Neurobiology of Disease 123 (2019) 127–136

Contents lists available at ScienceDirect

Neurobiology of Disease

journal homepage: www.elsevier.com/locate/ynbdi

Review

Big data sharing and analysis to advance research in post-traumatic epilepsy

Dominique Duncan[a,*], Paul Vespa[b], Asla Pitkänen[c], Adebayo Braimah[a], Niina Lapinlampi[c], Arthur W. Toga[a]

[a] Laboratory of Neuro Imaging, USC Stevens Neuroimaging and Informatics Institute, Keck School of Medicine of USC, University of Southern California, Los Angeles, CA, USA
[b] Division of Neurosurgery and Department of Neurology, University of California at Los Angeles School of Medicine, Los Angeles, CA, USA
[c] A.I. Virtanen Institute for Molecular Sciences, University of Eastern Finland, Kuopio, Finland

ARTICLE INFO

Keywords:
Biomarkers
EEG
Epilepsy
Epileptogenesis
Informatics
MRI
Neuroimaging
TBI

ABSTRACT

We describe the infrastructure and functionality for a centralized preclinical and clinical data repository and analytic platform to support importing heterogeneous multi-modal data, automatically and manually linking data across modalities and sites, and searching content. We have developed and applied innovative image and electrophysiology processing methods to identify candidate biomarkers from MRI, EEG, and multi-modal data. Based on heterogeneous biomarkers, we present novel analytic tools designed to study epileptogenesis in animal model and human with the goal of tracking the probability of developing epilepsy over time.

1. Introduction

The goal of the Epilepsy Bioinformatics Study for Antiepileptogenic Therapy (EpiBioS4Rx) is to identify relevant biomarkers of epileptogenesis after traumatic brain injury (TBI) and perform rigorous preclinical trials that permit the future design and performance of economically feasible full-scale clinical trials of antiepileptogenic therapies. A fundamental challenge in discovering these biomarkers of epileptogenesis is that this process is multifactorial and crosses multiple modalities. Rather than considering one type of data, we have been collecting and analyzing multi-modal data, including neuroimaging, electrophysiology, and molecular/serological/tissue. Furthermore, to facilitate analysis and collaboration among scientists from various centers around the world, we have created the informatics infrastructure needed for a large dataset of this size. We have also developed innovative analytic tools that are shared with the broader epilepsy research community, including any other interested researchers outside of the epilepsy research community who might find these data useful for their research, so that others may use our tools in addition to their own tools to advance research in this field in general, in addition to identifying biomarkers of epileptogenesis after TBI.

Investigators must have access to a large number of high quality, well-curated data points and study subjects in order for biomarker signals to be detectable above the noise inherent in complex phenomena, such as epileptogenesis, TBI, and conditions of data collection. Additionally, data generating and collecting sites are spread worldwide among different laboratories, clinical sites, heterogeneous data types, and formats, and across multi-center preclinical trials. Before the data can even be analyzed, a central platform is needed to standardize these data and provide tools for searching, viewing, annotating, and analyzing them. By centralizing an enduring data archive, biobank, and analytic tools, researchers may identify and validate biomarkers of epileptogenesis in studies using various types of data. Beyond creating a centralized data repository, we have pioneered innovative standardization/co-registration references, fully supported by novel image and electrophysiology processing methods to extract candidate biomarkers from the diverse data. Not only does a well-curated and standardized multi-modal dataset facilitate the development of models of epileptogenesis, but it also ensures that such models are statistically significant and can be validated.

1.1. EEG and MRI databases

There have been other efforts to create centralized data archives, but it has proven to be especially challenging for human neurophysiological data for many reasons, such as large file sizes, varying formats, privacy constraints, and funding. Two examples of centralized EEG databases that have been developed include Epilepsiae (Dourado et al., 2009; Ihle et al., 2012; Klatt et al., 2012; Schulze-Bonhage et al., 2010), a European Union-funded project, and IEEG.ORG, an NINDS-

* Corresponding author at: 2025 Zonal Ave. 208B, Los Angeles, CA 90033, USA.
E-mail address: dduncan@loni.usc.edu (D. Duncan).

https://doi.org/10.1016/j.nbd.2018.05.026
Received 16 April 2018; Received in revised form 24 May 2018; Accepted 31 May 2018
Available online 01 June 2018
0969-9961/

funded cloud-based platform (Kini et al., 2016). Epilepsiae stores recordings from 275 individuals with epilepsy, with a total recording time of more than 40,000 h. Investigators can export the data locally for analysis. IEEG.ORG (Brinkmann et al., 2009; Kini et al., 2016; Wagenaar et al., 2015) hosts academic and clinical datasets of scalp and intracranial EEG, just over 800 of which are shared publicly, from both animal models of epilepsy and patients. This platform uses Amazon cloud services. Access for Epilepsiae is restricted to scientific groups that financially contribute to the maintenance of the database, which has resulted in fewer people using the platform. IEEG.ORG is free and accessible to the epilepsy research community.

The Laboratory of Neuro Imaging (LONI) at the University of Southern California (USC) has experience with many major clinical consortia and big data projects, such as the Biomedical Informatics Resource Network (BIRN) (Astakhov et al., 2005; Helmer et al., 2011), the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Toga and Crawford, 2010a), the Michael J. Fox Foundation's Parkinson's Progressive Markers Initiative (PPMI) (Marek et al., 2011), the Human Connectome Project (HCP) (Marcus et al., 2013; Van Essen et al., 2013), the Big Data to Knowledge (BD2K) (Bourne et al., 2015; Margolis et al., 2014) program, the NIH Autism Centers of Excellence (ACE) (Rakap et al., 2015) program, The Enhancing Neuroimaging Genetics through Meta-Analysis (ENIGMA) Consortium (Thompson et al., 2014), and the Global Alzheimer's Association Interactive Network (GAAIN) (Toga et al., 2016), among others. The challenges that we have faced in these projects and the knowledge and experience that we have gained from them have informed and guided us to design an optimal platform to focus on clinical and preclinical TBI data in EpiBioS4Rx.

The number of large databases and related neurological disease-focused consortia around the world has grown rapidly in recent years (Lim, 2014), which demonstrates the importance of transparency in large-scale projects and the sharing of data that are collected. The larger datasets from preclinical studies, such as the one generated in EpiBioS4Rx are now emerging. Beyond sharing data, to encourage the most impactful outside collaborations and scientific discoveries, the data must be well organized and annotated (i.e., for EEG). Furthermore, the data sharing platform must be user friendly and straightforward to use. What makes our project unique is the ability to store and share disparate types of data, including imaging, electrophysiology, and clinical data, from both humans and animals, on one platform that includes not only options for data visualization but also a wide variety of analytic tools that are integrated across different programming languages.

2. LONI infrastructure, data storage, and processing

The total amount of data that are planned to be collected in EpiBioS4Rx is unprecedented: video-electroencephalography (EEG) from cohorts of animals after TBI (using a fluid percussion model) recorded continuously for six months, in addition to prolonged continuous intensive care unit (ICU) EEG recordings from 300 humans and intermittent sampling of brain images, blood, and tissue data (Vespa et al., 2018). The data, measured in tens to one hundred terabytes, represent investigation on a scale that was not possible until just recently. It leverages state of the art analysis tools to track candidate biomarkers and their statistical associations.

2.1. Study sites and patient population

The study sites include University of California, Los Angeles (UCLA – clinical coordinating center), University of California, Davis, Phoenix Childrens' Hospital, Yale University, Harvard University/Massachusetts General Hospital, University of Pennsylvania, University of Cincinnati, University of Miami, University of Pittsburgh, Johns Hopkins University, Columbia University, Royal Melbourne Hospital, The Alfred, and Childrens National Hospital. 300 total patients will be enrolled over 4 years, and they will be followed longitudinally for 2 years after injury. Patients admitted into the ICU after an acute moderate-severe TBI involving a frontal and/or temporal lobe hemorrhagic contusion will be screened (Vespa et al., 2018).

2.2. REDCap and LONI IDA online databases

ICU physiological data, demographic information, outcome measures, and prospective research data will be uploaded to the Research Electronic Data Capture (REDCap) data repository hosted by LONI. REDCap is a secure, web-based application designed to support data capture for research studies in a metadata-driven manner (Harris et al., 2009). The online database contains 26 electronic case report forms (eCRFs) designed by the UCLA Brain Injury Research Center. Common data elements have been expanded to collect EEG, MRI, and biosample information. All continuous EEG and neuroimaging data across sites will be uploaded and managed by the USC LONI Online Image and Data Archive (IDA) (Vespa et al., 2018).

2.3. Continuous scalp and depth EEG monitoring to detect early seizures

Enrolled patients will receive 24-h continuous EEG (cEEG) for 72 h minimum during the first 7 days after TBI. Scalp cEEG monitoring will be performed at the patient bedside using a 16–21 channel bipolar and referential composite montage implemented at each study center based on their established ICU EEG protocols. A subset of 100 patients will receive additional depth EEG monitoring using a 6-contact mini depth electrode during the first 7 days after TBI for higher resolution as well as pHFOs and repetitive high frequency oscillations and spikes (rHFOSs) detection. 24-h cEEG files will be de-identified and uploaded to the LONI IDA (Vespa et al., 2018).

