

# UCSF

## UC San Francisco Previously Published Works

### Title

FAIR Data Reuse in Traumatic Brain Injury: Exploring Inflammation and Age as Moderators of Recovery in the TRACK-TBI Pilot

### Permalink

<https://escholarship.org/uc/item/2vw6q438>

### Authors

Huie, J Russell  
Chou, Austin  
Torres-Espin, Abel  
et al.

### Publication Date

2021

### DOI

10.3389/fneur.2021.768735

### Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed



# FAIR Data Reuse in Traumatic Brain Injury: Exploring Inflammation and Age as Moderators of Recovery in the TRACK-TBI Pilot

J. Russell Huie<sup>1,2</sup>, Austin Chou<sup>1</sup>, Abel Torres-Espin<sup>1</sup>, Jessica L. Nielson<sup>3,4</sup>, Esther L. Yuh<sup>1,5</sup>, Raquel C. Gardner<sup>6</sup>, Ramon Diaz-Arrastia<sup>7</sup>, Geoff T. Manley<sup>1</sup>, Adam R. Ferguson<sup>1,2\*</sup> and the TRACK-TBI Investigators

## OPEN ACCESS

### Edited by:

Robert David Stevens,  
Johns Hopkins University,  
United States

### Reviewed by:

Pushpa Sharma,  
Uniformed Services University of the  
Health Sciences, United States  
Nicholas Morris,  
University of Maryland, United States

### \*Correspondence:

Adam R. Ferguson  
adam.ferguson@ucsf.edu

### Specialty section:

This article was submitted to  
Neurotrauma,  
a section of the journal  
Frontiers in Neurology

**Received:** 01 September 2021

**Accepted:** 01 October 2021

**Published:** 03 November 2021

### Citation:

Huie JR, Chou A, Torres-Espin A,  
Nielson JL, Yuh EL, Gardner RC,  
Diaz-Arrastia R, Manley GT,  
Ferguson AR and the TRACK-TBI  
Investigators (2021) FAIR Data Reuse  
in Traumatic Brain Injury: Exploring  
Inflammation and Age as Moderators  
of Recovery in the TRACK-TBI Pilot.  
*Front. Neurol.* 12:768735.  
doi: 10.3389/fneur.2021.768735

<sup>1</sup> Brain and Spinal Injury Center, Department of Neurosurgery, University of California, San Francisco, San Francisco, CA, United States, <sup>2</sup> San Francisco Veterans Affairs Medical Center, San Francisco, CA, United States, <sup>3</sup> Department of Psychiatry and Behavioral Sciences, University of Minnesota, Minneapolis, MN, United States, <sup>4</sup> Institute for Health Informatics, University of Minnesota, Minneapolis, MN, United States, <sup>5</sup> Department of Radiology, University of California, San Francisco, San Francisco, CA, United States, <sup>6</sup> Department of Neurology, University of California, San Francisco, San Francisco, CA, United States, <sup>7</sup> Department of Neurology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States

The guiding principle for data stewardship dictates that data be FAIR: findable, accessible, interoperable, and reusable. Data reuse allows researchers to probe data that may have been originally collected for other scientific purposes in order to gain novel insights. The current study reuses the Transforming Research and Clinical Knowledge for Traumatic Brain Injury (TRACK-TBI) Pilot dataset to build upon prior findings and ask new scientific questions. Specifically, we have previously used a multivariate analytics approach to multianalyte serum protein data from the TRACK-TBI Pilot dataset to show that an inflammatory ensemble of biomarkers can predict functional outcome at 3 and 6 months post-TBI. We and others have shown that there are quantitative and qualitative changes in inflammation that come with age, but little is known about how this interaction affects recovery from TBI. Here we replicate the prior proteomics findings with improved missing value analyses and non-linear principal component analysis and then expand upon this work to determine whether age moderates the effect of inflammation on recovery. We show that increased age correlates with worse functional recovery on the Glasgow Outcome Scale-Extended (GOS-E) as well as increased inflammatory signature. We then explore the interaction between age and inflammation on recovery, which suggests that inflammation has a more detrimental effect on recovery for older TBI patients.

**Keywords:** leveraging data science for traumatic brain injury prevention, evidence based healthcare, precision medicine, data sharing, aging, inflammation, proteomics, outcomes

## INTRODUCTION

As data science technologies rapidly advance, we are able to collect, store, and leverage more clinical traumatic brain injury (TBI) research data than ever before, opening new opportunities for precision medicine in neurotrauma. For example, machine learning has the potential to harness the full information contained within medical records to place individual patients into their own unique position in a multidimensional outcome space for TBI (1, 2). However, in order to realize the potential of biomedical “big-data,” there needs to be effective data management and governance that will enable responsible application of advanced analytics. In order to ensure that research data can be optimally utilized and shepherded appropriately, governance bodies have determined that biomedical data need to be made FAIR: findable, accessible, interoperable, and reusable (3). FAIR principles provide a roadmap for how data can be shared and reused in a way that is beneficial to the data originator, to the reuser, and ideally to the field as a whole. The end goal of FAIR data is to crowdsource secondary analyses, power machine-learning-based exploratory analyses, and accelerate novel hypothesis generation. In addition, data reuse encourages reproducibility and helps ensure that the scientific community extracts as much information as possible from costly federal investments in research dollars and data collection time (4–6). The Transforming Research and Clinical Knowledge in Traumatic Brain Injury (TRACK-TBI) study is an example of early adoption of large-scale FAIR data sharing with the goal of advancing knowledge discovery in clinical TBI (7–9). The initial TRACK-TBI Pilot project generated a dataset of 586 subjects from three level 1 trauma centers that were deeply phenotyped across 900+ variables spanning clinical variables, biofluid biomarkers, imaging biomarkers, and neuropsychiatric and cognitive outcomes. Primary analysis papers were purpose-curated around specific hypotheses, resulting in 38 papers (as of July 2021), and staged improvement in data curation quality over time. Ultimately, the multidimensional curation across variables began to reach sufficient maturity such that later versions of the dataset were primed for advanced multidimensional analytics, including application of machine learning and other artificial intelligence tools to predict outcomes at the level of individual subjects while taking into account ever larger numbers of features (10–12). In addition, the TRACK-TBI Pilot gave rise to an 18-center TRACK-TBI study that has generated 52 papers (and counting).

The goal of the current study is to demonstrate the real-world application of FAIR principles to TBI and expand on prior findings by combining and reusing TRACK-TBI Pilot data from two previously published studies (10, 11). Prior work demonstrated that a multianalyte blood biomarker panel correlates with TBI outcomes. Much of the variance in blood biomarkers reflected inflammatory pathways, and patients with higher pro-inflammatory expression were more likely to have been diagnosed by neuroradiologists as having at least one pathology on computed tomography scan (CT positive) and performed significantly worse on the California Verbal Learning Task and the Glasgow Outcome Scale-Extended

(GOS-E) than those expressing a less inflammatory expression profile (10).

