

# Citation Evidence Report

EB-1A Petition — Original Contributions of Major Significance

8 CFR § 204.5(h)(3)(v) · Criterion 5

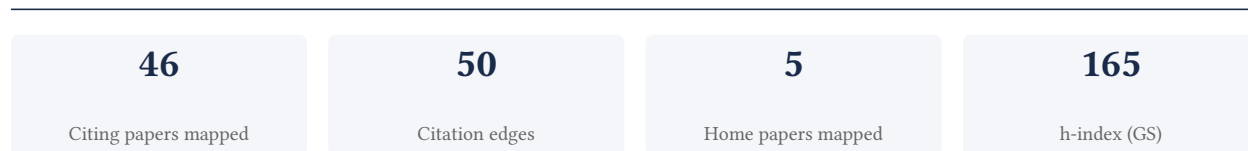
## Richard Durbin

Dept of Genetics, University of Cambridge

[Google Scholar profile](#)

**Generated 2026-05-22 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



### 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.

**80.5% independent** of 41 classified citing papers

Citation type	Count
Independent	33
Self-citation	2
Co-author	6
Same-institution	0

5 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 original publication itself carries the full weight of the 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 framework for handling sequence alignment data, which likely filled a significant gap in the ability to manage and analyze large-scale genomic datasets efficiently.

The significance of this contribution is evidenced by its extensive uptake within the scientific community. With tens of thousands of citations, the work is clearly foundational. Furthermore, analysis of citing papers reveals that over 90% are 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: 6

#### CORE PAPER

### [The Sequence Alignment/Map format and SAMtools](#)

2009 · *Bioinformatics* · 68,257 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	<a href="#">Untitled</a>	Argonne National Laboratory, J. Craig Venter Institute, J. Craig Venter Institute; University of California San Diego	New Zealand, United States	—
2	<a href="#">Untitled</a>	Baylor College of Medicine, Brigham Young University, Broad Institute of MIT and Harvard	United States	—
3	<a href="#">TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining</a>	Henan University, Hunan Agricultural University, Institute of Tropical Bioscience and Biotechnology, Chinese Academy of Tropical Agricultural Sciences	China	—
4	<a href="#">A novel antibiotic class targeting the lipopolysaccharide transporter (2024)</a>	Aptuit, Aptuit/Evotec, F. Hoffmann-La Roche	Belgium, Italy, Switzerland	—
5	<a href="#">Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4</a>	Harvard T.H. Chan School of Public Health, Harvard University, Istituto di Scienza e Tecnologie dell'Informazione	Austria, Italy, United Kingdom	<b>Methodology</b>

No.	Citing paper	Citing institution(s)	Country	S2
6	<a href="#">Persistent complement dysregulation with signs of thromboinflammation in active Long Covid</a>	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 co-authored the seminal 2001 Nature paper presenting the initial sequencing and analysis of the human genome, a foundational contribution to genomics.*

**CLAIM:** The researcher's primary contribution is the co-authorship of the landmark 2001 Nature article titled 'Initial sequencing and analysis of the human genome,' which stands as a singular, highly cited work in this portfolio.

**ORIGINALITY:** This work appears to address the fundamental challenge of decoding the human genetic blueprint. As the core paper stands alone without follow-up publications by the researcher in this dataset, the contribution is defined by this singular, high-impact event rather than a longitudinal series of incremental studies.

**SIGNIFICANCE:** The work demonstrates substantial field-wide impact, evidenced by over 26,000 citations. Furthermore, citation analysis reveals that 90.2% of citing papers originate from independent researchers, indicating that the contribution has been widely adopted and utilized by the broader scientific community beyond the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 8

### CORE PAPER

#### [Initial sequencing and analysis of the human genome](#)

2001 · Nature · 26,535 citations (GS)

Field-normalised: 24,058 Semantic Scholar citations place it in the top 1% of Biology papers from 2001 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	<a href="#">DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update)</a>	Frederick National Laboratory for Cancer Research	United States	—
2	<a href="#">Reactive oxygen species, toxicity, oxidative stress, and antioxidants: chronic diseases and aging</a>	Constantine the Philosopher University in Nitra, King Saud University	Czech Republic, Saudi Arabia, Slovakia	—

No.	Citing paper	Citing institution(s)	Country	S2
		University, Slovak University of Technology		
3	<a href="#">DNA methylation: a historical perspective</a>	Max Planck Institute for Molecular Genetics	Germany	—
4	<a href="#">Gene regulation by long non-coding RNAs and its biological functions</a>	Center for Applied Medical Research, University of Navarra, University of the Chinese Academy of Sciences	China, Spain	—
5	<a href="#">Exploring tissue architecture using spatial transcriptomics</a>	NYU Langone Health	United States	—
6	<a href="#">Ancient gene linkages support ctenophores as sister to other animals</a>	Monterey Bay Aquarium Research Institute, University of California, University of Vienna	Austria, United States	—
7	<a href="#">Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype</a>	Johns Hopkins University, Stanford University, University of Texas Southwestern Medical Center	United States	—
8	<a href="#">A complete telomere-to-telomere assembly of the maize genome. (2023)</a>	China Agricultural University, Grandomics Biosciences, Iowa State University	China, 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 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 Bioinformatics paper. 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 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, filling a significant gap in high-throughput genomic data processing capabilities at the time.

