SPECIFIC AIMS

Seven million hours of human labor were spent building the Empire State Building. The Panama Canal took twenty million human-hours to complete. The construction of these monumental structures required the coordinated efforts of tens of thousands of people over many years. By comparison, it has been estimated that nine *billion* human-hours are spent playing Solitaire annually. More generally, up to 150 billion hours (the equivalent of 17 million years of human effort) are spent playing games **every year** [1]. Obviously people play Solitaire and other games because they are enjoyable and fun. But aside from that enjoyment, the time spent playing these games largely results in no tangible benefit, neither to the individual nor to society at large.

Here, we propose to build several **"games with a purpose",** a class of games that focuses on collaboratively harnessing gamers for productive ends. This proposal will use scientific games to engage a broad community of individuals in advancing biological research. Specifically, our games will focus on creating biological value on three distinct levels. First, we will annotate the function and disease relevance of individual human genes. Second, we will construct a functional gene network that represents consensus biological knowledge. Third, we will discover multi-gene biomarker panels for human disease.

This proposal will provide a mechanism whereby the entire biological research community and motivated individuals from the society at large can participate in improving our understanding of gene function in health and disease. We will use online games as the mechanism to enable and encourage broad community participation. Our games will be based on existing game designs that are proven to be fun and engaging. If we can harness even a fraction of the "Solitaire time" from the gamer community, then already we will have access to a substantial, novel and *free* resource to advance scientific research. This proposal will test the hypothesis that biological games can substantially improve genome-scale science.

This objective will be achieved through the following three Specific Aims:

Specific Aim #1: Use biological games to annotate human genes. Identification of novel structured gene annotations for diseases and functions will improve many statistical analyses of genetic and genomic data.

Specific Aim #2: Use biological games to create a "community consensus network" of gene relationships. The resulting network of biological knowledge will define both the existence of gene-gene edges, as well as the specific nature of each gene-gene relationship.

Specific Aim #3: Use biological games to discover biomarkers of human disease. These games will identify molecular features for studying disease mechanisms, diagnosis, prognosis, and treatment selection.

As with any scientific proposal to build novel methodology, we strongly believe in the value of a **driving biological problem (DBP)**. Focusing on a concrete biomedical goal will improve the quality of the resulting data. Therefore, our team includes Professor Dan Salomon, an immunologist and expert in organ transplantation (see Letter of Support). Dr. Salomon has extensive experience with genetic and genomic profiling in the context of organ transplantation and rejection. Analysis and interpretation of these data have been hindered by the incompleteness of structured annotations. Moreover, as a physician-scientist, Dr. Salomon has a practical need for improved biomarkers that reflect his patients' immunological state, which in turn would provide objective measures for adjusting drug treatment. In recognition of this DBP, all of our Specific Aims will initially be focused on transplantation immunology. Once we are confident that our game designs are producing useful data for Dr. Salomon's research, these games will readily generalize to other therapeutic areas of interest.

CHALLENGE, INNOVATION AND IMPACT STATEMENT

Sequencing the human genome and characterizing the ~25,000 genes contained therein was a milestone achievement for biomedical science. Comprehensively annotating how those genes affect human health and disease represents an even bigger challenge, and one that is only at its earliest stages. As sequencing costs fall, many have noted that we are in an era of a "thousand-dollar genome, but a million-dollar interpretation".

Effectively interpreting large sets of genomic and genetic data represents a critical challenge in biomedical science. Meeting this challenge requires a firm foundation of biological knowledge. We propose that this foundation can be conceptually divided into three distinct levels: detailed and comprehensive annotation of gene function, understanding of how genes interact in biological networks, and discovery of molecular biomarkers for diagnosis and prognosis of biomedical phenotypes.

Traditionally, science has addressed massive challenges like these through the distributed and independent efforts of many research groups. Biomedical knowledge grows incrementally with each new publication. Recently, "**crowdsourcing**" has emerged as a complementary approach that directly harnesses the collaborative efforts of large communities of people. This proposal outlines a novel mechanism of crowdsourcing for addressing the massive challenges underlying a one-million-dollar genome interpretation.

RATIONALE

Crowdsourcing has been the foundation of many successful web-based applications, from YouTube (for producing video content) to Yelp (for reviewing businesses) to Wikipedia (for creating a free, online encyclopedia). This principle has also been applied to scientific challenges of massive scale. For example, the Galaxy Zoo initiative enables citizen astronomers to classify galaxies in large sets of celestial images [3, 4]. The Gene Wiki project, spearheaded by our group, engages the research community to create a gene-specific review article for every human gene. Similar initiatives have emerged for RNA families and biological pathways. The motivations for users to contribute to these efforts involve altruism, personal learning, developing a sense of community with other contributors, or some combination thereof.

One emerging trend among crowdsourcing initiatives is the use of games as a mechanism to attract contributors. Games add "fun" as another motivation to participate in crowdsourcing projects of massive scale. Why is "fun" such a powerful motivation? Because as a society we spend *billions* of hours each year playing games. If we can productively harness even a small fraction of that energy using **"games with a purpose" (GWAPs)**, then we can leverage an incredibly large human resource for addressing huge challenges.

One of the first and most successful GWAPs, called the **"ESP Game"**, had the ambitious goal of tagging all online images with informative keywords. Accurate keywords are essential for image searching and web-browsing for the visually-impaired. However, computational approaches to tagging are rudimentary, and the cost of manual tagging is prohibitively high. To address this challenge, the ESP game developers transformed this mundane task into an interactive and engaging game. This game resulted in 50 million labels produced by more than 200,000 players [5]. Similarly successful games were later developed to annotate music, text, and videos.

Arguably the most successful GWAP since the ESP Game was actually created to advance biomedical science. The "**Foldit**" game has harnessed the efforts of over 300,000 gamers to predict protein structure from primary sequence, to provide accurate structural models that led to the crystal structure of a previously intractable retroviral protease, and to design new protein folding strategies and algorithms [6]. These successes are discussed in more detail later in this proposal.

Crowdsourcing is crucial because it scales with the rate of modern science. Our team of investigators has successfully leveraged crowdsourcing for biology in the past. Here, we will create games to address fundamental challenges underlying the million-dollar genome interpretation, testing the novel hypothesis that games can be an effective collaborative platform to address biological challenges of massive scale.

APPROACH

Because the methodologies underlying this proposal may be unfamiliar to many reviewers, we include a rather extensive background section. This proposal builds on larger trends in society and the Internet. For those interested in that broader context, we suggest [1, 7, 8] and TED talks from Jane McGonigal and Clay Shirky.

Crowdsourcing science. Crowdsourcing harnesses the coordinated efforts of large communities of individuals toward addressing a common goal. Crowdsourcing has become an increasingly popular model for general-purpose websites. For example, Wikipedia, YouTube, Amazon Reviews, and Yelp are all examples where content is drawn from a large and diverse set of contributors. The most noteworthy feature in all these cases is that they are built on society's **"cognitive surplus"** [7]**.** These organizations do not pay contributors for their content, but have merely provided a platform for them to share their knowledge and effort for free.

