



Unlimited Hypothesis Research

James C. Engert

Genome Res. 2000 10: 271-272

Access the most recent version at doi:[10.1101/gr.10.3.271](https://doi.org/10.1101/gr.10.3.271)

References This article cites 1 articles, 1 of which can be accessed free at:
<http://genome.cshlp.org/content/10/3/271.full.html#ref-list-1>

License

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Cold Spring Harbor Laboratory Press

Unlimited Hypothesis Research

James C. Engert¹

Montreal Genome Centre, Montreal General Hospital Research Institute, Montreal, PQ, H3G 1A4, Canada

I read with great interest and a certain amount of perturbation the August Editorial by Laurie Goodman (1999). In sharp disagreement with the author, I believe that rather than being the “most efficient means of advancing our scientific understanding” hypothesis-free data collection would be completely devoid of any advancement at all. It is indeed the hypothesis that “follows the data”. No one would argue that a hypothesis springs *de novo*, but to believe that the data alone will formulate our further understanding is misguided.

Referring to the rendering of some monumental hypotheses as “discoveries of surprise,” Goodman further neglects to state that they were testable hypotheses and tested they were. Goodman states that some of these more famous scientific discoveries “met with resistance” because of the limitations provided by the current hypothesis. Hopefully the truly profound discoveries to come from high-throughput genomics research will also meet with resistance, for what a dull scientific world we would inhabit without the vigorous scientific debate that now accompanies the advancement of knowledge. The challenges to any new hypothesis and its limitations will only strengthen our understanding in the long run.

As Goodman points out, working too hard to fit or bend the data to a hypothesis has certainly always been a problem, as long as there have been scientists with egos or underlings yearning to please them. But most scientists are aware of the danger of attachment to ideas, and this drawback has certainly never stymied creative minds and those willing to take a fresh look.

¹E-MAIL jamie@med.mcgill.ca; FAX (514) 934-8353.

As the prime example of “hypothesis-free collections of data,” Goodman sites the Human Genome Project (HGP). This undertaking was certainly not initiated “without thought for what these data might be used.” Although we could not have envisioned all future uses, some were known at the outset. That this undertaking is “invaluable” cannot be argued, but this should not be confused with “at any cost.” Although it is true that the price of a data point for many high-throughput assays is dropping, it is incorrect to assume that the cost will ever be zero. For this example (HGP) and all to follow, the cost should be deemed too great if the resource and curatorial efforts take precedence over the design of experiments to test hypotheses. I do not think this will be the case.

Perhaps Goodman’s desire to dispense with hypothesis-driven science is born of a need to push current technological breakthroughs. *Genome Research* strives to be on the forefront of a revolution in technology, and this necessitates the defense of high-throughput methodologies. When we get over the excitement brought about by the avalanche of data, and it will only grow, the “era of pattern detection” will certainly be here. And exactly what will the era of pattern detection engender if not hypotheses? Some of these hypotheses will now concern which analysis method contains the most utility for the study of the genome and all biological systems. Many of these new hypotheses will come from an examination of methods that have already proved useful in other scientific disciplines. Other hypotheses will spring from perceptions impossible before the new era. Hopefully this era will not last too long if it prevents a new generation of young researchers from

watching their own hypotheses succeed and fail through rigorous experimentation.

Editor’s Response

The Editorial (Goodman 1999) was never meant to indicate that testable hypotheses should be completely eliminated from the scientific process—only that data collection should not be limited to only trying to collect information in pursuit of testing a hypothesis. This aspect was apparently not clearly presented in the Editorial. Thus, the points about generating testable hypotheses, as indicated by the author above, are important ones. [More points on how hypotheses should be tested are discussed in the next article (Łastowski and Makalowski 2000.)]

The author above also stresses that the costs of meaningless data may remain quite high if one were to pursue, wholeheartedly, a hypothesis-free data collection system in the sciences. The contention of the Editorial is that high-throughput methods may now make the costs of generating some meaningless data alongside scientifically useful data low enough that their impact on the overall costs will be negligible. This point, of course, remains debatable, as the author’s thoughts above clearly demonstrate, but I remain convinced that this is a debate worth pursuing. It should also be made clear that not all forms of data collection can be handled in a high-throughput manner and that the Editorial certainly did not mean to suggest that wholesale high-throughput data collection is how all scientific endeavors should proceed.

The author, in discussions with the editor after submission, also raised another important point: that lack of some forethought (i.e., hypotheses) on how data should be collected, can result in increased waste even in a high-throughput system—potentially to the point that none of the data generated

would be useful if the appropriate controls, etc., were not included at the outset. For example, the author indicated that it might be easy to initially obtain data from biopsies or other limited-supply samples or reagents, but then it would be difficult to return and obtain additional data afterwards (now with the hindsight of what would have really been useful). Agreed—but this can occur even with the best laid plans based on hypotheses (especially those that turn out to be erroneous). The author's point about

forethought when collecting data remains to be acknowledged; however, highly defined hypotheses about exactly what should be answered from such data may be equally counterproductive. The debate on this matter must continue as well.

Finally, although idealistically, lack of resistance to hypotheses that fly in the face of preconceived notions might make for "a dull scientific world" (it is indeed a fine romance when scientific "truth" at last conquers over falsehood), realistically (and, I

think, of an equally romantic nature) any scientific hypothesis is necessarily and excitingly debatable, as all hypotheses must always remain on the brink of falsification (see Łastowski and Makalowski 2000). We need never have fears of a dull scientific world.

Laurie Goodman

REFERENCES

- Goodman, L. 1999. *Genome Res.* **9**: 673–674.
Łastowski, K. and Makalowski, W. 2000. *Genome Res.* (this issue).