Citation Evidence Report

EB-1A Petition — Original Contributions of Major Significance

8 CFR § 204.5(h)(3)(v) · Criterion 5

Simon Anders

Bioquant Center, University of Heidelberg

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

15	17	2	49
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.

86.7% independent of 15 classified citing papers

Citation type	Count
Independent	13
Self-citation	0
Co-author	2
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 DESeq2, a seminal statistical framework for RNA-seq data analysis that enables moderated estimation of fold change and dispersion, establishing a standard tool for the field.

The researcher's primary contribution is the development of DESeq2, introduced in a 2014 paper published in Genome Biology. This work provides a robust statistical method for analyzing RNA-seq data, specifically focusing on the moderated estimation of fold change and dispersion. The titles indicate a focus on improving the accuracy and reliability of differential expression analysis in high-throughput sequencing experiments.

This line of work appears to address the challenge of handling variability and noise inherent in RNA-seq datasets. By introducing a method for moderated estimation, the researcher likely aimed to improve the detection of true biological signals amidst technical variation. The absence of follow-up papers by the same researcher suggests that this single publication established a complete and enduring solution that did not require further iterative refinement by the original author.

The significance of this contribution is evidenced by its extensive adoption within the scientific community. With over 99,000 citations, the work has become a foundational resource in genomics. Furthermore, the high proportion of independent citations indicates that researchers across diverse institutions and collaborations rely on this tool, confirming its broad impact and utility beyond the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 10

CORE PAPER

Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2

2014 · Genome Biology · 99,107 citations (GS)

Field-normalised: 10,669 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2014 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Ferroptosis surveillance independent of GPX4 and differentially regulated by sex hormones (2023)	Memorial Sloan Kettering Cancer Center	United States	—
2	TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining (2023)	Henan University, Hunan Agricultural University, Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences	China	—
3	Next-Generation Sequencing Technology: Current Trends and Advancements (2023)	miBiome Therapeutics, UMass Chan Medical School	India, United States	—
4	Best practices for single-cell analysis across modalities (2023)	Helmholtz Center Munich, German Research Center for Environmental Health, Helmholtz Munich, Technical University of Munich	Germany	—
5	Organ aging signatures in the plasma proteome track health and disease (2023)	Stanford University	United States	—
6	Identification of mobile genetic elements with geNomad (2023)	Lawrence Berkeley National Laboratory, Los Alamos National Laboratory	United States	—

No.	Citing paper	Citing institution(s)	Country	S2
7	Safety, efficacy and determinants of response of allogeneic CD19-specific CAR-NK cells in CD19+ B cell tumors: a phase 1/2 trial (2024)	The University of Texas MD Anderson Cancer Center	United States	—
8	Using clusterProfiler to characterize multiomics data (2024)	Guangdong Academy of Sciences, Southern Medical University	China	—
9	TTD: Therapeutic Target Database describing target druggability information (2023)	Ningbo University, Tsinghua University, Yale University	China	—
10	APOE4/4 is linked to damaging lipid droplets in Alzheimer's disease microglia (2024)	Gladstone Institute of Neurological Disease, Linköping University, Stanford University	Sweden, 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 HTSeq, a widely adopted Python framework that standardized the processing of high-throughput sequencing data, significantly advancing bioinformatics workflows.

The researcher's primary contribution is the development of HTSeq, a Python framework designed to facilitate work with high-throughput sequencing data. This work, published in Bioinformatics in 2015, serves as the foundational element of this research line, with no subsequent follow-up papers by the same author extending the specific framework.

This line of work appears to address the need for accessible, standardized tools in bioinformatics. By providing a dedicated Python framework, the researcher likely simplified the complex computational tasks associated with analyzing high-throughput sequencing data, offering a practical solution for researchers lacking specialized programming expertise.

The significance of this contribution is evidenced by its substantial citation count of 23,390, indicating widespread adoption within the scientific community. Furthermore, analysis of citing papers reveals that 86.7% originate from independent researchers, suggesting that HTSeq has become a standard, widely utilized tool across diverse institutions and research groups rather than a niche or self-referential publication.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 5

CORE PAPER

HTSeq—a Python framework to work with high-throughput sequencing data

2015 · Bioinformatics · 23,390 citations (GS)

Field-normalised: 19,151 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2015 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Next-Generation Sequencing Technology: Current Trends and Advancements (2023)	miBiome Therapeutics, UMass Chan Medical School	India, United States	—
2	RNA sequencing: the teenage years (2019)	AstraZeneca, Cancer Research UK Cambridge Institute, University of Cambridge	United Kingdom	—

No.	Citing paper	Citing institution(s)	Country	S2
3	B cells and tertiary lymphoid structures promote immunotherapy response (2020)	Broad Institute of the Massachusetts Institute of Technology, INSERM, Cordeliers Research Center, Massachusetts General Hospital Cancer Center	France, Netherlands, United States	—
4	Using clusterProfiler to characterize multiomics data (2024)	Guangdong Academy of Sciences, Southern Medical University	China	—
5	The single-cell transcriptional landscape of mammalian organogenesis (2019)	Max Planck Institute for Molecular Genetics, University of Washington	Germany, 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.

D. Citing-Institution Prestige & Geography

Top citing institutions

Institution	Country	World ranking	Citing papers
Stanford University	United States	SCImago #18 · THE =5 · QS 3	2
The University of Texas MD Anderson Cancer Center	United States	—	2
AstraZeneca	United Kingdom	SCImago #244	1
Helmholtz Munich	Germany	—	1
University of Washington	United States	SCImago #45 · THE 25 · QS 81	1
Los Alamos National Laboratory	United States	SCImago #1704	1
UMass Chan Medical School	United States	SCImago #1179	1
Tsinghua University	China	SCImago #8 · THE 12 · QS =17	1
Yale University	United States	SCImago #76 · THE 10 · QS 21	1
Hunan Agricultural University	China	SCImago #2787	1
Henan University	China	SCImago #1369	1
Zhejiang University	China	SCImago #6 · THE 39 · QS 49	1
Guangdong Academy of Sciences	China	SCImago #3810	1
Lawrence Berkeley National Laboratory	United States	SCImago #530	1
Max Planck Institute for Molecular Genetics	Germany	SCImago #287	1

Geographic distribution of citing authors

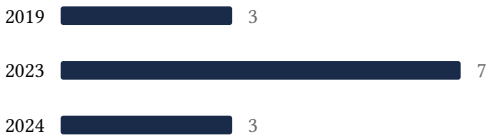
Country	Citing papers
United States	9
China	3
Australia	2

Country	Citing papers
Germany	2
India	1
Netherlands	1
Sweden	1
United Kingdom	1
France	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	Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2	10	8 CFR 204.5(h)(3)(v) — Criterion 5
Contribution 2	HTSeq—a Python framework to work with high-throughput sequencing data	5	8 CFR 204.5(h)(3)(v) — Criterion 5