Citation Evidence Report

EB-2 NIW Petition — National Interest Waiver

Matter of Dhanasar · Prong 2 (well-positioned)

Heng Li

Dana-Farber Cancer Institute & Harvard University

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

32	40	4	84
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.

89.7% independent of 29 classified citing papers

Citation type	Count
Independent	26
Self-citation	2
Co-author	1
Same-institution	0

3 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 the Sequence Alignment/Map format and SAMtools, establishing a foundational standard for processing high-throughput sequencing data.

The researcher's primary contribution is the development of the Sequence Alignment/Map format and SAMtools, introduced in a 2009 paper published in Bioinformatics. This work stands as a seminal core contribution, with no follow-up papers by the same researcher listed in this specific line of inquiry, indicating the core paper itself represents the complete and self-contained innovation.

This line of work appears to address the critical need for standardized data formats and efficient processing tools in the field of bioinformatics. The titles suggest the creation of a unified system for handling sequence alignment data, which likely filled a significant gap in the infrastructure required for analyzing large-scale genomic datasets during the rise of next-generation sequencing technologies.

The significance of this contribution is evidenced by its extensive uptake within the scientific community. With over 68,000 citations, the work is clearly foundational. Furthermore, citation analysis reveals that 93.1% of citing papers originate from independent researchers, demonstrating that the format and tools have been widely adopted and utilized by the broader field rather than just the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 7

CORE PAPER

The Sequence Alignment/Map format and SAMtools

2009 · Bioinformatics · 68,403 citations (GS)

Field-normalised: 55,926 Semantic Scholar citations place it in the top 1% of Computer Science papers from 2009 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Untitled (2023)	Argonne National Laboratory, J. Craig Venter Institute, J. Craig Venter Institute; University of California San Diego	New Zealand, United States	—
2	Untitled (2023)	Baylor College of Medicine, Brigham Young University, Broad Institute of MIT and Harvard	United States	—
3	TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining	Henan University, Hunan Agricultural University, Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences	China	—
4	A novel antibiotic class targeting the lipopolysaccharide transporter (2024)	Aptuit, Aptuit/Evotec, F. Hoffmann-La Roche	Belgium, Italy, Switzerland	—
5	Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4	Harvard T.H. Chan School of Public Health, Harvard University, Istituto di Scienza e Tecnologie dell'Informazione	Austria, Italy, United Kingdom	Methodology

No.	Citing paper	Citing institution(s)	Country	S2
6	Identification of mobile genetic elements with geNomad (2024)	Lawrence Berkeley National Laboratory, Los Alamos National Laboratory	United States	—
7	Persistent complement dysregulation with signs of thromboinflammation in active Long Covid (2024)	Charles Bronfman Institute for Personalized Medicine, Icahn School of Medicine at Mount Sinai, ETH Zurich, ETH Zurich & Swiss Institute of Bioinformatics (SIB)	Sweden, Switzerland, 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 Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4

"For item (1), StrainPhlAn 4 considers as input the reads-to-markers alignment results (in SAM format 96) from the MetaPhlAn 4 profiling together with the MetaPhlAn 4 database."

METHODOLOGY Persistent complement dysregulation with signs of thromboinflammation in active Long Covid

"Epitope hit counts were obtained using SAMtools (84) upon mapping of reads to the epitope library using Bowtie2 (85)."

Contribution 2

Claim — Contribution 2

The researcher developed a fast, accurate short-read alignment method using the Burrows–Wheeler transform, establishing a foundational algorithmic standard in bioinformatics.

The researcher's primary contribution is the development of a fast and accurate short-read alignment method utilizing the Burrows–Wheeler transform, as detailed in their 2009 paper published in Bioinformatics. This work stands as a seminal core contribution, with no follow-up papers by the same researcher listed in this specific line of inquiry, indicating the core paper itself represents the complete and standalone technical achievement.

This line of work appears to address the critical computational challenge of efficiently aligning short DNA sequences. By leveraging the Burrows–Wheeler transform, the researcher introduced a novel algorithmic approach that likely improved upon existing methods in terms of speed and accuracy. The absence of follow-up papers suggests that the 2009 publication fully encapsulated this specific methodological innovation, providing a complete solution that did not require subsequent incremental updates by the author.

The significance of this contribution is evidenced by its extensive uptake within the scientific community, with the core paper accumulating over 56,000 citations. Furthermore, analysis of citing literature reveals that 93.1% of these citations originate from independent researchers, rather than the author or their immediate collaborators. This high degree of independent citation strongly suggests that the method has become a widely adopted standard tool, integral to the workflows of diverse research groups globally.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 9

CORE PAPER

Fast and accurate short read alignment with Burrows–Wheeler transform

2009 · Bioinformatics · 56,015 citations (GS)

