

# The new date, new format, new goals and new sponsor of the Archon Genomics X PRIZE Competition

Larry Kedes & Grant Company

The Archon Genomics X PRIZE presented by MEDCO is a \$10 million prize for whole human genome sequencing that will test 100 samples and establish a defined method for determining quality and cost per genome. The contest will entail a head-to-head competition starting on 3 January 2013. The \$10 million grand prize will be awarded to the first team to meet all the quality standards. If there is no grand prize winner, lesser awards will be made for single achievements in the three categories of accuracy, completeness and haplotype phasing. Similar to the open comment period offered earlier this year for the validation protocol, the X PRIZE Foundation seeks commentary from the scientific community on the revised competition format and judging criteria prior to publishing final rules as part of a master agreement with competing teams next year. The official validation protocol will make the competition judging transparent and fair using multiple standard techniques and robust bioinformatics. With sponsorship from Medco Health Solutions (MEDCO), the 100 samples will be derived from genomes of appropriately consented centenarian subjects. At the conclusion of the contest, the data from the redundantly sequenced genomes will be deposited into a scientific database and provide a heretofore unrealized depth of sequence data based on multiple technologies. The sample set and the bioinformatic pipeline will also be made available to the scientific community.

First publicly proposed in October 2006, the Archon Genomics X PRIZE was designed to encourage radical breakthroughs, incentivizing a revolution for whole human genome sequencing technologies to become faster, more accurate and less expensive than any existing at that time. Because of the unprecedented success of the developers of sequencing technology, we are now able to make the competition more ambitious, revising the rules and goals to be challenging and the aims of the contest to be robust, while still fulfilling the X PRIZE Foundation's goals of benefiting humanity. To that end, on 3 January 2013, teams from around the world will compete against each other assessed by our official validation protocol (Supplementary Methods)—a new quality standard for whole genomic sequencing. The resulting reduplicated sequences may truly define a 'medical-grade' genome.

*Larry Kedes and Grant Company are at the X PRIZE Foundation, Playa Vista, California, USA.*

A head-to-head competition will take place from 3 January to 3 February 2013. The \$10 million grand prize will be awarded to the team(s) able to sequence 100 human genomes within 30 days to an accuracy of 1 error per 1,000,000 bases, with 98% completeness, identification of insertions, deletions and rearrangements, and a complete haplotype, at an audited total cost of \$1,000 per genome. The complete rules for the Archon Genomics X PRIZE presented by MEDCO are available as a supplement to this Commentary (Supplementary Note) and at the competition web site: <http://genomics.xprize.org/>.

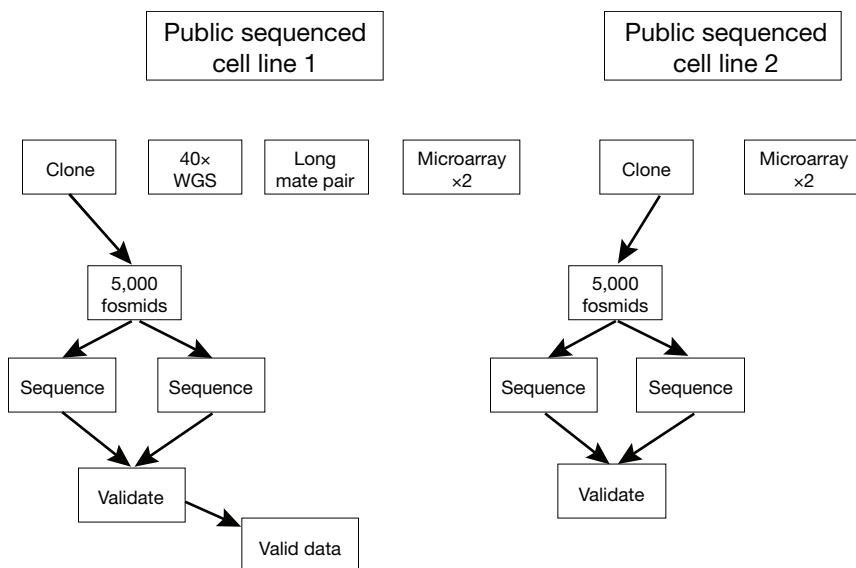
We changed the contest length from 10 to 30 days after discussions with potential competitors who felt the short timeframe was a barrier. In addition, dramatic reductions in advertised prices prompted us to lower the allowable per-genome cost accordingly. The X PRIZE Foundation may make additional changes to the rules as public commentary or circumstances dictate.

