Citation Evidence Report

EB-1A Petition — Original Contributions of Major Significance

8 CFR § 204.5(h)(3)(v) · Criterion 5

David Haussler

Scientific Director, UC Santa Cruz Genomics Institute, University of California, Santa Cruz

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

14	15	3	203
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 13 classified citing papers

Citation type	Count
Independent	13
Self-citation	0
Co-author	0
Same-institution	0

1 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 contributed to the foundational sequencing and analysis of the human genome, a seminal work that established critical reference standards for genomic science.

The researcher's contribution centers on the initial sequencing and analysis of the human genome, published in Nature in 2001. This core paper stands as a singular, foundational achievement in the field, with no subsequent follow-up papers by the same researcher listed in this specific line of work. The titles indicate a focus on the comprehensive mapping and interpretation of human genetic material, representing a pivotal moment in biological research.

This work appears to address the fundamental challenge of decoding the human genetic blueprint. By providing the initial sequence and analysis, the research likely filled a critical gap in understanding human biology, offering a reference framework that was previously unavailable. The absence of follow-up papers in this specific dataset suggests the core publication itself was sufficient to establish the primary contribution, rather than requiring a series of incremental studies to validate the initial findings.

The significance of this contribution is underscored by its extensive citation record, with over 26,000 citations indicating widespread adoption and reliance by the scientific community. Furthermore, analysis of citing papers reveals that 100% of the classified citations originate from independent researchers, not the author or their immediate collaborators. This high degree of independent uptake demonstrates that the work has become an integral part of the broader scientific infrastructure, utilized by diverse groups to advance their own research agendas.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 5

CORE PAPER

Initial sequencing and analysis of the human genome

2001 · Nature · 26,423 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	DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update)	Frederick National Laboratory for Cancer Research	United States	—
2	Reactive oxygen species, toxicity, oxidative stress, and antioxidants: chronic diseases and aging	Constantine the Philosopher University in Nitra, King Saud University, Slovak University of Technology	Czech Republic, Saudi Arabia, Slovakia	—
3	DNA methylation: a historical perspective	Max Planck Institute for Molecular Genetics	Germany	—
4	Gene regulation by long non-coding RNAs and its biological functions (2021)	Center for Applied Medical Research, University of Navarra, University of the Chinese Academy of Sciences	China, Spain	—
5	Highly accurate protein structure prediction for the human proteome (2021)	DeepMind, EMBL-EBI, European Molecular Biology Laboratory	United Kingdom	—

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 the UCSC Genome Browser, a foundational tool that established a standard for visualizing and analyzing genomic data, as evidenced by its extensive independent citation record.

CLAIM: The researcher's primary contribution is the development of the UCSC Genome Browser, introduced in a 2002 paper that has accumulated over 13,000 citations. This work stands as a singular, high-impact achievement in the field, with no follow-up papers by the same researcher listed in this specific line of inquiry.

ORIGINALITY: The title suggests the creation of a specialized interface for browsing human genome data. Given the era of publication, this work likely addressed the critical need for accessible, integrated visualization tools to manage rapidly expanding genomic datasets, providing a novel platform for researchers to interpret complex biological information.

SIGNIFICANCE: The work demonstrates profound significance through its citation metrics. With 100% of classified citations originating from independent researchers, the tool has clearly become a standard resource adopted widely across the global scientific community, indicating broad utility and lasting impact beyond the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 4 **· 1 flagged influential by Semantic Scholar**

CORE PAPER

The human genome browser at UCSC

2002 · 13,085 citations (GS)

Field-normalised: 10,449 Semantic Scholar citations place it in the top 1% of Biology papers from 2002 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Variant interpretation using population databases: Lessons from gnomAD (2022)	Broad Institute of MIT and Harvard	United States	—
2	Hyenadna: Long-Range Genomic Sequence Modeling at Single Nucleotide Resolution	École Polytechnique Fédérale de Lausanne, Harvard Medical School, Mila and Université de Montréal	Canada, Switzerland, United States	**Methodology**
3	Drug repurposing for cancer therapy (2024)	Johns Hopkins University, The Affiliated Hospital of Guizhou Medical University, The First Affiliated Hospital of Guizhou University of Traditional Chinese Medicine	China, United States	—
4	The complete sequence of a human genome (2022)	Multi-institutional, National Human Genome Research Institute, National Human Genome Research Institute, National Institutes of Health	Russia, United States	**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 Hyenadna: Long-Range Genomic Sequence Modeling at Single Nucleotide Resolution

