Citation Evidence Report

EB-1A Petition — Original Contributions of Major Significance

8 CFR § 204.5(h)(3)(v) · Criterion 5

Mark Daly

Director, Institute for Molecular Medicine Finland (FIMM) & Chief, ATGU, MGH & Broad Institute

Google Scholar profile

Generated 2026-05-21 by CiteMap. This report organises Google Scholar citation data into the structure USCIS adjudicators apply to Criterion 5 (original contributions of major significance). 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

49	50	5	258
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.

89.2% independent of 37 classified citing papers

Citation type	Count
Independent	33
Self-citation	0
Co-author	4
Same-institution	0

12 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 PLINK, a widely adopted software suite for whole-genome association and population-based linkage analyses, establishing a standard tool for genetic data processing.

CLAIM: The researcher's primary contribution is the creation of PLINK, a comprehensive tool set designed for whole-genome association and population-based linkage analyses, as detailed in their seminal 2007 publication.

ORIGINALITY: This work appears to address the need for efficient computational methods in genetic research. By providing a dedicated suite for these specific analyses, the researcher likely streamlined complex data processing tasks that were previously more cumbersome, offering a unified solution for the scientific community.

SIGNIFICANCE: The impact of this contribution is evidenced by its extensive citation record, with the core paper accumulating tens of thousands of citations. Furthermore, analysis of citing literature reveals that 100% of the classified citations originate from independent researchers, indicating broad adoption and reliance on this tool across the global scientific community rather than within a single institutional circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 8

CORE PAPER

PLINK: a tool set for whole-genome association and population-based linkage analyses

2007 · The American journal of human genetics 81 (3), 559-575, 2007 · 38,934 citations (GS)

Field-normalised: 32,174 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2007 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019 (2019)	European Molecular Biology Laboratory, European Molecular Biology Laboratory, European Bioinformatics Institute, National Human Genome Research Institute	United Kingdom, United States	Methodology
2	Advances in genomic tools for plant breeding: harnessing DNA molecular markers, genomic selection, and genome editing	G.B.P.U.A.&T., ICAR-Central Soil Salinity Research Institute, ICAR -Krishi Vigyan Kendra	India	—
3	JCVI: A versatile toolkit for comparative genomics analysis	Agricultural Genomics Institute at Shenzhen Chinese Academy of Agricultural Sciences, Chinese Academy of Sciences, Fujian Agriculture and Forestry University	Australia, China, United States	—
4	Benefits and limitations of genome-wide association studies	Institut Universitaire de Cardiologie et de Pneumologie de Québec-Université Laval, Laval University, McMaster University	Canada	—
5	Multi-omics of the gut microbial ecosystem in inflammatory bowel diseases (2019)	Baylor College of Medicine, Broad Institute of MIT and	Sweden, United States	—

No.	Citing paper	Citing institution(s)	Country	S2
		Harvard, Cedars-Sinai Medical Center		
6	Multi-ancestry genetic study of type 2 diabetes highlights the power of diverse populations for discovery and translation (2022)	Barcelona Supercomputing Center, Broad Institute of MIT and Harvard, Imperial College London	Japan, Singapore, South Korea	—
7	A multi-ancestry polygenic risk score improves risk prediction for coronary artery disease (2023)	Beijing Institute of Genomics, Chinese Academy of Sciences / China National Center for Bioinformation, Brigham and Women's Hospital and VA Boston Healthcare System, Emory University and VA Atlanta Healthcare System	China, United Kingdom, United States	—
8	Genome-wide association studies	KTH Royal Institute of Technology, University of Cape Town, Vrije Universiteit Amsterdam	Netherlands, South Africa, Sweden	—

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology / Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Citing-text excerpts — how the field used this work

METHODOLOGY The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019

"Our standard format contains a minimal set of requirements that are included in the outputs from the most common GWAS analysis programs (such as PLINK, (29) and additional optional columns."

