LETTER

No bull: Upholding community standards in public sharing of biological datasets

We write as representatives of a multi-institutional team working since 1997 to sequence the Y chromosomes of eight mammals (human, chimpanzee, rhesus macaque, marmoset, mouse, rat, bull, and opossum) to exacting standards of accuracy and completeness; we are pleased to say that this goal is in sight. Throughout the course of the project, we have publicly shared the immediate products of our efforts [the completed sequences of individual bacterial artificial chromosome (BAC) clones] via GenBank, in keeping with the 1996 Bermuda Principles (www.genome.gov/10506537), so the data could be used by the community. We have done so with the understanding that our research team's interests are protected by the 2003 Fort Lauderdale Agreement (www.genome.gov/pages/research/wellcomereport0303.pdf) and the 2009 Toronto Agreement (www.nature.com/nature/journal/v461/n7261/full/461168a.html), which govern the use of publicly deposited biological datasets. Specifically, the Fort Lauderdale and Toronto agreements protect the right of data generators to publish the first analyses of their publicly deposited datasets. Taken together, the Bermuda, Fort Lauderdale, and Toronto accords have promoted and enabled the immediate public sharing of large and valuable biological datasets. The record shows that these widely accepted principles have served the interests of the public and of the scientific community as a whole.

It is in this context that we were surprised to read the publication by Chang et al. (1), which presents an unauthorized annotation of the entirety of our preliminary bull Y chromosome sequence assembly (GenBank accession no. CM001061). We first announced the plan to assemble and annotate the bull Y chromosome in 2006 (2). Professor Liu made unauthorized use of our research team's unpublished data in publications in 2009 and 2011 (3–5), and we communicated our concerns to Professor Liu after reading these papers. We informed Professor Liu of our plans to publish a comprehensive and systematic description and analysis of the bull Y sequence on completion of the project, and we requested that he respect the Fort Lauderdale and Toronto guidelines. The recent publication (1) is decidedly at odds with these community principles, and as such poses a direct challenge to the practice of immediately and publicly sharing broadly useful datasets.

Beyond these concerns, we also write to raise issues regarding the scientific accuracy and validity of Chang et al.'s analyses and conclusions regarding the bull Y chromosome sequence. For example, we find that the majority of the 1,274 bull Y genomic sequences that Chang and colleagues identify as protein-coding genes are actually pseudogenes, many of which contain truncating mutations and many of which lack compelling evidence of transcriptional activity. It is of note here that the high-throughput mRNA sequencing datasets alluded to by Chang and colleagues have not been publicly released.

Our research team will in due course publish a rigorously grounded assembly, analysis, and annotation of the bull Y chromosome. In the meantime, our team will continue to publicly and immediately share the genomic sequence datasets that the team generates during the closing phases of this technically challenging project.

Jennifer F. Hughes[a], Helen Skaletsky[a,b], D. Winston Bellott[a], Bhanu P. Chowdhary[c], Wesley C. Warren[d], Kim C. Worley[e], Richard K. Wilson[d], Richard A. Gibbs[e], and David C. Page[a,b,f,1]

[a]Whitehead Institute, Cambridge, MA 02142; [b]Howard Hughes Medical Institute, Cambridge, MA 02142; [c]Department of Veterinary Integrative Biosciences, Texas A&M University, College Station, TX 77843; [d]The Genome Institute, Washington University School of Medicine, St. Louis, MO 63108; [e]Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030; and [f]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142

1 Chang T-C, Yang Y, Retzel EF, Liu W-S (2013) Male-specific region of the bovine Y chromosome is gene rich with a high transcriptomic activity in testis development. *Proc Natl Acad Sci USA* 110(30):12373–12378.
2 Rozen S, et al. (2006) Sequencing and annotating new mammalian Y chromosomes: a white paper proposal. Available at http://www.genome.gov/Pages/Research/Sequencing/SeqProposals/YChromosomeWP.pdf. Accessed July 1, 2006.
3 Liu WS, et al. (2009) Molecular characterization of the *DDX3Y* gene and its homologs in cattle. *Cytogenet Genome Res* 126(4):318–328.
4 Chang TC, et al. (2011) The expansion of the *PRAME* gene family in Eutheria. *PLoS ONE* 6(2):e16867.
5 Yang Y, et al. (2011) *ZNF280BY* and *ZNF280AY*: Autosome derived Y-chromosome gene families in Bovidae. *BMC Genomics* 12:13.

Author contributions: J.F.H., H.S., D.W.B., B.P.C., W.C.W., K.C.W., R.K.W., R.A.G., and D.C.P. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. E-mail: dcpage@wi.mit.edu.

