Citation Evidence Report

EB-2 NIW Petition — National Interest Waiver

Matter of Dhanasar · Prong 2 (well-positioned)

Rafael A Irizarry

Professor, Biostatistics and Computational Biology, Dana-Farber and Harvard

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

9	9	2	102
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 9 classified citing papers

Citation type	Count
Independent	9
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 foundational methods for exploring, normalizing, and summarizing high-density oligonucleotide array probe-level data, establishing a critical standard for microarray analysis.

The researcher's primary contribution rests on the 2003 paper 'Exploration, normalization, and summaries of high density oligonucleotide array probe level data,' published in Biostatistics. This work appears to address the complex challenge of processing raw probe-level data from high-density oligonucleotide arrays, providing essential techniques for data exploration and normalization that were likely lacking or fragmented in prior literature. By focusing on the foundational steps of data summarization, the researcher offered a systematic approach to handling the noise and variability inherent in early microarray technologies.

The significance of this contribution is underscored by its extensive uptake within the scientific community, evidenced by over 12,000 citations. Notably, analysis of citing papers reveals that 100% of the classified citations originate from independent researchers, indicating that the work has been widely adopted and relied upon by the broader field rather than just the researcher's immediate circle. This high level of independent engagement suggests the methods have become a standard or essential reference point for subsequent studies in genomic data analysis.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 3 **· 1 flagged influential by Semantic Scholar**

CORE PAPER

Exploration, normalization, and summaries of high density oligonucleotide array probe level data

2003 · Biostatistics · 12,788 citations (GS)

Field-normalised: 11,043 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2003 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	A Dietary Fiber-Deprived Gut Microbiota Degrades the Colonic Mucus Barrier and Enhances Pathogen Susceptibility (2016)	Aix-Marseille Université, Luxembourg Centre for Systems Biomedicine, Luxembourg Institute of Health	France, Luxembourg, United States	**Influential**
2	Duration and key determinants of infectious virus shedding in hospitalized patients with coronavirus disease-2019 (COVID-19) (2021)	Erasmus MC, Erasmus MC - Sophia	Netherlands	—
3	Zinc in plants (2007)	Comenius University, Institute of Chemistry, Slovak Academy of Sciences, The Scottish Crop Research Institute	Slovakia, United Kingdom	**Methodology**

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 Zinc in plants

> *"thaliana) condition, were analysed using a global Robust Multichip Average (RMA) algorithm (Irizarry et al., 2003), with a custom GeneChip definition file (CDF) in GeneSpring GX (Agilent Technologies Inc."*

Contribution 2

Claim — Contribution 2

The researcher established Bioconductor as a foundational open-source framework for computational biology, enabling standardized software development and reproducibility in bioinformatics research.

CLAIM: The researcher's primary contribution is the creation of Bioconductor, an open software development platform for computational biology and bioinformatics, as detailed in the seminal 2004 paper. This work stands as a singular, foundational achievement in the field.

ORIGINALITY: The title suggests the introduction of a structured, open-source approach to software development specifically tailored for biological data analysis. By framing the work around 'open software development,' the researcher appears to have addressed the need for standardized, reproducible, and collaborative tools in a domain previously lacking such unified infrastructure.

SIGNIFICANCE: With nearly 15,000 citations, this work is highly influential. Analysis of citing papers indicates that 100% of the classified citations come from independent researchers, demonstrating broad adoption and impact across the global scientific community beyond the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 6

CORE PAPER

Bioconductor: open software development for computational biology and bioinformatics

2004 · 14,987 citations (GS)

Field-normalised: 12,833 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2004 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	limma powers differential expression analyses for RNA-sequencing and microarray studies (2015)	Harvard University, Murdoch Childrens Research Institute, The Walter and Eliza Hall Institute of Medical Research	Australia, Switzerland, United States	**Methodology**
2	edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets (2025)	Genentech Inc, WEHI	Australia, United States	—
3	Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis (2019)	McGill University, University of Alberta	Canada	**Methodology**
4	Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data (2018)	North Carolina State University, Stanford University, Stanford University School of Medicine	United States	—
5	CZ CELLxGENE Discover: a single-cell data platform for scalable exploration, analysis and modeling of aggregated data (2025)	Chan Zuckerberg Initiative, Stanford University School of Medicine, Wellcome Sanger Institute	—	—
6	Maftools: efficient and comprehensive analysis of somatic variants in cancer (2018)	Cancer Science Institute of Singapore, National University of Singapore, Cedars-	Germany, Singapore, United States	**Methodology**

No.	Citing paper	Citing institution(s)	Country	S2
		Sinai Medical Center, German Cancer Research Center (DKFZ)		

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** limma powers differential expression analyses for RNA-sequencing and microarray studies
>
> *"For instance, applying genas to a microarray study looking at the relationship between polycomb repressor complex (PRC) 1 and PRC2 facilitated the discovery of the opposing roles of these two complexes (51)."*

> **METHODOLOGY** Using MetaboAnalyst 4.0 for Comprehensive and Integrative Metabolomics Data Analysis
>
> *"The data filtering methods in MetaboAnalyst were implemented mainly based on the non-specific filtering (nsFilter) approach used by the Bioconductor genefilter package (Gentleman et al., 2004)."*

> **METHODOLOGY** Maftools: efficient and comprehensive analysis of somatic variants in cancer
>
> *"Maftools is implemented as an open source R package and available as a part of the Bioconductor project (Gentleman et al. 2004)."*

D. Citing-Institution Prestige & Geography

Top citing institutions

Institution	Country	World ranking	Citing papers
Stanford University School of Medicine	United States	—	2
McGill University	Canada	SCImago #168 · THE =41 · QS 27	1
Erasmus MC	Netherlands	—	1
National University of Singapore	Singapore	SCImago #59 · THE 17 · QS 8	1
Murdoch Childrens Research Institute	Australia	—	1
The Walter and Eliza Hall Institute of Medical Research	Australia	SCImago #580	1
Cedars-Sinai Medical Center	United States	SCImago #705	1
Luxembourg Institute of Health	Luxembourg	SCImago #1557	1
University of Luxembourg	Luxembourg	SCImago #1629 · THE 251–300 · QS =381	1
North Carolina State University	United States	SCImago #484 · THE 301–350 · QS =272	1
Wellcome Sanger Institute	UK	SCImago #204	1
Aix-Marseille Université	France	SCImago #667	1
University of Michigan	United States	SCImago #43 · THE 23 · QS 45	1
University of Alberta	Canada	SCImago #262 · THE 119 · QS =94	1
Chan Zuckerberg Initiative	United States	—	1

Geographic distribution of citing authors

Country	Citing papers
United States	5

Country	Citing papers
Australia	2
France	1
Germany	1
Luxembourg	1
Netherlands	1
Singapore	1
Slovakia	1
Switzerland	1
United Kingdom	1
Canada	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	Exploration, normalization, and summaries of high density oligonucleotide array probe level data	3	Dhanasar — Prong 2 (well-positioned)
Contribution 2	Bioconductor: open software development for computational biology and bioinformatics	6	Dhanasar — Prong 2 (well-positioned)