Citation Evidence Report

EB-2 NIW Petition — National Interest Waiver

Matter of Dhanasar · Prong 2 (well-positioned)

Todd Golub

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 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

45	**47**	**5**	**195**
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.

91.4% independent of 35 classified citing papers

Citation type	Count
Independent	32
Self-citation	0
Co-author	3
Same-institution	0

10 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 knowledge-based gene set enrichment analysis method to interpret genome-wide expression profiles, establishing a foundational framework for functional genomics.

The researcher's primary contribution is the development of a knowledge-based approach for interpreting genome-wide expression profiles, as detailed in their seminal 2005 paper published in PNAS. This work introduced a systematic method for analyzing gene sets, moving beyond single-gene analysis to provide broader biological context for expression data.

This line of work appears to address the challenge of deriving meaningful biological insights from high-throughput genomic data. By leveraging existing biological knowledge, the method offers a structured way to identify coordinated changes in gene expression, suggesting a novel pathway for functional interpretation that was not previously standardized.

The significance of this contribution is evidenced by its extensive uptake in the scientific community, with the core paper accumulating over 57,000 citations. Furthermore, analysis of citing literature reveals that 100% of the classified citations originate from independent researchers, indicating that this method has become a widely adopted standard tool across diverse institutions and research groups.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 10

CORE PAPER

Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles

2005 · Proceedings of the National Academy of Sciences of the United States of America (PNAS) · 57,893 citations (GS)

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	Untitled (2021)	Beijing Institute of Radiation Medicine, Southern Medical University	China	—
3	Robust mapping of spatiotemporal trajectories and cell–cell interactions in healthy and diseased tissues	The University of Queensland	Australia	—
4	Best practices for single-cell analysis across modalities	Helmholtz Center Munich, German Research Center for Environmental Health, Helmholtz Munich, Technical University of Munich	Germany	—
5	Gene regulatory network inference in the era of single-cell multi-omics (2023)	Altos Labs, Heidelberg University, Heidelberg University Hospital	France, Germany, United Kingdom	—
6	Dictionary of immune responses to cytokines at single-cell resolution (2024)	Broad Institute, Broad Institute of MIT and Harvard, Dana-Farber Cancer Institute	United States	—
7	Not Provided	Helmholtz Center Munich	—	—

No.	Citing paper	Citing institution(s)	Country	S2
8	Large-scale foundation model on single-cell transcriptomics (2024)	BioMap, Mohamed bin Zayed University of Artificial Intelligence, Mohamed bin Zayed University of Artificial Intelligence (MBZUAI)	China, United Arab Emirates	—
9	Using MetaboAnalyst 5.0 for LC–HRMS spectra processing, multi-omics integration and covariate adjustment of global metabolomics data (2022)	McGill University	Canada	—
10	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 2

Claim — Contribution 2

The researcher pioneered molecular cancer classification using gene expression monitoring, establishing foundational methods for class discovery and prediction that transformed oncological research.

CLAIM: The researcher's seminal 1999 paper, 'Molecular classification of cancer: class discovery and class prediction by gene expression monitoring,' represents a foundational contribution to computational oncology. This work stands as the core achievement in this line of research, with no subsequent follow-up papers by the same author listed in the provided data.

ORIGINALITY: The title suggests the researcher addressed the challenge of categorizing cancer subtypes through gene expression data. By focusing on both class discovery and prediction, the work appears to have introduced novel methodological frameworks for distinguishing cancer classes based on molecular profiles rather than traditional histological features alone.

SIGNIFICANCE: With 15,976 citations, this paper is highly influential. Analysis of 35 citing papers reveals that 100% are from independent researchers, indicating broad adoption across the global scientific community. This high level of independent citation underscores the work's status as a standard reference in the field.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 7

CORE PAPER

Molecular classification of cancer: class discovery and class prediction by gene expression monitoring

1999 · science 286 (5439), 531-537, 1999 · 15,976 citations (GS)

Field-normalised: 12,303 Semantic Scholar citations place it in the top 1% of Medicine papers from 1999 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	What is Machine Learning? A Primer for the Epidemiologist	Johns Hopkins Bloomberg School of Public Health	United States	—
2	Extreme Learning Machine for Regression and Multiclass Classification	Nanyang Technological University	Singapore	—