## Winning options

The Archon Genomics X PRIZE presented by MEDCO also wants to recognize and celebrate the achievements of technology developers

**Table 1 Summary of criteria for judging the prize contest**

Competition categories	Minimum requirements	Best-in-class requirements
Cost	\$10,000 per genome	\$1,000 per genome or less
Speed	30 days	First to submit; must be 30 days or less
Accuracy	1 in 100,000 error	No more than 1 error per 1,000,000 bases
Completeness	95%	98%
Haplotype phasing	0%	Completely phased chromosomes



**Figure 1** Archon Genomic X PRIZE Validation Protocol (VP) Flow Chart. DNA from two previously sequenced and publicly available cell lines will be subjected to various characterizations including whole-genome sequencing and assembly. The analyses will be integrated, discrepancies will be evaluated by Sanger sequencing and the final data set will be compared to the previously sequenced public genomes. The publicly available genomes, and open access to the scoring bioinformatics pipeline, will be made available upon application to any laboratory desiring to examine either the VP or their own competency in advance of the contest itself. Labels are as follows: 40x WGS: Illumina short-insert paired-end sequencing. Long mate pair: long-insert mate pairs will be sequenced using Life Technologies SOLiD technology at 40-fold coverage. Microarray: each genome will be examined in duplicate using an Affymetrix 6.0 format or better. 5,000 fosmids: a random selection of 5,000 fosmid clones from the genome will be subjected to NGS in duplicate and assembled. Sequence: current next-generation technologies (Illumina, SOLiD, or Pac Bio or a combination) will be used to assemble the genome after sequencing at a minimum of 40-fold coverage. Validate: Sanger sequencing of PCR products, both randomly selected and targeted on the basis of discrepancies. The Bioinformatics group will choose the validation targets by comparing WGS, fosmid and microarray data.

and users who might be able to meet some, but not all, of these grand challenges. Thus, if the grand prize is not awarded, the \$10 million prize purse will be split between three category prizes for accuracy, completeness and haplotype phasing (see Table 1). For example, teams whose costs per genome are as much as \$10,000 may compete for lesser 'best-in-class' prizes by satisfying just one of the criteria for accuracy, completeness or haplotyping while achieving minimum requirements in the other competition categories. The goal is to be as inclusive as possible, allowing teams engaged in developing or disseminating sequencing technology to demonstrate their capabilities. Each team achieving a best-in-class outcome will be recognized, but cash awards to these teams will be made only if there is not a grand prize winner.

### Judging the genomes

Over the past two years, with the help and guidance of numerous collaborators, the Archon Genomics X PRIZE presented by MEDCO developed a method to fairly judge the accuracy and completeness of a whole

human genome assembly submitted by teams competing for the \$10 million prize. We call this method the Archon Genomics X PRIZE validation protocol (Supplementary Methods; hereafter referred to as the VP). The VP represents the first attempt to provide a universally applicable method that is independent of the technology used to sequence and is highly scalable. The initial announcement of a proposed VP (ref. 1) and the subsequent comments it engendered were instrumental in the crafting of the final approach.

The original VP (ref. 2) and some subsequently suggested modifications<sup>3</sup> use an in-depth sampling technique, duplicate sequencing of randomly selected DNA segments by complementary next-generation sequencing techniques, verification of discrepancies by Sanger sequencing, detection of indels through selective paired-end technologies, and extensive, duplicative genotyping. These data will be integrated in a bioinformatics pipeline. The first step in implementing the VP will be to test it on two publicly available genomes (one male and one female). A sum-

mary of the methods to be used is shown in Figure 1. The genomes will be identified and the validation data and the software pipeline will be made available well in advance of the competition.

Opportunities for feedback and public discussion will be implemented. The ultimate goal of this method is to be able to declare winners of the competition through a transparent, fair and robust process without controversy.

The VP and its associated bioinformatics pipeline will be used for analyzing the 100 contest genomes. We will use a scoring matrix for determining whether contestant genomic assemblies either meet or exceed the required standards. We do not intend to compare the scores of winning submissions against one another: all submissions meeting or exceeding the contest standards will be declared eligible for awards.

The competition rules require every one of the 100 genomes sequenced to satisfy the minimum requirements for each competition category. Therefore, the use of a sampling technique to disqualify a team because the scoring matrix identified one or more genomes to be of lesser quality will indisputably be a correct disqualification. Of lesser concern is the remote risk that the public's post-contest analysis of a team's submitted sequences might reveal that the winning teams' unsampled segments were in significant error, because these segments were excluded by the validation protocol procedures.

