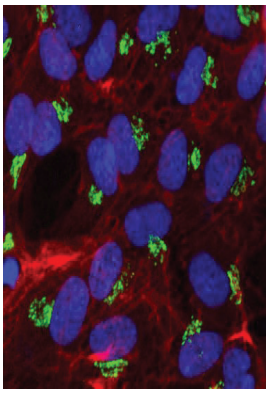
médecine/sciences 2015 ; 31 : 187-96

Chémobiologie (5)

Criblage phénotypique à haut contenu pour la chémobiologie et ses enjeux

Priscille Brodin[1], Elaine DelNery[2], Emmanuelle Soleilhac[3-5]

› Le *high content screening* (HCS), connu en France sous le terme de « criblage à haut contenu », est apparu au cours des années 1990 et représente une technologie innovante permettant l'évaluation des composés destinés à devenir des futurs médicaments. Le HCS fut initialement conçu pour élargir la compréhension du mécanisme d'action des composés en testant directement leur bioactivité sur des cellules cultivées en laboratoire. Au cours des dix dernières années, le champ d'utilisation du HCS s'est considérablement étendu au sein des laboratoires de recherche académiques. La technologie est ainsi devenue le point de départ de nombreux programmes d'études fonctionnelles visant à comprendre, pour une pathologie donnée, les cibles protéiques ou les grandes fonctions cellulaires sur lesquelles de futurs médicaments pourraient agir. Cette approche de criblage repose sur l'acquisition d'images en fluorescence de cellules marquées, ce qui génère d'énormes quantités de données qui doivent être traitées par analyse d'images automatisée, et stockées de manière appropriée. Le stockage, l'analyse et la visualisation des données associés à leur intégration et leur interprétation sont encore des défis omniprésents auxquels la méthodologie HCS doit faire face. ‹



[1] Inserm U1019, CNRS UMR8204, université de Lille-Nord de France, institut Pasteur de Lille, centre pour l'infection et l'immunité, 1, rue du professeur Calmette, 59000 Lille, France ;
[2] Institut Curie, centre de recherche, département de recherche translationnelle, 26, rue d'Ulm, 75005 Paris, France ;
[3] Université Grenoble Alpes, institut de recherches en technologies et sciences pour le vivant (iRTSV) - biologie à grande échelle (BGE), 38000 Grenoble, France ;
[4] CEA, iRTSV (Institut de recherches en technologies et sciences pour le vivant) - BGE (biologie à grande échelle) - criblages de molécules bioactives (CMBA), 38000 Grenoble, France ;
[5] Inserm, BGE, 38000 Grenoble, France.
emmanuelle.soleilhac@cea.fr
elaine.del.nery@curie.fr
priscille.brodin@inserm.fr

Contexte du développement du criblage phénotypique à haut contenu

Découvrir un nouveau médicament dans le contexte scientifique actuel est un processus laborieux et coûteux : les échecs apparaissent très souvent à des stades tardifs de la découverte d'un médicament (phases cliniques) et leurs taux sont relativement élevés. Ils sont principalement liés à des mécanismes d'adaptation des cellules cancéreuses, d'échappement aux traitements, à une importante toxicité ou à un manque d'efficacité [1]. La compréhension limitée de la fonction d'une cible moléculaire ou des mécanismes biologiques sous-jacents à l'évolution de la maladie traitée est probablement une des causes majeures d'échec des candidats-médicaments. Les cibles choisies peuvent, notamment, insuffisamment contribuer à la maladie, ou les effets secondaires des composés être plus importants que prévu [2]. De nouveaux champs d'investigation sont donc nécessaires afin de fournir des cibles biologiques fiables et validées pour une thérapie donnée, d'une part, et d'améliorer les phases précoces de recherche d'un médicament par de nouvelles méthodologies de criblage, d'autre part.

La révolution des approches « omiques » (génomique, transcriptomique, protéomique, métabolomique, etc.) fournit continuellement des informations précieuses pour juger de la pertinence des nouvelles cibles biologiques, générant une masse d'informations sur le rôle et les réseaux de régulation des gènes ou des protéines impliqués dans, ou directement responsables de, l'évolution d'une maladie. Parallèlement,

Cet article fait partie de la série « Chémobiologie » qui a débuté dans le n° 12, vol. 30, décembre 2014 (www.medecinesciences.org).

