Specific Aims

We have entered an age of unprecedented scale, resolution, and accessibility of digitized biomedical image data. The BigBrain[1] project is a publicly accessible digitized image data set of an entire human brain at a resolution approaching individual cells, painstakingly acquired and processed over the course of a decade, with a corresponding magnetic resonance image (MRI). The BigBrain offers an important opportunity to provide common data and standards to bring together neuroscience communities focusing on the macroscopic as well as the microscopic structure of the brain. Since the microstructural (cytoarchitectonic) boundaries cannot consistently be predicted from macroscopic boundaries[2], the BigBrain should be anatomically labeled at multiple scales to maximize its potential. Unlike MRI data, manually labeling a single such brain at high resolution would take an expert many years to complete, and it takes years to train an expert on a given labeling protocol. With the prospect of more high-resolution biomedical images becoming available, we need to scale up our ability to label cytoarchitectonic data such as in the BigBrain. Facing a similar challenge, EyeWire is a successful computer game created for humans to segment mouse brain scanning electron micrographs into three-dimensional neurons. **The primary objective of our project is to create a computer game for microstructural labeling of human brain imaging data.**

Our proposal joins the expertise of the following people: Dr. Arno Klein (Sage Bionetworks) has expertise in brain image processing and has led manual and automated human brain image labeling projects; Dr. Alan Evans (McGill University) helped to create the BigBrain and leads a team interested in making it more useful and more accessible, Dr. Paul Yushkevich (University of Pennsylvania) is a leader in hippocampal labeling of histology and MRI data; Dr. Jesse Himmelstein (Centre de Recherches Interdisciplinaires) creates video games for research and education and makes online platforms for sharing games; and Amy Robinson directs the EyeWire project.

Aim 1: To evaluate non-expert labels, experts will label a small subset of BigBrain hippocampal data. To provide training and to evaluate results of a crowdsourced task, it is necessary to establish gold standards. We will have experts manually label a portion of BigBrain hippocampal regions to establish these gold standards.

Aim 2: To scale up labeling efforts, we will break the labeling problem into smaller tasks within a computer game. Just as the EyeWire game breaks data into small cubes within which isolated tasks are performed by non-experts, we will break BigBrain data into smaller pieces to make labeling a parallel, distributed effort. To engage non-expert labelers, we will map individual tasks to elements of a game that a human would enjoy playing. If this game fails to engage sufficient numbers of players, we will enlist players through Amazon's Mechanical Turk.

Aim 3: Aggregate the results of game play to label the BigBrain. We will combine the thousands of results of playing the game to establish labels for the BigBrain hippocampal regions.

Exploratory Aim 1: Assess the potential for extending the game to all brain regions. The hippocampus is an important region to label, but ultimately we want to label the entire brain, so we will explore how our approach generalizes to all other brain regions.

Exploratory Aim 2: Train a supervised learning algorithm on the crowdsourced results. We propose to radically scale up our ability to label imaging data by training a supervised learning algorithm on the crowdsourced labels, and evaluate how well we can automate labeling.

Research Strategy

(A) SIGNIFICANCE

The importance of segmenting the BigBrain:

The BigBrain[1] represents a decade of work to construct a high-resolution 3-D model of a human brain at nearly cellular resolution (20µm isotropic) based on reconstruction of 7,404 histological sections stained for cell bodies (Fig.1). In 2014, we extended our processing of the BigBrain to include improved section-to-section intensity inhomogeneity correction, 3-D tissue classification, and cortical surface extraction[3]. The BigBrain is extremely important to the neuroscience community because it represents whole-brain histological data that match its accompanying magnetic resonance imaging (MRI) data. It is the best brain imaging data set in the world to establish a canonical space at both microscopic and macroscopic resolutions. However, for the cell-stained histology data to be truly useful, it needs to be segmented into cytoarchitectonic regions, as has been the standard since Korbinian Brodmann[4]. Cytoarchitectonic labeling is important for neuroscience and human health:

- Delineation of regions on the BigBrain, whether obtained from a manual or automated approach, will help improve statistics in probabilistic maps of neocortical and subcortical areas, as well as fiber tracts (although the stain used was for cell bodies, some fibers can be followed in areas of high contrast).
- These cytoarchitectonic maps can be warped to a common reference frame and used as a topographical reference for anatomical localisation of activations observed in functional imaging studies, or as a high-resolution 3-D neuroanatomical guide (atlas) for microsurgery.

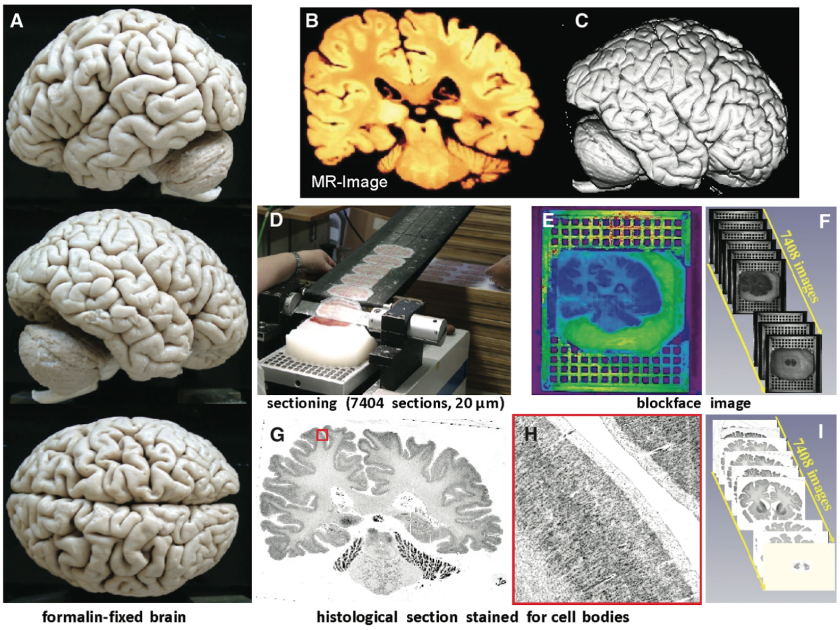

Fig.1. From Amunts, et al. (2013)[1].

Labeling the cytoarchitecture of the human brain requires great expertise acquired over many years of labeling. It is a slow and tedious process; it would take an expert many years to draw all of the major cytoarchitectonic boundaries for the 7,404 sections of the BigBrain, and unless one could convince many such experts to do it, the necessary inter-rater reliability statistics of consistency cannot be measured. In other words, it will be a long time before we see the cytoarchitectonic boundaries of the BigBrain comprehensively labeled by experts, diminishing the tremendous potential that this data set has to offer. *In this project, we propose to overcome the daunting challenge of labeling BigBrain cytoarchitectonic regions through crowdsourcing*.

