PROTOCOL

Open Access

Mining multi-site clinical data to develop machine learning MRI biomarkers: application to neonatal hypoxic ischemic encephalopathy

Rebecca J. Weiss[1†] , Sara V. Bates[1†], Ya'nan Song[2], Yue Zhang[2], Emily M. Herzberg[1], Yih-Chieh Chen[1], Maryann Gong[3], Isabel Chien[3], Lily Zhang[3], Shawn N. Murphy[4], Randy L. Gollub[5] , P. Ellen Grant[2,6*†] and Yangming Ou[2,6,7*†]

Abstract

Background: Secondary and retrospective use of hospital-hosted clinical data provides a time- and cost-efficient alternative to prospective clinical trials for biomarker development. This study aims to create a retrospective clinical dataset of Magnetic Resonance Images (MRI) and clinical records of neonatal hypoxic ischemic encephalopathy (HIE), from which clinically-relevant analytic algorithms can be developed for MRI-based HIE lesion detection and outcome prediction.

Methods: This retrospective study will use clinical registries and big data informatics tools to build a multi-site dataset that contains structural and diffusion MRI, clinical information including hospital course, short-term outcomes (during infancy), and long-term outcomes (~ 2 years of age) for at least 300 patients from multiple hospitals.

Discussion: Within machine learning frameworks, we will test whether the quantified deviation from our recently-developed normative brain atlases can detect abnormal regions and predict outcomes for individual patients as accurately as, or even more accurately, than human experts.

Trial Registration Not applicable. This study protocol mines existing clinical data thus does not meet the ICMJE definition of a clinical trial that requires registration

Keywords: Neonatal encephalopathy, Hypoxic ischemic encephalopathy, MRI, Biomarkers, Machine learning, Outcome prediction, Bioinformatics

Background

Hypoxic ischemic encephalopathy (HIE) affects 1–5/1000 of live births, and is a leading cause of morbidity and mortality in childhood [1, 2]. Although the implementation of therapeutic hypothermia (TH) reduces infant mortality and chronic disability (by 2 years of age) [3–5], neurodevelopmental impairments are still common in survivors [6–8]. Specific impairments vary across surviving patients, motivating the development of prognostic biomarkers. There is progress in developing clinical [9–11], biochemical [9–12], and serum [12, 13] biomarkers. However, it remains unclear whether or not MRI can serve as a non-invasive and highly sensitive biomarker to improve outcome prediction in the early postnatal period [14–16]. Indeed, in the 108 clinical trials that are ongoing for HIE worldwide [17] (Fig. 1), MRI is used in over half of them to assess all stages of HIE management including: diagnosis, prevention, prognosis, intervention, and rehabilitation.

Currently, expert scoring of neonatal MRI is used in clinical trials to predict 2-year outcomes [18–23], but

*Correspondence: Ellen.Grant@childrens.harvard.edu; Yangming.
Ou@childrens.harvard.edu
†Rebecca J. Weiss and Sara V. Bates—co-first authors
†P. Ellen Grant and Yangming Ou—co-senior authors
² Fetal Neonatal Neuroimaging and Developmental Science Center
(FNNDSC), Boston Children's Hospital, Harvard Medical School, 401 Park
Drive, Landmark Center 7022, Boston, MA 02115, USA
Full list of author information is available at the end of the article

Weiss *et al. J Transl Med* (2019) 17:385

Page 2 of 16

Fig. 1 Need for MRI in HIE-related clinical trials. Each icon notes a hospital/site where at least one HIE-related clinical trial is ongoing. Red icons are hospitals that use MRI and blue icons are those that do not use MRI in their trials. Among 108 ongoing clinical trials pertaining to HIE at hospitals from 33 countries in 5 continents, roughly half of the hospitals use MRI as part of their trials, highlighting the widespread need for MRI biomarkers that can detect HIE lesions at infancy and predict HIE outcomes at 2 years of age. This figure was created based on searching the key word "Hypoxic Ischemic Encephalopathy" in the public website for clinical trial registries (https://clinicaltrials.gov). The search was in June 2019. We manually added each site in all 108 resulting HIE trials on the Google My Map website

limitations exist. *First*, experts score MRI by looking for lesion presence or absence in selected key brain regions (e.g., thalamus, basal ganglia, etc.), and by looking at whether HIE lesions are unilateral or bilateral, locally confined or globally distributed, etc. Different scoring systems range from 6 score levels [19] to recently 57 score levels [20]. As scoring criteria get more complex, expert scoring takes more time, requires more training and becomes more uncertain (20–40% intra-/inter-reader variability [15]). *Second*, sensitivity can vary in multi-site data, with a detailed scoring system developed in 2018 reporting a 92.3% sensitivity in one cohort but 42.1% in another cohort [20]. *Third and more importantly*, up to 50% of HIE patients have MRIs visually interpreted as normal by experts [24, 25], despite 5–8% of them having adverse outcomes at 18–22 months [4, 11, 26]. This remains unexplained to date. To address these three limitations in expert MRI scoring systems, our goal is to develop quantitative, objective and machine learning-powered algorithms and software to detect HIE lesions during neonatal stages and predict 2-year HIE outcomes.

Machine learning (ML) approaches identify lesions, extract MRI injury features, and find the feature subset (i.e., patterns) that best predicts outcomes [27, 28]. Accuracy is measured by comparing the predicted outcome (assuming unknown during prediction) with the actually known outcome [29]. The promise has been shown in brain tumors [30–32], Alzheimer's Disease (AD) [33–36], neuroscience [37, 38], stroke [39], epilepsy [40], psychiatric disorders [41–43], traumatic brain injury (TBI) [44, 45], pediatric brain tumor [46], etc. HIE poses 3 unique challenges to ML: (i) lack of data: public data exists for hundreds of patients with brain tumors [47, 48], AD [49, 50], etc.; however, annotated MRIs with linked clinical and outcome data are rare in HIE and none exists publicly. (ii) unique difficulty in lesion detection: radiologists look for regions of abnormal signals in T1-/T2-weighted MRI, diffusion-weighted image (DWI) and Apparent Diffusion Coefficient (ADC) maps [19, 51]. However, HIE-induced T1, T2, DWI and ADC changes rapidly evolve and are entangled with rapid normal neonatal brain development [51–53]. This is not an issue in mature brains; (iii) normal MRI but adverse outcome: 20–50% HIE patients have normal MRI (i.e., no lesion detected visually) [24, 25], but 5–8% of them still develop adverse outcomes [3, 11, 26]. This cannot be explained by expert scores (standard for HIE) or current ML methods (designed for other diseases), which require explicit lesion detection [28, 29].

Our study has three novelties to address these three challenges. *First*, we will retrospectively collect multi-site clinical data that can be used to develop MRI analysis

Weiss *et al. J Transl Med* (2019) 17:385

Page 3 of 16

tools. Clinical data includes demographics, hospital assessments, treatment, MRI, as well as outcomes at neonatal intensive care unit (NICU) discharge and outcomes at 2 years. The secondary use of hospital-hosted clinical data has received increasing attention with value added clinical results [54–56]. We plan to use registry- and informatics-driven approaches to retrospectively pull and regularly update data from NICU and hospital-hosted clinical archives at Massachusetts General Hospital (MGH) and Boston Children's Hospital (BCH). This is different from many clinical trials, which require significant funding and a pre-set timeframe (often years) to prospectively collect patient data. In contrast, we anticipate a relatively low cost and shorter time frame for data collection; the data will be continuously evolving as new patients or visits are added. Moreover, starting with actual clinical data will facilitate the translation to clinical practice. We present our planned efforts in data collection and MRI analysis algorithm design. Special emphasis is placed on addressing the quality of the clinical data, especially multi-site data. *Second*, we propose a new ML framework specific for HIE lesion detection, where the uniqueness is to disentangle HIE-induced and normal-development-related ADC signals in ML. The specific hypothesis we plan to test is that the quantitative deviation from normative neonatal brain Apparent Diffusion Coefficient (ADC) atlases, which we developed recently [57], can facilitate quantitative and automated HIE lesion detection and outcome prediction at an accuracy comparable or higher than experts. We will use ADC maps derived from diffusion tensor MRI [58, 59], as they are commonly used to identify HIE lesions clinically in the first week of age [24, 60]. Radiologists identify lesions by searching for regions of abnormally low ADC values corresponding to decreased water diffusion [24, 60]. However, the normal ranges of ADC variations vary in space (different brain regions) and in time (as the brain develops rapidly in infancy), making expert interpretation error-prone [15, 61]. We recently developed the first-of-its-kind normative ADC atlases, which quantified the normal range of ADC variations in space and in time [57] (see Figs. 4 and 5). Based on this, we can objectively quantify the deviation of a patient's ADC values from

normal variations at every voxel (i.e., 3D pixel) in the brain [62]. Deep learning lesion detection frameworks still apply. But, instead of using voxels from training patients that ignored the normal variations in their spatiotemporal locations, which is not a problem in mature brains, we will feed new channels (we term Z_{ADC} map) to specifically separate HIE-related ADC changes from spatiotemporal ADC changes from normal neonatal brain development. *Third*, other than current radiomics approaches that mostly rely on explicit lesion detection, we will develop a novel "radiomics without lesion detection" approach, which relies on regional and tract-wise MRI features throughout the brain to address the unique issue in HIE that some neonates with clinically-normal MRI (no detectable lesions) may still develop adverse 2-year outcome.

Methods and design

Overview

This study is approved by the Institutional Review Board at MGH and BCH. Figure 2 outlines the three key components in our study.
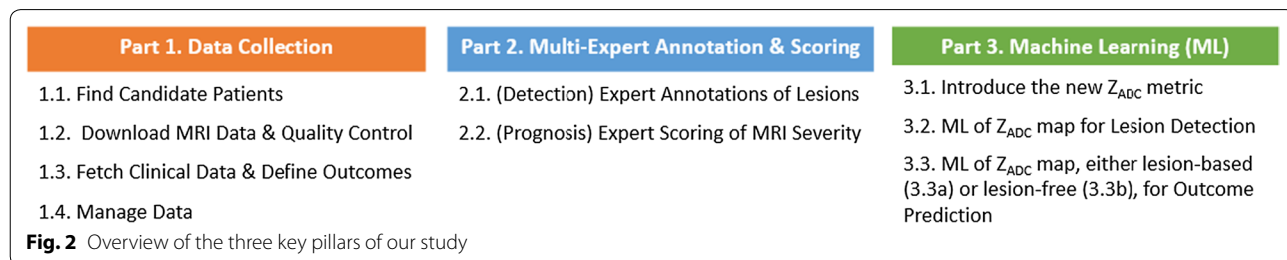
Part 1. Data collection

Figure 2 (Part 1) shows the major steps in data collection.

Part1.1. Find candidate patients

We will use two sources to find candidate patients. Our primary source is the NICU registry. The NICU registry contains patient's diagnosis, medical record numbers (MRNs), and demographic information (birth weight, gestational age, etc.). We will query the NICU registries for patients who

a. were term born (> 36 weeks gestation);
b. had a clinical diagnosis of HIE and were free of other major neurological disorders.

A second source is the hospital-wide database. In the big data era, an increasing number of hospitals around the world have informatics tools that allow authorized personnel to search patients by diagnosis, or by the International Classification of Diseases (ICD) codes [63–65].