Crowdsourcing is gaining momentum in science as well. The open source software movement has long used crowdsourcing principles to develop useful software for science. The "Galaxy Zoo" project has engaged more than 200,000 volunteers to contribute over 100 million galaxy classifications from astronomical images [4]. Our group develops the Gene Wiki, an effort to build a gene-specific review article for every human gene directly within the online encyclopedia Wikipedia. The Gene Wiki is collectively viewed over 50 million times and edited almost 20,000 times per year. The biocuration community also alluded to the power of crowdsourcing in a recent review article, noting that "*sooner or later, the research community will need to be involved in the annotation effort to scale up to the rate of data generation*" [9].

With the exception of the Gene Wiki and its Wikipedia-derived editor-base, crowdsourcing efforts in biology have struggled to generate a critical mass of contributors. In most cases, the altruistic incentives that presumably led to Wikipedia's success are simply not strong enough to motivate biologists to contribute to those resources. This proposal targets a completely different motivation to build a community of contributors. Specifically, this proposal appeals to contributors' sense of *fun*. We seek to convince users to contribute to our gene annotation effort simply because we make the process fun and enjoyable.

Games with a purpose. It has been estimated that nine billion human-hours are spent playing Solitaire each year [8]. More generally, industry groups estimate that as many as 73 billion human-hours are spent playing video and computer games in the United States alone [5], up to perhaps 150 billion gaming hours annually worldwide [1]. If just a fraction of this "Solitaire time" can be productively harnessed, then this human resource has the potential to address very big problems. One strategy to compete with Solitaire is to create **"games with a purpose" (GWAP)** [5]. GWAPs disguise useful work in the form of fun and engaging games.

One of the first GWAPs was the **ESP Game**, used to create text labels for images. Using a simple two-player game design, more than 200,000 people contributed more than 50 million image labels [5]. All of this work was willingly done for free. Google later licensed the ESP Game to improve their image search product. The ESP Game leveraged a powerful new resource by converting a boring repetitive task into an entertaining and social game. The ESP Game serves as the model for the games in **Specific Aim 1** and **Specific Aim 2**.

Foldit, the protein folding game. Perhaps the most successful GWAP to date actually addresses a fundamental biomedical challenge: computational protein folding. The Foldit game (http://fold.it) challenges players to find a protein's lowest-energy three-dimensional structure. Players visually manipulate the protein backbone and side chains, and Foldit gives continuous feedback to the player through a real-time score derived from a free energy approximation. For example, as more hydrophobic residues are buried and more hydrogen bonds established, the player's point total increases.

Leveraging their army of more than 300,000 game players, Foldit has demonstrated remarkable achievements thus far. Among their many successes, Foldit players have: generated better structure predictions than the state-of-the-art Rosetta structure prediction program, solved the structure of a retroviral protease that was previously intractable to both computational and experimental methods, and devised a new protein folding algorithm with performance that is competitive with the most recent professionally created solutions [6]. The Foldit game is particularly relevant to **Specific Aim 3**, and more details of Foldit are provided in that section.

Target audience. This proposal describes the use of biological GWAPs to crowdsource three massive challenges in genetics and genomics. The games described in **Specific Aim 1** and **Specific Aim 2** are targeted at professional biologists. To roughly estimate the amount of effort that could reasonably be leveraged, we started with the Department of Labor's estimate of 91,300 biological scientists in the United States. Capturing just 1.3 hours per year per biologist (the average *Solitaire time* worldwide) would result in ~117,000 human-hours of work or the equivalent of 56 full-time employees. Capturing 21 hours per year per person (the average *gaming* time) would be equivalent to almost 1,000 full-time employees. Although these estimates have huge error bars, we believe they qualitatively illustrate the tremendous potential of GWAPs.

The target audience for **Specific Aim 3** is different. Since we will harness players' innate problem solving and language comprehension skills, the player population is not limited to professional biologists. In short, the potential player community is almost infinitely *massive*. This game will compete for as much of the 150 billion hours of the world-wide, annual gaming time as possible. As Foldit clearly demonstrated, harnessing even the smallest sliver of that massive human resource can result in ground-breaking scientific achievements.

SPECIFIC AIMS

This proposal includes three Specific Aims, each of which describes the development of scientific GWAPs. In order, these aims will focus on generating detailed and comprehensive annotations of gene function, an understanding of how genes interact in biological networks, and discovering biomarkers for human disease using genetic and genomic features. These aims address key challenges in biomedical research.

Specific Aim #1: Use biological games to annotate human genes. Identification of novel structured gene annotations for diseases and functions will improve many statistical analyses of genetic and genomic data.

Background. Using the tools of high-throughput biology, scientists can quickly identify long lists of candidate genes that differ between two experimental conditions. To interpret these gene lists, biologists use structured annotations to discover fundamental properties like gene function and disease relevance. Moreover, these structured annotations are a foundation on which many bioinformatics and statistical analyses are built. For example, gene set enrichment, pathway modeling, and cross-genome comparisons are just a few of the analyses that depend on structured gene annotations [10]. And the importance of methods like these will only grow as the rate of genomic data generation increases.

Unfortunately, gene annotations are incomplete. Moreover, the biocuration community recognizes that without fundamental changes, the knowledge gap between what is known in the literature and what is structured in formal annotations will grow [9]. Biologists can easily observe this gap by searching for their favorite gene or pathway within the Gene Ontology database. For example, there are currently eleven human genes with a formally annotated role in "regulatory T cell differentiation" (GO:0045066). An expert in this area could (and did) readily identify many missing genes, including *RUNX3*, *IL10*, *IL4*, *IL13*, *CD25/IL2RA*, and *INHBA* (**Figure 1**). Structured annotations for diseases are even less complete. For example, despite the well-recognized role for glutamate decarboxylase (GAD) in Type 1 diabetes (T1D), this link is not found in any of the most popular

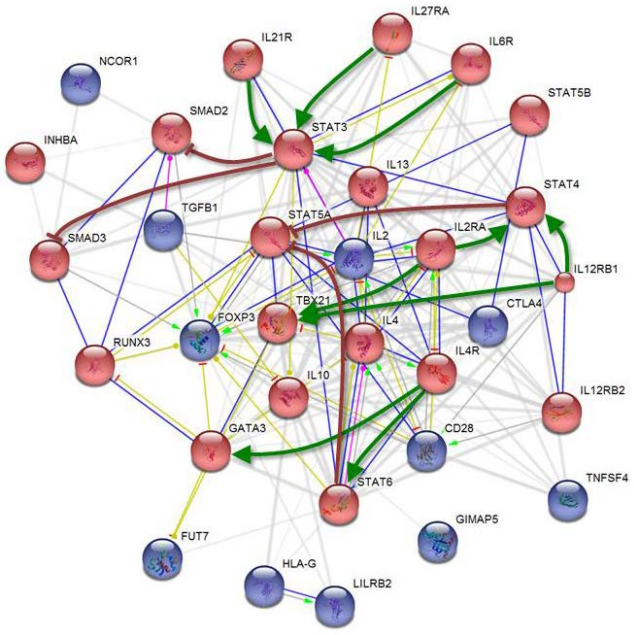

Figure 1. Network of genes involved in regulatory T cell differentiation. This pathway was chosen to illustrate the gaps in knowledge between structured databases and biological knowledge. Gene nodes colored blue had a previously-annotated role in this pathway. Genes colored red were readily identified by an expert as missing. The network diagram, including inter-gene edges, was originally generated using the STRING database [2]. Thick arrows indicating activation (green) and inhibition (red) have also been overlaid to illustrate expert-identified gene-gene relationships that are currently unannotated.

disease databases including PharmGKB, OMIM, HuGE Navigator and KEGG. These functional annotations are simply missing from the gene annotation databases. Their absence means that these fundamental pieces of knowledge are not incorporated into most computational and statistical analyses. Despite the popularity and widespread use of enrichment-based methods, we believe that the incompleteness in the underlying structured gene annotations is underappreciated.