Labeling the hippocampus:
The medial temporal lobe (MTL) is a complex brain region of enormous interest in research on memory, aging, psychiatric disorders, and neurodegenerative diseases. Within the MTL, the subfields of the hippocampus (cornu Ammonis fields CA1-CA4, dentate gyrus, subiculum) and the adjacent cortical subregions of the parahippocampal gyrus (entorhinal cortex, perirhinal cortex, and parahippocampal cortex) are understood to subserve different functions in the memory system[5–8]. Different psychiatric and neurological disorders are known to affect MTL subregions differentially, selectively, and in a complex progression[9–15]. The non-uniformity of MTL involvement in normal brain function and in disease makes in vivo interrogation of the structural and functional properties of MTL subregions highly desirable. More studies are adopting high-resolution MTL imaging, creating a growing need for accurate and reliable computational techniques for analyzing these data.

Dr. Yushkevich (University of Pennsylvania) is a world expert in hippocampal segmentation and works with hippocampal image data at both scales reflecting BigBrain modalities -- microstructure and macrostructure (see Fig.4). His group has been a major contributor to the MTL subregional imaging field, in particular to the effort to bring the MTL imaging and neuroanatomical communities together with the goal of developing a harmonized protocol for MTL subregion segmentation, and took a lead role on the first paper to emerge from this effort[16]. *In the proposed project, the UPenn team will provide reference segmentation of hippocampal subregions in a subset of slices from the BigBrain dataset as gold standard labels for evaluating crowdsourced labels, toward a rich ex vivo atlas that would greatly facilitate comparisons between different in vivo and ex vivo studies of the hippocampus.*

Computer games for solving scientific problems:
There has been increasing interest in and attention given to "citizen science" initiatives to crowdsource the solving of scientific problems through what is sometimes termed "human computation." Compared to computers, humans currently excel at face and pattern recognition, object identification, abstract symbol processing, etc. Examples of recent scientific computer games which have attracted many players include Foldit and EyeWire.

Foldit (http://fold.it) is an online puzzle video game where the objective is to fold the structure of selected proteins using various tools provided within the game. Foldit's players (now estimated at 240,000 people[17]): provided useful results that matched or outperformed algorithmically computed solutions in 2010[18], helped to decipher the crystal structure of an AIDS-causing monkey virus in 2011[19], and achieved the first crowdsourced redesign of a protein in 2012[20]. Some of the makers of Foldit have introduced two new scientific discovery games, one called EteRNA, in which a player designs RNA molecules to conform to desired shapes (http://eterna.cmu.edu/web/), and the other called Nanocrafter (http://nanocrafter.org/) about synthetic biology, where a player uses pieces of DNA to build everything from computer circuits to nanoscale machines.

EyeWire (http://eyewire.org) is a game to help map neural connections in the visual cortex of a mouse brain played by nearly 200,000 people worldwide. The player does this by linking together computer-rendered, 3-dimensional segments of tissue extracted from scanning electron microscope sections to make neural processes that start from one face of a small cube of data to another face of the cube. By breaking the problem down from something currently computationally intractable (segmentation of interconnected neurons from SEM data) to one that is not only tractable but engaging for players (building a scaffold to escape a cube), they are able to crowdsource a very difficult problem. *Likewise, the primary objective of our project is to crowdsource the segmentation of brain image data, in our case the cell-stained histological sections from a human rather than SEM sections from a mouse.*

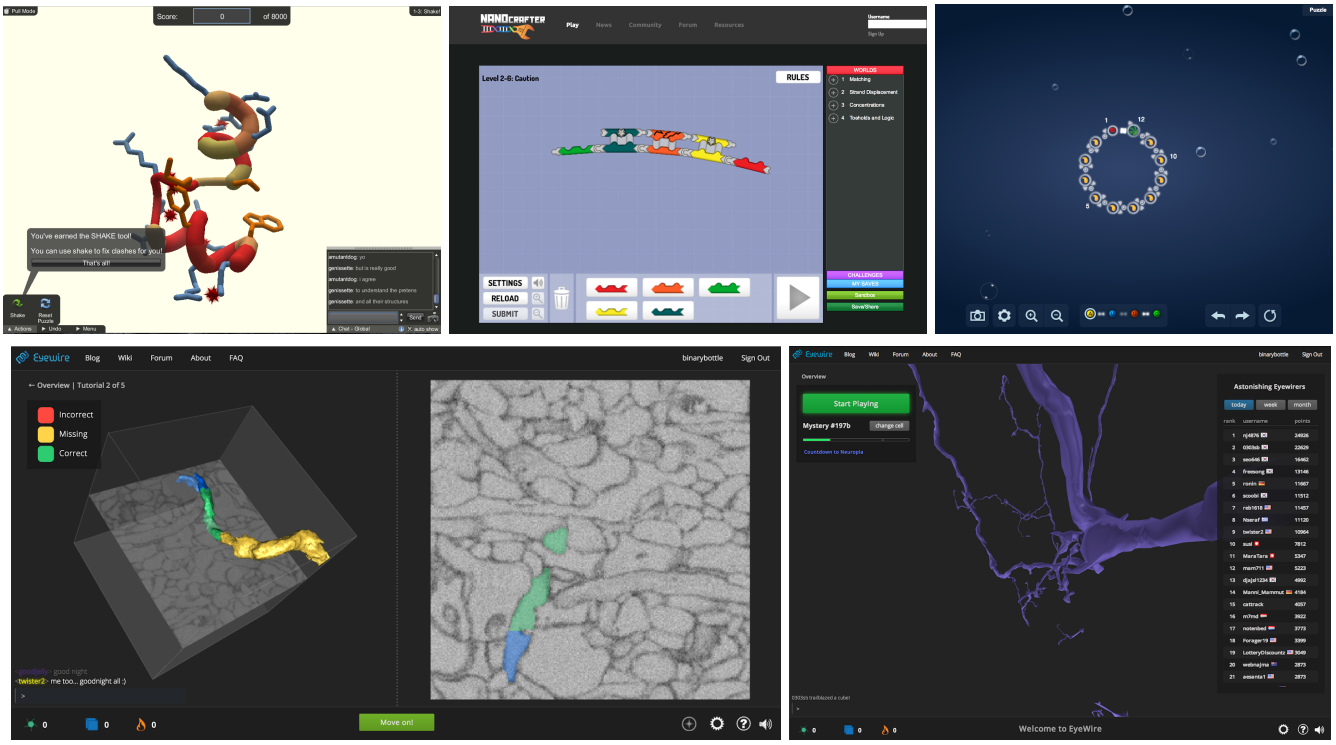

Fig.2. Screenshots of popular scientific games.
Top (left to right): FoldIt, Nanocrafter, and EteRNA tutorials, important for training and evaluating players and for providing scientific context, so that the player is further motivated to play.
Bottom: EyeWire has nearly 200,000 players who build neural connections from mouse scanning electron micrographs. Left: tutorial; Right: neural reconstruction with scoring leaderboard.

