
	Synthetic	11 (34)	21 (66)
Moderate	Original	3 (19)	13 (81)
	Synthetic	13 (42)	18 (58)
Suggestive	Original	6 (38)	10 (62)
	Synthetic	19 (59)	13 (41)

We were interested in whether the selection of the trait that was provided affected not only what peak-specific convincing connections were found but also what specific genes in those peaks annotators found to be of interest. To do this, we compared the annotation of the (AD, CAD) set to that of the (CAD, AD) set, examining whether the genes identified by the annotators were different or not between these related sets. Of the six peaks that overlapped between the two sets, Annotator 1 found genes with convincing connections for (AD, CAD) that matched the genes Annotators 3 and 4 found for (CAD, AD). The SNP regions and their corresponding genes were rs10160170 (*CXCL12*), rs11218343 (*SORL1*), rs646776 (*SORT1*), rs1752684 (*CR1*), and rs4977574 (*CDKN2A*) (Supporting Information Table S2).

4 | DISCUSSION

As far as we are aware, this is the first study to analyze annotators' ability to interpret GWAS synthetic peaks. In this study, we have shown that, given literature search results and a mixture of original and synthetic association peaks for a scanned trait, annotators are able to make a large number of convincing connections between synthetic peaks and the scanned trait, with 55% of the synthetic peaks being annotated as having convincing connections (Table 2A). As might be expected, annotators had a higher likelihood of wanting to further study original peaks as well as finding original peaks as having greater strength of evidence compared to synthetic peaks.

We found that convincing connection status was significantly associated with both peak significance category and original/synthetic peak status (Tables 2A–2C). This is as we expected as one would assume that more significantly associated peaks would have been studied further and, thus, have more evidence for annotators to find within the

literature. In fact, multiple papers cited by the annotators also cited the GWAS from which the summary statistics were obtained. One would also make the fair assumption that peaks that have previously been found to be significantly associated with a trait should be annotated as having convincing connections more often than peaks that are less significant.

We found that original peaks had significantly different strengths of evidence and likelihood of further study compared with synthetic peaks, with original peaks having significantly more counts of having very strong evidence compared with synthetic peaks (Figure 2). This makes sense as one would expect original peaks to have more evidence connecting them to the scanned trait within the literature than the synthetic peaks. Polytomous multinomial logistic regression showed significant differences between highly significant peaks and suggestively significant peaks for no strength of evidence and very strong evidence. Again, this is as we would expect as suggestively significant peaks are more likely to be false positives than highly significant peaks and, thus, have less evidence supporting their connections to the scanned trait within the literature.

When comparing the annotation results of the (CAD, AD) and (AD, CAD) related sets, not only did the annotators find convincing connections between a majority of the synthetic peaks and the scanned trait, but the genes of interest they selected for the scanned trait matched those selected for the original trait for five of the six overlapping peaks (Supporting Information Table S2). One possible explanation is that CAD and AD are not as genetically uncorrelated as proposed by Bulik-Sullivan et al. (2015) (and more recently by Pickrell et al., 2016), which could explain the matches seen at these peaks. For example, each of these five genes identified by the annotators is thought to be associated with inflammation, which may play a role in the development of both CAD and AD (Rea et al., 2018). *CXCL12* has been shown to be associated with neuroinflammation (Pappas et al., 2015), *SORT1* has been found to be associated with levels of progranulin, a protein which plays a role in inflammation during wound healing (Meeter et al., 2016), *CR1* is a member of the complement receptor family which drives inflammation (Morgan & Harris, 2015), cyclin-dependent kinases, which may be inhibited by *CDKN2A*, a cyclin-dependent kinase inhibitor, are involved in the regulation of inflammatory factors, and *SORL1* is associated with inflammatory intermediary molecules (Ogita et al., 2013). However, inflammation was not the primary reason annotators found these genes to be of interest (Supporting Information Table S2); they were merely one of the attributes contributing to some of their selections. If these genes are involved in multiple pathways and have multiple effects, this pleiotropy may make it easier for an annotator to discover a connection between a synthetic peak and the scanned trait.

