CBE—Life Sciences Education
Vol. 10, 318–327, Fall 2011

Article

A Comprehensive Analysis of High School Genetics Standards: Are States Keeping Pace with Modern Genetics?

M.J. Dougherty,* C. Pleasants,† L. Solow,‡ A. Wong,* and H. Zhang*

*American Society of Human Genetics, Bethesda, MD 20814; †Hampden-Sydney College, Hampden-Sydney, VA 23943; ‡Wesleyan University, Middletown, CT 06459

Submitted September 20, 2010; Revised April 4, 2011; Accepted April 4, 2011
Monitoring Editor: Elisa Stone

Science education in the United States will increasingly be driven by testing and accountability requirements, such as those mandated by the No Child Left Behind Act, which rely heavily on learning outcomes, or "standards," that are currently developed on a state-by-state basis. Those standards, in turn, drive curriculum and instruction. Given the importance of standards to teaching and learning, we investigated the quality of life sciences/biology standards with respect to genetics for all 50 states and the District of Columbia, using core concepts developed by the American Society of Human Genetics as normative benchmarks. Our results indicate that the states' genetics standards, in general, are poor, with more than 85% of the states receiving overall scores of Inadequate. In particular, the standards in virtually every state have failed to keep pace with changes in the discipline as it has become genomic in scope, omitting concepts related to genetic complexity, the importance of environment to phenotypic variation, differential gene expression, and the differences between inherited and somatic genetic disease. Clearer, more comprehensive genetics standards are likely to benefit genetics instruction and learning, help prepare future genetics researchers, and contribute to the genetic literacy of the U.S. citizenry.

INTRODUCTION

The first requirements of the No Child Left Behind Act (NCLB), which became law in 2002, led every state to evaluate the adequate yearly progress (AYP) of students in math and language arts based on specific content standards in those subjects. Assessment requirements extended (by statute, if not yet in practice) to science in 2007, where they encompass testing in three grade bands, 3–5, 6–9, and 10–12. In science, guidance for the production of content standards and curricula has been available through the *National Science Education Standards* (National Research Council [NRC], 1996) and *Project 2061 Benchmarks for Science Literacy* (American Association for the Advancement of Science [AAAS], 1993, 2009), and through state and district administrative offices.

The standards governing AYP in all subjects were developed on a state-by-state basis because, until recently at least, there has been little support at the state level for common national standards. That may be changing. Most states have adopted common core standards for reading and math, although tremendous work remains to be done to implement these standards (Paulson, 2010; for current list, see http://www.corestandards.org/in-the-states). The NRC recently initiated a process for developing common core standards for science.

The rationale behind standards and testing can be summarized by a familiar saying, invoked in the context of NCLB by former U.S. Secretary of Education Margaret Spellings: "What gets measured, gets done" (Spellings, 2005). In K–12 education, this can be further translated as standards drive testing and testing drives curriculum (Popham, 2004). There is some evidence that standards can also drive achievement, although here the details of structure and implementation are more critical (Swanson, 2006).

We make no argument in favor of or against standards-based instruction or NCLB; however, given that this is the

DOI: 10.1187/cbe.10-09-0122
Address correspondence to: M. J. Dougherty (mdougherty@ashg.org).

reality of K–12 public education in the United States at the current time, we do argue that understanding the quality of standards is useful and even necessary. Science will soon follow reading and math down the path of high-stakes outcomes and the drive to achieve AYP—through curriculum, instruction, and testing—and so it is reasonable to examine science standards for quality and consistency across different states. This is especially true in light of concerns that variations in defining proficiency allow states to manipulate standards to facilitate meeting AYP and to avoid sanctions (Kingsbury *et al.*, 2003; Aspen Institute, 2006). In fact, evidence from the National Center for Educational Statistics (NCES) demonstrates an incongruity between student proficiency at the state level and student performance on the gold standard of assessment, the National Assessment of Educational Progress (NAEP): "There is a strong negative correlation between the proportions of students meeting the states' proficiency standards and the NAEP score equivalents to those standards, suggesting that the observed heterogeneity in states' reported percents proficient can be largely attributed to differences in the stringency of their standards" (NCES, 2007). Such findings suggest worrisome weaknesses in math and reading standards and reason for concern apropos science standards.

To date, there have been few comprehensive evaluations of specific science content in state standards, and none by a professional scientific society, and their quality remains largely unknown (Gross *et al.*, 2005; Hoffman and Barstow, 2007). In genetics, lists of core principles have been developed to guide learning for audiences ranging from undergraduate nonscience majors to health professionals (Hott *et al.*, 2002; National Coalition for Health Professional Education in Genetics [NCHPEG], 2004; Smith *et al.*, 2008), but there are no similar lists for K–12. The American Society of Human Genetics (ASHG) has been interested in genetics education and K–12 standards since it began documenting misconceptions that frequently appeared in essays submitted to its National DNA Day Essay Contest (Shaw *et al.*, 2008). In this paper, we present the results of an evaluation of the life sciences standards of every state to assess their adequacy with respect to genetics coverage. As benchmarks, we developed and used a list of core genetics concepts that all students should understand by grade 12 as preparation for life in a world of healthcare and medicine that is increasingly informed by genetics. This analysis of extant genetics standards hopefully can inform science education researchers and policy makers during the normal revision process for existing standards, which occurs at different intervals for different states, and during the development of common science standards.

METHODS

Identifying Core Concepts

A list of "core concepts" in genetics appropriate for students up to grade 12 was developed using an iterative process. Initial drafts of concepts were adapted by ASHG staff from several previously published documents that included important or essential genetics content: "Development and evaluation of a genetics literacy assessment for undergraduates" (Bowling *et al.*, 2008); "The genetics concept assessment" (Smith *et al.*, 2008); "Closing the gap" (Dougherty, 2009); NCHPEG Core Principles of Genetics (2004); *Project 2061 Benchmarks for Science Literacy* (AAAS, 2009); *National Science Education Standards* (NRC, 1996); *AP Biology Course Description* (College Board, 2010); and undergraduate nonmajor biology courses (Hott *et al.*, 2002). Using these documents as a baseline, concepts were developed to adequately describe five major conceptual areas of genetics deemed essential to genetic literacy: nature of the genetic material, transmission genetics/patterns of inheritance, gene expression and regulation, genetic variation, and evolution.