Part 1. Data Collection	Part 2. Multi-Expert Annotation & Scoring	Part 3. Machine Learning (ML)
1.1. Find Candidate Patients	2.1. (Detection) Expert Annotations of Lesions	3.1. Introduce the new Z_{ADC} metric
1.2. Download MRI Data & Quality Control	2.2. (Prognosis) Expert Scoring of MRI Severity	3.2. ML of Z_{ADC} map for Lesion Detection
1.3. Fetch Clinical Data & Define Outcomes		3.3. ML of Z_{ADC} map, either lesion-based (3.3a) or lesion-free (3.3b), for Outcome Prediction
1.4. Manage Data		

Fig. 2 Overview of the three key pillars of our study

Weiss *et al. J Transl Med*   (2019) 17:385

Page 4 of 16

From our registry data, we identified a list of ICD codes commonly used for HIE (Table 1). We will use these ICD codes to search patients that are not captured in the registry.

Part1.2. Download MRI data and quality control

We will use medical record numbers (MRNs) of candidate patients to search and download their MRI data. MGH's mi2b2 workbench [66] allows authorized users to find and copy DICOM-format MRIs from the Radiology archives to a local cache. BCH's ChRIS platform offers the same function [103, 104]. From the candidate patients identified in the previous step, we will include those with:

a. both structural and diffusion MRIs. At BCH and MGH, the T1-weighted MPRAGE sequence typically has a 1 mm isotropic high resolution. Diffusion MRI typically has a 2 mm isotropic resolution, with at least 6 (often 20–35) gradient directions at b value 1000 s/mm^2.

b. reasonable image quality (e.g., no severe motion or artifacts in either MRI sequences), as visually reviewed by a trained assistant.

Part1.3. Fetch clinical data, define outcomes

Clinical variables   We will include the following maternal variables: maternal demographics, parity, significant medical history, prescription medications during pregnancy, alcohol/tobacco/elicit substance use, mode of delivery, complications around delivery (e.g. chorioamnionitis, prolonged 2nd stage of labor), or a sentinel event (e.g. fetal bradycardia, uterine rupture, umbilical cord prolapse), and placental pathology, if available. We will also include the following infant data: anthropometric measurements, APGAR scores, umbilical cord gas and/or the infant's initial blood gas (if available), medications administered during the initial hospitalization, the presence of clinical or electrographic seizures, mode of feeding at hospital discharge, abnormalities on the discharge physical examination, length of stay, and discharge dispo-

sition (e.g. deceased, home, transferred). A trained expert will obtain this information from the electronic health records (EHRs). Absence of explicit information will not be used as a negative finding.

Outcomes   We will also retrieve from EHR:

a. outcome at NICU discharge: deceased or survival;
b. outcomes at 2 years of age: neurocognition at 18–24 months, including both continuously-valued developmental assessments and categorical outcomes, as listed in Table 2.

Part1.4. Manage data

We will use REDCap [67], a HIPAA-compliant, secure, and user-friendly web application, to facilitate manual entry of clinical variables from the EHR and expert review. Entries into REDCap will be reviewed by collaborating neonatologists.

MRI data will be anonymized, stored, and analyzed using HIPAA-compliant computers and high-performance computer clusters as provided by MGH and BCH.

Part 2. Multi-expert annotation and scoring

Expert opinions will serve as references to validate the proposed MRI analysis tools (Fig. 2, Part 2).

Part 2.1. Annotations of lesions

We will use expert consensus as the Ref. [68], since postmortem histologic specimens are rarely available. We plan to have 3 experts (> 5 years of clinical pediatric neuroradiology experience) independently annotate HIE lesions on each ADC map. While there may be uncertainties among experts [15, 61], the consensus will be more accurate compared to individual expert annotation. We define consensus as found by the STAPLE tool, which, loosely speaking, is an improvement of majority voting by computing a probabilistic estimate of the true regional annotation from multiple experts' annotations [69]. Given that up to 50% of HIE patients may not

Table 1 ICD codes for HIE as a secondary source to query candidate patients

ICD-9	Meaning	ICD-10	Meaning
768.70	Hypoxic-ischemic encephalopathy, unspecified	*P91.60*	Hypoxic ischemic encephalopathy [HIE], unspecified
768.71	Mild hypoxic-ischemic encephalopathy	*P91.61*	Mild hypoxic ischemic encephalopathy [HIE]
768.72	Moderate hypoxic-ischemic encephalopathy	*P91.62*	Moderate hypoxic ischemic encephalopathy [HIE]
768.73	Severe hypoxic-ischemic encephalopathy	*P91.63*	Severe hypoxic ischemic encephalopathy [HIE]
779.2	Cerebral depression, coma, and other abnormal cerebral signs in fetus or newborn	*P91.4*	Neonatal cerebral depression

Italic plain font for ICD-9 (left half of the table) and italic font for ICD-10 codes (right half of the table)

Weiss *et al. J Transl Med*   (2019) 17:385

Page 5 of 16

Table 2 Definition of long-term neurocognitive outcomes at ~ 2 years of age

i. Continuously-valued outcome	Numerical domain scores
BSID-III (Bayley Scale of Infant Development, Version III)	Cognitive (ranging 50–150) Language (ranging 50–150) Motor (ranging 50–150)

ii. Binary-valued outcome	Definition
Developmental delay	YES (1) if BSID-III ≤ 85 in any domain OR any documentation by a physician that there is a significant delay in motor, cognitive, or language development that requires intervention; otherwise NO (0)
Cerebral palsy	As documented in the EHR by the medical provider
Motor impairment	YES (1) if BSID-III ≤ 85 in the motor domain OR any noted motor abnormality documented in the EHR by a medical provider; or NO (0)
Visual/hearing impairment	As documented in the EHR by a medical provider

Table 3 The NICHD–NRN scoring system [19]

NICHD–NRN Scores, 2012 [19]	Criteria
0	Normal
1A	Minimal cerebral lesions only, without basal ganglia thalamus (BGT), anterior limb of internal capsule (ALIC), posterior limb of internal capsule (PLIC) or watershed (WS) infarction
1B	More extensive cerebral lesions, without BGT, ALIC, PLIC or WS infarction
2A	Any BGT, ALIC, PLIC or WS infarction without any other cerebral lesions
2B	Either BGT, ALIC, PLIC or WS infarction AND any other cerebral lesions
3	Hemispheric devastation

Fig. 3 Flowchart for ML-driven lesion detection and outcome prediction

have clinically-detectable lesions in the neonatal MRI, an expert will leave a blank annotation if he/she does not find lesions in a patient.

Part 2.2. Expert scores to predict outcomes

Predicting outcomes requires a different approach. The "ground truth" is available from the clinical records. We will use at least 2 experts to score the severity of the neonatal MRI and to test whether the proposed algorithms/tools outperform expert scores in predicting the "ground-truth" outcomes. The expert scores will be based on the NICHD–NRN scoring criteria [19] (National Institute of Children Health and Human Development, Neonatal Research Network), as listed in Table 3.

Part 3. Developing machine learning algorithms/tools for lesion detection and outcome prediction

Figure 3 shows the flowchart of Part 3. We start from feature extraction (Part 3.1, Z_{ADC} calculation). The raw ADC map and the novel Z_{ADC} feature map will be fed to ML-based lesion detection (Part 3.2). Depending
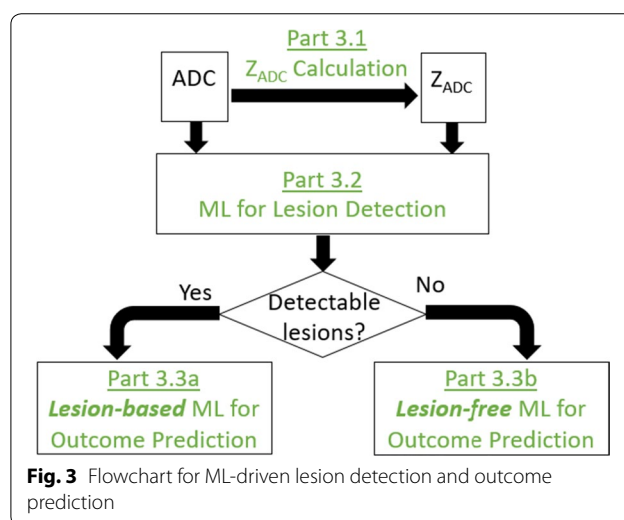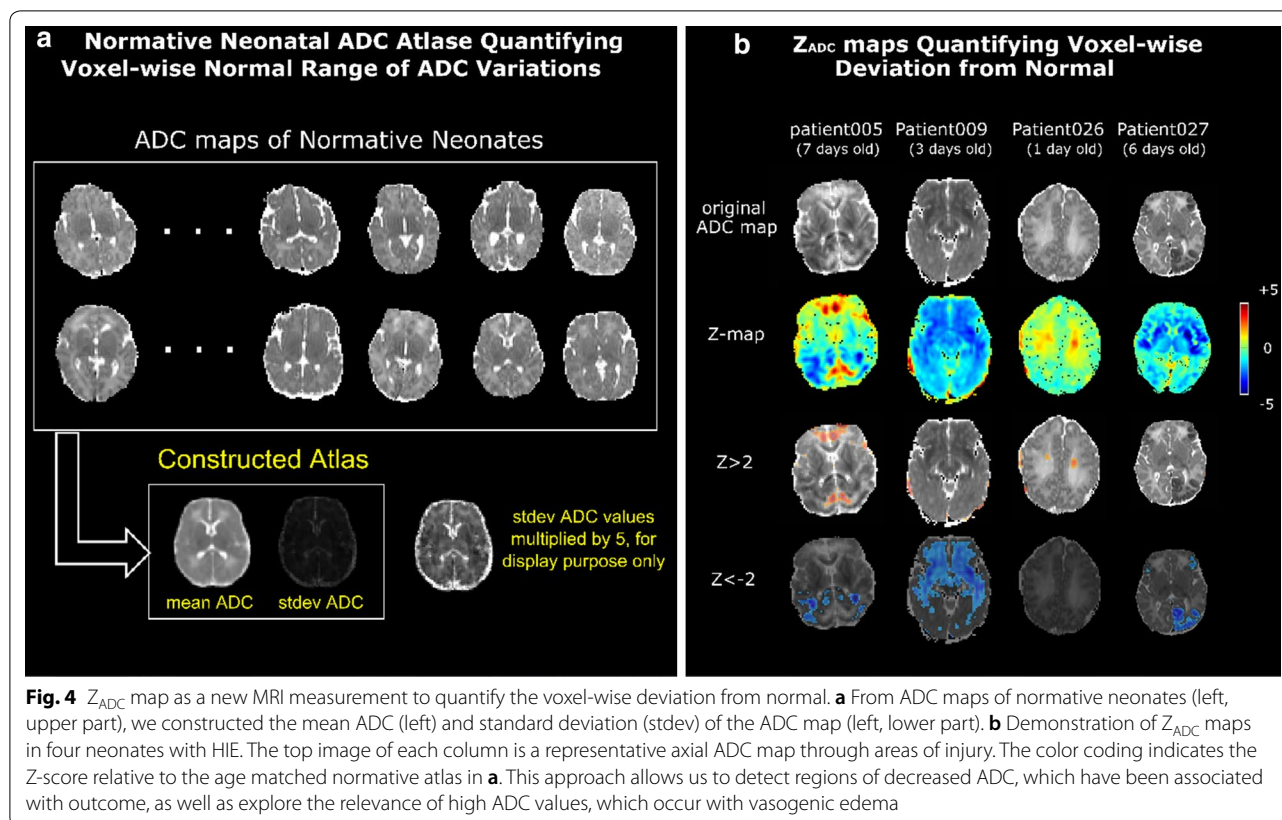
on whether there are detectable lesions in the patient, ML-driven outcome prediction (Part 3.3) will either go through lesion-based outcome prediction, which will use ADC, Z_{ADC} and detected lesions as input (Part 3.3a), or go through lesion-free outcome prediction, which will use ADC and Z_{ADC} for outcome prediction (Part 3.3b).

