Huda Y. Zoghbi
Howard Hughes Medical Institute
Baylor College of Medicine
Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital

I am honored to be able to participate in this meeting. In fact, all of us here in this room are enjoying a rather remarkable privilege. We few get to shape a momentous initiative, and I for one feel the weight of this enormous ethical responsibility. We must remember the interests of the larger community of neuroscientists who do not have a voice here today, including our trainees and young investigators whose careers and research paths will be disproportionately affected by this project for decades to come.

My goal this morning is to warm us up for the crucial conversations we will have over the next day and a half; to encourage us to ask hard questions and to think carefully about the consequences that projects of this magnitude can have, not only on scientific knowledge but on the way we practice science.

The brain initiative has often been compared to the Human Genome Project. We all know that the brain is orders of magnitude more complex than the genome, but before moving onto some concrete ideas to ground discussion of the brain mapping initiative from the perspective of a neuroscientist, I want to examine three points of comparison.

First, the genome project had a concrete goal.

I was a first-year postdoc attending a Cold Spring Harbor meeting on Molecular Biology of *Homo Sapiens* in June of 1986 when Jim Watson expressed strong support for sequencing the human genome. He unwittingly launched what turned into a heated debate about the merits of the project. On the one hand, the project had a very clear, finite objective: to read out a few billion base pairs. We knew what to do with the results, and there was a clear end-point. On the other hand, the DNA code was just the beginning of understanding genes.

Many people wondered what would sequencing the genome deliver? Who would benefit? Is it worth spending all this money on sequencing “junk” DNA?

Some overly optimistic types predicted an immediate revolution in clinical medicine; clearly that was before we realized that figuring out the DNA code was just the first step toward understanding gene function, before we really appreciated the complexities of epigenetics, small and non-coding RNAs, and so forth. But at the same time, there have been remarkable advances: cloning the gene for a sporadic disorder like Rett syndrome took my lab 16 years, whereas today it could be done in a matter of days. And our practice of medicine *has* changed: As a child neurologist I used to order dozens of tests to diagnose a mystery case, but today we order exome sequencing and can get an answer in 30-40% of the cases.

Moreover, the sequencing of genomes from other species was transformative, as it uncovered the cross-species conservation of so many genes and pathways. Then came the variances in genomes that led to sequencing of 1,000 genomes, which in turn revealed tens of thousands of single nucleotide polymorphisms whose effects remain to be determined. Not to mention the epigenome, whose metastability serves as the interface between the genome and the environment, and which we are only now beginning to understand.

So I am not troubled if we cannot predict all the outcomes of the brain initiative. However, we do need clear, concrete goals and milestones.

The second point of comparison has to do with technology.

Technology was a big part of sequencing the genome. In fact, some scientists who were in favor of the project argued nevertheless that we should not jump into massive sequencing using the technology that was then available. They thought that effort should be devoted to developing better technology before launching the project in earnest. Fluorescence labeling and the capillary sequencers were, in fact, huge technological advances. Yes, it was still Sanger sequencing, so the basic method didn't change until more recently, but it's hard to

argue that technology development wasn't a major part of the genome project. In addition, Venter's use of genome wide shotgun assembly was a revelation. It came mostly from improvements in algorithms, longer read lengths and computational power. The benefits of the influx of bioinformatics due to the genome project have spread well beyond genome science. I propose we also need to think about what technology needs to be developed, and what portion of funding needs to be set aside for technology development. But of course we cannot decide that until we know what we want to measure.

The third point of comparison has to do with effects on the sociology of science as a practice.

With all the benefits of the genome project that I just talked about, it did have one major side effect: it ushered in a period when biomedical research became mostly about sequencing, sequencing, and more sequencing. Functional studies, sophisticated biochemistry and physiology, and the creativity focused on developing new technologies in these disciplines faded for a while. Thankfully, functional studies are starting to have a comeback. I would argue that the impact of the genome project might have been far greater if the planning of the project dedicated separate funding for high throughput functional studies. Maybe we would have had gains beyond molecular diagnoses.

So before we start to discuss the goals of the brain mapping initiative, I want to stress that we should guard against the tendency to let one paradigm dominate biomedical research. We can't have one technology taking over neuroscience. We need to be sure that the project leaves room for new and creative ideas that none of us can think of today! We also have a responsibility to make certain that we do not take away from investigator-initiated research. We're meeting today under the shadow of the sequester, and at a time when science as a whole is under attack, we need to protect the ability of scientists to pursue the questions they deem most interesting and important.

So now I'd like to switch gears and share my thoughts about the brain initiative.

I am not here to suggest one goal or another, this is much better done collectively, but I thought I would share some thoughts for us to consider over the day and a half.

There has been a lot of talk about recording from every neuron in a circuit so that we can understand how the circuit functions. As we consider such an endeavor, however, we need to think about what it will take to decipher an emergent circuit state.

To understand how the brain works, I believe we need to know the following:

1. We need to know ALL of the components of the brain, not just neurons, and we need a much better idea than we currently have about what defines them at a molecular level.
2. We need to know how these components connect and communicate.
3. We need to figure out how to measure such communication.
4. And most importantly, we need to know how brain activity varies with experiences and external stimuli.

Let's start with the components of the brain. What molecules are present, at what levels, and in what spatial distribution? This will let us figure out how the cells got that way and what we can do to reprogram them. We are not even close to knowing all the different cell types and what defines them. We all know glia outnumber neurons, yet we have barely a concept about their diversity.

While we need better technology to define the molecular anatomy of the various brain cells, we have a framework to permit rapid progress in this area. The Allen Brain Institute has laid a great foundation for such efforts and so did Gensat. This is something we can start doing right away and the data can be disseminated so that the greater scientific community can begin to put the information to use. This is one example of a concrete goal that would require some investment in refining technologies but that would benefit the entire field and be widely used by the community.

