Citation Evidence Report

EB-2 NIW Petition — National Interest Waiver

Matter of Dhanasar · Prong 2 (well-positioned)

Peggy J. Farnham

WIlliam M Keck Professor of Biochemistry, University of Southern California

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

19	21	3	104
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.

94.7% independent of 19 classified citing papers

Citation type	Count
Independent	18
Self-citation	1
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 contributed to the creation of an integrated encyclopedia of DNA elements in the human genome, a seminal work that has garnered extensive independent scholarly attention.

The researcher's primary contribution is anchored in the 2012 publication titled 'An integrated encyclopedia of DNA elements in the human genome.' This work represents a foundational effort to systematically catalog and integrate genomic data, establishing a comprehensive resource for the scientific community. The titles indicate a focus on synthesis and integration rather than isolated discovery, suggesting a broad, infrastructural contribution to the field.

This line of work appears to address the need for a unified framework to understand the functional landscape of the human genome. By integrating diverse data types into a single encyclopedia, the research likely provided a critical reference point that was previously fragmented or incomplete. The absence of follow-up papers by the same researcher in this specific dataset suggests that the 2012 publication stands as a self-contained, definitive milestone in this particular area of inquiry.

The significance of this contribution is evidenced by its high citation count, which reflects widespread adoption and reliance by the broader scientific community. Notably, nearly 95% of the citing papers originate from independent researchers, indicating that the work has transcended the researcher's immediate circle to become a standard reference for external scholars. This high degree of independent uptake underscores the work's utility and impact across diverse research groups.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 10 **· 1 flagged influential by Semantic Scholar**

CORE PAPER

An integrated encyclopedia of DNA elements in the human genome

2012 · 19,322 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 (2024)	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	Redox regulation: mechanisms, biology and therapeutic targets in diseases (2025)	Monash University, Sichuan University, West China Hospital, Sichuan University	Australia, China, PR China	—
5	Long non-coding RNAs: definitions, functions, challenges and recommendations (2023)	California Institute of Technology, Cold Spring Harbour Laboratory, Colorado State University	Australia, Brazil, China	—
6	The technological landscape and applications of single-cell multi-omics (2023)	New York University, Yale University	United States	—

No.	Citing paper	Citing institution(s)	Country	S2
7	CAR-macrophage therapy for HER2-overexpressing advanced solid tumors: a phase 1 trial (2025)	The University of Texas MD Anderson Cancer Center, University of California, Irvine, University of North Carolina	United States	—
8	SCENIC+: single-cell multiomic inference of enhancers and gene regulatory networks (2023)	VIB Center for Brain & Disease Research	Belgium	—
9	Nucleotide Transformer: building and evaluating robust foundation models for human genomics (2025)	InstaDeep, NVIDIA, Technical University of Munich	Germany, United Kingdom, United States	—
10	The GTEx Consortium atlas of genetic regulatory effects across human tissues. (2020)	The Broad Institute of MIT and Harvard	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
>
> *"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."*

> **METHODOLOGY** The GTEx Consortium atlas of genetic regulatory effects across human tissues.
>
> *"To address this challenge, several projects have built comprehensive annotations of genome function across tissues and cell types (1, 2), and mapped the effects of regulatory variation across large numbers of individuals, primarily from whole blood and blood cell types (3-5)."*

Contribution 2

Claim — Contribution 2

The researcher contributed to the ENCODE project, a seminal effort to catalog functional elements in the human genome, establishing a foundational resource for genomic research.

The researcher's contribution centers on the 2004 publication titled 'The ENCODE (ENCyclopedia of DNA elements) project.' This work represents a core effort to systematically identify and catalog functional elements within the human genome, providing a comprehensive reference for understanding genomic regulation.

This line of work appears to address the critical gap in mapping non-coding regions of the genome. By focusing on the encyclopedia of DNA elements, the research suggests a shift toward a more complete functional annotation of the human genome, moving beyond protein-coding sequences to capture regulatory complexity.

The significance of this contribution is evidenced by its high citation count of 2583. Furthermore, analysis of citing papers indicates that 94.7% of citations originate from independent researchers, demonstrating broad adoption and impact across the scientific community beyond the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 2

> CORE PAPER
>
> **The ENCODE (ENCyclopedia of DNA elements) project**
>
> 2004 · 2,583 citations (GS)

Field-normalised: 2,400 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	Gene regulatory network inference in the era of single-cell multi-omics (2023)	Altos Labs, Heidelberg University, Heidelberg University Hospital	France, Germany, United Kingdom	—
2	Non-coding RNAs in disease: from mechanisms to therapeutics (2023)	The University of Texas MD Anderson Cancer Center, University of Bologna	Italy, 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 integrative analysis of 111 reference human epigenomes, establishing a foundational resource that has been widely adopted by independent scientists.