This incompleteness of gene annotations is, at least in part, due to inefficiencies in the translation of scientific knowledge into structured annotations. There are over 21 million articles indexed in PubMed, but only ~73,000 articles are cited in support of GO annotations. Even though the sum total of biological knowledge is potentially quite rich, structured representations of that knowledge are quite sparse. Currently, we rely on a few large biocuration groups to translate all of the peer-reviewed literature into structured annotations. However, these centralized efforts simply cannot keep up with the rate of biomedical science. The biocuration community itself has noted that "*the exponential growth in the amount of biological data means that revolutionary measures are needed for data management, analysis and accessibility*" [9]. Since funding agencies like the NIH are unlikely to exponentially increase funding for biocuration efforts, new models for biocuration need to be explored.

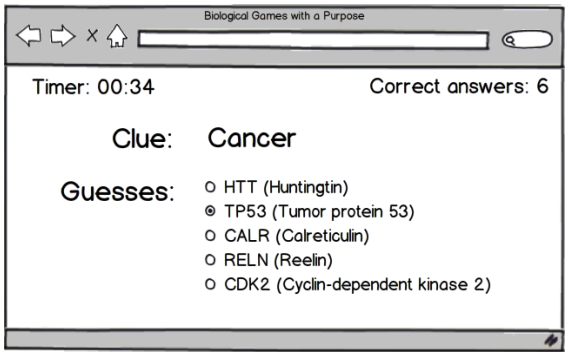

Figure 2. A one-player GWAP mockup for gene-disease annotations. Other "easy" examples for training include breast cancer/*BRCA1*, Alzheimer's Disease/*APP*, Huntington's disease/Huntington, and cystic fibrosis / *CFTR*. A crude prototype ("Dizeez") is described in the Preliminary Data section. This design also serves as the model for the first game in Specific Aim 2 to annotate gene-gene interactions. Training examples in that case could include *FOS/JUN*, *TNF/TNFR*, and *NFkB/IKB*.

Game summary. In this aim, we will build two online games that identify candidate novel gene annotations. We will first focus on gene-disease annotations since we have already built one very basic prototype (see the Preliminary Data section). These designs can then be easily generalized to gene-function annotations.

The first game will be a training game (**Figure 2**), a multiple choice quiz similar in design to "Trivial Pursuit". The player will be presented with a disease drawn from the Human Disease Ontology (the "Clue"). The player will also be shown a multiple-choice selector with five genes, only one of which has prior evidence linking it to the Clue disease. The player is asked to guess that gene. If "the Guess" is correct, the player is given points. The player attempts to accumulate as many points as possible in a fixed amount of time.

This training game serves a dual purpose. First, this game presents the user with relatively easy puzzles, enabling players to quickly accumulate points, reinforcing the positive and fun rewards of game play. Second, this game enables us to screen out players who have insufficient knowledge. If a player cannot easily pass these training tests, then it is likely they will not have sufficient knowledge to meaningfully contribute and their data can be ignored in data analyses.

The second game will be modeled after the successful ESP Game. Two players will be randomly and anonymously paired, and both players will be shown a disease (the "Clue") (**Figure 3**). Instead of a multiple choice selector, players are shown a free-text input box. Each player is instructed to type genes (the

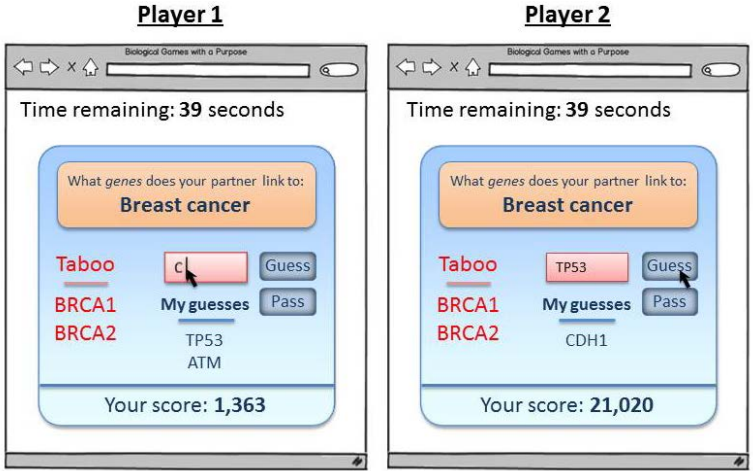

Figure 3. A two-player GWAP mockup for gene-disease annotations. Two independent players are shown a disease (the "Clue") and submit genes related to that disease as Guesses. Points are awarded when the players match on a Guess. As players win more points, the Clues test increasingly specialized knowledge. Many variants of this game are possible. Instead of a disease, the clue could also be a biological pathway (see "Game variants"), or a gene (see the first game of **Specific Aim 2**). Note that even though each player always appears to be playing in real time with another user, a realistic playing partner "script" can be retrieved from a past game with the same clue, allowing two people to play together asynchronously. The role of the Taboo list is described in the Game variants section.

"Guesses") that are related to the Clue disease. When one of the Guesses is named by both players, then the players "win" that round and receive points. Players are then shown a new Clue disease, and players attempt to win as many points as possible in a fixed amount of time.

A key feature of fun games is tailoring the difficulty of gameplay to the player's skill. We will start with games that test common knowledge that virtually all game players possess (see the examples in **Figure 2**) and become increasingly difficult. Once players have passed our introductory levels, the game will also allow players to select a specific area of biology (for example, by disease or protein family) that best matches their expertise (**Figure 4**). Alternatively, users may enter a PubMed search (on their name, for example) from which the system would infer the player's area of expertise. Within each biological domain, users will advance through progressively more difficult levels, where harder levels test less well-studied diseases.

