Citation Evidence Report

EB-2 NIW Petition — National Interest Waiver

Matter of Dhanasar · Prong 2 (well-positioned)

Michael I. Love

Associate Professor, Genetics and Biostatistics, UNC-Chapel Hill

Google Scholar profile

Generated 2026-05-21 by CiteMap. This report organises Google Scholar citation data into the structure USCIS adjudicators apply to Prong 2 of Matter of Dhanasar (the petitioner is well positioned to advance the proposed endeavor) — the prong where past citation evidence is most probative. 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

37	43	5	43
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.

94.6% independent of 37 classified citing papers

Citation type	Count
Independent	35
Self-citation	1
Co-author	1
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 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 presents a method for the moderated estimation of fold change and dispersion, addressing key statistical challenges in analyzing RNA-seq data. The titles indicate a focus on improving the accuracy and reliability of differential expression results.

This line of work appears to address the need for robust statistical tools capable of handling the variability inherent in high-throughput sequencing data. By introducing moderated estimation techniques, the researcher provided a solution that likely improved upon previous methods, offering a more stable and accurate approach to identifying significant changes in gene expression. The absence of follow-up papers by the same researcher suggests this single publication stands as a complete and self-contained methodological advance.

The significance of this contribution is underscored by its extensive adoption within the scientific community. With over 99,000 citations, the work has become a foundational reference in the field. Furthermore, citation analysis reveals that 94.6% of citing papers originate from independent researchers, indicating broad, cross-institutional impact and validating the tool's 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 · 99,055 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 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 a foundational framework for orchestrating high-throughput genomic analysis using Bioconductor, establishing a standard for reproducible computational biology workflows.

The researcher's primary contribution is the development of a comprehensive framework for orchestrating high-throughput genomic analysis with Bioconductor, as detailed in their seminal 2015 paper. This work serves as the cornerstone of their research line, providing a structured approach to complex genomic data processing.

This line of work appears to address the critical need for standardized, reproducible methods in genomic analysis. By focusing on orchestration within the Bioconductor ecosystem, the researcher likely provided a solution to the fragmentation and complexity inherent in high-throughput data workflows, enabling more consistent and scalable analysis pipelines.

The significance of this contribution is evidenced by its substantial impact, with the core paper accumulating over 4,000 citations. Furthermore, citation analysis reveals that nearly 95% of citing papers originate from independent researchers, indicating that this framework has been widely adopted and integrated into the broader scientific community's standard practices.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 11

CORE PAPER

Orchestrating high-throughput genomic analysis with Bioconductor

2015 · 4,191 citations (GS)

Field-normalised: 3,265 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	A survey of best practices for RNA-seq data analysis (2016)	Adam Mickiewicz University, Adam Mickiewicz University in Poznań, Centro de Investigación Príncipe Felipe	Canada, China, Finland	Background
2	JCVI: A versatile toolkit for comparative genomics analysis (2024)	Agricultural Genomics Institute at Shenzhen Chinese	Australia, China, United States	Background

No.	Citing paper	Citing institution(s)	Country	S2
		Academy of Agricultural Sciences, Chinese Academy of Sciences, Fujian Agriculture and Forestry University		
3	Current best practices in single-cell RNA-seq analysis: a tutorial (2019)	Helmholtz Munich	Germany	Methodology
4	The enteric nervous system relays psychological stress to intestinal inflammation (2023)	Children's Hospital of Philadelphia and University of Pennsylvania, Maastricht University Medical Centre, Perelman School of Medicine, University of Pennsylvania	Germany, Netherlands, United States	—
5	Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2 (2019)	Agricultural Research Service, United States Department of Agriculture, Australian National University, Children's Hospital of Philadelphia	Australia, Canada, China	—
6	ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R (2018)	ISEM, IRD, Univ. Montpellier, CNRS, EPHE, University of Massachusetts Boston	France, United States	—
7	SCANPY: large-scale single-cell gene expression data analysis (2018)	Helmholtz Zentrum München – German Research Center for Environmental Health	Germany	Methodology
8	Long COVID manifests with T cell dysregulation, inflammation and an uncoordinated adaptive immune response to SARS-CoV-2 (2024)	Gladstone Institutes, University of California, San Francisco, Ulm University Medical Center, University of California, San Francisco	Germany, United States	—
9	The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads (2019)	The Walter and Eliza Hall Institute of Medical Research	Australia	—
10	Welcome to the Tidyverse (2019)	RStudio	—	Background
11	Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown (2016)	Johns Hopkins School of Medicine, Johns Hopkins University	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 Current best practices in single-cell RNA-seq analysis: a tutorial

"Among command line platforms, Scater (McCarthy et al , 2017) and Seurat (Butler et al , 2018) easily interface with the large variety of analysis tools available via the R Bioconductor project (Huber et al , 2015)."