Because our interest was the quality and comprehensiveness of genetics content in the state standards, we did not develop benchmarks (or analyze standards) related to noncontent understandings or abilities, such as inquiry, the nature of science, or system standards. Advanced genetics topics, such as epigenetics, the regulatory roles of small RNAs, chromatin remodeling, and others, also were not included on the grounds that: 1) they were too complex or detailed and basic genetic literacy did not require them of all students; 2) many teachers would likely be unfamiliar with them; and 3) most state standards would not be expected to include them. Initial drafts were reviewed, revised through 11 iterations, and ultimately approved by ASHG's Information and Education Committee, a standing committee that advises ASHG's board of directors on issues related to genetics education. Its members have expertise in genetics content, science education, and teaching, including substantial expertise with high school curricula and pedagogy. A total of 15 concepts were tested in a pilot analysis, and a final list of 19 concepts served as the benchmark concepts against which state science standards were compared (Table 1).

Evaluating State Standards with Respect to Genetics

The most current state standards as of summer 2009 were identified from the websites of the departments of education for all 50 states and the District of Columbia. Standards related to genetics in any way (i.e., in all life or biological sciences categories) from grades 9–12 were assembled in state-specific pdf files to aid reviewers. Where states used more than one section to group content, for example, introductory paragraphs or rubrics to supplement content listings, we tried to be as inclusive as possible. In general, the only content excluded from consideration related to rare cases where content was explicitly identified as beyond the level at which all students should learn (i.e., exceeding the state's view of essential content and, therefore, less likely to be taught or tested). A simple online scoring guide was developed in-house to allow reviewers to open an appropriate pdf of the state's standards, review them, and then use drop-down menus to record a score for each concept. A box for written comments was also provided. In light of ongoing efforts to revise national standards documents, we chose not to analyze standards that appear likely to soon be replaced (e.g., the *National Science Education Standards* or *Project 2061 Benchmarks for Science Literacy*).

Pilot Evaluation

A pilot test was conducted to evaluate the feasibility and reliability of our rating system. We used a three-level grading system (0 = Not addressed/absent, 1 = Inadequate, 2 = Adequate) that was similar to methods used for evaluating earth

Table 1. Scoring results for ASHG's genetics concepts averaged across all states by concept and by concept category

Core concept	Average score (all states)/ adequacy	Concept category	Average by category/ adequacy
1. DNA is the genetic material for all species of living organisms.	1.7/Adequate	Nature of genetic material	1.6/Adequate
2. Genes are segments of DNA that encode information critical for development. DNA is organized into structures called chromosomes.	1.5/Adequate		
3. Genes exist in different forms called alleles.	1.1/Inadequate	Transmission/ patterns of inheritance	0.9/Inadequate
4. Alleles are passed from parent to offspring through the processes of replication, meiosis, and fertilization.	1.4/Inadequate		
5. For traits primarily influenced by single genes, certain combinations of alleles lead to predictable genotypic and phenotypic patterns of inheritance, illustrating Mendel's principles of segregation of alleles and independent assortment of genes.	1.5/Adequate		
6. Polygenic (or complex) traits (e.g., height, blood glucose) often show continuous variation within populations and are less predictable than single-gene traits.	0.3/Not addressed/ absent		
7. Polygenic traits are influenced by multiple genes and their products.	0.3/Not addressed/ absent		
8. The expression of genetic information generally flows from DNA to RNA to protein. This occurs through transcription of DNA into RNA and translation of mRNA into protein.	1.3/Inadequate	Gene expression and regulation	0.9/Inadequate
9. Virtually all cells within an organism contain the same genetic information.	0.9/Inadequate		
10. Different genes are turned on and off at specific times to form different types of cells and to influence the way different cells function.	0.6/Inadequate		
11. The functions of genes and their products can be affected by the environment and other genes at one or many steps involved in producing a trait.	0.7/Inadequate		
12. Mutations are changes in DNA sequence. They can occur spontaneously during DNA replication or they can be the result of damage by environmental factors.	1.3/Inadequate	Genetic variation	1.1/Inadequate
13. Mutations in DNA, and sorting and recombination during meiosis, result in genetic variation.	1.4/Inadequate		
14. Only mutations in the DNA of sex cells will be passed on to offspring. Mutations in somatic cells will be passed on only to descendant cells.	0.9/Inadequate		
15. Mutations may help, harm, or have little or no effect on an organism.	1.2/Inadequate		
16. One harmful effect of mutations is genetic disease. Some genetic diseases are inherited (e.g., Tay-Sachs), and others develop during life (e.g., cancer).	0.8/Inadequate		
17. Genetic variation and the phenotypic variation it leads to are the basis for evolution.	1.5/Adequate	Evolution	1.5/Adequate
18. Evolution by natural selection is a process by which inherited traits influence how likely an organism is to survive, reproduce, and pass those traits to its offspring.	1.7/Adequate		
19. The process of evolution occurs at a population level (e.g., not at the level of individual organisms), and takes place over generations (e.g., not within an individual organism's lifespan).	1.2/Inadequate		

Numerical scores: 0 = Not addressed/absent; 1 = Inadequately addressed; 2 = Adequate.

science standards (Hoffman and Barstow, 2007). Two project staff members serving as pilot reviewers independently evaluated the standards of 10 randomly selected states against a draft set of ASHG core genetics concepts that included 15 concepts ($n = 150$ pairwise comparisons). Interrater reliability at the concept level was determined by simple counting. Interrater reliability for averages across all concepts at a state-by-state level was calculated according to the methods

of Ebel (1951), using the open-access calculator developed by Solomon (2004). "Adequacy" grades were assigned by converting state or concept averages according to: Adequate = 1.5–2.0; Inadequate = 0.6–1.4; Not addressed/absent < 0.6.