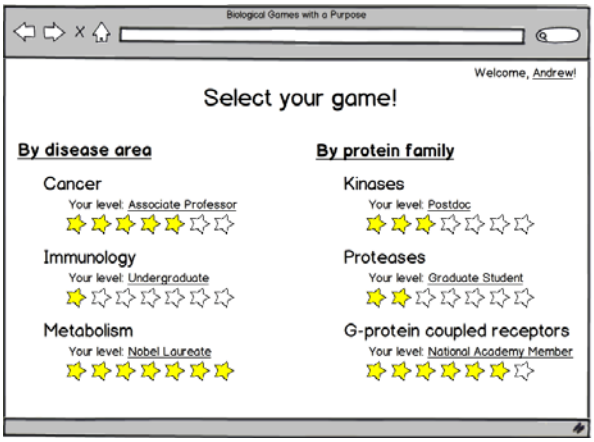

Figure 4. A mockup of an interface to allow users to select games and track progress in specific biological subdomains. For example, the user depicted above has advanced to the highest level in "Metabolism", but is only a beginner in "Immunology". Questions get progressively more difficult as the player advances through levels. This type of interface would apply to all games in this proposal.

To align this aim with our DBP, games will initially be developed with a specific focus on improving annotations in the area of post-transplant immunity. Dr. Salomon recently published two studies of chronic and acute rejection of kidney transplants [11, 12]. Although each study revealed candidate genes involved in transplant rejection, we will work with Dr. Salomon to improve gene annotations in those lists and in pathways related to transplantation immunology, thereby enabling a more comprehensive pathway-level analysis of those data.

Game variants. There are several variants that can further stimulate game play and increase the value of the game playing data. We briefly discuss several variants here (though full discussions are not possible given space constraints). As we design and test game prototypes, these variants will be explored to determine which most effectively capture novel candidate gene annotations.

- As mentioned previously, this game design can also identify gene-function annotations by using pathways and biological processes from the Gene Ontology as the Clue instead of diseases.
- When specific gene-disease annotations are well established and validated, that gene could be added to a "Taboo" list for that disease, prohibiting players from using that gene to match.
- The entity types for Clues and Guesses could be reversed. For example, genes could be presented as clues and diseases would be supplied as Guesses.
- In two-player "output agreement games", both players have the same task. We may also use an "Inversion-problem" game design [5], similar to common party games Taboo and Pictionary. In this model, Player 1's Clue is a disease, and Player 1's Guesses are genes related to the Clue. Player 2's Clues are Player 1's Guesses, and Player 2's Guesses attempt to name the disease in Player 1's Clue.
- For advanced users, the point scoring system can be amended to reward explicit evidence for an asserted annotation. For example, at the end of a round, additional points could be given to players who supply a PubMed ID for a novel assertion.

Evaluation. Data from these games can be mined for candidate novel annotations in the same way that the ESP Game results were used to label images. Every player Guess in the games above can be interpreted as an assertion of a putative gene annotation between the Clue (disease or pathway) and the Guesses (genes). To enrich for high-quality candidate annotations, we will employ several meta-analysis strategies. Most importantly, candidate annotations that are independently reported across multiple players will obtain the highest confidence scores. This strategy utilizes the value of independent replication, a property that is used extensively to improve results in related crowdsourcing initiatives [3]. Replication has even been shown to improve results from professional curators [13] but is seldom used as part of biocuration pipelines due to cost. Prioritization of candidate annotations can also be achieved by a player skill assessment (based, for example, on their ability to recall known annotations and to pass the training game).

Recognizing the important role of expert curators, we will also partner with the biocuration community to formally evaluate our high-scoring candidate annotations. First, we will partner with Doug Howe to assess gene-function candidate annotations. Dr. Howe is the Data Curation Manager for ZFIN, a member of the Gene Ontology Consortium. Second, we will partner with Warren Kibbe and Lynn Schriml to evaluate candidate gene-disease associations. Drs. Kibbe and Schriml are the PIs on the Human Disease Ontology project and have done extensive work mining annotations from the literature. (See the Letters of Support.) These interactions extend our published work assessing candidate novel annotations mined from the Gene Wiki [14].

Scientific gaming data will be complementary to, and not competitive with, expert curation efforts. For example, we recently demonstrated how crowdsourced gene annotations from our Gene Wiki effort, while not perfect, significantly improved the statistical power of gene set enrichment analyses [14]. Our goal is to leverage these collaborations to improve the efficiency with which scientific knowledge is represented in structured annotation databases.

Specific Aim #2: Use biological games to create a "community consensus network" of gene relationships. The resulting network of biological knowledge will define both the existence of gene-gene edges, as well as the specific nature of each gene-gene relationship.

Background. Over many decades, biological research has been incredibly successful studying genes (and their protein products) according to their individual functions. However, as genome-scale methods are now commonplace, the study of gene and protein function in the context of networks (or biological "systems") is also increasingly common. The integration, analysis, and visualization of these biological networks have produced many insights on how biological systems work and continue to be areas of active research. (For simplicity, we use the term "gene network" to encompass a broader umbrella where nodes also include mRNA and protein products. This simplified view will be sufficient for many current enrichment-based analyses, and these distinct entities can later be teased apart using similar games.)

Systems biology analyses often map experimental data on to reference gene networks, for example, to look for subnetworks that are enriched for differentially expressed genes. Most commonly, these reference gene networks are derived from large, high-throughput data sets like physical protein-protein interactions (PPI) or gene expression correlation. While these networks have proved to be useful, they also clearly have limitations due to substantial false-positive rates and lack of temporal and anatomical resolution. Perhaps equally importantly, the edge types in these existing networks typically reflect an experimental method, not a functional biological relationship. A network of functional relationships would be a useful scaffold that would be complementary (if not superior) for the integrative analyses described above. Although the STRING database attempts to mine functional relationships from the literature [2], the networks are far from complete. Again, an expert user can easily identify edges in the gene network that are not currently represented in any structured database (**Figure 1**).

Game summary. In this aim, we will use biological games to elucidate a **"community consensus network" (CCN)** that is a reflection of current biological knowledge. Existing networks are largely organized around the specific experimental or analytical method used to generate edges. In contrast, our CCN will focus on functional relationships regardless of what specific method was used to describe them. Edges in this network will be based on both high-throughput studies and single hypothesis-driven experiments. In addition to describing the *existence* of specific gene-gene edges, this aim will also accurately describe the *functional nature* of each gene-gene relationship. This gene network will be a more accurate and more useful scaffold on which systems biology analyses can be based. This aim will again be accomplished in a series of two games.

In the first game, we will assemble the structure of our CCN using an online, two-player game that is very similar to the second game in **Specific Aim 1** (**Figure 3**). Instead of a disease as the Clue, players will be shown a gene name. Each player submits Guesses in the form of other genes that functionally interact with the Clue gene. Like the previous game, points are awarded when the players match on a Guess, and players attempt to accumulate as many points as possible in a fixed amount of time. The definition of a "functional interaction" at this phase is completely up to the game player's interpretation.

In the second game, we will functionally characterize the relationship types of gene-gene edges. (We intend to use the CCN from the first game as a starting point, though any other gene network could be used as well.) For example, suppose a strong link exists between *TP53* and *MDM2* in the CCN, but the specific nature of that relationship is unclear. This second game again employs the same basic game design, except here the *TP53* - *MDM2* gene pair is the Clue (**Figure 5**). Players offer a Guess as to the nature of the relationship that best describes that gene pair. As before, if the two players' Guesses match then they earn points.