Part3.1. Feature extraction—introducing the Z_{ADC} measurement

Uncertainty in the visual interpretation of neonatal brain ADC maps arises for a number of reasons including the rapidity of brain development during infancy (i.e., temporal uncertainty) and the variation of normal ADC values across different brain regions (i.e., spatial uncertainty). Our recently developed normative ADC atlases quantified the mean and standard deviation (stdev) ADC values at every voxel in the brain in a normative cohort of 13 term-born neonates who were scanned in the first 2 weeks, with a median age of 4 days at the time of MRI scan (see Fig. 4a) [57]. This allows us to quantitatively compare a

Weiss *et al. J Transl Med*   (2019) 17:385

Page 6 of 16

Fig. 4 Z_{ADC} map as a new MRI measurement to quantify the voxel-wise deviation from normal. **a** From ADC maps of normative neonates (left, upper part), we constructed the mean ADC (left) and standard deviation (stdev) of the ADC map (left, lower part). **b** Demonstration of Z_{ADC} maps in four neonates with HIE. The top image of each column is a representative axial ADC map through areas of injury. The color coding indicates the Z-score relative to the age matched normative atlas in **a**. This approach allows us to detect regions of decreased ADC, which have been associated with outcome, as well as explore the relevance of high ADC values, which occur with vasogenic edema

patient's ADC value y(u) at a voxel u (first row of panel b) to the mean μ(v) and standard deviation σ(v) ADC values at the anatomically-corresponding location v on the atlas (Fig. 4a). Here the correspondence will be found by the extensively-validated [70] patient-to-atlas DRAMMS deformable registration [71]. This will convert a patient's ADC value into a Z_{ADC} value, i.e., $Z_{ADC}(u) = [y(u) - \mu(v)]/\sigma(v)$, for each voxel u in the patient space (second row of panel 7b). The Z_{ADC} value quantifies the deviation from normal at every voxel in a patient [62].

Our pilot results in 8 patients showed that regions with $Z_{ADC} < -2$ (i.e., last row in Fig. 4b) identified HIE lesions at an accuracy comparable to consensus of time consuming expert labelling (between experts Dice overlap at 71% and the average algorithm-and-expert Dice overlap at 69%) [72, 73]. The next steps will include testing the effects of the novel Z_{ADC} map for lesion detection in larger cohorts and in data from more institutions, as enabled by the planned dataset. Finding spatial lesions patterns informative of outcomes, and eventually, patient-centered outcome prediction is the goal (Fig. 4). Machine learning (ML) is well suited for these tasks.

Part3.2. Machine learning of Z_{ADC} and ADC for automated lesion detection

Preprocessing   We will convert series of 2D DICOM files into an integral 3D NIfTI file for each MR sequence. We will anonymize the patient name and the date of birth in the NIfTI header. We will perform N4 to correct for the inhomogeneity of MRI signals caused by inhomogeneity of the magnetic field in the scanner [74], we will do field-of-view normalization to make sure brain MRIs are of the same scope (from brain stem to the top of the brain, excluding any neck, shoulder or even upper chest that were included in the raw scan) [75], we will skull strip the brain MRI to keep only the brain and remove the eyes, faces, skull, etc. [76, 77]; we will do automated segmentation [78] on the MPRAGE structural image to parcellate the brain into 62 anatomic structures, and we will nonrigidly map the segmented anatomic regions to the diffusion MRI space [71] so that diffusion MRI including the ADC maps will also be segmented into structures.

Question 1: Can Z_{ADC} detect HIE lesions more accurately than single experts?   The calculation of Z_{ADC} map is essentially a feature calculation or feature extraction step, which we consider as part of machine learning in this paper. We will test the accuracy of atlas-based lesion detection by the sensitivity, specificity and Dice over-

Weiss *et al. J Transl Med* (2019) 17:385

Page 7 of 16

lap between computer-generated and expert consensus labelled lesion regions [72, 73]. One way of using our novel Z_{ADC} feature map is to simply threshold Z_{ADC} values at each voxel (Fig. 4). This is analogous to a Bayesian classifier—from normal controls we have built a Gaussian model of normal distributions of ADC values at each voxel, and given a new patient's ADC value at this value, Z_{ADC} is related to the likelihood of this patient's this voxel being lesioned or not. We will first test simply thresholding the Z_{ADC} map at various threshold values (e.g., -1, -1.5, -2, -2.5). Another way of using the novel Z_{ADC} map is to use more complex machine classifiers to identify whether every voxel in the brain is affected by HIE lesions (e.g., voxel-wise normal-vs-lesion machine classification), given the ADC and Z_{ADC} values at this voxel and in the geometric neighborhood of this voxel. We will test whether deep learning classifiers [79, 80] (e.g., 2D and 3D U-Net [81], V-Net [82]), which is free of hand-crafted features and characterizes each voxel by its multi-scale neighborhood information, offer additional advantages over classic classifiers that rely on hand-crafted features of a voxel (e.g., Support Vector Machine [30], and Random Forest [68]). We will test the effect of using the Z_{ADC} value at each voxel as inputs with and without the ADC values of the voxel, and quantify whether this improves lesion detection accuracy compared with using the ADC values alone [68]. We also plan to test whether post-processing based on prior knowledge can further improve the accuracy of lesion segmentation. One post-processing can be the opening (i.e., dilation) and/or closing (i.e., erosion) morphological operations. Another post-processing is to regularize the computer-detected regions with voxel-wise probability of lesion occurrence in HIE populations (see Question 4 and Fig. 5a). Frameworks for lesion-atlas guided/regularized lesion detection can be used [83].

We will use Dice overlap and receiver-operating-curves (ROCs) to quantify the accuracy with regard to expert consensus labelled lesion regions. We will compare our accuracy with the literature [68] and with multiple experts. We will conclude that our algorithm is more accurate than single experts if it achieves a higher Dice overlap with expert consensus than the Dice overlap between single experts and expert consensus [30, 84]. Here, the average Dice accuracy from leave-one-out cross validation will be used—training the algorithm on all but one subject and testing it on the left-out one to compute the algorithm-to-consensus Dice overlap of segmented lesion regions, and iterating until every patient has been left out once and only once.

Question 2: Can Z_{ADC} detect lesions in multi-site/scanner/protocol data? A fundamental problem is that the target patient may 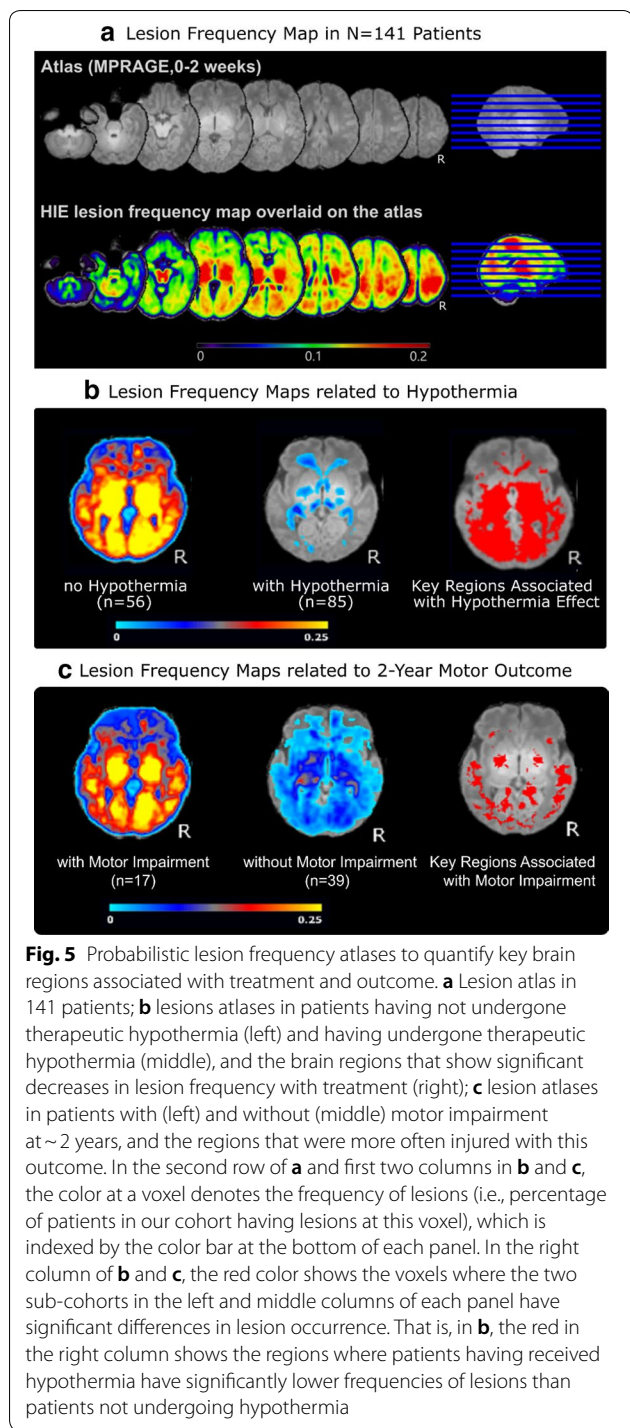have very different distribution of ADC and Z_{ADC} values than the training populations. We will design a self-adaptive mechanism to deal with differences of ADC maps acquired from different sites or scanners. The mechanism will first detect regions of abnormally ADC and Z_{ADC} values in the target patient using information learned from other patients. Then we will re-train the machine classifier on the ADC and Z_{ADC} maps from the target patient, using the tentatively-detected regions as training samples. The assumption is that, the knowledge of lesion voxel appearance as learned from training patients may not be completely suitable for a specific target patient, because of individual differences and sites/scanner/protocol differences. This is especially true for target image voxels that have probabilities of being lesioned just at the border line (e.g., those voxels that computer algorithm thought of having 49% or 51% percent of probabilities being lesioned). On the other hand, the target image voxels that computer algorithms assign very high probabilities of being lesioned (e.g., $>75\%$) are more reliable. These voxels can serve as "silver standard" to re-train the voxel-wise classifier, using the target image's features. Re-training using target image voxels for which the tentative results have high confidence to be lesioned or normative voxels can reduce bias in the classifier arising from training on other patients or other imaging protocols, as those "silver standard" voxels are from the same target patient and the same imaging protocol.

Part3.3. Machine learning of Z_{ADC} and ADC patterns for outcome prediction

Part3.3a. Lesion-based outcome prediction

Question 3: What are lesion patterns that are associated with neurocognitive outcomes at 2 years of age? The MRI scoring systems currently used in the hospital setting focus on injuries in certain key brain regions such as the thalamus, basal ganglia, internal capsules, etc. [19, 20, 85]. While the scores reflect the severity of HIE during infancy, the predictive power for outcomes by 2 years of age is not established.