Full Evaluation

To assemble the requisite experts to serve as evaluators of state standards, we recruited 167 members of ASHG's Genetics Education Outreach Network, a volunteer network of geneticists across the nation who work in K–12 classrooms, help grade DNA Day essays, and/or provide career talks. All have content expertise in genetics and varying levels of expertise with genetics education outreach. Of those participating, 77% held doctoral degrees (mostly PhD and MD), 20% held master's degrees, and just over 2% held bachelor's degrees. (ASHG's membership consists largely of academic basic researchers, clinical researchers, and genetic counselors.) Our rationale for using a large pool of experts external to the project team as reviewers was that the intent of a state's standards apropos a specific concept should be clear. If a specific concept in the standards is not apparent to these experts, it likely will not be apparent to many curriculum developers and teachers. Using our criteria, such standards would be graded as either Inadequate (i.e., present but not adequate) or Not addressed/absent. As Hoffman and Barstow (2007) have noted apropos standards, "Each state has its own development processes, generally involving a wide variety of people with different backgrounds, subject area expertise, teaching experience, special interests, and political agendas. These processes result in widely varying standards." Such variation is likely the case for those interpreting standards as well, thus arguing in favor of clarity. In addition, by using a large number of external reviewers, we hoped to avoid the effects of systematic bias that may creep into analyses conducted by small, close-knit teams.

The evaluation and grading of standards is necessarily a subjective process, but by averaging across a number of independent judges, we believe a reasonably accurate assessment was attained. Just over half (94) of the evaluators reviewed two states, and the remainder (73) reviewed one state, with sufficient overlap to yield an average of 5.11 reviewers per state. No state had fewer than four reviewers. To minimize the impact of outliers, we dropped the high and low scores from states having five or six reviewers. Thus, the analysis for all states was based on a minimum of three scores per concept (four for 14 states). Averages by state across all concepts (i.e., per-state averages) and averages by concept across all states (i.e., per-concept averages) were calculated. Interrater reliabilities were evaluated by the methods of Ebel and Solomon, as described above.

To establish a common reference frame for reviewers to use in their evaluation (acknowledging both the vagaries of the state standards and likely differing interpretations), the reviewers were instructed by email as follows:

> Please use your best judgment to match concepts with standards and do not assume too much. An incomplete standard that evokes for you (as an expert in genetics) a natural elaboration encompassing the ASHG core concept may not evoke that same connection for a non-expert high school teacher. We need to evaluate what each standard actually says. At the same time, if the intent of a standard clearly matches the intent of an ASHG core concept and differs only in wording, you should credit that standard appropriately. All states are being evaluated by multiple reviewers.

RESULTS

An important statistical measure for assessing the consistency of analysis (or scoring) by different reviewers is interrater reliability, which we calculated for the pilot and full evaluations. Values of 70% or higher are generally considered good to acceptable, and values of 80% or higher to be very good (Lombard *et al.*, 2010). Interrater reliability in the pilot phase (two reviewers), across 150 pairwise comparisons, was 63% at the per-concept level, and in only one instance (0.7%) did the two reviewers differ by 2 grading steps (i.e., 0, Not addressed/absent, and 2, Adequate). Interrater reliability at the level of per-state averages (i.e., average of scores across all 15 concepts used in the pilot) was 96%. The difference in reliability at the two levels reflects the fact that states with overall weak (or strong) standards tend to have many individually weak (or strong) standards, and averaging smooths out differences that may be more apparent at the finer-grained level of individual concepts. In addition, the very high degree of per-state reliability was not surprising, because the project staff who conducted the pilot analysis had spent considerable time assembling pdfs of the state standards and discussing scoring criteria and the meaning of agreement. Also, the pilot reviewers noted that a number of concepts were difficult to align with state standards, because the benchmarking concepts themselves contained more than one distinct idea. Based on that finding, several benchmarking concepts containing more than one idea were divided into simpler, separate concepts, yielding a final draft with 19 concepts (Table 1).

At the conclusion of the full evaluation, the interrater reliability for the same 10 states that were used in the pilot was 79%. The interrater reliability across all states analyzed (i.e., 50 states plus the District of Columbia) was 75%. The per-concept reliability average was 71%, with four concepts having a reliability under 60% (averaging 52%: concepts 1, 11, 18, and 19 in Table 1). However, there was virtually no correlation between interrater reliability (by concept) and concept score ($r^2 = 0.16$; unpublished data). Thus, although four concepts were more challenging for reviewers to evaluate in the standards (i.e., lower reliability), that difficulty did not translate into systematically higher or lower scores.

We first analyzed which genetics concepts were adequately addressed across the nation as a whole. The 19 core concepts yielded average scores (i.e., across all states) ranging from Not addressed/absent (2 concepts) to Inadequate (12 concepts) to Adequate (5 concepts), and the five broader genetics categories encompassing those concepts had averages of Adequate (2) and Inadequate (3; see Table 1). Overall, average scores for many concepts were low (e.g., 7 out of 19 were <1.0), and the range of scores was broad both across and within categories, especially for transmission/patterns of inheritance, which comprised five concepts ranging from 0.3–1.5. No concept scored above 1.7, and only five scored 1.5 or higher (DNA being common to all living organisms; genes as units of information; Mendel's laws of segregation

