# Wisdom of Crowds for Diagnostics Signature Challenge

Julia Hoeng[1], Marja Talikka[1], Pablo Meyer Rojas[2], Raquel Norel[2], John J Rice[2], Jörg Sprengel[3], Gustavo Stolovitzky[2], M. Peitsch[1], [1]Philip Morris Products SA, Neuchâtel, Switzerland; [2]IBM Computational Biology Center, Yorktown Heights, NY, USA; [3]IBM Life Sciences Division, Zurich, Switzerland

## Introduction

We propose a method for the development of computational algorithms and the verification of the intended performance of these algorithms. The method is based on the concept of collaborative competition and consists of presenting the research community with a scientific question or Challenge. These researchers, motivated by rewards to the best solution and by the importance and relevance of the challenge, work independently towards the answer of the same scientific question. Rigor in determining the best solution is attained through the objective assessment of the submissions by third party evaluation. Aggregation of the submitted results often outperform any of the individual submissions, a phenomenon that we call the wisdom of crowds

Considering cancer being the leading cause of death worldwide, we formulate the Diagnostics Signature Challenge to evaluate novel approaches for the identification of robust and predictive disease signatures. The goal of a Diagnostics Signature Challenge is to verify that transcriptomics data contains enough information for the determination and prognosis of certain human disease states that could profit from better diagnostics and prognosis signatures.