We hypothesize that through machine learning we can find a specific combination of sub-regions with abnormal Z_{ADC} values that better inform outcomes compared to expert scoring systems. One rationale comes from our preliminary results in Fig. 5, which show that injury may involve brain structures not specifically assessed in the current expert scoring systems. In Fig. 5a, we generated probabilistic lesion atlases for HIE. This was based on using our extensively-validated [70] non-rigid registration tool [71] to map individual patients' lesions into a neonatal atlas [57], where the lesion loads among 141 patients were averaged at the voxel level. For example, a voxel the constructed probabilistic HIE lesion atlas having a value of 0.1 means that 10% of the patients in our cohort had HIE

Weiss *et al. J Transl Med*   (2019) 17:385

Page 8 of 16

Fig. 5 Probabilistic lesion frequency atlases to quantify key brain regions associated with treatment and outcome. **a** Lesion atlas in 141 patients; **b** lesions atlases in patients having not undergone therapeutic hypothermia (left) and having undergone therapeutic hypothermia (middle), and the brain regions that show significant decreases in lesion frequency with treatment (right); **c** lesion atlases in patients with (left) and without (middle) motor impairment at ~2 years, and the regions that were more often injured with this outcome. In the second row of **a** and first two columns in **b** and **c**, the color at a voxel denotes the frequency of lesions (i.e., percentage of patients in our cohort having lesions at this voxel), which is indexed by the color bar at the bottom of each panel. In the right column of **b** and **c**, the red color shows the voxels where the two sub-cohorts in the left and middle columns of each panel have significant differences in lesion occurrence. That is, in **b**, the red in the right column shows the regions where patients having received hypothermia have significantly lower frequencies of lesions than patients not undergoing hypothermia

lesions at the same anatomical location. We further used voxel-wise lesion symptom mapping (VLSM [86]) to statistically test whether lesion occurrence in each voxel was significantly associated with treatment (Fig. 5b) and with motor impairment at 2 years (Fig. 5c). The significance was defined as $p < 0.05$ after 10,000 permutations correcting for potential false positives in multiple comparisons

[86]. Panel b shows the lesion atlases in patients who were treated without therapeutic hypothermia (N = 56, mean age at scan = 4 days) and with therapeutic hypothermia (N = 85, mean age at scan = 4 days), and the regions of significant difference between these two cohorts. This quantified the regional reduction in lesion load and shows a larger impact of treatment in the deep gray and posterior structures. Panel c shows atlases in cohorts with (N = 17, mean age 4 days at scan) and without (N = 39, mean age 5 days at scan) clinically-documented motor impairment at 2 years, highlighting that key brain regions associated with adverse motor outcomes involve the anatomic regions mainly along the corticospinal tract (CST) and other cortical structures in occipital and parietal lobes that are not specifically scored in many expert scoring systems.

Our second motivation is that machine learning algorithms can capture subtle patterns that may not be scored in expert scoring systems. For example, two patients may both have injury in thalamus, but may have different outcomes due to slight differences in the volume and location of the injury, in the actual ADC and Z_{ADC} distributions within the injured regions (e.g., whether average $Z_{ADC} = -3$ or -6 in the detected lesion regions as stated under Question 1), in the slight differences in the bi-lateralization, etc. These subtle patterns are not in current expert scoring systems, but can be captured by computer algorithms. In addition, regions that inform outcomes

Table 4 A list features to be used for lesion-based outcome prediction

Categories	Details of features
I. Lesion anatomy	I.1. Mass center in standard neonatal atlas space
	I.2. Percentage of the whole-brain volume and the volume of each of the 61 auto-segmented brain structures being injured [76, 78, 122]
	I.3. Ratios of volumetric injury in the same brain structures between the left and right hemisphere
	I.4. Percentage and distribution of HIE lesions in 28 major fiber tracts as defined in the JHU atlas [123]
II. Lesion geometry	II.1. Lesion volume
	II.2. Maximum diameter along different orthogonal directions, maximum surface of lesion, lesion compactness, lesion sphericity, surface-to-volume ratio
III. Lesion heterogeneity	III.1. Histogram analysis (0, 25, 50, 75 and 100-percentile) of T1, T2, DWI, ADC, Z_{T1}, Z_{T2}, Z_{DWI}, Z_{ADC} signal values within the lesion regions
	III.2. Skewness (asymmetry), kurtosis (flatness), uniformity and randomness (entropy and standard deviations) of T1, T2, DWI, ADC, Z_{T1}, Z_{T2}, Z_{DWI}, Z_{ADC} signal values within the lesion regions
IV. Lesion texture	IV.1. gray-level co-occurrence matrix (GLCM) features and gray-level run-length matrix (GLRLM) of T1, T2, DWI, ADC, Z_{T1}, Z_{T2}, Z_{DWI}, Z_{ADC} signal values within lesion regions
	IV.2. fractal analysis, Minkowski functionals, wavelet transform and Laplacian transforms of Gaussian-filtered images for the lesion regions

Weiss *et al. J Transl Med*   (2019) 17:385

Page 9 of 16

may include both regions with $Z_{ADC} < -2$ (abnormally low ADC, possibly ischemic necrosis) as well as regions with $Z_{ADC} > 2$ (abnormally high ADC, possibly vasogenic edema), but the latter is not considered in expert scoring systems. Table 4 lists all the MRI features (predictive variables extracted from MRI) we propose to use for lesion-based outcome prediction.

Part3.3b. Lesion-free outcome prediction

Question 4: Can the magnitude and pattern of Z_{ADC} augment outcome prediction? We hypothesize that outcomes are associated with the magnitude and pattern of Z_{ADC} values. For example, a patient with $Z_{ADC} = -5$ in every voxel in a region may have a worse outcome than another patient with values ranging between -5 and -2 in the same region.

To test this hypothesis, we will include textures of Z_{ADC} in the multivariate predictive model. The texture will include the entropy, skewness, kurtosis, and histogram analysis (0, 25, 50, 75, 100-percentile) of ADC and Z_{ADC} values in the injured regions. Each patient will be represented as a high-dimensional feature vector. Classification models (e.g., support vector machine (SVM) [87] or random forest (RF) [88]) will be explored for their ability to predict categorical outcomes, and regression models (e.g., support or relevance vector regression (SVR, RVR) [89]) to predict the continuous outcomes [29].

To reduce the risk of over-fitting, our iterative forward inclusion and backward elimination (FIBE) feature selection algorithm will be used [71, 90] to select the most informative subset of features. The FIBE feature selection algorithm will start from the single most informative feature (one with the smallest prediction error in the training set), and iteratively add one feature into the subset at a time, such as adding this feature leads to the maximum decrease of prediction errors with Support Vector Regression compared to adding any other feature, until no features can be added that further reduces the prediction error. The algorithm will then exclude features from the subset, one at a time, such that removing this feature leads to the maximum decrease of prediction errors compared to removing any other feature, until no other features can be removed from the subset that further reduces prediction error. The algorithm iterates between forward inclusion and backward elimination until no feature can be added or remove. The final subset is the selected subset of features that leads to the minimum prediction error. The FIBE algorithm does not start from the full feature set. At any time, the subset only contains a small fraction of all features, therefore it reduces the risk for overfitting.

One merit of using texture features of ADC and Z_{ADC} maps is that we may be able to predict outcomes in ADC maps that are read as clinically normal. It is well known that 30–50% of patients affected by HIE do not show visually explicit lesions [24, 25], which may be due to mild injury, pseudo-normalization (ADC values returning to normal, hiding the lesions before the lesion actually resolves [52]), the use of TH, or other reasons. In these scenarios, voxels may not survive the threshold of Z_{ADC} map by -2 (since lesions are invisible), but subtle pattern abnormalities can still be captured in the texture of ADC and Z_{ADC} and these patterns may contribute to outcome prediction. Table 5 lists all the MRI features (predictive variables extracted from MRI) that we plan to use for lesion-free outcome prediction. No explicit lesion detection is needed.

For both Part3.3a and Part3.3b

Question 5: Generality to multi-site data? We will test the hypothesis that certain predictive variables that are more robust to multi-site data will achieve more stable prediction accuracy. For example, lesion volume might be a more stable variable than the boundary irregularity, and the standard deviation of ADC and Z_{ADC} values should be more stable than the actual mean or median values. We will rank variables and variable combinations by their predictive power in data across sites and encourage the auto-selection of stable variable combinations for multi-site generalization.

Table 5 A list features to be used for lesion-free outcome prediction

Categories	Details of features
I. Fiber tract features	I.1. Histogram analysis (0, 25, 50, 75 and 100-percentile) of T1, T2, DWI, ADC, Z_{T1}, Z_{T2}, Z_{DWI}, Z_{ADC} signal values within each of the 28 major fiber bundles as defined in the JHU atlas [123] I.2. Skewness (asymmetry), kurtosis (flatness), uniformity and randomness (entropy and standard deviations) of T1, T2, DWI, ADC, Z_{T1}, Z_{T2}, Z_{DWI}, Z_{ADC} signal values in each brain structures
II. Regional anatomy features	II.1. Histogram analysis (0, 25, 50, 75 and 100-percentile) of T1, T2, DWI, ADC, Z_{T1}, Z_{T2}, Z_{DWI}, Z_{ADC} signal values within the brain and each of the 61 auto-segmented brain structures/regions II.2. Skewness (asymmetry), kurtosis (flatness), uniformity and randomness (entropy) of T1, T2, DWI, ADC, Z_{T1}, Z_{T2}, Z_{DWI}, Z_{ADC} signal values in the brain and 61 auto-segmented regions II.3. Volume of the 61 auto-segmented structures/regions as measured in T1 image II.4. Left/right asymmetry in features II.1–II.3

Weiss *et al. J Transl Med*   (2019) 17:385

Page 10 of 16

Question 6. Can clinical variables augment MRI metrics and further improve outcome prediction? We will test this by combining the MRI metrics mentioned above with clinical variables, such as: EEG [91], 1- and 5-min APGAR scores [92], umbilical cord arterial pH value [93], length of stay in neonatal intensive care unit (NICU) [94], as listed in Part 1.3. The combination of MRI and clinical variables will increase the length of the patient-wise feature vector. However, the same feature selection (prediction and accuracy assessment) will still apply.

Accuracy evaluation For predicting binary outcomes, we will measure accuracy by sensitivity and specificity in leave-one-out cross validations. That is, we will divide N patients into a cohort of $(N-1)$ training patients and 1 testing patient. We learn the MRI signatures from known outcomes in the training patients, apply the learned model to predict outcome in the testing patient, and then check whether the predicted outcome for the testing patient is correct or false compared to the actual outcome of this testing patient. We iterate this process N times, such that every patient has been left out once and only once as the testing patient. Sensitivity is measured as the percentage of patients who had adverse outcomes has been correctly predicted, and specificity is measured as the percentage of patients who had normal outcomes has been correctly predicted. Similarly, accuracy for predicting continuously-valued outcomes (Bayley scores) will be measured via the root mean squared error (RMSE) between the predicted and actual scores in cross validations. That is, in a leave-one-out cross validations, ML learns the predictive model from $(N-1)$ training patients and predicts the Bayley score of a testing patient. We will repeat this N times such that each patient has been left out once and only once as the testing patient. The average RMSE between the predicted and the actual Bayley scores will be used to quantify prediction accuracy. A smaller RMSE means a higher accuracy in prediction.

We will evaluate the accuracies mentioned above for different strategies (lesion-based and lesion-free outcome prediction), for different classifiers during binary outcome prediction (for comparing the accuracies of different classifiers such as SVM, RF, etc.), and for different regressors in during continuous-valued outcome prediction (for comparing the accuracies among regressors such as SVR, RVR, etc.).

