Robot Scientist: an Autonomous Platform for Systems Biology Discovery

Professor R.D. King, University of Wales, Aberystwyth, Department of Computer Science. Mike Benway, Caliper Life Sciences Inc, Hopkinton MA, USA

Abstract

A "Robot Scientist" applies techniques from artificial intelligence to: originate hypotheses to explain data, devise experiments to test these hypotheses, physically run these experiments using laboratory robotics, interpret the results, and then repeat the cycle. We recently described the first proof-of-principle Robot Scientist in Nature. In conjunction with Caliper Life Sciences we have now built the first fully automated Robot Scientist. This is designed to: select frozen yeast strains from a freezer, inoculate these strains into a rich medium, harvest a defined quantity of cells, inoculate these cells into specified media (base plus added metabolites and/or inhibitors), and measure growth rate. The new Robot Scientist autonomously initiates over 1,000 experiments a day composed of unique combinations of strain/defined growth-media, and measures accurate growth curves by observing growth every < 25 minutes. This results in >100,000 data measurements a day, plus a similar number of meta-data measurements.

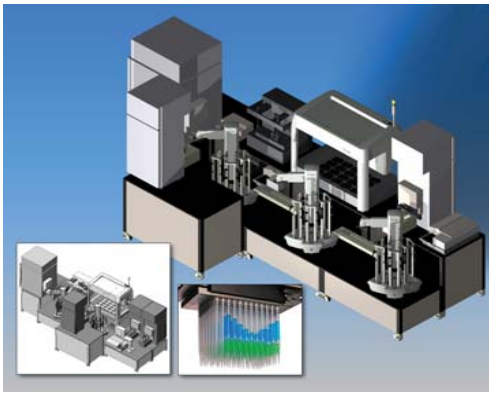
Discussion

A New Way to Do Science

Many interesting biological systems are combinatorial in nature; where a small number of factors combine in various ways to produce a large set of potential responses in an organism (see for example, Notes: 1,2,3,4,5 on the back page). Problems such as these are inherently unmanageable for human investigators beyond those involving only trivial subsets of experimental conditions. Advances in laboratory automation and informatics technologies have led to a situation in which the generation of hypotheses and the design of experiments can become new bottlenecks. By easily enabling 96 discrete, parallel dispense operations to be executed solely under machine control, and not requiring human knowledge of experimental conditions a priori, Caliper Sciclone i1000 helps to eliminate this bottleneck. While the current system is designed to generate new discoveries in *S. Cerevisiae* functional genomics, this general approach to autonomous experimentation could find applications in many other areas of cellular and molecular biology.

System Overview

Figure 1. The original Design for the Robot Scientist Platform. Since this was completed, a faster robot and an additional incubator were added to subsystem 3 in order to increase capacity and throughput In addition, some parts of subsystem 1 have been moved to subsystem 2 for potential future provisioning with temperature isolation zones. These changes are illustrated as an insert here. Not shown is the cage and HEPA filtration units built around the system to maintain sterile conditions. Caliper Sciclone i1000 manifold demonstrating its critical ability to dispense variable independent volumes using machine-readable input arguments is also shown in an insert.