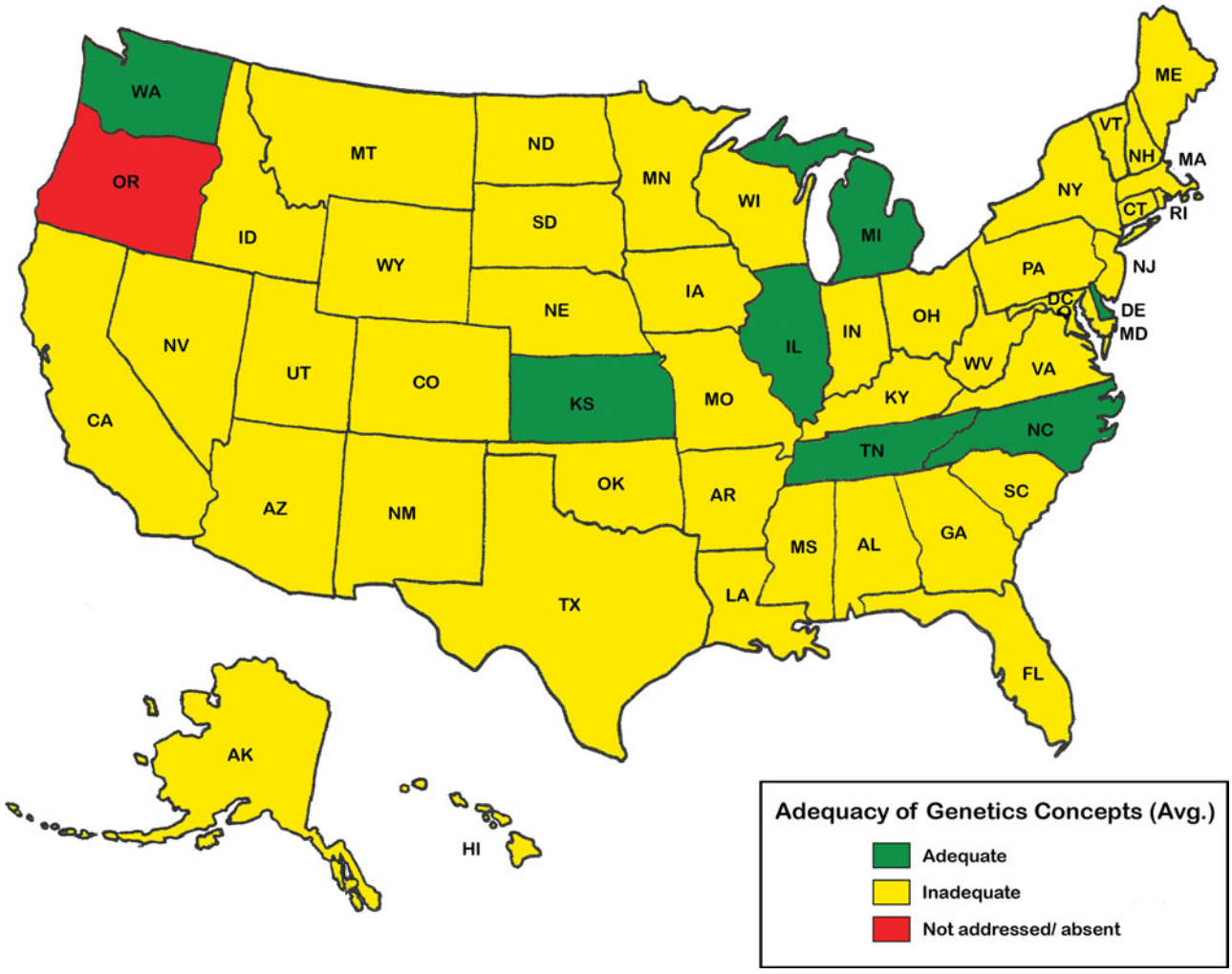

Figure 1. Map of the United States summarizing the average quality of genetics standards on a state-by-state basis.

and independent assortment; genetic and phenotypic variation as the substrate for evolution; and evolution by natural selection). The five lowest-scoring concepts ranged from 0.3–0.8 (continuous variation; polygeny; differential gene expression; multifactorial causation; and the connection between mutations and inherited vs. somatic genetic disease). The two lowest-scoring concepts (6 and 7, at 0.3—essentially not addressed) are both related to complex traits. When concepts were analyzed by category, the nature of the genetic material and evolution fared best, scoring Adequate, while transmission/patterns of inheritance, gene expression and regulation, and genetic variation all were deemed Inadequate.

We next analyzed individual states to determine which ones adequately addressed these genetics concepts. Average scores across all concepts were calculated by state and were normally distributed (mean = 1.11, median = 1.11, and range = 0.51–1.72). Figure 1 shows the results of each state's overall quality of genetics standards using a map depiction. Seven states (i.e., 14%) had overall genetics standards deemed Adequate by the reviewers (Delaware, Illinois, Kansas, Michigan, North Carolina, Tennessee, and Washington). Among those states, Michigan fared best, with an average of 1.72, followed by Delaware at 1.67. States judged to have Inadequate standards and at the bottom of the ranking included Alaska, Colorado, Idaho, Louisiana, North Dakota, Nebraska, New Jersey, Oregon (the only state to have an overall average of Not addressed/absent with respect to genetics coverage), and Wyoming, which all had average scores below 0.8. (For individual concept scores for each state, please see supplementary data at: www.ashg.org/education/pdf/StateConceptScores.pdf.)

We also examined states for their "comprehensiveness" with respect to the benchmark genetics concepts—that is, the number of individual genetics concepts (out of the total 19) for which a state received scores ranking it as Adequate (independent of overall average). Figure 2 shows a map of the specific results by state and Table 2 summarizes those results. Delaware and Michigan also fared well by this measure, each having 15 concepts that scored as Adequate. Interestingly, both states scored inadequately on the same four concepts (6, 7, 10, and 11), which are the four lowest-scoring concepts overall (see Table 1). Ten additional states were judged to have adequate coverage of at least 11 concepts. Seven states had adequate coverage of fewer than three concepts. Not