HTML5 technologies for browser-based scientific games:
The Web browser is doubtlessly the most widespread and cost-effective way to reach large numbers of potential players of scientific games. And yet, only a few years ago, technical difficulties prevented developers from delivering interactive and computationally intensive experiences over the Web. With the increasing uptake of the HTML5 standard by popular Web browsers, Web applications can now include 3-D graphics (WebGL), high-fidelity sound, and real-time communication with servers, all without browser plugins that require installation. Even better, many of these standards are being adopted by mobile Web browsers, meaning that the same application is available directly on a player's smartphone or tablet.

The Centre de Recherches Interdisciplinaires (CRI) developed two technologies specifically created for scientists to take advantage of HTML5 technologies for games. RedWire (http://redwire.io) is a new kind of game engine for remixing games. Game creators can take apart existing game designs, mix them with others, and add their own functionality. Scientists can create new games faster, and others can contribute variants of their games that may even outperform the original. The novelty of the engine lies in allowing developers to create imperative blocks of code within a functional framework, execute these blocks in parallel and merge the resulting changes, representing input and output as data buffers, and providing visual programming and debugging tools.

RedMetrics (https://redmetrics.io) is the first open data, open source game analytics service. Once a game is released, its creators want to track how players interact with it. Knowing where players get stuck or give up is invaluable for improving the player experience and encouraging players to continue. For a scientific video game, it is critical that teachers and researchers can track how players

interact with the game in order to follow the learning curve or gather experimental evidence. Since no open source solution currently exists to gather in-depth metrics on player behavior, RedMetrics is designed to fill that gap. As a Web service, it is capable of receiving information about any game, regardless of the platform it is developed on.

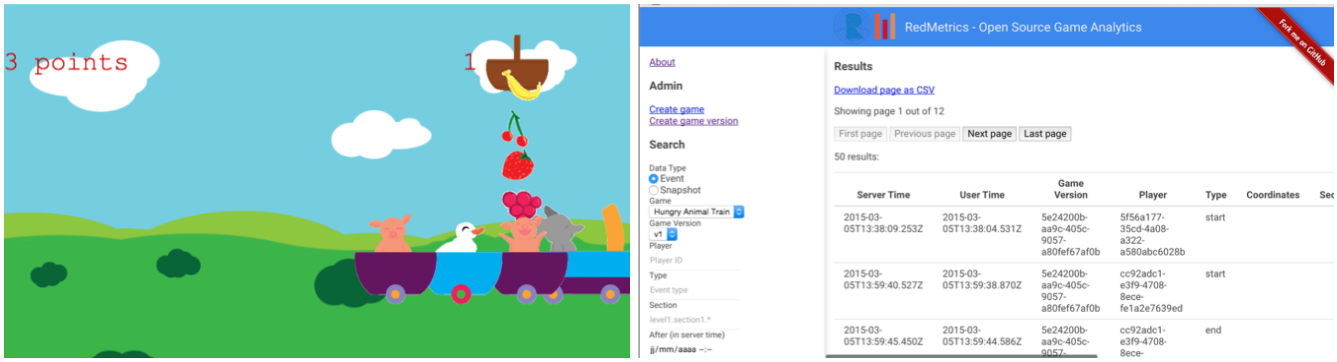

Fig.3. Screenshots of a RedWire game and its RedMetrics game analytics.
Hungry Animal Train (left) is a game created at the CRI Game Lab to study time estimation among children with autism. It was created entirely within RedWire, and publishes data about its use to the RedMetrics game analytics service (right), where researchers, developers, parents, and other interested partners can filter and download the data for statistical analysis.

(B) INNOVATION

Aim 1, while focusing on traditional neuroanatomical methods for segmenting brain cytoarchitecture, is nonetheless innovative since it will modify a hippocampal labeling protocol for use with ex vivo hippocampal data, and will make use of these data as gold-standard labels for evaluation and scoring of players of a computer game.

Aim 2 is innovative since it will be the first instance of crowdsourcing cytoarchitectonic processing or analysis, and that too by individuals who have no neuroanatomical expertise (in contrast with Aim 1). Moreover, this crowdsourcing will take the form of a computer game, and will be the first to deal with cytoarchitectonic data. Players' repeated segmentations will also be used to generate intra- and inter-rater reliability statistics for cytoarchitectonic segmentation, which will complement existing data on inter-brain anatomical variation of cytoarchitectonic boundaries[2]. The game will use the first open data, open source game analytics service (RedMetrics).

Aim 3 will aggregate results from the crowdsourcing above, and will constitute the first time cytoarchitectonic boundary delineation will have been undertaken on small portions of the image for which (macrostructural) anatomical context has been removed. Another innovative contribution will be the software written to aggregate and reconcile tens of thousands of drawn curves to construct 3-dimensional boundary surfaces.

Exploratory Aim 1 is innovative in that it proposes to generalize a game-based approach tailored to a specific domain (the hippocampus) to a larger context (the entire brain).

Exploratory Aim 2 proposes the use of crowdsourced segmentations to train a machine learning algorithm to perform the same task. If this algorithm performs well, this will introduce a new and powerful approach to inferring cytoarchitectonic boundaries of brain regions.

(C) APPROACH

Aim 1: To evaluate non-expert labels, experts will label a small subset of BigBrain hippocampal data.

Dr. Yushkevich will oversee the manual segmentation of BigBrain hippocampal sections for use as a gold standard to evaluate each player's accuracy and consistency, and will use this information to weight player contributions accordingly. Manual segmentation of the entire hippocampus at the resolution of the BigBrain is beyond the scope of this proposal. The UPenn team will label the left and right hippocampal regions in a subset of 60 slices each, taken in the coronal plane with approximately 1mm intervals. This represents 2% of the total number of slices spanning the hippocampus. We will store and release these segmentations and their boundaries as a publicly accessible project in Sage Bionetworks' Synapse platform (http://synapse.org).