Here we will describe the approach, the necessary operational steps and how we intend to engage the community to assess the applicability of this approach to discover and verify disease signatures.

The proposed scheme can be extended to verify complex industrial and academic research workflows, through a methodology that we call IMPROVER[1], for Industrial Methodology for Process Verification of Research

## Self Assessment of Computational Methods often Leads to Biases

Self assessment is suspect[2]
✓ Authors' method tends to be the best in an unreasonable majority of cases
  ✓ selective reporting of performance: inadvertent or disingenuous
  ✓ choice of only one, best metric
  ✓ information leakage
  ✓ overfitting
  ✓ parameter tinkering
Proposed solution:
  ✓ impartial assessment of predictions on unseen data



- First (self assessment) 87%
- First (independent assessment) 6%
- Not first (independent assessment) 7%
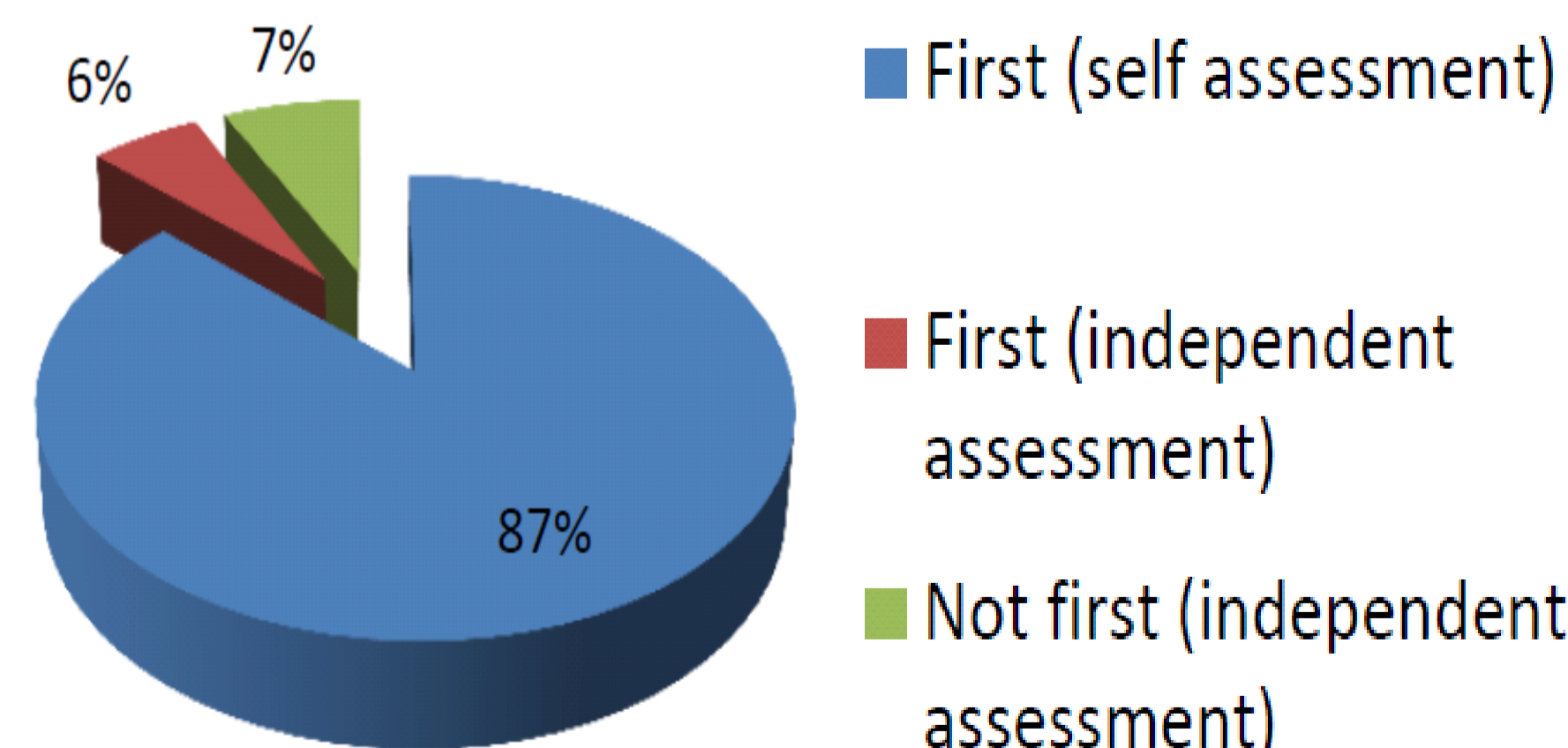
Percentage of papers in which the authors rank their method as first or not first using self- or independent assessment. Out of the 69 papers reviewed we did not find any self-assessment paper where the presented method was not top ranked in at least one metric.