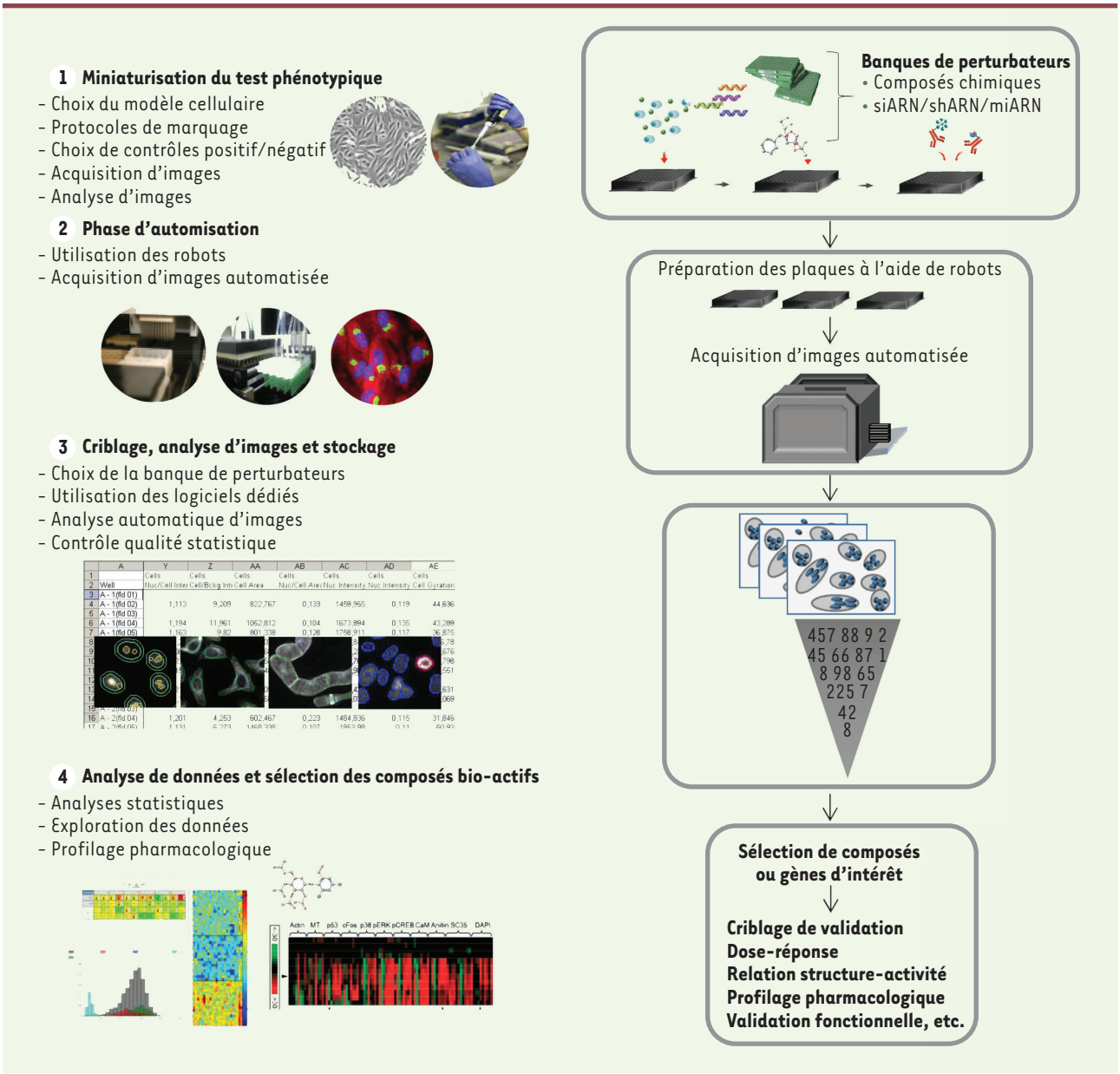


Figure 1. ***Processus de criblage HCS.*** Celui-ci se décompose généralement en quatre étapes : (***1***) la phase de développement et de miniaturisation d'un test phénotypique, avec détection des cellules par microscopie automatisée, (***2***) la phase d'automatisation, consistant en la robotisation de l'essai, (***3***) la phase de criblage, incluant l'acquisition et l'analyse d'images, ainsi que le stockage des données, et (***4***) une étape finale d'analyse statistique des données pour la sélection de modulateurs génétiques et chimiques, incluant les phases de validation secondaire et de caractérisation fonctionnelle (dose-réponse, étude de la relation structure-activité des composés, etc.).

les progrès apportés par l'automatisation de la microscopie optique, ainsi que le développement des sondes fluorescentes et des technologies de l'information (logiciels d'analyse d'images, bases de données, etc.) ont conduit à l'apparition du criblage à haut contenu (*high content screening*, HCS), qui repose sur l'acquisition et l'analyse automatisées d'images de cellules en culture. Autant dans l'industrie pharmaceutique qu'en milieu académique, ces nouvelles méthodes ont apporté une réelle valeur ajoutée aux criblages phénotypiques de composés chimiques (chémobiologie) ou de banques de petits ARN interférents [3-6].

En complément des tests biochimiques *in vitro*, cette puissante méthode de criblage permet d'améliorer la sélection et l'évaluation du profil d'activité de composés sur des modèles cellulaires. Ces modèles cellulaires peuvent être des cellules saines ou « contrôles », ou représentatives d'une condition pathologique. Grâce à la comparaison des profils d'activité obtenus à partir de cellules contrôles avec ceux obtenus à partir de cellules « pathologiques », il est désormais possible d'identifier

des composés prometteurs et d'avoir directement des informations sur leurs mécanismes d'action et, potentiellement, sur leurs cibles biologiques. De plus, les principaux avantages des tests phénotypiques sont de détecter simultanément la capacité des composés à atteindre leur cible dans un contexte cellulaire (biodisponibilité), ainsi que leur efficacité et leurs effets toxiques ou indésirables. Ainsi, le développement d'un composé dont la biodisponibilité n'est pas bonne, ou qui exerce une importante toxicité sur les cellules, n'atteindra pas les phases précliniques, ce qui permet d'épargner temps et argent [7, 8].

Outre son utilisation pour découvrir de nouveaux candidats-médicaments, le criblage phénotypique à haut contenu est également précieux en génétique fonctionnelle dans des projets de recherche fondamentale et biomédicale. En effet, il est primordial de déterminer la fonction des gènes et des protéines, aussi bien pour appréhender de façon toujours plus approfondie le fonctionnement du vivant, que pour déterminer la (ou les) cible(s) biologique(s) à l'origine de la genèse et de l'évolution des pathologies. En effet, l'analyse des conséquences de la perte de fonction d'un gène donné peut aboutir à la découverte de fonctions biologiques insoupçonnées et permettre de mieux comprendre son rôle dans les maladies humaines. Ainsi, de nombreuses études de génétique fonctionnelle sont menées à grande échelle grâce au criblage par HCS de banques de petits ARN interférents modulant l'expression des gènes humains (siARN et shARN, miARN), et grâce à l'évolution des méthodes de transfection à moyen et haut débit dans divers modèles cellulaires [4-6, 9, 10].

Qu'est-ce que le criblage à haut contenu ou HCS ?

Le terme criblage à haut contenu désigne une technologie innovante qui se fonde sur l'idée suivante : robotiser un microscope optique à fluorescence et le coupler à des logiciels informatiques capables de détecter et de quantifier l'ensemble des effets biologiques (physiologiques, pathologiques) survenant en réponse à un traitement, de manière reproductible et automatisée. Ce type de criblage s'appuie sur l'utilisation de tests phénotypiques miniaturisés réalisés en microplaques, qui permettent, grâce à des marquages fluorescents (anticorps, sondes fluorescentes, etc.), de visualiser les cellules, leur noyau et toutes structures subcellulaires telles que microtubules, actine, mitochondries, autophagosomes, endosomes, appareil de Golgi, etc. [3]. Le criblage HCS est une approche multiparamétrique, car il permet de détecter et corréler simultanément de nombreux paramètres à l'échelle d'une cellule unique, afin de comprendre la dynamique et l'hétérogénéité des réponses des populations cellulaires soumises à des perturbations telles que des traitements chimiques ou des extinctions géniques [6, 11-13].

Généralement lié aux activités des plates-formes de criblage à haut débit (HTS, *high throughput screening*), le HCS s'appuie souvent sur les outils robotiques de ces plates-formes. De plus, il utilise la puissance des outils issus du *data mining* (la fouille des données, qui regroupe l'analyse de données et les statistiques exploratoires) afin de comprendre et modéliser l'effet des composés testés. Les différentes phases d'un projet HCS sont détaillées dans la *Figure 1*.

