

A peer-reviewed version of this preprint was published in PeerJ on 8 November 2017.

[View the peer-reviewed version](https://doi.org/10.7717/peerj.3997) (peerj.com/articles/3997), which is the preferred citable publication unless you specifically need to cite this preprint.

Groth P, Cox J. 2017. Indicators for the use of robotic labs in basic biomedical research: a literature analysis. PeerJ 5:e3997
<https://doi.org/10.7717/peerj.3997>

Indicators for the use of Robotic Labs in Basic Biomedical Research

Paul Groth¹ and Jessica Cox²

^{1,2}Elsevier Labs

Corresponding author:

Paul Groth¹

Email address: p.groth@elsevier.com

ABSTRACT

Robotic Labs, in which experiments are carried out entirely by robots, have the potential to provide a reproducible and transparent foundation for performing basic biomedical laboratory experiments. In this article, we investigate whether these labs are applicable in current experimental practice. We do this by text mining 1628 papers for occurrences of methods that are supported by commercial robotic labs. We find that 62% of the papers have at least one of these methods. This and our other results provide indications that robotic labs can serve as the foundation for performing many lab-based experiments.

INTRODUCTION

The reproducibility of a scientific experiment is an important factor in both its credibility and overall usefulness to a given field. In recent years there has been an up-tick in discussion surrounding scientific reproducibility, and it is increasingly being called into question. For example, Baker conducted a 2016 survey of 1500 researchers for Nature in which 70% were unable to reproduce their colleagues experiments (Baker, 2016). Furthermore, over 50% of the same researchers agreed that there was a significant crisis in reproducibility. While these issues arise in all fields, special attention has been paid to reproducibility in cancer research. Major pharmaceutical companies like Bayer and Amgen have reported the inability to reproduce results in preclinical cancer studies, potentially explaining the failure of several costly oncology trials (Begley and Ellis, 2012).

Munafò et al. (2017) outline several potential threats to reproducible science including p-hacking, publication bias, failure to control for biases, low statistical power in study design, and poor quality control. To address these issues, the Reproducibility Project: Cancer Biology in its reproduction of 50 cancer biology papers, used commercial contract research organizations (CROs) as well as a number of other interventions, such as registered reports (Errington et al., 2014). They argue that CROs provide a better basis for replication as they are both skilled in the expertise area and independent, in turn reducing risk of bias.

Extending this approach to providing an industrialized basis for performing experiments, is the introduction of large amounts of automation into experimental processes. At the forefront of this move towards automation is the introduction of “robotic labs”. These are labs in which the entire experimental process is performed by robots and available remotely in the cloud (Bates et al., 2016). A pioneering example of this is King’s Robot Scientist (King et al., 2009), which completely encapsulates and connects all the necessary equipment in order to perform microbial batch experiments; only needing to be provided consumables. Companies such as Transcriptic (<http://transcriptic.com>) and Emerald Cloud Lab (<http://emeraldcloudlab.com>) are beginning to make this same infrastructure in a commercial form.

The promise of these labs is that they remove the issues of quality control from individual labs and provide greater transparency in their operation. Additionally, they allow for biomedical experiments to become more like computational experiments where code can be re-executed, interrogated, analyzed and reused. This ability to have a much more detailed computational view is critical for reproducibility as narrative descriptions of methods are known to be inadequate for this task as summarized in (Gil and Garijo, 2017). This lack of detail is illustrated compellingly in the work on reproducibility maps

where it took 280 hours to reproduce a single computational experiment in computational biology (Garijo et al., 2013). While there are still challenges to reproducibility even within computational environments (Fokkens et al., 2013), robotic labs potentially remove an important variable around infrastructure. They provide, in essence, a programming language for biomedical research.

While this promise is compelling, a key question is whether robotic labs would be widely applicable to current methods used in biomedical research. This question can be broken down into two parts:

1. does basic lab-based biomedical research reuse and assemble existing methods, or is it primarily focused on the development of new techniques? and;
2. what existing methods are covered by robotic labs?