Segmentation Protocol:
The segmentation protocols by Ding et al.[21] and Duvernoy[22] will be used as references for the BigBrain segmentation protocol. This protocol uses a combination of different histological stains (e.g., NISSL anatomical stain, AT8 tau stain, etc.) to define rules for distinct anatomical subfields in the hippocampus, with particular focus on the anterior hippocampus where the structure is very complex in shape and organization. Since the BigBrain only uses a single stain, the protocol will be modified to use the features available in the BigBrain data. Additionally, the UPenn group will stay abreast of the active effort to harmonize the hippocampal subfield segmentation protocols[16] (http://www.hippocampalsubfields.com). Laura Wisse, a postdoctoral fellow in the group is involved in the effort to develop harmonized boundary definition rules for the harmonized protocol. Although this protocol targets MRI, the boundary working group is also using histology data to test the veracity of proposed rules. A team of neuroanatomists is examining rules from prior histology and MRI segmentation protocols, and working to define rules that are consistent between them. When the rules adopted from Ding deviate from the rules put forward by this group of anatomists, the UPenn team will use the consensus rules to ensure better alignment with the harmonized protocol for MRI segmentation (expected in 2016-2017).

Boundaries:
Boundaries can be defined either by a difference in the size/shape/density of cells, or in the presence/absence of layers. The hippocampus is considered more primitive cortex and doesn't have the distinct laminar appearance of the cortex. However, parts of the head of the hippocampus do appear this way, and this is one of the ways an expert uses to differentiate regions. The head of the hippocampus also has multiple classes of cells within each subfield. For example, CA3 cells appear quite different in the uncal, vertical, and typical regions[23].

Substructures:
The segmentation protocol will label the following hippocampal subfields: Cornu Ammonis (CA) fields CA1, CA2, and CA3; Dentate gyrus hilar region (DG:H), which is also referred to as CA4 by some authors, and the dentate gyrus proper (DG); subiculum (SUB). The hippocampal layers stratum radiatum and stratum lacunoso-moleculare will be assigned a separate label (SRLM), which correspond to the prominent hypointense band in MRI images of the hippocampus.

Operationalization:
Segmentation will be performed by John Pluta, who has over 5 years of expertise in hippocampal segmentation including ex vivo MRI and histology. He will use the HistoloZee tool (see Figure 4), developed specifically for working with large 3D histology stacks and their segmentation. He will perform segmentation on the highest available resolution data from BigBrain and will use histology data from UPenn and published sources[21-22] for reference. Segmentations will be exported in vector graphics format allowing for future modification.

Reliability:
Inter-rater and intra-rater reliability for the protocol will be established. A set of 10 slices will be used for this purpose. John Pluta will resegment these slices after 1 month or more delay to establish intra-rater reliability. Laura Wisse will perform separate segmentations 1 month apart to establish her own intra-rater reliability and inter-rater reliability with Pluta. Reliability will be reported using the intraclass correlation coefficient, Dice similarity coefficient (a relative overlap measure) and boundary distance measures. The intra- and inter-rater reliability will be valuable data for answering hypotheses about the ability of non-expert gamers to replicate the work of anatomists.

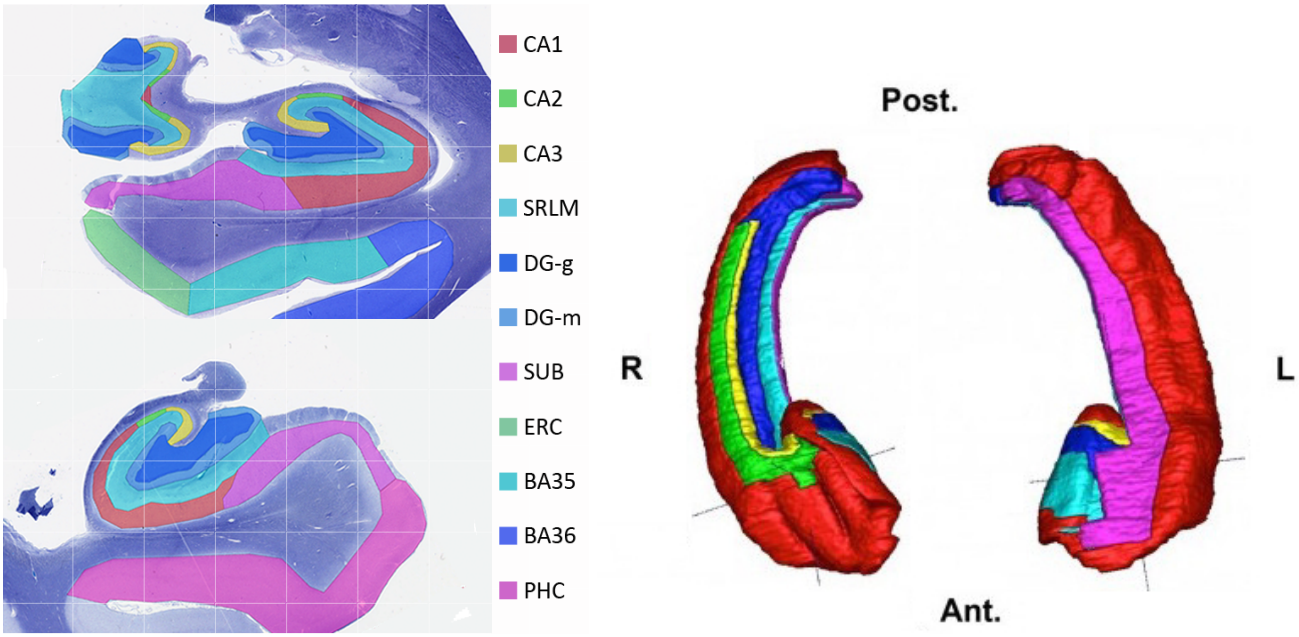

Fig.4. Example hippocampal subfield labels.
Left: Example of hippocampal subfield labeling. Right: The first work demonstrating hippocampal subfield labels in MRI space that are derived from ground-truth histological imaging[24].

Aim 2: To scale up labeling efforts, we will break the labeling problem into smaller tasks within a computer game.

We will break BigBrain data into small pieces to turn labeling into a parallel, distributed effort. Labeling each piece would constitute an element of a game played by non-experts.