Comparison with expert scoring systems We will quantitatively compare our Z_{ADC}-based predictive model with expert scores, in terms of the accuracy in predicting outcomes in a k-fold or leave-one-out cross validation. Expert scores will be independently determined by 2 pediatric neuroradiologists, using the NICHD–NRN scoring criteria [19].

Expected sample size and power analysis

The expected sample size is 300 patients with a complete set of MRI and clinical variables, including outcomes. Approximately 440 cases of neonatal HIE cases with accompanying MRI scans have been identified from patients treated in MGH and BCH during 2009–2019 thus far, and our ongoing expert review of clinical records has found that roughly 50–60% of them had outcome data. MGH and BCH admit ~ 50–60 HIE patients annually.

When developing Z_{ADC}-based thresholding and machine learning to detect HIE lesions, we will quantify the Dice overlap, sensitivity and specificity of the detected lesion with regard to expert consensus in the leave-one-out manner (under Question 1). Recent machine learning driven HIE lesion detection has a median algorithm-and-expert Dice overlap at 0.52 and from 20 HIE patients (2017) [68]. We can loosely consider their mean Dice as 0.52 (actually not reported [68]). Given a desired power of 0.8, and alpha = 0.05, and assuming the standard deviation of Dice overlaps at 0.15 (which is the case in our pilot data [72]), we need 56 (or 13) patients to say our algorithm has achieved a significantly higher Dice accuracy if our mean Dice is 0.6 (or 0.69 as in our pilot data [72]).

When correlating each MRI metric with outcome scores. We need 46 or 71 subjects to test whether two variables (MRI and outcome) are significantly correlated with a power at 0.95 (beta = 0.05) and 0.995 (beta = 0.005) respectively (|PCC| > 0.5, p < 0.05, where PCC is Pearson's Correlation Coefficient).

In developing multi-variate prediction models, the one-in-ten-rule [95, 96] states a minimum risk of over-fitting if the algorithm selects, from the anticipated cohort of 300 patients, no more than 30 features that jointly predict outcomes. This is often the case in our similar studies and can be strictly enforced by our feature selection tool [35, 42, 43, 90].

Discussion

This protocol describes our plans to build on our Z_{ADC} measurement, and to develop machine learning tools to detect lesions and predict outcomes for neonatal HIE patients.

Secondary use of hospital-hosted data has received increasing attention [54–56]. Recent years have seen a harmonization of public informatics platforms that facilitate the mining of clinical databases for big data research [97]. Retrospective data collection is possible because of the now mature hospital data registries and the dissemination of clinical informatics tools. NICU registries record comprehensive clinical information on infant patients. Complementary to registry data, big data search engines allow us to query hundreds of thousands

Weiss *et al. J Transl Med*   (2019) 17:385

Page 11 of 16

of patients by ICD codes and keywords [98]. Tools for this purpose include i2b2 [99], SHRINE [100], HiGHmed [101], tranSMART [102], etc.; many of which are being adopted in healthcare settings through the world [63–65]. Publicly available platforms such as mi2b2 [66] and ChRIS [103, 104] permit the download of MRI data from Radiology archives with patient MRNs obtained from registry or hospital-wide searches [66]. Experts further filter cases for eligibility and quality control, and review clinical records to ascertain clinical information (e.g. entering outcomes into the REDCap database) [105]. The registries in hospital departments and the public availability of hospital-wide search tools foster data collection that is reproducible among hospitals.

A typical clinical trial often requires significant funding and years to collect data. In contrast, we plan to use existing clinical database. With 1–2 years of effort and limited resources we have identified 440 candidates (141 from MGH and 299 from BCH), the equivalent of 10+ years of active enrollment at the two hospitals. While not all of these candidate patients have a full spectrum of data, the process demonstrates the feasibility of the proposed study to aggregate the necessary cohort from our two hospitals. In contrast, the "Whole Brain Cooling" trial collected data from 208 HIE patients over a 3–4 year time frame (2000–2003) across 16 sites [5, 106–108], the "Optimizing Cooling" trial collected data from 364 HIE patients over 6–7 years (2010–2016) from 19 sites [22, 109–111], the "Late Hypothermia" trial collected data from 168 HIE patients over 8–9 years (2008–2016) from 22 sites [112], and the BABY BAC II trial is aiming to collect data for 160 HIE patients over 3–4 years (2017–2020) from 12 sites [113]. The time and resources saved in data collection makes our protocol a useful complement to existing clinical trials.

Of specific utility to HIE investigations, we recently developed the first-of-its-kind normative pediatric ADC atlases. This valuable resource allows for the quantification of the deviation from normal at the voxel level. We will develop machine learning tools to fully explore and test this new MRI measurement (voxel-wise Z_{ADC}) in lesion detection and outcome prediction. This will supplement the interpretation of neonatal brain MRIs, which is currently a subjective assessment performed by experts, by adding data that is quantitative, objective, consistent, anatomically-created and generalizable across multiple sites (Fig. 4). Once validated, the novel and machine-learning-powered tools that build on our new Z_{ADC} measurement will pave the way for future preclinical trials involving patients with HIE (Fig. 1).

Another novelty in the study is to use statistically-rigorous lesion-symptom mapping to quantify key neural substrate for treatment and outcomes (Fig. 5).

We do note that the results in Fig. 5 are preliminary (N = 141) and will be updated when more retrospective data is gathered (N = 300 as planned). The update is needed in at least two aspects. The first is to have a larger sample size ideally more balanced between sub-cohorts. Right now, N = 17 patients with and N = 39 without motor impairment (not big number, and imbalanced) had led to findings of the red regions (right column of Fig. 5c) not fully within known motor tracts. The second improvement when sample size is bigger is to purify the data. Neurocognitive outcomes at ~ 2 years are often multifactorial, including impairment in multiple sub-domains that exceed motor and include hearing, visual, memory, development delay, etc. We will start from sub-cohorts with the overall normative versus adverse outcomes. When it comes to sub-domains, simply stratifying them into with and without impairment in one specific sub-domain function may be contaminated but comorbidities. For example, Fig. 5c the red regions in the right column are not all within known motor tracts. We will likely need to use clustering approaches of outcomes to find major branches of multifactorial outcomes, or, to factor out (statistically control for) comorbidities in outcomes. Nevertheless, Fig. 5, especially panels (a) and (b), has shown promise of quantitative and rigorous analysis of lesions and lesion-outcome mapping at the voxel level, which may add knowledge to the current experience-based subjective scoring systems.

Special attention has been paid to the importance of fostering multi-site collaborations and generalizability. We will design specific image analysis algorithms to deal with differences in imaging data from multiple sites (Questions 2 and 5). One example of multi-site/scanner differences is in the diffusion parameters and the number of diffusion directions. An ideal number of diffusion directions is a topic under investigation and some suggested at least 45 directions to construct satisfactory fibers in high-angular-resolution diffusion-weighted imaging (HARDI) [114]. However, existing clinical diffusion MRI protocols in our hospitals used 24–32 diffusion directions. Similar settings are adopted in other hospitals or clinical database for HIE populations [115–117]. One reason is that the purpose is not to construct high angular resolution fibers, but to only create ADC and FA maps for clinical neuroradiology interpretation [24, 60]. For another reason, not having more directions in clinical settings is not to extend MRI scan time on neonates (non-sedation) [24, 60]. The purpose of our study is to retrospectively gather data that has been acquired clinically. So, we cannot change the clinical imaging protocol for neonates. Nevertheless, we plan to record the accuracies of lesion detection and outcome prediction as a function

Weiss *et al. J Transl Med* (2019) 17:385

Page 12 of 16

of different sites/scanners, different imaging parameters (b values, number of diffusion directions, etc.). This will provide new and quantitative evidence for future search of an optimal imaging protocol that balances between the clinical considerations (scan time especially for non-sedated neonates, sufficiency for neuroradiology reads of ADC/FA maps, etc.) and higher quality and accuracy of diffusion tensor reconstruction.

When it comes to multi-site data, harmonizing diffusion protocols or images has received increasing interest, especially for normative data [118, 119]. Our study is to retrospectively gather existing data from clinical databases from multiple sites. As the data is from patients, and lesions can appear at varying locations and sizes in the patient data, we used an alternative approach to deal with multi-site data differences. The approach is to design an adaptive lesion segmentation algorithm (see Question 2), which first learns the appearances of lesion voxels from other patients, and re-trains itself using the target patient's own image voxels that have been deem highly probable to be lesioned or normative. The re-training phase is directly on the target patient, not on training patients which may be scanned in a different site. We will test whether this improves the lesion detection accuracy for multi-site data.

Future larger-scale collaborations and pre-clinical trials across more sites will need to address the limitations of this current protocol. These limitations include: sample size, the need to further test multi-site compatibility, retrospective versus prospective suitability, dealing with variability in treatment guidelines and protocols, and standardization of outcome definitions across institution. In addition, we recognize the need for a secure and stable data warehouse, free release and dissemination of the planned algorithms and software tools; inclusion of additional MRI sequences (e.g., spectroscopy [120, 121]); the exploration and incorporation of other clinical [9–11], biochemical [9–12], and serum [12, 13] biomarkers; and so on. Nevertheless, the current protocol is a novel and needed approach that will provide a basis for larger-scale, multi-site studies.

In summary, this paper describes a registry- and informatics-driven clinical dataset collection protocol to power next-generation machine-learning-based MRI analytics for HIE lesion detection and outcome prediction. This study should benefit HIE clinical trials that incorporate brain MRI. The same technical framework can be used for data collection and biomarker development in other pediatric and adult cohorts, such as those with stroke, tumor and other non-brain disorders.

Abbreviations
MRI: Magnetic Resonance Imaging; HIE: hypoxic ischemic encephalopathy; TH: therapeutic hypothermia; NICU: neonatal intensive care unit; ADC: Apparent Diffusion Coefficient; MRN: medical record number; ICD: International Classification of Diseases; EHR: electronic health record; BSID-III: Bayley Scale of Infant Development, Version III; NICHD–NRN: National Institute of Children Health and Human Development, Neonatal Research Network; BGT: basal ganglia thalamus; ALIC: anterior limb of internal capsule; PLIC: posterior limb of internal capsule; WS: watershed; Stdev: standard deviation; ML: machine learning; ROC: receiver-operating-curve; VLSM: voxel-wise lesion symptom mapping; CST: corticospinal tract; FIBE: forward inclusion and backward elimination; RMSE: root mean squared error.

Authors' contributions
RJW contributed to the study design, acquisition of clinical data and manuscript preparation. SVB contributed to the study design, provided expert clinical data interpretation, supervised acquisition and curation of clinical data, and assisted in manuscript preparation. YS, YG, MG, IC, LZ contributed to data analysis and interpretation. EH and YC assisted with clinical data acquisition and interpretation. SNM created and implemented data acquisition software tools. RLG provided bioinformatics expertise, supervised acquisition and analysis of data, and provided extensive edits to the manuscript. PEG provided neonatal neuroimaging expertise, assisted with the implementation of study concept and design, supervised data interpretation, and provided edits to the manuscript. YO developed standardized protocols for acquiring and analyzing the imaging data, supervised data interpretation, and contributed to the study design and manuscript preparation. All authors have reviewed version of the manuscript. All authors read and approved the final manuscript.

Availability of data and materials
The data that support the findings of this study are not yet publicly available, as these methods are continuing to be validated. This group is open to collaboration for future analyses on request from the corresponding authors [YO, PEG].

Ethics approval and consent to participate
The Institutional Review Board of Partners Healthcare has approved this study and a wavier for individual consent to participate.

Consent for publication
Not applicable.