The central analysis team reviews cEEG data using PERSYST Version 13 and MATLAB 8.1 to analyze spikes, pHFOs, and rHFOSs. A structured protocol will be performed to determine interictal epileptiform spike and seizure onset, location of epileptiform onset, spike morphology, spike repetition rate, clustering features, field size, and spread patterns. EEG data will be correlated with MRI data and metabolite plasma data over the first seven days of injury.

2.4. Multimodal MRI analysis for structural biomarkers

The initial three standards of care clinical CT scans over the first 24 h after trauma will be de-identified and uploaded to the LONI IDA to confirm initial injury characteristics. A high-resolution MRI, acquired on a 3 T scanner, will be performed on Day 14 (± 4 days) post-injury. MRI sequences acquired include: 3D T1, 2D resting state bold oxygen level dependent imaging (rs-BOLD), 2D diffusion tensor imaging (DTI), 3D Gradient Echo/Susceptibility Weighted Imaging (GRE/SWI), 3D T2, and 3D T2-weighted fluid attenuated inversion recovery (FLAIR). MRI acquisition parameters will be optimized across sites and scanner types to reduce inter-scanner variability.

MRI injury location and total hemorrhagic lesion load will be correlated with post-traumatic epilepsy (PTE) occurrence. T1-weighted MRI scans will be used for 1) a regional volumetric analysis, using FMRIB Software Library (FSL) utilities, and a 2) subcortical morphometric shape analysis, using a validated open-source pipeline robust to brain pathology.

2.5. Biosample collection and analysis

Study sites will collect blood samples through central lines or venous punctures on post-injury day 1, 3, 5, 15, 30 ± 10, 90 ± 10, and 180 ± 10. Biosample collection, processing, and shipping information will be inputted into the REDCap database. cEEG and/or depth EEG data will be correlated to the biomarker results to characterize the relationship between EEG epileptiform activity and appearance and time course of selected biomarkers.

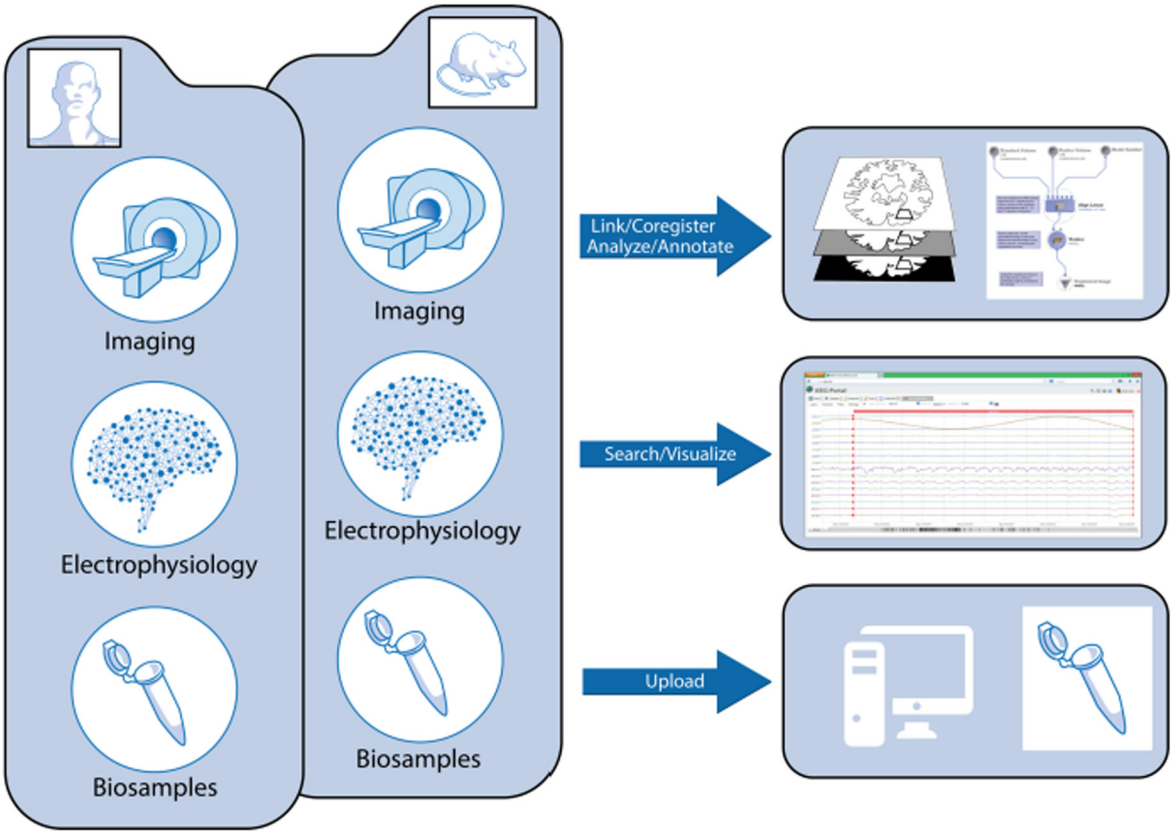

Fig. 1. The EpiBioS4Rx portal supports heterogeneous data and provides essential features including upload/ingest, search/visualize, link/co-register, and analyze/annotate for users.

All longitudinal follow up evaluations will be uploaded to the REDCap database.

We have established the informatics and analytics framework of EpiBioS4Rx at LONI. Continually working closely with data collection sites, we assist in consolidating the data and providing analytic tools that show potential to lead to the identification of biomarkers of epileptogenesis. Well-established communication with the data collection sites ensures that we are constantly improving and optimizing our infrastructure as we receive feedback from our collaborators and outside investigators to make the process more efficient and user friendly. Furthermore, detailed documentation and annotations enable researchers to link different data types easily. For example, a researcher can look at EEG and find the corresponding patient's imaging data to see spatially from where the EEG recordings were taken (and vice versa). Moreover, researchers will be able to compare clinical data over various time points with the EEG and MRI data. By combining these new data capabilities, which allow investigators to link various data modalities, we have been focusing on discovering quantitative methods, including dimensionality reduction and pattern recognition, of identifying epileptogenesis after TBI.

Moreover, biomarkers and models of epileptogenesis will help define preclinical trial populations, expedite interventions to prevent epilepsy after TBI, and document epilepsy before late seizures occur. Based on previous studies, it is likely that there are reproducible changes in biomarkers, such as occurrence of pathological high frequency oscillations (pHFOs) in the intracranial EEG, which identify the epileptogenic area before its overt clinical expression (Buja et al., 2009; Winden et al., 2015, 2011).

We have implemented new approaches for analyzing the collected data, including novel graphical methods to visualize multivariable interactions and to quantify patterns or variability in the data. Quantitative and data mining methods enable investigators to record and analyze gold-standard data and to create a shared bioinformatics resource for epilepsy research that will continue to exist after this study concludes. We are developing a wide variety of analytic tools for users and integrating multi-modal data in a way that transcends the capability of a single laboratory or center. We are providing a lasting and open platform for standardized biomarker research in both TBI and PTE. Furthermore, because of the existing data on the LONI IDA that have been collected from other projects, researchers may validate EpiBioS4Rx as specific for PTE or not.

Different data often have rich, high-dimensional levels of detail. As such, exploring and navigating through the full data set poses challenges, especially in providing investigators with tools that are easy to use and comprehend. Visualizing data helps to orient investigators and provides context to understand relationships and discover hidden insights within the underlying data. Although in recent years, database solutions have emerged to collect these types of data, they do not provide visualization or exploration functionalities. Applications such as tranSMART offer some analytic capability but require data to be laboriously imported and are not automatically updated when new data are archived. The LONI IDA data visualization interface unites the benefits of visual representations with a comprehensive, harmonized data set and allows creation of subject cohorts and to search, compare, and download data.

Data are being stored on the IDA, because LONI has the ability and experience to store and share petabytes of data (current data storage capacity is over 7 PB and increasing each year). In terms of compute hours, we currently limit each user to 129,024 CPU hours on the LONI server per week; this is 768 slots at any given moment (for 24 h and 7 days). This past year, the average monthly grid usage was approximately 800,000 h. We intend to continue increasing storage availability as needed, allowing each user more CPU hours to ensure that users can process their data with little to no wait time.

2.6. Streamlined data consolidation

Users upload their raw data files (of various file types) directly to an extended version of the LONI infrastructure, where they are automatically classified, converted, and annotated (Fig. 1). By automating much of this process, researchers uploading and downloading data are spared the time and effort previously involved in accessing and sharing epilepsy data. This streamlined data consolidation aims to increase the financial efficiency and scientific productivity of the broader epilepsy research community. In addition, physical samples are aggregated in a single biobank with our collaborators outside of LONI, further reducing coordination challenges.

