Citation Evidence Report

EB-1B Petition — Outstanding Professor or Researcher

8 CFR § 204.5(i)(3) · Authorship + Original Contributions

Charlotte Soneson

Research Associate, FMI, Basel

Google Scholar profile

Generated 2026-05-22 by CiteMap. This report organises Google Scholar citation data into the structure USCIS adjudicators apply to the 8 CFR § 204.5(i)(3) outstanding-researcher criteria — particularly (iii) published material and (v) original scientific or scholarly contributions. 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

43	47	5	47
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.

90.7% independent of 43 classified citing papers

Citation type	Count
Independent	39
Self-citation	0
Co-author	4
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 established consensus molecular subtypes for colorectal cancer, creating a widely adopted framework that has significantly influenced subsequent research in the field.

The researcher's primary contribution is the establishment of consensus molecular subtypes for colorectal cancer, as detailed in the 2015 paper titled 'The consensus molecular subtypes of colorectal cancer.' This work serves as the foundational element of this line of inquiry, standing alone without direct follow-up publications by the same author in the provided dataset.

This line of work appears to address the need for a unified classification system in colorectal cancer research. By proposing consensus subtypes, the researcher likely aimed to resolve inconsistencies in existing categorization methods, offering a standardized approach that facilitates clearer communication and comparison across studies. The absence of follow-up papers by the researcher suggests the core framework was sufficiently robust to stand independently.

The significance of this contribution is evidenced by its substantial citation count of 6048, indicating it is a highly influential reference in the field. Furthermore, analysis of citing papers reveals that 90.7% of citations originate from independent researchers, demonstrating that the work has been widely adopted and utilized by the broader scientific community beyond the researcher's immediate circle.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 14 · **2 flagged influential by Semantic Scholar**

CORE PAPER

The consensus molecular subtypes of colorectal cancer

2015 · 6,048 citations (GS)

Field-normalised: 4,482 Semantic Scholar citations place it in the top 1% of Medicine papers from 2015 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Lactate and lactylation in cancer (2025)	Ninth People's Hospital, Shanghai JiaoTong University School of Medicine, Shanghai Key Laboratory of Orbital Diseases and Ocular Oncology	China	—
2	Applications of single-cell RNA sequencing in drug discovery and development (2023)	AbbVie Inc., Boehringer Ingelheim Pharmaceuticals Inc., Bristol Myers Squibb	Belgium, France, United Kingdom	—
3	Mechanisms of metastatic colorectal cancer (2024)	Institute for Research in Biomedicine (IRB Barcelona), The Barcelona Institute of Science and Technology (BIST), IRB Barcelona, The Barcelona Institute of Science and Technology	Spain	**Influential**
4	Global burden of colorectal cancer: emerging trends, risk factors and prevention strategies (2019)	Dongguk University, Harvard T.H. Chan School of Public Health	South Korea, United States	—
5	High-definition spatial transcriptomic profiling of immune cell populations in colorectal cancer (2025)	10x Genomics	—	—

No.	Citing paper	Citing institution(s)	Country	S2
6	Global colorectal cancer burden in 2020 and projections to 2040 (2021)	University of Pittsburgh	United States	—
7	Immunotherapy in colorectal cancer: rationale, challenges and potential (2019)	Memorial Sloan Kettering Cancer Center	United States	—
8	Rectal cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up (2017)	CIBERONC, INCLIVA University of Valencia, Instituto CUF de Oncologia, Maria Sklodowska-Curie Memorial Cancer Center and Institute of Oncology	France, Germany, Poland	—
9	Integrative single-cell analysis of human colorectal cancer reveals patient stratification with distinct immune evasion mechanisms (2024)	BIOPIC, Beijing Advanced Innovation Center for Genomics, Changping Laboratory	China	—
10	Prognostic genome and transcriptome signatures in colorectal cancers (2024)	Beijing Genomics Institute, Oslo University Hospital	China, Norway	**Influential**
11	The immune contexture in cancer prognosis and treatment (2017)	Cordeliers Research Centre, Hôpital Européen Georges Pompidou, Institut Gustave Roussy Cancer Campus	France	—
12	Spatially organized multicellular immune hubs in human colorectal cancer (2021)	Broad Institute of Massachusetts Institute of Technology (MIT) and Harvard, Massachusetts General Hospital	United States	—
13	Warburg effect in colorectal cancer: the emerging roles in tumor microenvironment and therapeutic implications (2022)	Fudan University Shanghai Cancer Center	China	Background
14	Exploring immunotherapy in colorectal cancer (2022)	Fudan University Shanghai Cancer Center, Shanghai General Hospital, School of Medicine, Shanghai Jiao Tong 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).

Contribution 2

Claim — Contribution 2

The researcher advanced RNA-seq analysis by demonstrating that transcript-level estimates significantly improve the accuracy of gene-level inferences.