Le test cellulaire et la détection des cellules par microscopie automatisée

Lors du développement et de la miniaturisation d'un test phénotypique, maîtriser l'utilisation des sondes fluorescentes est un élément essentiel [14]. Ces dernières années, de nombreux outils ont été développés, tels que des anticorps révélés directement ou indirectement par des fluorochromes, des sondes fluorescentes lipidiques ou des indicateurs d'ions. Les fluorophores utilisés actuellement sont plus brillants et plus stables, avec des spectres d'excitation et d'émission plus étroits qui permettent d'éviter le chevauchement des spectres de fluorophores contigus. Le marquage est aussi obtenu par couplage d'un gène recombinant à une protéine fluorescente, comme la protéine GFP (*green fluorescent protein*). Associés à la puissance des systèmes HCS (plusieurs jeux de filtres, miroirs optiques, détecteurs), ces développements permettent le « multiplexage »[1] du test grâce à l'utilisation simultanée de fluorophores distincts et à l'acquisition séquentielle des images. L'acquisition simultanée d'images est possible avec les systèmes HCS équipés de plusieurs caméras. Le multiplexage est un atout important du HCS car, selon la spécificité des marqueurs utilisés, il est possible de connaître l'effet d'un composé sur plusieurs cibles d'intérêt et événements cellulaires simultanément, tels que la morphologie, la perméabilité membranaire, l'activation d'une cascade de phosphorylation, la translocation d'une protéine ou d'un élément subcellulaire, ou encore la toxicité du composé. En fonction des caractéristiques de la sonde utilisée, chaque signal est corrélé, directement ou non, à une propriété physiologique [13, 15].

Lors du développement du test et pour l'étape ultérieure de criblage, il est crucial de choisir des contrôles bio-inactifs et bio-actifs. Généralement, dans le cas du criblage de composés chimiques, le contrôle bio-inactif est le diméthylsulfoxyde (DMSO), solvant universel des molécules, et qui n'induit pas de phénotype particulier à la concentration utilisée ; le contrôle bio-actif, quant à lui, sera un composé connu pour induire le phénotype recherché. Ces contrôles sont absolument nécessaires au suivi de la reproductibilité et de la robustesse statistique du test tout au long du criblage (contrôles internes pour chaque plaque, plaque contrôle quotidienne, etc.) ; ils permettent d'avoir un contrôle qualité du test, d'éliminer les faux-positifs et de sélectionner avec fiabilité les composés bio-actifs [16].

[1] Le multiplexage est une technique qui consiste à faire passer plusieurs informations à travers un seul support de transmission.

Appareil HCS, société	Objectifs possibles	Fluorophores et caméras possibles[§]	Modules optionnels[€]
ArrayScan XTI, Thermo Scientific	x 2,5*, x 5, x 10, x 20, x 40*, x 63*	Sept canaux, une caméra CCD	*Brightfield*, confocal, LCC, pipetage, roue pour filtres d'émission (FRET, Q*dots*), bras robotique et *stacker*
CellInsight CX5, Thermo Scientific	x 2*, x 4*, x 10, x 20	Cinq canaux, une caméra CCD	Bras robotique et *stacker*
INCell Analyzer 2200, GE-Healthcare	x 2*, x 4*, x 10, x 20, x 40*, x 60*, x 100*	Sept canaux, une caméra sCMOS	Lumière transmise, LCC, pipetage, déconvolution, restauration images 3-D, module FRET
INCell Analyzer 6000, GE-Healthcare Confocal	x 2*, x 4*, x 10, x 20, x 40*, x 60*	Quatre lasers, huit filtres Em, une caméra sCMOS	Lumière transmise, LCC, pipetage
Operetta, Perkin Elmer	x 2*, x 10*, x 20, x 40*, x 60*, x 100*	Huit canaux	*Brightfield*, LCC, confocal
Opera QEHS/Opera Phenix, Perkin Elmer Confocal	x 10*, x 20, x 40*, x 60*	Quatre lasers, module UV, quatre caméras CCD	Objectifs à immersion (eau), lumière transmise, LCC, pipetage, *fluorescence lifetime imaging* (FLIM), bras robotique et *stacker*
Cell Voyager CV7000, Yokogawa Confocal	x 4 à x 60	Quatre lasers, caméra sCMOS (max 3)	Objectif x 60 à immersion (eau), pipetage, LCC, bras robotique et *stacker*, lecteur code-à-barre
ImageXpress Micro XLS, Molecular Devices	x 1 à x 100	Cinq canaux, une caméra (CCD ou sCMOS)	Objectifs à immersion (huile), lumière transmise, confocal, LCC, pipetage, bras robotique et *stacker*
ImageXpress Ultra, Molecular Devices Confocal	x 4 à x 100	Deux à quatre lasers, une caméra sCMOS	Objectifs à immersion (huile), lumière transmise, confocal, LCC, pipetage, bras robotique et *stacker*
Scan^R, Olympus	x 4*, x 20, x 40*, x 60*, x 100*	Cinq canaux, une caméra CCD	Laser IR pour autofocus, LCC, bras robotique et *stacker*

Tableau I. ***Microscopes automatisés dédiés au HCS.*** Liste non exhaustive. Certains objectifs sont fournis en série sur les appareils ; les autres* sont optionnels (exemple : grossissements x 2,5, x 4 , x 5, x 40, x 60, x 100, grandes ouvertures numériques [*high* NA], objectifs à immersion [eau, huile]) ; [§] : les systèmes décrits ont tous la capacité d'acquérir des images de marquages classiques type Hoechst/DAPI (4',6-diamidino-2-phénylindole), FITC (*fluorescein isothiocyanate*)/A488, TRITC (*tetramethylrhodamine*), cyanin (3 et 5) et Texas Red pour certains (Em : émission). [€] : certains systèmes peuvent être équipés de filtres optionnels pour le FRET (*Förster resonance energy transfer*). LCC : *live cell chamber* (module de contrôle de température et CO_2) ; bras robotique et *stacker* : pour l'alimentation robotique et augmenter le débit d'acquisition et d'analyse ; sCMOS : *scientific complementary metal-oxide-semiconductor*.

Instruments utilisés pour le HCS

Les instruments robotiques dédiés aux criblages HTS peuvent être utilisés pour les tests phénotypiques utilisant le HCS : (1) lors de l'ensemencement des cellules en microplaques (au niveau de la plate-forme robotique si elle est placée sous flux laminaire, sinon sous hotte en salle de culture) ; (2) pour le formatage des banques de composés et leur ajout dans les plaques de test ; (3) pour l'élaboration du protocole expérimental (fixation, perméabilisation, étapes de marquage ou de lavage, etc.) ; et (4) pour l'automatisation de l'ajout des plaques dans le microscope automatisé au moment de l'acquisition d'images. Éléments spécifiques du HCS, les systèmes de microscopie automatisée ont été développés pour répondre à la fois aux besoins de rapidité et de robustesse du criblage à haut débit (automatisation) et aux exigences de