Here we revisit these data and cross-curate them with age demographics to determine the extent to which age moderates the role of inflammation on recovery after TBI. As we and others have shown in previous work, the body reacts differently to TBI depending on age (13). Patients with TBI over the age of 40 are more likely to decline over the first 5 years after injury, and those over 55 have an increased risk of developing dementia and Parkinson’s disease (14–16). The patterns of expression for inflammatory patterns in particular are sensitive to age as has been shown in both pre-clinical and clinical studies of TBI, spinal cord injury (SCI), and stroke (17–21). Despite this, little is known about the nature of the interaction between age and inflammation on outcome after TBI.

The current data reuse study aims to first determine the effect of age on outcome in the TRACK-TBI Pilot cohort. We then sought to replicate the multivariate analyses of the protein biomarker panel using improved techniques for handling missing data and non-linear dimensionality reduction while verifying that the biomarker profile is associated with functional outcome. Finally, we explore the extent to which age moderates the ensemble biomarker profile as a predictor of recovery after TBI.

## METHODS

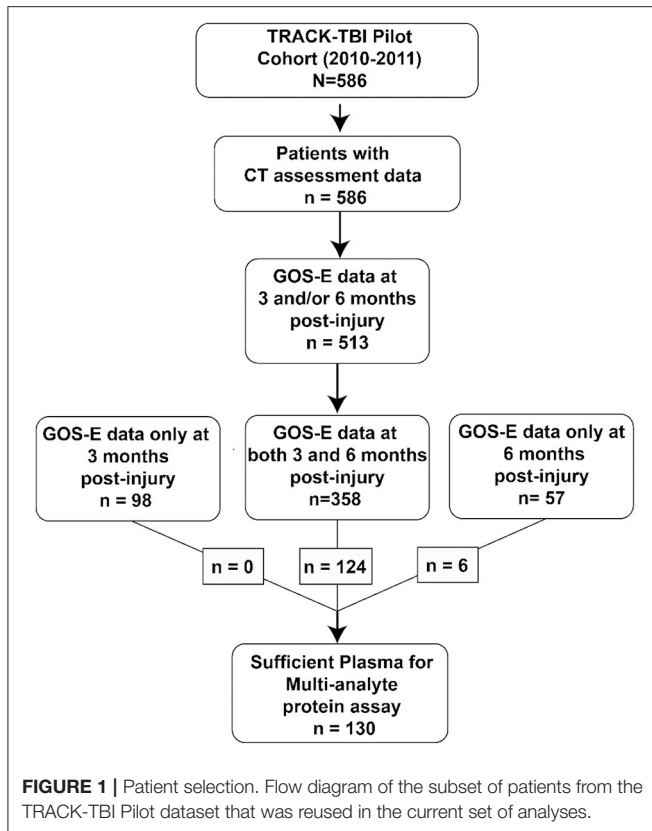
### Data Reuse

The current dataset was created by merging and cross-curating data from two previous subsets of the full TRACK-TBI Pilot dataset that were used previously (10, 11). This dataset was composed of TRACK-TBI participants ( $N = 586$ ) aged 17+ years who were enrolled between April 2010 and January 2011 from three level 1 trauma centers in the United States. Assessment of evidence for acute TBI within 24 h of injury by non-contrast head CT was used as the primary inclusion criterion. Exclusion criteria included non-fluency in English, contraindication to MRI, pregnancy, and current incarceration/legal detention or placement on psychiatric hold.

The current dataset utilizes only a small number of variables that were drawn from the previously published data versions, including age, presence of brain pathology on CT scan (CT positive), proteomic blood biomarker panel used in Huie et al. (10), and the GOS-E (3 and 6 months, **Figure 1**). This new merged data version has been recorded as a branch in the TRACK-TBI GitHub in order to capture this specific subset of patients, raw variables, and derived variables in perpetuity. This unique dataset version is available upon request.

### Proteomics Panel

Blood samples were collected at <24 h of injury, and plasma was prepared following TBI Common Data Elements Biospecimens and Biomarkers Working Group guidelines. Samples were centrifuged, aliquoted, and frozen at  $-80^{\circ}\text{C}$  for batch processing. Frozen plasma samples were analyzed by Myriad Rules-Based Medicine (Myriad RBM, Austin, TX) and quantified using a multiplexed fluorescent immunoassay profile (HumanMAP v2.0).



## CT Scan

A brain CT was performed on all patients within 24 h of their arrival in the emergency department (ED presentation). CT data collection and interpretation were performed in accordance with TBI-CDE Working Group (22), with a blinded board-certified neuroradiologist reviewing the de-identified, anonymized scans. Rotterdam and Marshall scores were also determined. A positive CT indicated any intracranial pathology as determined by the neuroradiologist.

## Outcome Measures

The GOS-E was used to assess overall functional disability at 3 and 6 months post-injury. For consistency with prior work (10, 23), GOS-E scores were dichotomized in two ways: good recovery (GOS-E > 4) vs. poor recovery (GOS-E ≤ 4) and complete recovery (GOS-E = 8) vs. incomplete recovery (GOS-E < 8). To determine if patients' recovery improved between 3 and 6 months post-injury, a difference score was calculated (GOS-E at 6 months – GOS-E at 3 months).

## Missing Values in Proteomics

Protein values were considered missing if they fell outside the range of detection. A threshold for missingness for any biomarker was then set so that if any particular protein was outside the limit of detection for more than 50 patients (i.e., 30% missingness), then that protein would be excluded from analysis. This thresholding resulted in 15 biomarkers removed from analysis

due to missingness. Fifty-seven biomarkers were subsequently included for analysis. For those remaining proteins, we then applied a constant-value imputation approach: for protein levels that were below the limit of quantification (LOQ), a value lower  $LOQ \times 0.5$  was imputed using the LOQ of the respective protein. For those above the LOQ, the upper LOQ value was used. By removing high-missingness variables prior to imputation with constant values, we reduce the likelihood of misrepresenting the relationship between the measured proteins and injury outcome that were the result from technical limitations of the assay.