### Transparent methodology

The Archon Genomics X Prize presented by MEDCO will assess the validation methodology, as well as the extensive bioinformatics pipeline needed for evaluation, on two public genomes (see Fig. 1). These materials will be available to all interested parties in the Fall of 2012. This may be of particular interest for teams considering registering for the competition (for their self-evaluation purposes). This approach is designed to make the validation and scoring methods as transparent as possible.

At the conclusion of the competition, we will make certain that the 100 test samples themselves, as well as the validation protocol software pipeline, are openly available to the scientific community. This will allow organizations to self-test the capabilities of new technologies or modifications of existing technologies well into the future.

It is our hope the validation protocol could become the industry-wide standard for self-evaluation. In addition, it may well become a model for standardization of technology and certification of proficiency by professional organizations.

### A 'medical-grade' genome

The 100 human genome samples to be sequenced in this competition will provide the research community with an unprecedented opportunity to analyze what will likely be the most deeply sequenced set of human genomes ever assembled. In addition, the redundant sequencing of the same set of genomes by multiple teams, using different technology platforms, will likely allow subsequent rapid corrections, generating a nearly error-free set of 100 human genomes.

Following the competition, all assembled genome sequences submitted by the teams will be deposited into a scientific database (dbGaP is currently being considered). Systematic errors inherent in any technology are likely to be complemented by other technologies. Discrepancies in both sequencing and genome assembly among competitors will be quickly identified, allowing follow-on experiments to rectify the differences as well as possibly clarify what led to the errors. Indeed, the outcome of such a large-scale approach will be close to the reality of 'medical-grade' genomes that could be used as models for clinical applications.

### Introducing the "MEDCO 100 Over 100"

To help to promote the global educational mission of the Archon Genomics X Prize, we partnered with Medco Health Solutions (MEDCO). MEDCO has become the presenting sponsor of the Archon Genomics X Prize and the title sponsor of the "MEDCO 100 Over 100" program designed to enroll 100 centenarians into providing the competition sample set. The Archon Genomics X Prize and MEDCO intend to develop innovative campaigns to increase public awareness and understanding of the future of medicine and, for the benefit of the educational mission, make heroes of our 100 genomic pioneers.

The choice of centenarians as the subject genomes was made after broad consultation. There are sound fundamental principles as to why the resulting data could lead to important scientific discovery<sup>4</sup>.

To best identify the ideal composition of the MEDCO 100 Over 100 sample set, The Archon Genomics X Prize sponsored a 1-day workshop in May 2011 at the National Institutes of Health in Bethesda, Maryland, USA under the generous hospitality of the National Institute of Aging (NIA) (the NIA's participation and

support does not imply endorsement by the NIA). The goals of the workshop were (i) to prioritize the characteristics of centenarians to be enrolled in the competition; (ii) to define the minimum and best practices for informed consent requirements for enrolled centenarians; (iii) to recognize scientific compromises that may have to be made to ensure a 100-genome sample set; (iv) to establish a minimum timeline for sample acquisition; and (v) to confirm the eventual sharing of contest data among sample contributors and scientific databases. The requirements for centenarians to be included in the MEDCO 100 Over 100 are detailed in **Box 1**.

### A realistic choice of centenarian genomes

The contest genomes will be donated from 100 individuals who have achieved 100 years of age. Scientists at several centenarian research groups eagerly embraced the opportunity to contribute Institutional Review Board protocol-approved samples for this purpose. Extensive phenotypic data on these individuals will also be available and will be released along with the sequences themselves, thus allowing potential use of the data to further search for explanations of exceptional longevity.

A working group of centenarian and genetics experts will advise the Archon Genomics X PRIZE and MEDCO as to the advisability of including each submitted sample as a component of the MEDCO 100 Over 100.

We recognize that accumulating 100 such subjects and samples may prove difficult. The Archon Genomics X PRIZE, MEDCO and their advisors need to determine realistically the likely makeup of the sample set so that educational and outreach programs can be appropriately presented.

The actual subjects might differ from the intended ideal in that some of the subjects may already be deceased. Age verification and the full consent required might be difficult or impossible to obtain from some collections outside North America, with the effect of reducing the international participation and affecting the ethnic and geographic diversity originally sought. Ideally each subject of the MEDCO 100 Over 100 will have consented in language that allows identification of the subject through name, age, ethnicity and geographic location, and, when available, medical and pharmaceutical history, photographs and social history. Some fraction of subjects may not be fully consented or willing to consent for such full phenotype revelation or for personal identification. For deceased subjects, informed consent by responsible next of kin or proxy will be considered on a case-by-case basis.

