DATA MANAGEMENT AND SHARING PLAN

Element 1: Data Type

A. Types and amount of scientific data expected to be generated in the project:
This study will generate measures of gene expression, protein synthesis, luminescence, biochemical composition from genetically modified iPSC cell-seeded gel constructs. Additionally, similar data will be recorded on similar constructs implanted in mouse models of RA, along with locomotor, pain/behavioral measurements, serum biomarkers, clinical scores, microCT, and histologic analysis of disease severity. From past experience, the amount of data will be approximately 100 GB, mostly arising from microCT and histology images.

B. Scientific data that will be preserved and shared, and the rationale for doing so:
Shared data generated from this project will be made available no later than the time of publication or the end of the funding period, whichever comes first. The duration of preservation and sharing the data will be a minimum of 10 year after the end of the funding period.

C. Metadata, other relevant data, and associated documentation:
Metadata and study protocols will be linked to the data files at the time of publication, as we have done previously in our recent studies in Nature Communications, Science Advances, eLife, etc. Detailed documentation of the data, including the study design, data collection methods, and any processing or cleaning that occurs, will be made available to other researchers along with the data.

Element 2: Related Tools, Software and/or Code:

The raw data for the imaging analysis will need ImageJ and code in Python to generate statistical analysis of the variables and the code will also be available. Additional code used for analysis will be uploaded to Github, as we have done in recent studies.

Element 3: Standards:

Data will be stored in common and open formats such as csv, MS-Word, PDF, or TIFF data for images. Information needed to make use of this data as bioluminescence time series, histology images, or microCT series, along with references to the sources will be included wherever applicable.

Element 4: Data Preservation, Access, and Associated Timelines

A. Repository where scientific data and metadata will be archived:
All data will be deposited upon collection to the Washington University Box starting immediately after the award begins. Read only access to the public will be provided by the WU Box permission platform. Datasets will be identified as complete versus in progress for clarity. Once published, data will be uploaded to Dryad and linked to the publication for access. All computer code will be deposited on Github, and similarly linked within the publication using html identifier.

B. How scientific data will be findable and identifiable:
We will use Persistent Unique Identifiers (PIDs) to improve data findability across all outputs. PIDs used will include people's ID, DOIs for outputs (e.g. datasets).

C. When and how long the scientific data will be made available:
Shared data generated from this project will be made available no later than the time of publication or the end of the funding period, whichever comes first. The duration of preservation and sharing the data will be a minimum of 10 year after the end of the funding period.

Element 5: Access, Distribution, or Reuse Considerations

A. Factors affecting subsequent access, distribution, or reuse of scientific data:
There are no anticipated factors or limitations that will affect access, distribution or reuse of the scientific data generated by the proposal.

B. Whether access to scientific data will be controlled:
Controlled access will not be used. The data that is shared will be shared by unrestricted download.

C. Protections for privacy, rights, and confidentiality of human research participants:
No human subjects are involved in this study.

Element 6: Oversight of Data Management and Sharing:

Data Management Training: All members of the research team will receive training on data management and sharing, and will be familiar with the NIH's DMS policy and requirements, describe above. Oversight will include the following points.
Data Collection and Storage: All data collected during the course of the study will be securely stored on a password-protected server, with regular backups to ensure data integrity.
Data Access and Sharing: Data will be made available to all members of the research team and any collaborators as needed for the conduct of the study. Data will also be made available to other researchers upon request, subject to ethical and legal considerations, such as participant confidentiality and informed consent.
Data Management: The principal investigators will be responsible for maintaining accurate and up-to-date records of all data collected, including documentation of any data cleaning or processing that occurs. Data management plans will be regularly reviewed and updated as necessary.
Oversight: A committee of 3 uninvolved investigators will examine the processes and protocols outlined here on an annual basis to ensure appropriate oversight of data management and sharing.