While LONI has many years of experience with large-scale neuroimaging studies and data integration (Crawford et al., 2016; Dinov et al., 2014; Torgerson et al., 2015; Van Horn and Toga, 2009), with data upload and download using the Image and Data Archive (IDA), we have expanded the types of data that can be uploaded to include EEG using the EDF+ format. Users have the option to select MRI or EEG when uploading or downloading data so that the process remains clear and easy to use. Furthermore, these features are being expanded for preclinical data so that users will be able to upload both preclinical and clinical data on the same platform. Besides the rich clinical data collected as part of EpiBioS4Rx, this will be the first prospective preclinical databank in the world.

Data transformation software (Barker-Haliski et al., 2014) automatically detects new data, validates them, maps the data to a common data model (where applicable), and pre-indexes the clinical data by features and values to aid in search and co-registration. Through a federated architecture, key components of the data may be distributed across the LONI platform. Quality control and provenance information is maintained with all raw data.

2.7. Global Unique Identifiers (GUIDs)

An essential requirement when federating data from multiple research studies is to prevent data from subjects who participate in more than one research study from being multiply-represented in the federated system. Since personally-identifying information, such as first and last names, is regularly removed from the data collected and shared by research studies to protect each subject's identity, determining what data belong to any given subject across datasets is a difficult, if not impossible, task for investigators who are analyzing the data. Global Unique Identifiers (GUIDs) are used to distinguish subjects uniquely across research studies while preventing the identities of the subjects from being discovered (Johnson et al., 2010).

We have designed and built a GUID system for the third phase of one of our other big data projects, the ADNI (Alzheimer's Disease Neuroimaging Initiative) (Schneider et al., 2011; Toga and Crawford, 2010b), which is being used to create GUIDs for ADNI subjects. This system is also applied to EpiBioS4Rx. To ensure compatibility with the GUIDs created by the NIH, we have encapsulated the GUID algorithm used by the National Database for Autism Research (NDAR) (Payakachat et al., 2016) system into the Global Alzheimer's Association Interactive Network (GAAIN) GUID generator, and it serves as our GUID generation engine. We have also developed an algorithm that allows for cross-comparisons of GUIDs between different GUID systems without revealing the internal hash codes used for GUID subject identification, including those stored in the NDAR and FITBIR systems.

2.8. Quality control

The LONI Neuroimaging Quality Control (QC) System is used for all multi-modal data, checked for quality, and reviewed by participating investigators who are collecting the data (Fig. 2). Since there is variability in data, annotation, and models among the various data collection sites, we have developed tools to normalize and harmonize signal, image, and other data. Images uploaded from participating centers are processed using LONI's multicenter data review and assessment system (https://qc.loni.usc.edu). This system allows automated pre-processing that generates vector statistics and derived images to assess data quality. Data are run through automated artifact detection algorithms in preparation for initial biomarker processing (Brinkmann et al., 2009; LeVan et al., 2006). This system is web-accessible, user friendly, simple to navigate, and provides a long-term resource for this field.

2.9. User-friendly data search and navigation

Data are searchable and accessible through web clients as well as programmatic (MATLAB, C, Python, Java, and R) interfaces. By converting data to consistent file formats and tagging that data with metadata, we have enabled Google-style search of all available epilepsy data. However, since data are interlinked and co-registered across datasets and modalities, the search functionality does not simply match data against individual items as Google does—rather, it will find interlinked combinations of data (even across modalities and data sources) that match the desired criteria (Talukdar et al., 2010). This enables sophisticated custom searches that match the functionality of predefined query forms. Users can browse data in their most appropriate visual representation and pivot from one data view or modality to another. Based on our experience with previous big data studies, we have learned what access control and sharing mechanisms are required by the community and how to effectively enable inter-project as well as community-scale data sharing, which we have now implemented into EpiBioS4Rx. Key components include giving users explicit access control for their data and results as well as providing project groups for larger-scale permissions management.

3. LONI analysis methods

3.1. Automated analysis

The LONI Pipeline (Dinov et al., 2010, 2014; MacKenzie-Graham and Payan, 2008) contains a common framework for visual and programmatic construction of data-driven workflows for electrophysiology, imaging, and biosample data. With the aid of LONI's workflow builder (Fig. 3), complex analyses are represented visually, further supporting researchers' investigations. Examples of LONI Pipeline applications include developing a unified coordinate space for seizure onset locations across various brains, including animal and human MRI, using string similarity and value overlap to predict that different contributor metadata fields are the same, and providing graphical interfaces for linking data. Co-registration algorithms are typically invoked at upload-time but may also be triggered later manually for further refinement. We provide MRI supervision and integration from different scanners and centers by supervising phantom studies, assessing quality, and fixing problems with heterogeneity. Robust workflow pipelines are provided for researchers to use on both humans and animal models. An example depicting steps of some MRI analysis using the LONI Pipeline is shown in Fig. 3.

3.2. Standardized sample collection, shipping, and biobank storage protocols

We have developed protocols based on our previous studies to define methods for harvesting, freezing, and storing tissue and other biosamples (i.e., cerebrospinal fluid). Parallel storage protocols are followed to ensure that parallel human and animal samples are stored and treated in a similar fashion to facilitate comparison of findings. These protocols are also meant to be foundational to future sample collection and transport among the wider epilepsy research community.

We describe examples from the initial analysis on the MRI and EEG data collected in EpiBioS4Rx from the 16 currently enrolled patients

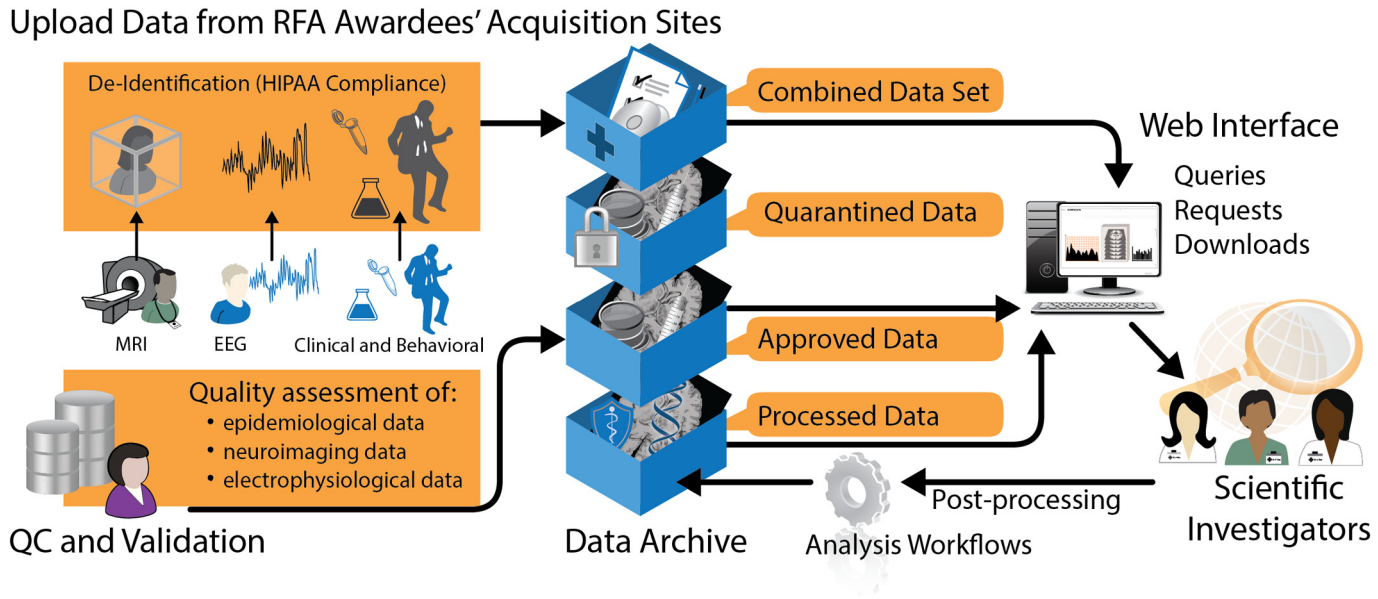

Fig. 2. The major elements and functions of our data ingestion and archive.

and animal data from 10 rats.

3.3. MRI analysis

The collected human MRI data consist of structural, functional (resting state), and diffusion weighted measures (Vespa et al., 2018). MRI analyses consist of structural analyses (performed in BrainSuite (Shattuck and Leahy, 2002)) to measure each subject's intracranial volumes as well as gray matter volumes and other anatomical measures. Functional analyses are conducted using Statistical Parametric Mapping (SPM) (Ashburner, 2012), a software suite of MATLAB, to ascertain brain activation in different regions. Functional connectivity analyses are performed in the CONN toolbox of MATLAB to examine network connectivity in comparison to non-TBI data, to determine abnormally active/inactive networks. Lastly, the diffusion weighted analyses consist of constructing each subject's fractional anisotropy (FA) maps, in addition to measuring each patient's apparent diffusion coefficient (ADC) to assess white matter integrity and connectivity in FMRIB Software Library (FSL) (Jenkinson et al., 2012). These FA maps of TBI data are compared to five normal, non-TBI data in a group analysis via tract based spatial statistics (TBSS) in FSL.