Weiss *et al. J Transl Med* (2019) 17:385

Page 13 of 16

Author details
[1] Division of Newborn Medicine, Department of Pediatrics, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA. [2] Fetal Neonatal Neuroimaging and Developmental Science Center (FNNDSC), Boston Children's Hospital, Harvard Medical School, 401 Park Drive, Landmark Center 7022, Boston, MA 02115, USA. [3] Computer Science & Artificial Intelligence Lab (CSAIL), Massachusetts Institute of Technology, Cambridge, MA 02139, USA. [4] Laboratory of Computer Science, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA. [5] Department of Psychiatry and Department of Radiology, Massachusetts General Hospital, Harvard Medical School, Boston, MA 02114, USA. [6] Neuroradiology Division, Department of Radiology, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA. [7] Computational Health Informatics Program (CHIP), Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA.

References
1. Finer NN, Robertson CM, Richards RT, Pinnell LE, Peters KL. Hypoxic-ischemic encephalopathy in term neonates: perinatal factors and outcome. J Pediatr. 1981;98:112–7.
2. Douglas-Escobar M, Weiss MD. Hypoxic-ischemic encephalopathy: a review for the clinician. JAMA Pediatr. 2015;169:397–403.
3. Azzopardi DV, Strohm B, Edwards AD, Dyet L, Halliday HL, Juszczak E, Kapellou O, Levene M, Marlow N, Porter E. Moderate hypothermia to treat perinatal asphyxial encephalopathy. N Engl J Med. 2009;361:1349–58.
4. Edwards AD, Brocklehurst P, Gunn AJ, Halliday H, Juszczak E, Levene M, Strohm B, Thoresen M, Whitelaw A, Azzopardi D. Neurological outcomes at 18 months of age after moderate hypothermia for perinatal hypoxic ischaemic encephalopathy: synthesis and meta-analysis of trial data. BMJ. 2010;340:c363–c363.
5. Shankaran S, Laptook AR, Ehrenkranz RA, Tyson JE, McDonald SA, Donovan EF, Fanaroff AA, Poole WK, Wright LL, Higgins RD. Whole-body hypothermia for neonates with hypoxic–ischemic encephalopathy. N Engl J Med. 2005;353:1574–84.
6. Vexler ZS, Ferriero DM. Molecular and biochemical mechanisms of perinatal brain injury. Seminars in Neonatology. 2001;6:99–108.
7. Ferriero DM. Neonatal brain injury. N Engl J Med. 2004;351:1985–95.
8. Perez A, Ritter S, Brotschi B, Werner H, Caflisch J, Martin E, Latal B. Long-term neurodevelopmental outcome with hypoxic-ischemic encephalopathy. J Pediatr. 2013;163:454–459.e1.
9. Ramaswamy V, Horton J, Vandermeer B, Buscemi N, Miller S, Yager J. Systematic review of biomarkers of brain injury in term neonatal encephalopathy. Pediatr Neurol. 2009;40:215–26.
10. van Laerhoven H, de Haan TR, Offringa M, Post B, van der Lee JH. Prognostic tests in term neonates with hypoxic-ischemic encephalopathy: a systematic review. Pediatrics. 2013;131:88–98.
11. Natarajan G, Pappas A, Shankaran S. Outcomes in childhood following therapeutic hypothermia for neonatal hypoxic-ischemic encephalopathy (HIE). Semin Perinatol. 2016;40:549–55.
12. Massaro AN, Jeromin A, Kadom N, Vezina G, Hayes RL, Wang KKW, Streeter J, Johnston MV. Serum biomarkers of MRI brain injury in neonatal hypoxic ischemic encephalopathy treated with whole-body hypothermia: a pilot study. Pediatr Crit Care Med. 2013;14:310–7.
13. Douglas-Escobar M, Weiss MD. Biomarkers of hypoxic-ischemic encephalopathy in newborns. Front Neurol. 2012;3:144.
14. Azzopardi D, David Edwards A. Magnetic resonance biomarkers of neuroprotective effects in infants with hypoxic ischemic encephalopathy. Semin Fetal Neonatal Med. 2010;15:261–9.
15. Goergen SK, Ang H, Wong F, Carse EA, Charlton M, Evans R, Whiteley G, Clark J, Shipp D, Jolley D. Early MRI in term infants with perinatal hypoxic–ischaemic brain injury: interobserver agreement and MRI predictors of outcome at 2 years. Clin Radiol. 2014;69:72–81.
16. Khong PL, Lam BCC, Tung HKS, Wong V, Chan FL, Ooi GC. MRI of neonatal encephalopathy. Clin Radiol. 2003;58:833–44.
17. ClinicalTrials.gov Study Database. http://clinicaltrials.gov. Accessed 10 Oct 2019.
18. Rutherford M, Malamateniou C, McGuinness A, Allsop J, Biarge MM, Counsell S. Magnetic resonance imaging in hypoxic-ischaemic encephalopathy. Early Human Dev. 2010;86:351–60.
19. Shankaran S, Barnes PD, Hintz SR, Laptook AR, Zaterka-Baxter KM, McDonald SA, Ehrenkranz RA, Walsh MC, Tyson JE, Donovan EF, Goldberg RN, Bara R, Das A, Finer NN, Sanchez PJ, Poindexter BB, Van Meurs KP, Carlo WA, Stoll BJ, Duara S, Guillet R, Higgins RD. Eunice kennedy shriver National Institute of Child Health and Human development neonatal research network. Brain injury following trial of hypothermia for neonatal hypoxic–ischaemic encephalopathy. Arch Dis Child Fetal Neonatal. 2012;97:F398–404.
20. Weeke LC, Groenendaal F, Mudigonda K, Blennow M, Lequin MH, Meiners LC, van Haastert IC, Benders MJ, Hallberg B, de Vries LS. A novel magnetic resonance imaging score predicts neurodevelopmental outcome after perinatal asphyxia and therapeutic hypothermia. J Pediatr. 2018;192:33–40.e2.
21. Barkovich AJ. The encephalopathic neonate: choosing the proper imaging technique. Am J Neuroradiol. 1997;18:1816–20.
22. Shankaran S, Laptook AR, Pappas A, McDonald SA, Das A, Tyson JE, Poindexter BB, Schibler K, Bell EF, Heyne RJ, Pedroza C, Bara R, Van Meurs KP, Huitema CMP, Grisby C, Devaskar U, Ehrenkranz RA, Harmon HM, Chalak LF, DeMauro SB, Garg M, Hartley-McAndrew ME, Khan AM, Walsh MC, Ambalavanan N, Brumbaugh JE, Watterberg KL, Shepherd EG, Hamrick SEG, Barks J, Cotten CM, Kilbride HW, Higgins RD. Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal Research Network. Effect of depth and duration of cooling on death or disability at age 18 months among neonates with hypoxic-ischemic encephalopathy: a randomized clinical trial. JAMA. 2017;318:57–67.
23. Trivedi SB, Vesoulis ZA, Rao R, Liao SM, Shimony JS, McKinstry RC, Mathur AM. A validated clinical MRI injury scoring system in neonatal hypoxic-ischemic encephalopathy. Pediatr Radiol. 2017;47:1491–9.
24. Groenendaal F, de Vries LS. Fifty years of brain imaging in neonatal encephalopathy following perinatal asphyxia. Pediatr Res. 2017;81:150–5.
25. Hayakawa K, Koshino S, Tanda K, Nishimura A, Sato O, Morishita H, Ito T. Diffusion pseudonormalization and clinical outcome in term neonates with hypoxic-ischemic encephalopathy. Pediatr Radiol. 2018;48:865–74.
26. Conway JM, Walsh BH, Boylan GB, Murray DM. Mild hypoxic ischaemic encephalopathy and long term neurodevelopmental outcome—a systematic review. Early Hum Dev. 2018;120:80–7.
27. Aerts HJ, Velazquez ER, Leijenaar RT, Parmar C, Grossmann P, Carvalho S, Bussink J, Monshouwer R, Haibe-Kains B, Rietveld D. Decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. Nat Commun. 2014;5:4006.
28. Lambin P, Rios-Velazquez E, Leijenaar R, Carvalho S, van Stiphout RGPM, Granton P, Zegers CML, Gillies R, Boellard R, Dekker A, Aerts HJWL. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012;48:441–6.
29. Zhou M, Scott J, Chaudhury B, Hall L, Goldgof D, Yeom KW, Iv M, Ou Y, Kalpathy-Cramer J, Napel S, Gillies R, Gevaert O, Gatenby R. Radiomics in brain tumor: image assessment, quantitative feature descriptors, and machine-learning approaches. Am J Neuroradiol. 2017. https://doi.org/10.3174/ajnr.A5391.
30. Verma R, Zacharaki EI, Ou Y, Cai H, Chawla S, Lee S-K, Melhem ER, Wolf R, Davatzikos C. Multiparametric tissue characterization of brain neoplasms and their recurrence using pattern classification of MR images. Acad Radiol. 2008;15:966–77.
31. Macyszyn L, Akbari H, Pisapia JM, Da X, Attiah M, Pigrish V, Bi Y, Pal S, Davuluri RV, Roccograndi L, Dahmane N, Martinez-Lage M, Biros G, Wolf RL, Bilello M, O'Rourke DM, Davatzikos C. Imaging patterns predict patient survival and molecular subtype in glioblastoma via machine learning techniques. Neuro-oncology. 2016;18:417–25.
32. Chang K, Zhang B, Guo X, Zong M, Rahman R, Sanchez D, Winder N, Reardon DA, Zhao B, Wen PY, Huang RY. Multimodal imaging patterns predict survival in recurrent glioblastoma patients treated with bevacizumab. Neuro-oncology. 2016;18:1680–7.

Weiss *et al. J Transl Med*   (2019) 17:385

Page 14 of 16

33. Moradi E, Pepe A, Gaser C, Huttunen H, Tohka J, Initiative ADN. Machine learning framework for early MRI-based Alzheimer's conversion prediction in MCI subjects. Neuroimage. 2015;104:398–412.

34. Gaser C, Franke K, Klöppel S, Koutsouleris N, Sauer H, Initiative ADN. BrainAGE in mild cognitive impaired patients: predicting the conversion to Alzheimer's disease. PLoS ONE. 2013;8:e67346.

35. Da X, Toledo JB, Zee J, Wolk DA, Xie SX, Ou Y, Shacklett A, Parmpi P, Shaw L, Trojanowski JQ. Integration and relative value of biomarkers for prediction of MCI to AD progression: spatial patterns of brain atrophy, cognitive scores, APOE genotype and CSF biomarkers. NeuroImage Clin. 2014;4:164–73.

36. Andrade de Oliveira A, Carthery-Goulart MT, Oliveira Júnior PP, Carrettiero DC, Sato JR. Defining multivariate normative rules for healthy aging using neuroimaging and machine learning: an application to Alzheimer's disease. J Alzheimers Dis. 2015;43:201–12.

37. Vu M-AT, Adalı T, Ba D, Buzsáki G, Carlson D, Heller K, Liston C, Rudin C, Sohal VS, Widge AS, Mayberg HS, Sapiro G, Dzirasa K. A shared vision for machine learning in neuroscience. J Neurosci. 2018;38:1601–7.

38. Akkus Z, Galimzianova A, Hoogi A, Rubin DL, Erickson BJ. Deep learning for brain MRI segmentation: state of the art and future directions. J Digit Imaging. 2017;30:449–59.