## BOX 1 SELECTING CENTENARIANS FOR SEQUENCING

The subjects will be age-verified centenarians (ideal age >105 years) with immortalized cell lines as the genome source, fully consented to allow for the inclusion of phenotypic information, medical and drug history, and in some cases willingness to be interviewed and filmed as part of the educational campaign.

Permission by informed consent from the subject or the designated proxy will also be obtained for each subject in order for the X PRIZE Foundation to use photographs and life stories of each individual to publicize the centenarian study.

In the case of deceased subjects, permission to obtain and use such information will be obtained from the designated proxy.

All subjects who are consented for the first time or are re-consented anew for the purpose of this competition must be demonstrated to be cognitively intact. A proxy can do this if cognitive function is impaired.

The consent procedures must satisfy each host institution, the Archon Genomics X PRIZE and MEDCO standards.

The subject will be ethnically diverse, represent both genders and be assembled internationally from as many parts of the globe as possible.

Established cell lines rather than whole blood samples will be used in order to meet the need for multiple samples for many contestants and for the scientific community after the competition. The National Institute of General Medical Sciences (NIGMS) Human Genetic Cell Repository at the Coriell Institute for Medical Research will immortalize fresh whole blood samples and will prepare a distribution stock of and perform quality-control measures on previously immortalized cell lines. All fresh blood samples and previously immortalized cell lines submitted to the NIGMS Repository for the use in the competition must comply with NIGMS Repository informed consent regulations. The competition cell lines must pass NIGMS Repository quality-control tests, including viability and absence of mycoplasma and other contaminants.

These immortalized cell lines will be maintained by the NIGMS Repository and will have consent to be released to legitimate scientific organizations for future analysis. Following the competition, interested, qualified investigators may obtain the samples from the NIGMS Repository upon completion of the NIGMS Repository Materials Transfer Agreement and Statement of Research Intent forms.

**To benefit humanity**

The working hypothesis is that centenarians carry genetic variations that avoid or repress common ailments; discovery of such variants may lead to the discovery of drug targets as well as disease and protective mechanisms. Although it is still a hypothesis, the goal is very egalitarian: that the data emerging from this competition be usable by investigators in the field to examine the health problems of ordinary people, thus amplifying the universal reverence for the elderly and the benefits of healthy aging.

By establishing a wide-ranging set of new goals, the Archon Genomics X PRIZE presented by MEDCO believes that its original concept of incentivizing a major transformation in whole-genome sequencing technologies can still be met while broadening the impact of the contest through discovery and

technology. Although most races can only have one winner, we believe that after this race, the competition will benefit an entire industry. The contestants will contribute to scientific knowledge and help prepare the groundwork for applied human medical genomics, thus paving the way for personalized medicine and medical breakthroughs to benefit all of us.

*Note: Supplementary information is available on the Nature Genetics website.*

**ACKNOWLEDGMENTS**

We thank the NIA and its director, Richard Hodes, for hosting our conference. Thomas Perls (Boston University Medical Center) and Nir Barzilai (Albert Einstein College of Medicine) have been instrumental in guiding the competition on centenarian selection criteria. We have been able to offset some of the costs of developing and implementing the validation protocol through generous support from the John Templeton Foundation and from the Presenting

Sponsorship of this PRIZE by MEDCO Health Solutions, Inc.

**COMPETING FINANCIAL INTERESTS**

The authors declare competing financial interests: details accompany the full-text HTML version of the paper at <http://www.nature.com/naturegenetics/>.

Reprints and permissions information is available online at <http://npg.nature.com/reprintsandpermissions>.

This paper is distributed under the terms of the Creative Commons Attribution Share Alike license cc-by-sa 3.0, and is freely available to all readers at <http://www.nature.com/naturegenetics/>.

1. Kedes, L. & Liu, E. *Nat. Genet.* **42**, 917–918 (2010).
2. Kedes, L., Sutton, G., Liu, E. & Jongeneel, V. Preprint available from *Nature Precedings* <<http://hdl.handle.net/10101/npre.2011.5731.1>> (2011).
3. Zaranek, A.W., Clegg, T., Vandewege, W. & Thakuria, J. Preprint available from *Nature Precedings* <<http://dx.doi.org/10.1038/npre.2011.5766.1>> (2011).
4. Perls, T. & Barzilai, N. Preprint available from *Nature Precedings* <<http://dx.doi.org/10.1038/npre.2011.5756.1>> (2011).