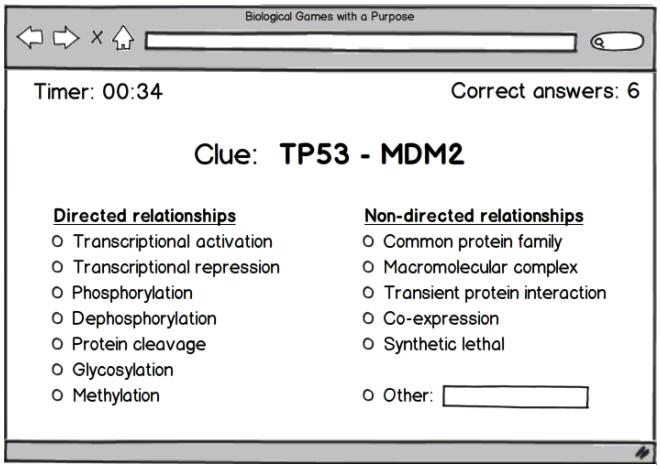

Figure 5. GWAP for labeling gene-gene edges in the CCN. Note that multiple correct answers can exist - the players are asked to select the relationship that they think best characterizes the link. Whether or not the link matches their partner's guess only determines whether points are awarded. All assertions about the nature of the gene-gene relationship will be recorded and tallied.

Players will select a Guess via a multiple-choice selector based on an ontology of relationship types. There is currently no universally accepted standard ontology for genetic relationships, although basic ontologies have been proposed [15]. Unless one universal standard emerges, we will define a simple ontology of well-known relationship types (**Figure 5**). In addition, users will be allowed to enter a new relationship type in a free-text entry box, and commonly-suggested relationships will be added to the ontology.

To align with our DBP, games will initially be focused on the CCN around genes related to post-transplant rejection [11, 12]. A better understanding of the functional gene network surrounding the genes that Dr. Salomon previously identified will aid the translation of those results to pathway-level testable hypotheses.

Game variants. Many game variants can be explored to stimulate game play and increase the value of the resulting data. Among them:

- Several of the variants from **Specific Aim 1** can also be applied to these games. Specifically, see the previous descriptions for "Taboo" list, "Inversion-problem" game design, and reward explicit evidence.
- Tune difficulty level and maximize fun by allowing players to select a specific area of biology (for example, by disease area or protein family) that best matches their expertise (**Figure 4**)
- Include and disambiguate related biomolecules (e.g., DNA, mRNA, ncRNAs, metabolites, proteins, drugs)
- Create additional games to resolve the temporal and anatomical contexts of network edges

Evaluation. The data from these games will be mined using the same principles as in **Specific Aim 1**. Every Guess given by the players represents an assertion that two genes share a functional relationship (in the first game), or a particular relationship type defines a gene-gene edge (in the second game). Aggregated over game play by many users, assertions that are highly replicated become more confident properties of our CCN. This network can then be utilized and visualized using existing network tools (e.g., Cytoscape). In addition, we expect new tools to be developed that can take advantage of the explicit typing of gene relationships and the quantitative confidence scores.

To assess the game-derived CCN, we will partner with experts in the field of network biology. Lars Juhl Jensen, the principal scientist behind the STRING protein interaction database, and Alex Pico, Executive Director of the National Resource for Network Biology (nrnb.org), will help evaluate the accuracy and precision of candidate gene-gene interactions (see Letters of Support).

Specific Aim #3: Use biological games to discover biomarkers of human disease. These games will identify molecular features for studying disease mechanisms, diagnosis, prognosis, and treatment selection.

Background. Identifying biomarkers for human disease is a top priority for biomedical research. Broadly defined, a biomarker is any biological molecule that serves as an indicator of biomedical state. While most

often used to help select an appropriate treatment strategy, biomarkers can also be used to molecularly characterize a patient's disease status, stratify patients into disease subclasses, or to predict a patient's prognosis [16]. In many cases, biomarker discovery can lead to new strategies to investigate disease mechanisms. Using high-throughput technologies, patient samples can be molecularly profiled for a large and rapidly expanding set of potential biomarkers. In total, a single biological sample could reasonably be described by millions of molecular variations in DNA, RNA, proteins, and metabolites. The process of picking which of these variables best serve as indicators of biological state is referred to as "feature selection" [17].

When faced with millions of possible features, feature selection by purely statistical methods has not been a particularly successful approach to biomarker discovery. One key issue in defining biomarker panels is combinatorial explosion. Consider that exhaustively testing even all 4-gene biomarker panels from a typical gene expression microarray alone results in 10^{18} tests. Brute force methods for testing these combinations lead to both computational challenges (CPU load) and statistical challenges (multiple testing, overfitting). Purely statistical approaches have been effective in some cases (e.g., [18, 19]), but such success is not the norm. Many such "gene expression signatures" have failed to replicate in subsequent validation studies [20]. While new feature selection techniques are continually emerging, no automated method consistently delivers good predictive models based on high dimensional biological data.

Not surprisingly, it is increasingly clear that human knowledge is an essential component to defining biomarker panels (e.g., [21-23]). Knowledge-based approaches offer many advantages relative to feature selection based on statistics alone. Because they are based on biological insight, they are more likely to generalize to external validation sets and less likely to overfit to training data. Because they bypass the combinatorial nature of brute force methods, they have lower computational burden and reduced multiple testing. They are also more interpretable in terms of their potential mechanistic role. The main disadvantage of knowledge-driven approaches is the sheer amount of human effort needed to organize the relevant information and data. However, harnessing large amounts of human capital is exactly the strength of crowdsourcing through games.

Game summary. In this aim, we will build a single game to identify biomarker panels for determining disease diagnosis and prognosis. This game, like Foldit, will be a puzzle-solving strategy game. Instead of folding protein structures, game players will use their language comprehension and problem solving skills to select biomarkers for a specific disease. Aggregated over the entire community of game players, we hypothesize that the "wisdom of the crowds" will produce biomarker panels of equal or higher quality than currently exist. Moreover, there is increasing evidence that "interactive machine learning" can achieve impressive results in classification problems, even when used by non-experts [24].

Relative to the two previous aims, the biomedical objective here is much more challenging. Indeed, deriving genetic and genomic biomarkers for diagnosis and prognosis is at the cutting edge of biomedical science and personalized medicine. Translating this scientific challenge into a fun and engaging game is therefore also much more challenging. So while we are confident in following the ESP Game model for the previous aims, **Specific Aim 3** will require spending the first 6-12 months on creating and refining a specific game design. As the Foldit authors emphasized [25], an iterative approach is vital to success for complex scientific games.