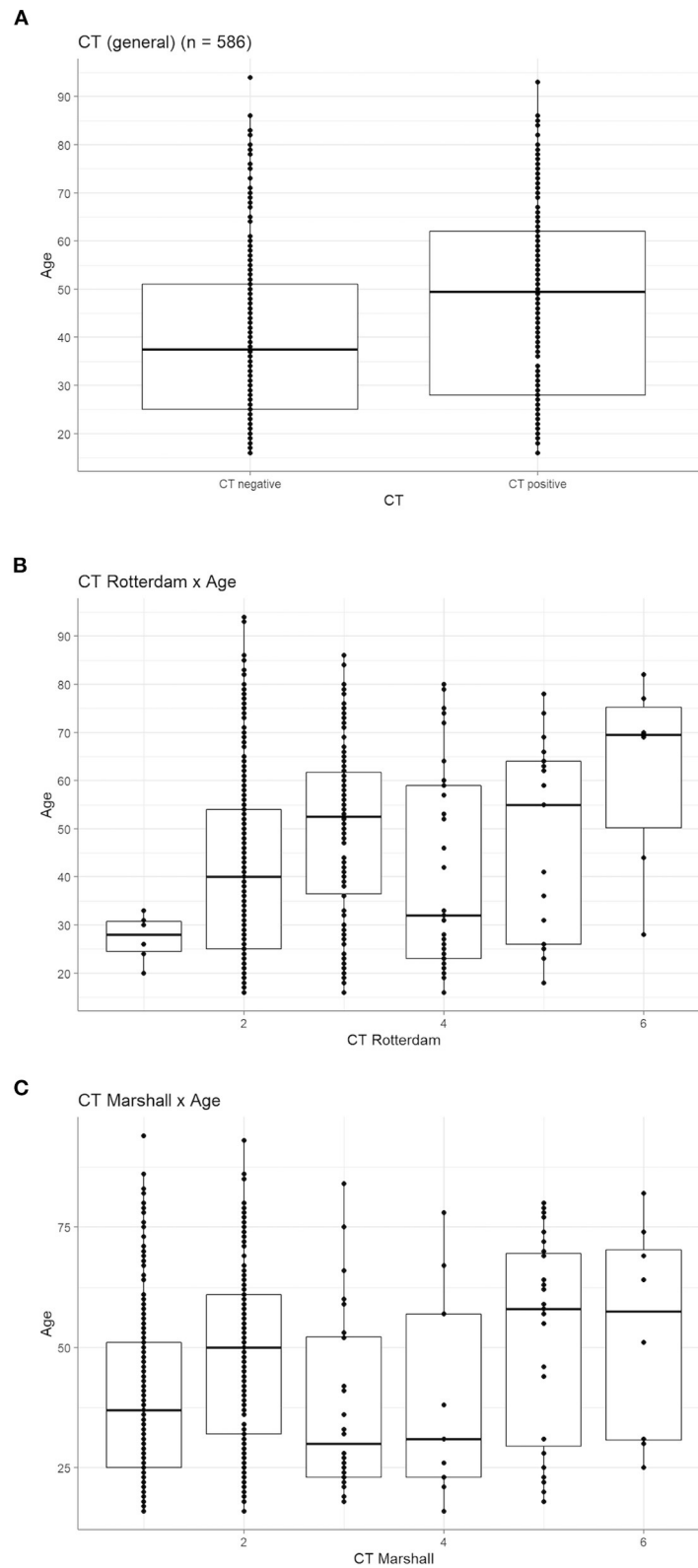
## Statistics and Data Visualization

All analyses were performed using the open-source programming language R (24) (version 3.6.3). The *tidyverse* (25) R package was utilized for dataframe manipulation and data wrangling. For general data visualization and 3d surface plots, the *ggplot* (26) and *plotly* (27) R packages were utilized. Missing data analysis and visualization were generated using the *naniar* (28) R package. After biomarkers that exceeded the threshold of missingness were removed, the protein biomarker data were corrected for batch effects by within-batch *z*-score standardization. Using the “*princals*” function from the *Gifi* (29) R package, we then performed non-linear principal component analysis (NL-PCA) on the batch-corrected protein biomarker data to reduce high dimensionality. Specifically, NL-PCA was performed to allow for 2° splines and 57 total principal components (PCs) in the solution (matching the 57 biomarkers used as input to the NL-PCA). Permutation tests were then run using the *syndRomics* (30) R package to identify the PCs with variance accounted for (VAF) and standardized loadings significantly above chance (adjusted *p*-value > 0.05). Bar maps were additionally generated using the *syndRomics* R package. Hypothesis testing was conducted using linear regression for continuous outcome variables or logistic regression for dichotomized outcome variables, and Welch's unequal variance *t*-test for two group differences. Regression models were produced using the “*lm*” and “*glm*” functions from the *stats* (24) R package with additional type-III analysis of variance (ANOVA) performed to derive significance of main effects and interactions using the *car* (31) R package. To determine the confidence interval of the marginal effect of P2 on the change in GOS-E outcome between 3 and 6 months post-injury, we utilized the “*deltamethod*” function from the *msm* (32) R package.