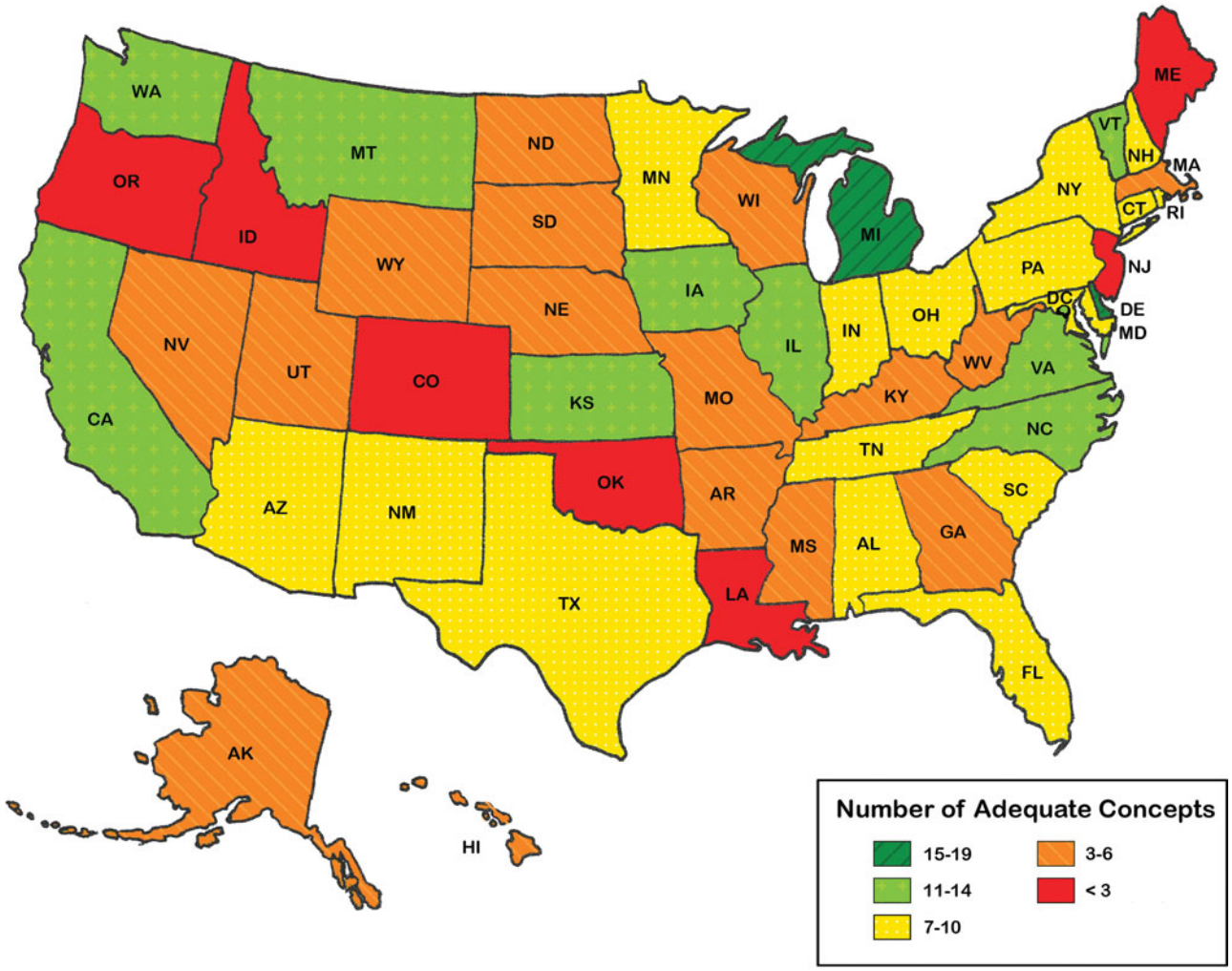

Figure 2. Map of the United States summarizing the comprehensiveness of genetics standards on a state-by-state basis. The ASHG benchmarks list included 19 concepts (see Table 1); colors indicate the total number of concepts rated by reviewers as Adequate.

Table 2. Summary of comprehensiveness of genetics standards across all states and comparison with an independent evaluation of the overall quality of science standards.

Number of adequate genetics concepts (out of 19)	Number of states (including District of Columbia)	Fordham scoring, science overall	Number of states
15–19	2	A	7
11–14	10	B	12
7–10	16	C	9
3–6	16	D	7
<3	7	F	15
Total	51		50

Columns 1 and 2: see Table 1. Columns 3 and 4: Data from the Thomas B. Fordham Institute (Gross *et al.*, 2005). The Fordham analysis used a letter grading system of A–F and considered noncontent criteria, such as "organization," "seriousness," and "inquiry." At the time of this study, Iowa was not included because it did not have its own state standards.

surprisingly, states with high overall averages tended to have the largest number of concepts that adequately addressed the genetics benchmarks ($r^2 = 0.80$; unpublished data). The few exceptions to this trend were relatively minor. For example, Kansas had a very high average of 1.6, but addressed only 11 of 19 concepts adequately. This is not far from the modal value of 10 Adequate concepts across all states. Similarly, Texas had a relatively high average score of 1.4 (just below Adequate), but had Adequate treatment of only seven concepts. In both cases, the concepts that were addressed adequately tended to

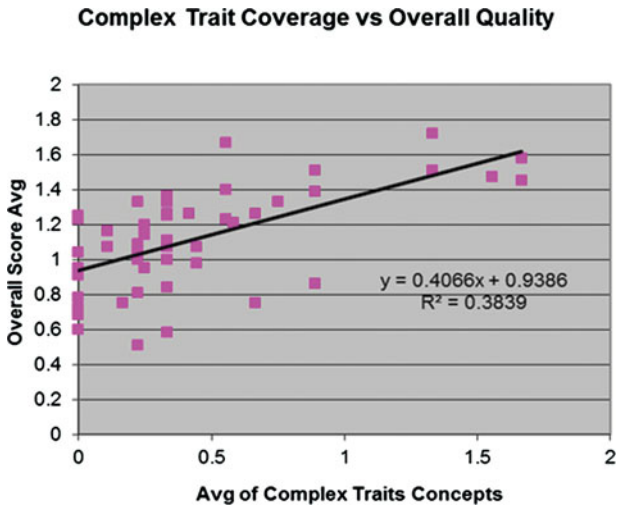

Figure 3. The relationship between the overall average quality of genetics standards to the average of the three concepts directly related to complex trait genetics (concepts 6, 7, and 11; Table 1).

have consistently high ratings from all reviewers. Across all states, the average number of benchmark concepts covered adequately was 7.4, with an SD of 4.0.