Logiciels d'analyse d'images	
Logiciels commercialisés avec les systèmes HCS	HCS Studio (Thermo Scientific), INCell Investigator (GE-Healthcare), Columbus-Acapella (Perkin Elmer), Harmony (Perkin Elmer), MetaXpress (Molecular Devices)
Logiciels libres (*open source*) développés pour le HCS	Cell Profiler, CellCognition, CellTracker
Autres logiciels commerciaux	Metamorph (Molecular Devices), Matlab (MathWorks), Volocity 3D (Perkin Elmer)
Autres logiciels libres (*open source*)	ImageJ, iCy, FijI
Logiciels d'analyse de données	
Logiciels commercialisés avec les systèmes HCS	HCS Explorer (Thermo Scientific), Columbus (Perkin Elmer), AcuityXpress (Molecular Devices)
Logiciels libres (*open source*) développés pour le HCS	Cell Profiler Analyst
Autres logiciels commerciaux	Tibco SpotFire (TIBCO), Pipeline Pilot (BIOVIA), Amadea (ISoft), GraphPad Prism (GraphPad Software), R (logiciel gratuit)
Gestion des données	
Solutions commercialisées avec les systèmes HCS	Store (Thermo Scientific), INCell Miner (GE-Healthcare), Columbus (Perkin Elmer), MDCStore (Molecular Devices)

Tableau II. ***Logiciels d'analyse automatisée d'images et d'analyse de données.*** Liste non exhaustive.

l'imagerie (images haute résolution, caméras). Les microscopes HCS sont équipés de systèmes d'autofocus instantané et de caméras CCD (*charge-coupled device*) ou sCMOS (*scientific complementary metal oxide semiconductor*) de haute définition, qui rendent la capture d'images extrêmement robuste et rapide. Une liste non exhaustive de systèmes dédiés au HCS est proposée dans le *Tableau I*. Ces appareils sont tous adaptés à l'acquisition d'images de tests cellulaires réalisés en microplaques, généralement de 96, 384 voire 1 536 puits. Ils sont entièrement motorisés (filtres excitation/émission, objectifs, platine porte-plaque) et permettent d'utiliser une large gamme de fluorophores et d'observer les cellules à différents grossissements. Afin de couvrir un large éventail de tests phénotypiques, les microscopes HCS peuvent pour la plupart être implémentés de modules optionnels, comme le module de contrôle de la température, du CO_2 et de l'humidité pour la réalisation de tests sur cellules vivantes, les modules de lumière transmise (*brightfield*, contraste de phase [DIC, *differential interference contrast*]), et le module confocal.

L'analyse d'images et l'extraction des informations phénotypiques

L'analyse d'images est une étape clé dans le HCS. Bien que complètement intégrée dans le processus HCS, cette étape est parfois spécifiquement appelée *high content analysis* (HCA). Il s'agit d'utiliser des fonctionnalités mathématiques de traitement d'images pour en extraire les informations qu'elles contiennent et générer ainsi de nouvelles connaissances sur le processus biologique étudié [17-20]. Grâce aux ordinateurs, l'analyse du contenu des images est automatisée. Il existe aujourd'hui des logiciels qui mettent en œuvre des algorithmes adaptés aux traitements du grand nombre d'images issues des cribles HCS. Ces logiciels ont des fonctionnalités variées et puissantes, pouvant être adaptées aux besoins spécifiques de chaque utilisateur [21, 22]. Une liste de solutions logicielles dédiées à l'analyse d'images, mais aussi au traitement des données HCS et à leur gestion, est présentée dans le *Tableau II*.

L'analyse d'images permettra, par exemple, de connaître le nombre de cellules, la distribution d'une protéine cible dans chaque cellule, les caractéristiques morphologiques (taille, forme, texture), l'identification des compartiments subcellulaires associée au calcul de leur diamètre ou de leur surface, ou encore le repérage du chevauchement entre différents objets (colocalisation). Toutes ces données sont accessibles pour la totalité de la population cellulaire, pour des sous-populations ou encore cellule par cellule [23]. Cette étape est illustrée dans la *Figure 2*.

Dans le cas de réponses cellulaires hétérogènes, un système de filtrage informatique des données rend possible le tri de cellules selon leur intensité de marquage, leur taille ou selon une combinaison de paramètres modélisés par un arbre décisionnel. Plus précisément, un arbre décisionnel est une combinaison de conditions qui permet de

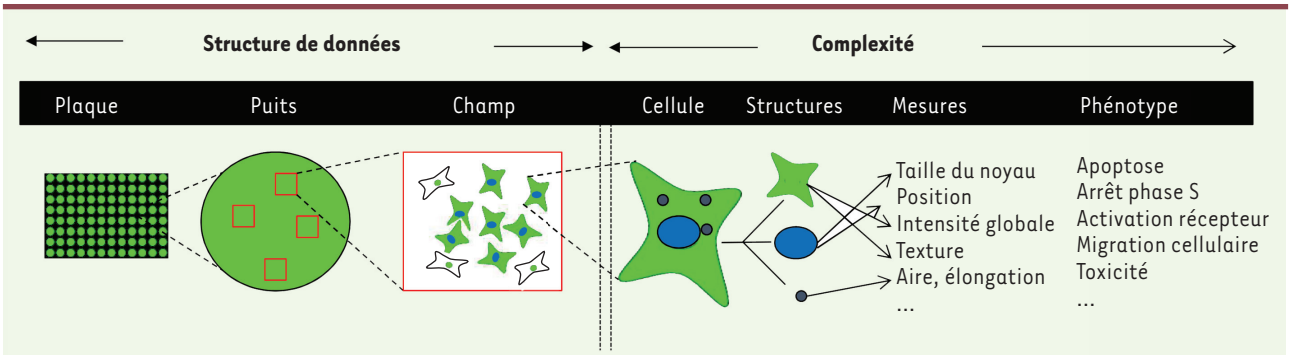


Figure 2. ***De l'acquisition en plaque à l'analyse du phénotype.*** Chaque image correspond à un champ d'un puits d'une microplaque. Le phénotype est extrait en segmentant chacune des cellules du champ puis, au sein de celles-ci, en déterminant la taille, la surface et la texture des structures intracellulaires, et en établissant une corrélation avec un processus biologique.

cibler l'étude sur une sous-population ayant des caractéristiques bien définies. Les critères de séparation sont déterminés par l'utilisateur : ainsi chaque cellule dont les caractéristiques correspondent à la définition établie est classée dans la catégorie adéquate. Un exemple de classification à partir d'un test de quantification de l'autophagie que nous avons développé (É. Soleilhac, données non publiées) est présenté dans la *Figure 3*. Cette analyse à l'échelle cellulaire est très performante lorsque l'on s'adresse à des sous-populations minoritaires, difficiles à identifier dans une analyse globale [24-26].