Nevertheless, we can describe some general game principles. Players will be presented with virtual patients and challenged to predict phenotype based on a list of biomedical features. For example, the simplest training level of this game may involve identifying which of 10 virtual patients has lung cancer, where the available features include the patient's height, eye color, smoking status, shoe size and blood pressure. The player would "win" the level if they identified smoking status as the best predictor of lung cancer. In subsequent training levels, players could be introduced to genetic features through Mendelian diseases. For example, another game may challenge a player to select the patients suffering from cystic fibrosis based on mutation data. The player would win if they chose the *CFTR* gene as the best predictor. As the player becomes more and more proficient at solving these known puzzles, they would eventually unlock the entire suite of genetic and genomic features for each patient.

How will game players be able to evaluate genetic and genomic features in terms of diseases? Lots of biomedical knowledge is accessible through news articles, blogs, Wikipedia, or even PubMed Central. Players

could access those biomedical texts either through tools provided within the game environment or through their computer's web browser. Gameplay will leverage the players' language comprehension skills, an area where humans still vastly outperform computers. The training phase of the game would involve increasingly complex challenges, starting with the link between cystic fibrosis and *CFTR*, to breast cancer and *BRCA1/2*, to diabetes and adiponectin. Training would also transition the player from single-gene biomarker to multi-gene biomarker panels. And again, interpretation of biomedical text does not have to be perfect given our ability to leverage replication across players, as described in **Specific Aim 1**.

At the most advanced levels, patient data will be divided into separate training and test sets. Players will develop their biomarker panels with full access to the training data, while the test data will be blinded to players and be used solely for scoring players' input. This system helps to prevent players from building biomarker panels that overfit the training data.

Following the Foldit model, we will incorporate social features that allow players to collaborate on designing biomarker panels. Foldit provided players a simple interface to build protein folding workflows (called "recipes"), which then could be shared and extended by other players. Using this system, over 500 players collectively wrote over 5,000 folding recipes [6]. In our biomarker discovery game, we intend to build a similar interface for online collaboration. For example, players will be able to directly extend others' biomarker panels. Alternatively, players might develop biomarkers for intermediate biological processes (e.g., T-cell activation, cell migration), which can then be utilized as meta-features for a variety of disease challenges.

To reiterate, these general principles will be translated into specific game mechanics during an initial 6-12 month design period. More generally, we offer three reasons why we are confident in this game design process and our ambitious goal of discovering biomarker panels for human disease.

First, the human effort that can be tapped has enormous potential. We previously described the massive scale of gaming effort (150 billion human hours per year globally). It is also notable that this gaming community is broad and diverse. Industry groups report that 69% of households play computer or video games. Female players account for 40% of all gamers, and a quarter of gamers are over the age of 50 [1]. Among the top contributors to Foldit, only ~15% were employed in science, and the largest sector came from the business/finance/legal sector (~20%). Approximately two-thirds of top players had no more than a high school level background in biochemistry. This unique group of gamer-scientists is made up of highly experienced problem solvers with extremely diverse backgrounds - an ideal recipe for creative group problem solving.

Second, the Foldit game proved that gamers can contribute to projects that are conceptually and intellectual complex. Most crowdsourcing efforts (like Galaxy Zoo and the ESP Game) decompose one massive project into a very large number of simple tasks, which are then distributed among a very large population of participants. However, the challenge of protein folding, like biomarker discovery, is neither simple nor easily decomposable. Foldit demonstrated for the first time that even conceptually and intellectually complex challenges can be addressed by gamers. In doing so, Foldit used crowdsourcing in a fundamentally new way. Foldit built a large community of average players as a mechanism to find the few exceptionally talented players. These were the individuals who then made tangible contributions to the complex scientific goals (for example, by exploring a new conformational space or writing a new folding recipe).

Third, our team includes experts in game design and biostatistics. Dr. Zoran Popović leads the development of the Foldit game as a Professor at the University of Washington. Mr. Josh Peay has a long history in the video game industry including a decade at Sony Computer Entertainment. Both have a firm understanding of video game mechanics, building engaging games, and leveraging mobile devices as gaming platforms. In addition, Dr. Nicholas Schork, an international expert in biostatistics and human genetics, will help ensure that the feature selection approach developed in our game accurately reflects real-world biomedical applications. The diverse and experienced background of this team provides confidence that the general ideas outlined above will result in a successful game design within the 6-12 month design period. (See Letters of Support).

Of this proposal's three aims, **Specific Aim 3** clearly has the highest risk/reward potential. The high reward stems from the close relation between the game output (biomarkers) and biomedical applicability, and also

from the incredibly large community of potential contributors. The high risk stems from the desire to engage the general public, not just professional biomedical researchers. Yet we are confident in this aim because people want to be engaged in science. When scientists challenged ordinary people to help with classifying astronomical images (Galaxy Zoo), over 100,000 people participated. When scientists wrapped the complex challenge of protein folding in a game, 300,000 people signed up. Given the nearly universal interest in human health and disease, we are confident that our games will receive a similar response.

Evaluation. Games will initially be developed in collaboration with Co-PI Dr. Salomon to differentiate kidney transplants with good and poor outcomes. Dr. Salomon has a large cohort of kidney and liver transplant patients with extensive clinical and genomic characterization (see Letter of Support and [11, 12]). Patient data will be separated into a fully-accessible training set and a blinded external test set that will be used to score the players' biomarker panel. In addition, we will work with Dr. Salomon to evaluate the most successful player-generated biomarker panels for mechanistic, diagnostic, or prognostic utility. This game design can then be extrapolated to any disease for which we have access to both clinical readouts and molecular profiles. For example, many suitable data sets are available from public repositories like the Gene Expression Omnibus (GEO) and dbGAP for phenotypes as diverse as breast cancer prognosis, heart disease and body mass index.

In addition to validation using published data, we will also participate in challenges in the context of the **Critical Assessment of Genome Interpretation (CAGI)**. CAGI organizers obtain data sets on a pre-publication basis, and challenge participants to predict phenotype based on molecular profiles. For example, in the last CAGI assessment, patients were provided exome sequencing data from 56 individuals, and participants were challenged to predict which suffered from Crohn's disease. Following Foldit's successful participation in the similar Critical Assessment for Structure Prediction (CASP), we will use CAGI's periodic blinded competitions to both motivate player participation and assess performance.

COMMON THEMES FOR ALL AIMS

Dissemination and outreach. Providing data to the research community in a useful form is critical for this proposal's success. We will provide all data in raw and processed form via existing standard file formats (e.g., GAF 2.0, XGMML, RDF) for maximal compatibility with bioinformatics tools and pipelines. We will also work with the NRNB to disseminate the CCN to their user community (see Letter of Support from Alex Pico).

The success of these aims depends on developing a robust community of gamers. To facilitate this outreach, we will publicize our games through word of mouth, through social media, and of course, through scientific publications. As the games become more robust and mature, we also will actively seek out contributions in targeted scientific areas. For example, establishing informal and fun competitions at the annual meetings of scientific societies would be an ideal way to stimulate game play. Although we have not emphasized it here, there are also many educational applications of these games that will encourage learning by gaming.

