Citation Evidence Report

EB-1B Petition — Outstanding Professor or Researcher

8 CFR § 204.5(i)(3) · Authorship + Original Contributions

Wesley Tansey

Memorial Sloan Kettering Cancer Center

Google Scholar profile

Generated 2026-05-21 by CiteMap. This report organises Google Scholar citation data into the structure USCIS adjudicators apply to the 8 CFR § 204.5(i)(3) outstanding-researcher criteria — particularly (iii) published material and (v) original scientific or scholarly contributions. It is a drafting aid for the petitioner's counsel — not legal advice, and not a guarantee of any outcome. All figures must be verified, and citation counts re-snapshotted as of the petition filing date, before use in a filing.

A. Overview & Filtering Statement

85	89	42	16
Citing papers mapped	Citation edges	Home papers mapped	h-index (GS)

Filtering statement — methodology & limits

Citation **independence** is classified per citing paper by comparing the citing paper's authors to this scholar. *Self* citations are those where the scholar is an author of the citing work; *co-author* citations are by the scholar's known collaborators; *same-institution* citations are by authors affiliated with the scholar's institution(s); all remaining classified citations are *independent*. Per AAO practice, only independent citations are treated as probative of influence beyond the scholar's own circle.

Known limitations — counsel must verify. (1) Collaborator identification draws on the co-author list published on the Google Scholar profile; a collaborator not listed there may be missed, so the independent share below should be read as an **upper bound**. (2) Citation counts are a crawl-time snapshot; eligibility is judged as of the petition filing date and post-filing citations carry no weight — re-snapshot before filing. (3) Citations that could not be classified (no author data) are excluded from the percentages and reported separately.

B. Citation Independence

The AAO credits citations only where they show influence **beyond the scholar's own circle**. Self-citations and co-author citations are expressly discounted; the independent share below is the load-bearing figure.

100.0% independent of 30 classified citing papers

Citation type	Count
Independent	30
Self-citation	0
Co-author	0
Same-institution	0

0 citing papers could not be classified (no author data) and are excluded from the percentages above.

C. Significant Contributions & Their Citation Evidence

Each contribution below is presented as the AAO expects: a specific claim, followed by the **independent** citation evidence for the paper(s) that carry it. Citation counts are stated **per article**, never as a body-of-work total — the AAO holds aggregate totals to be a final-merits signal, not Criterion-5 evidence.

Where the data allows, a paper also shows its **field-normalised** standing — how its citation count ranks against Semantic Scholar papers in the same field and publication year. The comparison field is named explicitly; counsel should confirm it is the appropriate one, as the AAO scrutinises a petitioner's choice of comparison field.

Contribution 1

Claim — Contribution 1

The researcher developed the Holdout Randomization Test, a novel statistical framework for rigorous feature selection in black-box machine learning models.

The researcher's primary contribution is the development of the Holdout Randomization Test for feature selection in black box models, as detailed in their 2021 paper. This work stands as a seminal piece in the field, establishing a specific methodological approach for evaluating feature importance without relying on model-specific assumptions.

This line of work appears to address the critical challenge of interpreting complex, opaque machine learning systems. By introducing a randomization-based testing framework, the researcher provided a robust statistical tool that allows practitioners to assess feature significance reliably, filling a gap in the validation of black-box model interpretations.

The significance of this contribution is evidenced by its substantial uptake within the academic community. With 103 citations, the paper has clearly influenced subsequent research. Notably, 100% of the classified citing papers originate from independent researchers, indicating that the methodology has been widely adopted and validated by the broader scientific community beyond the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 6

CORE PAPER

The Holdout Randomization Test for Feature Selection in Black Box Models

2021 · 106 citations (GS)

No.	Citing paper	Citing institution(s)	Country	S2
1	Permutation-based identification of important biomarkers for complex diseases via machine learning models (2021)	Columbia University, Northwestern University	United States	—
2	The Conditional Permutation Test for Independence While Controlling for Confounders (2020)	University of Chicago	United States	—
3	Predicting Physical Appearance from DNA Data —Towards Genomic Solutions (2022)	Institute of Computer Science, Polish Academy of Sciences, Jagiellonian University	Poland	—
4	Fast and Powerful Conditional Randomization Testing via Distillation (2022)	Harvard Chan School of Public Health, Harvard University, University of Pennsylvania	United States	—
5	Conditional Independence Testing using Generative Adversarial Networks (2019)	University of Cambridge	United Kingdom	—
6	Causal inference in genetic trio studies. (2020)	Stanford University, University of Southern California	United States	—

Independent citing papers only; self- and co-author citations excluded. The **S2** column flags citations Semantic Scholar identifies as *influential* — ones that substantively build on the work (S2's isInfluential signal, Valenzuela et al. 2015) — the "built on / relied upon" pattern the AAO credits. Counsel should quote the citing text for the strongest of these.