Seed boundaries:
The expertly drawn hippocampal subfield label boundaries will only cover a small portion (2%) of the BigBrain's hippocampus (and none of the rest of the BigBrain). We can supplement these with less accurate, automated label boundaries to get a rough sense where we should expect these boundaries to lie, by affine registering other labeled image volumes to a downsampled version of the BigBrain. We anticipate that the inter-subject anatomical variability[2] in the different images and the affine transformation will result in a set of non-overlapping "seed" boundaries in the space of the BigBrain. We can use these seed boundaries to select image tiles and to provide hints while playing the game.

Image tiles:
We will first present the labeling problem to players much as it is presented to expert labelers of neuroanatomy: as 2-D images upon which cytoarchitectonic boundaries are to be drawn. Beyond this, the experience will be quite different. We will not present an entire section to the player, but instead present a small portion (tile) containing gray matter, so there is no context of where in the hippocampus it came from. We will only show players image tiles that intersect seed boundaries, and for each original tile, we will generate two additional tiles translated in x or in y by half of the tile's width. This sliding window approach to tiling the image will account for cases where a label boundary lies along a tile's edge. Generating the tiles and serving them to remote clients will be done using the current BigBrain tile server mechanism developed by Dr. Evans' team. A multiresolution pyramidal 3D structure is generated with blocks of a chosen size from which remote clients can obtain tiles from arbitrary spatial plans at the resolution matching the proper zoom level. The gaming application will only require a subset of this flexible data structure; we will choose the appropriate spatial plans, tile size and resolution to present to the gamers. The current system and infrastructure, hosted on the CBRAIN platform[25], has been benchmarked at 16Gbps for extremely rapid tile read operations during real-time visualization, which is sufficient to serve single images to many concurrent gamers.

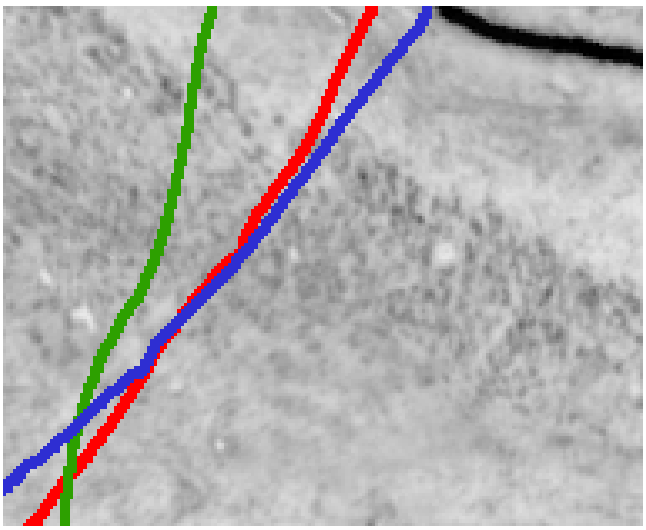

Fig.5. Example boundary estimates in the BigBrain hippocampus.
This figure shows example color line drawings atop a (low-resolution) image tile containing a small portion of the BigBrain's right hippocampus. The lines represent a novice's estimates of the CA1/subiculum cytoarchitectonic boundary. This boundary is extremely difficult, as it results in the greatest overall disagreement among hippocampal subfield labeling protocols[16].

Scoring players:
Players would draw a line or curve on an image tile, and would be scored in the following ways:

1. *Accuracy*: similarity to an expert's drawing
2. *Intra-rater reliability*: similarity to previous drawings by the player on the same tile
3. *Inter-rater reliability*: similarity to other players' drawings on the same tile
4. *Registration-based reliability*: similarity to the seed boundaries used to select the tile
5. *Independent reward*: a score that is independent of other lines

For 1-4 above, similarity between two drawings can be computed by the Hausdorff distance measure, or similarity between regions delineated by the drawings can be measured by various overlap measures[26]. An independent reward can either reflect drawing speed or some other aspect of game play that has nothing to do with the line, such as discovery of a random treasure. This could be used to motivate a player to contribute drawings from which intra- or inter-rater reliability measures can later be computed.

Training:
A new player will have no idea how to draw a boundary between cytoarchitectonic regions, and will rely on the game for initial training and assessment. A proven technique in these circumstances is to present players with "reference" puzzles where the solution is already known (see Figure 2). In this case, the game will show a player image tiles that have been previously labeled with cytoarchitectonic boundaries by experts, although the player does not see the previous labels. The player will draw a boundary and will receive an accuracy score, and based on the score, may receive further training. During actual game play, every so often an image tile from an expertly labeled section will be presented to the player to evaluate how well the player is performing, and perhaps to infer effects of learning, attention, and drift. These intermittent evaluations can be used to retrain the player, or weight the confidence in their contributions. Since the game will compute intra-rater reliability, inter-rater reliability, and registration-based reliability for every drawing, these scores can be used as a proxy for accuracy, or to determine whether the player should be evaluated for accuracy.

Game mechanics:
There are a number of possibilities for the central game mechanic. One promising idea is to represent the labeling process as a drawing or painting game. This approach could appeal to the artistic side of players, especially if the game supports touch controls on smartphones and tablets. Another idea is to approach the labeling as a game about cutting or slicing, which is both a popular mechanic among mobile games and may appeal to players as being akin to brain surgery. The "surgery" would introduce a goal that is independent of labeling: the player would find a path to reach a tumor that must be removed. If the path deviates too much from the "correct" path, the player would be penalized, as in the popular "Operation" game currently manufactured by Hasbro, whereas if the path is close, the surgery is successful and another patient is presented to the player. A "correct" path could reflect one of four scores above: accuracy, intra-rater reliability, inter-rater reliability, or registration-based reliability.

With any (or all) of the types of scores, it would be possible to set up the game as a competition among players, and to display a leaderboard with players' highest scores. Levels could be achieved based on labeling difficulty, determined by intra-rater and inter-rater reliability statistics obtained for

expert labelers or for all players. Alternatively, the game could reward cooperation among players. For example, players could try to guess where others have drawn their lines on a given image tile. This would promote inter-rater reliability and therefore more consistent labels. FoldIt and EyeWire have demonstrated the power of social tools within citizen science games. In addition to a cooperative mechanic, we will consider bringing social interaction to our game via chat, team play, or badges.

The power of crowdsourcing lies not just in the number of minds focused on a challenge, but in the different and unexpected solutions that can arise. Rather than assume that the original image data will be conducive to drawing boundaries, if we were to make available browser-based image filters to enhance contrast, alter colormaps, etc., players could explore different ways of seeing and segmenting cytoarchitectonic regions. The Cancer Digital Slide Archive (CDSA) tool (below) for browser-based visualization of high-resolution image data such as the BigBrain has image filtering and semi-automated segmentation capabilities. We will explore different ways of enabling players to filter the image tiles and evaluate whether image filtering improves labeling accuracy.

