Citation Evidence Report

EB-1A Petition — Original Contributions of Major Significance

8 CFR § 204.5(h)(3)(v) · Criterion 5

Wolfgang Huber

EMBL - European Molecular Biology Laboratory

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

14	15	3	111
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 14 classified citing papers

Citation type	Count
Independent	14
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 a foundational statistical framework for differential expression analysis in sequence count data, establishing a standard method widely adopted across the genomics community.

The researcher's primary contribution is the development of a robust statistical framework for differential expression analysis tailored to sequence count data, as detailed in the 2010 paper published in Genome Biology. This work addresses the critical need for accurate statistical methods capable of handling the unique distributional properties of high-throughput sequencing data, which differ significantly from traditional microarray data. By providing a specialized analytical approach, the researcher filled a methodological gap that was essential for the accurate interpretation of RNA-seq experiments during the rapid expansion of next-generation sequencing technologies. The significance of this contribution is evidenced by its extensive citation record, with over 18,000 citations indicating that the method has become a standard tool in the field. Furthermore, the fact that 100% of the classified citing papers originate from independent researchers underscores the broad, cross-institutional impact and widespread adoption of this framework by the global scientific community.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 5

CORE PAPER

Differential expression analysis for sequence count data

2010 · Genome Biology · 18,823 citations (GS)

No.	Citing paper	Citing institution(s)	Country	S2
1	Multivariable association discovery in population-scale meta-omics studies (2021)	Broad Institute of MIT and Harvard, CUNY Graduate School of Public Health and Health Policy, Genentech	United States	—
2	Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression (2019)	New York Genome Center	United States	—
3	Microbiome Datasets Are Compositional: And This Is Not Optional	Universitat de Girona, Universitat Politècnica de Catalunya, University of Western Ontario	Canada, Spain	—
4	Single-cell RNA sequencing technologies and bioinformatics pipelines (2018)	Kyung Hee University, Yonsei University	South Korea	—
5	The GTEx Consortium atlas of genetic regulatory effects across human tissues.	The Broad Institute of MIT and Harvard	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 2

Claim — Contribution 2

The researcher developed DESeq2, a seminal statistical framework for moderated estimation of fold change and dispersion in RNA-seq data, establishing a standard for differential expression analysis.

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

This line of work appears to address the need for more precise statistical modeling in RNA-seq analysis. By introducing moderated estimation techniques, the researcher likely aimed to reduce false positives and improve the detection of biologically significant changes. The absence of follow-up papers by the same researcher suggests that this single publication established a complete and enduring solution to the problem, rather than an iterative series of incremental improvements.

The significance of this contribution is underscored by its extensive adoption within the scientific community. With nearly 100,000 citations, the work has become a foundational tool in genomics. Furthermore, the fact that 100% of the classified citing papers originate from independent researchers demonstrates that the method has been widely validated and utilized across diverse institutions, confirming its broad impact and independence from the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 5

CORE PAPER

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

2014 · Genome Biology · 99,777 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	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	—
2	Next-Generation Sequencing Technology: Current Trends and Advancements	miBiome Therapeutics, UMass Chan Medical School	India, United States	—
3	Best practices for single-cell analysis across modalities	Helmholtz Center Munich, German Research Center for Environmental Health, Helmholtz Munich, Technical University of Munich	Germany	—
4	Organ aging signatures in the plasma proteome track health and disease	Stanford University	United States	—
5	Using clusterProfiler to characterize multiomics data (2024)	Guangdong Academy of Sciences, Southern Medical University	China	—

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 HTSeq, a Python framework for high-throughput sequencing data, establishing a widely adopted standard for bioinformatics analysis.

The researcher's primary contribution is the development of HTSeq, a Python framework designed to work with high-throughput sequencing data, as detailed in their 2015 publication in Bioinformatics. This work stands as a seminal core paper in the field, providing essential computational tools for genomic analysis.

This line of work appears to address the need for accessible and efficient methods to process complex sequencing data. By creating a dedicated Python framework, the researcher provided a standardized solution that likely simplified data handling tasks for the broader scientific community, filling a gap in available bioinformatics resources.

The significance of this contribution is evidenced by its extensive uptake, with the core paper accumulating over 23,000 citations. Furthermore, analysis of citing literature indicates that 100% of the classified citations originate from independent researchers, demonstrating that the work has been widely adopted and utilized by the global scientific community outside the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 5

CORE PAPER

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

2015 · Bioinformatics · 23,354 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	edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets	Genentech Inc, WEHI	Australia, United States	—
2	B cells and tertiary lymphoid structures promote immunotherapy response	Broad Institute of the Massachusetts Institute of Technology, INSERM, Cordeliers Research Center, Massachusetts General Hospital Cancer Center	France, Netherlands, United States	—
3	IL-10 constrains sphingolipid metabolism to limit inflammation (2024)	University of California, Los Angeles, Yale University, Yale University School of Medicine	United States	—
4	Genetics of circulating inflammatory proteins identifies drivers of immune-mediated disease risk and therapeutic targets (2023)	Consortium, Copenhagen University Hospital, German Diabetes Center	Denmark, Estonia, Germany	—
5	Using clusterProfiler to characterize multiomics data (2024)	Guangdong Academy of Sciences, Southern Medical University	China	—

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).

D. Citing-Institution Prestige & Geography

Top citing institutions

Institution	Country	World ranking	Citing papers
Helmholtz Munich	Germany	—	2
Yonsei University	South Korea	SCImago #238 · THE 86 · QS 50	1
University of Tartu	Estonia	SCImago #1820 · THE 301–350 · QS =362	1
Harvard T.H. Chan School of Public Health	United States	—	1
Universitat Politècnica de Catalunya	Spain	SCImago #624 · THE 601–800	1
Southern Medical University	China	SCImago #392 · THE 251–300	1
University of Cambridge	United Kingdom	SCImago #63 · THE =3 · QS 6	1
Copenhagen University Hospital	Denmark	SCImago #536	1
Uppsala University	Sweden	SCImago #349 · THE 128 · QS 93	1
German Diabetes Center	Germany	—	1
Universitat de Girona	Spain	SCImago #3154 · THE 1001–1200	1
Kyung Hee University	South Korea	SCImago #792 · THE 251–300 · QS =331	1
Yale University School of Medicine	United States	—	1
George Washington University	United States	SCImago #832 · THE 201–250 · QS =358	1
University of Pittsburgh	United States	SCImago #212 · QS =281	1

Geographic distribution of citing authors

Country	Citing papers
United States	8
China	2
Germany	2
Estonia	1
France	1
Australia	1
Netherlands	1
South Korea	1
Spain	1
Sweden	1
United Kingdom	1
India	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	Differential expression analysis for sequence count data	5	8 CFR 204.5(h)(3)(v) — Criterion 5
Contribution 2	Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2	5	8 CFR 204.5(h)(3)(v) — Criterion 5

Contribution	Core paper	Indep. cites	Supports
Contribution 3	HTSeq—a Python framework to work with high-throughput sequencing data	5	8 CFR 204.5(h)(3)(v) — Criterion 5