The collected rat MRI consists of structural and diffusion weighted measures. MRI processing and analyses mainly consist of each rat's diffusion weighted measures. Each rat's FA map is constructed in FSL. Additionally, TBSS is performed on the TBI rats and non-TBI animal data to measure group differences. Example structural and DTI data for a control rat (Sprague-Dawley) and TBI rat (left lateral fluid percussion injury) are shown in Fig. 4 using DSI Studio. In this example, the data were collected using a Bruker BioSpin MRI GmbH at the University of Eastern Finland, Kuopio using a dtiEpiT SpinEcho sequence.

We have introduced methods for TBI connectomics (Irimia et al., 2012a, 2012b; Torgerson et al., 2013; van Horn et al., 2012) to be used on the clinical data in this study. DTI is used to extract connectivity between all pairs of gyral and sulcal structures in the presence of brain trauma. Connectivity between all brain regions (165 in our scheme) is computed from DTI volumes acquired longitudinally from each patient. Diffusion tractography is used to determine connectivity properties (WM bundle length, connectivity density, and FA) and each subject's weighted connectivity matrix. WM fiber tracking of inter-regional connectivity is conducted using TrackVis (Wedeen et al., 2008) or other tractography tools. Connectivity between regions (such as thalamo-cortical connections and hippocampal connections) is assessed systematically within each patient using purpose-built workflows for multi-modal co-registration of MRI. This will be followed by calculation of (i) inter-regional connectivity matrices and (ii) longitudinal changes in connectivity topology using network-theoretic descriptors of nodal and network-wide segregation (clustering coefficient, modularity, etc.) and integration (characteristic path length, global efficiency, etc.). Additional network-theoretic measures (scale freedom, small worldness, robustness, centrality, degree distribution, and communication efficiency) (Eguiluz et al., 2005; Salvador et al., 2005; Stam, 2004; Wilcox, 2012) will be computed. Results and changes over time in each

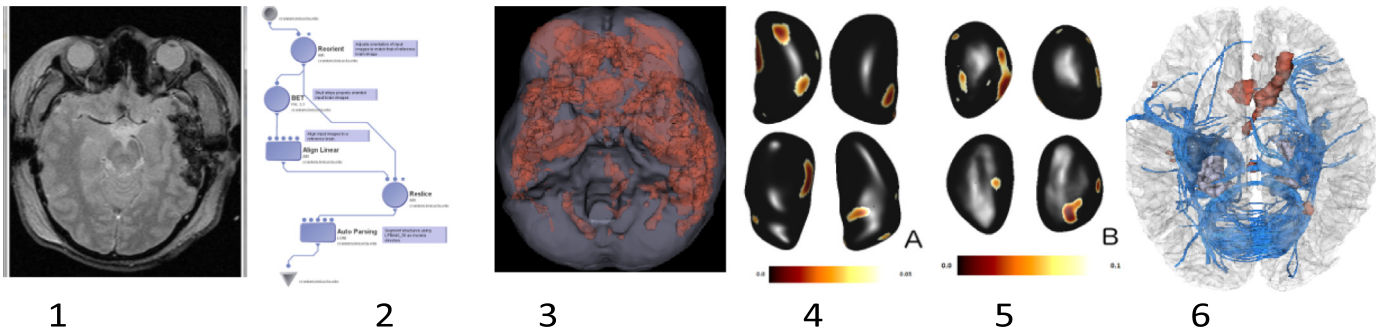

Fig. 3. Example of analysis performed using the LONI Pipeline, including MRI and DTI data with group analysis over 46 patients depicting common locations of hemorrhages across these patients and physiological group differences in PTE, making use of both modular and automated LONI Pipeline techniques.

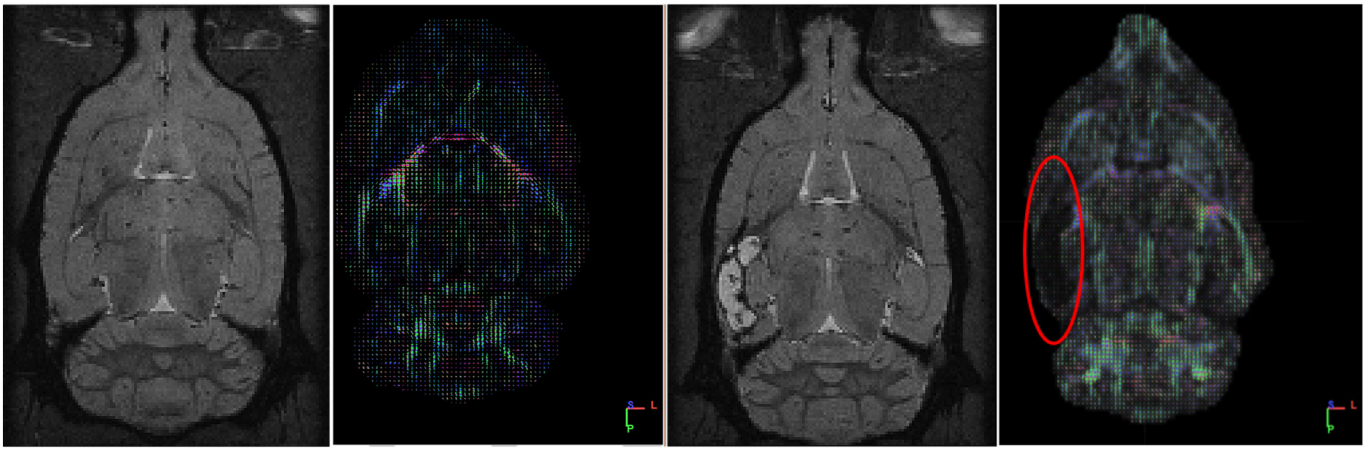

Fig. 4. Rat (male, 2 month-old Sprague Dawley rat, 300 g weight, courtesy of University of Eastern Finland, 7 T/16 cm Bruker Pharmascan) T1 MRI on the left and corresponding DTI on the right for a control rat in the first 2 images and a rat (left parietal LFPI model, 5 mm, severe injury, on a male, 2 month-old Sprague Dawley rat, 300 g weight, courtesy of University of Eastern Finland, 7 T/16 cm Bruker Pharmascan) in the third and fourth images (decreased FA map intensity circled in red); FA map used deterministic fiber tracking algorithm, anisotropy threshold was randomly selected, angular threshold was selected from 15 to 90, and fiber trajectories were smoothed by averaging propagation direction with percentage of previous direction. The images are in radiological orientation, so right and left are flipped. Colors correspond with the direction of the water/fluid flow in the WM tracts, in which blue is superior-inferior direction, red is right –left (lateral), and green is anterior-posterior. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

patient are visualized and analyzed using connectograms (Irimia et al., 2012a). Workflows are fully integrated with the LONI Pipeline (Dinov et al., 2010; MacKenzie-Graham and Payan, 2008).

3.4. Translational aspect of analysis

The comparison of human and animal neuroimaging data presents anatomical challenges; we aim to compare the WM tracts' characteristics and integrity of human and rat neuroimaging data. The TBI rats' TBSS and the patients' TBSS will be compared to examine WM tract similarities that could relate to network abnormalities in epileptic human patients. Specifically, we are analyzing the WM integrities in the animal model and how that may impact network performance or connectivity in humans. The translational aspect will be unique to EpiBioS4Rx with such a thorough dataset, including rats and patients, at LONI.

3.5. EEG analysis

We use Persyst software,(Sierra-Marcos et al., 2015) linked to data from the IDA, for EEG data visualization over multiple channels, export of artifact reduced waveform data, seizure and spike detection, wavelets, matching pursuit, correlation, FFT phase, period evolution, and other EEG analysis tools. Due to the sheer volume of EEG data due to the continuous recordings and number of electrode contacts used, we have applied a variety of dimensionality reduction techniques to the EEG for both preclinical and clinical data. To increase the ease of understanding the high dimensional data and outline trends in these collected samples, we apply these methods, assuming the data can be reduced to lie on a nonlinear manifold of lower, intrinsic dimensionality. Furthermore, we use these methods to remove excessive noise in the data, which is particularly a problem with scalp EEG as well as to look for patterns or features of epileptogenesis.