Preliminary Data. We recently released a game called "**Dizeez**" (http://sulab.scripps.edu/dizeez/), to identify gene-disease links. Dizeez follows a design similar to the first game of **Specific Aim 1**. Players are shown one gene name and five disease names, and the player tries to guess the disease to which the gene has already been linked.

We released Dizeez to the community, publicizing its existence only through our lab blog and twitter account. Within one week, over 350 games had been played by over 100 unique individuals. Summed over all games, players provided 3069 guesses that spanned 2762 unique gene-disease assertions. We focused our analysis on the 139 gene-disease assertions that were provided by game players more than once and that were not previously found in OMIM or PharmGKB. Six of these gene-disease assertions were asserted in four or more distinct games, and we found strong evidence for five of those six assertions in the literature (**Table 1**).

Gene Symbol (ID)	Disease Name (DOID)	Validated
DHH (50846)	polyneuropathy (DOID:1389)	yes
WBSCR22 (114049)	Williams syndrome (DOID:1928)	---
SOX8 (30812)	mental retardation (DOID:1059)	yes
ITGAL (3683)	leukocyte adhesion deficiency (DOID:6612)	yes
TG (7038)	Graves' disease (DOID:12361)	yes
BCL10 (8915)	lymphoma (DOID:353)	yes

Table 1. Candidate gene-disease annotations mined from the Dizeez prototype game.

These results demonstrate that even from limited game play, biological insights can be mined from game playing logs. Dizeez is clearly only a crude prototype of one the games we are intending to build. The quantity and quality of data will undoubtedly increase as the proposed features are added. Nevertheless, Dizeez received an enthusiastic response based on only limited effort building and promoting the game. This preliminary work provided strong evidence that our proposed games will translate to useful biological data.

Conclusion. We recognize that the methods described in this proposal are sufficiently unconventional as to invite skepticism from many different angles. We have heard and discussed many of these criticisms with our colleagues spanning expertise in bioinformatics, network biology, drug discovery, immunology, metabolism, oncology, and neurobiology. Combined with peer-review comments of grants and manuscripts on similar topics, we have summarized the most common criticisms about crowdsourcing and games in biology.

Criticism #1: Games will generate low quality data. The strongest counterargument is the overwhelming success of the Foldit game. By framing protein folding in the context of a game, even non-scientists were able play and contribute to several successes, both in terms of biomedical discovery and algorithm design. In addition, we have several mechanisms built into our game designs to ensure high quality. First, replication (across many games and many players) has been shown to outperform individual "experts". Second, player scores can be used to roughly quantify user knowledge, which in turn can be used to filter game playing data. Third, we will collaborate with expert biocurators to develop an appropriate annotation evaluation pipeline.

Criticism #2: Only visual game designs will work. Both the ESP Game and Foldit leverage game players' visual processing capabilities. While visual processing is undoubtedly one area where humans outperform computers, it is not the *only* skill for which that is true. The games proposed here largely leverage humans' ability to read, interpret, and *understand* language. The players' abilities to search for and understand biomedical information far exceeds computational natural language processing, especially since the Internet offers so much relevant text. **Specific Aim 3** also leverages humans' superior abilities at puzzle solving in much the same way that the Foldit "recipe" scripting system does.

Criticism #3: Nobody will have time to play these games. First, people are already playing games. The average person spends only one-third of their waking hours "at work", and gaming is a popular activity for the remaining "cognitive surplus". Second, given the sheer number of gamers, only a small percentage of potential players are needed to generate a substantial amount of data to be mined (especially relative to the number of professional biocurators, which likely number in the hundreds). Third, we believe that these games will harness the bits of lost time during the day (e.g., commuting, between experiments or meetings, during mental breaks) by offering an outlet that is more fun and more productive than other options like Solitaire and Facebook. For reference, Scripps (a biomedical research institute with 2,500 employees) already generates ~3 gigabits of network traffic to Facebook on the average day.

Criticism #4: Game players will not want to do "work" for fun. There is a common perception that gamers only play trivially "mindless" games. In truth, many gamers crave difficult mental challenges and deep engagement in their games. Consider **World of Warcraft ("WoW")**, a popular online role-playing game. In the complex virtual world of WoW, players can study a profession, train for new skills, collaborate with others, and of course, do battle with the enemy on quests. They also created the second largest wiki ("WoWWiki") in the world behind Wikipedia to share game notes and strategies. At its peak, Nielsen reported that the average player was spending over 17 hours a week playing WoW, almost the equivalent of a half-time job. Collectively, they have spent over 5.9 million human *years* playing this "job" [1]. WoW players spend considerable free time and effort because they connect with the "epic missions" in the game. Imagine how motivated they would be if we could challenge them with similarly epic missions to impact human health *in real life*.

In summary, addressing the challenge of a million-dollar genome interpretation requires a solid foundation of biological knowledge. We have described three aims that target key biomedical challenges that require human effort on a massive scale. Harnessing community effort through games represents an untapped and powerful human resource to address those challenges. We believe we have outlined an effective approach and assembled an ideal team of investigators to explore this innovative approach to genome-scale science.

APPROPRIATENESS FOR THE TRANSFORMATIVE RESEARCH AWARDS INITIATIVE

This proposal is perfectly suited to the TR01 program for several reasons. First, the aims address fundamental challenges in genome-scale biomedical science. The sum total of biological knowledge is growing exponentially, as is the volume of data that can be generated in any one experiment. This proposal challenges the conventional thinking that manual expert curation is the only mechanism to produce structured gene annotations. The NIH spends tens of millions of dollars per year on biocuration even as the biocuration community recognizes that the current model does not scale with the rate of science. It also addresses a fundamental challenge in biomedical science -- the discovery of biomarkers for studying disease mechanisms, diagnosis, prognosis, and treatment selection.

Second, the use of crowdsourcing to address massive challenges in genetics and genomics is innovative. This proposal to harness the power of crowdsourcing builds on recent trends in society, the media, and the Internet. Crowdsourcing has also claimed several recent successes in science, including the Gene Wiki project developed by our group. Using games as a mechanism for crowdsourcing has only seriously been tried once in biology (Foldit), and this proposal builds on that success to address modern challenges in genetics and genomics. Given our diverse backgrounds, we feel that this team of investigators is uniquely qualified to explore this area of science.

Third, successful completion of the aims will have a huge potential impact on biomedical science. The size of the game player population is simply enormous. The world spends 150 billion hours worldwide playing games every year, all in the selfish pursuit of fun. It is up to the scientific community to align our most daunting challenges with that pursuit of fun, thereby unlocking a human resource with almost limitless potential. The games in this proposal harness that effort to support and complement the professional research community in advancing biomedical science.

TIMELINE

This proposal will require funding for five years (**Table 2**). The games for each Specific Aim will follow a three-step development process of design, implementation, and data mining. For **Specific Aim 1** and **Specific Aim 2**, we will follow existing and successful game designs (the ESP Game). Therefore, we will immediately begin an initial implementation of these games. As mentioned for **Specific Aim 3**, the majority of the first year will be spent working out the exact game mechanics for biomarker selection. This process will be done within the development team and with the participation of our two key game design consultants, Professor Zoran Popović and Mr. Josh Peay. As emphasized previously, game design is a highly iterative process [25]. Therefore, even after the initial release, all games will undergo continual redesigns to explore game variants.