Certains logiciels d'analyse d'images peuvent être paramétrés pour la discrimination d'objets d'intérêt *via* l'apprentissage par classification supervisée ou *machine learning* [27-29]. La classification supervisée consiste à inférer, à partir d'un échantillon de données, à quelle classe appartient une observation, parmi un ensemble de classes connues. L'objectif est d'identifier les classes auxquelles appartiennent les cellules à partir d'éléments descriptifs (mesures, caractéristiques, *features*). Dans la classification supervisée, les classes sont connues et l'on dispose d'une population de cellules qui servira d'exemple pour chaque classe. La plupart des algorithmes d'apprentissage supervisés tentent donc de trouver un modèle (une fonction mathématique) qui explique le lien entre des données d'entrée (les cellules) et les classes de sortie (les phénotypes). En revanche, ces algorithmes ont l'inconvénient de nécessiter des connaissances préalables sur les phénotypes qui seront rencontrés lors du criblage ; or, un des atouts des criblages HCS est de permettre de découvrir des phénotypes nouveaux, non attendus. En utilisant la classification supervisée, le biologiste peut passer à côté de ces phénotypes et réduire ses chances de découvrir de nouveaux modes d'action.

Traitement et exploration des données HCS

La complexité de la recherche de modulateurs prometteurs réside dans le fait que les microscopes HCS génèrent énormément d'images. Le nombre d'images acquises lors du criblage phénotypique est très variable et dépend du format de la plaque utilisée (96 ou 384 puits), de la taille de la banque criblée, du nombre de fluorochromes utilisés, du nombre d'images/puits pour chaque marqueur, des sections optiques, etc. Une seule microplaque 384 puits peut aboutir à l'acquisition de plus de 6 000 images *(Figure 4)*. Dans ces conditions, un crible en triplicat de plusieurs milliers de composés produit 200 000 images et nécessite plus d'un To d'espace de stockage.

De plus, les données issues de criblages cellulaires sont particulièrement complexes à traiter, avec des variables comportant de multiples classes, des distributions non gaussiennes, des valeurs de paramètres pouvant varier sur plusieurs ordres de grandeurs. Afin d'extraire des résultats statistiquement fiables et robustes, des méthodes de normalisation développées pour le criblage HTS peuvent être utilisées [30-32]. Cependant, les données multiparamétriques des criblages HCS posent certaines difficultés supplémentaires, et l'extraction des paramètres pertinents et discriminants pour décrire un phénotype est souvent complexe. Le développement de méthodes de normalisation spécifiques aux criblages HCS demandera encore beaucoup d'efforts en bio-informatique et biostatistique.

Une méthode communément utilisée pour explorer les données HCS est le regroupement (*clustering*) des perturbateurs générant des phénotypes similaires dans des expériences données [33-35]. Les méthodes de regroupement sont innombrables et fortement dépendantes des choix de modélisation (calcul des distances, espace des paramètres, répartition des observations, etc.) ; si leur utilité théorique est évidente, il est difficile de proposer des modèles universels. La qualité informative des résultats obtenus dépendra grandement des phénotypes observés et de l'adéquation du modèle.

Le HCS en recherche académique

Le criblage cellulaire est aujourd'hui utilisé non seulement par les grandes sociétés pharmaceutiques,