We have applied and compared Principal Component Analysis (PCA), Diffusion Maps, Laplacian Eigenmaps, Kernel PCA, and Unsupervised Diffusion Component Analysis (UDCA), which are methods that can be used on both animal and human data. Each of these tools has its own benefits and weaknesses, so we are providing a variety of dimensionality reduction methods for researchers, because one method will not always be the best to use in every instance. PCA is a linear dimensionality reduction method (Wold et al., 1987); it is used by rotating data in a different orientation in the dimensional space by exposing the maximum variance. It detects and eliminates some noise and collects the redundancy of the data. Kernel PCA is an extension of PCA that uses techniques of kernel methods (Jade et al., 2003). Laplacian Eigenmaps is a nonlinear dimensionality reduction method that assumes that data lie in a low dimensional manifold within the high dimensional space; it grabs information from nearest neighbors of each data point (Belkin and Niyogi, 2003). Thus, a low dimensional dataset is produced by preserving local properties of the manifold and minimizing the distance between a data point and its nearest neighbor. Diffusion Mapping is another nonlinear dimensionality reduction method (Coifman and Lafon, 2006; Duncan et al., 2013). A family of embeddings of a dataset is computed into a low-dimensional Euclidean space whose coordinates can be computed from the eigenvectors and corresponding eigenvalues of a diffusion operator on the data. We have developed UDCA (Duncan and Strohmer, 2016), which is an extension and adaptation of diffusion maps. In this algorithm, coordinates are constructed that generate efficient geometric representations of the complex data. Additionally, this algorithm performs well by removing noise from the data (using the Mahalanobis distance measure with inverse covariance matrices (Talmon et al., 2012)) and is completely automatic. EEGLAB (Delorme and Makeig, 2004), a MATLAB toolbox and graphic user interface, is used to open EDF + EEG files in MATLAB.

Fig. 5 shows UDCA applied to a sample of pre-ictal data, where the algorithm separates preseizure features that are not apparent from visually inspecting the raw data, and then a method of plotting the Euclidean distances of the points from the embedding to the origin is shown to demonstrate how we can set a threshold of a chosen amplitude that can be used to automatically extract features of epileptogenesis after TBI. This method is useful for noisy, complex data, such as EEG and allows researchers to extract the underlying brain activity that may be associated with biomarkers of epileptogenesis (Duncan et al., 2018).

4. Discussion

Our central data repository, the LONI IDA, has been configured to receive preclinical and clinical EpiBioS4Rx data, including MRI, CT, and EEG data submitted by the 16 participating clinical and 4 preclinical centers of the study. The data upload process (depicted in Fig. 6) includes a data de-identification process that occurs before data

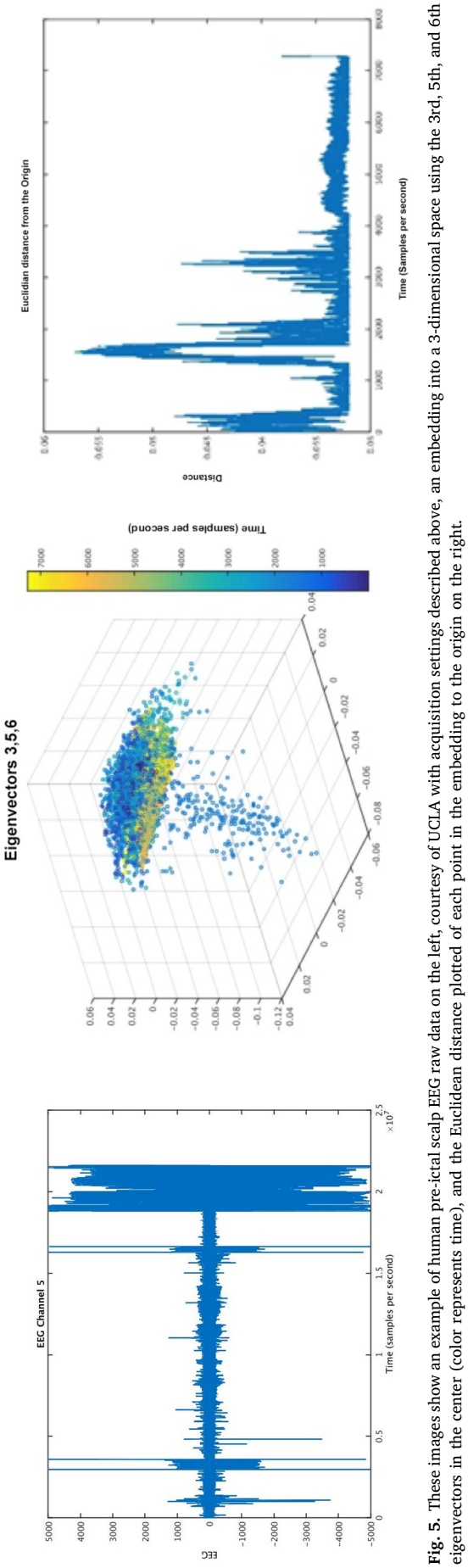

Fig. 5. These images show an example of human pre-ictal scalp EEG raw data on the left, courtesy of UCLA with acquisition settings described above, an embedding into a 3-dimensional space using the 3rd, 5th, and 6th eigenvectors in the center (color represents time), and the Euclidean distance plotted of each point in the embedding to the origin on the right.

are transferred to the central repository and is configured to work for multiple file formats, including DICOM, ECAT, HRRT, and EDF. Data arriving at the central repository are immediately and automatically checked in, and a subset of metadata attributes are extracted from the files and used to catalog and describe the data to support database searches. Check-in is typically completed within 3 min, at which time data become immediately available to investigators to import into the LONI QC and/or Pipeline workflow environments and/or to download for local analysis.

Some of the challenges that we have encountered are in managing the heterogeneity and scale of the potentially relevant data. This occurs (1) while integrating and interlinking the data such that it can be stored, accessed, searched, and analyzed; (2) while browsing and algorithmically analyzing the data in search of biomarkers, where relevant features are likely. We have worked to ensure rigorous experimental design for robust and unbiased results on our platform.

5. Conclusions

We have built upon decades of experience with big data projects at LONI to develop the informatics infrastructure needed for a large-scale study, such as EpiBioS4Rx.

We have created an infrastructure for EpiBioS4Rx investigators, collaborators, and the broader epilepsy clinical and research community. Additionally, we have established methods for mining the complex, multi-modal data collected in the study, and ultimately, to develop data-driven predictive mathematical models of the epileptogenic processes that represent sensitive and specific biomarkers to predict the development of epilepsy after TBI. We have described some of the analytic tools that we have developed and used in our search for biomarkers of epileptogenesis. Biomarkers that are discovered will be instrumental in our efforts to develop models aimed at predicting the probability of developing epilepsy in post-TBI subjects and identifying specific times, regions, and processes where intervention may be most beneficial. The online interface that we have established allows users to search across raw and processed data in a data mining way and enable visualization and analyses to test hypotheses and validate results. By sharing access to our data and analytic tools as well as our server for data processing, we hope to encourage collaborations among different centers around the world and to bring awareness to investigators about the data collected and analysis methods used by other teams. We have developed a data flow process that makes the data findable, accessible, interoperable and reusable for the epilepsy research community, thus following the Findable, Accessible, Interoperable, and Reusable (FAIR) guiding principles (Wilkinson et al., 2016). This infrastructure has the potential to spur the advancement and development of research directed toward translational or clinical development of additional disease-modifying or preventative therapies. The mechanisms that are revealed in our search for these biomarkers will be used as targets for pioneering antiepileptogenic treatments.

We will continue to extract features from neuroimaging, electrophysiologic, molecular, clinical, cognitive, and behavioral measures over time to identify candidate diagnostic biomarkers of epileptogenesis. Novel statistical tools, which we will continue to modify and improve, have been developed to visualize complex associations among multiple variables as they evolve over time during epileptogenesis. They will reveal processes, regions, and stages in epileptogenesis correlated with specific anatomical changes in imaging. Advanced statistical techniques will then be used with the goal of building models of epileptogenesis to predict the probability of epilepsy, based on biomarker inputs. The results of the predictive models will be validated by testing the robustness of the results in the presence of uncertainty.