Both the drawing and slicing mechanics above offer players the enjoyment of recognizing patterns and the sense of completion that comes with finishing a puzzle. These mechanics could be embellished by adding secondary mechanics such as time limits, treasures to discover, territory to surround and conquer, enemies to fight off, etc. For better or for worse, these "flashier" mechanics turn the focus away from the central scientific task, which may help to keep some players engaged and may turn others away who were attracted by the appeal of a game that purports to contribute to science and society. For this reason, we will need to strike a balance to discover which game mechanics best motivate players to contribute to the game.

Game development:
We will begin our game design process by prototyping different *game mechanics* (above) and testing them directly with players in order to evaluate which ideas provoke the most engagement and provide the best results. Building on this initial insight, we will use an agile software development approach to develop a polished game that we can test with larger and larger user groups. This will resemble the iterative process of design adopted by the FoldIt developers[27].

The ubiquity of the Web browser makes it the platform of choice for a citizen science game that wants to lower the barrier of entry to new users. We will build our game upon JavaScript and HTML5 technologies, ideally without additional plugins that require installation. This approach will enable us to draw upon Dr. David Gutman's CDSA image viewer (http://cancer.digitalslidearchive.net/; example viewing BigBrain data: http://node15.cci.emory.edu/BIGBRAIN/), Montreal Neurological Institute collaborator Dr. Alan Evans' BrainBrowser (https://brainbrowser.cbrain.mcgill.ca/), Dr. Roberto Toro's Seadragon-based drawing tool (https://vimeo.com/115817177), and/or the RedWire game engine, as needed.

Data handling:
Each image tile would have a unique identifier, as would each player, and each player's drawing. This information would be stored in a CouchDB or PostgreSQL database management system. As players play the game, image tiles would be loaded into the game (as described above), and players' vector-based drawings would be stored in the database as JSON objects. Regarding player interaction with the game, we will link the game to the RedMetrics open game analytics service, so that such data will be accessible to citizens, scientists, and game developers alike for analysis.

Aim 3: Aggregate the results of game play to label the BigBrain.

As described above, the proposed game will result in multiple manually drawn curves for each image tile, each one corresponding to a putative cytoarchitectonic boundary for a small portion of a given histological section containing hippocampus. Aggregation of the large collection of manually drawn curves will involve the following steps:

1. Position the multitude of curves in the original space of the BigBrain.
2. Weight the confidence in each curve based on the evaluated skill or estimated performance of each player at the time it was drawn.
3. Propagate label boundary assignments from expert-labeled curves to all other curves.
4. Create a 3-dimensional, probabilistic label assignment for each volume element (voxel) based on its position with respect to all label boundaries in 3-D.
5. Create an additional atlas with hard boundaries based on label fusion (Fig.6)[28], or on a modified version of the STAPLE algorithm ("Simultaneous Truth And Performance Level Estimation") that attempts to estimate ground-truth labels based on incomplete labeling by multiple labelers[29].

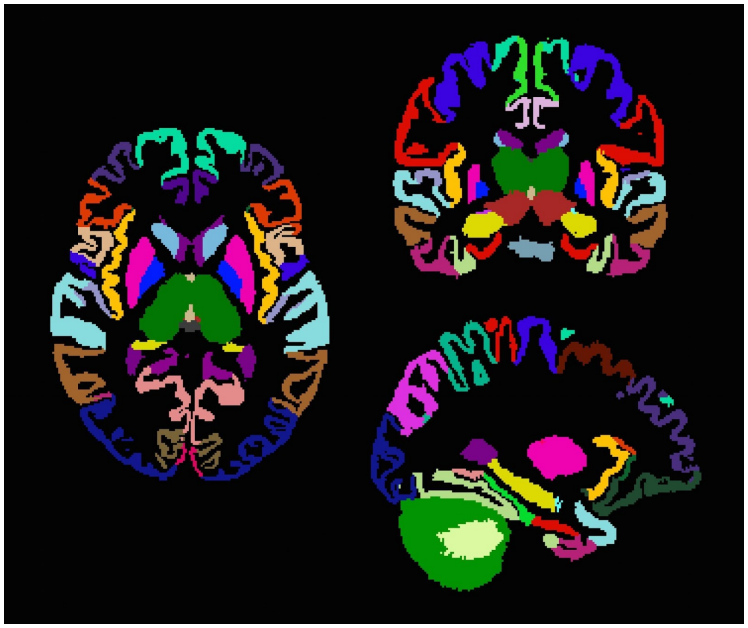

Fig.6. Example joint fusion of multiple label assignments. This macroscopic brain atlas was built from 20 individually labeled atlases, using joint fusion[28], after nonlinearly registering the 20 to a template (http://mindboggle.info/data.html).

Exploratory Aim 1: Assess the potential for extending the game to all brain regions.
Ultimately we want to label the entire brain, so we will explore how our approach generalizes to other brain regions. Since we have expert manual labels drawn by Dr. Evans' team for some limbic regions, we will start with these as gold standard data within the game. If there is time, we will evaluate how similar players are at delineating other regions for which we do not have expert labels. For example, we can test to see how consistently players assign boundaries for cortical regions throughout the BigBrain, and see how these boundaries compare with expected number and position of Brodmann label boundaries[4].

Exploratory Aim 2: Train a supervised learning algorithm on the crowdsourced results.
To explore the potential for learning from crowdsourced human results to automate labeling, we will train a supervised learning algorithm. We will use supervised machine learning methods in the open source scikit-learn Python package (http://scikit-learn.org/), the open source Dato Python package (https://dato.com/products/create/open_source.html), and our own in the R programming environment (http://www.r-project.org/). We have considerable experience building and training classifiers and regression models, and are very interested to try deep learning (convolutional neural network) approaches using the Dato or Theano Python packages. We have previously applied deep learning using Theano to automatically segment human brain MRI data at a macroscopic scale, and are in communication with developers of Dato for guidance in applying deep learning to the BigBrain's microscopic data.

Contingency plans and timeline:
Without training or initialization, it is clear in Figure 5 that a player would simply draw a line between a darker and a lighter region, or between layers, rather than through gray matter for this difficult boundary. The orientation can be initialized in several ways to avoid these tendencies, such as by providing a "starting line" on the outer edge of the hippocampus or by providing a hint, such as the seed boundaries used to select the tile.