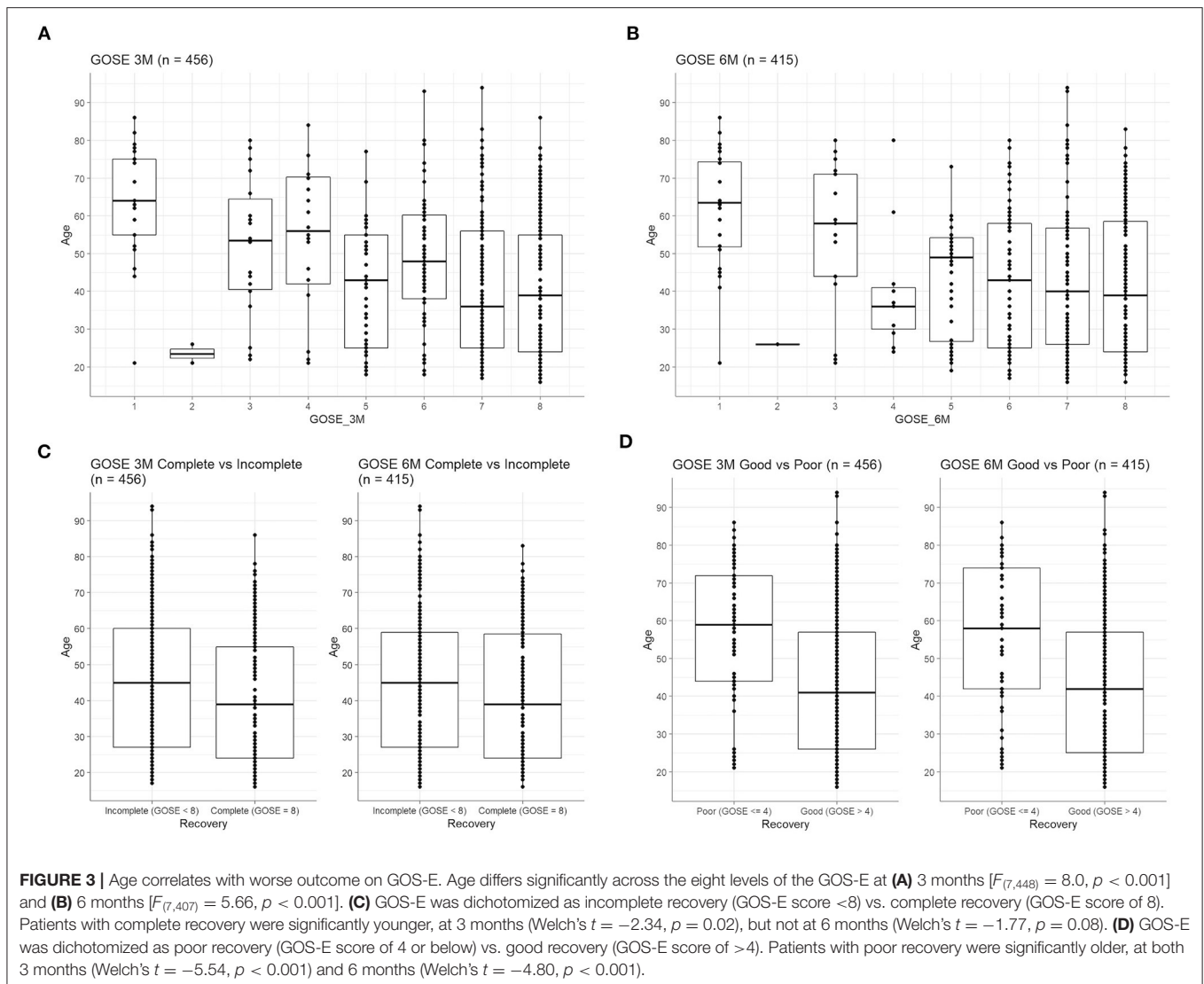
## RESULTS

### Age Correlates With Worse GOS-E Outcome and CT-Positive TBI

The first aim was to determine whether age predicts CT pathology and/or outcome on GOS-E at 3 and 6 months post-injury. Results showed that the patient population with positive CTs was significantly older than those who were CT negative (Figure 2A, Welch's  $t = 5.61$ ,  $p < 0.01$ ). Rotterdam scores also differed significantly by age [Figure 2B,  $F_{(1,584)} = 17.34$ ,  $p < 0.001$ ] as did Marshall scores [Figure 2C,  $F_{(1,584)} = 18.25$ ,  $p <$



**FIGURE 2 |** Age correlates with worse CT pathology. **(A)** Age is significantly higher in those with a positive diagnosis of CT pathology (Welch's  $t = 5.61$ ,  $p < 0.0$ ). **(B)** Age significantly differed across the six levels of the Rotterdam CT score [ $F_{(1,584)} = 17.34$ ,  $p < 0.001$ ] and the **(C)** six levels of the Marshall CT score [ $F_{(1,584)} = 18.25$ ,  $p < 0.001$ ]. For both scores, the two highest scores (indicative of more severe injury) have a higher median age than all other CT scores.



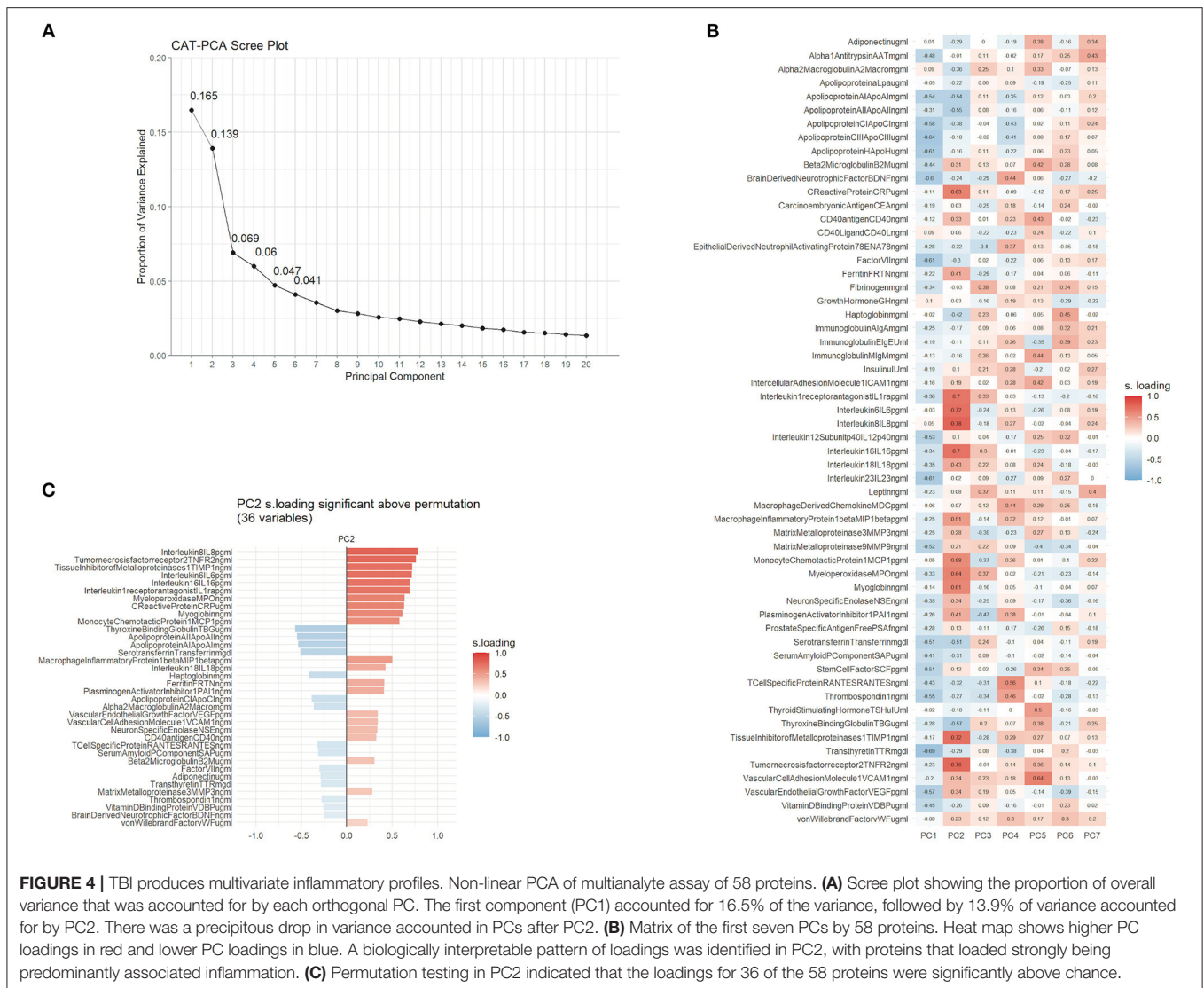
0.001]: patients with higher scores (i.e., more severe injury by CT) on both measures tended to be older.

Worse functional outcome as captured by lower GOS-E scores was also correlated with older patient age at both 3 months [Figure 3A,  $F_{(7,448)} = 8.0, p < 0.001$ ] and 6 months post-injury [Figure 3B,  $F_{(7,407)} = 5.66, p < 0.001$ ]. Similarly, patients with complete recovery at 3 months were significantly younger (Figure 3C, Welch's  $t = -2.34, p = 0.02$ ), but not at 6 months (Figure 3C, Welch's  $t = -1.77, p = 0.08$ ). When GOS-E was dichotomized as poor recovery ( $\leq 4$ ) or good recovery ( $> 4$ ), those with good recovery were also significantly younger (Figure 3D) at both 3 months (Welch's  $t = -5.54, p < 0.001$ ) and 6 months (Welch's  $t = -4.80, p < 0.001$ ).