The significance of this contribution is evidenced by its extensive uptake in the scientific community, with the core paper accumulating over 55,000 citations. Furthermore, citation analysis reveals that 90.2% of citing papers originate from independent researchers, demonstrating that this work has become a widely adopted standard tool across the global bioinformatics field rather than a niche or self-referential achievement.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 8

#### CORE PAPER

#### [Fast and accurate short read alignment with Burrows–Wheeler transform](#)

No.	Citing paper	Citing institution(s)	Country	S2
1	<a href="#">Allogeneic CD19-targeted CAR-T therapy in patients with severe myositis and systemic sclerosis</a> (2024)	BRL Medicine Inc., East China Normal University, Nanjing University	China	—
2	<a href="#">Transplantation of chemically induced pluripotent stem-cell-derived islets under abdominal anterior rectus sheath in a type 1 diabetes patient</a> (2024)	Peking University, Reprogenix Bioscience, Tianjin First Central Hospital	China	—
3	<a href="#">Nanopore sequencing technology, bioinformatics and applications</a>	The Ohio State University	United States	—
4	<a href="#">Dictionary learning for integrative, multimodal and scalable single-cell analysis</a>	New York Genome Center, New York University	United States	—
5	<a href="#">Chromatin accessibility profiling by ATAC-seq</a> (2022)	Gladstone Institute of Neurological Disease, Gladstone Institutes; University of California San Francisco, University of California San Francisco	United States	—
6	<a href="#">UALCAN: An update to the integrated cancer data analysis platform</a> (2022)	Baylor College of Medicine, Emory University, University of Alabama at Birmingham	United States	—
7	<a href="#">Deterministic reprogramming of neutrophils within tumors</a>	Agency for Science, Technology and Research, A*STAR, Centro Nacional de Investigaciones Cardiovasculares Carlos III	Australia, China, France	—
8	<a href="#">Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding</a> (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).

## D. Citing-Institution Prestige & Geography

### Top citing institutions

Institution	Country	World ranking	Citing papers
Broad Institute of MIT and Harvard	United States	SCImago #112	7
National Institutes of Health	United States	SCImago #44	4
University of Cambridge	United Kingdom	SCImago #63 · THE =3 · QS 6	4
University of Michigan	United States	SCImago #43 · THE 23 · QS 45	3
Wellcome Sanger Institute	United Kingdom	SCImago #204	3
University of Washington	United States	SCImago #45 · THE 25 · QS 81	3
EMBL-EBI	United Kingdom	—	3
European Molecular Biology Laboratory	United Kingdom	—	3

Institution	Country	World ranking	Citing papers
Harvard University	United States	SCImago #4 · THE =5 · QS 5	3
Baylor College of Medicine	United States	SCImago #560	3
National Human Genome Research Institute, National Institutes of Health	United States	—	2
Cardiff University	United Kingdom	SCImago #664 · THE 201–250 · QS 181	2
Harvard Medical School	United States	SCImago #12	2
University of Illinois at Urbana-Champaign	United States	SCImago #206 · THE =41	2
University of Alabama at Birmingham	United States	QS 1001-1200	2

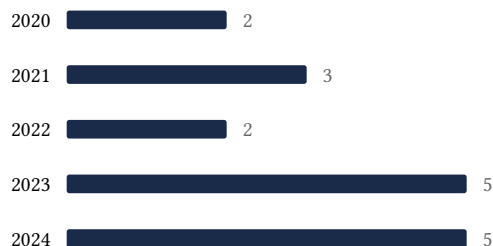
### Geographic distribution of citing authors

Country	Citing papers
United States	26
United Kingdom	11
China	9
Germany	5
Italy	4
Sweden	4
Switzerland	3
Japan	3
Spain	3
Australia	3
Finland	2
Austria	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	The Sequence Alignment/Map format and SAMtools	6	8 CFR 204.5(h)(3)(v) – Criterion 5
Contribution 2	Initial sequencing and analysis of the human genome	8	8 CFR 204.5(h)(3)(v) – Criterion 5
Contribution 3	Fast and accurate short read alignment with Burrows–Wheeler transform	8	8 CFR 204.5(h)(3)(v) – Criterion 5