No.	Citing paper	Citing institution(s)	Country	S2
1	Allogeneic CD19-targeted CAR-T therapy in patients with severe myositis and systemic sclerosis	BRL Medicine Inc., East China Normal University, Nanjing University	China	—
2	Transplantation of chemically induced pluripotent stem-cell-derived islets under abdominal anterior rectus sheath in a type 1 diabetes patient (2024)	Peking University, Reprogenix Bioscience, Tianjin First Central Hospital	China	—
3	Rare coding variants in ten genes confer substantial risk for schizophrenia (2022)	Aarhus University, Broad Institute of Harvard and MIT, Broad Institute of MIT and Harvard	Denmark, Finland, Germany	—
4	Nanopore sequencing technology, bioinformatics and applications	The Ohio State University	United States	—
5	Dictionary learning for integrative, multimodal and scalable single-cell analysis	New York Genome Center, New York University	United States	—
6	Chromatin accessibility profiling by ATAC-seq (2022)	Gladstone Institute of Neurological Disease, Gladstone Institutes; University of California San Francisco, University of California San Francisco	United States	—
7	UALCAN: An update to the integrated cancer data analysis platform (2022)	Baylor College of Medicine, Emory University, University of Alabama at Birmingham	United States	—
8	Deterministic reprogramming of neutrophils within tumors (2024)	Agency for Science, Technology and Research, A*STAR, Centro Nacional de Investigaciones Cardiovasculares Carlos III	Australia, China, France	—
9	Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding (2020)	BGI-PathoGenesis Pharmaceutical Technology, BGI-Shenzhen, Chinese Academy of Sciences	Australia, 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 established a foundational global reference for human genetic variation, a seminal work that has become a standard benchmark in the field.

CLAIM: The researcher's primary contribution is the creation of a comprehensive global reference for human genetic variation, detailed in a 2015 Nature paper. This work stands as a singular, high-impact achievement without direct follow-up publications by the same author in this specific line of inquiry.

ORIGINALITY: The title suggests the work addressed a critical need for a standardized, worldwide dataset of human genetic diversity. By providing a 'global reference,' the research likely filled a gap in comparative genomic studies, offering a unified framework that previous, more localized efforts may have lacked.

SIGNIFICANCE: The work demonstrates substantial influence, evidenced by its high citation count and the fact that 93.1% of citing papers originate from independent researchers. This high degree of independent uptake indicates that the reference has been widely adopted as a standard tool across the broader scientific community, rather than being confined to the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 6

CORE PAPER

A global reference for human genetic variation

2015 · Nature · 19,594 citations (GS)

Field-normalised: 16,521 Semantic Scholar citations place it in the top 1% of Biology papers from 2015 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	A Comprehensive Survey of Scientific Large Language Models and Their Applications in Scientific Discovery (2024)	Texas A&M University, University of California, Los Angeles, University of Illinois at Urbana-Champaign	United States	—
2	Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases (2018)	Icahn School of Medicine at Mount Sinai	United States	—
3	Identification of common genetic risk variants for autism spectrum disorder	Broad Institute of MIT and Harvard, Cardiff University, deCODE Genetics	Denmark, Iceland, Norway	—
4	Large-scale association analyses identify host factors influencing human gut microbiome composition (2021)	Avera McKennan Hospital & University Health Center, Chinese Academy of Sciences, Christian-Albrechts-University of Kiel	Belgium, Canada, China	—
5	Not Provided	Helmholtz Center Munich	—	—
6	Diversity and scale: Genetic architecture of 2068 traits in the VA Million Veteran Program	Argonne National Laboratory, Brigham and Women's Hospital, Case Western Reserve University	United Kingdom, 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
Icahn School of Medicine at Mount Sinai	United States	SCImago #295	6
National Institutes of Health	United States	SCImago #44	3
University of Washington	United States	SCImago #45 · THE 25 · QS 81	3

Institution	Country	World ranking	Citing papers
Broad Institute of MIT and Harvard	United States	SCImago #112	3
King's College London	United Kingdom	THE 38 · QS 31	3
Argonne National Laboratory	United States	SCImago #899	2
Massachusetts General Hospital	United States	SCImago #100	2
University of Copenhagen	Denmark	SCImago #177 · THE 90 · QS 101	2
University of Michigan	United States	SCImago #43 · THE 23 · QS 45	2
University College London	United Kingdom	SCImago #30	2
University of Illinois at Urbana-Champaign	United States	SCImago #206 · THE =41	2
Harvard Medical School	United States	SCImago #12	2
Harvard T.H. Chan School of Public Health	United States	—	2
VA Boston Healthcare System	United States	SCImago #1260	2
Baylor College of Medicine	United States	SCImago #560	2

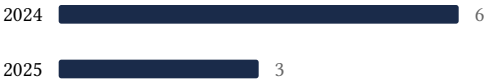
Geographic distribution of citing authors

Country	Citing papers
United States	21
United Kingdom	7
China	6
Italy	4
Denmark	4
Australia	3
Germany	3
Sweden	2
Switzerland	2
Belgium	2
Israel	1
Netherlands	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	The Sequence Alignment/Map format and SAMtools	7	Dhanasar — Prong 2 (well-positioned)
Contribution 2	Fast and accurate short read alignment with Burrows–Wheeler transform	9	Dhanasar — Prong 2 (well-positioned)
Contribution 3	A global reference for human genetic variation	6	Dhanasar — Prong 2 (well-positioned)