## TBI Produces Multivariate Inflammatory Profile

In order to reuse TRACK-TBI Pilot data to test whether age impacts proteomic expression, we first sought to replicate and

improve upon prior multivariate proteomic analyses from this dataset (10). While the prior analysis used a linear PCA with optimal scaling, the current model used a non-linear approach as well as handled missing data above and below the limit of quantification differently than the prior published work (see Methods). Here, a permutation test for significant PCs was additionally used to objectively determine the number of significant principal components. This permutation test of VAF revealed that the first six PCs accounted for variance above chance. A second permutation test of significant loadings found that 33 proteins in PC1 and 35 proteins in PC2 loaded strongly enough to be above chance (adjusted  $p$ -value  $< 0.05$ ). Although the VAF permutation determined that six PCs accounted for variance above chance (Figure 4A), no further PCs after PC2 had a protein loading that was significant (Figure 4B), suggesting that further analysis and interpretation should be limited to PC1 and PC2 only. The first two PCs together accounted for 28.7% of the variance (16.4% variance accounted for in PC1



**FIGURE 4 |** TBI produces multivariate inflammatory profiles. Non-linear PCA of multianalyte assay of 58 proteins. **(A)** Scree plot showing the proportion of overall variance that was accounted for by each orthogonal PC. The first component (PC1) accounted for 16.5% of the variance, followed by 13.9% of variance accounted for by PC2. There was a precipitous drop in variance accounted in PCs after PC2. **(B)** Matrix of the first seven PCs by 58 proteins. Heat map shows higher PC loadings in red and lower PC loadings in blue. A biologically interpretable pattern of loadings was identified in PC2, with proteins that loaded strongly being predominantly associated inflammation. **(C)** Permutation testing in PC2 indicated that the loadings for 36 of the 58 proteins were significantly above chance.

and 12.3% variance accounted for in PC2). PC2 revealed a biologically interpretable loading pattern that strongly resembled the pattern seen in Huie et al. (10). The proteins with the strongest positive loadings on PC2 included pro-inflammatory proteins IL-8, IL-6, TIMP1, and C-reactive protein, while those that loaded in the opposite direction included anti-inflammatory or protective proteins including apolipoproteins AI, AII, and CI, as well as BDNF and serotransferrin (Figure 4C). Taken as a whole, the pattern of strong loading proteins indicates a broadly inflammatory profile, as seen in the previously published PCA of the same protein assay (10).

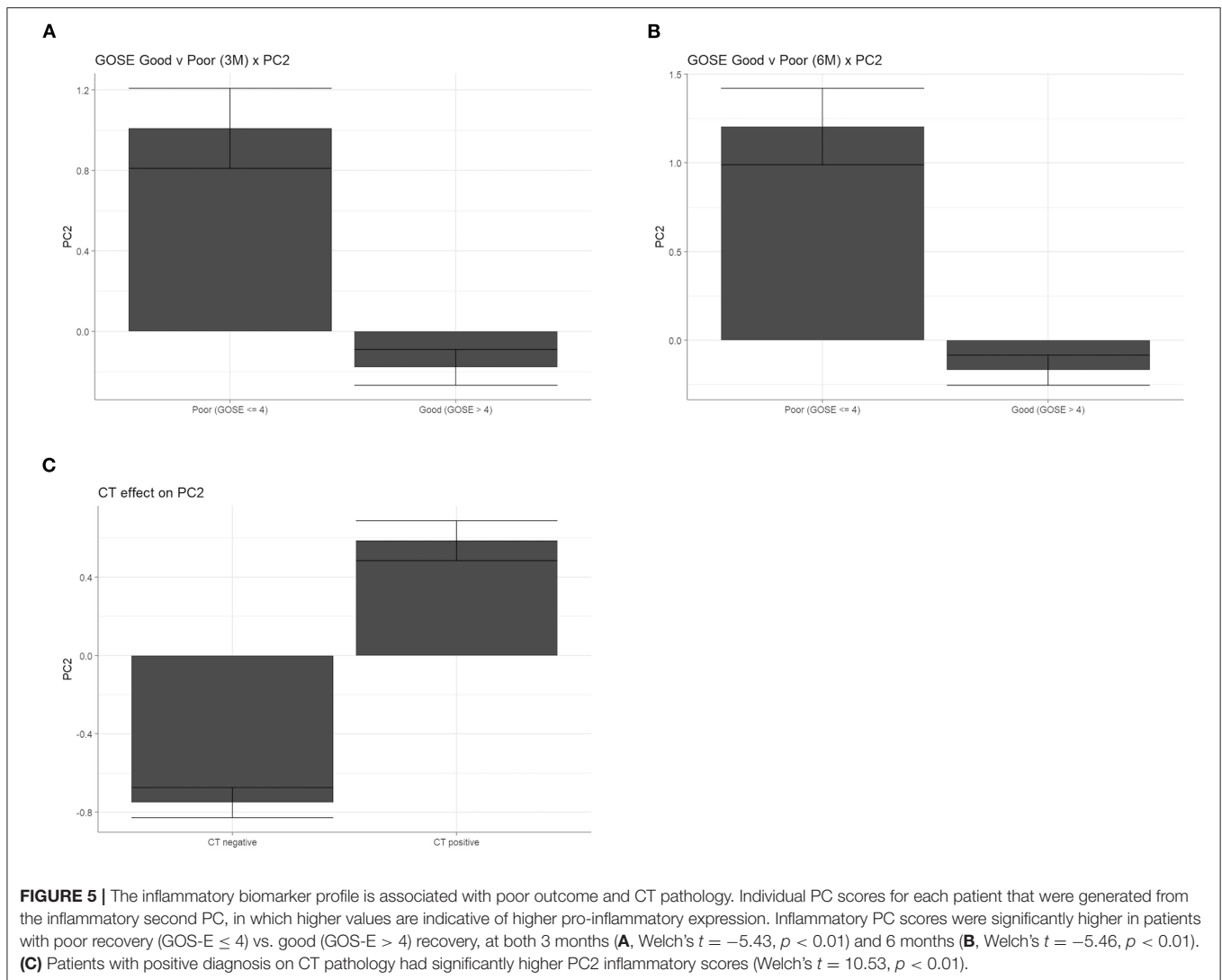
## Inflammatory Biomarker Profile Is Associated With Poor Outcome and CT-Positive TBI

The inflammatory biomarker profile revealed by proteomic PC2 is represented by PC scores for each patient, with higher scores indicating a more inflammatory proteomic composition and

lower scores indicating an anti-inflammatory composition. Using this score as an injury response measure, we then determined whether patients with different outcomes expressed different inflammatory biomarker values. Results showed that patients with poor (GOS-E  $\leq 4$ ) recovery had significantly higher inflammatory PC scores than those with good (GOS-E  $> 4$ ) recovery at both 3 months (Figure 5A, Welch's  $t = -5.43$ ,  $p < 0.01$ ) and 6 months (Figure 5B, Welch's  $t = -5.46$ ,  $p < 0.01$ ). Similarly, patients with CT-positive TBI had significantly higher inflammatory PC scores than those who were CT negative (Figure 5C, Welch's  $t = 10.53$ ,  $p < 0.01$ ). These findings replicate those seen in Huie et al. (10).