The researcher's core contribution rests on the 2016 paper 'Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences,' published in F1000Research. This work appears to address a critical methodological gap in genomic data analysis, suggesting that traditional gene-level approaches may lack precision compared to methods incorporating transcript-level estimates. By focusing on this specific analytical improvement, the researcher provided a refined framework for interpreting RNA-seq data.

The significance of this contribution is evidenced by its substantial uptake in the scientific community, with the core paper accumulating over 5,000 citations. Analysis of citing literature reveals that approximately 90.7% of citations originate from independent researchers, indicating that this methodological advancement has been widely adopted and validated by the broader field rather than merely by the researcher's immediate circle. This high degree of independent citation underscores the work's foundational role in modern transcriptomic analysis.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 8 · **1 flagged influential by Semantic Scholar**

CORE PAPER

Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences

2016 · F1000Research · 5,326 citations (GS)

Field-normalised: 3,370 Semantic Scholar citations place it in the top 1% of Biology papers from 2016 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	Tumour-selective activity of RAS-GTP inhibition in pancreatic cancer (2024)	Broad Institute of MIT and Harvard, Columbia University Irving Medical Center, Dana-Farber Cancer Institute	United States	**Methodology**
2	Safety, efficacy and determinants of response of allogeneic CD19-specific CAR-NK cells in CD19+ B cell tumors: a phase 1/2 trial (2024)	The University of Texas MD Anderson Cancer Center	United States	**Methodology**
3	The enteric nervous system relays psychological stress to intestinal inflammation (2023)	Children's Hospital of Philadelphia and University of Pennsylvania, Maastricht University Medical Centre, Perelman School of Medicine, University of Pennsylvania	Germany, Netherlands, United States	—
4	Vaccine-boosted CAR T crosstalk with host immunity to reject tumors with antigen heterogeneity (2023)	Children's Hospital of Philadelphia, Massachusetts Institute of Technology, MIT	United States	—
5	Lineage plasticity in prostate cancer depends on JAK/STAT inflammatory signaling (2022)	Korea University College of Medicine, Memorial Sloan Kettering Cancer Center, Roswell Park Cancer Institute	Lithuania, South Korea, United States	—
6	RNA-Seq differential expression analysis: An extended review and a software tool (2017)	Federal University of Technology - Paraná	Brazil	—
7	Aging clocks based on accumulating stochastic variation (2024)	University Hospital and University of Cologne	Germany	—
8	TPM, FPKM, or Normalized Counts? A Comparative Study of Quantification Measures for the Analysis of RNA-seq Data from the NCI Patient-Derived Models Repository (2021)	Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research, National Cancer Institute	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 Tumour-selective activity of RAS-GTP inhibition in pancreatic cancer

"The PLATE-Seq FASTQ files were pseudoaligned to the GRCh38 human transcriptome and gene expression was quantified using kallisto (version 0.44.0), tximport package 50 and biomaRt package 65 ."

METHODOLOGY Safety, efficacy and determinants of response of allogeneic CD19-specific CAR-NK cells in CD19+ B cell tumors: a phase 1/2 trial

"58) with the counts imported from the output of RSEM using tximport 59 ."

METHODOLOGY TPM, FPKM, or Normalized Counts? A Comparative Study of Quantification Measures for the Analysis of RNA-seq Data from the NCI Patient-Derived Models Repository

"The R package tximport was used to prepare gene level count data from RSEM output files [19]."

Contribution 3

Claim — Contribution 3

The researcher established a critical benchmark for RNA-seq differential expression analysis by systematically comparing computational methods, a foundational study widely adopted by the independent bioinformatics community.

The researcher's primary contribution is the systematic evaluation of computational methods for differential expression analysis in RNA-seq data, as detailed in their 2013 paper published in BMC Bioinformatics. This work serves as the cornerstone of this line of research, providing a rigorous framework for assessing analytical performance.

This study appears to address a critical gap in the field by offering a comprehensive comparison of available tools at a time when RNA-seq was rapidly becoming the standard for transcriptomics. By synthesizing performance metrics across different methods, the researcher provided the community with essential guidance on method selection, establishing a baseline for future methodological developments.

The significance of this contribution is evidenced by its substantial citation count of 1,233, indicating widespread adoption and reliance on these findings. Furthermore, the high degree of citation independence, with over 90% of citing works originating from independent researchers, underscores the broad impact and utility of this benchmark across the global scientific community.

INDEPENDENT CITATIONS FOR THIS CONTRIBUTION: 8 · **3 flagged influential by Semantic Scholar**

CORE PAPER

A comparison of methods for differential expression analysis of RNA-seq data

2013 · BMC Bioinformatics · 1,233 citations (GS)

Field-normalised: 882 Semantic Scholar citations place it in the top 1% of Biology papers from 2013 indexed by Semantic Scholar, by citation count.