39. Kamal H, Lopez V, Sheth SA. Machine learning in acute ischemic stroke neuroimaging. Front Neurol. 2018;9:945.

40. Moghim N, Corne DW. Predicting epileptic seizures in advance. PLoS ONE. 2014;9:e99334.

41. Bleich-Cohen M, Jamshy S, Sharon H, Weizman R, Intrator N, Poyurovsky M, Hendler T. Machine learning fMRI classifier delineates subgroups of schizophrenia patients. Schizophr Res. 2014;160:196–200.

42. Serpa MH, Ou Y, Schaufelberger MS, Doshi J, Ferreira LK, Machado-Vieira R, Menezes PR, Scazufca M, Davatzikos C, Busatto GF. Neuroanatomical classification in a population-based sample of psychotic major depression and bipolar I disorder with 1 year of diagnostic stability. BioMed Res Int. 2014;2014:706157.

43. Zanetti MV, Schaufelberger MS, Doshi J, Ou Y, Ferreira LK, Menezes PR, Scazufca M, Davatzikos C, Busatto GF. Neuroanatomical pattern classification in a population-based sample of first-episode schizophrenia. Prog Neuropsychopharmacol Biol Psychiatry. 2013;43:116–25.

44. Ledig C, Heckemann RA, Hammers A, Lopez JC, Newcombe VFJ, Makropoulos A, Lötjönen J, Menon DK, Rueckert D. Robust whole-brain segmentation: application to traumatic brain injury. Med Image Anal. 2015;21:40–58.

45. Benson RR, Gattu R, Sewick B, Kou Z, Zakariah N, Cavanaugh JM, Haacke EM. Detection of hemorrhagic and axonal pathology in mild traumatic brain injury using advanced MRI: implications for neurorehabilitation. NeuroRehabil Interdiscip J. 2012;31:261.

46. Spiteri M, Guillemaut J-Y, Windridge D, Avula S, Kumar R, Lewis E. Fully-automated identification of imaging biomarkers for post-operative cerebellar mutism syndrome using longitudinal paediatric MRI. Neuroinform. 2019. https://doi.org/10.1007/s12021-019-09427-w.

47. Menze B, Jakab A, Bauer S, Kalpathy-Cramer J, Farahani K, Kirby J, Burren Y, Porz N, Slotboom J, Wiest R. The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans Med Imaging. 2014;33:1993–2024.

48. Bakas S, Akbari H, Sotiras A, Bilello M, Rozycki M, Kirby JS, Freymann JB, Farahani K, Davatzikos C. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. Scientific Data. 2017;4:170117.

49. Jack CR, Bernstein MA, Fox NC, Thompson P, Alexander G, Harvey D, Borowski B, Britson PJ, Whitwell LJ, Ward C. The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods. J Magn Reson Imaging. 2008;27:685–91.

50. Mueller SG, Weiner MW, Thal LJ, Petersen RC, Jack CR, Jagust W, Trojanowski JQ, Toga AW, Beckett L. Ways toward an early diagnosis in Alzheimer's disease: the Alzheimer's Disease Neuroimaging Initiative (ADNI). Alzheimer's Dementia. 2005;1:55–66.

51. Liauw L, van Wezel-Meijler G, Veen S, van Buchem MA, van der Grond J. Do apparent diffusion coefficient measurements predict outcome in children with neonatal hypoxic-ischemic encephalopathy? AJNR Am J Neuroradiol. 2009;30:264–70.

52. Winter JD, Lee DS, Hung RM, Levin SD, Rogers JM, Thompson RT, Gelman N. Apparent diffusion coefficient pseudonormalization

53. Wolf RL, Zimmerman RA, Clancy R, Haselgrove JH. Quantitative apparent diffusion coefficient measurements in term neonates for early detection of hypoxic-ischemic brain injury: initial experience 1. Radiology. 2001;218:825–33.

54. Jensen PB, Jensen LJ, Brunak S. Mining electronic health records: towards better research applications and clinical care. Nat Rev Genet. 2012;13:395–405.

55. Martin-Sanchez FJ, Aguiar-Pulido V, Lopez-Campos GH, Peek N, Sacchi L. Secondary use and analysis of big data collected for patient care. Yearb Med Inform. 2017;26:28–37.

56. Raja K, Patrick M, Gao Y, Madu D, Yang Y, Tsoi LC. A review of recent advancement in integrating omics data with literature mining towards biomedical discoveries. Int J Genomics. 2017;2017:6213474.

57. Ou Y, Zollei L, Retzepi K, Victor C, Bates S, Pieper S, Andriole K, Murphy SN, Gollub RL, Grant PE. Using clinically-acquired MRI to construct age-specific ADC atlases: quantifying spatiotemporal ADC changes from birth to 6 years old. Hum Brain Mapp. 2017;38:3052–68.

58. Le Bihan D, Mangin JF, Poupon C, Clark CA, Pappata S, Molko N, Chabriat H. Diffusion tensor imaging: concepts and applications. J Magn Reson Imaging. 2001;13:534–46.

59. Beaulieu C. The basis of anisotropic water diffusion in the nervous system—a technical review. NMR Biomed. 2002;15:435–55.

60. Rutherford M, Srinivasan L, Dyet L, Ward P, Allsop J, Counsell S, Cowan F. Magnetic resonance imaging in perinatal brain injury: clinical presentation, lesions and outcome. Pediatr Radiol. 2006;36:582–92.

61. Ozturk A, Sasson AD, Farrell JAD, Landman BA, da Motta A, Aralasmak A, Yousem DM. Regional differences in diffusion tensor imaging measurements: assessment of intrarater and interrater variability. Am J Neuroradiol. 2008;29:1124–7.

62. Pinto ALR, Ou Y, Sahin M, Grant PE. Quantitative apparent diffusion coefficient mapping may predict seizure onset in children with Sturge-Weber syndrome. Pediatr Neurol. 2018;84:32–8.

63. Zapletal E, Rodon N, Grabar N, Degoulet P. Methodology of integration of a clinical data warehouse with a clinical information system: the HEGP case. Stud Health Technol Inform. 2010;160:193–7.

64. Prasser F, Kohlbacher O, Mansmann U, Bauer B, Kuhn KA. Data integration for future medicine (DIFUTURE). Methods Inf Med. 2018;57:e57–65.

65. Wagholikar KB, Mendis M, Dessai P, Sanz J, Law S, Gilson M, Sanders S, Vangala M, Bell DS, Murphy SN. Automating installation of the integrating biology and the bedside (i2b2) platform. Biomed Inform Insights. 2018;10:1178222618777749.

66. Murphy SN, Herrick C, Wang Y, Wang TD, Sack D, Andriole KP, Wei J, Reynolds N, Plesniak W, Rosen BR, Pieper S, Gollub RL. High through-put tools to access images from clinical archives for research. J Digit Imaging. 2014;28:1–11.

67. Patridge EF, Bardyn TP. Research electronic data capture (REDCap). JMLA. 2018;106:142.

68. Murphy K, Aa NE, Negro S, Groenendaal F, Vries LS, Viergever MA, Boylan GB, Benders MJ, Išgum I. Automatic quantification of ischemic injury on diffusion-weighted MRI of neonatal hypoxic ischemic encephalopathy. NeuroImage Clin. 2017;14:222–32.

69. Warfield SK, Zou KH, Wells WM. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Trans Med Imaging. 2004;23:903–21.

70. Ou Y, Akbari H, Bilello M, Da X, Davatzikos C. Comparative evaluation of registration algorithms in different brain databases with varying difficulty: results and insights. IEEE Trans Med Imaging. 2014;33:2039–65.

71. Ou Y, Sotiras A, Paragios N, Davatzikos C. DRAMMS: deformable registration via attribute matching and mutual-saliency weighting. Med Image Anal. 2011;15:622–39.

72. Ou Y, Gollub RL, Wang J, Fan Q, Bates S, Chou J, Weiss R, Retzepis K, Pieper S, Jaimes C, Murphy S, Zollei L, Grant PE. MRI detection of neonatal hypoxic ischemic encephalopathy: machine v.s. Radiologists. Organization for Human Brain Mapping (OHBM); 2017. https://archive.aievolution.com/2017/hbm1701/index.cfm?do=abs.viewAbs&abs=4200. Accessed 10 July 2019.

time in neonatal hypoxic-ischemic encephalopathy. Pediatr Neurol. 2007;37:255–62.

Weiss *et al. J Transl Med*   (2019) 17:385

Page 15 of 16

73. Song Y, Bates SV, Gollub RL, Weiss RJ, He S, Cobos CJ, Sotardi S, Zhang Y, Liu T, Grant PE, Ou Y. Probabilistic atlases of neonatal hypoxic ischemic injury. In: Proceedings at pediatric academic society (PAS), 24–30 April 2019; Baltimore.

74. Tustison NJ, Avants BB, Cook PA, Zheng Y, Egan A, Yushkevich PA, Gee JC. N4ITK: improved N3 bias correction. IEEE Trans Med Imaging. 2010;29:1310–20.

75. Ou Y, Zollei L, Da X, Retzepi K, Murphy SN, Gerstner ER, Rosen BR, Grant PE, Kalpathy-Cramer J, Gollub RL. Field of view normalization in multi-site brain MRI. Neuroinformatics. 2018;16:431–44.

76. Doshi J, Erus G, Ou Y, Gaonkar B, Davatzikos C. Multi-atlas skull-stripping. Acad Radiol. 2013;20:1566–76.

77. Ou Y, Gollub RL, Retzepi K, Reynold NA, Pienaar R, Murphy SN, Grant PE, Zöllei L. Brain extraction in pediatric ADC maps, toward characterizing neuro-development in multi-platform and multi-institution clinical images. NeuroImage. 2015;122:246–61.

78. Doshi J, Erus G, Ou Y, Resnick SM, Gur RC, Gur RE, Satterthwaite TD, Furth S, Davatzikos C, Initiative AN, et al. MUSE: MUlti-atlas region Segmentation utilizing Ensembles of registration algorithms and parameters, and locally optimal atlas selection. NeuroImage. 2015;127:186–95.

79. Pereira S, Pinto A, Alves V, Silva CA. Brain tumor segmentation using convolutional neural networks in MRI images. IEEE Trans Med Imaging. 2016;35:1240–51.

80. Havaei M, Davy A, Warde-Farley D, Biard A, Courville A, Bengio Y, Pal C, Jodoin P-M, Larochelle H. Brain tumor segmentation with Deep Neural Networks. Med Image Anal. 2017;35:18–31.

81. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. In: Navab N, Hornegger J, Wells WM, Frangi AF, editors. Medical image computing and computer-assisted intervention—MICCAI, vol. 9351. Berlin: Springer International Publishing; 2015. p. 234–41.

82. Milletari F, Navab N, Ahmadi SA. V-net: fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth international conference on 3D vision (3DV). IEEE; 2016. pp. 565–571.

83. Menze BH, Van Leemput K, Lashkari D, Riklin-Raviv T, Geremia E, Alberts E, Gruber P, Wegener S, Weber M-A, Szekely G, Ayache N, Golland P. A generative probabilistic model and discriminative extensions for brain lesion segmentation-with application to tumor and stroke. IEEE Trans Med Imaging. 2016;35:933–46.

84. Cai H, Verma R, Ou Y, Lee S, Melhem ER, Davatzikos C. Probabilistic segmentation of brain tumors based on multi-modality magnetic resonance images. In: 4th IEEE international symposium on biomedical imaging: from nano to macro, 2007. ISBI. New York: IEEE; 2007. p. 600–603.