Prifysgol Cymru Aberystwyth 1872 The University of Wales

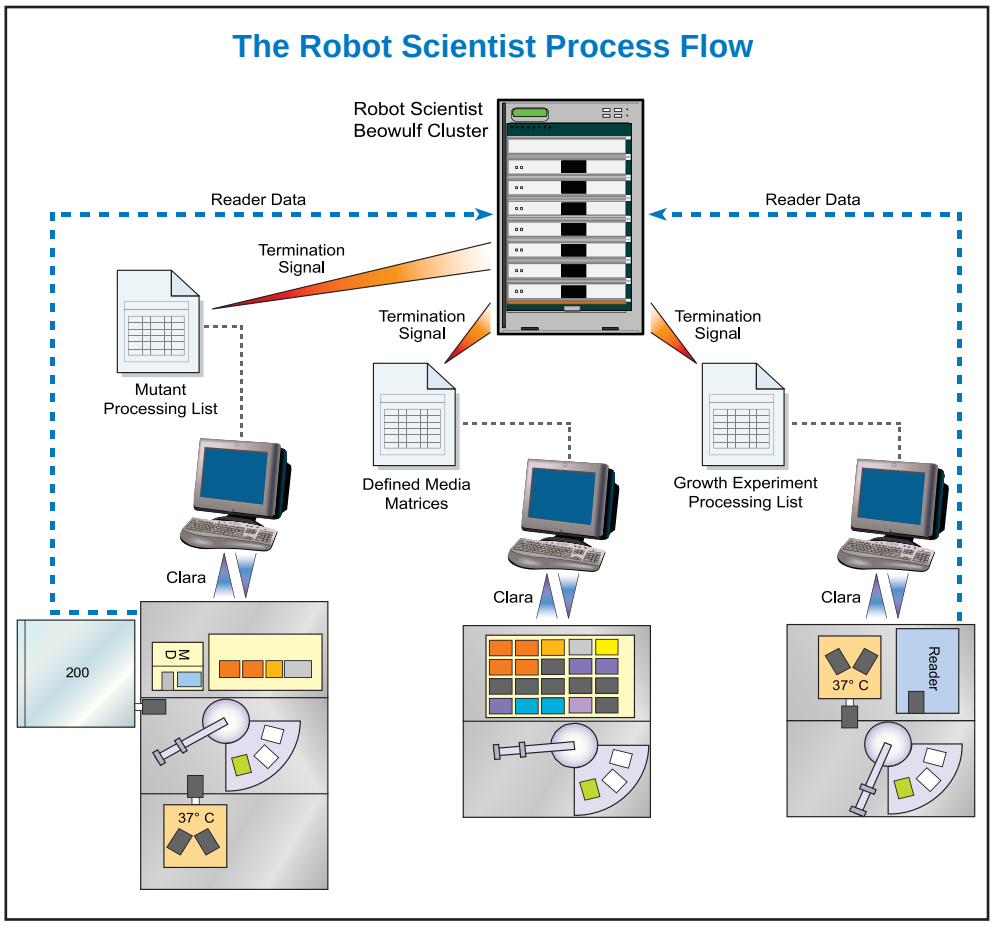
Caliper LifeSciences

Methods

A Closed-Loop Scientific Discovery Engine

The system is designed to support fully automated execution of auxotrophic growth experiments with yeast knockout mutants. This is a classical technique for investigating the function of metabolic genes using conditions in which wild type organisms will grow, but some mutant wills not (***Nature*** *427, 247-252*). The current system stores 6,000 frozen yeast knock-out strains (***Saccharomyces*** Genome Deletion Project). The Robot Scientist software emits a list of these strains to prepare, which are then retrieved, picked and inoculated into rich growth medium by the PLH (Process 1).

These picked colonies are incubated and monitored. A determination is made by the Robot Scientist Software whether to terminate this open-ended propagation process after each read. Once flagged as ready, colony plates are pelleted, washed, re-suspended and then quickly normalized in minimal medium in a single step using the Caliper Sciclone i1000. A set of bar code-associated matrix files for each experimental nutrient are deposited by the Robot Scientist Software which describe each distinct growth experiments' nutrient composition (the growth "cocktails") for each individual well. The Caliper Sciclone i1000 picks up these files to assemble 96 individual cocktails at a time, in parallel.

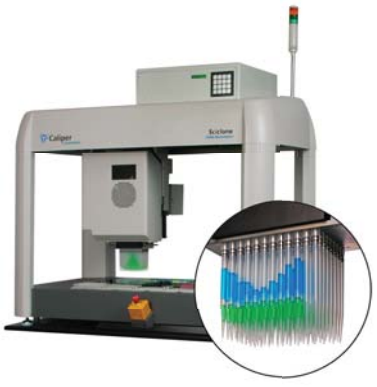


A standard injection of normalized yeast colonies is made into the associared cocktails. However, if any colony exhibited a terminal growth density too low to be normalized for a standard injection, the Caliper Sciclone i1000 will automatically increase inoculation volume and adjust the total volume of the corresponding nutrient cocktail accordingly. The Robot Scientist software is notified when a set of 96 growth experiments has begun (Process 2).

These arrayed growth experiments are incubated and read at intervals of <25 minutes in order to ensure that good rate measurements are made during exponential growth. A determination is made by the Robot Scientist Software whether to terminate this open-ended growth process after each read (Process 3).

The Robot Scientist software incorporates these results to re-weight its set of active hypotheses, and may order another round of experiments.

"Advances in laboratory automation and informatics technologies have led to a situation in which the generation of hypotheses and the design of experiments can become new bottlenecks. By easily enabling 96 discrete, parallel dispense operations to be executed solely under machine control, and not requiring human knowledge of experimental conditions a priori, Caliper Sciclone i1000 helps to eliminate this bottleneck"



All of these processes proceed independently and in parallel, and form a closed loop. The Caliper Sciclone i1000 tracks nutrient utilization, automatically tops up a working supply of each nutrient in lidded, cooled deck locations, enhances sterillity of cultures by using it's built-in gripper to manage lidded culture plates which are only ever de-lidded just at the moment of a liquid handling step, and provides notification of impending resource depletion to human maintainers. In essence, the inputs to the system are plates and media, and the outputs are true theories in functional genomics.

