



Hypothesis-Limited Research

Genome Res. 1999 9: 673-674

Access the most recent version at doi:[10.1101/gr.9.8.673](https://doi.org/10.1101/gr.9.8.673)

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](#).

An advertisement banner with a teal background. On the left, the text reads "CRISPR and RNAi Genetic Screening. Your new superpower." in white. In the center, there is a white-bordered box containing the text "LEARN MORE". On the right, there is a photograph of a woman wearing a red and white superhero costume with a red mask. To the right of the photo is the Cellecta logo, which consists of a green molecular structure and the word "CELLECTA" in white capital letters.

CRISPR and RNAi Genetic Screening.
Your new superpower.

LEARN MORE

CELLECTA

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

Cold Spring Harbor Laboratory Press

Hypothesis-Limited Research

Students are taught that the proper way to do science is through the following steps: First, devise a hypothesis, and then design experiments that will prove or disprove their theory. The conscientious scientist follows this with a deliberate collection of data from these carefully crafted experiments, and the data prove or disprove the aforementioned theory, thereby placing a new stepping stone in the path through the jungle of our unknown universe.

It seems obvious that this is the appropriate way of moving forward in science. In fact, in most published papers, the hypothesis is put forth, followed by the experimental proof, and ending with a restatement of the veracity of the theory and potential future steps. Grants for funding research are also presented in such a light: a pure statement of theoretical intent, with a description of the experiments designed to determine the accuracy of the hypothesis. It is clear that pursuing an answer to a preformed theory is certainly a more cerebral quest than a mindless gathering of data, and it is much more appealing to the ivory-tower mentality in us all. But more importantly, it is obvious that just collecting data, with no hypothesis in mind at all, would be, in a word, wasteful. Gathering tremendous amounts of information with no thought to purpose would certainly provide information, but most of it would be so much clutter that the tremendous amount of time, energy, and money that would be required to gather all of the data one could imagine on a particular subject would be far too extensive to be worthwhile.

However, the current growth of high-throughput technologies for collecting some types of data may be tipping the balance with regard to the amount of waste that is really generated with regard to time, effort, and money [for examples of high-throughput expression data collection and analysis, see Eisen et al. (1998), Spellman et al. (1998), Chambers et al. (1999), Iyer et al. (1999), and Rhee et al. (1999)]. Reports such as these illustrate the growing ease with which data collection can be done. With data acquisition becoming so fast and growing cheaper by the day, perhaps the time has come to let go of the hypothesis part and simply take in every possible bit of data one can, only cataloging how it was taken and what the results were of every measure done. Previously, clever experiments were designed to put a hypothesis on a knife edge for dissection but also to provide the most cost-effective, straightforward way to get at the answer. But is it now actually more cost-effective to have large pools of data—any data—created and stored, without thought for what these data might be used? The sequencing of

genomes is one of the largest hypothesis-free collections of data the biological community has so far. As a community resource it is invaluable, and if it were not being done by the community as a whole, it would be far more expensive, take much more time (if it were done at all), and certainly be more wasteful in the redundancy of laboratories providing overlapping information. The costs of other types of data acquisition may now also be low enough that even the collection of data that might never be used would add so little to the overall cost as to make hypothesis-free data collection still the most efficient means of advancing our scientific understanding of biological systems.

Once pooled, the data can be examined in any variety of ways by anyone in the community and can tell the story about what is there. Just a voyage of discovery—no preconceived notion of what one might find—not unlike mapping some uncharted terrain, previously thought to end, perhaps, at the precipitous edge of the earth. Yes, one would need carefully designed tools for exploring, arranging, comparing, and cataloging. But the goal would be to sift through the data to find patterns that are present, not to bend the data to fit a theory as many have done, often unknowingly, in the past; as if they were writing in the center of an uncharted continent, “Monsters be here” and then imagination and belief ended up making it more difficult to determine the truth.

Perhaps it is also time to admit that the time-honored belief that good science constitutes first devising a hypothesis and then collecting data for proof does have a flaw—often ignored—one that crops up again and again in many a philosophy of science course; that is, that in reality it is really the hypothesis that follows the data. Hypotheses devised today for papers and grants actually stand on a tremendous foundation of data; they are not sprung from midair. The design of experiments to follow certainly does enable more testing of the theory, of course, but perhaps a communally available pool of data could take the place of many labs doing the same experiments and provide more unexpected discoveries than can be devised on fewer data.

This is, perhaps, the greater concern about clinging to hypothesis-driven research—the waste caused by what is missed. Most major scientific breakthroughs are the result of seeing unexpected patterns in data already gathered—patterns that might have been missed if one is bent on a set goal. For just a few examples, reach all the way back to Copernicus and the earth revolving around the sun, to Darwin and the

theory of evolution, up through Barbara McClintock and the discovery of transposons, and on to Tom Cech and self-processing mRNA. All were discoveries of surprise that the data alone revealed, and some of these discoveries met with resistance because of the limitations provided by the current hypotheses.

So perhaps the time has come to just do some mindless gathering of data. The cost seems to be growing less. The usefulness as a resource to the community appears quite high. Is this a heretical idea—to ease up just a bit on our perhaps errant belief that we knew all along what the data would tell us? To give up on proposing theory first and collecting data after? To do so would require a great number of changes, including how and whether nonhypothesis research is funded. But science, after all, is often called heretical for one reason or another. It seems worth gathering data to test this hypothesis and move into an era of pattern-

detection research rather than continuing to do research that might be hypothesis limited.

REFERENCES

- Chambers, J., A. Angelo, D. Amaratunga, H. Guo, Y. Jiang, J.S. Wan, A. Bittner, K. Frueh, M.R. Jackson, P.A. Peterson, M.G. Erlander, and P. Ghazal. 1999. *J. Virol.* **73**,r 5757–5766.
- Eisen, M.B., P.T. Spellman, P.O. Brown, and D. Botstein. 1998. *Proc. Natl. Acad. Sci.* **95**: 14863–14868.
- Iyer, V.R., M.B. Eisen, D.T. Ross, G. Schuler, T. Moore, J.C.F. Lee, J.M. Trent, L.M. Staudt, J. Hudson, Jr., M.S. Boguski, D. Lashkari, D. Shalon, D. Botstein, and P.O. Brown. 1999. *Science* **283**: 83–87.
- Rhee, C.H., K. Hess, J. Jabbur, M. Ruiz, Y. Yang, S. Chen, A. Chenchik, G.N. Fuller, and W. Zhang. 1999. *Oncogene* **18**:12711–2717.
- Spellman, P.T., G. Sherlock, M.Q. Zhang, V.R. Iyer, K. Anders, M.B. Eisen, P.O. Brown, D. Botstein, and B. Futcher. 1998. *Mol. Cell. Biol.* **9**: 3273–3297.

Laurie Goodman