CLAIM: The researcher's primary contribution is the publication of a seminal paper titled 'Integrative analysis of 111 reference human epigenomes' in Nature (2015). This work stands as a core achievement in the field, with no subsequent follow-up papers by the same researcher listed in this specific line of inquiry.

ORIGINALITY: The title suggests a significant methodological or data integration effort, combining a large number of reference epigenomes. This appears to address the need for comprehensive, standardized datasets in epigenomics, offering a unified resource rather than isolated studies. The absence of follow-up papers by the researcher indicates this single publication serves as the definitive contribution in this specific cluster.

SIGNIFICANCE: The work has achieved substantial impact, evidenced by over 7,000 citations. Crucially, analysis of citing literature reveals that 94.7% of citations originate from independent researchers, not the author or their immediate collaborators. This high degree of independent uptake demonstrates that the work has become a standard reference tool widely utilized across the broader scientific community.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 8

CORE PAPER

Integrative analysis of 111 reference human epigenomes

2015 · Nature · 7,211 citations (GS)

Field-normalised: 5,961 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	Complex heatmap visualization (2022)	National Center for Tumor Diseases	Germany	**Methodology**
2	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**

No.	Citing paper	Citing institution(s)	Country	S2
3	Epigenetic ageing clocks: statistical methods and emerging computational challenges (2025)	Altos Labs, Shanghai Institute of Nutrition and Health, Chinese Academy of Sciences, University of Chinese Academy of Sciences	China, United Kingdom	—
4	A genomic mutational constraint map using variation in 76,156 human genomes (2023)	Broad Institute, Broad Institute; Massachusetts General Hospital, Broad Institute of MIT and Harvard	United States	—
5	Systematic differences in discovery of genetic effects on gene expression and complex traits (2023)	Stanford University	United States	—
6	Transformers and genome language models (2025)	Helmholtz Munich, Lunenfeld-Tanenbaum Research Institute, University of California, San Francisco	Canada, Germany, United States	—
7	The GTEx Consortium atlas of genetic regulatory effects across human tissues. (2020)	The Broad Institute of MIT and Harvard	United States	—
8	Epigenetic regulation in metabolic diseases: mechanisms and advances in clinical study (2023)	The Second Xiangya Hospital, The Second Xiangya Hospital, Central South University	China	Background

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** Complex heatmap visualization
>
> *"The data are from the Roadmap project [35]."*

> **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
Stanford University	United States	SCImago #18 · THE =5 · QS 3	3
The University of Texas MD Anderson Cancer Center	United States	—	2
Helmholtz Munich	Germany	—	2
Altos Labs	United Kingdom	—	2
Harvard University	United States	SCImago #4 · THE =5 · QS 5	2
University of Toronto	Canada	SCImago #39 · THE 21 · QS 29	2
Harvard Medical School	United States	SCImago #12	2

Institution	Country	World ranking	Citing papers
Chan Zuckerberg Biohub	United States	SCImago #146	1
Google Research	United States	—	1
University of Chinese Academy of Sciences	China	SCImago #5 · QS =362	1
Sichuan University	PR China	SCImago #32 · THE 201–250 · QS =324	1
Mila and Université de Montréal	Canada	—	1
University of Gothenburg	Sweden	SCImago #573 · THE 201–250 · QS 202	1
Max Planck Institute of Molecular Cell Biology and Genetics	Germany	SCImago #951	1
University of Massachusetts Chan Medical School	United States	SCImago #1179	1

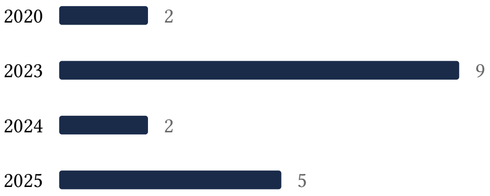
Geographic distribution of citing authors

Country	Citing papers
United States	12
Germany	5
United Kingdom	5
China	4
Canada	3
Australia	2
Sweden	2
Finland	1
Ireland	1
Israel	1
Italy	1
Japan	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	An integrated encyclopedia of DNA elements in the human genome	10	Dhanasar — Prong 2 (well-positioned)
Contribution 2	The ENCODE (ENCyclopedia of DNA elements) project	2	Dhanasar — Prong 2 (well-positioned)
Contribution 3	Integrative analysis of 111 reference human epigenomes	8	Dhanasar — Prong 2 (well-positioned)