LETTER

Reply to Hughes et al.: No bullying: Publications comply with international standards

Hughes et al. (1) claim we made unauthorized use of their unpublished data in our publications on the bovine Y chromosome (BTAY) 2–5). These accusations are troubling in their intent and unfounded. First, we carefully followed the international standards, including the Fort Lauderdale (www.genome.gov/pages/research/wellcomereport0303.pdf) and Toronto Agreements (www.nature.com/nature/journal/v461/n7261/full/461168a.html) for public sharing of biological datasets. In all cases, proper acknowledgments were made in our publications whenever publically available data were used. Second, we have studied BTAY for >15 years with financial support from the US Department of Agriculture (USDA). In 2007, we initiated a direct testis cDNA selection approach to study the transcriptome of BTAY using a Y-specific DNA library (6). We identified a number of unique transcripts (cDNAs), including those transcribed from PRAMEY, ZNF280AY, and ZNF280BY (3, 4). By BLAST searching with our cDNA sequences, we identified a large number of bovine bacterial artificial chromosomes (BACs) that were listed as unassigned in 2008. We informed Dr. Kim Worley and Dr. David Page about our discovery, requesting a potential collaboration, which was not welcomed, so we continued our work independently. We analyzed the PRAMEY-, ZNF280AY-, and ZNF280BY-linked BACs (51 of 58 were unassigned) and acknowledged the sequence data producer in our publications (3, 4). Third, the draft BTAY sequence assembly (GenBank accession no. CM001061) was used as a reference during our analysis of the BTAY transcriptome (5). We aligned the testis RNA-seq reads against the draft BTAY assembly to identify Y-linked reads for assembly of RNAs, which were used for the Y locus–specific expression study. Because the draft BTAY assembly was reported to GenBank in 2010, the use of the draft assembly as a reference in our work (5) is entirely consistent with the Toronto Agreement, which states the protected time period to allow the data producers to publish the analysis of their dataset is "ideally within one year." Fourth, we recently deposited the testis RNA-seq data generated for the BTAY transcriptome analysis into GenBank. The delay in release of these data was due to the death of our dear friend and coauthor, Dr. Ernest Retzel. Finally, we welcome any researchers to challenge our research findings regarding the number of protein coding genes on BTAY with data. Our predicted number of genes was based on de novo assembly of the Y-linked RNA-seq reads, and the alignment of the assembled cDNAs against the draft BTAY assembly to the predicted gene loci based on the splicing sites and ORF lengths (≥90 amino acids) (5).

Therefore, the claims of Hughes et al. are unjustified, set a bad precedent for the entire research community, and are inconsistent with the goals of large, publically funded projects, for which rapid public disclosure of sequence results is required. We agree with Hughes et al. that the release is required "so the data could be used by the community." Only with this approach will public research funds be used for maximum societal benefit.

Wan-Sheng Liu[1], Ti-Cheng Chang[2], and Yang Yang[3]

Department of Animal Science, Center for Reproductive Biology and Health, College of Agricultural Sciences, The Pennsylvania State University, University Park, PA 16802

1 Hughes JF, et al. (2013) No bull: Upholding community standards in public sharing of biological datasets. *Proc Natl Acad Sci USA* 110:E4277.
2 Liu W-S, et al. (2009) Molecular characterization of the DDX3Y gene and its homologs in cattle. *Cytogenet Genome Res* 126(4):318–328.
3 Chang TC, et al. (2011) The expansion of the PRAME gene family in Eutheria. *PLoS ONE* 6(2):e16867.
4 Yang Y, et al. (2011) ZNF280BY and ZNF280AY: Autosome derived Y-chromosome gene families in Bovidae. *BMC Genomics* 12:13.
5 Chang T-C, Yang Y, Retzel EF, Liu W-S (2013) Male-specific region of the bovine Y chromosome is gene rich with a high transcriptomic activity in testis development. *Proc Natl Acad Sci USA* 110(30): 12373–12378.
6 Liu W-S, Mariani P, Beattie CW, Alexander LJ, Ponce De León FA (2002) A radiation hybrid map for the bovine Y Chromosome. *Mamm Genome* 13(6):320–326.

Author contributions: W.-S.L., T.-C.C., and Y.Y. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. E-mail: wul12@psu.edu.

[2]Present address: Department of Plant Pathology, University of California, Davis, CA 95616.

[3]Present address: Department of Pathology and Laboratory Medicine, School of Medicine, University of North Carolina, Chapel Hill, NC 27599.

www.pnas.org/cgi/doi/10.1073/pnas.1316664110