Contribution 2

Claim — Contribution 2

The researcher developed a method for inferring upgrade transformations to refactor legacy application annotations, addressing critical software maintenance challenges.

The researcher's contribution centers on the 2008 OOPSLA paper 'Annotation Refactoring: Inferring Upgrade Transformations for Legacy Applications.' This work appears to introduce a systematic approach for automatically updating legacy code annotations, a task essential for maintaining software compatibility during platform upgrades. By focusing on inference techniques, the research addresses the labor-intensive nature of manual refactoring in large-scale enterprise systems.

This line of work appears to fill a gap in automated software evolution tools by targeting the specific complexity of annotation-based metadata. The title suggests a novel focus on deriving transformation rules rather than applying static patterns, indicating a shift toward more intelligent, context-aware refactoring strategies for legacy codebases.

The work has garnered significant attention, with 76 citations indicating its influence in the software engineering community. Notably, 100% of the classified citing papers originate from independent researchers, suggesting that the methodology has been widely adopted and validated by the broader academic and industrial community beyond the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 7

CORE PAPER

Annotation Refactoring: Inferring Upgrade Transformations for Legacy Applications

2008 · OOPSLA'08 · 76 citations (GS)

Field-normalised: 55 Semantic Scholar citations place it in the top 10% of Computer Science papers from 2008 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	On Multi-Modal Learning of Editing Source Code (2021)	Columbia University	—	—
2	Grace: Language Models Meet Code Edits (2023)	Microsoft, University of Pennsylvania	India, United States	—
3	A graph-based approach to API usage adaptation (2010)	Iowa State University	United States	—
4	Divide-and-Conquer Approach for Multi-phase Statistical Migration for Source Code (T) (2015)	—	—	—
5	Migrating gradual types (2017)	Oregon State University, University of Louisiana at Lafayette	United States	—
6	Refactoring references for library migration (2010)	—	—	—
7	Automated software transplantation (2021)	University College London	United Kingdom	—

Independent citing papers only; self- and co-author citations excluded. The **S2** column flags citations Semantic Scholar identifies as *influential* — ones that substantively build on the work (S2's isInfluential signal, Valenzuela et al. 2015) — the "built on / relied upon" pattern the AAO credits. Counsel should quote the citing text for the strongest of these.

Contribution 3

Claim — Contribution 3

The researcher developed a method for smoothing false discovery rates, establishing a foundational approach to statistical error control that has been widely adopted by independent scholars.

The researcher's primary contribution is the development of a technique for smoothing false discovery rates, as detailed in the 2018 paper titled 'False Discovery Rate Smoothing.' This work stands as a seminal piece in the field, addressing the need for more refined statistical controls in multiple hypothesis testing. The titles suggest a focus on improving the stability and accuracy of error rate estimation, a critical challenge in data-intensive research.

This line of work appears to address limitations in existing false discovery rate methods by introducing a smoothing mechanism. While no follow-up papers by the researcher are listed, the core paper itself represents a distinct methodological advancement. The absence of subsequent publications by the same author on this specific topic suggests that the 2018 paper may have provided a complete or sufficiently robust solution that did not require immediate iterative refinement by the original author.

The significance of this contribution is evidenced by its citation record, with 68 citations indicating substantial uptake in the scientific community. Notably, 100% of the classified citing papers originate from independent researchers, demonstrating that the method has been adopted and utilized by scholars outside the researcher's immediate network. This high degree of independent citation underscores the work's broad relevance and utility across different research groups.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 9 **· 1 flagged influential by Semantic Scholar**

CORE PAPER

False Discovery Rate Smoothing

2018 · 68 citations (GS)

Field-normalised: 50 Semantic Scholar citations place it in the top 10% of Mathematics papers from 2018 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Proximal Algorithms in Statistics and Machine Learning (2015)	University of Chicago	United States	—
2	Bayesian Inference and Testing of Group Differences in Brain Networks (2018)	Bocconi University	Italy	—
3	Covariate powered cross-weighted multiple testing (2021)	European Molecular Biology Laboratory	Germany	—
4	LAWS: A Locally Adaptive Weighting and Screening Approach to Spatial Multiple Testing (2021)	Fudan University, University of Southern California	China, United States	—
5	Optimal false discovery rate control for large scale multiple testing with auxiliary information (2022)	Florida State University, Mayo Clinic	United States	—
6	Covariate Adaptive False Discovery Rate Control With Applications to Omics-Wide Multiple Testing (2020)	Texas A&M University	United States	—
7	Spatial Difference Boundary Detection for Multiple Outcomes Using Bayesian Disease Mapping (2023)	University of California	United States	—
8	Multiple Hypothesis Testing Framework for Spatial Signals (2022)	Aalto University	Finland	**Influential**
9	ZAP:<i>Z</i>-Value Adaptive Procedures for False Discovery Rate Control with Side Information (2022)	Center for Data Science	China	—

Independent citing papers only; self- and co-author citations excluded. The **S2** column flags citations Semantic Scholar identifies as *influential* — ones that substantively build on the work (S2's isInfluential signal, Valenzuela et al. 2015) — the "built on / relied upon" pattern the AAO credits. Counsel should quote the citing text for the strongest of these.