Three of the four lowest-scoring concepts deal with complex traits (continuous variation, polygenic inheritance, and multifactorial causation [concepts 6, 7, and 11, respectively, in Table 1]). Only Kansas, North Carolina, and Tennessee were judged as having Adequate coverage of complex trait genetics concepts (unpublished data). There was a modest but significant positive correlation between the overall quality of a state's genetics standards and its treatment of complex traits ($r^2 = 0.38$, $p < 0.001$; Figure 3).

DISCUSSION

Our analysis is based on the premise that the discipline of genetics/genomics (and scientific disciplines, in general) is structured around an intellectual core of fundamental concepts. The dynamic nature of science has accustomed us to expect that our understanding of scientific details will change even as further research serves to strengthen the concepts at the core. Such is the case with genetics, where concepts such as the allelic nature of genes, basic processes of gene expression, mutations, genetic variation, and patterns of inheritance have been expanded and enhanced across more than a century. Other concepts, while less familiar, such as differential gene expression and the genetic basis of complex traits, have been recognized for decades as crucial, but only in the past 10–20 yr have advances in technology enabled researchers to elucidate the genetic mechanisms by which they occur.

The collection of concepts that make up our core genetics benchmarks is not definitive. Other groups (several of which are referenced in *Methods*) have developed their own lists for their own purposes. The ASHG list drew heavily from those previously published lists, but our choices were guided by the specific goal of capturing the minimum number of concepts, at the appropriate level of specificity, that the authors and ASHG's Information and Education Committee believe to be essential for genetic literacy and thus should be understood by all high school graduates. As with all such lists, our process necessitated choosing among concepts that should be included and concepts that should be omitted, and some readers may disagree with these choices. We believe that ASHG, a large and influential professional society in genetics, can make a contribution to the discussion of standards with its views on what constitutes a reasonably specific, comprehensive, and age-appropriate set of concepts in one of the major subdisciplines of biology. Consequently, our choices reflect a moderate level of specificity, and our benchmark list includes 19 concepts rather than, for example, 6 or 60. This acknowledges dramatic changes in genetics over the past two decades, such as the discipline's increasing impact on medicine and direct-to-consumer genetic testing, and the fact that an understanding of additional concepts is needed to raise the standard of scientific literacy among the nation's population. The new concepts we propose can be taught using many different subconcepts and examples, which we do not prescribe.

Overall, our analysis identifies substantial deficiencies in the treatment of genetics in state standards across the United States. Of ASHG's 19 benchmark concepts, 14 were treated inadequately (averaged across all states, including the District of Columbia), and 39 states had Adequate coverage of fewer than 11 concepts (Table 1 and Figure 2). Two concepts (continuous variation of complex traits and polygenic inheritance) were virtually absent from state standards. Three of five major categories of genetics concepts are covered inadequately in standards. Only evolution and the nature of the genetic material were adequately covered, while the broad categories of gene expression and regulation and genetic variation were not.

It is particularly troubling that the Inadequate and Not addressed/absent concepts represent ideas that are increasingly important as genetics/genomics matures and assumes practical importance in people's lives. As DNA sequencing and genotyping technologies have become more powerful and less expensive (at rates that vastly exceed Moore's law in computing), they have been applied more often in medical research. It is now possible to investigate the genetic contributions to virtually any disease or normal trait, and our ability to apply knowledge of individual genetic variation to clinical treatment and outcomes is advancing. Even now "personalized medicine" is becoming manifest. For example, roughly 10% of drugs approved by the Food and Drug Administration have labels with pharmacogenomics information, and genetic testing has already become common for certain conditions, such as breast cancer (Hamburg and Collins, 2010). As Lanie *et al.* have written, "Even if people grasp and understand these basic concepts [Mendelian genetic concepts], the impact of this knowledge will be limited. . . it will not go far in helping the layperson understand the barrage of genetic information to which they are exposed through the media—the vast majority of which deals with complex diseases and traits" (2004). Without adequate standards dealing with mutations, gene regulation, and non-Mendelian patterns of inheritance, such concepts may never make it into curricula and assessments.

The most-neglected concepts also demand higher orders of thinking than concepts that tended to be covered adequately. For example, unlike the nature of the genetic material (i.e.,

Table 3. Students' responses to genetics questions on the 2000 NAEP (grade 12) indicate poor understanding of essential genetics concepts

NAEP question (related concept, this work, from Table 1)	% Complete or essential answers
What is a gene? What is it made of? What is the major function of a gene? (2)	21
Briefly explain how the information that a gene provides to a cell results in the production of a molecule such as hemoglobin. (8)	1
[E]xplain why kidney cells do not make hemoglobin, even though they contain the hemoglobin gene. (10)	10
Give an example of an inherited disease that people might be able to treat by use of [recombinant DNA] technology, and explain how you think the technology might be used to treat this disease. (16)	5

DNA), which deals largely with descriptive biology, gene expression (concept 8) deals with the mechanisms of how genes operate and the functional consequences of those operations. The latter involve understanding positive and negative feedback (general ideas that also extend to engineering, computing, and economics) and require thinking at levels of application, interpretation, and analysis. In parallel with being poorly represented in state standards (Table 1), assessment questions related to the concepts of gene expression (concept 8) and differential gene expression (concepts 9 and 10) elicited very few responses indicative of complete understanding by grade 12 students on the 2000 NAEP (2006; Table 3).