The implementation phase will be executed by a four person development team, led by project leader Dr. Benjamin Good and technical leader Dr. Chunlei Wu. This four person team is a minimal and ideal size for an agile game development effort. We expect initial versions of the games in **Specific Aim 1** and **Specific Aim 2** to be complete within the first year, and an initial version of the game in **Specific Aim 3** to be complete by the end of Year 2. Implementation will continue through the course of the grant as new game designs are conceived and tested. Games will initially be released online and then later adapted to mobile devices.

	Year				
	1	2	3	4	5
Specific Aim 1 - Gene annotation games					
Iterative game design	x	x	x		
Implement game	x	x	x	x	
Data mining & dissemination		x	x	x	x
Specific Aim 2 - Gene network games					
Iterative game design	x	x	x	x	
Implement game	x	x	x	x	
Data mining & dissemination		x	x	x	x
Specific Aim 3 - Disease predictor game					
Iterative game design	x	x	x	x	x
Implement game		x	x	x	x
Data mining & dissemination			x	x	x

Table 2. Timeline for project completion.

Finally, the data mining phase will be led by Dr. Good in collaboration with our many expert partners, including Dr. Dan Salomon. This phase will focus on mining biologically useful data out of gaming logs. Moreover, this phase will also include significant effort devoted to outreach in order to reach as many game players as possible (including at scientific conferences).

REFERENCES

1. McGonigal J. **Reality is broken : why games make us better and how they can change the world**. New York: *Penguin Press*; 2011.
2. Szklarczyk D, Franceschini A, Kuhn M, Simonovic M, Roth A, Minguez P, et al. **The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored**. *Nucleic Acids Res*. 2011;**39**(Database issue):D561-8. (PMC3013807)
3. Lintott CJ, Schawinski K, Slosar A, Land K, Bamford S, Thomas D, et al. **Galaxy Zoo: morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey★**. *Monthly Notices of the Royal Astronomical Society*. 2008;**389**(3):1179-89.
4. Raddick MJ, Bracey G, Gay PL, Lintott CJ, Murray P, Schawinski K, et al. **Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers**. *Astronomy Education Review*. 2010;**9**(1):010103-18.
5. von Ahn L, Dabbish L. **Designing games with a purpose**. *Commun ACM*. 2008;**51**:58-67.
6. Good BM, Su AI. **Games with a scientific purpose**. *Genome Biol*. 2011;**12**(12):135.
7. Shirky C. **Cognitive surplus: Creativity and generosity in a connected age**. New York: *The Penguin Press*; 2010.
8. von Ahn L. **Google Tech Talk: Human Computation** 2006. Available from: http://www.youtube.com/watch?v=qlzM3zcd-lk&t=6m38s.
9. Howe D, Costanzo M, Fey P, Gojobori T, Hannick L, Hide W, et al. **Big data: The future of biocuration**. *Nature*. 2008;**455**(7209):47-50. (PMC2819144)
10. Huang da W, Sherman BT, Lempicki RA. **Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists**. *Nucleic Acids Res*. 2009;**37**(1):1-13. (PMC2615629)
11. Grigoryev YA, Kurian SM, Avnur Z, Borie D, Deng J, Campbell D, et al. **Deconvoluting post-transplant immunity: cell subset-specific mapping reveals pathways for activation and expansion of memory T, monocytes and B cells**. *PLoS One*. 2010;**5**(10):e13358. (PMC2954794)
12. Kurian SM, Heilman R, Mondala TS, Nakorchevsky A, Hewel JA, Campbell D, et al. **Biomarkers for early and late stage chronic allograft nephropathy by proteogenomic profiling of peripheral blood**. *PLoS One*. 2009;**4**(7):e6212. (PMC2703807)
13. Camon EB, Barrell DG, Dimmer EC, Lee V, Magrane M, Maslen J, et al. **An evaluation of GO annotation retrieval for BioCreAtIvE and GOA**. *BMC Bioinformatics*. 2005;**6 Suppl 1**:S17. (PMC1869009)
14. Good BM, Howe DG, Lin SM, Kibbe WA, Su AI. **Mining the Gene Wiki for functional genomic knowledge**. *BMC Genomics*. 2011;**12**(1):603.
15. Lu LJ, Sboner A, Huang YJ, Lu HX, Gianoulis TA, Yip KY, et al. **Comparing classical pathways and modern networks: towards the development of an edge ontology**. *Trends Biochem Sci*. 2007;**32**(7):320-31.
16. Bhattacharya S, Mariani TJ. **Array of hope: expression profiling identifies disease biomarkers and mechanism**. *Biochemical Society transactions*. 2009;**37**(Pt 4):855-62.
17. Guyon I, Elisseeff A. **An introduction to variable and feature selection**. *J Mach Learn Res*. 2003;**3**:1157-82.
18. Buyse M, Loi S, van't Veer L, Viale G, Delorenzi M, Glas AM, et al. **Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer**. *J Natl Cancer Inst*. 2006;**98**(17):1183-92.
19. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, et al. **Gene expression profiling predicts clinical outcome of breast cancer**. *Nature*. 2002;**415**(6871):530-6.
20. Xu JZ, Wong CW. **Hunting for robust gene signature from cancer profiling data: sources of variability, different interpretations, and recent methodological developments**. *Cancer Lett*. 2010;**296**(1):9-16.
21. Cobleigh MA, Tabesh B, Bitterman P, Baker J, Cronin M, Liu ML, et al. **Tumor gene expression and prognosis in breast cancer patients with 10 or more positive lymph nodes**. *Clin Cancer Res*. 2005;**11**(24 Pt 1):8623-31.
22. Rosenberg S, Elashoff MR, Beineke P, Daniels SE, Wingrove JA, Tingley WG, et al. **Multicenter validation of the diagnostic accuracy of a blood-based gene expression test for assessing obstructive coronary artery disease in nondiabetic patients**. *Ann Intern Med*. 2010;**153**(7):425-34.
23. Su J, Yoon B-J, Dougherty ER. **Accurate and Reliable Cancer Classification Based on Probabilistic Inference of Pathway Activity**. *PLoS One*. 2009;**4**(12):e8161. (PMC2781165)
24. Stumpf S, Rajaram V, Li L, Wong W-K, Burnett M, Dietterich T, et al. **Interacting meaningfully with machine learning systems: Three experiments**. *International Journal of Human-Computer Studies*. 2009;**67**(8):639-62.
25. Cooper S, Treuille A, Barbero J, Leaver-Fay A, Tuite K, Khatib F, et al. **The challenge of designing scientific discovery games**. *Proceedings of the Fifth International Conference on the Foundations of Digital Games; Monterey, California*. 1822354: ACM; 2010. p. 40-7.