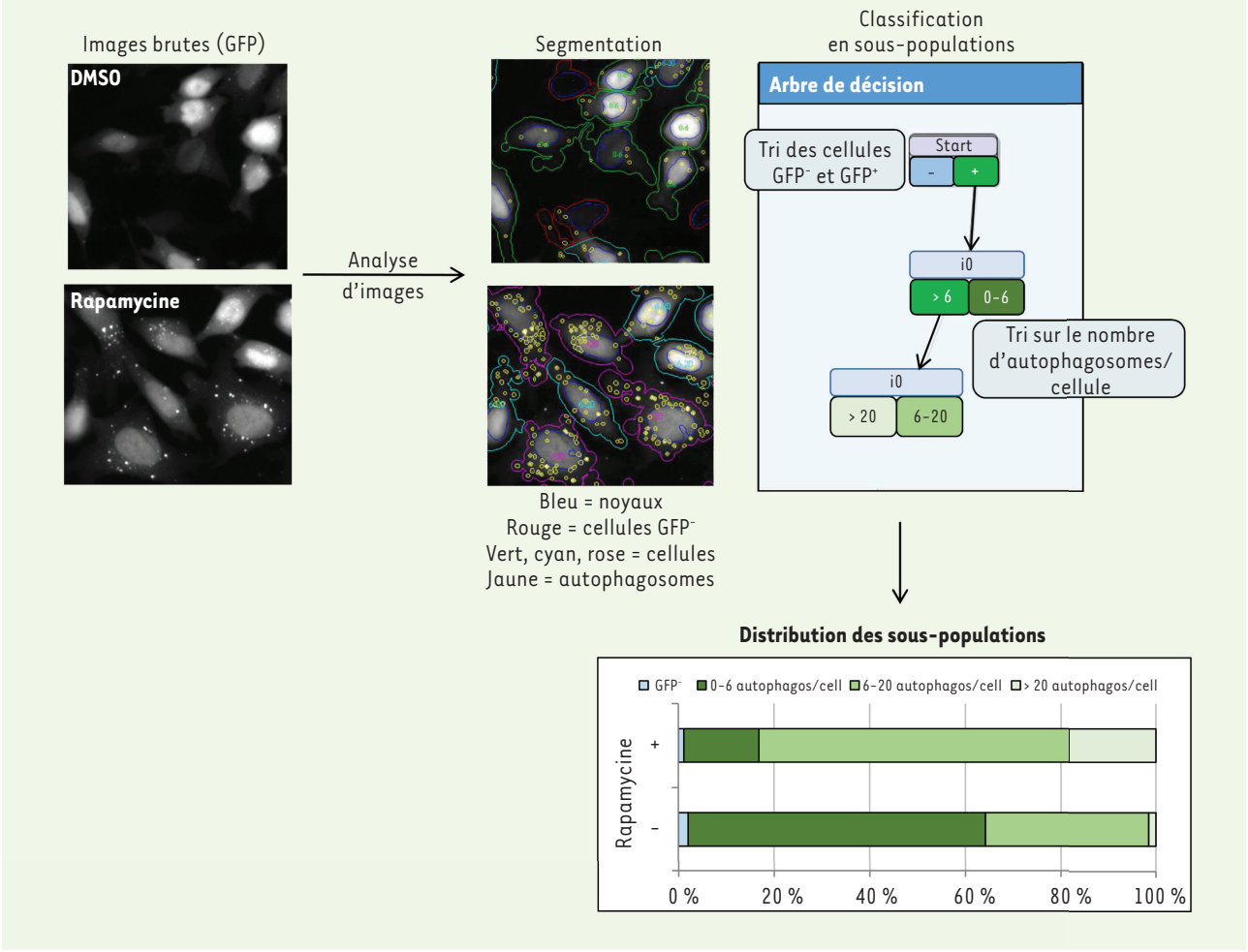


Figure 3. ***Arbre décisionnel appliqué à la quantification de l'autophagie.*** Images acquises en épifluorescence de cellules HeLa exprimant la protéine LC3 (*light chain* 3) (marqueur spécifique des autophagosomes) couplée à la GFP ; les cellules ont été traitées durant 2 h avec du DMSO (contrôle bio-inactif) ou de la rapamycine, inducteur de l'autophagie utilisé comme contrôle bio-actif. Les noyaux sont visualisés par marquage de l'ADN avec du Hoechst ; les autophagosomes visibles dans la cellule sous forme de spots brillants sont détectés dans le canal de la GFP. En présence de rapamycine, il y a une nette augmentation du nombre d'autophagosomes. L'analyse d'images permet, entre autres, d'accéder à l'intensité de fluorescence dans les différents canaux et au nombre d'autophagosomes par cellule. Un arbre de décision a été développé afin de trier les cellules en différentes sous-populations en fonction de leur intensité de fluorescence en GFP et de leur contenu en autophagosomes. L'histogramme présenté montre que, lors d'un traitement à la rapamycine, il y a une nette augmentation des sous-classes de cellules contenant plus de six vésicules, au détriment de la classe « 0-6 » (E. Soleilhac *et al.*, données non publiées). Autophagos : autophagosome.

mais de plus en plus par de petites entreprises de biotechnologie et des laboratoires de recherche académiques. Depuis peu, des associations à but non lucratif soutiennent financièrement des projets de criblage d'intérêt thérapeutique et orientent les équipes de recherche vers des plates-formes académiques réunissant équipements, compétences et expertises dans le domaine du criblage. Il faut noter que, depuis une dizaine d'années, les universités, en particulier aux États-Unis, et les grands pôles de recherche européens sont devenus des acteurs essentiels dans le développement des technologies de criblage automatisé à haut débit ou à haut contenu (http://eu-openscreen.eu/). En France, des plates-formes académiques de criblage sont structurées dans un consortium (GDR CNRS 3056 ChemBioScreen, http://www.chembioscreen.fr/) [3]. Les structures académiques de criblage ont un rôle important à jouer, non seulement dans la découverte des composés potentiellement thérapeutiques, mais également dans la découverte des composés qui serviront d'outils de recherche en sciences du vivant pour la description des fonctions biologiques essentielles. Le monde académique a donc tous les outils en mains pour émettre et vérifier des hypothèses sur la fonction des gènes, en couplant des données cellulaires issues du HCS avec des études en génomique fonctionnelle [4]. L'enjeu du HCS réside

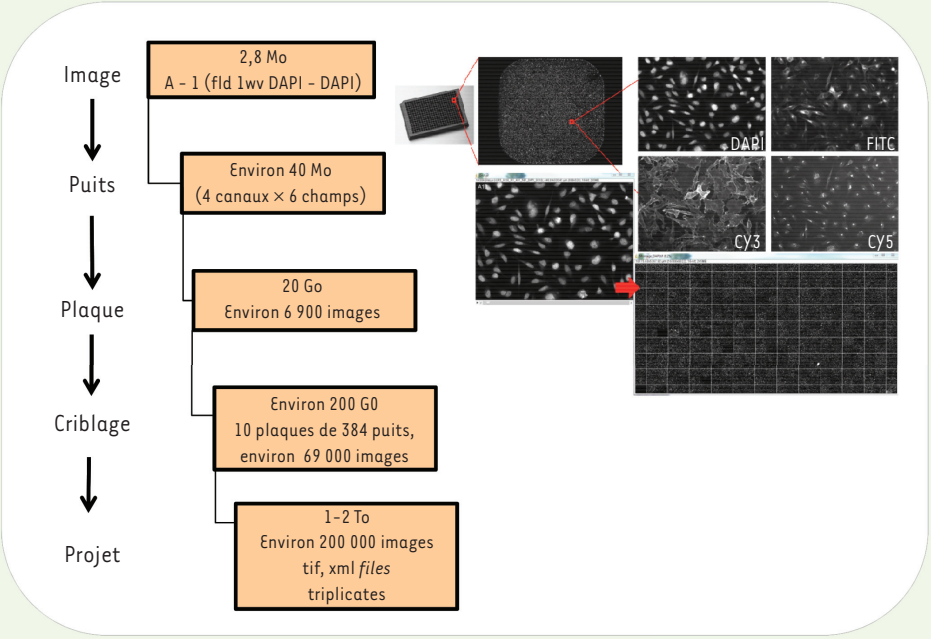


Figure 4. ***Stockage informatique nécessaire pour la gestion des images d'un projet de criblage.*** Acquisition d'images dans une microplaque 384 puits, avec quatre canaux de fluorescence et six images (champs) pour chaque canal et par puits. Dans ces conditions, 6 900 images seront acquises. Pour un criblage de 10 plaques, 69 000 images seront générées, représentant 200 Go de données. Un projet réalisé en triplicat aboutira au stockage de 200 000 images, soit 1 à 2 To de données de criblage. Mo : méga octet ; Go : giga octet ; To : téra octet.

finalement aussi, en recherche biomédicale, dans la caractérisation des nouvelles cibles potentiellement thérapeutiques et des voies de régulation dans lesquelles elles sont impliquées.

Conclusion

Plutôt que d'écrire une revue exhaustive sur le HCS, nous avons souhaité dresser ici un portrait de cette activité de plus en plus présente sur les plates-formes académiques de criblage. En effet, à travers des approches systématiques, le HCS est aujourd'hui accessible aux communautés scientifique et médicale pour (1) comprendre les mécanismes biologiques fondamentaux en effectuant des observations directes dans une cellule génétiquement modifiée ou traitée par une substance chimique, (2) identifier des gènes responsables de pathologies humaines, (3) définir de nouvelles cibles thérapeutiques, et, enfin, (4) isoler des petites molécules, qui pourraient être un jour des « têtes de série » de futurs médicaments. Cette technologie représente un véritable atout pour l'identification et la compréhension de l'activité de perturbateurs (chimiques, génétiques, ou autres) dans le contexte cellulaire. Le développement de protocoles d'analyse d'images, simples comme complexes, permet d'en extraire des informations biologiques pertinentes pour décrire l'activité des composés testés. Il faut souligner ici l'importance de l'œil critique et expert du biologiste et du chimiste dans l'analyse du phénotype cellulaire et du composé bio-actif, dans le cas d'un projet en chémobiologie. Approche multidisciplinaire, le HCS fait appel également à la rigueur du bio-informaticien pour l'extraction de l'information et l'analyse des données, et aux compétences du biostatisticien pour le développement de méthodes statistiques adaptées au traitement de grandes masses de données. Cette approche est aussi bien utilisée pour des criblages de banques de composés (de 1 000 à plusieurs milliers de composés, petit à haut débit de criblage), que pour les phases de validation et caractérisation des composés bio-actifs sélectionnés à l'issue d'un criblage HTS par exemple (étude dose-réponse, études de relation structure-activité, étude de l'effet du composé sur différents phénotypes cellulaires, etc.). Pour aller encore plus loin dans les possibilités offertes par cette technologie, des équipes l'appliquent à des organismes modèles, tels que *T. gondii*, *C. elegans*, le *zebrafish*, ou *Drosophila melanogaster* [5, 36-39]. De plus, afin de travailler sur des modèles d'étude *in vivo* plus proches de la réalité, de nombreux efforts sont réalisés dans la mise en place de modèles en trois dimensions (3D) comme les cultures de sphéroïdes [40] ou des cocultures en 3D [41]. Ces modèles sont plus représentatifs du vivant que des cellules cultivées en 2D, tout en conservant les caractéristiques des cellules humaines [42, 43].