To answer this question, we use an approach inspired by Vasilevsky (Vasilevsky et al., 2013) that used text analysis of the literature to identify resources (e.g. cell lines, reagents). Concretely, we automatically extract methods from a corpus of 1628 open access papers from a range of journals covering basic biomedical research. We identify which of those methods are currently supported by robotic labs. Our results show that that 62% of these papers have some methods that are currently supported by cloud-based robotic labs.¹

MATERIALS & METHODS

Article Corpus Construction

Our aim was to construct a meaningfully sized corpus that covered representative papers of basic lab-based biomedical research. Additionally, for reasons of processing efficiency we selected papers from Elsevier because we had access to the XML versions of the paper in a preprocessed fashion. To build our corpus, we first selected journals categorized under "Life Sciences" in ScienceDirect², specifically those marked under "Biochemistry, Genetics and Molecular Biology". We then filtered for journals categorized as "Biochemistry", "Biochemistry, Genetics and Molecular Biology", "Biophysics", "Cancer Research", "Cell Biology", "Developmental Biology", "Genetics", or "Molecular Biology". This returned a list of 412 journals. We then manually inspected each journal on this list. Journals were excluded if they were comprised of seminars or reviews, were non-English, primarily clinical studies, primarily new methods, population studies or a predecessor to another journal. ISSNs were returned for each title, for a final list of 143 journals. The list of journals selected with their ISSN are available at Groth and Cox (2017).

From these journals, we selected CC-BY licensed papers. The list of papers and their DOIs is available at Groth and Cox (2017) which includes a script to download the corpus.

Method Space Definition

To define the space of methods, we relied upon the 2015 edition of the National Library of Medicine's Medical Subject Headings (MeSH) controlled vocabulary. MeSH provides a number of benefits: One, it provides an independent definition of a set of possible methods. Two, it provides a computationally friendly definition covering multiple synonyms for the same method concept that researchers could potentially use. For example, it defines synonyms for Polymerase Chain Reaction such as PCR, Nested PCR, and Anchored Polymerase Chain Reaction. Third, because it is arranged hierarchically, it captures methods at different levels of granularity. For example, a researcher may use PCR but not identify the specific variant like Amplified Fragment Length Polymorphism Analysis. Thus, we took the Investigative Techniques [E05] branch of MeSH as defining the total space of methods. For use in our analysis, we extracted that branch from the Linked Data version of MeSH³ using a SPARQL query. This branch of MeSH contained 1036 total concepts. The SPARQL query, the branch reformatted as a CSV file and a link to the specific linked data version are available in Groth and Cox (2017).

To define what methods could be automated by a robot lab, we built a list of available and soon to be available methods from the Transcriptic and Emerald Cloud Lab websites as of March 10, 2017. This list contained 107 methods. We term methods that can be executed within a robotic lab a **robotic method**. We manually mapped those lists to MeSH concepts from the Investigative Techniques [E05] branch. We were able to map 71 methods to MeSH concepts. During the mapping procedure, we selected leaf nodes

¹Data and Code are available at <http://dx.doi.org/10.17632/g7bfzcgdyd.1> and referenced throughout.