Another specific area of weakness is the state standards' treatment of mutations. Frequently, state standards do not distinguish between germline and somatic mutations and their connections to hereditary versus nonhereditary genetic disease (concepts 14 and 16, respectively). Failure to make this distinction may contribute to students' conflating inherited disease with any disease influenced by genetics, a frequent student misconception (Shaw *et al*., 2008). Students' poor understanding of these concepts is also reflected in their responses to a related question on the NAEP (Table 3).

There are some bright spots. Perhaps not surprisingly, DNA as the genetic material (concept 1) is well covered by most states, as it should be nearly 60 yr after the elucidation of DNA's structure (average score of 1.7, Table 1). The organization of genetic information in the form of genes carried by chromosomes also fared well (1.5). Evolution by natural selection (concept 18) tied for best average score (1.7), and was supported by strong coverage of genetic and phenotypic variation as the substrate for evolution (concept 17; 1.5). These findings support earlier research indicating improvements in state standards' coverage of evolution (Mead and Mates, 2009). Unfortunately, state standards are not doing as well framing the population-level and generational timescale at which evolution operates (concept 19).

Learning goals related to single-gene inheritance patterns (i.e., Mendelian inheritance, concept 5) are also adequately addressed nationwide in the state standards (average score 1.5), an encouraging but not unexpected finding. Instruction about genes, alleles, and Mendelian segregation, including meiosis, is virtually ubiquitous in high school general biology, and these concepts continue to dominate instruction in equivalent undergraduate courses (Hott *et al*., 2002). These concepts also lend themselves easily to problem-based learning, which engages higher-level critical thinking. Unfortunately, there may be a downside to all this attention. In the context of the modern discipline's broader view of genetics, we may be skewing students' understanding of the genotype–phenotype connection (Dougherty, 2009). In essence, if we spend too much time on single-gene inheritance—at the expense of polygenic inheritance and complex traits—then we not only fail to convey modern genetics accurately, but we also risk giving students a false impression that most traits are inherited in the "simple" manner conveyed by the rare single-gene traits that are so often used as examples, such as cystic fibrosis and hemophilia. In fact, patterns-of-inheritance misconceptions represent the second most problematic conceptual category in genetics identified by Shaw *et al*. (2008). Not surprisingly, low scores for concepts related to complex traits and multifactorial causation (concepts 6, 7, and 11) contributed substantially to low average scores for most states.

To get some sense of how our specific results compared with the quality of state science standards as judged more generally, we compared them with the results of an evaluation conducted by the Thomas B. Fordham Institute (Table 2, columns three and four). Differences in methodology prevent direct comparison; however, weaknesses in state standards are apparent in both analyses.

Our study was limited in several ways. First, states were not scored blindly, which may have allowed some unintentional reviewer bias, although our method of using multiple reviewers and dropping the high and low scores for states with five or six reviewers should have helped minimize such bias. Second, for a small number of states, it is possible that some concepts may have been addressed only at lower grade levels, in which case this analysis would have missed them. The decision to focus exclusively on high school standards was based on a preliminary analysis that indicated that middle school standards were generally repeated in the high school standards for the states that included genetics concepts at both levels. The repetition of genetics concepts in middle and high school standards, in the cases where we observed it, bolsters the argument by Schmidt *et al*. (2005) and Daro *et al*. (2010) that the current U.S. curriculum does not achieve much depth in math and science. Third, our analysis represents a temporal slice through state genetics standards. Different states have different revision schedules (and processes) for their standards, and the performance of any given state with respect to genetics (or any other content) would be expected to change over time. Finally, our analysis considers only one variable (i.e., standards) in the complex system that constitutes genetics education.

Of course, the absence of concepts in a state's standards does not mean that those concepts will not be taught. Knowledgeable teachers can address such concepts even in situations where high-stakes exams do not require them. However, given the pressures of performing well on state exams and limited classroom time, teachers often prioritize the content specified by their states. Conversely, the fact that concepts are represented in a state's standards does not guarantee that students will learn them (Schmidt *et al*., 2005). For example, DNA as the genetic material and the nature of genes were the

best-covered concepts for any category on our benchmark list (concepts 1 and 2, Table 1); however, only a minority of students gave complete or essential answers to a question dealing with these ideas on the 2000 NAEP. Those concepts have been taught since long before standards achieved their current prominence in education, and yet student understanding lags. Thus, we recognize that improved standards are no panacea.

The fact that 44 states (86%) had genetics standards judged to be Inadequate and only 12 states (24%) had 11 or more individual concepts judged to be Adequate convincingly demonstrates (in our view) a need for improvement when states revise their life sciences standards. ASHG will encourage members of its Genetics Education Outreach Network to volunteer to assist with revisions in those states that allow such participation. Similar networks exist at other scientific professional societies and may offer a generalizable mechanism for involving scientists who are content experts, as well as knowledgeable about K–12 education, in both the evaluation and improvement of content coverage in state standards.

Standards are just one element of what should be an integrated and coherent teaching system, and changes to standards should be made only with full recognition of the effects those changes may have on other parts of the system. Crowded curricula, teacher professional development, and available instruction time, must all be taken into consideration. Within genetics, new conceptual frameworks may be necessary. For example, it may be possible to refocus instruction around complex traits, which have a greater capacity for carrying modern understandings of genetics (as well as traditional ones) than do the simpler, single-gene traits that are used so widely in high school classrooms now (Dougherty, 2009).

The NRC's move to generate common science standards offers a different leverage point for modernizing the genetics curriculum. A recent report supports the establishment of clear and common standards across states, noting that roughly one-half the teachers in all states agreed that the standards of their own states were not clear enough, and that 85% think "having tougher academic standards would make at least a moderate impact on improving academic achievement" (Scholastic and Bill and Melinda Gates Foundation, 2010). Sound science instruction requires expertise in teaching (content knowledge and pedagogy), exemplary curricula, strong assessment, and a supportive system (e.g., administration, professional development). In the current U.S. system, standards are the foundation upon which curricula and instruction are built, and our work shows that the foundation, at least in genetics, is in need of repair. We propose that the findings detailed in this paper are well positioned to guide the development of common core standards in science.