If players are unable to draw consistent boundaries, we will increase their training against gold-standard evaluation data and provide hints from better players and from co-registered label boundaries. If this doesn't work, we can simplify the task and ask players if there is, or is not, a boundary within a given tile. A game that can generate a consistent set of boundary-containing tiles will be a success, because we will be able to infer the probabilistic positions of label boundaries, which may be sufficient for most label-based applications, or could be used to help guide expert labelers so that they are much more efficient.

If we fail to attract many players, we will present the game through Amazon's Mechanical Turk (AMT, http://mturk.com). AMT provides an online, on-demand, scalable workforce, where each "human intelligence task" (HIT) is submitted by a Requester and is performed by one of tens of thousands of paid Workers around the world. AMT has been used in many scientific experiments, including cognitive behavioral experiments[30-31], and is considered a convenient, affordable way to attract many workers to perform tasks that a computer is currently poor at executing. The median wage is approximately one to two dollars per hour, and short tasks (around five minutes) are awarded around 10 cents. We would pay Workers a penny to play each round of the game. It would be straightforward to turn our game into AMT HITs using the open source psiTurk Python library (http://psiturk.org/, http://gureckislab.org/mtworkshop/), an open platform for setting up, testing code, posting HITs,

and paying Workers on Mechanical Turk. We would also share our open source code under an Apache v2.0 license with anyone via GitHub (http://github.com) and through psiTurk's Experiment Exchange (http://psiturk.org/ee/). We have considerable experience designing and building Web apps for visualizing and interacting with data, and are poised to modify the browser-based game so that it is accessible to AMT Workers.

Timeline

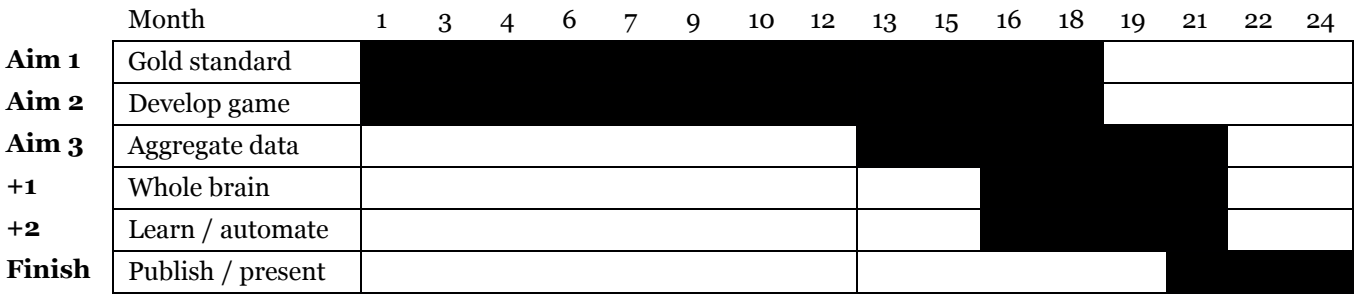

Fig.7. Timeline.
We will prepare the gold standard evaluation data (Aim 1) and develop and update the game (Aim 2) through the 18th month as we get feedback on its use. Year 2 will consist primarily of testing the aggregation of segmented data (Aim 3), exploring how well the game can generalize to segmentation of every region of the BigBrain (Exploratory Aim 1), and training and testing an automated approach that learns from the crowdsourced data (Exploratory Aim 2), as well as to publish and present our findings.

Bibliography & References Cited

[1] Amunts, K, Lepage, C, Borgeat, L, Mohlberg, H, Dickscheid, T, Rousseau, MÉ, Bludau, S, Bazin, PL, Lewis, LB, Oros-Peusquens, AM, Shah, NJ, Lippert, T, Zilles, K, Evans, AC. 2013. BigBrain: an Ultrahigh-Resolution 3D Human Brain Model. *Science* 340(6139):1472-1475. doi: 10.1126/science.1235381.

[2] Amunts, K, Kedo, O, Kindler, M, Pieperhoff, P, Mohlberg, H, Shah, NJ, Habel, U, Schneider, F, Zilles, K. 2005. Cytoarchitectonic mapping of the human amygdala, hippocampal region and entorhinal cortex: intersubject variability and probability maps. Anat Embryol (Berlin). 210(5-6):343-52. PMID: 16208455

[3] Lewis, L, LePage, C, Fournier, M, Zilles, K, Amunts, K, Evans, A. 2014. BigBrain: Initial Tissue Classification and Surface Extraction. *20th Annual Meeting of the Organization for Human Brain Mapping (Hamburg, Germany).*

[4] Brodmann K. 1909. Vergleichende Lokalisationslehre der Grosshirnrinde. Leipzig: Johann Ambrosius Bart.

[5] Squire, LR, Stark, CEL, Clark, RE. 2004. The medial temporal lobe. *Annu Rev Neurosci* 27:279–306.

[6] Moscovitch, M, Nadel, L, Winocur, G, Gilboa, A, and Rosenbaum, RS. 2006. The cognitive neuroscience of remote episodic, semantic and spatial memory. *Curr Opin Neurobiol* 16(2):179–190. PMID: 16564688.

[7] Bakker, A, Kirwan, CB, Miller, M, and Stark, CEL. 2008. Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science* 319(5870):1640–2. PMCID: PMC2829853.

[8] Wolk, DA, Dunfee, KL, Dickerson, BC, Aizenstein, HJ, and DeKosky, ST. 2011. A medial temporal lobe division of labor: insights from memory in aging and early Alzheimer disease. *Hippocampus* 21(5):461–6. PMCID: PMC2918673.

[9] Small, S, Schobel, S, Buxton, R, Witter, M, and Barnes, C. 2011. A pathophysiological framework of hippocampal dysfunction in ageing and disease. *Nature Reviews Neuroscience* 12(10):585–601.

[10] Braak, H and Braak, E. 1995. Staging of Alzheimer's disease-related neurofibrillary changes. *Neurobiol Aging* 16(3):271–8; discussion 278–84. PMID: 7566337.

[11] Arnold, SE, Franz, BR, Gur, RC, Gur, RE, Shapiro, RM, Moberg, PJ, and Trojanowski, JQ. 1995. Smaller neuron size in schizophrenia in hippocampal subfields that mediate cortical-hippocampal interactions. *Am J Psychiatry* 152(5):738–748. PMID: 7726314.