Acknowledgements

This research was supported by the National Institute of

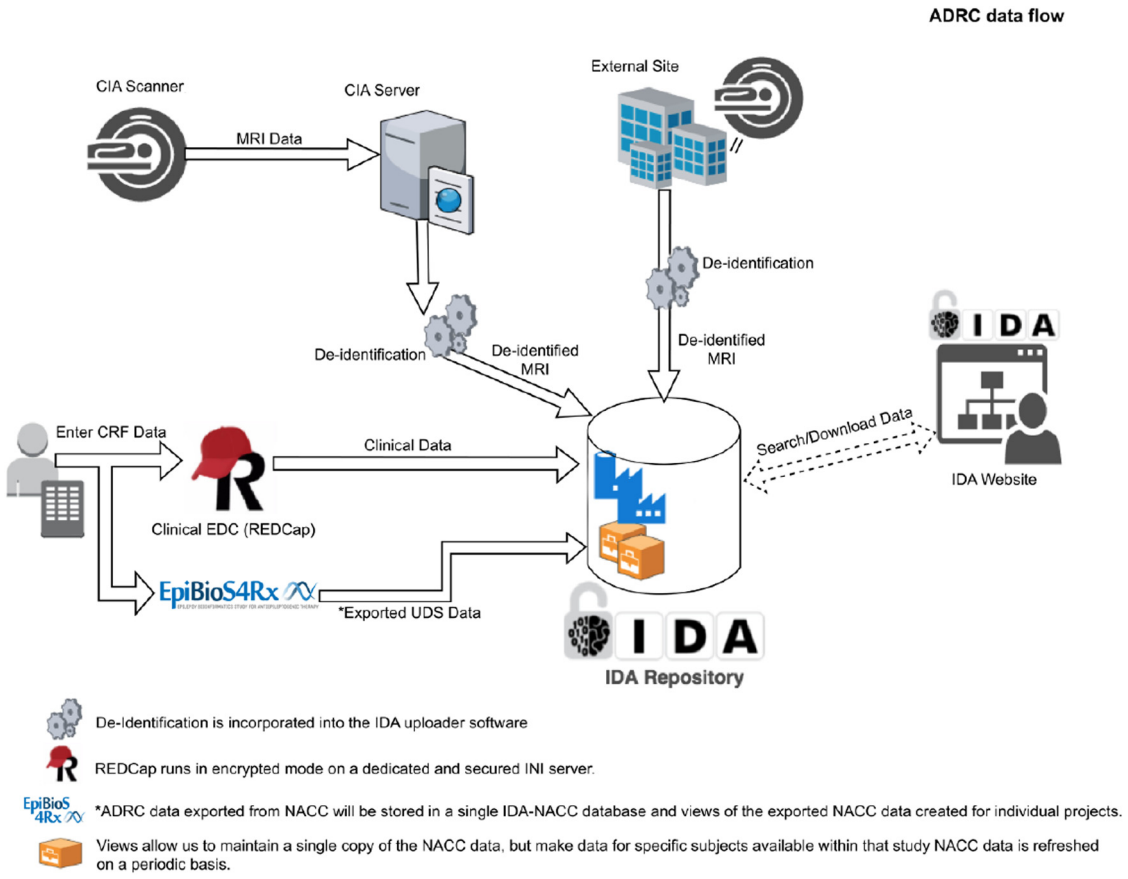

Fig. 6. Data flow process.

Neurological Disorders and Stroke (NINDS) of the National Institutes of Health (NIH) under Award Numbers U54NS100064 (EpiBioS4Rx), NIH P41-EB015922, and NIH U54-EB020406.

References

Ashburner, J., 2012. SPM: a history. NeuroImage. http://dx.doi.org/10.1016/j.neuroimage.2011.10.025.

Astakhov, V., Gupta, A., Santini, S., Grethe, J.S., 2005. Data integration in the biomedical informatics research network (BIRN). In: Data Integr. Life Sci. Proc. 3615. pp. 317–320.

Barker-Haliski, M., Friedman, D., White, H.S., French, J.A., 2014. How clinical development can, and should, inform translational science. Neuron. http://dx.doi.org/10.1016/j.neuron.2014.10.029.

Belkin, M., Niyogi, P., 2003. Laplacian Eigenmaps for dimensionality reduction and data representation. Neural Comput. 15, 1373–1396. http://dx.doi.org/10.1162/089976603321780317.

Bourne, P.E., Bonazzi, V., Dunn, M., Green, E.D., Guyer, M., Komatsoulis, G., Larkin, J., Russell, B., 2015. The NIH big data to knowledge (BD2K) initiative. J. Am. Med. Inform. Assoc. 22 (1114–1114). https://doi.org/10.1093/jamia/ocv136.

Brinkmann, B.H., Bower, M.R., Stengel, K.A., Worrell, G.A., Stead, M., 2009. Multiscale electrophysiology format: an open-source electrophysiology format using data compression, encryption, and cyclic redundancy check. In: Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine. 2009. EMBC, pp. 7083–7086. http://dx.doi.org/10.1109/IEMBS.2009.5332915.

Buja, A., Cook, D., Hofmann, H., Lawrence, M., Lee, E.-K., Swayne, D.F., Wickham, H., 2009. Statistical inference for exploratory data analysis and model diagnostics. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. 367, 4361–4383. http://dx.doi.org/10.1098/rsta.2009.0120.

Coifman, R.R., Lafon, S., 2006. Diffusion maps. Appl. Comput. Harmon. Anal. 21, 5–30. http://dx.doi.org/10.1016/j.acha.2006.04.006.

Crawford, K.L., Neu, S.C., Toga, A.W., 2016. The image and data archive at the laboratory of neuro imaging. NeuroImage 124, 1080–1083. http://dx.doi.org/10.1016/j.neuroimage.2015.04.067.

Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J. Neurosci. Methods 134, 9–21. http://dx.doi.org/10.1016/j.jneumeth.2003.10.009.

Dinov, I., Lozev, K., Petrosyan, P., Liu, Z., Eggert, P., 2010. Neuroimaging study designs, computational analyses and data provenance using the LONI pipeline. PLoS One 5, e13070. http://dx.doi.org/10.1371/journal.pone.0013070.

Dinov, I.D., Petrosyan, P., Liu, Z., Eggert, P., Zamanyan, A., Torri, F., Macciardi, F., Hobel, S., Moon, S.W., Sung, Y.H., Jiang, Z., Labus, J., Kurth, F., Ashe-McNalley, C., Mayer, E., Vespa, P.M., Van Horn, J.D., Toga, A.W., 2014. The perfect neuroimaging-genetics-computation storm: collision of petabytes of data, millions of hardware devices and thousands of software tools. Brain Imaging Behav. 8, 311–322. http://dx.doi.org/10.1007/s11682-013-9248-x.

Dourado, A., Le Van Quyen, M., Schelter, B., Favaro, G., Schulze-Bonhage, A., Sales, S., Navarro, V., 2009. EPILEPSIAE-evolving platform for improving living expectation of patients suffering from IctAl events. Epilepsia 50, 210–211.

Duncan, D., Strohmer, T., 2016. Classification of Alzheimer's disease using unsupervised diffusion component analysis. Math. Biosci. Eng. 13, 1119–1130. http://dx.doi.org/10.3934/mbe.2016033.

Duncan, D., Talmon, R., Zaveri, H.P., Coifman, R.R., 2013. Identifying preseizure state in intracranial EEG data using diffusion kernels. Math. Biosci. Eng. 10, 579–590. http://dx.doi.org/10.3934/mbe.2013.10.579.

Duncan, D., Toga, A.W., Vespa, P.M., 2018. Detecting features of epileptogenesis in EEG after TBI using unsupervised diffusion component analysis. Math. Biosci. Eng. 23.

Eguiluz, V.M., Chialvo, D., Cecchi, G.A., Baliki, M., Apkarian, A.V., 2005. Scale-free brain functional networks. Phys. Rev. Lett. 94, 018102.

Harris, P.A., Taylor, R., Thielke, R., Payne, J., Gonzalez, N., Conde, J.G., 2009. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. J. Biomed. Inform. 42 (2), 377–381. http://dx.doi.org/10.1016/j.jbi.2008.08.010.

Helmer, K.G., Ambite, J.L., Ames, J., Ananthakrishnan, R., Burns, G., Chervenak, A.L., Foster, I., Liming, L., Keator, D., Macciardi, F., Madduri, R., Navarro, J.-P., Potkin, S., Rosen, B., Ruffins, S., Schuler, R., Turner, J.A., Toga, A., Williams, C., Kesselman, C., 2011. Enabling collaborative research using the biomedical informatics research network (BIRN). J. Am. Med. Inform. Assoc. 18, 416–422. http://dx.doi.org/10.1136/amiajnl-2010-000032.

Ihle, M., Feldwisch-Drentrup, H., Teixeira, C.A., Witon, A., Schelter, B., Timmer, J., Schulze-Bonhage, A., 2012. EPILEPSIAE - a European epilepsy database. Comput. Methods Prog. Biomed. 106, 127–138. http://dx.doi.org/10.1016/j.cmpb.2010.08.011.

Irimia, A., Chambers, M.C., Torgerson, C.M., Van Horn, J.D., 2012a. Circular representation of human cortical networks for subject and population-level connectomic visualization. NeuroImage 60, 1340–1351. http://dx.doi.org/10.1016/j.neuroimage.2012.01.107.