## Challenges for Impartial Assessment of Computational Methods

A Challenge is a scientific problem presented to the community. It has the following requirements for its definition:
✓ Need of gold standard or ground truth: the solution to the challenge has to be known at least in one instance.
✓ At least one metric of performance for evaluation of the submitted challenge solutions has to be defined a priori.
✓ Not always clear which metric is best
Challenges have the following advantages:
✓ Nucleates a community around a given scientific problem.
✓ Allow for easy comparison of performance of different methods on the same data set.
✓ Allow to establish the state-of-the-art technology in a field, and find complementary methods to solve a problem.

## Wisdom of Crowds

Aggregation of solutions has been shown to dramatically improve results[3]
✓ Challenges may have many participants (crowds).
✓ Participants produce independent methods and submit different solutions which tackle different aspects of a complex problem.
✓ The assembly of the solutions of all participants often outperforms the best performing submissions, a phenomenon we call wisdom of crowds.

## Disease Signature Challenge for Diagnosis and Prognosis

The identification of clinically useful predictive signatures for cancer has proven challenging, and many initially promising signatures failed to translate into clinically useful applications. Why?
✓ Poor reproducibility in gene expression measurements
✓ Non-specific markers
✓ Poor statistical experimental design
✓ Poor computational methods
Given the importance of discovery effective cancer signatures we formulate the Diagnostic Signature challenge: Does transcriptomics data contain enough information for the determination and prognosis of certain cancers?
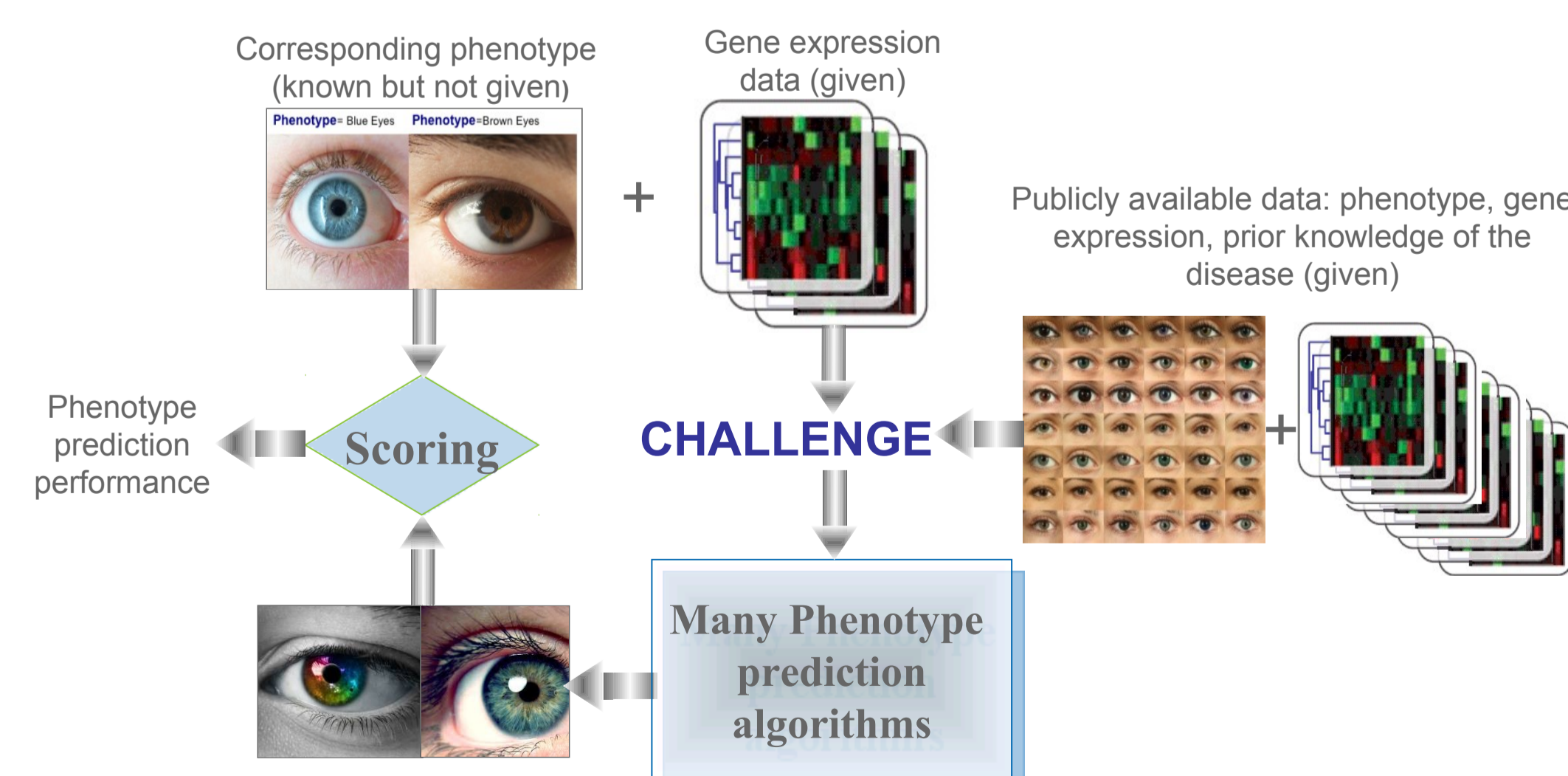


Scheme of the disease signature challenge. The eye color is a representation of the subject phenotype, which in the actual challenge will be survival, cancer subtype or stage.

## Signatures May be More than a List of Over and Under-Expressed Genes

If the challenge has interesting data, crowd sourcing should lead to many participants submitting predictions to our challenge. Amongst them we expect to see non-intuitive solutions, created from attacking the problem from out-of-the-box thinking. For example, we may find robust solutions such as:
✓ Generalized signatures, such as "master regulators". These are genes which causally drive the disease, even though they may not even be significantly expressed in the gene expression array.
✓ The signature may represent markers of an inferred gene regulatory network.
✓ Novel approaches may contribute to the discovery of more accurate and robust predictors of disease and clinical responses, that are independent of the measurement platform.