There were multiple limitations of the study design which could have reduced our power and are worth discussing. First, there was a small number of annotators which may have limited our ability to detect some significant associations. Second, two of the annotators were not blind to the name of the synthetic traits. However, as the focus was on annotating the peak with respect to the stated trait, knowing the name of the synthetic trait had little bearing on whether or not there was annotation showing a convincing connection to the stated trait. Third, the annotation process used in this study may be more superficial than what is actually done in a real study and may be dependent on the background and expertise of the annotator. In this study, we asked annotators to look at literature search results, but in a real study, annotators are also likely to use other sources such as RegulomeDB and other online databases during the annotation process (Boyle et al., 2012). Similarly, large research teams may collectively contain higher levels of annotation expertise than was applied here. As such, this study may not be as accurate a facsimile of annotation as possible. The fourth limitation is the subjective interpretation of whether a convincing connection was made or not. As mentioned earlier, determining whether a convincing connection was made depended on the paragraph written by the annotator. As the interpretation of these paragraphs is subjective, there may be discrepancies in what the annotator and the investigator considers a convincing connection; this could be addressed through the addition of more questions about the peak. Despite these limitations, we believe this represents an important first step in examining how easy it is to make convincing connections between a synthetic peak and the scanned trait.

It is important to remember that, if we were omniscient, we could design a study where all original peaks selected would be truly associated with the scanned trait and all synthetic peaks selected would be truly unassociated. This would allow us to analyze annotators' ability to make convincing connections without worry that the connections they make are in fact real. Because we are not omniscient, there is no guarantee that the synthetic peaks are truly

unassociated with the original trait. One way to improve synthetic peak selection might be to use the GWAS Catalog to select synthetic peaks that have not been found to be associated with the scanned trait (MacArthur et al., 2017). However, due to the interconnected nature of biology, we cannot be certain of the true nature of the peaks.

5 | CONCLUSIONS

To analyze the ability of annotators to find convincing connections between false-positive results and the scanned trait of a GWAS, we created synthetic peaks using summary statistics from GWAS of genetically uncorrelated traits and asked annotators to annotate a mix of original peaks and these synthetic peaks. We found that convincing connection status was significantly associated with both original/synthetic peak type and significance category. Annotators also found original peaks to have significantly stronger evidence than synthetic peaks and a significantly increased likelihood of further study, while significance category was significantly associated with the likelihood of further study. These results show that human annotators are easily able to make convincing connections between synthetic peaks and the scanned trait of a GWAS.

ACKNOWLEDGMENTS

This work is derived from Richard J. Biedrzycki's Master's Thesis, and we would like to thank Dr. John Shaffer and Dr. Wei Chen for their helpful comments and suggestions on the thesis. This work was supported by the University of Pittsburgh.

ORCID

Daniel E. Weeks http://orcid.org/0000-0001-9410-7228

REFERENCES

Boyle, A. P., Hong, E. L., Hariharan, M., Cheng, Y., Schaub, M. A., Kasowski, M., ... Snyder, M. (2012). Annotation of functional variation in personal genomes using RegulomeDB. *Genome Research*, *22*(9), 1790–1797. https://doi.org/10.1101/gr.137323.112

Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P.-R., ... Neale, B. M. (2015). An Atlas of genetic correlations across human diseases and traits. *Nature Genetics*, *47*(11), 1236–1241. https://doi.org/10.1038/ng.3406

Dowle, M. A. S. (2017). data.table: Extension of 'data.frame' (version 1.10.4). CRAN. https://CRAN.R-project.org/package=data.table

Ioannidis, J. P. A., Tarone, R., & McLaughlin, J. K. (2011). The false-positive to false-negative ratio in epidemiologic studies. *Epidemiology*, *22*(4), 450–456. https://doi.org/10.1097/EDE.0b013e31821b506e