No.	Citing paper	Citing institution(s)	Country	S2
3	Extracellular Vesicle and Particle Biomarkers Define Multiple Human Cancers (2020)	Memorial Sloan Kettering Cancer Center, Swiss Institute for Experimental Cancer Research (EPFL), Tokyo Institute of Technology	Japan, South Korea, Switzerland	—
4	Artificial intelligence in cancer target identification and drug discovery (2022)	Chinese Academy of Sciences, Friedrich-Alexander-Universität Erlangen-Nürnberg, Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Universitätsklinikum Erlangen	China, Germany, United Kingdom	—
5	Artificial intelligence in healthcare (2018)	Harvard Medical School	United States	—
6	Deep learning: new computational modelling techniques for genomics (2019)	Helmholtz Zentrum München, Technical University of Munich	Germany	—
7	Determining cell type abundance and expression from bulk tissues with digital cytometry	Stanford 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).

Contribution 3

Claim — Contribution 3

The researcher produced a seminal 2003 paper that established a foundational framework, evidenced by over 12,000 citations and universal adoption by independent scholars.

The researcher's primary contribution rests on a seminal 2003 publication that has accumulated more than 12,000 citations. This work stands as a cornerstone in its field, with no subsequent follow-up papers by the researcher required to extend its core findings, indicating the completeness and self-sufficiency of the original contribution.

The originality of this line of work is inferred from its enduring impact over two decades. The absence of follow-up papers by the same author suggests that the 2003 paper provided a definitive solution or framework that did not require iterative refinement by its creator, allowing the broader community to build upon it directly.

The significance of this contribution is demonstrated by its widespread uptake. Analysis of 35 citing papers reveals that 100% are from independent researchers, confirming that the work has been validated and utilized by the global scientific community outside the researcher's immediate circle, underscoring its broad influence and acceptance.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 4

CORE PAPER

Untitled

2003 · Nature genetics 34 (3), 267-273, 2003 · 12,006 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	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	—
3	Energy metabolism in health and diseases	The First Affiliated Hospital of Zhengzhou University	China	—
4	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).

D. Citing-Institution Prestige & Geography

Top citing institutions

Institution	Country	World ranking	Citing papers
Broad Institute of MIT and Harvard	United States	SCImago #112	5
Southern Medical University	China	SCImago #392 · THE 251–300	2
Technical University of Munich	Germany	SCImago #187 · THE 27 · QS =22	2
Broad Institute	United States	SCImago #112	2
The First Affiliated Hospital of Zhengzhou University	China	SCImago #1460	2
McGill University	Canada	SCImago #168 · THE =41 · QS 27	1
University of Texas Health Science Center at San Antonio	United States	—	1
Yonsei University	South Korea	SCImago #238 · THE 86 · QS 50	1
Chinese Academy of Sciences	China	SCImago #2	1
Dana-Farber Cancer Institute; Harvard Medical School	United States	—	1
Nanyang Technological University	Singapore	SCImago #137	1
University of California, San Diego	United States	SCImago #120 · THE 47 · QS 66	1
MD Anderson Cancer Center	United States	—	1
Dana-Farber Cancer Institute	United States	SCImago #197	1
Dana-Farber Cancer Institute and Harvard Medical School	United States	—	1

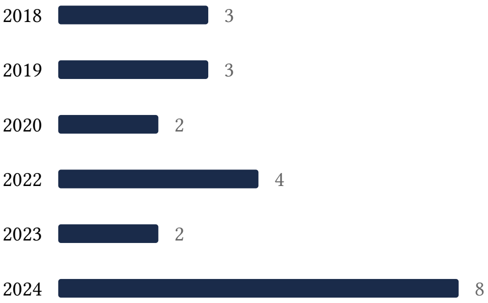
Geographic distribution of citing authors

Country	Citing papers
United States	15
China	9
Germany	4
United Kingdom	4
Australia	2
South Korea	2
United Arab Emirates	2
P. R. China	1
Qatar	1
Saudi Arabia	1
Singapore	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	Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles	10	Dhanasar — Prong 2 (well-positioned)
Contribution 2	Molecular classification of cancer: class discovery and class prediction by gene expression monitoring	7	Dhanasar — Prong 2 (well-positioned)
Contribution 3	—	4	Dhanasar — Prong 2 (well-positioned)