## Age Moderates the Effect of Inflammation on Outcome After TBI

To determine whether the association between inflammation and outcome is affected by age, we first asked whether there is an overall association between age and the multivariate



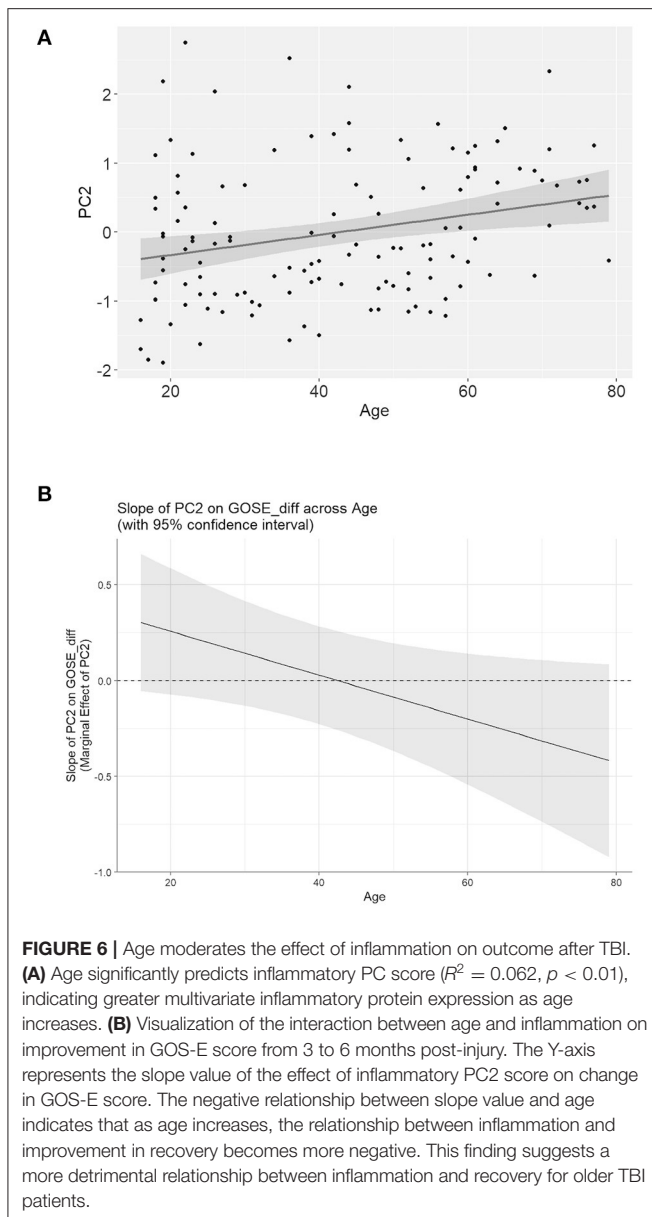
inflammatory biomarker panel. We found that age significantly correlates with inflammatory PC score, with a higher score (indicating a more pro-inflammatory profile) as patient age increases (**Figure 6A**,  $R^2 = 0.062$ ,  $p < 0.01$ ). Next, the goal was to determine whether age interacts with the inflammatory PC score on GOS-E outcome. There was no significant interaction between age and inflammatory PC score for good/poor or complete/incomplete GOS-E recovery at either 3 or 6 months. Interestingly, there was a significant interaction between age and inflammatory PC score on the *change* in GOS-E between 3 and 6 months (age  $\times$  PC score, LR chi-square = 5.31,  $p = 0.02$ ) even after controlling for CT positivity. To better understand the nature of this interaction, we explored how the effect of inflammation on GOS-E improvement might differ across age. **Figure 6B** shows a decrease in the slope of the relationship between inflammatory PC score and GOS-E improvement as age increases. This finding suggests that in younger patients, an improvement in GOS-E is associated with

increased inflammation; conversely, increased inflammation in older patients is associated with less robust recovery between 3 and 6 months post-injury.

## DISCUSSION

This work aimed to illustrate the utility of data reuse for insight discovery. The governing principles of data management and stewardship are based on data being FAIR: findable, accessible, interoperable, and reusable. By harmonizing subsets of data from previously published work, there is an opportunity to test new hypotheses, improve on prior techniques, and gain knowledge that may otherwise be missed. The reuse of data in new ways also accelerates the process of discovery, allowing for novel recombination of data that maximizes the return on the costly investment of data collection and curation. The reuse and recombination of shared data will ultimately bring light to the “long tail of dark data” (5, 10).





The results of this data reuse reproduced and expanded on prior work showing that recovery after TBI is often age dependent. We found that patients who exhibited complete recovery (GOS-E = 8) at 3 and 6 months after TBI were significantly younger than those with incomplete recovery. Similarly, those with a GOS-E of 4 and below were significantly older than those with a GOS-E > 4. Of course, the likelihood of confounding comorbidities is greater as age increases, and older age in and of itself may not be a causal factor. We and others have previously shown that even when controlling for possible confounders, poor outcome becomes more likely with age. Gardner et al. showed that for those over 55, moderate to severe TBI significantly increases the risk of developing dementia, and for those over 65, even mild TBI increased this risk. This result was seen despite adjustments for common predictors and

compared the TBI patients to non-TBI trauma controls, rather than the general population, to account for the possibility of reverse causality (e.g., early stage dementia leading to TBI in the first place) (15). Similarly, de la Plata et al. found that greater decline (as measured by the Disability Rating Scale) was seen in older TBI patients compared to younger TBI patients (14). Importantly, they noted that although age was a significant factor in decline after TBI, a relatively small proportion of the variance in the Disability Rating Scale was explained by age alone and that other factors and comorbidities may have a greater influence. Nevertheless, a large body of work has made it clear that age is an important factor that must be accounted for to understand the trajectory of recovery after TBI, especially since the elderly make up the highest proportion of TBI-related hospitalizations (33).