Kovalchik, S. (2017). *RISmed: Download content from NCBI databases (Version 2.1.7)* [Package]. CRAN: https://CRAN.R-project.org/package=RISmed

Lambert, J. -C., Ibrahim-Verbaas, C. A., Harold, D., Naj, A. C., Sims, R., Bellenguez, C., ... Amouyel, P. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature Genetics*, *45*(12), 1452–1458. https://doi.org/10.1038/ng.2802

Liu, J. Z., van Sommeren, S., Huang, H., Ng, S. C., Alberts, R., Takahashi, A., ... Weersma, R. K. (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nature Genetics*, *47*(9), 979–986. https://doi.org/10.1038/ng.3359

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*(Database issue), D896–D901. https://doi.org/10.1093/nar/gkw1133

Meeter, L. H. H., Patzke, H., Loewen, G., Dopper, E. G. P., Pijnenburg, Y. A. L., van Minkelen, R., & van Swieten, J. C. (2016). Progranulin levels in plasma and cerebrospinal fluid in granulin mutation carriers. *Dementia and Geriatric Cognitive Disorders Extra*, *6*(2), 330–340. https://doi.org/10.1159/000447738

Morgan, B. P., & Harris, C. L. (2015). Complement, a target for therapy in inflammatory and degenerative diseases. *Nature Reviews Drug Discovery*, *14*, 857–877. https://doi.org/10.1038/nrd4657

Ogita, M., Miyauchi, K., Dohi, T., Tsuboi, S., Miyazaki, T., Yokoyama, T., ... Daida, H. (2013). Increased circulating soluble LR11 in patients with acute coronary syndrome. *Clinica Chimica Acta*, *415*, 191–194. https://doi.org/10.1016/j.cca.2012.10.047

Pappas, A., Chaiworapongsa, T., Romero, R., Korzeniewski, S. J., Cortez, J. C., Bhatti, G., ... Tarca, A. L. (2015). Transcriptomics of maternal and fetal membranes can discriminate between gestational-age matched preterm neonates with and without cognitive impairment diagnosed at 18–24 months. *PLoS One*, *10*(3), e0118573. https://doi.org/10.1371/journal.pone.0118573

Pickrell, J. K., Berisa, T., Liu, J. Z., Ségurel, L., Tung, J. Y., & Hinds, D. A. (2016). Detection and interpretation of shared genetic influences on 42 human traits. *Nature Genetics*, *48*(7), 709–717. https://doi.org/10.1038/ng.3570

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

R Core Team. (2017). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org

Rea, I. M., Gibson, D. S., McGilligan, V., McNerlan, S. E., Alexander, H. D., & Ross, O. A. (2018). Age and age-related diseases: Role of inflammation triggers and cytokines. *Frontiers in Immunology*, *9*, 586. https://doi.org/10.3389/fimmu.2018.00586

Schizophrenia Working Group of the Psychiatric Genomics Consortium (2014). Biological insights from 108 schizophrenia-associated genetic

loci. *Nature*, *511*(7510), 421–427. https://doi.org/10.1038/nature13595

The Coronary Artery Disease (C4D) Genetics Consortium (2011). A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nature Genetics*, *43*(4), 339–344.

Wickham, H. (2016). rvest: Easily Harvest (Scrape) Web Pages (Version 0.3.2). CRAN: https://CRAN.R-project.org/package=rvest

Willer, C. J., Schmidt, E. M., Sengupta, S., Peloso, G. M., Gustafsson, S., Kanoni, S., ... Abecasis, G. R. (2013). Discovery and refinement of loci associated with lipid levels. *Nature Genetics*, *45*(11), 1274–1283. https://doi.org/10.1038/ng.2797

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: Biedrzycki RJ, Sier AE, Liu D, Dreikorn EN, Weeks DE. Spinning convincing stories for both true and false association signals. *Genet. Epidemiol.* 2019;43:356–364. https://doi.org/10.1002/gepi.22189