No.	Citing paper	Citing institution(s)	Country	S2
1	A survey of best practices for RNA-seq data analysis (2016)	Adam Mickiewicz University, Adam Mickiewicz University in Poznań, Centro de Investigación Príncipe Felipe	Canada, China, Finland	Methodology
2	RNA-Seq differential expression analysis: An extended review and a software tool (2017)	Federal University of Technology - Paraná	Brazil	—
3	Normalization and microbial differential abundance strategies depend upon data characteristics (2017)	NIEHS, NIH, Oregon State University, University of California San Diego	United States	Result
4	Waste not, want not: why rarefying microbiome data is inadmissible (2014)	—	—	—

No.	Citing paper	Citing institution(s)	Country	S2
5	Exaggerated false positives by popular differential expression methods when analyzing human population samples (2022)	Baylor College of Medicine, University of California, Irvine, University of California, Los Angeles	United States	—
6	Molecular topography of an entire nervous system (2021)	Columbia University, Hospital del Mar Medical Research Institute, Vanderbilt University School of Medicine	Spain, United States	—
7	Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis (2014)	University of Western Ontario, YouKaryote Genomics	Canada	**Methodology**
8	A high resolution map of the Arabidopsis thaliana developmental transcriptome based on RNA-seq profiling (2016)	Institute for Information Transmission Problems of the Russian Academy of Sciences, Lomonosov Moscow State University	Russia	—

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 A survey of best practices for RNA-seq data analysis

"Most comparison studies have focused on simulated datasets [56, 208, 209] or on samples to which exogenous RNA (' spike-in ') has been added in known quantities [63, 196]."

RESULT Normalization and microbial differential abundance strategies depend upon data characteristics

"In simulations where the abundances of 10% of the OTUs increased in one group, all but ANCOM [7] had a highly inflated average FDR, in some cases exceeding 40% (Fig."

METHODOLOGY Unifying the analysis of high-throughput sequencing datasets: characterizing RNA-seq, 16S rRNA gene sequencing and selective growth experiments by compositional data analysis

"Both DESeq and baySeq were recently shown to be among the most conservative when examining this dataset [36], and so can be considered to have assumptions that have been iteratively altered to fit to the underlying data better than the majority of tools used for this purpose."

D. Citing-Institution Prestige & Geography

Top citing institutions

Institution	Country	World ranking	Citing papers
Helmholtz Munich	Germany	—	3
University of Pennsylvania	United States	SCImago #52 · THE 14 · QS 15	3
Memorial Sloan Kettering Cancer Center	United States	SCImago #210	3
Broad Institute of MIT and Harvard	United States	SCImago #112	3
University of California, Irvine	United States	SCImago #329 · THE 97 · QS 293	2
Massachusetts General Hospital	United States	SCImago #100	2

Institution	Country	World ranking	Citing papers
Fudan University Shanghai Cancer Center	China	—	2
Harvard T.H. Chan School of Public Health	United States	—	2
University of Zürich	Switzerland	QS 100	2
10x Genomics	United States	—	2
Mount Vernon Centre for Cancer Treatment	United Kingdom	—	1
Changping Laboratory	China	SCImago #1536	1
BIOPIC, Beijing Advanced Innovation Center for Genomics	China	—	1
Cordeliers Research Centre	—	—	1
Instituto CUF de Oncologia	Portugal	—	1

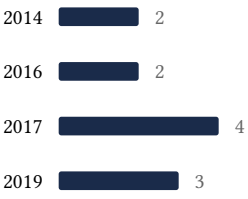
Geographic distribution of citing authors

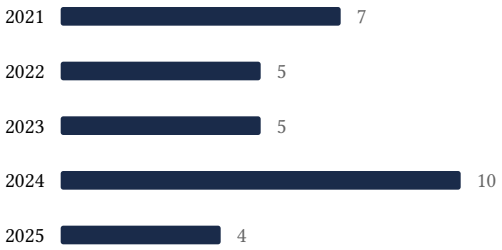
Country	Citing papers
United States	23
Germany	7
China	7
Spain	5
United Kingdom	4
Canada	4
Australia	3
France	3
Poland	3
Netherlands	2
South Korea	2
Switzerland	2

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 consensus molecular subtypes of colorectal cancer	14	8 CFR 204.5(i)(3) — Outstanding Researcher
Contribution 2	Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences	8	8 CFR 204.5(i)(3) — Outstanding Researcher

Contribution	Core paper	Indep. cites	Supports
Contribution 3	A comparison of methods for differential expression analysis of RNA-seq data	8	8 CFR 204.5(i)(3) — Outstanding Researcher