²<http://sciencedirect.com>

³<https://id.nlm.nih.gov/mesh/>

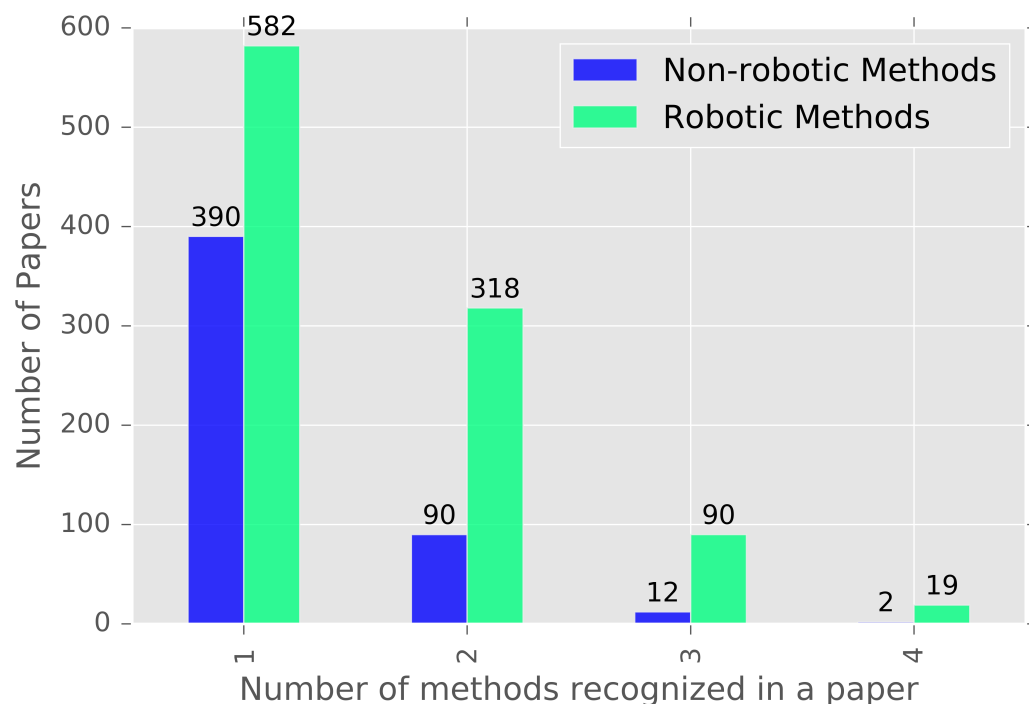


Figure 1. The distribution of the count of unique methods per paper categorized by type.

of the tree and used the MeSH preferred concept as the target. In some cases, this meant that a particular method was mapped to a more general method type. Our final list of robotic methods mapped to MeSH contains 59 concepts. The complete mapping is also available at Groth and Cox (2017).

Those methods that were not mapped to a robotic method but were tagged with a MeSH investigative technique are termed a **non-robotic method**.

Method Identification

To identify methods mentioned in the corpus, we use Solr Dictionary Annotator (SoDA) (Pal, 2015) - a flexible, scalable lexicon based annotator that provides convenient integrations with Apache Spark (a distributed computing environment). Using SoDA, we annotated all content paragraphs (excluding abstracts, titles, section headings, figures, and references) against the whole of MeSH 2015 using SoDA's exact setting. Adopting such a dictionary based approach translates to high precision in method identification, sacrificing recall. Using this approach means that we cannot determine complete coverage of all methods used in a paper.

After annotation, analysis was performed by matching the lists detailed above with the output annotations. The analysis procedure code is available in Groth and Cox (2017).

RESULTS

Within our 1628 article corpus, 1165 of those articles were identified to have at least one method as defined by matching to an MeSH investigative technique. In total, we identified 151 unique methods used across the corpus.

Using the mapping to robotic labs discussed above, we identified 1011 articles or roughly 62% of the total corpus have at least one method that can be executed within a known robotic lab. Of the 1165 papers where the procedure recognized a method, the mean number of robotic methods within an article is 1.5. Figure 1 shows the number of times a robotic method or a non-robotic method occur within a paper. For example, in roughly 19 papers, an robotic method occurs 4 times.

Table 1. Occurrence of robotic methods

Method Name	Count
Polymerase Chain Reaction	746
Enzyme-Linked Immunosorbent Assay	165
Chromatography, High Pressure Liquid	110
Transfection	101
Immunoprecipitation	71
Real-Time Polymerase Chain Reaction	67
Microscopy	64
Flow Cytometry	47
Cell Culture Techniques	39
Spectrometry, Mass, Matrix-Assisted Laser Desorption-Ionization	32
Crystallization	25
Microscopy, Electron	23
Blotting, Western	21

Table 2. Occurrence of non-robotic methods

Method Name	Count
Passive Cutaneous Anaphylaxis	47
Immunohistochemistry	45
In Situ Nick-End Labeling	36
Immunoblotting	34
Mass Spectrometry	31
Mutagenesis, Site-Directed	22
Animal Experimentation	22
Data Collection	19
Electrophoresis	15
Insemination, Artificial, Heterologous	15

Table 1 lists robotic methods that occur in more than 15 papers. Of the 59 potential robotic methods, 33 occurred within our corpus. We analyze this list in more detail later in the discussion section.