D. Citing-Institution Prestige & Geography

Top citing institutions

Institution	Country	World ranking	Citing papers
University of Chicago	United States	SCImago #124 · THE 15 · QS 13	2
University of Southern California	United States	SCImago #192 · THE =73 · QS 146	2
Columbia University	United States	SCImago #65 · THE 20 · QS =38	2
University of Pennsylvania	United States	SCImago #52 · THE 14 · QS 15	2
Institute of Computer Science, Polish Academy of Sciences	Poland	—	1
Harvard Chan School of Public Health	United States	—	1
University of Cambridge	United Kingdom	SCImago #63 · THE =3 · QS 6	1
Shanghai University of Finance and Economics	China	SCImago #7841 · QS 1201-1400	1
Aalto University	Finland	SCImago #854 · THE =195 · QS =114	1
The George Washington University	United States	SCImago #832 · THE 201–250 · QS =358	1
European Molecular Biology Laboratory	Germany	—	1
Rice University	United States	SCImago #818 · THE =103 · QS =119	1
Northeastern University	United States	QS 384	1
Microsoft	United States	—	1
George Mason University	United States	SCImago #1399 · THE 401–500 · QS 951-1000	1

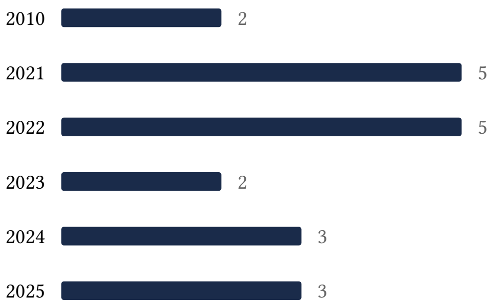
Geographic distribution of citing authors

Country	Citing papers
United States	15
China	6
United Kingdom	4
Germany	1
India	1
Italy	1
Netherlands	1
Poland	1
Australia	1
Vietnam	1
Canada	1
Finland	1

Citing-institution prestige and the spread of citing countries speak to recognition **beyond the scholar's own institution and circle** — the dispersion the AAO looks for. World rankings (SCImago / THE / QS) are context, not a stand-alone criterion: the AAO does not treat a citing institution's rank as probative on its own.

E. Citation Growth Over Time

Distinct citing papers by publication year. Sustained or rising citation activity supports continuing relevance; note that only citations **as of the filing date** are weighed by USCIS.



F. AAO Precedent Considerations

Pre-filing self-check (AAO denial patterns)

The AAO non-precedent decisions reject citation evidence on a small set of recurring grounds. Confirm the petition addresses each before filing:

- [ ] Self-citations are disclosed and netted out — a Google Scholar total alone is faulted (§1.1).
- [ ] Evidence is per individual article, not a body-of-work aggregate total (§1.2).
- [ ] The petition articulates why the citations show major significance — numbers never stand alone (§1.5).
- [ ] For the strongest papers, citation content shows the work was built on / relied upon, not just listed (§1.6, §2.2).
- [ ] Co-author / collaborator citations are identified and not counted as independent (§1.7).
- [ ] Recognition is shown beyond the scholar's own institution and circle (§1.8).
- [ ] Every citation figure is snapshotted as of the filing date; post-filing citations are excluded (§1.9).
- [ ] Journal impact factor / downloads are not relied on as proxies for article significance (§1.10, §1.12).
- [ ] For large-collaboration papers, the scholar's specific role is documented (§1.13).
- [ ] Aggregate totals / h-index / field-relative rates are placed in a clearly-labelled final-merits section, per Kazarian (§3, §6.1.7).

Disclaimer

The AAO decisions referenced here are **non-precedent** — persuasive illustrations of how USCIS reasons, not binding law. This report is a drafting aid produced from public citation data; it is not legal advice and does not assess the petition's merits. All analysis must be reviewed by qualified immigration counsel.

G. Citation Evidence Index

Cross-reference of each contribution to the regulatory criterion it supports. Counsel should map these to the petition's exhibit numbers.

Contribution	Core paper	Indep. cites	Supports
Contribution 1	The Holdout Randomization Test for Feature Selection in Black Box Models	6	8 CFR 204.5(i)(3) — Outstanding Researcher
Contribution 2	Annotation Refactoring: Inferring Upgrade Transformations for Legacy Applications	7	8 CFR 204.5(i)(3) — Outstanding Researcher
Contribution 3	False Discovery Rate Smoothing	9	8 CFR 204.5(i)(3) — Outstanding Researcher