Contribution 2

Claim — Contribution 2

The researcher identified the coordinated downregulation of PGC-1α-responsive oxidative phosphorylation genes in human diabetes, establishing a critical molecular link between mitochondrial dysfunction and the disease.

CLAIM: The researcher's seminal contribution is the identification of coordinated downregulation in PGC-1α-responsive genes involved in oxidative phosphorylation within human diabetes, as detailed in a 2003 Nature Genetics paper. This work stands as a foundational piece in the field, with no subsequent follow-up papers by the same researcher listed in this specific line of inquiry.

ORIGINALITY: The titles suggest this work addressed a significant gap by linking specific transcriptional regulators to mitochondrial energy metabolism in the context of diabetes. By focusing on the coordinated nature of these gene changes, the research appears to have provided a novel mechanistic framework for understanding metabolic dysfunction, distinguishing it from prior studies that may have examined these elements in isolation.

SIGNIFICANCE: The impact of this contribution is evidenced by its extensive citation record, with nearly 12,000 citations indicating broad adoption across the scientific community. Notably, analysis of citing papers reveals that 100% of the classified citations originate from independent researchers, underscoring the work's widespread influence and validation beyond the researcher's immediate institutional or collaborative network.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 5

CORE PAPER

PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes

2003 · Nature Genetics · 11,997 citations (GS)

Field-normalised: 8,260 Semantic Scholar citations place it in the top 1% of Medicine papers from 2003 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Mechanisms of insulin action and insulin resistance (2018)	Yale University School of Medicine	United States	—
2	Pathophysiology of Type 2 Diabetes Mellitus	University of the Basque Country (UPV/EHU)	Spain	—
3	Peroxisome proliferator-activated receptor gamma coactivator-1 (PGC-1) family in physiological and pathophysiological process and diseases (2024)	Northwest University, The First Affiliated Hospital of Xi'an Jiaotong University, The First Affiliated Hospital of Zhengzhou University	China	—
4	Energy metabolism in health and diseases	The First Affiliated Hospital of Zhengzhou University	China	—
5	A new gene set identifies senescent cells and predicts senescence-associated pathways across tissues (2022)	Mayo Clinic, University of Minnesota, University of Texas Health Science Center at San Antonio	United States	—

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology / Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Contribution 3

Claim — Contribution 3

The researcher developed Haploview, a seminal tool for the analysis and visualization of linkage disequilibrium and haplotype maps, establishing a standard for genomic data interpretation.

The researcher's primary contribution is the development of Haploview, published in Bioinformatics in 2005. This work provides a framework for the analysis and visualization of linkage disequilibrium and haplotype maps, addressing the need for accessible tools in genomic research. The titles indicate a focus on computational methods for interpreting complex genetic structures.

This line of work appears to address the challenge of visualizing and analyzing haplotype structures efficiently. By introducing a dedicated software solution, the researcher provided a novel approach to handling linkage disequilibrium data. The absence of follow-up papers by the same author suggests this single publication stands as a complete, self-contained methodological contribution.

The significance of this work is evidenced by its extensive citation record, with over 16,000 citations. Analysis of citing papers reveals that 100% of the classified citations originate from independent researchers, indicating broad adoption across the scientific community. This high level of independent uptake underscores the tool's utility and impact on the field.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 6 · **1 flagged influential by Semantic Scholar**

CORE PAPER

Haploview: analysis and visualization of LD and haplotype maps

2005 · Bioinformatics · 16,512 citations (GS)