Additionally, as discussed we identified the most common non-robotic methods. There were 118 unique non-robotic methods in total, and methods appearing in at least 15 papers are presented in table: nonautomethodnum.

We note that robotic methods appear more frequently in articles. For example, the most frequently occurring robotic method occurs in 15 times more articles than the most frequently occurring non-robotic method.

DISCUSSION

We return to our initial questions: 1) do basic biomedical papers reuse existing methods and, 2) if so, are those methods supported by robotic labs.

With respect to our first question, our analysis suggests that biomedical research papers do reuse existing methods. 71% of the papers had at least one known method as listed within MeSH. Interestingly, of the potential 1035 methods only 151 were recognized. One could take this skew as evidence that a small number of highly common methods are being employed in practice. However, this skew could be the result of a number of other variables including the recognition algorithm used, the level of reporting by scientists in their papers (e.g. ignoring methods that are widely used), and the coverage of method synonyms by MeSH. From a more qualitative perspective, we see that common techniques are recognized. For example, it is unsurprising that the most common robotic method is PCR, shown in Table 1. PCR is a relatively standardized and cost-effective method used ubiquitously in biomedical research. Its elegant yet straightforward protocol lends itself to be used in a variety of contexts within a biomedical lab: from gene expression measurement to cloning. Current thermocycler technology enables easy adjustment of

experimental parameters, relatively little sample handling and the use of commercialized master mixes. Combined with its pervasiveness in biomedical research labs, these factors make PCR an attractive choice for automation.

With the exception of cell culture, the other methods in Table 1 are also comprised of highly automatable tasks. Just as thermocycler technology is relatively standardized, so too are the equipments, kits and protocols used for methods like HPLC and ELISAs. Biomedical labs are using nearly identical protocols in many instances, yet introducing their own variability due to human use. In these cases, robotic automation would facilitate quick execution of the same method for all of these labs, increasing transparency and reproducibility. This argument can be extended to nearly all of the methods within the table. Simply stated, robots can pipette, measure and handle samples better than humans can, and in turn facilitate reproducible science.

Table 2 represents the most commonly identified non-robotic methods. We combed through the list of 119 unique methods and crossed them with the list of methods currently provided by Transcriptic and Emerald. Twenty-six of these methods (22%) are in fact supported by one of these cloud labs, exposing some "leakiness" in our procedure. Our annotation procedure does not attempt to generalize: the MeSH ID a method is labeled with is exactly what is returned in a search for the method without traversing the hierarchy of the method tree. This explains why Real-Time Polymerase Chain Reaction appears in this list, and is not grouped in with Polymerase Chain Reaction, the most common robotic method in Table 1. Additionally, many of the methods tagged are not applicable in the context of a biomedical laboratory pipeline. For example, "passive cutaneous anaphylaxis" refers to a clinical event, but reflects the nature of MeSH as a information management vocabulary as well as potential outliers in our document corpus.

In terms of the second question, our analysis suggests that the research represented by this corpus of literature has the potential for using robotic labs in at a least some aspects of the described experimental processes. Indeed, looking at the coverage of methods found, one sees that over half of the methods indexed have some automated equivalent. This figure is striking in that robotic labs are still just becoming available for use. This large overlap could have to do with what the skewed distribution of recognized methods as described above.

Looking more deeply at the actual methods identified, the top robotic methods in Table 1 are a mix of both workflow techniques (i.e. cell culture, transfection) and endpoint measurements (i.e. qPCR, ELISA). Roughly 6% of our corpus had more than 3 robotic methods within one paper, which we believe to be an underestimation. This qualitative view provides some support that robotic methods can execute the majority of an end to end biomedical workflow. One may argue that robotic labs do not lend themselves to the building of a disease model. Building a model requires extensive experimentation and parameter tweaking, and some argue that this kind of platform is more conducive to endpoint analysis after a model has been rigorously developed and tested, and not its actual development. However, we contend that with some more work, a robotic lab that does support every part of the workflow would actually accelerate model system development and allow researchers to spend more time testing and developing new hypotheses. This outcome would be the consequence of allowing essentially what is parameter search to be performed by the robot with minimal human interaction during experimental execution. This could accelerate the pace of discovery in entire fields, all while maintaining reproducibility.