Given the prominent roles of both age and inflammation on TBI prognosis, we sought to understand how these important factors may interact to affect recovery. We found that even when controlling for CT positivity, there remained a significant interaction between age and the multivariate inflammatory profile on improvement in GOS-E over time. Interestingly, this interaction appears to be characterized by a decrease in the value of the slope of the relationship between inflammation and GOS-E improvement across ages (**Figure 6B**). This finding suggests that for younger TBI patients, an increase in inflammation is associated with *greater* functional improvement, while in older patients, increased inflammation tends to predict less improvement. Previous work supports the notion that the effect of inflammation on recovery may be age dependent. Pre-clinical study of age and inflammation found that the infiltration of inflammatory macrophages to the central nervous system (CNS) after experimental TBI was significantly increased in older animals relative to younger animals (34). Similarly, we found that in a TBI aging model, age-dependent inflammation was a significant contributor to cognitive impairment (17). Recent work using deep machine learning has derived an algorithm that uses inflammatory biomarkers to determine one's "inflammatory age," based on the idea that the increase in chronic inflammation is an inevitable consequence of aging (35). This concept is related to the notion of "immunosenescence," the dysregulation of immune function due to aging (36, 37). From this perspective, the immune challenge presented by a TBI in aging patients may work to undermine recovery by inducing detrimental inflammation that is not as easily resolved as in younger patients.

It is important to note that the interaction between age and inflammation on recovery found here was highly variable, as seen in the wide confidence intervals in **Figure 5B**. Thus, it is likely that a number of other confounders may account for the large amount of variance observed. For instance, given that the interaction of inflammation (measured within 24 h of injury) with age appears to have an effect on the course of recovery from 3 to 6 months, another source of variance may be the regulation of the inflammatory proteomic response over time.

Future studies will be needed to determine the static and dynamic mitigating factors that are likely at work. Similarly, as with many clinical datasets, generalization may be limited by blind spots of underrepresented groups among those enrolled, including those patients who did not warrant ED assessment at a

high-level trauma center. It is also worth noting that although the unsupervised, data-driven approach of non-linear PCA revealed a predominantly inflammatory “identity” on PC2, the limitations of the assay were such that other proteins that would typically be candidate biomarkers for inflammation were found to be among the 15 proteins that were outside the level of detection, including TNF alpha, IL1-B, and macrophage inflammatory protein alpha. Thus, future work that is meant to specifically target the role of inflammation will require a more hypothesis-driven approach, with the *a priori* selection of biomarkers specific to inflammation. This kind of limitation may be typical for data reuse in the future and should be recognized and addressed: in the repurposing of prior data, there exists a possibility that the data may not always be ideal for comprehensively addressing a scientific question. But as long as the methodology and results are rigorous and transparent, building upon such findings can be undertaken in good faith.

## DATA AVAILABILITY STATEMENT

The data analyzed in this study is subject to the following licenses/restrictions: a subset of the data is publicly available at <https://doi.org/10.1371/journal.pone.0169490.s011>. Access to corresponding proteomics data will be granted upon approved request. Requests to access these datasets should be directed to [https://tracktbi.ucsf.edu/sites/tracktbi.ucsf.edu/files/TRACK-TBI%20Collaboration%20Proposal%20Request%20Form\\_0.pdf](https://tracktbi.ucsf.edu/sites/tracktbi.ucsf.edu/files/TRACK-TBI%20Collaboration%20Proposal%20Request%20Form_0.pdf).

## ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the University of California San Francisco

(UCSF) Institutional Review Board of record, the Committee on Human Research (CHR). All subjects gave written informed consent in accordance with the Declaration of Helsinki. The protocol was approved by the UCSF CHR #10-00011.

## AUTHOR CONTRIBUTIONS

JH, AC, AT-E, JN, and AF contributed to data curation and analysis. All authors contributed to interpretation of results, writing, and revising manuscript.

## FUNDING

This study was supported by National Institutes of Health Grants: R01NS088475 (AF); R01NS122888 (AF); UH3NS106899 (AF); and U24NS122732 (AF); Department of Veterans Affairs: 1I01RX002245 (AF) and I01RX002787 (AF); and Wings for Life Foundation and Craig H. Neilsen Foundation (AF), NIH/NINDS F32NS117728 (AC).

## ACKNOWLEDGMENTS

The authors thank Amy J. Markowitz, JD for editorial support and also like to thank the following TRACK-TBI Investigators: Pratik Mukherjee, MD PhD, University of California, San Francisco; Ross Zafonte, DO, Harvard/Spaulding Rehabilitation Hospital; Michael A. McCrea, PhD, Medical College of Wisconsin; Mary Vassar, RN MS, University of California, San Francisco; Neeraj Badjatia, MD, University of Maryland, Baltimore; and Claudia Robertson, MD, Baylor College of Medicine.

## REFERENCES

- Norgeot B, Glicksberg BS, Butte AJ. A call for deep-learning healthcare. *Nat Med.* (2019) 25:14–5. doi: 10.1038/s41591-018-0320-3
- Huie JR, Almeida CA, Ferguson AR. Neurotrauma as a big-data problem. *Curr Opin Neurol.* (2018) 31:702–8. doi: 10.1097/WCO.0000000000000614
- Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data.* (2016) 3:160018. doi: 10.1038/sdata.2016.18
- Hawkins BE, Huie JR, Almeida C, Chen J, Ferguson AR. Data dissemination: shortening the long tail of traumatic brain injury dark data. *J Neurotrauma.* (2020) 37:2414–23. doi: 10.1089/neu.2018.6192
- Ferguson AR, Nielson JL, Cragin MH, Bandrowski AE, Martone ME. Big data from small data: data-sharing in the “long tail” of neuroscience. *Nat Neurosci.* (2014) 17:1442–7. doi: 10.1038/nn.3838
- Wallis JC, Rolando E, Borgman CL. If we share data, will anyone use them? Data sharing and reuse in the long tail of science and technology. *PLoS ONE.* (2013) 8:e67332. doi: 10.1371/journal.pone.0067332
- Yue JK, Vassar MJ, Lingsma HF, Cooper SR, Okonkwo DO, Valadka AB, et al. Transforming research and clinical knowledge in traumatic brain injury pilot: multicenter implementation of the common data elements for traumatic brain injury. *J Neurotrauma.* (2013) 30:1831–44. doi: 10.1089/neu.2013.2970
- Manley GT, Maas AIR. Traumatic brain injury: an international knowledge-based approach. *JAMA.* (2013) 310:473–4. doi: 10.1001/jama.2013.169158
- Manley GT, Mac Donald CL, Markowitz AJ, Stephenson D, Robbins A, Gardner RC, et al. The traumatic brain injury endpoints development (TED) initiative: progress on a public-private regulatory collaboration to accelerate diagnosis and treatment of traumatic brain injury. *J Neurotrauma.* (2017) 34:2721–30. doi: 10.1089/neu.2016.4729
- Huie JR, Diaz-Arrastia R, Yue JK, Sorani MD, Puccio AM, Okonkwo DO, et al. Testing a multivariate proteomic panel for traumatic brain injury biomarker discovery: a Track-TBI pilot study. *J Neurotrauma.* (2019) 36:100–10. doi: 10.1089/neu.2017.5449
- Nielson JL, Cooper SR, Yue JK, Sorani MD, Inoue T, Yuh EL, et al. Uncovering precision phenotype-biomarker associations in traumatic brain injury using topological data analysis. *PLoS ONE.* (2017) 12:e0169490. doi: 10.1371/journal.pone.0169490
- Huie JR. The 3rd joint symposium of the international and national neurotrauma societies and AANS/CNS Section on neurotrauma and critical care august 11-16, 2018 Toronto, Canada. *J Neurotrauma.* (2018). 35:A-154. doi: 10.1089/neu.2018.29013.abstracts
- Gardner RC, Dams-O'Connor K, Morrissey MR, Manley GT. Geriatric traumatic brain injury: epidemiology, outcomes, knowledge gaps, and future directions. *J Neurotrauma.* (2018) 35:889–906. doi: 10.1089/neu.2017.5371
- Marquez de la Plata CD, Hart T, Hammond FM, Frol AB, Hudak A, Harper CR, et al. Impact of age on long-term recovery from traumatic brain injury. *Arch Phys Med Rehabil.* (2008) 89:896–903. doi: 10.1016/j.apmr.2007.12.030
- Gardner RC, Burke JF, Nettiksimmons J, Kaup A, Barnes DE, Yaffe K. Dementia risk after traumatic brain injury vs. nonbrain trauma: the role of age and severity. *JAMA Neurol.* (2014) 71:1490–7. doi: 10.1001/jamaneurol.2014.2668