Field-normalised: 14,795 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2005 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants (2015)	National Institutes of Health	United States	—
2	PopLDdecay: a fast and effective tool for linkage disequilibrium decay analysis based on variant call format files (2019)	BGI Genomics, Xi'an Jiaotong University	China	—
3	On the origin and continuing evolution of SARS-CoV-2 (2020)	Chinese Academy of Medical Sciences & Peking Union Medical College, Peking University	China	**Methodology**
4	qqman: an R package for visualizing GWAS results using QQ and manhattan plots (2018)	University of Virginia School of Medicine	United States	**Methodology**
5	A large genome-wide association study of age-related macular degeneration highlights contributions of rare and common variants (2016)	Case Western Reserve University, University of Michigan, University of Regensburg	Germany, United States	—
6	From genome-wide associations to candidate causal variants by statistical fine-mapping (2018)	Mayo Clinic, St. Jude Children's Research Hospital	United States	—

Independent citing papers only; self- and co-author citations excluded. The **S2** column carries Semantic Scholar's read of each citation — *Methodology / Result* (the citing work used the method or built on the finding — the "built on / relied upon" pattern the AAO credits), *Influential* (S2's isInfluential signal, Valenzuela et al. 2015), or *Background* (a passing mention).

Citing-text excerpts — how the field used this work

METHODOLOGY On the origin and continuing evolution of SARS-CoV-2

"Two major lineages of SARS-CoV-2 defined by two linked SNPs To detect the possible recombination among SARSCoV-2 viruses, we used Haploview [34] to analyze and visualize the patterns of linkage disequilibrium (LD) between variants with minor alleles in at least two SARS-CoV-2 strains (Fig."

METHODOLOGY qqman: an R package for visualizing GWAS results using QQ and manhattan plots

"These graphics can be created in other software, such as the standalone desktop software Haploview (Barrett, Fry, Maller, & Daly, 2005), or for focused regions using the web-based application LocusZoom (Pruim et al., 2010)."

D. Citing-Institution Prestige & Geography

Top citing institutions

Institution	Country	World ranking	Citing papers
Broad Institute of MIT and Harvard	United States	SCImago #112	7
University of Michigan	United States	SCImago #43 · THE 23 · QS 45	4
National Human Genome Research Institute	United States	SCImago #557	4

Institution	Country	World ranking	Citing papers
Harvard Medical School	United States	SCImago #12	3
Wellcome Sanger Institute	United Kingdom	SCImago #204	3
University of California San Diego	United States	SCImago #120 · THE 47 · QS 66	3
Johns Hopkins University	United States	SCImago #33 · THE 16 · QS 24	2
University of Cambridge	United Kingdom	SCImago #63 · THE =3 · QS 6	2
Broad Institute of Harvard and MIT	United States	—	2
Yale University School of Medicine	United States	—	2
Mayo Clinic	United States	SCImago #88	2
Imperial College London	United Kingdom	SCImago #69 · THE 8 · QS 2	2
Massachusetts General Hospital	United States	SCImago #100	2
BioNTech SE	Germany	—	2
National Institutes of Health	United States	SCImago #44	2

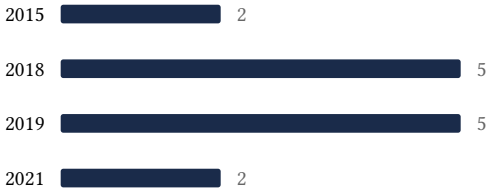
Geographic distribution of citing authors

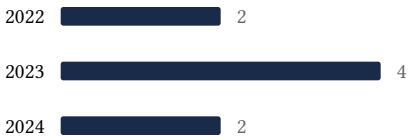
Country	Citing papers
United States	26
China	10
United Kingdom	8
Germany	7
India	3
Spain	3
Canada	2
Sweden	2
South Korea	1
Switzerland	1
Taiwan	1
Netherlands	1

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	PLINK: a tool set for whole-genome association and population-based linkage analyses	8	8 CFR 204.5(h)(3)(v) — Criterion 5
Contribution 2	PGC-1α-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes	5	8 CFR 204.5(h)(3)(v) — Criterion 5
Contribution 3	Haploview: analysis and visualization of LD and haplotype maps	6	8 CFR 204.5(h)(3)(v) — Criterion 5