## Verification by Challenge in Industry R&D and Large Collaborative Projects

✓ The concept of a challenge can be generalized to help discover better solutions and/or verify crucial steps in Industrial R&D or large collaborative projects.
✓ A complex research program is typically built upon research projects ("building blocks") that synergistically support each other towards a final goal.
✓ A building block is a standalone research process at the smaller scales of a complex workflow with a defined input, resulting in a defined output (e.g. Identify targets of a Transcription factor, or discover SNPs associated with disease.)
✓ We proposed The Industrial Methodology for Process Verification in Research[1] (IMPROVER), which is a methodology designed to
  ✓ Verify the robustness of R&D scientific practices in Systems Biology.
  ✓ Estimate the risk associated with the outcome of a research workflow by dividing it into building blocks which are verifiable by challenge.
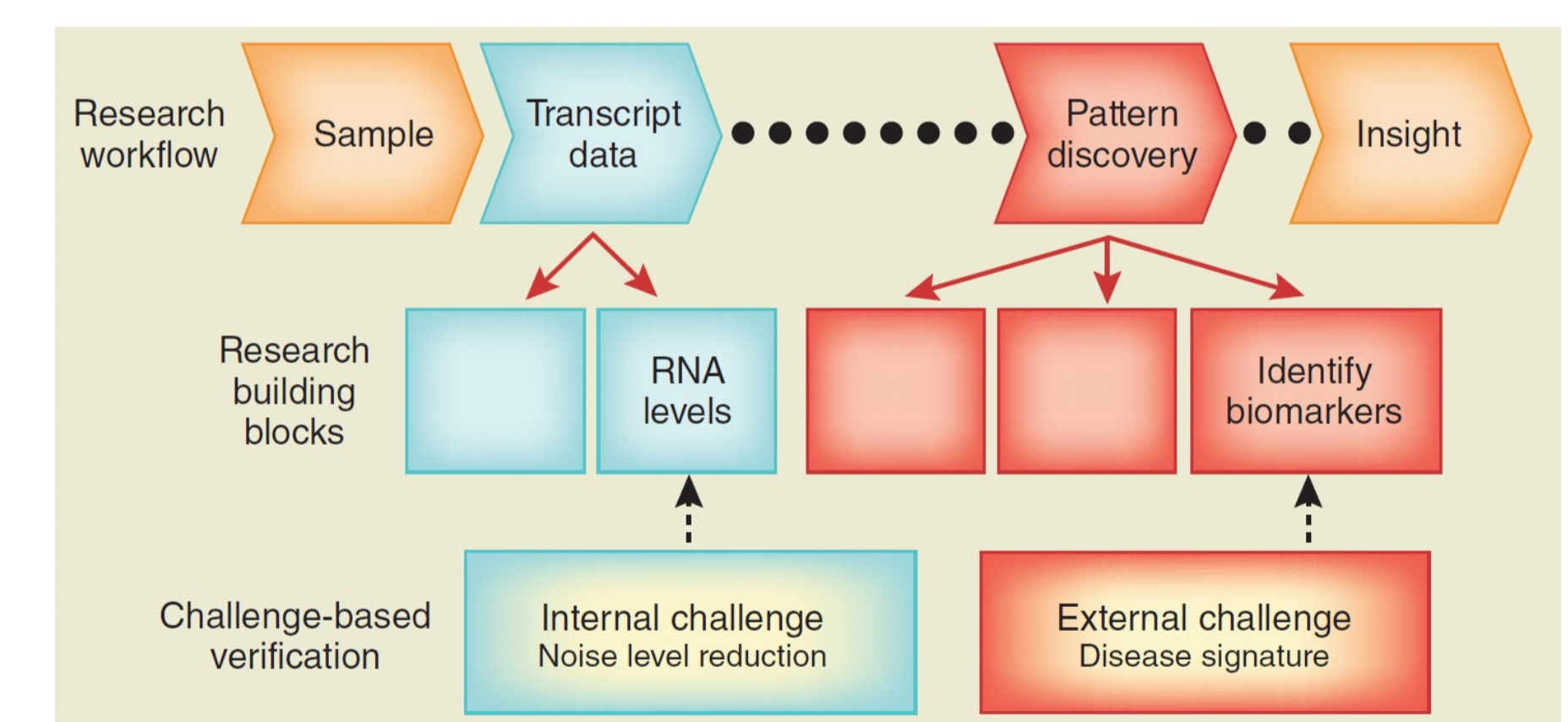


Example application of IMPROVER for verification of a plausible research workflow.

## References

1) Meyer P. et al (2011) submitted **Industrial Methodology for Process Verification in Research (IMPROVER): Towards Systems Biology Verification**

2) Norel R. et al (2011) in press, **The self-assessment trap: Can we all be better than average?** Mol. Sys. Bio.

3) Marbach D. et al., Proc Natl Acad Sci U S A 107, 6286 (2010) **Revealing strengths and weaknesses of methods for gene network inference**.

4) Prill R.J. et al., PLoS ONE, 5(2):e9202 (2010) **Towards a Rigorous Assessment of Systems Biology Models: The DREAM3 Challenges**.