16. Gardner RC, Burke JF, Nettiksimmons J, Goldman S, Tanner CM, Yaffe K. Traumatic brain injury in later life increases risk for Parkinson disease. *Ann Neurol.* (2015) 77:987–95. doi: 10.1002/ana.24396
17. Chou A, Krukowski K, Morganti JM, Riparip L-K, Rosi S. Persistent infiltration and impaired response of peripherally-derived monocytes after traumatic brain injury in the aged brain. *Int J Mol Sci.* (2018) 19:1616. doi: 10.3390/ijms19061616
18. Krukowski K, Chou A, Feng X, Tired B, Paladini M-S, Riparip L-K, et al. Traumatic brain injury in aged mice induces chronic microglia activation, synapse loss, and complement-dependent memory deficits. *Int J Mol Sci.* (2018) 19:3753. doi: 10.3390/ijms19123753
19. Thamisetty SS, Pedragosa J, Weng YC, Calon F, Planas A, Kriz J. Age-related deregulation of TDP-43 after stroke enhances NF- $\kappa$ B-mediated inflammation and neuronal damage. *J Neuroinflammation.* (2018) 15:1–15. doi: 10.1186/s12974-018-1350-y
20. Zhang B, Bailey WM, Braun KJ, Gensel JC. Age decreases macrophage IL-10 expression: implications for functional recovery and tissue repair in spinal cord injury. Age decreases macrophage IL-10 expression: implications for functional recovery and tissue repair in spinal cord injury. *Exp Neurol.* (2015) 273:83–91. doi: 10.1016/j.expneurol.2015.08.001
21. Hooshmand MJ, Galvan MD, Partida E, Anderson AJ. Characterization of recovery, repair, and inflammatory processes following contusion spinal cord injury in old female rats: is age a limitation? *Immun Ageing.* (2014) 11:1–13. doi: 10.1186/1742-4933-11-15
22. Manley GT, Diaz-Arrastia R, Brophy M, Engel D, Goodman C, Gwinn K, et al. Common data elements for traumatic brain injury: recommendations from the biospecimens and biomarkers working group. *Arch Phys Med Rehabil.* (2010) 91:1667–72. doi: 10.1016/j.apmr.2010.05.018
23. Diaz-Arrastia R, Wang KK, Papa L, Sorani MD, Yue JK, Puccio AM, et al. Acute biomarkers of traumatic brain injury: relationship between plasma levels of ubiquitin C-terminal hydrolase-L1 and glial fibrillary acidic protein. *J Neurotrauma.* (2014) 31:19–25. doi: 10.1089/neu.2013.3040
24. R Core Team. *R: A Language and Environment for Statistical Computing.* Vienna: R Foundation for Statistical Computing (2020).
25. Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, et al. Welcome to the Tidyverse. *J Open Source Software.* (2019) 4:1686. doi: 10.21105/joss.01686
26. Wickham H. *ggplot2: Elegant Graphics for Data Analysis.* New York, NY: Springer International Publishing (2016).
27. Sievert C. *Interactive Web-based Data Visualization with R, Plotly, and Shiny* Boca Raton, FL: CRC Press (2020).
28. Tierney NJ, Cook DH. Expanding tidy data principles to facilitate missing data exploration, visualization and assessment of imputations. *arXiv preprint arXiv:1809.02264* (2018).
29. Mair P, de Leeuw J. A general framework for multivariate analysis with optimal scaling: the R package aspect. *J Stat Softw.* (2010) 32:1–23.
30. Torres Espín A, Chou A, Huie JR, Kyritsis N, Upadhyayula PS, Ferguson AR. Reproducible analysis of disease space via principal components using the novel R package syndRomics. *Elife.* (2021) 10:e61812. doi: 10.7554/eLife.61812.sa2
31. Fox J, Weisberg S. *An {R} Companion to Applied Regression* (2019). Available online at: <https://socialsciences.mcmaster.ca/jfox/Books> (accessed October 18, 2021).
32. Jackson C. Multi-state models for panel data: the msm package for R. *J Stat Softw.* (2011) 38:1–28. doi: 10.18637/jss.v038.i08
33. Thurman DJ, Alverson C, Dunn KA, Guerrero J, Sniezek JE. Traumatic brain injury in the United States: a public health perspective. *J Head Trauma Rehabil.* (1999) 14:602–15. doi: 10.1097/00001199-199912000-00009
34. Morganti JM, Riparip L-K, Chou A, Liu S, Gupta N, Rosi S. Age exacerbates the CCR2/5-mediated neuroinflammatory response to traumatic brain injury. *J Neuroinflammation.* (2016) 13:1–12. doi: 10.1186/s12974-016-0547-1
35. Sayed N, Huang Y, Nguyen K, Krejciova-Rajaniemi Z, Grawe AP, Gao T, et al. An inflammatory aging clock (iAge) based on deep learning tracks multimorbidity, immunosenescence, frailty and cardiovascular aging. *Nat Aging.* (2021) 1:598–615. doi: 10.1038/s43587-021-00082-y
36. Goronzy JJ, Weyand CM. Understanding immunosenescence to improve responses to vaccines. *Nat Immunol.* (2013) 14:428–36. doi: 10.1038/ni.2588
37. Shaw AC, Goldstein DR, Montgomery RR. Age-dependent dysregulation of innate immunity. *Nat Rev Immunol.* (2013) 13:875–87. doi: 10.1038/nri3547

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Huie, Chou, Torres-Espín, Nielson, Yuh, Gardner, Diaz-Arrastia, Manley, Ferguson and the TRACK-TBI Investigators. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.