De plus, sont apparues récemment des technologies de *gene editing* [47], qui permettent de manipuler le génome de cellules somatiques ; plus particulièrement, le système CRISPR-Cas9 (*clustered regularly interspaced short palindromic repeat-CRISPR-associated nuclease* 9) ouvre de nouvelles possibilités de cribles génétiques, plus puissants que les cribles basés sur l'interférence d'ARN [48, 49] (➜). En effet, l'emploi de lentivirus pour introduire l'enzyme Cas9 et des sgARN (*single guide RNA*) permet de réaliser à grande échelle des *knock-out* (invalidation de gènes) plutôt que des *knock-down* générés avec des shARN (inhibition partielle de l'expression génique) [44, 45]. Le développement d'outils innovants de criblage cellulaire, couplé à l'utilisation de shARN et CRISPR, est aujourd'hui à l'avant-garde dans le domaine prometteur de la biologie cellulaire à haut débit, dans ses applications cognitives et thérapeutiques.

(➜) Voir les Nouvelles de H. Gilgenkrantz, *m/s* n° 12, décembre 2014, page 1066, et de C. Asensio, page 137 de ce numéro

Enfin, il est vrai que les méthodes de criblage ont considérablement modifié les délais d'obtention des données biologiques, mais elles ont surtout bouleversé l'échelle des données analysables. La phase d'analyse de données est particulièrement importante dans un projet HCS. Elle est cruciale pour la qualité des résultats et consomme typiquement de l'ordre de 70 % du temps d'une étude HCS. Le développement et l'intégration d'outils bio-informatiques puissants et de méthodes statistiques robustes, adaptés à l'analyse HCS, ainsi que le croisement des données avec celles issues d'autres sources d'information (génomique, protéomique, etc.) sont en train d'émerger, et révolutionneront les outils commerciaux et universitaires actuellement utilisés par la communauté du HTS/HCS [46]. Les liens fonctionnels établis seront une source de nouvelles connaissances qui permettront d'élucider les mécanismes moléculaires d'une maladie, de prédire des résultats thérapeutiques et, peut-être, de collaborer au développement d'une médecine personnalisée. Bien plus qu'une technique, le criblage est un domaine essentiel à maîtriser dans une ère marquée par une nouvelle approche des questions scientifiques : l'ère de la biologie des grands nombres. ◊

SUMMARY

High content screening in chemical biology: overview and main challenges

The last two decades have seen the development of high content screening (HCS) methodology and its adaptation for the evaluation of small molecules as drug candidates or their use as chemical tools for research purpose. HCS was initially set-up for the understanding of the mechanism of action of compounds by testing them on cell based-assays for pharmacological and toxicological studies. Since the last decade, the use of HCS has been extended to academic research laboratories and this technology has become the starting point for numerous projects aiming at the identification of molecular targets and cellular pathways for a given disease on which novel type of drugs could act. This screening approach relies on image capture of fluorescently labeled cells therefore generating a large amount of data that must be handled by appropriate automated image analysis methods and storage instrumentation. These latter in addition to the integration and data sharing are current challenges that HCS must still tackle. ◊

LIENS D'INTÉRÊT

Les auteurs déclarent n'avoir aucun lien d'intérêt concernant les données publiées dans cet article.

RÉFÉRENCES

1. Carragher NO, Brunton VG, Frame MC. Combining imaging and pathway profiling: an alternative approach to cancer drug discovery. *Drug Discov Today* 2012 ; 17 : 203-14.
2. Kola I, Landis J. Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov* 2004 ; 3 : 711-5.
3. Prudent R, Soleilhac É, Barette C, *et al.* Les criblages phénotypiques ou comment faire d'une pierre deux coups : découvrir la cible et la molécule pharmacologique capable de la réguler. *Med Sci (Paris)* 2013 ; 29 : 897-905.
4. Perrimon N, Friedman A, Mathey-Prevot B, Eggert US. Drug-target identification in *Drosophila* cells: combining high-throughout RNAi and small-molecule screens. *Drug Discov Today* 2007 ; 12 : 28-33.
5. Perrimon N, Mathey-Prevot B. Applications of high-throughput RNA interference screens to problems in cell and developmental biology. *Genetics* 2007 ; 175 : 7-16.
6. Rines DR, Tu B, Miraglia L, *et al.* High-content screening of functional genomic libraries. *Methods Enzymol* 2006 ; 414 : 530-65.
7. Zock JM. Applications of high content screening in life science research. *Comb Chem High Throughput Screen* 2009 ; 12 : 870-6.
8. Mouchet EH, Simpson PB. High-content assays in oncology drug discovery: opportunities and challenges. *IDrugs* 2008 ; 11 : 422-7.
9. Berns K, Hijmans EM, Mullenders J, *et al.* A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* 2004 ; 428 : 431-7.
10. Ganesan AK, Ho H, Bodemann B, *et al.* Genome-wide siRNA-based functional genomics of pigmentation identifies novel genes and pathways that impact melanogenesis in human cells. *PLoS Genet* 2008 ; 4 : e1000298.
11. Eggert US, Mitchison TJ. Small molecule screening by imaging. *Curr Opin Chem Biol* 2006 ; 10 : 232-7.
12. Korn K, Krausz E. Cell-based high-content screening of small-molecule libraries. *Curr Opin Chem Biol* 2007 ; 11 : 503-10.
13. Perlman ZE, Slack MD, Feng Y, *et al.* Multidimensional drug profiling by automated microscopy. *Science* 2004 ; 306 : 1194-8.
14. Giuliano KA. Optimizing the integration of immunoreagents and fluorescent probes for multiplexed high content screening assays. *Methods Mol Biol* 2007 ; 356 : 189-93.
15. Young DW, Bender A, Hoyt J, *et al.* Integrating high-content screening and ligand-target prediction to identify mechanism of action. *Nat Chem Biol* 2008 ; 4 : 59-68.
16. Maréchal E, Roy S, Lafanechère L. *Chemogénomique : des petites molécules pour explorer le vivant. Une introduction à l'usage des biologistes, chimistes et informaticiens*. Paris : EDP Sciences, 2007 : 258 p.
17. Walter T, Held M, Neumann B, *et al.* Automatic identification and clustering of chromosome phenotypes in a genome wide RNAi screen by time-lapse imaging. *J Struct Biol* 2010 ; 170 : 1-9.
18. Fenistein D, Lenseigne B, Christophe T, *et al.* A fast, fully automated cell segmentation algorithm for high-throughput and high-content screening. *Cytometry A* 2008 ; 73 : 958-64.
19. Shariff A, Kangas J, Coelho LP, *et al.* Automated image analysis for high-content screening and analysis. *J Biomol Screen* 2010 ; 15 : 726-34.
20. Horvath P, Wild T, Kutay U, Csucs G. Machine learning improves the precision and robustness of high-content screens: using nonlinear multiparametric methods to analyze screening results. *J Biomol Screen* 2011 ; 16 : 1059-67.
21. Carpenter AE, Jones TR, Lamprecht MR, *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol* 2006 ; 7 : R100.
22. Walter T, Shattuck DW, Baldock R, *et al.* Visualization of image data from cells to organisms. *Nat Methods* 2010 ; 7 : S26-41.
23. Soleilhac E, Nadon R, Lafanechere L. High-content screening for the discovery of pharmacological compounds: advantages, challenges and potential benefits of recent technological developments. *Expert Opin Drug Discov* 2010 ; 5 : 135-44.