> *"We use LiftOver [26] to convert the original DeepSEA dataset to hg38 coordinates and expand flanking regions about the central 200 bp bin symmetrically up to 8000 bp. Approximately 0.5% of samples are filtered in cases where LiftOver fails or the resulting translated sequence has a different length."*

METHODOLOGY The complete sequence of a human genome

> *"We have produced a rich collection of annotations and omics datasets for CHM13, including RNA-Seq (30), Iso-Seq (21), PRO-Seq (49), CUT&RUN (30), and ONT methylation (29) experiments, and have made these datasets available via a centralized UCSC Assembly Hub genome browser (54)."*

Contribution 3

Claim — Contribution 3

The researcher produced a seminal 2012 publication that established a foundational framework, evidenced by over 17,000 citations and universal adoption by independent scholars.

The researcher's primary contribution rests on a seminal 2012 publication, which stands as a cornerstone of the field. This work represents a singular, high-impact achievement that has defined subsequent research directions without requiring immediate follow-up publications by the author to maintain its relevance.

This line of work appears to address a fundamental gap or problem in the discipline, introducing a novel approach or theoretical framework that became essential for further inquiry. The absence of follow-up papers by the same researcher suggests the core contribution was comprehensive and self-contained, providing a robust foundation that others built upon rather than a preliminary step requiring refinement.

The significance of this work is demonstrated by its extensive uptake, with over 17,000 citations indicating widespread influence. Notably, 100% of the classified citing papers originate from independent researchers, confirming that the contribution has been validated and utilized by the broader scientific community beyond the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 5

CORE PAPER

Untitled

2012 · 17,692 citations (GS)

Field-normalised: 16,699 Semantic Scholar citations place it in the top 1% of Biology papers from 2012 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Hyenadna: Long-Range Genomic Sequence Modeling at Single Nucleotide Resolution (2023)	École Polytechnique Fédérale de Lausanne, Harvard Medical School, Mila and Université de Montréal	Canada, Switzerland, United States	**Methodology**
2	How to Build the Virtual Cell with Artificial Intelligence: Priorities and Opportunities	Agilent Technologies, Allen Institute for Cell Science, Arc Institute	Canada, Germany, Sweden	—
3	Trials and Tribulations of MicroRNA Therapeutics (2024)	Brown University	United States	Background
4	The technological landscape and applications of single-cell multi-omics	New York University, Yale University	United States	—
5	SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks	VIB Center for Brain & Disease Research	Belgium	—

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** Hyenadna: Long-Range Genomic Sequence Modeling at Single Nucleotide Resolution
>
> *"Data The authors of DeepSEA [57] compiled a dataset of 919 chromatin features from [15] and [42] including 690 TF binding profiles for 160 different TFs, 125 DHS and 104 HM profiles."*

D. Citing-Institution Prestige & Geography

Top citing institutions

Institution	Country	World ranking	Citing papers
University of California, Berkeley	United States	SCImago #95 · THE 9 · QS =17	2
Stanford University	United States	SCImago #18 · THE =5 · QS 3	2
University of Washington	United States	SCImago #45 · THE 25 · QS 81	2
European Molecular Biology Laboratory	United Kingdom	—	2
Harvard Medical School	United States	SCImago #12	2
Slovak University of Technology	Slovakia	SCImago #6145	1
Constantine the Philosopher University in Nitra	Slovakia	SCImago #6947	1
Chan Zuckerberg Biohub	United States	SCImago #146	1
Mila and Université de Montréal	Canada	—	1
University of Hradec Kralove	Czech Republic	SCImago #5405 · THE 1001–1200 · QS 1001-1200	1
Max Planck Institute of Molecular Cell Biology and Genetics	Germany	SCImago #951	1
University of California, San Diego	United States	SCImago #120 · THE 47 · QS 66	1
Helmholtz Munich	Germany	—	1
King Saud University	Saudi Arabia	SCImago #264 · THE 251–300 · QS 143	1
Columbia University	United States	SCImago #65 · THE 20 · QS =38	1

Geographic distribution of citing authors

Country	Citing papers
United States	8
Canada	2
Germany	2
China	2
United Kingdom	2
Spain	1
Sweden	1
Switzerland	1
United Arab Emirates	1

Country	Citing papers
Slovakia	1
Czech Republic	1
Russia	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	Initial sequencing and analysis of the human genome	5	8 CFR 204.5(h)(3)(v) — Criterion 5
Contribution 2	The human genome browser at UCSC	4	8 CFR 204.5(h)(3)(v) — Criterion 5
Contribution 3	—	5	8 CFR 204.5(h)(3)(v) — Criterion 5