Methods (cont)

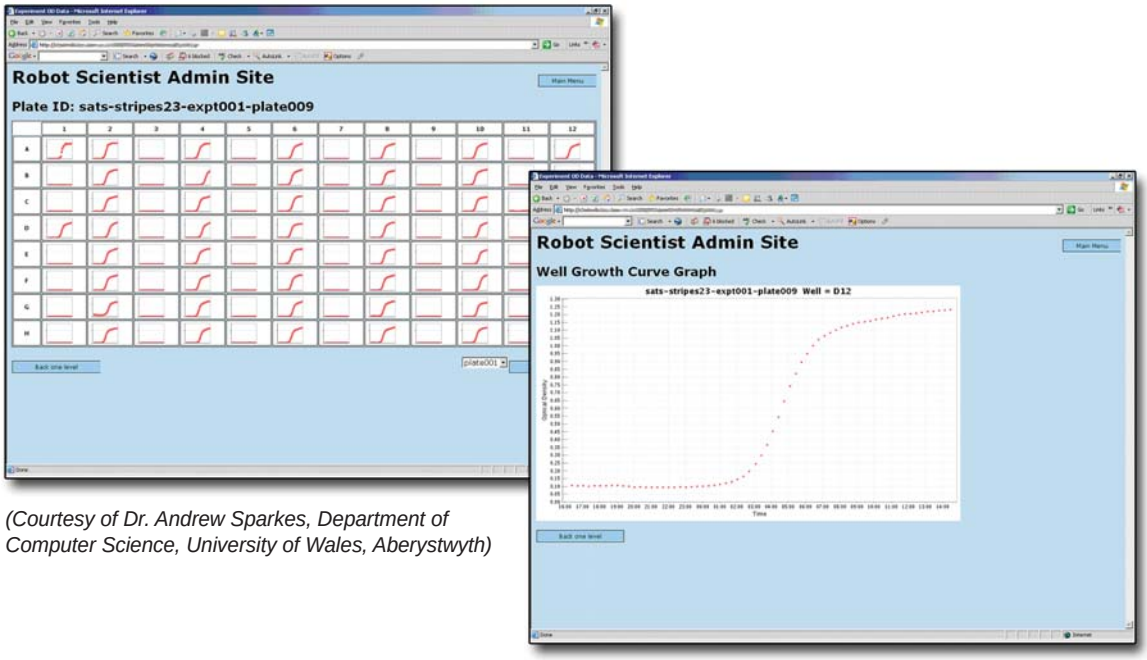


Figure 2. Web interface for browsing results generated by the Robot Scientist Platform for internal use. The interface is a development tool and is not strictly necessary since this data is interpreted entirely by artificial intelligence under normal operation.

(Courtesy of Dr. Andrew Sparkes, Department of Computer Science, University of Wales, Aberystwyth)

Loosely-Coupled Subsystems Achieve Independent, Dynamic Operation

Figure 3 The system is arranged as three distinct subsystems corresponding to distinct sub-processes which are loosely coupled. Each subsystem runs under a separate instance of iLink, and can execute a set of compatible, "atomic" operations conditionally. The three subsystems pass plates between them using a shuttle system, and maintain externally visible process status lists. This allows each subsystem to run independently while remaining aware of the state of the other subsystems and keeping the Robot Scientist software aware of the overall system state. The Robot Scientist (and the human maintainers) interact with and control the overall system by maintaining a master status list. Each subsystem conducts atomic operations in a way that is independent of, but sensitive to the state of the other subsystems, effectively achieving dynamic resource utilization.

Conclusions

We have implemented a system for an actively learning, artificially intelligent software which now orchestrates the autonomous execution of large-scale, hypotheses-driven, wet biology experimentation. The unique capabilities of Caliper Sciclone i1000 have helped enable this Robot Scientist. As life sciences move further into the exploration of more deeply complex phenomena, new ways of doing science that depend on more intelligent technologies such as these will become more essential and more common.

Notes:

1: Curr Opin Chem Biol. 2006 Aug;10(4):294-302. Epub 2006 Jul 5.
The application of systems biology to drug discovery
Cho CR, Labow M, Reinhardt M, van Oostrum J, Peitsch MC.
Department of Systems Biology, Genome and Proteome Sciences, Novartis Institutes of BioMedical Research, Cambridge MA 02139, USA

2: J Mol Biol. 2006 Jun 30;360(1):213-27. Epub 2006 May 3.
Comprehensive analysis of combinatorial regulation using the transcriptional regulatory network of yeast
Balaji S, Babu MM, Iyer LM, Luscombe NM, Aravind L.
National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda 20894, USA

3: BMC Bioinformatics. 2006 Mar 16;7:145.
Genetic algorithm learning as a robust approach to RNA editing site prediction
Thompson J, Gopal S. Department of Biological Sciences, Rochester Institute of Technology, Rochester, NY 14623, USA

4: Proteins. 2006 Apr 1;63(1):24-34.
Rough set-based proteochemometrics modeling of G-protein-coupled receptor-ligand interactions
Strombergsson H, Prusis P, Midelfart H, Lapinsh M, Wikberg JE, Komorowski J. Uppsala University,
The Linnaeus Centre for Bioinformatics, Uppsala, Sweden

5: BMC Bioinformatics. 2006 Jan 4;7:1.
Fast-Find: a novel computational approach to analyzing combinatorial motifs
Hamady M, Peden E, Knight R, Singh R. Department of Computer Science, University of Colorado, Boulder, CO 80309, USA

Caliper LifeSciences