[12] Simić, G, Kostović, I, Winblad, B, and Bogdanović, N. 1997. Volume and number of neurons of the human hip- pocampal formation in normal aging and Alzheimer's disease. *J Comp Neurol* 379(4):482–94. PMID: 9067838.

[13] de Lanerolle, NC, Kim, JH, Williamson, A, Spencer, SS, Zaveri, HP, Eid, T, Spencer, DD, and Eid, T. 2003. A retrospective analysis of hippocampal pathology in human temporal lobe epilepsy: evidence for distinctive patient subcategories. *Epilepsia* 44(5):677–87. PMID: 12752467.

[14] West, MJ, Kawas, CH, Stewart, WF, Rudow, GL, and Troncoso, JC. 2004, Hippocampal neurons in pre-clinical Alzheimer's disease. *Neurobiol Aging* 25(9):1205–1212. PMID: 15312966.

[15] Lucassen, PJ, Heine, VM, Muller, MB, van der Beek, EM, Wiegant, VM, De Kloet, ER, Joels, M, Fuchs, E, Swaab, DF, and Czeh, B. 2006. Stress, depression and hippocampal apoptosis. *CNS Neurol Disord Drug Targets* 5(5):531–46. PMID: 17073656.

[16] Yushkevich, PA, Amaral, RSC, Augustinack, JC, Bender, AR, Bernstein, JD, Boccardi, M, Bocchetta, M, Burggren, AC, Carr, VA, Chakravarty, MM, Chételat, G, Daugherty, AM, Davachi, L, Ding, SL, Ekstrom, A, Geerlings, MI, Hassan, A, Huang, Y, Iglesias, JE, La Joie, R, Kerchner, GA, LaRocque, KF, Libby, LA, Malykhin, N, Mueller, SG, Olsen, RK, Palombo, DJ, Parekh, MB, Pluta, JB, Preston, AR, Pruessner, JC, Ranganath, C, Raz, N, Schlichting, ML, Schoemaker, D, Singh, S, Stark, CEL, Suthana, N, Tompary, A, Turowski, MM, Van Leemput, K, Wagner, AD, Wang, L, Winterburn, JL, Wisse, LEM, Yassa, MA, Zeineh, MM, and for the Hippocampal Subfields Group (HSG). 2015. Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: Towards a harmonized segmentation protocol. *Neuroimage*. PMID: 25596463.

[17] Marshall, Jessica. 2012. "Online Gamers Achieve First Crowd-Sourced Redesign of Protein". Scientific American.

[18] Cooper, S, Khatib, F, Treuille, A, Barbero, J, Lee, J, Beenen, M, Leaver-Fay, A, Baker, D, Popović, Z, and Foldit players. 2010. Predicting protein structures with a multiplayer online game. *Nature* 466:756–760.

[19] Khatib, F, Dimaio, F, Cooper, S, Kazmierczyk, M, Gilski, M, Krzywda, S, Zabranska, H, Pichova, I, Thompson, J. 2011. Crystal structure of a monomeric retroviral protease solved by protein folding game players. *Nature Structural & Molecular Biology* 18(10):1175. doi:10.1038/nsmb.2119.

[20] Eiben, C, Siegel, J, Bale, J, Cooper, S, Khatib, F, Shen, B, Foldit players, Stoddard, B, Popovic, Z, Baker, D. 2012. Increased Diels-Alderase activity through backbone remodeling guided by Foldit players". *Nature Biotechnology* 30(2):190–192. doi:10.1038/nbt.2109. PMC 3566767. PMID 22267011.

[21] Ding, SL,, Van Hoesen, GW. 2015. Organization and detailed parcellation of human hippocampal head and body regions based on a combined analysis of cyto- and chemo-architecture. *J Comp Neurol.* Apr 14. doi: 10.1002/cne.23786.

[22] Duvernoy, HM. 1999. The Human Brain: Surface, Three-Dimensional Sectional Anatomy with MRI, and Blood Supply. Second Edition. Springer Wien, New York.

[23] Ding, SL, Van Hoesen, GW. 2015. Organization and detailed parcellation of human hippocampal head and body regions based on a combined analysis of cyto- and chemo-architecture. J Comp Neurol. [Epub ahead of print]. doi: 10.1002/cne.23786

[24] Adler, DH, Pluta, J, Kadivar, S, Craige, C, Gee, JC, Avants, BB, and Yushkevich, P. 2014. Histology-derived volumetric annotation of the human hippocampal subfields in postmortem MRI. *Neuroimage* 84(1):505-523.

[25] Sherif, T, Rioux, P, Rousseau, M-E, Kassis, N, Beck, N, Adalat, R, Das, S, Glatard, T, and Evans, AC. 2014. CBRAIN: a web-based, distributed computing platform for collaborative neuroimaging research. Front. Neuroinform. http://dx.doi.org/10.3389/fninf.2014.00054

[26] Klein, A, Andersson, J, Ardekani, BA, Ashburner, J, Avants, B, Chiang, M-C, Christensen, GE, Collins, DL, Gee, J, Hellier, P, Song, JH, Jenkinson, M, Lepage, C, Rueckert, D, Thompson, P,

Vercauteren, T, Woods, RP, Mann, JJ, Parsey, RV. 2009.
Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration.
NeuroImage. 46(3): 786-802. PMCID: PMC2747506

[27] Cooper, S, Treuille, A, Barbero, J, Leaver-Fay, A, Tuite, K, Khatib, F, Cho, K, Michael Beenen, S, Salesin, D, Baker, D, and Popović, Z. 2010. The challenge of designing scientific discovery games. In *Proceedings of the Fifth International Conference on the Foundations of Digital Games* (FDG '10). ACM, New York, NY, USA, 40-47. DOI=10.1145/1822348.1822354

[28] Wang, H, Suh, JW, Das, SR, Pluta, JB, Craige, C, Yushkevich, PA. 2013. Multi-atlas segmentation with joint label fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35(3):611-623.

[29] Landman, BA, Bogovic, JA, Prince, JL. 2010. Simultaneous Truth and Performance Level Estimation with Incomplete, Over-complete, and Ancillary Data. *Proceedings of SPIE--the International Society for Optical Engineering*. 7623:76231N-. doi:10.1117/12.844182.

[30] CogSci2011 workshop: "How to use Mechanical Turk for Cognitive Science Research" http://cognitivesciencesociety.org/uploads/2011-t4.pdf

[31] Mason, W, Suri, S. 2012. A Guide to Behavioral Experiments on Mechanical Turk. *Behavior Research Methods* 44(1):1-23. http://rd.springer.com/article/10.3758%2Fs13428-011-0124-6