Irimia, A., Wang, B., Aylward, S.R., Prastawa, M.W., Pace, D.F., Gerig, G., Hovda, D.A., Kikinis, R., Vespa, P.M., Van Horn, J.D., 2012b. Neuroimaging of structural

pathology and connectomics in traumatic brain injury: toward personalized outcome prediction. NeuroImage Clin. http://dx.doi.org/10.1016/j.nicl.2012.08.002.
Jade, A.M., Srikanth, B., Jayaraman, V.K., Kulkarni, B.D., Jog, J.P., Priya, L., 2003. Feature extraction and denoising using kernel PCA. Chem. Eng. Sci. 58, 4441–4448. http://dx.doi.org/10.1016/S0009-2509(03)00340-3.
Jenkinson, M., Beckmann, C.F., Behrens, T.E.J., Woolrich, M.W., Smith, S.M., 2012. FSL. NeuroImage 62, 782–790. http://dx.doi.org/10.1016/j.neuroimage.2011.09.015.
Johnson, S.B., Whitney, G., McAuliffe, M., Wang, H., Mccreedy, E., Rozenblit, L., Evans, C.C., 2010. Using global unique identifiers to link autism collections. J. Am. Med. Inform. Assoc. 17, 689–695. http://dx.doi.org/10.1136/jamia.2009.002063.
Kini, L.G., Davis, K.A., Wagenaar, J.B., 2016. Data integration: combined imaging and electrophysiology data in the cloud. NeuroImage 124, 1175–1181. http://dx.doi.org/10.1016/j.neuroimage.2015.05.075.
Klatt, J., Feldwisch-Drentrup, H., Ihle, M., Navarro, V., Neufang, M., Teixeira, C., Adam, C., Valderrama, M., Alvarado-Rojas, C., Witon, A., Le Van Quyen, M., Sales, F., Dourado, A., Timmer, J., Schulze-Bonhage, A., Schelter, B., 2012. The EPILEPSIAE database: an extensive electroencephalography database of epilepsy patients. Epilepsia 53, 1669–1676. http://dx.doi.org/10.1111/j.1528-1167.2012.03564.x.
Levan, P., Urrestarazu, E., Gotman, J., 2006. A system for automatic artifact removal in ictal scalp EEG based on independent component analysis and Bayesian classification. Clin. Neurophysiol. 117, 912–927. http://dx.doi.org/10.1016/j.clinph.2005.12.013.
Lim, M.D., 2014. Consortium sandbox: building and sharing resources. Sci. Transl. Med. http://dx.doi.org/10.1126/scitranslmed.3009024.
MacKenzie-Graham, A., Payan, A., 2008. Neuroimaging data provenance using the LONI pipeline workflow environment. In: Proven…, pp. 1–12. http://dx.doi.org/10.1007/978-3-540-89965-5_22.
Marcus, D.S., Harms, M.P., Snyder, A.Z., Jenkinson, M., Wilson, J.A., Glasser, M.F., Barch, D.M., Archie, K.A., Burgess, G.C., Ramaratnam, M., Hodge, M., Horton, W., Herrick, R., Olsen, T., McKay, M., House, M., Hileman, M., Reid, E., Harwell, J., Coalson, T., Schindler, J., Elam, J.S., Curtiss, S.W., Van Essen, D.C., 2013. Human connectome project informatics: quality control, database services, and data visualization. NeuroImage 80, 202–219. http://dx.doi.org/10.1016/j.neuroimage.2013.05.077.
Marek, K., Jennings, D., Lasch, S., Siderowf, A., Tanner, C., Simuni, T., Coffey, C., Kieburtz, K., Flagg, E., Chowdhury, S., Poewe, W., Mollenhauer, B., Sherer, T., Frasier, M., Meunier, C., Rudolph, A., Casaceli, C., Seibyl, J., Mendick, S., Schuff, N., Zhang, Y., Toga, A., Crawford, K., Ansbach, A., de Blasio, P., Piovella, M., Trojanowski, J., Shaw, L., Singleton, A., Hawkins, K., Eberling, J., Russell, D., Leary, L., Factor, S., Sommerfeld, B., Hogarth, P., Pighetti, E., Williams, K., Standaert, D., Guthrie, S., Hauser, R., Delgado, H., Jankovic, J., Hunter, C., Stern, M., Tran, B., Leverenz, J., Baca, M., Frank, S., Thomas, C.A., Richard, I., Deeley, C., Rees, L., Sprenger, F., Lang, E., Shill, H., Obradov, S., Fernandez, H., Winters, A., Berg, D., Gauss, K., Galasko, D., Fontaine, D., Mari, Z., Gerstenhaber, M., Brooks, D., Malloy, S., Barone, P., Longo, K., Comery, T., Ravina, B., Grachev, I., Gallagher, K., Collins, M., Widnell, K.L., Ostrowizki, S., Fontoura, P., La-Roche, F.H., Ho, T., Luthman, J., van der Brug, M., Reith, A.D., Taylor, P., 2011. The Parkinson progression marker initiative (PPMI). Prog. Neurobiol. http://dx.doi.org/10.1016/j.pneurobio.2011.09.005.
Margolis, R., Derr, L., Dunn, M., Huerta, M., Larkin, J., Sheehan, J., Guyer, M., Green, E.D., 2014. The National Institutes of Health's big data to knowledge (BD2K) initiative: capitalizing on biomedical big data. J. Am. Med. Inform. Assoc. 21, 957–958. http://dx.doi.org/10.1136/amiajnl-2014-002974.
Payakachat, N.N., Tilford, J.M., Ungar, W.J., 2016. National Database for Autism Research (NDAR): big data opportunities for health services research and health technology assessment. PharmacoEconomics 34, 127–138. http://dx.doi.org/10.1007/s40273-015-0331-6.
Rakap, S., Jones, H.A., Emery, A.K., 2015. Evaluation of a web-based professional development program (project ACE) for teachers of children with autism spectrum disorders. Teach. Educ. Spec. Educ. 38, 221–239. http://dx.doi.org/10.1177/0888406414535821.
Salvador, R., Suckling, J., Schwarzbauer, C., Bullmore, E., 2005. Undirected graphs of frequency-dependent functional connectivity in whole brain networks. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 360, 937–946. http://dx.doi.org/10.1098/rstb.2005.1645.
Schneider, L.S., Insel, P.S., Weiner, M.W., 2011. Treatment with cholinesterase inhibitors and memantine of patients in the Alzheimer's disease neuroimaging initiative. Arch. Neurol. 68, 58–66. http://dx.doi.org/10.1001/archneurol.2010.343.
Schulze-Bonhage, A., Ihle, M., Sales, F., Navarro, V., Dourado, A., 2010. A European EEG database of epilepsy patients EPILEPSIAE. Clin. Neurophysiol. 121, S200.
Shattuck, D.W., Leahy, R.M., 2002. Brainsuite: an automated cortical surface identification tool. Med. Image Anal. 6, 129–142. http://dx.doi.org/10.1016/S1361-8415(02)00054-3.
Sierra-Marcos, A., Scheuer, M.L., Rossetti, A.O., 2015. Seizure detection with automated EEG analysis: a validation study focusing on periodic patterns. Clin. Neurophysiol. 126, 456–462. http://dx.doi.org/10.1016/j.clinph.2014.06.025.
Stam, C.J., 2004. Functional connectivity patterns of human magnetoencephalographic recordings: a “small-world” network? Neurosci. Lett. 355, 25–28. http://dx.doi.org/10.1016/j.neulet.2003.10.063.
Talmon, R., Kushnir, D., Coifman, R.R., Cohen, I., Gannot, S., 2012. Parametrization of linear systems using diffusion kernels. IEEE Trans. Signal Process. 60, 1159–1173. http://dx.doi.org/10.1109/TSP.2011.2177973.
Talukdar, P.P., Ives, Z.G., Pereira, F., 2010. Automatically incorporating new sources in keyword search-based data integration. In: Proc. 2010 Int. Conf. Manag. data - SIGMOD. Vol. 10. pp. 387–398. http://dx.doi.org/10.1145/1807167.1807211.
Thompson, P.M., Stein, J.L., Medland, S.E., Hibar, D.P., Vasquez, A.A., Renteria, M.E., Toro, R., Jahanshad, N., Schumann, G., Franke, B., Wright, M.J., Martin, N.G., Agartz, I., Alda, M., Alhusaini, S., Almasy, L., Almeida, J., Alpert, K., Andreasen, N.C., Andreassen, O.A., Apostolova, L.G., Appel, K., Armstrong, N.J., Aribisala, B., Bastin, M.E., Bauer, M., Bearden, C.E., Bergmann, Ø., Binder, E.B., Blangero, J., Bockholt, H.J., Bøen, E., Bois, C., Boomsma, D.I., Booth, T., Bowman, I.J., Bralten, J., Brouwer, R.M., Brunner, H.G., Brohawn, D.G., Buckner, R.L., Buitelaar, J., Bulayeva, K., Bustillo, J.R., Calhoun, V.D., Cannon, D.M., Cantor, R.M., Carless, M.A., Caseras, X., Cavalleri, G.L., Chakravarty, M.M., Chang, K.D., Ching, C.R.K., Christoforou, A., Cichon, S., Clark, V.P., Conrod, P., Coppola, G., Crespo-Facorro, B., Curran, J.E., Czisch, M., Deary, I.J., de Geus, E.J.C., den Braber, A., Delvecchio, G., Depondt, C., de Haan, L., de Zubicaray, G.I., Dima, D., Dimitrova, R., Djurovic, S., Dong, H., Donohoe, G., Duggirala, R., Dyer, T.D., Ehrlich, S., Ekman, C.J., Elvsåshagen, T., Emsell, L., Erk, S., Espeseth, T., Fagerness, J., Fears, S., Fedko, I., Fernández, G., Fisher, S.E., Foroud, T., Fox, P.T., Francks, C., Frangou, S., Frey, E.M., Frodl, T., Frouin, V., Garavan, H., Giddaluru, S., Glahn, D.C., Godlewska, B., Goldstein, R.Z., Gollub, R.L., Grabe, H.J., Grimm, O., Gruber, O., Guadalupe, T., Gur, R.E., Gur, R.C., Göring, H.H.H., Hagenaars, S., Hajek, T., Hall, G.B., Hall, J., Hardy, J., Hartman, C.A., Hass, J., Hatton, S.N., Haukvik, U.K., Hegenscheid, K., Heinz, A., Hickie, I.B., Ho, B.C., Hoehn, D., Hoekstra, P.J., Hollinshead, M., Holmes, A.J., Homuth, G., Hoogman, M., Hong, L.E., Hosten, N., Hottenga, J.J., Hulshoff Pol, H.E., Hwang, K.S., Jack, C.R., Jenkinson, M., Johnston, C., Jönsson, E.G., Kahn, R.S., Kasperaviciute, D., Kelly, S., Kim, S., Kochunov, P., Koenders, L., Krämer, B., Kwok, J.B.J., Lagopoulos, J., Laje, G., Landen, M., Landman, B.A., Lauriello, J., Lawrie, S.M., Lee, P.H., Le Hellard, S., Lemaître, H., Leonardo, C.D., Li, C. Shan, Liberg, B., Liewald, D.C., Liu, X., Lopez, L.M., Loth, E., Lourdusamy, A., Luciano, M., Macciardi, F., Machielsen, M.W.J., MacQueen, G.M., Malt, U.F., Mandl, R., Manoach, D.S., Martinot, J.L., Matarin, M., Mather, K.A., Mattheisen, M., Mattingsdal, M., Meyer-Lindenberg, A., McDonald, C., McIntosh, A.M., McMahon, F.J., McMahon, K.L., Meisenzahl, E., Melle, I., Milaneschi, Y., Mohnke, S., Montgomery, G.W., Morris, D.W., Moses, E.K., Mueller, B.A., Muñoz Maniega, S., Mühleisen, T.W., Müller-Myhsok, B., Mwangi, B., Nauck, M., Nho, K., Nichols, T.E., Nilsson, L.G., Nugent, A.C., Nyberg, L., Olvera, R.L., Oosterlaan, J., Ophoff, R.A., Pandolfo, M., Papalampropoulou-Tsiridou, M., Papmeyer, M., Paus, T., Pausova, Z., Pearlson, G.D., Penninx, B.W., Peterson, C.P., Pfennig, A., Phillips, M., Pike, G.B., Poline, J.B., Potkin, S.G., Pütz, B., Ramasamy, A., Rasmussen, J., Rietschel, M., Rijpkema, M., Risacher, S.L., Roffman, J.L., Roiz-Santiañez, R., Romanczuk-Seiferth, N., Rose, E.J., Royle, N.A., Rujescu, D., Ryten, M., Sachdev, P.S., Salami, A., Satterthwaite, T.D., Savitz, J., Saykin, A.J., Scanlon, C., Schmaal, L., Schnack, H.G., Schork, A.J., Schulz, S.C., Schür, R., Seidman, L., Shen, L., Shoemaker, J.M., Simmons, A., Sisodiya, S.M., Smith, C., Smoller, J.W., Soares, J.C., Sponheim, S.R., Sprooten, E., Starr, J.M., Steen, V.M., Strakowski, S., Strike, L., Sussmann, J., Sämann, P.G., Teumer, A., Toga, A.W., Tordesillas-Gutierrez, D., Trabzuni, D., Trost, S., Turner, J., Van den Heuvel, M., van der Wee, N.J., van Eijk, K., van Erp, T.G.M., van Haren, N.E.M., van't Ent, D., van Tol, M.J., Valdés Hernández, M.C., Veltman, D.J., Versace, A., Völzke, H., Walker, R., Walter, H., Wang, L., Wardlaw, J.M., Weale, M.E., Weiner, M.W., Wen, W., Westlye, L.T., Whalley, H.C., Whelan, C.D., White, T., Winkler, A.M., Wittfeld, K., Woldehawariat, G., Wolf, C., Zilles, D., Zwiers, M.P., Thalamuthu, A., Schofield, P.R., Freimer, N.B., Lawrence, N.S., Drevets, W., 2014. The ENIGMA consortium: large-scale collaborative analyses of neuroimaging and genetic data. Brain Imaging Behav. 8, 153–182. http://dx.doi.org/10.1007/s11682-013-9269-5.
Toga, A.W., Crawford, K.L., 2010a. The informatics core of the Alzheimer's disease neuroimaging initiative. Alzheimers Dement. http://dx.doi.org/10.1016/j.jalz.2010.03.001.
Toga, A.W., Crawford, K.L., 2010b. The informatics core of the Alzheimer's disease neuroimaging initiative. Alzheimers Dement. 6, 247–256. http://dx.doi.org/10.1016/j.jalz.2010.03.001.
Toga, A.W., Neu, S.C., Bhatt, P., Crawford, K.L., Ashish, N., 2016. The Global Alzheimer's Association interactive network. Alzheimers Dement. 12, 49–54. http://dx.doi.org/10.1016/j.jalz.2015.06.1896.
Torgerson, C.M., Irimia, A., Leow, A.D., Bartzokis, G., Moody, T.D., Jennings, R.G., Alger, J.R., van Horn, J.D., Altshuler, L.L., 2013. DTI tractography and white matter fiber tract characteristics in euthymic bipolar I patients and healthy control subjects. Brain Imaging Behav. 7, 129–139. http://dx.doi.org/10.1007/s11682-012-9202-3.
Torgerson, C.M., Quinn, C., Dinov, I., Liu, Z., Petrosyan, P., Pelphrey, K., Haselgrove, C., Kennedy, D.N., Toga, A.W., Van Horn, J.D., 2015. Interacting with the National Database for Autism Research (NDAR) via the LONI pipeline workflow environment. Brain Imaging Behav. 9, 89–103. http://dx.doi.org/10.1007/s11682-015-9354-z.
Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E.J., Yacoub, E., Ugurbil, K., 2013. The WU-Minn human connectome project: an overview. NeuroImage 80, 62–79. http://dx.doi.org/10.1016/j.neuroimage.2013.05.041.
Van Horn, J.D., Toga, A.W., 2009. Is it time to re-prioritize neuroimaging databases and digital repositories? NeuroImage 47, 1720–1734. http://dx.doi.org/10.1016/j.neuroimage.2009.03.086.
van Horn, J.D., Irimia, A., Torgerson, C.M., Chambers, M.C., Kikinis, R., Toga, A.W., 2012. Mapping connectivity damage in the case of phineas gage. PLoS One 7. http://dx.doi.org/10.1371/journal.pone.0037454.
Vespa, P.M., Shrestha, V., Abend, N., Agoston, D., Au, A., Bell, M.J., Bleck, T.P., Buitrago Blanco, M., Claassen, J., Diaz-Arrastia, R., Duncan, D., Ellingson, B., Foreman, B., Gilmore, E.J., Hirsch, L., Hunn, M., Kamnaksh, A., McArthur, D., Morokoff, A., O'Brien, T., O'Phelan, K., Robertson, C.L., Rosenthal, E., Staba, R., Toga, A., Willyerd, F.A., Zimmermann, L., Real, C., Martinez, S., Yam, E., Engel Jr., J., Group, F. the E.S., 2018. The epilepsy bioinformatics epilepsy study for anti-epileptogenic therapy (Epibios4Rx) clinical biomarker: study design and protocol. Neurobiol. Dis.
Wagenaar, J.B., Worrell, G.A., Ives, Z., Matthias, D., Litt, B., Schulze-Bonhage, A., 2015. Collaborating and sharing data in epilepsy research. J. Clin. Neurophysiol. http://dx.doi.org/10.1097/WNP.0000000000000159.
Wedeen, V.J., Wang, R.P., Schmahmann, J.D., Benner, T., Tseng, W.Y.I., Dai, G., Pandya, D.N., Hagmann, P., D'Arceuil, H., de Crespigny, A.J., 2008. Diffusion spectrum