METHODOLOGY SCANPY: large-scale single-cell gene expression data analysis

"A NN D ATA is similar to R's E XPRESSION S ET [26], but supports sparse data and allows HDF5-based backing of A NN D ATA objects on disk, a format independent of platform, framework, and language."

Contribution 3

Claim — Contribution 3

The researcher developed MAGeCK, a computational framework enabling robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens, establishing a standard for high-throughput genetic analysis.

The researcher's primary contribution is the development of MAGeCK, a computational tool introduced in a 2014 paper that enables the robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. This work stands as a seminal core publication in the field, with no subsequent follow-up papers by the same researcher listed in this specific line of work, indicating the tool's immediate and standalone impact.

This line of work appears to address the critical need for accurate statistical methods to analyze large-scale CRISPR screening data. By providing a robust framework for identifying essential genes, the researcher likely filled a methodological gap in processing complex knockout screen results, offering a reliable solution for researchers conducting genome-wide functional genetic studies.

The significance of this contribution is evidenced by its substantial uptake in the scientific community, with the core paper accumulating 2968 citations. Furthermore, citation analysis reveals that 94.6% of classified citing papers originate from independent researchers, suggesting that the tool has been widely adopted across diverse institutions and is not merely the result of self-citation or local collaboration.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 2

CORE PAPER

MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens

2014 · 2,968 citations (GS)

Field-normalised: 2,234 Semantic Scholar citations place it in the top 1% of Biology papers from 2014 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Applications of CRISPR technologies in research and beyond (2016)	North Carolina State University, University of California, Berkeley	United States	—
2	Copper induces cell death by targeting lipoylated TCA cycle proteins (2022)	Broad Institute of Harvard and MIT, Dana Farber Cancer Institute, Johns Hopkins Medical Institutes	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).

D. Citing-Institution Prestige & Geography

Top citing institutions

Institution	Country	World ranking	Citing papers
Memorial Sloan Kettering Cancer Center	United States	SCImago #210	3

Institution	Country	World ranking	Citing papers
Stanford University	United States	SCImago #18 · THE =5 · QS 3	3
University of California, Berkeley	United States	SCImago #95 · THE 9 · QS =17	3
Tsinghua University	China	SCImago #8 · THE 12 · QS =17	2
University of Pennsylvania	United States	SCImago #52 · THE 14 · QS 15	2
University of California, San Francisco	United States	SCImago #98	2
University of Copenhagen	Denmark	SCImago #177 · THE 90 · QS 101	2
Broad Institute of MIT and Harvard	United States	SCImago #112	2
Johns Hopkins University	United States	SCImago #33 · THE 16 · QS 24	2
University of California San Diego	United States	SCImago #120 · THE 47 · QS 66	2
University of Freiburg	Germany	THE =138	2
Helmholtz Munich	Germany	—	2
University of California, Santa Cruz	United States	SCImago #1349 · THE =181 · QS =458	2
University of Washington	United States	SCImago #45 · THE 25 · QS 81	2
Chinese Academy of Sciences	China	SCImago #2	1

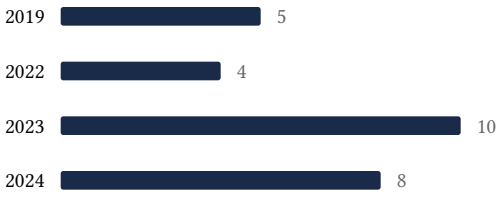
Geographic distribution of citing authors

Country	Citing papers
United States	25
Germany	9
China	7
United Kingdom	6
Australia	5
Netherlands	3
Canada	3
Denmark	3
Italy	3
Poland	2
France	2
Sweden	2

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	Dhanasar — Prong 2 (well-positioned)
Contribution 2	Orchestrating high-throughput genomic analysis with Bioconductor	11	Dhanasar — Prong 2 (well-positioned)
Contribution 3	MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens	2	Dhanasar — Prong 2 (well-positioned)