RÉFÉRENCES

24. Loo LH, Lin HJ, Steininger RJ 3rd, *et al.* An approach for extensibly profiling the molecular states of cellular subpopulations. *Nat Methods* 2009 ; 6 : 759-65.
25. Loo LH, Wu LF, Altschuler SJ. Image-based multivariate profiling of drug responses from single cells. *Nat Methods* 2007 ; 4 : 445-53.
26. Neumann B, Held M, Liebel U, *et al.* High-throughput RNAi screening by time-lapse imaging of live human cells. *Nat Methods* 2006 ; 3 : 385-90.
27. Hamilton NA, Pantelic RS, Hanson K, Teasdale RD. Fast automated cell phenotype image classification. *BMC Bioinformatics* 2007 ; 8 : 110.
28. Jones TR, Carpenter AE, Lamprecht MR, *et al.* Scoring diverse cellular morphologies in image-based screens with iterative feedback and machine learning. *Proc Natl Acad Sci USA* 2009 ; 106 : 1826-31.
29. Smith K, Horvath P. Active learning strategies for phenotypic profiling of high-content Screens. *J Biomol Screen* 2014 ; 19 : 685-95.
30. Brideau C, Gunter B, Pikounis B, Liaw A. Improved statistical methods for hit selection in high-throughput screening. *J Biomol Screen* 2003 ; 8 : 634-47.
31. Kevorkov D, Makarenkov V. Statistical analysis of systematic errors in high-throughput screening. *J Biomol Screen* 2005 ; 10 : 557-67.
32. Malo N, Hanley JA, Cerquozzi S, *et al.* Statistical practice in high-throughput screening data analysis. *Nat Biotechnol* 2006 ; 24 : 167-75.
33. Gagarin A, Makarenkov V, Zentilli P. Using clustering techniques to improve hit selection in high-throughput screening. *J Biomol Screen* 2006 ; 11 : 903-14.
34. Makarenkov V, Zentilli P, Kevorkov D, *et al.* An efficient method for the detection and elimination of systematic error in high-throughput screening. *Bioinformatics* 2007 ; 23 : 1648-57.
35. Kozak K, Agrawal A, Machuy N, Csucs G. Data mining techniques in high content screening: A survey. *J Comput Sci Syst Biol* 2009 ; 2 : 219-39.
36. Carey KL, Westwood NJ, Mitchison TJ, Ward GE. A small-molecule approach to studying invasive mechanisms of *Toxoplasma gondii*. *Proc Natl Acad Sci USA* 2004 ; 101 : 7433-8.
37. Peterson RT, Fishman MC. Discovery and use of small molecules for probing biological processes in zebrafish. *Methods Cell Biol* 2004 ; 76 : 569-91.
38. Simpson KJ, Davis GM, Boag PR. Comparative high-throughput RNAi screening methodologies in *C. elegans* and mammalian cells. *New Biotechnol* 2012 ; 29 : 459-70.
39. Wählby C, Kamentsky L, Liu ZH, *et al.* An image analysis toolbox for high-throughput *C. elegans* assays. *Nat Methods* 2012 ; 9 : 666-70.
40. LaBarbera DV, Reid BG, Yoo BH. The multicellular tumor spheroid model for high-throughput cancer drug discovery. *Expert Opin Drug Discov* 2012 ; 7 : 819-30.
41. Li Q, Chen C, Kapadia A, *et al.* 3D models of epithelial-mesenchymal transition in breast cancer metastasis: high-throughput screening assay development, validation, and pilot screen. *J Biomol Screen* 2011 ; 16 : 141-54.
42. Breslin S, O'Driscoll L. Three-dimensional cell culture: the missing link in drug discovery. *Drug Discov Today* 2013 ; 18 : 240-9.
43. Rimann M, Graf-Hausner U. Synthetic 3D multicellular systems for drug development. *Curr Opin Biotechnol* 2012 ; 23 : 803-9.
44. Wang T, Wei JJ, Sabatini DM, Lander ES. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 2014 ; 343 : 80-4.
45. Shalem O, Sanjana NE, Hartenian E, *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 2014 ; 343 : 84-7.
46. De Souza A, Bittker JA, Lahr DL, *et al.* An overview of the challenges in designing, Integrating, and delivering BARD: a public chemical-biology resource and query portal for multiple organizations, locations, and disciplines. *J Biomol Screen* 2014 ; 19 : 614-27.
47. Dupret B, Angrand PO. L'ingénierie des génomes par les TALEN. *Med Sci (Paris)* 2014 ; 30 : 186-93.
48. Gilgenkrantz H. La révolution des CRISPR est en marche. *Med Sci (Paris)* 2014 ; 30 : 1066-9.
49. Asensio C. Application de la méthode Cas9/SRISPR à l'étude de la fonction synaptique. *Med Sci (Paris)* 2015 ; 31 : 137-8.

TIRÉS À PART

E. Soleilhac