magnetic resonance imaging (DSI) tractography of crossing fibers. NeuroImage 41, 1267–1277. http://dx.doi.org/10.1016/j.neuroimage.2008.03.036.

Wilcox, R., 2012. Introduction to Robust Estimation and Hypothesis Testing. http://dx.doi.org/10.1016/B978-0-12-386983-8.00015-9.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M.ge, Baak, A., Blomberg, N., Boiten, J.W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., Mons, B., 2016. The FAIR guiding principles for scientific data management and stewardship. Sci Data. 3, 160018. http://dx.doi.org/10.1038/sdata.2016.18.

Winden, K.D., Karsten, S.L., Bragin, A., Kudo, L.C., Gehman, L., Ruidera, J., Geschwind, D.H., Engel, J., 2011. A systems level, functional genomics analysis of chronic epilepsy. PLoS One 6. http://dx.doi.org/10.1371/journal.pone.0020763.

Winden, K.D., Bragin, A., Engel, J., Geschwind, D.H., 2015. Molecular alterations in areas generating fast ripples in an animal model of temporal lobe epilepsy. Neurobiol. Dis. 78, 35–44. http://dx.doi.org/10.1016/j.nbd.2015.02.011.

Wold, S., Esbensen, K., Geladi, P., 1987. Principal component analysis. Chemom. Intell. Lab. Syst. 2, 37–52. http://dx.doi.org/10.1016/0169-7439(87)80084-9.