While these results provide salient indicators for the ability to move towards robotic labs, there are a number areas where our analysis could be improved.

Our analysis does not provide information about whether the given automated methods cover all aspects of the protocols described within an article. This incompleteness comes from four sources:

1. the identification algorithm biases towards precision rather than recall (e.g. it does not perform fuzzy matching);
2. the identification relies on a manually created list (i.e. MeSH that is necessarily incomplete);
3. the recognition algorithm does not determine how the methods/steps that are recognized join up to form a total protocol, this includes how materials are physically transferred between steps;
4. papers will frequently not mention steps or smaller parts of protocols that are necessary but are well known to trained researchers;

To address the above, we would need much more complex natural language processing techniques. Indeed, the state of the art in process/task detection (a similar task to method recognition) is only 0.44 F1⁴ that is not including recognizing the dependency relations between the tasks. In biology specific method extraction state-of-the-art ranges between roughly 0.6 and .7 F1 (Burns et al., 2016). Recent work by Dasigi et al. (2017) shows the effectiveness of deep learning approaches on the larger scientific discourse extraction, however, this was applied only to a small number of papers. In future work, we aim to apply these recent advances to deepen our analysis. Based on the challenges listed above, we believe that the numbers presented here are an underestimation of the total number of robotic methods that can be applied in biomedical research.

Finally, While we believe the selected corpus reflects the body of literature that would most likely use robotic labs, it could be argued that a much larger corpus would be more informative. This investigation is also left to future work.

CONCLUSION

Reproducibility is of increasing concern across the sciences. Robotic labs, particularly in biomedicine, provide the potential for reducing the quality control issues between experiments while increasing the transparency of reporting. In this article, we analyzed a subset of the biomedical literature and find that greater than 60% of the papers have some methods that are supported by existing commercial robotic labs. Furthermore, we find that basic methods are indeed “popular” and are increasingly being covered by robotic labs.

While there will always be labs that specialize in the development of new methods, given these indicators, we believe that robotic labs can provide the basis for performing a large percentage of basic biomedical research in a reproducible and transparent fashion.

Acknowledgements

We thank Ron Daniel and Brad Allen for their helpful comments.

REFERENCES

- Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, 533(7604):452–454, ISSN: 0028-0836, DOI: 10.1038/533452a, <http://www.nature.com/doi/10.1038/533452a>.
- Bates, M., Berliner, A. J., Lachoff, J., Jaschke, P. R. A., and Groban, E. S. (2016). Wet Lab Accelerator: A Web-Based Application Democratizing Laboratory Automation for Synthetic Biology. *ACS Synthetic Biology*, page acssynbio.6b00108, ISSN: 2161-5063, DOI: 10.1021/acssynbio.6b00108, <http://pubs.acs.org/doi/abs/10.1021/acssynbio.6b00108>.
- Begley, C. G. and Ellis, L. M. (2012). Drug development: Raise standards for preclinical cancer research. *Nature*, 483(7391):531–533, ISSN: 0028-0836, DOI: 10.1038/483531a, <http://www.nature.com/doi/10.1038/483531a>.
- Burns, G. A., Dasigi, P., de Waard, A., and Hovy, E. H. (2016). Automated detection of discourse segment and experimental types from the text of cancer pathway results sections. *Database*, 2016:baw122, ISSN: 1758-0463, DOI: 10.1093/database/baw122, <https://academic.oup.com/database/article-lookup/doi/10.1093/database/baw122>.
- Dasigi, P., Burns, G. A. P. C., Hovy, E., and De Waard, A. (2017). Experiment Segmentation in Scientific Discourse as Clause-level Structured Prediction using Recurrent Neural Networks, <https://arxiv.org/pdf/1702.05398.pdf>.
- Errington, T. M., Iorns, E., Gunn, W., Tan, F. E., Lomax, J., Nosek, B. A., Gilbert, K., Moore, J., Renaut, S., Rennison, D., Laitin, D., Madon, T., Nelson, L., Nosek, B., Petersen, M., Sedlmayr, R., Simmons, J., Simonsohn, U., Laan, M. V. d., Huguenard, J., Kelner, K., Koroshetz, W., Krainc, D., Lazic, S., Levine, M., Macleod, M., McCall, J., Moxley, R., Narasimhan, K., Nobel, L., Perrin, S., Porter, J., Steward, O., Unger, E., Utz, U., and Silberberg, S. (2014). An open investigation of the reproducibility of cancer biology research. *eLife*, 3:726–728, ISSN: 2050-084X, DOI: 10.7554/eLife.04333, <http://www.ncbi.nlm.nih.gov/pubmed/25490932><http://www.ncbi.nlm.nih.gov/pubmed/25490932>

