Φ-score, a Robust Cell-by-cell Score for Sensitive and Specific Hit Discovery in RNA interference High Content Screening

Laurent Guyon*†[1], Christian Lajaunie[2,3], Frédéric Fer[1], Ricky Bhajun[1], Eric Sulpice[1], Guillaume Pinna[4], Anna Campalans[5], J. Pablo Radicella[5], Stéphanie Combe[1], Patrica Obeid[1], Jean-Philippe Vert[2,3], and Xavier Gidrol‡[1]

[1]Laboratoire de Biologie à Grande Échelle (BGE - UMR S1038) – CEA, Inserm : U1038, Université Grenoble Alpes – CEA -Grenoble/IRTSV Laboratoire Biologie à Grande Echelle 17 Rue des Martyrs 38054 GRENOBLE CEDEX 9, France

[2]Centre de Bioinformatique (CBIO) – MINES ParisTech - École nationale supérieure des mines de Paris – 35 rue Saint-Honoré 77300 Fontainebleau, France, France

[3]Institut Curie – Institut Curie – 26 rue d'Ulm 75248 PARIS CEDEX 05, France

[4]Institut de Biologie et de Technologies de Saclay (IBITECS) – CEA – CEA Saclay bat 532 91191 Gif sur Yvette cedex, France

[5]Radiobiologie moléculaire et cellulaire (RMC) – CNRS : UMR217, CEA – avenue du général leclerc BP 6 92265 FONTENAY AUX ROSES CEDEX, France

Résumé

High Content Screening (HCS) has enabled great advances both in oncology and biology. It consists in visualizing phenotypes modification of cells after perturbation. This perturbation is generally achieved either by chemical compounds or RNA interference in a highly parallel manner (in 384 well-plates for example). RNA interference screening is an efficient methodology to investigate genotype-phenotype relations at a system level. HCS experiments produce huge amount of data, typically tables of millions of rows and tens of columns; each row corresponding to a cell whose phenotype is characterized with different metrics (each column). After analysis, hits, which are the biggest modifiers of the cell phenotype, are extracted.
A method of choice, Z-score, is often used after averaging the quantified phenotype for each cell modified by a given perturbation. It has proven great efficiency. However it faces a few drawbacks:

Z-score is very sensitive to artifacts (aberrant phenotype of just one cell will affect the score)

Low cell number for a given treatment will strongly increase Z-score variability

Z-score is associated to a p-value only when the distribution is Gaussian

*Intervenant
†Auteur correspondant: laurent.guyon@cea.fr
‡Auteur correspondant: xavier.gidrol@cea.fr

In situations where the consequences of perturbation are strong and overtake such "artifacts", Z-score finds hits with a low False Discovery Rate. However in conditions closer to the physiological reality (such as rare cells or cells from patient), one needs to overcome these pitfalls. Thus, in an attempt to improve the potential of discovery of HCS, we developed a so called Φ-score, which is based on the rank of the quantified phenotype and takes into account the number of cells in a given treatment with an appropriate model.

In order to compare the performance of Φ-score related to Z-score and other commonly used scores in HCS, we have performed simulations in various situations, an experiment with controlled conditions and both primary and secondary small interfering RNA (siRNA) screens targeting more than 7000 genes. The siRNA screens aimed at identifying designed to identify new factors participating in oxidized DNA repair, in particular those that affect the recruitment of the repair enzyme OGG1.
On one hand, both simulations and the dedicated experiment are performed in controlled conditions. The Φ-score is here shown to prove better efficiency and sensitivity. On the other hand, no conditions were available as internal control for the 'OGG1 screen'. Ontology enrichments and comparison between both screens permitted to show in which extent the Φ-score over-performed the others.