Let me share an example from my own work. We discovered that mutations in a gene that encodes protein called MeCP2 cause a terrible neurological disorder called Rett syndrome,

whose symptoms ranged from loss of language and social skills around two years of age to inability to control movement. It so happens that MeCP2 is expressed all over the brain. But the symptoms were so wide ranging that we decided to study MeCP2 by deleting it only in one specific group of neurons at a time. To our great surprise, we learned that loss of MeCP2 from inhibitory neurons causes almost all features of the disease, including premature death. Even more surprising is that loss of MeCP2 only partially disables neuronal function, by decreasing their signaling by about 30%; it doesn't kill the neurons or inactivate them. Who would have thought that partial disabling of inhibitory neurons is enough to cause a plethora of neuropsychiatric phenotypes and death? This was possible only by knowing molecules that define a group of cells and studying the effect of one protein. Somehow we must try to capitalize on such detailed molecular information and behavioral data as we study circuits.

Second, how do various cell types connect and communicate? To understand networks, we need to understand points of contact. Again, this is an area where we are just beginning to gain knowledge. While retrograde labeling techniques using rabies virus are successful, we can optimize retrograde mapping and develop anterograde mapping technology. Renaissance disciplines in neuroscience such as retrograde mapping and optogenetics relied on virology and microbial biology. We should remain open-minded to what other disciplines have to teach us. Having tools to study specific cells and their inputs and outputs will be extremely powerful.

Third, we need to figure out how to study how cells communicate. This is the most obvious area where we need disciplines such as physics, mathematics, bioengineering, and computational sciences. We need to develop technologies so that we can visualize and record from cells deep within the brain, not just on its surface. We also need to increase our ability to record from a large number of cells at time. Of course the real question is do we have to record from every single cell within a network. I suspect that a more complete understanding of each brain cell type and their inputs and outputs will help us determine the number we need to record from. Might we learn a lot from recording from just 10-20% of neurons? I'd like to offer an analogy. Merely possessing the Oxford English Dictionary does not enable one to speak English. When we learn a language, we have to learn its basic

structures, its rules of grammar, noun declensions, verb tenses, and so forth, along with its irregularities, which is usually helped by an understanding of its history. In other words, we want to know the *principles* of how language works so we can understand what it is capable of. In neuroscientific terms, we want to know the principles of neuronal function, not just a recording of their activity over a certain period of time.

Fourth, how does circuit activity vary over time, based on different experiences and external stimuli, not to mention developmental stage, age, sex, activity, time of day, and disease progression? We need to remember that brain activity does not happen in a vacuum. How would we interpret information from *all* spikes from *all* neurons? We couldn't -- unless we also recorded at the same time *all* external stimuli and *all* aspects of behavior. There would be no end in sight. Furthermore, we know that we lose hundreds of thousands of cortical neurons at a steady state from early childhood (this is another argument for why we might not want to record form every cell—many will disappear). A circuit in newborn animal with limited experiences is vastly different from that of a mature adult animal. Should we consider recording from such vastly different circuits? I think we should consider this for many reasons that we might debate during the next two days, but clearly we need to define not only what needs to be measured, and in which cells, but we need defined endpoints.

> [*Note: Immediately after presenting these thoughts, I learned that Partha Mitra had expressed the same criticism of the notion of recording from every neuron in an article published in *Scientific American* (What's Wrong with the Brain Activity Map Proposal," March 5, 2013). I would argue that the fact that Partha, I, and other neuroscientists question the initial ambition to record from every neuron only underscores the importance and transparency of this consideration].

I would posit that the map can only be understood if we know the identity of the cells whose activity we are recording, understand their unique biochemical properties, and determine how the activity of each of these cell types changes in response to differing conditions. The power of approaching the map combining molecular anatomy and physiology goes beyond our ability to interpret activity. It will help us connect our discipline with advances born of other disciplines. For example, human genome sequencing of individuals with neuropsychiatric disease will provide a list of thousands of genes that are associated with such disorders. Finding ways to develop functional relationships between disease-associated DNA changes,

molecules that define brain cells, and activity changes will bring us closer to understanding how a circuit might become vulnerable.

Biology, genome projects, and neuroscience have taught us the value of the model organism, we need to continue to capitalize on this. Take for example the fly gene atonal, which governs the development of hearing and proprioceptive organs. Its mammalian homolog specifies multiple components of the proprioceptive and auditory systems in mice, so here we have an example of a macrocircuit determined by one molecule! Knitting molecular anatomy, cross-species studies, and connectivity into the overall activity map will have a great impact.

Obviously, some behaviors we're interested in would be better studied in humans than animal models. Somehow we need to make the research more iterative between humans and the lab. Take for example, deep brain stimulation: DBS of Cingulate area 25, seems to be helping severe depression. Yet there are still some patients who do not respond to such DBS therapy while others do. What is different between them? Helen Mayberg used MRI on white matter to study the difference between responders and non-responders and found that responders showed activation of white matter pathways linking Cingulate 25 to three distinct remote regions. She is now testing a DBS paradigm that better activates a more distributed network. Such studies can guide lab work in animal models for fine mapping of such micro and macrocircuits and can then inform future clinical research.

Last but not least the success of this project will depend both on people who receive the funding and on the greater scientific community that does not directly participate in this project. Thus it is critical that all data generated from this project go public as soon as they are generated and be made accessible and user friendly. This was why the Genome project had such an enormous influence. Data came on line immediately and the larger scientific community put it to great use. Most of the significant insights came from the greater community and not the data generators, who were really performing an altruistic service for the larger community. We must adapt this model and have a well-designed Brain Browser that every scientist can use.

Thank you.