⁴See 2017 SemEval Task 10 <https://scienceie.github.io>

- 241 //www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4270077http:
242 //elifesciences.org/lookup/doi/10.7554/eLife.04333.
- 243 Fokkens, A., van Erp, M., Postma, M., Pedersen, T., Vossen, P., and Freire, N. (2013). Offspring from
244 Reproduction Problems: What Replication Failure Teaches Us. In *Proceedings of the 51st Annual*
245 *Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1691–
246 1701, Sofia, Bulgaria. Association for Computational Linguistics, <http://www.aclweb.org/anthology/P13-1166>.
- 247
248 Garijo, D., Kinnings, S., Xie, L., Xie, L., Zhang, Y., Bourne, P. E., and Gil, Y. (2013). Quantifying
249 Reproducibility in Computational Biology: The Case of the Tuberculosis Drugome. *PLoS ONE*,
250 8(11):e80278, ISSN: 1932-6203, DOI: 10.1371/journal.pone.0080278, <http://dx.plos.org/10.1371/journal.pone.0080278>.
- 251
252 Gil, Y. and Garijo, D. (2017). Towards Automating Data Narratives. In *Proceedings of the 22nd*
253 *International Conference on Intelligent User Interfaces - IUI '17*, pages 565–576, New York, New
254 York, USA. ACM Press, ISBN: 9781450343480, DOI: 10.1145/3025171.3025193, <http://dl.acm.org/citation.cfm?doid=3025171.3025193>.
- 255
256 Groth, P. and Cox, J. (2017). Datasets for Potential of Robotic Lab Methods Usage in Biomedical Papers.
257 DOI: <http://dx.doi.org/10.17632/gy7bfzcgzd.1>.
- 258
259 King, R. D., Rowland, J., Oliver, S. G., Young, M., Aubrey, W., Byrne, E., Liakata, M., Markham,
260 M., Pir, P., Soldatova, L. N., Sparkes, A., Whelan, K. E., and Clare, A. (2009). The Automation
261 of Science. *Science*, 324(5923):85 LP – 89, <http://science.sciencemag.org/content/324/5923/85.abstract>.
- 262
263 Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Per-
264 cie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., and Ioannidis, J. P. A.
265 (2017). A manifesto for reproducible science. *Nature Human Behaviour*, 1(1):0021,
266 ISSN: 2397-3374, DOI: 10.1038/s41562-016-0021, <http://www.nature.com/articles/s41562-016-0021>.
- 267
268 Pal, S. (2015). Solr Dictionary Annotator (SoDA). DOI: 10.5281/ZENODO.48974, <https://zenodo.org/record/48974#.WNFN7mVql69>.
- 269
270 Vasilevsky, N. A., Brush, M. H., Paddock, H., Ponting, L., Tripathy, S. J., LaRocca, G. M., and Haendel,
271 M. A. (2013). On the reproducibility of science: unique identification of research resources in the
272 biomedical literature. *PeerJ*, 1:e148, ISSN: 2167-8359, DOI: 10.7717/peerj.148, <https://peerj.com/articles/148>.