ACKNOWLEDGMENTS

The authors thank the 2009/2010 Information and Education Committee of ASHG for their help in generating and reviewing the core concepts used in this analysis. We also thank the members of the Genetics Education Outreach Network, without whose assistance this work could not have been completed. The work of C.P. and L.S. was supported, in part, by a generous grant from Life Technologies to ASHG's Undergraduate Summer Interns Program. This material is based upon work supported, in part, by the National Science Foundation under Grant No. EHR-0634296. Opinions and conclusions expressed here are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

American Association for the Advancement of Science (1993; 2009). Project 2061 Benchmarks for Science Literacy, Washington, DC. www.project2061.org/publications/bsl/online/index.php (accessed 15 September 2010).

Aspen Institute, Commission on No Child Left Behind (2006). State Standards: Assessing Differences in Quality and Rigor and How They Impact NCLB, Hearing Report. www.aspeninstitute.org/policy-work/no-child-left-behind/reports/state-standards-assessing-differences-qual (accessed 13 December 2010).

Bowling BV, Acra EE, Wang L, Myers MF, Dean GE, Markle GC, Moshalik CL, Huether CA (2008). Development and evaluation of a genetics literacy assessment instrument for undergraduates. Genetics *178*, 15–22.

College Board (2010). AP Biology Course Description, Effective Fall 2010. http://apcentral.collegeboard.com/apc/public/repository/ap-biology-course-description.pdf (accessed 17 May 2011).

Daro P, McCallum W, Zimba J (2010). Common U.S. math standards. Science *328*, 285.

Dougherty MJ (2009). Closing the gap: inverting the genetics curriculum to ensure an informed public. Amer J Hum Gen *85*, 1–7.

Ebel RL (1951). Estimation of the reliability of ratings. Psychometrika *16*, 407–424.

Gross PR, Goodenough U, Haack S, Lerner LS, Schwartz M, Schwartz R (2005). The State of State Science Standards, Washington, DC: Thomas B. Fordham Institute. www.fordhamfoundation.org/doc/Science%20Standards.FinalFinal.pdf (accessed 15 September 2010).

Hamburg MA, Collins FS (2010). The path to personalized medicine. N Engl J Med *363*, 301–304.

Hoffman M, Barstow D (2007). Revolutionizing Earth System Science Education for the 21st Century: Report and Recommendations from a 50-State Analysis of Earth Science Education Standards, Cambridge, MA: TERC. www.oesd.noaa.gov/noaa_terc_study_lowres.pdf (accessed 17 May 2011).

Hott AM, *et al.* (2002). Genetics content in introductory biology courses for non-science majors: theory and practice. BioScience *52*, 1024–1035.

Kingsbury GG, Olson A, Cronin J, Hauser C, Houser R (2003). The State of State Standards. Northwest Evaluation Association. www.ictliteracy.info/rf.pdf/NWEA-State-of-Stndrds.4.pdf (accessed 26 January 2011).

Lanie AD, Jayaratne TE, Sheldon JP, Karia SLR, Anderson ES, Feldbaum M, Petty EM (2004). Exploring the public understanding of basic genetics concepts. J Genet Couns *13*, 305–320.

Lombard M, Snyder-Duch J, Bracken CC (2010). Practical Resources for Assessing and Reporting Intercoder Reliability in Content Analysis Research Projects. http://astro.temple.edu/~lombard/reliability/ (accessed 20 January 2011).

Mead LS, Mates A (2009). Why science standards are important to a strong science curriculum and how states measure up. Evolution: Education and Outreach *2*, 359–371.

National Assessment of Educational Progress (2006). The Nation's Report Card. http://nationsreportcard.gov/2005_assessment/s0043.asp (accessed 15 September 2010).

National Center for Education Statistics (2007). Mapping 2005 State Proficiency Standards onto the NAEP Scales (NCES 2007–482), Washington, DC: U.S. Government Printing Office.

National Coalition for Health Professional Education in Genetics (2004). Core Principles in Genetics. www.nchpeg.org/index.php?option=com_content&view=article&id=95&Itemid=105 (accessed 15 September 2010).

National Research Council (1996). National Science Education Standards, Washington, DC: National Academies Press.

Paulson A (2010). Uniform education standards: momentum grows as more states sign on. Christian Science Monitor July 26. http://tinyurl.com/2dvexvn (accessed 15 September 2010).

Popham WJ (2004). Curriculum matters. Am Sch Board J *191*, 30–33.

Schmidt WH, Wang HC, McKnight CC (2005). Curriculum coherence: an examination of U.S. mathematics and science content standards from an international perspective. J Curr Studies *37*, 525–559.

Scholastic and Bill and Melinda Gates Foundation (2010). Primary Sources: America's Teachers on America's Schools. www.scholastic.com/primarysources/pdfs/Scholastic_Gates_0310.pdf (accessed 15 September 2010).

Shaw KRM, Van Horner K, Zhang H, Boughman J (2008). Essay contest reveals misconceptions of high school students in genetics content. Genetics *178*, 1157–1168.

Smith MK, Wood WB, Knight JK (2008). The genetics concept assessment: a new concept inventory for gauging student understanding of genetics. CBE Life Sci Educ *7*, 422–430.

Solomon DJ (2004). The rating reliability calculator. BMC Med Res Methodol *4*, 11. www.ncbi.nlm.nih.gov/pmc/articles/PMC411037/ (accessed 15 September 2010).

Spellings M (2005). Growing pains won't sidetrack No Child Left Behind. USA Today, May 1. www.usatoday.com/news/opinion/editorials/2005-05-01-spellings-edit_x.htm (accessed 17 May 2011).

Swanson CB (2006). Making the connection: a decade of standards-based reform and achievement. Editorial Projects in Education Research, January 5. www.edweek.org/media/ew/qc/2006/MakingtheConnection.pdf (accessed 15 September 2010).