85. Barkovich AJ, Hajnal BL, Vigneron D, Sola A, Partridge JC, Allen F, Ferriero DM. Prediction of neuromotor outcome in perinatal asphyxia: evaluation of MR scoring systems. AJNR Am J Neuroradiol. 1998;19:143–9.

86. Bates E, Wilson SM, Saygin AP, Dick F, Sereno MI, Knight RT, Dronkers NF. Voxel-based lesion–symptom mapping. Nat Neurosci. 2003;6(5):448–50.

87. Scholkopf B, Smola AJ. Learning with kernels: support vector machines, regularization, optimization, and beyond. Cambridge: MIT press; 2001.

88. Liaw A, Wiener M. Classification and regression by randomForest, vol. 23. Winston-Salem: Forest; 2001.

89. Tipping ME. Sparse Bayesian learning and the relevance vector machine. J Mach Learn Res. 2001;1:211–44.

90. Ou Y, Shen D, Zeng J, Sun L, Moul J, Davatzikos C. Sampling the spatial patterns of cancer: optimized biopsy procedures for estimating prostate cancer volume and Gleason Score. Med Image Anal. 2009;13:609–20.

91. Murray DM, Boylan GB, Ryan CA, Connolly S. Early EEG findings in hypoxic-ischemic encephalopathy predict outcomes at 2 years. Pediatrics. 2009;124:e459–67.

92. Eun S, Lee JM, Yi DY, Lee NM, Kim H, Yun SW, Lim I, Choi ES, Chae SA. Assessment of the association between Apgar scores and seizures in infants less than 1 year old. Seizure. 2016;37:48–54.

93. Yeh P, Emary K, Impey L. The relationship between umbilical cord arterial pH and serious adverse neonatal outcome: analysis of 51 519 consecutive validated samples. BJOG Int J Obstet Gynaecol. 2012;119:824–31.

94. Simbruner G, Mittal RA, Rohlmann F, Muche R. Systemic hypothermia after neonatal encephalopathy: outcomes of neo. nEURO. network RCT. Pediatrics. 2010;126:e771–8.

95. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med. 1996;15:361–87.

96. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol. 1996;49:1373–9.

97. Canuel V, Rance B, Avillach P, Degoulet P, Burgun A. Translational research platforms integrating clinical and omics data: a review of publicly available solutions. Brief Bioinform. 2015;16:280–90.

98. Nalichowski R, Keogh D, Chueh HC, Murphy SN. Calculating the benefits of a research patient data repository. In: AMIA annual symposium proceedings; 2006. p. 1044.

99. Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, Kohane I. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). J Am Med Inform Assoc. 2010;17:124–30.

100. Weber GM, Murphy SN, McMurry AJ, Macfadden D, Nigrin DJ, Churchill S, Kohane IS. The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories. J Am Med Inform Assoc. 2009;16:624–30.

101. Haarbrandt B, Schreiweis B, Rey S, Sax U, Scheithauer S, Rienhoff O, Knaup-Gregori P, Bavendiek U, Dieterich C, Brors B, Kraus I, Thoms CM, Jäger D, Ellenrieder V, Bergh B, Yahyapour R, Eils R, Consortium, H, Marschollek M. HiGHmed—an open platform approach to enhance care and research across institutional boundaries. Methods Inf Med. 2018;57:e66–81.

102. Szalma S, Koka V, Khasanova T, Perakslis ED. Effective knowledge management in translational medicine. J Transl Med. 2010;8:68.

103. Pienaar R, Rannou N, Bernal J, Hahn D, Grant PE. ChRIS—a web-based neuroimaging and informatics system for collecting, organizing, processing, visualizing and sharing of medical data. Conf Proc IEEE Eng Med Biol Soc. 2015;2015:206–9.

104. Bernal-Rusiel JL, Rannou N, Gollub RL, Pieper S, Murphy S, Robertson R, Grant PE, Pienaar R. Reusable client-side javascript modules for immersive web-based real-time collaborative neuroimage visualization. Front Neuroinform. 2017;11:32.

105. Harris PA, Taylor R, Thielke R, Payne J, Gonzalez N, Conde JG. Research electronic data capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support. J Biomed Inform. 2009;42:377–81.

106. Shankaran S, Laptook A, Wright LL, Ehrenkranz RA, Donovan EF, Fanaroff AA, Stark AR, Tyson JE, Poole K, Carlo WA, Lemons JA, Oh W, Stoll BJ, Papile L-A, Bauer CR, Stevenson DK, Korones SB, McDonald S. Whole-body hypothermia for neonatal encephalopathy: animal observations as a basis for a randomized, controlled pilot study in term infants. Pediatrics. 2002;110:377–85.

107. Laptook A, Tyson J, Shankaran S, McDonald S, Ehrenkranz R, Fanaroff A, Donovan E, Goldberg R, O'Shea TM, Higgins RD, Poole WK, National Institute of Child Health and Human Development Neonatal Research Network. Elevated temperature after hypoxic-ischemic encephalopathy: risk factor for adverse outcomes. Pediatrics. 2008;122:491–9.

108. Laptook AR, Shankaran S, Ambalavanan N, Carlo WA, McDonald SA, Higgins RD, Das A, Hypothermia Subcommittee of the NICHD Neonatal Research Network. Outcome of term infants using apgar scores at 10 minutes following hypoxic-ischemic encephalopathy. Pediatrics. 2009;124:1619–26.

109. Pedroza C, Tyson JE, Das A, Laptook A, Bell EF, Shankaran S, Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal Research Network. Advantages of Bayesian monitoring methods in deciding whether and when to stop a clinical trial: an example of a neonatal cooling trial. Trials. 2016;17:335.

110. Shankaran S, Laptook AR, Pappas A, McDonald SA, Das A, Tyson JE, Poindexter BB, Schibler K, Bell EF, Heyne RJ, Pedroza C, Bara R, Van Meurs KP, Grisby C, Huitema CMP, Garg M, Ehrenkranz RA, Shepherd EG, Chalak LF, Hamrick SEG, Khan AM, Reynolds AM, Laughon MM, Truog WE, Dysart KC, Carlo WA, Walsh MC, Watterberg KL, Higgins RD, Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal Research Network. Effect of depth and duration of cooling on

Weiss *et al. J Transl Med*   (2019) 17:385

Page 16 of 16

deaths in the NICU among neonates with hypoxic ischemic encephalopathy: a randomized clinical trial. JAMA. 2014;312:2629–39.

111. Shankaran S. Outcomes of hypoxic-ischemic encephalopathy in neonates treated with hypothermia. Clin Perinatol. 2014;41:149–59.

112. Laptook AR, Shankaran S, Tyson JE, Munoz B, Bell EF, Goldberg RN, Parikh NA, Ambalavanan N, Pedroza C, Pappas A, Das A, Chaudhary AS, Ehrenkranz RA, Hensman AM, Van Meurs KP, Chalak LF, Khan AM, Hamrick SEG, Sokol GM, Walsh MC, Poindexter BB, Faix RG, Watterberg KL, Frantz ID, Guillet R, Devaskar U, Truog WE, Chock VY, Wyckoff MH, McGowan EC, Carlton DP, Harmon HM, Brumbaugh JE, Cotten CM, Sánchez PJ, Hibbs AM, Higgins RD, for the Eunice Kennedy Shriver National Institute of Child Health and Human Development Neonatal Research Network. Effect of therapeutic hypothermia initiated after 6 hours of age on death or disability among newborns with hypoxic-ischemic encephalopathy: a randomized clinical trial. JAMA. 2017;318:1550.

113. Cotten CM, Murtha AP, Goldberg RN, Grotegut CA, Smith PB, Goldstein RF, Fisher KA, Gustafson KE, Waters-Pick B, Swamy GK, Rattray B, Tan S, Kurtzberg J. Feasibility of autologous cord blood cells for infants with hypoxic-ischemic encephalopathy. J Pediatr. 2014;164:973–979.e1.

114. Tournier J-D, Calamante F, Connelly A. Determination of the appropriate b value and number of gradient directions for high-angular-resolution diffusion-weighted imaging. NMR Biomed. 2013;26:1775–86.

115. Lock C, Kwok J, Kumar S, Ahmad-Annuar A, Narayanan V, Ng AS, Tan YJ, Kandiah N, Tan EK, Czosnyka Z, Czosnyka M, Pickard JD, Keong NC. DTI profiles for rapid description of cohorts at the clinical-research interface. Front Med (Lausanne). 2018;5:357.

116. Zhang K, Johnson B, Pennell D, Ray W, Sebastianelli W, Slobounov S. Are functional deficits in concussed individuals consistent with white matter structural alterations: combined FMRI & DTI study. Exp Brain Res. 2010;204:57–70.

117. Holdsworth SJ, Aksoy M, Newbould RD, Yeom K, Van AT, Ooi MB, Barnes PD, Bammer R, Skare S. Diffusion tensor imaging (DTI) with retrospective motion correction for large-scale pediatric imaging. J Magn Reson Imaging. 2012;36:961–71.

118. Fortin J-P, Parker D, Tunç B, Watanabe T, Elliott MA, Ruparel K, Roalf DR, Satterthwaite TD, Gur RC, Gur RE, Schultz RT, Verma R, Shinohara RT.

119. Pohl KM, Sullivan EV, Rohlfing T, Chu W, Kwon D, Nichols BN, Zhang Y, Brown SA, Tapert SF, Cummins K, Thompson WK, Brumback T, Colrain IM, Baker FC, Prouty D, De Bellis MD, Voyvodic JT, Clark DB, Schirda C, Nagel BJ, Pfefferbaum A. Harmonizing DTI measurements across scanners to examine the development of white matter microstructure in 803 adolescents of the NCANDA study. Neuroimage. 2016;130:194–213.

120. Lally PJ, Pauliah S, Montaldo P, Chaban B, Oliveira V, Bainbridge A, Soe A, Pattnayak S, Clarke P, Satodia P, Harigopal S, Abernethy LJ, Turner MA, Huertas-Ceballos A, Shankaran S, Thayyil S. Magnetic resonance biomarkers in neonatal encephalopathy (MARBLE): a prospective multi-country study. BMJ Open. 2015;5:e008912.

121. Lally PJ, Montaldo P, Oliveira V, Soe A, Swamy R, Bassett P, Mendoza J, Atreja G, Kariholu U, Pattnayak S, Sashikumar P, Harizaj H, Mitchell M, Ganesh V, Harigopal S, Dixon J, English P, Clarke P, Muthukumar P, Satodia P, Wayte S, Abernethy LJ, Yajamanyam K, Bainbridge A, Price D, Huertas A, Sharp DJ, Kalra V, Chawla S, Shankaran S, Thayyil S, MARBLE consortium. Magnetic resonance spectroscopy assessment of brain injury after moderate hypothermia in neonatal encephalopathy: a prospective multicentre cohort study. Lancet Neurol. 2019;18:35–45.

122. Morton S, Vyas R, Gagoski B, Vu C, Litt J, Larsen R, Kuchan MJ, Lasekan JB, Sutton BP, Grant PE, Ou Y. Maternal dietary intake of omega-3 fatty acids correlates positively with regional brain volumes in 1-month-old term infants. Cerebral Cortex; 2019. **(In Press)**.

123. Oishi K, Mori S, Donohue PK, Ernst T, Anderson L, Buchthal S, Faria A, Jiang H, Li X, Miller MI, van Zijl PCM, Chang L. Multi-contrast human neonatal brain atlas: application to normal neonate development analysis. NeuroImage. 2011;56:8–20.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.